BRINGING AI INTO YOUR EXISTING HPC ENVIRONMENT, AND SCALING IT UP

Introduction

Today's advancements in high performance computing (HPC) present new opportunities to tackle exceedingly complex workflows. HPC is well established as a key field for advanced computing. Many academic researchers, government agencies, and businesses have deployed HPC solutions, leading to breakthrough discoveries in fields like astrophysics, meteorology, seismology, engineering, genomics, medicine, and nuclear simulation.

The rapid pace of technology advancement means research organizations must consistently evaluate their HPC-enabled approaches. One of the most disruptive technologies of our time, artificial intelligence (AI), creates new opportunities for academics and government agencies to harness the power of HPC to solve unique problems. Those embracing AI capability will maintain their position on the forefront of science; those who ignore its potential run the risk of falling behind their peers.

In this eGuide, we offer practical considerations for HPC managers to incorporate AI into their HPC environment and scale those capabilities to accommodate emerging workloads and increasing end-user demand.

Introduction

Evolving HPC and AI Workloads

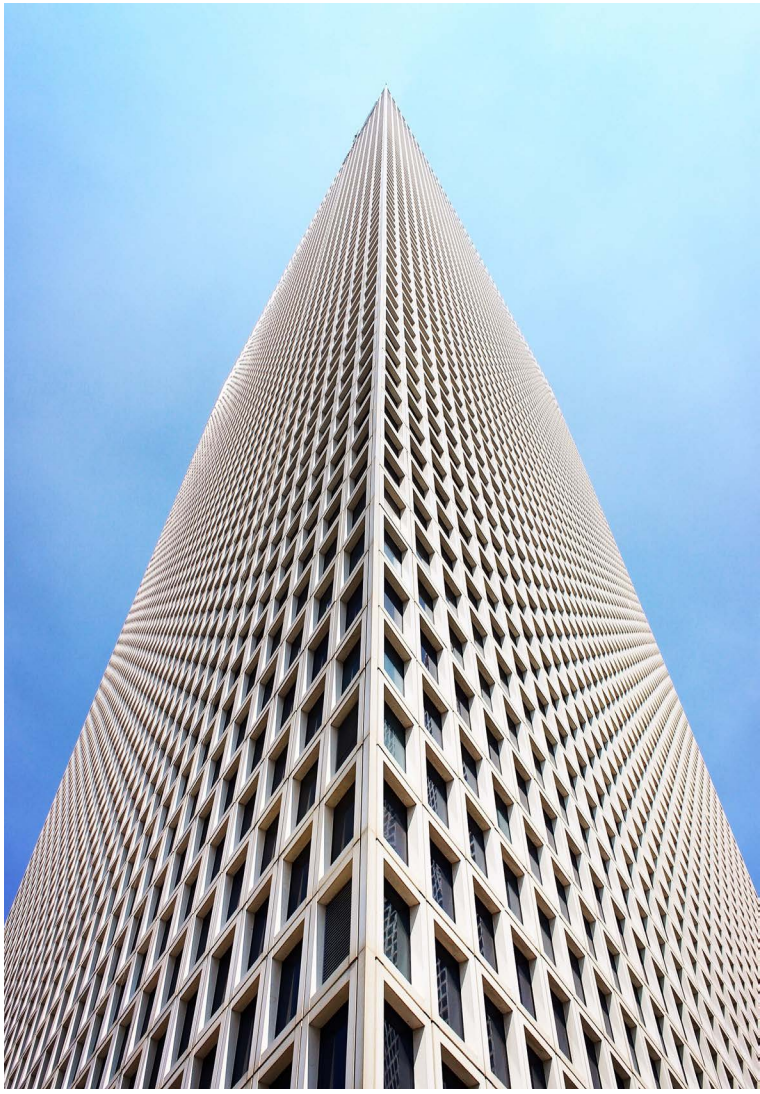
Government institutions and academic researchers push the limits of HPC solutions for advanced problem solving. Deep learning capabilities, and advanced neural networks, can address challenges unimagined only a couple of decades ago. Commonly used HPC workloads like analytics are augmented – or supplanted – by AI to generate more meaningful insights from data.

Unlike past approaches, AI empowers HPC systems beyond simplistic rule-based instructions. Instead, AI evaluates data using an instruction set of 'theories' and algorithms. By learning from these theories, AI can better predict and understand the context using inference to fill in data gaps. AI models complement more traditional HPC solutions to reveal insights faster, and more comprehensively, than data-processing and analytics-based applications can by themselves.

As HPC-driven capabilities like simulation, modeling, big data analytics, and AI converge, hybrid workloads are becoming more and more common. One step in a hybrid workload might involve modeling and simulation, while the next step requires deep learning capability. Traditional HPC use cases like modeling and simulation are more compute-intensive, while newer workloads like AI are more data-intensive. The varied nature of these workflows can represent a big challenge for non-optimized HPC systems. However, a balanced multi-node system excelling at both unleashes new possibilities for increasingly automated discovery and human benefit.

Deploying and Augmenting Scalable Solutions For HPC

If your organization plans updates to its current HPC infrastructure to accommodate advanced AI workloads, below are steps for consideration.



1. THINK HOLISTICALLY ABOUT YOUR STAKEHOLDERS AND HPC NEEDS

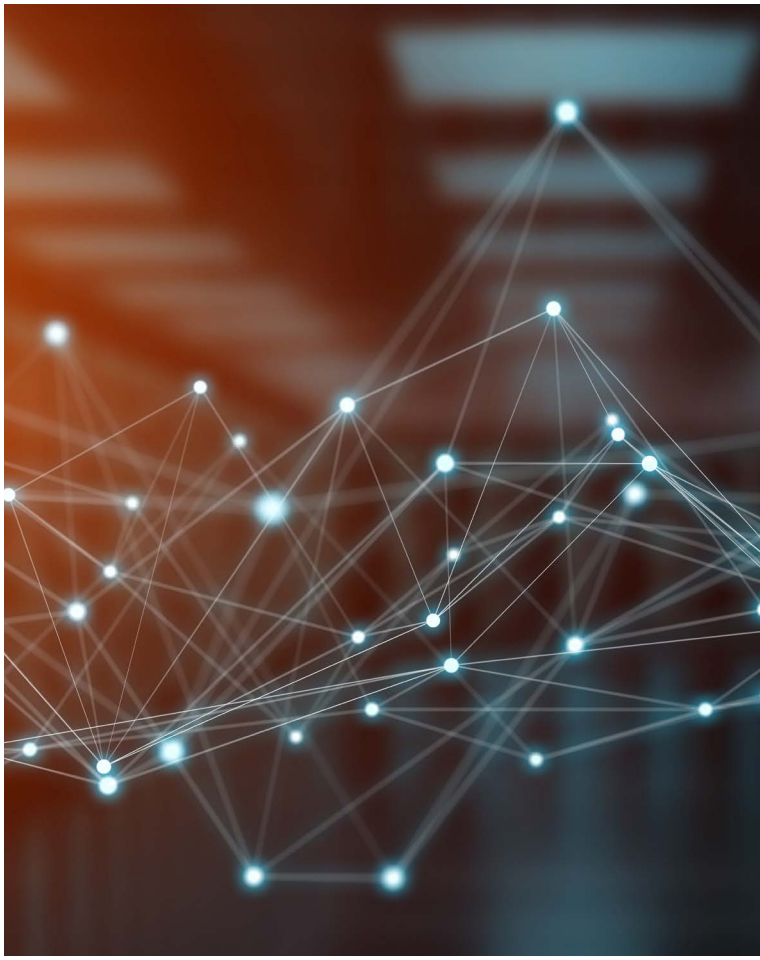
HPC systems are more than the sum of individual parts. To evolve and grow optimal HPC infrastructure for your organization's needs, hardware, software, and human skill sets must converge to develop a holistic HPC and AI implementation strategy. HPC managers seek to provide their users with the HPC infrastructure, support, and tools to make projects successful.

While enterprise organizations regularly adapt their HPC systems for defined workloads, academics and government institutions face a very different challenge. In many government and academic scenarios, HPC systems support a diverse group of individuals, each with unique research needs. To maintain quality of service (QoS) for stakeholders, an HPC system supporting a diverse array of workloads requires scalability, performance, adaptability, and future-proofing. HPC managers must always seek to understand the needs of their users so the HPC investment is used as much as possible by stakeholders.

AI can solve problems by "learning" from supervised and unsupervised examples rather than depending on a formalized set of equations or rules. For this reason, the role of a combined HPC and AI solution offers new capabilities and opportunities. By mapping your organization's HPC and AI needs against use cases and stakeholder demands, you can determine if your current HPC implementation meets its requirements, or what other upgrades or modifications must be prioritized to enable more complex, demanding future scenarios including AI.

Takeaways and Outcomes:

- User needs drive underlying HPC technology.
- Evaluate needs now, be prepared for the future.



2. THINK HOLISTICALLY ABOUT YOUR HPC SOLUTION

Software Selection

End-user experience is paramount. As your organization's needs grow, software capability needs to grow with it. Various scientific endeavors like modeling, simulation, visualization, and AI benefit from applications designed and optimized for those tasks. If you are unsure where to begin the process of identifying the ideal software solution, a good starting point is researching which software solutions best serve other experts driving similar workloads successfully.

Depending on the intended HPC usage scenarios, software resources from Intel, independent software vendors (ISV), and the open source community can offer a jumpstart toward AI-capable systems. By first considering the software solutions needed, it is easier to plan for a physical infrastructure that provides the ideal underpinnings for applications.

2. THINK HOLISTICALLY ABOUT YOUR HPC SOLUTION

Software Application and Development Environment:

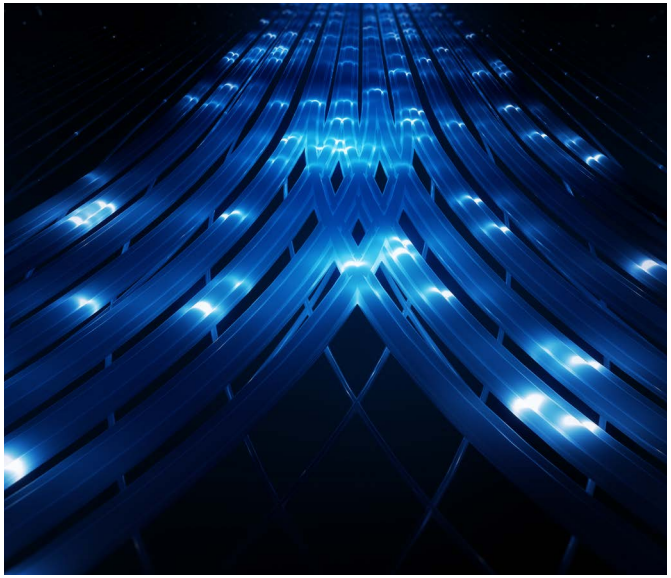
If ready-made applications are inadequate for your unique usage scenarios, you will need the right development skills to modify or create them. For scientific computing, the HPC community offers a large number of high performance libraries for HPC systems of all sizes.

However, optimizing some open source options will require developers with skills and experience in parallel computing or mathematical algorithms. Fortran*, C/C++*, Java*, and Python* code many HPC and AI applications. Intel's HPC interoperable framework, which supports all these languages, helps accelerate the development process. It also supports a common development model for improved code portability, plus software tools to accelerate HPC application modernization initiatives.

	HPC	Big Data (Hadoop/Spark)	AI
Primary Languages	C/C++/Fortran	Java/Python/Scala/R	Python/C++/R
Scaling	Strong	Weak	Both
Cluster Stack Management	Custom	Cloudera Manager/ Apache/ Ambari/ MapR Control System	NA
Schedulers	Slurm, PBS	YARN, Mesos	NA
File Systems	High-performance Distributed POSIX	Distributed Across Node Local Storage - HDFS	Posix or HDFS
OS	RedHat/CentOS SuSE	RedHat/CentOS SuSE	Ubuntu

2. THINK HOLISTICALLY ABOUT YOUR HPC SOLUTION

Deep learning frameworks have a central role in maximizing the potential of AI on HPC systems. Popular frameworks like TensorFlow* and Caffe*, optimized for performance on Intel® architecture-based platforms, speed both deep learning training and inference workflows. Tapping the Intel® Math Kernel Library (Intel® MKL) and Intel® Math Kernel Library for Deep Neural Networks (Intel® MKL-DNN) TensorFlow offers significant performance gains. Similar benefits occur when pairing the Intel® Distribution of Caffe* with Intel® Xeon® processors.



Physical Infrastructure

Before building upon your existing HPC infrastructure, determine first which ingredients perform at an optimal level, and which create bottlenecks that impede advanced workflows. A platform based on Intel® Xeon® Scalable processors and other innovative Intel® technologies for HPC offer the foundation for the most demanding workloads. Elements like processors, memory, storage, and fabric all weigh into the equation.

Wherever possible, preserve your existing infrastructure investments. However, if anticipated use requirements exceed your HPC system's hardware capability, upgrades may be necessary. Budget for the needed components and prioritize their implementation over time as workloads require it. By taking a staged approach, you can help your existing HPC infrastructure evolve and scale over time.

Flexibility of HPC systems for academic environments

The Texas Advanced Computing Center's (TACC) Stampede2[2] HPC system helps thousands of scientists and researchers achieve breakthroughs in AI, modeling, simulations, and much more. To accommodate the diversity of projects, Stampede's flexible and scalable HPC infrastructure supports more than 40 major science and engineering applications at any given time, alongside thousands of applications used by small groups or individual researchers. The current iteration Stampede2 ranks at #12 on the June 2017 Top500.org[3] listing of the fastest supercomputers worldwide.

Takeaways and Outcomes:

- Determine if current data and compute infrastructure generate desired insights.
- Map technology needs against today's business opportunities.
- Consider software and evaluate key AI frameworks early on.
- Choose a flexible, scalable HPC platform which can scale to accommodate complex AI workloads.

3. VALIDATE YOUR HPC TECHNOLOGY FIRST, THEN SCALE IT UP

Different organizations have different needs for HPC and AI deployment, so the process is not a one-size-fits-all proposition. Some supercomputers used in government and academic environments are supported by many internal staff members with all the expertise necessary to manage and grow their HPC environment. Organizations lacking dedicated human resources and expertise should consider working with Intel, an original equipment manufacturer (OEM), or a consultant who can help them accelerate system deployment.

With your team in place, validate your test system before completing a full-scale implementation. The testing process helps determine many factors like the adequacy of performance levels for challenging workloads, whether bottlenecks exist, and the value of data insights obtained. If the test system proves unable to meet your organization's needs, regroup to determine alternate, more effective solutions.

Once the chosen technology meets all your intended outcomes, a full-scale rollout is a next step. Whether an internal team or a third-party completes your HPC implementation, be sure you also identify experts to accommodate ongoing maintenance and administration. Having the right team — and the right skills — in place will reap the greatest results from longer-term HPC investments.

HPC implementation enables extreme weather modeling[4]

One organization dedicated to weather monitoring has applied a careful evaluation process to define its HPC infrastructure, and software needs to solve the major challenge of predicting hazardous weather conditions. Their task – quite literally – can be a matter of life or death. Driven by the need to predict the path of destructive hurricanes well in advance of landfall, researchers optimized their HPC systems for the important task. Through rapid analysis of available meteorological data, researchers can alert local authorities well in advance should weather conditions indicate the need for evacuation.

Takeaways and Outcomes:

- Identify your deployment team.
- Perform testing and validation to confirm system adequacy before rollout.
- Identify experts for ongoing administration, maintenance, and system growth.

Summary

Three practical steps can help you determine ideal infrastructure requirements for a future-proofed HPC system:

1. Think holistically about your stakeholders and HPC needs.
2. Think holistically about your HPC solution.
3. Validate your HPC technology first, then scale it up.

Learn More

To find out more about Intel's HPC technologies and how they can advance your organization's business and scale to meet the needs of increasingly complex workloads like AI, talk to your preferred system provider or learn more at intel.com/hpc.

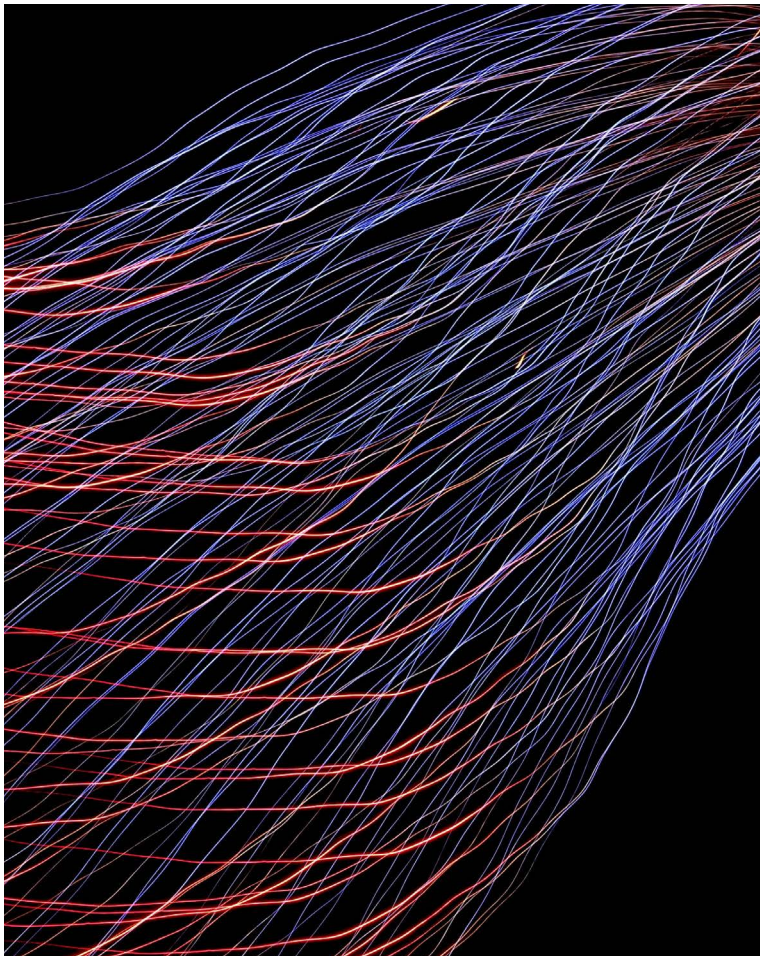
Other Resources

Intel® Omni-path architecture white paper.

AI-HPC happening now white paper.

Intel® Xeon® Scalable processors for HPC, business brief.

Accelerating AI with Intel® Omni-Path Architecture webinar.



Intel® technologies enable advanced AI workloads today

The ever-expanding demands of today's large and complex HPC workloads require powerful underlying technologies that deliver:

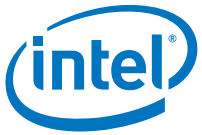
- Improved performance
- System resilience
- Energy efficiency
- Usage flexibility
- Code portability

Potential bottlenecks such as processor speed, memory, and storage must each be addressed in a holistic HPC system to derive the greatest benefit from each element. For vast multi-rack HPC systems, the latency of communication among nodes can also impact overall system speed. The interconnecting fabric can have a profound impact on overall system throughput too.

Intel's leadership in platform innovation, supported by a broad software ecosystem, is paving the way for future AI capabilities. Intel's HPC technologies include many innovations to help organizations get the most from their most demanding workloads:

- Intel® Xeon® Scalable processors and Intel® Xeon Phi™ processors
- Intel® Omni-Path Architecture
- Intel® FPGAs
- Intel® Optane™ technology
- Intel® 3D NAND SSDs
- Intel® Advanced Vector Extensions 512 (Intel® AVX-512)

Share this eGuide



[1] https://newsroom.intel.com/editorials/one-simple-truth-artificial-intelligence-healthcare-its-already-here/

[2] https://www.nextplatform.com/2017/07/25/texas-advanced-supercomputing-center-taps-latest-hpc-tech/

[3] https://www.top500.org/list/2017/06/

[4] https://www.hpcwire.com/2017/08/28/nersc-scales-deep-learning15-pflops/

0418/JO/CAT/PDF

337479-001EN