HYPERION RESEARCH

Market Forecast

Worldwide HPC-based Artificial Intelligence (AI) Market Forecast Update, 2016–2021

Steve Conway, Earl Joseph, Bob Sorensen, and Kevin Monroe
July 2017

HYPERION RESEARCH OPINION

Worldwide HPC-based Artificial (AI) Market Revenue Snapshot

Hyperion Research forecasts that the worldwide HPC server-based AI market will expand at a 29.5% CAGR to reach more than $1.26 billion in 2021, up more than three-fold from $346 million in 2016. We define the HPC AI market as a subset of the high-performance data analysis (HPDA) market that includes machine learning (ML), deep learning (DL) and other AI workloads running on HPC servers.

Hyperion Research treats machine learning, deep learning and AI as *methodologies* rather than distinct market segments such as bio-life sciences, defense, or weather/climate. ML, DL and AI are more analogous to methods used in the long-standing HPC modeling and simulation market, such as computation fluid dynamics (CFD) or finite element analysis (FEA). Each of these methodologies: ML, DL, CFD, FEA -- is applicable across multiple market segments. In addition, more than one of these methodologies (e.g., simulation and advanced analytics) may be used in combination to solve an HPC problem.

Interest in ML, DL and other AI work within the HPC community has grown large enough for us to size HPC server spending related to these methodologies and provide a five-year forecast in this document.

Note: The forecasts provided in this study include only server systems used in technical computing applications. Systems sold into commercial (nontechnical) applications and desktop technical computers are not included in this study.

Note: This page is intentionally blank.

IN THIS REPORT

This Hyperion Research study presents our first five-year forecast (2016-2021) for the HPC-based AI market, from both the worldwide and U.S. perspectives.

Definitions Used in This Forecast

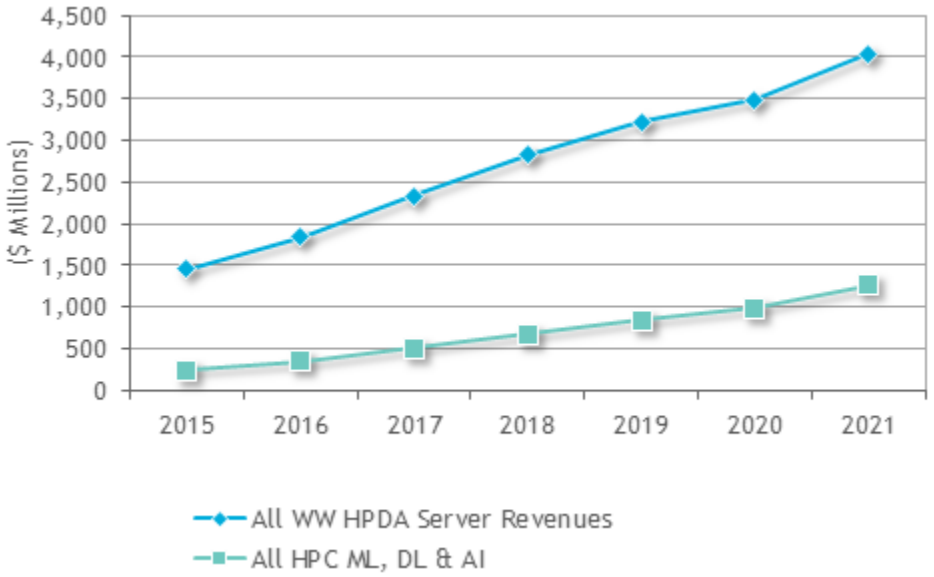
- **Artificial Intelligence (AI):** a broad, general term for the ability of computers to do things human thinking does (but NOT to think in the same way humans think. AI includes machine learning, deep learning (a.k.a. cognitive computing) and more minor methodologies.
- **Machine learning (ML)**: a process where examples are used to train computers to recognize specified patterns, such as human blue eyes or numerical patterns indicating fraud. The computers are unable to learn beyond their training and human oversight is needed in the recognition process.
- **Deep Learning (DL):** an advanced form of machine learning that uses digital neural networks to enable a computer to go beyond its training and learn on its own, without explicit programming or human oversight.

Worldwide HPC AI Server Revenues vs. All HPDA Server Revenues

Figure 1 displays Hyperion's five-years forecasts for high performance data analysis (HPDA) server revenues, along with our five-year projection for the subset of revenues for HPDA servers used primarily (>50% of cycles) for ML, DL or other AI workloads. We predict that during the period shown in Figure 1 (2015-2021), HPC server revenue for the whole AI category (ML, DL, et al.) will expand at a 29.5% CAGR to reach $1.2 billion, or about 31% of the $4.0 billion total for all HPDA server revenue.

FIGURE 1

Worldwide HPC AI Server Revenues vs. All HPDA Server Revenues



Source: Hyperion Research 2017

Table 1 shows the revenue figures associated with the Figure 1 graph. As Table 1 indicates, worldwide revenue for the HPC AI server market was $346 million in historical year 2016, representing 18.8% of worldwide HPDA server revenues. We predict that the CAGR for the AI portion of the HPDA server market (29.5%) will be substantially higher than the CAGR for the overall HPDA server market (17.0%) during the forecast period.

TABLE 1

Worldwide HPC AI Server Revenues vs. All HPDA Server Revenues ($ Millions)

	2015	2016	2017	2018	2019	2020	2021	CAGR 16-21
Total WW HPDA Server Revenues	$1,455	$1,845	$2,333	$2,830	$3,224	$3,488	$4,040	17.0%
Total HPC-Based AI (DL, ML, and Other)	$246	$346	$501	$673	$845	$986	$1,260	29.5%

Source: Hyperion Research 2017

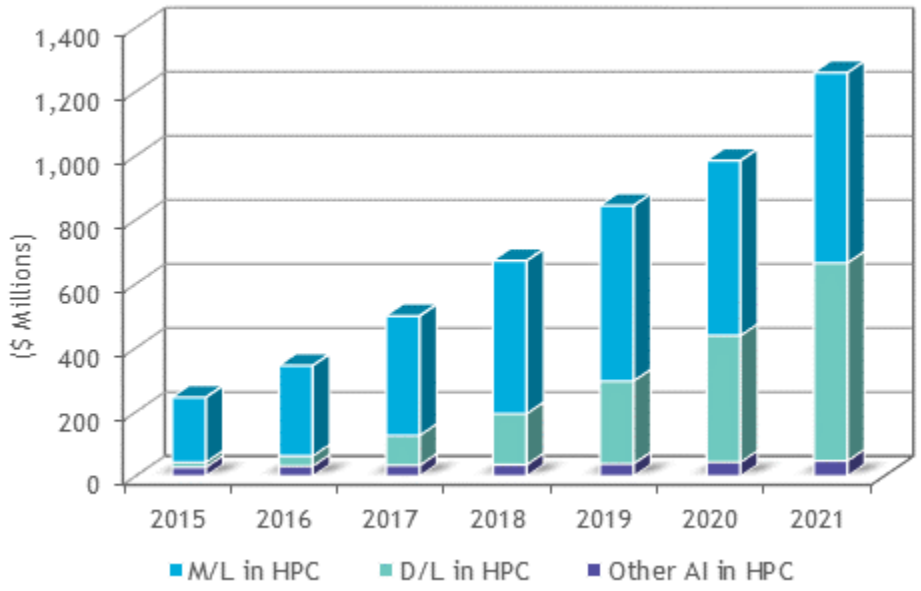
WORLDWIDE HPC AI SERVER MARKET FORECAST BY METHODOLOGY

Figure 2 and Table 2 show Hyperion's five-year forecast for worldwide HPC AI server revenues, split out by AI methodology areas.

- We expect DL to grow exceptionally fast from its tiny 2015 starting point ($15 million) to exceed $600 million in 2021 (81.5% CAGR), as the prerequisites for production-grade DL, especially access to adequate data volumes, begin to exist in more HPC-supported markets.
- ML growth should be impressive though less stellar (16.0% CAGR). As DL becomes practical for more use cases in more HPC market segments, some of the most challenging ML workloads will migrate to DL methods while some of the least challenging ML job types trickle down to desktop computers and over time to portable electronic devices.

FIGURE 2

Worldwide M/L, D/L & AI HPC-Based Revenues



Source: Hyperion Research 2017

TABLE 2

Worldwide HPC AI Server Revenues by Methodology ($ Millions)

	2015	2016	2017	2018	2019	2020	2021	CAGR 16-21
ML in HPC	$203	$282	$373	$478	$548	$546	$594	16.0%
DL in HPC	$15	$31	$93	$159	$258	$395	$618	81.5%
Other AI In HPC	$28	$32	$34	$36	$39	$44	$48	8.4%
Total	$246	$346	$501	$673	$845	$986	$1,260	29.5%

Source: Hyperion Research 2017

THE U.S. HPC AI MARKET FORECAST

U.S. HPC-related AI server sales (machine learning, deep learning and the catch-all "Other" category) generated revenues of $164 million in 2016 (Table 3). We project that revenues will expand at a strong CAGR of about 26% to exceed $500 million in 2021, or about 31% of the projected $1.7 billion figure for all HPDA server sales worldwide in that year.

U.S. HPC AI Server Revenues vs All U.S. HPDA Server Revenues

Table 3 displays historical (2015-2016) and forecast (2016-21) U.S. revenues for HPC AI servers as a subset of all U.S. HPDA server revenues. Hyperion Research projects that the U.S. AI component will expand during the forecast period (26.5% CAGR) at nearly double the growth rate (14.1%) of all U.S. HPDA servers, increasing from 16.9% of the HPDA category figure in 2016 to 31.1% in 2021.

TABLE 3

U.S. HPC AI Server Revenues vs All U.S. HPDA Server Revenues ($ Millions)

	2015	2016	2017	2018	2019	2020	2021	CAGR 16-21
U.S. All HPDA Server Revenues	$727	$876	$1,073	$1,259	$1,402	$1,482	$1,697	14.1%
U.S. HPC AI Server Revenues	$123	$164	$230	$299	$368	$419	$529	26.4%

Source: Hyperion Research 2017

U.S. HPC AI Server Forecast by Methodology Area

Within the U.S. HPC server market (Table 4), as in the equivalent worldwide market (Table 2), Hyperion Research forecasts that DL will be the fastest-growing AI methodology during the five-year forecast period, expanding at a stellar 77.1% CAGR from a mere $15 million in 2016 to $260 million in 2021. In 2021, both in the U.S. and globally, HPC DL will for the first time surpass HPC ML as the largest source of AI server revenues.

Fueling this accelerating growth will be DL's evolution from a largely exploratory methodology today to a production-grade approach for a growing number of HPC market segments and use cases. Prominent among these use cases, we predict, will be the development of autonomous vehicles in the automotive industry—a task that will pair established simulation with advanced analytics (AI) methods—and the use of HPC server systems as decision-support tools for patient diagnosis and treatment planning in the global health care industry.

TABLE 4

U.S. HPC AI Server Revenues by Methodology ($ Millions)

	2015	2016	2017	2018	2019	2020	2021	CAGR 16-21
ML in HPC	$101	$134	$172	$213	$238	$232	$249	13.2%
DL in HPC	$8	$15	$43	$71	$112	$168	$260	77.1%
Other AI In HPC	$14	$15	$16	$16	$17	$19	$20	5.8%
Total	$123	$164	$230	$299	$368	$419	$529	26.4%

Source: Hyperion Research 2017

Larger Growth Drivers

Hyperion Research expects several key factors to drive healthy growth across the worldwide HPC server market, enabling this market to continue expanding faster than the enterprise IT server market. These drivers include:

- Requirements for new HPC systems with a broad range of architectures to support development and operational capabilities in the artificial intelligence sector, especially in the area of deep learning.
- Many revenue gains in the HPC sector writ large will be driven by those within the HPDA sector. Hyperion Research analysts project that HPDA server revenues will grow at a CAGR of almost 17.0 % out to 2021, and new commercial analytics emerging within that space will see a CAGR of over 25% during the same time frame. Examples here include new big data applications that are running in non-traditional HPC environments but that use HPC hardware, such as in the financial services and cyber security sectors.
- New and rapidly growing opportunities to support the continued migration and expansion of enterprise HPC workloads to cloud-based ecosystems. In many cases, HPC in the cloud operations will be used not as a replacement scheme but instead to augment critical on premise HPCs capabilities. Additional growth will come as these HPC in the cloud offerings support a wider range of virtual environments targeted for key application sectors, lowering the barriers to entry for new HPC users.
- The expanding role and diversity of new big data analytics running in non-traditional HPC environments, especially in the financial services, healthcare and cyber security sectors. Of particular importance, will be the ability of HPC systems to empower big data analysis on a near-real time basis, an increasingly necessary requirement for many application spaces.
- The continued expansion of HPC servers into the traditional modeling and simulation environment as more and more commercial and government users turn to advanced computing to meet their toughest computational requirements for larger problem sizes, higher modeling fidelity, and more aggressive iteration methods, all designed to address the need for faster turnaround time.

FORECAST METHODOLOGY

The forecasts in this study are based on multiple information sources, including Hyperion Research's technical computing systems quarterly census database, vendor results for the historical years, discussions with vendors and users on future business directions and expectations, end-user studies, and in-depth interviews with users.

The forecasts were developed based on Hyperion Research's technical computing systems forecast model, which targets compute servers. This model initially considers competitive price band segments (supercomputers, technical divisional servers, technical departmental servers, and technical workgroup servers), forecasting system unit shipments, revenue, and average sales price by industry/application segment. The forecasts also take into consideration estimates for second-tier and new-entrant vendors selling into the HPC server market space.

The forecasts provided in this study include only server systems used in technical computing applications. Systems sold into commercial (nontechnical) applications and desktop technical computers are not included in this study.

About Hyperion Research, LLC

Hyperion Research, consisting of the former IDC high performance computing (HPC) analyst team, provides HPC information, analysis, and recommendations based on technology and market trends. 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.hpcuserforum.com and www.HPCatHyperion.com