HYPERION RESEARCH

Market Forecast

Worldwide HPC Server Market Forecast Update, 2017–2022

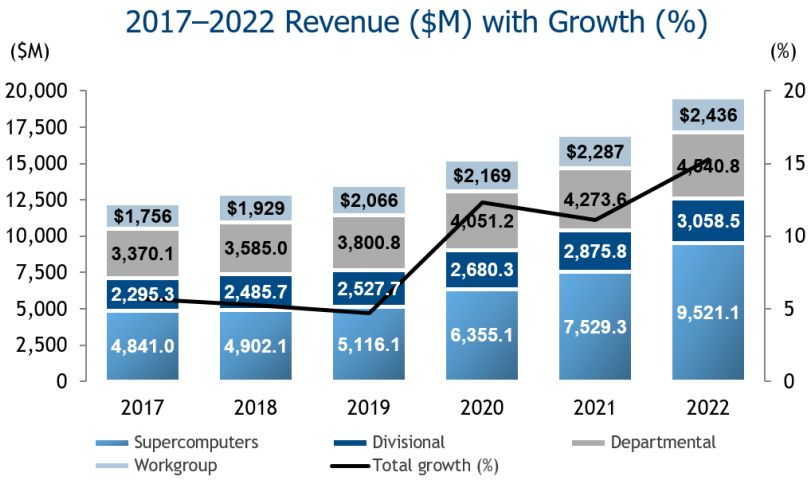
Earl Joseph, Steve Conway, Bob Sorensen, and Alex Norton
June 2018

HYPERION RESEARCH OPINION

This Hyperion Research study presents our latest five-year forecast (2017-2022) for HPC server systems. Worldwide revenue for the HPC technical server market grew 5.8% from 2016 to 2017 to a record $12.3 billion. Hyperion Research predicts the HPC technical server market will grow at a 9.8% CAGR between 2017 and 2022 to reach $19.5 billion by 2022.

- The Supercomputer market segment for HPC systems priced at $500,000 and up is projected to show the highest growth rate (14.5% CAGR), driven substantially by the global exascale race.
- The Divisional and Departmental segments will continue to exhibit healthy growth (around 6.0% five-year CAGR), and the Workgroup segment is rebounding to improved growth following several years of stagnation (6.8% CAGR).

Worldwide HPC Server Revenue Snapshot



Selected Segment Growth Rate

- Supercomputers CAGR 14.5%
- Divisional CAGR 5.9%
- Departmental CAGR 6.1%
- Workgroup CAGR 6.8%

Total Market CAGR 9.8%

Source: Hyperion Research 2018

Note: This page is intentionally blank.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

IN THIS REPORT

This Hyperion Research study presents our latest five-year forecast (2017-2022) for the HPC server space.

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

Worldwide revenue for the HPC technical server market grew 5.8% in 2017 to a record $12.3 billion, up from $11.6 billion in 2016. Current projections call for growth in 2018 of around 5%, resulting in an HPC technical server market total for 2018 of $12.9 billion.

The worldwide HPC server sector is projected to grow at a CAGR of 9.8% between 2017 and 2022, reaching $19.6 billion in 2022.

A number of factors are driving aggressive growth rates across all segments of the HPC sector and that growth will continue to outstrip the growth rate expected for the general-purpose enterprise IT server sector.

- The worldwide exascale race will likely see multiple, large, over $400 million supercomputers, installed in during the forecast period, plus a number of large, $100 million plus supercomputers in the same timeframe.
- Despite the extreme nascent state of quantum computing development today, activity in the highly visible quantum computing section will soon create additional markets for HPCs to be used as quantum simulators to help development of quantum computing algorithms. Here, the ability of HPCs to field large memory spaces will be a necessary feature to support digitally-based quantum simulation. In addition, quantum algorithm development will increasingly require HPC capability in order to validate and verify new or emergent quantum algorithms.
- New and rapidly growing opportunities are emerging to support the continued migration and expansion of enterprise HPC workloads to private and other cloud-based ecosystems. A key development going forward is the emergence of so-called containers that orchestrate computing, networking, and storage infrastructure on behalf of user workloads between on prem and cloud HPC platforms. Hyperion Research 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 HPC capabilities.
- The seemingly endless roll-out of new algorithms, applications and use cases of machine and deep learning will continue to drive interest in HPCs that can provide fast, capable performance for even the most aggressive deep learning training and operational sessions. As new deep learning networks continue to emerge, the ability of HPC systems to match the computational requirements of those networks will become an increasingly important driver of HPC development.
- 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.

ADVICE FOR TECHNOLOGY SUPPLIERS

The HPC sector experienced healthy growth in 2017 (5.8%) following 2016's strong growth rate (8.1%). This was due to a combination of demand for HPCs bound for the traditional modeling and simulation sectors as well new growth in areas that include the AI sector, particularly in deep learning, and the big data applications across a growing set of users.

Hyperion Research advises vendors to consider the following:

The range of processors available for use in HPC systems is wider than it has been in a long time. As such, vendors will need to move carefully as they decide which processors to support as well how they deploy variants of the processors for different market segments. Processor options relevant to the HPC sector today include x86 (both Intel and AMD), ARM, Power/OpenPower, SPARC, and others.

- Although commitment to a single processor allows a vendor to concentrate its resources on developing leading-edge capabilities, it could lock a vendor out of important sectors. Hyperion observes that as the spectrum of HPC workloads widens, especially with growth in advanced analytics, demand is growing for HPC systems that are more heterogeneous, including on the processor/accelerator side. We foresee cases in which it will make sense for a single system to include Intel x86 CPUs, AMD x86 CPUs, and ARM or other processors, as well as GPU or FPGA accelerators.
- Sophistication of the processor software stack ecosystem development will be a key user determinant of processor choice going forward.

The diversity and inclusion of GPUs, FPGAs, and other accelerators continues to increase, and their promise of potential for high performance are driving the growing trend towards heterogenous HPC architectures. 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 and third-party 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 users will increasingly be looking for in their HPC configurations. Although each offer both opportunities and challenges for HPC vendors looking to develop systems uniquely configured for key use cases, 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.

HPC users are increasingly looking to include some form of cloud-based computational or storage capability in their overall HPC set-up. HPC vendors will need to offer seamless support for the efficient migration of computational tasks and associated data between on-prem and cloud-based resources including private clouds, public clouds, virtual private clouds, and hybrid configurations.

Use cases that span traditional high performance computing (HPC) modeling and simulation jobs, big data analytics, and AI applications including deep learning are placing an increasing demand on file systems to store and deliver the right data to the right place in a cost-effective manner across a wide mix of storage technology options. Increasingly HPC vendors will need to be able to offer a storage options that provides more than simple data movement across the data storage infrastructure for efficient storage utilization. New features will center on efficient, seamless, and automatic support for a

tiered storage infrastructure that can consist of SSDs, fast nearline hard drives, archival hard drives, cloud-based storage, and tape.

In addition to technology, vendors need to consider other forces at play as they navigate within the changing HPC landscape. We offer these recommendations:

The convergence of the so-called three pillars of HPC, computational modeling and simulation, big data, and artificial intelligence centered on deep learning will be redefining the emphasis and diversity of workloads seen in many HPC sites. HPC hardware will necessarily become more diverse and heterogenous to support this wider range of computational demands and as such users will be looking for HPC vendors with the ability to offer effective systems and applications software to manage the complete HPC workload.

Hyperion Research predicts the emergence of HPCs whose resources are dynamically reconfigurable in real time, where software-defined partitioning enables a single system to operate effectively as if it were a collection of relatively tightly coupled, diverse systems.

Likewise, there may be a small but growing class of users who will seek to acquire bespoke HPCs targeted for a single or small set of critical applications. Although the hardware composition of these systems likely will be based on key performance capabilities of select processors, accelerators, interconnects, or memory systems, the ability to match an HPC design to a particular workload will be key. Successful HPC vendors will be able to provide options for 'mix and match' HPCs to meet a broad range of what likely will diverse sets of user technical specifications.

There will be growing opportunities for 'accidental' HPC users in the next few years, particularly those driven into the HPC world as a result of new big data analysis opportunities presented by the larger data sets that they are amassing. Many of these potential new users either soon will or have already outgrown the capabilities of their traditional enterprise hardware and software to effectively manage and analyze their data. Finding, educating, and then supporting such users will be costly at first, but could have significant payoff in the long-term.

Although more users will be seeking out options for cloud-based HPC opportunities, many are not in a strong position to understand the cost and performance dynamics of implementing support for a mixed on-prem/cloud HPC operation. HPC vendors could benefit significantly by building on their in-house HPC operational skills to help broker, set up, or even manage an integrated on-prem/cloud HPC resource.

The push towards exascale has significantly heightened the political visibility of the HPC sector, with a decided growing emphasis on indigenous HPC technology development for these planned high-end systems. HPC vendors need to consider a global HPC strategy that is not one-size-fits-all but that instead seeks to establish a credible technical or marketing presence across a board range of diverse and perhaps increasingly closed national or region boundaries. Being perceived as a domestic HPC developer or supplier could become an increasingly important feature of any major HPC vendor that seeks to sell across a wide range of county markets.

HPC MARKET FORECASTS

HPC used in traditional modeling and simulation applications, coupled with new demands for big data analytics, deep learning training systems, and AI will fuel continued growth in the sector well beyond the forecast period.

The government lab, finance, HPDA, government, cyber security, and the academic spaces are projected to be the brightest spots during the forecast period. As the exascale race around the world becomes fiercer, we expect to see more nations announcing their plans and more mega-systems being rolled out in the forecast period.

Overall Forecasts

Table 1 shows worldwide HPC server revenue, shipments, and average selling prices (ASP). Figures 1, 2, and 3 illustrate HPC server revenue, shipments, and average selling prices, respectively.

Table 1

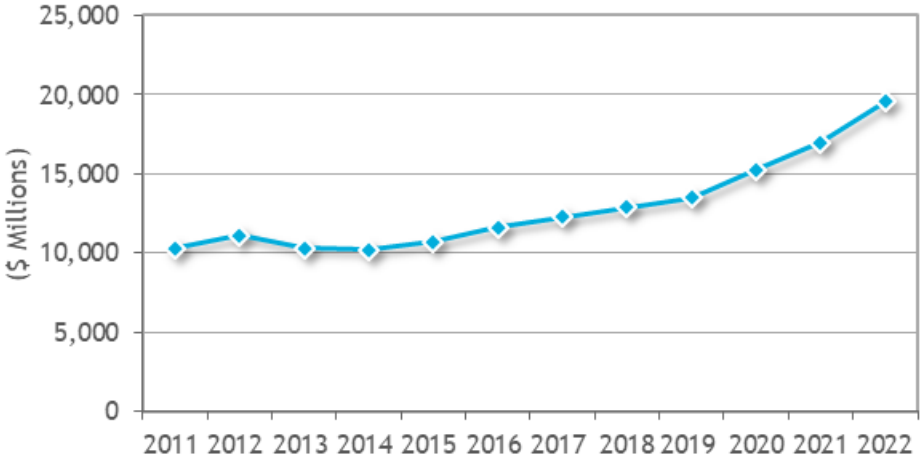
Worldwide Total High-Performance Technical Systems Total Market Forecast by Revenue, Units and ASP, 2011 - 2022

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	CAGR 17-22
Revenue $M	$10,300	$11,098	$10,299	$10,222	$10,727	$11,595	$12,262	$12,902	$13,510	$15,256	$16,966	$19,557	9.8%
Units	111,553	104,148	123,982	133,392	99,817	86,435	88,560	89,597	91,525	94,591	97,473	100,825	2.6%
ASP $000	$92	$107	$83	$77	$107	$134	$138	$144	$148	$161	$174	$194	7.0%

Source: Hyperion Research 2018

FIGURE 1

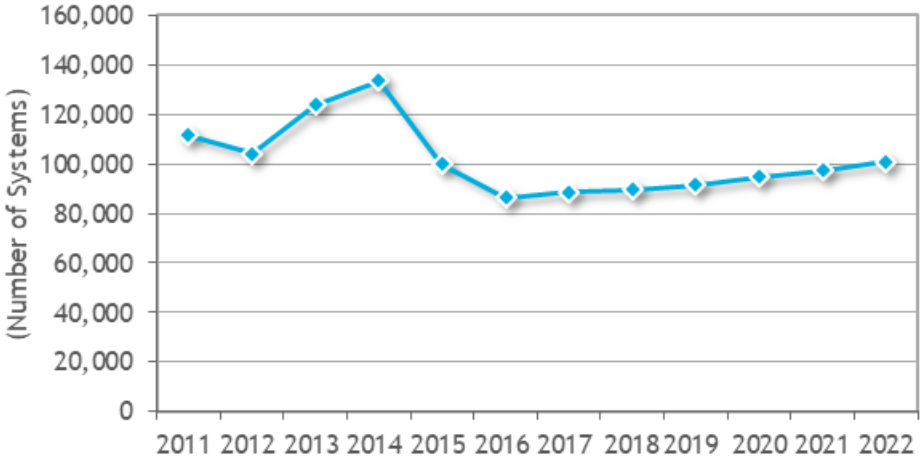
Worldwide High-Performance Technical Systems Total Market Revenue Forecast



Source: Hyperion Research 2018

FIGURE 2

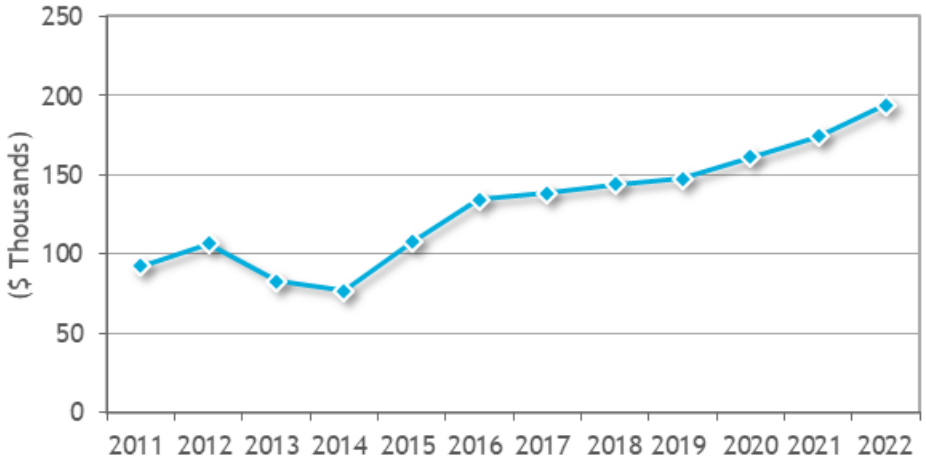
Worldwide High-Performance Technical Systems Total Market System Unit Forecast



Source: Hyperion Research 2018

FIGURE 3

Worldwide High-Performance Technical Systems Total Market Average Selling Price Forecast



Source: Hyperion Research 2018

Forecasts by Competitive Segments

Tables 2, 3, and 4 present HPC revenue, shipment, and ASP forecasts, respectively, by Hyperion Research's four main competitive price band segments. From a competitive segment perspective, between 2017 and 2022, we expect to see the highest growth in the supercomputer segment, with a CAGR of 14.5%. Next is the workgroup segment (6.8%) and then the departmental and divisional segments (around 6% CAGR).

Table 2

Worldwide Total Technical Computer Market Revenue Forecast by Competitive Segment

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	CAGR 17-22
Supercomputer	$4,370	$5,655	$3,995	$3,149	$3,241	$4,485	$4,841	$4,902	$5,116	$6,355	$7,529	$9,521	14.5%
Divisional	$1,237	$1,216	$1,355	$1,524	$1,922	$2,273	$2,295	$2,486	$2,528	$2,680	$2,876	$3,059	5.9%
Departmental	$3,467	$2,979	$3,363	$3,831	$3,757	$3,147	$3,370	$3,585	$3,801	$4,051	$4,274	$4,541	6.1%
Workgroup	$1,226	$1,247	$1,586	$1,718	$1,806	$1,689	$1,756	$1,929	$2,066	$2,169	$2,287	$2,436	6.8%
Total	$10,300	$11,098	$10,299	$10,222	$10,727	$11,595	$12,262	$12,902	$13,510	$15,256	$16,966	$19,557	9.8%

Source: Hyperion Research 2018

Table 3

Worldwide Total Technical Computer Market Shipments Forecast, by Competitive Segment

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	CAGR 17-22
Supercomputer	2,908	2,400	1,484	1,383	1,392	1,590	1,526	1,634	1,471	1,324	1,158	1,027	-7.6%
Divisional	3,724	3,663	4,271	4,788	6,033	5,936	5,597	6,225	6,409	6,435	6,460	6,487	3.0%
Departmental	20,624	16,981	20,246	22,378	23,267	19,442	20,616	20,349	21,213	22,576	23,719	25,243	4.1%
Workgroup	84,294	81,104	97,981	104,843	69,125	59,467	60,821	61,389	62,432	64,257	66,135	68,068	2.3%
Total	111,550	104,148	123,982	133,392	99,817	86,435	88,560	89,597	91,525	94,591	97,473	100,825	2.6%

Source: Hyperion Research 2018

Table 4

Worldwide Total Technical Computer Market Average Sales Price Forecast, by Competitive Segment ($K)

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	CAGR 17-22
Supercomputer	1,503	2,356	2,691	2,277	2,328	2,821	3,172	3,000	3,479	4,801	6,501	9,268	23.9%
Divisional	332	332	317	318	319	383	410	399	394	417	445	472	2.8%
Departmental	168	175	166	171	161	162	163	176	179	179	180	180	1.9%
Workgroup	15	15	16	16	26	28	29	31	33	34	35	36	4.4%
Total	92	107	83	77	107	134	138	144	148	161	174	194	7.0%

Source: Hyperion Research 2018

Major Assumptions

Technology and Markets Forces Drive Growth at the High End of the Market

The number and scope of recent announcements about new petaflops systems around the world are clear indications that the high end "exascale race" is alive and well, and spending growth will no doubt follow. Hyperion Research estimates that the supercomputer segment of the HPC market will be one of the fastest growing with a revenue CAGR of 14.5% between 2017 and 2022.

- It is interesting to note that for the same time period, unit shipments of supercomputer-class systems (systems purchased to support technical applications and sold for $500,000 or more) are projected to decline at a CAGR of -7.6%%, the lowest unit shipment growth rate of any class of HPC systems. This reflects the fact that the average sales price of a supercomputer-class system will be increasing at a faster rate than for any other HPC class of machines. Simply put, there will be a relative slowdown in the number of supercomputer-class HPCs sold, but each one on average will be sold at a higher price, especially as the early exascale systems are installed.

HPDA Growth Expands the Space

The fastest revenue gains in the HPC sector writ large will be driven by the HPDA-AI sector. HPDA-AI server revenues within established HPC market segments will grow at a CAGR of almost 17 % out to 2022, and new HPDA commercial analytics segments will see a CAGR of over 26% during the same time frame. Examples here include new big data applications that are running in non-traditional HPC environments but that use HPC hardware, such as in the financial services, health care and cyber security sectors.

- Commercial firms that currently are engaging in traditional enterprise business analytics are increasingly integrating HPCs into their enterprise data centers to address some of their more complex, time sensitive, or data rich problems. Despite this, many of these users likely will not strongly identify with or be strongly identified by the traditional HPC sector as part of the HPC universe.
- The process whereby these 'new' users enter the HPC universe will be interesting to watch as they will add their own unique experiences, expectations, and requirements to the mix.

The Rise of Deep Learning Applications

HPC has moved to the forefront of R&D for deep learning and early AI applications.

Definitions of Deep Learning and AI

- Deep Learning. What makes deep learning stand out from machine learning and other big data methods is deep learning's use of digital neural networks that enable the computer to learn on its own, to go beyond the training it received. Deep learning computers don't think in the same way humans think, but they can accomplish an expanding range of things human thinking accomplishes, only faster. Because deep learning computers can learn on their own, they will be able to perform complex tasks in rapidly changing environments, such as driving autonomous vehicles or diagnosing a baffling disease, without the need for explicit programming or human oversight.

- **Artificial Intelligence.** We see AI as a broader, more generic term for the ability of computers to do things human thinking does. Machine learning and deep learning both qualify as examples of early AI. Experts generally agree that it may take another decade or more for computers to meet the Turing Test, where a person listening to a natural-language conversation between a human and a machine can't tell which party is the machine.

Users everywhere are looking at hardware to support deep learning applications that emphasize high computational capability, large memory capacity, and strong interconnect schemes. Currently, the state of deep learning development is wide and varied with a host of large and small hardware, software and perhaps most dominantly, end users contributing to the state of art in deep learning development and applications.

- Some of the most aggressive and successful developers of deep learning expertise are end users, some of whom are designing and building systems to their own internal specifications.
- A number of HPC vendors have rolled out devices or systems for either one or both of the two main deep learning computational realms: the HPC-based training phase that requires high-end computation capabilities that can consume days or weeks of HPC time and the more local, widely deployed and less technically demanding inference engines where key insights developed during the training phase are translated into on the ground decision making, such as in speech recognition on a smart phones or autonomous vehicle navigation.

Even Google doesn't fully understand exactly how its DeepMind computer manages to defeat the best human competitors in the AlphaGo computer game, but when and if an autonomous vehicle crashes, you can bet that automakers and auto insurers will need to understand exactly what the computing systems did, or failed to do, that may have contributed to the collision. That will be crucial for consumer confidence in self-driving vehicles (and for insurance underwriting). Hyperion Research expects automakers to begin adding this task to existing uses for HPC simulation in upstream vehicle design.

Drivers and Inhibitors: Drivers

HPC Leadership Efforts

Assumption: National HPC exascale programs will have a material impact on the development of state-of-the-art HPC technology that will flow down into a wide range of HPC systems. This will lead to more costly early exascale systems, sometimes with prices above $500 million per system.

Government programs around the world are advancing the state of the art in HPC toward exascale computing capabilities in the 2020 and beyond timeframe, with a number of national programs already in place. These programs include:

- The US NSCI/Exascale Computing Project run out of the Department of Energy that is targeting a peak exascale system in 2021 and sustained exascale system capabilities by 2022-2023 at a cost of $300-600 million per system. Current US Government and vendors partners' investments are running at about $1 to $2 billion per year in R&D, including NRE. Plans call for the purchase of multiple exascale systems each year.
- The EU PRACE and ETP4HPC initiatives that are targeting pre-exascale systems in 2021-2022 and sustained exascale systems in the 2023-2024 time frame. These systems will likely be a mix of indigenous European and non-European technology such as ARM-based processors of European design. European planners are looking to spend $300-$350 million per system, and investments for exascale-related R&D are happening at both the EU and country levels.

- Japan's Post-K Project managed by Riken, one of Japan's leading government R&D centers, that is looking to develop a sustained exascale system in the 2022 time frame using Japanese commercial partners for system hardware and software. One of the more aggressive designs under consideration around the world, the expected cost of the largest system built under this project could go as high as $800 million to $1 billion. Program planners are also looking to spin off a number of smaller ($100 - $150 million) follow-on systems using the same technology base. RIKEN is actively collaborating with France's CEA organization.
- China's Exascale Development Project that targets a peak exascale system in 2020 and sustained exascale system by 2023, at a total cost of $350-500 million per system. China is currently funding three different pre-exascale systems to be evaluated in the late 2018 time frame before deciding on a final design and associated vendor. Spending over $1 billion a year in R&D, China is working hard to field exascale systems comprised primarily of indigenous technology.

Impact: Competition in the global high-end HPC server market will heat up as both hardware and software technology developed through these projects work their way into commercial products. As such, the impact that each of the individual national programs have on any single country or region to advance its position as a world-class developers and users of HPC could be significant.

Economic Growth and HPCs

Assumption According to a recent World Bank assessment, the global economy is experiencing a cyclical recovery, reflecting a rebound in investment, manufacturing activity, and trade. This improvement comes against the backdrop of benign global financing conditions, generally accommodative policies, rising confidence, and firming commodity prices. Global GDP growth is estimated to have picked up from 2.4 percent in 2016 to 3 percent in 2017. The upturn is broad based, with growth increasing in more than half of the world's economies.

- In advanced economies, growth in 2017 is estimated to have rebounded to 2.3 percent, driven by a pickup in capital spending, a turnaround in inventories, and strengthening external demand. While growth accelerated in all major economies, the improvement was markedly stronger than expected in the Euro Area, based on the same World Bank analysis.
- Growth among developing nations is estimated to have accelerated to 4.3 percent in 2017, reflecting firming activity in commodity exporters and continued solid growth in commodity importers. The improvement in economic activity among commodity exporters took place as key economies such as Brazil and the Russian Federation, emerged from recession, prices of most commodities rose, confidence improved, the drag from earlier policy tightening diminished, and investment growth bottomed out after a prolonged period of weakness, based on the same World Bank assessment.

Impact: In a time of optimistic overall growth projections for both developed and developing economies, the HPC sector is increasingly seen as one that can help fuel industrial and economic competitiveness in a time of growing global competition. In addition, HPC is increasingly seen as a key driver of such growth and as HPC server sales growth rates are projected above national economic averages. For less developed economies, particularly commodity suppliers that are looking to move up the technology food chain, HPC technology can be a necessary element to invigorate higher-margin manufacturing capabilities.

Drivers and Inhibitors: Inhibitors

Widespread Innovation in the HPC Sector Can Muddy the Waters

Assumption: The continual roll out of new HPC-related technology, particularly in the areas of deep learning, HPC in the cloud, and advanced big data analytics, all of which show great promise in becoming significant market drivers, will also create 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.

Impact: Both vendors and users need to balance carefully 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 that may not be sustainable in the long term. Likewise, moving too slowly opens the door for more nimble players to build-up an insurmountable technology or market lead.

- 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 serve as powerful mechanisms to help ensure that vendors, users, and technology trends remain in sync.

Significant Market Developments

A number of factors will drive growth rates across all segments of the HPC sector going forward including:

- The growing interest in roll-out of new pre-exascale and exascale systems will create significant interest at the highest end of the HPC sector. Much of the technology developed for those systems will find its way over time into smaller, less costly, and more widely applicable systems.
- Despite the extreme nascent state of quantum computing development today, activity in the highly visible quantum computing section will soon create additional markets for HPCs to be used as quantum simulators to help development of quantum computing algorithms. Here, the ability of HPCs to field large memory spaces will be a necessary feature to support digitally-based quantum simulation. In addition, quantum algorithm development will increasingly require HPC capability in order to validate and verify new or emergent quantum algorithms.
- New and rapidly growing opportunities are emerging to support the continued migration and expansion of enterprise HPC workloads to cloud-based ecosystems. A key development going forward is the emergence of so-called containers that orchestrate computing, networking, and storage infrastructure on behalf of user workload between on prem and cloud HPC platforms. 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 HPC capabilities.
- The seemingly endless roll-out of new algorithms, applications and use cases of deep learning will continue to drive interest in HPCs that can provide fast, capable performance for even the most aggressive deep learning training sessions. As new deep learning networks continue to emerge, the ability of HPC systems to match the computational requirements of those networks will become an increasingly important driver of HPC development.

Although Hyperion Research expects x86-based processors to remain dominant during this period, coprocessors and accelerators, especially NVIDIA GPGPUs, but also FPGAs, will see increased traction in the 2018-2022 period. Low-power processors, such as those based on ARM IP, also will begin finding their place in the HPC ecosystem. Open software initiatives are starting to gain market presence as well.

- Coprocessors and accelerators are rapidly gaining momentum in the HPC community today, and mainstream adoption is largely dependent on the programming models and application readiness.
- In addition, ISV support will be critical for the growth potential of these components in the overall HPC sector.

Within the overall HPC sector, HPC storage is increasingly playing a critical role in determining the overall performance and utility of a complete HPC system. There are a number of parallel trends contributing to this:

- Data access is rapidly becoming the primary limiting factor in HPC performance. Almost all processors today and going forward realize their increasing rates of computational performance by adding additional parallel computational units and require greater data access capabilities to run at their full potential.
- As such, the demands for more data at higher transfer rates and with lower wait times for reads and writes is shifting the burden of overall system performance away from the processor to the memory/storage system.
- The case is complicated by the proliferation of HPC designs that rely heavily on specialized processors such as graphic processing units (GPUs) or other custom accelerators that perform a limited class of specialized, typically data intensive operations.
- In addition, new algorithms, such as those in deep learning, are driving the need for variable precision floating point and integer operations ranging from 8 to 64 bit and beyond.

In response, HPC designers are increasingly looking to implement new storage/memory schemes to speed performance. Indeed, the inclusion of new features such as multi-level caches, burst buffer schemes, solid state disks, flash memory options, network supported data processing capabilities, and myriad disk options that span the price/performance spectrum are all becoming an integral part of the HPC storage mix.

- In such architectures, the ability to control the flow of data across a single job, system, or even a data center is becoming increasingly complex.
- Optimizing and simply managing that flow in a real-time environment will increasingly become the responsibly of a sophisticated hierarchical storage management system.

CHANGES FROM PRIOR FORECASTS

Table 5 and Figure 4 compare the previous Hyperion Research HPC server market forecast (see *Worldwide HPC Server 2016-2021 Forecast,* HR00001.4617, June 2017) with this new five-year forecast. The new forecast represents a major increase over Hyperion Research's April 2017 forecast.

Table 5

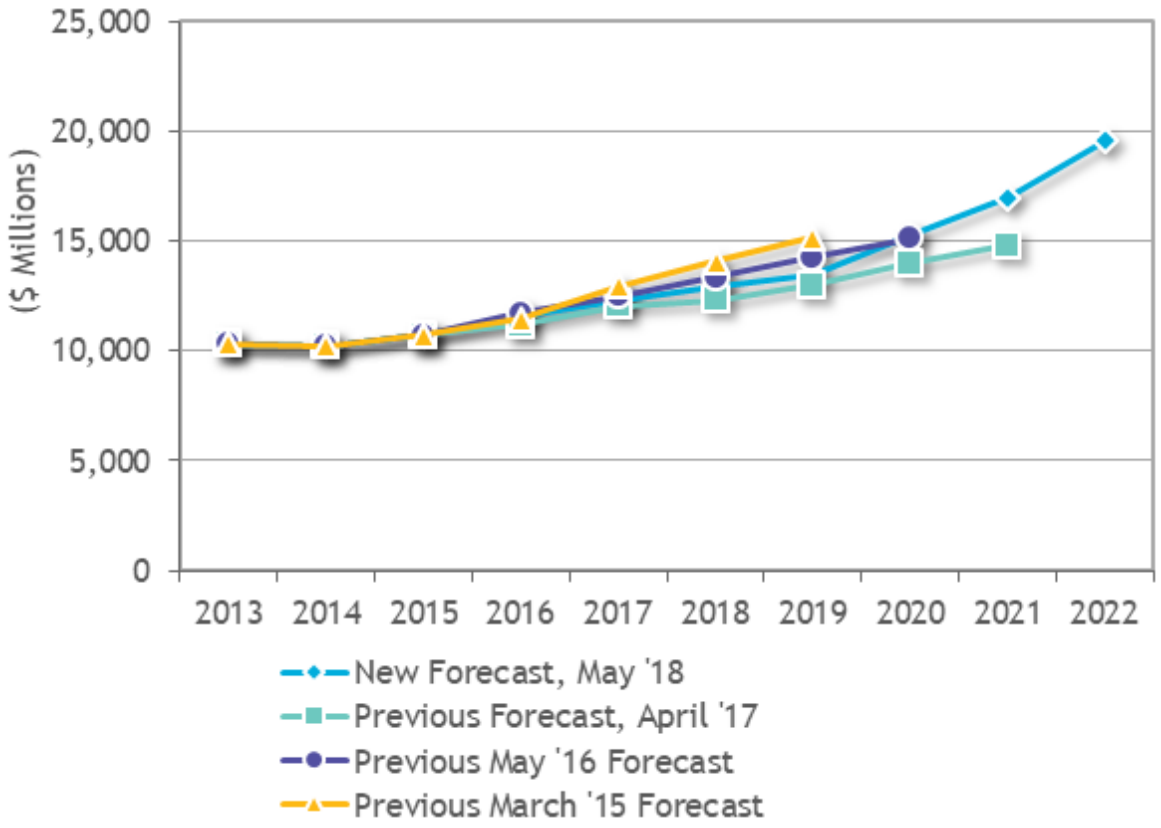
Worldwide Technical Computing Server Revenue: Comparison of New to Previous Forecasts ($M)

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
New Forecast, May '18	$10,300	$11,098	$10,299	$10,222	$10,727	$11,595	$12,262	$12,902	$13,510	$15,256	$16,966	$19,557
Previous Forecast, April '17	$10,300	$11,098	$10,299	$10,222	$10,727	$11,200	$12,019	$12,302	$12,988	$13,997	$14,819	
Previous May '16 Forecast	$10,300	$11,098	$10,299	$10,222	$10,727	$11,703	$12,505	$13,362	$14,254	$15,104		
Previous March '15 Forecast	$10,300	$11,098	$10,299	$10,222	$10,718	$11,467	$12,958	$14,073	$15,165			

Source: Hyperion Research 2018

FIGURE 4

Comparison of Old and New HPC Revenue Forecasts



Source: Hyperion Research 2018

APPENDIX: MARKET SEGMENT DEFINITIONS

This Hyperion Research study presents an overview of Hyperion Research's forecast for the HPC server market for the 2018–2022 period. The data in this study is based on Hyperion Research's segmentation of the technical market, which is as follows:

- **Supercomputers:** Systems purchased to support technical applications and sold for $500,000+
- **Technical divisional servers:** Systems purchased to support technical applications and sold for $250,000–$499,999
- **Technical departmental servers:** Systems purchased to support technical applications and sold for $100,000–$249,999
- **Technical workgroup servers:** Systems purchased to support technical applications and sold for under $100,000

APPENDIX: METHODOLOGY

The forecasts in this study are based on a number of Hyperion Research information sources, including our 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.

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
