LENOVO ENABLES SOLUTIONS FOR THE NEW HPC

Addison Snell

White paper
January 2019

EXECUTIVE SUMMARY

The 2019 version of High Performance Computing (HPC) has changed considerably from previous years'. Most notably, advancements in Artificial Intelligence (AI) has given rise to data-centric applications in machine learning. Machine learning often relies on HPC technologies, particularly in the early training phase, and in many cases, machine learning initiatives share infrastructure, budgets, and personnel with established HPC installations. A 2018 Intersect360 Research study found that the majority of experienced HPC users had also embarked on machine learning applications, usually overlapping HPC and AI.

In some ways, HPC and AI are merging, even at the workload level. Machine learning's ability to extrapolate likely outcomes or solutions based on previous experience can eliminate unnecessary, costly computation. Forward-thinking organizations are examined areas in which machine learning augments HPC, not where it replaces it.

Machine learning is not the only example of expanded use cases in the new HPC. Big data and analytics transformed enterprise computing in the first half of the decade, creating a new category of high-performance, data-centric workloads. As with machine learning, big data applications tended to share resources with other HPC applications, without a dramatic increase in budgets to accommodate new requirements.

The new HPC, inclusive of analytics and AI, and with its wide range of technology components and choices, presents significant challenges to a commercial enterprise. Performance is still paramount—it isn't called "Medium Performance Computing"—but organizations still need to consider their choices carefully before investing in any new technology or selecting an HPC vendor.

Lenovo has quickly established itself as a top-tier HPC provider. Since late 2014, when Lenovo acquired IBM's x86 server line of business, Lenovo has steadily increased its presence with HPC end users. In Intersect360 Research surveys over the past three surveys, Lenovo has increased its share of named systems vendors for HPC more than any other vendor. At the high end of the market, Lenovo was also the only vendor to increase its share of named, surveyed systems with over 1,000 nodes in each of the last two surveys.

Much of Lenovo's current momentum comes from its worldwide manufacturing capabilities and scale. The company has 36 manufacturing facilities at six sites around the world and claims it produces in excess of 100 servers every hour, giving Lenovo influence and reach with its suppliers. Lenovo owns and controls its own manufacturing, giving the company tight oversight of quality and security, with multiple levels of environmental, electromagnetic, and reliability testing. Lenovo has over 10,000 support personnel worldwide, with 51 major contact centers and support teams in 165 countries. Moreover, Lenovo's HPC and AI Innovation Centers are nexuses for advancement for Lenovo customers at the forefront of research and development.

Ultimately, the new HPC brings about opportunities that are worth the challenges. AI is promising for more than just parlor tricks and the Turing test, but the biggest breakthroughs may be hidden inside broader innovation initiatives, where machine learning blends with AI and HPC to unlock another secret of the universe. That's what the new HPC is all about, and Lenovo will be there for it.

MARKET DYNAMICS

The New HPC: New Workloads, New Markets, New Technologies

High performance computing (HPC) is an enduring market. As long as there are scientific discoveries to be uncovered or new innovations to be realized, there will be a need for the computational tools that make them possible. Unlike general enterprise IT markets, HPC is driven year after year by the need to provide more computational power and efficiency within capital and operational constraints. As time goes on, the perpetual gains in the HPC market bring about a corollary to a well-known proverb. In this case, the more things stay the same, the more they change.

AI and Big Data

HPC is a long-term growth market. Intersect360 Research forecasts worldwide spending on HPC technologies and services to have reached $38 billion in 2018, up 7.6% from the prior year, and showing 4.2% compound annual growth rate (CAGR) over the past five years.[1] This steady growth rate implies stability. In fact, HPC is an intensely dynamic market, with new workloads, new markets, and new technologies perpetually usurping old models. "The king is dead. Long live the king!"

The 2019 version of HPC has changed considerably from previous years'. Most notably, advancements in Artificial Intelligence (AI) has given rise to data-centric applications in machine learning (including deep learning, a recursive machine learning approach with ongoing algorithmic modification based on new data). Machine learning often relies on HPC technologies, particularly in the early training phase, and in many cases, machine learning initiatives share infrastructure, budgets, and personnel with established HPC installations. A 2018 Intersect360 Research study found that the majority of experienced HPC users had also embarked on machine learning applications, usually overlapping HPC and AI.[2]

In some ways, HPC and AI are merging, even at the workload level. Machine learning's ability to extrapolate likely outcomes or solutions based on previous experience can eliminate unnecessary, costly computation. Forward-thinking organizations are examined areas in which machine learning augments HPC, not where it replaces it.

Consider, for example, possibilities in computational steering. If AI can be taught to devise strategies for rules-based games like chess or go, it isn't hard to imagine it playing a different game, such as *optimize the aerodynamics of the minivan*. Given a set of constraints, a machine learning algorithm can explore possible solution spaces, reporting to the engineer ideas of where to look for likely improvements. Such usage of machine learning such not displace HPC; it complements it.

Machine learning is not the only example of expanded use cases in the new HPC. Big data and analytics transformed enterprise computing in the first half of the decade, creating a new category of high-performance, data-centric workloads. As with machine learning, big data applications tended to share resources with other HPC applications, without a dramatic increase in budgets to accommodate new requirements.

While the buzz around big data has faded, those workloads haven't gone away. Analytics became a new requirement within the high-performance data center. Where an organization may have been pursuing scientific and technical computing for years, it now incorporates big data and machine learning workloads into the same greater envelope.

Experienced HPC users' requirements have increased. Meanwhile, the new HPC extends to new users as well.

---

1 Intersect360 Research, "Worldwide HPC 2017 Total Market Model and Forecast: Products and Services," June 2018.
2 Intersect360 Research, "HPC User Budget Map Survey: Machine Learning's Impact on HPC Environments," September 2018.

Intersect360 Research
P.O. Box 60296 | Sunnyvale, CA 94088 | Tel. [888] 256-0124
www.Intersect360.com | info@Intersect360.com

Commercial HPC has been the primary engine of growth over the past several years, and the availability of reliable, supported, enterprise-ready solutions contributes to ongoing adoption.[3]

Computation, Data, and the Cloud

The influx of analytics and AI has done more than extending the scope of workloads in an HPC environment. It also has had a profound influence on the wealth of technologies available.

Big data brought about an intensified interest in technologies for maximizing data throughput. Flash memory solutions saw a corresponding boom in adoption. Some enterprises found they could pursue all-flash array implementations, while others looked toward flash tiers for hot data. Storage interconnects also started to change, with more end users adopting InfiniBand as a high-bandwidth option. Parallel file systems also migrated into enterprises seeking to maximize data bandwidth to large storage systems. Certainly, flash memory was available before big data became an industry buzzword, but the need for high-throughput analytics was the accelerant that launch it into a wide range of high-performance storage installations.

What big data did for storage architectures, machine learning is doing for computation. While Intel x86 processors still are the overwhelmingly dominant choice for HPC systems, by mid-decade, there were already myriad options available as accelerators to Intel Xeon systems. As AI took off, so did these alternatives. GPUs (graphics processing units) from NVIDIA are the most common accessory, but there are also roles for programmable FPGAs (field programmable gate arrays) and specialized AI inferencing processors.[4] And even the "standard" choice of Intel Xeon offers a multitude of options across a range of performance, suitable for a wide range of applications.

As if that confluence of technology choices didn't provide enough possible permutations in HPC architectures, the past few years have also allowed more choices for computing high-performance workloads in the cloud. In 2017, HPC cloud usage grew by 44% by revenue, crossing $1 billion in worldwide spending for the first time. This still represented only 3.1% of total HPC spending, but this percentage is expected to rise past 5% by 2022, and furthermore, it mostly represents hybrid cloud usage.[5] Just like a traveler who owns a car yet still sometimes takes a train, many HPC users predominantly consume on-premise infrastructure, leveraging cloud resources when it makes sense for particular workloads.

Becoming a High-Performance Enterprise

With so many choices available, at first blush it may seem straightforward to select a suitable system for HPC. But against a wide diversity of applications and workloads, a correspondingly wide diversity of technology choices can be daunting. No option is as simple as the selection of it. Users must carefully consider how to tier, migrate, and archive data for an optimal balance of performance and cost. Applications need to be optimized and possibly ported to take advantage of the computing, storage, and interconnect deployed.

Machine learning has introduced yet another variable. Most traditional scientific computing applications are deterministic: They take a set of inputs and calculate a specific answer (3.5 x 4.25 + 1.75 = 16.625). Typically, these are computed with 64-bit precision (or "double precision"); that is, with 64 binary digits of accuracy before truncating or rounding any remaining fractions. By contrast, AI is inherently experiential, not deterministic. It predicts that a given object is a stop sign based on other stop signs it has encountered in the past. The nature of these calculations is that they are often streamlined to run at 32-bit "single precision," 16-bit "half precision," or even less, particularly for inferencing tasks. These lower-precision calculations use fewer digits of accuracy and

[3] Intersect360 Research, "Worldwide HPC 2017 Total Market Model and 2018–2022 Forecast: Vertical Markets," September 2018.
[4] Intersect360 Research, "The Top of All Things in HPC, 2017 Site Census Snapshot Analysis," April 2018.
[5] Intersect360 Research, "Worldwide HPC 2017 Total Market Model and Forecast: Products and Services," June 2018; and "Worldwide HPC 2017 Total Market Model and Forecast: Cloud Categories," June 2018.

Intersect360 Research
P.O. Box 60296 | Sunnyvale, CA 94088 | Tel. [888] 256-0124
www.Intersect360.com | info@Intersect360.com

therefore run much faster. As AI melds with HPC, researchers may experiment with mixed-precision workloads, choosing where to spend valuable computational cycles on number-crunching accuracy, and where an application can be less exacting. Even within the classical boundaries of technical computing, this approach to selectively lowering precision may prove revolutionary to some applications.

The new HPC, inclusive of analytics and AI, and with its wide range of technology components and choices, presents significant challenges to a commercial enterprise. Performance is still paramount—it isn't called "Medium Performance Computing"—but organizations still need to consider their choices carefully before investing in any new technology or selecting an HPC vendor. Some of the primary considerations are:

- *Quality and reliability:* How quickly can a system be put to productive use, and how likely is it to stay online and working once up and running?
- *Support:* When there are problems, how quickly can a supplier respond to get things back up and running?
- *Manufacturing and delivery:* Does the supplier have access to the latest technologies and the ability to deliver within the required timeframe?
- *Experience in with relevant technologies and workloads:* Can the supplier confidently recommend solutions for a given set of workloads, based on experience with the possible technology choices?
- *Security:* How well will the supplier protect my sensitive data and intellectual property?
- *Acquisition and operating costs:* How much will the solution cost to acquire, and how much will it continue to cost after it is installed?

These questions can be critical, because the HPC environment continues to change. While the application of HPC leads to new insights and innovations, users still are leery of picking a technology or vendor that leaves them on the wrong tack, needing to beat back upwind as the competition sails on. The more complicated the new HPC gets, the more a solid HPC partner can be viewed as a necessity to taking full advantage of it.

LENOVO AND THE NEW HPC

Increasing Momentum

Lenovo has quickly established itself as a top-tier HPC provider. Since late 2014, when Lenovo acquired IBM's x86 server line of business, Lenovo has steadily increased its presence with HPC end users. In 2017, Lenovo introduced a completely revamped product line branded ThinkSystem, based on the new Intel Xeon Scalable Platform. ThinkSystem garnered positive reviews upon its introduction[6,7], and Lenovo also claimed 42 world-record performance benchmarks at the time of the ThinkSystem launch[8].

Most importantly, Lenovo used these developments as springboards to gain traction with HPC-using organizations. On an annual basis, Intersect360 Research surveys members of the HPC user community, spanning industry, government, and academia. Over the past three surveys, Lenovo has increased its share of named systems vendors for HPC more than any other vendor. (See figure below.)[9] At the high end of the market, Lenovo was also the only vendor to increase its share of named, surveyed systems with over 1,000 nodes in each of the last two HPC User Site Census surveys.

---

6 Forbes, "Lenovo Moves the Datacenter Ball Forward at NYC Transform Event," by Patrick Moorhead, June 26, 2017, https://www.forbes.com/sites/patrickmoorhead/2017/06/26/lenovo-moves-the-datacenter-ball-forward-at-nyc-transform-event.

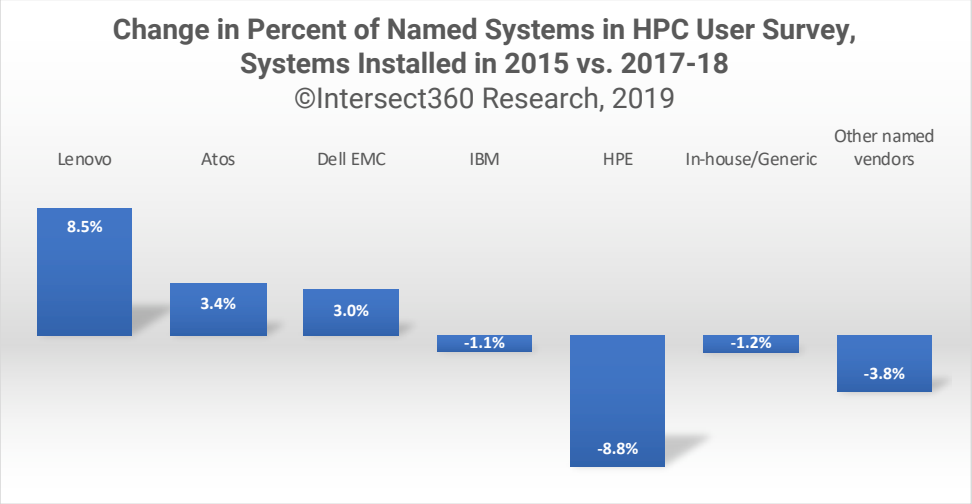
7 ZDNet, "Here's a Look at Lenovo's Master Plan for Grabbing Data Center Share," by Larry Dignan, June 22, 2017, http://www.zdnet.com/article/heres-a-look-at-lenovos-master-plan-for-grabbing-data-center-share.

8 https://news.lenovo.com/pressroom/press-releases/lenovo-launches-thinksystem-server-platform.htm.

9 Most recent survey conducted in the second quarter of 2018. This data compares systems installed in 2015 to those installed in 2017 or later.

Intersect360 Research
P.O. Box 60296 | Sunnyvale, CA 94088 | Tel. [888] 256-0124
www.Intersect360.com | info@Intersect360.com

This gain in share corresponds to strategic efforts at Lenovo focusing on supercomputing and AI. Lenovo has opened HPC and AI Innovation Centers in Stuttgart, Taipei, Beijing, and Morrisville, North Carolina, USA, and the company recently announced a partnership and joint venture with NetApp to provide high-performance storage solutions. In 2018, Lenovo mounted a campaign based on its server reliability and differentiation in manufacturing capabilities. In Intersect360 Research surveys, Lenovo-using organizations report spending an average of 4.2% of their HPC budgets on services and personnel for maintenance and repair, compared to the population-wide average of 4.8%, implying 12.5% savings on spending in those categories.



Lenovo Manufacturing and Support Scale

Much of Lenovo's current momentum comes from its worldwide manufacturing capabilities and scale. Lenovo was able to hit the ground running with the acquisition of an established HPC product line, deep engineering experience, and a broad partner community. Lenovo complemented those assets with its might in global manufacturing and services.

Lenovo is multinational company with over $45 billion in annual revenue and over 50,000 employees worldwide. The company has 36 manufacturing facilities at six sites around the world and claims it produces in excess of 100 servers every hour, giving Lenovo influence and reach with its suppliers. Lenovo owns and controls its own manufacturing, giving the company tight oversight of quality and security, with multiple levels of environmental, electromagnetic, and reliability testing. Lenovo cites independent end-user technology studies from ITIC and TBR, naming Lenovo the most reliable x86 server vendor, and the number-one vendor in client satisfaction on 19 out of 22 categories.[10]

Lenovo's client satisfaction score comes not only from reliability, but also from support. Lenovo has over 10,000 support personnel worldwide, with 51 major contact centers and support teams in 165 countries. Moreover, Lenovo's HPC and AI Innovation Centers are nexuses for advancement for Lenovo customers at the forefront of research and development.

Finally, Lenovo has endeavored to lead the field in security. Lenovo subjects itself to annual reviews by the Committee on Foreign Investment in the United States (CFIUS), with security assessments rotating among

[10] http://blog.lenovo.com/en/blog/lenovo-is-serious-about-reliability-and-ease-of-doing-business

Intersect360 Research
P.O. Box 60296 | Sunnyvale, CA 94088 | Tel. [888] 256-0124
www.Intersect360.com | info@Intersect360.com

three independent security firms. Lenovo runs a Trusted Supplier Program among its partners, with an auditable supply chain. And with its Morrisville and Hungary facilities, Lenovo can manufacture its systems entirely within the U.S. or Europe respectively, a potentially compelling point of differentiation for buyers.

INTERSECT360 RESEARCH ANALYSIS

When Lenovo acquired the x86 server business from IBM, some observers doubted whether Lenovo would be viewed as a credible, top-tier HPC vendor. Any doubts should be completely quelled now. While there are still certain HPC segments that have been slow to warm up (most notably, U.S. Government), Lenovo has nevertheless staked out a claim in the industry.

Lenovo has been making strategic moves to broaden its portfolio. Its embrace of AI has led the company to incorporate more processor options, including GPUs, into its server mix, and its partnership with NetApp affords far more options in deploying storage solutions matched to targeted workloads. Intersect360 Research believes the NetApp alliance to be one of the most significant developments in HPC in 2018.[11]

Lenovo's manufacturing muscle should not be underestimated. Intersect360 Research has visited the Lenovo plant in Shenzhen, China, where teams of line workers are given spiffy bonuses for finding quality and throughput improvements. These capabilities are world-class, and they have downstream benefits for customers who rely on trusted partners. Even the U.S. Government segment may eventually come around, with an authentic, auditable Made-in-Morrisville stamp available for HPC systems.

Ultimately, the new HPC brings about opportunities that are worth the challenges. AI is promising for more than just parlor tricks and the Turing test, but the biggest breakthroughs may be hidden inside broader innovation initiatives, where machine learning blends with AI and HPC to unlock another secret of the universe. That's what the new HPC is all about, and Lenovo will be there for it.

[11] https://soundcloud.com/this-week-in-hpc/episode-240-lenovo-and-netapp-partner-up-nvidia-puts-turing-into-tesla.

Intersect360 Research
P.O. Box 60296 | Sunnyvale, CA 94088 | Tel. [888] 256-0124
www.Intersect360.com | info@Intersect360.com