# Market Forecast

## Worldwide HPC Server Market Forecast Update, 2023-2028

Melissa Riddle and Earl Joseph
May 2024

### HYPERION RESEARCH OPINION

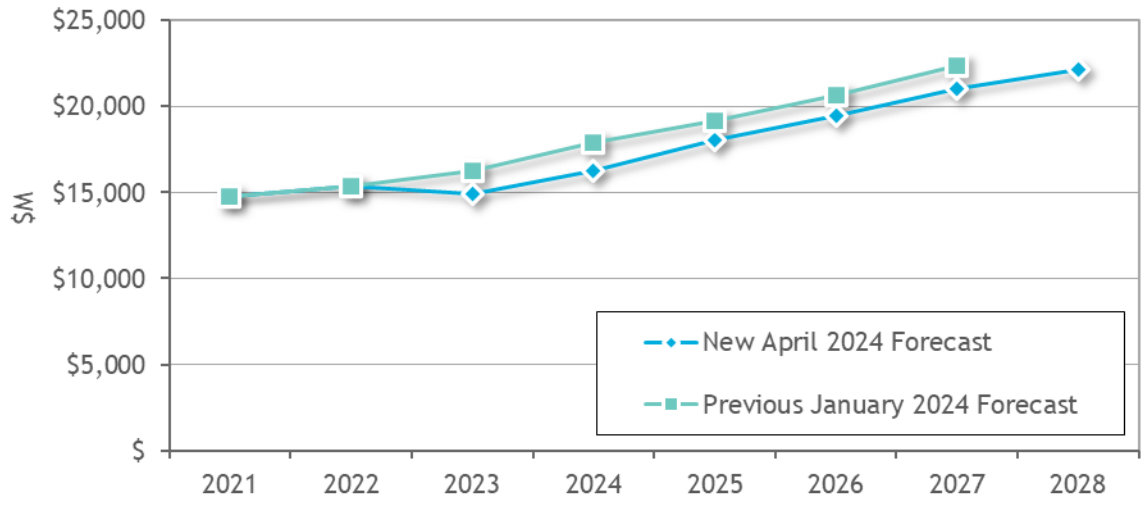
This Hyperion Research study presents the latest five-year forecast (2023-2028) for HPC on-premises technical servers. Worldwide revenue for the HPC technical server market in 2023 was $15.0 billion, representing a slight decline (-2.7%) over 2022 revenues. Hyperion Research now predicts that the HPC technical server market will grow at an 8.2% CAGR between 2023 and 2028 to reach $22.2 billion in 2028 (see Figure 1). Compared with the previous version of the forecast, there have been some downgrades to the revenues for 2023 through 2027, primarily due to rescheduling of some large exascale systems and ongoing supply chain delays.

The most important market drivers affecting the forecast are:

- Ongoing worldwide supply chain delays
- Planned acceptances (and related delays) of exascale and pre-exascale systems
- Emerging technologies such as LLMs, AI, ML, and DL
- The global economy and geopolitical forces

### FIGURE 1

**Updated Worldwide HPC Server Revenue Forecast ($M)**




Source: Hyperion Research, May 2024

*Note: This page is intentionally blank.*

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## IN THIS REPORT

This Hyperion Research study presents the latest five-year forecast (2023-2028) for HPC server systems. Worldwide revenue for the HPC technical server market declined -2.7% from 2022 to 2023 (see Figure 2). HPC server revenues are expected to be stronger in 2024, growing 8.7% from the previous year as the market recovery continues and exascale systems continue to ramp up. Looking further into the forecast, the HPC technical server market is projected to grow at an 8.2% CAGR overall during the forecast period (2023-2028) to reach $22.2 billion in 2028.

Compared to the previous version of the forecast, there have been some slight downgrades to the forecasted HPC server revenues for 2023 through 2027 (see Table 1). Each year of the forecast is expected to increase over the previous year, from 2024 through 2028.

### Table 1

### Updated Worldwide High-Performance Computing Server Market Forecast ($M)

| | 2022 | 2023 | 2024 | 2025 | 2026 | 2027 | 2028 | CAGR '22-'27 | CAGR '23-'28 |
|---|---|---|---|---|---|---|---|---|---|
| New April 2024 Forecast | $15,369 | $14,954 | $16,254 | $18,058 | $19,501 | $21,025 | $22,152 | 6.5% | 8.2% |
| Previous January 2024 Forecast | $15,369 | $16,267 | $17,873 | $19,184 | $20,694 | $22,370 | | 7.8% | |

Source: Hyperion Research, May 2024

## FORECAST ASSUMPTIONS

Table 2 summarizes the major forecast assumptions and their impacts on the HPC server market.

### Table 2

### Key Forecast Assumptions for the Worldwide HPC Market, 2023-2028

| Assumption | Impact |
|---|---|
| **Global economic conditions** point to a slow and perhaps fragile overall global economic environment for the next few years. Global GDP growth in 2024 is projected to be 3.1%. OECD estimates call for even lower growth rates for the US (2.6%) and the Euro area (0.7%) in 2024. Similarly, in 2025 global GDP is expected to increase 3.2% while the US increases only 1.8% and the Euro area 1.5%. | Negative impact. Lower to slightly lower than normal growth in many segments of the market. |
| **Global supply chain delays** and chip shortages are continuing to improve but will persist for several more years. | Negative impact. Significant decrease to the server forecast throughout the forecast period. |

## Table 2

## Key Forecast Assumptions for the Worldwide HPC Market, 2023-2028

| Assumption | Impact |
|---|---|
| Enthusiasm for the potential of **large language models (LLMs)** across a wide range of commercial sectors will drive many HPC end users to consider integrating LLMs into their overall compute environment, presenting challenges and opportunities for both general-purpose HPC and LLM-specific hardware and software suppliers in what could be a rapidly expanding segment of the overall HPC market. | Increased growth. The inclusion of LLMs will both increase the overall size of the HPC market as well as provide a strong growth driver for the total market. They will also add new types of buyers to the HPC market. |
| The roll out of previously announced **exascale systems** targeted for installations at large, typically government funded HPC centers, will continue as planned (with some delays), but subsequent systems will likely become more modular and designed to serve a variety of diverse workloads, challenging both HPC vendors and end users to accurately characterize their planned and anticipated workloads. | Growth. The average selling price of a high-end HPC will likely decline (after the first 4 to 6 exascale systems are installed) as these systems will increasingly comprise different hardware partitions added incrementally over time and augmented with frequent upgrades, resulting in a longer total time in service for any individual machine. However, the number of such systems is also expected to increase over the forecast period. |
| The HPC sector writ large is moving back to 'old school' aspirations for **vertical integration**, but with a modern twist, as varying parts of the HPC ecosystem seek to expand their technical capabilities beyond their traditional product mix. Examples today include HPC server suppliers offering cloud services, cloud service providers developing AI accelerators for integration into their proprietary HPC systems, and GPU suppliers expanding their capabilities into networks, related high-speed interconnects, and even the data center. | These HPC vendor forays into new areas across the board engender significant opportunities and challenges for HPC users to assemble the exact HPC best suited to their specific computational workload. |
| With the profusion of HPC use cases comes a profusion of **HPC metrics**. In addition to traditional HPC metrics like LINPACK, TPP and STREAMS, there are now considerations for green, sustainable, or renewable performance, time to solution, power to solution, CO2 production, ROI, and PUE along with specific metrics for AI workloads, many with mixed or low precision, and even benchmarks for quantum simulation capabilities. | HPC vendors must navigate a more complex set of benchmarks from which to target their designs, especially when trying to anticipate which metrics will carry the most weight in any given procurement. |
| **Open-standard or open-source assets** like the RISC-V ISA will enable flexible technological advancements as well as allow for developments in indigenous semiconductor and technology production. | This trend also has the potential to reduce supply-chain concerns or production bottlenecks but will require a full ecosystem in order to become mainstream. |
| The explosion of **generative AI** interest, especially in the public eye, will fuel efforts to acquire, regulate, police, and govern the development and use of AI models. Trust, reproducibility, and bias-reduction are some of the major focuses. | Increased growth. While many of the talking points may be similar in a global sense, regional and national differences contribute significantly to policy perspectives. |
| A widening **rift between China and Western nations** in scientific and economic cooperation will magnify calls for indigenous technology, slow | This trend will continue to raise the stakes on historically tense arrangements such as with semiconductor production. |

## Table 2

## Key Forecast Assumptions for the Worldwide HPC Market, 2023-2028

| Assumption | Impact |
|---|---|
| economic growth between traditionally aligned trading partners, and limit the accessibility of reliable information from China. | |
| The conversation about employing **cloud-based resources** has turned from an adversarial "cloud vs. on-premises" to a conversation about how to leverage both to provide a user with an optimal balance of performance, cost, features, and skillsets. | Stronger growth rates for spending on HPC resources in the cloud and slightly lower growth rates for on-premises. |
| Attention is being shifted from storage management to **data management**. Providing the data when it's needed, where it's needed, goes beyond just capacity and performance. Security, accessibility, and governance are becoming critical factors within an overall data management strategy. | Strong growth for storage solutions, both on-premises and cloud-based, that provide easy implementation of data management capabilities. |
| Many HPC sites are identifying the first steps towards improving **sustainability**, driving the widespread implementation of telemetry solutions to track energy use of hardware components and assess software cost efficiency. | Buyers will increasingly require more sustainable products and will consider efficiency and sustainability more frequently in procurements. |
| Introducing **AI, ML, DL, and HPDA** to new industries and various new application types is in part enabled by the increased popularity and widespread use of frameworks, which allow a speedier path to adoption for entry-level users. | Increased growth. Various types of AI frameworks will create a new market segment in frameworks and foundational models. |
| The **historically cyclical nature of the HPC market** has also been considered. The HPC market typically alternates between patterns of wide-scale procurements followed by periods of relative calm, or in some cases, limited contractions. | Healthy growth rates during the forecast period, although there could potentially be large growth spikes in certain areas. |

Source: Hyperion Research, May 2024

## THE ON-PREMISES HPC SERVER MARKET FORECAST BY COMPETITIVE SEGMENT

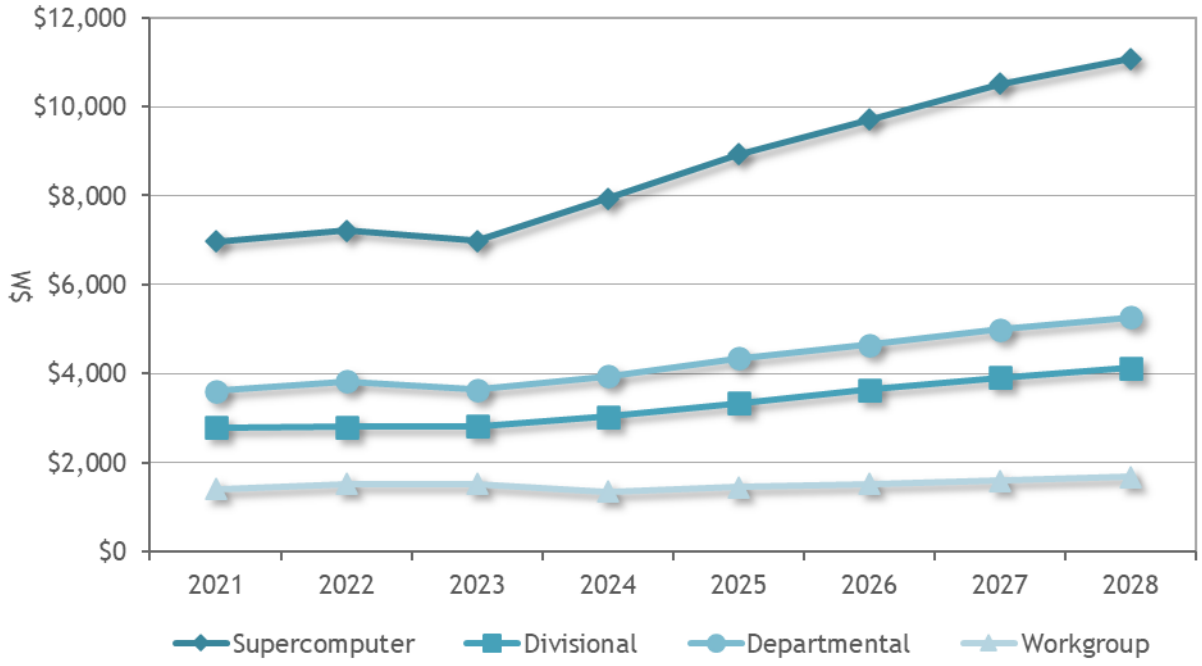
Figure 2 and Table 3, below, show HPC server revenues by competitive segment.

- The Supercomputer market segment is projected to be both the largest and fastest growing segment, reaching $11.1 billion in 2028 with 9.7% CAGR. This is heavily driven by the worldwide roll-out of large, exascale systems. Throughout the forecast, Supercomputers make up about half of all HPC server revenues.
- The Divisional segment is anticipated to have the second largest growth rate (8.0% CAGR), reaching $4.1 billion in 2028.
- The Departmental segment is expected to have five-year growth slightly lower than the HPC server market writ large (7.7% CAGR vs. 8.2% CAGR, respectively), reaching nearly $5.3 billion in 2028. After Supercomputers, Departmental systems are the next largest segment.

- The Workgroup market segment is projected to remain relatively flat over the forecast period (2.1% CAGR), representing about $1.7 billion in revenues in 2028. The Workgroup segment continues to experience erosion from HPC cloud usage and has been the slowest to recover from the economic slowdown.

## FIGURE 2

### Updated Worldwide HPC Server Revenue Forecast by Segment ($M)




Source: Hyperion Research, May 2024

## Table 3

### Worldwide Technical Server Market Forecast by Price Band ($M)

| | 2022 | 2023 | 2024 | 2025 | 2026 | 2027 | 2028 | CAGR '23-'28 |
|---|---|---|---|---|---|---|---|---|
| Supercomputer | $7,219 | $6,979 | $7,934 | $8,933 | $9,698 | $10,509 | $11,070 | 9.7% |
| Divisional | $2,805 | $2,812 | $3,028 | $3,336 | $3,638 | $3,915 | $4,126 | 8.0% |
| Departmental | $3,826 | $3,644 | $3,941 | $4,343 | $4,648 | $4,999 | $5,269 | 7.7% |
| Workgroup | $1,519 | $1,518 | $1,351 | $1,447 | $1,516 | $1,601 | $1,687 | 2.1% |
| Total HPC Server Revenues | $15,369 | $14,954 | $16,254 | $18,058 | $19,501 | $21,025 | $22,152 | 8.2% |

Source: Hyperion Research, May 2024

## Exascale and Pre-Exascale Systems

Exascale and pre-exascale systems make up a large portion of the Supercomputer and overall HPC Server forecasts. Although the number of systems installed per year is expected to rise each year in the forecast, the total revenue from these systems is expected to be somewhat lower in 2025 and beyond as system prices are expected to decrease from over $500 million each, to around $250 to $300 million each. The countries/regions that anticipate installing the most exascale and pre-exascale systems are EU/UK/Germany, China, and the United States.

## METHODOLOGY

The forecasts in this study are based on a number of Hyperion Research information sources, including conducting multiple surveys of buyers and vendors, an in-house technical computing systems quarterly census database, vendor results for the historical years, discussions with vendors and users on future business directions and expectations, and in-depth interviews with users.

The forecasts were developed using Hyperion Research's technical computing systems forecast model, which targets compute servers by 13 verticals and by 26 countries/regions. This model considers competitive 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 include estimates for second-tier and new-entrant vendors to the HPC server market.

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

*Note: Numbers in this document may not be exact due to rounding. Monetary values given in USD.*

## DEFINITIONS

### *Technical Computing (HPC)*

Hyperion Research uses the terms *technical computing* and *high-performance computing (HPC)* to encompass the entire market for computer servers used by scientists, engineers, analysts, and other groups using computationally and/or data-intensive modeling and simulation applications. An on-premises system primarily used for HPC workloads (at least 50%) can be referred to as an HPC system, or simply an HPC. Systems acquired by cloud service providers for the purpose of hosting cloud workloads are excluded, as Hyperion Research separately tracks spending for HPC usage in cloud environments.

Technical servers range from small servers costing less than $10,000 to the large-capability machines valued in hundreds of millions of dollars. In addition to scientific and engineering applications, technical computing includes related markets/applications areas such as economic analysis, financial analysis, animation, server-based gaming, digital content creation and management, business intelligence modeling, and homeland security database applications. These areas are included in the technical computing market based on a combination of historical development, applications type, computational intensity, and associations with traditional technical markets.

## *Accounting for Exascale and Pre-Exascale Systems*

Exascale and pre-exascale systems typically have high costs and often involve a large portion of non-recurring engineering (NRE) in the server contract. As with other HPC servers, Hyperion Research recognizes the entire cost of the server contract as server revenue.

Hyperion Research's accounting rules record these server revenues as a lump sum at the time of system acceptance (as indicated by the purchaser), regardless of when actual payments may have been made. This large dollar amount accepted all at once may give the appearance of extraordinary growth or decline in a single quarter or year. Whenever an exceptionally costly system is accepted in a particular period, this will be clearly noted.

When necessary (such as when multiple exascale systems are accepted in the same period), these exceptionally large systems may be pulled out as a separate line item within the supercomputer segment to preserve forecasting of the general HPC market.

## *Competitive Segments*

Based on input from HPC vendors and end users, Hyperion Research created four competitive segments to reflect the trends in the HPC technical server market. These competitive segments are based on average selling prices and defined as follows:

- **Supercomputers:** Technical servers that sell for $500,000 or more.
- **Divisional servers:** Technical servers that sell for $250,000–$499,999.
- **Departmental servers:** Technical servers that sell for $100,000–$249,999.
- **Workgroup servers:** Technical servers that sell for less than $100,000.

## 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 and www.HyperionResearch.com

---