PROWESS

DRIVE THE FUTURE OF THE FINANCIAL SERVICES INDUSTRY

Dell EMC™ Ready Solutions and Intel® technologies accelerate the evolution of HPC infrastructure to support AI workloads.

The traditional use of high-performance computing (HPC) has been for complex mathematical modeling and simulations, such as Monte Carlo simulations, in science and engineering as much as in financial services. Now, there is a new kind of workload coming to the forefront in a wide variety of business and industrial settings—artificial intelligence (AI)—which requires the same sorts of intensive computing resources as classic HPC workloads. The financial services industry is a leader in advancing from conventional HPC workloads to AI workloads to gain competitive business advantages. As the industry explores new AI use cases, it is also blazing a path toward a converged approach to HPC and AI.

How HPC Is Converging with AI

Until recently, the most compute-intensive AI processes, such as the training of machine learning (ML) and deep learning (DL) frameworks, were done on dedicated single-node hardware systems beefed up with multiple GPUs, rather than on the clusters of multiple nodes that typify HPC workload handling. Now, that's changing, as more AI work is moving to HPC infrastructures. It's a natural fit, because conventional workloads and AI workloads are complementary and converging.

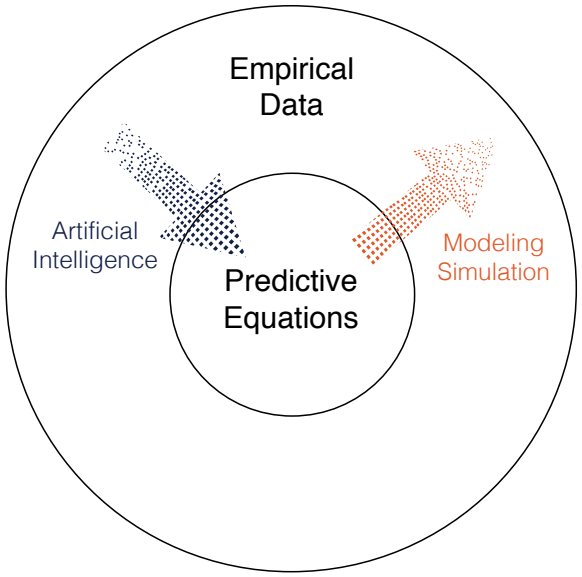

Figure 1. Inside-out HPC starts with equations, outside-in AI starts with data

Intel Fellow Pradeep Dubey describes conventional HPC workloads as mostly "inside out" computing, starting with human-generated equations and models and working to match observed data with high precision. AI workloads, on the other hand, are mostly "outside in"—starting with large amounts of data and working toward machine-generated predictive algorithms and inferences with lower precision requirements.[1] The software techniques differ, but the hardware requirements are similar: both require large amounts of computing power to deal with large amounts of data and analysis.

"HPC and the data-driven AI communities are converging as they are arguably running the same types of data and compute-intensive workloads on HPC hardware."

— Rob Farber, Intel[2]

There are two main reasons for the move away from specialized GPU-based systems for AI. First, the compute power of CPU-based processing clusters has expanded tremendously, and the ability of discrete CPU nodes to work in parallel creates a massive scalability advantage over single-node systems. This scalability can be critical for solving certain AI problems. But the second reason why organizations are moving toward HPC for their AI workloads is the efficiency and cost effectiveness of running AI alongside other workloads in a data center. Not only is it less complex than maintaining separate dedicated systems, but combining AI workloads with big data analytics and other workloads in the same HPC infrastructure allows for better resource utilization. Simply put, when HPC resources are not being used for DL workloads, they can be used for other kinds of work, which is much less true of dedicated GPU-based AI systems.

Financial Services, Data Analytics, and AI

The financial services industry has always shown leadership in the area of advanced data analytics for the purposes of customer service, security and fraud detection, and operational efficiency.[3] AI represents an opportunity to extend rather than replace data analytics. So, not surprisingly, banks, insurance companies, and other financial services companies have been pioneers in the early adoption of AI technologies. Banking alone is expected to spend $5.6 billion on AI-enabled solutions in 2019, according to IDC.[4] Some of the many use cases in which these companies see AI adding value include:

- Risk analysis, driven by increased regulations around controlling risk and understanding exposure
- Fraud detection that results in much better accuracy than older heuristic models
- Cyber security based on anomaly detection
- Trade optimization for algorithmic and high-frequency trading
- Credit risk assessment for mortgage and retail banking
- Marketing optimization by clustering or segmenting customer behavior to better serve and sell to groups identified as similar
- Automated customer service using chatbots and voice analytics to increase customer satisfaction and reduce churn

Among the many financial industry use cases, risk analysis is probably the most widely adopted for AI workloads. Traditional analytics are quite effective at evaluating risks based on structured data, but AI can achieve more with unstructured data—information that does not fit in a spreadsheet. AI-enabled processing of massive datasets takes place far more rapidly than purely human analysis ever could.

> "[T]he domain of risk management lends itself particularly well to cognitive computing capabilities, as typical risk issues often include unlikely and/or ambiguous events."
>
> — Deloitte[5]

Technical Challenges of AI

The evolution from HPC-based analytics to AI workloads in the financial services industry presents a number of technical challenges for organizations on that journey:[6]

- The sheer volume of data involved is made more challenging by the tendency of that data to be dispersed and difficult to use.
- There is an expertise gap, and not many organizations have had the time to develop the skills required to design, deploy, and manage advanced AI projects.[7] Specifically, AI on an HPC infrastructure can require low-level parallel programming expertise that is not common among data scientists.[2]
- The tools available for AI are evolving rapidly and are not always well integrated with each other, making the expertise gap even more difficult to close.
- The cost of IT infrastructure for AI is the top concern of those recently surveyed by research firm Enterprise Strategy Group (ESG).[8]
- Speed at scale is a challenge. The limitations of existing infrastructure can frustrate AI projects by producing results too slowly for the needs of the business.[9]

Because of these challenges, it is no surprise that 46 percent of data and analytics decision makers turn to purchasing commercially available packaged solutions with AI embedded in them.[10]

Dell EMC understands these challenges and offers solutions to meet the needs of financial services organizations on the HPC journey from advanced data analytics to AI. Dell EMC solutions featuring Intel® technologies let companies run ML and DL workloads alongside advanced analytics on the same HPC infrastructure.

Dell EMC Ready Solutions

Dell EMC Ready Solutions include software, servers, networking, storage, and services optimized for HPC and AI workloads, including both traditional and cognitive data analytics. Whether a highly integrated turnkey solution or a flexible customized solution, each Ready Solution is based on a scalable building block approach, so the infrastructure you buy today can grow to meet your needs in the future.

Beyond a complete portfolio of technologies included with each Ready Solution, Dell EMC also makes experts available to help create HPC configurations that can dramatically accelerate results in AI, ML, and DL environments. You can take a test drive with a proof of concept in one of the Dell Customer Solution Centers or in one of Dell EMC's worldwide HPC and AI Centers of Excellence.[11]

The following Ready Solutions are built on 2nd Generation Intel® Xeon® Scalable processors to support the most demanding applications:[12]

Which Solution Is Right for You?

If your situation is:	Consider this Ready Solution:
Already using Apache™ Hadoop® for your data	Ready Solutions for AI, Machine Learning with Hadoop
Seeking a pre-defined, integrated package for deep learning	Ready Solutions for AI, Deep Learning with Intel
Needing a custom solution that builds on your existing infrastructure	Ready Solutions for HPC

Next, let's take look at each of these Ready Solutions.

Ready Solutions for AI, Machine Learning with Hadoop

Organizations that use Apache™ Hadoop® for their big data storage and processing can adopt this Ready Solution to achieve fast implementation of ML with a hardware and software stack optimized for performance. This fully integrated design, created in partnership with Intel and Cloudera, includes data-science and framework optimization, so you can get up and running quickly.

The solution also enables data scientists to do DL experiments with BigDL—a distributed DL library for Apache Spark™. With BigDL and MLlib integration with Spark, data scientists can do both ML and DL projects on the same Hadoop/Spark cluster where data is stored.

The benefits and return on investment (ROI) of this solution have been proven. Forrester conducted a study of the Total Economic Impact™ of this solution on six companies that adopted it, and determined the ROI, net present value (NPV), and payback period for a composite organization's investment.[10]

Figure 2. ROI for six companies adopting Ready Solutions for AI, Machine Learning for Hadoop[10]

Ready Solutions for AI, Deep Learning with Intel

This recently released DL with Intel integrated design includes all the hardware, software, and services you need to get up and running quickly with near-bare-metal performance.

The hardware design is a scale-out cluster consisting of a single Dell EMC PowerEdge™ R740xd master/login node and 16 Dell EMC PowerEdge C6420 dense compute servers in four Dell PowerEdge C6000 chassis, featuring two Intel Xeon Gold 6148 processors per server (40 cores per server).

This Ready Solution also comes with deployment services to accelerate time to results and single-contact support for the complete hardware and software stack. The validated hardware and software stacks combine Dell EMC PowerEdge servers, Dell EMC Isilon™ storage, Dell EMC PowerSwitch networking, data-science software, AI libraries, and frameworks into preconfigured, scalable, balanced systems.

Dell EMC has integrated and validated Intel's open source AI initiative, Nauta, into the solution. Nauta simplifies the process for data scientists to get started with DL.[13] Rather than the multi-step process used in traditional AI infrastructures, with Nauta, a data scientist logs in to the system, specifies template parameters, and then submits the DL training job.

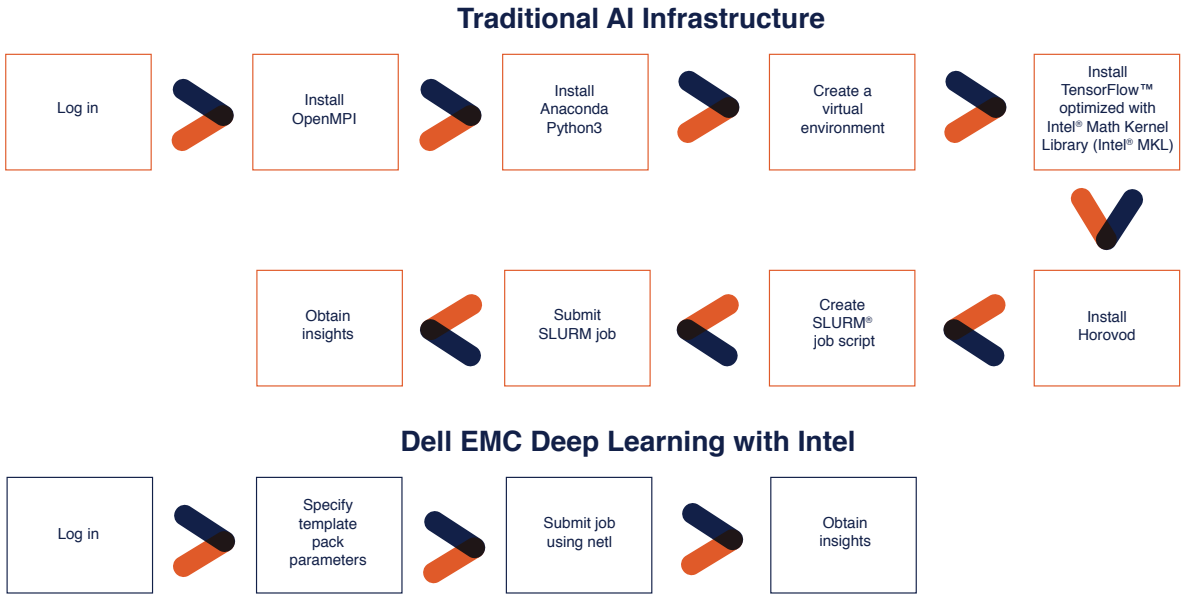

Figure 3. Nauta simplifies deep learning for data scientists[8]

This Ready Solution was validated by the third-party firm ESG, which evaluated the solution's total cost of ownership (TCO). The TCO for this Ready Solution was found to be favorable in comparison to both cloud-based alternatives or GPU-based on-premises solutions.[8]

Ready Solutions for HPC

Ready Solutions for HPC represent a flexible portfolio of high-performance technologies such as Intel® Xeon® Scalable processors and customizable Intel® Arria® 10 GX field-programmable gate arrays (FPGAs) that combine to create custom end-to-end solutions matching your specific target workload and use case.

Even small and medium enterprises can accelerate analytics and AI initiatives. Dell EMC experts take the time to understand your workloads and business goals to design an optimized solution of workstations, servers, networking, storage, and software, rounded out with the services and support you need to succeed.

Included Frameworks and Libraries

These Dell EMC Ready Solutions begin with hardware packages built on technology such as 2nd Generation Intel Xeon Scalable processors for the performance needed to enable AI-ready solutions. But Ready Solutions are much more than simply optimized hardware packages. They also provide the frameworks and libraries you need to get a quick start on your AI projects. Some of the frameworks included with Ready Solutions are:

- BigDL, a distributed DL library for Spark that can run directly on top of existing Spark or Hadoop clusters. BigDL can be used to write DL applications as Scala® or Python® programs.
- Caffe, a DL framework made with expression, speed, and modularity in mind. Caffe is developed by the Berkeley Vision and Learning Center (BVLC), in addition to community contributors, and it is popular for computer vision.[14]
- Intel® Math Kernel Library for Deep Neural Networks (Intel® MKL-DNN), an open source performance library for acceleration of DL frameworks on Intel architecture. It includes highly vectorized and threaded building blocks for implementation of convolutional neural networks (CNNs) with C and C++ interfaces.[15]
- Intel® Machine Learning Scaling Library (Intel® MLSL), a library that provides an efficient implementation of communication patterns used in DL. It is optimized to drive scalability of communication patterns.
- TensorFlow™, a software library for numerical computation using data-flow graphs, developed by Google's Machine Intelligence research organization.[16]
- Nauta, a distributed DL platform that makes use of Kubernetes® and Docker® technologies.[17]

Note that BigDL, Caffe, and TensorFlow have been optimized by Intel for 2nd Generation Intel Xeon Scalable processors.[18] Intel also has an optimized distribution for Python.[19]

Mastercard Is Fighting Fraud the Smart Way[20]

Mastercard is among the payment-processing companies fighting back against fraud every second of every day, using a war chest filled with sophisticated technologies. A case in point: To identify and stop fraudulent transactions, Mastercard leverages machine-learning algorithms running on HPC systems to process large data sets at lightning-fast speeds. …

While that's a challenging proposition for any company involved in retail sales, the scale at which Mastercard operates makes the problem all but unfathomable. According to [VP of the Big Data Consulting Practice at Mastercard Advisors Nick] Curcuru, Mastercard has 2 billion cards in use in more than 210 countries and territories. It processes 165 million transactions per hour, using machine-learning algorithms and applying 1.9 million rules to examine each transaction. It all happens in a matter of milliseconds.

The horsepower for this fraud-prevention engine is a secure, Payment Card Industry (PCI)-certified Apache™ Hadoop® cluster based on high-performance computing systems from Dell EMC. This fraud-detection machine-learning system uses supervised learning to look for established fraud patterns and unsupervised learning to identify emerging fraud patterns in real time.

Dell EMC Can Help Guide Your Journey

As the financial services industry journeys the challenging route from advanced analytics into AI-enabled processes, there is a growing need for high-performance hardware and optimized software suited for both kinds of workloads. Dell EMC Ready Solutions—based on Intel processors—can guide you along your best path to a cost-effective infrastructure that meets all your converging needs for HPC and AI.

- Learn more about Dell EMC Ready Solutions for AI at **www.dellemc.com/readyforai**.
- Learn more about Dell EMC Ready Solutions for HPC at **www.dellemc.com/hpc**.

INTEL® AI BUILDERS MEMBER

Dell EMC is a member of the Intel® AI Builders Program, an ecosystem of industry leading independent software vendors (ISVs), system integrators (SIs), original equipment manufacturers (OEMs), and enterprise end users, which have a shared mission to accelerate the adoption of artificial intelligence across Intel® platforms.

[1] InsideHPC. "How Intel is Driving the Convergence of HPC & Ai." November 2018. https://insidehpc.com/2018/11/intel-driving-convergence-hpc-ai/.

[2] Rob Farber. "AI-HPC is Happening Now." InsideHPC Special Report. www.advancedhpc.com/news-insights/whitepapers/ai-hpc-is-happening-now/.

[3] CIO Review. "Role of Data Analytics in Financial Services." https://bigdata.cioreview.com/cxoinsight/role-of-data-analytics-in-financial-services-nid-24894-cid-15.html.

[4] IDC. "Worldwide Spending on Artificial Intelligence Systems Will Grow to Nearly $35.8 Billion in 2019, According to New IDC Spending Guide." March 2019. www.idc.com/getdoc.jsp?containerId=prUS44911419.

[5] Deloitte. "Why artificial intelligence is a game changer for risk management." 2016. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/audit/us-ai-risk-powers-performance.pdf.

[6] Several of these points are discussed in: O'Reilly. "Accelerating Analytics and AI from the Edge to the Cloud." 2018. https://learning.oreilly.com/videos/strata-data-conference/9781492025955/9781492025955-video319004.

[7] PWC. "22nd Annual Global CEO Survey: CEO's curbed confidence spells caution." 2019. www.pwc.com/gx/en/ceo-survey/2019/report/pwc-22nd-annual-global-ceo-survey.pdf. And PWC. "Technology increases skills gap in financial services and threatens growth." February 2017. https://press.pwc.com/News-releases/Financial-services/technology-increases-skills-gap-in-financial-services-and-threatens-growth/s/e771eccb-14bf-4b46-bcad-0f86d0d9df6c.

[8] Enterprise Strategy Group. "Dell EMC Ready Solutions for AI: Deep Learning with Intel." April 2019. www.dellemc.com/en-us/solutions/data-analytics/machine-learning/ready-solutions-for-ai.htm#scroll=off&overlay=/en-us/collaterals/unauth/analyst-reports/products/ready-solutions/esg-technical-review-deep-learning-with-intel-april-2019.pdf.

[9] PWC. "Speed and sophistication: Building analytics into your workflows." May 2017. www.pwc.com/us/en/services/consulting/library/analytics-speed-sophistication.html.

[10] Forrester. "The Total Economic Impact™ Of Dell EMC Ready Solutions for AI, Machine Learning With Hadoop. August 2018. www.emc.com/collateral/analyst-report/forrester-total-economic-impact-study-dell-emc-ready-solutions-for-ai.pdf.

[11] Dell. "HPC and AI Centers of Excellence: Design solutions, test cutting-edge technologies, and collaborate with experts." www.dellemc.com/en-us/solutions/high-performance-computing/hpc-ai-centers-of-excellence.htm.

[12] Intel. "Intel® Xeon® Scalable Processors." www.intel.com/content/www/us/en/products/processors/xeon/scalable.html.

[13] Carlos Morales. "Introducing Nauta: A Distributed Deep Learning Platform for Kubernetes." www.intel.ai/introducing-nauta/.

[14] Intel. "Intel® Optimization for Caffe*." https://software.intel.com/en-us/ai/frameworks/caffe.

[15] Intel. "The Fastest and Most-Used Math Library for Intel®-Based Systems." https://software.intel.com/en-us/mkl.

[16] Intel. "Intel® Optimization for TensorFlow*." https://software.intel.com/en-us/ai/frameworks/tensorflow.

[17] Intel. "Nauta." www.intel.ai/nauta/.

[18] Intel. "Framework Optimization Training." https://software.intel.com/en-us/ai/frameworks.

[19] Intel. "Accelerate Python* Performance." https://software.intel.com/en-us/distribution-for-python.

[20] Excerpt from: Dell EMC. "Fighting fraud the smart way — with data analytics and artificial intelligence." www.dellemc.com/en-us/collaterals/unauth/white-papers/services/mastercard-fighting-fraud-the-smart-way.pdf.