THE CONVERGENCE OF HIGH PERFORMANCE COMPUTING, BIG DATA, AND MACHINE LEARNING

Summary of the Big Data and High End Computing Interagency Working Groups Joint Workshop
October 29-30, 2018

A report by the

BIG DATA INTERAGENCY WORKING GROUP *and the*
HIGH END COMPUTING INTERAGENCY WORKING GROUP

NETWORKING & INFORMATION TECHNOLOGY
RESEARCH & DEVELOPMENT SUBCOMMITTEE

COMMITTEE ON SCIENCE & TECHNOLOGY ENTERPRISE

of the

NATIONAL SCIENCE & TECHNOLOGY COUNCIL

SEPTEMBER 2019

About the National Science and Technology Council

The National Science and Technology Council (NSTC) is the principal means by which the Executive Branch coordinates science and technology policy across the diverse entities that make up the Federal research and development enterprise. A primary objective of the NSTC is to ensure that science and technology policy decisions and programs are consistent with the President's stated goals. The NSTC prepares research and development strategies that are coordinated across Federal agencies aimed at accomplishing multiple national goals. The work of the NSTC is organized under committees that oversee subcommittees and working groups focused on different aspects of science and technology. More information is available at https://www.whitehouse.gov/ostp/nstc.

About the Office of Science and Technology Policy

The Office of Science and Technology Policy (OSTP) was established by the National Science and Technology Policy, Organization, and Priorities Act of 1976 to provide the President and others within the Executive Office of the President with advice on the scientific, engineering, and technological aspects of the economy, national security, homeland security, health, foreign relations, the environment, and the technological recovery and use of resources, among other topics. OSTP leads interagency science and technology policy coordination efforts, assists the Office of Management and Budget with an annual review and analysis of Federal research and development in budgets, and serves as a source of scientific and technological analysis and judgment for the President with respect to major policies, plans, and programs of the Federal Government. More information is available at https://www.whitehouse.gov/ostps.

About the Networking and Information Technology Research and Development Program

The Networking and Information Technology Research and Development (NITRD) Program is the Nation's primary source of federally funded coordination of pioneering information technology (IT) research and development (R&D) in computing, networking, and software. The multiagency NITRD Program, guided by the NITRD Subcommittee of the NSTC Committee on Science and Technology Enterprise, seeks to provide the R&D foundations for ensuring continued U.S. technological leadership and meeting the needs of the Nation for advanced IT. The National Coordination Office (NCO) supports the NITRD Subcommittee and the Interagency Working Groups (IWGs) that report to it. More information is available at https://www.nitrd.gov/about/.

About the Big Data and High End Computing Interagency Working Groups

NITRD IWGs work to identify needs and opportunities across the Federal Government for R&D activities relevant to networking and IT and to offer opportunities for R&D coordination among agencies, academia, and the private sector.

The NITRD Big Data Interagency Working Group (BD IWG) focuses on R&D to improve the management and analysis of large-scale data—including mechanisms for data capture, curation, management, processing, and access—to develop the ability to extract knowledge and insight from large, diverse, and disparate data sources.

The NITRD High End Computing Interagency Working Group (HEC IWG) focuses on R&D to advance high-capability, revolutionary computing paradigms, and to provide the Nation with state-of-the-art computing, communication, software, and associated infrastructure to promote scientific discovery and innovation in the Federal, academic, and industry research communities.

More information is available at https://www.nitrd.gov/nitrdgroups/.

Acknowledgments

This workshop report was developed through contributions of the workshop committee, which included representatives from government, academia, and industry; NITRD Federal agency representatives; members of the Big Data and High End Computing IWGs; and staff of the NITRD NCO. Sincere thanks and appreciation go out to all contributors.

Copyright Information

This document is a work of the U.S. Government and is in the public domain (see 17 U.S.C. §105). It may be freely distributed, copied, and translated with acknowledgment to OSTP. Requests to use any images must be made to OSTP. Digital versions of this and other NITRD documents are available at https://www.nitrd.gov/publications/. Published in the United States of America, 2019.

Abbreviations

AI	artificial intelligence
BD	big data and Big Data NITRD IWG
CANDLE	CANcer Distributed Learning Environment
FPGA	field-programmable gate arrays
DL	deep learning
DOE	Department of Energy
FPGA	field-programmable gate array
GPU	graphic processing unit
HEC	High-End Computing (NITRD IWG)
HPC	high performance computing
IWG	Interagency Working Group
ML	machine learning
NCI	National Cancer Institute
NIH	National Institute of Health
NITRD	Networking and Information Technology Research and Development (Program)
TB	terabyte

Background

The high performance computing (HPC) and big data (BD) communities traditionally have pursued independent trajectories in the world of computational science. HPC has been synonymous with modeling and simulation, and BD with ingesting and analyzing data from diverse sources, including from simulations. However, both communities are evolving in response to changing user needs and advancing technological landscapes. Researchers are increasingly using machine learning (ML) not only for data analytics but also for modeling and simulation; science-based simulations are increasingly relying on embedded ML models not only to interpret results from massive data outputs but also to steer computations. Science-based models are being combined with data-driven models to represent complex systems and phenomena. There also is an increasing need for real-time data analytics, which requires large-scale computations to be performed closer to the data and data infrastructures, to adapt to HPC-like modes of operation. For example, in tactical mission support, where data comes from many different sources and the computational environment is varied and geographically distributed, new capabilities would include improved situational awareness and decision-making techniques such as imagery analysis to extract useful information from raw data; increased operating safety for aircraft, ships, and vehicles in complex, rapidly changing environments; and predictive maintenance and supply chain operations to predict the failure of critical parts, automate diagnostics, and plan maintenance based on data and equipment condition. This and other use cases create a vital need for HPC and BD systems to deal with simulations and data analytics in a more unified fashion.

To explore this need, the NITRD Big Data and High-End Computing R&D Interagency Working Groups held a workshop, The Convergence of High Performance Computing, Big Data, and Machine Learning, on October 29-30, 2018, in Bethesda, Maryland. The purposes of the workshop were to bring together representatives from the public, private, and academic sectors to share their knowledge and insights on integrating HPC, BD, and ML systems and approaches and to identify key research challenges and opportunities. Workshop participants represented a balanced cross-section of stakeholders involved in or impacted by this area of research. The workshop agenda, list of attendees, webcast, and other details are available at https://www.nitrd.gov/nitrdgroups/index.php?title=HPC-BD-Convergence.

Key Takeaways

There are four key takeaways from the joint workshop on the convergence of HPC, BD, and ML:

- Data is growing at an unprecedented rate, and science demands are driving the convergence of HPC, BD, and ML. It is not unusual to see petabytes of data being generated from one experimental instantiation. Data generation is no longer the research bottleneck it once was; it is now data management, analysis, and reasoning that are the bottlenecks.
- There will be increased heterogeneity in future systems—including specialized processors such as graphics processing units (GPUs) and field-programmable gate arrays (FPGAs)—as the performance improvements provided by semiconductor scaling diminish. Systems will need to be flexible and have low latency at all levels to effectively support new use cases. In addition, new tools and benchmarks will be needed to understand the common issues across simulation (HPC), big data, and ML applications, because there is little reliable data available at present.
- The computing ecosystems of tomorrow will not look like the computing ecosystems of today. Future computing will likely involve combinations of edge, cloud, and high performance computing. To make this a seamless ecosystem, new programming paradigms,[1] language compilers, and

[1] In this document, the term "programming" is not limited to hand-coding but is meant to reflect all levels, including auto-code development that will result in flexible, low defect software.

operating and runtime systems will be needed to provide new abstractions and services. "Smart computing at the edge," which involves intelligent data collection or data triage at the edge of the network (near the source of the data), is expected to become increasingly important.

- More collaboration between the HPC, BD, and ML communities is needed for rapid and efficient progress toward an ecosystem that effectively serves all three communities. While convergence of data analytics and HPC-based simulation has seen some progress, the software ecosystems supporting the HPC and BD communities remain distinctly different from one another, mainly due to technical and organizational differences.

Event Summary

The workshop began with an overview of the current landscape and use cases, which was followed by panel sessions and break-out discussions on the challenges and opportunities in three different aspects of convergence: hardware, modes of operation, and software. The four workshop sessions are summarized below.

Current Landscape, Use Cases or Applications, and Challenges

This session explored use cases from a variety of domains and applications to illustrate the current landscape, including what is currently possible and what new opportunities could emerge with convergence. Presentations highlighted the pervasiveness and unprecedented scale of data being generated and illustrated that convergence is already underway. (See also two recent convergence examples noted in the sidebar and referenced attachments at the end of this document.)

The session presentations showed that researchers are working together closely to build predictive models that both integrate a variety of experimental data and rely on ML to help steer new simulations and experiments. This form of convergence enables researchers to have a more dynamic view of domain sciences and optimize solutions with a significant reduction in compute requirements. However, there are several overarching challenges for convergence of HPC, BD, and ML:

Federal Collaborative Exascale HPC-BD-ML Projects

One example of HPC–BD-ML convergence is the Department of Energy (DOE) and National Institutes of Health (NIH) collaboration for the National Cancer Institute's Cancer Distributed Learning Environment (CANDLE). CANDLE focuses on bringing together data from three major challenge areas (molecular, drug response, and treatment strategy) to improve cancer patient outcomes. Each area involves distinct teams of experts using diverse forms of data at different scales, models, and simulations. The goal is to build a "single scalable deep neural network code that can be used to address all three challenges."[2] For more details on CANDLE, please see Attachment A.

Another convergence example is the DOE–industry-university collaboration Exascale Deep Learning (DL) for Climate Analytics, where researchers from multiple organizations used DOE's Summit supercomputing system to identify extreme weather patterns using a trained DL model. HPC resources are essential for handling the extreme data sizes and complexity in this application. For more details on this project, please see Attachment B.

- *Access to highly curated data and compute resources.* Although data is being generated at an unprecedented scale, there is a growing need for sound data management. Big data cannot be exploited for ML without well-curated, tagged datasets. Academia cannot keep pace with the

[2] https://candle.cels.anl.gov/

rapid advances made by industry without well-curated "gold standard" datasets and the scales of computing and hardware necessary to perform these computations.

- *Skilled workforce teams*. A converged HPC–BD–ML environment inherently requires using collaborative research teams of domain scientists, data scientists, and software engineers. Many domain scientists are unfamiliar with the new integrated technologies and require a team that includes software engineers and data scientists. This represents a shift from the traditional domain-based research teams consisting only of principal investigators and their graduate students; it also presents a new career path for data scientists and software engineers.
- *Scientific reproducibility*. Publication of scientific results should include the data and the software that support the results so that other scientists can evaluate the rigor of an experiment design and the quality of data results, as well as be better able to reproduce results as a basis for further research. With convergence comes the opportunity to reexamine and improve current processes to make reproducibility a reality.

Hardware Opportunities and Challenges

This session examined the various aspects of hardware convergence. Discussions highlighted the fact that both simulation and data analytics depend on the ability of computer systems to perform dense linear algebra efficiently. Because systems are designed not for specific application areas but instead for data structures and methods, some aspects of integration are not as difficult as previously imagined. Evidence to support this includes machines that presently support simulation, data analytics, and ML projects, such as DOE's Summit[3] and the National Science Foundation–funded Blue Waters[4] and Frontera[5] supercomputers. Despite this, there are performance issues that need to be addressed as hardware becomes more heterogeneous and flexible in response to changing user needs.

Major hardware challenges for achieving convergence include:

- *Interconnect efficiency at all levels:* More efficient interconnects are needed to facilitate better performance across nodes, including intra-node, inter-node, fabric, and inter-fabric. This is critical for large-scale applications. Today's hardware options are not efficient when off-node operations are required.
- *Innovative tools and common end-to-end benchmark suites:* Tools are needed to enable better understanding of compute workloads, performance, and bottlenecks to ensure effective and useful converged systems. There are no well-researched data, only anecdotes, to help identify common bottlenecks across simulation, big data, and ML applications.
- *Power efficiency:* The needs of the commercial sector will likely drive evolutionary approaches to improve power efficiency. However, work is needed to develop innovative fine-grain power-efficiency techniques—distinct from evolutionary steps—for both processers and memory.
- *Integrated memory:* Both simulation and data analytics are memory-bound, and research is needed for innovations in integrating memory and processing.
- *Scalable file systems:* HPC currently relies on file systems that do not scale well for new applications such as ML. Research is needed to identify or develop file system technologies that are effective for both HPC and BD.

[3] https://www.olcf.ornl.gov/summit/

[4] http://www.ncsa.illinois.edu/enabling/bluewaters

[5] https://www.tacc.utexas.edu/research-development/tacc-projects/frontera

- *Reliable networking:* There is a need for a low-cost, end-to-end wide area network that is reliable and more fully automated.
- *Balanced hardware development:* As specialized hardware such as GPUs and FPGAs are being employed to improve performance—particularly for applications involving machine learning and deep learning (DL)[6]—there is a need to balance R&D on different hardware aspects such as input and output, latency and bandwidth, memory-type, and the heterogeneity of the processor(s).

Modes of Operation Opportunities and Challenges

As noted above, new instrumentation, enormous data rates, and ever-increasing computing platform complexity are driving new computational use cases and requirements. This session explored the different modes of operation resulting from applications at the convergence of HPC, BD, and ML. Discussions highlighted that large-scale experiments, which have traditionally relied on local computing for data processing, are increasingly turning to HPC to produce timely results. Likewise, some ML or DL applications require HPC-scale resources for the training phases. Simulations are also reaching a scale and complexity such that a single application can take the form of a complex workflow of tasks and could benefit from using ML to automate those workloads. In addition, increased use of smart computing at the edge presents a use case where HPC, simulation, data analytics, and ML converge in the workflow across a distributed infrastructure.

Major challenges regarding modes of operation include:

- *Scalable tools and capabilities for ML and large-scale data analysis:* Large-scale complex simulations rely on scalable numerical libraries and software that have been optimized over recent decades. Innovations are needed to address the new world of ML and large-scale data analysis and may require changes in the underlying software stack.
- *New user training and support:* New users, whether from large-scale experiments or large-scale ML, are driving new HPC workloads. Services are needed to meet their needs, including support for a new and diverse set of software packages that are critical to their applications. In addition, more intuitive interfaces are needed for users who are not familiar with running applications at scale or interfacing with large-scale computing resources.
- *New tools and services for data:* Until recently, the HPC community has focused on simulation data and data management services local to the HPC centers. But with the data-driven nature of many of the new convergence applications, additional tools and services are needed for large-scale data management, curation, retention, and access.
- *Well-managed end-to-end solutions:* Whether complex simulations requiring embedded ML or complex distributed workflows, new applications will benefit from well-managed end-to-end solutions that reduce complexity and include reliable and elastic systems.

Software Opportunities and Challenges

This session explored the software stack and related convergence challenges and opportunities. Discussions revealed that the HPC community has embraced data analytics and that recent HPC systems are well equipped to combine the predictive capabilities of simulation with the analytic and optimization capabilities of machine learning. With the recent adoption of deep neural networks for machine learning, data analysis now has computational characteristics of traditional HPC workloads. Both HPC

[6] In layman's terms, machine learning refers to being able to train a computer to perform certain tasks without being explicitly programmed. DL is a type of machine learning where pattern recognition using stacked neural networks is used. Neural networks are modeled after the human brain using sensors and several layers of nodes.

and data analytic systems are adopting the use of accelerators such as GPUs to improve the performance of individual computing nodes, and this trend will continue as a response to the limited gains of scaling.

Major challenges regarding software include:

- *System design:* At the system level, there is a significant gap between how HPC systems are designed (tightly coupled collections of homogeneous nodes) versus BD systems (based on cloud computing data center architectures that consist of large numbers of loosely coupled and possibly heterogeneous computing nodes). These structural differences, in turn, have led to a split in the software stack that is both technological and cultural. The HPC stack relies on tools developed in government laboratories and academia. In contrast, the BD stack is much larger and more varied and is often driven by open-source projects, with the main contributors being commercial entities.
- *Edge Computing*, or smart computing at the edge*:* This was identified as a rapidly emerging key area—one requiring new abstractions, concepts, and tools, including the software architecture, runtime systems, and perhaps even new programming languages. The emerging combination of edge, cloud, and HPC will require software that makes these environments easier to program, debug, optimize, and interoperate in many future application areas.
- *System management:* Regardless of whether the computing is HPC or BD, launching massive jobs will require support to reduce job launch latency, monitor jobs in real time, and handle runtime node and other failures.
- *Common libraries:* Most domain scientists and BD users do not have the expertise to handle the complexities of emerging hardware. Having a common set of libraries would allow nonexperts to more easily use the systems, leaving the programming of these devices to experts.

Conclusion

The NITRD joint BD–HEC IWG workshop explored challenges and opportunities for convergence of HPC, BD, and ML. From the presentations and discussions, a vision emerged of a rich computational ecosystem consisting of heterogeneous combinations of edge, cloud, and high performance computing systems. This ecosystem would be flexible and be able to receive data from a variety of sources such as scientific and medical instruments, sensor networks, and security and infrastructure monitoring systems. It would have edge Internet of Things devices that would extract important features and convert data into forms suitable for ingesting and storing in the cloud. Large-scale data analytics would run in the cloud, combining ingested data with stored databases. HPC systems would perform more computationally intensive forms of analysis and optimization, as well as run simulations for predictive modeling.

Such a rich computing environment could provide capabilities well beyond those of today’s isolated systems. For biomedical and clinical research and healthcare, it would enable the use of clinical, laboratory, and even molecular data for patients and for researchers. Data sources could include smart health applications where patient outcomes are connected to an evidence-based computed model, thereby putting data as a “digital first” asset in the healthcare system. The computing environment would allow scientific and medical researchers to solve problems with many degrees of freedom in ways that allow data to inform models and simulations to build better models.

Achieving this vision of a rich computing ecosystem will require new capabilities in hardware (computing, network, and storage); management modes of operation; and software. Providing true convergence among current and future computing environments presents many technical and organization challenges, but it could provide capabilities in scientific research, national security, healthcare, and industry well beyond what is possible today.

Attachments

A: CANDLE: Exascale Deep Learning & Simulation-Enabled Precision Medicine for Cancer

CANDLE (CANcer Distributed Learning Environment) is an exascale computing project whose goal is to enable the most challenging deep learning problems in cancer research to run on the most capable supercomputers of the Department of Energy (DOE) and the National Institute of Health (NIH). It is designed to support three top challenges of the National Cancer Institute (NCI): understanding the molecular biology of key protein interactions, developing predictive models for drug response, and automating the analysis and extraction of information from millions of cancer patient records to determine optimal cancer treatment strategies.

By tackling these exemplar cancer problems, CANDLE is building a core set of cross-cutting technologies aimed at addressing common challenges at the convergence of HPC, big data, and artificial intelligence (AI) in science. For example, data processing and feature selection methods implemented in CANDLE allow experimental and derived datasets of multiple modalities to be harmonized and integrated in a machine learning framework. Representation learning methods in CANDLE compress very large input spaces such as raw simulation states into low dimensional representations that capture their scientific essence. Such encoded representations are then used to steer simulation, provide synthetic validation data, and guide the acquisition of new experimental samples in cancer research workflows.

CANDLE also aims to accelerate the many stages of DL workflows writ large, including feature engineering, parallel training, weight sharing in model populations, architectural search, hyperparameter optimization, and large-scale inference with uncertainty quantification. To accelerate large-scale model search experiments, ensembles, and uncertainty quantification, CANDLE features a set of DL benchmarks. These benchmarks are aimed at solving a problem associated with each of the cancer challenge problems, embody different DL approaches to problems in cancer biology, and are implemented in compliance with CANDLE standards. Combined, these techniques will support the application of DL to more scientific domains and prepare them for existing HPC resources and forthcoming DOE exascale platforms.

Implementations of CANDLE have been deployed on the DOE HPC systems Titan and Summit at Oak Ridge National Laboratory, Theta at Argonne National Laboratory, and Cori at the National Energy Research Scientific Computing Center at Lawrence Berkeley National Laboratory, as well as on the NIH Biowulf system. CANDLE computations use the full scale of these machines, using many thousands of nodes in parallel, requiring tens of terabytes (TBs) of input/training data, and producing many TBs of output data to analyze. In some cases, training data is harvested from petabytes of simulation data.

CANDLE software builds on open source DL frameworks and the project engages in collaborations with DOE computing centers, HPC vendors, and the DOE Exascale Computing Project (ECP) to both leverage and drive new advances in HPC software. Future release plans call for supporting experimental design, model acceleration, uncertainty-guided inference, network architecture search, synthetic data generation, and data modality conversion, as well as expanding into more scientific domain research areas.

For more details, see https://candle.cels.anl.gov/.

Attachment B: Exascale Deep Learning for Climate Analytics

In 2018, researchers from the DOE National Energy Research Scientific Computing Center at Lawrence Berkeley National Laboratory, a leading American technology company, the DOE Leadership Computing Facility at Oak Ridge National Laboratory (ORNL), and a leading American university achieved a major breakthrough when they successfully scaled a deep learning application on the DOE Summit supercomputing system at ORNL using 27,360 GPUs. The team developed an innovative convolutional segmentation architecture to automatically extract pixel-level masks of extreme weather patterns such as tropical cyclones and atmospheric rivers, thus enabling the climate science community to characterize the frequency and intensity of such events in the future. This project was awarded the prestigious Gordon Bell Prize at the Supercomputing 2018 conference.

The project overcame a number of technical challenges, most prominently in the area of storage and data management where the general parallel file system was unable to sustain the data and metadata rates required. HPC resources were essential for handling the extreme data sizes and complex learned network inherent in this climate application. The team processed a 20 TB climate dataset on 4560 Summit nodes, obtaining 1.13 Exaflops/second (EF/s) peak, and 0.999 EF/s sustained performance in half-precision mode.

The research team anticipates the co-design of future HPC systems to better support read-dominated AI workloads. Future DL frameworks will need to support optimized ingest pipelines for scientific datasets, supporting hybrid modes of data and model parallelism, and innovative methods for ensuring convergence at extreme scales.

For more details, see the paper from the IEEE Supercomputing 2018 conference proceedings, “Exascale Deep Learning for Climate Analytics” (Thorsten Kurth et al.): https://arxiv.org/pdf/1810.01993.pdf.