Technology Spotlight

The Cloud Steps Up To Tightly Coupled HPC Codes

Sponsored by Amazon Web Services

Steve Conway, Earl Joseph, Bob Sorensen and Alex Norton
January 2020

HYPERION RESEARCH OPINION

For much of the past decade, HPC jobs run in third-party clouds were effectively limited to "embarrassingly parallel" (EP) workloads that need little internode communication-I/O when the problems are running. The reason why, as the influential U.S. Department of Energy *Magellan Report on Cloud Computing for Science* explained in 2011, is that "I/O-intensive applications [a.k.a. tightly coupled codes] take a substantial hit when run inside virtual environments." Even with this limitation, however, more HPC sites turned to cloud services providers (CSPs). Between 2011 and 2018, the portion of all HPC sites worldwide that were running at least some jobs in third-party clouds jumped from 13% to 74%. The global pharmaceutical industry, for one, often found it faster and less expensive to run EP workloads in external clouds than on premises. Novartis tested 21 million small molecules as potential cancer-fighting agents on AWS in less than four hours and for only a few thousand dollars.

Major CSPs have substantially addressed the challenge of supporting tightly coupled codes by adding capable interconnects, along with lightweight containers that minimize the "virtualization tax" on these codes and bare metal instances that eliminate this penalty altogether. CSPs have taken these important steps to capture a broader range of HPC workloads for two main reasons. First, the HPC market is large and growing robustly, expanding to nearly $28 billion in 2018 and forecast by us to exceed $46 billion in 2023. Second, HPC is nearly indispensable at the forefront of R&D for economically important AI use cases, including automated driving, precision medicine, smart cities, the Internet of Things, and others.

CSPs' efforts are already paying off. Hyperion Research studies indicate that the portion of HPC workloads being sent to third-party clouds has doubled in the past 18 months, from 10% to 20%. Hyperion Research forecasts that HPC spending in third-party clouds will triple (24.6% CAGR) from $2.5 billion in 2018 to $7.4 billion in 2023.

The future looks promising for tightly coupled codes run in third-party clouds. A recent independent study of design engineers by Peerless Research Group found that 37% of the respondents already use third-party clouds for tightly coupled engineering simulations.[1] More surprising, a key takeaway from the "Banking Goes Cloud" conference hosted by the staid Swiss Bankers Association was that in a few years most banks will use third-party clouds, even for customer data.[2]

This paper discusses cloud support for tightly coupled codes and looks at Amazon Web Services, the market-leading CSP, to show how major players are addressing this large, important class of workloads.

[1] *Study on Cloud Computing for Engineering Simulation*. Peerless Research Group. February 2019

[2] *Cloud Leitfaden* [Cloud Guide]. SwissBanking. March 2019

Note: this page is intentionally blank.

NEW CAPABILITIES PUSH HPC CLOUD GROWTH PAST THE TIPPING POINT

From 2011 to 2017, the number of HPC sites running at least some jobs in third-party clouds skyrocketed from 11% to 74%, but the average portion of all their HPC workloads sent to these cloud services providers (CSPs) remained stuck at under 10 percent. HPC cloud use was becoming very wide, but not very deep.

What liberated HPC cloud growth, especially in the past 18 months, was the recognition by major CSPs that HPC is a market worth special attention. The market for HPC servers, storage, software and technical support expanded from about $2 billion in 1990 to $27.7 billion in 2018, en route to a Hyperion Research forecast $39.2 billion in 2023.

But that's not all, CSPs are also aware that HPC is an important factor for success in the emerging markets for artificial intelligence (AI) and other high performance data analysis (HPDA) applications. HPC is nearly indispensable today at the forefront of R&D for automated driving systems, precision medicine, affinity marketing, business intelligence, cyber security, smart cities and the Internet of Things. Today's HPC activity indicates where the mainstream HPDA and AI markets are headed in the future.

With the expanding HPC market opportunity in mind, including HPC-enabled AI, leading CSPs have been adding features, functions and partners designed to make their platforms run a broader spectrum of HPC related workloads time- and cost-effectively.

This special attention has already paid off. In Hyperion Research's latest global study, the surveyed users on average said they run 33% of all their HPC work in third-party clouds. Extrapolating from this group of admitted cloud users to the whole HPC community drops that average to 20%--still a big uptick from the 10% figure in Hyperion Research surveys not long ago. HPC cloud computing has reached a tipping point.

We forecast that HPC spending in third-party clouds will triple (24.6% CAGR) from $2.5 billion in 2018 to $7.4 billion in 2023 (Figure 1), bumping up our forecast for the whole HPC ecosystem market past $46 billion. The third-party cloud portion will represent an impressive 19% of that expanded total (Figure 2).

FIGURE 1

HPC Cloud Growth, 2018-2023 ($ Millions)

	2018	2019	2020	2021	2022	2023	CAGR '19-'23
NEW 2019 HPC Cloud Forecast	2,466	3,910	4,262	5,135	6,182	7,418	24.6%
Previous 2018 Cloud Forecast	2,466	3,031	3,771	4,585	5,520	-	-
HPC Server Forecast	13,706	14,495	15,780	17,356	19,011	19,979	7.8%

Source: Hyperion Research, 2019

FIGURE 2

HPC Worldwide Market Forecast, 2018-2023 ($000)

Revenues by the Broader HPC Market Areas			
	2018	2023	CAGR 18-23
Server	13,706,088	19,979,016	7.8%
Storage	5,547,188	7,771,184	7.0%
Middleware	1,582,892	2,217,801	7.0%
Applications	4,627,492	6,413,592	6.7%
Service	2,229,921	2,858,820	5.1%
Total Revenue	27,693,580	39,240,413	7.2%

HPC cloud (CSP) usage raises forecast to $46.6B)

Source: Hyperion Research, 2019

Supporting Tightly Coupled Codes

One of the most important capabilities added by major CSPs in recent years has been support for running tightly coupled software codes efficiently.

Before this, third-party clouds were widely criticized as fit only for "embarrassingly parallel" workloads and unsuitable for the many tightly coupled codes that depend on fast, efficient communication between compute nodes while jobs are running.

This important weakness was highlighted in the conclusion of the U.S. Department of Energy's influential 2011 *Magellan Report on Cloud Computing for Science*:

> Applications with minimal communication and I/O [requirements] are able to achieve similar performance in cloud environments as HPC environments. The [negative] performance impact for HPC applications comes from the absence of high bandwidth, low latency interconnects in virtualized environments. This a majority of current DOE HPC applications even in midrange concurrencies is unlikely to run efficiently in today's cloud environments. Similarly, I/O-intensive applications take a substantial hit when run inside virtual environments.[3]

Many HPC codes are designed to simulate (model) the structures and behaviors of dynamic physical systems, ranging from vehicle collisions and weather patterns to the cosmos itself, or quasi-physical systems such as financial risk calculations and social interactions. These workloads typically require

[3] *The Magellan Report on Cloud Computing for Science*. U.S. Department of Energy Office of Advanced Scientific Computing Research. December 2011

locally computed data to be distributed globally, over the HPC system interconnect, during the problem runtime; i.e., partial results of the solution must be shared in order for the solution to be usefully completed.

Designing a new car, for example, involves many interdependent factors; changing the body shape can harm the vehicle's fuel efficiency, crashworthiness and passenger cabin noise. Any change like this needs to be communicated to the nodes/processes optimizing these factors in order for the design of a competitive car to succeed.

Applications needing this fast communication and data distribution among running processes are called ***tightly coupled codes***. The most popular method for distributing data and related information during HPC workload runtimes is the message passing interface (MPI). In a recent Hyperion Research worldwide study, 91% of surveyed HPC sites said they use MPI and nearly half (47%) of the surveyed sites reported that a majority of their HPC workloads employ MPI. An important MPI alternative is the OpenMP parallel programming interface.

Tightly coupled applications are used in every HPC market segment, including computational fluid dynamics (CFD), weather prediction, oil and gas reservoir simulations, and artificial intelligence areas such as machine learning, deep learning and graph analytics.

Tightly coupled codes depend heavily on fast, reliable networks to distribute data among compute nodes involved in solution the problems. Even minor network inefficiencies or disruptions can cause major delays in solution times, with attendant losses of productivity.

Given the importance of tightly coupled codes, major CSPs set to work and added low-latency interconnects capable of providing 100 Gbps of dedicated aggregate network bandwidth. Not surprisingly, these interconnects have contributed significantly to the recent rapid growth in HPC use of CSP platforms.

Bypassing the Operating System

But while more capable interconnects speed up internode communications, they fail to remove another performance barrier. Tightly coupled codes are slowed down because each internode message has to access the operating system.

CSPs have been working to overcome this obstacle. Amazon Web Services (AWS), the global market leader in HPC cloud computing, is a case in point.

AMAZON DOUBLES DOWN ON CAPABILITIES FOR TIGHTLY COUPLED CODES

AWS has focused on network performance in the cloud for nearly a decade. AWS provided 10 Gbps instance types in 2011 and a step function in network performance with Enhanced Networking in 2013.

AWS was an early adopter of 25 Gbps networking in 2011 and in 2018 AWS released EC2 C5n instances with 100 Gbps networking both between instances and to other AWS infrastructure such as Amazon Simple Storage Service (S3).

Also in 2016, AWS introduced the AWS Nitro System, an IO/network accelerator plus security chip plus hypervisor all-in-one “card.” The company also debuted the Elastic Network Adapter (ENA) as part of its Nitro System. This enabled customers to select AWS instance types with bandwidths and price points designed to support their applications with minimal hypervisor overhead.

The Elastic Fabric Adapter (EFA)

In an effort to bring ENA-like flexibility and upgradability to HPC and ML workloads, in the first half of 2019 AWS introduced the Elastic Fabric Adapter (EFA). EFA was designed to provide three main benefits: bypassing the operating system (OS) to provide more direct access to Amazon EC2 networking hardware, in order to speed solution times and reduce overhead; continuing ENA improvements for enhanced bandwidth; and innovating in network transport (e.g., AWS's cloud-centric Scalable Reliable Datagram) for reliability.

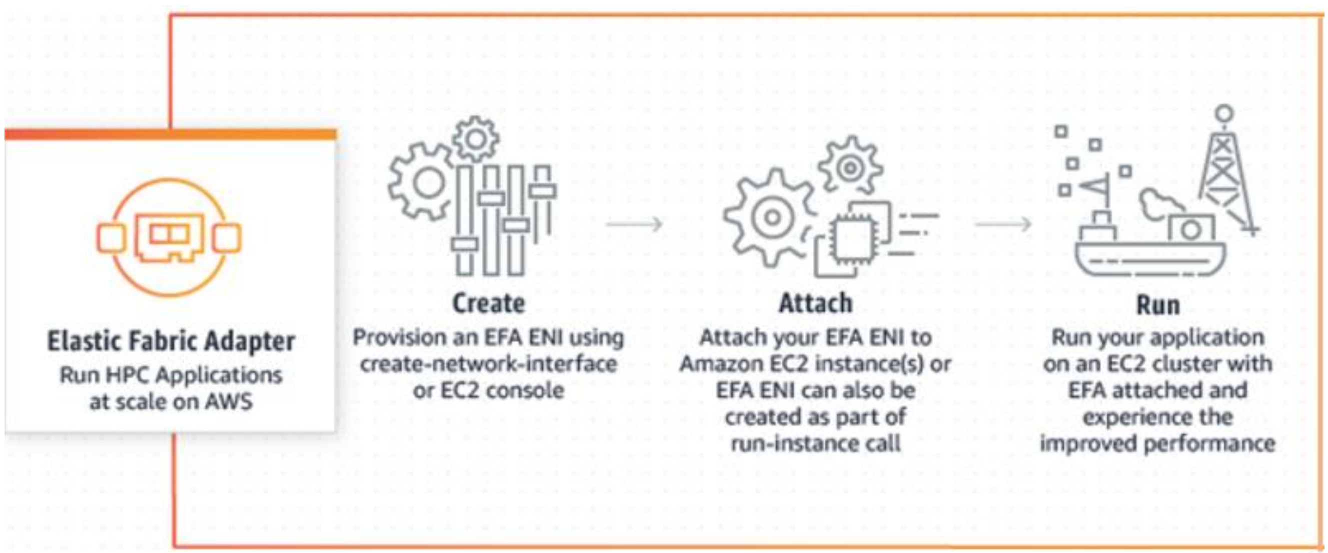
EFA supports applications that use MPI by employing the industry-standard libfabric APIs. As a result, Amazon reports, codes using a supported MPI library can get by with little or no source modifications.

According to AWS, the EFA technology will enable customers to scale their applications to tens of thousands of CPU cores. It's aimed at typical scalable HPC codes such as computational fluid dynamics, weather modeling, and reservoir simulation, as well as newer AI methods and codes.

AWS makes EFA available as an optional networking feature, at no charge, that can be enabled on a growing number of instance types. For now, EFA is available on the following instances types - c5n, c5n.metal, i3en, inf, m5dn, m5n, r5dn, r5n, and p3dn. EFA will soon be available on g4dn.metal, c5an.metal, and c5adn.metal instance types.

FIGURE 3

The Amazon Elastic Fabric Adapter (EFA)



Source: Amazon, 2019

Key EFA Features

- *Out-of-order delivery.* Relaxing the need for in-order message delivery eliminates head-of-line blocking. The messages become independent and in most cases the application/middleware can restore ordering if/when it's needed. To support applications that require such strong ordering guarantees, AWS implemented a packet reordering engine in the EFA user space software stack.
- *Equal-cost multi-path routing (ECMP).* There are hundreds of possible paths between two EFA-enabled instances. Amazon uses the consistent flow hashing properties of its large multi-path network and SRD's ability to rapidly react to network conditions to find the most effective paths for a message. Packet spraying aims to prevent hot spots and allow for fast and transparent recovery from network failures.
- *Fast packet drop response*. SRD is designed to respond much more quickly to packet drops than high-level protocols do. Occasional packet loss, especially for long-running HPC applications, is part and parcel of normal network operations, not an aberrant condition.
- *Scalable transport offload*. With SRD, unlike other reliable protocols such as InfiniBand Reliably-Connected (IBRC), applications do not maintain a connection with dedicated resources for each pair of communicating processes in the application. A single process can create and use a single Queue Pair (QP) to communicate with any number of peers.

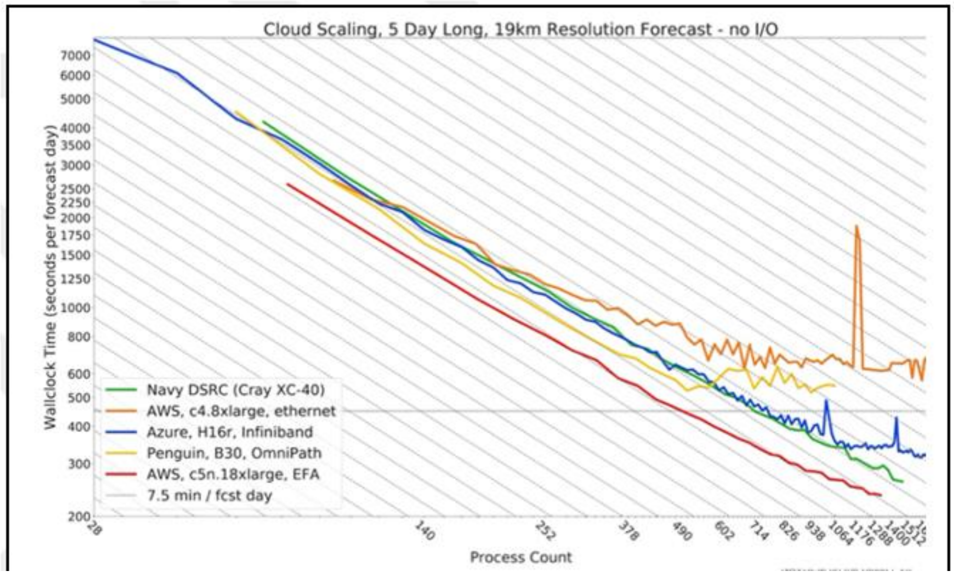
EFA Making a Difference

An evaluation study of HPC infrastructure for running the Navy Global Atmospheric Model (NAVGEM) weather prediction code exemplifies the bottom-line impact of the EFA advancement on tightly coupled applications. The US Naval Research Laboratory (NRL) concluded that for high resolution forecasts, using Amazon EC2 C5n Instances with the EFA network interface achieved results 74% faster with 27% lower costs than previous generation instances (C4) and network (TCP over ENA) on AWS.[4] On the other hand, note too, that at higher process count, NAVGEM on ethernet (without EFA), InfiniBand, and OmniPath stops scaling and wall clock time (i.e., time-to-results) exhibits undesired and counterproductive scaling behavior. Figure 4.

[4] *NAVGEM on the Cloud: Computational Evaluation of Commercial Cloud HPC with a Global Atmospheric Model*, U.S. Naval Research Laboratory, 2019.

FIGURE 4

Evaluation of HPC Infrastructure for Running NAVGEM



Source: Naval Research Laboratory (U.S.), 2019

Generational MPI bandwidth improvements. Figure 5 compares multi-stream bandwidth of MPI across multiple generations of EC2 instances using the Ohio State University (OSU) Benchmark Suite. . C5n results in orange in the right pane are without EFA.

FIGURE 5

Amazon EC2 MPI Bandwidth and Latency Gains Over Time

Source: Amazon, 2019

FUTURE OUTLOOK

Hyperion Research forecasts that the global market for running HPC workloads in third-party clouds will reach about $7.4 billion in 2023, representing about 19% of our projected $46.6 billion in overall spending on HPC in that year. CSPs have been turning more attention to the global HPC market because this market has become a sizable opportunity and because HPC is at the forefront of R&D for economically important, emerging HPDA-AI use cases.

The ability of CSPs to support tightly coupled codes will boost their share of all HPC spending. It will also increase their prospects in the AI market, because engineering simulation and other tightly coupled codes are important for many of the most economically important AI use cases. For instance, the RAND Corporation estimates that 8.8 billion miles of test driving will be needed for consumers to acquire 95% confidence in the safety of autonomous vehicles, and that physical testing for this many miles would take 400 years.[5] Experts indicate that the only way to instill confidence in 5-10 years is with simulation, using AI algorithms on high performance computers. Precision medicine, smart cities and the Internet of Things will also rely on the interplay of analytics and simulation.

Hyperion Research believes that Amazon Web Services' EFA provides important new support for running tightly coupled codes in a cloud environment, and that EFA increases AWS' potential to benefit from the robust growth we forecast for HPC cloud usage.

[5] https://www.rand.org/content/dam/rand/pubs/research_reports/RR1400/RR1478/RAND_RR1478.pdf

About Hyperion Research, LLC

Hyperion Research provides data driven research, analysis and recommendations for technologies, applications, and markets in high performance computing and emerging technology areas to help organizations worldwide make effective decisions and seize growth opportunities. Research includes market sizing and forecasting, share tracking, segmentation, technology and related trend analysis, and both user & vendor analysis for multiuser technical server technology used for HPC and HPDA (high performance data analysis). We provide thought leadership and practical guidance for users, vendors and other members of the HPC community by focusing on key market and technology trends across government, industry, commerce, and academia.

Headquarters

365 Summit Avenue
St. Paul, MN 55102
USA
612.812.5798
www.HyperionResearch.com and www.hpcuserforum.com