

## Market Forecast

# Worldwide HPC-based Artificial Intelligence (AI) Market Forecast Update, 2017-2022

Steve Conway, Earl Joseph, Bob Sorensen, and Alex Norton  
*October 2018*

### HYPERION RESEARCH OPINION

---

Hyperion Research forecasts that the worldwide HPC server-based AI market will expand at a robust 27.1% CAGR from 2017-2022 to reach \$1.6 billion in 2022, up more than four-and-a-half-fold from \$346 million in 2016, the first year in which we tracked this figure. 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 sold into public and private sector environments and do not include other elements of the HPDA and AI ecosystem, notably storage, software and technical support.*

*Note: This page is intentionally blank.*

## IN THIS REPORT

This Hyperion Research study presents our five-year forecast (2017-2022) for the HPC-based AI market, from both the worldwide and U.S. perspectives.

### Definitions Used in This Forecast

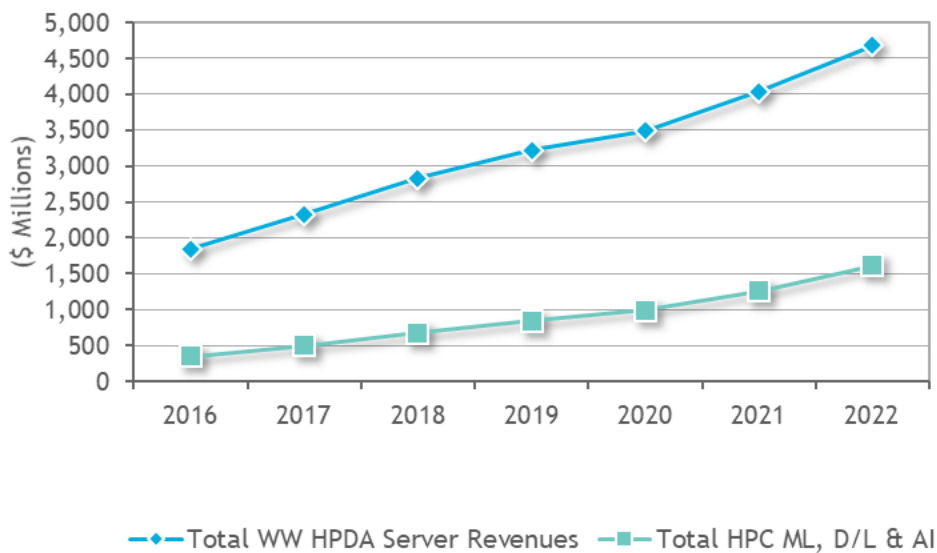
- **Artificial Intelligence (AI):** a broad, general term for the ability of computers to do things human thinking does (but NOT to think in the same way humans think). AI includes machine learning, deep learning (a.k.a. cognitive computing) and more minor methodologies.
- **Machine learning (ML):** 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 2017-2022, HPC server revenue for the whole AI category (ML, DL, et al.) will expand at a 27.1% CAGR to reach \$1.6 billion, or about 35% of the \$4.7 billion total for all HPDA server revenue (see Figure 1).

## FIGURE 1

### Worldwide Server-Based AI Revenues



Source: Hyperion Research 2018

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 \$501 million in 2017, representing 21.5% of worldwide HPDA server revenues. We predict that the CAGR for the AI portion of the HPDA server market (27.1%) will be substantially higher than the CAGR for the overall HPDA server market (14.9%) during the forecast period.

**TABLE 1**

**Worldwide HPC-Based AI Revenues vs Total HPDA Revenues (Millions)**

	2017	2018	2019	2020	2021	2022	CAGR 17-22
Total WW HPDA Server Revenues	2,333	2,830	3,224	3,488	4,040	4,680	14.9%
Total HPC-Based AI (DL, ML, and Other)	501	673	845	986	1,260	1,664	27.1%

Source: Hyperion Research 2018

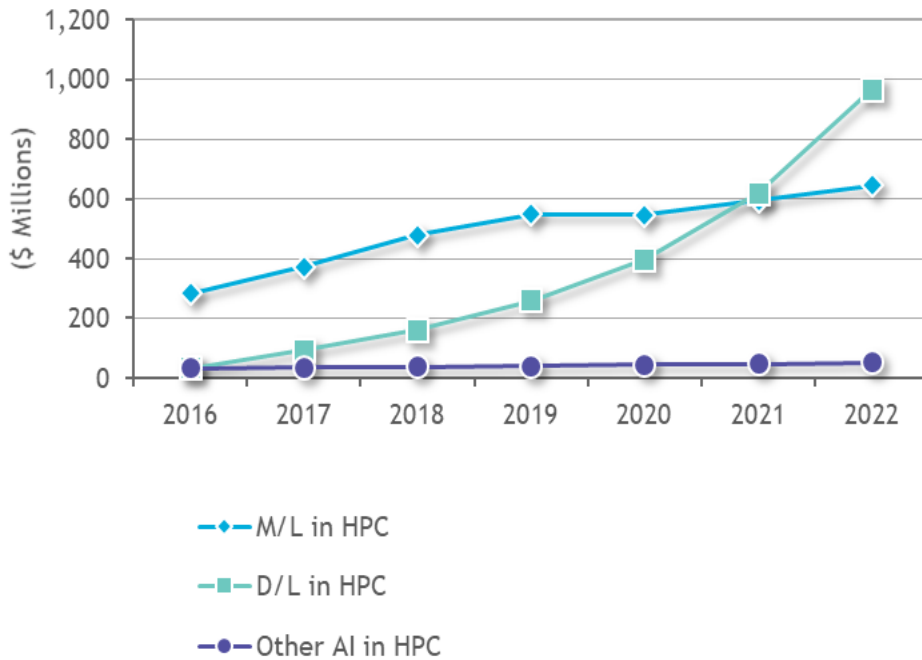
**WORLDWIDE HPC AI SERVER MARKET FORECAST BY METHODOLOGY**

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

- We expect DL to grow exceptionally fast from its 2017 starting point (\$93 million) to reach about \$966 million in 2022 (59.7% CAGR), as the prerequisites for production-grade DL, especially access to adequate usable data volumes, begin to exist in more markets needing HPC.
- ML growth should be impressive though less stellar (11.6% 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 standard enterprise servers, desktop computers and over time to portable electronic devices.

**FIGURE 2**

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



Source: Hyperion Research 2018

**TABLE 2**

**Worldwide M/L, D/L & AI HPC-Based Revenues (\$ Millions)**

	2017	2018	2019	2020	2021	2022	CAGR 17-22
M/L in HPC	373	478	548	546	594	646	11.6%
D/L in HPC	93	159	258	395	618	966	59.7%
Other AI In HPC	34	36	39	44	48	52	8.9%
<b>Total</b>	<b>501</b>	<b>673</b>	<b>845</b>	<b>986</b>	<b>1260</b>	<b>1664</b>	<b>27.1%</b>

Source: Hyperion Research 2018

## THE U.S. HPC AI MARKET FORECAST

U.S. HPC-related AI server sales (machine learning, deep learning, graph and the catch-all "Other" category) generated revenues of \$230 million in 2017 (Table 3). We project that revenues will expand at a strong CAGR (24.6%) to reach about \$692 million in 2022, or about 35% of the projected \$1.9 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 (2017) and forecast (2018-22) 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 (24.6% CAGR) at nearly double the growth rate (12.6%) of all U.S. HPDA servers, increasing from 21.4% of the HPDA category figure in 2017 to 35.6% in 2022.

**TABLE 3**

### **US HPC-Based AI Revenues vs Total HPDA Revenues (\$ Millions)**

	2017	2018	2019	2020	2021	2022	CAGR 17-22
Total WW HPDA Server Revenues	1,073	1,259	1,402	1,482	1,697	1,942	12.6%
Total HPC ML, D/L & AI	230	299	368	419	529	692	24.6%

Source: Hyperion Research 2018

### ***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. In 2021, both in the U.S. and globally, HPC DL for the first time surpassed 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. Both major use cases will employ coupled environments: local (on-premise) computing for tasks needing low latency (e.g., car-car communications in automated driving systems); and cloud computing for tasks that can tolerate higher latency communications (e.g., urban traffic management in ADS).

**TABLE 4****US M/L, D/L & AI HPC-Based Revenues (\$ Millions)**

	2017	2018	2019	2020	2021	2022	CAGR 17-22
M/L in HPC	172	213	238	232	249	268	9.4%
D/L in HPC	43	71	112	168	260	402	56.4%
Other AI In HPC	16	16	17	19	20	22	6.8%
Total	230	299	368	419	529	692	24.6%

Source: Hyperion Research 2018

## Larger Growth Drivers

Hyperion Research expects several key factors to drive healthy growth across the worldwide HPC 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, primarily for machine learning and graph analytics today but increasing in the area of deep learning.
- Many revenue gains in the HPC sector will be driven by those within the HPDA sector. 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, healthcare, retail 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' capacities and 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 continued expansion of HPC servers used for established modeling and simulation, as more commercial users and SMBs are driven by competitive forces to advance their computing capabilities on premise, in the cloud, or in both environments.

## 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 forecast provided in this study include only server systems acquired primarily (more than 50% of planned use) to run HPDA workloads, including AI workloads. Hyperion Research studies have consistently found that HPC is indispensable at the forefront of AI R&D. This forecast covers only the portion of the AI server market that employs HPC resources, in both public- and private sector environments.



## 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 and 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](http://www.hpcuserforum.com) and [www.HyperionResearch.com](http://www.HyperionResearch.com)

---

## Copyright Notice

Copyright 2018 Hyperion Research LLC. Reproduction is forbidden unless authorized. All rights reserved. Visit [www.hpcuserforum.com](http://www.hpcuserforum.com) or [www.HyperionResearch.com](http://www.HyperionResearch.com) to learn more. Please contact 612.812.5798 and/or email [info@hyperionres.com](mailto:info@hyperionres.com) for information on reprints, additional copies, web rights, or quoting permission.