White Paper

To Out-compute is to Out-compete: Competitive Threats and Opportunities Relative to U.S. Government HPC Leadership

Sponsored by: Dell Technologies and AMD

Mark Nossokoff, Bob Sorensen, and Earl Joseph
April 2022

HYPERION RESEARCH OPINION

Nations and companies have learned that to be competitive in global innovation, advancing science and R&D requires serious computing capabilities. For more than three decades the US Council on Competitiveness has brought together a broad spectrum of constituents to champion an agenda that fuels entrepreneurship and spurs the commercialization of new ideas. It coined the term "To Out-compute is to Out-compete" as a way of describing the need for advanced computing capabilities. This paper takes that sentiment and examines U.S. government investment in high performance computing (HPC) across its various departments and agencies.

Investments in HPC are recognized globally as a fundamental tool for conducting R&D and advancing the economic competitiveness of nations. More countries than ever are increasing their spending on HPC infrastructure and critical associated areas, including AI, applications development and optimization, and workforce development and retention. Nations who do not adequately invest in HPC infrastructure and workforce development run the very real risk of weakening their national defense and falling behind economically.

National investments in HPC are not just academic exercises. Hyperion Research studies have shown return on investment is very high and can reach $507 dollars in sales revenues per dollar invested in HPC, and $47 dollars in profits or cost savings per dollar invested in dedicated strategic HPC activities.

Returns relative to national security go well beyond national military defense. Advances by medical researchers working in disease detection, prevention, and cures in the interest of public health require access to HPC infrastructure. Scientists working on digital twins of the earth to supplement existing climate and weather modeling to protect public welfare rely on advanced technical computing. Engineers performing complex modelling and simulation depend on advanced computing to develop more fuel-efficient and aerodynamic cars and planes.

Government leadership relative to investing in HPC infrastructure is critical. While the US government is making large investments in leadership-class machines (e.g., Frontier at ORNL, El Capitan at LLNL, and Aurora at ANL), it is at risk of falling behind in maintaining and extending its HPC investments to address the next tier of divisional and departmental requirements. Through increased investments in HPC, many U.S. government departments and agencies can bolster their missions and deliver outcomes benefiting all. Failure to make these investments not only jeopardize missions, but could compromise the defense, health, and overall well-being of the entire nation.

Note: This page is intentionally blank.

EXECUTIVE SUMMARY

Why Nations are Investing More in HPC

High performance computing (oftentimes referred to as supercomputing) has made enormous contributions to national and global scientific, engineering, and industrial competitiveness, as well as to homeland security and other government missions, since its introduction in the 1960s.

Supercomputers have played crucial roles in numerous government agencies and departments. Their impact, however, goes far beyond that. Supercomputers have made cars and planes much safer, more fuel efficient and environmentally friendly. They are crucial aids in discovering and extracting new sources of oil and gas, and for developing alternative energy sources. They have enabled the weather community to create more accurate predictions of severe storms that can devastate lives and property. Industries ranging from financial services to medicine and health care, entertainment, consumer products, and more recently by Internet companies and cloud service providers rely heavily on HPC. They have helped mankind explore space and understand how the universe was formed.

The ROI from Investing More Heavily in HPC and R&D is Very Strong

HPC-based computer simulation has become a fundamental driver of scientific discovery in many disciplines and is often referred to as the "third pillar" of scientific discovery, complementing traditional theory and experimentation. Supercomputers are a tool for researching scientific areas in ways that were previously impossible to pursue. HPC typically allows for dramatically faster time-to-solution and time-to-discovery. Scientific and industrial organizations often find it difficult to quantify their returns from using HPC, yet some things are clear:

- A growing number of Nobel laureates have relied heavily on HPC for their achievements.
- In academia, HPC use has spread from its established strongholds in the physical sciences to the social sciences and the humanities.
- HPC use has repeatedly saved lives and property by predicting severe storms.
- In the automotive and aerospace industries, HPC has dramatically reduced the time-to-market and increased the safety and reliability of new vehicle designs.
- Some large industrial firms have cited savings of $50 billion or more from HPC usage.
- HPC has allowed the modeling of phenomena that are impossible or undesirable to test (e.g., nuclear accidents).

Hyperion Research has been tracking the ROI from investments in HPC since 2013 as part of a US Department of Energy (DOE) grant. The industrial returns from successfully applying HPC to industrial problems in this study have been large:

- Industrial firms have seen sales revenue returns of $452 for each dollar invested in HPC.
- Finance, manufacturing, life sciences, and transportation realized:
  - $504 sales revenue returns for each dollar invested in HPC.
  - $38 in profits or cost savings for each dollar invested in HPC.

Details on the study, including input data, summary slide deck, success stories and the full report can be found on the HPC User Forum website (www.hpcuserforum.com).

The U.S. HPC Sector: Accelerated by U.S. Government Involvement

For more than fifty years, the United States has led the world in the development, supply, and use of HPC. This capability was due in large part to a comprehensive program of sustained federal government investments in HPC-related R&D, targeted procurements of systems at the highest level of performance, and perhaps most important, a broad and continued partnership with U.S. HPC commercial vendors. The U.S. Government (USG) recognized early on that to meet its broad range of important science, engineering, and defense applications, it had to help foster a robust, secure, domestic supply of the most powerful systems in the world. The successful partnership between the U.S. government R&D community and its domestic HPC supply base has served as a model that has been replicated, albeit to a somewhat lesser extent, throughout the world. While this has served the U.S. government well relative to leadership class machines, there's more work to do.

Beyond HPC Leadership Class Machines

Leadership class machines receive a large amount of media attention due to the increasingly large price tags and impressive benchmarks they can deliver. However, many of the challenges scientists and engineers are faced with today don't require the full capabilities of leadership class machines (those that cost over $500K). Divisional (costing $250K to $500K) and departmental (costing $100K to $250K) class machines can often satisfy the demands of a broad array of scientific, engineering, AI, and (high performance data analysis) HPDA workloads. According to recent Hyperion Research studies, the divisional and departmental class machines comprised 43.7% of the market in 2021.

Requirements, designs, and procurements for leadership-class machines are driven largely by the DOE and several NSF-funded academic sites (e.g., TACC). Many other US government departments and agencies are constantly examining their current HPC infrastructure and future HPC resource needs to ensure they can continue to properly support the increasing demands workloads are placing on the systems. A small subset of these departments and agencies include:

- Department of Defense (DOD)
- Department of Homeland Security (DHS)
- Federal Bureau of Investigation (FBI)
- Intelligence Advanced Research Projects Activity (IARPA)
- National Aeronautics and Space Administration (NASA)
- National Institutes of Health (NIH)
- National Oceanic and Atmospheric Administration (NOAA)

Rapidly Evolving Innovations and Heterogeneity in HPC Ecosystem

Suppliers are responding to HPC users' increasingly demanding and complex requirements with a growing array of novel innovations across the ecosystem, including:

- New types of CPUs and general-purpose processors (e.g., AMD EPYC)
- Accelerators and special-purpose processors (e.g., AMD Instinct)
- Memory and storage convergence
- System design divergence to support a broader set of applications
- HPC resource access models

Increased investments in current technologies and innovations are required for departments and agencies to deliver on their missions. Those that fail to do so run the risk of failing to fully fulfill their missions and deliver on their committed outcomes.

HPC SOLUTIONS AND DELL TECHNOLOGIES

Navigating today's HPC ecosystem is no small task. A proper balance of performance, scale, applications, and support is required to provide users and their data center managers the appropriate infrastructure to satisfy their needs. Users need to partner with vendors with a diverse range of products, services, and support offerings who deeply understand user requirements.

With several systems in the Top20 of the Top500 list, Dell Technologies in an example of an HPC supplier that takes lessons-learned from leading HPC systems, applies them to solutions with broader HPC market appeal, and augments them with corresponding support and choices of access methods:

- **Dell PowerEdge Servers:** With a broad array of servers from which to choose, including the R6515 and R7515 based on the 3rd Gen AMD EPYC processor, Dell has an HPC node to satisfy whatever scale users' workloads demand.
- **Dell Technologies on Demand:** The cloud is increasingly being adopted as an HPC resource incremental to on-premises infrastructure. Dell partners with leading CSPs to offer HPC cloud-based solutions to support customers with cloud-native, hybrid-cloud, and multi-cloud applications.
- **HPC & AI Innovation Lab:** This team of engineers and subject matter experts collaborates with customers and partners to move beyond individual products and develop targeted solutions HPC & AI workloads. The Lab is available for customers to evaluate new technology or develop focused solutions for a specific outcome, or virtually via access on-line resources for best practices and benchmark results.
- **Customer Solution Centers:** Resourced with Dell personnel, these centers provide customer and partners free hands-on access to Dell infrastructure and the opportunity to interact directly with Dell for demos and testing before buying. Interaction with the HPC & AI Innovation Lab for advanced solution engineering and performance testing is also available through these centers.
- **HPC & AI Centers of Excellence:** With almost a dozen locations around the world, these third-party centers develop and maintain local partnerships, test new technologies, share best practices and function as entry-points for customers to provide feedback and influence future product roadmaps.
- **Dell HPC Community:** Pre-COVID-19, Dell facilitated several in-person gatherings throughout the year for worldwide community networking and collaboration. Successfully evolving this to an on-line virtual activity, the Dell HPC Community event is a vibrant weekly gathering led by a combination of industry subject matter experts and Dell HPC experts to provide insight and education across a wide variety of HPC topics.

SITUATIONAL ANALYSIS

The U.S. HPC Sector: Accelerated by U.S. Government Involvement

Key U.S. policymakers charged with ensuring U.S. government (USG) access to world-class HPCs understand that fostering a domestic supply base targeted exclusively to USG HPC needs would not be enough. HPC policies would also need to help their domestic HPC supply base offer systems that have relevant and timely capabilities targeted for the overall base of HPC commercial end users. The advantages of this strategy for domestic HPC suppliers were widespread:

- U.S. HPC suppliers were better able to effectively compete against foreign HPC counterparts. From an economic perspective, that ability meant US HPC suppliers were able to capture the largest possible share of the global HPC markets. From a national defense perspective, it helped ensure that US national security research programs had unique and controlled access to HPCs with performance superior to counterpart programs in adversarial nations.
- U.S. HPC suppliers, working off the guidance, experience, and in many cases specialized funding of advanced US government HPC users, were able to explore new technologies, in both hardware and software, that could also be incorporated in their next generation commercial offerings. This process advanced the state of the commercially available U.S. HPC systems while helping mitigate U.S. HPC suppliers' costs for new technology developments.
- U.S. government HPC programs, which often supported close public/private partnerships to facilitate the design and operation of their procured HPCs, helped grow a strong base of HPC experts within the US HPC commercial base as both ideas and personnel could flow between the public and private sector elements.

However, USG attention to the development of HPC had a positive impact well beyond that of the HPC sector and generated larger benefits across both broader national policy agendas as well as into the commercial space.

- US government programs often took the lead in driving HPC-based research in key areas of public good that initially had no direct commercial financial benefits, but that contributed to the general economic or societal status of both the domestic and sometimes global population base. Areas here include public health, climate change research, and multiple areas associated with sustainability and energy conservation. In addition, as many of these developments matured, commercial entities entered the space seeking to address the market demands these new technologies engendered.

Likewise, USG promotion of a world-class HPC supplier base ensured that a wide range of US commercial entities spread across verticals including automotive, aerospace, bio sciences, finance, and energy has access to world-class HPCs. The advantages of such access enabled innovation in both products and services development to the US industrial base within almost every major commercial vertical.

A Broad and Diverse Range of U.S. Government Involvement in HPC

Due to their sheer size, budget, and diversity in missions, the roles that individual US government agencies play in the development, procurement, and use of HPC varies widely. For example, in 2016, the National Strategic Computing Initiative, identified ten leading US government agencies and their specific HPC-related responsibilities based on their individual missions and capabilities.

The Department of Defense (DOD), Department of Energy (DOE), and National Science Foundation (NSF) were charged with the development and delivery of advanced integrated hardware and software capability, engaging in mutually supportive R&D on hardware and software, and developing the HPC workforce. HPC activities in these departments are on-going and continue to push the state of the art in HPC design, production, and use.

- The DOD has a broad research agenda targeting some of the most advanced national defense-related research priorities building on capabilities in structural mechanics, combustion and propulsion, computational chemistry, computational electromagnetics and acoustics, and integrated modeling and testing. DOD currently lists eleven HPC systems of varying sizes spread across six different HPC centers with announced plans for multiple, additional new HPC procurements in 2022 alone.
- DOE has assumed not only a US but a global role in developing the highest performant HPCs. It currently houses two of the fastest US-based HPCs with plans to introduce three additional exascale-class systems in the next two years. DOE mission responsibilities for HPC are divvied up among its six main research labs: three dedicated to ensuring the United States maintains a safe, secure, and reliable nuclear stockpile, prevents nuclear weapon proliferation, and reduces the threat of nuclear and radiological terrorism around the world; and three are dedicated to multidisciplinary research in areas including advanced and sustainable energy, AI and machine learning, genomics, HPCs, large-scale scientific instrumentation, and quantum information science.
- The NSF currently supports a wide range of HPC resources for academic research at facilities including the Pittsburg Supercomputing Center, the San Diego Supercomputing Center, the National Center for Supercomputing Applications, and the Texas Advanced Computing Center, home to the new LoneStar6 and Frontera HPC, which premiered at number five on the June 2019 Top 500 HPC as the fastest HPC system installed at an academic site anywhere in the world.

The Intelligence Advanced Research Projects Activity (IARPA) and National Institute of Standards and Technology (NIST) were charged with fundamental scientific discovery work and associated advances in engineering. For its part, IARPA invests in high-risk, high-payoff research programs to tackle some of the most difficult challenges of the agencies and disciplines in the Intelligence Community and includes research into AI, quantum computing, machine learning and synthetic biology. Recent NIST activities center on HPC research that span both leadership-class HPCs as well as more conventional production clustered systems.

The Department of Homeland Security (DHS), Federal Bureau of Investigation (FBI), National Aeronautics and Space Administration (NASA), National Institutes of Health (NIH), and National Oceanic and Atmospheric Administration (NOAA) were charged with the development of mission-based requirements to influence the early design stages of new HPC systems and seek viewpoints from the private sector and academia on HPC hardware, architectural, and software requirements.

While DHS and the FBI concentrate more on targeted HPC applications to meet their mission requirements, organizations such as NASA, NIH, and NOAA are actively involved in a broad range of HPC-enabled research and development and/or procurements of HPC systems on a regular basis.

NASA's High-End Computing (HEC) Program plans and provisions high-end computing systems and services to support NASA's mission needs. The program uses integrated management at two main centers, the NASA Advanced Supercomputing (NAS) Facility at Ames Research Center and the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center.

- The NAS facility at NASA Ames currently has three world-class supercomputers: Pleiades, currently capable of 7.09 petaflops of peak performance; Electra, an innovative modular supercomputing system currently capable of 8.32 petaflops peak; and Aitken, the newest modular system, capable of 10.76 petaflops peak.
- The NCCS at NASA hosts the Discover supercomputer, which links tens of thousands of processors to achieve eight petaflops peak performance. Discover-hosted simulations can span time scales from days (weather prediction) to seasons and years (short-term climate prediction) to decades and centuries (climate change projection).

The NIH HPC group plans, manages, and supports high performance computing systems specifically for use by the intramural NIH community. These systems include Biowulf, a 105,000+ processor Linux cluster; Helix, an interactive system for file transfer and management, and Helix web, which provides a number of web-based scientific tools. The group provides access to a wide range of computational applications for genomics, molecular and structural biology, mathematical and graphical analysis, image analysis, and other scientific fields.

NOAA HPC programs improve the accuracy and timeliness of short-term weather warnings, forecasts, and regional and global climate and ecosystem predictions. NOAA uses a mix of NOAA R&D systems for its internal research activities and an external R&D system that provides resources for scientific research and greater opportunities for collaboration with the academic community. NOAA currently hosts seven different HPC platforms distributed across its national network including the internal R&D Gaea system at Oak Ridge National Lab and the external R&D Orion system at Mississippi State University.

New Challenges for USG HPC Procurements

With the profusion of new HPC-based hardware, software, architectures, access models, and related technology developments across the IT space, both existing and aspiring HPC users, in the academic, commercial, and government sector face an increasingly daunting task in selecting, procuring, and operating a productive and cost effective HPC infrastructure.

This latest round of HPC reinvention arises from the confluence of a number of key trends in the sector that includes new developments in both general and special purpose processors, new HPC architectures optimized for a range of unique, sometimes clashing workloads, and changing programming and access models used to efficiently bring the appropriate HPC resources to bear on a range of computational problems.

Diverging HPC Architectures and the Workloads They Support

Until recently, HPC use centered on its ability to perform large numbers of mathematical calculations per second typical in the bulk of the modeling and simulation applications integral to both commercial and national security applications. Some of the largest HPCs currently available can deliver the results of over 10 to the 18th mathematical results per second, and the history of the sector has been one that focused on delivering more pure computational capability each year. Moreover, the performance of the world's fastest HPC increased by almost 70,000x in the past 20 years.

The continued expansion of HPCs bound for the traditional modeling and simulation environment remains critical as 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. However, a number of new workloads have emerged in the HPC space, presenting new challenges and new opportunities for both existing and new HPC suppliers and users in determining the optimal HPC configuration:

- High performance data analytics is a growing field that requires the special storage and interconnect capabilities of HPCs to effectively process large and often diverse data sets that may include voice, text, image, and instrumentation outputs to generate new insight appropriate to a wide range of sectors including medicine, finance, transportation, and manufacturing.
- Likewise, the rapidly growing field of artificial intelligence that includes both machine learning and deep learning, is creating new demands for HPC with its own application- specific technical requirements that include mixed precision and integer math.

The convergence of these three pillars of HPC is redefining the emphasis and diversity of HPC architectures required for at least the next three to five years. HPC architectures will necessarily become more diverse and heterogenous to support this wider range of computational demands, while users and vendors will need to work together to accurately characterize the type and scope of the workloads to ensure the system they procure is well suited to its operational environment.

HPC Memory/Storage: The New Performance Determinant

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. Almost all processors today and going forward realize their increasing rates of computational performance by adding additional parallel computational units that require greater data access capabilities to run a significant percentage of 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 GPUs or other custom accelerators that perform a limited class of specific, typically data-intensive operations and require sophisticated storage hierarchies to ensure that an HPC's overall computational capabilities are fully utilized.

Changing HPC Access Models

The HPC sector is in the midst of a watershed moment in providing user access to HPC resources, shifting from a predominately user-owned system model to one that offers options for cloud-based HPC access. For most of its existence, the HPC makers delivered their systems to user-owned and operated facilities, so-called on-premises sites. In this model, HPCs are bought or leased, and the additional costs of ownership such as power, cooling and other facility overhead, are borne by the user. Here, HPC owners need to balance the provisioning of their systems with diverse user demands for system access and resources, which could vary over time, sometimes leaving an HPC either idle or overscheduled.

Within the last few years, however, HPC workloads running in cloud-based environments have begun to proliferate as HPC on-premises user sites sought to distribute their workloads to the most cost effective, efficient, and performant hardware/software platform base. Cloud providers are able to offer a broad range of HPC resources on a pay-as-you-go basis, with the promise of reduced overall cost of ownership for many on-premises sites.

FUTURE OUTLOOK

Innovations and advancements across the entire HPC ecosystem continue accelerate. Users across the entire HPC ecosystem indicate the following reasons that are motivating their next generation HPC procurements:

- 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. In many cases, HPC in the cloud is being used not as a replacement scheme but instead to augment critical on-premises HPCs capabilities.
- There is an 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 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.
- There will be a continued expansion of HPCs into the traditional modeling and simulation environment as more commercial and government users, including small and medium-size organizations, 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.

U.S. government departments and agencies also need to invest in the most current HPC innovations. Failure to do so could jeopardize their ability to fulfill their missions and mute the impact and benefit of the science and research their users are attempting to deliver. The risks of under investment could then extend to decreased national economic competitiveness and sub-optimal national security.

Establishing relationships with key HPC partners will be critical for US government departments and agencies in providing their scientists, researchers, and engineers the resources that allow them to optimally deliver results. Partners should have a long track record of success within the HPC space while providing a wide range of solutions and services for users to leverage. Dell Technologies is one example of a potential partner. With multiple systems in the top of the Top500 list, a broad product line, on-premises and cloud-based solutions and support, and strong field support teams and resources, Dell Technologies has supported some of the most demanding HPC users in delivering their research and scientific breakthroughs and engineering achievements.

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 & vendor analysis for multi-user technical server technology used for HPC and HPDA (high performance data analysis). Hyperion Research provides 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.HyperionResearch.com and www.hpcuserforum.com
