White Paper

HPC Investments Bring High Returns

Sponsored by: Dell Technologies and Intel Corporation

Alex Norton and Earl Joseph
July 2020

HYPERION RESEARCH OPINION

HPC has been instrumental in scientific discoveries, product development, and furthering research in corporations, governments, and academic institutions for decades. Although HPC is sometimes considered synonymous with large government lab and academic systems, HPC is also crucial in product development and testing for automotive companies, aerospace companies, and financial institutions, to name just a few. Nevertheless, it can be challenging to fully grasp the value of the investments in HPC. The Department of Energy Office of Science recognized that challenge years ago and commissioned a study to create a methodology to measure the financial return on investment or return on research.

Based on over 400 HPC projects, spanning the areas of product development and research, the DOE study revealed that the return on investment and return on research are both very high for investments in HPC technologies. On the financial return side, organizations saw an average ROI of $44 of profit for every dollar invested in HPC. The data on the return on research further reinforced the high impact HPC has on the world.

HPC has evolved to include traditional workloads such as computational fluid dynamics, weather modeling, and large scale stochastic modeling for finance, as well as emergent AI application spaces. These AI applications require the scale and capability of HPC systems to produce their desired outcomes. The convergence of modeling and simulation applications with AI has driven the HPC-enabled AI market to boom at a 30% yearly growth rate, growing faster than any other sub-segment of the HPC market. AI has also become an integral consideration in future system design, leading to heterogeneous systems that can handle an array of applications.

Dell Technologies has years of experience helping HPC customers achieve the benefits of HPC systems for their applications and is one of the top HPC vendors in the global HPC market. Dell continues to offer consulting services as well as tailored solutions for HPC and AI applications. In addition, Dell has created numerous centers, including worldwide Customer Solution Centers, the HPC & AI Innovation Lab and Centers of Excellence for users to engage with subject matter experts to accelerate and enhance their deployment of HPC and AI applications free of charge. The assistance and resources are designed to enable users to realize the full potential of AI, HPC, and the convergence of the two, not to mention the financial return or scientific impact of their investments.

SITUATION OVERVIEW

Definitions for This Paper

For the purposes of this paper, the portion of the artificial intelligence (AI) market covered is HPC-enabled AI, including leading-edge R&D for AI. AI is more broadly defined as the ability of computers to perform complex tasks that otherwise require human intelligence. Within AI there are a number of subsegments, including machine learning, deep learning, and graph analysis.

All references to the cloud are referring to the public or external cloud. This does not include internal private clouds within an organization.

The Expanding HPC Ecosystem

The HPC ecosystem has seen an increased number of users as new application spaces have emerged and technology requirements for applications have shifted. Traditional HPC referenced the modeling and simulation workloads usually associated with large government labs, academic sites for research, and a few verticals of early HPC adoption such as aerospace and the energy sector. However, with the growth of big data, AI, and the need for larger-scale traditional modeling and simulation jobs, the HPC user base has expanded, and continues to expand. Some high growth areas include financial services, as well as emergent AI application spaces like automated driving systems and affinity marketing. These new sectors are adopting HPC technology to address their growing need for higher-performant systems to handle large data sets and scale-out their current applications. Although some users in these sectors have been users of HPC for many years, there is a movement of non-traditional HPC users to HPC technology.

ROI in HPC

Efforts to understand the impact and value of HPC and how that ties to economic and scientific leadership is a challenging yet valuable exercise. A few years ago, Hyperion Research, with the guidance and insights of the US Department of Energy, conducted a comprehensive investigation and analysis of the return on investment and return on research of HPC worldwide. This ROI (Return On Investment) and ROR (Return On Research) analysis was conducted to examine both the financial benefits as well as the impact of scientific discoveries.

Financial benefits were modeled based on revenues, cost-savings or profits, and jobs created as a result of HPC investments. Because it was considered impractical to calculate a financial return for many scientific innovations, a separate grading system, an innovation index, was established to rank, quantify, and understand the impact HPC has on projects that do not directly result in revenue, profits, or job creation. These innovation indexes were combined into an innovation "class" scale to rate the impact and value of scientific research achieved through HPC investments. This grading system takes into consideration the importance of furthering innovation and research outside of one organization and developing widely-used research tools and applications.

The financial ROI of HPC database consists of over 150 use cases worldwide that show an average revenue of $463 dollars per dollar of investment in HPC, as well as $44 of profit for every dollar of investment in HPC. To put that into context, for an HPC system purchased at $100,000 by a private corporation, the analysis estimates that the profit will be around $4.4 million.

Across the entire HPC ecosystem, the ROI of HPC is high, but sectors experience varying levels of financial return. As can be seen in the table below, which consists of data from the same ROI study,

the transportation sector reported the highest average revenue per dollar invested in HPC. But, the more impactful number in identifying the value of HPC is the profit or cost savings, where the finance and oil and gas sectors reported the two highest average profit or cost savings per dollar invested in HPC ($61 and $54 respectively). Thus, in finance use cases, a $100,000 system, considered to be part of the workgroup segment, could result in a profit or cost savings of more than $6 million.

The example of a workgroup system as a reference point for ROI is important. The most units shipped of the four price bands is in this lowest price band. The bulk of the HPC base is not comprised of the largest systems, although they may receive the most attention due to their price tag and public announcements.

TABLE 1

### Return on Investment of Select Industries

Industry	Average Revenue $ per HPC $	Average of Profit or Cost Saving $ per HPC $
Finance	$834	$61
Insurance	$71	-
Life Sciences	$160	$41
Manufacturing	$83	$20
Oil and Gas	$416	$54
Retail	$30	$12
Telecommunications	$211	$30
Transportation	$1,804	$16

Note: More details found in full Hyperion Research ROI Report

Source: Hyperion Research, 2020

FIGURE 1

The Innovation "Class" Scale
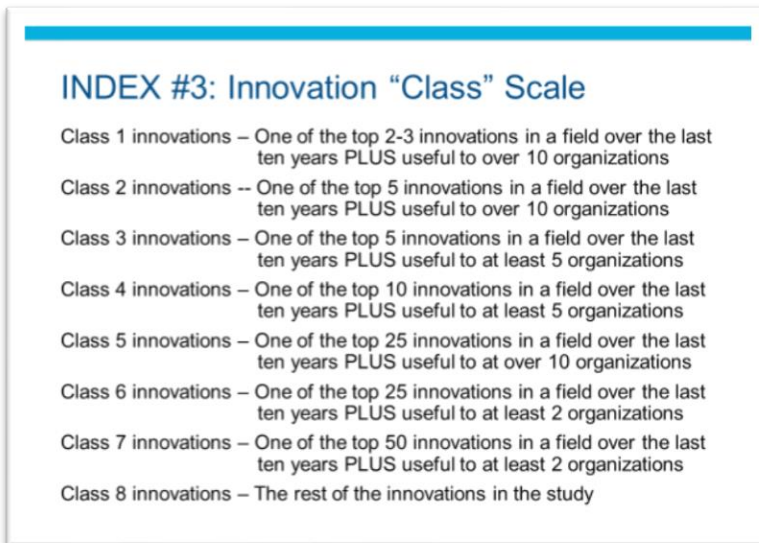
Source: Hyperion Research, 2020

For research projects geared at furthering science and not designed for financial returns, a separate grading system was created to judge the impact and value of those HPC investments. The Innovation "Class" Scale (Figure 1, above), takes into consideration the impact of an innovation on the specific field it is in, as well as the value it has to broader organizations. Overall, the research-oriented respondents in the study assigned a wide range of value to their work, leaning credence to a respondent's comparative contribution of any one effort to the specific field of interest. The data collected resulted in an average score of around class 3.6 on the innovation scale, with the most common responses falling into classes 3 and 4. Based on hundreds of surveys and interviews with researchers who did not expect a financial outcome of their HPC research, many see an extremely high value of HPC in research, with most experiencing top 5 to 10 innovations in the field, and relatively widespread usefulness of the discoveries.

HPDA/AI Growth and Convergence

High performance data analysis (HPDA), HPC-enabled artificial intelligence (AI), and all of the subsets including machine learning, deep learning, and graph analysis, have been growing at aggressive rates for a number of years now (Table 2). Compared with the larger HPC on-premise server market, which is growing at 8.7%, AI server growth typically exceeds 30% CAGR, approaching more than $3.5 billion of AI-dedicated HPC servers projected to be purchased in 2024.

Surveys and interviews with experts and end users conclude that a majority of the AI applications underway in the ecosystem today are machine learning based. A companion AI technology, deep learning, currently suffers from a lack of quality data necessary for training. Likewise, most deep learning programs cannot yet provide sufficient transparency, the ability to demonstrate why a specific outcome was generated, to drive widespread adoption. Deep learning experienced some missteps in use cases that have contributed to a negative perception, namely the inability to understand the decision making process conducted by the computer that led to a wrong decision.

TABLE 2

Forecast: Worldwide HPC-Based AI Server Revenues vs Total HPDA Server Revenues ($ Millions)

	2018	2019	2020	2021	2022	2023	2024	CAGR 19-24
WW HPC Server Revenues	13,683	13,713	14,484	15,658	18,457	19,940	20,844	8.7%
WW HPDA Server Revenues	3,153	3,598	3,932	4,737	5,467	6,480	7,478	15.8%
WW HPC-Based AI (ML, DL & Other) Server Revenues	747	918	1,094	1,399	1,810	2,745	3,555	31.1%

Source: Hyperion Research, 2020

In the above forecast, the HPC-based AI servers include dedicated AI servers only, but many servers deployed today and in the next few years will be required to handle both traditional modeling and simulation workloads as well as AI workloads. This is due to the growing heterogeneity of HPC workload portfolios, the convergence of traditional HPC workloads and the capabilities that integrating AI can accelerate or boost the modeling and simulation workloads. In the drive towards exascale, three pillars were established as requirements for future large systems: modeling and simulation, big data, and AI. Every system has to be able to handle all of those workloads, setting a precedent for much of the greater HPC ecosystem to work to incorporate HPDA and AI into their current workload portfolio.

The HPC-enabled AI forecast is further broken down (Table 3) into machine learning (ML), deep learning (DL), and other emerging methodologies, which is mainly comprised of graph analysis applications. ML applications comprise the largest portion of the AI-dedicated HPC server market, although deep learning growth (32.9% CAGR), a smaller component of the forecast, is growing more rapidly.

TABLE 3

Forecast: Worldwide ML, DL & Other AI HPC-Based Revenues ($ Millions)

	2018	2019	2020	2021	2022	2023	2024	CAGR 19-24
ML in HPC	532	667	771	986	1,285	1,960	2,538	30.6%
DL in HPC	177	209	265	342	443	665	866	32.9%
Other AI in HPC	38	42	58	70	83	120	150	29.0%
Total HPC-Enabled AI Server Revenue	747	918	1,094	1,399	1,810	2,745	3,555	31.1%

Source: Hyperion Research, 2020

As AI and HPC continue to converge in system design, highlighted by the rise of heterogeneous systems designed to handle both traditional modeling and simulation workloads as well as address emergent AI workloads, there is more need for tailored hardware and software solutions to support the diversity of users' workloads. The integration of AI methodologies into traditional modeling and simulation workloads further emphasizes the need for expertise and solutions geared toward optimizing the hardware and software to maximize the return on investment in the solution.

Where does HPC in the cloud fit with emerging HPC-enabled AI applications?

HPC in the cloud has been growing as well, stemming from early adoption by certain sectors of the HPC market as well as a concerted effort from CSPs and vendors to bring HPC expertise to the table. Furthermore, many HPC users see the cloud as a prime location to run their AI applications due to the variety of hardware and software solutions readily available, as well as the solutions that CSPs and vendors have created to aid users in adopting AI techniques.

Based on a number of studies of the AI market for HPC users, machine learning and deep learning workloads are the fastest growing set of applications in the cloud. Projecting out the next year and a half, Hyperion Research anticipates that 29% of HPC-based AI applications will be run in third party clouds, a higher portion than traditional modeling and simulation workloads.

DELL TECHNOLOGIES' READY SOLUTIONS FOR HPC AND AI

Dell Technologies has reacted to the dramatic rise in AI applications with hardware, software, and consulting solutions optimized to handle machine learning, deep learning and other AI applications for customers. The Dell EMC Ready Solutions for AI include the hardware stack (servers, networking, and storage), the software stack, and services, all geared towards helping customers address their AI needs with the best solutions possible. Among their hardware solutions are the Deep Learning with Intel solution, which is a near-bare-metal performance from a container-based platform, either on Kubernetes or Docker, and run on the Intel Xeon Scalable processors and the Nauta open-source platform. This lower cost, less complex platform allows users to scale their compute independent of the storage, and realize a more optimal balance between cost, capacity and performance. On top of the DL specific hardware solution, Dell provides their consulting, deployment and support services to help users maximize their investment and performance of their AI applications.

Not only are there solutions to allow users to take advantage of AI with little previous knowledge, Dell Technologies also provides HPC solutions for new entrants to the HPC ecosystem. There are many new sectors of the market in which users are being pushed to use HPC technology, such as the financial services sector, and Dell offers the ability to utilize HPC capabilities without previous experience in HPC on premises or on-demand.

Dell also offers an ML solution based on the Intel Xeon Gold scalable processors and Hadoop, in a partnership with Cloudera. This solution offers data science and framework optimization, as well as the ability to run both ML and DL applications on a single platform where the data is stored through Apache Spark. The ML with Hadoop solution provides a multiuser experience for collaboration between data scientists on AI applications.

In addition to the various solutions, services are also available to help achieve rapid adoption and optimization of AI environments. Dell Technologies Consulting offers ProConsult Advisory Services, which helps align an AI solution with the needs of the customer, considering their strategy, vision, and

ultimate end goal of incorporating AI. Dell also provides education services to teach and certify end users on ML, DL and other data science and analytics methodologies and techniques.

As more industrial or commercial customers begin to adopt and incorporate AI into their environments, expert advice, as well as testbeds for discerning best practices and optimal solutions, is paramount. Dell has established a network of sites tailored toward helping customers realize the full potential of their AI and HPC applications.

- Worldwide Customer Solution Centers help customers run proofs of concept, run workshops, and enable collaborations with IT experts to help speed-up implementation and reduce potential risks associated with new technology investments.
- The Dell Technologies HPC & AI Innovation Lab hosts a large datacenter for testing, as well as a dedicated group of engineers, computer scientists, and subject matter experts with years of HPC and AI experience to work with customers on optimizing their HPC and AI solutions.
- Worldwide HPC & AI Centers of Excellence deliver thought leadership, best practices, and testbeds for hardware and software.

Dell Technologies not only provides hardware and software platforms to run AI and HPC workloads at scale, but also a deep network of expertise to empower customers to maximize their investments and take full advantage of AI and HPC for their applications.

DELL TECHNOLOGIES CUSTOMER SUCCESS STORIES

Worldwide, Dell Technologies is the number one overall server vendor according to IDC and the number two HPC server vendor, according to Hyperion Research. It comes as no surprise, therefore, that Dell Technologies has many customer success stories. Highlighted below are four customers that exemplify the breadth of capabilities Dell Technologies provides for their customers. Dell's AI and HPC solutions span further than large academic and government HPC deployments and reach into industry to accelerate business applications requiring HPC and AI technologies.

Eni, An Energy Company

Eni, the Italian integrated energy company, with the assistance and deep expertise of Dell Technologies, deployed a 52 petaflop system, the HPC5, as a complementary system to their previously installed HPC4. The HPC5 comprises Dell EMC PowerEdge servers, each with Intel Gold processors, combined with accelerators, enabling a powerful, accelerated supercomputer - #6 on the TOP500 and the Green500. In addition, HPC5 uses an advanced parallel architecture and a hybrid programming model to take full advantage of Eni's internal algorithms for processing subsoil data.

Furthermore, Eni's HPC5 consumes energy produced by a large solar plant in Ferrera Erbognone, further accelerating the transition to non-fossil energy sources. Dell Technologies used their deep understanding of HPC, AI and system design to provide Eni with a cutting edge HPC system to tackle their hardest problems.

Mastercard, A Financial Company

Mastercard went to Dell Technologies with a problem that plagues the entire credit card world: fraud. A study conducted in 2018 by Experian estimates that more than $31 billion is lost each year due to credit card fraud. Mastercard turned to HPC to solve this problem. With the expertise of Dell Technologies, Mastercard implemented AI to combat the credit card fraud worldwide. For reference,

Mastercard processes 165 million transactions an hour, and thus the resulting HPC system had to be cutting edge and extremely powerful.

Mastercard leverages machine learning algorithms, <u>with almost 2 million rules</u>, to analyze credit card transactions and identify fraudulent uses. With each transaction, the algorithms analyze the details of the transaction against a plethora of characteristics of the individual card owner, including buying habits, geography, travel pattern, as well as comparing the transaction with other transactions on the same card.

Mastercard's solution to the fraud issue, termed MATCH (Mastercard Alert to Control High-risk Merchants), exploits the Cloudera Enterprise and Search that run on Dell EMC Ready Solutions for Hadoop. These HPC clusters, built on Intel Xeon Scalable processors, provide high performance, high speed, low latency capabilities for Mastercard's AI and HPC needs, delivering a powerful tool to fight credit card fraud. The resulting system allows for instant analysis of data to detect fraud while continuously training the model to improve the capabilities of recognizing user biometrics and fraudulent behaviors.

Zenuity, An Autonomous Vehicle Technology Company

Zenuity, a joint venture of Volvo Cars and Veoneer, is working to bring automated driving technology to the market in the form of a robust software stack. Zenuity needed an end to end solution to handle the variety of workloads, including machine learning and AI based applications, as well as performance-driven storage options to capture the variety of data types coming off the sensors in a car.

Dell Technologies is providing this HPC solution as a service, managing everything with Dell and Virtustream. With Zenuity allowing Dell Technologies to manage their HPC solution, Zenuity reduces their operational risk and can take full advantage of the expertise Dell brings to the relationship.

University of Florida

The University of Florida (UF) needed access to a cutting edge HPC system to address their computation needs for 1,300 diverse applications and more than 10,000 projects in areas such as medicine, agriculture, engineering and social sciences. They are also one of the leading universities in the nation in terms of creating and supporting startups.

UF Innovate comprises Tech Licensing, Ventures, and two business incubators, The Hub and Sid Martin Biotech. Together, those organizations move research discoveries from the laboratory to the market. UF Innovate connects innovators with entrepreneurs, investors and industry, incubates startups and growth companies, and fosters a resilient economy – all in an effort to make the world a better place.

In its latest 2.0 iteration, the evolving University of Florida HiPerGator system has a total of 46,000 cores and a 3-petabyte high performance Lustre storage system. The system has a theoretical maximum speed of 1.1 petaflops.

The University has implemented a platform called Open OnDemand. This platform allows access to the system through the web, enabling HiPerGator to be accessed the same way as a cloud.

FUTURE OUTLOOK

HPC, AI and the convergence of those two will drive business, innovation and research forward in the years to come. The relationship between the two is crucial to realizing the full potential of AI. We are just scratching the surface of what's possible with areas such as precision medicine, automated driving, fraud detection, and smart cities.

Dell Technologies addresses the evolving complexities of HPC and AI applications at all stages of a solution, from hardware implementation to application execution, services and consulting. Their position as the number one server vendor in the world, as well as decades of HPC experience put them in a position to provide high performance solutions for a wide breadth of applications, and the ability to expand their base of knowledge as the ecosystem evolves.

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). 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.HyperionResearch.com and www.hpcuserforum.com