HYPERION RESEARCH

Special Analysis

2022 HPC End Users Perspectives on AI, Big Data, and HPDA

Tom Sorensen and Earl Joseph
February 2023

HYPERION RESEARCH OPINION

Although artificial intelligence (AI) has seen growing adoption rates among a variety of HPC environments in the previous years, 2022 saw the technology reach a new tier of prominence both in adoption rates and application diversity. The 2022 iteration of Hyperion Research's annual *HPC Multi-Client Study: AI and HPDA Usage and Future Technology Trends* includes a breakout and analysis of this data. The key findings from that report are summarized in this document.

Hyperion Research conducts an annual end-user Multi-Client Study (MCS) to measure and track key trends across the spectrum of the HPC market. The latest iteration of the MCS encompassed 181 HPC end-user sites with 3,830 HPC systems. Reports produced as a result of the study span:

- AI and HPDA Usage and Future Technology Trends
- Vertical/Application Workload Areas and Technical Computing System Software and Middleware
- Use of Public/External Clouds for HPC Workloads, Trends, and Drivers
- Processors, Coprocessors/Accelerators, and HPC Budgets
- Trends and Forecasts in HPC Storage and Interconnects AI and HPDA Usage and Future Technology Trends

To put things into the proper context, it is useful to review the state of AI adoption and use among HPC sites. The study, which includes data collected from representatives of 181 global HPC sites, also shows that 90.6% of respondents indicated current use of AI technology. AI has grown to be a critical and significant portion of HPC datacenter workload portfolios over the last few years.

Overall, the HPC market is expected to grow by about 6% to 7% over the next five years and reach $40 billion by 2026. HPC storage will be the highest growth area of the on-premise HPC market, driven by the capacity demands of AI, machine learning, deep learning, and other analytics methods as well as increasing use of iterative modeling and simulation (M&S) methods. Within that environment, the key trends identified from this HPC end-user base survey include:

- The majority of respondents (94%) use or plan to use AI applications at least some of the time in the next 1-2 years, but on average AI jobs make up only 26% of all HPC workloads at the surveyed sites. In industry, AI was about equal to modeling and simulation as a percent of HPC workload (31% vs. 34% respectively).
- About half of the sites (49%) indicate AI expertise is the number one barrier to increased AI adoption and usage.
- Machine learning (81%) and deep learning (73%) are the top two methodologies expected to be used by the sites surveyed.
- GPUs continue to be the top anticipated HPC/AI/HPDA processing elements at a rate of 74%.

SELECT KEY FINDINGS

Select key findings in this report include the wide adoption rate of AI in the industry sector, a recognition among all HPC respondents of a need for expertise and development skills specific to AI technology, and a significantly low current use-rate of licensed AI-specific software.

Artificial Intelligence (AI) and High-performance Data Analysis (HPDA) are Widely Used in HPC, Especially in Industry

- The majority of respondents (94%) use or plan to use AI applications at least some of the time in the next 1-2 years, but on average AI jobs make up only 26% of all HPC workloads at the surveyed sites. In industry, AI was about equal to modeling and simulation as a percent of HPC workload (31% vs. 34%).
- Both Machine learning (ML) and deep learning (DL) are widely used. 81% of respondents are currently using or anticipating use of ML in the next 1-2 years and 73% indicating the same for DL. Other Hyperion Research studies consistently show that DL development has generally lagged expectations.
- A majority (67%) of respondents today run both simulation and analytics workloads on the same HPC system, and in 6-18 months this figure is not expected to drastically change (63%).
    - Important reasons are that relatively few user sites have managed to acquire budgets for separate systems dedicated to AI and HPDA.
    - Nearly a quarter (24%) of respondents indicated planned use of TPUs (Tensor Processing Units) within the next 12-18 months.
- More than half of the surveyed sites (63%) run at least some AI-HPDA workloads in external clouds.
- Expertise, experience, and applicable knowledge are highly valued among all respondents, indicating investments in AI expertise to be the most critical.

AI Expertise and Development Skills are Highly Valued

- Expertise is now a major concern for both HPC and AI, outranked only by budget and price concerns. A third of HPC sites (33%) reported that lack of knowledge or skilled support staff was one of their top three barriers to expanding HPC on-premises.
    - Only a third of respondents (35%) report that they do not have any staffing concerns within the next year.
- When asked about barriers to furthering AI capabilities, AI-specific expertise was a significant concern. Popular responses included access to AI expertise (49%), skills in AI model development (47%), and skills in AI programming (36%).

Industry Reported that AI Includes a Broad Plurality of Workloads

- Industry users report that AI/ML/DL job types will comprise the plurality of cycles for the first time, with nearly 34%. Not only does this anticipated shift indicate a significant moment in the journey of AI, but it also implies a heightened and continued galvanization of the applicability, pervasiveness, and trustworthiness of the technology. Notably, industry users are anticipating AI workloads will comprise over 18% more cycles than their government counterparts.

- Government users, conversely, expect traditional modeling and simulation workloads to continue to comprise over 60% of their cycles, over 30% more than the industry sector.

AI Software Licenses Only Starting to Gain Traction in Industry

- Academia and government respondents indicated there is no current use of AI-specific paid software licenses on-premise at a nearly unanimous 97%. This sentiment is mirrored in other questions, for example industry respondents indicated a higher rate of early adoption, diverse workload types, and interest in non-conventional means of optimization.

FUTURE OUTLOOK

Hyperion Research projects that revenue for HPC servers used mainly for machine learning will expand at a robust 16.6% CAGR (2020-2026), while those used primarily for deep learning will grow at an even higher 32.2% CAGR. Deep learning shows some of the highest potential among the various AI techniques, but machine learning is the most prominently used methodology today.

Hyperion Research anticipates that revenue for ML-dedicated machines will reach almost $1.8 billion in 2026, while DL-dedicated machines will account for more than $1.4 billion. Other subsegments of AI continue to grow as well, mainly the sometimes under-appreciated graph analytics methodology, which is expected to exhibit healthy growth and increased adoption over the next five years.

Hyperion Research forecasts that AI and HPDA will have particularly high growth, significantly outperforming the HPC market as a whole. Since AI-HPDA is a specialized field, it also has its own unique defined application areas.

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
