Update

High-Performance Computing Leadership Initiatives in the United States: HPC User Forum, April 7–9, 2014, Santa Fe, New Mexico

Steve Conway
Chirag Dekate, Ph.D.
Earl C. Joseph, Ph.D.

IN THIS UPDATE

This IDC update captures part of the proceedings at the 52nd HPC User Forum held in Santa Fe, New Mexico. The topic covered in detail in this document concerns high-performance computing (HPC) leadership initiatives across the United States. This update is a near-verbatim records of the talks given during the HPC User Forum session.

"Trinity Next-Generation Supercomputer for the ASC Program": Doug Doerfler, Sandia National Laboratories

The formal design review was completed in April 2013. The project just went out for a best and final offer. It takes a long time to spend $100 million of the government's money. It takes a lot of review and discussion.

ASC is the Advanced Simulation and Computing program within NNSA. This includes the tri-labs: Lawrence Livermore National Laboratory (LLNL), Los Alamos National Laboratory (LANL), and Sandia National Laboratories. Historically, the program has had capacity, capability, and advanced systems. The labs manage their own software stack. Capacity systems are clusters. Commodity systems are getting more capable and taking over the low end of what capability systems have handled, while advanced systems are taking over the top half of what capability systems have handled.

We do expect to do some NRE (e.g., for the transition on Trinity from multicore to manycore for our applications teams and also for integrating SSDs at large production scale and for advanced power management). The New Mexico Alliance for Computing at Extreme Scale (ACES) is a partnership between Los Alamos and Sandia. Our acquisition strategy allows us to buy several systems over time. The ACES MOU was signed in 2008 to co-develop and co-use HPC to meet the nuclear weapons stockpile stewardship needs. This is expected to produce a cost benefit. It also allows vendors to save money and time by responding to a single RFP and a single set of technical requirements. Another advantage of a single RFP for multiple labs is that the combined perspective reduces risk, avoiding tunnel vision by one lab. There will also be benefits in production, shared bug reports, and quarterly meetings, and there are less likely to be one-off systems.

April 2014, IDC #248119

NNSA and the Office of Science are collaborating as part of the road to exascale, which is a very expensive proposition. The Trinity-NERSC-8 collaboration will proceed with the joint RFP and selection, separate system awards, and attendance at each other's review sessions. It is a collaboration of two separate projects that will result in two distinct systems.

Target Specifications:

- Trinity memory 2–4PB, NERSC-8 1–2PB.
- We don't specify goals in peak flops but in capability gain over our current system. We are looking for a 8–10x gain with Trinity.
- Trinity: less than12+3MW; NERSC-8: less than 6MW
- The procurement will include 19 workloads from the tri-labs.
- Half of the current Cielo platform is dominated by very large jobs that use 32,000 cores and up and exploit 25–50% of the machine for a single job.
- Trinity is designed for large 3D weapons applications.
- The target delivery date is September 2015.

The number 1 challenge will be programming these systems, and we will collaborate on this with the vendors, right down to the chip vendors. Linpack is not a driving requirement, but it's a good stress test. There's a new Top 500 benchmark based on the sparse matrix equation.

"NERSC-8 Next-Generation Supercomputer": Katie Antypas, NERSC

We are very close to announcing the systems, maybe within a week.

Our challenge is moving to manycore architectures. NERSC directly supports the DOE Office of Science. We have 5,000 users across 700 projects. DOE-Science awards the time on our systems. We focus on the science productivity of our users. We supported three recent Nobel laureates.

Our two large systems are Hopper (a Cray XE6 with 1.3PF of peak and 150,000 cores) and Edison (Cray XC30 with 2PF of peak and 125,000 cores). Our strategy is to have two large systems on the floor at any time. We have a partnership with NNSA and focus on general-purpose architectures to support a wide range of applications.

Our main challenge will be transitioning our diverse workload to manycore architectures. Our platform also needs to start transitioning to more energy-efficient manycore architectures. We are looking for system delivery in 2015–2016. Workloads not ready to move to NERSC-8 can still run productively on Edison, which will still provide 20% of our cycles in 2015–2016. Our strategy is a smooth progression to exascale for our users, plus support for legacy codes and for a variety of programming models.

Case Study on Xeon Phi

At the NERSC testbed Babbage, we tested 45 Sandy Bridge nodes with Phi. We chose to test Phi as if it were a standalone processor (the next Phi version will have a self-hosted capability). FLASH is an adaptive mesh refinement code. We parallelized it and ran it in self-hosted first on Phi, then on 1 Sandy Bridge and 2 Sandy Bridge processors. Phi was slower, so we looked at what optimizations we could make. The best performance was with 120-way parallelism using MPI + OpenMP.

Next, we looked at how efficient our OpenMP implementation was. The speedup with more threads didn't make that much difference. Vectorization is another form of on-node parallelism. We tell our people, most of whom didn't grow up with vectorization, not to forget about it. Vectorization made no real difference on Phi. To get a vectorization gain, we had to restructure the code. This gave the best performance on Xeon Phi. This was good news. Then we looked at what happens when you run this on the original processors, and this was also encouraging.

"CORAL: A Collaboration of Oak Ridge, Argonne, and Lawrence Livermore to Procure Their Next-Generation Leadership Computing Systems": Buddy Bland, ORNL

The systems we have today at ORNL and Argonne are part of DOE-Science's Leadership Computing Facility (LCF) aimed at solving the most challenging scientific and engineering problems. Two centers come out of one budget line item. We jointly run a single allocation program for access to the HPC resources of the two centers. INCITE allocates about 60% of these resources, and ALCC allocates about 30% for DOE mission-specific work. We typically get requests for three times more cycles than we have available. LCF centers are focused on enabling science and engineering breakthroughs.

- Argonne has a 10PF Blue Gene Q with 768TB of system memory.
- ORNL has 27PF Cray XK7 with 710TB of system memory.
- ORNL and Argonne partner with LLNL, which has a 20PF IBM Blue Gene Q with 1.6PB of system memory.

CORAL was established at the DOE level to coordinate between the Office of Science and NNSA to procure three leadership computing systems for 2017–2018 delivery. One RFP leads to two NRE contracts and three computer procurement contracts. Two architectural paths will be selected for risk reduction. Both NRE contracts will be jointly managed by the three labs. Each lab manages and negotiates its own computer procurement contract.

We're awarding contracts in 2014 for machines to be delivered several years later, and we are prepared to see an evolution in the hardware during this period.

Why are we doing CORAL? Having diverse systems promotes price competition.

High-level requirements include 4x speedup on scalable benchmarks (full system size) and 6x speedup on throughput benchmarks (smaller jobs, but we run many copies). We're targeting minimum peak performance of 100PF, 4PB of memory, maximum 20MW, and mean time between application failure requiring human intervention of six days or better. There will also be data-centric capabilities, and delivery will be in 2017, with 2018 acceptance.

All three labs agreed on a common set of technical requirements. The CORAL benchmark suite consists of:

- Scalable science benchmarks
- Throughput benchmarks

- Data-centric benchmarks
- Skeleton benchmarks (network performance, threading overhead, etc.)
- Microbenchmarks (small codes to help vendors predict larger-scale performance)

The evaluation process has two steps. There will be eight teams of technical experts, with eight people on each team (e.g., project management, system software, and more).

Evaluation criteria:

- DOE mission requirements
- Technical proposal excellence
- Feasibility
- Diversity
- Overall price
- Supplier attributes
- Risk

The NRE contracts are coupled to build contracts. Evaluation results will not be announced until contracts have been negotiated (expected in late FY14).

"Preparing Applications for Next-Generation IO/Storage": Gary Grider, LANL

Drivers for change include the following:

- Scale (machine and storage sizes)
- Data volumes
- Bursty behavior
- Technology trends
- Economic trends

IO people have a small bag of tricks:

- Burst buffer economics say that at exascale, you have to meet two requirements: 100TBps burst from 1 billion processing elements, and 1EB of scratch capacity (30ish memories).
- Buying flash for capacity is inexpensive. Buying disk for capacity is also expensive, but with disk for capacity, you get bandwidth for free.
- The Trinity system will have 2–4PB of DRAM, 5–12PB of flash/burst buffer, and 100PB of disk or thereabouts. There is opportunity for in-transit analysis (before data goes to tape).
- Unlimited archiving will become cost prohibitive. Capacity is no longer the sole cost driver for archive as it was for 25 years. Bandwidth is now a major contributor to the TCO of archives.

- We need to figure out how to allow the storage system to fail. This challenge is in the FastForward Project. For reliability, a strategy is to put in checksums all along the way.

Regarding how your apps will change, there will be a need for non-blocking APIs, and more workflows will need to consider the expense of going to disk and especially to archive.

"The CREATE Ships Navy-Enhanced Sierra Mechanics (NESM) Project": Adam Hapij, Weidlinger Associates

Our sponsor is the Naval Surface Warfare Center, Carderock Division. NESM is intended for coupled simulation focused on ship responses to weapons effects. This involves coupled media-structure interactions (CFD and structural dynamics). We are a civil engineering firm focused on buildings, bridges, infrastructure, and applied science. We have long-standing relationships with U.S. Navy agencies.

My topic today is first-of-class shock trials for ships that have to be battle ready. Charges are deployed near the ships in offshore tests. This testing is expensive and takes the ships out of service for a long time. The Full Ship Shock Trial program is based on taking COTS technology (seismic air guns) to use for non-explosive shock trial testing. This testing can be done quickly, repeatedly, and in any weather.

A lot of modeling and simulation was needed using HPC. We modeled the discharge of the high-pressure air and did uncertainty quantification studies of the effects on the ships of the pressure discharged from the seismic air guns. The air gun source modeling uses the GEMINI code. So we are simulating virtual tests. It took 54 CPU years just for loading the source model library. We were very pleased with the HPC capabilities. On the uncertainty quantification side, we do full-scale-coupled fluid-structure interaction simulations with millions of elements and DOF, using Monte Carlo methods. The size and scale exceed the computational capabilities.

We performed 12 complete UQ studies in our 2012 test series, totaling 56 CPU years of calculation time. The HPC resources permit solutions to problems that were intractable a decade ago. It still takes longer to process the data than to fire off another air gun test.

LEARN MORE

Related Research

Additional research from IDC in the technical computing hardware program includes the following documents:

- *Worldwide High-Performance Data Analysis 2013–2017 Forecast* (IDC #241315, June 2013)
- *Experiences in HPC: HPC User Forum, April 29–May 1, 2013, Tucson, Arizona* (IDC #241455, June 2013)
- *HPC in Aerospace, Astrophysics, and Astronomy: HPC User Forum, April 29–May 1, 2013, Tucson, Arizona* (IDC #241451, June 2013)

- *Top Issues for HPC Sites: HPC User Forum, April 29–May 1, 2013, Tucson, Arizona* (IDC #241463, June 2013)
- *Potential Disruptive Technologies Panel: HPC User Forum, April 29–May 1, 2013, Tucson, Arizona* (IDC #241452, June 2013)
- *Advanced Visualization: HPC User Forum, April 29–May 1, 2013, Tucson, Arizona* (IDC #241446, June 2013)
- *Worldwide Technical Computing Server 2013–2017 Forecast* (IDC #241154, May 2013)
- *Supercomputers Exceed 50% of the HPC Server Market in 2012* (IDC #240426, April 2013)
- *Changing Market Dynamics: HPC Meeting Big Data and IDC's Projected Evolution of the Market* (IDC #240365, March 2013)
- *Livermore Lab Expands Industry Partnerships: Economic Security Is Vital for National Security* (IDC #240232, March 2013)
- *Worldwide Technical Computing 2013 Top 10 Predictions* (IDC #239421, February 2013)
- *High-Performance Data Analysis at NASA JPL* (IDC #238254, December 2012)
- *Advanced Research at the South Ural State University Supercomputing Center* (IDC #238225, December 2012)
- *The Economic Value of HPC in Science and Industry: HPC User Forum, September 2012, Dearborn, Michigan* (IDC #237182, October 2012)
- Big Data in HPC: HPC User Forum, September 2012, Dearborn, Michigan (IDC #237180, October 2012)
- *How Nations Are Applying High-End Petascale Supercomputers for Innovation and Economic Advancement in 2012* (IDC #236341, August 2012)
- *Tokyo Institute of Technology: Global Scientific Information and Computing Center* (IDC #236243, August 2012)
- *Shanghai Supercomputer Center* (IDC #236245, August 2012)
- *Petascale Supercomputing at the University of Tokyo* (IDC #236292, August 2012)
- *HPC Application Leadership at the Supercomputing Center of Chinese Academy of Sciences* (IDC #236281, August 2012)
- *India Broadening Access to Supercomputing* (IDC #lcUS23621912, July 2012)
- *Institute of Process Engineering, Chinese Academy of Sciences* (IDC #236204, July 2012)
- *How Fujitsu Built the World's Fastest Supercomputer in Record Time and Ahead of Schedule* (IDC #235733, July 2012)
- *Europe Sharpens Its Focus on Exascale Computing* (IDC #lcUS23555112, June 2012)
- *IBM Returns Supercomputer Crown to United States* (IDC #lcUS23547812, June 2012)
- *Potential Disruptive Technologies: HPC User Forum, April 2012, Richmond, Virginia* (IDC #234742, May 2012)

- *The Broader HPC Market: Servers, Storage, Software, Middleware, and Services* (IDC #234682, May 2012)
- *HPC End-User Site Update: RIKEN Advanced Institute for Computational Science* (IDC #233690, March 2012)
- *National Supercomputing Center in Tianjin* (IDC #233971, March 2012)
- Worldwide Data Intensive–Focused HPC Server Systems 2011–2015 Forecast (IDC #232572, February 2012)

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-insights-community.com
www.idc.com