aws

Guia de gerenciamento

Amazon EMR

Amazon EMR: Guia de gerenciamento

Table of Contents

O que é o Amazon EMR?

O Amazon EMR (anteriormente chamado de Amazon Elastic MapReduce) é uma plataforma de cluster gerenciada que simplifica a execução de estruturas de big data, como Apache Hadoop e Apache Spark, para processar e analisar grandes quantidades de dados. AWS Ao usar essas estruturas e projetos de código aberto relacionados, é possível processar dados para finalidades analíticas e workloads de inteligência de negócios. O Amazon EMR também permite transformar e mover grandes quantidades de dados de e para outros armazenamentos de dados e bancos de dados da AWS , como o Amazon Simple Storage Service (Amazon S3) e o Amazon DynamoDB.

Se você for um usuário iniciante do Amazon EMR, recomendamos começar com a leitura do seguinte material, além desta seção:

- Amazon EMR: esta página do serviço fornece destaques, detalhes do produto e informações sobre preços para o Amazon EMR.
- Tutorial: Começando com a Amazon EMR: este tutorial permite que você comece a usar o Amazon EMR rapidamente.

Nesta seção

Visão geral do Amazon EMR

Este tópico fornece uma visão geral dos clusters do Amazon EMR, incluindo como enviar trabalho para um cluster, como esses dados são processados e quais são os diversos estados pelos quais o cluster passa durante o processamento.

Neste tópico

Noções básicas sobre clusters e nós

O componente central do Amazon EMR é o cluster. Um cluster é um conjunto de instâncias do Amazon Elastic Compute Cloud (Amazon EC2). Cada instância do cluster é chamada de nó. Cada nó tem um perfil dentro do cluster, conhecido como tipo de nó. O Amazon EMR também instala diferentes componentes de software em cada tipo de nó, atribuindo a cada nó um perfil em uma aplicação distribuída, como o Apache Hadoop.

Os tipos de nós no Amazon EMR são os seguintes:

- Nó primário: um nó que gerencia o cluster ao executar componentes de software para coordenar a distribuição de dados e de tarefas entre outros nós para processamento. O nó primário rastreia o status das tarefas e monitora a integridade do cluster. Cada cluster tem um nó primário e é possível criar um cluster de nó único apenas com o nó primário.
- Nó core: nó com componentes de software que executam tarefas e armazenam dados no Hadoop Distributed File System (HDFS) do cluster. Clusters de vários nós têm pelo menos um nó core.
- Nó de tarefa: nó com componentes de software que apenas executa tarefas e não armazena dados no HDFS. Nós de tarefa são opcionais.

Envio de trabalhos para um cluster

Ao executar um cluster no Amazon EMR, você tem diversas opções sobre como especificar o trabalho que precisa ser feito.

- Forneça a definição completa do trabalho a ser feito nas funções que você especifica como etapas ao criar um cluster. Isto é normalmente feito para clusters que processam uma quantidade definida de dados e, em seguida, são encerrados quando o processamento é concluído.
- Crie um cluster de longa duração e use o console do Amazon EMR, a API do Amazon EMR ou AWS CLI o para enviar etapas, que podem conter um ou mais trabalhos. Para ter mais informações, consulte Enviar trabalhos a um cluster.
- Crie um cluster, conecte-se ao nó primário e a outros nós conforme necessário usando o SSH e use as interfaces que as aplicações instaladas fornecem para executar tarefas e enviar consultas, com scripts ou de forma interativa. Para obter mais informações, consulte o Guia de versão do Amazon EMR.

Processar dados

Ao executar o cluster, você escolhe as estruturas e os aplicativos a serem instalados para as suas necessidades de processamento de dados. Para processar dados no cluster do Amazon EMR, é possível enviar trabalhos ou consultas diretamente para aplicações instaladas ou executar etapas no cluster.

Envio de trabalhos diretamente para aplicações

Você pode enviar trabalhos e interagir diretamente com os softwares instalados no cluster do Amazon EMR. Para fazer isso, normalmente você se conecta ao nó primário usando uma conexão segura e acessa as interfaces e ferramentas que estão disponíveis para os softwares que são executados diretamente em seu cluster. Para ter mais informações, consulte Conectar-se a um cluster.

Execução de etapas para processar dados

Você pode enviar uma ou mais etapas ordenadas para um cluster do Amazon EMR. Cada etapa é uma unidade de trabalho que contém instruções para manipular dados para processamento pelos softwares instalados no cluster.

Veja a seguir um exemplo de processo utilizando quatro etapas:

1. Enviar um conjunto de dados de entrada para processamento.
2. Processar a saída da primeira etapa usando um programa Pig.
3. Processar um segundo conjunto de dados de entrada usando um programa Hive.
4. Gravar um conjunto de dados de saída.

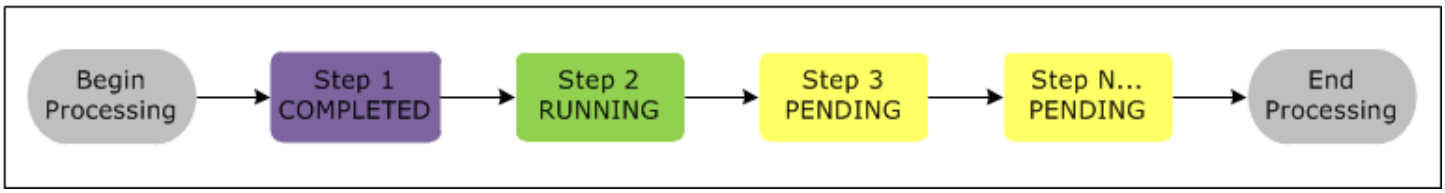
Geralmente, quando você processa dados no Amazon EMR, a entrada corresponde a dados armazenados como arquivos no sistema de arquivos subjacente escolhido, como o Amazon S3 ou o HDFS. Esses dados passam de uma etapa para a próxima na sequência de processamento. A etapa final grava os dados de saída em um local especificado, como um bucket do Amazon S3.

As etapas são executadas na seguinte sequência:

1. É enviada uma solicitação para iniciar as etapas de processamento.
2. O estado de todas as etapas é definido como PENDING (Pendente).

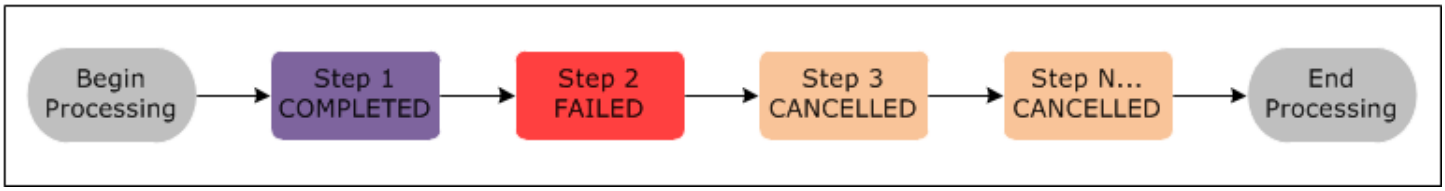
3. Quando a primeira etapa da sequência é iniciada, seu estado muda para RUNNING (Em execução). As outras etapas permanecem no estado PENDING (Pendente).
4. Após a conclusão da primeira etapa, seu estado muda para COMPLETED (Concluído).
5. A próxima etapa da sequência é iniciada, e seu estado muda para RUNNING (Em execução). Após a conclusão, seu estado muda para COMPLETED (Concluído).
6. Esse padrão repete-se para cada etapa, até todas elas estejam concluídas e o processamento seja encerrado.

O diagrama a seguir representa a sequência de etapas e mudança de estado para as etapas conforme elas são processadas.



Se uma etapa falhar durante o processamento, seu estado será alterado para FAILED. Você pode determinar o que acontece a seguir em cada etapa. Por padrão, todas as etapas restantes na sequência são definidas como CANCELLED e não são executadas se uma etapa anterior falhar. Você também pode optar por ignorar a falha e permitir que as etapas restantes continuem ou encerrem o cluster imediatamente.

O diagrama a seguir representa a sequência de etapas e a mudança de estado padrão quando uma etapa falha durante o processamento.



Noções básicas sobre o ciclo de vida do cluster

Um cluster do Amazon EMR com êxito segue este processo:

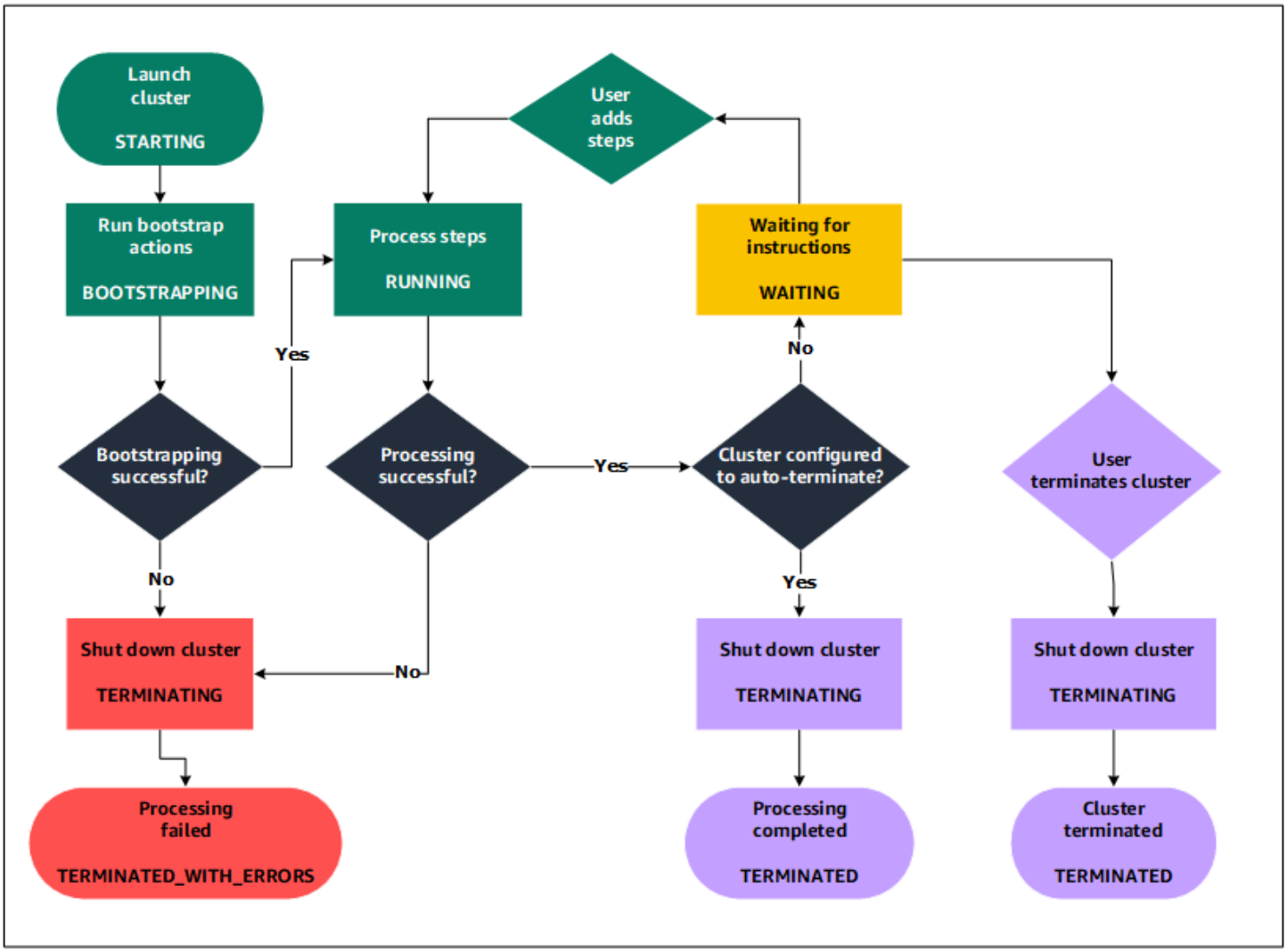
1. Primeiro, o Amazon EMR provisiona instâncias do EC2 no cluster para cada instância, de acordo com as suas especificações. Para ter mais informações, consulte Configurar o hardware e as redes do cluster. Para todas as instâncias, o Amazon EMR usa a AMI padrão para o Amazon EMR ou uma AMI personalizada do Amazon Linux especificada por você. Para ter mais

informações, consulte Usando um personalizado AMI. Durante essa fase, o estado do cluster é `STARTING`.

2. O Amazon EMR executa ações de bootstrap especificadas em cada instância. Você pode usar as ações de bootstrap para instalar aplicativos personalizados e executar as personalizações necessárias. Para ter mais informações, consulte Criar ações de bootstrap para instalar softwares adicionais. Durante essa fase, o estado do cluster é `BOOTSTRAPPING`.
3. O Amazon EMR instala as aplicações nativas especificadas ao criar o cluster, como o Hive, o Hadoop, o Spark e outros.
4. Depois que as ações de bootstrap forem concluídas com êxito e que os aplicativos nativos forem instalados, o estado do cluster será `RUNNING`. Nesse momento, você pode se conectar às instâncias de cluster, e o cluster executará de maneira sequencial todas as etapas que você especificou ao criar o cluster. Você pode enviar etapas adicionais, que serão executadas depois que as etapas anteriores forem concluídas. Para ter mais informações, consulte Enviar trabalhos a um cluster.
5. Depois que as etapas forem executadas com êxito, o cluster entrará no estado `WAITING`. Se um cluster estiver configurado para o encerramento automático após a conclusão da última etapa, ele entrará em um estado `TERMINATING` e, em seguida, no estado `TERMINATED`. Se o cluster estiver configurado para aguardar, você deverá encerrá-lo manualmente quando não precisar mais dele. Após encerrar manualmente o cluster, ele entrará no estado `TERMINATING` e, em seguida, no estado `TERMINATED`.

Uma falha durante o ciclo de vida do cluster faz com que o Amazon EMR encerre o cluster e todas as suas instâncias, a menos que você habilite a proteção contra encerramento. Se um cluster for encerrado devido a uma falha, todos os dados armazenados no cluster serão excluídos, e o estado do cluster será definido como `TERMINATED_WITH_ERRORS`. Se você tiver habilitado a proteção contra encerramento, você poderá recuperar os dados do seu cluster e, em seguida, remover a proteção contra encerramento e encerrá-lo. Para ter mais informações, consulte Usar a proteção contra término.

O diagrama a seguir representa o ciclo de vida de um cluster e como cada estágio desse ciclo é mapeado para um estado de cluster específico.



Benefícios do uso do Amazon EMR

Há muitos benefícios em usar o Amazon EMR. Esta seção fornece uma visão geral desses benefícios, além de links para informações adicionais para ajudá-lo a explorar ainda mais.

Tópicos

- Redução de custos
- AWS integração
- Implantação
- Escalabilidade e flexibilidade
- Confiabilidade
- Segurança

- Monitoramento
- Interfaces de gerenciamento

Redução de custos

Os preços do Amazon EMR dependem do tipo de instância e do número de instâncias do Amazon EC2 implantadas, bem como da região em que o seu cluster é iniciado. A definição de preço sob demanda oferece tarifas baixas, mas você pode reduzir os custos ainda mais comprando instâncias reservadas ou instâncias spot. As instâncias spot podem oferecer economias significativas. Em alguns casos, até um décimo dos preços sob demanda.

> **Note**
>
> Se você usar o Amazon S3, o Amazon Kinesis ou o DynamoDB com o cluster do EMR, haverá cobranças adicionais para os serviços faturados separadamente do uso do Amazon EMR.

> **Note**
>
> Ao configurar um cluster do Amazon EMR em uma sub-rede privada, recomendamos configurar também endpoints da VPC para o Amazon S3. Se o cluster do EMR estiver em uma sub-rede privada sem endpoints da VPC para o Amazon S3, você incorrerá em cobranças adicionais de gateway NAT associadas ao tráfego do S3, pois o tráfego entre o cluster do EMR e o S3 não permanecerá na VPC.

Para obter mais informações sobre as opções e os detalhes dos preços, consulte Preço do Amazon EMR.

AWS integração

O Amazon EMR se integra a outros AWS serviços para fornecer recursos e funcionalidades relacionados à rede, armazenamento, segurança, etc., para seu cluster. A lista a seguir fornece vários exemplos dessa integração:

- Amazon EC2 para as instâncias que compõem os nós do cluster.

- Amazon Virtual Private Cloud (Amazon VPC) para configurar a rede virtual na qual você inicia as instâncias.
- Amazon S3 para armazenar dados de entrada e de saída.
- Amazon CloudWatch monitorará o desempenho do cluster e configurará alarmes
- AWS Identity and Access Management (IAM) para configurar permissões
- AWS CloudTrail para auditar solicitações feitas ao serviço
- AWS Data Pipeline para programar e iniciar seus clusters
- AWS Lake Formation para descobrir, catalogar e proteger dados em um data lake do Amazon S3

Implantação

O cluster do EMR consiste de instâncias do EC2, que realizam o trabalho que você envia ao seu cluster. Ao executar o seu cluster, o Amazon EMR configura as instâncias com as aplicações que você escolher, como Apache Hadoop ou Spark. Escolha o tamanho de instância e o tipo que melhor se adequa às necessidades de processamento do seu cluster: processamento em lotes, consultas de baixa latência, dados de streaming ou armazenamento físico de dados grandes. Para obter mais informações sobre os tipos de instâncias disponíveis para o Amazon EMR, consulte Configurar o hardware e as redes do cluster.

O Amazon EMR oferece diversas maneiras de configurar softwares em seu cluster. Por exemplo, você pode instalar uma versão do Amazon EMR com um conjunto de aplicações escolhidas que pode incluir estruturas versáteis, como o Hadoop, e aplicações, como o Hive, o Pig ou o Spark. Também é possível instalar uma das diversas distribuições do MapR. O Amazon EMR usa o Amazon Linux, portanto, você também pode instalar softwares no cluster de forma manual ao usar o gerenciador de pacotes YUM ou a partir da origem. Para ter mais informações, consulte Configuração de software do cluster.

Escalabilidade e flexibilidade

O Amazon EMR oferece flexibilidade para aumentar ou reduzir a escala verticalmente do seu cluster conforme as necessidades de computação são alteradas. Você pode redimensionar seu cluster para adicionar instâncias para cargas de trabalho de pico e remover instâncias para controlar custos quando as cargas de pico diminuírem. Para ter mais informações, consulte Redimensionar manualmente um cluster em execução.

O Amazon EMR também oferece a opção de executar vários grupos de instâncias para que você possa usar instâncias sob demanda em um grupo para garantir a capacidade de processamento

em conjunto com instâncias spot em outro grupo para concluir os trabalhos com mais rapidez e custos mais baixos. Você também pode combinar diferentes tipos de instâncias para tirar proveito dos melhores preços por um tipo de instância spot sobre o outro. Para ter mais informações, consulte Quando você deve usar instâncias spot?.

Além disso, o Amazon EMR oferece flexibilidade para usar vários sistemas de arquivos para dados de entrada, de saída e intermediários. Por exemplo, você pode escolher o Sistema de Arquivos Distribuído do Hadoop (HDFS), que é executado nos nós primários e centrais do cluster para o processamento de dados que não precisam ser armazenados além do ciclo de vida do cluster. Você pode escolher o Sistema de Arquivos do EMR (EMRFS) para usar o Amazon S3 como uma camada de dados para aplicações em execução no cluster, com a finalidade de separar a computação e o armazenamento, e manter os dados persistentes de forma externa ao ciclo de vida do cluster. O EMRFS fornece o benefício adicional de permitir que você aumente ou diminua a escalabilidade independentemente, de acordo com as suas necessidades de computação e armazenamento. Você pode escalar suas necessidades de computação ao redimensionar o cluster e escalar as necessidades de armazenamento ao usar o Amazon S3. Para ter mais informações, consulte Trabalhar com armazenamento e sistemas de arquivos.

Confiabilidade

O Amazon EMR monitora nós no cluster e encerra e substitui automaticamente uma instância em caso de falha.

O Amazon EMR oferece opções de configuração que controlam se o cluster será encerrado automática ou manualmente. Se você configurar o cluster para ser automaticamente encerrado, isso acontecerá após a conclusão de todas as etapas. Ele é conhecido como cluster transitório. No entanto, você pode configurar o cluster para continuar a ser executado após o processamento, para poder optar por terminá-lo manualmente quando não precisar mais dele. Outra opção é criar um cluster, interagir diretamente com os aplicativos instalados e então terminá-lo manualmente quando você não precisar mais dele. Os clusters nestes exemplos são chamados de clusters de longa execução.

Além disso, você pode configurar a proteção contra encerramento para impedir que instâncias do seu cluster sejam terminadas devido a erros ou problemas durante o processamento. Quando a proteção contra encerramento está habilitada, você pode recuperar dados de instâncias antes do encerramento. As configurações padrão para essas opções são diferentes dependendo de você executar o cluster usando o console, a CLI ou a API. Para ter mais informações, consulte Usar a proteção contra término.

Segurança

O Amazon EMR utiliza outros AWS serviços, como IAM e Amazon VPC, e recursos como pares de chaves do Amazon EC2, para ajudar você a proteger seus clusters e dados.

IAM

O Amazon EMR se integra ao IAM para gerenciar permissões. Você define permissões usando políticas do IAM, que você anexa a usuários ou grupos do IAM. As permissões que você definir na política determinam as ações que esses usuários ou membros do grupo podem realizar, bem como os recursos que eles podem acessar. Para ter mais informações, consulte Como a Amazon EMR trabalha com IAM.

Além disso, o Amazon EMR usa perfis do IAM para o próprio serviço do Amazon EMR e o perfil de instância do EC2 para as instâncias. Essas funções concedem permissões para que o serviço e as instâncias acessem outros AWS serviços em seu nome. Há um perfil padrão para o serviço do Amazon EMR e um perfil padrão para o perfil de instância do EC2. As funções padrão usam políticas AWS gerenciadas, que são criadas automaticamente para você na primeira vez que você inicia um cluster do EMR a partir do console e escolhe as permissões padrão. Você também pode criar os perfis do IAM padrão usando a AWS CLI. Se quiser gerenciar as permissões em vez de AWS, você pode escolher funções personalizadas para o perfil do serviço e da instância. Para ter mais informações, consulte Configure funções IAM de serviço para EMR permissões da Amazon para AWS serviços e recursos.

Grupos de segurança

O Amazon EMR usa grupos de segurança para controlar o tráfego de entrada e de saída para as instâncias do EC2. Ao iniciar seu cluster, o Amazon EMR usa um grupo de segurança para a instância primária e um grupo de segurança para ser compartilhado pelas instâncias centrais e de tarefas. O Amazon EMR configura as regras do grupo de segurança para garantir a comunicação entre as instâncias do cluster. Como opção, é possível configurar grupos de segurança adicionais e atribuí-los às instâncias primárias, centrais e de tarefas para obter regras mais avançadas. Para ter mais informações, consulte Controle do tráfego de rede com grupos de segurança.

Criptografia

O Amazon EMR oferece suporte à opção de criptografia do lado do cliente e do servidor do Amazon S3 com EMRFS para ajudar a proteger os dados armazenados no Amazon S3. Com a criptografia do lado do servidor, o Amazon S3 criptografa seus dados após o upload.

Com a criptografia no lado do cliente, o processo de criptografia e descriptografia ocorre no cliente EMRFS, no seu cluster do EMR. Você gerencia a chave raiz para criptografia do lado do cliente usando o AWS Key Management Service (AWS KMS) ou seu próprio sistema de gerenciamento de chaves.

Para obter mais informações, consulte Specifying Amazon S3 encryption using EMRFS properties.

Amazon VPC

O Amazon EMR oferece suporte à execução de clusters em uma nuvem privada virtual (VPC) na Amazon VPC. Uma VPC é uma rede virtual isolada AWS que fornece a capacidade de controlar aspectos avançados da configuração e do acesso à rede. Para ter mais informações, consulte Configurar redes.

AWS CloudTrail

O Amazon EMR se integra CloudTrail para registrar informações sobre solicitações feitas por ou em nome de sua conta. AWS Com essas informações, você pode manter o controle de quem está acessando seu cluster, quando isso é feito e o endereço IP do qual a solicitação foi feita. Para ter mais informações, consulte Registro de EMR API chamadas da Amazon AWS CloudTrail.

Pares de chaves do Amazon EC2

Você pode monitorar e interagir com o seu cluster ao criar uma conexão segura entre o computador remoto e o nó primário. Você usa o protocolo de rede Secure Shell (SSH) nesta conexão ou usar o Kerberos para autenticação. Se você usar o SSH, um par de chaves do Amazon EC2 será necessário. Para ter mais informações, consulte Use um par de EC2 chaves para SSH credenciais.

Monitoramento

Você pode usar as interfaces de gerenciamento e os arquivos de log do Amazon EMR para solucionar problemas de cluster, como falhas ou erros. O Amazon EMR oferece a capacidade de arquivar arquivos de log no Amazon S3 para que você possa armazenar logs e solucionar problemas mesmo após o encerramento do cluster. O Amazon EMR também fornece uma ferramenta de depuração opcional no console do Amazon EMR para navegar nos arquivos de log com base em etapas, trabalhos e tarefas. Para ter mais informações, consulte Configurar registro em log e depuração do cluster.

O Amazon EMR se integra CloudWatch para monitorar métricas de desempenho do cluster e dos trabalhos dentro do cluster. Você pode configurar alarmes com base em várias métricas, por

exemplo, se o cluster está ocioso ou a porcentagem de armazenamento usado. Para ter mais informações, consulte Monitorando EMR métricas da Amazon com CloudWatch.

Interfaces de gerenciamento

Existem diversas maneiras de interagir com o Amazon EMR:

- Console: uma interface gráfica do usuário que você pode usar para iniciar e gerenciar clusters. Com ela, você preenche formulários da Web para especificar os detalhes dos clusters a serem executados, visualizar os detalhes de clusters existentes, depurar e encerrar clusters. Usar o console é a maneira mais fácil de começar a usar o Amazon EMR e nenhum conhecimento de programação é necessário. O console está disponível on-line em https://console.aws.amazon.com/elasticmapreduce/home.
- AWS Command Line Interface (AWS CLI) — Um aplicativo cliente que você executa em sua máquina local para se conectar ao Amazon EMR e criar e gerenciar clusters. O AWS CLI contém um conjunto rico em recursos de comandos específicos para o Amazon EMR. Com isso, você pode escrever scripts que automatizam o processo de execução e gerenciamento de clusters. Se você preferir trabalhar em uma linha de comando, usar o AWS CLI é a melhor opção. Para obter mais informações, consulte Amazon EMR em AWS CLI Command Reference.
- Kit de desenvolvimento de software (SDK): os SDKs fornecem funções que chamam o Amazon EMR para criar e gerenciar clusters. Com eles, você pode escrever aplicativos que automatizam o processo de criação e gerenciamento de clusters. Usar o SDK é a melhor opção para ampliar ou personalizar a funcionalidade do Amazon EMR. No momento, o Amazon EMR está disponível nos seguintes SDKs: Go, Java, .NET (C# e VB.NET), Node.js, PHP, Python e Ruby. Para obter mais informações sobre esses SDKs, consulte Ferramentas para criar com a AWS e códigos de exemplo e bibliotecas do Amazon EMR.
- API do serviço Web: uma interface de baixo nível que você pode usar para chamar o serviço Web diretamente, usando JSON. Usar a API é a melhor opção para criar um SDK personalizado que chame o Amazon EMR. Para obter mais informações, consulte a Referência da API do Amazon EMR.

Visão geral da arquitetura do Amazon EMR

A arquitetura de serviços do Amazon EMR consiste em várias camadas, cada uma delas fornecendo determinados recursos e funcionalidades ao cluster. Esta seção fornece uma visão geral das camadas e dos componentes de cada uma.

Neste tópico

Armazenamento

A camada de armazenamento inclui os diferentes sistemas de arquivos que são usados com o cluster. Existem vários tipos diferentes de opções de armazenamento, da seguinte maneira.

Hadoop Distributed File System (HDFS)

O Hadoop Distributed File System (HDFS) é um sistema de arquivos distribuído e escalável para o Hadoop. O HDFS distribui os dados armazenados entre as instâncias do cluster, armazenando várias cópias dos dados em instâncias diferentes para garantir que nenhum dos dados se perca caso uma das instâncias falhe. O HDFS é um armazenamento temporário que é reivindicado quando um cluster é encerrado. O HDFS é útil para armazenar em cache resultados intermediários durante o MapReduce processamento ou para cargas de trabalho com E/S aleatória significativa.

Para obter mais informações, consulte Armazenamento de instâncias neste guia ou acesse o HDFS User Guide no site do Apache Hadoop.

Sistema de arquivos do EMR (EMRFS)

Ao usar o Sistema de Arquivos do EMR (EMRFS), o Amazon EMR amplia o Hadoop para adicionar a capacidade de acessar diretamente os dados armazenados no Amazon S3, como se ele fosse um sistema de arquivos como o HDFS. Você pode usar o HDFS ou o Amazon S3 como o sistema de arquivos em seu cluster. Na maioria das vezes, o Amazon S3 é usado para armazenar dados de entrada e de saída, e os resultados intermediários são armazenados no HDFS.

Sistema de arquivos local

O sistema de arquivos local é a um disco conectado localmente. Quando você cria um cluster do Hadoop, cada nó é criado a partir de uma instância do Amazon EC2 que tem componentes configurados previamente para o armazenamento em disco anexado previamente, que é chamado de armazenamento de instância. Os dados nos volumes de armazenamento de instância persistem somente durante o ciclo de vida da instância do Amazon EC2.

Gerenciamento de recursos de cluster

A camada de gerenciamento de recursos é responsável por gerenciar os recursos de cluster e agendar os trabalhos para o processamento de dados.

Por padrão, o Amazon EMR usa o YARN (Yet Another Resource Negotiator), que é um componente introduzido no Apache Hadoop 2.0 para gerenciar centralmente os recursos de cluster para várias estruturas de processamento de dados. No entanto, existem outras estruturas e aplicações disponibilizadas no Amazon EMR que não usam o YARN como gerenciador de recursos. O Amazon EMR também tem um atendente em cada nó que administra os componentes do YARN, mantém o cluster íntegro e se comunica com o Amazon EMR.

Como as instâncias spot são frequentemente usadas para executar nós de tarefas, o Amazon EMR tem a funcionalidade padrão para programar trabalhos do YARN para que os trabalhos em execução não falhem quando os nós de tarefas em execução nas instâncias spot forem encerrados. O Amazon EMR faz isso ao permitir que processos principais de aplicações sejam executados somente em nós centrais. O processo principal da aplicação controla os trabalhos em execução e precisa permanecer ativo durante a vida útil do trabalho.

A versão 5.19.0 e as versões posteriores do Amazon EMR usam o recurso de rótulos de nós do YARN integrado para conseguir isso. (As versões anteriores usavam um patch de código). As propriedades nas classificações de configuração `yarn-site` e `capacity-scheduler` são configuradas por padrão para que o programador de capacidade e o programador justo do YARN aproveitem os rótulos de nós. O Amazon EMR rotula automaticamente os nós centrais com o rótulo `CORE` e define propriedades para que as aplicações principais sejam programadas somente em nós com o rótulo CORE. Modificar manualmente as propriedades relacionadas nas classificações de configuração yarn-site e docapacity-scheduler, ou diretamente nos arquivos XML associados, pode interromper esse recurso ou modificar essa funcionalidade.

Estruturas de processamento de dados

A camada de estruturas de processamento de dados é o mecanismo usado para processar e analisar dados. Existem muitas estruturas disponíveis que são executadas no YARN ou têm seu próprio gerenciamento de recursos. Estruturas diferentes estão disponíveis para tipos distintos de necessidades de processamento, como e lote, interativo, na memória, streaming e assim por diante. A estrutura que você escolhe depende do seu caso de uso. A escolha afeta as linguagens e interfaces disponíveis na camada de aplicativo, que é a camada usada para interagir com os dados que você deseja processar. As principais estruturas de processamento disponíveis para o Amazon EMR são o MapReduce Hadoop e o Spark.

Hadoop MapReduce

O Hadoop MapReduce é um modelo de programação de código aberto para computação distribuída. Ele simplifica o processo de gravação de aplicativos distribuídos em paralelo, manipulando toda a lógica, enquanto você fornece as funções Map e Reduce. A função Map mapeia dados para conjuntos de pares de chave/valor chamados de resultados intermediários. A função Reduce combina os resultados intermediários, aplica algoritmos adicionais e produz o resultado final. Existem várias estruturas disponíveis para MapReduce, como o Hive, que gera automaticamente os programas Map e Reduce.

Para obter mais informações, acesse How map and reduce operations are actually carried out no site Wiki do Apache Hadoop.

Apache Spark

O Spark é um modelo de programação e estrutura de cluster para o processamento de cargas de trabalho de Big Data. Como o Hadoop MapReduce, o Spark é um sistema de processamento distribuído de código aberto, mas usa gráficos acíclicos direcionados para planos de execução e armazenamento em cache na memória para conjuntos de dados. Ao executar o Spark no Amazon EMR, você pode usar o EMRFS para acessar diretamente os dados no Amazon S3. O Spark oferece suporte a vários módulos de consulta interativos, como o SparkSQL.

Para obter mais informações, consulte Apache Spark on Amazon EMR clusters no Guia de lançamento do Amazon EMR.

Aplicações e programas

O Amazon EMR oferece suporte para muitas aplicações, como o Hive, o Pig e a biblioteca Spark Streaming, para fornecer funcionalidades como o uso de linguagens de nível superior para criar workloads de processamento, o uso de algoritmos de machine learning, a criação de aplicações de processamento de fluxos e o desenvolvimento de data warehouses. Além disso, o Amazon EMR também oferece suporte a projetos de código aberto que têm suas próprias funcionalidades de gerenciamento de cluster em vez de usarem o YARN.

Você usa várias bibliotecas e linguagens para interagir com as aplicações executadas no Amazon EMR. Por exemplo, você pode usar Java, Hive ou Pig com MapReduce Spark Streaming, Spark SQL, MLlib e GraphX com o Spark.

Para obter mais informações, consulte o Guia de versão do Amazon EMR.

Configuração do Amazon EMR

Conclua as tarefas desta seção antes de iniciar um cluster do Amazon EMR pela primeira vez:

Antes de usar o Amazon EMR pela primeira vez, conclua as seguintes tarefas:

Inscreva-se para um Conta da AWS

Se você não tiver um Conta da AWS, conclua as etapas a seguir para criar um.

Para se inscrever em um Conta da AWS

1. Abra https://portal.aws.amazon.com/billing/signup.
2. Siga as instruções on-line.

   Parte do procedimento de inscrição envolve receber uma chamada telefônica e digitar um código de verificação no teclado do telefone.

   Quando você se inscreve em um Conta da AWS, um Usuário raiz da conta da AWSé criado. O usuário raiz tem acesso a todos os Serviços da AWS e recursos na conta. Como prática recomendada de segurança, atribua acesso administrativo a um usuário e use somente o usuário raiz para realizar tarefas que exijam acesso do usuário raiz.

AWS envia um e-mail de confirmação após a conclusão do processo de inscrição. A qualquer momento, é possível visualizar as atividades da conta atual e gerenciar sua conta acessando https://aws.amazon.com/ e selecionando Minha conta.

Crie um usuário com acesso administrativo

Depois de se inscrever em um Conta da AWS, proteja seu Usuário raiz da conta da AWS AWS IAM Identity Center, habilite e crie um usuário administrativo para que você não use o usuário root nas tarefas diárias.

Proteja seu Usuário raiz da conta da AWS

1. Faça login AWS Management Consolecomo proprietário da conta escolhendo Usuário raiz e inserindo seu endereço de Conta da AWS e-mail. Na próxima página, digite sua senha.

Para obter ajuda ao fazer login usando o usuário raiz, consulte Fazer login como usuário raiz no Guia do usuário do Início de Sessão da AWS .

2. Ative a autenticação multifator (MFA) para o usuário raiz.c

   Para obter instruções, consulte Habilitar um dispositivo de MFA virtual para seu usuário Conta da AWS raiz (console) no Guia do usuário do IAM.

Crie um usuário com acesso administrativo

1. Habilitar o IAM Identity Center.

   Para obter instruções, consulte Habilitar AWS IAM Identity Center no Guia do usuário do AWS IAM Identity Center .

2. No IAM Identity Center, conceda acesso administrativo a um usuário.

   Para ver um tutorial sobre como usar o Diretório do Centro de Identidade do IAM como fonte de identidade, consulte Configurar o acesso do usuário com o padrão Diretório do Centro de Identidade do IAM no Guia AWS IAM Identity Center do usuário.

Faça login como usuário com acesso administrativo

- Para fazer login com seu usuário do Centro de Identidade do IAM, use a URL de login que foi enviada ao seu endereço de e-mail quando você criou o usuário do Centro do Usuário do IAM.

  Para obter ajuda para fazer login usando um usuário do IAM Identity Center, consulte Como fazer login no portal de AWS acesso no Guia Início de Sessão da AWS do usuário.

Atribua acesso a usuários adicionais

1. No IAM Identity Center, crie um conjunto de permissões que siga as melhores práticas de aplicação de permissões com privilégios mínimos.

   Para obter instruções, consulte Criar um conjunto de permissões no Guia AWS IAM Identity Center do usuário.

2. Atribua usuários a um grupo e, em seguida, atribua acesso de login único ao grupo.

   Para obter instruções, consulte Adicionar grupos no Guia AWS IAM Identity Center do usuário.

Crie um par de chaves do Amazon EC2 para SSH

> **Note**
>
> Com as versões 5.10.0 ou posteriores do Amazon EMR, é possível configurar o Kerberos para autenticar usuários e conexões SSH com um cluster. Para ter mais informações, consulte Use o Kerberos para autenticação com a Amazon EMR.

Para realizar a autenticação e se conectar aos nós em um cluster por meio de um canal seguro usando o protocolo Secure Shell (SSH), crie um par de chaves do Amazon Elastic Compute Cloud (Amazon EC2) antes de iniciar o cluster. Também é possível criar um cluster sem par de chaves. Isso geralmente é feito com clusters transitórios que são iniciados, executam etapas e são encerrados automaticamente.

Se...	Então...
Você já tem um par de chaves do Amazon EC2 que deseja usar ou não tem a necessidade de se autenticar no cluster.	Pule esta etapa.
Você precisa criar um par de chaves.	Consulte Creating your key pair using Amazon EC2.

Próximas etapas

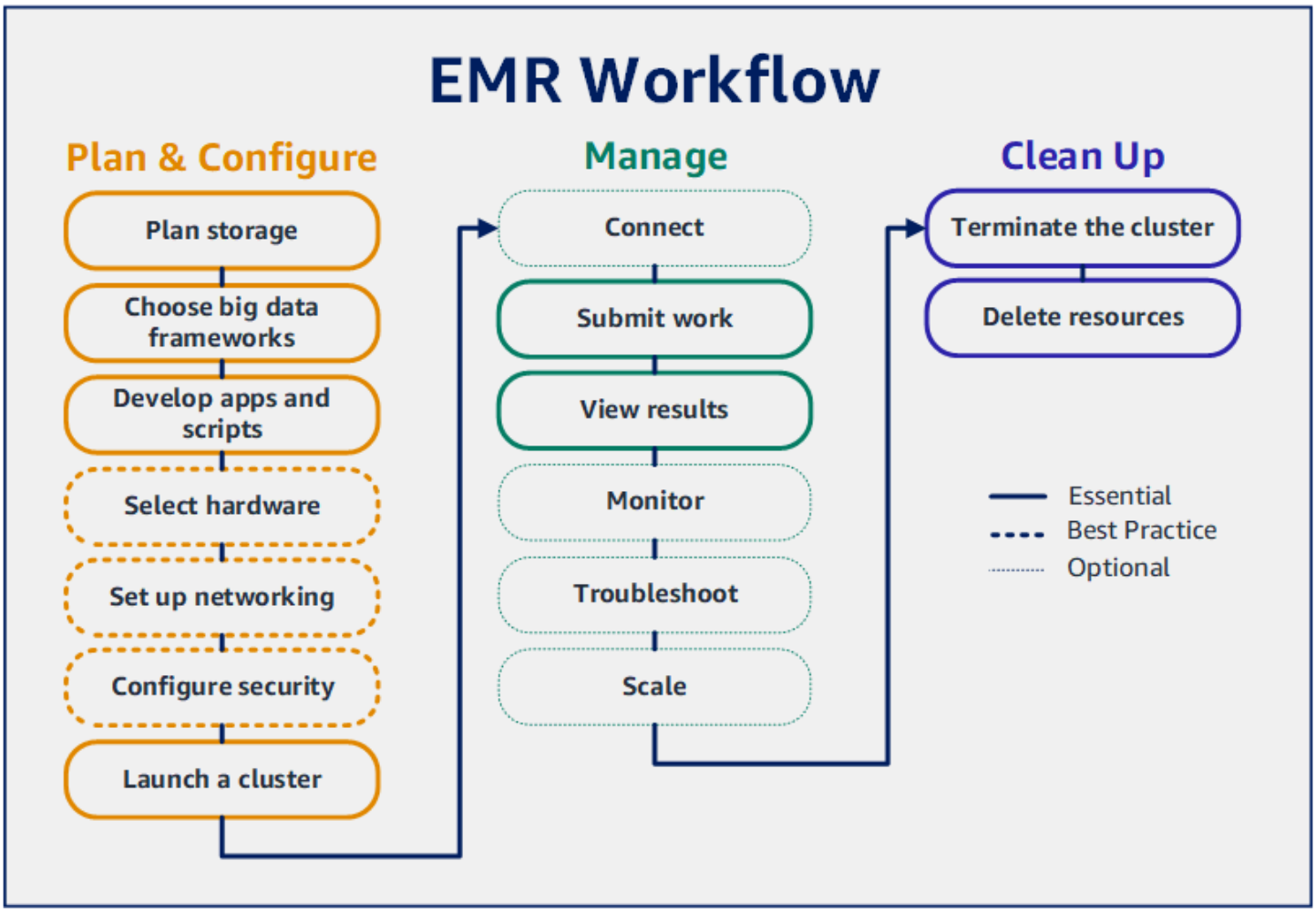
- Para obter orientação sobre como criar um cluster de exemplo, consulte Tutorial: Começando com a Amazon EMR.
- Para obter mais informações sobre como configurar um cluster personalizado e controlar o acesso a ele, consulte Planejar e configurar clusters e Segurança na Amazon EMR.

Tutorial: Começando com a Amazon EMR

Visão geral

Com a Amazon, EMR você pode configurar um cluster para processar e analisar dados com estruturas de big data em apenas alguns minutos. Este tutorial mostra como iniciar um cluster de amostra usando o Spark e como executar um PySpark script simples armazenado em um bucket do Amazon S3. Ele abrange EMR tarefas essenciais da Amazon em três categorias principais de fluxo de trabalho: planejar e configurar, gerenciar e limpar.

Você encontrará links para tópicos mais detalhados à medida que avança no tutorial e obterá ideias para etapas adicionais na seção Próximas etapas. Se você tiver dúvidas ou tiver dúvidas, entre em contato com a EMR equipe da Amazon em nosso fórum de discussão.



Pré-requisitos

- Antes de iniciar um EMR cluster da Amazon, certifique-se de concluir as tarefas emConfiguração do Amazon EMR.

Custo

- O exemplo de cluster que você criar será executado em um ambiente dinâmico. O cluster acumula cobranças mínimas. Para evitar cobranças adicionais, certifique-se de concluir as tarefas de limpeza na última etapa deste tutorial. As cobranças são acumuladas na taxa por segundo, de acordo com os preços da Amazon. EMR As cobranças também variam com base na região. Para obter mais informações, consulte os EMRpreços da Amazon.
- Cobranças mínimas podem ser acumuladas para arquivos pequenos armazenados no Amazon S3. Algumas ou todas as cobranças do Amazon S3 podem ser dispensadas se você estiver dentro dos limites de uso do AWS nível gratuito. Para obter mais informações, consulte Preço do Amazon S3 e nível gratuito da AWS.

Etapa 1: Planejar e configurar um EMR cluster da Amazon

Prepare o armazenamento para a Amazon EMR

Ao usar a AmazonEMR, você pode escolher entre uma variedade de sistemas de arquivos para armazenar dados de entrada, dados de saída e arquivos de log. Neste tutorial, você usa EMRFS para armazenar dados em um bucket do S3. EMRFSé uma implementação do sistema de arquivos Hadoop que permite ler e gravar arquivos regulares no Amazon S3. Para obter mais informações, consulte Trabalhar com armazenamento e sistemas de arquivos.

Para criar um bucket para este tutorial, siga as instruções em How do I create an S3 bucket? no Guia do usuário do console do Amazon Simple Storage Service. Crie o bucket na mesma AWS região em que você planeja lançar seu EMR cluster da Amazon. Por exemplo, Oeste dos EUA (Oregon) us-west-2.

Os buckets e pastas que você usa com a Amazon EMR têm as seguintes limitações:

- Os nomes podem consistir em letras minúsculas, números, pontos (.) e hifens (-).
- Os nomes não podem terminar em números.
- O nome do bucket deve ser exclusivo em todas as contas da AWS .

- Uma pasta de saída deve estar vazia.

Prepare um aplicativo com dados de entrada para a Amazon EMR

A forma mais comum de preparar um aplicativo para a Amazon EMR é fazer o upload do aplicativo e seus dados de entrada para o Amazon S3. Em seguida, ao enviar o trabalho para o cluster, você especifica os locais do Amazon S3 para o script e para os dados.

Nesta etapa, você carrega um PySpark script de amostra no seu bucket do Amazon S3. Fornecemos um PySpark script para você usar. O script processa os dados de inspeção de estabelecimentos alimentícios e retorna um arquivo de resultados em seu bucket do S3. O arquivo de resultados lista os dez principais estabelecimentos com mais violações do tipo “vermelho”.

Você também carrega dados de entrada de amostra para o Amazon S3 para que o PySpark script seja processado. Os dados de entrada correspondem a uma versão modificada dos resultados de inspeções do Departamento de Saúde no Condado de King, em Washington, de 2006 a 2020. Para obter mais informações, consulte King County Open Data: Food Establishment Inspection Data. Confira a seguir exemplos de linhas do conjunto de dados.

```
name, inspection_result, inspection_closed_business, violation_type, violation_points
100 LB CLAM, Unsatisfactory, FALSE, BLUE, 5
100 PERCENT NUTRICION, Unsatisfactory, FALSE, BLUE, 5
7-ELEVEN #2361-39423A, Complete, FALSE, , 0
```

Para preparar o PySpark script de exemplo para EMR

1. Copie o código de exemplo abaixo em um novo arquivo no editor de sua preferência.

```
import argparse

from pyspark.sql import SparkSession

def calculate_red_violations(data_source, output_uri):
    """
    Processes sample food establishment inspection data and queries the data to
 find the top 10 establishments
    with the most Red violations from 2006 to 2020.

    :param data_source: The URI of your food establishment data CSV, such as 's3://
DOC-EXAMPLE-BUCKET/food-establishment-data.csv'.
```

```
    :param output_uri: The URI where output is written, such as 's3://DOC-EXAMPLE-
BUCKET/restaurant_violation_results'.
    """
    with SparkSession.builder.appName("Calculate Red Health
 Violations").getOrCreate() as spark:
        # Load the restaurant violation CSV data
        if data_source is not None:
            restaurants_df = spark.read.option("header", "true").csv(data_source)

        # Create an in-memory DataFrame to query
        restaurants_df.createOrReplaceTempView("restaurant_violations")

        # Create a DataFrame of the top 10 restaurants with the most Red violations
        top_red_violation_restaurants = spark.sql("""SELECT name, count(*) AS
 total_red_violations
          FROM restaurant_violations
          WHERE violation_type = 'RED'
          GROUP BY name
          ORDER BY total_red_violations DESC LIMIT 10""")

        # Write the results to the specified output URI
        top_red_violation_restaurants.write.option("header",
 "true").mode("overwrite").csv(output_uri)

if __name__ == "__main__":
    parser = argparse.ArgumentParser()
    parser.add_argument(
        '--data_source', help="The URI for you CSV restaurant data, like an S3
 bucket location.")
    parser.add_argument(
        '--output_uri', help="The URI where output is saved, like an S3 bucket
 location.")
    args = parser.parse_args()

    calculate_red_violations(args.data_source, args.output_uri)
```

2. Salve o arquivo como `health_violations.py`.
3. Faça o upload de `health_violations.py` para o Amazon S3 no bucket criado para este tutorial. Para obter instruções, consulte Fazer upload de um objeto para o bucket no Guia de conceitos básicos do Amazon Simple Storage Service.

Para preparar os dados de entrada de amostra para EMR

1. Faça o download do arquivo zip food_establishment_data.zip.
2. Descompacte e salve `food_establishment_data.zip` como `food_establishment_data.csv` em sua máquina.
3. Faça upload do CSV arquivo no bucket do S3 que você criou para este tutorial. Para obter instruções, consulte Fazer upload de um objeto para o bucket no Guia de conceitos básicos do Amazon Simple Storage Service.

Para obter mais informações sobre como configurar dados paraEMR, consultePreparar dados de entrada.

Inicie um EMR cluster da Amazon

Depois de preparar um local de armazenamento e seu aplicativo, você pode iniciar um EMR cluster de amostra da Amazon. Nesta etapa, você executa um cluster Apache Spark usando a EMRversão mais recente da Amazon.

Console

Para iniciar um cluster com o Spark instalado com o console

1. Faça login no AWS Management Console e abra o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Na página Criar cluster, observe os valores padrão para Versão, Tipo de instância, Número de instâncias e Permissões. Esses campos são preenchidos automaticamente com valores que funcionam para clusters de uso geral.
4. No campo Nome do cluster, insira um nome de cluster exclusivo para ajudá-lo a identificar seu cluster, como *`My first cluster`*. O nome do cluster não pode conter os caracteres <, >, $, | ou `(crase).
5. Em Aplicações, escolha a opção Spark para instalar o Spark em seu cluster.

> **Note**
>
> Escolha os aplicativos que você deseja em seu EMR cluster da Amazon antes de iniciar o cluster. Não é possível adicionar ou remover aplicações de um cluster após a inicialização.

6. Em Logs do cluster, marque a caixa de seleção Publicar logs específicos do cluster no Amazon S3. Substitua o valor do local do Amazon S3 pelo bucket do Amazon S3 criado, seguido por **`/logs`**. Por exemplo, **`s3://DOC-EXAMPLE-BUCKET/logs`**. Adicionar **`/logs`** cria uma nova pasta chamada 'logs' em seu bucket, na qual a Amazon EMR pode copiar os arquivos de log do seu cluster.
7. Em Configuração e permissões de segurança, escolha seu EC2key pair. Na mesma seção, selecione o menu EMR suspenso Função de serviço para Amazon e escolha EMR _. DefaultRole Em seguida, selecione a IAMfunção do menu suspenso do perfil da instância e escolha EMR_ _ EC2. DefaultRole
8. Escolha Criar cluster para iniciar o cluster e abrir a página de detalhes do cluster.
9. Veja o Status do cluster próximo ao nome do cluster. O status muda de Starting to Running para Waiting à medida que a Amazon EMR provisiona o cluster. Pode ser necessário escolher o ícone de atualização à direita ou atualizar seu navegador para visualizar as atualizações de status.

O status do cluster é alterado para Aguardando quando o cluster está ativo, em execução e pronto para aceitar trabalhos. Para obter mais informações sobre como ler o resumo do cluster, consulte Visualizar o status e os detalhes do cluster. Para obter informações sobre o status do cluster, consulte Noções básicas sobre o ciclo de vida do cluster.

CLI

Para iniciar um cluster com o Spark instalado com o AWS CLI

1. Crie funções IAM padrão que você possa usar para criar seu cluster usando o comando a seguir.

```
aws emr create-default-roles
```

Para obter mais informações sobre `create-default-roles`, consulte AWS CLI Command Reference.

2. Crie um cluster do Spark com o comando a seguir. Insira um nome para seu cluster com a `--name` opção e especifique o nome do seu EC2 key pair com a `--ec2-attributes` opção.

```
aws emr create-cluster \
--name "<My First EMR Cluster>" \
--release-label <emr-5.36.2> \
--applications Name=Spark \
--ec2-attributes KeyName=<myEMRKeyPairName> \
--instance-type m5.xlarge \
--instance-count 3 \
--use-default-roles
```

Observe os outros valores necessários para `--instance-type`, `--instance-count` e `--use-default-roles`. Esses valores foram escolhidos para clusters de uso geral. Para obter mais informações sobre `create-cluster`, consulte AWS CLI Command Reference.

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

O resultado deverá ser parecido com o que segue. A saída mostra o `ClusterId` e o `ClusterArn` do seu novo cluster. Anote o seu `ClusterId`. Você usa o `ClusterId` para verificar o status do cluster e enviar trabalhos.

```
{
    "ClusterId": "myClusterId",
    "ClusterArn": "myClusterArn"
}
```

3. Verifique o status do seu cluster com o comando a seguir.

```
aws emr describe-cluster --cluster-id <myClusterId>
```

Você deverá visualizar uma saída semelhante à apresentada a seguir com o objeto `Status` para o seu novo cluster.

```
{
    "Cluster": {
        "Id": "myClusterId",
        "Name": "My First EMR Cluster",
        "Status": {
            "State": "STARTING",
            "StateChangeReason": {
                "Message": "Configuring cluster software"
            }
        }
     }
}
```

O `State` valor muda de `STARTING` para `RUNNING` à `WAITING` medida que a Amazon EMR provisiona o cluster.

O status do cluster é alterado para **`WAITING`** quando um cluster está ativo, em execução e pronto para aceitar trabalhos. Para obter informações sobre o status do cluster, consulte Noções básicas sobre o ciclo de vida do cluster.

Etapa 2: Gerencie seu EMR cluster da Amazon

Envie seu trabalho para a Amazon EMR

Após iniciar um cluster, você poderá enviar trabalhos ao cluster em execução para processar e analisar dados. Você envia o trabalho para um EMR cluster da Amazon como uma etapa. Uma etapa é uma unidade de trabalho composta por uma ou mais ações. Por exemplo, você pode enviar uma etapa para calcular valores ou para transferir e processar dados. É possível enviar etapas ao criar um cluster ou para um cluster em execução. Nesta parte do tutorial, você envia `health_violations.py` como uma etapa para o cluster em execução. Para saber mais sobre as etapas, consulte Enviar trabalhos a um cluster.

Console

Para enviar um aplicativo Spark como uma etapa com o console

1. Faça login no AWS Management Console e abra o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, selecione o cluster para o qual você deseja enviar o trabalho. O estado do cluster deve ser Aguardando.
3. Escolha a guia Etapas e, em seguida, escolha Adicionar etapa.
4. Configure a etapa de acordo com as seguintes diretrizes:

   - Para Tipo, escolha Aplicação do Spark. Você deverá visualizar campos adicionais para Modo de implantação, Local da aplicação e Opções de spark-submit.
   - Para Nome, insira um novo nome. Se você tiver muitas etapas em um cluster, nomear cada etapa ajudará a controlá-las.
   - Para Modo de implantação, deixe o valor padrão Modo de cluster. Para obter mais informações sobre os modos de implantação do Spark, consulte Cluster mode overview na documentação do Apache Spark.
   - Em Localização do aplicativo, insira a localização do seu `health_violations.py` script no Amazon S3, como *`s3://DOC-EXAMPLE-BUCKET/health_violations.py`*.
   - Deixe o campo Opções de spark-submit vazio. Para obter mais informações sobre as opções de `spark-submit`, consulte Launching applications with spark-submit.
   - No campo Argumentos, insira os seguintes argumentos e valores:

     ```
     --data_source s3://DOC-EXAMPLE-BUCKET/food_establishment_data.csv
     --output_uri s3://DOC-EXAMPLE-BUCKET/myOutputFolder
     ```

     Substituir *`s3://DOC-EXAMPLE-BUCKET/food_establishment_data.csv`* com o bucket S3 URI dos dados de entrada que você preparou. Prepare um aplicativo com dados de entrada para a Amazon EMR

     Substituir *`DOC-EXAMPLE-BUCKET`* com o nome do bucket que você criou para este tutorial e substitua *`myOutputFolder`* com um nome para a pasta de saída do cluster.
   - Para Ação se a etapa falhar, aceite a opção padrão Continuar. Dessa forma, se a etapa falhar, o cluster continuará em execução.

5. Escolha Adicionar para enviar a etapa. A etapa deve ser exibida no console com o status Pendente.
6. Monitore o status da etapa. Ele deve ser alterado de Pendente para Em execução e, por fim, para Concluído. Para atualizar o status no console, escolha o ícone de atualização à direita de Filtrar. O script demora cerca de um minuto para ser executado. Quando o status for alterado para Concluído, a etapa será concluída com êxito.

CLI

Para enviar uma inscrição no Spark como uma etapa com o AWS CLI

1. Certifique-se de ter o `ClusterId` do cluster iniciado em Inicie um EMR cluster da Amazon. Também é possível recuperar o ID do cluster com o comando apresentado a seguir.

```
aws emr list-clusters --cluster-states WAITING
```

2. Envie `health_violations.py` como uma etapa com o comando `add-steps` e seu `ClusterId`.

   - Você pode especificar um nome para sua etapa substituindo *"My Spark Application"*. Na `Args` matriz, substitua *s3://DOC-EXAMPLE-BUCKET/health_violations.py* com a localização do seu `health_violations.py` aplicativo.
   - Substituir *s3://DOC-EXAMPLE-BUCKET/food_establishment_data.csv* com a localização do seu `food_establishment_data.csv` conjunto de dados no S3.
   - Substituir *s3://DOC-EXAMPLE-BUCKET/MyOutputFolder* com o caminho S3 do bucket designado e um nome para a pasta de saída do cluster.
   - `ActionOnFailure=CONTINUE` significa que o cluster continuará em execução se a etapa falhar.

```
aws emr add-steps \
--cluster-id <myClusterId> \
--steps Type=Spark,Name="<My Spark
 Application>",ActionOnFailure=CONTINUE,Args=[<s3://DOC-EXAMPLE-
BUCKET/health_violations.py>,--data_source,<s3://DOC-EXAMPLE-BUCKET/
food_establishment_data.csv>,--output_uri,<s3://DOC-EXAMPLE-BUCKET/
MyOutputFolder>]
```

Para obter mais informações sobre o envio de etapas usando oCLI, consulte a Referência de AWS CLI Comandos.

Após enviar a etapa, você deverá visualizar uma saída como a apresentada a seguir, com uma lista de `StepIds`. Como você enviou uma etapa, verá somente um ID na lista. Copie o ID da etapa. Você usa o ID da etapa para verificar o status da etapa.

```
{
    "StepIds": [
        "s-1XXXXXXXXXXA"
    ]
}
```

3. Consulte o status da sua etapa com o comando `describe-step`.

```
aws emr describe-step --cluster-id <myClusterId> --step-id <s-1XXXXXXXXXXA>
```

Você deverá visualizar uma saída como a apresentada a seguir com informações sobre a etapa.

```
{
    "Step": {
        "Id": "s-1XXXXXXXXXXA",
        "Name": "My Spark Application",
        "Config": {
            "Jar": "command-runner.jar",
            "Properties": {},
            "Args": [
                "spark-submit",
                "s3://DOC-EXAMPLE-BUCKET/health_violations.py",
                "--data_source",
                "s3://DOC-EXAMPLE-BUCKET/food_establishment_data.csv",
                "--output_uri",
                "s3://DOC-EXAMPLE-BUCKET/myOutputFolder"
            ]
        },
        "ActionOnFailure": "CONTINUE",
        "Status": {
            "State": "COMPLETED"
        }
```

```
    }
}
```

O `State` da etapa é alterado de `PENDING` para `RUNNING` e para `COMPLETED`, conforme a etapa é executada. A etapa demora cerca de um minuto para ser executada, então pode ser necessário verificar o status algumas vezes.

Você saberá que a etapa obteve êxito quando o `State` for alterado para **`COMPLETED`**.

Para obter mais informações sobre o ciclo de vida da etapa, consulte Execução de etapas para processar dados.

Visualização dos resultados

Após a execução com êxito de uma etapa, você poderá visualizar os resultados de saída na pasta de saída do Amazon S3.

Visualizar os resultados de **`health_violations.py`**

1. Abra o console do Amazon S3 em. https://console.aws.amazon.com/s3/
2. Escolha o nome do bucket e, em seguida, a pasta de saída que você especificou ao enviar a etapa. Por exemplo, *`DOC-EXAMPLE-BUCKET`* e depois *`myOutputFolder`*.
3. Verifique se os seguintes itens aparecem na sua pasta de saída:
   - Um objeto de tamanho pequeno chamado `_SUCCESS`.
   - Um CSV arquivo que começa com o prefixo `part-` que contém seus resultados.
4. Escolha o objeto com seus resultados e, em seguida, escolha Fazer download para salvar os resultados em seu sistema de arquivos local.
5. Abra os resultados no editor de sua preferência. O arquivo de saída lista os dez principais estabelecimentos de alimentação com o maior número de violações vermelhas. O arquivo de saída também mostra o número total de violações vermelhas para cada estabelecimento.

   Confira a seguir um exemplo de resultados para `health_violations.py`.

```
name, total_red_violations
SUBWAY, 322
T-MOBILE PARK, 315
```

```
WHOLE FOODS MARKET, 299
PCC COMMUNITY MARKETS, 251
TACO TIME, 240
MCDONALD'S, 177
THAI GINGER, 153
SAFEWAY INC #1508, 143
TAQUERIA EL RINCONSITO, 134
HIMITSU TERIYAKI, 128
```

Para obter mais informações sobre a saída EMR do cluster da Amazon, consulteConfigurar um local de saída.

(Opcional) Conecte-se ao seu EMR cluster Amazon em execução

Ao usar a AmazonEMR, talvez você queira se conectar a um cluster em execução para ler arquivos de log, depurar o cluster ou usar CLI ferramentas como o shell do Spark. A Amazon EMR permite que você se conecte a um cluster usando o protocolo Secure Shell (SSH). Esta seção aborda como configurarSSH, conectar-se ao seu cluster e visualizar arquivos de log do Spark. Para obter mais informações sobre como se conectar a um cluster, consulte Autentique-se nos nós de EMR cluster da Amazon.

Autorize SSH conexões com seu cluster

Antes de se conectar ao cluster, você precisa modificar os grupos de segurança do cluster para autorizar conexões de entradaSSH. Os grupos EC2 de segurança da Amazon atuam como firewalls virtuais para controlar o tráfego de entrada e saída do seu cluster. Quando você criou seu cluster para este tutorial, a Amazon EMR criou os seguintes grupos de segurança em seu nome:

ElasticMapReduce-mestre

O grupo de segurança EMR gerenciado padrão da Amazon associado ao nó primário. Em um EMR cluster da Amazon, o nó primário é uma EC2 instância da Amazon que gerencia o cluster.

ElasticMapReduce-escravo

O grupo de segurança padrão associado aos nós centrais e de tarefa.

Console

Para permitir o SSH acesso a fontes confiáveis para o grupo de segurança primário com o console

Para editar seus grupos de segurança, você deve ter permissão para gerenciar grupos de segurança do cluster em VPC que o cluster está. Para obter mais informações, consulte Alteração de permissões para um usuário e o exemplo de política que permite gerenciar grupos de EC2 segurança no Guia IAM do usuário.

1. Faça login no AWS Management Console e abra o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha o cluster que você deseja atualizar. Isso abre a página de detalhes do cluster. A guia Propriedades nesta página deve estar pré-selecionada.
3. Em Rede, na guia Propriedades, selecione a seta ao lado EC2de grupos de segurança (firewall) para expandir essa seção. Em Nó primário, selecione o link do grupo de segurança. Ao concluir as etapas a seguir, você pode, opcionalmente, voltar a essa etapa, escolher os nós principais e de tarefas e repetir as etapas a seguir para permitir que o SSH cliente acesse os nós principais e de tarefas.
4. Isso abre o console do EC2. Escolha a guia Regras de entrada e, em seguida, Editar regras de entrada.
5. Verifique se há uma regra de entrada que permita acesso público com as configurações a seguir. Se existir, escolha Excluir para removê-la.

   - Tipo

     SSH
   - Porta

     22
   - Origem

     Personalizado 0.0.0.0/0

> **Warning**
>
> Antes de dezembro de 2020, o grupo de segurança ElasticMapReduce -master tinha uma regra pré-configurada para permitir tráfego de entrada na Porta 22 de todas as fontes. Essa regra foi criada para simplificar SSH as conexões iniciais com o nó principal. Recomendamos fortemente remover esta regra de entrada e restringir o tráfego para origens confiáveis.

6. Role até o final da lista de regras e escolha Adicionar regra.
7. Em Tipo, selecione SSH. A seleção insere SSH automaticamente TCPpara Protocolo e 22 para Intervalo de Portas.
8. Para a origem, selecione Meu IP para adicionar automaticamente seu endereço IP como o endereço de origem. Você também pode adicionar um intervalo personalizado de endereços IP de clientes confiáveis ou criar regras adicionais para outros clientes. Diversos ambientes de rede alocam endereços IP dinamicamente, portanto, pode ser necessário atualizar os endereços IP para clientes confiáveis no futuro.
9. Escolha Salvar.
10. Opcionalmente, escolha nós principais e de tarefas na lista e repita as etapas acima para permitir o acesso SSH do cliente aos nós principais e de tarefas.

Conecte-se ao seu cluster usando o AWS CLI

Independentemente do seu sistema operacional, você pode criar uma SSH conexão com seu cluster usando AWS CLI o.

Para se conectar ao seu cluster e visualizar os arquivos de log usando o AWS CLI

1. Use o comando a seguir para abrir uma SSH conexão com seu cluster. Substituir *<mykeypair.key>* com o caminho completo e o nome do arquivo do seu arquivo de key pair. Por exemplo, `C:\Users\<username>\.ssh\mykeypair.pem`.

   ```
   aws emr ssh --cluster-id <j-2AL4XXXXXX5T9> --key-pair-file <~/mykeypair.key>
   ```

2. Navegue até `/mnt/var/log/spark` para acessar os logs do Spark no nó principal do cluster. Em seguida, visualize os arquivos nesse local. Para obter uma lista de arquivos de log adicionais no nó principal, consulte Visualizar arquivos de log no nó primário.

```
cd /mnt/var/log/spark
ls
```

Etapa 3: limpe seus EMR recursos da Amazon

Encerramento do cluster

Agora que você enviou o trabalho para seu cluster e visualizou os resultados do seu PySpark aplicativo, você pode encerrar o cluster. O encerramento de um cluster interrompe todas as EMR cobranças e EC2 instâncias da Amazon associadas ao cluster.

Quando você encerra um cluster, a Amazon EMR retém metadados sobre o cluster por dois meses sem nenhum custo. Os metadados arquivados ajudam a clonar o cluster para um novo trabalho ou a revisitar a configuração do cluster para finalidades de referência. Os metadados não incluem dados que o cluster grava no S3 ou dados armazenados HDFS no cluster.

> **Note**
>
> O EMR console da Amazon não permite que você exclua um cluster da visualização de lista depois de encerrar o cluster. Um cluster encerrado desaparece do console quando a Amazon EMR limpa seus metadados.

Console

Para encerrar o cluster com o console

1. Faça login no AWS Management Console e abra o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. Escolha Clusters e, em seguida, selecione o cluster que você deseja encerrar.
3. No menu suspenso Ações, escolha Encerrar cluster.
4. Escolha Encerrar na caixa de diálogo. Dependendo da configuração do cluster, o encerramento pode demorar de cinco a dez minutos. Para obter mais informações sobre como criar EMR clusters da Amazon, consulteTerminar um cluster.

CLI

Para encerrar o cluster com o AWS CLI

1. Inicie o processo de encerramento do cluster com o comando a seguir. Substituir *<myClusterId>* com o ID do seu cluster de amostra. O comando não retorna uma saída.

```
aws emr terminate-clusters --cluster-ids <myClusterId>
```

2. Para verificar se o processo de encerramento do cluster está em andamento, verifique o status do cluster com o comando a seguir.

```
aws emr describe-cluster --cluster-id <myClusterId>
```

A seguir está um exemplo de saída em JSON formato. O `Status` do cluster deve ser alterado de **`TERMINATING`** para **`TERMINATED`**. O encerramento pode demorar de cinco a dez minutos, dependendo da configuração do cluster. Para obter mais informações sobre o encerramento de um EMR cluster da Amazon, consulteTerminar um cluster.

```
{
    "Cluster": {
        "Id": "j-xxxxxxxxxxxx",
        "Name": "My Cluster Name",
        "Status": {
            "State": "TERMINATED",
            "StateChangeReason": {
                "Code": "USER_REQUEST",
                "Message": "Terminated by user request"
            }
        }
    }
}
```

Exclusão de recursos do S3

Para evitar cobranças adicionais, você deve excluir o bucket do Amazon S3. Excluir o bucket remove todos os recursos do Amazon S3 deste tutorial. O bucket deve conter:

- O PySpark roteiro

- O conjunto de dados de entrada.
- Sua pasta de resultados de saída.
- Sua pasta de arquivos de log.

Talvez seja necessário tomar medidas adicionais para excluir os arquivos armazenados se você salvou o PySpark script ou a saída em um local diferente.

> **Note**
>
> O cluster deve ser encerrado antes de você excluir o bucket. Caso contrário, pode não ser possível esvaziar o bucket.

Para excluir seu bucket, siga as instruções apresentadas em How do I delete an S3 bucket? no Guia do usuário do Amazon Simple Storage Service.

Próximas etapas

Agora você lançou seu primeiro EMR cluster da Amazon do início ao fim. Você também concluiu EMR tarefas essenciais, como preparar e enviar aplicativos de big data, visualizar resultados e encerrar um cluster.

Use os tópicos a seguir para saber mais sobre como você pode personalizar seu EMR fluxo de trabalho na Amazon.

Explore aplicativos de big data para a Amazon EMR

Descubra e compare os aplicativos de big data que você pode instalar em um cluster no Amazon EMR Release Guide. O Guia de lançamento detalha cada versão de EMR lançamento e inclui dicas para usar estruturas como Spark e Hadoop na Amazon. EMR

Planejamento do hardware, das redes e da segurança do cluster

Neste tutorial, você criou um EMR cluster simples sem configurar opções avançadas. As opções avançadas permitem que você especifique os tipos de EC2 instância, a rede e a segurança do cluster da Amazon. Para obter mais informações sobre como planejar e iniciar um cluster que atenda aos seus requisitos, consulte Planejar e configurar clusters e Segurança na Amazon EMR.

Gerenciar clusters

Aprofunde-se no trabalho com clusters em execução em Gerenciar clusters. Para gerenciar um cluster, é possível se conectar ao cluster, depurar etapas e rastrear as atividades e a integridade do cluster. Você também pode ajustar os recursos do cluster em resposta às demandas de carga de trabalho com escalabilidade EMR gerenciada.

Uso de uma interface diferente

Além do EMR console da Amazon, você pode gerenciar a AWS Command Line Interface Amazon EMR usando o serviço API web ou um dos muitos compatíveis AWS SDKs. Para obter mais informações, consulte Interfaces de gerenciamento.

Você também pode interagir com aplicativos instalados nos EMR clusters da Amazon de várias maneiras. Algumas aplicações, como o Apache Hadoop, publicam interfaces da Web que você pode visualizar. Para obter mais informações, consulte Visualize interfaces web hospedadas em EMR clusters da Amazon.

Navegue pelo blog EMR técnico

Para exemplos de apresentações e discussões técnicas aprofundadas sobre os novos EMR recursos da Amazon, consulte o AWS blog de big data.

Console do Amazon EMR

O console oferece uma interface atualizada que fornece uma maneira intuitiva de gerenciar seu ambiente Amazon EMR e fornece acesso conveniente à documentação, informações sobre produtos e outros recursos.

Capacidades do console

O console do Amazon EMR está disponível no seguinte URL:

- URL do console — https://console.aws.amazon.com/emr

A tabela a seguir lista o status dos principais componentes do console do Amazon EMR.

Componente do console do Amazon EMR	Console	
EMR Studio	✓	
Criar e gerenciar clusters	✓	
Bloqueio de acesso público	✓	
Monitore CloudWatch eventos da Amazon	✓	
Configurações de segurança	✓	
Clusters virtuais (Amazon EMR no EKS)	✓	
Visualize e gerencie suas sub-redes da Amazon Virtual Private Cloud [1]	✓	
Cadernos [2]	✓	

[1] No console, você pode visualizar e gerenciar suas sub-redes da Amazon VPC na seção Rede ao criar um cluster.

[2] Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

Resumo das diferenças

Esta seção descreve os recursos da experiência do console do Amazon EMR. Esses recursos se enquadram nas seguintes categorias:

- Compatibilidade de clusters no console
- Criar clusters
- Visualizando ou editando detalhes do cluster
- Visualizando e pesquisando clusters
- Diferenças no trabalho com configurações de segurança

Compatibilidade de clusters no console

Em alguns casos, um cluster que você criou pode não ser compatível com o console. A lista a seguir descreve os requisitos de compatibilidade para o console do Amazon EMR.

- O console oferece suporte a clusters criados nas versões 5.20.1 e posteriores do Amazon EMR.
- Você pode clonar clusters que usam escalabilidade automática no console, mas você só pode criar novos clusters se quiser escalá-los manualmente ou usar escalabilidade gerenciada.

Para criar e trabalhar com clusters da versão 5.20.1 e anteriores, você pode usar o AWS Command Line Interface (AWS CLI) ou o AWS SDK.

Criar clusters

Recurso	Console	
Terminologia: tipos de nós de cluster do Amazon EMR	Primário, central e de tarefa	

Recurso	Console	
Versões do Amazon EMR com suporte[1]	Versão 5.20.1 e posterior do Amazon EMR	
Início rápido de um cluster	Use o botão Criar cluster no painel Resumo. O nome do cluster não pode conter os caractere s <, >, $, | ou `(crase).	
Configuração de um tempo limite para o provisionamento spot	Defina um período de tempo limite para o provision amento de instâncias para cada frota no cluster.	
Perfis de serviço e perfil para o perfil de instância do Amazon EC2	O console não cria funções padrão; você deve criar funções com o console do IAM ou selecionar uma função do IAM já criada	
Visibilidade do cluster	No console do Amazon EMR, não é possível tornar um cluster visível para todos os usuários. Sua política do IAM determina o acesso ao cluster.	
Redes: configuração de sub-redes privadas	Você deve configurar endpoints do Amazon S3 e gateways NAT usando os respectivos consoles do Amazon S3 e do Amazon VPC.	

Recurso	Console	
Visualização consistente do Sistema de Arquivos do EMR (EMRFS CV)	Com o lançamento da read-after-write consistência forte do Amazon S3 em 1º de dezembro de 2020, você não precisa usar o EMRFS CV com seus clusters do EMR	
Depuração	Você pode depurar trabalhos usando a interface do usuário da aplicação na página de detalhes do cluster.	

[1] Você não pode criar ou editar clusters usando versões anteriores ao Amazon EMR 5.20.1 no console, mas todos os clusters existentes criados usando versões anteriores à 5.20.1 continuarão funcionando. Para criar e editar clusters com versões do Amazon EMR anteriores à 5.20.1, use a API ou a CLI. Você pode visualizar todos os clusters usando o console, mas os consoles criados antes da versão 5.20.1 podem não ser compatíveis com os recursos mais recentes.

Visualizando e pesquisando clusters

A tabela a seguir destaca como você pode usar o console do Amazon EMR para visualizar, visualizar e pesquisar clusters.

> **Note**
>
> A aplicação de um filtro de dados à lista de clusters consulta todo o banco de dados. Entretanto, ao inserir uma string de texto na caixa de pesquisa, a pesquisa se aplica somente aos resultados que a lista carregou no lado do cliente.

Recurso	Console	
Visualização de detalhes do cluster	Você pode selecionar o ID do cluster para visualizar	

Recurso	Console	
	os detalhes completos do cluster, como as opções de configuração, as interfaces do usuário de aplicações persistentes e os logs.	
Pesquisa de clusters	Use um único campo de pesquisa para inserir consultas de pesquisa de texto e para criar e aplicar filtros de dados como “Status = qualquer status ativo”.	
Descoberta de clusters com falha	Para pesquisar clusters com falha, aplique o filtro Status = Encerrado com erros.	

Visualizando ou editando detalhes do cluster

Recurso	Console	
Visualização das instâncias em seus grupos de instâncias e frotas de instâncias, em conjunto com opções de escalabilidade, provisionamento, redimensionamento e encerramento.	Veja as opções e os detalhes da instância na guia Instâncias. Veja as opções de encerramento na guia Propriedades.	
Visualização de interfaces do usuário, logs e configurações de aplicações	Veja as configurações do cluster na guia Configurações. Inicie uma interface	

Recurso	Console	
(interface do usuário do Apache Spark, servidor de histórico do Spark, interface do usuário do Tez, servidor de linha do tempo do YARN)	do usuário da aplicação dinâmica e persistente para visualizar os logs de uma aplicação na guia Aplicações.	
Exportação de um cluster para a CLI	Opção disponível nos menus de detalhes e de visualização da listagem de Ações do cluster como “Visualizar comando para clonar cluster”.	

Diferenças no trabalho com configurações de segurança

Recurso	Console	
Clonagem de configurações de segurança	✓	
Governança federada ao usar Trino e Apache Ranger	✓	
Uso de um perfil de runtime para envio de trabalhos a um cluster[1]	✓	
Autorização de acesso aos dados do Sistema de Arquivos do EMR (EMRFS)	Pontos de acesso Amazon S3	
	Perfis de runtime	

Recurso	Console	
AWS Lake Formation controles de acesso		

[1] Para transmitir um perfil durante o envio da etapa, o cluster deve usar uma configuração de segurança com uma política de permissões do IAM anexada para que o usuário possa transmitir somente os perfis aprovados e os trabalhos possam acessar os recursos do Amazon EMR. Para ter mais informações, consulte Funções de tempo de execução para Amazon EMR Steps.

Amazon EMR Studio

O Amazon EMR Studio é um ambiente de desenvolvimento integrado (IDE) baseado na Web para cadernos Jupyter totalmente gerenciados que são executados em clusters do Amazon EMR. Você pode configurar um EMR Studio para que sua equipe desenvolva, visualize e depure aplicativos escritos em R, Python, Scala e. PySpark O EMR Studio é integrado ao AWS Identity and Access Management (IAM) e ao Centro de Identidade do IAM para que os usuários possam fazer login usando suas credenciais corporativas.

É possível criar um EMR Studio gratuitamente. As cobranças aplicáveis para o armazenamento do Amazon S3 e para os clusters do Amazon EMR se aplicam quando você usa o EMR Studio. Para obter detalhes e destaques do produto, consulte a página de serviços do Amazon EMR Studio.

Principais recursos do EMR Studio

O Amazon EMR Studio oferece os seguintes recursos:

- Autentique usuários com o AWS Identity and Access Management (IAM) ou o AWS IAM Identity Center com ou sem a propagação de identidade confiável e seu provedor de identidade empresarial.
- Acesse e execute clusters do Amazon EMR sob demanda para executar trabalhos do caderno Jupyter.
- Conexão aos clusters do Amazon EMR no EKS para enviar trabalhos à medida que o trabalho é executado.
- Navegação e salvamento de cadernos de exemplo. Para obter mais informações sobre exemplos de notebooks, consulte o repositório de exemplos de notebooks GitHub do EMR Studio.
- Analise dados usando Python, Spark Scala PySpark, Spark R ou SparkSQL e instale kernels e bibliotecas personalizados.
- Colaboração em tempo real com outros usuários no mesmo Workspace. Para ter mais informações, consulte Configuração da colaboração no Workspace.
- Uso do SQL Explorer do EMR Studio para navegar em seu catálogo de dados, executar consultas SQL e fazer download de resultados antes do trabalho com os dados em um caderno.
- Execução de cadernos parametrizados como parte dos fluxos de trabalho programados com uma ferramenta de orquestração, como o Apache Airflow ou o Amazon Managed Workflows for Apache

Airflow. Para obter mais informações, consulte Orchestrating analytics jobs on EMR Notebooks using MWAA no blog de Big Data da AWS.

- Vincule repositórios de código, como GitHub e. BitBucket
- Rastreamento e depuração de trabalhos usando o servidor de histórico do Spark, a interface do usuário do Tez ou o servidor de linha do tempo do YARN.

O EMR Studio também é elegível para a HIPAA e é certificado pela HITRUST CSF e pelo SOC 2. Para obter mais informações sobre a conformidade com a HIPAA para serviços da AWS, consulte https://aws.amazon.com/compliance/hipaa-compliance/. Para saber mais sobre a conformidade da HITRUST CSF para serviços da AWS, consulte https://aws.amazon.com/compliance/hitrust/. Para obter mais informações sobre outros programas de conformidade para serviços da AWS, consulte Serviços da AWS no escopo por programa de conformidade.

Histórico de recursos do Amazon EMR Studio

Esta tabela lista as atualizações na funcionalidade de ajuste de escala gerenciado do Amazon EMR.

Data de lançamento	Recurso
5 de janeiro de 2024	Foi adicionado suporte para o EMR Studio em AWS GovCloud (Leste dos EUA) e AWS GovCloud (Oeste dos EUA).
26 de novembro de 2023	Foi adicionado suporte à propagação de identidade confiável para o EMR Studio com a autenticação do Centro de Identidade do IAM.
26 de outubro de 2023	Capacidade adicional de criar uma aplicação do EMR Serverless com capacidade interativa.
28 de fevereiro de 2023	Adição de suporte para chaves gerenciadas pelo cliente do AWS KMS para o armazenamento de logs de aplicações para aplicações do EMR Sem Servidor.
23 de fevereiro de 2023	Adição da criação de perfil do IAM com um clique para envio de trabalhos do EMR Sem Servidor. Adição de pesquisa do ECR para quando você seleciona uma imagem personalizada para aplicações do EMR Sem Servidor.

Data de lançamento	Recurso
27 de janeiro de 2023	Os cadernos de execução descentralizados podem rastrear o progresso da execução de cada célula com a mágica `%execute_notebook` .
23 de janeiro de 2023	As aplicações persistentes foram otimizadas para a obtenção de tempos de inicialização mais rápidos.

Como o Amazon EMR Studio funciona

Um Amazon EMR Studio é um recurso do Amazon EMR criado para uma equipe de usuários. Cada Studio corresponde a um ambiente de desenvolvimento integrado que é independente e baseado na Web para cadernos Jupyter executados em clusters do Amazon EMR. Os usuários fazem login em um Studio usando credenciais corporativas.

Cada EMR Studio criado usa os seguintes recursos da AWS:

- Uma Amazon Virtual Private Cloud (VPC) com sub-redes: os usuários executam kernels e aplicações do Studio no Amazon EMR e clusters do Amazon EMR no EKS na VPC especificada. Um EMR Studio pode se conectar a qualquer cluster nas sub-redes especificadas na criação do Studio.
- Políticas de permissões e perfis do IAM: para gerenciar as permissões de usuários, você cria políticas de permissões do IAM que são anexadas à identidade do IAM de um usuário ou a um perfil de usuário. O EMR Studio também usa um perfil de serviço do IAM e grupos de segurança para interoperar com outros serviços da AWS. Para obter mais informações, consulte Controle de acesso e Defina grupos de segurança para controlar o tráfego de rede do EMR Studio.
- Grupos de segurança: o EMR Studio usa grupos de segurança para estabelecer um canal de rede seguro entre o Studio e um cluster do EMR.
- Um local de backup do Amazon S3: o EMR Studio salva o trabalho do caderno em um local do Amazon S3.

As seguintes etapas descrevem como criar e administrar um EMR Studio:

1. Crie um Studio em sua Conta da AWS com a autenticação do IAM ou do Centro de Identidade do IAM. Para obter instruções, consulte Configurar um Amazon EMR Studio.

2. Atribua usuários e grupos ao seu Studio. Use políticas de permissões para definir permissões detalhadas para cada usuário. Para obter mais informações, consulte o tópico Atribuir e gerenciar usuários do EMR Studio
3. Comece a monitorar as ações do EMR Studio com eventos do AWS CloudTrail. Para obter mais informações, consulte Monitore as ações do Amazon EMR Studio.
4. Forneça mais opções de cluster aos usuários do Studio com modelos de cluster e endpoints gerenciados do Amazon EMR no EKS.

Autenticação e login do usuário

O Amazon EMR Studio oferece suporte a dois modos de autenticação: o modo de autenticação do IAM e o modo de autenticação do Centro de Identidade do IAM. O modo do IAM usa o AWS Identity and Access Management (IAM), enquanto o modo do Centro de Identidade do IAM usa o AWS IAM Identity Center. Ao criar um EMR Studio, você escolhe o modo de autenticação para todos os usuários desse Studio.

Modo de autenticação do IAM

Com o modo de autenticação do IAM, você pode usar a autenticação do IAM ou a federação do IAM.

A autenticação do IAM permite gerenciar identidades do IAM, como usuários, grupos e perfis no IAM. Você concede aos usuários acesso a um Studio com as políticas de permissões do IAM e o controle de acesso por atributos (ABAC).

A federação do IAM permite estabelecer confiança entre um provedor de identidades (IdP) terceirizado e a AWS para que você possa gerenciar identidades de usuários por meio do seu IdP.

Modo de autenticação do Centro de Identidade do IAM

O modo de autenticação do Centro de Identidade do IAM permite conceder aos usuários o acesso federado a um EMR Studio. Você pode usar o Centro de Identidade do IAM para autenticar usuários e grupos do diretório do Centro de Identidade do IAM, do diretório corporativo existente ou de um IdP externo, como o Azure Active Directory (AD). Em seguida, você gerencia os usuários com o seu provedor de identidades (IdP).

O EMR Studio oferece suporte ao uso dos seguintes provedores de identidades para o Centro de Identidade do IAM:

- AWS Managed Microsoft AD e Active Directory autogerenciado: para obter mais informações, consulte Connect to your Microsoft AD directory.
- Provedores baseados em SAML: para obter uma lista completa, consulte Supported identity providers.
- O diretório do Centro de Identidade do IAM: para obter mais informações, consulte Gerenciamento de identidades no Centro de Identidade do IAM e Trusted identity propagation across applications no Guia do usuário do AWS IAM Identity Center.

Como a autenticação afeta o login e a atribuição de usuários

O modo de autenticação escolhido para o EMR Studio afeta como os usuários fazem login em um Studio, como você atribui um usuário a um Studio e como você autoriza (concede permissões) aos usuários para executar ações, como a criação de novos clusters do Amazon EMR.

A tabela a seguir resume os métodos de login do EMR Studio de acordo com o modo de autenticação.

Opções de login do EMR Studio por modo de autenticação

Modo de autenticação	Método de login	Descrição
• IAM (autenticação e federação) • IAM Identity Center	URL do EMR Studio	Os usuários fazem login em um Studio usando o URL de acesso ao Studio. Por exemplo, `https://xxxxxxxxxxxxxxxxxxxxxxxx.emrstudio-prod.us-east-1.amazonaws.com` . Os usuários inserem as credenciais do IAM quando você usa a autenticação do IAM. Quando você usa a federação do IAM ou o Centro de Identidade do IAM, o EMR Studio redireciona os usuários para o URL de login do seu provedor de identidades para a inserção das credenciais. No contexto da federação de identidades, esta opção de login é chamada de login iniciado com base no provedor de serviços (SP).

Modo de autenticação	Método de login	Descrição
• IAM (federação) • IAM Identity Center	Portal do provedor de identidades (IdP)	Os usuários fazem login no portal do seu provedor de identidades, como o portal do Azure, e iniciam o console do Amazon EMR. Após iniciarem o console do Amazon EMR, os usuários selecionam e abrem um Studio pela lista Studios. Você também pode configurar o EMR Studio como uma aplicação da SAML para que os usuários possam fazer login em um Studio específico usando o portal do seu provedor de identidades. Para obter instruções, consulte Para configurar um EMR Studio como uma aplicação da SAML em seu portal do IdP. No contexto da federação de identidades, esta opção de login é chamada de login iniciado com base no provedor de identidades (IdP).
• IAM (autenticação)	AWS Management Console	Os usuários fazem login no AWS Management Console usando as credenciais do IAM e abrem um Studio pela lista Studios no console do Amazon EMR.

A tabela a seguir descreve a atribuição e a autorização de usuários para o EMR Studio pelo modo de autenticação.

Atribuição e autorização de usuários do EMR Studio pelo modo de autenticação

Modo de autenticação	Atribuição de usuários	Autorização de usuários
IAM (autenticação e federação)	Permita a ação `CreateStudioPresignedUrl` em uma política de permissões do IAM	Defina políticas de permissões do IAM que permitem determinadas ações do EMR Studio.

Modo de autenticação	Atribuição de usuários	Autorização de usuários
	anexada a uma identidade do IAM (usuário, grupo ou perfil). Para usuários federados, permita a ação `CreateStudioPresignedUrl` em um IAM na política de permissões configurada para o perfil do IAM que é usado para a federação. Use o controle de acesso por atributo (ABAC) para especificar o Studio ou os Studios que o usuário pode acessar. Para obter instruções, consulte Atribuir um usuário ou grupo a um EMR estúdio.	Para usuários nativos, anexe a política de permissões do IAM a uma identidade do IAM (usuário, grupo ou perfil). Para usuários federados, permita as ações do Studio na política de permissões configurada para o perfil do IAM que é usado para a federação. Para obter mais informações, consulte Configurar permissões de usuário do EMR Studio para Amazon EC2 ou Amazon EKS.
IAM Identity Center	Para Studios criados com `IdCUserAssignment` definido como `REQUIRED`, mapeie os usuários para o Studio com uma política de sessão especificada. Para obter mais informações, consulte Atribuir um usuário ou grupo a um EMR estúdio. Para Studios criados com `IdCUserAssignment` definido como `OPTIONAL`, qualquer usuário ou grupo do Centro de Identidade pode acessar o Studio.	Opcional: defina políticas de sessão do IAM que permitam determinadas ações do EMR Studio. Mapeie uma política de sessão para um usuário ao atribuir o usuário a um Studio. Para obter mais informações, consulte Permissões de usuários para o modo de autenticação do Centro de Identidade do IAM.

Controle de acesso

No Amazon EMR Studio, você configura a autorização (permissões) de usuários com as políticas baseadas em identidade do AWS Identity and Access Management (IAM). Nessas políticas, você especifica as ações e os recursos permitidos, bem como as condições sob as quais as ações são permitidas.

Permissões de usuários para o modo de autenticação do IAM

Para definir as permissões de usuários ao usar a autenticação do IAM para o EMR Studio, você permite ações, como `elasticmapreduce:RunJobFlow`, em uma política de permissões do IAM. Você pode criar uma ou mais políticas de permissões para usar. Por exemplo, é possível criar uma política básica, que não permita que um usuário crie novos clusters do Amazon EMR, e outra política que permita a criação de clusters. Para obter uma lista de todas as ações do Studio, consulte AWS Identity and Access Management permissões para usuários do EMR Studio.

Permissões de usuários para o modo de autenticação do Centro de Identidade do IAM

Ao usar a autenticação do Centro de Identidade do IAM, você cria um único perfil de usuário do EMR Studio. O perfil de usuário corresponde a um perfil do IAM dedicado que um Studio assume quando um usuário faz login.

Você anexa políticas de sessão do IAM ao perfil de usuário do EMR Studio. Uma política de sessão é um tipo especial de política de permissões do IAM que limita o que um usuário federado pode fazer durante uma sessão de login do Studio. As políticas de sessão possibilitam definir permissões específicas para um usuário ou para um grupo sem a necessidade de criar diversos perfil de usuário para o EMR Studio.

Ao atribuir usuários e grupos a um Studio, você mapeia uma política de sessão para esse usuário ou grupo para a aplicação de permissões detalhadas. Você também pode atualizar a política de sessão de um usuário ou de um grupo a qualquer momento. O Amazon EMR armazena cada mapeamento de política de sessão criado.

Para obter mais informações sobre as políticas de sessão, consulte Políticas e permissões no Guia do usuário do AWS Identity and Access Management.

Workspaces

Os Workspaces são os principais componentes básicos do Amazon EMR Studio. Para organizar os cadernos, os usuários criam um ou mais Workspaces em um Studio. Para obter mais informações, consulte Compreensão das noções básicas do Workspace.

Semelhante aos espaços de trabalho no JupyterLab, um Workspace preserva o estado de trabalho do caderno. No entanto, a interface do usuário do Workspace amplia a interface do JupyterLab de código aberto com ferramentas adicionais para permitir que você crie e anexe clusters do EMR, execute trabalhos, explore cadernos de exemplo e vincule repositórios Git.

A seguinte lista inclui os principais recursos dos Workspaces do EMR Studio:

- A visibilidade do Workspace é baseada no Studio. Os Workspaces criados em um Studio não são visíveis em outros Studios.
- Por padrão, um Workspace é compartilhado e pode ser visualizado por todos os usuários do Studio. No entanto, somente um usuário pode abrir e trabalhar em um Workspace por vez. Para trabalhar simultaneamente com outros usuários, é possível realizar a Configuração da colaboração no Workspace.
- Você pode colaborar simultaneamente com outros usuários em um Workspace ao habilitar a colaboração no Workspace. Para obter mais informações, consulte Configuração da colaboração no Workspace.
- Os cadernos em um Workspace compartilham o mesmo cluster do EMR para a execução de comandos. Você pode anexar um Workspace a um cluster do Amazon EMR em execução no Amazon EC2 ou a um cluster virtual e a um endpoint gerenciado do Amazon EMR no EKS.
- Os Workspaces podem ser alternados para outra zona de disponibilidade associada às sub-redes de um Studio. Você pode interromper e reiniciar um Workspace para solicitar o processo de failover. Ao reiniciar um Workspace, o EMR Studio inicia o Workspace em uma zona de disponibilidade diferente na VPC do Studio quando o Studio está configurado com acesso a diversas zonas de disponibilidade. Se o Studio tiver somente uma zona de disponibilidade, o EMR Studio tentará iniciar o Workspace em uma sub-rede diferente. Para obter mais informações, consulte Resolução de problemas de conectividade do Workspace.
- Um Workspace pode se conectar a clusters em qualquer uma das sub-redes associadas a um Studio.

Para obter mais informações sobre como criar e configurar Workspaces do EMR Studio, consulte Compreensão das noções básicas do Workspace.

Armazenamento de cadernos no Amazon EMR Studio

Quando você usa um Workspace, o EMR Studio salva automaticamente as células em arquivos de cadernos em uma cadência regular no local do Amazon S3 associado ao seu Studio. Esse processo de backup preserva o trabalho entre as sessões para que você possa voltar a ele mais tarde sem a necessidade de confirmar as alterações em um repositório Git. Para obter mais informações, consulte Salvamento de conteúdo do Workspace.

Quando você exclui um arquivo de caderno de um Workspace, o EMR Studio exclui a versão de backup do Amazon S3 para você. No entanto, se você excluir um Workspace sem primeiro excluir os arquivos do cadernos, estes arquivos permanecerão no Amazon S3 e continuarão a acumular cobranças de armazenamento. Para saber mais, consulte Exclusão de um Workspace e de arquivos de cadernos.

Considerações sobre o EMR Studio

Considerações

Considere o seguinte ao trabalhar com o EMR Studio:

- O EMR Studio está disponível da seguinte forma: Regiões da AWS
  - Leste dos EUA (Ohio) (us-east-2)
  - Leste dos EUA (Norte da Virgínia) (us-east-1)
  - Oeste dos EUA (Norte da Califórnia) (us-west-1)
  - Oeste dos EUA (Oregon) (us-west-2)
  - África (Cidade do Cabo) (af-south-1)
  - Ásia-Pacífico (Hong Kong) (ap-east-1)
  - Ásia-Pacífico (Jacarta) (ap-southeast-3) *
  - Ásia-Pacífico (Melbourne) (ap-southeast-4) *
  - Ásia-Pacífico (Mumbai) (ap-south-1)
  - Ásia-Pacífico (Osaka) (ap-northeast-3) *
  - Ásia-Pacífico (Seul) (ap-northeast-2)
  - Ásia-Pacífico (Singapura) (ap-southeast-1)
  - Ásia-Pacífico (Sydney) (ap-southeast-2)
  - Ásia Pacific (Tóquio) (ap-northeast-1)

- Canadá (Central) (ca-central-1)
- Europa (Frankfurt) (eu-central-1)
- Europa (Irlanda) (eu-west-1)
- Europa (Londres) (eu-west-2)
- UE (Milão) (eu-south-1)
- Europa (Paris) (eu-west-3)
- Europa (Espanha) (eu-south-2)
- UE (Estocolmo) (eu-north-1)
- Europa (Zurique) (eu-central-2) *
- Israel (Tel Aviv) (il-central-1)*
- Oriente Médio (EAU) (me-central-1) *
- América do Sul (São Paulo) (sa-east-1)
- AWS GovCloud (Leste dos EUA) (gov-us-east-1)
- AWS GovCloud (Oeste dos EUA) (gov-us-west-1)

  * A interface ativa do Spark não é compatível com essas regiões.
- Para permitir que os usuários provisionem novos clusters do EMR em execução no Amazon EC2 para um Workspace, você pode associar um EMR Studio a um conjunto de modelos de cluster. Os administradores podem definir modelos de cluster com o Service Catalog e escolher se um usuário ou um grupo pode acessar os modelos de cluster, ou nenhum modelo de cluster, em um Studio.
- Ao definir permissões de acesso aos arquivos do notebook armazenados no Amazon S3 ou ler segredos AWS Secrets Manager, use a função de serviço do Amazon EMR. As políticas de sessão não são compatíveis com estas permissões.
- Você pode criar diversos EMR Studios para controlar o acesso a clusters do EMR em diferentes VPCs.
- Use o AWS CLI para configurar o Amazon EMR em clusters EKS. Em seguida, é possível usar a interface do Studio para anexar clusters a Workspaces com um endpoint gerenciado para executar trabalhos de cadernos.
- Há outras considerações ao usar a propagação de identidade confiável com o Amazon EMR que também se aplicam ao EMR Studio. Para ter mais informações, consulte Considerações e limitações para a Amazon EMR com a integração do Identity Center.
- O EMR Studio não oferece suporte aos seguintes comandos mágicos do Python:
  - `%alias`

  - `%alias_magic`
  - `%automagic`
  - `%macro`
  - `%%js`
  - `%%javascript`
  - Modificar `proxy_user` usando `%configure`
  - Modificar `KERNEL_USERNAME` usando `%env` ou `%set_env`
- O Amazon EMR em clusters EKS não oferece suporte a SparkMagic comandos para o EMR Studio.
- Para escrever instruções do Scala com várias linhas em células de cadernos, certifique-se de que todas as linhas, exceto a última, terminem com um ponto final. O exemplo a seguir usa a sintaxe adequada para instruções do Scala com várias linhas.

```
val df = spark.sql("SELECT * from table_name).
        filter("col1=='value'").
        limit(50)
```

- Para aumentar a segurança das aplicações fora do console que podem ser usadas com o Amazon EMR, os domínios de hospedagem das aplicações são registrados na Public Suffix List (PSL). Exemplos desses domínios de hospedagem incluem os seguintes: `emrstudio-prod.us-east-1.amazonaws.com`, `emrnotebooks-prod.us-east-1.amazonaws.com`, `emrappui-prod.us-east-1.amazonaws.com`. Para maior segurança, se precisar definir cookies confidenciais no nome de domínio padrão, recomendamos que você use cookies com um prefixo `__Host-`. Isso ajuda a defender seu domínio contra tentativas de falsificação de solicitação entre sites (CSRF). Para obter mais informações, consulte a página Set-Cookie em Mozilla Developer Network.

Problemas conhecidos

- Um EMR Studio que usa o Centro de Identidade do IAM com a propagação de identidade confiável habilitada só pode se associar a clusters do EMR que também usam a propagação de identidade confiável.
- Certifique-se de desativar as ferramentas de gerenciamento de proxy, como FoxyProxy ou SwitchyOmega, no navegador antes de criar um Studio. Os proxies ativos podem causar erros quando você escolhe Criar Studio e resultar em uma mensagem de erro de falha de rede.

- Os kernels executados em clusters do Amazon EMR no EKS podem falhar ao iniciar devido a problemas de tempo limite. Se você encontrar um erro ou problema ao iniciar o kernel, feche o arquivo de caderno, encerre o kernel e reabra o arquivo de caderno.
- A operação Reiniciar kernel não funciona conforme o esperado quando você usa um cluster do Amazon EMR no EKS. Após selecionar Reiniciar kernel, atualize o Workspace para que a reinicialização entre em vigor.
- Se um Workspace não estiver anexado a um cluster, uma mensagem de erro será exibida quando um usuário do Studio abrir um arquivo de caderno e tentar selecionar um kernel. Você pode ignorar essa mensagem de erro ao escolher OK, mas deve anexar o Workspace a um cluster e selecionar um kernel antes de poder executar o código do caderno.
- Ao usar o Amazon EMR 6.2.0 com uma configuração de segurança para definir a segurança do cluster, a interface do Workspace aparece em branco e não funciona conforme o esperado. Recomendamos usar uma versão diferente do Amazon EMR com suporte, se desejar configurar a criptografia de dados ou a autorização do Amazon S3 para o EMRFS em um cluster. O EMR Studio funciona com as versões 5.32.0 (série 5.x) e 6.2.0 (série 6.x) e superiores do Amazon EMR.
- Ao realizar a Depure a Amazon em EMR execução com trabalhos da Amazon EC2, os links para a interface do usuário do Spark no cluster podem não funcionar ou não aparecer. Para gerar os links novamente, crie uma nova célula de caderno e execute o comando `%%info`.
- O Jupyter Enterprise Gateway não limpa os kernels ociosos no nó primário de um cluster nas seguintes versões de liberação do Amazon EMR: 5.32.0, 5.33.0, 6.2.0 e 6.3.0. Os kernels ociosos consomem recursos de computação e podem causar falhas em clusters de longa execução. Você pode configurar a limpeza de kernels ociosos para o Jupyter Enterprise Gateway usando o script de exemplo a seguir. É possível Conecte-se ao nó primário usando SSH ou enviar o script como uma etapa. Para obter mais informações, consulte Run commands and scripts on an Amazon EMR cluster.

```
#!/bin/bash
sudo tee -a /emr/notebook-env/conf/jupyter_enterprise_gateway_config.py << EOF
c.MappingKernelManager.cull_connected = True
c.MappingKernelManager.cull_idle_timeout = 10800
c.MappingKernelManager.cull_interval = 300
EOF
sudo systemctl daemon-reload
sudo systemctl restart jupyter_enterprise_gateway
```

- Quando você usa uma política de encerramento automático com as versões 5.32.0, 5.33.0, 6.2.0 ou 6.3.0 do Amazon EMR, o Amazon EMR marca um cluster como ocioso e pode encerrá-lo

automaticamente mesmo se você tiver um kernel do Python3 ativo. Isso ocorre porque a execução de um kernel do Python3 não envia um trabalho do Spark no cluster. Para usar o encerramento automático com um kernel do Python3, recomendamos usar a versão 6.4.0 ou as versões posteriores do Amazon EMR. Para obter mais informações sobre o encerramento automático, consulte Usar uma política de término automático.

- Quando você costuma `%%display` exibir um Spark DataFrame em uma tabela, tabelas muito largas podem ficar truncadas. Você pode clicar com o botão direito do mouse na saída e selecionar Criar nova visualização para a saída para obter uma visualização da saída com rolagem.
- Iniciar um kernel baseado em Spark, como PySpark Spark ou SparkR, inicia uma sessão do Spark, e executar uma célula em um notebook coloca as tarefas do Spark em fila nessa sessão. Quando você interrompe uma célula em execução, o trabalho do Spark continua a ser executado. Para interromper o trabalho do Spark, você deve usar a interface do usuário do Spark no cluster. Para obter instruções sobre como se conectar à interface do usuário do Spark, consulte Depure aplicativos e trabalhos com EMR o Studio.
- Usar o Amazon EMR Studio Workspaces como usuário raiz em um Conta da AWS causa um erro. `403: Forbidden` Isso ocorre porque a configuração do Jupyter Enterprise Gateway no Amazon EMR não permite acesso ao usuário raiz. Recomendamos que você não use o usuário root para suas tarefas diárias. Para outras opções de autenticação, consulte AWS Identity and Access Management o Amazon EMR.

Limitações de recursos

O Amazon EMR Studio não oferece suporte aos seguintes recursos do Amazon EMR:

- Anexação e execução de trabalhos em clusters do EMR com uma configuração de segurança que especifica a autenticação do Kerberos.
- Clusters com vários nós primários.
- Clusters que usam instâncias do Amazon EC2 com base no AWS Graviton2 para versões 6.x do Amazon EMR inferiores a 6.9.0 e versões 5.x inferiores a 5.36.1

Os recursos a seguir não são compatíveis com um Studio que usa a propagação de identidade confiável:

- Criação de clusters do EMR sem um modelo.

- Uso de aplicações do EMR Sem Servidor.
- Execução de clusters do Amazon EMR no EKS.
- Uso de um perfil de runtime.
- Ativação da colaboração do SQL Explorer ou do Workspace.

Limites de serviço para o EMR Studio

A tabela a seguir exibe os limites de serviço para o EMR Studio.

Item	Limite
EMR Studios	Máximo de 100 por AWS conta
Subredes	Máximo de cinco associações para cada EMR Studio
Grupos do Centro de Identidade do IAM	Máximo de cinco atribuições para cada EMR Studio
Usuários do Centro de Identidade do IAM	Máximo de cem atribuições para cada EMR Studio

Práticas recomendadas para VPC e para sub-rede

Use as seguintes melhores práticas para configurar uma Amazon Virtual Private Cloud (Amazon VPC) com sub-redes para o EMR Studio:

- Você pode especificar, no máximo, cinco sub-redes em sua VPC para serem associadas ao Studio. Recomendamos fornecer várias sub-redes em diferentes zonas de disponibilidade para oferecer suporte à disponibilidade do Workspace e disponibilizar aos usuários do Studio o acesso a clusters em diferentes zonas de disponibilidade. Para saber mais sobre como trabalhar com VPCs, sub-redes e zonas de disponibilidade, consulte VPCs e sub-redes no Guia do usuário da Amazon Virtual Private Cloud .
- As sub-redes especificadas deverão ser capazes de se comunicar entre si.
- Para permitir que os usuários vinculem um Workspace a repositórios Git hospedados publicamente, você deve especificar somente sub-redes privadas que tenham acesso à Internet

através da conversão de endereços de rede (NAT). Para obter mais informações sobre como configurar uma sub-rede privada para o Amazon EMR, consulte Sub-redes privadas.

- Ao usar o Amazon EMR no EKS com o EMR Studio, deve haver, no mínimo, uma sub-rede em comum entre o Studio e o cluster do Amazon EKS usado para registrar um cluster virtual. Caso contrário, o endpoint gerenciado não aparecerá como uma opção nos Workspaces do Studio. Você pode criar um cluster do Amazon EKS e associá-lo a uma sub-rede que pertence ao Studio ou criar um Studio e especificar as sub-redes do seu cluster do EKS.
- Se você planeja usar o Amazon EMR no EKS com o EMR Studio, escolha a mesma VPC dos nós de processamento do cluster do Amazon EKS.

Requisitos de cluster para o Amazon EMR Studio

Clusters do Amazon EMR em execução no Amazon EC2

Todos os clusters do Amazon EMR em execução no Amazon EC2 criados para um Workspace do EMR Studio devem atender aos requisitos apresentados a seguir. Os clusters criados usando a interface do EMR Studio atendem automaticamente a esses requisitos.

- O cluster deve usar as versões 5.32.0 (Amazon EMR de série 5.x) ou 6.2.0 (Amazon EMR de série 6.x) ou posteriores do Amazon EMR. Você pode criar um cluster usando o console do Amazon EMR, ou SDK AWS Command Line Interface, e depois anexá-lo a um espaço de trabalho do EMR Studio. Os usuários do Studio também podem provisionar e anexar clusters ao criar ou trabalhar em um Workspace do Amazon EMR. Para ter mais informações, consulte Anexar uma computação a um espaço de trabalho do EMR Studio.
- O cluster deve estar em uma Amazon Virtual Private Cloud. A plataforma EC2-Classic não é compatível.
- O cluster deve ter o Spark, o Livy e o Jupyter Enterprise Gateway instalados. Se você planeja usar o cluster para o SQL Explorer, deverá instalar o Presto e o Spark.
- Para usar o SQL Explorer, o cluster deve usar a versão 5.34.0, ou versões posteriores, ou a versão 6.4.0, ou versões posteriores, do Amazon EMR e ter o Presto instalado. Se quiser especificar o AWS Glue Data Catalog como metastore do Hive para o Presto, você deve configurá-lo no cluster. Para obter mais informações, consulte Using Presto with the AWS Glue Data Catalog.
- O cluster deve estar em uma sub-rede privada com conversão de endereços de rede (NAT) para usar repositórios Git hospedados publicamente com o EMR Studio.

Recomendamos as configurações de cluster apresentadas a seguir ao trabalhar com o EMR Studio.

- Defina o modo de implantação das sessões do Spark para o modo de cluster. O modo de cluster coloca os processos principais de aplicações nos nós centrais e não no nó primário de um cluster. Isso alivia o nó primário de possíveis pressões de memória. Para obter mais informações, consulte o tópico de Visão geral do modo de cluster na documentação do Apache Spark.
- Altere o tempo limite do Livy do padrão de uma hora para seis horas, como no exemplo de configuração apresentado a seguir.

```
{
    "classification":"livy-conf",
        "Properties":{
            "livy.server.session.timeout":"6h",
            "livy.spark.deploy-mode":"cluster"
        }
}
```

- Crie diversas frotas de instâncias com até 30 instâncias e selecione vários tipos de instâncias em sua frota de instâncias spot. Por exemplo, é possível especificar os seguintes tipos de instâncias otimizadas para memória para workloads do Spark: r5.2x, r5.4x, r5.8x, r5.12x, r5.16x, r4.2x, r4.4x, r4.8x, r4.12 etc. Para ter mais informações, consulte Configurar frotas de instâncias.
- Use a estratégia de alocação com capacidade otimizada para instâncias spot com a finalidade de ajudar o Amazon EMR a fazer seleções de instâncias eficazes com base em insights de capacidade em tempo real do Amazon EC2. Para ter mais informações, consulte Estratégia de alocação para frotas de instâncias.
- Habilite o ajuste de escala gerenciado em seu cluster. Defina o parâmetro máximo de nós centrais para a capacidade persistente mínima que você planeja usar, e configure a escalabilidade em uma frota de tarefas bem diversificada que é executada em instâncias spot para economizar custos. Para ter mais informações, consulte Usando escalabilidade gerenciada na Amazon EMR.

Também recomendamos manter o bloqueio de acesso público do Amazon EMR habilitado e restringir o tráfego SSH de entrada para origens confiáveis. O acesso de entrada a um cluster permite que os usuários executem cadernos no cluster. Para obter mais informações, consulte Usando a Amazon, EMR bloqueie o acesso público e Controle do tráfego de rede com grupos de segurança.

Clusters do Amazon EMR no EKS

Além dos clusters do EMR em execução no Amazon EC2, você pode configurar e gerenciar clusters do Amazon EMR no EKS para o EMR Studio usando a AWS CLI. Configure os clusters do Amazon EMR no EKS usando as seguintes diretrizes:

- Crie um endpoint HTTPS gerenciado para o cluster do Amazon EMR no EKS. Os usuários anexam um Workspace a um endpoint gerenciado. O cluster do Amazon Elastic Kubernetes Service (EKS) usado para registrar um cluster virtual deve ter uma sub-rede privada para oferecer suporte a endpoints gerenciados.
- Use um cluster do Amazon EKS com, no mínimo, uma sub-rede privada e com conversão de endereços de rede (NAT) quando desejar usar repositórios Git hospedados publicamente.
- Evite usar ARM de AMIs do Amazon Linux otimizadas para o Amazon EKS, que não são compatíveis com os endpoints gerenciados pelo Amazon EMR no EKS.
- Evite usar clusters AWS Fargate somente do Amazon EKS, que não são compatíveis.

Configurar o Amazon EMR Studio

Esta seção é para administradores do EMR Studio. Ele aborda como configurar um EMR Studio para sua equipe e fornece instruções para tarefas como atribuição de usuários e grupos, configuração de modelos de cluster e otimização do Apache Spark for Studio. EMR

Tópicos

- Permissões de administrador para criar e gerenciar um EMR Studio
- Configurar um Amazon EMR Studio
- Gerencie um Amazon EMR Studio
- Criptografando cadernos e arquivos do espaço de trabalho do EMR Studio
- Defina grupos de segurança para controlar o tráfego de rede do EMR Studio
- Crie AWS CloudFormation modelos para o Amazon EMR Studio
- Estabelecimento de acesso e de permissões para repositórios baseados em Git
- Otimize as tarefas do Spark no Studio EMR

Permissões de administrador para criar e gerenciar um EMR Studio

As IAM permissões descritas nesta página permitem que você crie e gerencie um EMR estúdio. Para obter informações detalhadas sobre cada permissão obrigatória, consulte Permissões necessárias para gerenciar um EMR estúdio.

Permissões necessárias para gerenciar um EMR estúdio

A tabela a seguir lista as operações relacionadas à criação e ao gerenciamento de um EMR Studio. A tabela também exibe as permissões necessárias para cada operação.

> **Note**
>
> Você só precisa das `SessionMapping` ações do IAM Identity Center e do Studio ao usar o modo de autenticação do IAM Identity Center.

Permissões para criar e gerenciar um EMR estúdio

Operation	Permissões
Criar um Studio	`"elasticmapreduce:CreateStudio", "sso:CreateApplication", "sso:PutApplicationAuthenticationMethod", "sso:PutApplicationGrant", "sso:PutApplicationAccessScope", "sso:PutApplicationAssignmentConfiguration", "iam:PassRole"`
Descrever um Studio	`"elasticmapreduce:DescribeStudio", "sso:GetManagedApplicationInstance"`
Listar Studios	`"elasticmapreduce:ListStudios"`
Excluir um Studio	`"elasticmapreduce:DeleteStudio", "sso:DeleteApplication",`

Operation	Permissões
	`"sso:DeleteApplicationAuthentica tionMethod", "sso:DeleteApplicationAccessScope", "sso:DeleteApplicationGrant"`
Additional permissions required when you use IAM Identity Center mode	
Atribuir usuários ou grupos a um Studio	`"elasticmapreduce:CreateStudioSessio nMapping", "sso:GetProfile", "sso:ListDirectoryAssociations", "sso:ListProfiles", "sso:AssociateProfile", "sso-directory:SearchUsers", "sso-directory:SearchGroups", "sso-directory:DescribeUser", "sso-directory:DescribeGroup", "sso:ListInstances", "sso:CreateApplicationAssignment", "sso:DescribeInstance", "organizations:DescribeOrga nization", "organizations:ListDelegatedAdmini strators", "sso:CreateInstance", "sso:DescribeRegisteredRegions", "sso:GetSharedSsoConfiguration", "iam:ListPolicies"`
Recuperar detalhes de atribuição do Studio para um usuário ou um grupo específico	`"sso-directory:SearchUsers", "sso-directory:SearchGroups", "sso-directory:DescribeUser", "sso-directory:DescribeGroup", "sso:DescribeApplication", "elasticmapreduce:GetStudioSessio nMapping"`

Operation	Permissões
Listar todos os usuários e os grupos atribuídos a um Studio	`"elasticmapreduce:ListStudioSessionMappings"`
Atualizar a política de sessão anexada a um usuário ou a um grupo atribuído a um Studio	`"sso-directory:SearchUsers", "sso-directory:SearchGroups", "sso-directory:DescribeUser", "sso-directory:DescribeGroup", "sso:DescribeApplication", "sso:DescribeInstance", "elasticmapreduce:UpdateStudioSessionMapping"`
Remover um usuário ou um grupo de um Studio	`"elasticmapreduce:DeleteStudioSessionMapping", "sso-directory:SearchUsers", "sso-directory:SearchGroups", "sso-directory:DescribeUser", "sso-directory:DescribeGroup", "sso:ListDirectoryAssociations", "sso:GetProfile", "sso:DescribeApplication", "sso:DescribeInstance", "sso:ListProfiles", "sso:DisassociateProfile", "sso:DeleteApplicationAssignment", "sso:ListApplicationAssignments"`

Para criar uma política com permissões de administrador para o EMR Studio

1. Siga as instruções em Criação de IAM políticas para criar uma política usando um dos exemplos a seguir. As permissões de que você precisa dependem do seu modo de autenticação para o EMR Studio.

   Insira seus próprios valores para estes itens:

   - Substituir *<seu-recurso- >ARN* para especificar o Amazon Resource Name (ARN) do objeto ou objetos que a declaração abrange para seus casos de uso.

- Substituir *<region>* com o código de Região da AWS onde você planeja criar o Studio.
- Substituir *<aws-account_id>* com o ID da AWS conta do Studio.
- Substituir *<EMRStudio-Service-Role>* e *<EMRStudio-User-Role>* com os nomes da sua função de serviço do EMR Studio e da função de usuário do EMR Studio.

Example Exemplo de política: permissões de administrador quando você usa o modo de IAM autenticação

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:<region>:<aws-account-id>:studio/
*",
            "Action": [
                "elasticmapreduce:CreateStudio",
                "elasticmapreduce:DescribeStudio",
                "elasticmapreduce:DeleteStudio"
            ]
        },
        {
            "Effect": "Allow",
            "Resource": "<your-resource-ARN>",
            "Action": [
                "elasticmapreduce:ListStudios"
            ]
        },
        {
            "Effect": "Allow",
            "Resource": [
                "arn:aws:iam::<aws-account-id>:role/<EMRStudio-Service-Role>"
            ],
            "Action": "iam:PassRole"
        }
    ]
}
```

Example Exemplo de política: permissões de administrador ao usar o modo de autenticação do IAM Identity Center

> **Note**
>
> O Identity Center e o diretório do Identity Center APIs não suportam a especificação de um elemento ARN no recurso de uma declaração de IAM política. Para permitir o acesso ao IAM Identity Center e ao IAM Identity Center Directory, as permissões a seguir especificam todos os recursos, “Resource” :"*”, para ações do IAM Identity Center. Para obter mais informações, consulte Ações, recursos e chaves de condição para o IAM Identity Center Directory.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:<region>:<aws-account-id>:studio/
*",
            "Action": [
                "elasticmapreduce:CreateStudio",
                "elasticmapreduce:DescribeStudio",
                "elasticmapreduce:DeleteStudio",
                "elasticmapreduce:CreateStudioSessionMapping",
                "elasticmapreduce:GetStudioSessionMapping",
                "elasticmapreduce:UpdateStudioSessionMapping",
                "elasticmapreduce:DeleteStudioSessionMapping"
            ]
        },
        {
            "Effect": "Allow",
            "Resource": "<your-resource-ARN>",
            "Action": [
                "elasticmapreduce:ListStudios",
                "elasticmapreduce:ListStudioSessionMappings"
            ]
        },
        {
            "Effect": "Allow",
```

```
            "Resource": [
                "arn:aws:iam::<aws-account-id>:role/<EMRStudio-Service-Role>",
                "arn:aws:iam::<aws-account-id>:role/<EMRStudio-User-Role>"
            ],
            "Action": "iam:PassRole"
        },
        {
            "Effect": "Allow",
            "Resource": "*",
            "Action": [
                "sso:CreateApplication",
                "sso:PutApplicationAuthenticationMethod",
                "sso:PutApplicationGrant",
                "sso:PutApplicationAccessScope",
                "sso:PutApplicationAssignmentConfiguration",
                "sso:DescribeApplication",
                "sso:DeleteApplication",
                "sso:DeleteApplicationAuthenticationMethod",
                "sso:DeleteApplicationAccessScope",
                "sso:DeleteApplicationGrant",
                "sso:ListInstances",
                "sso:CreateApplicationAssignment",
                "sso:DeleteApplicationAssignment",
                "sso:ListApplicationAssignments",
                "sso:DescribeInstance",
                "sso:AssociateProfile",
                "sso:DisassociateProfile",
                "sso:GetProfile",
                "sso:ListDirectoryAssociations",
                "sso:ListProfiles",
                "sso-directory:SearchUsers",
                "sso-directory:SearchGroups",
                "sso-directory:DescribeUser",
                "sso-directory:DescribeGroup",
                "organizations:DescribeOrganization",
                "organizations:ListDelegatedAdministrators",
                "sso:CreateInstance",
                "sso:DescribeRegisteredRegions",
                "sso:GetSharedSsoConfiguration",
                "iam:ListPolicies"
            ]
        }
    ]
```

```
}
```

2. Anexe a política à sua IAM identidade (usuário, função ou grupo). Para obter instruções, consulte Adicionar e remover permissões de IAM identidade.

Configurar um Amazon EMR Studio

Conclua as etapas a seguir para configurar um Amazon EMR Studio.

Antes de começar

> **Note**
>
> Se você planeja usar o EMR Studio com o EMR Amazon EKS ativado, recomendamos que você primeiro configure o Amazon EMR on EKS for EMR Studio antes de configurar um Studio.

Antes de configurar um EMR Studio, verifique se você tem os seguintes itens:

- Um Conta da AWS. Para obter instruções, consulte Configuração do Amazon EMR.
- Permissões para criar e gerenciar um EMR estúdio. Para obter mais informações, consulte the section called “Permissões de administrador para criar um EMR estúdio”.
- Um bucket do Amazon S3 em que o EMR Studio pode fazer backup dos espaços de trabalho e dos arquivos do notebook em seu Studio. Para obter instruções, consulte Criação de um bucket no Guia do usuário do Amazon Simple Storage Service (S3).
- Se você quiser se conectar a um EKS cluster Amazon EMR on EC2 ou Amazon EMR on, ou usar repositórios Git, precisará de uma Amazon Virtual Private Cloud (VPC) para o Studio e de no máximo cinco sub-redes. Você não precisa VPC usar o EMR Studio com EMR Serverless. Para obter dicas sobre como configurar as redes, consulte Práticas recomendadas para VPC e para sub-rede.

Para configurar um EMR estúdio

1. Escolha um modo de autenticação para o Amazon EMR Studio
2. Crie os recursos do Studio a seguir.
   - Criar uma função de serviço do EMR Studio

- Configurar permissões de usuário do EMR Studio para Amazon EC2 ou Amazon EKS
- (Opcional) Defina grupos de segurança para controlar o tráfego de rede do EMR Studio.

3. Crie um EMR estúdio
4. Atribuir um usuário ou grupo a um EMR estúdio

Após concluir as etapas de configuração, você poderá Use um Amazon EMR Studio.

Escolha um modo de autenticação para o Amazon EMR Studio

EMRO Studio oferece suporte a dois modos de IAM autenticação: modo de autenticação e modo de autenticação do IAM Identity Center. IAMo modo usa AWS Identity and Access Management (IAM), enquanto o modo IAM Identity Center usa AWS IAM Identity Center. Ao criar um EMR Studio, você escolhe o modo de autenticação para todos os usuários desse Studio. Para obter mais informações sobre os diferentes modos de autenticação, consulte Autenticação e login do usuário.

Use a tabela a seguir para escolher um modo de autenticação para o EMR Studio.

Se você...	Recomendamos…
Já está familiarizado ou já configurou a IAM autenticação ou federação	O IAMmodo de autenticação, que oferece os seguintes benefícios: • Fornece configuração rápida para o EMR Studio se você já gerencia identidades, como usuários e grupos noIAM. • Funciona com provedores de identidade compatíveis com OpenID Connect (OIDC) ou Security Assertion Markup Language 2.0 (2.0). SAML • Oferecimento de suporte ao uso de diversos provedores de identidade com a mesma Conta da AWS. • Disponível em um grande número de Regiões da AWS. • Compatível com SOC 2.

Se você...	Recomendamos…
Novo na Amazon AWS EMR	O IAMModo de autenticação do Identity Center, que fornece os seguintes recursos: • Suporta fácil atribuição de AWS recursos por usuários e grupos. • Funciona com o Microsoft Active Directory e provedores de identidade SAML 2.0. • Facilita a configuração da federação de várias contas para que você não precise configurar separadamente a federação para cada uma Conta da AWS em sua organizaç ão.

Configurar o modo de IAM autenticação para o Amazon EMR Studio

Com o modo de IAM autenticação, você pode usar a IAM autenticação ou a IAM federação. IAMa autenticação permite gerenciar IAM identidades como usuários, grupos e funções emIAM. Você concede aos usuários acesso a um Studio com políticas de IAM permissões e controle de acesso baseado em atributos () ABAC. IAMa federação permite estabelecer confiança entre um provedor de identidade (IdP) terceirizado e gerenciar as identidades dos usuários AWS por meio do seu IdP.

> **Note**
>
> Se você já usa IAM para controlar o acesso aos AWS recursos ou se já configurou seu provedor de identidade (IdP) paraIAM, consulte Permissões de usuários para o modo de autenticação do IAM para definir permissões de usuário ao usar o modo de IAM autenticação do Studio. EMR

Use a IAM federação para o Amazon EMR Studio

Para usar a IAM federação para o EMR Studio, você cria uma relação de confiança entre você Conta da AWS e seu provedor de identidade (IdP) e permite que usuários federados acessem o. AWS Management Console As etapas executadas para criar essa relação de confiança variam com base no padrão de federação do seu IdP.

Em geral, você conclui as tarefas a seguir para configurar a federação com um IdP externo. Para obter instruções completas, consulte Habilitando o acesso de usuários federados SAML 2.0 AWS Management Console e Habilitando o acesso personalizado do agente de identidade ao AWS Management Console no Guia do AWS Identity and Access Management Usuário.

1. Reúna informações do seu IdP. Isso geralmente significa gerar um documento de metadados para validar as solicitações de SAML autenticação do seu IdP.
2. Crie uma IAM entidade provedora de identidade para armazenar informações sobre seu IdP. Para obter instruções, consulte Criação de provedores de IAM identidade.
3. Crie uma ou mais IAM funções para seu IdP. EMRO Studio atribui uma função a um usuário federado quando o usuário faz login. O perfil permite que seu IdP solicite credenciais de segurança temporárias para obter acesso à AWS. Para obter instruções, consulte Criar um perfil para um provedor de identidades de terceiros (federação). As políticas de permissões que você atribui à função determinam o que os usuários federados podem fazer em AWS e em um EMR Studio. Para obter mais informações, consulte Permissões de usuários para o modo de autenticação do IAM.
4. (Para SAML provedores) Complete a SAML confiança configurando seu IdP com AWS informações e as funções que você deseja que os usuários federados assumam. Esse processo de configuração cria confiança entre seu AWS IdP e. Para obter mais informações, consulte Configurando seu IdP SAML 2.0 com confiança de terceiros confiáveis e adicionando declarações.

Para configurar um EMR Studio como um SAML aplicativo em seu portal IdP

Você pode configurar um determinado EMR Studio como um SAML aplicativo usando um link direto para o Studio. Isso permite que os usuários façam login no seu portal do IdP e iniciem um Studio específico em vez de navegar pelo console da Amazon. EMR

- Use o formato a seguir para configurar um link direto para seu EMR Studio como destino URL após a verificação SAML da afirmação.

```
https://console.aws.amazon.com/emr/home?region=<aws-region>#studio/<your-studio-id>/start
```

Configurar o modo de autenticação do IAM Identity Center para o Amazon EMR Studio

Para se preparar AWS IAM Identity Center para o EMR Studio, você deve configurar sua fonte de identidade e provisionar usuários e grupos. O provisionamento é o processo de disponibilizar

informações de usuários e grupos para uso pelo IAM Identity Center e por aplicativos que usam o IAM Identity Center. Para obter mais informações, consulte User and group provisioning.

EMRO Studio oferece suporte ao uso dos seguintes provedores de identidade para o IAM Identity Center:

- AWS Managed Microsoft AD e Active Directory autogerenciado — Para obter mais informações, consulte Connect to your Microsoft AD directory.
- SAMLprovedores baseados — Para obter uma lista completa, consulte Provedores de identidade compatíveis.
- O diretório do IAM Identity Center — Para obter mais informações, consulte Gerenciar identidades no IAM Identity Center.

Para configurar o IAM Identity Center for EMR Studio

1. Para configurar o IAM Identity Center for EMR Studio, você precisa do seguinte:
   - Uma conta de gerenciamento em sua AWS organização se você usar várias contas em sua organização.

     > **Note**
     >
     > Você só deve usar sua conta de gerenciamento para habilitar o IAM Identity Center e provisionar usuários e grupos. Depois de configurar o IAM Identity Center, use uma conta de membro para criar um EMR Studio e atribuir usuários e grupos. Para saber mais sobre AWS terminologia, consulte AWS Organizations terminologia e conceitos.

   - Se você ativou o IAM Identity Center antes de 25 de novembro de 2019, talvez seja necessário habilitar aplicativos que usam o IAM Identity Center para as contas AWS da sua organização. Para obter mais informações, consulte Habilitar aplicativos integrados ao IAM Identity Center em AWS contas.
   - Verifique se você tem os pré-requisitos listados na página de pré-requisitos do IAMIdentity Center.
2. Siga as instruções em Ativar o IAM Identity Center para habilitar o IAM Identity Center no Região da AWS local em que você deseja criar o EMR Studio.
3. Conecte o IAM Identity Center ao seu provedor de identidade e provisione os usuários e grupos que você deseja atribuir ao Studio.

Se você usar...	Fazer isso...
Um diretório do Microsoft AD	1. Siga as instruções em Conectar-se ao seu diretório do Microsoft AD para conectar seu Active Directory ou AWS Managed Microsoft AD diretório autogerenciado usando AWS Directory Service. 2. Para provisionar usuários e grupos para o IAM Identity Center, você pode sincronizar dados de identidade do seu AD de origem com o IAM Identity Center. É possível sincronizar as identidades do seu AD de origem de várias maneiras. Uma das maneiras é atribuir usuários ou grupos do AD a uma conta da AWS na sua organizaç ão. Para obter instruções, consulte Single sign-on. A sincronização pode demorar até duas horas. Depois de concluir esta etapa, os usuários e os grupos sincronizados aparecerão no seu repositório de identidad es. **Note** Usuários e grupos não aparecem no Identity Store até que você sincronize as informações do usuário e do grupo ou use just-in-time (JIT) o provisionamento de usuários. Para obter mais informações, consulte Provision ing when users come from Active Directory.

Se você usar...	Fazer isso...
	3. (Opcional) Depois de sincronizar usuários e grupos do AD, você pode remover o acesso deles à sua AWS conta que você configurou na etapa anterior. Para obter instruções, consulte Remove user access.
Um provedor de identidades externo	Siga as instruções em Connect to your external identity provider.
O diretório do IAM Identity Center	Quando você cria usuários e grupos no IAM Identity Center, o provisionamento é automático. Para obter mais informaçõ es, consulte Gerenciar identidades no IAM Identity Center.

Agora você pode atribuir usuários e grupos da sua Identity Store a um EMR Studio. Para obter instruções, consulte Atribuir um usuário ou grupo a um EMR estúdio.

Criar uma função de serviço do EMR Studio

Sobre a função de serviço do EMR Studio

Cada EMR estúdio usa uma IAM função com permissões que permitem que o estúdio interaja com outros AWS serviços. Essa função de serviço deve incluir permissões que permitam ao EMR Studio estabelecer um canal de rede seguro entre espaços de trabalho e clusters, armazenar arquivos do notebook e acessá-los AWS Secrets Manager enquanto vincula um espaço de trabalho a um repositório Git. Amazon S3 Control

Use o perfil de serviço do Studio (em vez das políticas de sessão) para definir todas as permissões de acesso do Amazon S3 para armazenar arquivos de cadernos e para definir as permissões de acesso do AWS Secrets Manager .

Como criar uma função de serviço para o EMR Studio na Amazon EC2 ou na Amazon EKS

1. Siga as instruções em Criação de uma função para delegar permissões a um AWS serviço para criar a função de serviço com a seguinte política de confiança.

> **Important**
>
> A política de confiança a seguir inclui as chaves de condição `aws:SourceAccount`globais `aws:SourceArn`e as chaves de condição para limitar as permissões que você concede ao EMR Studio a recursos específicos em sua conta. Fazer isso pode proteger você contra o problema de "confused deputy".

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "elasticmapreduce.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "StringEquals": {
          "aws:SourceAccount": "<account-id>"
        },
        "ArnLike": {
          "aws:SourceArn": "arn:aws:elasticmapreduce:<region>:<account-id>:*"
        }
      }
    }
  ]
}
```

2. Remova as permissões de perfil padrão. Em seguida, inclua as permissões do exemplo de política de IAM permissões a seguir. Como alternativa, você pode criar uma política personalizada que use as EMRPermissões da função de serviço do Studio.

> **Important**
>
> - Para que o controle de acesso EC2 baseado em tags da Amazon funcione com o EMR Studio, você deve definir o acesso ao `ModifyNetworkInterfaceAttribute` API conforme mostrado na política a seguir.

- Para que o EMR Studio trabalhe com a função de serviço, você não deve alterar as seguintes declarações: `AllowAddingEMRTagsDuringDefaultSecurityGroupCreation` `AllowAddingTagsDuringEC2ENICreation` e.
- Para usar a política de exemplo, você deve etiquetar os recursos apresentados a seguir com a chave "**`for-use-with-amazon-emr-managed-policies`**" e o valor "**`true`**".
  - Sua Amazon Virtual Private Cloud (VPC) para EMR Studio.
  - Cada sub-rede que deseja usar com o Studio.
  - Qualquer grupo de segurança personalizado do EMR Studio. Você deve marcar todos os grupos de segurança criados durante o período de pré-visualização do EMR Studio se quiser continuar a usá-los.
  - Segredos mantidos nos AWS Secrets Manager quais os usuários do Studio usam para vincular repositórios Git a um espaço de trabalho.

  Você pode aplicar etiquetas aos recursos usando a guia Etiquetas na tela de recursos relevantes no AWS Management Console.

Quando aplicável, altere a * política a seguir para especificar o Amazon Resource Name (ARN) dos recursos que a declaração abrange para seu caso de uso. `"Resource":"*"`

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowEMRReadOnlyActions",
      "Effect": "Allow",
      "Action": [
        "elasticmapreduce:ListInstances",
        "elasticmapreduce:DescribeCluster",
        "elasticmapreduce:ListSteps"
      ],
      "Resource": "*"
    },
    {
      "Sid": "AllowEC2ENIActionsWithEMRTags",
      "Effect": "Allow",
```

```
      "Action": [
        "ec2:CreateNetworkInterfacePermission",
        "ec2:DeleteNetworkInterface"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:network-interface/*"
      ],
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowEC2ENIAttributeAction",
      "Effect": "Allow",
      "Action": [
        "ec2:ModifyNetworkInterfaceAttribute"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:instance/*",
        "arn:aws:ec2:*:*:network-interface/*",
        "arn:aws:ec2:*:*:security-group/*"
      ]
    },
    {
      "Sid": "AllowEC2SecurityGroupActionsWithEMRTags",
      "Effect": "Allow",
      "Action": [
        "ec2:AuthorizeSecurityGroupEgress",
        "ec2:AuthorizeSecurityGroupIngress",
        "ec2:RevokeSecurityGroupEgress",
        "ec2:RevokeSecurityGroupIngress",
        "ec2:DeleteNetworkInterfacePermission"
      ],
      "Resource": "*",
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowDefaultEC2SecurityGroupsCreationWithEMRTags",
```

```
      "Effect": "Allow",
      "Action": [
        "ec2:CreateSecurityGroup"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:security-group/*"
      ],
      "Condition": {
        "StringEquals": {
          "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowDefaultEC2SecurityGroupsCreationInVPCWithEMRTags",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateSecurityGroup"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:vpc/*"
      ],
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowAddingEMRTagsDuringDefaultSecurityGroupCreation",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateTags"
      ],
      "Resource": "arn:aws:ec2:*:*:security-group/*",
      "Condition": {
        "StringEquals": {
          "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true",
          "ec2:CreateAction": "CreateSecurityGroup"
        }
      }
    },
    {
      "Sid": "AllowEC2ENICreationWithEMRTags",
```

```
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterface"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:network-interface/*"
      ],
      "Condition": {
        "StringEquals": {
          "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowEC2ENICreationInSubnetAndSecurityGroupWithEMRTags",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterface"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:subnet/*",
        "arn:aws:ec2:*:*:security-group/*"
      ],
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowAddingTagsDuringEC2ENICreation",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateTags"
      ],
      "Resource": "arn:aws:ec2:*:*:network-interface/*",
      "Condition": {
        "StringEquals": {
          "ec2:CreateAction": "CreateNetworkInterface"
        }
      }
    },
    {
      "Sid": "AllowEC2ReadOnlyActions",
```

```
      "Effect": "Allow",
      "Action": [
        "ec2:DescribeSecurityGroups",
        "ec2:DescribeNetworkInterfaces",
        "ec2:DescribeTags",
        "ec2:DescribeInstances",
        "ec2:DescribeSubnets",
        "ec2:DescribeVpcs"
      ],
      "Resource": "*"
    },
    {
      "Sid": "AllowSecretsManagerReadOnlyActionsWithEMRTags",
      "Effect": "Allow",
      "Action": [
        "secretsmanager:GetSecretValue"
      ],
      "Resource": "arn:aws:secretsmanager:*:*:secret:*",
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowWorkspaceCollaboration",
      "Effect": "Allow",
      "Action": [
        "iam:GetUser",
        "iam:GetRole",
        "iam:ListUsers",
        "iam:ListRoles",
        "sso:GetManagedApplicationInstance",
        "sso-directory:SearchUsers"
      ],
      "Resource": "*"
    }
  ]
}
```

3. Dê à sua função de serviço acesso de leitura e gravação à sua localização do Amazon S3 para EMR o Studio. Use o conjunto mínimo de permissões apresentado a seguir. Para obter mais

informações, consulte o exemplo Amazon S3: permite acesso de leitura e gravação a objetos em um bucket do S3 de forma programática e no console.

```
"s3:PutObject",
"s3:GetObject",
"s3:GetEncryptionConfiguration",
"s3:ListBucket",
"s3:DeleteObject"
```

Se você criptografar seu bucket do Amazon S3, inclua as permissões a seguir para o AWS Key Management Service.

```
"kms:Decrypt",
"kms:GenerateDataKey",
"kms:ReEncryptFrom",
"kms:ReEncryptTo",
"kms:DescribeKey"
```

4. Se você quiser controlar o acesso aos segredos do Git no nível do usuário, adicione permissões baseadas em tags `secretsmanager:GetSecretValue` na política de função de usuário do EMR Studio e remova as permissões da `secretsmanager:GetSecretValue` política de função de serviço do EMR Studio. Para obter mais informações sobre como configurar as permissões refinadas de usuário, consulte Crie políticas de permissões para usuários do EMR Studio.

Função de serviço mínima para EMR Serverless

Se você quiser executar cargas de trabalho interativas com o EMR Serverless por meio de notebooks do EMR Studio, use a mesma política de confiança usada para configurar o EMR Studio na seção anterior,. Como criar uma função de serviço para o EMR Studio na Amazon EC2 ou na Amazon EKS

Para sua IAM política, a política mínima viável tem as seguintes permissões. Atualize *`bucket-name`* com o nome do bucket que você planeja usar ao configurar o EMR Studio e o Workspace. EMRO Studio usa o bucket para fazer backup dos espaços de trabalho e dos arquivos do notebook em seu Studio.

```
{
    "Version": "2012-10-17",
```

```
    "Statement": [
        {
            "Sid": "ObjectActions",
            "Effect": "Allow",
            "Action": [
               "s3:PutObject",
               "s3:GetObject",
               "s3:DeleteObject"
            ],
            "Resource": ["arn:aws:s3:::bucket-name/*"]
        },
        {
            "Sid": "BucketActions",
            "Effect": "Allow",
            "Action": [
               "s3:ListBucket",
               "s3:GetEncryptionConfiguration"
            ],
            "Resource": ["arn:aws:s3:::bucket-name"]
        }
    ]
}
```

Se usar um bucket criptografado do Amazon S3, inclua as seguintes permissões na sua política:

```
"kms:Decrypt",
"kms:GenerateDataKey",
"kms:ReEncryptFrom",
"kms:ReEncryptTo",
"kms:DescribeKey"
```

EMRPermissões da função de serviço do Studio

A tabela a seguir lista as operações que o EMR Studio executa usando a função de serviço, junto com as IAM ações necessárias para cada operação.

Operation	Ações
Estabeleça um canal de rede seguro entre um espaço de trabalho e um	`"ec2:CreateNetworkInterface", "ec2:CreateNetworkInterfacePermission", "ec2:DeleteNetworkInterface", "ec2:DeleteNetworkInterfacePermission",`

Operation	Ações

EMR cluster e execute as ações de limpeza necessárias.	`"ec2:DescribeNetworkInterfaces",`
	`"ec2:ModifyNetworkInterfaceAttribute",`
	`"ec2:AuthorizeSecurityGroupEgress",`
	`"ec2:AuthorizeSecurityGroupIngress",`
	`"ec2:CreateSecurityGroup",`
	`"ec2:DescribeSecurityGroups",`
	`"ec2:RevokeSecurityGroupEgress",`
	`"ec2:DescribeTags",`
	`"ec2:DescribeInstances",`
	`"ec2:DescribeSubnets",`
	`"ec2:DescribeVpcs",`
	`"elasticmapreduce:ListInstances",`
	`"elasticmapreduce:DescribeCluster",`
	`"elasticmapreduce:ListSteps"`

Use as credenciais do Git armazenadas para AWS Secrets Manager vincular os repositórios do Git a um espaço de trabalho.	`"secretsmanager:GetSecretValue"`
Aplique AWS tags à interface de rede e aos grupos de segurança padrão que o EMR Studio cria ao configurar o canal de rede seguro. Para obter mais informações, consulte Etiquetar recursos da AWS.	`"ec2:CreateTags"`

Operation	Ações

Acesso ou upload de arquivos e metadados de cadernos para o Amazon S3.	`"s3:PutObject",`
	`"s3:GetObject",`
	`"s3:GetEncryptionConfiguration",`
	`"s3:ListBucket",`
	`"s3:DeleteObject"`
	
	Se você usar um bucket criptografado do Amazon S3, inclua as permissões a seguir.
	
	`"kms:Decrypt",`
	`"kms:GenerateDataKey",`
	`"kms:ReEncryptFrom",`
	`"kms:ReEncryptTo",`
	`"kms:DescribeKey"`

Habilitação e configuração da colaboração no Workspace.	`"iam:GetUser",`
	`"iam:GetRole",`
	`"iam:ListUsers",`
	`"iam:ListRoles",`
	`"sso:GetManagedApplicationInstance",`
	`"sso-directory:SearchUsers"`

Criptografe cadernos e arquivos do espaço de trabalho do EMR Studio usando chaves gerenciadas pelo cliente () com CMK AWS Key Management Service	`"kms:Decrypt",`
	`"kms:GenerateDataKey",`
	`"kms:ReEncryptFrom",`
	`"kms:ReEncryptTo",`
	`"kms:DescribeKey"`

Configurar permissões de usuário do EMR Studio para Amazon EC2 ou Amazon EKS

Você deve configurar políticas de permissões de usuário para o Amazon EMR Studio para poder definir permissões refinadas de usuários e grupos. Para obter informações sobre como as permissões de usuário funcionam no EMR Studio, consulte Controle de acesso emComo o Amazon EMR Studio funciona.

> **Note**
>
> As permissões abordadas nesta seção não impõem controle de acesso a dados. Para gerenciar o acesso aos conjuntos de dados de entrada, você deve configurar permissões para os clusters que seu Studio usa. Para obter mais informações, consulte Segurança na Amazon EMR.

Crie uma função de usuário do EMR Studio para o modo de autenticação do IAM Identity Center

Você deve criar uma função de usuário do EMR Studio ao usar o modo de autenticação do IAM Identity Center.

Para criar uma função de usuário para o EMR Studio

1. Siga as instruções em Criação de uma função para delegar permissões a um AWS serviço no Guia do AWS Identity and Access Management usuário para criar uma função de usuário.

   Ao criar o perfil, use a política de relação de confiança apresentada a seguir.

   ```
   {
     "Version": "2008-10-17",
     "Statement": [
       {
         "Effect": "Allow",
         "Principal": {
           "Service": "elasticmapreduce.amazonaws.com"
         },
         "Action": [
                 "sts:AssumeRole",
                 "sts:SetContext"
                 ]
       }
     ]
   }
   ```

2. Remova as permissões e as políticas de perfil padrão.
3. Antes de atribuir usuários e grupos a um Studio, anexe suas políticas de sessão do EMR Studio à função de usuário. Para obter instruções sobre como criar políticas de sessão, consulte Crie políticas de permissões para usuários do EMR Studio.

Crie políticas de permissões para usuários do EMR Studio

Consulte as seções a seguir para criar políticas de permissões para o EMR Studio.

Tópicos

> **Note**
>
> Para definir permissões de acesso ao Amazon S3 para armazenar arquivos de notebook e para definir permissões de AWS Secrets Manager acesso para ler segredos ao vincular espaços de trabalho a repositórios Git, use a função de serviço do Studio. EMR

Criação das políticas de permissões

Crie uma ou mais políticas de IAM permissões que especifiquem quais ações um usuário pode realizar no seu Studio. Por exemplo, é possível criar três políticas separadas para tipos de usuários básicos, intermediários e avançados do Studio com os exemplos de políticas nesta página.

Para obter um detalhamento de cada operação do Studio que um usuário pode realizar e as IAM ações mínimas necessárias para realizar cada operação, consulteAWS Identity and Access Management permissões para usuários do EMR Studio. Para ver as etapas para criar as políticas, consulte Criação de IAM políticas no Guia IAM do usuário.

Sua política de permissões deve incluir as instruções apresentadas a seguir.

```
{
            "Sid": "AllowAddingTagsOnSecretsWithEMRStudioPrefix",
            "Effect": "Allow",
            "Action": "secretsmanager:TagResource",
            "Resource": "arn:aws:secretsmanager:*:*:secret:emr-studio-*"
},
{
            "Sid": "AllowPassingServiceRoleForWorkspaceCreation",
            "Action": "iam:PassRole",
```

```
            "Resource": [
                "arn:aws:iam::*:role/your-emr-studio-service-role"
            ],
            "Effect": "Allow"
}
```

Definição de propriedade para colaboração no Workspace

A colaboração no Workspace permite que vários usuários trabalhem simultaneamente no mesmo Workspace e pode ser configurada com o painel Colaboração na interface do usuário do Workspace. Para visualizar e usar o painel Colaboração, o usuário deve ter as permissões apresentadas a seguir. Qualquer usuário com essas permissões poderá visualizar e usar o painel Colaboração.

```
"elasticmapreduce:UpdateEditor",
"elasticmapreduce:PutWorkspaceAccess",
"elasticmapreduce:DeleteWorkspaceAccess",
"elasticmapreduce:ListWorkspaceAccessIdentities"
```

Para restringir o acesso ao painel Colaboração, é possível usar o controle de acesso por etiquetas. Quando um usuário cria um espaço de trabalho, o EMR Studio aplica uma tag padrão com uma chave `creatorUserId` cujo valor é o ID do usuário que está criando o espaço de trabalho.

> **Note**
>
> EMRO Studio adiciona a `creatorUserId` tag aos espaços de trabalho criados após 16 de novembro de 2021. Para restringir quem pode configurar a colaboração dos espaços de trabalhos criados antes dessa data, recomendamos adicionar manualmente a tag `creatorUserId` ao seu Workspace e, em seguida, usar o controle de acesso por tags nas suas políticas de permissões de usuários.

A instrução de exemplo a seguir permite que um usuário configure a colaboração para qualquer Workspace com a chave de etiqueta `creatorUserId` cujo valor corresponde ao ID do usuário (indicado pela variável de política `aws:userId`). Em outras palavras, a instrução permite que um usuário configure a colaboração para os Workspaces criados por ele. Para saber mais sobre variáveis de política, consulte elementos de IAM política: variáveis e tags no Guia IAM do usuário.

```
    {
        "Sid": "UserRolePermissionsForCollaboration",
        "Action": [
```

```
            "elasticmapreduce:UpdateEditor",
            "elasticmapreduce:PutWorkspaceAccess",
            "elasticmapreduce:DeleteWorkspaceAccess",
            "elasticmapreduce:ListWorkspaceAccessIdentities"
        ],
        "Resource": "*",
        "Effect": "Allow",
        "Condition": {
            "StringEquals": {
                "elasticmapreduce:ResourceTag/creatorUserId": "${aws:userid}"
            }
        }
    }
```

Criação de uma política de segredos do Git no nível de usuário

Tópicos

- Para usar permissões no nível de usuário
- Para fazer a transição de permissões do nível de serviço para permissões do nível de usuário
- Para usar permissões no nível de serviço

Para usar permissões no nível de usuário

EMRO Studio adiciona automaticamente a `for-use-with-amazon-emr-managed-user-policies` tag ao criar segredos do Git. Se você quiser controlar o acesso aos segredos do Git no nível do usuário, adicione permissões baseadas em tags à política de função de usuário do EMR Studio `secretsmanager:GetSecretValue` conforme mostrado na Para fazer a transição de permissões do nível de serviço para permissões do nível de usuário seção abaixo.

Se você tiver permissões existentes `secretsmanager:GetSecretValue` na política de função de serviço do EMR Studio, remova essas permissões.

Para fazer a transição de permissões do nível de serviço para permissões do nível de usuário

> **Note**
>
> A tag `for-use-with-amazon-emr-managed-user-policies` garante que as permissões da Etapa 1 abaixo concedam ao criador do espaço de trabalho acesso ao segredo do Git. No entanto, se você vinculou repositórios Git antes de 1.º de setembro de 2023, os segredos do Git correspondentes terão o acesso negado por não terem a

tag `for-use-with-amazon-emr-managed-user-policies` aplicada. Para aplicar permissões em nível de usuário, você deve recriar os segredos antigos JupyterLab e vincular os repositórios Git apropriados novamente.
Para obter mais informações sobre variáveis de política, consulte elementos de IAM política: variáveis e tags no Guia IAM do usuário.

1. Adicione as seguintes permissões à política de função de usuário do EMR Studio. A política usa a chave `for-use-with-amazon-emr-managed-user-policies` com valor `"${aws:userid}"`.

```
{
   "Sid": "AllowSecretsManagerReadOnlyActionsWithEMRTags",
    "Effect": "Allow",
    "Action": "secretsmanager:GetSecretValue",
    "Resource": "arn:aws:secretsmanager:*:*:secret:*",
    "Condition": {
        "StringEquals": {
            "secretsmanager:ResourceTag/for-use-with-amazon-emr-managed-user-
policies": "${aws:userid}"
        }
    }
}
```

2. Se presente, remova a seguinte permissão da política de função de serviço do EMR Studio. Como a política de perfil de serviço se aplica a todos os segredos definidos por cada usuário, você só precisa fazer isso uma vez.

```
{
    "Sid": "AllowSecretsManagerReadOnlyActionsWithEMRTags",
    "Effect": "Allow",
    "Action": [
        "secretsmanager:GetSecretValue"
     ],
    "Resource": "arn:aws:secretsmanager:*:*:secret:*",
    "Condition": {
        "StringEquals": {
            "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
        }
    }
}
```

Para usar permissões no nível de serviço

A partir de 1º de setembro de 2023, o EMR Studio adiciona automaticamente a `for-use-with-amazon-emr-managed-user-policies` tag para controle de acesso em nível de usuário. Como esse é um recurso adicional, você pode continuar usando o acesso em nível de serviço que está disponível por meio da `GetSecretValue` permissão na função de serviço do EMR Studio.

Para segredos criados antes de 1º de setembro de 2023, o EMR Studio não adicionou a `for-use-with-amazon-emr-managed-user-policies` tag. Para continuar usando as permissões de nível de serviço, basta manter sua função de serviço do EMR Studio e suas permissões de função de usuário existentes. No entanto, para restringir quem pode acessar um segredo individual, recomendamos seguir as etapas em Para usar permissões no nível de usuário para adicionar manualmente a tag `for-use-with-amazon-emr-managed-user-policies` aos seus segredos e, em seguida, usar o controle de acesso por tags nas suas políticas de permissões de usuários.

Para obter mais informações sobre variáveis de política, consulte elementos de IAM política: variáveis e tags no Guia IAM do usuário.

Anexe a política de permissões à sua IAM identidade

A tabela a seguir resume a qual IAM identidade você anexa uma política de permissões, dependendo do modo de autenticação do EMR Studio. Para obter instruções sobre como anexar uma política, consulte Adicionar e remover permissões de IAM identidade.

Se você usar...	Anexe a política para...
IAMautenticação	Suas IAM identidades (usuários, grupos de usuários ou funções). Por exemplo, é possível anexar uma política de permissões a um usuário em sua Conta da AWS.
IAMfederação com um provedor de identidade externo (IdP)	A IAM função ou funções que você cria para seu IdP externo. Por exemplo, uma federação IAM for SAML 2.0. EMRO Studio usa as permissões que você atribui às suas IAM funções para usuários com acesso federado a um Studio.

Se você usar...	Anexe a política para...
IAMCentro de identidade	Sua função de usuário do Amazon EMR Studio.

Exemplo de políticas de usuário

A política básica de usuário a seguir permite a maioria das ações do EMR Studio, mas não permite que um usuário crie novos EMR clusters da Amazon.

Política básica

> **Important**
>
> O exemplo de política não inclui a `CreateStudioPresignedUrl` permissão, que você deve conceder a um usuário ao usar o modo de IAM autenticação. Para obter mais informações, consulte Atribuir um usuário ou grupo a um EMR estúdio.

O exemplo de política inclui `Condition` elementos para impor o controle de acesso baseado em tags (TBAC) para que você possa usar a política com o exemplo de função de serviço do Studio. EMR Para obter mais informações, consulte Criar uma função de serviço do EMR Studio.

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Sid":"AllowDefaultEC2SecurityGroupsCreationInVPCWithEMRTags",
         "Effect":"Allow",
         "Action":[
            "ec2:CreateSecurityGroup"
         ],
         "Resource":[
            "arn:aws:ec2:*:*:vpc/*"
         ],
         "Condition":{
            "StringEquals":{
               "aws:ResourceTag/for-use-with-amazon-emr-managed-policies":"true"
            }
         }
```

```
    },
    {
       "Sid":"AllowAddingEMRTagsDuringDefaultSecurityGroupCreation",
       "Effect":"Allow",
       "Action":[
          "ec2:CreateTags"
       ],
       "Resource":"arn:aws:ec2:*:*:security-group/*",
       "Condition":{
          "StringEquals":{
             "aws:RequestTag/for-use-with-amazon-emr-managed-policies":"true",
             "ec2:CreateAction":"CreateSecurityGroup"
          }
       }
    },
    {
       "Sid":"AllowSecretManagerListSecrets",
       "Action":[
          "secretsmanager:ListSecrets"
       ],
       "Resource":"*",
       "Effect":"Allow"
    },
    {
       "Sid":"AllowSecretCreationWithEMRTagsAndEMRStudioPrefix",
       "Effect":"Allow",
       "Action":"secretsmanager:CreateSecret",
       "Resource":"arn:aws:secretsmanager:*:*:secret:emr-studio-*",
       "Condition":{
          "StringEquals":{
             "aws:RequestTag/for-use-with-amazon-emr-managed-policies":"true"
          }
       }
    },
    {
       "Sid":"AllowAddingTagsOnSecretsWithEMRStudioPrefix",
       "Effect":"Allow",
       "Action":"secretsmanager:TagResource",
       "Resource":"arn:aws:secretsmanager:*:*:secret:emr-studio-*"
    },
    {
       "Sid":"AllowPassingServiceRoleForWorkspaceCreation",
       "Action":"iam:PassRole",
       "Resource":[
```

```
            "arn:aws:iam::*:role/<your-emr-studio-service-role>"
         ],
         "Effect":"Allow"
      },
      {
         "Sid":"AllowS3ListAndLocationPermissions",
         "Action":[
            "s3:ListAllMyBuckets",
            "s3:ListBucket",
            "s3:GetBucketLocation"
         ],
         "Resource":"arn:aws:s3:::*",
         "Effect":"Allow"
      },
      {
         "Sid":"AllowS3ReadOnlyAccessToLogs",
         "Action":[
            "s3:GetObject"
         ],
         "Resource":[
            "arn:aws:s3:::aws-logs-<aws-account-id>-<region>/elasticmapreduce/*"
         ],
         "Effect":"Allow"
      },
      {
         "Sid":"AllowConfigurationForWorkspaceCollaboration",
         "Action":[
            "elasticmapreduce:UpdateEditor",
            "elasticmapreduce:PutWorkspaceAccess",
            "elasticmapreduce:DeleteWorkspaceAccess",
            "elasticmapreduce:ListWorkspaceAccessIdentities"
         ],
         "Resource":"*",
         "Effect":"Allow",
         "Condition":{
            "StringEquals":{
               "elasticmapreduce:ResourceTag/creatorUserId":"${aws:userId}"
            }
         }
      },
      {
         "Sid":"DescribeNetwork",
         "Effect":"Allow",
         "Action":[
```

```
            "ec2:DescribeVpcs",
            "ec2:DescribeSubnets",
            "ec2:DescribeSecurityGroups"
         ],
         "Resource":"*"
      },
      {
         "Sid":"ListIAMRoles",
         "Effect":"Allow",
         "Action":[
            "iam:ListRoles"
         ],
         "Resource":"*"
      }
   ]
}
```

A política de usuário intermediário a seguir permite a maioria das ações do EMR Studio e permite que um usuário crie novos EMR clusters da Amazon usando um modelo de cluster.

Política intermediária

> **Important**
>
> O exemplo de política não inclui a `CreateStudioPresignedUrl` permissão, que você deve conceder a um usuário ao usar o modo de IAM autenticação. Para obter mais informações, consulte Atribuir um usuário ou grupo a um EMR estúdio.

O exemplo de política inclui `Condition` elementos para impor o controle de acesso baseado em tags (TBAC) para que você possa usar a política com o exemplo de função de serviço do Studio. EMR Para obter mais informações, consulte Criar uma função de serviço do EMR Studio.

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Sid":"AllowEMRBasicActions",
         "Action":[
            "elasticmapreduce:CreateEditor",
            "elasticmapreduce:DescribeEditor",
            "elasticmapreduce:ListEditors",
```

```
            "elasticmapreduce:StartEditor",
            "elasticmapreduce:StopEditor",
            "elasticmapreduce:DeleteEditor",
            "elasticmapreduce:OpenEditorInConsole",
            "elasticmapreduce:AttachEditor",
            "elasticmapreduce:DetachEditor",
            "elasticmapreduce:CreateRepository",
            "elasticmapreduce:DescribeRepository",
            "elasticmapreduce:DeleteRepository",
            "elasticmapreduce:ListRepositories",
            "elasticmapreduce:LinkRepository",
            "elasticmapreduce:UnlinkRepository",
            "elasticmapreduce:DescribeCluster",
            "elasticmapreduce:ListInstanceGroups",
            "elasticmapreduce:ListBootstrapActions",
            "elasticmapreduce:ListClusters",
            "elasticmapreduce:ListSteps",
            "elasticmapreduce:CreatePersistentAppUI",
            "elasticmapreduce:DescribePersistentAppUI",
            "elasticmapreduce:GetPersistentAppUIPresignedURL",
            "elasticmapreduce:GetOnClusterAppUIPresignedURL"
        ],
        "Resource":"*",
        "Effect":"Allow"
    },
    {
        "Sid":"AllowEMRContainersBasicActions",
        "Action":[
            "emr-containers:DescribeVirtualCluster",
            "emr-containers:ListVirtualClusters",
            "emr-containers:DescribeManagedEndpoint",
            "emr-containers:ListManagedEndpoints",
            "emr-containers:DescribeJobRun",
            "emr-containers:ListJobRuns"
        ],
        "Resource":"*",
        "Effect":"Allow"
    },
    {
        "Sid": "AllowRetrievingManagedEndpointCredentials",
        "Effect": "Allow",
        "Action": [
            "emr-containers:GetManagedEndpointSessionCredentials"
        ],
```

```
            "Resource": [
                "arn:aws:emr-containers:<region>:<account-id>:/virtualclusters/<virtual-
cluster-id>/endpoints/<managed-endpoint-id>"
            ],
            "Condition": {
                "StringEquals": {
                    "emr-containers:ExecutionRoleArn": [
                        "arn:aws:iam::<account-id>:role/<emr-on-eks-execution-role>"
                    ]
                }
            }
        },
        {
            "Sid":"AllowSecretManagerListSecrets",
            "Action":[
               "secretsmanager:ListSecrets"
            ],
            "Resource":"*",
            "Effect":"Allow"
        },
        {
            "Sid":"AllowSecretCreationWithEMRTagsAndEMRStudioPrefix",
            "Effect":"Allow",
            "Action":"secretsmanager:CreateSecret",
            "Resource":"arn:aws:secretsmanager:*:*:secret:emr-studio-*",
            "Condition":{
               "StringEquals":{
                  "aws:RequestTag/for-use-with-amazon-emr-managed-policies":"true"
               }
            }
        },
        {
            "Sid":"AllowAddingTagsOnSecretsWithEMRStudioPrefix",
            "Effect":"Allow",
            "Action":"secretsmanager:TagResource",
            "Resource":"arn:aws:secretsmanager:*:*:secret:emr-studio-*"
        },
        {
            "Sid":"AllowClusterTemplateRelatedIntermediateActions",
            "Action":[
               "servicecatalog:DescribeProduct",
               "servicecatalog:DescribeProductView",
               "servicecatalog:DescribeProvisioningParameters",
               "servicecatalog:ProvisionProduct",
```

```
            "servicecatalog:SearchProducts",
            "servicecatalog:UpdateProvisionedProduct",
            "servicecatalog:ListProvisioningArtifacts",
            "servicecatalog:ListLaunchPaths",
            "servicecatalog:DescribeRecord",
            "cloudformation:DescribeStackResources"
        ],
        "Resource":"*",
        "Effect":"Allow"
    },
    {
        "Sid":"AllowPassingServiceRoleForWorkspaceCreation",
        "Action":"iam:PassRole",
        "Resource":[
            "arn:aws:iam::*:role/<your-emr-studio-service-role>"
        ],
        "Effect":"Allow"
    },
    {
        "Sid":"AllowS3ListAndLocationPermissions",
        "Action":[
            "s3:ListAllMyBuckets",
            "s3:ListBucket",
            "s3:GetBucketLocation"
        ],
        "Resource":"arn:aws:s3:::*",
        "Effect":"Allow"
    },
    {
        "Sid":"AllowS3ReadOnlyAccessToLogs",
        "Action":[
            "s3:GetObject"
        ],
        "Resource":[
            "arn:aws:s3:::aws-logs-<aws-account-id>-<region>/elasticmapreduce/*"
        ],
        "Effect":"Allow"
    },
    {
        "Sid":"AllowConfigurationForWorkspaceCollaboration",
        "Action":[
            "elasticmapreduce:UpdateEditor",
            "elasticmapreduce:PutWorkspaceAccess",
            "elasticmapreduce:DeleteWorkspaceAccess",
```

```
            "elasticmapreduce:ListWorkspaceAccessIdentities"
         ],
         "Resource":"*",
         "Effect":"Allow",
         "Condition":{
            "StringEquals":{
               "elasticmapreduce:ResourceTag/creatorUserId":"${aws:userId}"
            }
         }
      },
      {
         "Sid":"DescribeNetwork",
         "Effect":"Allow",
         "Action":[
            "ec2:DescribeVpcs",
            "ec2:DescribeSubnets",
            "ec2:DescribeSecurityGroups"
         ],
         "Resource":"*"
      },
      {
         "Sid":"ListIAMRoles",
         "Effect":"Allow",
         "Action":[
            "iam:ListRoles"
         ],
         "Resource":"*"
      },
      {
         "Sid": "AllowServerlessActions",
         "Action": [
            "emr-serverless:CreateApplication",
            "emr-serverless:UpdateApplication",
            "emr-serverless:DeleteApplication",
            "emr-serverless:ListApplications",
            "emr-serverless:GetApplication",
            "emr-serverless:StartApplication",
            "emr-serverless:StopApplication",
            "emr-serverless:StartJobRun",
            "emr-serverless:CancelJobRun",
            "emr-serverless:ListJobRuns",
            "emr-serverless:GetJobRun",
            "emr-serverless:GetDashboardForJobRun",
            "emr-serverless:AccessInteractiveEndpoints"
```

```
         ],
         "Resource": "*",
         "Effect": "Allow"
        },
        {
           "Sid": "AllowPassingRuntimeRoleForRunningServerlessJob",
           "Action": "iam:PassRole",
           "Resource": "arn:aws:iam::*:role/serverless-runtime-role",
           "Effect": "Allow"
        }
   ]
}
```

A política de usuário avançada a seguir permite todas as ações do EMR Studio e permite que um usuário crie novos EMR clusters da Amazon usando um modelo de cluster ou fornecendo uma configuração de cluster.

Política avançada

> **Important**
>
> O exemplo de política não inclui a `CreateStudioPresignedUrl` permissão, que você deve conceder a um usuário ao usar o modo de IAM autenticação. Para obter mais informações, consulte Atribuir um usuário ou grupo a um EMR estúdio.

O exemplo de política inclui `Condition` elementos para impor o controle de acesso baseado em tags (TBAC) para que você possa usar a política com o exemplo de função de serviço do Studio. EMR Para obter mais informações, consulte Criar uma função de serviço do EMR Studio.

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Sid":"AllowEMRBasicActions",
         "Action":[
            "elasticmapreduce:CreateEditor",
            "elasticmapreduce:DescribeEditor",
            "elasticmapreduce:ListEditors",
            "elasticmapreduce:StartEditor",
            "elasticmapreduce:StopEditor",
            "elasticmapreduce:DeleteEditor",
```

```
            "elasticmapreduce:OpenEditorInConsole",
            "elasticmapreduce:AttachEditor",
            "elasticmapreduce:DetachEditor",
            "elasticmapreduce:CreateRepository",
            "elasticmapreduce:DescribeRepository",
            "elasticmapreduce:DeleteRepository",
            "elasticmapreduce:ListRepositories",
            "elasticmapreduce:LinkRepository",
            "elasticmapreduce:UnlinkRepository",
            "elasticmapreduce:DescribeCluster",
            "elasticmapreduce:ListInstanceGroups",
            "elasticmapreduce:ListBootstrapActions",
            "elasticmapreduce:ListClusters",
            "elasticmapreduce:ListSteps",
            "elasticmapreduce:CreatePersistentAppUI",
            "elasticmapreduce:DescribePersistentAppUI",
            "elasticmapreduce:GetPersistentAppUIPresignedURL",
            "elasticmapreduce:GetOnClusterAppUIPresignedURL"
         ],
         "Resource":"*",
         "Effect":"Allow"
      },
      {
         "Sid":"AllowEMRContainersBasicActions",
         "Action":[
            "emr-containers:DescribeVirtualCluster",
            "emr-containers:ListVirtualClusters",
            "emr-containers:DescribeManagedEndpoint",
            "emr-containers:ListManagedEndpoints",
            "emr-containers:DescribeJobRun",
            "emr-containers:ListJobRuns"
         ],
         "Resource":"*",
         "Effect":"Allow"
      },
      {
          "Sid": "AllowRetrievingManagedEndpointCredentials",
          "Effect": "Allow",
          "Action": [
              "emr-containers:GetManagedEndpointSessionCredentials"
          ],
          "Resource": [
              "arn:aws:emr-containers:<region>:<account-id>:/virtualclusters/<virtual-
cluster-id>/endpoints/<managed-endpoint-id>"
```

```
            ],
            "Condition": {
                "StringEquals": {
                    "emr-containers:ExecutionRoleArn": [
                        "arn:aws:iam::<account-id>:role/<emr-on-eks-execution-role>"
                    ]
                }
            }
        },
        {
            "Sid":"AllowSecretManagerListSecrets",
            "Action":[
                "secretsmanager:ListSecrets"
            ],
            "Resource":"*",
            "Effect":"Allow"
        },
        {
            "Sid":"AllowSecretCreationWithEMRTagsAndEMRStudioPrefix",
            "Effect":"Allow",
            "Action":"secretsmanager:CreateSecret",
            "Resource":"arn:aws:secretsmanager:*:*:secret:emr-studio-*",
            "Condition":{
                "StringEquals":{
                    "aws:RequestTag/for-use-with-amazon-emr-managed-policies":"true"
                }
            }
        },
        {
            "Sid":"AllowAddingTagsOnSecretsWithEMRStudioPrefix",
            "Effect":"Allow",
            "Action":"secretsmanager:TagResource",
            "Resource":"arn:aws:secretsmanager:*:*:secret:emr-studio-*"
        },
        {
            "Sid":"AllowClusterTemplateRelatedIntermediateActions",
            "Action":[
                "servicecatalog:DescribeProduct",
                "servicecatalog:DescribeProductView",
                "servicecatalog:DescribeProvisioningParameters",
                "servicecatalog:ProvisionProduct",
                "servicecatalog:SearchProducts",
                "servicecatalog:UpdateProvisionedProduct",
                "servicecatalog:ListProvisioningArtifacts",
```

```
            "servicecatalog:ListLaunchPaths",
            "servicecatalog:DescribeRecord",
            "cloudformation:DescribeStackResources"
        ],
        "Resource":"*",
        "Effect":"Allow"
    },
    {
        "Sid":"AllowEMRCreateClusterAdvancedActions",
        "Action":[
            "elasticmapreduce:RunJobFlow"
        ],
        "Resource":"*",
        "Effect":"Allow"
    },
    {
        "Sid":"AllowPassingServiceRoleForWorkspaceCreation",
        "Action":"iam:PassRole",
        "Resource":[
            "arn:aws:iam::*:role/<your-emr-studio-service-role>",
            "arn:aws:iam::*:role/EMR_DefaultRole_V2",
            "arn:aws:iam::*:role/EMR_EC2_DefaultRole"
        ],
        "Effect":"Allow"
    },
    {
        "Sid":"AllowS3ListAndLocationPermissions",
        "Action":[
            "s3:ListAllMyBuckets",
            "s3:ListBucket",
            "s3:GetBucketLocation"
        ],
        "Resource":"arn:aws:s3:::*",
        "Effect":"Allow"
    },
    {
        "Sid":"AllowS3ReadOnlyAccessToLogs",
        "Action":[
            "s3:GetObject"
        ],
        "Resource":[
            "arn:aws:s3:::aws-logs-<aws-account-id>-<region>/elasticmapreduce/*"
        ],
        "Effect":"Allow"
```

```
        },
        {
            "Sid":"AllowConfigurationForWorkspaceCollaboration",
            "Action":[
                "elasticmapreduce:UpdateEditor",
                "elasticmapreduce:PutWorkspaceAccess",
                "elasticmapreduce:DeleteWorkspaceAccess",
                "elasticmapreduce:ListWorkspaceAccessIdentities"
            ],
            "Resource":"*",
            "Effect":"Allow",
            "Condition":{
                "StringEquals":{
                    "elasticmapreduce:ResourceTag/creatorUserId":"${aws:userId}"
                }
            }
        },
        {
            "Sid" : "SageMakerDataWranglerForEMRStudio",
            "Effect" : "Allow",
            "Action" : [
                "sagemaker:CreatePresignedDomainUrl",
                "sagemaker:DescribeDomain",
                "sagemaker:ListDomains",
                "sagemaker:ListUserProfiles"
            ],
            "Resource":"*"
        },
        {
            "Sid":"DescribeNetwork",
            "Effect":"Allow",
            "Action":[
                "ec2:DescribeVpcs",
                "ec2:DescribeSubnets",
                "ec2:DescribeSecurityGroups"
            ],
            "Resource":"*"
        },
        {
            "Sid":"ListIAMRoles",
            "Effect":"Allow",
            "Action":[
                "iam:ListRoles"
            ],
```

```
            "Resource":"*"
        },
        {
            "Sid": "AllowServerlessActions",
            "Action": [
                "emr-serverless:CreateApplication",
                "emr-serverless:UpdateApplication",
                "emr-serverless:DeleteApplication",
                "emr-serverless:ListApplications",
                "emr-serverless:GetApplication",
                "emr-serverless:StartApplication",
                "emr-serverless:StopApplication",
                "emr-serverless:StartJobRun",
                "emr-serverless:CancelJobRun",
                "emr-serverless:ListJobRuns",
                "emr-serverless:GetJobRun",
                "emr-serverless:GetDashboardForJobRun",
                "emr-serverless:AccessInteractiveEndpoints"
            ],
            "Resource": "*",
            "Effect": "Allow"
        },
        {
            "Sid": "AllowPassingRuntimeRoleForRunningServerlessJob",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::*:role/serverless-runtime-role",
            "Effect": "Allow"
        },
        {
            "Sid": "AllowCodeWhisperer",
            "Effect": "Allow",
            "Action": [ "codewhisperer:GenerateRecommendations" ],
            "Resource": "*"
        },
        {
            "Sid": "AllowAthenaSQL",
            "Action": [
                "athena:StartQueryExecution",
                "athena:StopQueryExecution",
                "athena:GetQueryExecution",
                "athena:GetQueryRuntimeStatistics",
                "athena:GetQueryResults",
                "athena:ListQueryExecutions",
                "athena:BatchGetQueryExecution",
```

```
            "athena:GetNamedQuery",
            "athena:ListNamedQueries",
            "athena:BatchGetNamedQuery",
            "athena:UpdateNamedQuery",
            "athena:DeleteNamedQuery",
            "athena:ListDataCatalogs",
            "athena:GetDataCatalog",
            "athena:ListDatabases",
            "athena:GetDatabase",
            "athena:ListTableMetadata",
            "athena:GetTableMetadata",
            "athena:ListWorkGroups",
            "athena:GetWorkGroup",
            "athena:CreateNamedQuery",
            "athena:GetPreparedStatement",
            "glue:CreateDatabase",
            "glue:DeleteDatabase",
            "glue:GetDatabase",
            "glue:GetDatabases",
            "glue:UpdateDatabase",
            "glue:CreateTable",
            "glue:DeleteTable",
            "glue:BatchDeleteTable",
            "glue:UpdateTable",
            "glue:GetTable",
            "glue:GetTables",
            "glue:BatchCreatePartition",
            "glue:CreatePartition",
            "glue:DeletePartition",
            "glue:BatchDeletePartition",
            "glue:UpdatePartition",
            "glue:GetPartition",
            "glue:GetPartitions",
            "glue:BatchGetPartition",
            "kms:ListAliases",
            "kms:ListKeys",
            "kms:DescribeKey",
            "lakeformation:GetDataAccess",
            "s3:GetBucketLocation",
            "s3:GetBucketLocation",
            "s3:GetObject",
            "s3:ListBucket",
            "s3:ListBucketMultipartUploads",
            "s3:ListMultipartUploadParts",
```

```
            "s3:AbortMultipartUpload",
            "s3:PutObject",
            "s3:PutBucketPublicAccessBlock",
            "s3:ListAllMyBuckets"
        ],
        "Resource": "*",
        "Effect": "Allow"
    }
  ]
}
```

A política de usuário a seguir contém as permissões mínimas de usuário necessárias para usar um aplicativo interativo EMR sem servidor com o EMR Studio Workspaces.

EMRPolítica interativa sem servidor

Neste exemplo de política que tem permissões de usuário para aplicativos interativos EMR sem servidor com o EMR Studio, substitua os espaços reservados para *`serverless-runtime-role`* e *`emr-studio-service-role`* com sua função de serviço correta do EMR Studio e função de tempo de execução EMR sem servidor.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowServerlessActions",
            "Action": [
                "emr-serverless:CreateApplication",
                "emr-serverless:UpdateApplication",
                "emr-serverless:DeleteApplication",
                "emr-serverless:ListApplications",
                "emr-serverless:GetApplication",
                "emr-serverless:StartApplication",
                "emr-serverless:StopApplication",
                "emr-serverless:StartJobRun",
                "emr-serverless:CancelJobRun",
                "emr-serverless:ListJobRuns",
                "emr-serverless:GetJobRun",
                "emr-serverless:GetDashboardForJobRun",
                "emr-serverless:AccessInteractiveEndpoints"
            ],
            "Resource": "*",
            "Effect": "Allow"
```

```
        },
        {
            "Sid": "AllowEMRBasicActions",
            "Action": [
                "elasticmapreduce:CreateEditor",
                "elasticmapreduce:DescribeEditor",
                "elasticmapreduce:ListEditors",
                "elasticmapreduce:UpdateStudio",
                "elasticmapreduce:StartEditor",
                "elasticmapreduce:StopEditor",
                "elasticmapreduce:DeleteEditor",
                "elasticmapreduce:OpenEditorInConsole",
                "elasticmapreduce:AttachEditor",
                "elasticmapreduce:DetachEditor",
                "elasticmapreduce:CreateStudio",
                "elasticmapreduce:DescribeStudio",
                "elasticmapreduce:DeleteStudio",
                "elasticmapreduce:ListStudios",
                "elasticmapreduce:CreateStudioPresignedUrl"
            ],
            "Resource": "*",
            "Effect": "Allow"
        },
        {
            "Sid": "AllowPassingRuntimeRoleForRunningEMRServerlessJob",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::*:role/serverless-runtime-role",
            "Effect": "Allow"
        },
        {
            "Sid": "AllowPassingServiceRoleForWorkspaceCreation",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::*:role/emr-studio-service-role",
            "Effect": "Allow"
        },
        {
            "Sid": "AllowS3ListAndGetPermissions",
            "Action": [
                "s3:ListAllMyBuckets",
                "s3:ListBucket",
                "s3:GetBucketLocation",
                "s3:GetObject"
            ],
            "Resource": "arn:aws:s3:::*",
```

```
            "Effect": "Allow"
        },
        {
            "Sid":"DescribeNetwork",
            "Effect":"Allow",
            "Action":[
                "ec2:DescribeVpcs",
                "ec2:DescribeSubnets",
                "ec2:DescribeSecurityGroups"
            ],
            "Resource":"*"
        },
        {
            "Sid": "ListIAMRoles",
            "Effect": "Allow",
            "Action": [
                "iam:ListRoles"
            ],
            "Resource": "*"
        }
    ]
}
```

AWS Identity and Access Management permissões para usuários do EMR Studio

A tabela a seguir inclui cada operação do Amazon EMR Studio que um usuário pode realizar e lista as IAM ações mínimas necessárias para realizar essa operação. Você permite essas ações em suas políticas de IAM permissões (quando você usa IAM autenticação) ou em suas políticas de sessão de função de usuário (quando você usa a autenticação do IAM Identity Center) para o EMR Studio.

A tabela também exibe as operações permitidas em cada exemplo de política de permissões para o EMR Studio. Para obter mais informações sobre exemplos de políticas de permissões, consulte Crie políticas de permissões para usuários do EMR Studio.

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
Criação e exclusão de Workspaces	Sim	Sim	Sim	`"elasticmapreduce:CreateEditor",`

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
				`"elasticmapreduce:DescribeEditor", "elasticmapreduce:ListEditors", "elasticmapreduce:DeleteEditor"`
Visualização do painel Colaboração, habilitação da colaboração no Workspace e adição de colaboradores. Para obter mais informações, consulte Definição de propriedade para colaboração no Workspace.	Sim	Sim	Sim	`"elasticmapreduce:UpdateEditor", "elasticmapreduce:PutWorkspaceAccess", "elasticmapreduce:DeleteWorkspaceAccess", "elasticmapreduce:ListWorkspaceAccessIdentities"`
Veja uma lista de buckets de Amazon S3 Control armazenamento na mesma conta do Studio ao criar um novo EMR cluster e acesse os registros do contêiner ao usar uma interface de usuário da web para depurar aplicativos	Sim	Sim	Sim	`"s3:ListAllMyBuckets", "s3:ListBucket", "s3:GetBucketLocation", "s3:GetObject"`

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
Acesso aos Workspaces.	Sim	Sim	Sim	`"elasticmapreduce:DescribeEditor", "elasticmapreduce:ListEditors", "elasticmapreduce:StartEditor", "elasticmapreduce:StopEditor", "elasticmapreduce:OpenEditorInConsole"`
Anexe ou desanexe EMR clusters existentes da Amazon associados ao Workspace	Sim	Sim	Sim	`"elasticmapreduce:AttachEditor", "elasticmapreduce:DetachEditor", "elasticmapreduce:ListClusters", "elasticmapreduce:DescribeCluster", "elasticmapreduce:ListInstanceGroups", "elasticmapreduce:ListBootstrapActions"`

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
Anexe ou desanexe a Amazon EMR em clusters EKS	Sim	Sim	Sim	`"elasticmapreduce:AttachEditor", "elasticmapreduce:DetachEditor", "emr-containers:ListVirtualClusters", "emr-containers:DescribeVirtualCluster", "emr-containers:ListManagedEndpoints", "emr-containers:DescribeManagedEndpoint", "emr-containers:GetManagedEndpointSessionCredentials"`

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
Anexar ou desanexar aplicativos EMR sem servidor associados ao espaço de trabalho	Não	Sim	Sim	`"elasticmapreduce:AttachEditor", "elasticmapreduce:DetachEditor", "emr-serverless:GetApplication", "emr-serverless:StartApplication", "emr-serverless:ListApplications", "emr-serverless:GetDashboardForJobRun", "emr-serverless:AccessInteractiveEndpoints", "iam:PassRole"` A `PassRole` permissão é necessária para passar a função de tempo de execução do trabalho EMR sem servidor. Para obter mais informações, consulte Job runtime roles no Amazon EMR Serverless User Guide.

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
Depure a Amazon EMR em EC2 trabalhos com interfaces de usuário de aplicativos persistentes	Sim	Sim	Sim	`"elasticmapreduce:CreatePersistentAppUI", "elasticmapreduce:DescribePersistentAppUI", "elasticmapreduce:GetPersistentAppUIPresignedURL", "elasticmapreduce:ListClusters", "elasticmapreduce:ListSteps", "elasticmapreduce:DescribeCluster", "s3:ListBucket", "s3:GetObject"`
Depure a Amazon EMR em EC2 trabalhos com interfaces de usuário de aplicativos em cluster	Sim	Sim	Sim	`"elasticmapreduce:GetOnClusterAppUIPresignedURL"`

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
Depure a Amazon EMR em execuções de EKS trabalhos usando o Spark History Server	Sim	Sim	Sim	`"elasticmapreduce:CreatePersistentAppUI", "elasticmapreduce:DescribePersistentAppUI", "elasticmapreduce:GetPersistentAppUIPresignedURL", "emr-containers:ListVirtualClusters", "emr-containers:DescribeVirtualCluster", "emr-containers:ListJobRuns", "emr-containers:DescribeJobRun", "s3:ListBucket", "s3:GetObject"`
Criação e exclusão de repositórios Git.	Sim	Sim	Sim	`"elasticmapreduce:CreateRepository", "elasticmapreduce:DeleteRepository", "elasticmapreduce:ListRepositories", "elasticmapreduce:DescribeRepository", "secretsmanager:CreateSecret", "secretsmanager:ListSecrets", "secretsmanager:TagResource"`

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
Vinculação e desvinculação de repositórios Git.	Sim	Sim	Sim	`"elasticmapreduce:LinkRepository", "elasticmapreduce:UnlinkRepository", "elasticmapreduce:ListRepositories", "elasticmapreduce:DescribeRepository"`
Criação de novos clusters a partir de modelos de cluster definidos previamente.	Não	Sim	Sim	`"servicecatalog:SearchProducts", "servicecatalog:DescribeProduct", "servicecatalog:DescribeProductView", "servicecatalog:DescribeProvisioningParameters", "servicecatalog:ProvisionProduct", "servicecatalog:UpdateProvisionedProduct", "servicecatalog:ListProvisioningArtifacts", "servicecatalog:DescribeRecord", "servicecatalog:ListLaunchPaths", "cloudformation:DescribeStackResources", "elasticmapreduce:ListClusters", "elasticmapreduce:DescribeCluster"`

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
Forneça uma configuração de cluster para criar clusters.	Não	Não	Sim	`"elasticmapreduce:RunJobFlow", "iam:PassRole", "elasticmapreduce:ListClusters", "elasticmapreduce:DescribeCluster"`
Atribua um usuário a um Studio ao usar o modo de IAM autenticação.	Não	Não	Não	`"elasticmapreduce:CreateStudioPresignedUrl"`
Descrição dos objetos das redes.	Sim	Sim	Sim	`{ "Version": "2012-10-17", "Statement": [ { "Sid": "DescribeNetwork", "Effect": "Allow", "Action": [ "ec2:DescribeVpcs", "ec2:DescribeSubnets", "ec2:DescribeSecurityGroups" ], "Resource": "*" } ] }`

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
Listar IAM funções.	Sim	Sim	Sim	`{ "Version": "2012-10-17", "Statement": [ { "Sid": "ListIAMRoles", "Effect": "Allow", "Action": [ "iam:ListRoles" ], "Resource": "*" } ] }`
Conecte-se ao EMR Studio a partir do Amazon SageMaker Studio e use a interface visual do Data Wrangler.	Não	Não	Sim	`"sagemaker:CreatePresignedDomainUrl", "sagemaker:DescribeDomain", "sagemaker:ListDomains", "sagemaker:ListUserProfiles"`
Use a Amazon CodeWhisperer em seu EMR estúdio.	Não	Não	Sim	`"codewhisperer:GenerateRecommendations"`

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
Acesse o SQL editor Amazon Athena a partir do seu EMR Studio. Essa lista pode não incluir todas as permissões necessárias para usar todos os recursos do Athena. Para ver a up-to-date lista completa, consulte a política de acesso total do Athena.	Não	Não	Sim	`"athena:StartQueryExecution", "athena:StopQueryExecution", "athena:GetQueryExecution", "athena:GetQueryRuntimeStatistics", "athena:GetQueryResults", "athena:ListQueryExecutions", "athena:BatchGetQueryExecution", "athena:GetNamedQuery", "athena:ListNamedQueries", "athena:BatchGetNamedQuery", "athena:UpdateNamedQuery", "athena:DeleteNamedQuery", "athena:ListDataCatalogs", "athena:GetDataCatalog", "athena:ListDatabases", "athena:GetDatabase", "athena:ListTableMetadata", "athena:GetTableMetadata", "athena:ListWorkGroups", "athena:GetWorkGroup", "athena:CreateNamedQuery", "athena:GetPreparedStatement", "glue:CreateDatabase",`

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
				"glue:DeleteDatabase", "glue:GetDatabase", "glue:GetDatabases", "glue:UpdateDatabase", "glue:CreateTable", "glue:DeleteTable", "glue:BatchDeleteTable", "glue:UpdateTable", "glue:GetTable", "glue:GetTables", "glue:BatchCreatePartition", "glue:CreatePartition", "glue:DeletePartition", "glue:BatchDeletePartition", "glue:UpdatePartition", "glue:GetPartition", "glue:GetPartitions", "glue:BatchGetPartition", "kms:ListAliases", "kms:ListKeys", "kms:DescribeKey", "lakeformation:GetDataAccess", "s3:GetBucketLocation", "s3:GetBucketLocation", "s3:GetObject", "s3:ListBucket", "s3:ListBucketMultipartUploads", "s3:ListMultipartUploadParts", "s3:AbortMultipartUpload", "s3:PutObject", "s3:PutBucketPublicAccessBlock",

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
				`"s3:ListAllMyBuckets"`

Crie um EMR estúdio

Você pode criar um EMR estúdio para sua equipe com o EMR console da Amazon ou AWS CLI o. A criação de uma instância do Studio faz parte da configuração do Amazon EMR Studio.

Pré-requisitos

Antes de criar um Studio, certifique-se de ter concluído as tarefas anteriores em Configurar um Amazon EMR Studio.

Para criar um Studio usando o AWS CLI, você deve ter a versão mais recente instalada. Para obter mais informações, consulte Instalar ou atualizar a versão mais recente da AWS CLI.

> **Important**
>
> Desative as ferramentas de gerenciamento de proxy, como FoxyProxy ou SwitchyOmega no navegador, antes de criar um Studio. Os proxies ativos podem resultar em uma mensagem de erro de falha de rede quando você escolhe Criar Studio.

EMRA Amazon oferece uma experiência de console simples para criar um Studio, para que você possa começar rapidamente com as configurações padrão para executar cargas de trabalho interativas ou trabalhos em lote com as configurações padrão. A criação de um EMR estúdio também cria um aplicativo EMR sem servidor pronto para seus trabalhos interativos.

Se quiser ter controle total sobre as configurações do seu Studio, você pode escolher Personalizado, que permite definir todas as configurações adicionais.

Interactive workloads

Para criar um EMR estúdio para cargas de trabalho interativas

1. Abra o EMR console da Amazon em https://console.aws.amazon.com/emr.

2. Em EMRStudio, no painel de navegação à esquerda, escolha Introdução. Você também pode criar um novo Studio na página Studios.
3. EMRA Amazon fornece configurações padrão para você se você estiver criando um EMR Studio para cargas de trabalho interativas, mas você pode editar essas configurações. As configurações configuráveis incluem o nome do EMR Studio, a localização do S3 para seu espaço de trabalho, a função de serviço a ser usada, o (s) espaço (s) de trabalho que você deseja usar, o nome do aplicativo EMR sem servidor e a função de tempo de execução associada.
4. Escolha Create Studio e inicie o Workspace para finalizar e navegar até a página Studios. Seu novo estúdio aparece na lista com detalhes como nome do estúdio, data de criação e acesso ao estúdio URL. Seu espaço de trabalho é aberto em uma nova guia no seu navegador.

Batch jobs

Para criar um EMR estúdio para cargas de trabalho interativas

1. Abra o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. Em EMRStudio, no painel de navegação à esquerda, escolha Introdução. Você também pode criar um novo Studio na página Studios.
3. EMRA Amazon fornece configurações padrão para você se você estiver criando um EMR Studio para trabalhos em lote, mas você pode editar essas configurações. As configurações configuráveis incluem o nome do EMR Studio, o nome do aplicativo EMR sem servidor e a função de tempo de execução associada.
4. Escolha Create Studio e inicie o Workspace para finalizar e navegar até a página Studios. Seu novo estúdio aparece na lista com detalhes como nome do estúdio, data de criação e acesso ao estúdio URL. Seu EMR estúdio é aberto em uma nova guia no seu navegador.

Custom settings

Para criar um EMR estúdio com configurações personalizadas

1. Abra o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. Em EMRStudio, no painel de navegação à esquerda, escolha Introdução. Você também pode criar um novo Studio na página Studios.
3. Escolha Criar um Studio para abrir a página Criar um Studio.

4. Insira o nome do estúdio.
5. Escolha criar um novo bucket do S3 ou usar um local existente.
6. Escolha o espaço de trabalho a ser adicionado ao Studio. Você pode adicionar até 3 espaços de trabalho.
7. Em Autenticação, escolha um modo de autenticação para o Studio e forneça informações de acordo com a tabela a seguir. Para saber mais sobre a autenticação do EMR Studio, consulteEscolha um modo de autenticação para o Amazon EMR Studio.

Se você usar...	Fazer isso...
IAMautenticação ou federação	O método de autenticação padrão é AWS Identity and Access Management (IAM). Na parte inferior da tela, você também pode adicionar tags para dar acesso ao Studio para usuários específicos, conforme descrito em Atribuir um usuário ou grupo a um EMR estúdio. Se você quiser que os usuários federados façam login usando o Studio URL e as credenciais do seu provedor de identidade (IdP), selecione seu IdP na lista suspensa e insira seu nome de login e parâmetro do provedor de identidade (IdP). URL RelayState Para obter uma lista de RelayState nomes URLs e autenticação de IdP, consulte. RelayState Parâmetros e autenticação do provedor de identidade URLs
IAMautenticação do Identity Center	Selecione sua função de serviço do EMR Studio e sua função de usuário. Para ter mais informações, consulte Criar uma função de serviço do EMR Studio e Crie uma função de usuário do EMR Studio para

Se você usar...	Fazer isso...
	o modo de autenticação do IAM Identity Center. Ao usar a autenticação do IAM Identity Center (antigo AWS Single Sign On) para o Studio, você pode optar por simplificar a experiência de login dos usuários com a opção Habilitar propagação de identidade confiável. Com a propagação de identidad e confiável, os usuários podem fazer login com suas credenciais do Identity Center e ter suas identidades propagadas para AWS serviços posteriores ao usarem o Studio. Na seção Acesso à aplicação, você também pode especificar se todos os usuários e grupos no seu Centro de Identidade devem ter acesso ao Studio ou se somente os usuários e grupos atribuídos que você escolher podem acessá-lo. Para obter mais informaçõesIntegre a Amazon EMR com AWS IAM Identity Center, consulte e também Propagação confiável de identidade entre aplicativos no Guia do Usuário do AWS IAM Identity Center.

8. ParaVPC, escolha uma Amazon Virtual Private Cloud (VPC) para o Studio na lista suspensa.
9. Em Sub-redes, selecione no máximo cinco sub-redes VPC para associar ao Studio. Você tem a opção de adicionar mais sub-redes após a criação do Studio.
10. Em Grupos de segurança, escolha os grupos de segurança padrão ou os grupos de segurança personalizados. Para obter mais informações, consulte Defina grupos de segurança para controlar o tráfego de rede do EMR Studio.

Se você escolher...	Fazer isso...
Os grupos de segurança padrão do EMR Studio	Para habilitar a vinculação de repositórios baseados em Git para o Studio, escolha Habilitar clusters, endpoints e o repositór io Git. Caso contrário, escolha Habilitar clusters e endpoints.
Os grupos de segurança personalizados para seu Studio	• Em Grupo de segurança de cluster e endpoint, selecione o grupo de segurança do mecanismo que você configurou usando a lista suspensa. Seu Studio usa esse grupo de segurança para permitir o acesso de entrada de Workspaces anexados. • Em Grupo de segurança do Workspace , selecione o grupo de segurança do Workspace que você configurou usando a lista suspensa. Seu estúdio usa esse grupo de segurança com o Workspace s para fornecer acesso externo a EMR clusters Amazon conectados e repositór ios Git hospedados publicamente.

11. Adicione tags ao seu Studio e a outros recursos. Para obter mais informações sobre tags, consulte Clusters de tags.
12. Escolha Create Studio e inicie o Workspace para finalizar e navegar até a página Studios. Seu novo estúdio aparece na lista com detalhes como nome do estúdio, data de criação e acesso ao estúdio URL.

Depois de criar um Studio, siga as instruções em Atribuir um usuário ou grupo a um EMR estúdio.

CLI

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

Example — Crie um EMR estúdio que use IAM para autenticação

O AWS CLI comando de exemplo a seguir cria um EMR Studio com modo de IAM autenticação. Ao usar IAM autenticação ou federação para o Studio, você não especifica um`--user-role`.

Para permitir que usuários federados façam login usando o Studio URL e as credenciais do seu provedor de identidade (IdP), especifique seu e. `--idp-auth-url` `--idp-relay-state-parameter-name` Para obter uma lista de RelayState nomes URLs e autenticação de IdP, consulte. RelayState Parâmetros e autenticação do provedor de identidade URLs

```
aws emr create-studio \
--name <example-studio-name> \
--auth-mode IAM \
--vpc-id <example-vpc-id> \
--subnet-ids <subnet-id-1> <subnet-id-2>... <subnet-id-5>  \
--service-role <example-studio-service-role-name> \
--user-role studio-user-role-name \
--workspace-security-group-id <example-workspace-sg-id> \
--engine-security-group-id <example-engine-sg-id> \
--default-s3-location <example-s3-location> \
--idp-auth-url <https://EXAMPLE/login/> \
--idp-relay-state-parameter-name <example-RelayState>
```

Example — Crie um EMR estúdio que use o Identity Center para autenticação

O comando de AWS CLI exemplo a seguir cria um EMR Studio que usa o modo de autenticação do IAM Identity Center. Ao usar a autenticação do IAM Identity Center, você deve especificar um`--user-role`.

Para obter mais informações sobre o modo de autenticação do IAM Identity Center, consulteConfigurar o modo de autenticação do IAM Identity Center para o Amazon EMR Studio.

```
aws emr create-studio \
--name <example-studio-name> \
--auth-mode SSO \
--vpc-id <example-vpc-id> \
--subnet-ids <subnet-id-1> <subnet-id-2>... <subnet-id-5>  \
--service-role <example-studio-service-role-name> \
--user-role <example-studio-user-role-name> \
--workspace-security-group-id <example-workspace-sg-id> \
--engine-security-group-id <example-engine-sg-id> \
--default-s3-location <example-s3-location>
--trusted-identity-propagation-enabled \
--idc-user-assignment OPTIONAL \
--idc-instance-arn <iam-identity-center-instance-arn>
```

Example — CLI saída para **`aws emr create-studio`**

A seguir, é apresentado um exemplo da saída que aparece após a criação de um Studio.

```
{
    StudioId: "es-123XXXXXXXXX",
    Url: "https://es-123XXXXXXXXX.emrstudio-prod.us-east-1.amazonaws.com"
}
```

Para obter mais informações sobre o comando `create-studio`, consulte AWS CLI Command Reference.

RelayState Parâmetros e autenticação do provedor de identidade URLs

Quando você usa a IAM federação e deseja que os usuários façam login usando seu Studio URL e as credenciais do seu provedor de identidade (IdP), você pode especificar o URL login do provedor de identidade (IdP) RelayStatee o nome do parâmetro quando quiser. Crie um EMR estúdio

A tabela a seguir mostra a autenticação padrão URL e o nome do RelayState parâmetro para alguns provedores de identidade populares.

Provedor de identidades	Parâmetro	Autentificação URL
Auth0	`RelayState`	`https://<sub_domain> .auth0.com/samlp/<app_id>`

Provedor de identidades	Parâmetro	Autentificação URL
Contas do Google	`RelayState`	`https://accounts.google.com/o/saml2/initsso?idpid= <idp_id>&spid=<sp_id>&forceauthn=false`
Microsoft Azure	`RelayState`	`https://myapps.microsoft.com/signin/ <app_name> /<app_id>?tenantId= <tenant_id>`
Okta	`RelayState`	`https://<sub_domain> .okta.com/app/<app_name> /<app_id>/sso/saml`
PingFederate	`TargetResource`	`https://<host>/idp/<idp_id>/startSSO.ping?PartnerSpId= <sp_id>`
PingOne	`TargetResource`	`https://sso.connect.pingidentity.com/sso/sp/initsso?saasid= <app_id>&idpid=<idp_id>`

Atribuir e gerenciar usuários do EMR Studio

Depois de criar um EMR Studio, você pode atribuir usuários e grupos a ele. O método usado para atribuir, atualizar e remover usuários depende do modo de autenticação do Studio.

- Ao usar o modo de IAM autenticação, você configura a atribuição e as permissões do usuário do EMR Studio em IAM ou com IAM seu provedor de identidade.
- Com o modo de autenticação do IAM Identity Center, você usa o console EMR de gerenciamento da Amazon ou o AWS CLI para gerenciar usuários.

Para saber mais sobre a autenticação para o Amazon EMR Studio, consulteEscolha um modo de autenticação para o Amazon EMR Studio.

Atribuir um usuário ou grupo a um EMR estúdio

IAM

Ao usarConfigurar o modo de IAM autenticação para o Amazon EMR Studio, você deve permitir a `CreateStudioPresignedUrl` ação na política de IAM permissões de um usuário e restringir o usuário a um estúdio específico. Você pode incluir `CreateStudioPresignedUrl` em suas Permissões de usuários para o modo de autenticação do IAM ou usar uma política separada.

Para restringir um usuário a um Studio (ou conjunto de estúdios), você pode usar o controle de acesso baseado em atributos (ABAC) ou especificar o Amazon Resource Name (ARN) de um estúdio no `Resource` elemento da política de permissões.

Example Atribuir um usuário a um estúdio usando um estúdio ARN

O exemplo de política a seguir dá ao usuário acesso a um EMR estúdio específico, permitindo a `CreateStudioPresignedUrl` ação e especificando o Amazon Resource Name (ARN) do estúdio no `Resource` elemento.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowCreateStudioPresignedUrl",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:CreateStudioPresignedUrl"
            ],
            "Resource": "arn:aws:elasticmapreduce:<region>:<account-
id>:studio/<studio-id>"
        }
    ]
}
```

Example Atribuir um usuário a um Studio com ABAC para IAM autenticação

Há várias maneiras de configurar o controle de acesso baseado em atributos (ABAC) para um Studio. Por exemplo, você pode anexar uma ou mais tags a um EMR estúdio e, em seguida, criar uma IAM política que restrinja a `CreateStudioPresignedUrl` ação a um determinado estúdio ou conjunto de estúdios com essas tags.

Você pode adicionar etiquetas durante ou após a criação do Studio. Para adicionar etiquetas a um Studio existente, você pode usar o comando AWS CLIemr add-tags. O exemplo a seguir adiciona uma tag com o par de valores-chave Team = Data Analytics a um EMR Studio.

```
aws emr add-tags --resource-id <example-studio-id> --tags Team="Data Analytics"
```

O exemplo de política de permissões a seguir permite a CreateStudioPresignedUrl ação do EMR Studios com o par chave-valor da tag. Team = DataAnalytics Para obter mais informações sobre como usar etiquetas para controlar o acesso, consulte Controle de acesso para usuários e funções do IAM usando etiquetas ou Controlar o acesso a recursos da AWS usando etiquetas.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowCreateStudioPresignedUrl",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:CreateStudioPresignedUrl"
            ],
            "Resource": "arn:aws:elasticmapreduce:<region>:<account-id>:studio/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/Team": "Data Analytics"
                }
            }
        }
    ]
}
```

Example Atribua um usuário a um Studio usando a chave de condição SourceIdentity global aws:

Ao usar a IAM federação, você pode usar a chave de condição global aws:SourceIdentity em uma política de permissões para dar aos usuários acesso ao Studio quando eles assumirem sua IAM função na federação.

Primeiro, você deve configurar seu provedor de identidade (IdP) para retornar uma sequência de caracteres de identificação, como um endereço de e-mail ou nome de usuário, quando um usuário se autentica e assume sua IAM função na federação. IAMdefine a chave de condição global aws:SourceIdentity para a string de identificação retornada pelo seu IdP.

Para obter mais informações, consulte a postagem do blog Como relacionar a atividade da IAM função à identidade corporativa no Blog de AWS Segurança e a SourceIdentity entrada aws: na referência global de chaves de condição.

O exemplo de política a seguir permite a `CreateStudioPresignedUrl` ação e fornece aos usuários um `aws:SourceIdentity` que corresponda ao *<example-source-identity>* acesso ao EMR estúdio especificado por *<example-studio-arn>*.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "elasticmapreduce:CreateStudioPresignedUrl",
      "Resource": "<example-studio-arn>",
      "Condition": {
        "StringLike": {
          "aws:SourceIdentity": "<example-source-identity>"
        }
      }
    }
  ]
}
```

IAM Identity Center

Ao atribuir um usuário ou grupo a um EMR Studio, você especifica uma política de sessão que define permissões refinadas, como a capacidade de criar um novo EMR cluster para esse usuário ou grupo. A Amazon EMR armazena esses mapeamentos de políticas de sessão. É possível atualizar a política de sessão de um usuário ou de um grupo após a atribuição.

> **Note**
>
> As permissões finais para um usuário ou grupo são uma interseção entre as permissões definidas na sua função de usuário do EMR Studio e as permissões definidas na política de sessão desse usuário ou grupo. Se um usuário pertencer a mais de um grupo atribuído ao Studio, o EMR Studio usa uma união de permissões para esse usuário.

Para atribuir usuários ou grupos a um EMR estúdio usando o EMR console da Amazon

1. Navegue até o novo EMR console da Amazon e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Escolha EMRStudio no painel de navegação à esquerda.
3. Escolha o nome do seu Studio na lista Studios ou selecione o Studio e escolha Visualizar detalhes para abrir a página de detalhes do Studio.
4. Escolha Adicionar usuários para visualizar a tabela de pesquisa para Usuários e Grupos.
5. Selecione a guia Usuários ou a guia Grupos e insira um termo de pesquisa na barra de pesquisa para localizar um usuário ou um grupo.
6. Selecione um ou mais usuários ou grupos na lista de resultados da pesquisa. É possível alternar entre as guias Usuários e Grupos.
7. Após selecionar os usuários e os grupos a serem adicionados ao Studio, escolha Adicionar. Você deve visualizar os usuários e os grupos na lista Usuários do Studio. Pode demorar alguns segundos para que a lista seja atualizada.
8. Siga as instruções em Atualizar permissões para um usuário ou um grupo atribuído a um Studio para aprimorar as permissões do Studio para um usuário ou um grupo.

Para atribuir um usuário ou grupo a um EMR estúdio usando o AWS CLI

Insira seus próprios valores para os argumentos `create-studio-session-mapping` a seguir. Para obter mais informações sobre o comando `create-studio-session-mapping`, consulte AWS CLI Command Reference.

- **`--studio-id`**: o ID do Studio ao qual você deseja atribuir o usuário ou o grupo. Para obter instruções sobre como recuperar um ID do Studio, consulte Visualização de detalhes do Studio.
- **`--identity-name`**: o nome do usuário ou do grupo no repositório de identidades. Para obter mais informações, consulte UserNamepara usuários e DisplayNamegrupos na APIReferência do Identity Store.
- **`--identity-type`**: use `USER` ou `GROUP` para especificar o tipo de identidade.
- **`--session-policy-arn`**— O Amazon Resource Name (ARN) para a política de sessão que você deseja associar ao usuário ou grupo. Por exemplo, **`arn:aws:iam::`*`<aws-account-id>`*`:policy/`*`EMRStudio_Advanced_User_Policy`***. Para obter mais informações, consulte Crie políticas de permissões para usuários do EMR Studio.

```
aws emr create-studio-session-mapping \
 --studio-id <example-studio-id> \
 --identity-name <example-identity-name> \
 --identity-type <USER-or-GROUP> \
 --session-policy-arn <example-session-policy-arn>
```

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

Use o comando `get-studio-session-mapping` para verificar a nova atribuição. Substituir *<example-identity-name>* com o nome do IAM Identity Center do usuário ou grupo que você atualizou.

```
aws emr get-studio-session-mapping \
 --studio-id <example-studio-id> \
 --identity-type <USER-or-GROUP> \
 --identity-name <user-or-group-name> \
```

Atualizar permissões para um usuário ou um grupo atribuído a um Studio

IAM

Para atualizar as permissões de usuário ou grupo ao usar o modo de IAM autenticação, use IAM para alterar as políticas de IAM permissões anexadas às suas IAM identidades (usuários, grupos ou funções).

Para obter mais informações, consulte Permissões de usuários para o modo de autenticação do IAM.

IAM Identity Center

Para atualizar as permissões do EMR Studio para um usuário ou grupo usando o console

1. Navegue até o novo EMR console da Amazon e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.

2. Escolha EMRStudio no painel de navegação à esquerda.
3. Escolha o nome do seu Studio na lista Studios ou selecione o Studio e escolha Visualizar detalhes para abrir a página de detalhes do Studio.
4. Na lista de Usuários do Studio na página de detalhes do Studio, pesquise o usuário ou o grupo que você deseja atualizar. É possível pesquisar por nome ou por tipo de identidade.
5. Selecione o usuário ou o grupo que deseja atualizar e escolha Atribuir política para abrir a caixa de diálogo Política de sessão.
6. Selecione uma política para aplicar ao usuário ou ao grupo escolhido na etapa 5 e escolha Aplicar política. A lista Usuários do Studio deve exibir o nome da política na coluna Política de sessão para o usuário ou para o grupo que você atualizou.

Para atualizar as permissões do EMR Studio para um usuário ou grupo usando o AWS CLI

Insira seus próprios valores para os argumentos `update-studio-session-mappings` a seguir. Para obter mais informações sobre o comando `update-studio-session-mappings`, consulte AWS CLI Command Reference.

```
aws emr update-studio-session-mapping \
 --studio-id <example-studio-id> \
 --identity-name <name-of-user-or-group-to-update> \
 --session-policy-arn <new-session-policy-arn-to-apply> \
 --identity-type <USER-or-GROUP> \
```

Use o comando `get-studio-session-mapping` para verificar a nova atribuição de política de sessão. Substituir *<example-identity-name>* com o nome do IAM Identity Center do usuário ou grupo que você atualizou.

```
aws emr get-studio-session-mapping \
 --studio-id <example-studio-id> \
 --identity-type <USER-or-GROUP> \
 --identity-name <user-or-group-name> \
```

Remover um usuário ou um grupo de um Studio

IAM

Para remover um usuário ou grupo de um EMR Studio ao usar o modo de IAM autenticação, você deve revogar o acesso do usuário ao Studio reconfigurando a política de permissões do IAM usuário.

No exemplo de política a seguir, suponha que você tenha um EMR Studio com o par chave-valor da tag. `Team = Quality Assurance` De acordo com a política, o usuário pode acessar os Studios etiquetados com a chave `Team` cujo valor é igual a `Data Analytics` ou `Quality Assurance`. Para remover o usuário do Studio etiquetado com `Team = Quality Assurance`, remova `Quality Assurance` da lista de valores de etiqueta.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowCreateStudioPresignedUrl",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:CreateStudioPresignedUrl"
            ],
            "Resource": "arn:aws:elasticmapreduce:<region>:<account-id>:studio/*",
            "Condition": {
                "StringEquals": {
                    "emr:ResourceTag/Team": [
                        "Data Analytics",
                        "Quality Assurance"
                  ]
                }
            }
        }
    ]
}
```

IAM Identity Center

Para remover um usuário ou grupo de um EMR Studio usando o console

1. Navegue até o novo EMR console da Amazon e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Escolha EMRStudio no painel de navegação à esquerda.
3. Escolha o nome do seu Studio na lista Studios ou selecione o Studio e escolha Visualizar detalhes para abrir a página de detalhes do Studio.
4. Na lista de Usuários do Studio na página de detalhes do Studio, localize o usuário ou o grupo que você deseja remover do Studio. É possível pesquisar por nome ou por tipo de identidade.
5. Selecione o usuário ou o grupo que deseja excluir, escolha Excluir e confirme. O usuário ou o grupo excluído desaparecerá da lista Usuários do Studio.

Para remover um usuário ou grupo de um EMR estúdio usando o AWS CLI

Insira seus próprios valores para os argumentos `delete-studio-session-mapping` a seguir. Para obter mais informações sobre o comando `delete-studio-session-mapping`, consulte AWS CLI Command Reference.

```
aws emr delete-studio-session-mapping \
 --studio-id <example-studio-id> \
 --identity-type <USER-or-GROUP> \
 --identity-name <name-of-user-or-group-to-delete> \
```

Gerencie um Amazon EMR Studio

Esta seção inclui instruções para ajudar você a monitorar, atualizar ou excluir um recurso do EMR Studio. Para obter informações sobre como atribuir usuários ou atualizar as permissões de usuários, consulte Atribuir e gerenciar usuários do EMR Studio.

Visualização de detalhes do Studio

Console

Para ver detalhes sobre um EMR Studio com o novo console

1. Abra o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. Em EMREstúdio, no painel de navegação à esquerda, escolha Estúdios.
3. Selecione o Studio na lista Studios para abrir a página de detalhes do Studio. A página de detalhes do Studio inclui informações de configuração do Studio, como a descrição do Studio e as sub-redes. VPC

CLI

Para recuperar detalhes de um EMR Studio by Studio ID usando o AWS CLI

Use o `describe-studio` AWS CLI comando a seguir para obter informações detalhadas sobre um EMR estúdio específico. Para obter mais informações, consulte Referência de comandos da AWS CLI .

```
aws emr describe-studio \
 --studio-id <id-of-studio-to-describe> \
```

Para recuperar uma lista de EMR estúdios usando o AWS CLI

Use o seguinte comando `list-studios` AWS CLI : Para obter mais informações, consulte Referência de comandos da AWS CLI .

```
aws emr list-studios
```

Veja a seguir um exemplo de valor de retorno para o `list-studios` comando em JSON formato.

```
{
    "Studios": [
        {
            "AuthMode": "IAM",
            "VpcId": "vpc-b21XXXXX",
            "Name": "example-studio-name",
```

```
            "Url": "https://es-7HWP74SNGDXXXXXXXXXXXXXX.emrstudio-prod.us-
east-1.amazonaws.com",
            "CreationTime": 1605672582.781,
            "StudioId": "es-7HWP74SNGDXXXXXXXXXXXXXX",
            "Description": "example studio description"
        }
    ]
}
```

Monitore as ações do Amazon EMR Studio

Veja o EMR estúdio e a API atividade

EMRO Studio é integrado com AWS CloudTrail, um serviço que fornece um registro das ações realizadas por um usuário, por uma IAM função ou por outro AWS serviço no EMR Studio. CloudTrail captura API chamadas para o EMR Studio como eventos. Você pode ver os eventos usando o CloudTrail console em https://console.aws.amazon.com/cloudtrail/.

EMROs eventos do Studio fornecem informações como qual estúdio ou IAM usuário faz uma solicitação e que tipo de solicitação é essa.

> **Note**
>
> As ações no cluster, como execução de trabalhos de cadernos, não são emitidas para o AWS CloudTrail.

Você também pode criar uma trilha para entrega contínua de CloudTrail eventos do EMR Studio em um bucket do Amazon S3. Para obter mais informações, consulte o Guia do usuário do AWS CloudTrail.

Exemplo de CloudTrail evento: um usuário chama o DescribeStudio API

Veja a seguir um exemplo de AWS CloudTrail evento criado quando um usuário`admin`,, chama DescribeStudioAPIo. CloudTrail registra o nome do usuário como`admin`.

> **Note**
>
> Para proteger os detalhes do Studio, o API evento EMR Studio para DescribeStudio exclui um valor para`responseElements`.

```
{
   "eventVersion":"1.08",
   "userIdentity":{
      "type":"IAMUser",
      "principalId":"AIDXXXXXXXXXXXXXXXXX",
      "arn":"arn:aws:iam::653XXXXXXXXX:user/admin",
      "accountId":"653XXXXXXXXX",
      "accessKeyId":"AKIAIOSFODNN7EXAMPLE",
      "userName":"admin"
   },
   "eventTime":"2021-01-07T19:13:58Z",
   "eventSource":"elasticmapreduce.amazonaws.com",
   "eventName":"DescribeStudio",
   "awsRegion":"us-east-1",
   "sourceIPAddress":"72.XX.XXX.XX",
   "userAgent":"aws-cli/1.18.188 Python/3.8.5 Darwin/18.7.0 botocore/1.19.28",
   "requestParameters":{
      "studioId":"es-9O5XXXXXXXXXXXXXXXXXXXX"
   },
   "responseElements":null,
   "requestID":"0fxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx",
   "eventID":"b0xxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx",
   "readOnly":true,
   "eventType":"AwsApiCall",
   "managementEvent":true,
   "eventCategory":"Management",
   "recipientAccountId":"653XXXXXXXXX"
}
```


Visualização de atividades de usuários e de trabalhos do Spark

Para visualizar a atividade de trabalho dos usuários do Amazon EMR Studio no Spark, você pode configurar a representação do usuário em um cluster. Com a representação do usuário, cada trabalho do Spark enviado de um Workspace é associado ao usuário do Studio que executou o código.

Quando a representação de usuário está ativada, a Amazon EMR cria um diretório de HDFS usuários no nó principal do cluster para cada usuário que executa código no espaço de trabalho. Por exemplo, se o usuário `studio-user-1@example.com` executar um código, você poderá se conectar ao nó primário e visualizar que `hadoop fs -ls /user` tem um diretório para `studio-user-1@example.com`.

Para configurar a representação do usuário do Spark, defina as propriedades abaixo nas seguintes classificações de configuração:

- `core-site`
- `livy-conf`

```
[
    {
        "Classification": "core-site",
        "Properties": {
          "hadoop.proxyuser.livy.groups": "*",
          "hadoop.proxyuser.livy.hosts": "*"
        }
    },
    {
        "Classification": "livy-conf",
        "Properties": {
          "livy.impersonation.enabled": "true"
        }
    }
]
```

Para visualizar as páginas do servidor de histórico, consulte Depure aplicativos e trabalhos com EMR o Studio. Você também pode se conectar ao nó primário do cluster usando SSH para visualizar as interfaces web do aplicativo. Para obter mais informações, consulte Visualize interfaces web hospedadas em EMR clusters da Amazon.

Atualizar um Amazon EMR Studio

Depois de criar um EMR Studio, você pode atualizar os seguintes atributos usando o AWS CLI:

- Nome
- Descrição
- Local do S3 padrão
- Sub-redes

Para atualizar um EMR Studio usando o AWS CLI

Use o `update-studio` AWS CLI comando para atualizar um EMR Studio. Para obter mais informações, consulte Referência de comandos da AWS CLI .

> **Note**
>
> Você pode associar um Studio a, no máximo, cinco sub-redes. Essas sub-redes devem pertencer às VPC mesmas do Studio. A lista de sub-redes IDs que você envia para o `update-studio` comando pode incluir uma nova sub-redeIDs, mas também deve incluir toda a sub-rede IDs que você já associou ao Studio. Não é possível remover sub-redes de um Studio.

```
aws emr update-studio \
 --studio-id <example-studio-id-to-update> \
 --name <example-new-studio-name> \
 --subnet-ids <old-subnet-id-1 old-subnet-id-2 old-subnet-id-3 new-subnet-id> \
```

Para verificar as alterações, use o `describe-studio` AWS CLI comando e especifique sua ID do estúdio. Para obter mais informações, consulte Referência de comandos da AWS CLI .

```
aws emr describe-studio \
 --studio-id <id-of-updated-studio> \
```

Excluir um Amazon EMR Studio e um Workspaces

Quando você exclui um Studio, o EMR Studio exclui todas as atribuições de usuários e grupos do IAM Identity Center associadas ao Studio.

> **Note**
>
> Quando você exclui um Studio, EMR a Amazon não exclui os Workspaces associados a esse Studio. Você deve excluir os Workspaces do seu Studio separadamente.

Exclusão de Workspaces

Console

Como cada EMR Studio Workspace é uma instância de EMR notebook, você pode usar o console de EMR gerenciamento da Amazon para excluir Workspaces. Você pode excluir espaços de trabalho usando o EMR console da Amazon antes ou depois de excluir seu Studio

Para excluir um espaço de trabalho usando o console da Amazon EMR

1. Navegue até o novo EMR console da Amazon e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Escolha Cadernos.
3. Selecione os Workspaces que você deseja excluir.
4. Escolha Excluir e, em seguida, selecione Excluir novamente para confirmar.
5. Siga as instruções para Excluir objetos no Guia do usuário do console do Amazon Simple Storage Service para remover os arquivos de cadernos associados ao Workspace excluído do Amazon S3.

EMR Studio UI

From the Workspace UI

Excluir um espaço de trabalho e seus arquivos de backup associados do Studio EMR

1. Faça login no seu EMR Studio com seu acesso ao Studio URL e escolha Workspaces no painel de navegação à esquerda.
2. Localize seu Workspace na lista e, em seguida, marque a caixa de seleção ao lado do nome. É possível selecionar vários Workspaces a serem excluídos ao mesmo tempo.
3. Escolha Excluir no canto superior direito da lista Workspaces e confirme que deseja excluir os Workspaces selecionados. Escolha Delete para confirmar.
4. Se você desejar remover os arquivos de cadernos associados ao Workspace excluído do Amazon S3, siga as instruções para Excluir objetos no Guia do usuário do console do Amazon Simple Storage Service. Se não foi você quem criou o Studio, consulte o administrador do Studio para determinar o local de backup do Amazon S3 para o Workspace excluído.

From the Workspaces list

Exclusão de um Workspace e dos arquivos de backup associados da lista Workspaces

1. Navegue até a lista Workspaces no console.
2. Selecione o Workspace que deseja excluir da lista e, em seguida, escolha Ações.
3. Escolha Excluir.
4. Se você desejar remover os arquivos de cadernos associados ao Workspace excluído do Amazon S3, siga as instruções para Excluir objetos no Guia do usuário do console do Amazon Simple Storage Service. Se não foi você quem criou o Studio, consulte o administrador do Studio para determinar o local de backup do Amazon S3 para o Workspace excluído.

Excluir um EMR estúdio

Console

Para excluir um EMR Studio com o novo console

1. Abra o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. Em EMREstúdio, no painel de navegação à esquerda, escolha Estúdios.
3. Selecione o Studio na lista Studios com o botão de alternância à esquerda do nome do Studio. Escolha Excluir.

Old console

Para excluir um EMR Studio com o console antigo

1. Abra o EMR console da Amazon em https://console.aws.amazon.com/elasticmapreduce/casa.
2. Escolha EMRStudio no painel de navegação à esquerda.
3. Selecione o Studio na lista Studios e escolha Excluir.

CLI

Para excluir um EMR estúdio com o AWS CLI

Use o `delete-studio` AWS CLI comando para excluir um EMR Studio. Para obter mais informações, consulte Referência de comandos da AWS CLI .

```
aws emr delete-studio --studio-id <id-of-studio-to-delete>
```

Criptografando cadernos e arquivos do espaço de trabalho do EMR Studio

No EMR Studio, você pode criar e configurar diferentes áreas de trabalho para organizar e executar notebooks. Esses espaços de trabalho armazenam cadernos e arquivos relacionados em seu bucket Amazon S3 especificado. Por padrão, esses arquivos são criptografados com chaves gerenciadas pelo Amazon S3 (SSE-S3) com criptografia do lado do servidor como nível básico de criptografia. Você também pode optar por usar KMS chaves gerenciadas pelo cliente (SSE-KMS) para criptografar seus arquivos. Você pode fazer isso usando o console EMR de gerenciamento da Amazon ou por meio do AWS CLI e AWS SDK ao criar um EMR Studio.

EMRA criptografia de armazenamento do espaço de trabalho do Studio está disponível em todas as regiões em que o EMR Studio está disponível.

Pré-requisitos

Antes de criptografar o caderno e os arquivos do espaço de trabalho do EMR Studio, você deve usar AWS Key Management Service para criar uma chave simétrica de gerente de clientes (CMK) na mesma Conta da AWS região do seu Studio. EMR

Sua política de recursos AWS KMS deve ter as permissões de acesso necessárias para a função de serviço do seu EMR Studio. Veja a seguir um exemplo de IAM política que concede permissões mínimas de acesso à criptografia de armazenamento do EMR Studio Workspace:

```
{
    "Sid": "AllowEMRStudioServiceRoleAccess",
    "Effect": "Allow",
    "Principal": {
        "AWS": "arn:aws:iam::<ACCOUNT_ID>:role/<ROLE_NAME>"
    },
    "Action": [
        "kms:Decrypt",
        "kms:GenerateDataKey",
        "kms:ReEncryptFrom",
        "kms:ReEncryptTo",
```

```
            "kms:DescribeKey"
        ],
        "Resource": "*",
        "Condition": {
            "StringEquals": {
                "kms:CallerAccount": "<ACCOUNT_ID>",
                "kms:EncryptionContext:aws:s3:arn": "arn:aws:s3:::<S3_BUCKET_NAME>",
                "kms:ViaService": "s3.<AWS_REGION>.amazonaws.com"
            }
        }
}
```

Sua função de serviço do EMR Studio também deve ter as permissões de acesso para usar sua AWS KMS chave. Veja a seguir um exemplo de IAM política que concede as permissões mínimas de acesso à criptografia de armazenamento do EMR Studio Workspace:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowEMRStudioWorkspaceStorageEncryptionAccess",
            "Effect": "Allow",
            "Action": [
               "kms:Decrypt",
               "kms:GenerateDataKey",
               "kms:ReEncryptFrom",
               "kms:ReEncryptTo",
               "kms:DescribeKey"
            ],
            "Resource": ["arn:aws:kms:<REGION>:<ACCOUNT_ID>:key/<KEY_IDENTIFIER>"]
        }
    ]
}
```

Configuração

Siga estas etapas para criar um novo EMR Studio que usa criptografia de armazenamento do espaço de trabalho.

1. Abra o EMR console da Amazon em https://console.aws.amazon.com/elasticmapreduce/.
2. Escolha Estúdios e, em seguida, escolha Criar estúdio.

3. Para a localização do S3 para armazenamento, insira ou escolha um caminho do Amazon S3. Esse é o local do Amazon S3 onde a Amazon EMR armazena cadernos e arquivos do espaço de trabalho.
4. Em Função de serviço, insira ou escolha uma IAM função. Esse é o IAM papel que a Amazon EMR assume.
5. Escolha Criptografar arquivos do Workspace com sua própria chave. AWS KMS
6. Insira ou escolha uma AWS KMS chave para usar para criptografar cadernos e arquivos do espaço de trabalho no Amazon S3.
7. Escolha Create Studio ou Create Studio e inicie espaços de trabalho.
8. Escolha Criptografar arquivos do Workspace com sua própria chave. AWS KMS
9. Insira ou escolha um AWS KMS para usar para criptografar cadernos e arquivos do espaço de trabalho no Amazon S3.
10. Escolha Save Changes (Salvar alterações).

As etapas a seguir demonstram como atualizar um EMR Studio e configurar a criptografia de armazenamento do espaço de trabalho.

1. Abra o EMR console da Amazon em https://console.aws.amazon.com/elasticmapreduce/.
2. Escolha um EMR Studio existente na lista e escolha Editar.
3. Escolha Criptografar arquivos do Workspace com sua própria chave. AWS KMS
4. Insira ou escolha um AWS KMS para usar para criptografar cadernos e arquivos do espaço de trabalho no Amazon S3.
5. Escolha Save Changes (Salvar alterações).

Defina grupos de segurança para controlar o tráfego de rede do EMR Studio

Sobre os grupos de segurança do EMR Studio

O Amazon EMR Studio usa dois grupos de segurança para controlar o tráfego de rede entre os espaços de trabalho no Studio e um EMR cluster conectado da Amazon executado na AmazonEC2:

- Um grupo de segurança do mecanismo que usa a porta 18888 para se comunicar com um EMR cluster conectado da Amazon em execução na Amazon. EC2

- Um grupo de segurança do Workspace associado aos Workspaces em um Studio. Esse grupo de segurança inclui uma HTTPS regra de saída para permitir que o espaço de trabalho direcione o tráfego para a Internet e deve permitir o tráfego de saída para a Internet na porta 443 para permitir a vinculação de repositórios Git a um espaço de trabalho.

EMRO Studio usa esses grupos de segurança além de quaisquer grupos de segurança associados a um EMR cluster anexado a um espaço de trabalho.

Você deve criar esses grupos de segurança ao usar o AWS CLI para criar um Studio.

> **Note**
>
> Você pode personalizar os grupos de segurança do EMR Studio com regras adaptadas ao seu ambiente, mas deve incluir as regras anotadas nesta página. O grupo de segurança do Workspace não pode permitir tráfego de entrada, e o grupo de segurança do mecanismo deve permitir o tráfego de entrada do grupo de segurança do Workspace.

Use os grupos de segurança padrão do EMR Studio

Ao usar o EMR console da Amazon, você pode escolher os seguintes grupos de segurança padrão. Os grupos de segurança padrão são criados pelo EMR Studio em seu nome e incluem as regras mínimas de entrada e saída exigidas para espaços de trabalho em um estúdio. EMR

- `DefaultEngineSecurityGroup`
- `DefaultWorkspaceSecurityGroupGit` ou `DefaultWorkspaceSecurityGroupWithoutGit`

Pré-requisitos

Para criar os grupos de segurança para o EMR Studio, você precisa de uma Amazon Virtual Private Cloud (VPC) para o Studio. Você escolhe isso VPC ao criar os grupos de segurança. Isso deve ser o mesmo VPC que você especifica ao criar o Studio. Se você planeja usar o Amazon Amazon EMR on EKS with EMR Studio, escolha o VPC para seus nós de trabalho de EKS cluster da Amazon.

Instruções

Siga as instruções em Criação de um grupo de segurança no Guia EC2 do usuário da Amazon para instâncias Linux para criar um grupo de segurança de mecanismos e um grupo de segurança de

espaço de trabalho em seuVPC. Os grupos de segurança devem incluir as regras resumidas nas tabelas a seguir.

Ao criar grupos de segurança para o EMR Studio, observe o IDs para ambos. Você especifica cada grupo de segurança usando o ID ao criar um Studio.

Grupo de segurança do mecanismo

EMRO Studio usa a porta 18888 para se comunicar com um cluster conectado.

Regras de entrada

Tipo	Protocolo	Port (Porta)	Destination (Destino)	Descrição
TCP	TCP	18888	Seu grupo de segurança do EMR Studio Workspace.	Permita o tráfego de qualquer recurso no grupo de segurança do Workspace para EMR Studio.

Grupo de segurança do Workspace

Esse grupo de segurança está associado aos espaços de trabalho em um EMR estúdio.

Regras de saída

Type	Protocolo	Port (Porta)	Destination (Destino)	Descrição
TCP	TCP	18888	Seu grupo de segurança do EMR Studio Engine.	Permita o tráfego para qualquer recurso no grupo de segurança do Engine para EMR Studio.
HTTPS	TCP	443	0.0.0.0/0	Permite que o tráfego da Internet vincule repositór ios Git hospedado

Type	Protocolo	Port (Porta)	Destination (Destino)	Descrição
				s publicamente aos Workspaces.

Crie AWS CloudFormation modelos para o Amazon EMR Studio

Sobre os modelos de cluster do EMR Studio

Você pode criar AWS CloudFormation modelos para ajudar os usuários do EMR Studio a lançar novos EMR clusters da Amazon em um espaço de trabalho. CloudFormation modelos são arquivos de texto formatados em JSON ouYAML. Em um modelo, você descreve uma pilha de AWS recursos e explica CloudFormation como provisionar esses recursos para você. Para o EMR Studio, você pode criar um ou mais modelos que descrevam um EMR cluster da Amazon.

Você organiza seus modelos em AWS Service Catalog. AWS Service Catalog permite criar e gerenciar serviços de TI comumente implantados, chamados de produtos em AWS. Você coleta seus modelos como produtos em um portfólio que você compartilha com os usuários do EMR Studio. Após criar modelos de cluster, os usuários do Studio poderão iniciar um novo cluster para um Workspace com um de seus modelos. Os usuários devem ter permissão para criar novos clusters usando os modelos. Você pode definir permissões de usuário nas políticas de permissões do EMR Studio.

Para saber mais sobre CloudFormation modelos, consulte Modelos no Guia do AWS CloudFormation usuário. Para obter mais informações sobre AWS Service Catalog, consulte O que é AWS Service Catalog.

O vídeo a seguir demonstra como configurar modelos de cluster no AWS Service Catalog EMR Studio. Você também pode saber mais na postagem do blog Crie um ambiente de autoatendimento para cada linha de negócios usando a Amazon EMR e o Service Catalog.

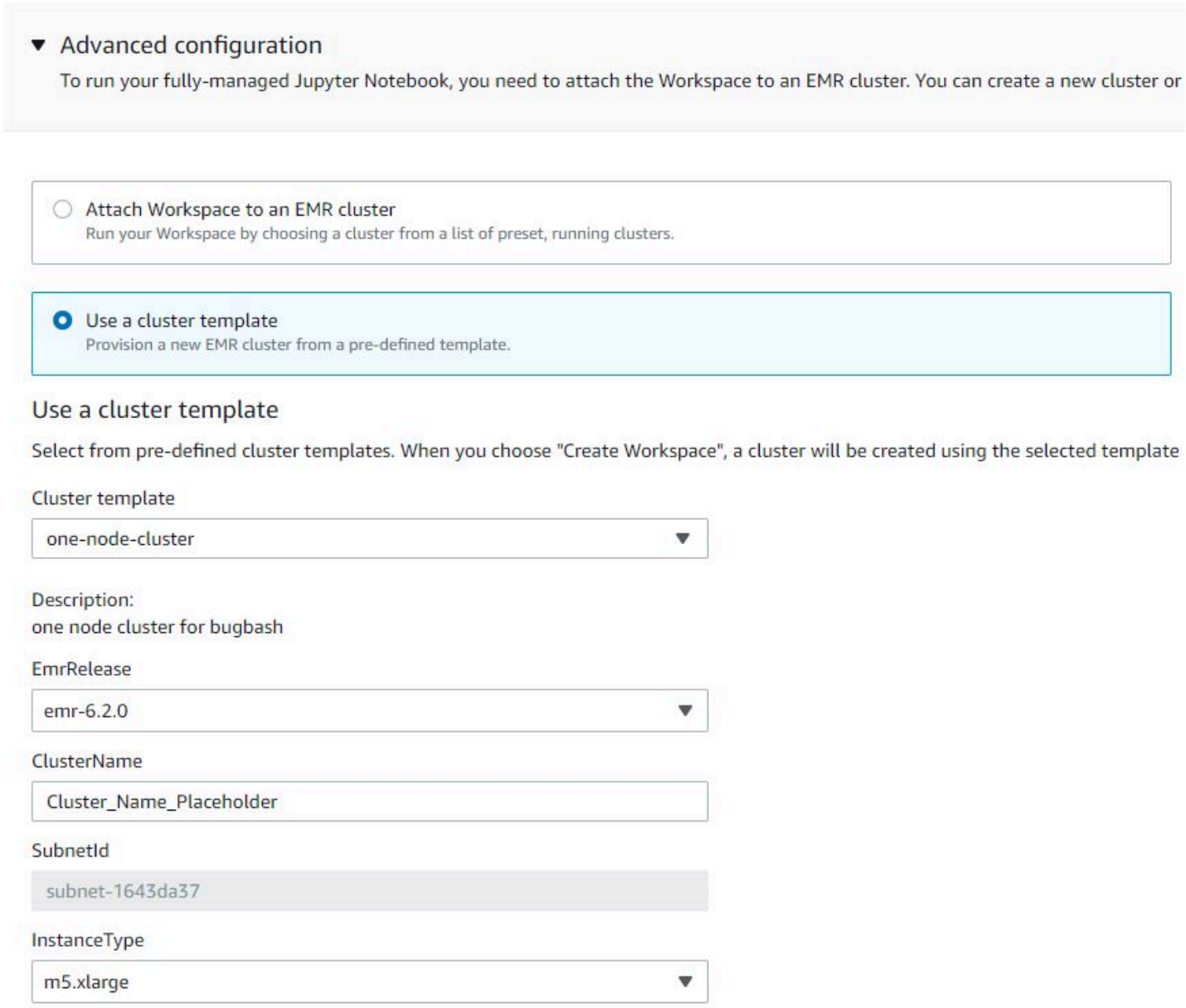
Parâmetros de modelo opcionais

Você pode incluir opções adicionais na seção `Parameters` do seu modelo. Os parâmetros permitem que os usuários do Studio insiram ou selecionem valores personalizados para um cluster. Por exemplo, você pode adicionar um parâmetro que permita aos usuários selecionar uma EMR versão específica da Amazon. Para obter mais informações, consulte Parâmetros no Guia do usuário do AWS CloudFormation .

O exemplo da seção `Parameters` a seguir define parâmetros de entrada adicionais, como o `ClusterName`, a versão de `EmrRelease` e o `ClusterInstanceType`.

```
Parameters:
  ClusterName:
    Type: "String"
    Default: "Cluster_Name_Placeholder"
  EmrRelease:
    Type: "String"
    Default: "emr-6.2.0"
    AllowedValues:
    - "emr-6.2.0"
    - "emr-5.32.0"
  ClusterInstanceType:
    Type: "String"
    Default: "m5.xlarge"
    AllowedValues:
    - "m5.xlarge"
    - "m5.2xlarge"
```

Ao adicionar parâmetros, os usuários do Studio visualizam opções adicionais de formulário após selecionar um modelo de cluster. A imagem a seguir mostra opções adicionais de formulário para a EmrReleaseversão ClusterName, InstanceTypee.



Pré-requisitos

Antes de criar um modelo de cluster, verifique se você tem IAM permissões para acessar a visualização do console do administrador do Service Catalog. Você também precisa das IAM permissões necessárias para realizar tarefas administrativas do Service Catalog. Para obter mais informações, consulte Grant permissions to Service Catalog administrators.

Instruções

Para criar modelos EMR de cluster usando o Service Catalog

1. Crie um ou mais CloudFormation modelos. O local em que você armazenará os modelos fica a seu critério. Como os modelos são arquivos de texto formatados, você pode fazer upload deles no Amazon S3 ou mantê-los em seu sistema de arquivos local. Para saber mais sobre CloudFormation modelos, consulte Modelos no Guia do AWS CloudFormation usuário.

   Use as regras apresentadas a seguir para nomear os modelos ou comparar os nomes em relação ao padrão `[a-zA-Z0-9][a-zA-Z0-9._-]*`.

   - Os nomes dos modelos devem começar com uma letra ou com um número.
   - Os nomes dos modelos podem consistir somente em letras, números, pontos (.), sublinhados (_) e hifens (-).

   Cada modelo de cluster criado deve incluir as seguintes opções:

   Parâmetros de entrada

   - ClusterName — Um nome para o cluster para ajudar os usuários a identificá-lo após o provisionamento.

   Saída

   - `ClusterId`— O ID do cluster recém-provisionadoEMR.

   Veja a seguir um exemplo AWS CloudFormation de modelo em YAML formato para um cluster com dois nós. O modelo de exemplo inclui as opções de modelo obrigatórias e define parâmetros de entrada adicionais para `EmrRelease` e `ClusterInstanceType`.

```
awsTemplateFormatVersion: 2010-09-09

Parameters:
  ClusterName:
    Type: "String"
    Default: "Example_Two_Node_Cluster"
  EmrRelease:
    Type: "String"
```

```
    Default: "emr-6.2.0"
    AllowedValues:
    - "emr-6.2.0"
    - "emr-5.32.0"
  ClusterInstanceType:
    Type: "String"
    Default: "m5.xlarge"
    AllowedValues:
    - "m5.xlarge"
    - "m5.2xlarge"

Resources:
  EmrCluster:
    Type: AWS::EMR::Cluster
    Properties:
      Applications:
      - Name: Spark
      - Name: Livy
      - Name: JupyterEnterpriseGateway
      - Name: Hive
      EbsRootVolumeSize: '10'
      Name: !Ref ClusterName
      JobFlowRole: EMR_EC2_DefaultRole
      ServiceRole: EMR_DefaultRole_V2
      ReleaseLabel: !Ref EmrRelease
      VisibleToAllUsers: true
      LogUri:
        Fn::Sub: 's3://aws-logs-${AWS::AccountId}-${AWS::Region}/elasticmapreduce/'
      Instances:
        TerminationProtected: false
        Ec2SubnetId: 'subnet-ab12345c'
        MasterInstanceGroup:
          InstanceCount: 1
          InstanceType: !Ref ClusterInstanceType
        CoreInstanceGroup:
          InstanceCount: 1
          InstanceType: !Ref ClusterInstanceType
          Market: ON_DEMAND
          Name: Core

Outputs:
  ClusterId:
    Value:
      Ref: EmrCluster
```

```
    Description: The ID of the  EMR cluster
```

2. Crie um portfólio para seus modelos de cluster na mesma AWS conta do seu Studio.

   a. Abra o AWS Service Catalog console em https://console.aws.amazon.com/servicecatalog/.

   b. Escolha Portfólios no menu de navegação à esquerda.

   c. Insira as informações solicitadas na página Criar portfólio.

   d. Escolha Criar. AWS Service Catalog cria o portfólio e exibe os detalhes do portfólio.

3. Use as etapas a seguir para adicionar seus modelos de cluster como produtos do AWS Service Catalog .

   a. Navegue até a página Produtos em Administração no console de gerenciamento do AWS Service Catalog .

   b. Escolha Fazer upload de novo produto.

   c. Insira um Nome do produto e um Proprietário.

   d. Especifique seu arquivo de modelo em Detalhes da versão.

   e. Escolha Analisar para analisar as configurações do produto e, em seguida, selecione Criar produto.

4. Conclua as etapas a seguir para adicionar os produtos ao seu portfólio.

   a. Navegue até a página Produtos no console de gerenciamento do AWS Service Catalog .

   b. Escolha seu produto, selecione Ações e, em seguida, clique em Adicionar produto ao portfólio.

   c. Escolha seu portfólio e, em seguida, escolha Adicionar produto ao portfólio.

5. Crie uma restrição de inicialização para seus produtos. Uma restrição de lançamento é uma IAM função que especifica as permissões do usuário para lançar um produto. Você pode personalizar suas restrições de lançamento, mas deve permitir permissões de uso, CloudFormation Amazon e. EMR AWS Service Catalog Para obter mais informações e instruções, consulte Service Catalog launch constraints.

6. Aplique a restrição de inicialização a cada produto do seu portfólio. Você deve aplicar a restrição de inicialização a cada produto individualmente.

   a. Selecione seu portfólio na página Portfólios no console de gerenciamento do AWS Service Catalog .

   b. Escolha a guia Constraints (Restrições) e Create constraint (Criar restrição).

c. Escolha seu produto e selecione Inicialização em Tipo de restrição. Escolha Continuar.

d. Selecione seu perfil de restrição de inicialização na seção Restrição de inicialização e, em seguida, escolha Criar.

7. Conceda acesso ao seu portfólio.

a. Selecione seu portfólio na página Portfólios no console de gerenciamento do AWS Service Catalog .

b. Expanda a guia Grupos, perfis e usuários e escolha Adicionar grupos, perfis e usuários.

c. Procure sua IAM função no EMR Studio na guia Funções, selecione sua função e escolha Adicionar acesso.

Se você usar...	Conceda acesso a...
IAMautenticação	Seus usuários nativos
IAMfederação	Seu IAM papel na federação
IAMFederação de Centros de Identidade	Sua função de usuário do EMR Studio

Estabelecimento de acesso e de permissões para repositórios baseados em Git

EMRO Studio oferece suporte aos seguintes serviços baseados em Git:

- AWS CodeCommit
- GitHub
- Bitbucket
- GitLab

Para permitir que os usuários do EMR Studio associem um repositório Git a um espaço de trabalho, configure os seguintes requisitos de acesso e permissões. Você também pode configurar repositórios baseados em Git hospedados em uma rede privada ao seguir as instruções em Configurar um repositório Git hospedado de forma privada para o Studio EMR.

Cluster com acesso à Internet

Tanto os EMR clusters da Amazon executados EMR na Amazon EC2 quanto os da Amazon em EKS clusters conectados ao Studio Workspaces devem estar em uma sub-rede privada que usa um gateway de tradução de endereços de rede (NAT), ou devem ser capazes de acessar a Internet por meio de um gateway privado virtual. Para obter mais informações, consulte VPCOpções da Amazon.

Os grupos de segurança que você usa com o EMR Studio também devem incluir uma regra de saída que permita ao Workspaces rotear o tráfego para a Internet a partir de um cluster conectadoEMR. Para obter mais informações, consulte Defina grupos de segurança para controlar o tráfego de rede do EMR Studio.

> **Important**
>
> Se a interface de rede estiver em uma sub-rede pública, ela não conseguirá se comunicar com a Internet por meio de um gateway de internet (IGW).

Permissões para AWS Secrets Manager

Para permitir que os usuários do EMR Studio acessem os repositórios Git com segredos armazenados AWS Secrets Manager, adicione uma política de permissões à função de serviço do EMR Studio que permita a operação. `secretsmanager:GetSecretValue`

Para obter informações sobre como vincular repositórios baseados em Git a Workspaces, consulte Vincular repositórios baseados em Git a um espaço de trabalho do Studio EMR.

Configurar um repositório Git hospedado de forma privada para o Studio EMR

Use as instruções a seguir para configurar repositórios hospedados de forma privada para o Amazon EMR Studio. Forneça um arquivo de configuração com informações sobre seus DNS servidores e Git. EMRO Studio usa essas informações para configurar espaços de trabalho que podem rotear o tráfego para seus repositórios autogerenciados.

> **Note**
>
> Se você configurar`DnsServerIpV4`, o EMR Studio usará seu DNS servidor para resolver seu EMR endpoint `GitServerDnsName` e o da Amazon, como`elasticmapreduce.us-`

`east-1.amazonaws.com`. Para configurar um endpoint para a AmazonEMR, conecte-se ao seu endpoint por meio do VPC que você está usando com seu Studio. Isso garante que o EMR endpoint da Amazon seja resolvido para um IP privado. Para obter mais informações, consulte Conecte-se à Amazon EMR usando um VPC endpoint de interface.

Pré-requisitos

Antes de configurar um repositório Git hospedado de forma privada EMR para o Studio, você precisa de um local de armazenamento do Amazon S3 EMR onde o Studio possa fazer backup dos espaços de trabalho e dos arquivos do notebook no Studio. Use o mesmo bucket do S3 especificado ao criar um Studio.

Para configurar um ou mais repositórios Git hospedados de forma privada para o Studio EMR

1. Crie um arquivo de configuração usando o modelo apresentado a seguir. Inclua os seguintes valores para cada servidor Git que deseja especificar em sua configuração:

   - **`DnsServerIpV4`**- O IPv4 endereço do seu DNS servidor. Se você fornecer valores para `DnsServerIpV4` e`GitServerIpV4List`, o valor para terá `DnsServerIpV4` precedência e o EMR Studio o usará `DnsServerIpV4` para resolver seu`GitServerDnsName`.

     > **Note**
     >
     > Para usar repositórios Git hospedados de forma privada, DNS seu servidor deve permitir o acesso de entrada do Studio. EMR Recomendamos que você proteja seu DNS servidor contra outros acessos não autorizados.

   - **`GitServerDnsName`**- O DNS nome do seu servidor Git. Por exemplo, `"git.example.com"`.
   - **`GitServerIpV4List`**- Uma lista de IPv4 endereços que pertencem aos seus servidores Git.

```
[
    {
        "Type": "PrivatelyHostedGitConfig",
        "Value": [
            {
                "DnsServerIpV4": "<10.24.34.xxx>",
                "GitServerDnsName": "<enterprise.git.com>",
```

```
                "GitServerIpV4List": [
                    "<xxx.xxx.xxx.xxx>",
                    "<xxx.xxx.xxx.xxx>"
                ]
            },
            {
                "DnsServerIpV4": "<10.24.34.xxx>",
                "GitServerDnsName": "<git.example.com>",
                "GitServerIpV4List": [
                    "<xxx.xxx.xxx.xxx>",
                    "<xxx.xxx.xxx.xxx>"
                ]
            }
        ]
    }
]
```

2. Salve seu arquivo de configuração como `configuration.json`.
3. Faça o upload do arquivo de configuração no local de armazenamento do Amazon S3 em uma pasta chamada `life-cycle-configuration`. Por exemplo, se o local padrão do S3 for `s3://`*`DOC-EXAMPLE-BUCKET`*`/studios`, seu arquivo de configuração estará em `s3://`*`DOC-EXAMPLE-BUCKET`*`/studios/`**`life-cycle-configuration/configuration.json`**.

> **Important**
>
> Recomendamos que você restrinja o acesso à sua `life-cycle-configuration` pasta aos administradores do Studio e à sua função de serviço do EMR Studio, e que você se proteja `configuration.json` contra acesso não autorizado. Para obter instruções, consulte Controlar o acesso a um bucket com políticas de usuário ou Práticas recomendadas de segurança para o Amazon S3.

Para obter instruções sobre como fazer o upload, consulte Criar uma pasta e Fazer upload de objetos no Guia do usuário do Amazon Simple Storage Service. Para aplicar sua configuração a um Workspace, feche e reinicie o Workspace após fazer o upload do arquivo de configuração para o Amazon S3.

Otimize as tarefas do Spark no Studio EMR

Ao executar um trabalho do Spark usando o EMR Studio, há algumas etapas que você pode seguir para ajudar a garantir que você esteja otimizando seus recursos de EMR cluster da Amazon.

Prolongamento da sessão do Livy

Se você usa o Apache Livy junto com o Spark em seu EMR cluster Amazon, recomendamos que você aumente o tempo limite da sessão do Livy fazendo o seguinte:

- Ao criar um EMR cluster da Amazon, defina essa classificação de configuração no campo Inserir configuração.

```
[
    {
        "Classification": "livy-conf",
        "Properties": {
          "livy.server.session.timeout": "8h"
        }
    }
]
```

- Para um cluster já em execução, conecte-se ao seu EMR cluster usando `ssh` e defina a classificação da `livy-conf` configuração em. `/etc/livy/conf/livy.conf`

```
[
    {
        "Classification": "livy-conf",
        "Properties": {
          "livy.server.session.timeout": "8h"
        }
    }
]
```

  Pode ser necessário reiniciar o Livy após alterar a configuração.

- Se você não deseja que sua sessão do Livy expire, defina a propriedade `livy.server.session.timeout-check` como `false` em `/etc/livy/conf/livy.conf`.

Execução do Spark no modo de cluster

No modo de cluster, o driver do Spark é executado em um nó central em vez de no nó primário, melhorando a utilização de recursos no nó primário.

Para executar seu aplicativo Spark no modo cluster em vez do modo cliente padrão, escolha o modo Cluster ao definir o modo Deploy enquanto configura sua etapa do Spark em seu novo cluster Amazon. EMR Para obter mais informações, consulte Cluster mode overview na documentação do Apache Spark.

Aumento da memória do driver do Spark

Para aumentar a memória do driver do Spark, configure sua sessão do Spark usando o comando `%%configure` mágico em seu EMR notebook, como no exemplo a seguir.

```
%%configure -f
{"driverMemory": "6000M"}
```

Use um Amazon EMR Studio

Esta seção contém tópicos que ajudam você a configurar e interagir com um Amazon EMR Studio.

O vídeo a seguir aborda informações práticas, como criar um novo espaço de trabalho e como lançar um novo EMR cluster da Amazon com um modelo de cluster. Além disso, o vídeo executa um caderno de exemplo.

Esta seção inclui os seguintes tópicos para ajudá-lo a trabalhar em um EMR estúdio:

- Compreensão das noções básicas do Workspace
- Configuração da colaboração no Workspace
- Execute um EMR Studio Workspace com uma função de tempo de execução
- Execução de cadernos do Workspace de forma programática
- Navegue pelos dados com o SQL Explorer
- Anexar uma computação a um espaço de trabalho do EMR Studio
- Vincular repositórios baseados em Git a um espaço de trabalho do Studio EMR
- Use o SQL editor Amazon Athena no Studio EMR

- CodeWhisperer Integração da Amazon com o EMR Studio Workspaces
- Depure aplicativos e trabalhos com EMR o Studio
- Instale kernels e bibliotecas em um Studio Workspace EMR
- Aprimoramento de kernels com comandos magic
- Use cadernos em várias linguagens com kernels do Spark

Compreensão das noções básicas do Workspace

Ao usar um EMR Studio, você pode criar e configurar diferentes espaços de trabalho para organizar e executar notebooks. Esta seção aborda como criar e trabalhar com Workspaces. Para obter uma visão geral conceitual, consulte Workspaces na página Como o Amazon EMR Studio funciona.

Esta seção aborda os seguintes tópicos para ajudá-lo a usar o EMR Studio Workspaces:

- Crie um espaço de trabalho de EMR estúdio
- Inicialização de um WorkSpace
- Compreensão da interface do usuário do Workspace
- Exploração de exemplos de cadernos
- Salvamento de conteúdo do Workspace
- Exclusão de um Workspace e de arquivos de cadernos
- Compreensão do status do Workspace
- Resolução de problemas de conectividade do Workspace

Crie um espaço de trabalho de EMR estúdio

Você pode criar espaços de trabalho do EMR Studio para executar o código do notebook usando a interface do EMR Studio.

Para criar um espaço de trabalho em um estúdio EMR

1. Faça login no seu EMR Studio.
2. Escolha Criar um Workspace.
3. Insira um Nome do WorkSpace e uma Descrição. Nomear um Workspace ajuda a identificá-lo na página Workspaces.

4. Se desejar trabalhar com outros usuários do Studio neste Workspace em tempo real, habilite a colaboração no Workspace. Você pode configurar colaboradores depois de iniciar o Workspace.
5. Se desejar anexar um cluster a um Workspace, expanda a seção Configuração avançada. Você pode anexar um cluster posteriormente, se preferir. Para obter mais informações, consulte Anexar uma computação a um espaço de trabalho do EMR Studio.

   > **Note**
   >
   > Para provisionar um novo cluster, você precisa receber permissões de acesso por parte do administrador.

   Escolha uma das opções de cluster para o Workspace e anexe o cluster. Para obter mais informações sobre o provisionamento de um cluster ao criar um Workspace, consulte Crie e anexe um novo EMR cluster a um EMR Studio Workspace.
6. Escolha Criar um Workspace no canto inferior direito da página.

Depois de criar um espaço de trabalho, o EMR Studio abrirá a página Espaços de trabalho. Você visualizará um banner verde representando o êxito na parte superior da página e poderá encontrar o Workspace recém-criado na lista.

Por padrão, um Workspace é compartilhado e pode ser visualizado por todos os usuários do Studio. No entanto, somente um usuário pode abrir e trabalhar em um Workspace por vez. Para trabalhar simultaneamente com outros usuários, é possível realizar a Configuração da colaboração no Workspace.

Inicialização de um WorkSpace

Para começar a trabalhar com arquivos de cadernos, inicie um Workspace para acessar o editor de caderno. A página Workspaces em um Studio lista todos os Workspaces aos quais você tem acesso com detalhes, incluindo Nome, Status, Horário de criação e Última modificação.

> **Note**
>
> Se você tinha EMR cadernos no antigo EMR console da Amazon, você pode encontrá-los no console como EMR Studio Workspaces. EMROs usuários de notebooks precisam de permissões adicionais de IAM função para acessar ou criar espaços de trabalho. Se você criou recentemente um notebook no console antigo, talvez seja necessário atualizar

a lista de espaços de trabalho para vê-lo no console. Para obter mais informações sobre a transição, consulte Os Amazon EMR Notebooks estão disponíveis como Amazon EMR Studio Workspaces no console e Console do Amazon EMR.

Iniciar um Workspace para edição e execução de cadernos

1. Na página Workspaces do seu Studio, localize o Workspace. Você pode filtrar a lista por palavra-chave ou por valor de coluna.
2. Escolha o nome do Workspace para iniciá-lo em uma nova guia do navegador. Pode demorar alguns minutos para o Workspace abrir, se ele estiver Ocioso. Como alternativa, selecione a linha para o Workspace e, em seguida, escolha Iniciar o Workspace. É possível escolher entre as seguintes opções de inicialização:

   - Início rápido: inicie rapidamente seu Workspace com as opções padrão. Escolha Início rápido se quiser anexar clusters ao espaço de trabalho em JupyterLab.
   - Início com opções: inicie seu Workspace com opções personalizadas. Você pode optar por iniciar no Jupyter ou JupyterLab anexar seu espaço de trabalho a um EMR cluster e selecionar seus grupos de segurança.

> **Note**
>
> Somente um usuário pode abrir e trabalhar em um Workspace por vez. Se você selecionar um espaço de trabalho que já esteja em uso, o EMR Studio exibirá uma notificação quando você tentar abri-lo. A coluna Usuário na página Workspaces mostra o usuário que está trabalhando no Workspace.

Compreensão da interface do usuário do Workspace

A interface de usuário do EMR Studio Workspace é baseada na JupyterLabinterface com guias indicadas por ícones na barra lateral esquerda. Ao colocar o cursor do mouse sobre um ícone, você visualizará uma descrição que mostra o nome da guia. Escolha as guias na barra lateral à esquerda para acessar os painéis apresentados a seguir.

- Navegador de arquivos: exibe os arquivos e diretórios no Workspace, bem como os arquivos e diretórios de repositórios Git vinculados.

- Kernels e terminais em execução: lista todos os kernels e os terminais em execução no Workspace. Para obter mais informações, consulte Gerenciando kernels e terminais na documentação oficial JupyterLab .
- Git: fornece uma interface gráfica do usuário para a execução de comandos nos repositórios Git anexados ao Workspace. Esse painel é uma JupyterLab extensão chamada jupyterlab-git. Para obter mais informações, consulte jupyterlab-git.
- EMRclusters — Permite anexar um cluster ou desanexar um cluster do Workspace para executar o código do notebook. O painel de configuração do EMR cluster também fornece opções avançadas de configuração para ajudá-lo a criar e anexar um novo cluster ao espaço de trabalho. Para obter mais informações, consulte Crie e anexe um novo EMR cluster a um EMR Studio Workspace.
- Repositório Amazon EMR Git — Ajuda você a vincular o espaço de trabalho a até três repositórios Git. Para obter detalhes e instruções, consulte Vincular repositórios baseados em Git a um espaço de trabalho do Studio EMR.
- Exemplos de cadernos: fornece uma lista de exemplos de cadernos que você pode salvar no Workspace. Você também pode acessar os exemplos ao escolher Exemplos de cadernos na página Inicializador do Workspace.
- Comandos — Oferece uma forma orientada pelo teclado de pesquisar e executar comandos. JupyterLab Para obter mais informações, consulte a página da paleta Command na JupyterLab documentação.
- Ferramentas do caderno: permite selecionar e definir opções, como o tipo de deslizamento da célula e os metadados. A opção Ferramentas do caderno aparece na barra lateral à esquerda depois que você abre um arquivo de caderno.
- Guias abertas: lista os documentos e as atividades abertos na área de trabalho principal para que você possa acessar uma guia aberta. Para obter mais informações, consulte a página do modo Tabulações e documento único na JupyterLab documentação.
- Colaboração: permite habilitar ou desabilitar a colaboração no Workspace e gerenciar colaboradores. Para visualizar o painel Colaboração, você deve ter as permissões necessárias. Para obter mais informações, consulte Definição de propriedade para colaboração no Workspace.

Exploração de exemplos de cadernos

Cada espaço de trabalho do EMR Studio inclui um conjunto de exemplos de cadernos que você pode usar para explorar os recursos do EMR Studio. Para editar ou executar um exemplo de caderno, você pode salvá-lo no Workspace.

Salvar um exemplo de caderno em um Workspace

1. Na barra lateral à esquerda, escolha a guia Exemplos de cadernos para abrir o painel Exemplos de cadernos. Você também pode acessar os exemplos ao escolher Exemplos de cadernos na página Inicializador do Workspace.
2. Escolha um exemplo de caderno para visualizá-lo previamente na área de trabalho principal. O exemplo é somente para leitura.
3. Para salvar o exemplo de caderno no Workspace, escolha Salvar no Workspace. EMRO Studio salva o exemplo em seu diretório inicial. Depois de salvar um exemplo de caderno no Workspace, você poderá renomeá-lo, editá-lo e executá-lo.

Para obter mais informações sobre os exemplos de notebooks, consulte o GitHub repositório de exemplos do EMR Studio Notebook.

Salvamento de conteúdo do Workspace

Quando você trabalha no editor de notebook de um espaço de trabalho, o EMR Studio salva o conteúdo das células do notebook e a saída para você no local do Amazon S3 associado ao Studio. Este processo de backup preserva o trabalho entre as sessões.

Você também pode salvar um caderno pressionando CTRL+S na guia aberta do caderno ou usando uma das opções de salvamento em Arquivo.

Outra maneira de fazer backup dos arquivos de cadernos em um Workspace é associar o Workspace a um repositório baseado em Git e sincronizar suas alterações com o repositório remoto. Isso também permite salvar e compartilhar cadernos com membros da equipe que usam um Workspace ou um Studio diferente. Para obter instruções, consulte Vincular repositórios baseados em Git a um espaço de trabalho do Studio EMR.

Exclusão de um Workspace e de arquivos de cadernos

Ao excluir um arquivo de notebook de um EMR Studio Workspace, você exclui o arquivo do Navegador de arquivos, e o EMR Studio remove sua cópia de backup no Amazon S3. Você não precisa tomar nenhuma medida adicional para evitar cobranças de armazenamento ao excluir um arquivo de um Workspace.

Quando você exclui um Workspace inteiro, seus arquivos e suas pastas de cadernos permanecerão no local de armazenamento do Amazon S3. Os arquivos continuam a acumular cobranças de

armazenamento. Para evitar cobranças de armazenamento, remova todos os arquivos e as pastas de backup associados ao Workspace excluído do Amazon S3.

Para excluir um arquivo de notebook de um EMR Studio Workspace

1. Selecione o painel Navegador de arquivos na barra lateral à esquerda do Workspace.
2. Selecione o arquivo ou a pasta que deseja excluir. Clique com o botão direito do mouse na sua seleção e escolha Excluir. O arquivo desaparecerá da lista. EMRO Studio remove o arquivo ou a pasta do Amazon S3 para você.

From the Workspace UI

Excluir um espaço de trabalho e seus arquivos de backup associados do Studio EMR

1. Faça login no seu EMR Studio com seu acesso ao Studio URL e escolha Workspaces no painel de navegação à esquerda.
2. Localize seu Workspace na lista e, em seguida, marque a caixa de seleção ao lado do nome. É possível selecionar vários Workspaces a serem excluídos ao mesmo tempo.
3. Escolha Excluir no canto superior direito da lista Workspaces e confirme que deseja excluir os Workspaces selecionados. Escolha Delete para confirmar.
4. Se você desejar remover os arquivos de cadernos associados ao Workspace excluído do Amazon S3, siga as instruções para Excluir objetos no Guia do usuário do console do Amazon Simple Storage Service. Se não foi você quem criou o Studio, consulte o administrador do Studio para determinar o local de backup do Amazon S3 para o Workspace excluído.

From the Workspaces list

Exclusão de um Workspace e dos arquivos de backup associados da lista Workspaces

1. Navegue até a lista Workspaces no console.
2. Selecione o Workspace que deseja excluir da lista e, em seguida, escolha Ações.
3. Escolha Excluir.
4. Se você desejar remover os arquivos de cadernos associados ao Workspace excluído do Amazon S3, siga as instruções para Excluir objetos no Guia do usuário do console do Amazon Simple Storage Service. Se não foi você quem criou o Studio, consulte o

administrador do Studio para determinar o local de backup do Amazon S3 para o Workspace excluído.

Compreensão do status do Workspace

Depois de criar um EMR Studio Workspace, ele aparece como uma linha na lista Workspaces em seu Studio com seu nome, status, hora de criação e data e hora da última modificação. A tabela a seguir descreve os status do Workspace.

Status	Descrição
Starting	O Workspace está sendo preparado, mas ainda não está pronto para uso. Não é possível abrir um Workspace quando o status for Iniciando.
Ready	Você pode abrir o espaço de trabalho para usar o editor do notebook, mas deve anexar o espaço de trabalho a um EMR cluster antes de poder executar o código do notebook.
Anexando	O Workspace está sendo anexado a um cluster.
Attached	O espaço de trabalho está conectado a um EMR cluster e pronto para você escrever e executar o código do notebook. Se o status de um Workspace não for Anexado, você deverá anexá-lo a um cluster antes de executar o código do caderno.
Ocioso	O Workspace foi interrompido. Para reativar um Workspace ocioso, selecione-o na lista Workspaces. O status é alterado de Ocioso para Iniciando e, em seguida, para Pronto quando você seleciona o Workspace.

Status	Descrição
Stopping	O Workspace está sendo encerrado e será definido como Ocioso. Quando você interromp e um Workspace, ele encerra todos os kernels de cadernos correspondentes. EMRO Studio interrompe notebooks que ficaram inativos por muito tempo.
Deleting	Quando você exclui um espaço de trabalho, o EMR Studio o marca para exclusão e inicia o processo de exclusão. Após a conclusão do processo de exclusão, o Workspace desaparec erá da lista. Quando você exclui um Workspace , os arquivos de cadernos permanecerão no local de armazenamento do Amazon S3.

Resolução de problemas de conectividade do Workspace

Para resolver problemas de conectividade do Workspace, você pode interromper e reiniciar um Workspace. Quando você reinicia um espaço de trabalho, o EMR Studio inicia o espaço de trabalho em uma zona de disponibilidade diferente ou em uma sub-rede diferente associada ao seu estúdio.

Para parar e reiniciar um EMR Studio Workspace

1. Feche o Workspace no seu navegador.
2. Navegue até a lista Workspace no console.
3. Selecione seu Workspace na lista e escolha Ações.
4. Escolha Interromper e aguarde até que o status do Workspace seja alterado de Interrompendo para Ocioso.
5. Escolha Ações novamente e, em seguida, selecione Iniciar para reiniciar o Workspace.
6. Aguarde até que o status do Workspace seja alterado de Iniciando para Pronto e, em seguida, escolha o nome do Workspace para abri-lo novamente em uma nova guia do navegador.

Configuração da colaboração no Workspace

A colaboração no Workspace permite escrever e executar códigos de cadernos simultaneamente com outros membros da sua equipe. Ao trabalhar no mesmo arquivo de caderno, você visualizará as alterações à medida que seus colaboradores as fizerem. É possível habilitar a colaboração ao criar um Workspace, ou habilitar e desabilitar a colaboração em um Workspace existente.

> **Note**
>
> EMRA colaboração do Studio Workspace não é compatível com aplicativos interativos EMR sem servidor ou se a propagação de identidade confiável estiver habilitada.

Pré-requisitos

Antes de configurar a colaboração para um Workspace, certifique-se de concluir as seguintes tarefas:

- Certifique-se de que seu administrador do EMR Studio tenha concedido as permissões necessárias. Por exemplo, a instrução apresentada a seguir permite que um usuário configure a colaboração para qualquer Workspace com a chave de etiqueta `creatorUserId` cujo valor corresponde ao ID do usuário (indicado pela variável de política `aws:userId`).

```
{
    "Sid": "UserRolePermissionsForCollaboration",
    "Action": [
        "elasticmapreduce:UpdateEditor",
        "elasticmapreduce:PutWorkspaceAccess",
        "elasticmapreduce:DeleteWorkspaceAccess",
        "elasticmapreduce:ListWorkspaceAccessIdentities"
    ],
    "Resource": "*",
    "Effect": "Allow",
    "Condition": {
        "StringEquals": {
            "elasticmapreduce:ResourceTag/creatorUserId": "${aws:userid}"
        }
    }
}
```

- Certifique-se de que a função de serviço associada ao seu EMR Studio tenha as permissões necessárias para habilitar e configurar a colaboração no Workspace, como no exemplo de declaração a seguir.

```
{
    "Sid": "AllowWorkspaceCollaboration",
    "Effect": "Allow",
    "Action": [
        "iam:GetUser",
        "iam:GetRole",
        "iam:ListUsers",
        "iam:ListRoles",
        "sso:GetManagedApplicationInstance",
        "sso-directory:SearchUsers"
    ],
    "Resource": "*"
}
```

Para obter mais informações, consulte Criar uma função de serviço do EMR Studio.

Habilitar a colaboração no Workspace e adicionar colaboradores

1. No seu Workspace, escolha o ícone Colaboração na tela do Inicializador ou na parte inferior do painel à esquerda.

> **Note**
>
> Não será possível visualizar o painel Colaboração, a menos que o administrador do Studio tenha concedido a você permissão para configurar a colaboração para o Workspace. Para obter mais informações, consulte Definição de propriedade para colaboração no Workspace.

2. Certifique-se de que o botão de alternância Permitir a colaboração no Workspace esteja na posição ativada. Ao habilitar a colaboração, somente você e os colaboradores adicionados poderão visualizar o Workspace na lista da página Workspaces do Studio.
3. Insira um Nome do colaborador. Seu Workspace pode ter, no máximo, cinco colaboradores, incluindo você. Um colaborador pode ser qualquer usuário com acesso ao seu EMR Studio. Se você não inserir um colaborador, o Workspace será considerado um Workspace privado acessível somente por você.

A tabela a seguir especifica os valores aplicáveis a serem inseridos para colaboradores com base no tipo de identidade do proprietário.

> **Note**
>
> Um proprietário pode convidar somente colaboradores com o mesmo tipo de identidade. Por exemplo, um usuário só pode adicionar outros usuários, e um usuário do IAM Identity Center só pode adicionar outros usuários do IAM Identity Center.

Modo de autenticação	Valor a ser inserido para o Nome do colaborador
IAMautenticação	Um nome de usuário. Este é o nome que um usuário visualiza quando faz login no AWS Management Console.
IAMfederação	O nome de uma IAM função e um nome de sessão opcional. Para adicionar todos os usuários federados que assumem a mesma IAM função, especifique o nome de uma IAM função para federação. Para adicionar um único usuário como colaborador, especifique um perfil e um nome de sessão. Por exemplo, `MyRoleName:MySessionName` .
SSO	Um nome de usuário do IAM Identity Center, como `user@example.com`.

4. Escolha Adicionar. Agora, o colaborador pode ver o espaço de trabalho na página do EMR Studio Workspaces e iniciar o espaço de trabalho para usá-lo em tempo real com você.

> **Note**
>
> Se você desabilitar a colaboração do Workspace, o Workspace retornará ao estado compartilhado e poderá ser visualizado por todos os usuários do Studio. No estado compartilhado, somente um usuário do Studio poderá abrir e trabalhar no Workspace por vez.

Execute um EMR Studio Workspace com uma função de tempo de execução

> **Note**
>
> A funcionalidade de função de tempo de execução descrita nesta página se aplica somente à Amazon em EMR execução na Amazon EC2 e não se refere à funcionalidade de função de tempo de execução em aplicativos interativos EMR sem servidor. Para saber mais sobre como usar funções de tempo de execução no EMR Serverless, consulte Job runtime roles no Amazon EMR Serverless User Guide.

Uma função de tempo de execução é uma função AWS Identity and Access Management (IAM) que você pode especificar ao enviar um trabalho ou uma consulta para um EMR cluster da Amazon. O trabalho ou consulta que você envia ao seu EMR cluster usa a função de tempo de execução para acessar AWS recursos, como objetos no Amazon S3.

Ao anexar um EMR Studio Workspace a um EMR cluster que usa o Amazon EMR 6.11 ou superior, você pode selecionar uma função de tempo de execução para o trabalho ou consulta que você envia para uso quando AWS acessa recursos. No entanto, se o EMR cluster não oferecer suporte a funções de tempo de execução, o EMR cluster não assumirá a função ao acessar AWS recursos.

Antes de usar uma função de tempo de execução com um Amazon EMR Studio Workspace, um administrador deve configurar as permissões do usuário para que o usuário do Studio possa invocar a função `elasticmapreduce:GetClusterSessionCredentials` API de tempo de execução. Em seguida, inicie um novo cluster com uma função de tempo de execução que você pode usar com seu Amazon EMR Studio Workspace.

Nesta página

- Inicialização de um novo cluster com um perfil de runtime
- Use o EMR cluster com uma função de tempo de execução no Workspaces
- Considerações

Configuração de permissões de usuários para o perfil de runtime

Configure as permissões do usuário para que o usuário do Studio possa chamar a `elasticmapreduce:GetClusterSessionCredentials` API função de tempo de execução que o usuário deseja usar. Você também deve configurar as the section called “Permissões de usuário do Studio (EC2,EKS)” antes que o usuário possa começar a usar o Studio.

> **Warning**
>
> Para conceder essa permissão, crie uma condição com base na chave de `elasticmapreduce:ExecutionRoleArn` contexto ao conceder a um chamador acesso para chamar o. `GetClusterSessionCredentials` APIs Os exemplos a seguir demonstram como fazer isso.

```
{
      "Sid": "AllowSpecificExecRoleArn",
      "Effect": "Allow",
      "Action": [
          "elasticmapreduce:GetClusterSessionCredentials"
      ],
      "Resource": "*",
      "Condition": {
          "StringEquals": {
              "elasticmapreduce:ExecutionRoleArn": [
                  "arn:aws:iam::111122223333:role/test-emr-demo1",
                  "arn:aws:iam::111122223333:role/test-emr-demo2"
              ]
          }
      }
  }
```

O exemplo a seguir demonstra como permitir que um IAM diretor use uma IAM função chamada função `test-emr-demo3` de tempo de execução. Além disso, o titular da apólice só poderá acessar os EMR clusters da Amazon com o ID do cluster`j-123456789`.

```
{
    "Sid":"AllowSpecificExecRoleArn",
    "Effect":"Allow",
    "Action":[
        "elasticmapreduce:GetClusterSessionCredentials"
    ],
    "Resource": [
          "arn:aws:elasticmapreduce:<region>:111122223333:cluster/j-123456789"
     ],
    "Condition":{
        "StringEquals":{
            "elasticmapreduce:ExecutionRoleArn":[
                "arn:aws:iam::111122223333:role/test-emr-demo3"
            ]
        }
    }
}
```

O exemplo a seguir permite que um IAM principal use qualquer IAM função com um nome começando com a string `test-emr-demo4` como a função de tempo de execução. Além disso, o titular da apólice só poderá acessar EMR clusters da Amazon marcados com o par de valores-chave. `tagKey: tagValue`

```
{
    "Sid":"AllowSpecificExecRoleArn",
    "Effect":"Allow",
    "Action":[
        "elasticmapreduce:GetClusterSessionCredentials"
    ],
    "Resource": "*",
    "Condition":{
        "StringEquals":{
             "elasticmapreduce:ResourceTag/tagKey": "tagValue"
        },
        "StringLike":{
            "elasticmapreduce:ExecutionRoleArn":[
                "arn:aws:iam::111122223333:role/test-emr-demo4*"
            ]
        }
    }
}
```

Inicialização de um novo cluster com um perfil de runtime

Agora que você tem as permissões necessárias, inicie um novo cluster com uma função de tempo de execução que você pode usar com seu Amazon EMR Studio Workspace.

Se você já iniciou um novo cluster com um perfil de runtime, poderá pular para a seção the section called “Uso do cluster com seu Workspace”.

1. Primeiro, conclua os pré-requisitos apresentados na seção Funções de tempo de execução para Amazon EMR Steps.
2. Em seguida, inicie um cluster com as seguintes configurações para usar funções de tempo de execução com o Amazon EMR Studio Workspaces. Para obter instruções sobre como iniciar seu cluster, consulte Especificar uma configuração de segurança para um cluster.
   - Escolha o rótulo de versão emr-6.11.0 ou posterior.
   - Selecione o Spark, o Livy e o Jupyter Enterprise Gateway como suas aplicações de cluster.
   - Use a configuração de segurança criada na etapa anterior.
   - Opcionalmente, você pode ativar o Lake Formation para seu EMR cluster. Para obter mais informações, consulte Habilite o Lake Formation com a Amazon EMR.

Depois de iniciar seu cluster, você estará pronto para usar o cluster habilitado para funções de tempo de execução com um EMR Studio Workspace.

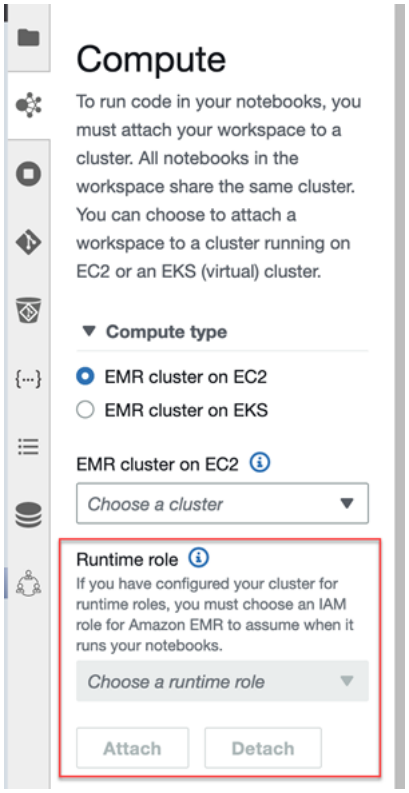
> **Note**
>
> No momento, o ExecutionRoleArnvalor não é suportado pela StartNotebookExecutionAPIoperação quando o `ExecutionEngineConfig.Type` valor é`EMR`.

Use o EMR cluster com uma função de tempo de execução no Workspaces

Depois de configurar e lançar seu cluster, você pode usar o cluster habilitado para funções de tempo de execução com seu EMR Studio Workspace.

1. Crie um novo Workspace ou inicie um Workspace existente. Para obter mais informações, consulte Crie um espaço de trabalho de EMR estúdio.

2. Escolha a guia EMRClusters na barra lateral esquerda do seu espaço de trabalho aberto, expanda a seção Tipo de computação e escolha seu cluster no menu Cluster no menu e a EMR função de tempo de execução no EC2 menu Função de tempo de execução.



3. Escolha Anexar para anexar o cluster com um perfil de runtime ao seu Workspace.

Considerações

Lembre-se das seguintes considerações ao usar um cluster habilitado para funções de tempo de execução com seu Amazon EMR Studio Workspace:

- Você só pode selecionar uma função de tempo de execução ao anexar um EMR Studio Workspace a um EMR cluster que usa a EMR versão 6.11 ou superior da Amazon.
- A funcionalidade de função de tempo de execução descrita nesta página só é compatível com a Amazon em EMR execução na Amazon EC2 e não com aplicativos interativos EMR sem servidor. Para saber mais sobre as funções de tempo de execução do EMR Serverless, consulte Job runtime roles no Amazon EMR Serverless User Guide.
- Embora você precise configurar permissões adicionais antes de especificar uma função de tempo de execução ao enviar um trabalho para um cluster, você não precisa de permissões adicionais

para acessar os arquivos gerados por um EMR Studio Workspace. As permissões para esses arquivos são semelhantes as dos arquivos gerados de clusters sem perfis de runtime.

- Você não pode usar o SQL Explorer em um EMR Studio Workspace com um cluster que tenha uma função de tempo de execução. A Amazon EMR desativa o SQL Explorer na interface do usuário quando um espaço de trabalho é anexado a um cluster habilitado para funções de tempo de EMR execução.
- Você não pode usar o modo de colaboração em um EMR Studio Workspace com um cluster que tenha uma função de tempo de execução. A Amazon EMR desativa os recursos de colaboração do Workspace quando um espaço de trabalho é anexado a um cluster habilitado para funções de tempo de execução. EMR O Workspace permanecerá acessível somente ao usuário que o anexou.
- Você não pode usar funções de tempo de execução em um Studio com a propagação de IAM identidade confiável do Identity Center ativada.
- Você pode encontrar um aviso “A página pode não ser segura!” da interface do usuário do Spark para um cluster habilitado para perfis de runtime. Se isso acontecer, ignore o alerta para continuar a visualizar a interface do usuário do Spark.

Execução de cadernos do Workspace de forma programática

> **Note**
>
> A execução programática de notebooks não é compatível com os aplicativos interativos Amazon EMR Serverless.

Você pode executar seus cadernos do Amazon EMR Studio Workspace programaticamente com um script ou no. AWS CLI Para saber como executar seu caderno de forma programática, consulte Exemplos de comandos para executar EMR Notebooks programaticamente.

Navegue pelos dados com o SQL Explorer

> **Note**
>
> SQLO Explorer for EMR Studio não é compatível com aplicativos interativos Amazon EMR Serverless ou em um Studio com a propagação de IAM identidade confiável do Identity Center ativada.

Este tópico fornece informações para ajudar você a começar a usar o SQL Explorer no Amazon EMR Studio. SQLO Explorer é uma ferramenta de página única em seu espaço de trabalho que ajuda você a entender as fontes de dados no catálogo de dados do seu EMR cluster. Você pode usar o SQL Explorer para pesquisar seus dados, executar SQL consultas para recuperar dados e baixar os resultados da consulta.

SQLO Explorer é compatível com o Presto. Antes de usar o SQL Explorer, verifique se você tem um cluster que usa a EMR versão 5.34.0 ou posterior da Amazon ou a versão 6.4.0 ou posterior com o Presto instalado. O Amazon EMR Studio SQL Explorer não é compatível com clusters do Presto que você configurou com criptografia em trânsito. Isso ocorre porque o Presto é executado no TLS modo nesses clusters.

Navegação pelo catálogo de dados do seu cluster

SQLO Explorer fornece uma interface de navegador de catálogos que você pode usar para explorar e entender como seus dados são organizados. Por exemplo, você pode usar o navegador do catálogo de dados para verificar os nomes das tabelas e colunas antes de escrever uma SQL consulta.

Navegar em seu catálogo de dados

1. Abra o SQL Explorer em seu espaço de trabalho.
2. Certifique-se de que seu espaço de trabalho esteja conectado a um EMR cluster em execução EC2 que usa a EMR versão 6.4.0 ou posterior da Amazon com o Presto instalado. Você pode escolher um cluster existente ou criar um novo. Para obter mais informações, consulte Anexar uma computação a um espaço de trabalho do EMR Studio.
3. Selecione um Banco de dados na lista suspensa para navegar.
4. Expanda uma tabela no seu banco de dados para visualizar os nomes das colunas da tabela. Também é possível inserir uma palavra-chave na barra de pesquisa para filtrar os resultados da tabela.

Execute uma SQL consulta para recuperar dados

Para recuperar dados com uma SQL consulta e baixar os resultados

1. Abra o SQL Explorer em seu espaço de trabalho.
2. Certifique-se de que seu espaço de trabalho esteja conectado a um EMR cluster em execução EC2 com o Presto e o Spark instalados. Você pode escolher um cluster existente ou criar um

novo. Para obter mais informações, consulte Anexar uma computação a um espaço de trabalho do EMR Studio.

3. Selecione Abrir editor para abrir uma nova guia do editor em seu Workspace.
4. Escreva sua SQL consulta na guia do editor.
5. Escolha Executar.
6. Visualize os resultados da consulta em Visualização do resultado. SQLO Explorer exibe os primeiros 100 resultados por padrão. Você pode escolher um número diferente de resultados a serem exibidos (até mil) usando o menu suspenso Visualizar os cem primeiros resultados da consulta.
7. Escolha Baixar resultados para baixar seus resultados em CSV formato. Você pode fazer download de até mil linhas de resultados.

Anexar uma computação a um espaço de trabalho do EMR Studio

O Amazon EMR Studio executa comandos do notebook usando um kernel em um EMR cluster. Antes de selecionar um kernel, você deve anexar o espaço de trabalho a um cluster que usa EC2 instâncias da Amazon, a um EKS cluster Amazon EMR on ou a um aplicativo sem EMR servidor. EMRO Studio permite anexar espaços de trabalho a clusters novos ou existentes e oferece a flexibilidade de alterar os clusters sem fechar o espaço de trabalho.

Esta seção aborda os tópicos a seguir para ajudá-lo a trabalhar e provisionar clusters para o EMR Studio:

- Anexar um EC2 cluster da Amazon a um EMR Studio Workspace
- Anexar um EKS cluster Amazon EMR on a um EMR Studio Workspace
- Anexe um aplicativo Amazon EMR Serverless a um Studio Workspace EMR
- Crie e anexe um novo EMR cluster a um EMR Studio Workspace
- Separar uma computação de um espaço de trabalho do Studio EMR

Anexar um EC2 cluster da Amazon a um EMR Studio Workspace

Você pode anexar um EMR cluster em execução na Amazon EC2 a um espaço de trabalho ao criar o espaço de trabalho ou anexar um cluster a um espaço de trabalho existente. Se você desejar criar e anexar um novo cluster, consulte Crie e anexe um novo EMR cluster a um EMR Studio Workspace.

> **Note**
>
> Um espaço de trabalho em um Studio que tenha a propagação de IAM identidade confiável do Identity Center ativada só pode ser anexada a um EMR cluster com uma configuração de segurança que tenha o Identity Center ativado.

On create

Conecte-se a um cluster de EMR computação da Amazon ao criar um espaço de trabalho

1. Na caixa de diálogo Criar um Workspace, certifique-se de já ter selecionado uma sub-rede para o novo Workspace. Expanda a seção Configuração avançada.
2. Escolha Anexar espaço de trabalho a um EMR cluster.
3. Na lista suspensa do EMRcluster, selecione um EMR cluster existente para anexar ao espaço de trabalho.

Depois de anexar um cluster, conclua a criação do Workspace. Ao abrir o novo espaço de trabalho pela primeira vez e escolher o painel de EMRclusters, você deverá ver o cluster selecionado anexado.

On launch

Conecte-se a um cluster de EMR computação da Amazon ao iniciar o Workspace

1. Navegue até a lista Workspaces e selecione a linha do Workspace que você deseja iniciar. Em seguida, selecione Iniciar o Workspace > Iniciar com opções.
2. Escolha um EMR cluster para anexar ao seu espaço de trabalho.

Depois de anexar um cluster, conclua a criação do Workspace. Ao abrir o novo espaço de trabalho pela primeira vez e escolher o painel de EMRclusters, você deverá ver o cluster selecionado anexado.

In JupyterLab

Anexe um espaço de trabalho a um cluster de EMR computação da Amazon em JupyterLab

1. Selecione seu Workspace e, em seguida, escolha Iniciar o Workspace > Início rápido.
2. Dentro JupyterLab, abra a guia Cluster na barra lateral esquerda.

3. Selecione o menu suspenso EMRon EC2 cluster ou selecione um Amazon EMR on EKS cluster.
4. Selecione Anexar para anexar o cluster ao seu Workspace.

Depois de anexar o cluster, conclua a criação do Workspace. Ao abrir o novo espaço de trabalho pela primeira vez e escolher o painel de EMRclusters, você deverá ver o cluster selecionado anexado.

In the Workspace UI

Anexe um espaço de trabalho a um cluster de EMR computação da Amazon a partir da interface de usuário do Workspace

1. No espaço de trabalho que você deseja anexar a um cluster, escolha o ícone de EMRclusters na barra lateral esquerda para abrir o painel Cluster.
2. Em Tipo de cluster, expanda a lista suspensa e selecione EMRcluster ativado. EC2
3. Escolha um cluster na lista suspensa. Pode ser necessário desanexar um cluster existente primeiro para habilitar a lista suspensa de seleção de cluster.
4. Escolha Anexar. Quando o cluster for anexado, você visualizará uma mensagem de êxito.

Anexar um EKS cluster Amazon EMR on a um EMR Studio Workspace

Além de usar EMR clusters da Amazon em execução na AmazonEC2, você pode anexar um espaço de trabalho a um EKS cluster Amazon EMR on para executar o código do notebook. Para obter mais informações sobre a Amazon EMR onEKS, consulte What is Amazon EMR on EKS.

Antes de conectar um espaço de trabalho a um EKS cluster Amazon EMR on, o administrador do Studio deve conceder a você permissões de acesso.

> **Note**
>
> Você não pode iniciar um EKS cluster Amazon EMR on em um EMR Studio que usa a propagação de IAM identidade confiável do Identity Center.

On create

Para anexar um Amazon EMR em um EKS cluster ao criar um espaço de trabalho

1. Na caixa de diálogo Criar um Workspace, expanda a seção Configuração avançada.
2. Escolha Anexar espaço de trabalho a um Amazon EMR on EKS cluster.
3. Em Amazon EMR on EKS cluster, escolha um cluster na lista suspensa.
4. Em Selecionar um endpoint, escolha um endpoint gerenciado para anexar ao Workspace. Um endpoint gerenciado é um gateway que permite que o EMR Studio se comunique com o cluster escolhido.
5. Escolha Criar um Workspace para concluir o processo de criação do Workspace e anexar o cluster selecionado.

Depois de anexar um cluster, você poderá concluir o processo de criação do Workspace. Ao abrir o novo espaço de trabalho pela primeira vez e escolher o painel de EMRclusters, você verá que o cluster selecionado está anexado.

In the Workspace UI

Para anexar um EKS cluster Amazon EMR on a partir da interface de usuário do Workspace

1. No espaço de trabalho que você deseja anexar a um cluster, escolha o ícone de EMRclusters na barra lateral esquerda para abrir o painel Cluster.
2. Expanda a lista suspensa Tipo de cluster e escolha EMRclusters em. EKS
3. Em EMRcluster ativado EKS, escolha um cluster na lista suspensa.
4. Em Endpoint, escolha um endpoint gerenciado para anexar ao Workspace. Um endpoint gerenciado é um gateway que permite que o EMR Studio se comunique com o cluster escolhido.
5. Escolha Anexar. Quando o cluster for anexado, você visualizará uma mensagem de êxito.

Anexe um aplicativo Amazon EMR Serverless a um Studio Workspace EMR

Você pode anexar um espaço de trabalho a um aplicativo EMR sem servidor para executar cargas de trabalho interativas. Para obter mais informações, consulte Usando notebooks para executar cargas de trabalho interativas com o EMR Serverless por meio do Studio. EMR

> **Note**
>
> Você não pode anexar um aplicativo EMR sem servidor a um EMR Studio que usa a propagação de IAM identidade confiável do Identity Center.

Example Anexe um espaço de trabalho a um aplicativo EMR sem servidor no JupyterLab

Antes de conectar um espaço de trabalho a um aplicativo EMR sem servidor, o administrador da sua conta deve conceder permissões de acesso conforme descrito em Permissões obrigatórias para cargas de trabalho interativas.

1. Navegue até o EMR Studio, selecione seu espaço de trabalho e, em seguida, selecione Launch Workspace > Início rápido.
2. Dentro JupyterLab, abra a guia Cluster na barra lateral esquerda.
3. Selecione EMRSem servidor como opção de computação e, em seguida, selecione um aplicativo EMR sem servidor e uma função de tempo de execução.
4. Selecione Anexar para anexar o cluster ao seu Workspace.

Agora, ao abrir esse Workspace, você deverá ver a aplicação selecionada anexada.

Crie e anexe um novo EMR cluster a um EMR Studio Workspace

Os usuários do Advanced EMR Studio podem provisionar novos EMR clusters em execução na Amazon EC2 para uso com um espaço de trabalho. O novo cluster tem todos os aplicativos de big data necessários para o EMR Studio instalados por padrão.

Para criar clusters, primeiro é necessário que o administrador do Studio conceda permissão a você usando uma política de sessão. Para obter mais informações, consulte Crie políticas de permissões para usuários do EMR Studio.

Você pode criar um novo cluster na caixa de diálogo Criar um Workspace ou no painel Cluster na interface do usuário do Workspace. De qualquer forma, você tem duas opções de criação de cluster:

1. Crie um EMR cluster — Crie um EMR cluster escolhendo o tipo e a contagem de EC2 instâncias da Amazon.
2. Usar um modelo de cluster: provisione um cluster ao selecionar um modelo de cluster definido previamente. Esta opção aparece se você tiver permissão para usar os modelos de cluster.

> **Note**
>
> Se você habilitou a propagação de IAM identidade confiável com o Identity Center for your Studio, deverá usar um modelo para criar um cluster.

Para criar um EMR cluster fornecendo uma configuração de cluster

1. Escolha um ponto de partida.

Para...	Fazer isso...
Criar o cluster ao criar um Workspace com a caixa de diálogo Criar um Workspace.	Expanda a seção Configuração avançada na caixa de diálogo Criar um espaço de trabalho e selecione Criar um EMR cluster.
Crie o cluster a partir do painel do EMRcluster na UI do Workspace depois de criar um Workspace.	Escolha a guia EMRClusters na barra lateral esquerda de um espaço de trabalho aberto, expanda a seção Configuração avançada e escolha Criar cluster.

2. Insira um Nome de cluster. Nomear o cluster ajuda você a encontrá-lo posteriormente na lista EMR Studio Clusters.
3. Para a EMRversão da Amazon, escolha uma EMR versão da Amazon para o cluster.
4. Em Instance, selecione o tipo e o número de EC2 instâncias da Amazon para o cluster. Para obter mais informações sobre como selecionar os tipos de instância, consulte Configurar EC2 instâncias da Amazon. Uma instância será usada como nó primário.
5. Selecione uma sub-rede na qual o EMR Studio possa iniciar o novo cluster. Cada opção de sub-rede é aprovada previamente pelo administrador do Studio, e seu Workspace deve ser capaz de se conectar a um cluster em qualquer sub-rede listada.
6. Escolha um S3 URI para armazenamento de registros.
7. Escolha Criar EMR cluster para provisionar o cluster. Se você usar a caixa de diálogo Criar um Workspace, escolha Criar um Workspace para criar o Workspace e provisionar o cluster. Depois que o EMR Studio provisiona o novo cluster, ele anexa o cluster ao espaço de trabalho.

Criar um cluster usando um modelo de cluster

1. Escolha um ponto de partida.

Para...	Fazer isso...
Criar o cluster ao criar um Workspace com a caixa de diálogo Criar um Workspace.	Expanda a seção Configuração avançada na caixa de diálogo Criar um Workspace e selecione Usar um modelo de cluster.
Crie o cluster a partir do painel do EMRcluster na interface do usuário do Workspace.	Escolha a guia EMRClusters na barra lateral esquerda de um espaço de trabalho aberto, expanda a seção Configuração avançada e escolha Modelo de cluster.

2. Selecione um modelo de cluster na lista suspensa. Cada modelo de cluster disponível inclui uma breve descrição para ajudar você a fazer uma seleção.
3. O modelo de cluster que você escolher pode ter parâmetros adicionais, como a versão de EMR lançamento da Amazon ou o nome do cluster. Você pode escolher ou inserir valores, ou usar os valores padrão selecionados pelo administrador.
4. Selecione uma sub-rede na qual o EMR Studio possa iniciar o novo cluster. Cada opção de sub-rede é aprovada previamente pelo administrador do Studio, e seu Workspace deve ser capaz de se conectar a um cluster em qualquer sub-rede.
5. Escolha Usar modelo de cluster para provisionar o cluster e anexá-lo ao Workspace. O EMR Studio levará alguns minutos para criar o cluster. Se você usar a caixa de diálogo Criar um Workspace, escolha Criar um Workspace para criar o Workspace e provisionar o cluster. Depois que o EMR Studio provisiona o novo cluster, ele anexa o cluster ao seu espaço de trabalho.

Separar uma computação de um espaço de trabalho do Studio EMR

Para trocar o cluster anexado a um Workspace, é possível desanexar um cluster da interface do usuário do Workspace.

Desanexar um cluster de um Workspace

1. No espaço de trabalho que você deseja separar de um cluster, escolha o ícone de EMRclusters na barra lateral esquerda para abrir o painel Cluster.

2. Em Selecionar cluster, escolha Desanexar e aguarde até que o EMR Studio desanexe o cluster. Quando o cluster for desanexado, você visualizará uma mensagem de êxito.

Para separar um aplicativo EMR sem servidor de um Studio Workspace EMR

Para trocar a computação anexada a um Workspace, é possível desanexar a aplicação da interface do usuário do Workspace.

1. No espaço de trabalho que você deseja separar de um cluster, escolha o ícone de EMRcomputação da Amazon na barra lateral esquerda para abrir o painel Computação.
2. Em Selecionar computação, escolha Desanexar e aguarde até que o EMR Studio desanexe o aplicativo. Quando a aplicação for desanexada, você visualizará uma mensagem de êxito.

Vincular repositórios baseados em Git a um espaço de trabalho do Studio EMR

Sobre os repositórios Git para Studio EMR

Você pode associar no máximo três repositórios Git a um EMR Studio Workspace. Por padrão, cada espaço de trabalho permite que você escolha em uma lista de repositórios Git associados à AWS mesma conta do Studio. Também é possível criar um novo repositório Git como um recurso para um Workspace.

Você pode executar comandos do Git, como os apresentados a seguir, usando um comando de terminal enquanto estiver conectado ao nó primário de um cluster.

```
!git pull origin <branch-name>
```

Como alternativa, você pode usar a extensão jupyterlab-git. Abra-o na barra lateral à esquerda ao escolher o ícone Git. Para obter informações sobre a extensão jupyterlab-git para, consulte jupyterlab-git. JupyterLab

Pré-requisitos

- Para associar um repositório Git a um Workspace, o Studio deve ser configurado para permitir a vinculação do repositório Git. O administrador do Studio deve tomar medidas para o Estabelecimento de acesso e de permissões para repositórios baseados em Git.

- Se você usa um CodeCommit repositório, deve usar as credenciais do Git e. HTTPS SSHchaves e HTTPS com o auxiliar de AWS Command Line Interface credenciais não são suportadas. CodeCommit também não suporta tokens de acesso pessoal (PATs). Para obter mais informações, consulte Usando IAM com CodeCommit no Guia do IAM usuário e Configuração para HTTPS usuários que usam credenciais do Git no Guia do AWS CodeCommit Usuário.

Instruções

Vincular um repositório Git associado a um Workspace

1. Abra o Workspace que você deseja vincular a um repositório na lista Workspaces no Studio.
2. Na barra lateral esquerda, escolha o ícone Amazon EMR Git Repository para abrir o painel de ferramentas do repositório Git.
3. Em Repositórios Git, expanda a lista suspensa e selecione, no máximo, três repositórios para vincular ao Workspace. EMRO Studio registra sua seleção e começa a vincular cada repositório.

Pode demorar algum tempo para que o processo de vinculação seja concluído. Você pode visualizar o status de cada repositório selecionado no painel de ferramentas Repositório Git. Depois que o EMR Studio vincular um repositório a um espaço de trabalho, você deverá ver os arquivos que pertencem a esse repositório aparecerem no painel Navegador de arquivos.

Adicionar um novo repositório Git a um Workspace como um recurso

1. Abra o Workspace que você deseja vincular a um repositório na lista Workspaces em seu Studio.
2. Na barra lateral esquerda, escolha o ícone Amazon EMR Git Repository para abrir o painel de ferramentas do repositório Git.
3. Escolha Adicionar novo repositório Git.
4. Em Nome do repositório, insira um nome descritivo para o repositório no Studio. EMR Os nomes podem conter somente caracteres alfanuméricos, hifens e sublinhados.
5. Para o repositório GitURL, insira o URL para o repositório. Quando você usa um CodeCommit repositório, esse é o URL que é copiado quando você escolhe Clonar URL e depois Clonar. HTTPS Por exemplo, `https://git-codecommit.us-west-2.amazonaws.com/v1/repos/[MyCodeCommitRepoName]`.
6. Em Filial, insira o nome de uma filial existente que você deseja conferir.

7. Em Credenciais do Git, escolha uma opção de acordo com as diretrizes apresentadas a seguir. EMRO Studio acessa suas credenciais do Git usando segredos armazenados no Secrets Manager.

> **Note**
>
> Se você usa um GitHub repositório, recomendamos que você use um token de acesso pessoal (PAT) para autenticar. A partir de 13 de agosto de 2021, GitHub exigirá autenticação baseada em tokens e não aceitará mais senhas ao autenticar operações do Git. Para obter mais informações, consulte a publicação Requisitos de autenticação de token para operações do Git no The GitHub Blog.

Opção	Descrição
Criar um novo segredo	Escolha essa opção para associar as credenciais existentes do Git a um novo segredo que será criado para você. AWS Secrets Manager Execute um dos seguintes procedimentos com base nas credenciais do Git que você usar para o repositório. Se você usar um nome de usuário e uma senha do Git para acessar o repositório, selecione Nome de usuário e senha, insira o Nome do segredo a ser usado no Secrets Manager e, em seguida, insira o Nome de usuário e a Senha a serem associados ao segredo. OU Se você usar um token de acesso pessoal para acessar o repositório, selecione Token de acesso pessoal (PAT), insira o nome secreto a ser usado no Secrets Manager e, em seguida, insira seu token de acesso pessoal. Para obter mais informações, consulte Criação de um token de acesso pessoal para a linha de comando GitHub e Tokens de acesso pessoal para o Bitbucket . CodeCommit os repositórios não oferecem suporte a essa opção.
Usar um repositório público sem credenciais	Escolha esta opção para acessar um repositório público.

Opção	Descrição
Use um AWS segredo existente	Escolha esta opção se você já salvou suas credenciais como um segredo no Secrets Manager e, em seguida, selecione o nome do segredo na lista. Se você selecionar um segredo associado a um nome de usuário e senha do Git, o segredo deverá estar no formato `{"gitUsername": " MyUserName ", "gitPassword": " MyPassword "}`.

8. Escolha Adicionar repositório para criar o novo repositório. Depois que o EMR Studio criar o novo repositório, você verá uma mensagem de sucesso. O novo repositório aparece na lista suspensa em Repositórios Git.
9. Para vincular o novo repositório ao seu Workspace, escolha-o na lista suspensa em Repositórios Git.

Pode demorar algum tempo para que o processo de vinculação seja concluído. Depois que o EMR Studio vincular o novo repositório ao Espaço de trabalho, você verá uma nova pasta com o mesmo nome do seu repositório aparecer no painel Navegador de arquivos.

Para abrir um repositório vinculado diferente, navegue até a pasta dele no Navegador de arquivos.

Use o SQL editor Amazon Athena no Studio EMR

Visão geral

Você pode usar o Amazon EMR Studio para desenvolver e executar consultas interativas no Amazon Athena. Isso significa que você pode realizar SQL análises no Athena a partir da mesma interface do EMR Studio usada para executar seu Spark, Scala e outras cargas de trabalho. Com essa integração, você pode usar o preenchimento automático para desenvolver consultas rapidamente, pesquisar dados em seu AWS Glue Data Catalog, criar consultas salvas, visualizar seu histórico de consultas e muito mais.

Para obter mais informações sobre o uso do Amazon Athena, consulte Como usar o Athena no Guia do usuário do SQL Amazon Athena.

Use o SQL editor Athena no Studio EMR

Use as etapas a seguir para desenvolver e executar consultas interativas no Amazon Athena a partir do EMR seu Studio:

1. Adicione as permissões necessárias ao perfil de usuário para os usuários que acessam o Workspaces nesse Studio. As permissões estão listadas na AWS Identity and Access Management permissões para usuários do EMR Studio tabela na coluna Acesse o SQL editor do Amazon Athena a partir do seu EMR Studio. Como alternativa, você pode optar por copiar o conteúdo da política avançada do Exemplo de políticas de usuário para conceder aos usuários permissões completas sobre os recursos do EMR Studio, incluindo este.
2. Configure e crie um EMR estúdio.
3. Navegue até seu Studio e selecione Editor de consultas na barra lateral.

Agora, você deve ver a interface do usuário familiar do editor do Athena. Para obter informações sobre como começar e usar o Athena SQL para executar consultas interativas, consulte Introdução e uso do Athena no Guia do usuário do SQL Amazon Athena.

> **Note**
>
> Se você habilitou a propagação de identidade confiável por meio do IAM Identity Center para seu EMR Studio, deverá usar os grupos de trabalho do Athena para controlar o acesso às consultas, e o grupo de trabalho que você usa também deve usar a propagação de identidade confiável. Para ver as etapas para configurar o Identity Center e permitir a propagação confiável de identidade para seu grupo de trabalho, consulte Como usar grupos de trabalho do Athena habilitados para o IAM Identity Center no Guia do usuário do Amazon Athena.

Considerações sobre o uso do editor SQL Athena no Studio EMR

- A integração com o Athena está disponível em todas as regiões comerciais em que o EMR Studio e o Athena estão disponíveis.
- Os seguintes recursos do Athena não estão disponíveis no Studio: EMR
  - Recursos administrativos, como a criação ou a atualização de grupos de trabalho, fontes de dados ou reservas de capacidade do Athena.
  - Athena para o Spark ou para cadernos do Spark

- DataZone Integração com a Amazon
- Otimizador baseado em custos () CBO
- Step Functions

CodeWhisperer Integração da Amazon com o EMR Studio Workspaces

Visão geral

Você pode usar a Amazon CodeWhisperer com o Amazon EMR Studio para obter recomendações em tempo real à medida que escreve código JupyterLab. CodeWhispererpode concluir seus comentários, concluir linhas únicas de código, fazer line-by-line recomendações e gerar funções totalmente formadas.

> **Note**
>
> Quando você usa o Amazon EMR Studio, AWS pode armazenar dados sobre seu uso e conteúdo para fins de melhoria do serviço. Para obter mais informações e instruções sobre como cancelar o compartilhamento de dados, consulte Compartilhando seus dados AWS no Guia CodeWhisperer do usuário da Amazon.

Considerações sobre o uso CodeWhisperer com espaços de trabalho

- CodeWhisperer a integração está disponível no mesmo Regiões da AWS local em que o EMR Studio está disponível, conforme documentado nas considerações do EMR Studio.
- O Amazon EMR Studio usa automaticamente o CodeWhisperer endpoint no Leste dos EUA (Norte da Virgínia) (us-east-1) para recomendações, independentemente da região em que seu estúdio esteja.
- CodeWhisperer suporta somente a linguagem Python para ETL scripts de codificação para trabalhos do Spark no Studio. EMR
- Uma opção de telemetria do lado do cliente quantifica seu uso de. CodeWhisperer Essa funcionalidade não é compatível com o EMR Studio.

Permissões necessárias para CodeWhisperer

Para usar CodeWhisperer, você deve anexar a seguinte política à sua função de IAM usuário no Amazon EMR Studio:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
          "Sid": "CodeWhispererPermissions",
          "Effect": "Allow",
          "Action": [ "codewhisperer:GenerateRecommendations" ],
          "Resource": "*"
        }
     ]
}
```

Use CodeWhisperer com espaços de trabalho

Para exibir o registro de CodeWhisperer referência JupyterLab, abra o CodeWhispererpainel na parte inferior da JupyterLab janela e escolha Abrir registro de referência de código.

A lista a seguir contém atalhos que você pode usar para interagir com CodeWhisperer sugestões:

- Recomendações de pausa — Use as sugestões automáticas de pausa nas configurações. CodeWhisperer
- Aceitar uma recomendação: pressione Tab no teclado.
- Rejeitar uma recomendação: pressione Esc no teclado.
- Navegar pelas recomendações: use as setas para cima e para baixo no teclado.
- Invocação manual: pressione Alt e C no teclado. Se estiver usando um Mac, pressione Cmd e C.

Você também pode usar CodeWhisperer para alterar configurações, como nível de registro, e obter sugestões de referências de código. Para obter mais informações, consulte Configuração CodeWhisperer JupyterLab e recursos no Guia do CodeWhisperer usuário da Amazon.

Depure aplicativos e trabalhos com EMR o Studio

Com o Amazon EMR Studio, você pode iniciar interfaces de aplicativos de dados para analisar aplicativos e execuções de trabalhos no navegador.

Você também pode iniciar as interfaces de usuário persistentes e fora do cluster para a Amazon, EMR executadas em EC2 clusters a partir do EMR console da Amazon. Para obter mais informações, consulte Visualizar interfaces do usuário de aplicações persistentes.

> **Note**
>
> Com base nas configurações do seu navegador, pode ser necessário habilitar pop-ups para a abertura da interface do usuário de uma aplicação.

Para obter informações sobre como configurar e usar as interfaces do aplicativo, consulte The YARN Timeline Server, Monitoring and instrumentation ou Visão geral da Tez UI.

Depure a Amazon em EMR execução com trabalhos da Amazon EC2

Workspace UI

Inicialização de uma interface do usuário no cluster usando um arquivo de caderno

Ao usar as EMR versões 5.33.0 e posteriores da Amazon, você pode iniciar a interface de usuário web do Spark (a interface do usuário do Spark ou o Spark History Server) a partir de um notebook no seu espaço de trabalho.

UIsTrabalhe em cluster com os PySpark kernels Spark ou SparkR. O tamanho máximo de arquivo visível para logs de eventos ou para logs de contêineres do Spark é de 10 MB. Se seus arquivos de log excederem 10 MB, recomendamos usar o servidor de histórico do Spark persistente em vez da interface do usuário do Spark no cluster para depurar trabalhos.

> **Important**
>
> Para que o EMR Studio execute interfaces de usuário de aplicativos em cluster a partir de um espaço de trabalho, um cluster deve ser capaz de se comunicar com o Amazon API Gateway. Você deve configurar o EMR cluster para permitir o tráfego de saída da rede para o Amazon API Gateway e garantir que o Amazon API Gateway possa ser acessado a partir do cluster.
> A interface do usuário do Spark acessa os logs de contêineres ao resolver nomes de host. Se você usar um nome de domínio personalizado, certifique-se de que os nomes de host dos nós do cluster possam ser resolvidos pela Amazon DNS ou pelo DNS servidor que você especificar. Para fazer isso, defina as opções do Dynamic Host Configuration

Protocol (DHCP) para a Amazon Virtual Private Cloud (VPC) associada ao seu cluster. Para obter mais informações sobre DHCP as opções, consulte os conjuntos de DHCP opções no Guia do usuário da Amazon Virtual Private Cloud.

1. No seu EMR Studio, abra o espaço de trabalho que você deseja usar e certifique-se de que ele esteja conectado a um EMR cluster da Amazon em EC2 execução. Para obter instruções, consulte Anexar uma computação a um espaço de trabalho do EMR Studio.
2. Abra um arquivo de notebook e use o PySpark kernel, Spark ou SparkR. Para selecionar um kernel, escolha o nome do kernel no canto superior direito da barra de ferramentas do caderno para abrir a caixa de diálogo Selecionar kernel. O nome aparecerá como Nenhum Kernel! se nenhum kernel tiver sido selecionado.
3. Execute o código do seu caderno. Quando você inicia o Spark Context, o apresentado a seguir aparece como a saída no caderno. Pode demorar alguns segundos para que a aparição ocorra. Se você iniciou o Spark Context, poderá executar o comando `%%info` para acessar um link para a interface do usuário do Spark a qualquer momento.

   > **Note**
   >
   > Se os links da interface do usuário do Spark não funcionarem ou não aparecerem após alguns segundos, crie uma nova célula de caderno e execute o comando `%%info` para gerar os links novamente.

4. Para iniciar a interface do usuário do Spark, escolha Link em IU do Spark. Se a aplicação do Spark estiver em execução, a interface do usuário do Spark será aberta em uma nova guia. Se aplicação estiver sido concluída, o servidor de histórico do Spark será aberto.

Depois de iniciar a interface do usuário do Spark, você pode modificá-la URL no navegador para abrir o YARN ResourceManager ou o Yarn Timeline Server. Adicione um dos caminhos apresentados a seguir depois de `amazonaws.com`.

Interface do usuário da Web	Path	Exemplo modificado URL
YARN ResourceManager	/rm	https://*j-examplebby5ij* .emrappui-prod.*eu-west-1* .amazonaws.com/*rm*
Servidor de linha do tempo do YARN	/yts	https://*j-examplebby5ij* .emrappui-prod.*eu-west-1* .amazonaws.com/*yts*
Servidor de histórico do Spark	/shs	https://*j-examplebby5ij* .emrappui-prod.*eu-west-1* .amazonaws.com/*shs*

Studio UI

Inicie o YARN Timeline Server persistente, o Spark History Server ou a interface do usuário Tez a partir da interface do Studio EMR

1. No seu EMR Studio, selecione Amazon EMR EC2 no lado esquerdo da página para abrir a lista Amazon EMR on EC2 clusters.
2. Filtre a lista de clusters por nome, estado ou ID ao inserir valores na caixa de pesquisa. Você também pode pesquisar por intervalo de tempo de criação.
3. Selecione um cluster e, em seguida, escolha UIsIniciar aplicativo para selecionar uma interface de usuário do aplicativo. A interface do usuário da aplicação abre em uma nova guia do navegador e pode demorar algum tempo para carregar.

Debug EMR Studio em execução sem servidor EMR

Semelhante à Amazon em EMR execução na AmazonEC2, você pode usar a interface de usuário do Workspace para analisar seus aplicativos sem EMR servidor. Na interface do usuário do Workspace,

ao usar as EMR versões 6.14.0 e superiores da Amazon, você pode iniciar a interface de usuário web do Spark (a interface do usuário do Spark ou o Spark History Server) a partir de um notebook no seu espaço de trabalho. Para sua conveniência, também fornecemos um link para o log do driver para acesso rápido aos logs do driver do Spark.

Depure a Amazon EMR em execuções de EKS trabalhos com o Spark History Server

Ao enviar uma execução de trabalho para um EKS cluster Amazon EMR on, você pode acessar os registros dessa execução de trabalho usando o Spark History Server. O Spark History Server fornece ferramentas para monitorar aplicativos Spark, como uma lista de etapas e tarefas do agendador, um resumo dos RDD tamanhos e do uso de memória e informações ambientais. Você pode iniciar o Spark History Server para Amazon EMR em execuções de EKS trabalhos das seguintes formas:

- Ao enviar um trabalho executado usando o EMR Studio com um endpoint EKS gerenciado Amazon EMR on, você pode iniciar o Spark History Server a partir de um arquivo de notebook em seu espaço de trabalho.
- Ao enviar uma execução de trabalho usando o AWS CLI ou AWS SDK para o Amazon EMR onEKS, você pode iniciar o Spark History Server a partir da interface do usuário do EMR Studio.

Para obter informações sobre como usar o servidor de histórico do Spark, consulte Monitoring and Instrumentation na documentação do Apache Spark. Para obter mais informações sobre a execução de trabalhos, consulte Conceitos e componentes no Amazon EMR on EKS Development Guide.

Para iniciar o Spark History Server a partir de um arquivo de caderno no seu EMR Studio Workspace

1. Abra um espaço de trabalho conectado a um EKS cluster Amazon EMR on.
2. Selecione e abra seu arquivo de caderno no Workspace.
3. Escolha IU do Spark na parte superior do arquivo de caderno para abrir o servidor de histórico do Spark persistente em uma nova guia.

Para iniciar o Spark History Server a partir da interface do usuário do EMR Studio

> **Note**
>
> A lista de trabalhos na interface do usuário do EMR Studio exibe somente as execuções de trabalhos que você envia usando o AWS CLI ou AWS SDK para a Amazon EMR emEKS.

1. No seu EMR Studio, selecione Amazon EMR EKS no lado esquerdo da página.
2. Pesquise a Amazon EMR no cluster EKS virtual que você usou para enviar sua execução de trabalho. É possível filtrar a lista de clusters por status ou ID ao inserir valores na caixa de pesquisa.
3. Selecione o cluster para abrir a página de detalhes dele. A página de detalhes exibe informações sobre o cluster, como o ID, o namespace e o status. A página também mostra uma lista com todas as execuções de trabalhos enviadas para esse cluster.
4. Na página de detalhes do cluster, selecione uma execução de trabalho para depurar.
5. No canto superior à direita da lista Trabalhos, escolha Iniciar servidor de histórico do Spark para abrir a interface da aplicação em uma nova guia do navegador.

Instale kernels e bibliotecas em um Studio Workspace EMR

Cada Amazon EMR Studio Workspace vem com um conjunto de bibliotecas e kernels pré-instalados.

Kernels e bibliotecas em clusters executados na Amazon EC2

Você também pode personalizar o ambiente do EMR Studio das seguintes maneiras ao usar EMR clusters em execução na AmazonEC2:

- Instalar kernels do caderno Jupyter e bibliotecas Python em um nó primário do cluster: ao instalar bibliotecas usando esta opção, todos os Workspaces anexados ao mesmo cluster compartilham essas bibliotecas. Você pode instalar kernels ou bibliotecas de dentro de uma célula do notebook ou enquanto estiver conectado usando SSH o nó primário de um cluster.
- Usar bibliotecas com escopo de cadernos: quando os usuários do Workspace instalam e usam bibliotecas a partir de uma célula de caderno, essas bibliotecas ficam disponíveis somente para esse caderno. Esta opção permite que diferentes cadernos que usam o mesmo cluster funcionem sem se preocupar com versões conflitantes da biblioteca.

EMROs espaços de trabalho do Studio têm a mesma arquitetura subjacente dos EMR notebooks. Você pode instalar e usar os kernels do Jupyter Notebook e as bibliotecas Python com o Studio da mesma forma que EMR faria com o Notebooks. EMR Para obter instruções, consulte Instalação e uso de kernels e bibliotecas.

Kernels e bibliotecas na Amazon EMR em clusters EKS

A Amazon EMR em EKS clusters inclui os kernels PySpark e Python 3.7 com um conjunto de bibliotecas pré-instaladas. O Amazon EMR on EKS não oferece suporte à instalação de bibliotecas ou clusters adicionais.

Cada EKS cluster Amazon EMR on vem com os seguintes Python e PySpark bibliotecas instaladas:

- Python – boto3, cffi, future, ggplot, jupyter, kubernetes, matplotlib, numpy, pandas, plotly, pycryptodomex, py4j, requests, scikit-learn, scipy, seaborn
- PySpark – ggplot, jupyter, matplotlib, numpy, pandas, plotly, pycryptodomex, py4j, requests, scikit-learn, scipy, seaborn

Kernels e bibliotecas em aplicativos sem servidor EMR

Cada aplicativo EMR sem servidor vem com o seguinte Python e bibliotecas instaladas: PySpark

- Python – ggplot, matplotlib, numpy, pandas, plotly, bokeh, scikit-learn, scipy, seaborn
- PySpark – ggplot, matplotlib,numpy, pandas, plotly, bokeh, scikit-learn, scipy, seaborn

Aprimoramento de kernels com comandos magic

Visão geral

EMRmagicComandos de suporte para Studio e EMR Notebooks. Magiccomandos, ou magics, são aprimoramentos que o IPython kernel fornece para ajudá-lo a executar e analisar dados. IPythoné um ambiente de shell interativo criado com Python.

A Amazon EMR também oferece suporte Sparkmagic a um pacote que fornece magic comandos específicos aos kernels relacionados ao Spark (kernels PySpark SparkR e Scala) e que usa o Livy no cluster para enviar trabalhos do Spark.

Você pode usar magic comandos desde que tenha um kernel Python em seu notebook. EMR De forma semelhante, qualquer kernel relacionado ao Spark oferece suporte aos comandos do Sparkmagic.

Os comandos Magic, também chamados de magics, têm duas variedades:

- Linhas magics: esses comandos magic são indicados por um prefixo `%` único e operam em uma única linha de código.
- Células magics: esses comandos magic são indicados por um prefixo `%%` duplo e operam em várias linhas de código.

Para saber todos os magics disponíveis, consulte Listar os comandos magic e Sparkmagic.

Considerações e limitações

- EMRO Serverless não suporta `%%sh` a execução. `spark-submit` Ele não é compatível com os EMR Notebooks magic s.
- A Amazon EMR em EKS clusters não oferece suporte a Sparkmagic comandos para o EMR Studio. Isso ocorre porque os kernels Spark que você usa com endpoints gerenciados são integrados ao Kubernetes e não são suportados pelo Sparkmagic e pelo Livy. Você pode definir a configuração do Spark diretamente no SparkContext objeto como uma solução alternativa, conforme demonstra o exemplo a seguir.

```
spark.conf.set("spark.driver.maxResultSize", '6g')
```

- Os seguintes magic comandos e ações são proibidos por AWS:
  - `%alias`
  - `%alias_magic`
  - `%automagic`
  - `%macro`
  - Modificar `proxy_user` com `%configure`
  - Modificar `KERNEL_USERNAME` com `%env` ou `%set_env`

Listar os comandos magic e Sparkmagic

Use os seguintes comandos para listar os comandos magic disponíveis:

- `%lsmagic` lista todas as funções magic disponíveis no momento.
- `%%help` lista as funções magic relacionadas ao Spark disponíveis no momento e fornecidas pelo pacote Sparkmagic.

Use `%%configure` para configurar o Spark

Um dos comandos mais úteis do Sparkmagic é o comando `%%configure`, que configura os parâmetros de criação da sessão. Ao usar as configurações `conf`, você pode definir qualquer configuração do Spark mencionada na documentação de configuração do Apache Spark.

Example Adicione um JAR arquivo externo aos EMR Notebooks a partir do repositório Maven ou do Amazon S3

Você pode usar a abordagem a seguir para adicionar uma dependência de JAR arquivo externo a qualquer kernel relacionado ao Spark que seja suportado pelo. Sparkmagic

```
%%configure -f
{"conf": {
    "spark.jars.packages": "com.jsuereth:scala-arm_2.11:2.0,ml.combust.bundle:bundle-
ml_2.11:0.13.0,com.databricks:dbutils-api_2.11:0.0.3",
    "spark.jars": "s3://DOC-EXAMPLE-BUCKET/my-jar.jar"
    }
}
```

Example : Configuração do Hudi

Você pode usar o editor do notebook para configurar seu EMR notebook para usar o Hudi.

```
%%configure
{ "conf": {
     "spark.jars": "hdfs://apps/hudi/lib/hudi-spark-bundle.jar,hdfs:///apps/hudi/lib/
spark-spark-avro.jar",
     "spark.serializer": "org.apache.spark.serializer.KryoSerializer",
     "spark.sql.hive.convertMetastoreParquet":"false"
     }
}
```

Use o `%%sh` para executar o `spark-submit`

A `%%sh` magic executa comandos de shell em um subprocesso em uma instância do cluster anexado. Normalmente, você usaria um dos kernels relacionados ao Spark para executar aplicações do Spark em seu cluster anexado. No entanto, se desejar usar um kernel do Python para enviar uma aplicação do Spark, você pode usar a magic apresentada a seguir, substituindo o nome do bucket pelo nome do seu bucket em letras minúsculas.

```
%%sh
spark-submit --master yarn --deploy-mode cluster s3://DOC-EXAMPLE-BUCKET/test.py
```

Neste exemplo, o cluster precisa de acesso ao local `s3://DOC-EXAMPLE-BUCKET/test.py` ou o comando falhará.

Você pode usar qualquer comando do Linux com a `%%sh` magic. Se você quiser executar qualquer Spark ou qualquer YARN comando, use uma das seguintes opções para criar um usuário do `emr-notebook` Hadoop e conceder ao usuário permissões para executar os comandos:

- Você pode criar explicitamente um novo usuário ao executar os comandos a seguir.

  ```
  hadoop fs -mkdir /user/emr-notebook
  hadoop fs -chown emr-notebook /user/emr-notebook
  ```

- Você pode ativar a representação do usuário no Livy, que cria o usuário automaticamente. Consulte Habilitação da representação do usuário para monitorar a atividade de usuários e trabalhos do Spark Para mais informações.

Use **`%%display`** para visualizar dataframes do Spark

Você pode usar o `%%display` magic para visualizar um dataframe do Spark. Para usar essa magic, execute o comando apresentado a seguir.

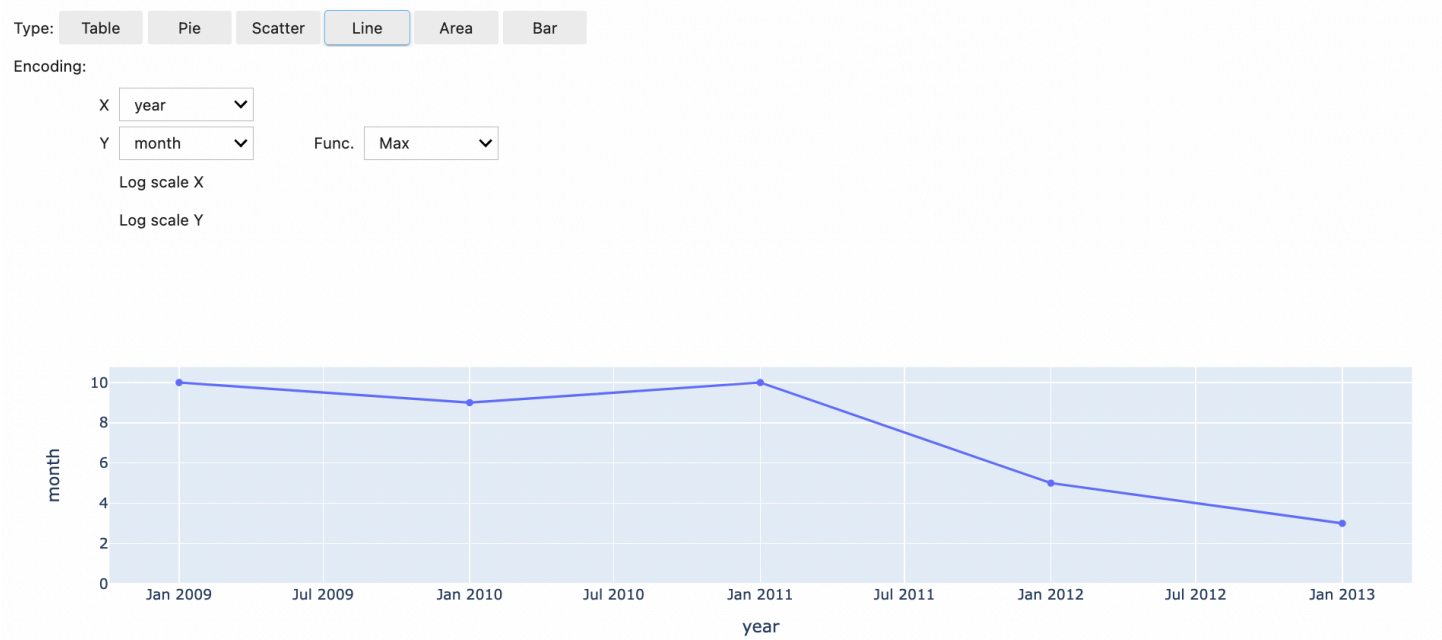
```
%%display df
```

Escolha visualizar os resultados em formato de tabela, como mostra a imagem a seguir.

Type: Table Pie Scatter Line Area Bar

year	month	total_passengers	total_trips
2012-01-01	3	26866837	16146923
2011-01-01	3	26091246	16066350
2013-01-01	3	26965079	15749228
2011-01-01	10	26287953	15707756
2009-01-01	10	26202049	15604551
2012-01-01	5	26278817	15567525
2011-01-01	5	25508952	15554868
2010-01-01	9	25533166	15540209
2010-01-01	5	26002858	15481351
2012-01-01	4	25900645	15477914

Você também pode optar por visualizar os dados com cinco tipos de gráficos. Suas opções incluem gráficos circular, de dispersão, de linha, de área e de barras.



Use EMR notebooks magic como

A Amazon EMR fornece os seguintes EMR notebooks que você pode usar com magic kernels baseados em Python3 e Spark:

- `%mount_workspace_dir`: monta seu diretório do Workspace em seu cluster para que você possa importar e executar códigos de outras aplicações em seu Workspace.

> **Note**
>
> Com `%mount_workspace_dir`, somente o kernel do Python 3 pode acessar seus sistemas de arquivos locais. Os executores do Spark não terão acesso ao diretório montado com este kernel.

- `%umount_workspace_dir`: desmonta seu diretório do Workspace do seu cluster.
- `%generate_s3_download_url`: gera um link de download temporário na saída do seu caderno para um objeto do Amazon S3.

Pré-requisitos

Antes de instalar o EMR Notebooks magic s, conclua as seguintes tarefas:

- Certifique-se de que seu Função de serviço para EC2 instâncias de cluster (perfil de EC2 instância) tenha acesso de leitura para o Amazon S3. O `EMR_EC2_DefaultRole` com a política gerenciada `AmazonElasticMapReduceforEC2Role` atende a esse requisito. Se você usar uma política ou um perfil personalizado, certifique-se de que ele tenha as permissões do S3 necessárias.

> **Note**
>
> EMROs notebooks magic são executados em um cluster como usuário do notebook e usam o perfil da EC2 instância para interagir com o Amazon S3. Quando você monta um diretório de espaço de trabalho em um EMR cluster, todos os espaços de trabalho e EMR notebooks com permissão para se conectar a esse cluster podem acessar o diretório montado.
>
> Por padrão, os diretórios são montados somente para leitura. Embora `s3fs-fuse` e `goofys` permitam montagens de leitura e de gravação, recomendamos fortemente que você não modifique os parâmetros de montagem para montar diretórios no modo de leitura

e de gravação. Se você permitir o acesso de gravação, todas as alterações realizadas no diretório serão gravadas no bucket do S3. Para evitar a exclusão ou a substituição acidentais, você pode habilitar o versionamento para seu bucket do S3. Para saber mais, consulte Usando o versionamento em buckets do S3.

- Execute um dos scripts a seguir em seu cluster para instalar as dependências do EMR Notebooks s. magic Para executar um script, você pode Usar ações de bootstrap personalizadas ou seguir as instruções em Executar comandos e scripts em um EMR cluster da Amazon quando você já tiver um cluster em execução.

  Você pode escolher qual dependência instalar. Tanto o s3fs-fuse quanto o goofys são ferramentas FUSE (sistema de arquivos no espaço do usuário) que permitem montar um bucket do Amazon S3 como um sistema de arquivos local em um cluster. A `s3fs` ferramenta oferece uma experiência semelhante POSIX a. A `goofys` ferramenta é uma boa opção quando você prefere desempenho a um sistema POSIX de arquivos compatível.

  A série Amazon EMR 7.x usa o Amazon Linux 2023, que não oferece suporte a EPEL repositórios. Se você estiver executando o Amazon EMR 7.x, siga as instruções do s3fs-fuse para instalar GitHub. `s3fs-fuse` Se você usa as séries 5.x ou 6.x, use os seguintes comandos para instalar. `s3fs-fuse`

  ```
  #!/bin/sh

  # Install the s3fs dependency for EMR Notebooks magics
  sudo amazon-linux-extras install epel -y
  sudo yum install s3fs-fuse -y
  ```

  OU

  ```
  #!/bin/sh

  # Install the goofys dependency for EMR Notebooks magics
  sudo wget https://github.com/kahing/goofys/releases/latest/download/goofys -P /usr/
  bin/
  sudo chmod ugo+x /usr/bin/goofys
  ```

Instale EMR Notebooks s magic

> **Note**
>
> Com as EMR versões 6.0 a 6.9.0 e 5.0 a 5.36.0 da Amazon, somente as versões de `emr-notebooks-magics` pacotes 0.2.0 e superiores oferecem suporte. `%mount_workspace_dir` magic

Conclua as etapas a seguir para instalar o EMR Notebooks magic s.

1. Em seu caderno, execute os comandos apresentados a seguir para instalar o pacote [emr-notebooks-magics].

   ```
   %pip install boto3 --upgrade
   %pip install botocore --upgrade
   %pip install emr-notebooks-magics --upgrade
   ```

2. Reinicie seu kernel para carregar os EMR Notebooks s. magic
3. Verifique a instalação com o comando a seguir, que deve exibir o texto de ajuda de saída para `%mount_workspace_dir`.

   ```
   %mount_workspace_dir?
   ```

Montagem de um diretório do Workspace com **`%mount_workspace_dir`**

O `%mount_workspace_dir` magic permite que você monte seu diretório do Workspace em seu EMR cluster para que você possa importar e executar outros arquivos, módulos ou pacotes armazenados em seu diretório.

O exemplo a seguir monta todo o diretório do Workspace em um cluster e especifica o opcional *<--tipo de fusível>* argumento para usar goofys para montar o diretório.

```
%mount_workspace_dir . <--fuse-type goofys>
```

Para verificar se o diretório do Workspace está montado, use o exemplo a seguir para exibir o diretório de trabalho atual com o comando `ls`. A saída deve exibir todos os arquivos em seu Workspace.

```
%%sh
ls
```

Quando você terminar de fazer as alterações no Workspace, poderá desmontar o diretório do Workspace com o seguinte comando:

> **Note**
>
> O diretório do Workspace permanece montado em seu cluster mesmo quando o Workspace é interrompido ou desanexado. Você deve desmontar explicitamente o diretório do Workspace.

```
%umount_workspace_dir
```

Download de um objeto do Amazon S3 com **`%generate_s3_download_url`**

O `generate_s3_download_url` comando cria um pré-assinado URL para um objeto armazenado no Amazon S3. Você pode usar o preassinado URL para baixar o objeto em sua máquina local. Por exemplo, você pode `generate_s3_download_url` executar o download do resultado de uma SQL consulta que seu código grava no Amazon S3.

O pré-assinado URL é válido por 60 minutos por padrão. Você pode alterar o tempo de expiração ao especificar um número de segundos para o sinalizador `--expires-in`. Por exemplo, `--expires-in 1800` cria um URL que é válido por 30 minutos.

O exemplo apresentado a seguir gera um link de download para um objeto ao especificar o caminho completo do Amazon S3: *`s3://EXAMPLE-DOC-BUCKET/path/to/my/object`*.

```
%generate_s3_download_url s3://EXAMPLE-DOC-BUCKET/path/to/my/object
```

Para saber mais sobre como usar `generate_s3_download_url`, execute o comando a seguir para exibir o texto de ajuda.

```
%generate_s3_download_url?
```

Execução de um caderno no modo descentralizado com `%execute_notebook`

Com a magic do `%execute_notebook`, você pode executar outro caderno no modo descentralizado e visualizar a saída de cada célula executada. Isso magic requer permissões adicionais para a função de instância que a Amazon EMR e a Amazon EC2 compartilham. Para obter mais detalhes sobre como conceder permissões adicionais, execute o comando `%execute_notebook?`.

Durante um trabalho de execução prolongada, seu sistema pode entrar em repouso devido à inatividade ou pode perder temporariamente a conectividade com a Internet. Isso pode interromper a conexão entre o seu navegador e o servidor Jupyter. Nesse caso, você poderá perder a saída das células que executou e enviou usando o servidor Jupyter.

Se você executar o notebook no modo headless com `%execute_notebook`magic, o EMR Notebooks capturará a saída das células que foram executadas, mesmo que a rede local sofra interrupções. EMRO Notebooks salva a saída de forma incremental em um novo notebook com o mesmo nome do notebook que você executou. EMREm seguida, os cadernos colocam o caderno em uma nova pasta dentro do espaço de trabalho. As execuções descentralizadas ocorrem no mesmo cluster e usam o perfil de serviço `EMR_Notebook_DefaultRole`, mas argumentos adicionais podem alterar os valores padrão.

Para executar um caderno no modo descentralizado, use o seguinte comando:

```
%execute_notebook <relative-file-path>
```

Para especificar um ID de cluster e um perfil de serviço para uma execução descentralizada, use o seguinte comando:

```
%execute_notebook <notebook_name>.ipynb --cluster-id <emr-cluster-id> --service-role
 <emr-notebook-service-role>
```

Quando a Amazon EMR e a Amazon EC2 compartilham uma função de instância, a função exige as seguintes permissões adicionais:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
```

```
                "elasticmapreduce:StartNotebookExecution",
                "elasticmapreduce:DescribeNotebookExecution",
                "ec2:DescribeInstances"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": "arn:aws:iam::<AccoundId>:role/EMR_Notebooks_DefaultRole"
        }
    ]
}
```

> **Note**
>
> Para usar a magic do `%execute_notebook`, instale a versão 0.2.3 ou superior do pacote `emr-notebooks-magics`.

Use cadernos em várias linguagens com kernels do Spark

Cada kernel de caderno Jupyter tem uma linguagem padrão. Por exemplo, o idioma padrão do kernel do Spark é o Scala e o idioma padrão do PySpark kernel é o Python. Com o Amazon EMR 6.4.0 e versões posteriores, o EMR Studio oferece suporte a notebooks em vários idiomas. Isso significa que cada kernel no EMR Studio pode oferecer suporte às seguintes linguagens, além da linguagem padrão: Python, Spark, R e Spark. SQL

Para ativar este atributo, especifique um dos comandos magic apresentados a seguir no início de qualquer célula.

Idioma	Comando
Python	`%%pyspark`
Scala	`%%scalaspark`
R	`%%rspark`

Idioma	Comando
	Não há suporte para cargas de trabalho interativas com o EMR Serverless.
Faísca SQL	`%%sql`

Quando invocados, esses comandos executam a célula inteira na mesma sessão do Spark usando o interpretador da linguagem correspondente.

A `%%pyspark` célula magic permite que os usuários escrevam PySpark código em todos os kernels do Spark.

```
%%pyspark
a = 1
```

A `%%sql` célula magic permite que os usuários executem o SQL código Spark em todos os kernels do Spark.

```
%%sql
SHOW TABLES
```

A célula magic do `%%rspark` permite que os usuários executem código SparkR em todos os kernels do Spark.

```
%%rspark
a <- 1
```

A célula magic do `%%scalaspark` permite que os usuários executem código Spark Scala em todos os kernels do Spark.

```
%%scalaspark
val a = 1
```

Compartilhamento de dados entre interpretadores de linguagens com tabelas temporárias

Você também pode compartilhar dados entre interpretadores de linguagem usando tabelas temporárias. O exemplo a seguir usa `%%pyspark` em uma célula para criar uma tabela temporária

em Python e usa `%%scalaspark` na célula a seguir para realizar a leitura de dados dessa tabela em Scala.

```
%%pyspark
df=spark.sql("SELECT * from nyc_top_trips_report LIMIT 20")
# create a temporary table called nyc_top_trips_report_view in python
df.createOrReplaceTempView("nyc_top_trips_report_view")
```

```
%%scalaspark
// read the temp table in scala
val df=spark.sql("SELECT * from nyc_top_trips_report_view")
df.show(5)
```

Visão geral dos Amazon EMR Notebooks

> **Note**
>
> EMROs notebooks estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar espaços de trabalho, os usuários do EMR Notebooks precisam de permissões adicionais de IAM função. Para obter mais informações, consulte Amazon EMR Notebooks são espaços de trabalho do Amazon EMR Studio no console e no console da Amazon EMR.

Você pode usar o Amazon EMR Notebooks junto com os EMR clusters da Amazon executando o Apache Spark para criar e abrir o Jupyter Notebook e interfaces JupyterLab dentro do console da Amazon. EMR Um EMR notebook é um notebook “sem servidor” que você pode usar para executar consultas e códigos. Ao contrário de um notebook tradicional, o conteúdo de um EMR caderno — equações, consultas, modelos, código e texto narrativo nas células do notebook — é executado em um cliente. Os comandos são executados usando um kernel no EMR cluster. O conteúdo do caderno também é salvo no Amazon S3 separadamente dos dados do cluster para maior durabilidade e reutilização flexível.

Você pode iniciar um cluster, anexar um EMR notebook para análise e, em seguida, encerrar o cluster. Você também pode fechar um bloco de anotações anexado a um cluster em execução e alternar para outro. Diversos usuários podem anexar cadernos ao mesmo cluster simultaneamente e compartilhar arquivos de cadernos no Amazon S3 entre si. Esses recursos permitem executar clusters sob demanda para economizar custos e reduzir o tempo gasto reconfigurando blocos de anotações para diferentes clusters e conjuntos de dados.

Você também pode executar um EMR notebook programaticamente usando a Amazon EMRAPI, sem a necessidade de interagir com o EMR console da Amazon (“execução sem cabeça”). Você precisa incluir uma célula no EMR notebook que tenha uma tag de parâmetros. Essa célula permite que um script transfira novos valores de entrada para o caderno. Cadernos parametrizados podem ser reutilizados com diferentes conjuntos de valores de entrada. Não há necessidade de fazer cópias do mesmo caderno para editar e executar com novos valores de entrada. EMRA Amazon cria e salva o notebook de saída no S3 para cada execução do notebook parametrizado. Para exemplos API de código de EMR notebook, consulteExemplos de comandos para executar EMR Notebooks programaticamente.

> **Important**
>
> O recurso EMR Notebooks oferece suporte a clusters que usam as EMR versões 5.18.0 e superiores da Amazon. Recomendamos que você use EMR Notebooks com clusters que usam a versão mais recente da AmazonEMR, ou pelo menos 5.30.0, 5.32.0 ou 6.2.0. Com essas versões, os kernels do Jupyter são executados no cluster anexado, em vez de em uma instância do Jupyter. Isso melhora a performance e aprimora sua capacidade de personalizar kernels e bibliotecas. Para obter mais informações, consulte Diferenças nas funcionalidades por versão de liberação do cluster.

Aplicam-se taxas aplicáveis para o armazenamento do Amazon S3 e para os EMR clusters da Amazon.

Os Amazon EMR Notebooks estão disponíveis como Amazon EMR Studio Workspaces no console

Fazendo a transição de EMR notebooks para espaços de trabalho

No novo EMR console da Amazon, mesclamos os EMR notebooks com o Amazon EMR Studio Workspaces em uma única experiência. Ao usar um EMR Studio, você pode criar e configurar diferentes espaços de trabalho para organizar e executar notebooks. Se você tinha EMR notebooks Amazon no console antigo, eles estão disponíveis como EMR Studio Workspaces no console.

A Amazon EMR criou esses novos EMR Studio Workspaces para você. O número de estúdios que criamos corresponde ao número de estúdios distintos VPCs que você usa EMR nos Notebooks. Por exemplo, se você se conectar a EMR clusters em dois computadores diferentes VPCs dos EMR Notebooks, criaremos dois novos EMR estúdios. Seus cadernos serão distribuídos entre os novos Studios.

> **Important**
>
> Desativamos a opção de criar novos cadernos no antigo EMR console da Amazon. Em vez disso, use Create Workspace no novo EMR console da Amazon.

Para obter mais informações sobre o Amazon EMR Studio Workspaces, consulteCompreensão das noções básicas do Workspace. Para uma visão geral conceitual do EMR Studio, consulte Workspaces na Como o Amazon EMR Studio funciona página.

O que você precisa fazer?

Embora você ainda possa usar seus notebooks existentes no console antigo, recomendamos que você use o Amazon EMR Studio Workspaces no console. Você deve configurar permissões de função adicionais para ativar os recursos do EMR Studio que não estão disponíveis nos EMR Notebooks.

> **Note**
>
> No mínimo, para visualizar os EMR Notebooks existentes como espaços de trabalho do EMR Studio e criar novos espaços de trabalho, os usuários devem ter `elasticmapreduce:CreateStudioPresignedUrl` permissões `elasticmapreduce:ListStudios` e permissões em suas funções. Para acessar todos os recursos do EMR Studio, consulte Ativando os recursos do EMR Studio para usuários de EMR notebooks a lista completa das permissões adicionais que os usuários do EMR Notebooks precisarão.

Recursos aprimorados no EMR Studio, além dos EMR notebooks

Com o Amazon EMR Studio, você pode configurar e usar os seguintes recursos que não estão disponíveis EMR nos Notebooks:

- Navegue e conecte-se a EMR clusters de dentro do Jupyterlab
- Navegue e EMR conecte-se aos clusters virtuais do Notebooks a partir do Jupyterlab
- Conectar-se aos repositórios Git internamente no JupyterLab
- Colaborar com outros membros da sua equipe para escrever e executar códigos de cadernos
- Procure dados com o SQL Explorer
- Provisionar EMR clusters com o Service Catalog

Para obter uma lista completa dos recursos do Amazon EMR Studio, consultePrincipais recursos do EMR Studio.

Ativando os recursos do EMR Studio para usuários de EMR notebooks

Os novos EMR estúdios que criaremos como parte dessa fusão usam a `EMR_Notebooks_DefaultRole` IAM função existente como função de serviço do EMR Studio.

Os usuários que fazem a transição do EMR Notebooks para o EMR Studio e desejam usar os recursos adicionais do EMR Studio precisam de várias novas permissões de função. Adicione as seguintes permissões às funções dos usuários do EMR Notebooks que planejam usar o EMR Studio.

> **Note**
>
> No mínimo, para visualizar os EMR Notebooks existentes como espaços de trabalho do EMR Studio e criar novos espaços de trabalho, os usuários devem ter `elasticmapreduce:CreateStudioPresignedUrl` permissões `elasticmapreduce:ListStudios` e permissões em suas funções. Para usar todos os recursos do EMR Studio, adicione todas as permissões listadas abaixo. Os usuários administradores também precisam de permissão para criar e gerenciar um EMR Studio. Para obter mais informações, consulte Permissões de administrador para criar e gerenciar um EMR Studio.

```
"elasticmapreduce:DescribeStudio",
"elasticmapreduce:ListStudios",
"elasticmapreduce:CreateStudioPresignedUrl",
"elasticmapreduce:UpdateEditor",
"elasticmapreduce:PutWorkspaceAccess",
"elasticmapreduce:DeleteWorkspaceAccess",
"elasticmapreduce:ListWorkspaceAccessIdentities",
"emr-containers:ListVirtualClusters",
"emr-containers:DescribeVirtualCluster",
"emr-containers:ListManagedEndpoints",
"emr-containers:DescribeManagedEndpoint",
"emr-containers:CreateAccessTokenForManagedEndpoint",
"emr-containers:ListJobRuns",
"emr-containers:DescribeJobRun",
"servicecatalog:SearchProducts",
"servicecatalog:DescribeProduct",
"servicecatalog:DescribeProductView",
"servicecatalog:DescribeProvisioningParameters",
"servicecatalog:ProvisionProduct",
"servicecatalog:UpdateProvisionedProduct",
```

```
"servicecatalog:ListProvisioningArtifacts",
"servicecatalog:DescribeRecord",
"servicecatalog:ListLaunchPaths",
"cloudformation:DescribeStackResources"
```

As permissões a seguir também são necessárias para usar os recursos de colaboração no EMR Studio, mas não eram necessárias com os EMR Notebooks.

```
"sso-directory:SearchUsers",
"iam:GetUser",
"iam:GetRole",
"iam:ListUsers",
"iam:ListRoles",
"sso:GetManagedApplicationInstance"
```

Considerações ao usar notebooks EMR

> **Note**
>
> EMROs notebooks estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar espaços de trabalho, os usuários do EMR Notebooks precisam de permissões adicionais de IAM função. Para obter mais informações, consulte Amazon EMR Notebooks são espaços de trabalho do Amazon EMR Studio no console e no console da Amazon EMR.

Considere os seguintes requisitos ao criar clusters e desenvolver soluções usando o EMR notebook.

Requisitos de cluster

- Habilite o Amazon EMR Block Public Access — O acesso de entrada a um cluster permite que os usuários do cluster executem kernels de notebooks. Garanta que somente usuários autorizados possam acessar o cluster. É altamente recomendável que você deixe o bloqueio de acesso público ativado e limite o SSH tráfego de entrada somente a fontes confiáveis. Para ter mais informações, consulte Usando a Amazon, EMR bloqueie o acesso público e Controle do tráfego de rede com grupos de segurança.
- Use um cluster compatível: um cluster conectado a um caderno deve atender aos seguintes requisitos:

- Somente clusters criados usando a Amazon EMR são compatíveis. Você pode criar um cluster de forma independente dentro da Amazon EMR e depois anexar um EMR notebook, ou você pode criar um cluster compatível ao criar um EMR notebook.
- Somente clusters criados usando a EMR versão 5.18.0 e posterior da Amazon são compatíveis. Consulte the section called “Diferenças nas funcionalidades por versão de liberação do cluster”.
- Clusters criados usando EC2 instâncias da Amazon com AMD EPYC processadores — por exemplo, tipos de instância m5a.* e r5a.* — não são compatíveis.
- EMROs notebooks funcionam somente com clusters criados com `VisibleToAllUsers` set to`true`. `VisibleToAllUsers`é `true` por padrão.
- O cluster deve ser lançado dentro de um EC2 -VPC. Sub-redes públicas e privadas têm suporte. A plataforma EC2 -Classic não é suportada.
- O cluster deve ser iniciado com o Hadoop, Spark e Livy instalados. Outros aplicativos podem ser instalados, mas os EMR Notebooks atualmente oferecem suporte somente aos clusters Spark.

  > **Important**
  >
  > Para as EMR versões 5.32.0 e posteriores da Amazon, ou 6.2.0 e posteriores, seu cluster também deve estar executando o aplicativo Jupyter Enterprise Gateway para funcionar com notebooks. EMR

- Clusters que usam a autenticação do Kerberos não são compatíveis.
- Clusters integrados AWS Lake Formation oferecem suporte somente à instalação de bibliotecas com escopo de notebook. A instalação de kernels e bibliotecas no cluster não é permitida.
- Clusters com vários nós primários não são compatíveis.
- Não há suporte para clusters que usam EC2 instâncias da Amazon com base no AWS Graviton2.

Diferenças nas funcionalidades por versão de liberação do cluster

É altamente recomendável que você use EMR Notebooks com clusters criados usando as versões 5.30.0, 5.32.0 ou posterior da Amazon EMR ou 6.2.0 ou posterior. Com essas versões, o EMR Notebooks executa kernels no cluster Amazon conectado. EMR Os kernels e as bibliotecas podem ser instalados diretamente no nó primário do cluster. Usar EMR notebooks com essas versões de cluster tem os seguintes benefícios:

- Desempenho aprimorado — os kernels do notebook são executados em clusters com os tipos de EC2 instância selecionados por você. As versões anteriores executam kernels em uma instância especializada que não pode ser redimensionada, acessada ou personalizada.
- Capacidade de adicionar e personalizar kernels: você pode se conectar ao cluster para instalar pacotes de kernel usando `conda` e `pip`. Além disso, a instalação de `pip` é compatível com o uso de comandos de terminal dentro de células do bloco de anotações. Nas versões anteriores, somente kernels pré-instalados estavam disponíveis (Python, PySpark Spark e SparkR). Para obter mais informações, consulte Instalação de kernels e de bibliotecas Python em um nó primário do cluster.
- Capacidade de instalar bibliotecas Python: você pode instalar bibliotecas Python no nó primário do cluster usando `conda` e `pip`. Recomendamos usar `conda`. Nas versões anteriores, somente bibliotecas com escopo de notebook são suportadas. PySpark

Recursos de EMR notebooks compatíveis por versão de cluster

Versão do cluster	Bibliotecas com escopo de notebooks para PySpark	Instalação do kernel no cluster	Instalação da bibliotec a Python no nó primário
Antes da versão 5.18.0	EMRNotebooks não suportados		
5.18.0 a 5.25.0	Não	Não	Não
5.26.0 a 5.29.0	Sim	Não	Não
5.30.0	Sim	Sim	Sim
6.0.0	Não	Não	Não
5.32.0 e posteriores e 6.2.0 e posteriores	Sim	Sim	Sim

Limites para notebooks conectados EMR simultaneamente

Ao criar um cluster compatível com notebooks, considere o tipo de EC2 instância do nó primário do cluster. As restrições de memória dessa EC2 instância determinam o número de notebooks que podem estar prontos simultaneamente para executar códigos e consultas no cluster.

Tipo de EC2 instância do nó primário	Número de EMR notebooks
*.medium	2
*.large	4
*.xlarge	8
*.2xlarge	16
*.4xlarge	24
*.8xlarge	24
*.16xlarge	24

Versões do caderno Jupyter e Python

EMRO Notebooks executa o Jupyter Notebook versão 6.0.2 e o Python 3.6.5, independentemente da versão de lançamento da Amazon EMR do cluster conectado.

Considerações sobre segurança

Usar locais criptografados do S3

> Se você especificar um local criptografado no Amazon S3 para armazenar arquivos de cadernos, deverá configurar o Função de serviço para EMR notebooks como usuário da chave. A função de serviço padrão é `EMR_Notebooks_DefaultRole`. Se você estiver usando uma AWS KMS chave para criptografia, consulte Como usar políticas de chaves AWS KMS no Guia do AWS Key Management Service desenvolvedor e no artigo de suporte para adicionar usuários de chaves.

Uso de cookies com domínios de hospedagem

Para aumentar a segurança dos aplicativos fora do console que você pode usar com a AmazonEMR, os domínios de hospedagem de aplicativos são registrados na Lista Pública de Sufixos (). PSL Exemplos desses domínios de hospedagem incluem os seguintes: `emrstudio-prod.us-east-1.amazonaws.com`, `emrnotebooks-prod.us-east-1.amazonaws.com`, `emrappui-prod.us-east-1.amazonaws.com`. Para maior segurança, se precisar definir cookies confidenciais no nome de domínio padrão, recomendamos que você use cookies com um prefixo `__Host-`. Isso ajuda a defender seu domínio contra tentativas de falsificação de solicitações entre sites ()CSRF. Para obter mais informações, consulte a página Set-Cookie em Mozilla Developer Network.

Criação de um bloco de anotações

> **Note**
>
> EMROs notebooks estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar espaços de trabalho, os usuários do EMR Notebooks precisam de permissões adicionais de IAM função. Para obter mais informações, consulte Amazon EMR Notebooks são espaços de trabalho do Amazon EMR Studio no console e no console da Amazon EMR.

Você cria um EMR notebook usando o antigo EMR console da Amazon. A criação de notebooks usando o AWS CLI ou a Amazon não EMR API é suportada.

Para criar um EMR caderno

1. Abra o EMR console da Amazon em https://console.aws.amazon.com/elasticmapreduce/.
2. Escolha Notebooks (Blocos de anotações, Create notebook (Criar bloco de anotações).
3. Insira um Notebook name (Nome do bloco de anotações) e uma Notebook description (Descrição do bloco de anotações) adicional.
4. Se você tiver um cluster ativo ao qual deseja anexar o caderno, deixe o padrão Escolher um cluster existente selecionado, clique em Escolher, selecione um cluster na lista e, em seguida, clique em Escolher cluster. Para obter informações sobre os requisitos de cluster para EMR notebooks, consulteConsiderações ao usar notebooks EMR.

—ou—

Escolha Criar um cluster, insira um Nome de cluster e escolha opções de acordo com as diretrizes a seguir. O cluster é criado no padrão VPC para a conta usando instâncias sob demanda.

Configuração	Descrição
Nome do cluster	O nome amigável usado para identificar o cluster.
Versão	Não pode ser modificado. O padrão é a versão de EMR lançamento mais recente da Amazon (5.36.2).
Aplicativos	Não pode ser modificado. Lista os aplicativos instalados no cluster.
Instância	Insira o número de instâncias e selecione o tipo de EC2 instância. Uma instância é usada para o nó primário. O resto é usado para nós core. O tipo de instância determina o número de blocos de anotações que podem ser anexados ao cluster simultaneamente. Para obter mais informações, consulte Limites para notebooks conectados EMR simultaneamente.
EMRpapel	Deixe o padrão ou escolha o link para especificar uma função de serviço personalizada para a AmazonEMR. Para obter mais informações, consulte Função de serviço para a Amazon EMR (EMRfunção).

Configuração	Descrição
EC2perfil de instância	Deixe o padrão ou escolha o link para especificar uma função de serviço personalizada para EC2 as instâncias. Para obter mais informações, consulte Função de serviço para EC2 instâncias de cluster (perfil de EC2 instância).
EC2key pair	Escolha um EC2 key pair para poder se conectar às instâncias do cluster. Para obter mais informações, consulte Conecte-se ao nó primário usando SSH.
Encerramento automático	A terminação automática é compatível com as EMR versões 5.30.0 e 6.1.0 e posteriores da Amazon. Marque a caixa de seleção para habilitar o encerramento automático e, em seguida, especifique o tempo de inatividade após o qual o cluster deverá ser desligado automaticamente. Para obter mais informações, consulte Usar uma política de término automático.

5. Em Security groups (Grupos de segurança), escolha Use default security groups (Usar grupos de segurança padrão). Como alternativa, escolha Escolher grupos de segurança e selecione grupos VPC de segurança personalizados que estejam disponíveis no cluster. Selecione um grupo para a instância primária e outro para a instância do cliente do caderno. Para obter mais informações, consulte the section called “Grupos de segurança para EMR notebooks”.
6. Em Perfil de serviço da AWS , deixe o padrão ou escolha um perfil personalizado na lista. A instância do cliente do bloco de anotações usa essa função. Para obter mais informações, consulte Função de serviço para EMR notebooks.
7. Em Local do caderno, escolha o local no Amazon S3 no qual o arquivo de caderno será salvo ou especifique seu próprio local. Se o bucket e a pasta não existirem, a Amazon os EMR criará.

EMRA Amazon cria uma pasta com o ID do Notebook como nome da pasta e salva o notebook em um arquivo chamado*NotebookName*.ipynb. Por exemplo, se você especificar o local do Amazon S3 s3://MyBucket/MyNotebooks para um caderno chamado MyFirstEMRManagedNotebook, o arquivo de caderno será salvo em s3://MyBucket/MyNotebooks/*NotebookID*/MyFirstEMRManagedNotebook.ipynb.

Se você especificar um local criptografado no Amazon S3, deverá configurar o Função de serviço para EMR notebooks como um usuário da chave. A função de serviço padrão é EMR_Notebooks_DefaultRole. Se você estiver usando uma AWS KMS chave para criptografia, consulte Como usar políticas de chaves AWS KMS no Guia do AWS Key Management Service desenvolvedor e no artigo de suporte para adicionar usuários de chaves.

8. Opcionalmente, se você adicionou um repositório baseado em Git na EMR Amazon que deseja associar a esse notebook, escolha repositório Git, selecione Escolher repositório e selecione um repositório na lista. Para obter mais informações, consulte Associando repositórios baseados em Git a notebooks EMR.
9. Opcionalmente, selecione Tags e, em seguida, adicione as tags de chave-valor adicionais para o bloco de anotações.

> **Important**
>
> Uma tag padrão com a string Key definida como creatorUserID e o valor definido como seu ID de IAM usuário é aplicada para fins de acesso. Recomendamos que você não altere nem remova essa tag, pois ela pode ser usada para controlar o acesso. Para obter mais informações, consulte Use tags de cluster e notebook com IAM políticas para controle de acesso.

10. Selecione Criar bloco de anotações.

Trabalhando com EMR notebooks

> **Note**
>
> EMROs notebooks estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar espaços de trabalho, os usuários do EMR Notebooks precisam de permissões

adicionais de IAM função. Para obter mais informações, consulte Amazon EMR Notebooks são espaços de trabalho do Amazon EMR Studio no console e no console da Amazon EMR.

Depois de criar um EMR caderno, o notebook demora um pouco para ser iniciado. O Status na lista Notebooks (Blocos de anotações) mostra Starting (Iniciando). Você pode abrir um bloco de anotações quando seu status for Ready (Pronto). Pode demorar um pouco mais para um bloco de anotações entrar no status Ready (Pronto) se você tiver criado um cluster com ele.

> **Tip**
>
> Atualize o navegador ou escolha o ícone de atualização acima da lista de blocos de anotações para atualizar o status do bloco de anotações.

Noções básicas sobre o status do caderno

Um EMR caderno de anotações pode ter o seguinte para Status na lista de cadernos.

Status	Significado
Ready	Você pode abrir o bloco de anotações usando o editor de blocos de anotações. Enquanto um bloco de anotações estiver no status Ready (Pronto), você poderá interrompê-lo ou excluí-lo. Para alterar clusters, você deve interromper o bloco de anotações primeiro. Se um bloco de anotações no status Ready (Pronto) ficar ocioso por muito tempo, ele será interrompido automaticamente.
Starting	O bloco de anotações está sendo criado e conectado ao cluster. Enquanto um bloco de anotações estiver sendo iniciado, você não poderá abrir o editor de blocos de anotações, interromper, excluir nem alterar clusters.

Status	Significado
Pendente	O bloco de anotações foi criado e está aguardando a integração com o cluster para ser concluído. O cluster ainda pode estar provisionamento recursos ou respondendo a outras solicitações. Você pode abrir o editor de blocos de anotações com o bloco de anotações no modo local. Qualquer código que se baseie em processos de cluster não será executado e falhará.
Parando	O bloco de anotações está sendo desligado ou o cluster ao qual o bloco de anotações está sendo anexado está sendo encerrado. Enquanto um bloco de anotações estiver sendo interrompido, você não poderá abrir o editor de blocos de anotações, interromper, excluir nem alterar clusters.
Interrompido	O bloco de anotações foi encerrado. Você pode iniciar o bloco de anotações no mesmo cluster, desde que o cluster ainda esteja em execução. Você pode alterar os clusters e excluir o cluster.
Excluindo	O cluster é removido da lista de clusters disponíveis. O arquivo de caderno *NotebookName* `.ipynb` permanece no Amazon S3 e continua acumulando as cobranças de armazenamento aplicáveis.

Como trabalhar com o editor de cadernos

Uma vantagem de usar um EMR notebook é que você pode iniciar o notebook no Jupyter ou JupyterLab diretamente do console.

Com o EMR Notebooks, o editor de notebook que você acessa do EMR console da Amazon é o conhecido editor de notebook Jupyter de código aberto ou. JupyterLab Como o editor do notebook é lançado no EMR console da Amazon, é mais eficiente configurar o acesso do que com um notebook hospedado em um EMR cluster da Amazon. Você não precisa configurar o cliente de um usuário para ter acesso à Web por meio SSH de regras de grupo de segurança e configurações de proxy. Se um usuário tiver permissões suficientes, ele pode simplesmente abrir o editor do notebook no EMR console da Amazon.

Somente um usuário pode ter um EMR notebook aberto por vez na AmazonEMR. Se outro usuário tentar abrir um EMR notebook que já está aberto, ocorrerá um erro.

> **Important**
>
> EMRA Amazon cria um exclusivo pré-assinado URL para cada sessão do editor de notebook, que é válido somente por um curto período de tempo. Recomendamos que você não compartilhe o editor do notebookURL. Fazer isso cria um risco de segurança porque os destinatários do URL adotam suas permissões para editar o notebook e executar o código do notebook durante toda a vida útil doURL. Se outras pessoas precisarem acessar um notebook, forneça permissões a seus usuários por meio de políticas de permissões e garanta que a função de serviço para EMR notebooks tenha acesso à localização do Amazon S3. Para ter mais informações, consulte the section called “Segurança” e Função de serviço para EMR notebooks.

Para abrir o editor de caderno para um EMR notebook

1. Selecione um bloco de anotações com um Status de Ready (Pronto) ou Pending (Pendente) na lista Notebooks (Blocos de anotações).
2. Escolha Abrir no Jupyter JupyterLab ou Abrir no Jupyter.

   Uma nova guia do navegador é aberta para o editor JupyterLab ou o editor do Jupyter Notebook.
3. No menu Kernel, escolha Change kernel (Alterar kernel) e, em seguida, selecione o kernel para sua linguagem de programação.

   Agora você está pronto para gravar e executar código de dentro do editor de blocos de anotações.

Como salvar o conteúdo de um caderno

Ao trabalhar no editor de cadernos, o conteúdo das células e as saídas do caderno são salvos automaticamente no arquivo de caderno no Amazon S3, de forma periódica. Um bloco de anotações que não tem alterações desde a última vez em que uma célula foi editada mostrará (autosaved) ao lado do nome do bloco de anotações no editor. Se as alterações ainda não foram salvas, unsaved changes (alterações não salvas) será exibido.

Você pode salvar um bloco de anotações manualmente. No menu Arquivo, escolha Salvar e Ponto de verificação ou pressione CTRL +S. Isso cria um arquivo nomeado *NotebookName*`.ipynb` em uma pasta de pontos de verificação dentro da pasta do notebook no Amazon S3. Por exemplo, `s3://`*MyBucket*`/`*MyNotebookFolder*`/`*NotebookID*`/checkpoints/`*NotebookName*`.ipynb`. Somente o arquivo de pontos de verificação mais recente é salvo nesse local.

Como alterar clusters

Você pode alterar o cluster ao qual um EMR notebook está conectado sem alterar o conteúdo do próprio notebook. Você pode alterar clusters apenas para os blocos de anotações que têm o status Stopped (Interrompido).

Para alterar o cluster de um EMR notebook

1. Se o bloco de anotações que você deseja alterar estiver em execução, selecione-o na lista Notebooks (Blocos de anotações) e escolha Stop (Interromper).
2. Quando o status do bloco de anotações for Stopped (interrompido), selecione o bloco de anotações na lista Notebooks (Blocos de anotações) e, em seguida, escolha View details (Exibir detalhes).
3. Escolha Change cluster (Alterar cluster).
4. Se você tiver um cluster ativo com o Hadoop, Spark e Livy em execução ao qual você deseje anexar o bloco de anotações, deixe o padrão e selecione um cluster na lista. Somente clusters que atendam aos requisitos listados.

   —ou—

   Selecione Create a cluster (Criar um cluster) e escolha as opções de cluster. Para obter mais informações, consulte Requisitos de cluster.
5. Escolha uma opção para Security groups (Grupos de segurança) e, em seguida, escolha Change cluster and start notebook (Alterar cluster e iniciar bloco de anotações).

Como excluir cadernos e arquivos de cadernos

Ao excluir um EMR notebook usando o EMR console da Amazon, você exclui o notebook da lista de cadernos disponíveis. No entanto, os arquivos de cadernos permanecem no Amazon S3 e continuam a acumular as cobranças de armazenamento.

Para excluir um bloco de anotações e remover arquivos associados

1. Abra o EMR console da Amazon em https://console.aws.amazon.com/elasticmapreduce/.
2. Escolha Notebooks (Blocos de anotações), selecione seu bloco de anotações na lista e, em seguida, escolha View details (Exibir detalhes).
3. Escolha o ícone da pasta ao lado da localização do Notebook e copie o URL, que está no padrão`s3://`*`MyNotebookLocationPath`*`/`*`NotebookID`*`/`.
4. Escolha Excluir.

   O bloco de anotações é removido da lista e os detalhes de bloco de anotações deixam de aparecer.
5. Siga as instruções fornecidas em How do I delete folders from an S3 bucket? no Guia do usuário do Amazon Simple Storage Service. Navegue até o bucket e a pasta na etapa 3.

   —ou—

   Se você tiver o AWS CLI instalado, abra um prompt de comando e digite o comando no final deste parágrafo. Substitua o local do Amazon S3 pelo local que você copiou acima. Certifique-se de que AWS CLI esteja configurado com as chaves de acesso de um usuário com permissões para excluir a localização do Amazon S3. Para obter mais informações, consulte Configuração da AWS CLI no Guia do usuário da AWS Command Line Interface .

   ```
   aws s3 rm s3://MyNotebookLocationPath/NotebookID
   ```

Como compartilhar arquivos de cadernos

Cada EMR notebook é salvo no Amazon S3 como um arquivo chamado. *`NotebookName`*`.ipynb` Desde que o arquivo do notebook seja compatível com a mesma versão do Jupyter Notebook na qual o EMR Notebooks se baseia, você pode abrir o notebook como um notebook. EMR

A maneira mais fácil de abrir um arquivo de notebook de outro usuário é salvar o arquivo*.ipynb de outro usuário no sistema de arquivos local e, em seguida, usar o recurso de upload no Jupyter e nos editores. JupyterLab

Você pode usar esse processo para usar EMR cadernos compartilhados por outras pessoas, cadernos compartilhados na comunidade Jupyter ou para restaurar um notebook que foi excluído do console quando você ainda tem o arquivo do notebook.

Para usar um arquivo de notebook diferente como base para um EMR notebook

1. Antes de continuar, feche o editor do notebook de todos os notebooks com os quais você trabalhará e, em seguida, interrompa o notebook se for um EMR notebook.
2. Crie um EMR caderno e insira um nome para ele. O nome que você inserir para o bloco de anotações será o nome do arquivo que você precisará substituir. O novo nome de arquivo deve corresponder exatamente ao nome desse arquivo.
3. Anote o local no Amazon S3 que você escolheu para o caderno. O arquivo que você substituir está em uma pasta com um caminho e nome de arquivo como o padrão a seguir: `s3://MyNotebookLocation/NotebookID/MyNotebookName.ipynb`.
4. Interrompa o bloco de anotações.
5. Substitua o antigo arquivo de caderno no local do Amazon S3 pelo novo, usando exatamente o mesmo nome.

   O AWS CLI comando a seguir para o Amazon S3 substitui um arquivo salvo em uma máquina local chamada `SharedNotebook.ipynb` para um EMR notebook com o nome MyNotebook, uma ID de `e-12A3BCDEFJHIJKLMNO45PQRST` e criado com o `MyBucket/MyNotebooksFolder` especificado no Amazon S3. Para obter informações sobre como usar o console do Amazon S3 para copiar e substituir arquivos, consulte Fazer upload, fazer download e trabalhar com objetos no Guia do usuário do Amazon Simple Storage Service.

   ```
   aws s3 cp SharedNotebook.ipynb s3://MyBucket/
   MyNotebooksFolder/-12A3BCDEFJHIJKLMNO45PQRST/MyNotebook.ipynb
   ```

Exemplos de comandos para executar EMR Notebooks programaticamente

> **Note**
>
> EMROs notebooks estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar espaços de trabalho, os usuários do EMR Notebooks precisam de permissões adicionais de IAM função. Para obter mais informações, consulte Amazon EMR Notebooks são espaços de trabalho do Amazon EMR Studio no console e no console da Amazon EMR.

Visão geral

Você pode executar EMR notebooks com execução a APIs partir de um script ou da linha de comando. Quando você inicia, interrompe, lista e descreve as execuções do EMR notebook fora do AWS console, você pode controlar programaticamente um notebook. EMR É possível transferir valores de parâmetros diferentes para um caderno com uma célula de caderno parametrizada. Isto elimina a necessidade de criar uma cópia do caderno para cada novo conjunto de valores de parâmetros. Para obter mais informações, consulte EMRAPIAções da Amazon.

Você pode agendar ou agrupar execuções de EMR cadernos com CloudWatch eventos da Amazon e. AWS Lambda Para obter mais informações, consulte Usando AWS Lambda com Amazon CloudWatch Events.

Permissões de perfil para a execução programática

Para usar a execução programática com EMR Notebooks, você deve configurar as permissões do usuário com as seguintes políticas:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowExecutionActions",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:StartNotebookExecution",
```

```
                "elasticmapreduce:DescribeNotebookExecution",
                "elasticmapreduce:ListNotebookExecutions"
            ],
            "Resource": "*"
        },
        {
            "Sid": "AllowPassingServiceRole",
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": "arn:aws:iam::account-id:role/EMR_Notebooks_DefaultRole"
        }
    ]
}
```

Ao executar programaticamente EMR Notebooks em um cluster de EMR Notebooks, você deve adicionar estas permissões adicionais:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowRetrievingManagedEndpointCredentials",
            "Effect": "Allow",
            "Action": [
                "emr-containers:GetManagedEndpointSessionCredentials"
            ],
            "Resource": [
                "arn:aws:emr-containers:region:account-id:/virtualclusters/virtual-
cluster-id/endpoints/managed-endpoint-id"
            ],
            "Condition": {
                "StringEquals": {
                    "emr-containers:ExecutionRoleArn": [
                        "arn:aws:iam::account-id:role/emr-on-eks-execution-role"
                    ]
                }
            }
        },
        {
            "Sid": "AllowDescribingManagedEndpoint",
            "Effect": "Allow",
```

```
            "Action": [
                "emr-containers:DescribeManagedEndpoint"
            ],
            "Resource": [
                "arn:aws:emr-containers:region:account-id:/virtualclusters/virtual-
cluster-id/endpoints/managed-endpoint-id"
            ]
        }
    ]
}
```

Limitações da execução programática

- Há suporte para um máximo de 100 execuções simultâneas Região da AWS por conta.
- Uma execução será encerrada se for executada por mais de 30 dias.
- A execução programática de notebooks não é compatível com os aplicativos interativos Amazon EMR Serverless.

Exemplos de execução programática de EMR cadernos

As seções a seguir fornecem vários exemplos de execução programática de EMR notebooks com o AWS CLI, Boto3 (SDKPython) e Ruby:

- Exemplos de CLI comandos de execução do notebook
- Exemplos de Python de execução de cadernos
- Exemplos de Ruby de execução de cadernos

Você também pode executar notebooks parametrizados como parte dos fluxos de trabalho programados com uma ferramenta de orquestração, como o Apache Airflow ou o Amazon Managed Workflows for Apache Airflow (). MWAA Para obter mais informações, consulte Orquestrando trabalhos de análise em EMR notebooks usando MWAA o AWS Big Data Blog.

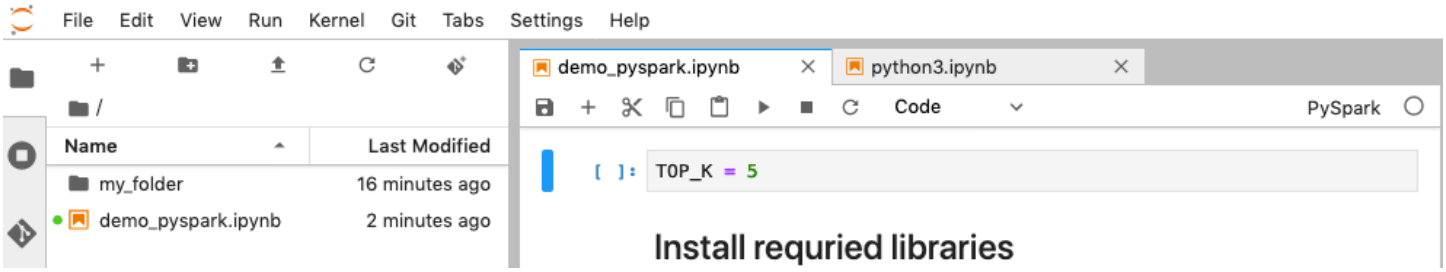
Exemplos de CLI comandos de execução do notebook

> **Note**
>
> EMROs notebooks estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar

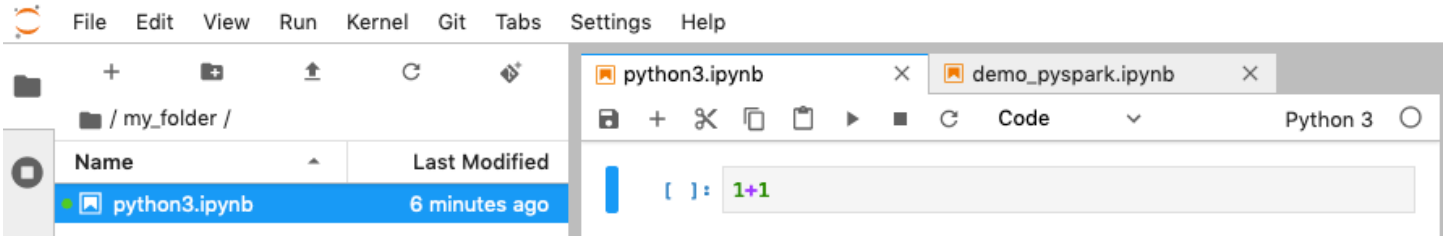
ou criar espaços de trabalho, os usuários do EMR Notebooks precisam de permissões adicionais de IAM função. Para obter mais informações, consulte Amazon EMR Notebooks são espaços de trabalho do Amazon EMR Studio no console e no console da Amazon EMR.

O exemplo a seguir usa o notebook de demonstração do console EMR Notebooks. Para localizar o caderno, use o caminho do arquivo relativo ao diretório inicial. Neste exemplo, há dois arquivos de cadernos que você pode executar: `demo_pyspark.ipynb` e `my_folder/python3.ipynb`.

O caminho relativo para o arquivo `demo_pyspark.ipynb` é `demo_pyspark.ipynb`, como apresentado abaixo.



O caminho relativo para `python3.ipynb` é `my_folder/python3.ipynb`, como apresentado abaixo.



Para obter informações sobre as EMR API `NotebookExecution` ações da Amazon, consulte EMRAPIAções da Amazon. .

Execução de um caderno

Você pode usar o AWS CLI para executar seu notebook com a `start-notebook-execution` ação, conforme demonstrado nos exemplos a seguir.

Example — Executando um EMR notebook em um EMR Studio Workspace com um cluster Amazon EMR (rodando na AmazonEC2)

```
aws emr --region us-east-1 \
start-notebook-execution \
```

```
--editor-id e-ABCDEFG123456 \
--notebook-params '{"input_param":"my-value", "good_superhero":["superman", "batman"]}'
 \
--relative-path test.ipynb \
--notebook-execution-name my-execution \
--execution-engine '{"Id" : "j-1234ABCD123"}' \
--service-role EMR_Notebooks_DefaultRole

{
    "NotebookExecutionId": "ex-ABCDEFGHIJ1234ABCD"
}
```

Example — Executando um EMR notebook em um EMR Studio Workspace com um EMR cluster de notebooks

```
aws emr start-notebook-execution \
    --region us-east-1 \
    --service-role EMR_Notebooks_DefaultRole \
    --environment-variables '{"KERNEL_EXTRA_SPARK_OPTS": "--conf
 spark.executor.instances=1", "KERNEL_LAUNCH_TIMEOUT": "350"}' \
    --output-notebook-format HTML \
    --execution-engine Id=arn:aws:emr-containers:us-west-2:account-id:/
virtualclusters/ABCDEFG/
endpoints/ABCDEF,Type=EMR_ON_EKS,ExecutionRoleArn=arn:aws:iam::account-
id:role/execution-role \
    --editor-id e-ABCDEFG \
    --relative-path EMRonEKS-spark_python.ipynb
```

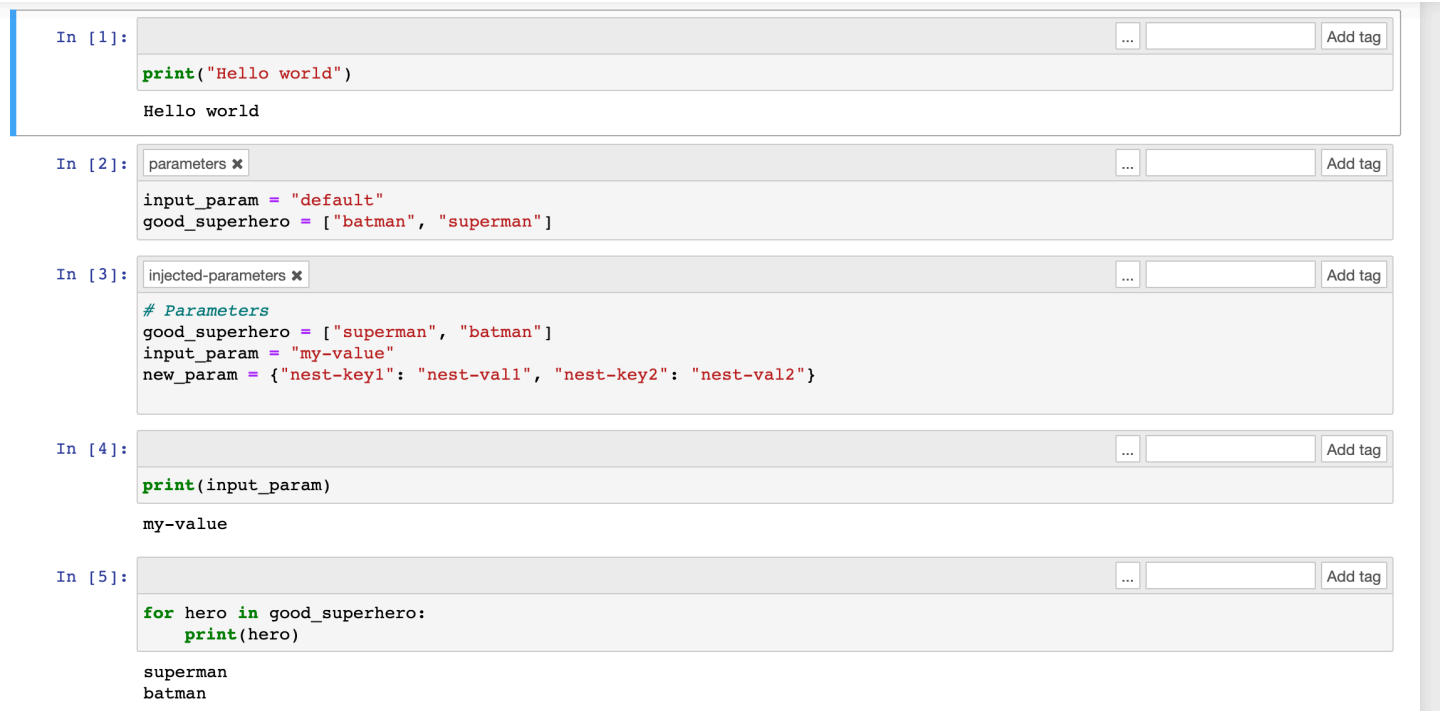
Example — Executando um EMR notebook especificando sua localização no Amazon S3

```
aws emr start-notebook-execution \
    --region us-east-1 \
    --notebook-execution-name my-execution-on-emr-on-eks-cluster \
    --service-role EMR_Notebooks_DefaultRole \
    --environment-variables '{"KERNEL_EXTRA_SPARK_OPTS": "--conf
 spark.executor.instances=1", "KERNEL_LAUNCH_TIMEOUT": "350"}' \
    --output-notebook-format HTML \
    --execution-engine Id=arn:aws:emr-containers:us-west-2:account-id:/
virtualclusters/ABCDEF/
endpoints/ABCDEF,Type=EMR_ON_EKS,ExecutionRoleArn=arn:aws:iam::account-
id:role/execution-role \
    --notebook-s3-location '{"Bucket": "your-s3-bucket","Key": "s3-prefix-to-notebook-
location/EMRonEKS-spark_python.ipynb"}' \
```

```
    --output-notebook-s3-location '{"Bucket": "your-s3-bucket","Key": "s3-prefix-for-storing-output-notebook"}'
```

Saída de bloco de anotações

Confira o resultado de um caderno de exemplo. A célula 3 mostra os valores dos parâmetros injetados recentemente.



Descrição de um caderno

Você pode usar a ação `describe-notebook-execution` para acessar informações sobre a execução de um caderno específico.

```
aws emr --region us-east-1 \
describe-notebook-execution --notebook-execution-id ex-IZWZZVR9DKQ9WQ7VZWXJZR29UGHTE

{
    "NotebookExecution": {
        "NotebookExecutionId": "ex-IZWZZVR9DKQ9WQ7VZWXJZR29UGHTE",
        "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
        "ExecutionEngine": {
            "Id": "j-2QMOV6JAX1TS2",
            "Type": "EMR",
            "MasterInstanceSecurityGroupId": "sg-05ce12e58cd4f715e"
```

```
        },
        "NotebookExecutionName": "my-execution",
        "NotebookParams": "{\"input_param\":\"my-value\", \"good_superhero\":
[\"superman\", \"batman\"]}",
        "Status": "FINISHED",
        "StartTime": 1593490857.009,
        "Arn": "arn:aws:elasticmapreduce:us-east-1:123456789012:notebook-execution/ex-
IZWZZVR9DKQ9WQ7VZWXJZR29UGHTE",
        "LastStateChangeReason": "Execution is finished for cluster j-2QMOV6JAX1TS2.",
        "NotebookInstanceSecurityGroupId": "sg-0683b0a39966d4a6a",
        "Tags": []
    }
}
```

Interrupção de um caderno

Se o seu caderno estiver executando uma execução que você gostaria de interromper, poderá fazer isso com o comando `stop-notebook-execution`.

```
# stop a running execution
aws emr --region us-east-1 \
stop-notebook-execution --notebook-execution-id ex-IZWZX78UVPAATC8LHJR129B1RBN4T


# describe it
aws emr --region us-east-1 \
describe-notebook-execution --notebook-execution-id ex-IZWZX78UVPAATC8LHJR129B1RBN4T

{
    "NotebookExecution": {
        "NotebookExecutionId": "ex-IZWZX78UVPAATC8LHJR129B1RBN4T",
        "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
        "ExecutionEngine": {
            "Id": "j-2QMOV6JAX1TS2",
            "Type": "EMR"
        },
        "NotebookExecutionName": "my-execution",
        "NotebookParams": "{\"input_param\":\"my-value\", \"good_superhero\":
[\"superman\", \"batman\"]}",
        "Status": "STOPPED",
        "StartTime": 1593490876.241,
        "Arn": "arn:aws:elasticmapreduce:us-east-1:123456789012:editor-execution/ex-
IZWZX78UVPAATC8LHJR129B1RBN4T",
```

```
        "LastStateChangeReason": "Execution is stopped for cluster j-2QMOV6JAX1TS2.
 Internal error",
        "Tags": []
    }
}
```

Listagem das execuções de um caderno por horário de início

Você pode transferir um parâmetro `--from` para `list-notebook-executions` com a finalidade de listar as execuções do caderno por horário de início.

```
# filter by start time
aws emr --region us-east-1 \
list-notebook-executions --from 1593400000.000

{
    "NotebookExecutions": [
        {
            "NotebookExecutionId": "ex-IZWZX78UVPAATC8LHJR129B1RBN4T",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "STOPPED",
            "StartTime": 1593490876.241
        },
        {
            "NotebookExecutionId": "ex-IZWZZVR9DKQ9WQ7VZWXJZR29UGHTE",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "RUNNING",
            "StartTime": 1593490857.009
        },
        {
            "NotebookExecutionId": "ex-IZWZYRS0M14L5V95WZ9OQ399SKMNW",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "STOPPED",
            "StartTime": 1593490292.995
        },
        {
            "NotebookExecutionId": "ex-IZX009ZK83IVY5E33VH8MDMELVK8K",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "FINISHED",
```

```
            "StartTime": 1593489834.765
        },
        {
            "NotebookExecutionId": "ex-IZWZXOZF88JWDF9J09GJ91R57VI0N",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "FAILED",
            "StartTime": 1593488934.688
        }
    ]
}
```

Listagem das execuções de um caderno por horário de início e status

O comando `list-notebook-executions` também pode usar um parâmetro `--status` para filtrar os resultados.

```
# filter by start time and status
aws emr --region us-east-1 \
list-notebook-executions --from 1593400000.000 --status FINISHED
{
    "NotebookExecutions": [
        {
            "NotebookExecutionId": "ex-IZWZZVR9DKQ9WQ7VZWXJZR29UGHTE",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "FINISHED",
            "StartTime": 1593490857.009
        },
        {
            "NotebookExecutionId": "ex-IZX009ZK83IVY5E33VH8MDMELVK8K",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "FINISHED",
            "StartTime": 1593489834.765
        }
    ]
}
```

Exemplos de Python de execução de cadernos

> **Note**
>
> EMROs notebooks estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar espaços de trabalho, os usuários do EMR Notebooks precisam de permissões adicionais de IAM função. Para obter mais informações, consulte Amazon EMR Notebooks são espaços de trabalho do Amazon EMR Studio no console e no console da Amazon EMR.

O exemplo de código a seguir é um arquivo SDK para Python (Boto3) chamado `demo.py` que mostra a execução do notebook. APIs

Para obter informações sobre as EMR API `NotebookExecution` ações da Amazon, consulte EMRAPIAções da Amazon.

```
import boto3,time

emr = boto3.client(
    'emr',
    region_name='us-west-1'
)

start_resp = emr.start_notebook_execution(
    EditorId='e-40AC8ZO6EGGCPJ4DLO48KGGGI',
    RelativePath='boto3_demo.ipynb',
    ExecutionEngine={'Id':'j-1HYZS6JQKV11Q'},
    ServiceRole='EMR_Notebooks_DefaultRole'
)

execution_id = start_resp["NotebookExecutionId"]
print(execution_id)
print("\n")

describe_response = emr.describe_notebook_execution(NotebookExecutionId=execution_id)

print(describe_response)
print("\n")

list_response = emr.list_notebook_executions()
```

```
print("Existing notebook executions:\n")
for execution in list_response['NotebookExecutions']:
    print(execution)
    print("\n")

print("Sleeping for 5 sec...")
time.sleep(5)

print("Stop execution " + execution_id)
emr.stop_notebook_execution(NotebookExecutionId=execution_id)
describe_response = emr.describe_notebook_execution(NotebookExecutionId=execution_id)
print(describe_response)
print("\n")
```

Confira o resultado da execução do arquivo `demo.py`.

```
ex-IZX56YJDW1D29Q1PHR32WABU2SAPK

{'NotebookExecution': {'NotebookExecutionId': 'ex-IZX56YJDW1D29Q1PHR32WABU2SAPK',
 'EditorId': 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'ExecutionEngine': {'Id':
 'j-1HYZS6JQKV11Q', 'Type': 'EMR'}, 'NotebookExecutionName': '', 'Status': 'STARTING',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 49, 19, 418000, tzinfo=tzlocal()),
 'Arn': 'arn:aws:elasticmapreduce:us-west-1:123456789012:notebook-execution/ex-
IZX56YJDW1D29Q1PHR32WABU2SAPK', 'LastStateChangeReason': 'Execution is starting
 for cluster j-1HYZS6JQKV11Q.', 'Tags': []}, 'ResponseMetadata': {'RequestId':
 '70f12c5f-1dda-45b7-adf6-964987d373b7', 'HTTPStatusCode': 200, 'HTTPHeaders': {'x-
amzn-requestid': '70f12c5f-1dda-45b7-adf6-964987d373b7', 'content-type': 'application/
x-amz-json-1.1', 'content-length': '448', 'date': 'Wed, 19 Aug 2020 00:49:22 GMT'},
 'RetryAttempts': 0}}

Existing notebook executions:

{'NotebookExecutionId': 'ex-IZX56YJDW1D29Q1PHR32WABU2SAPK', 'EditorId':
 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'NotebookExecutionName': '', 'Status': 'STARTING',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 49, 19, 418000, tzinfo=tzlocal())}


{'NotebookExecutionId': 'ex-IZX5ABS5PR1E5AHMFYEMX3JJIORRB', 'EditorId':
 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'NotebookExecutionName': '', 'Status': 'RUNNING',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 48, 36, 373000, tzinfo=tzlocal())}
```

```
{'NotebookExecutionId': 'ex-IZX5GLVXIU1HNI8BWVW057F6MF4VE', 'EditorId':
 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'NotebookExecutionName': '', 'Status': 'FINISHED',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 45, 14, 646000, tzinfo=tzlocal()),
 'EndTime': datetime.datetime(2020, 8, 19, 0, 46, 26, 543000, tzinfo=tzlocal())}


{'NotebookExecutionId': 'ex-IZX5CV8YDUO8JAIWMXN2VH32RUIT1', 'EditorId':
 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'NotebookExecutionName': '', 'Status': 'FINISHED',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 43, 5, 807000, tzinfo=tzlocal()),
 'EndTime': datetime.datetime(2020, 8, 19, 0, 44, 31, 632000, tzinfo=tzlocal())}


{'NotebookExecutionId': 'ex-IZX5AS0PPW55CEDEURZ9NSOWSUJZ6', 'EditorId':
 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'NotebookExecutionName': '', 'Status': 'FINISHED',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 42, 29, 265000, tzinfo=tzlocal()),
 'EndTime': datetime.datetime(2020, 8, 19, 0, 43, 48, 320000, tzinfo=tzlocal())}


{'NotebookExecutionId': 'ex-IZX57YF5Q53BKWLR4I5QZ14HJ7DRS', 'EditorId':
 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'NotebookExecutionName': '', 'Status': 'FINISHED',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 38, 37, 81000, tzinfo=tzlocal()),
 'EndTime': datetime.datetime(2020, 8, 19, 0, 40, 39, 646000, tzinfo=tzlocal())}

Sleeping for 5 sec...
Stop execution ex-IZX56YJDW1D29Q1PHR32WABU2SAPK
{'NotebookExecution': {'NotebookExecutionId': 'ex-IZX56YJDW1D29Q1PHR32WABU2SAPK',
 'EditorId': 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'ExecutionEngine': {'Id':
 'j-1HYZS6JQKV11Q', 'Type': 'EMR'}, 'NotebookExecutionName': '', 'Status': 'STOPPING',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 49, 19, 418000, tzinfo=tzlocal()),
 'Arn': 'arn:aws:elasticmapreduce:us-west-1:123456789012:notebook-execution/ex-
IZX56YJDW1D29Q1PHR32WABU2SAPK', 'LastStateChangeReason': 'Execution is being stopped
 for cluster j-1HYZS6JQKV11Q.', 'Tags': []}, 'ResponseMetadata': {'RequestId':
 '2a77ef73-c1c6-467c-a1d1-7204ab2f6a53', 'HTTPStatusCode': 200, 'HTTPHeaders': {'x-
amzn-requestid': '2a77ef73-c1c6-467c-a1d1-7204ab2f6a53', 'content-type': 'application/
x-amz-json-1.1', 'content-length': '453', 'date': 'Wed, 19 Aug 2020 00:49:30 GMT'},
 'RetryAttempts': 0}}
```

Exemplos de Ruby de execução de cadernos

> **Note**
>
> EMROs notebooks estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar

ou criar espaços de trabalho, os usuários do EMR Notebooks precisam de permissões adicionais de IAM função. Para obter mais informações, consulte Amazon EMR Notebooks são espaços de trabalho do Amazon EMR Studio no console e no console da Amazon EMR.

A seguir estão exemplos de código Ruby que demonstram o uso da execução API do notebook.

```
# prepare an Amazon EMR client

emr = Aws::EMR::Client.new(
  region: 'us-east-1',
  access_key_id: 'AKIA...JKPKA',
  secret_access_key: 'rLMeu...vU0OLrAC1',
)
```

Início da execução do caderno e obtenção do ID de execução

Neste exemplo, o editor e o EMR notebook Amazon S3 são. `s3://mybucket/notebooks/e-EA8VGAA429FEQTC8HC9ZHWISK/test.ipynb`

Para obter informações sobre as EMR API `NotebookExecution` ações da Amazon, consulte EMRAPIAções da Amazon.

```
start_response = emr.start_notebook_execution({
    editor_id: "e-EA8VGAA429FEQTC8HC9ZHWISK",
    relative_path: "test.ipynb",

    execution_engine: {id: "j-3U82I95AMALGE"},

    service_role: "EMR_Notebooks_DefaultRole",
})


notebook_execution_id = start_resp.notebook_execution_id
```

Descrição da execução do caderno e impressão dos detalhes

```
describe_resp = emr.describe_notebook_execution({
    notebook_execution_id: notebook_execution_id
})
puts describe_resp.notebook_execution
```

A saída dos comandos definidos acima será semelhante a apresentada a seguir.

```
{
:notebook_execution_id=>"ex-IZX3VTVZWVWPP27KUB90BZ7V9IEDG",
:editor_id=>"e-EA8VGAA429FEQTC8HC9ZHWISK",
:execution_engine=>{:id=>"j-3U82I95AMALGE", :type=>"EMR", :master_instance_security_group_id=>n
:notebook_execution_name=>"",
:notebook_params=>nil,
:status=>"STARTING",
:start_time=>2020-07-23 15:07:07 -0700,
:end_time=>nil,
:arn=>"arn:aws:elasticmapreduce:us-east-1:123456789012:notebook-execution/ex-
IZX3VTVZWVWPP27KUB90BZ7V9IEDG",
:output_notebook_uri=>nil,
:last_state_change_reason=>"Execution is starting for cluster
 j-3U82I95AMALGE.", :notebook_instance_security_group_id=>nil,
:tags=>[]
}
```

Filtros para cadernos

```
"EditorId": "e-XXXX",           [Optional]
"From" : "1593400000.000",    [Optional]
"To" :
```

Interrupção da execução do caderno

```
stop_resp = emr.stop_notebook_execution({
    notebook_execution_id: notebook_execution_id
})
```

Habilitação da representação do usuário para monitorar a atividade de usuários e trabalhos do Spark

> **Note**
>
> EMROs notebooks estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar

ou criar espaços de trabalho, os usuários do EMR Notebooks precisam de permissões adicionais de IAM função. Para obter mais informações, consulte Amazon EMR Notebooks são espaços de trabalho do Amazon EMR Studio no console e no console da Amazon EMR.

EMROs notebooks permitem que você configure a representação do usuário em um cluster do Spark. Esse recurso ajuda a rastrear a atividade do trabalho iniciado no editor de blocos de anotações. Além disso, o EMR Notebooks tem um widget Jupyter Notebook integrado para visualizar os detalhes do trabalho do Spark junto com a saída da consulta no editor do notebook. O widget está disponível por padrão e não requer configuração especial. No entanto, para visualizar os servidores de histórico, seu cliente deve estar configurado para visualizar as interfaces EMR web da Amazon que estão hospedadas no nó primário.

Configuração da representação do usuário do Spark

Por padrão, os trabalhos do Spark que os usuários enviam usando o editor de blocos de anotações parecem se originar de uma identidade de usuário `livy` indiscriminada. Você pode configurar a representação do usuário para o cluster para que esses trabalhos sejam associados à identidade de usuário que executou o código. HDFSdiretórios de usuário no nó primário são criados para cada identidade de usuário que executa código no notebook. Por exemplo, se o usuário `NbUser1` executar o código do editor de cadernos, é possível se conectar ao nó primário e ver que `hadoop fs -ls /user` mostra o diretório `/user/user_NbUser1`.

Para habilitar esse recurso, configure as propriedades nas classificações de configuração `livy-conf` e `core-site`. Esse recurso não está disponível por padrão quando você faz com que a Amazon EMR crie um cluster junto com um notebook. Para obter mais informações sobre o uso de classificações de configuração para personalizar aplicativos, consulte Configuração de aplicativos no Amazon EMR Release Guide.

Use as seguintes classificações e valores de configuração para permitir a representação do usuário em Notebooks: EMR

```
[
    {
        "Classification": "core-site",
        "Properties": {
          "hadoop.proxyuser.livy.groups": "*",
          "hadoop.proxyuser.livy.hosts": "*"
        }
```

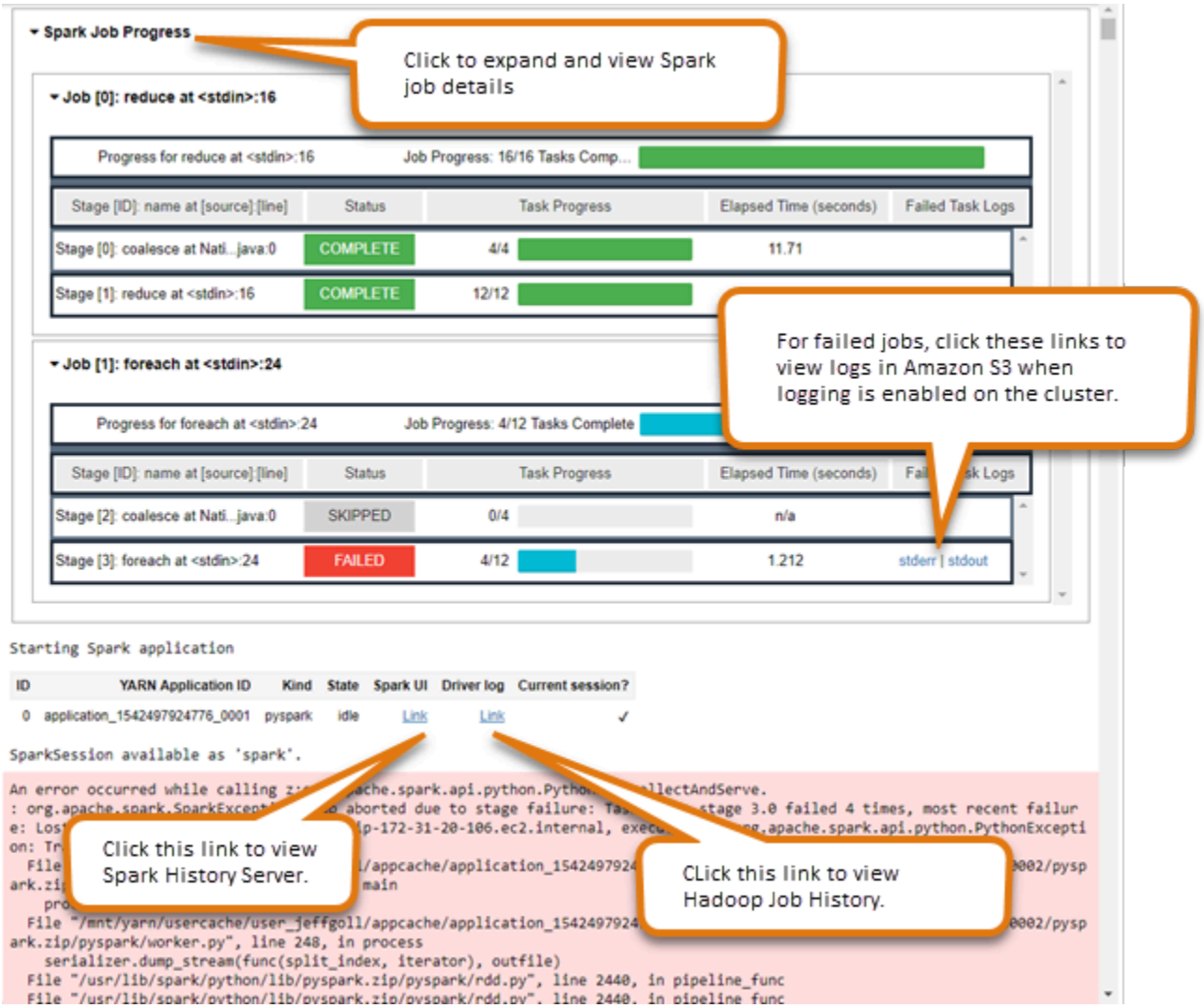
```
    },
    {
        "Classification": "livy-conf",
        "Properties": {
          "livy.impersonation.enabled": "true"
        }
    }
]
```

Uso do widget de monitoramento de trabalhos do Spark

Quando você executa um código no editor do notebook que executa trabalhos do Spark no EMR cluster, a saída inclui um widget do Jupyter Notebook para monitoramento de trabalhos do Spark. O widget fornece detalhes do trabalho e links úteis para a página do servidor de histórico do Spark e para a página de histórico de trabalhos do Hadoop, além de links convenientes para logs de trabalho no Amazon S3 para todos os trabalhos com falha.

Para visualizar as páginas do servidor de histórico no nó primário do cluster, você deve configurar um SSH cliente e um proxy conforme apropriado. Para obter mais informações, consulte Visualize interfaces web hospedadas em EMR clusters da Amazon. Para visualizar os logs no Amazon S3, o registro em log do cluster deve estar habilitado, que é o padrão para os novos clusters. Para obter mais informações, consulte Visualizar arquivos de log arquivados no Amazon S3.

A seguir é apresentado um exemplo de monitoramento de trabalhos do Spark.



EMRsegurança e controle de acesso de notebooks

> **Note**
>
> EMROs notebooks estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar espaços de trabalho, os usuários do EMR Notebooks precisam de permissões adicionais de IAM função. Para obter mais informações, consulte Amazon EMR Notebooks são espaços de trabalho do Amazon EMR Studio no console e no console da Amazon EMR.

Vários recursos estão disponíveis para ajudá-lo a personalizar a postura de segurança dos EMR notebooks. Isso ajuda a garantir que somente usuários autorizados tenham acesso a um EMR

notebook, possam trabalhar com notebooks e usar o editor do notebook para executar código no cluster. Esses recursos funcionam junto com os recursos de segurança disponíveis para a Amazon EMR e os EMR clusters da Amazon. Para obter mais informações, consulte Segurança na Amazon EMR.

- Você pode usar declarações AWS Identity and Access Management de política junto com etiquetas de caderno para limitar o acesso. Para ter mais informações, consulte Como a Amazon EMR trabalha com IAM e Exemplo de declarações de política baseadas em identidade para notebooks EMR.
- Os grupos EC2 de segurança da Amazon atuam como firewalls virtuais que controlam o tráfego de rede entre a instância primária do cluster e o editor do notebook. Você pode usar valores padrão ou personalizar esses grupos de segurança. Para obter mais informações, consulte Especificando grupos EC2 de segurança para notebooks EMR.
- Você especifica uma função de AWS serviço que determina quais permissões um EMR notebook tem ao interagir com outros AWS serviços. Para obter mais informações, consulte Função de serviço para EMR notebooks.

Instalação e uso de kernels e bibliotecas

> **Note**
>
> EMROs notebooks estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar espaços de trabalho, os usuários do EMR Notebooks precisam de permissões adicionais de IAM função. Para obter mais informações, consulte Amazon EMR Notebooks são espaços de trabalho do Amazon EMR Studio no console e no console da Amazon EMR.

Cada EMR notebook vem com um conjunto de bibliotecas e kernels pré-instalados. Você pode instalar bibliotecas e kernels adicionais em um EMR cluster se o cluster tiver acesso ao repositório em que os kernels e as bibliotecas estão localizados. Por exemplo, para clusters em sub-redes privadas, talvez seja necessário configurar a tradução de endereços de rede (NAT) e fornecer um caminho para que o cluster acesse o repositório público do PyPI para instalar uma biblioteca. Para obter mais informações sobre como configurar o acesso externo para diferentes configurações de rede, consulte Cenários e exemplos no Guia VPC do usuário da Amazon.

EMROs aplicativos sem servidor vêm com as seguintes bibliotecas pré-instaladas para Python e: PySpark

- Bibliotecas Python – ggplot, matplotlib, numpy, pandas, plotly, bokeh, scikit-learn, scipy, scipy
- PySpark bibliotecas — ggplotmatplotlib,numpy,pandas,plotly,bokeh,scikit-learn,scipy, scipy

Instalação de kernels e de bibliotecas Python em um nó primário do cluster

Com a EMR versão 5.30.0 e posterior da Amazon, excluindo a 6.0.0, você pode instalar bibliotecas e kernels adicionais do Python no nó primário do cluster. Após a instalação, esses kernels e bibliotecas ficam disponíveis para qualquer usuário executando um EMR notebook conectado ao cluster. As bibliotecas Python instaladas dessa forma estão disponíveis somente para processos em execução no nó primário. As bibliotecas não são instaladas nos nós principais ou de tarefas e não estão disponíveis para executores em execução nesses nós.

> **Note**
>
> Para EMR as versões 5.30.1, 5.31.0 e 6.1.0 da Amazon, você deve realizar etapas adicionais para instalar kernels e bibliotecas no nó primário de um cluster.
> Para habilitar o recurso, faça o seguinte:
>
> 1. Certifique-se de que a política de permissões anexada à função de serviço para EMR Notebooks permita a seguinte ação:
>
>    ```
>    elasticmapreduce:ListSteps
>    ```
>
>   Para obter mais informações, consulte Função de serviço para EMR notebooks.
> 2. Use o AWS CLI para executar uma etapa no cluster que configura os EMR Notebooks, conforme mostrado no exemplo a seguir. Você deve usar o nome da etapa `EMRNotebooksSetup`. Substituir *`us-east-1`* com a região em que seu cluster reside. Para obter mais informações, consulte Adding steps to a cluster using the AWS CLI.
>
>    ```
>    aws emr add-steps --cluster-id MyClusterID --steps
>     Type=CUSTOM_JAR,Name=EMRNotebooksSetup,ActionOnFailure=CONTINUE,Jar=s3://us-
>    east-1.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://
>    awssupportdatasvcs.com/bootstrap-actions/EMRNotebooksSetup/emr-notebooks-
>    setup.sh"]
>    ```

Você pode instalar kernels e bibliotecas usando `pip` ou `conda` no diretório `/emr/notebook-env/bin` no nó primário.

Example : instalação de bibliotecas Python

No kernel do Python3, execute a mágica `%pip` como um comando de dentro de uma célula de caderno para instalar bibliotecas Python.

```
%pip install pmdarima
```

Pode ser necessário reiniciar o kernel para usar os pacotes atualizados. Você também pode usar a mágica `%%sh` do Spark para invocar `pip`.

```
%%sh
/emr/notebook-env/bin/pip install -U matplotlib
/emr/notebook-env/bin/pip install -U pmdarima
```

Ao usar um PySpark kernel, você pode instalar bibliotecas no cluster usando `pip` comandos ou usar bibliotecas com escopo de notebook de dentro de um notebook. PySpark

Para executar `pip` comandos no cluster a partir do terminal, primeiro conecte-se ao nó primário usandoSSH, conforme demonstrado nos comandos a seguir.

```
sudo pip3 install -U matplotlib
sudo pip3 install -U pmdarima
```

Como alternativa, você pode usar bibliotecas com escopo de cadernos. Com bibliotecas com escopo de cadernos, a instalação da sua biblioteca é limitada ao escopo da sua sessão e ocorre em todos os executores do Spark. Para obter mais informações, consulte Uso de bibliotecas com escopo de cadernos.

Se você quiser empacotar várias bibliotecas Python em um PySpark kernel, você também pode criar um ambiente virtual Python isolado. Para obter exemplos, consulte Uso de Virtualenv.

Para criar um ambiente virtual Python em uma sessão, use a propriedade `spark.yarn.dist.archives` do Spark do comando mágico `%%configure` na primeira célula de um caderno, como demonstra o exemplo a seguir.

```
%%configure -f
{
```

```
   "conf": {
   "spark.yarn.appMasterEnv.PYSPARK_PYTHON":"./environment/bin/python",
   "spark.yarn.appMasterEnv.PYSPARK_DRIVER_PYTHON":"./environment/bin/python",
   "spark.yarn.dist.archives":"s3://DOC-EXAMPLE-BUCKET/prefix/
my_pyspark_venv.tar.gz#environment",
   "spark.submit.deployMode":"cluster"
   }
}
```

De forma semelhante, você pode criar um ambiente de executor do Spark.

```
%%configure -f
{
   "conf": {
   "spark.yarn.appMasterEnv.PYSPARK_PYTHON":"./environment/bin/python",
   "spark.yarn.appMasterEnv.PYSPARK_DRIVER_PYTHON":"./environment/bin/python",
   "spark.executorEnv.PYSPARK_PYTHON":"./environment/bin/python",
   "spark.yarn.dist.archives":"s3://DOC-EXAMPLE-BUCKET/prefix/
my_pyspark_venv.tar.gz#environment",
   "spark.submit.deployMode":"cluster"
   }
}
```

Você também pode usar `conda` para instalar bibliotecas Python. Você não precisa de acesso ao sudo para usar `conda`. Você deve se conectar ao nó primário com eSSH, em seguida, executar a `conda` partir do terminal. Para obter mais informações, consulte Conecte-se ao nó primário usando SSH.

Example : instalação de kernels

O seguinte exemplo demonstra a instalação do kernel do Kotlin usando um comando de terminal enquanto estiver conectado ao nó primário de um cluster:

```
sudo /emr/notebook-env/bin/conda install kotlin-jupyter-kernel -c jetbrains
```

> **Note**
>
> Estas instruções não instalam as dependências do kernel. Se o seu kernel tiver dependências de terceiros, talvez seja necessário realizar etapas adicionais de configuração antes de poder usar o kernel com o seu caderno.

Considerações e limitações com bibliotecas com escopo de cadernos

Considere o seguinte ao usar bibliotecas com escopo de cadernos:

- Bibliotecas com escopo de notebook estão disponíveis para clusters que você cria com as EMR versões 5.26.0 e superiores da Amazon.
- As bibliotecas com escopo de notebook devem ser usadas somente com o kernel. PySpark
- Qualquer usuário pode instalar bibliotecas adicionais com escopo de bloco de anotações de dentro de uma célula de bloco de anotações. Essas bibliotecas só estão disponíveis para esse usuário do bloco de anotações durante uma única sessão do bloco de anotações. Se outros usuários precisarem das mesmas bibliotecas ou o mesmo usuário precisar das mesmas bibliotecas em uma sessão diferente, a biblioteca deverá ser reinstalada.
- Você pode desinstalar somente as bibliotecas que foram instaladas com `install_pypi_package` API o. Não é possível desinstalar nenhuma biblioteca pré-instalada no cluster.
- Se as mesmas bibliotecas com versões diferentes estiverem instaladas no cluster e como bibliotecas com escopo de bloco de anotações, a versão da biblioteca com escopo de bloco de anotações substituirá a versão da biblioteca do cluster.

Como trabalhar com bibliotecas com escopo de cadernos

Para instalar bibliotecas, seu EMR cluster da Amazon deve ter acesso ao repositório PyPI onde as bibliotecas estão localizadas.

Os exemplos a seguir demonstram comandos simples para listar, instalar e desinstalar bibliotecas de dentro de uma célula do notebook usando o PySpark kernel e. APIs Para ver exemplos adicionais, consulte a postagem Instalar bibliotecas Python em um cluster em execução com EMR Notebooks no AWS Big Data Blog.

Example : listagem de bibliotecas atuais

O comando a seguir lista os pacotes Python disponíveis para a sessão de bloco de anotações Spark atual. Isso lista as bibliotecas instaladas no cluster e as bibliotecas com escopo de bloco de anotações.

```
sc.list_packages()
```

Example : instalação da biblioteca Celery

O comando a seguir instala a biblioteca Celery como uma biblioteca com escopo de bloco de anotações.

```
sc.install_pypi_package("celery")
```

Depois de instalar a biblioteca, o comando a seguir confirma que a biblioteca está disponível no driver e nos executores Spark.

```
import celery
sc.range(1,10000,1,100).map(lambda x: celery.__version__).collect()
```

Example : instalação da biblioteca Arrow com especificação de versão e de repositório

O comando a seguir instala a biblioteca Arrow como uma biblioteca com escopo de notebook, com uma especificação da versão da biblioteca e do repositório. URL

```
sc.install_pypi_package("arrow==0.14.0", "https://pypi.org/simple")
```

Example : desinstalação de uma biblioteca

O comando a seguir desinstala a biblioteca Arrow, removendo-a como uma biblioteca com escopo de bloco de anotações da sessão atual.

```
sc.uninstall_package("arrow")
```

Associando repositórios baseados em Git a notebooks EMR

> **Note**
>
> EMROs notebooks estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar espaços de trabalho, os usuários do EMR Notebooks precisam de permissões adicionais de IAM função. Para obter mais informações, consulte Amazon EMR Notebooks são espaços de trabalho do Amazon EMR Studio no console e no console da Amazon EMR.

Você pode associar repositórios baseados em Git aos seus EMR cadernos da Amazon para salvá-los em um ambiente com controle de versão. É possível associar até três repositórios a um bloco de anotações. Os seguintes serviços baseados em GIT são compatíveis:

- AWS CodeCommit
- GitHub
- Bitbucket
- GitLab

Associar repositórios baseados em GIT ao bloco de anotações tem os seguintes benefícios.

- Controle de versão: é possível registrar alterações de código em um sistema com controle de versão para poder analisar o histórico de alterações e revertê-las seletivamente.
- Colaboração: colegas que trabalham em diferentes cadernos podem compartilhar códigos por meio de repositórios remotos baseados em Git. Os blocos de anotações podem clonar ou mesclar código de repositórios remotos e retornar as alterações para esses repositórios remotos.
- Reutilização de código — Muitos notebooks Jupyter que demonstram técnicas de análise de dados ou aprendizado de máquina estão disponíveis em repositórios hospedados publicamente, como. GitHub É possível associar os blocos de anotações a um repositório para reutilizar os blocos de anotações Jupyter contidos em um repositório.

Para usar repositórios baseados em Git com EMR Notebooks, você adiciona os repositórios como recursos no EMR console da Amazon, associa credenciais para repositórios que exigem autenticação e os vincula aos seus cadernos. Você pode ver uma lista de repositórios que estão armazenados em sua conta e detalhes sobre cada repositório no console da AmazonEMR. Você pode associar um repositório baseado em GIT existente a um bloco de anotações ao criá-lo.

Tópicos

Pré-requisitos e considerações

> **Note**
>
> EMROs notebooks estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar espaços de trabalho, os usuários do EMR Notebooks precisam de permissões adicionais de IAM função. Para obter mais informações, consulte Amazon EMR Notebooks são espaços de trabalho do Amazon EMR Studio no console e no console da Amazon EMR.

Considere o seguinte ao planejar a integração de um repositório baseado em Git com Notebooks. EMR

AWS CodeCommit

Se você usa um CodeCommit repositório, deve usar as credenciais HTTPS do Git e com. CodeCommit SSHAs chaves e HTTPS com o auxiliar de AWS CLI credenciais não são suportadas. CodeCommit não suporta tokens de acesso pessoal (PATs). Para obter mais informações, consulte Usando IAM com CodeCommit: credenciais, SSH chaves e chaves de AWS acesso do Git no Guia do IAM usuário e Configuração para usuários que HTTPS usam credenciais do Git no Guia do usuário.AWS CodeCommit

Considerações sobre acesso e permissão

Antes de associar um repositório ao seu notebook, certifique-se de que seu cluster, IAM função para EMR Notebooks e grupos de segurança tenham as configurações e permissões corretas. Você também pode configurar repositórios baseados em Git hospedados em uma rede privada ao seguir as instruções em Configurar um repositório Git hospedado de forma privada para notebooks EMR.

- Acesso à Internet do cluster: a interface de rede iniciada tem somente um endereço IP privado. Isso significa que o cluster ao qual seu notebook se conecta deve estar em uma sub-rede privada com um gateway de tradução de endereços de rede (NAT) ou deve ser capaz de acessar a Internet por meio de um gateway privado virtual. Para obter mais informações, consulte as VPCopções da Amazon.

  Os grupos de segurança do bloco de anotações devem incluir uma regra de saída que permita ao bloco de anotações rotear tráfego para a Internet por meio do cluster. Recomendamos que você

crie seus próprios grupos de segurança. Para obter mais informações, consulte Especificando grupos EC2 de segurança para EMR notebooks.

> **Important**
>
> Se a interface de rede for iniciada em uma sub-rede pública, ela não conseguirá se comunicar com a Internet por meio de um gateway de internet (IGW).

- Permissões para AWS Secrets Manager — Se você usa o Secrets Manager para armazenar segredos usados para acessar um repositório, ele the section called "EMRFunção dos notebooks" deve ter uma política de permissões anexada que permita a `secretsmanager:GetSecretValue` ação.

Configurar um repositório Git hospedado de forma privada para notebooks EMR

Use as instruções a seguir para configurar repositórios hospedados de forma privada para Notebooks. EMR Você deve fornecer um arquivo de configuração com informações sobre seus servidores DNS e os do Git. A Amazon EMR usa essas informações para configurar EMR notebooks que podem direcionar o tráfego para seus repositórios hospedados de forma privada.

Pré-requisitos

Antes de configurar um repositório Git hospedado de forma privada EMR para Notebooks, você deve ter o seguinte:

- Um Amazon S3 Control local onde os arquivos do seu EMR notebook serão salvos.

Para configurar um ou mais repositórios Git hospedados de forma privada para Notebooks EMR

1. Crie um arquivo de configuração usando o modelo fornecido. Inclua os seguintes valores para cada servidor Git que deseja especificar em sua configuração:

   - **`DnsServerIpV4`**- O IPv4 endereço do seu DNS servidor. Se você fornecer valores para `DnsServerIpV4` e `GitServerIpV4List`, o valor para `DnsServerIpV4` terá precedência e será usado para resolver seu `GitServerDnsName`.

> **Note**
>
> Para usar repositórios Git hospedados de forma privada, DNS seu servidor deve permitir acesso de entrada a partir de Notebooks. EMR É altamente recomendável que você proteja seu DNS servidor contra outros acessos não autorizados.

- **`GitServerDnsName`**- O DNS nome do seu servidor Git. Por exemplo, `"git.example.com"`.
- **`GitServerIpV4List`**- Uma lista de IPv4 endereços que pertencem ao (s) seu (s) servidor (es) Git.

```
[
    {
        "Type": "PrivatelyHostedGitConfig",
        "Value": [
            {
                "DnsServerIpV4": "<10.24.34.xxx>",
                "GitServerDnsName": "<enterprise.git.com>",
                "GitServerIpV4List": [
                    "<xxx.xxx.xxx.xxx>",
                    "<xxx.xxx.xxx.xxx>"
                ]
            },
            {
                "DnsServerIpV4": "<10.24.34.xxx>",
                "GitServerDnsName": "<git.example.com>",
                "GitServerIpV4List": [
                    "<xxx.xxx.xxx.xxx>",
                    "<xxx.xxx.xxx.xxx>"
                ]
            }
        ]
    }
]
```

2. Salve seu arquivo de configuração como `configuration.json`.
3. Faça o upload do arquivo de configuração no local de armazenamento designado do Amazon S3 em uma pasta chamada `life-cycle-configuration`. Por exemplo, se o local padrão do S3 for `s3://DOC-EXAMPLE-BUCKET/notebooks`, seu arquivo de configuração deverá estar

localizado em `s3://DOC-EXAMPLE-BUCKET/notebooks/`**`life-cycle-configuration/configuration.json`**.

> **Important**
>
> É altamente recomendável que você restrinja o acesso à sua `life-cycle-configuration` pasta somente aos administradores do EMR Notebooks e à função de serviço do Notebooks. EMR Você também deve proteger `configuration.json` contra acesso não autorizado. Para obter instruções, consulte Controlar o acesso a um bucket com políticas de usuário ou Práticas recomendadas de segurança para o Amazon S3.

Para obter instruções sobre como fazer o upload, consulte Criar uma pasta e Fazer upload de objetos no Guia do usuário do Amazon Simple Storage Service.

Adicione um repositório baseado em Git à Amazon EMR

> **Note**
>
> EMROs notebooks estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar espaços de trabalho, os usuários do EMR Notebooks precisam de permissões adicionais de IAM função. Para obter mais informações, consulte Amazon EMR Notebooks são espaços de trabalho do Amazon EMR Studio no console e no console da Amazon EMR.

Consulte as seções a seguir para obter informações sobre como adicionar um repositório baseado em Git a um EMR notebook no console antigo ou a um EMR Studio Workspace no console.

Console

Como os EMR Notebooks são espaços de trabalho do EMR Studio no novo console, você pode seguir as instruções Vincular repositórios baseados em Git a um espaço de trabalho do Studio EMR para associar até três repositórios Git ao seu espaço de trabalho.

Como alternativa, você pode usar a extensão JupyterLab Git. Escolha o ícone Git na barra lateral esquerda do seu caderno JupyterLab para acessar a extensão. Para obter informações sobre a extensão, consulte o repositório GitHub jupyterlab-git.

Para associar um repositório Git a um Workspace, o administrador do Studio deve seguir etapas para configurar o Studio para permitir a vinculação do repositório Git. Para obter mais informações, consulte Estabelecimento de acesso e de permissões para repositórios baseados em Git.

Atualização ou exclusão de um repositório baseado em Git

> **Note**
>
> EMROs notebooks estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar espaços de trabalho, os usuários do EMR Notebooks precisam de permissões adicionais de IAM função. Para obter mais informações, consulte Amazon EMR Notebooks são espaços de trabalho do Amazon EMR Studio no console e no console da Amazon EMR.

Consulte as seções a seguir para obter informações sobre como excluir um repositório baseado em Git de um EMR notebook no console antigo ou de um EMR Studio Workspace no console.

Console

Como os EMR Notebooks são espaços de trabalho do EMR Studio no novo console, você pode consultar Vincular repositórios baseados em Git a um espaço de trabalho do Studio EMR para obter mais informações sobre como trabalhar com repositórios Git em seu espaço de trabalho. Mas, neste momento, não é possível excluir repositórios do Git do Workspaces.

Vinculação ou desvinculação de um repositório baseado em Git

> **Note**
>
> EMROs notebooks estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar espaços de trabalho, os usuários do EMR Notebooks precisam de permissões adicionais de IAM função. Para obter mais informações, consulte Amazon EMR Notebooks são espaços de trabalho do Amazon EMR Studio no console e no console da Amazon EMR.

Use as etapas a seguir para vincular ou desvincular um repositório baseado em Git a um EMR notebook no console antigo ou a um EMR Studio Workspace no console.

Console

Como os EMR Notebooks são espaços de trabalho do EMR Studio no novo console, você pode consultar Vincular repositórios baseados em Git a um espaço de trabalho do Studio EMR para obter mais informações sobre como trabalhar com repositórios Git em seu espaço de trabalho. Mas, neste momento, não é possível excluir repositórios do Git do Workspaces.

Noções básicas sobre o status do repositório

Um repositório Git pode ter qualquer um dos status a seguir na lista de repositórios. Para obter mais informações sobre como vincular EMR notebooks a repositórios Git, consulte. Vinculação ou desvinculação de um repositório baseado em Git

Status	Significado
Linking (Vinculando)	O repositório do Git está sendo vinculado ao bloco de anotações. Enquanto o repositório estiver Linking (Vinculando), não será possível interromper o bloco de anotações.
Linked (Vinculado)	O repositório do Git está vinculado ao bloco de anotações. Enquanto o repositório tiver um status Linked (Vinculado) ele estará conectado ao repositório remoto.
Link Failed (Falha ao vincular)	Ocorreu uma falha ao vincular o repositório do Git ao bloco de anotações. Você pode tentar vinculá-los novamente.
Unlinking (Desvinculando)	O repositório do Git está sendo desvinculado do bloco de anotações. Enquanto o repositór io estiver Unlinking (Desvinculando), não será possível interromper o bloco de anotações . Desvincular um repositório do Git de um bloco de anotações apenas o desconecta do

Status	Significado
	repositório remoto; isso não exclui nenhum código do bloco de anotações.
Unlink Failed (Falha ao desvincular)	Ocorreu uma falha ao desvincular o repositório do Git do bloco de anotações. Você pode tentar desvinculá-los novamente.

Criação de um novo Caderno com um repositório do Git associado

> **Note**
>
> EMROs notebooks estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar espaços de trabalho, os usuários do EMR Notebooks precisam de permissões adicionais de IAM função. Para obter mais informações, consulte Amazon EMR Notebooks são espaços de trabalho do Amazon EMR Studio no console e no console da Amazon EMR.

Para criar um notebook e associá-lo aos repositórios Git no antigo console da Amazon EMR

1. Siga as instruções em Criação de um bloco de anotações.
2. Em Security group (Grupo de segurança), selecione Use your own security group (Usar seu próprio grupo de segurança).

   > **Note**
   >
   > Os grupos de segurança do bloco de anotações devem incluir uma regra de saída que permita ao bloco de anotações rotear tráfego para a Internet por meio do cluster. Recomendamos que você crie seus próprios grupos de segurança. Para obter mais informações, consulte Especificando grupos EC2 de segurança para EMR notebooks.

3. Em Git repositories (Repositórios do Git), selecione Choose repository (Escolher repositório) para escolher qual repositório associar ao bloco de anotações.

   1. Escolha um repositório armazenado como um recurso na sua conta e selecione Save (Salvar).

2. Para adicionar um novo repositório como um recurso em sua conta, selecione add a new repository (adicionar um novo repositório). Conclua o fluxo de trabalho Add repository (Adicionar repositório) em uma nova janela.

Uso de repositórios do Git em um Caderno

> **Note**
>
> EMROs notebooks estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar espaços de trabalho, os usuários do EMR Notebooks precisam de permissões adicionais de IAM função. Para obter mais informações, consulte Amazon EMR Notebooks são espaços de trabalho do Amazon EMR Studio no console e no console da Amazon EMR.

Você pode escolher Abrir no Jupyter JupyterLab ou Abrir no Jupyter ao abrir um caderno.

Se você optar por abrir o bloco de anotações no Jupyter, será exibida uma lista de pastas e arquivos expansíveis dentro do bloco de anotações. É possível executar comandos do Git manualmente, como os apresentados a seguir em uma célula do bloco de anotações.

```
!git pull origin primary
```

Para abrir qualquer um dos outros repositórios, navegue até as outras pastas.

Se você optar por abrir o notebook com uma JupyterLab interface, poderá usar a extensão JupyterLab Git pré-instalada. Para obter informações sobre a extensão, consulte jupyterlab-git.

Planejar e configurar clusters

Esta seção explica as opções de configuração e as instruções para planejar, configurar e lançar clusters usando a AmazonEMR. Antes de executar um cluster, você faz escolhas sobre o seu sistema com base nos dados que está processando e nos seus requisitos de custo, velocidade, capacidade, disponibilidade, segurança e gerenciabilidade. Suas opções incluem:

- Em qual região executar um cluster, onde e como armazenar dados e como gerar a saída dos resultados. Consulte Configurar o armazenamento de dados físico e o local do cluster.
- Se você estiver executando EMR clusters da Amazon em Outposts ou Locais Zones. Consulte EMRclusters em AWS Outposts ou EMRclusters em AWS Locais Zones.
- Se um cluster é transitório ou de longa execução, e quais softwares ele executa. Consulte Configurar um cluster para continuar ou terminar após a execução da etapa e Configuração de software do cluster.
- Se um cluster tem um único nó primário ou três nós primários. Consulte Planejar e configurar nós primários.
- As opções de hardware e rede que otimizam o custo, o desempenho e a disponibilidade do seu aplicativo. Consulte Configurar o hardware e as redes do cluster.
- Como configurar clusters, para que você possa gerenciá-los com mais facilidade e monitorar as atividades, o desempenho e a integridade. Consulte Configurar registro em log e depuração do cluster e Clusters de etiqueta.
- Como autenticar e autorizar o acesso aos recursos do cluster e como criptografar os dados. Consulte Segurança na Amazon EMR.
- Como integrar-se com outros softwares e serviços. Consulte Integração de drivers e aplicações de terceiros.

Iniciar um cluster rapidamente

Para iniciar rapidamente um cluster com o console

1. Faça login no AWS Management Console e abra o EMR console da Amazon em https://console.aws.amazon.com/emr/clusters.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.

3. Na página Criar cluster, insira ou selecione os valores dos campos fornecidos. O painel de resumo persistente exibe uma visualização em tempo real das opções de cluster que estão selecionadas atualmente. Selecione um título no painel de resumo para navegar até a seção correspondente e fazer os ajustes. O nome do cluster não pode conter os caracteres <, >, $, | ou `(crase). Você deve concluir todas as configurações necessárias antes de escolher Criar cluster.
4. Escolha Criar cluster para aceitar a configuração, conforme o exemplo.
5. A página de detalhes do cluster é exibida. Veja o Status do cluster próximo ao nome do cluster. O status deve ser alterado de Starting para Running para Waiting durante o processo de criação do cluster. Talvez você precise escolher o ícone de atualização na parte superior à direita ou atualizar o navegador para receber atualizações.

   Quando o status muda para Esperando, seu cluster está ativo, em execução e pronto para aceitar etapas e SSH conexões.

Configurar o armazenamento de dados físico e o local do cluster

Esta seção descreve como configurar a região para um cluster, os diferentes sistemas de arquivos disponíveis quando você usa a Amazon EMR e como usá-los. Também aborda como preparar ou carregar dados para a Amazon, EMR se necessário, e também como preparar um local de saída para arquivos de log e quaisquer arquivos de dados de saída que você configurar.

Tópicos

Escolha uma AWS região

A Amazon Web Services é executada em servidores distribuídos em datacenters ao redor do mundo. Os datacenters são organizados por região geográfica. Ao iniciar um EMR cluster da Amazon, você deve especificar uma região. É possível escolher a região para reduzir a latência, minimizar custos ou atender a exigências regulamentares. Para ver a lista de regiões e endpoints suportados pela AmazonEMR, consulte Regiões e endpoints no. Referência geral da Amazon Web Services

Para obter a melhor performance, você deve executar o cluster na mesma região que os seus dados. Por exemplo, se o bucket do Amazon S3 que armazena seus dados de entrada estiver na região Oeste dos EUA (Oregon), você deverá executar seu cluster na região Oeste dos EUA (Oregon) para evitar taxas de transferência de dados entre regiões. Se você usar um bucket do Amazon S3 para receber a saída do cluster, também deverá criá-lo na região Oeste dos EUA (Oregon).

Se você planeja associar um par de EC2 chaves da Amazon ao cluster (necessário SSH para fazer login no nó principal), o par de chaves deve ser criado na mesma região do cluster. Da mesma forma, os grupos de segurança que a Amazon EMR cria para gerenciar o cluster são criados na mesma região do cluster.

Se você se inscreveu Conta da AWS em ou depois de 17 de maio de 2017, a região padrão ao acessar um recurso do AWS Management Console é Leste dos EUA (Ohio) (us-east-2); para contas mais antigas, a região padrão é Oeste dos EUA (Oregon) (us-west-2) ou Leste dos EUA (Norte da Virgínia) (us-east-1). Para mais informações, consulte Regiões e endpoints da .

Alguns AWS recursos estão disponíveis somente em regiões limitadas. Por exemplo, instâncias de computação em cluster estão disponíveis apenas na região Leste dos EUA (Norte da Virgínia), e a região Ásia-Pacífico (Sydney) apenas oferece suporte ao Hadoop 1.0.3 e versões posteriores. Ao escolher uma região, verifique se ela oferece suporte aos atributos que você deseja usar.

Para obter o melhor desempenho, use a mesma região para todos os AWS recursos que serão usados com o cluster. A tabela a seguir mapeia os nomes de regiões entre serviços. Para obter uma lista das EMR regiões da Amazon, consulte Regiões da AWS os endpoints no Referência geral da Amazon Web Services.

Escolher uma região usando o console

A região padrão é exibida à esquerda das informações da conta na barra de navegação. Para trocar de região no console novo ou no antigo, escolha o menu suspenso Região e selecione uma nova opção.

Especifique uma região com o AWS CLI

Especifique uma região padrão AWS CLI usando o aws configure comando ou a variável de `AWS_DEFAULT_REGION` ambiente. Para obter mais informações, consulte Configurando a AWS região no Guia do AWS Command Line Interface usuário.

Escolha uma região com um SDK ou o API

Para escolher uma região usando umSDK, configure seu aplicativo para usar o endpoint dessa região. Se você estiver criando um aplicativo cliente usando um AWS SDK, poderá alterar o endpoint do cliente chamando`setEndpoint`, conforme mostrado no exemplo a seguir:

```
client.setEndpoint("elasticmapreduce.us-west-2.amazonaws.com");
```

Depois que seu aplicativo especificar uma região definindo o endpoint, você pode definir a zona de disponibilidade para as EC2 instâncias do seu cluster. As zonas de disponibilidade são as localizações geográficas distintas que são criadas para serem isoladas das falhas em outras zonas da disponibilidade e fornecem rede de conectividade acessível e de baixa latência a outras zonas de disponibilidade da mesma região. Uma região contém uma ou mais zonas de disponibilidade. Para otimizar o desempenho e reduzir a latência, todos os recursos devem estar localizados na mesma zona de disponibilidade do cluster que os utiliza.

Trabalhar com armazenamento e sistemas de arquivos

A Amazon EMR e o Hadoop fornecem uma variedade de sistemas de arquivos que você pode usar ao processar etapas do cluster. Você especifica qual sistema de arquivos usar pelo prefixo do URI usado para acessar os dados. Por exemplo, `s3://amzn-s3-demo-bucket1/path` faz referência a um bucket do Amazon S3 usando. EMRFS A tabela a seguir lista os sistemas de arquivos disponíveis e inclui as recomendações sobre quando é melhor usar cada um deles.

A Amazon EMR e o Hadoop normalmente usam dois ou mais dos seguintes sistemas de arquivos ao processar um cluster. HDFSe EMRFS são os dois principais sistemas de arquivos usados com a AmazonEMR.

> **Important**
>
> A partir da EMR versão 5.22.0 da Amazon, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. As EMR versões anteriores da Amazon usam a AWS Signature Version 2 em alguns casos, a menos que as notas de lançamento indiquem que a Signature Version 4 é usada exclusivamente. Para obter mais informações, consulte Autenticação de solicitações (AWS assinatura versão 4) e Solicitações de autenticação (AWS assinatura versão 2) no Guia do desenvolvedor do Amazon Simple Storage Service.

Sistema de arquivos	Prefixo	Descrição
HDFS	`hdfs://` (ou sem prefixo)	HDFSé um sistema de arquivos distribuído, escalável e portátil para o Hadoop. Uma vantagem HDFS é o reconhecimento de dados entre os nós do cluster do Hadoop que gerenciam os clusters e os nós do cluster do Hadoop que gerenciam as etapas individuais. Para obter mais informações, consulte a documentação do Hadoop. HDFSé usado pelos nós principal e central. Uma de suas vantagens é a velocidade; uma desvantagem é ser um armazenamento temporário que é reivindicado quando o cluster é encerrado. É melhor usado para armazenamento em cache dos resultados intermedi ários produzidos pelas etapas de um fluxo de trabalho.
EMRFS	`s3://`	EMRFSé uma implementação do sistema de arquivos Hadoop usado para ler e gravar arquivos regulares da Amazon EMR diretamente no Amazon S3. EMRFSforn ece a conveniência de armazenar dados persisten tes no Amazon S3 para uso com o Hadoop, além de fornecer recursos como criptografia, consistência e consistência de listas no servidor do Amazon S3. read-after-write **Note** Anteriormente, a Amazon EMR usava `s3n` os sistemas de `s3a` arquivos e. Embora ambos ainda funcionem, recomendamos que você use o `s3` URI esquema para obter o melhor desempenho, segurança e confiabilidade.
sistema de arquivos local		O sistema de arquivos local é a um disco conectado localmente. Quando um cluster Hadoop é criado, cada nó é criado a partir de uma EC2 instância que vem

Sistema de arquivos	Prefixo	Descrição
		com um bloco pré-configurado de armazenamento em disco pré-conectado chamado armazenamento de instâncias. Os dados nos volumes de armazenam ento da instância persistem somente durante a vida útil da EC2 instância. Os volumes de armazenam ento da instância são ideais para armazenar dados temporários que mudam continuamente, tais como buffers, caches, dados temporários e outros conteúdos temporários. Para obter mais informações, consulte Armazenamento de EC2 instâncias da Amazon. O sistema de arquivos local é usado porHDFS, mas o Python também é executado a partir do sistema de arquivos local e você pode optar por armazenar arquivos de aplicativos adicionais em volumes de armazenamento de instâncias.
(Herdado) Sistema de arquivos de bloco do Amazon S3	`s3bfs://`	O sistema de arquivos de bloco do Amazon S3 é um sistema de armazenamento de arquivos herdado. Não recomendamos em hipótese alguma o uso de deste sistema. **Important** Não recomendamos que você use esse sistema de arquivos pois ele pode acionar um comportamento de disputa que pode causar uma falha no cluster. No entanto, ele pode ser necessário para aplicativos herdados.

Acessar sistemas de arquivo

Você especifica qual sistema de arquivos usar pelo prefixo do identificador uniforme de recursos (URI) usado para acessar os dados. Os procedimentos a seguir ilustram como fazer referência a vários tipos diferentes de sistemas de arquivos.

Para acessar um local HDFS

- Especifique o `hdfs:///` prefixo noURI. A Amazon EMR resolve caminhos que não especificam um prefixo no URI para o local. HDFS Por exemplo, as duas opções a seguir URIs seriam resolvidas no mesmo local emHDFS.

```
hdfs:///path-to-data

/path-to-data
```

Para acessar um controle remoto HDFS

- Inclua o endereço IP do nó principal noURI, conforme mostrado nos exemplos a seguir.

```
hdfs://master-ip-address/path-to-data

master-ip-address/path-to-data
```

Acessar o Amazon S3

- Use o prefixo `s3://`.

```
s3://bucket-name/path-to-file-in-bucket
```

Acessar o sistema de arquivos de bloco do Amazon S3

- Use apenas para aplicações herdadas que exigem o sistema de arquivos de bloco do Amazon S3. Para acessar ou armazenar dados com esse sistema de arquivos, use o `s3bfs://` prefixo noURI.

  O sistema de arquivos de bloco do Amazon S3 é um sistema de arquivos antigo que foi usado para oferecer suporte a carregamentos maiores do que 5 GB de tamanho para o Amazon S3.

Com a funcionalidade de upload em várias partes que a Amazon EMR fornece por meio do AWS JavaSDK, você pode fazer upload de arquivos de até 5 TB para o sistema de arquivos nativo do Amazon S3, e o sistema de arquivos em blocos do Amazon S3 está obsoleto.

> **Warning**
>
> Como esse sistema de arquivos legado pode criar condições de corrida que podem corromper o sistema de arquivos, você deve evitar esse formato e usá-lo EMRFS em seu lugar.

```
s3bfs://bucket-name/path-to-file-in-bucket
```

Preparar dados de entrada

A maioria dos clusters carrega dados de entrada e depois processa esses dados. Para carregar dados, eles precisam estar em um local que o cluster possa acessar e ter um formato que o cluster possa processar. O cenário mais comum é carregar dados de entrada no Amazon S3. EMRA Amazon fornece ferramentas para que seu cluster importe ou leia dados do Amazon S3.

O formato de entrada padrão no Hadoop é um arquivo de texto, embora você possa personalizar o Hadoop e usar ferramentas para importar dados armazenados em outros formatos.

Tópicos

- Tipos de entrada que a Amazon EMR pode aceitar
- Como colocar dados na Amazon EMR

Tipos de entrada que a Amazon EMR pode aceitar

O formato de entrada padrão para um cluster é um arquivo de texto, com cada linha separada por um caractere de nova linha (\n), que é o formato de entrada mais usado.

Se os seus dados de entrada estiverem em um formato diferente do formato de arquivo de texto padrão, você poderá usar a interface do Hadoop `InputFormat` para especificar outros tipos de entradas. Você pode até mesmo criar uma subclasse da classe `FileInputFormat` para tratar os

tipos de dados personalizados. Para obter mais informações, consulte http://hadoop.apache.org/docs/current/api/org/apache/hadoop/mapred/ InputFormat .html.

Se você estiver usando o Hive, você pode usar um serializador/desserializador (SerDe) para ler dados de um determinado formato em. HDFS Para obter mais informações, consulte https://cwiki.apache.org/confluence/display/Hive/ SerDe.

Como colocar dados na Amazon EMR

EMRA Amazon fornece várias maneiras de colocar dados em um cluster. A forma mais comum é fazer o upload dos dados para o Amazon S3 e usar os recursos integrados da Amazon EMR para carregar os dados no seu cluster. Você também pode usar o recurso DistributedCache do Hadoop para transferir arquivos de um sistema de arquivos distribuído para o sistema de arquivos local. A implementação do Hive fornecida pela Amazon EMR (Hive versão 0.7.1.1 e posterior) inclui funcionalidades que você pode usar para importar e exportar dados entre o DynamoDB e um cluster da Amazon. EMR Se tiver grandes quantidades de dados on-premises para processar, talvez considere o serviço AWS Direct Connect útil.

Tópicos

- Carregar dados no Amazon S3
- Carregar dados usando o AWS DataSync
- Importar arquivos com o cache distribuído
- Como processar arquivos compactados
- Importar dados do DynamoDB para o Hive
- Conectar-se aos dados usando o AWS Direct Connect
- Carregar grandes quantidades de dados usando o AWS Snowball

Carregar dados no Amazon S3

Para obter instruções sobre como carregar objetos no Amazon S3, consulte Add an object to your bucket no Guia do usuário do Amazon Simple Storage Service. Para obter mais informações sobre como usar o Amazon S3 com o Hadoop, consulte http://wiki.apache.org/hadoop/AmazonS3.

Tópicos

- Criar e configurar um bucket do Amazon S3
- Configurar o carregamento multiparte para o Amazon S3
- Práticas recomendadas

- Upload de dados no Amazon S3 Express One Zone

Criar e configurar um bucket do Amazon S3

A Amazon EMR usa o AWS SDK for Java com o Amazon S3 para armazenar dados de entrada, arquivos de log e dados de saída. O Amazon S3 se refere a esses locais de armazenamento como bucket. Os buckets têm certas restrições e limitações para se adequarem aos requisitos e aos requisitos do Amazon DNS S3. Para obter mais informações, consulte Restrições e limitações de bucket no Manual do usuário do Amazon Simple Storage Service.

Esta seção mostra como usar o Amazon S3 AWS Management Console para criar e depois definir permissões para um bucket do Amazon S3. Você também pode criar e definir permissões para um bucket do Amazon S3 usando o Amazon API S3 ou. AWS CLI Você também pode usar curl junto com uma modificação para transmitir os parâmetros de autenticação apropriados para o Amazon S3.

Consulte os recursos a seguir:

- Para criar um bucket usando o console, consulte Criação de um bucket, no Guia do usuário do Amazon S3.
- Para criar e trabalhar com buckets usando o AWS CLI, consulte Como usar comandos de alto nível do S3 com o AWS Command Line Interface no Guia do usuário do Amazon S3.
- Para criar um bucket usando umSDK, consulte Exemplos de criação de um bucket no Guia do usuário do Amazon Simple Storage Service.
- Para trabalhar com buckets usando Curl, consulte Amazon S3 authentication tool for curl.
- Para obter mais informações sobre buckets específicos para regiões, consulte Acesso a um bucket no Guia do usuário do Amazon Simple Storage Service.
- Para trabalhar com buckets usando Pontos de Acesso Amazon S3, consulte Usar um alias em estilo de bucket para seu ponto de acesso de bucket do S3 no Guia do usuário do Amazon S3. Você facilmente pode usar os Pontos de Acesso Amazon S3 com o alias do Ponto de Acesso Amazon S3 em vez do nome do bucket do Amazon S3. Use o Ponto de Acesso Amazon S3 para aplicações novas e já existentes, inclusive Spark, Hive, Presto e outros.

> **Note**
>
> Se você ativar o registro em um bucket, ele habilitará somente os logs de acesso ao bucket, não os logs EMR do cluster da Amazon.

Durante a criação do bucket ou depois, você pode definir as permissões apropriadas para acessar o bucket, dependendo de seu aplicativo. Normalmente, você atribui acesso de leitura e gravação para si mesmo (o proprietário) e atribui acesso de leitura para os usuários autenticados.

Os buckets do Amazon S3 obrigatórios devem existir para que você possa criar um cluster. Você deve carregar todos os scripts necessários ou dados referenciados no cluster no Amazon S3. A tabela a seguir descreve dados de exemplo, scripts e locais de arquivo de log.

Configurar o carregamento multiparte para o Amazon S3

A Amazon EMR suporta o upload de várias partes do Amazon S3 por meio do AWS SDK para Java. O multipart upload permite que você faça upload de um único objeto como um conjunto de partes. O upload dessas partes de objetos pode ser feito de maneira independente e em qualquer ordem. Se a transmissão de alguma parte falhar, você poderá retransmitir essa parte sem afetar outras partes. Depois que todas as partes do objeto forem carregadas, o Amazon S3 montará as partes e criará o objeto.

Para obter mais informações, consulte Visão geral do carregamento fracionado no Manual do usuário do Amazon Simple Storage Service.

Além disso, a Amazon EMR oferece propriedades que permitem que você controle com mais precisão a limpeza de peças com falha no upload de várias partes.

A tabela a seguir descreve as propriedades de EMR configuração da Amazon para upload de várias partes. Você pode configurar esses valores usando a classificação de configuração `core-site`. Para obter mais informações, consulte Configurar aplicativos no Amazon EMR Release Guide.

Nome do parâmetro de configuração	Valor padrão	Descrição
`fs.s3n.multipart.uploads.enabled`	`true`	Um tipo booleano que indica se os carregamentos multiparte devem ou não ser habilitados. Quando a visualização EMRFS consistente está ativada, os uploads de várias partes são ativados por padrão e a configuração desse valor como `false` é ignorada.
`fs.s3n.multipart.uploads.split.size`	`134217728`	Especifica o tamanho máximo de uma peça, em bytes, antes de EMRFS iniciar

Nome do parâmetro de configuração	Valor padrão	Descrição
		um novo carregamento de peça quando o carregamento de várias partes está ativado. O valor mínimo é `5242880` (5 MB). Se um valor menor for especificado, `5242880` será usado. O máximo é `5368709120` (5 GB). Se um valor maior for especificado, `5368709120` será usado. Se a criptografia EMRFS do lado do cliente estiver desativada e o Amazon S3 Optimized Committer também estiver desativado, esse valor também controlará o tamanho máximo que um arquivo de dados pode aumentar EMRFS até usar uploads de várias partes em vez `PutObject` de uma solicitação para carregar o arquivo. Para obter mais informações, consulte
`fs.s3n.ssl.enabled`	`true`	Um tipo booleano que indica se o http ou o https deve ser usado.
`fs.s3.buckets.create.enabled`	`false`	Um tipo booleano que indica se um bucket deve ser criado caso ele não exista. Configurar como `false` gera uma exceção em operações `CreateBucket` .
`fs.s3.multipart.clean.enabled`	`false`	Um tipo booliano que indica se deseja habilitar a limpeza periódica em segundo plano de carregamentos multiparte incompletos.
`fs.s3.multipart.clean.age.threshold`	`604800`	Um tipo longo que especifica a idade mínima de um multipart upload, em segundos, antes de ser considerado para limpeza. O padrão é uma semana.

Nome do parâmetro de configuração	Valor padrão	Descrição
`fs.s3.multipart.clean.jitter.max`	`10000`	Um tipo inteiro que especifica o valor máximo de atraso de oscilação aleatória em segundos adicionado ao atraso fixo de 15 minutos antes de programar a próxima execução de limpeza.

Desabilitar carregamentos multiparte

Console

Para desativar os carregamentos de várias partes com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Em Configurações do software, insira a seguinte configuração: `classification=core-site,properties=[fs.s3n.multipart.uploads.enabled=false]`.
4. Escolha qualquer outra opção que se aplique ao cluster.
5. Para iniciar o cluster, escolha Criar cluster.

CLI

Para desativar o upload de várias partes usando o AWS CLI

Esse procedimento explica como desabilitar o multipart upload usando a AWS CLI. Para desabilitar o multipart upload, digite o comando `create-cluster` com o parâmetro `--bootstrap-actions`.

1. Crie um arquivo, `myConfig.json`, com o seguinte conteúdo e salve-o no mesmo diretório onde você executa o comando:

```
[
  {
```

```
    "Classification": "core-site",
    "Properties": {
      "fs.s3n.multipart.uploads.enabled": "false"
    }
  }
]
```

2. Digite o seguinte comando e substitua *myKey* com o nome do seu EC2 key pair.

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --name "Test cluster" \
--release-label emr-7.2.0 --applications Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey --instance-type m5.xlarge \
--instance-count 3 --configurations file://myConfig.json
```

API

Para desativar o upload de várias partes usando o API

- Para obter informações sobre o uso programático de uploads de várias partes do Amazon S3, consulte Usando o AWS SDK for Java para upload de várias partes no Guia do usuário do Amazon Simple Storage Service.

  Para obter mais informações sobre o AWS SDK para Java, consulte AWS SDKpara Java.

Práticas recomendadas

A seguir estão recomendações para o uso de buckets do Amazon S3 com clusters. EMR

Habilitar o versionamento

O versionamento é uma configuração recomendada para o seu bucket do Amazon S3. Habilitando o versionamento, você garante que, mesmo que os dados sejam excluídos ou substituídos sem querer,

eles possam ser recuperados. Para obter mais informações, consulte Usando versionamento no Guia do usuário do Amazon Simple Storage Service.

Limpar carregamentos multiparte com falha

EMRos componentes do cluster usam uploads de várias partes por meio do AWS SDK for Java with Amazon APIs S3 para gravar arquivos de log e enviar dados para o Amazon S3 por padrão. Para obter informações sobre a alteração de propriedades relacionadas a essa configuração usando a AmazonEMR, consulteConfigurar o carregamento multiparte para o Amazon S3. Às vezes, carregar um arquivo grande pode resultar em um carregamento multiparte do Amazon S3 incompleto. Quando não é possível concluir com êxito um multipart upload em andamento, este continua a ocupar seu bucket e resulta em cobranças de armazenamento. Recomendamos as seguintes opções para evitar excesso de armazenamento de arquivos:

- Para buckets que você usa com a AmazonEMR, use uma regra de configuração de ciclo de vida no Amazon S3 para remover uploads incompletos de várias partes três dias após a data de início do upload. As regras de configuração de ciclo de vida permitem que você controle a classe de armazenamento e o tempo de vida dos objetos. Para obter mais informações, consulte Object lifecycle management e Aborting incomplete multipart uploads using a bucket lifecycle policy.
- Ative o recurso EMR de limpeza em várias partes da Amazon `fs.s3.multipart.clean.enabled` configurando `true` e ajustando outros parâmetros de limpeza. Esse recurso é útil em alto volume, grande escala e com clusters que tenham tempo limitado. Nesse caso, o parâmetro `DaysAfterIntitiation` de uma regra de configuração do ciclo de vida pode ser muito longo, mesmo se definido como o mínimo, causando picos no armazenamento do Amazon S3. A limpeza em várias partes EMR da Amazon permite um controle mais preciso. Para obter mais informações, consulte Configurar o carregamento multiparte para o Amazon S3.

Gerenciar marcadores de versão

Recomendamos que você habilite uma regra de configuração de ciclo de vida no Amazon S3 para remover marcadores de exclusão de objetos expirados para buckets versionados que você usa com a Amazon. EMR Ao excluir um objeto em um bucket com versionamento, um marcador de exclusão é criado. Se todas as versões anteriores do objeto expirarem posteriormente, um marcador de exclusão de objeto expirado será deixado no bucket. Embora você não seja cobrado pelos marcadores de exclusão, a remoção dos marcadores expirados pode melhorar o desempenho das solicitações. LIST Para obter mais informações, consulte Lifecycle configuration for a bucket with versioning no Guia do usuário do Amazon Simple Storage Service.

Práticas recomendadas de desempenho

Dependendo de suas cargas de trabalho, tipos específicos de uso de EMR clusters e aplicativos nesses clusters podem resultar em um grande número de solicitações em um bucket. Para obter mais informações, consulte Request rate and performance considerations no Guia do usuário do Amazon Simple Storage Service.

Upload de dados no Amazon S3 Express One Zone

Visão geral

Com o Amazon EMR 6.15.0 e superior, você pode usar o Amazon EMR com o Apache Spark em conjunto com a classe de armazenamento Amazon S3 Express One Zone para melhorar o desempenho em suas tarefas do Spark. As EMR versões 7.2.0 e superiores da Amazon também oferecem suporte ao HBase Flink e ao Hive, então você também pode se beneficiar do S3 Express One Zone se usar esses aplicativos. O S3 Express One Zone é uma classe de armazenamento do S3 para aplicações que acessam dados frequentemente com centenas de milhares de solicitações por segundo. Na hora da execução, o S3 Express One Zone oferece o armazenamento de objetos na nuvem com a menor latência e a maior performance do Amazon S3.

Pré-requisitos

- Permissões do S3 Express One Zone: quando o S3 Express One Zone inicialmente executa uma ação como `GET`, `LIST` ou `PUT` em um objeto do S3, a classe de armazenamento chama `CreateSession` em seu nome. Sua IAM política deve permitir a `s3express:CreateSession` permissão para que o S3A conector possa invocar o. `CreateSession` API Para obter um exemplo de política com essa permissão, consulte Conceitos básicos da classe Amazon S3 Express One Zone.
- Conector S3A: para configurar o cluster do Spark para acessar dados de um bucket do Amazon S3 que usa a classe de armazenamento S3 Express One Zone, você deve usar o conector S3A do Apache Hadoop. Para usar o conector, certifique-se de que todos os S3 URIs usem o `s3a` esquema. Caso contrário, você pode alterar a implementação do sistema de arquivos usado para os esquemas do `s3` e do `s3n`.

Para alterar o esquema do `s3`, especifique as seguintes configurações de cluster:

```
[
  {
    "Classification": "core-site",
```

```
    "Properties": {
      "fs.s3.impl": "org.apache.hadoop.fs.s3a.S3AFileSystem",
      "fs.AbstractFileSystem.s3.impl": "org.apache.hadoop.fs.s3a.S3A"
    }
  }
]
```

Para alterar o esquema do `s3n`, especifique as seguintes configurações de cluster:

```
[
  {
    "Classification": "core-site",
    "Properties": {
      "fs.s3n.impl": "org.apache.hadoop.fs.s3a.S3AFileSystem",
      "fs.AbstractFileSystem.s3n.impl": "org.apache.hadoop.fs.s3a.S3A"
    }
  }
]
```

Conceitos básicos da classe Amazon S3 Express One Zone

Tópicos

- Criação de uma política de permissões
- Criação e configuração de um cluster
- Visão geral das configurações

Criação de uma política de permissões

Antes de criar um cluster que usa o Amazon S3 Express One Zone, você deve criar uma IAM política para anexar ao perfil de EC2 instância da Amazon para o cluster. A política deve ter permissões para acessar a classe de armazenamento S3 Express One Zone. O exemplo de política a seguir mostra como conceder a permissão necessária. Depois de criar a política, anexe a política à função de perfil da instância que você usa para criar seu EMR cluster, conforme descrito na Criação e configuração de um cluster seção.

```
{
  "Version":"2012-10-17",
  "Statement": [
        {
            "Effect": "Allow",
```

```
            "Resource": "arn:aws:s3express:region-code:account-id:bucket/DOC-EXAMPLE-
BUCKET",
            "Action": [
                "s3express:CreateSession"
            ]
        }
    ]

}
```

Criação e configuração de um cluster

Em seguida, crie um cluster que execute o SparkHBase, o Flink ou o Hive com o S3 Express One Zone. As seguinte etapas descrevem uma visão geral de alto nível para criar um cluster no AWS Management Console:

1. Navegue até o EMR console da Amazon e selecione Clusters na barra lateral. Depois, selecione Criar cluster.
2. Se você usa o Spark, selecione a EMR versão Amazon `emr-6.15.0` ou superior. Se você usa HBase o Flink ou o Hive, selecione `emr-7.2.0` ou superior.
3. Selecione os aplicativos que você deseja incluir em seu cluster, como Spark ou Flink. HBase
4. Para habilitar o Amazon S3 Express One Zone, insira uma configuração semelhante ao exemplo a seguir na seção Configurações de software. As configurações e os valores recomendados estão descritos na seção Visão geral das configurações após esse procedimento.

```
[
  {
    "Classification": "core-site",
    "Properties": {
      "fs.s3a.aws.credentials.provider":
 "software.amazon.awssdk.auth.credentials.InstanceProfileCredentialsProvider",
      "fs.s3a.change.detection.mode": "none",
      "fs.s3a.endpoint.region": "aa-example-1",
      "fs.s3a.select.enabled": "false"
    }
  },
  {
    "Classification": "spark-defaults",
    "Properties": {
      "spark.sql.sources.fastS3PartitionDiscovery.enabled": "false"
    }
```

```
  }
]
```

5. Na EMR seção Perfil de EC2 instância da Amazon, escolha usar uma função existente e use uma função com a política anexada que você criou na Criação de uma política de permissões seção acima.
6. Defina o restante das configurações do cluster conforme apropriado para a sua aplicação e selecione Criar cluster.

Visão geral das configurações

As tabelas a seguir descrevem as configurações e os valores sugeridos que você deve especificar ao configurar um cluster que usa o S3 Express One Zone com a AmazonEMR, conforme descrito na Criação e configuração de um cluster seção.

Configurações do S3A

Parâmetro	Valor padrão	Valor sugerido	Explicação
`fs.s3a.aws.credentials.provider`	Se não for especificado, usa `AWSCredentialProviderList` na seguinte ordem: `TemporaryAWSCredentialsProvider` , `SimpleAWSCredentialsProvider` , `EnvironmentVariableCredentialsProvider` , `IAMInstanceCredentialsProvider` .	`software.amazon.awssdk.auth.credentials.InstanceProfileCredentialsProvider`	A função do perfil da EMR instância da Amazon deve ter a política que permite que o S3A sistema de arquivos faça chamadas. `s3express:CreateSession` Outros provedores de credenciais também funcionam se tiverem as permissões do S3 Express One Zone.

Parâmetro	Valor padrão	Valor sugerido	Explicação
`fs.s3a.endpoint.region`	nulo	O Região da AWS local onde você criou o bucket.	A lógica de resolução da região não funciona com a classe de armazenamento S3 Express One Zone.
`fs.s3a.select.enabled`	`true`	`false`	O valor `select` do Amazon S3 não é compatível com a classe de armazenamento S3 Express One Zone.
`fs.s3a.change.detection.mode`	`server`	nenhuma	A detecção de alterações pelo S3A funciona verificando `etags` baseadas em MD5. A classe de armazenamento S3 Express One Zone não é compatível com `checksums` de MD5.

Configurações do Spark

Parâmetro	Valor padrão	Valor sugerido	Explicação
`spark.sql.sources.fastS3Par`	`true`	`false`	A otimização interna usa um API parâmetro S3 que a

Parâmetro	Valor padrão	Valor sugerido	Explicação
`titionDis covery.enabled`			classe de armazenam ento S3 Express One Zone não suporta.

Considerações

Considere o seguinte ao integrar o Apache Spark na Amazon EMR com a classe de armazenamento S3 Express One Zone:

- O Amazon S3 Express One Zone é compatível com as EMR versões 6.15.0 e superiores da Amazon.
- O conector S3A é necessário para usar o S3 Express One Zone com a Amazon. EMR Somente o S3A tem os recursos e as classes de armazenamento necessários para interagir com o S3 Express One Zone. Para ver as etapas de configuração do conector, consulte the section called “Pré-requisitos”.
- A classe de armazenamento Amazon S3 Express One Zone só é compatível com o Spark em um EMR cluster da Amazon executado na Amazon. EC2
- A classe de armazenamento Amazon S3 Express One Zone só oferece suporte à criptografia SSE-S3. Para obter mais informações, consulte Criptografia do lado do servidor com chaves gerenciadas do Amazon S3 (-S3). SSE
- A classe de armazenamento Amazon S3 Express One Zone não oferece suporte a gravações com o `FileOutputCommitter` do S3A. As gravações com o `FileOutputCommitter` do S3A em buckets do S3 Express One Zone resultam em um erro: InvalidStorageClass: The storage class you specified is not valid.
- A classe de armazenamento Amazon S3 Express One Zone não é compatível com o Amazon EMR Serverless ou o Amazon on. EMR EKS

Carregar dados usando o AWS DataSync

AWS DataSync é um serviço de transferência de dados on-line que simplifica, automatiza e acelera o processo de movimentação de dados entre seus serviços de armazenamento e armazenamento locais ou entre serviços AWS de armazenamento. AWS DataSync oferece suporte a uma variedade de sistemas de armazenamento local, como o Hadoop Distributed File System (HDFS), servidores de NAS arquivos e armazenamento autogerenciado de objetos.

A maneira mais comum de colocar dados em um cluster é fazer o upload dos dados para o Amazon S3 e usar os recursos integrados da Amazon EMR para carregar os dados no seu cluster.

DataSync pode ajudá-lo a realizar as seguintes tarefas:

- Replique HDFS em seu cluster Hadoop para o Amazon S3 para continuidade dos negócios
- Copie HDFS para o Amazon S3 para preencher seus data lakes
- Transfira dados entre seu cluster Hadoop HDFS e o Amazon S3 para análise e processamento

Para fazer upload de dados para seu bucket do S3, primeiro você implanta um ou mais DataSync agentes na mesma rede do seu armazenamento local. O agente é uma máquina virtual (VM) usada para ler ou gravar dados em um local autogerenciado. Em seguida, você ativa seus agentes no bucket do S3 Conta da AWS e Região da AWS onde ele está localizado.

Depois que o agente é ativado, crie um local de origem para o armazenamento on-premises, um local de destino para o bucket do S3 e uma tarefa. Uma tarefa é um conjunto de dois locais (origem e destino) e um conjunto de opções padrão que você usa para controlar o comportamento da tarefa.

Finalmente, você executa sua DataSync tarefa de transferir dados da origem para o destino.

Para obter mais informações, consulte Conceitos básicos do AWS DataSync.

Importar arquivos com o cache distribuído

Tópicos

- Tipos de arquivos compatíveis
- Local dos arquivos em cache
- Acessar arquivos em cache de aplicações de transmissão
- Acessar arquivos em cache de aplicações de transmissão

O DistributedCache é um recurso do Hadoop que pode aumentar a eficiência quando uma tarefa map ou reduce precisa acessar dados comuns. Se o cluster depender de aplicações já existentes ou binárias que não estão instaladas quando o cluster é criado, você poderá usar o DistributedCache para importar esses arquivos. Esse recurso permite que um nó de cluster leia os arquivos importados do seu sistema de arquivos local, em vez de recuperar os arquivos de outros nós do cluster.

Para obter mais informações, acesse http://hadoop.apache.org/docs/stable/api/org/apache/hadoop/filecache/ DistributedCache .html.

Você invoca o DistributedCache ao criar o cluster. Os arquivos são armazenados em cache logo antes do início do trabalho do Hadoop e permanecem no cache pela duração do trabalho. Você pode armazenar em cache arquivos armazenados em qualquer sistema de arquivos compatível com Hadoop, por exemplo, ou no Amazon HDFS S3. O tamanho padrão do cache de arquivo é 10 GB. Para alterar o tamanho do cache, reconfigure o parâmetro Hadoop `local.cache.size` usando a ação de bootstrap. Para obter mais informações, consulte Criar ações de bootstrap para instalar softwares adicionais.

Tipos de arquivos compatíveis

O DistributedCache permite arquivos únicos e arquivamentos. Arquivos individuais são armazenados em cache como somente leitura. Executáveis e arquivos binários têm permissões de execução definidas.

Os arquivamentos são um ou mais arquivos empacotados por meio de um utilitário, como o `gzip`. O DistributedCache passa os arquivos compactados para cada nó central e descompacta o arquivo como parte do armazenamento em cache. O DistributedCache é compatível com os seguintes formatos de compactação:

- zip
- tgz
- tar.gz
- tar
- jar

Local dos arquivos em cache

O DistributedCache copia arquivos apenas para nós centrais. Se não houver um nó central no cluster, o DistributedCache copiará os arquivos para o nó primário.

O DistributedCache associa os arquivos de cache ao diretório de trabalho atual do mapper e do reducer usando links simbólicos. Um link simbólico é um alias para uma localização de arquivo, e não essa localização propriamente dita. O valor do parâmetro, `yarn.nodemanager.local-dirs` em `yarn-site.xml`, especifica a localização dos arquivos temporários. A Amazon EMR define esse parâmetro como`/mnt/mapred`, ou alguma variação, com base no tipo e na EMR versão da instância. Por exemplo, uma configuração pode ter `/mnt/mapred` e `/mnt1/mapred` porque o tipo de instância tem dois volumes temporários. Arquivos de cache estão localizados em um subdiretório da localização de arquivo temporária em `/mnt/mapred/taskTracker/archive`.

Se você armazenar um único arquivo em cache, o DistributedCache colocará o arquivo no diretório `archive`. Se você armazenar um arquivamento em cache, o DistributedCache descompactará o arquivo e criará um subdiretório em `/archive` com o mesmo nome do arquivamento. Os arquivos individuais estão localizados no novo subdiretório.

Você pode usar o DistributedCache somente ao utilizar o Streaming.

Acessar arquivos em cache de aplicações de transmissão

Para acessar os arquivos em cache dos seus aplicativos de mapeador ou redutor, certifique-se de ter adicionado o diretório de trabalho atual (. /) ao caminho do aplicativo e referenciado os arquivos em cache como se estivessem presentes no diretório de trabalho atual.

Acessar arquivos em cache de aplicações de transmissão

Você pode usar o AWS Management Console e o AWS CLI para criar clusters que usam o cache distribuído.

Console

Especificar arquivos de cache distribuído usando o novo console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Em Etapas, escolha Adicionar etapa. Isso abrirá a caixa de diálogo Adicionar etapa. No campo Argumentos, inclua os arquivos e arquivamentos para salvar no cache. O tamanho do arquivo (ou o tamanho total dos arquivos em um arquivamento) deve ser menor que o tamanho do cache alocado.

   Para adicionar um arquivo individual ao cache distribuído, especifique `-cacheFile` seguido do nome e do local do arquivo, do sinal de cerquilha (#) e do nome que você deseja dar ao arquivo quando ele for inserido no cache local. O exemplo a seguir demonstra como adicionar um arquivo individual ao cache distribuído.

   ```
   -cacheFile \
   s3://DOC-EXAMPLE-BUCKET/file-name#cache-file-name
   ```

Para adicionar um arquivo ao cache distribuído, insira `-cacheArchive` seguido da localização dos arquivos no Amazon S3, do sinal de cerquilha (#) e do nome que você deseja dar à coleção de arquivos no cache local. O exemplo a seguir demonstra como adicionar um arquivo de arquivamento ao cache distribuído.

```
-cacheArchive \
s3://DOC-EXAMPLE-BUCKET/archive-name#cache-archive-name
```

Insira os valores correspondentes nos outros campos da caixa de diálogo. As opções diferem dependendo do tipo de etapa. Para adicionar a etapa e sair da caixa de diálogo, escolha Adicionar etapa.

4. Escolha qualquer outra opção que se aplique ao cluster.
5. Para iniciar o cluster, escolha Criar cluster.

CLI

Para especificar arquivos de cache distribuídos com o AWS CLI

- Para enviar uma etapa de Streaming quando um cluster é criado, digite o comando `create-cluster` com o parâmetro `--steps`. Para especificar arquivos de cache distribuídos usando o AWS CLI, especifique os argumentos apropriados ao enviar uma etapa de streaming.

  Para adicionar um arquivo individual ao cache distribuído, especifique `-cacheFile` seguido do nome e do local do arquivo, do sinal de cerquilha (#) e do nome que você deseja dar ao arquivo quando ele for inserido no cache local.

  Para adicionar um arquivo ao cache distribuído, insira `-cacheArchive` seguido da localização dos arquivos no Amazon S3, do sinal de cerquilha (#) e do nome que você deseja dar à coleção de arquivos no cache local. O exemplo a seguir demonstra como adicionar um arquivo de arquivamento ao cache distribuído.

  Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consultehttps://docs.aws.amazon.com/cli/latest/reference/emr.

Example 1

Digite o seguinte comando para executar um cluster e envie um etapa de Streaming que use `-cacheFile` para adicionar um arquivo, `sample_dataset_cached.dat`, ao cache.

```
aws emr create-cluster --name "Test cluster" --release-label emr-4.0.0 --
applications Name=Hive Name=Pig --use-default-roles --ec2-attributes KeyName=myKey
 --instance-type m5.xlarge --instance-count 3 --steps Type=STREAMING,Name="Streaming
 program",ActionOnFailure=CONTINUE,Args=["--files","s3://my_bucket/my_mapper.py
 s3://my_bucket/my_reducer.py","-mapper","my_mapper.py","-reducer","my_reducer.py,"-
input","s3://my_bucket/my_input","-output","s3://my_bucket/my_output", "-
cacheFile","s3://my_bucket/sample_dataset.dat#sample_dataset_cached.dat"]
```

Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó primário é executado, e as instâncias restantes são executadas como nós centrais. Todos os nós usarão o tipo de instância especificado no comando.

Se você ainda não criou a função de EMR serviço e o perfil de EC2 instância padrão, digite `aws emr create-default-roles` para criá-los antes de digitar o `create-cluster` subcomando.

Example 2

O comando a seguir mostra a criação de um cluster de streaming e usa `-cacheArchive` para adicionar um arquivamento de arquivos ao cache.

```
aws emr create-cluster --name "Test cluster" --release-label emr-4.0.0 --
applications Name=Hive Name=Pig --use-default-roles --ec2-attributes KeyName=myKey
 --instance-type m5.xlarge --instance-count 3 --steps Type=STREAMING,Name="Streaming
 program",ActionOnFailure=CONTINUE,Args=["--files","s3://my_bucket/my_mapper.py
 s3://my_bucket/my_reducer.py","-mapper","my_mapper.py","-reducer","my_reducer.py,"-
input","s3://my_bucket/my_input","-output","s3://my_bucket/my_output", "-
cacheArchive","s3://my_bucket/sample_dataset.tgz#sample_dataset_cached"]
```

Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó primário é executado, e as instâncias restantes são executadas como nós centrais. Todos os nós usarão o tipo de instância especificado no comando.

Se você ainda não criou a função de EMR serviço e o perfil de EC2 instância padrão, digite `aws emr create-default-roles` para criá-los antes de digitar o `create-cluster` subcomando.

Como processar arquivos compactados

O Hadoop verifica a extensão do arquivo para detectar arquivos compactados. Os tipos de compactação suportados pelo Hadoop são: gzip, bzip2 e. LZO Você não precisa tomar medidas adicionais para extrair arquivos usando esses tipos de compactação; o Hadoop manipula o processo para você.

Para indexar LZO arquivos, você pode usar a biblioteca hadoop-lzo, que pode ser baixada do hadoop-lzo. https://github.com/kevinweil/ Observe que, como essa é uma biblioteca de terceiros, EMR a Amazon não oferece suporte ao desenvolvedor sobre como usar essa ferramenta. Para obter informações de uso, consulte o arquivo leiame da hadoop-lzo.

Importar dados do DynamoDB para o Hive

A implementação do Hive fornecida pela Amazon EMR inclui funcionalidades que você pode usar para importar e exportar dados entre o DynamoDB e um cluster da Amazon. EMR Isso é útil quando seus dados de entrada estão armazenados no DynamoDB. Para obter mais informações, consulte Exportar, importar, consultar e unir tabelas no DynamoDB usando a Amazon. EMR

Conectar-se aos dados usando o AWS Direct Connect

AWS Direct Connect é um serviço que você pode usar para estabelecer uma conexão de rede privada dedicada com a Amazon Web Services a partir do seu data center, escritório ou ambiente de colocation. Se você tiver grandes quantidades de dados de entrada, o uso AWS Direct Connect pode reduzir seus custos de rede, aumentar a taxa de transferência da largura de banda e fornecer uma experiência de rede mais consistente do que as conexões baseadas na Internet. Para obter mais informações, consulte o Guia do usuário do AWS Direct Connect.

Carregar grandes quantidades de dados usando o AWS Snowball

AWS Snowball é um serviço que você pode usar para transferir grandes quantidades de dados entre o Amazon Simple Storage Service (Amazon S3) e seu local de armazenamento de dados no local em alta velocidade. faster-than-internet O Snowball oferece suporte a dois tipos de trabalho: trabalhos de importação e trabalhos de exportação. Os trabalhos de importação envolvem a transferência de dados de uma fonte on-premises para um bucket do Amazon S3. Os trabalhos de exportação envolvem a transferência de dados de um bucket do Amazon S3 para uma fonte on-premises. Para ambos os tipos de trabalho, os dispositivos Snowball protegem seus dados, enquanto as transportadoras regionais os transportam entre o Amazon S3 e a local do armazenamento de dados. Os dispositivos Snowball são fisicamente robustos e protegidos pelo (). AWS Key

Management Service AWS KMS Para obter mais informações, consulte o Guia do desenvolvedor do AWS Snowball Edge.

Configurar um local de saída

O formato de saída mais comum de um EMR cluster da Amazon é como arquivos de texto, compactados ou não compactados. Normalmente, esse arquivo é gravado em um bucket do Amazon S3. Esse bucket deve ser criado antes de você iniciar o cluster. Você especifica o bucket do S3 como o local de saída quando inicia o cluster.

Para obter mais informações, consulte os tópicos a seguir.

Tópicos

- Criar e configurar um bucket do Amazon S3
- Quais formatos a Amazon pode EMR devolver?
- Como gravar dados em um bucket do Amazon S3 do qual você não é proprietário
- Compactar a saída do cluster

Criar e configurar um bucket do Amazon S3

A Amazon EMR (AmazonEMR) usa o Amazon S3 para armazenar dados de entrada, arquivos de log e dados de saída. O Amazon S3 se refere a esses locais de armazenamento como bucket. Os buckets têm certas restrições e limitações para se adequarem aos requisitos e aos requisitos do Amazon DNS S3. Para obter mais informações, acesse Restrições e limitações de bucket no Guia do desenvolvedor do Amazon Simple Storage Service.

Para criar um bucket do Amazon S3, siga as instruções da página Criação de um bucket no Guia do desenvolverdor do Amazon Simple Storage Service.

> **Note**
>
> Se você habilitar o registro em log no assistente Create a Bucket (Criar um bucket), ele só permitirá logs de acesso do bucket, e não logs de cluster.

> **Note**
>
> Para obter mais informações sobre a especificação de buckets específicos da região, consulte Buckets and Regions no Amazon Simple Storage Service Developer Guide e os endpoints regionais disponíveis para o. AWS SDKs

Depois de criar o bucket, você poderá definir as permissões apropriadas. Normalmente, você atribui a si (o proprietário) acesso de leitura e gravação. É altamente recomendável seguir as Práticas recomendadas de segurança para o Amazon S3 ao configurar o bucket.

Os buckets do Amazon S3 obrigatórios devem existir para que você possa criar um cluster. Você deve carregar todos os scripts necessários ou dados referenciados no cluster no Amazon S3. A tabela a seguir descreve dados de exemplo, scripts e locais de arquivo de log.

Informações	Exemplo de local no Amazon S3
script ou programa	`s3://amzn-s3-demo-bucket1/script/MapperScript.py`
arquivos de log	`s3://amzn-s3-demo-bucket1/logs`
dados de entrada	`s3://amzn-s3-demo-bucket1/input`
dados de saída	`s3://amzn-s3-demo-bucket1/output`

Quais formatos a Amazon pode EMR devolver?

O formato de saída padrão para um cluster é texto com pares de chave e valor gravados nas linhas individuais dos arquivos de texto. Este é o formato de saída mais comumente usado.

Se os dados de saída precisam ser gravados em um formato que não seja o de arquivos de texto padrão, você pode usar a interface do Hadoop `OutputFormat` para especificar outros tipos de saída. Você pode até mesmo criar uma subclasse da classe `FileOutputFormat` para tratar os tipos de dados personalizados. Para obter mais informações, consulte http://hadoop.apache.org/docs/current/api/org/apache/hadoop/mapred/ OutputFormat .html.

Se você estiver iniciando um cluster do Hive, poderá usar um serializador/desserializador (SerDe) para gerar dados de um determinado formato. HDFS Para obter mais informações, consulte https://cwiki.apache.org/confluence/display/Hive/ SerDe.

Como gravar dados em um bucket do Amazon S3 do qual você não é proprietário

Ao gravar um arquivo em um bucket do Amazon Simple Storage Service (Amazon S3), por padrão, você é o único usuário capaz de ler esse arquivo. A suposição é a de que você gravará arquivos em seus próprio buckets, e essa configuração padrão protege a privacidade desses arquivos.

No entanto, se você estiver executando um cluster e quiser que a saída seja gravada no bucket Amazon S3 de outro AWS usuário e quiser que esse outro AWS usuário possa ler essa saída, você deve fazer duas coisas:

- Faça com que o outro AWS usuário conceda a você permissões de gravação para o bucket do Amazon S3. O cluster que você executa é executado sob suas AWS credenciais, portanto, qualquer cluster que você iniciar também poderá gravar no bucket desse outro AWS usuário.
- Defina permissões de leitura para o outro AWS usuário nos arquivos que você ou o cluster gravam no bucket do Amazon S3. A maneira mais fácil de definir essas permissões de leitura é usar listas de controle de acesso predefinidas (ACLs), um conjunto de políticas de acesso predefinidas definidas pelo Amazon S3.

Para obter informações sobre como o outro AWS usuário pode conceder a você permissões para gravar arquivos no bucket do Amazon S3 do outro usuário, consulte Editando permissões do bucket no Guia do usuário do Amazon Simple Storage Service.

Para que seu cluster use o padrão ACLs ao gravar arquivos no Amazon S3, defina a opção de configuração `fs.s3.canned.acl` do cluster como o padrão a ACL ser usado. A tabela a seguir lista os enlatados ACLs atualmente definidos.

Enlatado ACL	Descrição
`AuthenticatedRead`	Especifica que o proprietário recebe `Permission.FullControl` e o favorecido do grupo `GroupGrantee.AuthenticatedUsers` recebe o acesso `Permission.Read` .

Enlatado ACL	Descrição
`BucketOwnerFullControl`	Especifica que o proprietário do bucket recebe `Permission.FullControl` . O proprietário do bucket não é necessariamente o proprietário do objeto.
`BucketOwnerRead`	Especifica que o proprietário do bucket recebe `Permission.Read` . O proprietário do bucket não é necessariamente o proprietário do objeto.
`LogDeliveryWrite`	Especifica que o proprietário recebe `Permissio n.FullControl` e o favorecido do grupo `GroupGran tee.LogDelivery` recebe o acesso `Permissio n.Write` , permitindo que logs de acesso sejam fornecidos.
`Private`	Especifica que o proprietário recebe `Permissio n.FullControl` .
`PublicRead`	Especifica que o proprietário recebe `Permissio n.FullControl` e o favorecido do grupo `GroupGran tee.AllUsers` recebe o acesso `Permissio n.Read` .
`PublicReadWrite`	Especifica que o proprietário recebe `Permissio n.FullControl` e o favorecido do grupo `GroupGran tee.AllUsers` recebe os acessos `Permissio n.Read` e `Permission.Write` .

Há muitas maneiras de definir opções de configuração do cluster, dependendo do tipo de cluster que você está executando. Os procedimentos a seguir mostram como definir a opção para casos comuns.

Para gravar arquivos usando canned ACLs no Hive

- No prompt de comando do Hive, defina a opção de `fs.s3.canned.acl` configuração como a configuração padrão ACL que você deseja que o cluster defina nos arquivos que ele grava no Amazon S3. Para acessar o prompt de comando do HiveSSH, conecte-se ao nó principal

usando e digite Hive no prompt de comando do Hadoop. Para obter mais informações, consulte Conecte-se ao nó primário usando SSH.

O exemplo a seguir define a configuração de opção `fs.s3.canned.acl` como `BucketOwnerFullControl`, que dá ao proprietário do bucket do Amazon S3 controle total sobre o arquivo. Observe que o comando definido faz distinção entre maiúsculas e minúsculas e não contém aspas ou espaços.

```
hive> set fs.s3.canned.acl=BucketOwnerFullControl;
create table acl (n int) location 's3://acltestbucket/acl/';
insert overwrite table acl select count(*) from acl;
```

As duas últimas linhas do exemplo criam uma tabela que é armazenada no Amazon S3 e gravam dados nessa tabela.

Para gravar arquivos usando canned ACLs in Pig

- No prompt de comando do Pig, defina a opção de `fs.s3.canned.acl` configuração como a configuração padrão ACL que você deseja que o cluster defina nos arquivos que ele grava no Amazon S3. Para acessar o prompt de comando do PigSSH, conecte-se ao nó principal usando e digite Pig no prompt de comando do Hadoop. Para obter mais informações, consulte Conecte-se ao nó primário usando SSH.

  O exemplo a seguir define a opção de `fs.s3.canned.acl` configuração como BucketOwnerFullControl, o que dá ao proprietário do bucket do Amazon S3 controle total sobre o arquivo. Observe que o comando set inclui um espaço antes do ACL nome padrão e não contém aspas.

  ```
  pig> set fs.s3.canned.acl BucketOwnerFullControl;
  store some data into 's3://acltestbucket/pig/acl';
  ```

Para gravar arquivos usando canned ACLs em um formato personalizado JAR

- Defina a opção de configuração `fs.s3.canned.acl` usando o Hadoop com o sinalizador -D. Isso é mostrado no exemplo a seguir.

```
hadoop jar hadoop-examples.jar wordcount
-Dfs.s3.canned.acl=BucketOwnerFullControl s3://mybucket/input s3://mybucket/output
```

Compactar a saída do cluster

Tópicos

Compactação de dados de saída

Isso comprime a saída de seu trabalho do Hadoop. Se você estiver usando TextOutputFormat o resultado é um arquivo de texto compactado com gzip. Se você estiver escrevendo para SequenceFiles , o resultado SequenceFile será comprimido internamente. Para habilitar isso, defina a configuração mapred.output.compress como true.

Se você estiver executando um trabalho de streaming, poderá habilitar isso transmitido esses argumentos ao trabalho em questão.

```
-jobconf mapred.output.compress=true
```

Você também pode usar uma ação de bootstrap para compactar automaticamente todas as saídas do trabalho. Veja a seguir como fazer isso com o cliente Ruby.

```
--bootstrap-actions s3://elasticmapreduce/bootstrap-actions/configure-hadoop \
--args "-s,mapred.output.compress=true"
```

Por fim, se você estiver escrevendo um JAR personalizado, poderá habilitar a compactação de saída com a seguinte linha ao criar seu trabalho.

```
FileOutputFormat.setCompressOutput(conf, true);
```

Compactação de dados intermediária

Se o seu trabalho embaralhar uma quantidade significativa de dados dos mapeadores para os reducers, você poderá ver uma melhoria de desempenho ao habilitar a compactação intermediária. Compacta a saída de mapa e a descompacta quando ela chega ao nó core. A definição de configuração é mapred.compress.map.output. Você pode habilitar isso da mesma forma que a compactação de saída.

Ao escrever um JAR personalizado, use o seguinte comando:

```
conf.setCompressMapOutput(true);
```

Usando a biblioteca Snappy com a Amazon EMR

A Snappy é uma biblioteca de compactação e descompactação otimizada para velocidade. Ele está disponível na Amazon EMR AMIs versão 2.0 e posterior e é usado como padrão para compactação intermediária. Para obter mais informações sobre a Snappy, acesse http://code.google.com/p/snappy/.

Planejar e configurar nós primários

Ao iniciar um EMR cluster da Amazon, você pode optar por ter um ou três nós primários em seu cluster. A alta disponibilidade, por exemplo, de frotas é suportada pelas EMR versões 5.36.1, 5.36.2, 6.8.1, 6.9.1, 6.10.1, 6.11.1, 6.12.0 e superiores da Amazon. Para grupos de instâncias, a alta disponibilidade é suportada nas EMR versões 5.23.0 e superiores da Amazon. Para melhorar ainda mais a disponibilidade do cluster, a Amazon EMR pode usar grupos de EC2 posicionamento da Amazon para garantir que os nós primários sejam colocados em um hardware subjacente distinto. Para obter mais informações, consulte EMRIntegração da Amazon com grupos de EC2 colocação.

Um EMR cluster da Amazon com vários nós primários oferece os seguintes benefícios:

- O nó primário não é mais um ponto único de falha. Se um nó primário falhar, o cluster usará os outros dois nós primários e executará sem interrupção. Enquanto isso, a Amazon substitui EMR automaticamente o nó primário com falha por um novo que é provisionado com as mesmas ações de configuração e bootstrap.
- A Amazon EMR habilita os recursos de alta disponibilidade do Hadoop HDFS NameNode YARN ResourceManager e oferece suporte à alta disponibilidade para alguns outros aplicativos de código aberto.

  Para obter mais informações sobre como um EMR cluster da Amazon com vários nós primários oferece suporte a aplicativos de código aberto e outros EMR recursos da Amazon, consulteAplicações e atributo compatíveis.

> **Note**
>
> O cluster pode residir apenas em uma zona de disponibilidade ou sub-rede.

Esta seção fornece informações sobre aplicativos e recursos compatíveis de um EMR cluster da Amazon com vários nós primários, bem como detalhes de configuração, melhores práticas e considerações para iniciar o cluster.

Tópicos

Aplicações e atributo compatíveis

Este tópico fornece informações sobre os recursos de alta disponibilidade do Hadoop de HDFS NameNode e em YARN ResourceManager um EMR cluster da Amazon e como os recursos de alta disponibilidade funcionam com aplicativos de código aberto e outros recursos da Amazon. EMR

Alta disponibilidade HDFS

Um EMR cluster da Amazon com vários nós primários habilita o recurso HDFS NameNode de alta disponibilidade no Hadoop. Para obter mais informações, consulte HDFSAlta disponibilidade.

Em um EMR cluster da Amazon, dois ou mais nós separados são configurados como NameNodes. Um NameNode está em um `active` estado e os outros estão em um `standby` estado. Se o nó com `active` NameNode falha, a Amazon EMR inicia um processo de HDFS failover automático. Um nó `standby` NameNode se torna `active` e assume todas as operações do cliente no cluster. A Amazon EMR substitui o nó com falha por um novo, que então se junta novamente como um. `standby`

> **Note**
>
> Nas EMR versões 5.23.0 da Amazon até 5.30.1, inclusive, apenas dois dos três nós principais são executados. HDFS NameNode

Se precisar descobrir qual NameNode é`active`, você pode usar SSH para se conectar a qualquer nó primário no cluster e executar o seguinte comando:

```
hdfs haadmin -getAllServiceState
```

A saída lista os nós em que NameNode está instalado e seu status. Por exemplo,

```
ip-##-#-#-##1.ec2.internal:8020 active
ip-##-#-#-##2.ec2.internal:8020 standby
ip-##-#-#-##3.ec2.internal:8020 standby
```

Alta disponibilidade YARN ResourceManager

Um EMR cluster da Amazon com vários nós primários habilita o recurso YARN ResourceManager de alta disponibilidade no Hadoop. Para obter mais informações, consulte ResourceManager Alta disponibilidade.

Em um EMR cluster da Amazon com vários nós primários, YARN ResourceManager é executado em todos os três nós primários. Um ResourceManager está no `active` estado e os outros dois estão no `standby` estado. Se o nó primário `active` ResourceManager falhar, EMR a Amazon iniciará um processo de failover automático. Um nó primário com `standby` ResourceManager a assume todas as operações. A Amazon EMR substitui o nó primário que falhou por um novo, que então volta ao ResourceManager quórum como um. `standby`

Você pode se conectar a “http://*master-public-dns-name*:8088/cluster” para qualquer nó primário, que o direciona automaticamente para o gerenciador de recursos. `active` Para descobrir

qual é o gerenciador de recursos`active`, use SSH para se conectar a qualquer nó primário no cluster. Depois, execute o seguinte comando para obter uma lista com os três nós primários e os status deles:

```
yarn rmadmin -getAllServiceState
```

Aplicativos compatíveis em um EMR cluster da Amazon com vários nós primários

Você pode instalar e executar os seguintes aplicativos em um EMR cluster da Amazon com vários nós primários. Para cada aplicação, o processo de failover do nó primário varia.

Aplicativo	Disponibilidade durante failover do nó primário	Observações
Flink	Disponibilidade não afetada pelo failover do nó primário	Os trabalhos do Flink na Amazon EMR são executados como YARN aplicativos. O Flink JobManagers funciona da mesma forma YARN que ApplicationMasters nos nós principais. O não JobManager é afetado pelo processo de failover do nó primário. Se você usa a Amazon EMR versão 5.27.0 ou anterior, esse JobManager é um único ponto de falha. Quando o JobManager falha, ele perde todos os estados de trabalho e não retoma os trabalhos em execução. Você pode habilitar a JobManager alta disponibi lidade configurando a contagem de tentativa s de aplicativos, o checkpoint e ativando o armazenamento ZooKeeper como estado para o Flink. Para obter mais informações, consulte Configurando o Flink em um EMR cluster da Amazon com vários nós primários. A partir da EMR versão 5.28.0 da Amazon, nenhuma configuração manual é necessári a para permitir a JobManager alta disponibi lidade.

Aplicativo	Disponibilidade durante failover do nó primário	Observações
Ganglia	Disponibilidade não afetada pelo failover do nó primário	O Ganglia está disponível em todos os nós primários e, portanto, continua em execução durante o processo de failover do nó primário.
Hadoop	Alta disponibilidade	HDFS NameNode e faça YARN ResourceM anager o failover automaticamente para o nó em espera quando o nó primário ativo falhar.
HBase	Alta disponibilidade	HBaseo failover automático para o nó em espera quando o nó primário ativo falha. Se você estiver se conectando HBase por meio de um servidor REST ou Thrift, deverá alternar para um nó primário diferente quando o nó primário ativo falhar.
HCatalog	Disponibilidade não afetada pelo failover do nó primário	HCatalogé construído sobre o metastore Hive, que existe fora do cluster. HCatalogpermanece disponível durante o processo de failover do nó primário.
JupyterHub	Alta disponibilidade	JupyterHub está instalado em todas as três instâncias principais. É altamente recomendá vel configurar a persistência do caderno para evitar a perda do caderno após uma falha do nó primário. Para obter mais informações, consulte Configuring persistence for notebooks in Amazon S3.

Aplicativo	Disponibilidade durante failover do nó primário	Observações
Livy	Alta disponibilidade	O Livy é instalado em todos os três nós primários. Quando o nó primário ativo falha, você perde o acesso à sessão Livy atual e precisa criar uma nova sessão em outro nó primário ou no novo nó de substituição.
Mahout	Disponibilidade não afetada pelo failover do nó primário	Como o Mahout não tem daemons, ele não é afetado pelo processo de failover do nó primário.
MXNet	Disponibilidade não afetada pelo failover do nó primário	Como não MXNet tem daemon, ele não é afetado pelo processo de failover do nó primário.
Phoenix	Alta disponibilidade	Phoenix QueryServer funciona apenas em um dos três nós primários. O Phoenix em todos os três mestres está configurado para conectar o Phoenix QueryServer. É possível encontrar o IP privado do servidor de consulta do Phoenix usando o arquivo `/etc/phoenix/conf/phoenix-env.sh`
Pig	Disponibilidade não afetada pelo failover do nó primário	Como o Pig não tem daemons, ele não é afetado pelo processo de failover do nó primário.
Spark	Alta disponibilidade	Todos os aplicativos Spark são executados em YARN contêineres e podem reagir ao failover do nó primário da mesma forma que os recursos de alta YARN disponibilidade.

Aplicativo	Disponibilidade durante failover do nó primário	Observações
Sqoop	Alta disponibilidade	Por padrão, sqoop-job e sqoop-metastore armazenam dados (descrições de trabalhos) no disco local do mestre que executa o comando. Se você deseja salvar dados do metastore no banco de dados externo, consulte a documenta ção do Apache Sqoop.
Tez	Alta disponibilidade	Como os contêineres Tez funcionamYARN, o Tez se comporta da mesma forma que YARN durante o processo de failover do nó primário.
TensorFlow	Disponibilidade não afetada pelo failover do nó primário	Como não TensorFlow tem daemon, ele não é afetado pelo processo de failover do nó primário.
Zeppelin	Alta disponibilidade	O Zeppelin é instalado em todos os três nós primários. O Zeppelin armazena notas e configurações do intérprete HDFS por padrão para evitar perda de dados. As sessões de intérprete são completamente isoladas em todas as três instâncias primárias. Os dados da sessão serão perdidos após uma falha da instância mestra. Recomenda-se não modificar a mesma nota simultaneamente em instâncias primárias diferentes.

Aplicativo	Disponibilidade durante failover do nó primário	Observações
ZooKeeper	Alta disponibilidade	ZooKeeper é a base do recurso de failover HDFS automático. ZooKeeper fornece um serviço altamente disponível para manter os dados de coordenação, notificar os clientes sobre alterações nesses dados e monitorar os clientes em busca de falhas. Para obter mais informações, consulte failover HDFS automático.

Para executar os seguintes aplicativos em um EMR cluster da Amazon com vários nós primários, você deve configurar um banco de dados externo. O banco de dados externo existe fora do cluster e torna os dados persistentes durante o processo de failover do nó primário. Para as aplicações a seguir, os componentes de serviço serão recuperados automaticamente durante o processo de failover do nó primário, mas os trabalhos ativos podem falhar e precisam ser repetidos.

Aplicativo	Disponibilidade durante failover do nó primário	Observações
Hive	Alta disponibilidade somente para componentes de serviço	É necessário um metastore externo para o Hive. Essa deve ser uma metastore SQL externa My, pois o Postgre não SQL é compatível com clusters de vários mestres. Para obter mais informações, consulte Configuring an external metastore for Hive.
Hue	Alta disponibilidade somente para componentes de serviço	É necessário um banco de dados externo para o Hue. Para obter mais informações, consulte Usando o Hue com um banco de dados remoto na Amazon RDS.
Oozie		É necessário um banco de dados externo para o Oozie. Para obter mais informações, consulte

Aplicativo	Disponibilidade durante failover do nó primário	Observações
	Alta disponibilidade somente para componentes de serviço	Usando o Oozie com um banco de dados remoto na Amazon RDS. O Oozie-server e o oozie-client são instalados nos três nós primários. Os oozie-clients são configurados para se conectar ao oozie-server correto por padrão.
PrestoDB ou Presto/Trino SQL	Alta disponibilidade somente para componentes de serviço	É necessário um metastore externo do Hive para o PrestoDB (Presto na Amazon EMR 6.1.0-6.3.0 ou SQL Trino na Amazon 6.4.0 e versões posteriores). EMR Você pode usar o Presto com o AWS Glue Data Catalog ou usar um SQL banco de dados externo My for Hive. O Presto CLI é instalado em todos os três nós primários para que você possa usá-lo para acessar o Coordenador do Presto a partir de qualquer um dos nós primários. O coordenador do Presto é instalado em apenas um nó primário. Você pode encontrar o DNS nome do nó primário em que o Presto Coordinator está instalado ligando para a Amazon EMR `describe-cluster` API e lendo o valor retornado do `MasterPublicDnsName` campo na resposta.

> **Note**
>
> Quando um nó primário falha, sua conectividade de banco de dados Java (JDBC) ou conectividade de banco de dados aberta (ODBC) encerra sua conexão com o nó primário. Você pode se conectar a qualquer um dos nós primários restantes para continuar o trabalho,

pois o daemon do Hive Metastore é executado em todos os nós primários. Ou você pode esperar a substituição do nó primário com falha.

Como os EMR recursos da Amazon funcionam em um cluster com vários nós primários

Conectando-se aos nós primários usando SSH

Você pode se conectar a qualquer um dos três nós principais em um EMR cluster da Amazon usando SSH da mesma forma que se conecta a um único nó primário. Para obter mais informações, consulte Conectar-se ao nó primário usando SSH.

Se um nó primário falhar, sua SSH conexão com esse nó primário será encerrada. Para que ela continue funcionando, você pode se conectar a um dos outros dois nós primários. Como alternativa, você pode acessar o novo nó primário depois que EMR a Amazon substituir o que falhou por um novo.

> **Note**
>
> O endereço IP privado para o nó primário de substituição permanece o mesmo que o anterior. O endereço IP público para o nó primário de substituição pode mudar. Você pode recuperar os novos endereços IP no console ou usando o `describe-cluster` comando no AWS CLI.
> NameNode só é executado em dois dos nós primários. No entanto, você pode executar `hdfs` CLI comandos e operar trabalhos para acessar HDFS em todos os três nós principais.

Trabalhando com etapas em um EMR cluster da Amazon com vários nós primários

Você pode enviar etapas para um EMR cluster da Amazon com vários nós primários da mesma forma que trabalha com etapas em um cluster com um único nó primário. Para obter mais informações, consulte Submit work to a cluster.

A seguir estão algumas considerações para trabalhar com etapas em um EMR cluster da Amazon com vários nós primários:

- Se um nó primário falhar, as etapas em execução no nó primário serão marcadas comoFAILED. Todos os dados que foram gravados localmente são perdidos. No entanto, o status FAILED pode não refletir o estado real das etapas.
- Se uma etapa em execução tiver iniciado um YARN aplicativo quando o nó primário falhar, a etapa poderá continuar e ser bem-sucedida devido ao failover automático do nó primário.
- Recomendamos que você verifique os status das etapas consultando a saída dos trabalhos. Por exemplo, os MapReduce trabalhos usam um _SUCCESS arquivo para determinar se o trabalho foi concluído com êxito.
- É recomendável que você defina o ActionOnFailure parâmetro comoCONTINUE, ou CANCEL _ AND _WAIT, em vez de TERMINATE JOB _ _ FLOW ou TERMINATE _CLUSTER.

Proteção automática de término

A Amazon habilita EMR automaticamente a proteção contra terminação para todos os clusters com vários nós primários e substitui todas as configurações de execução de etapas que você fornece ao criar o cluster. É possível desabilitar a proteção contra término depois que o cluster é iniciado. Consulte Configurar a proteção contra término para clusters em execução. Para desligar um cluster com múltiplos nós primários, primeiro é necessário modificar os atributos do cluster para desabilitar a proteção contra término. Para obter instruções, consulte Encerrar um EMR cluster da Amazon com vários nós primários.

Para obter mais informações sobre proteção contra término, consulte Usar a proteção contra término.

Recursos não suportados em um EMR cluster da Amazon com vários nós primários

No momento, os seguintes EMR recursos da Amazon não estão disponíveis em um EMR cluster da Amazon com vários nós primários:

- EMRCadernos
- Acesso com um clique ao servidor de histórico persistente do Spark
- Interfaces do usuário de aplicações persistentes
- Atualmente, o acesso com um clique às interfaces de usuário de aplicativos persistentes não está disponível para EMR clusters da Amazon com vários nós primários ou para EMR clusters da Amazon integrados ao AWS Lake Formation.

> **Note**
>
> Para usar a autenticação Kerberos em seu cluster, você deve configurar uma externa. KDC A partir da EMR versão 5.27.0 da Amazon, você pode configurar a criptografia HDFS transparente em um EMR cluster da Amazon com vários nós primários. Para obter mais informações, consulte Criptografia transparente HDFS na Amazon EMR.

Inicie um Amazon EMR Cluster com vários nós primários

Este tópico fornece detalhes de configuração e exemplos para iniciar um EMR cluster da Amazon com vários nós primários.

> **Note**
>
> A Amazon habilita EMR automaticamente a proteção de encerramento para todos os clusters que têm vários nós primários e substitui todas as configurações de encerramento automático que você fornece ao criar o cluster. Para desligar um cluster com múltiplos nós primários, primeiro é necessário modificar os atributos do cluster para desabilitar a proteção contra término. Para obter instruções, consulte Encerrar um EMR cluster da Amazon com vários nós primários.

Pré-requisitos

- Você pode iniciar um EMR cluster da Amazon com vários nós primários em VPC sub-redes públicas e privadas. EC2-Classic não é suportado. Para iniciar um EMR cluster da Amazon com vários nós primários em uma sub-rede pública, você deve permitir que as instâncias dessa sub-rede recebam um endereço IP público selecionando Atribuição automática IPv4 no console ou executando o comando a seguir. Substituir *22XXXX01* com seu ID de sub-rede.

  ```
  aws ec2 modify-subnet-attribute --subnet-id subnet-22XXXX01 --map-public-ip-on-launch
  ```

- Para executar o Hive, o Hue ou o Oozie em um EMR cluster da Amazon com vários nós primários, você deve criar um metastore externo. Para obter mais informações, consulte Configurando um metastore externo para o Hive, Usando o Hue com um banco de dados remoto na Amazon RDS ou Apache Oozie.

- Para usar a autenticação Kerberos em seu cluster, você deve configurar uma externa. KDC Para obter mais informações, consulte Configurando o Kerberos na Amazon Amazon. EMR

Inicie um Amazon EMR Cluster com vários nós primários

Você pode executar um cluster com vários nós primários ao usar grupos ou frotas de instâncias. Ao usar os grupos de instâncias com vários nós primários, é preciso especificar um valor `3` de contagem de instâncias para o grupo de instâncias do nó primário. Ao usar frotas de instâncias com vários nós primários, você deve especificar a `TargetOnDemandCapacity` de `3`, a `TargetSpotCapacity` de `0` para a frota de instâncias primária e a `WeightedCapacity` de `1` para cada tipo de instância que configurar para a frota principal.

Os exemplos a seguir demonstram como iniciar o cluster usando o padrão AMI ou um personalizado AMI com grupos de instâncias e frotas de instâncias:

> **Note**
>
> Você deve especificar o ID da sub-rede ao iniciar um EMR cluster da Amazon com vários nós primários usando o. AWS CLI Substituir *22XXXX01* e *22XXXX02* com seu ID de sub-rede nos exemplos a seguir.

Default AMI, instance groups

Example Exemplo — Lançamento de um cluster de grupos de EMR instâncias da Amazon com vários nós primários usando um padrão AMI

```
aws emr create-cluster \
--name "ha-cluster" \
--release-label emr-6.15.0 \
--instance-groups InstanceGroupType=MASTER,InstanceCount=3,InstanceType=m5.xlarge
 InstanceGroupType=CORE,InstanceCount=4,InstanceType=m5.xlarge \
--ec2-attributes
 KeyName=ec2_key_pair_name,InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-22XXXX01
 \
--service-role EMR_DefaultRole \
--applications Name=Hadoop Name=Spark
```

Default AMI, instance fleets

Example Exemplo — Lançamento de um cluster de frota de EMR instâncias da Amazon com vários nós primários usando um padrão AMI

```
aws emr create-cluster \
--name "ha-cluster" \
--release-label emr-6.15.0 \
--instance-fleets '[
    {
        "InstanceFleetType": "MASTER",
        "TargetOnDemandCapacity": 3,
        "TargetSpotCapacity": 0,
        "LaunchSpecifications": {
            "OnDemandSpecification": {
                "AllocationStrategy": "lowest-price"
            }
        },
        "InstanceTypeConfigs": [
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.xlarge"
            },
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.2xlarge"
            },
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.4xlarge"
            }
        ],
        "Name": "Master - 1"
    },
    {
        "InstanceFleetType": "CORE",
        "TargetOnDemandCapacity": 5,
        "TargetSpotCapacity": 0,
        "LaunchSpecifications": {
            "OnDemandSpecification": {
                "AllocationStrategy": "lowest-price"
```

```
            }
        },
        "InstanceTypeConfigs": [
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.xlarge"
            },
            {
                "WeightedCapacity": 2,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.2xlarge"
            },
            {
                "WeightedCapacity": 4,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.4xlarge"
            }
        ],
        "Name": "Core - 2"
    }
]' \
 --ec2-attributes '{"InstanceProfile":"EMR_EC2_DefaultRole","SubnetIds":
["subnet-22XXXX01", "subnet-22XXXX02"]}' \
--service-role EMR_DefaultRole \
--applications Name=Hadoop Name=Spark
```

Custom AMI, instance groups

Example Exemplo — Lançamento de um cluster de grupos de EMR instâncias da Amazon com vários nós primários usando um cluster personalizado AMI

```
aws emr create-cluster \
--name "custom-ami-ha-cluster" \
--release-label emr-6.15.0 \
--instance-groups InstanceGroupType=MASTER,InstanceCount=3,InstanceType=m5.xlarge
 InstanceGroupType=CORE,InstanceCount=4,InstanceType=m5.xlarge \
--ec2-attributes
 KeyName=ec2_key_pair_name,InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-22XXXX01
 \
--service-role EMR_DefaultRole \
--applications Name=Hadoop Name=Spark \
--custom-ami-id ami-MyAmiID
```

Custom AMI, instance fleets

Example Exemplo — Lançamento de um cluster de frota de EMR instâncias da Amazon com vários nós primários usando um cluster personalizado AMI

```
aws emr create-cluster \
--name "ha-cluster" \
--release-label emr-6.15.0 \
--instance-fleets '[
    {
        "InstanceFleetType": "MASTER",
        "TargetOnDemandCapacity": 3,
        "TargetSpotCapacity": 0,
        "LaunchSpecifications": {
            "OnDemandSpecification": {
                "AllocationStrategy": "lowest-price"
            }
        },
        "InstanceTypeConfigs": [
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.xlarge"
            },
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.2xlarge"
            },
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.4xlarge"
            }
        ],
        "Name": "Master - 1"
    },
    {
        "InstanceFleetType": "CORE",
        "TargetOnDemandCapacity": 5,
        "TargetSpotCapacity": 0,
        "LaunchSpecifications": {
            "OnDemandSpecification": {
                "AllocationStrategy": "lowest-price"
```

```
            }
        },
        "InstanceTypeConfigs": [
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.xlarge"
            },
            {
                "WeightedCapacity": 2,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.2xlarge"
            },
            {
                "WeightedCapacity": 4,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.4xlarge"
            }
        ],
        "Name": "Core - 2"
    }
]' \
--ec2-attributes '{"InstanceProfile":"EMR_EC2_DefaultRole","SubnetIds":
["subnet-22XXXX01", "subnet-22XXXX02"]}' \
--service-role EMR_DefaultRole \
--applications Name=Hadoop Name=Spark \
--custom-ami-id ami-MyAmiID
```

Encerrar um EMR cluster da Amazon com vários nós primários

Para encerrar um EMR cluster da Amazon com vários nós primários, você deve desativar a proteção contra encerramento antes de encerrar o cluster, conforme demonstra o exemplo a seguir. Substituir *j-3KVTXXXXXX7UG* com seu ID de cluster.

```
aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --no-termination-
protected
aws emr terminate-clusters --cluster-id j-3KVTXXXXXX7UG
```

EMRIntegração da Amazon com grupos de EC2 colocação

Ao iniciar um cluster de EMR vários nós primários da Amazon na AmazonEC2, você tem a opção de usar estratégias de grupos de posicionamento para especificar como deseja que as instâncias de nós primários sejam implantadas para se protegerem contra falhas de hardware.

As estratégias de grupos de posicionamento são suportadas a partir da EMR versão 5.23.0 da Amazon como uma opção para vários clusters de nós primários. Atualmente, somente os tipos de nós primários são compatíveis com a estratégia de grupo de posicionamento, e a estratégia `SPREAD` é aplicada a estes nós primários. A estratégia `SPREAD` posiciona um pequeno grupo de instâncias em um hardware subjacente separado para evitar a perda de múltiplos nós primários em caso de falha de hardware. Uma solicitação de inicialização de instância poderá falhar se não houver hardware exclusivo suficiente para atender à solicitação. Para obter mais informações sobre estratégias e limitações de EC2 posicionamento, consulte Grupos de posicionamento no Guia EC2 do usuário para instâncias Linux.

Há um limite inicial da Amazon EC2 de 500 clusters habilitados para estratégias de grupos de posicionamento que podem ser lançados por AWS região. Entre em contato com o AWS suporte para solicitar um aumento no número de grupos de colocação permitidos. Você pode identificar os grupos de EC2 posicionamento que a Amazon EMR cria rastreando o par de valores-chave que a Amazon EMR associa à estratégia de grupos de EMR posicionamento da Amazon. Para obter mais informações sobre tags de instância de EC2 cluster, consulteVeja instâncias de cluster na Amazon EC2.

Anexando a política gerenciada por grupos de posicionamento à Amazon EMRrole

A estratégia de grupos de posicionamento exige uma política gerenciada chamada`AmazonElasticMapReducePlacementGroupPolicy`, que permite EMR à Amazon criar, excluir e descrever grupos de posicionamento na AmazonEC2. Você deve `AmazonElasticMapReducePlacementGroupPolicy` se associar à função de serviço da Amazon EMR antes de iniciar um EMR cluster da Amazon com vários nós primários.

Como alternativa, você pode anexar a política `AmazonEMRServicePolicy_v2` gerenciada à função de EMR serviço da Amazon em vez da política gerenciada do grupo de posicionamento. `AmazonEMRServicePolicy_v2`permite o mesmo acesso aos grupos de colocação na Amazon EC2 que `AmazonElasticMapReducePlacementGroupPolicy` o. Para obter mais informações, consulte Função de serviço para a Amazon EMR (EMRfunção).

A política `AmazonElasticMapReducePlacementGroupPolicy` gerenciada é o seguinte JSON texto criado e administrado pela AmazonEMR.

> **Note**
>
> Como a política `AmazonElasticMapReducePlacementGroupPolicy` gerenciada é atualizada automaticamente, a política mostrada aqui pode ser out-of-date. Use o AWS Management Console para visualizar a política atual.

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Resource":"*",
         "Effect":"Allow",
         "Action":[
            "ec2:DeletePlacementGroup",
            "ec2:DescribePlacementGroups"
         ]
      },
      {
         "Resource":"arn:aws:ec2:*:*:placement-group/pg-*",
         "Effect":"Allow",
         "Action":[
            "ec2:CreatePlacementGroup"
         ]
      }
   ]
}
```

Inicie um EMR cluster da Amazon com vários nós primários usando a estratégia de grupos de posicionamento

Para iniciar um EMR cluster da Amazon que tenha vários nós primários com uma estratégia de grupo de posicionamento, anexe a política gerenciada do grupo de posicionamento `AmazonElasticMapReducePlacementGroupPolicy` à EMR função da Amazon. Para obter mais informações, consulte Anexando a política gerenciada por grupos de posicionamento à Amazon EMRrole.

Toda vez que você usa essa função para iniciar um EMR cluster da Amazon com vários nós primários, a Amazon EMR tenta lançar um cluster com a `SPREAD` estratégia aplicada aos seus nós primários. Se você usar uma função que não tenha a política gerenciada por grupos de posicionamento `AmazonElasticMapReducePlacementGroupPolicy` vinculada a ela, a Amazon EMR tentará lançar um EMR cluster da Amazon que tenha vários nós primários sem uma estratégia de grupos de posicionamento.

Se você iniciar um EMR cluster da Amazon que tenha vários nós primários com o `placement-group-configs` parâmetro usando o Amazon EMRAPI ouCLI, a Amazon EMR só iniciará o cluster se a Amazon EMRrole tiver a política gerenciada por grupos de posicionamento `AmazonElasticMapReducePlacementGroupPolicy` anexada. Se a Amazon EMRrole não tiver a política anexada, o EMR cluster da Amazon com vários nós primários iniciados falhará.

Amazon EMR API

Example Exemplo — Use uma estratégia de grupos de posicionamento para iniciar um cluster de grupos de instâncias com vários nós primários da Amazon EMR API

Ao usar a RunJobFlow ação para criar um EMR cluster da Amazon com vários nós primários, defina a `PlacementGroupConfigs` propriedade da seguinte forma. Atualmente, o perfil de instância `MASTER` usa automaticamente `SPREAD` como estratégia de grupo de posicionamento.

```
{
   "Name":"ha-cluster",
   "PlacementGroupConfigs":[
      {
         "InstanceRole":"MASTER"
      }
   ],
   "ReleaseLabel": emr-6.15.0,
   "Instances":{
      "ec2SubnetId":"subnet-22XXXX01",
      "ec2KeyName":"ec2_key_pair_name",
      "InstanceGroups":[
         {
            "InstanceCount":3,
            "InstanceRole":"MASTER",
            "InstanceType":"m5.xlarge"
         },
         {
            "InstanceCount":4,
            "InstanceRole":"CORE",
```

```
            "InstanceType":"m5.xlarge"
         }
      ]
   },
   "JobFlowRole":"EMR_EC2_DefaultRole",
   "ServiceRole":"EMR_DefaultRole"
}
```

- Substituir *ha-cluster* com o nome do seu cluster de alta disponibilidade.
- Substituir *subnet-22XXXX01* com seu ID de sub-rede.
- Substitua o *ec2_key_pair_name* com o nome do seu EC2 key pair para esse cluster. EC2o key pair é opcional e só é necessário se você quiser usá-lo SSH para acessar seu cluster.

AWS CLI

Example Exemplo: usar uma estratégia de grupos de posicionamento para executar um cluster de frotas de instâncias com vários nós primários da AWS Command Line Interface

Ao usar a RunJobFlow ação para criar um EMR cluster da Amazon com vários nós primários, defina a `PlacementGroupConfigs` propriedade da seguinte forma. Atualmente, o perfil de instância `MASTER` usa automaticamente `SPREAD` como estratégia de grupo de posicionamento.

```
aws emr create-cluster \
--name "ha-cluster" \
--placement-group-configs InstanceRole=MASTER \
--release-label emr-6.15.0 \
--instance-fleets '[
    {
        "InstanceFleetType": "MASTER",
        "TargetOnDemandCapacity": 3,
        "TargetSpotCapacity": 0,
        "LaunchSpecifications": {
            "OnDemandSpecification": {
                "AllocationStrategy": "lowest-price"
            }
        },
        "InstanceTypeConfigs": [
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.xlarge"
```

```
            },
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.2xlarge"
            },
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.4xlarge"
            }
        ],
        "Name": "Master - 1"
    },
    {
        "InstanceFleetType": "CORE",
        "TargetOnDemandCapacity": 5,
        "TargetSpotCapacity": 0,
        "LaunchSpecifications": {
            "OnDemandSpecification": {
                "AllocationStrategy": "lowest-price"
            }
        },
        "InstanceTypeConfigs": [
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.xlarge"
            },
            {
                "WeightedCapacity": 2,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.2xlarge"
            },
            {
                "WeightedCapacity": 4,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.4xlarge"
            }
        ],
        "Name": "Core - 2"
    }
]' \
--ec2-attributes '{
```

```
    "KeyName": "ec2_key_pair_name",
    "InstanceProfile": "EMR_EC2_DefaultRole",
    "SubnetIds": [
        "subnet-22XXXX01",
        "subnet-22XXXX02"
    ]
}' \
--service-role EMR_DefaultRole \
--applications Name=Hadoop Name=Spark
```

- Substituir *ha-cluster* com o nome do seu cluster de alta disponibilidade.
- Substitua o *ec2_key_pair_name* com o nome do seu EC2 key pair para esse cluster. EC2o key pair é opcional e só é necessário se você quiser usá-lo SSH para acessar seu cluster.
- Substituir *subnet-22XXXX01* e *subnet-22XXXX02*com sua sub-redeIDs.

Iniciar um cluster com múltiplos nós primários sem uma estratégia de grupos de posicionamento

Para que um cluster com múltiplos nós primários inicie nós primários sem a estratégia de grupos de posicionamento, é necessário:

- Remova a política gerenciada por grupos `AmazonElasticMapReducePlacementGroupPolicy` de posicionamento da Amazon EMRrole ou
- Inicie um cluster com vários nós primários com o `placement-group-configs` parâmetro usando a Amazon EMRAPI ou CLI escolhendo `NONE` como estratégia de grupo de posicionamento.

Amazon EMR API

Example — Lançamento de um cluster com vários nós primários sem estratégia de grupos de posicionamento usando a AmazonEMRAPI.

Ao usar a RunJobFlow ação para criar um cluster com vários nós primários, defina a `PlacementGroupConfigs` propriedade da seguinte forma.

```
{
   "Name":"ha-cluster",
   "PlacementGroupConfigs":[
      {
```

```
        "InstanceRole":"MASTER",
        "PlacementStrategy":"NONE"
      }
   ],
   "ReleaseLabel":"emr-5.30.1",
   "Instances":{
      "ec2SubnetId":"subnet-22XXXX01",
      "ec2KeyName":"ec2_key_pair_name",
      "InstanceGroups":[
         {
            "InstanceCount":3,
            "InstanceRole":"MASTER",
            "InstanceType":"m5.xlarge"
         },
         {
            "InstanceCount":4,
            "InstanceRole":"CORE",
            "InstanceType":"m5.xlarge"
         }
      ]
   },
   "JobFlowRole":"EMR_EC2_DefaultRole",
   "ServiceRole":"EMR_DefaultRole"
}
```

- Substituir *ha-cluster* com o nome do seu cluster de alta disponibilidade.
- Substituir *subnet-22XXXX01* com seu ID de sub-rede.
- Substitua o *ec2_key_pair_name* com o nome do seu EC2 key pair para esse cluster. EC2o key pair é opcional e só é necessário se você quiser usá-lo SSH para acessar seu cluster.

Amazon EMR CLI

Example — Lançamento de um cluster com vários nós primários sem uma estratégia de grupos de posicionamento usando a AmazonEMRCLI.

Ao usar a RunJobFlow ação para criar um cluster com vários nós primários, defina a `PlacementGroupConfigs` propriedade da seguinte forma.

```
aws emr create-cluster \
--name "ha-cluster" \
--placement-group-configs InstanceRole=MASTER,PlacementStrategy=NONE \
```

```
--release-label emr-5.30.1 \
--instance-groups InstanceGroupType=MASTER,InstanceCount=3,InstanceType=m5.xlarge
 InstanceGroupType=CORE,InstanceCount=4,InstanceType=m5.xlarge \
--ec2-attributes
 KeyName=ec2_key_pair_name,InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-22XXXX01
 \
--service-role EMR_DefaultRole \
--applications Name=Hadoop Name=Spark
```

- Substituir *ha-cluster* com o nome do seu cluster de alta disponibilidade.
- Substituir *subnet-22XXXX01* com seu ID de sub-rede.
- Substitua o *ec2_key_pair_name* com o nome do seu EC2 key pair para esse cluster. EC2o key pair é opcional e só é necessário se você quiser usá-lo SSH para acessar seu cluster.

Verificar a configuração da estratégia de grupos de posicionamento anexada ao cluster com múltiplos nós primários

Você pode usar o cluster Amazon EMR describe API para ver a configuração da estratégia do grupo de posicionamento anexada ao cluster com vários nós primários.

Example

```
aws emr describe-cluster --cluster-id "j-xxxxx"
{
   "Cluster":{
      "Id":"j-xxxxx",
      ...
      ...
      "PlacementGroups":[
         {
            "InstanceRole":"MASTER",
            "PlacementStrategy":"SPREAD"
         }
      ]
   }
}
```

Considerações e práticas recomendadas

Considere o seguinte ao criar um EMR cluster da Amazon com vários nós primários:

> **Important**
>
> Para lançar EMR clusters de alta disponibilidade com vários nós primários, é altamente recomendável que você use a EMR versão mais recente da Amazon. Isso garante que você obtenha o mais alto nível de resiliência e estabilidade para os seus clusters de alta disponibilidade.

- A alta disponibilidade, por exemplo, de frotas é suportada pelas EMR versões 5.36.1, 5.36.2, 6.8.1, 6.9.1, 6.10.1, 6.11.1, 6.12.0 e superiores da Amazon. Para grupos de instâncias, a alta disponibilidade é suportada nas EMR versões 5.23.0 e superiores da Amazon. Para saber mais, consulte Sobre os EMR lançamentos da Amazon.
- Em clusters de alta disponibilidade, a Amazon EMR só oferece suporte ao lançamento de nós primários com instâncias On Demand. Isso garante a maior disponibilidade para o seu cluster.
- Você ainda pode especificar vários tipos de instância para a frota primária, mas todos os nós primários de clusters de alta disponibilidade são executados com o mesmo tipo de instância, incluindo substituições de nós primários não íntegros.
- Para continuar as operações, um cluster de alta disponibilidade com vários nós primários exige que dois dos três nós primários estejam íntegros. Como resultado, se dois nós primários falharem simultaneamente, seu EMR cluster falhará.
- Todos os EMR clusters, incluindo clusters de alta disponibilidade, são lançados em uma única zona de disponibilidade. Portanto, eles não são tolerantes a falhas na zona de disponibilidade. Se houver uma interrupção na zona de disponibilidade, você perde o acesso ao cluster.
- Se você usar Se estiver usando uma função ou política de serviço personalizada ao iniciar um cluster dentro de uma frota de instâncias, poderá adicionar a `ec2:DescribeInstanceTypeOfferings` permissão para que a Amazon EMR possa filtrar zonas de disponibilidade (AZ) não suportadas. Quando a Amazon EMR filtra as AZs que não oferecem suporte a nenhum tipo de instância de nós primários, a Amazon EMR evita que as inicializações de cluster falhem devido a tipos de instância primária não suportados. Para obter mais informações, consulte Tipo de instância não suportado.
- A Amazon EMR não garante alta disponibilidade para aplicativos de código aberto além dos especificados emAplicativos compatíveis em um EMR cluster da Amazon com vários nós primários.
- Nas EMR versões 5.23.0 a 5.36.2 da Amazon, apenas dois dos três nós principais de um cluster de grupo de instâncias são executados. HDFS NameNode

- Nas EMR versões 6.x e posteriores da Amazon, todos os três nós principais de um grupo de instâncias são executadosHDFS NameNode.

Considerações para configurar a sub-rede:

- Um EMR cluster da Amazon com vários nós primários pode residir somente em uma zona de disponibilidade ou sub-rede. A Amazon EMR não pode substituir um nó primário com falha se a sub-rede for totalmente utilizada ou estiver com excesso de assinaturas no caso de um failover. Para evitar esse cenário, é recomendável que você dedique uma sub-rede inteira a um cluster da AmazonEMR. Além disso, certifique-se de que há endereços IP privados suficientes disponíveis na sub-rede.

Considerações para configurar nós core:

- Para também garantir a alta disponibilidade dos nós centrais, é recomendável executar pelo menos quatro nós centrais. Se você decidir iniciar um cluster menor com três ou menos nós principais, `dfs.replication parameter` defina como pelo menos 2 HDFS para ter DFS replicação suficiente. Para obter mais informações, consulte HDFSconfiguração.

> **Warning**
>
> 1. `dfs.replication`Definir como 1 em clusters com menos de quatro nós pode levar à perda de HDFS dados se um único nó ficar inativo. É recomendável usar um cluster com pelo menos quatro nós centrais para workloads de produção.
> 2. A Amazon não EMR permitirá que os clusters escalem os nós principais abaixo`dfs.replication`. Por exemplo, se `dfs.replication = 2`, o número mínimo de nós central será 2.
> 3. Ao usar o Ajuste de Escala Gerenciado, o Auto Scaling ou optar por redimensionar manualmente o cluster, é recomendável definir `dfs.replication` como 2 ou mais.

Considerações para configurar alarmes em métricas:

- A Amazon EMR não fornece métricas específicas de aplicativos sobre ou. HDFS YARN É recomendável definir alarmes para monitorar a contagem de instâncias para nós primários. Configure os alarmes usando as seguintes CloudWatch métricas da

Amazon:`MultiMasterInstanceGroupNodesRunning`,`MultiMasterInstanceGroupNodesRunning` ou`MultiMasterInstanceGroupNodesRequested`. CloudWatch notificará você em caso de falha e substituição do nó primário.

- Se `MultiMasterInstanceGroupNodesRunningPercentage` for menor que 1,0 e maior que 0,5, o cluster pode ter perdido um nó primário. Nessa situação, a Amazon EMR tenta substituir um nó primário.
- Se a `MultiMasterInstanceGroupNodesRunningPercentage` ficar abaixo de 0,5, dois nós primários podem ter falhado. Nesse caso, o quórum do cluster é perdido e não é possível recuperá-lo. É necessário migrar os dados desse cluster manualmente.

Para obter mais informações, consulte Setting alarms on metrics.

EMRclusters em AWS Outposts

A partir do Amazon EMR 5.28.0, você pode criar e executar EMR clusters no. AWS Outposts AWS Outposts habilita AWS serviços, infraestrutura e modelos operacionais nativos em instalações locais. Em AWS Outposts ambientes, você pode usar as mesmas AWS APIs ferramentas e infraestrutura que usa na AWS nuvem. O Amazon EMR on AWS Outposts é ideal para cargas de trabalho de baixa latência que precisam ser executadas nas proximidades de dados e aplicativos locais. Para obter mais informações sobre AWS Outposts, consulte o Guia AWS Outposts do usuário.

Pré-requisitos

A seguir estão os pré-requisitos para usar a Amazon em: EMR AWS Outposts

- Você deve ter instalado e configurado AWS Outposts em seu data center local.
- Você deve ter uma conexão de rede confiável entre seu ambiente Outpost e uma AWS região.
- Você deve ter capacidade suficiente para os tipos de instância EMR suportados pela Amazon disponíveis em seu Outpost.

Limitações

A seguir estão as limitações do uso da Amazon EMR em AWS Outposts:

- As instâncias sob demanda são a única opção compatível com as EC2 instâncias da Amazon. As instâncias spot não estão disponíveis para a Amazon EMR em AWS Outposts.

- Se você precisar de volumes adicionais EBS de armazenamento da Amazon, somente o General Purpose SSD (GP2) é suportado.
- Quando você usa AWS Outposts com as EMR versões 5.28 a 6.x da Amazon, você só pode usar buckets do S3 que armazenam objetos em um Região da AWS que você especificar. Com o Amazon EMR 7.0.0 e versões posteriores, o Amazon EMR on também AWS Outposts é compatível com o S3A prefixo do cliente do sistema de arquivos. `s3a://`
- Somente os seguintes tipos de instância são compatíveis com a Amazon EMR em AWS Outposts:

Classe de instância	Tipos de instância
Propósito geral	`m5.xlarge |m5.2xlarge | m5.4xlarge |m5.12xlarge | m5.24xlarge |m5d.xlarge | m5d.2xlarge | m5d.4xlarge | m5d.12xlarge | m5d.24xlarge`
Otimizada para computação	`c5.xlarge |c5.2xlarge | c5.4xlarge |c5.18xlarge | c5d.xlarge |c5d.2xlarge | c5d.4xlarge | c5d.18xlarge`
Otimizada para memória	`r5.xlarge |r5.2xlarge | r5.4xlarge |r5.12xlarge | r5d.xlarge |r5d.2xlarge | r5d.4xlarge | r5d.12xlarge | r5d.24xlarge`
Otimizada para armazenamento	`i3en.xlarge |i3en.2xlarge | i3en.3xlarge | i3en.6xlarge | i3en.12xlarge |i3en.24xlarge`

Considerações sobre a conectividade de rede

- Se a conectividade de rede entre seu Posto Avançado e sua AWS região for perdida, seus clusters continuarão funcionando. No entanto, você não pode criar novos clusters ou executar novas ações em clusters existentes enquanto que a conectividade não for restaurada. Em caso de falhas na instância, a instância não será substituída automaticamente. Além disso, ações como adicionar etapas a um cluster em execução, verificar o status de execução das etapas e enviar CloudWatch métricas e eventos serão adiadas.

- Recomendamos que você forneça conectividade de rede confiável e altamente disponível entre seu Posto Avançado e a AWS Região. Se a conectividade de rede entre seu Posto Avançado e sua AWS região for perdida por mais de algumas horas, os clusters que ativaram a proteção de encerramento continuarão funcionando e os clusters que desativaram a proteção de encerramento poderão ser encerrados.
- Se a conectividade da rede for afetada por uma manutenção de rotina, recomendamos que a proteção contra encerramento seja ativada proativamente. De modo geral, a interrupção da conectividade significa que quaisquer dependências externas que não sejam locais para a rede do cliente ou o para Outpost ficarão inacessíveis. Isso inclui Amazon S3, DynamoDB usado com EMRFS visualização de consistência e Amazon se uma instância na região RDS for usada para um cluster da Amazon com vários nós primários. EMR

Criação de um EMR cluster da Amazon em AWS Outposts

Criar um EMR cluster da Amazon no AWS Outposts é semelhante à criação de um EMR cluster da Amazon na AWS nuvem. Ao criar um EMR cluster da Amazon no AWS Outposts, você deve especificar uma EC2 sub-rede da Amazon associada ao seu Outpost.

Uma Amazon VPC pode abranger todas as zonas de disponibilidade em uma AWS região. AWS Outposts são extensões das zonas de disponibilidade, e você pode estender uma conta Amazon VPC em uma conta para abranger várias zonas de disponibilidade e locais associados do Outpost. Ao configurar seu Outpost, você associa uma sub-rede a ele para estender seu VPC ambiente regional às suas instalações locais. As instâncias do Outpost e os serviços relacionados aparecem como parte de sua regiãoVPC, semelhante a uma zona de disponibilidade com sub-redes associadas. Para obter mais informações, consulte o Guia do usuário do AWS Outposts.

Console

Para criar um novo EMR cluster da Amazon AWS Outposts com o AWS Management Console, especifique uma EC2 sub-rede da Amazon associada ao seu Outpost.

Console

Para criar um cluster AWS Outposts com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.

2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Em Configuração do cluster, selecione Grupos de instâncias ou Frotas de instâncias. Em seguida, escolha um tipo de instância no menu suspenso Escolher tipo de EC2 instância ou selecione Ações e escolha Adicionar EBS volumes. O Amazon EMR on AWS Outposts oferece suporte a tipos limitados EBS de volume e instância da Amazon.
4. Em Rede, selecione uma EC2 sub-rede com um Outpost ID neste formato: op-123456789.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

CLI

Para criar um cluster AWS Outposts com o AWS CLI

- Para criar um novo EMR cluster da Amazon AWS Outposts com o AWS CLI, especifique uma EC2 sub-rede associada ao seu Outpost, como no exemplo a seguir. Substituir *subnet-22XXXX01* com seu próprio ID de EC2 sub-rede da Amazon.

```
aws emr create-cluster \
--name "Outpost cluster" \
--release-label emr-7.2.0 \
--applications Name=Spark \
--ec2-attributes KeyName=myKey SubnetId=subnet-22XXXX01 \
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

EMRclusters em AWS Locais Zones

A partir da EMR versão 5.28.0 da Amazon, você pode criar e executar EMR clusters da Amazon em uma sub-rede de Zonas AWS Locais como uma extensão lógica de uma AWS região que suporta Zonas Locais. Uma zona local permite que os EMR recursos da Amazon e um subconjunto de AWS serviços, como serviços de computação e armazenamento, estejam localizados mais perto dos usuários para fornecer acesso de latência muito baixa aos aplicativos executados localmente. Para obter uma lista das zonas locais disponíveis, consulte Zonas locais da AWS. Para obter informações sobre como acessar as Zonas AWS Locais disponíveis, consulte Regiões, Zonas de Disponibilidade e zonas locais.

Tipos de instâncias compatíveis

Os seguintes tipos de instância estão disponíveis para EMR clusters da Amazon em Locais Zones. A disponibilidade do tipo de instância pode variar de acordo com a região.

Classe de instância	Tipos de instância
Propósito geral	m5.xlarge | m5.2xlarge | m5.4xlarge | m5.12xlarge | m5.24xlarge | m5d.xlarge | m5d.2xlarge | m5d.4xlarge | m5d.12xlarge | m5d.24xlarge
Otimizada para computação	c5.xlarge | c5.2xlarge | c5.4xlarge | c5.9xlarge | c5.18xlarge | c5d.xlarge | c5d.2xlarge | c5d.4xlarge| c5d.9xlarge | c5d.18xlarge
Otimizada para memória	r5.xlarge | r5.2xlarge | r5.4xlarge | r5.12xlarge | r5d.xlarge | r5d.2xlarge | r5d.4xlarge | r5d.12xlarge | r5d.24xlarge
Otimizada para armazenamento	i3en.xlarge | i3en.2xlarge | i3en.3xlarge | i3en.6xlarge | i3en.12xlarge | i3en.24xlarge

Criação de um EMR cluster da Amazon em Locais Zones

Crie um EMR cluster da Amazon em Zonas AWS Locais iniciando o EMR cluster da Amazon em uma VPC sub-rede da Amazon associada a uma Zona Local. É possível acessar o cluster usando o nome da zona local, como us-west-2-lax-1a, no console da região Oeste dos EUA (Oregon).

Atualmente, as Zonas Locais não oferecem suporte a Amazon EMR Notebooks ou conexões diretamente com a Amazon EMR usando a interface VPC endpoint ()AWS PrivateLink.

Console

Para criar um cluster em uma zona local com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.

3. Em Rede, selecione uma EC2 sub-rede com uma ID de zona local neste formato: subnet 123abc | us-west-2-lax-1a.
4. Escolha um tipo de instância ou adicione volumes de EBS armazenamento da Amazon para grupos de instâncias ou frotas de instâncias uniformes.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

CLI

Para criar um cluster em uma zona local com o AWS CLI

- Use o comando create-cluster, junto com o SubnetId para a zona local, conforme mostrado no exemplo a seguir. Substitua a subnet-22 XXXX1234567 pela zona local SubnetId e substitua outras opções conforme necessário. Para obter mais informações, consulte https://docs.aws.amazon.com/cli/latest/reference/emr/create-cluster.html.

```
aws emr create-cluster \
--name "Local Zones cluster" \
--release-label emr-5.29.0 \
--applications Name=Spark \
--ec2-attributes KeyName=myKey,SubnetId=subnet-22XXXX1234567 \
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

Configurar o Docker

O Amazon EMR 6.x é compatível com o Hadoop 3, o que permite YARN NodeManager lançar contêineres diretamente no EMR cluster da Amazon ou dentro de um contêiner Docker. Os contêineres do Docker fornecem ambientes de execução personalizados nos quais o código do aplicativo é executado. O ambiente de execução personalizado é isolado do ambiente de execução do YARN NodeManager e de outros aplicativos.

Os contêineres do Docker podem incluir bibliotecas especiais usadas pelo aplicativo e podem fornecer diferentes versões de ferramentas e bibliotecas nativas, como R e Python. É possível usar ferramentas familiares do Docker para definir bibliotecas e dependências de runtime para as aplicações.

Os clusters do Amazon EMR 6.x são configurados por padrão para permitir que YARN aplicativos, como o Spark, sejam executados usando contêineres Docker. Para personalizar a configuração do contêiner, edite as opções de suporte do Docker definidas nos arquivos `yarn-site.xml` e `container-executor.cfg` disponíveis no diretório `/etc/hadoop/conf`. Para obter detalhes sobre cada opção de configuração e como ela é usada, consulte Launching applications using Docker containers.

É possível optar por usar o Docker ao enviar um trabalho. Use as variáveis a seguir para especificar o runtime do Docker e a imagem do Docker.

- `YARN_CONTAINER_RUNTIME_TYPE=docker`
- `YARN_CONTAINER_RUNTIME_DOCKER_IMAGE={`*`DOCKER_IMAGE_NAME`*`}`

Quando você usa contêineres do Docker para executar seus YARN aplicativos, YARN baixa a imagem do Docker que você especifica ao enviar seu trabalho. YARNPara resolver essa imagem do Docker, ela deve ser configurada com um registro do Docker. As opções de configuração de um registro do Docker dependem se você implanta o cluster usando uma sub-rede pública ou privada.

Registros do Docker

Um registro do Docker é um sistema de armazenamento e distribuição de imagens do Docker. Para a Amazon, EMR recomendamos que você use a AmazonECR, que é um registro de contêineres do Docker totalmente gerenciado que permite criar suas próprias imagens personalizadas e hospedá-las em uma arquitetura altamente disponível e escalável.

Considerações de implantação

Os registros do Docker exigem acesso à rede de cada host no cluster. Isso ocorre porque cada host baixa imagens do registro do Docker quando seu YARN aplicativo está sendo executado no cluster. Esses requisitos de conectividade de rede podem limitar sua escolha de registro do Docker, dependendo se você implanta seu EMR cluster da Amazon em uma sub-rede pública ou privada.

Public subnet (Sub-rede pública)

Quando os EMR clusters são implantados em uma sub-rede pública, os nós em execução YARN NodeManager podem acessar diretamente qualquer registro disponível na Internet.

Sub-rede privada

Quando os EMR clusters são implantados em uma sub-rede privada, os nós em execução YARN NodeManager não têm acesso direto à Internet. As imagens do Docker podem ser hospedadas na Amazon ECR e acessadas por meio de AWS PrivateLink.

Para obter mais informações sobre como usar AWS PrivateLink para permitir o acesso à Amazon ECR em um cenário de sub-rede privada, consulte Configurando AWS PrivateLink a Amazon ECS e a Amazon ECR.

Configurar registros do Docker

Para usar os registros do Docker na AmazonEMR, você deve configurar o Docker para confiar no registro específico que você deseja usar para resolver imagens do Docker. Os registros de confiança padrão são locais (privados) e centOS. Para usar outros repositórios públicos ou a AmazonECR, você pode substituir `docker.trusted.registries` as configurações `/etc/hadoop/conf/container-executor.cfg` usando a EMR Classificação API com a chave de `container-executor` classificação.

O exemplo a seguir mostra como configurar o cluster para confiar tanto em um repositório público, nomeado`your-public-repo`, quanto em um endpoint de ECR registro. `123456789123.dkr.ecr.us-east-1.amazonaws.com` Se você usarECR, substitua esse endpoint pelo seu ECR endpoint específico.

```
[
  {
    "Classification": "container-executor",
    "Configurations": [
        {
            "Classification": "docker",
            "Properties": {
                "docker.trusted.registries": "local,centos,your-public-
repo,123456789123.dkr.ecr.us-east-1.amazonaws.com",
                "docker.privileged-containers.registries": "local,centos,your-public-
repo,123456789123.dkr.ecr.us-east-1.amazonaws.com"
            }
        }
    ]
  }
]
```

Para iniciar um cluster Amazon EMR 6.0.0 com essa configuração usando o AWS Command Line Interface (AWS CLI), crie um arquivo chamado `container-executor.json` com o conteúdo da

configuração anterior do executor-contêiner. JSON Depois, use os comandos a seguir para executar o cluster.

```
export KEYPAIR=<Name of your Amazon EC2 key-pair>
export SUBNET_ID=<ID of the subnet to which to deploy the cluster>
export INSTANCE_TYPE=<Name of the instance type to use>
export REGION=<Region to which to deploy the cluster>

aws emr create-cluster \
    --name "EMR-6.0.0" \
    --region $REGION \
    --release-label emr-6.0.0 \
    --applications Name=Hadoop Name=Spark \
    --service-role EMR_DefaultRole \
    --ec2-attributes KeyName=$KEYPAIR,InstanceProfile=EMR_EC2_DefaultRole,SubnetId=
$SUBNET_ID \
    --instance-groups InstanceGroupType=MASTER,InstanceCount=1,InstanceType=
$INSTANCE_TYPE InstanceGroupType=CORE,InstanceCount=2,InstanceType=$INSTANCE_TYPE \
    --configuration file://container-executor.json
```

Configurando YARN para acessar a Amazon ECR na EMR versão 6.0.0 e versões anteriores

Se você é novo na AmazonECR, siga as instruções em Introdução à Amazon ECR e verifique se você tem acesso à Amazon a ECR partir de cada instância em seu EMR cluster da Amazon.

Na EMR versão 6.0.0 e versões anteriores, para acessar a Amazon ECR usando o comando Docker, você deve primeiro gerar credenciais. Para verificar se YARN pode acessar imagens da AmazonECR, use a variável de ambiente do contêiner `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG` para passar uma referência às credenciais que você gerou.

Execute o comando a seguir em um dos nós principais para obter a linha de login da sua ECR conta.

```
aws ecr get-login --region us-east-1 --no-include-email
```

O `get-login` comando gera o CLI comando correto do Docker a ser executado para criar credenciais. Copie e execute a saída de `get-login`.

```
sudo docker login -u AWS -p <password> https://<account-id>.dkr.ecr.us-
east-1.amazonaws.com
```

Esse comando gera um arquivo `config.json` na pasta `/root/.docker`. Copie esse arquivo HDFS para que os trabalhos enviados ao cluster possam usá-lo para se autenticar na AmazonECR.

Execute os comandos a seguir para copiar o arquivo `config.json` no diretório inicial.

```
mkdir -p ~/.docker
sudo cp /root/.docker/config.json ~/.docker/config.json
sudo chmod 644 ~/.docker/config.json
```

Execute os comandos abaixo para colocar o config.json HDFS para que ele possa ser usado por trabalhos em execução no cluster.

```
hadoop fs -put ~/.docker/config.json /user/hadoop/
```

YARNpode acessar ECR como um registro de imagem do Docker e extrair contêineres durante a execução do trabalho.

Depois de configurar os registros do DockerYARN, você pode executar YARN aplicativos usando contêineres do Docker. Para obter mais informações, consulte Executar aplicativos Spark com o Docker usando o Amazon EMR 6.0.0.

Na EMR versão 6.1.0 e versões posteriores, você não precisa configurar manualmente a autenticação na AmazonECR. Se um ECR registro da Amazon for detectado na chave de `container-executor` classificação, o recurso de autenticação ECR automática da Amazon é ativado e YARN gerencia o processo de autenticação quando você envia um trabalho do Spark com uma ECR imagem. Você pode confirmar se a autenticação automática está habilitada verificando `yarn.nodemanager.runtime.linux.docker.ecr-auto-authentication.enabled` em yarn-site. A autenticação automática está ativada e a configuração de YARN autenticação é definida como `true` se `docker.trusted.registries` contiver um ECR registroURL.

Pré-requisitos para usar a autenticação automática na Amazon ECR

- EMRversão 6.1.0 ou posterior
- ECRo registro incluído na configuração está na mesma região do cluster
- IAMfunção com permissões para obter o token de autorização e extrair qualquer imagem

Consulte Configurando com a Amazon ECR para obter mais informações.

Como habilitar a autenticação automática

Siga Configurar registros do Docker para definir um ECR registro da Amazon como um registro confiável e garantir que o ECR repositório da Amazon e o cluster estejam na mesma região.

Para habilitar esse recurso mesmo quando o ECR registro não estiver definido no registro confiável, use a classificação de configuração `yarn.nodemanager.runtime.linux.docker.ecr-auto-authentication.enabled` para definir como`true`.

Como desabilitar a autenticação automática

Por padrão, a autenticação automática é desativada se nenhum ECR registro da Amazon for detectado no registro confiável.

Para desativar a autenticação automática, mesmo quando o ECR registro da Amazon estiver definido no registro confiável, use a classificação de configuração `yarn.nodemanager.runtime.linux.docker.ecr-auto-authentication.enabled` para definir`false`.

Como verificar se a autenticação automática está habilitada em um cluster

No nó principal, use um editor de texto, como `vi`, para visualizar o conteúdo do arquivo de log: `vi /etc/hadoop/conf.empty/yarn-site.xml`. Verifique o valor de `yarn.nodemanager.runtime.linux.docker.ecr-auto-authentication.enabled`.

Controle de término do cluster

Esta seção descreve suas opções para encerrar os EMR clusters da Amazon. Ele abrange a rescisão automática e a proteção contra rescisão e como elas interagem com outros EMR recursos da Amazon.

Você pode desligar um EMR cluster da Amazon das seguintes formas:

- Término após a execução da última etapa: crie um cluster transitório que será terminado após a conclusão de todas as etapas.
- Término automático (após tempo ocioso): crie um cluster com uma política de término automático que é desativado após um tempo ocioso especificado. Para obter mais informações, consulte Usar uma política de término automático.

- Término manual: crie um cluster de execução prolongada que continue em execução até você terminá-lo deliberadamente. Para obter mais informações sobre como encerrar um cluster manualmente, consulte Terminar um cluster.

Você também pode definir a proteção contra encerramento em um cluster para evitar o desligamento de EC2 instâncias por acidente ou erro.

Quando a Amazon EMR encerra seu cluster, todas as EC2 instâncias da Amazon no cluster são encerradas. Os dados no armazenamento e nos EBS volumes da instância não estão mais disponíveis e não podem ser recuperados. Entender e gerenciar o término do cluster é essencial para desenvolver uma estratégia para gerenciar e preservar dados gravando no Amazon S3 e equilibrando o custo.

Tópicos

Configurar um cluster para continuar ou terminar após a execução da etapa

Este tópico explica as diferenças entre usar um cluster de execução prolongada e criar um cluster transitório que é desativado após a execução da última etapa. Também aborda como configurar a execução de etapas em um cluster.

Criar um cluster de execução prolongada

Por padrão, os clusters que você cria com o console ou com o AWS CLI são de longa duração. Os clusters de execução prolongada continuam funcionando, aceitando trabalho e acumulando cobranças até você tomar medidas para desativá-los.

Um cluster de execução prolongada tem efeito nas seguintes situações:

- Quando você precisa consultar dados de forma interativa ou automática.
- Quando você precisa interagir continuamente com aplicações de big data hospedadas no cluster.
- Quando você processa periodicamente um conjunto de dados tão grande ou com tanta frequência que é ineficiente iniciar novos clusters e carregar dados todas as vezes.

Você também pode definir a proteção contra encerramento em um cluster de longa duração para evitar o desligamento de EC2 instâncias por acidente ou erro. Para obter mais informações, consulte Usar a proteção contra término.

> **Note**
>
> A Amazon habilita EMR automaticamente a proteção contra terminação para todos os clusters com vários nós primários e substitui todas as configurações de execução de etapas que você fornece ao criar o cluster. É possível desabilitar a proteção contra término depois que o cluster é iniciado. Consulte Configurar a proteção contra término para clusters em execução. Para desligar um cluster com múltiplos nós primários, primeiro é necessário modificar os atributos do cluster para desabilitar a proteção contra término. Para obter instruções, consulte Encerrar um EMR cluster da Amazon com vários nós primários.

Configurar um cluster para terminar após a execução da etapa

Quando você configura o término após a execução da etapa, o cluster é iniciado, executa ações de bootstrap e executa as etapas especificadas. Assim que a última etapa for concluída, a Amazon EMR encerrará as instâncias Amazon EC2 do cluster. Os clusters que você executa com a Amazon EMR API têm a execução em etapas ativada por padrão.

O término após a execução da etapa é eficaz para clusters que realizam uma tarefa de processamento periódico, como uma execução diária do processamento de dados. A execução de etapas também ajuda a garantir que você pague somente pelo tempo necessário para processar seus dados. Para mais informações sobre as etapas, consulte Enviar trabalhos a um cluster.

Console

Para ativar o encerramento após a execução da etapa com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Em Etapas, escolha Adicionar etapa. Na caixa de diálogo Adicionar etapa, insira os valores apropriados dos campos. As opções diferem dependendo do tipo de etapa. Para adicionar a etapa e sair da caixa de diálogo, escolha Adicionar etapa.

4. Em Término do cluster, marque a caixa de seleção Terminar cluster após a conclusão da última etapa.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

AWS CLI

Para ativar a rescisão após a execução da etapa com o AWS CLI

- Especifique o parâmetro `--auto-terminate` quando usar o comando `create-cluster` para criar um cluster transitório.

  O exemplo a seguir demonstra com usar o parâmetro `--auto-terminate`. Você pode digitar o seguinte comando e substituir *myKey* com o nome do seu EC2 key pair.

  > **Note**
  >
  > Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 \
--applications Name=Hive Name=Pig --use-default-roles --ec2-attributes
 KeyName=myKey \
--steps Type=PIG,Name="Pig Program",ActionOnFailure=CONTINUE,\
Args=[-f,s3://mybucket/scripts/pigscript.pig,-p,\
INPUT=s3://mybucket/inputdata/,-p,OUTPUT=s3://mybucket/outputdata/,\
$INPUT=s3://mybucket/inputdata/,$OUTPUT=s3://mybucket/outputdata/]
--instance-type m5.xlarge --instance-count 3 --auto-terminate
```

API

Para desativar o encerramento após a execução da etapa com o Amazon EMR API no cluster, execute

1. Ao usar a RunJobFlowação para criar um cluster, defina a KeepJobFlowAliveWhenNoStepspropriedade como`false`.

2. Para alterar sua configuração de encerramento após a execução da etapa com o lançamento do Amazon EMR API Post Cluster:

   Use SetKeepJobFlowAliveWhenNoSteps a ação.

Usar uma política de término automático

Uma política de término automático permite orquestrar a limpeza do cluster sem a necessidade de monitorar e terminar manualmente os clusters não utilizados. Ao adicionar uma política de término automático a um cluster, especifique a quantidade de tempo ocioso após o qual o cluster deverá ser desligado automaticamente.

Dependendo da versão de lançamento, a Amazon EMR usa critérios diferentes para marcar um cluster como ocioso. A tabela a seguir descreve como a Amazon EMR determina a ociosidade do cluster.

Quando você usa...	O cluster é considerado ocioso quando...
EMRVersões da Amazon 5.34.0 e posteriores e 6.4.0 e posteriores	• Não há YARN aplicativos ativos • HDFSa utilização está abaixo de 10% • Não há conexões ativas com o EMR notebook ou com o EMR Studio • Não há interfaces de usuário de aplicações no cluster em uso • Não há etapas pendentes
EMRVersões da Amazon 5.30.0 - 5.33.0 e 6.1.0 - 6.3.0	• Não há YARN aplicativos ativos • O cluster não tem trabalhos do Spark ativos

Quando você usa...	O cluster é considerado ocioso quando...
	Note A Amazon EMR marca um cluster como ocioso e pode encerrá-lo automaticamente mesmo se você tiver um kernel Python3 ativo. Isso ocorre porque a execução de um kernel do Python3 não envia um trabalho do Spark no cluster. Para usar a terminaçã o automática com um kernel Python3, recomendamos que você use a versão 6.4.0 ou posterior da AmazonEMR.

Note

EMRAs versões 6.4.0 e posteriores da Amazon oferecem suporte a um arquivo no cluster para detectar atividades no nó primário:. `/emr/metricscollector/isbusy` Ao usar um cluster para executar scripts de shell ou não YARN aplicativos, você pode tocar ou atualizar periodicamente `isbusy` para informar à Amazon EMR que o cluster não está ocioso.

É possível anexar uma política de término automático ao criar um cluster ou adicionar uma política a um cluster atual. Para alterar ou desabilitar o término automático, é possível atualizar ou remover a política.

Considerações

Leve em consideração os atributos e as limitações a seguir antes de usar uma política de término automático:

- A seguir Regiões da AWS, a EMR terminação automática da Amazon está disponível com o Amazon EMR 6.14.0 e superior:
  - Ásia-Pacífico (Hyderabad) (ap-south-2)
  - Ásia-Pacífico (Jacarta) (ap-southeast-3)

  - Europa (Espanha) (eu-south-2)
- A seguir Regiões da AWS, a EMR terminação automática da Amazon está disponível com o Amazon EMR 5.30.0 e 6.1.0 e versões superiores:
  - Leste dos EUA (Norte da Virgínia) (us-east-1)
  - Leste dos EUA (Ohio) (us-east-2)
  - Oeste dos EUA (Oregon) (us-west-2)
  - Oeste dos EUA (Norte da Califórnia) (us-west-1)
  - África (Cidade do Cabo) (af-south-1)
  - Ásia-Pacífico (Hong Kong) (ap-east-1)
  - Ásia-Pacífico (Mumbai) (ap-south-1)
  - Ásia-Pacífico (Seul) (ap-northeast-2)
  - Ásia-Pacífico (Singapura) (ap-southeast-1)
  - Ásia-Pacífico (Sydney) (ap-southeast-2)
  - Ásia Pacific (Tóquio) (ap-northeast-1)
  - Canadá (Central) (ca-central-1)
  - América do Sul (São Paulo) (sa-east-1)
  - Europa (Frankfurt) (eu-central-1)
  - Europa (Irlanda) (eu-west-1)
  - Europa (Londres) (eu-west-2)
  - UE (Milão) (eu-south-1)
  - Europa (Paris) (eu-west-3)
  - UE (Estocolmo) (eu-north-1)
  - China (Pequim) (cn-north-1)
  - China (Ningxia) (cn-northwest-1)
  - AWS GovCloud (Leste dos EUA) (us-gov-east-1)
  - AWS GovCloud (Oeste dos EUA) (us-gov-west-1)
- O tempo limite ocioso é padronizado para 60 minutos (uma hora) quando não há um valor especificado. Você pode especificar um tempo limite ocioso mínimo de um minuto e um tempo limite ocioso máximo de sete dias.
- Com EMR as versões 6.4.0 e posteriores da Amazon, o encerramento automático é ativado por padrão quando você cria um novo cluster com o console da AmazonEMR.

- A Amazon EMR publica Amazon CloudWatch métricas de alta resolução quando você ativa o encerramento automático de um cluster. Use essas métricas para monitorar a atividade e a ociosidade do cluster. Para obter mais informações, consulte Métricas de capacidade de cluster.
- A terminação automática não é suportada quando você usa aplicativos não YARN baseados, como Presto, Trino ou. HBase
- Para usar o encerramento automático, o processo coletor de métricas deve ser capaz de se conectar ao API endpoint público para o encerramento automático no Gateway. API Se você usar um DNS nome privado com Amazon Virtual Private Cloud, o encerramento automático não funcionará corretamente. Para garantir que o término automático funcione, é recomendável executar uma das seguintes ações:
  - Remova o VPC endpoint da interface API Gateway da sua AmazonVPC.
  - Siga as instruções em Por que recebo um erro HTTP 403 Forbidden ao me conectar ao meu API Gateway a APIs partir de um VPC? para desativar a configuração DNS do nome privado.
  - Em vez disso, inicie o cluster em sua sub-rede privada. Para obter mais informações, consulte o tópico em Sub-redes privadas.
- (EMR5.30.0 e versões posteriores) Se você remover a regra de saída padrão Permitir Todos para 0.0.0.0/ para o grupo de segurança primário, deverá adicionar uma regra que permita TCP conectividade de saída ao seu grupo de segurança para acesso ao serviço na porta 9443. Seu grupo de segurança para acesso ao serviço também deve permitir TCP tráfego de entrada na porta 9443 do grupo de segurança principal. Para obter mais informações sobre a configuração de grupos de segurança, consulte Grupo EMR de segurança gerenciado pela Amazon para a instância primária (sub-redes privadas).

Permissões para usar o término automático

Antes de aplicar e gerenciar políticas de encerramento automático para a AmazonEMR, você precisa anexar as permissões listadas no exemplo de política de IAM permissões a seguir aos IAM recursos que gerenciam seu EMR cluster.

```
{
  "Version": "2012-10-17",
  "Statement": {
      "Sid": "AllowAutoTerminationPolicyActions",
      "Effect": "Allow",
      "Action": [
        "elasticmapreduce:PutAutoTerminationPolicy",
        "elasticmapreduce:GetAutoTerminationPolicy",
```

```
        "elasticmapreduce:RemoveAutoTerminationPolicy"
      ],
      "Resource": "<your-resources>"
    }
}
```

Anexar, atualizar ou remover uma política de término automático

Esta seção inclui instruções para ajudá-lo a anexar, atualizar ou remover uma política de encerramento automático de um EMR cluster da Amazon. Antes de trabalhar com políticas de encerramento automático, verifique se você tem as IAM permissões necessárias. Consulte Permissões para usar o término automático.

Console

Para anexar uma política de encerramento automático ao criar um cluster com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Em Término do cluster, selecione Terminar cluster após tempo ocioso.
4. Especifique o número de horas e minutos ociosos que podem decorrer antes que o cluster seja terminado automaticamente. O tempo ocioso padrão é de uma hora.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

Para anexar, atualizar ou remover uma política de encerramento automático em um cluster em execução com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EMREC2Em Ativado, no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar.
3. Na guia Propriedades da página de detalhes do cluster, localize Término do cluster e selecione Editar.

4. Selecione ou desmarque Habilitar término automático para ativar ou desativar o atributo. Se você ativar o término automático, especifique o número de horas e minutos ociosos que podem decorrer antes que o cluster seja terminado automaticamente. Depois selecione Salvar alterações para confirmar.

AWS CLI

Antes de começar

Antes de trabalhar com políticas de término automático, é recomendável atualizar para a versão mais recente da AWS CLI. Para obter instruções, consulte Installing, updating, and uninstalling the AWS CLI.

Anexar ou atualizar uma política de término automático usando a AWS CLI

- Use o comando `aws emr put-auto-termination-policy` para anexar ou atualizar uma política de término automático em um cluster.

  O exemplo a seguir especifica 3600 segundos para *`IdleTimeout`*. Se você não especificar *`IdleTimeout`*, o valor padrão é uma hora.

```
aws emr put-auto-termination-policy \
--cluster-id <your-cluster-id> \
--auto-termination-policy IdleTimeout=3600
```

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

  Também é possível especificar um valor para `--auto-termination-policy` ao usar o comando `aws emr create-cluster`. Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consulte a Referência de AWS CLI comandos.

Para remover uma política de encerramento automático com o AWS CLI

- Use o comando `aws emr remove-auto-termination-policy` para remover uma política de término automático de um cluster. Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consulte a Referência de AWS CLI comandos.

```
aws emr remove-auto-termination-policy --cluster-id <your-cluster-id>
```

Usar a proteção contra término

A proteção contra encerramento protege seus clusters contra o encerramento acidental, o que pode ser especialmente útil para clusters de longa execução que processam cargas de trabalho críticas. Quando a proteção contra encerramento está habilitada em um cluster de longa execução, você ainda poderá encerrar o cluster, mas deverá removê-la explicitamente do cluster primeiro. Isso ajuda a garantir que as EC2 instâncias não sejam encerradas por acidente ou erro. Você pode habilitar a proteção contra encerramento ao criar um cluster e alterar a configuração em um cluster em execução.

Com a proteção contra rescisão ativada, a `TerminateJobFlows` ação na Amazon EMR API não funciona. Os usuários não podem encerrar o cluster usando isso API ou o `terminate-clusters` comando do AWS CLI. O API retorna um erro e CLI sai com um código de retorno diferente de zero. Quando você usa o EMR console da Amazon para encerrar um cluster, você recebe uma etapa extra para desativar a proteção contra encerramento.

> **Warning**
>
> A proteção contra encerramento não garante que os dados sejam retidos no caso de um erro humano ou de uma solução alternativa, por exemplo, se um comando de reinicialização for emitido pela linha de comando enquanto estiver conectado à instância usandoSSH, se um aplicativo ou script em execução na instância emitir um comando de reinicialização ou se a Amazon ou EC2 a Amazon forem usadas para desativar a proteção contra encerramento. EMR API Isso também é verdade se você estiver executando as EMR versões 7.1 e superiores da Amazon e uma instância ficar insalubre e irrecuperável. Mesmo com a proteção de terminação ativada, os dados salvos no armazenamento da instância, incluindo HDFS dados, podem ser perdidos. Grave a saída de dados nos locais do Amazon S3 e crie estratégias de backup conforme a necessidade de seus requisitos de continuidade de negócios.

A proteção contra encerramento não afeta sua capacidade de dimensionar recursos de cluster usando qualquer uma das seguintes ações:

- Redimensionando um cluster manualmente com o AWS Management Console ou AWS CLI. Para obter mais informações, consulte Redimensionar manualmente um cluster em execução.
- Removendo instâncias de um grupo de instâncias core ou de tarefa usando uma política de redução com a escalabilidade automática. Para obter mais informações, consulte Usar o ajuste de escala automático com uma política personalizada para grupos de instâncias.
- Removendo instâncias de uma frota de instâncias, reduzindo a capacidade de destino. Para obter mais informações, consulte Opções de frotas de instâncias.

Proteção contra rescisão e Amazon EC2

A configuração de proteção contra encerramento em um EMR cluster da Amazon corresponde ao `DisableApiTermination` atributo de todas as EC2 instâncias da Amazon no cluster. Por exemplo, se você habilitar a proteção contra rescisão em um EMR cluster, a Amazon EMR automaticamente define `DisableApiTermination` como true para todas as EC2 instâncias dentro do EMR cluster. O mesmo se aplica se você desativar a proteção contra rescisão. A Amazon define EMR automaticamente como false `DisableApiTermination` para todas as EC2 instâncias dentro do EMR cluster. Se você encerrar ou reduzir um cluster da Amazon EMR e EC2 as configurações da Amazon entrarem em conflito para uma EC2 instância, a Amazon EMR priorizará a EMR configuração da Amazon sobre as configurações `DisableApiStop` e `DisableApiTermination` na Amazon EC2 e continuará encerrando a instância. EC2

Por exemplo, você pode usar o EC2 console da Amazon para ativar a proteção de encerramento em uma EC2 instância da Amazon em um EMR cluster com a proteção de encerramento desativada. Se você encerrar ou reduzir o cluster com o EMR console da Amazon, o ou o Amazon AWS CLI, a Amazon EMR API EMR substituirá a `DisableApiTermination` configuração, a definirá como falsa e encerrará a instância junto com outras instâncias.

Você também pode usar o EC2 console da Amazon para ativar a proteção de parada em uma EC2 instância da Amazon em um EMR cluster com a proteção de encerramento desativada. Se você encerrar ou reduzir o cluster, a Amazon EMR definirá como `DisableApiStop` false na Amazon EC2 e encerrará a instância junto com outras instâncias.

A Amazon EMR substitui a `DisableApiStop` configuração somente quando você encerra ou reduz a escala de um cluster. Quando você ativa ou desativa a proteção contra rescisão em um

EMR cluster, a Amazon EMR não altera a `disableApiStop` configuração de nenhuma das EC2 instâncias no respectivo EMR cluster.

> **Important**
>
> Se você criar uma instância como parte de um EMR cluster da Amazon com proteção contra encerramento e usar os AWS CLI comandos Amazon EC2 API ou para modificar a instância dessa `DisableApiTermination` formafalse, e então a Amazon EC2 API ou os AWS CLI comandos executarem a `TerminateInstances` operação, a EC2 instância da Amazon será encerrada.

Proteção de terminação e nós insalubres YARN

A Amazon verifica EMR periodicamente o YARN status do Apache Hadoop dos nós em execução nas EC2 instâncias principais e de tarefas da Amazon em um cluster. O estado de saúde é relatado pelo serviço NodeManager de verificação de saúde. Se um nó reportarUNHEALTHY, o controlador de EMR instância da Amazon adiciona o nó a uma lista de negação e não aloca YARN contêineres até que ele volte a ser íntegro. Dependendo do status da proteção contra encerramento, da substituição de nós com problemas de integridade e da versão de EMR lançamento da Amazon, a Amazon EMR substituirá a instância não íntegra ou interromperá a alocação de controladores para a instância.

Proteção de rescisão e rescisão após a execução da etapa

Quando você ativa a rescisão após a execução da etapa e também ativa a proteção contra rescisão, a Amazon EMR ignora a proteção contra rescisão.

Ao enviar etapas para um cluster, você pode definir a propriedade `ActionOnFailure` para determinar o que acontecerá se não for possível executar a etapa devido a um erro. Os valores possíveis para essa configuração são `TERMINATE_CLUSTER` (`TERMINATE_JOB_FLOW` com versões anteriores) `CANCEL_AND_WAIT` e `CONTINUE`. Para obter mais informações, consulte Enviar trabalhos a um cluster.

Se falhar uma etapa configurada com `ActionOnFailure` set to`CANCEL_AND_WAIT`, se a terminação após a execução da etapa for ativada, o cluster será encerrado sem executar as etapas subsequentes.

Se ocorrer uma falha em uma etapa configurada com `ActionOnFailure` definida como `TERMINATE_CLUSTER`, use a tabela de configurações abaixo para determinar o resultado.

ActionOnFailure	Rescisão após a execução da etapa	Termination protection	Resultado
`TERMINATE_CLUSTER`	Habilitado	Desabilitado	O cluster é encerrado
	Habilitado	Habilitado	O cluster é encerrado
	Desabilitado	Habilitado	O cluster continua
	Desabilitado	Desabilitado	O cluster é encerrado

Proteção contra término e instâncias spot

A proteção contra EMR rescisão da Amazon não impede que uma instância EC2 spot da Amazon seja encerrada quando o preço spot subir acima do preço spot máximo.

Configurar a proteção contra término ao iniciar um cluster

Você pode ativar ou desativar a proteção contra encerramento ao iniciar um cluster usando o console AWS CLI, o ou API o.

Para clusters de nó único, as configurações padrão de proteção contra encerramento são as seguintes:

- Lançamento de um cluster pelo Amazon EMR Console — a Proteção de rescisão está desativada por padrão.
- A inicialização de um cluster por meio da AWS CLI `aws emr create-cluster` —Termination Protection está desativada, a menos que `--termination-protected` seja especificada.
- Lançamento de um cluster pelo EMR API RunJobFlowcomando Amazon — a Proteção de terminação é desativada, a menos que o valor `TerminationProtected` booleano esteja definido como. `true`

Para clusters de alta disponibilidade, as configurações padrão de proteção contra encerramento são as seguintes:

- Lançamento de um cluster pelo Amazon EMR Console — A Proteção de Terminação é ativada por padrão.
- A inicialização de um cluster por meio da AWS CLI `aws emr create-cluster` —Termination Protection está desativada, a menos que `--termination-protected` seja especificada.
- Lançamento de um cluster pelo EMR API RunJobFlowcomando Amazon — a Proteção de terminação é desativada, a menos que o valor `TerminationProtected` booleano esteja definido como. `true`

Console

Para ativar ou desativar a proteção contra encerramento ao criar um cluster com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Para a versão EMR de lançamento, escolha emr-6.6.0 ou posterior.
4. Em Encerramento de cluster e substituição de nós, verifique se a opção Usar proteção de encerramento está pré-selecionada ou desmarque a seleção para desativá-la.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

AWS CLI

Para ativar ou desativar a proteção contra encerramento ao criar um cluster usando o AWS CLI

- Com o AWS CLI, você pode iniciar um cluster com a proteção de encerramento ativada com o `create-cluster` comando com o `--termination-protected` parâmetro. Por padrão, a proteção contra encerramento é desativada.

  O exemplo a seguir cria um cluster com proteção contra encerramento habilitada:

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --name "TerminationProtectedCluster" --release-
label emr-7.2.0 \
--applications Name=Hadoop Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey --instance-type m5.xlarge \
--instance-count 3 --termination-protected
```

Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consultehttps://docs.aws.amazon.com/cli/latest/reference/emr.

Configurar a proteção contra término para clusters em execução

Você pode configurar a proteção contra término para um cluster em execução usando o console ou a AWS CLI.

Console

Para ativar ou desativar a proteção contra encerramento de um cluster em execução com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EMREC2Em Ativado, no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar.
3. Na guia Propriedades na página de detalhes do cluster, localize Término do cluster e selecione Editar.
4. Marque ou desmarque a caixa de seleção Usar proteção contra término para ativar ou desativar o atributo. Depois selecione Salvar alterações para confirmar.

AWS CLI

Para ativar ou desativar a proteção contra encerramento de um cluster em execução usando o AWS CLI

- Para habilitar a proteção contra término em um cluster em execução usando a AWS CLI, digite o comando `modify-cluster-attributes` com o parâmetro `--termination-protected`. Para desabilitá-la, use o parâmetro `--no-termination-protected`.

  O exemplo a seguir ativa a proteção contra encerramento no cluster com ID *j-3KVTXXXXXX7UG*:

  ```
  aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --termination-protected
  ```

  O exemplo a seguir desabilita a proteção contra encerramento no mesmo cluster:

  ```
  aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --no-termination-protected
  ```

Substituindo nós não íntegros

A Amazon usa EMR periodicamente o serviço de verificação de NodeManager saúde no Apache Hadoop para monitorar o status dos nós principais em seus clusters Amazon on Amazon. EMR EC2 Se um nó não estiver funcionando de forma ideal, o verificador de saúde reporta esse nó ao controlador da Amazon. EMR O EMR controlador da Amazon adiciona o nó a uma lista de negação, impedindo que o nó receba novos YARN aplicativos até que o status do nó melhore. Um motivo comum pelo qual um nó pode ficar insalubre é a utilização excessiva do disco. Para obter mais informações sobre a identificação de nós não íntegros e a recuperação, consulte Erros de recursos.

Você pode escolher se a Amazon EMR deve encerrar os nós não íntegros ou mantê-los no cluster. Se você desativar a substituição de nós não íntegros, os nós não íntegros permanecerão na lista de rejeição e continuarão a contar para a capacidade do cluster. Você ainda pode se conectar à sua instância EC2 principal da Amazon para configuração e recuperação, para poder redimensionar seu cluster para aumentar a capacidade. Observe que a Amazon EMR substituirá os nós não íntegros mesmo se a proteção de encerramento estiver ativada.

Se a substituição de nós não íntegros estiver ativada, a Amazon EMR encerrará o nó principal não íntegro e provisionará uma nova instância com base no número de instâncias no grupo de instâncias ou na capacidade alvo das frotas de instâncias. Se vários ou todos os nós principais ficarem insalubres por mais de 45 minutos, a Amazon os EMR substituirá normalmente.

> **Important**
>
> Para evitar a possibilidade de perda permanente de HDFS dados, já que a Amazon substitui EMR normalmente uma instância central não íntegra, recomendamos que você sempre faça backup de seus dados.

A Amazon EMR publica o Amazon CloudWatch Events para substituição de nós com problemas de integridade, para que você possa acompanhar o que está acontecendo com suas instâncias principais não íntegras. Para obter mais informações, consulte eventos de substituição de nós não íntegros.

Configurações padrão de substituição e proteção de terminação de nós

A substituição de nós não íntegros está disponível para todas as EMR versões da Amazon, mas as configurações padrão dependem da etiqueta de lançamento que você escolher. Você pode alterar qualquer uma dessas configurações configurando a substituição de nós não íntegra ao criar um novo cluster ou acessando a configuração do cluster a qualquer momento.

Se você estiver criando um cluster de nó único ou de alta disponibilidade que esteja executando a EMR versão 7.0 ou inferior da Amazon, a configuração padrão de substituição de nó não íntegra depende da proteção contra encerramento:

- Ativar a proteção de terminação desativa a substituição não íntegra do nó.
- A desativação da proteção de terminação permite a substituição de nós não íntegra.

Configurando a substituição de nós não íntegra ao iniciar um cluster

Você pode ativar ou desativar a substituição não íntegra de nós ao iniciar um cluster usando o console AWS CLI, o ou o. API

A configuração padrão de substituição de nós não íntegros depende de como você executa o cluster:

- EMRConsole da Amazon — a substituição de nós não íntegros é ativada por padrão.

- AWS CLI `aws emr create-cluster`— a substituição de nós não íntegros é ativada por padrão, a menos que você especifique`--no-unhealthy-node-replacement`.
- EMRRunJobFlow APIComando Amazon — a substituição de nós não íntegros é ativada por padrão, a menos que você defina o valor `UnhealthyNodeReplacement` booleano como `True` ou. `False`

Console

Para ativar ou desativar a substituição de nós não íntegros ao criar um cluster com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Para a versão de EMR lançamento, escolha a etiqueta de EMR lançamento da Amazon que você deseja.
4. Em Encerramento do cluster e substituição de nós, verifique se a substituição de nó não íntegra (recomendada) está pré-selecionada ou desmarque a seleção para desativá-la.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

AWS CLI

Para ativar ou desativar a substituição de nós não íntegros ao criar um cluster usando o AWS CLI

- Com o AWS CLI, você pode iniciar um cluster com a substituição de nós não íntegros ativada com o `create-cluster` comando com o `--unhealthy-node-replacement` parâmetro. A substituição de nós não íntegros está ativada por padrão.

  O exemplo a seguir cria um cluster com a substituição de nós não íntegros ativada:

  > **Note**
  >
  > Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --name "SampleCluster" --release-label emr-7.2.0 \
--applications Name=Hadoop Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey --instance-type m5.xlarge \
--instance-count 3 --unhealthy-node-replacement
```

Para obter mais informações sobre o uso de EMR comandos da Amazon no AWS CLI, consulte EMR AWS CLI Comandos da Amazon.

Configurando a substituição de nós não íntegra em um cluster em execução

Você pode ativar ou desativar a substituição de nós não íntegros de um cluster em execução usando o console AWS CLI, o ou o. API

Console

Para ativar ou desativar a substituição de nós não íntegros em um cluster em execução com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EMREC2Em Ativado, no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar.
3. Na guia Propriedades da página de detalhes do cluster, encontre Terminação do cluster e substituição do nó e selecione Editar.
4. Marque ou desmarque a caixa de seleção substituição de nó não íntegra para ativar ou desativar o recurso. Depois selecione Salvar alterações para confirmar.

AWS CLI

Para ativar ou desativar a substituição de nós não íntegros em um cluster em execução usando o AWS CLI

- Para ativar a substituição de nós não íntegros em um cluster em execução com o AWS CLI, use o `modify-cluster-attributes` comando com o `--unhealthy-node-`

`replacement` parâmetro. Para desabilitá-la, use o parâmetro `--no-unhealthy-node-replacement`.

O exemplo a seguir ativa a substituição de nós não íntegros no cluster com ID *j-3KVTXXXXXX7UG*:

```
aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --unhealthy-node-replacement
```

O exemplo a seguir desativa a substituição de nós não íntegros no mesmo cluster:

```
aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --no-unhealthy-node-replacement
```

Trabalhando com o Amazon Linux AMIs na Amazon EMR

Amazon Linux Amazon Machine Images (AMIs)

A Amazon EMR usa um Amazon Linux Amazon Machine Image (AMI) para inicializar EC2 instâncias da Amazon quando você cria e executa um cluster. O AMI contém o sistema operacional Amazon Linux, outros softwares e as configurações necessárias para cada instância hospedar seus aplicativos de cluster.

Por padrão, quando você cria um cluster, a Amazon EMR usa um Amazon Linux padrão AMI que é criado especificamente para a versão de EMR lançamento da Amazon que você usa. Para obter mais informações sobre o Amazon Linux padrãoAMI, consulteUsando o Amazon Linux padrão AMI para Amazon EMR. Ao usar o Amazon EMR 5.7.0 ou superior, você pode escolher especificar um Amazon Linux personalizado AMI em vez do Amazon Linux padrão para a AMI Amazon. EMR Um personalizado AMI permite criptografar o volume do dispositivo raiz e personalizar aplicativos e configurações como alternativa ao uso de ações de bootstrap. Você pode especificar um personalizado AMI para cada tipo de instância na configuração do grupo de instâncias ou da frota de instâncias de um EMR cluster da Amazon. O AMI suporte personalizado múltiplo oferece a flexibilidade de usar mais de um tipo de arquitetura em um cluster. Consulte Usando um personalizado AMI.

A Amazon anexa EMR automaticamente um SSD volume Amazon EBS General Purpose como dispositivo raiz para todosAMIs. EBS- apoiado por AMIs melhorar o desempenho. Para obter mais informações sobre o Amazon LinuxAMIs, consulte Amazon Machine Images (AMI). Para

obter mais informações sobre armazenamento de instâncias para EMR instâncias da Amazon, consulteArmazenamento de instâncias.

Usando o Amazon Linux padrão AMI para Amazon EMR

Cada versão de EMR lançamento da Amazon usa um Amazon Linux padrão AMI para Amazon, EMR a menos que você especifique um personalizadoAMI. A partir das versões Amazon EMR 5.36, Amazon EMR 6.6 e Amazon EMR 7.0, o comportamento padrão para atualizar o Amazon Linux 2 (AL2para EMR 5.x e 6.x, AL2 0.23 para 7.xEMR) em um padrão da Amazon é aplicar automaticamente AMI a versão mais recente do Amazon Linux à Amazon EMR padrão. EMR AMI

Atualizações automáticas do Amazon Linux para EMR versões da Amazon

Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 7.0 ou superior, 6.6 ou superior ou 5.36 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux para a Amazon padrão. EMR AMI Por exemplo:

- Onde há uma versão `x.x.0` e uma, a `x.x.1` `x.x.0` versão deixa de receber AMI atualizações quando é `x.x.1` lançada.
- Da mesma forma`x.x.1`, para de receber AMI atualizações ao ser `x.x.2` lançado.
- Posteriormente, quando `x.y.0` lançado, `x.x.[latest]` continua recebendo AMI atualizações ao lado`x.y.[latest]`.

Para ver se você está usando a versão de patch mais recente, conforme indicado pelo número após o segundo ponto decimal (`6.8.1`) para uma EMR versão da Amazon, consulte as versões disponíveis no Guia de EMR lançamento da Amazon, verifique a lista suspensa de lançamentos da Amazon EMR ao criar um cluster no console ou use a ação ou. `ListReleaseLabels`API`list-release-labels`CLI Para receber atualizações quando lançarmos uma nova EMR versão da Amazon, assine o RSS feed do What's new? página no Guia de lançamento.

Se quiser, você pode optar por iniciar seu cluster com a versão do Amazon Linux com a qual a versão da Amazon EMR foi enviada pela primeira vez. Para obter informações sobre como especificar a versão do Amazon Linux para o cluster, consulte Alterando a versão do Amazon Linux ao criar um EMR cluster.

Versões padrão do Amazon Linux

Tópicos

- Padrão AMIs para Amazon EMR 7.0 e superior
- Padrão AMIs para Amazon EMR 6.6 e superior
- Padrão AMIs para Amazon EMR 5.x

Padrão AMIs para Amazon EMR 7.0 e superior

A tabela a seguir lista as informações do Amazon Linux para a versão de patch mais recente das EMR versões 7.0 e superiores da Amazon.

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2023.5.2 240708.(	6.1.96-102.177.amzn2023	23 de julho de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-southeast-4 • ap-northeast-1 • ap-northeast-2

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1 • ca-central-1 • il-central-1 • ca-west-1 • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2023.3.2 240304.(	6.1.79-99.164.amzn2023	12 de março de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-southeast-4 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• ca-central-1 • il-central-1 • ca-west-1 • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2023.3.2 240219.(	6.1.77-99.164.amzn2023	1.º de março de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-southeast-4 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• ca-central-1 • il-central-1 • ca-west-1 • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2023.3.2 240205.(	6.1.75-99.163.amzn2023	19 de fevereiro de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-southeast-4 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• ca-central-1 • il-central-1 • ca-west-1 • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2023.3.2 240122.(	6.1.72-96.166.amzn2023	5 de fevereiro de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-southeast-4 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• ca-central-1 • il-central-1 • ca-west-1 • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2023.3.2 240108.(	6.1.72-96.166.amzn2023	24 de janeiro de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-southeast-4 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• ca-central-1 • il-central-1 • ca-west-1 • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2023.3.2 231211.4	6.1.66-91.160.amzn2023	19 de dezembro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-southeast-4 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• ca-central-1 • il-central-1 • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

Padrão AMIs para Amazon EMR 6.6 e superior

A tabela a seguir lista as informações do Amazon Linux para a versão de patch mais recente das EMR versões 6.6.x e superiores da Amazon.

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20240709.1	4.14.348	23 de julho de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10.1+) • eu-south-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• eu-south-2 (6.10.1+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10.1+) • ap-southeast-3 • ap-southeast-4 (6.8.1+ e 5.36,1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 (6.10.1+) • me-south-1 • ca-central-1 • il-central-1 (6.8.1+ e 5.36,1) • ca-west-1 (6.9.1+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20240223.0	4.14.336	8 de março de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10.1+) • eu-south-1 • eu-south-2 (6.10.1+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10.1+) • ap-southeast-3 • ap-southeast-4 (6.8.1+ e 5.36,1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10.1+) • me-south-1 • ca-central-1 • il-central-1 (6.8.1+ e 5.36,1) • ca-west-1 (6.9.1+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20240131.0	4.14.336	14 de fevereiro de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10.1+) • eu-south-1 • eu-south-2 (6.10.1+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10.1+) • ap-southeast-3 • ap-southeast-4 (6.8.1+ e 5.36,1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10.1+) • me-south-1 • ca-central-1 • il-central-1 (6.8.1+ e 5.36,1) • ca-west-1 (6.9.1+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-centra l-2 (6.10.1+) • eu-south-1 • eu-south-2 (6.10.1+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10.1+) • ap-southeast-3 • ap-southe ast-4 (6.8.1+ e 5.36,1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10.1+) • me-south-1 • ca-central-1 • il-central-1 (6.8.1+ e 5.36,1) • ca-west-1 (6.9.1+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20240109.0	4.14.334	24 de janeiro de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10.1+) • eu-south-1 • eu-south-2 (6.10.1+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10.1+) • ap-southeast-3 • ap-southeast-4 (6.8.1+ e 5.36,1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10.1+) • me-south-1 • ca-central-1 • il-central-1 (6.8.1+ e 5.36,1) • ca-west-1 (6.9.1+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20231218.0	4.14.330	2 de janeiro de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10+) • me-south-1 • ca-central-1 • il-central-1 (6.8+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20231206.0	4.14.330	22 de dezembro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10+) • me-south-1 • ca-central-1 • il-central-1 (6.8+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10+) • me-south-1 • ca-central-1 • il-central-1 (6.8+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20231101.0	4.14.327	17 de novembro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10+) • me-south-1 • ca-central-1 • il-central-1 (6.8+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20231020.1	4.14.326	7 de novembro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10+) • me-south-1 • ca-central-1 • il-central-1 (6.8+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230912.1	4.14.326	26 de outubro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10+) • me-south-1 • ca-central-1 • il-central-1 (6.8+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230926.0	4.14.322	19 de outubro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10+) • me-south-1 • ca-central-1 • il-central-1 (6.8+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20238906.0	4.14.322	4 de outubro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• `me-central-1` (6.10+) • `me-south-1` • `ca-central-1` • `il-central-1` (6.9+ e 5.36.1)

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230822.0	4.14.322	30 de agosto de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• `me-central-1` (6.10+) • `me-south-1` • `ca-central-1` • `il-central-1` (6.9+ e 5.36.1)

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2023 808.0	4.14.320	24 de agosto de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10+) • me-south-1 • ca-central-1 • il-central-1 (6.9+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230727.0	4.14.320	14 de agosto de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• `me-central-1` (6.10+) • `me-south-1` • `ca-central-1` • `il-central-1` (6.9+ e 5.36.1)

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230719.0	4.14.320	2 de agosto de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• `me-central-1` (6.10+) • `me-south-1` • `ca-central-1` • `il-central-1` (6.9+ e 5.36.1)

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2023 628.0	4.14.318	12 de julho de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 (6.10+)

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230612.0	4.14.314	23 de junho de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 (6.10+)

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-south-1 • ca-central-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2023 504.1	4.14.313	16 de maio de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230418.0	4.14.311	3 de maio de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10 somente) • eu-south-1 • eu-south-2 (6.10 somente) • ap-east-1 • ap-south-1 • ap-south-2 (6.10 somente) • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-south-1 • ca-central-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2023 404.1	4.14.311	18 de abril de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1 • ca-central-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2023 404.0	4.14.311	10 de abril de 2023	• `us-east-1` • `eu-west-3`

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2023 320.0	4.14.309	30 de março de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230307.0	4.14.305	15 de março de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1 • ca-central-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2023 207.0	4.14.304	3 de março de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20231119.1	4.14.301	9 de fevereiro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20221210.1	4.14.301	12 de janeiro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20221103.3	4.14.296	5 de dezembro de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2022 004.0	4.14.294	2 de novembro de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20220912.1	4.14.291	7 de outubro de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1
2.0.20220805.0	4.14.287	30 de agosto de 2022	• us-west-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20220719.0	4.14.287	10 de agosto de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2022 426.0	4.14.281	10 de junho de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2022 406.1	4.14.275	2 de maio de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

Padrão AMIs para Amazon EMR 5.x

A tabela a seguir lista as informações do Amazon Linux para a versão de patch mais recente do Amazon EMR 5.x, versões 5.36 e superiores.

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20240709.1	4.14.348	23 de julho de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10.1+) • eu-south-1 • eu-south-2 (6.10.1+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10.1+) • ap-southeast-3 • ap-southeast-4 (6.8.1+ e 5.36,1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10.1+) • me-south-1 • ca-central-1 • il-central-1 (6.8.1+ e 5.36,1) • ca-west-1 (6.9.1+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230504.1	4.14.313	16 de maio de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • ca-central-1 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • me-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230418.0	4.14.311	3 de maio de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • ca-central-1 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • me-central-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2023 404.1	4.14.311	18 de abril de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • ca-central-1 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1
2.0.2023 404.0	4.14.311	10 de abril de 2023	• us-east-1 • eu-west-3

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230320.0	4.14.309	30 de março de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • ca-central-1 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230307.0	4.14.305	15 de março de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • ca-central-1 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2023 207.0	4.14.304	3 de março de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20221210.1	4.14.301	12 de janeiro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20221103.3	4.14.296	5 de dezembro de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20220004.0	4.14.294	2 de novembro de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20220912.1	4.14.291	7 de outubro de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20220719.0	4.14.287	10 de agosto de 2022	• us-west-1 • eu-west-3 • eu-north-1 • eu-central-1 • ap-south-1 • me-south-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20220426.0	4.14.281	14 de junho de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

Considerações sobre a atualização de software

Observe os comportamentos padrão de atualização de software a seguir.

Amazon EMR 7.x — Amazon Linux 2023

A Amazon EMR lança versões 7.0 e superiores executadas no Amazon Linux 2023 (AL2023). O comportamento padrão para AL2 023 é AMIs bloquear uma versão específica do repositório de software Amazon Linux. Portanto, as atualizações de segurança não são aplicadas toda vez que você executa um cluster. Em vez disso, o comportamento padrão das versões EMR 7.x da Amazon é aplicar automaticamente a última versão AL2 023 à Amazon padrão EMR AMI somente quando você cria o cluster. Para receber as atualizações de segurança mais recentes, recomendamos que você recrie seu cluster periodicamente.

Amazon EMR 5.x e 6.x — Amazon Linux e Amazon Linux 2

Para EMR versões da Amazon inferiores à 7.0, quando uma EC2 instância da Amazon é inicializada pela primeira vez em um cluster baseado no Amazon Linux (AL) ou no Amazon Linux 2 (AL2) padrão AMI para AmazonEMR, ela verifica as atualizações de software que se aplicam à versão de lançamento nos repositórios de pacotes habilitados para AL e Amazon. EMR Assim como em outras AL e AL2 instâncias, atualizações de segurança críticas e importantes desses repositórios são instaladas automaticamente.

Observe também que, na sua configuração de rede, você deve permitir HTTP e sair HTTPS dos repositórios Amazon Linux no Amazon S3. Caso contrário, as atualizações de segurança falharão. Para obter mais informações, consulte Amazon Linux - Package repository no Amazon EC2 User Guide. Por padrão, outros pacotes de software e atualizações do kernel que exigem uma reinicialização, incluindo NVIDIA eCUDA, são excluídos do download automático na primeira inicialização.

Amazon EMR 5.35.0 e inferior e 6.5.0 e inferior — Amazon Linux bloqueado AMI para a versão de lançamento da Amazon EMR

Para o Amazon EMR 5.35.0 e inferior e 6.5.0 e inferior, o padrão AMI é baseado na maior parte do up-to-date Amazon Linux AMI disponível no momento do lançamento da Amazon. EMR Foi AMI testado quanto à compatibilidade com os aplicativos de big data e os EMR recursos da Amazon incluídos nessa versão de lançamento.

Cada versão de EMR lançamento do Amazon EMR 5.35.0 e inferior e 6.5.0 e inferior da Amazon está “bloqueada” para sua respectiva versão atribuída do Amazon Linux para manter a AMI compatibilidade. Por esse motivo, recomendamos que você use a versão mais recente da AmazonEMR, a menos que precise de uma versão inferior para fins de compatibilidade e não consiga migrar. Se você precisar usar uma versão inferior da Amazon EMR para fins de

compatibilidade, recomendamos que você use a versão mais recente de uma série. Por exemplo, se você precisar usar a série 5.12, use a 5.12.0 em vez da 5.12.2 5.12.1. Se uma nova versão se tornar disponível em uma série, considere a migração de seus aplicativos para a nova versão.

Para obter mais informações sobre o comportamento de atualização automática introduzido com o Amazon EMR 5.36.0 e superior e 6.6.0 e superior, consulte. Atualizações automáticas do Amazon Linux para EMR versões da Amazon

O comportamento de inicialização padrão exclui atualizações de kernel

Quando uma EC2 instância da Amazon em um cluster baseado no Amazon Linux padrão AMI para Amazon é EMR inicializada pela primeira vez, ela verifica os repositórios de pacotes habilitados para Amazon Linux e Amazon em EMR busca de atualizações de software que se apliquem à AMI versão. Assim como em outras EC2 instâncias da Amazon, atualizações de segurança críticas e importantes desses repositórios são instaladas automaticamente.

No entanto, se você estiver usando uma versão mais antiga do Amazon LinuxAMI, a atualização de segurança mais recente pode não ser instalada automaticamente. Isso ocorre porque os repositórios aos quais seu EMR cluster faz referência são fixos para cada versão do Amazon Linux. AMI

Observe também que, na sua configuração de rede, você deve permitir HTTP e sair HTTPS dos repositórios Amazon Linux no Amazon S3. Caso contrário, as atualizações de segurança falharão. Para obter mais informações, consulte Amazon Linux - Package repository no Amazon EC2 User Guide. Por padrão, outros pacotes de software e atualizações do kernel que exigem uma reinicialização, incluindo NVIDIA eCUDA, são excluídos do download automático na primeira inicialização.

> **Important**
>
> EMRclusters que executam AL2 023 usam o comportamento padrão do Amazon Linux, e suas Amazon Machine Images (AMIs) estão bloqueadas em uma versão específica do repositório Amazon Linux. Por padrão, seus clusters não receberão automaticamente as atualizações de segurança do software na execução. Seus clusters contêm somente as atualizações que estavam disponíveis na versão AL2 023 AMI que você escolheu quando criou seu cluster. Para obter mais informações, consulte Atualização do Amazon Linux 2023 no Guia do usuário do Amazon Linux 2023.

> **Important**
>
> EMRclusters que executam Amazon Linux ou Amazon Linux 2 Amazon Machine Images (AMIs) usam o comportamento padrão do Amazon Linux e não baixam e instalam automaticamente atualizações importantes e críticas do kernel que exigem uma reinicialização. Esse é o mesmo comportamento de outras EC2 instâncias da Amazon que executam o Amazon Linux padrãoAMI. Se novas atualizações de software Amazon Linux que exigem uma reinicialização (como kernel e CUDA atualizações) ficarem disponíveis após a disponibilização de uma EMR versão da Amazon, as instâncias de EMR cluster que executam o padrão AMI não baixam e instalam automaticamente essas atualizações. NVIDIA Para obter atualizações do kernel, você pode personalizar sua Amazon EMR AMI para usar o Amazon Linux AMI mais recente.

O cluster é iniciado com ou sem atualizações

Lembre-se de que, se não foi possível instalar atualizações de software porque os repositórios de pacotes estão inacessíveis na primeira inicialização do cluster, a instância do cluster ainda concluirá sua execução. Por exemplo, os repositórios podem estar inacessíveis porque o S3 está temporariamente indisponível, ou você pode ter VPC regras de firewall configuradas para bloquear o acesso.

Não executar **`sudo yum update`**

Quando você se conecta a uma instância de cluster usandoSSH, as primeiras linhas de saída de tela fornecem um link para as notas de lançamento do Amazon Linux AMI que a instância usa, um aviso da AMI versão mais recente do Amazon Linux, um aviso do número de pacotes disponíveis para atualização nos repositórios habilitados e uma diretiva para execução`sudo yum update`.

> **Important**
>
> É altamente recomendável que você não execute `sudo yum update` em instâncias de cluster enquanto estiver conectado SSH ou ao usar uma ação de bootstrap. Isso pode causar incompatibilidades porque todos os pacotes são instalados indiscriminadamente.

Práticas recomendadas de atualização de software

Práticas recomendadas de gerenciamento de atualizações de software

- Se você usa uma versão inferior da AmazonEMR, considere e teste uma migração para a versão mais recente antes de atualizar os pacotes de software.
- Se você migrar para uma versão posterior ou atualizar pacotes de software, teste primeiro a implementação em um ambiente que não seja de produção. A opção de clonar clusters com o EMR console da Amazon é útil para isso.
- Avalie as atualizações de software para seus aplicativos e para sua versão do AMI Amazon Linux individualmente. Somente teste e instale pacotes em ambientes de produção que você determinar que são totalmente necessários para sua postura de segurança, funcionalidade do aplicativo ou desempenho.
- Acompanhe o Amazon Linux Security Center para verificar se há atualizações.
- Evite instalar pacotes conectando-se a instâncias de cluster individuais usando SSH o. Em vez disso, use uma ação de bootstrap para instalar e atualizar pacotes em todas as instâncias de cluster conforme necessário. Isso requer que você encerre um cluster e reinicie-o. Para obter mais informações, consulte Criar ações de bootstrap para instalar softwares adicionais.

Usando um personalizado AMI

Ao usar o Amazon EMR 5.7.0 ou superior, você pode escolher especificar um Amazon Linux personalizado AMI em vez do Amazon Linux padrão para a AMI Amazon. EMR Um personalizado AMI é útil se você quiser fazer o seguinte:

- Pré-instale aplicativos e realize outras personalizações em vez de usar ações de bootstrap. Isso pode melhorar o tempo de início do cluster e agilizar o fluxo de trabalho de inicialização. Para obter mais informações e um exemplo, consulte Criação de um Amazon Linux personalizado AMI a partir de uma instância pré-configurada.
- Implementar configurações de cluster e o nó mais sofisticadas do que o permitido por ações de bootstrap.
- Criptografe os volumes do dispositivo EBS raiz (volumes de inicialização) das EC2 instâncias em seu cluster se você estiver usando uma EMR versão da Amazon inferior à 5.24.0. Assim como no padrãoAMI, o tamanho mínimo do volume raiz para um personalizado AMI é 10 GiB para as EMR versões 6.9 e inferiores da Amazon e 15 GiB para as versões 6.10 e superiores da

AmazonEMR. Para obter mais informações, consulte Criação de um volume personalizado AMI com um dispositivo EBS raiz criptografado da Amazon.

> **Note**
>
> A partir da EMR versão 5.24.0 da Amazon, você pode usar uma opção de configuração de segurança para criptografar o dispositivo EBS raiz e os volumes de armazenamento ao especificar AWS KMS como seu provedor de chaves. Para obter mais informações, consulte Criptografia de disco local.

Um personalizado AMI deve existir na mesma AWS região em que você cria o cluster. Também deve corresponder à arquitetura da EC2 instância. Por exemplo, uma instância m5.xlarge tem a arquitetura x86_64. Portanto, para provisionar um m5.xlarge usando um personalizadoAMI, seu personalizado também AMI deve ter a arquitetura x86_64. Da mesma forma, para provisionar uma instância m6g.xlarge, que tem a arquitetura arm64, sua instância personalizada AMI deve ter a arquitetura arm64. Para obter mais informações sobre como identificar um Linux AMI para seu tipo de instância, consulte Encontre um Linux AMI no Guia EC2 do usuário da Amazon.

> **Important**
>
> EMRclusters que executam Amazon Linux ou Amazon Linux 2 Amazon Machine Images (AMIs) usam o comportamento padrão do Amazon Linux e não baixam e instalam automaticamente atualizações importantes e críticas do kernel que exigem uma reinicialização. Esse é o mesmo comportamento de outras EC2 instâncias da Amazon que executam o Amazon Linux padrãoAMI. Se novas atualizações de software Amazon Linux que exigem uma reinicialização (como kernel e CUDA atualizações) ficarem disponíveis após a disponibilização de uma EMR versão da Amazon, as instâncias de EMR cluster que executam o padrão AMI não baixam e instalam automaticamente essas atualizações. NVIDIA Para obter atualizações do kernel, você pode personalizar sua Amazon EMR AMI para usar o Amazon Linux AMI mais recente.

Criação de um Amazon Linux personalizado AMI a partir de uma instância pré-configurada

As etapas básicas para pré-instalar o software e realizar outras configurações para criar um Amazon Linux personalizado para a AMI Amazon EMR são as seguintes:

- Execute uma instância a partir do Amazon Linux básicoAMI.
- Conecte-se à instância para instalar o software e realizar outras personalizações.
- Crie uma nova imagem (AMIinstantâneo) da instância que você configurou.

Depois de criar a imagem com base na sua instância personalizada, você pode copiar essa imagem para um destino criptografado conforme descrito em Criação de um volume personalizado AMI com um dispositivo EBS raiz criptografado da Amazon.

Tutorial: criar uma AMI a partir de uma instância com software personalizado instalado

Para iniciar uma EC2 instância com base no Amazon Linux mais recente AMI

1. Use o AWS CLI para executar o comando a seguir, que cria uma instância a partir de uma existenteAMI. *MyKeyName*Substitua pelo par de chaves que você usa para se conectar à instância e *MyAmiId* com o ID de um Amazon Linux apropriadoAMI. Para as mais recentes AMIIDs, consulte Amazon Linux AMI.

   > **Note**
   >
   > Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

   ```
   aws ec2 run-instances --image-id MyAmiID \
   --count 1 --instance-type m5.xlarge \
   --key-name MyKeyName --region us-west-2
   ```

   O valor de saída `InstanceId` é usado como *MyInstanceId* na próxima etapa.

2. Execute o seguinte comando:

   ```
   aws ec2 describe-instances --instance-ids MyInstanceId
   ```

   O valor de saída `PublicDnsName` é usado para se conectar à instância na próxima etapa.

Para se conectar à instância e instalar o software

1. Use uma SSH conexão que permita executar comandos shell na sua instância Linux. Para obter mais informações, consulte Conectando-se à sua instância Linux usando SSH o Amazon EC2 User Guide.
2. Realize todas as personalizações necessárias. Por exemplo:

```
sudo yum install MySoftwarePackage
sudo pip install MySoftwarePackage
```

Para criar um snapshot da imagem personalizada

- Depois de personalizar a instância, use o `create-image` comando para criar uma a AMI partir da instância.

```
aws ec2 create-image --no-dry-run --instance-id MyInstanceId --name MyEmrCustomAmi
```

O valor de saída `imageID` é usado quando você executa o cluster ou cria um snapshot criptografados. Para ter mais informações, consulte Use um único personalizado AMI em um EMR cluster e Criação de um volume personalizado AMI com um dispositivo EBS raiz criptografado da Amazon.

Como usar um personalizado AMI em um EMR cluster da Amazon

Você pode usar um customizado AMI para provisionar um EMR cluster da Amazon de duas maneiras:

- Use uma única personalização AMI para todas as EC2 instâncias no cluster.
- Use personalização diferente AMIs para os diferentes tipos de EC2 instância usados no cluster.

Você pode usar somente uma das duas opções ao provisionar um EMR cluster e não pode alterá-la depois que o cluster for iniciado.

Considerações sobre o uso de um ou vários personalizados AMIs em um cluster da Amazon EMR

Consideração	Personalização única AMI	Vários personalizados AMIs
Use os processadores x86 e Graviton2 personalizados AMIs no mesmo cluster	✗ Sem suporte	✓ Há suporte
AMIa personalização varia entre os tipos de instância	✗ Sem suporte	✓ Há suporte
Altere a personalização AMIs ao adicionar novos grupos/frotas de instâncias de tarefas a um cluster em execução. Observação: você não pode alterar o costume dos AMI grupos/frotas de instâncias existentes.	✗ Sem suporte	✓ Há suporte
Use o AWS console para iniciar um cluster	✓ Há suporte	✗ Sem suporte
Use AWS CloudFormation para iniciar um cluster	✓ Há suporte	✓ Há suporte

Use um único personalizado AMI em um EMR cluster

Para especificar uma AMI ID personalizada ao criar um cluster, use uma das seguintes opções:

- AWS Management Console
- AWS CLI
- Amazon EMR SDK
- Amazon EMR API RunJobFlow
- AWS CloudFormation (veja a `CustomAmiID` propriedade em Cluster InstanceGroupConfig, Cluster InstanceTypeConfig InstanceGroupConfig, Recurso ou Recurso InstanceFleetConfig - InstanceTypeConfig)

Amazon EMR console

Para especificar um único personalizado AMI no console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Em Nome e aplicações, localize Opções do sistema operacional. Escolha Personalizado AMI e insira seu AMI ID no AMI campo Personalizado.
4. Escolha qualquer outra opção que se aplique ao cluster.
5. Para iniciar o cluster, escolha Criar cluster.

AWS CLI

Para especificar um único personalizado AMI com o AWS CLI

- Use o `--custom-ami-id` parâmetro para especificar o AMI ID ao executar o `aws emr create-cluster` comando.

  O exemplo a seguir especifica um cluster que usa um único personalizado AMI com um volume de inicialização de 20 GiB. Para obter mais informações, consulte Personalizando o volume do dispositivo EBS raiz da Amazon.

  > **Note**
  >
  > Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --name "Cluster with My Custom AMI" \
--custom-ami-id MyAmiID --ebs-root-volume-size 20 \
--release-label emr-5.7.0 --use-default-roles \
--instance-count 2 --instance-type m5.xlarge
```

Use vários personalizados AMIs em um EMR cluster da Amazon

Para criar um cluster usando várias opções personalizadasAMIs, use uma das seguintes opções:

- AWS CLIversão 1.20.21 ou superior
- AWS SDK
- Amazon EMR RunJobFlowna Amazon EMR API Reference
- AWS CloudFormation (veja a `CustomAmiID` propriedade em Cluster InstanceGroupConfig, Cluster InstanceTypeConfig InstanceGroupConfig, Recurso ou Recurso InstanceFleetConfig - InstanceTypeConfig)

Atualmente, o AWS Management Console não oferece suporte à criação de um cluster usando vários modelos personalizadosAMIs.

Example - Use o AWS CLI para criar um cluster de grupos de instâncias usando vários modelos personalizados AMIs

Usando a AWS CLI versão 1.20.21 ou superior, você pode atribuir um único personalizado AMI a todo o cluster ou pode atribuir vários personalizados AMIs a cada nó da instância em seu cluster.

O exemplo a seguir mostra um cluster uniforme de grupos de instâncias criado com dois tipos de instância (m5.xlarge) usados em todos os tipos de nós (primário, central, de tarefa). Cada nó tem vários itens personalizadosAMIs. O exemplo ilustra vários recursos da AMI configuração personalizada múltipla:

- Não há AMI atribuição personalizada no nível do cluster. Isso é para evitar conflitos entre os vários personalizados AMIs e um único personalizadoAMI, o que faria com que a inicialização do cluster falhasse.
- O cluster pode ter vários nós de tarefas principais, centrais e individuais personalizadosAMIs. Isso permite AMI personalizações individuais, como aplicativos pré-instalados, configurações sofisticadas de cluster e volumes criptografados do dispositivo raiz da AmazonEBS.
- O nó principal do grupo de instâncias pode ter somente um tipo de instância e a personalização correspondenteAMI. Da mesma forma, o nó primário pode ter somente um tipo de instância e a personalização correspondenteAMI.
- O cluster pode ter múltiplos nós de tarefa.

```
aws emr create-cluster --instance-groups
```

```
InstanceGroupType=PRIMARY,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-123456
InstanceGroupType=CORE,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-234567
InstanceGroupType=TASK,InstanceType=m6g.xlarge,InstanceCount=1,CustomAmiId=ami-345678
InstanceGroupType=TASK,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-456789
```

Example - Use a AWS CLI versão 1.20.21 ou superior para adicionar um nó de tarefa a um cluster de grupos de instâncias em execução com vários tipos de instância e vários personalizados AMIs

Usando a AWS CLI versão 1.20.21 ou superior, você pode adicionar vários personalizados AMIs a um grupo de instâncias que você adiciona a um cluster em execução. O argumento `CustomAmiId` pode ser usado com o comando `add-instance-groups`, conforme mostrado no exemplo a seguir. Observe que a mesma AMI ID personalizada múltipla (ami-123456) é usada em mais de um nó.

```
aws emr create-cluster --instance-groups
InstanceGroupType=PRIMARY,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-123456
InstanceGroupType=CORE,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-123456
InstanceGroupType=TASK,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-234567

{
    "ClusterId": "j-123456",
    ...
}

aws emr add-instance-groups --cluster-id j-123456 --instance-groups
 InstanceGroupType=Task,InstanceType=m6g.xlarge,InstanceCount=1,CustomAmiId=ami-345678
```

Example - Use a AWS CLI versão 1.20.21 ou superior para criar um cluster de frota de instâncias, vários tipos de instância personalizadosAMIs, primário sob demanda, núcleo sob demanda, vários núcleos e nós de tarefas

```
aws emr create-cluster --instance-fleets
InstanceFleetType=PRIMARY,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarg
 CustomAmiId=ami-123456}']
InstanceFleetType=CORE,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarge,C
{InstanceType=m6g.xlarge, CustomAmiId=ami-345678}']
InstanceFleetType=TASK,TargetSpotCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarge,Custo
{InstanceType=m6g.xlarge, CustomAmiId=ami-567890}']
```

Example - Use a AWS CLI versão 1.20.21 ou superior para adicionar nós de tarefas a um cluster em execução com vários tipos de instância e vários personalizados AMIs

```
aws emr create-cluster --instance-fleets
InstanceFleetType=PRIMARY,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarg
 CustomAmiId=ami-123456}']
InstanceFleetType=CORE,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarge,C
{InstanceType=m6g.xlarge, CustomAmiId=ami-345678}']

{
    "ClusterId": "j-123456",
    ...
}

aws emr add-instance-fleet --cluster-id j-123456 --instance-fleet
InstanceFleetType=TASK,TargetSpotCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarge,Custo
{InstanceType=m6g.xlarge, CustomAmiId=ami-345678}']
```

Gerenciando atualizações AMI do repositório de pacotes

Na primeira inicialização, por padrão, o Amazon Linux AMIs se conecta aos repositórios de pacotes para instalar atualizações de segurança antes que outros serviços sejam iniciados. Dependendo dos seus requisitos, você pode optar por desativar essas atualizações ao especificar uma personalização AMI para a AmazonEMR. A opção de desativar esse recurso está disponível somente quando você usa um personalizadoAMI. Por padrão, as atualizações de kernel do Amazon Linux e de outros pacotes de software que exigem uma reinicialização não são atualizados. Observe que sua configuração de rede deve permitir HTTP e HTTPS acessar os repositórios do Amazon Linux no Amazon S3, caso contrário, as atualizações de segurança não serão bem-sucedidas.

> **Warning**
>
> É altamente recomendável que você opte por atualizar todos os pacotes instalados na reinicialização ao especificar um personalizadoAMI. Se os pacotes de atualização não forem atualizados, poderá haver riscos de segurança adicionais.

Com o AWS Management Console, você pode selecionar a opção de desativar as atualizações ao escolher Personalizado AMI.

Com o AWS CLI, você pode especificar `--repo-upgrade-on-boot NONE` junto com `--custom-ami-id` ao usar o create-cluster comando.

Com a Amazon EMRAPI, você pode especificar `NONE` o RepoUpgradeOnBootparâmetro.

Criação de um volume personalizado AMI com um dispositivo EBS raiz criptografado da Amazon

Para criptografar o volume do dispositivo EBS raiz Amazon de um Amazon Linux AMI para AmazonEMR, copie uma imagem instantânea de um destino não criptografado AMI para um destino criptografado. Para obter informações sobre a criação de EBS volumes criptografados, consulte a EBScriptografia da Amazon no Guia EC2 do usuário da Amazon. A origem AMI do snapshot pode ser o Amazon Linux básicoAMI, ou você pode copiar um snapshot de um AMI derivado do Amazon Linux básico AMI que você personalizou.

> **Note**
>
> A partir da EMR versão 5.24.0 da Amazon, você pode usar uma opção de configuração de segurança para criptografar o dispositivo EBS raiz e os volumes de armazenamento ao especificar AWS KMS como seu provedor de chaves. Para obter mais informações, consulte Criptografia de disco local.

Você pode usar um provedor de chave externo ou uma AWS KMS chave para criptografar o volume EBS raiz. A função de serviço que a Amazon EMR usa (geralmente a padrão`EMR_DefaultRole`) deve ter permissão para criptografar e descriptografar o volume, no mínimo, para que EMR a Amazon crie um cluster com o. AMI Ao usar AWS KMS como provedor de chaves, isso significa que as seguintes ações devem ser permitidas:

- `kms:encrypt`
- `kms:decrypt`
- `kms:ReEncrypt*`
- `kms:CreateGrant`
- `kms:GenerateDataKeyWithoutPlaintext"`
- `kms:DescribeKey"`

A maneira mais simples de fazer isso é adicionar o perfil como um usuário de chave, conforme descrito no seguinte tutorial. O exemplo a seguir de declaração de política é fornecido caso você precise personalizar as políticas de função.

```
{
"Version": "2012-10-17",
"Statement": [
   {
   "Sid": "EmrDiskEncryptionPolicy",
   "Effect": "Allow",
   "Action": [
      "kms:Encrypt",
      "kms:Decrypt",
      "kms:ReEncrypt*",
      "kms:CreateGrant",
      "kms:GenerateDataKeyWithoutPlaintext",
      "kms:DescribeKey"
      ],
   "Resource": [
      "*"
      ]
   }
]
}
```

Tutorial: Criando um volume personalizado AMI com um dispositivo raiz criptografado usando uma KMS chave

A primeira etapa neste exemplo é encontrar ARN a KMS chave ou criar uma nova. Para obter mais informações sobre como criar chaves, consulte Creating keys no Guia do desenvolvedor do AWS Key Management Service . O procedimento a seguir mostra como adicionar a função de serviço padrão, `EMR_DefaultRole`, como um usuário e chave à política de chave. Anote o ARNvalor da chave ao criá-la ou editá-la. Você usa o ARN superior, quando você cria AMI o.

Para adicionar a função de serviço da Amazon EC2 à lista de usuários de chaves de criptografia com o console

1. Faça login no console AWS Management Console e abra o AWS Key Management Service (AWS KMS) em https://console.aws.amazon.com/kms.
2. Para alterar o Região da AWS, use o seletor de região no canto superior direito da página.

3. Escolha o alias da KMS chave a ser usada.
4. Na página de detalhes da chave, em Key Users (Usuários de chaves), escolha Add (Adicionar).
5. Na caixa de diálogo Anexar, escolha a função EMR de serviço da Amazon. O nome da função padrão é `EMR_DefaultRole`.
6. Escolha Anexar.

Para criar um criptografado AMI com o AWS CLI

- Use o `aws ec2 copy-image` comando do AWS CLI para criar um AMI com um volume de dispositivo EBS raiz criptografado e a chave que você modificou. Substitua o `--kms-key-id` valor especificado pela totalidade ARN da chave que você criou ou modificou para baixo.

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws ec2 copy-image --source-image-id MyAmiId \
--source-region us-west-2 --name MyEncryptedEMRAmi \
--encrypted --kms-key-id arn:aws:kms:us-west-2:12345678910:key/xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx
```

A saída do comando fornece o ID do AMI que você criou, que você pode especificar ao criar um cluster. Para obter mais informações, consulte Use um único personalizado AMI em um EMR cluster. Você também pode optar por personalizar isso AMI instalando o software e executando outras configurações. Para obter mais informações, consulte Criação de um Amazon Linux personalizado AMI a partir de uma instância pré-configurada.

Práticas recomendadas e considerações

Ao criar um personalizado AMI para a AmazonEMR, considere o seguinte:

- A série Amazon EMR 7.x é baseada no Amazon Linux 2023. Para essas EMR versões da Amazon, você precisa usar imagens baseadas no Amazon Linux 2023 para personalizaçãoAMIs. Para encontrar uma base personalizadaAMI, consulte Encontrando um Linux AMI.
- Para EMR versões da Amazon inferiores a 7.x, o Amazon Linux 2023 não AMIs é suportado.
- O Amazon EMR 5.30.0 e superior e a série Amazon EMR 6.x são baseados no Amazon Linux 2. Para essas EMR versões da Amazon, você precisa usar imagens baseadas no Amazon Linux 2 para personalizaçãoAMIs. Para encontrar uma base personalizadaAMI, consulte Encontrando um Linux AMI.
- Para EMR versões da Amazon inferiores a 5.30.0 e 6.x, o Amazon Linux 2 não AMIs é suportado.
- Você deve usar um Amazon Linux de 64 bitsAMI. Não AMI há suporte para 32 bits.
- O Amazon Linux AMIs com vários EBS volumes da Amazon não é suportado.
- Baseie sua personalização no EBS Amazon Linux mais recente com suporte. AMI Para obter uma lista do Amazon Linux AMIs e seus correspondentes AMIIDs, consulte Amazon Linux AMI.
- Não copie um snapshot de uma EMR instância existente da Amazon para criar uma personalizadaAMI. Isto provoca erros.
- Somente o tipo de HVM virtualização e as instâncias compatíveis com a Amazon EMR são compatíveis. Certifique-se de selecionar a HVM imagem e um tipo de instância compatível com a Amazon EMR ao passar pelo processo de AMI personalização. Para conhecer os tipos de virtualização e instâncias compatíveis, consulte Tipos de instâncias compatíveis.
- Sua função de serviço deve ter permissões de lançamento noAMI, portanto, ela AMI deve ser pública ou você deve ser o proprietário da AMI ou compartilhá-la com você pelo proprietário.
- Criar usuários no AMI com o mesmo nome dos aplicativos causa erros (por exemplo,`hadoop`,`hdfs`,`yarn`, ou`spark`).
- O conteúdo de`/tmp`,`/var`, e `/emr` (se existirem noAMI) é movido para `/mnt/tmp`/`mnt/var`, e `/mnt/emr` respectivamente durante a inicialização. Os arquivos são preservados, mas, se houver uma grande quantidade de dados, a inicialização poderá demorar mais do que o esperado.
- Se você usa um Amazon Linux personalizado AMI baseado em um Amazon Linux AMI com uma data de criação de 11/08/2018, o servidor Oozie falhará ao iniciar. Se você usa o Oozie, crie um personalizado AMI com base em um Amazon Linux AMI ID com uma data de criação diferente. Você pode usar o AWS CLI comando a seguir para retornar uma lista de imagens IDs para todo o HVM Amazon Linux AMIs com uma versão 2018.03, junto com a data de lançamento, para que você possa escolher um Amazon Linux apropriado AMI como base. MyRegion Substitua pelo seu identificador de região, como us-west-2.

```
aws ec2 --region MyRegion describe-images --owner amazon --query 'Images[?
Name!=`null`]|[?starts_with(Name, `amzn-ami-hvm-2018.03`) == `true`].
[CreationDate,ImageId,Name]' --output text | sort -rk1
```

- Nos casos em que você usa um VPC com um nome de domínio não padrão e AmazonProvidedDNS, você não deve usar a `rotate` opção na DNS configuração de sistemas operacionais.

Para obter mais informações, consulte Criação de um Linux EBS com suporte da Amazon AMI no Guia do EC2 usuário da Amazon.

Alterando a versão do Amazon Linux ao criar um EMR cluster

Quando você executa um cluster usando o Amazon EMR 6.6.0 ou superior, ele usa automaticamente a versão mais recente do Amazon Linux 2 que foi validada para a Amazon padrão. EMR AMI Você pode especificar uma versão diferente do Amazon Linux para seu cluster com o EMR console da Amazon ou AWS CLI o.

Amazon EMR console

Para alterar a versão do Amazon Linux ao criar um cluster usando o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Para a EMRversão, escolha emr-6.6.0 ou superior.
4. Em Opções do sistema operacional, escolha Versão do Amazon Linux e marque a caixa de seleção Aplicar automaticamente as últimas atualizações do Amazon Linux.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

AWS CLI

Alterar a versão do Amazon Linux ao criar um cluster usando a AWS CLI

- Use o parâmetro `--os-release-label` para especificar a versão do Amazon Linux ao executar o comando `aws emr` create-cluster.

```
aws emr create-cluster --name "Cluster with Different Amazon Linux Release" \
--os-release-label 2.0.20210312.1 \
--release-label emr-6.6.0 --use-default-roles \
--instance-count 2 --instance-type m5.xlarge
```

Personalizando o volume do dispositivo EBS raiz da Amazon

EBSpadrões de volume raiz

Com o Amazon EMR 4.x e superior, você pode especificar o tamanho do volume raiz ao criar um cluster. Com as EMR versões 6.15.0 e superiores da Amazon, você também pode especificar o volume raiz IOPS e a taxa de transferência. Os atributos se aplicam somente ao volume do dispositivo EBS raiz da Amazon e se aplicam a todas as instâncias no cluster. Os atributos não se aplicam a volumes de armazenamento, que você especifica separadamente para cada tipo de instância ao criar o cluster.

- O tamanho padrão do volume raiz é de 15 GiB no Amazon EMR 6.10.0 e superior. O tamanho padrão do volume raiz das versões anteriores é de 10 GiB. Você pode ajustá-lo para até 100 GiB.
- O volume raiz padrão IOPS é 3000. Você pode ajustá-las para até 16.000.
- O volume raiz padrão tem 125 MiB/s de throughput. Você pode ajustá-lo para até 1000 Mib/s.

> **Note**
>
> O tamanho do volume raiz não IOPS pode ter uma proporção maior do que 1 volume para 500 IOPS (1:500), enquanto o volume raiz IOPS e a taxa de transferência não podem ter uma proporção maior que a taxa de transferência de 1 IOPS para 0,25 (1:0,25).

Para obter mais informações sobre a AmazonEBS, consulte Volume do dispositivo EC2 raiz da Amazon.

Tipo de volume do dispositivo raiz com o padrão AMI

Quando você usa o padrãoAMI, o tipo de volume do dispositivo raiz é determinado pela EMR versão da Amazon que você usa.

- Com as EMR versões 6.15.0 e superiores da Amazon, a Amazon EMR atribui o General Purpose SSD (gp3) como o tipo de volume do dispositivo raiz.
- Com EMR versões da Amazon inferiores à 6.15.0, a Amazon EMR atribui General Purpose SSD (gp2) como o tipo de volume do dispositivo raiz.

Tipo de volume do dispositivo raiz com o personalizado AMI

Um personalizado AMI pode ter diferentes tipos de volume do dispositivo raiz. A Amazon EMR sempre usa seu tipo AMI de volume personalizado.

- Com as EMR versões 6.15.0 e superiores da Amazon, você pode configurar o tamanho do volume raiz e a taxa de transferência para seu volume personalizadoAMI, desde que esses atributos sejam aplicáveis ao tipo de volume personalizadoAMI. IOPS
- Com EMR versões da Amazon inferiores à 6.15.0, você só pode configurar o tamanho do volume raiz para o seu personalizado. AMI

Se você não configurar o tamanho do volume raiz ou a taxa de transferência ao criar seu cluster, a Amazon EMR usará os valores do personalizado, AMI se aplicável. IOPS Se você decidir configurar esses valores ao criar seu cluster, a Amazon EMR usará os valores que você especificar, desde que os valores sejam compatíveis e suportados pelo volume AMI raiz personalizado. Para obter mais informações, consulte Usando um personalizado AMI.

Definição de preços do tamanho do volume raiz do dispositivo

O custo do volume do dispositivo EBS raiz é calculado proporcionalmente por hora, com base nas EBS cobranças mensais desse tipo de volume na região em que o cluster é executado. O mesmo é verdadeiro para volumes de armazenamento. As cobranças são feitas em GB, mas você especifica o tamanho do volume raiz em GiB, portanto, convém considerar isso nas suas estimativas (1 GB é igual a 0.931323 GiB).

O SSD gp2 e o gp3 de uso geral são cobrados de forma diferente. Para estimar as cobranças associadas aos volumes do dispositivo EBS raiz em seu cluster, use as seguintes fórmulas:

SSDGP2 de uso geral

O custo do gp2 inclui somente o tamanho do EBS volume em GB.

```
($EBS size in GB/month) * 0.931323 / 30 / 24 * EMR_EBSRootVolumesizeInGiB *
 InstanceCount
```

Por exemplo, pegue um cluster que tenha um nó primário, um nó principal, e use o Amazon Linux básicoAMI, com o volume padrão de 10 GiB do dispositivo raiz. Se o EBS custo na região for de 0,10 USD USD/GB/mês, isso resultará em aproximadamente 0,00129 USD por instância por hora e 0,00258 USD por hora para o cluster (0,10 USD/GB/mês dividido por 30 dias, dividido por 24 horas, multiplicado por 10 GB, multiplicado por 2 instâncias de cluster).

SSDGP3 de uso geral

O custo do gp3 inclui o tamanho do EBS volume em GB, IOPS acima de 3000 (3000 IOPS gratuitos) e a taxa de transferência acima de 125 MB/s (125 MB/s gratuitos).

```
($EBS size in GB/month) * 0.931323 / 30 / 24 * EMR_EBSRootVolumesizeInGiB *
 InstanceCount
+
($EBS IOPS/Month)/30/24* (EMR_EBSRootVolumeIops - 3000) * InstanceCount
+
($EBS throughput/Month)/30/24* (EMR_EBSRootVolumeThroughputInMb/s - 125) *
 InstanceCount
```

Por exemplo, pegue um cluster que tenha um nó primário, um nó principal e use o Amazon Linux básicoAMI, com o tamanho padrão do volume do dispositivo raiz de 15 GiBIOPS, 4000 e 140 taxas de transferência. Se o EBS custo na região for de 0,10 USD USD/GB/mês, 0,005/ provisionado/mês acima de 3000 e 0,040 USD/MB/mês provisionado IOPS acima de 125. Isso representa aproximadamente 0,009293 USD por instância à hora e 0,018586 USD por hora para o cluster.

Especificação de configurações personalizadas do volume raiz do dispositivo

> **Note**
>
> O tamanho do volume raiz não IOPS pode ter uma proporção maior do que 1 volume para 500 IOPS (1:500), enquanto o volume raiz IOPS e a taxa de transferência não podem ter uma proporção maior que a taxa de transferência de 1 IOPS para 0,25 (1:0,25).

Console

Para especificar atributos de volume do dispositivo EBS raiz da Amazon a partir do EMR console da Amazon

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Selecione a EMR versão 6.15.0 ou superior da Amazon.
4. Em Configuração do cluster, navegue até a seção do volume EBS raiz e insira um valor para qualquer um dos atributos que você deseja configurar.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

CLI

Para especificar os atributos de volume do dispositivo EBS raiz da Amazon com o AWS CLI

- Use os parâmetros `--ebs-root-volume-size`, `--ebs-root-volume-iops` e `--ebs-root-volume-throughput` do comando create-cluster conforme mostrado no exemplo a seguir.

  > **Note**
  >
  > Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --release-label emr-6.15.0\
--ebs-root-volume-size 20 \
--ebs-root-volume-iops 3000\
--ebs-root-volume-throughput 135\
--instance-groups InstanceGroupType=MASTER,\
InstanceCount=1,InstanceType=m5.xlarge
 InstanceGroupType=CORE,InstanceCount=2,InstanceType=m5.xlarge
```

Configuração de software do cluster

Quando você seleciona uma versão de software, a Amazon EMR usa uma Amazon Machine Image (AMI) com Amazon Linux para instalar o software que você escolhe ao iniciar seu cluster, como Hadoop, Spark e Hive. EMRA Amazon fornece novos lançamentos regularmente, adicionando novos recursos, novos aplicativos e atualizações gerais. Recomendamos que você use a versão mais recente para executar seu cluster, sempre que possível. A versão mais recente é a opção padrão quando você executa um cluster a partir do console.

Para obter mais informações sobre os EMR lançamentos da Amazon e as versões de software disponíveis em cada lançamento, acesse o Guia de EMR lançamento da Amazon. Para obter mais informações sobre como editar as configurações padrão de aplicativos e software instalados em seu cluster, acesse Configuração de aplicativos no Amazon EMR Release Guide. Algumas versões dos componentes de código aberto do ecossistema Hadoop e Spark que estão incluídos nas EMR versões da Amazon têm patches e melhorias, que estão documentados no Guia de lançamento da Amazon. EMR

Além do software padrão e dos aplicativos que estão disponíveis para instalação no seu cluster, você pode usar ações de bootstrap para instalar softwares personalizados. As ações de bootstrap são scripts executados nas instâncias quando o cluster é executado, e que são executados nos novos nós adicionados ao seu cluster quando eles são criados. As ações de bootstrap também são úteis para invocar AWS CLI comandos em cada nó para copiar objetos do Amazon S3 para cada nó em seu cluster.

> **Note**
>
> As ações do Bootstrap são usadas de forma diferente na EMR versão 4.x e posterior da Amazon. Para obter mais informações sobre essas diferenças em relação às EMR AMI

versões 2.x e 3.x da Amazon, acesse Diferenças introduzidas na versão 4.x no Guia de lançamento da AmazonEMR.

Criar ações de bootstrap para instalar softwares adicionais

Você pode usar uma ação de bootstrap para instalar softwares adicionais ou personalizar a configuração de instâncias de cluster. As ações de bootstrap são scripts que são executados no cluster depois que a Amazon EMR lança a instância usando o Amazon Linux Amazon Machine Image (AMI). As ações de bootstrap são executadas antes que a Amazon EMR instale os aplicativos que você especifica ao criar o cluster e antes que os nós do cluster comecem a processar os dados. Se você adicionar nós a um cluster em execução, as ações de bootstrap também serão executadas nesses nós da mesma forma. É possível criar ações de bootstrap personalizadas e especificá-las ao criar seu cluster.

A maioria das ações de bootstrap predefinidas para as EMR AMI versões 2.x e 3.x da Amazon não são suportadas nas versões 4.x da Amazon. EMR Por exemplo, `configure-Hadoop` e não `configure-daemons` são compatíveis com a EMR versão 4.x da Amazon. Em vez disso, a EMR versão 4.x da Amazon fornece essa funcionalidade de forma nativa. Para obter mais informações sobre como migrar ações de bootstrap das EMR AMI versões 2.x e 3.x da Amazon para a versão 4.x da EMR Amazon, acesse Personalizando a configuração de clusters e aplicativos com AMI versões anteriores da Amazon EMR no Amazon Release Guide. EMR

Noções básicas sobre ações de bootstrap

Ações de bootstrap são executadas como o usuário do Hadoop por padrão. Você pode executar uma ação de bootstrap com privilégios de root usando `sudo`.

Todas as interfaces EMR de gerenciamento da Amazon oferecem suporte a ações de bootstrap. Você pode especificar até 16 ações de bootstrap por cluster fornecendo vários `bootstrap-actions` parâmetros do console, AWS CLI, ouAPI.

No EMR console da Amazon, você pode, opcionalmente, especificar uma ação de bootstrap ao criar um cluster.

Ao usar oCLI, você pode passar referências aos scripts de ação de bootstrap para a Amazon EMR adicionando o `--bootstrap-actions` parâmetro ao criar o cluster usando o `create-cluster` comando.

```
--bootstrap-actions Path="s3://mybucket/filename",Args=[arg1,arg2]
```

Se a ação de bootstrap retornar um código de erro diferente de zero, a Amazon EMR tratará isso como uma falha e encerrará a instância. Se muitas instâncias falharem em suas ações de bootstrap, a Amazon EMR encerrará o cluster. Se apenas algumas instâncias falharem, a Amazon EMR tentará realocar as instâncias com falha e continuar. Use o código de erro de cluster `lastStateChangeReason` para identificar falhas causadas por uma ação de bootstrap.

Executar uma ação de bootstrap condicionalmente

Para executar apenas ações de bootstrap no nó principal, você pode usar uma ação de bootstrap personalizada com um pouco de lógica para determinar se o nó é principal.

```
#!/bin/bash
if grep isMaster /mnt/var/lib/info/instance.json | grep false;
then
    echo "This is not master node, do nothing,exiting"
    exit 0
fi
echo "This is master, continuing to execute script"
# continue with code logic for master node below
```

A saída a seguir será impressa de um nó central.

```
This is not master node, do nothing, exiting
```

A saída a seguir será impressa de um nó principal.

```
This is master, continuing to execute script
```

Para usar essa lógica, carregue a ação de bootstrap, incluindo o código acima, no bucket do Amazon S3. No AWS CLI, adicione o `--bootstrap-actions` parâmetro à `aws emr create-cluster` API chamada e especifique a localização do script de bootstrap como o valor de`Path`.

Ações de desligamento

Um script de ação de bootstrap pode criar uma ou mais ações de desligamento, escrevendo scripts no diretório `/mnt/var/lib/instance-controller/public/shutdown-actions/`. Quando um cluster é encerrado, todos os scripts nesse diretório são executados em paralelo. Cada script deve ser executado e concluído em até 60 segundos.

Scripts de ação de desligamento não terão garantia de execução se o nó for encerrado com um erro.

> **Note**
>
> Ao usar EMR as versões 4.0 e posteriores da Amazon, você deve criar manualmente o `/mnt/var/lib/instance-controller/public/shutdown-actions/` diretório no nó principal. Ele não existe por padrão. No entanto, depois de serem criados, os scripts nesse diretório são executados antes do desligamento. Para obter mais informações sobre como conectar-se ao nó principal para criar diretórios, consulte Conecte-se ao nó primário usando SSH.

Usar ações de bootstrap personalizadas

Você pode criar um script personalizado para executar uma ação de bootstrap personalizada. Qualquer uma das EMR interfaces da Amazon pode fazer referência a uma ação de bootstrap personalizada.

> **Note**
>
> Para obter o melhor desempenho, recomendamos que você armazene ações de bootstrap, scripts e outros arquivos personalizados que você deseja usar com a Amazon EMR em um bucket do Amazon S3 que esteja no Região da AWS mesmo que seu cluster.

Conteúdo

Adicionar ações de bootstrap personalizadas

Console

Para criar um cluster com uma ação de bootstrap com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.

2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Em Ações de bootstrap, escolha Adicionar para especificar um nome, um local do script e os argumentos opcionais para a ação. Selecione Adicionar ação de bootstrap.
4. Opcionalmente, adicione mais ações de bootstrap.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

CLI

Para criar um cluster com uma ação de bootstrap personalizada com o AWS CLI

Ao usar a ação AWS CLI para incluir uma ação de bootstrap, especifique `Path` e `Args` como uma lista separada por vírgulas. O exemplo a seguir não usa uma lista de argumentos.

- Para iniciar um cluster com uma ação de bootstrap personalizada, digite o comando a seguir, substituindo *myKey* com o nome do seu EC2 key pair. Inclua `--bootstrap-actions` como parâmetro e especifique o local do script de bootstrap como o valor de `Path`.
  - Usuários UNIX de Linux e Mac OS X:

```
aws emr create-cluster --name "Test cluster" --release-label emr-4.0.0 \
--use-default-roles --ec2-attributes KeyName=myKey \
--applications Name=Hive Name=Pig \
--instance-count 3 --instance-type m5.xlarge \
--bootstrap-actions Path="s3://elasticmapreduce/bootstrap-actions/download.sh"
```

  - Usuários do Windows:

```
aws emr create-cluster --name "Test cluster" --release-label emr-4.2.0 --use-
default-roles --ec2-attributes KeyName=myKey --applications Name=Hive Name=Pig
 --instance-count 3 --instance-type m5.xlarge --bootstrap-actions Path="s3://
elasticmapreduce/bootstrap-actions/download.sh"
```

Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó primário é executado, e as instâncias restantes são executadas como nós centrais. Todos os nós usarão o tipo de instância especificado no comando.

> **Note**
>
> Se você ainda não criou a função de EMR serviço e o perfil de EC2 instância padrão da Amazon, digite `aws emr create-default-roles` para criá-los antes de digitar o `create-cluster` subcomando.

Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consultehttps://docs.aws.amazon.com/cli/latest/reference/emr.

Usar uma ação de bootstrap personalizada para copiar um objeto do Amazon S3 para cada nó

Você pode usar uma ação de bootstrap para copiar objetos do Amazon S3 para cada nó no cluster antes que suas aplicações sejam instaladas. O AWS CLI é instalado em cada nó de um cluster, para que sua ação de bootstrap possa chamar AWS CLI comandos.

O exemplo a seguir demonstra um script de ação de bootstrap simples que copia um arquivo, `myfile.jar`, do Amazon S3 para uma pasta local, `/mnt1/myfolder`, em cada nó do cluster. O script é salvo no Amazon S3 com o nome de arquivo `copymyfile.sh` com os conteúdos a seguir.

```
#!/bin/bash
aws s3 cp s3://mybucket/myfilefolder/myfile.jar /mnt1/myfolder
```

Ao iniciar o cluster, você especifica o script. O AWS CLI exemplo a seguir demonstra isso:

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 \
--use-default-roles --ec2-attributes KeyName=myKey \
--applications Name=Hive Name=Pig \
--instance-count 3 --instance-type m5.xlarge \
--bootstrap-actions Path="s3://mybucket/myscriptfolder/copymyfile.sh"
```

Configurar o hardware e as redes do cluster

Uma consideração importante ao criar um EMR cluster da Amazon é como você configura as EC2 instâncias e as opções de rede da Amazon. Este capítulo aborda as opções a seguir e vincula todos eles em conjunto com as práticas recomendadas e diretrizes.

- Tipos de nós — EC2 As instâncias da Amazon em um EMR cluster são organizadas em tipos de nós. Existem três: nós primários, nós centrais e nós de tarefa. Cada tipo de nó realiza um conjunto de funções definidas pelos aplicativos distribuídos que você instala no cluster. Durante um trabalho do Hadoop MapReduce ou do Spark, por exemplo, componentes nos nós principais e de tarefas processam dados, transferem a saída para o Amazon S3 ou HDFS fornecem metadados de status de volta ao nó primário. Com um cluster de nó único, todos os componentes são executados no nó primário. Para obter mais informações, consulte Noções básicas sobre tipos de nó: nós primários, centrais e de tarefa.
- EC2instâncias — Ao criar um cluster, você faz escolhas sobre as EC2 instâncias da Amazon nas quais cada tipo de nó será executado. O tipo de EC2 instância determina o perfil de processamento e armazenamento do nó. A escolha da EC2 instância da Amazon para seus nós é importante porque determina o perfil de desempenho dos tipos de nós individuais em seu cluster. Para obter mais informações, consulte Configurar EC2 instâncias da Amazon.
- Rede — Você pode iniciar seu EMR cluster da Amazon VPC usando uma sub-rede pública, uma sub-rede privada ou uma sub-rede compartilhada. A configuração de redes determina como clientes e serviços podem se conectar aos clusters para realizar o trabalho, como os clusters se conectam aos armazenamentos de dados e outros recursos da AWS e as opções que você tem para controlar o tráfego nessas conexões. Para obter mais informações, consulte Configurar redes.
- Agrupamento de instâncias — o conjunto de EC2 instâncias que hospedam cada tipo de nó é chamado de frota de instâncias ou grupo de instâncias uniforme. A configuração de agrupamento de instâncias é uma escolha que deve ser feita ao criar um cluster. Essa escolha determina como você poderá adicionar nós ao cluster enquanto ele estiver em execução. A configuração se aplica a todos os tipos de nó. Não é possível alterá-lo mais tarde. Para obter mais informações, consulte Criar um cluster com frotas de instâncias ou grupos de instâncias uniformes.

> **Note**
>
> A configuração de frotas de instâncias está disponível somente nas EMR versões 4.8.0 e posteriores da Amazon, excluindo 5.0.0 e 5.0.3.

Noções básicas sobre tipos de nó: nós primários, centrais e de tarefa

Use esta seção para entender como a Amazon EMR usa cada um desses tipos de nós e como base para o planejamento da capacidade do cluster.

Nó primário

O nó primário gerencia o cluster e normalmente executa os componentes primários de aplicações distribuídas. Por exemplo, o nó primário executa o YARN ResourceManager serviço para gerenciar recursos para aplicativos. Ele também executa o HDFS NameNode serviço, rastreia o status dos trabalhos enviados ao cluster e monitora a integridade dos grupos de instâncias.

Para monitorar o progresso de um cluster e interagir diretamente com os aplicativos, você pode se conectar ao nó primário SSH como usuário do Hadoop. Para obter mais informações, consulte Conecte-se ao nó primário usando SSH. Conectar-se ao nó primário que você acesse diretórios e arquivos, como os arquivos de log do Hadoop, diretamente. Para obter mais informações, consulte Exibir arquivos de log do . Você também pode visualizar interfaces de usuário que as aplicações publicam como sites em execução no nó primário. Para obter mais informações, consulte Visualize interfaces web hospedadas em EMR clusters da Amazon.

> **Note**
>
> Com o Amazon EMR 5.23.0 e versões posteriores, você pode iniciar um cluster com três nós principais para oferecer suporte à alta disponibilidade de aplicativos como YARN Resource Manager, Spark HDFS NameNode, Hive e Ganglia. O nó primário não é mais um possível ponto de falha único com esse recurso. Se um dos nós primários falhar, a Amazon EMR automaticamente passa para um nó primário em espera e substitui o nó primário com falha por um novo com as mesmas ações de configuração e bootstrap. Para obter mais informações, consulte Plan and Configure Primary Nodes.

Nós centrais

Os nós centrais são gerenciados pelo nó primário. Os nós principais executam o daemon Data Node para coordenar o armazenamento de dados como parte do Hadoop Distributed File System (). HDFS Eles também executam o daemon Task Tracker e realizam outras tarefas de computação paralelas nos dados necessários pelos aplicativos instalados. Por exemplo, um nó central executa YARN NodeManager daemons, MapReduce tarefas do Hadoop e executores do Spark.

Há apenas um grupo de instâncias principais ou uma frota de instâncias por cluster, mas pode haver vários nós em execução em várias EC2 instâncias da Amazon no grupo de instâncias ou na frota de instâncias. Com grupos de instâncias, você pode adicionar e remover EC2 instâncias da Amazon enquanto o cluster está em execução. Também é possível configurar o ajuste de escala automático para adicionar instâncias com base no valor de uma métrica. Para obter mais informações sobre

como adicionar e remover EC2 instâncias da Amazon com a configuração de grupos de instâncias, consulteUsar ajuste de escala de clusters.

Com frotas de instâncias, você pode adicionar e remover instâncias efetivamente, modificando as capacidades de destino da frota de instâncias para sob demanda e spot, conforme necessário. Para obter mais informações sobre capacidades alvo, consulte Opções de frotas de instâncias.

> **Warning**
>
> A remoção de HDFS daemons de um nó principal em execução ou o encerramento de nós principais corre o risco de perda de dados. Tenha cuidado ao configurar nós core para usar instâncias spot. Para obter mais informações, consulte Quando você deve usar instâncias spot?.

Nós de tarefa

Você pode usar nós de tarefas para aumentar a potência de realizar tarefas de computação paralela em dados, como tarefas do Hadoop e executores do MapReduce Spark. Os nós de tarefas não executam o daemon Data Node nem armazenam dados nele. HDFS Assim como nos nós principais, você pode adicionar nós de tarefas a um cluster adicionando EC2 instâncias da Amazon a um grupo de instâncias uniforme existente ou modificando as capacidades de destino de uma frota de instâncias de tarefas.

Com a configuração de grupo de instâncias uniforme, você pode ter um total de 48 grupos de instâncias de tarefa. A capacidade de adicionar grupos de instâncias dessa forma permite combinar tipos de EC2 instâncias e opções de preços da Amazon, como instâncias sob demanda e instâncias spot. Isso proporciona a flexibilidade necessária para atender aos requisitos de workload de uma maneira econômica.

Com a configuração de frota de instâncias, a capacidade de combinar tipos de instâncias e opções de compra está integrada e, portanto, há apenas uma frota de instâncias de tarefa.

Como as Instâncias Spot são frequentemente usadas para executar nós de tarefas, a Amazon EMR tem a funcionalidade padrão para agendar YARN trabalhos para que os trabalhos em execução não falhem quando os nós de tarefas executados em Instâncias Spot forem encerrados. EMRA Amazon faz isso permitindo que os processos principais do aplicativo sejam executados somente nos nós principais. O processo principal da aplicação controla os trabalhos em execução e precisa permanecer ativo durante a vida útil do trabalho.

A EMR versão 5.19.0 e posterior da Amazon usa o recurso integrado de rótulos de YARN nós para fazer isso. (As versões anteriores usavam um patch de código). As propriedades nas classificações de `capacity-scheduler` configuração `yarn-site` e são configuradas por padrão para que o planejador YARN de capacidade e o agendador justo aproveitem os rótulos dos nós. A Amazon rotula EMR automaticamente os nós principais com o `CORE` rótulo e define as propriedades para que os mestres do aplicativo sejam programados somente nos nós com o CORE rótulo. A modificação manual das propriedades relacionadas nas classificações de configuração yarn-site e capacity-scheduler, ou diretamente nos XML arquivos associados, pode interromper esse recurso ou modificar essa funcionalidade.

A partir da série de lançamento Amazon EMR 6.x, o recurso de rótulos de YARN nós está desativado por padrão. Os processos primários da aplicação podem ser executados tanto nos nós centrais como nos nós de tarefa por padrão. Você pode ativar o recurso de rótulos de YARN nós configurando as seguintes propriedades:

- `yarn.node-labels.enabled: true`
- `yarn.node-labels.am.default-node-label-expression: 'CORE'`

Começando com a série de lançamentos Amazon EMR 7.x, a Amazon EMR atribui rótulos de YARN nós às instâncias de acordo com seu tipo de mercado, como On-Demand ou Spot. Você pode habilitar rótulos de nós e restringir os processos do aplicativo a ON_ DEMAND configurando as seguintes propriedades:

```
yarn.node-labels.enabled: true
yarn.node-labels.am.default-node-label-expression: 'ON_DEMAND'
```

Se você estiver usando o Amazon EMR 7.0 ou superior, poderá restringir o processo de inscrição aos nós com o `CODE` rótulo usando a seguinte configuração:

```
yarn.node-labels.enabled: true
yarn.node-labels.am.default-node-label-expression: 'CORE'
```

Para as EMR versões 7.2 e superiores da Amazon, se seu cluster usar escalabilidade gerenciada com rótulos de nós, a Amazon EMR tentará escalar o cluster com base no processo do aplicativo e na demanda do executor de forma independente.

Por exemplo, se você usa as EMR versões 7.2 ou superiores da Amazon e restringe o processo de aplicação a `ON_DEMAND` nós, a escalabilidade gerenciada aumenta a escala `ON_DEMAND` dos nós

se a demanda do processo de aplicação aumentar. Da mesma forma, se você restringir o processo do aplicativo aos `CORE` nós, o escalonamento gerenciado aumenta os `CORE` nós se a demanda do processo do aplicativo aumentar.

Para obter informações sobre as propriedades específicas, consulte EMRConfigurações da Amazon para evitar falhas no trabalho devido ao encerramento da instância spot do nó da tarefa.

Configurar EC2 instâncias da Amazon

EC2as instâncias vêm em configurações diferentes, conhecidas como tipos de instância. Os tipos de instância têm capacidades diferentes CPU de entrada/saída e de armazenamento. Além do tipo de instância, você pode escolher diferentes opções de compra para EC2 instâncias da Amazon. Você pode especificar diferentes tipos de instâncias e opções de compra em grupos de instâncias uniformes ou frotas de instâncias. Para obter mais informações, consulte Criar um cluster com frotas de instâncias ou grupos de instâncias uniformes. Para obter orientação sobre como escolher tipos de instância e opções de compra para sua aplicação, consulte Práticas recomendadas para configuração de clusters.

> **Important**
>
> Quando você escolhe um tipo de instância usando o AWS Management Console, o número de v CPU mostrado para cada tipo de instância é o número de YARN vcores desse tipo de instância, não o número desse tipo EC2 vCPUs de instância. Para obter mais informações sobre o número de vCPUs para cada tipo de instância, consulte Tipos de EC2 instância da Amazon.

Tópicos

- Tipos de instâncias compatíveis
- Configurar redes
- Criar um cluster com frotas de instâncias ou grupos de instâncias uniformes

Tipos de instâncias compatíveis

Esta seção descreve os tipos de instância que a Amazon EMR oferece suporte, organizados por Região da AWS. Para saber mais sobre os tipos de instância, consulte EC2Instâncias da Amazon e a matriz de tipos de AMI instância do Amazon Linux.

Nem todos os tipos de instância estão disponíveis em todas as regiões. A disponibilidade da instância está sujeita à disponibilidade e à demanda na região e zona de disponibilidade especificadas. A zona de disponibilidade da instância é determinada pela sub-rede usada para iniciar o cluster.

Considerações

Considere o seguinte ao escolher os tipos de instância para seu EMR cluster da Amazon.

> **Important**
>
> Quando você escolhe um tipo de instância usando o AWS Management Console, o número de v CPU mostrado para cada tipo de instância é o número de YARN vcores desse tipo de instância, não o número desse tipo EC2 vCPUs de instância. Para obter mais informações sobre o número de vCPUs para cada tipo de instância, consulte Tipos de EC2 instância da Amazon.

- Se você criar um cluster usando um tipo de instância que não está disponível na região e na zona de disponibilidade especificadas, o cluster poderá falhar ao tentar provisionar ou pode ficar preso no estado de provisionamento. Para obter informações sobre a disponibilidade de instâncias, consulte a página de EMR preços da Amazon ou veja as Tipos de instância compatíveis com Região da AWS tabelas nesta página.
- A partir da EMR versão 5.13.0 da Amazon, todas as instâncias usam HVM virtualização e armazenamento EBS garantido para volumes raiz. Ao usar versões de EMR lançamento da Amazon anteriores à 5.13.0, algumas instâncias da geração anterior usam PVM virtualização. Para obter mais informações, consulte Tipos de AMI virtualização do Linux.
- Devido à falta de suporte de hardware e de configurações padrão que podem levar à subutilização da memória e dos núcleos, não recomendamos que você use os tipos de instância`c7a`,,,,,`c7i`, `m7i` `m7i-flex` `r7a` `r7i` `r7izi4i.12xlarge`, `i4i.24xlarge` se você executar EMR versões da Amazon inferiores a 5.36.1 e 6.10.0. Se você executar esses tipos de instância nessas versões, poderá ter um desempenho inferior e não verá os benefícios esperados dos tipos de instância mais novos, como `c7i` vs. `c6i` Para otimizar a utilização dos recursos e o desempenho com esses tipos de desempenho, você deve executar a versão 5.36.1 e superior ou a versão 6.10.0 e superior para maximizar seus recursos.
- Alguns tipos de instâncias oferecem suporte a redes avançadas. Para obter mais informações, consulte Redes avançadas no Linux.

- NVIDIAe CUDA os drivers são instalados nos tipos de GPU instância por padrão.

Tipos de instância compatíveis com Região da AWS

As tabelas a seguir listam os tipos de EC2 instância da Amazon que a Amazon EMR suporta, organizados por Região da AWS. As tabelas também listam os primeiros EMR lançamentos da Amazon nas séries 5.x, 6.x e 7.x que oferecem suporte a cada tipo de instância.

Leste dos EUA (Norte da Virgínia) – us-east-1

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g6.xlarge	emr-7.2.0
	g6.2xlarge	emr-7.2.0
	g6.4xlarge	emr-7.2.0
	g6.8xlarge	emr-7.2.0
	g6.12xlarge	emr-7.2.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g6.16xlarge	emr-7.2.0
	g6.24xlarge	emr-7.2.0
	g6.48xlarge	emr-7.2.0
	gr6.4xlarge	emr-7.2.0
	gr6.8xlarge	emr-7.2.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p5.48xlarge	emr-6.14.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7iz.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	h1.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Leste dos EUA (Ohio): us-east-2

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7a.2xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7a.4xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7a.8xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7a.12xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7a.16xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7a.24xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7a.32xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7a.48xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i.2xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.6.0, emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.6.0, emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.6.0, emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.6.0, emr-5.0.1, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7a.xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7a.2xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7a.4xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7a.8xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7a.12xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7a.16xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7a.24xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7a.32xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7a.48xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7gn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g6.xlarge	emr-7.2.0
	g6.2xlarge	emr-7.2.0
	g6.4xlarge	emr-7.2.0
	g6.8xlarge	emr-7.2.0
	g6.12xlarge	emr-7.2.0
	g6.16xlarge	emr-7.2.0
	g6.24xlarge	emr-7.2.0
	g6.48xlarge	emr-7.2.0
	gr6.4xlarge	emr-7.2.0
	gr6.8xlarge	emr-7.2.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p5.48xlarge	emr-6.14.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7a.xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7a.2xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7a.4xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7a.8xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7a.12xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7a.16xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7a.24xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7a.32xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7a.48xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7i.16xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7iz.xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7iz.2xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7iz.4xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7iz.8xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7iz.12xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7iz.16xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7iz.32xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	h1.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Oeste dos EUA (Norte da Califórnia): us-west-1

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Oeste dos EUA (Oregon): us-west-2

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g6.xlarge	emr-7.2.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g6.2xlarge	emr-7.2.0
	g6.4xlarge	emr-7.2.0
	g6.8xlarge	emr-7.2.0
	g6.12xlarge	emr-7.2.0
	g6.16xlarge	emr-7.2.0
	g6.24xlarge	emr-7.2.0
	g6.48xlarge	emr-7.2.0
	gr6.4xlarge	emr-7.2.0
	gr6.8xlarge	emr-7.2.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p5.48xlarge	emr-6.14.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7iz.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	h1.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	h1.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

AWS GovCloud (Oeste dos EUA) - -1 us-gov-west

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.48xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.48xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

AWS GovCloud (Leste dos EUA) - -1 us-gov-east

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5.24xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5d.12xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m7i.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.48xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i-flex.xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i-flex.8xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7i.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.48xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r7i.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7i.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.48xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3.4xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

África (Cidade do Cabo): af-south-1

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5n.9xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3en.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Ásia-Pacífico (Hong Kong): ap-east-1

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5d.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5d.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Ásia-Pacífico (Jacarta): ap-southeast-3

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5.4xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5.8xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5.12xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5.16xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5.24xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5d.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5d.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5d.4xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5d.8xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5d.12xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5d.16xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5d.24xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m6g.xlarge	emr-5.30.2, emr-6.1.1, emr-7.0.0
	m6g.2xlarge	emr-5.30.2, emr-6.1.1, emr-7.0.0
	m6g.4xlarge	emr-5.30.2, emr-6.1.1, emr-7.0.0
	m6g.8xlarge	emr-5.30.2, emr-6.1.1, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6g.12xlarge	emr-5.30.2, emr-6.1.1, emr-7.0.0
	m6g.16xlarge	emr-5.30.2, emr-6.1.1, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5.4xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5.9xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5.12xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5.18xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5.24xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5d.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5d.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5d.4xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5d.9xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5d.12xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5d.18xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5d.24xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5n.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5n.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5n.4xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5n.9xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5n.18xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c6g.xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6g.2xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	c6g.4xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	c6g.8xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	c6g.12xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	c6g.16xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Computação acelerada	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5.4xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5.8xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5.12xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5.16xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5.24xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5d.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5d.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5d.4xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5d.8xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5d.12xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5d.16xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5d.24xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r6g.xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6g.2xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	r6g.4xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	r6g.8xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	r6g.12xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	r6g.16xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r7i.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7i.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r7i.4xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r7i.8xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r7i.12xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r7i.16xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r7i.24xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r7i.48xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenam ento	i3.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i3.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i3.4xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i3.8xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i3.16xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i3en.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i3en.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3en.3xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i3en.6xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i3en.12xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i3en.24xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Ásia-Pacífico (Mumbai): ap-south-1

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.6.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.6.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.6.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.6.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7i.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenam ento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Ásia-Pacífico (Hyderabad): ap-south-2

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6gd.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.32xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.9xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.18xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.9xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.18xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6i.32xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5d.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.32xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.3xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.6xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Asia Pacific (Osaka): ap-northeast-3

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Otimizada para armazenamento	i3.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Ásia-Pacífico (Seul): ap-northeast-2

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Ásia-Pacífico (Singapura): ap-southeast-1

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Ásia-Pacífico (Sydney) – ap-southeast-2

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Ásia-Pacífico (Tóquio) – ap-northeast-1

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Canadá (Central): ca-central-1

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.8.2, emr-5.0.2, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i-flex.2xlarge	emr-4.8.2, emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.8.2, emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.8.2, emr-5.0.2, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7i.xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Oeste do Canadá (Calgary): ca-west-1

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5d.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5d.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5d.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5d.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5d.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5d.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5d.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6g.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6g.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6g.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6g.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6g.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6g.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6gd.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6gd.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6gd.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6gd.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6gd.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6gd.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6i.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6i.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6i.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6i.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6i.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6i.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6i.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6id.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c5.9xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c5.18xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6g.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6g.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6g.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6g.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6g.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6g.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6gn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6gn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6gn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6gn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6gn.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6gn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6i.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6i.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6i.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6i.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6i.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6i.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6i.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Otimizadas para memória	r5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6g.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6g.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6g.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6g.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6g.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6g.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6i.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6i.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6i.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6i.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6i.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6i.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6i.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6i.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6id.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3en.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i3en.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i3en.3xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i3en.6xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i3en.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i3en.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

China (Ningxia): cn-northwest-1

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

China (Pequim): cn-north-1

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Europa (Frankfurt): eu-central-1

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7a.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.48xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7iz.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	d3en.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Europa (Zurique): eu-central-2

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.32xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.9xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.18xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5d.9xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.18xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6gd.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6gd.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gd.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6gd.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6gd.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6gd.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Otimizadas para memória	r5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5d.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6gd.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6gd.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6gd.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6gd.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6gd.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6gd.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.32xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	d3.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	d3.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	d3.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3en.3xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.6xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Europa (Irlanda): eu-west-1

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7gn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7iz.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	h1.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	h1.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Europa (Londres): eu-west-2

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i.2xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.8.2, emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.8.2, emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.8.2, emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.8.2, emr-5.0.3, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7i.xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7i.4xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7i.16xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Europa (Milão): eu-south-1

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5d.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5n.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5d.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Europa (Espanha): eu-south-2

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7a.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.48xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i.48xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i-flex.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i-flex.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i-flex.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i-flex.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5.9xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.18xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.9xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.18xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7a.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7a.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7a.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7a.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7a.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7a.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7a.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7a.32xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7a.48xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7i.48xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7a.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7a.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7a.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7a.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7a.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7a.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7a.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7a.32xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7a.48xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7i.48xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3en.3xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.6xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Europa (Paris): eu-west-3

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i.24xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.9.2, emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.9.2, emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.9.2, emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.9.2, emr-5.5.3, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.48xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r7i.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7i.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.48xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Europa (Estocolmo): eu-north-1

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5d.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.48xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7a.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7a.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7a.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7a.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7a.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7a.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7a.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7a.32xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7a.48xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7i.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	p5.48xlarge	emr-6.14.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7a.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7a.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7a.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7a.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7a.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7a.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7a.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7a.32xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7a.48xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Oriente Médio (Bahrein): me-south-1

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5d.xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5.9xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5n.2xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5d.8xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Oriente Médio (UAE) - me-central-1

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6gd.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.32xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.9xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.18xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.9xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.18xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Computação acelerada	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.32xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.3xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.6xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

América do Sul (São Paulo): sa-east-1

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	EMRVersão mínima suportada da Amazon (5.x, 6.x, 7.x)
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Instâncias da geração anterior

A Amazon EMR oferece suporte a instâncias da geração anterior para suportar aplicativos otimizados para essas instâncias e que ainda não foram atualizados. Para obter mais informações sobre esses tipos de instâncias e caminhos de atualização, consulte Instâncias de gerações anteriores.

Classe de instância	Tipos de instância
General Purpose	m1.small[1] | m1.medium[1] | m1.large[1] | m1.xlarge[1] | m3.xlarge[1] | m3.2xlarge[1] | m4.large | m4.xlarge | m4.2xlarge | m4.4xlarge | m4.10xlarge | m4.16xlarge
Compute Optimized	c1.medium[1] [2] | c1.xlarge[1] | c3.xlarge[1] | c3.2xlarge[1] | c3.4xlarge[1] | c3.8xlarge[1] | c4.large | c4.xlarge | c4.2xlarge | c4.4xlarge | c4.8xlarge
Memory Optimized	m2.xlarge[1] | m2.2xlarge[1] | m2.4xlarge[1] | r3.xlarge | r3.2xlarge | r3.4xlarge | r3.8xlarge | r4.xlarge | r4.2xlarge | r4.4xlarge | r4.8xlarge | r4.16xlarge
Storage Optimized	d2.xlarge | d2.2xlarge | d2.4xlarge | d2.8xlarge | i2.xlarge | i2.2xlarge | i2.4xlarge | i2.8xlarge

[1] Usa PVM virtualização AMI com versões de EMR lançamento da Amazon anteriores à 5.13.0. Para obter mais informações, consulte Tipos de AMI virtualização do Linux.

[2] Sem suporte na versão 5.15.0.

Opções de compra de instância

Ao configurar um cluster, você escolhe uma opção de compra para EC2 instâncias da Amazon. É possível escolher instâncias sob demanda, instâncias spot ou ambas. Os preços variam com base no tipo de instância e na região. O EMR preço da Amazon é um acréscimo ao EC2 preço da Amazon (o preço dos servidores subjacentes) e ao EBS preço da Amazon (se anexar EBS volumes da Amazon). Para obter os preços atuais, consulte Amazon EMR Pricing.

Sua opção para usar grupos de instâncias ou frotas de instâncias no cluster determina como você pode alterar opções de compra de instância enquanto um cluster está em execução. Se você

escolher grupos de instâncias uniformes, só poderá especificar a opção de compra para um grupo de instâncias ao criá-lo, e o tipo de instância e a opção de compra se aplicam a todas as EC2 instâncias da Amazon em cada grupo de instâncias. Se você optar por usar frotas de instâncias, poderá alterar as opções de compra após criar a frota de instância, e poderá combinar opções de compra para preencher uma capacidade alvo especificada por você. Para obter mais informações sobre essas configurações, consulte Criar um cluster com frotas de instâncias ou grupos de instâncias uniformes.

Instâncias sob demanda

Com instâncias sob demanda, você paga pela capacidade computacional por segundo. Opcionalmente, você pode fazer com que essas instâncias sob demanda usem as opções de compra de instâncias reservadas ou dedicadas. Com instâncias reservadas, você faz um pagamento único por uma instância para reservar capacidade. As instâncias dedicadas são fisicamente isoladas no nível do hardware do host das instâncias que pertencem a outras AWS contas. Para obter mais informações sobre as opções de compra, consulte Opções de compra por instância no Guia EC2 do usuário da Amazon.

Usar instâncias reservadas

Para usar instâncias reservadas na AmazonEMR, você usa EC2 a Amazon para comprar a instância reservada e especificar os parâmetros da reserva, incluindo o escopo da reserva aplicável a uma região ou a uma zona de disponibilidade. Para obter mais informações, consulte Amazon EC2 Reserved Instances e Buying Reserved Instances no Amazon EC2 User Guide. Depois de comprar uma Instância Reservada, se todas as condições a seguir forem verdadeiras, a Amazon EMR usará a Instância Reservada quando um cluster for iniciado:

- Uma instância sob demanda é especificada na configuração do cluster que corresponde à especificação da instância reservada.
- O cluster é executado no escopo da reserva de instância (a zona de disponibilidade ou região).
- A capacidade da Instância reservada ainda está disponível

Por exemplo, digamos que você compre uma instância reservada `m5.xlarge` com a reserva de instância direcionada à região US-East. Em seguida, você inicia um EMR cluster da Amazon no Leste dos EUA que usa duas `m5.xlarge` instâncias. A primeira instância é cobrada de acordo com a taxa da Instância reservada, e a outra de acordo com a taxa Sob demanda. A capacidade da Instância reservada é usada antes que as Instâncias sob demanda sejam criadas.

Usar instâncias dedicadas

Para usar instâncias dedicadas, você compra instâncias dedicadas usando a Amazon EC2 e, em seguida, cria uma VPC com o atributo de locação dedicada. Na AmazonEMR, você então especifica que um cluster deve ser iniciado nelaVPC. Todas as instâncias sob demanda no cluster que correspondem com a especificação de instâncias dedicadas usam as instâncias dedicadas disponíveis quando o cluster é executado.

> **Note**
>
> EMRA Amazon não oferece suporte à configuração do `dedicated` atributo em instâncias individuais.

Instâncias spot

As instâncias spot na Amazon EMR oferecem uma opção para você comprar capacidade de EC2 instância da Amazon a um custo reduzido em comparação com a compra sob demanda. A desvantagem de usar instâncias spot é que as instâncias podem ser terminadas se a capacidade spot ficar indisponível para o tipo de instância que você está executando. Para obter mais informações sobre quando usar instâncias spot pode ser apropriado para seu aplicativo, consulte Quando você deve usar instâncias spot?.

Quando a Amazon EC2 tem capacidade não utilizada, ela oferece EC2 instâncias a um custo reduzido, chamado de preço spot. Esse preço flutua com base na disponibilidade e na demanda e é estabelecido por região e zona de disponibilidade. Ao escolher instâncias spot, você especifica o preço spot máximo que você está disposto a pagar por cada tipo de EC2 instância. Quando o preço spot na zona de disponibilidade do cluster estiver abaixo do preço máximo especificado para esse tipo de instância, as instâncias serão executadas. Enquanto as instâncias forem executadas, você será cobrado de acordo com o preço spot atual e não o preço spot máximo.

> **Note**
>
> As instâncias spot com duração definida (também conhecidas como blocos spot) não estarão mais disponíveis para novos clientes a partir de 1.º de julho de 2021. Aos clientes que utilizaram o recurso anteriormente, continuaremos a oferecer suporte a instâncias spot com duração definida até 31 de dezembro de 2022.

Para obter os preços atuais, consulte os preços das instâncias EC2 spot da Amazon. Para obter mais informações, consulte Instâncias spot no Guia EC2 do usuário da Amazon. Ao criar e configurar um cluster, você especifica as opções de rede que, em última análise, determinam a Zona de disponibilidade na qual seu cluster é executado. Para obter mais informações, consulte Configurar redes.

> **Tip**
>
> Você pode ver o preço spot em tempo real no console ao passar o mouse sobre a dica de ferramenta de informações ao lado da opção de compra de Spot quando criar um cluster usando as Advanced Options (Opções avançadas). Os preços de cada zona de disponibilidade na região selecionada são exibidos. Os preços mais baixos estão nas linhas de cor verde. Devido à flutuação dos preços Spot entre as Zonas de disponibilidade, selecionar a Zona de disponibilidade com o menor preço inicial pode não resultar no menor preço durante a vigência do cluster. Para obter os melhores resultados, estude o histórico de preços da Zona de disponibilidade antes de escolher. Para obter mais informações, consulte o histórico de preços de instâncias spot no Guia EC2 do usuário da Amazon.

As opções de instâncias Spot dependem de você usar grupos de instâncias uniformes ou frotas de instâncias na sua configuração de cluster.

Instâncias Spot em grupos de instâncias uniformes

Quando você usar instâncias Spot em um grupo de instâncias uniforme, todas as instâncias desse grupo devem ser instâncias Spot. Você especifica uma única sub-rede ou Zona de disponibilidade para o cluster. Para cada grupo de instâncias, você especifica uma única instância spot e um preço spot máximo. As instâncias spot desse tipo serão executadas se o preço spot na região e na zona de disponibilidade do cluster estiver abaixo do preço spot máximo. As instâncias serão encerradas se o preço spot estiver acima do preço spot máximo. Você define o preço spot máximo somente ao configurar um grupo de instâncias. Não é possível alterá-lo mais tarde. Para obter mais informações, consulte Criar um cluster com frotas de instâncias ou grupos de instâncias uniformes.

Instâncias Spot em frotas de instâncias

Quando você usa a configuração de frotas de instâncias, opções adicionais dão maior controle sobre como as instâncias Spot são executadas e encerradas. Fundamentalmente, frotas de instâncias usam um método diferente daquele de grupos de instâncias uniformes para executar instâncias. Isso funciona porque estabelecer uma capacidade alvo para instâncias Spot (e instâncias

sob demanda) e até cinco tipos de instâncias. Você também pode especificar uma capacidade ponderada para cada tipo de instância ou usar o v CPU (YARNvcores) do tipo de instância como capacidade ponderada. Essa capacidade ponderada conta para a capacidade de destino quando uma instância desse tipo é provisionada. A Amazon EMR provisiona instâncias com ambas as opções de compra até que a capacidade desejada para cada meta seja atingida. Além disso, você pode definir uma variedade de zonas de disponibilidade para EMR a Amazon escolher ao iniciar instâncias. Você também fornece opções spot adicionais para cada frota, incluindo um tempo limite de provisionamento. Para obter mais informações, consulte Configurar frotas de instâncias.

Armazenamento de instâncias

Visão geral

O armazenamento de instâncias e o armazenamento de EBS volume da Amazon são usados para HDFS dados e para buffers, caches, dados temporários e outros conteúdos temporários que alguns aplicativos podem “espalhar” para o sistema de arquivos local.

A Amazon EBS funciona de forma diferente na Amazon EMR do que com EC2 instâncias regulares da Amazon. EBSOs volumes da Amazon anexados aos EMR clusters da Amazon são efêmeros: os volumes são excluídos após o encerramento do cluster e da instância (por exemplo, ao reduzir grupos de instâncias), portanto, você não deve esperar que os dados persistam. Embora os dados sejam efêmeros, é possível que os dados de entrada HDFS possam ser replicados dependendo do número e da especialização dos nós no cluster. Quando você adiciona volumes EBS de armazenamento da Amazon, eles são montados como volumes adicionais. Eles não fazem parte do volume de inicialização. YARNestá configurado para usar todos os volumes adicionais, mas você é responsável por alocar os volumes adicionais como armazenamento local (para arquivos de log locais, por exemplo).

Considerações

Lembre-se dessas considerações adicionais ao usar a Amazon EBS com EMR clusters:

- Você não pode capturar um EBS volume da Amazon e depois restaurá-lo na AmazonEMR. Para criar configurações personalizadas reutilizáveis, use uma personalizada AMI (disponível na Amazon EMR versão 5.7.0 e posterior). Para obter mais informações, consulte Usando um personalizado AMI.
- Um volume de dispositivo EBS raiz criptografado da Amazon é suportado somente quando se usa um volume personalizadoAMI. Para obter mais informações, consulte Criação de um volume personalizado AMI com um dispositivo EBS raiz criptografado da Amazon.

- Se você aplicar tags usando a Amazon EMRAPI, essas operações serão aplicadas aos EBS volumes.
- Existe um limite de 25 volumes por instância.
- Os EBS volumes da Amazon nos nós principais não podem ser inferiores a 5 GB.
- EBSA Amazon tem um limite fixo de 2.500 EBS volumes por solicitação de inicialização de instância. Esse limite também se aplica à Amazon EMR em EC2 clusters. Recomendamos que você inicie clusters com o número total de EBS volumes dentro desse limite e, em seguida, escale manualmente o cluster ou use a escalabilidade EMR gerenciada pela Amazon, conforme necessário. Para saber mais sobre o limite de EBS volume, consulte Cotas de serviço.

EBSArmazenamento padrão da Amazon para instâncias

Para EC2 instâncias que têm EBS somente armazenamento, a Amazon EMR aloca volumes de armazenamento EBS Amazon gp2 ou gp3 para instâncias. Quando você cria um cluster com as EMR versões 5.22.0 e superiores da Amazon, a quantidade padrão de EBS armazenamento da Amazon aumenta em relação ao tamanho da instância.

Dividimos qualquer aumento de armazenamento em vários volumes. Isso aumenta o IOPS desempenho e, por sua vez, aumenta o desempenho de algumas cargas de trabalho padronizadas. Se você quiser usar uma configuração diferente de armazenamento de EBS instâncias da Amazon, você pode especificar isso ao criar um EMR cluster ou adicionar nós a um cluster existente. Você pode usar volumes EBS gp2 ou gp3 da Amazon como volumes raiz e adicionar volumes gp2 ou gp3 como volumes adicionais. Para obter mais informações, consulte Especificação de volumes adicionais EBS de armazenamento.

A tabela a seguir identifica o número padrão de volumes, tamanhos e tamanhos totais de armazenamento Amazon EBS gp2 por tipo de instância. Para obter informações sobre volumes gp2 comparados aos gp3, consulte Comparando os tipos de EBS volume gp2 e gp3 da Amazon.

Volumes e tamanho de armazenamento padrão do Amazon EBS gp2 por tipo de instância para Amazon EMR 5.22.0 e superior

Tamanho da instância	Número de volumes	Tamanho do volume (GiB)	Tamanho total (GiB)
*.large	1	32	32
*.xlarge	2	32	64

Tamanho da instância	Número de volumes	Tamanho do volume (GiB)	Tamanho total (GiB)
*.2xlarge	4	32	128
*.4xlarge	4	64	256
*.8xlarge	4	128	512
*.9xlarge	4	144	576
*.10xlarge	4	160	640
*.12xlarge	4	192	768
*.16xlarge	4	256	1024
*.18xlarge	4	288	1152
*.24xlarge	4	384	1536

Volume EBS raiz padrão da Amazon para instâncias

Com as EMR versões 6.15 e superiores da Amazon, a Amazon anexa EMR automaticamente um Amazon EBS General Purpose SSD (gp3) como dispositivo raiz para melhorar o desempenho. AMIs Nas versões anteriores, a Amazon EMR atribui o EBS General Purpose SSD (gp2) como dispositivo raiz.

	6.15 e superior	6.14 e inferior
Tipo de volume raiz padrão		
Tamanho padrão		
Padrão IOPS		
Throughput padrão		

Para obter informações sobre como personalizar o volume do dispositivo EBS raiz da Amazon, consulteEspecificação de volumes adicionais EBS de armazenamento.

Especificação de volumes adicionais EBS de armazenamento

Ao configurar tipos de instância na AmazonEMR, você pode especificar EBS volumes adicionais para adicionar capacidade além do armazenamento de instâncias (se houver) e do EBS volume padrão. A Amazon EBS fornece os seguintes tipos de volume: General Purpose (SSD), Provisioned IOPS (SSD), Throughput Optimized (HDD), Cold (HDD) e Magnetic. Eles diferem em características de performance e preço, para que você possa adaptar seu armazenamento às necessidades analíticas e comerciais das suas aplicações. Por exemplo, algumas aplicações podem precisar ser transferidas para o disco, enquanto outras podem trabalhar com segurança na memória ou usando o Amazon S3.

Você só pode anexar EBS volumes da Amazon às instâncias no momento da inicialização do cluster e ao adicionar um grupo extra de instâncias de nós de tarefas. Se uma instância em um EMR cluster da Amazon falhar, tanto a instância quanto os EBS volumes anexados da Amazon serão substituídos por novos volumes. Consequentemente, se você separar manualmente um EBS volume da Amazon, a Amazon EMR tratará isso como uma falha e substituirá o armazenamento de instâncias (se aplicável) e os armazenamentos de volume.

A Amazon EMR não permite que você modifique seu tipo de volume de gp2 para gp3 para um cluster existente. EMR Para usar o gp3 para suas cargas de trabalho, inicie um novo cluster. EMR Além disso, não recomendamos que você atualize a taxa de transferência IOPS em um cluster que esteja em uso ou que esteja sendo provisionado, porque a Amazon EMR usa a taxa de transferência e IOPS os valores que você especifica no momento da inicialização do cluster para qualquer nova instância adicionada durante a expansão do cluster. Para ter mais informações, consulte Comparando os tipos de EBS volume gp2 e gp3 da Amazon e Seleção IOPS e taxa de transferência ao migrar para o gp3.

> **Important**
>
> Para usar um volume gp3 com seu EMR cluster, você deve iniciar um novo cluster.

Comparando os tipos de EBS volume gp2 e gp3 da Amazon

Veja aqui uma comparação dos custos entre os volumes gp2 e gp3 na região Leste dos EUA (Norte da Virgínia). Para obter as informações mais atualizadas, consulte a página do produto Amazon EBS General Purpose Volumes e a página de EBSpreços da Amazon.

Tipo de volume	gp3	gp2
Tamanho do volume	1 GiB – 16 TiB	1 GiB – 16 TiB
Padrão/Linha de base IOPS	3000	IOPS3/GiB (mínimo 100IOPS) até um máximo de 16.000. IOPS Volumes menores que 1 TiB também podem aumentar para 3.000. IOPS
Volume IOPS máximo/	16.000	16.000
Throughput padrão/de referência	125 MiB/s	O limite de throughput é entre 128 MiB/s e 250 MiB/s, dependendo do tamanho do volume.
Throughput máximo/volume	1.000 MiB/s	250 MiB/s
Preço	0,08 USD/GiB por mês 3.000 gratuitos e 0,005 USD/ mês provisionado por mês acima de 3.000; 125 MiB/s IOPS gratuitos e 0,04/MiB/s provisionados IOPS por mês acima de 125 MiB/s	USD 0,10 USD/Gib por mês

Seleção IOPS e taxa de transferência ao migrar para o gp3

Ao provisionar um volume gp2, você deve descobrir o tamanho do volume para obter a proporção e a taxa de transferência. IOPS Com o gp3, não necessário provisionar um volume maior para aumentar a performance. Você pode escolher o tamanho e a performance desejados de acordo com a necessidade da aplicação. Selecionar o tamanho certo e os parâmetros de desempenho corretos (IOPS, taxa de transferência) pode proporcionar a máxima redução de custos, sem afetar o desempenho.

Aqui está uma tabela para ajudar você a selecionar as opções de configuração do gp3:

Tamanho do volume	IOPS	Throughput
1-170 GiB	3000	125 MiB/s
170-334 GiB	3000	125 MiB/s se o tipo de EC2 instância escolhido suportar 125 MiB/s ou menos, use mais de acordo com o uso, máximo de 250 MiB/s*.
334-1000 GiB	3000	125 MiB/s se o tipo de EC2 instância escolhido suportar 125 MiB/s ou menos, use mais conforme o uso, máx. 250 MiB/s *.
1000+ GiB	Combine gp2 IOPS (tamanho em GiB x 3) ou máximo determinado pelo volume gp2 IOPS atual	125 MiB/s se o tipo de EC2 instância escolhido suportar 125 MiB/s ou menos, use mais conforme o uso, máx. 250 MiB/s *.

*O Gp3 tem a capacidade de fornecer throughput de até 1000 MiB/s. Como o gp2 fornece throughput máximo de 250 MiB/s, talvez não seja necessário ultrapassar esse limite ao usar o gp3.

Configurar redes

A maioria dos clusters é iniciada em uma rede virtual usando a Amazon Virtual Private Cloud (AmazonVPC). VPCA é uma rede virtual isolada AWS que está logicamente isolada em sua AWS conta. É possível configurar aspectos como intervalos de endereços IP privados, sub-redes, tabelas de roteamento e gateways de rede. Para obter mais informações, consulte o Guia VPC do usuário da Amazon.

VPCoferece os seguintes recursos:

- Processamento de dados confidenciais

Iniciar um cluster em um VPC é semelhante a iniciar o cluster em uma rede privada com ferramentas adicionais, como tabelas de roteamento e redeACLs, para definir quem tem acesso à rede. Se você estiver processando dados confidenciais em seu cluster, talvez queira o controle de acesso adicional que a inicialização do cluster em um VPC fornece. Além disso, você pode optar por executar seus recursos em uma sub-rede privada, em que nenhum deles tem conectividade direta com a Internet.

- Acesso a recursos em uma rede interna

  Se sua fonte de dados estiver localizada em uma rede privada, pode ser impraticável ou indesejável fazer o upload desses dados AWS para importação na AmazonEMR, seja por causa da quantidade de dados a serem transferidos ou devido à natureza confidencial dos dados. Em vez disso, você pode iniciar o cluster em um VPC e conectar seu data center ao seu VPC por meio de uma VPN conexão, permitindo que o cluster acesse recursos em sua rede interna. Por exemplo, se você tiver um banco de dados Oracle em seu data center, iniciar seu cluster em um VPC conectado a essa rede VPN possibilita que o cluster acesse o banco de dados Oracle.

Sub-redes públicas e privadas

Você pode iniciar EMR clusters da Amazon em VPC sub-redes públicas e privadas. Isso significa que você não precisa de conectividade com a Internet para executar um EMR cluster da Amazon; no entanto, talvez seja necessário configurar a conversão de endereços de rede (NAT) e VPN gateways para acessar serviços ou recursos localizados fora doVPC, por exemplo, em uma intranet corporativa ou endpoints de AWS serviço público, como. AWS Key Management Service

> **Important**
>
> A Amazon EMR só oferece suporte ao lançamento de clusters em sub-redes privadas na versão 4.2 e posterior.

Para obter mais informações sobre a AmazonVPC, consulte o Guia VPC do usuário da Amazon.

Tópicos

- VPCOpções da Amazon
- Configurar um VPC para hospedar clusters
- Inicie clusters em um VPC

- Política mínima do Amazon S3 para uma sub-rede privada
- Mais recursos para aprender sobre VPCs

VPCOpções da Amazon

Ao iniciar um EMR cluster da Amazon em umVPC, você pode iniciá-lo em uma sub-rede pública, privada ou compartilhada. Existem pequenas diferenças, porém significativas, na configuração, dependendo do tipo de sub-rede escolhido para um cluster.

Sub-redes públicas

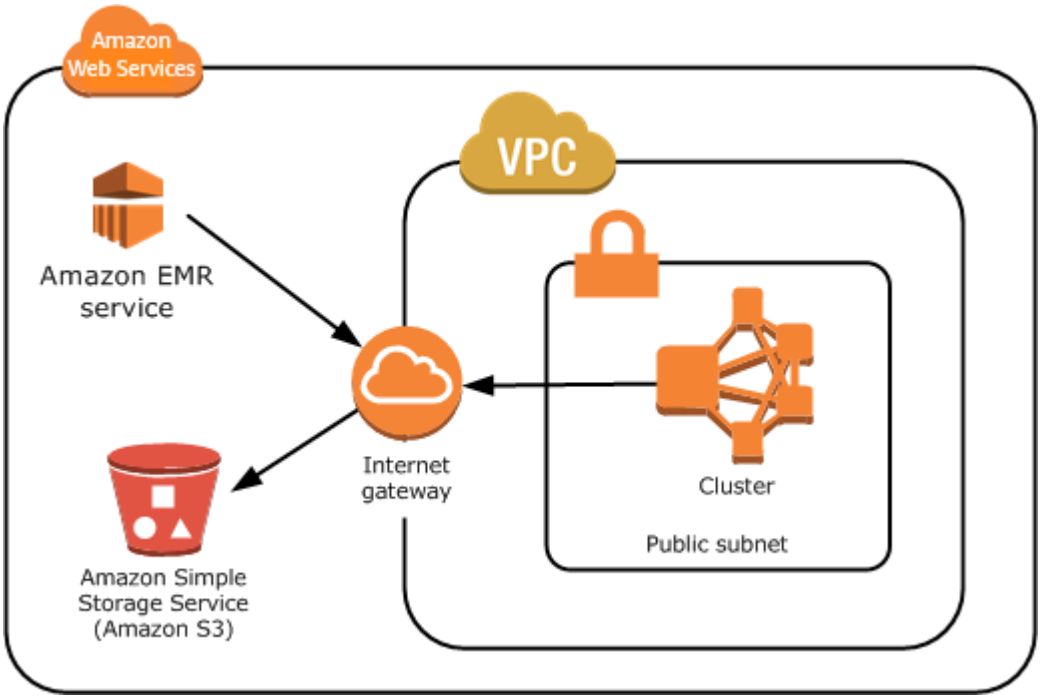
EMRclusters em uma sub-rede pública exigem um gateway de internet conectado. Isso ocorre porque os EMR clusters da Amazon devem acessar AWS os serviços e a AmazonEMR. Se um serviço, como o Amazon S3, fornecer a capacidade de criar um VPC endpoint, você poderá acessar esses serviços usando o endpoint em vez de acessar um endpoint público por meio de um gateway de internet. Além disso, a Amazon EMR não pode se comunicar com clusters em sub-redes públicas por meio de um dispositivo de tradução de endereços de rede (NAT). Um gateway de internet é necessário para essa finalidade, mas você ainda pode usar uma NAT instância ou gateway para outro tráfego em cenários mais complexos.

Todas as instâncias em um cluster se conectam ao Amazon S3 por meio de um VPC endpoint ou gateway de internet. Outros AWS serviços que atualmente não oferecem suporte a VPC endpoints usam apenas um gateway de internet.

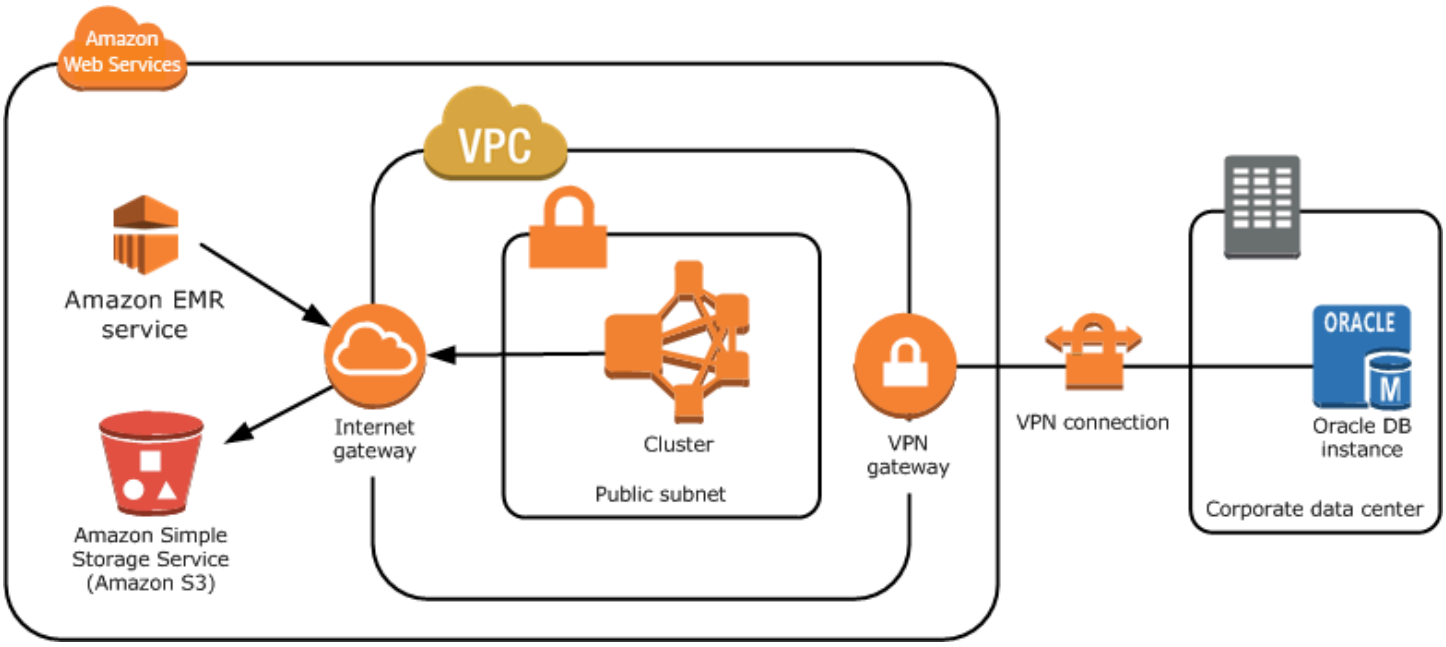
Se você tiver AWS recursos adicionais que não deseja conectar ao gateway da Internet, você pode iniciar esses componentes em uma sub-rede privada que você cria dentro da suaVPC.

Clusters em execução em uma sub-rede pública usam dois grupos de segurança: um para o nó primário e outro para os nós centrais e de tarefa. Para obter mais informações, consulte Controle do tráfego de rede com grupos de segurança.

O diagrama a seguir mostra como um EMR cluster da Amazon é executado VPC usando uma sub-rede pública. O cluster é capaz de se conectar a outros AWS recursos, como buckets do Amazon S3, por meio do gateway da Internet.



O diagrama a seguir mostra como configurar um VPC para que um cluster no VPC possa acessar recursos em sua própria rede, como um banco de dados Oracle.



Sub-redes privadas

Uma sub-rede privada permite que você inicie AWS recursos sem exigir que a sub-rede tenha um gateway de internet conectado. A Amazon EMR oferece suporte ao lançamento de clusters em sub-redes privadas com versões de lançamento 4.2.0 ou posteriores.

> **Note**
>
> Ao configurar um EMR cluster da Amazon em uma sub-rede privada, recomendamos que você também configure VPCendpoints para o Amazon S3. Se o seu EMR cluster estiver em uma sub-rede privada sem VPC endpoints para o Amazon S3, você incorrerá em cobranças NAT adicionais de gateway associadas ao tráfego do S3 porque o tráfego entre EMR seu cluster e o S3 não permanecerá dentro do seu. VPC

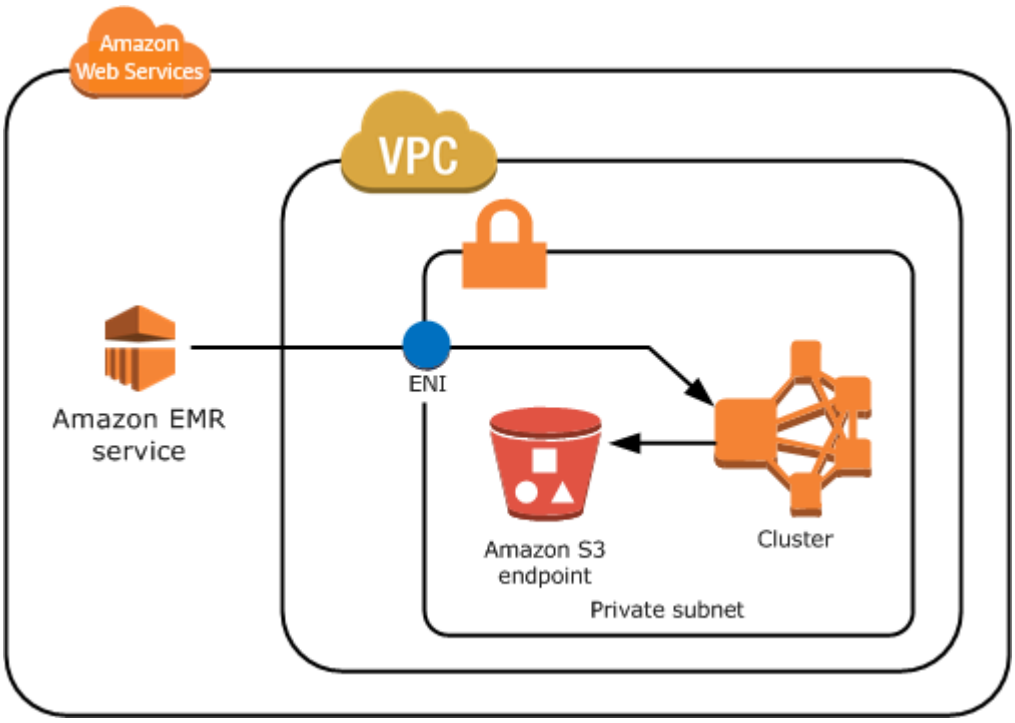
As sub-redes privadas diferem das sub-redes do seguinte modo:

- Para acessar AWS serviços que não fornecem um VPC endpoint, você ainda precisa usar uma NAT instância ou um gateway de internet.
- No mínimo, você deve fornecer uma rota para o bucket de registros de EMR serviços da Amazon e o repositório Amazon Linux no Amazon S3. Para ter mais informações, consulte Política mínima do Amazon S3 para uma sub-rede privada
- Se você usa EMRFS recursos, precisa ter um VPC endpoint do Amazon S3 e uma rota da sua sub-rede privada para o DynamoDB.
- A depuração só funciona se você fornecer uma rota da sua sub-rede privada para um endpoint público da Amazon. SQS
- A criação de uma configuração de sub-rede privada com uma NAT instância ou gateway em uma sub-rede pública só é suportada usando o. AWS Management Console A maneira mais fácil de adicionar e configurar NAT instâncias e VPC endpoints do Amazon S3 para EMR clusters da Amazon é usar a página Lista de VPC sub-redes no console da Amazon. EMR Para configurar NAT gateways, consulte NATGateways no Guia VPC do usuário da Amazon.
- Você não pode alterar uma sub-rede com um EMR cluster Amazon existente de pública para privada ou vice-versa. Para localizar um EMR cluster da Amazon em uma sub-rede privada, o cluster deve ser iniciado nessa sub-rede privada.

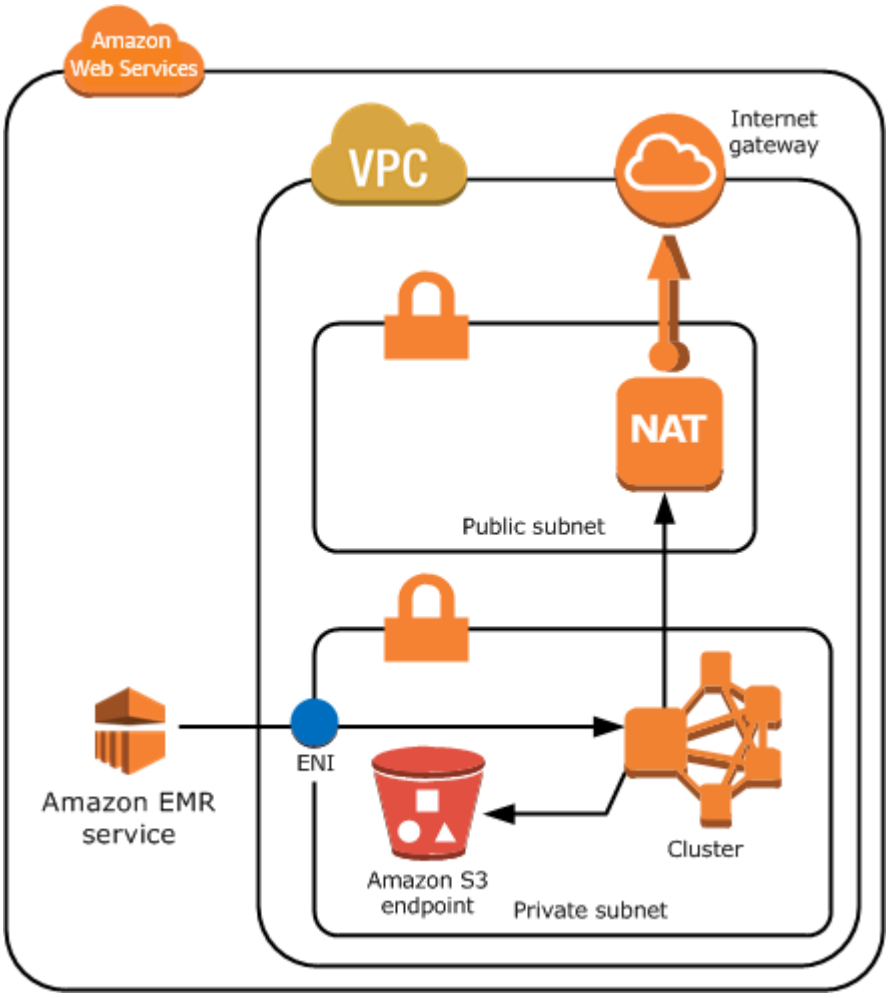
EMRA Amazon cria e usa diferentes grupos de segurança padrão para os clusters em uma sub-rede privada: ElasticMapReduce -Master-Private, ElasticMapReduce -Slave-Private e -. ElasticMapReduce ServiceAccess Para obter mais informações, consulte Controle do tráfego de rede com grupos de segurança.

Para obter uma lista completa NACLs do seu cluster, escolha Grupos de segurança para Primário e Grupos de segurança para Core & Task na página de detalhes do cluster do EMR console Amazon.

A imagem a seguir mostra como um EMR cluster da Amazon é configurado em uma sub-rede privada. A única comunicação fora da sub-rede é com a AmazonEMR.



A imagem a seguir mostra um exemplo de configuração para um EMR cluster da Amazon em uma sub-rede privada conectada a uma NAT instância que reside em uma sub-rede pública.



Sub-redes compartilhadas

VPCo compartilhamento permite que os clientes compartilhem sub-redes com outras AWS contas dentro da mesma AWS organização. Você pode iniciar EMR clusters da Amazon em sub-redes públicas compartilhadas e privadas, com as seguintes ressalvas.

O proprietário da sub-rede deve compartilhar uma sub-rede com você antes que você possa lançar um EMR cluster da Amazon nela. No entanto, as sub-redes compartilhadas podem deixar de ser compartilhadas posteriormente. Para obter mais informações, consulte Trabalhando com o Shared VPCs. Quando um cluster é lançado em uma sub-rede compartilhada e essa sub-rede compartilhada é descompartilhada, você pode observar comportamentos específicos com base no estado do EMR cluster da Amazon quando a sub-rede não é compartilhada.

- A sub-rede não é compartilhada antes que o cluster seja lançado com sucesso - Se o proprietário parar de compartilhar a Amazon VPC ou a sub-rede enquanto o participante estiver lançando um

cluster, o cluster poderá falhar ao iniciar ou ser parcialmente inicializado sem provisionar todas as instâncias solicitadas.

- A sub-rede não é compartilhada depois que o cluster é lançado com sucesso - Quando o proprietário parar de compartilhar uma sub-rede ou a Amazon VPC com o participante, os clusters do participante não poderão ser redimensionados para adicionar novas instâncias ou substituir instâncias não íntegras.

Quando você inicia um EMR cluster da Amazon, vários grupos de segurança são criados. Em uma sub-rede compartilhada, o participante da sub-rede controla esses grupos de segurança. O proprietário da sub-rede pode visualizar esses grupos de segurança, mas não pode executar nenhuma ação neles. Se o proprietário da sub-rede deseja remover ou modificar o grupo de segurança, o participante que criou o grupo de segurança deve realizar a ação.

Controle VPC as permissões com IAM

Por padrão, todos os usuários do podem ver todas as sub-redes da conta, e qualquer usuário pode executar um cluster em qualquer sub-rede.

Ao iniciar um cluster em umVPC, você pode usar AWS Identity and Access Management (IAM) para controlar o acesso aos clusters e restringir ações usando políticas, assim como faria com clusters lançados no Amazon EC2 Classic. Para obter mais informações sobreIAM, consulte o Guia IAM do usuário.

Você também pode usar IAM para controlar quem pode criar e administrar sub-redes. Por exemplo, você pode criar uma IAM função para administrar sub-redes e uma segunda função que pode iniciar clusters, mas não pode modificar as configurações da Amazon. VPC Para obter mais informações sobre como administrar políticas e ações na Amazon EC2 e na AmazonVPC, consulte IAMPolíticas para a Amazon EC2 no Guia do EC2 usuário da Amazon.

Configurar um VPC para hospedar clusters

Antes de iniciar clusters em umVPC, você deve criar uma VPC e uma sub-rede. Para sub-redes públicas, é necessário criar um gateway da Internet e anexá-lo à sub-rede. As instruções a seguir descrevem como criar um sistema VPC capaz de hospedar EMR clusters da Amazon.

Para criar um VPC com sub-redes para um cluster da Amazon EMR

1. Abra o VPC console da Amazon em https://console.aws.amazon.com/vpc/.
2. No canto superior direito da página, escolha o Região da AWSpara seuVPC.

3. Escolha Criar VPC.
4. Na página de VPCconfigurações, escolha VPCe muito mais.
5. Em Geração automática de etiquetas de nome, habilite Geração automática e insira um nome para sua. VPC Isso ajuda você a identificar a sub-rede VPC e no VPC console da Amazon depois de criá-la.
6. No campo de IPv4CIDRbloqueio, insira um espaço de endereço IP privado VPC para garantir a resolução adequada do DNS nome do host; caso contrário, você poderá enfrentar falhas no EMR cluster da Amazon. Isso inclui os seguintes intervalos de endereços IP:
   - 10.0.0.0 - 10.255.255.255
   - 172.16.0.0 - 172.31.255.255
   - 192.168.0.0 - 192.168.255.255
7. Em Número de zonas de disponibilidade (AZs), escolha o número de zonas de disponibilidade nas quais você deseja iniciar suas sub-redes.
8. Em Número de sub-redes públicas, escolha uma única sub-rede pública para adicionar à sua. VPC Se os dados usados pelo cluster estiverem disponíveis na Internet (por exemplo, no Amazon S3 ou na AmazonRDS), você só precisará usar uma sub-rede pública e não precisará adicionar uma sub-rede privada.
9. Em Número de sub-redes privadas, escolha o número de sub-redes privadas que você deseja adicionar à sua. VPC Selecione um ou mais se os dados da aplicação estiverem armazenados em sua própria rede (por exemplo, em um banco de dados Oracle). Para uma VPC sub-rede privada, todas as EC2 instâncias da Amazon devem ter, no mínimo, uma rota para a Amazon EMR por meio da interface de rede elástica. No console, isso é configurado automaticamente para você.
10. Em NATgateways, opcionalmente, escolha adicionar NAT gateways. Eles só são necessários se houver sub-redes privadas que precisam se comunicar com a Internet.
11. Em VPCendpoints, opcionalmente, escolha adicionar endpoints do Amazon S3 às suas sub-redes.
12. Verifique se Habilitar DNS nomes de host e Ativar DNS resolução estão marcadas. Para obter mais informações, consulte Usando DNS com seu VPC.
13. Escolha Criar VPC.
14. Uma janela de status mostra o trabalho em andamento. Quando o trabalho for concluído, escolha Exibir VPC para navegar até a VPCs página Sua, que exibe seu padrão VPC e o VPC

que você acabou de criar. O VPC que você criou não é padrãoVPC, portanto, a VPC coluna Padrão exibe Não.

15. Se você quiser associar seu VPC a uma DNS entrada que não inclui um nome de domínio, navegue até conjuntos de DHCP opções, escolha Criar conjunto de DHCP opções e omita um nome de domínio. Depois de criar seu conjunto de opções, navegue até o novoVPC, escolha Editar conjunto de DHCP opções no menu Ações e selecione o novo conjunto de opções. Você não pode editar o nome de domínio usando o console após a criação do conjunto de DNS opções.

    É uma prática recomendada com o Hadoop e aplicativos relacionados garantir a resolução do nome de domínio totalmente qualificado (FQDN) para os nós. Para garantir a DNS resolução adequada, configure uma VPC que inclua um conjunto de DHCP opções cujos parâmetros estejam definidos com os seguintes valores:

    - domain-name = **`ec2.internal`**

      Use **`ec2.internal`**, se a região for Leste dos EUA (Norte da Virgínia). Para outras regiões, use *`region-name`***`.compute.internal`**. Para exemplos em`us-west-2`, use**`us-west-2.compute.internal`**. Para a região AWS GovCloud (Oeste dos EUA), use**`us-gov-west-1.compute.internal`**.
    - domain-name-servers = **`AmazonProvidedDNS`**

    Para obter mais informações, consulte os conjuntos de DHCP opções no Guia VPC do usuário da Amazon.

16. Depois que o VPC for criado, vá até a página Sub-redes e anote o ID da sub-rede de uma das sub-redes da sua nova. VPC Você usa essas informações ao iniciar o EMR cluster da Amazon noVPC.

Inicie clusters em um VPC

Depois de ter uma sub-rede configurada para hospedar EMR clusters da Amazon, inicie o cluster nessa sub-rede especificando o identificador de sub-rede associado ao criar o cluster.

> **Note**
>
> A Amazon EMR oferece suporte a sub-redes privadas nas versões de lançamento 4.2 e superiores.

Quando o cluster é lançado, a Amazon EMR adiciona grupos de segurança com base no fato de o cluster estar sendo lançado em sub-redes públicas ou VPC privadas. Todos os grupos de segurança permitem a entrada na porta 8443 para se comunicar com o EMR serviço da Amazon, mas os intervalos de endereços IP variam para sub-redes públicas e privadas. EMRA Amazon gerencia todos esses grupos de segurança e pode precisar adicionar endereços IP adicionais ao AWS intervalo ao longo do tempo. Para obter mais informações, consulte Controle do tráfego de rede com grupos de segurança.

Para gerenciar o cluster em umVPC, a Amazon EMR conecta um dispositivo de rede ao nó primário e o gerencia por meio desse dispositivo. Você pode visualizar esse dispositivo usando a EC2 API ação da Amazon `DescribeInstances`. Se esse dispositivo for modificado de qualquer maneira, o cluster poderá falhar.

Console

Para iniciar um cluster em um VPC com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Em Rede, vá para o campo Nuvem privada virtual (VPC). Insira o nome do seu VPC ou escolha Procurar para selecionar seuVPC. Como alternativa, escolha Criar VPC para criar um VPC que você possa usar para seu cluster.
4. Escolha qualquer outra opção que se aplique ao cluster.
5. Para iniciar o cluster, escolha Criar cluster.

AWS CLI

Para iniciar um cluster em um VPC com o AWS CLI

> **Note**
>
> O AWS CLI não fornece uma maneira de criar uma NAT instância automaticamente e conectá-la à sua sub-rede privada. No entanto, para criar um endpoint S3 na sua sub-rede, você pode usar os comandos da Amazon. VPC CLI Use o console para criar NAT instâncias e executar clusters em uma sub-rede privada.

Depois de VPC configurado, você pode iniciar EMR clusters da Amazon nele usando o `create-cluster` subcomando com o `--ec2-attributes` parâmetro. Use o `--ec2-attributes` parâmetro para especificar a VPC sub-rede do seu cluster.

- Para criar um cluster em uma sub-rede específica, digite o seguinte comando, substitua *myKey* com o nome do seu par de EC2 chaves da Amazon e substitua *77XXXX03* com seu ID de sub-rede.

```
aws emr create-cluster --name "Test cluster" --release-label emr-4.2.0 --
applications Name=Hadoop Name=Hive Name=Pig --use-default-roles --ec2-attributes
 KeyName=myKey,SubnetId=subnet-77XXXX03 --instance-type m5.xlarge --instance-
count 3
```

  Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó primário é executado, e as instâncias restantes são executadas como nós centrais. Todos os nós usam o tipo de instância especificado no comando.

  > **Note**
  >
  > Se você ainda não criou a função de EMR serviço e o perfil de EC2 instância padrão da Amazon, digite `aws emr create-default-roles` para criá-los antes de digitar o `create-cluster` subcomando.

Garantindo endereços IP disponíveis para um EMR cluster em EC2

Para garantir que uma sub-rede com endereços IP livres suficientes esteja disponível quando você iniciar, a seleção de EC2 sub-rede verifica a disponibilidade de IP. O processo de criação usa uma sub-rede com a contagem necessária de endereços IP para iniciar os nós principais, primários e de tarefas conforme necessário, mesmo que, na criação inicial, somente os nós principais do cluster sejam criados. EMRverifica o número de endereços IP necessários para iniciar os nós primários e de tarefas durante a criação, bem como calcula separadamente o número de endereços IP necessários para iniciar os nós principais. O número mínimo de instâncias ou nós primários e de tarefas necessários é determinado automaticamente pela AmazonEMR.

> **Important**
>
> Se nenhuma sub-rede VPC tiver o suficiente disponível IPs para acomodar os nós essenciais, um erro será retornado e o cluster não será criado.

Na maioria dos casos de implantação, há uma diferença de tempo entre cada lançamento dos nós principais, primários e de tarefas. Além disso, é possível que vários clusters compartilhem uma sub-rede. Nesses casos, a disponibilidade do endereço IP pode flutuar e os lançamentos subsequentes dos nós de tarefas, por exemplo, podem ser limitados pelos endereços IP disponíveis.

Política mínima do Amazon S3 para uma sub-rede privada

Para sub-redes privadas, no mínimo, você deve fornecer à Amazon a capacidade de acessar os EMR repositórios Amazon Linux. Essa política de sub-rede privada faz parte das políticas de VPC endpoint para acessar o Amazon S3. Com o Amazon EMR 5.25.0 ou posterior, para permitir o acesso com um clique ao servidor de histórico persistente do Spark, você deve permitir que a Amazon EMR acesse o bucket do sistema que coleta os registros de eventos do Spark. Se você ativar o registro em log, forneça PUT permissões para um `aws157-logs-*` bucket. Para obter mais informações, consulte One-click access to persistent Spark History Server.

Cabe a você determinar as restrições da política que atendam às suas necessidades comerciais. Por exemplo, é possível especificar a região `packages.us-east-1.amazonaws.com` para evitar um nome ambíguo de bucket do Amazon S3. O exemplo de política a seguir fornece permissões para acessar os repositórios Amazon Linux e o bucket do EMR sistema Amazon para coletar registros de eventos do Spark. Substituir *`MyRegion`* com a região em que seus repositórios de log residem, por exemplo`us-east-1`.

Para obter mais informações sobre o uso de IAM políticas com VPC endpoints da Amazon, consulte Políticas de endpoint para o Amazon S3.

```
{
   "Version": "2008-10-17",
   "Statement": [
       {
           "Sid": "AmazonLinuxAMIRepositoryAccess",
           "Effect": "Allow",
           "Principal": "*",
           "Action": "s3:GetObject",
           "Resource": [
            "arn:aws:s3:::packages.MyRegion.amazonaws.com/*",
            "arn:aws:s3:::repo.MyRegion.amazonaws.com/*",
            "arn:aws:s3:::repo.MyRegion.emr.amazonaws.com/*"
           ]
       },
       {
           "Sid": "EnableApplicationHistory",
           "Effect": "Allow",
           "Principal": "*",
           "Action": [
               "s3:Put*",
               "s3:Get*",
               "s3:Create*",
               "s3:Abort*",
               "s3:List*"
           ],
           "Resource": [
            "arn:aws:s3:::prod.MyRegion.appinfo.src/*"
           ]
       }
   ]
}
```

O exemplo de política a seguir fornece as permissões necessárias para acessar repositórios do Amazon Linux 2. O Amazon Linux 2 AMI é o padrão.

```
{
   "Statement": [
       {
           "Sid": "AmazonLinux2AMIRepositoryAccess",
           "Effect": "Allow",
```

```
            "Principal": "*",
            "Action": "s3:GetObject",
            "Resource": [
             "arn:aws:s3:::amazonlinux.MyRegion.amazonaws.com/*",
             "arn:aws:s3:::amazonlinux-2-repos-MyRegion/*"
            ]
        }
    ]
}
```

Mais recursos para aprender sobre VPCs

Use os tópicos a seguir para saber mais sobre sub-redes VPCs e sub-redes.

- Sub-redes privadas em um VPC
  - Cenário 2: VPC com sub-redes públicas e privadas () NAT
  - NATInstâncias
  - Alta disponibilidade para VPC NAT instâncias da Amazon: um exemplo
- Sub-redes públicas em um VPC
  - Cenário 1: VPC com uma única sub-rede pública
- VPCInformações gerais
  - Guia VPC do usuário da Amazon
  - VPCEspreitando
  - Usando interfaces de rede elástica com seu VPC
  - Conecte-se com segurança a instâncias Linux em execução em um ambiente privado VPC

Criar um cluster com frotas de instâncias ou grupos de instâncias uniformes

Quando você cria um cluster e especifica a configuração do nó primário, dos nós centrais e dos nós de tarefa, existem opções de configuração. Você pode usar frotas de instâncias ou grupos de instâncias uniformes. A opção de configuração escolhida se aplica a todos os nós e pelo tempo de vida do cluster, e frotas de instâncias e grupos de instâncias não podem coexistir em um cluster. A configuração das frotas de instâncias está disponível na Amazon EMR versão 4.8.0 e posterior, excluindo as versões 5.0.x.

Você pode usar o EMR console da Amazon AWS CLI, o ou o Amazon EMR API para criar clusters com qualquer configuração. Ao usar o comando `create-cluster` a partir da AWS CLI, você usar

ambos os parâmetros `--instance-fleets` para criar o cluster usando frotas de instâncias ou, como alternativa, pode usar os parâmetros `--instance-groups` para criá-los usando grupos de instâncias uniformes.

O mesmo acontece usando a Amazon EMRAPI. Você usa também a configuração `InstanceGroups` para especificar uma matriz de objetos `InstanceGroupConfig` ou usa a configuração `InstanceFleets` para especificar uma matriz de objetos `InstanceFleetConfig`.

No novo EMR console da Amazon, você pode optar por usar grupos de instâncias ou frotas de instâncias ao criar um cluster, e você tem a opção de usar instâncias spot com cada um. Com o antigo EMR console da Amazon, se você usar as configurações padrão de opções rápidas ao criar seu cluster, a Amazon EMR aplica a configuração uniforme de grupos de instâncias ao cluster e usa instâncias sob demanda. Para instâncias spot com grupos de instâncias uniformes ou configurar frotas de instâncias e fazer outras personalizações, escolha Advanced Options (Opções avançadas).

Frotas de instâncias

A configuração de frotas de instâncias oferece a maior variedade de opções de provisionamento para instâncias da Amazon. EC2 Cada tipo de nó tem uma única frota de instâncias, e a frota de instâncias de tarefa é opcional. Você pode especificar até cinco tipos de EC2 instância por frota ou 30 tipos de EC2 instância por frota ao criar um cluster usando a Amazon AWS CLI ou a Amazon EMR API e uma estratégia de alocação para instâncias sob demanda e spot. Para as frotas de instâncias centrais e de tarefa, você atribui uma capacidade de destino para instâncias sob demanda e outra para instâncias spot. A Amazon EMR escolhe qualquer combinação dos tipos de instância especificados para atender às capacidades desejadas, provisionando tanto instâncias sob demanda quanto instâncias spot.

Para o tipo de nó primário, a Amazon EMR escolhe um único tipo de instância da sua lista de instâncias e você especifica se ele é provisionado como uma instância sob demanda ou spot.
As frotas de instâncias também oferecem outras opções para compras de instâncias spot e sob demanda. As opções de instância spot incluem um tempo limite que especifica uma ação a ser tomada, caso não seja possível provisionar a capacidade spot, e uma estratégia de alocação preferencial (otimizada para capacidade) para iniciar frotas de instâncias spot. Também é possível iniciar frotas de instâncias sob demanda usando a opção de estratégia de alocação (menor preço). Se você usar uma função de serviço que não seja a função de serviço EMR padrão ou usar uma política EMR gerenciada em sua função de serviço, precisará adicionar permissões adicionais à função de serviço de cluster personalizada para habilitar a opção de estratégia de alocação. Para obter mais informações, consulte Função de serviço para a Amazon EMR (EMRfunção).

Para obter mais informações sobre como configurar frotas de instâncias, consulte Configurar frotas de instâncias.

Grupos de instâncias uniformes

Os grupos de instâncias uniformes oferecem uma configuração mais simples do que as frotas de instâncias. Cada EMR cluster da Amazon pode incluir até 50 grupos de instâncias: um grupo de instâncias primário que contém uma EC2 instância da Amazon, um grupo de instâncias principais que contém uma ou mais EC2 instâncias e até 48 grupos de instâncias de tarefas opcionais. Cada grupo de instâncias principais e de tarefas pode conter qualquer número de EC2 instâncias da Amazon. Você pode escalar cada grupo de instâncias adicionando e removendo EC2 instâncias da Amazon manualmente ou pode configurar a escalabilidade automática. Para obter informações sobre como adicionar e remover instâncias, consulte Usar ajuste de escala de clusters.

Para obter mais informações sobre como configurar grupos de instâncias uniformes, consulte Configurar grupos de instâncias uniformes.

Trabalhar com frotas de instâncias e grupos de instâncias

Tópicos

- Configurar frotas de instâncias
- Usar reservas de capacidade com a frotas de instância
- Configurar grupos de instâncias uniformes
- Práticas recomendadas para flexibilidade de instâncias e de zona de disponibilidade
- Práticas recomendadas para configuração de clusters

Configurar frotas de instâncias

> **Note**
>
> A configuração de frotas de instâncias está disponível somente nas EMR versões 4.8.0 e posteriores da Amazon, excluindo 5.0.0 e 5.0.3.

A configuração da frota de instâncias para EMR clusters da Amazon permite que você selecione uma ampla variedade de opções de provisionamento para EC2 instâncias da Amazon e ajuda a desenvolver uma estratégia de recursos flexível e elástica para cada tipo de nó em seu cluster.

Em uma configuração de frota de instância, especifique uma capacidade de destino para instâncias sob demanda e instâncias spot em cada frota. Quando o cluster é iniciado, a Amazon EMR provisiona instâncias até que as metas sejam cumpridas. Quando a Amazon EC2 recupera uma instância spot em um cluster em execução devido a um aumento de preço ou falha na instância, a Amazon EMR tenta substituir a instância por qualquer um dos tipos de instância que você especificar. Isso facilita recuperar a capacidade durante um pico nos preços Spot.

Você pode especificar no máximo cinco tipos de EC2 instância da Amazon por frota para EMR a Amazon usar ao cumprir as metas, ou no máximo 30 tipos de EC2 instância da Amazon por frota ao criar um cluster usando a AWS CLI Amazon EMR API e uma estratégia de alocação para instâncias sob demanda e spot.

Você também pode selecionar várias sub-redes em diferentes zonas de disponibilidade. Quando a Amazon EMR lança o cluster, ela examina essas sub-redes para encontrar as instâncias e as opções de compra que você especifica. Se a Amazon EMR detectar um evento de AWS grande escala em uma ou mais zonas de disponibilidade, a Amazon EMR automaticamente tentará direcionar o tráfego para fora das zonas de disponibilidade afetadas e tentará lançar novos clusters que você cria em zonas de disponibilidade alternativas de acordo com suas seleções. A seleção da zona de disponibilidade do cluster ocorre somente na criação do cluster. Os nós de cluster já existentes não são reiniciados automaticamente em uma nova zona de disponibilidade em caso de interrupção na zona de disponibilidade.

Considerações para trabalhar com frotas de instâncias

Considere os itens a seguir ao usar frotas de instâncias com a AmazonEMR.

- Você pode ter apenas uma frota de instância por tipo de nó (primário, central, de tarefa). Você pode especificar até cinco tipos de EC2 instância da Amazon para cada frota no AWS Management Console (ou um máximo de 30 tipos por frota de instâncias ao criar um cluster usando o AWS CLI ou Amazon EMR API e umEstratégia de alocação para frotas de instâncias).
- A Amazon EMR escolhe qualquer um ou todos os tipos de EC2 instância da Amazon especificados para provisionar com opções de compra à vista e sob demanda.
- Você pode estabelecer capacidades de destino para instâncias spot e sob demanda para frota central e frota de tarefa. Use v CPU ou uma unidade genérica atribuída a cada EC2 instância da Amazon que conta para as metas. A Amazon EMR provisiona instâncias até que cada capacidade alvo seja totalmente cumprida. Para a frota primária, o destino é sempre um.
- Você pode escolher uma sub-rede (zona de disponibilidade) ou um intervalo. Se você escolher um intervalo, a Amazon EMR provisiona a capacidade na zona de disponibilidade mais adequada.

- Quando você especificar uma capacidade alvo para instâncias Spot:
  - Para cada tipo de instância, especifique um preço spot máximo. A Amazon EMR provisiona instâncias spot se o preço spot estiver abaixo do preço spot máximo. Você paga o preço spot e não necessariamente o preço spot máximo.
  - Para cada frota, defina um tempo limite para o provisionamento de instâncias Spot. Se a Amazon não EMR puder provisionar a capacidade spot, você poderá encerrar o cluster ou mudar para o provisionamento de capacidade sob demanda. Isso se aplica somente ao provisionamento de clusters, não ao redimensionamento deles. Se o período de tempo limite terminar durante o processo de redimensionamento do cluster, as solicitações spot não provisionadas serão anuladas sem serem transferidas para capacidade sob demanda.
- Para cada frota, você pode especificar uma das seguintes estratégias de alocação para suas Instâncias Spot: preço-capacidade otimizado, otimizado para capacidade capacity-optimized-prioritized, menor preço ou diversificada em todos os pools.
- Para cada frota, você pode aplicar as seguintes estratégias de alocação para suas instâncias sob demanda: a estratégia de menor preço ou a estratégia priorizada.
- Para cada frota com instâncias sob demanda, você pode optar por aplicar opções de reserva de capacidade.
- Se você usa a estratégia de alocação para frotas de exemplo, as seguintes considerações se aplicam ao escolher sub-redes para seu cluster: EMR
  - Quando a Amazon EMR provisiona um cluster com uma frota de tarefas, ela filtra as sub-redes que não têm endereços IP disponíveis suficientes para provisionar todas as instâncias do cluster solicitadoEMR. Isso inclui os endereços IP necessários para as frotas de instâncias primárias, principais e de tarefas durante a execução do cluster. A Amazon EMR então aproveita sua estratégia de alocação para determinar o pool de instâncias, com base no tipo de instância e nas sub-redes restantes com endereços IP suficientes, para iniciar o cluster.
  - Se a Amazon EMR não conseguir iniciar o cluster inteiro devido à insuficiência de endereços IP disponíveis, ela tentará identificar sub-redes com endereços IP livres suficientes para iniciar as frotas de instâncias essenciais (principais e primárias). Nesses cenários, sua frota de instâncias de tarefas entrará em um estado suspenso, em vez de encerrar o cluster com um erro.
  - Se nenhuma das sub-redes especificadas contiver endereços IP suficientes para provisionar as frotas essenciais de instância principal e primária, a inicialização do cluster falhará com um _. VALIDATION ERROR Isso aciona um evento de encerramento do cluster de CRITICALgravidade, notificando você de que o cluster não pode ser iniciado. Para evitar esse problema, recomendamos aumentar o número de endereços IP em suas sub-redes.

- Ao iniciar instâncias sob demanda, você pode usar reservas de capacidade abertas ou direcionadas para nós primários, principais e de tarefas em suas contas. Você pode ver capacidade insuficiente com instâncias sob demanda com estratégia de alocação para frotas de instâncias. Recomendamos que você especifique vários tipos de instância para diversificar e reduzir a chance de ter capacidade insuficiente. Para obter mais informações, consulte the section called “Usar reservas de capacidade com a frotas de instância”.

Opções de frotas de instâncias

Use as seguintes diretrizes para compreender as opções de frota de instância.

Tópicos

Definir capacidades de destino

Especifique as capacidades alvo que deseja para a frota de núcleo e de tarefa. Quando você faz isso, isso determina o número de instâncias sob demanda e instâncias spot que a Amazon EMR provisiona. Quando você especifica uma instância, você decide o quanto cada instância é considerada para o destino. Quando uma instância sob demanda é provisionada, ela é considerada para o destino sob demanda. O mesmo aplica-se para instâncias spot. Ao contrário de frotas centrais e de tarefa, a frota primária é sempre uma instância. Portanto, a capacidade de destino desta frota é sempre um.

Quando você usa o console, as vCPUs do tipo de EC2 instância da Amazon são usadas como a contagem das capacidades de destino por padrão. Você pode alterar isso para unidades genéricas e, em seguida, especificar a contagem para cada tipo de EC2 instância. Ao usar o AWS CLI, você atribui manualmente unidades genéricas para cada tipo de instância.

> **Important**
>
> Quando você escolhe um tipo de instância usando o AWS Management Console, o número de v CPU mostrado para cada tipo de instância é o número de YARN vcores desse tipo de instância, não o número desse tipo EC2 vCPUs de instância. Para obter mais informações

sobre o número de vCPUs para cada tipo de instância, consulte Tipos de EC2 instância da Amazon.

Para cada frota, você especifica até cinco tipos de EC2 instância da Amazon. Se você usa um Estratégia de alocação para frotas de instâncias e cria um cluster usando o AWS CLI ou o Amazon EMRAPI, você pode especificar até 30 tipos de EC2 instância por frota de instâncias. A Amazon EMR escolhe qualquer combinação desses tipos de EC2 instância para atender às capacidades desejadas. Como a Amazon EMR quer preencher completamente a capacidade alvo, um excesso pode acontecer. Por exemplo, se houver duas unidades não atendidas e a Amazon só EMR puder provisionar uma instância com uma contagem de cinco unidades, a instância ainda será provisionada, o que significa que a capacidade alvo é excedida em três unidades.

Se você reduzir a capacidade alvo para redimensionar um cluster em execução, a Amazon EMR tentará concluir as tarefas do aplicativo e encerrar as instâncias para atingir a nova meta. Para obter mais informações, consulte Terminar na conclusão de tarefas.

Opções de inicialização

Para instâncias spot, você pode especificar Preço spot máximo para cada tipo de instância da frota. Você pode definir esse preço como uma porcentagem do preço sob demanda ou como uma quantia em dólar. A Amazon EMR provisiona instâncias spot se o preço spot atual em uma zona de disponibilidade estiver abaixo do seu preço spot máximo. Você paga o preço spot e não necessariamente o preço spot máximo.

> **Note**
>
> As instâncias spot com duração definida (também conhecidas como blocos spot) não estarão mais disponíveis para novos clientes a partir de 1.º de julho de 2021. Aos clientes que utilizaram o recurso anteriormente, continuaremos a oferecer suporte a instâncias spot com duração definida até 31 de dezembro de 2022.

Disponível no Amazon EMR 5.12.1 e versões posteriores, você tem a opção de lançar frotas de instâncias spot e sob demanda com alocação de capacidade otimizada. Essa opção de estratégia de alocação pode ser definida na antiga AWS Management Console ou usando o. API `RunJobFlow` Não é possível personalizar a estratégia de alocação no novo console. Usar a opção de estratégia de alocação requer outras permissões de perfil de serviço. Se você usar a função de EMR serviço padrão da Amazon e a política gerenciada (`EMR_DefaultRole`e`AmazonEMRServicePolicy_v2`)

para o cluster, as permissões para a opção de estratégia de alocação já estão incluídas. Se você não estiver usando a função de EMR serviço padrão e a política gerenciada da Amazon, deverá adicioná-las para usar essa opção. Consulte Função de serviço para a Amazon EMR (EMRfunção).

Para obter mais informações sobre instâncias spot, consulte Instâncias spot no Guia EC2 do usuário da Amazon. Para obter mais informações sobre instâncias sob demanda, consulte Instâncias sob demanda no Guia EC2 do usuário da Amazon.

Ao escolher iniciar frotas de instâncias sob demanda com a estratégia de alocação de menor preço, você terá a opção de usar reservas de capacidade. As opções de reserva de capacidade podem ser definidas usando a Amazon EMR API`RunJobFlow`. As reservas de capacidade exigem outras permissões de perfil de serviço que você deve adicionar para usar essas opções. Consulte Permissões da estratégia de alocação. Não é possível personalizar as reservas de capacidade no novo console.

Várias opções de sub-rede (zonas de disponibilidade)

Ao usar frotas de instâncias, você pode especificar várias EC2 sub-redes da Amazon em umaVPC, cada uma correspondendo a uma zona de disponibilidade diferente. Se você usa EC2 -Classic, especifica explicitamente as zonas de disponibilidade. A Amazon EMR identifica a melhor zona de disponibilidade para iniciar instâncias de acordo com as especificações da sua frota. Instâncias são sempre provisionadas em apenas uma Zona de disponibilidade. Você pode selecionar sub-redes privadas ou públicas, mas não pode misturar as duas, e as sub-redes especificadas devem estar dentro das mesmas. VPC

Configuração do nó principal

Como a frota de instância primária é somente uma única instância, sua configuração é um pouco diferente de frotas de instâncias centrais e de tarefa. Você seleciona apenas sob demanda ou spot para a frota de instâncias primária, pois ela é formada por somente uma instância. Se você usar o console para criar a frota de instâncias, a capacidade alvo para a opção de compra que você selecionar será definida como 1. Se você usar o AWS CLI, sempre `TargetOnDemandCapacity` defina um `TargetSpotCapacity` ou como 1, conforme apropriado. Ainda é possível escolher até cinco tipos de instância para a frota de instâncias primárias (ou no máximo 30 ao usar a opção de estratégia de alocação para instâncias sob demanda ou spot). No entanto, diferentemente das frotas de instâncias principais e de tarefas, nas quais a Amazon EMR pode provisionar várias instâncias de diferentes tipos, a Amazon EMR seleciona um único tipo de instância para provisionar a frota de instâncias primárias.

Estratégia de alocação para frotas de instâncias

Com EMR as versões 5.12.1 e posteriores da Amazon, você pode usar a opção de estratégia de alocação com instâncias sob demanda e spot para cada nó do cluster. Ao criar um cluster usando o AWS CLI EMR console da Amazon EMR API ou da Amazon com uma estratégia de alocação, você pode especificar até 30 tipos de EC2 instância da Amazon por frota. Com a configuração padrão da frota de instâncias de EMR cluster da Amazon, você pode ter até 5 tipos de instância por frota. É recomendável usar a opção de estratégia de alocação para obter provisionamento mais rápido do cluster, alocação mais precisa de instâncias spot e menos interrupções de instâncias spot.

Tópicos

Estratégia de alocação para instâncias sob demanda

As seguintes estratégias de alocação estão disponíveis para suas instâncias sob demanda:

`lowest-price` (padrão)

A estratégia de alocação de menor preço lança instâncias sob demanda a partir do pool de menor preço que tem capacidade disponível. Se o pool de menor preço não tiver capacidade disponível, as instâncias sob demanda vêm do próximo pool de menor preço com capacidade disponível.

`prioritized`

A estratégia de alocação priorizada permite que você não especifique um valor prioritário para cada tipo de instância da sua frota de instâncias. A Amazon EMR lança suas instâncias sob demanda que têm a maior prioridade. Se você usar essa estratégia, deverá configurar a prioridade para pelo menos um tipo de instância. Se você não configurar o valor de prioridade para um tipo de instância, a Amazon EMR atribuirá a prioridade mais baixa a esse tipo de instância. Cada frota de instâncias (primária, principal ou tarefa) em um cluster pode ter um valor de prioridade diferente para um determinado tipo de instância.

> **Note**
>
> Se você usa a estratégia de alocação capacity-optimized-prioritizedspot, a Amazon EMR aplica as mesmas prioridades às suas instâncias sob demanda e às instâncias spot quando você define prioridades.

Estratégia de alocação com instâncias spot

Em Instâncias spot, você escolher uma destas estratégias de alocação:

`price-capacity-optimized` (recomendado)

A estratégia de alocação otimizada para preço-capacidade inicia instâncias spot com base nos grupos de instâncias spot que têm a maior capacidade disponível e o menor preço para o número de instâncias que estão sendo iniciadas. Como resultado, a estratégia otimizada de preço-capacidade normalmente tem uma chance maior de obter capacidade spot e oferece menores taxas de interrupção. Essa é a estratégia padrão para as versões 6.10.0 e superiores da AmazonEMR.

`capacity-optimized`

A estratégia de alocação otimizada para capacidade inicia instâncias spot nos grupos mais disponíveis com a menor chance de interrupção no curto prazo. Essa é uma boa opção para workloads que podem ter um custo maior de interrupção associado ao trabalho que é reiniciado. Essa é a estratégia padrão para as EMR versões 6.9.0 e inferiores da Amazon.

`capacity-optimized-prioritized`

A estratégia de capacity-optimized-prioritized alocação permite que você especifique um valor de prioridade para cada tipo de instância em sua frota de instâncias. A Amazon EMR otimiza primeiro a capacidade, mas respeita as prioridades do tipo de instância com base no melhor esforço possível, como se a prioridade não afetar significativamente a capacidade da frota de provisionar a capacidade ideal. Recomendamos essa opção se você tiver cargas de trabalho que precisam ter uma quantidade mínima de interrupção e ainda precisam de determinados tipos de instância. Se você usar essa estratégia, deverá configurar a prioridade para pelo menos um tipo de instância. Se você não configurar uma prioridade para nenhum tipo de instância, a Amazon EMR atribuirá o menor valor de prioridade a esse tipo de instância. Cada frota de instâncias (primária, principal ou tarefa) em um cluster pode ter um valor de prioridade diferente para um determinado tipo de instância.

> **Note**
>
> Se você usa a estratégia priorizada de alocação sob demanda, a Amazon EMR aplica o mesmo valor de prioridade às suas instâncias sob demanda e spot quando você define prioridades.

`diversified`

Com a estratégia de alocação diversificada, a Amazon EC2 distribui instâncias spot em todos os pools de capacidade spot.

`lowest-price`

A estratégia de alocação de menor preço inicia instâncias spot pelo grupo de menor preço que tenha capacidade disponível. Se o grupo com menor preço não tiver capacidade disponível, as instâncias spot virão do próximo grupo com menor preço que tiver capacidade disponível. Se um pool ficar sem capacidade antes de atender à capacidade solicitada, a EC2 frota da Amazon usa o próximo pool de menor preço para continuar atendendo à sua solicitação. Para garantir que a capacidade desejada seja atendida, é possível receber instâncias spot de vários grupos. Como essa estratégia considera apenas o preço da instância e não considera a disponibilidade de capacidade, ela pode resultar em altas taxas de interrupção.

Permissões da estratégia de alocação

A opção de estratégia de alocação requer várias IAM permissões que são incluídas automaticamente na função de EMR serviço padrão da Amazon e na política EMR gerenciada pela Amazon (`EMR_DefaultRoleeAmazonEMRServicePolicy_v2`). Ao usar um perfil de serviço personalizado ou uma política gerenciada para o cluster, você deverá adicionar essas permissões antes de criar o cluster. Para obter mais informações, consulte Permissões da estratégia de alocação.

As reservas opcionais de capacidade sob demanda (ODCRs) estão disponíveis quando você usa a opção de estratégia de alocação sob demanda. As opções de reserva de capacidade permitem que você especifique uma preferência para usar primeiro a capacidade reservada para EMR clusters da Amazon. Você pode usar isso para garantir que suas cargas de trabalho críticas usem a capacidade que você já reservou usando aberta ou direcionadaODCRs. Para workloads não essenciais, as preferências de reserva de capacidade permitem especificar se a capacidade reservada deverá ser consumida.

As reservas de capacidade só podem ser usadas por instâncias que correspondam a seus atributos (tipo de instância, plataforma e zona de disponibilidade). Por padrão, as reservas de capacidade aberta são usadas automaticamente pela Amazon EMR ao provisionar instâncias sob demanda que correspondam aos atributos da instância. Se você não tiver nenhuma instância em execução que corresponda aos atributos das reservas de capacidade, elas permanecerão não utilizadas até você iniciar uma instância com atributos correspondentes. Se você não quiser usar nenhuma reserva de capacidade ao iniciar o cluster, defina a preferência de reserva de capacidade como nenhuma nas opções de inicialização.

No entanto, também é possível destinar uma reserva de capacidade para workloads específicas. Isso permite que você controle explicitamente quais instâncias têm permissão para executar na capacidade reservada. Para obter mais informações sobre reservas de capacidade sob demanda, consulte Usar reservas de capacidade com a frotas de instância.

IAMPermissões necessárias para uma estratégia de alocação

O Função de serviço para a Amazon EMR (EMRfunção) precisa de outras permissões para criar um cluster que use a opção de estratégia de alocação para frotas de instâncias sob demanda ou spot.

Incluímos automaticamente essas permissões na função de EMR serviço padrão da Amazon `EMR_DefaultRole`e na política EMR gerenciada da Amazon `AmazonEMRServicePolicy_v2`.

Ao usar um perfil de serviço personalizado ou uma política gerenciada para o cluster, você deverá adicionar as seguintes permissões:

```
{
 "Version": "2012-10-17",
 "Statement": [
  {
   "Effect": "Allow",
   "Action": [
    "ec2:DeleteLaunchTemplate",
    "ec2:CreateLaunchTemplate",
    "ec2:DescribeLaunchTemplates",
    "ec2:CreateLaunchTemplateVersion",
    "ec2:CreateFleet"
   ],
   "Resource": "*"
  }
}
```

As permissões de perfil de serviço a seguir são necessárias para criar um cluster que usa reservas de capacidade abertas ou direcionadas. É necessário incluir essas permissões além das permissões necessárias para usar a opção de estratégia de alocação.

Example Documento de política para reservas de capacidade de perfil de serviço

Para usar reservas de capacidade aberta, é necessário incluir as permissões adicionais a seguir.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeCapacityReservations",
                "ec2:DescribeLaunchTemplateVersions",
                "ec2:DeleteLaunchTemplateVersions"
            ],
            "Resource": "*"
        }
    ]
}
```

Example

Para usar reservas de capacidade direcionada, é necessário incluir as permissões adicionais a seguir.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeCapacityReservations",
                "ec2:DescribeLaunchTemplateVersions",
                "ec2:DeleteLaunchTemplateVersions",
                "resource-groups:ListGroupResources"
            ],
            "Resource": "*"
        }
    ]
```

```
}
```

Configurar frotas de instâncias para o cluster

Console

Para criar um cluster com frotas de instâncias com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EMREC2Em Ativado, no painel de navegação esquerdo, escolha Clusters e escolha Criar cluster.
3. Em Configuração do cluster, escolha Frotas de instâncias.
4. Em cada Grupo de nós, selecione Adicionar tipo de instância e escolha até cinco tipos de instância para frotas de instâncias primárias e centrais e até quinze tipos de instância para frotas de instâncias de tarefa. A Amazon EMR pode provisionar qualquer combinação desses tipos de instância ao iniciar o cluster.
5. Para alterar essas configurações, em cada tipo de grupo de nós, escolha o menu suspenso Ações ao lado de cada instância:

   Adicionar EBS volumes

   Especifique EBS os volumes a serem anexados ao tipo de instância depois que a Amazon a EMR provisionar.

   Editar capacidade ponderada

   Para o grupo de nós centrais, altere esse valor para qualquer número de unidades adequado a suas aplicações. O número de YARN vCores para cada tipo de instância de frota é usado como unidades de capacidade ponderada padrão. Não é possível editar a capacidade ponderada do nó primário.

   Editar preço máximo spot

   Especifique um preço spot máximo para cada tipo de instância da frota. Você pode definir esse preço como uma porcentagem do preço sob demanda ou como uma quantia em dólar. Se o preço spot atual em uma zona de disponibilidade estiver abaixo do seu preço spot máximo, a Amazon EMR provisiona instâncias spot. Você paga o preço spot e não necessariamente o preço spot máximo.

6. Opcionalmente, para adicionar grupos de segurança aos seus nós, expanda os grupos de EC2 segurança (firewall) na seção Rede e selecione seu grupo de segurança para cada tipo de nó.
7. Opcionalmente, marque a caixa de seleção ao lado de Aplicar estratégia de alocação, se quiser usar a opção de estratégia de alocação, e selecione a estratégia de alocação que deseja especificar para as instâncias spot. Você não deve selecionar essa opção se sua função EMR de serviço da Amazon não tiver as permissões necessárias. Para obter mais informações, consulte Estratégia de alocação para frotas de instâncias.
8. Escolha qualquer outra opção que se aplique ao cluster.
9. Para iniciar o cluster, escolha Criar cluster.

AWS CLI

Para criar e executar um cluster com frotas de instâncias com o AWS CLI, siga estas diretrizes:

- Para criar e executar um cluster com frotas de instâncias, use o comando `create-cluster` com parâmetros `--instance-fleet`.
- Para obter detalhes sobre a configuração das frotas de instâncias em um cluster, use o comando `list-instance-fleets`.
- Para adicionar vários Amazon Linux personalizados AMIs a um cluster que você está criando, use a `CustomAmiId` opção com cada `InstanceType` especificação. Você pode configurar nós de frota de instâncias com vários tipos de instância e vários personalizados AMIs para atender às suas necessidades. Consulte Exemplos: criar um cluster com a configuração de frotas de instâncias.
- Para fazer alterações na capacidade alvo para uma frota de instâncias, use o comando `modify-instance-fleet`.
- Para adicionar uma frota de instância de tarefas a um cluster que ainda não tem uma, use o comando `add-instance-fleet`.
- Vários itens personalizados AMIs podem ser adicionados à frota de instâncias de tarefas usando o CustomAmiId argumento com o add-instance-fleet comando. Consulte Exemplos: criar um cluster com a configuração de frotas de instâncias.
- Para usar a opção de estratégia de alocação ao criar uma frota de instâncias, atualize o perfil de serviço de modo a incluir o exemplo de documento de política na seção a seguir.

- Para usar as opções de reservas de capacidade ao criar uma frota de instâncias com a estratégia de alocação sob demanda, atualize o perfil de serviço de modo a incluir o exemplo de documento de política na seção a seguir.
- As frotas de instâncias são incluídas automaticamente na função de EMR serviço padrão e na política EMR gerenciada da Amazon (`EMR_DefaultRoleeAmazonEMRServicePolicy_v2`). Ao usar um perfil de serviço personalizada ou uma política gerenciada personalizada para o cluster, você deverá adicionar as novas permissões para a estratégia de alocação na seção a seguir.

Exemplos: criar um cluster com a configuração de frotas de instâncias

Os exemplos a seguir demonstram comandos `create-cluster` com uma variedade de opções que você pode combinar.

> **Note**
>
> Se você ainda não criou a função de EMR serviço e o perfil de EC2 instância padrão da Amazon, use-os `aws emr create-default-roles` para criá-los antes de usar o `create-cluster` comando.

Example Exemplo: primário sob demanda, núcleo sob demanda com tipo de instância única, padrão VPC

```
aws emr create-cluster --release-label emr-5.3.1 --service-role EMR_DefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets \

 InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarg
 \

 InstanceFleetType=CORE,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarge}
```

Example Exemplo: Spot primário, Spot core com tipo de instância única, padrão VPC

```
aws emr create-cluster --release-label emr-5.3.1 --service-role EMR_DefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets \
    InstanceFleetType=MASTER,TargetSpotCapacity=1,\
```

```
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5}'] \
    InstanceFleetType=CORE,TargetSpotCapacity=1,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5}']
```

Example Exemplo: núcleo primário sob demanda, misto com tipo de instância única, sub-rede única EC2

```
aws emr create-cluster --release-label emr-5.3.1 --service-role EMR_DefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetIds=['subnet-ab12345c'] \
  --instance-fleets \
    InstanceFleetType=MASTER,TargetOnDemandCapacity=1,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge}'] \
    InstanceFleetType=CORE,TargetOnDemandCapacity=2,TargetSpotCapacity=6,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5,WeightedCapacity=2}']
```

Example Exemplo: primário sob demanda, núcleo spot com vários tipos de instância ponderada, tempo limite para spot, intervalo de sub-redes EC2

```
aws emr create-cluster --release-label emr-5.3.1 --service-role EMR_DefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetIds=['subnet-
ab12345c','subnet-de67890f'] \
  --instance-fleets \
    InstanceFleetType=MASTER,TargetOnDemandCapacity=1,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge}'] \
    InstanceFleetType=CORE,TargetSpotCapacity=11,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5,WeightedCapacity=3}',\
'{InstanceType=m4.2xlarge,BidPrice=0.9,WeightedCapacity=5}'],\
LaunchSpecifications={SpotSpecification='{TimeoutDurationMinutes=120,TimeoutAction=SWITCH_TO_ON
```

Example Exemplo: principal sob demanda, núcleo misto e tarefa com vários tipos de instância ponderada, tempo limite para instâncias spot principais, variedade de sub-redes EC2

```
aws emr create-cluster --release-label emr-5.3.1 --service-role EMR_DefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetIds=['subnet-
ab12345c','subnet-de67890f'] \
  --instance-fleets \

 InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarg
 \
    InstanceFleetType=CORE,TargetOnDemandCapacity=8,TargetSpotCapacity=6,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5,WeightedCapacity=3}',\
'{InstanceType=m4.2xlarge,BidPrice=0.9,WeightedCapacity=5}'],\
```

```
LaunchSpecifications={SpotSpecification='{TimeoutDurationMinutes=120,TimeoutAction=SWITCH_TO_ON
 \
    InstanceFleetType=TASK,TargetOnDemandCapacity=3,TargetSpotCapacity=3,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5,WeightedCapacity=3}']
```

Example Exemplo: Spot primário, sem núcleo ou tarefa, EBS configuração da Amazon, padrão VPC

```
aws emr create-cluster --release-label Amazon EMR 5.3.1 --service-role EMR_DefaultRole
 \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets \
    InstanceFleetType=MASTER,TargetSpotCapacity=1,\
LaunchSpecifications={SpotSpecification='{TimeoutDurationMinutes=60,TimeoutAction=TERMINATE_CLU
\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5,\
EbsConfiguration={EbsOptimized=true,EbsBlockDeviceConfigs=[{VolumeSpecification={VolumeType=gp2
\
SizeIn GB=100}},{VolumeSpecification={VolumeType=io1,SizeInGB=100,Iop
 s=100},VolumesPerInstance=4}]}}']
```

Example Exemplo: vários tipos de instância personalizadosAMIs, vários tipos de instância, primário sob demanda, núcleo sob demanda

```
aws emr create-cluster --release-label Amazon EMR 5.3.1 --service-role EMR_DefaultRole
 \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets \
    InstanceFleetType=MASTER,TargetOnDemandCapacity=1,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,CustomAmiId=ami-123456},
{InstanceType=m6g.xlarge, CustomAmiId=ami-234567}'] \
    InstanceFleetType=CORE,TargetOnDemandCapacity=1,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,CustomAmiId=ami-123456},
{InstanceType=m6g.xlarge, CustomAmiId=ami-234567}']
```

Example Exemplo: adicionar um nó de tarefa a um cluster em execução com vários tipos de instância e vários tipos personalizados AMIs

```
aws emr add-instance-fleet --cluster-id j-123456 --release-label Amazon EMR 5.3.1 \
  --service-role EMR_DefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleet \
    InstanceFleetType=Task,TargetSpotCapacity=1,\
```

```
InstanceTypeConfigs=['{InstanceType=m5.xlarge,CustomAmiId=ami-123456}',\
'{InstanceType=m6g.xlarge,CustomAmiId=ami-234567}']
```

Example Exemplo: usar um arquivo JSON de configuração

Você pode configurar os parâmetros da frota de instâncias em um JSON arquivo e, em seguida, referenciar o JSON arquivo como o único parâmetro para frotas de instâncias. Por exemplo, o comando a seguir faz referência a um arquivo de JSON configuração*my-fleet-config.json*:

```
aws emr create-cluster --release-label emr-5.30.0 --service-role EMR_DefaultRole \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole \
--instance-fleets file://my-fleet-config.json
```

A ferramenta *my-fleet-config.json* file especifica frotas de instâncias primárias, principais e de tarefas, conforme mostrado no exemplo a seguir. A frota de instâncias principais usa um preço spot máximo (`BidPrice`) como uma porcentagem do sob demanda, enquanto as frotas de tarefas e instâncias primárias usam um preço spot máximo (BidPriceAsPercentageofOnDemandPrice) como sequência de caracteres. USD

```
[
    {
        "Name": "Masterfleet",
        "InstanceFleetType": "MASTER",
        "TargetSpotCapacity": 1,
        "LaunchSpecifications": {
            "SpotSpecification": {
                "TimeoutDurationMinutes": 120,
                "TimeoutAction": "SWITCH_TO_ON_DEMAND"
            }
        },
        "InstanceTypeConfigs": [
            {
                "InstanceType": "m5.xlarge",
                "BidPrice": "0.89"
            }
        ]
    },
    {
        "Name": "Corefleet",
        "InstanceFleetType": "CORE",
        "TargetSpotCapacity": 1,
        "TargetOnDemandCapacity": 1,
```

```
        "LaunchSpecifications": {
          "OnDemandSpecification": {
            "AllocationStrategy": "lowest-price",
            "CapacityReservationOptions":
            {
                "UsageStrategy": "use-capacity-reservations-first",
                "CapacityReservationResourceGroupArn": "String"
            }
        },
            "SpotSpecification": {
                "AllocationStrategy": "capacity-optimized",
                "TimeoutDurationMinutes": 120,
                "TimeoutAction": "TERMINATE_CLUSTER"
            }
        },
        "InstanceTypeConfigs": [
            {
                "InstanceType": "m5.xlarge",
                "BidPriceAsPercentageOfOnDemandPrice": 100
            }
        ]
    },
    {
        "Name": "Taskfleet",
        "InstanceFleetType": "TASK",
        "TargetSpotCapacity": 1,
        "LaunchSpecifications": {
          "OnDemandSpecification": {
            "AllocationStrategy": "lowest-price",
            "CapacityReservationOptions":
            {
                "CapacityReservationPreference": "none"
            }
        },
            "SpotSpecification": {
                "TimeoutDurationMinutes": 120,
                "TimeoutAction": "TERMINATE_CLUSTER"
            }
        },
        "InstanceTypeConfigs": [
            {
                "InstanceType": "m5.xlarge",
                "BidPrice": "0.89"
            }
```

```
        ]
    }
]
```

Modificar capacidades de destino para uma frota de instâncias

Use o comando `modify-instance-fleet` para especificar novas capacidades alvo para uma frota de instâncias. Você deve especificar o ID de cluster e o ID de frota de instância. Use o `list-instance-fleets` comando para recuperar a frota IDs de instâncias.

```
aws emr modify-instance-fleet --cluster-id <cluster-id> \
  --instance-fleet \
    InstanceFleetId='<instance-fleet-id>',TargetOnDemandCapacity=1,TargetSpotCapacity=1
```

Adicionar uma frota de instâncias de tarefa a um cluster

Se um cluster tiver apenas frotas de instâncias primárias e centrais, você poderá usar o comando `add-instance-fleet` para adicionar uma frota de instâncias de tarefa. Isso só pode ser usado para adicionar frotas de instância de tarefa.

```
aws emr add-instance-fleet --cluster-id <cluster-id>
  --instance-fleet \
    InstanceFleetType=TASK,TargetSpotCapacity=1,\
LaunchSpecifications={SpotSpecification='{TimeoutDurationMinutes=20,TimeoutAction=TERMINATE_CLU
\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5}']
```

Obter detalhes da configuração de frotas de instâncias em um cluster

Use o comando `list-instance-fleets` para obter detalhes de configuração das frotas de instâncias em um cluster. O comando utiliza um ID de cluster como entrada. O exemplo a seguir demonstra o comando e sua saída para um cluster que contém um grupo de instâncias de tarefa primárias e um grupo de instâncias de tarefa centrais. Para obter a sintaxe completa da resposta, consulte ListInstanceFleetsna Amazon EMR API Reference.

```
list-instance-fleets --cluster-id <cluster-id>
```

```
{
    "InstanceFleets": [
        {
```

```
            "Status": {
                "Timeline": {
                    "ReadyDateTime": 1488759094.637,
                    "CreationDateTime": 1488758719.817
                },
                "State": "RUNNING",
                "StateChangeReason": {
                    "Message": ""
                }
            },
            "ProvisionedSpotCapacity": 6,
            "Name": "CORE",
            "InstanceFleetType": "CORE",
            "LaunchSpecifications": {
                "SpotSpecification": {
                    "TimeoutDurationMinutes": 60,
                    "TimeoutAction": "TERMINATE_CLUSTER"
                }
            },
            "ProvisionedOnDemandCapacity": 2,
            "InstanceTypeSpecifications": [
                {
                    "BidPrice": "0.5",
                    "InstanceType": "m5.xlarge",
                    "WeightedCapacity": 2
                }
            ],
            "Id": "if-1ABC2DEFGHIJ3"
        },
        {
            "Status": {
                "Timeline": {
                    "ReadyDateTime": 1488759058.598,
                    "CreationDateTime": 1488758719.811
                },
                "State": "RUNNING",
                "StateChangeReason": {
                    "Message": ""
                }
            },
            "ProvisionedSpotCapacity": 0,
            "Name": "MASTER",
            "InstanceFleetType": "MASTER",
            "ProvisionedOnDemandCapacity": 1,
```

```
            "InstanceTypeSpecifications": [
                {
                    "BidPriceAsPercentageOfOnDemandPrice": 100.0,
                    "InstanceType": "m5.xlarge",
                    "WeightedCapacity": 1
                }
            ],
           "Id": "if-2ABC4DEFGHIJ4"
        }
    ]
}
```

Usar reservas de capacidade com a frotas de instância

Para iniciar frotas de instâncias sob demanda com opções de reserva de capacidade, anexe outras permissões de perfil de serviço que são necessárias para usar as opções de reserva de capacidade. Como as opções de reserva de capacidade devem ser usadas junto com a estratégia de alocação sob demanda, também é necessário incluir as permissões necessárias para a estratégia de alocação em no perfil de serviço e na política gerenciada. Para obter mais informações, consulte Permissões da estratégia de alocação.

A Amazon EMR oferece suporte para reservas de capacidade abertas e direcionadas. Os tópicos a seguir mostram as configurações de frotas de instâncias que você pode usar com a ação `RunJobFlow` ou o comando `create-cluster` para iniciar frotas de instâncias usando reservas de capacidade sob demanda.

Usar reservas de capacidade aberta com base no melhor esforço

Se as instâncias sob demanda do cluster corresponderem aos atributos das reservas de capacidade aberta (tipo de instância, plataforma, locação e zona de disponibilidade) disponíveis na conta, as reservas de capacidade serão aplicadas automaticamente. No entanto, o uso das reservas de capacidade não é garantido. Para provisionar o cluster, a Amazon EMR avalia todos os pools de instâncias especificados na solicitação de execução e usa aquele com o menor preço que tenha capacidade suficiente para iniciar todos os nós principais solicitados. As reservas de capacidade aberta disponíveis que correspondem ao grupo de instâncias são aplicadas automaticamente. Se as reservas de capacidade aberta disponíveis não corresponderem ao grupos de instâncias, elas permanecerão inutilizadas.

Depois que os nós centrais são provisionados, a zona de disponibilidade é selecionada e corrigida. A Amazon EMR provisiona os nós de tarefas em grupos de instâncias, começando primeiro com os de menor preço, na zona de disponibilidade selecionada, até que todos os nós de tarefas sejam provisionados. As reservas de capacidade aberta disponíveis que correspondem aos grupos de instâncias são aplicadas automaticamente.

A seguir estão os casos de uso da lógica de alocação de EMR capacidade da Amazon para usar reservas de capacidade abertas com base no melhor esforço.

Exemplo 1: o grupo de instâncias de menor preço na solicitação de inicialização tem reservas de capacidade abertas disponíveis

Nesse caso, a Amazon EMR lança capacidade no pool de instâncias de menor preço com instâncias sob demanda. Suas reservas de capacidade aberta disponíveis nesse grupo de instâncias são usadas automaticamente.

Estratégia sob demanda	preço mais baixo		
Capacidade solicitada	100		
Tipo de instância	c5.xlarge	m5.xlarge	r5.xlarge
Reservas de capacidade aberta disponíveis	150	100	100
Preço sob demanda	$	$$	$$$
Instâncias provisionadas	100	-	-
Reserva de capacidade aberta utilizada	100	-	-
Reservas de capacidade aberta disponíveis	50	100	100

Depois que a frota de instâncias for iniciada, você poderá executar `describe-capacity-reservations` para ver quantas reservas de capacidade não utilizadas restam.

Exemplo 2: o grupo de instâncias de menor preço na solicitação de execução não tem reservas de capacidade abertas disponíveis

Nesse caso, a Amazon EMR lança capacidade no pool de instâncias de menor preço com instâncias sob demanda. No entanto, suas reservas de capacidade aberta permanecem inutilizadas.

Estratégia sob demanda	preço mais baixo		
Capacidade solicitada	100		
Tipo de instância	c5.xlarge	m5.xlarge	r5.xlarge
Reservas de capacidade aberta disponíveis	-	-	100
Preço sob demanda	$	$$	$$$
Instâncias provisionadas	100	-	-
Reserva de capacidade aberta utilizada	-	-	-
Reservas de capacidade aberta disponíveis	-	-	100

Configurar frotas de instâncias para usar reservas de capacidade aberta com base no melhor esforço

Ao usar a ação `RunJobFlow` para criar um cluster baseado em frota de instâncias, defina a estratégia de alocação sob demanda para `lowest-price` e `CapacityReservationPreference` para as opções de reservas de capacidade como `open`. Como alternativa, se você deixar esse campo em branco, a Amazon EMR padronizará a preferência de reserva de capacidade da instância sob demanda para. `open`

```
"LaunchSpecifications":
    {"OnDemandSpecification": {
        "AllocationStrategy": "lowest-price",
        "CapacityReservationOptions":
         {
            "CapacityReservationPreference": "open"
         }
        }
    }
```

Você também pode usar a Amazon EMR CLI para criar um cluster baseado em frota de instâncias usando reservas de capacidade aberta.

```
aws emr create-cluster \
 --name 'open-ODCR-cluster' \
 --release-label emr-5.30.0 \
 --service-role EMR_DefaultRole \
 --ec2-attributes SubnetId=subnet-22XXXX01,InstanceProfile=EMR_EC2_DefaultRole \
 --instance-fleets
 InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=c4.xlarg
 \

 InstanceFleetType=CORE,TargetOnDemandCapacity=100,InstanceTypeConfigs=['{InstanceType=c5.xlarg
{InstanceType=m5.xlarge},{InstanceType=r5.xlarge}'],\
   LaunchSpecifications={OnDemandSpecification='{AllocationStrategy=lowest-
price,CapacityReservationOptions={CapacityReservationPreference=open}}'}
```

Onde,

- Substitui-se `open-ODCR-cluster` pelo nome do cluster usando reservas de capacidade abertas.
- Substitui-se `subnet-22XXXX01` pelo ID da sub-rede.

Usar primeiro as reservas de capacidade aberta

Você pode optar por substituir a estratégia de alocação de menor preço e priorizar o uso das reservas de capacidade aberta disponíveis primeiro ao provisionar um cluster da Amazon. EMR Nesse caso, a Amazon EMR avalia todos os pools de instâncias com reservas de capacidade especificadas na solicitação de lançamento e usa aquele com o menor preço que tenha capacidade suficiente para iniciar todos os nós principais solicitados. Se nenhum dos pools de instâncias com reservas de capacidade tiver capacidade suficiente para os nós principais solicitados, a EMR

Amazon recorre ao caso de melhor esforço descrito no tópico anterior. Ou seja, a Amazon EMR reavalia todos os pools de instâncias especificados na solicitação de lançamento e usa aquele com o menor preço que tenha capacidade suficiente para iniciar todos os nós principais solicitados. As reservas de capacidade aberta disponíveis que correspondem ao grupo de instâncias são aplicadas automaticamente. Se as reservas de capacidade aberta disponíveis não corresponderem ao grupos de instâncias, elas permanecerão inutilizadas.

Depois que os nós centrais são provisionados, a zona de disponibilidade é selecionada e corrigida. A Amazon EMR provisiona os nós de tarefas em grupos de instâncias com reservas de capacidade, começando primeiro com os de menor preço, na zona de disponibilidade selecionada, até que todos os nós da tarefa sejam provisionados. A Amazon EMR usa primeiro as reservas de capacidade aberta disponíveis em cada pool de instâncias na zona de disponibilidade selecionada e, somente se necessário, usa a estratégia de menor preço para provisionar quaisquer nós de tarefas restantes.

A seguir estão os casos de uso da lógica de alocação de EMR capacidade da Amazon para usar primeiro as reservas de capacidade aberta.

Exemplo 1: o grupo de instâncias com reservas de capacidade aberta disponíveis na solicitação de inicialização tem capacidade suficiente para os nós centrais

Nesse caso, a Amazon EMR lança capacidade no pool de instâncias com reservas de capacidade aberta disponíveis, independentemente do preço do pool de instâncias. Como resultado, as reservas de capacidade aberta são usadas sempre que possível, até que todos os nós centrais sejam provisionados.

Estratégia sob demanda	preço mais baixo		
Capacidade solicitada	100		
Estratégia de uso	use-capacity-reservations-first		
Tipo de instância	c5.xlarge	m5.xlarge	r5.xlarge
Reservas de capacidade aberta disponíveis	-	-	150
Preço sob demanda	$	$$	$$$

Instâncias provisionadas	-	-	100
Reserva de capacidade aberta utilizada	-	-	100
Reservas de capacidade aberta disponíveis	-	-	50

Exemplo 2: o grupo de instâncias com reservas de capacidade aberta disponíveis na solicitação de inicialização não tem capacidade suficiente para os nós centrais

Nesse caso, a Amazon volta EMR a lançar nós principais usando a estratégia de menor preço com o melhor esforço para usar as reservas de capacidade.

Estratégia sob demanda	preço mais baixo		
Capacidade solicitada	100		
Estratégia de uso	use-capacity-reservations-first		
Tipo de instância	c5.xlarge	m5.xlarge	r5.xlarge
Reservas de capacidade aberta disponíveis	10	50	50
Preço sob demanda	$	$$	$$$
Instâncias provisionadas	100	-	-
Reserva de capacidade aberta utilizada	10	-	-

Reservas de capacidade aberta disponíveis	-	50	50

Depois que a frota de instâncias for iniciada, você poderá executar `describe-capacity-reservations` para ver quantas reservas de capacidade não utilizadas restam.

Configurar frotas de instâncias para usar primeiro reservas de capacidade aberta

Ao usar a ação `RunJobFlow` para criar um cluster baseado em frota de instâncias, defina a estratégia de alocação sob demanda para `lowest-price` e `UsageStrategy` para `CapacityReservationOptions` como `use-capacity-reservations-first`.

```
"LaunchSpecifications":
    {"OnDemandSpecification": {
        "AllocationStrategy": "lowest-price",
        "CapacityReservationOptions":
         {
            "UsageStrategy": "use-capacity-reservations-first"
         }
       }
    }
```

Você também pode usar a Amazon EMR CLI para criar um cluster baseado em frota de instâncias usando primeiro as reservas de capacidade.

```
aws emr create-cluster \
  --name 'use-CR-first-cluster' \
  --release-label emr-5.30.0 \
  --service-role EMR_DefaultRole \
  --ec2-attributes SubnetId=subnet-22XXXX01,InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets \

 InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=c4.xlarg
 \

 InstanceFleetType=CORE,TargetOnDemandCapacity=100,InstanceTypeConfigs=['{InstanceType=c5.xlarg
{InstanceType=m5.xlarge},{InstanceType=r5.xlarge}'],\
LaunchSpecifications={OnDemandSpecification='{AllocationStrategy=lowest-
price,CapacityReservationOptions={UsageStrategy=use-capacity-reservations-first}}'}
```

Onde,

- Substitui-se `use-CR-first-cluster` pelo nome do cluster usando reservas de capacidade abertas.
- Substitui-se `subnet-22XXXX01` pelo ID da sub-rede.

Usar primeiro as reservas de capacidade direcionadas

Ao provisionar um EMR cluster da Amazon, você pode optar por ignorar a estratégia de alocação de menor preço e priorizar primeiro o uso das reservas de capacidade direcionada disponíveis. Nesse caso, a Amazon EMR avalia todos os pools de instâncias com reservas de capacidade específicas especificadas na solicitação de lançamento e escolhe aquele com o menor preço que tenha capacidade suficiente para iniciar todos os nós principais solicitados. Se nenhum dos pools de instâncias com reservas de capacidade direcionadas tiver capacidade suficiente para os nós principais, a EMR Amazon recorre ao caso de melhor esforço descrito anteriormente. Ou seja, a Amazon EMR reavalia todos os pools de instâncias especificados na solicitação de execução e seleciona aquele com o menor preço que tenha capacidade suficiente para iniciar todos os nós principais solicitados. As reservas de capacidade aberta disponíveis que correspondem ao grupo de instâncias são aplicadas automaticamente. No entanto, as reservas de capacidade direcionadas permanecem inutilizadas.

Depois que os nós centrais são provisionados, a zona de disponibilidade é selecionada e corrigida. A Amazon EMR provisiona os nós de tarefas em grupos de instâncias com reservas de capacidade direcionadas, começando primeiro com os de menor preço, na zona de disponibilidade selecionada, até que todos os nós da tarefa sejam provisionados. A Amazon EMR tenta primeiro usar as reservas de capacidade direcionada disponíveis em cada pool de instâncias na zona de disponibilidade selecionada. Então, somente se necessário, a Amazon EMR usa a estratégia de menor preço para provisionar quaisquer nós de tarefas restantes.

A seguir estão os casos de uso da lógica de alocação de EMR capacidade da Amazon para usar primeiro as reservas de capacidade direcionadas.

Exemplo 1: O grupo de instâncias com reservas de capacidade direcionada disponíveis na solicitação de inicialização tem capacidade suficiente para os nós centrais

Nesse caso, a Amazon EMR lança capacidade no pool de instâncias com reservas de capacidade direcionadas disponíveis, independentemente do preço do pool de instâncias. Como resultado, as reservas de capacidade direcionada são usadas sempre que possível, até que todos os nós centrais sejam provisionados.

Estratégia sob demanda	preço mais baixo		
Estratégia de uso	use-capacity-reservations-first		
Capacidade solicitada	100		
Tipo de instância	c5.xlarge	m5.xlarge	r5.xlarge
Reservas de capacidade direcionada disponíveis	-	-	150
Preço sob demanda	$	$$	$$$
Instâncias provisionadas	-	-	100
Reserva de capacidade direcionada utilizada	-	-	100
Reservas de capacidade direcionada disponíveis	-	-	50

Example Exemplo 2: o grupo de instâncias com reservas de capacidade direcionada disponíveis na solicitação de inicialização não tem capacidade suficiente para os nós centrais

Estratégia sob demanda	preço mais baixo		
Capacidade solicitada	100		
Estratégia de uso	use-capacity-reservations-first		
Tipo de instância	c5.xlarge	m5.xlarge	r5.xlarge

Reservas de capacidade direcionada disponíveis	10	50	50
Preço sob demanda	$	$$	$$$
Instâncias provisionadas	100	-	-
Reservas de capacidade direcionada utilizadas	10	-	-
Reservas de capacidade direcionada disponíveis	-	50	50

Depois que a frota de instâncias for iniciada, você poderá executar `describe-capacity-reservations` para ver quantas reservas de capacidade não utilizadas restam.

Configurar frotas de instâncias para usar primeiro reservas de capacidade direcionada

Ao usar a ação `RunJobFlow` para criar um cluster baseado em frota de instâncias, defina a estratégia de alocação sob demanda como `lowest-price`, `UsageStrategy` de `CapacityReservationOptions` como `use-capacity-reservations-first` e `CapacityReservationResourceGroupArn` de `CapacityReservationOptions` como `<your resource group ARN>`. Para obter mais informações, consulte Trabalhar com reservas de capacidade no Guia EC2 do usuário da Amazon.

```
"LaunchSpecifications":
    {"OnDemandSpecification": {
        "AllocationStrategy": "lowest-price",
        "CapacityReservationOptions":
         {
            "UsageStrategy": "use-capacity-reservations-first",
            "CapacityReservationResourceGroupArn": "arn:aws:resource-groups:sa-
east-1:123456789012:group/MyCRGroup"
         }
       }
```

```
    }
```

Onde `arn:aws:resource-groups:sa-east-1:123456789012:group/MyCRGroup` é substituído pelo seu grupo de recursosARN.

Você também pode usar a Amazon EMR CLI para criar um cluster baseado em frota de instâncias usando reservas de capacidade direcionadas.

```
aws emr create-cluster \
  --name 'targeted-CR-cluster' \
  --release-label emr-5.30.0 \
  --service-role EMR_DefaultRole \
  --ec2-attributes SubnetId=subnet-22XXXX01,InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets
 InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=c4.xlarg
 \
    InstanceFleetType=CORE,TargetOnDemandCapacity=100,\
InstanceTypeConfigs=['{InstanceType=c5.xlarge},{InstanceType=m5.xlarge},
{InstanceType=r5.xlarge}'],\
LaunchSpecifications={OnDemandSpecification='{AllocationStrategy=lowest-
price,CapacityReservationOptions={UsageStrategy=use-capacity-reservations-
first,CapacityReservationResourceGroupArn=arn:aws:resource-groups:sa-
east-1:123456789012:group/MyCRGroup}}'}
```

Onde,

- Substitui-se `targeted-CR-cluster` pelo nome do cluster usando reservas de capacidade direcionadas.
- Substitui-se `subnet-22XXXX01` pelo ID da sub-rede.
- `arn:aws:resource-groups:sa-east-1:123456789012:group/MyCRGroup`é substituído pelo seu grupo de recursosARN.

Evitar usar reservas de capacidade aberta disponíveis

Example

Se você quiser evitar o uso inesperado de qualquer uma de suas reservas de capacidade aberta ao lançar um EMR cluster da Amazon, defina a estratégia de alocação sob demanda para e para`lowest-price`. `CapacityReservationPreference` `CapacityReservationOptions` none Caso contrário, a Amazon EMR padronizará a preferência de reserva de capacidade da

instância sob demanda `open` e tentará usar as reservas de capacidade aberta disponíveis com base no melhor esforço.

```
"LaunchSpecifications":
    {"OnDemandSpecification": {
        "AllocationStrategy": "lowest-price",
        "CapacityReservationOptions":
         {
            "CapacityReservationPreference": "none"
         }
       }
    }
```

Você também pode usar a Amazon EMR CLI para criar um cluster baseado em frota de instâncias sem usar nenhuma reserva de capacidade aberta.

```
aws emr create-cluster \
  --name 'none-CR-cluster' \
  --release-label emr-5.30.0 \
  --service-role EMR_DefaultRole \
  --ec2-attributes SubnetId=subnet-22XXXX01,InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets \

 InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=c4.xlarg
 \

 InstanceFleetType=CORE,TargetOnDemandCapacity=100,InstanceTypeConfigs=['{InstanceType=c5.xlarg
{InstanceType=m5.xlarge},{InstanceType=r5.xlarge}'],\
LaunchSpecifications={OnDemandSpecification='{AllocationStrategy=lowest-
price,CapacityReservationOptions={CapacityReservationPreference=none}}'}
```

Onde,

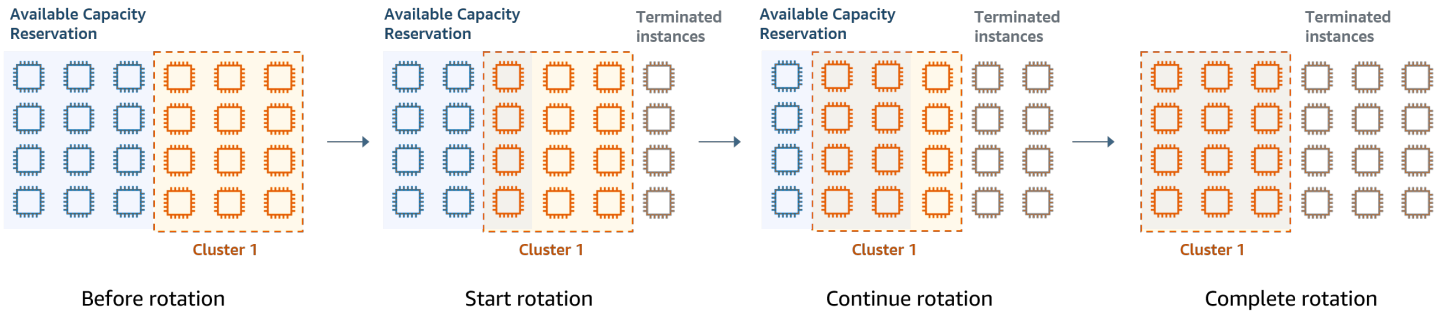
- Substitui-se `none-CR-cluster` pelo nome do cluster que não está usando reservas de capacidade abertas.
- Substitui-se `subnet-22XXXX01` pelo ID da sub-rede.

Cenários para o uso de reservas de capacidade

Você pode se beneficiar do uso de reservas de capacidade nos cenários a seguir.

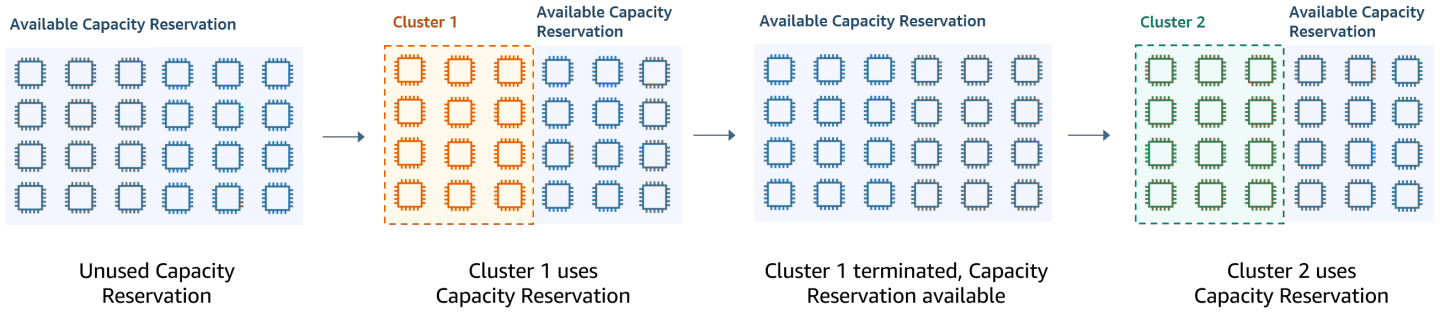
Cenário 1: alternar um cluster de execução prolongada usando reservas de capacidade

Ao alternar um cluster de execução prolongada, você poderá ter requisitos rígidos sobre os tipos de instância e as zonas de disponibilidade das novas instâncias provisionadas. Com as reservas de capacidade, você pode usar a garantia de capacidade para concluir a alternância do cluster sem interrupções.



Cenário 2: provisionar clusters sucessivos de curta duração usando reservas de capacidade

Também é possível usar reservas de capacidade para provisionar um grupo de clusters sucessivos e de curta duração para workloads individuais, de forma que, ao encerrar um cluster, o próximo cluster possa usar as reservas de capacidade. Você pode usar reservas de capacidade direcionadas para garantir que apenas os clusters pretendidos usem as reservas de capacidade.



Configurar grupos de instâncias uniformes

Com a configuração de grupos de instâncias, cada tipo de nó (principal, core ou tarefa) consiste no mesmo tipo de instância e na mesma opção de compra para instâncias: Sob demanda ou Spot. Você especifica essas configurações ao criar um grupo de instâncias. Não é possível alterá-las depois. No entanto, você pode adicionar instâncias do mesmo tipo e opção de compra a grupos de instâncias core e de tarefas. Você também pode remover instâncias.

Se as instâncias sob demanda do cluster corresponderem aos atributos das reservas de capacidade aberta (tipo de instância, plataforma, locação e zona de disponibilidade) disponíveis na conta, as reservas de capacidade serão aplicadas automaticamente. É possível usar reservas de capacidade

aberta para nós primários, centrais e de tarefa. No entanto, você não poderá usar reservas de capacidade direcionadas nem impedir que instâncias sejam iniciadas em reservas de capacidade abertas com atributos correspondentes ao provisionar clusters usando grupos de instâncias. Para usar reservas de capacidade direcionadas ou evitar que instâncias sejam iniciadas em reservas de capacidade abertas, use frotas de instâncias. Para obter mais informações, consulte Usar reservas de capacidade com a frotas de instância.

Para adicionar tipos de instâncias diferentes depois que um cluster for criado, é possível adicionar outros grupos de instâncias de tarefas. Você pode escolher diferentes tipos de instância e opções de compra para cada grupo de instância. Para obter mais informações, consulte Usar ajuste de escala de clusters.

Ao iniciar instâncias, a preferência da reserva de capacidade da instância sob demanda será padronizada como `open`, o que permitirá que ela seja executada em qualquer reserva de capacidade em aberto que tenha atributos correspondentes (tipo de instância, plataforma, zona de disponibilidade). Para obter mais informações sobre reservas de capacidade sob demanda, consulte Usar reservas de capacidade com a frotas de instância.

Esta seção discute a criação de um cluster com grupos de instâncias uniformes. Para obter mais informações sobre como modificar um grupo de instâncias existente, adicionando ou removendo instâncias manualmente ou com escalabilidade automática, consulte Gerenciar clusters.

Usar o console para configurar grupos de instâncias uniformes

Console

Criar um cluster com grupos de instâncias usando o novo console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EMREC2Em Ativado, no painel de navegação esquerdo, escolha Clusters e escolha Criar cluster.
3. Em Configuração do cluster, escolha Grupos de instâncias.
4. Em Grupos de nós, há uma seção para cada tipo de grupo de nós. Para o grupo de nós primários, marque a caixa de seleção Usar múltiplos nós primários se quiser ter três nós primários. Marque a caixa de seleção Usar a opção de compra spot se quiser usar a compra spot.
5. Para os grupos de nós primários e centrais, selecione Adicionar tipo de instância e escolha até cinco tipos de instância. Para o grupo de tarefa, selecione Adicionar tipo de instância e

escolha até 15 tipos de instância. A Amazon EMR pode provisionar qualquer combinação desses tipos de instância ao iniciar o cluster.

6. Para alterar essas configurações, em cada tipo de grupo de nós, escolha o menu suspenso Ações ao lado de cada instância:

   Adicionar EBS volumes

   Especifique EBS os volumes a serem anexados ao tipo de instância depois que a Amazon a EMR provisionar.

   Editar preço máximo spot

   Especifique um preço spot máximo para cada tipo de instância da frota. Você pode definir esse preço como uma porcentagem do preço sob demanda ou como uma quantia em dólar. Se o preço spot atual em uma zona de disponibilidade estiver abaixo do seu preço spot máximo, a Amazon EMR provisiona instâncias spot. Você paga o preço spot e não necessariamente o preço spot máximo.
7. Opcionalmente, expanda a configuração do Node para entrar em uma JSON configuração ou para carregar a JSON partir do Amazon S3.
8. Escolha qualquer outra opção que se aplique ao cluster.
9. Para iniciar o cluster, escolha Criar cluster.

Use o AWS CLI para criar um cluster com grupos de instâncias uniformes

Para especificar a configuração de grupos de instâncias para um cluster usando a AWS CLI, use o comando `create-cluster` junto com o parâmetro `--instance-groups`. A Amazon EMR assume a opção de instância sob demanda, a menos que você especifique o `BidPrice` argumento para um grupo de instâncias. Para obter exemplos de comandos `create-cluster` que executam grupos de instâncias uniformes com instâncias sob demanda e uma variedade de opções de cluster, digite `aws emr create-cluster help` na linha de comando, ou consulte create-cluster na AWS CLI Command Reference.

Você pode usar o AWS CLI para criar grupos de instâncias uniformes em um cluster que usa instâncias spot. O preço Spot oferecido depende da zona de disponibilidade. Ao usar o CLI ouAPI, você pode especificar a Zona de Disponibilidade com o `AvailabilityZone` argumento (se estiver usando uma rede EC2 -clássica) ou com o `SubnetID` argumento do `--ec2-attributes` parâmetro. A zona de disponibilidade ou sub-rede selecionada se aplica ao cluster e, portanto, é usada para todos os grupos de instâncias. Se você não especificar explicitamente uma zona de

disponibilidade ou sub-rede, a Amazon EMR selecionará a zona de disponibilidade com o menor preço spot ao iniciar o cluster.

O exemplo a seguir demonstra um comando `create-cluster` que cria um grupo de instâncias primárias, um grupo de instâncias centrais e dois grupos de instâncias de tarefa, todos usando instâncias spot. Substituir *myKey* com o nome do seu par de EC2 chaves da Amazon.

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --name "MySpotCluster" \
  --release-label emr-7.2.0 \
  --use-default-roles \
  --ec2-attributes KeyName=myKey \
  --instance-groups \
    InstanceGroupType=MASTER,InstanceType=m5.xlarge,InstanceCount=1,BidPrice=0.25 \
    InstanceGroupType=CORE,InstanceType=m5.xlarge,InstanceCount=2,BidPrice=0.03 \
    InstanceGroupType=TASK,InstanceType=m5.xlarge,InstanceCount=4,BidPrice=0.03 \
    InstanceGroupType=TASK,InstanceType=m5.xlarge,InstanceCount=2,BidPrice=0.04
```

Usando oCLI, você pode criar clusters de grupos de instâncias uniformes que especificam um personalizado exclusivo AMI para cada tipo de instância no grupo de instâncias. Assim, você pode usar arquiteturas de instância diferentes no mesmo grupo de instâncias. Cada tipo de instância deve usar uma arquitetura personalizada AMI com uma arquitetura correspondente. Por exemplo, você configuraria um tipo de instância m5.xlarge com uma arquitetura x86_64 personalizada AMI e um tipo de instância m6g.xlarge com uma arquitetura () personalizada correspondente. `AWS  AARCH64` ARM AMI

O exemplo a seguir mostra um cluster uniforme de grupos de instâncias criado com dois tipos de instância, cada um com sua própria personalizaçãoAMI. Observe que os personalizados AMIs são especificados somente no nível do tipo de instância, não no nível do cluster. Isso é para evitar conflitos entre o tipo de instância AMIs e an AMI no nível do cluster, o que faria com que a inicialização do cluster falhasse.

```
aws emr create-cluster
```

```
  --release-label emr-5.30.0 \
  --service-role EMR_DefaultRole \
  --ec2-attributes SubnetId=subnet-22XXXX01,InstanceProfile=EMR_EC2_DefaultRole \
  --instance-groups \

 InstanceGroupType=MASTER,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-123456
 \

 InstanceGroupType=CORE,InstanceType=m6g.xlarge,InstanceCount=1,CustomAmiId=ami-234567
```

Você pode adicionar vários personalizados AMIs a um grupo de instâncias que você adiciona a um cluster em execução. O argumento `CustomAmiId` pode ser usado com o comando `add-instance-groups`, conforme mostrado no exemplo a seguir.

```
aws emr add-instance-groups --cluster-id j-123456 \
  --instance-groups \

 InstanceGroupType=Task,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-123456
```

Use o Java SDK para criar um grupo de instâncias

Você instancia um objeto `InstanceGroupConfig` que especifica a configuração de um grupo de instâncias para um cluster. Para usar instâncias Spot, defina as propriedades `withBidPrice` e `withMarket` no objeto `InstanceGroupConfig`. O código a seguir mostra como definir grupos de instância primários, centrais e de tarefa que executam instâncias Spot.

```
InstanceGroupConfig instanceGroupConfigMaster = new InstanceGroupConfig()
 .withInstanceCount(1)
 .withInstanceRole("MASTER")
 .withInstanceType("m4.large")
 .withMarket("SPOT")
 .withBidPrice("0.25");

InstanceGroupConfig instanceGroupConfigCore = new InstanceGroupConfig()
 .withInstanceCount(4)
 .withInstanceRole("CORE")
 .withInstanceType("m4.large")
 .withMarket("SPOT")
 .withBidPrice("0.03");

InstanceGroupConfig instanceGroupConfigTask = new InstanceGroupConfig()
 .withInstanceCount(2)
```

```
  .withInstanceRole("TASK")
  .withInstanceType("m4.large")
  .withMarket("SPOT")
  .withBidPrice("0.10");
```

Práticas recomendadas para flexibilidade de instâncias e de zona de disponibilidade

Cada um Região da AWS tem vários locais isolados, conhecidos como zonas de disponibilidade. Ao iniciar uma instância, é possível especificar, opcionalmente, uma zona de disponibilidade (AZ) na Região da AWS utilizada. A flexibilidade da zona de disponibilidade é a distribuição de instâncias em váriasAZs. Se houver falha em uma instância, você poderá projetar sua aplicação para que uma instância em outra AZ possa lidar com as solicitações. Para obter mais informações sobre zonas de disponibilidade, consulte a documentação da região e das zonas no Guia EC2 do usuário da Amazon.

A flexibilidade da instância é o uso de múltiplos tipos de instância para atender aos requisitos de capacidade. Ao expressar flexibilidade com instâncias, é possível usar a capacidade agregada em todos os tamanhos, famílias e gerações de instâncias. Uma flexibilidade maior aumenta a chance de encontrar e alocar a quantidade necessária de capacidade computacional comparado a um cluster que usa um único tipo de instância.

A flexibilidade da instância e da zona de disponibilidade reduz erros de capacidade insuficientes (ICE) e interrupções pontuais quando comparada a um cluster com um único tipo de instância ou AZ. Use as práticas recomendadas abordadas aqui para determinar quais instâncias diversificar depois de conhecer a família e o tamanho iniciais da instância. Essa abordagem maximiza a disponibilidade dos pools de EC2 capacidade da Amazon com o mínimo de desempenho e variação de custo.

Ser flexível em relação às zonas de disponibilidade

Recomendamos que você configure todas as zonas de disponibilidade para uso em sua nuvem privada virtual (VPC) e que as selecione para seu EMR cluster. Os clusters devem existir em apenas uma zona de disponibilidade, mas com as frotas de EMR instâncias da Amazon, você pode selecionar várias sub-redes para diferentes zonas de disponibilidade. Quando a Amazon EMR lança o cluster, ela examina essas sub-redes para encontrar as instâncias e as opções de compra que você especifica. Quando você provisiona um EMR cluster para várias sub-redes, seu cluster pode acessar um pool de EC2 capacidade mais profundo da Amazon quando comparado aos clusters em uma única sub-rede.

Se você precisar priorizar um certo número de zonas de disponibilidade para uso em sua nuvem privada virtual (VPC) para seu EMR cluster, você pode aproveitar o recurso de pontuação de

posicionamento Spot com a Amazon. EC2 Com a pontuação de posicionamento spot, você especifica os requisitos de computação para suas instâncias spot e, em seguida, EC2 retorna as dez Regiões da AWS principais zonas de disponibilidade pontuadas em uma escala de 1 a 10. Uma pontuação de 10 indica que a solicitação spot tem alta probabilidade de êxito; uma pontuação de 1 indica que a solicitação spot provavelmente não terá êxito. Para obter mais informações sobre como usar a pontuação de posicionamento spot, consulte Pontuação de posicionamento spot no Guia do EC2 usuário da Amazon.

Ser flexível em relação aos tipos de instância

A flexibilidade da instância é o uso de múltiplos tipos de instância para atender aos requisitos de capacidade. A flexibilidade da instância beneficia tanto o uso do Amazon EC2 Spot quanto do On-Demand Instance. Com as Instâncias Spot, a flexibilidade da instância permite que a Amazon EC2 lance instâncias a partir de pools de capacidade mais profundos usando dados de capacidade em tempo real. Também prevê quais instâncias estão mais disponíveis. Isso oferece menos interrupções e pode reduzir o custo geral da workload. Com instâncias sob demanda, a flexibilidade da instância reduz erros de capacidade insuficientes (ICE) quando a capacidade total é provisionada em um número maior de pools de instâncias.

Para clusters de grupos de instâncias, você pode especificar até 50 tipos de EC2 instância. Para frotas de instâncias com estratégia de alocação, você pode especificar até 30 tipos de EC2 instância para cada grupo de nós primário, principal e de tarefas. Uma variedade maior de instâncias melhora os benefícios da flexibilidade da instância.

Expressar a flexibilidade da instância

Considere as práticas recomendadas a seguir para expressar a flexibilidade de instância da aplicação.

Tópicos

- Determinar a família e o tamanho da instância
- Incluir instâncias adicionais

Determinar a família e o tamanho da instância

A Amazon EMR oferece suporte a vários tipos de instância para diferentes casos de uso. Esses tipos de instância estão listados na documentação Tipos de instâncias compatíveis. Cada tipo de instância pertence a uma família de instâncias que descreve para qual aplicação o tipo é otimizado.

Para novas workloads, compare com os tipos de instância da família de uso geral, como `m5` ou `c5`. Em seguida, monitore o sistema operacional e YARN as métricas do Ganglia e determine os Amazon CloudWatch gargalos do sistema no pico de carga. Os gargalos incluem memóriaCPU, armazenamento e operações de E/S. Após identificar os gargalos, escolha otimizado para computação, otimizado para memória, otimizado para armazenamento ou outra família de instâncias apropriada para seus tipos de instância. Para obter mais detalhes, consulte a página Determine a infraestrutura certa para suas cargas de trabalho do Spark no guia de EMR melhores práticas da Amazon em. GitHub

Em seguida, identifique o menor YARN contêiner ou executor do Spark que seu aplicativo exige. Esse é o menor tamanho de instância adequado ao contêiner e o tamanho mínimo de instância para o cluster. Use essa métrica para determinar instâncias com as quais você poderá diversificar ainda mais. Uma instância menor permitirá maior flexibilidade da instância.

Para obter a máxima flexibilidade da instância, você deve aproveitar o maior número possível de instâncias. É recomendável diversificar com instâncias que tenham especificações de hardware semelhantes. Isso maximiza o acesso aos pools EC2 de capacidade com variação mínima de custo e desempenho. Diversifique em vários tamanhos. Para fazer isso, priorize antes o AWS Graviton e as gerações anteriores. Uma boa regra geral é tentar ser flexível para pelo menos 15 tipos de instância para cada workload. É recomendável começar com instâncias de uso geral, otimizadas para computação ou para memória. Esses tipos de instância fornecerão a maior flexibilidade.

Incluir instâncias adicionais

Para ter o máximo de diversidade, inclua outros tipos de instância. Priorize antes o tamanho da instância, o Graviton e a flexibilidade da geração. Isso permite o acesso a grupos EC2 de capacidade adicionais com perfis de custo e desempenho semelhantes. Se você precisar de mais flexibilidade devido a interrupções ICE ou interrupções pontuais, considere a flexibilidade de variantes e famílias. Cada abordagem tem vantagens e desvantagens conforme o caso de uso e os requisitos.

- Flexibilidade de tamanho: primeiro, diversifique com instâncias de tamanhos diferentes dentro da mesma família. As instâncias da mesma família oferecem o mesmo custo e performance, mas podem iniciar um número diferente de contêineres em cada host. Por exemplo, se o tamanho mínimo do executor necessário for de 2 V CPU e 8 Gb de memória, o tamanho mínimo da instância será. `m5.xlarge` Para flexibilidade de tamanho, inclua `m5.xlarge`, `m5.2xlarge`, `m5.4xlarge`, `m5.8xlarge`, `m5.12xlarge`, `m5.16xlarge` e `m5.24xlarge`.
- Flexibilidade do Graviton: além do tamanho, você pode diversificar com instâncias do Graviton. As instâncias Graviton são alimentadas por processadores AWS Graviton2 que oferecem a

melhor relação preço/desempenho para cargas de trabalho em nuvem na Amazon. EC2 Por exemplo, com o tamanho mínimo de instância de `m5.xlarge`, você pode incluir `m6g.xlarge`, `m6g.2xlarge`, `m6g.4xlarge`, `m6g.8xlarge` e `m6g.16xlarge` para flexibilidade do Graviton.

- Flexibilidade de geração: semelhante ao Graviton e à flexibilidade de tamanho, as instâncias das famílias da geração anterior compartilham as mesmas especificações de hardware. Isso resulta em um perfil de custo e desempenho semelhante, com um aumento no EC2 pool total acessível da Amazon. Para flexibilidade de geração, inclua `m4.xlarge`, `m4.2xlarge`, `m4.10xlarge` e `m4.16xlarge`.
- Flexibilidade de famílias e variantes
  - Capacidade: para otimizar a capacidade, recomendamos a flexibilidade de instâncias em todas as famílias de instâncias. Instâncias comuns de diferentes famílias de instâncias têm grupos de instâncias mais profundos que ajudam a atender aos requisitos de capacidade. No entanto, instâncias de famílias diferentes terão diferentes proporções de v CPU para memória. Isso resultará em subutilização se o contêiner da aplicação esperado for dimensionado para uma instância diferente. Por exemplo, com `m5.xlarge`, inclua instâncias otimizadas para computação, como `c5`, ou instâncias otimizadas para memória, como `r5` para obter flexibilidade de família de instâncias.
  - Custo: para otimizar o custo, recomenda-se a flexibilidade da instância em todas as variantes. Essas instâncias têm a mesma memória e CPU proporção v da instância inicial. A desvantagem com a flexibilidade de variantes é que essas instâncias têm grupos de capacidade menores, o que pode resultar em capacidade adicional limitada ou maiores interrupções spot. `m5.xlarge`Por exemplo, inclua instâncias AMD baseadas (`m5a`), instâncias SSD baseadas (`m5d`) ou instâncias otimizadas para rede (`m5n`) para flexibilidade de variantes de instância.

Práticas recomendadas para configuração de clusters

Use a orientação nesta seção para ajudá-lo a determinar os tipos de instância, as opções de compra e a quantidade de armazenamento a ser provisionada para cada tipo de nó em um EMR cluster.

Que tipo de instância você deve usar?

Há várias maneiras de adicionar EC2 instâncias da Amazon a um cluster. O método a ser escolhido depende se você usará a configuração de grupos de instâncias ou a configuração de frotas de instâncias para o cluster.

- Grupos de instâncias

  - Adicione manualmente instâncias do mesmo tipo a grupos de instâncias core e de tarefa existentes.
  - Adicione manualmente um grupo de instâncias de tarefa, que pode usar um tipo de instância diferente.
  - Configure a escalabilidade automática na Amazon EMR para um grupo de instâncias, adicionando e removendo instâncias automaticamente com base no valor de uma CloudWatch métrica da Amazon que você especificar. Para obter mais informações, consulte Usar ajuste de escala de clusters.
- Frotas de instâncias
  - Adicione uma única frota de instâncias de tarefa.
  - Altere a capacidade de destino para instâncias sob demanda e instâncias spot para as frotas de instâncias core e de tarefas existentes. Para obter mais informações, consulte Configurar frotas de instâncias.

Uma maneira de planejar as instâncias do seu cluster é executar um cluster de teste com um conjunto de dados de amostra representativo e monitorar a utilização dos nós nesse cluster. Para obter mais informações, consulte Visualizar e monitorar um cluster. Outra maneira é calcular a capacidade das instâncias que você está considerando e comparar esse valor com o tamanho dos seus dados.

Em geral, o tipo de nó principal, que atribui tarefas, não exige uma EC2 instância com muito poder de processamento; as EC2 instâncias da Amazon para o tipo de nó principal, que processam tarefas e armazenam dadosHDFS, precisam tanto de poder de processamento quanto de capacidade de armazenamento; as EC2 instâncias da Amazon para o tipo de nó de tarefa, que não armazenam dados, precisam apenas de poder de processamento. Para obter diretrizes sobre as EC2 instâncias disponíveis da Amazon e suas configurações, consulteConfigurar EC2 instâncias da Amazon.

As diretrizes a seguir se aplicam à maioria dos EMR clusters da Amazon.

- Há um CPU limite v para o número total de EC2 instâncias sob demanda da Amazon que você executa em uma AWS conta por Região da AWS. Para obter mais informações sobre o CPU limite v e como solicitar um aumento de limite para sua conta, consulte Instâncias sob demanda no Guia do EC2 usuário da Amazon para instâncias Linux.
- O nó primário normalmente não tem grandes requisitos de computação. Para clusters com um grande número de nós ou para clusters com aplicativos que são implantados especificamente no nó primário (JupyterHub, Hue etc.), um nó primário maior pode ser necessário e pode ajudar a

melhorar o desempenho do cluster. Por exemplo, considere usar uma instância m5.xlarge para clusters pequenos (até 50 nós) e aumentar para um tipo de instância maior para clusters maiores.

- As necessidades de computação dos nós core e de tarefas dependem do tipo de processamento realizado pelo seu aplicativo. Muitos trabalhos podem ser executados em tipos de instância de uso geral, que oferecem desempenho equilibrado em termos de CPU espaço em disco e entrada/saída. Clusters com uso intensivo de computação podem se beneficiar da execução em CPU instâncias altas, que têm proporcionalmente mais de. CPU RAM Aplicativos de banco de dados e de cache de memória podem se beneficiar com a execução em instâncias com mais memória. Aplicativos com uso intenso e CPU intenso de rede, como análise e aprendizado de máquinaNLP, podem se beneficiar da execução em instâncias de computação em cluster, que fornecem recursos proporcionalmente altos e maior desempenho da rede. CPU
- Se diferentes fases do seu cluster tiverem necessidades de capacidade diferentes, você pode começar com um pequeno número de nós core e aumentar ou diminuir o número de nós de tarefas para atender aos requisitos de capacidade variáveis do seu fluxo de trabalho.
- A quantidade de dados que você pode processar depende da capacidade de nós core e do tamanho dos seus dados como entrada, durante o processamento, e como saída. Os conjuntos de dados de entrada, intermediários e de saída residem todos no cluster durante o processamento.

Quando você deve usar instâncias spot?

Ao iniciar um cluster na AmazonEMR, você pode optar por iniciar instâncias primárias, principais ou de tarefas nas Instâncias Spot. Como cada tipo de grupo de instâncias desempenha um papel diferente no cluster, há implicações na execução de cada tipo de nó em instâncias spot. Você não pode alterar uma opção de compra de instância enquanto um cluster está em execução. Para alterar um grupo de instâncias sob demanda para instâncias spot, ou vice-versa, para nós primários e centrais, você deve terminar o cluster e iniciar um novo. Para nós de tarefa, você pode iniciar um novo grupo de instâncias de tarefa ou frota de instâncias e remover o antigo.

Tópicos

- EMRConfigurações da Amazon para evitar falhas no trabalho devido ao encerramento da instância spot do nó da tarefa
- Nó primário como uma instância spot
- Nós centrais em instâncias spot
- Nós de tarefa em instâncias spot
- Configurações de instâncias para cenários de aplicações

EMRConfigurações da Amazon para evitar falhas no trabalho devido ao encerramento da instância spot do nó da tarefa

Como as Instâncias Spot são frequentemente usadas para executar nós de tarefas, a Amazon EMR tem a funcionalidade padrão para agendar YARN trabalhos para que os trabalhos em execução não falhem quando os nós de tarefas executados em Instâncias Spot forem encerrados. EMRA Amazon faz isso permitindo que os processos principais do aplicativo sejam executados somente nos nós principais. O processo principal da aplicação controla os trabalhos em execução e precisa permanecer ativo durante a vida útil do trabalho.

A EMR versão 5.19.0 e posterior da Amazon usa o recurso integrado de rótulos de YARN nós para fazer isso. (As versões anteriores usavam um patch de código). As propriedades nas classificações de `capacity-scheduler` configuração `yarn-site` e são configuradas por padrão para que o planejador YARN de capacidade e o agendador justo aproveitem os rótulos dos nós. A Amazon rotula EMR automaticamente os nós principais com o `CORE` rótulo e define as propriedades para que os mestres do aplicativo sejam programados somente nos nós com o CORE rótulo. A modificação manual das propriedades relacionadas nas classificações de configuração yarn-site e capacity-scheduler, ou diretamente nos XML arquivos associados, pode interromper esse recurso ou modificar essa funcionalidade.

A Amazon EMR configura as seguintes propriedades e valores por padrão. Tenha cuidado ao configurar essas propriedades.

> **Note**
>
> A partir da série de lançamento Amazon EMR 6.x, o recurso de rótulos de YARN nós está desativado por padrão. Os processos primários da aplicação podem ser executados tanto nos nós centrais como nos nós de tarefa por padrão. Você pode ativar o recurso de rótulos de YARN nós configurando as seguintes propriedades:
>
> - `yarn.node-labels.enabled: true`
> - `yarn.node-labels.am.default-node-label-expression: 'CORE'`

- yarn-site (yarn-site.xml) Em todos os nós
  - `yarn.node-labels.enabled: true`
  - `yarn.node-labels.am.default-node-label-expression: 'CORE'`
  - `yarn.node-labels.fs-store.root-dir: '/apps/yarn/nodelabels'`

  - `yarn.node-labels.configuration-type: 'distributed'`
- yarn-site (yarn-site.xml) em nós primários e centrais
  - `yarn.nodemanager.node-labels.provider: 'config'`
  - `yarn.nodemanager.node-labels.provider.configured-node-partition: 'CORE'`
- capacity-scheduler (capacity-scheduler.xml) Em todos os nós
  - `yarn.scheduler.capacity.root.accessible-node-labels: '*'`
  - `yarn.scheduler.capacity.root.accessible-node-labels.CORE.capacity: 100`
  - `yarn.scheduler.capacity.root.default.accessible-node-labels: '*'`
  - `yarn.scheduler.capacity.root.default.accessible-node-labels.CORE.capacity: 100`

Nó primário como uma instância spot

O nó primário controla e direciona o cluster. Quando ela for terminada, o cluster será encerrado. Portanto, você só deve iniciar o nó primário como uma instância spot se você estiver executando um cluster em que o término repentino seja aceitável. Este pode ser o caso se você está testando uma nova aplicação, tem um cluster que periodicamente mantém a persistência de dados em um armazenamento externo, como o Amazon S3, ou está executando um cluster em que o custo é mais importante do que garantir a conclusão do cluster.

Quando você executa o grupo de instâncias primárias como uma instância spot, o cluster não é iniciado até que essa solicitação de instância spot seja atendida. Isso é algo a considerar ao selecionar seu preço spot máximo.

Você só pode adicionar um nó primário de instância spot ao iniciar o cluster. Não é possível adicionar ou remover nós primários de um cluster em execução.

Normalmente, você só executaria o nó primário como uma instância spot se estivesse executando o cluster inteiro (todos os grupos de instâncias) como instâncias spot.

Nós centrais em instâncias spot

Os nós principais processam dados e armazenam informações usandoHDFS. O encerramento de uma instância core representa risco de perda de dados. Por esse motivo, você só deve executar nós principais em instâncias spot quando a perda parcial de HDFS dados for tolerável.

Quando você inicia o grupo de instâncias principais como Instâncias Spot, a Amazon EMR espera até que possa provisionar todas as instâncias principais solicitadas antes de iniciar o grupo de

instâncias. Em outras palavras, se você solicitar seis EC2 instâncias da Amazon e apenas cinco estiverem disponíveis no preço spot máximo ou abaixo dele, o grupo de instâncias não será iniciado. EMRA Amazon continua esperando até que todas as seis EC2 instâncias da Amazon estejam disponíveis ou até que você encerre o cluster. Você pode alterar o número de instâncias spot em um grupo de instâncias core para adicionar capacidade a um cluster em execução. Para obter mais informações sobre como trabalhar com grupos de instâncias, e como as instâncias spot funcionam com frotas de instâncias, consulte the section called “Configurar frotas de instâncias ou grupos de instâncias”.

Nós de tarefa em instâncias spot

Os nós da tarefa processam dados, mas não armazenam dados persistentesHDFS. Se eles forem encerrados porque o preço spot ultrapassou seu preço spot máximo, não haverá perda de dados, e o efeito no seu cluster será mínimo.

Quando você executa um ou mais grupos de instâncias de tarefas como Instâncias Spot, a Amazon EMR provisiona o maior número possível de nós de tarefas, usando seu preço spot máximo. Isso significa que, se você solicitar um grupo de instâncias de tarefas com seis nós e apenas cinco instâncias spot estiverem disponíveis no preço spot máximo ou abaixo dele, a Amazon EMR lançará o grupo de instâncias com cinco nós, adicionando o sexto posteriormente, se possível.

A execução de grupos de instâncias de tarefas como instâncias Spot é uma maneira estratégica de expandir a capacidade do seu cluster e, ao mesmo tempo, minimizar os custos. Se você executar os grupos de instâncias primárias e centrais como instâncias sob demanda, a capacidade será garantida para a execução do cluster. Você pode adicionar instâncias de tarefa aos grupos de instâncias da tarefa conforme necessário, para processar picos de tráfego ou agilizar processamento de dados.

Você pode adicionar ou remover nós de tarefas usando o console, AWS CLI, ouAPI. Você também pode acrescentar grupos de tarefas adicionais, mas não poderá remover um grupo de tarefas depois de criado.

Configurações de instâncias para cenários de aplicações

A tabela a seguir é uma referência rápida às opções de compras de tipos de nó e configurações que são geralmente apropriadas para vários cenários de aplicativos. Escolha o link para exibir mais informações sobre cada tipo de cenário.

Cenário de aplicações	Opção de compra do nó primário	Opção de compra de nós centrais	Opção de compra de nós de tarefa
Clusters de execução prolongada e data warehouses	Sob demanda	Combinação de frotas de instâncias ou Sob demanda	Combinação de spot ou frota de instâncias
Cargas de trabalho com base no custo	Spot	Spot	Spot
Cargas de trabalho críticas para dados	Sob demanda	Sob demanda	Combinação de spot ou frota de instâncias
Testes de aplicativos	Spot	Spot	Spot

Há vários cenários nos quais as Instâncias Spot são úteis para executar um EMR cluster da Amazon.

Clusters de execução prolongada e data warehouses

Se você estiver executando um EMR cluster persistente da Amazon que tem uma variação previsível na capacidade computacional, como um data warehouse, você pode lidar com picos de demanda a um custo menor com instâncias spot. Você pode iniciar seus grupos de instâncias primárias e central como instâncias sob demanda para lidar com a capacidade normal e iniciar o grupo de instâncias de tarefa como instâncias spot para lidar com requisitos de carga de pico.

Cargas de trabalho com base no custo

Ao executar clusters transitórios para os quais um custo menor é mais importante do que o tempo para conclusão, e uma perda parcial do trabalho é aceitável, você pode executar o cluster inteiro (grupos de instâncias primárias, centrais e de tarefa) como instâncias spot para se beneficiar com a maior redução dos custos.

Cargas de trabalho críticas para dados

Se você estiver executando um cluster para o qual o menor custo é mais importante que o tempo para conclusão, mas uma perda parcial do trabalho não é aceitável, inicie os grupos de instâncias primárias e centrais como instâncias sob demanda e complemente-as com um ou mais grupos de instâncias de tarefa de instâncias spot. A execução dos grupos de instâncias principal e principal

como instâncias sob demanda garante que seus dados persistam HDFS e que o cluster esteja protegido contra o encerramento devido às flutuações do mercado spot, ao mesmo tempo em que proporciona economias decorrentes da execução dos grupos de instâncias de tarefas como instâncias spot.

Testes de aplicativos

Ao testar uma nova aplicação a fim de prepará-la para inicialização em um ambiente de produção, você pode executar o cluster inteiro (grupos de instâncias primárias, centrais e de tarefa) como instâncias spot para reduzir os custos de testes.

Calculando a HDFS capacidade necessária de um cluster

A quantidade de HDFS armazenamento disponível para seu cluster depende dos seguintes fatores:

- O número de EC2 instâncias da Amazon usadas para os nós principais.
- A capacidade do armazenamento de EC2 instâncias da Amazon para o tipo de instância usado. Para obter mais informações sobre volumes de armazenamento de instâncias, consulte Amazon Amazon EC2 Instance Store no Guia EC2 do usuário da Amazon.
- O número e o tamanho dos EBS volumes da Amazon anexados aos nós principais.
- Um fator de replicação, que explica como cada bloco de dados é armazenado HDFS para uma redundância RAID semelhante. Por padrão, o fator de replicação é de três para um cluster de 10 ou mais nós core, dois para um cluster com 4 a 9 nós core e um para um cluster de três nós ou menos.

Para calcular a HDFS capacidade de um cluster, para cada nó central, adicione a capacidade do volume de armazenamento da instância à capacidade EBS de armazenamento da Amazon (se usada). Multiplique o resultado pelo número de nós core e, em seguida, divida o total pelo fator de replicação com base no número de nós core. Por exemplo, um cluster com 10 nós principais do tipo i2.xlarge, que tem 800 GB de armazenamento de instância sem nenhum EBS volume vinculado da Amazon, tem um total de aproximadamente 2.666 GB disponíveis para HDFS (10 nós x 800 GB ÷ 3 fatores de replicação).

Se o valor da HDFS capacidade calculada for menor que seus dados, você poderá aumentar a quantidade de HDFS armazenamento das seguintes formas:

- Criar um cluster com EBS volumes adicionais da Amazon ou adicionar grupos de instâncias com EBS volumes Amazon anexados a um cluster existente

- Adicionando mais nós core
- Escolha de um tipo de EC2 instância da Amazon com maior capacidade de armazenamento
- Usando a compactação de dados
- Alterando as definições de configuração do Hadoop para reduzir o fator de replicação

A redução do fator de replicação deve ser usada com cuidado, pois reduz a redundância dos HDFS dados e a capacidade do cluster de se recuperar de blocos perdidos ou corrompidos. HDFS

Configurar registro em log e depuração do cluster

Uma das questões a ser decidida quando você planeja o seu cluster é quanto de suporte à depuração você deseja disponibilizar. Quando você está desenvolvendo um aplicativo de processamento de dados pela primeira vez, recomendamos testar o aplicativo em um cluster processando um subconjunto pequeno, mas representativo, dos seus dados. Ao fazer isso, você provavelmente desejará aproveitar todas as ferramentas de depuração que a Amazon EMR oferece, como o arquivamento de arquivos de log no Amazon S3.

Uma vez concluído o desenvolvimento, e com o aplicativo de processamento de dados totalmente em produção, você pode optar por reduzir a depuração. Dessa forma, você pode economizar no custo do armazenamento de arquivos de log no Amazon S3 e reduzir a carga de processamento no cluster, pois ele não precisa mais gravar o estado no Amazon S3. O risco, obviamente, é que se ocorrer algum problema, você terá menos ferramentas disponíveis para investigar o erro.

Arquivos de log padrão

Por padrão, cada cluster grava os arquivos de log no nó primário. Esses são gravados no diretório `/mnt/var/log/`. Você pode acessá-los usando SSH para se conectar ao nó primário, conforme descrito emConecte-se ao nó primário usando SSH. A Amazon EMR coleta determinados registros de sistemas e aplicativos gerados pelos EMR daemons da Amazon e outros EMR processos da Amazon para garantir operações de serviço eficazes.

> **Note**
>
> Se você usa a Amazon EMR versão 6.8.0 ou anterior, os arquivos de log são salvos no Amazon S3 durante o encerramento do cluster, portanto, você não pode acessar os arquivos de log após o término do nó primário. A Amazon EMR lança a versão 6.9.0 e versões

posteriores arquivam registros no Amazon S3 durante a redução do cluster, para que os arquivos de log gerados no cluster persistam mesmo após o término do nó.

Não é necessário habilitar nada para fazer com que os arquivos de log sejam gravados no nó primário. Esse é o comportamento padrão da Amazon EMR e do Hadoop.

Um cluster gera vários tipos de arquivos de log, incluindo:

- Registros de etapas — Esses registros são gerados pelo EMR serviço da Amazon e contêm informações sobre o cluster e os resultados de cada etapa. Os arquivos de log são armazenados no diretório `/mnt/var/log/hadoop/steps/` no nó primário. Cada etapa registra seus resultados em um subdiretório numerado separado: `/mnt/var/log/hadoop/steps/s-stepId1/` para a primeira etapa, `/mnt/var/log/hadoop/steps/s-stepId2/` para a segunda etapa, e assim por diante. Os identificadores de etapa de 13 caracteres (por exemplo, stepId 1, stepId 2) são exclusivos de um cluster.
- Hadoop e registros de YARN componentes — Os registros de componentes associados ao Apache YARN e MapReduce, por exemplo, estão contidos em pastas separadas em. `/mnt/var/log` Os locais dos arquivos de log para os componentes do Hadoop sob `/mnt/var/log` são os seguintes: hadoop-hdfs, hadoop-mapreduce, hadoop-httpfs e hadoop-yarn. O hadoop-state-pusher diretório é para a saída do processo de envio de estado do Hadoop.
- Logs de ação de bootstrap: se seu trabalho utiliza ações de bootstrap, os resultados dessas ações são registrados em logs. Os arquivos de log são armazenados em /mnt/var/log/bootstrap-actions no nó primário. Cada ação de bootstrap registra seus resultados em um subdiretório numerado separado: `/mnt/var/log/bootstrap-actions/1/` para a primeira ação de bootstrap, `/mnt/var/log/bootstrap-actions/2/` para a segunda ação de bootstrap, e assim por diante.
- Registros do estado da instância — Esses registros fornecem informações sobre o CPU estado da memória e os segmentos do coletor de lixo do nó. Os arquivos de log são armazenados em `/mnt/var/log/instance-state/` no nó primário.

Arquivamento dos arquivos de log no Amazon S3

> **Note**
>
> Atualmente não é possível usar a agregação de logs para o Amazon S3 com o utilitário `yarn logs`.

A Amazon EMR lança a versão 6.9.0 e versões posteriores arquivam registros no Amazon S3 durante a redução do cluster, para que os arquivos de log gerados no cluster persistam mesmo após o término do nó. Esse comportamento é habilitado automaticamente, então não é necessário ativá-lo. Para as EMR versões 6.8.0 e anteriores da Amazon, você pode configurar um cluster para arquivar periodicamente os arquivos de log armazenados no nó primário no Amazon S3. Isso garante que os arquivos de log estarão disponíveis depois que o cluster for terminado, seja por meio de desligamento normal, seja devido a um erro. A Amazon EMR arquiva os arquivos de log no Amazon S3 em intervalos de 5 minutos.

Para que os arquivos de log sejam arquivados no Amazon S3 para as versões 6.8.0 e anteriores da EMR Amazon, você deve habilitar esse recurso ao iniciar o cluster. Você pode fazer isso usando o consoleCLI, o ou API o. Por padrão, os clusters executados por meio do console têm a funcionalidade de registro em log habilitada. Para clusters lançados usando o CLI ouAPI, o registro no Amazon S3 deve ser habilitado manualmente.

Console

Arquivar arquivos de log no Amazon S3 usando o novo console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Em Logs do cluster, marque a caixa de seleção Publicar logs específicos do cluster no Amazon S3.
4. No campo Local do Amazon S3, digite (ou navegue até) o caminho do Amazon S3 onde os logs serão armazenados. Se você digitar o nome de uma pasta que não existe no bucket, o Amazon S3 a criará.

   Quando você define esse valor, a Amazon EMR copia os arquivos de log das EC2 instâncias no cluster para o Amazon S3. Isso evita que os arquivos de log sejam perdidos quando o cluster termina e EC2 encerra as instâncias que hospedam o cluster. Esses logs são úteis para auxiliar na solução de problemas. Para obter mais informações, consulte View log files.
5. Opcionalmente, marque a caixa de seleção Criptografar logs específicos do cluster. Em seguida, selecione uma AWS KMS chave na lista, insira uma chave ARN ou crie uma nova chave. Essa opção só está disponível com a Amazon EMR versão 5.30.0 e posterior, excluindo a versão 6.0.0. Para usar essa opção, adicione permissão AWS KMS para seu

perfil de EC2 instância e sua EMR função na Amazon. Para obter mais informações, consulte Para criptografar arquivos de log armazenados no Amazon S3 com AWS KMS uma chave gerenciada pelo cliente.

6. Escolha qualquer outra opção que se aplique ao cluster.
7. Para iniciar o cluster, escolha Criar cluster.

CLI

Para arquivar arquivos de log no Amazon S3 com o AWS CLI

Para arquivar arquivos de log no Amazon S3 usando o AWS CLI, digite o `create-cluster` comando e especifique o caminho de log do Amazon S3 usando o parâmetro. `--log-uri`

1. Para registrar arquivos no Amazon S3, digite o seguinte comando e substitua *myKey* com o nome do seu EC2 key pair.

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 --log-
uri s3://DOC-EXAMPLE-BUCKET/logs --applications Name=Hadoop Name=Hive Name=Pig
 --use-default-roles --ec2-attributes KeyName=myKey --instance-type m5.xlarge --
instance-count 3
```

2. Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó primário é executado, e as instâncias restantes são executadas como nós centrais. Todos os nós usarão o tipo de instância especificado no comando.

> **Note**
>
> Se você ainda não criou a função de EMR serviço e o perfil de EC2 instância padrão da Amazon, insira `aws emr create-default-roles` para criá-los antes de digitar o `create-cluster` subcomando.

Para criptografar arquivos de log armazenados no Amazon S3 com AWS KMS uma chave gerenciada pelo cliente

Com a Amazon EMR versão 5.30.0 e posterior (exceto a Amazon EMR 6.0.0), você pode criptografar arquivos de log armazenados no Amazon S3 com uma chave gerenciada pelo cliente. AWS KMS Para habilitar essa opção no console, siga as etapas em Arquivamento dos arquivos de log no

Amazon S3. Seu perfil de EC2 instância da Amazon e sua EMR função na Amazon devem atender aos seguintes pré-requisitos:

- O perfil de EC2 instância da Amazon usado para seu cluster deve ter permissão de usokms:GenerateDataKey.
- A EMR função da Amazon usada para seu cluster deve ter permissão de usokms:DescribeKey.
- O perfil da EC2 instância da Amazon e a EMR função da Amazon devem ser adicionados à lista de usuários-chave da chave gerenciada pelo AWS KMS cliente especificada, conforme demonstrado nas etapas a seguir:
  1. Abra o console AWS Key Management Service (AWS KMS) em https://console.aws.amazon.com/kms.
  2. Para alterar a AWS região, use o seletor de região no canto superior direito da página.
  3. Selecione o alias da KMS chave a ser modificada.
  4. Na página de detalhes da chave, em Key Users (Usuários de chaves), escolha Add (Adicionar).
  5. Na caixa de diálogo Adicionar usuários-chave, selecione seu perfil de EC2 instância da Amazon e sua EMR função na Amazon.
  6. Escolha Adicionar.

Para obter mais informações, consulte as funções de IAM serviço usadas pela Amazon EMR e Como usar as principais políticas no guia do desenvolvedor do AWS Key Management Service.

Agregar logs no Amazon S3 usando a AWS CLI

> **Note**
>
> Atualmente não é possível usar a agregação de logs com o utilitário `yarn logs`. Você só pode usar a agregação compatível com esse procedimento.

A agregação de logs (Hadoop 2.x) compila os logs de todos os contêineres de um aplicativo individual em um único arquivo. Para habilitar a agregação de logs para o Amazon S3 usando AWS CLI o, você usa uma ação de bootstrap na inicialização do cluster para habilitar a agregação de logs e especificar o bucket para armazenar os logs.

- Para habilitar a agregação de logs, crie o arquivo de configuração chamado `myConfig.json`, que contém o seguinte:

```
[
  {
    "Classification": "yarn-site",
    "Properties": {
      "yarn.log-aggregation-enable": "true",
      "yarn.log-aggregation.retain-seconds": "-1",
      "yarn.nodemanager.remote-app-log-dir": "s3:\/\/DOC-EXAMPLE-BUCKET\/logs"
    }
  }
]
```

Digite o seguinte comando e substitua *myKey* com o nome do seu EC2 key pair. Além disso, você pode substituir os textos em vermelho por suas próprias configurações.

```
aws emr create-cluster --name "Test cluster" \
--release-label emr-7.2.0 \
--applications Name=Hadoop \
--use-default-roles \
--ec2-attributes KeyName=myKey \
--instance-type m5.xlarge \
--instance-count 3 \
--configurations file://./myConfig.json
```

Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó primário é executado, e as instâncias restantes são executadas como nós centrais. Todos os nós usarão o tipo de instância especificado no comando.

> **Note**
>
> Se você ainda não criou a função de EMR serviço e o perfil de EC2 instância padrão, execute `aws emr create-default-roles` para criá-los antes de executar o `create-cluster` subcomando.

Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consulte Referência de AWS CLI comandos.

Locais de log

A lista a seguir inclui todos os tipos de log e seus locais no Amazon S3. Você pode usá-los para solucionar EMR problemas da Amazon.

Logs de etapa

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/steps/<step-id>/
```

Logs de aplicações

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/containers/
```

Esse local inclui contêiner `stderr` e `stdout`, `directory.info`, `prelaunch.out` e logs `launch_container.sh`.

Logs do gerenciador de recursos

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<leader-instance-id>/
applications/hadoop-yarn/
```

Hadoop HDFS

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<all-instance-id>/
applications/hadoop-hdfs/
```

Esse local inclui NameNode, DataNode, e YARN TimelineServer registros.

Logs do gerenciador de nós

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<all-instance-id>/
applications/hadoop-yarn/
```

Logs de estado de instância

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<all-instance-id>/daemons/
instance-state/
```

Registros de EMR provisionamento da Amazon

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<leader-instance-id>/
provision-node/*
```

Logs do Hive

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<leader-instance-id>/
applications/hive/*
```

- Para encontrar logs do Hive no cluster, remova o asterisco (*) e anexe `/var/log/hive/` ao link acima.
- Para encontrar HiveServer 2 registros, remova o asterisco (*) e anexe `var/log/hive/hiveserver2.log` ao link acima.
- Para encontrar CLI os registros do Hive, remova o asterisco (*) e anexe `/var/log/hive/user/hadoop/hive.log` ao link acima.
- Para encontrar os logs do Hive Metastore Server, remova o asterisco (*) e anexe `/var/log/hive/user/hive/hive.log` ao link acima.

Se sua falha estiver no nó primário ou no nó de tarefa da aplicação Tez, forneça logs do contêiner Hadoop apropriado.

Clusters de etiqueta

Pode ser conveniente categorizar seus AWS recursos de maneiras diferentes; por exemplo, por finalidade, proprietário ou ambiente. Você pode conseguir isso na Amazon EMR atribuindo metadados personalizados aos seus EMR clusters da Amazon usando tags. Uma tag consiste em uma chave e um valor, ambos definidos por você. Para a AmazonEMR, o cluster é o nível do recurso que você pode marcar. Por exemplo, é possível definir um conjunto de tags para os clusters da sua conta que ajuda a rastrear o proprietário de cada cluster ou identificar um cluster de produção versus um cluster de teste. Recomendamos criar um conjunto consistente de tags para atender às necessidades da sua organização.

Quando você adiciona uma tag a um EMR cluster da Amazon, a tag também é propagada para cada EC2 instância ativa da Amazon associada ao cluster. Da mesma forma, quando você remove uma tag de um EMR cluster da Amazon, essa tag é removida de cada EC2 instância ativa associada da Amazon.

> **Important**
>
> Use o EMR console da Amazon ou CLI para gerenciar tags em EC2 instâncias da Amazon que fazem parte de um cluster em vez do EC2 console da Amazon ouCLI, porque as alterações que você faz na Amazon EC2 não são sincronizadas de volta com o sistema de EMR marcação da Amazon.

Você pode identificar uma EC2 instância da Amazon que faz parte de um EMR cluster da Amazon procurando as seguintes tags do sistema. Neste exemplo, *CORE* é o valor da função do grupo de instâncias e *j-12345678* é um exemplo de valor de identificador de fluxo de trabalho (cluster):

- aws: elasticmapreduce: = instance-group-role*CORE*
- aws: elasticmapreduce: = job-flow-id*j-12345678*

> **Note**
>
> A Amazon EMR e a Amazon EC2 interpretam suas tags como uma sequência de caracteres sem significado semântico.

Você pode trabalhar com tags usando o AWS Management Console CLI, o e API o.

Você pode adicionar tags ao criar um novo EMR cluster da Amazon e pode adicionar, editar ou remover tags de um EMR cluster da Amazon em execução. Editar uma tag é um conceito que se aplica ao EMR console da Amazon. No entanto, usando o CLI eAPI, para editar uma tag, você remove a tag antiga e adiciona uma nova. Você pode editar chaves de tags e valores e pode remover tags de um recurso a qualquer momento durante a execução do cluster. No entanto, não é possível adicionar, editar ou remover tags de um cluster encerrado ou de instâncias encerradas que estavam anteriormente associadas a um cluster que ainda está ativo. Além disso, você pode definir o valor de uma tag como a string vazia, mas não pode definir valor de uma tag como nulo.

Se você estiver usando AWS Identity and Access Management (IAM) com suas EC2 instâncias da Amazon para obter permissões baseadas em recursos por tag, suas IAM políticas serão aplicadas às tags que a Amazon EMR propaga para as instâncias da Amazon de um cluster. EC2 Para que as EMR tags da Amazon se propaguem para suas EC2 instâncias da Amazon, sua IAM política para a Amazon EC2 precisa permitir permissões para chamar a Amazon EC2 CreateTags e. DeleteTags APIs Além disso, as tags propagadas podem afetar as permissões baseadas em recursos EC2 da Amazon. As tags propagadas para a Amazon EC2 podem ser lidas como condições em sua IAM política, assim como outras EC2 tags da Amazon. Lembre-se IAM de sua política ao adicionar tags aos seus EMR clusters da Amazon para evitar que os usuários tenham permissões incorretas para um cluster. Para evitar problemas, certifique-se de que suas IAM políticas não incluam condições nas tags que você também planeja usar em seus EMR clusters da Amazon. Para obter mais informações, consulte Controle do acesso aos EC2 recursos da Amazon.

Restrições de tags

As restrições básicas a seguir se aplicam a tags:

- As restrições que se aplicam aos EC2 recursos da Amazon também se aplicam à AmazonEMR. Para obter mais informações, consulte https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/Using_Tags.html#tag-restrictions.
- Não use o `aws:` prefixo nos nomes e valores das tags porque ele está reservado para AWS uso. Além disso, você não pode editar nem excluir nomes ou valores de tags com esse prefixo.
- Você não pode alterar ou editar tags em um cluster encerrado.
- Um valor de tag pode ser uma string vazia, mas não nula. Além disso, uma chave de tag não pode ser uma sequência vazia.
- Chaves e valores podem conter qualquer caractere alfabético em qualquer idioma, qualquer caractere numérico, espaço em branco, separadores invisíveis e os seguintes símbolos: _ . : / = + - @

Para obter mais informações sobre a marcação usando o AWS Management Console, consulte Como trabalhar com tags no console no Guia do EC2 usuário da Amazon. Para obter mais informações sobre a marcação usando a Amazon EC2API ou a linha de comando, consulte API uma CLIvisão geral no Guia do EC2 usuário da Amazon.

Recursos de tag para faturamento

Você pode usar etiquetas para organizar sua AWS fatura para refletir sua própria estrutura de custos. Para fazer isso, inscreva-se para receber a fatura AWS da sua conta com os valores-chave da tag incluídos. Dessa forma, você pode organizar suas informações de faturamento por valores de chave de tag, para ver o custo dos seus recursos combinados. Embora a Amazon EMR e a Amazon EC2 tenham extratos de cobrança diferentes, as tags em cada cluster também são colocadas em cada instância associada para que você possa usar tags para vincular EC2 os custos relacionados da Amazon EMR e da Amazon.

Por exemplo, é possível marcar vários recursos com um nome de aplicação específico, e depois organizar suas informações de faturamento para ver o custo total daquela aplicação em vários serviços. Veja mais informações sobre alocação de recursos e uso de etiquetas no Guia do usuário do AWS Billing .

Adicionar etiquetas a um cluster

Você pode adicionar etiquetas a um cluster ao criá-lo.

Console

Adicionar etiquetas ao criar um cluster usando o novo console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Em Etiquetas, escolha Adicionar nova etiqueta. Especifique uma etiqueta no campo Chave. Opcionalmente, especifique uma etiqueta no campo Valor.
4. Escolha qualquer outra opção que se aplique ao cluster.
5. Para iniciar o cluster, escolha Criar cluster.

AWS CLI

Para adicionar tags ao criar um cluster com o AWS CLI

O exemplo a seguir demonstra como adicionar uma tag a um novo cluster usando a AWS CLI. Para adicionar tags ao criar um cluster, digite o subcomando `create-cluster` com o parâmetro `--tags`.

- Para adicionar uma tag chamada *costCenter* com valor chave *marketing* ao criar um cluster, digite o seguinte comando e substitua *myKey* com o nome do seu EC2 key pair.

```
aws emr create-cluster --name "Test cluster" --release-label emr-4.0.0 --
applications Name=Hadoop Name=Hive Name=Pig --tags "costCenter=marketing" --
use-default-roles --ec2-attributes KeyName=myKey --instance-type m5.xlarge --
instance-count 3
```

Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó principal é executado, e as instâncias restantes são executadas como nós core. Todos os nós usarão o tipo de instância especificado no comando.

> **Note**
>
> Se você ainda não criou a função de EMR serviço e o perfil de EC2 instância padrão, digite aws `emr create-default-roles` para criá-los antes de digitar o `create-cluster` subcomando.

Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consultehttps://docs.aws.amazon.com/cli/latest/reference/emr.

Você também pode adicionar tags a um cluster existente.

Console

Adicionar etiquetas a um cluster existente usando o novo console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EMREC2Em Ativado, no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar.
3. Na guia Etiquetas na página de detalhes do cluster, selecione Gerenciar etiquetas. Especifique uma etiqueta no campo Chave. Opcionalmente, especifique uma etiqueta no campo Valor.
4. Selecione Save Changes (Salvar alterações). A guia Etiquetas é atualizada com o novo número de etiquetas que você tem no cluster. Por exemplo, se você agora tem duas etiquetas, o rótulo da sua guia é Etiquetas (2).

AWS CLI

Para adicionar tags a um cluster em execução com o AWS CLI

- Digite o subcomando `add-tags` com o parâmetro `--tag` para atribuir etiquetas a um ID do cluster. Você pode localizar o ID do cluster usando o console ou o comando `list-clusters`. Atualmente, o subcomando `add-tags` aceita apenas um ID de recurso.

Por exemplo, para adicionar duas tags a um cluster em execução, uma com uma chave chamada *costCenter* com um valor de *marketing* e outro chamado *other* com um valor de *accounting*, digite o seguinte comando e substitua *j-KT4XXXXXXXX1NM* com seu ID de cluster.

```
aws emr add-tags --resource-id j-KT4XXXXXXXX1NM --tag "costCenter=marketing" --
tag "other=accounting"
```

Observe que quando as tags são adicionadas usando o AWS CLI, não há saída do comando. Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consultehttps://docs.aws.amazon.com/cli/latest/reference/emr.

Visualizar etiquetas em um cluster

Para visualizar todas as etiquetas associadas a um cluster, você pode visualizá-las no console ou na AWS CLI.

Console

Visualizar etiquetas em um cluster usando o novo console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EMREC2Em Ativado, no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar.
3. Para visualizar todas as suas etiquetas, selecione a guia Etiquetas na página de detalhes do cluster.

AWS CLI

Para visualizar as tags em um cluster com o AWS CLI

Para visualizar as tags em um cluster usando o AWS CLI, digite o `describe-cluster` subcomando com o `--query` parâmetro.

- Para visualizar as tags de um cluster, digite o seguinte comando e substitua *j-KT4XXXXXXXX1NM* com seu ID de cluster.

```
aws emr describe-cluster --cluster-id j-KT4XXXXXX1NM --query Cluster.Tags
```

A saída exibe todas as informações de tag sobre o cluster, semelhantes às seguintes:

```
Value: accounting     Value: marketing
Key: other            Key: costCenter
```

Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consultehttps://docs.aws.amazon.com/cli/latest/reference/emr.

Remover etiquetas de um cluster

Se não precisa mais de uma tag, pode removê-la do cluster.

Console

Remover etiquetas de um cluster usando o novo console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EMREC2Em Ativado, no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar.
3. Na guia Etiquetas na página de detalhes do cluster, selecione Gerenciar etiquetas.
4. Escolha Remover para cada par de chave-valor que você deseja remover.
5. Escolha Salvar alterações.

AWS CLI

Para remover tags em um cluster com o AWS CLI

Digite o subcomando `remove-tags` com o parâmetro `--tag-keys`. Ao remover uma tag, apenas o nome da chave é necessário.

- Para remover uma tag de um cluster, digite o comando a seguir e substitua *j-KT4XXXXXXXX1NM* com seu ID de cluster.

```
aws emr remove-tags --resource-id j-KT4XXXXXX1NM --tag-keys "costCenter"
```

> **Note**
>
> No momento, não é possível remover várias tags usando um único comando.

Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consultehttps://docs.aws.amazon.com/cli/latest/reference/emr.

Integração de drivers e aplicações de terceiros

Você pode executar vários aplicativos populares de big data na Amazon EMR com preços de serviços públicos. Isso significa que você paga uma taxa por hora nominal adicional pelo aplicativo de terceiros enquanto seu cluster está em execução. Isso permite que você use o aplicativo sem precisar adquirir uma licença anual. As seções a seguir descrevem algumas das ferramentas com as quais você pode usarEMR.

Tópicos

Use ferramentas de inteligência de negócios com a Amazon EMR

Você pode usar ferramentas populares de business intelligence, como Microsoft Excel, MicroStrategyQlikView, e Tableau com EMR a Amazon, para explorar e visualizar seus dados. Muitas dessas ferramentas exigem um driver ODBC (Open Database Connectivity) ou JDBC (Java Database Connectivity). Para baixar e instalar os drivers mais recentes, consulte http://awssupportdatasvcs.com/bootstrap-actions/Simba/latest/.

Para encontrar versões mais antigas dos drivers, consulte http://awssupportdatasvcs.com/bootstrap-actions/Simba/.

Segurança na Amazon EMR

Segurança e conformidade são uma responsabilidade com a qual você compartilha AWS. Esse modelo de responsabilidade compartilhada pode ajudar a aliviar sua carga operacional, pois AWS opera, gerencia e controla os componentes do sistema operacional host e da camada de virtualização até a segurança física das instalações nas quais os EMR clusters operam. Você assume a responsabilidade, o gerenciamento e a atualização dos EMR clusters da Amazon, bem como configura o software do aplicativo e os controles de segurança AWS fornecidos. Essa diferenciação de responsabilidade é comumente chamada de segurança na nuvem versus segurança na nuvem.

- Segurança da nuvem — AWS é responsável por proteger a infraestrutura que é Serviços da AWS executada AWS. AWS também fornece serviços que você pode usar com segurança. Auditores de terceiros testam e verificam regularmente a eficácia da nossa segurança como parte dos programas de conformidade da AWS. Para saber mais sobre os programas de conformidade que se aplicam à AmazonEMR, consulte Serviços da AWS o escopo por programa de conformidade.
- Segurança na nuvem — você também é responsável por realizar todas as tarefas de configuração e gerenciamento de segurança necessárias para proteger um EMR cluster da Amazon. Os clientes que implantam um EMR cluster da Amazon são responsáveis pelo gerenciamento do software aplicativo instalado nas instâncias e pela configuração dos recursos AWS fornecidos, como grupos de segurança, criptografia e controle de acesso, de acordo com seus requisitos, leis e regulamentos aplicáveis.

Essa documentação ajuda você a entender como aplicar o modelo de responsabilidade compartilhada ao usar a AmazonEMR. Os tópicos deste capítulo mostram como configurar a Amazon EMR e usar outras Serviços da AWS para atender aos seus objetivos de segurança e conformidade.

Segurança de rede e infraestrutura

Como um serviço gerenciado, a Amazon EMR é protegida pelos procedimentos AWS globais de segurança de rede descritos no whitepaper Amazon Web Services: Visão geral dos processos de segurança. AWS os serviços de proteção de rede e infraestrutura oferecem proteções refinadas nos limites do host e da rede. Recursos de EMR suporte Serviços da AWS e aplicativos da Amazon que atendem aos requisitos de conformidade e proteção de rede.

- Os grupos EC2 de segurança da Amazon atuam como um firewall virtual para instâncias de EMR cluster da Amazon, limitando o tráfego de entrada e saída da rede. Para obter mais informações, consulte Controlar o tráfego de rede com grupos de segurança.
- O Amazon EMR block public access (BPA) impede que você lance um cluster em uma sub-rede pública se o cluster tiver uma configuração de segurança que permita tráfego de entrada de endereços IP públicos em uma porta. Para obter mais informações, consulte Como usar a Amazon para EMR bloquear o acesso público.
- O Secure Shell (SSH) ajuda a fornecer uma forma segura para os usuários se conectarem à linha de comando em instâncias de cluster. Você também pode usar SSH para visualizar interfaces da web que os aplicativos hospedam no nó principal de um cluster. Para obter mais informações, consulte Usar um par de EC2 chaves para SSH credenciais e Conectar-se a um cluster.

Atualizações para o Amazon Linux padrão AMI para Amazon EMR

> **Important**
>
> EMRclusters que executam Amazon Linux ou Amazon Linux 2 Amazon Machine Images (AMIs) usam o comportamento padrão do Amazon Linux e não baixam e instalam automaticamente atualizações importantes e críticas do kernel que exigem uma reinicialização. Esse é o mesmo comportamento de outras EC2 instâncias da Amazon que executam o Amazon Linux padrãoAMI. Se novas atualizações de software Amazon Linux que exigem uma reinicialização (como kernel e CUDA atualizações) ficarem disponíveis após a disponibilização de uma EMR versão da Amazon, as instâncias de EMR cluster que executam o padrão AMI não baixam e instalam automaticamente essas atualizações. NVIDIA Para obter atualizações do kernel, você pode personalizar sua Amazon EMR AMI para usar o Amazon Linux AMI mais recente.

Dependendo da postura de segurança de seu aplicativo e o período em que um cluster é executado, você pode optar por reinicializar periodicamente seu cluster para aplicar atualizações de segurança, ou criar uma ação de bootstrap para personalizar a instalação de pacotes e atualizações. Você também pode escolher testar e, em seguida, instalar determinadas atualizações de segurança nas instâncias de cluster em execução. Para obter mais informações, consulte Usando o Amazon Linux padrão AMI para Amazon EMR. Observe que sua configuração de rede deve permitir a HTTPS entrada HTTP e a saída para repositórios Linux no Amazon S3, caso contrário, as atualizações de segurança não serão bem-sucedidas.

AWS Identity and Access Management com a Amazon EMR

AWS Identity and Access Management (IAM) é um AWS serviço que ajuda o administrador a controlar com segurança o acesso aos AWS recursos. IAMos administradores controlam quem pode ser autenticado (conectado) e autorizado (tem permissões) para usar os recursos da AmazonEMR. IAMas identidades incluem usuários, grupos e funções. Uma IAM função é semelhante à de um IAM usuário, mas não está associada a uma pessoa específica e deve ser assumida por qualquer usuário que precise de permissões. Para obter mais informações, consulte AWS Identity and Access Management para a Amazon EMR. A Amazon EMR usa várias IAM funções para ajudar você a implementar controles de acesso para EMR clusters da Amazon. IAMé um AWS serviço que você pode usar sem custo adicional.

- IAMfunção para a Amazon EMR (EMRfunção) — controla como o EMR serviço da Amazon é capaz de acessar outros Serviços da AWS em seu nome, como provisionar EC2 instâncias da Amazon quando o EMR cluster da Amazon é lançado. Para obter mais informações, consulte Configurar funções de IAM serviço para EMR permissões Serviços da AWS e recursos da Amazon.
- IAMfunção para EC2 instâncias de cluster (perfil de EC2 instância) — uma função que é atribuída a cada EC2 instância no EMR cluster da Amazon quando a instância é iniciada. Os processos de aplicativos executados no cluster usam essa função para interagir com outros Serviços da AWS, como o Amazon S3. Para obter mais informações, consulte IAMpapel para EC2 instâncias do cluster.
- IAMfunção para aplicativos (função de tempo de execução) — uma IAM função que você pode especificar ao enviar um trabalho ou uma consulta para um EMR cluster da Amazon. O trabalho ou consulta que você envia ao seu EMR cluster da Amazon usa a função de tempo de execução para acessar AWS recursos, como objetos no Amazon S3. Você pode especificar funções de tempo de execução com a Amazon EMR para trabalhos do Spark e do Hive. Ao usar funções de tempo de execução, você pode isolar trabalhos em execução no mesmo cluster usando IAM funções diferentes. Para obter mais informações, consulte Usando a IAM função como função de tempo de execução com a Amazon EMR.

As identidades da força de trabalho se referem aos usuários que criam ou operam cargas de trabalho em. AWS A Amazon EMR fornece suporte para identidades da força de trabalho com o seguinte:

- AWS IAMo centro de identidade (Idc) é o recomendado Serviço da AWS para gerenciar o acesso do usuário aos AWS recursos. É um único local onde você pode atribuir identidades à sua força de

trabalho e acesso consistente a várias AWS contas e aplicativos. A Amazon EMR oferece suporte às identidades da força de trabalho por meio da propagação confiável de identidades. Com um recurso confiável de propagação de identidade, um usuário pode entrar no aplicativo e esse aplicativo pode passar a identidade do usuário Serviços da AWS para outra pessoa para autorizar o acesso a dados ou recursos. Para obter mais informações, consulte Habilitando o suporte para o centro de AWS IAM identidade com a Amazon EMR.

O Lightweight Directory Access Protocol (LDAP) é um protocolo de aplicativo aberto, independente de fornecedor e padrão do setor para acessar e manter informações sobre usuários, sistemas, serviços e aplicativos na rede. LDAPé comumente usado para autenticação de usuários em servidores de identidade corporativa, como Active Directory (AD) e OpenLDAP. Ao habilitar LDAP com EMR clusters, você permite que os usuários usem suas credenciais existentes para autenticar e acessar clusters. Para obter mais informações, consulte Ativar o suporte para LDAP com a Amazon EMR.

O Kerberos é um protocolo de autenticação de rede projetado para fornecer autenticação forte para aplicativos cliente/servidor usando criptografia de chave secreta. Quando você usa o Kerberos, a Amazon EMR configura o Kerberos para os aplicativos, componentes e subsistemas que ele instala no cluster para que eles sejam autenticados entre si. Para acessar um cluster com o Kerberos configurado, um kerberos principal deve estar presente no controlador de domínio Kerberos (). KDC Para obter mais informações, consulte Como ativar o suporte para Kerberos com a Amazon. EMR

Clusters de inquilino único e multilocatário

Por padrão, um cluster é configurado para uma única locação com o perfil da EC2 instância como IAM identidade. Em um cluster de inquilino único, cada trabalho tem acesso total e completo ao cluster e o acesso a todos os Serviços da AWS recursos é feito com base no perfil da EC2 instância. Em um cluster multilocatário, os inquilinos são isolados uns dos outros e não têm acesso total e completo aos clusters e às EC2 instâncias do cluster. A identidade em clusters multilocatários são as funções de tempo de execução ou as identificações da força de trabalho. Em um cluster multilocatário, você também pode ativar o suporte para controle de acesso refinado (FGAC) por meio do Apache Ranger. AWS Lake Formation Em um cluster com funções de tempo de execução ou FGAC habilitadas, o acesso ao perfil da EC2 instância também é desativado por meio de iptables.

> **Important**
>
> Qualquer usuário que tenha acesso a um cluster de locatário único pode instalar qualquer software no sistema operacional (SO) Linux, alterar ou remover componentes de software instalados pela Amazon EMR e impactar as EC2 instâncias que fazem parte do cluster. Se você quiser garantir que os usuários não possam instalar ou alterar as configurações de um EMR cluster da Amazon, recomendamos que você habilite a multilocação para o cluster. Você pode habilitar a multilocação em um cluster habilitando o suporte para função de tempo de execução, centro de AWS IAM identidade, Kerberos ou. LDAP

Proteção de dados

Com AWS, você controla seus dados usando Serviços da AWS ferramentas para determinar como os dados são protegidos e quem tem acesso a eles. Serviços como AWS Identity and Access Management (IAM) permitem que você gerencie com segurança o acesso Serviços da AWS e os recursos. AWS CloudTrail permite a detecção e a auditoria. A Amazon EMR facilita a criptografia de dados em repouso no Amazon S3 usando chaves gerenciadas por você ou totalmente gerenciadas AWS por você. A Amazon EMR também oferece suporte para habilitar a criptografia de dados em trânsito. Para obter mais informações, consulte criptografar dados em repouso e em trânsito.

Controle de acesso a dados

Com o controle de acesso aos dados, você pode controlar quais dados uma IAM identidade ou uma identidade da força de trabalho pode acessar. A Amazon EMR oferece suporte aos seguintes controles de acesso:

- IAMpolíticas baseadas em identidade — gerencie permissões para IAM funções que você usa na Amazon. EMR IAMas políticas podem ser combinadas com a marcação para controlar o acesso em uma cluster-by-cluster base. Para obter mais informações, consulte AWS Identity and Access Management para a Amazon EMR.
- AWS Lake Formationcentraliza o gerenciamento de permissões de seus dados e facilita o compartilhamento em toda a organização e externamente. Você pode usar o Lake Formation para permitir acesso refinado em nível de coluna a bancos de dados e tabelas no Glue Data Catalog. AWS Para obter mais informações, consulte Usando AWS Lake Formation com a Amazon EMR.
- O acesso ao Amazon S3 concede identidades de mapas e identidades de mapas em diretórios como o Active Directory, ou AWS Identity and Access Management (IAM) principals, para

conjuntos de dados no S3. Além disso, o acesso ao S3 concede ao log a identidade do usuário final e o aplicativo usado para acessar os dados do S3. AWS CloudTrail Para obter mais informações, consulte Usando concessões de acesso do Amazon S3 com a Amazon. EMR

- O Apache Ranger é uma estrutura para habilitar, monitorar e gerenciar a segurança abrangente de dados em toda a plataforma Hadoop. A Amazon EMR oferece suporte ao controle de acesso refinado baseado no Apache Ranger para o Apache Hive Metastore e o Amazon S3. Para obter mais informações, consulte Integrar o Apache Ranger com a Amazon. EMR

Usar configurações de segurança para definir a segurança do cluster

Você pode usar as configurações EMR de segurança da Amazon para configurar a criptografia de dados, a autenticação Kerberos e a autorização do Amazon S3 em seus clusters. EMRFS Primeiro, crie uma configuração de segurança. Em seguida, a configuração de segurança fica disponível para uso e reutilização ao criar clusters.

Você pode usar o AWS Management Console, o AWS Command Line Interface (AWS CLI) ou o AWS SDKs para criar configurações de segurança. Você também pode usar um AWS CloudFormation modelo para criar uma configuração de segurança. Para obter mais informações, consulte o Guia AWS CloudFormation do usuário e a referência do modelo para AWS:EMR:: SecurityConfiguration.

Tópicos

- Criar uma configuração de segurança
- Especificar uma configuração de segurança para um cluster

Criar uma configuração de segurança

Este tópico aborda os procedimentos gerais para criar uma configuração de segurança com o EMR console da Amazon e o AWS CLI, seguido por uma referência para os parâmetros que incluem criptografia, autenticação e IAM funções paraEMRFS. Para obter mais informações sobre esses recursos, consulte os tópicos a seguir:

- Criptografar dados em repouso e em trânsito
- Use o Kerberos para autenticação com a Amazon EMR
- Configurar IAM funções para EMRFS solicitações para o Amazon S3

Para criar uma configuração de segurança usando o console

1. Abra o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. No painel de navegação, escolha Security Configurations (Configurações de segurança), Create security configuration (Criar configuração de segurança).
3. Digite um nome em Name (Nome) para a configuração de segurança.
4. Escolha opções Criptografia e Autenticação conforme descrito nas seções abaixo e escolha Criar.

Para criar uma configuração de segurança usando o AWS CLI

- Use o comando `create-security-configuration` conforme mostrado no exemplo a seguir.
  - Para *`SecConfigName`*, especifique o nome da configuração de segurança. Trata-se do nome especificado por você ao criar um cluster que usa essa configuração de segurança.
  - Para*`SecConfigDef`*, especifique uma JSON estrutura embutida ou o caminho para um JSON arquivo local, como*`file://MySecConfig.json`*. Os JSON parâmetros definem opções de criptografia, IAMfunções para EMRFS acesso ao Amazon S3 e autenticação, conforme descrito nas seções abaixo.

```
aws emr create-security-configuration --name "SecConfigName" --security-
configuration SecConfigDef
```

Configurar criptografia de dados

Antes de configurar a criptografia em uma configuração de segurança, crie as chaves e os certificados usados na criptografia. Para ter mais informações, consulte Fornecendo chaves para criptografar dados em repouso com a Amazon EMR e Fornecendo certificados para criptografar dados em trânsito com a criptografia da Amazon EMR.

Ao criar uma configuração de segurança, você especifica dois conjuntos de opções de criptografia: a criptografia de dados em repouso e a criptografia de dados em trânsito. As opções para criptografia de dados em repouso incluem o Amazon S3 EMRFS com criptografia de disco local e a criptografia de disco local. As opções de criptografia em trânsito habilitam os recursos de criptografia de código aberto para determinados aplicativos que oferecem suporte ao Transport Layer Security ()TLS.

Opções em repouso e opções em trânsito podem ser habilitadas juntas ou separadamente. Para obter mais informações, consulte Criptografar dados em repouso e em trânsito.

> **Note**
>
> Quando você usa AWS KMS, cobranças são cobradas pelo armazenamento e uso de chaves de criptografia. Para obter mais informações, consulte Preços do AWS KMS.

Especificar opções de criptografia usando o console

Escolha as opções em Encryption (Criptografia) de acordo com as diretrizes a seguir.

- Escolha opções em At rest encryption (Criptografia em repouso) para criptografar os dados armazenados no sistema de arquivos.

  Você pode optar por criptografar dados no Amazon S3, em discos locais ou em ambos.
- Em Criptografia de dados do S3, em Modo de criptografia, escolha um valor para determinar como a Amazon EMR criptografa os dados do Amazon S3. EMRFS

  O que fazer em seguida depende do modo de criptografia escolhido:
  - SSE-S3

    Especifica a criptografia do lado do servidor com chaves de criptografia gerenciadas pelo Amazon S3. Você não precisa fazer mais nada, pois o Amazon S3 manipula as chaves para você.
  - SSE- KMS ou CSE- KMS

    Especifica a criptografia do lado do servidor com chaves AWS KMS gerenciadas (SSE-KMS) ou a criptografia do lado do cliente com chaves gerenciadas (-). AWS KMS CSE KMS Em AWS KMS key, selecione uma chave. A chave deve existir na mesma região do seu EMR cluster. Para conhecer os requisitos de chaves, consulte Usando AWS KMS keys para criptografia.
  - CSE-Personalizado

    Especifica a criptografia do lado do cliente usando uma chave raiz personalizada do lado do cliente (-custom). CSE Para o objeto S3, insira a localização no Amazon S3, ou no Amazon S3, do seu arquivo personalizado de ARN provedor de chaves. JAR Em seguida, em Key provider class, insira o nome completo da classe declarada em seu aplicativo que implementa a EncryptionMaterialsProvider interface.

- Em Local disk encryption (Criptografia de disco local), escolha um valor para Key provider type (Tipo de provedor de chave).
  - AWS KMS key

    Selecione essa opção para especificar uma AWS KMS key. Em AWS KMS key, selecione uma chave. A chave deve existir na mesma região do seu EMR cluster. Para obter mais informações sobre requisitos de chaves, consulte Usando AWS KMS keys para criptografia.

    EBSCriptografia

    Ao especificar AWS KMS como seu provedor de chaves, você pode ativar a EBS criptografia para criptografar o dispositivo EBS raiz e os volumes de armazenamento. Para habilitar essa opção, você deve conceder à função EMR `EMR_DefaultRole` de serviço da Amazon permissões para usar o AWS KMS key que você especificar. Para obter mais informações sobre requisitos de chaves, consulte Ativando a EBS criptografia fornecendo permissões adicionais para KMS chaves.
  - Custom (Personalizado)

    Selecione essa opção para especificar um provedor de chaves personalizado. Para o objeto S3, insira a localização no Amazon S3, ou no Amazon S3, do seu arquivo personalizado de ARN provedor de chaves. JAR Em Key provider class, insira o nome completo da classe declarada em seu aplicativo que implementa a EncryptionMaterialsProvider interface. O nome da classe que você fornece aqui deve ser diferente do nome da classe fornecido para CSE -Custom.
- Escolha a criptografia em trânsito para ativar os recursos de TLS criptografia de código aberto para dados em trânsito. Escolha um tipo de provedor certificado em Certificate provider type (Tipo de provedor de certificados), de acordo com as seguintes diretrizes:
  - PEM

    Selecione essa opção para usar PEM os arquivos que você fornece em um arquivo zip. Dois artefatos são necessários no arquivo zip: privateKey .pem e .pem. certificateChain Um terceiro arquivo, trustedCertificates .pem, é opcional. Para mais detalhes, consulte Fornecendo certificados para criptografar dados em trânsito com a criptografia da Amazon EMR. Para o objeto S3, especifique a localização no Amazon S3, ou no Amazon ARN S3, do campo do arquivo zip.
  - Custom (Personalizado)

Selecione essa opção para especificar um provedor de certificado personalizado e, em seguida, para o objeto S3, insira a localização no Amazon S3, ou no Amazon S3, do seu arquivo de provedor de certificado personalizado. ARN JAR Em Key provider class, insira o nome completo da classe declarada em seu aplicativo que implementa a TLSArtifactsProvider interface.

Especificando as opções de criptografia usando o AWS CLI

As seções a seguir usam exemplos de cenários para ilustrar o formato adequado --security-configuration JSON para diferentes configurações e provedores de chaves, seguidos por uma referência para os JSON parâmetros e valores apropriados.

Exemplo de opções de criptografia de dados em trânsito

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está habilitada, e a criptografia de dados em repouso está desabilitada.
- Um arquivo zip com certificados no Amazon S3 é usado como o provedor de chaves (consulte Fornecendo certificados para criptografar dados em trânsito com a criptografia da Amazon EMR para conhecer os requisitos de certificados).

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": true,
  "EnableAtRestEncryption": false,
  "InTransitEncryptionConfiguration": {
   "TLSCertificateConfiguration": {
    "CertificateProviderType": "PEM",
    "S3Object": "s3://MyConfigStore/artifacts/MyCerts.zip"
   }
  }
 }
}'
```

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está habilitada, e a criptografia de dados em repouso está desabilitada.
- Um provedor de chaves personalizado é usado (consulte Fornecendo certificados para criptografar dados em trânsito com a criptografia da Amazon EMR para conhecer os requisitos de certificados).

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": true,
  "EnableAtRestEncryption": false,
  "InTransitEncryptionConfiguration": {
   "TLSCertificateConfiguration": {
    "CertificateProviderType": "Custom",
    "S3Object": "s3://MyConfig/artifacts/MyCerts.jar",
    "CertificateProviderClass": "com.mycompany.MyCertProvider"
   }
  }
  }
}'
```

Exemplo de opções de criptografia de dados em repouso

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está desabilitada, e a criptografia de dados em repouso está habilitada.
- SSE-S3 é usado para criptografia Amazon S3.
- A criptografia de disco local é usada AWS KMS como provedor de chaves.

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": false,
  "EnableAtRestEncryption": true,
  "AtRestEncryptionConfiguration": {
   "S3EncryptionConfiguration": {
    "EncryptionMode": "SSE-S3"
   },
   "LocalDiskEncryptionConfiguration": {
    "EncryptionKeyProviderType": "AwsKms",
```

```
    "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
   }
  }
  }
}'
```

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está habilitada e faz referência a um arquivo zip com PEM certificados no Amazon S3, usando o. ARN
- SSE- KMS é usado para criptografia do Amazon S3.
- A criptografia de disco local é usada AWS KMS como provedor de chaves.

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": true,
  "EnableAtRestEncryption": true,
  "InTransitEncryptionConfiguration": {
   "TLSCertificateConfiguration": {
    "CertificateProviderType": "PEM",
    "S3Object": "arn:aws:s3:::MyConfigStore/artifacts/MyCerts.zip"
   }
  },
  "AtRestEncryptionConfiguration": {
   "S3EncryptionConfiguration": {
    "EncryptionMode": "SSE-KMS",
    "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
   },
   "LocalDiskEncryptionConfiguration": {
    "EncryptionKeyProviderType": "AwsKms",
    "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
   }
  }
 }
}'
```

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está habilitada e faz referência a um arquivo zip com PEM certificados no Amazon S3.
- CSE- KMS é usado para criptografia do Amazon S3.
- A criptografia de disco local usa um provedor de chave personalizado referenciado por itsARN.

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": true,
  "EnableAtRestEncryption": true,
  "InTransitEncryptionConfiguration": {
   "TLSCertificateConfiguration": {
    "CertificateProviderType": "PEM",
    "S3Object": "s3://MyConfigStore/artifacts/MyCerts.zip"
   }
  },
  "AtRestEncryptionConfiguration": {
   "S3EncryptionConfiguration": {
    "EncryptionMode": "CSE-KMS",
    "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
   },
   "LocalDiskEncryptionConfiguration": {
    "EncryptionKeyProviderType": "Custom",
    "S3Object": "arn:aws:s3:::artifacts/MyKeyProvider.jar",
    "EncryptionKeyProviderClass": "com.mycompany.MyKeyProvider"
   }
  }
 }
}'
```

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está habilitada com um provedor de chaves personalizado.
- CSE-Custom é usado para dados do Amazon S3.
- A criptografia do disco local usa um provedor de chaves personalizado.

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
```

```
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": "true",
  "EnableAtRestEncryption": "true",
  "InTransitEncryptionConfiguration": {
   "TLSCertificateConfiguration": {
    "CertificateProviderType": "Custom",
    "S3Object": "s3://MyConfig/artifacts/MyCerts.jar",
    "CertificateProviderClass": "com.mycompany.MyCertProvider"
   }
  },
  "AtRestEncryptionConfiguration": {
   "S3EncryptionConfiguration": {
    "EncryptionMode": "CSE-Custom",
    "S3Object": "s3://MyConfig/artifacts/MyCerts.jar",
    "EncryptionKeyProviderClass": "com.mycompany.MyKeyProvider"
    },
   "LocalDiskEncryptionConfiguration": {
    "EncryptionKeyProviderType": "Custom",
    "S3Object": "s3://MyConfig/artifacts/MyCerts.jar",
    "EncryptionKeyProviderClass": "com.mycompany.MyKeyProvider"
   }
  }
 }
}'
```

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está desabilitada, e a criptografia de dados em repouso está habilitada.
- A criptografia do Amazon S3 está habilitada com SSE -. KMS
- Várias AWS KMS chaves são usadas, uma por cada bucket do S3, e exceções de criptografia são aplicadas a esses buckets individuais do S3.
- A criptografia do disco local está desabilitada.

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
   "EncryptionConfiguration": {
     "AtRestEncryptionConfiguration": {
             "S3EncryptionConfiguration": {
           "EncryptionMode": "SSE-KMS",
           "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012",
```

```
           "Overrides": [
              {
                "BucketName": "sse-s3-bucket-name",
                "EncryptionMode": "SSE-S3"
              },
              {
                "BucketName": "cse-kms-bucket-name",
                "EncryptionMode": "CSE-KMS",
                "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
              },
              {
                "BucketName": "sse-kms-bucket-name",
                "EncryptionMode": "SSE-KMS",
                "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
              }
             ]
             }
           },
     "EnableInTransitEncryption": false,
      "EnableAtRestEncryption": true
  }
}'
```

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está desabilitada, e a criptografia de dados em repouso está habilitada.
- A criptografia do Amazon S3 está habilitada com SSE -S3 e a criptografia de disco local está desativada.

```
aws emr create-security-configuration --name "MyS3EncryptionConfig" --security-
configuration '{
    "EncryptionConfiguration": {
        "EnableInTransitEncryption": false,
        "EnableAtRestEncryption": true,
        "AtRestEncryptionConfiguration": {
            "S3EncryptionConfiguration": {
                "EncryptionMode": "SSE-S3"
            }
```

```
        }
     }
}'
```

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está desabilitada, e a criptografia de dados em repouso está habilitada.
- A criptografia de disco local é ativada AWS KMS como provedor de chaves e a criptografia do Amazon S3 está desativada.

```
aws emr create-security-configuration --name "MyLocalDiskEncryptionConfig" --security-
configuration '{
    "EncryptionConfiguration": {
        "EnableInTransitEncryption": false,
        "EnableAtRestEncryption": true,
        "AtRestEncryptionConfiguration": {
            "LocalDiskEncryptionConfiguration": {
                "EncryptionKeyProviderType": "AwsKms",
                "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
            }
        }
     }
}'
```

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está desabilitada, e a criptografia de dados em repouso está habilitada.
- A criptografia de disco local é ativada AWS KMS como provedor de chaves e a criptografia do Amazon S3 está desativada.
- EBSa criptografia está ativada.

```
aws emr create-security-configuration --name "MyLocalDiskEncryptionConfig" --security-
configuration '{
    "EncryptionConfiguration": {
        "EnableInTransitEncryption": false,
        "EnableAtRestEncryption": true,
```

```
        "AtRestEncryptionConfiguration": {
            "LocalDiskEncryptionConfiguration": {
                "EnableEbsEncryption": true,
                "EncryptionKeyProviderType": "AwsKms",
                "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
            }
        }
     }
}'
```

O exemplo abaixo ilustra o seguinte cenário:

SSE- EMR - WAL é usado para EMR WAL criptografia

```
aws emr create-security-configuration --name "MySecConfig" \
    --security-configuration '{
        "EncryptionConfiguration": {
            "EMRWALEncryptionConfiguration":{ },
            "EnableInTransitEncryption":false, "EnableAtRestEncryption":false
        }
    }'
```

`EnableInTransitEncryption`e `EnableAtRestEncryption` ainda pode ser verdade, se quiser habilitar a criptografia relacionada.

O exemplo abaixo ilustra o seguinte cenário:

- SSE- KMS - WAL é usado para EMR WAL criptografia
- A criptografia do lado do servidor é usada AWS Key Management Service como provedor principal

```
aws emr create-security-configuration --name "MySecConfig" \
    --security-configuration '{
        "EncryptionConfiguration": {
            "EMRWALEncryptionConfiguration":{
                "AwsKmsKey":"arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
                },
            "EnableInTransitEncryption":false, "EnableAtRestEncryption":false
        }
    }'
```

`EnableInTransitEncryption`e `EnableAtRestEncryption` ainda pode ser verdade, se quiser habilitar a criptografia relacionada.

JSONreferência para configurações de criptografia

A tabela a seguir lista os JSON parâmetros das configurações de criptografia e fornece uma descrição dos valores aceitáveis para cada parâmetro.

Parâmetro	Descrição
`"EnableInTransitEncryption" : true | false`	Especifique `true` para habilitar a criptografia em trânsito e `false` para desabilitá-la. Se omitido, `false` é assumido, e a criptografia em trânsito é desabilitada.
`"EnableAtRestEncryption": true | false`	Especifique `true` para habilitar a criptografia em repouso e `false` para desabilitá-la. Se omitido, `false` é assumido, e a criptografia em repouso é desabilitada.
Parâmetros de criptografia em trânsito	
`"InTransitEncryptionConfiguration" :`	Especifica uma coleção de valores usados para configurar a criptografia em trânsito quando `EnableInTransitEncryption` é `true`.
`"CertificateProviderType": "PEM" | "Custom"`	Especifica se é necessário usar `PEM`certificados referenciados com um arquivo compactado em zip ou um provedor de certificados `Custom`. Se `PEM` for especificado, `S3Object` deve ser uma referência à localização no Amazon S3 de um arquivo zip contendo os certificados. Se Personalizado for especificado, `S3Object` deve ser uma referência à localização de um JAR arquivo no Amazon S3, seguida por uma `CertificateProviderClass` entrada.
`"S3Object" : " ZipLocation " | "JarLocation "`	Fornece a localização no Amazon S3 para um arquivo zip quando `PEM` especificado

Parâmetro	Descrição
	ou para um JAR arquivo quando `Custom` especificado. O formato pode ser um caminho (por exemplo,`s3://MyConfig/artifacts/CertFiles.zip`) ou um ARN (por exemplo,`arn:aws:s3:::Code/MyCertProvider.jar)` . Se for especificado um arquivo zip, ele deverá conter arquivos exatamente denominados `privateKey.pem` e `certificateChain.pem` . Um arquivo denominado `trustedCertificates.pem` é opcional.
`"CertificateProviderClass" : "`*`MyClassID`*` "`	Obrigatório somente se `Custom` for especificado para`CertificateProviderType` . *`MyClassID`* especifica um nome de classe completo declarado no JAR arquivo, que implementa a TLSArtifactsProvider interface . Por exemplo, `com.mycompany.MyCertProvider` .
Parâmetros de criptografia em repouso	
`"AtRestEncryptionConfiguration" :`	Especifica uma coleção de valores para criptografia em repouso quando `EnableAtRestEncryption` estiver`true`, incluindo criptografia Amazon S3 e criptografia de disco local.
Parâmetros de criptografia do Amazon S3	
`"S3EncryptionConfiguration" :`	Especifica uma coleção de valores usados para a criptografia do Amazon S3 com o EMR Amazon File System EMRFS ().

Parâmetro	Descrição
`"EncryptionMode" : "SSE-S3" | "SSE-KMS" | "CSE-KMS" | "CSE-Custom"`	Especifica o tipo de criptografia do Amazon S3 a ser usada. Se `SSE-S3` for especificado, nenhum valor adicional de criptografia do Amazon S3 será necessário. Se um `SSE-KMS` ou `CSE-KMS` for especificado, an AWS KMS key ARN deverá ser especificado como o `AwsKmsKey` valor. Se `CSE-Custom` for especificado, os valores `S3Object` e `EncryptionKeyProviderClass` deverão ser especificados.
`"AwsKmsKey" : "` *`MyKeyARN`*`"`	Obrigatório somente quando um `SSE-KMS` ou `CSE-KMS` está especificado para`EncryptionMode` . *`MyKeyARN`*deve ser totalmente especificado ARN para uma chave (por exemplo,`arn:aws:kms:us-east-1:123456789012:key/12345678-1234-1234-1234-123456789012`).
`"S3Object" : "`*`JarLocation`*` "`	Obrigatório somente quando `CSE-Custom` é especificado para`CertificateProviderType` . *`JarLocation`* fornece a localização no Amazon S3 para um JAR arquivo. O formato pode ser um caminho (por exemplo,`s3://MyConfig/artifacts/MyKeyProvider.jar`) ou um ARN (por exemplo,`arn:aws:s3:::Code/MyKeyProvider.jar)` .
`"EncryptionKeyProviderClass" : "`*`MyS3KeyClassID`*` "`	Obrigatório somente quando `CSE-Custom` é especificado para`EncryptionMode` . *`MyS3KeyClassID`* especifica o nome completo de uma classe declarada no aplicativo que implementa a EncryptionMaterialsProvider interface; por exemplo,. *`com.mycompany.MyS3KeyProvider`*

Parâmetro	Descrição
Parâmetros de criptografia do disco local	
`"LocalDiskEncryptionConfiguration"`	Especifica o provedor de chaves e os valores correspondentes a serem usados para criptografia do disco local.
`"EnableEbsEncryption": true | false`	Especifique `true` para ativar a EBS criptografia. EBSa criptografia criptografa o volume do dispositivo EBS raiz e os volumes de armazenamento conectados. Para usar a EBS criptografia, você deve especificar `AwsKms` como seu`EncryptionKeyProviderType` .
`"EncryptionKeyProviderType": "AwsKms" | "Custom"`	Especifica o provedor de chaves. Se `AwsKms` for especificado, uma KMS chave ARN deverá ser especificada como `AwsKmsKey` valor. Se **`Custom`** for especificado, os valores `S3Object` e `EncryptionKeyProviderClass` deverão ser especificados.
`"AwsKmsKey : " MyKeyARN"`	Obrigatório somente quando `AwsKms` é especificado para`Type`. *`MyKeyARN`*deve ser totalmente especificado ARN para uma chave (por exemplo,`arn:aws:kms:us-east-1:123456789012:key/12345678-1234-1234-1234-456789012123`).
`"S3Object" : "JarLocation "`	Obrigatório somente quando `CSE-Custom` é especificado para`CertificateProviderType` . *`JarLocation`* fornece a localização no Amazon S3 para um JAR arquivo. O formato pode ser um caminho (por exemplo,`s3://MyConfig/artifacts/MyKeyProvider.jar`) ou um ARN (por exemplo,`arn:aws:s3:::Code/MyKeyProvider.jar)` .

Parâmetro	Descrição
`"EncryptionKeyProviderClass" : "MyLocalDiskKeyClassID "`	Obrigatório somente quando `Custom` é especificado para`Type`. *`MyLocalDiskKeyClassID`* especifica o nome completo de uma classe declarada no aplicativo que implementa a EncryptionMaterialsProvider interface; por exemplo,. *`com.mycompany.MyLocalDiskKeyProvider`*
EMRWALparâmetros de criptografia	
`"EMRWALEncryptionConfiguration"`	Especifica o valor da EMR WAL criptografia.
`"AwsKmsKey"`	Especifica a ID da CMK chave Arn.

Configurar a autenticação Kerberos

Uma configuração de segurança com definições Kerberos só pode ser usada por um cluster criado com atributos Kerberos, ou ocorrerá um erro. Para obter mais informações, consulte Use o Kerberos para autenticação com a Amazon EMR. O Kerberos só está disponível na EMR versão 5.10.0 e posterior da Amazon.

Especificar configurações do Kerberos usando o console

Escolha opções em Kerberos authentication (Autenticação Kerberos) de acordo com as diretrizes a seguir.

Parâmetro	Descrição
Kerberos	Especifica que o Kerberos está habilitado em clusters que usam essa configuração de segurança. Ao usar essa configuração de segurança, o cluster também deverá ter configurações Kerberos especificadas ou ocorrerá um erro.

Parâmetro		Descrição
Provedor	Dedicado ao cluster KDC	Especifica que a Amazon EMR cria um KDC no nó primário de qualquer cluster que usa essa configura ção de segurança. Você especifica o nome do territóri o e a senha KDC do administrador ao criar o cluster. Você pode referenciar isso KDC em outros clusters, se necessário. Crie esses clusters usando uma configura ção de segurança diferente, especifique uma externa KDC e use o nome do território e a senha de KDC administrador que você especifica para o cluster KDC dedicado.
	Externo KDC	Disponível somente com o Amazon EMR 5.20.0 e versões posteriores. Especifica que os clusters que usam essa configuração de segurança autenticam os principais do Kerberos usando um KDC servidor fora do cluster. A não KDC é criado no cluster. Ao criar o cluster, você especifica o nome do território e a senha KDC do administrador para o externoKDC.
Vida útil do tíquete		Opcional. Especifica o período durante o qual um tíquete Kerberos emitido pelo KDC é válido em clusters que usam essa configuração de segurança. Os ciclos de vida do tíquete são limitados por motivos de segurança. As aplicações e os serviços de cluster renovarão automaticamente os tíquetes quando perderem a validade. Os usuários que se conectam ao cluster SSH usando as credenciais do Kerberos precisam executar a `kinit` partir da linha de comando do nó primário para renovar após a expiração de um ticket.

Parâmetro		Descrição
Relação de confiança entre realms		Especifica uma relação de confiança entre regiões entre um cluster dedicado KDC em clusters que usam essa configuração de segurança e um em uma KDC região Kerberos diferente. As entidades principais (normalmente usuários) de outro realm são autenticados em clusters que usam essa configuração. É necessário ter configuração adicional no outro realm do Kerberos. Para obter mais informações, consulte Tutorial: configurar uma relação de confiança entre realms com um controlador de domínio do Active Directory.
Propriedades de confiança entre realms	Realm	Especifica o nome de realm Kerberos de outro realm na relação de confiança. Por convenção, os nomes de realm do Kerberos são iguais ao nome do domínio, mas em letras maiúsculas.
	Domínio	Especifica o nome de domínio de outro realm na relação de confiança.
	Servidor do administrador	Especifica o nome de domínio totalmente qualificado (FQDN) ou o endereço IP do servidor administrativo no outro domínio da relação de confiança. O servidor administrativo e o KDC servidor normalmente são executados na mesma máquina com a mesmaFQDN, mas se comunicam em portas diferentes. Se nenhuma porta especificada, a porta 749 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com:749`).

Parâmetro		Descrição
	KDCservidor	Especifica o nome de domínio totalmente qualificado (FQDN) ou endereço IP do KDC servidor no outro domínio da relação de confiança. O KDC servidor e o servidor de administração normalmente são executados na mesma máquina com a mesma portaFQDN, mas usam portas diferentes. Se nenhuma porta especificada, a porta 88 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com :88`).
Externo KDC		Especifica que os clusters externos KDC sejam usados pelo cluster.
KDCPropriedades externas	Servidor do administrador	Especifica o nome de domínio totalmente qualificado (FQDN) ou o endereço IP do servidor administrativo externo. O servidor administrativo e o KDC servidor normalmente são executados na mesma máquina com a mesmaFQDN, mas se comunicam em portas diferentes. Se nenhuma porta especificada, a porta 749 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com :749`).

Parâmetro			Descrição
	KDCservidor		Especifica o nome de domínio totalmente qualificado (FQDN) do KDC servidor externo. O KDC servidor e o servidor de administração normalmente são executados na mesma máquina com a mesma portaFQDN, mas usam portas diferentes. Se nenhuma porta especificada, a porta 88 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com :88`).
	Integração do Active Directory		Especifica que a autenticação da entidade principal do Kerberos está integrada a um domínio do Microsoft Active Directory.
	Propriedades de integração do Active Directory	Realm do Active Directory	Especifica o nome do realm do Kerberos do domínio do Active Directory. Por convenção, os nomes de realm do Kerberos geralmente são iguais ao nome do domínio, mas em letras maiúsculas.
		Domínio do Active Directory	Especifica o nome de domínio do Active Directory.
		Servidor do Active Directory	Especifica o nome de domínio totalmente qualificado (FQDN) do controlador de domínio do Microsoft Active Directory.

Especificando as configurações do Kerberos usando o AWS CLI

A tabela de referência a seguir mostra JSON os parâmetros das configurações do Kerberos em uma configuração de segurança. Para exemplos de configuração, consulte Exemplos de configuração.

Parâmetro	Descrição
"AuthenticationConfiguration": {	Obrigatório para o Kerberos. Especifica que uma configuração de autenticação faz parte dessa configuração de segurança.
"KerberosConfiguration": {	Obrigatório para o Kerberos. Especifica as propriedades de configuração do Kerberos.
"Provider": *"ClusterDedicatedKdc",* —ou— "Provider: *"ExternalKdc",*	*ClusterDedicatedKdc* especifica que a Amazon EMR cria um KDC no nó primário de qualquer cluster que usa essa configuração de segurança. Você especifica o nome do território e a senha KDC do administrador ao criar o cluster. Você pode referenciar isso KDC em outros clusters, se necessário. Crie esses clusters usando uma configuração de segurança diferente, especifique uma externa KDC e use o nome do território e a senha de KDC administrador que você especificou ao criar o cluster com o cluster KDC dedicado. *ExternalKdc* especifica que o cluster usa um externoKDC. EMRA Amazon não cria um KDC no nó primário. Um cluster que usa essa configuração de segurança deve especificar o nome do território e a senha de KDC administrador do externoKDC.

Parâmetro	Descrição
`"ClusterDedicatedKdcConfiguration": {`	Obrigatório quando *`ClusterDedicatedKdc`* for especificado.
`"TicketLifetimeInHours": 24,`	Opcional. Especifica o período durante o qual um tíquete Kerberos emitido pelo KDC é válido em clusters que usam essa configuração de segurança. Os ciclos de vida do tíquete são limitados por motivos de segurança . As aplicações e os serviços de cluster renovarão automaticamente os tíquetes quando perderem a validade. Os usuários que se conectam ao cluster SSH usando as credenciais do Kerberos precisam executar a `kinit` partir da linha de comando do nó primário para renovar após a expiração de um ticket.

Parâmetro				Descrição
		`"CrossRealmTrustConfiguration": {`		Especifica uma relação de confiança entre regiões entre um cluster dedicado KDC em clusters que usam essa configuração de segurança e um em uma KDC região Kerberos diferente. As entidades principais (normalmente usuários) de outro realm são autenticados em clusters que usam essa configuração. É necessário ter configuração adicional no outro realm do Kerberos. Para obter mais informações, consulte Tutorial: configurar uma relação de confiança entre realms com um controlador de domínio do Active Directory.
			`"Realm": "KDC2.COM",`	Especifica o nome de realm Kerberos de outro realm na relação de confiança. Por convenção, os nomes de realm do Kerberos são iguais ao nome do domínio, mas em letras maiúsculas.
			`"Domain": "kdc2.com",`	Especifica o nome de domínio de outro realm na relação de confiança.

Parâmetro	Descrição
"AdminServer": "*kdc.com:749* ",	Especifica o nome de domínio totalmente qualificado (FQDN) ou o endereço IP do servidor administrativo no outro domínio da relação de confiança. O servidor administrativo e o KDC servidor normalmente são executados na mesma máquina com a mesmaFQDN, mas se comunicam em portas diferentes. Se nenhuma porta especificada, a porta 749 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com` *:749*).
"KdcServer": "*kdc.com:88* "	Especifica o nome de domínio totalmente qualificado (FQDN) ou endereço IP do KDC servidor no outro domínio da relação de confiança. O KDC servidor e o servidor de administração normalmente são executados na mesma máquina com a mesma portaFQDN, mas usam portas diferentes. Se nenhuma porta especificada, a porta 88 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com` *:88*).

Parâmetro	Descrição
}	
}	
"ExternalKdcConfiguration": {	Obrigatório quando *ExternalKdc* for especificado.
"TicketLifetimeInHours": *24*,	Opcional. Especifica o período durante o qual um tíquete Kerberos emitido pelo KDC é válido em clusters que usam essa configuração de segurança. Os ciclos de vida do tíquete são limitados por motivos de segurança . As aplicações e os serviços de cluster renovarão automaticamente os tíquetes quando perderem a validade. Os usuários que se conectam ao cluster SSH usando as credenciais do Kerberos precisam executar a `kinit` partir da linha de comando do nó primário para renovar após a expiração de um ticket.
"KdcServerType": "Single",	Especifica que um único KDC servidor é referenciado. `Single`atualmente é o único valor suportado.

Parâmetro	Descrição
"AdminServer": "*kdc.com:749* ",	Especifica o nome de domínio totalmente qualificado (FQDN) ou o endereço IP do servidor administr ativo externo. O servidor administr ativo e o KDC servidor normalmen te são executados na mesma máquina com a mesmaFQDN, mas se comunicam em portas diferentes. Se nenhuma porta especificada, a porta 749 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com` *:749*).
"KdcServer": "*kdc.com:8 8* ",	Especifica o nome de domínio totalmente qualificado (FQDN) do KDC servidor externo. O KDC servidor e o servidor de administr ação normalmente são executados na mesma máquina com a mesma portaFQDN, mas usam portas diferentes. Se nenhuma porta especificada, a porta 88 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com` *:88*).

Parâmetro	Descrição
"AdIntegrationConfiguration": {	Especifica que a autenticação da entidade principal do Kerberos está integrada a um domínio do Microsoft Active Directory.
"AdRealm": "*AD.DOMAIN.COM* ",	Especifica o nome do realm do Kerberos do domínio do Active Directory. Por convenção, os nomes de realm do Kerberos geralmente são iguais ao nome do domínio, mas em letras maiúsculas.
"AdDomain": "*ad.domain.com* "	Especifica o nome de domínio do Active Directory.
"AdServer": "*ad.domain.com* "	Especifica o nome de domínio totalmente qualificado (FQDN) do controlador de domínio do Microsoft Active Directory.
}	
}	
}	
}	

Configurar IAM funções para EMRFS solicitações para o Amazon S3

IAMAs funções para EMRFS permitem que você forneça permissões diferentes aos EMRFS dados no Amazon S3. Você cria mapeamentos que especificam uma IAM função que é usada para permissões quando uma solicitação de acesso contém um identificador que você especifica. O identificador pode ser um usuário ou um perfil do Hadoop ou um prefixo do Amazon S3.

Para obter mais informações, consulte Configurar IAM funções para EMRFS solicitações para o Amazon S3.

Especificando IAM funções para EMRFS usar o AWS CLI

Veja a seguir um exemplo de JSON trecho para especificar IAM funções personalizadas EMRFS em uma configuração de segurança. Ele demonstra mapeamentos de perfil para os três tipos diferentes de identificadores, seguidos por uma referência de parâmetro.

```
{
  "AuthorizationConfiguration": {
    "EmrFsConfiguration": {
      "RoleMappings": [{
        "Role": "arn:aws:iam::123456789101:role/allow_EMRFS_access_for_user1",
        "IdentifierType": "User",
        "Identifiers": [ "user1" ]
      },{
        "Role": "arn:aws:iam::123456789101:role/allow_EMRFS_access_to_MyBuckets",
        "IdentifierType": "Prefix",
        "Identifiers": [ "s3://MyBucket/","s3://MyOtherBucket/" ]
      },{
        "Role": "arn:aws:iam::123456789101:role/allow_EMRFS_access_for_AdminGroup",
        "IdentifierType": "Group",
        "Identifiers": [ "AdminGroup" ]
      }]
    }
  }
}
```

Parâmetro	Descrição
`"AuthorizationConfiguration":`	Obrigatório.
`"EmrFsConfiguration":`	Obrigatório. Contém mapeamentos de perfil.
`"RoleMappings":`	Obrigatório. Contém uma ou mais definições de mapeamento de perfil. Os mapeamentos de perfil são avaliados na ordem em que aparecem, de cima para baixo. Se um mapeamento de função for avaliado como verdadeiro para uma EMRFS chamada

Parâmetro	Descrição
	de dados no Amazon S3, nenhum outro mapeamento de função será avaliado EMRFS e usará a função IAM especificada para a solicitação. O mapeamento de perfil tem os seguintes parâmetros obrigatórios:
`"Role":`	Especifica o ARN identificador de uma IAM função no formato`arn:aws:iam:: account-id :role/role-name` . Essa é a IAM função que a Amazon EMR assume se a EMRFS solicitação para o Amazon S3 corresponder a qualquer uma das `Identifiers` especificadas.
`"IdentifierType":`	Pode ser um dos seguintes: • `"User"` especifica que os identificadores são um ou mais usuários do Hadoop, que podem ser usuários de contas Linux ou entidades principais do Kerberos. Quando a EMRFS solicitação se origina com o usuário ou usuários especificados, a IAM função é assumida. • `"Prefix"` especifica que o identificador é um local do Amazon S3. A IAM função é assumida para chamadas para o local ou locais com os prefixos especificados. Por exemplo, o prefixo `s3://mybucket/` corresponde a `s3://mybucket/mydir` e `s3://mybucket/yetanotherdir` . • `"Group"` especifica que os identificadores são um ou mais grupos do Hadoop. A IAM função é assumida se a solicitação for originada de um usuário no grupo ou grupos especificados.

Parâmetro	Descrição
`"Identifiers":`	Especifica um ou mais identificadores do tipo de identificador adequado. Separe múltiplos identificadores por vírgulas sem espaços.

Configurar solicitações de serviço de metadados para instâncias da Amazon EC2

Os metadados da instância são dados sobre sua instância que é possível usar para configurar ou gerenciar a instância em execução. É possível acessar metadados de instância em uma instância em execução usando um dos seguintes métodos:

- Instance Metadata Service versão 1 (IMDSv1) — um método de solicitação/resposta
- Instance Metadata Service versão 2 (IMDSv2) — um método orientado a sessões

Enquanto a Amazon EC2 oferece suporte a ambos IMDSv1 eIMDSv2, a Amazon EMR oferece suporte IMDSv2 no Amazon EMR 5.23.1, 5.27.1, 5.32 ou posterior e 6.2 ou posterior. Nessas versões, EMR os componentes da Amazon são usados IMDSv2 para todas as IMDS chamadas. Para IMDS chamadas no código do seu aplicativo, você pode usar ambos IMDSv1 eIMDSv2, ou configurar o IMDS para uso somente IMDSv2 para aumentar a segurança. Quando você especifica que IMDSv2 deve ser usado, IMDSv1 não funciona mais.

Para obter mais informações, consulte Configurar o serviço de metadados da instância no Guia do EC2 usuário da Amazon.

> **Note**
>
> Nas versões anteriores do Amazon EMR 5.x ou 6.x, a desativação IMDSv1 causa falha na inicialização do cluster, pois os EMR componentes da Amazon são usados IMDSv1 para todas as IMDS chamadas. Ao desligarIMDSv1, certifique-se de que qualquer software personalizado que utilize IMDSv1 esteja atualizado paraIMDSv2.

Especificar a configuração do serviço de metadados da instância usando a AWS CLI

Veja a seguir um exemplo de JSON trecho para especificar o serviço de metadados de instância EC2 da Amazon IMDS () em uma configuração de segurança. Usar uma configuração de segurança personalizada é opcional.

```
{
  "InstanceMetadataServiceConfiguration" : {
      "MinimumInstanceMetadataServiceVersion": integer,
      "HttpPutResponseHopLimit": integer
   }
}
```

Parâmetro	Descrição
`"InstanceMetadataServiceConfiguration":`	Se você não especificar IMDS em uma configuração de segurança e usar uma EMR versão da Amazon que exijaIMDSv1, a Amazon usará IMDSv1 como EMR padrão a versão mínima do serviço de metadados de instância. Se você quiser usar sua própria configuração, os dois parâmetros a seguir são necessários.
`"MinimumInstanceMetadataServiceVersion":`	Obrigatório. Especifique 1 ou 2. Um valor 1 de IMDSv1 permissões IMDSv2 e. Um valor de 2 somente permissõesIMDSv2.
`"HttpPutResponseHopLimit":`	Obrigatório. O limite de salto de HTTP PUT resposta desejado para solicitações de metadados da instância. Quanto maior o número, mais as solicitações de metadados de instância podem viajar. Padrão: 1. Especifique um número inteiro de 1 a 64.

Especificar a configuração do serviço de metadados da instância usando o console

Você pode configurar o uso de IMDS para um cluster ao iniciá-lo no EMR console da Amazon.

Para configurar o IMDS uso do console:

1. Ao criar uma nova configuração de segurança na página Configurações de segurança, selecione Configurar serviço de metadados de EC2 instância na configuração Serviço de metadados de EC2 instância. Essa configuração é suportada somente no Amazon EMR 5.23.1, 5.27.1, 5.32 ou posterior e 6.2 ou posterior.
2. Na opção Versão mínima do serviço de metadados de instância, selecione:
   - Desative IMDSv1 e permita somente IMDSv2, se você quiser permitir somente IMDSv2 neste cluster. Consulte Transição para o uso do serviço de metadados de instância versão 2 no Guia do EC2 usuário da Amazon.
   - Permita ambos IMDSv1 e IMDSv2 no cluster, se você quiser permitir IMDSv1 e orientado por sessão nesse IMDSv2 cluster.
3. ParaIMDSv2, você também pode configurar o número permitido de saltos de rede para o token de metadados definindo o limite de salto de resposta HTTP put como um número inteiro entre e. 1 64

Para obter mais informações, consulte Configurar o serviço de metadados da instância no Guia do EC2 usuário da Amazon.

Consulte Configurar detalhes da instância e Configurar o serviço de metadados da instância no Guia do EC2 usuário da Amazon.

Especificar uma configuração de segurança para um cluster

Ao criar um cluster, você pode especificar configurações de criptografia definindo a configuração de segurança. Você pode usar o AWS Management Console ou AWS CLI o.

Console

Para especificar uma configuração de segurança com o console

1. Faça login no AWS Management Console e abra o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Em Configuração e permissões de segurança, localize o campo Configuração de segurança. Selecione o menu suspenso ou escolha Procurar para selecionar o nome de

uma configuração de segurança criada anteriormente. Como alternativa, escolha Criar configuração de segurança para criar uma configuração que você possa usar em seu cluster.

4. Escolha qualquer outra opção que se aplique ao cluster.
5. Para iniciar o cluster, escolha Criar cluster.

CLI

Para especificar uma configuração de segurança com o AWS CLI

- Use `aws emr create-cluster` para opcionalmente aplicar uma configuração de segurança usando `--security-configuration` *`MySecConfig`*, em que *`MySecConfig`* é o nome da configuração de segurança, como mostra o exemplo a seguir. O `--release-label` que você especificar deve ser 4.8.0 ou posterior e o `--instance-type` pode ser qualquer disponível.

```
aws emr create-cluster --instance-type m5.xlarge --release-label emr-5.0.0 --
security-configuration mySecConfig
```

Proteção de dados na Amazon EMR

O modelo de responsabilidade AWS compartilhada se aplica à proteção de dados na AmazonEMR. Conforme descrito neste modelo, AWS é responsável por proteger a infraestrutura global que executa toda a AWS nuvem. Você é responsável por manter o controle sobre seu conteúdo hospedado nessa infraestrutura. Esse conteúdo inclui as tarefas de configuração e gerenciamento de segurança do AWS que você usa. Para obter mais informações sobre privacidade de dados, consulte Privacidade de dados FAQ. Para obter informações sobre proteção de dados na Europa, consulte o modelo de responsabilidade compartilhada da Amazon e GDPR a postagem no blog AWS de segurança.

Para fins de proteção de dados, recomendamos que você proteja as credenciais da AWS conta e configure contas individuais com AWS Identity and Access Management. Dessa maneira, cada usuário receberá apenas as permissões necessárias para cumprir suas obrigações de trabalho. Recomendamos também que você proteja seus dados das seguintes formas:

- Use a autenticação multifator (MFA) com cada conta.
- Use TLS para se comunicar com AWS os recursos. Exigimos TLS 1,2.

- Configure API e registre as atividades do usuário com AWS CloudTrail.
- Use soluções AWS de criptografia, juntamente com todos os controles de segurança padrão nos AWS serviços.
- Use serviços gerenciados de segurança avançada, como o Amazon Macie, que ajuda a localizar e proteger dados pessoais armazenados no Amazon S3.
- Se você precisar de FIPS 140-2 módulos criptográficos validados ao acessar AWS por meio de uma interface de linha de comando ou umaAPI, use um endpoint. FIPS Para obter mais informações sobre os FIPS endpoints disponíveis, consulte Federal Information Processing Standard (FIPS) 140-2.

É altamente recomendável que você nunca coloque informações de identificação confidenciais, como números de conta dos seus clientes, em campos de formato livre, como um campo Nome. Isso inclui quando você trabalha com a Amazon EMR ou outros AWS serviços usando o console,API, AWS CLI, ou AWS SDKs. Todos os dados que você inserir na Amazon EMR ou em outros serviços podem ser coletados para inclusão nos registros de diagnóstico. Ao fornecer um URL para um servidor externo, não inclua informações de credenciais no URL para validar sua solicitação para esse servidor.

Criptografar dados em repouso e em trânsito

A criptografia de dados ajuda a impedir que usuários não autorizados leiam dados em um cluster e em sistemas de armazenamento físico de dados associados. Isso inclui dados salvos em mídias persistentes, conhecidos como dados em repouso, e dados que podem ser interceptados enquanto viajam pela rede, conhecidos como dados em trânsito.

A partir da EMR versão 4.8.0 da Amazon, você pode usar as configurações de EMR segurança da Amazon para definir as configurações de criptografia de dados para clusters com mais facilidade. As configurações de segurança oferecem configurações para permitir a segurança de dados em trânsito e dados em repouso nos volumes do Amazon Elastic Block Store (AmazonEBS) e no Amazon EMRFS S3.

Opcionalmente, começando com a EMR versão 4.1.0 e posterior da Amazon, você pode optar por configurar a criptografia transparenteHDFS, que não é configurada usando configurações de segurança. Para obter mais informações, consulte Criptografia transparente HDFS na Amazon EMR no Guia de EMR lançamento da Amazon.

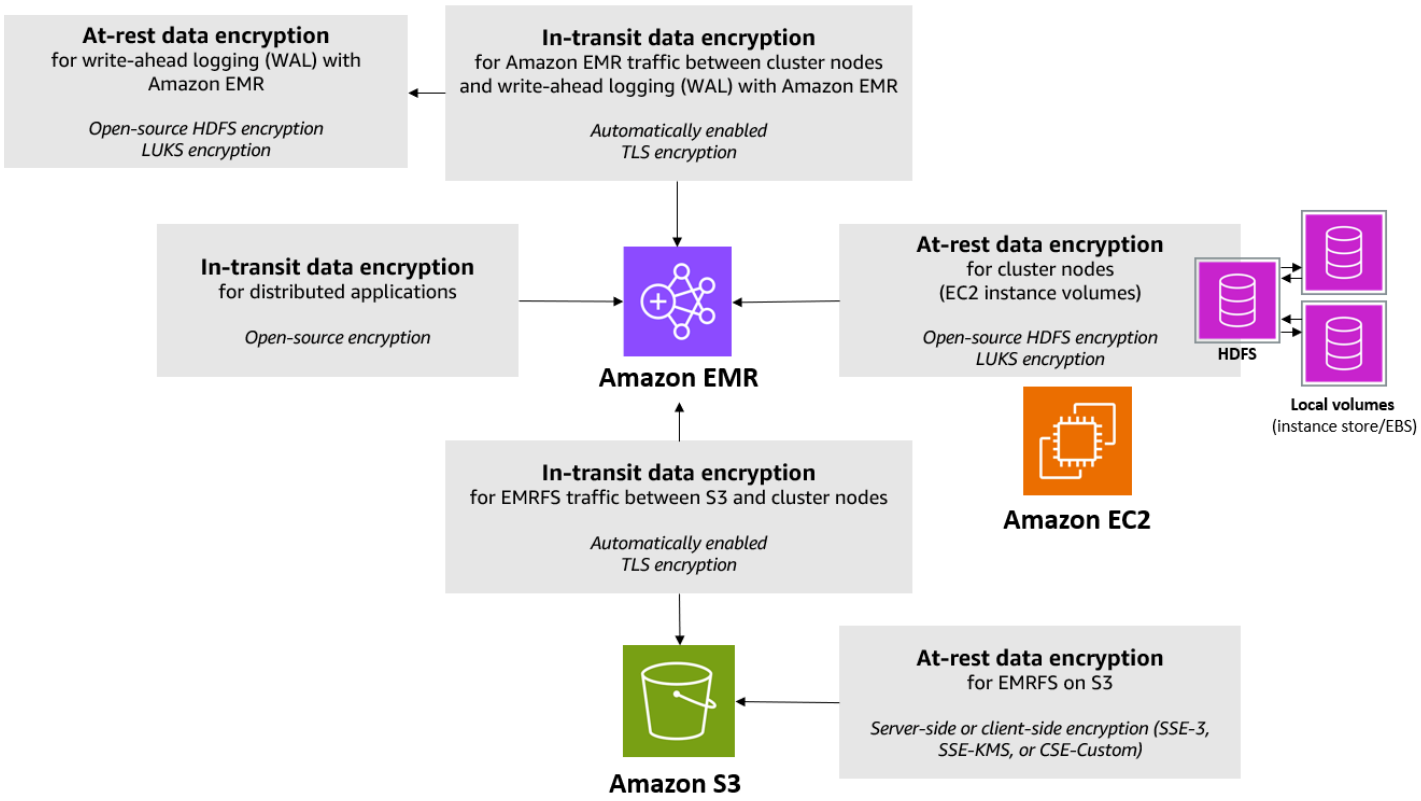
Tópicos

- Opções de criptografia
- Criar chaves e certificados para criptografia de dados

Opções de criptografia

Com as EMR versões 4.8.0 e superiores da Amazon, você pode usar uma configuração de segurança para especificar configurações para criptografar dados em repouso, dados em trânsito ou ambos. Ao habilitar a criptografia de dados em repouso, você pode optar por criptografar EMRFS dados no Amazon S3, dados em discos locais ou ambos. Cada configuração de segurança que você cria é armazenada na Amazon e EMR não na configuração do cluster, para que você possa facilmente reutilizar uma configuração para especificar as configurações de criptografia de dados sempre que criar um cluster. Para obter mais informações, consulte Criar uma configuração de segurança.

O diagrama a seguir mostra as diferentes opções de criptografia de dados disponíveis com as configurações de segurança.



As seguintes opções de criptografia também estão disponíveis e não são configuradas usando uma configuração de segurança:

- Opcionalmente, com EMR as versões 4.1.0 e posteriores da Amazon, você pode optar por configurar a criptografia transparente no. HDFS Para obter mais informações, consulte Criptografia transparente HDFS na Amazon EMR no Guia de EMR lançamento da Amazon.
- Se você estiver usando uma versão de lançamento da Amazon EMR que não suporta configurações de segurança, você pode configurar a criptografia para EMRFS dados no Amazon S3 manualmente. Para obter mais informações, consulte Especificação da criptografia do Amazon S3 usando EMRFS propriedades.
- Se você estiver usando uma EMR versão da Amazon anterior à 5.24.0, um volume de dispositivo EBS raiz criptografado é suportado somente ao usar um volume personalizado. AMI Para obter mais informações, consulte Criação de um volume personalizado AMI com um dispositivo EBS raiz criptografado da Amazon no Amazon EMR Management Guide.

> **Note**
>
> A partir da EMR versão 5.24.0 da Amazon, você pode usar uma opção de configuração de segurança para criptografar o dispositivo EBS raiz e os volumes de armazenamento ao especificar AWS KMS como seu provedor de chaves. Para obter mais informações, consulte Criptografia de disco local.

A criptografia de dados requer chaves e certificados. Uma configuração de segurança oferece a flexibilidade de escolher entre várias opções, incluindo chaves gerenciadas por AWS Key Management Service, chaves gerenciadas pelo Amazon S3 e chaves e certificados de fornecedores personalizados fornecidos por você. Ao usar AWS KMS como seu provedor de chaves, cobranças se aplicam pelo armazenamento e uso de chaves de criptografia. Para obter mais informações, consulte Definição de preço do AWS KMS.

Antes de especificar opções de criptografia, decida quais sistemas de gerenciamento de chaves e certificados você deseja usar, para poder primeiro criar as chaves e os certificados ou os provedores personalizados especificados como parte das configurações de criptografia.

Criptografia em repouso para EMRFS dados no Amazon S3

A criptografia do Amazon S3 funciona com os objetos EMR do Amazon File System (EMRFS) lidos e gravados no Amazon S3. Você especifica a criptografia do lado do servidor (SSE) ou do cliente () do Amazon S3 como o modo de criptografia padrão ao ativar a criptografia em repouso. CSE Opcionalmente, você pode especificar diferentes métodos de criptografia para buckets

individuais usando Per bucket encryption overrides (Substituições de criptografia por bucket). Independentemente de a criptografia do Amazon S3 estar habilitada, o Transport Layer Security (TLS) criptografa os EMRFS objetos em trânsito entre os nós do EMR cluster e o Amazon S3. Para obter mais informações sobre a criptografia do Amazon S3, consulte Proteção de dados usando criptografia no Guia do usuário do Amazon Simple Storage Service.

> **Note**
>
> Quando você usa AWS KMS, cobranças são cobradas pelo armazenamento e uso de chaves de criptografia. Para obter mais informações, consulte Preços do AWS KMS.

Criptografia do lado do servidor do Amazon S3

Quando você configura a criptografia do lado do servidor do Amazon S3, o Amazon S3 criptografa os dados no nível do objeto à medida que os grava no disco e os descriptografa quando são acessados. Para obter mais informações sobre issoSSE, consulte Proteção de dados usando criptografia do lado do servidor no Guia do usuário do Amazon Simple Storage Service.

Você pode escolher entre dois sistemas diferentes de gerenciamento de chaves ao especificar SSE na AmazonEMR:

- SSE-S3 — O Amazon S3 gerencia as chaves para você.
- SSE- KMS — Você usa um AWS KMS key para configurar políticas adequadas para a AmazonEMR. Para obter mais informações sobre os principais requisitos da AmazonEMR, consulte Usando AWS KMS keys para criptografia.

SSEcom chaves fornecidas pelo cliente (SSE-C) não está disponível para uso com a Amazon. EMR

Criptografia do lado do cliente do Amazon S3

Com a criptografia do lado do cliente do Amazon S3, a criptografia e a descriptografia do Amazon S3 ocorrem no cliente em seu cluster. EMRFS Os objetos são criptografados antes de serem carregados no Amazon S3 e descriptografados após serem baixados. O provedor especificado por você fornece a chave de criptografia que o cliente usa. O cliente pode usar chaves fornecidas por AWS KMS (CSE-KMS) ou uma classe Java personalizada que fornece a chave raiz do lado do cliente (CSE-C). As especificações da criptografia são ligeiramente diferentes entre CSE - KMS e CSE -C, dependendo do provedor especificado e dos metadados do objeto que está sendo

descriptografado ou criptografado. Para obter mais informações sobre essas diferenças, consulte Proteger dados usando a criptografia do lado do cliente no Guia do usuário do Amazon Simple Storage Service.

> **Note**
>
> O Amazon S3 CSE só garante que EMRFS os dados trocados com o Amazon S3 sejam criptografados; nem todos os dados nos volumes de instâncias de cluster são criptografados. Além disso, como o Hue não usaEMRFS, os objetos que o Navegador de arquivos Hue S3 grava no Amazon S3 não são criptografados.

Criptografia em repouso para dados na Amazon EMR WAL

Quando você configura a criptografia do lado do servidor (SSE) para registro antecipado de gravação ()WAL, a Amazon EMR criptografa os dados em repouso. Você pode escolher entre dois sistemas diferentes de gerenciamento de chaves ao especificar SSE na AmazonEMR:

SSE-EMR-WAL

A Amazon EMR gerencia as chaves para você. Por padrão, a Amazon EMR criptografa os dados EMR WAL com SSE-EMR-WAL os quais você armazenou na Amazon.

SSE-KMS-WAL

Você usa uma AWS KMS chave para configurar políticas que se aplicam à Amazon EMRWAL. Para obter mais informações sobre os principais requisitos da AmazonEMR, consulteUsando AWS KMS keys para criptografia.

Você não pode usar sua própria chave SSE ao habilitar WAL com a AmazonEMR. Para obter mais informações, consulte Write-ahead logs (WAL) para a Amazon. EMR

Criptografia de disco local

Os mecanismos a seguir trabalham juntos para criptografar discos locais quando você ativa a criptografia de disco local usando uma configuração de EMR segurança da Amazon.

Criptografia de código aberto HDFS

HDFStroca dados entre instâncias de cluster durante o processamento distribuído. Ele também lê e grava dados nos volumes de armazenamento de instâncias e nos EBS volumes anexados às

instâncias. As seguintes opções de criptografia Hadoop de código-fonte aberto são ativadas quando você habilita a criptografia do disco local:

- O Secure Hadoop RPC está definido como`Privacy`, que usa a Camada de Segurança de Autenticação Simples ()SASL.
- A criptografia de dados na transferência de dados em HDFS bloco está definida como `true` e está configurada para usar criptografia AES 256.

> **Note**
>
> Você pode ativar a criptografia adicional do Apache Hadoop habilitando a criptografia em trânsito. Para obter mais informações, consulte Criptografia em trânsito. Essas configurações de criptografia não ativam a criptografia HDFS transparente, que você pode configurar manualmente. Para obter mais informações, consulte Criptografia transparente HDFS na Amazon EMR no Guia de EMR lançamento da Amazon.

Criptografia de armazenamento de instância

Para tipos de EC2 instância que usam NVMe based SSDs como volume de armazenamento de instâncias, a NVMe criptografia é usada independentemente das configurações de EMR criptografia da Amazon. Para obter mais informações, consulte NVMeSSDos volumes no Guia EC2 do usuário da Amazon. Para outros volumes de armazenamento de instâncias, a Amazon EMR usa LUKS para criptografar o volume de armazenamento de instâncias quando a criptografia de disco local está ativada, independentemente de EBS os volumes serem criptografados usando EBS criptografia ouLUKS.

EBScriptografia de volume

Se você criar um cluster em uma região em que a EC2 criptografia de EBS volumes da Amazon esteja habilitada por padrão para sua conta, EBS os volumes serão criptografados mesmo que a criptografia de disco local não esteja habilitada. Para obter mais informações, consulte Criptografia por padrão no Guia EC2 do usuário da Amazon. Com a criptografia de disco local ativada em uma configuração de segurança, as EMR configurações da Amazon têm precedência sobre EC2 encryption-by-default as configurações da Amazon para EC2 instâncias de cluster.

As opções a seguir estão disponíveis para criptografar EBS volumes usando uma configuração de segurança:

- EBSCriptografia — A partir da EMR versão 5.24.0 da Amazon, você pode optar por ativar EBS a criptografia. A opção EBS de criptografia criptografa o volume do dispositivo EBS raiz e os volumes de armazenamento conectados. A opção de EBS criptografia está disponível somente quando você especifica AWS Key Management Service como seu provedor de chaves. Recomendamos o uso EBS de criptografia.
- LUKSCriptografia — Se você optar por usar a LUKS criptografia para EBS volumes da Amazon, a LUKS criptografia se aplicará somente aos volumes de armazenamento conectados, não ao volume do dispositivo raiz. Para obter mais informações sobre LUKS criptografia, consulte a especificação LUKS em disco.

  Para seu provedor de chaves, você pode configurar uma AWS KMS key com políticas adequadas para a Amazon EMR ou uma classe Java personalizada que forneça os artefatos de criptografia. Quando você usa AWS KMS, cobranças são cobradas pelo armazenamento e uso de chaves de criptografia. Para obter mais informações, consulte Definição de preço do AWS KMS.

> **Note**
>
> Para verificar se a EBS criptografia está habilitada em seu cluster, é recomendável usar `DescribeVolumes` API call. Para obter mais informações, consulte DescribeVolumes. A execução `lsblk` no cluster só verificará o status da LUKS criptografia, em vez da EBS criptografia.

Criptografia em trânsito

Vários mecanismos de criptografia estão habilitados com a criptografia em trânsito. Esses são recursos de código aberto, são específicos do aplicativo e podem variar de acordo com o lançamento da Amazon. EMR Os atributos de criptografia a seguir específicos da aplicação podem ser habilitados usando configurações de aplicação do Apache. Para obter mais informações, consulte Configure applications.

Hadoop

- Usos do shuffle MapReduce criptografado do Hadoop. TLS
- O Secure Hadoop RPC está configurado como “Privacidade” e usa SASL (ativado na Amazon EMR quando a criptografia em repouso está ativada).

- A criptografia de dados na transferência de dados em HDFS bloco usa AES 256 (ativada na Amazon EMR quando a criptografia em repouso está ativada na configuração de segurança).
- Para obter mais informações, consulte Hadoop in secure mode, na documentação do Apache Hadoop.

HBase

- Quando o Kerberos está habilitado, a propriedade `hbase.rpc.protection` é definida como `privacy` para comunicação criptografada.
- Para obter mais informações, consulte Configuração do lado do cliente para operação segura na documentação do HBase Apache.
- Para obter mais informações sobre o Kerberos com a AmazonEMR, consulte. Use o Kerberos para autenticação com a Amazon EMR

Hive

- JDBC/a comunicação ODBC do cliente com HiveServer 2 (HS2) é criptografada usando SSL configurações nas EMR versões 6.9.0 e posteriores da Amazon.
- Para obter mais informações, consulte a seção de SSLcriptografia da documentação do Apache Hive.

Spark

- RPCA comunicação interna entre os componentes do Spark, como o serviço de transferência de blocos e o serviço de embaralhamento externo, é criptografada usando a AES cifra -256 nas versões 5.9.0 e posteriores da Amazon. EMR Em versões anteriores, a RPC comunicação interna é criptografada usando SASL with DIGEST - MD5 como cifra.
- HTTPa comunicação de protocolo com interfaces de usuário, como o Spark History Server e servidores de arquivos HTTPS habilitados, é criptografada usando a configuração do Spark. SSL Para obter mais informações, consulte a SSLconfiguração na documentação do Spark.
- Para obter mais informações, consulte a seção Spark security settings da documentação do Apache Spark.

Tez

- O manipulador Tez shuffle usa (). TLS `tez.runtime.ssl.enable`

Presto

- A comunicação interna entre os nós do Presto usaSSL/TLS(somente na EMR versão 5.6.0 e posterior da Amazon).

Você especifica os artefatos de criptografia usados para a criptografia em trânsito de uma destas duas maneiras: fornecendo um arquivo compactado de certificados, que é carregado no Amazon S3, ou referenciando uma classe Java personalizada, que fornece artefatos de criptografia. Para obter mais informações, consulte Fornecendo certificados para criptografar dados em trânsito com a criptografia da Amazon EMR.

Criar chaves e certificados para criptografia de dados

Antes de especificar as opções de criptografia usando uma configuração de segurança, decida qual provedor você quer usar para as chaves e os artefatos criptográficos. Por exemplo, você pode usar AWS KMS ou um provedor personalizado criado por você. Depois, crie as chaves ou o provedor de chaves conforme descrito nesta seção.

Fornecendo chaves para criptografar dados em repouso com a Amazon EMR

Você pode usar AWS Key Management Service (AWS KMS) ou um provedor de chave personalizado para criptografia de dados em repouso na AmazonEMR. Quando você usa AWS KMS, cobranças são cobradas pelo armazenamento e uso de chaves de criptografia. Para obter mais informações, consulte Definição de preço do AWS KMS.

Este tópico fornece os principais detalhes da política de uma KMS chave a ser usada com a AmazonEMR, bem como diretrizes e exemplos de código para escrever uma classe de provedor de chaves personalizada para a criptografia do Amazon S3. Para obter mais informações sobre como criar chaves, consulte Creating keys no Guia do desenvolvedor do AWS Key Management Service .

Usando AWS KMS keys para criptografia

A chave de AWS KMS criptografia deve ser criada na mesma região da sua instância de EMR cluster da Amazon e dos buckets do Amazon S3 usados com. EMRFS Se a chave que você especificar estiver em uma conta diferente daquela usada para configurar um cluster, você deverá especificar a chave usando suaARN.

A função do perfil da EC2 instância da Amazon deve ter permissões para usar a KMS chave que você especificar. A função padrão para o perfil da instância na Amazon EMR é`EMR_EC2_DefaultRole`. Se você usar uma função diferente para o perfil da instância ou usar IAM funções para EMRFS solicitações ao Amazon S3, certifique-se de que cada função seja adicionada como um usuário chave, conforme apropriado. Isso dá à função permissões para usar a KMS chave. Para obter mais informações, consulte Usando políticas de chaves no Guia do AWS Key Management Service desenvolvedor e Configurar IAM funções para EMRFS solicitações ao Amazon S3.

Você pode usar o AWS Management Console para adicionar seu perfil de instância ou perfil de EC2 instância à lista de usuários de chave para a KMS chave especificada, ou você pode usar o AWS CLI ou an AWS SDK para anexar uma política de chaves apropriada.

Observe que a Amazon EMR oferece suporte somente a KMSchaves simétricas. Você não pode usar uma KMSchave assimétrica para criptografar dados em repouso em um cluster da Amazon. EMR Para obter ajuda para determinar se uma KMS chave é simétrica ou assimétrica, consulte Identificação de chaves simétricas e assimétricas. KMS

O procedimento abaixo descreve como adicionar o perfil de EMR instância padrão da Amazon, `EMR_EC2_DefaultRole` como um usuário chave usando AWS Management Console o. Ele pressupõe que você já tenha criado uma KMS chave. Para criar uma nova KMS chave, consulte Criação de chaves no Guia do AWS Key Management Service desenvolvedor.

Para adicionar o perfil de EC2 instância da Amazon EMR à lista de usuários da chave de criptografia

1. Faça login AWS Management Console e abra o console AWS Key Management Service (AWS KMS) em https://console.aws.amazon.com/kms.
2. Para alterar o Região da AWS, use o seletor de região no canto superior direito da página.
3. Selecione o alias da KMS chave a ser modificada.
4. Na página de detalhes da chave, em Key Users (Usuários de chaves), escolha Add (Adicionar).
5. Na caixa de diálogo Add key users (Adicionar usuários da chave) selecione a função apropriada. O nome da função padrão é `EMR_EC2_DefaultRole`.
6. Escolha Adicionar.

Ativando a EBS criptografia fornecendo permissões adicionais para KMS chaves

A partir da EMR versão 5.24.0 da Amazon, você pode criptografar o dispositivo EBS raiz e os volumes de armazenamento usando uma opção de configuração de segurança. Para ativar essa

opção, você deve especificar AWS KMS como seu provedor de chaves. Além disso, você deve conceder à função `EMR_DefaultRole` de serviço permissões para usar o AWS KMS key que você especificar.

Você pode usar o AWS Management Console para adicionar a função de serviço à lista de usuários principais da KMS chave especificada ou pode usar o AWS CLI ou an AWS SDK para anexar uma política de chaves apropriada.

O procedimento a seguir descreve como usar o AWS Management Console para adicionar a função de EMR serviço padrão da Amazon `EMR_DefaultRole` como um usuário chave. Ele pressupõe que você já tenha criado uma KMS chave. Para criar uma nova KMS chave, consulte Criação de chaves no Guia do AWS Key Management Service desenvolvedor.

Para adicionar a função EMR de serviço da Amazon à lista de usuários da chave de criptografia

1. Faça login AWS Management Console e abra o console AWS Key Management Service (AWS KMS) em https://console.aws.amazon.com/kms.
2. Para alterar o Região da AWS, use o seletor de região no canto superior direito da página.
3. Escolha Chaves gerenciadas pelo cliente na barra lateral esquerda.
4. Selecione o alias da KMS chave a ser modificada.
5. Na página de detalhes da chave, em Key Users (Usuários de chaves), escolha Add (Adicionar).
6. Na seção Adicionar usuários-chave, selecione a função apropriada. O nome da função de serviço padrão da Amazon EMR é`EMR_DefaultRole`.
7. Escolha Adicionar.

Criar um provedor de chaves personalizado

Ao usar uma configuração de segurança, você deve especificar um nome de classe de provedor diferente para a criptografia de disco local e para a criptografia do Amazon S3. Os requisitos para o provedor de chave personalizada dependem de você usar criptografia de disco local e criptografia Amazon S3, bem como a versão de EMR lançamento da Amazon.

Dependendo do tipo de criptografia que você usa ao criar um provedor de chave personalizado, o aplicativo também deve implementar EncryptionMaterialsProvider interfaces diferentes. Ambas as interfaces estão disponíveis na versão 1.11.0 e posterior AWS SDK para Java.

- Para implementar a criptografia do Amazon S3, use o com.amazonaws.services.s3.model.EncryptionMaterialsProvider interface.

- Para implementar a criptografia de disco local, use com.amazonaws.services.elasticmapreduce.spi.security. EncryptionMaterialsProvider interface.

Você pode usar qualquer estratégia para fornecer materiais de criptografia para a implementação. Por exemplo, você pode optar por fornecer materiais de criptografia estática ou integrá-los a um sistema de gerenciamento de chaves mais complexo.

Se você estiver usando a criptografia do Amazon S3, deverá usar os algoritmos de criptografia AES/GCM/NoPaddingpara materiais de criptografia personalizados.

Se você estiver usando criptografia de disco local, o algoritmo de criptografia a ser usado para materiais de criptografia personalizados varia de acordo com a EMR versão. Para o Amazon EMR 7.0.0 e versões anteriores, você deve usar AES/GCM/NoPadding. Para Amazon EMR 7.1.0 e superior, você deve usar AES.

A EncryptionMaterialsProvider classe obtém materiais de criptografia por contexto de criptografia. A Amazon EMR preenche as informações do contexto de criptografia em tempo de execução para ajudar o chamador a determinar os materiais de criptografia corretos a serem devolvidos.

Example Exemplo: Usando um provedor de chave personalizado para criptografia do Amazon S3 com EMRFS

Quando a Amazon EMR busca os materiais de criptografia da EncryptionMaterialsProvider classe para realizar a criptografia, EMRFS opcionalmente preenche o materialsDescription argumento com dois campos: o Amazon S3 URI para o objeto e o JobFlowId do cluster, que podem ser usados pela EncryptionMaterialsProvider classe para retornar materiais de criptografia seletivamente.

Por exemplo, o provedor pode retornar chaves diferentes para diferentes prefixos do Amazon S3URI. É a descrição dos materiais de criptografia retornados que são eventualmente armazenados com o objeto Amazon S3, em vez do materialsDescription valor que é gerado EMRFS e passado para o provedor. Ao descriptografar um objeto do Amazon S3, a descrição do material de criptografia é passada para a EncryptionMaterialsProvider classe, para que ela possa, novamente, retornar seletivamente a chave correspondente para descriptografar o objeto.

Uma implementação de EncryptionMaterialsProvider referência é fornecida abaixo. Outro provedor personalizado, EMRFSRSAEncryptionMaterialsProvider, está disponível em GitHub.

```
import com.amazonaws.services.s3.model.EncryptionMaterials;
import com.amazonaws.services.s3.model.EncryptionMaterialsProvider;
import com.amazonaws.services.s3.model.KMSEncryptionMaterials;
```

```
import org.apache.hadoop.conf.Configurable;
import org.apache.hadoop.conf.Configuration;

import java.util.Map;

/**
 * Provides KMSEncryptionMaterials according to Configuration
 */
public class MyEncryptionMaterialsProviders implements EncryptionMaterialsProvider,
 Configurable{
  private Configuration conf;
  private String kmsKeyId;
  private EncryptionMaterials encryptionMaterials;

  private void init() {
    this.kmsKeyId = conf.get("my.kms.key.id");
    this.encryptionMaterials = new KMSEncryptionMaterials(kmsKeyId);
  }

  @Override
  public void setConf(Configuration conf) {
    this.conf = conf;
    init();
  }

  @Override
  public Configuration getConf() {
    return this.conf;
  }

  @Override
  public void refresh() {

  }

  @Override
  public EncryptionMaterials getEncryptionMaterials(Map<String, String>
 materialsDescription) {
    return this.encryptionMaterials;
  }

  @Override
  public EncryptionMaterials getEncryptionMaterials() {
    return this.encryptionMaterials;
```

```
  }
}
```

Fornecendo certificados para criptografar dados em trânsito com a criptografia da Amazon EMR

Com a EMR versão 4.8.0 ou posterior da Amazon, você tem duas opções para especificar artefatos para criptografar dados em trânsito usando uma configuração de segurança:

- Você pode criar PEM certificados manualmente, incluí-los em um arquivo.zip e, em seguida, referenciar o arquivo.zip no Amazon S3.
- É possível implementar um provedor de certificados personalizado como uma classe Java. Você especifica o JAR arquivo do aplicativo no Amazon S3 e, em seguida, fornece o nome completo da classe do provedor conforme declarado no aplicativo. A classe deve implementar a TLSArtifactsProviderinterface disponível a partir da AWS SDK for Java versão 1.11.0.

A Amazon baixa EMR automaticamente os artefatos para cada nó no cluster e depois os usa para implementar os recursos de criptografia de código aberto em trânsito. Para obter mais informações sobre as opções disponíveis, consulte Criptografia em trânsito.

Usando PEM certificados

Quando você especifica um arquivo.zip para criptografia em trânsito, a configuração de segurança espera que PEM os arquivos dentro do arquivo.zip sejam nomeados exatamente como aparecem abaixo:

Certificados de criptografia em trânsito

Nome do arquivo	Obrigatório/opcional	Detalhes
privateKey.pem	Obrigatório	Chave privada
certificateChain.pem	Obrigatório	Cadeia de certificados
trustedCertificates.pem	Opcional	Necessário se o certificado fornecido não estiver assinado pela autoridade de certificação raiz confiável (AC) padrão Java ou por uma AC intermedi

Nome do arquivo	Obrigatório/opcional	Detalhes
		ária que possa estabelecer um vínculo com a AC raiz confiável Java padrão. A raiz confiável padrão do Java CAs pode ser encontrada em`jre/lib/security/cacerts` .

É provável que você queira configurar o PEM arquivo de chave privada para ser um certificado curinga que permita o acesso ao VPC domínio da Amazon no qual suas instâncias de cluster residem. Por exemplo, se o seu cluster reside em us-east-1 (Norte da Virgínia), você pode optar por especificar um nome comum na configuração do certificado que permita o acesso ao cluster, especificando `CN=*.ec2.internal` na definição de requerente do certificado. Se o seu cluster residir em us-west-2 (Oregon), poderá especificar `CN=*.us-west-2.compute.internal`.

Se o PEM arquivo fornecido no artefato de criptografia não tiver um caractere curinga no CN do domínio, você deverá alterar o valor de `hadoop.ssl.hostname.verifier` para. `ALLOW_ALL` Isso é feito com a classificação `core-site` ao enviar configurações para um cluster ou ao adicionar esse valor ao arquivo `core-site.xml`. Essa alteração é necessária porque o verificador de nome de host padrão não aceitará um nome de host sem curinga, resultando em um erro. Para obter mais informações sobre a configuração de EMR clusters em uma AmazonVPC, consulteConfigurar redes.

O exemplo a seguir demonstra como usar o Open SSL para gerar um certificado X.509 autoassinado com uma chave privada de 1024 bits. RSA A chave permite o acesso às instâncias de EMR cluster Amazon do emissor na região `us-west-2` (Oregon), conforme especificado pelo nome de *`*.us-west-2.compute.internal`* domínio como nome comum.

Outros itens de requerente opcionais como país (C), estado (S) e Localidade (L) são especificados. Como um certificado autoassinado é gerado, o segundo comando no exemplo copia o arquivo `certificateChain.pem` no arquivo `trustedCertificates.pem`. O terceiro comando usa `zip` para criar o arquivo `my-certs.zip` que contém os certificados.

> **Important**
>
> Este exemplo é apenas uma proof-of-concept demonstração. O uso de certificados autoassinados não é recomendado e apresenta um possível risco de segurança. Para

sistemas de produção, use uma autoridade de certificação (AC) confiável para emitir certificados.

```
$ openssl req -x509 -newkey rsa:1024 -keyout privateKey.pem -out certificateChain.pem
 -days 365 -nodes -subj '/C=US/ST=Washington/L=Seattle/O=MyOrg/OU=MyDept/CN=*.us-
west-2.compute.internal'
$ cp certificateChain.pem trustedCertificates.pem
$ zip -r -X my-certs.zip certificateChain.pem privateKey.pem trustedCertificates.pem
```

AWS Identity and Access Management para Amazon EMR

AWS Identity and Access Management (IAM) é uma ferramenta Serviço da AWS que ajuda o administrador a controlar com segurança o acesso aos AWS recursos. IAMos administradores controlam quem pode ser autenticado (conectado) e autorizado (tem permissões) para usar os recursos da AmazonEMR. IAMé um Serviço da AWS que você pode usar sem custo adicional.

Tópicos

Público

A forma como você usa AWS Identity and Access Management (IAM) difere, dependendo do trabalho que você faz na AmazonEMR.

Usuário do serviço — Se você usa o EMR serviço da Amazon para fazer seu trabalho, seu administrador fornecerá as credenciais e as permissões de que você precisa. À medida que você usa mais EMR recursos da Amazon para fazer seu trabalho, você pode precisar de permissões

adicionais. Entender como o acesso é gerenciado pode ajudar você a solicitar as permissões corretas ao seu administrador. Se você não conseguir acessar um recurso na AmazonEMR, consulteSolução de problemas de EMR identidade e acesso da Amazon.

Administrador de serviços — Se você é responsável pelos EMR recursos da Amazon em sua empresa, provavelmente tem acesso total à AmazonEMR. É seu trabalho determinar quais EMR recursos e recursos da Amazon seus usuários do serviço devem acessar. Em seguida, você deve enviar solicitações ao IAM administrador para alterar as permissões dos usuários do serviço. Revise as informações nesta página para entender os conceitos básicos doIAM. Para saber mais sobre como sua empresa pode usar IAM com a AmazonEMR, consulteComo a Amazon EMR trabalha com IAM.

IAMadministrador — Se você for IAM administrador, talvez queira saber detalhes sobre como criar políticas para gerenciar o acesso à AmazonEMR. Para ver exemplos de políticas EMR baseadas em identidade da Amazon que você pode usarIAM, consulte. Exemplos de políticas EMR baseadas em identidade da Amazon

Autenticando com identidades

A autenticação é como você faz login AWS usando suas credenciais de identidade. Você deve estar autenticado (conectado AWS) como IAM usuário ou assumindo uma IAM função. Usuário raiz da conta da AWS

Você pode entrar AWS como uma identidade federada usando credenciais fornecidas por meio de uma fonte de identidade. AWS IAM Identity Center Os usuários (do IAM Identity Center), a autenticação de login único da sua empresa e suas credenciais do Google ou do Facebook são exemplos de identidades federadas. Quando você entra como uma identidade federada, seu administrador configurou previamente a federação de identidades usando IAM funções. Ao acessar AWS usando a federação, você está assumindo indiretamente uma função.

Dependendo do tipo de usuário que você é, você pode entrar no AWS Management Console ou no portal de AWS acesso. Para obter mais informações sobre como fazer login em AWS, consulte Como fazer login Conta da AWS no Guia do Início de Sessão da AWS usuário.

Se você acessar AWS programaticamente, AWS fornece um kit de desenvolvimento de software (SDK) e uma interface de linha de comando (CLI) para assinar criptograficamente suas solicitações usando suas credenciais. Se você não usa AWS ferramentas, você mesmo deve assinar as solicitações. Para obter mais informações sobre como usar o método recomendado para você mesmo assinar solicitações, consulte Assinar AWS API solicitações no Guia IAM do usuário.

Independente do método de autenticação usado, também pode ser exigido que você forneça informações adicionais de segurança. Por exemplo, AWS recomenda que você use a autenticação multifator (MFA) para aumentar a segurança da sua conta. Para saber mais, consulte Autenticação multifator no Guia AWS IAM Identity Center do usuário e Uso da autenticação multifator (MFA) AWS no Guia do IAMusuário.

Conta da AWS usuário root

Ao criar uma Conta da AWS, você começa com uma identidade de login que tem acesso completo a todos Serviços da AWS os recursos da conta. Essa identidade é chamada de usuário Conta da AWS raiz e é acessada fazendo login com o endereço de e-mail e a senha que você usou para criar a conta. É altamente recomendável não usar o usuário raiz para tarefas diárias. Proteja as credenciais do usuário raiz e use-as para executar as tarefas que somente ele puder executar. Para ver a lista completa de tarefas que exigem que você faça login como usuário raiz, consulte Tarefas que exigem credenciais de usuário raiz no Guia do IAM usuário.

Identidade federada

Como prática recomendada, exija que usuários humanos, incluindo usuários que precisam de acesso de administrador, usem a federação com um provedor de identidade para acessar Serviços da AWS usando credenciais temporárias.

Uma identidade federada é um usuário do seu diretório de usuários corporativo, de um provedor de identidade da web AWS Directory Service, do diretório do Identity Center ou de qualquer usuário que acesse usando credenciais fornecidas Serviços da AWS por meio de uma fonte de identidade. Quando as identidades federadas acessam Contas da AWS, elas assumem funções, e as funções fornecem credenciais temporárias.

Para o gerenciamento de acesso centralizado, recomendamos usar o AWS IAM Identity Center. Você pode criar usuários e grupos no IAM Identity Center ou pode se conectar e sincronizar com um conjunto de usuários e grupos em sua própria fonte de identidade para uso em todos os seus Contas da AWS aplicativos. Para obter informações sobre o IAM Identity Center, consulte O que é o IAM Identity Center? no Guia do AWS IAM Identity Center usuário.

Grupos e usuários do IAM

Um IAMusuário é uma identidade dentro da sua Conta da AWS que tem permissões específicas para uma única pessoa ou aplicativo. Sempre que possível, recomendamos confiar em credenciais temporárias em vez de criar IAM usuários que tenham credenciais de longo prazo, como senhas

e chaves de acesso. No entanto, se você tiver casos de uso específicos que exijam credenciais de longo prazo com IAM os usuários, recomendamos que você alterne as chaves de acesso. Para obter mais informações, consulte Alterne as chaves de acesso regularmente para casos de uso que exigem credenciais de longo prazo no Guia do IAMusuário.

Um IAMgrupo é uma identidade que especifica uma coleção de IAM usuários. Não é possível fazer login como um grupo. É possível usar grupos para especificar permissões para vários usuários de uma vez. Os grupos facilitam o gerenciamento de permissões para grandes conjuntos de usuários. Por exemplo, você pode ter um grupo chamado IAMAdminse conceder a esse grupo permissões para administrar IAM recursos.

Usuários são diferentes de perfis. Um usuário é exclusivamente associado a uma pessoa ou a uma aplicação, mas um perfil pode ser assumido por qualquer pessoa que precisar dele. Os usuários têm credenciais permanentes de longo prazo, mas os perfis fornecem credenciais temporárias. Para saber mais, consulte Quando criar um IAM usuário (em vez de uma função) no Guia do IAM usuário.

IAMfunções

Uma IAMfunção é uma identidade dentro da sua Conta da AWS que tem permissões específicas. É semelhante a um IAM usuário, mas não está associado a uma pessoa específica. Você pode assumir temporariamente uma IAM função no AWS Management Console trocando de funções. Você pode assumir uma função chamando uma AWS API operação AWS CLI or ou usando uma personalizadaURL. Para obter mais informações sobre métodos de uso de funções, consulte Usando IAM funções no Guia IAM do usuário.

IAMfunções com credenciais temporárias são úteis nas seguintes situações:

- Acesso de usuário federado: para atribuir permissões a identidades federadas, você pode criar um perfil e definir permissões para ele. Quando uma identidade federada é autenticada, essa identidade é associada ao perfil e recebe as permissões definidas pelo mesmo. Para obter informações sobre funções para federação, consulte Criação de uma função para um provedor de identidade terceirizado no Guia IAM do usuário. Se você usa o IAM Identity Center, configura um conjunto de permissões. Para controlar o que suas identidades podem acessar após a autenticação, o IAM Identity Center correlaciona o conjunto de permissões a uma função em. IAM Para obter informações sobre conjuntos de permissões, consulte Conjuntos de Permissões no Manual do Usuário do AWS IAM Identity Center .
- Permissões temporárias IAM de IAM usuário — Um usuário ou função pode assumir uma IAM função para assumir temporariamente permissões diferentes para uma tarefa específica.

- Acesso entre contas — Você pode usar uma IAM função para permitir que alguém (um diretor confiável) em uma conta diferente acesse recursos em sua conta. Os perfis são a principal forma de conceder acesso entre contas. No entanto, com alguns Serviços da AWS, você pode anexar uma política diretamente a um recurso (em vez de usar uma função como proxy). Para saber a diferença entre funções e políticas baseadas em recursos para acesso entre contas, consulte Acesso a recursos entre contas IAM no Guia do IAM usuário.
- Acesso entre serviços — Alguns Serviços da AWS usam recursos em outros Serviços da AWS. Por exemplo, quando você faz uma chamada em um serviço, é comum que esse serviço execute aplicativos na Amazon EC2 ou armazene objetos no Amazon S3. Um serviço pode fazer isso usando as permissões do principal de chamada, usando um perfil de serviço ou um perfil vinculado a serviço.
  - Sessões de acesso direto (FAS) — Quando você usa um IAM usuário ou uma função para realizar ações em AWS, você é considerado principal. Ao usar alguns serviços, você pode executar uma ação que inicia outra ação em um serviço diferente. FASusa as permissões do diretor chamando um Serviço da AWS, combinadas com a solicitação Serviço da AWS para fazer solicitações aos serviços posteriores. FASas solicitações são feitas somente quando um serviço recebe uma solicitação que requer interações com outros Serviços da AWS ou com recursos para ser concluída. Nesse caso, você precisa ter permissões para executar ambas as ações. Para obter detalhes da política ao fazer FAS solicitações, consulte Encaminhar sessões de acesso.
  - Função de serviço — Uma função de serviço é uma IAMfunção que um serviço assume para realizar ações em seu nome. Um IAM administrador pode criar, modificar e excluir uma função de serviço internamenteIAM. Para obter mais informações, consulte Criação de uma função para delegar permissões a uma Serviço da AWS no Guia do IAM usuário.
  - Função vinculada ao serviço — Uma função vinculada ao serviço é um tipo de função de serviço vinculada a um. Serviço da AWS O serviço pode presumir a função de executar uma ação em seu nome. As funções vinculadas ao serviço aparecem em você Conta da AWS e são de propriedade do serviço. Um IAM administrador pode visualizar, mas não editar, as permissões das funções vinculadas ao serviço.
- Aplicativos em execução na Amazon EC2 — Você pode usar uma IAM função para gerenciar credenciais temporárias para aplicativos que estão sendo executados em uma EC2 instância e fazendo AWS CLI AWS API solicitações. Isso é preferível ao armazenamento de chaves de acesso na EC2 instância. Para atribuir uma AWS função a uma EC2 instância e disponibilizá-la para todos os aplicativos, você cria um perfil de instância anexado à instância. Um perfil de instância contém a função e permite que os programas em execução na EC2 instância recebam

credenciais temporárias. Para obter mais informações, consulte Como usar uma IAM função para conceder permissões a aplicativos executados em EC2 instâncias da Amazon no Guia IAM do usuário.

Para saber se usar IAM funções ou IAM usuários, consulte Quando criar uma IAM função (em vez de um usuário) no Guia do IAM usuário.

Gerenciando acesso usando políticas

Você controla o acesso AWS criando políticas e anexando-as a AWS identidades ou recursos. Uma política é um objeto AWS que, quando associada a uma identidade ou recurso, define suas permissões. AWS avalia essas políticas quando um principal (usuário, usuário raiz ou sessão de função) faz uma solicitação. As permissões nas políticas determinam se a solicitação será permitida ou negada. A maioria das políticas é armazenada AWS como JSON documentos. Para obter mais informações sobre a estrutura e o conteúdo dos documentos de JSON política, consulte Visão geral das JSON políticas no Guia IAM do usuário.

Os administradores podem usar AWS JSON políticas para especificar quem tem acesso ao quê. Ou seja, qual entidade principal pode executar ações em quais recursos e em que condições.

Por padrão, usuários e funções não têm permissões. Para conceder permissão aos usuários para realizar ações nos recursos de que precisam, um IAM administrador pode criar IAM políticas. O administrador pode então adicionar as IAM políticas às funções e os usuários podem assumir as funções.

IAMas políticas definem permissões para uma ação, independentemente do método usado para realizar a operação. Por exemplo, suponha que você tenha uma política que permite a ação `iam:GetRole`. Um usuário com essa política pode obter informações de função do AWS Management Console AWS CLI, do ou do AWS API.

Políticas baseadas em identidade

Políticas baseadas em identidade são documentos de políticas de JSON permissões que você pode anexar a uma identidade, como um IAM usuário, grupo de usuários ou função. Essas políticas controlam quais ações os usuários e perfis podem realizar, em quais recursos e em que condições. Para saber como criar uma política baseada em identidade, consulte Criação de IAM políticas no Guia do IAMusuário.

As políticas baseadas em identidade podem ser categorizadas ainda adicionalmente como políticas em linha ou políticas gerenciadas. As políticas em linha são anexadas diretamente a um único usuário, grupo ou perfil. As políticas gerenciadas são políticas autônomas que você pode associar a vários usuários, grupos e funções em seu Conta da AWS. As políticas AWS gerenciadas incluem políticas gerenciadas e políticas gerenciadas pelo cliente. Para saber como escolher entre uma política gerenciada ou uma política em linha, consulte Escolha entre políticas gerenciadas e políticas em linha no Guia do IAMusuário.

Políticas baseadas em recursos

Políticas baseadas em recursos são documentos JSON de política que você anexa a um recurso. Exemplos de políticas baseadas em recursos são políticas de confiança de IAM funções e políticas de bucket do Amazon S3. Em serviços que suportem políticas baseadas em recursos, os administradores de serviço podem usá-las para controlar o acesso a um recurso específico. Para o recurso ao qual a política está anexada, a política define quais ações um principal especificado pode executar nesse recurso e em que condições. Você deve especificar uma entidade principal em uma política baseada em recursos. Os diretores podem incluir contas, usuários, funções, usuários federados ou. Serviços da AWS

Políticas baseadas em recursos são políticas em linha localizadas nesse serviço. Você não pode usar políticas AWS gerenciadas de uma política baseada IAM em recursos.

Listas de controle de acesso (ACLs)

As listas de controle de acesso (ACLs) controlam quais diretores (membros da conta, usuários ou funções) têm permissões para acessar um recurso. ACLssão semelhantes às políticas baseadas em recursos, embora não usem o formato de documento JSON de política.

Amazon S3, AWS WAF, e Amazon VPC são exemplos de serviços que oferecem suporte. ACLs Para saber maisACLs, consulte a visão geral da lista de controle de acesso (ACL) no Guia do desenvolvedor do Amazon Simple Storage Service.

Outros tipos de política

AWS oferece suporte a tipos de políticas adicionais menos comuns. Esses tipos de política podem definir o máximo de permissões concedidas a você pelos tipos de política mais comuns.

- Limites de permissões — Um limite de permissões é um recurso avançado no qual você define as permissões máximas que uma política baseada em identidade pode conceder a uma IAM entidade

(IAMusuário ou função). É possível definir um limite de permissões para uma entidade. As permissões resultantes são a interseção das políticas baseadas em identidade de uma entidade com seus limites de permissões. As políticas baseadas em recurso que especificam o usuário ou o perfil no campo `Principal` não são limitadas pelo limite de permissões. Uma negação explícita em qualquer uma dessas políticas substitui a permissão. Para obter mais informações sobre limites de permissões, consulte Limites de permissões para IAM entidades no Guia IAM do usuário.

- Políticas de controle de serviço (SCPs) — SCPs são JSON políticas que especificam as permissões máximas para uma organização ou unidade organizacional (OU) em AWS Organizations. AWS Organizations é um serviço para agrupar e gerenciar centralmente vários Contas da AWS que sua empresa possui. Se você habilitar todos os recursos em uma organização, poderá aplicar políticas de controle de serviço (SCPs) a qualquer uma ou a todas as suas contas. Os SCP limites de permissões para entidades nas contas dos membros, incluindo cada uma Usuário raiz da conta da AWS. Para obter mais informações sobre Organizations eSCPs, consulte Políticas de controle de serviços no Guia AWS Organizations do Usuário.
- Políticas de sessão: são políticas avançadas que você transmite como um parâmetro quando cria de forma programática uma sessão temporária para um perfil ou um usuário federado. As permissões da sessão resultante são a interseção das políticas baseadas em identidade do usuário ou do perfil e das políticas de sessão. As permissões também podem ser provenientes de uma política baseada em atributo. Uma negação explícita em qualquer uma dessas políticas substitui a permissão. Para obter mais informações, consulte Políticas de sessão no Guia IAM do usuário.

Vários tipos de política

Quando vários tipos de política são aplicáveis a uma solicitação, é mais complicado compreender as permissões resultantes. Para saber como AWS determinar se uma solicitação deve ser permitida quando vários tipos de política estão envolvidos, consulte Lógica de avaliação de políticas no Guia IAM do usuário.

Como a Amazon EMR trabalha com IAM

Antes de usar IAM para gerenciar o acesso à AmazonEMR, saiba quais IAM recursos estão disponíveis para uso com a AmazonEMR.

IAMrecursos que você pode usar com a Amazon EMR

IAMrecurso	EMRSuporte da Amazon
Políticas baseadas em identidade	Sim
Políticas baseadas em atributos	Sim
Ações das políticas	Sim
Atributos de políticas	Sim
Chaves de condição de políticas	Sim
ACLs	Não
ABAC(tags nas políticas)	Sim
Credenciais temporárias	Sim
Permissões de entidade principal	Sim
Perfis de serviço	Não
Funções vinculadas ao serviço	Sim

Para ter uma visão geral de como a Amazon EMR e outros AWS serviços funcionam com a maioria dos IAM recursos, consulte AWS os serviços que funcionam com IAM no Guia do IAM usuário.

Políticas baseadas em identidade para a Amazon EMR

Compatível com políticas baseadas em identidade: Sim

Políticas baseadas em identidade são documentos de políticas de JSON permissões que você pode anexar a uma identidade, como um IAM usuário, grupo de usuários ou função. Essas políticas controlam quais ações os usuários e perfis podem realizar, em quais recursos e em que condições. Para saber como criar uma política baseada em identidade, consulte Criação de IAM políticas no Guia do IAMusuário.

Com políticas IAM baseadas em identidade, você pode especificar ações e recursos permitidos ou negados, bem como as condições sob as quais as ações são permitidas ou negadas. Você não

pode especificar a entidade principal em uma política baseada em identidade porque ela se aplica ao usuário ou perfil ao qual ela está anexada. Para saber mais sobre todos os elementos que você pode usar em uma JSON política, consulte a referência IAM JSON de elementos de política no Guia IAM do usuário.

Exemplos de políticas baseadas em identidade para a Amazon EMR

Para ver exemplos de políticas EMR baseadas em identidade da Amazon, consulte. Exemplos de políticas EMR baseadas em identidade da Amazon

Políticas baseadas em recursos na Amazon EMR

Compatível com políticas baseadas em recursos: sim

Políticas baseadas em recursos são documentos JSON de política que você anexa a um recurso. Exemplos de políticas baseadas em recursos são políticas de confiança de IAM funções e políticas de bucket do Amazon S3. Em serviços que suportem políticas baseadas em recursos, os administradores de serviço podem usá-las para controlar o acesso a um recurso específico. Para o recurso ao qual a política está anexada, a política define quais ações um principal especificado pode executar nesse recurso e em que condições. Você deve especificar uma entidade principal em uma política baseada em recursos. Os diretores podem incluir contas, usuários, funções, usuários federados ou. Serviços da AWS

Para habilitar o acesso entre contas, você pode especificar uma conta ou IAM entidades inteiras em outra conta como principal em uma política baseada em recursos. Adicionar uma entidade principal entre contas à política baseada em recurso é apenas metade da tarefa de estabelecimento da relação de confiança. Quando o principal e o recurso são diferentes Contas da AWS, um IAM administrador na conta confiável também deve conceder permissão à entidade principal (usuário ou função) para acessar o recurso. Eles concedem permissão ao anexar uma política baseada em identidade para a entidade. No entanto, se uma política baseada em recurso conceder acesso a uma entidade principal na mesma conta, nenhuma política baseada em identidade adicional será necessária. Para obter mais informações, consulte Acesso a recursos entre contas IAM no Guia do IAM usuário.

Ações políticas para a Amazon EMR

Compatível com ações de políticas: Sim

Os administradores podem usar AWS JSON políticas para especificar quem tem acesso ao quê. Ou seja, qual entidade principal pode executar ações em quais recursos, e em que condições.

O `Action` elemento de uma JSON política descreve as ações que você pode usar para permitir ou negar acesso em uma política. As ações de política geralmente têm o mesmo nome da AWS API operação associada. Há algumas exceções, como ações somente de permissão que não têm uma operação correspondente. API Algumas operações também exigem várias ações em uma política. Essas ações adicionais são chamadas de ações dependentes.

Incluem ações em uma política para conceder permissões para executar a operação associada.

Para ver uma lista de EMR ações da Amazon, consulte Ações, recursos e chaves de condição para a Amazon EMR na Referência de autorização de serviço.

As ações políticas na Amazon EMR usam o seguinte prefixo antes da ação:

```
EMR
```

Para especificar várias ações em uma única instrução, separe-as com vírgulas.

```
"Action": [
      "EMR:action1",
      "EMR:action2"
         ]
```

Para ver exemplos de políticas EMR baseadas em identidade da Amazon, consulte. Exemplos de políticas EMR baseadas em identidade da Amazon

Recursos de políticas para a Amazon EMR

Compatível com recursos de políticas: Sim

Os administradores podem usar AWS JSON políticas para especificar quem tem acesso ao quê. Ou seja, qual entidade principal pode executar ações em quais recursos, e em que condições.

O elemento `Resource` JSON de política especifica o objeto ou objetos aos quais a ação se aplica. As instruções devem incluir um elemento `Resource` ou `NotResource`. Como prática recomendada, especifique um recurso usando seu Amazon Resource Name (ARN). Isso pode ser feito para ações que oferecem compatibilidade com um tipo de recurso específico, conhecido como permissões em nível de recurso.

Para ações que não oferecem compatibilidade com permissões em nível de recurso, como operações de listagem, use um curinga (*) para indicar que a instrução se aplica a todos os recursos.

```
"Resource": "*"
```

Para ver uma lista dos tipos de EMR recursos da Amazon e seusARNs, consulte Recursos definidos pela Amazon EMR na Referência de autorização de serviço. Para saber quais ações você pode especificar para cada recurso, consulte Ações, recursos e chaves de condição para a Amazon EMR. ARN

Para ver exemplos de políticas EMR baseadas em identidade da Amazon, consulte. Exemplos de políticas EMR baseadas em identidade da Amazon

Chaves de condição de política para a Amazon EMR

Compatível com chaves de condição de política específicas de serviço: Sim

Os administradores podem usar AWS JSON políticas para especificar quem tem acesso ao quê. Ou seja, qual entidade principal pode executar ações em quais recursos, e em que condições.

O elemento `Condition` (ou bloco `Condition`) permite que você especifique condições nas quais uma instrução estiver em vigor. O elemento `Condition` é opcional. É possível criar expressões condicionais que usem agentes de condição, como “igual a” ou “menor que”, para fazer a condição da política corresponder aos valores na solicitação.

Se você especificar vários elementos `Condition` em uma instrução ou várias chaves em um único `Condition` elemento, a AWS os avaliará usando uma operação lógica `AND`. Se você especificar vários valores para uma única chave de condição, AWS avalia a condição usando uma `OR` operação lógica. Todas as condições devem ser atendidas antes que as permissões da instrução sejam concedidas.

Você também pode usar variáveis de espaço reservado ao especificar condições. Por exemplo, você pode conceder permissão a um IAM usuário para acessar um recurso somente se ele estiver marcado com o nome de IAM usuário. Para obter mais informações, consulte elementos de IAM política: variáveis e tags no Guia IAM do usuário.

AWS suporta chaves de condição globais e chaves de condição específicas do serviço. Para ver todas as chaves de condição AWS globais, consulte as chaves de contexto de condição AWS global no Guia IAM do usuário.

Para ver uma lista das chaves de EMR condição da Amazon e saber quais ações e recursos você pode usar uma chave de condição, consulte Ações, recursos e chaves de condição para a Amazon EMR na Referência de autorização de serviço.

Para ver exemplos de políticas EMR baseadas em identidade da Amazon, consulte. Exemplos de políticas EMR baseadas em identidade da Amazon

Listas de controle de acesso (ACLs) na Amazon EMR

SuportesACLs: Não

As listas de controle de acesso (ACLs) controlam quais diretores (membros da conta, usuários ou funções) têm permissões para acessar um recurso. ACLssão semelhantes às políticas baseadas em recursos, embora não usem o formato de documento JSON de política.

Controle de acesso baseado em atributos (ABAC) com a Amazon EMR

Suportes ABAC (tags nas políticas)	Sim

O controle de acesso baseado em atributos (ABAC) é uma estratégia de autorização que define permissões com base em atributos. Em AWS, esses atributos são chamados de tags. Você pode anexar tags a IAM entidades (usuários ou funções) e a vários AWS recursos. Marcar entidades e recursos é a primeira etapa doABAC. Em seguida, você cria ABAC políticas para permitir operações quando a tag do diretor corresponde à tag do recurso que ele está tentando acessar.

ABACé útil em ambientes que estão crescendo rapidamente e ajuda em situações em que o gerenciamento de políticas se torna complicado.

Para controlar o acesso baseado em tags, forneça informações sobre as tags no elemento de condição de uma política usando as `aws:ResourceTag/`*`key-name`*, `aws:RequestTag/`*`key-name`* ou chaves de condição `aws:TagKeys`.

Se um serviço oferecer suporte às três chaves de condição para cada tipo de recurso, o valor será Sim para o serviço. Se um serviço oferecer suporte às três chaves de condição somente para alguns tipos de recursos, o valor será Parcial

Para obter mais informações sobreABAC, consulte O que éABAC? no Guia do IAM usuário. Para ver um tutorial com etapas de configuraçãoABAC, consulte Usar controle de acesso baseado em atributos (ABAC) no Guia do IAMusuário.

Usando credenciais temporárias com a Amazon EMR

Compatível com credenciais temporárias: Sim

Alguns Serviços da AWS não funcionam quando você faz login usando credenciais temporárias. Para obter informações adicionais, incluindo quais Serviços da AWS funcionam com credenciais temporárias, consulte Serviços da AWS esse trabalho IAM no Guia do IAM usuário.

Você está usando credenciais temporárias se fizer login AWS Management Console usando qualquer método, exceto um nome de usuário e senha. Por exemplo, quando você acessa AWS usando o link de login único (SSO) da sua empresa, esse processo cria automaticamente credenciais temporárias. Você também cria automaticamente credenciais temporárias quando faz login no console como usuário e, em seguida, alterna perfis. Para obter mais informações sobre a troca de funções, consulte Alternando para uma função (console) no Guia IAM do usuário.

Você pode criar manualmente credenciais temporárias usando o AWS CLI ou AWS API. Em seguida, você pode usar essas credenciais temporárias para acessar AWS. AWS recomenda que você gere credenciais temporárias dinamicamente em vez de usar chaves de acesso de longo prazo. Para obter mais informações, consulte Credenciais de segurança temporárias emIAM.

Permissões principais entre serviços para a Amazon EMR

Suporta sessões de acesso direto (FAS): Sim

Quando você usa um IAM usuário ou uma função para realizar ações em AWS, você é considerado principal. Ao usar alguns serviços, você pode executar uma ação que inicia outra ação em um serviço diferente. FASusa as permissões do diretor chamando um Serviço da AWS, combinadas com a solicitação Serviço da AWS para fazer solicitações aos serviços posteriores. FASas solicitações são feitas somente quando um serviço recebe uma solicitação que requer interações com outros Serviços da AWS ou com recursos para ser concluída. Nesse caso, você precisa ter permissões para executar ambas as ações. Para obter detalhes da política ao fazer FAS solicitações, consulte Encaminhar sessões de acesso.

Funções de serviço para a Amazon EMR

Oferece suporte a perfis de serviço	Não

Funções vinculadas a serviços para a Amazon EMR

Oferece suporte a perfis vinculados ao serviço	Sim

Para obter detalhes sobre como criar ou gerenciar funções vinculadas a serviços, consulte AWS serviços que funcionam com. IAM Encontre um serviço na tabela que inclua um `Yes` na coluna Função vinculada ao serviço. Escolha o link Sim para visualizar a documentação do perfil vinculado a serviço desse serviço.

Use tags de cluster e notebook com IAM políticas para controle de acesso

A permissão para EMR ações da Amazon associadas a EMR notebooks e EMR clusters pode ser ajustada usando o controle de acesso baseado em tags com políticas baseadas em identidade. IAM Você pode usar chaves de condição em um elemento `Condition` (também denominado bloco `Condition`) para permitir determinadas ações somente quando um bloco de anotações, um cluster ou ambos têm uma chave de tag ou uma combinação de chave-valor. Você também pode limitar a `CreateEditor` ação (que cria um EMR notebook) e a `RunJobFlow` ação (que cria um cluster) para que uma solicitação de tag seja enviada quando o recurso for criado.

Na AmazonEMR, as chaves de condição que podem ser usadas em um `Condition` elemento se aplicam somente às EMR API ações da Amazon em que `ClusterID` ou `NotebookID` é um parâmetro de solicitação obrigatório. Por exemplo, a ModifyInstanceGroupsação não oferece suporte a chaves de contexto porque `ClusterID` é um parâmetro opcional.

Quando você cria um EMR notebook, uma tag padrão é aplicada com uma string chave `creatorUserId` definida para o valor da ID do IAM usuário que criou o notebook. Isso é útil para limitar as ações permitidas para o bloco de anotações apenas ao criador.

As seguintes chaves de condição estão disponíveis na AmazonEMR:

- Use a chave de contexto de condição `elasticmapreduce:ResourceTag/`*`TagKeyString`*, para permitir ou negar ações do usuário em clusters ou blocos de anotações com tags que tenham a *`TagKeyString`* especificada. Se uma ação passar `ClusterID` e `NotebookID`, a condição se aplicará ao cluster e ao bloco de anotações. Isso significa que ambos os recursos devem ter a string de chave de tag ou uma combinação de chave-valor que você especificar. Você pode usar o elemento `Resource` para limitar a declaração para que ela se aplique apenas a clusters ou blocos de anotações, conforme necessário. Para obter mais informações, consulte Exemplos de políticas EMR baseadas em identidade da Amazon.

- Use a chave de contexto de `elasticmapreduce:RequestTag/TagKeyString` condição para exigir uma tag específica com ações/chamadasAPI. Por exemplo, é possível usar essa chave de contexto de condição juntamente com a ação `CreateEditor` para exigir que uma chave com `TagKeyString` seja aplicada a um bloco de anotações quando ele é criado.

Exemplos

Para ver uma lista de EMR ações da Amazon, consulte Ações definidas pela Amazon EMR no Guia IAM do usuário.

Funções de tempo de execução para Amazon EMR Steps

Uma função de tempo de execução é uma função AWS Identity and Access Management (IAM) que você pode especificar ao enviar um trabalho ou uma consulta para um EMR cluster da Amazon. O trabalho ou consulta que você envia ao seu EMR cluster da Amazon usa a função de tempo de execução para acessar AWS recursos, como objetos no Amazon S3. Você pode especificar funções de tempo de execução com a Amazon EMR para trabalhos do Spark e do Hive.

Você também pode especificar funções de tempo de execução ao se conectar aos EMR clusters da Amazon em Amazon SageMaker e ao anexar um Amazon EMR Studio Workspace a um EMR cluster. Para obter mais informações, consulte Connect to an Amazon EMR cluster from Studio Execute um EMR Studio Workspace com uma função de tempo de execução e.

Anteriormente, EMR os clusters da Amazon executavam EMR trabalhos ou consultas da Amazon com permissões com base na IAM política anexada ao perfil de instância que você usou para iniciar o cluster. Isso significava que as políticas precisavam conter a união de todas as permissões para todos os trabalhos e consultas executados em um EMR cluster da Amazon. Com as funções de tempo de execução, agora você pode gerenciar o controle de acesso para cada trabalho ou consulta individualmente, em vez de compartilhar o perfil da EMR instância Amazon do cluster.

Em EMR clusters da Amazon com funções de tempo de execução, você também pode aplicar controle de acesso AWS Lake Formation baseado às tarefas e consultas do Spark, Hive e Presto em seus lagos de dados. Para saber mais sobre como fazer a integração com AWS Lake Formation, consulteIntegre a Amazon EMR com AWS Lake Formation.

> **Note**
>
> Quando você especifica uma função de tempo de execução para uma EMR etapa da Amazon, os trabalhos ou consultas que você envia só podem acessar AWS recursos que as

políticas anexadas à função de tempo de execução permitem. Esses trabalhos e consultas não podem acessar o Serviço de Metadados de Instância nas EC2 instâncias do cluster nem usar o perfil de EC2 instância do cluster para acessar quaisquer AWS recursos.

Pré-requisitos para lançar um EMR cluster da Amazon com uma função de tempo de execução

Tópicos

Etapa 1: definir configurações de segurança na Amazon EMR

Use a JSON estrutura a seguir para criar uma configuração de segurança no AWS Command Line Interface (AWS CLI) e `EnableApplicationScopedIAMRole` definir como`true`. Para obter mais informações sobre configurações de segurança, consulte Usar configurações de segurança para definir a segurança do cluster.

```
{
    "AuthorizationConfiguration":{
        "IAMConfiguration":{
            "EnableApplicationScopedIAMRole":true
        }
    }
}
```

É recomendável habilitar sempre as opções de criptografia em trânsito na configuração de segurança, para que os dados transferidos pela Internet sejam criptografados, em vez de em texto sem formatação. Você pode ignorar essas opções se não quiser se conectar aos EMR clusters da Amazon com funções de tempo de execução do SageMaker Runtime Studio ou do EMR Studio. Para configurar a criptografia de dados, consulte Configure data encryption.

Como alternativa, você pode criar uma configuração de segurança com configurações personalizadas usando o AWS Management Console.

Etapa 2: configurar um perfil de EC2 instância para o EMR cluster da Amazon

EMROs clusters da Amazon usam a função de perfil de EC2 instância da Amazon para assumir as funções de tempo de execução. Para usar funções de tempo de execução com EMR etapas da Amazon, adicione as seguintes políticas à IAM função que você planeja usar como função do perfil da instância. Para adicionar políticas a uma IAM função ou editar uma política embutida ou gerenciada existente, consulte Adicionar e remover permissões de IAM identidade.

```
{
    "Version":"2012-10-17",
    "Statement":[
        {
            "Sid":"AllowRuntimeRoleUsage",
            "Effect":"Allow",
            "Action":[
                "sts:AssumeRole",
                "sts:TagSession"
            ],
            "Resource":[
                <runtime-role-ARN>
            ]
        }
    ]
}
```

Etapa3: configurar uma política de confiança

Para cada IAM função que você planeja usar como função de tempo de execução, defina a seguinte política de confiança, `EMR_EC2_DefaultRole` substituindo-a pela função de perfil da instância. Para modificar a política de confiança de uma IAM função, consulte Modificação da política de confiança de uma função.

```
{
    "Sid":"AllowAssumeRole",
    "Effect":"Allow",
    "Principal":{
        "AWS":"arn:aws:iam::<AWS_ACCOUNT_ID>:role/EMR_EC2_DefaultRole"
    },
    "Action":"sts:AssumeRole"
}
```

Lance um EMR cluster da Amazon com controle de acesso baseado em funções

Depois de definir suas configurações, você pode iniciar um EMR cluster da Amazon com a configuração de segurança deEtapa 1: definir configurações de segurança na Amazon EMR. Para usar funções de tempo de execução com EMR as etapas da Amazon, use o rótulo de lançamento `emr-6.7.0` ou posterior e selecione Hive, Spark ou ambos como seu aplicativo de cluster. Para se conectar a partir do SageMaker Studio, use release `emr-6.9.0` ou posterior e selecione Livy, Spark, Hive ou Presto como seu aplicativo de cluster. Para obter instruções sobre como iniciar seu cluster, consulte Especificar uma configuração de segurança para um cluster.

Envie trabalhos do Spark usando as etapas da Amazon EMR

Veja a seguir um exemplo de como executar o HdfsTest exemplo incluído no Apache Spark. Essa API chamada só será bem-sucedida se a função de EMR tempo de execução fornecida pela Amazon puder acessar o. `S3_LOCATION`

```
RUNTIME_ROLE_ARN=<runtime-role-arn>
S3_LOCATION=<s3-path>
REGION=<aws-region>
CLUSTER_ID=<cluster-id>

aws emr add-steps --cluster-id $CLUSTER_ID \
--steps '[{ "Name": "Spark Example", "ActionOnFailure": "CONTINUE","HadoopJarStep":
 { "Jar":"command-runner.jar","Args" : ["spark-example","HdfsTest",
 "$S3_LOCATION"] } }]' \
--execution-role-arn $RUNTIME_ROLE_ARN \
--region $REGION
```

> **Note**
>
> Recomendamos que você desative o SSH acesso ao EMR cluster da Amazon e permita que somente EMR `AddJobFlowSteps` API a Amazon acesse o cluster.

Envie trabalhos do Hive usando as etapas da Amazon EMR

O exemplo a seguir usa o Apache Hive com EMR as etapas da Amazon para enviar um trabalho para executar o `QUERY_FILE.hql` arquivo. Essa consulta só será terá êxito se o perfil de runtime fornecido puder acessar o caminho do Amazon S3 do arquivo de consulta.

```
RUNTIME_ROLE_ARN=<runtime-role-arn>
REGION=<aws-region>
CLUSTER_ID=<cluster-id>

aws emr add-steps --cluster-id $CLUSTER_ID \
--steps '[{ "Name": "Run hive query using command-runner.jar - simple
 select","ActionOnFailure":"CONTINUE","HadoopJarStep": { "Jar": "command-
runner.jar","Args" :["hive -
f","s3://DOC_EXAMPLE_BUCKET/QUERY_FILE.hql"] } }]' \
--execution-role-arn $RUNTIME_ROLE_ARN \
--region $REGION
```

Conecte-se aos EMR clusters da Amazon com funções de tempo de execução a partir de um notebook SageMaker Studio

Você pode aplicar funções EMR de tempo de execução da Amazon às consultas que você executa nos EMR clusters da Amazon a partir do SageMaker Studio. Para isso, siga as etapas a seguir.

1. Siga as instruções em Inicie o Amazon SageMaker Studio para criar um SageMaker Studio.
2. Na interface do usuário do SageMaker Studio, inicie um notebook com kernels compatíveis. Por exemplo, inicie uma SparkMagic imagem com um PySpark kernel.
3. Escolha um EMR cluster da Amazon no SageMaker Studio e, em seguida, escolha Connect.
4. Escolha um perfil de runtime e escolha Conectar.

Isso criará uma célula de SageMaker notebook com comandos mágicos para se conectar ao seu EMR cluster da Amazon com a função de EMR tempo de execução da Amazon escolhida. Na célula do caderno, você pode inserir e executar consultas com perfil de runtime e controle de acesso baseado no Lake Formation. Para um exemplo mais detalhado, consulte Aplicar controles refinados de acesso a dados com e a AWS Lake Formation Amazon do EMR Amazon Studio. SageMaker

Controle o acesso à função de EMR tempo de execução da Amazon

Você pode controlar o acesso ao perfil de runtime usando a chave de condição `elasticmapreduce:ExecutionRoleArn`. A política a seguir permite que um IAM diretor use uma IAM função chamada`Caller`, ou qualquer IAM função que comece com a string`CallerTeamRole`, como a função de tempo de execução.

> **Important**
>
> Você deve criar uma condição com base na chave de `elasticmapreduce:ExecutionRoleArn` contexto ao conceder a um chamador acesso para chamar o `AddJobFlowSteps` ou `GetClusterSessionCredentials`APIs, conforme mostra o exemplo a seguir.

```
{
    "Sid":"AddStepsWithSpecificExecRoleArn",
    "Effect":"Allow",
    "Action":[
        "elasticmapreduce:AddJobFlowSteps"
    ],
    "Resource":"*",
    "Condition":{
        "StringEquals":{
            "elasticmapreduce:ExecutionRoleArn":[
                "arn:aws:iam::<AWS_ACCOUNT_ID>:role/Caller"
            ]
        },
        "StringLike":{
            "elasticmapreduce:ExecutionRoleArn":[
                "arn:aws:iam::<AWS_ACCOUNT_ID>:role/CallerTeamRole*"
            ]
        }
    }
}
```

Estabeleça confiança entre as funções de tempo de execução e os EMR clusters da Amazon

EMRA Amazon gera um identificador exclusivo `ExternalId` para cada configuração de segurança com autorização de função de tempo de execução ativada. Essa autorização permite que cada usuário tenha um conjunto de perfil de runtime para usar nos clusters que pertencem a eles. Por exemplo, em uma empresa, cada departamento pode usar o próprio ID externo para atualizar a política de confiança em seu próprio conjunto de perfis de runtime.

Você pode encontrar o ID externo com a Amazon EMR `DescribeSecurityConfiguration`API, conforme mostrado no exemplo a seguir.

```
aws emr describe-security-configuration --name 'iamconfig-with-lf'{"Name": "iamconfig-
with-lf",
    "SecurityConfiguration":
        "{\"AuthorizationConfiguration\":{\"IAMConfiguration\":
{\"EnableApplicationScopedIAMRole\
        ":true,\"ApplicationScopedIAMRoleConfiguration\":{\"PropagateSourceIdentity
\":true,\"Exter
        nalId\":\"FXH5TSACFDWUCDSR3YQE2O7ETPUSM4OBCGLYWODSCUZDNZ4Y\"}},\"Lake
        FormationConfiguration\":{\"AuthorizedSessionTagValue\":\"Amazon EMR\"}}}",
    "CreationDateTime": "2022-06-03T12:52:35.308000-07:00"
}
```

Para obter informações sobre como usar uma ID externa, consulte Como usar uma ID externa ao conceder acesso aos seus AWS recursos a terceiros.

Auditoria

Para monitorar e controlar as ações que os usuários finais realizam com as IAM funções, você pode ativar o recurso de identidade de origem. Para saber mais sobre a identidade de origem, consulte Monitorar e controlar ações realizadas com perfis assumidos.

Para rastrear a identidade de origem, defina `ApplicationScopedIAMRoleConfiguration/PropagateSourceIdentity` como `true` em sua configuração de segurança, como mostrado a seguir.

```
{
    "AuthorizationConfiguration":{
        "IAMConfiguration":{
            "EnableApplicationScopedIAMRole":true,
            "ApplicationScopedIAMRoleConfiguration":{
                "PropagateSourceIdentity":true
            }
        }
    }
}
```

Quando você define como`true`, `PropagateSourceIdentity` a Amazon EMR aplica a identidade de origem das credenciais de chamada a um trabalho ou sessão de consulta que você cria com a função de tempo de execução. Se nenhuma identidade de origem estiver presente nas credenciais de chamada, a Amazon EMR não definirá a identidade de origem.

Para usar essa propriedade, forneça permissões `sts:SetSourceIdentity` ao perfil de instância, como mostrado a seguir.

```
{ // PropagateSourceIdentity statement
    "Sid":"PropagateSourceIdentity",
    "Effect":"Allow",
    "Action":"sts:SetSourceIdentity",
    "Resource":[
        <runtime-role-ARN>
    ],
    "Condition":{
        "StringEquals":{
            "sts:SourceIdentity":<source-identity>
        }
    }
}
```

Também é necessário adicionar a instrução `AllowSetSourceIdentity` à política de confiança de seus perfis de runtime.

```
{ // AllowSetSourceIdentity statement
    "Sid":"AllowSetSourceIdentity",
    "Effect":"Allow",
    "Principal":{
        "AWS":"arn:aws:iam::<AWS_ACCOUNT_ID>:role/EMR_EC2_DefaultRole"
    },
    "Action":[
        "sts:SetSourceIdentity",
        "sts:AssumeRole"
    ],
    "Condition":{
        "StringEquals":{
            "sts:SourceIdentity":<source-identity>
        }
    }
}
```

Considerações adicionais

> **Note**
>
> Com o EMR lançamento da Amazon`emr-6.9.0`, você pode enfrentar falhas intermitentes ao se conectar aos EMR clusters da Amazon a partir do SageMaker Studio. Para resolver esse problema, instale o patch com uma ação de bootstrap ao iniciar o cluster. Para obter detalhes do patch, consulte Problemas conhecidos da EMR versão 6.9.0 da Amazon.

Além disso, considere o seguinte ao configurar funções de tempo de execução para a AmazonEMR.

- A Amazon EMR oferece suporte a funções de tempo de execução em todos os comerciais Regiões da AWS.
- EMRAs etapas da Amazon oferecem suporte a trabalhos do Apache Spark e do Apache Hive com funções de tempo de execução quando você usa a versão release ou posterior. `emr-6.7.0`
- SageMaker O Studio oferece suporte a consultas Spark, Hive e Presto com funções de tempo de execução quando você usa a versão ou posterior. `emr-6.9.0`
- Os seguintes kernels de notebook SageMaker oferecem suporte a funções de tempo de execução:
  - DataScience — Kernel Python 3
  - DataScience 2.0 — Kernel do Python 3
  - DataScience 3.0 — Kernel do Python 3
  - SparkAnalytics 1.0 — SparkMagic e PySpark grãos
  - SparkAnalytics 2.0 — SparkMagic e PySpark grãos
  - SparkMagic — PySpark núcleo
- A Amazon EMR oferece suporte a etapas que `RunJobFlow` são usadas somente no momento da criação do cluster. Isso API não suporta funções de tempo de execução.
- A Amazon EMR não oferece suporte a funções de tempo de execução em clusters que você configura para serem altamente disponíveis.
- Você deve escapar dos argumentos do comando Bash ao executar comandos com o `command-runner.jar` JAR arquivo:

```
aws emr add-steps --cluster-id <cluster-id> --steps '[{"Name":"sample-
step","ActionOnFailure":"CONTINUE","Jar":"command-runner.jar","Properties":"","Args":
```

```
["bash","-c","\"aws s3 ls\""],"Type":"CUSTOM_JAR"}]' --execution-role-
arn <IAM_ROLE_ARN>
```

- As funções de tempo de execução não oferecem suporte para controlar o acesso aos recursos do cluster, como HDFS e. HMS

Configure funções IAM de serviço para EMR permissões da Amazon para AWS serviços e recursos

A Amazon EMR e aplicativos como o Hadoop e o Spark precisam de permissões para acessar outros AWS recursos e realizar ações quando eles são executados. Cada cluster na Amazon EMR deve ter uma função de serviço e uma função para o perfil de EC2 instância da Amazon. Para obter mais informações, consulte IAMFunções e Como usar perfis de instância no Guia IAM do usuário. As IAM políticas anexadas a essas funções fornecem permissões para o cluster interoperar com outros AWS serviços em nome de um usuário.

Uma função adicional, a função Auto Scaling, é necessária se seu cluster usa escalabilidade automática na Amazon. EMR A função AWS de serviço para EMR Notebooks é necessária se você usa EMR Notebooks.

EMRA Amazon fornece funções padrão e políticas gerenciadas padrão que determinam as permissões para cada função. As políticas gerenciadas são criadas e mantidas por AWS, portanto, são atualizadas automaticamente se os requisitos de serviço mudarem. Consulte as políticas AWS gerenciadas no Guia IAM do usuário.

Se você estiver criando um cluster ou notebook pela primeira vez em uma conta, as funções para a Amazon ainda EMR não existem. Depois de criá-las, você pode visualizar as funções, as políticas associadas a elas e as permissões permitidas ou negadas pelas políticas no IAM console (https://console.aws.amazon.com/iam/). Você pode especificar funções padrão para EMR a Amazon criar e usar, você pode criar suas próprias funções e especificá-las individualmente ao criar um cluster para personalizar permissões, e você pode especificar funções padrão a serem usadas ao criar um cluster usando AWS CLI o. Para obter mais informações, consulte Personalize IAM funções.

Modificando políticas baseadas em identidade para obter permissões para passar funções de serviço para a Amazon EMR

As políticas EMR gerenciadas padrão de permissões completas da Amazon incorporam configurações `iam:PassRole` de segurança, incluindo as seguintes:

- `iam:PassRole`permissões somente para EMR funções padrão específicas da Amazon.
- `iam:PassedToService`condições que permitem que você use a política somente com AWS serviços específicos, como `elasticmapreduce.amazonaws.com` `ec2.amazonaws.com` e.

Você pode ver a JSON versão das políticas A mazonEMRFull AccessPolicy _v2 e A mazonEMRServicePolicy_v2 no console. IAM É recomendável criar novos clusters com políticas gerenciadas v2.

Resumo do perfil de serviço

A tabela a seguir lista as funções IAM de serviço associadas à Amazon EMR para referência rápida.

Função	Perfil padrão	Descrição	Política gerenciada padrão
Função de serviço para a Amazon EMR (EMRfunção)	`EMR_DefaultRole_V2`	Permite que EMR a Amazon chame outros AWS serviços em seu nome ao provisionar recursos e realizar ações de nível de serviço. Essa função é necessária para todos os clusters.	`AmazonEMRServicePolicy_v2` **Important** É necessário ter um perfil vinculado ao serviço para solicitar instâncias spot. Se essa função não existir, a função de EMR serviço da Amazon deve ter permissão para criá-la ou ocorrerá

Função	Perfil padrão	Descrição	Política gerenciada padrão
			um erro de permissão. Se você pretende solicitar instância s spot, é necessári o atualizar essa política para incluir uma instrução que permita a criação desse perfil vinculado ao serviço. Para obter mais informaçõ es, consulte Função de serviço para a Amazon EMR (EMRfunçã o) a função vinculada ao serviço para solicitações de instância s spot no Guia EC2 do usuário da Amazon.

Função	Perfil padrão	Descrição	Política gerenciada padrão
Função de serviço para EC2 instâncias de cluster (perfil de EC2 instância)	`EMR_EC2_DefaultRole`	Os processos de aplicativos que são executados no ecossistema do Hadoop em instâncias de cluster usam essa função quando chamam outros AWS serviços. Para acessar dados no Amazon S3 usandoEMRFS, você pode especificar diferentes funções a serem assumidas com base na localização dos dados no Amazon S3. Por exemplo, múltiplas equipes podem acessar uma única “conta de armazenamento” de dados do Amazon S3. Para obter mais informações, consulte Configurar IAM funções para EMRFS solicitações para o Amazon S3. Essa função é necessária para todos os clusters.	`AmazonElasticMapReduceforEC2Role` . Para obter mais informações, consulte Função de serviço para EC2 instâncias de cluster (perfil de EC2 instância).

Função	Perfil padrão	Descrição	Política gerenciada padrão
Função de serviço para escalabilidade automática na Amazon EMR (função Auto Scaling)	`EMR_AutoScaling_DefaultRole`	Permite ações adicionais para ambientes de escalabilidade dinâmica. Exigido somente para clusters que usam escalabilidade automática na AmazonEMR. Para obter mais informações, consulte Usar o ajuste de escala automático com uma política personalizada para grupos de instâncias.	`AmazonElasticMapReduceforAutoScalingRole` . Para obter mais informações, consulte Função de serviço para escalabilidade automática na Amazon EMR (função Auto Scaling).

Função	Perfil padrão	Descrição	Política gerenciada padrão
Função de serviço para EMR notebooks	`EMR_Notebooks_DefaultRole`	Fornece as permissões que um EMR notebook precisa para acessar outros AWS recursos e realizar ações. Exigido somente se os EMR Notebooks forem usados.	`AmazonElasticMapReduceEditorsRole` . Para obter mais informações, consulte Função de serviço para EMR notebooks. `S3FullAccessPolicy` também é anexado por padrão. Veja a seguir o conteúdo da política. `{ "Version": "2012-10-17", "Statement": [ { "Effect": "Allow", "Action": "s3:*", "Resource": "*" } ] }`

Função	Perfil padrão	Descrição	Política gerenciada padrão
Função vinculada ao serviço	`AWSServiceRoleForEMRCleanup`	A Amazon cria EMR automaticamente uma função vinculada ao serviço. Se o serviço da Amazon EMR tiver perdido a capacidade de limpar EC2 os recursos da Amazon, a Amazon EMR poderá usar essa função para limpar. Se um cluster usar instâncias spot, a política de permissões anexada ao Função de serviço para a Amazon EMR (EMRfunção) deverá permitir a criação de uma função vinculada ao serviço. Para obter mais informações, consulte Usando funções vinculadas a serviços para a Amazon EMR.	`AmazonEMRCleanupPolicy`

Tópicos

- IAMfunções de serviço usadas pela Amazon EMR
- Personalize IAM funções
- Configurar IAM funções para EMRFS solicitações para o Amazon S3
- Use políticas baseadas em recursos para que a Amazon EMR acesse o AWS Glue Data Catalog

- Use IAM funções com aplicativos que chamam AWS serviços diretamente
- Permitir que usuários e grupos criem e modifiquem perfis

IAMfunções de serviço usadas pela Amazon EMR

A Amazon EMR usa funções de IAM serviço para realizar ações em seu nome ao provisionar recursos de cluster, executar aplicativos, escalar recursos dinamicamente e criar e executar notebooks. EMR A Amazon EMR usa as seguintes funções ao interagir com outros AWS serviços. Cada função tem uma função exclusiva na AmazonEMR. Os tópicos desta seção descrevem o papel da função e fornecem as funções padrão e a política de permissões de cada função.

Se você tiver um código de aplicativo em seu cluster que chame AWS serviços diretamente, talvez seja necessário usar o SDK para especificar funções. Para obter mais informações, consulte Use IAM funções com aplicativos que chamam AWS serviços diretamente.

Tópicos

- Função de serviço para a Amazon EMR (EMRfunção)
- Função de serviço para EC2 instâncias de cluster (perfil de EC2 instância)
- Função de serviço para escalabilidade automática na Amazon EMR (função Auto Scaling)
- Função de serviço para EMR notebooks
- Usando funções vinculadas a serviços para a Amazon EMR

Função de serviço para a Amazon EMR (EMRfunção)

A EMR função da Amazon define as ações permitidas para a Amazon EMR quando ela provisiona recursos e executa tarefas de nível de serviço que não são executadas no contexto de uma EC2 instância da Amazon em execução em um cluster. Por exemplo, a função de serviço é usada para provisionar EC2 instâncias quando um cluster é iniciado.

- O nome de perfil padrão é `EMR_DefaultRole_V2`.
- A política EMR gerenciada padrão com escopo da Amazon anexada `EMR_DefaultRole_V2` é`AmazonEMRServicePolicy_v2`. Essa política v2 substitui a política gerenciada padrão defasada, `AmazonElasticMapReduceRole`.

`AmazonEMRServicePolicy_v2`depende do acesso limitado aos recursos que a Amazon EMR provisiona ou usa. Ao usar essa política, é necessário passar a etiqueta de usuário `for-use-with-`

`amazon-emr-managed-policies = true` ao provisionar o cluster. A Amazon EMR propagará automaticamente essas tags. Além disso, talvez seja necessário adicionar manualmente uma tag de usuário a tipos específicos de recursos, como grupos EC2 de segurança que não foram criados pela AmazonEMR. Consulte Etiquetar recursos para usar políticas gerenciadas.

> **Important**
>
> A Amazon EMR usa essa função EMR de serviço da Amazon e a `AWSServiceRoleForEMRCleanup` função para limpar recursos de cluster em sua conta que você não usa mais, como EC2 instâncias da Amazon. Você deve incluir ações nas políticas de perfil para excluir ou encerrar os recursos. Caso contrário, a Amazon não EMR poderá realizar essas ações de limpeza e você poderá incorrer em custos com recursos não utilizados que permanecem no cluster.

O exemplo a seguir mostra o conteúdo de uma política `AmazonEMRServicePolicy_v2` atual. Você também pode ver o conteúdo atual da política `AmazonEMRServicePolicy_v2`gerenciada no IAM console.

```
{
 "Version": "2012-10-17",
 "Statement": [
  {
   "Sid": "CreateInTaggedNetwork",
   "Effect": "Allow",
   "Action": [
    "ec2:CreateNetworkInterface",
    "ec2:RunInstances",
    "ec2:CreateFleet",
    "ec2:CreateLaunchTemplate",
    "ec2:CreateLaunchTemplateVersion"
   ],
   "Resource": [
    "arn:aws:ec2:*:*:subnet/*",
    "arn:aws:ec2:*:*:security-group/*"
   ],
   "Condition": {
    "StringEquals": {
     "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
    }
   }
```

```
  },
  {
   "Sid": "CreateWithEMRTaggedLaunchTemplate",
   "Effect": "Allow",
   "Action": [
    "ec2:CreateFleet",
    "ec2:RunInstances",
    "ec2:CreateLaunchTemplateVersion"
   ],
   "Resource": "arn:aws:ec2:*:*:launch-template/*",
   "Condition": {
    "StringEquals": {
     "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
    }
   }
  },
  {
   "Sid": "CreateEMRTaggedLaunchTemplate",
   "Effect": "Allow",
   "Action": "ec2:CreateLaunchTemplate",
   "Resource": "arn:aws:ec2:*:*:launch-template/*",
   "Condition": {
    "StringEquals": {
     "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
    }
   }
  },
  {
   "Sid": "CreateEMRTaggedInstancesAndVolumes",
   "Effect": "Allow",
   "Action": [
    "ec2:RunInstances",
    "ec2:CreateFleet"
   ],
   "Resource": [
    "arn:aws:ec2:*:*:instance/*",
    "arn:aws:ec2:*:*:volume/*"
   ],
   "Condition": {
    "StringEquals": {
     "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
    }
   }
  },
```

```
{
 "Sid": "ResourcesToLaunchEC2",
 "Effect": "Allow",
 "Action": [
  "ec2:RunInstances",
  "ec2:CreateFleet",
  "ec2:CreateLaunchTemplate",
  "ec2:CreateLaunchTemplateVersion"
 ],
 "Resource": [
  "arn:aws:ec2:*:*:network-interface/*",
  "arn:aws:ec2:*::image/ami-*",
  "arn:aws:ec2:*:*:key-pair/*",
  "arn:aws:ec2:*:*:capacity-reservation/*",
  "arn:aws:ec2:*:*:placement-group/pg-*",
  "arn:aws:ec2:*:*:fleet/*",
  "arn:aws:ec2:*:*:dedicated-host/*",
  "arn:aws:resource-groups:*:*:group/*"
 ]
},
{
 "Sid": "ManageEMRTaggedResources",
 "Effect": "Allow",
 "Action": [
  "ec2:CreateLaunchTemplateVersion",
  "ec2:DeleteLaunchTemplate",
  "ec2:DeleteNetworkInterface",
  "ec2:ModifyInstanceAttribute",
  "ec2:TerminateInstances"
 ],
 "Resource": "*",
 "Condition": {
  "StringEquals": {
   "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
  }
 }
},
{
 "Sid": "ManageTagsOnEMRTaggedResources",
 "Effect": "Allow",
 "Action": [
  "ec2:CreateTags",
  "ec2:DeleteTags"
 ],
```

```
    "Resource": [
     "arn:aws:ec2:*:*:instance/*",
     "arn:aws:ec2:*:*:volume/*",
     "arn:aws:ec2:*:*:network-interface/*",
     "arn:aws:ec2:*:*:launch-template/*"
    ],
    "Condition": {
     "StringEquals": {
      "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
     }
    }
   },
   {
    "Sid": "CreateNetworkInterfaceNeededForPrivateSubnet",
    "Effect": "Allow",
    "Action": [
     "ec2:CreateNetworkInterface"
    ],
    "Resource": [
     "arn:aws:ec2:*:*:network-interface/*"
    ],
    "Condition": {
     "StringEquals": {
      "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
     }
    }
   },
   {
    "Sid": "TagOnCreateTaggedEMRResources",
    "Effect": "Allow",
    "Action": [
     "ec2:CreateTags"
    ],
    "Resource": [
     "arn:aws:ec2:*:*:network-interface/*",
     "arn:aws:ec2:*:*:instance/*",
     "arn:aws:ec2:*:*:volume/*",
     "arn:aws:ec2:*:*:launch-template/*"
    ],
    "Condition": {
     "StringEquals": {
      "ec2:CreateAction": [
       "RunInstances",
       "CreateFleet",
```

```
      "CreateLaunchTemplate",
      "CreateNetworkInterface"
     ]
    }
   }
  },
  {
   "Sid": "TagPlacementGroups",
   "Effect": "Allow",
   "Action": [
    "ec2:CreateTags",
    "ec2:DeleteTags"
   ],
   "Resource": [
    "arn:aws:ec2:*:*:placement-group/pg-*"
   ]
  },
  {
   "Sid": "ListActionsForEC2Resources",
   "Effect": "Allow",
   "Action": [
    "ec2:DescribeAccountAttributes",
    "ec2:DescribeCapacityReservations",
    "ec2:DescribeDhcpOptions",
    "ec2:DescribeImages",
    "ec2:DescribeInstances",
    "ec2:DescribeInstanceTypeOfferings",
    "ec2:DescribeLaunchTemplates",
    "ec2:DescribeNetworkAcls",
    "ec2:DescribeNetworkInterfaces",
    "ec2:DescribePlacementGroups",
    "ec2:DescribeRouteTables",
    "ec2:DescribeSecurityGroups",
    "ec2:DescribeSubnets",
    "ec2:DescribeVolumes",
    "ec2:DescribeVolumeStatus",
    "ec2:DescribeVpcAttribute",
    "ec2:DescribeVpcEndpoints",
    "ec2:DescribeVpcs"
   ],
   "Resource": "*"
  },
  {
   "Sid": "CreateDefaultSecurityGroupWithEMRTags",
```

```
    "Effect": "Allow",
    "Action": [
     "ec2:CreateSecurityGroup"
    ],
    "Resource": [
     "arn:aws:ec2:*:*:security-group/*"
    ],
    "Condition": {
     "StringEquals": {
      "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
     }
    }
   },
   {
    "Sid": "CreateDefaultSecurityGroupInVPCWithEMRTags",
    "Effect": "Allow",
    "Action": [
     "ec2:CreateSecurityGroup"
    ],
    "Resource": [
     "arn:aws:ec2:*:*:vpc/*"
    ],
    "Condition": {
     "StringEquals": {
      "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
     }
    }
   },
   {
    "Sid": "TagOnCreateDefaultSecurityGroupWithEMRTags",
    "Effect": "Allow",
    "Action": [
     "ec2:CreateTags"
    ],
    "Resource": "arn:aws:ec2:*:*:security-group/*",
    "Condition": {
     "StringEquals": {
      "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true",
      "ec2:CreateAction": "CreateSecurityGroup"
     }
    }
   },
   {
    "Sid": "ManageSecurityGroups",
```

```
    "Effect": "Allow",
    "Action": [
     "ec2:AuthorizeSecurityGroupEgress",
     "ec2:AuthorizeSecurityGroupIngress",
     "ec2:RevokeSecurityGroupEgress",
     "ec2:RevokeSecurityGroupIngress"
    ],
    "Resource": "*",
    "Condition": {
     "StringEquals": {
      "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
     }
    }
   },
   {
    "Sid": "CreateEMRPlacementGroups",
    "Effect": "Allow",
    "Action": [
     "ec2:CreatePlacementGroup"
    ],
    "Resource": "arn:aws:ec2:*:*:placement-group/pg-*"
   },
   {
    "Sid": "DeletePlacementGroups",
    "Effect": "Allow",
    "Action": [
     "ec2:DeletePlacementGroup"
    ],
    "Resource": "*"
   },
   {
    "Sid": "AutoScaling",
    "Effect": "Allow",
    "Action": [
     "application-autoscaling:DeleteScalingPolicy",
     "application-autoscaling:DeregisterScalableTarget",
     "application-autoscaling:DescribeScalableTargets",
     "application-autoscaling:DescribeScalingPolicies",
     "application-autoscaling:PutScalingPolicy",
     "application-autoscaling:RegisterScalableTarget"
    ],
    "Resource": "*"
   },
   {
```

```
    "Sid": "ResourceGroupsForCapacityReservations",
    "Effect": "Allow",
    "Action": [
     "resource-groups:ListGroupResources"
    ],
    "Resource": "*"
   },
   {
    "Sid": "AutoScalingCloudWatch",
    "Effect": "Allow",
    "Action": [
     "cloudwatch:PutMetricAlarm",
     "cloudwatch:DeleteAlarms",
     "cloudwatch:DescribeAlarms"
    ],
    "Resource": "arn:aws:cloudwatch:*:*:alarm:*_EMR_Auto_Scaling"
   },
   {
    "Sid": "PassRoleForAutoScaling",
    "Effect": "Allow",
    "Action": "iam:PassRole",
    "Resource": "arn:aws:iam::*:role/EMR_AutoScaling_DefaultRole",
    "Condition": {
     "StringLike": {
      "iam:PassedToService": "application-autoscaling.amazonaws.com*"
     }
    }
   },
   {
    "Sid": "PassRoleForEC2",
    "Effect": "Allow",
    "Action": "iam:PassRole",
    "Resource": "arn:aws:iam::*:role/EMR_EC2_DefaultRole",
    "Condition": {
     "StringLike": {
      "iam:PassedToService": "ec2.amazonaws.com*"
     }
    }
   }
  ]
}
```

Seu perfil de serviço deve usar a seguinte política de confiança.

> **Important**
>
> A política de confiança a seguir inclui as chaves de condição `aws:SourceAccount` globais `aws:SourceArn` e as chaves de condição, que limitam as permissões que você concede EMR à Amazon para recursos específicos em sua conta. O uso delas pode proteger você contra o problema de "confused deputy".

```
{
    "Version": "2008-10-17",
    "Statement": [
        {
            "Sid": "",
            "Effect": "Allow",
            "Principal": {
                "Service": "elasticmapreduce.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEquals": {
                    "aws:SourceAccount": "<account-id>"
                },
                "ArnLike": {
                    "aws:SourceArn": "arn:aws:elasticmapreduce:<region>:<account-id>:*"
                }
            }
        }
    ]
}
```

Função de serviço para EC2 instâncias de cluster (perfil de EC2 instância)

A função de serviço para EC2 instâncias de cluster (também chamada de perfil de EC2 instância para a AmazonEMR) é um tipo especial de função de serviço que é atribuída a cada EC2 instância em um EMR cluster da Amazon quando a instância é iniciada. Os processos de aplicação que são executados no ecossistema do Hadoop assumem esse perfil para que as permissões interajam com outros serviços da AWS .

Para obter mais informações sobre funções de serviço para EC2 instâncias, consulte Como usar uma IAM função para conceder permissões a aplicativos executados em EC2 instâncias da Amazon no Guia IAM do usuário.

> **Important**
>
> A função de serviço padrão para EC2 instâncias de cluster e sua política gerenciada AWS padrão associada `AmazonElasticMapReduceforEC2Role` estão em vias de descontinuação, sem nenhuma política AWS gerenciada substituta fornecida. Será necessário criar e especificar um perfil de instância para substituir o perfil e a política padrão defasados.

Perfil padrão e política gerenciada

- O nome de perfil padrão é `EMR_EC2_DefaultRole`.
- A política gerenciada `EMR_EC2_DefaultRole` padrão, `AmazonElasticMapReduceforEC2Role`, está chegando ao fim do suporte. Em vez de usar uma política gerenciada padrão para o perfil da EC2 instância, aplique políticas baseadas em recursos aos buckets do S3 e outros recursos que a Amazon EMR precisa, ou use sua própria política gerenciada pelo cliente com uma IAM função como perfil de instância. Para obter mais informações, consulte Criação de uma função de serviço para EC2 instâncias de cluster com permissões de privilégio mínimo.

Veja a seguir o conteúdo da versão 3 de `AmazonElasticMapReduceforEC2Role`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Resource": "*",
            "Action": [
                "cloudwatch:*",
                "dynamodb:*",
                "ec2:Describe*",
                "elasticmapreduce:Describe*",
                "elasticmapreduce:ListBootstrapActions",
                "elasticmapreduce:ListClusters",
                "elasticmapreduce:ListInstanceGroups",
                "elasticmapreduce:ListInstances",
                "elasticmapreduce:ListSteps",
                "kinesis:CreateStream",
```

```
                "kinesis:DeleteStream",
                "kinesis:DescribeStream",
                "kinesis:GetRecords",
                "kinesis:GetShardIterator",
                "kinesis:MergeShards",
                "kinesis:PutRecord",
                "kinesis:SplitShard",
                "rds:Describe*",
                "s3:*",
                "sdb:*",
                "sns:*",
                "sqs:*",
                "glue:CreateDatabase",
                "glue:UpdateDatabase",
                "glue:DeleteDatabase",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:CreateTable",
                "glue:UpdateTable",
                "glue:DeleteTable",
                "glue:GetTable",
                "glue:GetTables",
                "glue:GetTableVersions",
                "glue:CreatePartition",
                "glue:BatchCreatePartition",
                "glue:UpdatePartition",
                "glue:DeletePartition",
                "glue:BatchDeletePartition",
                "glue:GetPartition",
                "glue:GetPartitions",
                "glue:BatchGetPartition",
                "glue:CreateUserDefinedFunction",
                "glue:UpdateUserDefinedFunction",
                "glue:DeleteUserDefinedFunction",
                "glue:GetUserDefinedFunction",
                "glue:GetUserDefinedFunctions"
            ]
        }
    ]
}
```

Seu perfil de serviço deve usar a seguinte política de confiança.

```
{
    "Version": "2008-10-17",
    "Statement": [
        {
            "Sid": "",
            "Effect": "Allow",
            "Principal": {
                "Service": "ec2.amazonaws.com"
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

Criação de uma função de serviço para EC2 instâncias de cluster com permissões de privilégio mínimo

Como prática recomendada, é altamente recomendável que você crie uma função de serviço para EC2 instâncias de cluster e uma política de permissões que tenha as permissões mínimas para outros AWS serviços exigidos pelo seu aplicativo.

A política gerenciada padrão, `AmazonElasticMapReduceforEC2Role`, fornece permissões que facilitam a execução de um cluster inicial. No entanto, `AmazonElasticMapReduceforEC2Role` está em vias de descontinuação e a Amazon não EMR fornecerá uma política padrão AWS gerenciada substituta para a função obsoleta. Para iniciar um cluster inicial, é necessário fornecer uma política gerenciada pelo cliente baseada em recursos ou baseada em ID.

As declarações de política a seguir fornecem exemplos das permissões necessárias para diferentes recursos da AmazonEMR. Recomendamos que você use essas permissões para criar uma política de permissões que restrinja o acesso somente a esses recursos e aos recursos que o cluster exige. Todos os exemplos de declarações de política usam o *`us-west-2`* Região e o ID fictício AWS da conta *`123456789012`*. Substitua-os conforme apropriado para seu cluster.

Para obter mais informações sobre como criar e especificar funções personalizadas, consulte Personalize IAM funções.

> **Note**
>
> Se você criar um EMR papel personalizado paraEC2, siga o fluxo de trabalho básico, que cria automaticamente um perfil de instância com o mesmo nome. A Amazon EC2 permite que você crie perfis e funções de instância com nomes diferentes, mas a Amazon EMR não

oferece suporte a essa configuração e isso resulta em um erro de “perfil de instância inválido” quando você cria o cluster.

Lendo e gravando dados no Amazon S3 usando EMRFS

Quando um aplicativo executado em um EMR cluster da Amazon faz referência a dados usando o `s3://mydata` formato, a Amazon EMR usa o perfil da EC2 instância para fazer a solicitação. Os clusters normalmente leem e gravam dados no Amazon S3 dessa forma, e a Amazon EMR usa as permissões associadas à função de serviço para EC2 instâncias de cluster por padrão. Para obter mais informações, consulte Configurar IAM funções para EMRFS solicitações para o Amazon S3.

Como as IAM funções de EMRFS retornarão às permissões associadas à função de serviço para EC2 instâncias de cluster, como melhor prática, recomendamos que você use IAM funções e limite as permissões do EMRFS Amazon S3 associadas à função de serviço para instâncias de clusterEC2. EMRFS

O exemplo de declaração abaixo demonstra as permissões EMRFS necessárias para fazer solicitações ao Amazon S3.

- *my-data-bucket-in-s3- -escreve for-emrfs-reads-and* especifica o bucket no Amazon S3 em que o cluster lê e grava dados e todas as subpastas usando */**. Adicione somente os buckets e pastas que seu aplicativo exige.
- A declaração de política que permite dynamodb ações é necessária somente se a visualização EMRFS consistente estiver ativada. *E mrFSMetadata* especifica a pasta padrão para uma visualização EMRFS consistente.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:AbortMultipartUpload",
                "s3:CreateBucket",
                "s3:DeleteObject",
                "s3:GetBucketVersioning",
                "s3:GetObject",
                "s3:GetObjectTagging",
```

```
                "s3:GetObjectVersion",
                "s3:ListBucket",
                "s3:ListBucketMultipartUploads",
                "s3:ListBucketVersions",
                "s3:ListMultipartUploadParts",
                "s3:PutBucketVersioning",
                "s3:PutObject",
                "s3:PutObjectTagging"
            ],
            "Resource": [
                "arn:aws:s3:::my-data-bucket-in-s3-for-emrfs-reads-and-writes",
                "arn:aws:s3:::my-data-bucket-in-s3-for-emrfs-reads-and-writes/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "dynamodb:CreateTable",
                "dynamodb:BatchGetItem",
                "dynamodb:BatchWriteItem",
                "dynamodb:PutItem",
                "dynamodb:DescribeTable",
                "dynamodb:DeleteItem",
                "dynamodb:GetItem",
                "dynamodb:Scan",
                "dynamodb:Query",
                "dynamodb:UpdateItem",
                "dynamodb:DeleteTable",
                "dynamodb:UpdateTable"
            ],
            "Resource": "arn:aws:dynamodb:us-west-2:123456789012:table/EmrFSMetadata"
        },
        {
            "Effect": "Allow",
            "Action": [
                "cloudwatch:PutMetricData",
                "dynamodb:ListTables",
                "s3:ListBucket"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
```

```
                "sqs:GetQueueUrl",
                "sqs:ReceiveMessage",
                "sqs:DeleteQueue",
                "sqs:SendMessage",
                "sqs:CreateQueue"
            ],
            "Resource": "arn:aws:sqs:us-west-2:123456789012:EMRFS-Inconsistency-*"
        }
    ]
}
```

Arquivar arquivos de log no Amazon S3

A declaração de política a seguir permite que o EMR cluster da Amazon arquive arquivos de log no local especificado do Amazon S3. No exemplo abaixo, quando o cluster foi criado, *s3://MyLoggingBucketyEMRCluster/M Logs* foi especificado usando a localização da pasta Log S3 no console, usando a `--log-uri` opção do AWS CLI ou usando o `LogUri` parâmetro no `RunJobFlow` comando. Para obter mais informações, consulte Arquivamento dos arquivos de log no Amazon S3.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "s3:PutObject",
            "Resource": "arn:aws:s3:::MyLoggingBucket/MyEMRClusterLogs/*"
        }
    ]
}
```

Usando o AWS Glue Data Catalog

A declaração de política a seguir permite ações que são necessárias se você usar o AWS Glue Data Catalog como metastore para aplicativos. Para obter mais informações, consulte Usando o AWS Glue Data Catalog como metastore para o Spark SQL, Usando o AWS Glue Data Catalog como metastore para o Hive e Usando o Presto com o Glue AWS Data Catalog no Amazon Release Guide. EMR

```
{
```

```
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:CreateDatabase",
                "glue:UpdateDatabase",
                "glue:DeleteDatabase",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:CreateTable",
                "glue:UpdateTable",
                "glue:DeleteTable",
                "glue:GetTable",
                "glue:GetTables",
                "glue:GetTableVersions",
                "glue:CreatePartition",
                "glue:BatchCreatePartition",
                "glue:UpdatePartition",
                "glue:DeletePartition",
                "glue:BatchDeletePartition",
                "glue:GetPartition",
                "glue:GetPartitions",
                "glue:BatchGetPartition",
                "glue:CreateUserDefinedFunction",
                "glue:UpdateUserDefinedFunction",
                "glue:DeleteUserDefinedFunction",
                "glue:GetUserDefinedFunction",
                "glue:GetUserDefinedFunctions"
            ],
            "Resource": "*",
        }
    ]
}
```

Função de serviço para escalabilidade automática na Amazon EMR (função Auto Scaling)

A função Auto Scaling da Amazon EMR executa uma função semelhante à função de serviço, mas permite ações adicionais para ambientes de escalabilidade dinâmica.

- O nome de perfil padrão é `EMR_AutoScaling_DefaultRole`.
- A política gerenciada padrão anexada a `EMR_AutoScaling_DefaultRole` é `AmazonElasticMapReduceforAutoScalingRole`.

O conteúdo da versão 1 da `AmazonElasticMapReduceforAutoScalingRole` é mostrado a seguir.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "cloudwatch:DescribeAlarms",
                "elasticmapreduce:ListInstanceGroups",
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Allow",
            "Resource": "*"
        }
    ]
}
```

Seu perfil de serviço deve usar a seguinte política de confiança.

> **Important**
>
> A política de confiança a seguir inclui as chaves de condição `aws:SourceAccount` globais `aws:SourceArn` e as chaves de condição, que limitam as permissões que você concede EMR à Amazon para recursos específicos em sua conta. O uso delas pode proteger você contra o problema de "confused deputy".

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "application-autoscaling.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEquals": {
                    "aws:SourceAccount": "<account-id>"
                },
```

```
                "ArnLike": {
                    "aws:SourceArn": "arn:aws:application-
autoscaling:<region>:<account-id>:scalable-target/*"
                }
            }
        }
    ]
}
```

Função de serviço para EMR notebooks

Cada EMR notebook precisa de permissões para acessar outros AWS recursos e realizar ações. As IAM políticas anexadas a essa função de serviço fornecem permissões para que o notebook interopere com outros AWS serviços. Ao criar um notebook usando o AWS Management Console, você especifica uma função AWS de serviço. É possível usar a função padrão, `EMR_Notebooks_DefaultRole`, ou especificar uma função criada por você. Se um bloco de anotações não foi criado anteriormente, é possível optar por criar a função padrão.

- O nome de perfil padrão é `EMR_Notebooks_DefaultRole`.
- As políticas gerenciadas listadas anexadas a `EMR_Notebooks_DefaultRole` são `AmazonElasticMapReduceEditorsRole` e `S3FullAccessPolicy`.

Seu perfil de serviço deve usar a seguinte política de confiança.

> **Important**
>
> A política de confiança a seguir inclui as chaves de condição `aws:SourceAccount`globais `aws:SourceArn`e as chaves de condição, que limitam as permissões que você concede EMR à Amazon para recursos específicos em sua conta. O uso delas pode proteger você contra o problema de “confused deputy”.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "elasticmapreduce.amazonaws.com"
```

```
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEquals": {
                    "aws:SourceAccount": "<account-id>"
                },
                "ArnLike": {
                    "aws:SourceArn": "arn:aws:elasticmapreduce:<region>:<account-id>:*"
                }
            }
        }
    ]
}
```

O conteúdo da versão 1 de `AmazonElasticMapReduceEditorsRole` é o seguinte.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ec2:AuthorizeSecurityGroupEgress",
                "ec2:AuthorizeSecurityGroupIngress",
                "ec2:CreateSecurityGroup",
                "ec2:DescribeSecurityGroups",
                "ec2:RevokeSecurityGroupEgress",
                "ec2:CreateNetworkInterface",
                "ec2:CreateNetworkInterfacePermission",
                "ec2:DeleteNetworkInterface",
                "ec2:DeleteNetworkInterfacePermission",
                "ec2:DescribeNetworkInterfaces",
                "ec2:ModifyNetworkInterfaceAttribute",
                "ec2:DescribeTags",
                "ec2:DescribeInstances",
                "ec2:DescribeSubnets",
                "ec2:DescribeVpcs",
                "elasticmapreduce:ListInstances",
                "elasticmapreduce:DescribeCluster",
                "elasticmapreduce:ListSteps"
            ],
            "Resource": "*"
        },
```

```
        {
            "Effect": "Allow",
            "Action": "ec2:CreateTags",
            "Resource": "arn:aws:ec2:*:*:network-interface/*",
            "Condition": {
                "ForAllValues:StringEquals": {
                    "aws:TagKeys": [
                        "aws:elasticmapreduce:editor-id",
                        "aws:elasticmapreduce:job-flow-id"
                    ]
                }
            }
        }
    ]
}
```

Veja a seguir o conteúdo de `S3FullAccessPolicy`. `S3FullAccessPolicy`Isso permite que sua função de serviço para EMR notebooks execute todas as ações do Amazon S3 em objetos em seu. Conta da AWS Ao criar uma função de serviço personalizada para EMR Notebooks, você deve conceder permissões ao Amazon S3 para sua função de serviço.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "s3:*",
            "Resource": "*"
        }
    ]
}
```

Você pode restringir o acesso de leitura e gravação do perfil de serviço ao local do Amazon S3 em que deseja salvar os arquivos do caderno. Use o conjunto mínimo de permissões do Amazon S3 apresentado a seguir.

```
"s3:PutObject",
"s3:GetObject",
"s3:GetEncryptionConfiguration",
"s3:ListBucket",
"s3:DeleteObject"
```

Se o seu bucket do Amazon S3 estiver criptografado, você deverá incluir as seguintes permissões para o AWS Key Management Service.

```
"kms:Decrypt",
"kms:GenerateDataKey",
"kms:ReEncryptFrom",
"kms:ReEncryptTo",
"kms:DescribeKey"
```

Ao vincular repositórios Git ao seu notebook e precisar criar um segredo para o repositório, você deve adicionar a `secretsmanager:GetSecretValue` permissão na IAM política anexada à função de serviço dos notebooks da Amazon. EMR Um exemplo de política é demonstrado a seguir:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "VisualEditor0",
            "Effect": "Allow",
            "Action": "secretsmanager:GetSecretValue",
            "Resource": "*"
        }
    ]
}
```

EMRPermissões da função de serviço de notebooks

Esta tabela lista as ações que o EMR Notebooks executa usando a função de serviço, junto com as permissões necessárias para cada ação.

Ação	Permissões
Estabeleça um canal de rede seguro entre um notebook e um EMR cluster da Amazon e execute as ações de limpeza necessárias.	`"ec2:CreateNetworkInterface", "ec2:CreateNetworkInterfacePermission", "ec2:DeleteNetworkInterface", "ec2:DeleteNetworkInterfacePermission", "ec2:DescribeNetworkInterfaces", "ec2:ModifyNetworkInterfaceAttribute", "ec2:AuthorizeSecurityGroupEgress", "ec2:AuthorizeSecurityGroupIngress", "ec2:CreateSecurityGroup",`

Ação	Permissões
	"ec2:DescribeSecurityGroups", "ec2:RevokeSecurityGroupEgress", "ec2:DescribeTags", "ec2:DescribeInstances", "ec2:DescribeSubnets", "ec2:DescribeVpcs", "elasticmapreduce:ListInstances", "elasticmapreduce:DescribeCluster", "elasticmapreduce:ListSteps"
Uso das credenciais do Git armazenadas no AWS Secrets Manager para vincular repositórios Git a um caderno.	"secretsmanager:GetSecretValue"
Aplique AWS tags à interface de rede e aos grupos de segurança padrão que o EMR Notebooks cria ao configurar o canal de rede seguro. Para obter mais informaçõ es, consulte Etiquetar recursos da AWS.	"ec2:CreateTags"
Acesso ou upload de arquivos e metadados de cadernos para o Amazon S3.	"s3:PutObject", "s3:GetObject", "s3:GetEncryptionConfiguration", "s3:ListBucket", "s3:DeleteObject" As permissões a seguir serão necessárias somente se você usar um bucket criptografado do Amazon S3. "kms:Decrypt", "kms:GenerateDataKey", "kms:ReEncryptFrom", "kms:ReEncryptTo", "kms:DescribeKey"

EMRAtualizações de notebooks para políticas AWS gerenciadas

Veja detalhes sobre as atualizações das políticas AWS gerenciadas para EMR notebooks desde 1º de março de 2021.

Alteração	Descrição	Data
`AmazonElasticMapReduceEditorsRole - Added permissions`	EMRNotebooks adicionados `ec2:describeVPCs` e `elastmicmapreduce:ListSteps` permissões para`AmazonElasticMapReduceEditorsRole` .	8 de fevereiro de 2023
EMROs notebooks começaram a monitorar as alterações	EMROs notebooks começaram a monitorar as alterações em suas políticas AWS gerenciadas.	8 de fevereiro de 2023

Usando funções vinculadas a serviços para a Amazon EMR

A Amazon EMR usa AWS Identity and Access Management (IAM) funções vinculadas ao serviço. Uma função vinculada a serviços é um tipo exclusivo de IAM função vinculada diretamente à Amazon. EMR As funções vinculadas ao serviço são predefinidas pela Amazon EMR e incluem todas as permissões que o serviço exige para ligar para outros AWS serviços em seu nome.

Tópicos

- Usando funções vinculadas a serviços para limpeza
- Usando funções vinculadas ao serviço para registro antecipado

Para obter informações sobre outros serviços que oferecem suporte a funções vinculadas a serviços, consulte AWS serviços que funcionam com IAM e procure os serviços que têm Sim na coluna Funções vinculadas ao serviço. Escolha um Sim com um link para visualizar a documentação da função vinculada a esse serviço.

Usando funções vinculadas a serviços para limpeza

A Amazon EMR usa AWS Identity and Access Management (IAM) funções vinculadas ao serviço. Uma função vinculada a serviços é um tipo exclusivo de IAM função vinculada diretamente à Amazon. EMR As funções vinculadas ao serviço são predefinidas pela Amazon EMR e incluem todas as permissões que o serviço exige para ligar para outros AWS serviços em seu nome.

As funções vinculadas ao serviço funcionam em conjunto com a função de EMR serviço da Amazon e o perfil de EC2 instância da Amazon para a Amazon. EMR Para obter mais informações sobre a função de serviço e o perfil da instância, consulte Configure funções IAM de serviço para EMR permissões da Amazon para AWS serviços e recursos.

Uma função vinculada ao serviço facilita a configuração da Amazon EMR porque você não precisa adicionar manualmente as permissões necessárias. A Amazon EMR define as permissões de suas funções vinculadas ao serviço e, a menos que seja definido de outra forma, somente a Amazon EMR pode assumir suas funções. As permissões definidas incluem a política de confiança e a política de permissões, e essa política de permissões não pode ser anexada a nenhuma outra IAM entidade.

Você pode excluir essa função vinculada ao serviço da Amazon EMR somente depois de excluir todos os recursos relacionados e encerrar todos os EMR clusters na conta. Isso protege seus EMR recursos da Amazon para que você não possa remover inadvertidamente a permissão para acessar os recursos.

Usando funções vinculadas a serviços para limpeza

A Amazon EMR usa a AWSServiceRoleForEMRCleanupfunção baseada em serviços para EMR conceder permissão à Amazon para encerrar e excluir EC2 recursos da Amazon em seu nome se a função EMR vinculada ao serviço da Amazon perder essa capacidade. EMRA Amazon cria a função vinculada ao serviço automaticamente durante a criação do cluster, caso ela ainda não exista.

A função AWSServiceRoleForEMRCleanup vinculada ao serviço confia nos seguintes serviços para assumir a função:

- `elasticmapreduce.amazonaws.com`

A política de permissões AWSServiceRoleForEMRCleanup de funções vinculadas ao serviço permite EMR que a Amazon conclua as seguintes ações nos recursos especificados:

- Ação: `DescribeInstances` em `ec2`

- Ação: `DescribeSpotInstanceRequests` em `ec2`
- Ação: `ModifyInstanceAttribute` em `ec2`
- Ação: `TerminateInstances` em `ec2`
- Ação: `CancelSpotInstanceRequests` em `ec2`
- Ação: `DeleteNetworkInterface` em `ec2`
- Ação: `DescribeInstanceAttribute` em `ec2`
- Ação: `DescribeVolumeStatus` em `ec2`
- Ação: `DescribeVolumes` em `ec2`
- Ação: `DetachVolume` em `ec2`
- Ação: `DeleteVolume` em `ec2`

Você deve configurar permissões para permitir que uma IAM entidade (como um usuário, grupo ou função) crie, edite ou exclua uma função vinculada ao serviço.

Criação de uma função vinculada a serviços para a Amazon EMR

Você não precisa criar a AWSServiceRoleForEMRCleanup função manualmente. Quando você inicia um cluster, seja pela primeira vez ou quando a função AWSServiceRoleForEMRCleanup vinculada ao serviço não está presente, a Amazon EMR cria a função AWSServiceRoleForEMRCleanup vinculada ao serviço para você. Você deve ter permissões para criar uma função vinculada ao serviço. Para obter um exemplo de declaração que adiciona esse recurso à política de permissões de uma IAM entidade (como um usuário, grupo ou função), consulteUsando funções vinculadas a serviços para limpeza.

> **Important**
>
> Se você usou a Amazon EMR antes de 24 de outubro de 2017, quando as funções vinculadas ao serviço não eram suportadas, a Amazon EMR criou a função AWSServiceRoleForEMRCleanup vinculada ao serviço em sua conta. Para obter mais informações, consulte Uma nova função apareceu em minha IAM conta.

Editando uma função vinculada ao serviço para a Amazon EMR

A Amazon EMR não permite que você edite a função AWSServiceRoleForEMRCleanup vinculada ao serviço. Depois de criar uma função vinculada ao serviço, você não pode alterar o nome da função

vinculada ao serviço porque várias entidades podem fazer referência à função vinculada ao serviço. No entanto, você pode editar a descrição da função vinculada ao serviço usando. IAM

Editando uma descrição de função vinculada ao serviço (console) IAM

Você pode usar o IAM console para editar a descrição de uma função vinculada ao serviço.

Para editar a descrição de uma função vinculada ao serviço (console)

1. No painel de navegação do IAM console, escolha Funções.
2. Escolha o nome da função a ser modificada.
3. À direita de Descrição do perfil, escolha Editar.
4. Insira uma nova descrição na caixa e escolha Save changes (Salvar alterações).

Editando uma descrição de função vinculada ao serviço () IAM CLI

Você pode usar IAM os comandos do AWS Command Line Interface para editar a descrição de uma função vinculada ao serviço.

Para alterar a descrição de uma função vinculada ao serviço () CLI

1. (Opcional) Para visualizar a descrição atual de a uma função, use um dos comandos a seguir:

   ```
   $ aws iam get-role --role-name role-name
   ```

   Use o nome da função, não oARN, para se referir às funções com os CLI comandos. Por exemplo, se uma função tem o seguinteARN:`arn:aws:iam::123456789012:role/myrole`, você se refere à função como**`myrole`**.
2. Para atualizar a descrição de uma função vinculada a serviço, use um dos comandos a seguir:

   ```
   $ aws iam update-role-description --role-name role-name --description description
   ```

Editando uma descrição de função vinculada ao serviço () IAM API

Você pode usar o IAM API para editar a descrição de uma função vinculada ao serviço.

Para alterar a descrição de uma função vinculada ao serviço () API

1. (Opcional) Para visualizar a descrição atual de uma função, use o comando a seguir:

IAM API: GetRole

2. Para atualizar a descrição de uma função, use o comando a seguir:

   IAM API: UpdateRoleDescription

Excluindo uma função vinculada ao serviço para a Amazon EMR

Se você não precisar mais usar um recurso ou serviço que exija uma função vinculada ao serviço, recomendamos que você exclua essa função vinculada ao serviço. Dessa forma, você não terá uma entidade não utilizada que não seja monitorada ativamente ou mantida. No entanto, você deve limpar sua função vinculada ao serviço antes de excluí-la.

Limpar um perfil vinculado ao serviço

Antes de poder usar IAM para excluir uma função vinculada ao serviço, você deve primeiro confirmar se a função vinculada ao serviço não tem sessões ativas e remover todos os recursos usados pela função vinculada ao serviço.

Para verificar se a função vinculada ao serviço tem uma sessão ativa no console IAM

1. Abra o IAM console em https://console.aws.amazon.com/iam/.
2. No painel de navegação, escolha Perfis. Selecione o nome (não a caixa de seleção) da função AWSServiceRoleForEMRCleanup vinculada ao serviço.
3. Na página Resumo da função vinculada ao serviço selecionada, escolha Access Advisor.
4. Na guia Consultor de acesso, revise a atividade recente para a função vinculada ao serviço.

> **Note**
>
> Se você não tiver certeza se a Amazon EMR está usando a função AWSServiceRoleForEMRCleanup vinculada ao serviço, você pode tentar excluir a função vinculada ao serviço. Se o serviço estiver usando a função vinculada ao serviço, a exclusão falhará e você poderá visualizar as regiões em que a função vinculada ao serviço está sendo usada. Se a função vinculada ao serviço estiver sendo usada, você deverá aguardar o término da sessão antes de excluir a função vinculada ao serviço. Não é possível revogar a sessão de uma função vinculada a um serviço.

Para remover EMR os recursos da Amazon usados pelo AWSServiceRoleForEMRCleanup

- Encerre todos os clusters em sua conta. Para obter mais informações, consulte Terminar um cluster.

Excluindo uma função vinculada ao serviço (console) IAM

Você pode usar o IAM console para excluir uma função vinculada ao serviço.

Para excluir uma função vinculada ao serviço (console)

1. Abra o IAM console em https://console.aws.amazon.com/iam/.
2. No painel de navegação, escolha Perfis. Marque a caixa de seleção ao lado AWSServiceRoleForEMRCleanup, não o nome ou a linha em si.
3. Em ações de Função na parte superior da página, escolha a função Excluir.
4. Na caixa de diálogo de confirmação, revise os últimos dados acessados do serviço, que mostram quando cada uma das funções selecionadas acessou um AWS serviço pela última vez. Isso ajuda a confirmar se a função está ativa no momento. Para prosseguir, selecione Yes, Delete.
5. Assista às notificações do IAM console para monitorar o progresso da exclusão da função vinculada ao serviço. Como a exclusão da função IAM vinculada ao serviço é assíncrona, depois de enviar a função vinculada ao serviço para exclusão, a tarefa de exclusão pode ser bem-sucedida ou falhar. Se a tarefa obtiver êxito, você poderá escolher Visualizar Detalhes ou Visualizar Recursos a partir das notificações para saber por que a exclusão falhou. Se houve falha na exclusão porque há recursos no serviço que estão sendo usados pela função, o motivo da falha incluirá uma lista de recursos.

Excluindo uma função vinculada ao serviço () IAM CLI

Você pode usar IAM os comandos do AWS Command Line Interface para excluir uma função vinculada ao serviço. Como uma função vinculada ao serviço não podem ser excluída se estiver sendo usada ou tiver recursos associados, você deverá enviar uma solicitação de exclusão. Se essas condições não forem atendidas, essa solicitação poderá ser negada.

Para excluir uma função vinculada ao serviço () CLI

1. Para verificar o status da tarefa de exclusão, você deve capturar o `deletion-task-id` da resposta. Digite o seguinte comando para enviar uma solicitação de exclusão de função vinculada ao serviço:

   ```
   $ aws iam delete-service-linked-role --role-name AWSServiceRoleForEMRCleanup
   ```

2. Digite o seguinte comando para verificar o estado da tarefa de exclusão:

   ```
   $ aws iam get-service-linked-role-deletion-status --deletion-task-id deletion-task-id
   ```

   O status da tarefa de exclusão pode ser `NOT_STARTED`, `IN_PROGRESS`, `SUCCEEDED`, ou `FAILED`. Se a exclusão falhar, a chamada informará o motivo de falha para que você possa solucionar o problema.

Excluindo uma função vinculada ao serviço () IAM API

Você pode usar o IAM API para excluir uma função vinculada ao serviço. Como uma função vinculada ao serviço não podem ser excluída se estiver sendo usada ou tiver recursos associados, você deverá enviar uma solicitação de exclusão. Se essas condições não forem atendidas, essa solicitação poderá ser negada.

Para excluir uma função vinculada ao serviço () API

1. Para enviar uma solicitação de exclusão para uma função vinculada ao serviço, ligue. DeleteServiceLinkedRole Na solicitação, especifique o nome da AWSServiceRoleForEMRCleanup função.

   Para verificar o status da tarefa de exclusão, você deve capturar o `DeletionTaskId` da resposta.

2. Para verificar o status da exclusão, ligue GetServiceLinkedRoleDeletionStatus. Na solicitação, especifique o `DeletionTaskId`.

   O status da tarefa de exclusão pode ser `NOT_STARTED`, `IN_PROGRESS`, `SUCCEEDED`, ou `FAILED`. Se a exclusão falhar, a chamada informará o motivo de falha para que você possa solucionar o problema.

Regiões suportadas para AWSServiceRoleForEMRCleanup

A Amazon EMR oferece suporte ao uso da função AWSServiceRoleForEMRCleanup vinculada ao serviço nas seguintes regiões.

Nome da região	Identidade da região	Support na Amazon EMR
Leste dos EUA (Norte da Virgínia)	us-east-1	Sim
Leste dos EUA (Ohio)	us-east-2	Sim
Oeste dos EUA (N. da Califórnia)	us-west-1	Sim
Oeste dos EUA (Oregon)	us-west-2	Sim
Ásia-Pacífico (Mumbai)	ap-south-1	Sim
Ásia-Pacífico (Osaka)	ap-northeast-3	Sim
Ásia-Pacífico (Seul)	ap-northeast-2	Sim
Ásia-Pacífico (Singapura)	ap-southeast-1	Sim
Ásia-Pacífico (Sydney)	ap-southeast-2	Sim
Ásia-Pacífico (Tóquio)	ap-northeast-1	Sim
Canadá (Central)	ca-central-1	Sim
Europa (Frankfurt)	eu-central-1	Sim
Europa (Irlanda)	eu-west-1	Sim
Europa (Londres)	eu-west-2	Sim
Europa (Paris)	eu-west-3	Sim
América do Sul (São Paulo)	sa-east-1	Sim

Usando funções vinculadas ao serviço para registro antecipado

A Amazon EMR usa AWS Identity and Access Management (IAM) funções vinculadas ao serviço. Uma função vinculada a serviços é um tipo exclusivo de IAM função vinculada diretamente à Amazon. EMR As funções vinculadas ao serviço são predefinidas pela Amazon EMR e incluem todas as permissões que o serviço exige para ligar para outros AWS serviços em seu nome.

As funções vinculadas ao serviço funcionam em conjunto com a função de EMR serviço da Amazon e o perfil de EC2 instância da Amazon para a Amazon. EMR Para obter mais informações sobre a função de serviço e o perfil da instância, consulte Configure funções IAM de serviço para EMR permissões da Amazon para AWS serviços e recursos.

Uma função vinculada ao serviço facilita a configuração da Amazon EMR porque você não precisa adicionar manualmente as permissões necessárias. A Amazon EMR define as permissões de suas funções vinculadas ao serviço e, a menos que seja definido de outra forma, somente a Amazon EMR pode assumir suas funções. As permissões definidas incluem a política de confiança e a política de permissões, e essa política de permissões não pode ser anexada a nenhuma outra IAM entidade.

Você pode excluir essa função vinculada ao serviço da Amazon EMR somente depois de excluir os recursos relacionados e encerrar todos os EMR clusters na conta. Isso protege seus EMR recursos da Amazon para que você não possa remover inadvertidamente a permissão para acessar os recursos.

Permissões de função vinculadas ao serviço para registro antecipado () WAL

A Amazon EMR usa a função vinculada ao serviço AWSServiceRoleForEMRWALpara recuperar o status de um cluster.

A função AWSServiceRoleForEMRWAL vinculada ao serviço confia nos seguintes serviços para assumir a função:

- `emrwal.amazonaws.com`

A política de `EMRDescribeClusterPolicyForEMRWAL`permissões para a função vinculada ao serviço permite que EMR a Amazon conclua as seguintes ações nos recursos especificados:

- Ação: `DescribeCluster` em `*`

Você deve configurar permissões para permitir que uma IAM entidade (neste caso, a Amazon EMRWAL) crie, edite ou exclua uma função vinculada ao serviço. Adicione as seguintes declarações conforme necessário à política de permissões do seu perfil de instância:

CreateServiceLinkedRole

Para permitir que uma IAM entidade crie a função vinculada ao AWSServiceRoleForEMRWAL serviço

Adicione a seguinte declaração à política de permissões da IAM entidade que precisa criar a função vinculada ao serviço:

```
{
    "Effect": "Allow",
    "Action": [
        "iam:CreateServiceLinkedRole",
        "iam:PutRolePolicy"
    ],
    "Resource": "arn:aws:iam::*:role/aws-service-role/emrwal.amazonaws.com*/
AWSServiceRoleForEMRWAL*",
    "Condition": {
        "StringLike": {
            "iam:AWSServiceName": [
                "emrwal.amazonaws.com",
                "elasticmapreduce.amazonaws.com.cn"
            ]
        }
    }
}
```

UpdateRoleDescription

Para permitir que uma IAM entidade edite a descrição da função vinculada ao AWSServiceRoleForEMRWAL serviço

Adicione a seguinte declaração à política de permissões da IAM entidade que precisa editar a descrição de uma função vinculada ao serviço:

```
{
    "Effect": "Allow",
    "Action": [
        "iam:UpdateRoleDescription"
    ],
```

```
    "Resource": "arn:aws:iam::*:role/aws-service-role/emrwal.amazonaws.com*/
AWSServiceRoleForEMRWAL*",
    "Condition": {
        "StringLike": {
            "iam:AWSServiceName": [
                "emrwal.amazonaws.com",
                "elasticmapreduce.amazonaws.com.cn"
            ]
        }
    }
}
```

DeleteServiceLinkedRole

Para permitir que uma IAM entidade exclua a função AWSServiceRoleForEMRWAL vinculada ao serviço

Adicione a seguinte declaração à política de permissões da IAM entidade que precisa excluir uma função vinculada ao serviço:

```
{
    "Effect": "Allow",
    "Action": [
        "iam:DeleteServiceLinkedRole",
        "iam:GetServiceLinkedRoleDeletionStatus"
    ],
    "Resource": "arn:aws:iam::*:role/aws-service-role/elasticmapreduce.amazonaws.com*/
AWSServiceRoleForEMRCleanup*",
    "Condition": {
        "StringLike": {
            "iam:AWSServiceName": [
                "emrwal.amazonaws.com",
                "elasticmapreduce.amazonaws.com.cn"
            ]
        }
    }
}
```

Criação de uma função vinculada a serviços para a Amazon EMR

Você não precisa criar a AWSServiceRoleForEMRWAL função manualmente. EMRA Amazon cria essa função vinculada ao serviço automaticamente quando você cria um WAL espaço de trabalho com o EMRWAL CLI ou de AWS CloudFormation, ou HBase criará a função vinculada ao serviço

quando você configura um espaço de trabalho para a Amazon EMR WAL e a função vinculada ao serviço ainda não existe. Você deve ter permissões para criar uma função vinculada ao serviço. Por exemplo, declarações que adicionam esse recurso à política de permissões de uma IAM entidade (como um usuário, grupo ou função), consulte a seção anterior,Permissões de função vinculadas ao serviço para registro antecipado () WAL.

Editando uma função vinculada ao serviço para a Amazon EMR

A Amazon EMR não permite que você edite a função AWSServiceRoleForEMRWAL vinculada ao serviço. Depois de criar uma função vinculada ao serviço, você não pode alterar o nome da função vinculada ao serviço porque várias entidades podem fazer referência à função vinculada ao serviço. No entanto, você pode editar a descrição da função vinculada ao serviço usando. IAM

Editando uma descrição de função vinculada ao serviço (console) IAM

Você pode usar o IAM console para editar a descrição de uma função vinculada ao serviço.

Para editar a descrição de uma função vinculada ao serviço (console)

1. No painel de navegação do IAM console, escolha Funções.
2. Escolha o nome da função a ser modificada.
3. À direita de Descrição do perfil, escolha Editar.
4. Insira uma nova descrição na caixa e escolha Save changes (Salvar alterações).

Editando uma descrição de função vinculada ao serviço () IAM CLI

Você pode usar IAM os comandos do AWS Command Line Interface para editar a descrição de uma função vinculada ao serviço.

Para alterar a descrição de uma função vinculada ao serviço () CLI

1. (Opcional) Para visualizar a descrição atual de a uma função, use um dos comandos a seguir:

   ```
   $ aws iam get-role --role-name role-name
   ```

   Use o nome da função, não oARN, para se referir às funções com os CLI comandos. Por exemplo, se uma função tem o seguinteARN:`arn:aws:iam::123456789012:role/myrole`, você se refere à função como**`myrole`**.
2. Para atualizar a descrição de uma função vinculada a serviço, use um dos comandos a seguir:

```
$ aws iam update-role-description --role-name role-name --description description
```

Editando uma descrição de função vinculada ao serviço () IAM API

Você pode usar o IAM API para editar a descrição de uma função vinculada ao serviço.

Para alterar a descrição de uma função vinculada ao serviço () API

1. (Opcional) Para visualizar a descrição atual de uma função, use o comando a seguir:

   IAM API: GetRole
2. Para atualizar a descrição de uma função, use o comando a seguir:

   IAM API: UpdateRoleDescription

Excluindo uma função vinculada ao serviço para a Amazon EMR

Se você não precisar mais usar um recurso ou serviço que exija uma função vinculada ao serviço, recomendamos que você exclua essa função vinculada ao serviço. Dessa forma, você não terá uma entidade não utilizada que não seja monitorada ativamente ou mantida. No entanto, você deve limpar sua função vinculada ao serviço antes de excluí-la.

> **Note**
>
> A operação de registro de gravação antecipada não será afetada se você excluir a AWSServiceRoleForEMRWAL função, mas a Amazon EMR não excluirá automaticamente os registros criados quando o cluster for encerrado. EMR Portanto, você precisará excluir manualmente os EMR WAL registros da Amazon se excluir a função vinculada ao serviço.

Limpar um perfil vinculado ao serviço

Antes de poder usar IAM para excluir uma função vinculada ao serviço, você deve primeiro confirmar se a função não tem sessões ativas e remover os recursos usados pela função.

Para verificar se a função vinculada ao serviço tem uma sessão ativa no console IAM

1. Abra o IAM console em https://console.aws.amazon.com/iam/.

2. No painel de navegação, escolha Perfis. Selecione o nome (não a caixa de seleção) da AWSServiceRoleForEMRWAL função.
3. Na página Summary (Resumo) da função selecionada, selecione a guia Access Advisor (Consultor de acesso).
4. Na guia Consultor de acesso, revise a atividade recente para a função vinculada ao serviço.

> **Note**
>
> Se você não tiver certeza se a Amazon EMR está usando a AWSServiceRoleForEMRWAL função, você pode tentar excluir a função vinculada ao serviço. Se o serviço estiver usando a função, a exclusão falhará e você poderá visualizar as regiões em que a função vinculada ao serviço está sendo usada. Se a função vinculada ao serviço estiver sendo usada, você deverá aguardar o término da sessão antes de excluir a função vinculada ao serviço. Não é possível revogar a sessão de uma função vinculada a um serviço.

Para remover EMR os recursos da Amazon usados pelo AWSServiceRoleForEMRWAL

- Encerre todos os clusters em sua conta. Para obter mais informações, consulte Terminar um cluster.

Excluindo uma função vinculada ao serviço (console) IAM

Você pode usar o IAM console para excluir uma função vinculada ao serviço.

Para excluir uma função vinculada ao serviço (console)

1. Abra o IAM console em https://console.aws.amazon.com/iam/.
2. No painel de navegação, escolha Perfis. Marque a caixa de seleção ao lado AWSServiceRoleForEMRWAL, não o nome ou a linha em si.
3. Em ações de Função na parte superior da página, escolha a função Excluir.
4. Na caixa de diálogo de confirmação, revise os últimos dados acessados do serviço, que mostram quando cada uma das funções selecionadas acessou um AWS serviço pela última vez. Isso ajuda a confirmar se a função está ativa no momento. Para prosseguir, selecione Yes, Delete.

5. Assista às notificações do IAM console para monitorar o progresso da exclusão da função vinculada ao serviço. Como a exclusão da função IAM vinculada ao serviço é assíncrona, depois de enviar a função para exclusão, a tarefa de exclusão pode ser bem-sucedida ou falhar. Se a tarefa obtiver êxito, você poderá escolher Visualizar Detalhes ou Visualizar Recursos a partir das notificações para saber por que a exclusão falhou. Se houve falha na exclusão porque há recursos no serviço que estão sendo usados pela função, o motivo da falha incluirá uma lista de recursos.

Excluindo uma função vinculada ao serviço () IAM CLI

Você pode usar IAM os comandos do AWS Command Line Interface para excluir uma função vinculada ao serviço. Como uma função vinculada ao serviço não podem ser excluída se estiver sendo usada ou tiver recursos associados, você deverá enviar uma solicitação de exclusão. Se essas condições não forem atendidas, essa solicitação poderá ser negada.

Para excluir uma função vinculada ao serviço () CLI

1. Para verificar o status da tarefa de exclusão, você deve capturar o `deletion-task-id` da resposta. Digite o seguinte comando para enviar uma solicitação de exclusão de função vinculada ao serviço:

```
$ aws iam delete-service-linked-role --role-name AWSServiceRoleForEMRWAL
```

2. Digite o seguinte comando para verificar o estado da tarefa de exclusão:

```
$ aws iam get-service-linked-role-deletion-status --deletion-task-id deletion-task-id
```

   O status da tarefa de exclusão pode ser `NOT_STARTED`, `IN_PROGRESS`, `SUCCEEDED`, ou `FAILED`. Se a exclusão falhar, a chamada informará o motivo de falha para que você possa solucionar o problema.

Excluindo uma função vinculada ao serviço () IAM API

Você pode usar o IAM API para excluir uma função vinculada ao serviço. Como uma função vinculada ao serviço não podem ser excluída se estiver sendo usada ou tiver recursos associados, você deverá enviar uma solicitação de exclusão. Se essas condições não forem atendidas, essa solicitação poderá ser negada.

Para excluir uma função vinculada ao serviço () API

1. Para enviar uma solicitação de exclusão para uma função vinculada ao serviço, ligue. DeleteServiceLinkedRole Na solicitação, especifique o nome da AWSServiceRoleForEMRWAL função.

   Para verificar o status da tarefa de exclusão, você deve capturar o `DeletionTaskId` da resposta.

2. Para verificar o status da exclusão, ligue GetServiceLinkedRoleDeletionStatus. Na solicitação, especifique o `DeletionTaskId`.

   O status da tarefa de exclusão pode ser `NOT_STARTED`, `IN_PROGRESS`, `SUCCEEDED`, ou `FAILED`. Se a exclusão falhar, a chamada informará o motivo de falha para que você possa solucionar o problema.

Regiões suportadas para AWSServiceRoleForEMRWAL

A Amazon EMR oferece suporte ao uso da função AWSServiceRoleForEMRWAL vinculada ao serviço nas seguintes regiões.

Nome da região	Identidade da região	Support na Amazon EMR
Leste dos EUA (Norte da Virgínia)	us-east-1	Sim
Leste dos EUA (Ohio)	us-east-2	Sim
Oeste dos EUA (N. da Califórnia)	us-west-1	Sim
Oeste dos EUA (Oregon)	us-west-2	Sim
Ásia-Pacífico (Mumbai)	ap-south-1	Sim
Ásia-Pacífico (Singapura)	ap-southeast-1	Sim
Ásia-Pacífico (Sydney)	ap-southeast-2	Sim
Ásia-Pacífico (Tóquio)	ap-northeast-1	Sim
Europa (Frankfurt)	eu-central-1	Sim

Nome da região	Identidade da região	Support na Amazon EMR
Europa (Irlanda)	eu-west-1	Sim

Personalize IAM funções

Talvez você queira personalizar as funções e permissões do IAM serviço para limitar os privilégios de acordo com seus requisitos de segurança. Para personalizar as permissões, recomendamos que você crie novas funções e políticas. Comece com as permissões nas políticas gerenciadas para as funções padrão (por exemplo, `AmazonElasticMapReduceforEC2Role` e `AmazonElasticMapReduceRole`). Em seguida, copie e cole o conteúdo em novas declarações de política, modifique as permissões conforme apropriado e anexe as políticas de permissões modificadas às funções que criar. Você deve ter as IAM permissões apropriadas para trabalhar com funções e políticas. Para obter mais informações, consulte Permitir que usuários e grupos criem e modifiquem perfis.

Se você criar um EMR papel personalizado paraEC2, siga o fluxo de trabalho básico, que cria automaticamente um perfil de instância com o mesmo nome. A Amazon EC2 permite que você crie perfis e funções de instância com nomes diferentes, mas a Amazon EMR não oferece suporte a essa configuração e isso resulta em um erro de “perfil de instância inválido” quando você cria o cluster.

> **Important**
>
> As políticas em linha não são atualizadas automaticamente quando os requisitos do serviço são alterados. Se você criar e anexar políticas em linha, lembre-se de que podem ocorrer atualizações de serviço que causem erros de permissão repentinamente. Para obter mais informações, consulte Políticas gerenciadas e políticas embutidas no Guia IAM do usuário e. Especifique IAM funções personalizadas ao criar um cluster

Para obter mais informações sobre como trabalhar com IAM funções, consulte os seguintes tópicos no Guia IAM do usuário:

- Criação de uma função para delegar permissões a um serviço AWS
- Modificar uma função
- Excluir um perfil

Especifique IAM funções personalizadas ao criar um cluster

Você especifica a função de serviço para a Amazon EMR e a função para o perfil de EC2 instância da Amazon ao criar um cluster. O usuário que está criando clusters precisa de permissões para recuperar e atribuir funções à Amazon EMR e às EC2 instâncias. Caso contrário, ocorrerá um EC2 erro na conta não autorizada a fazer chamadas. Para obter mais informações, consulte Permitir que usuários e grupos criem e modifiquem perfis.

Usar o console para especificar perfis personalizados

Ao criar um cluster, você pode especificar uma função de serviço personalizada para a AmazonEMR, uma função personalizada para o perfil da EC2 instância e uma função personalizada do Auto Scaling usando as opções avançadas. Quando você usa as opções rápidas, a função de serviço padrão e a função padrão para o perfil da EC2 instância são especificadas. Para obter mais informações, consulte IAMfunções de serviço usadas pela Amazon EMR.

Console

Para especificar IAM funções personalizadas com o console

Ao criar um cluster com o console, você deve especificar uma função de serviço personalizada para a Amazon EMR e uma função personalizada para o perfil da EC2 instância. Para obter mais informações, consulte IAMfunções de serviço usadas pela Amazon EMR.

1. Faça login no AWS Management Console e abra o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Em Configuração e permissões de segurança, encontre a IAMfunção para o perfil da instância e a função do serviço para EMR os campos da Amazon. Para cada tipo de função, selecione uma função na lista. Apenas as funções em sua conta que têm a política de confiança apropriada para esse tipo de função são listadas.
4. Escolha qualquer outra opção que se aplique ao cluster.
5. Para iniciar o cluster, escolha Criar cluster.

Use o AWS CLI para especificar funções personalizadas

Você pode especificar explicitamente uma função de serviço para a Amazon EMR e uma função de serviço para EC2 instâncias de cluster usando opções com o `create-cluster` comando

do AWS CLI. Use a opção `--service-role` para especificar a função de serviço. Use o `InstanceProfile` argumento da `--ec2-attributes` opção para especificar a função do perfil da EC2 instância.

O perfil do Auto Scaling é especificado usando uma opção separada, `--auto-scaling-role`. Para obter mais informações, consulte Usar o ajuste de escala automático com uma política personalizada para grupos de instâncias.

Para especificar IAM funções personalizadas usando o AWS CLI

- O comando a seguir especifica a função de serviço personalizada, *MyCustomServiceRoleForEMR*e uma função personalizada para o perfil da EC2 instância, *MyCustomServiceRoleForClusterEC2Instances*, ao iniciar um cluster. Este exemplo usa a EMR função padrão da Amazon.

  > **Note**
  >
  > Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 \
--applications Name=Hive Name=Pig --service-role MyCustomServiceRoleForEMR \
--ec2-attributes InstanceProfile=MyCustomServiceRoleForClusterEC2Instances,\
KeyName=myKey --instance-type m5.xlarge --instance-count 3
```

Você pode usar essas opções para especificar funções padrão explicitamente em vez de usar a opção `--use-default-roles`. A `--use-default-roles` opção especifica a função de serviço e a função do perfil de EC2 instância definido no `config` arquivo para o. AWS CLI

O exemplo a seguir demonstra o conteúdo de um `config` arquivo para AWS CLI as funções personalizadas especificadas para a Amazon. EMR Com esse arquivo de configuração, quando a `--use-default-roles` opção é especificada, o cluster é criado usando o *MyCustomServiceRoleForEMR* e *MyCustomServiceRoleForClusterEC2Instances*. Por padrão, o `config` arquivo especifica o padrão `service_role` como `AmazonElasticMapReduceRole` e o padrão `instance_profile` como`EMR_EC2_DefaultRole`.

```
[default]
output = json
region = us-west-1
aws_access_key_id = myAccessKeyID
aws_secret_access_key = mySecretAccessKey
emr =
     service_role = MyCustomServiceRoleForEMR
     instance_profile = MyCustomServiceRoleForClusterEC2Instances
```

Configurar IAM funções para EMRFS solicitações para o Amazon S3

> **Note**
>
> A capacidade de mapeamento de EMRFS funções descrita nesta página foi aprimorada com a introdução do Amazon S3 Access Grants no Amazon EMR 6.15.0. Para uma solução de controle de acesso escalável para seus dados no Amazon S3, recomendamos que você use o S3 Access Grants com a Amazon. EMR

Quando um aplicativo executado em um cluster faz referência a dados usando o `s3://mydata` formato, a Amazon EMR usa EMRFS para fazer a solicitação. Para interagir com o Amazon S3, EMRFS assume as políticas de permissões que estão anexadas ao seu perfil de instância da Amazon EC2. O mesmo perfil de EC2 instância da Amazon é usado independentemente do usuário ou grupo que executa o aplicativo ou da localização dos dados no Amazon S3.

Se você tiver um cluster com vários usuários que precisam de diferentes níveis de acesso aos dados no Amazon S3EMRFS, você pode definir uma configuração de segurança com IAM funções para. EMRFS EMRFSpode assumir uma função de serviço diferente para EC2 instâncias de cluster com base no usuário ou grupo que faz a solicitação ou com base na localização dos dados no Amazon S3. Cada IAM função do EMRFS pode ter permissões diferentes para acesso a dados no Amazon S3. Para obter mais informações sobre a função de serviço para EC2 instâncias de cluster, consulteFunção de serviço para EC2 instâncias de cluster (perfil de EC2 instância).

O uso de IAM funções personalizadas para EMRFS é suportado nas EMR versões 5.10.0 e posteriores da Amazon. Se você usa uma versão anterior ou tem requisitos além das IAM funções a serem EMRFS fornecidas, você pode criar um provedor de credenciais personalizado. Para obter mais informações, consulte Autorização do acesso aos EMRFS dados no Amazon S3.

Ao usar uma configuração de segurança para especificar IAM funçõesEMRFS, você configura mapeamentos de funções. Cada mapeamento de função especifica uma IAM função que corresponde aos identificadores. Esses identificadores determinam a base para o acesso ao Amazon EMRFS S3 por meio de. Os identificadores podem ser usuários, grupos ou prefixos do Amazon S3 que indicam um local de dados. Quando EMRFS faz uma solicitação ao Amazon S3, se a solicitação corresponder à base de acesso, EMRFS as EC2 instâncias de cluster assumem a IAM função correspondente para a solicitação. As IAM permissões associadas a essa função se aplicam em vez das IAM permissões anexadas à função de serviço para EC2 instâncias de cluster.

Os usuários e os grupos em um mapeamento de função são usuários e grupos do Hadoop definidos no cluster. Usuários e grupos são transmitidos EMRFS no contexto do aplicativo que os usa (por exemplo, representação de YARN usuário). O prefixo do Amazon S3 pode ser um especificador do bucket de qualquer profundidade (por exemplo `s3://mybucket` ou `s3://mybucket/myproject/mydata`). Você pode especificar vários identificadores em um único mapeamento de função, mas todos devem ser do mesmo tipo.

> **Important**
>
> IAMfunções para EMRFS fornecer isolamento em nível de aplicativo entre os usuários do aplicativo. Isso não fornece isolamento no nível do host entre os usuários no host. Qualquer usuário com acesso ao cluster pode ignorar o isolamento para assumir qualquer uma das funções.

Quando um aplicativo de cluster faz uma solicitação ao Amazon S3 por meio deleEMRFS, EMRFS avalia os mapeamentos de funções na ordem de cima para baixo em que aparecem na configuração de segurança. Se uma solicitação feita por meio de EMRFS nenhum identificador corresponder a nenhum identificador, EMRFS volte a usar a função de serviço para EC2 instâncias de cluster. Por esse motivo, recomendamos que as políticas anexadas ao perfil limitem as permissões ao Amazon S3. Para obter mais informações, consulte Função de serviço para EC2 instâncias de cluster (perfil de EC2 instância).

Configurar funções do

Antes de definir uma configuração de segurança com IAM funções paraEMRFS, planeje e crie as funções e as políticas de permissão a serem anexadas às funções. Para obter mais informações, consulte Como funcionam as funções das EC2 instâncias? no Guia do IAM usuário. Ao criar políticas de permissões, recomendamos que você comece com a política gerenciada anexada à EMR

função padrão da Amazon eEC2, em seguida, edite essa política de acordo com seus requisitos. O nome de perfil padrão é `EMR_EC2_DefaultRole`, e a política gerenciada padrão a ser editada é `AmazonElasticMapReduceforEC2Role`. Para obter mais informações, consulte Função de serviço para EC2 instâncias de cluster (perfil de EC2 instância).

Atualizar políticas de confiança para permissões para assumir perfil

Cada função EMRFS usada deve ter uma política de confiança que permita que a EMR função Amazon do cluster EC2 a assuma. Da mesma forma, a EMR função do cluster na Amazon EC2 deve ter uma política de confiança que permita que as EMRFS funções a assumam.

O exemplo de política de confiança a seguir está anexado às funções deEMRFS. A declaração permite que a EMR função padrão da Amazon assuma EC2 a função. Por exemplo, se você tiver duas EMRFS funções fictícias `EMRFSRole_First` e `EMRFSRole_Second` essa declaração de política for adicionada às políticas de confiança de cada uma delas.

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "AWS":"arn:aws:iam::AWSAcctID:role/EMR_EC2_DefaultRole"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

Além disso, o exemplo de declaração de política de confiança a seguir é adicionado ao `EMR_EC2_DefaultRole` para permitir que as duas EMRFS funções fictícias a assumam.

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "AWS": ["arn:aws:iam::AWSAcctID:role/EMRFSRole_First",
 "arn:aws:iam::AWSAcctID:role/EMRFSRole_Second"]
         },
         "Action":"sts:AssumeRole"
```

```
      }
   ]
}
```

Para atualizar a política de confiança de uma IAM função

Abra o IAM console em https://console.aws.amazon.com/iam/.

1. Selecione Roles (funções), insira o nome da função em Search (Pesquisar) e, em seguida, selecione o Role name (Nome da função).
2. Escolha Trust relationships (Relacionamentos de confiança), Edit trust relationship (Editar relacionamento de confiança).
3. Adicione uma instrução de confiança de acordo com o Documento da política, de acordo com as diretrizes acima e selecione Atualizar política de confiança.

Especificar um perfil como um usuário da chave

Se o perfil permitir acesso a um local no Amazon S3 que é criptografado usando uma AWS KMS key, especifique o perfil como um usuário de chaves. Isso dá permissão à função para usar a KMS chave. Para obter mais informações, consulte Políticas de chaves no AWS KMS no Guia do desenvolvedor do AWS Key Management Service .

Defina uma configuração de segurança com IAM funções para EMRFS

> **Important**
>
> Se nenhuma das IAM funções especificadas se aplicar, EMRFS volte para a EMR função da Amazon paraEC2. EMRFS Considere a possibilidade de personalizar esse perfil para restringir permissões para o Amazon S3 conforme apropriado para suas aplicação e especificar esse perfil personalizado em vez de `EMR_EC2_DefaultRole` ao criar um cluster. Para ter mais informações, consulte Personalize IAM funções e Especifique IAM funções personalizadas ao criar um cluster.

Para especificar IAM funções para EMRFS solicitações ao Amazon S3 usando o console

1. Crie uma configuração de segurança que especifica os mapeamentos de função:
   a. No EMR console da Amazon, selecione Configurações de segurança, Criar.

b. Digite um nome em Name (Nome) para a configuração de segurança. Esse nome é usado para especificar a configuração de segurança ao criar um cluster.

c. Escolha Usar IAM funções para EMRFS solicitações ao Amazon S3.

d. Selecione uma IAMfunção a ser aplicada e, em Base para acesso, selecione um tipo de identificador (usuários, grupos ou prefixos S3) na lista e insira os identificadores correspondentes. Se você usar vários identificadores, separe-os com uma vírgula e sem espaço. Para obter mais informações sobre cada tipo de identificador, consulte a JSON configuration reference abaixo.

e. Escolha Add role (Adicionar função) para configurar mapeamentos de funções adicionais, conforme descrito na etapa anterior.

f. Configure outras opções de configuração de segurança conforme apropriado e escolha Create (Criar). Para obter mais informações, consulte Criar uma configuração de segurança.

2. Especifique a configuração de segurança criada acima ao criar um cluster. Para obter mais informações, consulte Especificar uma configuração de segurança para um cluster.

Para especificar IAM funções para EMRFS solicitações ao Amazon S3 usando o AWS CLI

1. Use o `aws emr create-security-configuration` comando, especificando um nome para a configuração de segurança e os detalhes da configuração de segurança no JSON formato.

   O comando de exemplo mostrado a seguir cria uma configuração de segurança com o nome `EMRFS_Roles_Security_Configuration`. Ele é baseado em uma JSON estrutura no arquivo`MyEmrfsSecConfig.json`, que é salva no mesmo diretório em que o comando é executado.

   ```
   aws emr create-security-configuration --name EMRFS_Roles_Security_Configuration --
   security-configuration file://MyEmrFsSecConfig.json.
   ```

   Use as diretrizes a seguir para a estrutura do arquivo `MyEmrFsSecConfig.json`. Você pode especificar essa estrutura juntamente com estruturas de outras opções de configuração de segurança. Para obter mais informações, consulte Criar uma configuração de segurança.

   Veja a seguir um exemplo de JSON trecho para especificar IAM funções personalizadas EMRFS em uma configuração de segurança. Ele demonstra mapeamentos de perfil para os três tipos diferentes de identificadores, seguidos por uma referência de parâmetro.

```
{
  "AuthorizationConfiguration": {
    "EmrFsConfiguration": {
      "RoleMappings": [{
        "Role": "arn:aws:iam::123456789101:role/allow_EMRFS_access_for_user1",
        "IdentifierType": "User",
        "Identifiers": [ "user1" ]
      },{
        "Role": "arn:aws:iam::123456789101:role/allow_EMRFS_access_to_MyBuckets",
        "IdentifierType": "Prefix",
        "Identifiers": [ "s3://MyBucket/","s3://MyOtherBucket/" ]
      },{
        "Role": "arn:aws:iam::123456789101:role/allow_EMRFS_access_for_AdminGroup",
        "IdentifierType": "Group",
        "Identifiers": [ "AdminGroup" ]
      }]
    }
  }
}
```

Parâmetro	Descrição
`"AuthorizationConfiguration":`	Obrigatório.
`"EmrFsConfiguration":`	Obrigatório. Contém mapeamentos de perfil.
`"RoleMappings":`	Obrigatório. Contém uma ou mais definições de mapeamento de perfil. Os mapeamentos de perfil são avaliados na ordem em que aparecem, de cima para baixo. Se um mapeamento de função for avaliado como verdadeiro para uma EMRFS chamada de dados no Amazon S3, nenhum outro mapeamento de função será avaliado EMRFS e usará a função IAM especificada para a solicitação. O mapeamento de perfil tem os seguintes parâmetros obrigatórios:

Parâmetro	Descrição
`"Role":`	Especifica o ARN identificador de uma IAM função no formato`arn:aws:iam:: account-id :role/role-name` . Essa é a IAM função que a Amazon EMR assume se a EMRFS solicitação para o Amazon S3 corresponder a qualquer uma das `Identifiers` especificadas.
`"IdentifierType":`	Pode ser um dos seguintes: • `"User"` especifica que os identificadores são um ou mais usuários do Hadoop, que podem ser usuários de contas Linux ou entidades principais do Kerberos. Quando a EMRFS solicitação se origina com o usuário ou usuários especificados, a IAM função é assumida. • `"Prefix"` especifica que o identificador é um local do Amazon S3. A IAM função é assumida para chamadas para o local ou locais com os prefixos especificados. Por exemplo, o prefixo `s3://mybucket/` corresponde a `s3://mybucket/mydir` e `s3://mybucket/yetanotherdir` . • `"Group"` especifica que os identificadores são um ou mais grupos do Hadoop. A IAM função é assumida se a solicitação for originada de um usuário no grupo ou grupos especificados.
`"Identifiers":`	Especifica um ou mais identificadores do tipo de identificador adequado. Separe múltiplos identificadores por vírgulas sem espaços.

2. Use o comando `aws emr create-cluster` para criar um cluster e especifique a configuração de segurança que você criou na etapa anterior.

   O exemplo a seguir cria um cluster com aplicativos Hadoop de núcleo padrão instalados. O cluster usa a configuração de segurança criada acima como `EMRFS_Roles_Security_Configuration` e também usa uma EMR função personalizada da Amazon para EC2`EC2_Role_EMR_Restrict_S3`,, que é especificada usando o `InstanceProfile` argumento do `--ec2-attributes` parâmetro.

   > **Note**
   >
   > Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --name MyEmrFsS3RolesCluster \
--release-label emr-7.2.0 --ec2-attributes
 InstanceProfile=EC2_Role_EMR_Restrict_S3,KeyName=MyKey \
--instance-type m5.xlarge --instance-count 3 \
--security-configuration EMRFS_Roles_Security_Configuration
```

Use políticas baseadas em recursos para que a Amazon EMR acesse o AWS Glue Data Catalog

Se você usa o AWS Glue em conjunto com o Hive, o Spark ou o Presto na AmazonEMR, o AWS Glue oferece suporte a políticas baseadas em recursos para controlar o acesso aos recursos do Catálogo de Dados. Esses recursos incluem bancos de dados, tabelas, conexões e funções definidas pelo usuário. Para obter mais informações, consulte Políticas baseadas em recursos no AWS Glue no Guia do desenvolvedor do AWS Glue.

Ao usar políticas baseadas em recursos para limitar o acesso ao AWS Glue na AmazonEMR, o principal que você especifica na política de permissões deve ser a função ARN associada ao perfil de EC2 instância especificado quando um cluster é criado. Por exemplo, para uma política baseada em recursos anexada a um catálogo, você pode especificar a função da função de serviço padrão ARN para instâncias de clusterEC2, *EMR_EC2_DefaultRole* como o`Principal`, usando o formato mostrado no exemplo a seguir:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

A ferramenta *acct-id* pode ser diferente do ID da conta AWS Glue. Isso permite o acesso a partir de EMR clusters em contas diferentes. Você pode especificar várias entidades principais, cada uma de uma conta diferente.

Use IAM funções com aplicativos que chamam AWS serviços diretamente

Os aplicativos executados nas EC2 instâncias de um cluster podem usar o perfil da EC2 instância para obter credenciais de segurança temporárias ao chamar AWS serviços.

As versões do Hadoop disponíveis com a EMR versão 2.3.0 e posterior da Amazon já foram atualizadas para fazer uso das funções. IAM Se seu aplicativo é executado estritamente sobre a arquitetura do Hadoop e não chama diretamente nenhum serviço AWS, ele deve funcionar com IAM funções sem modificação.

Se seu aplicativo chamar serviços AWS diretamente, você precisará atualizá-lo para aproveitar as IAM funções. Isso significa que, em vez de obter as credenciais da conta `/etc/hadoop/conf/core-site.xml` nas EC2 instâncias do cluster, seu aplicativo usa an SDK para acessar os recursos usando IAM funções ou chama os metadados da EC2 instância para obter as credenciais temporárias.

Para acessar AWS recursos com IAM funções usando um SDK

- Os tópicos a seguir mostram como usar vários deles AWS SDKs para acessar credenciais temporárias usando IAM funções. Cada tópico começa com uma versão de um aplicativo que não usa IAM funções e, em seguida, orienta você pelo processo de conversão desse aplicativo em IAM funções de uso.
  - Usando IAM funções para EC2 instâncias da Amazon com o SDK for Java no Guia do AWS SDK for Java desenvolvedor
  - Usando IAM funções para EC2 instâncias da Amazon com o SDK for. NETno Guia do AWS SDK for .NET desenvolvedor
  - Usando IAM funções para EC2 instâncias da Amazon com o SDK for PHP no Guia do AWS SDK for PHP desenvolvedor
  - Usando IAM funções para EC2 instâncias da Amazon com o SDK for Ruby no Guia do AWS SDK for Ruby desenvolvedor

Para obter credenciais temporárias dos metadados da EC2 instância

- Chame o seguinte URL de uma EC2 instância que está sendo executada com a IAM função especificada, que retorna as credenciais de segurança temporárias associadas (AccessKeyId, SecretAccessKey SessionToken, e Expiração). O exemplo a seguir usa o perfil de instância padrão da AmazonEMR,`EMR_EC2_DefaultRole`.

```
GET http://169.254.169.254/latest/meta-data/iam/security-
credentials/EMR_EC2_DefaultRole
```

Para obter mais informações sobre como criar aplicativos que usam IAM funções, consulte Conceder acesso a AWS recursos para aplicativos executados em EC2 instâncias da Amazon.

Para obter mais informações sobre credenciais de segurança temporárias, consulte Using temporary security credentials no guia Using Temporary Security Credentials.

Permitir que usuários e grupos criem e modifiquem perfis

IAMdiretores (usuários e grupos) que criam, modificam e especificam funções para um cluster, incluindo funções padrão, devem ter permissão para realizar as seguintes ações. Para obter detalhes sobre cada ação, consulte Ações na IAMAPIReferência.

- `iam:CreateRole`
- `iam:PutRolePolicy`
- `iam:CreateInstanceProfile`
- `iam:AddRoleToInstanceProfile`
- `iam:ListRoles`
- `iam:GetPolicy`
- `iam:GetInstanceProfile`
- `iam:GetPolicyVersion`
- `iam:AttachRolePolicy`
- `iam:PassRole`

A permissão `iam:PassRole` permite a criação do cluster. As permissões restantes permitem a criação das funções padrão.

Para obter informações sobre como atribuir permissões a um usuário, consulte Alteração de permissões para um usuário no Guia do IAM usuário.

Exemplos de políticas EMR baseadas em identidade da Amazon

Por padrão, usuários e funções não têm permissão para criar ou modificar EMR recursos da Amazon. Eles também não podem realizar tarefas usando o AWS Management Console, AWS CLI, ou AWS API. Um IAM administrador deve criar IAM políticas que concedam aos usuários e funções permissão para realizar API operações específicas nos recursos especificados de que precisam. O administrador deve anexar essas políticas aos usuários ou grupos que exigem essas permissões.

Para saber como criar uma política IAM baseada em identidade usando esses exemplos de documentos JSON de política, consulte Criação de políticas na JSON guia do IAMusuário.

Tópicos

- Melhores práticas políticas para a Amazon EMR
- Permitir que usuários visualizem suas próprias permissões
- Políticas EMR gerenciadas pela Amazon
- IAMpolíticas para acesso baseado em tags a clusters e notebooks EMR
- Negando a ação ModifyInstanceGroup
- Solução de problemas de EMR identidade e acesso da Amazon

Melhores práticas políticas para a Amazon EMR

As políticas baseadas em identidade são muito eficientes. Eles determinam se alguém pode criar, acessar ou excluir EMR recursos da Amazon em sua conta. Essas ações podem gerar custos para sua AWS conta. Ao criar ou editar políticas baseadas em identidade, siga estas diretrizes e recomendações:

- Comece a usar políticas AWS gerenciadas — Para começar a usar a Amazon EMR rapidamente, use políticas AWS gerenciadas para dar aos seus funcionários as permissões de que precisam. Essas políticas já estão disponíveis em sua conta e são mantidas e atualizadas pela AWS. Para obter mais informações, consulte Comece a usar permissões com políticas AWS gerenciadas no Guia IAM do usuário Políticas EMR gerenciadas pela Amazon e.

- Conceder privilégio mínimo: ao criar políticas personalizadas, conceda apenas as permissões necessárias para executar uma tarefa. Comece com um conjunto mínimo de permissões e conceda permissões adicionais conforme necessário. Fazer isso é mais seguro do que começar com permissões que são muito lenientes e tentar restringi-las superiormente. Para obter mais informações, consulte Conceder privilégios mínimos no Guia do IAM usuário.
- Habilitar MFA operações confidenciais — Para maior segurança, exija que os usuários usem a autenticação multifator (MFA) para acessar recursos ou API operações confidenciais. Para obter mais informações, consulte Usando a autenticação multifator (MFA) AWS no Guia do IAM usuário.
- Usar condições de política para segurança adicional: na medida do possível, defina as condições sob as quais suas políticas baseadas em identidade permitem o acesso a um recurso. Por exemplo, você pode gravar condições para especificar um intervalo de endereços IP permitidos do qual a solicitação deve partir. Você também pode criar condições para permitir solicitações somente dentro de um intervalo de data ou hora especificado, ou para exigir o uso de SSL ouMFA. Para obter mais informações, consulte Elementos IAM JSON da política: Condição no Guia IAM do usuário.

Permitir que usuários visualizem suas próprias permissões

Este exemplo mostra como você pode criar uma política que permite que os usuários visualizem as políticas gerenciadas e em linha anexadas as respectivas identidades de usuário. Essa política inclui permissões para concluir essa ação no console ou programaticamente usando o AWS CLI ou. AWS API

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Sid":"ViewOwnUserInfo",
         "Effect":"Allow",
         "Action":[
            "iam:GetUser",
            "iam:GetUserPolicy",
            "iam:ListAttachedUserPolicies",
            "iam:ListGroupsForUser",
            "iam:ListUserPolicies"
         ],
         "Resource":[
            "arn:aws:iam::*:user/${aws:username}"
         ]
```

```
      },
      {
         "Sid":"NavigateInConsole",
         "Effect":"Allow",
         "Action":[
            "iam:GetGroupPolicy",
            "iam:GetPolicy",
            "iam:GetPolicyVersion",
            "iam:ListAttachedGroupPolicies",
            "iam:ListGroupPolicies",
            "iam:ListGroups",
            "iam:ListPolicies",
            "iam:ListPolicyVersions",
            "iam:ListUsers"
         ],
         "Resource":"*"
      }
   ]
}
```

Políticas EMR gerenciadas pela Amazon

A maneira mais fácil de conceder acesso total ou acesso somente de leitura às EMR ações necessárias da Amazon é usar as políticas IAM gerenciadas da Amazon. EMR Políticas gerenciadas oferecem o benefício de serem atualizadas automaticamente se os requisitos de permissões forem alterados. Se você usar políticas em linha, podem ocorrer alterações de serviço que provoquem erros de permissão.

A Amazon EMR substituirá as políticas gerenciadas existentes (políticas v1) em favor de novas políticas gerenciadas (políticas v2). As novas políticas gerenciadas foram reduzidas para se alinharem às melhores práticas. AWS Depois que as políticas gerenciadas v1 existentes forem descontinuadas, você não poderá anexar essas políticas a nenhuma nova IAM função ou usuário. Os perfis e os usuários existentes que usam políticas defasadas podem continuar a usá-las. As políticas gerenciadas v2 restringem o acesso usando etiquetas. Eles permitem somente EMR ações específicas da Amazon e exigem recursos de cluster marcados com uma chave EMR específica. É recomendável analisar cuidadosamente a documentação antes de usar as novas políticas v2.

As políticas v1 serão marcadas como obsoletas com um ícone de aviso ao lado delas na lista de políticas no console. IAM As políticas defasadas terão as seguintes características:

- Continuarão funcionando para todos os usuários, grupos e perfis atualmente conectados. Nada é rompido.
- Não é possível anexar a novos usuários, grupos ou perfis. Se você desvincular uma das políticas de uma entidade atual, não poderá anexá-la novamente.
- Após separar uma política v1 de todas as entidades atuais, a política não estará mais visível e não poderá mais ser usada.

A tabela a seguir resume as alterações entre as políticas atuais (v1) e v2.

Alterações na política EMR gerenciada pela Amazon

Tipo de política	Nomes de política	Finalidade da política	Alterações na política v2
Função EMR de serviço padrão e política gerenciada anexada	Nome da função: EMR_ DefaultRole Política V1 (a ser descontinuada): AmazonElasticMapReduceRole(Função de serviço) EMR Nome da política v2 (com escopo reduzido): **AmazonEMRServicePolicy_v2**	Permite que EMR a Amazon chame outros AWS serviços em seu nome ao provisionar recursos e realizar ações de nível de serviço. Essa função é necessária para todos os clusters.	A política adiciona a nova permissão `"ec2:DescribeInstanceTypeOfferings"` . Essa API operação retorna uma lista de tipos de instância que são compatíveis com uma lista de determinadas zonas de disponibilidade.
IAMpolítica gerenciada para EMR acesso total à Amazon por usuário, função ou grupo vinculado	Nome da política v2 (no escopo): **AmazonEMRServicePolicy_v2**	Permite aos usuários permissões totais para EMR ações. Inclui iam: PassRole permissões para recursos.	A política adiciona o pré-requisito de que os usuários devem adicionar etiquetas de usuário aos recursos para poderem usar essa política. Consulte

Tipo de política	Nomes de política	Finalidade da política	Alterações na política v2
			Etiquetar recursos para usar políticas gerenciadas. iam: a PassRole ação requer iam: PassedToService condição definida para o serviço especificado. O acesso à AmazonEC2 , Amazon S3 e outros serviços não é permitido por padrão. Consulte Política IAM gerenciada para acesso total (política padrão gerenciada v2).
IAMpolítica gerenciada para acesso somente de leitura por usuário, função ou grupo anexado	Política v1 (a ser defasada) : **AmazonElasticMapReduceReadOnlyAccess** Nome da política v2 (no escopo): **AmazonEMRReadOnlyAccessPolicy_v2**	Permite aos usuários permissões somente de leitura para ações da AmazonEMR.	As permissões concedem somente ações específicas de leitura do elasticmapreduce. O acesso ao Amazon S3 é um acesso não concedido por padrão. Consulte Política IAM gerenciada para acesso somente leitura (política padrão gerenciada v2).

Tipo de política	Nomes de política	Finalidade da política	Alterações na política v2
Função EMR de serviço padrão e política gerenciada anexada	Nome da função: EMR_ DefaultRole Política V1 (a ser descontinuada): AmazonElasticMapReduceRole(Função de serviço) EMR Nome da política v2 (com escopo reduzido): **AmazonEMRServicePolicy_v2**	Permite que EMR a Amazon chame outros AWS serviços em seu nome ao provisionar recursos e realizar ações de nível de serviço. Essa função é necessári a para todos os clusters.	O perfil de serviço v2 e a política padrão v2 substituem o perfil e a política defasados . A política adiciona o pré-requisito de que os usuários devem adicionar etiquetas de usuário aos recursos para poderem usar essa política. Consulte Etiquetar recursos para usar políticas gerenciadas. Consulte Função de serviço para a Amazon EMR (EMRfunção).

Tipo de política	Nomes de política	Finalidade da política	Alterações na política v2
Função de serviço para EC2 instâncias de cluster (perfil de EC2 instância)	Nome da função: EMR_ EC2 _ DefaultRole Nome da política obsoleta: AmazonElasticMapReduceforEC2Role	Permite que aplicativos executados em um EMR cluster acessem outros AWS recursos, como o Amazon S3. Por exemplo, se você executar trabalhos do Apache Spark que processam dados do Amazon S3, a política precisará permitir o acesso a esses recursos.	Tanto o perfil padrão como a política padrão estão prestes a serem defasados. Não há nenhuma função ou política gerenciada AWS padrão de substituição. É necessário fornecer uma política baseada em recursos ou em identidade. Isso significa que, por padrão, os aplicativos executados em um EMR cluster não têm acesso ao Amazon S3 ou a outros recursos, a menos que você os adicione manualmente à política. Consulte Perfil padrão e política gerenciada.

Tipo de política	Nomes de política	Finalidade da política	Alterações na política v2
Outras políticas EC2 de função de serviço	Nomes de políticas atuais: AmazonElasticMapReduceforAutoScalingRole AmazonElasticMapReduceEditorsRole,, Uma mazonEMRCleanup política	Fornece as permissões que a Amazon EMR precisa para acessar outros AWS recursos e realizar ações se estiver usando escalabilidade automática, notebooks ou para limpar EC2 recursos.	Nenhuma alteração na v2.

Objetivo de proteção: PassRole

As políticas EMR gerenciadas padrão de permissões completas da Amazon incorporam configurações `iam:PassRole` de segurança, incluindo as seguintes:

- `iam:PassRole`permissões somente para EMR funções padrão específicas da Amazon.
- `iam:PassedToService`condições que permitem que você use a política somente com AWS serviços específicos, como `elasticmapreduce.amazonaws.com` `ec2.amazonaws.com` e.

Você pode ver a JSON versão das políticas A mazonEMRFull AccessPolicy _v2 e A mazonEMRServicePolicy_v2 no console. IAM É recomendável criar novos clusters com políticas gerenciadas v2.

Para criar políticas personalizadas, recomendamos que você comece com as políticas gerenciadas e edite-as de acordo com seus requisitos.

Para obter informações sobre como anexar políticas a usuários (diretores), consulte Como trabalhar com políticas gerenciadas usando o AWS Management Console no Guia do IAM Usuário.

Etiquetar recursos para usar políticas gerenciadas

A mazonEMRService Policy_v2 e A mazonEMRFullAccessPolicy_v2 dependem do acesso limitado aos recursos que a Amazon provisiona ou usa. EMR Obtém-se o escopo reduzido restringindo o acesso somente aos recursos que têm uma etiqueta de usuário predefinida associada a eles. Ao

usar qualquer uma dessas duas políticas, é necessário passar a etiqueta de usuário predefinida `for-use-with-amazon-emr-managed-policies = true` ao provisionar o cluster. A Amazon então EMR propagará automaticamente essa tag. Além disso, é necessário adicionar uma etiqueta de usuário aos recursos listados na seção a seguir. Se você usa o EMR console da Amazon para iniciar seu cluster, consulte[Considerações sobre o uso do EMR console da Amazon para lançar clusters com políticas gerenciadas v2](#).

Para usar políticas gerenciadas, passe a tag de usuário `for-use-with-amazon-emr-managed-policies = true` ao provisionar um cluster com o CLISDK,, ou outro método.

Quando você passa a tag, a Amazon EMR propaga a tag para a sub-rede privadaENI, a EC2 instância e EBS os volumes que ela cria. A Amazon EMR também marca automaticamente os grupos de segurança que ela cria. No entanto, se você quiser que EMR a Amazon seja lançada com um determinado grupo de segurança, você deve marcá-lo. Para recursos que não são criados pela AmazonEMR, você deve adicionar tags a esses recursos. Por exemplo, você deve marcar EC2 sub-redes, grupos de EC2 segurança da Amazon (se não forem criados pela AmazonEMR) e VPCs (se quiser que EMR a Amazon crie grupos de segurança). Para iniciar clusters com políticas gerenciadas v2VPCs, você deve marcá-los VPCs com a tag de usuário predefinida. Consulte [Considerações sobre o uso do EMR console da Amazon para lançar clusters com políticas gerenciadas v2](#).

Marcação propagada especificada pelo usuário

A Amazon EMR marca os recursos que ela cria usando as EMR tags da Amazon que você especifica ao criar um cluster. A Amazon EMR aplica tags aos recursos que cria durante a vida útil do cluster.

A Amazon EMR propaga tags de usuário para os seguintes recursos:

- Sub-rede privada ENI (interfaces de rede elástica de acesso ao serviço)
- EC2Instâncias
- EBSVolumes
- EC2Modelo de lançamento

Grupos de segurança etiquetados automaticamente

A Amazon EMR marca os grupos de EC2 segurança que ela cria com a tag necessária para políticas gerenciadas v2 para a Amazon EMR`for-use-with-amazon-emr-managed-policies`, independentemente das tags que você especificar no comando create cluster. Para um grupo de segurança criado antes da introdução das políticas gerenciadas v2, a Amazon EMR não

marca automaticamente o grupo de segurança. Para usar políticas gerenciadas v2 com os grupos de segurança padrão que já existem na conta, você precisa etiquetar os grupos de segurança manualmente com `for-use-with-amazon-emr-managed-policies = true`.

Recursos de cluster etiquetados manualmente

Você deve marcar manualmente alguns recursos do cluster para que eles possam ser acessados pelas funções EMR padrão da Amazon.

- Você deve marcar manualmente grupos EC2 de segurança e EC2 sub-redes com a etiqueta de política EMR gerenciada da Amazon. `for-use-with-amazon-emr-managed-policies`
- Você deve marcar manualmente um VPC se quiser que EMR a Amazon crie grupos de segurança padrão. EMRtentará criar um grupo de segurança com a tag específica se o grupo de segurança padrão ainda não existir.

A Amazon marca EMR automaticamente os seguintes recursos:

- EMR-criados grupos EC2 de segurança

Você deve etiquetar manualmente os seguintes recursos:

- EC2Sub-rede
- EC2Grupos de segurança

Opcionalmente, você pode etiquetar manualmente os seguintes recursos:

- VPC- somente quando você quiser que EMR a Amazon crie grupos de segurança

Considerações sobre o uso do EMR console da Amazon para lançar clusters com políticas gerenciadas v2

Você pode provisionar clusters com políticas gerenciadas v2 usando o EMR console da Amazon. Aqui estão algumas considerações ao usar o console para iniciar EMR clusters da Amazon.

- Não é necessário passar a etiqueta predefinida. A Amazon adiciona EMR automaticamente a tag e a propaga para os componentes apropriados.
- Para componentes que precisam ser marcados manualmente, o EMR console antigo da Amazon tenta marcá-los automaticamente se você tiver as permissões necessárias para marcar recursos.

Se você não tiver as permissões para marcar recursos ou se quiser usar o console, peça ao administrador que marque esses recursos.

- Não é possível iniciar clusters com políticas gerenciadas v2 sem que todos os pré-requisitos sejam atendidos.
- O antigo EMR console da Amazon mostra quais recursos (VPC/Sub-redes) precisam ser marcados.

IAMpolítica gerenciada para acesso total (política padrão gerenciada v2)

As políticas gerenciadas EMR padrão com escopo v2 concedem privilégios de acesso específicos aos usuários. Eles exigem uma tag de EMR recurso predefinida da Amazon e chaves de `iam:PassRole` condição para recursos usados pela AmazonEMR, como o `Subnet` e `SecurityGroup` que você usa para iniciar seu cluster.

Para conceder as ações necessárias com escopo definido para a AmazonEMR, anexe a política `AmazonEMRFullAccessPolicy_v2` gerenciada. Essa política gerenciada padrão atualizada substitui a política gerenciada `AmazonElasticMapReduceFullAccess`.

`AmazonEMRFullAccessPolicy_v2`depende do acesso limitado aos recursos que a Amazon EMR provisiona ou usa. Ao usar essa política, é necessário passar a etiqueta de usuário `for-use-with-amazon-emr-managed-policies = true` ao provisionar o cluster. A Amazon EMR propagará automaticamente a tag. Além disso, talvez seja necessário adicionar manualmente uma tag de usuário a tipos específicos de recursos, como grupos EC2 de segurança que não foram criados pela AmazonEMR. Para obter mais informações, consulte Etiquetar recursos para usar políticas gerenciadas.

A política `AmazonEMRFullAccessPolicy_v2` protege os recursos fazendo o seguinte:

- Requer que os recursos sejam marcados com a tag predefinida de políticas EMR gerenciadas da Amazon `for-use-with-amazon-emr-managed-policies` para criação de clusters e EMR acesso à Amazon.
- Restringe a ação `iam:PassRole` a perfis padrão específicos e acesso `iam:PassedToService` a serviços específicos.
- Não fornece mais acesso à AmazonEC2, Amazon S3 e outros serviços por padrão.

Veja a seguir o conteúdo dessa política.

> **Note**
>
> Você também pode usar o link do console AmazonEMRFullAccessPolicy_v2 para visualizar a política.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "RunJobFlowExplicitlyWithEMRManagedTag",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:RunJobFlow"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
                }
            }
        },
        {
            "Sid": "ElasticMapReduceActions",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:AddInstanceFleet",
                "elasticmapreduce:AddInstanceGroups",
                "elasticmapreduce:AddJobFlowSteps",
                "elasticmapreduce:AddTags",
                "elasticmapreduce:CancelSteps",
                "elasticmapreduce:CreateEditor",
                "elasticmapreduce:CreateSecurityConfiguration",
                "elasticmapreduce:DeleteEditor",
                "elasticmapreduce:DeleteSecurityConfiguration",
                "elasticmapreduce:DescribeCluster",
                "elasticmapreduce:DescribeEditor",
                "elasticmapreduce:DescribeJobFlows",
                "elasticmapreduce:DescribeSecurityConfiguration",
                "elasticmapreduce:DescribeStep",
                "elasticmapreduce:DescribeReleaseLabel",
                "elasticmapreduce:GetBlockPublicAccessConfiguration",
```

```
                "elasticmapreduce:GetManagedScalingPolicy",
                "elasticmapreduce:GetAutoTerminationPolicy",
                "elasticmapreduce:ListBootstrapActions",
                "elasticmapreduce:ListClusters",
                "elasticmapreduce:ListEditors",
                "elasticmapreduce:ListInstanceFleets",
                "elasticmapreduce:ListInstanceGroups",
                "elasticmapreduce:ListInstances",
                "elasticmapreduce:ListSecurityConfigurations",
                "elasticmapreduce:ListSteps",
                "elasticmapreduce:ListSupportedInstanceTypes",
                "elasticmapreduce:ModifyCluster",
                "elasticmapreduce:ModifyInstanceFleet",
                "elasticmapreduce:ModifyInstanceGroups",
                "elasticmapreduce:OpenEditorInConsole",
                "elasticmapreduce:PutAutoScalingPolicy",
                "elasticmapreduce:PutBlockPublicAccessConfiguration",
                "elasticmapreduce:PutManagedScalingPolicy",
                "elasticmapreduce:RemoveAutoScalingPolicy",
                "elasticmapreduce:RemoveManagedScalingPolicy",
                "elasticmapreduce:RemoveTags",
                "elasticmapreduce:SetTerminationProtection",
                "elasticmapreduce:StartEditor",
                "elasticmapreduce:StopEditor",
                "elasticmapreduce:TerminateJobFlows",
                "elasticmapreduce:ViewEventsFromAllClustersInConsole"
            ],
            "Resource": "*"
        },
        {
            "Sid": "ViewMetricsInEMRConsole",
            "Effect": "Allow",
            "Action": [
                "cloudwatch:GetMetricStatistics"
            ],
            "Resource": "*"
        },
        {
            "Sid": "PassRoleForElasticMapReduce",
            "Effect": "Allow",
            "Action": "iam:PassRole",
            "Resource": [
                "arn:aws:iam::*:role/EMR_DefaultRole",
                "arn:aws:iam::*:role/EMR_DefaultRole_V2"
```

```
            ],
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": "elasticmapreduce.amazonaws.com*"
                }
            }
        },
        {
            "Sid": "PassRoleForEC2",
            "Effect": "Allow",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::*:role/EMR_EC2_DefaultRole",
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": "ec2.amazonaws.com*"
                }
            }
        },
        {
            "Sid": "PassRoleForAutoScaling",
            "Effect": "Allow",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::*:role/EMR_AutoScaling_DefaultRole",
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": "application-autoscaling.amazonaws.com*"
                }
            }
        },
        {
            "Sid": "ElasticMapReduceServiceLinkedRole",
            "Effect": "Allow",
            "Action": "iam:CreateServiceLinkedRole",
            "Resource": "arn:aws:iam::*:role/aws-service-role/
elasticmapreduce.amazonaws.com*/AWSServiceRoleForEMRCleanup*",
            "Condition": {
                "StringEquals": {
                    "iam:AWSServiceName": [
                        "elasticmapreduce.amazonaws.com",
                        "elasticmapreduce.amazonaws.com.cn"
                    ]
                }
            }
        },
```

```
        {
            "Sid": "ConsoleUIActions",
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeAccountAttributes",
                "ec2:DescribeAvailabilityZones",
                "ec2:DescribeImages",
                "ec2:DescribeKeyPairs",
                "ec2:DescribeNatGateways",
                "ec2:DescribeRouteTables",
                "ec2:DescribeSecurityGroups",
                "ec2:DescribeSubnets",
                "ec2:DescribeVpcs",
                "ec2:DescribeVpcEndpoints",
                "s3:ListAllMyBuckets",
                "iam:ListRoles"
            ],
            "Resource": "*"
        }
    ]
}
```

IAMpolítica gerenciada para acesso total (a caminho da suspensão de uso)

As políticas gerenciadas `AmazonElasticMapReduceFullAccess` e `AmazonEMRFullAccessPolicy_v2` AWS Identity and Access Management (IAM) concedem todas as ações necessárias para a Amazon EMR e outros serviços.

> **Important**
>
> A política `AmazonElasticMapReduceFullAccess` gerenciada está prestes a ser descontinuada e não é mais recomendada para uso com a Amazon. EMR Em seu lugar, use `AmazonEMRFullAccessPolicy_v2`. Quando o IAM serviço eventualmente descontinuar a política v1, você não poderá vinculá-la a uma função. No entanto, você pode anexar um perfil já existente a um cluster mesmo que esse perfil use a política defasada.

As políticas EMR gerenciadas padrão de permissões completas da Amazon incorporam configurações `iam:PassRole` de segurança, incluindo as seguintes:

- `iam:PassRole`permissões somente para EMR funções padrão específicas da Amazon.

- `iam:PassedToService`condições que permitem que você use a política somente com AWS serviços específicos, como `elasticmapreduce.amazonaws.com` `ec2.amazonaws.com` e.

Você pode ver a JSON versão das políticas A mazonEMRFull AccessPolicy _v2 e A mazonEMRServicePolicy_v2 no console. IAM É recomendável criar novos clusters com políticas gerenciadas v2.

Você pode ver o conteúdo da política v1 obsoleta no endereço. AWS Management Console `AmazonElasticMapReduceFullAccess` A `ec2:TerminateInstances` ação na política concede permissão a um usuário ou função para encerrar qualquer uma das EC2 instâncias da Amazon associadas à IAM conta. Isso inclui instâncias que não fazem parte de um EMR cluster.

IAMpolítica gerenciada para acesso somente leitura (política padrão gerenciada v2)

Para conceder privilégios de somente leitura à AmazonEMR, anexe a política gerenciada A mazonEMRRead OnlyAccessPolicy _v2. Essa política gerenciada padrão substitui a política gerenciada `AmazonElasticMapReduceReadOnlyAccess`.
O conteúdo dessa instrução de política é mostrado no trecho a seguir. Em comparação com a política `AmazonElasticMapReduceReadOnlyAccess`, a política `AmazonEMRReadOnlyAccessPolicy_v2` não usa caracteres curinga para o elemento `elasticmapreduce`. Em vez disso, a política v2 padrão define o escopo das ações `elasticmapreduce` que podem ser permitidas.

> **Note**
>
> Você também pode usar o AWS Management Console link `AmazonEMRReadOnlyAccessPolicy_v2`para ver a política.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "ElasticMapReduceActions",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:DescribeCluster",
                "elasticmapreduce:DescribeEditor",
                "elasticmapreduce:DescribeJobFlows",
```

```
                "elasticmapreduce:DescribeSecurityConfiguration",
                "elasticmapreduce:DescribeStep",
                "elasticmapreduce:DescribeReleaseLabel",
                "elasticmapreduce:GetBlockPublicAccessConfiguration",
                "elasticmapreduce:GetManagedScalingPolicy",
                "elasticmapreduce:GetAutoTerminationPolicy",
                "elasticmapreduce:ListBootstrapActions",
                "elasticmapreduce:ListClusters",
                "elasticmapreduce:ListEditors",
                "elasticmapreduce:ListInstanceFleets",
                "elasticmapreduce:ListInstanceGroups",
                "elasticmapreduce:ListInstances",
                "elasticmapreduce:ListSecurityConfigurations",
                "elasticmapreduce:ListSteps",
                "elasticmapreduce:ListSupportedInstanceTypes",
                "elasticmapreduce:ViewEventsFromAllClustersInConsole"
            ],
            "Resource": "*"
        },
        {
            "Sid": "ViewMetricsInEMRConsole",
            "Effect": "Allow",
            "Action": [
                "cloudwatch:GetMetricStatistics"
            ],
            "Resource": "*"
        }
    ]
}
```

IAMpolítica gerenciada para acesso somente leitura (a caminho da suspensão de uso)

A política gerenciada `AmazonElasticMapReduceReadOnlyAccess` está prestes a ser defasada. Você não pode anexar essa política ao iniciar novos clusters. `AmazonElasticMapReduceReadOnlyAccess`foi substituída [AmazonEMRReadOnlyAccessPolicy_v2](#)pela política gerenciada EMR padrão da Amazon. O conteúdo dessa instrução de política é mostrado no trecho a seguir. Caracteres curinga para o elemento `elasticmapreduce` especificam que somente as ações que comecem com as strings especificadas serão permitidas. Lembre-se de que, como essa política não nega explicitamente as ações, uma declaração de política diferente ainda pode ser usada para conceder acesso a ações específicas.

> **Note**
>
> Você também pode usar o AWS Management Console para visualizar a política.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "elasticmapreduce:Describe*",
        "elasticmapreduce:List*",
        "elasticmapreduce:ViewEventsFromAllClustersInConsole",
        "s3:GetObject",
        "s3:ListAllMyBuckets",
        "s3:ListBucket",
        "sdb:Select",
        "cloudwatch:GetMetricStatistics"
      ],
      "Resource": "*"
    }
  ]
}
```

AWS política gerenciada: EMRDescribeClusterPolicyForEMRWAL

Você não pode se vincular `EMRDescribeClusterPolicyForEMRWAL` às suas IAM entidades. Essa política está vinculada a uma função vinculada ao serviço que permite EMR à Amazon realizar ações em seu nome. Para obter mais informações sobre essa função vinculada ao serviço, consulte. Usando funções vinculadas ao serviço para registro antecipado

Essa política concede permissões somente de leitura que permitem que o WAL serviço da Amazon EMR encontre e retorne o status de um cluster. Para obter mais informações sobre a Amazon EMRWAL, consulte Write-ahead logs (WAL) para a Amazon. EMR

Detalhes de permissão

Esta política inclui as seguintes permissões:

- `emr`— Permite que os diretores descrevam o status do cluster da AmazonEMR. Isso é necessário para que a Amazon EMR possa confirmar quando um cluster foi encerrado e, depois de trinta dias, limpar todos WAL os registros deixados pelo cluster.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "elasticmapreduce:DescribeCluster"
      ],
      "Resource": "*"
    }
  ]
}
```

AWS políticas gerenciadas para a Amazon EMR

Uma política AWS gerenciada é uma política autônoma criada e administrada por AWS. AWS as políticas gerenciadas são projetadas para fornecer permissões para muitos casos de uso comuns, para que você possa começar a atribuir permissões a usuários, grupos e funções.

Lembre-se de que as políticas AWS gerenciadas podem não conceder permissões de privilégio mínimo para seus casos de uso específicos porque elas estão disponíveis para uso de todos os AWS clientes. Recomendamos que você reduza ainda mais as permissões definindo políticas gerenciadas pelo cliente da específicas para seus casos de uso.

Você não pode alterar as permissões definidas nas políticas AWS gerenciadas. Se AWS atualizar as permissões definidas em uma política AWS gerenciada, a atualização afetará todas as identidades principais (usuários, grupos e funções) às quais a política está anexada. AWS é mais provável que atualize uma política AWS gerenciada quando uma nova Serviço da AWS é lançada ou novas API operações são disponibilizadas para os serviços existentes.

Para obter mais informações, consulte políticas AWS gerenciadas no Guia IAM do usuário.

EMRAtualizações da Amazon para políticas AWS gerenciadas

Veja detalhes sobre as atualizações das políticas AWS gerenciadas da Amazon EMR desde que esse serviço começou a monitorar essas mudanças.

Alteração	Descrição	Data
EMRDescribeClusterPolicyForEMRWAL – Nova política	Foi adicionada uma nova política para que a Amazon EMR possa determinar o status do cluster para WAL limpeza trinta dias após o encerramento do cluster.	10 de agosto de 2023
AmazonEMRFullAccessPolicy_v2 e **AmazonEMRReadOnlyAccessPolicy_v2**: atualizar para uma política já existente	Adicionadas `elasticmapreduce:DescribeReleaseLabel` e `elasticmapreduce:GetAutoTerminationPolicy`.	21 de abril de 2022
AmazonEMRFullAccessPolicy_v2: atualizar para uma política existente	Adicionou-se `ec2:DescribeImages` para Usando um personalizado AMI.	15 de fevereiro de 2022
Políticas EMR gerenciadas pela Amazon	Atualizado para esclarecer o uso de etiquetas de usuário predefinidas. Foi adicionada uma seção sobre como usar o AWS console para iniciar clusters com políticas gerenciadas v2.	29 de setembro de 2021
AmazonEMRFullAccessPolicy_v2: atualizar para uma política existente	Foram alteradas as ações `PassRoleForAutoScaling` e `PassRoleForEC2` para usar o operador de condição `StringLike` para corresponder a `"iam:PassedToService":"application-autoscaling.amazonaws.com*"`	20 de maio de 2021

Alteração	Descrição	Data
	e `"iam:PassedToService":"ec2.amazonaws.com*"` , respectivamente.	
`AmazonEMRFullAccessPolicy_v2`: atualizar para uma política existente	A ação inválida `s3:ListBuckets` foi removida e substituída pela ação `s3:ListAllMyBuckets` . Criação atualizada da função vinculada ao serviço (SLR) para ser explicitamente reduzida à única que a SLR Amazon EMR tem com Princípios de Serviço explícitos. Os SLRs que podem ser criados são exatamente os mesmos de antes dessa alteração.	23 de março de 2021

Alteração	Descrição	Data
AmazonEMRFullAccessPolicy_v2 – Nova política	A Amazon EMR adicionou novas permissões para definir o escopo do acesso aos recursos e adicionar um pré-requisito de que os usuários devem adicionar uma tag de usuário predefinida aos recursos antes de poderem usar as políticas gerenciadas da AmazonEMR. A ação `iam:PassRole` requer uma condição `iam:PassedToService` definida para o serviço especificado. O acesso à AmazonEC2, Amazon S3 e outros serviços não é permitido por padrão.	11 de março de 2021
AmazonEMRServicePolicy_v2 – Nova política	Adiciona o pré-requisito de que os usuários devem adicionar etiquetas de usuário aos recursos para poderem usar essa política.	11 de março de 2021
AmazonEMRReadOnlyAccessPolicy_v2 – Nova política	As permissões concedem somente ações específicas de leitura do elasticmapreduce. O acesso ao Amazon S3 é um acesso não concedido por padrão.	11 de março de 2021

Alteração	Descrição	Data
A Amazon EMR começou a monitorar as mudanças	A Amazon EMR começou a monitorar as mudanças em suas políticas AWS gerenciadas.	11 de março de 2021

IAMpolíticas para acesso baseado em tags a clusters e notebooks EMR

Você pode usar condições em sua política baseada em identidade para controlar o acesso a clusters e EMR notebooks com base em tags.

Para obter mais informações sobre como adicionar tags a clusters, consulte Como marcar EMR clusters.

Os exemplos a seguir demonstram diferentes cenários e formas de usar operadores de condição com as chaves de EMR condição da Amazon. Essas declarações IAM de política são destinadas apenas para fins de demonstração e não devem ser usadas em ambientes de produção. Há várias maneiras de combinar declarações de políticas para conceder e negar permissões de acordo com seus requisitos. Para obter mais informações sobre IAM políticas de planejamento e teste, consulte o Guia IAM do usuário.

> **Important**
>
> Recusar, explicitamente, permissões para ações de uso de tags é uma consideração importante. Isso evita que os usuários façam a marcação de um recurso e, assim, concedam a si mesmos permissões que você não pretendia conceder. Se você não negar as ações de marcação de um recurso, o usuário poderá modificar as etiquetas e contornar a intenção das políticas baseadas em etiquetas.

Exemplo de instruções de políticas baseadas em identidade para clusters

Os exemplos a seguir demonstram políticas de permissões baseadas em identidade que são usadas para controlar as ações permitidas com EMR clusters.

> **Important**
>
> A `ModifyInstanceGroup` ação na Amazon EMR não exige que você especifique um ID de cluster. Por isso, negar essa ação com base em etiquetas de cluster requer mais atenção. Para obter mais informações, consulte Negando a ação ModifyInstanceGroup .

Tópicos

- Permitir ações somente em clusters com determinados valores de etiqueta
- Exigir a marcação do cluster quando um cluster é criado
- Permitir ações em clusters com uma etiqueta específica, independentemente do valor da etiqueta

Permitir ações somente em clusters com determinados valores de etiqueta

Os exemplos a seguir demonstram uma política que permite ao usuário executar ações com base na etiqueta de cluster *`department`* com o valor *`dev`* e também permite que o usuário atribua etiquetas a clusters com a mesma etiqueta. O exemplo final da política demonstra como negar privilégios para marcar EMR clusters com qualquer coisa, menos a mesma tag.

No exemplo de política a seguir, o operador de condição `StringEquals` tenta corresponder *`dev`* com o valor da tag *`department`*. Se a tag *`department`* ainda não tiver sido adicionada ao cluster ou não contiver o valor *`dev`*, a política não se aplicará e as ações não serão permitidas por essa política. Se nenhuma outra declaração de política permitir as ações, o usuário poderá somente trabalhar com clusters que tenham essa tag com esse valor.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Stmt12345678901234",
      "Effect": "Allow",
      "Action": [
        "elasticmapreduce:DescribeCluster",
        "elasticmapreduce:ListSteps",
        "elasticmapreduce:TerminateJobFlows",
        "elasticmapreduce:SetTerminationProtection",
        "elasticmapreduce:ListInstances",
        "elasticmapreduce:ListInstanceGroups",
        "elasticmapreduce:ListBootstrapActions",
```

```
        "elasticmapreduce:DescribeStep"
      ],
      "Resource": [
        "*"
      ],
      "Condition": {
        "StringEquals": {
          "elasticmapreduce:ResourceTag/department": "dev"
        }
      }
    }
  ]
}
```

Você também pode especificar vários valores de tag usando um operador de condição. Por exemplo, para permitir todas as ações em clusters em que a tag *department* contenha o valor *dev* ou *test*, você poderia substituir o bloco condicional no exemplo anterior com o seguinte.

```
            "Condition": {
              "StringEquals": {
                "elasticmapreduce:ResourceTag/department":["dev", "test"]
              }
            }
```

Exigir a marcação do cluster quando um cluster é criado

Como no exemplo anterior, o exemplo de política a seguir procura a mesma etiqueta correspondente: o valor *dev* para a etiqueta *department*. Mas neste exemplo, a chave de condição `RequestTag` especifica que a política se aplica durante a criação da etiqueta. Portanto, é necessário criar um cluster com uma etiqueta que corresponda ao valor especificado.

Para criar um cluster com uma etiqueta, também é necessário ter permissão para a ação `elasticmapredue:AddTags`. Para essa declaração, a chave de `elasticmapreduce:ResourceTag` condição garante que IAM somente conceda acesso aos recursos da tag com o valor *dev* na *department* tag. O elemento `Resource` é usado para limitar essa permissão aos recursos do cluster.

Para os `PassRole` recursos, você deve fornecer o ID ou alias da AWS conta, o nome da função de serviço na `PassRoleForEMR` declaração e o nome do perfil da instância na `PassRoleForEC2`

declaração. Para obter mais informações sobre o IAM ARN formato, consulte IAMARNso Guia IAM do usuário.

Para obter mais informações sobre a correspondência de valores de chave de tag, consulte o `aws:RequestTag/tag-key`Guia do IAMusuário.

```
{
    "Version": "2012-10-17",
    "Statement": [
      {
        "Sid": "RunJobFlowExplicitlyWithTag",
        "Effect": "Allow",
        "Action": [
          "elasticmapreduce:RunJobFlow"
        ],
        "Resource": "*",
        "Condition": {
          "StringEquals": {
            "aws:RequestTag/department": "dev"
          }
        }
      },
      {
        "Sid": "AddTagsForDevClusters",
        "Effect": "Allow",
        "Action": "elasticmapreduce:AddTags",
        "Resource": "arn:aws:elasticmapreduce:*:*:cluster/*",
        "Condition": {
          "StringEquals": {
            "elasticmapreduce:ResourceTag/department": "dev"
          }
        }
      },
      {
        "Sid": "PassRoleForEMR",
        "Effect": "Allow",
        "Action": "iam:PassRole",
        "Resource": "arn:aws:iam::AccountId:role/Role-Name-With-Path",
        "Condition": {
          "StringLike": {
            "iam:PassedToService": "elasticmapreduce.amazonaws.com*"
          }
        }
```

```
      },
      {
        "Sid": "PassRoleForEC2",
        "Effect": "Allow",
        "Action": "iam:PassRole",
        "Resource": "arn:aws:iam::AccountId:role/Role-Name-With-Path",
        "Condition": {
          "StringLike": {
            "iam:PassedToService": "ec2.amazonaws.com*"
          }
        }
      }
    ]
}
```

Permitir ações em clusters com uma etiqueta específica, independentemente do valor da etiqueta

Você também pode permitir ações somente em clusters que tenham uma determinada tag, independentemente do valor da tag. Para fazer isso, você pode usar o operador `Null`. Para obter mais informações, consulte Operador de condição para verificar a existência de chaves de condição no Guia IAM do usuário. Por exemplo, para permitir ações somente em EMR clusters que tenham a *`department`* tag, independentemente do valor que ela contenha, você pode substituir os blocos de condição no exemplo anterior pelo seguinte. O `Null` operador procura a presença da tag *`department`* em um EMR cluster. Se a tag existir, a instrução `Null` será avaliada como falsa, correspondendo à condição especificada nesta declaração de política e as ações apropriadas serão permitidas.

```
"Condition": {
  "Null": {
    "elasticmapreduce:ResourceTag/department":"false"
  }
}
```

A declaração de política a seguir permite que um usuário crie um EMR cluster somente se o cluster tiver uma *`department`* tag, que pode conter qualquer valor. Para o `PassRole` recurso, você precisa fornecer o ID ou alias da AWS conta e o nome da função de serviço. Para obter mais informações sobre o IAM ARN formato, consulte IAMARNso Guia IAM do usuário.

Para obter mais informações especificando o operador de condição nulo (“falso”), consulte Operador de condição para verificar a existência de chaves de condição no Guia do IAMusuário.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "CreateClusterTagNullCondition",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:RunJobFlow"
            ],
            "Resource": [
                "*"
            ],
            "Condition": {
                "Null": {
                    "aws:RequestTag/department": "false"
                }
            }
        },
        {
            "Sid": "AddTagsNullCondition",
            "Effect": "Allow",
            "Action": "elasticmapreduce:AddTags",
            "Resource": "arn:aws:elasticmapreduce:*:*:cluster/*",
            "Condition": {
                "Null": {
                    "elasticmapreduce:ResourceTag/department": "false"
                }
            }
        },
        {
            "Sid": "PassRoleForElasticMapReduce",
            "Effect": "Allow",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::AccountId:role/Role-Name-With-Path",
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": "elasticmapreduce.amazonaws.com*"
                }
            }
        },
```

```
        {
            "Sid": "PassRoleForEC2",
            "Effect": "Allow",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::AccountId:role/Role-Name-With-Path",
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": "ec2.amazonaws.com*"
                }
            }
        }
    ]
}
```

Exemplo de declarações de política baseadas em identidade para notebooks EMR

Os exemplos IAM de declarações de política nesta seção demonstram cenários comuns de uso de chaves para limitar as ações permitidas usando EMR Notebooks. Desde que nenhuma outra política associada à entidade principal (usuário) permita as ações, as chaves de contexto de condição limitam as ações permitidas conforme indicado.

Example — Permitir acesso somente aos EMR Notebooks que um usuário cria com base na marcação

A instrução de política de exemplo a seguir, quando anexada a um perfil ou usuário, permite que o usuário trabalhe apenas com cadernos criados por ele. Esta declaração de política usa a tag padrão aplicada quando um bloco de anotações é criado.

No exemplo, o operador de condição `StringEquals` tenta combinar uma variável que representa o ID do usuário atual (`{aws:userId}`) com o valor de etiqueta `creatorUserID`. Se a tag `creatorUserID` ainda não tiver sido adicionada ao bloco de anotações ou não contiver o valor do ID do usuário atual, a política não se aplicará e as ações não serão permitidas por essa política. Se nenhuma outra declaração de política permitir as ações, o usuário só poderá trabalhar com blocos de anotações que tenham essa tag com esse valor.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:DescribeEditor",
                "elasticmapreduce:StartEditor",
```

```
                "elasticmapreduce:StopEditor",
                "elasticmapreduce:DeleteEditor",
                "elasticmapreduce:OpenEditorInConsole"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/creatorUserId": "${aws:userId}"
                }
            }
        }
    ]
}
```

Example –Exigir marcação de caderno quando um caderno é criado

Neste exemplo, a chave de contexto `RequestTag` é usada. A ação `CreateEditor` será permitida somente se o usuário não alterar nem excluir a tag `creatorUserID` é que é adicionada por padrão. A variável $ {aws:userId} especifica a ID de usuário do usuário atualmente ativo, que é o valor padrão da tag.

A declaração de política pode ser usada para ajudar a garantir que os usuários não removam a tag `createUserId` tag nem alterem seu valor.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:CreateEditor"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:RequestTag/creatorUserId": "${aws:userid}"
                }
            }
        }
    ]
}
```

Este exemplo requer que o usuário crie o cluster com uma tag com a string de chave `dept` e um valor definido como um dos seguintes: `datascience`, `analytics`, `operations`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:CreateEditor"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:RequestTag/dept": [
                        "datascience",
                        "analytics",
                        "operations"
                    ]
                }
            }
        }
    ]
}
```

Example –Limitar a criação do caderno para clusters marcados e exigir etiquetas de caderno

Este exemplo permite a criação do bloco de anotações somente se o bloco de anotações for criado com uma tag que tenha a string de chave `owner` definida como um dos valores especificados. Além disso, o bloco de anotações poderá ser criado somente se o cluster tiver uma tag com a string de chave `department` definida como um dos valores especificados.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:CreateEditor"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
```

```
                "StringEquals": {
                    "elasticmapreduce:RequestTag/owner": [
                        "owner1",
                        "owner2",
                        "owner3"
                    ],
                    "elasticmapreduce:ResourceTag/department": [
                        "dep1",
                        "dep3"
                    ]
                }
            }
        }
    ]
}
```

Example –Limitar a capacidade de iniciar um caderno com base em etiquetas

Este exemplo limita a capacidade de iniciar blocos de anotações àqueles que tenham uma tag com a string de chave `owner` definida como um dos valores especificados. Como o elemento `Resource` é usado para especificar apenas o `editor`, a condição não se aplica ao cluster e não precisa ser marcado.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:StartEditor"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:editor/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/owner": [
                        "owner1",
                        "owner2"
                    ]
                }
            }
        }
    ]
```

```
}
```

Este exemplo é semelhante ao exposto acima. No entanto, o limite se aplica apenas a clusters com tags, e não a blocos de anotações.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:StartEditor"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:cluster/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/department": [
                        "dep1",
                        "dep3"
                    ]
                }
            }
        }
    ]
}
```

Este exemplo usa um conjunto diferente de tags de cluster e bloco de anotações. Ele permite que um bloco de anotações seja iniciado somente se:

- O bloco de anotações tiver uma tag com a string de chave `owner` definida como qualquer um dos valores especificados

  —e—
- O cluster tiver uma tag com a string de chave `department` definida como qualquer um dos valores especificados

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
```

```
                "elasticmapreduce:StartEditor"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:editor/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/owner": [
                        "user1",
                        "user2"
                    ]
                }
            }
        },
        {
            "Action": [
                "elasticmapreduce:StartEditor"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:cluster/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/department": [
                        "datascience",
                        "analytics"
                    ]
                }
            }
        }
    ]
}
```

Example –Limitar a capacidade de abrir o editor de caderno com base em etiquetas

Este exemplo permite que o editor de blocos de anotações seja aberto somente se:

- O bloco de anotações tiver uma tag com a string de chave `owner` definida como qualquer um dos valores especificados.

  —e—

- O cluster tiver uma tag com a string de chave `department` definida como qualquer um dos valores especificados.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:OpenEditorInConsole"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:editor/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/owner": [
                        "user1",
                        "user2"
                    ]
                }
            }
        },
        {
            "Action": [
                "elasticmapreduce:OpenEditorInConsole"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:cluster/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/department": [
                        "datascience",
                        "analytics"
                    ]
                }
            }
        }
    ]
}
```

Negando a ação ModifyInstanceGroup

A ModifyInstanceGroupsação na Amazon EMR não exige que você forneça um ID de cluster com a ação. Em vez disso, você pode especificar apenas um ID de grupo de instâncias. Por isso, uma política de negação aparentemente simples para essa ação com base no ID do cluster ou em uma etiqueta do cluster pode não ter o efeito pretendido. Considere a seguinte política de exemplo.

```
{
 "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Allow",
            "Resource": "*"
        },
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Deny",
            "Resource": "arn:aws:elasticmapreduce:us-east-1:123456789012:cluster/
j-12345ABCDEFG67"
        }
    ]

}
```

Se um usuário com essa política anexada realizar uma ação `ModifyInstanceGroup` e especificar somente o ID do grupo de instâncias, a política não se aplicará. Como a ação é permitida em todos os outros recursos, ela tem êxito.

Uma solução para esse problema é anexar uma declaração de política à identidade que usa um NotResourceelemento para negar qualquer `ModifyInstanceGroup` ação emitida sem um ID de cluster. O exemplo de política a seguir adiciona essa instrução de negação para que qualquer solicitação `ModifyInstanceGroups` falhe, a menos que um ID de cluster esteja especificado. Como uma identidade deve especificar um ID de cluster com a ação, as instruções de negação com base no ID do cluster são, portanto, efetivas.

```
{
 "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Allow",
            "Resource": "*"
```

```
        },
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Deny",
            "Resource": "arn:aws:elasticmapreduce:us-east-1:123456789012:cluster/
j-12345ABCDEFG67"
        },
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Deny",
            "NotResource": "arn:*:elasticmapreduce:*:*:cluster/*"
        }
    ]

}
```

Um problema semelhante ocorre quando você deseja negar a ação `ModifyInstanceGroups` com base no valor associado a uma etiqueta de cluster. A solução é semelhante. Além de uma instrução de negação que especifica o valor da etiqueta, é possível adicionar uma instrução de política que nega a ação `ModifyInstanceGroup` se a etiqueta especificada não estiver presente, qualquer que seja o valor.

O exemplo a seguir demonstra uma política que, quando anexada a uma identidade, nega à identidade a ação `ModifyInstanceGroups` de qualquer cluster com a etiqueta `department` definida como `dev`. Essa instrução só é efetiva por causa da instrução de negação que usa a condição `StringNotLike` para negar a ação, a menos que a etiqueta `department` esteja presente.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Allow",
            "Resource": "*"
        },
```

```
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Condition": {
                "StringEquals": {
                    "aws:ResourceTag/department": "dev"
                }
            },
            "Effect": "Deny",
            "Resource": "*"
        },
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Condition": {
                "StringNotLike": {
                    "aws:ResourceTag/department": "?*"
                }
            },
            "Effect": "Deny",
            "Resource": "*"
        }
    ],
}
```

Solução de problemas de EMR identidade e acesso da Amazon

Use as informações a seguir para ajudá-lo a diagnosticar e corrigir problemas comuns que você pode encontrar ao trabalhar com a Amazon EMR e. IAM

Tópicos

- Não estou autorizado a realizar uma ação na Amazon EMR
- Não estou autorizado a realizar iam: PassRole
- Quero permitir que pessoas fora da minha AWS conta acessem meus EMR recursos da Amazon

Não estou autorizado a realizar uma ação na Amazon EMR

Se isso AWS Management Console indicar que você não está autorizado a realizar uma ação, entre em contato com o administrador para obter ajuda. O administrador é a pessoa que forneceu o seu nome de usuário e senha.

O erro do exemplo a seguir ocorre quando o usuário `mateojackson` tenta usar o console para visualizar detalhes sobre um recurso do *`my-example-widget`* fictício, mas não tem as permissões fictícias do `EMR:`*`GetWidget`*.

```
User: arn:aws:iam::123456789012:user/mateojackson is not authorized to perform:
 EMR:GetWidget on resource: my-example-widget
```

Neste caso, Mateo pede ao administrador para atualizar suas políticas e permitir o acesso ao recurso *`my-example-widget`* usando a ação `EMR:`*`GetWidget`*.

Não estou autorizado a realizar iam: PassRole

Se você receber um erro informando que não está autorizado a realizar a `iam:PassRole` ação, suas políticas devem ser atualizadas para permitir que você passe uma função para a AmazonEMR.

Alguns Serviços da AWS permitem que você passe uma função existente para esse serviço em vez de criar uma nova função de serviço ou uma função vinculada ao serviço. Para fazer isso, é preciso ter permissões para passar o perfil para o serviço.

O exemplo de erro a seguir ocorre quando um IAM usuário chamado `marymajor` tenta usar o console para realizar uma ação na AmazonEMR. No entanto, a ação exige que o serviço tenha permissões concedidas por um perfil de serviço. Mary não tem permissões para passar o perfil para o serviço.

```
User: arn:aws:iam::123456789012:user/marymajor is not authorized to perform:
 iam:PassRole
```

Nesse caso, as políticas de Mary devem ser atualizadas para permitir que ela realize a ação `iam:PassRole`.

Se precisar de ajuda, entre em contato com seu AWS administrador. Seu administrador é a pessoa que forneceu suas credenciais de login.

Quero permitir que pessoas fora da minha AWS conta acessem meus EMR recursos da Amazon

Você pode criar um perfil que os usuários de outras contas ou pessoas fora da sua organização podem usar para acessar seus recursos. Você pode especificar quem é confiável para assumir o perfil. Para serviços que oferecem suporte a políticas baseadas em recursos ou listas de controle de acesso (ACLs), você pode usar essas políticas para conceder às pessoas acesso aos seus recursos.

Para saber mais, consulte:

- Para saber se a Amazon EMR oferece suporte a esses recursos, consulteComo a Amazon EMR trabalha com IAM.
- Para saber como fornecer acesso aos seus recursos em todos os Contas da AWS que você possui, consulte Fornecer acesso a um IAM usuário em outro Conta da AWS de sua propriedade no Guia do IAM usuário.
- Para saber como fornecer acesso aos seus recursos a terceiros Contas da AWS, consulte Fornecer Contas da AWS acesso a terceiros no Guia do IAM usuário.
- Para saber como fornecer acesso por meio da federação de identidades, consulte Fornecendo acesso a usuários autenticados externamente (federação de identidades) no Guia do IAMusuário.
- Para saber a diferença entre usar funções e políticas baseadas em recursos para acesso entre contas, consulte Acesso a recursos entre contas IAM no Guia do IAM usuário.

Usando as concessões de acesso do Amazon S3 com a Amazon EMR

Visão geral do S3 Access Grants para a Amazon EMR

Com as EMR versões 6.15.0 e superiores da Amazon, o Amazon S3 Access Grants fornece uma solução de controle de acesso escalável que você pode usar para aumentar o acesso aos dados do Amazon S3 da Amazon. EMR Se você tiver uma configuração de permissão complexa ou grande para os dados do S3, poderá usar a funcionalidade Access Grants para escalar as permissões de dados do S3 para usuários, perfis e aplicações no seu cluster.

Use o S3 Access Grants para aumentar o acesso aos dados do Amazon S3 além das permissões concedidas pela função de tempo de execução ou IAM das funções associadas às identidades com acesso ao seu cluster. EMR Para obter mais informações, consulte Gerenciar o acesso com o S3 Access Grants no Guia do usuário do Amazon S3.

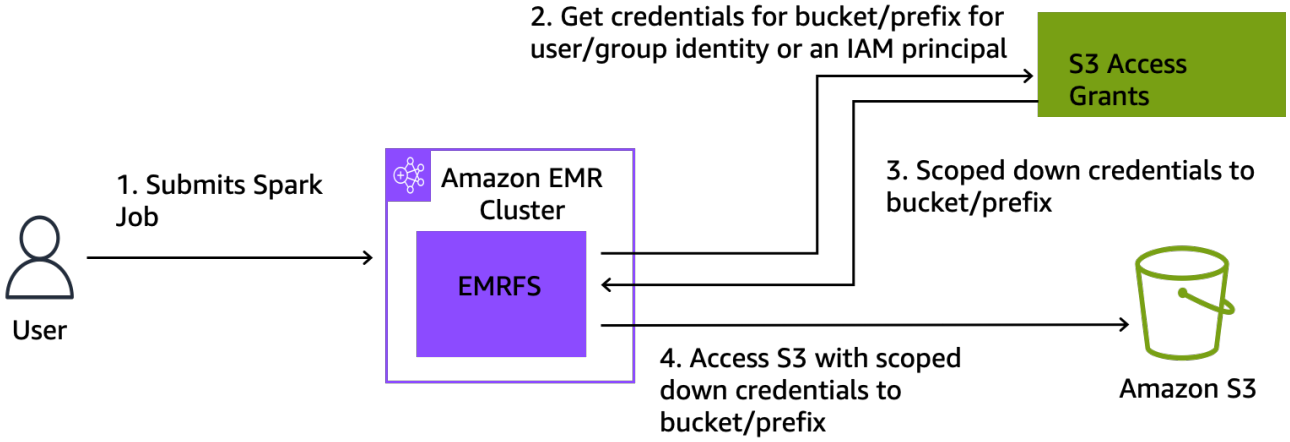
Para ver as etapas de uso do S3 Access Grants com outras EMR implantações da Amazon, consulte a seguinte documentação:

- Usando o S3 Access Grants com a Amazon EMR em EKS
- Usando o S3 Access Grants com o Amazon EMR Serverless

Como a Amazon EMR trabalha com o S3 Access Grants

As EMR versões 6.15.0 e superiores da Amazon fornecem uma integração nativa com o S3 Access Grants. Você pode ativar o S3 Access Grants na Amazon EMR e executar trabalhos do Spark. Quando um trabalho do Spark faz uma solicitação de dados do S3, o Amazon S3 fornece credenciais temporárias que têm como escopo o bucket, prefixo ou objeto específico.

A seguir está uma visão geral de alto nível de como a Amazon EMR obtém acesso aos dados protegidos pelo S3 Access Grants.



1. Um usuário envia um trabalho do Amazon EMR Spark que usa dados armazenados no Amazon S3.
2. A Amazon EMR solicita o S3 Access Grants para permitir o acesso ao bucket, prefixo ou objeto em nome desse usuário.
3. O Amazon S3 retorna credenciais temporárias na forma de um token AWS Security Token Service (STS) para o usuário. O escopo do token é acessar o bucket, prefixo ou objeto do S3.
4. A Amazon EMR usa o STS token para recuperar dados do S3.
5. A Amazon EMR recebe os dados do S3 e retorna os resultados ao usuário.

Considerações sobre o S3 Access Grants com a Amazon EMR

Observe os seguintes comportamentos e limitações ao usar o S3 Access Grants com a AmazonEMR.

Suporte a recursos

- O S3 Access Grants é compatível com as EMR versões 6.15.0 e superiores da Amazon.
- O Spark é o único mecanismo de consulta compatível quando você usa o S3 Access Grants com a Amazon. EMR
- Delta Lake e Hudi são os únicos formatos de tabela aberta compatíveis quando você usa o S3 Access Grants com a Amazon. EMR
- Os seguintes EMR recursos da Amazon não são compatíveis com o S3 Access Grants:
  - Tabelas Apache Iceberg
  - LDAPautenticação nativa
  - Autenticação nativa do Apache Ranger
  - AWS CLI solicitações para o Amazon S3 que usam funções IAM
  - Acesso ao S3 por meio do protocolo de código aberto do S3A
- A `fallbackToIAM` opção não é compatível com EMR clusters que usam propagação de identidade confiável com o IAM Identity Center.
- O S3 Access Grants with só AWS Lake Formationé compatível com EMR clusters da Amazon que são executados na AmazonEC2.

Considerações comportamentais

- A integração nativa do Apache Ranger com a Amazon EMR possui funcionalidade congruente com o S3 Access Grants como parte do plug-in S3 Apache Ranger. EMRFS Se você usa o Apache Ranger para controle de acesso refinado (FGAC), recomendamos que você use esse plug-in em vez do S3 Access Grants.
- EMRA Amazon fornece um cache de credenciais EMRFS para garantir que o usuário não precise fazer solicitações repetidas das mesmas credenciais em um trabalho do Spark. Portanto, a Amazon EMR sempre solicita o privilégio de nível padrão quando solicita credenciais. Para obter mais informações, consulte Solicitação de acesso aos dados do S3 no Guia do usuário do Amazon S3.

- No caso de um usuário realizar uma ação que o S3 Access Grants não suporta, a Amazon EMR está configurada para usar a IAM função que foi especificada para a execução do trabalho. Para obter mais informações, consulte Volte aos IAM papéis.

Lance um EMR cluster da Amazon com o S3 Access Grants

Esta seção descreve como iniciar um EMR cluster que é executado na Amazon EC2 e usa o S3 Access Grants para gerenciar o acesso aos dados no Amazon S3. Para ver as etapas de uso do S3 Access Grants com outras EMR implantações da Amazon, consulte a seguinte documentação:

- Usando o S3 Access Grants com a Amazon EMR em EKS
- Usando o S3 Access Grants com EMR o Serverless

Use as etapas a seguir para iniciar um EMR cluster executado na Amazon EC2 e usar o S3 Access Grants para gerenciar o acesso aos dados no Amazon S3.

1. Configure uma função de execução de trabalhos para seu EMR cluster. Inclua IAM as permissões necessárias para executar trabalhos do Spark `s3:GetDataAccess` e `s3:GetAccessGrantsInstanceForPrefix`:

```
{
    "Effect": "Allow",
    "Action": [
    "s3:GetDataAccess",
    "s3:GetAccessGrantsInstanceForPrefix"
    ],
    "Resource": [     //LIST ALL INSTANCE ARNS THAT THE ROLE IS ALLOWED TO QUERY
         "arn:aws_partition:s3:Region:account-id1:access-grants/default",
         "arn:aws_partition:s3:Region:account-id2:access-grants/default"
    ]
}
```

> **Note**
>
> Com a AmazonEMR, os S3 Access Grants aumentam as permissões definidas nas IAM funções. Se as IAM funções que você especificar para a execução do trabalho

contiverem permissões para acessar o S3 diretamente, os usuários poderão acessar mais dados do que apenas os dados definidos no S3 Access Grants.

2. Em seguida, use o AWS CLI para criar um cluster com o Amazon EMR 6.15 ou superior e a `emrfs-site` classificação para habilitar o S3 Access Grants, semelhante ao exemplo a seguir:

```
aws emr create-cluster
  --release-label emr-6.15.0 \
  --instance-count 3 \
  --instance-type m5.xlarge \
  --configurations '[{"Classification":"emrfs-site",
 "Properties":{"fs.s3.s3AccessGrants.enabled":"true",
 "fs.s3.s3AccessGrants.fallbackToIAM":"false"}}]'
```

Concessões de acesso ao S3 com AWS Lake Formation

Se você usa a Amazon EMR com a AWS Lake Formation integração, você pode usar o Amazon S3 Access Grants para acesso direto ou tabular aos dados no Amazon S3.

> **Note**
>
> O S3 Access Grants with só AWS Lake Formation é compatível com EMR clusters da Amazon que são executados na AmazonEC2.

Acesso direto

O acesso direto envolve todas as chamadas para acessar dados do S3 que não invocam o serviço For the API AWS Glue que a Lake Formation usa como metastore na EMR Amazon, por exemplo, para ligar para: `spark.read`

```
spark.read.csv("s3://...")
```

Quando você usa o S3 Access Grants AWS Lake Formation na AmazonEMR, todos os padrões de acesso direto passam pelo S3 Access Grants para obter credenciais temporárias do S3.

Acesso tabular

O acesso tabular ocorre quando o Lake Formation invoca o metastore API para acessar sua localização no S3, por exemplo, para consultar dados da tabela:

```
spark.sql("select * from test_tbl")
```

Quando você usa o S3 Access Grants AWS Lake Formation na AmazonEMR, todos os padrões de acesso tabulares passam pelo Lake Formation.

Volte aos IAM papéis

Se um usuário tentar realizar uma ação que o S3 Access Grants não suporta, a Amazon usa como EMR padrão a IAM função que foi especificada para execução do trabalho quando a `fallbackToIAM` configuração estiver. `true` Isso permite que os usuários efetuem o fallback do perfil de execução de trabalhos para fornecer credenciais de acesso ao S3 em cenários não cobertos pelo S3 Access Grants.

Com a opção `fallbackToIAM` habilitada, os usuários podem acessar os dados que o Access Grant permite. Se não houver um token do S3 Access Grants para os dados de destino, a Amazon EMR verifica a permissão em sua função de execução do trabalho.

> **Note**
>
> Recomendamos que você teste suas permissões de acesso com a configuração `fallbackToIAM` habilitada, mesmo que planeje desabilitar a opção para workloads de produção. Com os trabalhos do Spark, há outras maneiras pelas quais os usuários podem acessar todos os conjuntos de permissões com suas IAM credenciais. Quando ativadas em EMR clusters, as concessões do S3 dão às tarefas do Spark acesso às localizações do S3. Você deve garantir a proteção desses locais do S3 contra o acesso externo. EMRFS Por exemplo, você deve proteger as localizações do S3 contra o acesso de clientes do S3 usados em notebooks ou por aplicações sem o suporte do S3 Access Grants, como Hive ou Presto.

Autentique-se nos nós de EMR cluster da Amazon

SSHos clientes podem usar um par de EC2 chaves da Amazon para se autenticar em instâncias de cluster. Como alternativa, com as EMR versões 5.10.0 e superiores da Amazon, você pode configurar o Kerberos para autenticar usuários e SSH conexões com o nó primário. E com as EMR versões 5.12.0 e superiores da Amazon, você pode se autenticar com. LDAP

Tópicos

Use um par de EC2 chaves para SSH credenciais

Os nós de EMR cluster da Amazon são executados em EC2 instâncias da Amazon. Você pode se conectar aos nós do cluster da mesma forma que se conecta às EC2 instâncias da Amazon. Você pode usar EC2 a Amazon para criar um par de chaves ou importar um par de chaves. Ao criar um cluster, você pode especificar o par de EC2 chaves da Amazon que será usado para SSH conexões com todas as instâncias do cluster. Também é possível criar um cluster sem par de chaves. Isso geralmente é feito com clusters transitórios que são iniciados, executam etapas e são encerrados automaticamente.

O SSH cliente que você usa para se conectar ao cluster precisa usar o arquivo de chave privada associado a esse par de chaves. Esse é um arquivo.pem para SSH clientes que usam Linux, Unix e macOS. Você deve definir permissões para que apenas o proprietário da chave tenha permissão para acessar o arquivo. Esse é um arquivo.ppk para SSH clientes que usam o Windows, e o arquivo.ppk geralmente é criado a partir do arquivo.pem.

- Para obter mais informações sobre a criação de um par de EC2 chaves da Amazon, consulte os pares de EC2 chaves da Amazon no Guia EC2 do usuário da Amazon.
- Para obter instruções sobre como usar P uTTYgen para criar um arquivo.ppk a partir de um arquivo.pem, consulte Convertendo sua chave privada usando P no Guia do usuário uTTYgen da Amazon. EC2
- Para obter mais informações sobre como definir permissões de arquivos.pem e como se conectar ao nó primário de um EMR cluster usando métodos diferentes, inclusive `ssh` do Linux ou macOS, TTY do Pu do Windows ou AWS CLI de qualquer sistema operacional compatível, consulte. Conecte-se ao nó primário usando SSH

Use o Kerberos para autenticação com a Amazon EMR

As EMR versões 5.10.0 e superiores da Amazon oferecem suporte ao Kerberos. O Kerberos é um protocolo de autenticação de rede que usa criptografia segredo-chave para fornecer autenticação

forte, de maneira que senhas ou outras credenciais não sejam enviadas pela rede em um formato não criptografado.

No Kerberos, os serviços e os usuários que precisam se autenticar são conhecidos como entidades principais. As entidades principais existem em um realm Kerberos. Dentro do reino, um servidor Kerberos conhecido como centro de distribuição de chaves (KDC) fornece os meios para que os diretores se autenticem. KDCIsso é feito emitindo tíquetes para autenticação. O KDC mantém um banco de dados dos diretores em seu domínio, suas senhas e outras informações administrativas sobre cada diretor. A também KDC pode aceitar credenciais de autenticação de diretores em outros domínios, o que é conhecido como confiança entre reinos. Além disso, um EMR cluster pode usar um externo KDC para autenticar os principais.

Um cenário comum para estabelecer uma relação de confiança entre regiões ou usar uma interface externa KDC é autenticar usuários de um domínio do Active Directory. Isso permite que os usuários acessem um EMR cluster com sua conta de domínio quando usam SSH para se conectar a um cluster ou trabalhar com aplicativos de big data.

Quando você usa a autenticação Kerberos, a Amazon EMR configura o Kerberos para os aplicativos, componentes e subsistemas que ele instala no cluster para que eles sejam autenticados entre si.

> **Important**
>
> EMRA Amazon não oferece suporte AWS Directory Service for Microsoft Active Directory em uma relação de confiança entre regiões ou de forma externa. KDC

Antes de configurar o Kerberos usando a AmazonEMR, recomendamos que você se familiarize com os conceitos do Kerberos, os serviços que são executados em um KDC e as ferramentas para administrar os serviços Kerberos. Para obter mais informações, consulte a documentação do MIT Kerberos, publicada pelo consórcio Kerberos.

Tópicos

- Aplicações compatíveis
- Opções de arquitetura do Kerberos
- Configurando o Kerberos na Amazon EMR
- Usando SSH para se conectar a clusters Kerberizados
- Tutorial: Configurar um cluster dedicado KDC

- Tutorial: configurar uma relação de confiança entre realms com um controlador de domínio do Active Directory

Aplicações compatíveis

Em um EMR cluster, os principais do Kerberos são os serviços e subsistemas de aplicativos de big data que são executados em todos os nós do cluster. A Amazon EMR pode configurar os aplicativos e componentes listados abaixo para usar o Kerberos. Cada aplicativo tem uma entidade principal de usuário Kerberos associada.

A Amazon EMR não oferece suporte a relações de confiança entre reinos com. AWS Directory Service for Microsoft Active Directory

A Amazon configura EMR somente os recursos de autenticação Kerberos de código aberto para os aplicativos e componentes listados abaixo. Todos os outros aplicativos instalados não são Kerberizados, o que pode resultar em uma incapacidade de se comunicar com componentes Kerberizados e causar erros de aplicativo. Os aplicativos e os componentes não Kerberizados não têm autenticação ativada. Os aplicativos e componentes compatíveis podem variar para diferentes EMR versões da Amazon.

A interface de usuário do Livy é a única interface de usuário da Web hospedada no cluster que é kerberizada.

- Hadoop MapReduce
- Hbase
- HCatalog
- HDFS
- Hive
  - Não habilite o Hive com LDAP autenticação. Isso pode causar problemas de comunicação com o YARN Kerberized.
- Hue
  - A autenticação do usuário do Hue não é definida automaticamente e pode ser configurada usando a configuraçãoAPI.
  - O servidor do Hue é Kerberizado. O front-end (UI) do Hue não está configurado para autenticação. LDAPa autenticação pode ser configurada para a interface do usuário do Hue.
- Livy

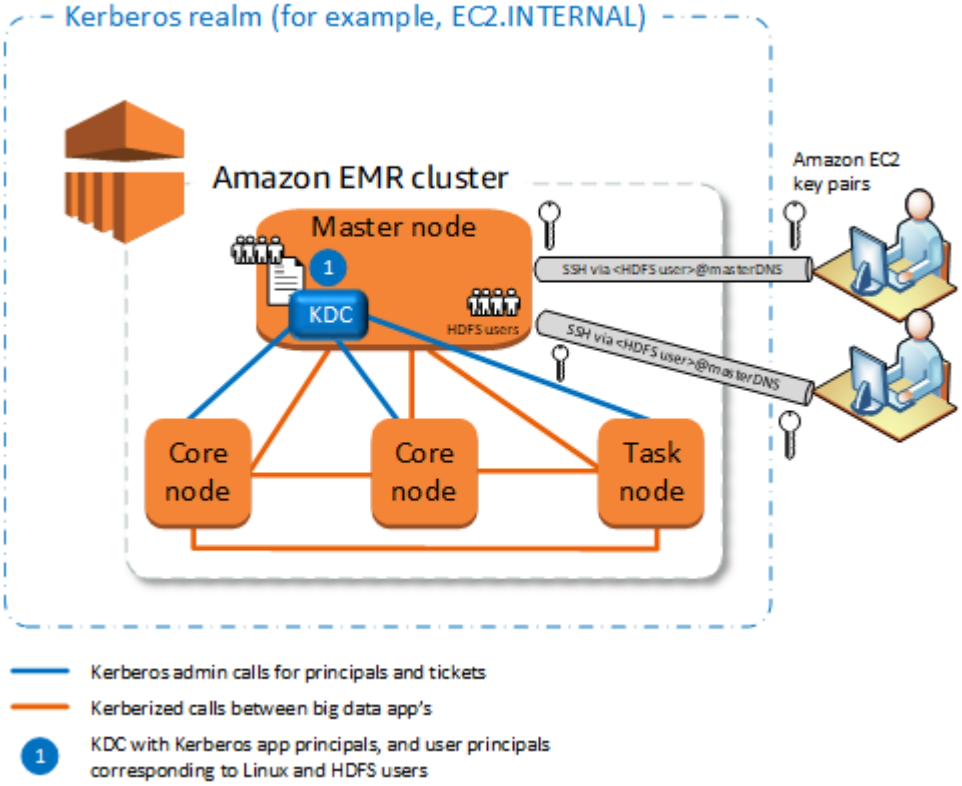
  - A representação de Livy com clusters Kerberizados é suportada nas versões 5.22.0 e superiores da AmazonEMR.
- Oozie
- Phoenix
- Presto
  - O Presto oferece suporte à autenticação Kerberos nas EMR versões 6.9.0 e superiores da Amazon.
  - Para usar a autenticação Kerberos com o Presto, é necessário habilitar a criptografia em trânsito.
- Spark
- Tez
- Trino
  - O Trino oferece suporte à autenticação Kerberos nas EMR versões 6.11.0 e superiores da Amazon.
  - Para usar a autenticação Kerberos com o Trino, é necessário habilitar a criptografia em trânsito.
- YARN
- Zeppelin
  - O Zeppelin é configurado somente para usar o Kerberos com o intérprete do Spark. Ele não é configurado para outros intérpretes.
  - A representação de usuário não oferece suporte a intérpretes kerberizados do Zeppelin além do Spark.
- Zookeeper
  - O cliente do Zookeeper não é compatível.

Opções de arquitetura do Kerberos

Ao usar o Kerberos com a AmazonEMR, você pode escolher entre as arquiteturas listadas nesta seção. Independentemente da arquitetura que escolhida, você pode configurar o Kerberos usando as mesmas etapas. Você cria uma configuração de segurança, especifica a configuração de segurança e as opções Kerberos compatíveis específicas do cluster ao criar o cluster e cria HDFS diretórios para usuários Linux no cluster que correspondem aos principais usuários no. KDC Para obter uma explicação sobre as opções de configuração e configurações de exemplo para cada arquitetura, consulte Configurando o Kerberos na Amazon EMR.

Dedicado ao cluster KDC (KDCno nó primário)

Essa configuração está disponível com as EMR versões 5.10.0 e posteriores da Amazon.



Vantagens

- A Amazon EMR tem propriedade total daKDC.
- O KDC no EMR cluster é independente de KDC implementações centralizadas, como Microsoft Active Directory ou. AWS Managed Microsoft AD
- O impacto no desempenho é mínimo porque KDC gerencia a autenticação somente para nós locais dentro do cluster.
- Opcionalmente, outros clusters Kerberizados podem referenciar o KDC como externo. KDC Para obter mais informações, consulte Externo KDC — nó primário em um cluster diferente.

Considerações e limitações

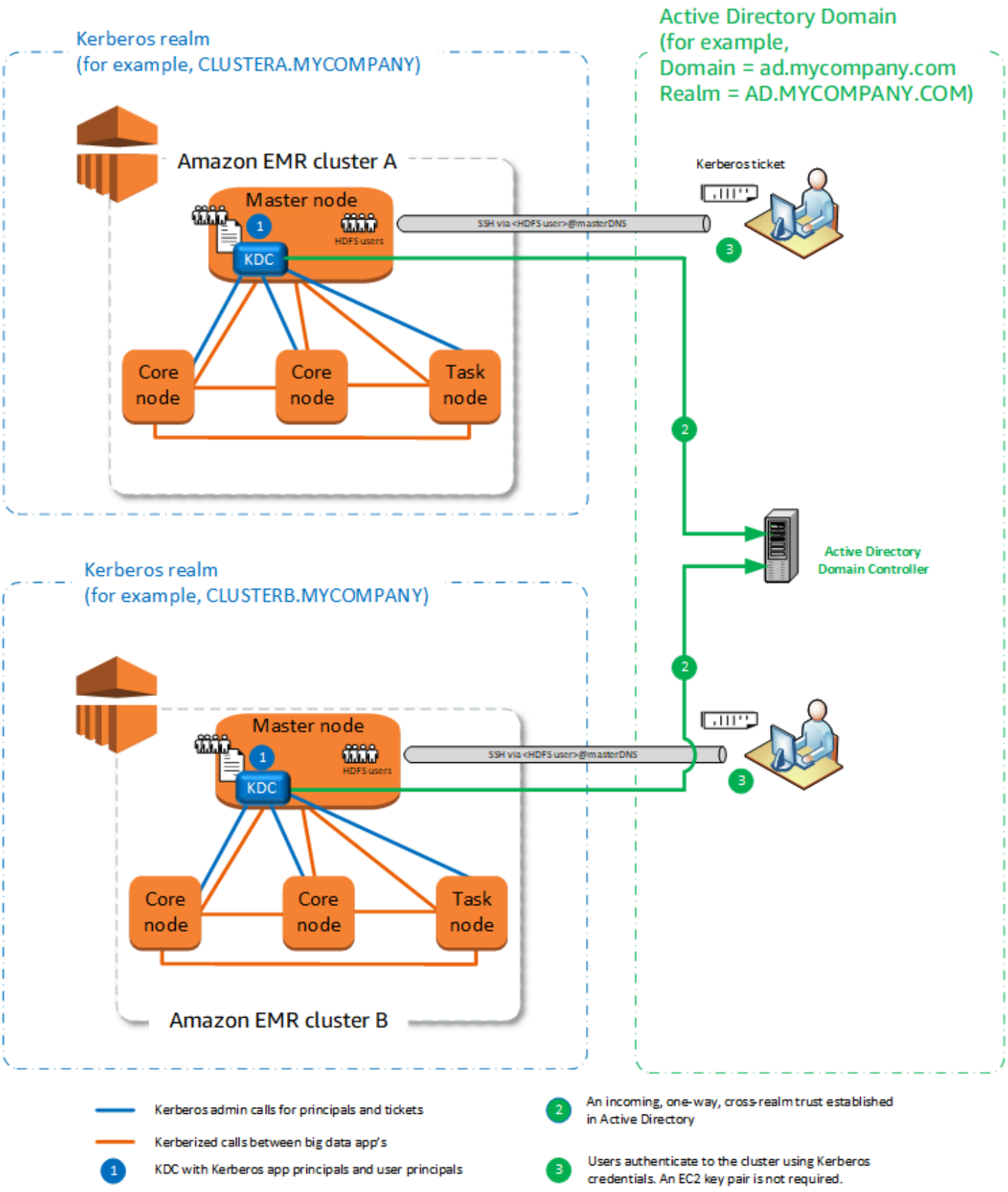
- Os clusters Kerberizados não podem autenticar uns ao outros, portanto, os aplicativos não podem interoperar. Se os aplicativos de cluster precisarem interoperar, você deverá estabelecer uma relação de confiança entre os clusters ou configurar um cluster como externo KDC para outros

clusters. Se uma relação de confiança entre reinos for estabelecida, eles KDCs devem ter reinos Kerberos diferentes.

- Você deve criar usuários Linux na EC2 instância do nó primário que correspondam aos principais KDC usuários, junto com os HDFS diretórios de cada usuário.
- Os usuários principais devem usar um arquivo de chave EC2 privada e `kinit` credenciais para se conectar ao cluster usando. SSH

Relação de confiança entre realms

Nessa configuração, os principais (geralmente usuários) de uma região Kerberos diferente se autenticam nos componentes do aplicativo em um cluster KerberizadoEMR, que tem seu próprio. KDC O KDC no nó primário estabelece uma relação de confiança com outro KDC usando um princípio entre domínios que existe em ambos. KDCs O nome principal e a senha coincidem exatamente em cada umKDC. Relações de confianças entre realms são mais comuns com implementações do Active Directory, conforme mostrado no diagrama a seguir. Confianças entre regiões com um cluster externo MIT KDC ou em KDC outro EMR cluster da Amazon também são suportadas.


Active Directory Domain
(for example,
Domain = ad.mycompany.com
Realm = AD.MYCOMPANY.COM)
Kerberos realm
(for example, CLUSTERA.MYCOMPANY)
Amazon EMR cluster A
Master node
1
HDFS users
KDC
SSH via <HDFS user>@masterDNS
Kerberos ticket
3
2
Core node
Core node
Task node
Active Directory Domain Controller
Kerberos realm
(for example, CLUSTERB.MYCOMPANY)
Master node
1
HDFS users
KDC
SSH via <HDFS user>@masterDNS
2
3
Core node
Core node
Task node
Amazon EMR cluster B
Kerberos admin calls for principals and tickets
Kerberized calls between big data app's
1
KDC with Kerberos app principals and user principals
2
An incoming, one-way, cross-realm trust established in Active Directory
3
Users authenticate to the cluster using Kerberos credentials. An EC2 key pair is not required.

Vantagens

- O EMR cluster no qual o KDC está instalado mantém a propriedade total doKDC.
- Com o Active Directory, a Amazon cria EMR automaticamente usuários Linux que correspondem aos principais usuários doKDC. Você ainda precisa criar HDFS diretórios para cada usuário. Além disso, usuários principais no domínio do Active Directory podem acessar clusters Kerberizados usando `kinit` credenciais, sem o EC2 arquivo de chave privada. Isso elimina a necessidade de compartilhar o arquivo de chave privada entre os usuários do cluster.
- Como cada cluster KDC gerencia a autenticação dos nós no cluster, os efeitos da latência da rede e da sobrecarga de processamento de um grande número de nós nos clusters são minimizados.

Considerações e limitações

- Se estiver estabelecendo uma relação de confiança com um domínio do Active Directory, você deverá fornecer um nome de usuário e senha do Active Directory com permissões para se juntar aos principais do domínio ao criar o cluster.
- As relações de confiança entre realms não podem ser estabelecidas entre realms do Kerberos com o mesmo nome.
- As relações de confiança entre realms deve ser estabelecidas explicitamente. Por exemplo, se o Cluster A e o Cluster B estabelecerem uma relação de confiança entre regiões com aKDC, eles não confiam inerentemente um no outro e seus aplicativos não podem se autenticar uns nos outros para interoperar.
- KDCsdevem ser mantidas de forma independente e coordenada para que as credenciais dos usuários principais correspondam com precisão.

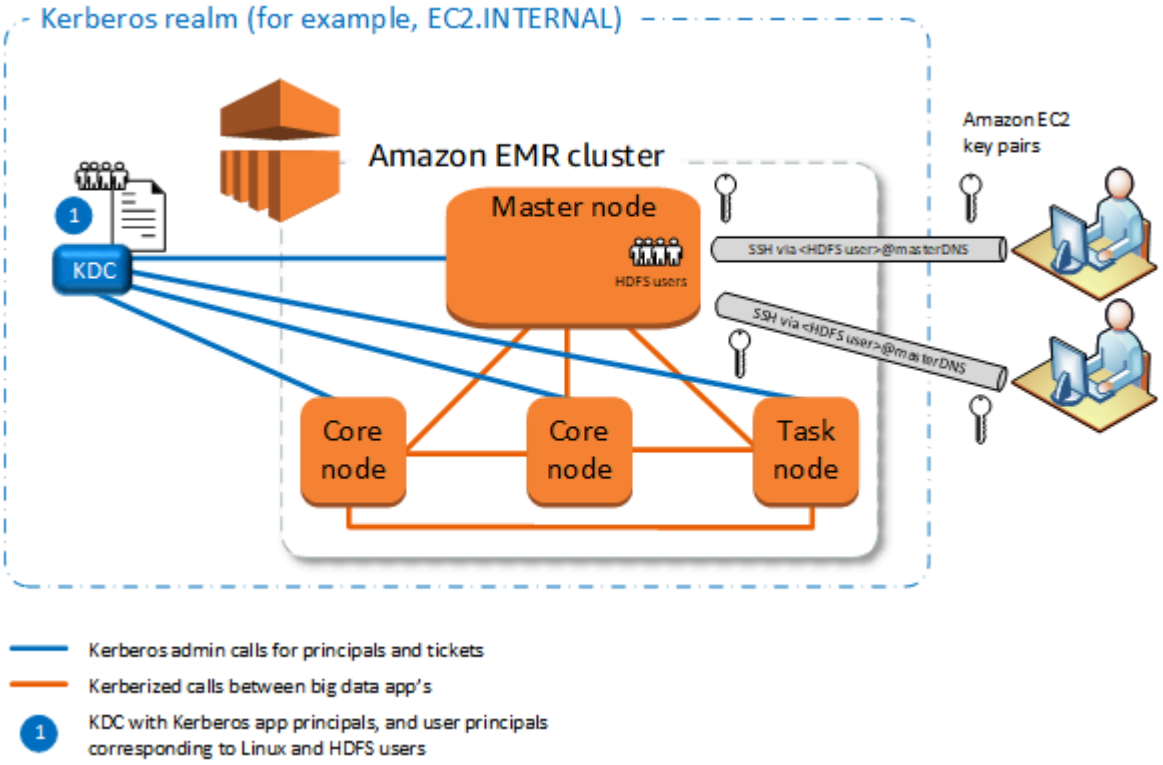
Externo KDC

As configurações com um externo KDC são compatíveis com o Amazon EMR 5.20.0 e versões posteriores.

- Externo KDC — MIT KDC
- Externo KDC — nó primário em um cluster diferente
- Externo KDC — cluster KDC em um cluster diferente com confiança entre regiões do Active Directory

Externo KDC — MIT KDC

Essa configuração permite que um ou mais EMR clusters usem os principais definidos e mantidos em um MIT KDC servidor.



Vantagens

- Os diretores administrativos são consolidados em um único. KDC
- Vários clusters podem usar o mesmo KDC no mesmo reino Kerberos. Para obter mais informações, consulte Requisitos para usar vários clusters com o mesmo KDC.
- O nó primário em um cluster Kerberizado não tem a carga de desempenho associada à manutenção do. KDC

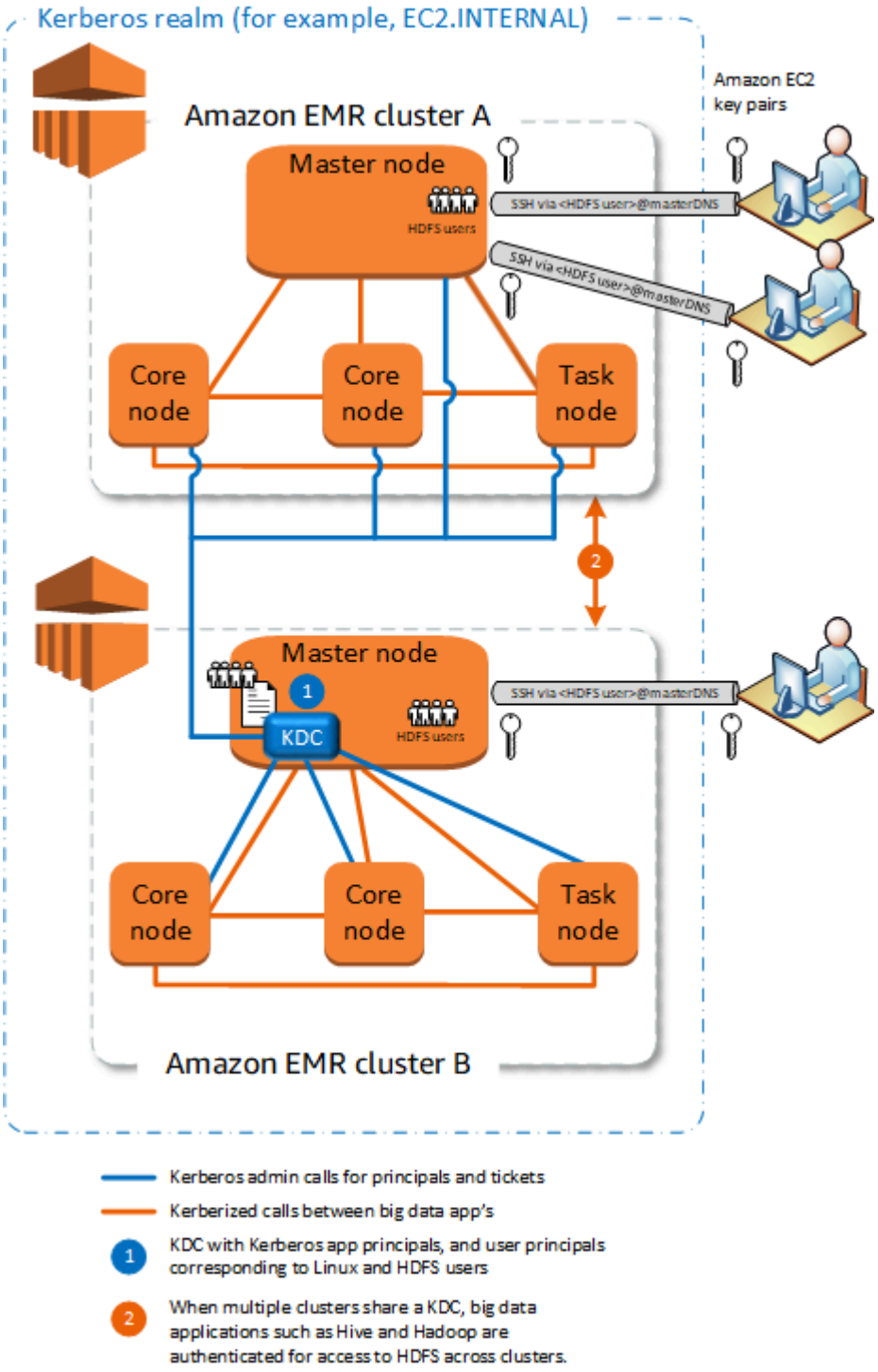
Considerações e limitações

- Você deve criar usuários Linux na EC2 instância do nó primário de cada cluster Kerberizado que correspondam aos principais usuários, junto com os HDFS diretórios de cada KDC usuário.
- Os usuários principais devem usar um arquivo de chave EC2 privada e `kinit` credenciais para se conectar aos clusters Kerberizados usando. SSH
- Cada nó em EMR clusters Kerberizados deve ter uma rota de rede para o. KDC

- Cada nó em clusters Kerberizados coloca uma carga de autenticação no externoKDC, portanto, a configuração do KDC afeta o desempenho do cluster. Ao configurar o hardware do KDC servidor, considere o número máximo de EMR nós da Amazon a serem suportados simultaneamente.
- O desempenho do cluster depende da latência da rede entre os nós nos clusters Kerberizados e o. KDC
- A solução de problemas pode ser mais difícil devido a interdependências.

Externo KDC — nó primário em um cluster diferente

Essa configuração é quase idêntica à MIT KDC implementação externa acima, exceto pelo fato de KDC estar no nó primário de um EMR cluster. Para ter mais informações, consulte Dedicado ao cluster KDC (KDCno nó primário) e Tutorial: configurar uma relação de confiança entre realms com um controlador de domínio do Active Directory.



Vantagens

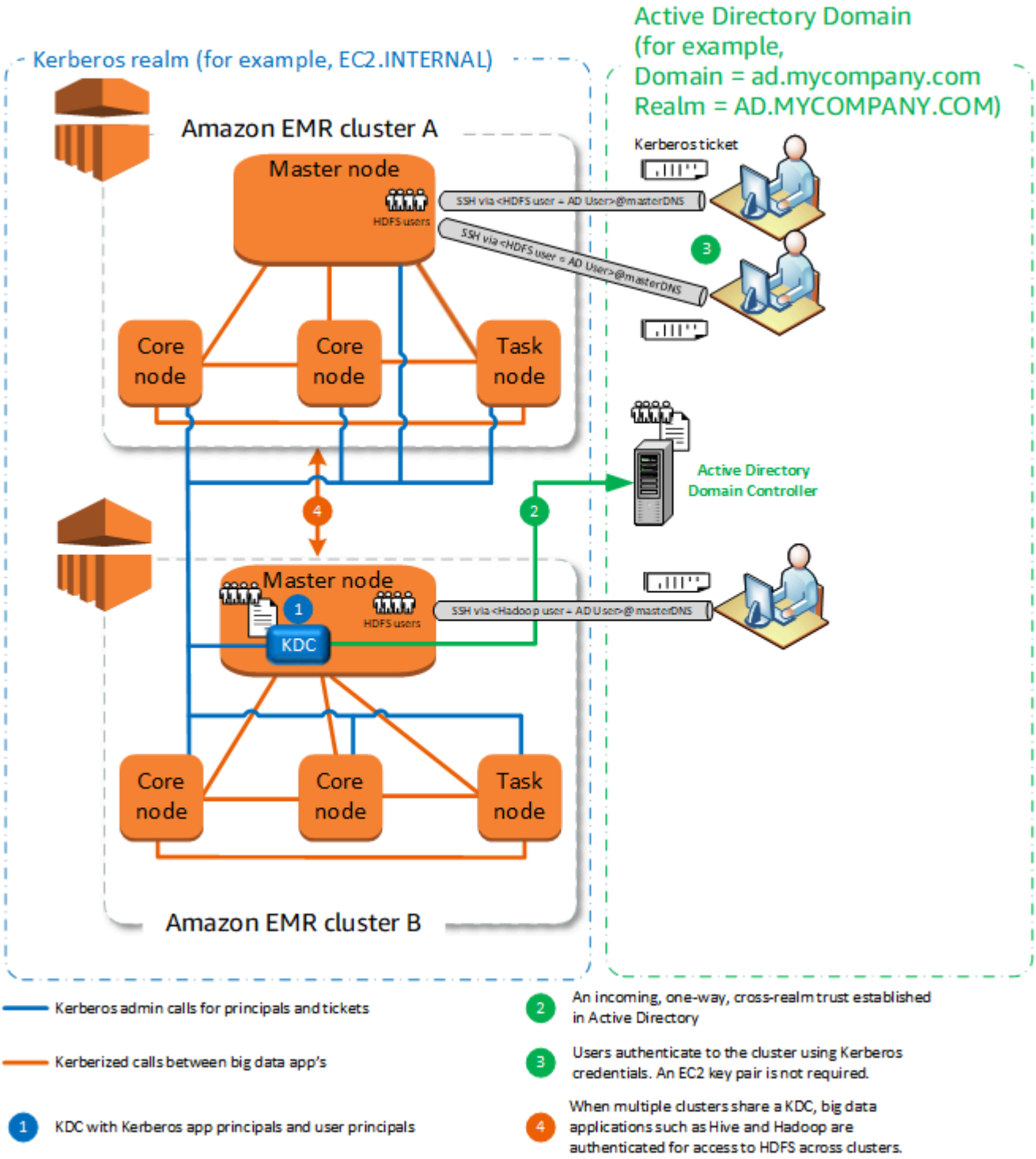
- Os diretores administrativos são consolidados em um único. KDC
- Vários clusters podem usar o mesmo KDC no mesmo reino Kerberos. Para obter mais informações, consulte Requisitos para usar vários clusters com o mesmo KDC.

Considerações e limitações

- Você deve criar usuários Linux na EC2 instância do nó primário de cada cluster Kerberizado que correspondam aos principais usuários, junto com os HDFS diretórios de cada KDC usuário.
- Os usuários principais devem usar um arquivo de chave EC2 privada e `kinit` credenciais para se conectar aos clusters Kerberizados usando. SSH
- Cada nó em cada EMR cluster deve ter uma rota de rede para KDC o.
- Cada EMR nó da Amazon em clusters Kerberizados coloca uma carga de autenticação no externoKDC, portanto, a configuração do KDC afeta o desempenho do cluster. Ao configurar o hardware do KDC servidor, considere o número máximo de EMR nós da Amazon a serem suportados simultaneamente.
- O desempenho do cluster depende da latência da rede entre os nós nos clusters e o. KDC
- A solução de problemas pode ser mais difícil devido a interdependências.

Externo KDC — cluster KDC em um cluster diferente com confiança entre regiões do Active Directory

Nessa configuração, primeiro você cria um cluster com um cluster dedicado KDC que tem uma relação de confiança unidirecional entre regiões com o Active Directory. Para ver um tutorial detalhado, consulte Tutorial: configurar uma relação de confiança entre realms com um controlador de domínio do Active Directory. Em seguida, você executa clusters adicionais, referenciando o cluster KDC que tem a confiança como externoKDC. Para ver um exemplo, consulte Cluster externo KDC com confiança entre regiões do Active Directory. Isso permite que cada EMR cluster da Amazon que usa o externo KDC autentique os principais definidos e mantidos em um domínio do Microsoft Active Directory.



Vantagens

- O gerenciamento de principais é consolidado no domínio do Active Directory.

- A Amazon EMR se junta ao reino do Active Directory, o que elimina a necessidade de criar usuários Linux que correspondam aos usuários do Active Directory. Você ainda precisa criar HDFS diretórios para cada usuário.
- Vários clusters podem usar o mesmo KDC no mesmo reino Kerberos. Para obter mais informações, consulte Requisitos para usar vários clusters com o mesmo KDC.
- Os usuários principais no domínio do Active Directory podem acessar clusters Kerberizados usando `kinit` credenciais, sem o EC2 arquivo de chave privada. Isso elimina a necessidade de compartilhar o arquivo de chave privada entre os usuários do cluster.
- Somente um nó EMR primário da Amazon tem a responsabilidade de manter oKDC, e somente esse cluster deve ser criado com as credenciais do Active Directory para a confiança entre regiões entre o KDC e o Active Directory.

Considerações e limitações

- Cada nó em cada EMR cluster deve ter uma rota de rede para o KDC e para o controlador de domínio do Active Directory.
- Cada EMR nó da Amazon coloca uma carga de autenticação no externoKDC, portanto, a configuração do KDC afeta o desempenho do cluster. Ao configurar o hardware do KDC servidor, considere o número máximo de EMR nós da Amazon a serem suportados simultaneamente.
- O desempenho do cluster depende da latência da rede entre os nós nos clusters e o KDC servidor.
- A solução de problemas pode ser mais difícil devido a interdependências.

Requisitos para usar vários clusters com o mesmo KDC

Vários clusters podem usar o mesmo KDC no mesmo reino Kerberos. No entanto, se os clusters forem executados simultaneamente, eles poderão falhar se usarem ServicePrincipal nomes Kerberos conflitantes.

Se você tiver vários clusters simultâneos com o mesmo externoKDC, certifique-se de que os clusters usem realms Kerberos diferentes. Se os clusters precisarem usar o mesmo reino Kerberos, certifique-se de que os clusters estejam em sub-redes diferentes e que seus CIDR intervalos não se sobreponham.

Configurando o Kerberos na Amazon EMR

Esta seção fornece detalhes da configuração e exemplos para configurar o Kerberos com arquiteturas comuns. Independentemente da arquitetura escolhida, as noções básicas de

configuração são as mesmas e a configuração é feita em três etapas. Se você usa uma relação de confiança externa KDC ou configura uma relação de confiança entre regiões, deve garantir que cada nó em um cluster tenha uma rota de rede para o externoKDC, incluindo a configuração de grupos de segurança aplicáveis para permitir tráfego Kerberos de entrada e saída.

Etapa 1: Criar uma configuração de segurança com propriedades do Kerberos

A configuração de segurança especifica detalhes sobre o Kerberos KDC e permite que a configuração do Kerberos seja reutilizada sempre que você cria um cluster. Você pode criar uma configuração de segurança usando o EMR console da Amazon AWS CLI, o ou EMR API o. A configuração de segurança também pode conter outras opções de segurança, como criptografia. Para obter mais informações sobre como criar configurações de segurança e especificar uma configuração de segurança ao criar um cluster, consulte Usar configurações de segurança para definir a segurança do cluster. Para obter informações sobre as propriedades do Kerberos em uma configuração de segurança, consulte Configurações do Kerberos para configurações de segurança.

Etapa 2: Criar um cluster e especificar os atributos do Kerberos específicos do cluster

Ao criar um cluster, você especifica uma configuração de segurança do Kerberos juntamente com e as opções do Kerberos específicas do cluster. Quando você usa o EMR console da Amazon, somente as opções Kerberos compatíveis com a configuração de segurança especificada estão disponíveis. Ao usar o AWS CLI ou a Amazon EMRAPI, certifique-se de especificar as opções do Kerberos compatíveis com a configuração de segurança especificada. Por exemplo, se você especificar uma senha principal para uma relação de confiança entre regiões ao criar um cluster usando o. e a CLI configuração de segurança especificada não estiver configurada com parâmetros de confiança entre regiões, ocorrerá um erro. Para obter mais informações, consulte Configurações do Kerberos para clusters.

Etapa 3: configurar o nó primário do cluster

Dependendo dos requisitos de sua arquitetura e implantação, configuração adicional no cluster pode ser necessária. Você pode fazer isso depois de criá-lo ou usando etapas ou ações de bootstrap durante o processo de criação.

Para cada usuário autenticado pelo Kerberos que se conecta ao cluster usandoSSH, você deve garantir que sejam criadas contas Linux que correspondam ao usuário Kerberos. Se os principais usuários forem fornecidos por um controlador de domínio do Active Directory, seja como externo KDC ou por meio de uma relação de confiança entre regiões, a Amazon EMR cria contas Linux automaticamente. Se o Active Directory não for usado, você deverá criar principais para cada usuário

que correspondam ao usuário do Linux. Para obter mais informações, consulte Configurando um cluster para usuários e conexões autenticados pelo Kerberos HDFS SSH.

Cada usuário também deve ter um diretório de HDFS usuários de sua propriedade, que você deve criar. Além disso, SSH deve ser configurado com GSSAPI enabled para permitir conexões de usuários autenticados pelo Kerberos. GSSAPIdeve estar habilitado no nó primário e o SSH aplicativo cliente deve estar configurado para usoGSSAPI. Para obter mais informações, consulte Configurando um cluster para usuários e conexões autenticados pelo Kerberos HDFS SSH.

Configuração de segurança e configurações de cluster para Kerberos na Amazon EMR

Ao criar um cluster Kerberizado, você especifica a configuração de segurança com atributos do Kerberos específicos do cluster. Você não pode especificar um conjunto sem o outro, ou ocorrerá um erro.

Este tópico fornece uma visão geral dos parâmetros de configuração disponíveis para o Kerberos quando você cria uma configuração de segurança e um cluster. Além disso, CLI exemplos para criar configurações de segurança e clusters compatíveis são fornecidos para arquiteturas comuns.

Configurações do Kerberos para configurações de segurança

Você pode criar uma configuração de segurança que especifique os atributos do Kerberos usando o EMR console da Amazon AWS CLI, o ou o. EMR API A configuração de segurança também pode conter outras opções de segurança, como criptografia. Para obter mais informações, consulte Criar uma configuração de segurança.

Use as referências a seguir para compreender as definições de configuração de segurança disponíveis para a arquitetura do Kerberos que você escolher. As configurações EMR do console Amazon são mostradas. Para obter CLI as opções correspondentes, consulte Especificando as configurações do Kerberos usando o AWS CLI ouExemplos de configuração.

Parâmetro	Descrição
Kerberos	Especifica que o Kerberos está habilitado em clusters que usam essa configuração de segurança. Ao usar essa configuração de segurança, o cluster também deverá ter configurações Kerberos especificadas ou ocorrerá um erro.

Parâmetro		Descrição
Provedor	Dedicado ao cluster KDC	Especifica que a Amazon EMR cria um KDC no nó primário de qualquer cluster que usa essa configura ção de segurança. Você especifica o nome do territóri o e a senha KDC do administrador ao criar o cluster. Você pode referenciar isso KDC em outros clusters, se necessário. Crie esses clusters usando uma configura ção de segurança diferente, especifique uma externa KDC e use o nome do território e a senha de KDC administrador que você especifica para o cluster KDC dedicado.
	Externo KDC	Disponível somente com o Amazon EMR 5.20.0 e versões posteriores. Especifica que os clusters que usam essa configuração de segurança autenticam os principais do Kerberos usando um KDC servidor fora do cluster. A não KDC é criado no cluster. Ao criar o cluster, você especifica o nome do território e a senha KDC do administrador para o externoKDC.
Vida útil do tíquete		Opcional. Especifica o período durante o qual um tíquete Kerberos emitido pelo KDC é válido em clusters que usam essa configuração de segurança. Os ciclos de vida do tíquete são limitados por motivos de segurança. As aplicações e os serviços de cluster renovarão automaticamente os tíquetes quando perderem a validade. Os usuários que se conectam ao cluster SSH usando as credenciais do Kerberos precisam executar a `kinit` partir da linha de comando do nó primário para renovar após a expiração de um ticket.

Parâmetro		Descrição
Relação de confiança entre realms		Especifica uma relação de confiança entre regiões entre um cluster dedicado KDC em clusters que usam essa configuração de segurança e um em uma KDC região Kerberos diferente. As entidades principais (normalmente usuários) de outro realm são autenticados em clusters que usam essa configuração. É necessário ter configuração adicional no outro realm do Kerberos. Para obter mais informações, consulte Tutorial: configurar uma relação de confiança entre realms com um controlador de domínio do Active Directory.
Propriedades de confiança entre realms	Realm	Especifica o nome de realm Kerberos de outro realm na relação de confiança. Por convenção, os nomes de realm do Kerberos são iguais ao nome do domínio, mas em letras maiúsculas.
	Domínio	Especifica o nome de domínio de outro realm na relação de confiança.
	Servidor do administrador	Especifica o nome de domínio totalmente qualificado (FQDN) ou o endereço IP do servidor administrativo no outro domínio da relação de confiança. O servidor administrativo e o KDC servidor normalmente são executados na mesma máquina com a mesmaFQDN, mas se comunicam em portas diferentes. Se nenhuma porta especificada, a porta 749 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com:749`).

Parâmetro		Descrição
	KDCservidor	Especifica o nome de domínio totalmente qualificado (FQDN) ou o endereço IP do KDC servidor no outro domínio da relação de confiança. O KDC servidor e o servidor de administração normalmente são executados na mesma máquina com a mesma portaFQDN, mas usam portas diferentes. Se nenhuma porta especificada, a porta 88 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com :88`).
Externo KDC		Especifica que os clusters externos KDC sejam usados pelo cluster.
KDCPropriedades externas	Servidor do administrador	Especifica o nome de domínio totalmente qualificado (FQDN) ou o endereço IP do servidor administrativo externo. O servidor administrativo e o KDC servidor normalmente são executados na mesma máquina com a mesmaFQDN, mas se comunicam em portas diferentes. Se nenhuma porta especificada, a porta 749 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com :749`).

Parâmetro			Descrição
	KDCservidor		Especifica o nome de domínio totalmente qualificado (FQDN) do KDC servidor externo. O KDC servidor e o servidor de administração normalmente são executados na mesma máquina com a mesma portaFQDN, mas usam portas diferentes. Se nenhuma porta especificada, a porta 88 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com :88`).
	Integração do Active Directory		Especifica que a autenticação da entidade principal do Kerberos está integrada a um domínio do Microsoft Active Directory.
	Propriedades de integração do Active Directory	Realm do Active Directory	Especifica o nome do realm do Kerberos do domínio do Active Directory. Por convenção, os nomes de realm do Kerberos geralmente são iguais ao nome do domínio, mas em letras maiúsculas.
		Domínio do Active Directory	Especifica o nome de domínio do Active Directory.
		Servidor do Active Directory	Especifica o nome de domínio totalmente qualificado (FQDN) do controlador de domínio do Microsoft Active Directory.

Configurações do Kerberos para clusters

Você pode especificar as configurações do Kerberos ao criar um cluster usando o EMR console da Amazon AWS CLI, o ou o. EMR API

Use as referências a seguir para compreender as definições de configuração de cluster disponíveis para a arquitetura do Kerberos que você escolher. As configurações EMR do console Amazon são mostradas. Para obter CLI as opções correspondentes, consulteExemplos de configuração.

Parâmetro	Descrição
Realm	O nome do realm do Kerberos para o cluster. A convenção do Kerberos deve ser a mesma do nome de domínio, mas em maiúsculas. Por exemplo, para o domínio `ec2.internal` , usando `EC2.INTERNAL` como o nome do realm.
KDCsenha de administrador	A senha usada dentro do cluster para `kadmin` ou `kadmin.local` . Essas são interfaces de linha de comando para o sistema de administração do Kerberos V5, que mantém os principais do Kerberos, as políticas de senha e os keytabs do cluster.
Senha do principal da relação de confiança entre realms (opcional)	Obrigatório quando se estabelece uma relação de confiança entre realms. A senha do principal entre realms, que deve ser idêntica em todos os realms. Use uma senha forte.
Usuário de inclusão no domínio do Active Directory (opcional)	Obrigatório ao usar o Active Directory em uma relação de confiança entre realms. Este é o nome de logon de usuário de uma conta do Active Directory com permissão para integrar computadores ao domínio. A Amazon EMR usa essa identidade para unir o cluster ao domínio. Para obter mais informações, consulte the section called “Etapa 3: Adicionar contas ao domínio para o EMR cluster”.
Senha de inclusão no domínio do Active Directory (opcional)	A senha para o usuário de inclusão no domínio do Active Directory. Para obter mais informaçõ es, consulte the section called “Etapa 3:

Parâmetro	Descrição
	Adicionar contas ao domínio para o EMR cluster".

Exemplos de configuração

Os exemplos a seguir demonstram configurações de segurança e configurações de cluster para cenários comuns. AWS CLI os comandos são mostrados para fins de concisão.

Local KDC

Os comandos a seguir criam um cluster com um cluster dedicado em KDC execução no nó primário. Configurações adicionais no cluster podem ser necessárias. Para obter mais informações, consulte Configurando um cluster para usuários e conexões autenticados pelo Kerberos HDFS SSH.

Criar configuração de segurança

```
aws emr create-security-configuration --name LocalKDCSecurityConfig \
--security-configuration '{"AuthenticationConfiguration": \
{"KerberosConfiguration": {"Provider": "ClusterDedicatedKdc",\
"ClusterDedicatedKdcConfiguration": {"TicketLifetimeInHours": 24 }}}}'
```

Criar cluster

```
aws emr create-cluster --release-label emr-7.2.0 \
--instance-count 3 --instance-type m5.xlarge \
--applications Name=Hadoop Name=Hive --ec2-attributes
 InstanceProfile=EMR_EC2_DefaultRole,KeyName=MyEC2Key \
--service-role EMR_DefaultRole \
--security-configuration LocalKDCSecurityConfig \
--kerberos-attributes Realm=EC2.INTERNAL,KdcAdminPassword=MyPassword
```

Dedicado ao cluster KDC com confiança entre regiões do Active Directory

Os comandos a seguir criam um cluster com um cluster dedicado em KDC execução no nó primário com uma relação de confiança entre regiões em um domínio do Active Directory. Configuração adicional no cluster e no Active Directory é necessária. Para obter mais informações, consulte Tutorial: configurar uma relação de confiança entre realms com um controlador de domínio do Active Directory.

Criar configuração de segurança

```
aws emr create-security-configuration --name LocalKDCWithADTrustSecurityConfig \
--security-configuration '{"AuthenticationConfiguration": \
{"KerberosConfiguration": {"Provider": "ClusterDedicatedKdc", \
"ClusterDedicatedKdcConfiguration": {"TicketLifetimeInHours": 24, \
"CrossRealmTrustConfiguration": {"Realm":"AD.DOMAIN.COM", \
"Domain":"ad.domain.com", "AdminServer":"ad.domain.com", \
"KdcServer":"ad.domain.com"}}}}}'
```

Criar cluster

```
aws emr create-cluster --release-label emr-7.2.0 \
--instance-count 3 --instance-type m5.xlarge --applications Name=Hadoop Name=Hive \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,KeyName=MyEC2Key \
--service-role EMR_DefaultRole --security-configuration KDCWithADTrustSecurityConfig \
--kerberos-attributes Realm=EC2.INTERNAL,KdcAdminPassword=MyClusterKDCAdminPassword,\
ADDomainJoinUser=ADUserLogonName,ADDomainJoinPassword=ADUserPassword,\
CrossRealmTrustPrincipalPassword=MatchADTrustPassword
```

KDCExterno em um cluster diferente

Os comandos a seguir criam um cluster que faz referência a um cluster dedicado KDC no nó primário de um cluster diferente para autenticar os principais. Configurações adicionais no cluster podem ser necessárias. Para obter mais informações, consulte Configurando um cluster para usuários e conexões autenticados pelo Kerberos HDFS SSH.

Criar configuração de segurança

```
aws emr create-security-configuration --name ExtKDCOnDifferentCluster \
--security-configuration '{"AuthenticationConfiguration": \
{"KerberosConfiguration": {"Provider": "ExternalKdc", \
"ExternalKdcConfiguration": {"KdcServerType": "Single", \
"AdminServer": "MasterDNSOfKDCMaster:749", \
"KdcServer": "MasterDNSOfKDCMaster:88"}}}}'
```

Criar cluster

```
aws emr create-cluster --release-label emr-7.2.0 \
--instance-count 3 --instance-type m5.xlarge \
```

```
--applications Name=Hadoop Name=Hive \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,KeyName=MyEC2Key \
--service-role EMR_DefaultRole --security-configuration ExtKDCOnDifferentCluster \
--kerberos-attributes Realm=EC2.INTERNAL,KdcAdminPassword=KDCOnMasterPassword
```

Cluster externo KDC com confiança entre regiões do Active Directory

Os comandos a seguir criam um cluster semKDC. O cluster faz referência a um cluster dedicado em KDC execução no nó primário de outro cluster para autenticar os principais. Isso KDC tem uma relação de confiança entre regiões com um controlador de domínio do Active Directory. É necessária uma configuração adicional no nó KDC primário com o. Para obter mais informações, consulte Tutorial: configurar uma relação de confiança entre realms com um controlador de domínio do Active Directory.

Criar configuração de segurança

```
aws emr create-security-configuration --name ExtKDCWithADIntegration \
--security-configuration '{"AuthenticationConfiguration": \
{"KerberosConfiguration": {"Provider": "ExternalKdc", \
"ExternalKdcConfiguration": {"KdcServerType": "Single", \
"AdminServer": "MasterDNSofClusterKDC:749", \
"KdcServer": "MasterDNSofClusterKDC.com:88", \
"AdIntegrationConfiguration": {"AdRealm":"AD.DOMAIN.COM", \
"AdDomain":"ad.domain.com", \
"AdServer":"ad.domain.com"}}}}'
```

Criar cluster

```
aws emr create-cluster --release-label emr-7.2.0 \
--instance-count 3 --instance-type m5.xlarge --applications Name=Hadoop Name=Hive \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,KeyName=MyEC2Key \
--service-role EMR_DefaultRole --security-configuration ExtKDCWithADIntegration \
--kerberos-attributes Realm=EC2.INTERNAL,KdcAdminPassword=KDCOnMasterPassword,\
ADDomainJoinUser=MyPrivilegedADUserName,ADDomainJoinPassword=PasswordForADDomainJoinUser
```

Configurando um cluster para usuários e conexões autenticados pelo Kerberos HDFS SSH

EMRA Amazon cria clientes de usuário autenticados pelo Kerberos para os aplicativos que são executados no cluster, por exemplo, o usuário, o `hadoop` usuário e outros. `spark` Você também pode adicionar usuários autenticados em processos de cluster usando o Kerberos. Os usuários

autenticados podem se conectar ao cluster usando as credenciais do Kerberos e trabalhar com os aplicativos. Para que um usuário faça autenticação no cluster, as seguintes configurações são necessárias:

- Uma conta Linux que corresponda ao principal do Kerberos KDC deve existir no cluster. EMRA Amazon faz isso automaticamente em arquiteturas que se integram ao Active Directory.
- Você deve criar um diretório de HDFS usuários no nó primário para cada usuário e conceder permissões ao usuário para o diretório.
- Você deve configurar o SSH serviço para que ele GSSAPI seja ativado no nó primário. Além disso, os usuários devem ter um SSH cliente com GSSAPI habilitado.

Adicionar usuários do Linux e entidades principais do Kerberos ao nó primário

Se você não usa o Active Directory, deve criar contas Linux no nó primário do cluster e adicionar os principais desses usuários Linux aoKDC. Isso inclui um principal no KDC nó primário. Além dos principais usuários, a KDC execução no nó primário precisa de um principal para o host local.

Quando sua arquitetura inclui a integração com o Active Directory, os usuários e diretores do Linux no localKDC, se aplicável, são criados automaticamente. Você pode ignorar esta etapa. Para ter mais informações, consulte Relação de confiança entre realms e Externo KDC — cluster KDC em um cluster diferente com confiança entre regiões do Active Directory.

> **Important**
>
> OKDC, junto com o banco de dados de principais, é perdido quando o nó primário é encerrado porque o nó primário usa armazenamento efêmero. Se você criar usuários para SSH conexões, recomendamos estabelecer uma relação de confiança entre regiões com um externo KDC configurado para alta disponibilidade. Como alternativa, se você criar usuários para SSH conexões usando contas Linux, automatize o processo de criação da conta usando ações e scripts de bootstrap para que ele possa ser repetido ao criar um novo cluster.

Enviar uma etapa ao cluster depois de criá-la ou ao criar o cluster é a maneira mais fácil de adicionar usuários e KDC principais. Como alternativa, você pode se conectar ao nó primário usando um EC2 key pair como `hadoop` usuário padrão para executar os comandos. Para obter mais informações, consulte Conecte-se ao nó primário usando SSH.

O exemplo a seguir envia um script bash `configureCluster.sh` para um cluster que já existe, fazendo referência ao ID do cluster. O script é salvo no Amazon S3.

```
aws emr add-steps --cluster-id <j-2AL4XXXXXX5T9> \
--steps Type=CUSTOM_JAR,Name=CustomJAR,ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-runner.jar,\
Args=["s3://DOC-EXAMPLE-BUCKET/configureCluster.sh"]
```

O exemplo a seguir demonstra o conteúdo do script `configureCluster.sh`. O script também trata da criação de diretórios de HDFS usuário e da ativação GSSAPI deSSH, que são abordados nas seções a seguir.

```
#!/bin/bash
#Add a principal to the KDC for the primary node, using the primary node's returned
 host name
sudo kadmin.local -q "ktadd -k /etc/krb5.keytab host/`hostname -f`"
#Declare an associative array of user names and passwords to add
declare -A arr
arr=([lijuan]=pwd1 [marymajor]=pwd2 [richardroe]=pwd3)
for i in ${!arr[@]}; do
    #Assign plain language variables for clarity
     name=${i}
     password=${arr[${i}]}

     # Create a principal for each user in the primary node and require a new password
 on first logon
     sudo kadmin.local -q "addprinc -pw $password +needchange $name"

     #Add hdfs directory for each user
     hdfs dfs -mkdir /user/$name

     #Change owner of each user's hdfs directory to that user
     hdfs dfs -chown $name:$name /user/$name
done

# Enable GSSAPI authentication for SSH and restart SSH service
sudo sed -i 's/^.*GSSAPIAuthentication.*$/GSSAPIAuthentication yes/' /etc/ssh/
sshd_config
sudo sed -i 's/^.*GSSAPICleanupCredentials.*$/GSSAPICleanupCredentials yes/' /etc/ssh/
sshd_config
sudo systemctl restart sshd
```

Adicionar HDFS diretórios de usuários

Para permitir que seus usuários façam login no cluster para executar trabalhos do Hadoop, você deve adicionar diretórios de HDFS usuários às contas Linux e conceder a cada usuário a propriedade do diretório.

Enviar uma etapa ao cluster depois de criá-la ou ao criar o cluster é a maneira mais fácil de criar HDFS diretórios. Como alternativa, você pode se conectar ao nó primário usando um EC2 key pair como `hadoop` usuário padrão para executar os comandos. Para obter mais informações, consulte Conecte-se ao nó primário usando SSH.

O exemplo a seguir envia um script bash `AddHDFSUsers.sh` para um cluster que já existe, fazendo referência ao ID do cluster. O script é salvo no Amazon S3.

```
aws emr add-steps --cluster-id <j-2AL4XXXXXX5T9> \
--steps Type=CUSTOM_JAR,Name=CustomJAR,ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://DOC-EXAMPLE-BUCKET/AddHDFSUsers.sh"]
```

O exemplo a seguir demonstra o conteúdo do script `AddHDFSUsers.sh`.

```
#!/bin/bash
# AddHDFSUsers.sh script

# Initialize an array of user names from AD, or Linux users created manually on the
 cluster
ADUSERS=("lijuan" "marymajor" "richardroe" "myusername")

# For each user listed, create an HDFS user directory
# and change ownership to the user

for username in ${ADUSERS[@]}; do
     hdfs dfs -mkdir /user/$username
     hdfs dfs -chown $username:$username /user/$username
done
```

Habilitando GSSAPI para SSH

Para que os usuários autenticados pelo Kerberos se conectem ao nó primário usandoSSH, o SSH serviço deve ter a autenticação habilitada. GSSAPI Para habilitarGSSAPI, execute os comandos

a seguir na linha de comando do nó primário ou use uma etapa para executá-los como um script. Depois de reconfigurarSSH, você deve reiniciar o serviço.

```
sudo sed -i 's/^.*GSSAPIAuthentication.*$/GSSAPIAuthentication yes/' /etc/ssh/
sshd_config
sudo sed -i 's/^.*GSSAPICleanupCredentials.*$/GSSAPICleanupCredentials yes/' /etc/ssh/
sshd_config
sudo systemctl restart sshd
```

Usando SSH para se conectar a clusters Kerberizados

Esta seção demonstra as etapas para que um usuário autenticado pelo Kerberos se conecte ao nó primário de um cluster. EMR

Cada computador usado para uma SSH conexão deve ter aplicativos SSH cliente e cliente Kerberos instalados. Os computadores Linux provavelmente incluem esses aplicativos por padrão. Por exemplo, o Open SSH está instalado na maioria dos sistemas operacionais Linux, Unix e macOS. Você pode verificar se há um SSH cliente digitando ssh na linha de comando. Se o computador não reconhecer o comando, instale um SSH cliente para se conectar ao nó primário. O SSH projeto Open fornece uma implementação gratuita do conjunto completo de SSH ferramentas. Para obter mais informações, consulte o SSH site da Open. Os usuários do Windows podem usar aplicativos como o Pu TTY como SSH cliente.

Para obter mais informações sobre SSH conexões, consulteConectar-se a um cluster.

SSHusa GSSAPI para autenticar clientes Kerberos, e você deve habilitar a GSSAPI autenticação para o SSH serviço no nó primário do cluster. Para obter mais informações, consulte Habilitando GSSAPI para SSH. SSHos clientes também devem usarGSSAPI.

Nos exemplos a seguir, para *MasterPublicDNS* use o valor que aparece para Master public DNS na guia Resumo do painel de detalhes do cluster — por exemplo, *ec2-11-222-33-44.compute-1.amazonaws.com*.

Pré-requisito para krb5.conf (que não é do Active Directory)

Ao usar uma configuração sem integração com o Active Directory, além dos aplicativos SSH cliente e cliente Kerberos, cada computador cliente deve ter uma cópia do `/etc/krb5.conf` arquivo que corresponda ao `/etc/krb5.conf` arquivo no nó primário do cluster.

Para copiar o arquivo krb5.conf

1. Use SSH para se conectar ao nó primário usando um EC2 key pair e o `hadoop` usuário padrão, por exemplo,. `hadoop@`*`MasterPublicDNS`* Para obter instruções detalhadas, consulte Conectar-se a um cluster.
2. No nó primário, copie o conteúdo do arquivo `/etc/krb5.conf`. Para obter mais informações, consulte Conectar-se a um cluster.
3. Em cada computador cliente que será usado para se conectar ao cluster, crie um arquivo `/etc/krb5.conf` idêntico com base na cópia feita na etapa anterior.

Usando kinit e SSH

Cada vez que um usuário se conecta a partir de um computador cliente usando credenciais do Kerberos, o usuário deve primeiro renovar tíquetes Kerberos para seu usuário no computador cliente. Além disso, o SSH cliente deve estar configurado para usar a GSSAPI autenticação.

Para usar SSH para se conectar a um cluster Kerberizado EMR

1. Use `kinit` para renovar os tíquetes Kerberos, conforme mostrado no exemplo a seguir

   ```
   kinit user1
   ```

2. Use um `ssh` cliente junto com o principal que você criou no nome de usuário dedicado ao cluster KDC ou do Active Directory. Certifique-se de que a GSSAPI autenticação esteja habilitada conforme mostrado nos exemplos a seguir.

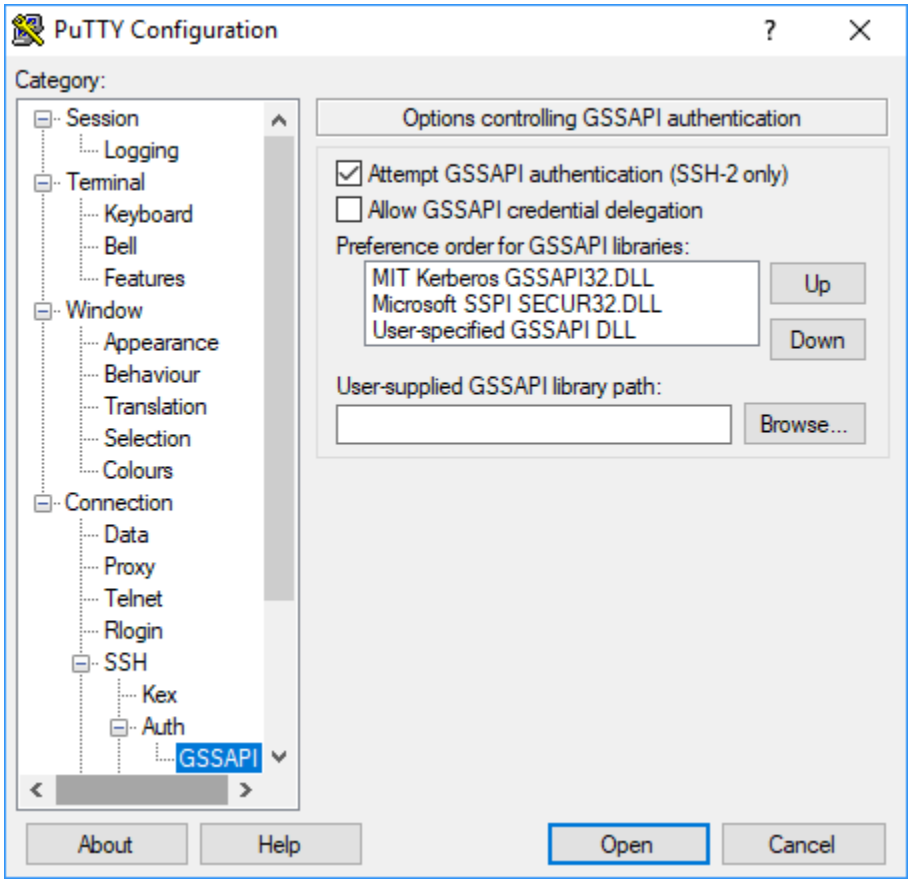
   Exemplo: usuários do Linux

   A `-K` opção especifica a GSSAPI autenticação.

   ```
   ssh -K user1@MasterPublicDNS
   ```

   Exemplo: usuários do Windows (PuTTY)

   Certifique-se de que a opção de GSSAPI autenticação para a sessão esteja ativada conforme mostrado:



Tutorial: Configurar um cluster dedicado KDC

Este tópico orienta você na criação de um cluster com um centro de distribuição de chaves dedicado ao cluster (KDC), na adição manual de contas Linux a todos os nós do cluster, na adição de entidades do Kerberos no nó primário e KDC na garantia de que os computadores clientes tenham um cliente Kerberos instalado.

Para obter mais informações sobre o EMR suporte da Amazon para Kerberos eKDC, além de links para a documentação do MIT Kerberos, consulte. Use o Kerberos para autenticação com a Amazon EMR

Etapa 1: criar o cluster kerberizado

1. Crie uma configuração de segurança que permita o Kerberos. O exemplo a seguir demonstra um `create-security-configuration` comando usando o AWS CLI que especifica a configuração de segurança como uma estrutura JSON embutida. Você também pode fazer referência a um arquivo salvo localmente.

```
aws emr create-security-configuration --name MyKerberosConfig \
--security-configuration '{"AuthenticationConfiguration": {"KerberosConfiguration":
{"Provider": "ClusterDedicatedKdc", "ClusterDedicatedKdcConfiguration":
 {"TicketLifetimeInHours": 24}}}}'
```

2. Crie um cluster que faça referência à configuração de segurança, estabeleça os atributos do Kerberos para o cluster e adicione contas do Linux usando uma ação de bootstrap. O exemplo a seguir demonstra um comando `create-cluster` usando a AWS CLI. O comando faz referência à configuração de segurança criada por você acima, `MyKerberosConfig`. Ele também faz referência a um script simples, `createlinuxusers.sh`, como uma ação de bootstrap, que você cria e carrega no Amazon S3 antes de criar o cluster.

   ```
   aws emr create-cluster --name "MyKerberosCluster" \
   --release-label emr-7.2.0 \
   --instance-type m5.xlarge \
   --instance-count 3 \
   --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,KeyName=MyEC2KeyPair \
   --service-role EMR_DefaultRole \
   --security-configuration MyKerberosConfig \
   --applications Name=Hadoop Name=Hive Name=Oozie Name=Hue Name=HCatalog Name=Spark \
   --kerberos-attributes Realm=EC2.INTERNAL,\
   KdcAdminPassword=MyClusterKDCAdminPwd \
   --bootstrap-actions Path=s3://DOC-EXAMPLE-BUCKET/createlinuxusers.sh
   ```

   O código a seguir demonstra o conteúdo do script `createlinuxusers.sh`, que adiciona user1, user2 e user3 a cada nó no cluster. Na próxima etapa, você adiciona esses usuários como KDC principais.

   ```
   #!/bin/bash
   sudo adduser user1
   sudo adduser user2
   sudo adduser user3
   ```

Etapa 2: Adicionar diretórios de usuários aoKDC, criar diretórios de HDFS usuários e configurar SSH

A KDC execução no nó primário precisa de um principal adicionado para o host local e para cada usuário que você cria no cluster. Você também pode criar HDFS diretórios para cada usuário se ele precisar se conectar ao cluster e executar trabalhos do Hadoop. Da mesma forma, configure o

SSH serviço para habilitar a GSSAPI autenticação, que é necessária para o Kerberos. Depois de habilitarGSSAPI, reinicie o SSH serviço.

A maneira mais fácil de realizar essas tarefas é enviar uma etapa para o cluster. O exemplo a seguir envia um `configurekdc.sh` de script bash para o cluster que você criou na etapa anterior, referenciando o ID do cluster. O script é salvo no Amazon S3. Como alternativa, você pode se conectar ao nó primário usando um EC2 key pair para executar os comandos ou enviar a etapa durante a criação do cluster.

```
aws emr add-steps --cluster-id <j-2AL4XXXXXX5T9> --steps
 Type=CUSTOM_JAR,Name=CustomJAR,ActionOnFailure=CONTINUE,Jar=s3://
myregion.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://DOC-EXAMPLE-
BUCKET/configurekdc.sh"]
```

O código a seguir demonstra o conteúdo do script `configurekdc.sh`.

```
#!/bin/bash
#Add a principal to the KDC for the primary node, using the primary node's returned
 host name
sudo kadmin.local -q "ktadd -k /etc/krb5.keytab host/`hostname -f`"
#Declare an associative array of user names and passwords to add
declare -A arr
arr=([user1]=pwd1 [user2]=pwd2 [user3]=pwd3)
for i in ${!arr[@]}; do
    #Assign plain language variables for clarity
     name=${i}
     password=${arr[${i}]}

     # Create principal for sshuser in the primary node and require a new password on
 first logon
     sudo kadmin.local -q "addprinc -pw $password +needchange $name"

     #Add user hdfs directory
     hdfs dfs -mkdir /user/$name

     #Change owner of user's hdfs directory to user
     hdfs dfs -chown $name:$name /user/$name
done

# Enable GSSAPI authentication for SSH and restart SSH service
sudo sed -i 's/^.*GSSAPIAuthentication.*$/GSSAPIAuthentication yes/' /etc/ssh/
sshd_config
```

```
sudo sed -i 's/^.*GSSAPICleanupCredentials.*$/GSSAPICleanupCredentials yes/' /etc/ssh/
sshd_config
sudo systemctl restart sshd
```

Os usuários que você adicionou agora devem poder se conectar ao cluster usando SSH o. Para obter mais informações, consulte Usando SSH para se conectar a clusters Kerberizados.

Tutorial: configurar uma relação de confiança entre realms com um controlador de domínio do Active Directory

Ao configurar uma relação de confiança entre regiões, você permite que os principais (geralmente usuários) de um realm Kerberos diferente se autentiquem nos componentes do aplicativo no cluster. EMR O centro de distribuição de chaves dedicado ao cluster (KDC) estabelece uma relação de confiança com outro KDC usando um princípio entre domínios que existe em ambos. KDCs O nome do principal e a senha coincidem precisamente.

Uma relação de confiança entre domínios exige que eles KDCs possam entrar em contato uns com os outros pela rede e resolver os nomes de domínio uns dos outros. As etapas para estabelecer uma relação de confiança entre regiões com um controlador de domínio do Microsoft AD em execução como uma EC2 instância são fornecidas abaixo, junto com um exemplo de configuração de rede que fornece a conectividade necessária e a resolução do nome de domínio. Qualquer configuração de rede que permita o tráfego de rede necessário KDCs é aceitável.

Opcionalmente, depois de estabelecer uma relação de confiança entre regiões com o Active Directory usando um cluster KDC em um, você pode criar outro cluster usando uma configuração de segurança diferente para referenciar o KDC primeiro cluster como externo. KDC Para obter um exemplo de configuração de segurança e a configuração do cluster, consulte Cluster externo KDC com confiança entre regiões do Active Directory.

Para obter mais informações sobre o EMR suporte da Amazon para Kerberos eKDC, além de links para a documentação do MIT Kerberos, consulte. Use o Kerberos para autenticação com a Amazon EMR

> **Important**
>
> A Amazon EMR não oferece suporte a relações de confiança entre reinos com. AWS Directory Service for Microsoft Active Directory

Etapa 1: configurar a VPC sub-rede e

Etapa 1: configurar a VPC sub-rede e

As etapas a seguir demonstram a criação de uma sub-rede VPC e para que o cluster dedicado KDC possa acessar o controlador de domínio do Active Directory e resolver seu nome de domínio. Nessas etapas, a resolução do nome de domínio é fornecida referenciando o controlador de domínio do Active Directory como o servidor de nomes de domínio no conjunto de opções. DHCP Para obter mais informações, consulte Etapa 5: Usar um conjunto de DHCP opções para especificar o controlador de domínio do Active Directory como VPC DNS servidor.

O KDC e o controlador de domínio do Active Directory devem ser capazes de resolver os nomes de domínio uns dos outros. Isso permite que EMR a Amazon associe computadores ao domínio e configure automaticamente contas e SSH parâmetros Linux correspondentes em instâncias de cluster.

Se a Amazon não EMR conseguir resolver o nome de domínio, você poderá referenciar a confiança usando o endereço IP do controlador de domínio do Active Directory. No entanto, você deve adicionar manualmente contas Linux, adicionar os principais correspondentes ao cluster dedicado e configurarKDC. SSH

Para configurar a VPC sub-rede e

1. Crie uma Amazon VPC com uma única sub-rede pública. Para obter mais informações, consulte Etapa 1: Crie o VPC no Guia de conceitos VPC básicos da Amazon.

   > **Important**
   >
   > Ao usar um controlador de domínio do Microsoft Active Directory, escolha um CIDR bloco para o EMR cluster de forma que todos os IPv4 endereços tenham menos de

nove caracteres (por exemplo, 10.0.0.0/16). Isso ocorre porque os DNS nomes dos computadores de cluster são usados quando os computadores ingressam no diretório do Active Directory. AWS atribui DNSnomes de host com base no IPv4 endereço de forma que endereços IP mais longos possam resultar em DNS nomes com mais de 15 caracteres. O Active Directory tem um limite de 15 caracteres para registrar nomes de computador adicionados e trunca nomes mais longos, o que pode causar erros imprevisíveis.

2. Remova o conjunto de DHCP opções padrão atribuído aoVPC. Para obter mais informações, consulte Alterando a VPC para usar Nenhuma DHCP opção. Posteriormente, você adiciona um novo que especifica o controlador de domínio do Active Directory como DNS servidor.
3. Confirme se o DNS suporte está habilitado para oVPC, ou seja, se DNS os nomes de host e a DNS resolução estão habilitados. Por padrão, as transições estão ativadas. Para obter mais informações, consulte Atualização do DNS suporte para seu VPC.
4. Confirme se você VPC tem um gateway de internet conectado, que é o padrão. Para mais informações, consulte Criar e anexar um gateway da Internet.

> **Note**
>
> Um gateway de internet é usado neste exemplo porque você está estabelecendo um novo controlador de domínio para VPC o. O gateway da Internet talvez não seja necessário para o aplicativo. O único requisito é que o cluster dedicado KDC possa acessar o controlador de domínio do Active Directory.

5. Crie uma tabela de rotas personalizada, adicione uma rota com o gateway da Internet como destino e a anexe à sub-rede. Para obter mais informações, consulte Criar uma tabela de rotas personalizada.
6. Quando você executa a EC2 instância do controlador de domínio, ela deve ter um IPv4 endereço público estático para você se conectar a ela usandoRDP. A maneira mais fácil de fazer isso é configurar sua sub-rede para atribuir endereços públicos IPv4 automaticamente. Não se trata da configuração padrão quando uma sub-rede é criada. Para obter mais informações, consulte Modificação do atributo de IPv4 endereçamento público da sua sub-rede. Você também pode atribuir o endereço ao iniciar a instância. Para obter mais informações, consulte Atribuição de um IPv4 endereço público durante a execução da instância.
7. Ao terminar, anote sua sub-rede VPC IDs e. Você os usará depois quando iniciar o controlador de domínio do Active Directory e o cluster.

Etapa 2: iniciar e instalar o controlador de domínio do Active Directory

1. Execute uma EC2 instância com base no Microsoft Windows Server 2016 BaseAMI. Recomendamos um tipo de instância m4.xlarge ou melhor. Para obter mais informações, consulte Lançamento de uma AWS Marketplace instância no Guia EC2 do usuário da Amazon.
2. Anote o ID do grupo de segurança associado à EC2 instância. Você precisa dele para o Etapa 6: iniciar um cluster kerberizado EMR. Nós usamos *sg-012xrlmdomain345*. Como alternativa, você pode especificar grupos de segurança diferentes para o EMR cluster e essa instância que permite o tráfego entre eles. Para obter mais informações, consulte Grupos EC2 de segurança da Amazon para instâncias Linux no Guia EC2 do usuário da Amazon.
3. Conecte-se à EC2 instância usando RDP o. Para obter mais informações, consulte Conectando-se à sua instância do Windows no Guia EC2 do usuário da Amazon.
4. Inicie o Server Manager para instalar e configurar o perfil Active Directory Domain Services no servidor. Promova o servidor para um controlador de domínio e atribua um nome de domínio (o exemplo que usamos aqui é *ad.domain.com*). Anote o nome do domínio porque você precisará dele posteriormente ao criar a configuração de EMR segurança e o cluster. Se estiver começando a configurar o Active Directory, você poderá seguir as instruções em How to setup Active Directory (AD) In Windows Server 2016.

   A instância será reiniciada quando você terminar.

Etapa 3: Adicionar contas ao domínio para o EMR cluster

RDPao controlador de domínio do Active Directory para criar contas em Usuários e Computadores do Active Directory para cada usuário do cluster. Para obter mais informações, consulte Create a User Account in Active Directory Users and Computers no site Microsoft Learn. Anote o User logon name (Nome de logon do usuário) de cada usuário. Você precisará dele mais tarde ao configurar o cluster.

Além disso, crie uma conta com privilégios suficientes para integrar computadores ao domínio. Você especifica essa conta ao criar um cluster. A Amazon o EMR usa para unir instâncias de cluster ao domínio. Você especifica essa conta e a senha em Etapa 6: iniciar um cluster kerberizado EMR. Para delegar privilégios de integração do computador à conta, recomendamos criar um grupo com privilégios de junção e, em seguida, atribuir o usuário ao grupo. Para obter instruções, consulte Delegating directory join privileges no Guia de administração AWS Directory Service .

Etapa 4: configurar uma relação de confiança recebida no controlador de domínio do Active Directory

Os comandos de exemplo abaixo criam uma relação de confiança no Active Directory, que é uma relação de confiança de região unidirecional, de entrada e não transitiva com o cluster dedicado. KDC O exemplo que usamos para no realm do cluster é *EC2.INTERNAL*. Substitua o *KDC-FQDN* com o DNS nome público listado para o nó EMR primário da Amazon que hospeda KDC o. O parâmetro `passwordt` especifica a cross-realm principal password (senha da entidade principal entre realms), determinada por você com o realm do cluster ao criar um cluster. O nome do realm deriva do nome de domínio padrão em `us-east-1` para o cluster. O `Domain` é o domínio do Active Directory no qual você está criando a confiança, que é em minúscula por convenção. O exemplo usa *ad.domain.com*

Abra o prompt de comando do Windows com privilégios de administrador e digite os seguintes comandos para criar a relação de confiança no controlador de domínio do Active Directory:

```
C:\Users\Administrator> ksetup /addkdc EC2.INTERNAL KDC-FQDN
C:\Users\Administrator> netdom trust EC2.INTERNAL /Domain:ad.domain.com /add /realm /
passwordt:MyVeryStrongPassword
C:\Users\Administrator> ksetup /SetEncTypeAttr EC2.INTERNAL AES256-CTS-HMAC-SHA1-96
```

Etapa 5: Usar um conjunto de DHCP opções para especificar o controlador de domínio do Active Directory como VPC DNS servidor

Agora que o controlador de domínio do Active Directory está configurado, você deve configurá-lo VPC para usá-lo como um servidor de nome de domínio para resolução de nomes em seuVPC. Para fazer isso, anexe um conjunto DHCP de opções. Especifique o Nome do domínio como o nome de domínio do cluster. Por exemplo, `ec2.internal` caso o cluster esteja em us-east-1 ou *region*`.compute.internal` para outras regiões. Para servidores de nomes de domínio, você deve especificar o endereço IP do controlador de domínio do Active Directory (que deve ser acessível a partir do cluster) como a primeira entrada, seguido por AmazonProvidedDNS(por exemplo, ***xx.xx.xx.xx***, AmazonProvided DNS). Para obter mais informações, consulte Alteração dos conjuntos de DHCP opções.

Etapa 6: iniciar um cluster kerberizado EMR

1. Na AmazonEMR, crie uma configuração de segurança que especifique o controlador de domínio do Active Directory que você criou nas etapas anteriores. Um comando de exemplo é mostrado abaixo. Substitua o domínio, *ad.domain.com*, pelo nome do domínio especificado por você em Etapa 2: iniciar e instalar o controlador de domínio do Active Directory.

```
aws emr create-security-configuration --name MyKerberosConfig \
--security-configuration '{
  "AuthenticationConfiguration": {
    "KerberosConfiguration": {
      "Provider": "ClusterDedicatedKdc",
      "ClusterDedicatedKdcConfiguration": {
        "TicketLifetimeInHours": 24,
        "CrossRealmTrustConfiguration": {
          "Realm": "AD.DOMAIN.COM",
          "Domain": "ad.domain.com",
          "AdminServer": "ad.domain.com",
          "KdcServer": "ad.domain.com"
        }
      }
    }
  }
}'
```

2. Crie o cluster com os seguintes atributos:

   - Use a opção `--security-configuration` para especificar a configuração de segurança que você criou. Nós usamos *`MyKerberosConfig`* no exemplo.
   - Use a propriedade `SubnetId` da `--ec2-attributes option` para especificar a sub-rede que você criou em Etapa 1: configurar a VPC sub-rede e. Nós usamos *`step1-subnet`* no exemplo.
   - Use `AdditionalMasterSecurityGroups` e `AdditionalSlaveSecurityGroups` da opção `--ec2-attributes` para especificar que o grupo de segurança associado ao controlador de domínio AD do Etapa 2: iniciar e instalar o controlador de domínio do Active Directory está associado ao nó primário do cluster, bem como aos nós centrais e de tarefa. Nós usamos *`sg-012xrlmdomain345`* no exemplo.

   Use `--kerberos-attributes` para especificar os seguintes atributos Kerberos específicos ao cluster:

   - O realm do cluster especificado por você ao configurar o controlador de domínio do Active Directory.
   - A senha da entidade principal da relação de confiança entre realms especificada por você como `passwordt` em Etapa 4: configurar uma relação de confiança recebida no controlador de domínio do Active Directory.
   - A`KdcAdminPassword`, que você pode usar para administrar o cluster KDC dedicado.

- O nome de logon do usuário e a senha da conta do Active Directory com privilégios de ingresso no computador criados por você em Etapa 3: Adicionar contas ao domínio para o EMR cluster.

O exemplo a seguir inicia um cluster kerberizado.

```
aws emr create-cluster --name "MyKerberosCluster" \
--release-label emr-5.10.0 \
--instance-type m5.xlarge \
--instance-count 3 \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,KeyName=MyEC2KeyPair,\
SubnetId=step1-subnet, AdditionalMasterSecurityGroups=sg-012xrlmdomain345,
AdditionalSlaveSecurityGroups=sg-012xrlmdomain345\
--service-role EMR_DefaultRole \
--security-configuration MyKerberosConfig \
--applications Name=Hadoop Name=Hive Name=Oozie Name=Hue Name=HCatalog Name=Spark \
--kerberos-attributes Realm=EC2.INTERNAL,\
KdcAdminPassword=MyClusterKDCAdminPwd,\
ADDomainJoinUser=ADUserLogonName,ADDomainJoinPassword=ADUserPassword,\
CrossRealmTrustPrincipalPassword=MatchADTrustPwd
```

Etapa 7: criar HDFS usuários e definir permissões no cluster para contas do Active Directory

Ao configurar uma relação de confiança com o Active Directory, a Amazon EMR cria usuários Linux no cluster para cada conta do Active Directory. Por exemplo, o nome de logon de usuário `LiJuan` no Active Directory tem uma conta do Linux de `lijuan`. Os nomes de usuário do Active Directory podem conter letras maiúsculas, mas o Linux não segue o uso de maiúsculas e minúsculas do Active Directory.

Para permitir que seus usuários façam login no cluster para executar trabalhos do Hadoop, você deve adicionar diretórios de HDFS usuários às contas Linux e conceder a cada usuário a propriedade do diretório. Para isso, recomendamos executar um script salvo no Amazon S3 como uma etapa de cluster. Você também pode executar os comandos no script abaixo da linha de comando no nó primário. Use o par de EC2 chaves que você especificou ao criar o cluster para se conectar ao nó primário SSH como usuário do Hadoop. Para obter mais informações, consulte Use um par de EC2 chaves para SSH credenciais.

Execute o comando a seguir para adicionar uma etapa ao cluster que executa um script, *AddHDFSUsers.sh*.

```
aws emr add-steps --cluster-id <j-2AL4XXXXXX5T9> \
```

```
--steps Type=CUSTOM_JAR,Name=CustomJAR,ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://DOC-EXAMPLE-BUCKET/AddHDFSUsers.sh"]
```

O conteúdo do arquivo *AddHDFSUsers.sh* é o seguinte.

```
#!/bin/bash
# AddHDFSUsers.sh script

# Initialize an array of user names from AD or Linux users and KDC principals created
 manually on the cluster
ADUSERS=("lijuan" "marymajor" "richardroe" "myusername")

# For each user listed, create an HDFS user directory
# and change ownership to the user

for username in ${ADUSERS[@]}; do
     hdfs dfs -mkdir /user/$username
     hdfs dfs -chown $username:$username /user/$username
done
```

Grupos do Active Directory mapeados para grupos do Hadoop

A Amazon EMR usa o System Security Services Daemon (SSD) para mapear grupos do Active Directory para grupos do Hadoop. Para confirmar mapeamentos de grupos, depois de fazer login no nó primário, conforme descrito em Usando SSH para se conectar a clusters Kerberizados, você poderá usar o comando `hdfs groups` para confirmar que os grupos do Active Directory aos quais sua conta do Active Directory pertence foram mapeados para os grupos do Hadoop para o usuário correspondente do Hadoop no cluster. Você também pode verificar mapeamentos de grupos de outros usuários especificando um ou mais nomes de usuário usando, por exemplo, o comando `hdfs groups lijuan`. Para obter mais informações, consulte grupos no Guia de HDFS Comandos do Apache.

Use o Active Directory ou LDAP servidores para autenticação com a Amazon EMR

Com as EMR versões 6.12.0 e superiores da Amazon, você pode usar o protocolo LDAP over SSL (LDAPS) para iniciar um cluster que se integra nativamente ao seu servidor de identidade corporativo. LDAP(Lightweight Directory Access Protocol) é um protocolo de aplicativo aberto e independente de fornecedor que acessa e mantém dados. LDAPé comumente usado para

autenticação de usuários em servidores de identidade corporativa hospedados em aplicativos como o Active Directory (AD) e o OpenLDAP. Com essa integração nativa, você pode usar seu LDAP servidor para autenticar usuários na AmazonEMR.

Os destaques da EMR LDAP integração com a Amazon incluem:

- A Amazon EMR configura os aplicativos compatíveis para serem autenticados com LDAP autenticação em seu nome.
- A Amazon EMR configura e mantém a segurança dos aplicativos compatíveis com o protocolo Kerberos. Não é necessário inserir nenhum comando ou script.
- Você obtém controle de acesso refinado (FGAC) por meio da autorização do Apache Ranger para bancos de dados e tabelas do Hive Metastore. Consulte Integre a Amazon EMR com o Apache Ranger Para mais informações.
- Quando você precisa de LDAP credenciais para acessar um cluster, você obtém um controle de acesso refinado (FGAC) sobre quem pode acessar seus clusters. EMR SSH

As páginas a seguir fornecem uma visão geral conceitual, pré-requisitos e etapas para iniciar um cluster com a integração EMR com a Amazon. EMR LDAP

Tópicos

Visão geral de LDAP com a Amazon EMR

O Lightweight Directory Access Protocol (LDAP) é um protocolo de software que os administradores de rede usam para gerenciar e controlar o acesso aos dados autenticando usuários na rede de uma empresa. O LDAP protocolo armazena informações em uma estrutura hierárquica de diretórios em árvore. Para obter mais informações, consulte LDAPConceitos básicos em LDAP.com.

Na rede de uma empresa, muitos aplicativos podem usar o LDAP protocolo para autenticar usuários. Com a EMR LDAP integração com a Amazon, EMR os clusters podem usar nativamente o mesmo LDAP protocolo com uma configuração de segurança adicional.

Há duas implementações principais do LDAP protocolo que a Amazon EMR oferece suporte: Active Directory e Open LDAP. Embora outras implementações sejam possíveis, a maioria se encaixa nos mesmos protocolos de autenticação do Active Directory ou do OpenLDAP.

Active Directory (AD)

O Active Directory (AD) é um serviço de diretório da Microsoft para redes de domínio Windows. O AD está incluído na maioria dos sistemas operacionais Windows Server e pode se comunicar com clientes por meio dos LDAPS protocolos LDAP e. Para autenticação, a Amazon EMR tenta vincular o usuário à sua instância do AD com o nome principal do usuário (UPN) como nome e senha distintos. O UPN usa o formato padrão`username@domain_name`.

Aberto LDAP

Open LDAP é uma implementação gratuita e de código aberto do LDAP protocolo. Para autenticação, a Amazon EMR tenta vincular o usuário à sua LDAP instância Open com o nome de domínio totalmente qualificado (FQDN) como nome e senha distintos. O FQDN usa o formato padrão`username_attribute=username,LDAP_user_search_base`. Normalmente, o valor de `username_attribute` é `uid`, e o valor de `LDAP_user_search_base` contém os atributos da árvore que leva ao usuário. Por exemplo, `ou=People,dc=example,dc=com`.

Outras implementações gratuitas e de código aberto do LDAP protocolo geralmente seguem um protocolo semelhante ao FQDN Open LDAP para os nomes distintos de seus usuários.

LDAPcomponentes para Amazon EMR

Você pode usar seu LDAP servidor para se autenticar na Amazon EMR e em qualquer aplicativo que o usuário utilize diretamente no EMR cluster por meio dos seguintes componentes.

Agente secreto

O agente secreto é um processo no cluster que autentica todas as solicitações do usuário. O agente secreto cria o vínculo do usuário ao seu LDAP servidor em nome dos aplicativos suportados no EMR cluster. O agente secreto é executado como o usuário `emrsecretagent` e grava logs no diretório `/emr/secretagent/log`. Esses logs fornecem detalhes sobre o estado da solicitação de autenticação de cada usuário e os erros que possam surgir durante a autenticação do usuário.

Daemon de serviços de segurança do sistema () SSSD

SSSDé um daemon executado em cada nó de um cluster LDAP EMR habilitado. SSSDcria e gerencia um UNIX usuário para sincronizar sua identidade corporativa remota com cada nó.

YARNaplicativos baseados em Hive e Spark exigem que exista um UNIX usuário local em cada nó que executa uma consulta para um usuário.

Suporte de aplicativos e considerações com LDAP a Amazon EMR

Aplicativos compatíveis com LDAP para a Amazon EMR

> **Important**
>
> Os aplicativos listados nesta página são os únicos para os quais a Amazon EMR oferece suporteLDAP. Para garantir a segurança do cluster, você só pode incluir aplicativos LDAP compatíveis ao criar um EMR cluster com LDAP habilitado. Se você tentar instalar outros aplicativos sem suporte, a Amazon EMR rejeitará sua solicitação de um novo cluster.

As EMR versões 6.12 e superiores da Amazon oferecem suporte à LDAP integração com os seguintes aplicativos:

- Apache Livy
- Apache Hive até HiveServer 2 () HS2
- Trino
- Presto
- Hue

Você também pode instalar os seguintes aplicativos em um EMR cluster e configurá-los para atender às suas necessidades de segurança:

- Apache Spark
- Apache Hadoop

Recursos compatíveis com LDAP para a Amazon EMR

Você pode usar os seguintes EMR recursos da Amazon com a LDAP integração:

> **Note**
>
> Para manter LDAP as credenciais seguras, você deve usar criptografia em trânsito para proteger o fluxo de dados dentro e fora do cluster. Para obter mais informações sobre criptografia em trânsito, consulte Criptografar dados em repouso e em trânsito.

- Criptografia em trânsito (obrigatório) e em repouso
- Grupos de instâncias, frotas de instâncias e instâncias spot
- Reconfiguração de aplicações em um cluster em execução
- EMRFScriptografia do lado do servidor () SSE

Atributos não compatíveis

Considere as seguintes limitações ao usar a EMR LDAP integração com a Amazon:

- A Amazon EMR desativa as etapas para clusters com LDAP habilitado.
- A Amazon EMR não oferece suporte a funções e AWS Lake Formation integrações de tempo de execução para clusters LDAP habilitados.
- A Amazon EMR não oferece suporte LDAP com o StartTLS.
- A Amazon EMR não oferece suporte ao modo de alta disponibilidade (clusters com vários nós primários) para clusters com o modo LDAP ativado.
- Você não pode alternar credenciais ou certificados de vinculação para clusters com habilitado. LDAP Se algum desses campos tiver sido alternado, é recomendável iniciar um novo cluster com as credenciais ou certificados de vinculação atualizados.
- Você deve usar bases de pesquisa exatas comLDAP. A base de pesquisa de LDAP usuários e grupos não suporta filtros LDAP de pesquisa.

Configure e execute um EMR cluster com LDAP

Esta seção aborda como configurar a Amazon EMR para uso com LDAP autenticação.

Tópicos

- Adicione AWS Secrets Manager permissões à função de EMR instância da Amazon
- Crie a configuração EMR de segurança da Amazon para LDAP integração

- Inicie um EMR cluster que se autentique com LDAP

Adicione AWS Secrets Manager permissões à função de EMR instância da Amazon

A Amazon EMR usa uma função IAM de serviço para realizar ações em seu nome para provisionar e gerenciar clusters. A função de serviço para EC2 instâncias de cluster, também chamada de perfil de EC2 instância para a Amazon EMR, é um tipo especial de função de serviço que a Amazon EMR atribui a cada EC2 instância em um cluster no lançamento.

Para definir permissões para que um EMR cluster interaja com dados do Amazon S3 e outros AWS serviços, defina um perfil personalizado de EC2 instância da Amazon em vez de `EMR_EC2_DefaultRole` quando você executa seu cluster. Para ter mais informações, consulte Função de serviço para EC2 instâncias de cluster (perfil de EC2 instância) e Personalize IAM funções.

Adicione as seguintes declarações ao perfil de EC2 instância padrão para permitir que EMR a Amazon marque sessões e acesse a AWS Secrets Manager que armazena LDAP certificados.

```
    {
      "Sid": "AllowAssumeOfRolesAndTagging",
      "Effect": "Allow",
      "Action": ["sts:TagSession", "sts:AssumeRole"],
      "Resource": [
        "arn:aws:iam::111122223333:role/LDAP_DATA_ACCESS_ROLE_NAME",
        "arn:aws:iam::111122223333:role/LDAP_USER_ACCESS_ROLE_NAME"
      ]
    },
    {
        "Sid": "AllowSecretsRetrieval",
        "Effect": "Allow",
        "Action": "secretsmanager:GetSecretValue",
        "Resource": [
            "arn:aws:secretsmanager:us-east-1:111122223333:secret:LDAP_SECRET_NAME*",
            "arn:aws:secretsmanager:us-
east-1:111122223333:secret:ADMIN_LDAP_SECRET_NAME*"
        ]
    }
```

> **Note**
>
> Suas solicitações de cluster falharão se você esquecer o caractere curinga * no final do nome do segredo ao definir as permissões do Secrets Manager. O curinga representa as versões do segredo.
> Você também deve limitar o escopo da AWS Secrets Manager política somente aos certificados que seu cluster precisa para provisionar instâncias.

Crie a configuração EMR de segurança da Amazon para LDAP integração

Antes de iniciar um EMR cluster com LDAP integração, use as etapas Criar uma configuração de segurança para criar uma configuração de EMR segurança da Amazon para o cluster. Complete as seguintes configurações no `LDAPConfiguration` bloco abaixo `AuthenticationConfiguration` ou nos campos correspondentes na seção Configurações de segurança do EMR console Amazon:

`EnableLDAPAuthentication`

Opção de console: Protocolo de autenticação: LDAP

Para usar a LDAP integração, defina essa opção `true` ou selecione-a como seu protocolo de autenticação ao criar um cluster no console. Por padrão, `EnableLDAPAuthentication` é `true` quando você cria uma configuração de segurança no EMR console da Amazon.

`LDAPServerURL`

Opção de console: localização LDAP do servidor

A localização do LDAP servidor, incluindo o prefixo:`ldaps://`*`location_of_server`*.

`BindCertificateARN`

Opção de console: LDAPSSLcertificado

O AWS Secrets Manager ARN que contém o certificado para assinar o SSL certificado que o LDAP servidor usa. Se seu LDAP servidor for assinado por uma Autoridade Certificadora (CA) pública, você poderá fornecer um AWS Secrets Manager ARN com um arquivo em branco. Para obter mais informações sobre como armazenar seu certificado no Secrets Manager, consulte Armazene TLS certificados em AWS Secrets Manager.

`BindCredentialsARN`

Opção de console: LDAPcredenciais de vinculação ao servidor

E AWS Secrets Manager ARN que contém as credenciais vinculadas do usuário LDAP administrador. As credenciais são armazenadas como um JSON objeto. Há somente um par de chave-valor nesse segredo; a chave no par é o nome de usuário e o valor é a senha. Por exemplo, `{"uid=admin,cn=People,dc=example,dc=com": "AdminPassword1"}`. Esse é um campo opcional, a menos que você habilite o SSH login para seu EMR cluster. Em muitas configurações, as instâncias do Active Directory exigem credenciais de associação para permitir SSSD a sincronização dos usuários.

`LDAPAccessFilter`

Opção de console: filtro de LDAP acesso

Especifica o subconjunto de objetos em seu LDAP servidor que podem ser autenticados. Por exemplo, se você quiser conceder acesso a todos os usuários com a classe de `posixAccount` objeto em seu LDAP servidor, defina o filtro de acesso como`(objectClass=posixAccount)`.

`LDAPUserSearchBase`

Opção de console: base LDAP de pesquisa de usuários

A base de pesquisa à qual seus usuários pertencem em seu LDAP servidor. Por exemplo, `cn=People,dc=example,dc=com`.

`LDAPGroupSearchBase`

Opção de console: base LDAP de pesquisa de grupos

A base de pesquisa à qual seus grupos pertencem em seu LDAP servidor. Por exemplo, `cn=Groups,dc=example,dc=com`.

`EnableSSHLogin`

Opção de console: SSHlogin

Especifica se a autenticação por senha com LDAP credenciais deve ou não ser permitida. Não é recomendável habilitar essa opção. Os pares de chaves são uma rota mais segura para permitir o acesso aos EMR clusters. Esse campo é opcional e usa o padrão `false`.

`LDAPServerType`

Opção de console: tipo de LDAP servidor

Especifica o tipo de LDAP servidor ao qual a Amazon EMR se conecta. As opções suportadas são Active Directory e OpenLDAP. Outros tipos de LDAP servidor podem funcionar, mas a Amazon EMR não oferece suporte oficial a outros tipos de servidor. Para obter mais informações, consulte LDAPcomponentes para Amazon EMR.

`ActiveDirectoryConfigurations`

Um sub-bloco necessário para configurações de segurança que utilizam o tipo de servidor Active Directory.

`ADDomain`

Opção do console: domínio do Active Directory

O nome de domínio usado para criar o Nome Principal do Usuário (UPN) para autenticação do usuário com configurações de segurança que usam o tipo de servidor Active Directory.

Considerações sobre configurações de segurança com LDAP a Amazon EMR

- Para criar uma configuração de segurança com a EMR LDAP integração com a Amazon, você deve usar criptografia em trânsito. Para obter informações sobre criptografia em trânsito, consulte Criptografar dados em repouso e em trânsito.
- Não é possível definir a configuração do Kerberos na mesma configuração de segurança. A Amazon EMR provisiona um KDC thar dedicado automaticamente e gerencia a senha do administrador para issoKDC. Os usuários não poderão acessar essa senha de administrador.
- Você não pode definir funções IAM de tempo de execução e AWS Lake Formation na mesma configuração de segurança.
- `LDAPServerURL` deve ter o protocolo `ldaps://` em seu valor.
- `LDAPAccessFilter` não pode estar vazio.

Use LDAP com a integração do Apache Ranger para a Amazon EMR

Com a LDAP integração com a AmazonEMR, você pode se integrar ainda mais com o Apache Ranger. Ao inserir .your LDAP users no Ranger, você pode então associar esses usuários a um servidor de políticas Apache Ranger para integração com a Amazon EMR e outros aplicativos. Para fazer isso, defina o `RangerConfiguration` campo dentro `AuthorizationConfiguration` da configuração de segurança que você usa com seu LDAP cluster. Para obter mais informações sobre como definir a configuração de segurança, consulte Crie a configuração EMR de segurança.

Ao usar LDAP com a AmazonEMR, você não precisa fornecer uma EMR integração `KerberosConfiguration` com a Amazon para o Apache Ranger.

Inicie um EMR cluster que se autentique com LDAP

Use as etapas a seguir para iniciar um EMR cluster com o LDAP Active Directory.

1. Configure o ambiente:

   - Certifique-se de que os nós em seu EMR cluster possam se comunicar com o Amazon S3 e. AWS Secrets Manager Para obter mais informações sobre como modificar sua função de perfil de EC2 instância para se comunicar com esses serviços, consulteAdicione AWS Secrets Manager permissões à função de EMR instância da Amazon.
   - Se você planeja executar seu EMR cluster em uma sub-rede privada, você deve usar VPC endpoints da Amazon ou usar AWS PrivateLink a tradução de endereço de rede (NAT) para configurá-lo para se comunicar com o S3 e o VPC Secrets Manager. Para obter mais informações, consulte VPCendpoints AWS PrivateLink e NAT instâncias no Amazon VPC Getting Started Guide.
   - Verifique se há conectividade de rede entre seu EMR cluster e o LDAP servidor. Seus EMR clusters devem acessar seu LDAP servidor pela rede. Os nós primário, principal e de tarefas do cluster se comunicam com o LDAP servidor para sincronizar os dados do usuário. Se o seu LDAP servidor for executado na AmazonEC2, atualize o grupo de EC2 segurança para aceitar o tráfego do EMR cluster. Para obter mais informações, consulte Adicione AWS Secrets Manager permissões à função de EMR instância da Amazon.

2. Crie uma configuração EMR de segurança da Amazon para a LDAP integração. Para obter mais informações, consulte Crie a configuração EMR de segurança da Amazon para LDAP integração.
3. Agora que você está configurado, use as etapas descritas em Inicie um EMR cluster da Amazon para iniciar o cluster com as seguintes configurações:

   - Selecione a EMR versão 6.12 ou superior da Amazon. Recomendamos que você use a EMR versão mais recente da Amazon.
   - Somente especifique ou selecione aplicativos compatíveis com seu clusterLDAP. Para obter uma lista LDAP de aplicativos compatíveis com a AmazonEMR, consulteSuporte de aplicativos e considerações com LDAP a Amazon EMR.
   - Aplique a configuração de segurança criada na etapa anterior.

Exemplos de uso LDAP com a Amazon EMR

Depois de provisionar um EMR cluster que usa LDAP integração, você pode fornecer suas LDAP credenciais para qualquer aplicativo compatível por meio do mecanismo de autenticação de nome de usuário e senha incorporado. Esta página mostra alguns exemplos.

Usando a LDAP autenticação com o Apache Hive

Example - Apache Hive

O comando de exemplo a seguir inicia uma sessão do Apache Hive por meio de HiveServer 2 e Beeline:

```
beeline -u "jdbc:hive2://$HOSTNAME:10000/default;ssl=true;sslTrustStore=
$TRUSTSTORE_PATH;trustStorePassword=$TRUSTSTORE_PASS"  -n LDAP_USERNAME -
p LDAP_PASSWORD
```

Usando a LDAP autenticação com o Apache Livy

Example - Apache Livy

O comando de exemplo a seguir inicia uma sessão do Livy por meio de c. URL Substitua *ENCODED-KEYPAIR* com uma string codificada em Base64 por `username:password`.

```
curl -X POST --data '{"proxyUser":"LDAP_USERNAME","kind": "pyspark"}' -H "Content-Type:
 application/json" -H "Authorization: Basic ENCODED-KEYPAIR" DNS_OF_PRIMARY_NODE:8998/
sessions
```

Usando a LDAP autenticação com o Presto

Example - Presto

O comando de exemplo a seguir inicia uma sessão do Presto por meio do CLI Presto:

```
presto-cli --user "LDAP_USERNAME" --password --catalog hive
```

Depois de executar esse comando, digite a LDAP senha no prompt.

Usando a LDAP autenticação com Trino

Example - Trino

O comando de exemplo a seguir inicia uma sessão do Trino por meio do CLI Trino:

```
trino-cli --user "LDAP_USERNAME" --password --catalog hive
```

Depois de executar esse comando, digite a LDAP senha no prompt.

Usando a LDAP autenticação com o Hue

Você pode acessar a interface do usuário do Hue por meio de um SSH túnel criado no cluster ou pode configurar um servidor proxy para transmitir publicamente a conexão com o Hue. Como o Hue não é executado no HTTPS modo por padrão, recomendamos que você use uma camada de criptografia adicional para garantir que a comunicação entre os clientes e a interface do usuário do Hue seja criptografada com. HTTPS Isso reduz a chance de expor acidentalmente as credenciais do usuário em texto sem formatação.

Para usar a interface do usuário do Hue, abra a interface do usuário do Hue no seu navegador e digite seu LDAP nome de usuário e senha para fazer login. Se as credenciais estiverem corretas, o Hue fará login e usará sua identidade para autenticar você em todas as aplicações compatíveis.

Usando SSH para autenticação por senha e tíquetes Kerberos para outros aplicativos

> **Important**
>
> Não recomendamos que você use a autenticação por senha para SSH entrar em um EMR cluster.

Você pode usar suas LDAP credenciais SSH para um EMR cluster. Para fazer isso, defina a `EnableSSHLogin` configuração como `true` na configuração de EMR segurança da Amazon que você usa para iniciar o cluster. Em seguida, use o comando a seguir SSH para acessar o cluster depois que ele for iniciado:

```
ssh username@EMR_PRIMARY_DNS_NAME
```

Depois de executar esse comando, digite a LDAP senha no prompt.

A Amazon EMR inclui um script no cluster que permite aos usuários gerar um arquivo keytab Kerberos e um ticket para usar com aplicativos compatíveis que não aceitam credenciais diretamente. LDAP Alguns desses aplicativos incluem`spark-submit`, Spark SQL e. PySpark

Execute `ldap-kinit` e siga as instruções. Se a autenticação tiver êxito, o arquivo keytab do Kerberos será exibido no diretório inicial com um tíquete do Kerberos válido. Use o tíquete do Kerberos para executar aplicações como você faria em qualquer ambiente kerberizado.

Integre a Amazon EMR com AWS IAM Identity Center

Com as EMR versões 6.15.0 e superiores da Amazon, você pode usar identidades de AWS IAM Identity Center para se autenticar em um cluster da Amazon. EMR As seções a seguir fornecem uma visão geral conceitual, os pré-requisitos e as etapas necessárias para iniciar um EMR cluster com a integração do Identity Center.

Tópicos

Visão geral

A propagação confiável de IAM identidades por meio do Identity Center pode ajudá-lo a criar ou conectar com segurança suas identidades de força de trabalho e gerenciar centralmente o acesso entre contas e aplicativos. AWS Com esse recurso, um usuário pode entrar no aplicativo que usa propagação de identidade confiável e esse aplicativo pode transmitir a identidade do usuário nas solicitações que ele faz para acessar dados em AWS serviços que também usam propagação de identidade confiável. Como o acesso é gerenciado com base na identidade do usuário, os usuários não precisam usar as credenciais do usuário local do banco de dados nem assumir uma IAM função para acessar os dados.

O Identity Center é a abordagem recomendada para autenticação e autorização da força de trabalho em AWS organizações de qualquer tamanho e tipo. Com o Identity Center, você pode criar e gerenciar identidades de usuários ou conectar sua fonte de identidade existente, incluindo Microsoft Active Directory, Okta, Ping Identity JumpCloud, Google Workspace e Microsoft Entra ID (antigo Azure AD). AWS

Para obter mais informações, consulte O que é AWS IAM Identity Center? e propagação confiável de identidade entre aplicativos no Guia do AWS IAM Identity Center usuário.

Atributos e benefícios

A EMR integração da Amazon com o IAM Identity Center oferece os seguintes benefícios:

- EMRA Amazon fornece credenciais para retransmitir sua identidade do Identity Center para um EMR cluster.
- A Amazon EMR configura todos os aplicativos compatíveis para serem autenticados com as credenciais do cluster.
- A Amazon EMR configura e mantém a segurança do aplicativo compatível com o protocolo Kerberos e sem comandos ou scripts exigidos por você.
- A capacidade de aplicar a autorização no nível de prefixo do Amazon S3 com as identidades do Centro de Identidade em prefixos do S3 gerenciados pelo S3 Access Grants.
- A capacidade de aplicar a autorização em nível de tabela com identidades do Identity Center em tabelas Glue AWS Lake Formation gerenciadas. AWS

Começando com a AWS IAM Identity Center integração com a Amazon EMR

Esta seção ajuda você a configurar EMR a Amazon para integração com AWS IAM Identity Center.

Tópicos

Criação de uma instância do Centro de Identidade

Se você ainda não tiver uma, crie uma instância do Identity Center no Região da AWS local em que você deseja iniciar seu EMR cluster. Uma instância do Centro de Identidade só pode existir em uma única região para uma Conta da AWS.

Use o AWS CLI comando a seguir para criar uma nova instância chamada`MyInstance`:

```
aws sso-admin create-instance --name MyInstance
```

Crie uma IAM função para o Identity Center

Para integrar a Amazon EMR AWS IAM Identity Center, crie uma IAM função que se autentique com o Identity Center a partir do EMR cluster. Nos bastidores, a Amazon EMR usa SigV4 credenciais para retransmitir a identidade do Identity Center para serviços posteriores, como. AWS Lake Formation Seu perfil também deve ter as respectivas permissões para invocar os serviços downstream.

Ao criar o perfil, use a seguinte política de permissões:

```
{
  "Statement": [
    {
      "Sid": "IdCPermissions",
      "Effect": "Allow",
      "Action": [
        "sso-oauth:*"
      ],
      "Resource": "*"
    },
    {
      "Sid": "GlueandLakePermissions",
      "Effect": "Allow",
      "Action": [
        "glue:*",
        "lakeformation:GetDataAccess"
      ],
      "Resource": "*"
    },
    {
      "Sid": "AccessGrantsPermissions",
      "Effect": "Allow",
      "Action": [
        "s3:GetDataAccess",
        "s3:GetAccessGrantsInstanceForPrefix"
      ],
      "Resource": "*"
    }
  ]
}
```

A política de confiança desse perfil permite que o perfil InstanceProfile deixe-o assumir o perfil.

```
{
    "Sid": "AssumeRole",
    "Effect": "Allow",
    "Principal": {
        "AWS": "arn:aws:iam::12345678912:role/EMR_EC2_DefaultRole"
    },
    "Action": [
        "sts:AssumeRole",
        "sts:SetContext"
    ]
}
```

Se a função não tiver credenciais confiáveis e acessar uma tabela protegida por Lake Formation, a EMR Amazon definirá automaticamente `principalId` a função assumida como. *`userID`*-`untrusted` A seguir está um trecho de um CloudTrail evento que exibe o. `principalId`

```
{
    "eventVersion": "1.09",
    "userIdentity": {
        "type": "AssumedRole",
        "principalId": "ABCDEFGH1JKLMNO2PQR3TU:5000-untrusted",
        "arn": "arn:aws:sts::123456789012:assumed-role/EMR_TIP/5000-untrusted",
        "accountId": "123456789012",
        "accessKeyId": "ABCDEFGH1IJKLMNOPQ7R3"
        ...
```

Criação de uma configuração de segurança habilitada para o Centro de Identidade

Para iniciar um EMR cluster com integração com o IAM Identity Center, use o comando de exemplo a seguir para criar uma configuração de EMR segurança da Amazon que tenha o Identity Center ativado. Cada configuração é explicada abaixo.

```
aws emr create-security-configuration --name "IdentityCenterConfiguration-with-lf-
accessgrants" --region "us-west-2" --security-configuration '{
 "AuthenticationConfiguration":{
  "IdentityCenterConfiguration":{
   "EnableIdentityCenter":true,
   "IdentityCenterApplicationAssigmentRequired":false,
   "IdentityCenterInstanceARN": "arn:aws:sso:::instance/ssoins-123xxxxxxxxx789",
```

```
   "IAMRoleForEMRIdentityCenterApplicationARN": "arn:aws:iam::123456789012:role/tip-
role"
     }
 },
 "AuthorizationConfiguration": {
  "LakeFormationConfiguration": {
   "EnableLakeFormation": true
  }
 },
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": true,
  "EnableAtRestEncryption": false,
  "InTransitEncryptionConfiguration": {
   "TLSCertificateConfiguration": {
    "CertificateProviderType": "PEM",
    "S3Object": "s3://my-bucket/cert/my-certs.zip"
   }
  }
 }
}'
```

- **`EnableIdentityCenter`**: (obrigatório) habilita a integração do Centro de Identidade.
- **`IdentityCenterApplicationARN`**— (obrigatório) A instância do Identity CenterARN.
- **`IAMRoleForEMRIdentityCenterApplicationARN`**— (obrigatório) A IAM função que adquire os tokens do Identity Center do cluster.
- **`IdentityCenterApplicationAssignmentRequired`** : (booleano) determina se uma atribuição será necessária para usar a aplicação do Centro de Identidade. O valor padrão é `true`.
- **`AuthorizationConfiguration`**/**`LakeFormationConfiguration`**— Opcionalmente, configure a autorização:
  - **`EnableLakeFormation`**: habilite a autorização do Lake Formation no cluster.

Para habilitar a integração do Identity Center com a AmazonEMR, você deve especificar `EncryptionConfiguration` `IntransitEncryptionConfiguration` e.

Criação e execução de um cluster habilitado para o Centro de Identidade

Agora que você configurou a IAM função que se autentica com o Identity Center e criou uma configuração de EMR segurança da Amazon com o Identity Center ativado, você pode criar e iniciar seu cluster com reconhecimento de identidade. Para ver as etapas de execução do cluster com a

configuração de segurança necessária, consulte Especificar uma configuração de segurança para um cluster.

Opcionalmente, consulte a seção a seguir se quiser usar seu cluster habilitado para o Identity Center com outras opções de segurança suportadas pela AmazonEMR:

- Trabalhando com o S3 Access Grants em um cluster habilitado EMR para o IAM Identity Center
- Configurar o Lake Formation para um EMR cluster habilitado para o IAM Identity Center

Configurar o Lake Formation para um EMR cluster habilitado para o IAM Identity Center

Você pode se integrar AWS Lake Formationao seu EMR cluster AWS IAM Identity Center habilitado.

Primeiro, certifique-se de ter uma instância do Centro de Identidade configurada na mesma região do cluster. Para obter mais informações, consulte Criação de uma instância do Centro de Identidade. Você pode encontrar a instância ARN no console do IAM Identity Center ao visualizar os detalhes da instância ou usar o comando a seguir para ver os detalhes de todas as suas instâncias noCLI:

```
aws sso-admin list-instances
```

Em seguida, use o ARN e o ID da sua AWS conta com o comando a seguir para configurar o Lake Formation para ser compatível com o IAM Identity Center:

```
aws lakeformation create-lake-formation-identity-center-configuration --cli-input-json
 file://create-lake-fromation-idc-config.json
json input:
{
    "CatalogId": "account-id/org-account-id",
    "InstanceArn": "identity-center-instance-arn"
}
```

Agora, chame `put-data-lake-settings` e habilite `AllowFullTableExternalDataAccess` com o Lake Formation:

```
aws lakeformation put-data-lake-settings --cli-input-json file://put-data-lake-
settings.json
json input:
```

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier": "admin-ARN"
            }
        ],
        "CreateDatabaseDefaultPermissions": [...],
        "CreateTableDefaultPermissions": [...],
        "AllowExternalDataFiltering": true,
        "AllowFullTableExternalDataAccess": true
    }
}
```

Por fim, conceda permissões completas de tabela ARN à identidade do usuário que acessa o EMR cluster. O ARN contém a ID do usuário do Identity Center. Navegue até o Centro de Identidade no console, selecione Usuários e, em seguida, o usuário para visualizar as configurações de Informações gerais.

Copie o ID do usuário e cole-o no seguinte ARN para*user-id*:

```
arn:aws:identitystore:::user/user-id
```

> **Note**
>
> As consultas no EMR cluster só funcionam se a IAM identidade do Identity Center tiver acesso total à tabela protegida do Lake Formation. Se a identidade não tiver acesso total à tabela, a consulta falhará.

Use o seguinte comando para conceder ao usuário acesso total à tabela:

```
aws lakeformation grant-permissions --cli-input-json file://grantpermissions.json
json input:
{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:identitystore:::user/user-id"
    },
    "Resource": {
        "Table": {
```

```
            "DatabaseName": "tip_db",
            "Name": "tip_table"
        }
    },
    "Permissions": [
        "ALL"
    ],
    "PermissionsWithGrantOption": [
        "ALL"
    ]
}
```

Trabalhando com o S3 Access Grants em um cluster habilitado EMR para o IAM Identity Center

Você pode integrar o S3 Access Grants ao seu EMR cluster AWS IAM Identity Center habilitado.

Use o S3 Access Grants para autorizar o acesso aos seus conjuntos de dados de clusters que usam o Centro de Identidade. Crie concessões para aumentar as permissões que você define para IAM usuários, grupos, funções ou para um diretório corporativo. Para obter mais informações, consulte Usando o S3 Access Grants com a Amazon EMR.

Tópicos

- Como criar uma instância e localização da funcionalidade S3 Access Grants
- Como criar concessões para identidades do Centro de Identidade

Como criar uma instância e localização da funcionalidade S3 Access Grants

Se você ainda não tiver uma, crie uma instância do S3 Access Grants no Região da AWS local em que você deseja iniciar seu EMR cluster.

Use o AWS CLI comando a seguir para criar uma nova instância chamada*MyInstance*:

```
aws s3control-access-grants create-access-grants-instance \
--account-id 12345678912 \
--identity-center-arn "identity-center-instance-arn" \
```

Em seguida, crie uma localização do S3 Access Grants, substituindo os valores vermelhos pelos seus próprios:

```
aws s3control-access-grants create-access-grants-location \
--account-id 12345678912 \
--location-scope s3:// \
--iam-role-arn "access-grant-role-arn" \
--region aa-example-1
```

> **Note**
>
> Defina o `iam-role-arn` parâmetro como `accessGrantRole` ARN o.

Como criar concessões para identidades do Centro de Identidade

Por fim, crie as concessões das identidades que têm acesso ao seu cluster:

```
aws s3control-access-grants create-access-grant \
--account-id 12345678912 \
--access-grants-location-id "default" \
--access-grants-location-configuration S3SubPrefix="s3-bucket-prefix"
--permission READ \
--grantee GranteeType=DIRECTORY_USER,GranteeIdentifier="your-identity-center-user-id"
```

Exemplo de saída:

```
{
"CreatedAt": "2023-09-21T23:47:24.870000+00:00",
"AccessGrantId": "1234-12345-1234-1234567",
"AccessGrantArn": "arn:aws:s3:aa-example-1-1:123456789012:access-grants/default/grant/
xxxx1234-1234-5678-1234-1234567890",
"Grantee": {
"GranteeType": "DIRECTORY_USER",
"GranteeIdentifier": "5678-56789-5678-567890"
},
"AccessGrantsLocationId": "default",
"AccessGrantsLocationConfiguration": {
"S3SubPrefix": "myprefix/*"
},
"Permission": "READ",
"GrantScope": "s3://myprefix/*"
}
```

Considerações e limitações para a Amazon EMR com a integração do Identity Center

Considere os seguintes pontos ao usar o IAM Identity Center com a AmazonEMR:

- A propagação de identidade confiável por meio do Identity Center é suportada no Amazon EMR 6.15.0 e superior, e somente com o Apache Spark.
- Para habilitar EMR clusters com propagação de identidade confiável, você deve usar o AWS CLI para criar uma configuração de segurança que tenha a propagação de identidade confiável ativada e usar essa configuração de segurança ao iniciar seu cluster. Para obter mais informações, consulte Criação de uma configuração de segurança habilitada para o Centro de Identidade.
- EMRclusters que usam propagação de identidade confiável só podem invocar serviços que também usam propagação de identidade confiável.
- Somente o controle de acesso em nível de tabela baseado em AWS Lake Formation está disponível para EMR clusters que usam propagação de identidade confiável.
- Com EMR clusters que usam propagação de identidade confiável, as operações que oferecem suporte ao controle de acesso com base no Lake Formation com o Apache Spark incluem`SELECT`, e. `ALTER TABLE` `DROP TABLE`
- Com EMR clusters que usam propagação de identidade confiável, os controles de acesso baseados em Lake Formation que não são compatíveis com o Apache Spark incluem instruções. `INSERT`
- A propagação de identidade confiável com a Amazon EMR é suportada no seguinte: Regiões da AWS
  - `ap-east-1`: Ásia-Pacífico (Hong Kong)
  - `ap-northeast-1`: Ásia-Pacífico (Tóquio)
  - `ap-northeast-2`: Ásia-Pacífico (Seul)
  - `ap-south-1`: Ásia-Pacífico (Mumbai)
  - `ap-southeast-1`: Ásia-Pacífico (Singapura)
  - `ap-southeast-2`: Ásia-Pacífico (Sydney)
  - `ca-central-1`: Canadá (Central)
  - `eu-central-1`: Europa (Frankfurt)
  - `eu-north-1`: Europa (Estocolmo)
  - `eu-west-1`: Europa (Irlanda)

- `eu-west-2`: Europa (Londres)
- `eu-west-3`: Europa (Paris)
- `me-south-1`: Oriente Médio (Bahrein)
- `sa-east-1`: América do Sul (São Paulo)
- `us-east-1`: Leste dos EUA (Norte da Virgínia)
- `us-east-2`: Leste dos EUA (Ohio)
- `us-west-1`: Oeste dos EUA (Norte da Califórnia)
- `us-west-2`: Oeste dos EUA (Oregon)

Integre a Amazon EMR com AWS Lake Formation

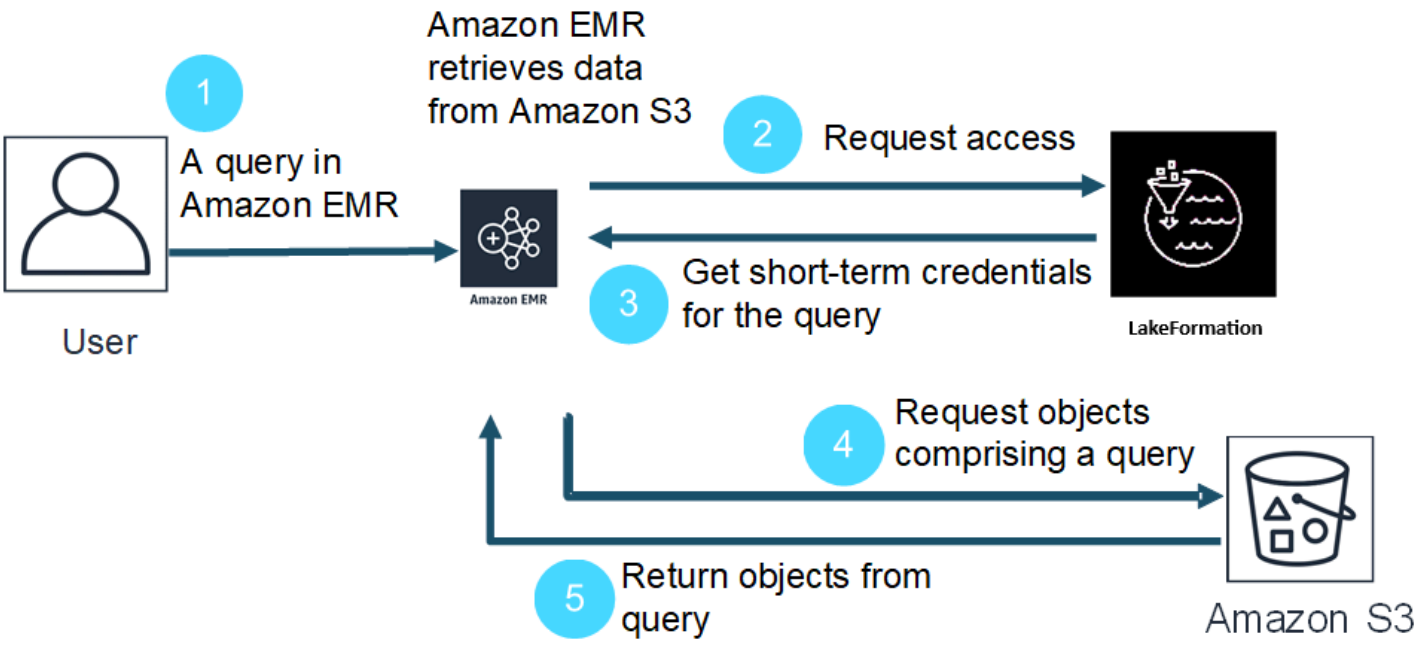
AWS Lake Formation é um serviço gerenciado que ajuda você a descobrir, catalogar, limpar e proteger dados em um data lake do Amazon Simple Storage Service (S3). O Lake Formation fornece acesso refinado em nível de coluna a bancos de dados e tabelas no Glue Data Catalog. AWS Para ter mais informações, consulte O que é o AWS Lake Formation?

Com a EMR versão 6.7.0 e posterior da Amazon, você pode aplicar o controle de acesso baseado em Lake Formation às tarefas do Spark, Hive e Presto que você envia aos clusters da Amazon. EMR Para se integrar ao Lake Formation, você deve criar um EMR cluster com uma função de tempo de execução. Uma função de tempo de execução é uma função AWS Identity and Access Management (IAM) que você associa a EMR trabalhos ou consultas da Amazon. A Amazon EMR então usa essa função para acessar AWS recursos. Para obter mais informações, consulte Funções de tempo de execução para Amazon EMR Steps.

Como a Amazon EMR trabalha com a Lake Formation

Depois de integrar a Amazon EMR com o Lake Formation, você pode executar consultas nos EMR clusters da Amazon com o `StepAPI`ou com o SageMaker Studio. Em seguida, o Lake Formation fornece acesso aos dados por meio de credenciais temporárias para a AmazonEMR. Esse processo chamado de fornecimento de credenciais. Para ter mais informações, consulte O que é o AWS Lake Formation?

A seguir, uma visão geral de alto nível de como a Amazon EMR obtém acesso aos dados protegidos pelas políticas de segurança do Lake Formation.



1. Um usuário envia uma EMR consulta à Amazon para obter dados no Lake Formation.
2. A Amazon EMR solicita credenciais temporárias do Lake Formation para dar acesso aos dados do usuário.
3. O Lake Formation retorna credenciais temporárias.
4. EMRA Amazon envia a solicitação de consulta para recuperar dados do Amazon S3.
5. A Amazon EMR recebe os dados do Amazon S3, os filtra e retorna os resultados com base nas permissões de usuário que o usuário definiu no Lake Formation.

Para obter mais informações sobre como adicionar usuários e grupos às políticas do Lake Formation, consulte Granting Data Catalog permissions.

Pré-requisitos

Você deve atender aos seguintes requisitos antes de integrar a Amazon EMR e a Lake Formation:

- Ative a autorização da função de tempo de execução no seu EMR cluster da Amazon.
- Use o AWS Glue Data Catalog como seu armazenamento de metadados.
- Defina e gerencie permissões no Lake Formation para acessar bancos de dados, tabelas e colunas no AWS Glue Data Catalog. Para ter mais informações, consulte O que é o AWS Lake Formation?

Tópicos

Habilite o Lake Formation com a Amazon EMR

Com o Amazon EMR 6.15.0 e superior, quando você executa trabalhos do Spark na Amazon EMR em EC2 clusters que acessam dados no AWS Glue Data Catalog, você pode usar AWS Lake Formation para aplicar permissões em nível de tabela, linha, coluna e célula em tabelas baseadas em Hudi, Iceberg ou Delta Lake.

Nesta seção, abordamos como criar uma configuração de segurança e configurar o Lake Formation para trabalhar com a AmazonEMR. Também veremos como iniciar um cluster com a configuração de segurança criada para o Lake Formation.

Etapa 1: configurar uma função de tempo de execução para seu EMR cluster

Para usar uma função de tempo de execução para seu EMR cluster, você deve criar uma configuração de segurança. Com uma configuração de segurança, você pode aplicar opções consistentes de segurança, autorização e autenticação nos clusters.

1. Crie um arquivo chamado `lf-runtime-roles-sec-cfg.json` com a configuração a seguir.

```
{
    "AuthorizationConfiguration": {
        "IAMConfiguration": {
            "EnableApplicationScopedIAMRole": true,
            "ApplicationScopedIAMRoleConfiguration": {
                "PropagateSourceIdentity": true
            }
        },
        "LakeFormationConfiguration": {
            "AuthorizedSessionTagValue": "Amazon EMR"
        }
    },
```

```
    "EncryptionConfiguration": {
            "EnableInTransitEncryption": true,
            "InTransitEncryptionConfiguration": {
            "TLSCertificateConfiguration": {<certificate-configuration>}
        }
    }
}
```

2. Em seguida, para garantir que a etiqueta da sessão possa autorizar o Lake Formation, defina a propriedade `LakeFormationConfiguration/AuthorizedSessionTagValue` como `Amazon EMR`.
3. Use o comando a seguir para criar a configuração EMR de segurança da Amazon.

```
aws emr create-security-configuration \
--name 'iamconfig-with-iam-lf' \
--security-configuration file://lf-runtime-roles-sec-cfg.json
```

Como alternativa, você pode usar o EMRconsole da Amazon para criar uma configuração de segurança com configurações personalizadas.

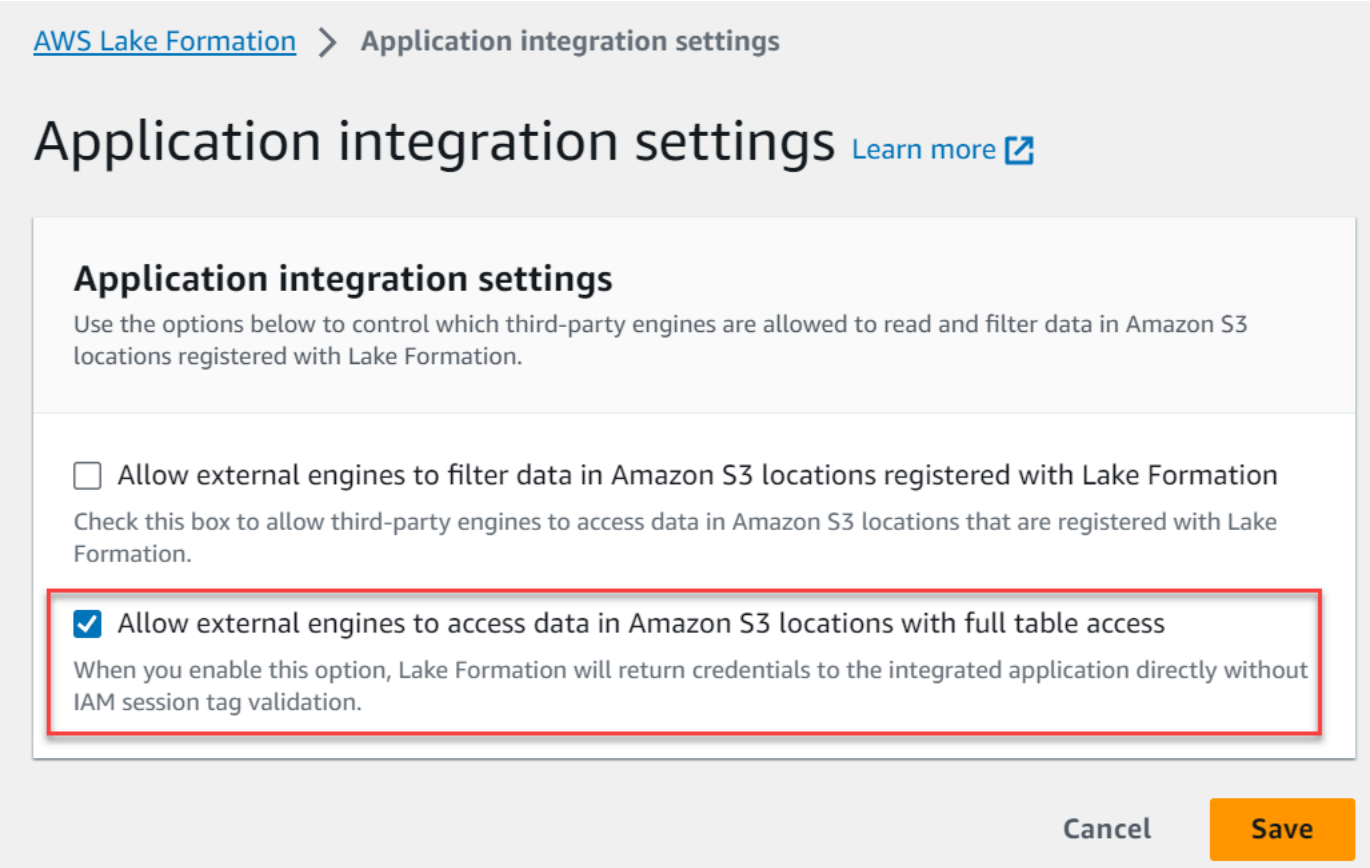
Etapa 2: iniciar um EMR cluster da Amazon

Agora você está pronto para iniciar um EMR cluster com a configuração de segurança criada na etapa anterior. Para obter mais informações sobre configurações de segurança, consulte Usar configurações de segurança para definir a segurança do cluster e Funções de tempo de execução para Amazon EMR Steps.

Etapa 3a: Configurar permissões em nível de tabela baseadas no Lake Formation com funções de tempo de execução da Amazon EMR

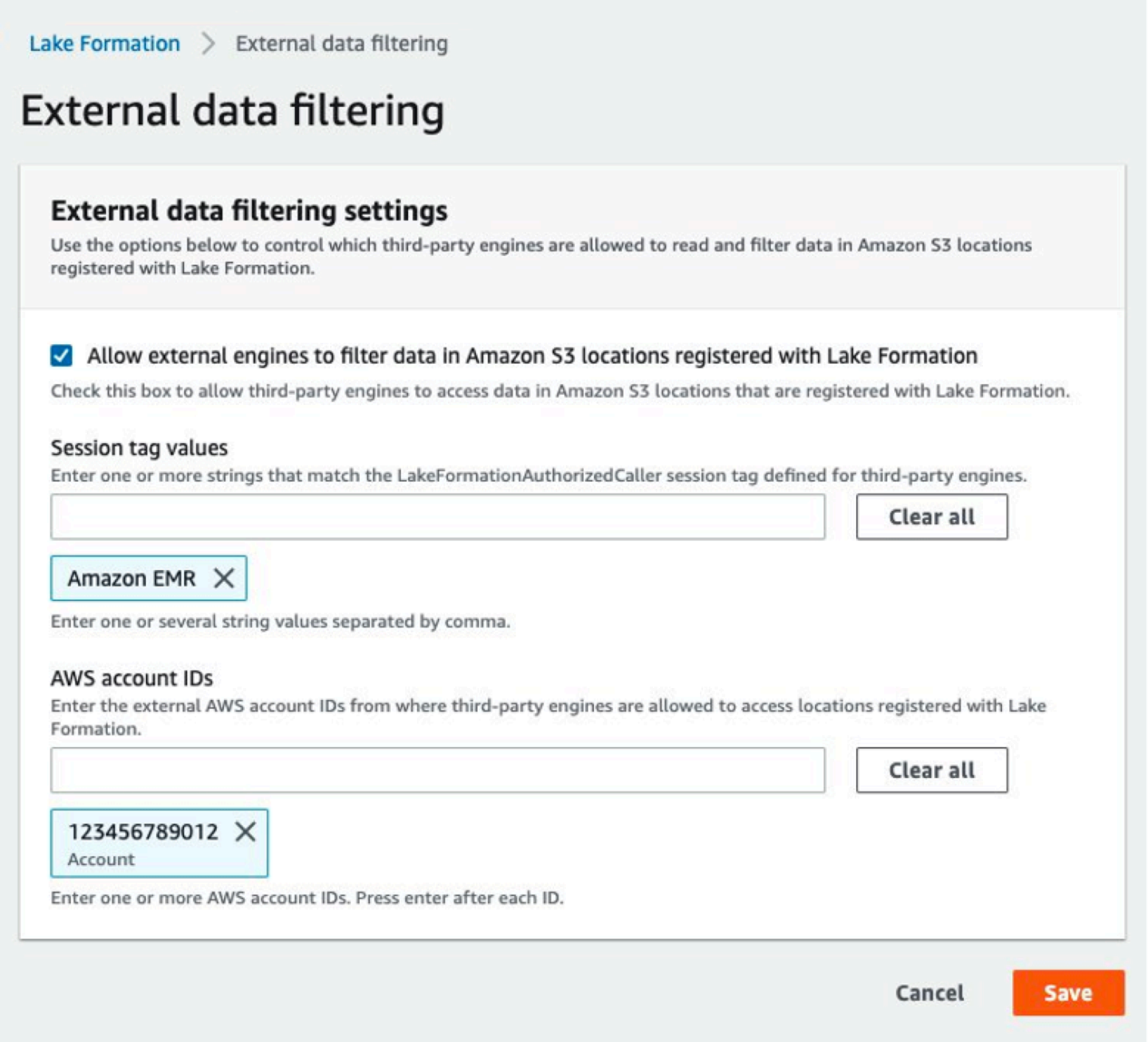
Se você não precisar de um controle de acesso refinado no nível de coluna, linha ou célula, poderá configurar permissões no nível de tabela com o Glue Data Catalog. Para habilitar o acesso em nível de tabela, navegue até o AWS Lake Formation console e selecione a opção Configurações de integração de aplicativos na seção Administração na barra lateral. Em seguida, habilite a seguinte opção e escolha Salvar:

Permitir que mecanismos externos acessem dados em locais do Amazon S3 com acesso total à tabela



Etapa 3b: Configurar permissões em nível de coluna, linha ou célula baseadas no Lake Formation com funções de tempo de execução da Amazon EMR

Para aplicar permissões no nível de tabela e coluna com o Lake Formation, o administrador do data lake no Lake Formation deve definir o `Amazon EMR` como o valor da configuração da tag de sessão, `AuthorizedSessionTagValue`. O Lake Formation usa essa etiqueta de sessão para autorizar os chamadores e fornecer acesso ao data lake. Você pode definir essa etiqueta de sessão na seção Filtragem de dados externos do console do Lake Formation. Substituir *`123456789012`* com sua própria Conta da AWS identidade.



Etapa 4: Configurar subsídios do AWS Glue e do Lake Formation para funções EMR de tempo de execução da Amazon

Para continuar com a configuração do controle de acesso baseado em Lake Formation com funções de tempo de EMR execução da Amazon, você deve configurar subsídios do AWS Glue e do Lake Formation para funções EMR de tempo de execução da Amazon. Para permitir que suas funções IAM de tempo de execução interajam com o Lake Formation, conceda a elas acesso com `lakeformation:GetDataAccess` `glue:Get*` e.

As permissões do Lake Formation controlam o acesso aos recursos do AWS Glue Data Catalog, aos locais do Amazon S3 e aos dados subjacentes nesses locais. IAMas permissões controlam o acesso ao Lake Formation, ao AWS Glue APIs e aos recursos. Embora você possa ter a permissão do Lake Formation para acessar uma tabela no catálogo de dados (SELECT), sua operação falhará se você não tiver a IAM permissão no `glue:Get*`API. Para obter mais detalhes sobre o controle de acesso do Lake Formation, consulte a visão geral do controle de acesso do Lake Formation.

1. Crie o arquivo `emr-runtime-roles-lake-formation-policy.json` com o conteúdo a seguir.

```
{
    "Version": "2012-10-17",
    "Statement": {
        "Sid": "LakeFormationManagedAccess",
        "Effect": "Allow",
        "Action": [
            "lakeformation:GetDataAccess",
            "glue:Get*",
            "glue:Create*",
            "glue:Update*"
        ],
        "Resource": "*"
    }
}
```

2. Crie a IAM política relacionada.

```
aws iam create-policy \
--policy-name emr-runtime-roles-lake-formation-policy \
--policy-document file://emr-runtime-roles-lake-formation-policy.json
```

3. Para atribuir essa política às suas funções IAM de tempo de execução, siga as etapas em Gerenciamento de AWS Lake Formation permissões.

Já é possível usar perfis de runtime e o Lake Formation para aplicar permissões em nível de tabela e coluna. Você também pode usar uma identidade de origem para controlar ações e monitorar operações com AWS CloudTrail. Para obter um end-to-end exemplo detalhado, consulte Introdução às funções de tempo de execução para EMR as etapas da Amazon.

Apache Hudi e Lake Formation

As EMR versões 6.15.0 e superiores da Amazon incluem suporte para controle de acesso refinado baseado no Apache Hudi quando você lê e grava dados AWS Lake Formation com o Spark. SQL A Amazon EMR oferece suporte ao controle de acesso em nível de tabela, linha, coluna e célula com o Apache Hudi. Com esse recurso, você pode executar consultas de instantâneos em copy-on-write tabelas para consultar o instantâneo mais recente da tabela em um determinado instante de confirmação ou compactação.

Atualmente, um cluster da EMR Amazon habilitado para Lake Formation deve recuperar a coluna de tempo de confirmação de Hudi para realizar consultas incrementais e consultas de viagem no tempo. Ele não suporta a `timestamp as of` sintaxe e a função do Spark. `Spark.read()` A sintaxe correta é `select * from table where _hoodie_commit_time <= point_in_time`. Para obter mais informações, consulte Consultas de viagem no tempo pontual na tabela Hudi.

A matriz de suporte a seguir lista alguns dos principais recursos do Apache Hudi com o Lake Formation:

	Copiar na gravação	mesclar na leitura
Consultas de instantâneos - Spark SQL	✓	✓
Consultas otimizadas para leitura - Spark SQL	✓	✓
Consultas incrementais	✓	✓
Consultas de viagem no tempo	✓	✓
Tabelas de metadados	✓	✓
DML**`INSERT`**comandos	✓	✓
DDLcomandos		
Consultas de fontes de dados do Spark		
Gravações na fonte de dados do Spark		

Consultar tabelas do Hudi

Esta seção mostra como você pode executar as consultas com suporte descritas acima em um cluster habilitado para Lake Formation. A tabela deve ser uma tabela de catálogo registrada.

1. Para iniciar o shell Spark, use os comandos a seguir.

```
spark-sql
--jars /usr/lib/hudi/hudi-spark-bundle.jar \
--conf spark.serializer=org.apache.spark.serializer.KryoSerializer \
--conf
spark.sql.catalog.spark_catalog=org.apache.spark.sql.hudi.catalog.HoodieCatalog \
--conf
spark.sql.extensions=org.apache.spark.sql.hudi.HoodieSparkSessionExtension,com.amazonaws.em
 \
--conf spark.sql.catalog.spark_catalog.lf.managed=true
```

   Se você quiser que o Lake Formation use o servidor de registros para gerenciar seu catálogo do Spark, `spark.sql.catalog.<managed_catalog_name>.lf.managed` defina como true.

2. Para consultar o instantâneo mais recente das copy-on-write tabelas, use os comandos a seguir.

```
SELECT * FROM my_hudi_cow_table
```

```
spark.read.table("my_hudi_cow_table")
```

3. Para consultar os dados compactados mais recentes das tabelas `MOR`, você pode consultar a tabela otimizada para leitura que tem o sufixo `_ro`:

```
SELECT * FROM my_hudi_mor_table_ro
```

```
spark.read.table("my_hudi_mor_table_ro")
```

> **Note**
>
> A performance das leituras nos clusters do Lake Formation pode ser mais lenta devido às otimizações sem suporte. Esses atributos incluem listagem de arquivos com base nos

metadados do Hudi e salto de dados. É recomendável testar a performance da aplicação para garantir que ela atenda aos seus requisitos.

Apache Iceberg e Lake Formation

As EMR versões 6.15.0 e superiores da Amazon incluem suporte para controle de acesso refinado baseado no Apache Iceberg quando você lê e grava dados AWS Lake Formation com o Spark. SQL A Amazon EMR oferece suporte ao controle de acesso em nível de tabela, linha, coluna e célula com o Apache Iceberg. Com esse recurso, você pode executar consultas de instantâneos em copy-on-write tabelas para consultar o instantâneo mais recente da tabela em um determinado instante de confirmação ou compactação.

Se você quiser usar o formato Iceberg, defina as configurações a seguir. Substitua *DB_LOCATION* pelo caminho do Amazon S3 onde suas tabelas do Iceberg estão localizadas e os espaços reservados para a região e o ID da conta por seus próprios valores.

```
spark-sql \
--conf
 spark.sql.extensions=org.apache.iceberg.spark.extensions.IcebergSparkSessionExtensions,com.ama

--conf spark.sql.catalog.iceberg_catalog=org.apache.iceberg.spark.SparkCatalog
--conf spark.sql.catalog.iceberg_catalog.warehouse=s3://DB_LOCATION
--conf spark.sql.catalog.iceberg_catalog.catalog-
impl=org.apache.iceberg.aws.glue.GlueCatalog
--conf spark.sql.catalog.iceberg_catalog.io-impl=org.apache.iceberg.aws.s3.S3FileIO
--conf spark.sql.catalog.iceberg_catalog.glue.account-id=ACCOUNT_ID
--conf spark.sql.catalog.iceberg_catalog.glue.id=ACCOUNT_ID
--conf spark.sql.catalog.iceberg_catalog.client.assume-role.region=AWS_REGION
--conf spark.sql.secureCatalog=iceberg_catalog
```

Se você quiser que o Lake Formation use o servidor de registros para gerenciar seu catálogo do Spark, `spark.sql.catalog.<managed_catalog_name>.lf.managed` defina como true.

Você também deve ter o cuidado NOT de passar as seguintes configurações de assumir função:

```
--conf spark.sql.catalog.my_catalog.client.assume-role.region
--conf spark.sql.catalog.my_catalog.client.assume-role.arn
```

```
--conf spark.sql.catalog.my_catalog.client.assume-
role.tags.LakeFormationAuthorizedCaller
```

A matriz de apoio a seguir lista alguns dos principais recursos do Apache Iceberg com o Lake Formation:

	Copiar na gravação	mesclar na leitura
Consultas de instantâneos - Spark SQL	✓	✓
Consultas otimizadas para leitura - Spark SQL	✓	✓
Consultas incrementais	✓	✓
Consultas de viagem no tempo	✓	✓
Tabelas de metadados	✓	✓
DML**`INSERT`**comandos	✓	✓
DDLcomandos		
Consultas de fontes de dados do Spark		
Gravações na fonte de dados do Spark		

Delta Lake e Lake Formation

As EMR versões 6.15.0 e superiores da Amazon incluem suporte para controle de acesso refinado AWS Lake Formation com base no Delta Lake quando você lê e grava dados com o Spark. SQL A Amazon EMR oferece suporte ao controle de acesso em nível de tabela, linha, coluna e célula com o Delta Lake. Com esse recurso, você pode executar consultas de instantâneos em copy-on-write tabelas para consultar o instantâneo mais recente da tabela em um determinado instante de confirmação ou compactação.

Para usar o Delta Lake com o Lake Formation, execute o comando a seguir.

```
spark-sql \
```

```
--conf
 spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension,com.amazonaws.emr.recordserver.co
 \
--conf spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog
 \
--conf spark.sql.catalog.spark_catalog.lf.managed=true
```

Se você quiser que o Lake Formation use o servidor de registros para gerenciar seu catálogo do Spark, `spark.sql.catalog.<managed_catalog_name>.lf.managed` defina como true.

A seguinte matriz de apoio lista alguns dos principais recursos do Delta Lake com o Lake Formation:

	Copiar na gravação	mesclar na leitura
Consultas de instantâneos - Spark SQL	✓	✓
Consultas otimizadas para leitura - Spark SQL	✓	✓
Consultas incrementais	Sem compatibilidade	Sem compatibilidade
Consultas de viagem no tempo	Sem compatibilidade	Sem compatibilidade
Tabelas de metadados	✓	✓
DML**`INSERT`**comandos	✓	✓
DDLcomandos		
Consultas de fontes de dados do Spark		
Gravações na fonte de dados do Spark		

Criação de uma tabela Delta Lake no AWS Glue Data Catalog

O Amazon EMR with Lake Formation não suporta DDL comandos e criação de tabelas Delta. Siga estas etapas para criar tabelas no AWS Glue Data Catalog.

1. Use o exemplo a seguir para criar uma tabela Delta. Certifique-se de que sua localização no S3 exista.

```
spark-sql \
--conf "spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension" \
--conf
 "spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog"

> CREATE DATABASE if not exists <DATABASE_NAME> LOCATION 's3://<S3_LOCATION>/
transactionaldata/native-delta/<DATABASE_NAME>/';
> CREATE TABLE <TABLE_NAME> (x INT, y STRING, z STRING) USING delta;
> INSERT INTO <TABLE_NAME> VALUES (1, 'a1', 'b1');
```

2. Para ver os detalhes da sua mesa, acesse https://console.aws.amazon.com/glue/.
3. No painel de navegação à esquerda, expanda Catálogo de Dados, escolha Tabelas e escolha a tabela que você criou. Em Esquema, você verá que a tabela Delta que você criou com o Spark armazena todas as colunas em um tipo de dados do `array<string>` AWS Glue.
4. Para definir filtros em nível de coluna e célula no Lake Formation, remova a `col` coluna do seu esquema e, em seguida, adicione as colunas que estão no esquema da tabela. Neste exemplo, adicione as colunas `x`, `y`, `z` e.

Considerações para a Amazon EMR com Lake Formation

Considere o seguinte ao usar a Amazon EMR com AWS Lake Formation.

- O controle de acesso em nível de tabela está disponível em clusters com EMR versões 6.13 e superiores da Amazon.
- O controle de acesso refinado em nível de linha, coluna e célula está disponível em clusters com EMR versões 6.15 e superiores da Amazon.
- Os usuários com acesso a uma tabela podem acessar todas as propriedades da tabela. Se você tiver controle de acesso baseado no Lake Formation em uma tabela, revise a tabela para garantir que as propriedades não contenham dados ou informações sigilosas.
- EMROs clusters da Amazon com Lake Formation não suportam a alternativa do Spark para HDFS quando o Spark coleta estatísticas de tabelas. Isso normalmente ajuda a otimizar a performance da consulta.
- As operações que oferecem suporte a controles de acesso baseados no Lake Formation com tabelas não governadas do Apache Spark incluem `INSERT INTO` e `INSERT OVERWRITE`.

- As operações que oferecem suporte a controles de acesso baseados no Lake Formation com Apache Spark e Apache Hive incluem `SELECT`, `DESCRIBE`, `SHOW DATABASE`, `SHOW TABLE`, `SHOW COLUMN` e `SHOW PARTITION`.
- A Amazon EMR não oferece suporte ao controle de acesso às seguintes operações baseadas em Lake Formation:
  - Grava em tabelas controladas
  - A Amazon EMR não oferece suporte`CREATE TABLE`. Suporte para Amazon EMR 6.10.0 e versões superiores. `ALTER TABLE`
  - DMLdeclarações que não sejam `INSERT` comandos.
- Há diferenças de performance entre a mesma consulta com e sem controle de acesso baseado no Lake Formation.
- Você só pode usar a Amazon EMR com o Lake Formation para trabalhos no Spark.

Integre a Amazon EMR com o Apache Ranger

A partir do Amazon EMR 5.32.0, você pode iniciar um cluster que se integra nativamente com o Apache Ranger. O Apache Ranger é uma estrutura de código aberto para habilitar, monitorar e gerenciar uma segurança de dados abrangente em toda a plataforma Hadoop. Para obter mais informações, consulte Apache Ranger. Com a integração nativa, você pode trazer seu próprio Apache Ranger para impor um controle refinado de acesso aos dados na Amazon. EMR

Esta seção fornece uma visão geral conceitual da EMR integração da Amazon com o Apache Ranger. Também inclui os pré-requisitos e as etapas necessárias para iniciar um EMR cluster da Amazon integrado ao Apache Ranger.

A integração nativa da Amazon EMR com o Apache Ranger oferece os seguintes benefícios principais:

- Controle de acesso refinado aos bancos de dados e tabelas do Hive Metastore, que permite definir políticas de filtragem de dados no nível de banco de dados, tabela e coluna para aplicações Apache Spark e Apache Hive. A filtragem em nível de linha e o mascaramento de dados são compatíveis com aplicações Hive.
- A capacidade de usar suas políticas existentes do Hive diretamente com os aplicativos Amazon EMR for Hive.
- Controle de acesso aos dados do Amazon S3 no nível do prefixo e do objeto, o que permite definir políticas de filtragem de dados para acesso aos dados do S3 usando o sistema de arquivos. EMR

- A capacidade de usar o CloudWatch Logs para auditoria centralizada.
- A Amazon EMR instala e gerencia os plug-ins do Apache Ranger em seu nome.

Apache Ranger

O Apache Ranger é um framework para habilitar, monitorar e gerenciar uma segurança de dados abrangente em toda a plataforma Hadoop.

O Apache Ranger tem os seguintes atributos:

- Administração de segurança centralizada para gerenciar todas as tarefas relacionadas à segurança em uma interface de usuário central ou usando RESTAPIs.
- Autorização refinada para realizar uma ação ou operação específica usando um componente ou ferramenta do Hadoop, gerenciada por meio de uma ferramenta de administração central.
- Um método de autorização padronizado em todos os componentes do Hadoop.
- Suporte aprimorado para diversos métodos de autorização.
- Auditoria centralizada do acesso do usuário e das ações administrativas (relacionadas à segurança) em todos os componentes do Hadoop.

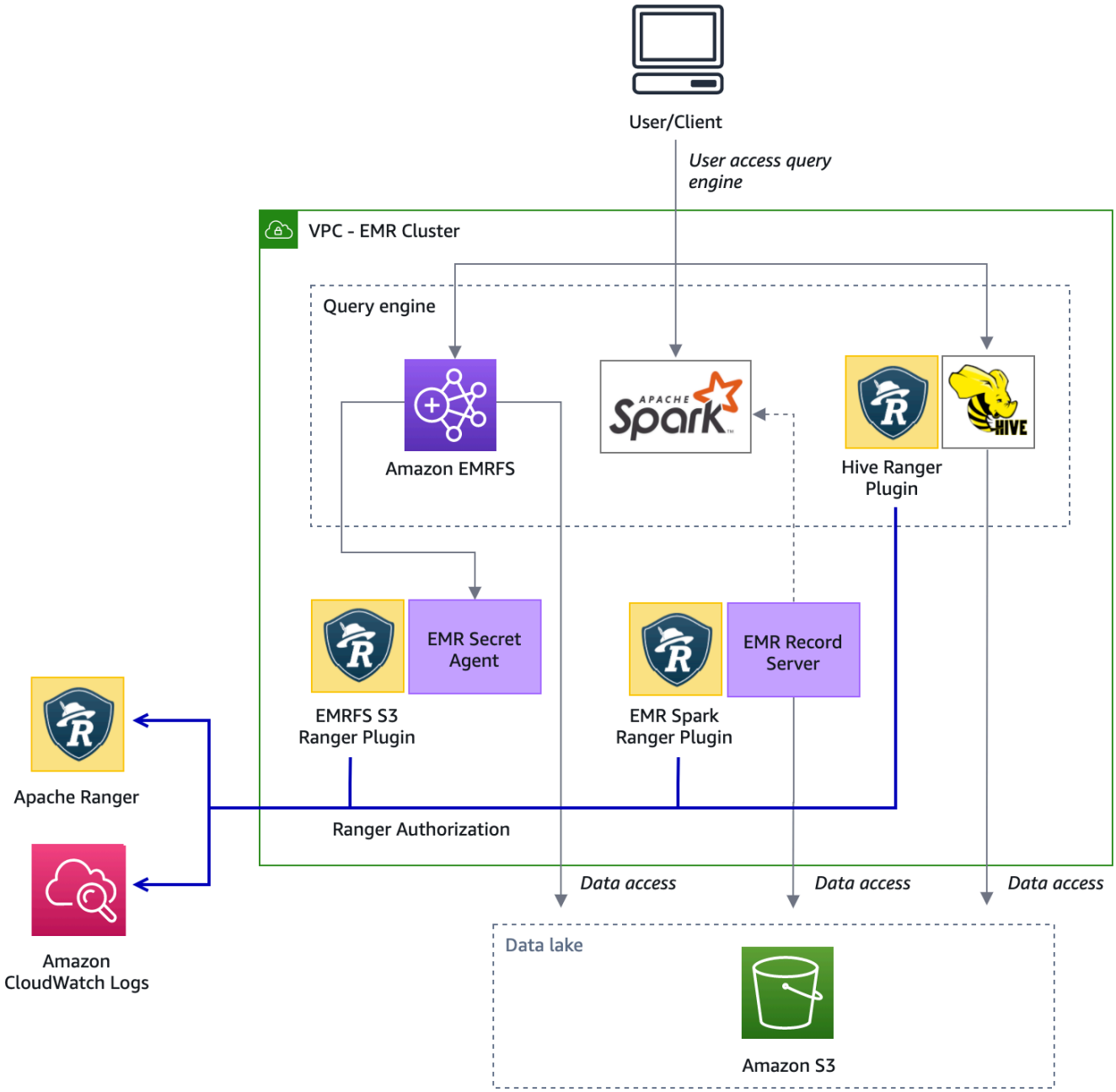
O Apache Ranger usa dois componentes principais para autorização:

- Servidor de administração de políticas Apache Ranger: esse servidor permite definir as políticas de autorização para aplicações Hadoop. Ao se integrar à AmazonEMR, você pode definir e aplicar políticas para que o Apache Spark e o Hive acessem o Hive Metastore e acessem o sistema de arquivos de dados do Amazon S3 (). EMR EMRFS Você pode configurar um novo servidor de administração de políticas Apache Ranger ou usar um existente para se integrar à Amazon. EMR
- Plug-in Apache Ranger: esse plug-in valida o acesso de um usuário em relação às políticas de autorização definidas no servidor de administração de políticas do Apache Ranger. A Amazon EMR instala e configura automaticamente o plug-in Apache Ranger para cada aplicativo Hadoop selecionado na configuração do Apache Ranger.

Tópicos

- Arquitetura da EMR integração da Amazon com o Apache Ranger
- EMRComponentes da Amazon

Arquitetura da EMR integração da Amazon com o Apache Ranger



EMRComponentes da Amazon

A Amazon EMR permite um controle de acesso refinado com o Apache Ranger por meio dos seguintes componentes. Veja o diagrama de arquitetura para uma representação visual desses EMR componentes da Amazon com os plug-ins Apache Ranger.

Agente secreto — O agente secreto armazena segredos com segurança e distribui segredos para outros EMR componentes ou aplicativos da Amazon. Os segredos podem incluir credenciais temporárias de usuário, chaves de criptografia ou tickets Kerberos. O agente secreto é executado em todos os nós do cluster e intercepta chamadas ao serviço de metadados da instância. Para solicitações às credenciais da função de perfil da instância, o agente secreto vende as credenciais dependendo do usuário solicitante e dos recursos solicitados após autorizar a solicitação com o plug-in S3 Ranger. EMRFS O agente secreto é executado como o usuário *`emrsecretagent`* e grava logs no diretório /emr/secretagent/log. O processo depende de um conjunto específico de regras `iptables` para funcionar. É importante garantir que não `iptables` esteja desabilitado. Se você personalizar a `iptables` configuração, as regras da NAT tabela deverão ser preservadas e deixadas inalteradas.

EMRservidor de registros — O servidor de registros recebe solicitações para acessar dados do Spark. Em seguida, ele autoriza as solicitações encaminhando os recursos solicitados para o plug-in Spark Ranger da Amazon. EMR O servidor de registros lê dados do Amazon S3 e retorna dados filtrados que o usuário está autorizado a acessar com base na política do Ranger. O servidor de registros é executado em cada nó do cluster como usuário emr_record_server e grava os registros no diretório /var/log/. emr-record-server

Suporte a aplicações e limitações

Aplicações compatíveis

A integração entre a Amazon EMR e o Apache Ranger, na qual EMR instala os plug-ins Ranger, atualmente, oferece suporte aos seguintes aplicativos:

- Apache Spark (disponível com EMR 5.32+ e 6.3+) EMR
- Apache Hive (disponível com EMR 5.32+ e 6.3+) EMR
- S3 Access por meio de EMRFS (disponível com EMR 5.32+ e 6.3+) EMR

Os aplicativos a seguir podem ser instalados em um EMR cluster e talvez precisem ser configurados para atender às suas necessidades de segurança:

- Apache Hadoop (disponível com EMR 5.32+ e 6.3+, incluindo e) EMR YARN HDFS
- Apache Livy (disponível com EMR 5.32+ e 6.3+) EMR
- Apache Zeppelin (disponível com EMR 5.32+ e 6.3+) EMR
- Apache Hue (disponível com EMR 5.32+ e 6.3+) EMR

- Ganglia (disponível com EMR 5.32+ e 6.3+) EMR
- HCatalog(Disponível com EMR 5,32+ e EMR 6,3+)
- Mahout (disponível com EMR 5.32+ e 6.3+) EMR
- MXNet(Disponível com EMR 5,32+ e EMR 6,3+)
- TensorFlow (Disponível com EMR 5,32+ e EMR 6,3+)
- Tez (disponível com EMR 5.32+ e 6.3+) EMR
- Trino (Disponível com EMR 6.7+)
- ZooKeeper (Disponível com EMR 5,32+ e EMR 6,3+)

> **Important**
>
> As aplicações listadas acima são as únicas com suporte no momento. Para garantir a segurança do cluster, você tem permissão para criar um EMR cluster somente com os aplicativos na lista acima quando o Apache Ranger está ativado.
> No momento, não há suporte para outros aplicativos. Para garantir a segurança do cluster, tentar instalar outras aplicações causará a rejeição do cluster.

Atributos compatíveis

Os seguintes EMR recursos da Amazon podem ser usados com a Amazon EMR e o Apache Ranger:

- Criptografia de dados em repouso e em trânsito
- Autenticação Kerberos (obrigatória)
- Grupos de instâncias, frotas de instâncias e instâncias spot
- Reconfiguração de aplicações em um cluster em execução
- EMRFScriptografia do lado do servidor () SSE

> **Note**
>
> As configurações EMR de criptografia da Amazon governam. SSE Para obter mais informações, consulte Encryption Options.

Limitações de aplicação

Há várias limitações que você deve ter em mente ao integrar a Amazon EMR e o Apache Ranger:

- No momento, você não pode usar o console para criar uma configuração de segurança que especifique a opção de integração do AWS Ranger no. AWS GovCloud (US) Region A configuração de segurança pode ser feita usando CLI o.
- O Kerberos precisa estar instalado no cluster.
- Aplicativos UIs (interfaces de usuário), como a interface do usuário, a interface do usuário e a HDFS NameNode interface do usuário do Livy do YARN Resource Manager, não são configurados com autenticação por padrão.
- As permissões HDFS padrão `umask` são configuradas para que os objetos criados sejam `world wide readable` definidos como padrão.
- A Amazon EMR não oferece suporte ao modo de alta disponibilidade (múltiplo primário) com o Apache Ranger.
- Para ver outras limitações, consulte as limitações de cada aplicação.

> **Note**
>
> As configurações EMR de criptografia da Amazon governam. SSE Para obter mais informações, consulte Encryption Options.

Limitações de plug-in

Cada plug-in tem limitações específicas. Para ver as limitações do plug-in Apache Hive, consulte as limitações do plug-in Apache Hive. Para ver as limitações do plug-in Apache Spark, consulte as limitações do plug-in Apache Spark. Para as limitações do plug-in EMRFS S3, consulte Limitações do plug-in EMRFS S3.

Configurar a Amazon EMR para o Apache Ranger

Antes de instalar o Apache Ranger, revise as informações nesta seção para garantir que a Amazon EMR esteja configurada corretamente.

Tópicos

- Configurar o servidor do Ranger Admin

- IAMfunções para integração nativa com o Apache Ranger
- Crie a configuração EMR de segurança
- Armazene TLS certificados em AWS Secrets Manager
- Iniciar um EMR cluster
- Configure o Zeppelin para clusters Amazon habilitados para Apache Ranger EMR
- Problemas conhecidos

Configurar o servidor do Ranger Admin

Para a EMR integração com a Amazon, os plug-ins do aplicativo Apache Ranger devem se comunicar com o servidor de administração usandoTLS/. SSL

Pré-requisito: Ativação do Ranger Admin Server SSL

O Apache Ranger na Amazon EMR exige SSL comunicação bidirecional entre os plug-ins e o servidor Ranger Admin. Para garantir que os plug-ins se comuniquem com o servidor Apache RangerSSL, habilite o seguinte atributo em ranger-admin-site .xml no servidor Ranger Admin.

```
<property>
    <name>ranger.service.https.attrib.ssl.enabled</name>
    <value>true</value>
</property>
```

Além disso, as configurações a seguir são necessárias.

```
<property>
    <name>ranger.https.attrib.keystore.file</name>
    <value>_<PATH_TO_KEYSTORE>_</value>
</property>

<property>
    <name>ranger.service.https.attrib.keystore.file</name>
    <value>_<PATH_TO_KEYSTORE>_</value>
</property>

<property>
    <name>ranger.service.https.attrib.keystore.pass</name>
    <value>_<KEYSTORE_PASSWORD>_</value>
</property>
```

```
<property>
    <name>ranger.service.https.attrib.keystore.keyalias</name>
    <value><PRIVATE_CERTIFICATE_KEY_ALIAS></value>
</property>

<property>
    <name>ranger.service.https.attrib.clientAuth</name>
    <value>want</value>
</property>

<property>
    <name>ranger.service.https.port</name>
    <value>6182</value>
</property>
```

TLScertificados

A integração do Apache Ranger com a Amazon EMR exige que o tráfego EMR dos nós da Amazon para o servidor Ranger Admin seja criptografado usandoTLS, e que os plug-ins do Ranger sejam autenticados no servidor Apache Ranger usando autenticação mútua bidirecional. TLS O EMR serviço Amazon precisa do certificado público do seu servidor Ranger Admin (especificado no exemplo anterior) e do certificado privado.

Certificados de plug-in do Apache Ranger

TLSOs certificados públicos do plug-in Apache Ranger devem estar acessíveis ao servidor Apache Ranger Admin para validar quando os plug-ins se conectam. Há três métodos diferentes para isso.

Método 1: configurar um armazenamento confiável no servidor Apache Ranger Admin

Preencha as seguintes configurações em ranger-admin-site .xml para configurar um armazenamento confiável.

```
<property>
    <name>ranger.truststore.file</name>
    <value><LOCATION TO TRUSTSTORE></value>
</property>

<property>
    <name>ranger.truststore.password</name>
    <value><PASSWORD FOR TRUSTSTORE></value>
</property>
```

Método 2: carregar o certificado no Java cacerts truststore

Se o seu servidor Ranger Admin não especificar um armazenamento confiável em suas JVM opções, você poderá colocar os certificados públicos do plug-in no armazenamento cacerts padrão.

Método 3: Criar um armazenamento confiável e especificar como parte das Opções JVM

Em `{RANGER_HOME_DIRECTORY}/ews/ranger-admin-services.sh`, modifique `JAVA_OPTS` para incluir "`-Djavax.net.ssl.trustStore=<TRUSTSTORE_LOCATION>`" e "`-Djavax.net.ssl.trustStorePassword=<TRUSTSTORE_PASSWORD>`". Por exemplo, adicione a linha a seguir após o JAVA _ existenteOPTS.

```
JAVA_OPTS=" ${JAVA_OPTS} -Djavax.net.ssl.trustStore=${RANGER_HOME}/truststore/
truststore.jck -Djavax.net.ssl.trustStorePassword=changeit"
```

> **Note**
>
> Essa especificação pode expor a senha do truststore se algum usuário conseguir fazer login no servidor Apache Ranger Admin e ver os processos em execução, como ao usar o comando `ps`.

Usar certificados autoassinados

Não é recomendável usar certificados autoassinados como certificados. Os certificados autoassinados não podem ser revogados e podem não estar em conformidade com os requisitos internos de segurança.

Instalação da definição de serviço

Uma definição de serviço é usada pelo servidor Ranger Admin para descrever os atributos das políticas de uma aplicação. As políticas são então armazenadas em um repositório de políticas para que os clientes baixem.

Para poder configurar as definições de serviço, as REST chamadas devem ser feitas para o servidor Ranger Admin. Consulte Apache Ranger P ublicAPIsv 2 para obter informações APIs obrigatórias na seção a seguir.

Instalar a definição de serviço do Apache Spark

Para instalar a definição de serviço do Apache Spark, consulte Plug-in Apache Spark.

Instalando a definição EMRFS de serviço

Para instalar a definição de serviço do S3 para a AmazonEMR, consulteEMRFSPlug-in S3.

Usar a definição de serviço do Hive

O Apache Hive pode usar a definição de serviço do Ranger já existente que vem com o Apache Ranger 2.0 e versões posteriores. Para obter mais informações, consulte Plug-in Apache Hive.

Regras de tráfego da rede

Quando o Apache Ranger é integrado ao seu EMR cluster, o cluster precisa se comunicar com servidores adicionais e. AWS

Todos os EMR nós da Amazon, incluindo os nós principais e de tarefas, devem ser capazes de se comunicar com os servidores Apache Ranger Admin para baixar as políticas. Se o seu administrador do Apache Ranger estiver em execução na AmazonEC2, você precisará atualizar o grupo de segurança para poder receber tráfego do EMR cluster.

Além de se comunicar com o servidor Ranger Admin, todos os nós precisam ser capazes de se comunicar com os seguintes serviços: AWS

- Amazon S3
- AWS KMS (se estiver usando EMRFS SSE -KMS)
- Amazon CloudWatch
- AWS STS

Se você planeja executar seu EMR cluster em uma sub-rede privada, configure-o VPC para poder se comunicar com esses serviços usando um dos VPCendpoints no Guia VPC do Usuário da Amazon ou usando a instância de tradução de endereço de rede (NAT) no Guia VPC do Usuário da Amazon.AWS PrivateLink

IAMfunções para integração nativa com o Apache Ranger

A integração entre a Amazon EMR e o Apache Ranger depende de três funções principais que você deve criar antes de iniciar seu cluster:

- Um perfil de EC2 instância personalizado da Amazon para a Amazon EMR
- Um IAM papel para os motores Apache Ranger

- Uma IAM função para outros AWS serviços

Esta seção fornece uma visão geral dessas funções e das políticas que você precisa incluir para cada IAM função. Para obter mais informações sobre como criar esses perfis, consulte Configurar o servidor do Ranger Admin.

EC2perfil de instância

A Amazon EMR usa uma função IAM de serviço para realizar ações em seu nome para provisionar e gerenciar clusters. A função de serviço para EC2 instâncias de cluster, também chamada de perfil de EC2 instância para a AmazonEMR, é um tipo especial de função de serviço atribuída a cada EC2 instância em um cluster na inicialização.

Para definir permissões para interação EMR do cluster com dados do Amazon S3 e com o metastore Hive protegido pelo Apache Ranger e outros AWS serviços, defina um perfil de EC2 instância personalizado para usar em vez do quando você iniciar seu cluster. `EMR_EC2_DefaultRole`

Para ter mais informações, consulte Função de serviço para EC2 instâncias de cluster (perfil de EC2 instância) e Personalize IAM funções.

Você precisa adicionar as seguintes declarações ao perfil de EC2 instância padrão da Amazon EMR para poder marcar sessões e acessar o AWS Secrets Manager que armazena TLS certificados.

```
    {
      "Sid": "AllowAssumeOfRolesAndTagging",
      "Effect": "Allow",
      "Action": ["sts:TagSession", "sts:AssumeRole"],
      "Resource": [
        "arn:aws:iam::<AWS_ACCOUNT_ID>:role/<RANGER_ENGINE-
PLUGIN_DATA_ACCESS_ROLE_NAME>",
        "arn:aws:iam::<AWS_ACCOUNT_ID>:role/<RANGER_USER_ACCESS_ROLE_NAME>"
      ]
    },
    {
        "Sid": "AllowSecretsRetrieval",
        "Effect": "Allow",
        "Action": "secretsmanager:GetSecretValue",
        "Resource": [

 "arn:aws:secretsmanager:<REGION>:<AWS_ACCOUNT_ID>:secret:<PLUGIN_TLS_SECRET_NAME>*",

 "arn:aws:secretsmanager:<REGION>:<AWS_ACCOUNT_ID>:secret:<ADMIN_RANGER_SERVER_TLS_SECRET_NAME>
```

```
        ]
    }
```

> **Note**
>
> Para obter as permissões do Secrets Manager, não esqueça o caractere curinga (“*”) no final do nome do segredo, senão as solicitações falharão. O curinga serve para versões de segredo.

> **Note**
>
> Limite o escopo da AWS Secrets Manager política somente aos certificados necessários para o provisionamento.

IAMpapel do Apache Ranger

Essa função fornece credenciais para mecanismos de execução confiáveis, como o Apache Hive e o Amazon EMR Record Server, para acessar os dados do Amazon S3. Use somente essa função para acessar os dados do Amazon S3, incluindo quaisquer KMS chaves, se você estiver usando o SSE S3 -. KMS

Esse perfil deve ser criado com a política mínima indicada no exemplo a seguir.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "CloudwatchLogsPermissions",
      "Action": [
        "logs:CreateLogGroup",
        "logs:CreateLogStream",
        "logs:PutLogEvents"
      ],
      "Effect": "Allow",
      "Resource": [

 "arn:aws:logs:<REGION>:<AWS_ACCOUNT_ID>:<CLOUDWATCH_LOG_GROUP_NAME_IN_SECURITY_CONFIGURATION>:
      ]
    },
```

```
    {
      "Sid": "BucketPermissionsInS3Buckets",
      "Action": [
        "s3:CreateBucket",
        "s3:DeleteBucket",
        "s3:ListAllMyBuckets",
        "s3:ListBucket"
      ],
      "Effect": "Allow",
      "Resource": [
 *"arn:aws:s3:::bucket1",
        "arn:aws:s3:::bucket2"*
        ]
    },
    {
      "Sid": "ObjectPermissionsInS3Objects",
      "Action": [
        "s3:GetObject",
        "s3:DeleteObject",
        "s3:PutObject"
      ],
      "Effect": "Allow",
      "Resource": [
 *"arn:aws:s3:::bucket1/*",
          "arn:aws:s3:::bucket2/*"
*        ]
    }
  ]
}
```

Important

O asterisco “*” no final do recurso de CloudWatch log deve ser incluído para fornecer permissão para gravar nos fluxos de log.

Note

Se você estiver usando a visualização de EMRFS consistência ou a SSE criptografia S3, adicione permissões às tabelas KMS e chaves do DynamoDB para que os mecanismos de execução possam interagir com esses mecanismos.

A IAM função do Apache Ranger é assumida pela função do perfil da EC2 instância. Use o exemplo a seguir para criar uma política de confiança que permita que a IAM função do Apache Ranger seja assumida pela função do perfil da EC2 instância.

```
   {
     "Sid": "",
     "Effect": "Allow",
     "Principal": {
       "AWS": "arn:aws:iam::<AWS_ACCOUNT_ID>:role/<EC2 INSTANCE PROFILE ROLE NAME eg.
EMR_EC2_DefaultRole>"
     },
     "Action": ["sts:AssumeRole", "sts:TagSession"]
   }
```

IAMfunção para outros AWS serviços

Essa função fornece aos usuários que não são mecanismos de execução confiáveis credenciais para interagir com os AWS serviços, se necessário. Não use essa IAM função para permitir o acesso aos dados do Amazon S3, a menos que sejam dados que devam ser acessados por todos os usuários.

Essa função será assumida pela função de perfil da EC2 instância. Use o exemplo a seguir para criar uma política de confiança que permita que a IAM função do Apache Ranger seja assumida pela função do perfil da EC2 instância.

```
   {
     "Sid": "",
     "Effect": "Allow",
     "Principal": {
       "AWS": "arn:aws:iam::<AWS_ACCOUNT_ID>:role/<EC2 INSTANCE PROFILE ROLE NAME eg.
EMR_EC2_DefaultRole>"
     },
     "Action": ["sts:AssumeRole", "sts:TagSession"]
   }
```

Validar permissões

Consulte Solução de problemas do Apache Ranger para ver instruções sobre como validar permissões.

Crie a configuração EMR de segurança

Criação de uma configuração EMR de segurança da Amazon para o Apache Ranger

Antes de iniciar um EMR cluster da Amazon integrado ao Apache Ranger, crie uma configuração de segurança.

Console

Criar uma configuração de segurança que especifique a opção Integração do AWS Ranger

1. No EMR console da Amazon, selecione Configurações de segurança e, em seguida, Criar.
2. Digite um nome em Name (Nome) para a configuração de segurança. Esse nome é usado para especificar a configuração de segurança ao criar um cluster.
3. Em Integração do AWS Ranger, selecione Habilitar controle de acesso granular gerenciado pelo Apache Ranger.
4. Selecione sua IAMfunção para o Apache Ranger se candidatar. Para obter mais informações, consulte IAMfunções para integração nativa com o Apache Ranger.
5. Selecione sua IAMfunção para que outros AWS serviços se inscrevam.
6. Configure os plug-ins para se conectar ao servidor Ranger Admin inserindo o Secret Manager ARN do servidor Admin e o endereço.
7. Selecione as aplicações para configurar os plug-ins do Ranger. Preencha o Gerenciador Secreto ARN que contém o TLS certificado privado do plug-in.

   Se você não configurar o Apache Spark ou o Apache Hive e eles forem selecionados como uma aplicação para seu cluster, a solicitação falhará.
8. Configure outras opções de configuração de segurança conforme apropriado e escolha Create (Criar). Você deve habilitar a autenticação Kerberos usando o cluster dedicado ou externo. KDC

> **Note**
>
> No momento, você não pode usar o console para criar uma configuração de segurança que especifique a opção de integração do AWS Ranger no. AWS GovCloud (US) Region A configuração de segurança pode ser feita usando CLI o.

CLI

Criar uma configuração de segurança para integração do Apache Ranger

1. *<ACCOUNT ID>*Substitua pelo ID AWS da sua conta.
2. Substitua *<REGION>* pela região em que o recurso está.
3. Especifique um valor `TicketLifetimeInHours` para determinar o período durante o qual um tíquete Kerberos emitido pela KDC é válido.
4. Especifique o endereço do servidor Ranger Admin para `AdminServerURL`.

```
{
    "AuthenticationConfiguration": {
        "KerberosConfiguration": {
            "Provider": "ClusterDedicatedKdc",
            "ClusterDedicatedKdcConfiguration": {
                "TicketLifetimeInHours": 24
            }
        }
    },
    "AuthorizationConfiguration":{
      "RangerConfiguration":{
         "AdminServerURL":"https://_<RANGER ADMIN SERVER IP>_:6182",
         "RoleForRangerPluginsARN":"arn:aws:iam::_<ACCOUNT ID>_:role/_<RANGER PLUGIN
 DATA ACCESS ROLE NAME>_",
         "RoleForOtherAWSServicesARN":"arn:aws:iam::_<ACCOUNT ID>_:role/_<USER
 ACCESS ROLE NAME>_",
         "AdminServerSecretARN":"arn:aws:secretsmanager:_<REGION>_:_<ACCOUNT
 ID>_:secret:_<SECRET NAME THAT PROVIDES ADMIN SERVERS PUBLIC TLS CERTIFICATE
 WITHOUT VERSION>_",
         "RangerPluginConfigurations":[
            {
               "App":"Spark",
               "ClientSecretARN":"arn:aws:secretsmanager:_<REGION>_:_<ACCOUNT
 ID>_:secret:_<SECRET NAME THAT PROVIDES SPARK PLUGIN PRIVATE TLS CERTIFICATE
 WITHOUT VERSION>_",
               "PolicyRepositoryName":"<SPARK SERVICE NAME eg. amazon-emr-spark>"
            },
            {
               "App":"Hive",
```

```
                "ClientSecretARN":"arn:aws:secretsmanager:_<REGION>_:_<ACCOUNT
 ID>_:secret:_<SECRET NAME THAT PROVIDES Hive PLUGIN PRIVATE TLS CERTIFICATE WITHOUT
 VERSION>_",
                "PolicyRepositoryName":"<HIVE SERVICE NAME eg. Hivedev>"
            },
            {
                "App":"EMRFS-S3",
                "ClientSecretARN":"arn:aws:secretsmanager:_<REGION>_:_<ACCOUNT
 ID>_:secret:_<SECRET NAME THAT PROVIDES EMRFS S3 PLUGIN PRIVATE TLS CERTIFICATE
 WITHOUT VERSION>_",
                "PolicyRepositoryName":"<EMRFS S3 SERVICE NAME eg amazon-emr-emrfs>"
            },
       {
                "App":"Trino",
                "ClientSecretARN":"arn:aws:secretsmanager:_<REGION>_:_<ACCOUNT
 ID>_:secret:_<SECRET NAME THAT PROVIDES TRINO PLUGIN PRIVATE TLS CERTIFICATE
 WITHOUT VERSION>_",
                "PolicyRepositoryName":"<TRINO SERVICE NAME eg amazon-emr-trino>"
            }
         ],
         "AuditConfiguration":{
            "Destinations":{
               "AmazonCloudWatchLogs":{
                  "CloudWatchLogGroup":"arn:aws:logs:<REGION>:_<ACCOUNT ID>_:log-
group:_<LOG GROUP NAME FOR AUDIT EVENTS>_"
               }
            }
         }
      }
   }
}
```

Esses PolicyRespositoryNames são os nomes dos serviços especificados em seu Apache Ranger Admin.

Crie uma configuração EMR de segurança da Amazon com o comando a seguir. Substitua a configuração de segurança por um nome a sua escolha. Selecione a configuração pelo nome ao criar o cluster.

```
aws emr create-security-configuration \
--security-configuration file://./security-configuration.json \
--name security-configuration
```

Configurar atributos de segurança adicionais

Para integrar EMR com segurança a Amazon ao Apache Ranger, configure os seguintes recursos de segurança: EMR

- Ative a autenticação Kerberos usando o cluster dedicado ou externo. KDC Para obter instruções, consulte Use o Kerberos para autenticação com a Amazon EMR.
- (Opcional) Habilite a criptografia em trânsito ou em repouso. Para obter mais informações, consulte Opções de criptografia.

Para obter mais informações, consulte Segurança na Amazon EMR.

Armazene TLS certificados em AWS Secrets Manager

Os plug-ins Ranger instalados em um EMR cluster da Amazon e o servidor Ranger Admin devem se comunicar TLS para garantir que os dados da política e outras informações enviadas não possam ser lidos se forem interceptados. EMRtambém exige que os plug-ins se autentiquem no servidor Ranger Admin fornecendo seu próprio TLS certificado e realizem a autenticação bidirecional. TLS Essa configuração exigiu a criação de quatro certificados: dois pares de TLS certificados privados e públicos. Para obter instruções sobre como instalar o certificado no servidor Ranger Admin, consulte Configurar o servidor do Ranger Admin. Para concluir a configuração, os plug-ins Ranger instalados no EMR cluster precisam de dois certificados: o TLS certificado público do seu servidor de administração e o certificado privado que o plug-in usará para se autenticar no servidor Ranger Admin. Para fornecer esses TLS certificados, eles devem estar no AWS Secrets Manager e ser fornecidos em uma Configuração EMR de Segurança.

> **Note**
>
> É altamente recomendável, mas não obrigatório, criar um par de certificados para cada uma das aplicações para limitar o impacto se um dos certificados do plug-in for comprometido.

> **Note**
>
> É necessário rastrear e alternar os certificados antes da data de vencimento.

Formato do certificado

A importação dos certificados para o AWS Secrets Manager é a mesma, independentemente de ser o certificado de plug-in privado ou o certificado de administrador público do Ranger. Antes de importar os TLS certificados, eles devem estar no formato 509xPEM.

Este é o formato de um exemplo de certificado público:

```
-----BEGIN CERTIFICATE-----
...Certificate Body...
-----END CERTIFICATE-----
```

Este é o formato de um exemplo de certificado privado:

```
-----BEGIN PRIVATE KEY-----
...Private Certificate Body...
-----END PRIVATE KEY-----
-----BEGIN CERTIFICATE-----
...Trust Certificate Body...
-----END CERTIFICATE-----
```

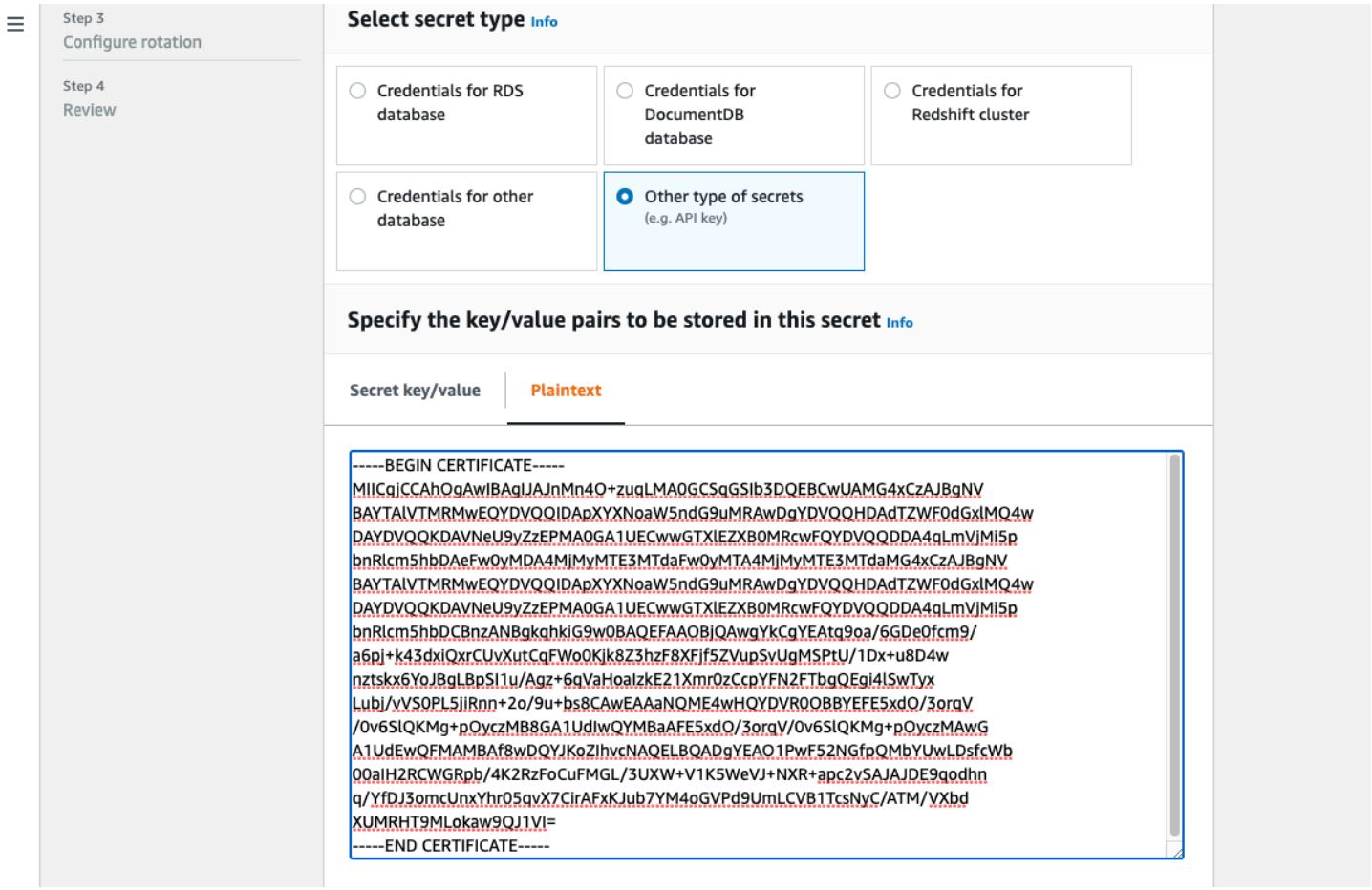
O certificado privado também deverá conter um certificado de confiança.

É possível validar se os certificados estão no formato correto executando o seguinte comando:

```
openssl x509 -in <PEM FILE> -text
```

Importar um certificado para o AWS Secrets Manager

Ao criar seu segredo no Secrets Manager, escolha Outro tipo de segredos em Tipo secreto e cole seu certificado PEM codificado no campo Texto simples.



Iniciar um EMR cluster

Antes de iniciar um EMR cluster da Amazon com o Apache Ranger, certifique-se de que cada componente atenda aos seguintes requisitos mínimos de versão:

- Amazon EMR 5.32.0 ou posterior, ou 6.3.0 ou posterior. Recomendamos que você use a EMR versão mais recente da Amazon.
- Servidor Apache Ranger Admin 2.x.

Execute as etapas a seguir.

- Instale o Apache Ranger, caso ainda não tenha instalado. Para obter mais informações sobre a instalação, consulte Apache Ranger 0.5.0 installation.
- Certifique-se de que haja conectividade de rede entre seu EMR cluster Amazon e o servidor Apache Ranger Admin. Consulte Configurar o servidor do Ranger Admin

- Crie as IAM funções necessárias. Consulte IAMfunções para integração nativa com o Apache Ranger.
- Crie uma configuração EMR de segurança para a instalação do Apache Ranger. Veja mais informações em Crie a configuração EMR de segurança.

Configure o Zeppelin para clusters Amazon habilitados para Apache Ranger EMR

O tópico aborda como configurar o Apache Zeppelin para um EMR cluster Amazon habilitado para Apache Ranger para que você possa usar o Zeppelin como um notebook para exploração interativa de dados. O Zeppelin está incluído nas versões EMR 5.0.0 e posteriores da Amazon. As versões anteriores incluem o Zeppelin como uma aplicação sandbox. Para obter mais informações, consulte as versões de lançamento da Amazon EMR 4.x no Amazon EMR Release Guide.

Por padrão, o Zeppelin é configurado com um login e uma senha padrão que não são seguros em um ambiente multilocatário.

Para configurar o Zeppelin, siga as etapas a seguir.

1. Modificar o mecanismo de autenticação.

   Modifique o arquivo `shiro.ini` para implementar o mecanismo de autenticação de sua preferência. O Zeppelin é compatível com Active Directory, LDAPPAM, e Knox. SSO Consulte Apache Shiro authentication for Apache Zeppelin para obter mais informações.
2. Configurar o Zeppelin para representar o usuário final

   Quando você permite que o Zeppelin represente o usuário final, os trabalhos enviados pelo Zeppelin podem ser executados como esse usuário final. Adicione o seguinte à configuração de `core-site.xml`:

```
[
  {
    "Classification": "core-site",
    "Properties": {
      "hadoop.proxyuser.zeppelin.hosts": "*",
      "hadoop.proxyuser.zeppelin.groups": "*"
    },
    "Configurations": [
    ]
  }
```

```
]
```

Em seguida, adicione a seguinte configuração a `hadoop-kms-site.xml` localizado em `/etc/hadoop/conf`:

```
[
  {
    "Classification": "hadoop-kms-site",
    "Properties": {
      "hadoop.kms.proxyuser.zeppelin.hosts": "*",
      "hadoop.kms.proxyuser.zeppelin.groups": "*"
    },
    "Configurations": [
    ]
  }
]
```

Você também pode adicionar essas configurações ao seu EMR cluster da Amazon usando o console seguindo as etapas em Reconfigurar um grupo de instâncias no console.

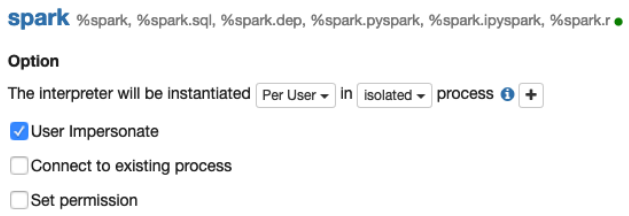
3. Permitir que o Zeppelin se torne o usuário final

   Crie um arquivo `/etc/sudoers.d/90-zeppelin-user` que contenha:

   ```
   zeppelin ALL=(ALL) NOPASSWD:ALL
   ```

4. Modificar as configurações dos intérpretes para executar trabalhos do usuário em seus próprios processos.

   Configure todos os intérpretes para instanciar os intérpretes “Por usuário” em processos “isolados”.

5. Modificar **`zeppelin-env.sh`**

Adicione isto a `zeppelin-env.sh` que o Zeppelin comece a iniciar intérpretes como usuário final:

```
ZEPPELIN_IMPERSONATE_USER=`echo ${ZEPPELIN_IMPERSONATE_USER} | cut -d @ -f1`
export ZEPPELIN_IMPERSONATE_CMD='sudo -H -u ${ZEPPELIN_IMPERSONATE_USER} bash -c'
```

Adicione isto a `zeppelin-env.sh` para alterar as permissões padrão de caderno para somente leitura para o criador:

```
export ZEPPELIN_NOTEBOOK_PUBLIC="false"
```

Por fim, adicione o seguinte `zeppelin-env.sh` para incluir o caminho da EMR RecordServer classe após a primeira `CLASSPATH` declaração:

```
export CLASSPATH="$CLASSPATH:/usr/share/aws/emr/record-server/lib/aws-emr-record-
server-connector-common.jar:/usr/share/aws/emr/record-server/lib/aws-emr-record-
server-spark-connector.jar:/usr/share/aws/emr/record-server/lib/aws-emr-record-
server-client.jar:/usr/share/aws/emr/record-server/lib/aws-emr-record-server-
common.jar:/usr/share/aws/emr/record-server/lib/jars/secret-agent-interface.jar"
```

6. Reiniciar o Zeppelin.

   Execute o seguinte comando para reiniciar o Zeppelin:

   ```
   sudo systemctl restart zeppelin
   ```

Problemas conhecidos

Problemas conhecidos

Há um problema conhecido na EMR versão 5.32 da Amazon em que as permissões `hive-site.xml` foram alteradas para que somente usuários privilegiados possam lê-la, pois pode haver credenciais armazenadas nela. Isso pode impedir que o Hue leia `hive-site.xml` e fazer com que as páginas da Web sejam recarregadas continuamente. Se você tiver esse problema, adicione esta configuração para corrigir o problema:

```
[
  {
```

```
    "Classification": "hue-ini",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "desktop",
        "Properties": {
          "server_group":"hive_site_reader"
         },
        "Configurations":[
        ]
      }
    ]
  }
]
```

Há um problema conhecido de que o plug-in EMRFS S3 para Apache Ranger atualmente não suporta o recurso Security Zone do Apache Ranger. As restrições de controle de acesso definidas usando o recurso Security Zone não são aplicadas aos seus EMR clusters da Amazon.

Aplicação UIs

Por padrão, as IUs de aplicações não realizam autenticação. Isso inclui a ResourceManager interface do usuário, a NodeManager interface do usuário, a interface do usuário Livy, entre outras. Além disso, qualquer usuário que tenha a capacidade de acessar o UIs é capaz de visualizar informações sobre os trabalhos de todos os outros usuários.

Se esse comportamento não for desejado, você deve garantir que um grupo de segurança seja usado para restringir o acesso dos usuários UIs ao aplicativo.

HDFSPermissões padrão

Por padrão, os objetos que os usuários criam HDFS recebem permissões legíveis em todo o mundo. Isso poderá tornar os dados legíveis por usuários que não deveriam ter acesso a eles. Para alterar esse comportamento de modo que as permissões de arquivo padrão sejam definidas para leitura e gravação somente pelo criador do trabalho, execute as etapas a seguir.

Ao criar seu EMR cluster, forneça a seguinte configuração:

```
[
  {
    "Classification": "hdfs-site",
```

```
    "Properties": {
      "dfs.namenode.acls.enabled": "true",
      "fs.permissions.umask-mode": "077",
      "dfs.permissions.superusergroup": "hdfsadmingroup"
    }
  }
]
```

Além disso, execute esta ação de bootstrap:

```
--bootstrap-actions Name='HDFS UMask Setup',Path=s3://elasticmapreduce/hdfs/umask/
umask-main.sh
```

Plug-ins Apache Ranger

Os plug-ins Apache Ranger validam o acesso de um usuário em relação às políticas de autorização definidas no servidor de administração de políticas do Apache Ranger.

Tópicos

- Plug-in Apache Hive
- Plug-in Apache Spark
- EMRFSPlug-in S3
- Plug-in Trino

Plug-in Apache Hive

O Apache Hive é um mecanismo de execução bastante usado dentro do ecossistema Hadoop. EMRA Amazon fornece um plug-in Apache Ranger para poder fornecer controles de acesso refinados para o Hive. O plug-in é compatível com o servidor Apache Ranger Admin de código aberto versão 2.0 e posteriores.

Tópicos

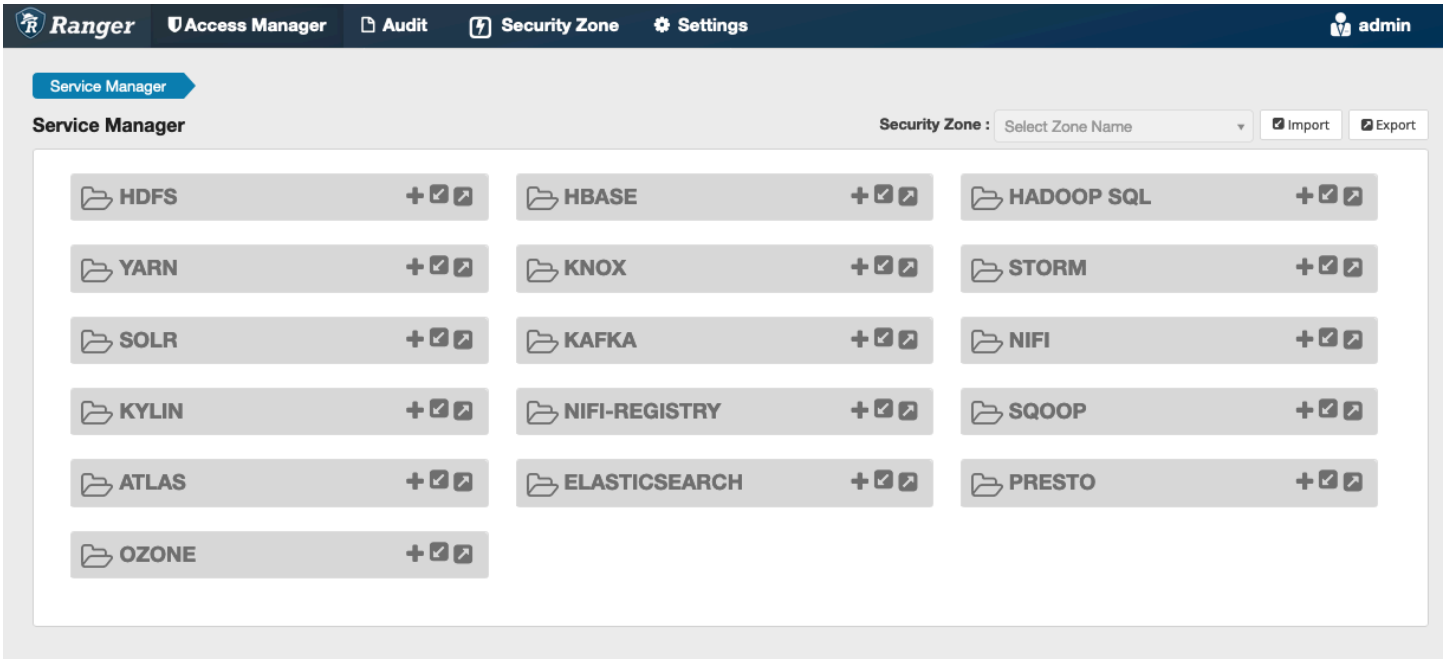
- Atributos compatíveis
- Instalação da configuração de serviço
- Considerações
- Limitações

Atributos compatíveis

O plug-in Apache Ranger para Hive on EMR suporta todas as funcionalidades do plug-in de código aberto, que inclui controles de acesso em nível de banco de dados, tabela e coluna, filtragem de linhas e mascaramento de dados. Para ver uma tabela dos comandos do Hive e das permissões associadas do Ranger, consulte Hive commands to Ranger permission mapping.

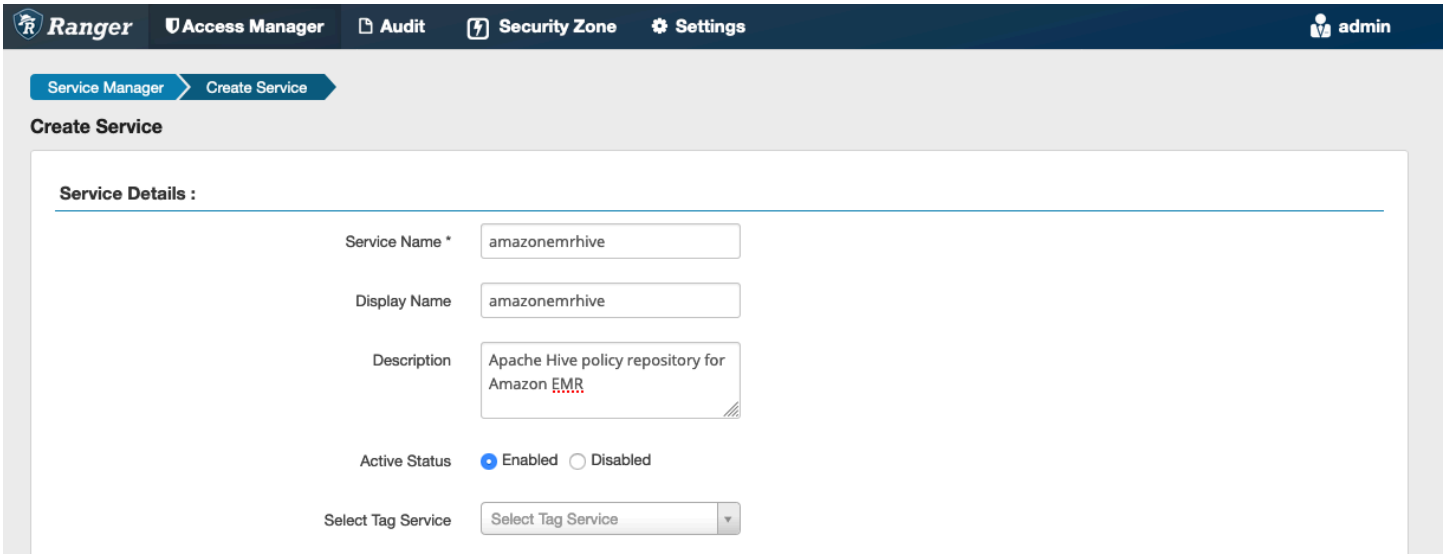
Instalação da configuração de serviço

O plug-in Apache Hive é compatível com a definição de serviço Hive existente no Apache Hive Hadoop. SQL



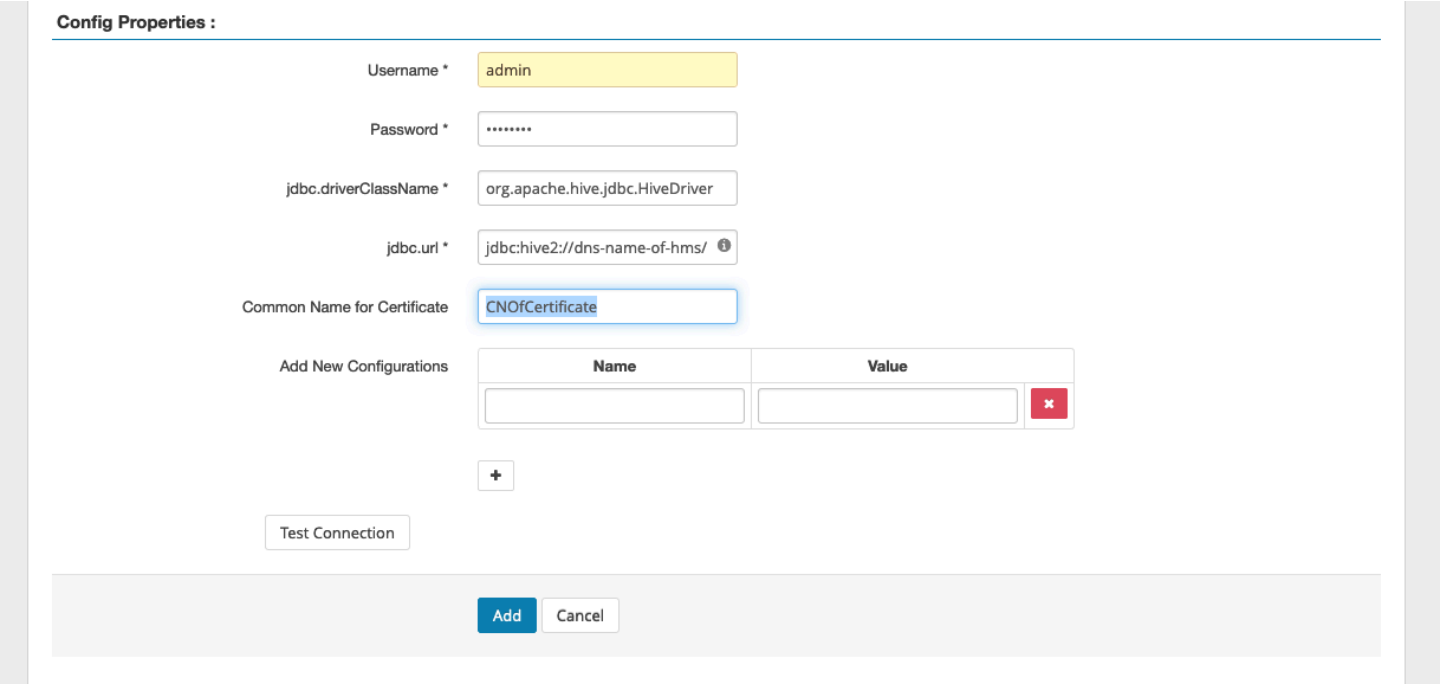
Se você não tiver uma instância do serviço no HadoopSQL, como mostrado acima, você pode criar uma. Clique no + ao lado do HadoopSQL.

1. Nome do serviço (se for exibido): insira o nome do serviço. O valor sugerido é **`amazonemrhive`**. Anote esse nome de serviço -- ele é necessário ao criar uma configuração EMR de segurança.
2. Nome de exibição: insira o nome a ser exibido para o serviço. O valor sugerido é **`amazonemrhive`**.

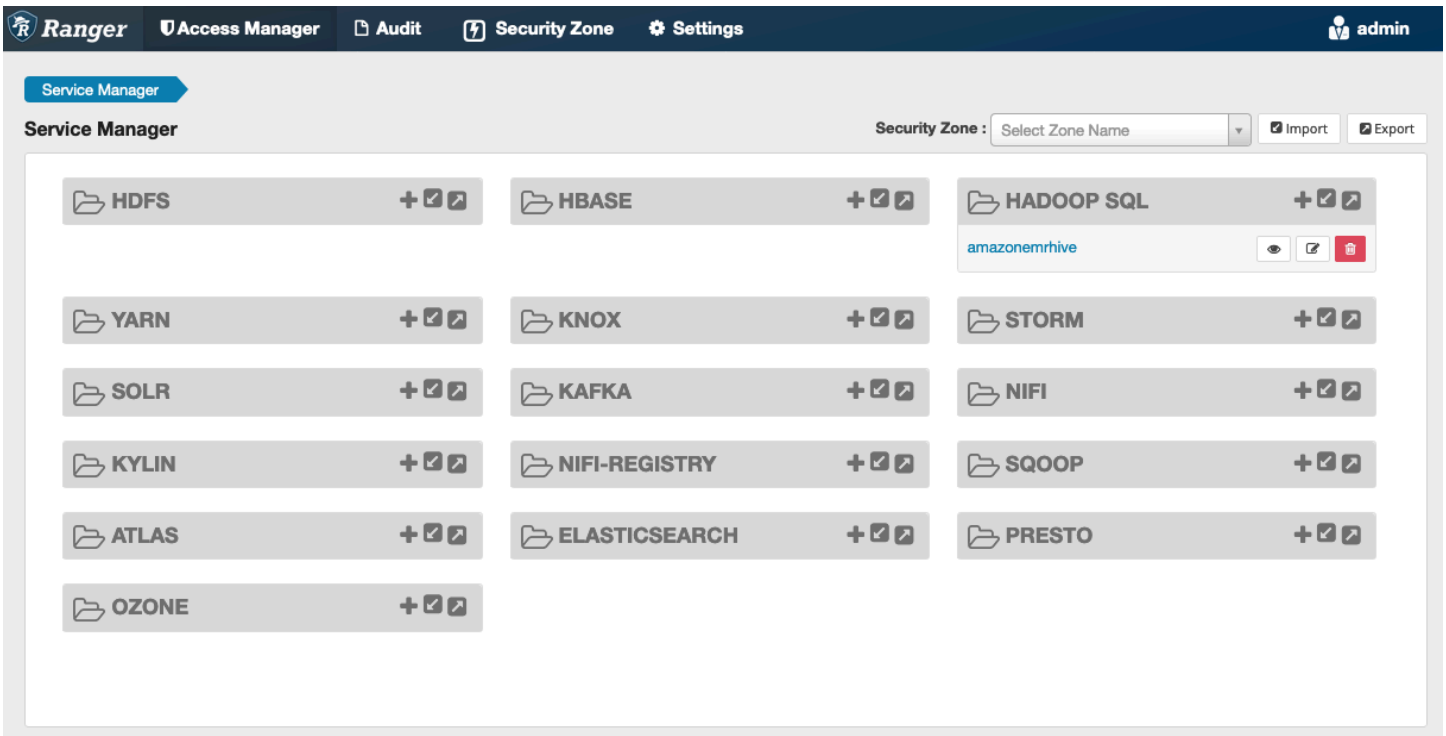


As propriedades de configuração do Apache Hive são usadas para estabelecer uma conexão com seu servidor Apache Ranger Admin com um 2 HiveServer para implementar o preenchimento automático ao criar políticas. As propriedades abaixo não precisam ser precisas se você não tiver um processo persistente HiveServer 2 e puderem ser preenchidas com qualquer informação.

- Nome de usuário: insira um nome de usuário para a JDBC conexão com uma instância de uma instância HiveServer 2.
- Senha: insira a senha do nome de usuário acima.
- jdbc. driver. ClassName: insira o nome da JDBC classe para a conectividade do Apache Hive. O valor padrão pode ser usado.
- jdbc.url: insira a string de JDBC conexão a ser usada ao se conectar a 2. HiveServer
- Nome comum para certificado: o campo CN dentro do certificado usado para se conectar ao servidor de administração com base em um plug-in cliente. Esse valor deve corresponder ao campo CN em seu TLS certificado que foi criado para o plug-in.



O botão Testar conexão testa se os valores acima podem ser usados para se conectar com êxito à instância HiveServer 2. Depois que o serviço for criado com êxito, o Service Manager deverá ficará semelhante a isto:



Considerações

Servidor de metadados Hive

O servidor de metadados Hive só pode ser acessado por mecanismos confiáveis, especificamente o Hive e `emr_record_server`, para proteção contra acesso não autorizado. O servidor de metadados Hive também é acessado por todos os nós do cluster. A porta 9083 necessária fornece acesso de todos os nós ao nó principal.

Autenticação

Por padrão, o Apache Hive está configurado para se autenticar usando Kerberos conforme configurado na configuração de Segurança. EMR HiveServer2 também pode ser configurado para autenticar usuários usandoLDAP. Consulte LDAPImplementação da autenticação para o Hive em um cluster EMR Amazon multilocatário para obter informações.

Limitações

A seguir estão as limitações atuais do plug-in Apache Hive na Amazon EMR 5.x:

- Não há suporte para perfis do Hive atualmente. Não há suporte para instruções Grant e Revoke.
- O Hive não CLI é suportado. JDBC/Beeline é a única forma autorizada de conectar o Hive.
- `hive.server2.builtin.udf.blacklista` configuração deve ser preenchida com o UDFs que você considera inseguro.

Plug-in Apache Spark

EMRA Amazon se integrou EMR RecordServer para fornecer controle de acesso refinado para o Spark. SQL EMR's RecordServer é um processo privilegiado executado em todos os nós em um cluster habilitado para Apache Ranger. Quando um driver ou executor do Spark executa uma SQL instrução do Spark, todos os metadados e solicitações de dados passam pelo. RecordServer Para saber mais EMR RecordServer, consulte a EMRComponentes da Amazon página.

Tópicos

- Atributos compatíveis
- Reimplante a definição de serviço para usar INSERTALTER, ou instruções DDL
- Instalação da definição de serviço

- Criação de políticas do Spark SQL
- Considerações
- Limitações

Atributos compatíveis

SQLDeclaração/ação do ranger	STATUS	EMRVersão suportada
SELECT	Compatível	A partir da 5.32
SHOW DATABASES	Compatível	A partir da 5.32
SHOW COLUMNS	Compatível	A partir da 5.32
SHOW TABLES	Compatível	A partir da 5.32
SHOW TABLE PROPERTIES	Compatível	A partir da 5.32
DESCRIBE TABLE	Compatível	A partir da 5.32
INSERT OVERWRITE	Compatível	A partir da 5.34 e 6.4
INSERT INTO	Compatível	A partir da 5.34 e 6.4
ALTER TABLE	Compatível	A partir da 6.4
CREATE TABLE	Compatível	A partir da 5.35 e 6.7
CREATE DATABASE	Compatível	A partir da 5.35 e 6.7
DROP TABLE	Compatível	A partir da 5.35 e 6.7

SQLDeclaração/ação do ranger	STATUS	EMRVersão suportada
DROP DATABASE	Compatível	A partir da 5.35 e 6.7
DROP VIEW	Compatível	A partir da 5.35 e 6.7
CREATE VIEW	Sem suporte	

Os seguintes recursos são compatíveis ao usar o Spark: SQL

- Controle de acesso refinado em tabelas dentro do Hive Metastore, e é possível criar políticas em nível de banco de dados, tabela e coluna.
- As políticas do Apache Ranger podem incluir políticas de concessão e políticas de negação para usuários e grupos.
- Os eventos de auditoria são enviados para o CloudWatch Logs.

Reimplante a definição de serviço para usar INSERTALTER, ou instruções DDL

> **Note**
>
> A partir do Amazon EMR 6.4, você pode usar o Spark SQL com as declarações: INSERTINTO, INSERTOVERWRITE, ou. ALTER TABLE A partir do Amazon EMR 6.7, você pode usar o Spark SQL para criar ou eliminar bancos de dados e tabelas. Se você já tiver uma instalação no servidor Apache Ranger com definições de serviço Apache Spark implantadas, use o código a seguir para reimplantar as definições de serviço.

```
# Get existing Spark service definition id calling Ranger REST API and JSON
 processor
curl --silent -f -u <admin_user_login>:<password_for_ranger_admin_user> \
-H "Accept: application/json" \
-H "Content-Type: application/json" \
-k 'https://*<RANGER SERVER ADDRESS>*:6182/service/public/v2/api/servicedef/
name/amazon-emr-spark' | jq .id

# Download the latest Service definition
```

```
wget https://s3.amazonaws.com/elasticmapreduce/ranger/service-definitions/
version-2.0/ranger-servicedef-amazon-emr-spark.json

# Update the service definition using the Ranger REST API
curl -u <admin_user_login>:<password_for_ranger_admin_user> -X PUT -d @ranger-
servicedef-amazon-emr-spark.json \
-H "Accept: application/json" \
-H "Content-Type: application/json" \
-k 'https://*<RANGER SERVER ADDRESS>*:6182/service/public/v2/api/
servicedef/<Spark service definition id from step 1>'
```

Instalação da definição de serviço

A instalação da definição EMR do serviço Apache Spark exige que o servidor Ranger Admin seja configurado. Consulte Configurar o servidor do Ranger Admin.

Siga estas etapas para instalar a definição de serviço Apache Spark:

Etapa 1: SSH no servidor Apache Ranger Admin

Por exemplo:

```
ssh ec2-user@ip-xxx-xxx-xxx-xxx.ec2.internal
```

Etapa 2: baixar a definição de serviço e o plug-in do servidor Apache Ranger Admin

Em um diretório temporário, baixe a definição de serviço. Essa definição de serviço é compatível com as versões Ranger 2.x.

```
mkdir /tmp/emr-spark-plugin/
cd /tmp/emr-spark-plugin/

wget https://s3.amazonaws.com/elasticmapreduce/ranger/service-definitions/version-2.0/
ranger-spark-plugin-2.x.jar
wget https://s3.amazonaws.com/elasticmapreduce/ranger/service-definitions/version-2.0/
ranger-servicedef-amazon-emr-spark.json
```

Etapa 3: instalar o plug-in Apache Spark para a Amazon EMR

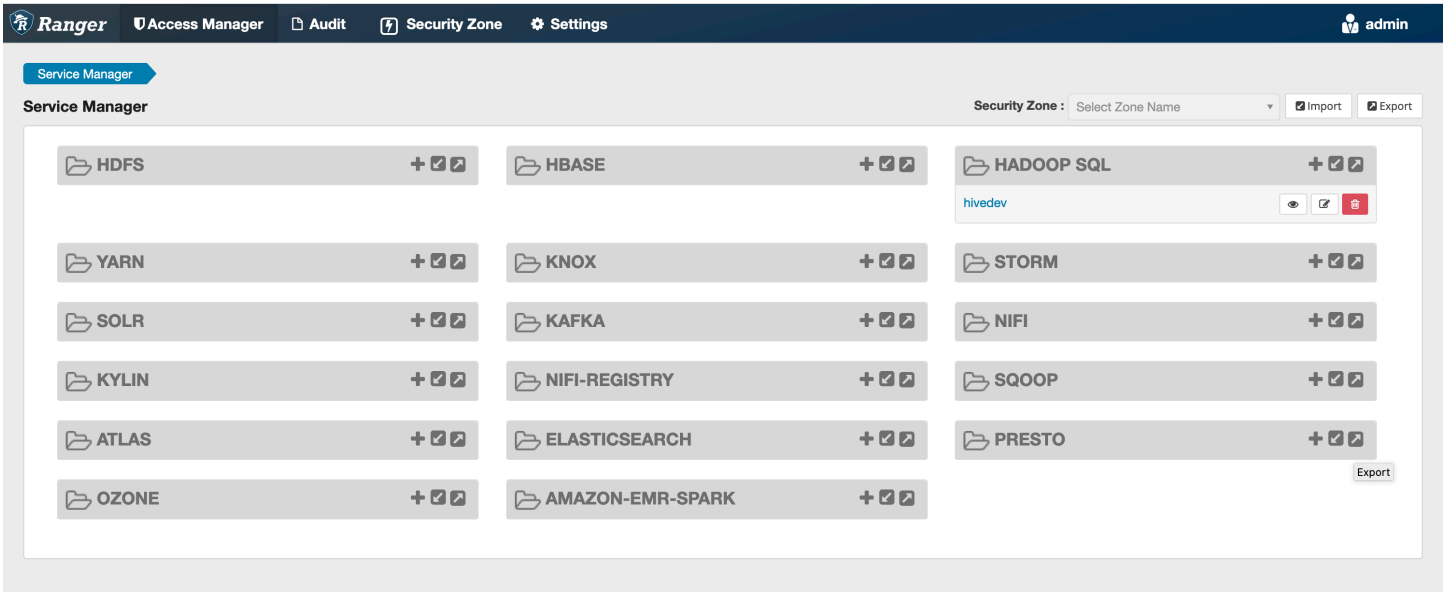
```
export RANGER_HOME=.. # Replace this Ranger Admin's home directory eg /usr/lib/ranger/
ranger-2.0.0-admin
```

```
mkdir $RANGER_HOME/ews/webapp/WEB-INF/classes/ranger-plugins/amazon-emr-spark
mv ranger-spark-plugin-2.x.jar $RANGER_HOME/ews/webapp/WEB-INF/classes/ranger-plugins/
amazon-emr-spark
```

Etapa 4: registrar a definição do serviço Apache Spark para a Amazon EMR

```
curl -u *<admin users login>*:*_<_**_password_ **_for_** _ranger admin user_**_>_* -X
 POST -d @ranger-servicedef-amazon-emr-spark.json \
-H "Accept: application/json" \
-H "Content-Type: application/json" \
-k 'https://*<RANGER SERVER ADDRESS>*:6182/service/public/v2/api/servicedef'
```

Se esse comando for executado com êxito, você verá um novo serviço em sua interface de usuário do Ranger Admin chamado "AMAZON- EMR - SPARK “, conforme mostrado na imagem a seguir (a versão 2.0 do Ranger é mostrada).

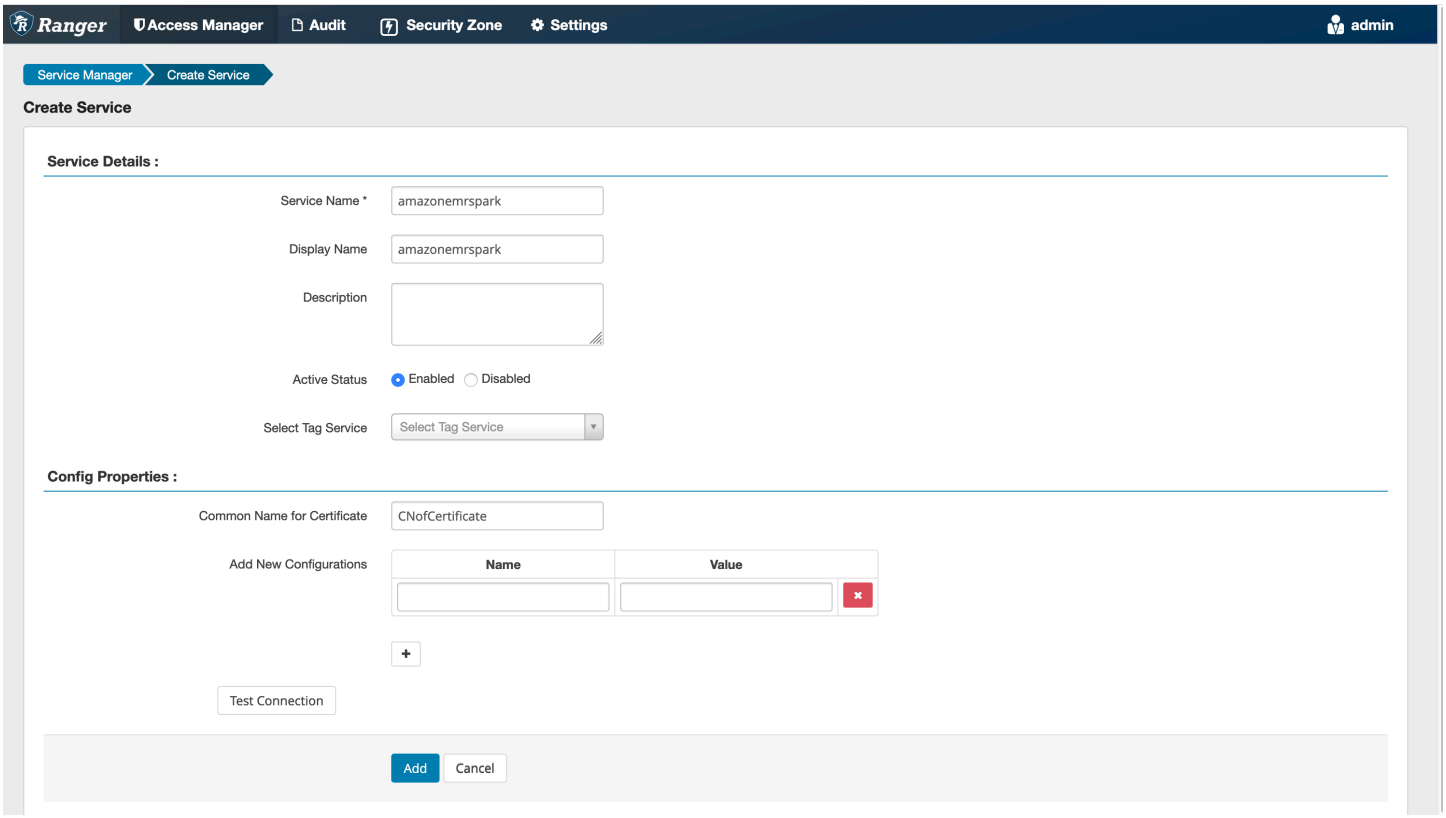


Etapa 5: criar uma instância do SPARK aplicativo AMAZON EMR - -

Nome do serviço (se for exibido): o nome do serviço que será usado. O valor sugerido é **`amazonemrspark`**. Anote esse nome de serviço, pois ele será necessário ao criar uma configuração EMR de segurança.

Nome de exibição: o nome a ser exibido para a instância. O valor sugerido é **`amazonemrspark`**.

Nome comum para certificado: o campo CN dentro do certificado usado para se conectar ao servidor de administração com base em um plug-in cliente. Esse valor deve corresponder ao campo CN em seu TLS certificado que foi criado para o plug-in.



> **Note**
>
> O TLS certificado desse plug-in deveria ter sido registrado no armazenamento confiável do servidor Ranger Admin. Consulte TLScertificados para obter mais detalhes.

Criação de políticas do Spark SQL

Ao criar uma nova política, os campos a serem preenchidos são:

Nome da política: o nome da política.

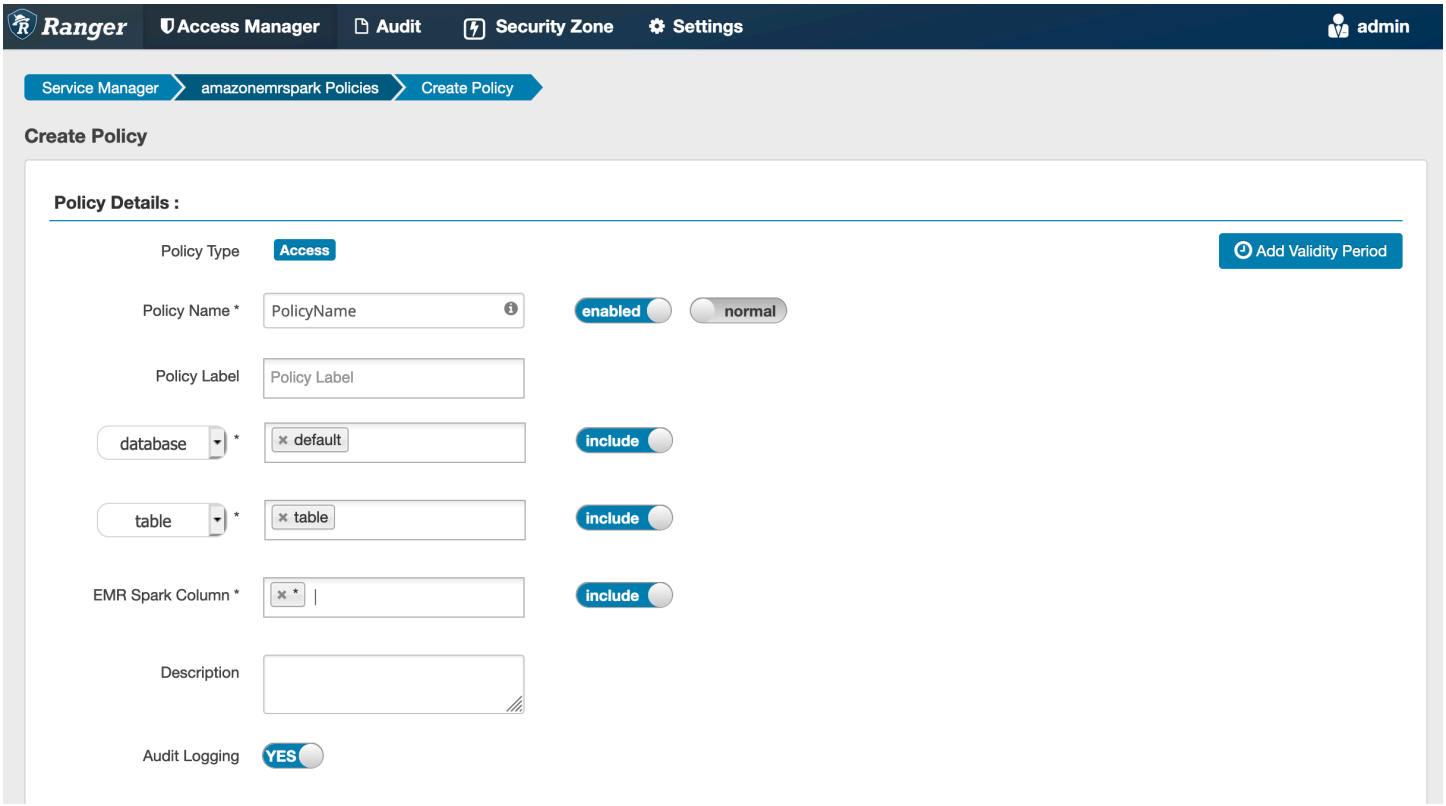
Rótulo de política: um rótulo que você pode colocar na política.

Banco de dados: o banco de dados ao qual a política se aplica. O caractere curinga "*" representa todos os bancos de dados.

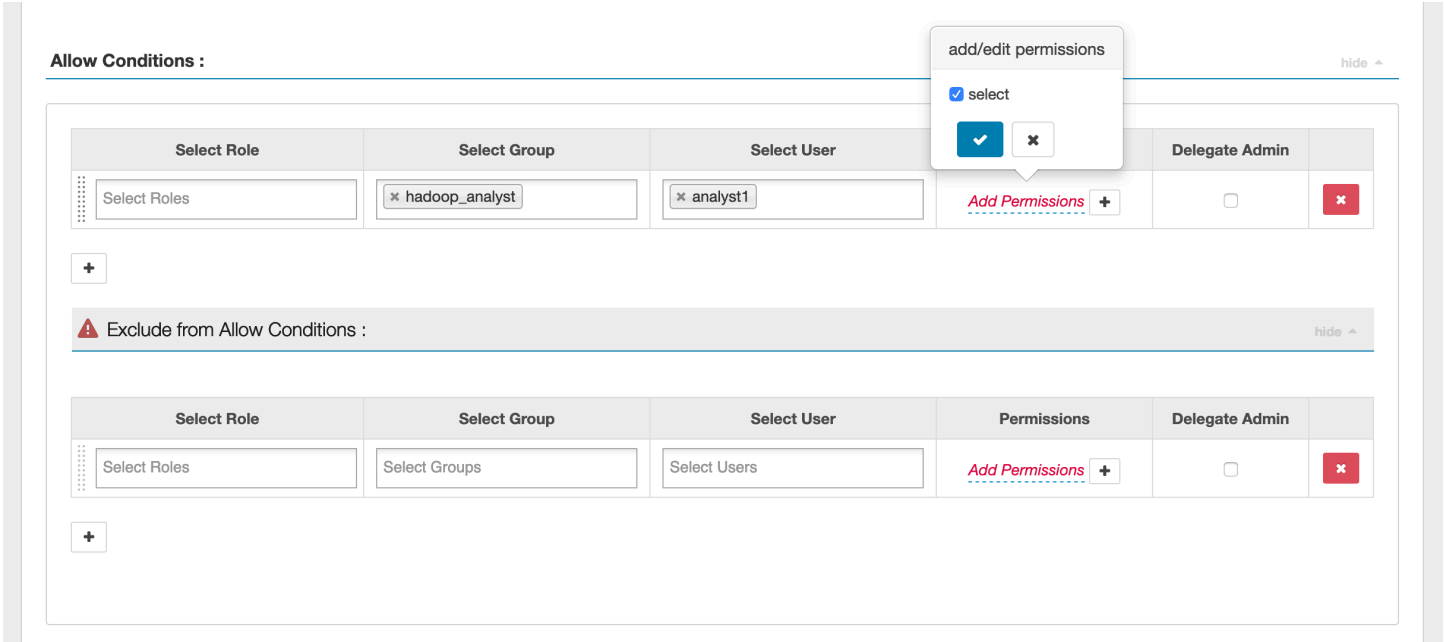
Tabela: as tabelas às quais a política se aplica. O caractere curinga "*" representa todas as tabelas.

EMRColuna do Spark: as colunas às quais essa política se aplica. O caractere curinga "*" representa todas as colunas.

Descrição: uma descrição da política.



Para especificar usuários e grupos, insira os usuários e grupos abaixo para conceder permissões. Também é possível especificar exclusões para as condições de permissão e negação.



Após especificar as condições de permitir e negar, clique em Salvar.

Considerações

Cada nó dentro do EMR cluster deve ser capaz de se conectar ao nó principal na porta 9083.

Limitações

Estas são as limitações atuais do plug-in Apache Spark:

- O Record Server sempre se conectará à HMS execução em um EMR cluster da Amazon. Configure HMS para se conectar ao Modo Remoto, se necessário. Você não deve colocar valores de configuração no arquivo de configuração Hive-site.xml do Apache Spark.
- Tabelas criadas usando fontes de dados do Spark no CSV ou Avro não podem ser lidas usando. EMR RecordServer Utilize o Hive para criar e gravar dados e ler usando Record.
- Não há suporte para tabelas Delta Lake e Hudi.
- Os usuários precisam ter acesso ao banco de dados padrão. Esse é um requisito do Apache Spark.
- O servidor Ranger Admin não oferece suporte ao preenchimento automático.
- O SQL plug-in Spark para Amazon EMR não oferece suporte a filtros de linha ou mascaramento de dados.
- Ao usar ALTER TABLE com o SparkSQL, um local de partição deve ser o diretório filho de um local de tabela. Não há suporte para inserção de dados em uma partição em que a localização da partição seja diferente da localização da tabela.

EMRFSPlug-in S3

Para facilitar o fornecimento de controles de acesso contra objetos no S3 em um cluster multilocatário, o plug-in do EMRFS S3 fornece controles de acesso aos dados no S3 ao acessá-los por meio dele. EMRFS Você pode permitir acesso aos recursos do S3 em nível de usuário e grupo.

Para conseguir isso, quando seu aplicativo tenta acessar dados no S3, EMRFS envia uma solicitação de credenciais para o processo do Agente Secreto, onde a solicitação é autenticada e autorizada em um plug-in Apache Ranger. Se a solicitação for autorizada, o Agente Secreto assume a IAM função de Apache Ranger Engines com uma política restrita para gerar credenciais que só têm acesso à política Ranger que permitiu o acesso. As credenciais são então repassadas EMRFS para acessar o S3.

Tópicos

Atributos compatíveis

EMRFSO plug-in S3 fornece autorização de nível de armazenamento. Políticas podem ser criadas para conceder acesso a usuários e grupos a buckets e prefixos do S3. A autorização é feita somente contraEMRFS.

Instalação da configuração de serviço

Para instalar a definição do EMRFS serviço, você deve configurar o servidor Ranger Admin. Para configurar o servidor, consulteConfigurar o servidor do Ranger Admin.

Siga estas etapas para instalar a definição EMRFS de serviço.

Etapa 1: SSH no servidor Apache Ranger Admin.

Por exemplo:

```
ssh ec2-user@ip-xxx-xxx-xxx-xxx.ec2.internal
```

Etapa 2: Baixe a definição do EMRFS serviço.

Em um diretório temporário, baixe a definição de EMR serviço da Amazon. Essa definição de serviço é compatível com as versões Ranger 2.x.

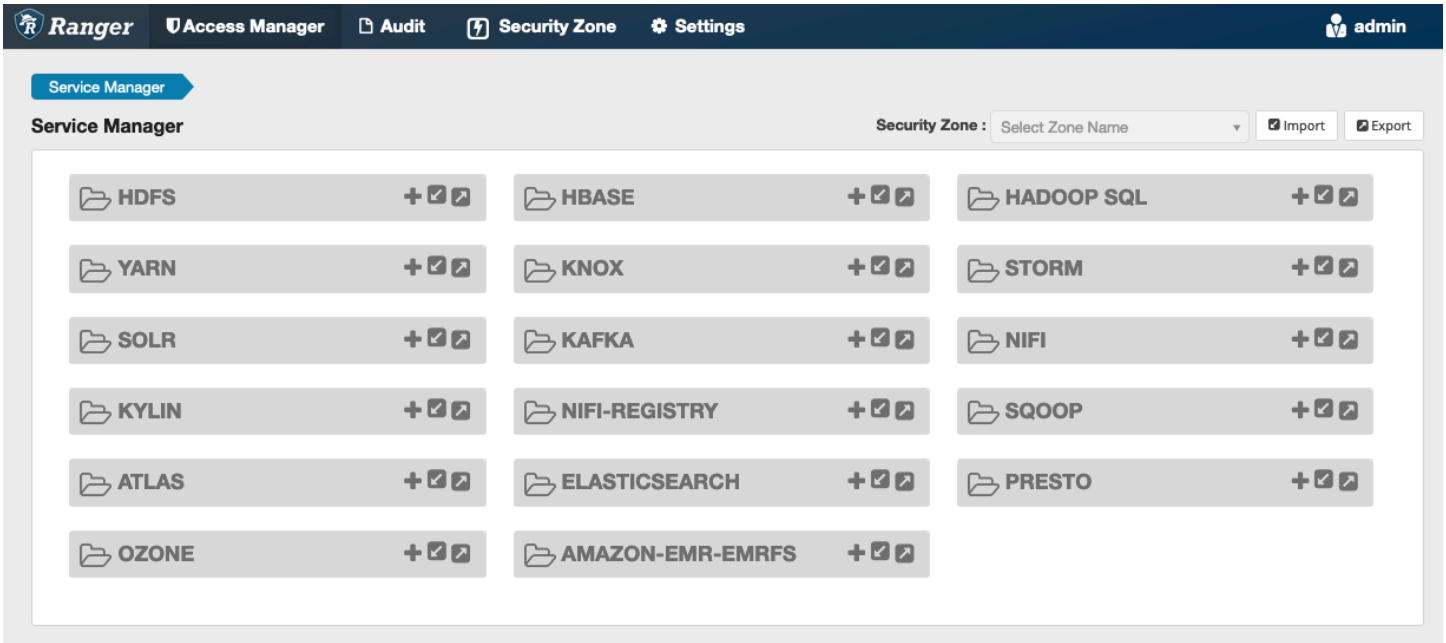
```
wget https://s3.amazonaws.com/elasticmapreduce/ranger/service-definitions/version-2.0/
ranger-servicedef-amazon-emr-emrfs.json
```

Etapa 3: Registrar a definição do serviço EMRFS S3.

```
curl -u *<admin users login>*:*_<_**_password_ **_for_** _ranger admin user_**_>_* -X
 POST -d @ranger-servicedef-amazon-emr-emrfs.json \
```

```
-H "Accept: application/json" \
-H "Content-Type: application/json" \
-k 'https://*<RANGER SERVER ADDRESS>*:6182/service/public/v2/api/servicedef'
```

Se esse comando for executado com êxito, você verá um novo serviço na interface do usuário do Ranger Admin chamado "AMAZON- EMR -S3", conforme mostrado na imagem a seguir (a versão 2.0 do Ranger é mostrada).



Etapa 4: Crie uma instância do EMRFS aplicativo AMAZON EMR - -.

Crie uma instância da definição de serviço.

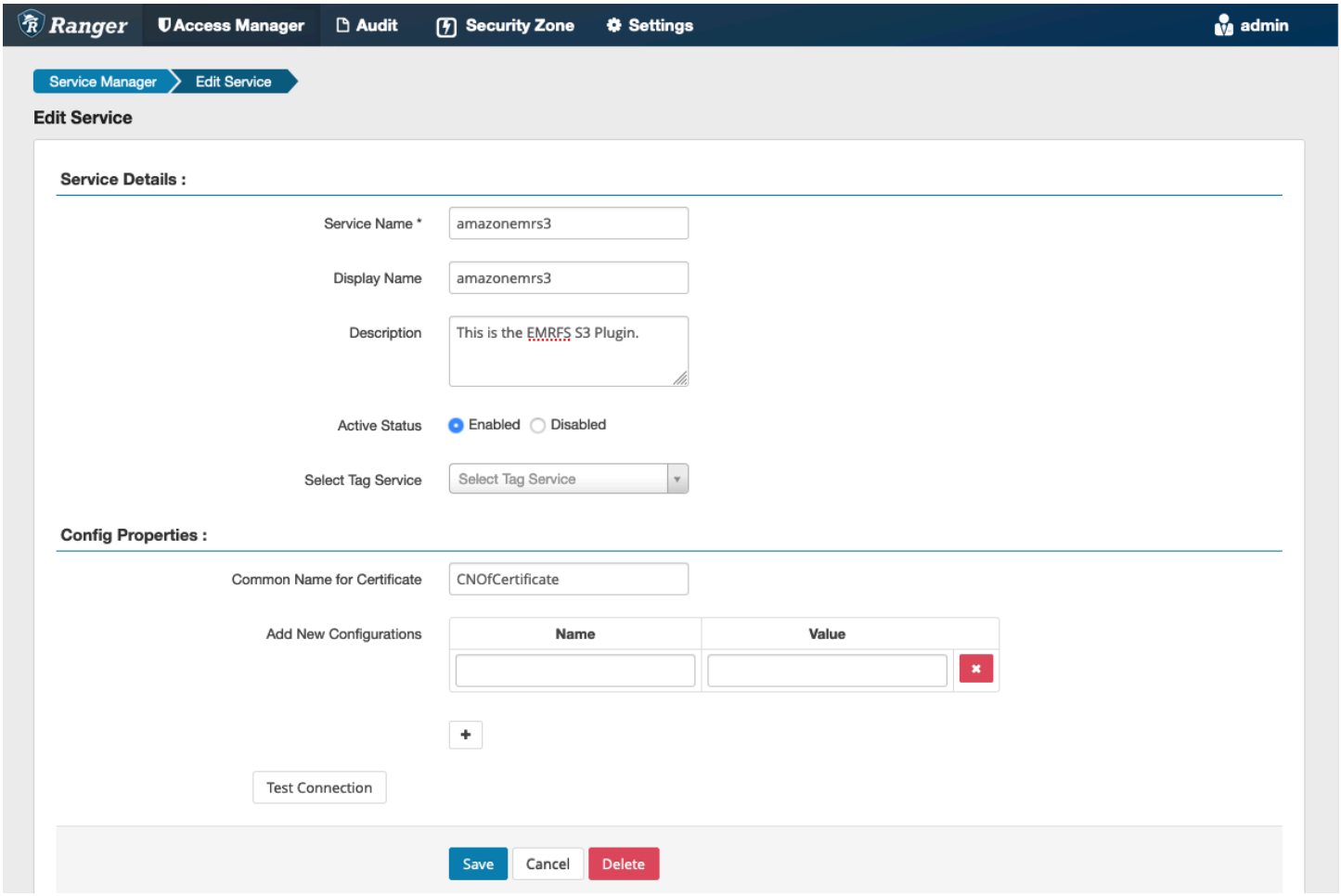
- Clique no + ao lado de AMAZON - EMR -EMRFS.

Preencha os seguintes campos:

Nome do serviço (se for exibido): o valor sugerido é **`amazonemrspark`**. Anote esse nome de serviço, pois ele será necessário ao criar uma configuração EMR de segurança.

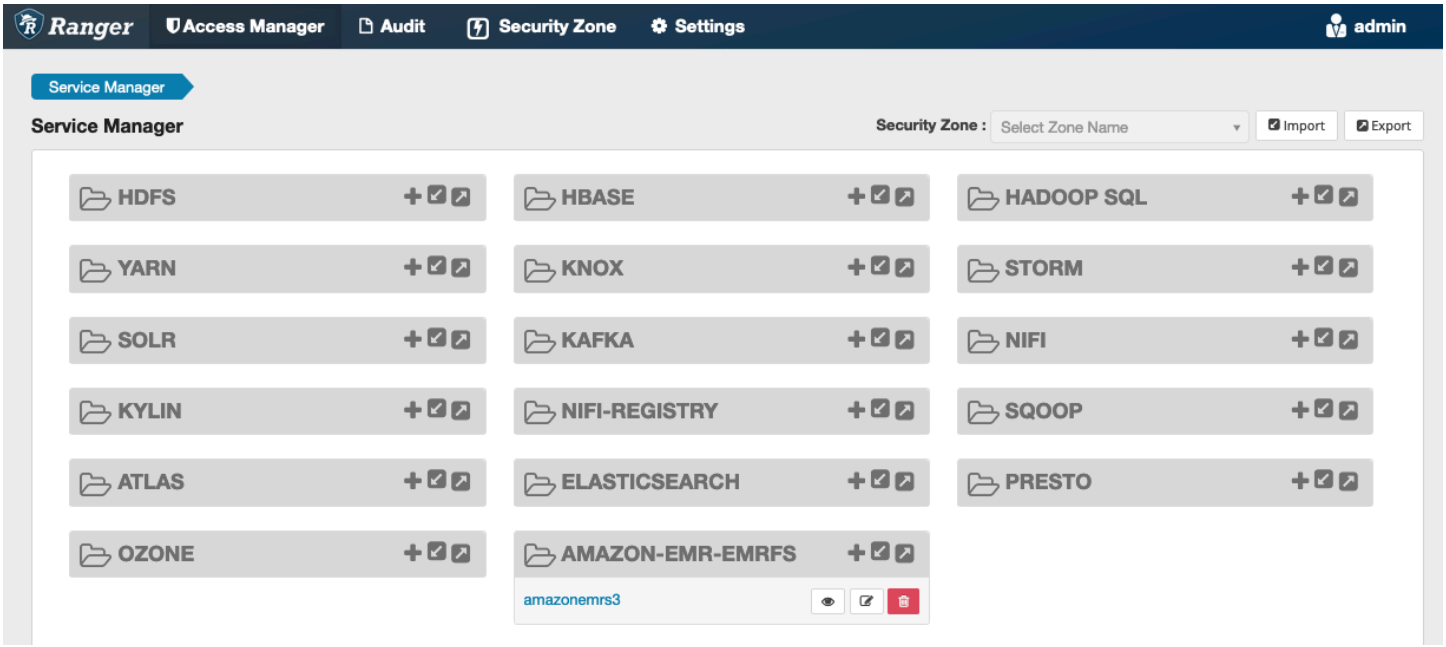
Nome de exibição: o nome exibido para o serviço. O valor sugerido é **`amazonemrspark`**.

Nome comum para certificado: o campo CN dentro do certificado usado para se conectar ao servidor de administração com base em um plug-in cliente. Esse valor deve corresponder ao campo CN no TLS certificado que foi criado para o plug-in.



> **Note**
>
> O TLS certificado desse plug-in deveria ter sido registrado no armazenamento confiável do servidor Ranger Admin. Consulte TLScertificados para obter mais detalhes.

Quando o serviço é criado, o Service Manager inclui "AMAZON- EMR - EMRFS “, conforme mostrado na imagem a seguir.



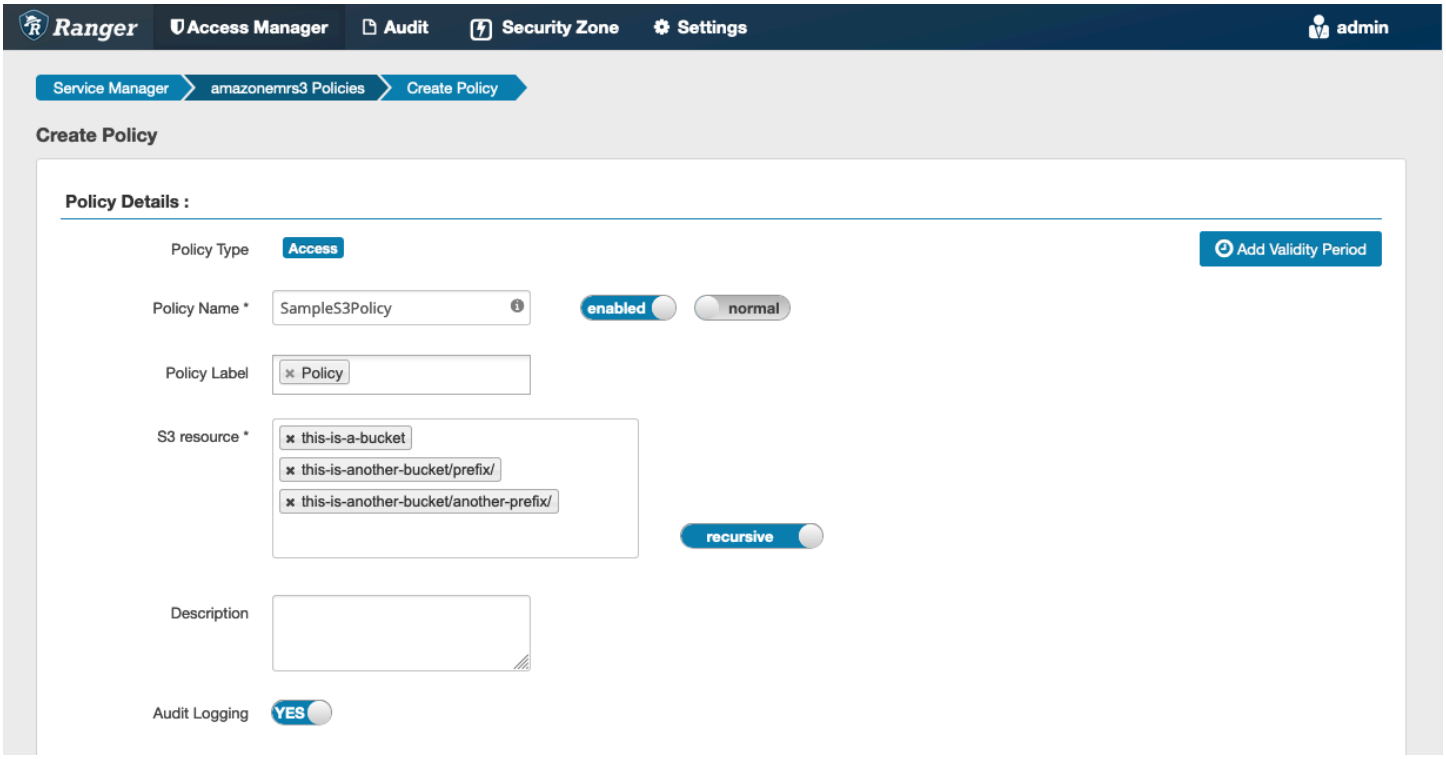
Criação de EMRFS políticas do S3

Para criar uma nova política na página Criar política do Service Manager, preencha os campos a seguir.

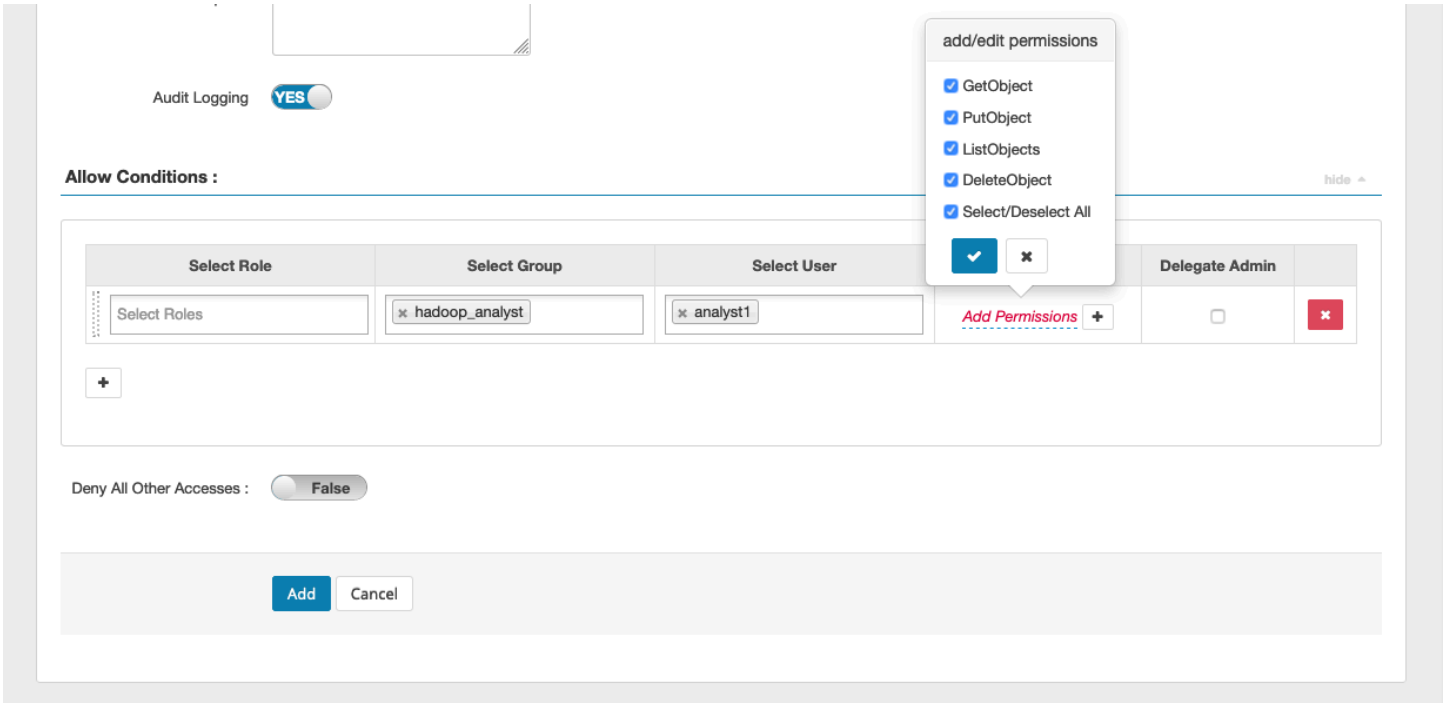
Nome da política: o nome da política.

Rótulo de política: um rótulo que você pode colocar na política.

Recurso do S3: um recurso que começa com o bucket e o prefixo opcional. Consulte EMRFSNotas de uso das políticas do S3 para obter informações sobre práticas recomendadas. Os recursos no servidor Ranger Admin não devem conter **`s3://`**, **`s3a://`** ou **`s3n://`**.



É possível especificar usuários e grupos para conceder permissões. Também é possível especificar exclusões para condições de permissão e negação.



> **Note**
>
> São permitidos no máximo três recursos por política. Adicionar mais de três recursos pode resultar em um erro quando essa política é usada em um EMR cluster. Adicionar mais de três políticas exibirá um lembrete sobre o limite da política.

EMRFSNotas de uso das políticas do S3

Ao criar políticas do S3 no Apache Ranger, atente para algumas considerações sobre o uso.

Permissões para múltiplos objetos do S3

É possível usar políticas recursivas e expressões curinga para conceder permissões a vários objetos do S3 com prefixos comuns. As políticas recursivas concedem permissões a todos os objetos com um prefixo comum. As expressões curinga selecionam múltiplos prefixos. Juntos, eles concedem permissões a todos os objetos com múltiplos prefixos comuns, conforme mostrado nos exemplos a seguir.

Example Usar uma política recursiva

Suponha que você queira permissões para listar todos os arquivos parquet em um bucket do S3 organizado da forma a seguir.

```
s3://sales-reports/americas/
    +- year=2000
    |      +- data-q1.parquet
    |      +- data-q2.parquet
    +- year=2019
    |      +- data-q1.json
    |      +- data-q2.json
    |      +- data-q3.json
    |      +- data-q4.json
    |
    +- year=2020
    |      +- data-q1.parquet
    |      +- data-q2.parquet
    |      +- data-q3.parquet
    |      +- data-q4.parquet
    |      +- annual-summary.parquet
    +- year=2021
```

Primeiro, considere os arquivos parquet que tenham o prefixo `s3://sales-reports/americas/year=2000`. Você pode conceder GetObject permissões a todos eles de duas maneiras:

Usar políticas não recursivas: uma opção é usar duas políticas não recursivas separadas, uma para o diretório e outra para os arquivos.

A primeira política concede permissão ao prefixo `s3://sales-reports/americas/year=2020` (não há / final).

```
- S3 resource = "sales-reports/americas/year=2000"
- permission = "GetObject"
- user = "analyst"
```

A segunda política usa a expressão curinga para conceder permissões a todos os arquivos com prefixo `sales-reports/americas/year=2020/` (observe o / final).

```
- S3 resource = "sales-reports/americas/year=2020/*"
- permission = "GetObject"
- user = "analyst"
```

Usar uma política recursiva: uma alternativa mais conveniente é usar uma única política recursiva e conceder permissão recursiva ao prefixo.

```
 - S3 resource = "sales-reports/americas/year=2020"
 - permission = "GetObject"
 - user = "analyst"
 - is recursive = "True"
```

Até agora, apenas os arquivos parquet com o prefixo `s3://sales-reports/americas/year=2000` foram incluídos. Também já é possível incluir os arquivos parquet com outro prefixo, `s3://sales-reports/americas/year=2020`, na mesma política recursiva introduzindo uma expressão curinga da forma a seguir.

```
 - S3 resource = "sales-reports/americas/year=20?0"
 - permission = "GetObject"
 - user = "analyst"
 - is recursive = "True"
```

Políticas PutObject e DeleteObject permissões

Escrever políticas `PutObject` e `DeleteObject` permissões para arquivos EMRFS precisa de cuidados especiais porque, diferentemente das GetObject permissões, elas precisam de permissões recursivas adicionais concedidas ao prefixo.

Example Políticas PutObject e DeleteObject permissões

Por exemplo, excluir o arquivo `annual-summary.parquet` requer não apenas uma DeleteObject permissão para o arquivo real.

```
- S3 resource = "sales-reports/americas/year=2020/annual-summary.parquet"
- permission = "DeleteObject"
- user = "analyst"
```

Também requer uma política que conceda permissões `GetObject` e `PutObject` recursivas para o prefixo.

Da mesma forma, modificar o arquivo `annual-summary.parquet` requer não apenas uma permissão `PutObject` para o arquivo real.

```
- S3 resource = "sales-reports/americas/year=2020/annual-summary.parquet"
- permission = "PutObject"
- user = "analyst"
```

Também requer uma política que conceda a permissão `GetObject` recursiva para o prefixo.

```
- S3 resource = "sales-reports/americas/year=2020"
- permission = "GetObject"
- user = "analyst"
- is recursive = "True"
```

Curingas em políticas

Há duas áreas em que é possível especificar caracteres curingas. Ao especificar um recurso do S3, pode-se usar “*” e “?”. O “*” faz correspondência com um caminho do S3 e corresponde a tudo que está depois do prefixo. Por exemplo, a política a seguir.

```
S3 resource = "sales-reports/americas/*"
```

Isso corresponde aos caminhos do S3 a seguir.

```
sales-reports/americas/year=2020/
sales-reports/americas/year=2019/
sales-reports/americas/year=2019/month=12/day=1/afile.parquet
sales-reports/americas/year=2018/month=6/day=1/afile.parquet
sales-reports/americas/year=2017/afile.parquet
```

O curinga “?” corresponde a apenas um caractere. Por exemplo, para a política.

```
S3 resource = "sales-reports/americas/year=201?/"
```

Isso corresponde aos caminhos do S3 a seguir.

```
sales-reports/americas/year=2019/
sales-reports/americas/year=2018/
sales-reports/americas/year=2017/
```

Curingas em usuários

Há dois curingas integrados ao atribuir usuários para fornecer acesso aos usuários. O primeiro é o caractere curinga “{USER}” que fornece acesso a todos os usuários. O segundo caractere curinga é “{OWNER}”, que fornece acesso direto ao proprietário de um objeto específico. No entanto, o caractere curinga “{USER}” não é suportado atualmente.

Limitações

A seguir estão as limitações atuais do plug-in EMRFS S3:

- As políticas do Apache Ranger podem conter no máximo três políticas.
- O acesso ao S3 deve ser feito por meio de aplicativos relacionados ao Hadoop EMRFS e pode ser usado com eles. Não há suporte para:

  - Bibliotecas Boto3

  - AWS SDK e AWK CLI

  - Conector de código aberto S3A
- Não há suporte para políticas de negação do Apache Ranger.
- Operações no S3 com chaves com KMS criptografia CSE - atualmente não são suportadas.
- O suporte entre regiões não é compatível.

- Não há suporte para o atributo de zona de segurança do Apache Ranger. As restrições de controle de acesso definidas usando o recurso Security Zone não são aplicadas aos seus EMR clusters da Amazon.
- O usuário do Hadoop não gera nenhum evento de auditoria, pois o Hadoop sempre acessa o Perfil da Instância. EC2
- É recomendável que você desative o Amazon EMR Consistency View. O S3 do tem um alto nível de consistência e, portanto, isso não é mais necessário. Para obter mais informações, consulte Amazon S3 strong consistency.
- O plug-in EMRFS S3 faz várias STS chamadas. É recomendável que você faça testes de carga em uma conta de desenvolvimento e monitore o volume de STS chamadas. Também é recomendável que você faça uma STS solicitação para aumentar os limites do AssumeRole serviço.
- O servidor Ranger Admin não oferece suporte ao preenchimento automático.

Plug-in Trino

O Trino (anteriormente PrestoSQL) é um mecanismo de SQL consulta que você pode usar para executar consultas em fontes de dados como armazenamento de objetosHDFS, bancos de dados relacionais e nenhum banco de dados. SQL Ele elimina a necessidade de migrar dados para um local central e permite que você consulte os dados de onde quer que estejam. EMRA Amazon fornece um plug-in Apache Ranger para fornecer controles de acesso refinados para o Trino. O plug-in é compatível com o servidor Apache Ranger Admin de código aberto versão 2.0 e posteriores.

Tópicos

Atributos compatíveis

O plug-in Apache Ranger para Trino na Amazon EMR oferece suporte a todas as funcionalidades do mecanismo de consulta Trino, que é protegido por um controle de acesso refinado. Isso inclui controles de acesso em nível de banco de dados, de tabela e de coluna, filtragem de linhas e

mascaramento de dados. As políticas do Apache Ranger podem incluir políticas de concessão e políticas de negação para usuários e grupos. Os eventos de auditoria também são enviados aos CloudWatch registros.

Instalação da configuração de serviço

A instalação da definição de serviço Trino requer que o servidor Ranger Admin esteja configurado. Para configurar o servidor Ranger Admin, consulte Configurar o servidor do Ranger Admin.

Siga estas etapas para instalar a definição de serviço do Trino.

1. SSHno servidor Apache Ranger Admin.

```
ssh ec2-user@ip-xxx-xxx-xxx-xxx.ec2.internal
```

2. Desinstale o plug-in do servidor Presto, se houver. Execute o seguinte comando . Se isso ocorrer com o erro “Serviço não encontrado”, significa que o plug-in do servidor Presto não foi instalado no servidor. Prossiga para a próxima etapa.

```
curl -f -u *<admin users login>*:*_<_**_password_ **_for_** _ranger admin
 user_**_>_* -X DELETE -k 'https://*<RANGER SERVER ADDRESS>*:6182/service/public/
v2/api/servicedef/name/presto'
```

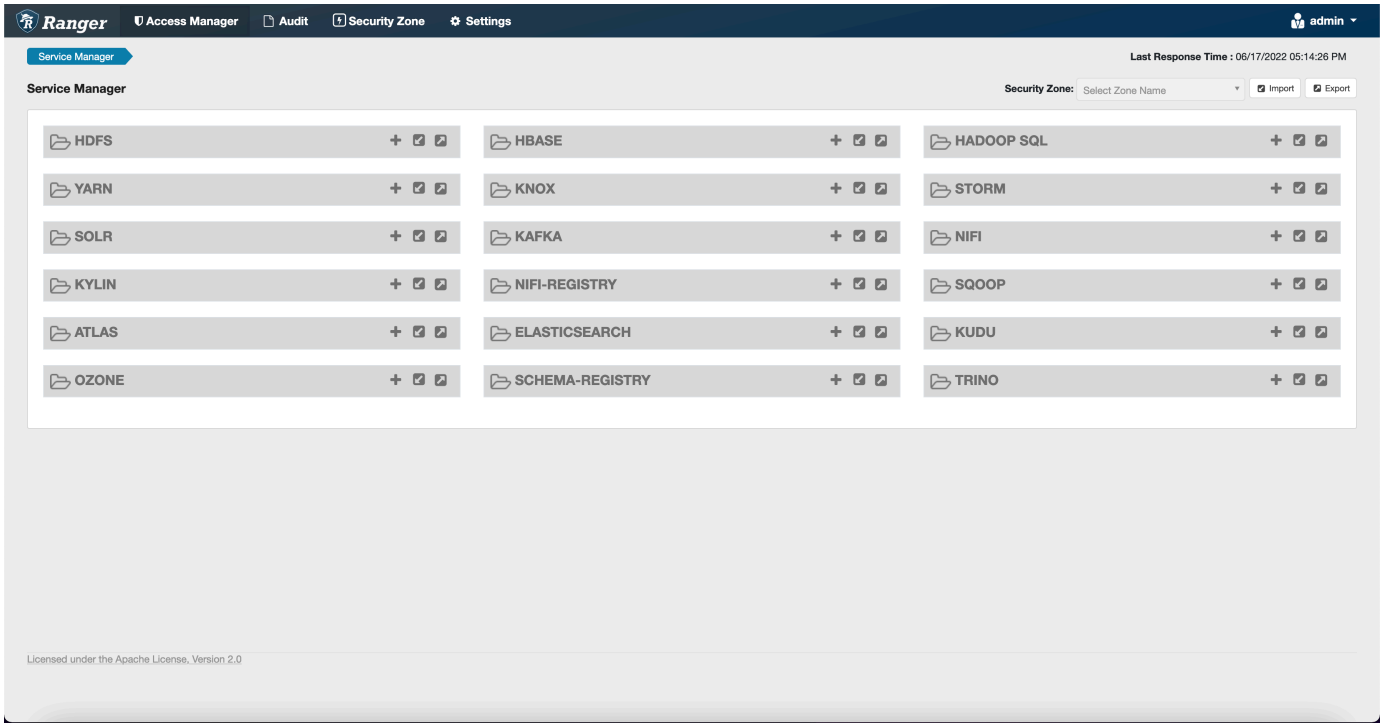
3. Baixe a definição de serviço e o plug-in do servidor Apache Ranger Admin. Em um diretório temporário, baixe a definição de serviço. Essa definição de serviço é compatível com as versões Ranger 2.x.

```
wget https://s3.amazonaws.com/elasticmapreduce/ranger/service-definitions/
version-2.0/ranger-servicedef-amazon-emr-trino.json
```

4. Registre a definição do serviço Apache Trino para a Amazon. EMR

```
curl -u *<admin users login>*:*_<_**_password_ **_for_** _ranger admin user_**_>_*
 -X POST -d @ranger-servicedef-amazon-emr-trino.json \
-H "Accept: application/json" \
-H "Content-Type: application/json" \
-k 'https://*<RANGER SERVER ADDRESS>*:6182/service/public/v2/api/servicedef'
```

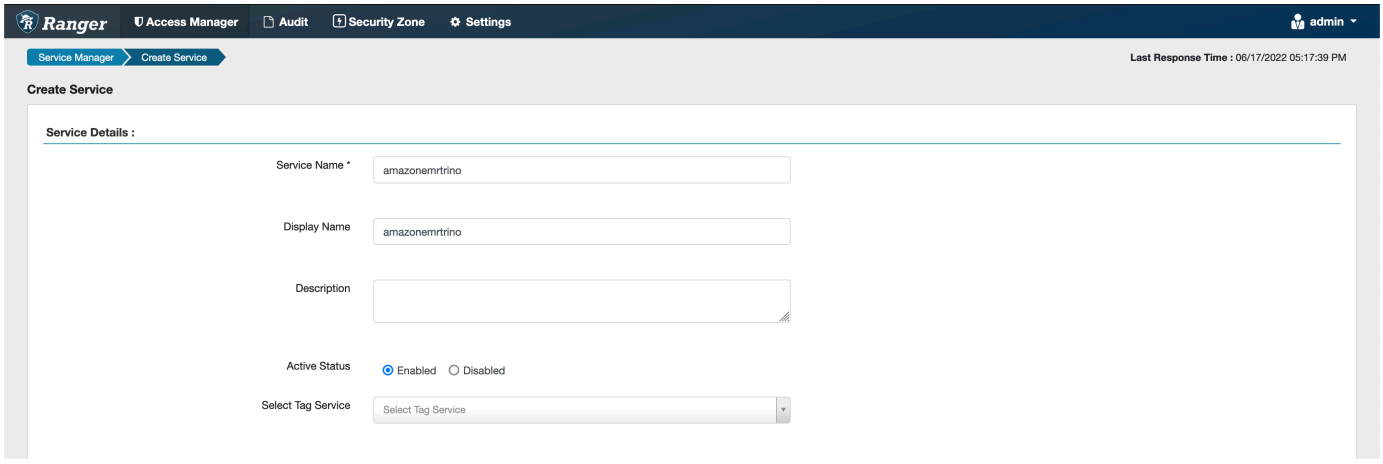
   Se esse comando for executado com êxito, você verá um novo serviço na IU do Ranger Admin chamado `TRINO`, conforme mostrado na imagem.



5. Crie uma instância da aplicação `TRINO`, inserindo as informações a seguir.

   Nome do serviço: o nome do serviço que você usará. O valor sugerido é `amazonemrtrino`. Anote esse nome de serviço, pois ele será necessário ao criar uma configuração de EMR segurança da Amazon.

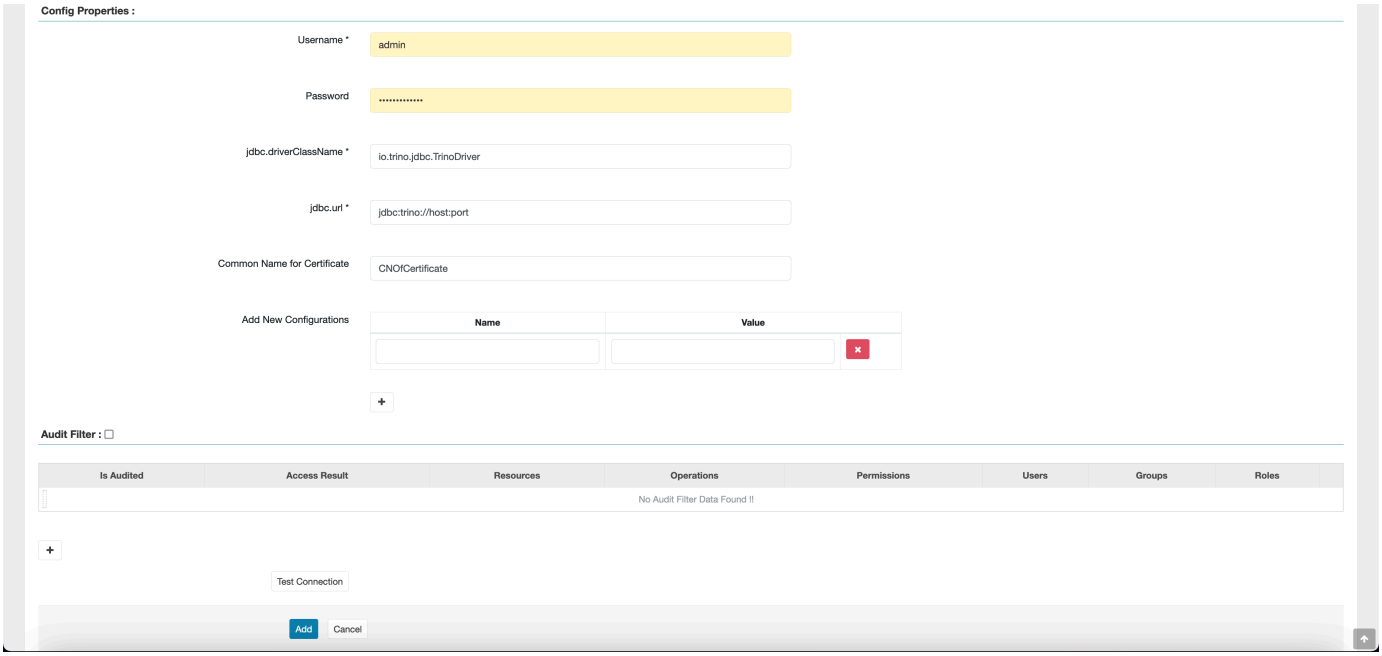
   Nome de exibição: o nome a ser exibido para a instância. O valor sugerido é `amazonemrtrino`.



   jdbc. driver. ClassName: O nome da JDBC classe para conectividade Trino. Você pode usar o valor padrão.

   jdbc.url: A string de JDBC conexão a ser usada ao se conectar ao coordenador Trino.

Nome comum para certificado: o campo CN dentro do certificado usado para se conectar ao servidor de administração com base em um plug-in cliente. Esse valor deve corresponder ao campo CN em seu TLS certificado que foi criado para o plug-in.



Observe que o TLS certificado desse plug-in deveria ter sido registrado no armazenamento confiável do servidor Ranger Admin. Para obter mais informações, consulte TLScertificados.

Criar políticas do Trino

Ao criar uma nova política, preencha os campos a seguir.

Nome da política: o nome da política.

Rótulo de política: um rótulo que você pode colocar na política.

Catálogo: o catálogo ao qual a política se aplica. O curinga “*” representa todos os catálogos.

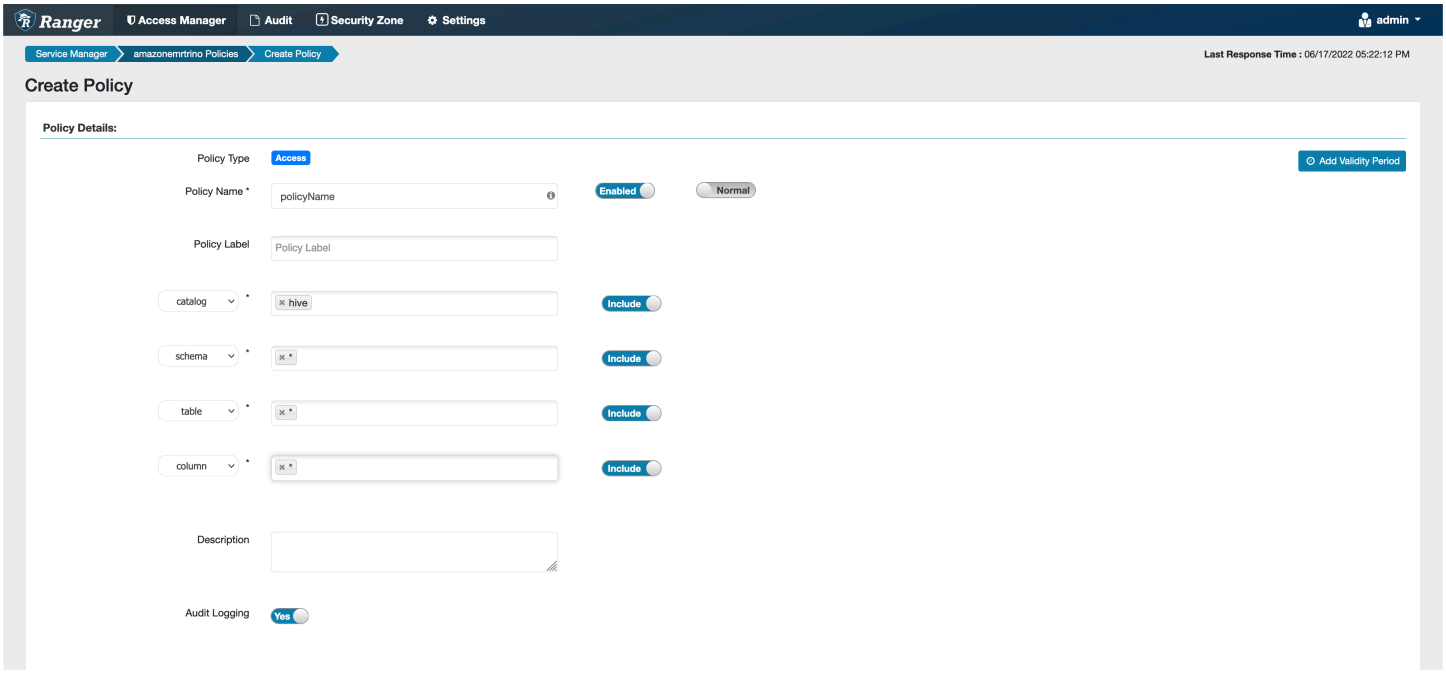
Esquema: os esquemas aos quais a política se aplica. O curinga “*” representa todos os esquemas.

Tabela: as tabelas às quais a política se aplica. O caractere curinga “*” representa todas as tabelas.

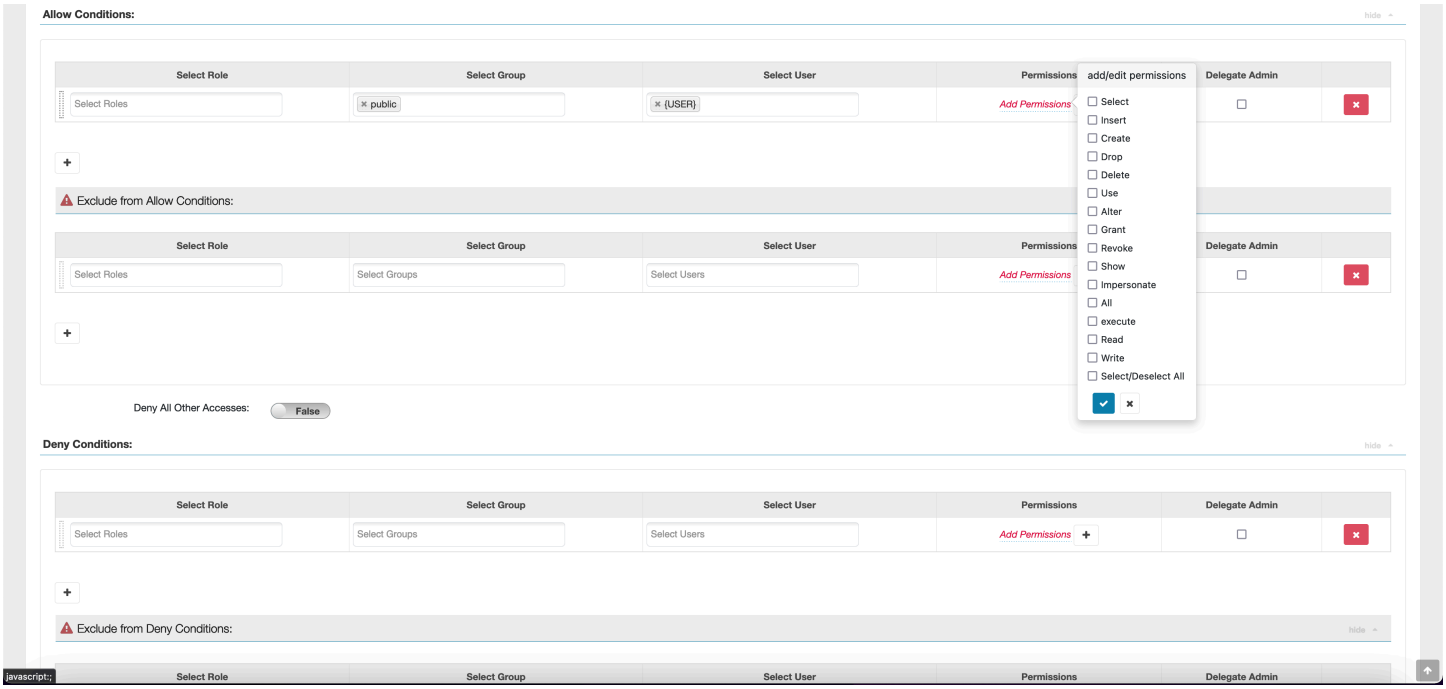
Coluna: as colunas às quais a política se aplica. O caractere curinga “*” representa todas as colunas.

Descrição: uma descrição da política.

Existem outros tipos de políticas para o usuário Trino (para acesso à representação do usuário), a propriedade do sistema/sessão Trino (para alterar o sistema do mecanismo ou as propriedades da sessão), funções/procedimentos (para permitir chamadas de função ou procedimento) e URL(para conceder acesso de leitura/gravação ao mecanismo em locais de dados).



Para conceder permissões a usuários e grupos específicos, insira os usuários e grupos. Também é possível especificar exclusões para condições de permissão e negação.



Após especificar as condições de permitir e negar, escolha Salvar.

Considerações

Ao criar políticas do Trino no Apache Ranger, atente para algumas considerações sobre o uso.

Servidor de metadados Hive

O servidor de metadados Hive só pode ser acessado por mecanismos confiáveis, especificamente o mecanismo Trino, para proteção contra acesso não autorizado. O servidor de metadados Hive também é acessado por todos os nós do cluster. A porta 9083 necessária fornece acesso de todos os nós ao nó principal.

Autenticação

Por padrão, o Trino está configurado para se autenticar usando o Kerberos conforme configurado na configuração de segurança da Amazon. EMR

A criptografia em trânsito é obrigatória

O plug-in Trino exige que você tenha a criptografia em trânsito ativada na configuração de EMR segurança da Amazon. Para ativar a criptografia, consulte Criptografia em trânsito.

Limitações

Estas são as limitações atuais do plug-in Trino:

- O servidor Ranger Admin não oferece suporte ao preenchimento automático.

Solução de problemas do Apache Ranger

Aqui estão alguns problemas diagnosticados com frequência relacionados ao uso do Apache Ranger.

Recomendações

- Teste usando um único cluster de nó principal: clusters principais de nó único são provisionados mais rapidamente do que um cluster de múltiplos nós, o que pode diminuir o tempo de cada iteração de teste.
- Defina o modo de desenvolvimento no cluster. Ao iniciar seu EMR cluster, defina o `--additional-info"` parâmetro como:

```
'{"clusterType":"development"}'
```

Esse parâmetro só pode ser definido por meio do AWS CLI ou AWS SDK e não está disponível no EMR console da Amazon. Quando esse sinalizador é definido e o mestre falha no provisionamento, o EMR serviço da Amazon mantém o cluster ativo por algum tempo antes de desativá-lo. Esse momento é muito útil para testar vários arquivos de log antes que o cluster seja terminado.

EMRfalha no provisionamento do cluster

Há vários motivos pelos quais um EMR cluster da Amazon pode falhar ao iniciar. Veja aqui algumas maneiras de diagnosticar o problema.

Verifique os registros de EMR aprovisionamento

A Amazon EMR usa o Puppet para instalar e configurar aplicativos em um cluster. A análise dos logs fornecerá detalhes sobre a ocorrência de erros durante a fase de provisionamento de um cluster. Os logs podem ser acessados no cluster ou no S3 se os logs estiverem configurados para serem enviados ao S3.

Os logs são armazenados em `/var/log/provision-node/apps-phase/0/{UUID}/puppet.log` no disco e em `s3://<LOG LOCATION>/<CLUSTER ID>/node/<EC2 INSTANCE ID>/provision-node/apps-phase/0/{UUID}/puppet.log.gz`.

Mensagens de erro comuns

Mensagem de erro	Causa
Puppet (err): Falha na inicialização do Systemd! emr-record-server registro journalctl para: emr-record-server	EMRFalha ao iniciar o servidor de registros. Consulte os EMR registros do Record Server abaixo.
Puppet (err): Falha na inicialização do Systemd! emr-record-server registro journalctl para emrsecretagent:	EMRO agente secreto não conseguiu iniciar. Veja abaixo os logs do agente secreto.

Mensagem de erro	Causa
/Stage [main] /Ranger_plugins: :ranger_h ive_plugin/Ranger_plugins: :prepare_two_way_t ls [configure bidirecional TLS no plug-in Hive] /Exec [crie armazenamento de chaves e armazenamento confiável para o plug-in Ranger Hive] /returns (aviso): 140408606 197664:error:0906d06c: rotinas: _read_bio_: sem linha de partida: pem_lib.c:707:Esperando: PEM PEM ANY PRIVATE KEY	O TLS certificado privado no Secret Manager para o certificado do plug-in Apache Ranger não está no formato correto ou não é um certificado privado. Consulte TLScertificados para ver os formatos de certificado.
/Stage [main] /Ranger_plugins: :ranger_s 3_plugin/Ranger_plugins: :prepare_two_way_t ls [configure bidirecional TLS no plugin Ranger s3] /Exec [criar keystore e truststore para o plugin Ranger 3] /returns (aviso): Ocorreu um erro () ao chamar a operação: Usuário: arn:aws:sts: ::assumed-role/ _ _ /i- não está autorizado a executar: secretsmanager: on resource: amazon-emr-s arn:aws:secretsman ager:us-east-1 ::secret: - AccessDeniedExcept ion GetSecretValue XXXXXXXXXXX EMR EC2 DefaultRole XXXXXXXXXXX GetSecretValue XXXXXXXXXX AdminServer XXXXX	A função de perfil da EC2 instância não tem as permissões corretas para recuperar os TLS certificados do Secrets Agent.

Verifique SecretAgent os registros

Os registros do Secret Agent estão localizados `/emr/secretagent/log/` em um EMR nó ou no `s3://<LOG LOCATION>/<CLUSTER ID>/node/<EC2 INSTANCE ID>/daemons/secretagent/` diretório no S3.

Mensagens de erro comuns

Mensagem de erro	Causa

Mensagem de erro	Causa
Exceção no tópico “main” com.amazonaws.services.securitytoken.model. AWSSecurityTokenServiceException: O usuário: arn:aws:sts: ::assumed-role/ XXXXXXXXXXXX _ EMR _ EC2 /i- XXXXXXXXXXXXXX não DefaultRole está autorizado a executar: sts: no AssumeRole recurso: arn:aws:iam: ::role/* XXXXXXXXX XXX (Serviço:; Código de status: 403; Código de erro:; ID da solicitação: - - - -; Proxy: null) RangerPluginDataAccessRole AWSSecurityTokenService AccessDenied XXXXXXXX XXXX XXXX XXXX XXXXXXXXXXX	A exceção acima significa que a função do perfil da EMR EC2 instância não tem permissões para assumir a função RangerPluginDataAccessRole. Consulte IAMfunções para integração nativa com o Apache Ranger.
ERRORqtp54617902-149: Ocorreu uma exceção no aplicativo Web javax.ws.rs. NotAllowedException: Método HTTP 405 não permitido	Esses erros podem ser ignorados com segurança.

Verifique os registros do servidor (para o SparkSQL)

EMROs registros do Record Server estão disponíveis em /var/log/emr-record-server/em um EMR nó ou podem ser encontrados no diretório s3: //< LOG LOCATION >/< id>/node/< CLUSTER id>/daemons//no S3. EC2 INSTANCE emr-record-server

Mensagens de erro comuns

Mensagem de erro	Causa
InstanceMetadataServiceResourceFetcher:105 - [] Falha ao recuperar o token com.amazonaws. SdkClientException: Falha na conexão com o endpoint de serviço	Ele EMR SecretAgent não apareceu ou está tendo um problema. Inspecione os SecretAgent registros em busca de erros e o script de marionete para determinar se houve algum erro de provisionamento.

As consultas estão falhando inesperadamente

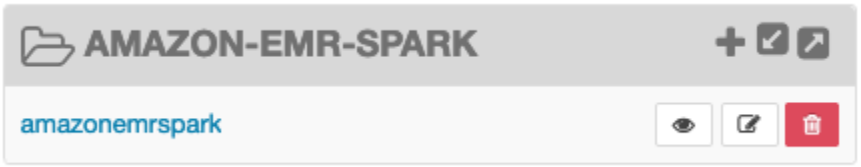
Verifique os registros do plug-in Apache Ranger (registros do Apache Hive,, EMR RecordServer EMR SecretAgent, etc.)

Esta seção é comum a todos os aplicativos que se integram ao plug-in Ranger, como Apache Hive, EMR Record Server e. EMR SecretAgent

Mensagens de erro comuns

Mensagem de erro	Causa
ERROR PolicyRefresher:272 - [] PolicyRef resher (serviceName=policy-repository): falha ao encontrar o serviço. Limpará o cache local de políticas (-1)	Essas mensagens de erro significam que o nome do serviço fornecido na configuração de EMR segurança não corresponde a um repositório de políticas de serviço no servidor Ranger Admin.

Se, no servidor Ranger Admin, seu SPARK serviço AMAZON EMR - - tiver a seguinte aparência, você deverá inserir **`amazonemrspark`** o nome do serviço.



Trabalhando com visualizações do AWS Glue Data Catalog (pré-visualização)

> **Note**
>
> AWS As visualizações do Glue Data Catalog na Amazon EMR estão em versão prévia e estão sujeitas a alterações. O recurso é fornecido como um serviço de Pré-visualização, conforme definido nos Termos de Serviço da AWS.

Você pode criar e gerenciar visualizações comuns únicas no AWS Glue Data Catalog. Visualizações comuns únicas são úteis porque oferecem suporte a vários mecanismos de SQL consulta, para que você possa acessar a mesma visualização em diferentes Serviços da AWS, como Amazon EMR Amazon Athena e Amazon Redshift.

Ao criar uma exibição no Catálogo de Dados, você pode usar concessões de recursos e controles de acesso baseados em tags AWS Lake Formation para conceder acesso a uma exibição do Catálogo de Dados. Usando esse método de controle de acesso, você não precisa configurar acesso adicional às tabelas referenciadas ao criar a exibição. Esse método de concessão de permissões é chamado de semântica definidora, e essas visualizações são chamadas de visualizações definidoras. Para obter mais informações sobre controle de acesso no Lake Formation, consulte Conceder e revogar permissões nos recursos do Catálogo de Dados. no Guia do AWS Lake Formation desenvolvedor.

As visualizações do Catálogo de Dados são úteis para os seguintes casos de uso:

- Controle de acesso granular — crie uma visualização que restrinja o acesso aos dados com base nas permissões de que o usuário precisa. Por exemplo, você pode usar visualizações no Catálogo de Dados para impedir que funcionários que não trabalham no departamento de RH vejam informações de identificação pessoal ()PII.
- Definição completa da visualização — ao aplicar determinados filtros à sua exibição no Catálogo de Dados, você garante que os registros de dados dentro de uma exibição no Catálogo de Dados estejam sempre completos.
- Segurança aprimorada — a definição da consulta usada para criar a exibição deve estar completa. Esse benefício significa que as visualizações no Catálogo de Dados são menos suscetíveis a SQL comandos de jogadores mal-intencionados.
- Compartilhamento simples de dados — compartilhe dados com outras Contas da AWS pessoas sem mover nenhum dado. Para obter mais informações, consulte Compartilhamento de dados entre contas no Lake Formation.

Criação de uma visualização do Catálogo de Dados

> **Important**
>
> Durante esta versão prévia, a Amazon EMR não valida o Spark SQL que você usa ao criar a visualização. Para reduzir os riscos, recomendamos que você limite os usuários aos quais você concede permissões de criação de visualizações.

Para criar uma exibição do Catálogo de Dados, você deve usar uma IAM função que tenha a `SELECT` permissão total com `Grantable` opções em todas as tabelas que você deseja referenciar ao criar a exibição. Essa função é chamada de função definidora. Para obter uma lista completa das permissões e pré-requisitos necessários para criar uma visualização do Catálogo de Dados, consulte Como trabalhar com exibições no Guia do AWS Lake Formation Desenvolvedor. Você deve usar o AWS CLI para configurar sua IAM função. Consulte Usar uma IAM função no AWS CLI para obter mais informações.

Siga estas etapas para criar uma exibição do Catálogo de Dados.

> **Note**
>
> Para acessar uma visualização do catálogo de dados do Apache Spark na AmazonEMR, você deve definir o dialeto para `SPARK` e para. `DialectVersion` `3.4.1-amzn-2`

1. Primeiro, baixe o modelo de pré-visualização.

```
aws s3 cp s3://emr-data-access-control-us-east-1/beta/glue-views/model/
service-2.json
```

2. Configure o AWS CLI para usar o modelo de pré-visualização.

```
aws configure add-model --service-model file:///<path-to-preview-model>/
service-2.json --service-name glue-views
```

3. Crie a visualização.

```
aws glue-views create-table --cli-input-json '{
  "DatabaseName": "<database>",
  "TableInput": {
    "Name": "<view>",
    "StorageDescriptor": {
      "Columns": [
        {
          "Name": "<col1>",
          "Type": "<data-type>"
        },
        ...
        {
          "Name": "<colN>",
```

```
          "Type": "<data-type>"
        }
      ]
    },
    "ViewDefinition": {
      "SubObjects": [
        "arn:aws:glue:<aws-region;>:<aws-account-id>:table/<database>/<referenced-
table1>",
        ...
        "arn:aws:glue:<aws-region>:<aws-account-id>:table/<database>/<referenced-
tableN>",
       ],
       "IsProtected": true,
      "Representations": [
        {
          "Dialect": "SPARK",
          "DialectVersion": "3.4.1-amzn-2",
          "ViewOriginalText": "<Spark-SQL>",
          "ViewExpandedText": "<Spark-SQL>"
        }
      ]
    }
  }
}'
```

Habilitando o acesso a uma visualização do Catálogo de Dados

> **Important**
>
> Recomendamos que você habilite o acesso às visualizações do Catálogo de Dados somente com EMR clusters em ambientes de teste e não em ambientes de produção.

Para acessar a visualização do catálogo de dados do Apache Spark na AmazonEMR, você deve primeiro habilitar o suporte para Lake Formation e usar o script abaixo para habilitar o suporte para visualizações com o Spark na Amazon. EMR Para obter mais informações sobre como habilitar o suporte, consulte Habilitar o Lake Formation com a Amazon EMR e Usar ações de bootstrap personalizadas.

```
# Download the script and upload it to Amazon S3
wget https://emr-data-access-control-us-east-1.s3.amazonaws.com/beta/glue-views/ba/
enable-mdv.sh /Users/$USER/enable-mdv.sh
aws s3 cp /Users/$USER/enable-views.sh s3://<bucket>/<prefix>/enable-views.sh

# EMR Security Configuration
cat <<EOT > /Users/$USER/lakeformation-protection.json
{
    "AuthorizationConfiguration":{
        "IAMConfiguration":{
            "EnableApplicationScopedIAMRole":true
        },
        "LakeFormationConfiguration":{
            "AuthorizedSessionTagValue":"Amazon EMR"
        }
    },
     "EncryptionConfiguration": {
            "EnableInTransitEncryption": true,
            "InTransitEncryptionConfiguration": {
            "TLSCertificateConfiguration": {
                 "CertificateProviderType": "PEM",
                 "S3Object": "s3://<BUCKET>/<PREFIX>/certificates.zip"
            }
        }
    }
}
EOT

SECURITY_CONFIG="RuntimeRolesWithAWSLakeFormation"

aws emr create-security-configuration \
--name $SECURITY_CONFIG \
--security-configuration file:///Users/$USER/lakeformation-protection.json

# EMR Cluster version
RELEASE_LABEL="emr-6.15.0"
```

Em seguida, use o AWS CLI comando a seguir que usa a ação bootstrap para criar um EMR cluster compatível com visualizações do Catálogo de Dados.

```
aws emr create-cluster \
...
--release-label $RELEASE_LABEL \
```

```
--security-configuration $SECURITY_CONFIG \
--bootstrap-actions \
 Name='Enable Views',Path="s3://<bucket>/<prefix>/enable-views.sh"
```

Consulta de uma visualização do Catálogo de Dados

> **Important**
>
> Durante esta versão prévia, recomendamos que você acesse exibições somente de fontes confiáveis. Na versão prévia, a Amazon EMR tem uma quantidade limitada de validações que protegem seu EMR cluster.

Depois de criar uma exibição do Catálogo de Dados, agora você pode usar uma IAM função para consultar a exibição. A IAM função deve ter a `SELECT` permissão na exibição do Catálogo de Dados. Você não precisa conceder acesso às tabelas subjacentes mencionadas na exibição. Você deve usar essa IAM função como uma função de tempo de execução. Você pode acessar a visualização de um EMR cluster usando uma função de tempo de execução do Amazon EMR Steps, EMR Studio e SageMaker Studio. Para obter mais informações sobre funções de tempo de execução, consulte Funções de tempo de execução para EMR etapas da Amazon.

Depois de configurar tudo, você pode consultar sua visualização. Por exemplo, depois de anexar o EMR cluster ao seu espaço de trabalho no EMR Studio, você pode executar a consulta a seguir para acessar uma visualização.

```
SELECT * from <database>.<glue-data-catalog-view> LIMIT 10
```

Limitações

Considere as seguintes limitações ao usar as exibições do Catálogo de Dados.

- Você só pode criar visualizações do catálogo de dados com o Amazon EMR 6.15.0.
- Você só pode referenciar até 10 tabelas na definição da exibição.
- Você só pode criar visualizações do Catálogo de `PROTECTED` Dados. `UNPROTECTED`visualizações não são suportadas.
- Você não pode referenciar tabelas em outras Conta da AWS nas visualizações do Catálogo de Dados.

- As funções definidas pelo usuário (UDFs) não são suportadas.
- Você não pode referenciar formatos de tabela aberta, como Apache Hudi ou Apache Iceberg, nas visualizações do Catálogo de Dados.
- Você não pode referenciar outras visualizações nas visualizações do Catálogo de Dados.

Controle do tráfego de rede com grupos de segurança

Os grupos de segurança atuam como firewalls virtuais para EC2 instâncias em seu cluster para controlar o tráfego de entrada e saída. Cada grupo de segurança tem um conjunto de regras que controla o tráfego de entrada para as instâncias e um conjunto de regras separado para controlar o tráfego de saída. Para obter mais informações, consulte Grupos EC2 de segurança da Amazon para instâncias Linux no Guia EC2 do usuário da Amazon.

Você usa duas classes de grupos de segurança com a AmazonEMR: grupos de segurança EMR gerenciados pela Amazon e grupos de segurança adicionais.

Cada cluster tem grupos de segurança gerenciados associados a ele. Você pode usar os grupos de segurança gerenciados padrão que a Amazon EMR cria ou especificar grupos de segurança gerenciados personalizados. De qualquer forma, a Amazon adiciona EMR automaticamente regras aos grupos de segurança gerenciados que um cluster precisa para se comunicar entre instâncias e AWS serviços do cluster.

Grupos de segurança adicionais são opcionais. Você pode especificá-los além dos grupos de segurança gerenciados para personalizar o acesso às instâncias do cluster. Os grupos de segurança adicionais contêm apenas regras definidas por você. A Amazon EMR não os modifica.

As regras que a Amazon EMR cria em grupos de segurança gerenciados permitem que o cluster se comunique entre os componentes internos. Para permitir que usuários e aplicativos acessem um cluster de fora do cluster, você pode editar regras em grupos de segurança gerenciados, criar grupos de segurança adicionais com regras adicionais ou ambos.

> **Important**
>
> A edição de regras em grupos de segurança gerenciados pode ter consequências inesperadas. Você pode bloquear acidentalmente o tráfego necessário para os clusters funcionarem corretamente e causar erros porque os nós estão inacessíveis. Tenha cuidado ao planejar e testar configurações de grupo de segurança antes da implementação.

Você pode especificar grupos de segurança somente ao criar um cluster. Eles não podem ser adicionados a um cluster ou instâncias do cluster enquanto um cluster está em execução, mas é possível editar, adicionar e remover regras de grupos de segurança existentes. As regras entram em vigor assim que você as salva.

Grupos de segurança são restritivos por padrão. A menos que seja adicionada uma regra que permita o tráfego, ele é rejeitado. Se houver mais de uma regra que se aplique ao mesmo tráfego e à mesma origem, a regra mais permissiva será aplicada. Por exemplo, se você tiver uma regra que permite a SSH partir do endereço IP 192.0.2.12/32 e outra regra que permite acesso a todo o tráfego do intervalo 192.0.2.0/24, a regra que permite todo o TCP tráfego do intervalo que inclui 192.0.2.12 tem precedência. TCP Nesse caso, o cliente em 192.0.2.12 pode ter mais acesso do que o desejado.

> **Important**
>
> Tome cuidado ao editar as regras de grupo de segurança para portas abertas. Adicione regras que só permitam tráfego de clientes confiáveis e autenticados para os protocolos e portas que sejam necessários para executar suas workloads.

Você pode configurar o EMR bloqueio de acesso público da Amazon em cada região que você usa para evitar a criação de clusters se uma regra permitir acesso público em qualquer porta que você não adicione a uma lista de exceções. Para AWS contas criadas após julho de 2019, o EMR bloqueio de acesso público da Amazon está ativado por padrão. Para AWS contas que criaram um cluster antes de julho de 2019, o EMR bloqueio de acesso público da Amazon está desativado por padrão. Para obter mais informações, consulte Usando a Amazon, EMR bloqueie o acesso público.

Tópicos

- Trabalhando com grupos de segurança EMR gerenciados pela Amazon
- Trabalhar com grupos de segurança adicionais
- Especificação de grupos de EMR segurança adicionais e gerenciados pela Amazon
- Especificando grupos EC2 de segurança para notebooks EMR
- Usando a Amazon, EMR bloqueie o acesso público

> **Note**
>
> A Amazon EMR pretende usar alternativas inclusivas para termos industriais potencialmente ofensivos ou não inclusivos, como “mestre” e “escravo”. Fizemos a transição para uma nova

terminologia para promover uma experiência mais inclusiva e facilitar a compreensão dos componentes do serviço.
Agora descrevemos “nós” como instâncias e descrevemos os tipos de EMR instância da Amazon como instâncias primárias, centrais e de tarefas. Durante a transição, você ainda pode encontrar referências antigas a termos desatualizados, como aqueles que dizem respeito aos grupos de segurança da AmazonEMR.

Trabalhando com grupos de segurança EMR gerenciados pela Amazon

> **Note**
>
> A Amazon EMR pretende usar alternativas inclusivas para termos industriais potencialmente ofensivos ou não inclusivos, como “mestre” e “escravo”. Fizemos a transição para uma nova terminologia para promover uma experiência mais inclusiva e facilitar a compreensão dos componentes do serviço.
> Agora descrevemos “nós” como instâncias e descrevemos os tipos de EMR instância da Amazon como instâncias primárias, centrais e de tarefas. Durante a transição, você ainda pode encontrar referências antigas a termos desatualizados, como aqueles que dizem respeito aos grupos de segurança da AmazonEMR.

Diferentes grupos de segurança gerenciados estão associados à instância primária e às instâncias centrais e de tarefa em um cluster. Um grupo de segurança gerenciado adicional para acesso de serviço é necessário quando você cria um cluster em uma sub-rede privada. Para obter mais informações sobre a função de grupos de segurança gerenciados com respeito à configuração de sua rede, consulte VPCOpções da Amazon.

Ao especificar grupos de segurança gerenciados para um cluster, você deve usar o mesmo tipo de grupo de segurança, padrão ou personalizado, para todos os grupos de segurança gerenciados. Por exemplo, você não pode especificar um grupo de segurança personalizado para a instância primária e, em seguida, não especificar um grupo de segurança personalizado para instâncias centrais e de tarefa.

Se você usar grupos de segurança gerenciados padrão, não será necessário especificá-los ao criar um cluster. A Amazon usa EMR automaticamente os padrões. Além disso, se os padrões VPC ainda não existirem no cluster, a Amazon os EMR cria. A Amazon EMR também os cria se você os especificar explicitamente e eles ainda não existirem.

É possível editar regras em grupos de segurança gerenciados depois que os clusters forem criados. Quando você cria um novo cluster, a Amazon EMR verifica as regras nos grupos de segurança gerenciados que você especifica e, em seguida, cria todas as regras de entrada ausentes que o novo cluster precisa, além das regras que podem ter sido adicionadas anteriormente. Salvo indicação específica em contrário, cada regra para grupos de segurança padrão EMR gerenciados pela Amazon também é adicionada aos grupos de segurança personalizados EMR gerenciados pela Amazon que você especificar.

Os grupos de segurança gerenciados padrão são os seguintes:

- ElasticMapReduce-primário

  Para regras nesse grupo de segurança, consulte Grupo EMR de segurança gerenciado pela Amazon para a instância primária (sub-redes públicas).
- ElasticMapReduce-núcleo

  Para regras nesse grupo de segurança, consulte Grupo EMR de segurança gerenciado pela Amazon para instâncias principais e de tarefas (sub-redes públicas).
- ElasticMapReduce- Primário - Privado

  Para regras nesse grupo de segurança, consulte Grupo EMR de segurança gerenciado pela Amazon para a instância primária (sub-redes privadas).
- ElasticMapReduce-Núcleo-Privado

  Para regras nesse grupo de segurança, consulte Grupo EMR de segurança gerenciado pela Amazon para instâncias principais e de tarefas (sub-redes privadas).
- ElasticMapReduce-ServiceAccess

  Para regras nesse grupo de segurança, consulte Grupo EMR de segurança gerenciado pela Amazon para acesso a serviços (sub-redes privadas).

Grupo EMR de segurança gerenciado pela Amazon para a instância primária (sub-redes públicas)

O grupo de segurança gerenciado padrão para a instância primária em sub-redes públicas tem o nome do grupo de ElasticMapReduce -primary. Tem as regras a seguir. Se você especificar um grupo de segurança gerenciado personalizado, a Amazon EMR adicionará as mesmas regras ao seu grupo de segurança personalizado.

Tipo	Protoc	Interva de portas	Origem	Detalhes
Regras de entrada				
Tudo ICMP - IPv4	Todos	N/D	O ID do grupo de segurança gerenciado da instância principal. Em outras palavras, o mesmo grupo de segurança em que a regra é exibida.	Essas regras reflexivas permitem o tráfego de entrada de qualquer instância associada ao grupo de segurança especificado. Usar o padrão `ElasticMapReduce-primary` para vários clusters permite que os nós principais e de tarefas desses clusters se comuniquem entre si por ICMP meio de qualquer UDP porta TCP ou porta. Especifiq ue grupos de segurança gerenciados personalizados para restringir o acesso entre clusters.
Tudo TCP	TCP	Todos		
Tudo UDP	UDP	Todos		
Tudo ICMP - IPV4	Todos	N/D	O ID de grupo de segurança gerenciado especificado para nós core e de tarefa.	Essas regras permitem todo o ICMP tráfego de entrada e o tráfego por qualquer UDP porta TCP ou por qualquer instância principal e de qualquer instância de tarefa associada ao grupo de segurança especificado, mesmo que as instâncias estejam em clusters diferentes.
Tudo TCP	TCP	Todos		
Tudo UDP	UDP	Todos		
Personali zar	TCP	8443	Vários intervalos de endereços IP da Amazon	Essas regras permitem que o gerenciador de clusters se comunique com o nó primário.

Para conceder SSH acesso a fontes confiáveis ao grupo de segurança principal com o console

Para editar seus grupos de segurança, você deve ter permissão para gerenciar grupos de segurança do cluster em VPC que o cluster está. Para obter mais informações, consulte Alteração

de permissões para um usuário e o exemplo de política que permite gerenciar grupos de EC2 segurança no Guia IAM do usuário.

1. Faça login no AWS Management Console e abra o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. Escolha Clusters. Escolha o ID do cluster que você deseja modificar.
3. No painel Rede e segurança, expanda a lista suspensa Grupos EC2 de segurança (firewall).
4. Em Nó primário, escolha seu grupo de segurança.
5. Escolha Editar regras de entrada.
6. Verifique se há uma regra de entrada que permita acesso público com as configurações a seguir. Se existir, escolha Excluir para removê-la.

   - Tipo

     SSH
   - Porta

     22
   - Origem

     Personalizado 0.0.0.0/0

   > **Warning**
   >
   > Antes de dezembro de 2020, havia uma regra pré-configurada para permitir tráfego de entrada na porta 22 de todas as fontes. Essa regra foi criada para simplificar SSH as conexões iniciais com o nó primário. Recomendamos fortemente remover esta regra de entrada e restringir o tráfego para origens confiáveis.

7. Role até o final da lista de regras e escolha Adicionar regra.
8. Em Tipo, selecione SSH.

   A seleção insere SSH automaticamente TCPpara Protocolo e 22 para Intervalo de Portas.
9. Para a origem, selecione Meu IP para adicionar automaticamente seu endereço IP como o endereço de origem. Você também pode adicionar um intervalo personalizado de endereços IP de clientes confiáveis ou criar regras adicionais para outros clientes. Diversos ambientes de rede

alocam endereços IP dinamicamente, portanto, pode ser necessário atualizar os endereços IP para clientes confiáveis no futuro.

10. Escolha Salvar.
11. Opcionalmente, escolha o outro grupo de segurança em nós principais e de tarefas no painel Rede e segurança e repita as etapas acima para permitir que o SSH cliente acesse os nós principais e de tarefas.

Grupo EMR de segurança gerenciado pela Amazon para instâncias principais e de tarefas (sub-redes públicas)

O grupo de segurança gerenciado padrão para instâncias principais e de tarefas em sub-redes públicas tem o nome do ElasticMapReduce grupo -core. O grupo de segurança gerenciado padrão tem as seguintes regras, e a Amazon EMR adiciona as mesmas regras se você especificar um grupo de segurança gerenciado personalizado.

Tipo	Protoc	Interva de portas	Origem	Detalhes
Regras de entrada				
Tudo ICMP - IPV4	Todos	N/D	O ID do grupo de segurança gerenciado para instâncias core e de tarefa. Em outras palavras, o mesmo grupo de segurança em que a regra é exibida.	Essas regras reflexivas permitem o tráfego de entrada de qualquer instância associada ao grupo de segurança especificado. Usar o padrão `ElasticMapReduce-core` para vários clusters permite que as instância s principais e de tarefas desses clusters se comuniquem entre si por ICMP ou por qualquer TCP UDP porta. Especifique grupos de segurança gerenciados personalizados para restringir o acesso entre clusters.
Tudo TCP	TCP	Todos		
Tudo UDP	UDP	Todos		
Tudo ICMP - IPV4	Todos	N/D	O ID do grupo de segurança gerenciado da instância principal.	Essas regras permitem todo o ICMP tráfego de entrada e o tráfego por qualquer UDP porta TCP ou por qualquer instância primária associada ao grupo de segurança especific

Tipo	Protoc	Interva de portas	Origem	Detalhes
Tudo TCP	TCP	Todos		ado, mesmo que as instâncias estejam em clusters diferentes.
Tudo UDP	UDP	Todos		

Grupo EMR de segurança gerenciado pela Amazon para a instância primária (sub-redes privadas)

O grupo de segurança gerenciado padrão para a instância primária em sub-redes privadas tem o nome do grupo de ElasticMapReduce -Primary-Private. O grupo de segurança gerenciado padrão tem as seguintes regras, e a Amazon EMR adiciona as mesmas regras se você especificar um grupo de segurança gerenciado personalizado.

Tipo	Protoc	Interva de portas	Origem	Detalhes
Regras de entrada				
Tudo ICMP - IPv4	Todos	N/D	O ID do grupo de segurança gerenciado da instância principal. Em outras palavras, o mesmo grupo de segurança em que a regra é exibida.	Essas regras reflexivas permitem o tráfego de entrada de todas as instâncias associadas com o grupo de segurança especificado e acessíveis a partir da sub-rede privada. Usar o padrão `ElasticMapReduce-Primary-Private` para vários clusters permite que os nós principais e de tarefas desses clusters se comuniquem entre si por ICMP meio de qualquer UDP porta TCP ou porta. Especifique grupos de segurança gerenciados personalizados para restringir o acesso entre clusters.
Tudo TCP	TCP	Todos		
Tudo UDP	UDP	Todos		

Tipo	Protoc	Interva de portas	Origem	Detalhes
Tudo ICMP - IPV4	Todos	N/D	O ID do grupo de segurança gerenciado para nós core e de tarefa.	Essas regras permitem que todo o ICMP tráfego de entrada e o tráfego por qualquer UDP porta TCP ou por qualquer instância central e de tarefa estejam associados ao grupo de segurança especificado e possam ser acessados de dentro da sub-rede privada, mesmo que as instâncias estejam em clusters diferentes.
Tudo TCP	TCP	Todos		
Tudo UDP	UDP	Todos		
HTTPS(843)	TCP	8443	O ID do grupo de segurança gerenciado para acesso de serviço em uma sub-rede privada.	Essa regra permite que o gerenciador de clusters se comunique com o nó primário.
Regras de saída				
Todo o tráfego	Tudo	Tudo	0.0.0.0/0	Fornece acesso de saída à Internet.
Personalizado TCP	TCP	9443	O ID do grupo de segurança gerenciado para acesso de serviço em uma sub-rede privada.	Se a regra de saída padrão “Todo o tráfego” acima for removida, essa regra é um requisito mínimo para o Amazon EMR 5.30.0 e versões posteriores. **Note** EMRA Amazon não adiciona essa regra quando você usa um grupo de segurança gerenciado personalizado.

Tipo	Protoc	Interva de portas	Origem	Detalhes
Personali zado TCP	TCP	80 (http) ou 443 (https)	O ID do grupo de segurança gerenciado para acesso de serviço em uma sub-rede privada.	Se a regra de saída padrão “Todo o tráfego” acima for removida, essa regra é um requisito mínimo para que o Amazon EMR 5.30.0 e versões posteriores se conectem ao Amazon S3 por https. **Note** EMRA Amazon não adiciona essa regra quando você usa um grupo de segurança gerenciado personalizado.

Grupo EMR de segurança gerenciado pela Amazon para instâncias principais e de tarefas (sub-redes privadas)

O grupo de segurança gerenciado padrão para instâncias principais e de tarefas em sub-redes privadas tem o nome do ElasticMapReduce grupo -Core-Private. O grupo de segurança gerenciado padrão tem as seguintes regras, e a Amazon EMR adiciona as mesmas regras se você especificar um grupo de segurança gerenciado personalizado.

Tipo	Protoc	Interva de portas	Origem	Detalhes
Regras de entrada				
Tudo ICMP - IPV4	Todos	N/D	O ID do grupo de segurança gerenciado para instâncias core e de tarefa. Em outras palavras, o	Essas regras reflexivas permitem o tráfego de entrada de qualquer instância associada ao grupo de segurança especificado. Usar o padrão `ElasticMapReduce-core` para vários clusters permite que as instância s principais e de tarefas desses clusters
Tudo TCP	TCP	Todos		

Tipo	Protoc	Interva de portas	Origem	Detalhes
Tudo UDP	UDP	Todos	mesmo grupo de segurança em que a regra é exibida.	se comuniquem entre si por ICMP ou por qualquer TCP UDP porta. Especifique grupos de segurança gerenciados personalizados para restringir o acesso entre clusters.
Tudo ICMP - IPV4	Todos	N/D	O ID do grupo de segurança gerenciado da instância principal.	Essas regras permitem todo o ICMP tráfego de entrada e o tráfego por qualquer UDP porta TCP ou por qualquer instância primária associada ao grupo de segurança especific ado, mesmo que as instâncias estejam em clusters diferentes.
Tudo TCP	TCP	Todos		
Tudo UDP	UDP	Todos		
HTTPS(843)	TCP	8443	O ID do grupo de segurança gerenciado para acesso de serviço em uma sub-rede privada.	Essa regra permite que o gerenciador de clusters se comunique com os nós core e de tarefa.
Regras de saída				
Todo o tráfego	Tudo	Tudo	0.0.0.0/0	Consulte Editar regras de saída abaixo.

Tipo	Protoc	Interva de portas	Origem	Detalhes
Personali zado TCP	TCP	80 (http) ou 443 (https)	O ID do grupo de segurança gerenciado para acesso de serviço em uma sub-rede privada.	Se a regra de saída padrão “Todo o tráfego” acima for removida, essa regra é um requisito mínimo para que o Amazon EMR 5.30.0 e versões posteriores se conectem ao Amazon S3 por https. **Note** EMRA Amazon não adiciona essa regra quando você usa um grupo de segurança gerenciado personalizado.

Editar regras de saída

Por padrão, a Amazon EMR cria esse grupo de segurança com regras de saída que permitem todo o tráfego de saída em todos os protocolos e portas. A opção de permitir todo o tráfego de saída é selecionada porque vários aplicativos da Amazon EMR e de clientes que podem ser executados em EMR clusters da Amazon podem exigir regras de saída diferentes. EMRA Amazon não consegue antecipar essas configurações específicas ao criar grupos de segurança padrão. Você pode reduzir o escopo da saída em seus grupos de segurança para incluir somente as regras adequadas a seus casos de uso e políticas de segurança. No mínimo, esse grupo de segurança exige as regras de saída a seguir, mas algumas aplicações podem precisar de saída adicional.

Tipo	Protoc	Interva de portas	Destination (Destino)	Detalhes
Tudo TCP	TCP	Todos	pl-*xxxxxxxx*	Lista gerenciada de prefixos do Amazon S3 `com.amazonaws.` *`MyRegion`*`.s3`.
Todo o tráfego	Tudo	Todos	sg-*xxxxxxxxx xxxxxxxx*	O ID do grupo de segurança `ElasticMa pReduce-Core-Private` .

Tipo	Protoc	Interva de portas	Destination (Destino)	Detalhes
Todo o tráfego	Tudo	Todos	sg-*xxxxxxxxx xxxxxxxx*	O ID do grupo de segurança `ElasticMa pReduce-Primary-Private` .
Personali zado TCP	TCP	9443	sg-*xxxxxxxxx xxxxxxxx*	O ID do grupo de segurança `ElasticMa pReduce-ServiceAccess` .

Grupo EMR de segurança gerenciado pela Amazon para acesso a serviços (sub-redes privadas)

O grupo de segurança gerenciado padrão para acesso ao serviço em sub-redes privadas tem o nome do grupo de ElasticMapReduce -. ServiceAccess Ele tem regras de entrada e regras de saída que permitem o tráfego HTTPS (porta 8443, porta 9443) para outros grupos de segurança gerenciados em sub-redes privadas. Essas regras permitem que o gerenciador de clusters se comunique com o nó primário e com os nós centrais e de tarefa. As mesmas regras serão necessárias se você estiver usando grupos de segurança personalizados.

Tipo	Protoc	Interva de portas	Origem	Detalhes
Regras de entrada necessárias para EMR clusters da Amazon com a EMR versão 5.30.0 e versões posteriores da Amazon.				
Personali zado TCP	TCP	9443	O ID do grupo de segurança gerenciado para a instância primária.	Essa regra permite a comunicação entre o grupo de segurança da instância principal e o grupo de segurança de acesso ao serviço.
Regras de saída obrigatórias para todos os clusters da Amazon EMR				
Personali zado TCP	TCP	8443	O ID do grupo de segurança	Essas regras permitem que o gerenciador de clusters se comunique com o nó primário e com os nós centrais e de tarefa.

Tipo	Protoc	Interva de portas	Origem	Detalhes
			gerenciado para a instância primária.	
Personali zado TCP	TCP	8443	O ID do grupo de segurança gerenciado para instâncias core e de tarefa.	Essas regras permitem que o gerenciador de clusters se comunique com o nó primário e com os nós centrais e de tarefa.

Trabalhar com grupos de segurança adicionais

Independentemente de você usar os grupos de segurança gerenciados padrão ou especificar grupos de segurança gerenciados personalizados, é possível usar grupos de segurança adicionais. Os grupos de segurança adicionais oferecem a você a flexibilidade para adaptar o acesso entre diferentes clusters e de clientes externos, recursos e aplicativos.

Considere os seguintes cenários como um exemplo. Você tem vários clusters que precisam se comunicar entre si, mas deseja permitir o SSH acesso de entrada à instância primária somente para um subconjunto específico de clusters. Para fazer isso, você pode usar o mesmo conjunto de grupos de segurança gerenciados para os clusters. Em seguida, você cria grupos de segurança adicionais que permitem SSH acesso de entrada de clientes confiáveis e especifica os grupos de segurança adicionais para a instância primária de cada cluster no subconjunto.

Você pode aplicar até 15 grupos de segurança adicionais para a instância primária, 15 para instâncias principais e de tarefas e 15 para acesso ao serviço (em sub-redes privadas). Se necessário, você pode especificar o mesmo grupo de segurança adicional para instâncias primárias, instâncias centrais e de tarefa e acesso de serviço. O número máximo de grupos de segurança e regras em sua conta está sujeito a limites da conta. Para obter mais informações, consulte Limites de grupos de segurança no Guia VPC do usuário da Amazon.

Especificação de grupos de EMR segurança adicionais e gerenciados pela Amazon

Você pode especificar grupos de segurança usando o AWS Management Console AWS CLI, o ou o Amazon EMRAPI. Se você não especificar grupos de segurança, a Amazon EMR cria grupos de segurança padrão. A especificação de grupos de segurança adicionais é opcional. Você pode atribuir grupos de segurança adicionais para instâncias primárias, instâncias centrais e de tarefa e acesso de serviço (somente sub-redes privadas).

Console

Para especificar grupos de segurança com o console

1. Faça login no AWS Management Console e abra o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Em Rede, selecione a seta ao lado EC2de grupos de segurança (firewall) para expandir essa seção. Em Nó primário e nós principais e de tarefas, os grupos de segurança EMR gerenciados padrão da Amazon são selecionados por padrão. Se você usa uma sub-rede privada, também tem a opção de selecionar um grupo de segurança em Acesso ao serviço.
4. Para alterar seu grupo de segurança EMR gerenciado pela Amazon, use o menu suspenso Escolher grupos de segurança para selecionar uma opção diferente na lista de opções do grupo EMR de segurança gerenciado pela Amazon. Você tem um grupo de segurança EMR gerenciado pela Amazon para o nó primário e os nós principais e de tarefas.
5. Para adicionar grupos de segurança personalizados, use o mesmo menu suspenso Escolher grupos de segurança para selecionar até quatro grupos de segurança personalizados na lista de opções Grupo de segurança personalizado. Você pode ter até quatro grupos de segurança personalizados para o nó primário e os nós centrais e de tarefa.
6. Escolha qualquer outra opção que se aplique ao cluster.
7. Para iniciar o cluster, escolha Criar cluster.

Especificar grupos de segurança com a AWS CLI

Para especificar grupos de segurança usando o, AWS CLI você usa o `create-cluster` comando com os seguintes parâmetros da `--ec2-attributes` opção:

Parâmetro	Descrição
`EmrManagedPrimarySecurityGroup`	Use esse parâmetro para especificar um grupo de segurança gerenciado personalizado para a instância primária. Se esse parâmetro for especificado, `EmrManagedCoreSecurityGroup` também deve ser especificado. Para clusters em sub-redes privadas, `ServiceAccessSecurityGroup` também deverá ser especificado.
`EmrManagedCoreSecurityGroup`	Use esse parâmetro para especificar um grupo de segurança gerenciado personalizado para instâncias core e de tarefa. Se esse parâmetro for especificado, `EmrManagedPrimarySecurityGroup` também deve ser especificado. Para clusters em sub-redes privadas, `ServiceAccessSecurityGroup` também deverá ser especificado.
`ServiceAccessSecurityGroup`	Use esse parâmetro para especificar um grupo de segurança gerenciado personalizado para acesso de serviço, o que se aplica apenas a clusters em sub-redes privadas. O grupo de segurança que você especificar não `ServiceAccessSecurityGroup` deve ser usado para nenhuma outra finalidade e também deve ser reservado para a AmazonEMR. Se esse parâmetro for especificado, `EmrManagedPrimarySecurityGroup` também deve ser especificado.
`AdditionalPrimarySecurityGroups`	

Parâmetro	Descrição
	Use esse parâmetro para especificar até quatro grupos de segurança adicionais para a instância primária.
`AdditionalCoreSecurityGroups`	Use esse parâmetro para especificar até quatro grupos de segurança adicionais para instâncias core e de tarefa.

Example — especificar grupos de segurança personalizados EMR gerenciados pela Amazon e grupos de segurança adicionais

O exemplo a seguir especifica grupos de segurança personalizados EMR gerenciados pela Amazon para um cluster em uma sub-rede privada, vários grupos de segurança adicionais para a instância primária e um único grupo de segurança adicional para instâncias principais e de tarefas.

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --name "ClusterCustomManagedAndAdditionalSGs" \
--release-label emr-emr-7.2.0 --applications Name=Hue Name=Hive \
Name=Pig --use-default-roles --ec2-attributes \
SubnetIds=subnet-xxxxxxxxxxxx,KeyName=myKey,\
ServiceAccessSecurityGroup=sg-xxxxxxxxxxxx,\
EmrManagedPrimarySecurityGroup=sg-xxxxxxxxxxxx,\
EmrManagedCoreSecurityGroup=sg-xxxxxxxxxxx,\
AdditionalPrimarySecurityGroups=['sg-xxxxxxxxxxx',\
'sg-xxxxxxxxxxx','sg-xxxxxxxxxx'],\
AdditionalCoreSecurityGroups=sg-xxxxxxxxxxx \
--instance-type m5.xlarge
```

Para obter mais informações, consulte create-cluster na AWS CLI Command Reference.

Especificando grupos EC2 de segurança para notebooks EMR

Quando você cria um EMR notebook, dois grupos de segurança são usados para controlar o tráfego de rede entre o EMR notebook e o EMR cluster da Amazon quando você usa o editor do notebook. Os grupos de segurança padrão têm regras mínimas que permitem somente o tráfego de rede entre o serviço EMR Notebooks e os clusters aos quais os notebooks estão conectados.

Um EMR notebook usa o Apache Livy para se comunicar com o cluster por meio de um proxy pela TCP porta 18888. Ao criar grupos de segurança personalizados com regras personalizadas para seu ambiente, você pode limitar o tráfego de rede para que apenas um subconjunto de cadernos possa executar código no editor de cadernos em determinados clusters. O cluster usa segurança personalizada, além dos grupos de segurança padrão do cluster. Para obter mais informações, consulte Controle o tráfego de rede com grupos de segurança no Amazon EMR Management Guide Especificando grupos EC2 de segurança para notebooks EMR e.

Grupo EC2 de segurança padrão para a instância primária

O grupo EC2 de segurança padrão da instância primária está associado à instância primária, além dos grupos de segurança do cluster para a instância primária.

Nome do grupo: ElasticMapReduceEditors-Livy

Regras

- Entrada

  Permita a TCP porta 18888 de qualquer recurso no grupo de EC2 segurança padrão para notebooks EMR
- Saída

  Nenhum

Grupo EC2 de segurança padrão para EMR notebooks

O grupo EC2 de segurança padrão do EMR notebook está associado ao editor do notebook de qualquer EMR notebook ao qual ele esteja atribuído.

Nome do grupo: ElasticMapReduceEditors-Editor

Regras

- Entrada

  Nenhum

- Saída

  Permita que a TCP porta 18888 acesse qualquer recurso no grupo de EC2 segurança padrão para EMR notebooks.

Grupo EC2 de segurança personalizado para EMR Notebooks ao associar Notebooks a repositórios Git

Para vincular um repositório Git ao seu notebook, o grupo de segurança do EMR notebook deve incluir uma regra de saída para que o notebook possa rotear o tráfego para a Internet. É recomendável criar um grupo de segurança para essa finalidade. A atualização do grupo de segurança padrão ElasticMapReduceEditors-Editor pode fornecer as mesmas regras de saída para outros notebooks anexados a esse grupo de segurança.

Regras

- Entrada

  Nenhum

- Saída

  Permita que o caderno encaminhe o tráfego para a Internet por meio do cluster, como demonstra o exemplo a seguir. Utiliza-se o valor 0.0.0.0/0 para fins de exemplo. É possível modificar essa regra para especificar os endereços IP dos repositórios baseados em Git.

Tipo	Protocolo	Intervalo de portas	Destination (Destino)
TCPRegra personalizada	TCP	18888	SG-
HTTPS	TCP	443	0.0.0.0/0

Usando a Amazon, EMR bloqueie o acesso público

O Amazon EMR block public access (BPA) impede que você lance um cluster em uma sub-rede pública se o cluster tiver uma configuração de segurança que permita tráfego de entrada de endereços IP públicos em uma porta.

> **Important**
>
> O bloqueio de acesso público é habilitado por padrão. Para aumentar a proteção da conta, é recomendável mantê-la habilitada.

Noções básicas do bloqueio ao acesso público

Você pode usar a configuração em nível de conta de bloqueio de acesso público para gerenciar centralmente o acesso à rede pública aos clusters da Amazon. EMR

Quando um usuário do seu Conta da AWS executa um cluster, a Amazon EMR verifica as regras de porta no grupo de segurança do cluster e as compara com suas regras de tráfego de entrada. Se o grupo de segurança tiver uma regra de entrada que abre portas para os endereços IP públicos IPv4 0.0.0.0/0 ouIPv6:: /0, e essas portas não forem especificadas como exceções para sua conta, a EMR Amazon não permitirá que o usuário crie o cluster.

Se um usuário modificar as regras do grupo de segurança de um cluster em execução em uma sub-rede pública para ter uma regra de acesso público que viole a BPA configuração da sua conta, a Amazon EMR revogará a nova regra se tiver permissão para fazer isso. Se a Amazon EMR não tiver permissão para revogar a regra, ela cria um evento no AWS Health painel que descreve a violação. Para conceder a permissão da regra de revogação à AmazonEMR, consulteConfigure EMR a Amazon para revogar as regras do grupo de segurança.

O bloqueio de acesso público é habilitado por padrão para todos os clusters em cada Região da AWS de sua Conta da AWS. BPAse aplica a todo o ciclo de vida de um cluster, mas não se aplica aos clusters que você cria em sub-redes privadas. Você pode configurar exceções à BPA regra; a porta 22 é uma exceção por padrão. Para obter mais informações sobre como configurar exceções, consulte Configurar o bloqueio de acesso público.

Configurar o bloqueio de acesso público

Você pode atualizar os grupos de segurança e a configuração de bloqueio de acesso público de suas contas a qualquer momento.

Você pode ativar e desativar as configurações de bloqueio de acesso público (BPA) com o AWS Management Console, o AWS Command Line Interface (AWS CLI) e o Amazon EMRAPI. As configurações se aplicam à sua conta com base na Região. Para manter a segurança do cluster, recomendamos que você useBPA.

Console

Para configurar, bloquear o acesso público com o console

1. Faça login no e AWS Management Console, em seguida, abra o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. Na barra de navegação superior, selecione a região que você deseja configurar, se ainda não estiver selecionada.
3. Em EMREC2Ativado no painel de navegação esquerdo, escolha Bloquear acesso público.
4. Em Block public access settings (Configurações de bloqueio de acesso público), conclua as etapas a seguir.

Para...	Fazer isso...
Ativar ou desativar o bloqueio de acesso público	Escolha Editar, escolha Ativar ou Desativar, conforme o caso, e escolha Salvar.
Editar portas na lista de exceções	1. Escolha Editar e encontre a seção Exceções do intervalo de portas. 2. Para adicionar portas à lista de exceções, escolha Add a port range (Adicionar um intervalo de portas) e insira uma nova porta ou um intervalo de portas. Repita para cada porta ou intervalo de portas a ser adicionado. 3. Para remover uma porta ou um intervalo de portas, escolha Remover ao lado da entrada na lista de intervalos de portas.

Para...	Fazer isso...
	4. Escolha Salvar.

AWS CLI

Para configurar, bloquear o acesso público usando o AWS CLI

- Use o comando `aws emr put-block-public-access-configuration` para configurar o bloqueio de acesso público, conforme mostrado nos exemplos a seguir.

Para...	Fazer isso...
Ativar o bloqueio de acesso público	Defina `BlockPublicSecurityGroupRules` como `true`, conforme mostrado no exemplo a seguir. Para que o cluster seja iniciado, nenhum grupo de segurança associado a um cluster pode ter uma regra de entrada que permita acesso público. `aws emr put-block-public-access-configuration --block-public-access-configuration BlockPublicSecurityGroupRules=true`

Para...	Fazer isso...
Desativar o bloqueio de acesso público	Defina `BlockPublicSecurityGroupRules` como `false`, conforme mostrado no exemplo a seguir. Os grupos de segurança associados a um cluster podem ter regras de entrada que permitam o acesso público em qualquer porta. Não recomendamos essa configuração. `aws emr put-block-public-access-configuration --block-public-access-configuration BlockPublicSecurityGroupRules=false`
Ativar o bloqueio de acesso público e especificar portas como exceções	O exemplo a seguir ativa o bloqueio de acesso público e especifica a porta 22 e as portas 100-101 como exceções. Isso permite que os clusters sejam criados se um grupo de segurança associado tiver uma regra de entrada que permita o acesso público na porta 22, na porta 100 ou na porta 101. `aws emr put-block-public-access-configuration --block-public-access-configuration '{ "BlockPublicSecurityGroupRules": true, "PermittedPublicSecurityGroupRuleRanges": [ { "MinRange": 22, "MaxRange": 22 }, { "MinRange": 100, "MaxRange": 101 } ] }'`

Configure EMR a Amazon para revogar as regras do grupo de segurança

A Amazon EMR precisa de permissão para revogar as regras do grupo de segurança e cumprir sua configuração de bloqueio de acesso público. Você pode usar uma das seguintes abordagens para dar à Amazon EMR a permissão de que ela precisa:

- (Recomendado) Anexe a política gerenciada `AmazonEMRServicePolicy_v2` ao perfil de serviço. Para obter mais informações, consulte Função de serviço para a Amazon EMR (EMRfunção).
- Crie uma nova política em linha que permita a ação `ec2:RevokeSecurityGroupIngress` em grupos de segurança. Para obter mais informações sobre como modificar uma política de permissões de função, consulte Modificar uma política de permissões de função com o IAMconsole e AWS CLIno Guia do IAM usuário. AWS API

Resolver violações ao bloqueio de acesso público

Se ocorrer uma violação ao bloqueio de acesso público, você poderá mitigá-la com uma destas ações:

- Se você quiser acessar uma interface web em seu cluster, use uma das opções descritas em Visualize interfaces web hospedadas em EMR clusters da Amazon Para acessar a interface por meio da SSH (porta 22).
- Para permitir o tráfego no cluster com base em endereços IP específicos em vez do endereço IP público, adicione uma regra de grupo de segurança. Para obter mais informações, consulte Adicionar regras a um grupo de segurança no Amazon EC2 Getting Started Guide.
- (Não recomendado) Você pode configurar EMR BPA as exceções da Amazon para incluir a porta ou o intervalo de portas desejado. Ao especificar uma BPA exceção, você introduz riscos com uma porta desprotegida. Se você pretende especificar uma exceção, remova a exceção assim que ela não for mais necessária. Para obter mais informações, consulte Configurar o bloqueio de acesso público.

Identificar clusters associados às regras do grupo de segurança

Talvez seja necessário identificar todos os clusters associados a determinada regra de grupo de segurança ou encontrar a regra de grupo de segurança de determinado cluster.

- Se você conhece o grupo de segurança, poderá identificar os clusters associados se encontrar as interfaces de rede do grupo de segurança. Para obter mais informações, consulte Como posso

encontrar os recursos associados a um grupo de EC2 segurança da Amazon? ligado AWS re:Post. As EC2 instâncias da Amazon que estão conectadas a essas interfaces de rede serão marcadas com o ID do cluster ao qual elas pertencem.

- Se você quiser encontrar os grupos de segurança de um cluster conhecido, siga as etapas descritas em Visualizar o status e os detalhes do cluster. Você pode encontrar os grupos de segurança do cluster no painel Rede e segurança no console ou no campo `Ec2InstanceAttributes` da AWS CLI.

Validação de conformidade para a Amazon EMR

Auditores terceirizados avaliam a segurança e a conformidade da Amazon EMR como parte de vários programas de AWS conformidade. Isso incluiSOC, PCI RAMPHIPAA, Fed e outros.

Para obter uma lista de AWS serviços no escopo de programas de conformidade específicos, consulte AWS serviços no escopo por programa de conformidade. Para obter informações gerais, consulte Programas de conformidade da AWS.

Você pode baixar relatórios de auditoria de terceiros usando AWS Artifact. Para obter mais informações, consulte Baixar relatórios em AWS Artifact.

Sua responsabilidade de conformidade ao usar a Amazon EMR é determinada pela confidencialidade de seus dados, pelos objetivos de conformidade da sua empresa e pelas leis e regulamentos aplicáveis. Se seu uso da Amazon EMR estiver sujeito à conformidade com padrões comoHIPAA,PCI, ou o FedRAMP, AWS fornece recursos para ajudar a:

- Guias de início rápido sobre segurança e conformidade — Esses guias de implantação discutem considerações arquitetônicas e fornecem etapas para a implantação de ambientes básicos focados em segurança e conformidade em. AWS
- Documento técnico sobre arquitetura para HIPAA segurança e conformidade — Este whitepaper descreve como as empresas podem usar AWS para criar aplicativos compatíveis. HIPAA
- AWS recursos de conformidade — essa coleção de pastas de trabalho e guias pode ser aplicada ao seu setor e local.
- AWS Config— Esse AWS serviço avalia se suas configurações de recursos estão em conformidade com as práticas internas, as diretrizes e os regulamentos do setor.
- AWS Security Hub— Esse AWS serviço fornece uma visão abrangente do seu estado de segurança interno, AWS que ajuda você a verificar sua conformidade com os padrões e as melhores práticas do setor de segurança.

Resiliência na Amazon EMR

A infraestrutura AWS global é construída em torno de AWS regiões e zonas de disponibilidade. AWS As regiões fornecem várias zonas de disponibilidade fisicamente separadas e isoladas, conectadas a redes de baixa latência, alta taxa de transferência e alta redundância. Com as Zonas de Disponibilidade, é possível projetar e operar aplicações e bancos de dados que executem o failover automaticamente entre as Zonas de Disponibilidade sem interrupção. As zonas de disponibilidade são mais altamente disponíveis, tolerantes a falhas e escaláveis que uma ou várias infraestruturas de datacenter tradicionais.

Para obter mais informações sobre AWS regiões e zonas de disponibilidade, consulte infraestrutura AWS global.

Além da infraestrutura AWS global, a Amazon EMR oferece vários recursos para ajudar a suportar suas necessidades de resiliência e backup de dados.

- Integração com o Amazon S3 por meio de EMRFS
- Suporte a vários nós principais

Segurança da infraestrutura na Amazon EMR

Como um serviço gerenciado, a Amazon EMR é protegida pela segurança de rede AWS global. Para obter informações sobre serviços AWS de segurança e como AWS proteger a infraestrutura, consulte AWS Cloud Security. Para projetar seu AWS ambiente usando as melhores práticas de segurança de infraestrutura, consulte Proteção de infraestrutura no Security Pillar AWS Well-Architected Framework.

Você usa API chamadas AWS publicadas para acessar a Amazon EMR pela rede. Os clientes devem oferecer suporte para:

- Segurança da camada de transporte (TLS). Exigimos TLS 1,2 e recomendamos TLS 1,3.
- Suítes de criptografia com sigilo direto perfeito (), como (Ephemeral PFS Diffie-Hellman) ou DHE (Elliptic Curve Ephemeral Diffie-Hellman). ECDHE A maioria dos sistemas modernos, como Java 7 e versões posteriores, comporta esses modos.

Além disso, as solicitações devem ser assinadas usando uma ID de chave de acesso e uma chave de acesso secreta associada a um IAM principal. Ou você pode usar o AWS Security Token Service (AWS STS) para gerar credenciais de segurança temporárias para assinar solicitações.

Tópicos

Conecte-se à Amazon EMR usando um VPC endpoint de interface

Você pode se conectar diretamente à Amazon EMR usando um VPCendpoint de interface (AWS PrivateLink) em sua Virtual Private Cloud (VPC) em vez de se conectar pela Internet. Quando você usa um VPC endpoint de interface, a comunicação entre você VPC e a Amazon EMR é conduzida inteiramente dentro da AWS rede. Cada VPC endpoint é representado por uma ou mais interfaces de rede elástica (ENIs) com endereços IP privados em suas VPC sub-redes.

O VPC endpoint da interface conecta você VPC diretamente à Amazon EMR sem um gateway de internet, NAT dispositivo, VPN conexão ou AWS Direct Connect conexão. As instâncias em seu VPC não precisam de endereços IP públicos para se comunicar com a Amazon EMRAPI.

Para usar a Amazon EMR por meio do seuVPC, você deve se conectar a partir de uma instância que esteja dentro da VPC ou conectar sua rede privada à sua VPC usando uma Amazon Virtual Private Network (VPN) ou AWS Direct Connect. Para obter informações sobre a AmazonVPN, consulte VPNas conexões no Guia do usuário da Amazon Virtual Private Cloud. Para obter informações sobre AWS Direct Connect, consulte Criação de uma conexão no Guia AWS Direct Connect do usuário.

Você pode criar um VPC endpoint de interface para se conectar à Amazon EMR usando o AWS console ou os comandos AWS Command Line Interface (AWS CLI). Para obter mais informações, consulte Creating an interface endpoint (Criação de um endpoint de interface).

Depois de criar um VPC endpoint de interface, se você habilitar DNS nomes de host privados para o endpoint, o endpoint padrão da Amazon será resolvido para seu EMR endpoint. VPC O endpoint de nome de serviço padrão da Amazon EMR está no seguinte formato.

```
elasticmapreduce.Region.amazonaws.com
```

Se você não habilitar DNS nomes de host privados, a Amazon VPC fornecerá um nome de DNS endpoint que você pode usar no seguinte formato.

```
VPC_Endpoint_ID.elasticmapreduce.Region.vpce.amazonaws.com
```

Para obter mais informações, consulte Interface VPC endpoints (AWS PrivateLink) no Guia do VPC usuário da Amazon.

A Amazon EMR oferece suporte para fazer chamadas para todas as suas APIações dentro do seuVPC.

Você pode anexar políticas de VPC endpoint a um VPC endpoint para controlar o acesso dos principais. IAM Você também pode associar grupos de segurança a um VPC endpoint para controlar o acesso de entrada e saída com base na origem e no destino do tráfego de rede, como uma variedade de endereços IP. Para obter mais informações, consulte Controle do acesso a serviços com VPC endpoints.

Crie uma política de VPC endpoint para a Amazon EMR

Você pode criar uma política para VPC endpoints da Amazon EMR para especificar o seguinte:

- O principal que pode ou não executar ações
- As ações que podem ser executadas
- Os recursos nos quais as ações podem ser executadas

Para obter mais informações, consulte Controle do acesso a serviços com VPC endpoints no Guia do VPC usuário da Amazon.

Example — política de VPC endpoint para negar todo o acesso de uma conta especificada AWS

A seguinte política de VPC endpoint nega AWS a conta *123456789012* todo o acesso aos recursos usando o endpoint.

```
{
    "Statement": [
        {
            "Action": "*",
            "Effect": "Allow",
            "Resource": "*",
            "Principal": "*"
        },
        {
```

```
            "Action": "*",
            "Effect": "Deny",
            "Resource": "*",
            "Principal": {
                "AWS": [
                    "123456789012"
                ]
            }
        }
    ]
}
```

Example — política de VPC endpoint para permitir VPC acesso somente a um IAM principal especificado (usuário)

A política de VPC endpoint a seguir permite acesso total somente ao usuário a. *lijuan* na AWS conta *123456789012*. Todos os outros IAM diretores têm acesso negado usando o endpoint.

```
{
    "Statement": [
        {
            "Action": "*",
            "Effect": "Allow",
            "Resource": "*",
            "Principal": {
                "AWS": [
                    "arn:aws:iam::123456789012:user/lijuan"
                ]
            }
        }]
}
```

Example — política VPC de endpoint para permitir operações somente de leitura EMR

A política de VPC endpoint a seguir permite somente AWS uma conta *123456789012* para realizar as EMR ações especificadas da Amazon.

As ações especificadas fornecem o equivalente ao acesso somente de leitura para a Amazon. EMR Todas as outras ações no VPC são negadas para a conta especificada. Todas as outras contas terão acesso negado. Para obter uma lista de EMR ações da Amazon, consulte Ações, recursos e chaves de condição para a Amazon EMR.

```
{
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:DescribeSecurityConfiguration",
                "elasticmapreduce:GetBlockPublicAccessConfiguration",
                "elasticmapreduce:ListBootstrapActions",
                "elasticmapreduce:ViewEventsFromAllClustersInConsole",
                "elasticmapreduce:ListSteps",
                "elasticmapreduce:ListInstanceFleets",
                "elasticmapreduce:DescribeCluster",
                "elasticmapreduce:ListInstanceGroups",
                "elasticmapreduce:DescribeStep",
                "elasticmapreduce:ListInstances",
                "elasticmapreduce:ListSecurityConfigurations",
                "elasticmapreduce:DescribeEditor",
                "elasticmapreduce:ListClusters",
                "elasticmapreduce:ListEditors"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Principal": {
                "AWS": [
                    "123456789012"
                ]
            }
        }
    ]
}
```

Example — política VPC de endpoint negando acesso a um cluster especificado

A política de VPC endpoint a seguir permite acesso total a todas as contas e diretores, mas nega qualquer acesso à conta AWS *123456789012* às ações realizadas no EMR cluster da Amazon com ID do cluster *j-A1B2CD34EF5G*. Outras EMR ações da Amazon que não oferecem suporte a permissões em nível de recursos para clusters ainda são permitidas. Para obter uma lista das EMR ações da Amazon e seus tipos de recursos correspondentes, consulte Ações, recursos e chaves de condição para a Amazon EMR.

```
{
    "Statement": [
        {
```

```
            "Action": "*",
            "Effect": "Allow",
            "Resource": "*",
            "Principal": "*"
        },
        {
            "Action": "*",
            "Effect": "Deny",
            "Resource": "arn:aws:elasticmapreduce:us-west-2:123456789012:cluster/j-
A1B2CD34EF5G",
            "Principal": {
                "AWS": [
                    "123456789012"
                ]
            }
        }
    ]
}
```

Gerenciar clusters

Após iniciar seu cluster, você pode monitorá-lo e gerenciá-lo. EMRA Amazon fornece várias ferramentas que você pode usar para se conectar e controlar seu cluster.

Tópicos

Conectar-se a um cluster

Quando você executa um EMR cluster da Amazon, geralmente tudo o que você precisa fazer é executar um aplicativo para analisar seus dados e, em seguida, coletar a saída de um bucket do Amazon S3. Às vezes, você pode querer interagir com o nó primário enquanto o cluster está em execução. Por exemplo, talvez você queira se conectar ao nó primário para executar consultas interativas, verificar arquivos de log, depurar um problema com o cluster, monitorar a performance usando uma aplicação como o Ganglia, que é executada no nó primário e assim por diante. As seções a seguir descrevem técnicas que você pode usar para conectar-se ao nó primário.

Em um EMR cluster, o nó primário é uma EC2 instância da Amazon que coordena as EC2 instâncias que estão sendo executadas como tarefas e nós principais. O nó primário expõe um DNS nome público que você pode usar para se conectar a ele. Por padrão, a Amazon EMR cria regras de grupo de segurança para o nó primário e para os nós principais e de tarefas, que determinam como você acessa os nós.

> **Note**
>
> Você pode conectar-se ao nó primário somente enquanto o cluster está em execução. Quando o cluster é encerrado, a EC2 instância que atua como nó primário é encerrada e não está mais disponível. Para se conectar ao nó primário, você também deve se autenticar

para o cluster. Você pode usar o Kerberos para autenticação ou especificar uma EC2 chave privada de par de chaves da Amazon ao iniciar o cluster. Para obter mais informações sobre como configurar o Kerberos e se conectar, consulte Use o Kerberos para autenticação com a Amazon EMR. Quando você executa um cluster a partir do console, a EC2 chave privada do par de chaves da Amazon é especificada na seção Segurança e acesso na página Criar cluster.

Por padrão, o grupo de segurança ElasticMapReduce -master não permite acesso de entradaSSH. Talvez seja necessário adicionar uma regra de entrada que permita o SSH acesso (TCPporta 22) das fontes que você deseja acessar. Para obter mais informações sobre a modificação das regras do grupo de segurança, consulte Adicionar regras a um grupo de segurança no Guia do EC2 usuário da Amazon.

Important

Não modifique as regras restantes no grupo de segurança ElasticMapReduce -master. Modificar essas regras pode interferir com o funcionamento do cluster.

Tópicos

Antes de se conectar: autorize o tráfego de entrada

Antes de se conectar a um EMR cluster da Amazon, você deve autorizar o SSH tráfego de entrada (porta 22) de clientes confiáveis, como o endereço IP do seu computador. Para isso, edite as regras do grupo de segurança gerenciado para os nós aos quais deseja se conectar. Por exemplo, as instruções a seguir mostram como adicionar uma regra de entrada para SSH acessar o grupo de segurança ElasticMapReduce -master padrão.

Para obter mais informações sobre o uso de grupos de segurança com a AmazonEMR, consulteControle do tráfego de rede com grupos de segurança.

Console

Para conceder SSH acesso a fontes confiáveis ao grupo de segurança principal com o console

Para editar seus grupos de segurança, você deve ter permissão para gerenciar grupos de segurança do cluster em VPC que o cluster está. Para obter mais informações, consulte Alteração de permissões para um usuário e o exemplo de política que permite gerenciar grupos de EC2 segurança no Guia IAM do usuário.

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha o cluster que você deseja atualizar. Isso abre a página de detalhes do cluster. A guia Propriedades da página será pré-selecionada.
3. Em Rede, na guia Propriedades, selecione a seta ao lado EC2de grupos de segurança (firewall) para expandir essa seção. Em Nó primário, selecione o link do grupo de segurança. Isso abre o console do EC2.
4. Escolha a guia Regras de entrada e escolha Editar regras.
5. Verifique se há uma regra de entrada que permita acesso público com as configurações a seguir. Se existir, escolha Excluir para removê-la.

   - Tipo

     SSH
   - Porta

     22
   - Origem

     Personalizado 0.0.0.0/0

> **Warning**
>
> Antes de dezembro de 2020, o grupo de segurança ElasticMapReduce -master tinha uma regra pré-configurada para permitir tráfego de entrada na Porta 22 de todas as fontes. Essa regra foi criada para simplificar SSH as conexões iniciais com o nó

primário. Recomendamos fortemente remover esta regra de entrada e restringir o tráfego para origens confiáveis.

6. Role até o final da lista de regras e escolha Adicionar regra.
7. Em Tipo, selecione SSH. Essa seleção entra automaticamente em TCPProtocolo e 22 em Intervalo de Portas.
8. Para a origem, selecione Meu IP para adicionar automaticamente seu endereço IP como o endereço de origem. Você também pode adicionar um intervalo personalizado de endereços IP de clientes confiáveis ou criar regras adicionais para outros clientes. Diversos ambientes de rede alocam endereços IP dinamicamente, portanto, pode ser necessário atualizar os endereços IP para clientes confiáveis no futuro.
9. Escolha Salvar.
10. Opcionalmente, retorne à Etapa 3, escolha os Nós centrais e de tarefa e repita as Etapas 4 a 8. Isso concede acesso ao SSH cliente aos nós principais e de tarefas.

Conecte-se ao nó primário usando SSH

O Secure Shell (SSH) é um protocolo de rede que você pode usar para criar uma conexão segura com um computador remoto. Depois de estabelecer uma conexão, o terminal no computador local se comporta como se estivesse em execução no computador remoto. Os comandos que você emitir localmente serão executados no computador remoto, e a saída do comando do computador remoto será exibida na janela do terminal.

Ao usar SSH com AWS, você está se conectando a uma EC2 instância, que é um servidor virtual executado na nuvem. Ao trabalhar com a AmazonEMR, o uso mais comum SSH é conectar-se à EC2 instância que está atuando como o nó principal do cluster.

O uso SSH para se conectar ao nó primário permite monitorar e interagir com o cluster. Você pode emitir comandos do Linux no nó primário, executar aplicações como o Hive e o Pig interativamente, pesquisar diretórios, ler arquivos de log e assim por diante. Você também pode criar um túnel na sua SSH conexão para visualizar as interfaces web hospedadas no nó primário. Para obter mais informações, consulte Visualize interfaces web hospedadas em EMR clusters da Amazon.

Para se conectar ao nó primário usandoSSH, você precisa do DNS nome público do nó primário. Além disso, o grupo de segurança associado ao nó primário deve ter uma regra de entrada que permita o tráfego SSH (TCPporta 22) de uma origem que inclua o cliente de origem da SSH

conexão. Talvez seja necessário adicionar uma regra para permitir uma SSH conexão do seu cliente. Para obter mais informações sobre a modificação das regras do grupo de segurança, consulte Controle do tráfego de rede com grupos de segurança Adicionar regras a um grupo de segurança no Guia do EC2 usuário da Amazon.

Recupere o DNS nome público do nó primário

Você pode recuperar o DNS nome público principal usando o EMR console da Amazon e o. AWS CLI

Console

Para recuperar o DNS nome público do nó primário com o novo console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, selecione o cluster em que você deseja recuperar o nome públicoDNS.
3. Observe o DNS valor público do nó primário na seção Resumo da página de detalhes do cluster.

CLI

Para recuperar o DNS nome público do nó primário com o AWS CLI

1. Para recuperar o identificador do cluster, digite o seguinte comando.

```
aws emr list-clusters
```

A saída lista seus clusters, incluindo o clusterIDs. Observe o ID do cluster ao qual você está se conectando.

```
"Status": {
    "Timeline": {
        "ReadyDateTime": 1408040782.374,
        "CreationDateTime": 1408040501.213
    },
    "State": "WAITING",
    "StateChangeReason": {
        "Message": "Waiting after step completed"
```

```
    }
},
"NormalizedInstanceHours": 4,
"Id": "j-2AL4XXXXXX5T9",
"Name": "My cluster"
```

2. Para listar as instâncias do cluster, incluindo o DNS nome público do cluster, digite um dos comandos a seguir. Substituir *j-2AL4XXXXXX5T9* com o ID do cluster retornado pelo comando anterior.

```
aws emr list-instances --cluster-id j-2AL4XXXXXX5T9
```

Ou:

```
aws emr describe-cluster --cluster-id j-2AL4XXXXXX5T9
```

A saída lista as instâncias do cluster, incluindo DNS nomes e endereços IP. Observe o valor para `PublicDnsName`.

```
"Status": {
    "Timeline": {
        "ReadyDateTime": 1408040779.263,
        "CreationDateTime": 1408040515.535
    },
    "State": "RUNNING",
    "StateChangeReason": {}
},
"Ec2InstanceId": "i-e89b45e7",
"PublicDnsName": "ec2-###-##-##-###.us-west-2.compute.amazonaws.com"

"PrivateDnsName": "ip-###-##-##-###.us-west-2.compute.internal",
"PublicIpAddress": "##.###.###.##",
"Id": "ci-12XXXXXXXXFMH",
"PrivateIpAddress": "###.##.#.###"
```

Para obter mais informações, consulte EMRos comandos da Amazon no AWS CLI.

Conecte-se ao nó primário usando SSH uma chave EC2 privada da Amazon no Linux, Unix e Mac OS X

Para criar uma SSH conexão autenticada com um arquivo de chave privada, você precisa especificar a EC2 chave privada do par de chaves da Amazon ao iniciar um cluster. Para obter mais informações sobre como acessar seu par de chaves, consulte os pares de EC2 chaves da Amazon no Guia EC2 do usuário da Amazon.

Seu computador Linux provavelmente inclui um SSH cliente por padrão. Por exemplo, o Open SSH está instalado na maioria dos sistemas operacionais Linux, Unix e macOS. Você pode verificar se há um SSH cliente digitando ssh na linha de comando. Se o seu computador não reconhecer o comando, instale um SSH cliente para se conectar ao nó primário. O SSH projeto Open fornece uma implementação gratuita do conjunto completo de SSH ferramentas. Para obter mais informações, consulte o SSH site da Open.

As instruções a seguir demonstram como abrir uma SSH conexão com o nó EMR primário da Amazon no Linux, Unix e Mac OS X.

Para configurar as permissões do arquivo de chave privada do par de chaves

Antes de usar sua EC2 chave privada de par de chaves Amazon para criar uma SSH conexão, você deve definir permissões no `.pem` arquivo para que somente o proprietário da chave tenha permissão para acessá-lo. Isso é necessário para criar uma SSH conexão usando o terminal ou AWS CLI o.

1. Verifique se você permitiu o SSH tráfego de entrada. Para obter instruções, consulte Antes de se conectar: autorize o tráfego de entrada.
2. Localize seu arquivo `.pem`. Estas instruções pressupõem que o arquivo se chame `mykeypair.pem` e esteja armazenado no diretório inicial do usuário atual.
3. Digite o seguinte comando para definir as permissões. Substituir *~/mykeypair.pem* com o caminho completo e o nome do arquivo de chave privada do par de chaves. Por exemplo, `C:/Users/<username>/.ssh/mykeypair.pem`.

   ```
   chmod 400 ~/mykeypair.pem
   ```

   Se você não definir permissões no arquivo `.pem`, receberá um erro indicando que o arquivo de chave está desprotegido e que a chave será rejeitada. Para conectar, você só precisa definir permissões no arquivo de chave privada do par de chaves ao usá-lo pela primeira vez.

Conectar-se ao nó primário usando o terminal

1. Abra uma janela do terminal. No Mac OS X, escolha Applications > Utilities > Terminal (Aplicativos > Utilitários > Terminal). Em outras distribuições do Linux, o terminal está normalmente localizado em Applications > Accessories > Terminal (Aplicativos > Acessórios > Terminal).
2. Para estabelecer uma conexão com o nó primário, digite o comando a seguir. Substituir *ec2-###-##-##-###.compute-1.amazonaws.com* com o DNS nome público primário do seu cluster e substitua *~/mykeypair.pem* com o caminho completo e o nome do `.pem` arquivo. Por exemplo, `C:/Users/<username>/.ssh/mykeypair.pem`.

   ```
   ssh hadoop@ec2-###-##-##-###.compute-1.amazonaws.com -i ~/mykeypair.pem
   ```

   > **Important**
   >
   > Você deve usar o nome de login `hadoop` ao se conectar ao nó EMR primário da Amazon; caso contrário, você poderá ver um erro semelhante `Server refused our key` a.

3. Um aviso afirma que não foi possível verificar a autenticidade do host ao qual você está se conectando. Digite `yes` para continuar.
4. Quando terminar de trabalhar no nó primário, digite o comando a seguir para fechar a SSH conexão.

   ```
   exit
   ```

Se você estiver tendo dificuldades para se conectar SSH ao seu nó primário, consulte Solucionar problemas de conexão com sua instância.

Conecte-se ao nó primário usando SSH no Windows

Os usuários do Windows podem usar um SSH cliente como o Pu TTY para se conectar ao nó primário. Antes de se conectar ao nó EMR primário da Amazon, você deve baixar e instalar o Pu TTY e o uTTYgen P. Você pode baixar essas ferramentas na página de TTY download do Pu.

TTYO Pu não oferece suporte nativo ao formato de arquivo de chave privada do par de chaves (`.pem`) gerado pela AmazonEC2. Você usa P uTTYgen para converter seu arquivo de chave para o

TTY formato Pu necessário (`.ppk`). Você deve converter sua chave nesse formato (`.ppk`) antes de tentar se conectar ao nó primário usando PuTTY.

Para obter mais informações sobre como converter sua chave, consulte Convertendo sua chave privada usando P uTTYgen no Guia EC2 do usuário da Amazon.

Para se conectar ao nó primário usando Pu TTY

1. Verifique se você permitiu o SSH tráfego de entrada. Para obter instruções, consulte Antes de se conectar: autorize o tráfego de entrada.
2. Abra o `putty.exe`. Você também pode iniciar o Pu na lista TTY de programas do Windows.
3. Se necessário, na lista Category (Categoria), escolha Session (Sessão).
4. Em Nome do host (ou endereço IP), digite `hadoop@`*`MasterPublicDNS`*. Por exemplo: `hadoop@`*`ec2-###-##-##-###.compute-1.amazonaws.com`*.
5. Na lista Categoria, escolha Conexão > SSH, Auth.
6. Para Private key file for authentication (Arquivo de chave privada para autenticação), escolha Browse (Procurar) e selecione o arquivo `.ppk` que você gerou.
7. Escolha Abrir e depois Sim para ignorar o alerta de TTY segurança Pu.

> **Important**
>
> Quando fizer login no nó primário, digite `hadoop` se for solicitado a especificar um nome de usuário.

8. Quando terminar de trabalhar no nó primário, você pode fechar a SSH conexão fechando o PuTTY.

> **Note**
>
> Para evitar que a SSH conexão atinja o tempo limite, você pode escolher Conexão na lista Categoria e selecionar a opção Ativar TCP _keepalives. Se você tiver uma SSH sessão ativa no PuTTY, poderá alterar suas configurações abrindo o contexto (clique com o botão direito do mouse) da barra de TTY título do Pu e escolhendo Alterar configurações.

Se você estiver tendo dificuldades para se conectar SSH ao seu nó primário, consulte Solucionar problemas de conexão com sua instância.

Conectar-se ao nó primário usando a AWS CLI

Você pode criar uma SSH conexão com o nó primário usando o AWS CLI no Windows e no Linux, Unix e Mac OS X. Independentemente da plataforma, você precisa do DNS nome público do nó primário e da EC2 chave privada do par de chaves da Amazon. Se você estiver usando o AWS CLI no Linux, Unix ou Mac OS X, também deverá definir permissões no arquivo (`.pemou.ppk`) chave privada, conforme mostrado emPara configurar as permissões do arquivo de chave privada do par de chaves.

Para se conectar ao nó primário usando o AWS CLI

1. Verifique se você permitiu o SSH tráfego de entrada. Para obter instruções, consulte Antes de se conectar: autorize o tráfego de entrada.
2. Para recuperar o identificador de cluster, digite:

   ```
   aws emr list-clusters
   ```

   A saída lista seus clusters, incluindo o clusterIDs. Observe o ID do cluster ao qual você está se conectando.

   ```
   "Status": {
       "Timeline": {
           "ReadyDateTime": 1408040782.374,
           "CreationDateTime": 1408040501.213
       },
       "State": "WAITING",
       "StateChangeReason": {
           "Message": "Waiting after step completed"
       }
   },
   "NormalizedInstanceHours": 4,
   "Id": "j-2AL4XXXXXX5T9",
   "Name": "AWS CLI cluster"
   ```

3. Digite o comando a seguir para abrir uma SSH conexão com o nó primário. No exemplo a seguir, substitua *j-2AL4XXXXXX5T9* com o ID do cluster e substitua *~/mykeypair.key* com

o caminho completo e o nome do `.pem` arquivo (para Linux, Unix e Mac OS X) ou `.ppk` arquivo (para Windows). Por exemplo, `C:\Users\<username>\.ssh\mykeypair.pem`.

```
aws emr ssh --cluster-id j-2AL4XXXXXX5T9 --key-pair-file ~/mykeypair.key
```

4. Quando terminar de trabalhar no nó primário, feche a AWS CLI janela.

Para obter mais informações, consulte EMRos comandos da Amazon no AWS CLI. Se você estiver tendo dificuldades para se conectar SSH ao seu nó primário, consulte Solucionar problemas de conexão com sua instância.

Portas EMR de serviço da Amazon

> **Note**
>
> A seguir estão as interfaces e portas de serviço para componentes na AmazonEMR. Esta não é uma lista completa de portas de serviço. Serviços não padrão, como SSL portas e diferentes tipos de protocolos, não estão listados.

> **Important**
>
> Tome cuidado ao editar as regras de grupo de segurança para portas abertas. Adicione regras que só permitam tráfego de clientes confiáveis e autenticados para os protocolos e portas que sejam necessários para executar suas workloads.

Componente	Service description (Descrição do serviço)	Serviço em execução por padrão	Port (Porta)	Chave de configuração
Hadoop	HTTP KMS REST API	Sim	9600	hadoop.kms.http.port
HDFS	IU da Web do Namenode	Sim	9870	dfs.namenode.http-address

Componente	Service description (Descrição do serviço)	Serviço em execução por padrão	Port (Porta)	Chave de configuração
	Nó de nome RPC	Sim	8020	dfs.namenode.rpc-address
	DataNode UI da Web	Sim	9864	dfs.datanode.http.address
	Datanode HTTP para transferência de dados	Sim	986	dfs.datanode.address
	Datanode RPC para transferência de dados	Sim	9867	dfs.datanode.ipc.address
Hive	HiveServer2 Parcimônia	Sim	10000	hive.server2.thrift.port
	HiveServer2 HTTP	Não	10001	hive.server2.thrift.http.port
	HiveServer2 Interface de usuário da Web	Sim	10002	hive.server2.webui.port
	Hive Metastore	Sim	9083	hive.metastore.port / metastore.thrift.port
	W ebHCat	Não	50111	templeton.port
	LLAPserviço de gerenciamento de daemon () RPC	Não	15004	hive.llap.management.rpc.port

Componente	Service description (Descrição do serviço)	Serviço em execução por padrão	Port (Porta)	Chave de configuração
	YARNporta shuffle para LLAP - daemon-hosted shuffle	Não	1551	hive.llap.daemon.yarn.shuffle.port
	O LLAP daemon RPC	Não	Dinâmico	hive.llap.daemon.rpc.port
	LLAPdaemon Web UI	Não	15002	hive.llap.daemon.web.port
	LLAPserviço de saída daemon	Não	15003	hive.llap.daemon.output.service.port
Oozie		Sim	11000	
Tez	IU Tez	Sim	8080	
YARN	Shuffle	Sim	13562	mapreduce.shuffle.port
	Localizador RPC	Sim	8040	yarn.nodemanager.localizer.address
		Sim	8041	
	Endereço do NM Webapp	Sim	8042	yarn.nodemanager.webapp.address

Componente	Service description (Descrição do serviço)	Serviço em execução por padrão	Port (Porta)	Chave de configuração
	Aplicação Web RM	Sim	8088	yarn.resourcemanager.webapp.address
		Sim	8025	
	Scheduler	Sim	8030	yarn.resourcemanager.scheduler.address
	interface do gerenciador de aplicações	Sim	8032	yarn.resourcemanager.address
	Interface do administrador do RM	Sim	8033	yarn.resourcemanager.admin.address
	JobHistory UI da Web do servidor	Sim	1988	mapreduce.jobhistory.webapp.address
	JobHistory UI da Web para administrador do servidor	Sim	10033	mapreduce.jobhistory.admin.address
	JobHistory Servidor (RPC)	Sim	10020	mapreduce.jobhistory.address

Componente	Service description (Descrição do serviço)	Serviço em execução por padrão	Port (Porta)	Chave de configuração
	Servidor de cronograma do aplicativo () RPC	Sim	10200	yarn.timeline-service.address
	UI da HTTP Web do servidor Timeline do aplicativo	Sim	8188	yarn.timeline-service.webapp.address
	UI da HTTPS Web do servidor Timeline do aplicativo	Não	8190	yarn.timeline-service.webapp.https.address
		Sim	2088	
Zookeeper	Porta de cliente	Sim	2181	
		Sim	37301	
		Sim	8341	

Visualize interfaces web hospedadas em EMR clusters da Amazon

> **Important**
>
> É possível configurar um grupo de segurança personalizado para permitir acesso de entrada a essas interfaces da Web. Lembre-se de que qualquer porta na qual você permita o tráfego de entrada representa uma possível vulnerabilidade de segurança. Revise atentamente os grupos de segurança personalizados para minimizar vulnerabilidades. Para obter mais informações, consulte Controle do tráfego de rede com grupos de segurança.

O Hadoop e outros aplicativos que você instala no seu EMR cluster publicam interfaces de usuário como sites hospedados no nó primário. Por motivos de segurança, ao usar o Amazon EMR Managed Security Groups, esses sites só estão disponíveis no servidor web local do nó primário. Por isso, é necessário se conectar ao nó primário para visualizar as interfaces Web. Para obter mais informações, consulte Conecte-se ao nó primário usando SSH. O Hadoop também publica interfaces de usuário como sites hospedados nos nós core e escravos. Esses sites também só estão disponíveis em servidores Web locais nos nós.

A tabela a seguir lista as interfaces da web que você pode visualizar nas instâncias do cluster. Essas interfaces Hadoop estão disponíveis em todos os clusters. Para as interfaces da instância mestre, substitua *master-public-dns-name* com o público principal DNS listado na guia Resumo do cluster no EMR console da Amazon. Para interfaces principais e de instância de tarefas, substitua *coretask-public-dns-name* com o DNSnome público listado para a instância. Para encontrar o DNSnome público de uma instância, no EMR console da Amazon, escolha seu cluster na lista, escolha a guia Hardware, escolha o ID do grupo de instâncias que contém a instância à qual você deseja se conectar e, em seguida, anote o DNSnome público listado para a instância.

Nome da interface	URI
Servidor de histórico Flink (EMRversão 5.33 e posterior)	http://*master-public-dns-name* : 808/
Ganglia	http://*master-public-dns-name* /gânglios/
Hadoop HDFS NameNode (EMRversão pré-6.x)	https://*master-public-dns-name* : 50470/
Hadoop HDFS NameNode	http://*master-public-dns-name* : 50070/
Hadoop HDFS DataNode	http://*coretask-public-dns-name* : 500 75/
Hadoop HDFS NameNode (EMRversão 6.x)	https://*master-public-dns-name* : 9870/
Hadoop HDFS DataNode (EMRversão pré-6.x)	https://*coretask-public-dns-name* : 50 475/

Nome da interface	URI
Hadoop HDFS DataNode (EMRversão 6.x)	https://*coretask-public-dns-name* : 98 65/
HBase	http://*master-public-dns-name* : 16010/
Hue	http://*master-public-dns-name* :88/8
JupyterHub	https://*master-public-dns-name* : 9443/
Livy	http://*master-public-dns-name* : 8998/
Faísca HistoryServer	http://*master-public-dns-name* : 18080/
Tez	http://*master-public-dns-name* :8080/tez-ui
YARN NodeManager	http://*coretask-public-dns-name* : 804/
YARN ResourceManager	http://*master-public-dns-name* : 808/
Zeppelin	http://*master-public-dns-name* 8:890/

Como há várias interfaces específicas de aplicativos disponíveis no nó primário que não estão disponíveis nos nós principais e de tarefas, as instruções neste documento são específicas para o nó primário da AmazonEMR. O acesso as interfaces Web em todos os nós centrais e de tarefa pode ser feito da mesma maneira como você acessaria as interfaces Web no nó primário.

Existem várias maneiras de acessar as interfaces Web no nó primário. O método mais fácil e rápido é conectar-se SSH ao nó primário e usar o navegador baseado em texto, o Lynx, para visualizar os sites em seu cliente. SSH No entanto, o Lynx é um navegador baseado em texto com uma interface de usuário limitada que não pode exibir gráficos. O exemplo a seguir mostra como abrir a ResourceManager interface do Hadoop usando o Lynx (o Lynx também URLs é fornecido quando você faz login no nó primário usando). SSH

```
lynx http://ip-###-##-##-###.us-west-2.compute.internal:8088/
```

Existem duas opções restantes para acessar interfaces Web no nó primário que fornecem funcionalidade de navegador completa. Escolha uma das seguintes opções:

- Opção 1 (recomendada para usuários mais técnicos): use um SSH cliente para se conectar ao nó primário, configure o SSH tunelamento com encaminhamento de porta local e use um navegador da Internet para abrir interfaces da Web hospedadas no nó primário. Esse método permite configurar o acesso à interface da web sem usar um SOCKS proxy.
- Opção 2 (recomendada para novos usuários): use um SSH cliente para se conectar ao nó primário, configure o SSH tunelamento com encaminhamento dinâmico de portas e configure seu navegador da Internet para usar um complemento, como o Firefox ou SwitchyOmega o Chrome, FoxyProxy para gerenciar suas configurações de proxy. SOCKS Esse método permite filtrar automaticamente URLs com base em padrões de texto e limitar as configurações de proxy a domínios que correspondam ao formato do DNS nome do nó primário. Para obter mais informações sobre como configurar o FoxyProxy Firefox e o Google Chrome, consulteOpção 2, parte 2: configurar definições de proxy para visualizar sites hospedados no nó primário.

> **Note**
>
> Se você modificar a porta em que um aplicativo é executado por meio da configuração do cluster, o hiperlink para a porta não será atualizado no EMR console da Amazon. Isso ocorre porque o console não tem a funcionalidade de ler a configuração `server.port`.

Com a Amazon EMR versão 5.25.0 ou posterior, você pode acessar a interface do servidor de histórico do Spark a partir do console sem configurar um proxy web por meio de uma conexão. SSH Para obter mais informações, consulte One-click access to persistent Spark history server.

Tópicos

Opção 1: configurar um SSH túnel para o nó primário usando o encaminhamento de porta local

Para se conectar ao servidor web local no nó primário, você cria um SSH túnel entre o computador e o nó primário. Isso também é conhecido como encaminhamento de portas. Se você não quiser usar um SOCKS proxy, poderá configurar um SSH túnel para o nó primário usando o encaminhamento de porta local. Com o encaminhamento de portas locais, você pode especificar portas locais que são utilizadas para encaminhar o tráfego a portas remotas específicas no servidor Web local do nó primário.

Configurar um SSH túnel usando o encaminhamento de porta local requer o DNS nome público do nó primário e o arquivo de chave privada do seu par de chaves. Para obter informações sobre como localizar o DNS nome público principal, consulteRecupere o DNS nome público do nó primário. Para obter mais informações sobre como acessar seu par de chaves, consulte os pares de EC2 chaves da Amazon no Guia EC2 do usuário da Amazon. Para obter mais informações sobre os sites que você pode querer visualizar no nó primário, consulte Visualize interfaces web hospedadas em EMR clusters da Amazon.

Configure um SSH túnel para o nó primário usando o encaminhamento de porta local com Open SSH

Para configurar um SSH túnel usando o encaminhamento de porta local no terminal

1. Verifique se você permitiu o SSH tráfego de entrada. Para obter instruções, consulte Antes de se conectar: autorize o tráfego de entrada.
2. Abra uma janela do terminal. No Mac OS X, escolha Applications > Utilities > Terminal (Aplicativos > Utilitários > Terminal). Em outras distribuições do Linux, o terminal está normalmente localizado em Applications > Accessories > Terminal (Aplicativos > Acessórios > Terminal).
3. Digite o comando a seguir para abrir um SSH túnel na sua máquina local. Este exemplo de comando acessa a interface ResourceManager da web encaminhando o tráfego na porta local 8157 (uma porta local não usada escolhida aleatoriamente) para a porta 8088 no servidor web local do nó principal.

   No comando, substitua *~/mykeypair.pem* com a localização e o nome do `.pem` arquivo e substitua *ec2-###-##-##-###.compute-1.amazonaws.com* com o DNS nome público principal do seu cluster. Para acessar uma interface da Web diferente, `8088` substitua pelo número de porta apropriado. Por exemplo, `8088` substitua `8890` pela interface do Zeppelin.

```
ssh -i ~/mykeypair.pem -N -L 8157:ec2-###-##-##-
###.compute-1.amazonaws.com:8088 hadoop@ec2-###-##-##-###.compute-1.amazonaws.com
```

`-L` significa o uso do encaminhamento de portas locais, que permite especificar uma porta local usada para encaminhar dados à porta remota identificada no servidor Web local do nó principal.

Após a emissão desse comando, o terminal permanece aberto e não retorna uma resposta.

4. Para abrir a interface ResourceManager da web em seu navegador, digite `http://localhost:8157/` na barra de endereço.
5. Quando terminar de trabalhar com as interfaces Web no nó primário, feche as janelas do terminal.

Opção 2, parte 1: configurar um SSH túnel para o nó primário usando o encaminhamento dinâmico de portas

Para se conectar ao servidor web local no nó primário, você cria um SSH túnel entre o computador e o nó primário. Isso também é conhecido como encaminhamento de portas. Se você criar seu SSH túnel usando o encaminhamento dinâmico de portas, todo o tráfego roteado para uma porta local não utilizada especificada será encaminhado para o servidor web local no nó primário. Isso cria um SOCKS proxy. Em seguida, você pode configurar seu navegador da Internet para usar um complemento, como FoxyProxy ou SwitchyOmega para gerenciar suas configurações de SOCKS proxy.

O uso de um complemento de gerenciamento de proxy permite filtrar automaticamente URLs com base em padrões de texto e limitar as configurações de proxy a domínios que correspondam ao formato do nome público DNS do nó primário. O complemento do navegador manipula automaticamente a ativação e desativação do proxy quando você alterna entre visualizar sites hospedados no nó primário e aqueles na Internet.

Antes de começar, você precisa do DNS nome público do nó primário e do arquivo de chave privada do seu par de chaves. Para obter informações sobre como localizar o DNS nome público principal, consulteRecupere o DNS nome público do nó primário. Para obter mais informações sobre como acessar seu par de chaves, consulte os pares de EC2 chaves da Amazon no Guia EC2 do usuário da Amazon. Para obter mais informações sobre os sites que você pode querer visualizar no nó primário, consulte Visualize interfaces web hospedadas em EMR clusters da Amazon.

Configure um SSH túnel para o nó primário usando o encaminhamento dinâmico de portas com o Open SSH

Para configurar um SSH túnel usando o encaminhamento dinâmico de portas com Open SSH

1. Verifique se você permitiu o SSH tráfego de entrada. Para obter instruções, consulte Antes de se conectar: autorize o tráfego de entrada.
2. Abra uma janela do terminal. No Mac OS X, escolha Applications > Utilities > Terminal (Aplicativos > Utilitários > Terminal). Em outras distribuições do Linux, o terminal está normalmente localizado em Applications > Accessories > Terminal (Aplicativos > Acessórios > Terminal).
3. Digite o comando a seguir para abrir um SSH túnel na sua máquina local. Substituir *~/mykeypair.pem* com a localização e o nome do `.pem` arquivo, substitua *8157* com um número de porta local não utilizado e substitua *ec2-###-##-##-###.compute-1.amazonaws.com* com o DNS nome público principal do seu cluster.

   ```
   ssh -i ~/mykeypair.pem -N -D 8157 hadoop@ec2-###-##-##-###.compute-1.amazonaws.com
   ```

   Após a execução desse comando, o terminal permanece aberto e não retorna uma resposta.

   > **Note**
   >
   > `-D` significa o uso do encaminhamento de portas dinâmicas, que permite especificar uma porta local usada para encaminhar dados a todas as portas remotas identificadas no servidor Web local do nó primário. O encaminhamento dinâmico de portas cria um SOCKS proxy local escutando na porta especificada no comando.

4. Depois que o túnel estiver ativo, configure um SOCKS proxy para seu navegador. Para obter mais informações, consulte Opção 2, parte 2: configurar definições de proxy para visualizar sites hospedados no nó primário.
5. Quando terminar de trabalhar com as interfaces Web no nó primário, feche a janela do terminal.

Configure um SSH túnel usando o encaminhamento dinâmico de portas com o AWS CLI

Você pode criar uma SSH conexão com o nó primário usando o AWS CLI no Windows e no Linux, Unix e Mac OS X. Se você estiver usando o AWS CLI no Linux, Unix ou Mac OS X, deverá definir as permissões no `.pem` arquivo conforme mostrado em. Para configurar as permissões do arquivo

de chave privada do par de chaves Se você estiver usando o AWS CLI no Windows, Pu TTY deverá aparecer na variável de ambiente path ou você poderá receber um erro como Open SSH ou Pu TTY not available.

Para configurar um SSH túnel usando o encaminhamento dinâmico de portas com o AWS CLI

1. Verifique se você permitiu o SSH tráfego de entrada. Para obter instruções, consulte Antes de se conectar: autorize o tráfego de entrada.
2. Crie uma SSH conexão com o nó primário, conforme mostrado emConectar-se ao nó primário usando a AWS CLI.
3. Para recuperar o identificador de cluster, digite:

   ```
   aws emr list-clusters
   ```

   A saída lista seus clusters, incluindo o clusterIDs. Observe o ID do cluster ao qual você está se conectando.

   ```
   "Status": {
       "Timeline": {
           "ReadyDateTime": 1408040782.374,
           "CreationDateTime": 1408040501.213
       },
       "State": "WAITING",
       "StateChangeReason": {
           "Message": "Waiting after step completed"
       }
   },
   "NormalizedInstanceHours": 4,
   "Id": "j-2AL4XXXXXX5T9",
   "Name": "AWS CLI cluster"
   ```

4. Digite o comando a seguir para abrir um SSH túnel para o nó primário usando o encaminhamento dinâmico de portas. No exemplo a seguir, substitua *j-2AL4XXXXXX5T9* com o ID do cluster e substitua *~/mykeypair.key* com a localização e o nome do `.pem` arquivo (para Linux, Unix e Mac OS X) ou `.ppk` arquivo (para Windows).

   ```
   aws emr socks --cluster-id j-2AL4XXXXXX5T9 --key-pair-file ~/mykeypair.key
   ```

> **Note**
>
> O comando socks configura automaticamente o encaminhamento de portas dinâmicas na porta local 8157. Atualmente, essa configuração não pode ser modificada.

5. Depois que o túnel estiver ativo, configure um SOCKS proxy para seu navegador. Para obter mais informações, consulte Opção 2, parte 2: configurar definições de proxy para visualizar sites hospedados no nó primário.
6. Quando terminar de trabalhar com as interfaces da Web no nó primário, feche a AWS CLI janela.

   Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consultehttps://docs.aws.amazon.com/cli/latest/reference/emr.

Configure um SSH túnel para o nó primário usando Pu TTY

Os usuários do Windows podem usar um SSH cliente como o Pu TTY para criar um SSH túnel para o nó primário. Antes de se conectar ao nó EMR primário da Amazon, você deve baixar e instalar o Pu TTY e o uTTYgen P. Você pode baixar essas ferramentas na página de TTY download do Pu.

TTYO Pu não oferece suporte nativo ao formato de arquivo de chave privada do par de chaves (`.pem`) gerado pela AmazonEC2. Você usa P uTTYgen para converter seu arquivo de chave para o TTY formato Pu necessário (`.ppk`). Você deve converter sua chave nesse formato (`.ppk`) antes de tentar se conectar ao nó primário usando PuTTY.

Para obter mais informações sobre como converter sua chave, consulte Convertendo sua chave privada usando P uTTYgen no Guia EC2 do usuário da Amazon.

Para configurar um SSH túnel usando o encaminhamento dinâmico de portas usando Pu TTY

1. Verifique se você permitiu o SSH tráfego de entrada. Para obter instruções, consulte Antes de se conectar: autorize o tráfego de entrada.
2. Clique duas vezes `putty.exe` para iniciar o PuTTY. Você também pode iniciar o Pu na lista TTY de programas do Windows.

> **Note**
>
> Se você já tiver uma SSH sessão ativa com o nó primário, poderá adicionar um túnel clicando com o botão direito do mouse na barra de TTY título do Pu e escolhendo Alterar configurações.

3. Se necessário, na lista Category (Categoria), escolha Session (Sessão).
4. No campo Nome do host, digite **hadoop@***MasterPublicDNS*. Por exemplo: **hadoop@***ec2-###-##-##-###.compute-1.amazonaws.com*.
5. Na lista Categoria, expanda SSHConexão > e escolha Auth.
6. Para Private key file for authentication (Arquivo de chave privada para autenticação), escolha Browse (Procurar) e selecione o arquivo `.ppk` que você gerou.

> **Note**
>
> TTYO Pu não oferece suporte nativo ao formato de arquivo de chave privada do par de chaves (`.pem`) gerado pela AmazonEC2. Você usa P uTTYgen para converter seu arquivo de chave para o TTY formato Pu necessário (`.ppk`). Você deve converter sua chave nesse formato (`.ppk`) antes de tentar se conectar ao nó primário usando PuTTY.

7. Na lista Categoria, expanda Conexão > SSH e escolha Túneis.
8. No campo Porta de origem, digite `8157` (uma porta local não utilizada) e escolha Adicionar.
9. Deixe o campo Destination (Destino) em branco.
10. Selecione as opções Dynamic (Dinâmico) e Auto.
11. Escolha Open (Abrir).
12. Escolha Sim para ignorar o alerta de TTY segurança Pu.

> **Important**
>
> Ao fazer login no nó primário, digite `hadoop` se for solicitado um nome de usuário.

13. Depois que o túnel estiver ativo, configure um SOCKS proxy para seu navegador. Para obter mais informações, consulte Opção 2, parte 2: configurar definições de proxy para visualizar sites hospedados no nó primário.
14. Quando terminar de trabalhar com as interfaces da Web no nó primário, feche a TTY janela Pu.

Opção 2, parte 2: configurar definições de proxy para visualizar sites hospedados no nó primário

Se você usar um SSH túnel com encaminhamento dinâmico de portas, deverá usar um complemento de gerenciamento de SOCKS proxy para controlar as configurações de proxy no seu navegador. O uso de uma ferramenta de gerenciamento de SOCKS proxy permite filtrar automaticamente URLs com base em padrões de texto e limitar as configurações de proxy a domínios que correspondam ao formato do DNS nome público do nó primário. O complemento do navegador manipula automaticamente a ativação e desativação do proxy quando você alterna entre visualizar sites hospedados no nó primário e aqueles na Internet. Para gerenciar suas configurações de proxy, configure seu navegador para usar um complemento como FoxyProxy ou SwitchyOmega.

Para obter mais informações sobre a criação de um SSH túnel, consulteOpção 2, parte 1: configurar um SSH túnel para o nó primário usando o encaminhamento dinâmico de portas. Para obter mais informações sobre as interfaces Web disponíveis, consulte Visualize interfaces web hospedadas em EMR clusters da Amazon.

Inclua as seguintes configurações ao definir o complemento de proxy:

- Use localhost como endereço do host.
- Use o mesmo número de porta local que você selecionou para estabelecer o SSH túnel com o nó primárioOpção 2, parte 1: configurar um SSH túnel para o nó primário usando o encaminhamento dinâmico de portas. Por exemplo, porta *8157*. Essa porta também deve corresponder ao número da porta que você usa no Pu TTY ou em qualquer outro emulador de terminal usado para se conectar.
- Especifique o protocolo SOCKSv5. SOCKSA v5 permite que você configure opcionalmente a autorização do usuário.
- URLPadrões

  Os URL padrões a seguir devem ser listados como permitidos e especificados com um tipo de padrão curinga:

  - O *ec2*. *compute os padrões *.amazonaws.com* e *10*.amazonaws.com* para corresponder ao nome público dos clusters nas regiões dos EUA. DNS
  - Os padrões *ec2*.compute* e *10*.compute* correspondem ao nome público dos clusters em todas as outras regiões. DNS
  - UM 10. * padrão para fornecer acesso aos arquivos de JobTracker log no Hadoop. Altere esse filtro se ele entrar em conflito com seu plano de acesso de rede.

- Os padrões *.ec2.internal* e *.compute.internal* correspondem aos DNS nomes privados (internos) dos clusters na região e em todas as outras regiões, respectivamente. `us-east-1`

Exemplo: Configurar FoxyProxy para o Firefox

O exemplo a seguir demonstra uma configuração FoxyProxy padrão (versão 7.5.1) para o Mozilla Firefox.

FoxyProxy fornece um conjunto de ferramentas de gerenciamento de proxy. Ele permite URLs que você use um servidor proxy para esses padrões de correspondência correspondentes aos domínios usados pelas EC2 instâncias da Amazon em seu EMR cluster da Amazon.

Para instalar e configurar FoxyProxy usando o Mozilla Firefox

1. No Firefox, acesse https://addons.mozilla.org/, pesquise por FoxyProxy Padrão e siga as instruções para adicionar FoxyProxy ao Firefox.
2. Usando um editor de texto, crie um JSON arquivo com o nome `foxyproxy-settings.json` do exemplo de configuração a seguir.

```
{
  "k20d21508277536715": {
    "active": true,
    "address": "localhost",
    "port": 8157,
    "username": "",
    "password": "",
    "type": 3,
    "proxyDNS": true,
    "title": "emr-socks-proxy",
    "color": "#0055E5",
    "index": 9007199254740991,
    "whitePatterns": [
      {
        "title": "*ec2*.*compute*.amazonaws.com*",
        "active": true,
        "pattern": "*ec2*.*compute*.amazonaws.com*",
        "importedPattern": "*ec2*.*compute*.amazonaws.com*",
        "type": 1,
        "protocols": 1
      },
      {
```

```
        "title": "*ec2*.compute*",
        "active": true,
        "pattern": "*ec2*.compute*",
        "importedPattern": "*ec2*.compute*",
        "type": 1,
        "protocols": 1
      },
      {
        "title": "10.*",
        "active": true,
        "pattern": "10.*",
        "importedPattern": "http://10.*",
        "type": 1,
        "protocols": 2
      },
      {
        "title": "*10*.amazonaws.com*",
        "active": true,
        "pattern": "*10*.amazonaws.com*",
        "importedPattern": "*10*.amazonaws.com*",
        "type": 1,
        "protocols": 1
      },
      {
        "title": "*10*.compute*",
        "active": true,
        "pattern": "*10*.compute*",
        "importedPattern": "*10*.compute*",
        "type": 1,
        "protocols": 1
      },
      {
        "title": "*.compute.internal*",
        "active": true,
        "pattern": "*.compute.internal*",
        "importedPattern": "*.compute.internal*",
        "type": 1,
        "protocols": 1
      },
      {
        "title": "*.ec2.internal* ",
        "active": true,
        "pattern": "*.ec2.internal*",
        "importedPattern": "*.ec2.internal*",
```

```
        "type": 1,
        "protocols": 1
      }
    ],
    "blackPatterns": []
  },
  "logging": {
    "size": 100,
    "active": false
  },
  "mode": "patterns",
  "browserVersion": "68.12.0",
  "foxyProxyVersion": "7.5.1",
  "foxyProxyEdition": "standard"
}
```

3. Abra a página Gerenciamento de extensões do Firefox (acesse about:addons e escolha Extensões).
4. Escolha FoxyProxy Padrão e, em seguida, escolha o botão de mais opções (o botão que parece uma elipse).
5. Selecione Opções no menu suspenso.
6. Escolha Importar configurações no menu esquerdo.
7. Na página Configurações de importação, escolha Importar configurações em Importar configurações do FoxyProxy 6.0+, navegue até o local do **`foxyproxy-settings.json`** arquivo que você criou, selecione o arquivo e escolha Abrir.
8. Escolha OK quando solicitado para substituir as configurações atuais e salvar a nova configuração.

Exemplo: Configurar SwitchyOmega para o Chrome

O exemplo a seguir demonstra como configurar a SwitchyOmega extensão para o Google Chrome. SwitchyOmega permite configurar, gerenciar e alternar entre vários proxies.

Para instalar e configurar SwitchyOmega usando o Google Chrome

1. Vá para https://chrome.google.com/webstore/categoria/extensões, pesquise por Proxy SwitchyOmega e adicione-o ao Chrome.
2. Escolha Novo perfil e insira `emr-socks-proxy` como nome do perfil.

3. Escolha o PACperfil e, em seguida, Criar. Os arquivos de configuração automática de proxy (PAC) ajudam você a definir uma lista de permissões para solicitações do navegador que devem ser encaminhadas para um servidor proxy da web.
4. No campo PACScript, substitua o conteúdo pelo script a seguir, que define o que URLs deve ser encaminhado por meio do seu servidor proxy da web. Se você especificou um número de porta diferente ao configurar seu SSH túnel, substitua *8157* com o número da sua porta.

```
function FindProxyForURL(url, host) {
    if (shExpMatch(url, "*ec2*.*compute*.amazonaws.com*")) return 'SOCKS5
 localhost:8157';
    if (shExpMatch(url, "*ec2*.compute*")) return 'SOCKS5 localhost:8157';
    if (shExpMatch(url, "http://10.*")) return 'SOCKS5 localhost:8157';
    if (shExpMatch(url, "*10*.compute*")) return 'SOCKS5 localhost:8157';
    if (shExpMatch(url, "*10*.amazonaws.com*")) return 'SOCKS5 localhost:8157';
    if (shExpMatch(url, "*.compute.internal*")) return 'SOCKS5 localhost:8157';
    if (shExpMatch(url, "*ec2.internal*")) return 'SOCKS5 localhost:8157';
    return 'DIRECT';
}
```

5. Em Ações, escolha Aplicar alterações para salvar as configurações de proxy.
6. Na barra de ferramentas do Chrome, escolha SwitchyOmega e selecione o `emr-socks-proxy` perfil.

Acessar uma interface da Web no navegador

Para abrir uma interface da web, insira o DNS nome público do seu nó principal ou principal seguido pelo número da porta da interface escolhida na barra de endereço do navegador. O exemplo a seguir mostra o URL que você digitaria para se conectar ao Spark HistoryServer.

```
http://master-public-dns-name:18080/
```

Para obter instruções sobre como recuperar o DNS nome público de um nó, consulteRecupere o DNS nome público do nó primário. Para obter uma lista completa da interface da webURLs, consulteVisualize interfaces web hospedadas em EMR clusters da Amazon.

Enviar trabalhos a um cluster

Esta seção descreve os métodos que você pode usar para enviar trabalhos para um EMR cluster da Amazon. Para enviar trabalhos, é possível adicionar etapas ou enviar trabalhos do Hadoop de forma interativa para o nó primário.

Considere estas regras de comportamento de etapas ao enviar etapas para um cluster:

- Um ID de etapa pode conter até 256 caracteres.
- Você pode ter até 256 PENDING e RUNNING etapas em um cluster.
- Mesmo com 256 etapas ativas em execução no cluster, é possível enviar trabalhos de forma interativa ao nó primário. Você pode enviar um número ilimitado de etapas durante a vida útil de um cluster de longa execução, mas somente 256 etapas podem ser realizadas RUNNING ou a qualquer PENDING momento.
- Com EMR as versões 4.8.0 e posteriores da Amazon, exceto a versão 5.0.0, você pode cancelar etapas pendentes. Para obter mais informações, consulte Cancelar etapas.
- Com EMR as versões 5.28.0 e posteriores da Amazon, você pode cancelar as etapas pendentes e em execução. Você também pode optar por executar várias etapas em paralelo para melhorar a utilização de cluster e economizar custos. Para obter mais informações, consulte Considerações sobre a execução de várias etapas em paralelo.

> **Note**
>
> Para obter o melhor desempenho, recomendamos que você armazene ações de bootstrap, scripts e outros arquivos personalizados que você deseja usar com a Amazon EMR em um bucket do Amazon S3 que esteja no Região da AWS mesmo que seu cluster.

Tópicos

Adicionar etapas a um cluster com o Amazon EMR Management Console

Realize os procedimentos a seguir para adicionar etapas a um cluster usando o AWS Management Console. Para obter informações detalhadas sobre como enviar etapas para aplicativos específicos de big data, consulte as seguintes seções do Amazon EMR Release Guide:

- Envie uma JAR etapa personalizada
- Enviar uma etapa de transmissão do Hadoop
- Enviar uma etapa do Spark
- Enviar uma etapa do Pig
- Executar um comando ou script como etapa
- Transmitir valores em etapas para executar scripts do Hive

Adicionar etapas durante a criação do cluster

A partir do AWS Management Console, você pode adicionar etapas ao criar um cluster.

Console

Para adicionar etapas ao criar um cluster com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Em Etapas, escolha Adicionar etapa. Insira os valores apropriados nos campos da caixa de diálogo Adicionar etapa. Para obter informações sobre como formatar argumentos de etapa, consulte Adicionar argumentos de etapas. As opções diferem dependendo do tipo de etapa. Para adicionar a etapa e sair da caixa de diálogo, selecione Adicionar etapa.
4. Escolha qualquer outra opção que se aplique ao cluster.
5. Para iniciar o cluster, escolha Criar cluster.

Adicionar etapas a um cluster em execução

Com o AWS Management Console, você pode adicionar etapas a um cluster com a opção de encerramento automático desativada.

Console

Para adicionar etapas a um cluster em execução com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EMREC2Em Ativado, no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar.
3. Na guia Etapas da página de detalhes do cluster, selecione a guia Adicionar etapa. Para clonar uma etapa já existente, escolha o menu suspenso Ações e selecione Clonar etapa.
4. Insira os valores apropriados nos campos da caixa de diálogo Adicionar etapa. As opções diferem dependendo do tipo de etapa. Para adicionar a etapa e sair da caixa de diálogo, escolha Adicionar etapa.

Modificar o nível de simultaneidade da etapa em um cluster em execução

Com o AWS Management Console, você pode modificar o nível de simultaneidade de etapas em um cluster em execução.

> **Note**
>
> Você só pode executar várias etapas em paralelo com a Amazon EMR versão 5.28.0 e posterior.

Console

Para modificar a simultaneidade de etapas em um cluster em execução com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EMREC2Em Ativado, no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar. O cluster deve estar em execução para alterar o respectivo atributo de simultaneidade.
3. Na guia Etapas da página de detalhes do cluster, encontre a seção Atributos. Selecione Editar para alterar a simultaneidade. Insira um valor entre 1 e 256.

Adicionar argumentos de etapas

Ao usar o AWS Management Console para adicionar uma etapa ao seu cluster, você pode especificar argumentos para essa etapa no campo Argumentos. É necessário separar argumentos com espaço em branco e cercar com aspas argumentos de sequência de caracteres formados por caracteres e espaços em branco.

Example : Argumentos corretos

Os argumentos de exemplo a seguir estão formatados corretamente para o AWS Management Console, com aspas ao redor do argumento final da string.

```
bash -c "aws s3 cp s3://DOC-EXAMPLE-BUCKET/my-script.sh ."
```

Também é possível colocar cada argumento em uma linha separada para facilitar a leitura, como mostra o exemplo a seguir.

```
bash
-c
"aws s3 cp s3://DOC-EXAMPLE-BUCKET/my-script.sh ."
```

Example : Argumentos incorretos

Os argumentos de exemplo a seguir estão formatados incorretamente para o AWS Management Console. Observa-se que o argumento final da string ,`aws s3 cp s3://DOC-EXAMPLE-BUCKET/my-script.sh .`, contém espaços em branco e não está entre aspas.

```
bash -c aws s3 cp s3://DOC-EXAMPLE-BUCKET/my-script.sh .
```

Adicionando etapas a um cluster com o AWS CLI

Os procedimentos a seguir demonstram como adicionar etapas a um cluster recém-criado e a um cluster em execução com a AWS CLI. Ambos os exemplos usam o subcomando `--steps` para adicionar etapas ao cluster.

Para adicionar etapas durante a criação do cluster

- Digite o seguinte comando para criar um cluster e adicionar uma etapa do Apache Pig. Certifique-se de substituir *myKey* com o nome do seu par de EC2 chaves da Amazon.

```
aws emr create-cluster --name "Test cluster" \
--applications Name=Spark \
--use-default-roles \
--ec2-attributes KeyName=myKey \
--instance-groups InstanceGroupType=PRIMARY,InstanceCount=1,InstanceType=m5.xlarge
 InstanceGroupType=CORE,InstanceCount=2,InstanceType=m5.xlarge \
--steps '[{"Args":["spark-submit","--deploy-mode","cluster","--
class","org.apache.spark.examples.SparkPi","/usr/lib/spark/examples/jars/spark-
examples.jar","5"],"Type":"CUSTOM_JAR","ActionOnFailure":"CONTINUE","Jar":"command-
runner.jar","Properties":"","Name":"Spark application"}]'
```

> **Note**
>
> A lista de argumentos muda dependendo do tipo de etapa.

Por padrão, o nível de simultaneidade da etapa é 1. É possível definir o nível de simultaneidade da etapa usando o parâmetro `StepConcurrencyLevel` ao criar um cluster.

A saída é um identificador de cluster semelhante ao seguinte:

```
{
    "ClusterId": "j-2AXXXXXXGAPLF"
}
```

Para adicionar uma etapa a um cluster em execução

- Digite o seguinte comando para adicionar uma etapa a um cluster em execução. Substitua *j-2AXXXXXXGAPLF* por seu próprio ID do cluster.

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps '[{"Args":["spark-submit","--deploy-mode","cluster","--
class","org.apache.spark.examples.SparkPi","/usr/lib/spark/examples/jars/spark-
examples.jar","5"],"Type":"CUSTOM_JAR","ActionOnFailure":"CONTINUE","Jar":"command-
runner.jar","Properties":"","Name":"Spark application"}]'
```

A saída é um identificador de etapa semelhante ao seguinte:

```
{
    "StepIds": [
 "s-Y9XXXXXXAPMD"
    ]
}
```

Para modificar o StepConcurrencyLevel em um cluster em execução

1. Em um cluster em execução, você pode modificar o `StepConcurrencyLevel` com `ModifyCluster` API o. Por exemplo, digite o seguinte comando para aumentar o `StepConcurrencyLevel` para `10`. Substitua *`j-2AXXXXXXGAPLF`* pelo nome do ID do cluster.

```
aws emr modify-cluster --cluster-id j-2AXXXXXXGAPLF --step-concurrency-level 10
```

2. A saída é semelhante à seguinte.

```
{
"StepConcurrencyLevel": 10
}
```

Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consulte a Referência de AWS CLI comandos.

Considerações sobre a execução de várias etapas em paralelo

- As etapas executadas em paralelo podem ser concluídas em qualquer ordem, mas as etapas pendentes na fila passam para o estado de execução na ordem em que são enviadas.
- Ao selecionar um nível de simultaneidade da etapa para o cluster, você deve considerar se o tipo de instância de nó primário atende ou não aos requisitos de memória das workloads do usuário. O processo executor da etapa principal é executado no nó primário de cada etapa. A execução de várias etapas em paralelo exige mais memória e CPU utilização do nó primário do que executar uma etapa por vez.
- Para obter um agendamento complexo e o gerenciamento de recursos de etapas simultâneas, você pode usar recursos de YARN agendamento, como ou. `FairScheduler` `CapacityScheduler` Por exemplo, você pode usar o `FairScheduler` com um conjunto

`queueMaxAppsDefault` para impedir que mais de um determinado número de trabalhos seja executado por vez.

- O nível de simultaneidade da etapa está sujeito às configurações dos gerenciadores de recursos. Por exemplo, se YARN estiver configurado com apenas um paralelismo de`5`, você só poderá ter cinco YARN aplicativos em execução em paralelo, mesmo que `StepConcurrencyLevel` esteja definido como. `10` Para obter mais informações sobre a configuração de gerenciadores de recursos, consulte Configurar aplicativos no Amazon EMR Release Guide.
- Você não pode adicionar uma etapa com `ActionOnFailure` outra CONTINUE enquanto o nível de simultaneidade de etapas do cluster for maior que 1.
- Se o nível de simultaneidade de etapas do cluster for maior que 1, o atributo `ActionOnFailure` da etapa não será ativado.
- Se o nível de simultaneidade de etapas do cluster for 1, mas houver várias etapas em execução, `TERMINATE_CLUSTER ActionOnFailure` poderá ser ativada, mas não `CANCEL_AND_WAIT` `ActionOnFailure` será. Esse caso extremo ocorre quando o nível de simultaneidade da etapa do cluster é maior que 1, mas diminui durante a execução de várias etapas.
- Você pode usar o escalonamento EMR automático para aumentar e diminuir a escala com base nos YARN recursos para evitar a contenção de recursos. Para obter mais informações, consulte Como usar a escalabilidade automática com uma política personalizada para grupos de instâncias no Amazon EMR Management Guide.
- Quando você diminui o nível simultâneo da etapa, EMR permite que todas as etapas em execução sejam concluídas antes de reduzir o número de etapas. Se os recursos estiverem esgotados porque o cluster está executando muitas etapas simultâneas, recomendamos cancelar as etapas em execução manualmente para liberar recursos.

Visualizar etapas

Você pode ver até 10.000 etapas que a Amazon EMR concluiu nos últimos sete dias. Você também pode ver 1.000 etapas que a Amazon EMR concluiu a qualquer momento. Esse total inclui tanto etapas do sistema quanto etapas enviadas pelo usuário.

Se você enviar novas etapas quando o cluster atingir o limite de registros de 1.000 etapas, a Amazon EMR excluirá as etapas inativas enviadas pelo usuário cujos status foram COMPLETED ou FAILED há mais de sete dias. CANCELLED Se você enviar etapas além do limite de registro de 10.000 etapas, a Amazon EMR excluirá os registros de etapas inativas enviadas pelo usuário, independentemente de sua duração inativa. A Amazon EMR não remove esses registros dos

arquivos de log. A Amazon os EMR remove do AWS console e eles não são retornados quando você usa o AWS CLI ou API para recuperar informações do cluster. Registros de etapas do sistema nunca são removidos.

As informações de etapas que você pode visualizar dependem do mecanismo usado para recuperar informações do cluster. A tabela a seguir indica as informações de etapa retornadas por cada uma das opções disponíveis.

Opção	DescribeJobFlow ou --describe --jobflow	ListSteps ou lista-etapas
SDK	256 etapas	Até 10.000 etapas
Amazon EMR CLI	256 etapas	N/D
AWS CLI	N/D	Até 10.000 etapas
API	256 etapas	Até 10.000 etapas

Cancelar etapas

Você pode cancelar etapas pendentes e em execução da AWS Management Console AWS CLI, da ou da Amazon EMRAPI.

Console

Para cancelar etapas com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EMREC2Em Ativado, no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar.
3. Na guia Etapas da página de detalhes do cluster, marque a caixa de seleção ao lado da etapa que você deseja cancelar. Escolha o menu suspenso Ações e selecione Cancelar etapas.
4. Na caixa de diálogo Cancelar a etapa, escolha entre cancelar a etapa e esperar a saída ou cancelar a etapa e forçar a saída. Depois, selecione Confirm (Confirmar).

5. O status das etapas na tabela Etapas é alterado para `CANCELLED`.

CLI

Para cancelar usando o AWS CLI

- Use o comando `aws emr cancel-steps`, especificando o cluster e as etapas a serem canceladas. O exemplo a seguir demonstra um comando da AWS CLI para cancelar duas etapas.

```
aws emr cancel-steps --cluster-id j-2QUAXXXXXXXXX \
--step-ids s-3M8DXXXXXXXXX s-3M8DXXXXXXXXX \
--step-cancellation-option SEND_INTERRUPT
```

Com a EMR versão 5.28.0 da Amazon, você pode escolher uma das duas opções de cancelamento a seguir como `StepCancellationOption` parâmetro ao cancelar etapas.

- `SEND_INTERRUPT`: essa é a opção padrão. Quando uma solicitação de cancelamento de etapa é recebida, EMR envia um `SIGTERM` sinal para a etapa. Adicione um manipulador de `SIGTERM` sinal à sua lógica de etapas para capturar esse sinal e encerrar os processos da etapa descendente ou esperar que eles sejam concluídos.
- `TERMINATE_PROCESS`— Quando essa opção é selecionada, EMR envia um `SIGKILL` sinal para a etapa e todos os seus processos descendentes, o que os encerra imediatamente.

Considerações sobre o cancelamento de etapas

- Cancelar uma etapa em execução ou pendente removerá a etapa da contagem de etapas ativas.
- Cancelar uma etapa em execução não permite que uma etapa pendente comece a ser executada, supondo que não haja alteração em `stepConcurrencyLevel`.
- O cancelamento de uma etapa em execução não aciona a etapa `ActionOnFailure`.
- Para EMR 5.32.0 e versões posteriores, `SEND_INTERRUPT StepCancellationOption` envia um `SIGTERM` sinal para o processo Step Child. Observe esse sinal e faça uma limpeza e desligue-o normalmente. `TERMINATE_PROCESS StepCancellationOption` envia um sinal `SIGKILL` para o processo filho da etapa e para todos os seus processos descendentes; mas os processos assíncronos não são afetados.

Visualizar e monitorar um cluster

EMRA Amazon fornece várias ferramentas que você pode usar para coletar informações sobre seu cluster. Você pode acessar informações sobre o cluster a partir do console CLI ou de forma programática. As interfaces Web padrão do Hadoop e os arquivos de log estão disponíveis no nó primário. Você também pode usar serviços de monitoramento, como o CloudWatch Ganglia, para monitorar o desempenho do seu cluster.

O histórico do aplicativo também está disponível no console usando o aplicativo UIs “persistente” do Spark History Server a partir do Amazon EMR 5.25.0. Com o Amazon EMR 6.x, o servidor de YARN cronograma persistente e as interfaces de usuário do Tez também estão disponíveis. Esses serviços são hospedados fora do cluster, para que você possa acessar o histórico do aplicativo por 30 dias após o encerramento do cluster, sem a necessidade de uma SSH conexão ou proxy web. Consulte Visualizar o histórico da aplicação.

Tópicos

Visualizar o status e os detalhes do cluster

Depois de criar um cluster, você pode monitorar seu status e obter informações detalhadas sobre sua execução e erros que podem ter ocorrido, mesmo depois de ele ter sido terminado. A Amazon EMR salva metadados sobre clusters encerrados para sua referência por dois meses, após os quais os metadados são excluídos. Você não pode excluir clusters do histórico de clusters, mas, usando o AWS Management Console, você pode usar o Filter (Filtro) e, usando a AWS CLI, você pode usar opções com o comando `list-clusters` para focalizar nos clusters que interessam a você.

Você pode acessar o histórico do aplicativo armazenado no cluster por uma semana a partir de sua gravação, independentemente de se o cluster está em execução ou encerrado. Além disso, as

interfaces do usuário de aplicativos persistentes armazenam o histórico de aplicativos fora do cluster por 30 dias após o encerramento de um cluster. Consulte Visualizar o histórico da aplicação.

Para obter mais informações sobre estados de cluster, como Waiting e Running, consulte Noções básicas sobre o ciclo de vida do cluster.

Visualizar os detalhes do cluster usando o AWS Management Console

A lista Clusters no https://console.aws.amazon.com/emr lista todos os clusters em sua conta e AWS região, incluindo clusters encerrados. A lista mostra o seguinte para cada cluster: o nome e a ID, os detalhes de status e status, a hora de criação, o tempo decorrido em que o cluster estava em execução e as horas de instância normalizada que foram acumuladas para todas as EC2 instâncias no cluster. Essa lista é o ponto de partida para monitorar o status dos clusters. Ela foi criada para que você possa analisar detalhadamente cada cluster para análise e solução de problemas.

Console

Para visualizar as informações do cluster com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EMREC2Em Ativado, no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja visualizar.
3. Use o painel Resumo para ver os conceitos básicos da configuração do seu cluster, como o status do cluster, os aplicativos de código aberto que a Amazon EMR instalou no cluster e a versão da Amazon EMR que você usou para criar o cluster. Use cada guia abaixo do Resumo para visualizar informações, conforme descrito na tabela a seguir.

Visualize os detalhes do cluster usando o AWS CLI

Os exemplos a seguir demonstram como recuperar detalhes de cluster usando a AWS CLI. Para obter mais informações sobre os comandos disponíveis, consulte a Referência de AWS CLI comandos da Amazon EMR. Você pode usar o comando describe-cluster para ver detalhes em nível de cluster, incluindo status, configuração de hardware e software, VPC configurações, ações de bootstrap, grupos de instâncias e assim por diante. Para obter mais informações sobre estados de cluster, consulte Noções básicas sobre o ciclo de vida do cluster. O exemplo a seguir demonstra o uso do comando `describe-cluster`, seguido por exemplos do comando list-clusters.

Example Visualizar o status do cluster

Para usar o comando `describe-cluster`, você precisa do ID do cluster. Este exemplo demonstra como usar para obter uma lista de clusters criados em um determinado intervalo de datas e, em seguida, usar um dos clusters IDs retornados para listar mais informações sobre o status de um cluster individual.

O comando a seguir descreve o cluster *j-1K48XXXXXXHCB*, que você substitui pelo ID do cluster.

```
aws emr describe-cluster --cluster-id j-1K48XXXXXXHCB
```

A saída do comando é semelhante à seguinte.

```
{
    "Cluster": {
        "Status": {
            "Timeline": {
                "ReadyDateTime": 1438281058.061,
                "CreationDateTime": 1438280702.498
            },
            "State": "WAITING",
            "StateChangeReason": {
                "Message": "Waiting for steps to run"
            }
        },
        "Ec2InstanceAttributes": {
            "EmrManagedMasterSecurityGroup": "sg-cXXXXX0",
            "IamInstanceProfile": "EMR_EC2_DefaultRole",
            "Ec2KeyName": "myKey",
            "Ec2AvailabilityZone": "us-east-1c",
            "EmrManagedSlaveSecurityGroup": "sg-example"
        },
        "Name": "Development Cluster",
        "ServiceRole": "EMR_DefaultRole",
        "Tags": [],
        "TerminationProtected": false,
        "ReleaseLabel": "emr-4.0.0",
        "NormalizedInstanceHours": 16,
        "InstanceGroups": [
            {
                "RequestedInstanceCount": 1,
                "Status": {
                    "Timeline": {
```

```
                    "ReadyDateTime": 1438281058.101,
                    "CreationDateTime": 1438280702.499
                },
                "State": "RUNNING",
                "StateChangeReason": {
                    "Message": ""
                }
            },
            "Name": "CORE",
            "InstanceGroupType": "CORE",
            "Id": "ig-2EEXAMPLEXXP",
            "Configurations": [],
            "InstanceType": "m5.xlarge",
            "Market": "ON_DEMAND",
            "RunningInstanceCount": 1
        },
        {
            "RequestedInstanceCount": 1,
            "Status": {
                "Timeline": {
                    "ReadyDateTime": 1438281023.879,
                    "CreationDateTime": 1438280702.499
                },
                "State": "RUNNING",
                "StateChangeReason": {
                    "Message": ""
                }
            },
            "Name": "MASTER",
            "InstanceGroupType": "MASTER",
            "Id": "ig-2A1234567XP",
            "Configurations": [],
            "InstanceType": "m5.xlarge",
            "Market": "ON_DEMAND",
            "RunningInstanceCount": 1
        }
    ],
    "Applications": [
        {
            "Version": "1.0.0",
            "Name": "Hive"
        },
        {
            "Version": "2.6.0",
```

```
                "Name": "Hadoop"
            },
            {
                "Version": "0.14.0",
                "Name": "Pig"
            },
            {
                "Version": "1.4.1",
                "Name": "Spark"
            }
        ],
        "BootstrapActions": [],
        "MasterPublicDnsName": "ec2-X-X-X-X.compute-1.amazonaws.com",
        "AutoTerminate": false,
        "Id": "j-jobFlowID",
        "Configurations": [
            {
                "Properties": {
                    "hadoop.security.groups.cache.secs": "250"
                },
                "Classification": "core-site"
            },
            {
                "Properties": {
                    "mapreduce.tasktracker.reduce.tasks.maximum": "5",
                    "mapred.tasktracker.map.tasks.maximum": "2",
                    "mapreduce.map.sort.spill.percent": "90"
                },
                "Classification": "mapred-site"
            },
            {
                "Properties": {
                    "hive.join.emit.interval": "1000",
                    "hive.merge.mapfiles": "true"
                },
                "Classification": "hive-site"
            }
        ]
    }
}
```

Example Listar clusters por data de criação

Para recuperar clusters criados em um intervalo de dados específicos, use o comando `list-clusters` com os parâmetros `--created-after` e `--created-before`.

O comando a seguir lista todos os clusters criados entre 9 e 12 de outubro de 2019.

```
aws emr list-clusters --created-after 2019-10-09T00:12:00 --created-
before 2019-10-12T00:12:00
```

Example Listar clusters por estado

Para listar clusters por estado, use o comando `list-clusters` com o parâmetro `--cluster-states`. Os estados de cluster válidos incluem: STARTING BOOTSTRAPPINGRUNNING,WAITING,TERMINATING,TERMINATED, e TERMINATED _ WITH _ERRORS.

```
aws emr list-clusters --cluster-states TERMINATED
```

Você também pode usar os seguintes parâmetros de atalho para listar todos os clusters nos estados especificados:

- `--active`filtra clusters nos TERMINATING estados STARTING BOOTSTRAPPINGRUNNING,WAITING,, ou.
- `--terminated`filtra clusters no TERMINATED estado.
- `--failed`o parâmetro filtra clusters no ERRORS estado TERMINATED WITH _ _.

As seguintes comandos retornam o mesmo resultado.

```
aws emr list-clusters --cluster-states TERMINATED
```

```
aws emr list-clusters --terminated
```

Para obter mais informações sobre estados de cluster, consulte Noções básicas sobre o ciclo de vida do cluster.

Etapa aprimorada de depuração

Se uma EMR etapa da Amazon falhar e você enviar seu trabalho usando a API operação Step com uma AMI versão 5.x ou posterior, a Amazon EMR poderá identificar e retornar a causa raiz da falha da etapa em alguns casos, junto com o nome do arquivo de log relevante e uma parte do rastreamento da pilha do aplicativo via. API Por exemplo, as seguintes falhas podem ser identificadas:

- Um erro do Hadoop comum, como o diretório de saída já existe, o diretório de entrada não existe ou um aplicativo ficou sem memória.
- Erros de Java, como um aplicativo que foi compilado com uma versão incompatível do Java ou executado com uma classe principal não encontrada.
- Um problema ao acessar objetos armazenados no Amazon S3.

Essas informações estão disponíveis usando DescribeStepas ListStepsAPIoperações e. O FailureDetailscampo do StepSummaryretornado por essas operações. Para acessar as FailureDetails informações, use o AWS CLI console ou AWS SDK.

Console

O novo EMR console da Amazon não oferece depuração por etapas. No entanto, é possível visualizar os detalhes do encerramento do cluster realizando as etapas a seguir.

Para ver os detalhes da falha com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, selecione o cluster que você deseja visualizar.
3. Observe o valor do Status na seção Resumo da página de detalhes do cluster. Se o status for Terminado com erros, passe o mouse sobre o texto para visualizar os detalhes da falha do cluster.

CLI

Para ver os detalhes da falha com o AWS CLI

- Para obter detalhes da falha de uma etapa com o AWS CLI, use o `describe-step` comando.

```
aws emr describe-step --cluster-id j-1K48XXXXXHCB --step-id s-3QM0XXXXXM1W
```

A saída será semelhante à seguinte:

```
{
  "Step": {
    "Status": {
      "FailureDetails": {
        "LogFile": "s3://myBucket/logs/j-1K48XXXXXHCB/steps/s-3QM0XXXXXM1W/
stderr.gz",
        "Message": "org.apache.hadoop.mapred.FileAlreadyExistsException: Output
 directory s3://myBucket/logs/beta already exists",
        "Reason": "Output directory already exists."
      },
      "Timeline": {
        "EndDateTime": 1469034209.143,
        "CreationDateTime": 1469033847.105,
        "StartDateTime": 1469034202.881
      },
      "State": "FAILED",
      "StateChangeReason": {}
    },
    "Config": {
      "Args": [
        "wordcount",
        "s3://myBucket/input/input.txt",
        "s3://myBucket/logs/beta"
      ],
      "Jar": "s3://myBucket/jars/hadoop-mapreduce-examples-2.7.2-amzn-1.jar",
      "Properties": {}
    },
    "Id": "s-3QM0XXXXXM1W",
    "ActionOnFailure": "CONTINUE",
    "Name": "ExampleJob"
  }
```

```
}
```

Visualizar o histórico da aplicação

Você pode visualizar os detalhes do Spark History Server e YARN do aplicativo de serviço de cronograma com a página de detalhes do cluster no console. O histórico de EMR aplicativos da Amazon facilita a solução de problemas e a análise de trabalhos ativos e do histórico de trabalhos.

> **Note**
>
> Para aumentar a segurança dos aplicativos fora do console que você pode usar com a AmazonEMR, os domínios de hospedagem de aplicativos são registrados na Lista Pública de Sufixos (). PSL Exemplos desses domínios de hospedagem incluem os seguintes: `emrstudio-prod.us-east-1.amazonaws.com`, `emrnotebooks-prod.us-east-1.amazonaws.com`, `emrappui-prod.us-east-1.amazonaws.com`. Para maior segurança, se precisar definir cookies confidenciais no nome de domínio padrão, recomendamos que você use cookies com um prefixo `__Host-`. Isso ajuda a defender seu domínio contra tentativas de falsificação de solicitações entre sites ()CSRF. Para obter mais informações, consulte a página Set-Cookie em Mozilla Developer Network.

A seção Interfaces de usuário da aplicação da guia Aplicações fornece várias opções de visualização, conforme o status do cluster e das aplicações instaladas no cluster.

- Acesso fora do cluster a interfaces de usuário de aplicativos persistentes — A partir da EMR versão 5.25.0 da Amazon, links de interface de usuário de aplicativos persistentes estão disponíveis para a interface de usuário do Spark e o Spark History Service. Com a Amazon EMR versão 5.30.1 e posterior, a interface de usuário do Tez e o servidor de YARN linha do tempo também têm interfaces de usuário de aplicativos persistentes. O servidor da YARN linha do tempo e a interface do usuário do Tez são aplicativos de código aberto que fornecem métricas para clusters ativos e encerrados. A interface de usuário do Spark fornece detalhes sobre estágios e tarefas do agendador, RDD tamanhos e uso de memória, informações ambientais e informações sobre os executores em execução. UIsOs aplicativos persistentes são executados fora do cluster, portanto, as informações e os registros do cluster ficam disponíveis por 30 dias após o término do aplicativo. Diferentemente das interfaces de usuário de aplicativos em cluster, os aplicativos persistentes UIs não exigem que você configure um proxy web por meio de uma SSH conexão.

- Interfaces de usuário de aplicações no cluster: há uma variedade de interfaces de usuário de histórico de aplicações que podem ser executadas em um cluster. As interfaces de usuário no cluster são hospedadas no nó principal e exigem que você configure uma SSH conexão com o servidor web. As interfaces do usuário de aplicativos no cluster mantêm o histórico de aplicativos por uma semana após o encerramento do aplicativo. Para obter mais informações e instruções sobre como configurar um SSH túnel, consulteVisualize interfaces web hospedadas em EMR clusters da Amazon.

  Com exceção do Spark History Server, do servidor de YARN linha do tempo e dos aplicativos Hive, o histórico de aplicativos no cluster só pode ser visualizado enquanto o cluster está em execução.

Visualizar interfaces do usuário de aplicações persistentes

A partir da EMR versão 5.25.0 da Amazon, você pode se conectar aos detalhes persistentes do aplicativo Spark History Server hospedado fora do cluster usando a página de resumo do cluster ou a guia Interfaces de usuário do aplicativo no console. As interfaces de aplicativos persistentes do Tez UI e do YARN Timeline Server estão disponíveis a partir da EMR versão 5.30.1 da Amazon. O acesso com um clique, por meio de um link, ao histórico de aplicativos persistente fornece os seguintes benefícios:

- Você pode analisar e solucionar problemas de trabalhos ativos e histórico de trabalhos rapidamente sem configurar um proxy da web por meio de uma SSH conexão.
- Você pode acessar o histórico de aplicativos e os arquivos de log relevantes para clusters ativos e encerrados. Os logs ficam disponíveis por 30 dias após o aplicativo ser encerrado.

Navegue até os detalhes do seu cluster no console e selecione a guia Aplicações. Selecione a interface do usuário da aplicação que você deseja após a inicialização do cluster. A interface do usuário da aplicação abre em uma nova guia do navegador. Para obter mais informações, consulte Monitoring and instrumentation.

Você pode visualizar os registros do YARN contêiner por meio dos links no servidor de histórico do Spark, no servidor da YARN linha do tempo e na interface do usuário do Tez.

> **Note**
>
> Para acessar os registros de YARN contêineres do servidor de histórico do Spark, do servidor da YARN linha do tempo e da interface do usuário do Tez, você deve habilitar

o registro no Amazon S3 para seu cluster. Se você não ativar o registro, os links para os registros de YARN contêineres não funcionarão.

Coleta de logs

Para permitir o acesso com um clique às interfaces de usuário de aplicativos persistentes, a Amazon EMR coleta dois tipos de registros:

- Os registros de eventos do aplicativo são coletados em um bucket EMR do sistema. Os registros de eventos são criptografados em repouso usando criptografia do lado do servidor com chaves gerenciadas do Amazon S3 (-S3). SSE Se você usar uma sub-rede privada para o cluster, inclua `“arn:aws:s3:::prod.MyRegion.appinfo.src/*”` na lista de recursos da política do Amazon S3 da sub-rede privada. Para obter mais informações, consulte Minimum Amazon S3 policy for private subnet.
- YARNos registros de contêineres são coletados em um bucket do Amazon S3 que você possui. Você deve ativar o registro em log para que seu cluster acesse os registros do YARN contêiner. Para obter mais informações, consulte Configurar registro em log e depuração do cluster.

Se você precisar desabilitar esse recurso por motivos de privacidade, será possível interromper o daemon usando um script de bootstrap ao criar um cluster, como demonstra o exemplo a seguir.

```
aws emr create-cluster --name "Stop Application UI Support" --release-label emr-7.2.0 \
--applications Name=Hadoop Name=Spark --ec2-attributes KeyName=<myEMRKeyPairName> \
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,InstanceType=m3.xlarge
 InstanceGroupType=CORE,InstanceCount=1,InstanceType=m3.xlarge
 InstanceGroupType=TASK,InstanceCount=1,InstanceType=m3.xlarge \
--use-default-roles --bootstrap-actions Path=s3://region.elasticmapreduce/bootstrap-
actions/run-if,Args=["instance.isMaster=true","echo Stop Application UI | sudo tee /
etc/apppusher/run-apppusher; sudo systemctl stop apppusher || exit 0"]
```

Depois de executar esse script de bootstrap, a Amazon não EMR coletará nenhum registro de eventos do Spark History Server ou YARN do servidor de cronograma no bucket do EMR sistema. Nenhuma informação do histórico do aplicativo estará disponível na guia Interfaces do usuário do aplicativo e você perderá acesso a todas as interfaces do usuário do aplicativo do console.

Arquivos grandes de log de eventos do Spark

Em alguns casos, trabalhos de longa duração do Spark, como streaming do Spark, e trabalhos grandes, como SQL consultas do Spark, podem gerar grandes registros de eventos. Com grandes registros de eventos, você pode rapidamente usar espaço em disco nas instâncias de computação e encontrar `OutOfMemory` erros ao carregar o PersistentUIs. Para evitar esses problemas, recomenda-se ativar o atributo de rolagem e compactação do log de eventos do Spark. Esse recurso está disponível nas EMR versões emr-6.1.0 e posteriores da Amazon. Para obter mais detalhes sobre rolagem e compactação, consulte Applying compaction on rolling event log files na documentação do Spark.

Para ativar o atributo de rolagem e compactação do log de eventos do Spark, ative as configurações do Spark a seguir.

- `spark.eventLog.rolling.enabled`: ativa a rolagem do log de eventos com base no tamanho. Essa configuração é desativada por padrão.
- `spark.eventLog.rolling.maxFileSize`: quando a rolagem é ativada, especifica o tamanho máximo do arquivo de log de eventos antes da rolagem. O padrão é 128 MB.
- `spark.history.fs.eventLog.rolling.maxFilesToRetain`: especifica o número máximo de arquivos de log de eventos não compactados a serem retidos. Por padrão, todos os arquivos de log de eventos são mantidos. Defina com um número menor para compactar logs de eventos mais antigos. O valor mais baixo é 1.

A compactação tenta excluir eventos com arquivos de log de eventos desatualizados, como os apresentados a seguir. Se ele descartar eventos, eles não serão mais exibidos na interface do Spark History Server.

- Eventos para trabalhos concluídos e eventos relacionados de preparação ou de tarefa.
- Eventos para executores terminados.
- Eventos para SQL consultas concluídas e eventos relacionados a cargos, estágios e tarefas.

Para iniciar um cluster com rolagem e compactação habilitadas

1. Crie um arquivo `spark-configuration.json` com a configuração a seguir.

```
[
   {
```

```
    "Classification": "spark-defaults",
       "Properties": {
          "spark.eventLog.rolling.enabled": true,
          "spark.history.fs.eventLog.rolling.maxFilesToRetain": 1
       }
   }
]
```

2. Crie o cluster com a configuração de compactação contínua do Spark da forma exibida a seguir.

```
aws emr create-cluster \
--release-label emr-6.6.0 \
--instance-type m4.large \
--instance-count 2 \
--use-default-roles \
--configurations file://spark-configuration.json
```

Considerações e limitações

No momento, o acesso com um clique às interfaces do usuário de aplicações persistentes tem as limitações a seguir.

- Haverá um atraso de pelo menos dois minutos quando os detalhes da aplicação forem exibidos na interface do Spark History Server.
- Esse recurso funciona somente quando o diretório de registro de eventos do aplicativo está ativoHDFS. Por padrão, a Amazon EMR armazena registros de eventos em um diretório deHDFS. Se você alterar o diretório padrão para um sistema de arquivos diferente, como, por exemplo o Amazon S3, esse atributo não funcionará.
- No momento, esse recurso não está disponível para EMR clusters com vários nós principais ou para EMR clusters integrados com AWS Lake Formation o.
- Para permitir o acesso com um clique às interfaces de usuário persistentes do aplicativo, você deve ter permissão para a `DescribeCluster` ação da AmazonEMR. Se você negar a permissão de um IAM diretor para essa ação, levará aproximadamente cinco minutos para que a alteração da permissão se propague.
- Se você reconfigurar os aplicativos em um cluster em execução, o histórico do aplicativo não estará disponível na interface do usuário do aplicativo.
- Para cada um Conta da AWS, o limite padrão para o aplicativo ativo UIs é 200.

- A seguir Regiões da AWS, você pode acessar o aplicativo a UIs partir do console com o Amazon EMR 6.14.0 e superior:
  - Ásia-Pacífico (Jacarta) (ap-southeast-3)
  - Europa (Espanha) (eu-south-2)
  - Ásia-Pacífico (Melbourne) (ap-southeast-4)
  - Israel (Tel Aviv) (il-central-1)
  - Oriente Médio (UAE) (me-central-1)
- A seguir Regiões da AWS, você pode acessar o aplicativo a UIs partir do console com o Amazon EMR 5.25.0 e superior:
  - Leste dos EUA (Norte da Virgínia) (us-east-1)
  - Oeste dos EUA (Oregon) (us-west-2)
  - Ásia-Pacífico (Mumbai) (ap-south-1)
  - Ásia-Pacífico (Seul) (ap-northeast-2)
  - Ásia-Pacífico (Singapura) (ap-southeast-1)
  - Ásia-Pacífico (Sydney) (ap-southeast-2)
  - Ásia Pacific (Tóquio) (ap-northeast-1)
  - Canadá (Central) (ca-central-1)
  - América do Sul (São Paulo) (sa-east-1)
  - Europa (Frankfurt) (eu-central-1)
  - Europa (Irlanda) (eu-west-1)
  - Europa (Londres) (eu-west-2)
  - Europa (Paris) (eu-west-3)
  - UE (Estocolmo) (eu-north-1)
  - China (Pequim) (cn-north-1)
  - China (Ningxia) (cn-northwest-1)

Visualizar um histórico de aplicações de alto nível

> **Note**
>
> Recomenda-se usar a interface da aplicação persistente para melhorar a experiência do usuário e reter o histórico da aplicação por até 30 dias. O histórico de aplicativos de alto

nível descrito nesta página não está disponível no novo EMR console da Amazon (https://console.aws.amazon.com/emr). Para obter mais informações, consulte Visualizar interfaces do usuário de aplicações persistentes.

Com as EMR versões 5.8.0 a 5.36.0 e 6.x até 6.8.0 da Amazon, você pode visualizar um histórico de aplicativos de alto nível na guia Interfaces de usuário do aplicativo no antigo console da Amazon. EMR A interface de usuário do EMR aplicativo Amazon mantém o resumo do histórico do aplicativo por 7 dias após a conclusão do aplicativo.

Considerações e limitações

Considere as seguintes limitações ao usar a guia Interfaces de usuário do aplicativo no antigo EMR console da Amazon.

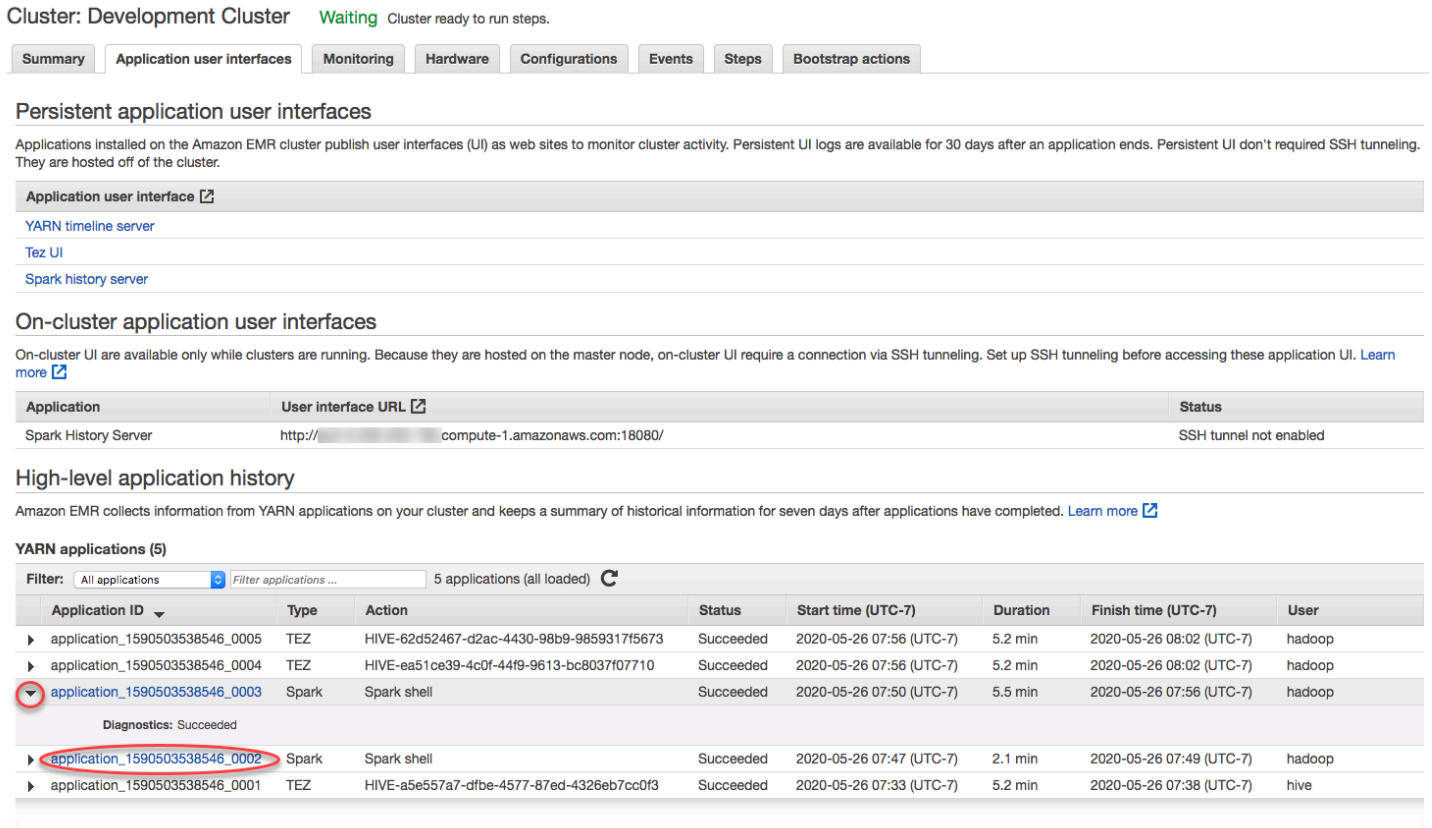
- Você só pode acessar o recurso de histórico de aplicativos de alto nível ao usar as EMR versões 5.8.0 a 5.36.0 e 6.x da Amazon até 6.8.0. A partir de 23 de janeiro de 2023, a Amazon EMR descontinuará o histórico de aplicativos de alto nível para todas as versões. Se você usa a Amazon EMR versão 5.25.0 ou superior, recomendamos que você use a interface de usuário do aplicativo persistente em vez disso.
- O atributo de histórico de aplicações de alto nível não é compatível com aplicações Spark Streaming.
- Atualmente, o acesso com um clique às interfaces de usuário de aplicativos persistentes não está disponível para EMR clusters da Amazon com vários nós principais ou para EMR clusters da Amazon integrados com AWS Lake Formation.

Exemplo: visualizar um histórico de aplicações de alto nível

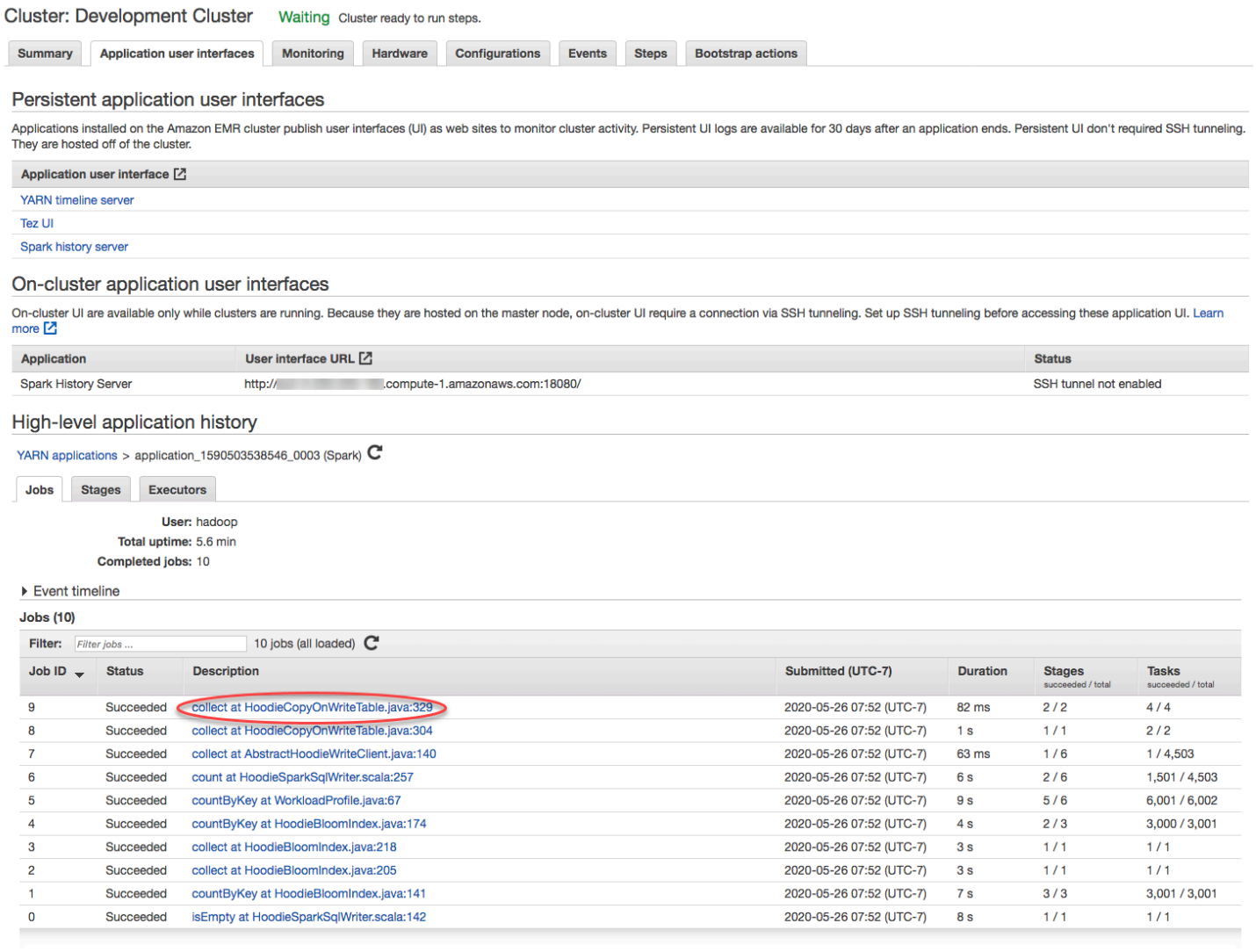
A sequência a seguir demonstra um detalhamento por meio de um Spark ou YARN aplicativo nos detalhes do trabalho usando a guia Interfaces de usuário do aplicativo na página de detalhes do cluster do console antigo.

Para visualizar detalhes do cluster, selecione o Nome de um cluster na lista Clusters. Para visualizar informações sobre registros de YARN contêineres, você deve habilitar o registro em seu cluster. Para obter mais informações, consulte Configurar registro em log e depuração do cluster. Para o histórico da aplicação Spark, as informações fornecidas na tabela de resumo são apenas um subconjunto das informações disponíveis pela IU do servidor de histórico do Spark.

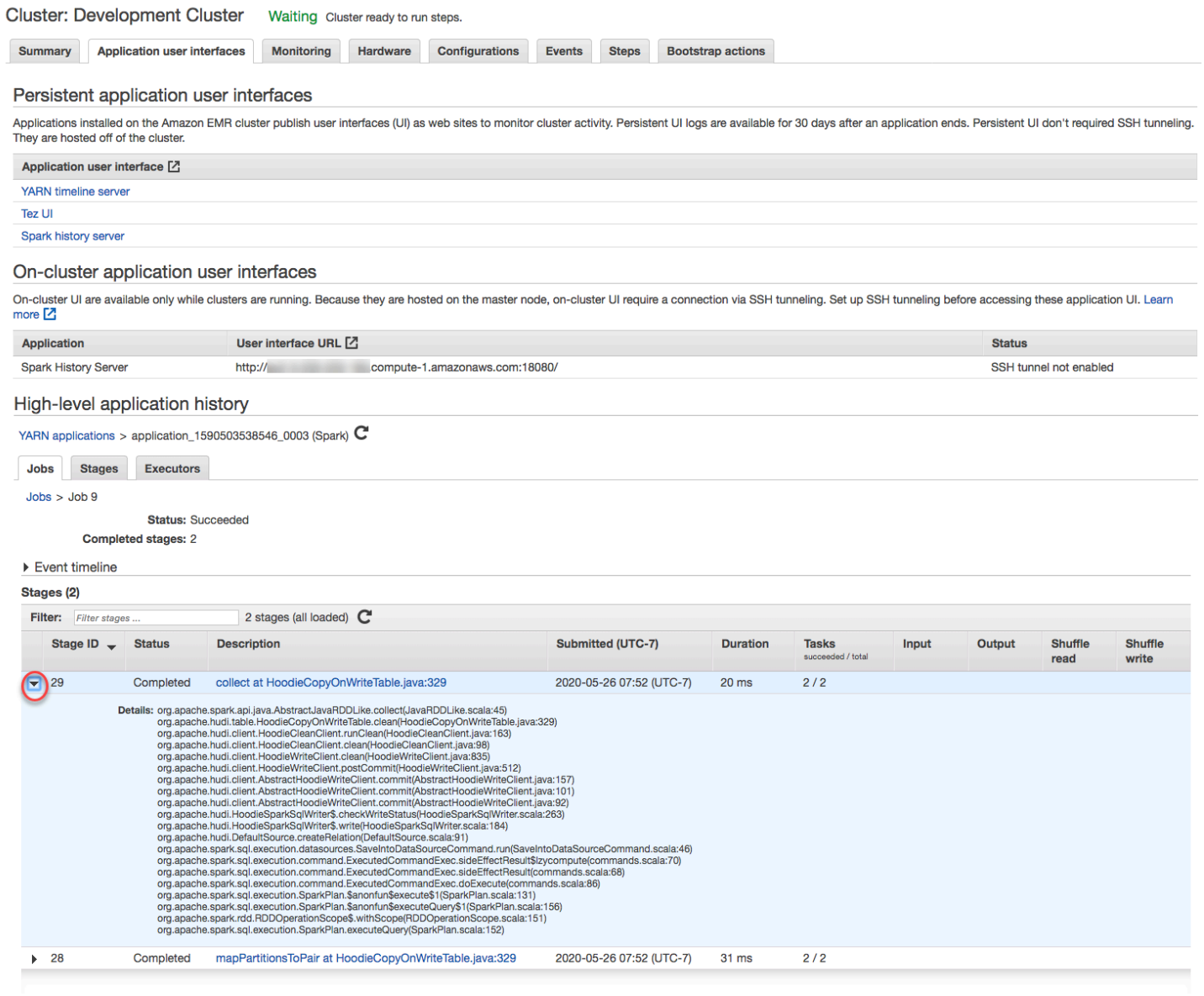
Na guia Interfaces de usuário da aplicação, em Histórico da aplicação de alto nível, você pode expandir uma linha para exibir o resumo do diagnóstico de uma aplicação Spark ou selecionar um link de ID da aplicação para visualizar detalhes sobre outra aplicação.



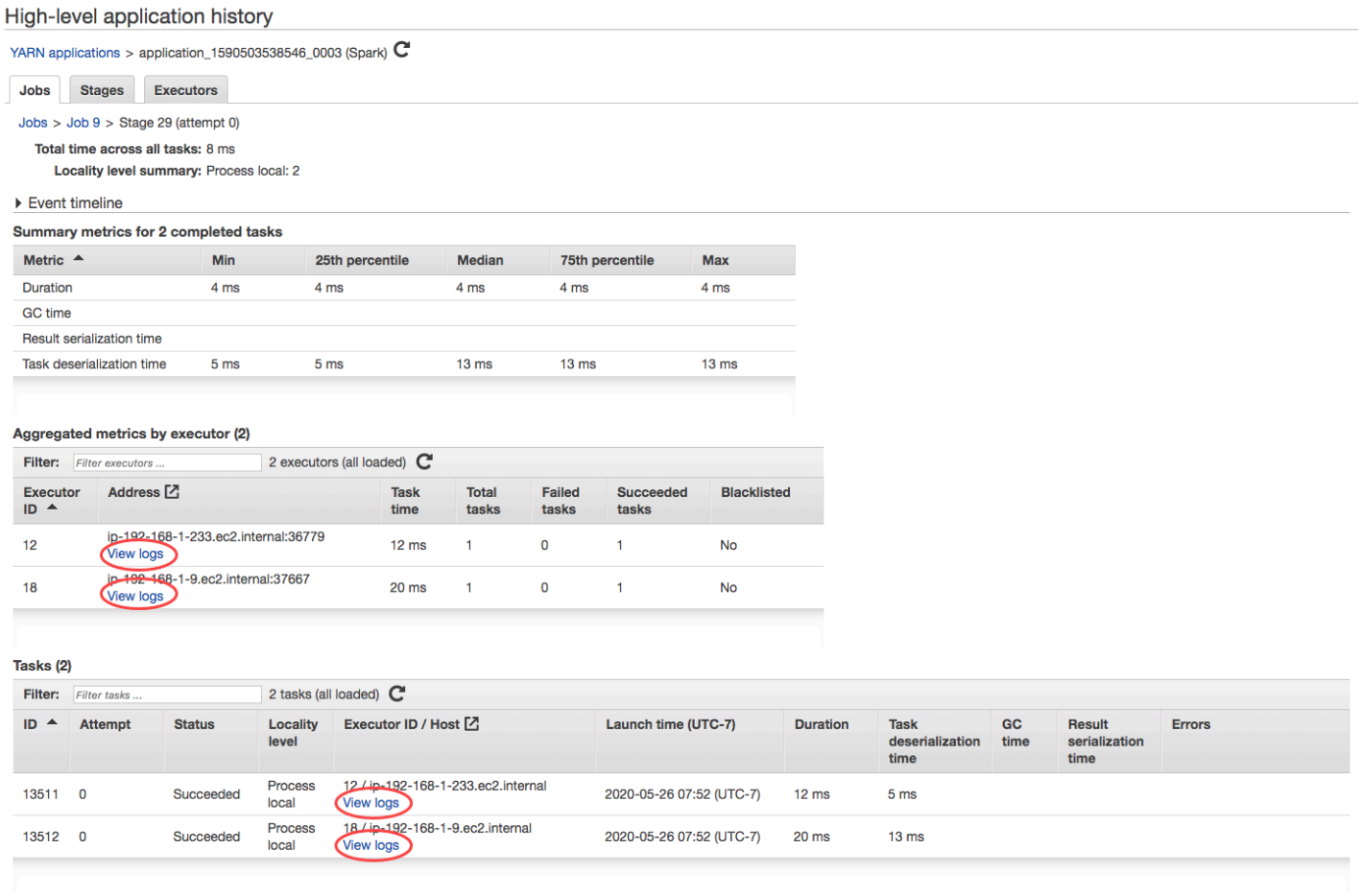
Quando você seleciona um link de ID do aplicativo, a interface do YARN usuário muda para mostrar os detalhes do aplicativo. Na guia Trabalhos dos detalhes do YARNaplicativo, você pode escolher o link Descrição de um trabalho para exibir os detalhes desse trabalho.



Na página de detalhes do trabalho, você pode expandir as informações sobre a preparação do trabalho individual e selecionar o link Descrição para ver os detalhes da preparação.



Na página de detalhes da preparação, você pode visualizar as principais métricas para tarefas e executores do preparação. Também é possível visualizar os logs de tarefa e do executor usando os links Visualizar logs.



Exibir arquivos de log do

Tanto a Amazon EMR quanto o Hadoop produzem arquivos de log que relatam o status no cluster. Por padrão, esses são gravados no nó primário, no diretório `/mnt/var/log/`. Dependendo de como você configurou seu cluster quando o executou, esses logs também podem ser arquivados no Amazon S3 e podem ser visualizados na ferramenta de depuração gráfica.

Há muitos tipos de logs gravados no nó primário. A Amazon EMR grava registros de etapas, ações de bootstrap e estados da instância. O Apache Hadoop grava logs para informar o processamento de trabalhos, tarefas e tentativas de tarefas. O Hadoop também registra logs de seus daemons. Para obter mais informações sobre os registros escritos pelo Hadoop, acesse http://hadoop.apache.org/docs/stable/ hadoop-project-dist /hadoop-common/ ClusterSetup .html.

Visualizar arquivos de log no nó primário

A tabela a seguir lista alguns dos arquivos de log que você encontrará no nó primário.

Local	Descrição
/emr/instance-controller/log/bootstrap-actions	Logs gravados durante o processamento das ações de bootstrap.
/mnt/var/log/ hadoop-state-pusher	Logs gravados pelo processo de agente de envio de estado do Hadoop.
/emr/instance-controller/log	Logs do controlador de instâncias.
/emr/instance-state	Logs de estado de instância. Eles contêm informações sobre o CPU estado da memória e os segmentos do coletor de lixo do nó.
/emr/service-nanny	Logs gravados pelo processo nanny de serviço.
/mnt/var/log/*application*	Logs específicos de um aplicativo, como o Hadoop, o Spark ou o Hive.
/mnt/var/log/hadoop/etapas/*N*	Logs de etapa que contêm informações sobre o processamento da etapa. O valor de *N* indica o stepId atribuído pela AmazonEMR . Por exemplo, um cluster tem duas etapas: `s-1234ABCDEFGH` e `s-5678IJKLMNOP` . A primeiro etapa está localizado em `/mnt/var/log/hadoop/steps/s-1234ABCDEFGH/` e segundo etapa, em `/mnt/var/log/hadoop/steps/s-5678IJKLMNOP/` . Os registros de etapas escritos pela Amazon EMR são os seguintes. • controller: informações sobre o processamento da etapa. Se a etapa falhar durante o carregamento, você encontrará o rastreamento da pilha nesse log.

Local	Descrição
	• syslog: descreve a execução dos trabalhos do Hadoop na etapa. • stderr: o canal de erro padrão do Hadoop enquanto ele processa a etapa. • stdout: o canal de saída padrão do Hadoop enquanto ele processa a etapa.

Para visualizar arquivos de log no nó primário usando a AWS CLI.

1. Use SSH para se conectar ao nó primário conforme descrito emConecte-se ao nó primário usando SSH.
2. Navegue até o diretório que contém as informações do arquivo de log que você deseja visualizar. A tabela anterior fornece uma lista dos tipos de arquivos de log que estão disponíveis e onde você os encontrará. O exemplo a seguir mostra o comando para navegar até o log de etapas com um ID `s-1234ABCDEFGH`.

   ```
   cd /mnt/var/log/hadoop/steps/s-1234ABCDEFGH/
   ```

3. Use um visualizador de arquivos de sua preferência para visualizar o arquivo de log. O exemplo a seguir usa o comando `less` do Linux para visualizar o arquivo de log `controller`.

   ```
   less controller
   ```

Visualizar arquivos de log arquivados no Amazon S3

Por padrão, os EMR clusters da Amazon lançados usando o console arquivam automaticamente os arquivos de log no Amazon S3. Você pode especificar seu próprio caminho de log ou pode permitir que o console gere automaticamente um caminho de log para você. Para clusters lançados usando o CLI ouAPI, você deve configurar o arquivamento de logs do Amazon S3 manualmente.

Quando a Amazon EMR está configurada para arquivar arquivos de log no Amazon S3, ela armazena os arquivos no local do S3 que você especificou, no/*cluster-id*/pasta, onde *cluster-id* é o ID do cluster.

A tabela a seguir lista alguns dos arquivos de log que você encontrará no Amazon S3.

Local	Descrição
/*cluster-id* /nó/	Logs de nós, incluindo logs de ações de bootstrap, estado da instância e aplicativo para o nó. Os registros de cada nó são armazenados em uma pasta rotulada com o identificador da EC2 instância desse nó.
/*cluster-id* /nó/*instance-id* /*application*	Os logs criados por cada aplicativo ou daemon associado a um aplicativo. Por exemplo, o log do servidor Hive está localizado em *cluster-id* /node/*instance-id* /hive/hive-server.log .
/*cluster-id* /etapas/*step-id*/	Logs de etapa que contêm informações sobre o processamento da etapa. O valor de *step-id* indica o ID da etapa atribuído pela AmazonEMR. Por exemplo, um cluster tem duas etapas: s-1234ABCDEFGH e s-5678IJKLMNOP . A primeiro etapa está localizado em /mnt/var/log/hadoop/steps/s-1234ABCDEFGH/ e segundo etapa, em /mnt/var/log/hadoop/steps/s-5678IJKLMNOP/ . Os registros de etapas escritos pela Amazon EMR são os seguintes. • controller: informações sobre o processamento da etapa. Se a etapa falhar durante o carregamento, você encontrará o rastreamento da pilha nesse log. • syslog: descreve a execução dos trabalhos do Hadoop na etapa. • stderr: o canal de erro padrão do Hadoop enquanto ele processa a etapa.

Local	Descrição
	• stdout: o canal de saída padrão do Hadoop enquanto ele processa a etapa.
/*cluster-id* containers/	Logs de contêiner de aplicativo. Os registros de cada YARN aplicativo são armazenados nesses locais.
/*cluster-id* /hadoop-mapreduce/	Os registros que contêm informações sobre detalhes de configuração e histórico de MapReduce trabalhos.

Visualizar os arquivos de log arquivados no Amazon S3 usando o console do Amazon S3

1. Faça login no AWS Management Console e abra o console do Amazon S3 em. https://console.aws.amazon.com/s3/
2. Abra o bucket do S3 especificado quando você configurou o cluster para arquivar arquivos de log no Amazon S3.
3. Navegue até o arquivo de log que contém as informações a serem exibidas. A tabela anterior fornece uma lista dos tipos de arquivos de log que estão disponíveis e onde você os encontrará.
4. Baixe o objeto do arquivo de log para visualizá-lo. Para obter instruções, consulte Fazer download de um objeto.

Veja instâncias de cluster na Amazon EC2

Para ajudá-lo a gerenciar seus recursos, a Amazon EC2 permite que você atribua metadados aos recursos na forma de tags. Cada EC2 tag da Amazon consiste em uma chave e um valor. As tags permitem que você categorize seus EC2 recursos da Amazon de maneiras diferentes: por exemplo, por finalidade, proprietário ou ambiente.

Você pode pesquisar e filtrar recursos com base nessas tags. As tags que você atribui aos recursos por meio de sua AWS conta estão disponíveis somente para você. Outras contas que compartilham o recurso não podem visualizar suas etiquetas.

A Amazon marca EMR automaticamente cada EC2 instância que executa com pares de valores-chave. As chaves identificam o cluster e o grupo de instâncias ao qual a instância pertence.

Isso facilita a filtragem de suas EC2 instâncias para mostrar, por exemplo, somente aquelas que pertencem a um cluster específico ou para mostrar todas as instâncias atualmente em execução no grupo de instâncias da tarefa. Isso é especialmente útil se você executa vários clusters simultaneamente ou gerencia um grande número de EC2 instâncias.

Esses são os pares de valores-chave predefinidos que a Amazon atribui: EMR

Chave	Valor	Definição de valor
aws:elasticmapreduce:job-flow-id	*job-flow-identifier*	O ID do cluster para o qual a instância está provisionada. Ele é exibido no formato `j-XXXXXXXXXXXXX` e pode conter até 256 caracteres.
aws:elasticmapreduce:instance-group-role	*group-role*	O tipo de grupo de instâncias, inserido como um destes valores: `master`, `core` ou `task`.

Você pode visualizar e filtrar as tags que a Amazon EMR adiciona. Para obter mais informações, consulte Usando tags no Guia do EC2 usuário da Amazon. Como as tags definidas pela Amazon EMR são tags do sistema e não podem ser editadas ou excluídas, as seções sobre exibição e filtragem de tags são as mais relevantes.

> **Note**
>
> A Amazon EMR adiciona tags à EC2 instância quando seu status é atualizado para Running. Se a latência ocorrer entre o momento em que a EC2 instância é provisionada e o momento em que seu status é definido como Running, as tags EMR definidas pela Amazon aparecerão quando a instância for iniciada. Se você não vir as tags, aguarde alguns minutos e atualize a exibição.

CloudWatch eventos e métricas

Use eventos e métricas para monitorar a atividade e a integridade de um EMR cluster da Amazon. Eventos são úteis para monitorar uma ocorrência específica em um cluster. Por exemplo, quando um cluster muda do estado iniciando para em execução. As métricas são úteis para monitorar um valor

específico, por exemplo, a porcentagem de espaço em disco disponível que HDFS está sendo usado em um cluster.

Para obter mais informações sobre CloudWatch eventos, consulte o Guia do usuário do Amazon CloudWatch Events. Para obter mais informações sobre CloudWatch métricas, consulte Uso de CloudWatch métricas da Amazon e Criação de CloudWatch alarmes da Amazon no Guia do CloudWatch usuário da Amazon.

Tópicos

- Monitorando EMR métricas da Amazon com CloudWatch
- Monitorando EMR eventos da Amazon com CloudWatch
- Respondendo a eventos CloudWatch

Monitorando EMR métricas da Amazon com CloudWatch

As métricas são atualizadas a cada cinco minutos e coletadas e enviadas automaticamente CloudWatch para cada EMR cluster da Amazon. Esse intervalo não é configurável. Não há cobrança pelas EMR métricas da Amazon relatadas em CloudWatch. Essas métricas de ponto de dados de cinco minutos são arquivadas por 63 dias, e os dados são descartados após esse período.

Como faço para usar as EMR métricas da Amazon?

A tabela a seguir mostra os usos comuns das métricas relatadas pela AmazonEMR. Essas são sugestões para você começar, e não uma lista abrangente. Para obter uma lista completa das métricas relatadas pela AmazonEMR, consulteMétricas relatadas pela Amazon EMR em CloudWatch.

Como eu faço para...	Métricas relevantes
Controlar o progresso do meu cluster	Examine as métricas `RunningMapTasks` , `RemainingMapTasks` , `RunningReduceTasks` e `RemainingReduceTasks` .
Detectar clusters ociosos	A métrica `IsIdle` controla se um cluster está ativo, mas não executando tarefas no momento. Você pode definir um alarme a ser disparado quando o cluster permanecer

Como eu faço para...	Métricas relevantes
	ocioso por um determinado tempo, como trinta minutos.
Detectar quando um nó fica sem armazenamento	A `MRUnhealthyNodes` métrica rastreia quando um ou mais nós principais ou de tarefas ficam sem armazenamento em disco local e fazem a transição para um `UNHEALTHY` YARN estado. Por exemplo, os nós centrais ou de tarefa estão com pouco espaço em disco e não poderão executar tarefas.
Detectar quando um cluster fica sem armazenamento	A `HDFSUtilization` métrica monitora a HDFS capacidade combinada do cluster e pode exigir o redimensionamento do cluster para adicionar mais nós principais. Por exemplo, a HDFS utilização é alta, o que pode afetar as tarefas e a integridade do cluster.
Detectar quando um cluster está em execução com capacidade reduzida	A métrica `MRLostNodes` rastreia quando um ou mais nós centrais ou de tarefa não conseguem se comunicar com o nó principal . Por exemplo, o nó principal não consegue acessar o nó central ou de tarefa.

Para obter mais informações, consulte O cluster termina com NO_ _ LEFT e nós principais SLAVE _BY_ FAILED MASTER e AWSSupport-A nalyzeEMRLogs.

CloudWatch Métricas de acesso para a Amazon EMR

Você pode visualizar as métricas que a Amazon EMR reporta CloudWatch usando o EMR console da Amazon ou o CloudWatch console. Você também pode recuperar métricas usando o CloudWatch CLI comando `mon-get-stats` ou o. CloudWatch `GetMetricStatistics` API Para obter mais informações sobre como visualizar ou recuperar métricas para EMR uso da Amazon CloudWatch, consulte o Guia do CloudWatch usuário da Amazon.

Console

Para visualizar métricas com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha o cluster do qual você deseja visualizar as métricas. Isso abrirá a página de detalhes do cluster.
3. Selecione a guia Monitoramento da página de detalhes do cluster. Escolha qualquer uma das opções Status do cluster, Status do nó ou Entradas e saídas para carregar os relatórios sobre o progresso e a integridade do cluster.
4. Após escolher uma métrica para visualizar, você poderá aumentar cada grafo. Para filtrar o período de tempo do grafo, selecione uma opção pré-preenchida ou escolha Personalizado.

Métricas relatadas pela Amazon EMR em CloudWatch

As tabelas a seguir listam as métricas que a Amazon EMR reporta no console e para CloudWatch as quais envia.

EMRMétricas da Amazon

EMRA Amazon envia dados de várias métricas para CloudWatch. Todos os EMR clusters da Amazon enviam métricas automaticamente em intervalos de cinco minutos. As métricas são arquivadas por duas semanas. Depois desse período, os dados serão descartados.

O namespace `AWS/ElasticMapReduce` inclui as métricas a seguir.

> **Note**
>
> A Amazon EMR extrai métricas de um cluster. Se um cluster torna-se inacessível, nenhuma métrica é relatada até que o cluster fique disponível novamente.

As métricas a seguir estão disponíveis para clusters que executam o Hadoop versões 2.x.

Métrica	Descrição
Status do cluster	
IsIdle	Indica que um cluster não está mais executando nenhum trabalho, mas ainda está ativo e acumulando cobranças. É definido como 1 se nenhuma tarefa ou nenhum trabalho estiver em execução, caso contrário, é definido como 0. Esse valor é verificado em intervalos de 5 minutos, sendo que um valor de 1 indica somente que o cluster estava ocioso no momento da verificação, e não que ele ficou ocioso durante todo o período de 5 minutos. Para evitar falsos positivos, você deve gerar um alerta quando esse valor for 1 em mais de uma verificação consecutiva de 5 minutos. Por exemplo, você pode gerar um alerta para esse valor se ele for 1 por 30 minutos ou mais. Caso de uso: monitorar a performance do cluster Unidade: booliano
ContainerAllocated	O número de contêineres de recursos alocados peloResourceManager. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
ContainerReserved	O número de contêineres reservados. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
ContainerPending	O número de contêineres na fila que ainda não foram alocados. Caso de uso: monitorar o progresso do cluster Unidades: Contagem

Métrica	Descrição
ContainerPendingRatio	A proporção de contêineres pendentes em relação aos contêineres alocados (ContainerPendingRatio = Container Pending / ContainerAllocated). Se ContainerAllocated = 0, então ContainerPendingRatio =ContainerPending. O valor de ContainerPendingRatio representa um número, não uma porcentagem. Esse valor é útil para escalonar recursos de cluster com base no comportamento de alocação do contêiner. Unidades: Contagem
AppsCompleted	O número de inscrições enviadas YARN que foram concluídas. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
AppsFailed	O número de inscrições enviadas YARN que não foram concluídas. Caso de uso: monitorar o progresso do cluster, monitorar a integridade do cluster Unidades: Contagem
AppsKilled	O número de inscrições enviadas YARN a ela foram eliminadas. Caso de uso: monitorar o progresso do cluster, monitorar a integridade do cluster Unidades: Contagem

Métrica	Descrição
AppsPending	O número de solicitações enviadas YARN a ela está em um estado pendente. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
AppsRunning	O número de inscrições enviadas para YARN isso estão em execução. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
AppsSubmitted	O número de inscrições enviadas paraYARN. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
Status do nó	
CoreNodesRunning	O número de nós core em funcionamento. Os pontos de dados para essa métrica são relatados somente quando existe um grupo de instâncias correspondente. Caso de uso: monitorar a integridade do cluster Unidades: Contagem
CoreNodesPending	O número de nós core aguardando atribuição. Todos os nós core solicitados podem não estar disponíveis imediatam ente; essa métrica reporta as solicitações pendentes. Os pontos de dados para essa métrica são relatados somente quando existe um grupo de instâncias correspondente. Caso de uso: monitorar a integridade do cluster Unidades: Contagem

Métrica	Descrição
LiveDataNodes	A porcentagem de nós de dados que estão recebendo trabalho do Hadoop. Caso de uso: monitorar a integridade do cluster Unidade: percentual
MRTotalNodes	O número de nós atualmente disponíveis para MapReduce trabalhos. Equivalente à YARN métrica`mapred.resourcemanager.TotalNodes` . Caso de uso: monitorar o progresso do cluster Unidades: Contagem
MRActiveNodes	O número de nós que estão executando MapReduce tarefas ou trabalhos no momento. Equivalente à YARN métrica`mapred.resourcemanager.NoOfActiveNodes` . Caso de uso: monitorar o progresso do cluster Unidades: Contagem
MRLostNodes	O número de nós alocados MapReduce que foram marcados em um LOST estado. Equivalente à YARN métrica`mapred.resourcemanager.NoOfLostNodes` . Caso de uso: monitorar a integridade do cluster, monitorar o progresso do cluster Unidades: Contagem

Métrica	Descrição
MRUnhealthyNodes	O número de nós disponíveis para MapReduce trabalhos marcados em um UNHEALTHY estado. Equivalente à YARN métrica`mapred.resourcemanager.NoOfUnhealthyNodes` . Caso de uso: monitorar o progresso do cluster Unidades: Contagem
MRDecommissionedNodes	O número de nós alocados para MapReduce aplicativos que foram marcados em um DECOMMISSIONED estado. Equivalente à YARN métrica`mapred.resourceman ager.NoOfDecommissionedNodes` . Caso de uso: monitorar a integridade do cluster, monitorar o progresso do cluster Unidades: Contagem
MRRebootedNodes	O número de nós disponíveis MapReduce que foram reinicializados e marcados em um REBOOTED estado. Equivalente à YARN métrica`mapred.resourceman ager.NoOfRebootedNodes` . Caso de uso: monitorar a integridade do cluster, monitorar o progresso do cluster Unidades: Contagem
MultiMasterInstanceGroupNodesRunning	O número de nós principais em execução. Caso de uso: monitorar falhas do nó principal e substituição Unidades: Contagem

Métrica	Descrição
MultiMasterInstanceGroupNodesRunningPercentage	A porcentagem de nós principais em execução sobre a contagem solicitada de instâncias de nós principais. Caso de uso: monitorar falhas do nó principal e substituição Unidade: percentual
MultiMasterInstanceGroupNodesRequested	O número de nós principais solicitados. Caso de uso: monitorar falhas do nó principal e substituição Unidades: Contagem
IO	
S3 BytesWritten	O número de bytes gravados no Amazon S3. Essa métrica agrega somente MapReduce trabalhos e não se aplica a outras cargas de trabalho na Amazon. EMR Caso de uso: analisar a performance do cluster, monitorar o progresso do cluster Unidades: Contagem
S3 BytesRead	O número de bytes lidos no Amazon S3. Essa métrica agrega somente MapReduce trabalhos e não se aplica a outras cargas de trabalho na Amazon. EMR Caso de uso: analisar a performance do cluster, monitorar o progresso do cluster Unidades: Contagem
HDFSUtilization	A porcentagem de HDFS armazenamento usada atualmente. Caso de uso: analisar a performance do cluster Unidade: percentual

Métrica	Descrição
HDFSBytesRead	O número de bytes lidos deHDFS. Essa métrica agrega somente MapReduce trabalhos e não se aplica a outras cargas de trabalho na Amazon. EMR Caso de uso: analisar a performance do cluster, monitorar o progresso do cluster Unidades: Contagem
HDFSBytesWritten	O número de bytes gravados emHDFS. Essa métrica agrega somente MapReduce trabalhos e não se aplica a outras cargas de trabalho na Amazon. EMR Caso de uso: analisar a performance do cluster, monitorar o progresso do cluster Unidades: Contagem
MissingBlocks	O número de blocos nos quais não HDFS tem réplicas. Esses podem ser blocos danificados. Caso de uso: monitorar a integridade do cluster Unidades: Contagem
CorruptBlocks	O número de blocos que são HDFS relatados como corrompidos. Caso de uso: monitorar a integridade do cluster Unidades: Contagem
TotalLoad	O número total de transferências simultâneas de dados. Caso de uso: monitorar a integridade do cluster Unidades: Contagem

Métrica	Descrição
MemoryTotalMB	A quantidade total de memória no cluster. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
MemoryReservedMB	A quantidade de memória reservada. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
MemoryAvailableMB	A quantidade de memória disponível para ser alocada. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
YARNMemoryAvailablePercentage	A porcentagem de memória restante disponível para YARN (YARNMemoryAvailablePercentage= MemoryAvailable MB/ MemoryTotalMB). Esse valor é útil para escalar os recursos do cluster com base no uso da YARN memória. Unidade: percentual
MemoryAllocatedMB	A quantidade de memória alocada para o cluster. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
PendingDeletionBlocks	O número de blocos marcados para exclusão. Caso de uso: monitorar o progresso do cluster, monitorar a integridade do cluster Unidades: Contagem

Métrica	Descrição
UnderReplicatedBlocks	O número de blocos que precisam ser replicados uma ou mais vezes. Caso de uso: monitorar o progresso do cluster, monitorar a integridade do cluster Unidades: Contagem
DfsPendingReplicationBlocks	O status da replicação de bloco: blocos sendo replicados, idade das solicitações de replicação e solicitações de replicação sem sucesso. Caso de uso: monitorar o progresso do cluster, monitorar a integridade do cluster Unidades: Contagem
CapacityRemainingGB	A quantidade de capacidade restante HDFS do disco. Caso de uso: monitorar o progresso do cluster, monitorar a integridade do cluster Unidades: Contagem

Veja a seguir as métricas do Hadoop 1:

Métrica	Descrição
Status do cluster	
IsIdle	Indica que um cluster não está mais executando nenhum trabalho, mas ainda está ativo e acumulando cobranças. É definido como 1 se nenhuma tarefa ou nenhum trabalho estiver em execução, caso contrário, é definido como 0. Esse valor é verificado em intervalos de 5 minutos, sendo que um valor de 1 indica somente que o cluster estava ocioso no momento da verificação, e não que ele ficou

Métrica	Descrição
	ocioso durante todo o período de 5 minutos. Para evitar falsos positivos, você deve gerar um alerta quando esse valor for 1 em mais de uma verificação consecutiva de 5 minutos. Por exemplo, você pode gerar um alerta para esse valor se ele for 1 por 30 minutos ou mais. Caso de uso: monitorar a performance do cluster Unidade: booliano
JobsRunning	O número de trabalhos no cluster que estão em execução no momento. Caso de uso: monitorar a integridade do cluster Unidades: Contagem
JobsFailed	O número de trabalhos no cluster que apresentaram falha. Caso de uso: monitorar a integridade do cluster Unidades: Contagem
Map/Reduce	
MapTasksRunning	O número de tarefas de mapeamento em execução para cada trabalho. Se você tiver um programador instalado e vários trabalhos em execução, vários gráficos são gerados. Caso de uso: monitorar o progresso do cluster Unidades: Contagem

Métrica	Descrição
MapTasksRemaining	O número de tarefas de mapeamento restantes para cada trabalho. Se você tiver um programador instalado e vários trabalhos em execução, vários gráficos são gerados. Uma tarefa de mapeamento restante não está em nenhum dos seguintes estados: Running, Killed ou Completed. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
MapSlotsOpen	A capacidade não utilizada da tarefa de mapeament o. É calculado como o número máximo de tarefas de mapeamento para um determinado cluster, menos o número total de tarefas de mapeamento em execução no momento nesse cluster. Caso de uso: analisar a performance do cluster Unidades: Contagem
RemainingMapTasksPerSlot	A razão entre o total de tarefas de mapeamento restantes e o total de slots de mapeamento disponíveis no cluster. Caso de uso: analisar a performance do cluster Unidade: razão
ReduceTasksRunning	O número de tarefas de redução em execução para cada trabalho. Se você tiver um programador instalado e vários trabalhos em execução, vários gráficos são gerados. Caso de uso: monitorar o progresso do cluster Unidades: Contagem

Métrica	Descrição
ReduceTasksRemaining	O número de tarefas de redução restantes para cada trabalho. Se você tiver um programador instalado e vários trabalhos em execução, vários gráficos são gerados. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
ReduceSlotsOpen	Capacidade não utilizada das tarefas de redução. É calculado como a capacidade máxima da tarefa de redução para um determinado cluster, menos o número total de tarefas de redução em execução no momento nesse cluster. Caso de uso: analisar a performance do cluster Unidades: Contagem
Status do nó	
CoreNodesRunning	O número de nós core em funcionamento. Os pontos de dados para essa métrica são relatados somente quando existe um grupo de instâncias correspondente. Caso de uso: monitorar a integridade do cluster Unidades: Contagem
CoreNodesPending	O número de nós core aguardando atribuição. Todos os nós core solicitados podem não estar disponíveis imediatam ente; essa métrica reporta as solicitações pendentes. Os pontos de dados para essa métrica são relatados somente quando existe um grupo de instâncias correspondente. Caso de uso: monitorar a integridade do cluster Unidades: Contagem

Métrica	Descrição
LiveDataNodes	A porcentagem de nós de dados que estão recebendo trabalho do Hadoop. Caso de uso: monitorar a integridade do cluster Unidade: percentual
TaskNodesRunning	O número de nós da tarefa trabalhando. Os pontos de dados para essa métrica são relatados somente quando existe um grupo de instâncias correspondente. Caso de uso: monitorar a integridade do cluster Unidades: Contagem
TaskNodesPending	O número de nós de tarefa aguardando atribuição. Todos os nós de tarefa solicitados podem não estar disponíveis imediatamente; essa métrica reporta as solicitações pendentes. Os pontos de dados para essa métrica são relatados somente quando existe um grupo de instâncias correspondente. Caso de uso: monitorar a integridade do cluster Unidades: Contagem
LiveTaskTrackers	O percentual dos rastreadores de tarefas que estão funcionando. Caso de uso: monitorar a integridade do cluster Unidade: percentual
IO	

Métrica	Descrição
S3 BytesWritten	O número de bytes gravados no Amazon S3. Essa métrica agrega somente MapReduce trabalhos e não se aplica a outras cargas de trabalho na Amazon. EMR Caso de uso: analisar a performance do cluster, monitorar o progresso do cluster Unidades: Contagem
S3 BytesRead	O número de bytes lidos no Amazon S3. Essa métrica agrega somente MapReduce trabalhos e não se aplica a outras cargas de trabalho na Amazon. EMR Caso de uso: analisar a performance do cluster, monitorar o progresso do cluster Unidades: Contagem
HDFSUtilization	A porcentagem de HDFS armazenamento usada atualment e. Caso de uso: analisar a performance do cluster Unidade: percentual
HDFSBytesRead	O número de bytes lidos deHDFS. Caso de uso: analisar a performance do cluster, monitorar o progresso do cluster Unidades: Contagem
HDFSBytesWritten	O número de bytes gravados emHDFS. Caso de uso: analisar a performance do cluster, monitorar o progresso do cluster Unidades: Contagem

Métrica	Descrição
MissingBlocks	O número de blocos nos quais não HDFS tem réplicas. Esses podem ser blocos danificados. Caso de uso: monitorar a integridade do cluster Unidades: Contagem
TotalLoad	O número total atual de leitores e escritores relatados por todos DataNodes em um cluster. Caso de uso: diagnostique até que ponto a alta taxa de E/S pode continuar contribuindo para o desempenho insatisfatório da execução do trabalho. Os nós de trabalho que executam o DataNode daemon também devem realizar tarefas de mapeamento e redução. TotalLoad Valores persistentemente altos ao longo do tempo podem indicar que a alta E/S pode ser um fator que contribui para o baixo desempenho. Os picos ocasionais nesse valor são típicos e geralmente não são indícios de problema. Unidades: Contagem

Métricas de capacidade de cluster

As métricas a seguir indicam as capacidades atuais ou de destino de um cluster. Essas métricas só estão disponíveis quando o ajuste de escala gerenciado ou o término automático estão habilitados.

Para clusters compostos por frotas de instâncias, as métricas de capacidade de cluster são medidas em `Units`. Para clusters compostos por grupos de instâncias, as métricas de capacidade de cluster são medidas em `Nodes` ou `VCPU` com base no tipo de unidade usado na política de escalabilidade gerenciada. Para obter mais informações, consulte Usando escalabilidade EMR gerenciada no Amazon EMR Management Guide.

Métrica	Descrição
• `TotalUnitsRequested` • `TotalNodesRequested` • `TotalVCPURequested`	O número total desejado de vCPUs unidades/nós/em um cluster, conforme determinado pelo escalonamento gerenciado. Unidades: Contagem
• `TotalUnitsRunning` • `TotalNodesRunning` • `TotalVCPURunning`	O número total atual de vCPUs unidades/nós/disponíveis em um cluster em execução. Quando um redimensi onamento de cluster for solicitado, essa métrica será atualizada depois que as novas instâncias forem adicionad as ou removidas do cluster. Unidades: Contagem
• `CoreUnitsRequested` • `CoreNodesRequested` • `CoreVCPURequested`	O número alvo de CORE vCPUs unidades/nós/em um cluster, conforme determinado pelo escalonamento gerenciado. Unidades: Contagem
• `CoreUnitsRunning` • `CoreNodesRunning` • `CoreVCPURunning`	O número atual de CORE unidades/nós/em vCPUs execução em um cluster. Unidades: Contagem
• `TaskUnitsRequested` • `TaskNodesRequested` • `TaskVCPURequested`	O número alvo de TASK vCPUs unidades/nós/em um cluster, conforme determinado pelo escalonamento gerenciado. Unidades: Contagem

Métrica	Descrição
• `TaskUnitsRunning` • `TaskNodesRunning` • `TaskVCPURunning`	O número atual de TASK unidades/nós/em vCPUs execução em um cluster. Unidades: Contagem

A Amazon EMR emite as seguintes métricas em uma granularidade de um minuto quando você ativa o encerramento automático usando uma política de encerramento automático. Algumas métricas estão disponíveis somente para EMR as versões 6.4.0 e posteriores da Amazon. Para saber mais sobre término automático, consulte Usar uma política de término automático.

Métrica	Descrição
`TotalNotebookKernels`	O total de kernels de cadernos em execução e ociosos no cluster. Essa métrica está disponível somente para EMR as versões 6.4.0 e posteriores da Amazon.
`AutoTerminationIsClusterIdle`	Indica se o cluster está em uso. O valor 0 indica que o cluster está sendo usado ativamente por um destes componentes: • Um YARN aplicativo • HDFS • Um caderno • Uma interface de usuário no cluster, como Spark History Server

Métrica	Descrição
	O valor 1 indica que o cluster está ocioso. A Amazon EMR verifica a ociosidade contínua do cluster (`AutoTerminationIsClusterIdle = 1`). Quando o tempo ocioso de um cluster é igual ao `IdleTimeout` valor em sua política de encerramento automático, a Amazon EMR encerra o cluster.

Dimensões para EMR métricas da Amazon

EMROs dados da Amazon podem ser filtrados usando qualquer uma das dimensões na tabela a seguir.

Dimensão	Descrição
JobFlowId	Igual ao ID do cluster, que é o identificador exclusivo de um cluster no formato `j-XXXXXXXXXXXX` . Encontre esse valor clicando no cluster no EMR console da Amazon.

Monitorando EMR eventos da Amazon com CloudWatch

A Amazon EMR rastreia eventos e mantém informações sobre eles por até sete dias no EMR console da Amazon. A Amazon EMR registra eventos quando há uma mudança no estado de clusters, grupos de instâncias, frotas de instâncias, políticas de escalabilidade automática ou etapas. Os eventos capturam a data e a hora em que o evento ocorreu, detalhes sobre os elementos afetados e outros pontos de dados essenciais.

A tabela a seguir lista EMR os eventos da Amazon, junto com o estado ou a mudança de estado que o evento indica, a gravidade do evento, o tipo de evento, o código do evento e as mensagens do evento. EMRA Amazon representa eventos como JSON objetos e os envia automaticamente para um stream de eventos. O JSON objeto é importante quando você configura regras para processamento de CloudWatch eventos usando Eventos porque as regras buscam corresponder aos

padrões do JSON objeto. Para obter mais informações, consulte Eventos e padrões de eventos e EMReventos da Amazon no Guia do usuário do Amazon CloudWatch Events.

> **Note**
>
> Para garantir que forneceremos as informações mais pertinentes, refinamos continuamente nossas mensagens de erro. Por isso, não é recomendável analisar o texto das mensagens para iniciar as próximas ações do fluxo de trabalho.

Eventos de início de cluster

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`CREATING`	`WARN`	EMRprovisionamento de frota de instâncias	EC2provisionamento — Capacidade de instância insuficiente	Não podemos criar seu EMR cluster `ClusterId (ClusterName)` da Amazon para a frota de instâncias. EC2 A `InstanceFleetID` Amazon tem capacidade spot insuficiente para o tipo de instância `[Instancetype1, Instancetype2]` e capacidade

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
				sob demanda insuficiente para o tipo de instância `[Instancetype3, Instancetype4]` na zona `[AvailabilityZone1, AvaliabilityZone2]` de disponibilidade. Confira aqui a documentação para obter mais informações sobre como responder a esse evento.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
CREATING	WARN	EMRaprovisionamento de grupos de instâncias	EC2provisionamento — Capacidade de instância insuficiente	Não podemos criar seu EMR cluster ClusterId (ClusterName) da Amazon para o grupo de instâncias. EC2 A InstanceGroupID Amazon tem capacidade spot insuficiente para o tipo de instância [Instancetype1, Instancetype2] e capacidade sob demanda insuficiente para o tipo de instância [Instancetype3, Instancetype4] na zona [AvailabilityZone1

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
				, AvaliabilityZone 2] de disponibilidade. Confira aqui a documentação para obter mais informações sobre como responder a esse evento.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`CREATING`	`WARN`	EMRprovisionamento de frota de instâncias	EC2provisionamento - Endereços livres insuficientes na sub-rede	Não podemos criar o EMR cluster da Amazon `ClusterId (ClusterName)` que você solicitou, por exemplo, a frota `InstanceFleetID` porque a sub-rede especificada `[Subnet1, Subnet2]` não contém endereços IP privados gratuitos suficientes para atender à sua solicitação. Use a `DescribeSubnets` operação para ver quantos endereços IP estão disponíveis (não usados) na sua sub-rede. Para obter informações

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
				sobre como responder a esse evento, consulte Códigos de erro da Amazon EC2 API

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`CREATING`	`WARN`	EMRaprovisionamento de grupos de instâncias	EC2provisionamento - Endereços livres insuficientes na sub-rede	Não podemos criar o EMR cluster da Amazon `ClusterId (ClusterName)` que você solicitou para o grupo de instâncias `InstanceGroupID` porque a sub-rede especificada `[Subnet1, Subnet2]` não contém endereços IP privados gratuitos suficientes para atender à sua solicitação. Use a `DescribeSubnets` operação para ver quantos endereços IP estão disponíveis (não usados) na sua sub-rede. Para obter

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
				informações sobre como responder a esse evento, consulte Códigos de erro da Amazon EC2 API
`CREATING`	`WARN`	EMRprovisionamento de frota de instâncias	EC2Provisionamento — v CPU Limite excedido	O fornecimento `InstanceFleetID` no EMR cluster da Amazon `ClusterId (ClusterName)` está atrasado porque você atingiu o limite do número de vCPUs (unidades de processamento virtual) atribuídas às instâncias em execução em seu`account (accountId)` . Para obter mais informações, Códigos de erro para a Amazon EC2 API

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`CREATING`	`WARN`	EMRaprovisionamento de grupos de instâncias	EC2Provisionamento — v CPU Limite excedido	O fornecimento do grupo de instâncias `InstanceGroupID` no EMR cluster da Amazon `ClusterId` está atrasado porque você atingiu o limite do número de vCPUs (unidades de processamento virtual) atribuídas às instâncias em execução na sua conta(`accountId`) . Para obter mais informações, Códigos de erro para a Amazon EC2 API

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`CREATING`	`WARN`	EMRprovisionamento de frota de instâncias	EC2Provisionamento — Limite de contagem de instâncias spot excedido	O fornecimento da frota de instâncias `InstanceFleetID` no EMR cluster da Amazon `ClusterID (ClusterName)` está atrasado porque você atingiu o limite do número de instâncias spot que você pode iniciar no seu`account (accountId)` . Para obter mais informações, consulte Códigos de erro para a Amazon EC2 API.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`CREATING`	`WARN`	EMRaprovisionamento de grupos de instâncias	EC2Provisionamento — Limite de contagem de instâncias spot excedido	O fornecimento do grupo de instâncias `InstanceGroupID` no EMR cluster da Amazon `ClusterID (ClusterName)` está atrasado porque você atingiu o limite do número de instâncias spot que você pode iniciar no seu`account (accountId)` . Para obter mais informações, consulte Códigos de erro para a Amazon EC2 API.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`CREATING`	`WARN`	EMRprovisionamento de frota de instâncias	EC2Provisionamento — limite de instâncias excedido	O fornecimento da frota de instâncias `InstanceFleetID` no EMR cluster da Amazon `ClusterId (ClusterName)` está atrasado porque você atingiu o limite do número de instâncias que você pode executar simultaneamente no seu`account (accountID)` . Para obter mais informações sobre os limites EC2 de serviço da Amazon, consulte Códigos de erro para a Amazon EC2 API.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`CREATING`	`WARN`	EMRaprovisionamento de grupos de instâncias	EC2Provisionamento — limite de instâncias excedido	O fornecimento do grupo de instâncias `InstanceGroupID` no EMR cluster da Amazon `ClusterId (ClusterName)` está atrasado porque você atingiu o limite do número de instâncias que você pode executar simultaneamente no seu`account (accountID)` . Para obter mais informações sobre os limites EC2 de serviço da Amazon, consulte Códigos de erro para a Amazon EC2 API.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`CREATING`	`WARN`	EMRaprovisionamento de grupos de instâncias	none	O EMR cluster da Amazon `ClusterId (ClusterName)` foi criado em `Time` e está pronto para uso. - ou - O EMR cluster da Amazon `ClusterId (ClusterName)` concluiu a execução de todas as etapas pendentes em`Time`. **Note** Um cluster no estado `WAITING` pode ainda estar processando

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
				trabalhos .
`STARTING`	`INFO`	EMRmudança de estado do cluster	none	O EMR cluster da Amazon `ClusterId (ClusterName)` foi solicitado em `Time` e está sendo criado.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`STARTING`	`INFO`	EMRmudança de estado do cluster	none	**Note** Aplica-se somente a clusters com a configuração de frotas de instâncias e várias zonas de disponibilidade selecionadas na AmazonEC2. `ClusterId (ClusterName)` O EMR cluster da Amazon está sendo criado na zona (`AvailabilityZoneID`), que foi escolhida

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
				entre as opções de zona de disponibilidade especificadas.
`STARTING`	`INFO`	EMRmudança de estado do cluster	none	O EMR cluster da Amazon `ClusterId (ClusterName)` começou a executar etapas em`Time`.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`WAITING`	`INFO`	EMRmudança de estado do cluster	none	O EMR cluster da Amazon `ClusterId (ClusterName)` foi criado em `Time` e está pronto para uso. - ou - O EMR cluster da Amazon `ClusterId (ClusterName)` concluiu a execução de todas as etapas pendentes em`Time`. **Note** Um cluster no estado `WAITING` pode ainda estar processando

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
				trabalhos .

> **Note**
>
> Os eventos com código de evento são `EC2 provisioning - Insufficient Instance Capacity` emitidos periodicamente quando seu EMR cluster encontra um erro de capacidade insuficiente da Amazon EC2 para sua frota de instâncias ou grupo de instâncias durante a criação do cluster ou a operação de redimensionamento. Para obter informações sobre como responder a esses eventos, consulte Respondendo aos eventos de capacidade insuficiente EMR de instâncias do cluster da Amazon.

Eventos de término de clusters

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`TERMINATED`	A gravidade depende do motivo da mudança de estado, conforme mostrado a seguir: • **`CRITICAL`** se o cluster terminou com	EMRmudança de estado do cluster	none	O Amazon EMR Cluster `ClusterId (ClusterName)` foi encerrado em `Time` por um motivo de`StateChangeReason: Code` .

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
	qualquer um dos seguintes motivos de mudança de estado: `INTERNAL_ERROR`, `VALIDATION_ERROR`, `INSTANCE_FAILURE`, `BOOTSTRAP_FAILURE` ou `STEP_FAILURE`. • **`INFO`** se o cluster terminou com qualquer um dos seguintes motivos de mudança de estado: `USER_REQUEST` ou `ALL_STEPS_COMPLETED`.			

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`TERMINATED_WITH_ERRORS`	`CRITICAL`	EMRmudança de estado do cluster	none	O Amazon EMR Cluster `ClusterId (ClusterName)` foi encerrado com erros por um motivo de`StateChangeReason: Code . Time`
`TERMINATED_WITH_ERRORS`	`CRITICAL`	EMRmudança de estado do cluster	none	O Amazon EMR Cluster `ClusterId (ClusterName)` foi encerrado com erros por um motivo de`StateChangeReason: Code . Time`

Eventos de alteração de estado da frota de instâncias

> **Note**
>
> A configuração de frotas de instâncias está disponível somente nas EMR versões 4.8.0 e posteriores da Amazon, excluindo 5.0.0 e 5.0.3.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
De `PROVISIONING` até `WAITING`	`INFO`		none	O provisionamento, por exemplo, da frota no EMR cluster `InstanceFleetID` da Amazon `ClusterId` (`ClusterName`) está concluído. O provisionamento começou às `Time` e levou `Num` minutos. Agora, a frota de instâncias tem capacidade sob demanda de `Num` e capacidade spot de `Num`. A capacidade sob demanda de destino era `Num`, e a capacidade spot de destino era `Num`.
De `WAITING` até `RESIZING`	`INFO`		none	Um redimensionamento, por exemplo, da

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
				frota `InstanceFleetID` no EMR cluster da Amazon `ClusterId` (`ClusterName`) começou em`Time`. A frota de instâncias está sendo redimensionada de uma capacidade sob demanda de `Num` para um destino de `Num` e de uma capacidade spot de `Num` para um destino de `Num`.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
De `RESIZING` até `WAITING`	`INFO`		none	A operação de redimensionamento, por exemplo, da frota `InstanceFleetID` no EMR cluster da Amazon `ClusterId` (`Cluster Name`) está concluída. O redimensionamento começou às `Time` e durou `Num` minutos. Agora, a frota de instâncias tem capacidade sob demanda de `Num` e capacidade spot de `Num`. A capacidade sob demanda de destino era `Num`, e a capacidade spot de destino era `Num`.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
De `RESIZING` até `WAITING`	`INFO`		none	A operação de redimensionamento, por exemplo, da frota `InstanceFleetID` no EMR cluster da Amazon `ClusterId (ClusterName)` atingiu o tempo limite e foi interrompida. O redimensionamento começou às `Time` e foi interrompido após `Num` minutos. Agora, a frota de instâncias tem capacidade sob demanda de `Num` e capacidade spot de `Num`. A capacidade sob demanda de destino era `Num`, e a capacidade spot de destino era `Num`.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
SUSPENDED	ERROR		none	A frota de instâncias InstanceFleetID no EMR cluster da Amazon ClusterId (ClusterName) foi presa Time pelo seguinte motivo:ReasonDesc .
RESIZING	WARNING		none	A operação de redimensionamento, por exemplo, da frota InstanceFleetID no EMR cluster da Amazon ClusterId (ClusterName) está paralisada pelo seguinte motivo:ReasonDesc .

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`WAITING` ou `Running`	`INFO`		none	A operação de redimensionamento, por exemplo, da frota `InstanceFleetID` no EMR cluster da Amazon não `ClusterId (ClusterName)` pôde ser concluída enquanto a Amazon EMR adicionava capacidade spot na zona `AvailabilityZone` de disponibilidade. Cancelamos sua solicitação para provisionar uma capacidade spot maior. Para ver as ações recomendadas, verifique Práticas recomendadas para flexibilidade de instâncias e de zona de

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
				disponibilidade e tente novamente .
`WAITING` ou `Running`	`INFO`		none	Uma operação de redimensionamento, por exemplo, da frota `InstanceFleetID` no EMR cluster da Amazon, `ClusterId (ClusterName)` foi iniciada por `Entity` at`Time`.

Eventos de redimensionamento da frota de instâncias

Tipo de evento	Gravidade	Código do evento	Message
EMRredimensionamento da frota de instâncias	`ERROR`	Tempo limite de provisionamento spot	A operação de redimensionamento da frota de instâncias `InstanceFleetID` no EMR cluster da Amazon não `ClusterId (ClusterName)` pôde ser concluída durante a aquisição

Tipo de evento	Gravidade	Código do evento	Message
			da capacidade spot em AZ. `AvailabilityZone` Já cancelamos a solicitação e paramos de tentar provisionar qualquer capacidade spot adicional, e a frota de instâncias provisionou a capacidade spot de `num`. A capacidade spot de destino era `num`. Para obter mais informações e ações recomendadas, consulte a página de documentação aqui e tente novamente.

Tipo de evento	Gravidade	Código do evento	Message
EMRredimensionamento da frota de instâncias	ERROR	Tempo limite de provisionamento sob demanda	A operação de redimensionamento da frota de instâncias `InstanceFleetID` no EMR cluster da Amazon não `ClusterId (ClusterName)` pôde ser concluída durante a aquisição de capacidade sob demanda em AZ. `AvailabilityZone` Já cancelamos a solicitação e paramos de tentar provisionar qualquer capacidade sob demanda adicional, e a frota de instâncias provision ou a capacidade sob demanda de num. A capacidade sob demanda de destino era num. Para obter mais informações e ações recomendadas, consulte a página de documentação aqui e tente novamente.

Tipo de evento	Gravidade	Código do evento	Message
EMRredimensionamento da frota de instâncias	`WARNING`	EC2provisionamento — Capacidade de instância insuficiente	Não podemos concluir a operação de redimensionamento da frota de instâncias `InstanceFleetID` no EMR cluster, `ClusterId (ClusterName)` pois a Amazon EC2 tem capacidade spot insuficiente para tipos de instância `[Instancetype1, Instancetype2]` e capacidade sob demanda insuficiente para tipos de instância `[Instancetype3, Instancetype4]` na zona de disponibilidade. `[AvailabilityZone1]` Até agora, a frota de instâncias provision ou a capacidade sob demanda de `num`, e a capacidade sob demanda de destino era `num`. A capacidade spot provisionada é `num`, e a capacidade spot de destino era `num`. Confira

Tipo de evento	Gravidade	Código do evento	Message
			aqui a documentação para obter mais informações sobre como responder a esse evento.

Tipo de evento	Gravidade	Código do evento	Message
EMRredimensionamento da frota de instâncias	`WARNING`	Tempo limite de provisionamento spot: redimensionamento contínuo	Ainda estamos provisionando a capacidade spot para a operação de redimensionamento da frota de instâncias que foi iniciada, `time` por exemplo, no ID da frota no EMR cluster `InstanceFleetID ClusterId (ClusterName)` da Amazon ou em AZ. `[Instancetype1, Instancetype2] AvailabilityZone` Para a operação de redimensionamento anterior iniciada em`time`, o período de tempo limite expirou, então a Amazon EMR parou de provisionar a capacidade spot após adicionar as `num` instâncias solicitadas à sua `num` frota de instâncias. Para obter mais informações e ações recomendadas, confira a página de documentação aqui.

Tipo de evento	Gravidade	Código do evento	Message
EMRredimensionamento da frota de instâncias	`WARNING`	Tempo limite de provisionamento sob demanda: redimensionamento contínuo	Ainda estamos provisionando a capacidade sob demanda para a operação de redimensionamento da frota de instâncias que foi iniciada, `time` por exemplo, no ID da frota no EMR cluster `InstanceFleetID ClusterId (ClusterName)` da Amazon ou em AZ. `[Instancetype1, Instancetype2] AvailabilityZone` Para a operação de redimensionamento anterior, iniciada em`time`, o período de tempo limite expirou, então a Amazon EMR parou de provisionar a capacidade sob demanda após adicionar as instâncias solicitadas `num` à sua frota `num` de instâncias. Para obter mais informações e ações recomendadas,

Tipo de evento	Gravidade	Código do evento	Message
			confira a página de documentação aqui.
EMRredimensionamento da frota de instâncias	`WARNING`	EC2Provisionamento - Endereço livre insuficiente na sub-rede	Não podemos concluir a operação de redimensionamento, por exemplo, da frota `InstanceFleetID` no EMR cluster da Amazon `ClusterId (ClusterName)` porque a sub-rede especificada [Subnet1, Subnet2] não contém endereços IP privados gratuitos suficientes para atender à sua solicitação. Use a `DescribeSubnets` operação para ver quantos endereços IP estão disponíveis (não usados) na sua sub-rede. Para obter informações sobre como responder a esse evento, consulte Códigos de erro da Amazon EC2 API.

Tipo de evento	Gravidade	Código do evento	Message
EMRredimensionamento da frota de instâncias	`WARNING`	EC2Provisionamento - v CPU Limite excedido	O redimensionamento da frota de instâncias `InstanceFleetID` no EMR cluster da Amazon `ClusterName` está atrasado porque você atingiu o limite do número de vCPUs (unidades de processamento virtual) atribuídas às instâncias em execução no seu `account (accountId)` . Para obter mais informações, consulte Códigos de erro para a Amazon EC2 API.

Tipo de evento	Gravidade	Código do evento	Message
EMRredimensionamento da frota de instâncias	`WARNING`	EC2Provisionamento - Limite de contagem de instâncias spot excedido	O fornecimento da frota de instâncias `InstanceFleetID` no EMR cluster da Amazon `ClusterID (ClusterName)` está atrasado porque você atingiu o limite do número de instâncias spot que você pode iniciar no seu`account (accountId)` . Para obter mais informações, consulte Códigos de erro para a Amazon EC2 API.

Tipo de evento	Gravidade	Código do evento	Message
EMRredimensionamento da frota de instâncias	`WARNING`	EC2Provisionamento — Limite de instâncias excedido	O fornecimento da frota de instâncias `InstanceFleetID` no EMR cluster da Amazon `ClusterID` `(ClusterName)` está atrasado porque você atingiu o limite do número de instâncias sob demanda que você pode executar no seu`account` `(accountId)` . Para obter mais informações sobre [códigos de erro para a Amazon EC2 API].

> **Note**
>
> Os eventos de tempo limite de provisionamento são emitidos quando a EMR Amazon interrompe o provisionamento de capacidade spot ou sob demanda para a frota após o tempo limite expirar. Para obter informações sobre como responder a esses eventos, consulte [Respondendo aos eventos de tempo limite de redimensionamento da frota de instâncias de EMR cluster da Amazon].

Eventos de instâncias de grupos

Tipo de evento	Gravidade	Código do evento	Message
De `RESIZING` até `Running`	`INFO`	none	A operação de redimensionamento

Tipo de evento	Gravidade	Código do evento	Message
			do grupo de instâncias `InstanceGroupID` no EMR cluster da Amazon `ClusterId (ClusterName)` está concluída . Agora, ele tem uma contagem de instâncias de `Num`. O redimensionamento começou às `Time` e levou `Num` minutos para ser concluído.
De `RUNNING` até `RESIZING`	`INFO`	none	Um grupo de redimensionamento, por exemplo, `InstanceGroupID` no EMR cluster da Amazon `ClusterId (ClusterName)` começou em`Time`. Ele está sendo redimensionado de uma contagem de instâncias de `Num` a `Num`.

Tipo de evento	Gravidade	Código do evento	Message
SUSPENDED	ERROR	none	O grupo de instâncias InstanceGroupID no EMR cluster da Amazon ClusterId (ClusterName) foi preso Time pelo seguinte motivo:ReasonDesc .
RESIZING	WARNING	none	A operação de redimensionamento do grupo de instâncias InstanceGroupID no EMR cluster da Amazon ClusterId (ClusterName) está travada pelo seguinte motivo:ReasonDesc .

Tipo de evento	Gravidade	Código do evento	Message
EMRredimensionamento do grupo de instâncias	`WARNING`	EC2provisionamento — Capacidade de instância insuficiente	Não conseguimos concluir a operação de redimensionamento que começou em `time` for Instance Group `InstanceGroupID` no EMR cluster, `ClusterId (ClusterName)` pois a Amazon não EC2 tem `Spot/On Demand` capacidade suficiente para o tipo de instância `[Instancetype]` na zona `[AvailabilityZone1]` de disponibilidade. Até agora, o grupo de instâncias tem uma contagem de instâncias em execução de `num`, e a contagem de instâncias solicitadas era `num`. Confira aqui a documentação para obter mais informações sobre como responder a esse evento.

Tipo de evento	Gravidade	Código do evento	Message
EMRredimensionamento do grupo de instâncias	`WARNING`	EC2Provisionamento - Endereço livre insuficiente na sub-rede	Não podemos concluir a operação de redimensionamento do grupo de instâncias `InstanceGroupID` no EMR cluster da Amazon `ClusterId(ClusterName)` porque a sub-rede especificada [Subnet1, Subnet2] não contém endereços IP privados gratuitos suficientes para atender à sua solicitação. Use a `DescribeSubnets` operação para ver quantos endereços IP estão disponíveis (não usados) na sua sub-rede. Para obter informações sobre como responder a esse evento, consulte Códigos de erro da Amazon EC2 API.

Tipo de evento	Gravidade	Código do evento	Message
EMRredimensionamento do grupo de instâncias	`WARNING`	EC2Provisionamento - v CPU Limite excedido	O redimensionamento do grupo de instâncias `InstanceGroupID` no EMR cluster da Amazon `ClusterName` está atrasado porque você atingiu o limite do número de vCPUs (unidades de processamento virtual) atribuídas às instâncias em execução no seu`account (accountId)` . Para obter mais informações, consulte Códigos de erro para a Amazon EC2 API.

Tipo de evento	Gravidade	Código do evento	Message
EMRredimensionamento do grupo de instâncias	`WARNING`	EC2Provisionamento - Limite de contagem de instâncias spot excedido	O fornecimento do grupo de instâncias `InstanceGroupID` no EMR cluster da Amazon `ClusterID (ClusterName)` está atrasado porque você atingiu o limite do número de instâncias spot que você pode iniciar no seu`account (accountId)` . Para obter mais informações, consulte Códigos de erro para a Amazon EC2 API.

Tipo de evento	Gravidade	Código do evento	Message
EMRredimensionamento do grupo de instâncias	`WARNING`	EC2Provisionamento — Limite de instâncias excedido	O fornecimento do grupo de instâncias `InstanceGroupID` no EMR cluster da Amazon `ClusterID` `(ClusterName)` está atrasado porque você atingiu o limite do número de instâncias sob demanda que você pode executar no seu`account` `(accountId)` . Para obter mais informações sobre códigos de erro para a Amazon EC2 API.
De `RUNNING` até `RESIZING`	`INFO`	none	Um redimensionamento para o grupo de instâncias `InstanceGroupID` no EMR cluster da Amazon `ClusterId` `(ClusterName)` foi iniciado por `Entity` at`Time`.

> **Note**
>
> Com a Amazon EMR versão 5.21.0 e posterior, você pode substituir as configurações do cluster e especificar classificações de configuração adicionais para cada grupo de instâncias em um cluster em execução. Você faz isso usando o EMR console da Amazon, o AWS

Command Line Interface (AWS CLI) ou AWS SDK o. Para obter mais informações, consulte Supplying a Configuration for an Instance Group in a Running Cluster.

A tabela a seguir lista EMR os eventos da Amazon para a operação de reconfiguração, junto com o estado ou a mudança de estado que o evento indica, a gravidade do evento e as mensagens do evento.

Estado ou alteração de estado	Gravidade	Message
`RUNNING`	`INFO`	Uma reconfiguração para o grupo de instâncias `InstanceGroupID` no EMR cluster da Amazon `ClusterId (ClusterName)` foi iniciada pelo usuário em`Time`. A versão da configuração solicitada é `Num`.
De `RECONFIGURING` até `Running`	`INFO`	A operação de reconfiguração do grupo de instâncias `InstanceGroupID` no EMR cluster da Amazon `ClusterId (ClusterName)` está concluída. A reconfiguração começou às `Time` e levou `Num` minutos para ser concluída. A versão de configuração atual é `Num`.
De `RUNNING` até `RECONFIGURING` em	`INFO`	Uma reconfiguração para o grupo de instâncias `InstanceGroupID` no EMR cluster da Amazon `ClusterId (ClusterName)` começou em`Time`. Ela é a configuração do número

Estado ou alteração de estado	Gravidade	Message
		da versão Num ao número da versão Num.
RESIZING	INFO	A operação de reconfiguração para a versão de configura ção do Num grupo de instância s InstanceGroupID no EMR cluster da Amazon ClusterId (ClusterN ame) está temporariamente bloqueada Time porque o grupo de instâncias está ativado. State
RECONFIGURING	INFO	A operação de redimensi onamento em relação à contagem Num de instância s do grupo de instâncias InstanceGroupID no EMR cluster da Amazon ClusterId (ClusterN ame) está temporariamente bloqueada Time porque o grupo de instâncias está dentroState.
RECONFIGURING	WARNING	A operação de reconfigu ração do grupo de instância s InstanceGroupID no EMR cluster da Amazon ClusterId (ClusterN ame) falhou Time e levou Num minutos para falhar. A versão de configuração com falha é Num.

Estado ou alteração de estado	Gravidade	Message
RECONFIGURING	INFO	As configurações estão revertendo para o número da versão anterior bem-sucedida do grupo Num de instâncias InstanceGroupID no EMR cluster ClusterId (ClusterName) da Amazon em. Time A nova versão de configuração é Num.
De RECONFIGURING até Running	INFO	As configurações foram revertidas com sucesso para a versão anterior bem-sucedida do Num grupo de instâncias InstanceGroupID no EMR cluster ClusterId (ClusterName) da Amazon em. Time A nova versão de configuração é Num.
De RECONFIGURING até SUSPENDED	CRITICAL	Falha ao reverter para a versão anterior bem-sucedida do Num grupo de instâncias InstanceGroupID no EMR cluster da Amazon ClusterId (ClusterName) emTime.

Eventos de política do Auto Scaling

Estado ou alteração de estado	Gravidade	Message
PENDING	INFO	Uma política de Auto Scaling foi adicionada ao grupo de instâncias no EMR cluster InstanceGroupID ClusterId (Cluster Name) da Amazon em. Time A política tem um anexo pendente. - ou - A política de Auto Scaling para o grupo de instância s no EMR cluster da InstanceGroupID Amazon ClusterId (Cluster Name) foi atualizada em. Time A política tem um anexo pendente.
ATTACHED	INFO	A política de Auto Scaling, por exemplo, para o grupo no EMR cluster InstanceG roupID da Amazon, ClusterId (Cluster Name) foi anexada em. Time
DETACHED	INFO	A política de Auto Scaling para o grupo de instâncias no EMR cluster InstanceGroupID da Amazon ClusterId (ClusterName) foi desanexada em. Time

Estado ou alteração de estado	Gravidade	Message
FAILED	ERROR	A política de Auto Scaling para o grupo de instâncias no EMR cluster InstanceGroupID da Amazon não ClusterId (ClusterName) pôde ser anexada e falhou em. Time - ou - A política de Auto Scaling para o grupo de instâncias no EMR cluster da InstanceGroupID Amazon não ClusterId (ClusterName) pôde ser separada e falhou em. Time

Eventos de etapa

Estado ou alteração de estado	Gravidade	Message
PENDING	INFO	A etapa StepID (StepName) foi adicionada ao EMR cluster da Amazon ClusterId (ClusterName) em Time e está pendente de execução.
CANCEL_PENDING	WARN	A etapa StepID (StepName) no EMR cluster da Amazon ClusterId (ClusterName) foi cancelada em Time

Estado ou alteração de estado	Gravidade	Message
		e está pendente de cancelamento.
RUNNING	INFO	A etapa StepID (StepName) no EMR cluster da Amazon ClusterId (ClusterName) começou a ser executada emTime.
COMPLETED	INFO	A etapa StepID (StepName) no EMR cluster da Amazon ClusterId (ClusterName) concluiu a execução emTime. A etapa começou a ser executada às Time e levou Num minutos para ser concluída.
CANCELLED	WARN	A solicitação de cancelamento foi bem-sucedida para a etapa de cluster StepID (StepName) no EMR cluster da Amazon ClusterId (ClusterName) emTime, e a etapa agora está cancelada.
FAILED	ERROR	A etapa StepID (StepName) no EMR cluster da Amazon ClusterId (ClusterName) falhou emTime.

Eventos de substituição de nós não íntegros

Tipo de evento	Gravidade	Código do evento	Message	
Substituição EMR insalubre de nós da Amazon	`INFO`	Detectado nó central não íntegro	EMRA Amazon identificou que a instância principal `[instanceID (Instance Name)] InstanceGroup/Fleet` no EMR cluster da Amazon `clusterID (ClusterName)` é`UNHEALTHY` . A Amazon EMR tentará recuperar ou substituir a instância sem problemas. `UNHEALTHY`	
Substituição EMR insalubre de nós da Amazon	`INFO`	Nó central não íntegro - substituição desativada	EMRA Amazon identificou que a instância principal `[instanceID (Instance Name)] InstanceGroup/Fleet` no EMR cluster	

Tipo de evento	Gravidade	Código do evento	Message	
			da Amazon `(clusterID)` `(ClusterName)` é`UNHEALTHY` . Ative a substituição normal de nós principais não íntegros em seu cluster para permitir que a Amazon substitua as `UNHEALTHY` instâncias EMR sem problemas, caso elas não possam ser recuperadas.	

Tipo de evento	Gravidade	Código do evento	Message
Substituição EMR insalubre de nós da Amazon	WARN	O nó central não íntegro não foi substituído	A Amazon não EMR pode substituir sua instância UNHEALTHY principal [instanceID (InstanceName)] InstanceGroup/Fleet no EMR cluster da Amazon clusterID (ClusterName) por um motivo. Note O motivo pelo qual a Amazon não EMR pode substituir seu nó principal difere dependendo do seu cenário. Por

Tipo de evento	Gravidade	Código do evento	Message	
			exemplo, uma razão pela qual a Amazon não EMR pode excluir um nó é porque um cluster não teria nenhum nó principal restante.	

Tipo de evento	Gravidade	Código do evento	Message	
Substituição EMR insalubre de nós da Amazon	INFO	Nó central não íntegro recuperado	EMRA Amazon recuperou suas instâncias UNHEALTHY principais [instanceID (InstanceName)] InstanceGroup/Fleet no EMR cluster da Amazon clusterID (ClusterName)	

Para obter mais informações sobre a substituição de nós não íntegros, consulte Substituindo nós não íntegros.

Visualização de eventos com o EMR console da Amazon

Para cada cluster, você pode visualizar uma lista simples de eventos no painel de detalhes, que lista os eventos em ordem decrescente de ocorrência. Você também pode visualizar todos os eventos para todos os clusters de uma região em ordem decrescente de ocorrência.

Se não quiser que um usuário veja todos os eventos de cluster para uma região, adicione uma instrução que negue permissão (`"Effect": "Deny"`) para a ação `elasticmapreduce:ViewEventsFromAllClustersInConsole` a uma política anexada a esse usuário.

Para visualizar eventos de todos os clusters em uma região com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EMREm EC2 Ativado no painel de navegação esquerdo, escolha Eventos.

Para visualizar eventos de um determinado cluster com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha um cluster.
3. Para visualizar todos os seus eventos, selecione a guia Eventos na página de detalhes do cluster.

Respondendo a eventos CloudWatch

Esta seção descreve várias maneiras de responder a eventos acionáveis que a Amazon EMR emite como mensagens de CloudWatch eventos.

Tópicos

- Criação de regras para EMR eventos da Amazon com CloudWatch
- Configurando alarmes nas métricas CloudWatch
- Respondendo aos eventos de capacidade insuficiente EMR de instâncias do cluster da Amazon
- Respondendo aos eventos de tempo limite de redimensionamento da frota de instâncias de EMR cluster da Amazon

Criação de regras para EMR eventos da Amazon com CloudWatch

A Amazon envia eventos EMR automaticamente para um stream de CloudWatch eventos. Você pode criar regras que correspondem eventos de acordo com um padrão especificado e rotear esses eventos para destinos a fim de realizar ações, como enviar uma notificação por e-mail. Os padrões são comparados com o JSON objeto do evento. Para obter mais informações sobre os detalhes dos EMR eventos da Amazon, consulte EMREventos da Amazon no Guia do usuário do Amazon CloudWatch Events.

Para obter informações sobre como configurar regras de CloudWatch eventos, consulte Criação de uma CloudWatch regra que é acionada em um evento.

Configurando alarmes nas métricas CloudWatch

A Amazon EMR envia métricas para a Amazon CloudWatch. Em resposta, você pode usar CloudWatch para definir alarmes em suas EMR métricas da Amazon. Por exemplo, você pode

configurar um alarme CloudWatch para enviar um e-mail sempre que a HDFS utilização ultrapassar 80%. Para obter instruções detalhadas, consulte Criar ou editar um CloudWatch alarme no Guia do CloudWatch usuário da Amazon.

Respondendo aos eventos de capacidade insuficiente EMR de instâncias do cluster da Amazon

Visão geral

EMROs clusters da Amazon retornam o código do evento `EC2 provisioning - Insufficient Instance Capacity` quando a zona de disponibilidade selecionada não tem capacidade suficiente para atender à solicitação de início ou redimensionamento do cluster. O evento é emitido periodicamente com grupos de instâncias e frotas de instâncias se a Amazon encontrar EMR repetidamente exceções de capacidade insuficientes e não puder atender à sua solicitação de provisionamento para iniciar ou redimensionar o cluster.

Esta página descreve como você pode responder melhor a esse tipo de evento quando ele ocorre em seu EMR cluster.

Solução recomendada a um evento de capacidade insuficiente

Recomendamos responder a um evento de capacidade insuficiente de uma das seguintes maneiras:

- Aguarde a recuperação da capacidade. Como a capacidade muda com frequência, uma exceção de capacidade insuficiente pode se recuperar sozinha. Seus clusters começarão ou terminarão de ser redimensionados assim que a EC2 capacidade da Amazon estiver disponível.
- Como alternativa, você pode encerrar o cluster, modificar as configurações de tipo de instância e criar um novo cluster com a solicitação de configuração de cluster atualizada. Para obter mais informações, consulte Práticas recomendadas para flexibilidade de instâncias e de zona de disponibilidade.

Também é possível configurar regras ou respostas automatizadas para um evento de capacidade insuficiente, conforme descrito na próxima seção.

Recuperação automatizada de um evento de capacidade insuficiente

Você pode criar automação em resposta a EMR eventos da Amazon, como aqueles com código de evento`EC2 provisioning - Insufficient Instance Capacity`. Por exemplo, a AWS Lambda função a seguir encerra um EMR cluster com um grupo de instâncias que usa instâncias sob demanda e, em seguida, cria um novo EMR cluster com um grupo de instâncias que contém tipos de instância diferentes da solicitação original.

Estas condições acionam a ocorrência do processo automatizado:

- O evento de capacidade insuficiente foi emitido para nós primários ou centrais durante mais de 20 minutos.
- O cluster não está em um WAITINGestado READYou. Para obter mais informações sobre estados de EMR cluster, consulteNoções básicas sobre o ciclo de vida do cluster.

> **Note**
>
> Ao criar um processo automatizado para uma exceção de capacidade insuficiente, considere que o evento de capacidade insuficiente é recuperável. A capacidade geralmente muda e seus clusters retomarão o redimensionamento ou iniciarão a operação assim que a EC2 capacidade da Amazon estiver disponível.

Example função para responder ao evento de capacidade insuficiente

```
// Lambda code with Python 3.10 and handler is lambda_function.lambda_handler
// Note: related IAM role requires permission to use Amazon EMR

import json
import boto3
import datetime
from datetime import timezone

INSUFFICIENT_CAPACITY_EXCEPTION_DETAIL_TYPE = "EMR Instance Group Provisioning"
INSUFFICIENT_CAPACITY_EXCEPTION_EVENT_CODE = (
    "EC2 provisioning - Insufficient Instance Capacity"
)
ALLOWED_INSTANCE_TYPES_TO_USE = [
    "m5.xlarge",
    "c5.xlarge",
    "m5.4xlarge",
    "m5.2xlarge",
    "t3.xlarge",
]
CLUSTER_START_ACCEPTABLE_STATES = ["WAITING", "RUNNING"]
CLUSTER_START_SLA = 20

CLIENT = boto3.client("emr", region_name="us-east-1")
```

```
# checks if the incoming event is 'EMR Instance Fleet Provisioning' with eventCode 'EC2
 provisioning - Insufficient Instance Capacity'
def is_insufficient_capacity_event(event):
    if not event["detail"]:
        return False
    else:
        return (
            event["detail-type"] == INSUFFICIENT_CAPACITY_EXCEPTION_DETAIL_TYPE
            and event["detail"]["eventCode"]
            == INSUFFICIENT_CAPACITY_EXCEPTION_EVENT_CODE
        )


# checks if the cluster is eligible for termination
def is_cluster_eligible_for_termination(event, describeClusterResponse):
    # instanceGroupType could be CORE, MASTER OR TASK
    instanceGroupType = event["detail"]["instanceGroupType"]
    clusterCreationTime = describeClusterResponse["Cluster"]["Status"]["Timeline"][
        "CreationDateTime"
    ]
    clusterState = describeClusterResponse["Cluster"]["Status"]["State"]

    now = datetime.datetime.now()
    now = now.replace(tzinfo=timezone.utc)
    isClusterStartSlaBreached = clusterCreationTime < now - datetime.timedelta(
        minutes=CLUSTER_START_SLA
    )

    # Check if instance group receiving Insufficient capacity exception is CORE or
 PRIMARY (MASTER),
    # and it's been more than 20 minutes since cluster was created but the cluster
 state and the cluster state is not updated to RUNNING or WAITING
    if (
        (instanceGroupType == "CORE" or instanceGroupType == "MASTER")
        and isClusterStartSlaBreached
        and clusterState not in CLUSTER_START_ACCEPTABLE_STATES
    ):
        return True
    else:
        return False


# Choose item from the list except the exempt value
```

```
def choice_excluding(exempt):
    for i in ALLOWED_INSTANCE_TYPES_TO_USE:
        if i != exempt:
            return i


# Create a new cluster by choosing different InstanceType.
def create_cluster(event):
    # instanceGroupType cloud be CORE, MASTER OR TASK
    instanceGroupType = event["detail"]["instanceGroupType"]

    # Following two lines assumes that the customer that created the cluster already
 knows which instance types they use in original request
    instanceTypesFromOriginalRequestMaster = "m5.xlarge"
    instanceTypesFromOriginalRequestCore = "m5.xlarge"

    # Select new instance types to include in the new createCluster request
    instanceTypeForMaster = (
        instanceTypesFromOriginalRequestMaster
        if instanceGroupType != "MASTER"
        else choice_excluding(instanceTypesFromOriginalRequestMaster)
    )
    instanceTypeForCore = (
        instanceTypesFromOriginalRequestCore
        if instanceGroupType != "CORE"
        else choice_excluding(instanceTypesFromOriginalRequestCore)
    )

    print("Starting to create cluster...")
    instances = {
        "InstanceGroups": [
            {
                "InstanceRole": "MASTER",
                "InstanceCount": 1,
                "InstanceType": instanceTypeForMaster,
                "Market": "ON_DEMAND",
                "Name": "Master",
            },
            {
                "InstanceRole": "CORE",
                "InstanceCount": 1,
                "InstanceType": instanceTypeForCore,
                "Market": "ON_DEMAND",
                "Name": "Core",
```

```
            },
        ]
    }
    response = CLIENT.run_job_flow(
        Name="Test Cluster",
        Instances=instances,
        VisibleToAllUsers=True,
        JobFlowRole="EMR_EC2_DefaultRole",
        ServiceRole="EMR_DefaultRole",
        ReleaseLabel="emr-6.10.0",
    )

    return response["JobFlowId"]


# Terminated the cluster using clusterId received in an event
def terminate_cluster(event):
    print("Trying to terminate cluster, clusterId: " + event["detail"]["clusterId"])
    response = CLIENT.terminate_job_flows(JobFlowIds=[event["detail"]["clusterId"]])
    print(f"Terminate cluster response: {response}")


def describe_cluster(event):
    response = CLIENT.describe_cluster(ClusterId=event["detail"]["clusterId"])
    return response


def lambda_handler(event, context):
    if is_insufficient_capacity_event(event):
        print(
            "Received insufficient capacity event for instanceGroup, clusterId: "
            + event["detail"]["clusterId"]
        )

        describeClusterResponse = describe_cluster(event)

        shouldTerminateCluster = is_cluster_eligible_for_termination(
            event, describeClusterResponse
        )
        if shouldTerminateCluster:
            terminate_cluster(event)

            clusterId = create_cluster(event)
            print("Created a new cluster, clusterId: " + clusterId)
```

```
        else:
            print(
                "Cluster is not eligible for termination, clusterId: "
                + event["detail"]["clusterId"]
            )

    else:
        print("Received event is not insufficient capacity event, skipping")
```

Respondendo aos eventos de tempo limite de redimensionamento da frota de instâncias de EMR cluster da Amazon

Visão geral

EMROs clusters da Amazon emitem eventos durante a execução da operação de redimensionamento, por exemplo, clusters de frotas. Os eventos de tempo limite de provisionamento são emitidos quando a EMR Amazon interrompe o provisionamento de capacidade spot ou sob demanda para a frota após o tempo limite expirar. O usuário pode configurar a duração do tempo limite como parte das especificações de redimensionamento das frotas de instâncias. Em cenários de redimensionamento consecutivo para a mesma frota de instâncias, a Amazon EMR emite os `On-Demand provisioning timeout - continuing resize` eventos `Spot provisioning timeout - continuing resize` ou quando o tempo limite da operação de redimensionamento atual expira. Em seguida, começa a provisionar capacidade para a próxima operação de redimensionamento da frota.

Responder a eventos de tempo limite de redimensionamento da frota de instâncias

Recomendamos responder a um evento de tempo limite de aprovisionamento de uma das seguintes maneiras:

- Revisite as especificações de redimensionamento e repita a operação de redimensionamento. Como a capacidade muda com frequência, seus clusters serão redimensionados com sucesso assim que a EC2 capacidade da Amazon estiver disponível. Recomendamos que os clientes configurem valores mais baixos para a duração do tempo limite dos trabalhos que exigem mais rigorSLAs.
- Como alternativa, você pode:
  - Iniciar um novo cluster com tipos de instância diversificados com base nas práticas recomendadas para instâncias e na flexibilidade da zona de disponibilidade ou
  - Iniciar um cluster com capacidade sob demanda

- Para o evento de tempo limite de provisionamento e redimensionamento contínuo, você também pode aguardar o processamento das operações de redimensionamento. A Amazon EMR continuará processando sequencialmente as operações de redimensionamento acionadas para a frota, respeitando as especificações de redimensionamento configuradas.

Também é possível configurar regras ou respostas automatizadas para este evento, conforme descrito na próxima seção.

Recuperação automatizada de um evento de tempo limite de provisionamento

Você pode criar automação em resposta aos EMR eventos da Amazon com o código do `Spot Provisioning timeout` evento. Por exemplo, a AWS Lambda função a seguir encerra um EMR cluster com uma frota de instâncias que usa instâncias spot para nós de tarefas e, em seguida, cria um novo EMR cluster com uma frota de instâncias que contém tipos de instância mais diversificados do que a solicitação original. Neste exemplo, o evento `Spot Provisioning timeout` emitido para os nós de tarefa acionará a execução da função do Lambda.

Example Exemplo de função para responder ao evento **`Spot Provisioning timeout`**

```
// Lambda code with Python 3.10 and handler is lambda_function.lambda_handler
// Note: related IAM role requires permission to use Amazon EMR

import json
import boto3
import datetime
from datetime import timezone

SPOT_PROVISIONING_TIMEOUT_EXCEPTION_DETAIL_TYPE = "EMR Instance Fleet Resize"
SPOT_PROVISIONING_TIMEOUT_EXCEPTION_EVENT_CODE = (
    "Spot Provisioning timeout"
)

CLIENT = boto3.client("emr", region_name="us-east-1")

# checks if the incoming event is 'EMR Instance Fleet Resize' with eventCode 'Spot
 provisioning timeout'
def is_spot_provisioning_timeout_event(event):
    if not event["detail"]:
        return False
    else:
        return (
```

```
            event["detail-type"] == SPOT_PROVISIONING_TIMEOUT_EXCEPTION_DETAIL_TYPE
            and event["detail"]["eventCode"]
            == SPOT_PROVISIONING_TIMEOUT_EXCEPTION_EVENT_CODE
        )


# checks if the cluster is eligible for termination
def is_cluster_eligible_for_termination(event, describeClusterResponse):
    # instanceFleetType could be CORE, MASTER OR TASK
    instanceFleetType = event["detail"]["instanceFleetType"]

    # Check if instance fleet receiving Spot provisioning timeout event is TASK
    if (instanceFleetType == "TASK"):
        return True
    else:
        return False


# create a new cluster by choosing different InstanceType.
def create_cluster(event):
    # instanceFleetType cloud be CORE, MASTER OR TASK
    instanceFleetType = event["detail"]["instanceFleetType"]

    # the following two lines assumes that the customer that created the cluster
 already knows which instance types they use in original request
    instanceTypesFromOriginalRequestMaster = "m5.xlarge"
    instanceTypesFromOriginalRequestCore = "m5.xlarge"

    # select new instance types to include in the new createCluster request
    instanceTypesForTask = [
        "m5.xlarge",
        "m5.2xlarge",
        "m5.4xlarge",
        "m5.8xlarge",
        "m5.12xlarge"
    ]

    print("Starting to create cluster...")
    instances = {
        "InstanceFleets": [
            {
                "InstanceFleetType":"MASTER",
                "TargetOnDemandCapacity":1,
                "TargetSpotCapacity":0,
```

```
                "InstanceTypeConfigs":[
                    {
                        'InstanceType': instanceTypesFromOriginalRequestMaster,
                        "WeightedCapacity":1,
                    }
                ]
            },
            {
                "InstanceFleetType":"CORE",
                "TargetOnDemandCapacity":1,
                "TargetSpotCapacity":0,
                "InstanceTypeConfigs":[
                    {
                        'InstanceType': instanceTypesFromOriginalRequestCore,
                        "WeightedCapacity":1,
                    }
                ]
            },
            {
                "InstanceFleetType":"TASK",
                "TargetOnDemandCapacity":0,
                "TargetSpotCapacity":100,
                "LaunchSpecifications":{},
                "InstanceTypeConfigs":[
                    {
                        'InstanceType': instanceTypesForTask[0],
                        "WeightedCapacity":1,
                    },
                    {
                        'InstanceType': instanceTypesForTask[1],
                        "WeightedCapacity":2,
                    },
                    {
                        'InstanceType': instanceTypesForTask[2],
                        "WeightedCapacity":4,
                    },
                    {
                        'InstanceType': instanceTypesForTask[3],
                        "WeightedCapacity":8,
                    },
                    {
                        'InstanceType': instanceTypesForTask[4],
                        "WeightedCapacity":12,
                    }
```

```
                ],
                "ResizeSpecifications": {
                    "SpotResizeSpecification": {
                        "TimeoutDurationMinutes": 30
                    }
                }
            }
        ]
    }
    response = CLIENT.run_job_flow(
        Name="Test Cluster",
        Instances=instances,
        VisibleToAllUsers=True,
        JobFlowRole="EMR_EC2_DefaultRole",
        ServiceRole="EMR_DefaultRole",
        ReleaseLabel="emr-6.10.0",
    )

    return response["JobFlowId"]


# terminated the cluster using clusterId received in an event
def terminate_cluster(event):
    print("Trying to terminate cluster, clusterId: " + event["detail"]["clusterId"])
    response = CLIENT.terminate_job_flows(JobFlowIds=[event["detail"]["clusterId"]])
    print(f"Terminate cluster response: {response}")


def describe_cluster(event):
    response = CLIENT.describe_cluster(ClusterId=event["detail"]["clusterId"])
    return response


def lambda_handler(event, context):
    if is_spot_provisioning_timeout_event(event):
        print(
            "Received spot provisioning timeout event for instanceFleet, clusterId: "
            + event["detail"]["clusterId"]
        )

        describeClusterResponse = describe_cluster(event)

        shouldTerminateCluster = is_cluster_eligible_for_termination(
            event, describeClusterResponse
```

```
        )
        if shouldTerminateCluster:
            terminate_cluster(event)

            clusterId = create_cluster(event)
            print("Created a new cluster, clusterId: " + clusterId)
        else:
            print(
                "Cluster is not eligible for termination, clusterId: "
                + event["detail"]["clusterId"]
            )

    else:
        print("Received event is not spot provisioning timeout event, skipping")
```

Visualizar métricas para aplicações de cluster com o Ganglia

O Ganglia está disponível nas EMR versões da Amazon entre 4.2 e 6.15. O Ganglia é um projeto de código aberto que é um sistema distribuído e escalável projetado para monitorar clusters e grades e, ao mesmo tempo, minimizar o impacto no desempenho. Quando você habilita o Ganglia no seu cluster, pode gerar relatórios e visualizar o desempenho do cluster como um todo, bem como inspecionar o desempenho de instâncias de nós individuais. O Ganglia também é configurado para analisar e visualizar as métricas do Hadoop e do Spark. Para obter mais informações, consulte Ganglia no Guia de EMRlançamento da Amazon.

Registro de EMR API chamadas da Amazon AWS CloudTrail

EMRA Amazon está integrada com AWS CloudTrail, um serviço que fornece um registro das ações realizadas por um usuário, função ou AWS serviço na AmazonEMR. CloudTrail captura todas as API chamadas para a Amazon EMR como eventos. As chamadas capturadas incluem chamadas do EMR console da Amazon e chamadas de código para as EMR API operações da Amazon. Se você criar uma trilha, poderá habilitar a entrega contínua de CloudTrail eventos para um bucket do Amazon S3, incluindo eventos para a Amazon. EMR Se você não configurar uma trilha, ainda poderá ver os eventos mais recentes no CloudTrail console no Histórico de eventos. Usando as informações coletadas por CloudTrail, você pode determinar a solicitação que foi feita à AmazonEMR, o endereço IP do qual a solicitação foi feita, quem fez a solicitação, quando ela foi feita e detalhes adicionais.

Para saber mais sobre isso CloudTrail, consulte o Guia AWS CloudTrail do usuário.

EMRInformações da Amazon em CloudTrail

CloudTrail é ativado em sua AWS conta quando você cria a conta. Quando a atividade ocorre na AmazonEMR, essa atividade é registrada em um CloudTrail evento junto com outros eventos AWS de serviço no histórico de eventos. Você pode visualizar, pesquisar e baixar eventos recentes em sua AWS conta. Para obter mais informações, consulte Visualização de eventos com histórico de CloudTrail eventos.

Para um registro contínuo de eventos em sua AWS conta, incluindo eventos para a AmazonEMR, crie uma trilha. Uma trilha permite CloudTrail entregar arquivos de log para um bucket do Amazon S3. Por padrão, quando você cria uma trilha no console, a trilha se aplica a todas as AWS regiões. A trilha registra eventos de todas as regiões na AWS partição e entrega os arquivos de log ao bucket do Amazon S3 que você especificar. Além disso, você pode configurar outros AWS serviços para analisar e agir com base nos dados de eventos coletados nos CloudTrail registros. Para obter mais informações, consulte as informações a seguir.

- Visão geral da criação de uma trilha
- CloudTrail serviços e integrações suportados
- Configurando SNS notificações da Amazon para CloudTrail
- Recebendo arquivos de CloudTrail log de várias regiões e Recebendo arquivos de CloudTrail log de várias contas

Todas as EMR ações da Amazon são registradas CloudTrail e documentadas na EMRAPIReferência da Amazon. Por exemplo, chamadas para o `RunJobFlow` `ListCluster` e `DescribeCluster` as ações geram entradas nos arquivos de CloudTrail log.

Cada entrada de log ou evento contém informações sobre quem gerou a solicitação. As informações de identidade ajudam a determinar:

- Se a solicitação foi feita com credenciais de usuário root ou AWS Identity and Access Management (IAM).
- Se a solicitação foi feita com credenciais de segurança temporárias de um perfil ou de um usuário federado.
- Se a solicitação foi feita por outro AWS serviço.

No caso de um processo, em vez de um usuário criar um cluster, você pode usar o identificador `principalId` para determinar o usuário associado à criação do cluster. Para obter mais informações, consulte o CloudTrail userIdentityelemento.

Exemplo: entradas do arquivo EMR de log da Amazon

Uma trilha é uma configuração que permite a entrega de eventos como arquivos de log para um bucket do Amazon S3 que você especificar. CloudTrail os arquivos de log contêm uma ou mais entradas de log. Um evento representa uma única solicitação de qualquer fonte e inclui informações sobre a ação solicitada, a data e a hora da ação, os parâmetros da solicitação e assim por diante. CloudTrail os arquivos de log não são um rastreamento de pilha ordenado das API chamadas públicas, portanto, eles não aparecem em nenhuma ordem específica.

O exemplo a seguir mostra uma entrada de CloudTrail registro que demonstra a RunJobFlowação.

```
{
 "Records": [
 {
         "eventVersion":"1.01",
         "userIdentity":{
            "type":"IAMUser",
            "principalId":"EX_PRINCIPAL_ID",
            "arn":"arn:aws:iam::123456789012:user/temporary-user-xx-7M",
            "accountId":"123456789012",
            "userName":"temporary-user-xx-7M"
         },
         "eventTime":"2018-03-31T17:59:21Z",
         "eventSource":"elasticmapreduce.amazonaws.com",
         "eventName":"RunJobFlow",
         "awsRegion":"us-west-2",
         "sourceIPAddress":"192.0.2.1",
         "userAgent":"aws-sdk-java/unknown-version Linux/xx Java_HotSpot(TM)_64-
Bit_Server_VM/xx",
         "requestParameters":{
            "tags":[
               {
                  "value":"prod",
                  "key":"domain"
               },
               {
                  "value":"us-west-2",
```

```
                  "key":"realm"
               },
               {
                  "value":"VERIFICATION",
                  "key":"executionType"
               }
            ],
            "instances":{
               "slaveInstanceType":"m5.xlarge",
               "ec2KeyName":"emr-integtest",
               "instanceCount":1,
               "masterInstanceType":"m5.xlarge",
               "keepJobFlowAliveWhenNoSteps":true,
               "terminationProtected":false
            },
            "visibleToAllUsers":false,
            "name":"MyCluster",
            "ReleaseLabel":"emr-5.16.0"
         },
         "responseElements":{
            "jobFlowId":"j-2WDJCGEG4E6AJ"
         },
         "requestID":"2f482daf-b8fe-11e3-89e7-75a3d0e071c5",
         "eventID":"b348a38d-f744-4097-8b2a-e68c9b424698"
      },
  ...additional entries
   ]
}
```


Usar ajuste de escala de clusters

Você pode ajustar o número de EC2 instâncias da Amazon disponíveis para um EMR cluster da Amazon de forma automática ou manual em resposta às cargas de trabalho que têm demandas variadas. Há duas opções para usar a escalabilidade automática. Você pode ativar a escalabilidade EMR gerenciada da Amazon ou criar uma política personalizada de escalabilidade automática. A tabela a seguir descreve as diferenças entre as duas opções.

	Escalabilidade EMR gerenciada pela Amazon	Escalabilidade automática personalizada
Políticas e regras de escalabilidade	Nenhuma política necessária. A Amazon EMR gerencia a atividade de escalabilidade automática avaliando continuamente as métricas do cluster e tomando decisões de escalabilidade otimizadas.	É necessário definir e gerenciar as políticas e as regras de ajuste de escala automático, como as condições específicas que acionam ações de ajuste de escala, períodos de avaliação, períodos de esfriamento etc.
EMRLançamentos compatíveis da Amazon	Amazon EMR versão 5.30.0 e superior (exceto Amazon EMR versão 6.0.0)	Amazon EMR versão 4.0.0 e superior
Composição de cluster compatível	Grupos de instâncias ou frotas de instâncias	Somente grupos de instâncias
Configuração de limites de escalabilidade	Os limites de escalabilidade são configurados para todo o cluster.	Os limites de escalabilidade só podem ser configurados para cada grupo de instâncias.
Frequência da avaliação de métricas	A cada 5 a 10 segundos A avaliação mais frequente das métricas permite que EMR a Amazon tome decisões de escalabilidade mais precisas.	É possível definir os períodos de avaliação apenas em incrementos de cinco minutos.
Aplicações compatíveis	Somente YARN aplicativos são suportados, como Spark, Hadoop, Hive, Flink. A escalabilidade EMR gerenciada da Amazon não oferece suporte a aplicativos que não	Você pode escolher quais aplicativos são compatíveis ao definir as regras de escalabilidade automática.

	Escalabilidade EMR gerenciada pela Amazon	Escalabilidade automática personalizada
	sejam baseados emYARN, como Presto ou. HBase	

Considerações

- Um EMR cluster da Amazon sempre compreende um ou três nós primários. Depois de configurar o cluster inicialmente, você só pode escalar os nós centrais e de tarefas. Você não pode escalar o número de nós primários para o cluster.
- Para grupos de instâncias, as operações de reconfiguração e redimensionamento ocorrem consecutivamente e não simultaneamente. Se você iniciar uma reconfiguração enquanto um grupo de instâncias estiver sendo redimensionado, a reconfiguração será iniciada quando o grupo de instâncias concluir o redimensionamento em andamento. Por outro lado, se você iniciar uma operação de redimensionamento enquanto uma instância agrupa sua reconfiguração.

Usando escalabilidade gerenciada na Amazon EMR

> **Important**
>
> É altamente recomendável que você use a EMR versão mais recente da Amazon (Amazon EMR 7.2.0) para escalabilidade gerenciada. Em versões anteriores, você pode enfrentar falhas de aplicações intermitentes ou atrasos no ajuste de escala. A Amazon EMR resolveu esse problema com as versões 5.x 5.30.2, 5.31.1, 5.32.1, 5.33.1 e superiores, e com as versões 6.x 6.1.1, 6.2.1, 6.3.1 e superiores. Para ter mais informações sobre Regiões e disponibilidade de versões, consulte Disponibilidade gerenciada de ajuste de escala.

Visão geral

Com EMR as versões 5.30.0 e superiores da Amazon (exceto a Amazon EMR 6.0.0), você pode habilitar a escalabilidade gerenciada da AmazonEMR. Com o ajuste de escala gerenciado, é possível aumentar ou diminuir automaticamente o número de instâncias ou unidades no cluster com base na workload. A Amazon avalia EMR continuamente as métricas do cluster para tomar decisões de

escalabilidade que otimizem seus clusters em termos de custo e velocidade. O ajuste de escala gerenciado está disponível para clusters compostos por grupos de instâncias ou frotas de instâncias.

Disponibilidade gerenciada de ajuste de escala

- A seguir Regiões da AWS, a escalabilidade EMR gerenciada da Amazon está disponível com o Amazon EMR 6.14.0 e superior:
  - Ásia-Pacífico (Hyderabad) (ap-south-2)
  - Ásia-Pacífico (Jacarta) (ap-southeast-3)
  - Europa (Espanha) (eu-south-2)
- A seguir Regiões da AWS, a escalabilidade EMR gerenciada da Amazon está disponível com o Amazon EMR 5.30.0 e 6.1.0 e superior:
  - Leste dos EUA (Norte da Virgínia) (us-east-1)
  - Leste dos EUA (Ohio) (us-east-2)
  - Oeste dos EUA (Oregon) (us-west-2)
  - Oeste dos EUA (Norte da Califórnia) (us-west-1)
  - África (Cidade do Cabo) (af-south-1)
  - Ásia-Pacífico (Hong Kong) (ap-east-1)
  - Ásia-Pacífico (Mumbai) (ap-south-1)
  - Ásia-Pacífico (Seul) (ap-northeast-2)
  - Ásia-Pacífico (Singapura) (ap-southeast-1)
  - Ásia-Pacífico (Sydney) (ap-southeast-2)
  - Ásia Pacific (Tóquio) (ap-northeast-1)
  - Canadá (Central) (ca-central-1)
  - América do Sul (São Paulo) (sa-east-1)
  - Europa (Frankfurt) (eu-central-1)
  - Europa (Irlanda) (eu-west-1)
  - Europa (Londres) (eu-west-2)
  - UE (Milão) (eu-south-1)
  - Europa (Paris) (eu-west-3)
  - UE (Estocolmo) (eu-north-1)
  - China (Pequim) (cn-north-1)

  - China (Ningxia) (cn-northwest-1)
  - AWS GovCloud (Leste dos EUA) (us-gov-east-1)
  - AWS GovCloud (Oeste dos EUA) (us-gov-west-1)
- A escalabilidade EMR gerenciada da Amazon só funciona com YARN aplicativos, como Spark, Hadoop, Hive e Flink. Ele não oferece suporte a aplicativos que não são baseados emYARN, como Presto e. HBase

Parâmetros do ajuste de escala gerenciado

É necessário configurar os parâmetros a seguir para ajuste de escala gerenciado. O limite só se aplica aos nós core e de tarefa. Não é possível escalar o nó primário após a configuração inicial.

- Minimum (`MinimumCapacityUnits`) — O limite inferior da EC2 capacidade permitida em um cluster. Ele é medido por meio de núcleos ou instâncias da unidade central de processamento virtual (vCPU) para grupos de instâncias. É medido por meio de unidades para frotas de instâncias.
- Máximo (`MaximumCapacityUnits`) — O limite superior da EC2 capacidade permitida em um cluster. Ele é medido por meio de núcleos ou instâncias da unidade central de processamento virtual (vCPU) para grupos de instâncias. É medido por meio de unidades para frotas de instâncias.
- Limite sob demanda (`MaximumOnDemandCapacityUnits`) (opcional) — O limite superior da EC2 capacidade permitida para o tipo de mercado sob demanda em um cluster. Se este parâmetro não for especificado, o valor `MaximumCapacityUnits` será usado como padrão.
  - Esse parâmetro é usado para dividir a alocação de capacidade entre instâncias sob demanda e spot. Por exemplo, se você definir o parâmetro mínimo como 2 instâncias, o parâmetro máximo como 100 instâncias, o limite sob demanda como 10 instâncias, a escalabilidade EMR gerenciada da Amazon escalará até 10 instâncias sob demanda e aloca a capacidade restante para instâncias spot. Para obter mais informações, consulte Cenários de alocação de nós.
- Máximo de nós principais (`MaximumCoreCapacityUnits`) (opcional) — O limite superior da EC2 capacidade permitida para o tipo de nó principal em um cluster. Se este parâmetro não for especificado, o valor `MaximumCapacityUnits` será usado como padrão.
  - Esse parâmetro é usado para dividir a alocação de capacidade entre nós de centrais e de tarefa. Por exemplo, se você definir o parâmetro mínimo como 2 instâncias, o máximo como 100 instâncias, o nó principal máximo como 17 instâncias, a escalabilidade EMR gerenciada da

Amazon escalará até 17 nós principais e alocará as 83 instâncias restantes aos nós de tarefas. Para obter mais informações, consulte Cenários de alocação de nós.

Para obter mais informações sobre parâmetros de ajuste de escala gerenciado, consulte `ComputeLimits`.

Considerações sobre a escalabilidade EMR gerenciada da Amazon

- A escalabilidade gerenciada é suportada em EMR versões limitadas Regiões da AWS e da Amazon. Para obter mais informações, consulte Disponibilidade gerenciada de ajuste de escala.
- Você deve configurar os parâmetros necessários para a escalabilidade EMR gerenciada da Amazon. Para obter mais informações, consulte Parâmetros do ajuste de escala gerenciado.
- Para usar o escalonamento gerenciado, o processo coletor de métricas deve ser capaz de se conectar ao API endpoint público para escalabilidade gerenciada no Gateway. API Se você usar um DNS nome privado com Amazon Virtual Private Cloud, o escalonamento gerenciado não funcionará corretamente. Para garantir que o ajuste de escala gerenciado funcione, é recomendável executar uma das seguintes ações:
  - Remova o VPC endpoint da interface API Gateway da sua AmazonVPC.
  - Siga as instruções em Por que recebo um erro HTTP 403 Forbidden ao me conectar ao meu API Gateway a APIs partir de um VPC? para desativar a configuração DNS do nome privado.
  - Em vez disso, inicie o cluster em sua sub-rede privada. Para obter mais informações, consulte o tópico em Sub-redes privadas.
- Se suas YARN tarefas ficarem intermitentemente lentas durante a redução e os registros do YARN Resource Manager mostrarem que a maioria dos seus nós foi negada durante esse período, você poderá ajustar o limite de tempo limite de descomissionamento.

  Reduza `spark.blacklist.decommissioning.timeout` de uma hora para um minuto para disponibilizar o nó para que outros contêineres pendentes continuem o processamento de tarefa.

  Você também deve definir um valor maior `YARN.resourcemanager.nodemanager-graceful-decommission-timeout-secs` para garantir que a Amazon EMR não force o encerramento do nó enquanto a “Tarefa Spark” mais longa ainda estiver em execução no nó. O padrão atual é 60 minutos, o que significa que o contêiner é YARN encerrado à força após 60 minutos, quando o nó entra no estado de descomissionamento.

  O exemplo de linha de registro do YARN Resource Manager a seguir mostra os nós adicionados ao estado de descomissionamento:

```
2021-10-20 15:55:26,994 INFO
 org.apache.hadoop.YARN.server.resourcemanager.DefaultAMSProcessor
 (IPC Server handler 37 on default port 8030): blacklist are updated in
 Scheduler.blacklistAdditions: [ip-10-10-27-207.us-west-2.compute.internal,
 ip-10-10-29-216.us-west-2.compute.internal, ip-10-10-31-13.us-
west-2.compute.internal, ... , ip-10-10-30-77.us-west-2.compute.internal],
 blacklistRemovals: []
```

Veja mais detalhes sobre como a Amazon EMR se integra à listagem de YARN negação durante o descomissionamento de nós, casos em que nós na Amazon EMR podem ser listados como negados e como configurar o comportamento de desativação de nós do Spark.

- A utilização excessiva de EBS volumes pode causar problemas de escalabilidade gerenciada. Recomendamos que você mantenha o EBS volume abaixo de 90% de utilização. Para obter mais informações, consulte Armazenamento de instâncias.
- CloudWatch As métricas da Amazon são essenciais para a operação do escalonamento EMR gerenciado da Amazon. Recomendamos que você monitore de perto CloudWatch as métricas da Amazon para garantir que os dados não estejam ausentes. Para obter mais informações sobre como você pode configurar CloudWatch alarmes para detectar métricas ausentes, consulte Usando CloudWatch alarmes da Amazon.
- As operações de ajuste de escala gerenciado nos clusters das versões 5.30.0 e 5.30.1 sem o Presto instalado podem causar falhas na aplicação ou fazer com que um grupo de instâncias ou uma frota de instâncias uniforme permaneça no estado `ARRESTED`, sobretudo quando uma operação de redução da escala verticalmente logo é seguida por uma operação de aumento da escala verticalmente.

  Como solução alternativa, escolha o Presto como um aplicativo a ser instalado ao criar um cluster com as EMR versões 5.30.0 e 5.30.1 da Amazon, mesmo que seu trabalho não exija o Presto.
- Ao definir o nó principal máximo e o limite sob demanda para a escalabilidade EMR gerenciada da Amazon, considere as diferenças entre grupos de instâncias e frotas de instâncias. Cada grupo de instâncias consiste no mesmo tipo de instância e na mesma opção de compra para instâncias: sob demanda ou spot. Para cada frota de instâncias, você pode especificar até cinco tipos de instâncias, que podem ser configurados como instâncias sob demanda e spot. Para obter mais informações, consulte Create a cluster with instance fleets or uniform instance groups, Instance fleet options e Cenários de alocação de nós.
- Com o Amazon EMR 5.30.0 e versões posteriores, se você remover a regra de saída Allow All padrão para 0.0.0.0/ para o grupo de segurança principal, deverá adicionar uma regra que permita

a TCP conectividade de saída ao seu grupo de segurança para acesso ao serviço na porta 9443. Seu grupo de segurança para acesso ao serviço também deve permitir TCP tráfego de entrada na porta 9443 do grupo de segurança mestre. Para obter mais informações sobre a configuração de grupos de segurança, consulte Grupo EMR de segurança gerenciado pela Amazon para a instância primária (sub-redes privadas).

- Você pode usar AWS CloudFormation para configurar a escalabilidade EMR gerenciada da Amazon. Para obter mais informações, consulte AWS::EMR: :Cluster no Guia do AWS CloudFormation Usuário.
- Se você estiver usando nós Spot, considere usar rótulos de nós para evitar que a Amazon EMR remova processos de aplicação quando a Amazon EMR remover nós Spot. Para obter mais informações sobre rótulos de nós, consulte Nós de tarefas.
- A rotulagem de nós não é suportada por padrão nas EMR versões 6.15 ou inferiores da Amazon. Para obter mais informações, consulte Compreender os tipos de nós: nós primários, principais e de tarefas.
- Se você estiver usando as EMR versões 6.15 ou inferiores da Amazon, só poderá atribuir rótulos de nós por tipo de nó, como nós principais e de tarefas. No entanto, se você estiver usando a EMR versão 7.0 ou superior da Amazon, poderá configurar rótulos de nós por tipo de nó e tipo de mercado, como On-Demand e Spot.
- Se a demanda do processo do aplicativo aumentar e a demanda do executor diminuir ao restringir o processo do aplicativo aos nós principais, você poderá adicionar novamente os nós principais e remover os nós da tarefa na mesma operação de redimensionamento. Para obter mais informações, consulte Compreendendo a estratégia e os cenários de alocação de nós.
- A Amazon EMR não rotula os nós de tarefas, então você não pode definir as YARN propriedades para restringir os processos de aplicativos somente para nós de tarefas. No entanto, se você quiser usar tipos de mercado como rótulos de nós, poderá usar os `SPOT` rótulos `ON_DEMAND` ou para posicionamento do processo de inscrição. Não recomendamos o uso de nós Spot para processos primários de aplicativos.
- Ao usar rótulos de nós, o total de unidades em execução no cluster pode exceder temporariamente a computação máxima definida em sua política de escalabilidade gerenciada, enquanto a Amazon EMR descomissiona algumas de suas instâncias. O total de unidades solicitadas sempre permanecerá igual ou abaixo do cálculo máximo da sua apólice.
- O escalonamento gerenciado suporta apenas os rótulos dos nós `ON_DEMAND` `CORE` e `SPOT` ou e. `TASK` Não há suporte para rótulos de nós personalizados.
- EMRA Amazon cria rótulos de nós ao criar o cluster e provisionar recursos. A Amazon EMR não suporta a adição de rótulos de nós quando você reconfigura o cluster. Você também não pode

modificar os rótulos dos nós ao configurar o escalonamento gerenciado após a inicialização do cluster.

- O escalonamento gerenciado dimensiona os nós principais e de tarefas de forma independente, com base no processo do aplicativo e na demanda do executor. Para evitar problemas de perda de HDFS dados durante a redução da escala do núcleo, siga a prática padrão para os nós principais. Para saber mais sobre as melhores práticas sobre nós principais e HDFS replicação, consulte Considerações e melhores práticas.
- Você não pode colocar o processo do aplicativo e os executores somente no nó `core` ou no `ON_DEMAND` nó. Se você quiser adicionar o processo do aplicativo e os executores em um dos nós, não use a `yarn.node-labels.am.default-node-label-expression` configuração.

  Por exemplo, para colocar o processo do aplicativo e os executores em `ON_DEMAND` nós, defina a computação máxima como a máxima no `ON_DEMAND` nó. Remova também a `yarn.node-labels.am.default-node-label-expression` configuração.

  Para adicionar o processo do aplicativo e os executores nos `core` nós, remova a `yarn.node-labels.am.default-node-label-expression` configuração.
- Ao usar o escalonamento gerenciado com rótulos de nós, defina a propriedade `yarn.scheduler.capacity.maximum-am-resource-percent: 1` se você planeja executar vários aplicativos em paralelo. Isso garante que seus processos de aplicação utilizem totalmente os `ON_DEMAND` nós `CORE` ou os nós disponíveis.
- Se você usa escalabilidade gerenciada com rótulos de nós, defina `yarn.resourcemanager.decommissioning.timeout` a propriedade com um valor maior do que o aplicativo de execução mais longa em seu cluster. Isso reduz a chance de o escalonamento EMR gerenciado da Amazon precisar reprogramar seus aplicativos para recomissionamento ou nós. `CORE` `ON_DEMAND`

Histórico de recursos

Esta tabela lista as atualizações da capacidade de escalabilidade EMR gerenciada da Amazon.

Data de lançamento	Recurso	EMRVersões da Amazon
20 de agosto de 2024	Agora, os rótulos de nós estão disponíveis no escalonamento gerenciado, para que você	7.2.0 e superior

Data de lançamento	Recurso	EMRVersões da Amazon
	possa rotular suas instâncias com base no tipo de mercado ou no tipo de nó para melhorar a escalabilidade automática.	
31 de março de 2024	O escalonamento gerenciado está disponível na região `ap-south-2` Ásia-Pacífico (Hyderabad).	6.14.0 e posterior
13 de fevereiro de 2024	O escalonamento gerenciado está disponível na região `eu-south-2` Europa (Espanha).	6.14.0 e posterior
10 de outubro de 2023	Ajuste de Escala Gerenciado disponível na região `ap-southeast-3` Ásia-Pacífico (Jacarta).	6.14.0 e posterior
28 de julho de 2023	Escalabilidade gerenciada aprimorada para mudar para um grupo de instâncias de tarefas diferente na expansão quando a Amazon EMR experimenta um atraso na expansão com o grupo de instâncias atual.	5.34.0 e posteriores, 6.4.0 e posteriores

Data de lançamento	Recurso	EMRVersões da Amazon
16 de junho de 2023	O ajuste de escala gerenciado foi aprimorado para reconhecer os nós que executam a aplicação principal, de forma que esses nós não sejam reduzidos. Para obter mais informações, consulte Noções básicas da estratégia e dos cenários de alocação de nós.	5.34.0 e posteriores, 6.4.0 e posteriores
21 de março de 2022	Foi adicionado o reconhecimento de dados de shuffle do Spark usado ao reduzir a escala verticalmente de clusters. Para EMR clusters da Amazon com o Apache Spark e o recurso de escalabilidade gerenciada ativado, a Amazon monitora EMR continuamente os executores do Spark e os locais intermediários de dados aleatórios. Usando essas informações, a Amazon EMR reduz somente instâncias subutilizadas que não contêm dados aleatórios usados ativamente. Isso evita o recálculo de dados de shuffle perdidos, ajudando a reduzir custos e melhorar a performance do trabalho. Para obter mais informações, consulte o Spark Programming Guide.	5.34.0 e posteriores, 6.4.0 e posteriores

Configurando a escalabilidade gerenciada para a Amazon EMR

As seções a seguir explicam como iniciar um EMR cluster que usa escalabilidade gerenciada com o AWS Management Console AWS SDK for Java, o ou o. AWS Command Line Interface

Tópicos

Use o AWS Management Console para configurar o escalonamento gerenciado

Você pode usar o EMR console da Amazon para configurar a escalabilidade gerenciada ao criar um cluster ou para alterar uma política de escalabilidade gerenciada para um cluster em execução.

Console

Para configurar o escalonamento gerenciado ao criar um cluster com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Escolha uma EMR versão da Amazon emr-5.30.0 ou posterior, exceto a versão emr-6.0.0.
4. Na opção Escalabilidade e provisionamento de clusters, escolha Usar EMR escalabilidade gerenciada. Especifique o número mínimo e máximo de instâncias, o máximo de instâncias do nó central e o máximo de instâncias sob demanda.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

Para configurar o escalonamento gerenciado em um cluster existente com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EMREC2Em Ativado, no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar.

3. Na guia Instâncias da página de detalhes do cluster, encontre a seção Configurações do grupo de instâncias. Na seção Editar ajuste de escala do cluster, especifique novos valores para os números Mínimo e Máximo de instâncias e o limite Sob demanda.

Use o AWS CLI para configurar o escalonamento gerenciado

Você pode usar AWS CLI comandos da Amazon EMR para configurar a escalabilidade gerenciada ao criar um cluster. Você pode usar uma sintaxe abreviada, especificando a JSON configuração em linha nos comandos relevantes, ou pode referenciar um arquivo contendo a configuração. JSON Também é possível aplicar uma política de escalabilidade gerenciada a um cluster existente e remover uma política de escalabilidade gerenciada que foi aplicada anteriormente. Além disso, você pode recuperar os detalhes da configuração de uma política de escalabilidade de um cluster em execução.

Habilitar a escalabilidade gerenciada durante a execução do cluster

É possível habilitar a escalabilidade gerenciada durante a execução do cluster, conforme demonstra o exemplo a seguir.

```
aws emr create-cluster \
 --service-role EMR_DefaultRole \
 --release-label emr-7.2.0 \
 --name EMR_Managed_Scaling_Enabled_Cluster \
 --applications Name=Spark Name=Hbase \
 --ec2-attributes KeyName=keyName,InstanceProfile=EMR_EC2_DefaultRole \
 --instance-groups InstanceType=m4.xlarge,InstanceGroupType=MASTER,InstanceCount=1
 InstanceType=m4.xlarge,InstanceGroupType=CORE,InstanceCount=2 \
 --region us-east-1 \
 --managed-scaling-policy
 ComputeLimits='{MinimumCapacityUnits=2,MaximumCapacityUnits=4,UnitType=Instances}'
```

Você também pode especificar uma configuração de política gerenciada usando a managed-scaling-policy opção -- ao usar`create-cluster`.

Aplicar uma política de escalabilidade gerenciada a um cluster existente

É possível aplicar uma política de escalabilidade gerenciada a um cluster existente, conforme demonstra o exemplo a seguir.

```
aws emr put-managed-scaling-policy
--cluster-id j-123456
```

```
--managed-scaling-policy ComputeLimits='{MinimumCapacityUnits=1,
MaximumCapacityUnits=10,  MaximumOnDemandCapacityUnits=10, UnitType=Instances}'
```

Também é possível aplicar uma política de escalabilidade gerenciada a um cluster existente usando o comando `aws emr put-managed-scaling-policy`. O exemplo a seguir usa uma referência a um JSON arquivo,`managedscaleconfig.json`, que especifica a configuração da política de escalabilidade gerenciada.

```
aws emr put-managed-scaling-policy --cluster-id j-123456 --managed-scaling-policy
 file://./managedscaleconfig.json
```

O exemplo a seguir mostra o conteúdo do arquivo `managedscaleconfig.json`, que define a política de escalabilidade gerenciada.

```
{
    "ComputeLimits": {
        "UnitType": "Instances",
        "MinimumCapacityUnits": 1,
        "MaximumCapacityUnits": 10,
        "MaximumOnDemandCapacityUnits": 10
    }
}
```

Recuperar uma configuração de política de escalabilidade gerenciada

O comando `GetManagedScalingPolicy` recupera a configuração da política. Por exemplo, o comando a seguir recupera a configuração de um cluster com o ID `j-123456`.

```
aws emr get-managed-scaling-policy --cluster-id j-123456
```

Este comando gera o seguinte exemplo de saída.

```
{
   "ManagedScalingPolicy": {
      "ComputeLimits": {
         "MinimumCapacityUnits": 1,
         "MaximumOnDemandCapacityUnits": 10,
         "MaximumCapacityUnits": 10,
         "UnitType": "Instances"
      }
   }
```

```
}
```

Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consultehttps://docs.aws.amazon.com/cli/latest/reference/emr.

Remover a política de escalabilidade gerenciada

O comando `RemoveManagedScalingPolicy` remove a configuração da política. Por exemplo, o comando a seguir remove a configuração de um cluster com o ID `j-123456`.

```
aws emr remove-managed-scaling-policy --cluster-id j-123456
```

Use AWS SDK for Java para configurar o escalonamento gerenciado

O trecho do programa a seguir mostra como configurar a escalabilidade gerenciada usando o AWS SDK for Java:

```
package com.amazonaws.emr.sample;

import java.util.ArrayList;
import java.util.List;

import com.amazonaws.AmazonClientException;
import com.amazonaws.auth.AWSCredentials;
import com.amazonaws.auth.AWSStaticCredentialsProvider;
import com.amazonaws.auth.profile.ProfileCredentialsProvider;
import com.amazonaws.regions.Regions;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduce;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduceClientBuilder;
import com.amazonaws.services.elasticmapreduce.model.Application;
import com.amazonaws.services.elasticmapreduce.model.ComputeLimits;
import com.amazonaws.services.elasticmapreduce.model.ComputeLimitsUnitType;
import com.amazonaws.services.elasticmapreduce.model.InstanceGroupConfig;
import com.amazonaws.services.elasticmapreduce.model.JobFlowInstancesConfig;
import com.amazonaws.services.elasticmapreduce.model.ManagedScalingPolicy;
import com.amazonaws.services.elasticmapreduce.model.RunJobFlowRequest;
import com.amazonaws.services.elasticmapreduce.model.RunJobFlowResult;

public class CreateClusterWithManagedScalingWithIG {

 public static void main(String[] args) {
  AWSCredentials credentialsFromProfile = getCreadentials("AWS-Profile-Name-Here");
```

```
 /**
  * Create an Amazon EMR client with the credentials and region specified in order to
create the cluster
  */
 AmazonElasticMapReduce emr = AmazonElasticMapReduceClientBuilder.standard()
  .withCredentials(new AWSStaticCredentialsProvider(credentialsFromProfile))
  .withRegion(Regions.US_EAST_1)
  .build();

 /**
  * Create Instance Groups - Primary, Core, Task
  */
 InstanceGroupConfig instanceGroupConfigMaster = new InstanceGroupConfig()
   .withInstanceCount(1)
   .withInstanceRole("MASTER")
   .withInstanceType("m4.large")
   .withMarket("ON_DEMAND");

 InstanceGroupConfig instanceGroupConfigCore = new InstanceGroupConfig()
  .withInstanceCount(4)
  .withInstanceRole("CORE")
  .withInstanceType("m4.large")
  .withMarket("ON_DEMAND");

 InstanceGroupConfig instanceGroupConfigTask = new InstanceGroupConfig()
  .withInstanceCount(5)
  .withInstanceRole("TASK")
  .withInstanceType("m4.large")
  .withMarket("ON_DEMAND");

 List<InstanceGroupConfig> igConfigs = new ArrayList<>();
 igConfigs.add(instanceGroupConfigMaster);
 igConfigs.add(instanceGroupConfigCore);
 igConfigs.add(instanceGroupConfigTask);

       /**
        *  specify applications to be installed and configured when Amazon EMR creates
the cluster
        */
 Application hive = new Application().withName("Hive");
 Application spark = new Application().withName("Spark");
 Application ganglia = new Application().withName("Ganglia");
 Application zeppelin = new Application().withName("Zeppelin");
```

```
 /**
  * Managed Scaling Configuration -
        * Using UnitType=Instances for clusters composed of instance groups
  *
        * Other options are:
        * UnitType = VCPU ( for clusters composed of instance groups)
        * UnitType = InstanceFleetUnits ( for clusters composed of instance fleets)
        **/
 ComputeLimits computeLimits = new ComputeLimits()
   .withMinimumCapacityUnits(1)
   .withMaximumCapacityUnits(20)
   .withUnitType(ComputeLimitsUnitType.Instances);

 ManagedScalingPolicy managedScalingPolicy = new ManagedScalingPolicy();
 managedScalingPolicy.setComputeLimits(computeLimits);

 // create the cluster with a managed scaling policy
 RunJobFlowRequest request = new RunJobFlowRequest()
         .withName("EMR_Managed_Scaling_TestCluster")
         .withReleaseLabel("emr-7.2.0")          // Specifies the version label for
the Amazon EMR release; we recommend the latest release
         .withApplications(hive,spark,ganglia,zeppelin)
         .withLogUri("s3://path/to/my/emr/logs")  // A URI in S3 for log files is
required when debugging is enabled.
         .withServiceRole("EMR_DefaultRole")      // If you use a custom IAM service
role, replace the default role with the custom role.
         .withJobFlowRole("EMR_EC2_DefaultRole")  // If you use a custom Amazon EMR
role for EC2 instance profile, replace the default role with the custom Amazon EMR
role.
         .withInstances(new JobFlowInstancesConfig().withInstanceGroups(igConfigs)
             .withEc2SubnetId("subnet-123456789012345")
             .withEc2KeyName("my-ec2-key-name")
             .withKeepJobFlowAliveWhenNoSteps(true))
         .withManagedScalingPolicy(managedScalingPolicy);
    RunJobFlowResult result = emr.runJobFlow(request);

    System.out.println("The cluster ID is " + result.toString());
}

public static AWSCredentials getCredentials(String profileName) {
 // specifies any named profile in .aws/credentials as the credentials provider
 try {
  return new ProfileCredentialsProvider("AWS-Profile-Name-Here")
     .getCredentials();
```

```
        } catch (Exception e) {
            throw new AmazonClientException(
                    "Cannot load credentials from .aws/credentials file. " +
                    "Make sure that the credentials file exists and that the profile
 name is defined within it.",
                    e);
        }
 }

 public CreateClusterWithManagedScalingWithIG() { }
}
```

Noções básicas da estratégia e dos cenários de alocação de nós

Esta seção fornece uma visão geral da estratégia de alocação de nós e dos cenários comuns de escalabilidade que você pode usar com a escalabilidade EMR gerenciada da Amazon.

Estratégia de alocação de nós

A escalabilidade EMR gerenciada da Amazon aloca nós principais e de tarefas com base nas seguintes estratégias de expansão e redução:

Estratégia de aumento da escala verticalmente

- Para as EMR versões 7.2 e superiores da Amazon, a escalabilidade gerenciada primeiro adiciona nós com base nos rótulos dos nós e na propriedade de restrição YARN do processo de aplicação.
- Para as EMR versões 7.2 e posteriores da Amazon, se você habilitou rótulos de nós e restringiu os processos de aplicativos aos `CORE` nós, a escalabilidade EMR gerenciada da Amazon ampliará os nós principais e os nós de tarefas se a demanda do processo do aplicativo aumentar e a demanda do executor aumentar. Da mesma forma, se você habilitou os rótulos de nós e restringiu os processos de aplicativos aos `ON_DEMAND` nós, o escalonamento gerenciado ampliará os nós sob demanda se a demanda do processo do aplicativo aumentar e escalará os nós spot se a demanda do executor aumentar.
- Se os rótulos de nós não estiverem habilitados, o posicionamento do processo de aplicativo não estará restrito a nenhum nó ou tipo de mercado.
- Ao usar rótulos de nós, o escalonamento gerenciado pode aumentar e reduzir diferentes grupos de instâncias e frotas de instâncias na mesma operação de redimensionamento. Por exemplo, em um cenário em que `instance_group1` tem um `ON_DEMAND` nó e `instance_group2` tem um `SPOT` nó, os rótulos dos nós estão habilitados e os processos do aplicativo são restritos aos nós com

o `ON_DEMAND` rótulo. O escalonamento gerenciado diminuirá `instance_group1` e aumentará `instance_group2` se a demanda do processo de aplicativos diminuir e a demanda do executor aumentar.

- Quando a Amazon EMR sofre um atraso na escalabilidade com o grupo de instâncias atual, os clusters que usam escalabilidade gerenciada mudam automaticamente para um grupo de instâncias de tarefas diferente.
- Se o `MaximumCoreCapacityUnits` parâmetro for definido, a Amazon EMR escalará os nós principais até que as unidades principais atinjam o limite máximo permitido. Toda a capacidade restante é adicionada aos nós de tarefa.
- Se o `MaximumOnDemandCapacityUnits` parâmetro for definido, a Amazon EMR escalará o cluster usando as instâncias sob demanda até que as unidades sob demanda atinjam o limite máximo permitido. Toda a capacidade restante é adicionada usando instâncias spot.
- Se os `MaximumOnDemandCapacityUnits` parâmetros `MaximumCoreCapacityUnits` e forem definidos, a Amazon EMR considerará os dois limites durante a escalabilidade.

  Por exemplo, se `MaximumCoreCapacityUnits` for menor que`MaximumOnDemandCapacityUnits`, a Amazon EMR primeiro escalará os nós principais até que o limite de capacidade do núcleo seja atingido. Para a capacidade restante, a Amazon EMR primeiro usa instâncias sob demanda para escalar nós de tarefas até que o limite sob demanda seja atingido e, em seguida, usa instâncias spot para nós de tarefas.

Estratégia de redução da escala verticalmente

- Semelhante à estratégia de expansão, a Amazon EMR remove nós com base nos rótulos dos nós. Para obter mais informações sobre rótulos de nós, consulte Compreender os tipos de nós: nós primários, principais e de tarefas.
- Se você não habilitou os rótulos de nós, o escalonamento gerenciado remove os nós de tarefas e, em seguida, remove os nós principais até atingir a capacidade desejada de redução de escala. O escalonamento gerenciado nunca reduz o cluster abaixo das restrições mínimas especificadas na política de escalabilidade gerenciada.
- EMRAs versões 5.34.0 e superiores da Amazon e EMR as versões 6.4.0 e superiores da Amazon oferecem suporte à escalabilidade gerenciada que reconhece os dados aleatórios do Spark (dados que o Spark redistribui entre partições para realizar operações específicas). Para obter mais informações sobre operações de shuffle, consulte o Guia de programação do Spark. O ajuste de escala gerenciado reduz somente as instâncias que são subutilizadas e que não contêm dados de shuffle usados ativamente. Esse ajuste de escala inteligente evita a perda não intencional de

dados de shuffle, evitando a necessidade de novas tentativas de trabalho e recálculo de dados intermediários.

- O escalonamento gerenciado primeiro remove os nós de tarefas e, em seguida, remove os nós principais até atingir a capacidade desejada de redução de escala. O cluster nunca é escalado abaixo das restrições mínimas especificadas na política de escalabilidade gerenciada.
- Para clusters lançados com o Amazon EMR 5.x versões 5.34.0 e superiores e 6.x versões 6.4.0 e superiores, a escalabilidade EMR gerenciada pela Amazon não reduz os nós que o Apache Spark tem `ApplicationMaster` em execução neles. Isso minimiza falhas e novas tentativas de trabalho, o que ajuda a melhorar a performance do trabalho e reduzir custos. Para confirmar quais nós do cluster estão executando `ApplicationMaster`, acesse o Spark History Server e filtre o driver na guia Executores do ID da aplicação Spark.

Se o cluster não tiver nenhuma carga, a Amazon EMR cancela a adição de novas instâncias de uma avaliação anterior e executa operações de redução. Se o cluster tiver uma carga pesada, a Amazon EMR cancela a remoção de instâncias e executa operações de aumento de escala.

Considerações sobre alocação de nós

Recomendamos que você use a opção de compra sob demanda para os nós principais para evitar a perda de HDFS dados em caso de recuperação do Spot. Você pode usar a opção de compra spot para nós de tarefa para reduzir custos e obter uma execução mais rápida do trabalho quando mais instâncias spot são adicionadas aos nós de tarefa.

Cenários de alocação de nós

É possível criar vários cenários de ajuste de escala com base em suas necessidades configurando os parâmetros máximo, mínimo, limite sob demanda e nó central máximo em combinações diferentes.

Cenário 1: escalar somente os nós centrais

Para escalar somente os nós centrais, os parâmetros do ajuste de escala gerenciado devem atender aos seguintes requisitos:

- O limite sob demanda é igual ao limite máximo.
- O nó central máximo é igual ao limite máximo.

Quando o limite sob demanda e os parâmetros máximos do nó central não estão especificados, ambos os parâmetros assumem o limite máximo como padrão.

Esse cenário não é aplicável se você usa escalabilidade gerenciada com rótulos de nós e restringe seus processos de aplicativo para serem executados somente em `CORE` nós, porque a escalabilidade gerenciada dimensiona os nós de tarefas para acomodar a demanda do executor.

Os exemplos a seguir demonstram o cenário de ajuste de escala somente para os nós centrais.

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
Grupos de instâncias Central: uma sob demanda De tarefa: uma sob demanda e uma spot	`UnitType`: instâncias `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 20	Escale de 1 a 20 instâncias ou unidades da frota de instâncias nos nós centrais usando o tipo sob demanda. Sem ajuste de escala nos nós de tarefa. Quando você usa escalabilidade gerenciada com rótulos de nós e restringe seus processos de aplicação a `ON_DEMAND` nós, o cluster escalará de 1 a 20 instâncias ou unidades de
Frotas de instâncias Central: uma sob demanda De tarefa: uma sob demanda e uma spot	UnitType: InstanceFleetUnits `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 20	

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
		frota de instâncias em `CORE` nós usando o `Spot` tipo `On-Demand` or, dependendo do tipo de demanda.

Cenário 2: escalar somente nós de tarefa

Para escalar somente os nós de tarefa, os parâmetros do ajuste de escala gerenciado devem atender ao seguinte requisito:

- O nó central máximo deve ser igual ao limite mínimo.

Os exemplos a seguir demonstram o cenário de ajuste de escala somente para os nós de tarefa.

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
Grupos de instâncias Central: duas sob demanda De tarefa: uma spot	`UnitType`: instâncias `MinimumCapacityUnits` : 2 `MaximumCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 2	Mantenha os nós centrais estáveis em 2 e escale somente os nós de tarefa de 0 a 18 instâncias ou unidades da frota de instâncias. A capacidade entre os limites mínimo
Frotas de instâncias Central: duas sob demanda	`UnitType`: InstanceFleetUnits `MinimumCapacityUnits` : 2	

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
De tarefa: uma spot	`MaximumCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 2	e máximo é adicionada somente aos nós de tarefa. Quando você usa escalabilidade gerenciada com rótulos de DEMAND nós e restringe seus processos de aplicação aos nós ON_, o cluster mantém os nós principais estáveis em 2 e só escalará os nós de tarefas entre 0 e 18 instâncias ou unidades de frota de instâncias que usam o `Spot` tipo `On-demand` or, dependendo do tipo de demanda.

Cenário 3: somente instâncias sob demanda no cluster

Para ter somente instâncias sob demanda, o cluster e os parâmetros de ajuste de escala gerenciado devem atender aos seguintes requisitos:

- O limite sob demanda é igual ao limite máximo.

  Quando o limite sob demanda não é especificado, o valor do parâmetro assume o limite máximo como padrão. O valor padrão indica que a Amazon EMR escala somente instâncias sob demanda.

Se o nó central máximo for menor que o limite máximo, o parâmetro do nó central máximo poderá ser usado para dividir a alocação de capacidade entre os nós centrais e os nós de tarefa.

Para habilitar esse cenário em um cluster composto por grupos de instâncias, todos os grupos de nós do cluster devem usar o tipo de mercado sob demanda durante a configuração inicial.

Esse cenário não é aplicável se você usa escalabilidade gerenciada com rótulos de nós e restringe seus processos de aplicativo para serem executados somente em `ON_DEMAND` nós, porque a escalabilidade gerenciada dimensiona `Spot` os nós para acomodar a demanda do executor.

Os exemplos a seguir demonstram o cenário de ter instâncias sob demanda em todo o cluster.

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
Grupos de instâncias Central: uma sob demanda De tarefa: uma sob demanda	`UnitType`: instâncias `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 12	Escale de 1 a 12 instâncias ou unidades da frota de instâncias nos nós centrais usando o tipo sob demanda. Escale a capacidade restante usando sob demanda em nós de tarefa. Sem
Frotas de instâncias Central: uma sob demanda	`UnitType`: InstanceFleetUnits `MinimumCapacityUnits` : 1	

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
De tarefa: uma sob demanda	`MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 12	ajuste de escala usando instâncias spot. Quando você usa escalabilidade gerenciada com rótulos de nós e restringe seus processos de aplicação a `CORE` nós, o cluster escala entre 1 a 20 instâncias ou unidades de frota de instâncias em `CORE` nós ou `task` nós usando o `ON_DEMAND` tipo, dependendo do tipo de demanda. O escalonamento nos nós principais não excederá 12 instâncias ou unidades da frota de instâncias.

Cenário 4: somente instâncias spot no cluster

Para ter somente instâncias spot, os parâmetros de ajuste de escala gerenciado devem atender aos seguintes requisitos:

- O limite sob demanda está definido como 0.

Se o nó central máximo for menor que o limite máximo, o parâmetro do nó central máximo poderá ser usado para dividir a alocação de capacidade entre os nós centrais e os nós de tarefa.

Para habilitar esse cenário em um cluster composto por grupos de instâncias, o grupo de instâncias central deve usar a opção de compra spot durante a configuração inicial. Se não houver nenhuma instância spot no grupo de instâncias de tarefas, a escalabilidade EMR gerenciada da Amazon cria um grupo de tarefas usando instâncias spot quando necessário.

Esse cenário não é aplicável se você usar o escalonamento gerenciado com rótulos de nós e restringir seus processos de aplicativo para serem executados somente em `ON_DEMAND` nós, porque o escalonamento gerenciado dimensiona `ON_DEMAND` os nós para acomodar a demanda do processo de aplicativo.

Os exemplos a seguir demonstram o cenário de ter instâncias spot em todo o cluster.

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
Grupos de instâncias Central: uma spot De tarefa: uma spot	`UnitType`: instâncias `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 0	Escale de 1 a 20 instâncias ou unidades da frota de instâncias nos nós centrais usando spot. Sem ajuste de escala usando o tipo sob demanda.
Frotas de instâncias Central: uma spot	`UnitType`: InstanceFleetUnits `MinimumCapacityUnits` : 1	

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
De tarefa: uma spot	`MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 0	Quando você usa escalabilidade gerenciada com rótulos de nós e restringe seus processos de aplicativos a `CORE` nós, o cluster se expande entre 1 a 20 instâncias ou unidades da frota de instâncias `CORE` ou `TASK` nós usando o Spot, dependendo do tipo de demanda. A Amazon EMR não escala usando o `ON_DEMAND` tipo.

Cenário 5: escalar instâncias sob demanda nos nós centrais e instâncias spot nos nós de tarefa

Para escalar instâncias sob demanda em nós centrais e instâncias spot em nós de tarefa, os parâmetros de ajuste de escala gerenciado devem atender aos seguintes requisitos:

- O limite sob demanda deve ser igual ao nó central máximo.
- Tanto o limite sob demanda como o nó central máximo devem ser menores que o limite máximo.

Para habilitar esse cenário em um cluster composto por grupos de instâncias, o grupo de nós centrais deve usar a opção de compra sob demanda.

Esse cenário não é aplicável se você usa escalabilidade gerenciada com rótulos de nós e restringe seus processos de aplicativo para serem executados somente em `ON_DEMAND` nós ou `CORE` nós.

Os exemplos a seguir demonstram o cenário de ajuste de escala de instâncias sob demanda nos nós centrais e instâncias spot nos nós de tarefa.

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
Grupos de instâncias Central: uma sob demanda De tarefa: uma sob demanda e uma spot	`UnitType`: instâncias `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 7 `MaximumCoreCapacityUnits` : 7	Escale até 6 unidades sob demanda no nó central, pois já existe 1 unidade sob demanda no nó de tarefa, e o limite máximo para sob demanda é 7. Depois escale até 13 unidades spot nos nós de tarefa.
Frotas de instâncias Central: uma sob demanda De tarefa: uma sob demanda e uma spot	`UnitType`: InstanceFleetUnits `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 7 `MaximumCoreCapacityUnits` : 7	

Cenário 6: escale **`CORE`** as instâncias para a demanda do processo do aplicativo e **`TASK`** as instâncias para a demanda do executor.

Esse cenário só é aplicável se você usar escalabilidade gerenciada com rótulos de nós e restringir os processos do aplicativo para serem executados somente em CORE nós.

Para escalar CORE os nós com base na demanda do processo do aplicativo e TASK os nós com base na demanda do executor, você deve definir as seguintes configurações na inicialização do cluster:

- yarn.node-labels.enabled:true
- yarn.node-labels.am.default-node-label-expression: 'CORE'

Se você não especificar o ON_DEMAND limite e os parâmetros máximos do CORE nó, ambos os parâmetros usarão como padrão o limite máximo.

Se o ON_DEMAND nó máximo for menor que o limite máximo, o escalonamento gerenciado usa o parâmetro do ON_DEMAND nó máximo para dividir a alocação de capacidade entre os ON_DEMAND nós. SPOT Se você definir o parâmetro máximo do CORE nó como menor ou igual ao parâmetro de capacidade mínima, CORE os nós permanecerão estáticos na capacidade máxima do núcleo.

Os exemplos a seguir demonstram o cenário de escalabilidade de CORE instâncias com base na demanda do processo do aplicativo e TASK instâncias com base na demanda do executor.

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
Grupos de instâncias Central: uma sob demanda De tarefa: uma sob demanda	UnitType: instâncias MinimumCapacityUnits : 1 MaximumCapacityUnits : 20 MaximumOnDemandCapacityUnits : 10 MaximumCoreCapacityUnits : 20	Dimensiona os CORE nós entre 1 e 20 nós com base na demanda do processo de aplicação do cluster usando o tipo de mercado sob demanda ou spot. Dimension a TASK os nós com base na
Frotas de instâncias Central: uma sob demanda	UnitType: InstanceFleetUnits MinimumCapacityUnits : 1	

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
De tarefa: uma sob demanda	`MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 10 `MaximumCoreCapacityUnits` : 20	demanda do executor e na capacidade disponível restante após a EMR `CORE` alocação dos nós pela Amazon. A soma dos `TASK` nós solicitados `CORE` e não excederá 20. `maximumCapacity` A soma dos nós principais solicitados e dos nós de tarefas sob demanda não excederá 10`maximumOnDemandCapacity` . Os nós principais ou de tarefas adicionais usam o tipo de mercado spot.

Cenário 7: escale **`ON_DEMAND`** as instâncias para a demanda do processo do aplicativo e **`SPOT`** as instâncias para a demanda do executor.

Esse cenário só é aplicável se você usar escalabilidade gerenciada com rótulos de nós e restringir os processos do aplicativo para serem executados somente em ON_DEMAND nós.

Para escalar ON_DEMAND os nós com base na demanda do processo do aplicativo e SPOT os nós com base na demanda do executor, você deve definir as seguintes configurações na inicialização do cluster:

- `yarn.node-labels.enabled:true`
- `yarn.node-labels.am.default-node-label-expression: 'ON_DEMAND'`

Se você não especificar o ON_DEMAND limite e os parâmetros máximos do CORE nó, ambos os parâmetros usarão como padrão o limite máximo.

Se o CORE nó máximo for menor que o limite máximo, o escalonamento gerenciado usa o parâmetro do CORE nó máximo para dividir a alocação de capacidade entre os CORE nós. TASK Se você definir o parâmetro máximo do CORE nó como menor ou igual ao parâmetro de capacidade mínima, CORE os nós permanecerão estáticos na capacidade máxima do núcleo.

Os exemplos a seguir demonstram o cenário de escalabilidade de instâncias sob demanda com base na demanda do processo de aplicativos e instâncias spot com base na demanda do executor.

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
Grupos de instâncias Central: uma sob demanda De tarefa: uma sob demanda	UnitType: instâncias MinimumCapacityUnits : 1 MaximumCapacityUnits : 20 MaximumOnDemandCapacityUnits : 20 MaximumCoreCapacityUnits : 10	Dimensiona os ON_DEMAND nós entre 1 e 20 nós com base na demanda do processo de aplicação do cluster usando o tipo de TASK nó CORE ou. Dimensiona SPOT os nós
Frotas de instâncias	UnitType: InstanceFleetUnits	

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
Central: uma sob demanda De tarefa: uma sob demanda	`MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 10	com base na demanda do executor e na capacidade disponível restante após a EMR `ON_DEMAND` alocação dos nós pela Amazon. A soma dos `SPOT` nós solicitados `ON_DEMAND` e não excederá 20. `maximumCapacity` A soma dos nós principais sob demanda e dos nós centrais spot solicitados não excederá 10. `maximumCoreCapacity` Outros nós sob demanda ou spot usam o tipo de `TASK` nó.

Noções básicas sobre métricas de ajuste de escala gerenciado

A Amazon EMR publica métricas de alta resolução com dados em uma granularidade de um minuto quando a escalabilidade gerenciada está habilitada para um cluster. Você pode visualizar eventos em cada iniciação e conclusão de redimensionamento controlados pela escalabilidade gerenciada com o console da Amazon ou o EMR console da Amazon. CloudWatch CloudWatch as métricas são essenciais para a operação do escalonamento EMR gerenciado da Amazon. Recomendamos que você monitore de perto CloudWatch as métricas para garantir que os dados não estejam ausentes. Para obter mais informações sobre como você pode configurar CloudWatch alarmes para detectar métricas ausentes, consulte Usando CloudWatch alarmes da Amazon. Para obter mais informações sobre o uso de CloudWatch eventos com a AmazonEMR, consulte Monitorar CloudWatch eventos.

As métricas a seguir indicam as capacidades atuais ou de destino de um cluster. Essas métricas só estão disponíveis quando a escalabilidade gerenciada está habilitada. Para clusters compostos por frotas de instâncias, as métricas de capacidade de cluster são medidas em `Units`. Para clusters compostos por grupos de instâncias, as métricas de capacidade de cluster são medidas em `Nodes` ou `vCPU` com base no tipo de unidade usado na política de escalabilidade gerenciada.

Métrica	Descrição
• `TotalUnitsRequested` • `TotalNodesRequested` • `TotalVCPURequested`	O número total desejado de vCPUs unidades/nós/em um cluster, conforme determinado pelo escalonamento gerenciado. Unidades: Contagem
• `TotalUnitsRunning` • `TotalNodesRunning` • `TotalVCPURunning`	O número total atual de vCPUs unidades/nós/disponíveis em um cluster em execução. Quando um redimensi onamento de cluster for solicitado, essa métrica será atualizada depois que as novas instâncias forem adicionad as ou removidas do cluster. Unidades: Contagem
• `CoreUnitsRequested`	

Métrica	Descrição
• `CoreNodesRequested` • `CoreVCPURequested`	O número alvo de CORE vCPUs unidades/nós/em um cluster, conforme determinado pelo escalonamento gerenciado. Unidades: Contagem
• `CoreUnitsRunning` • `CoreNodesRunning` • `CoreVCPURunning`	O número atual de CORE unidades/nós/em vCPUs execução em um cluster. Unidades: Contagem
• `TaskUnitsRequested` • `TaskNodesRequested` • `TaskVCPURequested`	O número alvo de TASK vCPUs unidades/nós/em um cluster, conforme determinado pelo escalonamento gerenciado. Unidades: Contagem
• `TaskUnitsRunning` • `TaskNodesRunning` • `TaskVCPURunning`	O número atual de TASK unidades/nós/em vCPUs execução em um cluster. Unidades: Contagem

As métricas a seguir indicam o status de uso do cluster e dos aplicativos. Essas métricas estão disponíveis para todos os EMR recursos da Amazon, mas são publicadas em uma resolução maior com dados em uma granularidade de um minuto quando a escalabilidade gerenciada é habilitada para um cluster. É possível correlacionar as métricas a seguir com as métricas de capacidade do cluster na tabela anterior para entender as decisões de escalabilidade gerenciada.

Métrica	Descrição

Métrica	Descrição
`AppsCompleted`	O número de inscrições enviadas YARN que foram concluídas. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
`AppsPending`	O número de solicitações enviadas YARN a ela está em um estado pendente. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
`AppsRunning`	O número de inscrições enviadas para YARN isso estão em execução. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
`ContainerAllocated`	O número de contêineres de recursos alocados peloResourceManager. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
`ContainerPending`	O número de contêineres na fila que ainda não foram alocados. Caso de uso: monitorar o progresso do cluster Unidades: Contagem

Métrica	Descrição
`ContainerPendingRatio`	A proporção de contêineres pendentes em relação aos contêineres alocados (ContainerPendingRatio = Container Pending / ContainerAllocated). Se ContainerAllocated = 0, então ContainerPendingRatio =ContainerPending. O valor de ContainerPendingRatio representa um número, não uma porcentagem. Esse valor é útil para escalonar recursos de cluster com base no comportamento de alocação do contêiner. Unidades: Contagem
`HDFSUtilization`	A porcentagem de HDFS armazenamento usada atualment e. Caso de uso: analisar a performance do cluster Unidade: percentual
`IsIdle`	Indica que um cluster não está mais executando nenhum trabalho, mas ainda está ativo e acumulando cobranças. É definido como 1 se nenhuma tarefa ou nenhum trabalho estiver em execução, caso contrário, é definido como 0. Esse valor é verificado em intervalos de 5 minutos, sendo que um valor de 1 indica somente que o cluster estava ocioso no momento da verificação, e não que ele ficou ocioso durante todo o período de 5 minutos. Para evitar falsos positivos, é necessário gerar um alarme quando esse valor for 1 em mais de uma verificação consecutiva de cinco minutos. Por exemplo, você pode gerar um alerta para esse valor se ele for 1 por 30 minutos ou mais. Caso de uso: monitorar a performance do cluster Unidade: booliano

Métrica	Descrição
`MemoryAvailableMB`	A quantidade de memória disponível para ser alocada. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
`MRActiveNodes`	O número de nós que estão executando MapReduce tarefas ou trabalhos no momento. Equivalente à YARN métrica`mapred.resourcemanager.NoOfActiveNodes` . Caso de uso: monitorar o progresso do cluster Unidades: Contagem
`YARNMemoryAvailablePercentage`	A porcentagem de memória restante disponível para YARN (YARNMemoryAvailablePercentage= MemoryAvailable MB/ MemoryTotalMB). Esse valor é útil para escalar os recursos do cluster com base no uso da YARN memória. Unidade: percentual

Criar grafos de métricas de ajuste de escala gerenciado

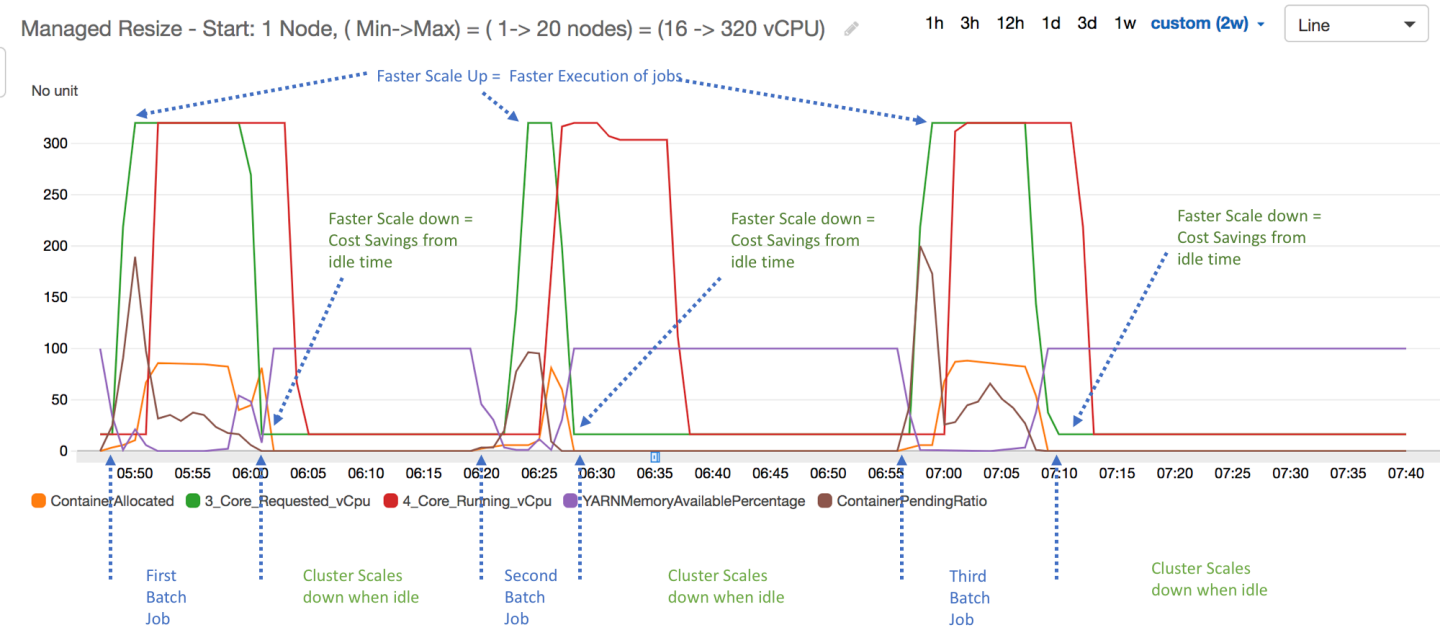
Você pode representar graficamente as métricas para visualizar os padrões de carga de trabalho do seu cluster e as decisões de escalabilidade correspondentes tomadas pelo escalonamento EMR gerenciado da Amazon, conforme demonstrado nas etapas a seguir.

Para representar graficamente as métricas de escalabilidade gerenciadas no console CloudWatch

1. Abra o CloudWatchconsole.
2. No painel de navegação, escolha Amazon EMR. Você pode pesquisar com base no identificador do cluster para monitoramento.
3. Role para baixo até a métrica para exibição em gráfico. Abra uma métrica para exibir o gráfico.

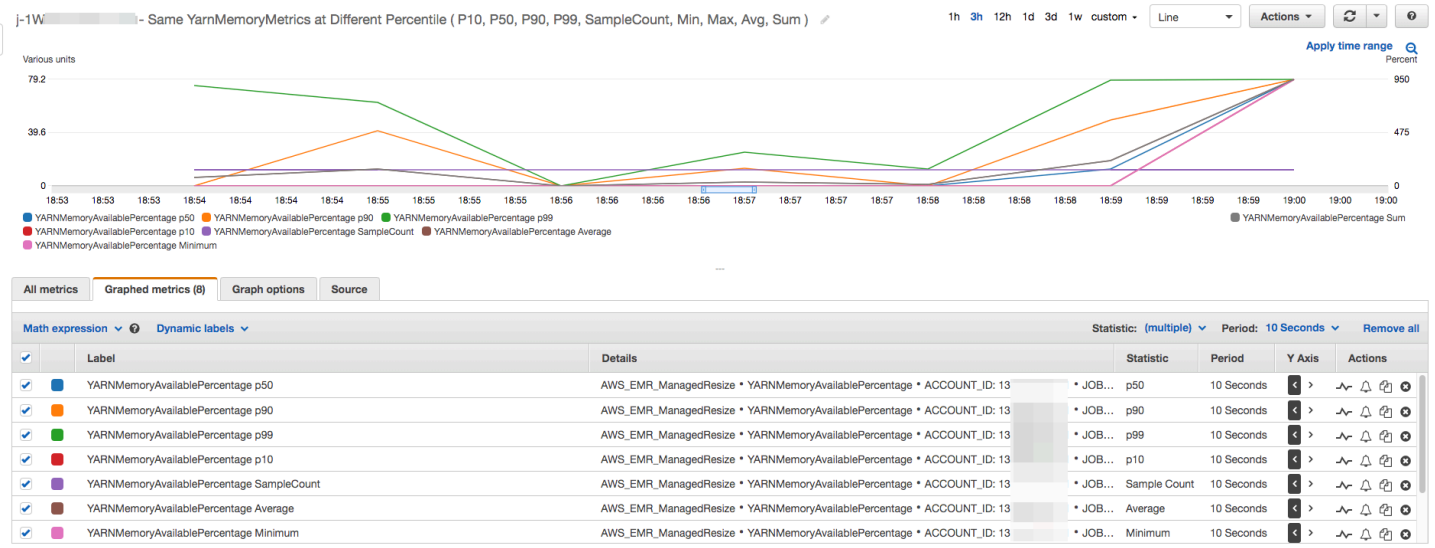
4. Para criar um gráfico de uma ou mais métricas, marque a caixa de seleção ao lado de cada métrica.

O exemplo a seguir ilustra a atividade de escalabilidade EMR gerenciada pela Amazon de um cluster. O gráfico mostra três períodos de redução automática, que economizam custos quando há uma workload menos ativa.



Todas as métricas de capacidade e uso do cluster são publicadas em intervalos de um minuto. As informações estatísticas adicionais também estão associadas a cada dado de um minuto, o que permite representar várias funções como `Percentiles`, `Min`, `Max`, `Sum`, `Average` e `SampleCount`.

Por exemplo, o gráfico a seguir representa graficamente a mesma métrica `YARNMemoryAvailablePercentage` em percentis diferentes, P10, P50, P90 e P99, juntamente com `Sum`, `Average`, `Min` e `SampleCount`.



Usar o ajuste de escala automático com uma política personalizada para grupos de instâncias

A escalabilidade automática com uma política personalizada nas EMR versões 4.0 e superiores da Amazon permite que você escale programaticamente os nós principais e os nós de tarefas com base em uma CloudWatch métrica e em outros parâmetros que você especifica em uma política de escalabilidade. A escalabilidade automática com uma política personalizada está disponível com a configuração de grupos de instâncias e não está disponível ao usar frotas de instâncias. Para obter mais informações sobre os grupos de instâncias e frotas de instâncias, consulte Criar um cluster com frotas de instâncias ou grupos de instâncias uniformes.

> **Note**
>
> Para usar a escalabilidade automática com um recurso de política personalizada na AmazonEMR, você deve definir `true` o `VisibleToAllUsers` parâmetro ao criar um cluster. Para obter mais informações, consulte SetVisibleToAllUsers.

A política de escalabilidade é parte da configuração de um grupo de instâncias. Você pode especificar uma política durante a configuração inicial de um grupo de instâncias ou pode modificar um grupo de instâncias de um cluster existente, mesmo que esse grupo de instâncias esteja ativo. Cada grupo de instâncias em um cluster, com exceção do grupo de instâncias primário, pode ter sua própria política de ajuste de escala, que consiste em regras de aumento ou redução da escala na

horizontal. As regras de expansão e redução podem ser configuradas de forma independente, com parâmetros diferentes para cada regra.

Você pode configurar políticas de escalabilidade com a AWS Management Console AWS CLI, a ou com a Amazon EMRAPI. Ao usar o AWS CLI ou a Amazon EMRAPI, você especifica a política de escalabilidade no JSON formato. Além disso, quando estiver usando o AWS CLI ou o Amazon EMRAPI, você pode especificar CloudWatch métricas personalizadas. As métricas personalizadas não estão disponíveis para seleção ao usar o AWS Management Console. Quando você cria inicialmente uma política de ajuste de escala usando o console, uma política padrão adequada para muitas aplicações é pré-configurada para ajudar você a começar. Você pode excluir ou modificar as regras padrão.

Embora o escalonamento automático permita ajustar a capacidade do EMR cluster on-the-fly, você ainda deve considerar os requisitos básicos de carga de trabalho e planejar suas configurações de nós e grupos de instâncias. Para obter mais informações, consulte Cluster configuration guidelines.

> **Note**
>
> Para a maioria das cargas de trabalho, a configuração de ambas as regras de expansão e redução é desejável para otimizar a utilização de recursos. Definir uma regra sem a outra significa que você precisaria manualmente redimensionar o número de instâncias após uma ação de escalabilidade. Em outras palavras, isso definiria uma política "unidirecional" automática de expansão ou redução com uma reinicialização manual.

Criando a IAM função para o escalonamento automático

A escalabilidade automática na Amazon EMR exige uma IAM função com permissões para adicionar e encerrar instâncias quando as atividades de escalabilidade são acionadas. Uma função padrão `EMR_AutoScaling_DefaultRole`, configurada com as políticas de função e de confiança adequadas, está disponível para esse objetivo. Quando você cria um cluster com uma política de escalabilidade pela primeira vez com o AWS Management Console, a Amazon EMR cria a função padrão e anexa a política gerenciada padrão para permissões,. `AmazonElasticMapReduceforAutoScalingRole`

Ao criar um cluster com uma política de escalabilidade automática com o AWS CLI, você deve primeiro garantir que a IAM função padrão exista ou que você tenha uma IAM função personalizada com uma política anexada que forneça as permissões apropriadas. Para criar a função padrão,

você pode executar o comando `create-default-roles` antes de criar um cluster. Em seguida, você pode especificar a opção `--auto-scaling-role EMR_AutoScaling_DefaultRole` ao criar um cluster. Como alternativa, você pode criar uma função personalizada de escalabilidade automática e, em seguida, especificá-la ao criar um cluster, por exemplo, `--auto-scaling-role` *`MyEMRAutoScalingRole`*. Se você criar uma função de escalabilidade automática personalizada para a AmazonEMR, recomendamos que você baseie as políticas de permissões para sua função personalizada com base na política gerenciada. Para obter mais informações, consulte Configure funções IAM de serviço para EMR permissões da Amazon para AWS serviços e recursos.

Noções básicas sobre as regras de ajuste de escala automático

Quando uma regra de expansão aciona uma atividade de escalabilidade para um grupo de instâncias, as instâncias da Amazon são adicionadas ao grupo de EC2 instâncias de acordo com suas regras. Novos nós podem ser usados por aplicativos como Apache Spark, Apache Hive e Presto assim que a instância da Amazon EC2 entrar no estado. `InService` Você também pode configurar uma regra de redução que encerra as instâncias e remove os nós. Para obter mais informações sobre o ciclo de vida das EC2 instâncias da Amazon que escalam automaticamente, consulte o ciclo de vida do Auto Scaling no Guia do usuário do Amazon EC2 Auto Scaling.

Você pode configurar como um cluster encerra as EC2 instâncias da Amazon. Você pode optar por rescindir no limite de EC2 horas de faturamento da instância da Amazon ou após a conclusão da tarefa. Esta configuração se aplica tanto ao Auto Scaling quanto ao redimensionamento manual de operações. Para obter mais informações sobre essa configuração, consulte Redução da escala verticalmente do cluster.

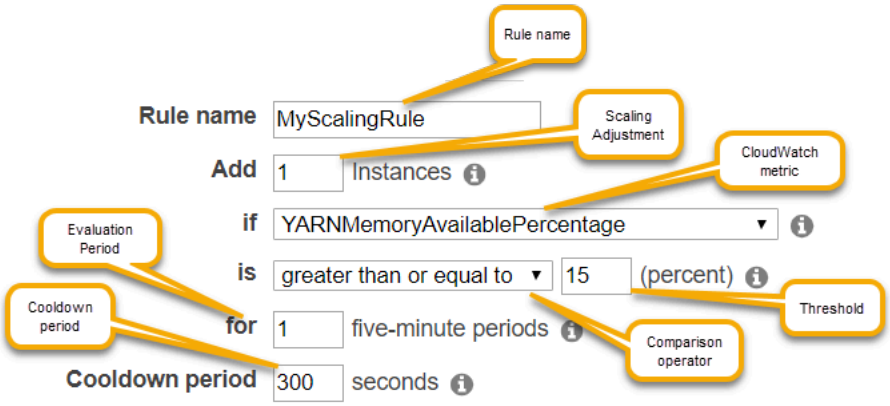
Os parâmetros a seguir se referem às regras das políticas e determinam o comportamento do Auto Scaling.

> **Note**
>
> Os parâmetros listados aqui são baseados nos AWS Management Console da AmazonEMR. Quando você usa o AWS CLI ou a Amazon EMRAPI, opções adicionais de configuração avançada estão disponíveis. Para obter mais informações sobre opções avançadas, consulte SimpleScalingPolicyConfigurationna Amazon EMR API Reference.

- Números máximo e mínimo de instâncias. A restrição Máximo de instâncias especifica o número máximo de EC2 instâncias da Amazon que podem estar no grupo de instâncias e se aplica a todas

as regras de escalabilidade. Da mesma forma, a restrição de instâncias mínimas especifica o número mínimo de EC2 instâncias da Amazon e se aplica a todas as regras de escalabilidade.

- O Rule name (Nome da regra), que deve ser único dentro da política.
- O ajuste de escalabilidade, que determina o número de EC2 instâncias a serem adicionadas (para regras de expansão horizontal) ou encerradas (para regras de expansão) durante a atividade de escalabilidade acionada pela regra.
- A CloudWatch métrica, que é observada em busca de uma condição de alarme.
- Um operador de comparação, usado para comparar a CloudWatch métrica com o valor limite e determinar uma condição de gatilho.
- Um período de avaliação, em incrementos de cinco minutos, durante o qual a CloudWatch métrica deve estar em uma condição de gatilho antes que a atividade de escalabilidade seja acionada.
- Um Cooldown period (Desaquecimento), que determina a quantidade de tempo que deve se passar entre uma ação de escalabilidade iniciada por uma regra e o início da próxima ação de escalabilidade, independentemente da regra que o aciona. Quando um grupo de instâncias conclui uma atividade de escalabilidade e atinge seu estado de pós-escala, o período de espera oferece uma oportunidade para que as CloudWatch métricas que podem acionar as atividades de escalabilidade subsequentes se estabilizem. Para obter mais informações, consulte os cooldowns do Auto Scaling no Guia do usuário do Amazon Auto Scaling. EC2



Considerações e limitações

- CloudWatch As métricas da Amazon são essenciais para a operação da escalabilidade EMR automática da Amazon. Recomendamos que você monitore de perto CloudWatch as métricas da Amazon para garantir que os dados não estejam ausentes. Para obter mais informações sobre como você pode configurar CloudWatch os alarmes da Amazon para detectar métricas ausentes, consulte Usando CloudWatch alarmes da Amazon.

- A utilização excessiva de EBS volumes pode causar problemas de escalabilidade gerenciada. Recomendamos que você monitore de perto EBS o uso do volume para garantir que o EBS volume esteja abaixo de 90% de utilização. Consulte Armazenamento de instâncias para obter informações sobre como especificar EBS volumes adicionais.
- A escalabilidade automática com uma política personalizada nas EMR versões 5.18 a 5.28 da Amazon pode apresentar falhas de escalabilidade causadas pela falta intermitente de dados nas métricas da Amazon. CloudWatch Recomendamos que você use as EMR versões mais recentes da Amazon para melhorar o escalonamento automático. Você também pode entrar em contato com o AWS Support para obter um patch se precisar usar uma EMR versão da Amazon entre 5.18 e 5.28.

Usando o AWS Management Console para configurar o escalonamento automático

Ao criar um cluster, você configura uma política de ajuste de escala para os grupos de instâncias usando as opções de configuração avançadas do cluster. Você também pode criar ou modificar uma política de escalabilidade para um grupo de instâncias em serviço modificando os grupos de instâncias nas configurações de Hardware de um cluster existente.

1. Navegue até o novo EMR console da Amazon e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Se você estiver criando um cluster, no EMR console da Amazon, selecione Criar cluster, selecione Ir para opções avançadas, escolha as opções para a Etapa 1: Software e Etapas e, em seguida, vá para a Etapa 2: Configuração de hardware.

   - ou -

   Se você estiver modificando um grupo de instâncias em um cluster em execução, selecione o seu cluster na lista de clusters e, em seguida, expanda a seção Hardware.
3. Na seção Opção de ajuste de escala e provisionamento de clusters, selecione Habilitar ajuste de escala de clusters. Selecione Criar uma política personalizada de escalabilidade automática.

   Na tabela de Políticas personalizadas de escalabilidade automática, clique no ícone de lápis que aparece na linha do grupo de instâncias que você deseja configurar. A tela Regras do Auto Scaling é exibida.
4. Digite o número de Maximum instances (Máximo de instâncias) que você deseja que o grupo de instâncias tenha quando houver uma expansão e digite o número de Minimum instances

(Mínimo de instâncias) que deseja que o grupo de instâncias tenha quando houver uma redução.

5. Clique no lápis para editar os parâmetros das regras, clique em X para remover uma regra da política e clique em Add rule (Adicionar regra) para acrescentar regras adicionais.
6. Escolha os parâmetros para as regras como descrevemos anteriormente neste tópico. Para obter descrições das CloudWatch métricas disponíveis para a AmazonEMR, consulte EMRas métricas e dimensões da Amazon no Guia CloudWatch do usuário da Amazon.

Usando o AWS CLI para configurar o escalonamento automático

Você pode usar AWS CLI comandos da Amazon EMR para configurar a escalabilidade automática ao criar um cluster e ao criar um grupo de instâncias. Você pode usar uma sintaxe abreviada, especificando a JSON configuração em linha nos comandos relevantes, ou pode referenciar um arquivo contendo a configuração. JSON Você também pode aplicar uma política de Auto Scaling para um grupo de instâncias existente e remover uma política de Auto Scaling que foi aplicada anteriormente. Além disso, você pode recuperar os detalhes da configuração de uma política de escalabilidade de um cluster em execução.

> **Important**
>
> Ao criar um cluster que tenha uma política de escalabilidade automática, você deve usar o `--auto-scaling-role` *`MyAutoScalingRole`* comando para especificar a IAM função da escalabilidade automática. A função padrão é *`EMR_AutoScaling_DefaultRole`* e pode ser criada com o comando `create-default-roles`. Esta função só pode ser adicionada quando o cluster é criado e não em um cluster existente.

Para obter uma descrição detalhada dos parâmetros disponíveis ao configurar uma política de escalabilidade automática, consulte PutAutoScalingPolicyna Amazon EMR API Reference.

Criar um cluster com uma política do Auto Scaling aplicada a um grupo de instâncias

Você pode especificar uma configuração de escalabilidade automática dentro da opção `--instance-groups` do comando `aws emr create-cluster`. O exemplo a seguir ilustra um comando create-cluster em que uma política de Auto Scaling para o grupo de instâncias core é fornecida na linha. O comando cria uma configuração de escalabilidade equivalente à política de escalabilidade horizontal padrão que aparece quando você cria uma política de escalabilidade

automática com o for Amazon. AWS Management Console EMR Para não estender a explicação, não mostramos uma política de redução. Não é recomendável criar uma regra de expansão sem uma regra de redução.

```
aws emr create-cluster --release-label emr-5.2.0 --service-role
 EMR_DefaultRole --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole
 --auto-scaling-role EMR_AutoScaling_DefaultRole  --instance-groups
 Name=MyMasterIG,InstanceGroupType=MASTER,InstanceType=m5.xlarge,InstanceCount=1
 'Name=MyCoreIG,InstanceGroupType=CORE,InstanceType=m5.xlarge,InstanceCount=2,AutoScalingPolicy
scale-out,Description=Replicates the default scale-out rule in the
 console.,Action={SimpleScalingPolicyConfiguration={AdjustmentType=CHANGE_IN_CAPACITY,ScalingAd
ElasticMapReduce,Period=300,Statistic=AVERAGE,Threshold=15,Unit=PERCENT,Dimensions=[{Key=JobFlo
```

O comando a seguir ilustra como usar a linha de comando para fornecer a definição da política do Auto Scaling como parte de um arquivo de configuração de grupo de instâncias chamado *instancegroupconfig.json*.

```
aws emr create-cluster --release-label emr-5.2.0 --service-role EMR_DefaultRole --ec2-
attributes InstanceProfile=EMR_EC2_DefaultRole --instance-groups file://your/path/to/
instancegroupconfig.json --auto-scaling-role EMR_AutoScaling_DefaultRole
```

O conteúdo do arquivo de configuração é o seguinte:

```
[
{
  "InstanceCount": 1,
  "Name": "MyMasterIG",
  "InstanceGroupType": "MASTER",
  "InstanceType": "m5.xlarge"
},
{
  "InstanceCount": 2,
  "Name": "MyCoreIG",
  "InstanceGroupType": "CORE",
  "InstanceType": "m5.xlarge",
  "AutoScalingPolicy":
    {
     "Constraints":
      {
       "MinCapacity": 2,
```

```
       "MaxCapacity": 10
      },
     "Rules":
     [
      {
       "Name": "Default-scale-out",
       "Description": "Replicates the default scale-out rule in the console for YARN
 memory.",
       "Action":{
        "SimpleScalingPolicyConfiguration":{
          "AdjustmentType": "CHANGE_IN_CAPACITY",
          "ScalingAdjustment": 1,
          "CoolDown": 300
        }
       },
       "Trigger":{
        "CloudWatchAlarmDefinition":{
          "ComparisonOperator": "LESS_THAN",
          "EvaluationPeriods": 1,
          "MetricName": "YARNMemoryAvailablePercentage",
          "Namespace": "AWS/ElasticMapReduce",
          "Period": 300,
          "Threshold": 15,
          "Statistic": "AVERAGE",
          "Unit": "PERCENT",
          "Dimensions":[
             {
               "Key" : "JobFlowId",
               "Value" : "${emr.clusterId}"
             }
          ]
        }
       }
      }
     ]
   }
}
]
```

Adicionar um grupo de instâncias com uma política do Auto Scaling a um cluster

Você pode especificar uma configuração de política de ajuste de escala usando a opção `--instance-groups` com o comando `add-instance-groups` da mesma maneira

com que usa o `create-cluster`. O exemplo a seguir usa uma referência a um JSON arquivo,*instancegroupconfig.json*, com a configuração do grupo de instâncias.

```
aws emr add-instance-groups --cluster-id j-1EKZ3TYEVF1S2 --instance-groups file://your/
path/to/instancegroupconfig.json
```

Aplicar uma política de ajuste de escala automático a um grupo de instâncias atual ou modificar uma política aplicada

Use o comando `aws emr put-auto-scaling-policy` para aplicar uma política de Auto Scaling a um grupo de instâncias existente. O grupo de instâncias precisa fazer parte de um cluster que usa a IAM função de escalabilidade automática. O exemplo a seguir usa uma referência a um JSON arquivo,*autoscaleconfig.json*, que especifica a configuração automática da política de escalabilidade.

```
aws emr put-auto-scaling-policy --cluster-id j-1EKZ3TYEVF1S2 --instance-group-
id ig-3PLUZBA6WLS07 --auto-scaling-policy file://your/path/to/autoscaleconfig.json
```

O conteúdo do arquivo `autoscaleconfig.json`, que define a mesma regra de expansão apresentada no exemplo anterior, é mostrado a seguir.

```
{
          "Constraints": {
                  "MaxCapacity": 10,
                  "MinCapacity": 2
          },
          "Rules": [{
                  "Action": {
                          "SimpleScalingPolicyConfiguration": {
                                  "AdjustmentType": "CHANGE_IN_CAPACITY",
                                  "CoolDown": 300,
                                  "ScalingAdjustment": 1
                          }
                  },
                  "Description": "Replicates the default scale-out rule in the console
 for YARN memory",
                  "Name": "Default-scale-out",
                  "Trigger": {
                          "CloudWatchAlarmDefinition": {
                                  "ComparisonOperator": "LESS_THAN",
                                  "Dimensions": [{
```

```
                                    "Key": "JobFlowId",
                                    "Value": "${emr.clusterID}"
                                }],
                                "EvaluationPeriods": 1,
                                "MetricName": "YARNMemoryAvailablePercentage",
                                "Namespace": "AWS/ElasticMapReduce",
                                "Period": 300,
                                "Statistic": "AVERAGE",
                                "Threshold": 15,
                                "Unit": "PERCENT"
                            }
                        }
                }]
    }
```

Remover uma política de ajuste de escala automático de um grupo de instâncias

```
aws emr remove-auto-scaling-policy --cluster-id j-1EKZ3TYEVF1S2 --instance-group-
id ig-3PLUZBA6WLS07
```

Recuperar configuração de uma política de ajuste de escala automático

O `describe-cluster` comando recupera a configuração da política no InstanceGroup bloco. Por exemplo, o comando a seguir recupera a configuração de um cluster com o ID `j-1CWOHP4PI30VJ`.

```
aws emr describe-cluster --cluster-id j-1CWOHP4PI30VJ
```

Este comando gera o seguinte exemplo de saída.

```
{
    "Cluster": {
        "Configurations": [],
        "Id": "j-1CWOHP4PI30VJ",
        "NormalizedInstanceHours": 48,
        "Name": "Auto Scaling Cluster",
        "ReleaseLabel": "emr-5.2.0",
        "ServiceRole": "EMR_DefaultRole",
        "AutoTerminate": false,
        "TerminationProtected": true,
```

```
        "MasterPublicDnsName": "ec2-54-167-31-38.compute-1.amazonaws.com",
        "LogUri": "s3n://aws-logs-232939870606-us-east-1/elasticmapreduce/",
        "Ec2InstanceAttributes": {
            "Ec2KeyName": "performance",
            "AdditionalMasterSecurityGroups": [],
            "AdditionalSlaveSecurityGroups": [],
            "EmrManagedSlaveSecurityGroup": "sg-09fc9362",
            "Ec2AvailabilityZone": "us-east-1d",
            "EmrManagedMasterSecurityGroup": "sg-0bfc9360",
            "IamInstanceProfile": "EMR_EC2_DefaultRole"
        },
        "Applications": [
            {
                "Name": "Hadoop",
                "Version": "2.7.3"
            }
        ],
        "InstanceGroups": [
            {
                "AutoScalingPolicy": {
                    "Status": {
                        "State": "ATTACHED",
                        "StateChangeReason": {
                            "Message": ""
                        }
                    },
                    "Constraints": {
                        "MaxCapacity": 10,
                        "MinCapacity": 2
                    },
                    "Rules": [
                        {
                            "Name": "Default-scale-out",
                            "Trigger": {
                                "CloudWatchAlarmDefinition": {
                                    "MetricName": "YARNMemoryAvailablePercentage",
                                    "Unit": "PERCENT",
                                    "Namespace": "AWS/ElasticMapReduce",
                                    "Threshold": 15,
                                    "Dimensions": [
                                        {
                                            "Key": "JobFlowId",
                                            "Value": "j-1CWOHP4PI30VJ"
                                        }
```

```
                        ],
                        "EvaluationPeriods": 1,
                        "Period": 300,
                        "ComparisonOperator": "LESS_THAN",
                        "Statistic": "AVERAGE"
                    }
                },
                "Description": "",
                "Action": {
                    "SimpleScalingPolicyConfiguration": {
                        "CoolDown": 300,
                        "AdjustmentType": "CHANGE_IN_CAPACITY",
                        "ScalingAdjustment": 1
                    }
                }
            },
            {
                "Name": "Default-scale-in",
                "Trigger": {
                    "CloudWatchAlarmDefinition": {
                        "MetricName": "YARNMemoryAvailablePercentage",
                        "Unit": "PERCENT",
                        "Namespace": "AWS/ElasticMapReduce",
                        "Threshold": 75,
                        "Dimensions": [
                            {
                                "Key": "JobFlowId",
                                "Value": "j-1CWOHP4PI30VJ"
                            }
                        ],
                        "EvaluationPeriods": 1,
                        "Period": 300,
                        "ComparisonOperator": "GREATER_THAN",
                        "Statistic": "AVERAGE"
                    }
                },
                "Description": "",
                "Action": {
                    "SimpleScalingPolicyConfiguration": {
                        "CoolDown": 300,
                        "AdjustmentType": "CHANGE_IN_CAPACITY",
                        "ScalingAdjustment": -1
                    }
                }
```

```
                    }
                ]
            },
            "Configurations": [],
            "InstanceType": "m5.xlarge",
            "Market": "ON_DEMAND",
            "Name": "Core - 2",
            "ShrinkPolicy": {},
            "Status": {
                "Timeline": {
                    "CreationDateTime": 1479413437.342,
                    "ReadyDateTime": 1479413864.615
                },
                "State": "RUNNING",
                "StateChangeReason": {
                    "Message": ""
                }
            },
            "RunningInstanceCount": 2,
            "Id": "ig-3M16XBE8C3PH1",
            "InstanceGroupType": "CORE",
            "RequestedInstanceCount": 2,
            "EbsBlockDevices": []
        },
        {
            "Configurations": [],
            "Id": "ig-OP62I28NSE8M",
            "InstanceGroupType": "MASTER",
            "InstanceType": "m5.xlarge",
            "Market": "ON_DEMAND",
            "Name": "Master - 1",
            "ShrinkPolicy": {},
            "EbsBlockDevices": [],
            "RequestedInstanceCount": 1,
            "Status": {
                "Timeline": {
                    "CreationDateTime": 1479413437.342,
                    "ReadyDateTime": 1479413752.088
                },
                "State": "RUNNING",
                "StateChangeReason": {
                    "Message": ""
                }
            },
```

```
                "RunningInstanceCount": 1
            }
        ],
        "AutoScalingRole": "EMR_AutoScaling_DefaultRole",
        "Tags": [],
        "BootstrapActions": [],
        "Status": {
            "Timeline": {
                "CreationDateTime": 1479413437.339,
                "ReadyDateTime": 1479413863.666
            },
            "State": "WAITING",
            "StateChangeReason": {
                "Message": "Cluster ready after last step completed."
            }
        }
    }
}
```

Redimensionar manualmente um cluster em execução

Você pode adicionar e remover instâncias de grupos de instâncias principais e de tarefas e frotas de instâncias em um cluster em execução com o AWS Management Console AWS CLI, ou o Amazon EMRAPI. Se um cluster usa grupos de instâncias, você altera explicitamente a contagem de instâncias. Se o cluster usa frotas de instâncias, você pode alterar as unidades de destino para instâncias sob demanda e instâncias spot. A frota de instâncias, em seguida, adiciona e remove instâncias para corresponder ao novo destino. Para obter mais informações, consulte Opções de frotas de instâncias. Os aplicativos podem usar EC2 instâncias Amazon recém-provisionadas para hospedar nós assim que as instâncias estiverem disponíveis. Quando as instâncias são removidas, a Amazon EMR encerra as tarefas de uma forma que não interrompe os trabalhos e protege contra a perda de dados. Para obter mais informações, consulte Terminar na conclusão de tarefas.

Redimensionar um cluster usando o console

Você pode usar o EMR console da Amazon para redimensionar um cluster em execução.

Console

Alterar a contagem de instâncias para um cluster existente usando o novo console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EMREC2Em Ativado, no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar. O cluster deve estar em execução, e não é possível redimensionar um cluster provisionado ou terminado.
3. Na guia Instâncias da página de detalhes do cluster, visualize o painel Grupos de instâncias.
4. Para redimensionar um grupo de instâncias já existente, selecione o botão de opção ao lado do grupo de instâncias central ou de tarefa que você deseja redimensionar e escolha Redimensionar grupo de instâncias. Especifique o novo número de instâncias do grupo de instâncias e selecione Redimensionar.

   > **Note**
   >
   > Se você optar por reduzir o tamanho de um grupo de instâncias em execução, a Amazon EMR selecionará de forma inteligente as instâncias a serem removidas do grupo para perda mínima de dados. Para um controle mais granular da ação de redimensionamento, você pode selecionar o ID do grupo de instâncias, escolher as instâncias que deseja remover e usar a opção Terminar. Para obter mais informações sobre o comportamento inteligente de redução da escala verticalmente, consulte Redução da escala verticalmente do cluster.

5. Para cancelar a ação de redimensionamento, selecione o botão de opção para um grupo de instâncias com o status Resizing e escolha Interromper redimensionamento na lista de ações.
6. Para adicionar um ou mais grupos de instâncias de tarefa ao cluster em resposta ao aumento da workload, escolha Adicionar grupo de instâncias de tarefa na lista de ações. Escolha o tipo de EC2 instância da Amazon, insira o número de instâncias para o grupo de tarefas e selecione Adicionar grupo de instâncias de tarefas para retornar ao painel Grupos de instâncias do seu cluster.

Quando você altera o número de nós, o Status do grupo de instâncias é atualizado. Quando a alteração solicitada estiver concluída, o Status muda para Running (Em execução).

Redimensionar um cluster com o AWS CLI

Você pode usar o AWS CLI para redimensionar um cluster em execução. Você pode aumentar ou diminuir o número de nós de tarefa, e pode aumentar o número de nós core de um cluster em execução. Também é possível encerrar uma instância no grupo de instâncias principal com o AWS CLI ou API o. Isso deve ser feito com cuidado. Desativar uma instância no grupo de instâncias centrais expõe você ao risco de perda de dados, e a instância não é substituída automaticamente.

Além de redimensionar os grupos centrais e de tarefa, você também pode adicionar um ou mais grupos de instâncias de tarefa a um cluster em execução usando a AWS CLI.

Para redimensionar um cluster alterando a contagem de instâncias com o AWS CLI

Você pode adicionar instâncias ao grupo principal ou ao grupo de tarefas e remover instâncias do grupo de tarefas com o AWS CLI `modify-instance-groups` subcomando com o `InstanceCount` parâmetro. Para adicionar instâncias aos grupos core ou de tarefas, aumente o `InstanceCount`. Para reduzir o número de instâncias no grupo de tarefas, diminua o `InstanceCount`. Alterar o número de instâncias do grupo de tarefas para 0 remove todas as instâncias, mas não o grupo de instâncias.

- Para aumentar o número de instâncias no grupo de instâncias da tarefa de 3 para 4, digite o comando a seguir e substitua: *`ig-31JXXXXXXBT0`* com o ID do grupo de instâncias.

  ```
  aws emr modify-instance-groups --instance-groups
   InstanceGroupId=ig-31JXXXXXXBT0,InstanceCount=4
  ```

  Para recuperar o `InstanceGroupId`, use o subcomando `describe-cluster`. A saída é um JSON objeto chamado `Cluster` que contém o ID de cada grupo de instâncias. Para usar esse comando, você precisa do ID do cluster (que pode ser recuperado usando o comando `aws emr list-clusters` ou pelo console). Para recuperar o ID do grupo de instâncias, digite o comando a seguir e substitua *`j-2AXXXXXXGAPLF`* com o ID do cluster.

  ```
  aws emr describe-cluster --cluster-id j-2AXXXXXXGAPLF
  ```

  Com o AWS CLI, você também pode encerrar uma instância no grupo de instâncias principal com o `--modify-instance-groups` subcomando.

> **Warning**
>
> A especificação de `EC2InstanceIdsToTerminate` deve ser feita com cuidado. As instâncias são encerradas imediatamente, independentemente do status dos aplicativos em execução nelas, e as instâncias não são substituídas automaticamente. Isso é verdadeiro, independentemente da configuração de Scale down behavior (Comportamento da escalabilidade vertical) do cluster. O encerramento de uma instância dessa forma tem o risco de perda de dados e de comportamento imprevisível do cluster.

Para encerrar uma instância específica, você precisa do ID do grupo de instâncias (retornado pelo `aws emr describe-cluster --cluster-id` subcomando) e do ID da instância (retornado pelo `aws emr list-instances --cluster-id` subcomando), digite o comando a seguir, substitua *`ig-6RXXXXXX07SA`* com o ID do grupo de instâncias e substitua *`i-f9XXXXf2`* com o ID da instância.

```
aws emr modify-instance-groups --instance-groups
 InstanceGroupId=ig-6RXXXXXX07SA,EC2InstanceIdsToTerminate=i-f9XXXXf2
```

Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consultehttps://docs.aws.amazon.com/cli/latest/reference/emr.

Para redimensionar um cluster adicionando grupos de instâncias de tarefas com o AWS CLI

Com o AWS CLI, você pode adicionar de 1 a 48 grupos de instâncias de tarefas a um cluster com o `--add-instance-groups` subcomando. Os grupos de instâncias de tarefa só podem ser adicionados a um cluster contendo um grupo de instâncias primárias e um grupo de instâncias centrais. Ao usar o AWS CLI, você pode adicionar até cinco grupos de instâncias de tarefas sempre que usar o `--add-instance-groups` subcomando.

1. Para adicionar um único grupo de instâncias de tarefas a um cluster, digite o comando a seguir e substitua *`j-JXBXXXXXX37R`* com o ID do cluster.

   ```
   aws emr add-instance-groups --cluster-id j-JXBXXXXXX37R --instance-groups
    InstanceCount=6,InstanceGroupType=task,InstanceType=m5.xlarge
   ```

2. Para adicionar vários grupos de instâncias de tarefas a um cluster, digite o comando a seguir e substitua *j-JXBXXXXXX37R* com o ID do cluster. Você pode adicionar até cinco grupos de instâncias de tarefas em um único comando.

```
aws emr add-instance-groups --cluster-id j-JXBXXXXXX37R --instance-
groups InstanceCount=6,InstanceGroupType=task,InstanceType=m5.xlarge
 InstanceCount=10,InstanceGroupType=task,InstanceType=m5.xlarge
```

   Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consultehttps://docs.aws.amazon.com/cli/latest/reference/emr.

Interromper um redimensionamento

Usando a EMR versão 4.1.0 ou posterior da Amazon, você pode emitir um redimensionamento em meio a uma operação de redimensionamento existente. Além disso, você pode interromper uma solicitação de redimensionamento enviada anteriormente ou enviar uma nova solicitação para substituir uma solicitação anterior, antes mesmo que ela seja concluída. Você também pode interromper um redimensionamento existente no console ou com a `ModifyInstanceGroups` API chamada com a contagem atual como a contagem de destino do cluster.

A imagem a seguir mostra um grupo de instâncias de tarefas que está sendo redimensionado mas pode ser interrompido pela opção de Stop (Interromper).

Para interromper um redimensionamento com o AWS CLI

Você pode usar o AWS CLI para interromper o redimensionamento com o `modify-instance-groups` subcomando. Suponha que você tem seis instâncias em um grupo de instâncias e deseja aumentar este número para 10. E mais tarde você decide cancelar essa solicitação:

- A solicitação inicial:

```
aws emr modify-instance-groups --instance-groups
 InstanceGroupId=ig-myInstanceGroupId,InstanceCount=10
```

   A segunda solicitação para interromper a primeira solicitação:

```
aws emr modify-instance-groups --instance-groups
 InstanceGroupId=ig-myInstanceGroupId,InstanceCount=6
```

> **Note**
>
> Como esse processo é assíncrono, você pode ver a contagem de instâncias mudar em relação às API solicitações anteriores antes que as solicitações subsequentes sejam atendidas. Em caso de redução, se você tiver um trabalho em execução nos nós, é possível que o grupo de instâncias não seja reduzido até que os nós tenham concluído seu trabalho.

Estado suspenso

Um grupo de instâncias entra em estado suspenso se encontrar muitos erros durante a tentativa de iniciar os novos nós do cluster. Por exemplo, se novos nós falharem ao realizar ações de bootstrap, o grupo de instâncias entrará em um SUSPENDEDestado, em vez de provisionar continuamente novos nós. Depois de resolver o problema básico, redefina o número desejado de nós no grupo de instâncias do cluster e, em seguida, o grupo de instâncias reiniciará a alocação de nós. A modificação de um grupo de instâncias instrui EMR a Amazon a tentar provisionar nós novamente. Os nós em execução não são reiniciados ou encerrados.

No AWS CLI, o `list-instances` subcomando retorna todas as instâncias e seus estados, assim como o `describe-cluster` subcomando. Se a Amazon EMR detectar uma falha em um grupo de instâncias, ela mudará o estado do grupo paraSUSPENDED.

Para redefinir um cluster em um SUSPENDED estado com o AWS CLI

Digite o subcomando `describe-cluster` com o parâmetro `--cluster-id` para visualizar o estado das instâncias no cluster.

- Para ver informações sobre todas as instâncias e grupos de instâncias em um cluster, digite o comando a seguir e substitua *j-3KVXXXXXXY7UG* com o ID do cluster.

  ```
  aws emr describe-cluster --cluster-id j-3KVXXXXXXY7UG
  ```

  A saída exibe informações sobre os grupos de instâncias e o estado das instâncias:

```
{
    "Cluster": {
        "Status": {
            "Timeline": {
                "ReadyDateTime": 1413187781.245,
                "CreationDateTime": 1413187405.356
            },
            "State": "WAITING",
            "StateChangeReason": {
                "Message": "Waiting after step completed"
            }
        },
        "Ec2InstanceAttributes": {
            "Ec2AvailabilityZone": "us-west-2b"
        },
        "Name": "Development Cluster",
        "Tags": [],
        "TerminationProtected": false,
        "RunningAmiVersion": "3.2.1",
        "NormalizedInstanceHours": 16,
        "InstanceGroups": [
            {
                "RequestedInstanceCount": 1,
                "Status": {
                    "Timeline": {
                        "ReadyDateTime": 1413187775.749,
                        "CreationDateTime": 1413187405.357
                    },
                    "State": "RUNNING",
                    "StateChangeReason": {
                        "Message": ""
                    }
                },
                "Name": "MASTER",
                "InstanceGroupType": "MASTER",
                "InstanceType": "m5.xlarge",
                "Id": "ig-3ETXXXXXXFYV8",
                "Market": "ON_DEMAND",
                "RunningInstanceCount": 1
            },
            {
                "RequestedInstanceCount": 1,
                "Status": {
```

```
                    "Timeline": {
                        "ReadyDateTime": 1413187781.301,
                        "CreationDateTime": 1413187405.357
                    },
                    "State": "RUNNING",
                    "StateChangeReason": {
                        "Message": ""
                    }
                },
                "Name": "CORE",
                "InstanceGroupType": "CORE",
                "InstanceType": "m5.xlarge",
                "Id": "ig-3SUXXXXXQ9ZM",
                "Market": "ON_DEMAND",
                "RunningInstanceCount": 1
            }
...
}
```

Para exibir as informações sobre um grupo de instâncias específico, digite o subcomando `list-instances` com os parâmetros `--cluster-id` e `--instance-group-types`. Você pode visualizar as informações para grupos primários, centrais ou de tarefa.

```
aws emr list-instances --cluster-id j-3KVXXXXXXY7UG --instance-group-types "CORE"
```

Use o subcomando `modify-instance-groups` com o parâmetro `--instance-groups` para redefinir um cluster no estado SUSPENDED. O ID do grupo de instâncias é obtido pelo subcomando `describe-cluster`.

```
aws emr modify-instance-groups --instance-groups
 InstanceGroupId=ig-3SUXXXXXQ9ZM,InstanceCount=3
```

Considerações ao reduzir o tamanho do cluster

Se você optar por reduzir o tamanho de um cluster em execução, considere o seguinte EMR comportamento e as melhores práticas da Amazon:

- Para reduzir o impacto nas tarefas em andamento, a Amazon seleciona de EMR forma inteligente as instâncias a serem removidas. Para obter mais informações sobre o comportamento de redução

da escala de clusters, consulte Terminar na conclusão de tarefas o Amazon EMR Management Guide.

- Quando você reduz o tamanho de um cluster, a Amazon EMR copia os dados das instâncias que ela remove para as instâncias que permanecem. Verifique se há capacidade de armazenamento suficiente para esses dados nas instâncias que permanecem no grupo.
- A Amazon EMR tenta HDFS descomissionar as instâncias do grupo. Antes de reduzir o tamanho de um cluster, recomendamos que você minimize a E/S de HDFS gravação.
- Para obter o controle mais granular ao reduzir o tamanho de um cluster, é possível visualizar o cluster no console e navegar até a guia Instâncias. Selecione o ID do grupo de instâncias que você deseja redimensionar. Em seguida, use a opção Terminar para as instâncias específicas que você deseja remover.

Configurar os tempos limite para a capacidade de provisionamento

Ao usar frotas de instâncias, é possível configurar os tempos limite de provisionamento. Um tempo limite de provisionamento instrui a Amazon EMR a interromper o provisionamento da capacidade da instância se o cluster exceder um limite de tempo especificado durante a inicialização do cluster ou as operações de escalabilidade do cluster. Os tópicos a seguir abordam como configurar um tempo limite de provisionamento para a inicialização do cluster e para operações de aumento da escala verticalmente do cluster.

Tópicos

- Configure os tempos limite de provisionamento para o lançamento do cluster na Amazon EMR
- Personalize um período de tempo limite de provisionamento para redimensionamento do cluster na Amazon EMR

Configure os tempos limite de provisionamento para o lançamento do cluster na Amazon EMR

Você pode definir um período de tempo limite para provisionar instâncias spot para cada frota do cluster. Se a Amazon não EMR puder provisionar a capacidade spot, você poderá optar por encerrar o cluster ou provisionar a capacidade sob demanda. Se o período de tempo limite terminar durante o processo de redimensionamento do cluster, a Amazon EMR cancelará solicitações spot não provisionadas. As instâncias spot que não foram provisionadas não são transferidas para a capacidade sob demanda.

Execute as etapas a seguir para personalizar um período de tempo limite de provisionamento para o lançamento do cluster com o console da Amazon. EMR

Console

Para configurar o tempo limite de provisionamento ao criar um cluster com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Na página Criar cluster, navegue até Configuração do cluster e selecione Frotas de instâncias.
4. Em Opção de ajuste de escala e provisionamento de clusters, especifique o tamanho do spot para suas frotas centrais e de tarefa.
5. Em Configuração de tempo limite spot, selecione Terminar cluster após o tempo limite spot ou Alternar para sob demanda após tempo limite spot. Em seguida, especifique o período de tempo limite para provisionamento de instâncias spot. O valor padrão é uma hora.
6. Escolha qualquer outra opção que se aplique ao cluster.
7. Para iniciar o cluster com o tempo limite configurado, escolha Criar cluster.

AWS CLI

Especificar um tempo limite de provisionamento com o comando **`create-cluster`**

```
aws emr create-cluster \
--release-label emr-5.35.0 \
--service-role EMR_DefaultRole \
--ec2-attributes '{"InstanceProfile":"EMR_EC2_DefaultRole","SubnetIds":["subnet-
XXXXX"]}' \
--instance-fleets
 '[{"InstanceFleetType":"MASTER","TargetOnDemandCapacity":1,"TargetSpotCapacity":0,"LaunchSp
{"OnDemandSpecification":{"AllocationStrategy":"lowest-
price"}},"InstanceTypeConfigs":[{"WeightedCapacity":1,"EbsConfiguration":
{"EbsBlockDeviceConfigs":[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemand
 - 1"},
{"InstanceFleetType":"CORE","TargetOnDemandCapacity":1,"TargetSpotCapacity":1,"LaunchSpecifi
{"SpotSpecification":
```

```
{"TimeoutDurationMinutes":120,"TimeoutAction":"SWITCH_TO_ON_DEMAND"},"OnDemandSpecification"
{"AllocationStrategy":"lowest-price"}},"InstanceTypeConfigs":
[{"WeightedCapacity":1,"EbsConfiguration":{"EbsBlockDeviceConfigs":
[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemand
 - 2"}]'
```

Personalize um período de tempo limite de provisionamento para redimensionamento do cluster na Amazon EMR

É possível definir um período de tempo limite para provisionar instâncias spot para cada frota do cluster. Se a Amazon não EMR puder provisionar a capacidade spot, ela cancelará a solicitação de redimensionamento e interromperá suas tentativas de provisionar capacidade spot adicional. Ao criar um cluster, é possível configurar o tempo limite. Em um cluster em execução, é possível adicionar ou atualizar um tempo limite.

Quando o período de tempo limite expira, a Amazon envia EMR automaticamente os eventos para um stream do Amazon CloudWatch Events. Com CloudWatch, você pode criar regras que correspondam aos eventos de acordo com um padrão especificado e, em seguida, rotear os eventos aos alvos para que sejam executadas ações. Por exemplo, é possível configurar uma regra para enviar uma notificação por e-mail. Para obter mais informações sobre como criar regras, consulte Criação de regras para EMR eventos da Amazon com CloudWatch. Para obter mais informações sobre diferentes detalhes de evento, consulte Eventos de alteração de estado da frota de instâncias.

Exemplos de tempos limite de provisionamento para redimensionamento de clusters

Especifique um tempo limite de provisionamento para redimensionar usando a AWS CLI

O exemplo a seguir usa o comando `create-cluster` para adicionar um tempo limite de provisionamento para redimensionamento.

```
aws emr create-cluster \
--release-label emr-5.35.0 \
--service-role EMR_DefaultRole \
--ec2-attributes '{"InstanceProfile":"EMR_EC2_DefaultRole","SubnetIds":["subnet-
XXXXX"]}' \
--instance-fleets
 '[{"InstanceFleetType":"MASTER","TargetOnDemandCapacity":1,"TargetSpotCapacity":0,"InstanceTyp
[{"WeightedCapacity":1,"EbsConfiguration":{"EbsBlockDeviceConfigs":
[{"VolumeSpecification":
```

```
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemandPri
 - 1"},
{"InstanceFleetType":"CORE","TargetOnDemandCapacity":1,"TargetSpotCapacity":1,"LaunchSpecificat
{"SpotSpecification":
{"TimeoutDurationMinutes":120,"TimeoutAction":"SWITCH_TO_ON_DEMAND"},"OnDemandSpecification":
{"AllocationStrategy":"lowest-price"}},"ResizeSpecifications":
{"SpotResizeSpecification":{"TimeoutDurationMinutes":20},"OnDemandResizeSpecification":
{"TimeoutDurationMinutes":25}},"InstanceTypeConfigs":
[{"WeightedCapacity":1,"EbsConfiguration":{"EbsBlockDeviceConfigs":
[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemandPri
 - 2"}]'
```

O exemplo a seguir usa o comando `modify-instance-fleet` para adicionar um tempo limite de provisionamento para redimensionamento.

```
aws emr modify-instance-fleet \
--cluster-id j-XXXXXXXXXXXX \
--instance-fleet '{"InstanceFleetId":"if-XXXXXXXXXXX","ResizeSpecifications":
{"SpotResizeSpecification":{"TimeoutDurationMinutes":30},"OnDemandResizeSpecification":
{"TimeoutDurationMinutes":60}}}' \
--region us-east-1
```

O exemplo a seguir usa `add-instance-fleet-command` para adicionar um tempo limite de provisionamento para redimensionamento.

```
aws emr add-instance-fleet \
--cluster-id j-XXXXXXXXXXXX \
--instance-fleet
 '{"InstanceFleetType":"TASK","TargetOnDemandCapacity":1,"TargetSpotCapacity":0,"InstanceTypeCo
[{"WeightedCapacity":1,"EbsConfiguration":{"EbsBlockDeviceConfigs":
[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemandPri
{"SpotResizeSpecification":{"TimeoutDurationMinutes":30},"OnDemandResizeSpecification":
{"TimeoutDurationMinutes":35}}}' \
--region us-east-1
```

Especifique um tempo limite de provisionamento para redimensionar e iniciar com o AWS CLI

O exemplo a seguir usa o comando `create-cluster` para adicionar um tempo limite de provisionamento para redimensionamento e inicialização.

```
aws emr create-cluster \
--release-label emr-5.35.0 \
--service-role EMR_DefaultRole \
--ec2-attributes '{"InstanceProfile":"EMR_EC2_DefaultRole","SubnetIds":["subnet-
XXXXX"]}' \
--instance-fleets
 '[{"InstanceFleetType":"MASTER","TargetOnDemandCapacity":1,"TargetSpotCapacity":0,"LaunchSpeci
{"OnDemandSpecification":{"AllocationStrategy":"lowest-price"}},"InstanceTypeConfigs":
[{"WeightedCapacity":1,"EbsConfiguration":{"EbsBlockDeviceConfigs":
[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemandPri
 - 1"},
{"InstanceFleetType":"CORE","TargetOnDemandCapacity":1,"TargetSpotCapacity":1,"LaunchSpecificat
{"SpotSpecification":
{"TimeoutDurationMinutes":120,"TimeoutAction":"SWITCH_TO_ON_DEMAND"},"OnDemandSpecification":
{"AllocationStrategy":"lowest-price"}},"ResizeSpecifications":
{"SpotResizeSpecification":{"TimeoutDurationMinutes":20},"OnDemandResizeSpecification":
{"TimeoutDurationMinutes":25}},"InstanceTypeConfigs":
[{"WeightedCapacity":1,"EbsConfiguration":{"EbsBlockDeviceConfigs":
[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemandPri
 - 2"}]'
```

Considerações para redimensionamento dos tempos limite de provisionamento

Ao configurar os tempos limite de provisionamento de clusters para suas frotas de instâncias, leve em consideração os comportamentos a seguir.

- É possível configurar os tempos limite de provisionamento para instâncias spot e sob demanda. O tempo limite mínimo de provisionamento é de cinco minutos. O tempo limite máximo de provisionamento é de sete dias.
- Você só pode configurar tempos limite de provisionamento para um EMR cluster que usa frotas de instâncias. É necessário configurar cada frota central e de tarefa separadamente.
- Ao criar um cluster, você pode configurar os tempos limite de provisionamento. É possível adicionar um tempo limite ou atualizar um tempo limite atual para um cluster em execução.
- Se você enviar várias operações de redimensionamento, a Amazon EMR rastreará os tempos limite de provisionamento para cada operação de redimensionamento. Por exemplo, defina o tempo limite de provisionamento em um cluster como *60* minutos. Em seguida, envie uma operação de redimensionamento *R1* na hora *T1*. Enviar uma segunda operação de

redimensionamento *R2* na hora *T2*. O tempo limite de provisionamento para R1 expira em *T1 + 60 minutes*. O tempo limite de provisionamento para R2 expira em *T2 + 60 minutes*.

- Se você enviar uma nova operação de redimensionamento de expansão antes que o tempo limite expire, a Amazon EMR continuará sua tentativa de provisionar capacidade para seu cluster. EMR

Redução da escala verticalmente do cluster

> **Note**
>
> As opções de comportamento de redução de escala não são mais suportadas desde a EMR versão 5.10.0 da Amazon. Devido à introdução do faturamento por segundo na AmazonEC2, o comportamento padrão de redução de escala para EMR clusters da Amazon agora é encerrado na conclusão da tarefa.

Com as EMR versões 5.1.0 a 5.9.1 da Amazon, há duas opções para o comportamento de redução: encerrar no limite da hora da instância para o faturamento da Amazon ou encerrar na conclusão da tarefa. EC2 A partir da EMR versão 5.10.0 da Amazon, a configuração para rescisão no limite da hora da instância está obsoleta devido à introdução do faturamento por segundo na Amazon. EC2 Não recomendamos especificar o encerramento no limite de tempo de execução da instância em que a opção está disponível.

> **Warning**
>
> Se você usar o AWS CLI para emitir um `modify-instance-groups` com`EC2InstanceIdsToTerminate`, essas instâncias serão encerradas imediatamente, sem considerar essas configurações e independentemente do status dos aplicativos em execução nelas. O encerramento de uma instância dessa forma tem o risco de perda de dados e de comportamento imprevisível do cluster.

Quando a conclusão da tarefa é especificada, a Amazon EMR Deny lista e drena as tarefas dos nós antes de encerrar as instâncias da Amazon. EC2 Com nenhum dos comportamentos especificados, EMR a Amazon não encerra EC2 instâncias da Amazon em grupos de instâncias principais se isso puder levar à HDFS corrupção.

Terminar na conclusão de tarefas

A Amazon EMR permite que você reduza seu cluster sem afetar sua carga de trabalho. A Amazon EMR YARN descomissiona normalmente e outros daemons nos nós principais e de tarefas durante uma operação de redimensionamento sem perder dados ou interromper trabalhos. HDFS A Amazon EMR só reduz o tamanho do grupo de instâncias se o trabalho atribuído aos grupos tiver sido concluído e eles estiverem ociosos. Para o YARN NodeManager Graceful Decommission, você pode ajustar manualmente o tempo que um nó espera pelo descomissionamento.

Este tempo é definido usando uma propriedade na a classificação de configuração `YARN-site`. Usando a EMR versão 5.12.0 e superior da Amazon, especifique a `YARN.resourcemanager.nodemanager-graceful-decommission-timeout-secs` propriedade. Usando EMR versões anteriores da Amazon, especifique a `YARN.resourcemanager.decommissioning.timeout` propriedade.

Se ainda houver contêineres ou YARN aplicativos em execução quando o tempo limite de descomissionamento passar, o nó será forçado a ser descomissionado e YARN reprograma os contêineres afetados em outros nós. O valor padrão é de 3.600 segundos (uma hora). Você pode definir esse tempo limite com um valor arbitrariamente alto para forçar a redução amigável a esperar mais tempo. Para obter mais informações, consulte Graceful Decommission of YARN nodes na documentação do Apache Hadoop.

Grupos de nós de tarefa

A Amazon seleciona de EMR forma inteligente instâncias que não têm tarefas executadas em nenhuma etapa ou aplicativo e primeiro remove essas instâncias de um cluster. Se todas as instâncias do cluster estiverem em uso, a Amazon EMR espera que as tarefas sejam concluídas em uma instância antes de removê-la do cluster. O tempo de espera padrão é 1 hora. Esse valor pode ser alterado com a configuração `YARN.resourcemanager.decommissioning.timeout`. A Amazon usa EMR dinamicamente a nova configuração. Você pode definir isso como um número arbitrariamente grande para garantir que a Amazon EMR não encerre nenhuma tarefa enquanto reduz o tamanho do cluster.

Grupos de nós centrais

Nos nós principais, ambos YARN NodeManager e os HDFS DataNode daemons devem ser desativados para que o grupo de instâncias seja reduzido. PoisYARN, a redução gradual garante que um nó marcado para descomissionamento só seja transferido para o `DECOMMISSIONED` estado se não houver contêineres ou aplicativos pendentes ou incompletos. A desativação termina imediatamente se não há contêineres em execução no nó no início da desativação.

PoisHDFS, uma redução suave garante que a capacidade alvo de HDFS seja grande o suficiente para caber em todos os blocos existentes. Se a capacidade de destino não for grande o suficiente, somente uma quantidade parcial de instâncias principais será desativada para que os nós restantes possam lidar com os dados atuais que residem. HDFS Você deve garantir HDFS capacidade adicional para permitir mais descomissionamento. Tente também minimizar a E/S de gravação antes de tentar reduzir os grupos de instâncias. O excesso de E/S de gravação poderá atrasar a conclusão da operação de redimensionamento.

Outro limite é o fator de replicação padrão `dfs.replication` no `/etc/hadoop/conf/hdfs-site`. Ao criar um cluster, a Amazon EMR configura o valor com base no número de instâncias no cluster: 1 com 1 a 3 instâncias, 2 para clusters com 4 a 9 instâncias e 3 para clusters com mais de 10 instâncias.

Warning

1. `dfs.replication`Definir como 1 em clusters com menos de quatro nós pode levar à perda de HDFS dados se um único nó ficar inativo. É recomendável usar um cluster com pelo menos quatro nós centrais para workloads de produção.
2. A Amazon não EMR permitirá que os clusters escalem os nós principais abaixo`dfs.replication`. Por exemplo, se `dfs.replication = 2`, o número mínimo de nós central será 2.
3. Ao usar o Ajuste de Escala Gerenciado, o Auto Scaling ou optar por redimensionar manualmente o cluster, é recomendável definir `dfs.replication` como 2 ou mais.

A redução suave não permite que você reduza os nós principais abaixo do fator de HDFS replicação. Isso permite fechar arquivos devido HDFS à insuficiência de réplicas. Para contornar esse limite, diminua o fator de replicação e reinicie o daemon. NameNode

Configurar o comportamento de redução de EMR escala da Amazon

Note

A opção de comportamento de redução de encerramento na hora da instância não é mais compatível com a EMR versão 5.10.0 e superior da Amazon. As seguintes opções de comportamento de redução de escala só aparecem no EMR console da Amazon nas versões 5.1.0 a 5.9.1.

Você pode usar o AWS Management Console AWS CLI, o ou o Amazon EMR API para configurar o comportamento de redução ao criar um cluster.

Console

Para configurar o comportamento de redução com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EC2Em EMRAtivado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Na seção Opções de escalabilidade e provisionamento de clusters, escolha Usar escalabilidade automática personalizada. Em Políticas personalizadas de escalabilidade automática, escolha o botão de ação adicional para adicionar escala nas políticas. Recomendamos que você adicione as políticas Scale In e Scale Out. Adicionar apenas um conjunto de políticas significa que a Amazon EMR executará apenas o escalonamento unidirecional e você precisará executar manualmente as outras ações.
4. Escolha qualquer outra opção que se aplique ao cluster.
5. Para iniciar o cluster, escolha Criar cluster.

AWS CLI

Para configurar o comportamento de redução de escala com o AWS CLI

- Use a opção `--scale-down-behavior` para especificar `TERMINATE_AT_INSTANCE_HOUR` ou `TERMINATE_AT_TASK_COMPLETION`.

Terminar um cluster

Esta seção descreve os métodos de encerramento de um cluster. Para obter informações sobre como habilitar a proteção contra encerramento e encerrar clusters automaticamente, consulte Controle de término do cluster. Você pode encerrar clusters nos estados `STARTING`, `RUNNING`ou `WAITING`. Um cluster no estado `WAITING` deve ser encerrado ou ele será executado indefinidamente, gerando encargos para sua conta. Você pode encerrar um cluster que não sai do estado `STARTING` ou que não consegue concluir uma etapa.

Se quiser encerrar um cluster que possui proteção de encerramento, deve primeiro desativar essa proteção antes de encerrar o cluster. Os clusters podem ser encerrados usando o console AWS CLI, o ou programaticamente usando o. `TerminateJobFlows` API

Dependendo da configuração do cluster, pode levar de 5 a 20 minutos para que o cluster seja completamente encerrado e libere os recursos alocados, como EC2 instâncias.

> **Note**
>
> Você não pode reiniciar um cluster terminado, mas pode clonar um cluster terminado para reutilizar a configuração dele em um novo cluster. Para obter mais informações, consulte Clonar um cluster usando o console.

> **Important**
>
> A Amazon EMR usa a função EMR de serviço da Amazon e a `AWSServiceRoleForEMRCleanup` função para limpar recursos de cluster em sua conta que você não usa mais, como EC2 instâncias da Amazon. Você deve incluir ações nas políticas de perfil para excluir ou encerrar os recursos. Caso contrário, a Amazon não EMR poderá realizar essas ações de limpeza e você poderá incorrer em custos com recursos não utilizados que permanecem no cluster.

Encerrrar um cluster com o console

Você pode encerrar um ou mais clusters usando o EMR console da Amazon. As etapas para encerrar um cluster no console variam de acordo com o estado da proteção de encerramento, ou seja, se a proteção está ativada ou não. Para encerrar um cluster protegido, você deve primeiro desativar a proteção de encerramento.

Console

Para encerrar um cluster com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. Escolha Clusters e, em seguida, selecione o cluster que você deseja encerrar.

3. No menu suspenso Ações, escolha Terminar cluster para abrir o prompt Terminar cluster.
4. No prompt, escolha Terminar. Dependendo da configuração do cluster, o encerramento pode demorar de cinco a dez minutos. Para obter mais informações sobre como criar EMR clusters da Amazon, consulteTerminar um cluster.

Encerrar um cluster com a AWS CLI

Para encerrar um cluster desprotegido usando o AWS CLI

Para encerrar um cluster desprotegido usando o AWS CLI, use o `terminate-clusters` subcomando com o parâmetro --cluster-ids.

- Digite o comando a seguir para encerrar um único cluster e substituir *`j-3KVXXXXXXX7UG`* com seu ID de cluster.

  ```
  aws emr terminate-clusters --cluster-ids j-3KVXXXXXXX7UG
  ```

  Para encerrar vários clusters, digite o seguinte comando e substitua *`j-3KVXXXXXXX7UG`* e *`j-WJ2XXXXXX8EU`* com seu clusterIDs.

  ```
  aws emr terminate-clusters --cluster-ids j-3KVXXXXXXX7UG j-WJ2XXXXXX8EU
  ```

  Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consultehttps://docs.aws.amazon.com/cli/latest/reference/emr.

Para encerrar um cluster protegido usando o AWS CLI

Para encerrar um cluster protegido usando o AWS CLI, primeiro desative a proteção de encerramento usando o `modify-cluster-attributes` subcomando com o `--no-termination-protected` parâmetro. Em seguida, use o subcomando `terminate-clusters` com o parâmetro `--cluster-ids` para encerrá-lo.

1. Digite o comando a seguir para desativar a proteção de terminação e substituir *`j-3KVTXXXXXX7UG`* com seu ID de cluster.

   ```
   aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --no-termination-protected
   ```

2. Para encerrar o cluster, digite o seguinte comando e substitua *j-3KVXXXXXXX7UG* com seu ID de cluster.

```
aws emr terminate-clusters --cluster-ids j-3KVXXXXXXX7UG
```

Para encerrar vários clusters, digite o seguinte comando e substitua *j-3KVXXXXXXX7UG* e *j-WJ2XXXXXX8EU* com seu clusterIDs.

```
aws emr terminate-clusters --cluster-ids j-3KVXXXXXXX7UG j-WJ2XXXXXX8EU
```

Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consultehttps://docs.aws.amazon.com/cli/latest/reference/emr.

Encerrar um cluster com a API

A `TerminateJobFlows` operação encerra o processamento da etapa, carrega todos os dados de log da Amazon EC2 para o Amazon S3 (se configurado) e encerra o cluster Hadoop. Um cluster também é encerrado automaticamente se você definir `KeepJobAliveWhenNoSteps` como `False` em uma solicitação `RunJobFlows`.

Você pode usar essa ação para encerrar um único cluster ou uma lista de clusters por clusterIDs.

Para obter mais informações sobre os parâmetros de entrada exclusivos de`TerminateJobFlows`, consulte TerminateJobFlows. Para obter mais informações sobre os parâmetros genéricos na solicitação, consulte Common request parameters.

Clonar um cluster usando o console

Você pode usar o EMR console da Amazon para clonar um cluster, que faz uma cópia da configuração do cluster original para usar como base para um novo cluster.

Console

Para clonar um cluster com o console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em https://console.aws.amazon.com/emr.
2. EMREm EC2 Ativado, no painel de navegação esquerdo, escolha Clusters.

3. Clonar um cluster da lista de clusters

    a. Use as opções de pesquisa e de filtro para encontrar o cluster que você deseja clonar na visualização de lista.

    b. Marque a caixa de seleção à esquerda da linha do cluster que deseja clonar.

    c. A opção Clonar já estará disponível na parte superior da visualização da lista. Selecione Clonar para iniciar o processo de clonagem. Se o cluster tiver etapas configuradas, escolha Incluir etapas e Continuar para clonar as etapas junto com as outras configurações do cluster.

    d. Revise as configurações do novo cluster que foram copiadas do cluster clonado. Ajuste as configurações, se necessário. Quando a configuração do novo cluster estiver satisfatória, selecione Criar cluster para iniciar o novo cluster.

4. Clonar um cluster na página de detalhes do cluster

    a. Para navegar até a página de detalhes do cluster que você deseja clonar, selecione o ID do cluster na visualização da lista de clusters.

    b. Na parte superior da página de detalhes do cluster, selecione Clonar cluster no menu Ações para iniciar o processo de clonagem. Se o cluster tiver etapas configuradas, escolha Incluir etapas e Continuar para clonar as etapas junto com as outras configurações do cluster.

    c. Revise as configurações do novo cluster que foram copiadas do cluster clonado. Ajuste as configurações, se necessário. Quando a configuração do novo cluster estiver satisfatória, selecione Criar cluster para iniciar o novo cluster.

Automatizar clusters recorrentes usando o AWS Data Pipeline

AWS Data Pipeline é um serviço que automatiza a movimentação e a transformação dos dados. Você pode usá-lo para programar a movimentação de dados de entrada para o Amazon S3 e para programar a inicialização de clusters para processar dados. Por exemplo, considere o caso em que você tenha um servidor web gravando logs de tráfego. Se você quiser executar um cluster semanal para analisar os dados de tráfego, você pode usá-lo AWS Data Pipeline para programar esses clusters. AWS Data Pipeline é um fluxo de trabalho orientado por dados, de modo que uma tarefa (iniciar o cluster) pode depender de outra tarefa (mover os dados de entrada para o Amazon S3). Ele também tem uma funcionalidade de novas tentativas robusta.

Para obter mais informações sobre AWS Data Pipeline, consulte o Guia do AWS Data Pipeline desenvolvedor, especialmente os tutoriais sobre a Amazon: EMR

- Tutorial: Inicie um fluxo de EMR trabalho da Amazon
- Introdução: processe registros da web com AWS Data Pipeline a Amazon e EMR o Hive
- Tutorial: importação e exportação do Amazon DynamoDB usando AWS Data Pipeline

Solução de problemas de clusters

Um EMR cluster é executado em um ecossistema complexo que inclui software de código aberto, código de aplicativo personalizado e. Serviços da AWS Quando ocorre um problema com qualquer uma dessas partes, o cluster pode falhar ou levar mais tempo do que o esperado para conclusão. Os tópicos a seguir podem ajudar a identificar problemas com o cluster e como corrigi-los.

Tópicos

Ao desenvolver uma nova aplicação Hadoop, é recomendável habilitar a depuração e processar um subconjunto pequeno, mas representativo, de seus dados para testar a aplicação. Talvez você também queira executar o aplicativo step-by-step para testar cada etapa separadamente. Para ter mais informações, consulte Configurar registro em log e depuração do cluster e Etapa 5: testar o cluster passo a passo.

Que ferramentas estão disponíveis para a solução de problemas?

Para identificar e corrigir erros de cluster, use as ferramentas descritas nesta página. Talvez seja necessário inicializar algumas ferramentas ao iniciar o cluster. Outras ferramentas estão disponíveis para todos os clusters por padrão.

Tópicos

Exibir detalhes EMR do cluster

Você pode usar o AWS Management Console, AWS CLI, ou EMR API para recuperar informações detalhadas sobre um EMR cluster e a execução de trabalhos. Para obter mais informações sobre como usar o AWS Management Console e AWS CLI, consulteVisualizar o status e os detalhes do cluster.

Painel de detalhes do EMR console Amazon

Na lista de clusters no EMR console da Amazon, você pode ver informações de alto nível sobre o status de cada cluster em sua conta e. Região da AWS A lista exibe todos os clusters ativos e terminados que você iniciou nos últimos dois meses. Na lista Clusters, você pode selecionar um Name (Nome) de cluster para visualizar detalhes do cluster. Essas informações são organizadas em diferentes categorias para facilitar a navegação.

As interfaces do usuário da aplicação disponíveis na página de detalhes do cluster podem ser para solucionar problemas de cluster. Ele fornece o status dos YARN aplicativos e, para alguns, como os aplicativos Spark, você pode detalhar diferentes métricas e facetas, como tarefas, estágios e executores. Para obter mais informações, consulte Visualizar o histórico da aplicação. Esse recurso está disponível somente para as EMR versões 5.8.0 e superiores da Amazon.

Interface de linha de EMR comando da Amazon

Você pode localizar detalhes sobre um cluster usando o `--describe` argumento AWS CLI with the.

Amazon EMR API

Você pode localizar detalhes sobre um cluster API usando a `DescribeJobFlows` ação.

Exibir detalhes EMR do erro do cluster

Quando um EMR cluster termina com um erro, o `DescribeCluster` e `ListClusters` APIs retorna um código de erro e uma mensagem de erro. Para erros de cluster selecionados, a matriz de dados `ErrorDetail` pode ajudar a solucionar a falha.

Para obter uma lista de códigos de erro que incluam dados `ErrorDetail`, consulte Códigos de erro com ErrorDetail informações.

> **Note**
>
> Refinamos continuamente nossas mensagens de erro para você receber as informações mais recentes e pertinentes. Não é recomendável analisar o texto de `ErrorMessage` porque ele está sujeito a alterações.

Execute scripts e configure EMR processos da Amazon

Como parte do processo de solução de problemas, talvez seja útil executar scripts personalizados no cluster ou visualizar e configurar processos de cluster.

Visualizar e reiniciar processos da aplicação

Pode ser útil visualizar os processos em execução no cluster para diagnosticar possíveis problemas. Você pode interromper e reiniciar os processos do cluster conectando-se ao nó principal do cluster. Para obter mais informações, consulte Visualize e reinicie a Amazon EMR e os processos de aplicativos (daemons).

Execute comandos e scripts sem uma SSH conexão

Para executar um comando ou script em seu cluster como uma etapa, você pode usar as `script-runner.jar` ferramentas `command-runner.jar` ou sem estabelecer uma SSH conexão com o nó principal. Para obter mais informações, consulte Executar comandos e scripts em um EMR cluster da Amazon.

Exibir arquivos de log do

Tanto a Amazon EMR quanto o Hadoop geram arquivos de log à medida que o cluster é executado. Você pode acessar esses arquivos de log de várias ferramentas diferentes, dependendo da configuração especificada ao iniciar o cluster. Para obter mais informações, consulte Configurar registro em log e depuração do cluster.

Arquivos de log no nó principal

Cada cluster publica arquivos de logs no diretório /mnt/var/log/ do nó principal. Esses arquivos de log estão disponíveis apenas enquanto o cluster está em execução.

Arquivos de log arquivados no Amazon S3

Se você executar o cluster e especificar um caminho de log do Amazon S3, o cluster copiará os arquivos de log armazenados em /mnt/var/log/ no nó principal para o Amazon S3 em intervalos de cinco minutos. Isso garante que você terá acesso aos arquivos de log, mesmo depois que o cluster for encerrado. Como os arquivos são arquivados em intervalos de 5 minutos, os últimos minutos de um cluster repentinamente encerrado podem não estar disponíveis.

Monitore o desempenho EMR do cluster

EMRA Amazon fornece várias ferramentas para monitorar o desempenho do seu cluster.

Interfaces Web do Hadoop

Cada cluster publica um conjunto de interfaces Web no nó principal que contêm informações sobre o cluster. Você pode acessar essas páginas da Web usando um SSH túnel para conectá-las ao nó principal. Para obter mais informações, consulte Visualize interfaces web hospedadas em EMR clusters da Amazon.

CloudWatch métricas

Cada cluster reporta métricas para CloudWatch. CloudWatch é um serviço da web que rastreia métricas e que você pode usar para definir alarmes sobre essas métricas. Para obter mais informações, consulte Monitorando EMR métricas da Amazon com CloudWatch.

Visualize e reinicie a Amazon EMR e os processos de aplicativos (daemons)

Ao solucionar problemas em um cluster, você pode relacionar os processos em execução. Também pode ser útil interromper ou reiniciar processos. Por exemplo, você pode reiniciar os processos após alterar uma configuração ou observar um problema com um determinado processo após a análise de arquivos de log e mensagens de erro.

Há dois tipos de processos que são executados em um cluster: EMR processos da Amazon (por exemplo, controlador de instância e Log Pusher) e processos associados aos aplicativos instalados no cluster (por exemplo, hadoop-hdfs-namenode e). hadoop-yarn-resourcemanager

Para trabalhar com os processos diretamente em um cluster, primeiro é necessário conectar-se ao nó principal. Para obter mais informações, consulte Conectar-se a um cluster.

Visualizar processos em execução

O método que você usa para visualizar os processos em execução em um cluster difere de acordo com a EMR versão da Amazon que você usa.

EMR 5.30.0 and 6.0.0 and later

Example : Listar todos os processos em execução

O exemplo a seguir usa `systemctl` e especifica `--type` para visualizar todos os processos.

```
systemctl --type=service
```

Example : Listar processos específicos

O exemplo a seguir lista todos os processos com nomes que contenham `hadoop`.

```
systemctl --type=service | grep -i hadoop
```

Resultado do exemplo:

```
 hadoop-hdfs-namenode.service           loaded active running Hadoop namenode
 hadoop-httpfs.service                  loaded active running Hadoop httpfs
 hadoop-kms.service                     loaded active running Hadoop kms
 hadoop-mapreduce-historyserver.service loaded active running Hadoop historyserver
 hadoop-state-pusher.service            loaded active running Daemon process that
 processes and serves EMR metrics data.
 hadoop-yarn-proxyserver.service        loaded active running Hadoop proxyserver
 hadoop-yarn-resourcemanager.service    loaded active running Hadoop resourcemanager
 hadoop-yarn-timelineserver.service     loaded active running Hadoop timelineserver
```

Example : Ver um relatório de status detalhado de um processo específico

O exemplo a seguir exibe um relatório de status detalhado do serviço `hadoop-hdfs-namenode`.

```
sudo systemctl status hadoop-hdfs-namenode
```

Resultado do exemplo:

```
hadoop-hdfs-namenode.service - Hadoop namenode
```

```
   Loaded: loaded (/etc/systemd/system/hadoop-hdfs-namenode.service; enabled; vendor
 preset: disabled)
   Active: active (running) since Wed 2021-08-18 21:01:46 UTC; 26min ago
 Main PID: 9733 (java)
    Tasks: 0
   Memory: 1.1M
   CGroup: /system.slice/hadoop-hdfs-namenode.service
           # 9733 /etc/alternatives/jre/bin/java -Dproc_namenode -Xmx1843m -server -
XX:OnOutOfMemoryError=kill -9 %p ...

Aug 18 21:01:37 ip-172-31-20-123 systemd[1]: Starting Hadoop namenode...
Aug 18 21:01:37 ip-172-31-20-123 su[9715]: (to hdfs) root on none
Aug 18 21:01:37 ip-172-31-20-123 hadoop-hdfs-namenode[9683]: starting namenode,
 logging to /var/log/hadoop-hdfs/ha...out
Aug 18 21:01:46 ip-172-31-20-123 hadoop-hdfs-namenode[9683]: Started Hadoop
 namenode:[  OK  ]
Aug 18 21:01:46 ip-172-31-20-123 systemd[1]: Started Hadoop namenode.
Hint: Some lines were ellipsized, use -l to show in full.
```

EMR 4.x - 5.29.0

Example : Listar todos os processos em execução

O exemplo a seguir lista todos os processos que estão em execução.

```
initctl list
```

EMR 2.x - 3.x

Example : Listar todos os processos em execução

O exemplo a seguir lista todos os processos que estão em execução.

```
ls /etc/init.d/
```

Interromper e reiniciar processos

Depois de determinar quais processos estão em execução, você pode interrompê-los e reiniciá-los, se necessário.

EMR 5.30.0 and 6.0.0 and later

Example : Interromper um processo

O exemplo a seguir interrompe o processo `hadoop-hdfs-namenode`.

```
sudo systemctl stop hadoop-hdfs-namenode
```

Consulte o `status` para verificar se o processo foi interrompido.

```
sudo systemctl status hadoop-hdfs-namenode
```

Resultado do exemplo:

```
hadoop-hdfs-namenode.service - Hadoop namenode
  Loaded: loaded (/etc/systemd/system/hadoop-hdfs-namenode.service; enabled; vendor
 preset: disabled)
  Active: failed (Result: exit-code) since Wed 2021-08-18 21:37:50 UTC; 8s ago
Main PID: 9733 (code=exited, status=143)
```

Example : Iniciar um processo

O exemplo a seguir inicia o processo `hadoop-hdfs-namenode`.

```
sudo systemctl start hadoop-hdfs-namenode
```

Consulte o status para verificar se o processo está em execução.

```
sudo systemctl status hadoop-hdfs-namenode
```

Resultado do exemplo:

```
hadoop-hdfs-namenode.service - Hadoop namenode
   Loaded: loaded (/etc/systemd/system/hadoop-hdfs-namenode.service; enabled; vendor
 preset: disabled)
   Active: active (running) since Wed 2021-08-18 21:38:24 UTC; 2s ago
  Process: 13748 ExecStart=/etc/init.d/hadoop-hdfs-namenode start (code=exited,
 status=0/SUCCESS)
 Main PID: 13800 (java)
    Tasks: 0
   Memory: 1.1M
```

```
  CGroup: /system.slice/hadoop-hdfs-namenode.service
          # 13800 /etc/alternatives/jre/bin/java -Dproc_namenode -Xmx1843m -server
 -XX:OnOutOfMemoryError=kill -9 %p...
```

EMR 4.x - 5.29.0

Example : Interromper um processo em execução

O exemplo a seguir interrompe o serviço `hadoop-hdfs-namenode`.

```
sudo stop hadoop-hdfs-namenode
```

Example : Reiniciar um processo interrompido

O exemplo a seguir reinicia o serviço `hadoop-hdfs-namenode`. Você deve usar o comando `start` em vez de `restart`.

```
sudo start hadoop-hdfs-namenode
```

Example : Verificar o status do processo

O exemplo a seguir busca o status de `hadoop-hdfs-namenode`. Você pode usar o comando `status` para verificar se o processo foi interrompido ou iniciado.

```
sudo status hadoop-hdfs-namenode
```

EMR 2.x - 3.x

Example : Interromper um processo da aplicação

O exemplo a seguir interrompe o `hadoop-hdfs-namenode` serviço, que está associado à versão da Amazon EMR instalada no cluster.

```
sudo /etc/init.d/hadoop-hdfs-namenode stop
```

Example : Reiniciar um processo da aplicação

O exemplo de comando a seguir reinicia o processo `hadoop-hdfs-namenode`:

```
sudo /etc/init.d/hadoop-hdfs-namenode start
```

Example : Interrompa um EMR processo da Amazon

O exemplo a seguir interrompe um processo, como instance-controller, que não está associado à versão da Amazon EMR no cluster.

```
sudo /sbin/stop instance-controller
```

Example : Reinicie um EMR processo da Amazon

O exemplo a seguir reinicia um processo, como instance-controller, que não está associado à versão da Amazon EMR no cluster.

```
sudo /sbin/start instance-controller
```

> **Note**
>
> Os comandos `/sbin/start, stop` e `restart` são symlinks para `/sbin/intictl`. Para obter mais informações sobre `initctl`, consulte a página do manual do initctl digitando `man initctl` no prompt de comando.

Erros comuns na Amazon EMR

Às vezes, os clusters falham ou demoram para processar os dados. As seções a seguir listam alguns problemas comuns de cluster com sugestões para corrigi-los.

Tópicos

- AWS GovCloud Erros (Oeste dos EUA)
- Encontrar um cluster ausente

Códigos de erro com ErrorDetail informações

Quando um EMR cluster termina com um erro, o `DescribeCluster` e `ListClusters` APIs retorna um código de erro e uma mensagem de erro. Para alguns erros de cluster, a matriz de dados `ErrorDetail` pode ajudar a solucionar a falha.

Os erros que incluem uma matriz `ErrorDetail` fornecem os seguintes detalhes:

`ErrorCode`

Um código de erro exclusivo que pode ser usado para acesso programático.

`ErrorData`

Uma lista de identificadores em pares de chave-valor que podem ser usados para programação ou pesquisa manual. Para obter descrições dos valores de `ErrorData` que um código de erro inclui, consulte a página de solução de problemas do código de erro.

`ErrorMessage`

Descrição do erro com um link para mais informações na EMR documentação da Amazon.

> **Note**
>
> Não é recomendável analisar o texto de `ErrorMessage` porque está sujeito a alterações.

Códigos de erro por categoria

- Códigos de erro de falha de bootstrap
- Códigos de erro internos
- Códigos de erro de falha de validação

Códigos de erro de falha de bootstrap

As seções a seguir fornecem informações sobre solução de problemas para códigos de erro de falha de bootstrap.

Tópicos

BOOTSTRAP_FAILURE_PRIMARY_WITH_NON_ZERO_CODE

Visão geral

Quando um cluster é terminado com um erro `BOOTSTRAP_FAILURE_PRIMARY_WITH_NON_ZERO_CODE`, uma ação de bootstrap falhou na instância primária. Para obter mais informações sobre ações de bootstrap, consulte Criar ações de bootstrap para instalar softwares adicionais.

Resolução

Para resolver esse erro, revise os detalhes retornados no API erro, modifique seu script de ação de bootstrap e crie um novo cluster com a ação de bootstrap atualizada.

Para solucionar o problema do EMR cluster com falha, consulte as `ErrorDetail` informações retornadas do `DescribeCluster` e. `ListClusters` APIs Para obter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`primary-instance-id`

O ID da instância primária em que a ação de bootstrap falhou.

`bootstrap-action`

O número ordinal da ação de bootstrap com falha. Um script com um valor `bootstrap-action` de 1 é a primeira ação de bootstrap a ser executada na instância.

`return-code`

O código de retorno para a ação de bootstrap com falha.

`amazon-s3-path`

O local da ação de bootstrap com falha no Amazon S3.

`public-doc`

O público URL da documentação do código de erro.

Etapas a serem executadas

Execute as etapas a seguir para identificar e corrigir a causa raiz do erro de ação de bootstrap. Em seguida, inicie um novo cluster.

1. Analise os arquivos de log de ações de bootstrap no Amazon S3 para identificar a causa raiz da falha. Para saber mais sobre como visualizar os EMR registros da Amazon, consulteExibir arquivos de log do .
2. Se você ativou os logs do cluster ao criar a instância, consulte o log `stdout` para obter mais informações. Você encontra o log `stdout` da ação de bootstrap neste local do Amazon S3:

   ```
   s3://EXAMPLE-BUCKET/logs/Your_Cluster_Id/node/Primary_Instance_Id/bootstrap-actions/Failed_Bootstrap_Action_Number/stdout.gz
   ```

   Para obter mais informações sobre logs de clusters, consulte Configurar registro em log e depuração do cluster.
3. Para determinar a falha na ação de bootstrap, revise as exceções nos logs `stdout` e o valor `return-code` em `ErrorData`.
4. Use suas descobertas da etapa anterior para revisar a ação de bootstrap para que ela evite exceções ou consiga lidar com exceções normalmente quando elas ocorrerem.
5. Inicie um novo cluster com a ação de bootstrap atualizada.

BOOTSTRAP_ FAILURE _BAR_ _ _ DOWNLOAD FAILED PRIMARY

Visão geral

Um cluster é encerrado com o erro `BOOTSTRAP_FAILURE_BA_DOWNLOAD_FAILED_PRIMARY` quando a instância primária não consegue baixar um script de ação de bootstrap no local do Amazon S3 especificado. As possíveis causas incluem:

- O arquivo de script de ação de bootstrap não está no local especificado do Amazon S3.
- A função de serviço para EC2 instâncias da Amazon no cluster (também chamada de perfil de EC2 instância da Amazon EMR) não tem permissões para acessar o bucket do Amazon S3 onde

reside o script de ação de bootstrap. Para obter mais informações sobre perfis de serviço, consulte Função de serviço para EC2 instâncias de cluster (perfil de EC2 instância).

Para obter mais informações sobre ações de bootstrap, consulte Criar ações de bootstrap para instalar softwares adicionais.

Resolução

Para resolver esse erro, certifique-se de que a instância primária tem o devido acesso ao script de ação de bootstrap.

Para solucionar o problema do EMR cluster com falha, consulte as `ErrorDetail` informações retornadas do `DescribeCluster` e. `ListClusters` APIs Para obter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`primary-instance-id`

O ID da instância primária em que a ação de bootstrap falhou.

`bootstrap-action`

O número ordinal da ação de bootstrap com falha. Um script com um valor `bootstrap-action` de 1 é a primeira ação de bootstrap a ser executada na instância.

`amazon-s3-path`

O local da ação de bootstrap com falha no Amazon S3.

`public-doc`

O público URL da documentação do código de erro.

Etapas a serem executadas

Execute as etapas a seguir para identificar e corrigir a causa raiz do erro de ação de bootstrap. Em seguida, inicie um novo cluster.

Etapas de solução de problemas

1. Use o valor `amazon-s3-path` da matriz `ErrorData` para encontrar o script de ação de bootstrap relevante no Amazon S3.

2. Se você ativou os logs do cluster ao criar a instância, consulte o log `stdout` para obter mais informações. Você encontra o log `stdout` da ação de bootstrap neste local do Amazon S3:

   ```
   s3://EXAMPLE-BUCKET/logs/Your_Cluster_Id/node/Primary_Instance_Id/bootstrap-actions/Failed_Bootstrap_Action_Number/stdout.gz
   ```

   Para obter mais informações sobre logs de clusters, consulte Configurar registro em log e depuração do cluster.
3. Para determinar a falha na ação de bootstrap, revise as exceções nos logs `stdout` e o valor `return-code` em `ErrorData`.
4. Use suas descobertas da etapa anterior para revisar a ação de bootstrap para que ela evite exceções ou consiga lidar com exceções normalmente quando elas ocorrerem.
5. Inicie um novo cluster com a ação de bootstrap atualizada.

BOOTSTRAP_FAILURE_FILE_NOT_FOUND_PRIMARY

Visão geral

O erro `BOOTSTRAP_FAILURE_FILE_NOT_FOUND_PRIMARY` indica que a instância primária não consegue encontrar o script de ação de bootstrap que a instância acabou de baixar no bucket do Amazon S3 especificado.

Resolução

Para resolver esse erro, confirme se a instância primária tem o devido acesso ao script de ação de bootstrap.

Para solucionar o problema do EMR cluster com falha, consulte as `ErrorDetail` informações retornadas do `DescribeCluster` e. `ListClusters` APIs Para obter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`primary-instance-id`

O ID da instância primária em que a ação de bootstrap falhou.

`bootstrap-action`

O número ordinal da ação de bootstrap com falha. Um script com um valor `bootstrap-action` de 1 é a primeira ação de bootstrap a ser executada na instância.

`amazon-s3-path`

O local da ação de bootstrap com falha no Amazon S3.

`public-doc`

O público URL da documentação do código de erro.

Etapas a serem executadas

Execute as etapas a seguir para identificar e corrigir a causa raiz do erro de ação de bootstrap. Em seguida, inicie um novo cluster.

1. Para encontrar o script de ação de bootstrap relevante no Amazon S3, use o valor `amazon-s3-path` da matriz `ErrorData`.
2. Analise os arquivos de log de ações de bootstrap no Amazon S3 para identificar a causa raiz da falha. Para saber mais sobre como visualizar os EMR registros da Amazon, consulteExibir arquivos de log do .

   > **Note**
   >
   > Se você não ativou os logs do cluster, será necessário criar um novo cluster com as mesmas configurações e ações de bootstrap. Para verificar se os logs do cluster estão ativados, consulte Configurar registro em log e depuração do cluster.

3. Analise o logs `stdout` de suas ações de bootstrap e confirme se não há processos personalizados que excluam arquivos na pasta `/emr/instance-controller/lib/bootstrap-actions` em suas instâncias primárias. Você encontra o log `stdout` da ação de bootstrap neste local do Amazon S3:

   ```
   s3://EXAMPLE-BUCKET/logs/Your_Cluster_Id/node/Primary_Instance_Id/bootstrap-actions/Failed_Bootstrap_Action_Number/stdout.gz
   ```

4. Inicie um novo cluster com a ação de bootstrap atualizada.

Códigos de erro internos

As seções a seguir fornecem informações sobre solução de problemas para códigos de erro interno.

Tópicos

- INTERNAL_ ERROR _ _ EC2 INSUFFICIENT _ CAPACITY _AZ
- INTERNAL_ERROR_SPOT_PRICE_INCREASE_PRIMARY
- INTERNAL_ ERROR _ SPOT CAPACITY _NÃO_ _ PRIMARY

INTERNAL_ ERROR _ _ EC2 INSUFFICIENT _ CAPACITY _AZ

Visão geral

Um cluster é encerrado com um `INTERNAL_ERROR_EC2_INSUFFICIENT_CAPACITY_AZ` erro quando a zona de disponibilidade selecionada não tem capacidade suficiente para atender à sua solicitação de tipo de EC2 instância da Amazon. A sub-rede que você selecionou para um cluster determina a zona de disponibilidade. Para obter mais informações sobre sub-redes da AmazonEMR, consulte. Configurar redes

Resolução

Para resolver esse erro, modifique as configurações de tipo de instância e crie um novo cluster com a solicitação atualizada.

Para solucionar o problema do EMR cluster com falha, consulte as `ErrorDetail` informações retornadas do `DescribeCluster` e. `ListClusters` APIs Para obter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`instance-type`

O tipo de instância que está fora da capacidade.

`availability-zone`

A zona de disponibilidade para a qual sua sub-rede é resolvida.

`public-doc`

O público URL da documentação do código de erro.

Etapas a serem executadas

Execute estas etapas para identificar e corrigir a causa raiz do erro de configuração do cluster:

- Analise as práticas recomendadas para conexão de cluster. Veja Práticas recomendadas para configuração de clusters no Guia EMR de gerenciamento da Amazon.

- Solucionar os problemas de inicialização e revisar a configuração. Consulte Solucionar problemas de lançamento de instâncias no Guia do EC2 usuário da Amazon.
- Inicie um novo cluster com a configuração de cluster atualizada.

INTERNAL_ERROR_SPOT_PRICE_INCREASE_PRIMARY

Visão geral

Um cluster é encerrado com um `INTERNAL_ERROR_SPOT_PRICE_INCREASE_PRIMARY` erro quando a Amazon não EMR consegue atender à sua solicitação de instância spot para o nó primário porque as instâncias não estão disponíveis no seu preço spot máximo ou abaixo dele. Para obter mais informações, consulte Instâncias spot no Guia EC2 do usuário da Amazon.

Resolução

Para resolver esse erro, especifique os tipos de instância do cluster que estejam dentro da meta de preço ou aumente o limite de preço para o mesmo tipo de instância.

Para solucionar o problema do EMR cluster com falha, consulte as `ErrorDetail` informações retornadas do `DescribeCluster` e. `ListClusters` APIs Para obter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`primary-instance-id`

O ID da instância primária do cluster que falhou.

`instance-type`

O tipo de instância que está fora da capacidade.

`availability-zone`

A zona de disponibilidade em que a sub-rede reside.

`public-doc`

O público URL da documentação do código de erro.

Etapas a serem executadas

Execute as etapas a seguir para solucionar problemas da estratégia de configuração de cluster e iniciar um novo cluster:

1. Analise as melhores práticas para Amazon EC2 Spot Instances e analise sua estratégia de configuração de cluster. Para obter mais informações, consulte as melhores práticas para o EC2 Spot no Guia EC2 do usuário da Amazon Práticas recomendadas para configuração de clusters e.
2. Modifique as configurações de tipo de instância ou zona de disponibilidade e crie um novo cluster com a solicitação atualizada.
3. Se o problema persistir, use a capacidade sob demanda para a instância primária.

INTERNAL_ ERROR _ SPOT CAPACITY _NÃO_ _ PRIMARY

Visão geral

Um cluster é terminado com um erro `INTERNAL_ERROR_SPOT_NO_CAPACITY_PRIMARY` quando não há capacidade suficiente para atender a uma solicitação de instância spot para o nó primário. Para obter mais informações, consulte Instâncias spot no Guia EC2 do usuário da Amazon.

Resolução

Para resolver esse erro, especifique os tipos de instância do cluster que estejam dentro da meta de preço ou aumente o limite de preço para o mesmo tipo de instância.

Para solucionar o problema do EMR cluster com falha, consulte as `ErrorDetail` informações retornadas do `DescribeCluster` e. `ListClusters` APIs Para obter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`primary-instance-id`

O ID da instância primária do cluster que falhou.

`instance-type`

O tipo de instância que está fora da capacidade.

`availability-zone`

A zona de disponibilidade para a qual sua sub-rede é resolvida.

`public-doc`

O público URL da documentação do código de erro.

Etapas a serem executadas

Execute as etapas a seguir para solucionar problemas da estratégia de configuração de cluster e iniciar um novo cluster:

1. Analise as melhores práticas para Amazon EC2 Spot Instances e analise sua estratégia de configuração de cluster. Para obter mais informações, consulte as melhores práticas para o EC2 Spot no Guia EC2 do usuário da Amazon Práticas recomendadas para configuração de clusters e.
2. Modifique as configurações de tipo de instância e crie um novo cluster com a solicitação atualizada.
3. Se o problema persistir, use a capacidade sob demanda para a instância primária.

Códigos de erro de falha de validação

As seções a seguir fornecem informações sobre solução de problemas para códigos de erro de falha de validação.

Tópicos

- VALIDATION_ERROR_SUBNET_NOT_FROM_ONE_VPC
- VALIDATION_ERROR_SECURITY_GROUP_NOT_FROM_ONE_VPC
- VALIDATION_ERROR_INVALID_SSH_KEY_NAME
- VALIDATION_ERROR_INSTANCE_TYPE_NOT_SUPPORTED

VALIDATION_ERROR_SUBNET_NOT_FROM_ONE_VPC

Visão geral

Quando seu cluster e as sub-redes que você faz referência para seu cluster pertencem a diferentes nuvens privadas virtuais (VPCs), o cluster é encerrado com um erro. `VALIDATION_ERROR_SUBNET_NOT_FROM_ONE_VPC` Você pode iniciar clusters com a Amazon EMR com a configuração de frotas de instâncias entre sub-redes em um. VPC Para obter mais informações sobre frotas de instâncias, consulte Configurar frotas de instâncias o Amazon EMR Management Guide.

Resolução

Para resolver esse erro, use sub-redes que pertençam ao VPC mesmo cluster.

Para solucionar o problema do EMR cluster com falha, consulte as `ErrorDetail` informações retornadas do `DescribeCluster` e. `ListClusters` APIs Para obter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`vpc`

Para cada sub-rede: VPC par, o ID do ao VPC qual a sub-rede pertence.

`subnet`

Para cada sub-rede: VPC par, o ID da sub-rede.

`public-doc`

O público URL da documentação do código de erro.

Etapas a serem executadas

Realize as etapas a seguir para identificar e corrigir o erro:

1. Examine IDs as sub-redes listadas na `ErrorData` matriz e confirme se elas pertencem ao VPC local em que você deseja iniciar o EMR cluster.
2. Modifique as configurações de sub-rede. Você pode usar um dos métodos a seguir para encontrar todas as sub-redes públicas e privadas disponíveis em um. VPC

   - Navegue até o Amazon VPC Console. Escolha Sub-redes e liste todas as sub-redes que residem dentro do Região da AWS seu cluster. Para encontrar somente sub-redes públicas ou privadas, aplique o filtro de atribuição automática de endereço público IPv4. Para encontrar e selecionar sub-redes nas VPC que seu cluster usa, use a opção Filtrar por VPC. Para obter mais informações sobre como criar sub-redes, consulte Criar uma sub-rede no Guia do usuário da Amazon Virtual Private Cloud.
   - Use o AWS CLI para encontrar todas as sub-redes públicas e privadas disponíveis nas VPC que seu cluster usa. Para obter mais informações, consulte as sub-redes descritivasAPI. Para criar novas sub-redes em umVPC, consulte create-subnet. API
3. Inicie um novo cluster com sub-redes do mesmo VPC cluster.

VALIDATION_ERROR_SECURITY_GROUP_NOT_FROM_ONE_VPC

Visão geral

Quando seu cluster e os grupos de segurança que você atribui ao seu cluster pertencem a diferentes nuvens privadas virtuais (VPCs), o cluster é encerrado com um `VALIDATION_ERROR_SECURITY_GROUP_NOT_FROM_ONE_VPC` erro. Para obter mais informações sobre grupos de segurança, consulte Especificação de grupos de EMR segurança adicionais e gerenciados pela Amazon e Controle do tráfego de rede com grupos de segurança.

Resolução

Para resolver esse erro, use grupos de segurança que pertençam ao VPC mesmo cluster.

Para solucionar o problema do EMR cluster com falha, consulte as `ErrorDetail` informações retornadas do `DescribeCluster` e. `ListClusters` APIs Para obter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`vpc`

Para cada grupo de segurança: VPC par, o ID do ao VPC qual o grupo de segurança pertence.

`security-group`

Para cada grupo de segurança: VPC par, o ID do grupo de segurança.

`public-doc`

O público URL da documentação do código de erro.

Etapas a serem executadas

Realize as etapas a seguir para identificar e corrigir o erro:

1. Examine IDs os grupos de segurança listados na `ErrorData` matriz e confirme se eles pertencem ao VPC local em que você deseja iniciar o EMR cluster.
2. Navegue até o Amazon VPC Console. Escolha Grupos de segurança para listar todos os grupos de segurança da selecionada. Encontre os grupos de segurança do VPC mesmo cluster e modifique a configuração do grupo de segurança.
3. Inicie um novo cluster com grupos de segurança do VPC mesmo cluster.

VALIDATION_ERROR_INVALID_SSH_KEY_NAME

Visão geral

Um cluster é encerrado com um `VALIDATION_ERROR_INVALID_SSH_KEY_NAME` erro quando você usa um par de EC2 chaves da Amazon que não é válido SSH para a instância primária. O nome do par de chaves pode estar incorreto ou o par de chaves pode não existir na solicitação Região da AWS. Para obter mais informações sobre pares de chaves, consulte pares de EC2 chaves da Amazon e instâncias Linux no Guia EC2 do usuário da Amazon.

Resolução

Para resolver esse erro, crie um novo cluster com um nome de par de SSH chaves válido.

Para solucionar o problema do EMR cluster com falha, consulte as `ErrorDetail` informações retornadas do `DescribeCluster` e. `ListClusters` APIs Para obter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`ssh-key`

O nome do par de SSH chaves que você forneceu ao criar o cluster.

`public-doc`

O público URL da documentação do código de erro.

Etapas a serem executadas

Realize as etapas a seguir para identificar e corrigir o erro:

1. Verifique seu *`keypair`*Arquivo.pem e confirme se ele corresponde ao nome da SSH chave que você vê no console da AmazonEMR.
2. Navegue até o EC2 console da Amazon. Verifique se o nome da SSH chave que você usou está disponível no Região da AWS que seu cluster usa. Você pode encontrar seu Região da AWS próximo ID de conta na parte superior do AWS Management Console.
3. Inicie um novo cluster com um nome de SSH chave válido.

VALIDATION_ERROR_INSTANCE_TYPE_NOT_SUPPORTED

Visão geral

Um cluster é terminado com um erro `VALIDATION_ERROR_INSTANCE_TYPE_NOT_SUPPORTED` quando as Região da AWS e as zonas de disponibilidade do cluster não oferecem suporte ao tipo de instância especificado para um ou mais grupos de instâncias. A Amazon EMR pode oferecer suporte a um tipo de instância em uma zona de disponibilidade dentro de uma região, mas não em outra. A sub-rede selecionada para um cluster determina a zona de disponibilidade na região. Para obter uma lista dos tipos de instância e regiões que a Amazon EMR oferece suporte, consulteTipos de instâncias compatíveis.

Resolução

Para resolver esse erro, especifique os tipos de instância para seu cluster que a Amazon EMR suporta na região e na zona de disponibilidade em que você solicita o cluster.

Para solucionar o problema do EMR cluster com falha, consulte as `ErrorDetail` informações retornadas do `DescribeCluster` e. `ListClusters` APIs Para obter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`instance-types`

A lista de tipos de instância com suporte.

`availability-zones`

A lista de zonas de disponibilidade para a qual sua sub-rede é resolvida.

`public-doc`

O público URL da documentação do código de erro.

Etapas a serem executadas

Realize as etapas a seguir para identificar e corrigir o erro:

1. Use o AWS CLI para recuperar os tipos de instância disponíveis em uma zona de disponibilidade. Para fazer isso, você pode usar o `ec2 describe-instance-type-offerings` comando para filtrar os tipos de instância disponíveis por local (Região da AWS ou zona de disponibilidade). Por exemplo, o comando a seguir retorna os tipos de instância que são oferecidos na AZ especificada, *`us-east-2a`*.

```
aws ec2 describe-instance-type-offerings --location-type "availability-
zone" --filters Name=location,Values=us-east-2a --region us-east-2 --query
 "InstanceTypeOfferings[*].[InstanceType]" --output text | sort
```

Para saber mais sobre como descobrir os tipos de instância disponíveis, consulte Encontre um tipo de EC2 instância da Amazon.

2. Após determinar os tipos de instância que estão disponíveis na mesma região e zona de disponibilidade do cluster, escolha uma das seguintes resoluções para continuar:

   a. Crie um novo cluster e escolha uma sub-rede para o cluster que está em uma zona de disponibilidade em que o tipo de instância que você selecionou está disponível e é suportado pela AmazonEMR.

   b. Crie um novo cluster na mesma região e EC2 sub-rede da Amazon do cluster que falhou, mas com um tipo de instância compatível com esse local pela AmazonEMR.

Para obter uma lista dos tipos de instância e regiões que a Amazon EMR oferece suporte, consulteTipos de instâncias compatíveis. Para comparar as capacidades dos tipos de instância, consulte Tipos de EC2instância da Amazon.

Erros de recursos

Os seguintes erros são geralmente causados pela restrição de recursos no cluster.

Tópicos

- O cluster termina com NO_ _ LEFT e nós principais SLAVE _BY_ FAILED MASTER
- Não é possível replicar os blocos, só foi possível replicar para zero nós.
- EC2 QUOTA EXCEEDED
- Muitas falhas de busca
- O arquivo pode ser replicado somente para 0 nós em vez de 1
- Negar deny-listed
- Erros de controle de utilização
- Tipo de instância sem suporte
- EC2está fora da capacidade
- HDFSerro do fator de replicação

- HDFSerro de espaço insuficiente

O cluster termina com NO_ _ LEFT e nós principais SLAVE _BY_ FAILED MASTER

Geralmente, isso acontece porque a proteção contra encerramento está desabilitada, e todos os nós core excedem a capacidade de armazenamento em disco, conforme especificado por um limite de utilização máxima na classificação de configuração `yarn-site`, que corresponde ao arquivo `yarn-site.xml`. Esse valor é 90%, por padrão. Quando a utilização do disco de um nó principal excede o limite de utilização, o serviço de YARN NodeManager saúde relata o nó como. `UNHEALTHY` Enquanto estiver nesse estado, a Amazon EMR deny lista o nó e não aloca YARN contêineres para ele. Se o nó permanecer inoperante por 45 minutos, a Amazon EMR marcará a EC2 instância associada da Amazon para encerramento como`FAILED_BY_MASTER`. Quando todas as EC2 instâncias da Amazon associadas aos nós principais são marcadas para encerramento, o cluster é encerrado com o status `NO_SLAVE_LEFT` porque não há recursos para executar trabalhos.

Ultrapassar a utilização de disco em um nó core pode causar uma reação em cadeia. Se um único nó exceder o limite de utilização do disco por causa dissoHDFS, é provável que outros nós também estejam próximos do limite. O primeiro nó excede o limite de utilização do disco, então o Amazon EMR Deny o lista. Isso aumenta a carga de utilização do disco para os nós restantes, pois eles começam a replicar entre si os HDFS dados que perderam no nó da lista negada. Cada nó subsequentemente entra no status `UNHEALTHY` da mesma maneira, e o cluster por fim é encerrado.

Práticas recomendadas e orientações

Configurar o hardware do cluster com armazenamento adequado

Ao criar um cluster, certifique-se de que haja nós principais suficientes e que cada um tenha um armazenamento de instâncias e volumes EBS de armazenamento adequadosHDFS. Para obter mais informações, consulte Calculando a HDFS capacidade necessária de um cluster. Você também pode adicionar instâncias core aos grupos de instâncias existentes manualmente ou usando a escalabilidade automática. As novas instâncias têm a mesma configuração de armazenamento que outras instâncias no grupo de instâncias. Para obter mais informações, consulte Usar ajuste de escala de clusters.

Habilitar a proteção contra encerramento

Habilitar a proteção contra encerramento. Dessa forma, se um nó principal estiver listado como negado, você poderá se conectar à EC2 instância associada da Amazon usando SSH para solucionar problemas e recuperar dados. Se você ativar a proteção contra rescisão, saiba que

EMR a Amazon não substitui a EC2 instância da Amazon por uma nova instância. Para obter mais informações, consulte Usar a proteção contra término.

Crie um alarme para a MRUnhealthyNodes CloudWatch métrica

Essa métrica informa o número de nós com o status UNHEALTHY. É equivalente à YARN métrica`mapred.resourcemanager.NoOfUnhealthyNodes`. Você pode configurar uma notificação desse alarme para avisá-lo de nós não íntegros antes que o limite de 45 minutos seja atingido. Para obter mais informações, consulte Monitorando EMR métricas da Amazon com CloudWatch.

Ajustar as configurações com yarn-site

As configurações a seguir podem ser ajustadas de acordo com os requisitos do aplicativo. Por exemplo, talvez você queira aumentar o limite de utilização de disco onde um nó informa UNHEALTHY ao aumentar o valor de `yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-percentage`.

Você pode definir esses valores ao criar um cluster usando a classificação de configuração `yarn-site`. Para obter mais informações, consulte Configuração de aplicativos no Amazon EMR Release Guide. Você também pode se conectar às EC2 instâncias da Amazon associadas aos nós principais usando eSSH, em seguida, adicionar os valores `/etc/hadoop/conf.empty/yarn-site.xml` usando um editor de texto. Depois de fazer a alteração, você deve reiniciar hadoop-yarn-nodemanager conforme mostrado abaixo.

> **Important**
>
> Quando você reinicia o NodeManager serviço, os YARN contêineres ativos são eliminados, a menos que `yarn.nodemanager.recovery.enabled` esteja configurado para `true` usar a classificação de `yarn-site` configuração ao criar o cluster. Você também deve especificar o diretório no qual armazenar um estado de contêiner usando a propriedade `yarn.nodemanager.recovery.dir`.

```
sudo /sbin/stop hadoop-yarn-nodemanager
sudo /sbin/start hadoop-yarn-nodemanager
```

Para obter mais informações sobre as `yarn-site` propriedades atuais e os valores padrão, consulte as configurações YARN padrão na documentação do Apache Hadoop.

Propriedade	Valor padrão	Descrição
yarn.nodemanager. disk-health-checker.intervalo-ms	120000	A frequência (em segundos) em que o verificador de integridade do disco é executado.
yarn.nodemanager. disk-health-checker. min-healthy-disks	0.25	A fração mínima do número de discos que devem estar íntegros NodeManager para lançar novos contêineres. Isso corresponde tanto a yarn.nodemanager.local-dirs (por padrão, na Amazon) quanto a yarn.nodemanager.log-dirs (por padrão, que tem um link simbólico `/mnt/yarn` na EMR Amazon). `/var/log/hadoop-yarn/containers` `mnt/var/log/hadoop-yarn/containers` EMR
`yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-percentage`	90.0	A porcentagem máxima de utilização de espaço em disco permitido depois que um disco é marcado como inválido. Os valores variam de 0,0 a 100,0. Se o valor for maior ou igual a 100, NodeManager verificará se há um disco cheio. Isso se aplica a `yarn-nodemanager.local-dirs` e a `yarn.nodemanager.log-dirs` .

Propriedade	Valor padrão	Descrição
`yarn.nodemanager.disk-health-checker.min-free-space-per-disk-mb`	0	O espaço mínimo que deve estar disponível em um disco para que ele seja usado. Isso se aplica a `yarn-nodemanager.local-dirs` e a `yarn.nodemanager.log-dirs` .

Não é possível replicar os blocos, só foi possível replicar para zero nós.

O erro: “Não é possível replicar os blocos, só foi possível replicar para zero nós”. normalmente ocorre quando um cluster não tem HDFS armazenamento suficiente. Esse erro ocorre quando você gera mais dados no cluster do que os que podem ser armazenadosHDFS. Você vê esse erro somente enquanto o cluster está em execução, porque quando o trabalho termina, ele libera o HDFS espaço que estava usando.

A quantidade de HDFS espaço disponível para um cluster depende do número e do tipo de EC2 instâncias da Amazon que são usadas como nós principais. Os nós de tarefas não são usados para HDFS armazenamento. Todo o espaço em disco em cada EC2 instância da Amazon, incluindo os volumes EBS de armazenamento anexados, está disponível paraHDFS. Para obter mais informações sobre a quantidade de armazenamento local para cada tipo de EC2 instância, consulte Tipos e famílias de instâncias no Guia EC2 do usuário da Amazon.

O outro fator que pode afetar a quantidade de HDFS espaço disponível é o fator de replicação, que é o número de cópias de cada bloco de dados que são armazenadas HDFS para redundância. O fator de replicação aumenta de acordo com o número de nós no cluster: são 3 cópias de cada bloco de dados para um cluster com 10 ou mais nós, 2 cópias de cada bloco para um cluster com 4 a 9 nós e 1 cópia (sem redundância) para clusters com 3 ou menos nós. O HDFS espaço total disponível é dividido pelo fator de replicação. Em alguns casos, como aumentar o número de nós de 9 para 10, o aumento no fator de replicação pode, na verdade, fazer com que a quantidade de HDFS espaço disponível diminua.

Por exemplo, um cluster com dez nós principais do tipo m1.large teria 2833 GB de espaço disponível para HDFS ((10 nós X 850 GB por nó) /fator de replicação de 3).

Se seu cluster exceder a quantidade de espaço disponívelHDFS, você poderá adicionar nós principais adicionais ao seu cluster ou usar a compactação de dados para criar mais HDFS espaço. Se o seu cluster puder ser interrompido e reiniciado, você pode considerar o uso de nós principais de um tipo maior de EC2 instância da Amazon. Você também deve considerar um ajuste no fator de replicação. No entanto, esteja ciente de que diminuir o fator de replicação reduz a redundância dos HDFS dados e a capacidade do seu cluster de se recuperar de blocos perdidos ou corrompidos. HDFS

EC2 QUOTA EXCEEDED

Se uma mensagem `EC2 QUOTA EXCEEDED` for exibida, pode haver várias causas. Dependendo das diferenças na configuração, pode demorar entre 5 a 20 minutos para que clusters anteriores sejam encerrados totalmente e liberem os recursos alocados. Se você está recebendo um erro `EC2 QUOTA EXCEEDED` ao tentar iniciar um cluster, pode ser que os recursos de um cluster recém-encerrado ainda não tenham sido liberados. Essa mensagem também pode ser causada pelo redimensionamento de um grupo ou frota de instâncias para um tamanho de destino maior do que a cota de instâncias atual da conta. Isso pode acontecer manualmente ou automaticamente por meio de escalabilidade automática.

Considere as opções a seguir para resolver o problema:

- Siga as instruções descritas em AWS service quotas no Referência geral da Amazon Web Services para solicitar um aumento do limite de serviço. Para algunsAPIs, organizar um CloudWatch evento pode ser uma opção melhor do que aumentar os limites. Para obter mais detalhes, consulte Quando configurar eventos do EMR em CloudWatch.
- Se um ou mais clusters em execução não estiverem na capacidade, redimensione os grupos de instâncias ou reduza as capacidades de destino nas frotas de instâncias para os clusters em execução.
- Crie clusters com menos EC2 instâncias ou capacidade alvo reduzida.

Muitas falhas de busca

A presença de mensagens de erro "Too many fetch-failures (Excesso de falhas de busca)" ou "Error reading task output (Erro ao ler a saída da tarefa)" nas etapas ou em logs de tentativas de tarefas indica que a tarefa em execução está dependendo da saída de uma outra tarefa. Isso geralmente ocorre quando uma tarefa é colocada na fila de execução e necessita da saída de uma ou mais tarefas de mapeamento, e essa saída ainda não está disponível.

Há vários motivos pelos quais a saída pode não estar disponível:

- A tarefa de pré-requisito ainda está em processamento. Essa geralmente é uma tarefa de mapeamento.
- Os dados podem estar indisponíveis devido à conectividade de rede ruim, se os dados estiverem localizados em uma instância diferente.
- Se HDFS for usado para recuperar a saída, pode haver um problema comHDFS.

A causa mais comum deste erro é que a tarefa anterior ainda está em processamento. Isso é mais provável se os erros estão ocorrendo quando as tarefas de redução estão sendo executadas pela primeira vez. Você pode verificar se é esse o caso examinando o log do syslog para a etapa do cluster que está gerando o erro. Se o syslog mostra que ambas as tarefas de mapeamento e redução estão em andamento, isso indica que a fase de redução foi iniciada e, ao mesmo tempo, há tarefas de mapeamento que ainda não foram concluídas.

Um item a ser pesquisado nos logs é a porcentagem de andamento do mapeamento que vai até 100% e, em seguida, cai para um valor mais baixo. Quando a porcentagem está em 100%, isso não significa que todas as tarefas de mapeamento foram concluídas. Isto significa simplesmente que o Hadoop está executando todas as tarefas de mapeamento. Se esse valor voltar a ficar abaixo de 100%, isso significa que uma tarefa de mapeamento falhou e, dependendo da configuração, o Hadoop pode tentar reprogramar a tarefa. Se a porcentagem do mapa permanecer em 100% nos registros, observe as CloudWatch métricas, especificamente`RunningMapTasks`, para verificar se a tarefa do mapa ainda está sendo processada. Você também pode encontrar essas informações usando a interface da web do Hadoop no nó principal.

Se você está vendo esse problema, pode tentar várias ações:

- Inclua instruções na fase de redução para esperar mais antes de iniciar. Você pode fazer isso alterando a definição da configuração do Hadoop mapred.reduce.slowstart.completed.maps para um tempo maior. Para obter mais informações, consulte Criar ações de bootstrap para instalar softwares adicionais.
- Iguale a contagem de reducers com a capacidade total de reducers do cluster. Você pode fazer isso ajustando a definição de configuração do Hadoop mapred.reduce.tasks de acordo com o trabalho.
- Use um código de classe de combiner para minimizar o número de saídas que precisam ser obtidas.

- Verifique se não há problemas com o EC2 serviço da Amazon que estejam afetando o desempenho da rede do cluster. Você pode fazer isso usando o Painel de status dos serviços.
- Analise os recursos CPU e de memória das instâncias em seu cluster para garantir que o processamento de dados não esteja sobrecarregando os recursos dos seus nós. Para obter mais informações, consulte Configurar o hardware e as redes do cluster.
- Verifique a versão da Amazon Machine Image (AMI) usada em seu EMR cluster da Amazon. Se a versão estiver entre a 2.3.0 e a 2.4.4, ambas incluídas, atualize para uma versão mais recente. AMIas versões no intervalo especificado usam uma versão do Jetty que pode falhar em fornecer a saída da fase do mapa. O erro de busca ocorre quando os reducers não conseguem obter uma saída da fase de mapeamento.

  O Jetty é um HTTP servidor de código aberto usado para comunicações entre máquinas em um cluster Hadoop.

O arquivo pode ser replicado somente para 0 nós em vez de 1

Quando um arquivo é gravadoHDFS, ele é replicado em vários nós principais. Quando você vê esse erro, significa que o NameNode daemon não tem nenhuma DataNode instância disponível para gravar dados. HDFS Em outras palavras, a replicação de blocos não está sendo realizada. Esse erro pode ser causado por vários problemas:

- O HDFS sistema de arquivos pode ter ficado sem espaço. Esta é a causa mais provável.
- DataNode as instâncias podem não estar disponíveis quando o trabalho foi executado.
- DataNode as instâncias podem ter sido bloqueadas de se comunicar com o nó principal.
- As instâncias no grupo de instâncias core podem não estar disponíveis.
- Podem estar faltando permissões. Por exemplo, o JobTracker daemon pode não ter permissões para criar informações do rastreador de tarefas.
- A configuração do espaço reservado para uma DataNode instância pode ser insuficiente. Verifique se esse é o caso, examinando a definição da configuração de dfs.datanode.du.reserved.

Para verificar se esse problema é causado pela HDFS falta de espaço em disco, veja a `HDFSUtilization` métrica em CloudWatch. Se o valor for muito alto, você pode adicionar mais nós core ao cluster. Se você tem um cluster que acha que pode ficar sem espaço em HDFS disco, você pode configurar um alarme CloudWatch para alertá-lo quando o valor de `HDFSUtilization` subir

acima de um determinado nível. Para ter mais informações, consulte Redimensionar manualmente um cluster em execução e Monitorando EMR métricas da Amazon com CloudWatch.

Se HDFS a falta de espaço não for o problema, verifique os registros, os DataNode NameNode registros e a conectividade de rede em busca de outros problemas que poderiam ter impedido HDFS a replicação dos dados. Para obter mais informações, consulte Exibir arquivos de log do .

Negar deny-listed

O NodeManager daemon é responsável por lançar e gerenciar contêineres nos nós principais e de tarefas. Os contêineres são alocados ao NodeManager daemon pelo ResourceManager daemon executado no nó principal. Ele ResourceManager monitora o NodeManager nó por meio de um batimento cardíaco.

Há algumas situações em que o ResourceManager daemon deny lista uma NodeManager, removendo-a do pool de nós disponíveis para processar tarefas:

- Se o não NodeManager tiver enviado uma pulsação ao ResourceManager daemon nos últimos 10 minutos (600.000 milissegundos). Esse intervalo de tempo pode ser configurado usando a definição da configuração `yarn.nm.liveness-monitor.expiry-interval-ms`. Para obter mais informações sobre como alterar as configurações do Yarn, consulte Como configurar aplicativos no Amazon EMR Release Guide.
- NodeManager verifica a integridade dos discos determinada por `yarn.nodemanager.local-dirs` e. `yarn.nodemanager.log-dirs` As verificações incluem permissões e espaço livre em disco (< 90%). Se um disco falhar na verificação, ele para de NodeManager usar esse disco específico, mas ainda informa o status do nó como íntegro. Se vários discos falharem na verificação, o nó será reportado como não íntegro ResourceManager e os novos contêineres não serão atribuídos ao nó.

O mestre do aplicativo também pode negar a lista de um NodeManager nó se ele tiver mais de três tarefas com falha. Você pode aumentar esse valor usando o parâmetro de configuração `mapreduce.job.maxtaskfailures.per.tracker`. Outras definições de configuração que você pode alterar controlam o número de tentativas para uma tarefa antes de marcá-la como falha: `mapreduce.map.max.attempts` para tarefas de mapeamento e `mapreduce.reduce.maxattempts` para tarefas de redução. Para obter mais informações sobre como alterar as configurações, consulte Como configurar aplicativos no Amazon EMR Release Guide.

Erros de controle de utilização

Os erros “Throttled from *Amazon EC2* ao iniciar o cluster” e “Falha ao provisionar instâncias devido à limitação de *Amazon EC2*“ocorre quando a Amazon EMR não consegue concluir uma solicitação porque outro serviço limitou a atividade. A Amazon EC2 é a fonte mais comum de erros de limitação, mas outros serviços podem ser a causa desses erros. AWS os limites de serviço se aplicam por região para melhorar o desempenho, e um erro de limitação indica que você excedeu o limite de serviço da sua conta nessa região.

Possíveis causas

A fonte mais comum de erros de EC2 limitação da Amazon é um grande número de instâncias de cluster iniciadas para que seu limite de serviço para EC2 instâncias seja excedido. As instâncias do cluster podem ser executadas pelos seguintes motivos:

- Novos clusters são criados.
- Os clusters são redimensionados manualmente. Para obter mais informações, consulte Redimensionar manualmente um cluster em execução.
- Os grupos de instâncias em um cluster adicionam instâncias (expandem) como resultado de uma regra de escalabilidade automática. Para obter mais informações, consulte Noções básicas sobre as regras de ajuste de escala automático.
- As frotas de instâncias em um cluster adicionam instâncias para atender a uma maior capacidade de destino. Para obter mais informações, consulte Configurar frotas de instâncias.

Também é possível que a frequência ou o tipo de API solicitação feita à Amazon EC2 cause erros de limitação. Para obter mais informações sobre como a Amazon EC2 controla as API solicitações, consulte Taxa de API solicitação de consulta na Referência da Amazon EC2 API.

Soluções

Considere as seguintes soluções:

- Siga as instruções descritas em AWS service quotas no Referência geral da Amazon Web Services para solicitar um aumento do limite de serviço. Para algunsAPIs, organizar um CloudWatch evento pode ser uma opção melhor do que aumentar os limites. Para obter mais detalhes, consulte Quando configurar eventos do EMR em CloudWatch.
- Se você tiver clusters são executados no mesmo agendamento (por exemplo, no começo da hora) considere intercalar os horários de início.

- Se tiver clusters que são dimensionados para picos de demanda, e você periodicamente tiver capacidade de instância, considere especificar a escalabilidade automática para adicionar e remover instâncias sob demanda. Dessa forma, as instâncias serão usadas de forma mais eficiente e, dependendo do perfil de demanda, menos instâncias poderão ser solicitadas em um determinado momento em uma conta. Para obter mais informações, consulte Usar o ajuste de escala automático com uma política personalizada para grupos de instâncias.

Tipo de instância sem suporte

Se você criar um cluster e ele falhar com a mensagem de erro “O tipo de instância solicitada *InstanceType* não é suportado na zona de disponibilidade solicitada”, significa que você criou o cluster e especificou um tipo de instância para um ou mais grupos de instâncias que não é suportado pela Amazon EMR na região e na zona de disponibilidade em que o cluster foi criado. A Amazon EMR pode oferecer suporte a um tipo de instância em uma zona de disponibilidade dentro de uma região e não em outra. A sub-rede selecionada para um cluster determina a Zona de disponibilidade na região.

Solução

Determine os tipos de instância disponíveis em uma zona de disponibilidade usando o AWS CLI

- Use o comando `ec2 run-instances` com a opção `--dry-run`. No exemplo abaixo, substitua *m5.xlarge* com o tipo de instância que você deseja usar, *ami-035be7bafff33b6b6* com o AMI associado a esse tipo de instância, e *subnet-12ab3c45* com uma sub-rede na zona de disponibilidade que você deseja consultar.

  ```
  aws ec2 run-instances --instance-type m5.xlarge --dry-run --image-
  id ami-035be7bafff33b6b6 --subnet-id subnet-12ab3c45
  ```

  Para obter instruções sobre como encontrar um AMI ID, consulte Encontre um Linux AMI. Para encontrar um ID de sub-rede, você pode usar o comando describe-subnets.

Para saber mais sobre como descobrir os tipos de instância disponíveis, consulte Encontre um tipo de EC2 instância da Amazon.

Depois de determinar os tipos de instâncias disponíveis, você pode fazer o seguinte:

- Crie o cluster na mesma região e EC2 sub-rede e escolha um tipo de instância diferente com recursos semelhantes aos da sua escolha inicial. Para obter uma lista dos tipos de instâncias compatíveis, consulte Tipos de instâncias compatíveis. Para comparar as capacidades dos tipos de EC2 instância, consulte os tipos de EC2 instância da Amazon.
- Escolha uma sub-rede para o cluster em uma zona de disponibilidade em que o tipo de instância esteja disponível e seja suportado pela AmazonEMR.

Mitigue as falhas de lançamento de clusters da frota de instâncias devido a tipos de instâncias primárias não compatíveis na Amazon EMR

Os nós primários são essenciais nos EMR clusters da Amazon. A inicialização de um EMR cluster pode falhar com um `instance type not supported` erro em que a Amazon EMR tenta iniciar o cluster em uma zona de disponibilidade, caso o tipo de instância primária não seja suportado. A seleção aprimorada da zona de disponibilidade para clusters de frotas de instâncias na Amazon filtra EMR automaticamente os tipos de instâncias primárias que você especificou na configuração do cluster. AZs Isso significa que a Amazon EMR não escolherá uma zona de disponibilidade em que os tipos de instância primária configurados não sejam suportados, o que evita falhas na inicialização do cluster devido a tipos de instância não compatíveis.

Para permitir essa melhoria, adicione a permissão necessária à função ou política de serviço do seu cluster. A versão mais recente do `AmazonEMRServicePolicy_v2` inclui essa permissão, portanto, se você usar essa política, a melhoria já estará disponível. Se você usa uma função ou política de serviço personalizada, adicione a permissão `ec2:DescribeInstanceTypeOfferings` ao iniciar seu cluster.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": "ec2:DescribeInstanceTypeOfferings",
            "Effect": "Allow",
            "Resource": "*"
        }
    ]
}
```

EC2está fora da capacidade

Um "EC2está fora da capacidade para *`InstanceType`*“o erro ocorre quando você tenta criar um cluster ou adicionar instâncias a um cluster em uma zona de disponibilidade que não tem mais do tipo de EC2 instância especificado. A sub-rede que você selecionou para um cluster determina a zona de disponibilidade.

Para criar um cluster, siga um destes procedimentos:

- Especificar outro tipo de instância com recursos semelhantes
- Criar o cluster em outra região
- Selecione uma sub-rede em uma zona de disponibilidade em que o tipo de instância desejado possa estar disponível.

Para adicionar instância a um cluster em execução, realize uma destas ações:

- Modifique as configurações do grupo de instâncias ou as configurações da frota de instâncias para adicionar os tipos de instância disponíveis com recursos semelhantes. Para obter uma lista dos tipos de instâncias compatíveis, consulte Tipos de instâncias compatíveis. Para comparar as capacidades dos tipos de EC2 instância, consulte os tipos de EC2 instância da Amazon.
- Termine o cluster e o recrie em uma região e zona de disponibilidade em que o tipo de instância está disponível.

HDFSerro do fator de replicação

Quando você remove um nó principal de um grupo de instâncias principais ou de uma frota de instâncias, a Amazon EMR pode se deparar com um erro de HDFS replicação. Esse erro ocorre quando você remove os nós principais e o número de nós principais fica abaixo do fator de replicação dfs.replication configurado para o Hadoop Distributed File System (). HDFS Dessa forma, a Amazon não EMR pode realizar a operação com segurança. Para determinar o valor padrão da `dfs.replication` configuração, HDFSconfiguração.

Possíveis causas

Veja a seguir as possíveis causas do erro do fator de HDFS replicação:

- Se você redimensionar manualmente um grupo de instâncias principais ou uma frota de instâncias abaixo do `dfs.replication` fator configurado.

- Suas políticas de escalabilidade gerenciada ou escalonamento automático podem permitir que o escalonamento reduza o número de nós principais abaixo do limite de. `dfs.replication`
- Esse erro também pode ocorrer se a Amazon EMR tentar substituir um nó principal não íntegro quando um cluster tem o número mínimo de nós principais definido por `dfs.replication`.

Soluções e melhores práticas

Veja a seguir as soluções e as melhores práticas:

- Ao redimensionar manualmente um EMR cluster da Amazon, não diminua a escala, `dfs.replication` pois a Amazon não EMR pode concluir o redimensionamento com segurança.
- Ao usar o escalonamento gerenciado ou o escalonamento automático, certifique-se de que a capacidade mínima do seu cluster não seja inferior ao fator. `dfs.replication`
- O número de instâncias principais deve ser pelo menos `dfs.replication` mais uma. Isso garante que a Amazon EMR possa substituir com sucesso um nó principal não íntegro se você habilitar a substituição de núcleo não íntegro.

> **Important**
>
> A falha de um único nó central pode levar à perda de HDFS dados se você `dfs.replication` definir como 1. Se seu cluster tiver HDFS armazenamento, recomendamos que você configure o cluster com pelo menos quatro nós principais para cargas de trabalho de produção para evitar perda de dados e também definir um `dfs.replication` fator de pelo menos 2.

HDFSerro de espaço insuficiente

Um erro de espaço insuficiente do Hadoop Distributed File System (HDFS) pode ocorrer se você tentar remover um nó central, mas a Amazon não EMR pode concluir a operação com segurança devido à falta de espaço no. HDFS Antes que a Amazon EMR remova um nó principal, todos os HDFS dados no nó devem ser transferidos para outros nós principais para garantir a redundância dos dados. No entanto, se não houver espaço suficiente nos outros nós principais para replicação, a Amazon não EMR poderá descomissionar o nó normalmente.

Possíveis causas

Veja a seguir uma lista das possíveis causas do erro de espaço HDFS insuficiente:

- Se você reduzir manualmente um grupo de instâncias principais ou uma frota de instâncias quando não houver HDFS espaço suficiente nos nós restantes para a replicação de dados antes da redução.
- O escalonamento gerenciado ou o escalonamento automático reduzem um grupo principal de instâncias ou uma frota de instâncias quando não há HDFS espaço suficiente para a replicação de dados.
- A Amazon EMR tenta substituir um nó central não íntegro, mas não consegue substituí-lo com segurança devido ao HDFS espaço insuficiente.

Soluções e melhores práticas

Veja a seguir as soluções e as melhores práticas:

- Aumente o número de nós principais em seu EMR cluster da Amazon. Se você usa escalabilidade gerenciada ou escalonamento automático, aumente a capacidade mínima dos seus nós principais.
- Use EBS volumes maiores para seus nós principais ao criar seu EMR cluster.
- Exclua HDFS dados desnecessários em seu EMR cluster. Recomendamos que você configure CloudWatch alarmes para monitorar a `HDFSUtilization` métrica em seu cluster para saber se seu EMR cluster está com pouco espaço.

Erros de entrada e saída

Os erros a seguir são comuns em operações de entrada e saída do cluster.

Tópicos

O caminho para o Amazon Simple Storage Service (Amazon S3) com pelo menos três barras?

Ao especificar um bucket do Amazon S3, você deve incluir uma barra final no final do. URL Por exemplo, em vez de referenciar um bucket como “s3n: //amzn-s3-demo-bucket1", você deve usar “s3n: //amzn-s3-demo-bucket1/”, caso contrário, o Hadoop falhará em seu cluster na maioria dos casos.

Você está tentando, recursivamente, desviar diretórios de entrada?

O Hadoop não pesquisa recursivamente diretórios de entrada para arquivos. Se você tiver uma estrutura de diretório como /corpus/01/01.txt, /corpus/01/02.txt, /corpus/02/01.txt, etc. e especificar / corpus/ como o parâmetro de entrada para seu cluster, o Hadoop não localizará nenhum arquivo de entrada porque o diretório /corpus/está vazio, e o Hadoop não verificará o conteúdo dos subdiretórios. Da mesma forma, o Hadoop não verifica recursivamente os subdiretórios de buckets do Amazon S3.

Os arquivos de entrada devem estar diretamente no diretório de entrada ou no bucket do Amazon S3 que você especificar, e não nos subdiretórios.

Seu diretório de saída já existe?

Se você especificar um caminho de saída que já existe, seu cluster apresentará falha no Hadoop na maioria dos casos. Isso significa que, se você executar um cluster uma vez e, em seguida, executá-lo novamente com, exatamente, os mesmos parâmetros ele, provavelmente, funcionará na primeira vez e depois nunca mais. Após a primeira execução, o caminho de saída passa a existir e isso faz com que haja falha em todas as execuções sucessivas.

Você está tentando especificar um recurso usando um HTTPURL?

O Hadoop não aceita locais de recursos especificados usando o prefixo http://. Você não pode referenciar um recurso usando um HTTPURL. Por exemplo, passar http://mysite/myjar.jar como JAR parâmetro faz com que o cluster falhe.

Você está referenciando um bucket do Amazon S3 usando um nome de formato inválido?

Se você tentar usar um nome de bucket como “amzn-s3-demo-bucket1.1" com a AmazonEMR, seu cluster falhará porque a EMR Amazon exige que os nomes de bucket sejam 2396 nomes de host RFC válidos; o nome não pode terminar com um número. Além disso, devido aos requisitos do

Hadoop, os nomes de bucket do Amazon S3 usados com a EMR Amazon devem conter somente letras minúsculas, números, pontos (.) e hífens (-). Para obter informações sobre como formatar nomes de buckets do Amazon S3, consulte Restrições e limitações do bucket no guia do usuário do Amazon Simple Storage Service.

Você está tendo problemas para carregar dados para carregar ou descarregar dados do Amazon S3?

O Amazon S3 é a fonte de entrada e saída mais popular da Amazon. EMR Um erro comum é tratar o Amazon S3 como um sistema de arquivos típico. Há diferenças entre o Amazon S3 e um sistema de arquivos que você precisa levar em conta ao executar seu cluster.

- Se ocorrer um erro interno no Amazon S3, sua aplicação deverá lidar com isso normalmente e repetir a operação.
- Se as chamadas para o Amazon S3 levam muito tempo para retornar, talvez seja necessário reduzir a frequência com que a aplicação chama o Amazon S3.
- Listar todos os objetos em um bucket do Amazon S3 é uma chamada de alto custo. O aplicativo deve minimizar o número de vezes que faz isso.

Há várias maneiras de melhorar como seu cluster interage com o Amazon S3.

- Inicie seu cluster usando a versão mais recente da AmazonEMR.
- Use o S3 DistCp para mover objetos para dentro e para fora do Amazon S3. O S3 DistCp implementa tratamento de erros, novas tentativas e recuos para atender aos requisitos do Amazon S3. Para obter mais informações, consulte Cópia distribuída usando o S3 DistCp.
- Projete seu aplicativo com consistência eventual em mente. Use HDFS para armazenamento intermediário de dados enquanto o cluster está em execução e o Amazon S3 somente para inserir os dados iniciais e gerar os resultados finais.
- Se os seus clusters confirmarem 200 ou mais transações por segundo para o Amazon S3, entre em contato com o suporte para preparar seu bucket para mais transações por segundo e considere usar estratégias de partição de chave, descritas em Amazon S3 performance tips and tricks.
- Defina a configuração io.file.buffer.size do Hadoop como 65536. Isso faz com que o Hadoop gaste menos tempo procurando entre objetos do Amazon S3.
- Considere desabilitar o atributo de execução especulativa do Hadoop, se o cluster estiver enfrentando problemas de simultaneidade do Amazon S3. Isso também é útil quando você estiver

solucionando problemas de um cluster lento. Você pode fazer isso definindo as propriedades `mapreduce.reduce.speculative` e `mapreduce.map.speculative` como `false`. Ao executar um cluster, você pode definir esses valores usando a classificação de configuração `mapred-env`. Para obter mais informações, consulte Configurando aplicativos no Amazon EMR Release Guide.

- Se você estiver executando um cluster do Hive, consulte Você está tendo problemas para carregar ou descarregar dados do Amazon S3 no Hive?.

Para obter informações adicionais, consulte Práticas recomendadas com relação a erros do Amazon S3 no Guia do usuário do Amazon Simple Storage Service.

Erros de permissão

Os seguintes erros são comuns quando se utiliza permissões ou credenciais.

Tópicos

- Você está passando as credenciais corretas paraSSH?
- Se você estiver usandoIAM, você tem as EC2 políticas adequadas da Amazon definidas?

Você está passando as credenciais corretas paraSSH?

Se você não conseguir usar SSH para se conectar ao nó principal, provavelmente é um problema com suas credenciais de segurança.

Primeiro, verifique se o arquivo.pem que contém sua SSH chave tem as permissões adequadas. Você pode usar o chmod para alterar as permissões de seu arquivo .pem, como mostrado no exemplo a seguir, onde você deve substituir mykey.pem pelo nome do seu próprio arquivo .pem.

```
chmod og-rwx mykey.pem
```

A segunda possibilidade é você não estar usando o par de chaves especificado quando o cluster foi criado. Isso é fácil de acontecer se você tiver criado vários pares de chaves. Verifique os detalhes do cluster no EMR console da Amazon (ou use a `--describe` opção noCLI) para obter o nome do par de chaves que foi especificado quando o cluster foi criado.

Depois de verificar se você está usando o par de chaves correto e se as permissões estão definidas corretamente no arquivo.pem, você pode usar o seguinte comando para se conectar SSH ao nó principal, onde você substituiria mykey.pem pelo nome do seu arquivo.pem e `hadoop@ec2-01-001-001-1.compute-1.amazonaws.com` pelo DNS nome público do nó principal (disponível por meio da opção `--describe` no console da Amazon ou por meio dele). CLI EMR

> **Important**
>
> Você deve usar o nome de login `hadoop` ao se conectar a um nó de EMR cluster da Amazon, caso contrário, poderá ocorrer um `Server refused our key` erro semelhante ao erro.

```
ssh -i mykey.pem hadoop@ec2-01-001-001-1.compute-1.amazonaws.com
```

Para obter mais informações, consulte Conecte-se ao nó primário usando SSH.

Se você estiver usandoIAM, você tem as EC2 políticas adequadas da Amazon definidas?

Como a Amazon EMR usa EC2 instâncias como nós, os usuários da Amazon EMR também precisam ter determinadas EC2 políticas da Amazon definidas para que EMR a Amazon possa gerenciar essas instâncias em nome do usuário. Se você não tiver as permissões necessárias definidas, a Amazon EMR retornará o erro: “a conta não está autorizada a ligar”EC2.

Para obter mais informações sobre as EC2 políticas da Amazon que sua IAM conta precisa definir para administrar a AmazonEMR, consulteComo a Amazon EMR trabalha com IAM.

Erros de cluster do Hive

Geralmente, você pode encontrar a causa de um erro do Hive no arquivo `syslog`, que você vincula a partir do painel Steps (Etapas). Se você não conseguir determinar o problema lá, verifique a mensagem de erro de tentativa de tarefa do Hadoop. Vincule-se a ela no painel Task Attempts (Tentativas da tarefa).

Os erros a seguir são comuns em clusters do Hive.

Tópicos

Você está usando a versão mais recente do Hive?

A versão mais recente do Hive tem todos os patches e correções de erros atuais e pode resolver o problema.

Você encontrou um erro de sintaxe no script do Hive?

Se houver falha em uma etapa, examine o arquivo `stdout` de logs para a etapa que executou o script do Hive. Se o erro não estiver lá, examine o arquivo `syslog` dos logs das tentativas de tarefa que tiveram falha. Para obter mais informações, consulte Exibir arquivos de log do .

Houve falha em um trabalho quando executado interativamente?

Se você estiver executando o Hive interativamente no nó principal e houver falha no cluster, veja as entradas do `syslog` no log de tentativas de tarefa para a tentativa de tarefas com falha. Para obter mais informações, consulte Exibir arquivos de log do .

Você está tendo problemas para carregar ou descarregar dados do Amazon S3 no Hive?

Se você estiver com problemas para acessar dados no Amazon S3, verifique primeiro as possíveis causas listadas em Você está tendo problemas para carregar dados para carregar ou descarregar dados do Amazon S3?. Se nenhum desses problemas for a causa, considere as opções a seguir específicas para o Hive.

- Verifique se você está usando a versão mais recente do Hive, que tem todos os patches e correções de erros atuais e pode resolver o problema. Para obter mais informações, consulte Apache Hive.
- Usar `INSERT OVERWRITE` exige a listagem do conteúdo do bucket ou pasta do Amazon S3. Isso é uma operação cara. Se possível, remova manualmente o caminho, em vez de fazer com que o Hive liste e exclua os objetos existentes.

- Se você usa versões de EMR lançamento da Amazon anteriores à 5.0, você pode usar o seguinte comando no HiveQL para pré-armazenar em cache os resultados de uma operação de lista do Amazon S3 localmente no cluster:

```
set hive.optimize.s3.query=true;
```

- Use partições estáticas sempre que possível.
- Em algumas versões do Hive e da AmazonEMR, é possível que o uso ALTER TABLES falhe porque a tabela está armazenada em um local diferente do esperado pelo Hive. A solução é adicionar ou atualizar o seguinte no `/home/hadoop/conf/core-site.xml`:

```
<property>
    <name>fs.s3n.endpoint</name>
    <value>s3.amazonaws.com</value>
</property>
```

VPCerros

Os seguintes erros são comuns na VPC configuração na AmazonEMR.

Tópicos

- Configuração de sub-rede inválida
- Conjunto de DHCP opções ausentes
- Erros de permissão
- Erros que resultam em START_FAILED
- Cluster Terminated with errors e NameNode falha ao iniciar

Configuração de sub-rede inválida

Na página Cluster Details (Detalhes do cluster), no campo Status, será exibida uma mensagem de erro semelhante ao seguinte:

`The subnet configuration was invalid: Cannot find route to InternetGateway in main RouteTable` *`rtb-id`* `for vpc` *`vpc-id`*`.`

Para resolver esse problema, você deve criar um Internet Gateway e anexá-lo ao seuVPC. Para obter mais informações, consulte Adicionar um gateway de internet ao seu VPC.

Como alternativa, verifique se você configurou seu VPC com a opção Ativar DNS resolução e Ativar suporte ao DNS nome do host ativadas. Para obter mais informações, consulte Usando DNS com seu VPC.

Conjunto de DHCP opções ausentes

Você verá uma falha de etapa no syslog (log do sistema) do cluster com uma mensagem de erro semelhante ao seguinte:

```
ERROR org.apache.hadoop.security.UserGroupInformation
(main): PriviledgedActionException as:hadoop (auth:SIMPLE)
cause:java.io.IOException:
org.apache.hadoop.yarn.exceptions.ApplicationNotFoundException: Application
with id 'application_id' doesn't exist in RM.
```

ou

```
ERROR org.apache.hadoop.streaming.StreamJob (main): Error Launching job :
org.apache.hadoop.yarn.exceptions.ApplicationNotFoundException: Application
with id 'application_id' doesn't exist in RM.
```

Para resolver esse problema, você deve configurar um VPC que inclua um Conjunto de DHCP Opções cujos parâmetros sejam definidos com os seguintes valores:

> **Note**
>
> Se você usar a região AWS GovCloud (Oeste dos EUA), defina domain-name como **`us-gov-west-1.compute.internal`** em vez do valor usado no exemplo a seguir.

- domain-name = **`ec2.internal`**

  Use **`ec2.internal`**, se a região for Leste dos EUA (Norte da Virgínia). Para outras regiões, use *`region-name`***`.compute.internal`**. Por exemplo, em us-west-2, use domain-name =. **`us-west-2.compute.internal`**
- domain-name-servers = **`AmazonProvidedDNS`**

Para obter mais informações, consulte Conjuntos DHCP de opções.

Erros de permissão

Uma falha no log `stderr` para uma etapa indica que um recurso do Amazon S3 não tem as permissões apropriadas. Este é um erro 403 e a mensagem de erro é semelhante a algo como:

```
Exception in thread "main" com.amazonaws.services.s3.model.AmazonS3Exception: Access
 Denied (Service: Amazon S3; Status Code: 403; Error Code: AccessDenied; Request
 ID: REQUEST_ID
```

Se ActionOnFailure for definido como`TERMINATE_JOB_FLOW`, isso resultará no encerramento do cluster com o estado,`SHUTDOWN_COMPLETED_WITH_ERRORS`.

Algumas maneiras de solucionar esse problema incluem:

- Se você estiver usando uma política de bucket do Amazon S3 em um`VPC`, certifique-se de dar acesso a todos os buckets criando um VPC endpoint e selecionando Permitir tudo na opção Política ao criar o endpoint.
- Certifique-se de que todas as políticas associadas aos recursos do S3 incluam aquela VPC na qual você executa o cluster.
- Tente executar o seguinte comando a partir de seu cluster, para verificar se você pode acessar o bucket

  ```
  hadoop fs -copyToLocal s3://path-to-bucket /tmp/
  ```

- Você pode obter mais informações de depuração específicas, ao configurar o parâmetro `log4j.logger.org.apache.http.wire` como `DEBUG` no arquivo `/home/hadoop/conf/log4j.properties` no cluster. Você pode verificar o arquivo de log `stderr` depois de tentar acessar o bucket a partir do cluster. O arquivo de log fornecerá informações mais detalhadas:

  ```
  Access denied for getting the prefix for bucket - us-west-2.elasticmapreduce with
   path samples/wordcount/input/
  15/03/25 23:46:20 DEBUG http.wire: >> "GET /?prefix=samples%2Fwordcount%2Finput
  %2F&delimiter=%2F&max-keys=1 HTTP/1.1[\r][\n]"
  15/03/25 23:46:20 DEBUG http.wire: >> "Host: us-
  west-2.elasticmapreduce.s3.amazonaws.com[\r][\n]"
  ```

Erros que resultam em **START_FAILED**

Antes da AMI versão 3.7.0, para VPCs onde um nome de host é especificado, a Amazon EMR mapeia os nomes de host internos da sub-rede com endereços de domínio personalizados da seguinte forma:. `ip-`*`X.X.X.X.customdomain.com`*`.tld` Por exemplo, se o nome do host fosse `ip-10.0.0.10` e tivesse a VPC opção de nome de domínio definida como customdomain.com, o nome de host resultante mapeado pela Amazon seria. EMR `ip-10.0.1.0.customdomain.com` Uma entrada é incluída em `/etc/hosts` para resolver o nome do host como 10.0.0.10. Esse comportamento foi alterado com a AMI versão 3.7.0 e agora a Amazon EMR respeita totalmente a DHCP configuração do. VPC Anteriormente, os clientes também podiam usar uma ação de bootstrap para especificar um mapeamento de nome de host.

Se quiser preservar esse comportamento, você deve fornecer a configuração DNS de resolução futura necessária para o domínio personalizado.

Cluster **Terminated with errors** e NameNode falha ao iniciar

Ao iniciar um EMR cluster em um VPC que usa um nome de DNS domínio personalizado, seu cluster pode falhar com a seguinte mensagem de erro no console:

```
Terminated with errors  On the master instance(instance-id), bootstrap action 1
 returned a  non-zero return code
```

A falha é resultado da NameNode impossibilidade de inicialização. Isso resultará no seguinte erro encontrado nos NameNode registros, cujo Amazon S3 URI tem o formato: `s3://`*`mybucket`*`/`*`logs`*`/`*`cluster-id`*`/daemons/`*`master instance-id`*`/hadoop-hadoop-namenode-`*`master node hostname`*`.log.gz`

```
2015-07-23 20:17:06,266 WARN
      org.apache.hadoop.hdfs.server.namenode.FSNamesystem (main): Encountered
 exception
      loading fsimage  java.io.IOException: NameNode is not formatted.
      at

 org.apache.hadoop.hdfs.server.namenode.FSImage.recoverTransitionRead(FSImage.java:212)
          at

 org.apache.hadoop.hdfs.server.namenode.FSNamesystem.loadFSImage(FSNamesystem.java:1020)
          at

 org.apache.hadoop.hdfs.server.namenode.FSNamesystem.loadFromDisk(FSNamesystem.java:739)
```

```
        at
    org.apache.hadoop.hdfs.server.namenode.NameNode.loadNamesystem(NameNode.java:537)
        at
    org.apache.hadoop.hdfs.server.namenode.NameNode.initialize(NameNode.java:596)

    at  org.apache.hadoop.hdfs.server.namenode.NameNode.<init>(NameNode.java:765)
        at
    org.apache.hadoop.hdfs.server.namenode.NameNode.<init>(NameNode.java:749)
    at

org.apache.hadoop.hdfs.server.namenode.NameNode.createNameNode(NameNode.java:1441)
        at
    org.apache.hadoop.hdfs.server.namenode.NameNode.main(NameNode.java:1507)
```

Isso ocorre devido a um possível problema em que uma EC2 instância pode ter vários conjuntos de nomes de domínio totalmente qualificados ao iniciar EMR clusters em umVPC, o que faz uso de um DNS servidor AWS fornecido e de um servidor personalizado fornecido pelo usuárioDNS. Se o DNS servidor fornecido pelo usuário não fornecer nenhum registro pointer (PTR) para nenhum registro A usado para designar nós em um EMR cluster, os clusters falharão na inicialização quando configurados dessa forma. A solução é adicionar 1 PTR registro para cada registro A criado quando uma EC2 instância é iniciada em qualquer uma das sub-redes do. VPC

Erros em clusters de transmissão

Em geral, você pode encontrar a causa de um erro de streaming em um arquivo `syslog`. Estabeleça um link para ela no painel Steps (Etapas).

Os seguintes erros são comuns em clusters de streaming.

Tópicos

- Os dados estão sendo enviados ao mapeador no formato errado?
- Seu script está perdendo a validade?
- Você está transmitindo argumentos de streaming inválidos?
- Seu script foi encerrado com um erro?

Os dados estão sendo enviados ao mapeador no formato errado?

Para verificar se esse é o caso, procure uma mensagem de erro no arquivo `syslog` de uma tentativa de tarefa com falha nos logs de tentativas de tarefas. Para obter mais informações, consulte Exibir arquivos de log do .

Seu script está perdendo a validade?

O tempo limite padrão para um script de mapeador ou reducer é de 600 segundos. Se o script demorar mais do que isso, a tentativa de tarefa falhará. Você pode verificar se esse é o caso consultando o arquivo `syslog` de uma tentativa de tarefa com falha nos logs de tentativas de tarefas. Para obter mais informações, consulte Exibir arquivos de log do .

Você pode alterar o limite de tempo definindo um novo valor para a definição de configuração `mapred.task.timeout`. Essa configuração especifica o número de milissegundos após os quais a Amazon EMR encerrará uma tarefa que não leu a entrada, gravou a saída ou atualizou sua string de status. Você pode atualizar esse valor transmitindo um argumento de streaming adicional `-jobconf mapred.task.timeout=800000`.

Você está transmitindo argumentos de streaming inválidos?

O streaming do Hadoop oferece suporte apenas aos seguintes argumentos. Se você transmitir argumentos diferentes dos listados abaixo, o cluster falhará.

```
-blockAutoGenerateCacheFiles
-cacheArchive
-cacheFile
-cmdenv
-combiner
-debug
-input
-inputformat
-inputreader
-jobconf
-mapper
-numReduceTasks
-output
-outputformat
-partitioner
-reducer
```

```
-verbose
```

Além disso, o streaming do Hadoop só reconhece argumentos transmitidos usando a sintaxe Java; ou seja, precedidos por um único hífen. Se você transmitir argumentos precedidos de um hífen duplo, o cluster falhará.

Seu script foi encerrado com um erro?

Se a saída do seu script de mapeador ou reducer for gerada com um erro, você poderá localizar esse erro no arquivo `stderr` dos logs de tentativas da tarefa com falha. Para obter mais informações, consulte Exibir arquivos de log do .

Erros JAR de cluster personalizados

Os erros a seguir são comuns em JAR clusters personalizados.

Tópicos

- Você está JAR lançando uma exceção antes de criar um emprego?
- Você está JAR lançando um erro dentro de uma tarefa de mapa?

Você está JAR lançando uma exceção antes de criar um emprego?

Se o programa principal do seu programa personalizado JAR gerar uma exceção ao criar a tarefa do Hadoop, o melhor lugar para procurar é o `syslog` arquivo dos registros das etapas. Para obter mais informações, consulte Exibir arquivos de log do .

Você está JAR lançando um erro dentro de uma tarefa de mapa?

Se o seu JAR customizador e o mapeador gerarem uma exceção ao processar os dados de entrada, o melhor lugar para procurar é o `syslog` arquivo dos registros de tentativas de tarefas. Para obter mais informações, consulte Exibir arquivos de log do .

AWS GovCloud Erros (Oeste dos EUA)

A região AWS GovCloud (Oeste dos EUA) difere de outras regiões em sua segurança, configuração e configurações padrão. Como resultado, use a lista de verificação a seguir para solucionar EMR erros da Amazon que são específicos da região AWS GovCloud (Oeste dos EUA) antes de usar recomendações mais gerais de solução de problemas.

- Verifique se suas IAM funções estão configuradas corretamente. Para obter mais informações, consulte Configure funções IAM de serviço para EMR permissões da Amazon para AWS serviços e recursos.
- Certifique-se de que sua VPC configuração tenha configurado corretamente o suporte à DNS resolução/nome do host, o Internet Gateway e DHCP os parâmetros do Conjunto de Opções. Para obter mais informações, consulte VPCerros.

Se essas etapas não resolverem o problema, continue com as etapas para solucionar EMR erros comuns da Amazon. Para obter mais informações, consulte Erros comuns na Amazon EMR.

Encontrar um cluster ausente

Se seu cluster estiver ausente da lista de consoles ou `ListClusters`API, verifique o seguinte:

- Confirme se a idade do cluster, a partir do momento da conclusão, é inferior a dois meses. A Amazon EMR preserva as informações de metadados dos clusters concluídos por dois meses sem nenhum custo. Você não pode excluir clusters concluídos do console. Em vez disso, a Amazon EMR limpa os clusters concluídos automaticamente após dois meses.
- Confirme que você tem permissões de perfil para visualizar o cluster.
- Confirme se você está visualizando o mesmo Região da AWS local em que o cluster reside.

Solucionar problemas em um cluster com falha

Esta seção orienta você durante o processo de solução de problemas de um cluster que apresentou falha. Isso significa que o cluster foi encerrado com um código de erro.

> **Note**
>
> Quando um EMR cluster termina com um erro, o `DescribeCluster` e `ListClusters` APIs retorna um código de erro e uma mensagem de erro. Para alguns erros de cluster, a matriz de dados `ErrorDetail` também ajuda a solucionar a falha. Para obter mais informações, consulte Códigos de erro com ErrorDetail informações.

Se o cluster é executado, mas leva muito tempo para retornar resultados, consulte Solucionar problemas com um cluster lento.

Tópicos

Etapa 1: coletar dados sobre o problema

A primeira etapa para solucionar problemas de um cluster é coletar informações sobre o que deu errado e o status e a configuração atuais do cluster. Essas informações serão usadas nas etapas a seguir para confirmar ou descartar as possíveis causas do problema.

Definir o problema

Começamos fazendo uma definição clara do problema. Algumas perguntas para se fazer:

- O que eu esperava que acontecesse? O que aconteceu em vez disso?
- Quando o problema ocorreu pela primeira vez? Com que frequência ele ocorreu desde então?
- Alguma coisa mudou na forma como eu configuro ou executo o cluster?

Detalhes do cluster

Os detalhes do cluster a seguir são úteis para ajudar a monitorar problemas. Para obter mais informações sobre como reunir essas informações, consulte Visualizar o status e os detalhes do cluster.

- Identificador do cluster. (Também chamado de identificador de fluxo de trabalho.)
- Região da AWS e na Zona de Disponibilidade em que o cluster foi lançado.
- Estado do cluster, inclusive detalhes da última alteração de estado.
- Tipo e número de EC2 instâncias especificados para os nós principal, principal e de tarefa.

Etapa 2: verificar o ambiente

EMRA Amazon opera como parte de um ecossistema de serviços web e software de código aberto. Coisas que afetam essas dependências podem afetar o desempenho da AmazonEMR.

Tópicos

Verificar a existência de interrupções de serviço

A Amazon EMR usa vários Amazon Web Services internamente. Ele executa servidores virtuais na AmazonEC2, armazena dados e scripts no Amazon S3 e reporta métricas para. CloudWatch Eventos que interrompem esses serviços são raros, mas quando ocorrem podem causar problemas na AmazonEMR.

Antes de avançar, verifique o Painel de status dos serviços. Verifique a região onde você iniciou o cluster para saber se há um eventos de interrupção em qualquer um desses serviços.

Verificar os limites de uso

Se você estiver iniciando um cluster grande, tiver lançado vários clusters simultaneamente ou se for um usuário compartilhando um Conta da AWS com outros usuários, o cluster pode ter falhado porque você excedeu um limite de AWS serviço.

A Amazon EC2 limita o número de instâncias de servidores virtuais em execução em uma única AWS região a 20 instâncias sob demanda ou reservadas. Se você iniciar um cluster com mais de 20 nós ou executar um cluster que faça com que o número total de EC2 instâncias ativas em você Conta da AWS exceda 20, o cluster não poderá executar todas as EC2 instâncias necessárias e poderá falhar. Quando isso acontece, a Amazon EMR retorna um `EC2 QUOTA EXCEEDED` erro. Você pode solicitar o AWS aumento do número de EC2 instâncias que podem ser executadas em sua conta enviando uma solicitação para aumentar o limite de EC2 instâncias da Amazon.

Outra coisa que pode fazer você exceder os limites de uso é o atraso entre quando um cluster é encerrado e quando ele libera todos os recursos. Dependendo da configuração, pode demorar de 5 a

20 minutos para um cluster ser encerrado totalmente e liberar os recursos alocados. Se você estiver recebendo um erro `EC2 QUOTA EXCEEDED` ao tentar iniciar um cluster, isso poderá acontecer porque os recursos de um cluster recém-encerrado talvez ainda não tenham sido liberados. Nesse caso, você pode solicitar que sua EC2 cota da Amazon seja aumentada ou esperar vinte minutos e reiniciar o cluster.

O Amazon S3 limita a cem o número de buckets criados em uma conta. Se o cluster criar um bucket novo que exceda esse limite, haverá falha na criação do bucket e poderá fazer com que haja uma falha no cluster.

Verificar a versão

Compare a etiqueta de lançamento que você usou para iniciar o cluster com a EMR versão mais recente da Amazon. Cada versão da Amazon EMR inclui melhorias, como novos aplicativos, recursos, patches e correções de erros. O problema que está afetando o cluster já pode ter sido corrigido na versão mais recente. Se possível, execute o cluster novamente usando a versão da mais recente.

Verifique a configuração da VPC sub-rede da Amazon

Se seu cluster foi lançado em uma VPC sub-rede da Amazon, a sub-rede precisa ser configurada conforme descrito em. Configurar redes Além disso, verifique se a sub-rede na qual o cluster é iniciado tem endereços IP elásticos livres suficientes para atribuir um a cada nó do cluster.

Etapa 3: conferir a última alteração de estado

A última alteração de estado fornece informações sobre o que ocorreu na última vez em que o estado do cluster foi alterado. Isso, geralmente, tem informações que podem determinar o que deu errado, conforme o estado de um cluster muda para `FAILED`. Por exemplo, se você iniciar um cluster de transmissão e especificar um local de saída que já exista no Amazon S3, haverá falha no cluster com uma última alteração de estado de “Streaming output directory already exists”.

Você pode localizar o valor da última alteração de estado no console visualizando o painel de detalhes do cluster, CLI usando os `describe-cluster` argumentos `list-steps` ou ou ou API usando as `ListSteps` ações `DescribeCluster` e. Para obter mais informações, consulte Visualizar o status e os detalhes do cluster.

Etapa 4: examinar os arquivos de log

A próxima etapa é examinar os arquivos de log para localizar um código de erro ou outra indicação do problema que o cluster enfrentou. Para obter informações sobre os arquivos de log disponíveis, onde encontrá-los e como visualizá-los, consulte Exibir arquivos de log do .

Pode ser necessário realizar algum trabalho investigativo para determinar o que aconteceu. O Hadoop executa o trabalho em tentativas de tarefa em múltiplos nós do cluster. A Amazon EMR pode iniciar tentativas de tarefas especulativas, encerrando as outras tentativas de tarefa que não forem concluídas primeiro. Isso gera uma atividade considerável que é registrada nos arquivos de log controller, stderr e syslog quando isso acontece. Além disso, várias tentativas de tarefa são executadas simultaneamente, mas um arquivo de log só pode exibir os resultados de forma linear.

Comece verificando os logs de ações de bootstrap em busca de erros ou alterações inesperadas na configuração durante a inicialização do cluster. A partir daí, consulte os logs de etapas para identificar trabalhos do Hadoop iniciados como parte de uma etapa com erros. Examine os logs de trabalhos do Hadoop para identificar as tentativas de tarefa com falha. O logs de tentativas de tarefa conterá detalhes sobre o que causou a falha de uma tentativa de tarefa.

As seções a seguir descrevem como usar os diversos arquivos de log para identificar erros no cluster.

Verificar os logs de ação de bootstrap

As ações de bootstrap executam scripts no cluster quando ele é iniciado. Geralmente são usados para instalar outros softwares no cluster ou para alterar as configurações com base nos valores padrão. Verificar esses logs pode fornecer insights sobre os erros que ocorreram durante a configuração do cluster, bem como das alterações nas configurações que podem afetar a performance.

Verificar os logs de etapa

Há quatro tipos de logs de etapas.

- controller— Contém arquivos gerados pela Amazon EMR (AmazonEMR) que surgem de erros encontrados ao tentar executar sua etapa. Se a etapa falhar durante o carregamento, você encontrará o rastreamento da pilha nesse log. Os erros ao carregar ou acessar a aplicação muitas vezes são descritos aqui, assim como os erros ausentes do arquivo do mapeador.

- stderr: contém mensagens de erro que ocorreram durante o processamento da etapa. Os erros de carregamento de aplicações muitas vezes são descritos aqui. Às vezes, esse log contém um rastreamento de pilha.
- stdout: contém o status gerado pelos executáveis do mapeador e do redutor. Os erros de carregamento de aplicações muitas vezes são descritos aqui. Às vezes, o log contém mensagens de erro da aplicação.
- syslog: contém registros de softwares que não são da Amazon, como Apache e Hadoop. Os erros de transmissão muitas vezes são descritos aqui.

Verifique se há erros óbvios em stderr. Se stderr exibe uma pequena lista de erros, a etapa foi interrompida rapidamente com um erro gerado. Isso geralmente é causado por um erro nas aplicações mapeadoras e redutoras que estão sendo executadas no cluster.

Verifique se há em avisos de erros ou falhas nas últimas linhas do controller e do syslog. Siga todos os avisos sobre tarefas que falharam, sobretudo se estiver escrito “Job Failed”.

Verificar os logs de tentativa de tarefas

Se a análise anterior dos logs de etapas retornou uma ou mais tarefas com falha, investigue os logs das tentativas de tarefa correspondentes para obter informações mais detalhadas sobre o erro.

Etapa 5: testar o cluster passo a passo

Uma técnica útil quando você está tentando rastrear a origem de um erro é reiniciar o cluster e enviar as etapas a ele uma por uma. Isso permite que você verifique os resultados de cada etapa antes de processar a seguinte e dá a você a oportunidade de corrigir e executar novamente uma etapa que tenha apresentado falha. Isso também permite que você carregue seus dados de entrada somente uma vez.

Para testar um cluster passo a passo

1. Execute um novo cluster, com as proteções de encerramento e keep alive ativadas. A proteção keep alive mantém o cluster em execução após ter processado todas as suas etapas pendentes. A proteção de encerramento impede que um cluster seja encerrado no caso de um erro. Para ter mais informações, consulte Configurar um cluster para continuar ou terminar após a execução da etapa e Usar a proteção contra término.
2. Envie uma etapa para o cluster. Para obter mais informações, consulte Enviar trabalhos a um cluster.

3. Quando a etapa for concluída, verifique se há erros de processamento nos arquivos de log da etapa. Para obter mais informações, consulte Etapa 4: examinar os arquivos de log. A maneira mais rápida de localizar esses arquivos de log é estabelecer uma conexão com o nó principal e exibir os arquivos de log. Os arquivos de log da etapa não serão exibidos até que a etapa seja executada por algum tempo, seja concluída ou apresente uma falha.
4. Se a etapa for concluída com êxito, execute a próxima etapa. Se houver erros, investigue o erro nos arquivos de log. Se houve um erro em seu código, faça a correção e execute novamente a etapa. Continue até que todas as etapas sejam executadas sem erros.
5. Quando você terminar a depuração do cluster e quiser encerrá-lo, deverá fazê-lo manualmente. Isso é necessário porque o cluster foi iniciado com a proteção de encerramento ativada. Para obter mais informações, consulte Usar a proteção contra término.

Solucionar problemas com um cluster lento

Esta seção orienta você durante o processo de solução de problemas com um cluster que ainda está em execução, mas está demorando muito para retornar resultados. Para obter mais informações sobre o que fazer se o cluster tiver sido encerrado com um código de erro, consulte Solucionar problemas em um cluster com falha

A Amazon EMR permite que você especifique o número e o tipo de instâncias no cluster. Essas especificações são os principais meios de afetar a velocidade com a qual o processamento dos seus dados é concluída. Uma coisa que você pode considerar é executar novamente o cluster, desta vez especificando EC2 instâncias com mais recursos ou especificando um número maior de instâncias no cluster. Para obter mais informações, consulte Configurar o hardware e as redes do cluster.

Os tópicos a seguir você orientam você durante o processo de identificar as causas alternativas de um cluster lento.

Tópicos

Etapa 1: coletar dados sobre o problema

A primeira etapa para solucionar problemas de um cluster é coletar informações sobre o que deu errado e o status e a configuração atuais do cluster. Essas informações serão usadas nas etapas a seguir para confirmar ou descartar as possíveis causas do problema.

Definir o problema

Começamos fazendo uma definição clara do problema. Algumas perguntas para se fazer:

- O que eu esperava que acontecesse? O que aconteceu em vez disso?
- Quando o problema ocorreu pela primeira vez? Com que frequência ele ocorreu desde então?
- Alguma coisa mudou na forma como eu configuro ou executo o cluster?

Detalhes do cluster

Os detalhes do cluster a seguir são úteis para ajudar a monitorar problemas. Para obter mais informações sobre como reunir essas informações, consulte Visualizar o status e os detalhes do cluster.

- Identificador do cluster. (Também chamado de identificador de fluxo de trabalho.)
- Região da AWS e na Zona de Disponibilidade em que o cluster foi lançado.
- Estado do cluster, inclusive detalhes da última alteração de estado.
- Tipo e número de EC2 instâncias especificados para os nós principal, principal e de tarefa.

Etapa 2: verificar o ambiente

Tópicos

Verificar a existência de interrupções de serviço

A Amazon EMR usa vários Amazon Web Services internamente. Ele executa servidores virtuais na AmazonEC2, armazena dados e scripts no Amazon S3 e reporta métricas para. CloudWatch Eventos que interrompem esses serviços são raros, mas quando ocorrem podem causar problemas na AmazonEMR.

Antes de avançar, verifique o Painel de status dos serviços. Verifique a região onde você iniciou o cluster para saber se há um eventos de interrupção em qualquer um desses serviços.

Verificar os limites de uso

Se você estiver iniciando um cluster grande, tiver lançado vários clusters simultaneamente ou se for um usuário compartilhando um Conta da AWS com outros usuários, o cluster pode ter falhado porque você excedeu um limite de AWS serviço.

A Amazon EC2 limita o número de instâncias de servidores virtuais em execução em uma única AWS região a 20 instâncias sob demanda ou reservadas. Se você iniciar um cluster com mais de 20 nós ou executar um cluster que faça com que o número total de EC2 instâncias ativas em você Conta da AWS exceda 20, o cluster não poderá executar todas as EC2 instâncias necessárias e poderá falhar. Quando isso acontece, a Amazon EMR retorna um `EC2 QUOTA EXCEEDED` erro. Você pode solicitar o AWS aumento do número de EC2 instâncias que podem ser executadas em sua conta enviando uma solicitação para aumentar o limite de EC2 instâncias da Amazon.

Outra coisa que pode fazer você exceder os limites de uso é o atraso entre quando um cluster é encerrado e quando ele libera todos os recursos. Dependendo da configuração, pode demorar de 5 a 20 minutos para um cluster ser encerrado totalmente e liberar os recursos alocados. Se você estiver recebendo um erro `EC2 QUOTA EXCEEDED` ao tentar iniciar um cluster, isso poderá acontecer porque os recursos de um cluster recém-encerrado talvez ainda não tenham sido liberados. Nesse caso, você pode solicitar que sua EC2 cota da Amazon seja aumentada ou esperar vinte minutos e reiniciar o cluster.

O Amazon S3 limita a cem o número de buckets criados em uma conta. Se o cluster criar um bucket novo que exceda esse limite, haverá falha na criação do bucket e poderá fazer com que haja uma falha no cluster.

Verifique a configuração da VPC sub-rede da Amazon

Se seu cluster foi lançado em uma VPC sub-rede da Amazon, a sub-rede precisa ser configurada conforme descrito em. Configurar redes Além disso, verifique se a sub-rede na qual o cluster é iniciado tem endereços IP elásticos livres suficientes para atribuir um a cada nó do cluster.

Reiniciar o cluster

A lentidão no processamento pode ser causada por uma condição transitória. Considere encerrar e reiniciar o cluster para ver se o desempenho melhora.

Etapa 3: examinar os arquivos de log

A próxima etapa é examinar os arquivos de log para localizar um código de erro ou outra indicação do problema que o cluster enfrentou. Para obter informações sobre os arquivos de log disponíveis, onde encontrá-los e como visualizá-los, consulte Exibir arquivos de log do .

Pode ser necessário realizar algum trabalho investigativo para determinar o que aconteceu. O Hadoop executa o trabalho em tentativas de tarefa em múltiplos nós do cluster. A Amazon EMR pode iniciar tentativas de tarefas especulativas, encerrando as outras tentativas de tarefa que não forem concluídas primeiro. Isso gera uma atividade considerável que é registrada nos arquivos de log controller, stderr e syslog quando isso acontece. Além disso, várias tentativas de tarefa são executadas simultaneamente, mas um arquivo de log só pode exibir os resultados de forma linear.

Comece verificando os logs de ações de bootstrap em busca de erros ou alterações inesperadas na configuração durante a inicialização do cluster. A partir daí, consulte os logs de etapas para identificar trabalhos do Hadoop iniciados como parte de uma etapa com erros. Examine os logs de trabalhos do Hadoop para identificar as tentativas de tarefa com falha. O logs de tentativas de tarefa conterá detalhes sobre o que causou a falha de uma tentativa de tarefa.

As seções a seguir descrevem como usar os diversos arquivos de log para identificar erros no cluster.

Verificar os logs de ação de bootstrap

As ações de bootstrap executam scripts no cluster quando ele é iniciado. Geralmente são usados para instalar outros softwares no cluster ou para alterar as configurações com base nos valores padrão. Verificar esses logs pode fornecer insights sobre os erros que ocorreram durante a configuração do cluster, bem como das alterações nas configurações que podem afetar a performance.

Verificar os logs de etapa

Há quatro tipos de logs de etapas.

- controller— Contém arquivos gerados pela Amazon EMR (AmazonEMR) que surgem de erros encontrados ao tentar executar sua etapa. Se a etapa falhar durante o carregamento, você encontrará o rastreamento da pilha nesse log. Os erros ao carregar ou acessar a aplicação muitas vezes são descritos aqui, assim como os erros ausentes do arquivo do mapeador.
- stderr: contém mensagens de erro que ocorreram durante o processamento da etapa. Os erros de carregamento de aplicações muitas vezes são descritos aqui. Às vezes, esse log contém um rastreamento de pilha.
- stdout: contém o status gerado pelos executáveis do mapeador e do redutor. Os erros de carregamento de aplicações muitas vezes são descritos aqui. Às vezes, o log contém mensagens de erro da aplicação.
- syslog: contém registros de softwares que não são da Amazon, como Apache e Hadoop. Os erros de transmissão muitas vezes são descritos aqui.

Verifique se há erros óbvios em stderr. Se stderr exibe uma pequena lista de erros, a etapa foi interrompida rapidamente com um erro gerado. Isso geralmente é causado por um erro nas aplicações mapeadoras e redutoras que estão sendo executadas no cluster.

Verifique se há em avisos de erros ou falhas nas últimas linhas do controller e do syslog. Siga todos os avisos sobre tarefas que falharam, sobretudo se estiver escrito “Job Failed”.

Verificar os logs de tentativa de tarefas

Se a análise anterior dos logs de etapas retornou uma ou mais tarefas com falha, investigue os logs das tentativas de tarefa correspondentes para obter informações mais detalhadas sobre o erro.

Verificar os logs de daemons do Hadoop

Em casos raros, o Hadoop em si poderá falhar. Para ver se esse é o caso, é necessário examinar os logs do Hadoop. Eles estão localizados em cada nó do `/var/log/hadoop/`.

Você pode usar os JobTracker registros para mapear uma tentativa de tarefa malsucedida para o nó em que ela foi executada. Depois de conhecer o nó associado à tentativa de tarefa, você pode verificar a integridade da EC2 instância que hospeda esse nó para ver se houve algum problema, como falta de memória CPU ou falta de memória.

Etapa 4: verificar a integridade do cluster e das instâncias

Um EMR cluster da Amazon é composto por nós executados em EC2 instâncias da Amazon. Se essas instâncias ficarem limitadas a recursos (como falta de memória CPU ou falta de memória), apresentarem problemas de conectividade de rede ou forem encerradas, a velocidade do processamento do cluster será prejudicada.

Existem até três tipos de nós em um cluster:

- nó principal: gerencia o cluster. Se ele sofrer um problema de desempenho, todo o cluster será afetado.
- nós principais — processa tarefas de redução de mapas e mantém o sistema de arquivos distribuído do Hadoop (). HDFS Se um desses nós tiver um problema de desempenho, ele poderá desacelerar HDFS as operações e reduzir o processamento de mapas. Você pode adicionar outros nós core a um cluster para melhorar o desempenho, mas não pode remover nós core. Para obter mais informações, consulte Redimensionar manualmente um cluster em execução.
- nós de tarefa: processam tarefas map/reduce. Estes são recursos puramente de computação e não armazenam dados. Você pode adicionar nós de tarefas a um cluster para acelerar o desempenho ou pode remover nós de tarefas que não são necessários. Para obter mais informações, consulte Redimensionar manualmente um cluster em execução.

Ao examinar a integridade de um cluster, você deve considerar o desempenho do cluster como um todo, bem como o desempenho de instâncias individuais. Existem várias ferramentas que pode ser usadas:

Verifique a integridade do cluster com CloudWatch

Cada EMR cluster da Amazon reporta métricas para CloudWatch. Essas métricas fornecem informações resumidas de desempenho sobre o cluster, como carga total, HDFS utilização, tarefas em execução, tarefas restantes, blocos corrompidos e muito mais. A análise das CloudWatch métricas fornece uma visão geral do que está acontecendo com seu cluster e pode fornecer informações sobre o que está causando a lentidão no processamento. Além de usar CloudWatch para analisar um problema de desempenho existente, você pode definir alarmes que CloudWatch causem alertas caso ocorra um problema de desempenho futuro. Para obter mais informações, consulte Monitorando EMR métricas da Amazon com CloudWatch.

Verifique o status do trabalho e HDFS a saúde

Use a guia Interfaces de usuário do aplicativo na página de detalhes do cluster para visualizar os detalhes do YARN aplicativo. Para determinados aplicativos, você pode analisar diretamente os logs de acesso em mais detalhes. Isso é útil principalmente para aplicativos Spark. Para obter mais informações, consulte Visualizar o histórico da aplicação.

O Hadoop fornece uma série de interfaces Web que você pode usar para visualizar informações. Para obter mais informações sobre como acessar essas interfaces Web, consulte Visualize interfaces web hospedadas em EMR clusters da Amazon.

- JobTracker — fornece informações sobre o progresso do trabalho que está sendo processado pelo cluster. Você pode usar essa interface para identificar quando um trabalho ficou preso.
- HDFS NameNode — fornece informações sobre a porcentagem de HDFS utilização e o espaço disponível em cada nó. Você pode usar essa interface para identificar quando HDFS está se tornando dependente de recursos e requer capacidade adicional.
- TaskTracker — fornece informações sobre as tarefas do trabalho que está sendo processado pelo cluster. Você pode usar essa interface para identificar quando uma tarefa ficou presa.

Verifique a integridade da instância com a Amazon EC2

Outra forma de procurar informações sobre o status das instâncias em seu cluster é usar o EC2 console da Amazon. Como cada nó no cluster é executado em uma EC2 instância, você pode usar ferramentas fornecidas pela Amazon EC2 para verificar seu status. Para obter mais informações, consulte Veja instâncias de cluster na Amazon EC2.

Etapa 5: verificar se há grupos suspensos

Um grupo de instâncias fica suspenso quando encontra muitos erros ao tentar executar nós. Por exemplo, se novos nós falharem repetidamente durante a execução de ações de bootstrap, depois de algum tempo, o grupo de instâncias entrará no estado `SUSPENDED` em vez de tentar provisionar continuamente novos nós.

Um nó poderá falhar se:

- O Hadoop ou o cluster estiver de alguma forma com problemas e não aceitar um novo nó no cluster
- Uma ação de bootstrap falhar no novo nó

- O nó não estava funcionando corretamente e não conseguir fazer check-in no Hadoop

Se um grupo de instâncias estiver no estado `SUSPENDED`, e o cluster estiver em um estado `WAITING`, você poderá adicionar uma etapa de cluster para redefinir o número desejado de nós core e de tarefa. Adicionar a etapa retoma o processamento do cluster e coloca o grupo de instâncias em um estado `RUNNING`.

Para obter mais informações sobre como redefinir um cluster em um estado suspenso, consulte Estado suspenso.

Etapa 6: revisar as configurações

Definições de configuração especificam detalhes sobre como um cluster é executado, como quantas vezes uma tarefa deve ser repetida e quanta memória está disponível para classificação. Quando você inicia um cluster usando a AmazonEMR, há configurações EMR específicas da Amazon, além das configurações padrão do Hadoop. As definições de configuração são armazenadas no nó principal do cluster. Você pode verificar as definições de configuração para garantir que o cluster tenha os recursos necessários para um funcionamento eficiente.

A Amazon EMR define as configurações padrão do Hadoop que usa para iniciar um cluster. Os valores são baseados no AMI e no tipo de instância que você especifica para o cluster. Você pode modificar os valores padrão das definições de configuração usando uma ação de bootstrap ou especificando novos valores em parâmetros de execução de trabalho. Para obter mais informações, consulte Criar ações de bootstrap para instalar softwares adicionais. Para determinar se uma ação de bootstrap alterou as definições de configuração, verifique os logs dessa ação.

A Amazon EMR registra as configurações do Hadoop usadas para executar cada trabalho. Os dados de log são armazenados em um arquivo nomeado no `job_`*`job-id`*`_conf.xml` `/mnt/var/log/hadoop/history/` diretório do nó principal, onde *`job-id`* é substituído pelo identificador do trabalho. Se você habilitou o arquivamento de logs, esses dados são copiados para o Amazon S3 na `logs/`*`date`*`/`*`jobflow-id`*`/jobs` pasta, onde *`date`* é a data em que o trabalho foi executado e *`jobflow-id`* é o identificador do cluster.

As seguintes definições de configuração de trabalhos do Hadoop são especialmente úteis para investigar problemas de desempenho. Para obter mais informações sobre as definições de configuração do Hadoop e como elas afetam o comportamento do Hadoop, acesse http://hadoop.apache.org/docs/.

⚠ Warning

1. `dfs.replication`Definir como 1 em clusters com menos de quatro nós pode levar à perda de HDFS dados se um único nó ficar inativo. É recomendável usar um cluster com pelo menos quatro nós centrais para workloads de produção.
2. A Amazon não EMR permitirá que os clusters escalem os nós principais abaixo`dfs.replication`. Por exemplo, se `dfs.replication = 2`, o número mínimo de nós central será 2.
3. Ao usar o Ajuste de Escala Gerenciado, o Auto Scaling ou optar por redimensionar manualmente o cluster, é recomendável definir `dfs.replication` como 2 ou mais.

Definição da configuração	Descrição
dfs.replication	O número de HDFS nós para os quais um único bloco (como o bloco do disco rígido) é copiado para produzir um ambiente RAID semelhante. Determina o número de HDFS nós que contêm uma cópia do bloco.
io.sort.mb	Total de memória disponível para classificação. Esse valor deve ser 10x io.sort.factor. Essa configuração também pode ser usada para calcular o total de memória usado pelo nó de tarefas, contando io.sort.mb multiplic ado por mapred.tasktracker.ap.tasks.maximum.
io.sort.spill.percent	Usado durante a classificação, no momento em que o disco começa a ser usado porque a memória de classific ação alocada está ficando cheia.
mapred.child.java.opts	Suspenso. Em vez disso, use mapred.map.child.java.opts e mapred.reduce.child.java.opts. As opções Java são TaskTracker usadas ao iniciar uma JVM tarefa para execução interna. Um parâmetro comum é “-Xmx” para configurar o tamanho máximo da memória.
mapred.map.child.java.opts	As opções Java são TaskTracker usadas ao iniciar uma JVM tarefa de mapeamento para execução interna. Um

Definição da configuração	Descrição
	parâmetro comum é “-Xmx” para configurar o tamanho máximo do heap de memória.
mapred.map.tasks.speculative.execution	Determina se tentativas de tarefas map da mesma tarefa podem ser executadas em paralelo.
mapred.reduce.tasks.speculative.execution	Determina se tentativas de tarefas reduce da mesma tarefa podem ser executadas em paralelo.
mapred.map.max.attempts	O número máximo de vezes que uma tarefa map pode ser tentada. Se tudo falhar, a tarefa map será marcada como falha.
mapred.reduce.child.java.opts	As opções Java são TaskTracker usadas ao iniciar uma JVM tarefa para reduzir a execução. Um parâmetro comum é “-Xmx” para configurar o tamanho máximo do heap de memória.
mapred.reduce.max.attempts	O número máximo de vezes que uma tarefa reduce pode ser tentada. Se tudo falhar, a tarefa map será marcada como falha.
mapred.reduce.slowstart.completed.maps	A quantidade de tarefas map que devem ser concluídas antes que tarefas reduce sejam tentadas. Uma espera insuficiente pode causar erros “Too many fetch-failure” em tentativas.
mapred.reuse.jvm.num.tasks	Uma tarefa é executada em uma únicaJVM. Especifica quantas tarefas podem ser reutilizadas. JVM
mapred.tasktracker.map.tasks.maximum	A quantidade máxima de tarefas que podem ser executadas em paralelo por nó de tarefa durante o mapeamento.
mapred.tasktracker.reduce.tasks.maximum	A quantidade máxima de tarefas que podem ser executadas em paralelo por nó de tarefa durante a redução.

Se as suas tarefas de cluster consumirem muita memória, você poderá melhorar o desempenho usando menos tarefas por nó core e reduzindo seu tamanho do heap do rastreador de trabalhos.

Etapa 7: examinar dados de entrada

Observe seus dados de entrada. Eles estão distribuídos uniformemente entre seus valores de chave? Se os seus dados estiverem fortemente desviados para um ou alguns valores de chave, a carga de processamento pode estar mapeada para um pequeno número de nós, enquanto outros nós estão ociosos. Essa distribuição desequilibrada de trabalho pode resultar em tempos de processamento mais lentos.

Um exemplo de um conjunto de dados desequilibrado seria executar um cluster para colocar palavras em ordem alfabética, mas ter um conjunto de dados contendo apenas palavras que começam com a letra "a". Quando o trabalho fosse mapeado, o nó processando valores que começam com "a" seria sobrecarregado, enquanto os nós processando palavras que começam com outras letras ficariam ociosos.

Solucionar problemas de um cluster do Lake Formation

Esta seção mostra o processo de solução de problemas comuns ao usar a Amazon EMR com AWS Lake Formation.

O acesso ao data lake não é permitido

Você deve optar explicitamente pela filtragem de dados nos EMR clusters da Amazon antes de poder analisar e processar dados em seu data lake. Quando o acesso aos dados falhar, você verá uma mensagem genérica `Access is not allowed` na saída das entradas do caderno.

Para se inscrever e permitir a filtragem de dados na AmazonEMR, consulte Permitir filtragem de dados EMR na Amazon no Guia do AWS Lake Formation desenvolvedor para obter instruções.

Expiração da sessão

O tempo limite da sessão para EMR Notebooks e Zeppelin é controlado pela configuração do IAM Role for Lake Formation. `Maximum CLI/API session duration` O valor padrão para essa configuração é uma hora. Quando ocorrer um tempo limite de sessão, você verá a seguinte mensagem na saída das entradas do seu notebook ao tentar executar os comandos do SparkSQL.

```
Error 401    HTTP ERROR: 401 Problem accessing /sessions/2/statements.
```

```
Reason:  JWT token included in request failed validation.
Powered by Jetty:// 9.3.24.v20180605
 org.springframework.web.client.HttpClientErrorException: 401 JWT token included in
 request failed validation…
```

Para validar sua sessão, atualize a página. Será solicitado que você faça a autenticação novamente usando seu IdP e seja redirecionado de volta para o bloco de anotações. Você pode continuar a executar consultas após a nova autenticação.

Não há permissões para o usuário na tabela solicitada

Ao tentar acessar uma tabela à qual você não tem acesso, você verá a seguinte exceção na saída das entradas do seu notebook ao tentar executar os comandos do SparkSQL.

```
org.apache.spark.sql.AnalysisException:
 org.apache.hadoop.hive.ql.metadata.HiveException: Unable to fetch table table.
Resource does not exist or requester is not authorized to access requested
 permissions.
(Service: AWSGlue; Status Code: 400; Error Code: AccessDeniedException; Request ID: …
```

Para acessar a tabela, você deve conceder acesso ao usuário atualizando as permissões associadas a essa tabela no Lake Formation.

Consultar dados de várias contas compartilhados com o Lake Formation

Quando você usa EMR a Amazon para acessar dados compartilhados com você de outra conta, algumas bibliotecas do Spark tentarão chamar a `Glue:GetUserDefinedFunctions` API operação. Como as versões 1 e 2 das permissões AWS RAM gerenciadas não oferecem suporte a essa ação, você recebe a seguinte mensagem de erro:

`"ERROR: User: arn:aws:sts::012345678901:assumed-role/my-spark-role/i-06ab8c2b59299508a is not authorized to perform: glue:GetUserDefinedFunctions on resource: arn:exampleCatalogResource because no resource-based policy allows the glue:GetUserDefinedFunctions action"`

Para resolver esse erro, o administrador do data lake que criou o compartilhamento de recursos deve atualizar as permissões AWS RAM gerenciadas anexadas ao compartilhamento de recursos. A versão 3 das permissões gerenciadas pelo AWS RAM permite que as entidades principais executem a ação `glue:GetUserDefinedFunctions`.

Se você criar um novo compartilhamento de recursos, o Lake Formation aplicará a versão mais recente da permissão AWS RAM gerenciada por padrão, e nenhuma ação será exigida por você. Para habilitar o acesso a dados entre contas para compartilhamentos de recursos existentes, você precisa atualizar as permissões AWS RAM gerenciadas para a versão 3.

Você pode ver as AWS RAM permissões atribuídas aos recursos compartilhados com você em AWS RAM. As permissões incluídas na versão 3 são estas:

```
Databases
  AWSRAMPermissionGlueDatabaseReadWriteForCatalog
  AWSRAMPermissionGlueDatabaseReadWrite

Tables
  AWSRAMPermissionGlueTableReadWriteForCatalog
  AWSRAMPermissionGlueTableReadWriteForDatabase

AllTables
  AWSRAMPermissionGlueAllTablesReadWriteForCatalog
  AWSRAMPermissionGlueAllTablesReadWriteForDatabase
```

Para atualizar a versão de permissões AWS RAM gerenciadas dos compartilhamentos de recursos existentes

Você (administrador do data lake) pode atualizar as permissões AWS RAM gerenciadas para uma versão mais recente seguindo as instruções no Guia do AWS RAM usuário ou revogar todas as permissões existentes para o tipo de recurso e concedê-las novamente. Se você revogar as permissões, AWS RAM excluirá o compartilhamento AWS RAM de recursos associado ao tipo de recurso. Quando você concede permissões novamente, AWS RAM cria novos compartilhamentos de recursos anexando a versão mais recente das permissões AWS RAM gerenciadas.

Inserir, criar e alterar tabelas

Não há suporte para a inserção, a criação ou a alteração de tabelas em bancos de dados protegidos por políticas do Lake Formation. Ao realizar essas operações, você verá a seguinte exceção na saída das entradas do seu notebook ao tentar executar os SQL comandos do Spark:

```
java.io.IOException:
 com.amazon.ws.emr.hadoop.fs.shaded.com.amazonaws.services.s3.model.AmazonS3Exception:
```

```
            Access Denied (Service: Amazon S3; Status Code: 403; Error Code:
AccessDenied; Request ID: …
```

Para obter mais informações, consulte Limitações da EMR integração da Amazon com AWS Lake Formation.

Escrita de aplicações que iniciam e gerenciam clusters

Tópicos

Você pode acessar a funcionalidade fornecida pela API do Amazon EMR chamando funções de wrapper em um dos SDKs. AWS Os AWS SDKs fornecem funções específicas de linguagem que envolvem a API do serviço web e simplificam a conexão com o serviço web, lidando com muitos dos detalhes da conexão para você. Para obter mais informações sobre como chamar o Amazon EMR usando um dos SDKs, consulte Uso de SDKs para chamar APIs do Amazon EMR.

> **Important**
>
> A taxa máxima de solicitações para o Amazon EMR é de uma solicitação a cada dez segundos.

E Exemplo de código-fonte Java do nd-to-end Amazon EMR

Os desenvolvedores podem chamar a API do Amazon EMR usando código Java personalizado para fazer as mesmas coisas que fariam com o console ou com a CLI do Amazon EMR. Esta seção fornece as end-to-end etapas necessárias para instalar AWS Toolkit for Eclipse e executar uma amostra de código-fonte Java totalmente funcional que adiciona etapas a um cluster do Amazon EMR.

> **Note**
>
> Este exemplo se concentra em Java, mas o Amazon EMR também oferece suporte a diversas linguagens de programação com uma coleção de SDKs do Amazon EMR. Para ter mais informações, consulte Uso de SDKs para chamar APIs do Amazon EMR.

Este exemplo de código-fonte Java demonstra como executar as seguintes tarefas usando a API do Amazon EMR:

- Recupere AWS as credenciais e envie-as ao Amazon EMR para fazer chamadas de API
- Configurar uma nova etapa personalizada e uma nova etapa predefinida
- Adicionar novas etapas a um cluster existente do Amazon EMR
- Recuperar os IDs das etapas de um cluster em execução

> **Note**
>
> Este exemplo demonstra como adicionar etapas a um cluster existente e, portanto, requer um cluster ativo na sua conta.

Antes de começar, instale a versão do Eclipse IDE for Java EE Developers (Eclipse IDE para desenvolvedores de Java EE) que corresponda a sua plataforma do computador. Para obter mais informações, acesse a página de downloads do Eclipse.

Em seguida, instale o plug-in de desenvolvimento de banco de dados para o Eclipse.

Instalar o plug-in de desenvolvimento de banco de dados para o Eclipse

1. Abra o Eclipse IDE.
2. Escolha Help (Ajuda) e Install New Software (Instalar novo software).
3. No campo Work with: (Trabalhar com:), digite **`http://download.eclipse.org/releases/kepler`** ou o caminho que corresponda ao número da versão do seu Eclipse IDE.
4. Na lista de itens, escolha Database Development (Desenvolvimento de banco de dados) e Finish (Concluir).
5. Reinicie o Eclipse quando solicitado.

Em seguida, instale o kit de ferramentas para Eclipse a fim de disponibilizar os modelos de projeto de código-fonte úteis e configurados previamente.

Instalar o kit de ferramentas para Eclipse

1. Abra o Eclipse IDE.

2. Escolha Help (Ajuda) e Install New Software (Instalar novo software).
3. No campo Work with: (Trabalhar com:), digite **`https://aws.amazon.com/eclipse`**.
4. Na lista de itens, escolha AWS Toolkit for Eclipse e Finish.
5. Reinicie o Eclipse quando solicitado.

Em seguida, crie um novo projeto AWS Java e execute o código-fonte Java de amostra.

Para criar um novo projeto AWS Java

1. Abra o Eclipse IDE.
2. Escolha File (Arquivo), New (Novo) e Other (Outros).
3. Na caixa de diálogo Select a wizard, escolha AWS Java Project e Next.
4. Na caixa de diálogo Novo projeto AWS Java, no **`Project name:`** campo, insira o nome do seu novo projeto, por exemplo**`EMR-sample-code`**.
5. Escolha Configurar AWS contas..., insira suas chaves de acesso públicas e privadas e escolha Concluir. Para obter mais informações sobre a criação de chaves de acesso, consulte How do I get security credentials? na Referência geral da Amazon Web Services.

   > **Note**
   >
   > Você não deve incorporar chaves de acesso diretamente no código. O SDK do Amazon EMR permite colocar as chaves de acesso em locais conhecidos para que você não precise mantê-las em código.

6. No novo projeto Java, clique com o botão direito do mouse na pasta src e, em seguida, escolha New (Novo) e Class (Classe).
7. Na caixa de diálogo Java Class (Classe Java), no campo Name (Nome), insira um nome para sua nova classe, por exemplo, **`main`**.
8. Na seção Which method stubs would you like to create? (Quais stubs de método você gostaria de criar?) escolha public static void main (String [] args) e Finish (Concluir).
9. Insira o código-fonte em Java dentro de sua nova classe e adicione as declarações adequadas de import (importação)para as classes e métodos no exemplo. Para sua conveniência, a listagem do código-fonte completo é mostrada abaixo.

> **Note**
>
> No código de exemplo a seguir, substitua o exemplo de ID de cluster (JobFlowId) *j-xxxxxxxxxxxx*, por um ID de cluster válido em sua conta encontrado no AWS Management Console ou usando o seguinte AWS CLI comando:
>
> ```
> aws emr list-clusters --active | grep "Id"
> ```
>
> Além disso, substitua o caminho de exemplo do Amazon S3, *s3://path/to/my/jarfolder*, pelo caminho válido para o seu JAR. Por fim, substitua o nome da classe do exemplo, *com.my.Main1*, pelo nome correto da classe em seu JAR, se aplicável.

```
import com.amazonaws.AmazonClientException;
import com.amazonaws.auth.AWSCredentials;
import com.amazonaws.auth.AWSStaticCredentialsProvider;
import com.amazonaws.auth.profile.ProfileCredentialsProvider;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduce;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduceClientBuilder;
import com.amazonaws.services.elasticmapreduce.model.*;
import com.amazonaws.services.elasticmapreduce.util.StepFactory;

public class Main {

 public static void main(String[] args) {
  AWSCredentials credentials_profile = null;
  try {
   credentials_profile = new
 ProfileCredentialsProvider("default").getCredentials();
  } catch (Exception e) {
   throw new AmazonClientException(
     "Cannot load credentials from .aws/credentials file. " +
       "Make sure that the credentials file exists and the profile name is
 specified within it.",
     e);
  }

  AmazonElasticMapReduce emr = AmazonElasticMapReduceClientBuilder.standard()
    .withCredentials(new AWSStaticCredentialsProvider(credentials_profile))
    .withRegion(Regions.US_WEST_1)
```

```
    .build();

  // Run a bash script using a predefined step in the StepFactory helper class
  StepFactory stepFactory = new StepFactory();
  StepConfig runBashScript = new StepConfig()
    .withName("Run a bash script")
    .withHadoopJarStep(stepFactory.newScriptRunnerStep("s3://jeffgoll/emr-scripts/
create_users.sh"))
    .withActionOnFailure("CONTINUE");

  // Run a custom jar file as a step
  HadoopJarStepConfig hadoopConfig1 = new HadoopJarStepConfig()
    .withJar("s3://path/to/my/jarfolder") // replace with the location of the jar
 to run as a step
    .withMainClass("com.my.Main1") // optional main class, this can be omitted if
 jar above has a manifest
    .withArgs("--verbose"); // optional list of arguments to pass to the jar
  StepConfig myCustomJarStep = new StepConfig("RunHadoopJar", hadoopConfig1);

  AddJobFlowStepsResult result = emr.addJobFlowSteps(new AddJobFlowStepsRequest()
    .withJobFlowId("j-xxxxxxxxxxxx") // replace with cluster id to run the steps
    .withSteps(runBashScript, myCustomJarStep));

  System.out.println(result.getStepIds());

 }
}
```

10. Escolha Run (Executar), Run As (Executar como) e Java Application (Aplicativo Java).
11. Se o exemplo for executado corretamente, uma lista de IDs para as novas etapas aparece na janela do console do Eclipse IDE. A saída correta é semelhante à seguinte:

```
[s-39BLQZRJB2E5E, s-1L6A4ZU2SAURC]
```

Conceitos comuns para chamadas de API

Tópicos

- Endpoints para o Amazon EMR
- Especificação dos parâmetros de cluster no Amazon EMR
- Zonas de disponibilidade no Amazon EMR

- Como usar arquivos e bibliotecas adicionais em clusters do Amazon EMR

Ao escrever uma aplicação que chama a API do Amazon EMR, há vários conceitos que se aplicam ao chamar uma das funções wrapper de um SDK.

Endpoints para o Amazon EMR

Um endpoint é um URL que é o ponto de entrada para um serviço da Web. Toda solicitação de serviço da web deve conter um endpoint. O endpoint especifica a AWS região em que os clusters são criados, descritos ou encerrados. Ele tem o formato `elasticmapreduce.`*`regionname`*`.amazonaws.com`. Se você especificar o endpoint geral (`elasticmapreduce.amazonaws.com`), o Amazon EMR direcionará sua solicitação para um endpoint na região padrão. Para contas criadas a partir de 8 de março de 2013, a região padrão é us-west-2; para contas mais antigas, a região padrão é us-east-1.

Para obter mais informações sobre os endpoints do Amazon EMR, consulte Regions and endpoints na Referência geral da Amazon Web Services.

Especificação dos parâmetros de cluster no Amazon EMR

Os parâmetros `Instances` permitem que você configure os tipos e o número de instâncias do EC2 para criar os nós que vão processar os dados. O Hadoop distribui o processamento dos dados entre os vários nós do cluster. O nó principal é responsável por acompanhar a integridade dos nós core e de tarefas e por sondar os nós para obter o status dos resultados de trabalhos. Os nós core e de tarefa realizam o processamento real dos dados. Se você tem um cluster com um único nó, este nó serve como nó principal e também como nó core.

O parâmetro `KeepJobAlive` em uma solicitação `RunJobFlow` determina se um cluster deve ser encerrado quando não tem mais etapas para executar. Defina este valor como `False` quando você sabe que o cluster está sendo executado como esperado. Quando você estiver tentando resolver problemas no fluxo de trabalho e adicionando etapas enquanto a execução do cluster é suspensa, defina este valor como `True`. Isso reduz a quantidade de tempo e as despesas de upload dos resultados para o Amazon Simple Storage Service (Amazon S3), apenas para repetir o processo após a modificação de uma etapa para reiniciar o cluster.

Em caso `KeepJobAlive` `true` afirmativo, depois de fazer com que o cluster conclua seu trabalho, você deve enviar uma `TerminateJobFlows` solicitação ou o cluster continuará em execução e gerará AWS cobranças.

Para obter mais informações sobre parâmetros que são exclusivos de`RunJobFlow`, consulte RunJobFlow. Para obter mais informações sobre os parâmetros genéricos na solicitação, consulte Common request parameters.

Zonas de disponibilidade no Amazon EMR

O Amazon EMR usa instâncias do EC2 como nós para o processamento de clusters. Essas instâncias do EC2 têm locais que são compostos por regiões e zonas de disponibilidade. As regiões são dispersas e localizadas em diferentes áreas geográficas. As zonas de disponibilidade são locais distintos dentro de uma região, que são isolados de falhas que ocorrem em outras zonas de disponibilidade. Cada zona de disponibilidade fornece conectividade de rede de baixa latência e custo reduzido para outras zonas de disponibilidade na mesma região. Para obter uma lista de regiões e endpoints para o Amazon EMR, consulte Regions and endpoints na Referência geral da Amazon Web Services.

O parâmetro `AvailabilityZone` especifica a localização geral do cluster. Esse parâmetro é opcional e, em geral, não recomendamos o seu uso. Quando `AvailabilityZone` não é especificado, o Amazon EMR escolhe automaticamente o melhor valor de `AvailabilityZone` para o cluster. Esse parâmetro pode ser útil se você desejar compartilhar os locais de suas instâncias com outras instâncias existentes em execução e seu cluster precisar ler ou gravar dados dessas instâncias. Para obter mais informações, consulte o Guia do usuário do Amazon EC2.

Como usar arquivos e bibliotecas adicionais em clusters do Amazon EMR

Em algumas ocasiões você pode querer usar arquivos adicionais ou bibliotecas personalizadas com seus aplicativos de mapeador ou reducer. Por exemplo, você pode querer usar uma biblioteca que converte um arquivo PDF em texto simples.

Para armazenar um arquivo em cache a ser usado pelo mapeador ou reducer quando usarem o streaming do Hadoop

- No campo `args` do JAR, adicione o seguinte argumento:

  ```
  -cacheFile s3://bucket/path_to_executable#local_path
  ```

  O arquivo `local_path` está no diretório de trabalho do mapeador, que pode referenciar o arquivo.

Uso de SDKs para chamar APIs do Amazon EMR

Tópicos

Os AWS SDKs fornecem funções que envolvem a API e cuidam de muitos detalhes da conexão, como calcular assinaturas, lidar com novas tentativas de solicitação e tratamento de erros. Os SDKs também contêm exemplos de código, tutoriais e outros recursos para ajudar você a começar a criar aplicativos que chamam. AWS Chamar as funções do wrapper em um SDK pode simplificar muito o processo de criação de um AWS aplicativo.

Para obter mais informações sobre como baixar e usar os AWS SDKs, consulte SDKs em Tools for Amazon Web Services.

Usando o AWS SDK for Java para criar um cluster do Amazon EMR

AWS SDK for Java Ele fornece três pacotes com a funcionalidade do Amazon EMR:

- com.amazonaws.services.elasticmapreduce
- com.amazonaws.services.elasticmapreduce.model
- com.amazonaws.services.elasticmapreduce.util

Para obter mais informações sobre esses pacotes, consulte Referência da API do AWS SDK for Java.

O exemplo a seguir ilustra como os SDKs podem simplificar a programação com o Amazon EMR. O exemplo de código apresentado abaixo usa o objeto `StepFactory`, uma classe auxiliar para criar tipos de etapas comuns do Amazon EMR, para criar um cluster do Hive interativo com a depuração habilitada.

```
import com.amazonaws.AmazonClientException;
import com.amazonaws.auth.AWSCredentials;
import com.amazonaws.auth.AWSStaticCredentialsProvider;
import com.amazonaws.auth.profile.ProfileCredentialsProvider;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduce;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduceClientBuilder;
import com.amazonaws.services.elasticmapreduce.model.*;
import com.amazonaws.services.elasticmapreduce.util.StepFactory;
```

```
public class Main {

 public static void main(String[] args) {
  AWSCredentialsProvider profile = null;
  try {
   credentials_profile = new ProfileCredentialsProvider("default"); // specifies any
 named profile in
                                    // .aws/credentials as the credentials provider
  } catch (Exception e) {
   throw new AmazonClientException(
     "Cannot load credentials from .aws/credentials file. " +
       "Make sure that the credentials file exists and that the profile name is defined
 within it.",
     e);
  }

  // create an EMR client using the credentials and region specified in order to
  // create the cluster
  AmazonElasticMapReduce emr = AmazonElasticMapReduceClientBuilder.standard()
    .withCredentials(credentials_profile)
    .withRegion(Regions.US_WEST_1)
    .build();

  // create a step to enable debugging in the AWS Management Console
  StepFactory stepFactory = new StepFactory();
  StepConfig enabledebugging = new StepConfig()
    .withName("Enable debugging")
    .withActionOnFailure("TERMINATE_JOB_FLOW")
    .withHadoopJarStep(stepFactory.newEnableDebuggingStep());

  // specify applications to be installed and configured when EMR creates the
  // cluster
  Application hive = new Application().withName("Hive");
  Application spark = new Application().withName("Spark");
  Application ganglia = new Application().withName("Ganglia");
  Application zeppelin = new Application().withName("Zeppelin");

  // create the cluster
  RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("MyClusterCreatedFromJava")
    .withReleaseLabel("emr-5.20.0") // specifies the EMR release version label, we
 recommend the latest release
    .withSteps(enabledebugging)
```

```
    .withApplications(hive, spark, ganglia, zeppelin)
    .withLogUri("s3://path/to/my/emr/logs") // a URI in S3 for log files is required
 when debugging is enabled
    .withServiceRole("EMR_DefaultRole") // replace the default with a custom IAM
 service role if one is used
    .withJobFlowRole("EMR_EC2_DefaultRole") // replace the default with a custom EMR
 role for the EC2 instance
                        // profile if one is used
    .withInstances(new JobFlowInstancesConfig()
      .withEc2SubnetId("subnet-12ab34c56")
      .withEc2KeyName("myEc2Key")
      .withInstanceCount(3)
      .withKeepJobFlowAliveWhenNoSteps(true)
      .withMasterInstanceType("m4.large")
      .withSlaveInstanceType("m4.large"));

  RunJobFlowResult result = emr.runJobFlow(request);
  System.out.println("The cluster ID is " + result.toString());

 }

}
```

No mínimo, você deve passar uma função de serviço e uma função de fluxo de trabalho correspondentes ao EMR_ DefaultRole e ao EMR_EC2_, respectivamente. DefaultRole Você pode fazer isso invocando esse AWS CLI comando para a mesma conta. Primeiro, verifique se as funções já existem:

```
aws iam list-roles | grep EMR
```

Tanto o perfil da instância (EMR_EC2_DefaultRole) quanto a função de serviço (EMR_DefaultRole) serão exibidos se existirem:

```
"RoleName": "EMR_DefaultRole",
            "Arn": "arn:aws:iam::AccountID:role/EMR_DefaultRole"
            "RoleName": "EMR_EC2_DefaultRole",
            "Arn": "arn:aws:iam::AccountID:role/EMR_EC2_DefaultRole"
```

Se os perfis padrão não existirem, você poderá usar o seguinte comando para criá-los:

```
aws emr create-default-roles
```

Gerenciamento de cotas de serviço do Amazon EMR

Tópicos

Os tópicos desta seção descrevem as cotas de serviço do EMR (anteriormente chamadas de limites de serviço), como gerenciá-las no e quando é vantajoso usar CloudWatch eventos em vez de cotas de serviço para monitorar clusters e acionar ações. AWS Management Console

O que são as cotas de serviço do Amazon EMR

Sua AWS conta tem cotas de serviço padrão, também conhecidas como limites, para cada AWS serviço. O serviço EMR tem dois tipos de limites:

- Limites de recursos: você pode usar o EMR para criar recursos do EC2. Contudo, estes recursos do EC2 estão sujeitos a cotas de serviço. As limitações de recursos nesta categoria são:
  - O número máximo de clusters ativos que podem ser executados ao mesmo tempo.
  - O número máximo de instâncias ativas por grupo de instâncias.
- Limites de APIs: ao usar APIs do EMR, os dois tipos de limitações são:
  - Limite de intermitência: este é o número máximo de chamadas de API que você pode fazer de uma só vez. Por exemplo, o número máximo de solicitações de AddInstanceFleet API que você pode fazer por segundo é definido como 5 chamadas/segundo como padrão. Isso significa que o limite de intermitência da AddInstanceFleet API é de 5 chamadas/segundo ou que, a qualquer momento, você pode fazer no máximo 5 AddInstanceFleet chamadas de API. Entretanto, depois de usar o limite de intermitência, as chamadas subsequentes serão limitadas pelo limite de taxa.
  - Limite de taxa: esta é a taxa de reabastecimento da capacidade de expansão da API. Por exemplo, a taxa de reabastecimento de AddInstanceFleet chamadas é definida como 0,5 chamadas/segundo como padrão. Isso significa que, depois de atingir o limite de intermitência, você terá que esperar, no mínimo, dois segundos (0,5 chamadas por segundo X 2 segundos = 1 chamada) para chamar a API. Se você fizer uma chamada antes disso, sofrerá o controle de utilização pelo serviço Web do EMR. A qualquer momento, você pode fazer somente a quantidade de chamadas correspondente à capacidade de expansão sem sofrer o controle de

utilização. A cada segundo adicional que você espera, a capacidade de expansão aumenta em 0,5 chamadas até atingir o limite máximo de cinco, que corresponde ao limite de intermitência.

Como gerenciar cotas de serviço do Amazon EMR

O Service Quotas é um AWS recurso que você pode usar para visualizar e gerenciar suas cotas ou limites de serviços do Amazon EMR a partir de um local central usando a API ou a AWS Management Console CLI. Para saber mais sobre como visualizar quotas e solicitar aumentos, consulte AWS service quotas na Referência geral da Amazon Web Services.

Para algumas APIs, configurar um CloudWatch evento pode ser uma opção melhor do que aumentar as cotas de serviço. Você também pode economizar tempo usando CloudWatch para definir alarmes e acionar solicitações de aumento de forma proativa, antes de atingir a cota de serviço. Para obter mais detalhes, consulte Quando configurar eventos do EMR em CloudWatch.

Quando configurar eventos do EMR em CloudWatch

Para algumas APIs de pesquisa, como DescribeCluster,, e DescribeStep ListClusters, configurar um CloudWatch evento pode reduzir o tempo de resposta às mudanças e liberar suas cotas de serviço. Por exemplo, se você tiver uma função do Lambda configurada para ser executada quando o estado de um cluster for alterado, como quando uma etapa for concluída ou um cluster for encerrado, você poderá usar esse acionador para iniciar a próxima ação em seu fluxo de trabalho em vez de aguardar pela próxima sondagem. Caso contrário, se você tiver instâncias dedicadas do Amazon EC2 ou funções do Lambda sondando constantemente a API do EMR em busca de alterações, você não somente desperdiçará recursos de computação, mas também poderá atingir sua cota de serviço.

A seguir são apresentados alguns casos nos quais você pode se beneficiar ao migrar para uma arquitetura orientada a eventos.

Caso 1: Sondagem do EMR DescribeCluster usando chamadas de API para conclusão da etapa

Example Pesquisando o EMR DescribeCluster usando chamadas de API para conclusão da etapa

Um padrão comum é enviar uma etapa para um cluster em execução e consultar o Amazon EMR para obter o status da etapa, normalmente usando DescribeCluster as DescribeStep APIs ou. Essa tarefa também pode ser realizada com atraso mínimo ao se conectar ao evento de alteração de etapa ou de status do Amazon EMR.

Este evento inclui as informações apresentadas a seguir em sua carga útil.

```
{
  "version": "0",
  "id": "999cccaa-eaaa-0000-1111-123456789012",
  "detail-type": "EMR Step Status Change",
  "source": "aws.emr",
  "account": "123456789012",
  "time": "2016-12-16T20:53:09Z",
  "region": "us-east-1",
  "resources": [],
  "detail": {
    "severity": "ERROR",
    "actionOnFailure": "CONTINUE",
    "stepId": "s-ZYXWVUTSRQPON",
    "name": "CustomJAR",
    "clusterId": "j-123456789ABCD",
    "state": "FAILED",
    "message": "Step s-ZYXWVUTSRQPON (CustomJAR) in Amazon EMR cluster j-123456789ABCD
 (Development Cluster) failed at 2016-12-16 20:53 UTC."
  }
}
```

No mapa detalhado, uma função do Lambda pode analisar “state”, “stepId” ou “clusterId” para localizar informações pertinentes.

Caso 2: sondagem do EMR para clusters disponíveis para a execução de fluxos de trabalho

Example Sondagem do EMR para clusters disponíveis para a execução de fluxos de trabalho

Um padrão para clientes que executam vários clusters é executar fluxos de trabalho em clusters assim que estiverem disponíveis. Se houver muitos clusters em execução e um fluxo de trabalho precisar ser executado em um cluster que está aguardando, um padrão pode ser pesquisar o EMR DescribeCluster usando ListClusters chamadas de API para os clusters disponíveis. Outra maneira de reduzir o atraso em saber quando um cluster está pronto para uma etapa seria processar o evento de alteração de estado do cluster do Amazon EMR.

Este evento inclui as informações apresentadas a seguir em sua carga útil.

```
{
```

```
  "version": "0",
  "id": "999cccaa-eaaa-0000-1111-123456789012",
  "detail-type": "EMR Cluster State Change",
  "source": "aws.emr",
  "account": "123456789012",
  "time": "2016-12-16T20:43:05Z",
  "region": "us-east-1",
  "resources": [],
  "detail": {
    "severity": "INFO",
    "stateChangeReason": "{\"code\":\"\"}",
    "name": "Development Cluster",
    "clusterId": "j-123456789ABCD",
    "state": "WAITING",
    "message": "Amazon EMR cluster j-123456789ABCD ..."
  }
}
```

Para este evento, uma função do Lambda pode ser configurada para enviar imediatamente um fluxo de trabalho em espera para um cluster assim que seu status for alterado para WAITING.

Caso 3: sondagem do EMR para o encerramento de um cluster

Example Sondagem do EMR para o encerramento de um cluster

Um padrão comum para clientes que executam vários clusters do EMR é sondar o Amazon EMR em busca de clusters encerrados para que o trabalho não seja mais enviado a eles. Você pode implementar esse padrão com as chamadas de ListClusters API DescribeCluster e usando o evento Amazon EMR Cluster State Change em.

Após o encerramento do cluster, o evento emitido é semelhante ao exemplo apresentado a seguir.

```
{
  "version": "0",
  "id": "1234abb0-f87e-1234-b7b6-000000123456",
  "detail-type": "EMR Cluster State Change",
  "source": "aws.emr",
  "account": "123456789012",
  "time": "2016-12-16T21:00:23Z",
  "region": "us-east-1",
  "resources": [],
  "detail": {
    "severity": "INFO",
```

```
    "stateChangeReason": "{\"code\":\"USER_REQUEST\",\"message\":\"Terminated by user
 request\"}",
    "name": "Development Cluster",
    "clusterId": "j-123456789ABCD",
    "state": "TERMINATED",
    "message": "Amazon EMR Cluster jj-123456789ABCD (Development Cluster) has
 terminated at 2016-12-16 21:00 UTC with a reason of USER_REQUEST."
  }
}
```

A seção “Detalhes” da carga útil inclui o clusterId e o estado que podem ser utilizados.

Glossário do AWS

Para obter a terminologia mais recente da AWS, consulte o glossário da AWS na Referência do Glossário da AWS.

As traduções são geradas por tradução automática. Em caso de conflito entre o conteúdo da tradução e da versão original em inglês, a versão em inglês prevalecerá.