

White Paper

Bringing HPC Expertise to Cloud Computing

Sponsored by: Dell Technologies and AMD

Alex Norton, Steve Conway and Earl Joseph
April 2020

HYPERION RESEARCH OPINION

HPC in the cloud has undergone significant growth in the past few years, attracting new users and more fully supporting existing users. Cloud service providers (CSPs) are key to this growth as they are taking a stronger interest in the HPC market, focusing on delivering capabilities that more closely match on-premise systems in performance, security, and in some cases, price. Prior to 2018, HPC users were running, on average, under 10 percent of their workloads in the cloud. More recently however, there has been a major shift in the sector, prompting Hyperion Research to revise its HPC cloud forecast to include an additional \$1 billion in 2019 spending and increased growth in cloud spending over the next five years to reach over \$9 billion in 2024. For both cloud and on-premise users, storage will continue to be a high growth area.

A major driver of increased cloud-based HPC usage is the recent availability of hybrid and multicloud options offered by many CSPs that are designed to provide HPC users a near seamless environment between on-premise hardware and counterpart cloud resources. As HPC applications grow in complexity, as in the case with the addition of new AI techniques to a wide range of application portfolios, the ability to easily burst into the cloud provides a more flexible, responsive computing environment. Likewise, multicloud capabilities are increasingly appealing to many HPC sites as they allow users to manage multiple third-party cloud resources from a single "pane of glass," a feature especially attractive to global corporations that use different CSPs scattered across different geographic locations.

Meanwhile, many CSPs are moving to provide support specifically targeted for HPC workloads, adding in-house HPC experts, and offering services to help HPC users better understand the process of running their applications in the cloud. With the added HPC expertise from CSPs, HPC users have more resources at their disposal to effectively and properly use the right instances on the right platform for each of their jobs. In addition, the added capabilities of multicloud options provide each application in a given portfolio the ability to run in a separate environment tailored to the technical needs of each application.

Realizing the fullest potential of any user's HPC in the cloud workload at the best possible price and performance requires a strong partnership between HPC users and the CSP that supports them. HPC applications carry with them a specific set of technical requirements, and in many cases, HPC users can benefit from the expert guidance and insight available from their CSP partners. Dell Technologies, with years of experience in supplying HPC systems around the world, currently offers a wide range of support, in hardware, software, and HPC cloud expertise, to HPC users looking to maximize the performance of their cloud-based HPC workloads.

SITUATION OVERVIEW

Key HPC Definitions

Hyperion Research divides the HPC sector into four major price bands.

- Supercomputer: systems priced at \$500,000 and above
- Divisional: systems priced at \$250,000 to \$449,000
- Departmental: systems priced at \$100,000 to \$249,000
- Workgroup: systems priced below \$100,000

For the purposes of this paper it is important to note that when the term "cloud" is used, it is referring to the public cloud, or third-party hosted resources. A private cloud that is owned and run internally is not included in this definition although it is an important portion of the overall HPC cloud market.

Growth of the HPC Market

The HPC market continues to grow: 2018 showed higher growth than expected and 2019 was a continuation of the strong growth the HPC market has seen in recent years, although showing slight growth from 2018. Looking forward to 2020 and beyond, Hyperion Research projects a CAGR of 10 percent year over year growth for the next five years for the broader HPC market (including servers, software, storage and repair services). All four price bands of the HPC market show growth, with the supercomputer group (systems priced at \$500,000 and above) exhibiting the largest.

Hyperion Research believes that the workgroup segment, tagged as the slowest growth sector in the overall on-prem HPC market, is the most aggressive in transitioning its overall HPC workload from on-prem to cloud based. The workgroup segment represents the largest portion of HPC systems installed worldwide.

New Capabilities Push HPC Cloud Growth Higher

Prior to 2018, the percentage of HPC workloads that ran in the public cloud had plateaued at around 10 percent of the overall HPC workload portfolio. In general, many users were running experiments in the cloud or using the cloud for so-called embarrassingly parallel workloads. In 2019, based on Hyperion Research surveys and studies, the percentage of HPC workloads run in the public cloud jumped to about 20 percent. This new data was the driving force behind the updated Hyperion Research HPC in the cloud forecast released in late 2019, shown below in Table 1, that highlights an almost \$1 billion increase in 2019 HPC cloud revenue. *(As a point of note, the Hyperion Research HPC cloud forecast is from the end-user spend perspective. It does not measure the purchase of HPC resources by CSPs.)*

- There were many factors behind the increase in public cloud spend from HPC users: the cloud offerings became much more capable for HPC applications; CSPs and other companies provided easier access models to clouds; and a larger set of application software became available on the cloud.

Another driving factor is that HPC in the cloud platforms are becoming an increasingly important enabler in the emerging and high growth markets of artificial intelligence and other high performance data analysis applications. Critical applications here include automated driving systems, precision medicine, affinity marketing, business intelligence, cyber security, smart cities and the internet of things.

- Cloud-based platforms are well suited to handle many of these AI applications, and for some applications the necessary data required for training are either collected in the cloud or are already stored there, offering significant cost reductions moving data into the cloud for processing.
- In addition, CSPs have started to offer bare metal instances to remove some of the virtualization penalties commonly associated with cloud computing.

Finally, hybrid solutions, as well as multicloud offerings, have grown in interest and deployment over the past few years. These two offerings allow for flexibility in compute resources, especially for HPC users with diverse workloads.

- Hybrid cloud solutions allow for more flexible and responsive burst or surge computing capabilities, providing an HPC user the ability to fine tune their compute/storage resource pool as needed. For specific workloads and types of users, hybrid environments allow for more efficient computing with the most appropriate hardware and software configuration, which may not be available on-premise.
- Multicloud allows for similar flexibility in choosing the optimal platform, including hardware, software, middleware, and other aspects of the compute instances.

Looking to the future, spending for HPC applications in the cloud shows strong growth out to 2024. Hyperion Research is anticipating major growth in the market segment in the coming years as more of the barriers of entry to HPC cloud usage are addressed by CSPs and associated hardware and software vendors.

One key issue for many HPC in the cloud users and potential users is data security. Many HPC users elect not to run their mission critical workloads in the cloud for fear of data breaches or data loss. This issue has been a constant concern within the HPC user base for many years, despite the major strides CSPs and vendors have made in addressing the problems. Most CSPs now offer specific security certifications on their platforms, like ITAR or FedRAMP, as well as make their security and data protection capabilities more transparent to the end user.

- Security concerns, although better addressed today than in the past, remain a constant friction point for HPC end users and are crucial to the next major growth spurt in this market.

TABLE 1

HPC Cloud Forecast 2018-2023 (\$M)

	2018	2019	2020	2021	2022	2023	CAGR '18-'23
NEW 2019 HPC Cloud Forecast	2,466	3,910	4,600	5,135	6,182	7,418	24.6%
Previous 2018 Cloud Forecast	2,466	3,031	3,771	4,585	5,520	-	23.8%
HPC Server Forecast	13,683	13,713	14,484	15,658	18,457	19,940	7.8%

Source: Hyperion Research, 2020

Table 2 shows the broader HPC market for each segment of the overall HPC sector. Note that service refers to the repair and maintenance services on the systems. As HPC cloud revenues have continued to climb, they have now been included in the overall HPC market numbers. HPC storage enjoys the largest projected growth of the HPC sectors tracked. With the addition of the cloud sector, the total HPC market is projected to top \$52 billion in 2024.

TABLE 2

Revenues by the Broader HPC Market Areas (\$M)

	2018	2019	2020	2021	2022	2023	2024	CAGR '19-'24
Server	\$13,683	\$13,713	\$14,484	\$15,658	\$18,457	\$19,940	\$20,844	8.7%
Storage	\$5,568	\$5,629	\$6,065	\$6,761	\$8,272	\$9,238	\$9,977	12.1%
Middleware	\$1,591	\$1,618	\$1,716	\$1,863	\$2,218	\$2,411	\$2,531	9.4%
Applications	\$4,655	\$4,695	\$4,973	\$5,274	\$6,143	\$6,525	\$6,692	7.3%
Service	\$2,250	\$2,238	\$2,324	\$2,419	\$2,767	\$2,883	\$2,899	5.3%
Public Cloud Spending	\$2,466	\$3,910	\$4,600	\$5,135	\$6,182	\$7,418	\$9,243	18.8%
Total Revenue	\$30,212	\$31,803	\$34,163	\$37,110	\$44,040	\$48,416	\$52,187	10.4%

Source: Hyperion Research, 2020

HPC Cloud and AI/ML/DL Workloads

Machine learning, deep learning, inferencing, training, graph analytics, and other AI methods, as well as high performance data analytics (HPDA) have all seen robust growth in the past few years, both in budget allocations as well as organizational focus. Growing at more than 30 percent, HPC-enabled AI is projected to be a \$3.5 billion market in 2024. The top table below shows the sub-segmentation of HPDA server revenue, as well as the HPC-enabled AI portion. It is important to note that HPC-enabled AI is a subset of the HPDA market revenue below. In the second table below, the AI portion of the server market is further divided into machine learning, deep learning and other AI, which is made up primarily of graph analytics right now. Machine learning and deep learning are growing at similar rates; however, machine learning is currently a more significant portion of the HPC-enabled AI market. Although we are still at the beginning of AI, the cloud will play a key role in the future of AI and HPC. HPC is crucial to AI developments, and the growth of HPC in the cloud, as well as the attention paid to HPC by CSPs, has resulted in many AI workflows being created and run in cloud environments. For some AI workflows, access to enough data for training can be challenging for on-premise execution or expensive to transfer from a cloud environment. In these cases, data residing or collected in a cloud environment, from sensors or the combination of multiple different data sets, can be better run where the data exists.

TABLE 3**Forecast: Worldwide HPC-Based AI Revenues vs Total HPDA Revenues (\$M)**

	2018	2019	2020	2021	2022	2023	2024	CAGR '19-'24
WW HPC Server Revenues	13,683	13,713	14,484	15,658	18,457	19,940	20,844	8.7%
WW HPDA (including AI) Server Revenues	3,153	3,598	3,932	4,737	5,467	6,480	7,478	15.8%
Total HPC-Based AI (ML, DL & Other)	747	918	1,094	1,399	1,810	2,745	3,555	31.1%

Source: Hyperion Research, 2020

TABLE 4**Forecast: Worldwide ML, DL & Other AI HPC-Based Revenues (\$M)**

	2018	2019	2020	2021	2022	2023	2024	CAGR '19-'24
ML in HPC	532	667	771	986	1,285	1,960	2,538	30.6%
DL in HPC	177	209	265	342	443	665	866	32.9%
Other AI in HPC	38	42	58	70	83	120	150	29.0%
Total	747	918	1,094	1,399	1,810	2,745	3,555	31.1%

Source: Hyperion Research, 2020

From a recent study on the global HPC end user ecosystem, HPC users reported that within the next year and a half, on average 20 percent of their AI workloads will be running in the cloud. Access to hardware and software infrastructure in the cloud not available from their on-premise systems is a major driver for moving certain workloads to the cloud. A third of the respondents indicated that access to specialty hardware or software was a major driving force in their decision to use the cloud. The complexities of the evolving AI workloads today and in the future necessitates utilization of specialized infrastructure.

- In recent years, x86 technology has been delivered by one vendor in the HPC market with little competition. Recently, AMD has shown increased and healthy growth around their processor technology, especially in their performance on AI workloads.

- A major indicator of this shift is the US DoE procurements for exascale machines, which will use the AMD processor to address the growing diversity of applications planned to be run, including AI applications.

DELL TECHNOLOGIES CLOUD SOLUTIONS

HPC in the cloud has been a growing market segment in the past and has recently undergone a major growth year. A major shift occurred in the market: the average percent of user workloads running in the cloud nearly doubled from 10-12 percent in 2018 to 20 percent in 2019. HPC users, especially those in the workgroup segment, look to the cloud as a viable option for compute resources for elastic, burst computing jobs, or as a way to run new workloads on hardware or software not available on-premise.

Dell Technologies has a long history in the HPC market and is the fastest growing major HPC OEM, providing on-premise solutions to some of the largest end user sites in the world. This experience, which led to an understanding of HPC user needs, places Dell Technologies in a strong position to help end users accomplish their workloads in an efficient environment. Dell Technologies offers scalable and elastic cloud resources to accomplish both HPC and AI workloads, with the opportunity to expand and integrate with current on-premise systems. Through partnerships with AWS, Microsoft Azure, and Google Cloud, Dell Technologies brings a long understanding of HPC, as well as deep domain expertise within the various segments of HPC, to provide targeted solutions to HPC customers on the top cloud platforms.

Dell Cloud's partnerships with other CSPs for multicloud enable end users to run their applications across a variety of platforms for optimized performance. Verne Global, for example, has the capability, through Dell Technologies, to implement an HPC colocation solution with the powerADVANCE solution. This solution offers cost effective high performance computing hosted in the same location or campus. With VMware Cloud on AWS, end users have access to more than 165 AWS services including compute, AI, and analytics. In addition, VMware Cloud on AWS provides the same VMware Cloud technologies across on-premise and cloud infrastructures, enabling a near seamless environment with the capability of moving applications from on-premise to AWS without the need to buy any new hardware, change operations, or rewrite applications.

Many end users are beginning to understand how AI, and its many subsegments, can aid their current workloads, whether through running AI in conjunction with traditional modeling and simulation workloads, or in entirely new application spaces with AI. With their HPC & AI Innovation Lab, Dell brings together a robust hardware and software infrastructure, combined with a staff of experts, including engineers, computer scientists, and subject matter experts. These experts are all deeply engaged in the HPC community and understand the best practices for solving HPC and AI problems.

With Dell's services for HPC, AI and HPDA hybrid clouds, Dell helps end users in all stages of deploying a hybrid cloud solution, starting at designing the solution and through the process to systems management. By bringing their expertise to the equation, Dell provides support to properly implement a hybrid cloud solution, including access to HPC experts and HPC cluster assistance and HPC solution-level support. Dell also provides extensive HPC consulting capabilities, including:

- Code migration and optimization
- Cluster optimization, tuning and management
- x86 storage servers: selection of Linux, OpenFabrics
- HPC cluster management processes and other consulting requests

With Dell Remote Cluster Management (RCM), end users can use the experts at Dell to manage and maintain their HPC clusters and applications, reducing the need to hire in-house expertise. As many HPC users are still learning how to effectively take advantage of the cloud and implement the best solution for their business, Dell offers their proficiency in HPC cluster management and HPC deployments in the cloud to ease the transition for the end user and provide the best solution for the job.

Moreover, Dell provides hybrid cloud security through Secureworks, a company that works to integrate intelligence with security. Secureworks allows end users to monitor their deployments and detect anomalies in the data as well as manage the vulnerabilities of the cloud deployment and reinforce the security in place with targeted adversarial testing. Testing the security of the deployment is achieved without waiting for a third party to infiltrate the system.

In addition to Secureworks, Dell Technologies also incorporates the Boomi AtomSphere platform to address security at critical points: the data level to the application and platform layer as well as the network and facilities infrastructure. This integration platform as a service (iPaaS) provides strong security for the data from transfer threats to security within the data center itself.

As a staple at the top of the HPC server business for years, Dell has developed a deep understanding around what HPC jobs need and how to solve some of the most complex HPC problems. With the rise of cloud computing for HPC, Dell has stepped up to help users understand how to deploy workloads in the cloud. Dell's hybrid cloud offerings, combined with their extensive partnerships with many CSPs, enable end users to take full advantage of the cloud to enhance their compute capabilities. Furthermore, with the added services and solutions Dell has tailored specifically to HPC and AI, end users can leverage Dell's knowledge and skills to deploy secure and powerful hybrid cloud solutions for their entire workload portfolio, including traditional modeling and simulation workloads as well as the new AI workloads in HPC.

FUTURE OUTLOOK

HPC in the cloud continues to expand as CSPs develop a growing portfolio of solutions that better address HPC workloads and users, offering options for better cost effectiveness for cloud computing, better queue times than on-premise data centers, and access to a wide range of hardware.

HPC in the cloud will continue to grow as barriers to entry, namely data security, data movement and cost, are addressed by CSPs, ultimately enabling end users to take full advantage of the cloud. Today, the cloud is primarily a complementary solution to on-premise resources, especially with hybrid cloud offerings, but that could change in coming years as cloud capabilities increase.

The Dell Technologies Cloud is well positioned to help support HPC end users to maximize their compute capabilities through hybrid cloud and multicloud solutions, bringing their extensive and well-known expertise in the HPC market to address the needs of HPC end users who want to capitalize on the capabilities offered by cloud solutions. With the growing excitement and investment in AI, as well as the continued growth of HPC, the cloud will play a major role in current and future compute considerations.

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 multi-user 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

Copyright Notice

Copyright 2020 Hyperion Research LLC. Reproduction is forbidden unless authorized. All rights reserved. Visit www.HyperionResearch.com to learn more. Please contact 612.812.5798 and/or email info@hyperionres.com for information on reprints, additional copies, web rights, or quoting permission.