Intersect360 Research White Paper: AI-AUGMENTED HPC FROM DELL DRIVES CUSTOMER SUCCESSES

MARKET DYNAMICS

AI and the New HPC

Science and engineering rarely stand still. Once one problem is solved, it unlocks the approach to a new, harder category of problem, which itself spurs investment to solve it. Until we reach the end of science (not likely any time soon), there will be more scientific discoveries to be made and new engineering enhancements to be enabled.

As a result, High Performance Computing (HPC) is a long-term growth market, as it unlocks new insights and reduces time to discovery in a wide range of domains, such as manufacturing, energy, electronics, and pharmaceuticals. Even during the COVID-19 pandemic, which caused worldwide economic disruption, major segments of HPC grew, as computational bio-sciences applications flourished, in understanding the virus and its spread and in pursuit of treatments and vaccines. Worldwide spending on HPC is poised to exceed $60 billion worldwide in 2025.[1]

The unending reinvention of scientific discovery demands new tools to enable it. New HPC applications, architectures, and approaches are driven by the perpetual need to solve the next big problem. But even in an arena in which perpetual reinvention and the pursuit of new horizons is the norm, we are now in the early stages of revolutionary advancement, thanks to the rapid maturation of a new tool: artificial intelligence (AI). Thanks to advancements in the availability of data, the scale of HPC, and investment by hyperscale companies, there has been a revolution in using AI—or more specifically, machine learning—as a complement or component of HPC.

Machine learning can be deployed any time there is a wealth of data to draw on, coupled with a reward from making more intelligent inferences based on that data. Most domains within science and engineering fit this pattern. HPC-using organizations in general are trending toward the incorporation of AI. In a 2021 survey, 81% of HPC users said they are already running machine learning as part of their environments or planning to implement machine learning within a year.[2]

[1] Intersect360 Research, *Worldwide HPC Market Model and Forecast*, 2021.

[2] Intersect360 Research, HPC User Budget Map survey data, 2021.

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

AI-Augmented HPC

While AI can be pursued independently from traditional HPC approaches, more commonly it is deployed in conjunction with deterministic scientific applications. HPC is in itself a key driver for AI, as certain aspects of machine learning, particularly the neural networks deployed for AI training, lean on high-performance architectures. Beyond HPC enabling AI, organizations are finding they can use AI to augment HPC modeling and simulation.

Traditionally, most HPC applications have been *deterministic*; given a set of inputs, the computer program performs calculations to determine an answer. Machine learning represents another type of applications that is *experiential*; the application makes predictions about new or current data based on patterns seen in the past. We call this "artificial intelligence" because it approximates how we learn as humans. After all, a child does not become adept at catching a ball by learning advanced physics and mathematics, but through repetition and experience, learning how the ball behaves in most situations, and applying past knowledge to new situations.

Both approaches have a future in HPC. Predicting the path of a storm, for example, can be done both by modeling weather conditions and by examining the behavior of past storms. The ability to compare the results of two disparate approaches already advances the state of the art in scientific computing.

Where things really have the potential to take off is in the combination of techniques, with AI-augmented HPC. Humans have been directing areas of HPC computation and inquiry for years. With AI, there is the opportunity to expand the reach of experimentation and discovery. Consider: If AI can be fed the rules of a game such a poker or go and find unexpected winning strategies, could it also be taught to play different "games," such as Discover the Drug or Optimize the Airplane Wing?

One such role for AI is in target reduction. If a pharmaceutical company, for example, has hundreds of thousands of potential target molecules for interacting with a protein in solution, testing them all would be a daunting computational task (and impossible to do in a wet lab in practice). Machine learning algorithms could potentially strip the field down to only hundreds that are the likeliest, based on a wealth of data reflecting past experience. HPC simulations can do in-depth analysis of the rest, yielding perhaps a dozen most promising candidates for synthesis and physical testing.

Still, target reduction represents AI and HPC essentially side-by-side. Another more integrated example of AI-augmented HPC is computational steering, using machine learning to guide ongoing simulations. In manufacturing, for example, multi-physics models might simultaneously examine a car design's crashworthiness (with finite element analysis or FEA applications), aerodynamics (computational fluid dynamics, CFD), and interior noise (noise-vibration-harshness, NVH). Engineers can make design changes to optimize any of these

With the power of AI-augmented HPC, the universe of potential solutions expands.

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

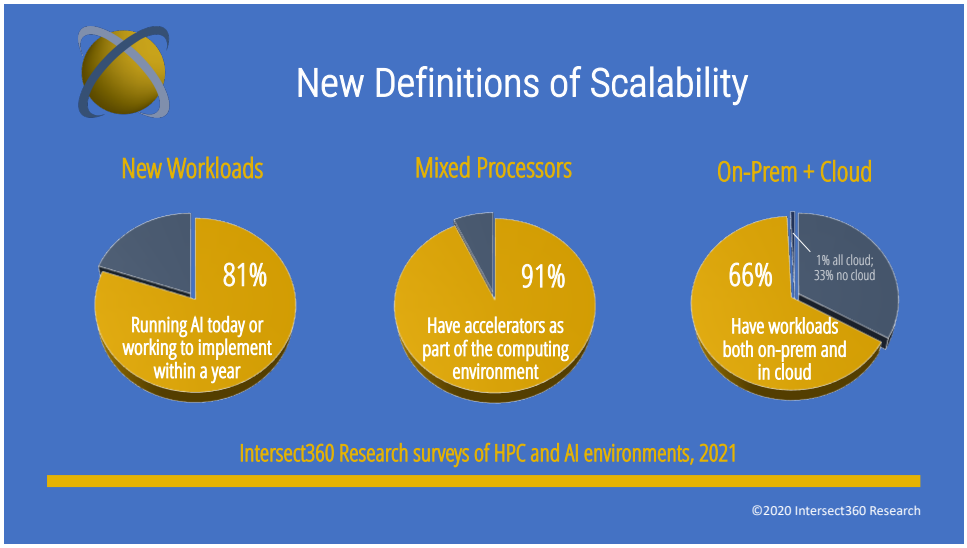
dimensions, but the human latency in looking at all three is high. Given a set of rules and targets, AI could be left to play, exploring iteration after iteration of designs, and perhaps finding potential design elements the human engineer might have missed. The HPC is still needed for validation, and the human engineer still needs to be in charge, but with the power of AI-augmented HPC, the universe of potential solutions expands.

New Dimensions of Scalability

The drive to incorporate new workloads has expanded the traditional notion of a fundamental HPC concept: scalability. Traditionally, scalability in HPC has essentially meant *more*. More cores, more flops, more memory, more bandwidth, more throughput: if you can make a bigger computer, that's scalability.

As workloads expand to include more workloads, today's HPC environments are uncovering new dimensions of scalability. To meet the specific demands of various applications, heterogeneous computing is becoming commonplace. 91% of users incorporate accelerators into some portion of their HPC environments. HPC sites often find they need to incorporate a wide range of processing capabilities to serve expanding workloads.

Furthermore, no discussion of scalable computing would be complete without mention of cloud computing, which has rapidly become a viable alternative to on-premises infrastructure. Our most recent surveys indicate that over 60% of academic HPC sites make use of cloud computing, but typically for only a small portion of their overall HPC workloads. Most HPC sites find it advantageous to run on-prem when they can, for reasons of cost and data locality, bursting to cloud only for extra scalability, capability, or capacity when needed. These hybrid-cloud HPC environments bring new scalability challenges, such as managing data sovereignty, availability, and optimization across the total ecosystem.



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

INTERSECT360 RESEARCH ANALYSIS

Today's HPC environments need to scale in new ways, combining multiple technology elements to merge scientific computing with large-scale data analytics and artificial intelligence. The potential of AI-augmented HPC means finding HPC solutions that are flexibly capable of multiple approaches. Successful solution providers will be those that have technologies that can span this expanded environment, combined with domain-specific expertise in helping organizations achieve meaningful breakthroughs.

This is the role of Dell Technologies. With trusted products across both computation and data management, Dell Technologies is the industry leader in total HPC solution revenue[3] (See chart below), and Dell was the most-cited vendor in a survey of HPC users in 2021.[4] Dell Technologies leverages this breadth of offerings with converged solutions that incorporate HPC, data analytics, and AI[5] and offers tailored solutions incorporating the latest in AI for a wide range of scientific domains.[6]

Internally, Dell Technologies pursues advancements in HPC and AI through its HPC & AI Innovation Lab, where the company's engineers test and optimize new generations of technologies for processing, networking, and storage.[7] Globally, Dell Technologies also operates its HPC & AI Centers of Excellence, which showcase the latest solutions and provide community collaboration opportunities with the wider HPC community[8], and its Worldwide Customer Solution Centers, which offer remote access capabilities for testing and optimizing customer-specific workloads in collaboration with solution specialists.[9]

Most importantly, Dell Technologies is helping its customers derive real value from their investments in computation and machine learning, including the potential of AI-augmented HPC. Dell Technologies is well-positioned to do this, based on its breadth of offerings and its actions to broaden HPC access, such as through its Dell HPC Community[10] and direct communications from its HPC engineering teams.[11] This direct-access, domain-specific support is critical in empowering innovation for those users investing in their next breakthroughs.

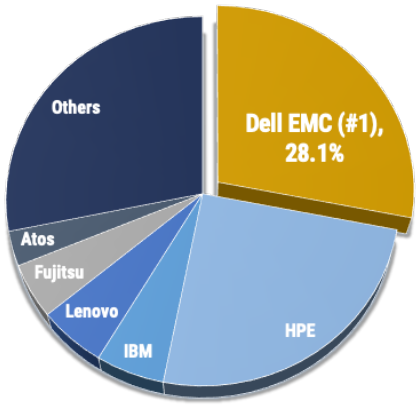
Dell Technologies is the industry leader in total HPC solution revenue, and Dell was the most-cited vendor in a survey of HPC users in 2021.

[3] 2020 HPC market share by revenue. Intersect360 Research HPC market model and forecast data, 2021.
[4] Intersect360 Research, HPC Technology Survey, http://www.intersect360.com/LiteratureRetrieve.aspx?ID=159114
[5] https://www.dellemc.com/en-us/collaterals/unauth/brochures/solutions/hpc_ai_convergence_brochure.pdf.
[6] https://www.delltechnologies.com/en-us/solutions/data-analytics/machine-learning/ready-solutions-for-ai.htm.
[7] http://delltechnologies.com/innovationlab.
[8] http://delltechnologies.com/coe.
[9] http://delltechnologies.com/csc.
[10] http://www.dellhpc.org.
[11] http://www.hpcatdell.com.

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

Combined HPC Server and Storage Revenue Share, 2020

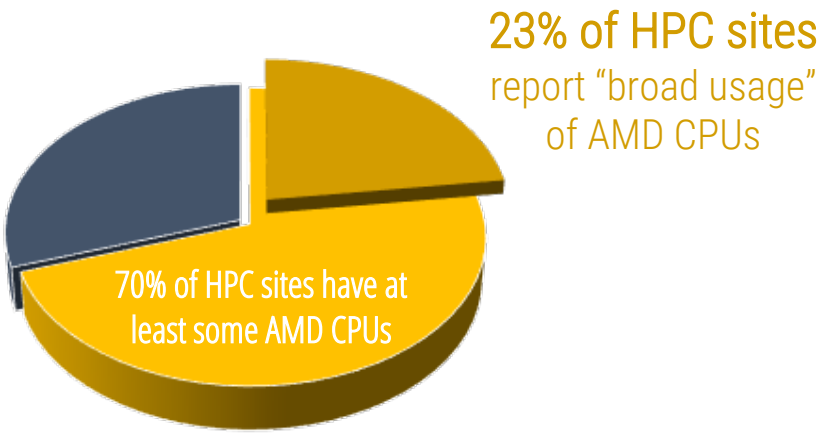
Intersect360 Research, 2021



Dell Technologies provides systems, solutions, and support for AI-augmented HPC, including servers incorporating the newest processing technologies from AMD. With its third-generation AMD EPYC™ processors and a steady drumbeat of HPC application benchmarks, AMD has strutted back to the forefront of many HPC users' minds. AMD EPYC processors are now in use at 70% of surveyed HPC sites, and in broad usage at 23%, representing a tremendous gain from previous years. (See chart below.)

Percentage of Users Using AMD CPUs as Part of HPC Environment [12]

Intersect360 Research, 2021



Furthermore, AMD is poised to become the first processor vendor to enter the HPC market with both CPUs and GPUs in an integrated package, connected by AMD Infinity Fabric™ technology. This combination promises to improve CPU-GPU communications, improving

AMD processors are now in use at 70% of surveyed HPC sites, and in broad usage at 23%, representing a tremendous gain from previous years.

[12] Intersect360 Research HPC Technology Survey, 2021.

performance for machine learning and other accelerated applications. In addition, direct memory addressing can simplify programming and porting of applications.

AI-Augmented HPC with Dell

As one example of Dell working with the community to expand the capabilities of AI-augmented HPC, Dell partnered with researchers from Carnegie Mellon University, Los Alamos National Laboratory, Massachusetts Institute of Technology (MIT), and the University of Massachusetts to explore how machine learning methods could optimize a CFD problem, to determine the optimal aerodynamic placement of lidar sensors on the outside of an autonomous driving vehicle. According to a white paper describing the partnership and its research:

> *To make a determination like this, design teams need to simulate the airflow around the vehicle, which is a computationally-intensive process with workloads that might run for an entire day in a high-end HPC system. While design teams would ideally run hundreds or even thousands of simulations to determine the optimal placement for a lidar array, the computational time requirements make this number of iterations highly impractical, if not impossible.*
>
> *To remove this barrier, and to enable faster fluid dynamics simulations and more design iterations, our team explored the use of machine learning techniques to approximate the air flow around a vehicle and calculate a drag coefficient without having to do direct simulation.* [13]

This partnership is an excellent example of Dell contributing not only domain-specific expertise, but specialized resources in its HPC & AI Innovation Lab. One of the systems in the lab that is available for AI-augmented HPC is Minerva, built in collaboration with AMD to include 150 Dell PowerEdge C6525 server nodes powered by second-generation AMD EPYC processors and over 2.8 Petabytes (combined) of Dell Ready Solution for HPC BeeGFS High Capacity Storage. According to Dell, "Engineering, partners and customers access the system for benchmarking, application characterizations, and solution optimization."[14]

Dell and AMD are also behind the new "Anvil" supercomputer at Purdue University, "a powerful new supercomputer that will provide advanced computing capabilities to support a wide range of computational and data-intensive research spanning from traditional high-performance computing to modern artificial intelligence applications," according to Purdue.[15] Funded by a grant from the National Science Foundation (NSF), Anvil will be part of the

---

[13] Dell Technologies, "Using Neural Networks to Quickly Approximate the Results of Computationally-Intensive Simulations," https://www.delltechnologies.com/asset/en-us/products/ready-solutions/industry-market/dell-neural-nets-computational-fluid-dynamics.pdf.

[14] Dell Technologies, HPC & AI Innovation Lab brochure, https://www.delltechnologies.com/asset/ru-ru/products/storage/briefs-summaries/h16221-hpc-lab-brochure.pdf.

[15] https://www.rcac.purdue.edu/anvil.

Extreme Science and Engineering Discovery Environment (XSEDE), furthering scientific discover at Purdue and across the wider research community.

In one more example, the 3.5 Petaflop Zaratan supercomputer at the University of Baltimore has over 50,000 computational cores, based on AMD EPYC processors. The system was configured to incorporate emerging and evolving technologies, to meet expanded use cases across HPC. According to Tripti Sinha, Assistant Vice President and Chief Technology Officer of the UMD Division of Information Technology, "We take the enablement of science using cutting-edge resources very seriously, and pushing the envelope of science and discovery is aided by computational resources and other prevailing technologies. That's our role, helping our researchers push the frontiers of science."[16]

The new HPC is fueled by analytics and AI. Across domains, there is a deep wealth of data. If harnessed, it can unlock new discoveries, which themselves will spawn new areas of research. While the fundamental drivers of research remain unchanged, these converged, high-performance solutions are critical to enabling new generations of insight. With its domain-specific knowledge and technology solutions across data management and computation for AI-augmented HPC, Dell Technologies is well-positioned to help is customers achieve new insights for the next generation.

To learn more, please visit delltechnologies.com/hpc.

[16] Dell Technologies customer case study, "Pushing the Frontiers of Science."

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