

LLNL, Intel, Cray produce big data machine

November 5 2013, by Donald B Johnston



Catalyst is a unique high performance computing (HPC) cluster that will serve research scientists and provide a proving ground for new HPC and Big Data technologies and architectures. It was recently installed at Lawrence Livermore National Laboratory.

(Phys.org) —Lawrence Livermore National Laboratory in partnership with Intel and Cray, today announced a unique high-performance computing (HPC) cluster that will serve research scientists at all three institutions and provide a proving ground for new HPC and Big Data technologies and architectures.

"As the name implies, Catalyst aims to accelerate HPC simulation and

big data innovation, as well as collaborations between the three institutions," said Matt Leininger, deputy of Advanced Technology Projects for LLNL. "The partnership between Intel, Cray and LLNL allows us to explore different approaches for utilizing large amounts of high performance [non-volatile memory](#) in HPC simulation and Big Data analytics."

The Catalyst resource, a Cray CS300 [cluster](#) supercomputer, will be shared between the three partners with access rights based on level of investment. System access will be managed through LLNL's High Performance Computing Innovation Center (HPCIC), whose mission is to work with industrial partners in the development of computing solutions for America to compete effectively in the 21st century global economy.

Delivered to LLNL in late October, Catalyst is expected to be available for limited use this month and general use by December. The Catalyst cluster consists of two scalable units (SUs), and represents an upgrade from the Appro clusters acquired under the Tri-Lab Linux Capacity Cluster (TLCC-2) procurement of a few years ago (Appro has since been acquired by Cray). TLCC aggregates the HPC capacity computing needs of the three weapons laboratories that serve the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program - Lawrence Livermore, Los Alamos and Sandia national laboratories - to procure commodity cluster systems more cost effectively.

The 150 teraflop/s (trillion floating operations per second) Catalyst cluster has 324 nodes, 7,776 cores and employs the latest-generation 12-core Intel Xeon E5-2695v2 processors. Catalyst runs the NNSA-funded Tri-lab Open Source Software (TOSS) that provides a common user environment across NNSA Tri-lab clusters. Catalyst features include 128 gigabytes (GB) of dynamic random access memory

(DRAM) per node, 800 GB of non-volatile memory (NVRAM) per compute node, 3.2 terabytes (TB) of NVRAM per Lustre router node, and improved cluster networking with dual rail Quad Data Rate (QDR-80) Intel TrueScale fabrics. The addition of an expanded node local NVRAM storage tier based on PCIe high-bandwidth Intel Solid State Drives (SSD) allows for the exploration of new approaches to application checkpointing, in-situ visualization, out-of-core algorithms and big data analytics.

"Big Data unlocks an entirely new method of discovery by deriving the solution to a problem from the massive sets of data itself. To research new ways of translating Big Data into knowledge, we had to design a one-of-a-kind system," said Raj Hazra, Intel vice president and general manager of the Technical Computing Group. "Equipped with the most powerful Intel processors, fabrics and SSDs, the Catalyst cluster will become a critical tool, providing insights into the technologies required to fuel innovation for the next decade."

The Catalyst architecture is expected to provide insights into the kind of technologies the ASC program will require over the next 5-10 years to meet high performance simulation and Big Data computing mission needs. The increased storage capacity of the system (in both volatile and nonvolatile memory) represents the major departure from classic simulation-based computing architectures common at DOE laboratories and opens new opportunities for exploring the potential of combining floating-point-focused capability with data analysis in one environment. Consequently, the insights provided by Catalyst could become a basis for future commodity technology procurements.

HPCIC at LLNL will offer access to Catalyst and the expected Big Data innovations it enables as new options for its ongoing collaborations with American companies and research institutions. The machine's expanded DRAM and fast, persistent NVRAM are well suited to big data problems

(i.e, bioinformatics, business analytics, machine learning and natural language processing), as well as meeting the increasingly demanding simulation requirements of ASC. Catalyst should extend the range of possibilities for the processing, analysis and management of the ever larger and more complex datasets that many areas of business and science now confront.

"We expect this collaboration to serve as a model for the kind of research and development that promotes HPC innovation," said Fred Streitz, director of the HPCIC. "Such innovation is critical to maintaining U.S. leadership in HPC."

"Cray firmly believes that collaboration is a vital element of introducing new innovations to the worlds of Big Data and supercomputing, and we are honored that a Cray CS300 cluster supercomputer is the foundation of this important project with our partners at Intel and Lawrence Livermore," said Daniel Kim, senior vice president and general manager of cluster solutions at Cray.

Provided by Lawrence Livermore National Laboratory

Citation: LLNL, Intel, Cray produce big data machine (2013, November 5) retrieved 7 May 2024 from <https://phys.org/news/2013-11-llnl-intel-cray-big-machine.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
