



API Reference

Amazon Bedrock



Amazon Bedrock: API Reference

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

Table of Contents

Amazon Bedrock API Reference	1
Amazon Bedrock endpoints	1
AWS Command Line Interface references	2
AWS SDK references	2
Actions	4
Amazon Bedrock	8
Agents for Amazon Bedrock	143
Agents for Amazon Bedrock Runtime	451
Amazon Bedrock Runtime	497
Data Types	544
Amazon Bedrock	556
Agents for Amazon Bedrock	646
Agents for Amazon Bedrock Runtime	858
Amazon Bedrock Runtime	1017
Common Parameters	1104
Common Errors	1106

Amazon Bedrock API Reference

This document provides detailed information about the Bedrock API actions and their parameters. For more information about setting up the Amazon Bedrock APIs, see [Set up the Amazon Bedrock API](#).

For information about the IAM access control permissions you need to use the APIs, see [Identity-based policy examples for Amazon Bedrock](#).

Amazon Bedrock endpoints

To connect programmatically to an AWS service, you use an endpoint. Refer to the [Amazon Bedrock endpoints and quotas](#) chapter in the AWS General Reference for information about the endpoints that you can use for Amazon Bedrock.

Amazon Bedrock provides the following service endpoints.

- `bedrock` – Contains control plane APIs for managing, training, and deploying models. For more information, see [Amazon Bedrock Actions](#) and [Amazon Bedrock Data Types](#).
- `bedrock-runtime` – Contains data plane APIs for making inference requests for models hosted in Amazon Bedrock. For more information, see [Amazon Bedrock Runtime Actions](#) and [Amazon Bedrock Runtime Data Types](#).
- `bedrock-agent` – Contains control plane APIs for creating and managing agents and knowledge bases. For more information, see [Agents for Amazon Bedrock Actions](#) and [Agents for Amazon Bedrock Data Types](#).
- `bedrock-agent-runtime` – Contains data plane APIs for invoking agents and querying knowledge bases. For more information, see [Agents for Amazon Bedrock Runtime Actions](#) and [Agents for Amazon Bedrock Runtime Data Types](#).

Note

Check that you're using the correct endpoint when making an API request.

AWS Command Line Interface references

Refer to the following references for AWS CLI commands and operations:

- [Amazon Bedrock CLI commands](#)
- [Amazon Bedrock Runtime CLI commands](#)
- [Agents for Amazon Bedrock CLI commands](#)
- [Agents for Amazon Bedrock Runtime CLI commands](#)

AWS SDK references

AWS software development kits (SDKs) are available for many popular programming languages. Each SDK provides an API, code examples, and documentation that make it easier for developers to build applications in their preferred language. SDKs automatically perform useful tasks for you, such as:

- Cryptographically sign your service requests
- Retry requests
- Handle error responses

Refer to the following table to find general information about and code examples for each SDK, as well as the Amazon Bedrock API references for each SDK. You can also find code examples at [Code examples for Amazon Bedrock using AWS SDKs](#).

SDK documentation	Code examples	Amazon Bedrock prefix	Amazon Bedrock runtime prefix	Agents for Amazon Bedrock prefix	Agents for Amazon Bedrock runtime prefix
AWS SDK for C++	AWS SDK for C++ code examples	bedrock	bedrock-runtime	bedrock-agent	bedrock-agent-runtime

SDK documentation	Code examples	Amazon Bedrock prefix	Amazon Bedrock runtime prefix	Agents for Amazon Bedrock prefix	Agents for Amazon Bedrock runtime prefix
AWS SDK for Go	AWS SDK for Go code examples	bedrock	bedrockruntime	bedrockagent	bedrockagentruntime
AWS SDK for Java	AWS SDK for Java code examples	bedrock	bedrockruntime	bedrockagent	bedrockagentruntime
AWS SDK for JavaScript	AWS SDK for JavaScript code examples	bedrock	bedrock-runtime	bedrock-agent	bedrock-agent-runtime
AWS SDK for Kotlin	AWS SDK for Kotlin code examples	bedrock	bedrockruntime	bedrockagent	bedrockagentruntime
AWS SDK for .NET	AWS SDK for .NET code examples	Bedrock	BedrockRuntime	BedrockAgent	BedrockAgentRuntime
AWS SDK for PHP	AWS SDK for PHP code examples	Bedrock	BedrockRuntime	BedrockAgent	BedrockAgentRuntime
AWS SDK for Python (Boto3)	AWS SDK for Python (Boto3) code examples	bedrock	bedrock-runtime	bedrock-agent	bedrock-agent-runtime

SDK documentation	Code examples	Amazon Bedrock prefix	Amazon Bedrock runtime prefix	Agents for Amazon Bedrock prefix	Agents for Amazon Bedrock runtime prefix
AWS SDK for Ruby	AWS SDK for Ruby code examples	Bedrock	BedrockRuntime	BedrockAgent	BedrockAgentRuntime
AWS SDK for Rust	AWS SDK for Rust code examples	aws-sdk-bedrock	aws-sdk-bedrockruntime	aws-sdk-bedrockagent	aws-sdk-bedrockagentruntime
AWS SDK for SAP ABAP	AWS SDK for SAP ABAP code examples	BDK	BDR	BDA	BDZ
AWS SDK for Swift	AWS SDK for Swift code examples	AWSBedrock	AWSBedrockRuntime	AWSBedrockAgent	AWSBedrockAgentRuntime

Topics

- [Actions](#)
- [Data Types](#)
- [Common Parameters](#)
- [Common Errors](#)

Actions

The following actions are supported by Amazon Bedrock:

- [CreateEvaluationJob](#)
- [CreateGuardrail](#)

- [CreateGuardrailVersion](#)
- [CreateModelCustomizationJob](#)
- [CreateProvisionedModelThroughput](#)
- [DeleteCustomModel](#)
- [DeleteGuardrail](#)
- [DeleteModelInvocationLoggingConfiguration](#)
- [DeleteProvisionedModelThroughput](#)
- [GetCustomModel](#)
- [GetEvaluationJob](#)
- [GetFoundationModel](#)
- [GetGuardrail](#)
- [GetModelCustomizationJob](#)
- [GetModelInvocationLoggingConfiguration](#)
- [GetProvisionedModelThroughput](#)
- [ListCustomModels](#)
- [ListEvaluationJobs](#)
- [ListFoundationModels](#)
- [ListGuardrails](#)
- [ListModelCustomizationJobs](#)
- [ListProvisionedModelThroughputs](#)
- [ListTagsForResource](#)
- [PutModelInvocationLoggingConfiguration](#)
- [StopEvaluationJob](#)
- [StopModelCustomizationJob](#)
- [TagResource](#)
- [UntagResource](#)
- [UpdateGuardrail](#)
- [UpdateProvisionedModelThroughput](#)

The following actions are supported by Agents for Amazon Bedrock:

- [AssociateAgentKnowledgeBase](#)
- [CreateAgent](#)
- [CreateAgentActionGroup](#)
- [CreateAgentAlias](#)
- [CreateDataSource](#)
- [CreateFlow](#)
- [CreateFlowAlias](#)
- [CreateFlowVersion](#)
- [CreateKnowledgeBase](#)
- [CreatePrompt](#)
- [CreatePromptVersion](#)
- [DeleteAgent](#)
- [DeleteAgentActionGroup](#)
- [DeleteAgentAlias](#)
- [DeleteAgentVersion](#)
- [DeleteDataSource](#)
- [DeleteFlow](#)
- [DeleteFlowAlias](#)
- [DeleteFlowVersion](#)
- [DeleteKnowledgeBase](#)
- [DeletePrompt](#)
- [DisassociateAgentKnowledgeBase](#)
- [GetAgent](#)
- [GetAgentActionGroup](#)
- [GetAgentAlias](#)
- [GetAgentKnowledgeBase](#)
- [GetAgentVersion](#)
- [GetDataSource](#)
- [GetFlow](#)
- [GetFlowAlias](#)

- [GetFlowVersion](#)
- [GetIngestionJob](#)
- [GetKnowledgeBase](#)
- [GetPrompt](#)
- [ListAgentActionGroups](#)
- [ListAgentAliases](#)
- [ListAgentKnowledgeBases](#)
- [ListAgents](#)
- [ListAgentVersions](#)
- [ListDataSources](#)
- [ListFlowAliases](#)
- [ListFlows](#)
- [ListFlowVersions](#)
- [ListIngestionJobs](#)
- [ListKnowledgeBases](#)
- [ListPrompts](#)
- [ListTagsForResource](#)
- [PrepareAgent](#)
- [PrepareFlow](#)
- [StartIngestionJob](#)
- [TagResource](#)
- [UntagResource](#)
- [UpdateAgent](#)
- [UpdateAgentActionGroup](#)
- [UpdateAgentAlias](#)
- [UpdateAgentKnowledgeBase](#)
- [UpdateDataSource](#)
- [UpdateFlow](#)
- [UpdateFlowAlias](#)
- [UpdateKnowledgeBase](#)

- [UpdatePrompt](#)

The following actions are supported by Agents for Amazon Bedrock Runtime:

- [DeleteAgentMemory](#)
- [GetAgentMemory](#)
- [InvokeAgent](#)
- [InvokeFlow](#)
- [Retrieve](#)
- [RetrieveAndGenerate](#)

The following actions are supported by Amazon Bedrock Runtime:

- [ApplyGuardrail](#)
- [Converse](#)
- [ConverseStream](#)
- [InvokeModel](#)
- [InvokeModelWithResponseStream](#)

Amazon Bedrock

The following actions are supported by Amazon Bedrock:

- [CreateEvaluationJob](#)
- [CreateGuardrail](#)
- [CreateGuardrailVersion](#)
- [CreateModelCustomizationJob](#)
- [CreateProvisionedModelThroughput](#)
- [DeleteCustomModel](#)
- [DeleteGuardrail](#)
- [DeleteModelInvocationLoggingConfiguration](#)
- [DeleteProvisionedModelThroughput](#)
- [GetCustomModel](#)

- [GetEvaluationJob](#)
- [GetFoundationModel](#)
- [GetGuardrail](#)
- [GetModelCustomizationJob](#)
- [GetModelInvocationLoggingConfiguration](#)
- [GetProvisionedModelThroughput](#)
- [ListCustomModels](#)
- [ListEvaluationJobs](#)
- [ListFoundationModels](#)
- [ListGuardrails](#)
- [ListModelCustomizationJobs](#)
- [ListProvisionedModelThroughputs](#)
- [ListTagsForResource](#)
- [PutModelInvocationLoggingConfiguration](#)
- [StopEvaluationJob](#)
- [StopModelCustomizationJob](#)
- [TagResource](#)
- [UntagResource](#)
- [UpdateGuardrail](#)
- [UpdateProvisionedModelThroughput](#)

CreateEvaluationJob

Service: Amazon Bedrock

API operation for creating and managing Amazon Bedrock automatic model evaluation jobs and model evaluation jobs that use human workers. To learn more about the requirements for creating a model evaluation job see, [Model evaluation](#).

Request Syntax

```
POST /evaluation-jobs HTTP/1.1
Content-type: application/json

{
  "clientRequestToken": "string",
  "customerEncryptionKeyId": "string",
  "evaluationConfig": { ... },
  "inferenceConfig": { ... },
  "jobDescription": "string",
  "jobName": "string",
  "jobTags": [
    {
      "key": "string",
      "value": "string"
    }
  ],
  "outputDataConfig": {
    "s3Uri": "string"
  },
  "roleArn": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

clientRequestToken

A unique, case-sensitive identifier to ensure that the API request completes no more than one time. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#).

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])*$`

Required: No

customerEncryptionKeyId

Specify your customer managed key ARN that will be used to encrypt your model evaluation job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:kms:[a-zA-Z0-9-]*:[0-9]{12}:((key/[a-zA-Z0-9-]{36})|(alias/[a-zA-Z0-9-_/]+))$`

Required: No

evaluationConfig

Specifies whether the model evaluation job is automatic or uses human worker.

Type: [EvaluationConfig](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: Yes

inferenceConfig

Specify the models you want to use in your model evaluation job. Automatic model evaluation jobs support a single model, and model evaluation job that use human workers support two models.

Type: [EvaluationInferenceConfig](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: Yes

jobDescription

A description of the model evaluation job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Pattern: `^\.+`

Required: No

jobName

The name of the model evaluation job. Model evaluation job names must be unique with your AWS account, and your account's AWS region.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^[a-z0-9](-*[a-z0-9]){0,62}`

Required: Yes

jobTags

Tags to attach to the model evaluation job.

Type: Array of [Tag](#) objects

Array Members: Minimum number of 0 items. Maximum number of 200 items.

Required: No

outputDataConfig

An object that defines where the results of model evaluation job will be saved in Amazon S3.

Type: [EvaluationOutputDataConfig](#) object

Required: Yes

roleArn

The Amazon Resource Name (ARN) of an IAM service role that Amazon Bedrock can assume to perform tasks on your behalf. The service role must have Amazon Bedrock as the service principal, and provide access to any Amazon S3 buckets specified in the `EvaluationConfig` object. To pass this role to Amazon Bedrock, the caller of this API must have the `iam:PassRole` permission. To learn more about the required permissions, see [Required permissions](#).

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:iam::([0-9]{12})?:role/\.+\$`

Required: Yes

Response Syntax

```
HTTP/1.1 202
Content-type: application/json

{
  "jobArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 202 response.

The following data is returned in JSON format by the service.

jobArn

The ARN of the model evaluation job.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:([0-9]{12}):evaluation-job/[a-z0-9]{12}\$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

Error occurred because of a conflict while performing an operation.

HTTP Status Code: 400

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 400

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

CreateGuardrail

Service: Amazon Bedrock

Creates a guardrail to block topics and to implement safeguards for your generative AI applications.

You can configure the following policies in a guardrail to avoid undesirable and harmful content, filter out denied topics and words, and remove sensitive information for privacy protection.

- **Content filters** - Adjust filter strengths to block input prompts or model responses containing harmful content.
- **Denied topics** - Define a set of topics that are undesirable in the context of your application. These topics will be blocked if detected in user queries or model responses.
- **Word filters** - Configure filters to block undesirable words, phrases, and profanity. Such words can include offensive terms, competitor names etc.
- **Sensitive information filters** - Block or mask sensitive information such as personally identifiable information (PII) or custom regex in user inputs and model responses.

In addition to the above policies, you can also configure the messages to be returned to the user if a user input or model response is in violation of the policies defined in the guardrail.

For more information, see [Guardrails for Amazon Bedrock](#) in the *Amazon Bedrock User Guide*.

Request Syntax

```
POST /guardrails HTTP/1.1
Content-type: application/json

{
  "blockedInputMessaging": "string",
  "blockedOutputsMessaging": "string",
  "clientRequestToken": "string",
  "contentPolicyConfig": {
    "filtersConfig": [
      {
        "inputStrength": "string",
        "outputStrength": "string",
        "type": "string"
      }
    ]
  }
}
```

```
},
"contextualGroundingPolicyConfig": {
  "filtersConfig": [
    {
      "threshold": number,
      "type": "string"
    }
  ]
},
"description": "string",
"kmsKeyId": "string",
"name": "string",
"sensitiveInformationPolicyConfig": {
  "piiEntitiesConfig": [
    {
      "action": "string",
      "type": "string"
    }
  ],
  "regexesConfig": [
    {
      "action": "string",
      "description": "string",
      "name": "string",
      "pattern": "string"
    }
  ]
},
"tags": [
  {
    "key": "string",
    "value": "string"
  }
],
"topicPolicyConfig": {
  "topicsConfig": [
    {
      "definition": "string",
      "examples": [ "string" ],
      "name": "string",
      "type": "string"
    }
  ]
},
}
```

```
"wordPolicyConfig": {
  "managedWordListsConfig": [
    {
      "type": "string"
    }
  ],
  "wordsConfig": [
    {
      "text": "string"
    }
  ]
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

blockedInputMessaging

The message to return when the guardrail blocks a prompt.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 500.

Required: Yes

blockedOutputsMessaging

The message to return when the guardrail blocks a model response.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 500.

Required: Yes

clientRequestToken

A unique, case-sensitive identifier to ensure that the API request completes no more than once. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#) in the *Amazon S3 User Guide*.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])*$`

Required: No

contentPolicyConfig

The content filter policies to configure for the guardrail.

Type: [GuardrailContentPolicyConfig](#) object

Required: No

contextualGroundingPolicyConfig

The contextual grounding policy configuration used to create a guardrail.

Type: [GuardrailContextualGroundingPolicyConfig](#) object

Required: No

description

A description of the guardrail.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

kmsKeyId

The ARN of the AWS KMS key that you use to encrypt the guardrail.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:kms:[a-zA-Z0-9-]*:[0-9]{12}:((key/[a-zA-Z0-9-]{36})|(alias/[a-zA-Z0-9-_/]+))$`

Required: No

name

The name to give the guardrail.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 50.

Pattern: `^[0-9a-zA-Z-_$]`

Required: Yes

sensitiveInformationPolicyConfig

The sensitive information policy to configure for the guardrail.

Type: [GuardrailSensitiveInformationPolicyConfig](#) object

Required: No

tags

The tags that you want to attach to the guardrail.

Type: Array of [Tag](#) objects

Array Members: Minimum number of 0 items. Maximum number of 200 items.

Required: No

topicPolicyConfig

The topic policies to configure for the guardrail.

Type: [GuardrailTopicPolicyConfig](#) object

Required: No

wordPolicyConfig

The word policy you configure for the guardrail.

Type: [GuardrailWordPolicyConfig](#) object

Required: No

Response Syntax

```
HTTP/1.1 202
Content-type: application/json

{
  "createdAt": "string",
  "guardrailArn": "string",
  "guardrailId": "string",
  "version": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 202 response.

The following data is returned in JSON format by the service.

[createdAt](#)

The time at which the guardrail was created.

Type: Timestamp

[guardrailArn](#)

The ARN of the guardrail.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[\d-]{12}:guardrail/[a-z0-9-]+$`

[guardrailId](#)

The unique identifier of the guardrail that was created.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 64.

Pattern: `^[a-z0-9]+$`

version

The version of the guardrail that was created. This value will always be DRAFT.

Type: String

Length Constraints: Fixed length of 5.

Pattern: `^DRAFT$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

Error occurred because of a conflict while performing an operation.

HTTP Status Code: 400

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 400

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

TooManyTagsException

The request contains more tags than can be associated with a resource (50 tags per resource). The maximum number of tags includes both existing tags and those included in your current request.

HTTP Status Code: 400

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

CreateGuardrailVersion

Service: Amazon Bedrock

Creates a version of the guardrail. Use this API to create a snapshot of the guardrail when you are satisfied with a configuration, or to compare the configuration with another version.

Request Syntax

```
POST /guardrails/guardrailIdentifier HTTP/1.1
Content-type: application/json
```

```
{
  "clientRequestToken": "string",
  "description": "string"
}
```

URI Request Parameters

The request uses the following URI parameters.

guardrailIdentifier

The unique identifier of the guardrail. This can be an ID or the ARN.

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^(([a-z0-9]+)|(arn:aws(-[^\:]+)? :bedrock:[a-z0-9-]{1,20}:[0-9]{12}:guardrail/[a-z0-9]+))$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

clientRequestToken

A unique, case-sensitive identifier to ensure that the API request completes no more than once. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#) in the *Amazon S3 User Guide*.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])*$`

Required: No

description

A description of the guardrail version.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

Response Syntax

```
HTTP/1.1 202
Content-type: application/json

{
  "guardrailId": "string",
  "version": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 202 response.

The following data is returned in JSON format by the service.

guardrailId

The unique identifier of the guardrail.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 64.

Pattern: `^[a-z0-9]+$`

version

The number of the version of the guardrail.

Type: String

Pattern: `^[1-9][0-9]{0,7}$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

Error occurred because of a conflict while performing an operation.

HTTP Status Code: 400

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 400

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

CreateModelCustomizationJob

Service: Amazon Bedrock

Creates a fine-tuning job to customize a base model.

You specify the base foundation model and the location of the training data. After the model-customization job completes successfully, your custom model resource will be ready to use. Amazon Bedrock returns validation loss metrics and output generations after the job completes.

For information on the format of training and validation data, see [Prepare the datasets](#).

Model-customization jobs are asynchronous and the completion time depends on the base model and the training/validation data size. To monitor a job, use the `GetModelCustomizationJob` operation to retrieve the job status.

For more information, see [Custom models](#) in the Amazon Bedrock User Guide.

Request Syntax

```
POST /model-customization-jobs HTTP/1.1
```

```
Content-type: application/json
```

```
{
  "baseModelIdentifier": "string",
  "clientRequestToken": "string",
  "customizationType": "string",
  "customModelKmsKeyId": "string",
  "customModelName": "string",
  "customModelTags": [
    {
      "key": "string",
      "value": "string"
    }
  ],
  "hyperParameters": {
    "string" : "string"
  },
  "jobName": "string",
  "jobTags": [
    {
      "key": "string",
      "value": "string"
    }
  ]
}
```

```

],
"outputDataConfig": {
  "s3Uri": "string"
},
"roleArn": "string",
"trainingDataConfig": {
  "s3Uri": "string"
},
"validationDataConfig": {
  "validators": [
    {
      "s3Uri": "string"
    }
  ]
},
"vpcConfig": {
  "securityGroupIds": [ "string" ],
  "subnetIds": [ "string" ]
}
}

```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

baseModelIdentifier

Name of the base model.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^(arn:aws(-[^\:]+)?):bedrock:[a-z0-9-]{1,20}:((([0-9]{12}:custom-model/([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2})/[a-z0-9]{12})|(:foundation-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([:][a-z0-9-]{1,63}){0,2}))|([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2})|((([0-9a-zA-Z][_-]?)+))$`

Required: Yes

clientRequestToken

A unique, case-sensitive identifier to ensure that the API request completes no more than one time. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#).

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])*$`

Required: No

customizationType

The customization type.

Type: String

Valid Values: FINE_TUNING | CONTINUED_PRE_TRAINING

Required: No

customModelKmsKeyId

The custom model is encrypted at rest using this key.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:kms:[a-zA-Z0-9-]*:[0-9]{12}:((key/[a-zA-Z0-9-]{36})|(alias/[a-zA-Z0-9-_/]+))$`

Required: No

customModelName

A name for the resulting custom model.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^([\0-9a-zA-Z][_-]?)+$`

Required: Yes

customModelTags

Tags to attach to the resulting custom model.

Type: Array of [Tag](#) objects

Array Members: Minimum number of 0 items. Maximum number of 200 items.

Required: No

hyperParameters

Parameters related to tuning the model. For details on the format for different models, see [Custom model hyperparameters](#).

Type: String to string map

Required: Yes

jobName

A name for the fine-tuning job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9\+\-\.\])*$`

Required: Yes

jobTags

Tags to attach to the job.

Type: Array of [Tag](#) objects

Array Members: Minimum number of 0 items. Maximum number of 200 items.

Required: No

[outputDataConfig](#)

S3 location for the output data.

Type: [OutputDataConfig](#) object

Required: Yes

[roleArn](#)

The Amazon Resource Name (ARN) of an IAM service role that Amazon Bedrock can assume to perform tasks on your behalf. For example, during model training, Amazon Bedrock needs your permission to read input data from an S3 bucket, write model artifacts to an S3 bucket. To pass this role to Amazon Bedrock, the caller of this API must have the `iam:PassRole` permission.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:iam::([0-9]{12})?:role/\.+\$`

Required: Yes

[trainingDataConfig](#)

Information about the training dataset.

Type: [TrainingDataConfig](#) object

Required: Yes

[validationDataConfig](#)

Information about the validation dataset.

Type: [ValidationDataConfig](#) object

Required: No

[vpcConfig](#)

VPC configuration (optional). Configuration parameters for the private Virtual Private Cloud (VPC) that contains the resources you are using for this job.

Type: [VpcConfig](#) object

Required: No

Response Syntax

```
HTTP/1.1 201
Content-type: application/json

{
  "jobArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 201 response.

The following data is returned in JSON format by the service.

jobArn

Amazon Resource Name (ARN) of the fine tuning job

Type: String

Length Constraints: Minimum length of 0. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:model-customization-job/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([a-z0-9-]{1,63}[.]){0,2}[a-z0-9-]{1,63}([:][a-z0-9-]{1,63}){0,2}/[a-z0-9]{12}$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

Error occurred because of a conflict while performing an operation.

HTTP Status Code: 400

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 400

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

TooManyTagsException

The request contains more tags than can be associated with a resource (50 tags per resource). The maximum number of tags includes both existing tags and those included in your current request.

HTTP Status Code: 400

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

CreateProvisionedModelThroughput

Service: Amazon Bedrock

Creates dedicated throughput for a base or custom model with the model units and for the duration that you specify. For pricing details, see [Amazon Bedrock Pricing](#). For more information, see [Provisioned Throughput](#) in the Amazon Bedrock User Guide.

Request Syntax

```
POST /provisioned-model-throughput HTTP/1.1
```

```
Content-type: application/json
```

```
{
  "clientRequestToken": "string",
  "commitmentDuration": "string",
  "modelId": "string",
  "modelUnits": number,
  "provisionedModelName": "string",
  "tags": [
    {
      "key": "string",
      "value": "string"
    }
  ]
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

clientRequestToken

A unique, case-sensitive identifier to ensure that the API request completes no more than one time. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#) in the Amazon S3 User Guide.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])*$`

Required: No

commitmentDuration

The commitment duration requested for the Provisioned Throughput. Billing occurs hourly and is discounted for longer commitment terms. To request a no-commit Provisioned Throughput, omit this field.

Custom models support all levels of commitment. To see which base models support no commitment, see [Supported regions and models for Provisioned Throughput](#) in the Amazon Bedrock User Guide

Type: String

Valid Values: OneMonth | SixMonths

Required: No

modelId

The Amazon Resource Name (ARN) or name of the model to associate with this Provisioned Throughput. For a list of models for which you can purchase Provisioned Throughput, see [Amazon Bedrock model IDs for purchasing Provisioned Throughput](#) in the Amazon Bedrock User Guide.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:((\[[0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}(([:][a-z0-9-]{1,63}){0,2})?/[a-z0-9]{12})|(:foundation-model/([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2})))|((\[[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2}))|((\[[0-9a-zA-Z][_]?)+))$`

Required: Yes

modelUnits

Number of model units to allocate. A model unit delivers a specific throughput level for the specified model. The throughput level of a model unit specifies the total number of input and output tokens that it can process and generate within a span of one minute. By default, your account has no model units for purchasing Provisioned Throughputs with commitment. You must first visit the [AWS support center](#) to request MUs.

For model unit quotas, see [Provisioned Throughput quotas](#) in the Amazon Bedrock User Guide.

For more information about what an MU specifies, contact your AWS account manager.

Type: Integer

Valid Range: Minimum value of 1.

Required: Yes

provisionedModelName

The name for this Provisioned Throughput.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: $^([\text{0-9a-zA-Z}][\text{-}]?)\text{\$}$

Required: Yes

tags

Tags to associate with this Provisioned Throughput.

Type: Array of [Tag](#) objects

Array Members: Minimum number of 0 items. Maximum number of 200 items.

Required: No

Response Syntax

```
HTTP/1.1 201
```

```
Content-type: application/json

{
  "provisionedModelArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 201 response.

The following data is returned in JSON format by the service.

provisionedModelArn

The Amazon Resource Name (ARN) for this Provisioned Throughput.

Type: String

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:provisioned-model/[a-z0-9]{12}$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 400

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

TooManyTagsException

The request contains more tags than can be associated with a resource (50 tags per resource). The maximum number of tags includes both existing tags and those included in your current request.

HTTP Status Code: 400

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

DeleteCustomModel

Service: Amazon Bedrock

Deletes a custom model that you created earlier. For more information, see [Custom models](#) in the Amazon Bedrock User Guide.

Request Syntax

```
DELETE /custom-models/modelIdentifier HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

modelIdentifier

Name of the model to delete.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}(([:][a-z0-9-]{1,63}){0,2})?/[a-z0-9]{12})|(:foundation-model/([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2})))|((([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2}))|((([0-9a-zA-Z][_]?)+))$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

Error occurred because of a conflict while performing an operation.

HTTP Status Code: 400

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

DeleteGuardrail

Service: Amazon Bedrock

Deletes a guardrail.

- To delete a guardrail, only specify the ARN of the guardrail in the `guardrailIdentifier` field. If you delete a guardrail, all of its versions will be deleted.
- To delete a version of a guardrail, specify the ARN of the guardrail in the `guardrailIdentifier` field and the version in the `guardrailVersion` field.

Request Syntax

```
DELETE /guardrails/guardrailIdentifier?guardrailVersion=guardrailVersion HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

guardrailIdentifier

The unique identifier of the guardrail. This can be an ID or the ARN.

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^(([a-z0-9]+)|(arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:guardrail/[a-z0-9]+))$`

Required: Yes

guardrailVersion

The version of the guardrail.

Pattern: `^[1-9][0-9]{0,7}$`

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 202
```


Response Elements

If the action is successful, the service sends back an HTTP 202 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

Error occurred because of a conflict while performing an operation.

HTTP Status Code: 400

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

DeleteModelInvocationLoggingConfiguration

Service: Amazon Bedrock

Delete the invocation logging.

Request Syntax

```
DELETE /logging/modelinvocations HTTP/1.1
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

DeleteProvisionedModelThroughput

Service: Amazon Bedrock

Deletes a Provisioned Throughput. You can't delete a Provisioned Throughput before the commitment term is over. For more information, see [Provisioned Throughput](#) in the Amazon Bedrock User Guide.

Request Syntax

```
DELETE /provisioned-model-throughput/provisionedModelId HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

provisionedModelId

The Amazon Resource Name (ARN) or name of the Provisioned Throughput.

Pattern: `^((([0-9a-zA-Z] [_-]?)+) | (arn:aws(- [^:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:provisioned-model/[a-z0-9]{12}))$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

Error occurred because of a conflict while performing an operation.

HTTP Status Code: 400

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)

- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetCustomModel

Service: Amazon Bedrock

Get the properties associated with a Amazon Bedrock custom model that you have created. For more information, see [Custom models](#) in the Amazon Bedrock User Guide.

Request Syntax

```
GET /custom-models/modelIdentifier HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

modelIdentifier

Name or Amazon Resource Name (ARN) of the custom model.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}(([:][a-z0-9-]{1,63}){0,2})?/[a-z0-9-]{12})|(:foundation-model/([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]{1}[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2})))|((([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]{1}[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2}))|((([0-9a-zA-Z][_]?)+))$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "baseModelArn": "string",
```



```

"creationTime": "string",
"customizationType": "string",
"hyperParameters": {
  "string" : "string"
},
"jobArn": "string",
"jobName": "string",
"modelArn": "string",
"modelKmsKeyArn": "string",
"modelName": "string",
"outputDataConfig": {
  "s3Uri": "string"
},
"trainingDataConfig": {
  "s3Uri": "string"
},
"trainingMetrics": {
  "trainingLoss": number
},
"validationDataConfig": {
  "validators": [
    {
      "s3Uri": "string"
    }
  ]
},
"validationMetrics": [
  {
    "validationLoss": number
  }
]
}

```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

baseModelArn

Amazon Resource Name (ARN) of the base model.

Type: String

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}/[a-z0-9]{12})|(:foundation-model/[a-z0-9-]{1,63}[.]{1}([a-z0-9-]{1,63}[.]){0,2}[a-z0-9-]{1,63}([:][a-z0-9-]{1,63}){0,2}))$`

creationTime

Creation time of the model.

Type: Timestamp

customizationType

The type of model customization.

Type: String

Valid Values: FINE_TUNING | CONTINUED_PRE_TRAINING

hyperParameters

Hyperparameter values associated with this model. For details on the format for different models, see [Custom model hyperparameters](#).

Type: String to string map

jobArn

Job Amazon Resource Name (ARN) associated with this model.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:([0-9]{12}:model-customization-job/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([a-z0-9-]{1,63}[.]){0,2}[a-z0-9-]{1,63}([:][a-z0-9-]{1,63}){0,2}/[a-z0-9]{12})$`

jobName

Job name associated with this model.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9\+\-\.\.])*$`

modelArn

Amazon Resource Name (ARN) associated with this model.

Type: String

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:((([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}/[a-z0-9]{12})|(:foundation-model/[a-z0-9-]{1,63}[.]{1}([a-z0-9-]{1,63}[.]{1}){0,2}[a-z0-9-]{1,63}([:][a-z0-9-]{1,63}){0,2}))$`

modelKmsKeyArn

The custom model is encrypted at rest using this key.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

modelName

Model name associated with this model.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^([0-9a-zA-Z][_-]?)+$`

outputDataConfig

Output data configuration associated with this custom model.

Type: [OutputDataConfig](#) object

trainingDataConfig

Contains information about the training dataset.

Type: [TrainingDataConfig](#) object

[trainingMetrics](#)

Contains training metrics from the job creation.

Type: [TrainingMetrics](#) object

[validationDataConfig](#)

Contains information about the validation dataset.

Type: [ValidationDataConfig](#) object

[validationMetrics](#)

The validation metrics from the job creation.

Type: Array of [ValidatorMetric](#) objects

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetEvaluationJob

Service: Amazon Bedrock

Retrieves the properties associated with a model evaluation job, including the status of the job. For more information, see [Model evaluation](#).

Request Syntax

```
GET /evaluation-jobs/jobIdentifier HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

jobIdentifier

The Amazon Resource Name (ARN) of the model evaluation job.

Length Constraints: Minimum length of 0. Maximum length of 1011.

Pattern: `^(arn:aws(-[^\:]+)?):bedrock:[a-z0-9-]{1,20}:[0-9]{12}:evaluation-job/[a-z0-9]{12})$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "creationTime": "string",
  "customerEncryptionKeyId": "string",
  "evaluationConfig": { ... },
  "failureMessages": [ "string" ],
  "inferenceConfig": { ... },
  "jobArn": "string",
  "jobDescription": "string",
```

```
"jobName": "string",
"jobType": "string",
"lastModifiedTime": "string",
"outputDataConfig": {
  "s3Uri": "string"
},
"roleArn": "string",
"status": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

creationTime

When the model evaluation job was created.

Type: Timestamp

customerEncryptionKeyId

The Amazon Resource Name (ARN) of the customer managed key specified when the model evaluation job was created.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[:]+)?:kms:[a-zA-Z0-9-]*:[0-9]{12}:((key/[a-zA-Z0-9-]{36})|(alias/[a-zA-Z0-9-_/]+))$`

evaluationConfig

Contains details about the type of model evaluation job, the metrics used, the task type selected, the datasets used, and any custom metrics you defined.

Type: [EvaluationConfig](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

failureMessages

An array of strings the specify why the model evaluation job has failed.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 20 items.

Length Constraints: Minimum length of 0. Maximum length of 2048.

[inferenceConfig](#)

Details about the models you specified in your model evaluation job.

Type: [EvaluationInferenceConfig](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

[jobArn](#)

The Amazon Resource Name (ARN) of the model evaluation job.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:evaluation-job/[a-z0-9]{12}$`

[jobDescription](#)

The description of the model evaluation job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Pattern: `^\.+`\$

[jobName](#)

The name of the model evaluation job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^[a-z0-9](-*[a-z0-9]){0,62}$`

[jobType](#)

The type of model evaluation job.

Type: String

Valid Values: Human | Automated

lastModifiedTime

When the model evaluation job was last modified.

Type: Timestamp

outputDataConfig

Amazon S3 location for where output data is saved.

Type: [EvaluationOutputDataConfig](#) object

roleArn

The Amazon Resource Name (ARN) of the IAM service role used in the model evaluation job.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:iam::([0-9]{12})?:role/\.+\$`

status

The status of the model evaluation job.

Type: String

Valid Values: InProgress | Completed | Failed | Stopping | Stopped

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetFoundationModel

Service: Amazon Bedrock

Get details about a Amazon Bedrock foundation model.

Request Syntax

```
GET /foundation-models/modelIdentifier HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

modelIdentifier

The model identifier.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[:]+)?:bedrock:[a-z0-9-]{1,20}:(([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}(([:][a-z0-9-]{1,63}){0,2})?/[a-z0-9]{12})|(:foundation-model/([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2})))|(([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2}))|(([0-9a-zA-Z][_]?)+)$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "modelDetails": {
    "customizationsSupported": [ "string" ],
    "inferenceTypesSupported": [ "string" ],
    "inputModalities": [ "string" ],
```

```
"modelArn": "string",
"modelId": "string",
"modelLifecycle": {
  "status": "string"
},
"modelName": "string",
"outputModalities": [ "string" ],
"providerName": "string",
"responseStreamingSupported": boolean
}
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

modelDetails

Information about the foundation model.

Type: [FoundationModelDetails](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetGuardrail

Service: Amazon Bedrock

Gets details about a guardrail. If you don't specify a version, the response returns details for the DRAFT version.

Request Syntax

```
GET /guardrails/guardrailIdentifier?guardrailVersion=guardrailVersion HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

guardrailIdentifier

The unique identifier of the guardrail for which to get details. This can be an ID or the ARN.

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^(([a-z0-9]+)|(arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:guardrail/[a-z0-9]+))$`

Required: Yes

guardrailVersion

The version of the guardrail for which to get details. If you don't specify a version, the response returns details for the DRAFT version.

Pattern: `^(([1-9][0-9]{0,7})|(DRAFT))$`

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "blockedInputMessaging": "string",
  "blockedOutputsMessaging": "string",
```

```
"contentPolicy": {
  "filters": [
    {
      "inputStrength": "string",
      "outputStrength": "string",
      "type": "string"
    }
  ]
},
"contextualGroundingPolicy": {
  "filters": [
    {
      "threshold": number,
      "type": "string"
    }
  ]
},
"createdAt": "string",
"description": "string",
"failureRecommendations": [ "string" ],
"guardrailArn": "string",
"guardrailId": "string",
"kmsKeyArn": "string",
"name": "string",
"sensitiveInformationPolicy": {
  "piiEntities": [
    {
      "action": "string",
      "type": "string"
    }
  ],
  "regexes": [
    {
      "action": "string",
      "description": "string",
      "name": "string",
      "pattern": "string"
    }
  ]
},
"status": "string",
"statusReasons": [ "string" ],
"topicPolicy": {
  "topics": [
```

```

    {
      "definition": "string",
      "examples": [ "string" ],
      "name": "string",
      "type": "string"
    }
  ]
},
"updatedAt": "string",
"version": "string",
"wordPolicy": {
  "managedWordLists": [
    {
      "type": "string"
    }
  ],
  "words": [
    {
      "text": "string"
    }
  ]
}
}

```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

[blockedInputMessaging](#)

The message that the guardrail returns when it blocks a prompt.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 500.

[blockedOutputsMessaging](#)

The message that the guardrail returns when it blocks a model response.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 500.

contentPolicy

The content policy that was configured for the guardrail.

Type: [GuardrailContentPolicy](#) object

contextualGroundingPolicy

The contextual grounding policy used in the guardrail.

Type: [GuardrailContextualGroundingPolicy](#) object

createdAt

The date and time at which the guardrail was created.

Type: Timestamp

description

The description of the guardrail.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

failureRecommendations

Appears if the status of the guardrail is FAILED. A list of recommendations to carry out before retrying the request.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 100 items.

Length Constraints: Minimum length of 1. Maximum length of 200.

guardrailArn

The ARN of the guardrail.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:guardrail/[a-z0-9]+$`

guardrailId

The unique identifier of the guardrail.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 64.

Pattern: `^[a-z0-9]+$`

kmsKeyArn

The ARN of the AWS KMS key that encrypts the guardrail.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[[:^:]]+)?:kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

name

The name of the guardrail.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 50.

Pattern: `^[0-9a-zA-Z-_$]+`

sensitiveInformationPolicy

The sensitive information policy that was configured for the guardrail.

Type: [GuardrailSensitiveInformationPolicy](#) object

status

The status of the guardrail.

Type: String

Valid Values: CREATING | UPDATING | VERSIONING | READY | FAILED | DELETING

statusReasons

Appears if the status is FAILED. A list of reasons for why the guardrail failed to be created, updated, versioned, or deleted.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 100 items.

Length Constraints: Minimum length of 1. Maximum length of 200.

topicPolicy

The topic policy that was configured for the guardrail.

Type: [GuardrailTopicPolicy](#) object

updatedAt

The date and time at which the guardrail was updated.

Type: Timestamp

version

The version of the guardrail.

Type: String

Pattern: `^(([1-9][0-9]{0,7})|(DRAFT))$`

wordPolicy

The word policy that was configured for the guardrail.

Type: [GuardrailWordPolicy](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetModelCustomizationJob

Service: Amazon Bedrock

Retrieves the properties associated with a model-customization job, including the status of the job. For more information, see [Custom models](#) in the Amazon Bedrock User Guide.

Request Syntax

```
GET /model-customization-jobs/jobIdentifier HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

jobIdentifier

Identifier for the customization job.

Length Constraints: Minimum length of 0. Maximum length of 1011.

Pattern: `^(arn:aws(-[^\:]+)??:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:model-customization-job/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([a-z0-9-]{1,63}[.]){0,2}[a-z0-9-]{1,63}(:[a-z0-9-]{1,63}){0,2}/[a-z0-9]{12})|([a-zA-Z0-9](-*[a-zA-Z0-9\+\-\.\.])*)$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "baseModelArn": "string",
  "clientRequestToken": "string",
  "creationTime": "string",
  "customizationType": "string",
  "endTime": "string",
```

```

"failureMessage": "string",
"hyperParameters": {
  "string" : "string"
},
"jobArn": "string",
"jobName": "string",
"lastModifiedTime": "string",
"outputDataConfig": {
  "s3Uri": "string"
},
"outputModelArn": "string",
"outputModelKmsKeyArn": "string",
"outputModelName": "string",
"roleArn": "string",
"status": "string",
"trainingDataConfig": {
  "s3Uri": "string"
},
"trainingMetrics": {
  "trainingLoss": number
},
"validationDataConfig": {
  "validators": [
    {
      "s3Uri": "string"
    }
  ]
},
"validationMetrics": [
  {
    "validationLoss": number
  }
],
"vpcConfig": {
  "securityGroupIds": [ "string" ],
  "subnetIds": [ "string" ]
}
}

```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

baseModelArn

Amazon Resource Name (ARN) of the base model.

Type: String

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}::foundation-model/[a-z0-9-]{1,63}[\.]{1}([a-z0-9-]{1,63}[\.]){0,2}[a-z0-9-]{1,63}(:[a-z0-9-]{1,63}){0,2}$`

clientRequestToken

The token that you specified in the `CreateCustomizationJob` request.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])*$`

creationTime

Time that the resource was created.

Type: Timestamp

customizationType

The type of model customization.

Type: String

Valid Values: FINE_TUNING | CONTINUED_PRE_TRAINING

endTime

Time that the resource transitioned to terminal state.

Type: Timestamp

failureMessage

Information about why the job failed.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

hyperParameters

The hyperparameter values for the job. For details on the format for different models, see [Custom model hyperparameters](#).

Type: String to string map

jobArn

The Amazon Resource Name (ARN) of the customization job.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:model-customization-job/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([a-z0-9-]{1,63}[.]){0,2}[a-z0-9-]{1,63}(:[a-z0-9-]{1,63}){0,2}/[a-z0-9]{12}$`

jobName

The name of the customization job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9\+\-\.\.])*$`

lastModifiedTime

Time that the resource was last modified.

Type: Timestamp

outputDataConfig

Output data configuration

Type: [OutputDataConfig](#) object

outputModelArn

The Amazon Resource Name (ARN) of the output model.

Type: String

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:custom-model/[a-z0-9-]{1,63}[\.]{1}[a-z0-9-]{1,63}([a-z0-9-]{1,63}[\.]){0,2}[a-z0-9-]{1,63}(:[a-z0-9-]{1,63}){0,2}/[a-z0-9]{12}$`

outputModelKmsKeyArn

The custom model is encrypted at rest using this key.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

outputModelName

The name of the output model.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^([0-9a-zA-Z][_ -]?)+$`

roleArn

The Amazon Resource Name (ARN) of the IAM role.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:iam::([0-9]{12})?:role/.*+$`

status

The status of the job. A successful job transitions from in-progress to completed when the output model is ready to use. If the job failed, the failure message contains information about why the job failed.

Type: String

Valid Values: `InProgress` | `Completed` | `Failed` | `Stopping` | `Stopped`

[trainingDataConfig](#)

Contains information about the training dataset.

Type: [TrainingDataConfig](#) object

[trainingMetrics](#)

Contains training metrics from the job creation.

Type: [TrainingMetrics](#) object

[validationDataConfig](#)

Contains information about the validation dataset.

Type: [ValidationDataConfig](#) object

[validationMetrics](#)

The loss metric for each validator that you provided in the createjob request.

Type: Array of [ValidatorMetric](#) objects

[vpcConfig](#)

VPC configuration for the custom model job.

Type: [VpcConfig](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetModelInvocationLoggingConfiguration

Service: Amazon Bedrock

Get the current configuration values for model invocation logging.

Request Syntax

```
GET /logging/modelinvocations HTTP/1.1
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "loggingConfig": {
    "cloudWatchConfig": {
      "largeDataDeliveryS3Config": {
        "bucketName": "string",
        "keyPrefix": "string"
      },
      "logGroupName": "string",
      "roleArn": "string"
    },
    "embeddingDataDeliveryEnabled": boolean,
    "imageDataDeliveryEnabled": boolean,
    "s3Config": {
      "bucketName": "string",
      "keyPrefix": "string"
    },
    "textDataDeliveryEnabled": boolean
  }
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

[loggingConfig](#)

The current configuration values.

Type: [LoggingConfig](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)

- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetProvisionedModelThroughput

Service: Amazon Bedrock

Returns details for a Provisioned Throughput. For more information, see [Provisioned Throughput](#) in the Amazon Bedrock User Guide.

Request Syntax

```
GET /provisioned-model-throughput/provisionedModelId HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

provisionedModelId

The Amazon Resource Name (ARN) or name of the Provisioned Throughput.

Pattern: `^((([0-9a-zA-Z] [_-]?) +) | (arn:aws(- [^:]+)? :bedrock:[a-z0-9-]{1,20} : [0-9]{12} :provisioned-model/[a-z0-9]{12})))$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "commitmentDuration": "string",
  "commitmentExpirationTime": "string",
  "creationTime": "string",
  "desiredModelArn": "string",
  "desiredModelUnits": number,
  "failureMessage": "string",
  "foundationModelArn": "string",
  "lastModifiedTime": "string",
  "modelArn": "string",
```

```

    "modelUnits": number,
    "provisionedModelArn": "string",
    "provisionedModelName": "string",
    "status": "string"
  }

```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

commitmentDuration

Commitment duration of the Provisioned Throughput.

Type: String

Valid Values: OneMonth | SixMonths

commitmentExpirationTime

The timestamp for when the commitment term for the Provisioned Throughput expires.

Type: Timestamp

creationTime

The timestamp of the creation time for this Provisioned Throughput.

Type: Timestamp

desiredModelArn

The Amazon Resource Name (ARN) of the model requested to be associated to this Provisioned Throughput. This value differs from the `modelArn` if updating hasn't completed.

Type: String

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}/[a-z0-9]{12})|(:foundation-model/[a-z0-9-]{1,63}[.]{1}([a-z0-9-]{1,63}[.]{1}){0,2}[a-z0-9-]{1,63}([:][a-z0-9-]{1,63}){0,2}))$`

desiredModelUnits

The number of model units that was requested for this Provisioned Throughput.

Type: Integer

Valid Range: Minimum value of 1.

failureMessage

A failure message for any issues that occurred during creation, updating, or deletion of the Provisioned Throughput.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

foundationModelArn

The Amazon Resource Name (ARN) of the base model for which the Provisioned Throughput was created, or of the base model that the custom model for which the Provisioned Throughput was created was customized.

Type: String

Pattern: `^arn:aws(-[:]+)?:bedrock:[a-z0-9-]{1,20}::foundation-model/[a-z0-9-]{1,63}[.]{1}([a-z0-9-]{1,63}[.]){0,2}[a-z0-9-]{1,63}(:[a-z0-9-]{1,63}){0,2}$`

lastModifiedTime

The timestamp of the last time that this Provisioned Throughput was modified.

Type: Timestamp

modelArn

The Amazon Resource Name (ARN) of the model associated with this Provisioned Throughput.

Type: String

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: `^arn:aws(-[:]+)?:bedrock:[a-z0-9-]{1,20}:(([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}/[a-z0-9-]{12})|(:foundation-`

```
model/[a-z0-9-]{1,63}[.]{1}([a-z0-9-]{1,63}[.]){0,2}[a-z0-9-]{1,63}([:]
[a-z0-9-]{1,63}){0,2}))$
```

modelUnits

The number of model units allocated to this Provisioned Throughput.

Type: Integer

Valid Range: Minimum value of 1.

provisionedModelArn

The Amazon Resource Name (ARN) of the Provisioned Throughput.

Type: String

Pattern: `^arn:aws(-[:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:provisioned-model/[a-z0-9]{12}$`

provisionedModelName

The name of the Provisioned Throughput.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^([0-9a-zA-Z][_]?)+$`

status

The status of the Provisioned Throughput.

Type: String

Valid Values: `Creating | InService | Updating | Failed`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListCustomModels

Service: Amazon Bedrock

Returns a list of the custom models that you have created with the `CreateModelCustomizationJob` operation.

For more information, see [Custom models](#) in the Amazon Bedrock User Guide.

Request Syntax

```
GET /custom-models?
baseModelArnEquals=baseModelArnEquals&creationTimeAfter=creationTimeAfter&creationTimeBefore=creationTimeBefore
HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

[baseModelArnEquals](#)

Return custom models only if the base model Amazon Resource Name (ARN) matches this parameter.

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}/[a-z0-9]{12})|(:foundation-model/[a-z0-9-]{1,63}[.]{1}([a-z0-9-]{1,63}[.]{0,2}[a-z0-9-]{1,63}([[:a-z0-9-]{1,63}){0,2})))$`

[creationTimeAfter](#)

Return custom models created after the specified time.

[creationTimeBefore](#)

Return custom models created before the specified time.

[foundationModelArnEquals](#)

Return custom models only if the foundation model Amazon Resource Name (ARN) matches this parameter.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}::foundation-model/[a-z0-9-]{1,63}[.]{1}([a-z0-9-]{1,63}[.]){0,2}[a-z0-9-]{1,63}([:][a-z0-9-]{1,63}){0,2}$`

maxResults

Maximum number of results to return in the response.

Valid Range: Minimum value of 1. Maximum value of 1000.

nameContains

Return custom models only if the job name contains these characters.

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^([0-9a-zA-Z][_]?)+$`

nextToken

Continuation token from the previous response, for Amazon Bedrock to list the next set of results.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

sortBy

The field to sort by in the returned list of models.

Valid Values: `CreationTime`

sortOrder

The sort order of the results.

Valid Values: `Ascending` | `Descending`

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json
```

```
{
  "modelSummaries": [
    {
      "baseModelArn": "string",
      "baseModelName": "string",
      "creationTime": "string",
      "customizationType": "string",
      "modelArn": "string",
      "modelName": "string"
    }
  ],
  "nextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

modelSummaries

Model summaries.

Type: Array of [CustomModelSummary](#) objects

nextToken

Continuation token for the next request to list the next set of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: $^\backslash S^*\$$

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListEvaluationJobs

Service: Amazon Bedrock

Lists model evaluation jobs.

Request Syntax

```
GET /evaluation-jobs?  
creationTimeAfter=creationTimeAfter&creationTimeBefore=creationTimeBefore&maxResults=maxResults  
HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

[creationTimeAfter](#)

A filter that includes model evaluation jobs created after the time specified.

[creationTimeBefore](#)

A filter that includes model evaluation jobs created prior to the time specified.

[maxResults](#)

The maximum number of results to return.

Valid Range: Minimum value of 1. Maximum value of 1000.

[nameContains](#)

Query parameter string for model evaluation job names.

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^[a-z0-9](-*[a-z0-9]){0,62}$`

[nextToken](#)

Continuation token from the previous response, for Amazon Bedrock to list the next set of results.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

sortBy

Allows you to sort model evaluation jobs by when they were created.

Valid Values: `CreationTime`

sortOrder

How you want the order of jobs sorted.

Valid Values: `Ascending` | `Descending`

statusEquals

Only return jobs where the status condition is met.

Valid Values: `InProgress` | `Completed` | `Failed` | `Stopping` | `Stopped`

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "jobSummaries": [
    {
      "creationTime": "string",
      "evaluationTaskTypes": [ "string" ],
      "jobArn": "string",
      "jobName": "string",
      "jobType": "string",
      "modelIdentifiers": [ "string" ],
      "status": "string"
    }
  ],
  "nextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

jobSummaries

A summary of the model evaluation jobs.

Type: Array of [EvaluationSummary](#) objects

Array Members: Minimum number of 1 item. Maximum number of 5 items.

nextToken

Continuation token from the previous response, for Amazon Bedrock to list the next set of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListFoundationModels

Service: Amazon Bedrock

Lists Amazon Bedrock foundation models that you can use. You can filter the results with the request parameters. For more information, see [Foundation models](#) in the Amazon Bedrock User Guide.

Request Syntax

```
GET /foundation-models?  
byCustomizationType=byCustomizationType&byInferenceType=byInferenceType&byOutputModality=byOutputModality  
HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

[byCustomizationType](#)

Return models that support the customization type that you specify. For more information, see [Custom models](#) in the Amazon Bedrock User Guide.

Valid Values: FINE_TUNING | CONTINUED_PRE_TRAINING

[byInferenceType](#)

Return models that support the inference type that you specify. For more information, see [Provisioned Throughput](#) in the Amazon Bedrock User Guide.

Valid Values: ON_DEMAND | PROVISIONED

[byOutputModality](#)

Return models that support the output modality that you specify.

Valid Values: TEXT | IMAGE | EMBEDDING

[byProvider](#)

Return models belonging to the model provider that you specify.

Pattern: `^[A-Za-z0-9-]{1,63}$`

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "modelSummaries": [
    {
      "customizationsSupported": [ "string" ],
      "inferenceTypesSupported": [ "string" ],
      "inputModalities": [ "string" ],
      "modelArn": "string",
      "modelId": "string",
      "modelLifecycle": {
        "status": "string"
      },
      "modelName": "string",
      "outputModalities": [ "string" ],
      "providerName": "string",
      "responseStreamingSupported": boolean
    }
  ]
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

modelSummaries

A list of Amazon Bedrock foundation models.

Type: Array of [FoundationModelSummary](#) objects

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListGuardrails

Service: Amazon Bedrock

Lists details about all the guardrails in an account. To list the DRAFT version of all your guardrails, don't specify the `guardrailIdentifier` field. To list all versions of a guardrail, specify the ARN of the guardrail in the `guardrailIdentifier` field.

You can set the maximum number of results to return in a response in the `maxResults` field. If there are more results than the number you set, the response returns a `nextToken` that you can send in another `ListGuardrails` request to see the next batch of results.

Request Syntax

```
GET /guardrails?  
guardrailIdentifier=guardrailIdentifier&maxResults=maxResults&nextToken=nextToken  
HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

guardrailIdentifier

The unique identifier of the guardrail. This can be an ID or the ARN.

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^(([a-z0-9]+)|(arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:guardrail/[a-z0-9]+))$`

maxResults

The maximum number of results to return in the response.

Valid Range: Minimum value of 1. Maximum value of 1000.

nextToken

If there are more results than were returned in the response, the response returns a `nextToken` that you can send in another `ListGuardrails` request to see the next batch of results.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "guardrails": [
    {
      "arn": "string",
      "createdAt": "string",
      "description": "string",
      "id": "string",
      "name": "string",
      "status": "string",
      "updatedAt": "string",
      "version": "string"
    }
  ],
  "nextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

guardrails

A list of objects, each of which contains details about a guardrail.

Type: Array of [GuardrailSummary](#) objects

Array Members: Minimum number of 0 items. Maximum number of 1000 items.

nextToken

If there are more results than were returned in the response, the response returns a `nextToken` that you can send in another `ListGuardrails` request to see the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListModelCustomizationJobs

Service: Amazon Bedrock

Returns a list of model customization jobs that you have submitted. You can filter the jobs to return based on one or more criteria.

For more information, see [Custom models](#) in the Amazon Bedrock User Guide.

Request Syntax

```
GET /model-customization-jobs?  
creationTimeAfter=creationTimeAfter&creationTimeBefore=creationTimeBefore&maxResults=maxResults  
HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

[creationTimeAfter](#)

Return customization jobs created after the specified time.

[creationTimeBefore](#)

Return customization jobs created before the specified time.

[maxResults](#)

Maximum number of results to return in the response.

Valid Range: Minimum value of 1. Maximum value of 1000.

[nameContains](#)

Return customization jobs only if the job name contains these characters.

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9\+\-\.\])*$`

[nextToken](#)

Continuation token from the previous response, for Amazon Bedrock to list the next set of results.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

sortBy

The field to sort by in the returned list of jobs.

Valid Values: `CreationTime`

sortOrder

The sort order of the results.

Valid Values: `Ascending` | `Descending`

statusEquals

Return customization jobs with the specified status.

Valid Values: `InProgress` | `Completed` | `Failed` | `Stopping` | `Stopped`

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "modelCustomizationJobSummaries": [
    {
      "baseModelArn": "string",
      "creationTime": "string",
      "customizationType": "string",
      "customModelArn": "string",
      "customModelName": "string",
      "endTime": "string",
      "jobArn": "string",
      "jobName": "string",
      "lastModifiedTime": "string",
      "status": "string"
    }
  ]
}
```

```
  ],  
  "nextToken": "string"  
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

modelCustomizationJobSummaries

Job summaries.

Type: Array of [ModelCustomizationJobSummary](#) objects

nextToken

Page continuation token to use in the next request.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: $^\backslash S^*\$$

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListProvisionedModelThroughputs

Service: Amazon Bedrock

Lists the Provisioned Throughputs in the account. For more information, see [Provisioned Throughput](#) in the Amazon Bedrock User Guide.

Request Syntax

```
GET /provisioned-model-throughputs?  
creationTimeAfter=creationTimeAfter&creationTimeBefore=creationTimeBefore&maxResults=maxResults  
HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

[creationTimeAfter](#)

A filter that returns Provisioned Throughputs created after the specified time.

[creationTimeBefore](#)

A filter that returns Provisioned Throughputs created before the specified time.

[maxResults](#)

The maximum number of results to return in the response. If there are more results than the number you specified, the response returns a `nextToken` value. To see the next batch of results, send the `nextToken` value in another list request.

Valid Range: Minimum value of 1. Maximum value of 1000.

[modelArnEquals](#)

A filter that returns Provisioned Throughputs whose model Amazon Resource Name (ARN) is equal to the value that you specify.

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(([\0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}/[a-z0-9]{12})|(:foundation-model/[a-z0-9-]{1,63}[.]{1}([a-z0-9-]{1,63}[.]{0,2}[a-z0-9-]{1,63}(:[a-z0-9-]{1,63}){0,2})))$`

nameContains

A filter that returns Provisioned Throughputs if their name contains the expression that you specify.

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^([0-9a-zA-Z][_]?)+$`

nextToken

If there are more results than the number you specified in the `maxResults` field, the response returns a `nextToken` value. To see the next batch of results, specify the `nextToken` value in this field.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

sortBy

The field by which to sort the returned list of Provisioned Throughputs.

Valid Values: `CreationTime`

sortOrder

The sort order of the results.

Valid Values: `Ascending` | `Descending`

statusEquals

A filter that returns Provisioned Throughputs if their statuses matches the value that you specify.

Valid Values: `Creating` | `InService` | `Updating` | `Failed`

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json
```

```
{
  "nextToken": "string",
  "provisionedModelSummaries": [
    {
      "commitmentDuration": "string",
      "commitmentExpirationTime": "string",
      "creationTime": "string",
      "desiredModelArn": "string",
      "desiredModelUnits": number,
      "foundationModelArn": "string",
      "lastModifiedTime": "string",
      "modelArn": "string",
      "modelUnits": number,
      "provisionedModelArn": "string",
      "provisionedModelName": "string",
      "status": "string"
    }
  ]
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

nextToken

If there are more results than the number you specified in the `maxResults` field, this value is returned. To see the next batch of results, include this value in the `nextToken` field in another list request.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

provisionedModelSummaries

A list of summaries, one for each Provisioned Throughput in the response.

Type: Array of [ProvisionedModelSummary](#) objects

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)

- [AWS SDK for Ruby V3](#)

ListTagsForResource

Service: Amazon Bedrock

List the tags associated with the specified resource.

For more information, see [Tagging resources](#) in the Amazon Bedrock User Guide.

Request Syntax

```
POST /listTagsForResource HTTP/1.1
```

```
Content-type: application/json
```

```
{  
  "resourceARN": "string"  
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

resourceARN

The Amazon Resource Name (ARN) of the resource.

Type: String

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: (^[a-zA-Z0-9][a-zA-Z0-9\-_]*\$)|(^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:([0-9]{12}|)(:(fine-tuning-job|model-customization-job|custom-model)/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([a-z0-9-]{1,63}[.]){0,2}[a-z0-9-]{1,63}(:[a-z0-9-]{1,63}){0,2}(/[a-z0-9]{12})\$)|(:provisioned-model/[a-z0-9]{12}\$)|(:guardrail/[a-z0-9]+\$)|(:evaluation-job/[a-z0-9]{12}\$))

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "tags": [
    {
      "key": "string",
      "value": "string"
    }
  ]
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

tags

An array of the tags associated with this resource.

Type: Array of [Tag](#) objects

Array Members: Minimum number of 0 items. Maximum number of 200 items.

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

PutModelInvocationLoggingConfiguration

Service: Amazon Bedrock

Set the configuration values for model invocation logging.

Request Syntax

```
PUT /logging/modelinvocations HTTP/1.1
Content-type: application/json

{
  "loggingConfig": {
    "cloudWatchConfig": {
      "largeDataDeliveryS3Config": {
        "bucketName": "string",
        "keyPrefix": "string"
      },
      "logGroupName": "string",
      "roleArn": "string"
    },
    "embeddingDataDeliveryEnabled": boolean,
    "imageDataDeliveryEnabled": boolean,
    "s3Config": {
      "bucketName": "string",
      "keyPrefix": "string"
    },
    "textDataDeliveryEnabled": boolean
  }
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

loggingConfig

The logging configuration values to set.

Type: [LoggingConfig](#) object

Required: Yes

Response Syntax

```
HTTP/1.1 200
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)

- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

StopEvaluationJob

Service: Amazon Bedrock

Stops an in progress model evaluation job.

Request Syntax

```
POST /evaluation-job/jobIdentifier/stop HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

jobIdentifier

The ARN of the model evaluation job you want to stop.

Length Constraints: Minimum length of 0. Maximum length of 1011.

Pattern: `^(arn:aws(-[^\:]+)?):bedrock:[a-z0-9-]{1,20}:[0-9]{12}:evaluation-job/[a-z0-9]{12})$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

Error occurred because of a conflict while performing an operation.

HTTP Status Code: 400

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)

- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

StopModelCustomizationJob

Service: Amazon Bedrock

Stops an active model customization job. For more information, see [Custom models](#) in the Amazon Bedrock User Guide.

Request Syntax

```
POST /model-customization-jobs/jobIdentifier/stop HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

jobIdentifier

Job identifier of the job to stop.

Length Constraints: Minimum length of 0. Maximum length of 1011.

Pattern: `^(arn:aws(-[^\:]+)??:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:model-customization-job/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([a-z0-9-]{1,63}[.]){0,2}[a-z0-9-]{1,63}(:[a-z0-9-]{1,63}){0,2}/[a-z0-9]{12})|([a-zA-Z0-9](-*[a-zA-Z0-9\+\-\.\.])*)$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

Error occurred because of a conflict while performing an operation.

HTTP Status Code: 400

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)

- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

TagResource

Service: Amazon Bedrock

Associate tags with a resource. For more information, see [Tagging resources](#) in the Amazon Bedrock User Guide.

Request Syntax

```
POST /tagResource HTTP/1.1
Content-type: application/json
```

```
{
  "resourceARN": "string",
  "tags": [
    {
      "key": "string",
      "value": "string"
    }
  ]
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

resourceARN

The Amazon Resource Name (ARN) of the resource to tag.

Type: String

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: (^[a-zA-Z0-9][a-zA-Z0-9\-_]*\$)|(^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:([0-9]{12})|((:(fine-tuning-job|model-customization-job|custom-model)/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([a-z0-9-]{1,63}[.]){0,2}[a-z0-9-]{1,63}([:][a-z0-9-]{1,63}){0,2}(/[a-z0-9]{12}))\$)|(:provisioned-

```
model/[a-z0-9]{12}$)|(:guardrail/[a-z0-9]+$)|(:evaluation-job/[a-z0-9]{12}$)))
```

Required: Yes

tags

Tags to associate with the resource.

Type: Array of [Tag](#) objects

Array Members: Minimum number of 0 items. Maximum number of 200 items.

Required: Yes

Response Syntax

```
HTTP/1.1 200
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

TooManyTagsException

The request contains more tags than can be associated with a resource (50 tags per resource). The maximum number of tags includes both existing tags and those included in your current request.

HTTP Status Code: 400

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

UntagResource

Service: Amazon Bedrock

Remove one or more tags from a resource. For more information, see [Tagging resources](#) in the Amazon Bedrock User Guide.

Request Syntax

```
POST /untagResource HTTP/1.1
Content-type: application/json
```

```
{
  "resourceARN": "string",
  "tagKeys": [ "string" ]
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

[resourceARN](#)

The Amazon Resource Name (ARN) of the resource to untag.

Type: String

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: (^[a-zA-Z0-9][a-zA-Z0-9\-_]*\$)|(^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:([0-9]{12})|(::(fine-tuning-job|model-customization-job|custom-model)/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([a-z0-9-]{1,63}[.]){0,2}[a-z0-9-]{1,63}(:[a-z0-9-]{1,63}){0,2}(/[a-z0-9]{12})\$)|(:provisioned-model/[a-z0-9]{12}\$)|(:guardrail/[a-z0-9]+\$)|(:evaluation-job/[a-z0-9]{12}\$)))

Required: Yes

[tagKeys](#)

Tag keys of the tags to remove from the resource.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 200 items.

Length Constraints: Minimum length of 1. Maximum length of 128.

Pattern: `^[a-zA-Z0-9\s._:/=+@-]*$`

Required: Yes

Response Syntax

```
HTTP/1.1 200
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

UpdateGuardrail

Service: Amazon Bedrock

Updates a guardrail with the values you specify.

- Specify a name and optional description.
- Specify messages for when the guardrail successfully blocks a prompt or a model response in the `blockedInputMessaging` and `blockedOutputsMessaging` fields.
- Specify topics for the guardrail to deny in the `topicPolicyConfig` object. Each [GuardrailTopicConfig](#) object in the `topicsConfig` list pertains to one topic.
 - Give a name and description so that the guardrail can properly identify the topic.
 - Specify DENY in the `type` field.
 - (Optional) Provide up to five prompts that you would categorize as belonging to the topic in the `examples` list.
- Specify filter strengths for the harmful categories defined in Amazon Bedrock in the `contentPolicyConfig` object. Each [GuardrailContentFilterConfig](#) object in the `filtersConfig` list pertains to a harmful category. For more information, see [Content filters](#). For more information about the fields in a content filter, see [GuardrailContentFilterConfig](#).
 - Specify the category in the `type` field.
 - Specify the strength of the filter for prompts in the `inputStrength` field and for model responses in the `strength` field of the [GuardrailContentFilterConfig](#).
- (Optional) For security, include the ARN of a AWS KMS key in the `kmsKeyId` field.

Request Syntax

```
PUT /guardrails/guardrailIdentifier HTTP/1.1
```

```
Content-type: application/json
```

```
{  
  "blockedInputMessaging": "string",  
  "blockedOutputsMessaging": "string",  
  "contentPolicyConfig": {  
    "filtersConfig": [  
      {  
        "inputStrength": "string",  
        "outputStrength": "string",  
        "type": "string"  
      }  
    ]  
  }  
}
```

```

    }
  ]
},
"contextualGroundingPolicyConfig": {
  "filtersConfig": [
    {
      "threshold": number,
      "type": "string"
    }
  ]
},
"description": "string",
"kmsKeyId": "string",
"name": "string",
"sensitiveInformationPolicyConfig": {
  "piiEntitiesConfig": [
    {
      "action": "string",
      "type": "string"
    }
  ],
  "regexesConfig": [
    {
      "action": "string",
      "description": "string",
      "name": "string",
      "pattern": "string"
    }
  ]
},
"topicPolicyConfig": {
  "topicsConfig": [
    {
      "definition": "string",
      "examples": [ "string" ],
      "name": "string",
      "type": "string"
    }
  ]
},
"wordPolicyConfig": {
  "managedWordListsConfig": [
    {
      "type": "string"
    }
  ]
}
}

```



```
    }
  ],
  "wordsConfig": [
    {
      "text": "string"
    }
  ]
}
```

URI Request Parameters

The request uses the following URI parameters.

[guardrailIdentifier](#)

The unique identifier of the guardrail. This can be an ID or the ARN.

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^(([a-z0-9]+)|(arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:guardrail/[a-z0-9]+))$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

[blockedInputMessaging](#)

The message to return when the guardrail blocks a prompt.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 500.

Required: Yes

[blockedOutputsMessaging](#)

The message to return when the guardrail blocks a model response.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 500.

Required: Yes

contentPolicyConfig

The content policy to configure for the guardrail.

Type: [GuardrailContentPolicyConfig](#) object

Required: No

contextualGroundingPolicyConfig

The contextual grounding policy configuration used to update a guardrail.

Type: [GuardrailContextualGroundingPolicyConfig](#) object

Required: No

description

A description of the guardrail.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

kmsKeyId

The ARN of the AWS KMS key with which to encrypt the guardrail.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:kms:[a-zA-Z0-9-]*:[0-9]{12}:((key/[a-zA-Z0-9-]{36})|(alias/[a-zA-Z0-9-_/]+))$`

Required: No

name

A name for the guardrail.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 50.

Pattern: `^[0-9a-zA-Z-_]+$`

Required: Yes

[sensitiveInformationPolicyConfig](#)

The sensitive information policy to configure for the guardrail.

Type: [GuardrailSensitiveInformationPolicyConfig](#) object

Required: No

[topicPolicyConfig](#)

The topic policy to configure for the guardrail.

Type: [GuardrailTopicPolicyConfig](#) object

Required: No

[wordPolicyConfig](#)

The word policy to configure for the guardrail.

Type: [GuardrailWordPolicyConfig](#) object

Required: No

Response Syntax

```
HTTP/1.1 202
Content-type: application/json

{
  "guardrailArn": "string",
  "guardrailId": "string",
  "updatedAt": "string",
  "version": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 202 response.

The following data is returned in JSON format by the service.

guardrailArn

The ARN of the guardrail.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[\0-9]{12}:guardrail/[a-z0-9]+$`

guardrailId

The unique identifier of the guardrail

Type: String

Length Constraints: Minimum length of 0. Maximum length of 64.

Pattern: `^[a-z0-9]+$`

updatedAt

The date and time at which the guardrail was updated.

Type: Timestamp

version

The version of the guardrail.

Type: String

Length Constraints: Fixed length of 5.

Pattern: `^DRAFT$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

Error occurred because of a conflict while performing an operation.

HTTP Status Code: 400

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 400

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)

- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

UpdateProvisionedModelThroughput

Service: Amazon Bedrock

Updates the name or associated model for a Provisioned Throughput. For more information, see [Provisioned Throughput](#) in the Amazon Bedrock User Guide.

Request Syntax

```
PATCH /provisioned-model-throughput/provisionedModelId HTTP/1.1
Content-type: application/json

{
  "desiredModelId": "string",
  "desiredProvisionedModelName": "string"
}
```

URI Request Parameters

The request uses the following URI parameters.

[provisionedModelId](#)

The Amazon Resource Name (ARN) or name of the Provisioned Throughput to update.

Pattern: `^((([0-9a-zA-Z] [_-]?) +) | (arn:aws(- [^:]+)? :bedrock: [a-z0-9-]{1,20} : [0-9]{12} :provisioned-model/[a-z0-9]{12})))$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

[desiredModelId](#)

The Amazon Resource Name (ARN) of the new model to associate with this Provisioned Throughput. You can't specify this field if this Provisioned Throughput is associated with a base model.

If this Provisioned Throughput is associated with a custom model, you can specify one of the following options:

- The base model from which the custom model was customized.
- Another custom model that was customized from the same base model as the custom model.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(([\0-9]{12}:custom-model/[a-z0-9-]{1,63}[\.]{1}[a-z0-9-]{1,63}(([:][a-z0-9-]{1,63}){0,2})?/[a-z0-9-]{12})|(:foundation-model/([a-z0-9-]{1,63}[\.]{1}[a-z0-9-]{1,63}([\.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2})))|((([a-z0-9-]{1,63}[\.]{1}[a-z0-9-]{1,63}([\.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2})))|(([\0-9a-zA-Z][_]?)+)$`

Required: No

desiredProvisionedModelName

The new name for this Provisioned Throughput.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^([\0-9a-zA-Z][_]?)+$`

Required: No

Response Syntax

```
HTTP/1.1 200
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

Agents for Amazon Bedrock

The following actions are supported by Agents for Amazon Bedrock:

- [AssociateAgentKnowledgeBase](#)
- [CreateAgent](#)
- [CreateAgentActionGroup](#)
- [CreateAgentAlias](#)
- [CreateDataSource](#)
- [CreateFlow](#)
- [CreateFlowAlias](#)
- [CreateFlowVersion](#)
- [CreateKnowledgeBase](#)
- [CreatePrompt](#)
- [CreatePromptVersion](#)
- [DeleteAgent](#)
- [DeleteAgentActionGroup](#)
- [DeleteAgentAlias](#)
- [DeleteAgentVersion](#)
- [DeleteDataSource](#)
- [DeleteFlow](#)
- [DeleteFlowAlias](#)
- [DeleteFlowVersion](#)
- [DeleteKnowledgeBase](#)
- [DeletePrompt](#)
- [DisassociateAgentKnowledgeBase](#)
- [GetAgent](#)
- [GetAgentActionGroup](#)
- [GetAgentAlias](#)
- [GetAgentKnowledgeBase](#)
- [GetAgentVersion](#)

- [GetDataSource](#)
- [GetFlow](#)
- [GetFlowAlias](#)
- [GetFlowVersion](#)
- [GetIngestionJob](#)
- [GetKnowledgeBase](#)
- [GetPrompt](#)
- [ListAgentActionGroups](#)
- [ListAgentAliases](#)
- [ListAgentKnowledgeBases](#)
- [ListAgents](#)
- [ListAgentVersions](#)
- [ListDataSources](#)
- [ListFlowAliases](#)
- [ListFlows](#)
- [ListFlowVersions](#)
- [ListIngestionJobs](#)
- [ListKnowledgeBases](#)
- [ListPrompts](#)
- [ListTagsForResource](#)
- [PrepareAgent](#)
- [PrepareFlow](#)
- [StartIngestionJob](#)
- [TagResource](#)
- [UntagResource](#)
- [UpdateAgent](#)
- [UpdateAgentActionGroup](#)
- [UpdateAgentAlias](#)
- [UpdateAgentKnowledgeBase](#)
- [UpdateDataSource](#)

- [UpdateFlow](#)
- [UpdateFlowAlias](#)
- [UpdateKnowledgeBase](#)
- [UpdatePrompt](#)

AssociateAgentKnowledgeBase

Service: Agents for Amazon Bedrock

Associates a knowledge base with an agent. If a knowledge base is associated and its `indexState` is set to `Enabled`, the agent queries the knowledge base for information to augment its response to the user.

Request Syntax

```
PUT /agents/agentId/agentversions/agentVersion/knowledgebases/ HTTP/1.1
```

```
Content-type: application/json
```

```
{
  "description": "string",
  "knowledgeBaseId": "string",
  "knowledgeBaseState": "string"
}
```

URI Request Parameters

The request uses the following URI parameters.

agentId

The unique identifier of the agent with which you want to associate the knowledge base.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentVersion

The version of the agent with which you want to associate the knowledge base.

Length Constraints: Fixed length of 5.

Pattern: `^DRAFT$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

description

A description of what the agent should use the knowledge base for.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: Yes

knowledgeBaseId

The unique identifier of the knowledge base to associate with the agent.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

knowledgeBaseState

Specifies whether to use the knowledge base or not when sending an [InvokeAgent](#) request.

Type: String

Valid Values: ENABLED | DISABLED

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "agentKnowledgeBase": {
    "agentId": "string",
    "agentVersion": "string",
    "createdAt": "string",
    "description": "string",
    "knowledgeBaseId": "string",
    "knowledgeBaseState": "string",
    "updatedAt": "string"
  }
}
```

```
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

[agentKnowledgeBase](#)

Contains details about the knowledge base that has been associated with the agent.

Type: [AgentKnowledgeBase](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

CreateAgent

Service: Agents for Amazon Bedrock

Creates an agent that orchestrates interactions between foundation models, data sources, software applications, user conversations, and APIs to carry out tasks to help customers.

- Specify the following fields for security purposes.
 - `agentResourceRoleArn` – The Amazon Resource Name (ARN) of the role with permissions to invoke API operations on an agent.
 - (Optional) `customerEncryptionKeyArn` – The Amazon Resource Name (ARN) of a AWS KMS key to encrypt the creation of the agent.
 - (Optional) `idleSessionTTLInSeconds` – Specify the number of seconds for which the agent should maintain session information. After this time expires, the subsequent `InvokeAgent` request begins a new session.
- To enable your agent to retain conversational context across multiple sessions, include a `memoryConfiguration` object. For more information, see [Configure memory](#).
- To override the default prompt behavior for agent orchestration and to use advanced prompts, include a `promptOverrideConfiguration` object. For more information, see [Advanced prompts](#).
- If your agent fails to be created, the response returns a list of `failureReasons` alongside a list of `recommendedActions` for you to troubleshoot.

Request Syntax

```
PUT /agents/ HTTP/1.1
Content-type: application/json

{
  "agentName": "string",
  "agentResourceRoleArn": "string",
  "clientToken": "string",
  "customerEncryptionKeyArn": "string",
  "description": "string",
  "foundationModel": "string",
  "guardrailConfiguration": {
    "guardrailIdentifier": "string",
    "guardrailVersion": "string"
  },
}
```

```

    "idleSessionTTLInSeconds": number,
    "instruction": "string",
    "memoryConfiguration": {
      "enabledMemoryTypes": [ "string" ],
      "storageDays": number
    },
    "promptOverrideConfiguration": {
      "overrideLambda": "string",
      "promptConfigurations": [
        {
          "basePromptTemplate": "string",
          "inferenceConfiguration": {
            "maxLength": number,
            "stopSequences": [ "string" ],
            "temperature": number,
            "topK": number,
            "topP": number
          },
          "parserMode": "string",
          "promptCreationMode": "string",
          "promptState": "string",
          "promptType": "string"
        }
      ]
    },
    "tags": {
      "string" : "string"
    }
  }
}

```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

agentName

A name for the agent that you create.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?)\{1,100\}$`

Required: Yes

agentResourceRoleArn

The Amazon Resource Name (ARN) of the IAM role with permissions to invoke API operations on the agent.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:iam::([0-9]{12})?:role/.+&`

Required: No

clientToken

A unique, case-sensitive identifier to ensure that the API request completes no more than one time. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#).

Type: String

Length Constraints: Minimum length of 33. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,256\}$`

Required: No

customerEncryptionKeyArn

The Amazon Resource Name (ARN) of the AWS KMS key with which to encrypt the agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]\{36\}$`

Required: No

description

A description of the agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

foundationModel

The foundation model to be used for orchestration by the agent you create.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(([\0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}(([:][a-z0-9-]{1,63}){0,2})?/[a-z0-9]{12})|(:foundation-model/([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2})))|((([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2}))|(([\0-9a-zA-Z][_]?)+))$`

Required: No

guardrailConfiguration

The unique Guardrail configuration assigned to the agent when it is created.

Type: [GuardrailConfiguration](#) object

Required: No

idleSessionTTLInSeconds

The number of seconds for which Amazon Bedrock keeps information about a user's conversation with the agent.

A user interaction remains active for the amount of time specified. If no conversation occurs during this time, the session expires and Amazon Bedrock deletes any data provided before the timeout.

Type: Integer

Valid Range: Minimum value of 60. Maximum value of 3600.

Required: No

instruction

Instructions that tell the agent what it should do and how it should interact with users.

Type: String

Length Constraints: Minimum length of 40. Maximum length of 4000.

Required: No

memoryConfiguration

Contains the details of the memory configured for the agent.

Type: [MemoryConfiguration](#) object

Required: No

promptOverrideConfiguration

Contains configurations to override prompts in different parts of an agent sequence. For more information, see [Advanced prompts](#).

Type: [PromptOverrideConfiguration](#) object

Required: No

tags

Any tags that you want to attach to the agent.

Type: String to string map

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Key Pattern: `^[a-zA-Z0-9\s._:/+=@-]*$`

Value Length Constraints: Minimum length of 0. Maximum length of 256.

Value Pattern: `^[a-zA-Z0-9\s._:/+=@-]*$`

Required: No

Response Syntax

HTTP/1.1 202

Content-type: application/json

```
{
  "agent": {
    "agentArn": "string",
    "agentId": "string",
    "agentName": "string",
    "agentResourceRoleArn": "string",
    "agentStatus": "string",
    "agentVersion": "string",
    "clientToken": "string",
    "createdAt": "string",
    "customerEncryptionKeyArn": "string",
    "description": "string",
    "failureReasons": [ "string" ],
    "foundationModel": "string",
    "guardrailConfiguration": {
      "guardrailIdentifier": "string",
      "guardrailVersion": "string"
    },
    "idleSessionTTLInSeconds": number,
    "instruction": "string",
    "memoryConfiguration": {
      "enabledMemoryTypes": [ "string" ],
      "storageDays": number
    },
    "preparedAt": "string",
    "promptOverrideConfiguration": {
      "overrideLambda": "string",
      "promptConfigurations": [
        {
          "basePromptTemplate": "string",
          "inferenceConfiguration": {
            "maximumLength": number,
            "stopSequences": [ "string" ],
            "temperature": number,
            "topK": number,
            "topP": number
          }
        }
      ],
      "parserMode": "string",
      "promptCreationMode": "string",
```

```
        "promptState": "string",
        "promptType": "string"
    }
  ],
  },
  "recommendedActions": [ "string" ],
  "updatedAt": "string"
}
}
```

Response Elements

If the action is successful, the service sends back an HTTP 202 response.

The following data is returned in JSON format by the service.

agent

Contains details about the agent created.

Type: [Agent](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of CreateAgent.

```
PUT /agents/ HTTP/1.1
Content-type: application/json

{
  "agentName": "o1nvve1",
  "agentResourceRoleArn": "arn:aws:iam::123456789012:role/
AmazonBedrockExecutionRoleForAgents_user",
  "instruction": "You are an IT agent who solves customer's problems",
  "description": "Description is here",
  "idleSessionTTLInSeconds": 900,
  "foundationModel": "anthropic.claude-v2"
}
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)

- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

CreateAgentActionGroup

Service: Agents for Amazon Bedrock

Creates an action group for an agent. An action group represents the actions that an agent can carry out for the customer by defining the APIs that an agent can call and the logic for calling them.

To allow your agent to request the user for additional information when trying to complete a task, add an action group with the `parentActionGroupSignature` field set to `AMAZON.UserInput`.

To allow your agent to generate, run, and troubleshoot code when trying to complete a task, add an action group with the `parentActionGroupSignature` field set to `AMAZON.CodeInterpreter`.

You must leave the `description`, `apiSchema`, and `actionGroupExecutor` fields blank for this action group. During orchestration, if your agent determines that it needs to invoke an API in an action group, but doesn't have enough information to complete the API request, it will invoke this action group instead and return an [Observation](#) reprompting the user for more information.

Request Syntax

```
PUT /agents/agentId/agentversions/agentVersion/actiongroups/ HTTP/1.1
```

```
Content-type: application/json
```

```
{
  "actionGroupExecutor": { ... },
  "actionGroupName": "string",
  "actionGroupState": "string",
  "apiSchema": { ... },
  "clientToken": "string",
  "description": "string",
  "functionSchema": { ... },
  "parentActionGroupSignature": "string"
}
```

URI Request Parameters

The request uses the following URI parameters.

[agentId](#)

The unique identifier of the agent for which to create the action group.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentVersion

The version of the agent for which to create the action group.

Length Constraints: Fixed length of 5.

Pattern: `^DRAFT$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

actionGroupExecutor

The Amazon Resource Name (ARN) of the Lambda function containing the business logic that is carried out upon invoking the action or the custom control method for handling the information elicited from the user.

Type: [ActionGroupExecutor](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

actionGroupName

The name to give the action group.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

Required: Yes

actionGroupState

Specifies whether the action group is available for the agent to invoke or not when sending an [InvokeAgent](#) request.

Type: String

Valid Values: ENABLED | DISABLED

Required: No

apiSchema

Contains either details about the S3 object containing the OpenAPI schema for the action group or the JSON or YAML-formatted payload defining the schema. For more information, see [Action group OpenAPI schemas](#).

Type: [APISchema](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

clientToken

A unique, case-sensitive identifier to ensure that the API request completes no more than one time. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#).

Type: String

Length Constraints: Minimum length of 33. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,256}$`

Required: No

description

A description of the action group.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

functionSchema

Contains details about the function schema for the action group or the JSON or YAML-formatted payload defining the schema.

Type: [FunctionSchema](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

[parentActionGroupSignature](#)

To allow your agent to request the user for additional information when trying to complete a task, set this field to `AMAZON.UserInput`. You must leave the `description`, `apiSchema`, and `actionGroupExecutor` fields blank for this action group.

To allow your agent to generate, run, and troubleshoot code when trying to complete a task, set this field to `AMAZON.CodeInterpreter`. You must leave the `description`, `apiSchema`, and `actionGroupExecutor` fields blank for this action group.

During orchestration, if your agent determines that it needs to invoke an API in an action group, but doesn't have enough information to complete the API request, it will invoke this action group instead and return an [Observation](#) reprompting the user for more information.

Type: String

Valid Values: `AMAZON.UserInput` | `AMAZON.CodeInterpreter`

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "agentActionGroup": {
    "actionGroupExecutor": { ... },
    "actionGroupId": "string",
    "actionGroupName": "string",
    "actionGroupState": "string",
    "agentId": "string",
    "agentVersion": "string",
    "apiSchema": { ... },
    "clientToken": "string",
    "createdAt": "string",
```

```
"description": "string",
"functionSchema": { ... },
"parentActionSignature": "string",
"updatedAt": "string"
}
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

agentActionGroup

Contains details about the action group that was created.

Type: [AgentActionGroup](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Create an action group using an OpenAPI schema and a Lambda function

The following example creates an action group using an OpenAPI schema uploaded to an Amazon S3 bucket and sends the information elicited from the user to a Lambda function.

```
PUT /agents/AGENT12345/agentversions/DRAFT/actiongroups/ HTTP/1.1
Content-type: application/json

{
  "actionGroupName": "Test Action",
  "actionGroupState": "ENABLED",
  "apiSchema": {
    "s3": {
      "s3BucketName": "apischema-s3",
      "s3ObjectKey": "it_agent_openapi.json"
    }
  },
  "description": "Testing latest IT Management action",
  "actionGroupExecutor": {
    "lambda": "arn:aws:lambda:us-west-2:123456789012:function:ItAgentLambda"
  }
}
```

Create an action group using an OpenAPI schema and return control

The following example creates an action group using an OpenAPI schema uploaded to an Amazon S3 bucket and returns control by sending the information in the InvokeAgent response.

```
{
  "actionGroupName": "WeatherAPIs",
  "description": "Actions to get current weather and historical trends for a
location",
  "actionGroupState": "ENABLED",
  "apiSchema": {
    "s3": {
      "s3BucketName": "openapi-spec-iad",
      "s3ObjectKey": "get_weather_openapi.yaml"
    }
  },
  "actionGroupExecutor": {
    "customControl": "RETURN_CONTROL"
  }
}
```

Create an action group using function details and return control

The following example creates an action group using function details and returns control by sending the information in the InvokeAgent response

```
PUT /agents/AGENT12345/agentversions/DRAFT/actiongroups/ HTTP/1.1
Content-type: application/json

{
  "actionGroupName": "OrderManagementAction",
  "description": "Action to get the order history, product details, product
availability and to update the order",
  "actionGroupState": "ENABLED",
  "actionGroupExecutor": {
    "customControl": "RETURN_CONTROL"
  },
  "functionSchema": {
    "functions": [{
      "name": "GetOrderDetails",
      "description": "Retrieves the order history for a given OrderId and returns
productId, color, productName, size, productType, quantity, and status."
      "parameters": {
```



```
        "orderId": {
            "type": "string",
            "required": true
        }
    }
}]
}
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

CreateAgentAlias

Service: Agents for Amazon Bedrock

Creates an alias of an agent that can be used to deploy the agent.

Request Syntax

```
PUT /agents/agentId/agentaliases/ HTTP/1.1
```

```
Content-type: application/json
```

```
{
  "agentAliasName": "string",
  "clientToken": "string",
  "description": "string",
  "routingConfiguration": [
    {
      "agentVersion": "string",
      "provisionedThroughput": "string"
    }
  ],
  "tags": {
    "string" : "string"
  }
}
```

URI Request Parameters

The request uses the following URI parameters.

agentId

The unique identifier of the agent.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

agentAliasName

The name of the alias.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

Required: Yes

clientToken

A unique, case-sensitive identifier to ensure that the API request completes no more than one time. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#).

Type: String

Length Constraints: Minimum length of 33. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,256}$`

Required: No

description

A description of the alias of the agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

routingConfiguration

Contains details about the routing configuration of the alias.

Type: Array of [AgentAliasRoutingConfigurationListItem](#) objects

Array Members: Minimum number of 0 items. Maximum number of 1 item.

Required: No

tags

Any tags that you want to attach to the alias of the agent.

Type: String to string map

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Key Pattern: `^[a-zA-Z0-9\s._:/=+@-]*$`

Value Length Constraints: Minimum length of 0. Maximum length of 256.

Value Pattern: `^[a-zA-Z0-9\s._:/=+@-]*$`

Required: No

Response Syntax

```
HTTP/1.1 202
Content-type: application/json

{
  "agentAlias": {
    "agentAliasArn": "string",
    "agentAliasHistoryEvents": [
      {
        "endDate": "string",
        "routingConfiguration": [
          {
            "agentVersion": "string",
            "provisionedThroughput": "string"
          }
        ],
        "startDate": "string"
      }
    ],
    "agentAliasId": "string",
    "agentAliasName": "string",
    "agentAliasStatus": "string",
    "agentId": "string",
    "clientToken": "string",
    "createdAt": "string",
    "description": "string",
    "failureReasons": [ "string" ],
    "routingConfiguration": [
      {
        "agentVersion": "string",
```

```
        "provisionedThroughput": "string"
    }
],
"updatedAt": "string"
}
}
```

Response Elements

If the action is successful, the service sends back an HTTP 202 response.

The following data is returned in JSON format by the service.

agentAlias

Contains details about the alias that was created.

Type: [AgentAlias](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of `CreateAgentAlias`.

```
PUT /agents/ABCDEFGHIJ/agentaliases/ HTTP/1.1
Content-type: application/json

{
  "agentAliasName": "TestName",
  "description": "Alias is test"
}
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)

- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

CreateDataSource

Service: Agents for Amazon Bedrock

Creates a data source connector for a knowledge base.

Important

You can't change the chunkingConfiguration after you create the data source connector.

Request Syntax

```
PUT /knowledgebases/knowledgeBaseId/datasources/ HTTP/1.1
```

```
Content-type: application/json
```

```
{
  "clientToken": "string",
  "dataDeletionPolicy": "string",
  "dataSourceConfiguration": {
    "confluenceConfiguration": {
      "crawlerConfiguration": {
        "filterConfiguration": {
          "patternObjectFilter": {
            "filters": [
              {
                "exclusionFilters": [ "string" ],
                "inclusionFilters": [ "string" ],
                "objectType": "string"
              }
            ]
          },
          "type": "string"
        }
      },
      "sourceConfiguration": {
        "authType": "string",
        "credentialsSecretArn": "string",
        "hostType": "string",
        "hostUrl": "string"
      }
    },
  },
}
```



```
"s3Configuration": {
  "bucketArn": "string",
  "bucketOwnerAccountId": "string",
  "inclusionPrefixes": [ "string" ]
},
"salesforceConfiguration": {
  "crawlerConfiguration": {
    "filterConfiguration": {
      "patternObjectFilter": {
        "filters": [
          {
            "exclusionFilters": [ "string" ],
            "inclusionFilters": [ "string" ],
            "objectType": "string"
          }
        ]
      },
      "type": "string"
    }
  },
  "sourceConfiguration": {
    "authType": "string",
    "credentialsSecretArn": "string",
    "hostUrl": "string"
  }
},
"sharePointConfiguration": {
  "crawlerConfiguration": {
    "filterConfiguration": {
      "patternObjectFilter": {
        "filters": [
          {
            "exclusionFilters": [ "string" ],
            "inclusionFilters": [ "string" ],
            "objectType": "string"
          }
        ]
      },
      "type": "string"
    }
  },
  "sourceConfiguration": {
    "authType": "string",
    "credentialsSecretArn": "string",
```

```

        "domain": "string",
        "hostType": "string",
        "siteUrls": [ "string" ],
        "tenantId": "string"
    }
},
"type": "string",
"webConfiguration": {
    "crawlerConfiguration": {
        "crawlerLimits": {
            "rateLimit": number
        },
        "exclusionFilters": [ "string" ],
        "inclusionFilters": [ "string" ],
        "scope": "string"
    },
    "sourceConfiguration": {
        "urlConfiguration": {
            "seedUrls": [
                {
                    "url": "string"
                }
            ]
        }
    }
}
},
"description": "string",
"name": "string",
"serverSideEncryptionConfiguration": {
    "kmsKeyArn": "string"
},
"vectorIngestionConfiguration": {
    "chunkingConfiguration": {
        "chunkingStrategy": "string",
        "fixedSizeChunkingConfiguration": {
            "maxTokens": number,
            "overlapPercentage": number
        },
        "hierarchicalChunkingConfiguration": {
            "levelConfigurations": [
                {
                    "maxTokens": number
                }
            ]
        }
    }
}
}

```

```

    ],
    "overlapTokens": number
  },
  "semanticChunkingConfiguration": {
    "breakpointPercentileThreshold": number,
    "bufferSize": number,
    "maxTokens": number
  }
},
"customTransformationConfiguration": {
  "intermediateStorage": {
    "s3Location": {
      "uri": "string"
    }
  },
  "transformations": [
    {
      "stepToApply": "string",
      "transformationFunction": {
        "transformationLambdaConfiguration": {
          "lambdaArn": "string"
        }
      }
    }
  ]
},
"parsingConfiguration": {
  "bedrockFoundationModelConfiguration": {
    "modelArn": "string",
    "parsingPrompt": {
      "parsingPromptText": "string"
    }
  },
  "parsingStrategy": "string"
}
}
}

```

URI Request Parameters

The request uses the following URI parameters.

knowledgeBaseId

The unique identifier of the knowledge base to which to add the data source.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

clientToken

A unique, case-sensitive identifier to ensure that the API request completes no more than one time. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#).

Type: String

Length Constraints: Minimum length of 33. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,256}$`

Required: No

dataDeletionPolicy

The data deletion policy for the data source.

You can set the data deletion policy to:

- **DELETE:** Deletes all underlying data belonging to the data source from the vector store upon deletion of a knowledge base or data source resource. Note that the vector store itself is not deleted, only the underlying data. This flag is ignored if an AWS account is deleted.
- **RETAIN:** Retains all underlying data in your vector store upon deletion of a knowledge base or data source resource.

Type: String

Valid Values: RETAIN | DELETE

Required: No

[dataSourceConfiguration](#)

The connection configuration for the data source.

Type: [DataSourceConfiguration](#) object

Required: Yes

[description](#)

A description of the data source.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

[name](#)

The name of the data source.

Type: String

Pattern: $^([\text{0-9a-zA-Z}][\text{-}]?)\{1,100\}\$$

Required: Yes

[serverSideEncryptionConfiguration](#)

Contains details about the server-side encryption for the data source.

Type: [ServerSideEncryptionConfiguration](#) object

Required: No

[vectorIngestionConfiguration](#)

Contains details about how to ingest the documents in the data source.

Type: [VectorIngestionConfiguration](#) object

Required: No

Response Syntax

HTTP/1.1 200

Content-type: application/json

```
{
  "dataSource": {
    "createdAt": "string",
    "dataDeletionPolicy": "string",
    "dataSourceConfiguration": {
      "confluenceConfiguration": {
        "crawlerConfiguration": {
          "filterConfiguration": {
            "patternObjectFilter": {
              "filters": [
                {
                  "exclusionFilters": [ "string" ],
                  "inclusionFilters": [ "string" ],
                  "objectType": "string"
                }
              ]
            }
          },
          "type": "string"
        }
      },
      "sourceConfiguration": {
        "authType": "string",
        "credentialsSecretArn": "string",
        "hostType": "string",
        "hostUrl": "string"
      }
    },
    "s3Configuration": {
      "bucketArn": "string",
      "bucketOwnerAccountId": "string",
      "inclusionPrefixes": [ "string" ]
    },
    "salesforceConfiguration": {
      "crawlerConfiguration": {
        "filterConfiguration": {
          "patternObjectFilter": {
            "filters": [
              {
                "exclusionFilters": [ "string" ],
```

```

        "inclusionFilters": [ "string" ],
        "objectType": "string"
    }
]
},
"type": "string"
}
},
"sourceConfiguration": {
    "authType": "string",
    "credentialsSecretArn": "string",
    "hostUrl": "string"
}
},
"sharePointConfiguration": {
    "crawlerConfiguration": {
        "filterConfiguration": {
            "patternObjectFilter": {
                "filters": [
                    {
                        "exclusionFilters": [ "string" ],
                        "inclusionFilters": [ "string" ],
                        "objectType": "string"
                    }
                ]
            },
            "type": "string"
        }
    },
    "sourceConfiguration": {
        "authType": "string",
        "credentialsSecretArn": "string",
        "domain": "string",
        "hostType": "string",
        "siteUrls": [ "string" ],
        "tenantId": "string"
    }
},
"type": "string",
"webConfiguration": {
    "crawlerConfiguration": {
        "crawlerLimits": {
            "rateLimit": number
        }
    },

```

```

        "exclusionFilters": [ "string" ],
        "inclusionFilters": [ "string" ],
        "scope": "string"
    },
    "sourceConfiguration": {
        "urlConfiguration": {
            "seedUrls": [
                {
                    "url": "string"
                }
            ]
        }
    }
},
"dataSourceId": "string",
"description": "string",
"failureReasons": [ "string" ],
"knowledgeBaseId": "string",
"name": "string",
"serverSideEncryptionConfiguration": {
    "kmsKeyArn": "string"
},
"status": "string",
"updatedAt": "string",
"vectorIngestionConfiguration": {
    "chunkingConfiguration": {
        "chunkingStrategy": "string",
        "fixedSizeChunkingConfiguration": {
            "maxTokens": number,
            "overlapPercentage": number
        }
    },
    "hierarchicalChunkingConfiguration": {
        "levelConfigurations": [
            {
                "maxTokens": number
            }
        ],
        "overlapTokens": number
    },
    "semanticChunkingConfiguration": {
        "breakpointPercentileThreshold": number,
        "bufferSize": number,
        "maxTokens": number
    }
}

```



```

    }
  },
  "customTransformationConfiguration": {
    "intermediateStorage": {
      "s3Location": {
        "uri": "string"
      }
    },
    "transformations": [
      {
        "stepToApply": "string",
        "transformationFunction": {
          "transformationLambdaConfiguration": {
            "lambdaArn": "string"
          }
        }
      }
    ]
  },
  "parsingConfiguration": {
    "bedrockFoundationModelConfiguration": {
      "modelArn": "string",
      "parsingPrompt": {
        "parsingPromptText": "string"
      }
    },
    "parsingStrategy": "string"
  }
}

```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

dataSource

Contains details about the data source.

Type: [DataSource](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

CreateFlow

Service: Agents for Amazon Bedrock

Creates a prompt flow that you can use to send an input through various steps to yield an output. Configure nodes, each of which corresponds to a step of the flow, and create connections between the nodes to create paths to different outputs. For more information, see [How it works](#) and [Create a flow in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Request Syntax

```
POST /flows/ HTTP/1.1
Content-type: application/json

{
  "clientToken": "string",
  "customerEncryptionKeyArn": "string",
  "definition": {
    "connections": [
      {
        "configuration": { ... },
        "name": "string",
        "source": "string",
        "target": "string",
        "type": "string"
      }
    ],
    "nodes": [
      {
        "configuration": { ... },
        "inputs": [
          {
            "expression": "string",
            "name": "string",
            "type": "string"
          }
        ],
        "name": "string",
        "outputs": [
          {
            "name": "string",
            "type": "string"
          }
        ]
      }
    ]
  }
}
```

```

        "type": "string"
    }
]
},
"description": "string",
"executionRoleArn": "string",
"name": "string",
"tags": {
    "string" : "string"
}
}

```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

clientToken

A unique, case-sensitive identifier to ensure that the API request completes no more than one time. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#).

Type: String

Length Constraints: Minimum length of 33. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,256}$`

Required: No

customerEncryptionKeyArn

The Amazon Resource Name (ARN) of the KMS key to encrypt the flow.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

Required: No

definition

A definition of the nodes and connections between nodes in the flow.

Type: [FlowDefinition](#) object

Required: No

description

A description for the flow.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

executionRoleArn

The Amazon Resource Name (ARN) of the service role with permissions to create and manage a flow. For more information, see [Create a service role for flows in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:iam::([0-9]{12})?:role/(service-role/)?\.\+$`

Required: Yes

name

A name for the flow.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?){1,100}$`

Required: Yes

tags

Any tags that you want to attach to the flow. For more information, see [Tagging resources in Amazon Bedrock](#).

Type: String to string map

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Key Pattern: `^[a-zA-Z0-9\s._:/=+@-]*$`

Value Length Constraints: Minimum length of 0. Maximum length of 256.

Value Pattern: `^[a-zA-Z0-9\s._:/=+@-]*$`

Required: No

Response Syntax

```
HTTP/1.1 201
Content-type: application/json

{
  "arn": "string",
  "createdAt": "string",
  "customerEncryptionKeyArn": "string",
  "definition": {
    "connections": [
      {
        "configuration": { ... },
        "name": "string",
        "source": "string",
        "target": "string",
        "type": "string"
      }
    ],
    "nodes": [
      {
        "configuration": { ... },
        "inputs": [
          {
            "expression": "string",
            "name": "string",
            "type": "string"
          }
        ],
        "name": "string",
        "outputs": [
```

```

        {
            "name": "string",
            "type": "string"
        }
    ],
    "type": "string"
}
]
},
"description": "string",
"executionRoleArn": "string",
"id": "string",
"name": "string",
"status": "string",
"updatedAt": "string",
"version": "string"
}

```

Response Elements

If the action is successful, the service sends back an HTTP 201 response.

The following data is returned in JSON format by the service.

arn

The Amazon Resource Name (ARN) of the flow.

Type: String

Pattern: `^arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10}$`

createdAt

The time at which the flow was created.

Type: Timestamp

customerEncryptionKeyArn

The Amazon Resource Name (ARN) of the KMS key that you encrypted the flow with.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

definition

A definition of the nodes and connections between nodes in the flow.

Type: [FlowDefinition](#) object

description

The description of the flow.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

executionRoleArn

The Amazon Resource Name (ARN) of the service role with permissions to create a flow. For more information, see [Create a service role for flows in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:iam::([0-9]{12})?:role/(service-role/)?\.\+$`

id

The unique identifier of the flow.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

name

The name of the flow.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?){1,100}$`

status

The status of the flow. When you submit this request, the status will be NotPrepared. If creation fails, the status becomes Failed.

Type: String

Valid Values: Failed | Prepared | Preparing | NotPrepared

updatedAt

The time at which the flow was last updated.

Type: Timestamp

version

The version of the flow. When you create a flow, the version created is the DRAFT version.

Type: String

Length Constraints: Fixed length of 5.

Pattern: ^DRAFT\$

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

CreateFlowAlias

Service: Agents for Amazon Bedrock

Creates an alias of a flow for deployment. For more information, see [Deploy a flow in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Request Syntax

```
POST /flows/flowIdentifier/aliases HTTP/1.1
```

```
Content-type: application/json
```

```
{
  "clientToken": "string",
  "description": "string",
  "name": "string",
  "routingConfiguration": [
    {
      "flowVersion": "string"
    }
  ],
  "tags": {
    "string" : "string"
  }
}
```

URI Request Parameters

The request uses the following URI parameters.

[flowIdentifier](#)

The unique identifier of the flow for which to create an alias.

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10})|([0-9a-zA-Z]{10})$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

clientToken

A unique, case-sensitive identifier to ensure that the API request completes no more than one time. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#).

Type: String

Length Constraints: Minimum length of 33. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,256}$`

Required: No

description

A description for the alias.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

name

A name for the alias.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

Required: Yes

routingConfiguration

Contains information about the version to which to map the alias.

Type: Array of [FlowAliasRoutingConfigurationListItem](#) objects

Array Members: Fixed number of 1 item.

Required: Yes

tags

Any tags that you want to attach to the alias of the flow. For more information, see [Tagging resources in Amazon Bedrock](#).

Type: String to string map

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Key Pattern: `^[a-zA-Z0-9\s._:/=+@-]*$`

Value Length Constraints: Minimum length of 0. Maximum length of 256.

Value Pattern: `^[a-zA-Z0-9\s._:/=+@-]*$`

Required: No

Response Syntax

```
HTTP/1.1 201
Content-type: application/json

{
  "arn": "string",
  "createdAt": "string",
  "description": "string",
  "flowId": "string",
  "id": "string",
  "name": "string",
  "routingConfiguration": [
    {
      "flowVersion": "string"
    }
  ],
  "updatedAt": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 201 response.

The following data is returned in JSON format by the service.

arn

The Amazon Resource Name (ARN) of the alias.

Type: String

Pattern: `^arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10}/alias/(TSTALIASID|[0-9a-zA-Z]{10})$`

createdAt

The time at which the alias was created.

Type: Timestamp

description

The description of the alias.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

flowId

The unique identifier of the flow that the alias belongs to.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

id

The unique identifier of the alias.

Type: String

Pattern: `^(TSTALIASID|[0-9a-zA-Z]{10})$`

name

The name of the alias.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?){1,100}$`

routingConfiguration

Contains information about the version that the alias is mapped to.

Type: Array of [FlowAliasRoutingConfigurationListItem](#) objects

Array Members: Fixed number of 1 item.

updatedAt

The time at which the alias of the flow was last updated.

Type: Timestamp

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

CreateFlowVersion

Service: Agents for Amazon Bedrock

Creates a version of the flow that you can deploy. For more information, see [Deploy a flow in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Request Syntax

```
POST /flows/flowIdentifier/versions HTTP/1.1
Content-type: application/json
```

```
{
  "clientToken": "string",
  "description": "string"
}
```

URI Request Parameters

The request uses the following URI parameters.

flowIdentifier

The unique identifier of the flow that you want to create a version of.

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10})|([0-9a-zA-Z]{10})$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

clientToken

A unique, case-sensitive identifier to ensure that the API request completes no more than one time. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#).

Type: String

Length Constraints: Minimum length of 33. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,256}$`

Required: No

description

A description of the version of the flow.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

Response Syntax

```
HTTP/1.1 201
Content-type: application/json

{
  "arn": "string",
  "createdAt": "string",
  "customerEncryptionKeyArn": "string",
  "definition": {
    "connections": [
      {
        "configuration": { ... },
        "name": "string",
        "source": "string",
        "target": "string",
        "type": "string"
      }
    ],
    "nodes": [
      {
        "configuration": { ... },
        "inputs": [
          {
            "expression": "string",
            "name": "string",
            "type": "string"
          }
        ]
      }
    ],
    "name": "string",
```

```
    "outputs": [  
      {  
        "name": "string",  
        "type": "string"  
      }  
    ],  
    "type": "string"  
  }  
]  
},  
"description": "string",  
"executionRoleArn": "string",  
"id": "string",  
"name": "string",  
"status": "string",  
"version": "string"  
}
```

Response Elements

If the action is successful, the service sends back an HTTP 201 response.

The following data is returned in JSON format by the service.

arn

The Amazon Resource Name (ARN) of the flow.

Type: String

Pattern: `^arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10}$`

createdAt

The time at which the flow was created.

Type: Timestamp

customerEncryptionKeyArn

The KMS key that the flow is encrypted with.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

definition

A definition of the nodes and connections in the flow.

Type: [FlowDefinition](#) object

description

The description of the flow version.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

executionRoleArn

The Amazon Resource Name (ARN) of the service role with permissions to create a flow. For more information, see [Create a service role for flows in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:iam::([0-9]{12})?:role/(service-role/)?\.\+$`

id

The unique identifier of the flow.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

name

The name of the flow version.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?){1,100}$`

status

The status of the flow.

Type: String

Valid Values: Failed | Prepared | Preparing | NotPrepared

version

The version of the flow that was created. Versions are numbered incrementally, starting from 1.

Type: String

Pattern: `^[0-9]{1,5}$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

CreateKnowledgeBase

Service: Agents for Amazon Bedrock

Creates a knowledge base that contains data sources from which information can be queried and used by LLMs. To create a knowledge base, you must first set up your data sources and configure a supported vector store. For more information, see [Set up your data for ingestion](#).

Note

If you prefer to let Amazon Bedrock create and manage a vector store for you in Amazon OpenSearch Service, use the console. For more information, see [Create a knowledge base](#).

- Provide the name and an optional description.
- Provide the Amazon Resource Name (ARN) with permissions to create a knowledge base in the `roleArn` field.
- Provide the embedding model to use in the `embeddingModelArn` field in the `knowledgeBaseConfiguration` object.
- Provide the configuration for your vector store in the `storageConfiguration` object.
 - For an Amazon OpenSearch Service database, use the `opensearchServerlessConfiguration` object. For more information, see [Create a vector store in Amazon OpenSearch Service](#).
 - For an Amazon Aurora database, use the `RdsConfiguration` object. For more information, see [Create a vector store in Amazon Aurora](#).
 - For a Pinecone database, use the `pineconeConfiguration` object. For more information, see [Create a vector store in Pinecone](#).
 - For a Redis Enterprise Cloud database, use the `redisEnterpriseCloudConfiguration` object. For more information, see [Create a vector store in Redis Enterprise Cloud](#).

Request Syntax

```
PUT /knowledgebases/ HTTP/1.1
Content-type: application/json

{
  "clientToken": "string",
  "description": "string",
```



```
"knowledgeBaseConfiguration": {
  "type": "string",
  "vectorKnowledgeBaseConfiguration": {
    "embeddingModelArn": "string",
    "embeddingModelConfiguration": {
      "bedrockEmbeddingModelConfiguration": {
        "dimensions": number
      }
    }
  }
},
"name": "string",
"roleArn": "string",
"storageConfiguration": {
  "mongoDbAtlasConfiguration": {
    "collectionName": "string",
    "credentialsSecretArn": "string",
    "databaseName": "string",
    "endpoint": "string",
    "endpointServiceName": "string",
    "fieldMapping": {
      "metadataField": "string",
      "textField": "string",
      "vectorField": "string"
    },
    "vectorIndexName": "string"
  },
  "opensearchServerlessConfiguration": {
    "collectionArn": "string",
    "fieldMapping": {
      "metadataField": "string",
      "textField": "string",
      "vectorField": "string"
    },
    "vectorIndexName": "string"
  },
  "pineconeConfiguration": {
    "connectionString": "string",
    "credentialsSecretArn": "string",
    "fieldMapping": {
      "metadataField": "string",
      "textField": "string"
    },
    "namespace": "string"
  }
}
```

```

    },
    "rdsConfiguration": {
      "credentialsSecretArn": "string",
      "databaseName": "string",
      "fieldMapping": {
        "metadataField": "string",
        "primaryKeyField": "string",
        "textField": "string",
        "vectorField": "string"
      },
      "resourceArn": "string",
      "tableName": "string"
    },
    "redisEnterpriseCloudConfiguration": {
      "credentialsSecretArn": "string",
      "endpoint": "string",
      "fieldMapping": {
        "metadataField": "string",
        "textField": "string",
        "vectorField": "string"
      },
      "vectorIndexName": "string"
    },
    "type": "string"
  },
  "tags": {
    "string" : "string"
  }
}

```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

clientToken

A unique, case-sensitive identifier to ensure that the API request completes no more than one time. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#).

Type: String

Length Constraints: Minimum length of 33. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,256}$`

Required: No

description

A description of the knowledge base.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

knowledgeBaseConfiguration

Contains details about the embeddings model used for the knowledge base.

Type: [KnowledgeBaseConfiguration](#) object

Required: Yes

name

A name for the knowledge base.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

Required: Yes

roleArn

The Amazon Resource Name (ARN) of the IAM role with permissions to invoke API operations on the knowledge base.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^:]+)?:iam::([0-9]{12})?:role/.$`

Required: Yes

storageConfiguration

Contains details about the configuration of the vector database used for the knowledge base.

Type: [StorageConfiguration](#) object

Required: Yes

tags

Specify the key-value pairs for the tags that you want to attach to your knowledge base in this object.

Type: String to string map

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Key Pattern: `^[a-zA-Z0-9\s._:/+=@-]*$`

Value Length Constraints: Minimum length of 0. Maximum length of 256.

Value Pattern: `^[a-zA-Z0-9\s._:/+=@-]*$`

Required: No

Response Syntax

```
HTTP/1.1 202
Content-type: application/json

{
  "knowledgeBase": {
    "createdAt": "string",
    "description": "string",
    "failureReasons": [ "string" ],
    "knowledgeBaseArn": "string",
    "knowledgeBaseConfiguration": {
      "type": "string",
      "vectorKnowledgeBaseConfiguration": {
        "embeddingModelArn": "string",
        "embeddingModelConfiguration": {
          "bedrockEmbeddingModelConfiguration": {
            "dimensions": number
          }
        }
      }
    }
  }
}
```

```
    }
  },
  "knowledgeBaseId": "string",
  "name": "string",
  "roleArn": "string",
  "status": "string",
  "storageConfiguration": {
    "mongoDbAtlasConfiguration": {
      "collectionName": "string",
      "credentialsSecretArn": "string",
      "databaseName": "string",
      "endpoint": "string",
      "endpointServiceName": "string",
      "fieldMapping": {
        "metadataField": "string",
        "textField": "string",
        "vectorField": "string"
      },
      "vectorIndexName": "string"
    },
    "opensearchServerlessConfiguration": {
      "collectionArn": "string",
      "fieldMapping": {
        "metadataField": "string",
        "textField": "string",
        "vectorField": "string"
      },
      "vectorIndexName": "string"
    },
    "pineconeConfiguration": {
      "connectionString": "string",
      "credentialsSecretArn": "string",
      "fieldMapping": {
        "metadataField": "string",
        "textField": "string"
      },
      "namespace": "string"
    },
    "rdsConfiguration": {
      "credentialsSecretArn": "string",
      "databaseName": "string",
      "fieldMapping": {
        "metadataField": "string",
        "primaryKeyField": "string",
```

```

        "textField": "string",
        "vectorField": "string"
    },
    "resourceArn": "string",
    "tableName": "string"
},
"redisEnterpriseCloudConfiguration": {
    "credentialsSecretArn": "string",
    "endpoint": "string",
    "fieldMapping": {
        "metadataField": "string",
        "textField": "string",
        "vectorField": "string"
    },
    "vectorIndexName": "string"
},
"type": "string"
},
"updatedAt": "string"
}
}

```

Response Elements

If the action is successful, the service sends back an HTTP 202 response.

The following data is returned in JSON format by the service.

knowledgeBase

Contains details about the knowledge base.

Type: [KnowledgeBase](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)

- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

CreatePrompt

Service: Agents for Amazon Bedrock

Creates a prompt in your prompt library that you can add to a flow. For more information, see [Prompt management in Amazon Bedrock](#), [Create a prompt using Prompt management](#) and [Prompt flows in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Request Syntax

```
POST /prompts/ HTTP/1.1
Content-type: application/json

{
  "clientToken": "string",
  "customerEncryptionKeyArn": "string",
  "defaultVariant": "string",
  "description": "string",
  "name": "string",
  "tags": {
    "string" : "string"
  },
  "variants": [
    {
      "inferenceConfiguration": { ... },
      "modelId": "string",
      "name": "string",
      "templateConfiguration": { ... },
      "templateType": "string"
    }
  ]
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

clientToken

A unique, case-sensitive identifier to ensure that the API request completes no more than one time. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#).

Type: String

Length Constraints: Minimum length of 33. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,256}$`

Required: No

customerEncryptionKeyArn

The Amazon Resource Name (ARN) of the KMS key to encrypt the prompt.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

Required: No

defaultVariant

The name of the default variant for the prompt. This value must match the name field in the relevant [PromptVariant](#) object.

Type: String

Pattern: `^([0-9a-zA-Z][_-]?){1,100}$`

Required: No

description

A description for the prompt.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

name

A name for the prompt.

Type: String

Pattern: $^([\text{0-9a-zA-Z}][\text{-}]?)\{1,100\}\$$

Required: Yes

tags

Any tags that you want to attach to the prompt. For more information, see [Tagging resources in Amazon Bedrock](#).

Type: String to string map

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Key Pattern: $^[a-zA-Z0-9\ _ \. \: \: / = + @ -] * \$$

Value Length Constraints: Minimum length of 0. Maximum length of 256.

Value Pattern: $^[a-zA-Z0-9\ _ \. \: \: / = + @ -] * \$$

Required: No

variants

A list of objects, each containing details about a variant of the prompt.

Type: Array of [PromptVariant](#) objects

Array Members: Minimum number of 0 items. Maximum number of 3 items.

Required: No

Response Syntax

```
HTTP/1.1 201
Content-type: application/json
```

```

{
  "arn": "string",
  "createdAt": "string",
  "customerEncryptionKeyArn": "string",
  "defaultVariant": "string",
  "description": "string",
  "id": "string",
  "name": "string",
  "updatedAt": "string",
  "variants": [
    {
      "inferenceConfiguration": { ... },
      "modelId": "string",
      "name": "string",
      "templateConfiguration": { ... },
      "templateType": "string"
    }
  ],
  "version": "string"
}

```

Response Elements

If the action is successful, the service sends back an HTTP 201 response.

The following data is returned in JSON format by the service.

arn

The Amazon Resource Name (ARN) of the prompt.

Type: String

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:prompt/[0-9a-zA-Z]{10}(?::[0-9]{1,5})?)$`

createdAt

The time at which the prompt was created.

Type: Timestamp

customerEncryptionKeyArn

The Amazon Resource Name (ARN) of the KMS key that you encrypted the prompt with.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

defaultVariant

The name of the default variant for your prompt.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

description

The description of the prompt.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

id

The unique identifier of the prompt.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

name

The name of the prompt.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

updatedAt

The time at which the prompt was last updated.

Type: Timestamp

variants

A list of objects, each containing details about a variant of the prompt.

Type: Array of [PromptVariant](#) objects

Array Members: Minimum number of 0 items. Maximum number of 3 items.

[version](#)

The version of the prompt. When you create a prompt, the version created is the DRAFT version.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: `^(DRAFT|[0-9]{0,4}[1-9][0-9]{0,4})$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

CreatePromptVersion

Service: Agents for Amazon Bedrock

Creates a static snapshot of your prompt that can be deployed to production. For more information, see [Deploy prompts using Prompt management by creating versions](#) in the Amazon Bedrock User Guide.

Request Syntax

```
POST /prompts/promptIdentifier/versions HTTP/1.1
Content-type: application/json

{
  "clientToken": "string",
  "description": "string",
  "tags": {
    "string" : "string"
  }
}
```

URI Request Parameters

The request uses the following URI parameters.

[promptIdentifier](#)

The unique identifier of the prompt that you want to create a version of.

Pattern: `^([\0-9a-zA-Z]{10})|(arn:aws:bedrock:[a-z0-9-]{1,20}:[\0-9]{12}:prompt/[\0-9a-zA-Z]{10})(?::[\0-9]{1,5})?&`

Required: Yes

Request Body

The request accepts the following data in JSON format.

[clientToken](#)

A unique, case-sensitive identifier to ensure that the API request completes no more than one time. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#).

Type: String

Length Constraints: Minimum length of 33. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,256}$`

Required: No

description

A description for the version of the prompt.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

tags

Any tags that you want to attach to the version of the prompt. For more information, see [Tagging resources in Amazon Bedrock](#).

Type: String to string map

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Key Pattern: `^[a-zA-Z0-9\s._:/=+@-]*$`

Value Length Constraints: Minimum length of 0. Maximum length of 256.

Value Pattern: `^[a-zA-Z0-9\s._:/=+@-]*$`

Required: No

Response Syntax

```
HTTP/1.1 201
Content-type: application/json

{
  "arn": "string",
  "createdAt": "string",
  "customerEncryptionKeyArn": "string",
```

```

    "defaultVariant": "string",
    "description": "string",
    "id": "string",
    "name": "string",
    "updatedAt": "string",
    "variants": [
      {
        "inferenceConfiguration": { ... },
        "modelId": "string",
        "name": "string",
        "templateConfiguration": { ... },
        "templateType": "string"
      }
    ],
    "version": "string"
  }

```

Response Elements

If the action is successful, the service sends back an HTTP 201 response.

The following data is returned in JSON format by the service.

arn

The Amazon Resource Name (ARN) of the version of the prompt.

Type: String

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:prompt/[0-9a-zA-Z]{10}(?::[0-9]{1,5})?)$`

createdAt

The time at which the prompt was created.

Type: Timestamp

customerEncryptionKeyArn

The Amazon Resource Name (ARN) of the KMS key to encrypt the version of the prompt.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

defaultVariant

The name of the default variant for the prompt. This value must match the name field in the relevant [PromptVariant](#) object.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?)\{1,100\}$`

description

A description for the prompt version.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

id

The unique identifier of the prompt.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

name

The name of the prompt version.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?)\{1,100\}$`

updatedAt

The time at which the prompt was last updated.

Type: Timestamp

variants

A list of objects, each containing details about a variant of the prompt.

Type: Array of [PromptVariant](#) objects

Array Members: Minimum number of 0 items. Maximum number of 3 items.

version

The version of the prompt that was created. Versions are numbered incrementally, starting from 1.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: `^(DRAFT|[0-9]{0,4}[1-9][0-9]{0,4})$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

DeleteAgent

Service: Agents for Amazon Bedrock

Deletes an agent.

Request Syntax

```
DELETE /agents/agentId?skipResourceInUseCheck=skipResourceInUseCheck HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

agentId

The unique identifier of the agent to delete.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

skipResourceInUseCheck

By default, this value is `false` and deletion is stopped if the resource is in use. If you set it to `true`, the resource will be deleted even if the resource is in use.

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 202
Content-type: application/json

{
  "agentId": "string",
  "agentStatus": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 202 response.

The following data is returned in JSON format by the service.

agentId

The unique identifier of the agent that was deleted.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

agentStatus

The status of the agent.

Type: String

Valid Values: CREATING | PREPARING | PREPARED | NOT_PREPARED | DELETING | FAILED | VERSIONING | UPDATING

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of DeleteAgent.

```
DELETE /agents/ABCDEFGHIJ/ HTTP/1.1
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

DeleteAgentActionGroup

Service: Agents for Amazon Bedrock

Deletes an action group in an agent.

Request Syntax

```
DELETE /agents/agentId/agentversions/agentVersion/actiongroups/actionGroupId?  
skipResourceInUseCheck=skipResourceInUseCheck HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

actionGroupId

The unique identifier of the action group to delete.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentId

The unique identifier of the agent that the action group belongs to.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentVersion

The version of the agent that the action group belongs to.

Length Constraints: Fixed length of 5.

Pattern: `^DRAFT$`

Required: Yes

skipResourceInUseCheck

By default, this value is `false` and deletion is stopped if the resource is in use. If you set it to `true`, the resource will be deleted even if the resource is in use.

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 204
```

Response Elements

If the action is successful, the service sends back an HTTP 204 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of DeleteAgentActionGroup.

```
DELETE /agents/ABCDEFGHIJ/agentversions/1/actiongroups/ABCDEFGHIJ/ HTTP/1.1
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

DeleteAgentAlias

Service: Agents for Amazon Bedrock

Deletes an alias of an agent.

Request Syntax

```
DELETE /agents/agentId/agentaliases/agentAliasId/ HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

agentAliasId

The unique identifier of the alias to delete.

Length Constraints: Fixed length of 10.

Pattern: $^(\backslash\text{TSTALIASID}\backslash\text{b} | [0-9a-zA-Z]+)\$$

Required: Yes

agentId

The unique identifier of the agent that the alias belongs to.

Pattern: $^[0-9a-zA-Z]{10}\$$

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 202
Content-type: application/json

{
  "agentAliasId": "string",
  "agentAliasStatus": "string",
```

```
"agentId": "string"  
}
```

Response Elements

If the action is successful, the service sends back an HTTP 202 response.

The following data is returned in JSON format by the service.

agentAliasId

The unique identifier of the alias that was deleted.

Type: String

Length Constraints: Fixed length of 10.

Pattern: $^(\backslash\text{TSTALIASID}\backslash\text{b} | [0-9a-zA-Z]+)\$$

agentAliasStatus

The status of the alias.

Type: String

Valid Values: CREATING | PREPARED | FAILED | UPDATING | DELETING

agentId

The unique identifier of the agent that the alias belongs to.

Type: String

Pattern: $^[0-9a-zA-Z]{10}\$$

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of `DeleteAgentAlias`.

```
DELETE /agents/ABCDEFGHIJ/agentaliases/ABCDEFGHIJ/ HTTP/1.1
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)

- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

DeleteAgentVersion

Service: Agents for Amazon Bedrock

Deletes a version of an agent.

Request Syntax

```
DELETE /agents/agentId/agentversions/agentVersion?  
skipResourceInUseCheck=skipResourceInUseCheck HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

agentId

The unique identifier of the agent that the version belongs to.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentVersion

The version of the agent to delete.

Pattern: `^[0-9]{1,5}$`

Required: Yes

skipResourceInUseCheck

By default, this value is `false` and deletion is stopped if the resource is in use. If you set it to `true`, the resource will be deleted even if the resource is in use.

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 202  
Content-type: application/json
```



```
{
  "agentId": "string",
  "agentStatus": "string",
  "agentVersion": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 202 response.

The following data is returned in JSON format by the service.

agentId

The unique identifier of the agent that the version belongs to.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

agentStatus

The status of the agent version.

Type: String

Valid Values: CREATING | PREPARING | PREPARED | NOT_PREPARED | DELETING | FAILED | VERSIONING | UPDATING

agentVersion

The version that was deleted.

Type: String

Pattern: `^[0-9]{1,5}$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of `DeleteAgentVersion`.

```
DELETE /agents/ABCDEFGHIJ/agentversions/1/?skipResourceInUseCheck=true HTTP/1.1
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

DeleteDataSource

Service: Agents for Amazon Bedrock

Deletes a data source from a knowledge base.

Request Syntax

```
DELETE /knowledgebases/knowledgeBaseId/datasources/dataSourceId HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

dataSourceId

The unique identifier of the data source to delete.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

knowledgeBaseId

The unique identifier of the knowledge base from which to delete the data source.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 202
Content-type: application/json

{
  "dataSourceId": "string",
  "knowledgeBaseId": "string",
  "status": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 202 response.

The following data is returned in JSON format by the service.

dataSourceId

The unique identifier of the data source that was deleted.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

knowledgeBaseId

The unique identifier of the knowledge base to which the data source that was deleted belonged.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

status

The status of the data source.

Type: String

Valid Values: AVAILABLE | DELETING | DELETE_UNSUCCESSFUL

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

DeleteFlow

Service: Agents for Amazon Bedrock

Deletes a flow.

Request Syntax

```
DELETE /flows/flowIdentifier?skipResourceInUseCheck=skipResourceInUseCheck HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

flowIdentifier

The unique identifier of the flow.

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10})|([0-9a-zA-Z]{10})$`

Required: Yes

skipResourceInUseCheck

By default, this value is `false` and deletion is stopped if the resource is in use. If you set it to `true`, the resource will be deleted even if the resource is in use.

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "id": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

id

The unique identifier of the flow.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

DeleteFlowAlias

Service: Agents for Amazon Bedrock

Deletes an alias of a flow.

Request Syntax

```
DELETE /flows/flowIdentifier/aliases/aliasIdentifier HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

aliasIdentifier

The unique identifier of the alias to be deleted.

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10}/alias/[0-9a-zA-Z]{10})|(TSTALIASID|[0-9a-zA-Z]{10})$`

Required: Yes

flowIdentifier

The unique identifier of the flow that the alias belongs to.

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10})|([0-9a-zA-Z]{10})$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "flowId": "string",
```

```
"id": "string"  
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

flowId

The unique identifier of the flow that the alias belongs to.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

id

The unique identifier of the flow.

Type: String

Pattern: `^(TSTALIASID|[0-9a-zA-Z]{10})$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

DeleteFlowVersion

Service: Agents for Amazon Bedrock

Deletes a version of a flow.

Request Syntax

```
DELETE /flows/flowIdentifier/versions/flowVersion?  
skipResourceInUseCheck=skipResourceInUseCheck HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

flowIdentifier

The unique identifier of the flow whose version that you want to delete

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10})|([0-9a-zA-Z]{10})$`

Required: Yes

flowVersion

The version of the flow that you want to delete.

Pattern: `^[0-9]{1,5}$`

Required: Yes

skipResourceInUseCheck

By default, this value is `false` and deletion is stopped if the resource is in use. If you set it to `true`, the resource will be deleted even if the resource is in use.

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
```

```
Content-type: application/json

{
  "id": "string",
  "version": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

id

The unique identifier of the flow.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

version

The version of the flow being deleted.

Type: String

Pattern: `^[0-9]{1,5}$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

DeleteKnowledgeBase

Service: Agents for Amazon Bedrock

Deletes a knowledge base. Before deleting a knowledge base, you should disassociate the knowledge base from any agents that it is associated with by making a [DisassociateAgentKnowledgeBase](#) request.

Request Syntax

```
DELETE /knowledgebases/knowledgeBaseId HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

knowledgeBaseId

The unique identifier of the knowledge base to delete.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 202
Content-type: application/json

{
  "knowledgeBaseId": "string",
  "status": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 202 response.

The following data is returned in JSON format by the service.

knowledgeBaseId

The unique identifier of the knowledge base that was deleted.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

status

The status of the knowledge base and whether it has been successfully deleted.

Type: String

Valid Values: CREATING | ACTIVE | DELETING | UPDATING | FAILED | DELETE_UNSUCCESSFUL

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

DeletePrompt

Service: Agents for Amazon Bedrock

Deletes a prompt or a prompt version from the Prompt management tool. For more information, see [Delete prompts from the Prompt management tool](#) and [Delete a version of a prompt from the Prompt management tool](#) in the Amazon Bedrock User Guide.

Request Syntax

```
DELETE /prompts/promptIdentifier?promptVersion=promptVersion HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

promptIdentifier

The unique identifier of the prompt.

Pattern: `^([0-9a-zA-Z]{10})|(arn:aws:bedrock:[a-z0-9-]{1,20}: [0-9]{12}:prompt/[0-9a-zA-Z]{10})(?::[0-9]{1,5})?$`

Required: Yes

promptVersion

The version of the prompt to delete.

Pattern: `^[0-9]{1,5}$`

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "id": "string",
  "version": "string"
}
```

```
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

id

The unique identifier of the prompt that was deleted.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

version

The version of the prompt that was deleted.

Type: String

Pattern: `^[0-9]{1,5}$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

DisassociateAgentKnowledgeBase

Service: Agents for Amazon Bedrock

Disassociates a knowledge base from an agent.

Request Syntax

```
DELETE /agents/agentId/agentversions/agentVersion/knowledgebases/knowledgeBaseId/
HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

agentId

The unique identifier of the agent from which to disassociate the knowledge base.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentVersion

The version of the agent from which to disassociate the knowledge base.

Length Constraints: Fixed length of 5.

Pattern: `^DRAFT$`

Required: Yes

knowledgeBaseId

The unique identifier of the knowledge base to disassociate.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 204
```

Response Elements

If the action is successful, the service sends back an HTTP 204 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetAgent

Service: Agents for Amazon Bedrock

Gets information about an agent.

Request Syntax

```
GET /agents/agentId/ HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

agentId

The unique identifier of the agent.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "agent": {
    "agentArn": "string",
    "agentId": "string",
    "agentName": "string",
    "agentResourceRoleArn": "string",
    "agentStatus": "string",
    "agentVersion": "string",
    "clientToken": "string",
    "createdAt": "string",
    "customerEncryptionKeyArn": "string",
```

```

    "description": "string",
    "failureReasons": [ "string" ],
    "foundationModel": "string",
    "guardrailConfiguration": {
      "guardrailIdentifier": "string",
      "guardrailVersion": "string"
    },
    "idleSessionTTLInSeconds": number,
    "instruction": "string",
    "memoryConfiguration": {
      "enabledMemoryTypes": [ "string" ],
      "storageDays": number
    },
    "preparedAt": "string",
    "promptOverrideConfiguration": {
      "overrideLambda": "string",
      "promptConfigurations": [
        {
          "basePromptTemplate": "string",
          "inferenceConfiguration": {
            "maxLength": number,
            "stopSequences": [ "string" ],
            "temperature": number,
            "topK": number,
            "topP": number
          },
          "parserMode": "string",
          "promptCreationMode": "string",
          "promptState": "string",
          "promptType": "string"
        }
      ]
    },
    "recommendedActions": [ "string" ],
    "updatedAt": "string"
  }
}

```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

[agent](#)

Contains details about the agent.

Type: [Agent](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of `GetAgent`.

```
GET /agents/ABCDEFGHIJ/ HTTP/1.1
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetAgentActionGroup

Service: Agents for Amazon Bedrock

Gets information about an action group for an agent.

Request Syntax

```
GET /agents/agentId/agentversions/agentVersion/actiongroups/actionGroupId/ HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

actionGroupId

The unique identifier of the action group for which to get information.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentId

The unique identifier of the agent that the action group belongs to.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentVersion

The version of the agent that the action group belongs to.

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: `^(DRAFT|[0-9]{0,4}[1-9][0-9]{0,4})$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "agentActionGroup": {
    "actionGroupExecutor": { ... },
    "actionGroupId": "string",
    "actionGroupName": "string",
    "actionGroupState": "string",
    "agentId": "string",
    "agentVersion": "string",
    "apiSchema": { ... },
    "clientToken": "string",
    "createdAt": "string",
    "description": "string",
    "functionSchema": { ... },
    "parentActionSignature": "string",
    "updatedAt": "string"
  }
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

[agentActionGroup](#)

Contains details about the action group.

Type: [AgentActionGroup](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of `GetAgentActionGroup`.

```
GET /agents/AGENT12345/agentversions/1/actiongroups/ACTION1234/ HTTP/1.1
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)

- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetAgentAlias

Service: Agents for Amazon Bedrock

Gets information about an alias of an agent.

Request Syntax

```
GET /agents/agentId/agentaliases/agentAliasId/ HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

agentAliasId

The unique identifier of the alias for which to get information.

Length Constraints: Fixed length of 10.

Pattern: $^(\backslash\text{TSTALIASID}\backslash\text{b} | [0-9a-zA-Z]^+)\$$

Required: Yes

agentId

The unique identifier of the agent to which the alias to get information belongs.

Pattern: $^[0-9a-zA-Z]{10}\$$

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
```

```

"agentAlias": {
  "agentAliasArn": "string",
  "agentAliasHistoryEvents": [
    {
      "endDate": "string",
      "routingConfiguration": [
        {
          "agentVersion": "string",
          "provisionedThroughput": "string"
        }
      ],
      "startDate": "string"
    }
  ],
  "agentAliasId": "string",
  "agentAliasName": "string",
  "agentAliasStatus": "string",
  "agentId": "string",
  "clientToken": "string",
  "createdAt": "string",
  "description": "string",
  "failureReasons": [ "string" ],
  "routingConfiguration": [
    {
      "agentVersion": "string",
      "provisionedThroughput": "string"
    }
  ],
  "updatedAt": "string"
}
}

```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

agentAlias

Contains information about the alias.

Type: [AgentAlias](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of `GetAgentAlias`.

```
GET /agents/ABCDEFGHIJ/agentaliases/ABCDEFGHIJ/ HTTP/1.1
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetAgentKnowledgeBase

Service: Agents for Amazon Bedrock

Gets information about a knowledge base associated with an agent.

Request Syntax

```
GET /agents/agentId/agentversions/agentVersion/knowledgebases/knowledgeBaseId/ HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

agentId

The unique identifier of the agent with which the knowledge base is associated.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentVersion

The version of the agent with which the knowledge base is associated.

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: `^(DRAFT|[0-9]{0,4}[1-9][0-9]{0,4})$`

Required: Yes

knowledgeBaseId

The unique identifier of the knowledge base associated with the agent.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "agentKnowledgeBase": {
    "agentId": "string",
    "agentVersion": "string",
    "createdAt": "string",
    "description": "string",
    "knowledgeBaseId": "string",
    "knowledgeBaseState": "string",
    "updatedAt": "string"
  }
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

agentKnowledgeBase

Contains details about a knowledge base attached to an agent.

Type: [AgentKnowledgeBase](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetAgentVersion

Service: Agents for Amazon Bedrock

Gets details about a version of an agent.

Request Syntax

```
GET /agents/agentId/agentversions/agentVersion/ HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

agentId

The unique identifier of the agent.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentVersion

The version of the agent.

Pattern: `^[0-9]{1,5}$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "agentVersion": {
    "agentArn": "string",
    "agentId": "string",
    "agentName": "string",
    "agentResourceRoleArn": "string",
```

```

    "agentStatus": "string",
    "createdAt": "string",
    "customerEncryptionKeyArn": "string",
    "description": "string",
    "failureReasons": [ "string" ],
    "foundationModel": "string",
    "guardrailConfiguration": {
      "guardrailIdentifier": "string",
      "guardrailVersion": "string"
    },
    "idleSessionTTLInSeconds": number,
    "instruction": "string",
    "memoryConfiguration": {
      "enabledMemoryTypes": [ "string" ],
      "storageDays": number
    },
    "promptOverrideConfiguration": {
      "overrideLambda": "string",
      "promptConfigurations": [
        {
          "basePromptTemplate": "string",
          "inferenceConfiguration": {
            "maxLength": number,
            "stopSequences": [ "string" ],
            "temperature": number,
            "topK": number,
            "topP": number
          },
          "parserMode": "string",
          "promptCreationMode": "string",
          "promptState": "string",
          "promptType": "string"
        }
      ]
    },
    "recommendedActions": [ "string" ],
    "updatedAt": "string",
    "version": "string"
  }
}

```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

agentVersion

Contains details about the version of the agent.

Type: [AgentVersion](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of `GetAgentVersion`.

```
GET /agents/agentId/agentversions/agentVersion/ HTTP/1.1
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetDataSource

Service: Agents for Amazon Bedrock

Gets information about a data source.

Request Syntax

```
GET /knowledgebases/knowledgeBaseId/datasources/dataSourceId HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

dataSourceId

The unique identifier of the data source.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

knowledgeBaseId

The unique identifier of the knowledge base that the data source was added to.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "dataSource": {
    "createdAt": "string",
    "dataDeletionPolicy": "string",
    "dataSourceConfiguration": {
      "confluenceConfiguration": {
```

```
"crawlerConfiguration": {
  "filterConfiguration": {
    "patternObjectFilter": {
      "filters": [
        {
          "exclusionFilters": [ "string" ],
          "inclusionFilters": [ "string" ],
          "objectType": "string"
        }
      ]
    },
    "type": "string"
  }
},
"sourceConfiguration": {
  "authType": "string",
  "credentialsSecretArn": "string",
  "hostType": "string",
  "hostUrl": "string"
}
},
"s3Configuration": {
  "bucketArn": "string",
  "bucketOwnerAccountId": "string",
  "inclusionPrefixes": [ "string" ]
},
"salesforceConfiguration": {
  "crawlerConfiguration": {
    "filterConfiguration": {
      "patternObjectFilter": {
        "filters": [
          {
            "exclusionFilters": [ "string" ],
            "inclusionFilters": [ "string" ],
            "objectType": "string"
          }
        ]
      },
      "type": "string"
    }
  }
},
"sourceConfiguration": {
  "authType": "string",
  "credentialsSecretArn": "string",
```

```

    "hostUrl": "string"
  }
},
"sharePointConfiguration": {
  "crawlerConfiguration": {
    "filterConfiguration": {
      "patternObjectFilter": {
        "filters": [
          {
            "exclusionFilters": [ "string" ],
            "inclusionFilters": [ "string" ],
            "objectType": "string"
          }
        ]
      },
      "type": "string"
    }
  },
  "sourceConfiguration": {
    "authType": "string",
    "credentialsSecretArn": "string",
    "domain": "string",
    "hostType": "string",
    "siteUrls": [ "string" ],
    "tenantId": "string"
  }
},
"type": "string",
"webConfiguration": {
  "crawlerConfiguration": {
    "crawlerLimits": {
      "rateLimit": number
    },
    "exclusionFilters": [ "string" ],
    "inclusionFilters": [ "string" ],
    "scope": "string"
  },
  "sourceConfiguration": {
    "urlConfiguration": {
      "seedUrls": [
        {
          "url": "string"
        }
      ]
    }
  }
}
]

```

```

    }
  }
}
},
"dataSourceId": "string",
"description": "string",
"failureReasons": [ "string" ],
"knowledgeBaseId": "string",
"name": "string",
"serverSideEncryptionConfiguration": {
  "kmsKeyArn": "string"
},
"status": "string",
"updatedAt": "string",
"vectorIngestionConfiguration": {
  "chunkingConfiguration": {
    "chunkingStrategy": "string",
    "fixedSizeChunkingConfiguration": {
      "maxTokens": number,
      "overlapPercentage": number
    },
    "hierarchicalChunkingConfiguration": {
      "levelConfigurations": [
        {
          "maxTokens": number
        }
      ],
      "overlapTokens": number
    },
    "semanticChunkingConfiguration": {
      "breakpointPercentileThreshold": number,
      "bufferSize": number,
      "maxTokens": number
    }
  },
  "customTransformationConfiguration": {
    "intermediateStorage": {
      "s3Location": {
        "uri": "string"
      }
    },
    "transformations": [
      {
        "stepToApply": "string",

```



```
        "transformationFunction": {
          "transformationLambdaConfiguration": {
            "lambdaArn": "string"
          }
        }
      ],
    },
    "parsingConfiguration": {
      "bedrockFoundationModelConfiguration": {
        "modelArn": "string",
        "parsingPrompt": {
          "parsingPromptText": "string"
        }
      },
      "parsingStrategy": "string"
    }
  }
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

dataSource

Contains details about the data source.

Type: [DataSource](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetFlow

Service: Agents for Amazon Bedrock

Retrieves information about a flow. For more information, see [Manage a flow in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Request Syntax

```
GET /flows/flowIdentifier/ HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

flowIdentifier

The unique identifier of the flow.

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10})|([0-9a-zA-Z]{10})$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "arn": "string",
  "createdAt": "string",
  "customerEncryptionKeyArn": "string",
  "definition": {
    "connections": [
      {
        "configuration": { ... },
        "name": "string",
```

```

        "source": "string",
        "target": "string",
        "type": "string"
    }
],
"nodes": [
    {
        "configuration": { ... },
        "inputs": [
            {
                "expression": "string",
                "name": "string",
                "type": "string"
            }
        ],
        "name": "string",
        "outputs": [
            {
                "name": "string",
                "type": "string"
            }
        ],
        "type": "string"
    }
]
},
"description": "string",
"executionRoleArn": "string",
"id": "string",
"name": "string",
"status": "string",
"updatedAt": "string",
"validations": [
    {
        "message": "string",
        "severity": "string"
    }
],
"version": "string"
}

```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

arn

The Amazon Resource Name (ARN) of the flow.

Type: String

Pattern: `^arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10}$`

createdAt

The time at which the flow was created.

Type: Timestamp

customerEncryptionKeyArn

The Amazon Resource Name (ARN) of the KMS key that the flow is encrypted with.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

definition

The definition of the nodes and connections between the nodes in the flow.

Type: [FlowDefinition](#) object

description

The description of the flow.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

executionRoleArn

The Amazon Resource Name (ARN) of the service role with permissions to create a flow. For more information, see [Create a service row for flows](#) in the Amazon Bedrock User Guide.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^:]+)?:iam::([0-9]{12})?:role/(service-role/)?\.\+$`

id

The unique identifier of the flow.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

name

The name of the flow.

Type: String

Pattern: `^([0-9a-zA-Z][_ -?]){1,100}$`

status

The status of the flow. The following statuses are possible:

- **NotPrepared** – The flow has been created or updated, but hasn't been prepared. If you just created the flow, you can't test it. If you updated the flow, the DRAFT version won't contain the latest changes for testing. Send a [PrepareFlow](#) request to package the latest changes into the DRAFT version.
- **Preparing** – The flow is being prepared so that the DRAFT version contains the latest changes for testing.
- **Prepared** – The flow is prepared and the DRAFT version contains the latest changes for testing.
- **Failed** – The last API operation that you invoked on the flow failed. Send a [GetFlow](#) request and check the error message in the `validations` field.

Type: String

Valid Values: `Failed | Prepared | Preparing | NotPrepared`

updatedAt

The time at which the flow was last updated.

Type: Timestamp

validations

A list of validation error messages related to the last failed operation on the flow.

Type: Array of [FlowValidation](#) objects

Array Members: Minimum number of 0 items. Maximum number of 100 items.

version

The version of the flow for which information was retrieved.

Type: String

Length Constraints: Fixed length of 5.

Pattern: ^DRAFT\$

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetFlowAlias

Service: Agents for Amazon Bedrock

Retrieves information about a flow. For more information, see [Deploy a flow in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Request Syntax

```
GET /flows/flowIdentifier/aliases/aliasIdentifier HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

aliasIdentifier

The unique identifier of the alias for which to retrieve information.

Pattern: $^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10}/alias/[0-9a-zA-Z]{10})|(TSTALIASID|[0-9a-zA-Z]{10})\$$

Required: Yes

flowIdentifier

The unique identifier of the flow that the alias belongs to.

Pattern: $^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10})|([0-9a-zA-Z]{10})\$$

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
```

```
"arn": "string",
"createdAt": "string",
"description": "string",
"flowId": "string",
"id": "string",
"name": "string",
"routingConfiguration": [
  {
    "flowVersion": "string"
  }
],
"updatedAt": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

arn

The Amazon Resource Name (ARN) of the flow.

Type: String

Pattern: `^arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10}/alias/(TSTALIASID|[0-9a-zA-Z]{10})$`

createdAt

The time at which the flow was created.

Type: Timestamp

description

The description of the flow.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

flowId

The unique identifier of the flow that the alias belongs to.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

id

The unique identifier of the alias of the flow.

Type: String

Pattern: `^(TSTALIASID|[0-9a-zA-Z]{10})$`

name

The name of the flow alias.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

routingConfiguration

Contains information about the version that the alias is mapped to.

Type: Array of [FlowAliasRoutingConfigurationListItem](#) objects

Array Members: Fixed number of 1 item.

updatedAt

The time at which the flow alias was last updated.

Type: Timestamp

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetFlowVersion

Service: Agents for Amazon Bedrock

Retrieves information about a version of a flow. For more information, see [Deploy a flow in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Request Syntax

```
GET /flows/flowIdentifier/versions/flowVersion/ HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

flowIdentifier

The unique identifier of the flow for which to get information.

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10})|([0-9a-zA-Z]{10})$`

Required: Yes

flowVersion

The version of the flow for which to get information.

Pattern: `^[0-9]{1,5}$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "arn": "string",
```

```

"createdAt": "string",
"customerEncryptionKeyArn": "string",
"definition": {
  "connections": [
    {
      "configuration": { ... },
      "name": "string",
      "source": "string",
      "target": "string",
      "type": "string"
    }
  ],
  "nodes": [
    {
      "configuration": { ... },
      "inputs": [
        {
          "expression": "string",
          "name": "string",
          "type": "string"
        }
      ],
      "name": "string",
      "outputs": [
        {
          "name": "string",
          "type": "string"
        }
      ],
      "type": "string"
    }
  ]
},
"description": "string",
"executionRoleArn": "string",
"id": "string",
"name": "string",
"status": "string",
"version": "string"
}

```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

arn

The Amazon Resource Name (ARN) of the flow.

Type: String

Pattern: `^arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10}$`

createdAt

The time at which the flow was created.

Type: Timestamp

customerEncryptionKeyArn

The Amazon Resource Name (ARN) of the KMS key that the version of the flow is encrypted with.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

definition

The definition of the nodes and connections between nodes in the flow.

Type: [FlowDefinition](#) object

description

The description of the flow.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

executionRoleArn

The Amazon Resource Name (ARN) of the service role with permissions to create a flow. For more information, see [Create a service role for flows in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:iam::([0-9]{12})?:role/(service-role/)?\.\+$`

id

The unique identifier of the flow.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

name

The name of the flow version.

Type: String

Pattern: `^([0-9a-zA-Z][_ -?]){1,100}$`

status

The status of the flow.

Type: String

Valid Values: Failed | Prepared | Preparing | NotPrepared

version

The version of the flow for which information was retrieved.

Type: String

Pattern: `^[0-9]{1,5}$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetIngestionJob

Service: Agents for Amazon Bedrock

Gets information about a ingestion job, in which a data source is added to a knowledge base.

Request Syntax

```
GET /knowledgebases/knowledgeBaseId/datasources/dataSourceId/
ingestionjobs/ingestionJobId HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

dataSourceId

The unique identifier of the data source in the ingestion job.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

ingestionJobId

The unique identifier of the ingestion job.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

knowledgeBaseId

The unique identifier of the knowledge base for which the ingestion job applies.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
```

Content-type: application/json

```
{
  "ingestionJob": {
    "dataSourceId": "string",
    "description": "string",
    "failureReasons": [ "string" ],
    "ingestionJobId": "string",
    "knowledgeBaseId": "string",
    "startedAt": "string",
    "statistics": {
      "numberOfDocumentsDeleted": number,
      "numberOfDocumentsFailed": number,
      "numberOfDocumentsScanned": number,
      "numberOfMetadataDocumentsModified": number,
      "numberOfMetadataDocumentsScanned": number,
      "numberOfModifiedDocumentsIndexed": number,
      "numberOfNewDocumentsIndexed": number
    },
    "status": "string",
    "updatedAt": "string"
  }
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

ingestionJob

Contains details about the ingestion job.

Type: [IngestionJob](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetKnowledgeBase

Service: Agents for Amazon Bedrock

Gets information about a knowledge base.

Request Syntax

```
GET /knowledgebases/knowledgeBaseId HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

knowledgeBaseId

The unique identifier of the knowledge base for which to get information.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "knowledgeBase": {
    "createdAt": "string",
    "description": "string",
    "failureReasons": [ "string ],
    "knowledgeBaseArn": "string",
    "knowledgeBaseConfiguration": {
      "type": "string",
      "vectorKnowledgeBaseConfiguration": {
        "embeddingModelArn": "string",
        "embeddingModelConfiguration": {
          "bedrockEmbeddingModelConfiguration": {
            "dimensions": number
          }
        }
      }
    }
  }
}
```

```

    }
  }
},
"knowledgeBaseId": "string",
"name": "string",
"roleArn": "string",
"status": "string",
"storageConfiguration": {
  "mongoDbAtlasConfiguration": {
    "collectionName": "string",
    "credentialsSecretArn": "string",
    "databaseName": "string",
    "endpoint": "string",
    "endpointServiceName": "string",
    "fieldMapping": {
      "metadataField": "string",
      "textField": "string",
      "vectorField": "string"
    },
    "vectorIndexName": "string"
  },
  "opensearchServerlessConfiguration": {
    "collectionArn": "string",
    "fieldMapping": {
      "metadataField": "string",
      "textField": "string",
      "vectorField": "string"
    },
    "vectorIndexName": "string"
  },
  "pineconeConfiguration": {
    "connectionString": "string",
    "credentialsSecretArn": "string",
    "fieldMapping": {
      "metadataField": "string",
      "textField": "string"
    },
    "namespace": "string"
  },
  "rdsConfiguration": {
    "credentialsSecretArn": "string",
    "databaseName": "string",
    "fieldMapping": {

```



```

        "metadataField": "string",
        "primaryKeyField": "string",
        "textField": "string",
        "vectorField": "string"
    },
    "resourceArn": "string",
    "tableName": "string"
},
"redisEnterpriseCloudConfiguration": {
    "credentialsSecretArn": "string",
    "endpoint": "string",
    "fieldMapping": {
        "metadataField": "string",
        "textField": "string",
        "vectorField": "string"
    },
    "vectorIndexName": "string"
},
"type": "string"
},
"updatedAt": "string"
}
}

```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

knowledgeBase

Contains details about the knowledge base.

Type: [KnowledgeBase](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetPrompt

Service: Agents for Amazon Bedrock

Retrieves information about a prompt or a version of it. For more information, see [View information about prompts using Prompt management](#) and [View information about a version of your prompt](#) in the Amazon Bedrock User Guide.

Request Syntax

```
GET /prompts/promptIdentifier?promptVersion=promptVersion HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

promptIdentifier

The unique identifier of the prompt.

Pattern: `^([0-9a-zA-Z]{10})|(arn:aws:bedrock:[a-z0-9-]{1,20}::[0-9]{12}:prompt/[0-9a-zA-Z]{10})(?::[0-9]{1,5})?$`

Required: Yes

promptVersion

The version of the prompt about which you want to retrieve information.

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: `^(DRAFT|[0-9]{0,4}[1-9][0-9]{0,4})$`

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200  
Content-type: application/json
```

```

{
  "arn": "string",
  "createdAt": "string",
  "customerEncryptionKeyArn": "string",
  "defaultVariant": "string",
  "description": "string",
  "id": "string",
  "name": "string",
  "updatedAt": "string",
  "variants": [
    {
      "inferenceConfiguration": { ... },
      "modelId": "string",
      "name": "string",
      "templateConfiguration": { ... },
      "templateType": "string"
    }
  ],
  "version": "string"
}

```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

arn

The Amazon Resource Name (ARN) of the prompt.

Type: String

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:prompt/[0-9a-zA-Z]{10}(?::[0-9]{1,5})?)$`

createdAt

The time at which the prompt was created.

Type: Timestamp

customerEncryptionKeyArn

The Amazon Resource Name (ARN) of the KMS key that the prompt is encrypted with.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

defaultVariant

The name of the default variant for the prompt. This value must match the name field in the relevant [PromptVariant](#) object.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

description

The description of the prompt.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

id

The unique identifier of the prompt.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

name

The name of the prompt.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

updatedAt

The time at which the prompt was last updated.

Type: Timestamp

variants

A list of objects, each containing details about a variant of the prompt.

Type: Array of [PromptVariant](#) objects

Array Members: Minimum number of 0 items. Maximum number of 3 items.

version

The version of the prompt.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: `^(DRAFT|[0-9]{0,4}[1-9][0-9]{0,4})$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListAgentActionGroups

Service: Agents for Amazon Bedrock

Lists the action groups for an agent and information about each one.

Request Syntax

```
POST /agents/agentId/agentversions/agentVersion/actiongroups/ HTTP/1.1  
Content-type: application/json
```

```
{  
  "maxResults": number,  
  "nextToken": "string"  
}
```

URI Request Parameters

The request uses the following URI parameters.

agentId

The unique identifier of the agent.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentVersion

The version of the agent.

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: `^(DRAFT|[0-9]{0,4}[1-9][0-9]{0,4})$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

maxResults

The maximum number of results to return in the response. If the total number of results is greater than this value, use the token returned in the response in the `nextToken` field when making another request to return the next batch of results.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 1000.

Required: No

nextToken

If the total number of results is greater than the `maxResults` value provided in the request, enter the token returned in the `nextToken` field in the response in this field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "actionGroupSummaries": [
    {
      "actionGroupId": "string",
      "actionGroupName": "string",
      "actionGroupState": "string",
      "description": "string",
      "updatedAt": "string"
    }
  ],
  "nextToken": "string"
```

```
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

actionGroupSummaries

A list of objects, each of which contains information about an action group.

Type: Array of [ActionGroupSummary](#) objects

Array Members: Minimum number of 0 items. Maximum number of 10 items.

nextToken

If the total number of results is greater than the `maxResults` value provided in the request, use this token when making another request in the `nextToken` field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of ListAgentActionGroups.

```
POST /agents/AGENT12345/agentversions/1/actiongroups/ HTTP/1.1
Content-type: application/json

{
  "maxResults": 10
}
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)

- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListAgentAliases

Service: Agents for Amazon Bedrock

Lists the aliases of an agent and information about each one.

Request Syntax

```
POST /agents/agentId/agentaliases/ HTTP/1.1
Content-type: application/json
```

```
{
  "maxResults": number,
  "nextToken": "string"
}
```

URI Request Parameters

The request uses the following URI parameters.

agentId

The unique identifier of the agent.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

maxResults

The maximum number of results to return in the response. If the total number of results is greater than this value, use the token returned in the response in the `nextToken` field when making another request to return the next batch of results.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 1000.

Required: No

nextToken

If the total number of results is greater than the `maxResults` value provided in the request, enter the token returned in the `nextToken` field in the response in this field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "agentAliasSummaries": [
    {
      "agentAliasId": "string",
      "agentAliasName": "string",
      "agentAliasStatus": "string",
      "createdAt": "string",
      "description": "string",
      "routingConfiguration": [
        {
          "agentVersion": "string",
          "provisionedThroughput": "string"
        }
      ],
      "updatedAt": "string"
    }
  ],
  "nextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

agentAliasSummaries

A list of objects, each of which contains information about an alias of the agent.

Type: Array of [AgentAliasSummary](#) objects

Array Members: Minimum number of 0 items. Maximum number of 10 items.

nextToken

If the total number of results is greater than the `maxResults` value provided in the request, use this token when making another request in the `nextToken` field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of ListAgentAliases.

```
POST /agents/ABCDEFGHIJ/agentaliases/ HTTP/1.1
Content-type: application/json

{
  "maxResults": 10
}
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListAgentKnowledgeBases

Service: Agents for Amazon Bedrock

Lists knowledge bases associated with an agent and information about each one.

Request Syntax

```
POST /agents/agentId/agentversions/agentVersion/knowledgebases/ HTTP/1.1
Content-type: application/json
```

```
{
  "maxResults": number,
  "nextToken": "string"
}
```

URI Request Parameters

The request uses the following URI parameters.

agentId

The unique identifier of the agent for which to return information about knowledge bases associated with it.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentVersion

The version of the agent for which to return information about knowledge bases associated with it.

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: `^(DRAFT|[0-9]{0,4}[1-9][0-9]{0,4})$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

maxResults

The maximum number of results to return in the response. If the total number of results is greater than this value, use the token returned in the response in the nextToken field when making another request to return the next batch of results.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 1000.

Required: No

nextToken

If the total number of results is greater than the maxResults value provided in the request, enter the token returned in the nextToken field in the response in this field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "agentKnowledgeBaseSummaries": [
    {
      "description": "string",
      "knowledgeBaseId": "string",
      "knowledgeBaseState": "string",
      "updatedAt": "string"
    }
  ],
  "nextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

agentKnowledgeBaseSummaries

A list of objects, each of which contains information about a knowledge base associated with the agent.

Type: Array of [AgentKnowledgeBaseSummary](#) objects

Array Members: Minimum number of 0 items. Maximum number of 10 items.

nextToken

If the total number of results is greater than the `maxResults` value provided in the request, use this token when making another request in the `nextToken` field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListAgents

Service: Agents for Amazon Bedrock

Lists the agents belonging to an account and information about each agent.

Request Syntax

```
POST /agents/ HTTP/1.1
Content-type: application/json

{
  "maxResults": number,
  "nextToken": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

maxResults

The maximum number of results to return in the response. If the total number of results is greater than this value, use the token returned in the response in the nextToken field when making another request to return the next batch of results.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 1000.

Required: No

nextToken

If the total number of results is greater than the maxResults value provided in the request, enter the token returned in the nextToken field in the response in this field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "agentSummaries": [
    {
      "agentId": "string",
      "agentName": "string",
      "agentStatus": "string",
      "description": "string",
      "guardrailConfiguration": {
        "guardrailIdentifier": "string",
        "guardrailVersion": "string"
      },
      "latestAgentVersion": "string",
      "updatedAt": "string"
    }
  ],
  "nextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

agentSummaries

A list of objects, each of which contains information about an agent.

Type: Array of [AgentSummary](#) objects

Array Members: Minimum number of 0 items. Maximum number of 10 items.

nextToken

If the total number of results is greater than the `maxResults` value provided in the request, use this token when making another request in the `nextToken` field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of `ListAgents`.

```
POST /agents/ HTTP/1.1
Content-type: application/json

{
  "maxResults": 10
}
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListAgentVersions

Service: Agents for Amazon Bedrock

Lists the versions of an agent and information about each version.

Request Syntax

```
POST /agents/agentId/agentversions/ HTTP/1.1
Content-type: application/json
```

```
{
  "maxResults": number,
  "nextToken": "string"
}
```

URI Request Parameters

The request uses the following URI parameters.

agentId

The unique identifier of the agent.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

maxResults

The maximum number of results to return in the response. If the total number of results is greater than this value, use the token returned in the response in the `nextToken` field when making another request to return the next batch of results.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 1000.

Required: No

nextToken

If the total number of results is greater than the `maxResults` value provided in the request, enter the token returned in the `nextToken` field in the response in this field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "agentVersionSummaries": [
    {
      "agentName": "string",
      "agentStatus": "string",
      "agentVersion": "string",
      "createdAt": "string",
      "description": "string",
      "guardrailConfiguration": {
        "guardrailIdentifier": "string",
        "guardrailVersion": "string"
      },
      "updatedAt": "string"
    }
  ],
  "nextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

agentVersionSummaries

A list of objects, each of which contains information about a version of the agent.

Type: Array of [AgentVersionSummary](#) objects

Array Members: Minimum number of 0 items. Maximum number of 10 items.

nextToken

If the total number of results is greater than the `maxResults` value provided in the request, use this token when making another request in the `nextToken` field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of ListAgentVersions.

```
POST /agents/agentId/agentversions/ HTTP/1.1
Content-type: application/json

{
  "maxResults": 10
}
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListDataSources

Service: Agents for Amazon Bedrock

Lists the data sources in a knowledge base and information about each one.

Request Syntax

```
POST /knowledgebases/knowledgeBaseId/datasources/ HTTP/1.1  
Content-type: application/json
```

```
{  
  "maxResults": number,  
  "nextToken": "string"  
}
```

URI Request Parameters

The request uses the following URI parameters.

knowledgeBaseId

The unique identifier of the knowledge base for which to return a list of information.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

maxResults

The maximum number of results to return in the response. If the total number of results is greater than this value, use the token returned in the response in the `nextToken` field when making another request to return the next batch of results.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 1000.

Required: No

nextToken

If the total number of results is greater than the `maxResults` value provided in the request, enter the token returned in the `nextToken` field in the response in this field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "dataSourceSummaries": [
    {
      "dataSourceId": "string",
      "description": "string",
      "knowledgeBaseId": "string",
      "name": "string",
      "status": "string",
      "updatedAt": "string"
    }
  ],
  "nextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

dataSourceSummaries

A list of objects, each of which contains information about a data source.

Type: Array of [DataSourceSummary](#) objects

[nextToken](#)

If the total number of results is greater than the `maxResults` value provided in the request, use this token when making another request in the `nextToken` field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListFlowAliases

Service: Agents for Amazon Bedrock

Returns a list of aliases for a flow.

Request Syntax

```
GET /flows/flowIdentifier/aliases?maxResults=maxResults&nextToken=nextToken HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

flowIdentifier

The unique identifier of the flow for which aliases are being returned.

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10})|([0-9a-zA-Z]{10})$`

Required: Yes

maxResults

The maximum number of results to return in the response. If the total number of results is greater than this value, use the token returned in the response in the `nextToken` field when making another request to return the next batch of results.

Valid Range: Minimum value of 1. Maximum value of 1000.

nextToken

If the total number of results is greater than the `maxResults` value provided in the request, enter the token returned in the `nextToken` field in the response in this field to return the next batch of results.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "flowAliasSummaries": [
    {
      "arn": "string",
      "createdAt": "string",
      "description": "string",
      "flowId": "string",
      "id": "string",
      "name": "string",
      "routingConfiguration": [
        {
          "flowVersion": "string"
        }
      ],
      "updatedAt": "string"
    }
  ],
  "nextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

flowAliasSummaries

A list, each member of which contains information about a flow alias.

Type: Array of [FlowAliasSummary](#) objects

Array Members: Minimum number of 0 items. Maximum number of 10 items.

nextToken

If the total number of results is greater than the `maxResults` value provided in the request, use this token when making another request in the `nextToken` field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListFlows

Service: Agents for Amazon Bedrock

Returns a list of flows and information about each flow. For more information, see [Manage a flow in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Request Syntax

```
GET /flows/?maxResults=maxResults&nextToken=nextToken HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

maxResults

The maximum number of results to return in the response. If the total number of results is greater than this value, use the token returned in the response in the nextToken field when making another request to return the next batch of results.

Valid Range: Minimum value of 1. Maximum value of 1000.

nextToken

If the total number of results is greater than the maxResults value provided in the request, enter the token returned in the nextToken field in the response in this field to return the next batch of results.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
```

```
"flowSummaries": [  
  {  
    "arn": "string",  
    "createdAt": "string",  
    "description": "string",  
    "id": "string",  
    "name": "string",  
    "status": "string",  
    "updatedAt": "string",  
    "version": "string"  
  }  
],  
"nextToken": "string"  
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

flowSummaries

A list, each member of which contains information about a flow.

Type: Array of [FlowSummary](#) objects

Array Members: Minimum number of 0 items. Maximum number of 10 items.

nextToken

If the total number of results is greater than the `maxResults` value provided in the request, use this token when making another request in the `nextToken` field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListFlowVersions

Service: Agents for Amazon Bedrock

Returns a list of information about each flow. For more information, see [Deploy a flow in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Request Syntax

```
GET /flows/flowIdentifier/versions?maxResults=maxResults&nextToken=nextToken HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

[flowIdentifier](#)

The unique identifier of the flow.

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10})|([0-9a-zA-Z]{10})$`

Required: Yes

[maxResults](#)

The maximum number of results to return in the response. If the total number of results is greater than this value, use the token returned in the response in the `nextToken` field when making another request to return the next batch of results.

Valid Range: Minimum value of 1. Maximum value of 1000.

[nextToken](#)

If the total number of results is greater than the `maxResults` value provided in the request, enter the token returned in the `nextToken` field in the response in this field to return the next batch of results.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "flowVersionSummaries": [
    {
      "arn": "string",
      "createdAt": "string",
      "id": "string",
      "status": "string",
      "version": "string"
    }
  ],
  "nextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

flowVersionSummaries

A list, each member of which contains information about a flow.

Type: Array of [FlowVersionSummary](#) objects

Array Members: Minimum number of 0 items. Maximum number of 10 items.

nextToken

If the total number of results is greater than the `maxResults` value provided in the request, use this token when making another request in the `nextToken` field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)

- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListIngestionJobs

Service: Agents for Amazon Bedrock

Lists the ingestion jobs for a data source and information about each of them.

Request Syntax

```
POST /knowledgebases/knowledgeBaseId/datasources/dataSourceId/ingestionjobs/ HTTP/1.1
Content-type: application/json
```

```
{
  "filters": [
    {
      "attribute": "string",
      "operator": "string",
      "values": [ "string" ]
    }
  ],
  "maxResults": number,
  "nextToken": "string",
  "sortBy": {
    "attribute": "string",
    "order": "string"
  }
}
```

URI Request Parameters

The request uses the following URI parameters.

dataSourceId

The unique identifier of the data source for which to return ingestion jobs.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

knowledgeBaseId

The unique identifier of the knowledge base for which to return ingestion jobs.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

filters

Contains a definition of a filter for which to filter the results.

Type: Array of [IngestionJobFilter](#) objects

Array Members: Fixed number of 1 item.

Required: No

maxResults

The maximum number of results to return in the response. If the total number of results is greater than this value, use the token returned in the response in the nextToken field when making another request to return the next batch of results.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 1000.

Required: No

nextToken

If the total number of results is greater than the maxResults value provided in the request, enter the token returned in the nextToken field in the response in this field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Required: No

sortBy

Contains details about how to sort the results.

Type: [IngestionJobSortBy](#) object

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "ingestionJobSummaries": [
    {
      "dataSourceId": "string",
      "description": "string",
      "ingestionJobId": "string",
      "knowledgeBaseId": "string",
      "startedAt": "string",
      "statistics": {
        "numberOfDocumentsDeleted": number,
        "numberOfDocumentsFailed": number,
        "numberOfDocumentsScanned": number,
        "numberOfMetadataDocumentsModified": number,
        "numberOfMetadataDocumentsScanned": number,
        "numberOfModifiedDocumentsIndexed": number,
        "numberOfNewDocumentsIndexed": number
      },
      "status": "string",
      "updatedAt": "string"
    }
  ],
  "nextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

[ingestionJobSummaries](#)

A list of objects, each of which contains information about an ingestion job.

Type: Array of [IngestionJobSummary](#) objects

[nextToken](#)

If the total number of results is greater than the `maxResults` value provided in the request, use this token when making another request in the `nextToken` field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListKnowledgeBases

Service: Agents for Amazon Bedrock

Lists the knowledge bases in an account and information about each of them.

Request Syntax

```
POST /knowledgebases/ HTTP/1.1
Content-type: application/json
```

```
{
  "maxResults": number,
  "nextToken": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

maxResults

The maximum number of results to return in the response. If the total number of results is greater than this value, use the token returned in the response in the nextToken field when making another request to return the next batch of results.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 1000.

Required: No

nextToken

If the total number of results is greater than the maxResults value provided in the request, enter the token returned in the nextToken field in the response in this field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "knowledgeBaseSummaries": [
    {
      "description": "string",
      "knowledgeBaseId": "string",
      "name": "string",
      "status": "string",
      "updatedAt": "string"
    }
  ],
  "nextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

[knowledgeBaseSummaries](#)

A list of objects, each of which contains information about a knowledge base.

Type: Array of [KnowledgeBaseSummary](#) objects

[nextToken](#)

If the total number of results is greater than the `maxResults` value provided in the request, use this token when making another request in the `nextToken` field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)

- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListPrompts

Service: Agents for Amazon Bedrock

Returns a list of prompts from the Prompt management tool and information about each prompt. For more information, see [View information about prompts using Prompt management](#) in the Amazon Bedrock User Guide.

Request Syntax

```
GET /prompts/?  
maxResults=maxResults&nextToken=nextToken&promptIdentifier=promptIdentifier HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

maxResults

The maximum number of results to return in the response. If the total number of results is greater than this value, use the token returned in the response in the nextToken field when making another request to return the next batch of results.

Valid Range: Minimum value of 1. Maximum value of 1000.

nextToken

If the total number of results is greater than the maxResults value provided in the request, enter the token returned in the nextToken field in the response in this field to return the next batch of results.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

promptIdentifier

The unique identifier of the prompt.

Pattern: `^([\0-9a-zA-Z]{10})|(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:prompt/[\0-9a-zA-Z]{10})(?::[0-9]{1,5})?$`

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "nextToken": "string",
  "promptSummaries": [
    {
      "arn": "string",
      "createdAt": "string",
      "description": "string",
      "id": "string",
      "name": "string",
      "updatedAt": "string",
      "version": "string"
    }
  ]
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

nextToken

If the total number of results is greater than the `maxResults` value provided in the request, use this token when making another request in the `nextToken` field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

promptSummaries

A list, each member of which contains information about a prompt using Prompt management.

Type: Array of [PromptSummary](#) objects

Array Members: Minimum number of 0 items. Maximum number of 10 items.

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)

- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ListTagsForResource

Service: Agents for Amazon Bedrock

List all the tags for the resource you specify.

Request Syntax

```
GET /tags/resourceArn HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

resourceArn

The Amazon Resource Name (ARN) of the resource for which to list tags.

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: (^arn:aws:bedrock:[a-zA-Z0-9-]+:/d{12}:(agent|agent-alias|knowledge-base|flow|prompt)/[A-Z0-9]{10}(?:/[A-Z0-9]{10})?\$\$|^arn:aws:bedrock:[a-zA-Z0-9-]+:/d{12}:flow/([A-Z0-9]{10})/alias/([A-Z0-9]{10})\$\$|^arn:aws:bedrock:[a-zA-Z0-9-]+:/d{12}:prompt/([A-Z0-9]{10})?(?:/d+)?\$\$)

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "tags": {
    "string" : "string"
  }
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

tags

The key-value pairs for the tags associated with the resource.

Type: String to string map

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Key Pattern: `^[a-zA-Z0-9\s._:/=+@-]*$`

Value Length Constraints: Minimum length of 0. Maximum length of 256.

Value Pattern: `^[a-zA-Z0-9\s._:/=+@-]*$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of ListTagsForResource.

```
GET /tags/arn:aws:bedrock:us-west-2:123456789012:agent/ABCDEFGHIJ HTTP/1.1
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

PrepareAgent

Service: Agents for Amazon Bedrock

Creates a DRAFT version of the agent that can be used for internal testing.

Request Syntax

```
POST /agents/agentId/ HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

agentId

The unique identifier of the agent for which to create a DRAFT version.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 202
Content-type: application/json

{
  "agentId": "string",
  "agentStatus": "string",
  "agentVersion": "string",
  "preparedAt": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 202 response.

The following data is returned in JSON format by the service.

agentId

The unique identifier of the agent for which the DRAFT version was created.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

agentStatus

The status of the DRAFT version and whether it is ready for use.

Type: String

Valid Values: CREATING | PREPARING | PREPARED | NOT_PREPARED | DELETING | FAILED | VERSIONING | UPDATING

agentVersion

The version of the agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: `^(DRAFT|[0-9]{0,4}[1-9][0-9]{0,4})$`

preparedAt

The time at which the DRAFT version of the agent was last prepared.

Type: Timestamp

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of PrepareAgent.

```
POST /agents/ABCDEFGHIJ/ HTTP/1.1
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

PrepareFlow

Service: Agents for Amazon Bedrock

Prepares the DRAFT version of a flow so that it can be invoked. For more information, see [Test a flow in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Request Syntax

```
POST /flows/flowIdentifier/ HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

flowIdentifier

The unique identifier of the flow.

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10})|([0-9a-zA-Z]{10})$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 202
Content-type: application/json

{
  "id": "string",
  "status": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 202 response.

The following data is returned in JSON format by the service.

id

The unique identifier of the flow.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

status

The status of the flow. When you submit this request, the status will be `NotPrepared`. If preparation succeeds, the status becomes `Prepared`. If it fails, the status becomes `FAILED`.

Type: String

Valid Values: `Failed` | `Prepared` | `Preparing` | `NotPrepared`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

StartIngestionJob

Service: Agents for Amazon Bedrock

Begins an ingestion job, in which a data source is added to a knowledge base.

Request Syntax

```
PUT /knowledgebases/knowledgeBaseId/datasources/dataSourceId/ingestionjobs/ HTTP/1.1
Content-type: application/json
```

```
{
  "clientToken": "string",
  "description": "string"
}
```

URI Request Parameters

The request uses the following URI parameters.

[dataSourceId](#)

The unique identifier of the data source to ingest.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

[knowledgeBaseId](#)

The unique identifier of the knowledge base to which to add the data source.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

[clientToken](#)

A unique, case-sensitive identifier to ensure that the API request completes no more than one time. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#).

Type: String

Length Constraints: Minimum length of 33. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,256}$`

Required: No

description

A description of the ingestion job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

Response Syntax

```
HTTP/1.1 202
Content-type: application/json

{
  "ingestionJob": {
    "dataSourceId": "string",
    "description": "string",
    "failureReasons": [ "string" ],
    "ingestionJobId": "string",
    "knowledgeBaseId": "string",
    "startedAt": "string",
    "statistics": {
      "numberOfDocumentsDeleted": number,
      "numberOfDocumentsFailed": number,
      "numberOfDocumentsScanned": number,
      "numberOfMetadataDocumentsModified": number,
      "numberOfMetadataDocumentsScanned": number,
      "numberOfModifiedDocumentsIndexed": number,
      "numberOfNewDocumentsIndexed": number
    },
    "status": "string",
    "updatedAt": "string"
  }
}
```

```
}
```

Response Elements

If the action is successful, the service sends back an HTTP 202 response.

The following data is returned in JSON format by the service.

[ingestionJob](#)

An object containing information about the ingestion job.

Type: [IngestionJob](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

TagResource

Service: Agents for Amazon Bedrock

Associate tags with a resource. For more information, see [Tagging resources](#) in the Amazon Bedrock User Guide.

Request Syntax

```
POST /tags/resourceArn HTTP/1.1
Content-type: application/json
```

```
{
  "tags": {
    "string" : "string"
  }
}
```

URI Request Parameters

The request uses the following URI parameters.

[resourceArn](#)

The Amazon Resource Name (ARN) of the resource to tag.

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: (^arn:aws:bedrock:[a-zA-Z0-9-]+:/d{12}:(agent|agent-alias|knowledge-base|flow|prompt)/[A-Z0-9]{10}(?:/[A-Z0-9]{10})?|\$|^arn:aws:bedrock:[a-zA-Z0-9-]+:/d{12}:flow/([A-Z0-9]{10})/alias/([A-Z0-9]{10})\$|^arn:aws:bedrock:[a-zA-Z0-9-]+:/d{12}:prompt/([A-Z0-9]{10})?(?:/d+)?\$)

Required: Yes

Request Body

The request accepts the following data in JSON format.

[tags](#)

An object containing key-value pairs that define the tags to attach to the resource.

Type: String to string map

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Key Pattern: `^[a-zA-Z0-9\s._:/=+@-]*$`

Value Length Constraints: Minimum length of 0. Maximum length of 256.

Value Pattern: `^[a-zA-Z0-9\s._:/=+@-]*$`

Required: Yes

Response Syntax

```
HTTP/1.1 200
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of TagResource.

```
POST /tags/arn:aws:bedrock:us-west-2:123456789012:agent/ABCDEFGHIJ HTTP/1.1
Content-type: application/json

{
  "tags": {
    "cost-center" : "Tech"
  }
}
```

Example response

This example illustrates one usage of TagResource.

```
HTTP/1.1 200
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

UntagResource

Service: Agents for Amazon Bedrock

Remove tags from a resource.

Request Syntax

```
DELETE /tags/resourceArn?tagKeys=tagKeys HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

resourceArn

The Amazon Resource Name (ARN) of the resource from which to remove tags.

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: (^arn:aws:bedrock:[a-zA-Z0-9-]+:/d{12}:(agent|agent-alias|knowledge-base|flow|prompt)/[A-Z0-9]{10}(?:/[A-Z0-9]{10})?\$/|^arn:aws:bedrock:[a-zA-Z0-9-]+:/d{12}:flow/([A-Z0-9]{10})/alias/([A-Z0-9]{10})\$/|^arn:aws:bedrock:[a-zA-Z0-9-]+:/d{12}:prompt/([A-Z0-9]{10})?(?:/d+)?/)\$)

Required: Yes

tagKeys

A list of keys of the tags to remove from the resource.

Array Members: Minimum number of 0 items. Maximum number of 200 items.

Length Constraints: Minimum length of 1. Maximum length of 128.

Pattern: ^[a-zA-Z0-9\s._:/=+@-]*\$/

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of UntagResource.

```
DELETE /tags/arn:aws:bedrock:us-west-2:123456789012:agent/ABCDEFGHIJ HTTP/1.1
```

Example response

This example illustrates one usage of UntagResource.

```
HTTP/1.1 200
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

UpdateAgent

Service: Agents for Amazon Bedrock

Updates the configuration of an agent.

Request Syntax

```
PUT /agents/agentId/ HTTP/1.1
Content-type: application/json

{
  "agentName": "string",
  "agentResourceRoleArn": "string",
  "customerEncryptionKeyArn": "string",
  "description": "string",
  "foundationModel": "string",
  "guardrailConfiguration": {
    "guardrailIdentifier": "string",
    "guardrailVersion": "string"
  },
  "idleSessionTTLInSeconds": number,
  "instruction": "string",
  "memoryConfiguration": {
    "enabledMemoryTypes": [ "string" ],
    "storageDays": number
  },
  "promptOverrideConfiguration": {
    "overrideLambda": "string",
    "promptConfigurations": [
      {
        "basePromptTemplate": "string",
        "inferenceConfiguration": {
          "maxLength": number,
          "stopSequences": [ "string" ],
          "temperature": number,
          "topK": number,
          "topP": number
        },
      }
    ],
    "parserMode": "string",
    "promptCreationMode": "string",
    "promptState": "string",
    "promptType": "string"
  }
}
```



```
    ]  
  }  
}
```

URI Request Parameters

The request uses the following URI parameters.

[agentId](#)

The unique identifier of the agent.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

[agentName](#)

Specifies a new name for the agent.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?) {1,100}$`

Required: Yes

[agentResourceRoleArn](#)

The Amazon Resource Name (ARN) of the IAM role with permissions to invoke API operations on the agent.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^:]+)? :iam::([0-9]{12})?:role/.+ $`

Required: Yes

customerEncryptionKeyArn

The Amazon Resource Name (ARN) of the AWS KMS key with which to encrypt the agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

Required: No

description

Specifies a new description of the agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

foundationModel

Specifies a new foundation model to be used for orchestration by the agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(((0-9){12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}(([:][a-z0-9-]{1,63}){0,2})?/[a-z0-9-]{12})|(:foundation-model/([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2})))|((([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2}))|((0-9a-zA-Z)[_]?)+)$`

Required: Yes

guardrailConfiguration

The unique Guardrail configuration assigned to the agent when it is updated.

Type: [GuardrailConfiguration](#) object

Required: No

idleSessionTTLInSeconds

The number of seconds for which Amazon Bedrock keeps information about a user's conversation with the agent.

A user interaction remains active for the amount of time specified. If no conversation occurs during this time, the session expires and Amazon Bedrock deletes any data provided before the timeout.

Type: Integer

Valid Range: Minimum value of 60. Maximum value of 3600.

Required: No

instruction

Specifies new instructions that tell the agent what it should do and how it should interact with users.

Type: String

Length Constraints: Minimum length of 40. Maximum length of 4000.

Required: No

memoryConfiguration

Specifies the new memory configuration for the agent.

Type: [MemoryConfiguration](#) object

Required: No

promptOverrideConfiguration

Contains configurations to override prompts in different parts of an agent sequence. For more information, see [Advanced prompts](#).

Type: [PromptOverrideConfiguration](#) object

Required: No

Response Syntax

HTTP/1.1 202

Content-type: application/json

```
{
  "agent": {
    "agentArn": "string",
    "agentId": "string",
    "agentName": "string",
    "agentResourceRoleArn": "string",
    "agentStatus": "string",
    "agentVersion": "string",
    "clientToken": "string",
    "createdAt": "string",
    "customerEncryptionKeyArn": "string",
    "description": "string",
    "failureReasons": [ "string" ],
    "foundationModel": "string",
    "guardrailConfiguration": {
      "guardrailIdentifier": "string",
      "guardrailVersion": "string"
    },
    "idleSessionTTLInSeconds": number,
    "instruction": "string",
    "memoryConfiguration": {
      "enabledMemoryTypes": [ "string" ],
      "storageDays": number
    },
    "preparedAt": "string",
    "promptOverrideConfiguration": {
      "overrideLambda": "string",
      "promptConfigurations": [
        {
          "basePromptTemplate": "string",
          "inferenceConfiguration": {
            "maximumLength": number,
            "stopSequences": [ "string" ],
            "temperature": number,
            "topK": number,
            "topP": number
          }
        }
      ],
      "parserMode": "string",
      "promptCreationMode": "string",
```

```
        "promptState": "string",
        "promptType": "string"
    }
]
},
"recommendedActions": [ "string" ],
"updatedAt": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 202 response.

The following data is returned in JSON format by the service.

agent

Contains details about the agent that was updated.

Type: [Agent](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of UpdateAgent.

```
PUT /agents/ABCDEFGHIJ/ HTTP/1.1
Content-type: application/json

{
  "agentName": "TestName",
  "agentResourceRoleArn": "arn:aws:iam::123456789012:role/
AmazonBedrockExecutionRoleForAgents_user",
  "instruction": "You are an IT agent who solves customer's problems",
  "description": "Description is here",
  "idleSessionTTLInSeconds": 900,
  "foundationModel": "anthropic.claude-v2"
}
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

UpdateAgentActionGroup

Service: Agents for Amazon Bedrock

Updates the configuration for an action group for an agent.

Request Syntax

```
PUT /agents/agentId/agentversions/agentVersion/actiongroups/actionGroupId/ HTTP/1.1  
Content-type: application/json
```

```
{  
  "actionGroupExecutor": { ... },  
  "actionGroupName": "string",  
  "actionGroupState": "string",  
  "apiSchema": { ... },  
  "description": "string",  
  "functionSchema": { ... },  
  "parentActionGroupSignature": "string"  
}
```

URI Request Parameters

The request uses the following URI parameters.

actionGroupId

The unique identifier of the action group.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentId

The unique identifier of the agent for which to update the action group.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentVersion

The unique identifier of the agent version for which to update the action group.

Length Constraints: Fixed length of 5.

Pattern: ^DRAFT\$

Required: Yes

Request Body

The request accepts the following data in JSON format.

actionGroupExecutor

The Amazon Resource Name (ARN) of the Lambda function containing the business logic that is carried out upon invoking the action.

Type: [ActionGroupExecutor](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

actionGroupName

Specifies a new name for the action group.

Type: String

Pattern: ^([0-9a-zA-Z][_]?){1,100}\$

Required: Yes

actionGroupState

Specifies whether the action group is available for the agent to invoke or not when sending an [InvokeAgent](#) request.

Type: String

Valid Values: ENABLED | DISABLED

Required: No

apiSchema

Contains either details about the S3 object containing the OpenAPI schema for the action group or the JSON or YAML-formatted payload defining the schema. For more information, see [Action group OpenAPI schemas](#).

Type: [APISchema](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

[description](#)

Specifies a new name for the action group.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

[functionSchema](#)

Contains details about the function schema for the action group or the JSON or YAML-formatted payload defining the schema.

Type: [FunctionSchema](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

[parentActionGroupSignature](#)

To allow your agent to request the user for additional information when trying to complete a task, set this field to `AMAZON.UserInput`. You must leave the `description`, `apiSchema`, and `actionGroupExecutor` fields blank for this action group.

During orchestration, if your agent determines that it needs to invoke an API in an action group, but doesn't have enough information to complete the API request, it will invoke this action group instead and return an [Observation](#) reprompting the user for more information.

Type: String

Valid Values: `AMAZON.UserInput` | `AMAZON.CodeInterpreter`

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "agentActionGroup": {
    "actionGroupExecutor": { ... },
    "actionGroupId": "string",
    "actionGroupName": "string",
    "actionGroupState": "string",
    "agentId": "string",
    "agentVersion": "string",
    "apiSchema": { ... },
    "clientToken": "string",
    "createdAt": "string",
    "description": "string",
    "functionSchema": { ... },
    "parentActionSignature": "string",
    "updatedAt": "string"
  }
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

[agentActionGroup](#)

Contains details about the action group that was updated.

Type: [AgentActionGroup](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of UpdateAgentActionGroup.

```
PUT /agents/AGENT12345/agentversions/1/actiongroups/ACTION1234/ HTTP/1.1
```

```
Content-type: application/json
```

```
{
  "actionGroupName": "bedrock-temp-actions",
  "actionGroupState": "ENABLED",
  "description": "Testing = latest IT Management action",
  "apiSchema": {
    "s3": {
      "s3BucketName": "apischema-s3",
      "s3objectKey": "it_agent_openapi.json"
    }
  },
  "actionGroupExecutor": {
    "lambda": "arn:aws:lambda:us-west-2:123456789012:function:ItAgentLambda"
  }
}
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

UpdateAgentAlias

Service: Agents for Amazon Bedrock

Updates configurations for an alias of an agent.

Request Syntax

```
PUT /agents/agentId/agentaliases/agentAliasId/ HTTP/1.1
Content-type: application/json
```

```
{
  "agentAliasName": "string",
  "description": "string",
  "routingConfiguration": [
    {
      "agentVersion": "string",
      "provisionedThroughput": "string"
    }
  ]
}
```

URI Request Parameters

The request uses the following URI parameters.

agentAliasId

The unique identifier of the alias.

Length Constraints: Fixed length of 10.

Pattern: $^(\backslash\text{bTSTALIASID}\backslash\text{b} | [0-9a-zA-Z]^+)$$

Required: Yes

agentId

The unique identifier of the agent.

Pattern: $^[0-9a-zA-Z]{10}$$

Required: Yes

Request Body

The request accepts the following data in JSON format.

agentAliasName

Specifies a new name for the alias.

Type: String

Pattern: `^([\u0000-\u0009a-zA-Z][\u005C\u005C_-]?)\{1,100\}$`

Required: Yes

description

Specifies a new description for the alias.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

routingConfiguration

Contains details about the routing configuration of the alias.

Type: Array of [AgentAliasRoutingConfigurationListItem](#) objects

Array Members: Minimum number of 0 items. Maximum number of 1 item.

Required: No

Response Syntax

```
HTTP/1.1 202
Content-type: application/json

{
  "agentAlias": {
    "agentAliasArn": "string",
    "agentAliasHistoryEvents": [
      {
```

```

    "endDate": "string",
    "routingConfiguration": [
      {
        "agentVersion": "string",
        "provisionedThroughput": "string"
      }
    ],
    "startDate": "string"
  }
],
"agentAliasId": "string",
"agentAliasName": "string",
"agentAliasStatus": "string",
"agentId": "string",
"clientToken": "string",
"createdAt": "string",
"description": "string",
"failureReasons": [ "string" ],
"routingConfiguration": [
  {
    "agentVersion": "string",
    "provisionedThroughput": "string"
  }
],
"updatedAt": "string"
}
}

```

Response Elements

If the action is successful, the service sends back an HTTP 202 response.

The following data is returned in JSON format by the service.

agentAlias

Contains details about the alias that was updated.

Type: [AgentAlias](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example request

This example illustrates one usage of UpdateAgentAlias.

```
PUT /agents/ABCDEFGHIJ/agentaliases/ABCDEFGHIJ/ HTTP/1.1
Content-type: application/json
```

```
{
  "agentAliasName": "TestName",
  "description": "Updated description",
  "routingConfiguration": [
    {
      "agentVersion": "2"
    }
  ]
}
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

UpdateAgentKnowledgeBase

Service: Agents for Amazon Bedrock

Updates the configuration for a knowledge base that has been associated with an agent.

Request Syntax

```
PUT /agents/agentId/agentversions/agentVersion/knowledgebases/knowledgeBaseId/ HTTP/1.1  
Content-type: application/json
```

```
{  
  "description": "string",  
  "knowledgeBaseState": "string"  
}
```

URI Request Parameters

The request uses the following URI parameters.

agentId

The unique identifier of the agent associated with the knowledge base that you want to update.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentVersion

The version of the agent associated with the knowledge base that you want to update.

Length Constraints: Fixed length of 5.

Pattern: `^DRAFT$`

Required: Yes

knowledgeBaseId

The unique identifier of the knowledge base that has been associated with an agent.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

description

Specifies a new description for the knowledge base associated with an agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

knowledgeBaseState

Specifies whether the agent uses the knowledge base or not when sending an [InvokeAgent](#) request.

Type: String

Valid Values: ENABLED | DISABLED

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "agentKnowledgeBase": {
    "agentId": "string",
    "agentVersion": "string",
    "createdAt": "string",
    "description": "string",
    "knowledgeBaseId": "string",
    "knowledgeBaseState": "string",
    "updatedAt": "string"
  }
}
```

```
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

[agentKnowledgeBase](#)

Contains details about the knowledge base that has been associated with an agent.

Type: [AgentKnowledgeBase](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

UpdateDataSource

Service: Agents for Amazon Bedrock

Updates the configurations for a data source connector.

Important

You can't change the chunkingConfiguration after you create the data source connector. Specify the existing chunkingConfiguration.

Request Syntax

```
PUT /knowledgebases/knowledgeBaseId/datasources/dataSourceId HTTP/1.1
```

```
Content-type: application/json
```

```
{
  "dataDeletionPolicy": "string",
  "dataSourceConfiguration": {
    "confluenceConfiguration": {
      "crawlerConfiguration": {
        "filterConfiguration": {
          "patternObjectFilter": {
            "filters": [
              {
                "exclusionFilters": [ "string" ],
                "inclusionFilters": [ "string" ],
                "objectType": "string"
              }
            ]
          },
          "type": "string"
        }
      },
      "sourceConfiguration": {
        "authType": "string",
        "credentialsSecretArn": "string",
        "hostType": "string",
        "hostUrl": "string"
      }
    },
    "s3Configuration": {
```

```

    "bucketArn": "string",
    "bucketOwnerAccountId": "string",
    "inclusionPrefixes": [ "string" ]
  },
  "salesforceConfiguration": {
    "crawlerConfiguration": {
      "filterConfiguration": {
        "patternObjectFilter": {
          "filters": [
            {
              "exclusionFilters": [ "string" ],
              "inclusionFilters": [ "string" ],
              "objectType": "string"
            }
          ]
        },
        "type": "string"
      }
    },
    "sourceConfiguration": {
      "authType": "string",
      "credentialsSecretArn": "string",
      "hostUrl": "string"
    }
  },
  "sharePointConfiguration": {
    "crawlerConfiguration": {
      "filterConfiguration": {
        "patternObjectFilter": {
          "filters": [
            {
              "exclusionFilters": [ "string" ],
              "inclusionFilters": [ "string" ],
              "objectType": "string"
            }
          ]
        },
        "type": "string"
      }
    },
    "sourceConfiguration": {
      "authType": "string",
      "credentialsSecretArn": "string",
      "domain": "string",

```



```

        "hostType": "string",
        "siteUrls": [ "string" ],
        "tenantId": "string"
    }
},
"type": "string",
"webConfiguration": {
    "crawlerConfiguration": {
        "crawlerLimits": {
            "rateLimit": number
        },
        "exclusionFilters": [ "string" ],
        "inclusionFilters": [ "string" ],
        "scope": "string"
    },
    "sourceConfiguration": {
        "urlConfiguration": {
            "seedUrls": [
                {
                    "url": "string"
                }
            ]
        }
    }
}
},
"description": "string",
"name": "string",
"serverSideEncryptionConfiguration": {
    "kmsKeyArn": "string"
},
"vectorIngestionConfiguration": {
    "chunkingConfiguration": {
        "chunkingStrategy": "string",
        "fixedSizeChunkingConfiguration": {
            "maxTokens": number,
            "overlapPercentage": number
        },
        "hierarchicalChunkingConfiguration": {
            "levelConfigurations": [
                {
                    "maxTokens": number
                }
            ]
        }
    }
},

```

```

    "overlapTokens": number
  },
  "semanticChunkingConfiguration": {
    "breakpointPercentileThreshold": number,
    "bufferSize": number,
    "maxTokens": number
  }
},
"customTransformationConfiguration": {
  "intermediateStorage": {
    "s3Location": {
      "uri": "string"
    }
  },
  "transformations": [
    {
      "stepToApply": "string",
      "transformationFunction": {
        "transformationLambdaConfiguration": {
          "lambdaArn": "string"
        }
      }
    }
  ]
},
"parsingConfiguration": {
  "bedrockFoundationModelConfiguration": {
    "modelArn": "string",
    "parsingPrompt": {
      "parsingPromptText": "string"
    }
  },
  "parsingStrategy": "string"
}
}
}

```

URI Request Parameters

The request uses the following URI parameters.

dataSourceId

The unique identifier of the data source.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

knowledgeBaseId

The unique identifier of the knowledge base for the data source.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

dataDeletionPolicy

The data deletion policy for the data source that you want to update.

Type: String

Valid Values: RETAIN | DELETE

Required: No

dataSourceConfiguration

The connection configuration for the data source that you want to update.

Type: [DataSourceConfiguration](#) object

Required: Yes

description

Specifies a new description for the data source.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

name

Specifies a new name for the data source.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?)\{1,100\}$`

Required: Yes

serverSideEncryptionConfiguration

Contains details about server-side encryption of the data source.

Type: [ServerSideEncryptionConfiguration](#) object

Required: No

vectorIngestionConfiguration

Contains details about how to ingest the documents in the data source.

Type: [VectorIngestionConfiguration](#) object

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "dataSource": {
    "createdAt": "string",
    "dataDeletionPolicy": "string",
    "dataSourceConfiguration": {
      "confluenceConfiguration": {
        "crawlerConfiguration": {
          "filterConfiguration": {
            "patternObjectFilter": {
              "filters": [
                {
                  "exclusionFilters": [ "string" ],
                  "inclusionFilters": [ "string" ],
                  "objectType": "string"
                }
              ]
            }
          },
          "type": "string"
        }
      }
    }
  }
}
```

```

    }
  },
  "sourceConfiguration": {
    "authType": "string",
    "credentialsSecretArn": "string",
    "hostType": "string",
    "hostUrl": "string"
  }
},
"s3Configuration": {
  "bucketArn": "string",
  "bucketOwnerAccountId": "string",
  "inclusionPrefixes": [ "string" ]
},
"salesforceConfiguration": {
  "crawlerConfiguration": {
    "filterConfiguration": {
      "patternObjectFilter": {
        "filters": [
          {
            "exclusionFilters": [ "string" ],
            "inclusionFilters": [ "string" ],
            "objectType": "string"
          }
        ]
      },
      "type": "string"
    }
  },
  "sourceConfiguration": {
    "authType": "string",
    "credentialsSecretArn": "string",
    "hostUrl": "string"
  }
},
"sharePointConfiguration": {
  "crawlerConfiguration": {
    "filterConfiguration": {
      "patternObjectFilter": {
        "filters": [
          {
            "exclusionFilters": [ "string" ],
            "inclusionFilters": [ "string" ],
            "objectType": "string"
          }
        ]
      }
    }
  }
}

```

```

        }
      ]
    },
    "type": "string"
  }
},
"sourceConfiguration": {
  "authType": "string",
  "credentialsSecretArn": "string",
  "domain": "string",
  "hostType": "string",
  "siteUrls": [ "string" ],
  "tenantId": "string"
}
},
"type": "string",
"webConfiguration": {
  "crawlerConfiguration": {
    "crawlerLimits": {
      "rateLimit": number
    },
    "exclusionFilters": [ "string" ],
    "inclusionFilters": [ "string" ],
    "scope": "string"
  },
  "sourceConfiguration": {
    "urlConfiguration": {
      "seedUrls": [
        {
          "url": "string"
        }
      ]
    }
  }
}
},
"dataSourceId": "string",
"description": "string",
"failureReasons": [ "string" ],
"knowledgeBaseId": "string",
"name": "string",
"serverSideEncryptionConfiguration": {
  "kmsKeyArn": "string"
}
},

```

```
"status": "string",
"updatedAt": "string",
"vectorIngestionConfiguration": {
  "chunkingConfiguration": {
    "chunkingStrategy": "string",
    "fixedSizeChunkingConfiguration": {
      "maxTokens": number,
      "overlapPercentage": number
    },
    "hierarchicalChunkingConfiguration": {
      "levelConfigurations": [
        {
          "maxTokens": number
        }
      ],
      "overlapTokens": number
    },
    "semanticChunkingConfiguration": {
      "breakpointPercentileThreshold": number,
      "bufferSize": number,
      "maxTokens": number
    }
  },
  "customTransformationConfiguration": {
    "intermediateStorage": {
      "s3Location": {
        "uri": "string"
      }
    },
    "transformations": [
      {
        "stepToApply": "string",
        "transformationFunction": {
          "transformationLambdaConfiguration": {
            "lambdaArn": "string"
          }
        }
      }
    ]
  },
  "parsingConfiguration": {
    "bedrockFoundationModelConfiguration": {
      "modelArn": "string",
      "parsingPrompt": {
```

```
        "parsingPromptText": "string"
      },
    },
    "parsingStrategy": "string"
  }
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

dataSource

Contains details about the data source.

Type: [DataSource](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

UpdateFlow

Service: Agents for Amazon Bedrock

Modifies a flow. Include both fields that you want to keep and fields that you want to change. For more information, see [How it works](#) and [Create a flow in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Request Syntax

```
PUT /flows/flowIdentifier/ HTTP/1.1
Content-type: application/json

{
  "customerEncryptionKeyArn": "string",
  "definition": {
    "connections": [
      {
        "configuration": { ... },
        "name": "string",
        "source": "string",
        "target": "string",
        "type": "string"
      }
    ],
    "nodes": [
      {
        "configuration": { ... },
        "inputs": [
          {
            "expression": "string",
            "name": "string",
            "type": "string"
          }
        ],
        "name": "string",
        "outputs": [
          {
            "name": "string",
            "type": "string"
          }
        ],
        "type": "string"
      }
    ]
  }
}
```

```
    ]  
  },  
  "description": "string",  
  "executionRoleArn": "string",  
  "name": "string"  
}
```

URI Request Parameters

The request uses the following URI parameters.

flowIdentifier

The unique identifier of the flow.

Pattern: $^{\wedge}(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10})|([0-9a-zA-Z]{10})$$

Required: Yes

Request Body

The request accepts the following data in JSON format.

customerEncryptionKeyArn

The Amazon Resource Name (ARN) of the KMS key to encrypt the flow.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: $^{\wedge}arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$$

Required: No

definition

A definition of the nodes and the connections between the nodes in the flow.

Type: [FlowDefinition](#) object

Required: No

description

A description for the flow.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

executionRoleArn

The Amazon Resource Name (ARN) of the service role with permissions to create and manage a flow. For more information, see [Create a service role for flows in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:iam::([0-9]{12})?:role/(service-role/)?\.\+$`

Required: Yes

name

A name for the flow.

Type: String

Pattern: `^([0-9a-zA-Z][_-]?){1,100}$`

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "arn": "string",
  "createdAt": "string",
```

```

"customerEncryptionKeyArn": "string",
"definition": {
  "connections": [
    {
      "configuration": { ... },
      "name": "string",
      "source": "string",
      "target": "string",
      "type": "string"
    }
  ],
  "nodes": [
    {
      "configuration": { ... },
      "inputs": [
        {
          "expression": "string",
          "name": "string",
          "type": "string"
        }
      ],
      "name": "string",
      "outputs": [
        {
          "name": "string",
          "type": "string"
        }
      ],
      "type": "string"
    }
  ]
},
"description": "string",
"executionRoleArn": "string",
"id": "string",
"name": "string",
"status": "string",
"updatedAt": "string",
"version": "string"
}

```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

arn

The Amazon Resource Name (ARN) of the flow.

Type: String

Pattern: `^arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10}$`

createdAt

The time at which the flow was created.

Type: Timestamp

customerEncryptionKeyArn

The Amazon Resource Name (ARN) of the KMS key that the flow was encrypted with.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

definition

A definition of the nodes and the connections between nodes in the flow.

Type: [FlowDefinition](#) object

description

The description of the flow.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

executionRoleArn

The Amazon Resource Name (ARN) of the service role with permissions to create a flow. For more information, see [Create a service role for flows in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:iam::([0-9]{12})?:role/(service-role/)?\.\+$`

id

The unique identifier of the flow.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

name

The name of the flow.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?)\{1,100\}$`

status

The status of the flow. When you submit this request, the status will be `NotPrepared`. If updating fails, the status becomes `Failed`.

Type: String

Valid Values: `Failed` | `Prepared` | `Preparing` | `NotPrepared`

updatedAt

The time at which the flow was last updated.

Type: Timestamp

version

The version of the flow. When you update a flow, the version updated is the `DRAFT` version.

Type: String

Length Constraints: Fixed length of 5.

Pattern: `^DRAFT$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

UpdateFlowAlias

Service: Agents for Amazon Bedrock

Modifies the alias of a flow. Include both fields that you want to keep and ones that you want to change. For more information, see [Deploy a flow in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Request Syntax

```
PUT /flows/flowIdentifier/aliases/aliasIdentifier HTTP/1.1
```

```
Content-type: application/json
```

```
{
  "description": "string",
  "name": "string",
  "routingConfiguration": [
    {
      "flowVersion": "string"
    }
  ]
}
```

URI Request Parameters

The request uses the following URI parameters.

aliasIdentifier

The unique identifier of the alias.

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10}/alias/[0-9a-zA-Z]{10})|(TSTALIASID|[0-9a-zA-Z]{10})$`

Required: Yes

flowIdentifier

The unique identifier of the flow.

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10})|([0-9a-zA-Z]{10})$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

description

A description for the flow alias.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

name

The name of the flow alias.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

Required: Yes

routingConfiguration

Contains information about the version to which to map the alias.

Type: Array of [FlowAliasRoutingConfigurationListItem](#) objects

Array Members: Fixed number of 1 item.

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "arn": "string",
  "createdAt": "string",
  "description": "string",
  "flowId": "string",
```

```
"id": "string",  
"name": "string",  
"routingConfiguration": [  
  {  
    "flowVersion": "string"  
  }  
],  
"updatedAt": "string"  
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

arn

The Amazon Resource Name (ARN) of the flow.

Type: String

Pattern: `^arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10}/alias/(TSTALIASID|[0-9a-zA-Z]{10})$`

createdAt

The time at which the flow was created.

Type: Timestamp

description

The description of the flow.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

flowId

The unique identifier of the flow.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

id

The unique identifier of the alias.

Type: String

Pattern: `^(TSTALIASID|[0-9a-zA-Z]{10})$`

name

The name of the flow alias.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?){1,100}$`

routingConfiguration

Contains information about the version that the alias is mapped to.

Type: Array of [FlowAliasRoutingConfigurationListItem](#) objects

Array Members: Fixed number of 1 item.

updatedAt

The time at which the flow alias was last updated.

Type: Timestamp

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)

- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

UpdateKnowledgeBase

Service: Agents for Amazon Bedrock

Updates the configuration of a knowledge base with the fields that you specify. Because all fields will be overwritten, you must include the same values for fields that you want to keep the same.

You can change the following fields:

- name
- description
- roleArn

You can't change the `knowledgeBaseConfiguration` or `storageConfiguration` fields, so you must specify the same configurations as when you created the knowledge base. You can send a [GetKnowledgeBase](#) request and copy the same configurations.

Request Syntax

```
PUT /knowledgebases/knowledgeBaseId HTTP/1.1
Content-type: application/json

{
  "description": "string",
  "knowledgeBaseConfiguration": {
    "type": "string",
    "vectorKnowledgeBaseConfiguration": {
      "embeddingModelArn": "string",
      "embeddingModelConfiguration": {
        "bedrockEmbeddingModelConfiguration": {
          "dimensions": number
        }
      }
    }
  },
  "name": "string",
  "roleArn": "string",
  "storageConfiguration": {
    "mongoDbAtlasConfiguration": {
      "collectionName": "string",
      "credentialsSecretArn": "string",
      "databaseName": "string",
```



```
"endpoint": "string",
"endpointServiceName": "string",
"fieldMapping": {
  "metadataField": "string",
  "textField": "string",
  "vectorField": "string"
},
"vectorIndexName": "string"
},
"opensearchServerlessConfiguration": {
  "collectionArn": "string",
  "fieldMapping": {
    "metadataField": "string",
    "textField": "string",
    "vectorField": "string"
  },
  "vectorIndexName": "string"
},
"pineconeConfiguration": {
  "connectionString": "string",
  "credentialsSecretArn": "string",
  "fieldMapping": {
    "metadataField": "string",
    "textField": "string"
  },
  "namespace": "string"
},
"rdsConfiguration": {
  "credentialsSecretArn": "string",
  "databaseName": "string",
  "fieldMapping": {
    "metadataField": "string",
    "primaryKeyField": "string",
    "textField": "string",
    "vectorField": "string"
  },
  "resourceArn": "string",
  "tableName": "string"
},
"redisEnterpriseCloudConfiguration": {
  "credentialsSecretArn": "string",
  "endpoint": "string",
  "fieldMapping": {
    "metadataField": "string",
```

```
        "textField": "string",
        "vectorField": "string"
    },
    "vectorIndexName": "string"
},
"type": "string"
}
```

URI Request Parameters

The request uses the following URI parameters.

knowledgeBaselId

The unique identifier of the knowledge base to update.

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

description

Specifies a new description for the knowledge base.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

knowledgeBaseConfiguration

Specifies the configuration for the embeddings model used for the knowledge base. You must use the same configuration as when the knowledge base was created.

Type: [KnowledgeBaseConfiguration](#) object

Required: Yes

name

Specifies a new name for the knowledge base.

Type: String

Pattern: `^[0-9a-zA-Z][_]?{1,100}$`

Required: Yes

roleArn

Specifies a different Amazon Resource Name (ARN) of the IAM role with permissions to invoke API operations on the knowledge base.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^:]+)?:iam::([0-9]{12})?:role/.*$`

Required: Yes

storageConfiguration

Specifies the configuration for the vector store used for the knowledge base. You must use the same configuration as when the knowledge base was created.

Type: [StorageConfiguration](#) object

Required: Yes

Response Syntax

```
HTTP/1.1 202
Content-type: application/json

{
  "knowledgeBase": {
    "createdAt": "string",
    "description": "string",
    "failureReasons": [ "string" ],
    "knowledgeBaseArn": "string",
    "knowledgeBaseConfiguration": {
      "type": "string",
```

```

    "vectorKnowledgeBaseConfiguration": {
      "embeddingModelArn": "string",
      "embeddingModelConfiguration": {
        "bedrockEmbeddingModelConfiguration": {
          "dimensions": number
        }
      }
    },
    "knowledgeBaseId": "string",
    "name": "string",
    "roleArn": "string",
    "status": "string",
    "storageConfiguration": {
      "mongoDbAtlasConfiguration": {
        "collectionName": "string",
        "credentialsSecretArn": "string",
        "databaseName": "string",
        "endpoint": "string",
        "endpointServiceName": "string",
        "fieldMapping": {
          "metadataField": "string",
          "textField": "string",
          "vectorField": "string"
        },
        "vectorIndexName": "string"
      },
      "opensearchServerlessConfiguration": {
        "collectionArn": "string",
        "fieldMapping": {
          "metadataField": "string",
          "textField": "string",
          "vectorField": "string"
        },
        "vectorIndexName": "string"
      },
      "pineconeConfiguration": {
        "connectionString": "string",
        "credentialsSecretArn": "string",
        "fieldMapping": {
          "metadataField": "string",
          "textField": "string"
        },
        "namespace": "string"
      }
    }
  }
}

```

```

    },
    "rdsConfiguration": {
      "credentialsSecretArn": "string",
      "databaseName": "string",
      "fieldMapping": {
        "metadataField": "string",
        "primaryKeyField": "string",
        "textField": "string",
        "vectorField": "string"
      },
      "resourceArn": "string",
      "tableName": "string"
    },
    "redisEnterpriseCloudConfiguration": {
      "credentialsSecretArn": "string",
      "endpoint": "string",
      "fieldMapping": {
        "metadataField": "string",
        "textField": "string",
        "vectorField": "string"
      },
      "vectorIndexName": "string"
    },
    "type": "string"
  },
  "updatedAt": "string"
}
}

```

Response Elements

If the action is successful, the service sends back an HTTP 202 response.

The following data is returned in JSON format by the service.

knowledgeBase

Contains details about the knowledge base.

Type: [KnowledgeBase](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

UpdatePrompt

Service: Agents for Amazon Bedrock

Modifies a prompt in your prompt library. Include both fields that you want to keep and fields that you want to replace. For more information, see [Prompt management in Amazon Bedrock](#) and [Edit prompts in your prompt library](#) in the Amazon Bedrock User Guide.

Request Syntax

```
PUT /prompts/promptIdentifier/ HTTP/1.1
Content-type: application/json

{
  "customerEncryptionKeyArn": "string",
  "defaultVariant": "string",
  "description": "string",
  "name": "string",
  "variants": [
    {
      "inferenceConfiguration": { ... },
      "modelId": "string",
      "name": "string",
      "templateConfiguration": { ... },
      "templateType": "string"
    }
  ]
}
```

URI Request Parameters

The request uses the following URI parameters.

[promptIdentifier](#)

The unique identifier of the prompt.

Pattern: `^([0-9a-zA-Z]{10})|(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:prompt/[0-9a-zA-Z]{10})(?::[0-9]{1,5})?$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

customerEncryptionKeyArn

The Amazon Resource Name (ARN) of the KMS key to encrypt the prompt.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

Required: No

defaultVariant

The name of the default variant for the prompt. This value must match the name field in the relevant [PromptVariant](#) object.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

Required: No

description

A description for the prompt.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

name

A name for the prompt.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

Required: Yes

variants

A list of objects, each containing details about a variant of the prompt.

Type: Array of [PromptVariant](#) objects

Array Members: Minimum number of 0 items. Maximum number of 3 items.

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "arn": "string",
  "createdAt": "string",
  "customerEncryptionKeyArn": "string",
  "defaultVariant": "string",
  "description": "string",
  "id": "string",
  "name": "string",
  "updatedAt": "string",
  "variants": [
    {
      "inferenceConfiguration": { ... },
      "modelId": "string",
      "name": "string",
      "templateConfiguration": { ... },
      "templateType": "string"
    }
  ],
  "version": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

arn

The Amazon Resource Name (ARN) of the prompt.

Type: String

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:prompt/[0-9a-zA-Z]{10}(?::[0-9]{1,5})?)$`

createdAt

The time at which the prompt was created.

Type: Timestamp

customerEncryptionKeyArn

The Amazon Resource Name (ARN) of the KMS key to encrypt the prompt.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

defaultVariant

The name of the default variant for the prompt. This value must match the name field in the relevant [PromptVariant](#) object.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

description

The description of the prompt.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

id

The unique identifier of the prompt.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

name

The name of the prompt.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

updatedAt

The time at which the prompt was last updated.

Type: Timestamp

variants

A list of objects, each containing details about a variant of the prompt.

Type: Array of [PromptVariant](#) objects

Array Members: Minimum number of 0 items. Maximum number of 3 items.

version

The version of the prompt. When you update a prompt, the version updated is the DRAFT version.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: `^(DRAFT|[0-9]{0,4}[1-9][0-9]{0,4})$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

ConflictException

There was a conflict performing an operation.

HTTP Status Code: 409

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 402

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)

- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

Agents for Amazon Bedrock Runtime

The following actions are supported by Agents for Amazon Bedrock Runtime:

- [DeleteAgentMemory](#)
- [GetAgentMemory](#)
- [InvokeAgent](#)
- [InvokeFlow](#)
- [Retrieve](#)
- [RetrieveAndGenerate](#)

DeleteAgentMemory

Service: Agents for Amazon Bedrock Runtime

Deletes memory from the specified memory identifier.

Request Syntax

```
DELETE /agents/agentId/agentAliases/agentAliasId/memories?memoryId=memoryId HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

agentAliasId

The unique identifier of an alias of an agent.

Length Constraints: Minimum length of 0. Maximum length of 10.

Pattern: `^[0-9a-zA-Z]+$`

Required: Yes

agentId

The unique identifier of the agent to which the alias belongs.

Length Constraints: Minimum length of 0. Maximum length of 10.

Pattern: `^[0-9a-zA-Z]+$`

Required: Yes

memoryId

The unique identifier of the memory.

Length Constraints: Minimum length of 2. Maximum length of 100.

Pattern: `^[0-9a-zA-Z._:-]+$`

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 202
```

Response Elements

If the action is successful, the service sends back an HTTP 202 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions. Check your permissions and retry your request.

HTTP Status Code: 403

BadGatewayException

There was an issue with a dependency due to a server issue. Retry your request.

HTTP Status Code: 502

ConflictException

There was a conflict performing an operation. Resolve the conflict and retry your request.

HTTP Status Code: 409

DependencyFailedException

There was an issue with a dependency. Check the resource configurations and retry the request.

HTTP Status Code: 424

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 400

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

GetAgentMemory

Service: Agents for Amazon Bedrock Runtime

Gets the sessions stored in the memory of the agent.

Request Syntax

```
GET /agents/agentId/agentAliases/agentAliasId/memories?  
maxItems=maxItems&memoryId=memoryId&memoryType=memoryType&nextToken=nextToken HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

agentAliasId

The unique identifier of an alias of an agent.

Length Constraints: Minimum length of 0. Maximum length of 10.

Pattern: $^{\wedge}[\text{0-9a-zA-Z}]^{\wedge}$

Required: Yes

agentId

The unique identifier of the agent to which the alias belongs.

Length Constraints: Minimum length of 0. Maximum length of 10.

Pattern: $^{\wedge}[\text{0-9a-zA-Z}]^{\wedge}$

Required: Yes

maxItems

The maximum number of items to return in the response. If the total number of results is greater than this value, use the token returned in the response in the nextToken field when making another request to return the next batch of results.

Valid Range: Minimum value of 1. Maximum value of 1000.

memoryId

The unique identifier of the memory.

Length Constraints: Minimum length of 2. Maximum length of 100.

Pattern: `^[0-9a-zA-Z._: -]+$`

Required: Yes

memoryType

The type of memory.

Valid Values: SESSION_SUMMARY

Required: Yes

nextToken

If the total number of results is greater than the `maxItems` value provided in the request, enter the token returned in the `nextToken` field in the response in this field to return the next batch of results.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "memoryContents": [
    { ... }
  ],
  "nextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

memoryContents

Contains details of the sessions stored in the memory

Type: Array of [Memory](#) objects

nextToken

If the total number of results is greater than the `maxItems` value provided in the request, use this token when making another request in the `nextToken` field to return the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions. Check your permissions and retry your request.

HTTP Status Code: 403

BadGatewayException

There was an issue with a dependency due to a server issue. Retry your request.

HTTP Status Code: 502

ConflictException

There was a conflict performing an operation. Resolve the conflict and retry your request.

HTTP Status Code: 409

DependencyFailedException

There was an issue with a dependency. Check the resource configurations and retry the request.

HTTP Status Code: 424

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 400

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)

- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

InvokeAgent

Service: Agents for Amazon Bedrock Runtime

Note

The AWS CLI doesn't support streaming operations in Amazon Bedrock, including InvokeAgent.

Sends a prompt for the agent to process and respond to. Note the following fields for the request:

- To continue the same conversation with an agent, use the same `sessionId` value in the request.
- To activate trace enablement, turn `enableTrace` to `true`. Trace enablement helps you follow the agent's reasoning process that led it to the information it processed, the actions it took, and the final result it yielded. For more information, see [Trace enablement](#).
- End a conversation by setting `endSession` to `true`.
- In the `sessionState` object, you can include attributes for the session or prompt or, if you configured an action group to return control, results from invocation of the action group.

The response is returned in the `bytes` field of the chunk object.

- The `attribution` object contains citations for parts of the response.
- If you set `enableTrace` to `true` in the request, you can trace the agent's steps and reasoning process that led it to the response.
- If the action predicted was configured to return control, the response returns parameters for the action, elicited from the user, in the `returnControl` field.
- Errors are also surfaced in the response.

Request Syntax

```
POST /agents/agentId/agentAliases/agentAliasId/sessions/sessionId/text HTTP/1.1
```

```
Content-type: application/json
```

```
{  
  "enableTrace": boolean,  
  "endSession": boolean,  
  "inputText": "string",
```

```
"memoryId": "string",
"sessionState": {
  "files": [
    {
      "name": "string",
      "source": {
        "byteContent": {
          "data": blob,
          "mediaType": "string"
        },
        "s3Location": {
          "uri": "string"
        },
        "sourceType": "string"
      },
      "useCase": "string"
    }
  ],
  "invocationId": "string",
  "knowledgeBaseConfigurations": [
    {
      "knowledgeBaseId": "string",
      "retrievalConfiguration": {
        "vectorSearchConfiguration": {
          "filter": { ... },
          "numberOfResults": number,
          "overrideSearchType": "string"
        }
      }
    }
  ],
  "promptSessionAttributes": {
    "string" : "string"
  },
  "returnControlInvocationResults": [
    { ... }
  ],
  "sessionAttributes": {
    "string" : "string"
  }
}
```


URI Request Parameters

The request uses the following URI parameters.

[agentAliasId](#)

The alias of the agent to use.

Length Constraints: Minimum length of 0. Maximum length of 10.

Pattern: `^[0-9a-zA-Z]+$`

Required: Yes

[agentId](#)

The unique identifier of the agent to use.

Length Constraints: Minimum length of 0. Maximum length of 10.

Pattern: `^[0-9a-zA-Z]+$`

Required: Yes

[sessionId](#)

The unique identifier of the session. Use the same value across requests to continue the same conversation.

Length Constraints: Minimum length of 2. Maximum length of 100.

Pattern: `^[0-9a-zA-Z._:-]+$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

[enableTrace](#)

Specifies whether to turn on the trace or not to track the agent's reasoning process. For more information, see [Trace enablement](#).

Type: Boolean

Required: No

endSession

Specifies whether to end the session with the agent or not.

Type: Boolean

Required: No

inputText

The prompt text to send the agent.

Note

If you include `returnControlInvocationResults` in the `sessionState` field, the `inputText` field will be ignored.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 25000000.

Required: No

memoryId

The unique identifier of the agent memory.

Type: String

Length Constraints: Minimum length of 2. Maximum length of 100.

Pattern: `^[0-9a-zA-Z._: -]+$`

Required: No

sessionState

Contains parameters that specify various attributes of the session. For more information, see [Control session context](#).

Note

If you include `returnControlInvocationResults` in the `sessionState` field, the `inputText` field will be ignored.

Type: [SessionState](#) object

Required: No

Response Syntax

```
HTTP/1.1 200
x-amzn-bedrock-agent-content-type: contentType
x-amz-bedrock-agent-memory-id: memoryId
x-amz-bedrock-agent-session-id: sessionId
Content-type: application/json
```

```
{
  "accessDeniedException": {
  },
  "badGatewayException": {
  },
  "chunk": {
    "attribution": {
      "citations": [
        {
          "generatedResponsePart": {
            "textResponsePart": {
              "span": {
                "end": number,
                "start": number
              },
              "text": "string"
            }
          },
          "retrievedReferences": [
            {
              "content": {
                "text": "string"
              },
              "location": {
```

```

        "confluenceLocation": {
            "url": "string"
        },
        "s3Location": {
            "uri": "string"
        },
        "salesforceLocation": {
            "url": "string"
        },
        "sharePointLocation": {
            "url": "string"
        },
        "type": "string",
        "webLocation": {
            "url": "string"
        }
    },
    "metadata": {
        "string" : JSON value
    }
}
]
}
],
"bytes": blob
},
"conflictException": {
},
"dependencyFailedException": {
},
"files": {
    "files": [
        {
            "bytes": blob,
            "name": "string",
            "type": "string"
        }
    ]
},
"internalServerErrorException": {
},
"resourceNotFoundException": {
},

```

```

"returnControl": {
  "invocationId": "string",
  "invocationInputs": [
    { ... }
  ]
},
"serviceQuotaExceededException": {
},
"throttlingException": {
},
"trace": {
  "agentAliasId": "string",
  "agentId": "string",
  "agentVersion": "string",
  "sessionId": "string",
  "trace": { ... }
},
"validationException": {
}
}

```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The response returns the following HTTP headers.

contentType

The MIME type of the input data in the request. The default value is `application/json`.

memoryId

The unique identifier of the agent memory.

Length Constraints: Minimum length of 2. Maximum length of 100.

Pattern: `^[0-9a-zA-Z._:-]+$`

sessionId

The unique identifier of the session with the agent.

Length Constraints: Minimum length of 2. Maximum length of 100.

Pattern: `^[0-9a-zA-Z._: -]+$`

The following data is returned in JSON format by the service.

[accessDeniedException](#)

The request is denied because of missing access permissions. Check your permissions and retry your request.

Type: Exception

HTTP Status Code: 403

[badGatewayException](#)

There was an issue with a dependency due to a server issue. Retry your request.

Type: Exception

HTTP Status Code: 502

[chunk](#)

Contains a part of an agent response and citations for it.

Type: [PayloadPart](#) object

[conflictException](#)

There was a conflict performing an operation. Resolve the conflict and retry your request.

Type: Exception

HTTP Status Code: 409

[dependencyFailedException](#)

There was an issue with a dependency. Check the resource configurations and retry the request.

Type: Exception

HTTP Status Code: 424

[files](#)

Contains intermediate response for code interpreter if any files have been generated.

Type: [FilePart](#) object

[internalServerErrorException](#)

An internal server error occurred. Retry your request.

Type: Exception

HTTP Status Code: 500

[resourceNotFoundException](#)

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

Type: Exception

HTTP Status Code: 404

[returnControl](#)

Contains the parameters and information that the agent elicited from the customer to carry out an action. This information is returned to the system and can be used in your own setup for fulfilling the action.

Type: [ReturnControlPayload](#) object

[serviceQuotaExceededException](#)

The number of requests exceeds the service quota. Resubmit your request later.

Type: Exception

HTTP Status Code: 400

[throttlingException](#)

The number of requests exceeds the limit. Resubmit your request later.

Type: Exception

HTTP Status Code: 429

[trace](#)

Contains information about the agent and session, alongside the agent's reasoning process and results from calling actions and querying knowledge bases and metadata about the trace. You can use the trace to understand how the agent arrived at the response it provided the customer. For more information, see [Trace events](#).

Type: [TracePart](#) object

validationException

Input validation failed. Check your request parameters and retry the request.

Type: Exception

HTTP Status Code: 400

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions. Check your permissions and retry your request.

HTTP Status Code: 403

BadGatewayException

There was an issue with a dependency due to a server issue. Retry your request.

HTTP Status Code: 502

ConflictException

There was a conflict performing an operation. Resolve the conflict and retry your request.

HTTP Status Code: 409

DependencyFailedException

There was an issue with a dependency. Check the resource configurations and retry the request.

HTTP Status Code: 424

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 400

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Example simple request

The following example inquires the agent to get the weather for Seattle.

```
POST https://bedrock-agent-runtime.us-east-1.amazonaws.com/agents/AGENT12345/
agentAliases/TSTALIASID/sessions/abb/text

{
  "inputText": "give me the weather for seattle",
  "enableTrace": true
}
```

Example response (action group defined with OpenAPI schema, control returned)

The following example shows a response from an agent that has invoked an action group that was configured as follows:

- Defined with an OpenAPI schema
- Configured to return control to the agent developer

```
HTTP/1.1 200
```

```
x-amzn-bedrock-agent-content-type: application/json
x-amzn-bedrock-agent-session-id: session0
Content-type: application/json

{
  "invocationInputs": [{
    "apiInvocationInput": {
      "actionGroup": "WeatherAPIs",
      "apiPath": "/get-weather",
      "httpMethod": "get",
      "parameters": [
        {
          "name": "location",
          "type": "string",
          "value": "seattle"
        },
        {
          "name": "date",
          "type": "string",
          "value": "2024-09-15"
        }
      ]
    }
  ]},
  "invocationId": "337cb2f6-ec74-4b49-8141-00b8091498ad"
}
```

Example request using results from returned control (action group defined with OpenAPI schema)

The following example shows a request in which the results returned in the `InvokeAgent` response from an agent are passed to the `sessionState` of a new request. The results were returned from an agent that has invoked an action group that was configured as follows:

- Defined with an OpenAPI schema
- Configured to return control to the agent developer

The `invocationId` must match the `invocationId` that was returned in the response.

```
POST https://bedrock-agent-runtime.us-east-1.amazonaws.com/agents/AGENT12345/agentAliases/TSTALIASID/sessions/abb/text
```

```
{
  "enableTrace": true,
  "sessionState": {
    "invocationId": "337cb2f6-ec74-4b49-8141-00b8091498ad",
    "returnControlInvocationResults": [{
      "apiResult": {
        "actionGroup": "WeatherAPIs",
        "httpMethod": "get",
        "apiPath": "/get-weather",
        "responseBody": {
          "application/json": {
            "body": "It's rainy in Seattle today."
          }
        }
      }
    ]
  }
}
```

Example response (action group defined with function details, control returned)

The following example shows a response from an agent that has invoked an action group that was configured as follows:

- Defined with function details
- Configured to return control to the agent developer

```
HTTP/1.1 200
x-amzn-bedrock-agent-content-type: application/json
x-amz-bedrock-agent-session-id: session0
Content-type: application/json
```

```
{
  "invocationInputs": [{
    "functionInvocationInput": {
      "actionGroup": "WeatherAPIs",
      "function": "getWeather",
      "parameters": [
        {
          "name": "location",
          "type": "string",
```

```
        "value": "seattle"
      },
      {
        "name": "date",
        "type": "string",
        "value": "2024-09-15"
      }
    ]
  }
}],
"invocationId": "79e0feaa-c6f7-49bf-814d-b7c498505172"
}
```

Example request using results from returned control (action group defined with function details)

The following example shows a request in which the results returned in the `InvokeAgent` response from an agent are passed to the `sessionState` of a new request. The results were returned from an agent that has invoked an action group that was configured as follows:

- Defined with function details
- Configured to return control to the agent developer

The `invocationId` must match the `invocationId` that was returned in the response.

```
POST https://bedrock-agent-runtime.us-east-1.amazonaws.com/agents/AGENT12345/agentAliases/TSTALIASID/sessions/abb/text
```

```
{
  "enableTrace": true,
  "sessionState": {
    "invocationId": "79e0feaa-c6f7-49bf-814d-b7c498505172",
    "returnControlInvocationResults": [{
      "functionResult": {
        "actionGroup": "WeatherAPIs",
        "function": "getWeather",
        "responseBody": {
          "TEXT": {
            "body": "It's rainy in Seattle today."
          }
        }
      }
    ]
  }
}
```

```
    }  
  }  
}
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

InvokeFlow

Service: Agents for Amazon Bedrock Runtime

Invokes an alias of a flow to run the inputs that you specify and return the output of each node as a stream. If there's an error, the error is returned. For more information, see [Test a flow in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Request Syntax

```
POST /flows/flowIdentifier/aliases/flowAliasIdentifier HTTP/1.1
```

```
Content-type: application/json
```

```
{
  "inputs": [
    {
      "content": { ... },
      "nodeName": "string",
      "nodeOutputName": "string"
    }
  ]
}
```

URI Request Parameters

The request uses the following URI parameters.

[flowAliasIdentifier](#)

The unique identifier of the flow alias.

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10}/alias/[0-9a-zA-Z]{10})|(\bTSTALIASID\b|[0-9a-zA-Z]+)$`

Required: Yes

[flowIdentifier](#)

The unique identifier of the flow.

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10})|([0-9a-zA-Z]{10})$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

inputs

A list of objects, each containing information about an input into the flow.

Type: Array of [FlowInput](#) objects

Array Members: Fixed number of 1 item.

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "accessDeniedException": {
  },
  "badGatewayException": {
  },
  "conflictException": {
  },
  "dependencyFailedException": {
  },
  "flowCompletionEvent": {
    "completionReason": "string"
  },
  "flowOutputEvent": {
    "content": { ... },
    "nodeName": "string",
    "nodeType": "string"
  },
  "internalServerErrorException": {
```

```
  },  
  "resourceNotFoundException": {  
  },  
  "serviceQuotaExceededException": {  
  },  
  "throttlingException": {  
  },  
  "validationException": {  
  }  
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

[accessDeniedException](#)

The request is denied because of missing access permissions. Check your permissions and retry your request.

Type: Exception

HTTP Status Code: 403

[badGatewayException](#)

There was an issue with a dependency due to a server issue. Retry your request.

Type: Exception

HTTP Status Code: 502

[conflictException](#)

There was a conflict performing an operation. Resolve the conflict and retry your request.

Type: Exception

HTTP Status Code: 409

[dependencyFailedException](#)

There was an issue with a dependency. Check the resource configurations and retry the request.

Type: Exception

HTTP Status Code: 424

flowCompletionEvent

Contains information about why the flow completed.

Type: [FlowCompletionEvent](#) object

flowOutputEvent

Contains information about an output from flow invocation.

Type: [FlowOutputEvent](#) object

internalServerErrorException

An internal server error occurred. Retry your request.

Type: Exception

HTTP Status Code: 500

resourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

Type: Exception

HTTP Status Code: 404

serviceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

Type: Exception

HTTP Status Code: 400

throttlingException

The number of requests exceeds the limit. Resubmit your request later.

Type: Exception

HTTP Status Code: 429

validationException

Input validation failed. Check your request parameters and retry the request.

Type: Exception

HTTP Status Code: 400

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions. Check your permissions and retry your request.

HTTP Status Code: 403

BadGatewayException

There was an issue with a dependency due to a server issue. Retry your request.

HTTP Status Code: 502

ConflictException

There was a conflict performing an operation. Resolve the conflict and retry your request.

HTTP Status Code: 409

DependencyFailedException

There was an issue with a dependency. Check the resource configurations and retry the request.

HTTP Status Code: 424

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 400

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

Retrieve

Service: Agents for Amazon Bedrock Runtime

Queries a knowledge base and retrieves information from it.

Request Syntax

```
POST /knowledgebases/knowledgeBaseId/retrieve HTTP/1.1
Content-type: application/json
```

```
{
  "nextToken": "string",
  "retrievalConfiguration": {
    "vectorSearchConfiguration": {
      "filter": { ... },
      "numberOfResults": number,
      "overrideSearchType": "string"
    }
  },
  "retrievalQuery": {
    "text": "string"
  }
}
```

URI Request Parameters

The request uses the following URI parameters.

knowledgeBaseId

The unique identifier of the knowledge base to query.

Length Constraints: Minimum length of 0. Maximum length of 10.

Pattern: `^[0-9a-zA-Z]+$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

nextToken

If there are more results than can fit in the response, the response returns a `nextToken`. Use this token in the `nextToken` field of another request to retrieve the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^\S*$`

Required: No

retrievalConfiguration

Contains configurations for the knowledge base query and retrieval process. For more information, see [Query configurations](#).

Type: [KnowledgeBaseRetrievalConfiguration](#) object

Required: No

retrievalQuery

Contains the query to send the knowledge base.

Type: [KnowledgeBaseQuery](#) object

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "nextToken": "string",
  "retrievalResults": [
    {
      "content": {
        "text": "string"
      },
      "location": {
        "confluenceLocation": {
```

```

        "url": "string"
    },
    "s3Location": {
        "uri": "string"
    },
    "salesforceLocation": {
        "url": "string"
    },
    "sharePointLocation": {
        "url": "string"
    },
    "type": "string",
    "webLocation": {
        "url": "string"
    }
},
"metadata": {
    "string" : JSON value
},
"score": number
}
]
}

```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

nextToken

If there are more results than can fit in the response, the response returns a nextToken. Use this token in the nextToken field of another request to retrieve the next batch of results.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: $^{\backslash}S^*\$$

retrievalResults

A list of results from querying the knowledge base.

Type: Array of [KnowledgeBaseRetrievalResult](#) objects

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions. Check your permissions and retry your request.

HTTP Status Code: 403

BadGatewayException

There was an issue with a dependency due to a server issue. Retry your request.

HTTP Status Code: 502

ConflictException

There was a conflict performing an operation. Resolve the conflict and retry your request.

HTTP Status Code: 409

DependencyFailedException

There was an issue with a dependency. Check the resource configurations and retry the request.

HTTP Status Code: 424

InternalServerError

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 400

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Send a basic query

The following example queries a knowledge base.

```
POST /knowledgebases/KB12345678/retrieve HTTP/1.1
Content-type: application/json

{
  "retrievalQuery": {
    "text": "What is AWS?"
  }
}
```

Send a query and include filters

To include filters in a knowledge base query, at least one of the data source files must include a `.metadata.json` file. For example, if you had a data source of articles called `articles.pdf`, accompanied by a metadata file called `articles.metadata.json`, you could tag it for genre, year, and author. In the Retrieve request, you could apply the following filter to return all entertainment articles written after 2018, in addition to cooking or sports articles written by authors starting with C.


```
POST /knowledgebases/KB12345678/retrieve HTTP/1.1
```

```
Content-type: application/json
```

```
{
  "retrievalQuery": {
    "text": "What is AWS?",
  },
  "vectorSearchConfiguration": {
    "numberOfResults": 5,
    "filter": {
      "orAll": [
        {
          "andAll": [
            {
              "equals": {
                "key": "genre",
                "value": "entertainment"
              }
            },
            {
              "greaterThan": {
                "key": "year",
                "value": 2018
              }
            }
          ]
        },
        {
          "andAll": [
            {
              "in": {
                "key": "genre",
                "value": ["cooking", "sports"]
              }
            },
            {
              "startsWith": {
                "key": "author",
                "value": "C"
              }
            }
          ]
        }
      ]
    }
  }
}
```

```
    ]  
  }  
}
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

RetrieveAndGenerate

Service: Agents for Amazon Bedrock Runtime

Queries a knowledge base and generates responses based on the retrieved results. The response only cites sources that are relevant to the query.

Request Syntax

```
POST /retrieveAndGenerate HTTP/1.1
Content-type: application/json

{
  "input": {
    "text": "string"
  },
  "retrieveAndGenerateConfiguration": {
    "externalSourcesConfiguration": {
      "generationConfiguration": {
        "additionalModelRequestFields": {
          "string" : JSON value
        },
        "guardrailConfiguration": {
          "guardrailId": "string",
          "guardrailVersion": "string"
        },
        "inferenceConfig": {
          "textInferenceConfig": {
            "maxTokens": number,
            "stopSequences": [ "string" ],
            "temperature": number,
            "topP": number
          }
        },
        "promptTemplate": {
          "textPromptTemplate": "string"
        }
      },
      "modelArn": "string",
      "sources": [
        {
          "byteContent": {
            "contentType": "string",
            "data": blob,

```

```

        "identifier": "string"
    },
    "s3Location": {
        "uri": "string"
    },
    "sourceType": "string"
}
]
},
"knowledgeBaseConfiguration": {
    "generationConfiguration": {
        "additionalModelRequestFields": {
            "string" : JSON value
        },
        "guardrailConfiguration": {
            "guardrailId": "string",
            "guardrailVersion": "string"
        },
        "inferenceConfig": {
            "textInferenceConfig": {
                "maxTokens": number,
                "stopSequences": [ "string" ],
                "temperature": number,
                "topP": number
            }
        },
        "promptTemplate": {
            "textPromptTemplate": "string"
        }
    },
    "knowledgeBaseId": "string",
    "modelArn": "string",
    "orchestrationConfiguration": {
        "queryTransformationConfiguration": {
            "type": "string"
        }
    },
    "retrievalConfiguration": {
        "vectorSearchConfiguration": {
            "filter": { ... },
            "numberOfResults": number,
            "overrideSearchType": "string"
        }
    }
}

```

```
    },
    "type": "string"
  },
  "sessionConfiguration": {
    "kmsKeyArn": "string"
  },
  "sessionId": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

input

Contains the query to be made to the knowledge base.

Type: [RetrieveAndGenerateInput](#) object

Required: Yes

retrieveAndGenerateConfiguration

Contains configurations for the knowledge base query and retrieval process. For more information, see [Query configurations](#).

Type: [RetrieveAndGenerateConfiguration](#) object

Required: No

sessionConfiguration

Contains details about the session with the knowledge base.

Type: [RetrieveAndGenerateSessionConfiguration](#) object

Required: No

sessionId

The unique identifier of the session. When you first make a `RetrieveAndGenerate` request, Amazon Bedrock automatically generates this value. You must reuse this value for all

subsequent requests in the same conversational session. This value allows Amazon Bedrock to maintain context and knowledge from previous interactions. You can't explicitly set the `sessionId` yourself.

Type: String

Length Constraints: Minimum length of 2. Maximum length of 100.

Pattern: `^[0-9a-zA-Z._: -]+$`

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "citations": [
    {
      "generatedResponsePart": {
        "textResponsePart": {
          "span": {
            "end": number,
            "start": number
          },
          "text": "string"
        }
      },
      "retrievedReferences": [
        {
          "content": {
            "text": "string"
          },
          "location": {
            "confluenceLocation": {
              "url": "string"
            },
            "s3Location": {
              "uri": "string"
            },
            "salesforceLocation": {
              "url": "string"
            }
          }
        }
      ]
    }
  ]
}
```

```

    },
    "sharePointLocation": {
      "url": "string"
    },
    "type": "string",
    "webLocation": {
      "url": "string"
    }
  },
  "metadata": {
    "string" : JSON value
  }
}
]
}
],
"guardrailAction": "string",
"output": {
  "text": "string"
},
"sessionId": "string"
}

```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

[citations](#)

A list of segments of the generated response that are based on sources in the knowledge base, alongside information about the sources.

Type: Array of [Citation](#) objects

[guardrailAction](#)

Specifies if there is a guardrail intervention in the response.

Type: String

Valid Values: INTERVENED | NONE

output

Contains the response generated from querying the knowledge base.

Type: [RetrieveAndGenerateOutput](#) object

sessionId

The unique identifier of the session. When you first make a `RetrieveAndGenerate` request, Amazon Bedrock automatically generates this value. You must reuse this value for all subsequent requests in the same conversational session. This value allows Amazon Bedrock to maintain context and knowledge from previous interactions. You can't explicitly set the `sessionId` yourself.

Type: String

Length Constraints: Minimum length of 2. Maximum length of 100.

Pattern: `^[0-9a-zA-Z._: -]+$`

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions. Check your permissions and retry your request.

HTTP Status Code: 403

BadGatewayException

There was an issue with a dependency due to a server issue. Retry your request.

HTTP Status Code: 502

ConflictException

There was a conflict performing an operation. Resolve the conflict and retry your request.

HTTP Status Code: 409

DependencyFailedException

There was an issue with a dependency. Check the resource configurations and retry the request.

HTTP Status Code: 424

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

HTTP Status Code: 400

ThrottlingException

The number of requests exceeds the limit. Resubmit your request later.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Send a basic query

The following example uses the minimally required fields to generate a response after querying a knowledge base.

```
POST /retrieveAndGenerate HTTP/1.1
Content-type: application/json

{
  "input": {
```

```
    "text": "What is AWS?"
  },
  "retrieveAndGenerateConfiguration": {
    "knowledgeBaseConfiguration": {
      "knowledgeBaseId": "KB12345678",
      "modelArn": "anthropic.claude-v2:1"
    },
    "type": "KNOWLEDGE_BASE"
  }
}
```

Send a query and include filters

To include filters in a knowledge base query, at least one of the data source files must include a `.metadata.json` file. For example, if you had a data source of articles called `articles.pdf`, accompanied by a metadata file called `articles.metadata.json`, you could tag it for genre, year, and author. In the Retrieve request, you could apply the following filter to return all entertainment articles written after 2018, in addition to cooking or sports articles written by authors starting with C.

```
POST /retrieveAndGenerate HTTP/1.1
Content-type: application/json

{
  "input": {
    "text": "What is AWS?",
  },
  "retrieveAndGenerateConfiguration": {
    "knowledgeBaseConfiguration": {
      "knowledgeBaseId": "KB12345678",
      "modelArn": "anthropic.claude-v2:1",
      "retrievalConfiguration": {
        "vectorSearchConfiguration": {
          "numberOfResults": 5,
          "filter": {
            "orAll": [
              {
                "andAll": [
                  {
                    "equals": {
                      "key": "genre",
                      "value": "entertainment"
                    }
                  }
                ]
              }
            ]
          }
        }
      }
    }
  }
}
```

```
    }
  },
  {
    "greaterThan": {
      "key": "year",
      "value": 2018
    }
  }
],
},
{
  "andAll": [
    {
      "in": {
        "key": "genre",
        "value": ["cooking", "sports"]
      }
    },
    {
      "startsWith": {
        "key": "author",
        "value": "C"
      }
    }
  ]
}
}
}
},
"type": "KNOWLEDGE_BASE"
}
}
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)

- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

Amazon Bedrock Runtime

The following actions are supported by Amazon Bedrock Runtime:

- [ApplyGuardrail](#)
- [Converse](#)
- [ConverseStream](#)
- [InvokeModel](#)
- [InvokeModelWithResponseStream](#)

ApplyGuardrail

Service: Amazon Bedrock Runtime

The action to apply a guardrail.

Request Syntax

```
POST /guardrail/guardrailIdentifier/version/guardrailVersion/apply HTTP/1.1
Content-type: application/json

{
  "content": [
    { ... }
  ],
  "source": "string"
}
```

URI Request Parameters

The request uses the following URI parameters.

guardrailIdentifier

The guardrail identifier used in the request to apply the guardrail.

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^(([a-z0-9]+)|(arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:guardrail/[a-z0-9]+))$`

Required: Yes

guardrailVersion

The guardrail version used in the request to apply the guardrail.

Pattern: `^(([1-9][0-9]{0,7})|(DRAFT))$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

content

The content details used in the request to apply the guardrail.

Type: Array of [GuardrailContentBlock](#) objects

Required: Yes

source

The source of data used in the request to apply the guardrail.

Type: String

Valid Values: INPUT | OUTPUT

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "action": "string",
  "assessments": [
    {
      "contentPolicy": {
        "filters": [
          {
            "action": "string",
            "confidence": "string",
            "type": "string"
          }
        ]
      },
      "contextualGroundingPolicy": {
        "filters": [
          {
            "action": "string",
            "score": number,
            "threshold": number,
            "type": "string"
          }
        ]
      }
    }
  ]
}
```

```
    ]
  },
  "sensitiveInformationPolicy": {
    "piiEntities": [
      {
        "action": "string",
        "match": "string",
        "type": "string"
      }
    ],
    "regexes": [
      {
        "action": "string",
        "match": "string",
        "name": "string",
        "regex": "string"
      }
    ]
  },
  "topicPolicy": {
    "topics": [
      {
        "action": "string",
        "name": "string",
        "type": "string"
      }
    ]
  },
  "wordPolicy": {
    "customWords": [
      {
        "action": "string",
        "match": "string"
      }
    ],
    "managedWordLists": [
      {
        "action": "string",
        "match": "string",
        "type": "string"
      }
    ]
  }
}
```

```
],
  "outputs": [
    {
      "text": "string"
    }
  ],
  "usage": {
    "contentPolicyUnits": number,
    "contextualGroundingPolicyUnits": number,
    "sensitiveInformationPolicyFreeUnits": number,
    "sensitiveInformationPolicyUnits": number,
    "topicPolicyUnits": number,
    "wordPolicyUnits": number
  }
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

action

The action taken in the response from the guardrail.

Type: String

Valid Values: NONE | GUARDRAIL_INTERVENED

assessments

The assessment details in the response from the guardrail.

Type: Array of [GuardrailAssessment](#) objects

outputs

The output details in the response from the guardrail.

Type: Array of [GuardrailOutputContent](#) objects

usage

The usage details in the response from the guardrail.

Type: [GuardrailUsage](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ResourceNotFoundException

The specified resource ARN was not found. Check the ARN and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

Your request exceeds the service quota for your account. You can view your quotas at [Viewing service quotas](#). You can resubmit your request later.

HTTP Status Code: 400

ThrottlingException

Your request was throttled because of service-wide limitations. Resubmit your request later or in a different region. You can also purchase [Provisioned Throughput](#) to increase the rate or number of tokens you can process.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

Converse

Service: Amazon Bedrock Runtime

Sends messages to the specified Amazon Bedrock model. `Converse` provides a consistent interface that works with all models that support messages. This allows you to write code once and use it with different models. If a model has unique inference parameters, you can also pass those unique parameters to the model.

Amazon Bedrock doesn't store any text, images, or documents that you provide as content. The data is only used to generate the response.

For information about the `Converse` API, see [Use the Converse API](#). To use a guardrail, see [Use a guardrail with the Converse API](#). To use a tool with a model, see [Tool use \(Function calling\)](#).

For example code, see [Converse API examples](#).

This operation requires permission for the `bedrock:InvokeModel` action.

Request Syntax

```
POST /model/modelId/converse HTTP/1.1
Content-type: application/json

{
  "additionalModelRequestFields": JSON value,
  "additionalModelResponseFieldPaths": [ "string" ],
  "guardrailConfig": {
    "guardrailIdentifier": "string",
    "guardrailVersion": "string",
    "trace": "string"
  },
  "inferenceConfig": {
    "maxTokens": number,
    "stopSequences": [ "string" ],
    "temperature": number,
    "topP": number
  },
  "messages": [
    {
      "content": [
        { ... }
      ]
    }
  ],
}
```

```

    "role": "string"
  }
],
"system": [
  { ... }
],
"toolConfig": {
  "toolChoice": { ... },
  "tools": [
    { ... }
  ]
}
}

```

URI Request Parameters

The request uses the following URI parameters.

modelId

The identifier for the model that you want to call.

The modelId to provide depends on the type of model that you use:

- If you use a base model, specify the model ID or its ARN. For a list of model IDs for base models, see [Amazon Bedrock base model IDs \(on-demand throughput\)](#) in the Amazon Bedrock User Guide.
- If you use a provisioned model, specify the ARN of the Provisioned Throughput. For more information, see [Run inference using a Provisioned Throughput](#) in the Amazon Bedrock User Guide.
- If you use a custom model, first purchase Provisioned Throughput for it. Then specify the ARN of the resulting provisioned model. For more information, see [Use a custom model in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^(arn:aws(-[^\:]+)?):bedrock:[a-z0-9-]{1,20}:((([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}/[a-z0-9]{12})|(:foundation-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.:]?[a-z0-9-]{1,63}))|([0-9]{12}:provisioned-model/[a-z0-9]{12})))|([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.:]?[a-z0-9-]{1,63}))|(([0-9a-zA-Z][_]?)+)$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

additionalModelRequestFields

Additional inference parameters that the model supports, beyond the base set of inference parameters that Converse supports in the `inferenceConfig` field. For more information, see [Model parameters](#).

Type: JSON value

Required: No

additionalModelResponseFieldPaths

Additional model parameters field paths to return in the response. Converse returns the requested fields as a JSON Pointer object in the `additionalModelResponseFields` field. The following is example JSON for `additionalModelResponseFieldPaths`.

```
[ "/stop_sequence" ]
```

For information about the JSON Pointer syntax, see the [Internet Engineering Task Force \(IETF\)](#) documentation.

Converse rejects an empty JSON Pointer or incorrectly structured JSON Pointer with a 400 error code. If the JSON Pointer is valid, but the requested field is not in the model response, it is ignored by Converse.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 10 items.

Length Constraints: Minimum length of 1. Maximum length of 256.

Required: No

guardrailConfig

Configuration information for a guardrail that you want to use in the request.

Type: [GuardrailConfiguration](#) object

Required: No

[inferenceConfig](#)

Inference parameters to pass to the model. Converse supports a base set of inference parameters. If you need to pass additional parameters that the model supports, use the `additionalModelRequestFields` request field.

Type: [InferenceConfiguration](#) object

Required: No

[messages](#)

The messages that you want to send to the model.

Type: Array of [Message](#) objects

Required: Yes

[system](#)

A system prompt to pass to the model.

Type: Array of [SystemContentBlock](#) objects

Required: No

[toolConfig](#)

Configuration information for the tools that the model can use when generating a response.

Note

This field is only supported by Anthropic Claude 3, Cohere Command R, Cohere Command R+, and Mistral Large models.

Type: [ToolConfiguration](#) object

Required: No

Response Syntax

HTTP/1.1 200

Content-type: application/json

```
{
  "additionalModelResponseFields": JSON value,
  "metrics": {
    "latencyMs": number
  },
  "output": { ... },
  "stopReason": "string",
  "trace": {
    "guardrail": {
      "inputAssessment": {
        "string" : {
          "contentPolicy": {
            "filters": [
              {
                "action": "string",
                "confidence": "string",
                "type": "string"
              }
            ]
          },
          "contextualGroundingPolicy": {
            "filters": [
              {
                "action": "string",
                "score": number,
                "threshold": number,
                "type": "string"
              }
            ]
          },
          "sensitiveInformationPolicy": {
            "piiEntities": [
              {
                "action": "string",
                "match": "string",
                "type": "string"
              }
            ]
          },
          "regexes": [
```

```
        {
            "action": "string",
            "match": "string",
            "name": "string",
            "regex": "string"
        }
    ]
},
"topicPolicy": {
    "topics": [
        {
            "action": "string",
            "name": "string",
            "type": "string"
        }
    ]
},
"wordPolicy": {
    "customWords": [
        {
            "action": "string",
            "match": "string"
        }
    ],
    "managedWordLists": [
        {
            "action": "string",
            "match": "string",
            "type": "string"
        }
    ]
}
},
"modelOutput": [ "string" ],
"outputAssessments": {
    "string" : [
        {
            "contentPolicy": {
                "filters": [
                    {
                        "action": "string",
                        "confidence": "string",
                        "type": "string"
                    }
                ]
            }
        }
    ]
}
```



```
    }
  ]
},
"contextualGroundingPolicy": {
  "filters": [
    {
      "action": "string",
      "score": number,
      "threshold": number,
      "type": "string"
    }
  ]
},
"sensitiveInformationPolicy": {
  "piiEntities": [
    {
      "action": "string",
      "match": "string",
      "type": "string"
    }
  ],
  "regexes": [
    {
      "action": "string",
      "match": "string",
      "name": "string",
      "regex": "string"
    }
  ]
},
"topicPolicy": {
  "topics": [
    {
      "action": "string",
      "name": "string",
      "type": "string"
    }
  ]
},
"wordPolicy": {
  "customWords": [
    {
      "action": "string",
      "match": "string"
    }
  ]
}
```


Note: This object is a Union. Only one member of this object can be specified or returned.

stopReason

The reason why the model stopped generating output.

Type: String

Valid Values: end_turn | tool_use | max_tokens | stop_sequence | guardrail_intervened | content_filtered

trace

A trace object that contains information about the Guardrail behavior.

Type: [ConverseTrace](#) object

usage

The total number of tokens used in the call to `Converse`. The total includes the tokens input to the model and the tokens generated by the model.

Type: [TokenUsage](#) object

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ModelErrorException

The request failed due to an error while processing the model.

HTTP Status Code: 424

ModelNotReadyException

The model specified in the request is not ready to serve inference requests.

HTTP Status Code: 429

ModelTimeoutException

The request took too long to process. Processing time exceeded the model timeout length.

HTTP Status Code: 408

ResourceNotFoundException

The specified resource ARN was not found. Check the ARN and try your request again.

HTTP Status Code: 404

ThrottlingException

Your request was throttled because of service-wide limitations. Resubmit your request later or in a different region. You can also purchase [Provisioned Throughput](#) to increase the rate or number of tokens you can process.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Send a message to a model

Send a message to Anthropic Claude Sonnet with Converse.

```
POST /model/anthropic.claude-3-sonnet-20240229-v1:0/converse HTTP/1.1
Content-type: application/json

{
  "messages": [
    {
      "role": "user",
```

```
        "content": [
            {
                "text": "Write an article about impact of high inflation to GDP of
a country"
            }
        ]
    },
    "system": [{"text" : "You are an economist with access to lots of data"}]
    "inferenceConfig": {
        "maxTokens": 1000,
        "temperature": 0.5
    }
}
```

Example response

Response for the above request.

```
HTTP/1.1 200
Content-type: application/json

{
    "output": {
        "message": {
            "content": [
                {
                    "text": "<text generated by the model>"
                }
            ],
            "role": "assistant"
        }
    },
    "stopReason": "end_turn",
    "usage": {
        "inputTokens": 30,
        "outputTokens": 628,
        "totalTokens": 658
    },
    "metrics": {
        "latencyMs": 1275
    }
}
```

Send a message with additional model fields

In the following example, the request passes a field (`top_k`) that the `Converse` field doesn't support. You pass the additional field in the `additionalModelRequestFields` field. The example also shows how to set the paths for the additional fields sent in the response from the model.

```
POST /model/anthropic.claude-3-sonnet-20240229-v1:0/converse HTTP/1.1
Content-type: application/json

{
  "messages": [
    {
      "role": "user",
      "content": [
        {
          "text": "Provide general steps to debug a BSOD on a Windows
laptop."
        }
      ]
    }
  ],
  "system": [{"text" : "You are a tech support expert who helps resolve technical
issues. Signal 'SUCCESS' if you can resolve the issue, otherwise 'FAILURE'"}],
  "inferenceConfig": {
    "stopSequences": [ "SUCCESS", "FAILURE" ]
  },
  "additionalModelRequestFields": {
    "top_k": 200
  },
  "additionalModelResponseFieldPaths": [
    "/stop_sequence"
  ]
}
```

Example response

Response for the above example.

```
HTTP/1.1 200
Content-type: application/json

{
```

```
"output": {
  "message": {
    "content": [
      {
        "text": "<text generated by the model>"
      }
    ],
    "role": "assistant"
  }
},
"additionalModelResponseFields": {
  "stop_sequence": "SUCCESS"
},
"stopReason": "stop_sequence",
"usage": {
  "inputTokens": 51,
  "outputTokens": 442,
  "totalTokens": 493
},
"metrics": {
  "latencyMs": 7944
}
}
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

ConverseStream

Service: Amazon Bedrock Runtime

Sends messages to the specified Amazon Bedrock model and returns the response in a stream. `ConverseStream` provides a consistent API that works with all Amazon Bedrock models that support messages. This allows you to write code once and use it with different models. Should a model have unique inference parameters, you can also pass those unique parameters to the model.

To find out if a model supports streaming, call [GetFoundationModel](#) and check the `responseStreamingSupported` field in the response.

Note

The AWS CLI doesn't support streaming operations in Amazon Bedrock, including `ConverseStream`.

Amazon Bedrock doesn't store any text, images, or documents that you provide as content. The data is only used to generate the response.

For information about the Converse API, see [Use the Converse API](#). To use a guardrail, see [Use a guardrail with the Converse API](#). To use a tool with a model, see [Tool use \(Function calling\)](#).

For example code, see [Conversation streaming example](#).

This operation requires permission for the `bedrock:InvokeModelWithResponseStream` action.

Request Syntax

```
POST /model/modelId/converse-stream HTTP/1.1
Content-type: application/json

{
  "additionalModelRequestFields": JSON value,
  "additionalModelResponseFieldPaths": [ "string" ],
  "guardrailConfig": {
    "guardrailIdentifier": "string",
    "guardrailVersion": "string",
    "streamProcessingMode": "string",
    "trace": "string"
  }
}
```



```
},
  "inferenceConfig": {
    "maxTokens": number,
    "stopSequences": [ "string" ],
    "temperature": number,
    "topP": number
  },
  "messages": [
    {
      "content": [
        { ... }
      ],
      "role": "string"
    }
  ],
  "system": [
    { ... }
  ],
  "toolConfig": {
    "toolChoice": { ... },
    "tools": [
      { ... }
    ]
  }
}
```

URI Request Parameters

The request uses the following URI parameters.

modelId

The ID for the model.

The modelId to provide depends on the type of model that you use:

- If you use a base model, specify the model ID or its ARN. For a list of model IDs for base models, see [Amazon Bedrock base model IDs \(on-demand throughput\)](#) in the Amazon Bedrock User Guide.
- If you use a provisioned model, specify the ARN of the Provisioned Throughput. For more information, see [Run inference using a Provisioned Throughput](#) in the Amazon Bedrock User Guide.

- If you use a custom model, first purchase Provisioned Throughput for it. Then specify the ARN of the resulting provisioned model. For more information, see [Use a custom model in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^(arn:aws(-[^\:]+)?):bedrock:[a-z0-9-]{1,20}:((([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}/[a-z0-9]{12})|(:foundation-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.:]?[a-z0-9-]{1,63}))|([0-9]{12}:provisioned-model/[a-z0-9]{12})))|([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.:]?[a-z0-9-]{1,63}))|((([0-9a-zA-Z][_]?)+))$`

Required: Yes

Request Body

The request accepts the following data in JSON format.

[additionalModelRequestFields](#)

Additional inference parameters that the model supports, beyond the base set of inference parameters that `ConverseStream` supports in the `inferenceConfig` field.

Type: JSON value

Required: No

[additionalModelResponseFieldPaths](#)

Additional model parameters field paths to return in the response. `ConverseStream` returns the requested fields as a JSON Pointer object in the `additionalModelResponseFields` field. The following is example JSON for `additionalModelResponseFieldPaths`.

```
[ "/stop_sequence" ]
```

For information about the JSON Pointer syntax, see the [Internet Engineering Task Force \(IETF\)](#) documentation.

`ConverseStream` rejects an empty JSON Pointer or incorrectly structured JSON Pointer with a `400` error code. If the JSON Pointer is valid, but the requested field is not in the model response, it is ignored by `ConverseStream`.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 10 items.

Length Constraints: Minimum length of 1. Maximum length of 256.

Required: No

guardrailConfig

Configuration information for a guardrail that you want to use in the request.

Type: [GuardrailStreamConfiguration](#) object

Required: No

inferenceConfig

Inference parameters to pass to the model. `ConverseStream` supports a base set of inference parameters. If you need to pass additional parameters that the model supports, use the `additionalModelRequestFields` request field.

Type: [InferenceConfiguration](#) object

Required: No

messages

The messages that you want to send to the model.

Type: Array of [Message](#) objects

Required: Yes

system

A system prompt to send to the model.

Type: Array of [SystemContentBlock](#) objects

Required: No

toolConfig

Configuration information for the tools that the model can use when generating a response.

Note

This field is only supported by Anthropic Claude 3 models.

Type: [ToolConfiguration](#) object

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
  "contentBlockDelta": {
    "contentBlockIndex": number,
    "delta": { ... }
  },
  "contentBlockStart": {
    "contentBlockIndex": number,
    "start": { ... }
  },
  "contentBlockStop": {
    "contentBlockIndex": number
  },
  "internalServerError": {
  },
  "messageStart": {
    "role": "string"
  },
  "messageStop": {
    "additionalModelResponseFields": JSON value,
    "stopReason": "string"
  },
  "metadata": {
    "metrics": {
      "latencyMs": number
    },
    "trace": {
      "guardrail": {
        "inputAssessment": {
          "string": {
            "contentPolicy": {
              "filters": [
                {
                  "action": "string",
                  "confidence": "string",
```

```
        "type": "string"
      }
    ]
  },
  "contextualGroundingPolicy": {
    "filters": [
      {
        "action": "string",
        "score": number,
        "threshold": number,
        "type": "string"
      }
    ]
  },
  "sensitiveInformationPolicy": {
    "piiEntities": [
      {
        "action": "string",
        "match": "string",
        "type": "string"
      }
    ],
    "regexes": [
      {
        "action": "string",
        "match": "string",
        "name": "string",
        "regex": "string"
      }
    ]
  },
  "topicPolicy": {
    "topics": [
      {
        "action": "string",
        "name": "string",
        "type": "string"
      }
    ]
  },
  "wordPolicy": {
    "customWords": [
      {
        "action": "string",
```

```

        "match": "string"
      }
    ],
    "managedWordLists": [
      {
        "action": "string",
        "match": "string",
        "type": "string"
      }
    ]
  }
},
"modelOutput": [ "string" ],
"outputAssessments": {
  "string" : [
    {
      "contentPolicy": {
        "filters": [
          {
            "action": "string",
            "confidence": "string",
            "type": "string"
          }
        ]
      },
      "contextualGroundingPolicy": {
        "filters": [
          {
            "action": "string",
            "score": number,
            "threshold": number,
            "type": "string"
          }
        ]
      },
      "sensitiveInformationPolicy": {
        "piiEntities": [
          {
            "action": "string",
            "match": "string",
            "type": "string"
          }
        ]
      }
    ]
  },

```

```
    "regexes": [
      {
        "action": "string",
        "match": "string",
        "name": "string",
        "regex": "string"
      }
    ],
  },
  "topicPolicy": {
    "topics": [
      {
        "action": "string",
        "name": "string",
        "type": "string"
      }
    ]
  },
  "wordPolicy": {
    "customWords": [
      {
        "action": "string",
        "match": "string"
      }
    ],
    "managedWordLists": [
      {
        "action": "string",
        "match": "string",
        "type": "string"
      }
    ]
  }
},
"usage": {
  "inputTokens": number,
  "outputTokens": number,
  "totalTokens": number
},
}
```

```
"modelStreamErrorException": {  
},  
"throttlingException": {  
},  
"validationException": {  
}  
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

contentBlockDelta

The messages output content block delta.

Type: [ContentBlockDeltaEvent](#) object

contentBlockStart

Start information for a content block.

Type: [ContentBlockStartEvent](#) object

contentBlockStop

Stop information for a content block.

Type: [ContentBlockStopEvent](#) object

internalServerErrorException

An internal server error occurred. Retry your request.

Type: Exception

HTTP Status Code: 500

messageStart

Message start information.

Type: [MessageStartEvent](#) object

messageStop

Message stop information.

Type: [MessageStopEvent](#) object

metadata

Metadata for the converse output stream.

Type: [ConverseStreamMetadataEvent](#) object

modelStreamErrorException

A streaming error occurred. Retry your request.

Type: Exception

HTTP Status Code: 424

throttlingException

The number of requests exceeds the limit. Resubmit your request later.

Type: Exception

HTTP Status Code: 429

validationException

Input validation failed. Check your request parameters and retry the request.

Type: Exception

HTTP Status Code: 400

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ModelErrorException

The request failed due to an error while processing the model.

HTTP Status Code: 424

ModelNotReadyException

The model specified in the request is not ready to serve inference requests.

HTTP Status Code: 429

ModelTimeoutException

The request took too long to process. Processing time exceeded the model timeout length.

HTTP Status Code: 408

ResourceNotFoundException

The specified resource ARN was not found. Check the ARN and try your request again.

HTTP Status Code: 404

ThrottlingException

Your request was throttled because of service-wide limitations. Resubmit your request later or in a different region. You can also purchase [Provisioned Throughput](#) to increase the rate or number of tokens you can process.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Send a message to a model and stream the response.

Send a message to Anthropic Claude Sonnet with `ConverseStream` and stream the response.

```
POST /model/anthropic.claude-3-sonnet-20240229-v1:0/converse-stream HTTP/1.1
```

```
{
  "messages": [
    {
      "role": "user",
      "content": [
        {
          "text": "Write an article about impact of high inflation to GDP of
a country"
        }
      ]
    }
  ],
  "system": [{"text" : "You are an economist with access to lots of data"}],
  "inferenceConfig": {
    "maxTokens": 1000,
    "temperature": 0.5
  }
}
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

InvokeModel

Service: Amazon Bedrock Runtime

Invokes the specified Amazon Bedrock model to run inference using the prompt and inference parameters provided in the request body. You use model inference to generate text, images, and embeddings.

For example code, see [Invoke model code examples](#).

This operation requires permission for the `bedrock:InvokeModel` action.

Request Syntax

```
POST /model/modelId/invoke HTTP/1.1
Accept: accept
Content-Type: contentType
X-Amzn-Bedrock-GuardrailIdentifier: guardrailIdentifier
X-Amzn-Bedrock-GuardrailVersion: guardrailVersion
X-Amzn-Bedrock-Trace: trace
```

body

URI Request Parameters

The request uses the following URI parameters.

[accept](#)

The desired MIME type of the inference body in the response. The default value is `application/json`.

[contentType](#)

The MIME type of the input data in the request. You must specify `application/json`.

[guardrailIdentifier](#)

The unique identifier of the guardrail that you want to use. If you don't provide a value, no guardrail is applied to the invocation.

An error will be thrown in the following situations.

- You don't provide a guardrail identifier but you specify the `amazon-bedrock-guardrailConfig` field in the request body.

- You enable the guardrail but the `contentType` isn't `application/json`.
- You provide a guardrail identifier, but `guardrailVersion` isn't specified.

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^(([a-z0-9]+)|(arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[\0-9]{12}:guardrail/[a-z0-9]+))$`

guardrailVersion

The version number for the guardrail. The value can also be `DRAFT`.

Pattern: `^(([1-9][0-9]{0,7})|(DRAFT))$`

modelId

The unique identifier of the model to invoke to run inference.

The `modelId` to provide depends on the type of model that you use:

- If you use a base model, specify the model ID or its ARN. For a list of model IDs for base models, see [Amazon Bedrock base model IDs \(on-demand throughput\)](#) in the Amazon Bedrock User Guide.
- If you use a provisioned model, specify the ARN of the Provisioned Throughput. For more information, see [Run inference using a Provisioned Throughput](#) in the Amazon Bedrock User Guide.
- If you use a custom model, first purchase Provisioned Throughput for it. Then specify the ARN of the resulting provisioned model. For more information, see [Use a custom model in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^(arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(([\0-9]{12}:custom-model/[a-z0-9-]{1,63}[\.]{1}[a-z0-9-]{1,63}/[a-z0-9]{12})|(:foundation-model/[a-z0-9-]{1,63}[\.]{1}[a-z0-9-]{1,63}([\.:]?[a-z0-9-]{1,63}))|([\0-9]{12}:provisioned-model/[a-z0-9]{12})))|([a-z0-9-]{1,63}[\.]{1}[a-z0-9-]{1,63}([\.:]?[a-z0-9-]{1,63}))|(([\0-9a-zA-Z][_]?)+)$`

Required: Yes

trace

Specifies whether to enable or disable the Bedrock trace. If enabled, you can see the full Bedrock trace.

Valid Values: ENABLED | DISABLED

Request Body

The request accepts the following binary data.

body

The prompt and inference parameters in the format specified in the `contentType` in the header. You must provide the body in JSON format. To see the format and content of the request and response bodies for different models, refer to [Inference parameters](#). For more information, see [Run inference](#) in the Bedrock User Guide.

Length Constraints: Minimum length of 0. Maximum length of 25000000.

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-Type: contentType

body
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The response returns the following HTTP headers.

contentType

The MIME type of the inference result.

The response returns the following as the HTTP body.

body

Inference response from the model in the format specified in the `contentType` header. To see the format and content of the request and response bodies for different models, refer to [Inference parameters](#).

Length Constraints: Minimum length of 0. Maximum length of 25000000.

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ModelErrorException

The request failed due to an error while processing the model.

HTTP Status Code: 424

ModelNotReadyException

The model specified in the request is not ready to serve inference requests.

HTTP Status Code: 429

ModelTimeoutException

The request took too long to process. Processing time exceeded the model timeout length.

HTTP Status Code: 408

ResourceNotFoundException

The specified resource ARN was not found. Check the ARN and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

Your request exceeds the service quota for your account. You can view your quotas at [Viewing service quotas](#). You can resubmit your request later.

HTTP Status Code: 400

ThrottlingException

Your request was throttled because of service-wide limitations. Resubmit your request later or in a different region. You can also purchase [Provisioned Throughput](#) to increase the rate or number of tokens you can process.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Run inference on a text model

Send an invoke request to run inference on a Titan Text G1 - Express model. We set the accept parameter to accept any content type in the response.

```
POST https://bedrock-runtime.us-east-1.amazonaws.com/model/amazon.titan-text-express-v1/invoke

-H accept: */*
-H content-type: application/json

Payload
{"inputText": "Hello world"}
```

Example response

Response for the above request.

```
-H content-type: application/json

Payload
<the model response>
```

Run inference on an image model

In the following example, the request sets the accept parameter to image/png.


```
POST https://bedrock-runtime.us-east-1.amazonaws.com/model/stability.stable-diffusion-
xl-v1/invoke
```

```
-H accept: image/png
-H content-type: application/json
```

```
Payload
{"inputText": "Picture of a bird"}
```

Example response

Response for the above example.

```
-H content-type: image/png
```

```
Payload
<image bytes>
```

Use a guardrail

This example shows how to use a guardrail with `InvokeModel`.

```
POST /model/modelId/invoke HTTP/1.1
Accept: accept
Content-Type: contentType
X-Amzn-Bedrock-GuardrailIdentifier: guardrailIdentifier
X-Amzn-Bedrock-GuardrailVersion: guardrailVersion
X-Amzn-Bedrock-GuardrailTrace: guardrailTrace
X-Amzn-Bedrock-Trace: trace
```

```
body
```

```
// body
{
  "amazon-bedrock-guardrailConfig": {
    "tagSuffix": "string"
  }
}
```

Example response

This is an example response from `InvokeModel` when using a guardrail.

```
HTTP/1.1 200
Content-Type: contentType

body

// body
{
  "amazon-bedrock-guardrailAction": "INTERVENED | NONE"
  "amazon-bedrock-trace": {
    "guardrails": {
      // Detailed guardrail trace
    }
  }
}
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

InvokeModelWithResponseStream

Service: Amazon Bedrock Runtime

Invoke the specified Amazon Bedrock model to run inference using the prompt and inference parameters provided in the request body. The response is returned in a stream.

To see if a model supports streaming, call [GetFoundationModel](#) and check the `responseStreamingSupported` field in the response.

Note

The AWS CLI doesn't support streaming operations in Amazon Bedrock, including `InvokeModelWithResponseStream`.

For example code, see [Invoke model with streaming code example](#).

This operation requires permissions to perform the `bedrock:InvokeModelWithResponseStream` action.

Request Syntax

```
POST /model/modelId/invoke-with-response-stream HTTP/1.1
X-Amzn-Bedrock-Accept: accept
Content-Type: contentType
X-Amzn-Bedrock-GuardrailIdentifier: guardrailIdentifier
X-Amzn-Bedrock-GuardrailVersion: guardrailVersion
X-Amzn-Bedrock-Trace: trace

body
```

URI Request Parameters

The request uses the following URI parameters.

[accept](#)

The desired MIME type of the inference body in the response. The default value is `application/json`.

[contentType](#)

The MIME type of the input data in the request. You must specify `application/json`.

guardrailIdentifier

The unique identifier of the guardrail that you want to use. If you don't provide a value, no guardrail is applied to the invocation.

An error is thrown in the following situations.

- You don't provide a guardrail identifier but you specify the `amazon-bedrock-guardrailConfig` field in the request body.
- You enable the guardrail but the `contentType` isn't `application/json`.
- You provide a guardrail identifier, but `guardrailVersion` isn't specified.

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^(([a-z0-9]+)|(arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:guardrail/[a-z0-9]+))$`

guardrailVersion

The version number for the guardrail. The value can also be DRAFT.

Pattern: `^(([1-9][0-9]{0,7})|(DRAFT))$`

modelId

The unique identifier of the model to invoke to run inference.

The `modelId` to provide depends on the type of model that you use:

- If you use a base model, specify the model ID or its ARN. For a list of model IDs for base models, see [Amazon Bedrock base model IDs \(on-demand throughput\)](#) in the Amazon Bedrock User Guide.
- If you use a provisioned model, specify the ARN of the Provisioned Throughput. For more information, see [Run inference using a Provisioned Throughput](#) in the Amazon Bedrock User Guide.
- If you use a custom model, first purchase Provisioned Throughput for it. Then specify the ARN of the resulting provisioned model. For more information, see [Use a custom model in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^(arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}/[a-z0-9]{12})|(:foundation-`

```
model/[a-z0-9-]{1,63}[\.]{1}[a-z0-9-]{1,63}([\.:]?[a-z0-9-]{1,63}))|([0-9]{12}:provisioned-model/[a-z0-9]{12}))|([a-z0-9-]{1,63}[\.]{1}[a-z0-9-]{1,63}([\.:]?[a-z0-9-]{1,63}))|(([0-9a-zA-Z][_]?)+)$
```

Required: Yes

trace

Specifies whether to enable or disable the Bedrock trace. If enabled, you can see the full Bedrock trace.

Valid Values: ENABLED | DISABLED

Request Body

The request accepts the following binary data.

body

The prompt and inference parameters in the format specified in the `contentType` in the header. You must provide the body in JSON format. To see the format and content of the request and response bodies for different models, refer to [Inference parameters](#). For more information, see [Run inference](#) in the Bedrock User Guide.

Length Constraints: Minimum length of 0. Maximum length of 25000000.

Required: Yes

Response Syntax

```
HTTP/1.1 200
X-Amzn-Bedrock-Content-Type: contentType
Content-type: application/json

{
  "chunk": {
    "bytes": blob
  },
  "internalServerErrorException": {
  },
  "modelStreamErrorException": {
  },
}
```

```
"modelTimeoutException": {  
},  
"throttlingException": {  
},  
"validationException": {  
}  
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The response returns the following HTTP headers.

[contentType](#)

The MIME type of the inference result.

The following data is returned in JSON format by the service.

[chunk](#)

Content included in the response.

Type: [PayloadPart](#) object

[internalServerErrorException](#)

An internal server error occurred. Retry your request.

Type: Exception

HTTP Status Code: 500

[modelStreamErrorException](#)

An error occurred while streaming the response. Retry your request.

Type: Exception

HTTP Status Code: 424

[modelTimeoutException](#)

The request took too long to process. Processing time exceeded the model timeout length.

Type: Exception

HTTP Status Code: 408

throttlingException

Your request was throttled because of service-wide limitations. Resubmit your request later or in a different region. You can also purchase [Provisioned Throughput](#) to increase the rate or number of tokens you can process.

Type: Exception

HTTP Status Code: 429

validationException

Input validation failed. Check your request parameters and retry the request.

Type: Exception

HTTP Status Code: 400

Errors

For information about the errors that are common to all actions, see [Common Errors](#).

AccessDeniedException

The request is denied because of missing access permissions.

HTTP Status Code: 403

InternalServerErrorException

An internal server error occurred. Retry your request.

HTTP Status Code: 500

ModelErrorException

The request failed due to an error while processing the model.

HTTP Status Code: 424

ModelNotReadyException

The model specified in the request is not ready to serve inference requests.

HTTP Status Code: 429

ModelStreamErrorException

An error occurred while streaming the response. Retry your request.

HTTP Status Code: 424

ModelTimeoutException

The request took too long to process. Processing time exceeded the model timeout length.

HTTP Status Code: 408

ResourceNotFoundException

The specified resource ARN was not found. Check the ARN and try your request again.

HTTP Status Code: 404

ServiceQuotaExceededException

Your request exceeds the service quota for your account. You can view your quotas at [Viewing service quotas](#). You can resubmit your request later.

HTTP Status Code: 400

ThrottlingException

Your request was throttled because of service-wide limitations. Resubmit your request later or in a different region. You can also purchase [Provisioned Throughput](#) to increase the rate or number of tokens you can process.

HTTP Status Code: 429

ValidationException

Input validation failed. Check your request parameters and retry the request.

HTTP Status Code: 400

Examples

Run inference with streaming on a text model

For streaming, you can set `x-amzn-bedrock-accept-type` in the header to contain the desired content type of the response. In this example, we set it to accept any content type. The default value is `application/json`.


```
POST https://bedrock-runtime.us-east-1.amazonaws.com/model/amazon.titan-text-express-
v1/invoke-with-response-stream
```

```
-H accept: application/vnd.amazon.eventstream
-H content-type: application/json
-H x-amzn-bedrock-accept: */*
```

Payload

```
{"inputText": "Hello world"}
```

Example response

For streaming, the content type in the response is always set to `application/vnd.amazon.eventstream`. The response includes an additional header (`x-amzn-bedrock-content-type`), which contains the actual content type of the response.

```
-H content-type: application/vnd.amazon.eventstream
-H x-amzn-bedrock-content-type: application/json
```

Payload (stream events)

```
<response chunk>
```

Use a guardrail

This examples show how to use a guardrail with `InvokeModelWithResponseStream`.

```
POST /model/modelId/invoke-with-response-stream HTTP/1.1
X-Amzn-Bedrock-Accept: accept
Content-Type: contentType
X-Amzn-Bedrock-GuardrailIdentifier: guardrailIdentifier
X-Amzn-Bedrock-GuardrailVersion: guardrailVersion
X-Amzn-Bedrock-GuardrailTrace: guardrailTrace
X-Amzn-Bedrock-Trace: trace
```

body

```
// body
{
  "amazon-bedrock-guardrailConfig": {
    "tagSuffix": "string",
    "streamProcessingMode": "string"
```

```
}  
}
```

Example response

This examples shows the response from a call to `InvokeModelWithResponseStream` when using a guardrail.

```
HTTP/1.1 200  
X-Amzn-Bedrock-Content-Type: contentType  
Content-type: application/json  
  
// chunk 1  
{  
  "completion": "...",  
  "amazon-bedrock-guardrailAction": "INTERVENED | NONE"  
}  
  
// chunk 2  
{  
  "completion": "...",  
  "amazon-bedrock-guardrailAction": "INTERVENED | NONE"  
}  
  
// last chunk  
{  
  "completion": "...",  
  "amazon-bedrock-guardrailAction": "INTERVENED | NONE",  
  "amazon-bedrock-trace": {  
    "guardrail": {  
      ... // Detailed guardrail trace  
    }  
  }  
}
```

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)

- [AWS SDK for C++](#)
- [AWS SDK for Go v2](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

Data Types

The following data types are supported by Amazon Bedrock:

- [AutomatedEvaluationConfig](#)
- [CloudWatchConfig](#)
- [CustomModelSummary](#)
- [EvaluationBedrockModel](#)
- [EvaluationConfig](#)
- [EvaluationDataset](#)
- [EvaluationDatasetLocation](#)
- [EvaluationDatasetMetricConfig](#)
- [EvaluationInferenceConfig](#)
- [EvaluationModelConfig](#)
- [EvaluationOutputDataConfig](#)
- [EvaluationSummary](#)
- [FoundationModelDetails](#)
- [FoundationModelLifecycle](#)
- [FoundationModelSummary](#)
- [GuardrailContentFilter](#)
- [GuardrailContentFilterConfig](#)
- [GuardrailContentPolicy](#)
- [GuardrailContentPolicyConfig](#)

- [GuardrailContextualGroundingFilter](#)
- [GuardrailContextualGroundingFilterConfig](#)
- [GuardrailContextualGroundingPolicy](#)
- [GuardrailContextualGroundingPolicyConfig](#)
- [GuardrailManagedWords](#)
- [GuardrailManagedWordsConfig](#)
- [GuardrailPiiEntity](#)
- [GuardrailPiiEntityConfig](#)
- [GuardrailRegex](#)
- [GuardrailRegexConfig](#)
- [GuardrailSensitiveInformationPolicy](#)
- [GuardrailSensitiveInformationPolicyConfig](#)
- [GuardrailSummary](#)
- [GuardrailTopic](#)
- [GuardrailTopicConfig](#)
- [GuardrailTopicPolicy](#)
- [GuardrailTopicPolicyConfig](#)
- [GuardrailWord](#)
- [GuardrailWordConfig](#)
- [GuardrailWordPolicy](#)
- [GuardrailWordPolicyConfig](#)
- [HumanEvaluationConfig](#)
- [HumanEvaluationCustomMetric](#)
- [HumanWorkflowConfig](#)
- [LoggingConfig](#)
- [ModelCustomizationJobSummary](#)
- [OutputDataConfig](#)
- [ProvisionedModelSummary](#)
- [S3Config](#)
- [Tag](#)

- [TrainingDataConfig](#)
- [TrainingMetrics](#)
- [ValidationDataConfig](#)
- [Validator](#)
- [ValidatorMetric](#)
- [VpcConfig](#)

The following data types are supported by Agents for Amazon Bedrock:

- [ActionGroupExecutor](#)
- [ActionGroupSummary](#)
- [Agent](#)
- [AgentActionGroup](#)
- [AgentAlias](#)
- [AgentAliasHistoryEvent](#)
- [AgentAliasRoutingConfigurationListItem](#)
- [AgentAliasSummary](#)
- [AgentFlowNodeConfiguration](#)
- [AgentKnowledgeBase](#)
- [AgentKnowledgeBaseSummary](#)
- [AgentSummary](#)
- [AgentVersion](#)
- [AgentVersionSummary](#)
- [APISchema](#)
- [BedrockEmbeddingModelConfiguration](#)
- [BedrockFoundationModelConfiguration](#)
- [ChunkingConfiguration](#)
- [CollectorFlowNodeConfiguration](#)
- [ConditionFlowNodeConfiguration](#)
- [ConfluenceCrawlerConfiguration](#)
- [ConfluenceDataSourceConfiguration](#)

- [ConfluenceSourceConfiguration](#)
- [CrawlFilterConfiguration](#)
- [CustomTransformationConfiguration](#)
- [DataSource](#)
- [DataSourceConfiguration](#)
- [DataSourceSummary](#)
- [EmbeddingModelConfiguration](#)
- [FixedSizeChunkingConfiguration](#)
- [FlowAliasRoutingConfigurationListItem](#)
- [FlowAliasSummary](#)
- [FlowCondition](#)
- [FlowConditionalConnectionConfiguration](#)
- [FlowConnection](#)
- [FlowConnectionConfiguration](#)
- [FlowDataConnectionConfiguration](#)
- [FlowDefinition](#)
- [FlowNode](#)
- [FlowNodeConfiguration](#)
- [FlowNodeInput](#)
- [FlowNodeOutput](#)
- [FlowSummary](#)
- [FlowValidation](#)
- [FlowVersionSummary](#)
- [Function](#)
- [FunctionSchema](#)
- [GuardrailConfiguration](#)
- [HierarchicalChunkingConfiguration](#)
- [HierarchicalChunkingLevelConfiguration](#)
- [InferenceConfiguration](#)
- [IngestionJob](#)

- [IngestionJobFilter](#)
- [IngestionJobSortBy](#)
- [IngestionJobStatistics](#)
- [IngestionJobSummary](#)
- [InputFlowNodeConfiguration](#)
- [IntermediateStorage](#)
- [IteratorFlowNodeConfiguration](#)
- [KnowledgeBase](#)
- [KnowledgeBaseConfiguration](#)
- [KnowledgeBaseFlowNodeConfiguration](#)
- [KnowledgeBaseSummary](#)
- [LambdaFunctionFlowNodeConfiguration](#)
- [LexFlowNodeConfiguration](#)
- [MemoryConfiguration](#)
- [MongoDbAtlasConfiguration](#)
- [MongoDbAtlasFieldMapping](#)
- [OpenSearchServerlessConfiguration](#)
- [OpenSearchServerlessFieldMapping](#)
- [OutputFlowNodeConfiguration](#)
- [ParameterDetail](#)
- [ParsingConfiguration](#)
- [ParsingPrompt](#)
- [PatternObjectFilter](#)
- [PatternObjectFilterConfiguration](#)
- [PineconeConfiguration](#)
- [PineconeFieldMapping](#)
- [PromptConfiguration](#)
- [PromptFlowNodeConfiguration](#)
- [PromptFlowNodeInlineConfiguration](#)
- [PromptFlowNodeResourceConfiguration](#)

- [PromptFlowNodeSourceConfiguration](#)
- [PromptInferenceConfiguration](#)
- [PromptInputVariable](#)
- [PromptModelInferenceConfiguration](#)
- [PromptOverrideConfiguration](#)
- [PromptSummary](#)
- [PromptTemplateConfiguration](#)
- [PromptVariant](#)
- [RdsConfiguration](#)
- [RdsFieldMapping](#)
- [RedisEnterpriseCloudConfiguration](#)
- [RedisEnterpriseCloudFieldMapping](#)
- [RetrievalFlowNodeConfiguration](#)
- [RetrievalFlowNodeS3Configuration](#)
- [RetrievalFlowNodeServiceConfiguration](#)
- [S3DataSourceConfiguration](#)
- [S3Identifier](#)
- [S3Location](#)
- [SalesforceCrawlerConfiguration](#)
- [SalesforceDataSourceConfiguration](#)
- [SalesforceSourceConfiguration](#)
- [SeedUrl](#)
- [SemanticChunkingConfiguration](#)
- [ServerSideEncryptionConfiguration](#)
- [SharePointCrawlerConfiguration](#)
- [SharePointDataSourceConfiguration](#)
- [SharePointSourceConfiguration](#)
- [StorageConfiguration](#)
- [StorageFlowNodeConfiguration](#)
- [StorageFlowNodeS3Configuration](#)

- [StorageFlowNodeServiceConfiguration](#)
- [TextPromptTemplateConfiguration](#)
- [Transformation](#)
- [TransformationFunction](#)
- [TransformationLambdaConfiguration](#)
- [UrlConfiguration](#)
- [ValidationExceptionField](#)
- [VectorIngestionConfiguration](#)
- [VectorKnowledgeBaseConfiguration](#)
- [WebCrawlerConfiguration](#)
- [WebCrawlerLimits](#)
- [WebDataSourceConfiguration](#)
- [WebSourceConfiguration](#)

The following data types are supported by Agents for Amazon Bedrock Runtime:

- [ActionGroupInvocationInput](#)
- [ActionGroupInvocationOutput](#)
- [ApiInvocationInput](#)
- [ApiParameter](#)
- [ApiRequestBody](#)
- [ApiResponse](#)
- [Attribution](#)
- [ByteContentDoc](#)
- [ByteContentFile](#)
- [Citation](#)
- [CodeInterpreterInvocationInput](#)
- [CodeInterpreterInvocationOutput](#)
- [ContentBody](#)
- [ExternalSource](#)
- [ExternalSourcesGenerationConfiguration](#)

- [ExternalSourcesRetrieveAndGenerateConfiguration](#)
- [FailureTrace](#)
- [FilePart](#)
- [FileSource](#)
- [FilterAttribute](#)
- [FinalResponse](#)
- [FlowCompletionEvent](#)
- [FlowInput](#)
- [FlowInputContent](#)
- [FlowOutputContent](#)
- [FlowOutputEvent](#)
- [FlowResponseStream](#)
- [FunctionInvocationInput](#)
- [FunctionParameter](#)
- [FunctionResult](#)
- [GeneratedResponsePart](#)
- [GenerationConfiguration](#)
- [GuardrailAssessment](#)
- [GuardrailConfiguration](#)
- [GuardrailContentFilter](#)
- [GuardrailContentPolicyAssessment](#)
- [GuardrailCustomWord](#)
- [GuardrailManagedWord](#)
- [GuardrailPiiEntityFilter](#)
- [GuardrailRegexFilter](#)
- [GuardrailSensitiveInformationPolicyAssessment](#)
- [GuardrailTopic](#)
- [GuardrailTopicPolicyAssessment](#)
- [GuardrailTrace](#)
- [GuardrailWordPolicyAssessment](#)

- [InferenceConfig](#)
- [InferenceConfiguration](#)
- [InputFile](#)
- [InvocationInput](#)
- [InvocationInputMember](#)
- [InvocationResultMember](#)
- [KnowledgeBaseConfiguration](#)
- [KnowledgeBaseLookupInput](#)
- [KnowledgeBaseLookupOutput](#)
- [KnowledgeBaseQuery](#)
- [KnowledgeBaseRetrievalConfiguration](#)
- [KnowledgeBaseRetrievalResult](#)
- [KnowledgeBaseRetrieveAndGenerateConfiguration](#)
- [KnowledgeBaseVectorSearchConfiguration](#)
- [Memory](#)
- [MemorySessionSummary](#)
- [ModelInvocationInput](#)
- [Observation](#)
- [OrchestrationConfiguration](#)
- [OrchestrationTrace](#)
- [OutputFile](#)
- [Parameter](#)
- [PayloadPart](#)
- [PostProcessingModelInvocationOutput](#)
- [PostProcessingParsedResponse](#)
- [PostProcessingTrace](#)
- [PreProcessingModelInvocationOutput](#)
- [PreProcessingParsedResponse](#)
- [PreProcessingTrace](#)
- [PromptTemplate](#)

- [PropertyParameters](#)
- [QueryTransformationConfiguration](#)
- [Rationale](#)
- [RepromptResponse](#)
- [RequestBody](#)
- [ResponseStream](#)
- [RetrievalFilter](#)
- [RetrievalResultConfluenceLocation](#)
- [RetrievalResultContent](#)
- [RetrievalResultLocation](#)
- [RetrievalResultS3Location](#)
- [RetrievalResultSalesforceLocation](#)
- [RetrievalResultSharePointLocation](#)
- [RetrievalResultWebLocation](#)
- [RetrieveAndGenerateConfiguration](#)
- [RetrieveAndGenerateInput](#)
- [RetrieveAndGenerateOutput](#)
- [RetrieveAndGenerateSessionConfiguration](#)
- [RetrievedReference](#)
- [ReturnControlPayload](#)
- [S3ObjectDoc](#)
- [S3ObjectFile](#)
- [SessionState](#)
- [Span](#)
- [TextInferenceConfig](#)
- [TextResponsePart](#)
- [Trace](#)
- [TracePart](#)

The following data types are supported by Amazon Bedrock Runtime:

- [AnyToolChoice](#)
- [AutoToolChoice](#)
- [ContentBlock](#)
- [ContentBlockDelta](#)
- [ContentBlockDeltaEvent](#)
- [ContentBlockStart](#)
- [ContentBlockStartEvent](#)
- [ContentBlockStopEvent](#)
- [ConverseMetrics](#)
- [ConverseOutput](#)
- [ConverseStreamMetadataEvent](#)
- [ConverseStreamMetrics](#)
- [ConverseStreamOutput](#)
- [ConverseStreamTrace](#)
- [ConverseTrace](#)
- [DocumentBlock](#)
- [DocumentSource](#)
- [GuardrailAssessment](#)
- [GuardrailConfiguration](#)
- [GuardrailContentBlock](#)
- [GuardrailContentFilter](#)
- [GuardrailContentPolicyAssessment](#)
- [GuardrailContextualGroundingFilter](#)
- [GuardrailContextualGroundingPolicyAssessment](#)
- [GuardrailConverseContentBlock](#)
- [GuardrailConverseTextBlock](#)
- [GuardrailCustomWord](#)
- [GuardrailManagedWord](#)
- [GuardrailOutputContent](#)
- [GuardrailPiiEntityFilter](#)

- [GuardrailRegexFilter](#)
- [GuardrailSensitiveInformationPolicyAssessment](#)
- [GuardrailStreamConfiguration](#)
- [GuardrailTextBlock](#)
- [GuardrailTopic](#)
- [GuardrailTopicPolicyAssessment](#)
- [GuardrailTraceAssessment](#)
- [GuardrailUsage](#)
- [GuardrailWordPolicyAssessment](#)
- [ImageBlock](#)
- [ImageSource](#)
- [InferenceConfiguration](#)
- [Message](#)
- [MessageStartEvent](#)
- [MessageStopEvent](#)
- [PayloadPart](#)
- [ResponseStream](#)
- [SpecificToolChoice](#)
- [SystemContentBlock](#)
- [TokenUsage](#)
- [Tool](#)
- [ToolChoice](#)
- [ToolConfiguration](#)
- [ToolInputSchema](#)
- [ToolResultBlock](#)
- [ToolResultContentBlock](#)
- [ToolSpecification](#)
- [ToolUseBlock](#)
- [ToolUseBlockDelta](#)
- [ToolUseBlockStart](#)

Amazon Bedrock

The following data types are supported by Amazon Bedrock:

- [AutomatedEvaluationConfig](#)
- [CloudWatchConfig](#)
- [CustomModelSummary](#)
- [EvaluationBedrockModel](#)
- [EvaluationConfig](#)
- [EvaluationDataset](#)
- [EvaluationDatasetLocation](#)
- [EvaluationDatasetMetricConfig](#)
- [EvaluationInferenceConfig](#)
- [EvaluationModelConfig](#)
- [EvaluationOutputDataConfig](#)
- [EvaluationSummary](#)
- [FoundationModelDetails](#)
- [FoundationModelLifecycle](#)
- [FoundationModelSummary](#)
- [GuardrailContentFilter](#)
- [GuardrailContentFilterConfig](#)
- [GuardrailContentPolicy](#)
- [GuardrailContentPolicyConfig](#)
- [GuardrailContextualGroundingFilter](#)
- [GuardrailContextualGroundingFilterConfig](#)
- [GuardrailContextualGroundingPolicy](#)
- [GuardrailContextualGroundingPolicyConfig](#)
- [GuardrailManagedWords](#)
- [GuardrailManagedWordsConfig](#)
- [GuardrailPiiEntity](#)
- [GuardrailPiiEntityConfig](#)

- [GuardrailRegex](#)
- [GuardrailRegexConfig](#)
- [GuardrailSensitiveInformationPolicy](#)
- [GuardrailSensitiveInformationPolicyConfig](#)
- [GuardrailSummary](#)
- [GuardrailTopic](#)
- [GuardrailTopicConfig](#)
- [GuardrailTopicPolicy](#)
- [GuardrailTopicPolicyConfig](#)
- [GuardrailWord](#)
- [GuardrailWordConfig](#)
- [GuardrailWordPolicy](#)
- [GuardrailWordPolicyConfig](#)
- [HumanEvaluationConfig](#)
- [HumanEvaluationCustomMetric](#)
- [HumanWorkflowConfig](#)
- [LoggingConfig](#)
- [ModelCustomizationJobSummary](#)
- [OutputDataConfig](#)
- [ProvisionedModelSummary](#)
- [S3Config](#)
- [Tag](#)
- [TrainingDataConfig](#)
- [TrainingMetrics](#)
- [ValidationDataConfig](#)
- [Validator](#)
- [ValidatorMetric](#)
- [VpcConfig](#)

AutomatedEvaluationConfig

Service: Amazon Bedrock

Use to specify a automatic model evaluation job. The `EvaluationDatasetMetricConfig` object is used to specify the prompt datasets, task type, and metric names.

Contents

`datasetMetricConfigs`

Specifies the required elements for an automatic model evaluation job.

Type: Array of [EvaluationDatasetMetricConfig](#) objects

Array Members: Minimum number of 1 item. Maximum number of 5 items.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

CloudWatchConfig

Service: Amazon Bedrock

CloudWatch logging configuration.

Contents

logGroupName

The log group name.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 512.

Required: Yes

roleArn

The role Amazon Resource Name (ARN).

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:iam::([0-9]{12})?:role/.[+]`

Required: Yes

largeDataDeliveryS3Config

S3 configuration for delivering a large amount of data.

Type: [S3Config](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)

- [AWS SDK for Ruby V3](#)

CustomModelSummary

Service: Amazon Bedrock

Summary information for a custom model.

Contents

baseModelArn

The base model Amazon Resource Name (ARN).

Type: String

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}/[a-z0-9]{12})|(:foundation-model/[a-z0-9-]{1,63}[.]{1}([a-z0-9-]{1,63}[.]{0,2}[a-z0-9-]{1,63}([[:]a-z0-9-]{1,63}){0,2})))$`

Required: Yes

baseModelName

The base model name.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}$`

Required: Yes

creationTime

Creation time of the model.

Type: Timestamp

Required: Yes

modelArn

The Amazon Resource Name (ARN) of the custom model.

Type: String

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:custom-model/[a-z0-9-]{1,63}[\.]{1}[a-z0-9-]{1,63}([a-z0-9-]{1,63}[\.]?){0,2}[a-z0-9-]{1,63}([:][a-z0-9-]{1,63}){0,2}/[a-z0-9]{12}$`

Required: Yes

modelName

The name of the custom model.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^([0-9a-zA-Z][_]?)+$`

Required: Yes

customizationType

Specifies whether to carry out continued pre-training of a model or whether to fine-tune it. For more information, see [Custom models](#).

Type: String

Valid Values: FINE_TUNING | CONTINUED_PRE_TRAINING

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

EvaluationBedrockModel

Service: Amazon Bedrock

Contains the ARN of the Amazon Bedrock models specified in your model evaluation job. Each Amazon Bedrock model supports different inferenceParams. To learn more about supported inference parameters for Amazon Bedrock models, see [Inference parameters for foundation models](#).

The inferenceParams are specified using JSON. To successfully insert JSON as string make sure that all quotations are properly escaped. For example, "temperature": "0.25" key value pair would need to be formatted as \"temperature\": \"0.25\" to successfully accepted in the request.

Contents

inferenceParams

Each Amazon Bedrock support different inference parameters that change how the model behaves during inference.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1023.

Required: Yes

modelIdentifier

The ARN of the Amazon Bedrock model specified.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(([\0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}(([:][a-z0-9-]{1,63}){0,2})?/[a-z0-9-]{12})|(:foundation-model/([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63} ([.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2})))|(([a-z0-9-]{1,63} [.] {1} [a-z0-9-]{1,63} ([.]? [a-z0-9-]{1,63}) ([:] [a-z0-9-]{1,63}) {0,2})))|(([\0-9a-zA-Z][_-]?)+)$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

EvaluationConfig

Service: Amazon Bedrock

Used to specify either a `AutomatedEvaluationConfig` or `HumanEvaluationConfig` object.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

automated

Used to specify an automated model evaluation job. See `AutomatedEvaluationConfig` to view the required parameters.

Type: [AutomatedEvaluationConfig](#) object

Required: No

human

Used to specify a model evaluation job that uses human workers. See `HumanEvaluationConfig` to view the required parameters.

Type: [HumanEvaluationConfig](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

EvaluationDataset

Service: Amazon Bedrock

Used to specify the name of a built-in prompt dataset and optionally, the Amazon S3 bucket where a custom prompt dataset is saved.

Contents

name

Used to specify supported built-in prompt datasets. Valid values are `Builtin.Bold`, `Builtin.BoolQ`, `Builtin.NaturalQuestions`, `Builtin.Gigaword`, `Builtin.RealToxicityPrompts`, `Builtin.TriviaQa`, `Builtin.T-Rex`, `Builtin.WomensEcommerceClothingReviews` and `Builtin.Wikitext2`.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^[0-9a-zA-Z-_.]+$`

Required: Yes

datasetLocation

For custom prompt datasets, you must specify the location in Amazon S3 where the prompt dataset is saved.

Type: [EvaluationDatasetLocation](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

EvaluationDatasetLocation

Service: Amazon Bedrock

The location in Amazon S3 where your prompt dataset is stored.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

s3Uri

The S3 URI of the S3 bucket specified in the job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1024.

Pattern: `^s3://[a-z0-9][\.\-a-z0-9]{1,61}[a-z0-9](/\.*)?$`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

EvaluationDatasetMetricConfig

Service: Amazon Bedrock

Defines the built-in prompt datasets, built-in metric names and custom metric names, and the task type.

Contents

dataset

Specifies the prompt dataset.

Type: [EvaluationDataset](#) object

Required: Yes

metricNames

The names of the metrics used. For automated model evaluation jobs valid values are "Builtin.Accuracy", "Builtin.Robustness", and "Builtin.Toxicity". In human-based model evaluation jobs the array of strings must match the name parameter specified in `HumanEvaluationCustomMetric`.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 10 items.

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^[0-9a-zA-Z-_.]+$`

Required: Yes

taskType

The task type you want the model to carry out.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^[A-Za-z0-9]+$`

Valid Values: Summarization | Classification | QuestionAndAnswer | Generation | Custom

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

EvaluationInferenceConfig

Service: Amazon Bedrock

Used to define the models you want used in your model evaluation job. Automated model evaluation jobs support only a single model. In a human-based model evaluation job, your annotator can compare the responses for up to two different models.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

models

Used to specify the models.

Type: Array of [EvaluationModelConfig](#) objects

Array Members: Minimum number of 1 item. Maximum number of 2 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

EvaluationModelConfig

Service: Amazon Bedrock

Defines the models used in the model evaluation job.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

bedrockModel

Defines the Amazon Bedrock model and inference parameters you want used.

Type: [EvaluationBedrockModel](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

EvaluationOutputDataConfig

Service: Amazon Bedrock

The Amazon S3 location where the results of your model evaluation job are saved.

Contents

s3Uri

The Amazon S3 URI where the results of model evaluation job are saved.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1024.

Pattern: `^s3://[a-z0-9][\.\-a-z0-9]{1,61}[a-z0-9](/*.*)?$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

EvaluationSummary

Service: Amazon Bedrock

A summary of the model evaluation job.

Contents

creationTime

When the model evaluation job was created.

Type: Timestamp

Required: Yes

evaluationTaskTypes

What task type was used in the model evaluation job.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 5 items.

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^[A-Za-z0-9]+$`

Valid Values: Summarization | Classification | QuestionAndAnswer | Generation | Custom

Required: Yes

jobArn

The Amazon Resource Name (ARN) of the model evaluation job.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:evaluation-job/[a-z0-9]{12}$`

Required: Yes

jobName

The name of the model evaluation job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^[a-z0-9](-*[a-z0-9]){0,62}$`

Required: Yes

jobType

The type, either human or automatic, of model evaluation job.

Type: String

Valid Values: Human | Automated

Required: Yes

modelIdentifiers

The Amazon Resource Names (ARNs) of the model(s) used in the model evaluation job.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 2 items.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(([\0-9]{12}:custom-model/[a-z0-9-]{1,63}[\.]{1}[a-z0-9-]{1,63}(([:][a-z0-9-]{1,63}){0,2})?/[a-z0-9]{12})|(:foundation-model/([a-z0-9-]{1,63}[\.]{1}[a-z0-9-]{1,63}([\.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2})))|((([a-z0-9-]{1,63}[\.]{1}[a-z0-9-]{1,63}([\.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2}))|(([\0-9a-zA-Z][_]?)+))$`

Required: Yes

status

The current status of the model evaluation job.

Type: String

Valid Values: InProgress | Completed | Failed | Stopping | Stopped

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FoundationModelDetails

Service: Amazon Bedrock

Information about a foundation model.

Contents

modelArn

The model Amazon Resource Name (ARN).

Type: String

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}::foundation-model/[a-z0-9-]{1,63}[.]{1}([a-z0-9-]{1,63}[.]){0,2}[a-z0-9-]{1,63}([:][a-z0-9-]{1,63}){0,2}$`

Required: Yes

modelId

The model identifier.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 140.

Pattern: `^[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([a-z0-9-]{1,63}[.]){0,2}[a-z0-9-]{1,63}([:][a-z0-9-]{1,63}){0,2}(/[a-z0-9]{12}|)$`

Required: Yes

customizationsSupported

The customization that the model supports.

Type: Array of strings

Valid Values: FINE_TUNING | CONTINUED_PRE_TRAINING

Required: No

inferenceTypesSupported

The inference types that the model supports.

Type: Array of strings

Valid Values: ON_DEMAND | PROVISIONED

Required: No

inputModalities

The input modalities that the model supports.

Type: Array of strings

Valid Values: TEXT | IMAGE | EMBEDDING

Required: No

modelLifecycle

Contains details about whether a model version is available or deprecated

Type: [FoundationModelLifecycle](#) object

Required: No

modelName

The model name.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 20.

Pattern: ^.*\$

Required: No

outputModalities

The output modalities that the model supports.

Type: Array of strings

Valid Values: TEXT | IMAGE | EMBEDDING

Required: No

providerName

The model's provider name.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 20.

Pattern: `^\.*$`

Required: No

responseStreamingSupported

Indicates whether the model supports streaming.

Type: Boolean

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FoundationModelLifecycle

Service: Amazon Bedrock

Details about whether a model version is available or deprecated.

Contents

status

Specifies whether a model version is available (ACTIVE) or deprecated (LEGACY).

Type: String

Valid Values: ACTIVE | LEGACY

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FoundationModelSummary

Service: Amazon Bedrock

Summary information for a foundation model.

Contents

modelArn

The Amazon Resource Name (ARN) of the foundation model.

Type: String

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}::foundation-model/[a-z0-9-]{1,63}[\.]?{1}([a-z0-9-]{1,63}[\.]?){0,2}[a-z0-9-]{1,63}([:][a-z0-9-]{1,63}){0,2}$`

Required: Yes

modelId

The model ID of the foundation model.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 140.

Pattern: `^[a-z0-9-]{1,63}[\.]?{1}[a-z0-9-]{1,63}([a-z0-9-]{1,63}[\.]?){0,2}[a-z0-9-]{1,63}([:][a-z0-9-]{1,63}){0,2}(/[a-z0-9]{12}|)$`

Required: Yes

customizationsSupported

Whether the model supports fine-tuning or continual pre-training.

Type: Array of strings

Valid Values: FINE_TUNING | CONTINUED_PRE_TRAINING

Required: No

inferenceTypesSupported

The inference types that the model supports.

Type: Array of strings

Valid Values: ON_DEMAND | PROVISIONED

Required: No

inputModalities

The input modalities that the model supports.

Type: Array of strings

Valid Values: TEXT | IMAGE | EMBEDDING

Required: No

modelLifecycle

Contains details about whether a model version is available or deprecated.

Type: [FoundationModelLifecycle](#) object

Required: No

modelName

The name of the model.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 20.

Pattern: `^.*$`

Required: No

outputModalities

The output modalities that the model supports.

Type: Array of strings

Valid Values: TEXT | IMAGE | EMBEDDING

Required: No

providerName

The model's provider name.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 20.

Pattern: `^\.*$`

Required: No

responseStreamingSupported

Indicates whether the model supports streaming.

Type: Boolean

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailContentFilter

Service: Amazon Bedrock

Contains filter strengths for harmful content. Guardrails support the following content filters to detect and filter harmful user inputs and FM-generated outputs.

- **Hate** – Describes language or a statement that discriminates, criticizes, insults, denounces, or dehumanizes a person or group on the basis of an identity (such as race, ethnicity, gender, religion, sexual orientation, ability, and national origin).
- **Insults** – Describes language or a statement that includes demeaning, humiliating, mocking, insulting, or belittling language. This type of language is also labeled as bullying.
- **Sexual** – Describes language or a statement that indicates sexual interest, activity, or arousal using direct or indirect references to body parts, physical traits, or sex.
- **Violence** – Describes language or a statement that includes glorification of or threats to inflict physical pain, hurt, or injury toward a person, group or thing.

Content filtering depends on the confidence classification of user inputs and FM responses across each of the four harmful categories. All input and output statements are classified into one of four confidence levels (NONE, LOW, MEDIUM, HIGH) for each harmful category. For example, if a statement is classified as *Hate* with HIGH confidence, the likelihood of the statement representing hateful content is high. A single statement can be classified across multiple categories with varying confidence levels. For example, a single statement can be classified as *Hate* with HIGH confidence, *Insults* with LOW confidence, *Sexual* with NONE confidence, and *Violence* with MEDIUM confidence.

For more information, see [Guardrails content filters](#).

This data type is used in the following API operations:

- [GetGuardrail response body](#)

Contents

inputStrength

The strength of the content filter to apply to prompts. As you increase the filter strength, the likelihood of filtering harmful content increases and the probability of seeing harmful content in your application reduces.

Type: String

Valid Values: NONE | LOW | MEDIUM | HIGH

Required: Yes

outputStrength

The strength of the content filter to apply to model responses. As you increase the filter strength, the likelihood of filtering harmful content increases and the probability of seeing harmful content in your application reduces.

Type: String

Valid Values: NONE | LOW | MEDIUM | HIGH

Required: Yes

type

The harmful category that the content filter is applied to.

Type: String

Valid Values: SEXUAL | VIOLENCE | HATE | INSULTS | MISCONDUCT | PROMPT_ATTACK

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailContentFilterConfig

Service: Amazon Bedrock

Contains filter strengths for harmful content. Guardrails support the following content filters to detect and filter harmful user inputs and FM-generated outputs.

- **Hate** – Describes language or a statement that discriminates, criticizes, insults, denounces, or dehumanizes a person or group on the basis of an identity (such as race, ethnicity, gender, religion, sexual orientation, ability, and national origin).
- **Insults** – Describes language or a statement that includes demeaning, humiliating, mocking, insulting, or belittling language. This type of language is also labeled as bullying.
- **Sexual** – Describes language or a statement that indicates sexual interest, activity, or arousal using direct or indirect references to body parts, physical traits, or sex.
- **Violence** – Describes language or a statement that includes glorification of or threats to inflict physical pain, hurt, or injury toward a person, group or thing.

Content filtering depends on the confidence classification of user inputs and FM responses across each of the four harmful categories. All input and output statements are classified into one of four confidence levels (NONE, LOW, MEDIUM, HIGH) for each harmful category. For example, if a statement is classified as *Hate* with HIGH confidence, the likelihood of the statement representing hateful content is high. A single statement can be classified across multiple categories with varying confidence levels. For example, a single statement can be classified as *Hate* with HIGH confidence, *Insults* with LOW confidence, *Sexual* with NONE confidence, and *Violence* with MEDIUM confidence.

For more information, see [Guardrails content filters](#).

Contents

inputStrength

The strength of the content filter to apply to prompts. As you increase the filter strength, the likelihood of filtering harmful content increases and the probability of seeing harmful content in your application reduces.

Type: String

Valid Values: NONE | LOW | MEDIUM | HIGH

Required: Yes

outputStrength

The strength of the content filter to apply to model responses. As you increase the filter strength, the likelihood of filtering harmful content increases and the probability of seeing harmful content in your application reduces.

Type: String

Valid Values: NONE | LOW | MEDIUM | HIGH

Required: Yes

type

The harmful category that the content filter is applied to.

Type: String

Valid Values: SEXUAL | VIOLENCE | HATE | INSULTS | MISCONDUCT | PROMPT_ATTACK

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailContentPolicy

Service: Amazon Bedrock

Contains details about how to handle harmful content.

This data type is used in the following API operations:

- [GetGuardrail response body](#)

Contents

filters

Contains the type of the content filter and how strongly it should apply to prompts and model responses.

Type: Array of [GuardrailContentFilter](#) objects

Array Members: Minimum number of 1 item. Maximum number of 6 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailContentPolicyConfig

Service: Amazon Bedrock

Contains details about how to handle harmful content.

Contents

filtersConfig

Contains the type of the content filter and how strongly it should apply to prompts and model responses.

Type: Array of [GuardrailContentFilterConfig](#) objects

Array Members: Minimum number of 1 item. Maximum number of 6 items.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailContextualGroundingFilter

Service: Amazon Bedrock

The details for the guardrails contextual grounding filter.

Contents

threshold

The threshold details for the guardrails contextual grounding filter.

Type: Double

Valid Range: Minimum value of 0.

Required: Yes

type

The filter type details for the guardrails contextual grounding filter.

Type: String

Valid Values: GROUNDING | RELEVANCE

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailContextualGroundingFilterConfig

Service: Amazon Bedrock

The filter configuration details for the guardrails contextual grounding filter.

Contents

threshold

The threshold details for the guardrails contextual grounding filter.

Type: Double

Valid Range: Minimum value of 0.

Required: Yes

type

The filter details for the guardrails contextual grounding filter.

Type: String

Valid Values: GROUNDING | RELEVANCE

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailContextualGroundingPolicy

Service: Amazon Bedrock

The details for the guardrails contextual grounding policy.

Contents

filters

The filter details for the guardrails contextual grounding policy.

Type: Array of [GuardrailContextualGroundingFilter](#) objects

Array Members: Minimum number of 1 item.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailContextualGroundingPolicyConfig

Service: Amazon Bedrock

The policy configuration details for the guardrails contextual grounding policy.

Contents

filtersConfig

The filter configuration details for the guardrails contextual grounding policy.

Type: Array of [GuardrailContextualGroundingFilterConfig](#) objects

Array Members: Minimum number of 1 item.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailManagedWords

Service: Amazon Bedrock

The managed word list that was configured for the guardrail. (This is a list of words that are pre-defined and managed by guardrails only.)

Contents

type

ManagedWords\$type The managed word type that was configured for the guardrail. (For now, we only offer profanity word list)

Type: String

Valid Values: PROFANITY

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailManagedWordsConfig

Service: Amazon Bedrock

The managed word list to configure for the guardrail.

Contents

type

The managed word type to configure for the guardrail.

Type: String

Valid Values: PROFANITY

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailPiiEntity

Service: Amazon Bedrock

The PII entity configured for the guardrail.

Contents

action

The configured guardrail action when PII entity is detected.

Type: String

Valid Values: BLOCK | ANONYMIZE

Required: Yes

type

The type of PII entity. For example, Social Security Number.

Type: String

Valid Values: ADDRESS | AGE | AWS_ACCESS_KEY | AWS_SECRET_KEY | CA_HEALTH_NUMBER | CA_SOCIAL_INSURANCE_NUMBER | CREDIT_DEBIT_CARD_CVV | CREDIT_DEBIT_CARD_EXPIRY | CREDIT_DEBIT_CARD_NUMBER | DRIVER_ID | EMAIL | INTERNATIONAL_BANK_ACCOUNT_NUMBER | IP_ADDRESS | LICENSE_PLATE | MAC_ADDRESS | NAME | PASSWORD | PHONE | PIN | SWIFT_CODE | UK_NATIONAL_HEALTH_SERVICE_NUMBER | UK_NATIONAL_INSURANCE_NUMBER | UK_UNIQUE_TAXPAYER_REFERENCE_NUMBER | URL | USERNAME | US_BANK_ACCOUNT_NUMBER | US_BANK_ROUTING_NUMBER | US_INDIVIDUAL_TAX_IDENTIFICATION_NUMBER | US_PASSPORT_NUMBER | US_SOCIAL_SECURITY_NUMBER | VEHICLE_IDENTIFICATION_NUMBER

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)

- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailPiiEntityConfig

Service: Amazon Bedrock

The PII entity to configure for the guardrail.

Contents

action

Configure guardrail action when the PII entity is detected.

Type: String

Valid Values: BLOCK | ANONYMIZE

Required: Yes

type

Configure guardrail type when the PII entity is detected.

The following PII types are used to block or mask sensitive information:

- **General**

- **ADDRESS**

A physical address, such as "100 Main Street, Anytown, USA" or "Suite #12, Building 123". An address can include information such as the street, building, location, city, state, country, county, zip code, precinct, and neighborhood.

- **AGE**

An individual's age, including the quantity and unit of time. For example, in the phrase "I am 40 years old," Guardrails recognizes "40 years" as an age.

- **NAME**

An individual's name. This entity type does not include titles, such as Dr., Mr., Mrs., or Miss. Guardrails doesn't apply this entity type to names that are part of organizations or addresses. For example, Guardrails recognizes the "John Doe Organization" as an organization, and it recognizes "Jane Doe Street" as an address.

- **EMAIL**

An email address, such as *marymajor@email.com*.

- **PHONE**

A phone number. This entity type also includes fax and pager numbers.

- **USERNAME**

A user name that identifies an account, such as a login name, screen name, nick name, or handle.

- **PASSWORD**

An alphanumeric string that is used as a password, such as `"*very20special#pass*"`.

- **DRIVER_ID**

The number assigned to a driver's license, which is an official document permitting an individual to operate one or more motorized vehicles on a public road. A driver's license number consists of alphanumeric characters.

- **LICENSE_PLATE**

A license plate for a vehicle is issued by the state or country where the vehicle is registered. The format for passenger vehicles is typically five to eight digits, consisting of upper-case letters and numbers. The format varies depending on the location of the issuing state or country.

- **VEHICLE_IDENTIFICATION_NUMBER**

A Vehicle Identification Number (VIN) uniquely identifies a vehicle. VIN content and format are defined in the *ISO 3779* specification. Each country has specific codes and formats for VINs.

- **Finance**

- **REDIT_DEBIT_CARD_CVV**

A three-digit card verification code (CVV) that is present on VISA, MasterCard, and Discover credit and debit cards. For American Express credit or debit cards, the CVV is a four-digit numeric code.

- **CREDIT_DEBIT_CARD_EXPIRY**

The expiration date for a credit or debit card. This number is usually four digits long and is often formatted as *month/year* or *MM/YY*. Guardrails recognizes expiration dates such as *01/21*, *01/2021*, and *Jan 2021*.

- **CREDIT_DEBIT_CARD_NUMBER**

The number for a credit or debit card. These numbers can vary from 13 to 16 digits in length. However, Amazon Comprehend also recognizes credit or debit card numbers when only the last four digits are present.

- **PIN**

A four-digit personal identification number (PIN) with which you can access your bank account.

- **INTERNATIONAL_BANK_ACCOUNT_NUMBER**

An International Bank Account Number has specific formats in each country. For more information, see www.iban.com/structure.

- **SWIFT_CODE**

A SWIFT code is a standard format of Bank Identifier Code (BIC) used to specify a particular bank or branch. Banks use these codes for money transfers such as international wire transfers.

SWIFT codes consist of eight or 11 characters. The 11-digit codes refer to specific branches, while eight-digit codes (or 11-digit codes ending in 'XXX') refer to the head or primary office.

- **IT**

- **IP_ADDRESS**

An IPv4 address, such as *198.51.100.0*.

- **MAC_ADDRESS**

A *media access control* (MAC) address is a unique identifier assigned to a network interface controller (NIC).

- **URL**

A web address, such as *www.example.com*.

- **AWS_ACCESS_KEY**

A unique identifier that's associated with a secret access key; you use the access key ID and secret access key to sign programmatic AWS requests cryptographically.

- **AWS_SECRET_KEY**

A unique identifier that's associated with an access key. You use the access key ID and secret access key to sign programmatic AWS requests cryptographically.

- **USA specific**

- **US_BANK_ACCOUNT_NUMBER**

A US bank account number, which is typically 10 to 12 digits long.

- **US_BANK_ROUTING_NUMBER**

A US bank account routing number. These are typically nine digits long,

- **US_INDIVIDUAL_TAX_IDENTIFICATION_NUMBER**

A US Individual Taxpayer Identification Number (ITIN) is a nine-digit number that starts with a "9" and contain a "7" or "8" as the fourth digit. An ITIN can be formatted with a space or a dash after the third and forth digits.

- **US_PASSPORT_NUMBER**

A US passport number. Passport numbers range from six to nine alphanumeric characters.

- **US_SOCIAL_SECURITY_NUMBER**

A US Social Security Number (SSN) is a nine-digit number that is issued to US citizens, permanent residents, and temporary working residents.

- **Canada specific**

- **CA_HEALTH_NUMBER**

A Canadian Health Service Number is a 10-digit unique identifier, required for individuals to access healthcare benefits.

- **CA_SOCIAL_INSURANCE_NUMBER**

A Canadian Social Insurance Number (SIN) is a nine-digit unique identifier, required for individuals to access government programs and benefits.

The SIN is formatted as three groups of three digits, such as *123-456-789*. A SIN can be validated through a simple check-digit process called the [Luhn algorithm](#).

- **UK Specific**

- **UK_NATIONAL_HEALTH_SERVICE_NUMBER**

A UK National Health Service Number is a 10-17 digit number, such as *485 777 3456*. The current system formats the 10-digit number with spaces after the third and sixth digits. The final digit is an error-detecting checksum.

- **UK_NATIONAL_INSURANCE_NUMBER**

A UK National Insurance Number (NINO) provides individuals with access to National Insurance (social security) benefits. It is also used for some purposes in the UK tax system.

The number is nine digits long and starts with two letters, followed by six numbers and one letter. A NINO can be formatted with a space or a dash after the two letters and after the second, fourth, and sixth digits.

- **UK_UNIQUE_TAXPAYER_REFERENCE_NUMBER**

A UK Unique Taxpayer Reference (UTR) is a 10-digit number that identifies a taxpayer or a business.

- **Custom**

- **Regex filter** - You can use a regular expressions to define patterns for a guardrail to recognize and act upon such as serial number, booking ID etc..

Type: String

Valid Values: ADDRESS | AGE | AWS_ACCESS_KEY | AWS_SECRET_KEY | CA_HEALTH_NUMBER | CA_SOCIAL_INSURANCE_NUMBER | CREDIT_DEBIT_CARD_CVV | CREDIT_DEBIT_CARD_EXPIRY | CREDIT_DEBIT_CARD_NUMBER | DRIVER_ID | EMAIL | INTERNATIONAL_BANK_ACCOUNT_NUMBER | IP_ADDRESS | LICENSE_PLATE | MAC_ADDRESS | NAME | PASSWORD | PHONE | PIN | SWIFT_CODE | UK_NATIONAL_HEALTH_SERVICE_NUMBER | UK_NATIONAL_INSURANCE_NUMBER | UK_UNIQUE_TAXPAYER_REFERENCE_NUMBER | URL | USERNAME | US_BANK_ACCOUNT_NUMBER | US_BANK_ROUTING_NUMBER | US_INDIVIDUAL_TAX_IDENTIFICATION_NUMBER | US_PASSPORT_NUMBER | US_SOCIAL_SECURITY_NUMBER | VEHICLE_IDENTIFICATION_NUMBER

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailRegex

Service: Amazon Bedrock

The regular expression configured for the guardrail.

Contents

action

The action taken when a match to the regular expression is detected.

Type: String

Valid Values: BLOCK | ANONYMIZE

Required: Yes

name

The name of the regular expression for the guardrail.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 100.

Required: Yes

pattern

The pattern of the regular expression configured for the guardrail.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 500.

Required: Yes

description

The description of the regular expression for the guardrail.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1000.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailRegexConfig

Service: Amazon Bedrock

The regular expression to configure for the guardrail.

Contents

action

The guardrail action to configure when matching regular expression is detected.

Type: String

Valid Values: BLOCK | ANONYMIZE

Required: Yes

name

The name of the regular expression to configure for the guardrail.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 100.

Required: Yes

pattern

The regular expression pattern to configure for the guardrail.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 500.

Required: Yes

description

The description of the regular expression to configure for the guardrail.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1000.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailSensitiveInformationPolicy

Service: Amazon Bedrock

Contains details about PII entities and regular expressions configured for the guardrail.

Contents

piiEntities

The list of PII entities configured for the guardrail.

Type: Array of [GuardrailPiiEntity](#) objects

Array Members: Minimum number of 1 item.

Required: No

regexes

The list of regular expressions configured for the guardrail.

Type: Array of [GuardrailRegex](#) objects

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailSensitiveInformationPolicyConfig

Service: Amazon Bedrock

Contains details about PII entities and regular expressions to configure for the guardrail.

Contents

piiEntitiesConfig

A list of PII entities to configure to the guardrail.

Type: Array of [GuardrailPiiEntityConfig](#) objects

Array Members: Minimum number of 1 item.

Required: No

regexesConfig

A list of regular expressions to configure to the guardrail.

Type: Array of [GuardrailRegexConfig](#) objects

Array Members: Minimum number of 1 item. Maximum number of 10 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailSummary

Service: Amazon Bedrock

Contains details about a guardrail.

This data type is used in the following API operations:

- [ListGuardrails response body](#)

Contents

arn

The ARN of the guardrail.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:guardrail/[a-z0-9]+$`

Required: Yes

createdAt

The date and time at which the guardrail was created.

Type: Timestamp

Required: Yes

id

The unique identifier of the guardrail.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 64.

Pattern: `^[a-z0-9]+$`

Required: Yes

name

The name of the guardrail.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 50.

Pattern: `^[0-9a-zA-Z-_]+$`

Required: Yes

status

The status of the guardrail.

Type: String

Valid Values: CREATING | UPDATING | VERSIONING | READY | FAILED | DELETING

Required: Yes

updatedAt

The date and time at which the guardrail was last updated.

Type: Timestamp

Required: Yes

version

The version of the guardrail.

Type: String

Pattern: `^(([1-9][0-9]{0,7})|(DRAFT))$`

Required: Yes

description

A description of the guardrail.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailTopic

Service: Amazon Bedrock

Details about topics for the guardrail to identify and deny.

This data type is used in the following API operations:

- [GetGuardrail response body](#)

Contents

definition

A definition of the topic to deny.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: Yes

name

The name of the topic to deny.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 100.

Pattern: `^[0-9a-zA-Z- _ !? .]+`

Required: Yes

examples

A list of prompts, each of which is an example of a prompt that can be categorized as belonging to the topic.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 5 items.

Length Constraints: Minimum length of 1. Maximum length of 100.

Required: No

type

Specifies to deny the topic.

Type: String

Valid Values: DENY

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailTopicConfig

Service: Amazon Bedrock

Details about topics for the guardrail to identify and deny.

Contents

definition

A definition of the topic to deny.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: Yes

name

The name of the topic to deny.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 100.

Pattern: `^[0-9a-zA-Z- _ !? .]+$`

Required: Yes

type

Specifies to deny the topic.

Type: String

Valid Values: DENY

Required: Yes

examples

A list of prompts, each of which is an example of a prompt that can be categorized as belonging to the topic.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 5 items.

Length Constraints: Minimum length of 1. Maximum length of 100.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailTopicPolicy

Service: Amazon Bedrock

Contains details about topics that the guardrail should identify and deny.

This data type is used in the following API operations:

- [GetGuardrail response body](#)

Contents

topics

A list of policies related to topics that the guardrail should deny.

Type: Array of [GuardrailTopic](#) objects

Array Members: Minimum number of 1 item. Maximum number of 30 items.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailTopicPolicyConfig

Service: Amazon Bedrock

Contains details about topics that the guardrail should identify and deny.

Contents

topicsConfig

A list of policies related to topics that the guardrail should deny.

Type: Array of [GuardrailTopicConfig](#) objects

Array Members: Minimum number of 1 item. Maximum number of 30 items.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailWord

Service: Amazon Bedrock

A word configured for the guardrail.

Contents

text

Text of the word configured for the guardrail to block.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 100.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailWordConfig

Service: Amazon Bedrock

A word to configure for the guardrail.

Contents

text

Text of the word configured for the guardrail to block.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 100.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailWordPolicy

Service: Amazon Bedrock

Contains details about the word policy configured for the guardrail.

Contents

managedWordLists

A list of managed words configured for the guardrail.

Type: Array of [GuardrailManagedWords](#) objects

Required: No

words

A list of words configured for the guardrail.

Type: Array of [GuardrailWord](#) objects

Array Members: Minimum number of 1 item. Maximum number of 10000 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailWordPolicyConfig

Service: Amazon Bedrock

Contains details about the word policy to configured for the guardrail.

Contents

managedWordListsConfig

A list of managed words to configure for the guardrail.

Type: Array of [GuardrailManagedWordsConfig](#) objects

Required: No

wordsConfig

A list of words to configure for the guardrail.

Type: Array of [GuardrailWordConfig](#) objects

Array Members: Minimum number of 1 item. Maximum number of 10000 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

HumanEvaluationConfig

Service: Amazon Bedrock

Specifies the custom metrics, how tasks will be rated, the flow definition ARN, and your custom prompt datasets. Model evaluation jobs use human workers *only* support the use of custom prompt datasets. To learn more about custom prompt datasets and the required format, see [Custom prompt datasets](#).

When you create custom metrics in `HumanEvaluationCustomMetric` you must specify the metric's name. The list of names specified in the `HumanEvaluationCustomMetric` array, must match the `metricNames` array of strings specified in `EvaluationDatasetMetricConfig`. For example, if in the `HumanEvaluationCustomMetric` array you specified the names "accuracy", "toxicity", "readability" as custom metrics *then* the `metricNames` array would need to look like the following ["accuracy", "toxicity", "readability"] in `EvaluationDatasetMetricConfig`.

Contents

`datasetMetricConfigs`

Use to specify the metrics, task, and prompt dataset to be used in your model evaluation job.

Type: Array of [EvaluationDatasetMetricConfig](#) objects

Array Members: Minimum number of 1 item. Maximum number of 5 items.

Required: Yes

`customMetrics`

A `HumanEvaluationCustomMetric` object. It contains the names the metrics, how the metrics are to be evaluated, an optional description.

Type: Array of [HumanEvaluationCustomMetric](#) objects

Array Members: Minimum number of 1 item. Maximum number of 10 items.

Required: No

`humanWorkflowConfig`

The parameters of the human workflow.

Type: [HumanWorkflowConfig](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

HumanEvaluationCustomMetric

Service: Amazon Bedrock

In a model evaluation job that uses human workers you must define the name of the metric, and how you want that metric rated `ratingMethod`, and an optional description of the metric.

Contents

name

The name of the metric. Your human evaluators will see this name in the evaluation UI.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^[0-9a-zA-Z-_.]+$`

Required: Yes

ratingMethod

Choose how you want your human workers to evaluation your model. Valid values for rating methods are `ThumbsUpDown`, `IndividualLikertScale`, `ComparisonLikertScale`, `ComparisonChoice`, and `ComparisonRank`

Type: String

Length Constraints: Minimum length of 1. Maximum length of 100.

Pattern: `^[0-9a-zA-Z-_.]+$`

Required: Yes

description

An optional description of the metric. Use this parameter to provide more details about the metric.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^.+`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

HumanWorkflowConfig

Service: Amazon Bedrock

Contains SageMakerFlowDefinition object. The object is used to specify the prompt dataset, task type, rating method and metric names.

Contents

flowDefinitionArn

The Amazon Resource Number (ARN) for the flow definition

Type: String

Length Constraints: Minimum length of 0. Maximum length of 1024.

Pattern: `^arn:aws(-[^\:]+)?:sagemaker:[a-z0-9-]{1,20}:[0-9]{12}:flow-definition/.*$`

Required: Yes

instructions

Instructions for the flow definition

Type: String

Length Constraints: Minimum length of 1. Maximum length of 5000.

Pattern: `^[\\S\\s]+$`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

LoggingConfig

Service: Amazon Bedrock

Configuration fields for invocation logging.

Contents

cloudWatchConfig

CloudWatch logging configuration.

Type: [CloudWatchConfig](#) object

Required: No

embeddingDataDeliveryEnabled

Set to include embeddings data in the log delivery.

Type: Boolean

Required: No

imageDataDeliveryEnabled

Set to include image data in the log delivery.

Type: Boolean

Required: No

s3Config

S3 configuration for storing log data.

Type: [S3Config](#) object

Required: No

textDataDeliveryEnabled

Set to include text data in the log delivery.

Type: Boolean

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ModelCustomizationJobSummary

Service: Amazon Bedrock

Information about one customization job

Contents

baseModelArn

Amazon Resource Name (ARN) of the base model.

Type: String

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}/[a-z0-9]{12})|(:foundation-model/[a-z0-9-]{1,63}[.]{1}([a-z0-9-]{1,63}[.]{0,2}[a-z0-9-]{1,63}(:[a-z0-9-]{1,63}){0,2})))$`

Required: Yes

creationTime

Creation time of the custom model.

Type: Timestamp

Required: Yes

jobArn

Amazon Resource Name (ARN) of the customization job.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:([0-9]{12}:model-customization-job/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([a-z0-9-]{1,63}[.]{0,2}[a-z0-9-]{1,63}(:[a-z0-9-]{1,63}){0,2})/[a-z0-9]{12}$`

Required: Yes

jobName

Name of the customization job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9\+\-\.\.])*$`

Required: Yes

status

Status of the customization job.

Type: String

Valid Values: `InProgress` | `Completed` | `Failed` | `Stopping` | `Stopped`

Required: Yes

customizationType

Specifies whether to carry out continued pre-training of a model or whether to fine-tune it. For more information, see [Custom models](#).

Type: String

Valid Values: `FINE_TUNING` | `CONTINUED_PRE_TRAINING`

Required: No

customModelArn

Amazon Resource Name (ARN) of the custom model.

Type: String

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([a-z0-9-]{1,63}[.]{1}){0,2}[a-z0-9-]{1,63}([:][a-z0-9-]{1,63}){0,2}/[a-z0-9]{12}$`

Required: No

customModelName

Name of the custom model.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^[0-9a-zA-Z][_ -]?$`

Required: No

endTime

Time that the customization job ended.

Type: Timestamp

Required: No

lastModifiedTime

Time that the customization job was last modified.

Type: Timestamp

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

OutputDataConfig

Service: Amazon Bedrock

S3 Location of the output data.

Contents

s3Uri

The S3 URI where the output data is stored.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1024.

Pattern: `^s3://[a-z0-9][\.\-a-z0-9]{1,61}[a-z0-9](/.*)?$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ProvisionedModelSummary

Service: Amazon Bedrock

A summary of information about a Provisioned Throughput.

This data type is used in the following API operations:

- [ListProvisionedThroughputs response](#)

Contents

creationTime

The time that the Provisioned Throughput was created.

Type: Timestamp

Required: Yes

desiredModelArn

The Amazon Resource Name (ARN) of the model requested to be associated to this Provisioned Throughput. This value differs from the `modelArn` if updating hasn't completed.

Type: String

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(([\0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}/[a-z0-9]{12})|(:foundation-model/[a-z0-9-]{1,63}[.]{1}([a-z0-9-]{1,63}[.]){\0,2}[a-z0-9-]{1,63}([:][a-z0-9-]{1,63}){\0,2}))$`

Required: Yes

desiredModelUnits

The number of model units that was requested to be allocated to the Provisioned Throughput.

Type: Integer

Valid Range: Minimum value of 1.

Required: Yes

foundationModelArn

The Amazon Resource Name (ARN) of the base model for which the Provisioned Throughput was created, or of the base model that the custom model for which the Provisioned Throughput was created was customized.

Type: String

Pattern: `^arn:aws(-[:]+)?:bedrock:[a-z0-9-]{1,20}::foundation-model/[a-z0-9-]{1,63}[.]{1}([a-z0-9-]{1,63}[.]){0,2}[a-z0-9-]{1,63}([:][a-z0-9-]{1,63}){0,2}$`

Required: Yes

lastModifiedTime

The time that the Provisioned Throughput was last modified.

Type: Timestamp

Required: Yes

modelArn

The Amazon Resource Name (ARN) of the model associated with the Provisioned Throughput.

Type: String

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: `^arn:aws(-[:]+)?:bedrock:[a-z0-9-]{1,20}:(([\0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}/[a-z0-9-]{12})|(:foundation-model/[a-z0-9-]{1,63}[.]{1}([a-z0-9-]{1,63}[.]){0,2}[a-z0-9-]{1,63}([:][a-z0-9-]{1,63}){0,2}))$`

Required: Yes

modelUnits

The number of model units allocated to the Provisioned Throughput.

Type: Integer

Valid Range: Minimum value of 1.

Required: Yes

provisionedModelArn

The Amazon Resource Name (ARN) of the Provisioned Throughput.

Type: String

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:provisioned-model/[a-z0-9]{12}$`

Required: Yes

provisionedModelName

The name of the Provisioned Throughput.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 63.

Pattern: `^([0-9a-zA-Z][_]?)+$`

Required: Yes

status

The status of the Provisioned Throughput.

Type: String

Valid Values: `Creating` | `InService` | `Updating` | `Failed`

Required: Yes

commitmentDuration

The duration for which the Provisioned Throughput was committed.

Type: String

Valid Values: `OneMonth` | `SixMonths`

Required: No

commitmentExpirationTime

The timestamp for when the commitment term of the Provisioned Throughput expires.

Type: Timestamp

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

S3Config

Service: Amazon Bedrock

S3 configuration for storing log data.

Contents

bucketName

S3 bucket name.

Type: String

Length Constraints: Minimum length of 3. Maximum length of 63.

Required: Yes

keyPrefix

S3 prefix.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 1024.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

Tag

Service: Amazon Bedrock

Definition of the key/value pair for a tag.

Contents

key

Key for the tag.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 128.

Pattern: `^[a-zA-Z0-9\s._:/+=@-]*$`

Required: Yes

value

Value for the tag.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

Pattern: `^[a-zA-Z0-9\s._:/+=@-]*$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

TrainingDataConfig

Service: Amazon Bedrock

S3 Location of the training data.

Contents

s3Uri

The S3 URI where the training data is stored.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1024.

Pattern: `^s3://[a-z0-9][\.\-a-z0-9]{1,61}[a-z0-9](/\.*)?$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

TrainingMetrics

Service: Amazon Bedrock

Metrics associated with the custom job.

Contents

trainingLoss

Loss metric associated with the custom job.

Type: Float

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ValidationDataConfig

Service: Amazon Bedrock

Array of up to 10 validators.

Contents

validators

Information about the validators.

Type: Array of [Validator](#) objects

Array Members: Minimum number of 0 items. Maximum number of 10 items.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

Validator

Service: Amazon Bedrock

Information about a validator.

Contents

s3Uri

The S3 URI where the validation data is stored.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1024.

Pattern: `^s3://[a-z0-9][\.\-a-z0-9]{1,61}[a-z0-9](/.*)?$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ValidatorMetric

Service: Amazon Bedrock

The metric for the validator.

Contents

validationLoss

The validation loss associated with this validator.

Type: Float

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

VpcConfig

Service: Amazon Bedrock

VPC configuration.

Contents

securityGroupIds

VPC configuration security group Ids.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 5 items.

Length Constraints: Minimum length of 0. Maximum length of 32.

Pattern: `^[-0-9a-zA-Z]+$`

Required: Yes

subnetIds

VPC configuration subnets.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 16 items.

Length Constraints: Minimum length of 0. Maximum length of 32.

Pattern: `^[-0-9a-zA-Z]+$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

Agents for Amazon Bedrock

The following data types are supported by Agents for Amazon Bedrock:

- [ActionGroupExecutor](#)
- [ActionGroupSummary](#)
- [Agent](#)
- [AgentActionGroup](#)
- [AgentAlias](#)
- [AgentAliasHistoryEvent](#)
- [AgentAliasRoutingConfigurationListItem](#)
- [AgentAliasSummary](#)
- [AgentFlowNodeConfiguration](#)
- [AgentKnowledgeBase](#)
- [AgentKnowledgeBaseSummary](#)
- [AgentSummary](#)
- [AgentVersion](#)
- [AgentVersionSummary](#)
- [APISchema](#)
- [BedrockEmbeddingModelConfiguration](#)
- [BedrockFoundationModelConfiguration](#)
- [ChunkingConfiguration](#)
- [CollectorFlowNodeConfiguration](#)
- [ConditionFlowNodeConfiguration](#)
- [ConfluenceCrawlerConfiguration](#)
- [ConfluenceDataSourceConfiguration](#)
- [ConfluenceSourceConfiguration](#)
- [CrawlFilterConfiguration](#)
- [CustomTransformationConfiguration](#)
- [DataSource](#)
- [DataSourceConfiguration](#)

- [DataSourceSummary](#)
- [EmbeddingModelConfiguration](#)
- [FixedSizeChunkingConfiguration](#)
- [FlowAliasRoutingConfigurationListItem](#)
- [FlowAliasSummary](#)
- [FlowCondition](#)
- [FlowConditionalConnectionConfiguration](#)
- [FlowConnection](#)
- [FlowConnectionConfiguration](#)
- [FlowDataConnectionConfiguration](#)
- [FlowDefinition](#)
- [FlowNode](#)
- [FlowNodeConfiguration](#)
- [FlowNodeInput](#)
- [FlowNodeOutput](#)
- [FlowSummary](#)
- [FlowValidation](#)
- [FlowVersionSummary](#)
- [Function](#)
- [FunctionSchema](#)
- [GuardrailConfiguration](#)
- [HierarchicalChunkingConfiguration](#)
- [HierarchicalChunkingLevelConfiguration](#)
- [InferenceConfiguration](#)
- [IngestionJob](#)
- [IngestionJobFilter](#)
- [IngestionJobSortBy](#)
- [IngestionJobStatistics](#)
- [IngestionJobSummary](#)
- [InputFlowNodeConfiguration](#)

- [IntermediateStorage](#)
- [IteratorFlowNodeConfiguration](#)
- [KnowledgeBase](#)
- [KnowledgeBaseConfiguration](#)
- [KnowledgeBaseFlowNodeConfiguration](#)
- [KnowledgeBaseSummary](#)
- [LambdaFunctionFlowNodeConfiguration](#)
- [LexFlowNodeConfiguration](#)
- [MemoryConfiguration](#)
- [MongoDbAtlasConfiguration](#)
- [MongoDbAtlasFieldMapping](#)
- [OpenSearchServerlessConfiguration](#)
- [OpenSearchServerlessFieldMapping](#)
- [OutputFlowNodeConfiguration](#)
- [ParameterDetail](#)
- [ParsingConfiguration](#)
- [ParsingPrompt](#)
- [PatternObjectFilter](#)
- [PatternObjectFilterConfiguration](#)
- [PineconeConfiguration](#)
- [PineconeFieldMapping](#)
- [PromptConfiguration](#)
- [PromptFlowNodeConfiguration](#)
- [PromptFlowNodeInlineConfiguration](#)
- [PromptFlowNodeResourceConfiguration](#)
- [PromptFlowNodeSourceConfiguration](#)
- [PromptInferenceConfiguration](#)
- [PromptInputVariable](#)
- [PromptModelInferenceConfiguration](#)
- [PromptOverrideConfiguration](#)

- [PromptSummary](#)
- [PromptTemplateConfiguration](#)
- [PromptVariant](#)
- [RdsConfiguration](#)
- [RdsFieldMapping](#)
- [RedisEnterpriseCloudConfiguration](#)
- [RedisEnterpriseCloudFieldMapping](#)
- [RetrievalFlowNodeConfiguration](#)
- [RetrievalFlowNodeS3Configuration](#)
- [RetrievalFlowNodeServiceConfiguration](#)
- [S3DataSourceConfiguration](#)
- [S3Identifier](#)
- [S3Location](#)
- [SalesforceCrawlerConfiguration](#)
- [SalesforceDataSourceConfiguration](#)
- [SalesforceSourceConfiguration](#)
- [SeedUrl](#)
- [SemanticChunkingConfiguration](#)
- [ServerSideEncryptionConfiguration](#)
- [SharePointCrawlerConfiguration](#)
- [SharePointDataSourceConfiguration](#)
- [SharePointSourceConfiguration](#)
- [StorageConfiguration](#)
- [StorageFlowNodeConfiguration](#)
- [StorageFlowNodeS3Configuration](#)
- [StorageFlowNodeServiceConfiguration](#)
- [TextPromptTemplateConfiguration](#)
- [Transformation](#)
- [TransformationFunction](#)
- [TransformationLambdaConfiguration](#)

- [UrlConfiguration](#)
- [ValidationExceptionField](#)
- [VectorIngestionConfiguration](#)
- [VectorKnowledgeBaseConfiguration](#)
- [WebCrawlerConfiguration](#)
- [WebCrawlerLimits](#)
- [WebDataSourceConfiguration](#)
- [WebSourceConfiguration](#)

ActionGroupExecutor

Service: Agents for Amazon Bedrock

Contains details about the Lambda function containing the business logic that is carried out upon invoking the action or the custom control method for handling the information elicited from the user.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

customControl

To return the action group invocation results directly in the InvokeAgent response, specify RETURN_CONTROL.

Type: String

Valid Values: RETURN_CONTROL

Required: No

lambda

The Amazon Resource Name (ARN) of the Lambda function containing the business logic that is carried out upon invoking the action.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:(aws[a-zA-Z-]*)?:lambda:[a-z]{2}(-gov)?-[a-z]+-\d{1}:\d{12}:function:[a-zA-Z0-9-_\.\.]+(:(\$LATEST|[a-zA-Z0-9-_\.\.]+))?$`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ActionGroupSummary

Service: Agents for Amazon Bedrock

Contains details about an action group.

Contents

actionGroupId

The unique identifier of the action group.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

actionGroupName

The name of the action group.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?) {1,100}$`

Required: Yes

actionGroupState

Specifies whether the action group is available for the agent to invoke or not when sending an [InvokeAgent](#) request.

Type: String

Valid Values: ENABLED | DISABLED

Required: Yes

updatedAt

The time at which the action group was last updated.

Type: Timestamp

Required: Yes

description

The description of the action group.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

Agent

Service: Agents for Amazon Bedrock

Contains details about an agent.

Contents

agentArn

The Amazon Resource Name (ARN) of the agent.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:agent/[0-9a-zA-Z]{10}$`

Required: Yes

agentId

The unique identifier of the agent.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentName

The name of the agent.

Type: String

Pattern: `^([0-9a-zA-Z][_ -?]){1,100}$`

Required: Yes

agentResourceRoleArn

The Amazon Resource Name (ARN) of the IAM role with permissions to invoke API operations on the agent.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:iam::([0-9]{12})?:role/.*+$`

Required: Yes

agentStatus

The status of the agent and whether it is ready for use. The following statuses are possible:

- **CREATING** – The agent is being created.
- **PREPARING** – The agent is being prepared.
- **PREPARED** – The agent is prepared and ready to be invoked.
- **NOT_PREPARED** – The agent has been created but not yet prepared.
- **FAILED** – The agent API operation failed.
- **UPDATING** – The agent is being updated.
- **DELETING** – The agent is being deleted.

Type: String

Valid Values: **CREATING** | **PREPARING** | **PREPARED** | **NOT_PREPARED** | **DELETING** | **FAILED** | **VERSIONING** | **UPDATING**

Required: Yes

agentVersion

The version of the agent.

Type: String

Length Constraints: Fixed length of 5.

Pattern: `^DRAFT$`

Required: Yes

createdAt

The time at which the agent was created.

Type: Timestamp

Required: Yes

idleSessionTTLInSeconds

The number of seconds for which Amazon Bedrock keeps information about a user's conversation with the agent.

A user interaction remains active for the amount of time specified. If no conversation occurs during this time, the session expires and Amazon Bedrock deletes any data provided before the timeout.

Type: Integer

Valid Range: Minimum value of 60. Maximum value of 3600.

Required: Yes

updatedAt

The time at which the agent was last updated.

Type: Timestamp

Required: Yes

clientToken

A unique, case-sensitive identifier to ensure that the API request completes no more than one time. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#).

Type: String

Length Constraints: Minimum length of 33. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,256}$`

Required: No

customerEncryptionKeyArn

The Amazon Resource Name (ARN) of the AWS KMS key that encrypts the agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

Required: No

description

The description of the agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

failureReasons

Contains reasons that the agent-related API that you invoked failed.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 2048 items.

Length Constraints: Minimum length of 0. Maximum length of 2048.

Required: No

foundationModel

The foundation model used for orchestration by the agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}(([:][a-z0-9-]{1,63}){0,2})?/[a-z0-9-]{12})|(:foundation-model/([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2})))|(([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2}))|(([0-9a-zA-Z][_-]?)+)$`

Required: No

guardrailConfiguration

Details about the guardrail associated with the agent.

Type: [GuardrailConfiguration](#) object

Required: No

instruction

Instructions that tell the agent what it should do and how it should interact with users.

Type: String

Length Constraints: Minimum length of 40. Maximum length of 4000.

Required: No

memoryConfiguration

Contains memory configuration for the agent.

Type: [MemoryConfiguration](#) object

Required: No

preparedAt

The time at which the agent was last prepared.

Type: Timestamp

Required: No

promptOverrideConfiguration

Contains configurations to override prompt templates in different parts of an agent sequence. For more information, see [Advanced prompts](#).

Type: [PromptOverrideConfiguration](#) object

Required: No

recommendedActions

Contains recommended actions to take for the agent-related API that you invoked to succeed.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 2048 items.

Length Constraints: Minimum length of 0. Maximum length of 2048.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

AgentActionGroup

Service: Agents for Amazon Bedrock

Contains details about an action group.

Contents

actionGroupId

The unique identifier of the action group.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

actionGroupName

The name of the action group.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

Required: Yes

actionGroupState

Specifies whether the action group is available for the agent to invoke or not when sending an [InvokeAgent](#) request.

Type: String

Valid Values: ENABLED | DISABLED

Required: Yes

agentId

The unique identifier of the agent to which the action group belongs.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentVersion

The version of the agent to which the action group belongs.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: `^(DRAFT|[0-9]{0,4}[1-9][0-9]{0,4})$`

Required: Yes

createdAt

The time at which the action group was created.

Type: Timestamp

Required: Yes

updatedAt

The time at which the action group was last updated.

Type: Timestamp

Required: Yes

actionGroupExecutor

The Amazon Resource Name (ARN) of the Lambda function containing the business logic that is carried out upon invoking the action or the custom control method for handling the information elicited from the user.

Type: [ActionGroupExecutor](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

apiSchema

Contains either details about the S3 object containing the OpenAPI schema for the action group or the JSON or YAML-formatted payload defining the schema. For more information, see [Action group OpenAPI schemas](#).

Type: [APISchema](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

clientToken

A unique, case-sensitive identifier to ensure that the API request completes no more than one time. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#).

Type: String

Length Constraints: Minimum length of 33. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,256}$`

Required: No

description

The description of the action group.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

functionSchema

Defines functions that each define parameters that the agent needs to invoke from the user. Each function represents an action in an action group.

Type: [FunctionSchema](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

parentActionSignature

If this field is set as `AMAZON.UserInput`, the agent can request the user for additional information when trying to complete a task. The `description`, `apiSchema`, and `actionGroupExecutor` fields must be blank for this action group.

During orchestration, if the agent determines that it needs to invoke an API in an action group, but doesn't have enough information to complete the API request, it will invoke this action group instead and return an [Observation](#) reprompting the user for more information.

Type: String

Valid Values: AMAZON.UserInput | AMAZON.CodeInterpreter

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

AgentAlias

Service: Agents for Amazon Bedrock

Contains details about an alias of an agent.

Contents

agentAliasArn

The Amazon Resource Name (ARN) of the alias of the agent.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:agent-alias/[0-9a-zA-Z]{10}/[0-9a-zA-Z]{10}$`

Required: Yes

agentAliasId

The unique identifier of the alias of the agent.

Type: String

Length Constraints: Fixed length of 10.

Pattern: `^(\\bTSTALIASID\\b|[0-9a-zA-Z]+)$`

Required: Yes

agentAliasName

The name of the alias of the agent.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

Required: Yes

agentAliasStatus

The status of the alias of the agent and whether it is ready for use. The following statuses are possible:

- CREATING – The agent alias is being created.
- PREPARED – The agent alias is finished being created or updated and is ready to be invoked.
- FAILED – The agent alias API operation failed.
- UPDATING – The agent alias is being updated.
- DELETING – The agent alias is being deleted.

Type: String

Valid Values: CREATING | PREPARED | FAILED | UPDATING | DELETING

Required: Yes

agentId

The unique identifier of the agent.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

createdAt

The time at which the alias of the agent was created.

Type: Timestamp

Required: Yes

routingConfiguration

Contains details about the routing configuration of the alias.

Type: Array of [AgentAliasRoutingConfigurationListItem](#) objects

Array Members: Minimum number of 0 items. Maximum number of 1 item.

Required: Yes

updatedAt

The time at which the alias was last updated.

Type: Timestamp

Required: Yes

agentAliasHistoryEvents

Contains details about the history of the alias.

Type: Array of [AgentAliasHistoryEvent](#) objects

Array Members: Minimum number of 0 items. Maximum number of 10 items.

Required: No

clientToken

A unique, case-sensitive identifier to ensure that the API request completes no more than one time. If this token matches a previous request, Amazon Bedrock ignores the request, but does not return an error. For more information, see [Ensuring idempotency](#).

Type: String

Length Constraints: Minimum length of 33. Maximum length of 256.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,256}$`

Required: No

description

The description of the alias of the agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

failureReasons

Information on the failure of Provisioned Throughput assigned to an agent alias.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 2048 items.

Length Constraints: Minimum length of 0. Maximum length of 2048.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

AgentAliasHistoryEvent

Service: Agents for Amazon Bedrock

Contains details about the history of the alias.

Contents

endDate

The date that the alias stopped being associated to the version in the `routingConfiguration` object

Type: Timestamp

Required: No

routingConfiguration

Contains details about the version of the agent with which the alias is associated.

Type: Array of [AgentAliasRoutingConfigurationListItem](#) objects

Array Members: Minimum number of 0 items. Maximum number of 1 item.

Required: No

startDate

The date that the alias began being associated to the version in the `routingConfiguration` object.

Type: Timestamp

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

AgentAliasRoutingConfigurationListItem

Service: Agents for Amazon Bedrock

Contains details about the routing configuration of the alias.

Contents

agentVersion

The version of the agent with which the alias is associated.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: `^(DRAFT|[0-9]{0,4}[1-9][0-9]{0,4})$`

Required: No

provisionedThroughput

Information on the Provisioned Throughput assigned to an agent alias.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^((([0-9a-zA-Z][_-]?) {1,63})|(arn:aws(-[^\:]+)? :bedrock:[a-z0-9-] {1,20}:[0-9]{12}:provisioned-model/[a-z0-9]{12})))$`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

AgentAliasSummary

Service: Agents for Amazon Bedrock

Contains details about an alias of an agent.

Contents

agentAliasId

Contains details about

Type: String

Length Constraints: Fixed length of 10.

Pattern: $^(\backslash\text{bTSTALIASID}\backslash\text{b} | [\text{0-9a-zA-Z}]^+)\text{\$}$

Required: Yes

agentAliasName

The name of the alias.

Type: String

Pattern: $^([\text{0-9a-zA-Z}] [_ -]^?)\{1, 100\}\text{\$}$

Required: Yes

agentAliasStatus

The status of the alias.

Type: String

Valid Values: CREATING | PREPARED | FAILED | UPDATING | DELETING

Required: Yes

createdAt

The time at which the alias of the agent was created.

Type: Timestamp

Required: Yes

updatedAt

The time at which the alias was last updated.

Type: Timestamp

Required: Yes

description

The description of the alias.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

routingConfiguration

Contains details about the version of the agent with which the alias is associated.

Type: Array of [AgentAliasRoutingConfigurationListItem](#) objects

Array Members: Minimum number of 0 items. Maximum number of 1 item.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

AgentFlowNodeConfiguration

Service: Agents for Amazon Bedrock

Defines an agent node in your flow. You specify the agent to invoke at this point in the flow. For more information, see [Node types in Amazon Bedrock works](#) in the Amazon Bedrock User Guide.

Contents

agentAliasArn

The Amazon Resource Name (ARN) of the alias of the agent to invoke.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:agent-alias/[0-9a-zA-Z]{10}/[0-9a-zA-Z]{10}$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

AgentKnowledgeBase

Service: Agents for Amazon Bedrock

Contains details about a knowledge base that is associated with an agent.

Contents

agentId

The unique identifier of the agent with which the knowledge base is associated.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentVersion

The version of the agent with which the knowledge base is associated.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: `^(DRAFT|[0-9]{0,4}[1-9][0-9]{0,4})$`

Required: Yes

createdAt

The time at which the association between the agent and the knowledge base was created.

Type: Timestamp

Required: Yes

description

The description of the association between the agent and the knowledge base.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: Yes

knowledgeBaseId

The unique identifier of the association between the agent and the knowledge base.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

knowledgeBaseState

Specifies whether to use the knowledge base or not when sending an [InvokeAgent](#) request.

Type: String

Valid Values: ENABLED | DISABLED

Required: Yes

updatedAt

The time at which the association between the agent and the knowledge base was last updated.

Type: Timestamp

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

AgentKnowledgeBaseSummary

Service: Agents for Amazon Bedrock

Contains details about a knowledge base associated with an agent.

Contents

knowledgeBaseId

The unique identifier of the knowledge base associated with an agent.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

knowledgeBaseState

Specifies whether the agent uses the knowledge base or not when sending an [InvokeAgent](#) request.

Type: String

Valid Values: ENABLED | DISABLED

Required: Yes

updatedAt

The time at which the knowledge base associated with an agent was last updated.

Type: Timestamp

Required: Yes

description

The description of the knowledge base associated with an agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

AgentSummary

Service: Agents for Amazon Bedrock

Contains details about an agent.

Contents

agentId

The unique identifier of the agent.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentName

The name of the agent.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?)\{1,100\}$`

Required: Yes

agentStatus

The status of the agent.

Type: String

Valid Values: CREATING | PREPARING | PREPARED | NOT_PREPARED | DELETING | FAILED | VERSIONING | UPDATING

Required: Yes

updatedAt

The time at which the agent was last updated.

Type: Timestamp

Required: Yes

description

The description of the agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

guardrailConfiguration

Details about the guardrail associated with the agent.

Type: [GuardrailConfiguration](#) object

Required: No

latestAgentVersion

The latest version of the agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: $^(DRAFT|[0-9]{0,4}[1-9][0-9]{0,4})\$$

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

AgentVersion

Service: Agents for Amazon Bedrock

Contains details about a version of an agent.

Contents

agentArn

The Amazon Resource Name (ARN) of the agent that the version belongs to.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:agent/[0-9a-zA-Z]{10}$`

Required: Yes

agentId

The unique identifier of the agent that the version belongs to.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

agentName

The name of the agent that the version belongs to.

Type: String

Pattern: `^([0-9a-zA-Z][_ -?]){1,100}$`

Required: Yes

agentResourceRoleArn

The Amazon Resource Name (ARN) of the IAM role with permissions to invoke API operations on the agent.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:iam::([0-9]{12})?:role/.*+$`

Required: Yes

agentStatus

The status of the agent that the version belongs to.

Type: String

Valid Values: CREATING | PREPARING | PREPARED | NOT_PREPARED | DELETING | FAILED | VERSIONING | UPDATING

Required: Yes

createdAt

The time at which the version was created.

Type: Timestamp

Required: Yes

idleSessionTTLInSeconds

The number of seconds for which Amazon Bedrock keeps information about a user's conversation with the agent.

A user interaction remains active for the amount of time specified. If no conversation occurs during this time, the session expires and Amazon Bedrock deletes any data provided before the timeout.

Type: Integer

Valid Range: Minimum value of 60. Maximum value of 3600.

Required: Yes

updatedAt

The time at which the version was last updated.

Type: Timestamp

Required: Yes

version

The version number.

Type: String

Pattern: `^[0-9]{1,5}$`

Required: Yes

customerEncryptionKeyArn

The Amazon Resource Name (ARN) of the AWS KMS key that encrypts the agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

Required: No

description

The description of the version.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

failureReasons

A list of reasons that the API operation on the version failed.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 2048 items.

Length Constraints: Minimum length of 0. Maximum length of 2048.

Required: No

foundationModel

The foundation model that the version invokes.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}(([:][a-z0-9-]{1,63}){0,2})?/[a-z0-9-]{12})|(:foundation-model/([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2})))|((([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2}))|((([0-9a-zA-Z][_]?)+))$`

Required: No

guardrailConfiguration

Details about the guardrail associated with the agent.

Type: [GuardrailConfiguration](#) object

Required: No

instruction

The instructions provided to the agent.

Type: String

Length Constraints: Minimum length of 40. Maximum length of 4000.

Required: No

memoryConfiguration

Contains details of the memory configuration on the version of the agent.

Type: [MemoryConfiguration](#) object

Required: No

promptOverrideConfiguration

Contains configurations to override prompt templates in different parts of an agent sequence. For more information, see [Advanced prompts](#).

Type: [PromptOverrideConfiguration](#) object

Required: No

recommendedActions

A list of recommended actions to take for the failed API operation on the version to succeed.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 2048 items.

Length Constraints: Minimum length of 0. Maximum length of 2048.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

AgentVersionSummary

Service: Agents for Amazon Bedrock

Contains details about a version of an agent.

Contents

agentName

The name of the agent to which the version belongs.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

Required: Yes

agentStatus

The status of the agent to which the version belongs.

Type: String

Valid Values: CREATING | PREPARING | PREPARED | NOT_PREPARED | DELETING | FAILED | VERSIONING | UPDATING

Required: Yes

agentVersion

The version of the agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: `^(DRAFT|[0-9]{0,4}[1-9][0-9]{0,4})$`

Required: Yes

createdAt

The time at which the version was created.

Type: Timestamp

Required: Yes

updatedAt

The time at which the version was last updated.

Type: Timestamp

Required: Yes

description

The description of the version of the agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

guardrailConfiguration

Details about the guardrail associated with the agent.

Type: [GuardrailConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

APISchema

Service: Agents for Amazon Bedrock

Contains details about the OpenAPI schema for the action group. For more information, see [Action group OpenAPI schemas](#). You can either include the schema directly in the payload field or you can upload it to an S3 bucket and specify the S3 bucket location in the s3 field.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

payload

The JSON or YAML-formatted payload defining the OpenAPI schema for the action group. For more information, see [Action group OpenAPI schemas](#).

Type: String

Required: No

s3

Contains details about the S3 object containing the OpenAPI schema for the action group. For more information, see [Action group OpenAPI schemas](#).

Type: [S3Identifier](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

BedrockEmbeddingModelConfiguration

Service: Agents for Amazon Bedrock

The vector configuration details for the Bedrock embeddings model.

Contents

dimensions

The dimensions details for the vector configuration used on the Bedrock embeddings model.

Type: Integer

Valid Range: Minimum value of 0. Maximum value of 4096.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

BedrockFoundationModelConfiguration

Service: Agents for Amazon Bedrock

Settings for a foundation model used to parse documents for a data source.

Contents

modelArn

The model's ARN.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}::foundation-model/([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]?[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2})$`

Required: Yes

parsingPrompt

Instructions for interpreting the contents of a document.

Type: [ParsingPrompt](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ChunkingConfiguration

Service: Agents for Amazon Bedrock

Details about how to chunk the documents in the data source. A *chunk* refers to an excerpt from a data source that is returned when the knowledge base that it belongs to is queried.

Contents

chunkingStrategy

Knowledge base can split your source data into chunks. A *chunk* refers to an excerpt from a data source that is returned when the knowledge base that it belongs to is queried. You have the following options for chunking your data. If you opt for `NONE`, then you may want to pre-process your files by splitting them up such that each file corresponds to a chunk.

- `FIXED_SIZE` – Amazon Bedrock splits your source data into chunks of the approximate size that you set in the `fixedSizeChunkingConfiguration`.
- `HIERARCHICAL` – Split documents into layers of chunks where the first layer contains large chunks, and the second layer contains smaller chunks derived from the first layer.
- `SEMANTIC` – Split documents into chunks based on groups of similar content derived with natural language processing.
- `NONE` – Amazon Bedrock treats each file as one chunk. If you choose this option, you may want to pre-process your documents by splitting them into separate files.

Type: String

Valid Values: `FIXED_SIZE` | `NONE` | `HIERARCHICAL` | `SEMANTIC`

Required: Yes

fixedSizeChunkingConfiguration

Configurations for when you choose fixed-size chunking. If you set the `chunkingStrategy` as `NONE`, exclude this field.

Type: [FixedSizeChunkingConfiguration](#) object

Required: No

hierarchicalChunkingConfiguration

Settings for hierarchical document chunking for a data source. Hierarchical chunking splits documents into layers of chunks where the first layer contains large chunks, and the second layer contains smaller chunks derived from the first layer.

Type: [HierarchicalChunkingConfiguration](#) object

Required: No

semanticChunkingConfiguration

Settings for semantic document chunking for a data source. Semantic chunking splits a document into smaller documents based on groups of similar content derived from the text with natural language processing.

Type: [SemanticChunkingConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

CollectorFlowNodeConfiguration

Service: Agents for Amazon Bedrock

Defines a collector node in your flow. This node takes an iteration of inputs and consolidates them into an array in the output. For more information, see [Node types in Amazon Bedrock works](#) in the Amazon Bedrock User Guide.

Contents

The members of this exception structure are context-dependent.

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ConditionFlowNodeConfiguration

Service: Agents for Amazon Bedrock

Defines a condition node in your flow. You can specify conditions that determine which node comes next in the flow. For more information, see [Node types in Amazon Bedrock works](#) in the Amazon Bedrock User Guide.

Contents

conditions

An array of conditions. Each member contains the name of a condition and an expression that defines the condition.

Type: Array of [FlowCondition](#) objects

Array Members: Minimum number of 1 item. Maximum number of 5 items.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ConfluenceCrawlerConfiguration

Service: Agents for Amazon Bedrock

The configuration of the Confluence content. For example, configuring specific types of Confluence content.

Contents

filterConfiguration

The configuration of filtering the Confluence content. For example, configuring regular expression patterns to include or exclude certain content.

Type: [CrawlFilterConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ConfluenceDataSourceConfiguration

Service: Agents for Amazon Bedrock

The configuration information to connect to Confluence as your data source.

Contents

sourceConfiguration

The endpoint information to connect to your Confluence data source.

Type: [ConfluenceSourceConfiguration](#) object

Required: Yes

crawlerConfiguration

The configuration of the Confluence content. For example, configuring specific types of Confluence content.

Type: [ConfluenceCrawlerConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ConfluenceSourceConfiguration

Service: Agents for Amazon Bedrock

The endpoint information to connect to your Confluence data source.

Contents

authType

The supported authentication type to authenticate and connect to your Confluence instance.

Type: String

Valid Values: BASIC | OAUTH2_CLIENT_CREDENTIALS

Required: Yes

credentialsSecretArn

The Amazon Resource Name of an AWS Secrets Manager secret that stores your authentication credentials for your Confluence instance URL. For more information on the key-value pairs that must be included in your secret, depending on your authentication type, see [Confluence connection configuration](#).

Type: String

Pattern: `^arn:aws(|-cn|-us-gov):secretsmanager:[a-z0-9-]{1,20}:([0-9]{12}|):secret:[a-zA-Z0-9!/_+=.@-]{1,512}$`

Required: Yes

hostType

The supported host type, whether online/cloud or server/on-premises.

Type: String

Valid Values: SAAS

Required: Yes

hostUrl

The Confluence host URL or instance URL.

Type: String

Pattern: `^https://[A-Za-z0-9][^\s]*$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

CrawlFilterConfiguration

Service: Agents for Amazon Bedrock

The configuration of filtering the data source content. For example, configuring regular expression patterns to include or exclude certain content.

Contents

type

The type of filtering that you want to apply to certain objects or content of the data source. For example, the PATTERN type is regular expression patterns you can apply to filter your content.

Type: String

Valid Values: PATTERN

Required: Yes

patternObjectFilter

The configuration of filtering certain objects or content types of the data source.

Type: [PatternObjectFilterConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

CustomTransformationConfiguration

Service: Agents for Amazon Bedrock

Settings for customizing steps in the data source content ingestion pipeline.

You can configure the data source to process documents with a Lambda function after they are parsed and converted into chunks. When you add a post-chunking transformation, the service stores chunked documents in an S3 bucket and invokes a Lambda function to process them.

To process chunked documents with a Lambda function, define an S3 bucket path for input and output objects, and a transformation that specifies the Lambda function to invoke. You can use the Lambda function to customize how chunks are split, and the metadata for each chunk.

Contents

intermediateStorage

An S3 bucket path for input and output objects.

Type: [IntermediateStorage](#) object

Required: Yes

transformations

A Lambda function that processes documents.

Type: Array of [Transformation](#) objects

Array Members: Fixed number of 1 item.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

DataSource

Service: Agents for Amazon Bedrock

Contains details about a data source.

Contents

createdAt

The time at which the data source was created.

Type: Timestamp

Required: Yes

dataSourceConfiguration

The connection configuration for the data source.

Type: [DataSourceConfiguration](#) object

Required: Yes

dataSourceId

The unique identifier of the data source.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

knowledgeBaseId

The unique identifier of the knowledge base to which the data source belongs.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

name

The name of the data source.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

Required: Yes

status

The status of the data source. The following statuses are possible:

- Available – The data source has been created and is ready for ingestion into the knowledge base.
- Deleting – The data source is being deleted.

Type: String

Valid Values: AVAILABLE | DELETING | DELETE_UNSUCCESSFUL

Required: Yes

updatedAt

The time at which the data source was last updated.

Type: Timestamp

Required: Yes

dataDeletionPolicy

The data deletion policy for the data source.

Type: String

Valid Values: RETAIN | DELETE

Required: No

description

The description of the data source.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

failureReasons

The detailed reasons on the failure to delete a data source.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 2048 items.

Length Constraints: Minimum length of 0. Maximum length of 2048.

Required: No

serverSideEncryptionConfiguration

Contains details about the configuration of the server-side encryption.

Type: [ServerSideEncryptionConfiguration](#) object

Required: No

vectorIngestionConfiguration

Contains details about how to ingest the documents in the data source.

Type: [VectorIngestionConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

DataSourceConfiguration

Service: Agents for Amazon Bedrock

The connection configuration for the data source.

Contents

type

The type of data source.

Type: String

Valid Values: S3 | WEB | CONFLUENCE | SALESFORCE | SHAREPOINT

Required: Yes

confluenceConfiguration

The configuration information to connect to Confluence as your data source.

Note

Confluence data source connector is in preview release and is subject to change.

Type: [ConfluenceDataSourceConfiguration](#) object

Required: No

s3Configuration

The configuration information to connect to Amazon S3 as your data source.

Type: [S3DataSourceConfiguration](#) object

Required: No

salesforceConfiguration

The configuration information to connect to Salesforce as your data source.

Note


Salesforce data source connector is in preview release and is subject to change.

Type: [SalesforceDataSourceConfiguration](#) object

Required: No

sharePointConfiguration

The configuration information to connect to SharePoint as your data source.

 **Note**


SharePoint data source connector is in preview release and is subject to change.

Type: [SharePointDataSourceConfiguration](#) object

Required: No

webConfiguration

The configuration of web URLs to crawl for your data source. You should be authorized to crawl the URLs.

 **Note**

Crawling web URLs as your data source is in preview release and is subject to change.

Type: [WebDataSourceConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

DataSourceSummary

Service: Agents for Amazon Bedrock

Contains details about a data source.

Contents

dataSourceId

The unique identifier of the data source.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

knowledgeBaseId

The unique identifier of the knowledge base to which the data source belongs.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

name

The name of the data source.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?) {1,100}$`

Required: Yes

status

The status of the data source.

Type: String

Valid Values: AVAILABLE | DELETING | DELETE_UNSUCCESSFUL

Required: Yes

updatedAt

The time at which the data source was last updated.

Type: Timestamp

Required: Yes

description

The description of the data source.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

EmbeddingModelConfiguration

Service: Agents for Amazon Bedrock

The configuration details for the embeddings model.

Contents

bedrockEmbeddingModelConfiguration

The vector configuration details on the Bedrock embeddings model.

Type: [BedrockEmbeddingModelConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FixedSizeChunkingConfiguration

Service: Agents for Amazon Bedrock

Configurations for when you choose fixed-size chunking. If you set the chunkingStrategy as NONE, exclude this field.

Contents

maxTokens

The maximum number of tokens to include in a chunk.

Type: Integer

Valid Range: Minimum value of 1.

Required: Yes

overlapPercentage

The percentage of overlap between adjacent chunks of a data source.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 99.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowAliasRoutingConfigurationListItem

Service: Agents for Amazon Bedrock

Contains information about a version that the alias maps to.

Contents

flowVersion

The version that the alias maps to.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: `^(DRAFT|[0-9]{0,4}[1-9][0-9]{0,4})$`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowAliasSummary

Service: Agents for Amazon Bedrock

Contains information about an alias of a flow.

This data type is used in the following API operations:

- [ListFlowAliases response](#)

Contents

arn

The Amazon Resource Name (ARN) of the flow alias.

Type: String

Pattern: `^arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10}/alias/(TSTALIASID|[0-9a-zA-Z]{10})$`

Required: Yes

createdAt

The time at which the alias was created.

Type: Timestamp

Required: Yes

flowId

The unique identifier of the flow.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

id

The unique identifier of the alias of the flow.

Type: String

Pattern: `^(TSTALIASID|[0-9a-zA-Z]{10})$`

Required: Yes

name

The name of the alias.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

Required: Yes

routingConfiguration

A list of configurations about the versions that the alias maps to. Currently, you can only specify one.

Type: Array of [FlowAliasRoutingConfigurationListItem](#) objects

Array Members: Fixed number of 1 item.

Required: Yes

updatedAt

The time at which the alias was last updated.

Type: Timestamp

Required: Yes

description

A description of the alias.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowCondition

Service: Agents for Amazon Bedrock

Defines a condition in the condition node.

Contents

name

A name for the condition that you can reference.

Type: String

Pattern: `^[a-zA-Z]([_]?[0-9a-zA-Z]){1,50}$`

Required: Yes

expression

Defines the condition. You must refer to at least one of the inputs in the condition. For more information, expand the Condition node section in [Node types in prompt flows](#).

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowConditionalConnectionConfiguration

Service: Agents for Amazon Bedrock

The configuration of a connection between a condition node and another node.

Contents

condition

The condition that triggers this connection. For more information about how to write conditions, see the **Condition** node type in the [Node types](#) topic in the Amazon Bedrock User Guide.

Type: String

Pattern: `^[a-zA-Z]([_]?[0-9a-zA-Z]){1,50}$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowConnection

Service: Agents for Amazon Bedrock

Contains information about a connection between two nodes in the flow.

Contents

name

A name for the connection that you can reference.

Type: String

Pattern: `^[a-zA-Z]([_]?[0-9a-zA-Z]){1,100}$`

Required: Yes

source

The node that the connection starts at.

Type: String

Pattern: `^[a-zA-Z]([_]?[0-9a-zA-Z]){1,50}$`

Required: Yes

target

The node that the connection ends at.

Type: String

Pattern: `^[a-zA-Z]([_]?[0-9a-zA-Z]){1,50}$`

Required: Yes

type

Whether the source node that the connection begins from is a condition node (`Conditional`) or not (`Data`).

Type: String

Valid Values: `Data` | `Conditional`

Required: Yes

configuration

The configuration of the connection.

Type: [FlowConnectionConfiguration](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowConnectionConfiguration

Service: Agents for Amazon Bedrock

The configuration of the connection.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

conditional

The configuration of a connection originating from a Condition node.

Type: [FlowConditionalConnectionConfiguration](#) object

Required: No

data

The configuration of a connection originating from a node that isn't a Condition node.

Type: [FlowDataConnectionConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowDataConnectionConfiguration

Service: Agents for Amazon Bedrock

The configuration of a connection originating from a node that isn't a Condition node.

Contents

sourceOutput

The name of the output in the source node that the connection begins from.

Type: String

Pattern: `^[a-zA-Z]([_]?[0-9a-zA-Z]){1,50}$`

Required: Yes

targetInput

The name of the input in the target node that the connection ends at.

Type: String

Pattern: `^[a-zA-Z]([_]?[0-9a-zA-Z]){1,50}$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowDefinition

Service: Agents for Amazon Bedrock

The definition of the nodes and connections between nodes in the flow.

Contents

connections

An array of connection definitions in the flow.

Type: Array of [FlowConnection](#) objects

Array Members: Minimum number of 0 items. Maximum number of 20 items.

Required: No

nodes

An array of node definitions in the flow.

Type: Array of [FlowNode](#) objects

Array Members: Minimum number of 0 items. Maximum number of 20 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowNode

Service: Agents for Amazon Bedrock

Contains configurations about a node in the flow.

Contents

name

A name for the node.

Type: String

Pattern: `^[a-zA-Z]([_]?[0-9a-zA-Z]){1,50}$`

Required: Yes

type

The type of node. This value must match the name of the key that you provide in the configuration you provide in the `FlowNodeConfiguration` field.

Type: String

Valid Values: `Input | Output | KnowledgeBase | Condition | Lex | Prompt | LambdaFunction | Storage | Agent | Retrieval | Iterator | Collector`

Required: Yes

configuration

Contains configurations for the node.

Type: [FlowNodeConfiguration](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

inputs

An array of objects, each of which contains information about an input into the node.

Type: Array of [FlowNodeInput](#) objects

Array Members: Minimum number of 0 items. Maximum number of 5 items.

Required: No

outputs

A list of objects, each of which contains information about an output from the node.

Type: Array of [FlowNodeOutput](#) objects

Array Members: Minimum number of 0 items. Maximum number of 5 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowNodeConfiguration

Service: Agents for Amazon Bedrock

Contains configurations for a node in your flow. For more information, see [Node types in Amazon Bedrock works](#) in the Amazon Bedrock User Guide.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

agent

Contains configurations for an agent node in your flow. Invokes an alias of an agent and returns the response.

Type: [AgentFlowNodeConfiguration](#) object

Required: No

collector

Contains configurations for a collector node in your flow. Collects an iteration of inputs and consolidates them into an array of outputs.

Type: [CollectorFlowNodeConfiguration](#) object

Required: No

condition

Contains configurations for a Condition node in your flow. Defines conditions that lead to different branches of the flow.

Type: [ConditionFlowNodeConfiguration](#) object

Required: No

input

Contains configurations for an input flow node in your flow. The first node in the flow. `inputs` can't be specified for this node.

Type: [InputFlowNodeConfiguration](#) object

Required: No

iterator

Contains configurations for an iterator node in your flow. Takes an input that is an array and iteratively sends each item of the array as an output to the following node. The size of the array is also returned in the output.

The output flow node at the end of the flow iteration will return a response for each member of the array. To return only one response, you can include a collector node downstream from the iterator node.

Type: [IteratorFlowNodeConfiguration](#) object

Required: No

knowledgeBase

Contains configurations for a knowledge base node in your flow. Queries a knowledge base and returns the retrieved results or generated response.

Type: [KnowledgeBaseFlowNodeConfiguration](#) object

Required: No

lambdaFunction

Contains configurations for a Lambda function node in your flow. Invokes an AWS Lambda function.

Type: [LambdaFunctionFlowNodeConfiguration](#) object

Required: No

lex

Contains configurations for a Lex node in your flow. Invokes an Amazon Lex bot to identify the intent of the input and return the intent as the output.

Type: [LexFlowNodeConfiguration](#) object

Required: No

output

Contains configurations for an output flow node in your flow. The last node in the flow. outputs can't be specified for this node.

Type: [OutputFlowNodeConfiguration](#) object

Required: No

prompt

Contains configurations for a prompt node in your flow. Runs a prompt and generates the model response as the output. You can use a prompt from Prompt management or you can configure one in this node.

Type: [PromptFlowNodeConfiguration](#) object

Required: No

retrieval

Contains configurations for a Retrieval node in your flow. Retrieves data from an Amazon S3 location and returns it as the output.

Type: [RetrievalFlowNodeConfiguration](#) object

Required: No

storage

Contains configurations for a Storage node in your flow. Stores an input in an Amazon S3 location.

Type: [StorageFlowNodeConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)

- [AWS SDK for Ruby V3](#)

FlowNodeInput

Service: Agents for Amazon Bedrock

Contains configurations for an input to a node.

Contents

expression

An expression that formats the input for the node. For an explanation of how to create expressions, see [Expressions in Prompt flows in Amazon Bedrock](#).

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Required: Yes

name

A name for the input that you can reference.

Type: String

Pattern: `^[a-zA-Z]([_]?[0-9a-zA-Z]){1,50}$`

Required: Yes

type

The data type of the input. If the input doesn't match this type at runtime, a validation error will be thrown.

Type: String

Valid Values: `String` | `Number` | `Boolean` | `Object` | `Array`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowNodeOutput

Service: Agents for Amazon Bedrock

Contains configurations for an output from a node.

Contents

name

A name for the output that you can reference.

Type: String

Pattern: `^[a-zA-Z]([_]?[0-9a-zA-Z]){1,50}$`

Required: Yes

type

The data type of the output. If the output doesn't match this type at runtime, a validation error will be thrown.

Type: String

Valid Values: `String` | `Number` | `Boolean` | `Object` | `Array`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowSummary

Service: Agents for Amazon Bedrock

Contains the definition of a flow.

Contents

arn

The Amazon Resource Name (ARN) of the flow.

Type: String

Pattern: `^arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10}$`

Required: Yes

createdAt

The time at which the flow was created.

Type: Timestamp

Required: Yes

id

The unique identifier of the flow.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

name

The name of the flow.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?){1,100}$`

Required: Yes

status

The status of the flow. The following statuses are possible:

- **NotPrepared** – The flow has been created or updated, but hasn't been prepared. If you just created the flow, you can't test it. If you updated the flow, the DRAFT version won't contain the latest changes for testing. Send a [PrepareFlow](#) request to package the latest changes into the DRAFT version.
- **Preparing** – The flow is being prepared so that the DRAFT version contains the latest changes for testing.
- **Prepared** – The flow is prepared and the DRAFT version contains the latest changes for testing.
- **Failed** – The last API operation that you invoked on the flow failed. Send a [GetFlow](#) request and check the error message in the `validations` field.

Type: String

Valid Values: Failed | Prepared | Preparing | NotPrepared

Required: Yes

updatedAt

The time at which the flow was last updated.

Type: Timestamp

Required: Yes

version

The latest version of the flow.

Type: String

Length Constraints: Fixed length of 5.

Pattern: ^DRAFT\$

Required: Yes

description

A description of the flow.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowValidation

Service: Agents for Amazon Bedrock

Contains information about validation of the flow.

This data type is used in the following API operations:

- [GetFlow response](#)
- [GetFlowVersion response](#)

Contents

message

A message describing the validation error.

Type: String

Pattern: `^\s\S]+$`

Required: Yes

severity

The severity of the issue described in the message.

Type: String

Valid Values: `Warning` | `Error`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowVersionSummary

Service: Agents for Amazon Bedrock

Contains information about the flow version.

This data type is used in the following API operations:

- [ListFlowVersions response](#)

Contents

arn

The Amazon Resource Name (ARN) of the flow that the version belongs to.

Type: String

Pattern: `^arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:flow/[0-9a-zA-Z]{10}$`

Required: Yes

createdAt

The time at the flow version was created.

Type: Timestamp

Required: Yes

id

The unique identifier of the flow.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

status

The status of the flow.

Type: String

Valid Values: Failed | Prepared | Preparing | NotPrepared

Required: Yes

version

The version of the flow.

Type: String

Pattern: `^[0-9]{1,5}$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

Function

Service: Agents for Amazon Bedrock

Defines parameters that the agent needs to invoke from the user to complete the function. Corresponds to an action in an action group.

This data type is used in the following API operations:

- [CreateAgentActionGroup request](#)
- [CreateAgentActionGroup response](#)
- [UpdateAgentActionGroup request](#)
- [UpdateAgentActionGroup response](#)
- [GetAgentActionGroup response](#)

Contents

name

A name for the function.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?)\{1,100\}$`

Required: Yes

description

A description of the function and its purpose.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1200.

Required: No

parameters

The parameters that the agent elicits from the user to fulfill the function.

Type: String to [ParameterDetail](#) object map

Key Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FunctionSchema

Service: Agents for Amazon Bedrock

Defines functions that each define parameters that the agent needs to invoke from the user. Each function represents an action in an action group.

This data type is used in the following API operations:

- [CreateAgentActionGroup request](#)
- [CreateAgentActionGroup response](#)
- [UpdateAgentActionGroup request](#)
- [UpdateAgentActionGroup response](#)
- [GetAgentActionGroup response](#)

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

functions

A list of functions that each define an action in the action group.

Type: Array of [Function](#) objects

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailConfiguration

Service: Agents for Amazon Bedrock

Details about the guardrail associated with an agent.

Contents

guardrailIdentifier

The unique identifier of the guardrail.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^(([a-z0-9]+)|(arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:guardrail/[a-z0-9]+))$`

Required: No

guardrailVersion

The version of the guardrail.

Type: String

Pattern: `^(([0-9]{1,8})|(DRAFT))$`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

HierarchicalChunkingConfiguration

Service: Agents for Amazon Bedrock

Settings for hierarchical document chunking for a data source. Hierarchical chunking splits documents into layers of chunks where the first layer contains large chunks, and the second layer contains smaller chunks derived from the first layer.

You configure the number of tokens to overlap, or repeat across adjacent chunks. For example, if you set overlap tokens to 60, the last 60 tokens in the first chunk are also included at the beginning of the second chunk. For each layer, you must also configure the maximum number of tokens in a chunk.

Contents

levelConfigurations

Token settings for each layer.

Type: Array of [HierarchicalChunkingLevelConfiguration](#) objects

Array Members: Fixed number of 2 items.

Required: Yes

overlapTokens

The number of tokens to repeat across chunks in the same layer.

Type: Integer

Valid Range: Minimum value of 1.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

HierarchicalChunkingLevelConfiguration

Service: Agents for Amazon Bedrock

Token settings for a layer in a hierarchical chunking configuration.

Contents

maxTokens

The maximum number of tokens that a chunk can contain in this layer.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 8192.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

InferenceConfiguration

Service: Agents for Amazon Bedrock

Contains inference parameters to use when the agent invokes a foundation model in the part of the agent sequence defined by the `promptType`. For more information, see [Inference parameters for foundation models](#).

Contents

maxLength

The maximum number of tokens to allow in the generated response.

Type: Integer

Valid Range: Minimum value of 0. Maximum value of 4096.

Required: No

stopSequences

A list of stop sequences. A stop sequence is a sequence of characters that causes the model to stop generating the response.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 4 items.

Required: No

temperature

The likelihood of the model selecting higher-probability options while generating a response. A lower value makes the model more likely to choose higher-probability options, while a higher value makes the model more likely to choose lower-probability options.

Type: Float

Valid Range: Minimum value of 0. Maximum value of 1.

Required: No

topK

While generating a response, the model determines the probability of the following token at each point of generation. The value that you set for `topK` is the number of most-likely

candidates from which the model chooses the next token in the sequence. For example, if you set `topK` to 50, the model selects the next token from among the top 50 most likely choices.

Type: Integer

Valid Range: Minimum value of 0. Maximum value of 500.

Required: No

topP

While generating a response, the model determines the probability of the following token at each point of generation. The value that you set for `Top P` determines the number of most-likely candidates from which the model chooses the next token in the sequence. For example, if you set `topP` to 80, the model only selects the next token from the top 80% of the probability distribution of next tokens.

Type: Float

Valid Range: Minimum value of 0. Maximum value of 1.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

IngestionJob

Service: Agents for Amazon Bedrock

Contains details about an ingestion job, which converts a data source to embeddings for a vector store in knowledge base.

This data type is used in the following API operations:

- [StartIngestionJob response](#)
- [GetIngestionJob response](#)
- [ListIngestionJob response](#)

Contents

dataSourceId

The unique identifier of the ingested data source.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

ingestionJobId

The unique identifier of the ingestion job.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

knowledgeBaseId

The unique identifier of the knowledge base to which the data source is being added.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

startedAt

The time at which the ingestion job started.

Type: Timestamp

Required: Yes

status

The status of the ingestion job.

Type: String

Valid Values: STARTING | IN_PROGRESS | COMPLETE | FAILED

Required: Yes

updatedAt

The time at which the ingestion job was last updated.

Type: Timestamp

Required: Yes

description

The description of the ingestion job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

failureReasons

A list of reasons that the ingestion job failed.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 2048 items.

Length Constraints: Minimum length of 0. Maximum length of 2048.

Required: No

statistics

Contains statistics about the ingestion job.

Type: [IngestionJobStatistics](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

IngestionJobFilter

Service: Agents for Amazon Bedrock

Defines a filter by which to filter the results.

Contents

attribute

The attribute by which to filter the results.

Type: String

Valid Values: STATUS

Required: Yes

operator

The operation to carry out between the attribute and the values.

Type: String

Valid Values: EQ

Required: Yes

values

A list of values for the attribute.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 10 items.

Length Constraints: Minimum length of 0. Maximum length of 100.

Pattern: `^.*$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

IngestionJobSortBy

Service: Agents for Amazon Bedrock

Parameters by which to sort the results.

Contents

attribute

The attribute by which to sort the results.

Type: String

Valid Values: STATUS | STARTED_AT

Required: Yes

order

The order by which to sort the results.

Type: String

Valid Values: ASCENDING | DESCENDING

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

IngestionJobStatistics

Service: Agents for Amazon Bedrock

Contains the statistics for the ingestion job.

Contents

numberOfDocumentsDeleted

The number of source documents that was deleted.

Type: Long

Required: No

numberOfDocumentsFailed

The number of source documents that failed to be ingested.

Type: Long

Required: No

numberOfDocumentsScanned

The total number of source documents that were scanned. Includes new, updated, and unchanged documents.

Type: Long

Required: No

numberOfMetadataDocumentsModified

The number of metadata files that were updated or deleted.

Type: Long

Required: No

numberOfMetadataDocumentsScanned

The total number of metadata files that were scanned. Includes new, updated, and unchanged files.

Type: Long

Required: No

numberOfModifiedDocumentsIndexed

The number of modified source documents in the data source that were successfully indexed.

Type: Long

Required: No

numberOfNewDocumentsIndexed

The number of new source documents in the data source that were successfully indexed.

Type: Long

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

IngestionJobSummary

Service: Agents for Amazon Bedrock

Contains details about an ingestion job.

Contents

dataSourceId

The unique identifier of the data source in the ingestion job.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

ingestionJobId

The unique identifier of the ingestion job.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

knowledgeBaseId

The unique identifier of the knowledge base to which the data source is added.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

startedAt

The time at which the ingestion job was started.

Type: Timestamp

Required: Yes

status

The status of the ingestion job.

Type: String

Valid Values: STARTING | IN_PROGRESS | COMPLETE | FAILED

Required: Yes

updatedAt

The time at which the ingestion job was last updated.

Type: Timestamp

Required: Yes

description

The description of the ingestion job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

statistics

Contains statistics for the ingestion job.

Type: [IngestionJobStatistics](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

InputFlowNodeConfiguration

Service: Agents for Amazon Bedrock

Contains configurations for the input flow node for a flow. This node takes the input from flow invocation and passes it to the next node in the data type that you specify.

Contents

The members of this exception structure are context-dependent.

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

IntermediateStorage

Service: Agents for Amazon Bedrock

A location for storing content from data sources temporarily as it is processed by custom components in the ingestion pipeline.

Contents

s3Location

An S3 bucket path.

Type: [S3Location](#) object

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

IteratorFlowNodeConfiguration

Service: Agents for Amazon Bedrock

Contains configurations for an iterator node in a flow. Takes an input that is an array and iteratively sends each item of the array as an output to the following node. The size of the array is also returned in the output.

The output flow node at the end of the flow iteration will return a response for each member of the array. To return only one response, you can include a collector node downstream from the iterator node.

Contents

The members of this exception structure are context-dependent.

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

KnowledgeBase

Service: Agents for Amazon Bedrock

Contains information about a knowledge base.

Contents

createdAt

The time at which the knowledge base was created.

Type: Timestamp

Required: Yes

knowledgeBaseArn

The Amazon Resource Name (ARN) of the knowledge base.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 128.

Pattern: `^arn:aws(|-cn|-us-gov):bedrock:[a-zA-Z0-9-]*:[0-9]{12}:knowledge-base/[0-9a-zA-Z]+$`

Required: Yes

knowledgeBaseConfiguration

Contains details about the embeddings configuration of the knowledge base.

Type: [KnowledgeBaseConfiguration](#) object

Required: Yes

knowledgeBaseId

The unique identifier of the knowledge base.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

name

The name of the knowledge base.

Type: String

Pattern: `^([\u0000-9a-zA-Z][_-]?){1,100}$`

Required: Yes

roleArn

The Amazon Resource Name (ARN) of the IAM role with permissions to invoke API operations on the knowledge base.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:iam::([\u0000-9]{12})?:role/.*+$`

Required: Yes

status

The status of the knowledge base. The following statuses are possible:

- **CREATING** – The knowledge base is being created.
- **ACTIVE** – The knowledge base is ready to be queried.
- **DELETING** – The knowledge base is being deleted.
- **UPDATING** – The knowledge base is being updated.
- **FAILED** – The knowledge base API operation failed.

Type: String

Valid Values: **CREATING** | **ACTIVE** | **DELETING** | **UPDATING** | **FAILED** | **DELETE_UNSUCCESSFUL**

Required: Yes

storageConfiguration

Contains details about the storage configuration of the knowledge base.

Type: [StorageConfiguration](#) object

Required: Yes

updatedAt

The time at which the knowledge base was last updated.

Type: Timestamp

Required: Yes

description

The description of the knowledge base.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

failureReasons

A list of reasons that the API operation on the knowledge base failed.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 2048 items.

Length Constraints: Minimum length of 0. Maximum length of 2048.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

KnowledgeBaseConfiguration

Service: Agents for Amazon Bedrock

Contains details about the embeddings configuration of the knowledge base.

Contents

type

The type of data that the data source is converted into for the knowledge base.

Type: String

Valid Values: VECTOR

Required: Yes

vectorKnowledgeBaseConfiguration

Contains details about the embeddings model that's used to convert the data source.

Type: [VectorKnowledgeBaseConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

KnowledgeBaseFlowNodeConfiguration

Service: Agents for Amazon Bedrock

Contains configurations for a knowledge base node in a flow. This node takes a query as the input and returns, as the output, the retrieved responses directly (as an array) or a response generated based on the retrieved responses. For more information, see [Node types in Amazon Bedrock works](#) in the Amazon Bedrock User Guide.

Contents

knowledgeBaseId

The unique identifier of the knowledge base to query.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10.

Pattern: `^[0-9a-zA-Z]+$`

Required: Yes

modelId

The unique identifier of the model to use to generate a response from the query results. Omit this field if you want to return the retrieved results as an array.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}(([:][a-z0-9-]{1,63}){0,2})?/[a-z0-9]{12})|(:foundation-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]{1}[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2})))|(([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]{1}[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2})))|(([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.]{1}[a-z0-9-]{1,63})([:][a-z0-9-]{1,63}){0,2})))|(([0-9a-zA-Z][_]?)+)$`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

KnowledgeBaseSummary

Service: Agents for Amazon Bedrock

Contains details about a knowledge base.

Contents

knowledgeBaseId

The unique identifier of the knowledge base.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

name

The name of the knowledge base.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?) {1,100}$`

Required: Yes

status

The status of the knowledge base.

Type: String

Valid Values: CREATING | ACTIVE | DELETING | UPDATING | FAILED | DELETE_UNSUCCESSFUL

Required: Yes

updatedAt

The time at which the knowledge base was last updated.

Type: Timestamp

Required: Yes

description

The description of the knowledge base.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

LambdaFunctionFlowNodeConfiguration

Service: Agents for Amazon Bedrock

Contains configurations for a Lambda function node in the flow. You specify the Lambda function to invoke and the inputs into the function. The output is the response that is defined in the Lambda function. For more information, see [Node types in Amazon Bedrock works](#) in the Amazon Bedrock User Guide.

Contents

lambdaArn

The Amazon Resource Name (ARN) of the Lambda function to invoke.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:(aws[a-zA-Z-]*)?:lambda:[a-z]{2}(-gov)?-[a-z]+\-\d{1}:\d{12}:function:[a-zA-Z0-9-_\.\.]+(:(\$\{LATEST|[a-zA-Z0-9-_\.\.]+))?$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

LexFlowNodeConfiguration

Service: Agents for Amazon Bedrock

Contains configurations for a Lex node in the flow. You specify a Amazon Lex bot to invoke. This node takes an utterance as the input and returns as the output the intent identified by the Amazon Lex bot. For more information, see [Node types in Amazon Bedrock works](#) in the Amazon Bedrock User Guide.

Contents

botAliasArn

The Amazon Resource Name (ARN) of the Amazon Lex bot alias to invoke.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 78.

Pattern: `^arn:aws(|-us-gov):lex:[a-z]{2}(-gov)?-[a-z]+-\d{1}:\d{12}:bot-alias/[0-9a-zA-Z]+/[0-9a-zA-Z]+$`

Required: Yes

localeId

The Region to invoke the Amazon Lex bot in.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 10.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

MemoryConfiguration

Service: Agents for Amazon Bedrock

Details of the memory configuration.

Contents

enabledMemoryTypes

The type of memory that is stored.

Type: Array of strings

Array Members: Fixed number of 1 item.

Valid Values: SESSION_SUMMARY

Required: Yes

storageDays

The number of days the agent is configured to retain the conversational context.

Type: Integer

Valid Range: Minimum value of 0. Maximum value of 30.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

MongoDbAtlasConfiguration

Service: Agents for Amazon Bedrock

Contains details about the storage configuration of the knowledge base in MongoDB Atlas.

Contents

collectionName

The collection name of the knowledge base in MongoDB Atlas.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 63.

Pattern: `^.*$`

Required: Yes

credentialsSecretArn

The Amazon Resource Name (ARN) of the secret that you created in AWS Secrets Manager that contains user credentials for your MongoDB Atlas cluster.

Type: String

Pattern: `^arn:aws(|-cn|-us-gov):secretsmanager:[a-z0-9-]{1,20}:([0-9]{12}|):secret:[a-zA-Z0-9!/_+=.@-]{1,512}$`

Required: Yes

databaseName

The database name in your MongoDB Atlas cluster for your knowledge base.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 63.

Pattern: `^.*$`

Required: Yes

endpoint

The endpoint URL of your MongoDB Atlas cluster for your knowledge base.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^.*$`

Required: Yes

fieldMapping

Contains the names of the fields to which to map information about the vector store.

Type: [MongoDbAtlasFieldMapping](#) object

Required: Yes

vectorIndexName

The name of the MongoDB Atlas vector search index.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^.*$`

Required: Yes

endpointServiceName

The name of the VPC endpoint service in your account that is connected to your MongoDB Atlas cluster.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 255.

Pattern: `^(?:arn:aws(?:-us-gov|-cn|-iso|-iso-[a-z])*:.*:\d+.*(?:/.*$|[a-zA-Z0-9*]+[a-zA-Z0-9._-]*)$`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

MongoDbAtlasFieldMapping

Service: Agents for Amazon Bedrock

Contains the names of the fields to which to map information about the vector store.

Contents

metadataField

The name of the field in which Amazon Bedrock stores metadata about the vector store.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^.*$`

Required: Yes

textField

The name of the field in which Amazon Bedrock stores the raw text from your data. The text is split according to the chunking strategy you choose.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^.*$`

Required: Yes

vectorField

The name of the field in which Amazon Bedrock stores the vector embeddings for your data sources.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^.*$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

OpenSearchServerlessConfiguration

Service: Agents for Amazon Bedrock

Contains details about the storage configuration of the knowledge base in Amazon OpenSearch Service. For more information, see [Create a vector index in Amazon OpenSearch Service](#).

Contents

collectionArn

The Amazon Resource Name (ARN) of the OpenSearch Service vector store.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:aws:aoss:[a-z]{2}(-gov)?-[a-z]+\-\d{1}:\d{12}:collection/[a-z0-9-]{3,32}$`

Required: Yes

fieldMapping

Contains the names of the fields to which to map information about the vector store.

Type: [OpenSearchServerlessFieldMapping](#) object

Required: Yes

vectorIndexName

The name of the vector store.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^\.*$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

OpenSearchServerlessFieldMapping

Service: Agents for Amazon Bedrock

Contains the names of the fields to which to map information about the vector store.

Contents

metadataField

The name of the field in which Amazon Bedrock stores metadata about the vector store.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^.*$`

Required: Yes

textField

The name of the field in which Amazon Bedrock stores the raw text from your data. The text is split according to the chunking strategy you choose.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^.*$`

Required: Yes

vectorField

The name of the field in which Amazon Bedrock stores the vector embeddings for your data sources.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^.*$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

OutputFlowNodeConfiguration

Service: Agents for Amazon Bedrock

Contains configurations for an output flow node in the flow. You specify the data type expected for the input into the node in the `type` field and how to return the final output in the `expression` field.

Contents

The members of this exception structure are context-dependent.

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ParameterDetail

Service: Agents for Amazon Bedrock

Contains details about a parameter in a function for an action group.

This data type is used in the following API operations:

- [CreateAgentActionGroup request](#)
- [CreateAgentActionGroup response](#)
- [UpdateAgentActionGroup request](#)
- [UpdateAgentActionGroup response](#)
- [GetAgentActionGroup response](#)

Contents

type

The data type of the parameter.

Type: String

Valid Values: `string` | `number` | `integer` | `boolean` | `array`

Required: Yes

description

A description of the parameter. Helps the foundation model determine how to elicit the parameters from the user.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 500.

Required: No

required

Whether the parameter is required for the agent to complete the function for action group invocation.

Type: Boolean

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ParsingConfiguration

Service: Agents for Amazon Bedrock

Settings for parsing document contents. By default, the service converts the contents of each document into text before splitting it into chunks. To improve processing of PDF files with tables and images, you can configure the data source to convert the pages of text into images and use a model to describe the contents of each page.

To use a model to parse PDF documents, set the parsing strategy to `BEDROCK_FOUNDATION_MODEL` and specify the model to use by ARN. You can also override the default parsing prompt with instructions for how to interpret images and tables in your documents. The following models are supported.

- Anthropic Claude 3 Sonnet - `anthropic.claude-3-sonnet-20240229-v1:0`
- Anthropic Claude 3 Haiku - `anthropic.claude-3-haiku-20240307-v1:0`

You can get the ARN of a model with the [ListFoundationModels](#) action. Standard model usage charges apply for the foundation model parsing strategy.

Contents

parsingStrategy

The parsing strategy for the data source.

Type: String

Valid Values: `BEDROCK_FOUNDATION_MODEL`

Required: Yes

bedrockFoundationModelConfiguration

Settings for a foundation model used to parse documents for a data source.

Type: [BedrockFoundationModelConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ParsingPrompt

Service: Agents for Amazon Bedrock

Instructions for interpreting the contents of a document.

Contents

`parsingPromptText`

Instructions for interpreting the contents of a document.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 10000.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PatternObjectFilter

Service: Agents for Amazon Bedrock

The specific filters applied to your data source content. You can filter out or include certain content.

Contents

objectType

The supported object type or content type of the data source.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 50.

Required: Yes

exclusionFilters

A list of one or more exclusion regular expression patterns to exclude certain object types that adhere to the pattern. If you specify an inclusion and exclusion filter/pattern and both match a document, the exclusion filter takes precedence and the document isn't crawled.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 25 items.

Length Constraints: Minimum length of 1. Maximum length of 1000.

Required: No

inclusionFilters

A list of one or more inclusion regular expression patterns to include certain object types that adhere to the pattern. If you specify an inclusion and exclusion filter/pattern and both match a document, the exclusion filter takes precedence and the document isn't crawled.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 25 items.

Length Constraints: Minimum length of 1. Maximum length of 1000.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PatternObjectFilterConfiguration

Service: Agents for Amazon Bedrock

The configuration of filtering certain objects or content types of the data source.

Contents

filters

The configuration of specific filters applied to your data source content. You can filter out or include certain content.

Type: Array of [PatternObjectFilter](#) objects

Array Members: Minimum number of 1 item. Maximum number of 25 items.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PineconeConfiguration

Service: Agents for Amazon Bedrock

Contains details about the storage configuration of the knowledge base in Pinecone. For more information, see [Create a vector index in Pinecone](#).

Contents

connectionString

The endpoint URL for your index management page.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^.*$`

Required: Yes

credentialsSecretArn

The Amazon Resource Name (ARN) of the secret that you created in AWS Secrets Manager that is linked to your Pinecone API key.

Type: String

Pattern: `^arn:aws(|-cn|-us-gov):secretsmanager:[a-z0-9-]{1,20}:([0-9]{12}|):secret:[a-zA-Z0-9!/_+=.@-]{1,512}$`

Required: Yes

fieldMapping

Contains the names of the fields to which to map information about the vector store.

Type: [PineconeFieldMapping](#) object

Required: Yes

namespace

The namespace to be used to write new data to your database.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^.*$`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PineconeFieldMapping

Service: Agents for Amazon Bedrock

Contains the names of the fields to which to map information about the vector store.

Contents

metadataField

The name of the field in which Amazon Bedrock stores metadata about the vector store.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^.*$`

Required: Yes

textField

The name of the field in which Amazon Bedrock stores the raw text from your data. The text is split according to the chunking strategy you choose.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^.*$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PromptConfiguration

Service: Agents for Amazon Bedrock

Contains configurations to override a prompt template in one part of an agent sequence. For more information, see [Advanced prompts](#).

Contents

basePromptTemplate

Defines the prompt template with which to replace the default prompt template. You can use placeholder variables in the base prompt template to customize the prompt. For more information, see [Prompt template placeholder variables](#). For more information, see [Configure the prompt templates](#).

Type: String

Length Constraints: Minimum length of 1. Maximum length of 100000.

Required: No

inferenceConfiguration

Contains inference parameters to use when the agent invokes a foundation model in the part of the agent sequence defined by the promptType. For more information, see [Inference parameters for foundation models](#).

Type: [InferenceConfiguration](#) object

Required: No

parserMode

Specifies whether to override the default parser Lambda function when parsing the raw foundation model output in the part of the agent sequence defined by the promptType. If you set the field as OVERRIDEN, the `overrideLambda` field in the [PromptOverrideConfiguration](#) must be specified with the ARN of a Lambda function.

Type: String

Valid Values: DEFAULT | OVERRIDDEN

Required: No

promptCreationMode

Specifies whether to override the default prompt template for this promptType. Set this value to `OVERRIDDEN` to use the prompt that you provide in the `basePromptTemplate`. If you leave it as `DEFAULT`, the agent uses a default prompt template.

Type: String

Valid Values: `DEFAULT` | `OVERRIDDEN`

Required: No

promptState

Specifies whether to allow the agent to carry out the step specified in the promptType. If you set this value to `DISABLED`, the agent skips that step. The default state for each promptType is as follows.

- `PRE_PROCESSING` – `ENABLED`
- `ORCHESTRATION` – `ENABLED`
- `KNOWLEDGE_BASE_RESPONSE_GENERATION` – `ENABLED`
- `POST_PROCESSING` – `DISABLED`

Type: String

Valid Values: `ENABLED` | `DISABLED`

Required: No

promptType

The step in the agent sequence that this prompt configuration applies to.

Type: String

Valid Values: `PRE_PROCESSING` | `ORCHESTRATION` | `POST_PROCESSING` | `KNOWLEDGE_BASE_RESPONSE_GENERATION`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PromptFlowNodeConfiguration

Service: Agents for Amazon Bedrock

Contains configurations for a prompt node in the flow. You can use a prompt from Prompt management or you can define one in this node. If the prompt contains variables, the inputs into this node will fill in the variables. The output from this node is the response generated by the model. For more information, see [Node types in Amazon Bedrock works](#) in the Amazon Bedrock User Guide.

Contents

sourceConfiguration

Specifies whether the prompt is from Prompt management or defined inline.

Type: [PromptFlowNodeSourceConfiguration](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PromptFlowNodeInlineConfiguration

Service: Agents for Amazon Bedrock

Contains configurations for a prompt defined inline in the node.

Contents

modelId

The unique identifier of the model to run inference with.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^(arn:aws(-[^\:]+)?):bedrock:[a-z0-9-]{1,20}:((([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}/[a-z0-9]{12}))|(:foundation-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.:]?[a-z0-9-]{1,63}))|([0-9]{12}:provisioned-model/[a-z0-9]{12})))|([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.:]?[a-z0-9-]{1,63}))|(([0-9a-zA-Z][_]?)+)$`

Required: Yes

templateConfiguration

Contains a prompt and variables in the prompt that can be replaced with values at runtime.

Type: [PromptTemplateConfiguration](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: Yes

templateType

The type of prompt template.

Type: String

Valid Values: TEXT

Required: Yes

inferenceConfiguration

Contains inference configurations for the prompt.

Type: [PromptInferenceConfiguration](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PromptFlowNodeResourceConfiguration

Service: Agents for Amazon Bedrock

Contains configurations for a prompt from Prompt management to use in a node.

Contents

promptArn

The Amazon Resource Name (ARN) of the prompt from Prompt management.

Type: String

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:prompt/[0-9a-zA-Z]{10}(?::[0-9]{1,5})?)$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PromptFlowNodeSourceConfiguration

Service: Agents for Amazon Bedrock

Contains configurations for a prompt and whether it is from Prompt management or defined inline.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

inline

Contains configurations for a prompt that is defined inline

Type: [PromptFlowNodeInlineConfiguration](#) object

Required: No

resource

Contains configurations for a prompt from Prompt management.

Type: [PromptFlowNodeResourceConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PromptInferenceConfiguration

Service: Agents for Amazon Bedrock

Contains inference configurations for the prompt.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

text

Contains inference configurations for a text prompt.

Type: [PromptModelInferenceConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PromptInputVariable

Service: Agents for Amazon Bedrock

Contains information about a variable in the prompt.

Contents

name

The name of the variable.

Type: String

Pattern: `^([0-9a-zA-Z][_]?){1,100}$`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PromptModelInferenceConfiguration

Service: Agents for Amazon Bedrock

Contains inference configurations related to model inference for a prompt. For more information, see [Inference parameters](#).

Contents

maxTokens

The maximum number of tokens to return in the response.

Type: Integer

Valid Range: Minimum value of 0. Maximum value of 4096.

Required: No

stopSequences

A list of strings that define sequences after which the model will stop generating.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 4 items.

Required: No

temperature

Controls the randomness of the response. Choose a lower value for more predictable outputs and a higher value for more surprising outputs.

Type: Float

Valid Range: Minimum value of 0. Maximum value of 1.

Required: No

topK

The number of most-likely candidates that the model considers for the next token during generation.

Type: Integer

Valid Range: Minimum value of 0. Maximum value of 500.

Required: No

topP

The percentage of most-likely candidates that the model considers for the next token.

Type: Float

Valid Range: Minimum value of 0. Maximum value of 1.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PromptOverrideConfiguration

Service: Agents for Amazon Bedrock

Contains configurations to override prompts in different parts of an agent sequence. For more information, see [Advanced prompts](#).

Contents

promptConfigurations

Contains configurations to override a prompt template in one part of an agent sequence. For more information, see [Advanced prompts](#).

Type: Array of [PromptConfiguration](#) objects

Array Members: Minimum number of 0 items. Maximum number of 10 items.

Required: Yes

overrideLambda

The ARN of the Lambda function to use when parsing the raw foundation model output in parts of the agent sequence. If you specify this field, at least one of the `promptConfigurations` must contain a `parserMode` value that is set to `OVERRIDDEN`. For more information, see [Parser Lambda function in Agents for Amazon Bedrock](#).

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:(aws[a-zA-Z-]*)?:lambda:[a-z]{2}(-gov)?-[a-z]+\d{1}:\d{12}:function:[a-zA-Z0-9-_\.\.]+(:(\$\{LATEST|[a-zA-Z0-9-_\.\.]+))?$`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)

- [AWS SDK for Ruby V3](#)

PromptSummary

Service: Agents for Amazon Bedrock

Contains information about a prompt in your Prompt management tool.

This data type is used in the following API operations:

- [ListPrompts response](#)

Contents

arn

The Amazon Resource Name (ARN) of the prompt.

Type: String

Pattern: `^(arn:aws:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:prompt/[0-9a-zA-Z]{10}(?::[0-9]{1,5})?)$`

Required: Yes

createdAt

The time at which the prompt was created.

Type: Timestamp

Required: Yes

id

The unique identifier of the prompt.

Type: String

Pattern: `^[0-9a-zA-Z]{10}$`

Required: Yes

name

The name of the prompt.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?)\{1,100\}$`

Required: Yes

updatedAt

The time at which the prompt was last updated.

Type: Timestamp

Required: Yes

version

The version of the prompt that this summary applies to.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: `^(DRAFT|[0-9]{0,4}[1-9][0-9]{0,4})$`

Required: Yes

description

The description of the prompt.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PromptTemplateConfiguration

Service: Agents for Amazon Bedrock

Contains the message for a prompt. For more information, see [Prompt management in Amazon Bedrock](#).

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

text

Contains configurations for the text in a message for a prompt.

Type: [TextPromptTemplateConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PromptVariant

Service: Agents for Amazon Bedrock

Contains details about a variant of the prompt.

Contents

name

The name of the prompt variant.

Type: String

Pattern: `^([0-9a-zA-Z][_ -]?)\{1,100\}$`

Required: Yes

templateType

The type of prompt template to use.

Type: String

Valid Values: TEXT

Required: Yes

inferenceConfiguration

Contains inference configurations for the prompt variant.

Type: [PromptInferenceConfiguration](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

modelId

The unique identifier of the model with which to run inference on the prompt.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^(arn:aws(-[^\:]+)??:bedrock:[a-z0-9-]{1,20}:(([\0-9]{12}:custom-model/[a-z0-9-]{1,63}[\.]{1}[a-z0-9-]{1,63}/[a-z0-9]{12})|(:foundation-model/[a-z0-9-]{1,63}[\.]{1}[a-z0-9-]{1,63}([\.:]?[a-z0-9-]{1,63}))|([\0-9]{12}:provisioned-model/[a-z0-9]{12})))|([a-z0-9-]{1,63}[\.]{1}[a-z0-9-]{1,63}([\.:]?[a-z0-9-]{1,63}))|(([\0-9a-zA-Z][_]?)+)$`

Required: No

templateConfiguration

Contains configurations for the prompt template.

Type: [PromptTemplateConfiguration](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RdsConfiguration

Service: Agents for Amazon Bedrock

Contains details about the storage configuration of the knowledge base in Amazon RDS. For more information, see [Create a vector index in Amazon RDS](#).

Contents

credentialsSecretArn

The Amazon Resource Name (ARN) of the secret that you created in AWS Secrets Manager that is linked to your Amazon RDS database.

Type: String

Pattern: `^arn:aws(|-cn|-us-gov):secretsmanager:[a-z0-9-]{1,20}:([0-9]{12}|):secret:[a-zA-Z0-9!/_+=.@-]{1,512}$`

Required: Yes

databaseName

The name of your Amazon RDS database.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 63.

Pattern: `^[a-zA-Z0-9_\-]+$`

Required: Yes

fieldMapping

Contains the names of the fields to which to map information about the vector store.

Type: [RdsFieldMapping](#) object

Required: Yes

resourceArn

The Amazon Resource Name (ARN) of the vector store.

Type: String

Pattern: `^arn:aws(|-cn|-us-gov):rds:[a-zA-Z0-9-]*:[0-9]{12}:cluster:[a-zA-Z0-9-]{1,63}$`

Required: Yes

tableName

The name of the table in the database.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 63.

Pattern: `^[a-zA-Z0-9_\. \-]+$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RdsFieldMapping

Service: Agents for Amazon Bedrock

Contains the names of the fields to which to map information about the vector store.

Contents

metadataField

The name of the field in which Amazon Bedrock stores metadata about the vector store.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 63.

Pattern: `^[a-zA-Z0-9_\-]+$`

Required: Yes

primaryKeyField

The name of the field in which Amazon Bedrock stores the ID for each entry.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 63.

Pattern: `^[a-zA-Z0-9_\-]+$`

Required: Yes

textField

The name of the field in which Amazon Bedrock stores the raw text from your data. The text is split according to the chunking strategy you choose.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 63.

Pattern: `^[a-zA-Z0-9_\-]+$`

Required: Yes

vectorField

The name of the field in which Amazon Bedrock stores the vector embeddings for your data sources.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 63.

Pattern: `^[a-zA-Z0-9_\-\]+$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RedisEnterpriseCloudConfiguration

Service: Agents for Amazon Bedrock

Contains details about the storage configuration of the knowledge base in Redis Enterprise Cloud. For more information, see [Create a vector index in Redis Enterprise Cloud](#).

Contents

credentialsSecretArn

The Amazon Resource Name (ARN) of the secret that you created in AWS Secrets Manager that is linked to your Redis Enterprise Cloud database.

Type: String

Pattern: `^arn:aws(|-cn|-us-gov):secretsmanager:[a-z0-9-]{1,20}:([0-9]{12}|):secret:[a-zA-Z0-9!/_+=.@-]{1,512}$`

Required: Yes

endpoint

The endpoint URL of the Redis Enterprise Cloud database.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^.*$`

Required: Yes

fieldMapping

Contains the names of the fields to which to map information about the vector store.

Type: [RedisEnterpriseCloudFieldMapping](#) object

Required: Yes

vectorIndexName

The name of the vector index.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^.*$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RedisEnterpriseCloudFieldMapping

Service: Agents for Amazon Bedrock

Contains the names of the fields to which to map information about the vector store.

Contents

metadataField

The name of the field in which Amazon Bedrock stores metadata about the vector store.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^.*$`

Required: Yes

textField

The name of the field in which Amazon Bedrock stores the raw text from your data. The text is split according to the chunking strategy you choose.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^.*$`

Required: Yes

vectorField

The name of the field in which Amazon Bedrock stores the vector embeddings for your data sources.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^.*$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RetrievalFlowNodeConfiguration

Service: Agents for Amazon Bedrock

Contains configurations for a Retrieval node in a flow. This node retrieves data from the Amazon S3 location that you specify and returns it as the output.

Contents

serviceConfiguration

Contains configurations for the service to use for retrieving data to return as the output from the node.

Type: [RetrievalFlowNodeServiceConfiguration](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RetrievalFlowNodeS3Configuration

Service: Agents for Amazon Bedrock

Contains configurations for the Amazon S3 location from which to retrieve data to return as the output from the node.

Contents

bucketName

The name of the Amazon S3 bucket from which to retrieve data.

Type: String

Length Constraints: Minimum length of 3. Maximum length of 63.

Pattern: `^[a-z0-9][\.\-a-z0-9]{1,61}[a-z0-9]$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RetrievalFlowNodeServiceConfiguration

Service: Agents for Amazon Bedrock

Contains configurations for the service to use for retrieving data to return as the output from the node.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

s3

Contains configurations for the Amazon S3 location from which to retrieve data to return as the output from the node.

Type: [RetrievalFlowNodeS3Configuration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

S3DataSourceConfiguration

Service: Agents for Amazon Bedrock

The configuration information to connect to Amazon S3 as your data source.

Contents

bucketArn

The Amazon Resource Name (ARN) of the S3 bucket that contains your data.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):s3:::[a-z0-9][a-z0-9.-]{1,61}[a-z0-9]$`

Required: Yes

bucketOwnerAccountId

The account ID for the owner of the S3 bucket.

Type: String

Length Constraints: Fixed length of 12.

Pattern: `^[0-9]{12}$`

Required: No

inclusionPrefixes

A list of S3 prefixes to include certain files or content. For more information, see [Organizing objects using prefixes](#).

Type: Array of strings

Array Members: Fixed number of 1 item.

Length Constraints: Minimum length of 1. Maximum length of 300.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

S3Identifier

Service: Agents for Amazon Bedrock

The identifier information for an Amazon S3 bucket.

Contents

s3BucketName

The name of the S3 bucket.

Type: String

Length Constraints: Minimum length of 3. Maximum length of 63.

Pattern: `^[a-z0-9][\.\-a-z0-9]{1,61}[a-z0-9]$`

Required: No

s3ObjectKey

The S3 object key for the S3 resource.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1024.

Pattern: `^[\\.\-\\!*_\\'\\(\\)a-zA-Z0-9][\\.\-\\!*_\\'\\(\\)\\/a-zA-Z0-9]*$`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

S3Location

Service: Agents for Amazon Bedrock

An Amazon S3 location.

Contents

uri

The location's URI. For example, `s3://my-bucket/chunk-processor/`.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^s3://.{1,128}$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

SalesforceCrawlerConfiguration

Service: Agents for Amazon Bedrock

The configuration of the Salesforce content. For example, configuring specific types of Salesforce content.

Contents

filterConfiguration

The configuration of filtering the Salesforce content. For example, configuring regular expression patterns to include or exclude certain content.

Type: [CrawlFilterConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

SalesforceDataSourceConfiguration

Service: Agents for Amazon Bedrock

The configuration information to connect to Salesforce as your data source.

Contents

sourceConfiguration

The endpoint information to connect to your Salesforce data source.

Type: [SalesforceSourceConfiguration](#) object

Required: Yes

crawlerConfiguration

The configuration of the Salesforce content. For example, configuring specific types of Salesforce content.

Type: [SalesforceCrawlerConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

SalesforceSourceConfiguration

Service: Agents for Amazon Bedrock

The endpoint information to connect to your Salesforce data source.

Contents

authType

The supported authentication type to authenticate and connect to your Salesforce instance.

Type: String

Valid Values: OAUTH2_CLIENT_CREDENTIALS

Required: Yes

credentialsSecretArn

The Amazon Resource Name of an AWS Secrets Manager secret that stores your authentication credentials for your Salesforce instance URL. For more information on the key-value pairs that must be included in your secret, depending on your authentication type, see [Salesforce connection configuration](#).

Type: String

Pattern: `^arn:aws(|-cn|-us-gov):secretsmanager:[a-z0-9-]{1,20}:([0-9]{12}|):secret:[a-zA-Z0-9!/_+=.@-]{1,512}$`

Required: Yes

hostUrl

The Salesforce host URL or instance URL.

Type: String

Pattern: `^https://[A-Za-z0-9][^\s]*$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

SeedUrl

Service: Agents for Amazon Bedrock

The seed or starting point URL. You should be authorized to crawl the URL.

Contents

url

A seed or starting point URL.

Type: String

Pattern: `^https?://[A-Za-z0-9][^\s]*$`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

SemanticChunkingConfiguration

Service: Agents for Amazon Bedrock

Settings for semantic document chunking for a data source. Semantic chunking splits a document into into smaller documents based on groups of similar content derived from the text with natural language processing.

With semantic chunking, each sentence is compared to the next to determine how similar they are. You specify a threshold in the form of a percentile, where adjacent sentences that are less similar than that percentage of sentence pairs are divided into separate chunks. For example, if you set the threshold to 90, then the 10 percent of sentence pairs that are least similar are split. So if you have 101 sentences, 100 sentence pairs are compared, and the 10 with the least similarity are split, creating 11 chunks. These chunks are further split if they exceed the max token size.

You must also specify a buffer size, which determines whether sentences are compared in isolation, or within a moving context window that includes the previous and following sentence. For example, if you set the buffer size to 1, the embedding for sentence 10 is derived from sentences 9, 10, and 11 combined.

Contents

breakpointPercentileThreshold

The dissimilarity threshold for splitting chunks.

Type: Integer

Valid Range: Minimum value of 50. Maximum value of 99.

Required: Yes

bufferSize

The buffer size.

Type: Integer

Valid Range: Minimum value of 0. Maximum value of 1.

Required: Yes

maxTokens

The maximum number of tokens that a chunk can contain.

Type: Integer

Valid Range: Minimum value of 1.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ServerSideEncryptionConfiguration

Service: Agents for Amazon Bedrock

Contains the configuration for server-side encryption.

Contents

kmsKeyArn

The Amazon Resource Name (ARN) of the AWS KMS key used to encrypt the resource.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

SharePointCrawlerConfiguration

Service: Agents for Amazon Bedrock

The configuration of the SharePoint content. For example, configuring specific types of SharePoint content.

Contents

filterConfiguration

The configuration of filtering the SharePoint content. For example, configuring regular expression patterns to include or exclude certain content.

Type: [CrawlFilterConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

SharePointDataSourceConfiguration

Service: Agents for Amazon Bedrock

The configuration information to connect to SharePoint as your data source.

Contents

sourceConfiguration

The endpoint information to connect to your SharePoint data source.

Type: [SharePointSourceConfiguration](#) object

Required: Yes

crawlerConfiguration

The configuration of the SharePoint content. For example, configuring specific types of SharePoint content.

Type: [SharePointCrawlerConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

SharePointSourceConfiguration

Service: Agents for Amazon Bedrock

The endpoint information to connect to your SharePoint data source.

Contents

authType

The supported authentication type to authenticate and connect to your SharePoint site/sites.

Type: String

Valid Values: OAUTH2_CLIENT_CREDENTIALS

Required: Yes

credentialsSecretArn

The Amazon Resource Name of an AWS Secrets Manager secret that stores your authentication credentials for your SharePoint site/sites. For more information on the key-value pairs that must be included in your secret, depending on your authentication type, see [SharePoint connection configuration](#).

Type: String

Pattern: `^arn:aws(|-cn|-us-gov):secretsmanager:[a-z0-9-]{1,20}:([0-9]{12})|):secret:[a-zA-Z0-9!/_+=.@-]{1,512}$`

Required: Yes

domain

The domain of your SharePoint instance or site URL/URLs.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 50.

Required: Yes

hostType

The supported host type, whether online/cloud or server/on-premises.

Type: String

Valid Values: ONLINE

Required: Yes

siteUrls

A list of one or more SharePoint site URLs.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 100 items.

Pattern: `^https://[A-Za-z0-9][^\s]*$`

Required: Yes

tenantId

The identifier of your Microsoft 365 tenant.

Type: String

Length Constraints: Fixed length of 36.

Pattern: `^[0-9a-f]{8}-[0-9a-f]{4}-[0-9a-f]{4}-[0-9a-f]{4}-[0-9a-f]{12}$`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

StorageConfiguration

Service: Agents for Amazon Bedrock

Contains the storage configuration of the knowledge base.

Contents

type

The vector store service in which the knowledge base is stored.

Type: String

Valid Values: OPENSEARCH_SERVERLESS | PINECONE | REDIS_ENTERPRISE_CLOUD | RDS | MONGO_DB_ATLAS

Required: Yes

mongoDbAtlasConfiguration

Contains the storage configuration of the knowledge base in MongoDB Atlas.

Type: [MongoDbAtlasConfiguration](#) object

Required: No

opensearchServerlessConfiguration

Contains the storage configuration of the knowledge base in Amazon OpenSearch Service.

Type: [OpenSearchServerlessConfiguration](#) object

Required: No

pineconeConfiguration

Contains the storage configuration of the knowledge base in Pinecone.

Type: [PineconeConfiguration](#) object

Required: No

rdsConfiguration

Contains details about the storage configuration of the knowledge base in Amazon RDS. For more information, see [Create a vector index in Amazon RDS](#).

Type: [RdsConfiguration](#) object

Required: No

redisEnterpriseCloudConfiguration

Contains the storage configuration of the knowledge base in Redis Enterprise Cloud.

Type: [RedisEnterpriseCloudConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

StorageFlowNodeConfiguration

Service: Agents for Amazon Bedrock

Contains configurations for a Storage node in a flow. This node stores the input in an Amazon S3 location that you specify.

Contents

serviceConfiguration

Contains configurations for the service to use for storing the input into the node.

Type: [StorageFlowNodeServiceConfiguration](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

StorageFlowNodeS3Configuration

Service: Agents for Amazon Bedrock

Contains configurations for the Amazon S3 location in which to store the input into the node.

Contents

bucketName

The name of the Amazon S3 bucket in which to store the input into the node.

Type: String

Length Constraints: Minimum length of 3. Maximum length of 63.

Pattern: `^[a-z0-9][\.-a-z0-9]{1,61}[a-z0-9]$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

StorageFlowNodeServiceConfiguration

Service: Agents for Amazon Bedrock

Contains configurations for the service to use for storing the input into the node.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

s3

Contains configurations for the Amazon S3 location in which to store the input into the node.

Type: [StorageFlowNodeS3Configuration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

TextPromptTemplateConfiguration

Service: Agents for Amazon Bedrock

Contains configurations for a text prompt template. To include a variable, enclose a word in double curly braces as in `{{variable}}`.

Contents

text

The message for the prompt.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200000.

Required: Yes

inputVariables

An array of the variables in the prompt template.

Type: Array of [PromptInputVariable](#) objects

Array Members: Minimum number of 0 items. Maximum number of 5 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

Transformation

Service: Agents for Amazon Bedrock

A custom processing step for documents moving through a data source ingestion pipeline. To process documents after they have been converted into chunks, set the step to apply to `POST_CHUNKING`.

Contents

`stepToApply`

When the service applies the transformation.

Type: String

Valid Values: `POST_CHUNKING`

Required: Yes

`transformationFunction`

A Lambda function that processes documents.

Type: [TransformationFunction](#) object

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

TransformationFunction

Service: Agents for Amazon Bedrock

A Lambda function that processes documents.

Contents

transformationLambdaConfiguration

The Lambda function.

Type: [TransformationLambdaConfiguration](#) object

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

TransformationLambdaConfiguration

Service: Agents for Amazon Bedrock

A Lambda function that processes documents.

Contents

lambdaArn

The function's ARN identifier.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^arn:(aws[a-zA-Z-]*)?:lambda:[a-z]{2}(-gov)?-[a-z]+\-\d{1}:\d{12}:function:[a-zA-Z0-9-_\.\.]+(:(\$\{LATEST|[a-zA-Z0-9-_\.\.]+)))?$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

UrlConfiguration

Service: Agents for Amazon Bedrock

The configuration of web URLs that you want to crawl. You should be authorized to crawl the URLs.

Contents

seedUrls

One or more seed or starting point URLs.

Type: Array of [SeedUrl](#) objects

Array Members: Minimum number of 1 item. Maximum number of 100 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ValidationExceptionField

Service: Agents for Amazon Bedrock

Stores information about a field passed inside a request that resulted in a validation error.

Contents

message

A message describing why this field failed validation.

Type: String

Pattern: `^[\s\S]+$`

Required: Yes

name

The name of the field.

Type: String

Pattern: `^[\s\S]+$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

VectorIngestionConfiguration

Service: Agents for Amazon Bedrock

Contains details about how to ingest the documents in a data source.

Contents

chunkingConfiguration

Details about how to chunk the documents in the data source. A *chunk* refers to an excerpt from a data source that is returned when the knowledge base that it belongs to is queried.

Type: [ChunkingConfiguration](#) object

Required: No

customTransformationConfiguration

A custom document transformer for parsed data source documents.

Type: [CustomTransformationConfiguration](#) object

Required: No

parsingConfiguration

A custom parser for data source documents.

Type: [ParsingConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

VectorKnowledgeBaseConfiguration

Service: Agents for Amazon Bedrock

Contains details about the model used to create vector embeddings for the knowledge base.

Contents

embeddingModelArn

The Amazon Resource Name (ARN) of the model used to create vector embeddings for the knowledge base.

Type: String

Length Constraints: Minimum length of 20. Maximum length of 2048.

Pattern: `^(arn:aws(-[^\:]+)?):bedrock:[a-z0-9-]{1,20}:((([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}/[a-z0-9]{12})|(:foundation-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.:]?[a-z0-9-]{1,63}))|([0-9]{12}:provisioned-model/[a-z0-9]{12})))|([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.:]?[a-z0-9-]{1,63}))|(([0-9a-zA-Z][_]?)+)$`

Required: Yes

embeddingModelConfiguration

The embeddings model configuration details for the vector model used in Knowledge Base.

Type: [EmbeddingModelConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

WebCrawlerConfiguration

Service: Agents for Amazon Bedrock

The configuration of web URLs that you want to crawl. You should be authorized to crawl the URLs.

Contents

crawlerLimits

The configuration of crawl limits for the web URLs.

Type: [WebCrawlerLimits](#) object

Required: No

exclusionFilters

A list of one or more exclusion regular expression patterns to exclude certain URLs. If you specify an inclusion and exclusion filter/pattern and both match a URL, the exclusion filter takes precedence and the web content of the URL isn't crawled.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 25 items.

Length Constraints: Minimum length of 1. Maximum length of 1000.

Required: No

inclusionFilters

A list of one or more inclusion regular expression patterns to include certain URLs. If you specify an inclusion and exclusion filter/pattern and both match a URL, the exclusion filter takes precedence and the web content of the URL isn't crawled.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 25 items.

Length Constraints: Minimum length of 1. Maximum length of 1000.

Required: No

scope

The scope of what is crawled for your URLs.

You can choose to crawl only web pages that belong to the same host or primary domain. For example, only web pages that contain the seed URL "https://docs.aws.amazon.com/bedrock/latest/userguide/" and no other domains. You can choose to include sub domains in addition to the host or primary domain. For example, web pages that contain "aws.amazon.com" can also include sub domain "docs.aws.amazon.com".

Type: String

Valid Values: HOST_ONLY | SUBDOMAINS

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

WebCrawlerLimits

Service: Agents for Amazon Bedrock

The rate limits for the URLs that you want to crawl. You should be authorized to crawl the URLs.

Contents

rateLimit

The max rate at which pages are crawled, up to 300 per minute per host.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 300.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

WebDataSourceConfiguration

Service: Agents for Amazon Bedrock

The configuration details for the web data source.

Contents

sourceConfiguration

The source configuration details for the web data source.

Type: [WebSourceConfiguration](#) object

Required: Yes

crawlerConfiguration

The Web Crawler configuration details for the web data source.

Type: [WebCrawlerConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

WebSourceConfiguration

Service: Agents for Amazon Bedrock

The configuration of the URL/URLs for the web content that you want to crawl. You should be authorized to crawl the URLs.

Contents

urlConfiguration

The configuration of the URL/URLs.

Type: [UrlConfiguration](#) object

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

Agents for Amazon Bedrock Runtime

The following data types are supported by Agents for Amazon Bedrock Runtime:

- [ActionGroupInvocationInput](#)
- [ActionGroupInvocationOutput](#)
- [ApiInvocationInput](#)
- [ApiParameter](#)
- [ApiRequestBody](#)
- [ApiResponse](#)
- [Attribution](#)
- [ByteContentDoc](#)

- [ByteContentFile](#)
- [Citation](#)
- [CodeInterpreterInvocationInput](#)
- [CodeInterpreterInvocationOutput](#)
- [ContentBody](#)
- [ExternalSource](#)
- [ExternalSourcesGenerationConfiguration](#)
- [ExternalSourcesRetrieveAndGenerateConfiguration](#)
- [FailureTrace](#)
- [FilePart](#)
- [FileSource](#)
- [FilterAttribute](#)
- [FinalResponse](#)
- [FlowCompletionEvent](#)
- [FlowInput](#)
- [FlowInputContent](#)
- [FlowOutputContent](#)
- [FlowOutputEvent](#)
- [FlowResponseStream](#)
- [FunctionInvocationInput](#)
- [FunctionParameter](#)
- [FunctionResult](#)
- [GeneratedResponsePart](#)
- [GenerationConfiguration](#)
- [GuardrailAssessment](#)
- [GuardrailConfiguration](#)
- [GuardrailContentFilter](#)
- [GuardrailContentPolicyAssessment](#)
- [GuardrailCustomWord](#)
- [GuardrailManagedWord](#)

- [GuardrailPiiEntityFilter](#)
- [GuardrailRegexFilter](#)
- [GuardrailSensitiveInformationPolicyAssessment](#)
- [GuardrailTopic](#)
- [GuardrailTopicPolicyAssessment](#)
- [GuardrailTrace](#)
- [GuardrailWordPolicyAssessment](#)
- [InferenceConfig](#)
- [InferenceConfiguration](#)
- [InputFile](#)
- [InvocationInput](#)
- [InvocationInputMember](#)
- [InvocationResultMember](#)
- [KnowledgeBaseConfiguration](#)
- [KnowledgeBaseLookupInput](#)
- [KnowledgeBaseLookupOutput](#)
- [KnowledgeBaseQuery](#)
- [KnowledgeBaseRetrievalConfiguration](#)
- [KnowledgeBaseRetrievalResult](#)
- [KnowledgeBaseRetrieveAndGenerateConfiguration](#)
- [KnowledgeBaseVectorSearchConfiguration](#)
- [Memory](#)
- [MemorySessionSummary](#)
- [ModelInvocationInput](#)
- [Observation](#)
- [OrchestrationConfiguration](#)
- [OrchestrationTrace](#)
- [OutputFile](#)
- [Parameter](#)
- [PayloadPart](#)

- [PostProcessingModelInvocationOutput](#)
- [PostProcessingParsedResponse](#)
- [PostProcessingTrace](#)
- [PreProcessingModelInvocationOutput](#)
- [PreProcessingParsedResponse](#)
- [PreProcessingTrace](#)
- [PromptTemplate](#)
- [PropertyParameters](#)
- [QueryTransformationConfiguration](#)
- [Rationale](#)
- [RepromptResponse](#)
- [RequestBody](#)
- [ResponseStream](#)
- [RetrievalFilter](#)
- [RetrievalResultConfluenceLocation](#)
- [RetrievalResultContent](#)
- [RetrievalResultLocation](#)
- [RetrievalResultS3Location](#)
- [RetrievalResultSalesforceLocation](#)
- [RetrievalResultSharePointLocation](#)
- [RetrievalResultWebLocation](#)
- [RetrieveAndGenerateConfiguration](#)
- [RetrieveAndGenerateInput](#)
- [RetrieveAndGenerateOutput](#)
- [RetrieveAndGenerateSessionConfiguration](#)
- [RetrievedReference](#)
- [ReturnControlPayload](#)
- [S3ObjectDoc](#)
- [S3ObjectFile](#)
- [SessionState](#)

- [Span](#)
- [TextInferenceConfig](#)
- [TextResponsePart](#)
- [Trace](#)
- [TracePart](#)

ActionGroupInvocationInput

Service: Agents for Amazon Bedrock Runtime

Contains information about the action group being invoked. For more information about the possible structures, see the InvocationInput tab in [OrchestrationTrace](#) in the Amazon Bedrock User Guide.

Contents

actionGroupName

The name of the action group.

Type: String

Required: No

apiPath

The path to the API to call, based off the action group.

Type: String

Required: No

executionType

How fulfillment of the action is handled. For more information, see [Handling fulfillment of the action](#).

Type: String

Valid Values: LAMBDA | RETURN_CONTROL

Required: No

function

The function in the action group to call.

Type: String

Required: No

invocationId

The unique identifier of the invocation. Only returned if the `executionType` is `RETURN_CONTROL`.

Type: String

Required: No

parameters

The parameters in the Lambda input event.

Type: Array of [Parameter](#) objects

Required: No

requestBody

The parameters in the request body for the Lambda input event.

Type: [RequestBody](#) object

Required: No

verb

The API method being used, based off the action group.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ActionGroupInvocationOutput

Service: Agents for Amazon Bedrock Runtime

Contains the JSON-formatted string returned by the API invoked by the action group.

Contents

text

The JSON-formatted string returned by the API invoked by the action group.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ApiInvocationInput

Service: Agents for Amazon Bedrock Runtime

Contains information about the API operation that the agent predicts should be called.

This data type is used in the following API operations:

- In the `returnControl` field of the [InvokeAgent response](#)

Contents

actionGroup

The action group that the API operation belongs to.

Type: String

Required: Yes

apiPath

The path to the API operation.

Type: String

Required: No

httpMethod

The HTTP method of the API operation.

Type: String

Required: No

parameters

The parameters to provide for the API request, as the agent elicited from the user.

Type: Array of [ApiParameter](#) objects

Required: No

requestBody

The request body to provide for the API request, as the agent elicited from the user.

Type: [ApiRequestBody](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ApiParameter

Service: Agents for Amazon Bedrock Runtime

Information about a parameter to provide to the API request.

This data type is used in the following API operations:

- [InvokeAgent response](#)

Contents

name

The name of the parameter.

Type: String

Required: No

type

The data type for the parameter.

Type: String

Required: No

value

The value of the parameter.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)

- [AWS SDK for Ruby V3](#)

ApiRequestBody

Service: Agents for Amazon Bedrock Runtime

The request body to provide for the API request, as the agent elicited from the user.

This data type is used in the following API operations:

- [InvokeAgent response](#)

Contents

content

The content of the request body. The key of the object in this field is a media type defining the format of the request body.

Type: String to [PropertyParameters](#) object map

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ApiResponse

Service: Agents for Amazon Bedrock Runtime

Contains information about the API operation that was called from the action group and the response body that was returned.

This data type is used in the following API operations:

- In the `returnControlInvocationResults` of the [InvokeAgent request](#)

Contents

actionGroup

The action group that the API operation belongs to.

Type: String

Required: Yes

apiPath

The path to the API operation.

Type: String

Required: No

httpMethod

The HTTP method for the API operation.

Type: String

Required: No

httpStatusCode

http status code from API execution response (for example: 200, 400, 500).

Type: Integer

Required: No

responseBody

The response body from the API operation. The key of the object is the content type (currently, only TEXT is supported). The response may be returned directly or from the Lambda function.

Type: String to [ContentBody](#) object map

Required: No

responseState

Controls the final response state returned to end user when API/Function execution failed. When this state is FAILURE, the request would fail with dependency failure exception. When this state is REPROMPT, the API/function response will be sent to model for re-prompt

Type: String

Valid Values: FAILURE | REPROMPT

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

Attribution

Service: Agents for Amazon Bedrock Runtime

Contains citations for a part of an agent response.

Contents

citations

A list of citations and related information for a part of an agent response.

Type: Array of [Citation](#) objects

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ByteContentDoc

Service: Agents for Amazon Bedrock Runtime

This property contains the document to chat with, along with its attributes.

Contents

contentType

The MIME type of the document contained in the wrapper object.

Type: String

Pattern: `[a-z]{1,20}/.{1,20}`

Required: Yes

data

The byte value of the file to upload, encoded as a Base-64 string.

Type: Base64-encoded binary data object

Length Constraints: Minimum length of 1. Maximum length of 10485760.

Required: Yes

identifier

The file name of the document contained in the wrapper object.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1024.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)

- [AWS SDK for Ruby V3](#)

ByteContentFile

Service: Agents for Amazon Bedrock Runtime

The property contains the file to chat with, along with its attributes.

Contents

data

The byte value of the file to attach, encoded as Base-64 string. The maximum size of all files that is attached is 10MB. You can attach a maximum of 5 files.

Type: Base64-encoded binary data object

Length Constraints: Minimum length of 1. Maximum length of 10485760.

Required: Yes

mediaType

The MIME type of data contained in the file used for chat.

Type: String

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

Citation

Service: Agents for Amazon Bedrock Runtime

An object containing a segment of the generated response that is based on a source in the knowledge base, alongside information about the source.

This data type is used in the following API operations:

- [InvokeAgent response](#) – in the citations field
- [RetrieveAndGenerate response](#) – in the citations field

Contents

generatedResponsePart

Contains the generated response and metadata

Type: [GeneratedResponsePart](#) object

Required: No

retrievedReferences

Contains metadata about the sources cited for the generated response.

Type: Array of [RetrievedReference](#) objects

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

CodeInterpreterInvocationInput

Service: Agents for Amazon Bedrock Runtime

Contains information about the code interpreter being invoked.

Contents

code

The code for the code interpreter to use.

Type: String

Required: No

files

Files that are uploaded for code interpreter to use.

Type: Array of strings

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

CodeInterpreterInvocationOutput

Service: Agents for Amazon Bedrock Runtime

Contains the JSON-formatted string returned by the API invoked by the code interpreter.

Contents

executionError

Contains the error returned from code execution.

Type: String

Required: No

executionOutput

Contains the successful output returned from code execution

Type: String

Required: No

executionTimeout

Indicates if the execution of the code timed out.

Type: Boolean

Required: No

files

Contains output files, if generated by code execution.

Type: Array of strings

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)

- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ContentBody

Service: Agents for Amazon Bedrock Runtime

Contains the body of the API response.

This data type is used in the following API operations:

- In the `returnControlInvocationResults` field of the [InvokeAgent request](#)

Contents

body

The body of the API response.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ExternalSource

Service: Agents for Amazon Bedrock Runtime

The unique external source of the content contained in the wrapper object.

Contents

sourceType

The source type of the external source wrapper object.

Type: String

Valid Values: S3 | BYTE_CONTENT

Required: Yes

byteContent

The identifier, contentType, and data of the external source wrapper object.

Type: [ByteContentDoc](#) object

Required: No

s3Location

The S3 location of the external source wrapper object.

Type: [S3ObjectDoc](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ExternalSourcesGenerationConfiguration

Service: Agents for Amazon Bedrock Runtime

Contains the generation configuration of the external source wrapper object.

Contents

additionalModelRequestFields

Additional model parameters and their corresponding values not included in the `textInferenceConfig` structure for an external source. Takes in custom model parameters specific to the language model being used.

Type: String to JSON value map

Key Length Constraints: Minimum length of 1. Maximum length of 100.

Required: No

guardrailConfiguration

The configuration details for the guardrail.

Type: [GuardrailConfiguration](#) object

Required: No

inferenceConfig

Configuration settings for inference when using `RetrieveAndGenerate` to generate responses while using an external source.

Type: [InferenceConfig](#) object

Required: No

promptTemplate

Contain the `textPromptTemplate` string for the external source wrapper object.

Type: [PromptTemplate](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ExternalSourcesRetrieveAndGenerateConfiguration

Service: Agents for Amazon Bedrock Runtime

The configurations of the external source wrapper object in the retrieveAndGenerate function.

Contents

modelArn

The modelArn used with the external source wrapper object in the retrieveAndGenerate function.

Type: String

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}/[a-z0-9]{12})|(:foundation-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}))$`

Required: Yes

sources

The document used with the external source wrapper object in the retrieveAndGenerate function.

Type: Array of [ExternalSource](#) objects

Array Members: Fixed number of 1 item.

Required: Yes

generationConfiguration

The prompt used with the external source wrapper object with the retrieveAndGenerate function.

Type: [ExternalSourcesGenerationConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FailureTrace

Service: Agents for Amazon Bedrock Runtime

Contains information about the failure of the interaction.

Contents

failureReason

The reason the interaction failed.

Type: String

Required: No

traceId

The unique identifier of the trace.

Type: String

Length Constraints: Minimum length of 2. Maximum length of 16.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FilePart

Service: Agents for Amazon Bedrock Runtime

Contains intermediate response for code interpreter if any files have been generated.

Contents

files

Files containing intermediate response for the user.

Type: Array of [OutputFile](#) objects

Array Members: Minimum number of 0 items. Maximum number of 5 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FileSource

Service: Agents for Amazon Bedrock Runtime

The source file of the content contained in the wrapper object.

Contents

sourceType

The source type of the files to attach.

Type: String

Valid Values: S3 | BYTE_CONTENT

Required: Yes

byteContent

The data and the text of the attached files.

Type: [ByteContentFile](#) object

Required: No

s3Location

The s3 location of the files to attach.

Type: [S3ObjectFile](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FilterAttribute

Service: Agents for Amazon Bedrock Runtime

Specifies the name that the metadata attribute must match and the value to which to compare the value of the metadata attribute. For more information, see [Query configurations](#).

This data type is used in the following API operations:

- [RetrieveAndGenerate request](#)

Contents

key

The name that the metadata attribute must match.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 100.

Required: Yes

value

The value to which to compare the value of the metadata attribute.

Type: JSON value

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FinalResponse

Service: Agents for Amazon Bedrock Runtime

Contains details about the response to the user.

Contents

text

The text in the response to the user.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowCompletionEvent

Service: Agents for Amazon Bedrock Runtime

Contains information about why a flow completed.

This data type is used in the following API operations:

- [InvokeFlow response](#)

Contents

completionReason

The reason that the flow completed.

Type: String

Valid Values: SUCCESS

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowInput

Service: Agents for Amazon Bedrock Runtime

Contains information about an input into the flow and what to do with it.

This data type is used in the following API operations:

- [InvokeFlow request](#)

Contents

content

Contains information about an input into the flow.

Type: [FlowInputContent](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: Yes

nodeName

A name for the input of the flow input node.

Type: String

Pattern: `^[a-zA-Z]([_]?[0-9a-zA-Z]){0,99}$`

Required: Yes

nodeOutputName

A name for the output of the flow input node.

Type: String

Pattern: `^[a-zA-Z]([_]?[0-9a-zA-Z]){0,99}$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowInputContent

Service: Agents for Amazon Bedrock Runtime

Contains information about an input into the flow.

This data type is used in the following API operations:

- [InvokeFlow request](#)

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

document

The input for the flow input node.

Type: JSON value

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowOutputContent

Service: Agents for Amazon Bedrock Runtime

Contains information about the output node.

This data type is used in the following API operations:

- [InvokeFlow request](#)

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

document

A name for the output of the flow.

Type: JSON value

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowOutputEvent

Service: Agents for Amazon Bedrock Runtime

Contains information about an output from flow invocation.

This data type is used in the following API operations:

- [InvokeFlow response](#)

Contents

content

The output of the node.

Type: [FlowOutputContent](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: Yes

nodeName

The name of the node to which input was provided.

Type: String

Pattern: `^[a-zA-Z]([_]?[0-9a-zA-Z]){0,99}$`

Required: Yes

nodeType

The type of node to which input was provided.

Type: String

Valid Values: `FlowInputNode | FlowOutputNode | LambdaFunctionNode | KnowledgeBaseNode | PromptNode | ConditionNode | LexNode`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FlowResponseStream

Service: Agents for Amazon Bedrock Runtime

The output of the flow.

This data type is used in the following API operations:

- [InvokeFlow response](#)

Contents

accessDeniedException

The request is denied because of missing access permissions. Check your permissions and retry your request.

Type: Exception

HTTP Status Code: 403

Required: No

badGatewayException

There was an issue with a dependency due to a server issue. Retry your request.

Type: Exception

HTTP Status Code: 502

Required: No

conflictException

There was a conflict performing an operation. Resolve the conflict and retry your request.

Type: Exception

HTTP Status Code: 409

Required: No

dependencyFailedException

There was an issue with a dependency. Check the resource configurations and retry the request.

Type: Exception

HTTP Status Code: 424

Required: No

flowCompletionEvent

Contains information about why the flow completed.

Type: [FlowCompletionEvent](#) object

Required: No

flowOutputEvent

Contains information about an output from flow invocation.

Type: [FlowOutputEvent](#) object

Required: No

internalServerErrorException

An internal server error occurred. Retry your request.

Type: Exception

HTTP Status Code: 500

Required: No

resourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

Type: Exception

HTTP Status Code: 404

Required: No

serviceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

Type: Exception

HTTP Status Code: 400

Required: No

throttlingException

The number of requests exceeds the limit. Resubmit your request later.

Type: Exception

HTTP Status Code: 429

Required: No

validationException

Input validation failed. Check your request parameters and retry the request.

Type: Exception

HTTP Status Code: 400

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

FunctionInvocationInput

Service: Agents for Amazon Bedrock Runtime

Contains information about the function that the agent predicts should be called.

This data type is used in the following API operations:

- In the `returnControl` field of the [InvokeAgent response](#)

Contents

actionGroup

The action group that the function belongs to.

Type: String

Required: Yes

function

The name of the function.

Type: String

Required: No

parameters

A list of parameters of the function.

Type: Array of [FunctionParameter](#) objects

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)

- [AWS SDK for Ruby V3](#)

FunctionParameter

Service: Agents for Amazon Bedrock Runtime

Contains information about a parameter of the function.

This data type is used in the following API operations:

- In the `returnControl` field of the [InvokeAgent response](#)

Contents

name

The name of the parameter.

Type: String

Required: No

type

The data type of the parameter.

Type: String

Required: No

value

The value of the parameter.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)

- [AWS SDK for Ruby V3](#)

FunctionResult

Service: Agents for Amazon Bedrock Runtime

Contains information about the function that was called from the action group and the response that was returned.

This data type is used in the following API operations:

- In the `returnControlInvocationResults` of the [InvokeAgent request](#)

Contents

actionGroup

The action group that the function belongs to.

Type: String

Required: Yes

function

The name of the function that was called.

Type: String

Required: No

responseBody

The response from the function call using the parameters. The key of the object is the content type (currently, only TEXT is supported). The response may be returned directly or from the Lambda function.

Type: String to [ContentBody](#) object map

Required: No

responseState

Controls the final response state returned to end user when API/Function execution failed. When this state is FAILURE, the request would fail with dependency failure exception. When this state is REPROMPT, the API/function response will be sent to model for re-prompt

Type: String

Valid Values: FAILURE | REPROMPT

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GeneratedResponsePart

Service: Agents for Amazon Bedrock Runtime

Contains metadata about a part of the generated response that is accompanied by a citation.

This data type is used in the following API operations:

- [InvokeAgent response](#) – in the generatedResponsePart field
- [RetrieveAndGenerate response](#) – in the generatedResponsePart field

Contents

textResponsePart

Contains metadata about a textual part of the generated response that is accompanied by a citation.

Type: [TextResponsePart](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GenerationConfiguration

Service: Agents for Amazon Bedrock Runtime

Contains configurations for response generation based on the knowledge base query results.

This data type is used in the following API operations:

- [RetrieveAndGenerate request](#)

Contents

additionalModelRequestFields

Additional model parameters and corresponding values not included in the `textInferenceConfig` structure for a knowledge base. This allows users to provide custom model parameters specific to the language model being used.

Type: String to JSON value map

Key Length Constraints: Minimum length of 1. Maximum length of 100.

Required: No

guardrailConfiguration

The configuration details for the guardrail.

Type: [GuardrailConfiguration](#) object

Required: No

inferenceConfig

Configuration settings for inference when using `RetrieveAndGenerate` to generate responses while using a knowledge base as a source.

Type: [InferenceConfig](#) object

Required: No

promptTemplate

Contains the template for the prompt that's sent to the model for response generation.

Type: [PromptTemplate](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailAssessment

Service: Agents for Amazon Bedrock Runtime

Assessment details of the content analyzed by Guardrails.

Contents

contentPolicy

Content policy details of the Guardrail.

Type: [GuardrailContentPolicyAssessment](#) object

Required: No

sensitiveInformationPolicy

Sensitive Information policy details of Guardrail.

Type: [GuardrailSensitiveInformationPolicyAssessment](#) object

Required: No

topicPolicy

Topic policy details of the Guardrail.

Type: [GuardrailTopicPolicyAssessment](#) object

Required: No

wordPolicy

Word policy details of the Guardrail.

Type: [GuardrailWordPolicyAssessment](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)

- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailConfiguration

Service: Agents for Amazon Bedrock Runtime

The configuration details for the guardrail.

Contents

guardrailId

The unique identifier for the guardrail.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 64.

Pattern: `^[a-z0-9]+$`

Required: Yes

guardrailVersion

The version of the guardrail.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: `^(([1-9][0-9]{0,7})|(DRAFT))$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailContentFilter

Service: Agents for Amazon Bedrock Runtime

Details of the content filter used in the Guardrail.

Contents

action

The action placed on the content by the Guardrail filter.

Type: String

Valid Values: BLOCKED

Required: No

confidence

The confidence level regarding the content detected in the filter by the Guardrail.

Type: String

Valid Values: NONE | LOW | MEDIUM | HIGH

Required: No

type

The type of content detected in the filter by the Guardrail.

Type: String

Valid Values: INSULTS | HATE | SEXUAL | VIOLENCE | MISCONDUCT | PROMPT_ATTACK

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)

- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailContentPolicyAssessment

Service: Agents for Amazon Bedrock Runtime

The details of the policy assessment in the Guardrails filter.

Contents

filters

The filter details of the policy assessment used in the Guardrails filter.

Type: Array of [GuardrailContentFilter](#) objects

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailCustomWord

Service: Agents for Amazon Bedrock Runtime

The custom word details for the filter in the Guardrail.

Contents

action

The action details for the custom word filter in the Guardrail.

Type: String

Valid Values: BLOCKED

Required: No

match

The match details for the custom word filter in the Guardrail.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailManagedWord

Service: Agents for Amazon Bedrock Runtime

The managed word details for the filter in the Guardrail.

Contents

action

The action details for the managed word filter in the Guardrail.

Type: String

Valid Values: BLOCKED

Required: No

match

The match details for the managed word filter in the Guardrail.

Type: String

Required: No

type

The type details for the managed word filter in the Guardrail.

Type: String

Valid Values: PROFANITY

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailPiiEntityFilter

Service: Agents for Amazon Bedrock Runtime

The Guardrail filter to identify and remove personally identifiable information (PII).

Contents

action

The action of the Guardrail filter to identify and remove PII.

Type: String

Valid Values: BLOCKED | ANONYMIZED

Required: No

match

The match to settings in the Guardrail filter to identify and remove PII.

Type: String

Required: No

type

The type of PII the Guardrail filter has identified and removed.

Type: String

Valid Values: ADDRESS | AGE | AWS_ACCESS_KEY | AWS_SECRET_KEY | CA_HEALTH_NUMBER | CA_SOCIAL_INSURANCE_NUMBER | CREDIT_DEBIT_CARD_CVV | CREDIT_DEBIT_CARD_EXPIRY | CREDIT_DEBIT_CARD_NUMBER | DRIVER_ID | EMAIL | INTERNATIONAL_BANK_ACCOUNT_NUMBER | IP_ADDRESS | LICENSE_PLATE | MAC_ADDRESS | NAME | PASSWORD | PHONE | PIN | SWIFT_CODE | UK_NATIONAL_HEALTH_SERVICE_NUMBER | UK_NATIONAL_INSURANCE_NUMBER | UK_UNIQUE_TAXPAYER_REFERENCE_NUMBER | URL | USERNAME | US_BANK_ACCOUNT_NUMBER | US_BANK_ROUTING_NUMBER | US_INDIVIDUAL_TAX_IDENTIFICATION_NUMBER | US_PASSPORT_NUMBER | US_SOCIAL_SECURITY_NUMBER | VEHICLE_IDENTIFICATION_NUMBER

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailRegexFilter

Service: Agents for Amazon Bedrock Runtime

The details for the regex filter used in the Guardrail.

Contents

action

The action details for the regex filter used in the Guardrail.

Type: String

Valid Values: BLOCKED | ANONYMIZED

Required: No

match

The match details for the regex filter used in the Guardrail.

Type: String

Required: No

name

The name details for the regex filter used in the Guardrail.

Type: String

Required: No

regex

The regex details for the regex filter used in the Guardrail.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailSensitiveInformationPolicyAssessment

Service: Agents for Amazon Bedrock Runtime

The details of the sensitive policy assessment used in the Guardrail.

Contents

piiEntities

The details of the PII entities used in the sensitive policy assessment for the Guardrail.

Type: Array of [GuardrailPiiEntityFilter](#) objects

Required: No

regexes

The details of the regexes used in the sensitive policy assessment for the Guardrail.

Type: Array of [GuardrailRegexFilter](#) objects

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailTopic

Service: Agents for Amazon Bedrock Runtime

The details for a specific topic defined in the Guardrail.

Contents

action

The action details on a specific topic in the Guardrail.

Type: String

Valid Values: BLOCKED

Required: No

name

The name details on a specific topic in the Guardrail.

Type: String

Required: No

type

The type details on a specific topic in the Guardrail.

Type: String

Valid Values: DENY

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailTopicPolicyAssessment

Service: Agents for Amazon Bedrock Runtime

The details of the policy assessment used in the Guardrail.

Contents

topics

The topic details of the policy assessment used in the Guardrail.

Type: Array of [GuardrailTopic](#) objects

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailTrace

Service: Agents for Amazon Bedrock Runtime

The trace details used in the Guardrail.

Contents

action

The trace action details used with the Guardrail.

Type: String

Valid Values: INTERVENED | NONE

Required: No

inputAssessments

The details of the input assessments used in the Guardrail Trace.

Type: Array of [GuardrailAssessment](#) objects

Required: No

outputAssessments

The details of the output assessments used in the Guardrail Trace.

Type: Array of [GuardrailAssessment](#) objects

Required: No

traceId

The details of the trace Id used in the Guardrail Trace.

Type: String

Length Constraints: Minimum length of 2. Maximum length of 16.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailWordPolicyAssessment

Service: Agents for Amazon Bedrock Runtime

The assessment details for words defined in the Guardrail filter.

Contents

customWords

The custom word details for words defined in the Guardrail filter.

Type: Array of [GuardrailCustomWord](#) objects

Required: No

managedWordLists

The managed word lists for words defined in the Guardrail filter.

Type: Array of [GuardrailManagedWord](#) objects

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

InferenceConfig

Service: Agents for Amazon Bedrock Runtime

The configuration for inference settings when generating responses using RetrieveAndGenerate.

Contents

textInferenceConfig

Configuration settings specific to text generation while generating responses using RetrieveAndGenerate.

Type: [TextInferenceConfig](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

InferenceConfiguration

Service: Agents for Amazon Bedrock Runtime

Specifications about the inference parameters that were provided alongside the prompt. These are specified in the [PromptOverrideConfiguration](#) object that was set when the agent was created or updated. For more information, see [Inference parameters for foundation models](#).

Contents

maximumLength

The maximum number of tokens allowed in the generated response.

Type: Integer

Valid Range: Minimum value of 0. Maximum value of 4096.

Required: No

stopSequences

A list of stop sequences. A stop sequence is a sequence of characters that causes the model to stop generating the response.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 4 items.

Required: No

temperature

The likelihood of the model selecting higher-probability options while generating a response. A lower value makes the model more likely to choose higher-probability options, while a higher value makes the model more likely to choose lower-probability options.

Type: Float

Valid Range: Minimum value of 0. Maximum value of 1.

Required: No

topK

While generating a response, the model determines the probability of the following token at each point of generation. The value that you set for topK is the number of most-likely

candidates from which the model chooses the next token in the sequence. For example, if you set `topK` to 50, the model selects the next token from among the top 50 most likely choices.

Type: Integer

Valid Range: Minimum value of 0. Maximum value of 500.

Required: No

topP

While generating a response, the model determines the probability of the following token at each point of generation. The value that you set for `Top P` determines the number of most-likely candidates from which the model chooses the next token in the sequence. For example, if you set `topP` to 80, the model only selects the next token from the top 80% of the probability distribution of next tokens.

Type: Float

Valid Range: Minimum value of 0. Maximum value of 1.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

InputFile

Service: Agents for Amazon Bedrock Runtime

Contains details of the source files.

Contents

name

The name of the source file.

Type: String

Required: Yes

source

Specifies where the files are located.

Type: [FileSource](#) object

Required: Yes

useCase

Specifies how the source files will be used by the code interpreter.

Type: String

Valid Values: CODE_INTERPRETER | CHAT

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

InvocationInput

Service: Agents for Amazon Bedrock Runtime

Contains information pertaining to the action group or knowledge base that is being invoked.

Contents

actionGroupInvocationInput

Contains information about the action group to be invoked.

Type: [ActionGroupInvocationInput](#) object

Required: No

codeInterpreterInvocationInput

Contains information about the code interpreter to be invoked.

Type: [CodeInterpreterInvocationInput](#) object

Required: No

invocationType

Specifies whether the agent is invoking an action group or a knowledge base.

Type: String

Valid Values: ACTION_GROUP | KNOWLEDGE_BASE | FINISH | ACTION_GROUP_CODE_INTERPRETER

Required: No

knowledgeBaseLookupInput

Contains details about the knowledge base to look up and the query to be made.

Type: [KnowledgeBaseLookupInput](#) object

Required: No

traceId

The unique identifier of the trace.

Type: String

Length Constraints: Minimum length of 2. Maximum length of 16.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

InvocationInputMember

Service: Agents for Amazon Bedrock Runtime

Contains details about the API operation or function that the agent predicts should be called.

This data type is used in the following API operations:

- In the `returnControl` field of the [InvokeAgent response](#)

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

apiInvocationInput

Contains information about the API operation that the agent predicts should be called.

Type: [ApiInvocationInput](#) object

Required: No

functionInvocationInput

Contains information about the function that the agent predicts should be called.

Type: [FunctionInvocationInput](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

InvocationResultMember

Service: Agents for Amazon Bedrock Runtime

A result from the invocation of an action. For more information, see [Return control to the agent developer](#) and [Control session context](#).

This data type is used in the following API operations:

- [InvokeAgent request](#)

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

apiResult

The result from the API response from the action group invocation.

Type: [ApiResult](#) object

Required: No

functionResult

The result from the function from the action group invocation.

Type: [FunctionResult](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)

- [AWS SDK for Ruby V3](#)

KnowledgeBaseConfiguration

Service: Agents for Amazon Bedrock Runtime

Configurations to apply to a knowledge base attached to the agent during query. For more information, see [Knowledge base retrieval configurations](#).

Contents

knowledgeBaseId

The unique identifier for a knowledge base attached to the agent.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10.

Pattern: `^[0-9a-zA-Z]+$`

Required: Yes

retrievalConfiguration

The configurations to apply to the knowledge base during query. For more information, see [Query configurations](#).

Type: [KnowledgeBaseRetrievalConfiguration](#) object

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

KnowledgeBaseLookupInput

Service: Agents for Amazon Bedrock Runtime

Contains details about the knowledge base to look up and the query to be made.

Contents

knowledgeBaseId

The unique identifier of the knowledge base to look up.

Type: String

Required: No

text

The query made to the knowledge base.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

KnowledgeBaseLookupOutput

Service: Agents for Amazon Bedrock Runtime

Contains details about the results from looking up the knowledge base.

Contents

retrievedReferences

Contains metadata about the sources cited for the generated response.

Type: Array of [RetrievedReference](#) objects

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

KnowledgeBaseQuery

Service: Agents for Amazon Bedrock Runtime

Contains the query made to the knowledge base.

This data type is used in the following API operations:

- [Retrieve request](#) – in the `retrievalQuery` field

Contents

text

The text of the query made to the knowledge base.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 1000.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

KnowledgeBaseRetrievalConfiguration

Service: Agents for Amazon Bedrock Runtime

Contains configurations for knowledge base query. For more information, see [Query configurations](#).

This data type is used in the following API operations:

- [Retrieve request](#) – in the `retrievalConfiguration` field
- [RetrieveAndGenerate request](#) – in the `retrievalConfiguration` field

Contents

vectorSearchConfiguration

Contains details about how the results from the vector search should be returned. For more information, see [Query configurations](#).

Type: [KnowledgeBaseVectorSearchConfiguration](#) object

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

KnowledgeBaseRetrievalResult

Service: Agents for Amazon Bedrock Runtime

Details about a result from querying the knowledge base.

This data type is used in the following API operations:

- [Retrieve response](#) – in the `retrievalResults` field

Contents

content

Contains a chunk of text from a data source in the knowledge base.

Type: [RetrievalResultContent](#) object

Required: Yes

location

Contains information about the location of the data source.

Type: [RetrievalResultLocation](#) object

Required: No

metadata

Contains metadata attributes and their values for the file in the data source. For more information, see [Metadata and filtering](#).

Type: String to JSON value map

Map Entries: Maximum number of items.

Key Length Constraints: Minimum length of 1. Maximum length of 100.

Required: No

score

The level of relevance of the result to the query.

Type: Double

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

KnowledgeBaseRetrieveAndGenerateConfiguration

Service: Agents for Amazon Bedrock Runtime

Contains details about the resource being queried.

This data type is used in the following API operations:

- [Retrieve request](#) – in the `knowledgeBaseConfiguration` field
- [RetrieveAndGenerate request](#) – in the `knowledgeBaseConfiguration` field

Contents

`knowledgeBaseId`

The unique identifier of the knowledge base that is queried and the foundation model used for generation.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10.

Pattern: `^[0-9a-zA-Z]+$`

Required: Yes

`modelArn`

The ARN of the foundation model used to generate a response.

Type: String

Length Constraints: Minimum length of 20. Maximum length of 1011.

Pattern: `^arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:(([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}/[a-z0-9]{12})|(:foundation-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}))$`

Required: Yes

`generationConfiguration`

Contains configurations for response generation based on the knowledge base query results.

Type: [GenerationConfiguration](#) object

Required: No

orchestrationConfiguration

Settings for how the model processes the prompt prior to retrieval and generation.

Type: [OrchestrationConfiguration](#) object

Required: No

retrievalConfiguration

Contains configurations for how to retrieve and return the knowledge base query.

Type: [KnowledgeBaseRetrievalConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

KnowledgeBaseVectorSearchConfiguration

Service: Agents for Amazon Bedrock Runtime

Configurations for how to perform the search query and return results. For more information, see [Query configurations](#).

This data type is used in the following API operations:

- [Retrieve request](#) – in the `vectorSearchConfiguration` field
- [RetrieveAndGenerate request](#) – in the `vectorSearchConfiguration` field

Contents

filter

Specifies the filters to use on the metadata in the knowledge base data sources before returning results. For more information, see [Query configurations](#).

Type: [RetrievalFilter](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

numberOfResults

The number of source chunks to retrieve.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

overrideSearchType

By default, Amazon Bedrock decides a search strategy for you. If you're using an Amazon OpenSearch Serverless vector store that contains a filterable text field, you can specify whether to query the knowledge base with a HYBRID search using both vector embeddings and raw text, or SEMANTIC search using only vector embeddings. For other vector store configurations, only SEMANTIC search is available. For more information, see [Test a knowledge base](#).

Type: String

Valid Values: HYBRID | SEMANTIC

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

Memory

Service: Agents for Amazon Bedrock Runtime

Contains sessions summaries.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

sessionSummary

Contains summary of a session.

Type: [MemorySessionSummary](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

MemorySessionSummary

Service: Agents for Amazon Bedrock Runtime

Contains details of a session summary.

Contents

memoryId

The unique identifier of the memory where the session summary is stored.

Type: String

Length Constraints: Minimum length of 2. Maximum length of 100.

Pattern: `^[0-9a-zA-Z._:-]+$`

Required: No

sessionExpiryTime

The time when the memory duration for the session is set to end.

Type: Timestamp

Required: No

sessionId

The identifier for this session.

Type: String

Length Constraints: Minimum length of 2. Maximum length of 100.

Pattern: `^[0-9a-zA-Z._:-]+$`

Required: No

sessionStartTime

The start time for this session.

Type: Timestamp

Required: No

summaryText

The summarized text for this session.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 25000000.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ModelInvocationInput

Service: Agents for Amazon Bedrock Runtime

The input for the pre-processing step.

- The type matches the agent step.
- The text contains the prompt.
- The `inferenceConfiguration`, `parserMode`, and `overrideLambda` values are set in the [PromptOverrideConfiguration](#) object that was set when the agent was created or updated.

Contents

`inferenceConfiguration`

Specifications about the inference parameters that were provided alongside the prompt. These are specified in the [PromptOverrideConfiguration](#) object that was set when the agent was created or updated. For more information, see [Inference parameters for foundation models](#).

Type: [InferenceConfiguration](#) object

Required: No

`overrideLambda`

The ARN of the Lambda function to use when parsing the raw foundation model output in parts of the agent sequence.

Type: String

Required: No

`parserMode`

Specifies whether to override the default parser Lambda function when parsing the raw foundation model output in the part of the agent sequence defined by the `promptType`.

Type: String

Valid Values: DEFAULT | OVERRIDDEN

Required: No

promptCreationMode

Specifies whether the default prompt template was OVERRIDDEN. If it was, the `basePromptTemplate` that was set in the [PromptOverrideConfiguration](#) object when the agent was created or updated is used instead.

Type: String

Valid Values: DEFAULT | OVERRIDDEN

Required: No

text

The text that prompted the agent at this step.

Type: String

Required: No

traceId

The unique identifier of the trace.

Type: String

Length Constraints: Minimum length of 2. Maximum length of 16.

Required: No

type

The step in the agent sequence.

Type: String

Valid Values: PRE_PROCESSING | ORCHESTRATION |
KNOWLEDGE_BASE_RESPONSE_GENERATION | POST_PROCESSING

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

Observation

Service: Agents for Amazon Bedrock Runtime

Contains the result or output of an action group or knowledge base, or the response to the user.

Contents

actionGroupInvocationOutput

Contains the JSON-formatted string returned by the API invoked by the action group.

Type: [ActionGroupInvocationOutput](#) object

Required: No

codeInterpreterInvocationOutput

Contains the JSON-formatted string returned by the API invoked by the code interpreter.

Type: [CodeInterpreterInvocationOutput](#) object

Required: No

finalResponse

Contains details about the response to the user.

Type: [FinalResponse](#) object

Required: No

knowledgeBaseLookupOutput

Contains details about the results from looking up the knowledge base.

Type: [KnowledgeBaseLookupOutput](#) object

Required: No

repromptResponse

Contains details about the response to reprompt the input.

Type: [RepromptResponse](#) object

Required: No

traceId

The unique identifier of the trace.

Type: String

Length Constraints: Minimum length of 2. Maximum length of 16.

Required: No

type

Specifies what kind of information the agent returns in the observation. The following values are possible.

- ACTION_GROUP – The agent returns the result of an action group.
- KNOWLEDGE_BASE – The agent returns information from a knowledge base.
- FINISH – The agent returns a final response to the user with no follow-up.
- ASK_USER – The agent asks the user a question.
- REPROMPT – The agent prompts the user again for the same information.

Type: String

Valid Values: ACTION_GROUP | KNOWLEDGE_BASE | FINISH | ASK_USER | REPROMPT

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

OrchestrationConfiguration

Service: Agents for Amazon Bedrock Runtime

Settings for how the model processes the prompt prior to retrieval and generation.

Contents

queryTransformationConfiguration

To split up the prompt and retrieve multiple sources, set the transformation type to `QUERY_DECOMPOSITION`.

Type: [QueryTransformationConfiguration](#) object

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

OrchestrationTrace

Service: Agents for Amazon Bedrock Runtime

Details about the orchestration step, in which the agent determines the order in which actions are executed and which knowledge bases are retrieved.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

invocationInput

Contains information pertaining to the action group or knowledge base that is being invoked.

Type: [InvocationInput](#) object

Required: No

modelInvocationInput

The input for the orchestration step.

- The type is ORCHESTRATION.
- The text contains the prompt.
- The `inferenceConfiguration`, `parserMode`, and `overrideLambda` values are set in the [PromptOverrideConfiguration](#) object that was set when the agent was created or updated.

Type: [ModelInvocationInput](#) object

Required: No

observation

Details about the observation (the output of the action group Lambda or knowledge base) made by the agent.

Type: [Observation](#) object

Required: No

rationale

Details about the reasoning, based on the input, that the agent uses to justify carrying out an action group or getting information from a knowledge base.

Type: [Rationale](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

OutputFile

Service: Agents for Amazon Bedrock Runtime

Contains details of the response from code interpreter.

Contents

bytes

The byte count of files that contains response from code interpreter.

Type: Base64-encoded binary data object

Length Constraints: Minimum length of 0. Maximum length of 1000000.

Required: No

name

The name of the file containing response from code interpreter.

Type: String

Required: No

type

The type of file that contains response from the code interpreter.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

Parameter

Service: Agents for Amazon Bedrock Runtime

A parameter for the API request or function.

Contents

name

The name of the parameter.

Type: String

Required: No

type

The type of the parameter.

Type: String

Required: No

value

The value of the parameter.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PayloadPart

Service: Agents for Amazon Bedrock Runtime

Contains a part of an agent response and citations for it.

Contents

attribution

Contains citations for a part of an agent response.

Type: [Attribution](#) object

Required: No

bytes

A part of the agent response in bytes.

Type: Base64-encoded binary data object

Length Constraints: Minimum length of 0. Maximum length of 1000000.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PostProcessingModelInvocationOutput

Service: Agents for Amazon Bedrock Runtime

The foundation model output from the post-processing step.

Contents

parsedResponse

Details about the response from the Lambda parsing of the output of the post-processing step.

Type: [PostProcessingParsedResponse](#) object

Required: No

traceId

The unique identifier of the trace.

Type: String

Length Constraints: Minimum length of 2. Maximum length of 16.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PostProcessingParsedResponse

Service: Agents for Amazon Bedrock Runtime

Details about the response from the Lambda parsing of the output from the post-processing step.

Contents

text

The text returned by the parser.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PostProcessingTrace

Service: Agents for Amazon Bedrock Runtime

Details about the post-processing step, in which the agent shapes the response.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

modelInvocationInput

The input for the post-processing step.

- The type is POST_PROCESSING.
- The text contains the prompt.
- The inferenceConfiguration, parserMode, and overrideLambda values are set in the [PromptOverrideConfiguration](#) object that was set when the agent was created or updated.

Type: [ModelInvocationInput](#) object

Required: No

modelInvocationOutput

The foundation model output from the post-processing step.

Type: [PostProcessingModelInvocationOutput](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)

- [AWS SDK for Ruby V3](#)

PreProcessingModelInvocationOutput

Service: Agents for Amazon Bedrock Runtime

The foundation model output from the pre-processing step.

Contents

parsedResponse

Details about the response from the Lambda parsing of the output of the pre-processing step.

Type: [PreProcessingParsedResponse](#) object

Required: No

traceId

The unique identifier of the trace.

Type: String

Length Constraints: Minimum length of 2. Maximum length of 16.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PreProcessingParsedResponse

Service: Agents for Amazon Bedrock Runtime

Details about the response from the Lambda parsing of the output from the pre-processing step.

Contents

isValid

Whether the user input is valid or not. If `false`, the agent doesn't proceed to orchestration.

Type: Boolean

Required: No

rationale

The text returned by the parsing of the pre-processing step, explaining the steps that the agent plans to take in orchestration, if the user input is valid.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PreProcessingTrace

Service: Agents for Amazon Bedrock Runtime

Details about the pre-processing step, in which the agent contextualizes and categorizes user inputs.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

modelInvocationInput

The input for the pre-processing step.

- The type is PRE_PROCESSING.
- The text contains the prompt.
- The inferenceConfiguration, parserMode, and overrideLambda values are set in the [PromptOverrideConfiguration](#) object that was set when the agent was created or updated.

Type: [ModelInvocationInput](#) object

Required: No

modelInvocationOutput

The foundation model output from the pre-processing step.

Type: [PreProcessingModelInvocationOutput](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)

- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PromptTemplate

Service: Agents for Amazon Bedrock Runtime

Contains the template for the prompt that's sent to the model for response generation. For more information, see [Knowledge base prompt templates](#).

This data type is used in the following API operations:

- [RetrieveAndGenerate request](#) – in the `filter` field

Contents

textPromptTemplate

The template for the prompt that's sent to the model for response generation. You can include prompt placeholders, which become replaced before the prompt is sent to the model to provide instructions and context to the model. In addition, you can include XML tags to delineate meaningful sections of the prompt template.

For more information, see the following resources:

- [Knowledge base prompt templates](#)
- [Use XML tags with Anthropic Claude models](#)

Type: String

Length Constraints: Minimum length of 1. Maximum length of 4000.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PropertyParameters

Service: Agents for Amazon Bedrock Runtime

Contains the parameters in the request body.

Contents

properties

A list of parameters in the request body.

Type: Array of [Parameter](#) objects

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

QueryTransformationConfiguration

Service: Agents for Amazon Bedrock Runtime

To split up the prompt and retrieve multiple sources, set the transformation type to `QUERY_DECOMPOSITION`.

Contents

type

The type of transformation to apply to the prompt.

Type: String

Valid Values: `QUERY_DECOMPOSITION`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

Rationale

Service: Agents for Amazon Bedrock Runtime

Contains the reasoning, based on the input, that the agent uses to justify carrying out an action group or getting information from a knowledge base.

Contents

text

The reasoning or thought process of the agent, based on the input.

Type: String

Required: No

traceId

The unique identifier of the trace step.

Type: String

Length Constraints: Minimum length of 2. Maximum length of 16.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RepromptResponse

Service: Agents for Amazon Bedrock Runtime

Contains details about the agent's response to reprompt the input.

Contents

source

Specifies what output is prompting the agent to reprompt the input.

Type: String

Valid Values: ACTION_GROUP | KNOWLEDGE_BASE | PARSER

Required: No

text

The text reprompting the input.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RequestBody

Service: Agents for Amazon Bedrock Runtime

The parameters in the API request body.

Contents

content

The content in the request body.

Type: String to array of [Parameter](#) objects map

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ResponseStream

Service: Agents for Amazon Bedrock Runtime

The response from invoking the agent and associated citations and trace information.

Contents

accessDeniedException

The request is denied because of missing access permissions. Check your permissions and retry your request.

Type: Exception

HTTP Status Code: 403

Required: No

badGatewayException

There was an issue with a dependency due to a server issue. Retry your request.

Type: Exception

HTTP Status Code: 502

Required: No

chunk

Contains a part of an agent response and citations for it.

Type: [PayloadPart](#) object

Required: No

conflictException

There was a conflict performing an operation. Resolve the conflict and retry your request.

Type: Exception

HTTP Status Code: 409

Required: No

dependencyFailedException

There was an issue with a dependency. Check the resource configurations and retry the request.

Type: Exception
HTTP Status Code: 424

Required: No

files

Contains intermediate response for code interpreter if any files have been generated.

Type: [FilePart](#) object

Required: No

internalServerErrorException

An internal server error occurred. Retry your request.

Type: Exception
HTTP Status Code: 500

Required: No

resourceNotFoundException

The specified resource Amazon Resource Name (ARN) was not found. Check the Amazon Resource Name (ARN) and try your request again.

Type: Exception
HTTP Status Code: 404

Required: No

returnControl

Contains the parameters and information that the agent elicited from the customer to carry out an action. This information is returned to the system and can be used in your own setup for fulfilling the action.

Type: [ReturnControlPayload](#) object

Required: No

serviceQuotaExceededException

The number of requests exceeds the service quota. Resubmit your request later.

Type: Exception
HTTP Status Code: 400

Required: No

throttlingException

The number of requests exceeds the limit. Resubmit your request later.

Type: Exception
HTTP Status Code: 429

Required: No

trace

Contains information about the agent and session, alongside the agent's reasoning process and results from calling actions and querying knowledge bases and metadata about the trace. You can use the trace to understand how the agent arrived at the response it provided the customer. For more information, see [Trace events](#).

Type: [TracePart](#) object

Required: No

validationException

Input validation failed. Check your request parameters and retry the request.

Type: Exception
HTTP Status Code: 400

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RetrievalFilter

Service: Agents for Amazon Bedrock Runtime

Specifies the filters to use on the metadata attributes in the knowledge base data sources before returning results. For more information, see [Query configurations](#). See the examples below to see how to use these filters.

This data type is used in the following API operations:

- [Retrieve request](#) – in the `filter` field
- [RetrieveAndGenerate request](#) – in the `filter` field

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

andAll

Knowledge base data sources are returned if their metadata attributes fulfill all the filter conditions inside this list.

Type: Array of [RetrievalFilter](#) objects

Array Members: Minimum number of 2 items. Maximum number of 5 items.

Required: No

equals

Knowledge base data sources are returned if they contain a metadata attribute whose name matches the `key` and whose value matches the `value` in this object.

The following example would return data sources with an `animal` attribute whose value is `cat`:

```
"equals": { "key": "animal", "value": "cat" }
```

Type: [FilterAttribute](#) object

Required: No

greaterThan

Knowledge base data sources are returned if they contain a metadata attribute whose name matches the key and whose value is greater than the value in this object.

The following example would return data sources with an year attribute whose value is greater than 1989:

```
"greaterThan": { "key": "year", "value": 1989 }
```

Type: [FilterAttribute](#) object

Required: No

greaterThanOrEqualTo

Knowledge base data sources are returned if they contain a metadata attribute whose name matches the key and whose value is greater than or equal to the value in this object.

The following example would return data sources with an year attribute whose value is greater than or equal to 1989:

```
"greaterThanOrEqualTo": { "key": "year", "value": 1989 }
```

Type: [FilterAttribute](#) object

Required: No

in

Knowledge base data sources are returned if they contain a metadata attribute whose name matches the key and whose value is in the list specified in the value in this object.

The following example would return data sources with an animal attribute that is either cat or dog:

```
"in": { "key": "animal", "value": ["cat", "dog"] }
```

Type: [FilterAttribute](#) object

Required: No

lessThan

Knowledge base data sources are returned if they contain a metadata attribute whose name matches the key and whose value is less than the `value` in this object.

The following example would return data sources with an `year` attribute whose value is less than to 1989.

```
"lessThan": { "key": "year", "value": 1989 }
```

Type: [FilterAttribute](#) object

Required: No

lessThanOrEquals

Knowledge base data sources are returned if they contain a metadata attribute whose name matches the key and whose value is less than or equal to the `value` in this object.

The following example would return data sources with an `year` attribute whose value is less than or equal to 1989.

```
"lessThanOrEquals": { "key": "year", "value": 1989 }
```

Type: [FilterAttribute](#) object

Required: No

listContains

Knowledge base data sources are returned if they contain a metadata attribute whose name matches the key and whose value is a list that contains the `value` as one of its members.

The following example would return data sources with an `animals` attribute that is a list containing a `cat` member (for example `["dog", "cat"]`).

```
"listContains": { "key": "animals", "value": "cat" }
```

Type: [FilterAttribute](#) object

Required: No

notEquals

Knowledge base data sources that contain a metadata attribute whose name matches the key and whose value doesn't match the `value` in this object are returned.

The following example would return data sources that don't contain an `animal` attribute whose value is `cat`.

```
"notEquals": { "key": "animal", "value": "cat" }
```

Type: [FilterAttribute](#) object

Required: No

notIn

Knowledge base data sources are returned if they contain a metadata attribute whose name matches the key and whose value isn't in the list specified in the `value` in this object.

The following example would return data sources whose `animal` attribute is neither `cat` nor `dog`.

```
"notIn": { "key": "animal", "value": ["cat", "dog"] }
```

Type: [FilterAttribute](#) object

Required: No

orAll

Knowledge base data sources are returned if their metadata attributes fulfill at least one of the filter conditions inside this list.

Type: Array of [RetrievalFilter](#) objects

Array Members: Minimum number of 2 items. Maximum number of 5 items.

Required: No

startsWith

Knowledge base data sources are returned if they contain a metadata attribute whose name matches the key and whose value starts with the `value` in this object. This filter is currently only supported for Amazon OpenSearch Serverless vector stores.

The following example would return data sources with an `animal` attribute starts with `ca` (for example, `cat` or `camel`).

```
"startsWith": { "key": "animal", "value": "ca" }
```

Type: [FilterAttribute](#) object

Required: No

stringContains

Knowledge base data sources are returned if they contain a metadata attribute whose name matches the key and whose value is one of the following:

- A string that contains the value as a substring. The following example would return data sources with an `animal` attribute that contains the substring `at` (for example `cat`).

```
"stringContains": { "key": "animal", "value": "at" }
```

- A list with a member that contains the value as a substring. The following example would return data sources with an `animals` attribute that is a list containing a member that contains the substring `at` (for example `["dog", "cat"]`).

```
"stringContains": { "key": "animals", "value": "at" }
```

Type: [FilterAttribute](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RetrievalResultConfluenceLocation

Service: Agents for Amazon Bedrock Runtime

The Confluence data source location.

Contents

url

The Confluence host URL for the data source location.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RetrievalResultContent

Service: Agents for Amazon Bedrock Runtime

Contains the cited text from the data source.

This data type is used in the following API operations:

- [Retrieve response](#) – in the content field
- [RetrieveAndGenerate response](#) – in the content field
- [InvokeAgent response](#) – in the content field

Contents

text

The cited text from the data source.

Type: String

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RetrievalResultLocation

Service: Agents for Amazon Bedrock Runtime

Contains information about the data source location.

This data type is used in the following API operations:

- [Retrieve response](#) – in the location field
- [RetrieveAndGenerate response](#) – in the location field
- [InvokeAgent response](#) – in the location field

Contents

type

The type of data source location.

Type: String

Valid Values: S3 | WEB | CONFLUENCE | SALESFORCE | SHAREPOINT

Required: Yes

confluenceLocation

The Confluence data source location.

Type: [RetrievalResultConfluenceLocation](#) object

Required: No

s3Location

The S3 data source location.

Type: [RetrievalResultS3Location](#) object

Required: No

salesforceLocation

The Salesforce data source location.

Type: [RetrievalResultSalesforceLocation](#) object

Required: No

sharePointLocation

The SharePoint data source location.

Type: [RetrievalResultSharePointLocation](#) object

Required: No

webLocation

The web URL/URLs data source location.

Type: [RetrievalResultWebLocation](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RetrievalResultS3Location

Service: Agents for Amazon Bedrock Runtime

The S3 data source location.

This data type is used in the following API operations:

- [Retrieve response](#) – in the s3Location field
- [RetrieveAndGenerate response](#) – in the s3Location field
- [InvokeAgent response](#) – in the s3Location field

Contents

uri

The S3 URI for the data source location.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RetrievalResultSalesforceLocation

Service: Agents for Amazon Bedrock Runtime

The Salesforce data source location.

Contents

url

The Salesforce host URL for the data source location.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RetrievalResultSharePointLocation

Service: Agents for Amazon Bedrock Runtime

The SharePoint data source location.

Contents

url

The SharePoint site URL for the data source location.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RetrievalResultWebLocation

Service: Agents for Amazon Bedrock Runtime

The web URL/URLs data source location.

Contents

url

The web URL/URLs for the data source location.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RetrieveAndGenerateConfiguration

Service: Agents for Amazon Bedrock Runtime

Contains details about the resource being queried.

This data type is used in the following API operations:

- [RetrieveAndGenerate request](#) – in the `retrieveAndGenerateConfiguration` field

Contents

type

The type of resource that is queried by the request.

Type: String

Valid Values: KNOWLEDGE_BASE | EXTERNAL_SOURCES

Required: Yes

externalSourcesConfiguration

The configuration used with the external source wrapper object in the `retrieveAndGenerate` function.

Type: [ExternalSourcesRetrieveAndGenerateConfiguration](#) object

Required: No

knowledgeBaseConfiguration

Contains details about the resource being queried.

Type: [KnowledgeBaseRetrieveAndGenerateConfiguration](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RetrieveAndGenerateInput

Service: Agents for Amazon Bedrock Runtime

Contains the query made to the knowledge base.

This data type is used in the following API operations:

- [RetrieveAndGenerate request](#) – in the `input` field

Contents

text

The query made to the knowledge base.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 1000.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RetrieveAndGenerateOutput

Service: Agents for Amazon Bedrock Runtime

Contains the response generated from querying the knowledge base.

This data type is used in the following API operations:

- [RetrieveAndGenerate response](#) – in the output field

Contents

text

The response generated from querying the knowledge base.

Type: String

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RetrieveAndGenerateSessionConfiguration

Service: Agents for Amazon Bedrock Runtime

Contains configuration about the session with the knowledge base.

This data type is used in the following API operations:

- [RetrieveAndGenerate request](#) – in the sessionConfiguration field

Contents

kmsKeyArn

The ARN of the AWS KMS key encrypting the session.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(|-cn|-us-gov):kms:[a-zA-Z0-9-]*:[0-9]{12}:key/[a-zA-Z0-9-]{36}$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

RetrievedReference

Service: Agents for Amazon Bedrock Runtime

Contains metadata about a source cited for the generated response.

This data type is used in the following API operations:

- [RetrieveAndGenerate response](#) – in the `retrievedReferences` field
- [InvokeAgent response](#) – in the `retrievedReferences` field

Contents

content

Contains the cited text from the data source.

Type: [RetrievalResultContent](#) object

Required: No

location

Contains information about the location of the data source.

Type: [RetrievalResultLocation](#) object

Required: No

metadata

Contains metadata attributes and their values for the file in the data source. For more information, see [Metadata and filtering](#).

Type: String to JSON value map

Map Entries: Maximum number of items.

Key Length Constraints: Minimum length of 1. Maximum length of 100.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ReturnControlPayload

Service: Agents for Amazon Bedrock Runtime

Contains information to return from the action group that the agent has predicted to invoke.

This data type is used in the following API operations:

- [InvokeAgent response](#)

Contents

invocationId

The identifier of the action group invocation.

Type: String

Required: No

invocationInputs

A list of objects that contain information about the parameters and inputs that need to be sent into the API operation or function, based on what the agent determines from its session with the user.

Type: Array of [InvocationInputMember](#) objects

Array Members: Minimum number of 1 item. Maximum number of 5 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

S3ObjectDoc

Service: Agents for Amazon Bedrock Runtime

The unique wrapper object of the document from the S3 location.

Contents

uri

The file location of the S3 wrapper object.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1024.

Pattern: `^s3://[a-z0-9][a-z0-9.-]{1,61}[a-z0-9]/.{1,1024}$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

S3ObjectFile

Service: Agents for Amazon Bedrock Runtime

Contains details of the s3 object where the source file is located.

Contents

uri

The uri of the s3 object.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1024.

Pattern: `^s3://[a-z0-9][a-z0-9.-]{1,61}[a-z0-9]/.{1,1024}$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

SessionState

Service: Agents for Amazon Bedrock Runtime

Contains parameters that specify various attributes that persist across a session or prompt. You can define session state attributes as key-value pairs when writing a [Lambda function](#) for an action group or pass them when making an [InvokeAgent](#) request. Use session state attributes to control and provide conversational context for your agent and to help customize your agent's behavior. For more information, see [Control session context](#).

Contents

files

Contains information about the files used by code interpreter.

Type: Array of [InputFile](#) objects

Required: No

invocationId

The identifier of the invocation of an action. This value must match the `invocationId` returned in the `InvokeAgent` response for the action whose results are provided in the `returnControlInvocationResults` field. For more information, see [Return control to the agent developer](#) and [Control session context](#).

Type: String

Required: No

knowledgeBaseConfigurations

An array of configurations, each of which applies to a knowledge base attached to the agent.

Type: Array of [KnowledgeBaseConfiguration](#) objects

Array Members: Minimum number of 1 item.

Required: No

promptSessionAttributes

Contains attributes that persist across a prompt and the values of those attributes. These attributes replace the `$prompt_session_attributes$` placeholder variable in the orchestration prompt template. For more information, see [Prompt template placeholder variables](#).

Type: String to string map

Required: No

returnControlInvocationResults

Contains information about the results from the action group invocation. For more information, see [Return control to the agent developer](#) and [Control session context](#).

Note

If you include this field, the `inputText` field will be ignored.

Type: Array of [InvocationResultMember](#) objects

Array Members: Minimum number of 1 item. Maximum number of 5 items.

Required: No

sessionAttributes

Contains attributes that persist across a session and the values of those attributes.

Type: String to string map

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

Span

Service: Agents for Amazon Bedrock Runtime

Contains information about where the text with a citation begins and ends in the generated output.

This data type is used in the following API operations:

- [RetrieveAndGenerate response](#) – in the span field
- [InvokeAgent response](#) – in the span field

Contents

end

Where the text with a citation ends in the generated output.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

start

Where the text with a citation starts in the generated output.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

TextInferenceConfig

Service: Agents for Amazon Bedrock Runtime

Configuration settings for text generation using a language model via the RetrieveAndGenerate operation. Includes parameters like temperature, top-p, maximum token count, and stop sequences.

Note

The valid range of maxTokens depends on the accepted values for your chosen model's inference parameters. To see the inference parameters for your model, see [Inference parameters for foundation models](#).

Contents

maxTokens

The maximum number of tokens to generate in the output text. Do not use the minimum of 0 or the maximum of 65536. The limit values described here are arbitrary values, for actual values consult the limits defined by your specific model.

Type: Integer

Valid Range: Minimum value of 0. Maximum value of 65536.

Required: No

stopSequences

A list of sequences of characters that, if generated, will cause the model to stop generating further tokens. Do not use a minimum length of 1 or a maximum length of 1000. The limit values described here are arbitrary values, for actual values consult the limits defined by your specific model.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 4 items.

Length Constraints: Minimum length of 1. Maximum length of 1000.

Required: No

temperature

Controls the random-ness of text generated by the language model, influencing how much the model sticks to the most predictable next words versus exploring more surprising options. A lower temperature value (e.g. 0.2 or 0.3) makes model outputs more deterministic or predictable, while a higher temperature (e.g. 0.8 or 0.9) makes the outputs more creative or unpredictable.

Type: Float

Valid Range: Minimum value of 0. Maximum value of 1.

Required: No

topP

A probability distribution threshold which controls what the model considers for the set of possible next tokens. The model will only consider the top p% of the probability distribution when generating the next token.

Type: Float

Valid Range: Minimum value of 0. Maximum value of 1.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

TextResponsePart

Service: Agents for Amazon Bedrock Runtime

Contains the part of the generated text that contains a citation, alongside where it begins and ends.

This data type is used in the following API operations:

- [RetrieveAndGenerate response](#) – in the textResponsePart field
- [InvokeAgent response](#) – in the textResponsePart field

Contents

span

Contains information about where the text with a citation begins and ends in the generated output.

Type: [Span](#) object

Required: No

text

The part of the generated text that contains a citation.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

Trace

Service: Agents for Amazon Bedrock Runtime

Contains one part of the agent's reasoning process and results from calling API actions and querying knowledge bases. You can use the trace to understand how the agent arrived at the response it provided the customer. For more information, see [Trace enablement](#).

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

failureTrace

Contains information about the failure of the interaction.

Type: [FailureTrace](#) object

Required: No

guardrailTrace

The trace details for a trace defined in the Guardrail filter.

Type: [GuardrailTrace](#) object

Required: No

orchestrationTrace

Details about the orchestration step, in which the agent determines the order in which actions are executed and which knowledge bases are retrieved.

Type: [OrchestrationTrace](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

postProcessingTrace

Details about the post-processing step, in which the agent shapes the response..

Type: [PostProcessingTrace](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

preProcessingTrace

Details about the pre-processing step, in which the agent contextualizes and categorizes user inputs.

Type: [PreProcessingTrace](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

TracePart

Service: Agents for Amazon Bedrock Runtime

Contains information about the agent and session, alongside the agent's reasoning process and results from calling API actions and querying knowledge bases and metadata about the trace. You can use the trace to understand how the agent arrived at the response it provided the customer. For more information, see [Trace enablement](#).

Contents

agentAliasId

The unique identifier of the alias of the agent.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10.

Pattern: `^[0-9a-zA-Z]+$`

Required: No

agentId

The unique identifier of the agent.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10.

Pattern: `^[0-9a-zA-Z]+$`

Required: No

agentVersion

The version of the agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 5.

Pattern: `^(DRAFT|[0-9]{0,4}[1-9][0-9]{0,4})$`

Required: No

sessionId

The unique identifier of the session with the agent.

Type: String

Length Constraints: Minimum length of 2. Maximum length of 100.

Pattern: `^[0-9a-zA-Z._:-]+$`

Required: No

trace

Contains one part of the agent's reasoning process and results from calling API actions and querying knowledge bases. You can use the trace to understand how the agent arrived at the response it provided the customer. For more information, see [Trace enablement](#).

Type: [Trace](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

Amazon Bedrock Runtime

The following data types are supported by Amazon Bedrock Runtime:

- [AnyToolChoice](#)
- [AutoToolChoice](#)
- [ContentBlock](#)

- [ContentBlockDelta](#)
- [ContentBlockDeltaEvent](#)
- [ContentBlockStart](#)
- [ContentBlockStartEvent](#)
- [ContentBlockStopEvent](#)
- [ConverseMetrics](#)
- [ConverseOutput](#)
- [ConverseStreamMetadataEvent](#)
- [ConverseStreamMetrics](#)
- [ConverseStreamOutput](#)
- [ConverseStreamTrace](#)
- [ConverseTrace](#)
- [DocumentBlock](#)
- [DocumentSource](#)
- [GuardrailAssessment](#)
- [GuardrailConfiguration](#)
- [GuardrailContentBlock](#)
- [GuardrailContentFilter](#)
- [GuardrailContentPolicyAssessment](#)
- [GuardrailContextualGroundingFilter](#)
- [GuardrailContextualGroundingPolicyAssessment](#)
- [GuardrailConverseContentBlock](#)
- [GuardrailConverseTextBlock](#)
- [GuardrailCustomWord](#)
- [GuardrailManagedWord](#)
- [GuardrailOutputContent](#)
- [GuardrailPiiEntityFilter](#)
- [GuardrailRegexFilter](#)
- [GuardrailSensitiveInformationPolicyAssessment](#)
- [GuardrailStreamConfiguration](#)

- [GuardrailTextBlock](#)
- [GuardrailTopic](#)
- [GuardrailTopicPolicyAssessment](#)
- [GuardrailTraceAssessment](#)
- [GuardrailUsage](#)
- [GuardrailWordPolicyAssessment](#)
- [ImageBlock](#)
- [ImageSource](#)
- [InferenceConfiguration](#)
- [Message](#)
- [MessageStartEvent](#)
- [MessageStopEvent](#)
- [PayloadPart](#)
- [ResponseStream](#)
- [SpecificToolChoice](#)
- [SystemContentBlock](#)
- [TokenUsage](#)
- [Tool](#)
- [ToolChoice](#)
- [ToolConfiguration](#)
- [ToolInputSchema](#)
- [ToolResultBlock](#)
- [ToolResultContentBlock](#)
- [ToolSpecification](#)
- [ToolUseBlock](#)
- [ToolUseBlockDelta](#)
- [ToolUseBlockStart](#)

AnyToolChoice

Service: Amazon Bedrock Runtime

The model must request at least one tool (no text is generated). For example, {"any" : {}}.

Contents

The members of this exception structure are context-dependent.

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

AutoToolChoice

Service: Amazon Bedrock Runtime

The Model automatically decides if a tool should be called or whether to generate text instead. For example, {"auto" : {}}.

Contents

The members of this exception structure are context-dependent.

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ContentBlock

Service: Amazon Bedrock Runtime

A block of content for a message that you pass to, or receive from, a model with the [Converse](#) or [ConverseStream](#) API operations.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

document

A document to include in the message.

Type: [DocumentBlock](#) object

Required: No

guardContent

Contains the content to assess with the guardrail. If you don't specify guardContent in a call to the Converse API, the guardrail (if passed in the Converse API) assesses the entire message.

For more information, see [Use a guardrail with the Converse API](#).

Type: [GuardrailConverseContentBlock](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

image

Image to include in the message.

Note

This field is only supported by Anthropic Claude 3 models.

Type: [ImageBlock](#) object

Required: No

text

Text to include in the message.

Type: String

Required: No

toolResult

The result for a tool request that a model makes.

Type: [ToolResultBlock](#) object

Required: No

toolUse

Information about a tool use request from a model.

Type: [ToolUseBlock](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ContentBlockDelta

Service: Amazon Bedrock Runtime

A block of content in a streaming response.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

text

The content text.

Type: String

Required: No

toolUse

Information about a tool that the model is requesting to use.

Type: [ToolUseBlockDelta](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ContentBlockDeltaEvent

Service: Amazon Bedrock Runtime

The content block delta event.

Contents

contentBlockIndex

The block index for a content block delta event.

Type: Integer

Valid Range: Minimum value of 0.

Required: Yes

delta

The delta for a content block delta event.

Type: [ContentBlockDelta](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ContentBlockStart

Service: Amazon Bedrock Runtime

Content block start information.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

toolUse

Information about a tool that the model is requesting to use.

Type: [ToolUseBlockStart](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ContentBlockStartEvent

Service: Amazon Bedrock Runtime

Content block start event.

Contents

contentBlockIndex

The index for a content block start event.

Type: Integer

Valid Range: Minimum value of 0.

Required: Yes

start

Start information about a content block start event.

Type: [ContentBlockStart](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ContentBlockStopEvent

Service: Amazon Bedrock Runtime

A content block stop event.

Contents

contentBlockIndex

The index for a content block.

Type: Integer

Valid Range: Minimum value of 0.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ConverseMetrics

Service: Amazon Bedrock Runtime

Metrics for a call to [Converse](#).

Contents

latencyMs

The latency of the call to `Converse`, in milliseconds.

Type: Long

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ConverseOutput

Service: Amazon Bedrock Runtime

The output from a call to [Converse](#).

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

message

The message that the model generates.

Type: [Message](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ConverseStreamMetadataEvent

Service: Amazon Bedrock Runtime

A conversation stream metadata event.

Contents

metrics

The metrics for the conversation stream metadata event.

Type: [ConverseStreamMetrics](#) object

Required: Yes

usage

Usage information for the conversation stream event.

Type: [TokenUsage](#) object

Required: Yes

trace

The trace object in the response from [ConverseStream](#) that contains information about the guardrail behavior.

Type: [ConverseStreamTrace](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ConverseStreamMetrics

Service: Amazon Bedrock Runtime

Metrics for the stream.

Contents

latencyMs

The latency for the streaming request, in milliseconds.

Type: Long

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ConverseStreamOutput

Service: Amazon Bedrock Runtime

The messages output stream

Contents

contentBlockDelta

The messages output content block delta.

Type: [ContentBlockDeltaEvent](#) object

Required: No

contentBlockStart

Start information for a content block.

Type: [ContentBlockStartEvent](#) object

Required: No

contentBlockStop

Stop information for a content block.

Type: [ContentBlockStopEvent](#) object

Required: No

internalServerErrorException

An internal server error occurred. Retry your request.

Type: Exception

HTTP Status Code: 500

Required: No

messageStart

Message start information.

Type: [MessageStartEvent](#) object

Required: No

messageStop

Message stop information.

Type: [MessageStopEvent](#) object

Required: No

metadata

Metadata for the converse output stream.

Type: [ConverseStreamMetadataEvent](#) object

Required: No

modelStreamErrorException

A streaming error occurred. Retry your request.

Type: Exception

HTTP Status Code: 424

Required: No

throttlingException

The number of requests exceeds the limit. Resubmit your request later.

Type: Exception

HTTP Status Code: 429

Required: No

validationException

Input validation failed. Check your request parameters and retry the request.

Type: Exception

HTTP Status Code: 400

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ConverseStreamTrace

Service: Amazon Bedrock Runtime

The trace object in a response from [ConverseStream](#). Currently, you can only trace guardrails.

Contents

guardrail

The guardrail trace object.

Type: [GuardrailTraceAssessment](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ConverseTrace

Service: Amazon Bedrock Runtime

The trace object in a response from [Converse](#). Currently, you can only trace guardrails.

Contents

guardrail

The guardrail trace object.

Type: [GuardrailTraceAssessment](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

DocumentBlock

Service: Amazon Bedrock Runtime

A document to include in a message.

Contents

format

The format of a document, or its extension.

Type: String

Valid Values: pdf | csv | doc | docx | xls | xlsx | html | txt | md

Required: Yes

name

A name for the document. The name can only contain the following characters:

- Alphanumeric characters
- Whitespace characters (no more than one in a row)
- Hyphens
- Parentheses
- Square brackets

Note

This field is vulnerable to prompt injections, because the model might inadvertently interpret it as instructions. Therefore, we recommend that you specify a neutral name.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 200.

Required: Yes

source

Contains the content of the document.

Type: [DocumentSource](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

DocumentSource

Service: Amazon Bedrock Runtime

Contains the content of a document.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

bytes

The raw bytes for the document. If you use an AWS SDK, you don't need to encode the bytes in base64.

Type: Base64-encoded binary data object

Length Constraints: Minimum length of 1.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailAssessment

Service: Amazon Bedrock Runtime

A behavior assessment of the guardrail policies used in a call to the Converse API.

Contents

contentPolicy

The content policy.

Type: [GuardrailContentPolicyAssessment](#) object

Required: No

contextualGroundingPolicy

The contextual grounding policy used for the guardrail assessment.

Type: [GuardrailContextualGroundingPolicyAssessment](#) object

Required: No

sensitiveInformationPolicy

The sensitive information policy.

Type: [GuardrailSensitiveInformationPolicyAssessment](#) object

Required: No

topicPolicy

The topic policy.

Type: [GuardrailTopicPolicyAssessment](#) object

Required: No

wordPolicy

The word policy.

Type: [GuardrailWordPolicyAssessment](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailConfiguration

Service: Amazon Bedrock Runtime

Configuration information for a guardrail that you use with the [Converse](#) operation.

Contents

guardrailIdentifier

The identifier for the guardrail.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^(([a-z0-9]+)|(arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:guardrail/[a-z0-9]+))$`

Required: Yes

guardrailVersion

The version of the guardrail.

Type: String

Pattern: `^(([1-9][0-9]{0,7})|(DRAFT))$`

Required: Yes

trace

The trace behavior for the guardrail.

Type: String

Valid Values: `enabled` | `disabled`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailContentBlock

Service: Amazon Bedrock Runtime

The content block to be evaluated by the guardrail.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

text

Text within content block to be evaluated by the guardrail.

Type: [GuardrailTextBlock](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailContentFilter

Service: Amazon Bedrock Runtime

The content filter for a guardrail.

Contents

action

The guardrail action.

Type: String

Valid Values: BLOCKED

Required: Yes

confidence

The guardrail confidence.

Type: String

Valid Values: NONE | LOW | MEDIUM | HIGH

Required: Yes

type

The guardrail type.

Type: String

Valid Values: INSULTS | HATE | SEXUAL | VIOLENCE | MISCONDUCT | PROMPT_ATTACK

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)

- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailContentPolicyAssessment

Service: Amazon Bedrock Runtime

An assessment of a content policy for a guardrail.

Contents

filters

The content policy filters.

Type: Array of [GuardrailContentFilter](#) objects

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailContextualGroundingFilter

Service: Amazon Bedrock Runtime

The details for the guardrails contextual grounding filter.

Contents

action

The action performed by the guardrails contextual grounding filter.

Type: String

Valid Values: BLOCKED | NONE

Required: Yes

score

The score generated by contextual grounding filter.

Type: Double

Valid Range: Minimum value of 0. Maximum value of 1.

Required: Yes

threshold

The threshold used by contextual grounding filter to determine whether the content is grounded or not.

Type: Double

Valid Range: Minimum value of 0. Maximum value of 1.

Required: Yes

type

The contextual grounding filter type.

Type: String

Valid Values: GROUNDING | RELEVANCE

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailContextualGroundingPolicyAssessment

Service: Amazon Bedrock Runtime

The policy assessment details for the guardrails contextual grounding filter.

Contents

filters

The filter details for the guardrails contextual grounding filter.

Type: Array of [GuardrailContextualGroundingFilter](#) objects

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailConverseContentBlock

Service: Amazon Bedrock Runtime

A content block for selective guarding with the [Converse](#) or [ConverseStream](#) API operations.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

text

The text to guard.

Type: [GuardrailConverseTextBlock](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailConverseTextBlock

Service: Amazon Bedrock Runtime

A text block that contains text that you want to assess with a guardrail. For more information, see [GuardrailConverseContentBlock](#).

Contents

text

The text that you want to guard.

Type: String

Required: Yes

qualifiers

The qualifier details for the guardrails contextual grounding filter.

Type: Array of strings

Valid Values: `grounding_source` | `query` | `guard_content`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailCustomWord

Service: Amazon Bedrock Runtime

A custom word configured in a guardrail.

Contents

action

The action for the custom word.

Type: String

Valid Values: BLOCKED

Required: Yes

match

The match for the custom word.

Type: String

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailManagedWord

Service: Amazon Bedrock Runtime

A managed word configured in a guardrail.

Contents

action

The action for the managed word.

Type: String

Valid Values: BLOCKED

Required: Yes

match

The match for the managed word.

Type: String

Required: Yes

type

The type for the managed word.

Type: String

Valid Values: PROFANITY

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailOutputContent

Service: Amazon Bedrock Runtime

The output content produced by the guardrail.

Contents

text

The specific text for the output content produced by the guardrail.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailPiiEntityFilter

Service: Amazon Bedrock Runtime

A Personally Identifiable Information (PII) entity configured in a guardrail.

Contents

action

The PII entity filter action.

Type: String

Valid Values: ANONYMIZED | BLOCKED

Required: Yes

match

The PII entity filter match.

Type: String

Required: Yes

type

The PII entity filter type.

Type: String

Valid Values: ADDRESS | AGE | AWS_ACCESS_KEY | AWS_SECRET_KEY | CA_HEALTH_NUMBER | CA_SOCIAL_INSURANCE_NUMBER | CREDIT_DEBIT_CARD_CVV | CREDIT_DEBIT_CARD_EXPIRY | CREDIT_DEBIT_CARD_NUMBER | DRIVER_ID | EMAIL | INTERNATIONAL_BANK_ACCOUNT_NUMBER | IP_ADDRESS | LICENSE_PLATE | MAC_ADDRESS | NAME | PASSWORD | PHONE | PIN | SWIFT_CODE | UK_NATIONAL_HEALTH_SERVICE_NUMBER | UK_NATIONAL_INSURANCE_NUMBER | UK_UNIQUE_TAXPAYER_REFERENCE_NUMBER | URL | USERNAME | US_BANK_ACCOUNT_NUMBER | US_BANK_ROUTING_NUMBER | US_INDIVIDUAL_TAX_IDENTIFICATION_NUMBER | US_PASSPORT_NUMBER | US_SOCIAL_SECURITY_NUMBER | VEHICLE_IDENTIFICATION_NUMBER

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailRegexFilter

Service: Amazon Bedrock Runtime

A Regex filter configured in a guardrail.

Contents

action

The region filter action.

Type: String

Valid Values: ANONYMIZED | BLOCKED

Required: Yes

match

The regex filter match.

Type: String

Required: No

name

The regex filter name.

Type: String

Required: No

regex

The regex query.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailSensitiveInformationPolicyAssessment

Service: Amazon Bedrock Runtime

The assessment for a Personally Identifiable Information (PII) policy.

Contents

piiEntities

The PII entities in the assessment.

Type: Array of [GuardrailPiiEntityFilter](#) objects

Required: Yes

regexes

The regex queries in the assessment.

Type: Array of [GuardrailRegexFilter](#) objects

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailStreamConfiguration

Service: Amazon Bedrock Runtime

Configuration information for a guardrail that you use with the [ConverseStream](#) action.

Contents

guardrailIdentifier

The identifier for the guardrail.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Pattern: `^(([a-z0-9]+)|(arn:aws(-[^\:]+)?:bedrock:[a-z0-9-]{1,20}:[0-9]{12}:guardrail/[a-z0-9]+))$`

Required: Yes

guardrailVersion

The version of the guardrail.

Type: String

Pattern: `^(([1-9][0-9]{0,7})|(DRAFT))$`

Required: Yes

streamProcessingMode

The processing mode.

The processing mode. For more information, see [Configure streaming response behavior](#).

Type: String

Valid Values: sync | async

Required: No

trace

The trace behavior for the guardrail.

Type: String

Valid Values: enabled | disabled

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailTextBlock

Service: Amazon Bedrock Runtime

The text block to be evaluated by the guardrail.

Contents

text

The input text details to be evaluated by the guardrail.

Type: String

Required: Yes

qualifiers

The qualifiers describing the text block.

Type: Array of strings

Valid Values: `grounding_source` | `query` | `guard_content`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailTopic

Service: Amazon Bedrock Runtime

Information about a topic guardrail.

Contents

action

The action the guardrail should take when it intervenes on a topic.

Type: String

Valid Values: BLOCKED

Required: Yes

name

The name for the guardrail.

Type: String

Required: Yes

type

The type behavior that the guardrail should perform when the model detects the topic.

Type: String

Valid Values: DENY

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailTopicPolicyAssessment

Service: Amazon Bedrock Runtime

A behavior assessment of a topic policy.

Contents

topics

The topics in the assessment.

Type: Array of [GuardrailTopic](#) objects

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailTraceAssessment

Service: Amazon Bedrock Runtime

A Top level guardrail trace object. For more information, see [ConverseTrace](#).

Contents

inputAssessment

The input assessment.

Type: String to [GuardrailAssessment](#) object map

Required: No

modelOutput

The output from the model.

Type: Array of strings

Required: No

outputAssessments

the output assessments.

Type: String to array of [GuardrailAssessment](#) objects map

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailUsage

Service: Amazon Bedrock Runtime

The details on the use of the guardrail.

Contents

contentPolicyUnits

The content policy units processed by the guardrail.

Type: Integer

Required: Yes

contextualGroundingPolicyUnits

The contextual grounding policy units processed by the guardrail.

Type: Integer

Required: Yes

sensitiveInformationPolicyFreeUnits

The sensitive information policy free units processed by the guardrail.

Type: Integer

Required: Yes

sensitiveInformationPolicyUnits

The sensitive information policy units processed by the guardrail.

Type: Integer

Required: Yes

topicPolicyUnits

The topic policy units processed by the guardrail.

Type: Integer

Required: Yes

wordPolicyUnits

The word policy units processed by the guardrail.

Type: Integer

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

GuardrailWordPolicyAssessment

Service: Amazon Bedrock Runtime

The word policy assessment.

Contents

customWords

Custom words in the assessment.

Type: Array of [GuardrailCustomWord](#) objects

Required: Yes

managedWordLists

Managed word lists in the assessment.

Type: Array of [GuardrailManagedWord](#) objects

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ImageBlock

Service: Amazon Bedrock Runtime

Image content for a message.

Contents

format

The format of the image.

Type: String

Valid Values: png | jpeg | gif | webp

Required: Yes

source

The source for the image.

Type: [ImageSource](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ImageSource

Service: Amazon Bedrock Runtime

The source for an image.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

bytes

The raw image bytes for the image. If you use an AWS SDK, you don't need to encode the image bytes in base64.

Type: Base64-encoded binary data object

Length Constraints: Minimum length of 1.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

InferenceConfiguration

Service: Amazon Bedrock Runtime

Base inference parameters to pass to a model in a call to [Converse](#) or [ConverseStream](#). For more information, see [Inference parameters for foundation models](#).

If you need to pass additional parameters that the model supports, use the `additionalModelRequestFields` request field in the call to `Converse` or `ConverseStream`. For more information, see [Model parameters](#).

Contents

maxTokens

The maximum number of tokens to allow in the generated response. The default value is the maximum allowed value for the model that you are using. For more information, see [Inference parameters for foundation models](#).

Type: Integer

Valid Range: Minimum value of 1.

Required: No

stopSequences

A list of stop sequences. A stop sequence is a sequence of characters that causes the model to stop generating the response.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 4 items.

Length Constraints: Minimum length of 1.

Required: No

temperature

The likelihood of the model selecting higher-probability options while generating a response. A lower value makes the model more likely to choose higher-probability options, while a higher value makes the model more likely to choose lower-probability options.

The default value is the default value for the model that you are using. For more information, see [Inference parameters for foundation models](#).

Type: Float

Valid Range: Minimum value of 0. Maximum value of 1.

Required: No

topP

The percentage of most-likely candidates that the model considers for the next token. For example, if you choose a value of 0.8 for topP, the model selects from the top 80% of the probability distribution of tokens that could be next in the sequence.

The default value is the default value for the model that you are using. For more information, see [Inference parameters for foundation models](#).

Type: Float

Valid Range: Minimum value of 0. Maximum value of 1.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

Message

Service: Amazon Bedrock Runtime

A message input, or returned from, a call to [Converse](#) or [ConverseStream](#).

Contents

content

The message content. Note the following restrictions:

- You can include up to 20 images. Each image's size, height, and width must be no more than 3.75 MB, 8000 px, and 8000 px, respectively.
- You can include up to five documents. Each document's size must be no more than 4.5 MB.
- If you include a `ContentBlock` with a `document` field in the array, you must also include a `ContentBlock` with a `text` field.
- You can only include images and documents if the `role` is `user`.

Type: Array of [ContentBlock](#) objects

Required: Yes

role

The role that the message plays in the message.

Type: String

Valid Values: `user` | `assistant`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

MessageStartEvent

Service: Amazon Bedrock Runtime

The start of a message.

Contents

role

The role for the message.

Type: String

Valid Values: user | assistant

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

MessageStopEvent

Service: Amazon Bedrock Runtime

The stop event for a message.

Contents

stopReason

The reason why the model stopped generating output.

Type: String

Valid Values: end_turn | tool_use | max_tokens | stop_sequence | guardrail_intervened | content_filtered

Required: Yes

additionalModelResponseFields

The additional model response fields.

Type: JSON value

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

PayloadPart

Service: Amazon Bedrock Runtime

Payload content included in the response.

Contents

bytes

Base64-encoded bytes of payload data.

Type: Base64-encoded binary data object

Length Constraints: Minimum length of 0. Maximum length of 1000000.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ResponseStream

Service: Amazon Bedrock Runtime

Definition of content in the response stream.

Contents

chunk

Content included in the response.

Type: [PayloadPart](#) object

Required: No

internalServerErrorException

An internal server error occurred. Retry your request.

Type: Exception

HTTP Status Code: 500

Required: No

modelStreamErrorException

An error occurred while streaming the response. Retry your request.

Type: Exception

HTTP Status Code: 424

Required: No

modelTimeoutException

The request took too long to process. Processing time exceeded the model timeout length.

Type: Exception

HTTP Status Code: 408

Required: No

throttlingException

Your request was throttled because of service-wide limitations. Resubmit your request later or in a different region. You can also purchase [Provisioned Throughput](#) to increase the rate or number of tokens you can process.

Type: Exception

HTTP Status Code: 429

Required: No

validationException

Input validation failed. Check your request parameters and retry the request.

Type: Exception

HTTP Status Code: 400

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

SpecificToolChoice

Service: Amazon Bedrock Runtime

The model must request a specific tool. For example, {"tool" : {"name" : "Your tool name"}}.

Note

This field is only supported by Anthropic Claude 3 models.

Contents

name

The name of the tool that the model must request.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: `^[a-zA-Z][a-zA-Z0-9_]*$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

SystemContentBlock

Service: Amazon Bedrock Runtime

A system content block.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

guardContent

A content block to assess with the guardrail. Use with the [Converse](#) or [ConverseStream](#) API operations.

For more information, see [Use a guardrail with the Converse API](#).

Type: [GuardrailConverseContentBlock](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

text

A system prompt for the model.

Type: String

Length Constraints: Minimum length of 1.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)

- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

TokenUsage

Service: Amazon Bedrock Runtime

The tokens used in a message API inference call.

Contents

inputTokens

The number of tokens sent in the request to the model.

Type: Integer

Valid Range: Minimum value of 0.

Required: Yes

outputTokens

The number of tokens that the model generated for the request.

Type: Integer

Valid Range: Minimum value of 0.

Required: Yes

totalTokens

The total of input tokens and tokens generated by the model.

Type: Integer

Valid Range: Minimum value of 0.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)

- [AWS SDK for Ruby V3](#)

Tool

Service: Amazon Bedrock Runtime

Information about a tool that you can use with the Converse API. For more information, see [Tool use \(function calling\)](#) in the Amazon Bedrock User Guide.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

toolSpec

The specification for the tool.

Type: [ToolSpecification](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ToolChoice

Service: Amazon Bedrock Runtime

Determines which tools the model should request in a call to `Converse` or `ConverseStream`. `ToolChoice` is only supported by Anthropic Claude 3 models and by Mistral AI Mistral Large.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

any

The model must request at least one tool (no text is generated).

Type: [AnyToolChoice](#) object

Required: No

auto

(Default). The Model automatically decides if a tool should be called or whether to generate text instead.

Type: [AutoToolChoice](#) object

Required: No

tool

The Model must request the specified tool. Only supported by Anthropic Claude 3 models.

Type: [SpecificToolChoice](#) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ToolConfiguration

Service: Amazon Bedrock Runtime

Configuration information for the tools that you pass to a model. For more information, see [Tool use \(function calling\)](#) in the Amazon Bedrock User Guide.

Note

This field is only supported by Anthropic Claude 3, Cohere Command R, Cohere Command R+, and Mistral Large models.

Contents

tools

An array of tools that you want to pass to a model.

Type: Array of [Tool](#) objects

Array Members: Minimum number of 1 item.

Required: Yes

toolChoice

If supported by model, forces the model to request a tool.

Type: [ToolChoice](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)

- [AWS SDK for Ruby V3](#)

ToolInputSchema

Service: Amazon Bedrock Runtime

The schema for the tool. The top level schema type must be object.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

json

The JSON schema for the tool. For more information, see [JSON Schema Reference](#).

Type: JSON value

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ToolResultBlock

Service: Amazon Bedrock Runtime

A tool result block that contains the results for a tool request that the model previously made.

Contents

content

The content for tool result content block.

Type: Array of [ToolResultContentBlock](#) objects

Required: Yes

toolUseId

The ID of the tool request that this is the result for.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: `^[a-zA-Z0-9_-]+$`

Required: Yes

status

The status for the tool result content block.

Note

This field is only supported Anthropic Claude 3 models.

Type: String

Valid Values: `success` | `error`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ToolResultContentBlock

Service: Amazon Bedrock Runtime

The tool result content block.

Contents

Important

This data type is a UNION, so only one of the following members can be specified when used or returned.

document

A tool result that is a document.

Type: [DocumentBlock](#) object

Required: No

image

A tool result that is an image.

Note

This field is only supported by Anthropic Claude 3 models.

Type: [ImageBlock](#) object

Required: No

json

A tool result that is JSON format data.

Type: JSON value

Required: No

text

A tool result that is text.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ToolSpecification

Service: Amazon Bedrock Runtime

The specification for the tool.

Contents

inputSchema

The input schema for the tool in JSON format.

Type: [ToolInputSchema](#) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: Yes

name

The name for the tool.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: `^[a-zA-Z][a-zA-Z0-9_]*$`

Required: Yes

description

The description for the tool.

Type: String

Length Constraints: Minimum length of 1.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ToolUseBlock

Service: Amazon Bedrock Runtime

A tool use content block. Contains information about a tool that the model is requesting be run., The model uses the result from the tool to generate a response.

Contents

input

The input to pass to the tool.

Type: JSON value

Required: Yes

name

The name of the tool that the model wants to use.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: `^[a-zA-Z][a-zA-Z0-9_]*$`

Required: Yes

toolUseId

The ID for the tool request.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: `^[a-zA-Z0-9_-]+$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ToolUseBlockDelta

Service: Amazon Bedrock Runtime

The delta for a tool use block.

Contents

input

The input for a requested tool.

Type: String

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

ToolUseBlockStart

Service: Amazon Bedrock Runtime

The start of a tool use block.

Contents

name

The name of the tool that the model is requesting to use.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: `^[a-zA-Z][a-zA-Z0-9_]*$`

Required: Yes

toolUseId

The ID for the tool request.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: `^[a-zA-Z0-9_-]+$`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)

Common Parameters

The following list contains the parameters that all actions use for signing Signature Version 4 requests with a query string. Any action-specific parameters are listed in the topic for that action. For more information about Signature Version 4, see [Signing AWS API requests](#) in the *IAM User Guide*.

Action

The action to be performed.

Type: string

Required: Yes

Version

The API version that the request is written for, expressed in the format YYYY-MM-DD.

Type: string

Required: Yes

X-Amz-Algorithm

The hash algorithm that you used to create the request signature.

Condition: Specify this parameter when you include authentication information in a query string instead of in the HTTP authorization header.

Type: string

Valid Values: AWS4-HMAC-SHA256

Required: Conditional

X-Amz-Credential

The credential scope value, which is a string that includes your access key, the date, the region you are targeting, the service you are requesting, and a termination string ("aws4_request"). The value is expressed in the following format: *access_key/YYYYMMDD/region/service/aws4_request*.

For more information, see [Create a signed AWS API request](#) in the *IAM User Guide*.

Condition: Specify this parameter when you include authentication information in a query string instead of in the HTTP authorization header.

Type: string

Required: Conditional

X-Amz-Date

The date that is used to create the signature. The format must be ISO 8601 basic format (YYYYMMDD'T'HHMMSS'Z'). For example, the following date time is a valid X-Amz-Date value: 20120325T120000Z.

Condition: X-Amz-Date is optional for all requests; it can be used to override the date used for signing requests. If the Date header is specified in the ISO 8601 basic format, X-Amz-Date is not required. When X-Amz-Date is used, it always overrides the value of the Date header. For more information, see [Elements of an AWS API request signature](#) in the *IAM User Guide*.

Type: string

Required: Conditional

X-Amz-Security-Token

The temporary security token that was obtained through a call to AWS Security Token Service (AWS STS). For a list of services that support temporary security credentials from AWS STS, see [AWS services that work with IAM](#) in the *IAM User Guide*.

Condition: If you're using temporary security credentials from AWS STS, you must include the security token.

Type: string

Required: Conditional

X-Amz-Signature

Specifies the hex-encoded signature that was calculated from the string to sign and the derived signing key.

Condition: Specify this parameter when you include authentication information in a query string instead of in the HTTP authorization header.

Type: string

Required: Conditional

X-Amz-SignedHeaders

Specifies all the HTTP headers that were included as part of the canonical request. For more information about specifying signed headers, see [Create a signed AWS API request](#) in the *IAM User Guide*.

Condition: Specify this parameter when you include authentication information in a query string instead of in the HTTP authorization header.

Type: string

Required: Conditional

Common Errors

This section lists the errors common to the API actions of all AWS services. For errors specific to an API action for this service, see the topic for that API action.

AccessDeniedException

You do not have sufficient access to perform this action.

HTTP Status Code: 400

IncompleteSignature

The request signature does not conform to AWS standards.

HTTP Status Code: 400

InternalFailure

The request processing has failed because of an unknown error, exception or failure.

HTTP Status Code: 500

InvalidAction

The action or operation requested is invalid. Verify that the action is typed correctly.

HTTP Status Code: 400

InvalidClientId

The X.509 certificate or AWS access key ID provided does not exist in our records.

HTTP Status Code: 403

NotAuthorized

You do not have permission to perform this action.

HTTP Status Code: 400

OptInRequired

The AWS access key ID needs a subscription for the service.

HTTP Status Code: 403

RequestExpired

The request reached the service more than 15 minutes after the date stamp on the request or more than 15 minutes after the request expiration date (such as for pre-signed URLs), or the date stamp on the request is more than 15 minutes in the future.

HTTP Status Code: 400

ServiceUnavailable

The request has failed due to a temporary failure of the server.

HTTP Status Code: 503

ThrottlingException

The request was denied due to request throttling.

HTTP Status Code: 400

ValidationError

The input fails to satisfy the constraints specified by an AWS service.

HTTP Status Code: 400