2017 End-user MCS Study

Technical Computing: Servers, Processors, and Coprocessors, Current Usage Levels, Drivers, and Forecasts

Bob Sorensen, Steve Conway, and Earl Joseph
September 2017

HYPERION RESEARCH OPINION

This special study is part of the fifth edition of Hyperion Research's high-performance computing (HPC) end-user-based tracking of cluster and other server systems, processors, and coprocessors used in the HPC marketplace.

High performance computing continues to be an integral component in driving progress in a wide range of scientific and engineering applications around the world that include foundational research, aerospace, automotive, biotechnology, manufacturing, pharmaceutical workloads, and a critical set of national security missions. In addition, the increasing importance of HPC in emerging use cases such as big data analytics, including deep learning, is driving further growth in HPC revenue.

Worldwide revenue for the HPC technical server market grew 4.4% in 2016 to a record $11.2 billion, up from $10.7 billion in 2015. Current projections call for growth in 2017 of 7.3%, resulting in a total HPC technical server market for 2017 worth over $12 billion.

Hyperion Research is also anticipating steady HPC revenue growth in the out years as well. Hyperion Research projects that the worldwide HPC server sector will grow at a CAGR of 5.8% between 2016 and 2021, reaching total revenues of $14.7 billion by 2021.

- Additional HPC-related technologies including storage, middleware, applications, and service software nearly doubled the total revenue for the HPC sector to $22.4 billion in 2016, with a projected CAGR of 6.2% between 2016-2021, resulting in a total HPC revenue base of over $30 billion in 2021.
- Within the HPC broader market, storage is the fastest growing segment. Hyperion Research reported that 2016 HPC-related storage revenue was $5.1B and forecasts it will grow to $6.7B in 2019 – a 9.4% CAGR, comprising almost a quarter of all HPC revenues.

Note: this page is intentionally blank.

Executive Summary

For this fifth edition of the HPC end-user Special Study, Hyperion Research conducted in-depth interviews with one or more representatives of 111 HPC sites around the world. These sites had a total of 972 technical server systems installed, and the average in the present study was 8.8 systems per site.

Worldwide revenue for the HPC technical server market grew 4.4% in 2016 to a record $11.2 billion, up from $10.7 billion in 2015. Current Hyperion Research projections call for growth in 2017 of 7.3%, resulting in a total HPC technical server market for 2017 with over $12 billion.

- Hyperion Research is anticipating healthy HPC revenue growth in the out years as well. Indeed, Hyperion Research projects that the worldwide HPC server sector will grow at a CAGR of 5.8% between 2016 and 2021, reaching total revenues of $14.7 billion by 2021.

The worldwide technical server market will continue to grow at a healthy 5.8% CAGR, to reach about $14.7 billion in 2021.

Hyperion Research analysts are expecting a number of factors to drive growth rates across all segments of the HPC technical server sector. Key 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.
- New and rapidly growing opportunities to support the continued migration and expansion of enterprise HPC workloads to cloud-based ecosystems. Hyperion expects that 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 a host of new HPC users.
- The expanding role and diversity of new big data analytics running in non-traditional HPC environments, especially for fraud and anomaly detection, affinity marketing, business intelligence, personalized medicine, 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 HPCs into the traditional modeling and simulation environment as more and more commercial and government users, including SMEs, turn to advanced computing to meet their toughest computational requirements for larger problem sizes, higher modeling fidelity, and more aggressive iteration methods, all operating under the requirement for faster turnaround time.

Hyperion Research expects x86-based processors to remain dominant through 2021, the end of our five-year forecast period. Hyperion Research's supply-side tracking showed that X86 processors comprised over 96% of the roughly 3.5 million processors used in all technical servers in 2016 but expects new growth in alternate chip options including POWER, SPARC, and ARM. Of all the sites surveyed, almost 70% indicated that they used some form of GPU or coprocessor in at least one application. The largest segment, at about one quarter of all sites surveyed, indicated that they use GPUs or coprocessors, but for less than 5% of their applications.

- Historic data from other Hyperion Research studies indicates that NVIDIA has in the past three years extended its dominance in the supply of GPU/coprocessors boards to the broader HPC user set. In 2014, the firm supplied about 70% of all products in this category and expanded that share to over 80% by 2016, racking up a 3 year CAGR of 35.7%.

The near continual roll out of new HPC-related technology, much of which shows great promise in becoming significant market drivers, will also create no small amount of confusion in the sector as vendors try to sort out the right mix of hardware, software, and applications to meet specific use cases in what is a rapidly evolving field with little clear guidance on how best to proceed. In addition, many new users will be drawn to the promise of these new HPC capabilities, and they will be looking for support in how best to incorporate that technology into their R&D or related business processes.

Both vendors and users need to balance carefully the timing of how and when to commit to these growing technology options. Moving too quickly, or in the wrong direction, runs the risk of committing to a technology choice that may not be sustainable in the long term. Likewise, moving too slowly opens the door for more nimble players to build-up what could become an insurmountable technology or market lead.

- In any case, HPC vendors and users should be working to develop closer collaborative relationships with each other to better understand the priorities of the market and to help users make the most of the technology of the day.
- HPC-based collaboration activities, such as OpenHPC, OpenPOWER, or the ARM Foundation, can server as powerful mechanisms to help ensure that vendors, users, and technology trends remain in sync.
- In addition, Hyperion Research is watching closely promising efforts in the EU, China, and Japan to HPCs that use ARM-based processors.

Key Findings

There were a total of 353,658 nodes installed in the 972 servers in the present study, and the total number of processors (physical socket parts) installed was 542,293.

- The total processor core count was 6,287,210, yielding an average of 11.6 cores per processor, 1.53 processors per node, and 17.7 cores per node.
- On average, there were 8.8 systems per site with 557 CPUs per system, 6,468 cores per system, and 363 nodes per system.
- The total peak performance installed summed across all 972 servers was 565 petaflops, yielding an average system peak performance of 0.58 petaflops, placing that average system at about 350 on the most recent Top 500 list.

Systems had an average of 11.6 cores per processor, and 17.7 cores per node.

HPC Servers

Worldwide revenue for the HPC technical server market grew 4.7% from 2015 to 2016 to a record $11.2 billion. Hyperion Research predicts CAGR growth of 5.8% to $14.8 billion in 2021.

- The Supercomputer market segment for HPC systems priced at $500,000 and up will show the highest growth rate during that period (6.9% CAGR), driven substantially by the global exascale race.
- The Divisional and Departmental segments will continue to exhibit healthy growth, and the Workgroup segment will rebound to robust growth following several years of decline.

According to this survey, nearly one-third of participants' 2017 current hardware budget will be allocated to server hardware, followed by storage hardware (15.8%) and applications software (14.5%). Overall total server software - consisting of OS (6.6%), middleware, compilers, and file systems (8.7%) and applications (14.5%) - is responsible for almost 30% of the respondent's' HPC

budgets. In addition, support for big data analysis and/or analytics now comprises 7.7% of respondent's budgets.

Perhaps more important, when asked about what percent changes respondents expect to have in their next year's budget, the single largest increase projected was in big data analysis and/or analytics (14%), followed by server hardware (7.0%) and systems software (all middleware, compilers, file sys, etc. (at 6.6%). No respondent surveyed indicated that they would maintaining or paring back any major segment of their budget in the next year

Hyperion Research predicts that in 2021, the largest workload segments will remain the same three that were biggest in 2016: government labs, university/academic and defense. We also project that these three segments will be the fastest growing during this period, with CAGRs of 7.0%, 6.7%, and 5.8%, respectively.

Processors

Hyperion Research expects x86-based processors to remain dominant through 2021, the end of our five-year forecast period. Hyperion Research's supply-side tracking showed that X86 processors comprised over 96% of the roughly 3.5 million processors used in all technical servers in 2016, but there may be some changes in the air.

- When asked about processor choices for their next server, an overwhelming percentage identified an x86 processor as their preferred choice going forward. Intel was selected by 73.8% and AMD's x86 processor was preferred by 7.8%, pointing to a continued dominance of x86 but with some weakening in Intel's hold on the sector.
- In addition, a significant minority of users seemed interested in exploring the range of non-x86 options currently available: 8.7% of survey respondents expressed an interest in Power and Open Power processors, 6% cited a preference for SPARC processors, and almost 4% stated a preference for other custom processors, such as those from Japanese and Chinese developers.

The x86 processor was the overwhelming processor choice for next server acquisition: Intel was selected by 73.8% and AMD's x86 processor by 7.8%.

Coprocessors

Of all the sites surveyed, almost 70% indicated that they used some form of GPU or coprocessor in at least one application. The largest segment, at about one quarter of all sites surveyed, indicated that they use GPUs or coprocessors, but for less than 5% of their applications. Only about one in fifteen sites indicated that they use GPUs or coprocessors in 50% or more of their applications.

- Historic data from other Hyperion Research studies indicates that NVIDIA has in the past three years extended its dominance in the supply of GPU/coprocessors boards to the broader HPC user set. In 2014, the firm supplied about 70% of all products in this category and expanded that share to over 80% by 2016, racking up a 3 year CAGR of 35.7%.
- The Intel Phi board, which in 2016 supplied about 15.2% of coprocessors, had a CAGR of 5.1% during the same time period.
- Alternate GPU/coprocessor suppliers – holding less than 4% of the 2016 coprocessor/accelerator market by offering FPGAs, DSPs, and custom chips - have maintained their absolute market size with a modest 3 year CAGR of 0.7%, but they have not kept pace with the rapidly growing revenues of their competitors.

When asked about their coprocessor/accelerator preference for their next server, 36.9% of participants identified the NVIDIA GPU as their likely preference for a next purchase. Intel's MIC/Xeon Phi follows, with a positive vote from 18.4% of respondents. FPGAs garnered almost 10% of the votes for next technical server acquisition. However, almost one-quarter (23.3%) of all sites surveyed indicated that that they would not be using any kind of GPU/accelerator/coprocessor in their next technical server.

Service and Support

Hyperion Research asked participants about where they acquire their services and support capabilities and found that sites typically use a combination of services that include acquiring services and support from traditional hardware vendors (71.6%), in-house staff (53.9%) and VAR/integrators (32.4%).

- In addition, about one third of participants plan to expand their in-house staff with additional hires, and just one-fifth plan to outsource service and support (e.g., via cloud).

Top 3 Applications

When asked about the key characteristics for the top 3 applications running at their sites, respondents' statistics support the conclusion that the most important jobs receive the most attention when it comes to tuning for optimal performance and minimal run time: the most important application typically had the shortest run time of the top three codes and also had the highest level of parallelism in the maximum core point.

For the most important applications, the average run time was almost 48 hours, used an average of almost 2,500 cores and achieved a maximum core count of 14,717.

- For the most important application, across all users responding, the average run time was almost 48 hours, used an average of almost 2,500 cores and achieved a maximum core count of14,717.
- For the second most important application across all users responding, the average run time was almost double at 84.9 hours, used an average of over 3,000 cores and achieved a maximum core count of over 10,000.
- For the third most important application across all users responding, the average run time was even higher at 91.5 hours, used an average of over about 2,500 cores and achieved a maximum core count of about 9,800.

TABLE OF CONTENTS

TABLE OF CONTENTS — Continued

LIST OF TABLES

LIST OF TABLES — Continued

LIST OF FIGURES

SITUATION OVERVIEW

In This Report

This special study is part of the fifth edition of Hyperion Research's high-performance computing (HPC) end-user-based tracking of servers, processors, coprocessors, and other server systems used in the HPC marketplace.

Introduction

For this fifth edition of the HPC end-user special study, Hyperion Research conducted in-depth interviews with one or more representatives of 111 HPC sites around the world. These sites had a total of 972 technical server systems installed, and the average in the present study was 8.8 systems per site.

There were a total 353,658 nodes installed the 972 servers in the present study, and the total number of processors (physical socket parts) installed was 542,293.

- The total processor core count was 6,287,210, yielding an average of 11.6 cores per processor, 1.53 processors per node, and 17.7 cores per node.
- On average, there were 8.8 systems per site with 557 CPUs per system, 6,468 cores per system, and 363 nodes per system.
- The total peak performance installed summed across all 972 servers was 565 petaflops, yielding an average system peak performance of 0.58 petaflops, placing that average system at about 350 on the most recent Top 500 list.

Systems had an average of 11.6 cores per processor, and 17.7 cores per node.

Overall HPC Market Trends

Hyperion Research analysts are expecting a number of factors to drive aggressive growth rates across all segments of the HPC sector going forward and that growth will likely far outstrip the growth rate expected for the general-purpose enterprise IT server sector. 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.
- New and rapidly growing opportunities to support the continued migration and expansion of enterprise HPC workloads to cloud-based ecosystems. Hyperion expects that 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 a host of new HPC users.
- The expanding role and diversity of new big data analytics running in non-traditional HPC environments, especially for fraud and anomaly detection, affinity marketing, business intelligence, personalized medicine, and cyber security sectors. Of particular import 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 HPCs 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 operating under the requirement for faster turnaround time.

After the HPC sector experienced a two-year slowdown (2013 and 2014), primarily driven by the slowdown at the very high end and the IBM-Lenovo transition, 2015 and 2016 exhibited stronger growth rates (4.9 and 4.4%, respectively). This reversal was due to a combination of returning demand for HPCs bound for the traditional modeling and simulation sectors as well new growth in areas that include deep learning, and the big data applications across a growing set of users. With these changing – and in many case newly emerging - HPC application spaces, Hyperion Research advises both HPC vendors and users to consider the following:

- The range of processors that are available for use in HPC systems is as wide as it has been in a long time. As such, vendors need to be prepared to move carefully in their support of any number of processor bases including x86, Power, SPARC, and ARM as well as different configurations for each processor flavor targeted for a different market segment. Although commitment to a single processor allows a vendor to concentrate its resources on developing leading-edge capabilities, such narrow commitment may lock a vendor out of key sectors. Likewise, a lack of focus here could dilute in-house resources that are stretched too thin to successfully compete across a broad range of processor options.
- The inclusion of GPUs, FPGAs, and other accelerators that offer the potential for high performance in a wide variety of heterogenous HPC architectures places new and sometime technically aggressive demands on an HPC system's overall ability to provide sufficient bandwidth and low latency. In many cases, users will be unsure about the programming and related application development requirements for new co-processors/accelerators, and they will require significant vendor support to help realize the full performance potential of the systems they purchase.
- New high bandwidth memory configurations, the inclusion of burst buffers, larger and more configurable caches at almost every level, and NVRAM storage class memory are options that increasingly users will be looking for in their HPC configurations. Each offers both opportunities and challenges for HPC vendors looking to develop systems uniquely configured for key use cases. However many users will be looking to the HPC supplier – or related technology firms within a vendor's overall HPC ecosystem - to help them integrate, manage, and ultimately optimize the data flow complexities that the use of such technology engenders.
- The range of options for high performance interconnects (including InfiniBand, OmniPath and 100Gb Ethernet) provides new opportunities for HPC vendors to meet the diversity of data transfer requirements across the span of HPC use cases. Complicating the issue will the growing trend in HPC toward more processors in both numbers and diversity - and cores – in a single system that can place significant, less regular, and more diverse data transfer demands on the system interconnect to support highly parallel computation.
- HPC storage is increasingly playing a critical role in determining the overall performance and utility of a complete HPC system. HPC applications are relying on an increasingly diverse set of data storage types, each of which must be handled efficiently and effectively by the data storage infrastructure. Likewise, new HPC use cases are broadening the base of requirements to which a well-balanced storage system must respond. Of particular significance is the growing demand for real-time analysis, such as in financial or fraud detection applications, new and typically unpredictable data access patterns for critical big data jobs such as graph analytics, and emerging requirements for heterogenous data such as those involved with analytics applications drawing on social media inputs.

Methodology

This server/processor report is a component of the fifth edition of Hyperion Research's demand-side tracking of the worldwide HPC market and augments Hyperion Research's established supply-side tracking. This study evaluates on-site HPC systems, requirements, purchasing criteria, and budgeting patterns of buyers. For this edition, Hyperion Research conducted in-depth interviews with purchasing decision-makers and influencers from HPC sites worldwide in the government, industry and academic sectors. We collected data on 972 HPC systems. To promote candor, Hyperion Research promised not to associate survey responses with individual or organizational identities.

As in earlier editions, Hyperion Research strove to ensure that the industrial sector was especially well represented; in this edition, this segment represents 45.9% of the surveyed sites. This strong participation enables Hyperion Research to drill down meaningfully into individual sub-segments of the industrial sector.

Hyperion Research's demand-side study provides a deeper and more direct view into the living, evolving world of HPC data centers. Where supply-side tracking sees an HPC system's growth through multiple additions as multiple systems (i.e., separate shipments), the demand-side perspective recognizes the expanded resource as a *single* system and also identifies after-market additions such as interconnects, storage, value-added software, and accelerator technology. (For more on distinctions between supply-side and demand-side perspectives, see the following section, *The Demand Side Perspective*.)

Hyperion Research makes results of its comprehensive demand-side study available through a series of six reports. Topics include:

- Processors, Clusters and System Attributes
- System Software and Middleware
- Storage and Interconnects
- Running HPC Workloads in External Clouds
- HPDA: Big Data in HPC
- Workloads by Industry/Applications
- New: HPC Budget Mix and Key Purchase Criteria

The Demand-Side Perspective: An Important Complementary View of the HPC Market

For more than 25 years, Hyperion Research team (formerly the IDC HPC team) has tracked worldwide HPC server revenues based on detailed quarterly supply-side reporting of vendors' factory shipments. Our aim, not perfectly realizable, has been to track every HPC-bound server system leaving a vendor's or subcontractor's manufacturing facility, and to characterize each system across numerous attributes related to actual ("street") pricing, component technologies, intended usage, and other characteristics of interest to Hyperion Research clients.

In 2008, we began augmenting vendor-centric, supply-side tracking and analysis with a new view of the HPC market: the buyer-centric, demand-side perspective. This complementary perspective is valuable for multiple reasons, as captured in the chart that follows:

Demand-side tracking	Supply-side tracking
"Sees" the reality of a single system growing in stages, expanded via multiple shipments and installments.	Treats each shipment leaving the factory as separate.
Captures information about "dark" clusters – sales vendors have not reported as part of HPC revenue, often because such clusters were purchased as components from computer-supply websites or outlets.	Does not capture specific information about "dark" clusters.
Follows the life cycle of a system configuration as it evolves at the user site. Identifies not only expanded processing power, but also on-site, after-purchase additions of storage, interconnects, software, alternative processors, and other components, as well as decommissioning of systems.	Focuses on initial system configuration, including internal storage bundled into a factory-shipped system.
Explicitly identifies the sites' largest and second-largest HPC systems, along with other installed HPC equipment. This helps identify emerging trends in system attributes and technology adoption.	Does not explicitly identify the largest systems.
Because reports are based on qualitative and quantitative user responses, reports capture valuable information about the intent of HPC buyers, including acquisition rationales and methods, along with budgetary and spending intent, including attributes for which buyers would pay a premium.	Does not provide qualitative information.

This 2017 fifth edition of Hyperion Research's demand-side tracking report captures information about sites representing 972 HPC systems. Demand-side tracking complements – but does not replace – our supply-side tracking, which keeps close tabs on more than 100,000 HPC systems per year. Hyperion Research conducts more deeply focused demand-side studies for private sponsors, and maintains close contact with demand-side requirements and trends through the Hyperion Research-operated HPC User Forum.

SURVEY DEMOGRAPHICS SAMPLE IN SERVERS, PROCESSORS, NODES, AND PEAK PF

For this fifth edition of the HPC end-user Special Study, Hyperion Research conducted in-depth interviews with one or more representatives of 111 HPC sites around the world. These sites had a total of 972 technical server systems installed, and the average in the present study was 8.8 systems per site.

As Table 1 shows, the total number of processors (physical socket parts) installed in the present study was 542,293. The aggregate core count was 6,287,210, yielding an average of 11.6 cores per processor. There were 353,658 nodes among all installed systems, or an average of 1.53 processors per node and 17.7 cores per node.

There are a number of interesting statistics revealed in Table 1.

The total peak performance of the 111 sites surveyed was 565 petaflops, representing 972 system. That number – 564 petaflops - is roughly the same as the combined total peak performance of both the top 50 HPCs on the Top 500 list and the bottom 450 systems on that same list. The peak performance of the average system of those surveyed for the present study was 0.58 petaflops, placing that average systems at about 350 on the most recent Top 500 list.

- The HPCs in the survey were scattered across a wide range of performance and were not unduly concentrated on the small class of more powerful, but less widespread, highest-end HPC systems.

The average number of HPCs at a site, as noted, was 8.8. This means that typically HPC sites surveyed are hosting a number of HPC that likely have different architectural configurations, hardware characteristics, software stacks, aging and reliability issues, and that come from a number of vendors.

- Many of these sites also have a wide range of computing requirements, and they likely are using different HPCs to serve different mission needs.
- As such, these sites likely are continually balancing the need to have the right HPC resource applied against a key application while keeping the number - and associated complexity of managing those HPCs – to a minimum.

Table 1

HPC End User MCS Sample Profile

Attribute	Sample Total
Number of Servers in the Sample	972
Number of Sites in the Sample	111
Average HPC Servers per Site	8.8
Total Processor Count (sockets)	542,293
Average CPUs per system	557.9
Total Processor Core Count	6,287,210
Average Cores per system	6,468.32
* Average Cores per CPU	11.6
Total System Node Count	353,658
Average Nodes per system	363.8
Total Peak PF Installed	565
Average PFlops per system	0.58
N = 111	

Source: Hyperion Research, 2017

Table 2 contains per-site information on the study's collection of 111 sites and 972 installed systems. On average, the sites had 4,886 processors (physical socket parts), 56,642 cores, and 3,186 nodes. Table 2 also shows how some key HPC architectural features have changed between 2015 (Hyperion's last edition of this survey) and 2017. Some of these changes are quite dramatic, even within a sector that is known for adopting technology at a relatively rapid pace.

- In the past two years, the average number of processor sockets per site has increased by 64% from 2,966 to 4,886. As the average number of cores per site has grown by only about 7 %, most of this growth was driven by the dramatic increase in the number of nodes per site, which more than doubled.
- The average length of time that an existing system has been installed dropped from 4.4 years to 3.62 years, a decline of over 18%, or more than nine months.

Table 2

Sample Characteristics: Changes in Per System Overall Averages

	2015 Value	2017 Value
Average number of processor sockets/site	2,966	4,886
Average number of cores/site	52,617	56,642
Average number of nodes/site	1,539	3,186
Average length of time that a system is installed	4.41	3.62
N = 111		

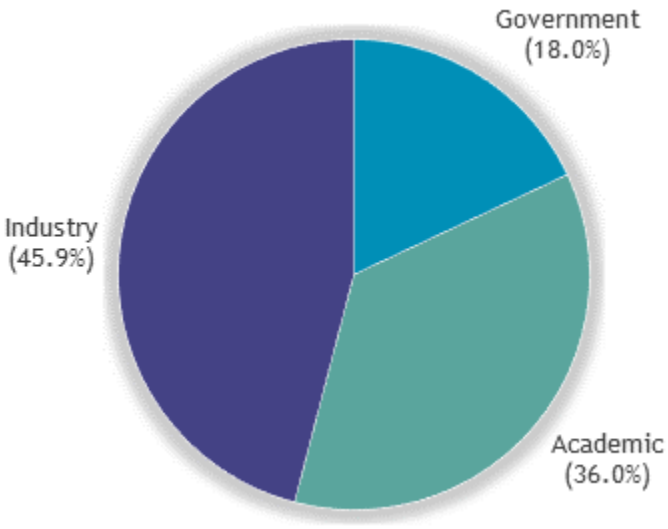
Source: Hyperion Research, 2017

Sample by Sector

Figure 1 and Table 3 show that Hyperion Research deliberately selected the sample of 111 sites, hosting 972 HPC systems, in favor of industry participants (45.9% of the sites), to allow meaningful explorations of individual industry segments (see Figure 2). Academic sites were also heavily represented in the study (36.0%). Fewer government sites were profiled (18.0%), but these sites were carefully chosen to reflect a wide range of government users within the global HPC community.

FIGURE 1

Sample by Major Sector



Source: Hyperion Research, 2017

Table 3

Sample by Major Sector

Q1.a. Is your organization in government, commerce/industry, or academia?		
Sector	Number of Responses	Percentage of Sample
Government	20	18.0%
Academic	40	36.0%
Industry	51	45.9%
Total	111	
N = 111		

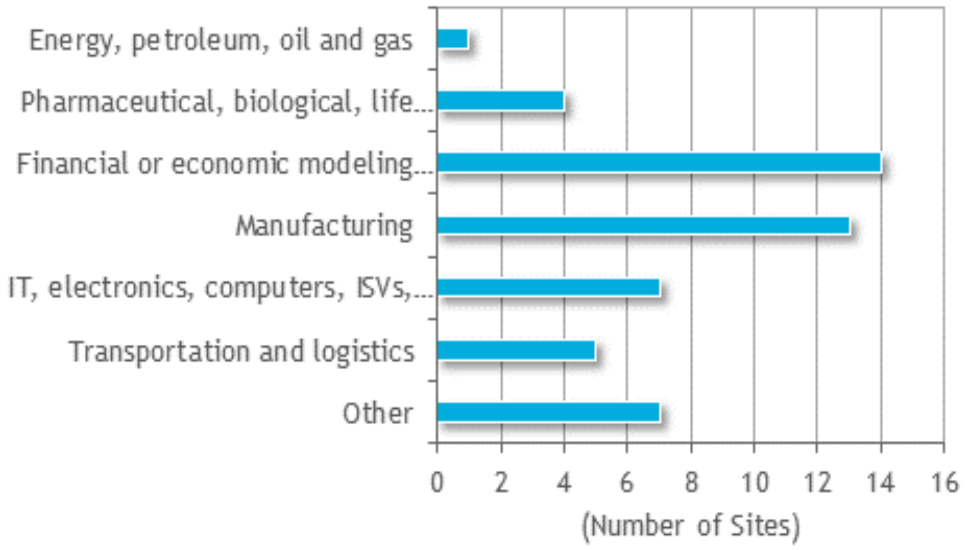
Source: Hyperion Research, 2017

Sample by Industries

As noted previously, Hyperion Research deliberately selected the sample in favor of industry. Figure 2 and Table 4 show the breakdown of industrial sites by industry segment. The financial services and manufacturing segments were most heavily represented within the population of industrial sites, accounting for about half of all industrial sites surveys. Hyperion Research was pleasantly surprised at the number of financial institutions willing to provide information for this study, given that sector's historical reluctance to participate in third-party surveys.

FIGURE 2

Industrial Mix in the Sample



Source: Hyperion Research, 2017

Table 4

Sample by Major Sector and by Industry

Q. If commercial or industry, which one best fits your organization?

Sector	Number of Responses	Percentage of Sample
Energy, petroleum, oil and gas	1	2.0%
Pharmaceutical, biological, life sciences, healthcare	4	7.8%
Financial or economic modeling and BI	14	27.5%
Manufacturing	13	25.5%
IT, electronics, computers, ISVs, ITSP, cloud provider, telco	7	13.7%
Transportation and logistics	5	9.8%
Other	7	13.7%
Total Industry	51	100.0%
N = 111		

Source: Hyperion Research, 2017

Number of Servers at the Sites

We asked respondents, "How many HPC server systems of all types (clusters, SMP systems, etc.) does your organization have?" As Table 5 shows, slightly fewer than half of the surveyed sites - and the biggest category in the group - said they have 2-5 HPC systems, and about one in five operated more than 15 systems.

Table 5

Number of Servers Per Site

Q. How many HPC server systems of all types (clusters, SMP systems, etc.) does your organization have?

Attribute	Number of Responses	Percentage of Sample
1	12	10.8%
2 to 5	46	41.4%
6 to 10	24	21.6%
11 to 15	5	4.5%
More than 15	24	21.6%
Total	111	100.0%
N = 111		

Source: Hyperion Research, 2017

Growth Driver: Pent-Up Demand

We asked respondents to estimate the amount of pent-up demand that they had for their HPC systems, defined as work the users would like to run but could not due to limitations in HPC resources. Note that the answers given here are based on the additional percent of resources needed over and above the existing workload that is run on their HPCs.

For 32.3 % of users – the largest segment chosen - the pent-up HPC demand was less than or equal to 24%. However, for more than 45 % of all respondents, there was pent-up HPC demand greater than 24% of their current existing workloads, and over 14 % of respondents had pent up demand that was equal to or greater than their existing workloads.

- Only about one in five respondents indicated that they had adequate resources to meet their existing computational requirements. Based on these numbers, the number of HPCs at many of these sites could easily be increased by 50 % or more with minimal concerns about maintaining high system utilization.

Table 6

Estimated HPC Growth Potential: Pent-up Demand at End User Sites

Estimated amount of "pent-up demand" that exists for HPC systems (work that users would like to run, but can't due to limited resources)		
Response	Number of Responses	Percentage of Sample
None, we are still in the mode of filling up the system	21	21.9%
≤ 24%	31	32.3%
25-50%	14	14.6%
51-100%	17	17.7%
101-200%	11	11.5%
201-300%	1	1.0%
> 300%	1	1.0%
N = 96		

Source: Hyperion Research, 2017

HPC SERVER MARKET SIZE, FORECASTS, AND TRENDS

The HPC Market in Revenues, Units and ASP

As seen in Table 7, worldwide revenue for the HPC technical server market grew 4.4% in 2016 to a record $11.2 billion, up from $10.7 billion in 2015. Current projections call for growth in 2017 of 7.3%, resulting in a total HPC technical server market for 2017 of over $12 billion. Hyperion Research is also anticipating healthy HPC revenue growth in the out years as well. Indeed, Hyperion Research projects that the worldwide HPC server sector will grow at a CAGR of 5.8% between 2016 and 2021, reaching total revenues of $14.8 billion by 2021.

- Additional HPC-related technologies including storage, middleware, applications, and service software nearly double the total revenue for the HPC sector to over $22 billion in 2016, with a projected CAGR of 6.2% between 2016-2021, resulting in a total HPC revenue base of over $30 billion in 2021.
- Within the HPC market, storage is the fastest growing segment. Hyperion Research estimates 2016 HPC-related storage revenue at $5.1B, growing to $6.7B in 2019 – a 9.4% CAGR, comprising almost a quarter of all HPC revenues.

Hyperion Research analysts are expecting a number of factors to drive aggressive growth rates across all segments of the HPC technical server sector going forward and that growth will likely far outstrip the growth rate expected for the general-purpose enterprise IT server sector. These drivers include:

- Requirements for new HPCs with a broad range of architectures to support development and operational capabilities in the artificial intelligence sector—especially in the area of deep learning,
- New and rapidly growing opportunities to support the continued migration and expansion of enterprise HPC workloads to cloud-based ecosystems. Hyperion expects that in many cases, HPC in the cloud operations will be used not as a replacement scheme but instead to augment critical on-prem 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 a host of new HPC users.
- The expanding role and diversity of new big data analytics running in non-traditional HPC environments, especially in the finance, personalized medicine, and cyber security sectors. Of particular import will be the ability of HPCs to enable big data analysis on a near-real time basis, an increasingly necessary requirement for many application spaces.
- The continued expansion of HPCs 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 operating under the requirement for faster turnaround time.

As Table 7 also indicates, Hyperion Research forecasts technical server unit shipments will grow robustly (4.3% CAGR) but more slowly that server revenue, during the forecast period, to exceed 103,203 units in 2021.

Average selling prices (ASPs) during the forecast period will be relatively flat (1.4% CAGR), as continued revenue growth of high-end supercomputer systems (priced at $500,000 and up) is offset by resumed growth of entry-level and midrange systems. Combined, it is clear that the bulk of increased HPC server revenues in the forecast period will come not from any increase in the unit price of a single HPC technical system, but through an increase in the number of systems sold.

Table 7

Worldwide HPC Server Market Forecast by Revenue, Units and ASP

	2015	2016	2017	2018	2019	2020	2021	2016/2021 CAGR
Revenue $Millions	$10,727	$11,200	$12,019	$12,302	$12,988	$13,997	$14,819	5.8%
Units	99,817	83,519	87,130	90,897	94,827	98,926	103,203	4.3%
ASP $000	$107	$134	$138	$135	$137	$141	$144	1.4%

Source: Hyperion Research, 2017

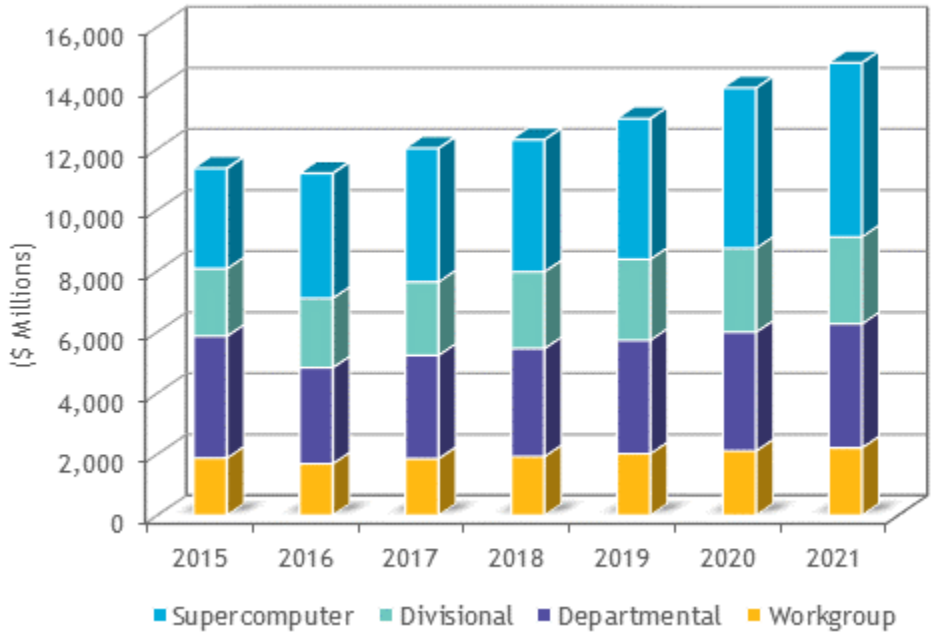
HPC Server Revenues by Competitive Segments

Through our HPC QView supply-side service, Hyperion Research closely tracks quarterly revenues for HPC server systems at all price points. Figure 3 and Table 8 display our forecasts for worldwide HPC server revenue by Hyperion Research's competitive segments, which correspond to price bands.

- The Supercomputer segment for HPC systems priced at $500,000 and up will expand fastest during the forecast period (6.9% CAGR), Hyperion Research predicts, and will reach $5.7 billion, or 38.5% of all HPC server revenues, by 2021.
- The Divisional segment for HPC systems sold for $250,000 to $449,000 will grow at a healthy 4.6% CAGR to $2.8 billion in 2021, or 19.2% of all HPC server revenues.
- The Departmental segment for systems in the $100,000 to $249,000 price range will expand at 5.2% CAGR to $4.1 billion in 2021, representing 27.4% of all HPC server revenues.
- The Workgroup segment for sub-$100,000 HPC systems will grow at a 5.5% CAGR, producing revenues of $2.2 billion in 2021, or 14.9% of HPC server revenues worldwide.

FIGURE 3

Worldwide HPC Technical Server REVENUE Forecast In $M



Source: Hyperion Research, 2017

Table 8

Worldwide HPC Server REVENUE Forecast, by Competitive Segment ($Millions)

	2015	2016	2017	2018	2019	2020	2021	2016/2021 CAGR
Supercomputer	$3,284	$4,091	$4,375	$4,324	$4,601	$5,249	$5,700	6.9%
Divisional	$2,212	$2,273	$2,404	$2,514	$2,655	$2,750	$2,848	4.6%
Departmental	$3,994	$3,147	$3,384	$3,538	$3,717	$3,885	$4,062	5.2%
Workgroup	$1,874	$1,689	$1,856	$1,925	$2,016	$2,113	$2,208	5.5%
Total	$10,727	$11,200	$12,019	$12,302	$12,988	$13,997	$14,819	5.8%

Source: Hyperion Research, 2017

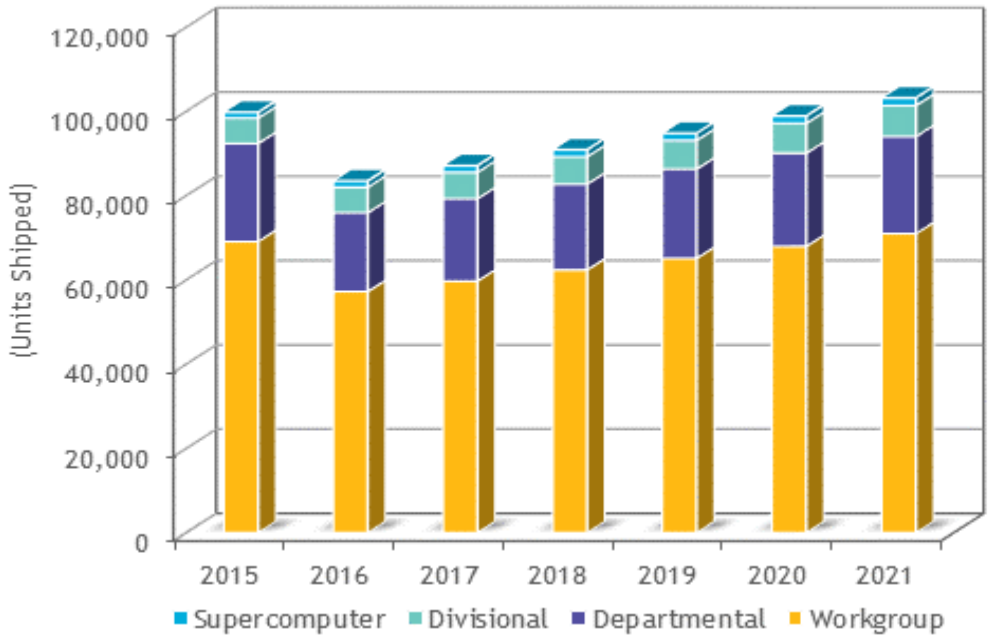
HPC System Units by Competitive Segments

As Figure 4 and Table 9 indicate, Hyperion Research forecasts unit shipments in each competitive segment will exhibit healthy growth, with the Workgroup segment taking a small lead with forecasted growth of 4.4%. The Supercomputing segment, which was essentially flat in earlier forecasts, has rebounded to 3.5% CAGR.

Where unit shipments are concerned, the technical server market is a wide-based pyramid in which the great majority of shipments occur in the sub-$100,000 Workgroup segment. For example, in 2016, 68.50% of all units shipped belonged to the Workgroup segment and only 1.8% in the Supercomputers segment. Hyperion Research predicts that for 2021 the distribution among the four server classes will remain roughly the same as it is today.

FIGURE 4

Worldwide HPC Technical Server SYSTEM UNIT Forecast



Source: Hyperion Research, 2017

Table 9

Worldwide HPC Server UNITS SHIPPED Forecast, by Competitive Segment

	2015	2016	2017	2018	2019	2020	2021	2016/2021 CAGR
Supercomputer	1,392	1,526	1,579	1,634	1,691	1,750	1,811	3.5%
Divisional	6,033	6,029	6,272	6,525	6,789	7,063	7,348	4.0%
Departmental	23,267	18,724	19,520	20,349	21,213	22,114	23,053	4.2%
Workgroup	69,125	57,240	59,759	62,389	65,134	68,000	70,991	4.4%
Total	99,817	83,519	87,130	90,897	94,827	98,926	103,203	4.3%

Source: Hyperion Research, 2017

HPC ASPs by Competitive Segments

As Table 10 indicates, the average selling price in each competitive segment will exhibit slow growth, with the Supercomputer segment taking the lead at forecasted growth of 3.3% CAGR between 2016 and 2021, itself a relatively modest price increase. The Divisional segment will show the slowest growth in price with a projected faint 0.6% CAGR between 2016 and 2021. It should be noted that the bulk of growth for ASP in the supercomputer segment will come during the 2019-2021 time frame as a few big and expensive exascale systems are shipped. Where average sales prices are concerned, the technical server market is an inverted pyramid in which the largest ASP is for the Supercomputers segment valued at $2.7 million per server; averaging about 90 times that of the average selling price of a workgroup server in the same year.

Table 10

Worldwide HPC Average Sales Price Forecast, by Competitive Segment ($K)

	2015	2016	2017	2018	2019	2020	2021	2016/2021 CAGR
Supercomputer	2,359	2,681	2,771	2,646	2,721	3,000	3,148	3.3%
Divisional	367	377	383	385	391	389	388	0.6%
Departmental	172	168	173	174	175	176	176	0.9%
Workgroup	27	30	31	31	31	31	31	1.1%
Overall Average ASP	107	134	138	135	137	141	144	1.4%

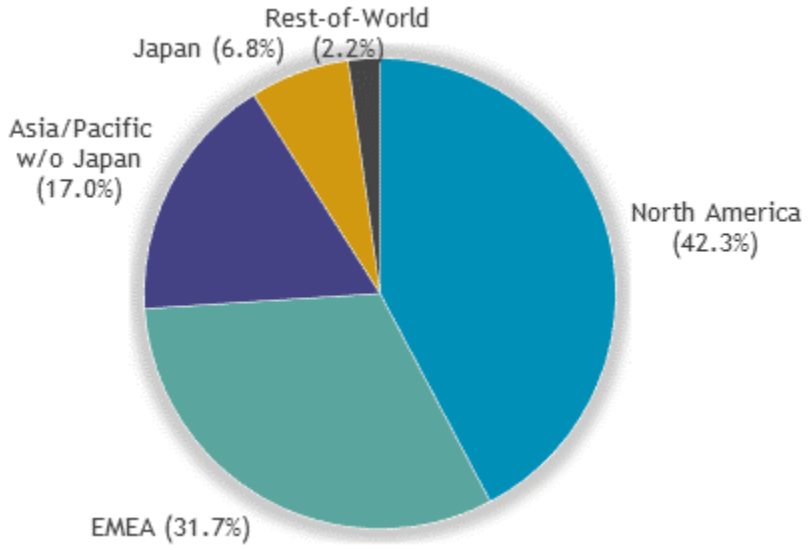
Source: Hyperion Research, 2017

HPC Servers by Region

As Figure 5 and Table 11 indicate, Hyperion Research forecasts that during the period 2016-2021, the HPC server markets in the Asia-Pacific, EMEA and North America will all experience continued healthy expansion. The U.S. market will continue to constitute about half of the global market. Growth in the Japanese HPC server market will be low, reflecting Japan's challenging economic situation, but we expect Japan's HPC server market to ramp up substantially in 2022-2023, when the RIKEN post-K supercomputer is installed.

FIGURE 5

Worldwide HPC Server Market, by Region in 2016 ($11.2 Billion)



Source: Hyperion Research, 2017

Table 11

Worldwide HPC Server Revenue Forecast, by Region ($Millions)

	2015	2016	2017	2018	2019	2020	2021	2016/2021 CAGR
North America	4,564	4,732	4,987	5,094	5,342	5,909	6,160	5.4%
EMEA	3,336	3,552	3,798	3,979	4,205	4,480	4,795	6.2%
Asia/Pacific w/o Japan	1,833	1,908	2,169	2,170	2,376	2,496	2,716	7.3%
Japan	751	764	788	795	807	817	827	1.6%
Rest-of-World	243	244	277	264	258	295	320	5.6%
Total	10,727	11,200	12,019	12,302	12,988	13,997	14,819	5.8%

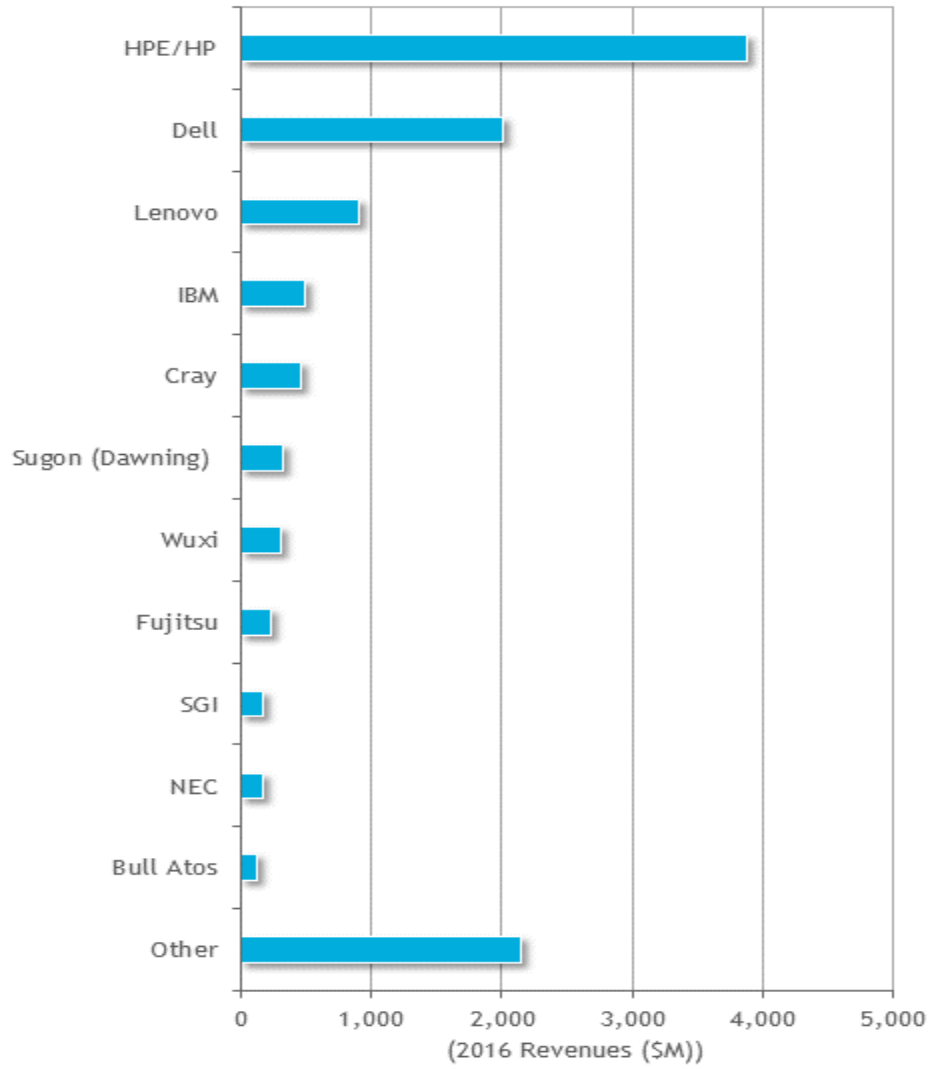
Source: Hyperion Research, 2017

HPC Servers Vendor Market Share in 2016

As Figure 6 and Table 12 indicate, HPE/HP has solidified its lead in the overall server market with 34.6% market share and sales revenues worth over $3.8 billion in 2016, followed by Dell at 18% ($2.01 billion), Lenovo at 8.1% ($908 million) and eight other vendors including Cray, IBM and Dawning, holding between 4% and 1% server market share, all with sales below $500 million. The others category, representing about one-fifth of the entire server supplier base, consists of firms that each have less than 1% share, with sales of $100 million or less.

FIGURE 6

Worldwide HPC Vendor Server Market Shares in 2016 ($M)



Source: Hyperion Research, 2017

Table 12

Worldwide HPC Vendor Server Market Shares in 2016 ($K)

	2016 Revenues (K)	2016 Market Share
HPE/HP	3,877,593	34.6%
Dell	2,013,824	18.0%
Lenovo	908,532	8.1%
IBM	491,959	4.4%
Cray	460,830	4.1%
Sugon (Dawning)	315,048	2.8%
Wuxi	300,000	2.7%
Fujitsu	226,396	2.0%
SGI	168,650	1.5%
NEC	165,758	1.5%
Bull Atos	117,717	1.1%
Other	2,153,466	19.2%
Total	11,199,773	100.0%

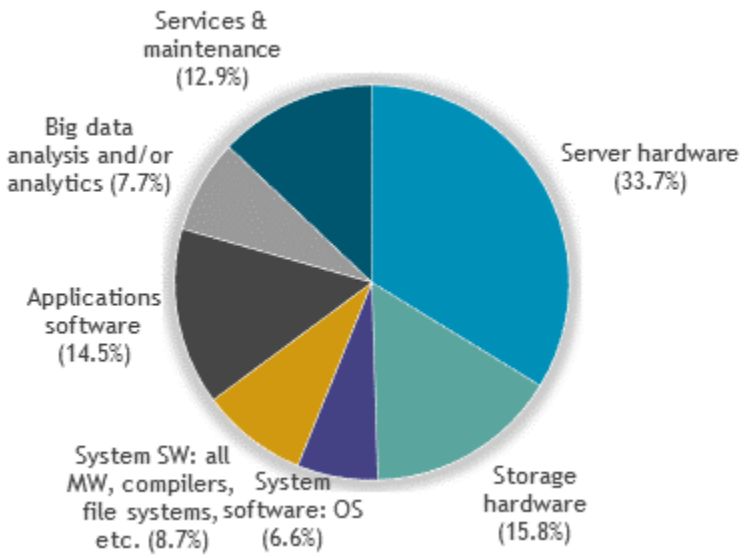
Source: Hyperion Research, 2017

HPC Budget Mix in 2017

As Figure 7 and Table 13 reveal, nearly one-third of participants' 2017 current hardware budget is allocated to server hardware, followed by storage hardware (15.8%) and applications software (14.5%). Total server software - consisting of OS (6.6%), middleware, compilers, and file systems (8.7%) and applications (14.5%) - is responsible for almost 30% of the overall respondent's' budgets. In addition, support for big data analysis and/or analytics now comprises 7.7% of respondents' budgets.

FIGURE 7

HPC Budget Mix in 2017



Source: Hyperion Research, 2017

Table 13

HPC Budget Mix in 2017

	Percent of HPC Budget
Server hardware	33.7%
Storage hardware	15.8%
System software: OS	6.6%
System SW: all MW, compilers, file systems, etc.	8.7%
Applications software	14.5%
Big data analysis and/or analytics	7.7%
Services & maintenance	12.9%
N=102	

Source: Hyperion Research, 2017

As indicated in Table 14, there are clear and distinct priorities for budget allocation across the three main categories of government, academic, and commercial sites. Academic sites devote the highest portion of their overall budget to server hardware (47.2%) at more than twice the rate spent at commercial sites (23.7%). In contrast, commercial sites commit 18.9% of their budgets to applications software, more than twice that seen in the academic sector (9.3%) and more than 75% larger than government's expenditures (10.8%). Indeed, for all of the software sectors studied, academia typically spends much less proportionally (18.3% for all software) than does either the commercial (38.5%) or government (28.5%) counterparts.

- This is likely due to the reliance of open source and other free software within the academic community, while the commercial sector, and to a lesser extent the government sector, rely more heavily on the purchase of third-party commercials software.
- As such, the academic sector can devote a larger percentage of its overall budget to HPC hardware, while the commercial and government sectors must have a more careful balance of software and hardware allocations to meet their computational requirements.

Table 14

HPC Budget Mix in 2017 (By Sector)

	Government Sites	Academic Sites	Commercial Sites	Overall
Server hardware	36.1%	47.1%	23.7%	33.7%
Storage hardware	16.6%	16.5%	15.0%	15.8%
System software: OS	7.2%	3.7%	8.7%	6.6%
System SW: all MW, compilers, file systems, etc.	10.5%	5.3%	10.9%	8.7%
Applications software	10.8%	9.3%	18.9%	14.5%
Big data analysis and/or analytics	8.1%	5.0%	9.3%	7.7%
Services & maintenance	10.7%	13.0%	13.3%	12.9%
N=102				

Source: Hyperion Research, 2017

As can be seen in Table 15, when asked about the percentage changes respondents expect to have in their next year's budget, the single largest increase projected was in big data analysis and/or analytics (14%), followed by server hardware (7.0%) and systems software all middleware, compilers, file system, etc. (6.6%). No respondent surveyed indicated that they would maintaining or paring back any major segment of their budget in the next year.

Table 15

HPC Budget Growth Rates

	Growth Percentage
Server hardware - Next Year % Up or Flat or % Down	7.0%
Storage hardware - Next Year % Up or Flat or % Down	5.3%
System software: OS - Next Year % Up or Flat or % Down	3.5%
System SW: all MW, compilers, file sys, etc. - Next Year % Up or Flat or % Down	6.6%
Applications software - Next Year % Up or Flat or % Down	4.6%
Big data analysis and/or analytics - Next Year % Up or Flat or % Down	14.0%
Services & maintenance - Next Year % Up or Flat or % Down	3.0%
N=102	

Source: Hyperion Research, 2017

As can be seen in Table 16, when asked about what percent changes respondents expect to have in their next year's budget, projections by sector varied to a great degree. Overall, commercial sites anticipated the highest growth rates for every category, generally followed by academic sites, and then government sites. For example, commercial sites looked for server hardware budgets to increase by 9.8% while the academic and government respondents foresaw increases of only 4.2% and 1.2% respectively.

- Despite the variance across the three sectors, no respondent surveyed indicated that they would maintaining or paring back any major segment of their budget in the next year, although the average government site expected some slow growth areas such as in server hardware (1.2%)and systems software: OS (1.2%).

Table 16

HPC Budget Growth Rates (By Sector)

	Government Sites	Academic Sites	Commercial Sites	Overall
Server hardware - Next Year % Up or Flat or % Down	1.2%	4.2%	9.8%	7.0%
Storage hardware - Next Year % Up or Flat or % Down	2.2%	2.2%	8.0%	5.3%
System software: OS - Next Year % Up or Flat or % Down	1.2%	5.5%	2.3%	3.5%
System SW: all MW, compilers, file sys, etc. - Next Year % Up or Flat or % Down	5.2%	2.9%	9.7%	6.6%
Applications software - Next Year % Up or Flat or % Down	3.2%	0.6%	7.6%	4.6%
Big data analysis and/or analytics - Next Year % Up or Flat or % Down	5.2%	10.7%	17.9%	14.0%
Services & maintenance - Next Year % Up or Flat or % Down	1.2%	4.7%	2.1%	3.0%
N=102				

Source: Hyperion Research, 2017

The Broader HPC Market

Figure 8 and Table 17 show Hyperion Research's five-year revenue forecast for the broader HPC market that includes not just technical servers, but also storage, middleware and operating systems (OS), applications software, and technical support services.

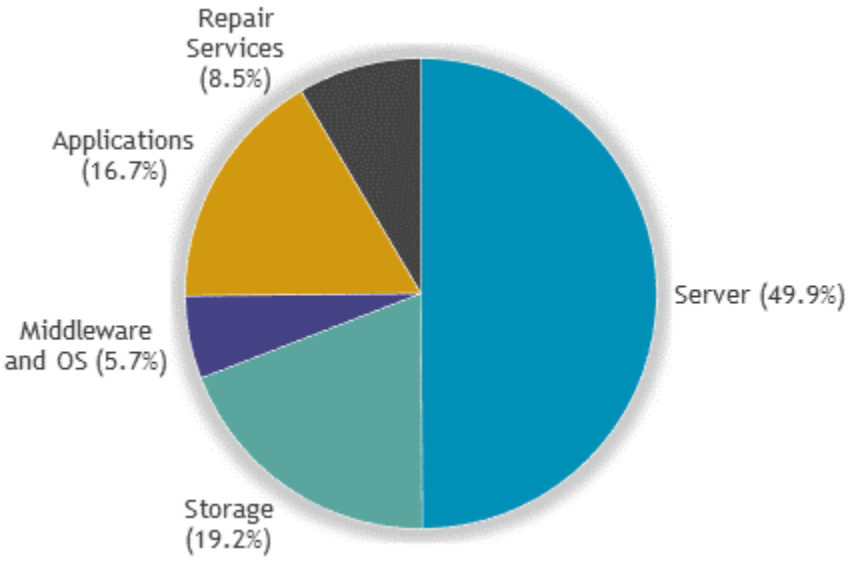
To put this in historical perspective, worldwide revenues for the broader HPC market have grown from about $2 billion in 1990 to $22.4 billion in 2016. This 10-fold leap – owing much to the introduction of standards-based clusters in 2001-2002 – means HPC has been one of the fastest-growing IT markets in recent decades.

As Table 8 illustrates, in 2021 Hyperion Research expects broader HPC revenues to reach $30.3 billion, with nearly half of that total (48.8%) spent on computer servers, followed by storage (20.8%) and applications software (16.8%). Repair services (7.6%) and middleware/OS (5.9%) represent less significant spending categories. As it has been in recent years, storage will have the highest CAGR (7.8%) during the forecast period, owing in large part to the rise of data volumes for both simulation and analytics HPC workloads.

- Spending for each category may vary greatly by scientific/engineering domain, however. For example, recent Hyperion research revealed that manufacturing firms may spend 4-5 times more per year on third-party applications software licenses than on computer servers.

FIGURE 8

Worldwide HPC Broader Market in 2016 ($22.4 Billion)



Source: Hyperion Research, 2017

Table 17

The Broader HPC Market

Worldwide HPC Compute, Storage, Middleware, Application and Service Revenues ($Billions)								
	2015	2016	2017	2018	2019	2020	2021	2016/2021 CAGR
Server	$10.7	1120%	$12.0	$12.3	$13.0	$14.0	$14.8	5.8%
Storage	$4.3	$4.3	$4.5	$5.0	$5.3	$5.8	$6.3	7.8%
Middleware and OS	$1.3	$1.3	$1.3	$1.4	$1.5	$1.7	$1.8	6.9%
Applications	$3.7	$3.7	$3.9	$4.2	$4.4	$4.8	$5.1	6.3%
Repair Services	$1.9	$1.9	$2.0	$2.0	$2.1	$2.2	$2.3	3.9%
Total	$21.9	$22.4	$23.7	$24.9	$26.3	$28.5	$30.3	6.2%

Source: Hyperion Research, 2017

Historic Processor Types Installed at HPC Sites

As shown in Table 18, the x86-64 processor base has solidified its overwhelming lead as the processor of choice for the HPC sites surveyed. Within the last two years. the number of RISC processors identified at respondent sites has declined by over 40%, while the reported use of vector processors has essentially disappeared. It is worth noting the appearance of a new albeit small presence of custom processors in the survey, reflecting some of the new custom chips being used in experimental HPCs around the world, such as those in leading-edge Chinese-built systems.

Table 18

HPC Base Processor Historic Installations

CPU Type	2014	2015	2016
RISC	139,028	89,431	86,131
Vector	7,220	7,680	
x86-64	3,025,692	3,219,689	3,371,227
Custom			40,960
Total Sum of WW Processor Package Volume	3,171,940	3,316,800	3,498,318

Source: Hyperion Research, 2017

As shown in Table 19, the pattern of growing dominance of x86-64 is evident as well when viewed from the perspective of core count as well as processor count. In addition, for x86-64 processors, there is a general upward trend in the number of cores shipped per processor. For example, in 2016 the average cores per processor for x86-64 was 8.9, two years later that number had risen to over 10.7 cores per processor, a 20% increase. However, for RISC chips, in the same two year period, cores per processor count had dropped slightly from 7.75 to 7.73.

Table 19

HPC Base Processor Historic Installations: Number of Cores

CPU Type	2014	2015	2016
RISC	1,077,504	678,568	666,006
Vector	64,980	61,440	
x86-64	27,015,414	32,675,020	36,148,315
Custom			10,649,600
Total Sum of Core Volume	28,157,898	33,415,028	47,463,921

Source: Hyperion Research, 2017

Historic Processor Installed at HPC Sites, By Competitive Segments

As shown in Table 20, for 2016, the Supercomputer segment used the most processors (34%) of all the roughly 3.5 million processors deployed in the technical server sector. Departmental systems were a close second with 29%, followed by Divisional and Workgroup with 20% and 16%, respectively. In addition, between 2014 and 2016, the number of processors used in the Divisional segment rose by almost 47%, while the Supercomputer and Workgroup processor use rate rose by 39 and 18% respectively. The Departmental sector, however, actually saw its overall use of processor decline by almost 25% during the same time period.

Table 20

HPC Historic Installations: Number of Base Processors by Competitive Segment

	2014	2015	2016
Supercomputer	852,397	945,607	1,190,586
Divisional	488,992	611,064	717,358
Departmental	1,357,651	1,245,607	1,028,889
Workgroup	472,900	514,522	561,484
Total Nodes Installed	3,171,940	3,316,800	3,498,318

Source: Hyperion Research, 2017

Historic Number of NODES Installed at HPC Sites, By Competitive Segments

As shown in Table 21, for 2016, the Supercomputer segment used the most nodes (34%) of all the roughly 1.8 million nodes used across the technical server sector. Departmental systems were a close second with 29.9%, followed by Divisional and Workgroup with 21% and 15% respectively. In addition, between 2014 and 2016, the number of nodes used in the Divisional segment rose by almost 66%, while the Supercomputer and Workgroup processor use rate both rose by about 38% respectively. The Departmental sector, however, actually saw its overall use of processors decline by almost 12% during the same time period.

Table 21

HPC Historic Installations: Number of Nodes by Competitive Segment

	2014	2015	2016
Supercomputer	455,177	478,453	629,824
Divisional	229,639	307,038	382,712
Departmental	598,464	622,825	528,070
Workgroup	204,561	249,088	283,017
Total Nodes Installed	1,487,840	1,657,404	1,823,623

Source: Hyperion Research, 2017

Historic Number of CPUs/Node Installed at HPC Sites, By Competitive Segments

As shown in Table 22, for 2016, the average number of CPUs/node across all segments of the technical server sector was 1.92 with the highest ratio for Workgroup (1.98), followed by Departmental (1.95), Supercomputer (1.89) and Divisional (1.87). Between 2014 and 2016, the average number of CPUs/node has declined by over 11% from 2.13 to 1.92 with the largest declines being the Departmental and Workgroup segments. In contrast, the CPUs/node count on in the Supercomputer sector did increase, but only by about 1% during the same time period.

Table 22

HPC Historic Installations: Number of CPUs/Node by Competitive Segment

	2014	2015	2016
Supercomputer	1.87	1.98	1.89
Divisional	2.13	1.99	1.87
Departmental	2.27	2.00	1.95
Workgroup	2.31	2.07	1.98
Total Nodes Installed	2.13	2.00	1.92

Source: Hyperion Research, 2017

Historic Number of Cores Installed at HPC Sites, By Competitive Segments

As shown in Table 23, for 2016, the Supercomputer segment used almost half (48%) of the roughly 47.4 million cores used across the technical server sector. Departmental systems were a close second with 22.6% of the cores used followed by Divisional and Workgroup with 15.7% and 12.6% respectively. In addition, between 2014 and 2016, the number of nodes used in the Supercomputer segment rose by more than a factor of three, while the Divisional and Workgroup processor use rate rose by about 68% and 41% respectively. The Departmental sector, however, actually saw its overall use of cores decline by almost 9% during the same time period.

Table 23

HPC Historic Installations: Number of Cores by Competitive Segment

	2014	2015	2016
Supercomputer	7,720,780	9,537,834	23,244,834
Divisional	4,442,711	6,088,958	7,491,896
Departmental	11,787,168	12,458,297	10,758,641
Workgroup	4,207,239	5,329,939	5,968,550
Total Nodes Installed	28,157,898	33,415,028	47,463,921

Source: Hyperion Research, 2017

Historic Number of Cores/CPU Installed at HPC Sites, By Competitive Segments

As shown in Table 24, for 2016, the average number of cores/CPU across all segments of the technical server sector was 13.6 with the highest ratio for Supercomputers (19.5), followed by Workgroup (10.6), Departmental (10.5), and Divisional (10.4). Between 2014 and 2016, the average number of cores/CPU had increased by over 52% from 8.9 to 13.6 with the largest increase seen in the Supercomputer segment (up by a factor of 2.1), followed by Departmental (up 20.4%), Workgroup up 19.4%) and Divisional (up 14.95%) segments.

Table 24

HPC Historic Installations: Average Number of Cores/CPU by Competitive Segment

	2014	2015	2016
Supercomputer	9.1	10.1	19.5
Divisional	9.1	10.0	10.4
Departmental	8.7	10.0	10.5
Workgroup	8.9	10.4	10.6
Total Nodes Installed	8.9	10.1	13.6

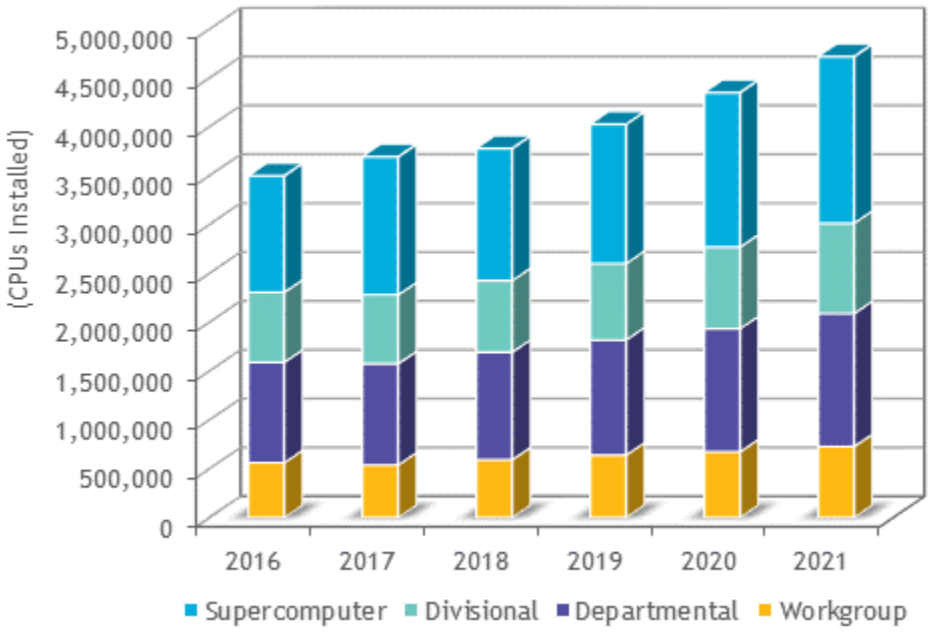
Source: Hyperion Research, 2017

HPC Processor Forecast, By Competitive Segments

As shown in Figure 9 and Table 25, the number of processors expected to be used within the overall technical server market will grow at a CAGR of 6.2% between 2016 and 2021, from the approximately 3.5 million processors base in 2016 to 4.7 million in 2021. The largest segment for growth will be in Supercomputers with a 7.5 % CAGR, and by 2021, Hyperion Research expects that this segment will use about 36.1% of all processors, followed by Departmental (28%), Divisional (19.5%) and Workgroup (15.4%).

FIGURE 9

Forecasted HPC Base Processors by Competitive Segment



Source: Hyperion Research, 2017

Table 25

Forecasted HPC Base Processors by Competitive Segment

	2016	2017	2018	2019	2020	2021	CAGR 16-21
Supercomputer	1,190,586	1,413,134	1,352,010	1,424,196	1,581,433	1,705,452	7.5%
Divisional	717,358	705,404	732,716	783,561	835,806	924,101	5.2%
Departmental	1,028,889	1,035,741	1,101,038	1,178,212	1,259,315	1,360,806	5.8%
Workgroup	561,484	539,640	592,485	640,336	674,316	728,225	5.3%
Total	3,498,318	3,693,919	3,778,248	4,026,306	4,350,870	4,718,583	6.2%

Source: Hyperion Research, 2017

Processor Preferences for Next System Purchase

As shown in Table 26, nearly three-quarters of respondents identified Intel x86 as their processor preference for their next technical server purchase. Another 7.8% indicated a preference for the AMD x86, giving the x86 option a combined 81.6% share of next processor preferences. POWER and Open Power options were preferred by 8.7% of respondents, and SPARC was selected by 5.8%. The other category – which likely includes both custom processors and ARM processor variants - were the preferred processor by 3.9% of those surveyed.

Table 26

Processor Preferences for Their NEXT Technical Server Purchase

	Number of Responses	Percentage of Sample
Intel (x86)	76	73.8%
AMD (x86)	8	7.8%
Power & Open Power	9	8.7%
SPARC (all SPARC-based processors)	6	5.8%
Other	4	3.9%
N=103		

Source: Hyperion Research, 2017

As shown in Table 27, although there remains a strong preference for x86 processors (primarily offerings from Intel) there are some small distinctions among processor preferences across the major user sectors.

- Academic sites have the strongest user preference for Intel chips (77.5%), followed by the commercial users (72.5%) and government (66.7%).
- In contrast, government sites are most open to considering POWER and Open Power variants going forward (16.7%), more so than academic sites (10.0%) and commercial sites (8.7%). In general, government sites are more committed (25%) to exploring non-x86 options than any other sector.
- It is worth noting that the commercial sector is most open to considering SPARC processors for their next system (9.8%), which is likely due to their demand for large database systems that often use SPARC processors. Although Oracle is winding down its involvement in SPARC and Solaris, Fujitsu appears to have a long-term commitment to offering SPARC-based servers.

Table 27

Processor Preferences for Their NEXT Technical Server Purchase (By Sector)

	Government Sites	Academic Sites	Commercial Sites	Overall
Intel (x86)	66.7%	77.5%	72.5%	73.8%
AMD (x86)	8.3%	7.5%	7.8%	7.8%
Power & Open Power	16.7%	10.0%	5.9%	8.7%
SPARC (all SPARC-based processors)	0.0%	2.5%	9.8%	5.8%
Other	8.3%	2.5%	3.9%	3.9%
N=103				

Source: Hyperion Research, 2017

Vendor Market Shares, by Processors Installed in 2016

As shown in Table 28, Hyperion Research's supply-side research confirmed that HPE/HP leads all other vendors in base processors installed in 2016, with 38.0% market share of the roughly 3.5 million processors used in the technical server market, followed by Dell (21.4%), Lenovo (10.6%), and Cray (4.1%) The installation at the National Supercomputing Center in Wuxi China of the Sunway TaihuLight, was large enough to account for 1.1% of all processors installed in 2016.

Table 28

Worldwide HPC Vendor PROCESSOR Market Shares in 2016

	Base Processors Installed	2016 Market Share
HPE/HP	1,328,294	38.0%
Dell	749,643	21.4%
Lenovo	370,830	10.6%
Cray	141,748	4.1%
Sugon (Dawning)	119,249	3.4%
IBM	71,006	2.0%
Fujitsu	43,995	1.3%
Wuxi	40,960	1.2%
SGI	38,091	1.1%
Bull Atos	27,742	0.8%
NEC	23,040	0.7%
Other	543,720	15.5%
Total	3,498,318	100.0%

Source: Hyperion Research, 2017

Mix of Processor Counts at HPC Sites

As shown in Table 29, for all of the sites surveyed, there was a relatively broad mix in the number of processors at any given site. The largest category was for the 129-512 processors segment with 20.4%, followed by the 1,000-4,999 segment with 16.5% and the 5,000-9,999 segment with 12.6%. Four different sites indicated that they used over 100,000 processors on site.

Table 29

Mix of Number of Processors, of All Types at the Sites Surveyed

	Number of Responses	Percentage of Sample
≤ 16	7	6.8%
17 to 32	5	4.9%
33 to 64	7	6.8%
65 to 128	9	8.7%
129 to 512	21	20.4%
513 to 999	10	9.7%
1,000 to 4,999	17	16.5%
5,000 to 9,999	13	12.6%
10,000 to 24,999	6	5.8%
25,000 to 100,000	4	3.9%
> 100,000	4	3.9%
N=103		

Source: Hyperion Research, 2017

Table 30 breaks out the number of HPC/technical computing server processors of all types, across all servers by three sectors: government, academia and industry.

- The segment 1,000 to 4,999 processors was the largest within the government and academic sector, found at 27.3% and 15.5% of all sites surveyed, respectively. Commercial sites identified this segment only 13.7% of the time.
- Among academic sites there was a wide range of segment preference: six of the 11 potential segments were identified at academic sites between 10-15% of the time and no site preferred any single segment by more than 15.5%.
- The processor segment most identified by industry was the129-512 segment (20.4%), followed by the 1,000-4,999 segment (16.5%). Industry was the only sector that identified sites with more than 100,00 processors, consisting of 7.8% of all industry sites.

Table 30

Mix of Number of Processors, of All Types at the Sites Surveyed (By Sector)

	Government Sites	Academic Sites	Commercial Sites	Overall
≤ 16	0.0%	5.0%	9.8%	6.8%
17 to 32	9.1%	0.0%	7.8%	4.9%
33 to 64	18.2%	7.5%	3.9%	6.8%
65 to 128	0.0%	12.5%	7.8%	8.7%
129 to 512	18.2%	15.0%	25.5%	20.4%
513 to 999	9.1%	10.0%	9.8%	9.7%
1,000 to 4,999	27.3%	17.5%	13.7%	16.5%
5,000 to 9,999	18.2%	15.0%	9.8%	12.6%
10,000 to 24,999	0.0%	12.5%	2.0%	5.8%
25,000 to 100,000	9.1%	5.0%	2.0%	3.9%
> 100,000	0.0%	0.0%	7.8%	3.9%
N=103				

Source: Hyperion Research, 2017

Mix of Node Counts at HPC Sites

As shown in Table 31, when asked about the number of nodes at their sites, roughly one in five respondents confirmed their organizations use between 129 and 512 HPC/technical computing server nodes of all types, across all servers. About 45% of all respondents reported having node courts between 129 and 4,999, and 31% between less than 16 and 64 nodes. 87% of all respondents reported having less than 4,999 nodes at their sites and 68% had less than 999.

Table 31

Mix of Number of NODES at the Sites Surveyed

	Number of Responses	Percentage of Sample
≤ 16	10	9.7%
17 to 32	10	9.7%
33 to 64	14	13.6%
65 to 128	9	8.7%
129 to 512	18	17.5%
513 to 999	14	13.6%
1,000 to 4,999	15	14.6%
5,000 to 9,999	4	3.9%
10,000 to 24,999	4	3.9%
25,000 to 100,000	2	1.9%
> 100,000	3	2.9%
N=103		

Source: Hyperion Research, 2017

Table 32 breaks down the data from Table 31 by sector.

- At government sites, two segments account for half of all possible configurations, 33 to 64 (25%) and 513 to 999 (25%). There were no government respondents that had more than 9,999 nodes at any site.
- The single largest segment identified by academic sites was the segment 1,000 to 4,999 (25%), and over 57% of the sites were clustered between the three segments that spanned 129 and 4,999 nodes. Academic respondents reported no sites with node counts that exceeded 24,999.
- The single largest segment identified by the commercial sector was the 129-512 option, chosen by almost one in five industry respondents. Indeed, over 70% of commercial respondents reported having fewer than 512 nodes at their sites, but industry also reported having almost 10% of their node configurations exceeding 25,000.

Table 32

Mix of Number of NODES at the Sites Surveyed (By Sector)

	Government Sites	Academic Sites	Commercial Sites	Overall
≤ 16	8.3%	5.0%	13.7%	9.7%
17 to 32	8.3%	5.0%	13.7%	9.7%
33 to 64	25.0%	12.5%	11.8%	13.6%
65 to 128	0.0%	7.5%	11.8%	8.7%
129 to 512	8.3%	17.5%	19.6%	17.5%
513 to 999	25.0%	15.0%	9.8%	13.6%
1,000 to 4,999	16.7%	25.0%	5.9%	14.6%
5,000 to 9,999	8.3%	5.0%	2.0%	3.9%
10,000 to 24,999	0.0%	7.5%	2.0%	3.9%
25,000 to 100,000	0.0%	0.0%	3.9%	1.9%
> 100,000	0.0%	0.0%	5.9%	2.9%
N=103				

Source: Hyperion Research, 2017

Use of GPUs or Coprocessors on HPC Applications in 2017

Table 33 shows the percentage of HPC applications that use GPUs or coprocessors at all sites surveyed. For all sites surveyed, almost 70% indicated that they used some form of GPU or coprocessor in at least one application. The largest segment, at about one quarter of all sites surveyed, indicated that they do use GPUs or coprocessors, but for less than 5% of their applications. About one in fifteen sites indicated that they use GPUs or coprocessors in 50% or more of their applications.

Table 33

Percentage of HPC Applications that Use GPUs or Coprocessors

Usage Level	Percentage
0	30.9%
≤ 5%	23.6%
6-10%	11.8%
11-25%	19.1%
26-50%	8.2%
> 50%	6.4%
N=110	

Source: Hyperion Research, 2017

Use of GPUs or Coprocessors on HPC Applications in 2017 By Sector

Table 34 breaks down the data from Table 33 by sector.

- Government sites are the most aggressive users of GPUs and coprocessors as a percentage of their overall HPC application base, albeit only by a small margin with 74% of government sites surveyed using at GPUs on at least one HPC application, followed by the academic sector (70%) and the commercial sector (69.1%). In addition, the government sector was the most aggressive user of GPUs or coprocessors, with almost half (47.50%) indicating that they used GPUs or coprocessors in more than 11% of their HPC applications.
- For respondents from the academic sector that indicated they had HPC applications that used GPUs or accelerators, the most chosen selection (30%) was for GPUs or coprocessor usage in less than 5% of their HPC application base. However, almost one in eight indicated that GPUs or coprocessors were used in 25% or more of their HPC applications.

- Although the industrial sector was the least likely of all sectors surveyed to use GPUs or coprocessors in their HPC application set (66%), more than half (52.9%) indicated use of these devices ranging from 6 to 25% of their HPC application base.

Table 34

Percentage of the Applications that use Accelerators or Coprocessor or Other Specialized Processors (i.e. FPGAs) in 2017 (By Sector)

	Government Sites	Academic Sites	Industrial Sites	Overall
0	26.3%	30.0%	33.3%	30.9%
≤ 5%	21.1%	30.0%	19.6%	23.6%
6-10%	5.3%	7.5%	17.6%	11.8%
11-25%	31.6%	17.5%	15.7%	19.1%
26-50%	10.5%	7.5%	7.8%	8.2%
> 50%	5.3%	7.5%	5.9%	6.4%
N=110				

Source: Hyperion Research, 2017

Historic GPU Board Shipments in HPC

As Table 35 indicates, NVIDIA has in the past three years extended its dominance in the supply of GPU/coprocessors boards to the overall HPC user set. In 2014, the firm supplied about 70% of all products in this category and expanded that share to over 80% by 2016, racking up a 3 year CAGR of 35.7% - 7 times the growth rate of the Intel Phi board during the same time period. Alternate GPU/coprocessor suppliers have maintained their absolute market size to some extend with a modest 3 year CAGR of 0.7%, but they have not kept pace with the rapidly growing revenue of their competitors.

Table 35

Historic GPU and Coprocessor Board Shipments in HPC

	2014	2015	2016	3yr CAGR
Total Sum of Nvidia Board Count	46,347	70,723	85,359	35.7%
Total Sum of Intel Phi Board Count	14,453	14,268	15,957	5.1%
Total Sum of Other Acce/CoProc Board Count	4,292	3,132	4,354	0.7%

Source: Hyperion Research, 2017

Table 36 is a break out of Table 35 by competitive segment.

- For NVIDIA, its largest market segment is in supercomputers, where in 2016 the firm shipped over 52% of all its boards to the technical server market, building off a three year CAGR of 30%. Despite that, the firm saw the most impressive growth rates between 2014 and 2016 in the Divisional and Departmental segments with 3 year CAGRs of 60.4 and 45.4%, respectively. In contrast, the Workgroup segment was the slowest growth area for NVIDIA boards with a 3 year CAGR of only 6%.
- For the Intel Phi boards, the greatest 3 year CAGR was in the Divisional segment at 15.2%. Conversely, Intel Phi boards saw only a 1% 3 year CAGR in the Departmental segment and a 3 year CAGR decline of 3.1% in the Workgroup segment.
- For the others categories, almost all of the growth (a somewhat impressive 34.8% 3 year CAGR) was limited to the Departmental segment with all other segments showing a negative CAGR.

Table 36

Historic GPU and Coprocessor Board Shipments in HPC, By Competitive Segment

	Competitive Segment	2014	2015	2016	3yr CAGR
Sum of Nvidia Board Count	Supercomputer	26,656	36,236	45,061	30.0%
	Divisional	7,893	12,928	20,303	60.4%
	Departmental	6,796	14,440	14,374	45.4%
	Workgroup	5,002	7,119	5,621	6.0%
Sum of Intel Phi Board Count	Supercomputer	6,002	5,777	6,601	4.9%
	Divisional	3,055	3,356	4,054	15.2%
	Departmental	2,915	2,957	2,971	1.0%
	Workgroup	2,481	2,178	2,331	-3.1%
Sum of Other Acce/CoProc Board Count	Supercomputer	1,859	1,615	1,838	-0.6%
	Divisional	980	753	838	-7.5%
	Departmental	608	480	1,104	34.8%
	Workgroup	845	284	574	-17.6%
Total Sum of Nvidia Board Count		46,347	70,723	85,359	35.7%
Total Sum of Intel Phi Board Count		14,453	14,268	15,957	5.1%
Total Sum of Other Acce/CoProc Board Count		4,292	3,132	4,354	0.7%

Source: Hyperion Research, 2017

GPU/Accelerator/Coprocessor Preference for their Next Technical Server

As shown in Table 37, in the accelerator/co-processor space, NVIDIA takes the lead, with 36.9% of participants identifying the NVIDIA GPU as their likely preference for a next purchase. Intel's MIC/Xeon Phi follows, winning a positive vote from 18.4% of respondents. FPGAs garnered almost 10% of the choice for next technical server acquisition. However, almost one-quarter (23.3%) of all sites surveyed indicated that that they would not be using any kind of GPU/accelerator/coprocessor in their next technical server.

Table 37

GPU/Accelerator/Coprocessor Preference for their Next Technical Server

	Responses	Percentage of Sites
None	24	23.3%
NVIDIA GPU	38	36.9%
Intel Phi	19	18.4%
ATI (ATI/AMD Fusion Processors)	5	4.9%
FPGA	10	9.7%
TI/DSP	3	2.9%
Other	4	3.9%
N=103		

Source: Hyperion Research, 2017

Table 38 is a break out of Table 37 by competitive sector and reveals some variations in accelerator/co-processor preference by sector:

- Within the government sector, Intel's MIC/Xeon Phi is the preferred choice, with 25.0% of respondents. Only 16.7% of respondents picked the NVIDIA GPU. Also, of all segments surveyed, government sites were most likely (33.3%) to have no plans to any GPU/accelerator/ coprocessor in their next technical server. However, Government buyers were also the most open-minded to alternate devices citing equal preference (8.3%) for FPGAs, TI/DSPs, and other custom devices for their next technical server acquisition.
- Among academic sites, NVIDIA's GPU was the overwhelming preference for next technical server procurements (52.5%), with Intel's MIC/Xeon Phi a distant second at 17.5%. Academic sites were also the most interested in the ATI/AMD Fusion (7.5%), compared with only 3.9% interest in the industrial sector, and zero interest in the government sector. Of all segments surveyed, the academic segment was the most likely (87.5%) to use some form of GPU/accelerator/ coprocessor hardware in their next system.
- The industry sector equally favors the NVIDIA GPU and the Intel MIC/Xeon Ph at 29.4%. Industry also expressed the most interest (13.7%) in using FPGAs in their next generation servers.

Table 38

GPU/Accelerator/Coprocessor Preference for their Next Technical Server

	Government Sites	Academic Sites	Industrial Sites	Overall
Don't plan to buy them in next server	33.3%	12.5%	29.4%	23.3%
NVIDIA GPU	16.7%	52.5%	29.4%	36.9%
Intel Phi	25.0%	17.5%	17.6%	18.4%
ATI (ATI/AMD Fusion Processors)	0.0%	7.5%	3.9%	4.9%
FPGA	8.3%	5.0%	13.7%	9.7%
TI/DSP	8.3%	2.5%	2.0%	2.9%
Other	8.3%	2.5%	3.9%	3.9%
N=103				

Source: Hyperion Research, 2017

CHARACTERISTICS OF TOP HPC APPLICATIONS

Table 39 summaries key characteristics for the top 3 applications running at respondents' sites. In general, the statistics support the conclusion that the most important jobs receive the most attention when it comes to tuning for optimal performance and minimal run time; the most important application typically had the shortest run time, and also had the highest level of parallelism in the maximum core point.

- For the most important application, across all users responding, the average run time was almost 48 hours, used an average of almost 2,500 cores and achieved a maximum core count of 14,717.
- For the second most important application across all users responding, the average run time was almost double at 84.9 hours, used an average of over 3,000 cores and achieved a maximum core count of over 10,000.
- For the third most important application across all users responding, the average run time was even higher at 91.5 hours, used an average of over about 2,500 cores and achieved a maximum core count of about 9,800.

Table 39

Characteristics of Top 3 Applications

	Characteristic
#1 - Application	
#1 - Typical run time (hours)	43.8
#1 - Parallelism: average core count	2,484
#1 - Parallelism: maximum core count	14,717
#2 - Application	
#2 - Typical run time (hours)	84.9
#2 - Parallelism: average core count	3,063
#2 - Parallelism: maximum core count	10,337
#3 - Application	
#3 - Typical run time (hours)	91.5
#3 - Parallelism: average core count	2,493
#3 - Parallelism: maximum core count	9,809
N=109	

Source: Hyperion Research, 2017

Table 40 is a break out of Table 39 by competitive sector and reveals some variations in the characteristics of top 3 applications.

- For government sites, typical runs times for all three applications averaged a little over 10 hours, well below the overall average run time across all sectors at 74.4 hours. Likewise, government sites had lower average core counts (456 cores) and maximum cores counts (2252) than did their academic (5,713 and 2,1482 respectively) or commercial (696 and 5042 respectively) counterparts.
- For academic sites, top application there used the highest number of average and maximum core counts seen across the three sectors, in many cases by a factor of 20.
- For industrial sites, top applications had some of the highest typical runs times found in any sector, with second and third tier applications using as much as 20X more run time as government or academic counterparts.

Table 40

Characteristics of Top 3 Applications (by Sector)

	Government Sites	Academic Sites	Industrial Sites	Overall
#1 - Application				
#1 - Typical run time (hours)	17.3	62.5	35.5	43.8
#1 - Parallelism: average core count	645	4,930	1,119	2,484
#1 - Parallelism: maximum core count	5,379	21,092	12,227	14,717
#2 - Application				
#2 - Typical run time (hours)	6.1	41.7	126.1	84.9
#2 - Parallelism: average core count	379	7,023	521	3,063
#2 - Parallelism: maximum core count	1,025	22,605	2,896	10,337
#3 - Application				
#3 - Typical run time (hours)	8.0	75.2	114.4	91.5
#3 - Parallelism: average core count	344	5,726	450	2,493
#3 - Parallelism: maximum core count	353	20,749	3,063	9,809
N=109				

Source: Hyperion Research, 2017

HPC Server Utilization

Hyperion Research asked participants about server utilization rates – that is, as a yearly average, what percentage of their HPC server nodes were actively running user jobs and applications (see Table 14). Data is presented for the three major sectors (government, industry, and academia). Overall average utilization of HPC servers across the surveyed group was 66.0%. The government sector indicated a 69.0% average utilization rate, while the academic sector came in at 68.4% and the industrial sector at 63.3%. Hyperion research notes, however, that many sites are becoming less interested in simply keeping their systems running at full capacity and are instead concentrating more on the overall utility of their complete workflow.

Table 41

Server Utilization Levels in 2017

	Government Sites	Academic Sites	Industrial Sites	Overall
On average, throughout the year, what percent of your HPC server nodes are actively running user jobs/applications?	69.0%	68.4%	63.3%	66.0%
N=112				

Source: Hyperion Research, 2017

WHERE HPC SITES ACQUIRE THEIR SERVICES AND SUPPORT

Hyperion Research asked participants about where they acquire their services and support capabilities. As Table 42 shows, sites typically use a combination of services that include acquiring services and support from traditional hardware vendors (71.6%), in-house staff (53.9%) and VAR/integrators (32.4%).

In addition, about one third of participants plan to expand their in-house staff with additional hires, and just one-fifth plan to outsource service and support (e.g., via cloud).

Table 42

Where HPC Sites Acquire Their Services and Support

	Responses	Percentage of Sites
From traditional hardware vendor	73	71.6%
From a VAR/Integrator	33	32.4%
Use existing in-house staff	55	53.9%
Expand in-house staff with additional hires	35	34.3%
Outsource: going into Cloud, use computing on demand services, etc.	12	11.8%
Other	2	2.0%
N=102		

Note: Multiple responses were allowed.

Source: Hyperion Research, 2017

Table 43 is a break out of Table 42 by competitive sector and reveals some variations in where HPC sites acquire their services and support. Academic sites look to traditional hardware vendors for support and services (80%) at a higher rate than either government (72.7%) or industry sites (64.7%) and expressed a higher interest in expanding in-house staff (40%) than the two other segments. When asked about outsourcing in cloud-based computing services, academic sites were the most interested (15.0%) followed by industrial sites (11.8%). In contrast, no government sites expressed an interest in pursuing cloud-based computing on demand services.

Table 43

Where HPC Sites Acquire Their Services and Support by Sector

	Government Sites	Academic Sites	Industrial Sites	Overall
From traditional hardware vendor	72.7%	80.0%	64.7%	71.6%
From a VAR/Integrator	36.4%	35.0%	29.4%	32.4%
Use existing in-house staff	45.5%	52.5%	56.9%	53.9%
Expand in-house staff with additional hires	27.3%	40.0%	31.4%	34.3%
Outsource: going into Cloud, use computing on demand services, etc.	0.0%	15.0%	11.8%	11.8%
Other	0.0%	2.5%	2.0%	2.0%
N=102				

Note: Multiple responses were allowed.

Source: Hyperion Research, 2017

KEY HPC GROWTH DRIVERS

Hyperion Research is expecting a number of factors to drive healthy growth rates across all segments of the HPC sector going forward and that growth will likely outstrip the growth rate expected for the general-purpose enterprise IT server sector.

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.
- New and rapidly growing opportunities to support the continued migration and expansion of enterprise HPC workloads to cloud-based ecosystems. Hyperion expects that 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 a host of new HPC users.
- The expanding role and diversity of new big data analytics running in non-traditional HPC environments, especially for fraud and anomaly detection, affinity marketing, business intelligence, personalized medicine, 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 HPCs 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 operating under the requirement for faster turnaround time.

Forecast Methodology

The forecasts in this study are based on a number of 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 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 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.

Note: All numbers in this document may not be exact due to rounding.

Synopsis

This Hyperon Research study is this fifth edition of the HPC End-user Special Study, Hyperion Research conducted in-depth interviews with one or more representatives of 111 HPC sites around the world. These sites had a total of 972 technical server systems installed, and the average in the present study was 8.8 systems per site. Worldwide revenue for the HPC technical server market grew 4.4% in 2016 to a record $11.2 billion, up from $10.7 billion in 2015. Current projections call for growth in 2017 of 7.3%, resulting in a total HPC technical server market for 2017 with over $12 billion.

According to Bob Sorensen, Hyperion Research's Vice President for Research and Technology. "Hyperion Research analysts are expecting a number of factors to drive growth rates across all segments of the HPC technical server sector going forward and that growth will likely far outstrip the growth rate expected for the general-purpose enterprise IT server market. This growth will be driven by new demands for deep learning computational platforms and HPDA-driven systems. In addition, Hyperion Research sees a continued expansion of traditional HPC modeling and simulation demand as more governments and commercial firms recognize the economic and competitive advantages of HPC in their business operations."

APPENDIX: SITES SURVEYED IN THIS STUDY

Table 44

Appendix: List of Organizations In The Study

Organization	Organization	Organization
Abbott Laboratories	Iowa State University	Peer Technologies GmbH & Co KG
Air Force Research Laboratory	J.P. Morgan	Pfizer
ANV Syndicates	Jefferson Lab	Philly Shipyard
Bank of America	JetBlue	Princeton University
Bank of America Merrill Lynch	John Wiley & Sons	Procter & Gamble
Berkeley Lab	JPL/NASA	Rothschild
Boston College	Kiln Group	RSC Group
Boston University	Kinder Morgan	RWTH Aachen University
Briggs & Stratton	L&L Products	Sandia National Laboratories
Brunel University	Leibniz Supercomputing Center LRZ	Scottish Borders Council
Carrefour Retail	London School of Economics	Shanghai Jiao Tong University
CERFACS	Manchester City Council	Shared Services Canada
CINECA	Marks and Spencer	Singapore Airlines
Citi	Mayo Clinic	Singtel
City of Chicago	Mellanox	Skanska
City of Westminster	Miami University	SUNY
City University of New York	Michigan State University	Tesco
CNA	MIT / CSAIL	TUI Europe
DataSwing	MITRE	Turner Construction
Dept. Of Defense	NASA Goddard Space Flight Center	Universität Hamburg, Scientific Computing (Informatics)
DFJ	NASA Langley Research Center	Universiti Putra Malaysia
Diagenode	Nationwide	University of Coimbra

Table 44

Appendix: List of Organizations In The Study

Organization	Organization	Organization
DKRZ GmbH (German Climate Computing Centre)	Navy Research Laboratory	University of Edinburgh
Dubai Trade (Government)	NCAR/UCAR	University of Liverpool
Emaar	Neoanyang Technological University	University of Maryland
EverBank	Netflix	University of Minnesota
eXact-lab srl	New York Life Insurance Company	University of Tennessee
Frobozz Magic	NIIT Technologies	University of Tennessee
GigaBack	NOAA	University of Wales
Grayhill	Northwestern University	University of Warwick
Hammersmith and Fulham Council	Nous Infosystems	UPS
Hewlett Packard Enterprises	NYC Department Of Transportation	Vanderbilt University - ACCRE
HLRS	Oak Ridge National Laboratory	Wells Fargo
HSBC	Onx	Western Michigan University
Hughes Network Systems LLC	Pacific Northwest National Laboratory	Western Oregon University
Institute Process Engineering, Chinese Academy of Sciences	Partnership	WWU Münster
International Academy, Research, and Industry Association - Advanced Computing	PDC	XPO Logistics
N = 111		

Source: Hyperion Research, 2017

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
