

Special Study Results

Cloud Resources are Employed at Nearly Half of HPC Sites and are Expected to Gain Additional Popularity

Jaclyn Ludema and Tom Sorensen
May 2023

HYPERION RESEARCH OPINION

The use of cloud resources continues to gain popularity amongst HPC sites, with about half of sites currently using public cloud services for HPC workloads and 60% of sites planning to use public cloud services in the next 6-18 months.

Pursuing the use of cloud services is of interest to HPC sites particularly for experimental research, and projects with compute and storage requirements that continue to evolve. Experimental research can take advantage of the ease at which cloud services are scaled, with less overhead than on-premises computing. Additionally, the cloud offers access to cutting-edge accelerators and components without the financial commitment of ownership. For many sites, the cloud is a great playground for experimental workloads before adopting production-ready workloads to sites' on-premises systems.

While the trend of using the cloud in some capacity at HPC sites continues to gain popularity, it is important to note that based on runtime, most sites only use the cloud for 20% or less of their workloads. This suggests that most sites are employing cloud resources to serve tactical purposes in a hybrid infrastructure, not to migrate away from on-premises infrastructures.

This data is from the eighth annual study of Hyperion Research's high-performance computing (HPC) end-user-based tracking of the HPC marketplace. It included 181 HPC end-user sites with 3,830 HPC systems.

STUDY FINDINGS

Today, the majority of HPC sites are either currently using the cloud for HPC workloads or are planning to do so soon (see Table 1, below), in line with other Hyperion Research studies. The sector with the highest utilization of cloud for HPC is the private (industry) sector, while government and academic sites are less rapidly adopting cloud for HPC. However, a small portion of both academic sites and government sites are anticipating adopting the cloud in the next six to 18 months. Growth for the cloud in academia, specifically, has been slow for many years, due in large part to the structure of academic funding for HPC resources and the cost structure of the cloud.

The table below shows the current and planned utilization of public cloud resources for sites overall and by sector.

TABLE 1**Current and Planned Public Cloud Use by Sector**

Q: *What best describes your use of the public cloud for your running your HPC workloads today OR in the next 6-18 months?*

	Overall Percent	Industry Percent	Government Percent	Academia Percent
I currently use the public cloud and plan to continue using it in the next 6-18 months.	49.4%	66.7%	20.0%	26.9%
I do not currently use the public cloud, but I plan on using the public cloud in the next 6-18 months.	10.3%	6.9%	15.0%	15.4%
I currently use the public cloud, but do NOT plan on continuing to use it in the next 6-18 months.	5.2%	3.9%	5.0%	7.7%
I do not currently use the public cloud, nor do I plan on using the public cloud in the next 6-18 months.	35.1%	22.5%	60.0%	50.0%

n = 174; 102; 20; 52

Source: Hyperion Research, 2023

FUTURE OUTLOOK

The trend towards increased cloud utilization at HPC sites is expected to continue beyond the next six to 18 months, as the demand for HPC resources for AI/ML/HPDA workloads continues to increase. New AI/ML workloads training on the larger, higher-fidelity models often requires the newest iterations of CPUs/GPUs/xPUs. HPC sites are looking to the cloud to avoid the cost of purchasing every new iteration of accelerators necessary to keep up with user demand. Public cloud offerings can meet the scalability, price, and performance needs for a variety of workload types.

Also, new study findings indicate the movement to cloud resources is motivated by the current issue of rising energy costs worldwide, a growing concern for many HPC decision-makers. The cloud offers the ability to outsource the issue of energy use to CSPs, which can be a smart business strategy for sites that are experiencing prohibitive energy costs or are required to address site sustainability due to new government regulations.

Cloud resources have the potential to democratize access to HPC resources, providing entrance to scientists and researchers around the world, regardless of location or HPC expertise. HPC sites that value the inclusion of new ideas and new partnerships should consider strengthening their relationships with cloud service providers.

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). Hyperion Research provides 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 2023 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.