

First-of-a-kind supercomputer at Lawrence Livermore available for collaborative research

May 7 2014, by Donald B Johnston



The Catalyst supercomputer at Lawrence Livermore employs a Cray CS300 architecture modified specifically for data-intensive computing. The system is now available for collaborative research with industry and academia.

Catalyst, a first-of-a-kind supercomputer at Lawrence Livermore National Laboratory (LLNL), is available to industry collaborators to test big data technologies, architectures and applications.

Developed by a partnership of Cray, Intel and Lawrence Livermore, this

Cray CS300 [high performance computing](#) (HPC) cluster is available for collaborative projects with industry through Livermore's High Performance Computing Innovation Center (HPCIC).

"Over the next decade, global data volume is forecasted to reach more than 35 zettabytes," (a zettabyte is a trillion gigabytes) said Fred Streitz, director of the HPCIC. "That enormous amount of unstructured data provides an opportunity. But how do we extract value and inform better decisions out of that wealth of raw information?"

A resource for the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) program, 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 (Los Alamos, Sandia and Lawrence Livermore national labs).

"The opportunity to work with Cray and Intel to design and deploy Catalyst, a novel computing platform optimized for HPC-end applications, has been very exciting," said Robin Goldstone, Livermore HPC Solutions architect. "We have modified the Cray CS300 architecture in ways that make Catalyst an outstanding HPC platform for data-intensive computing."

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 check-pointing, in-situ visualization, out-of-core algorithms and big data analytics. NVRAM is familiar to anyone who uses USB sticks or an MP3 player; it is simply memory that is persistent and that remains on files even when the power is off, hence "non-volatile."

Deployed in October 2013, the Catalyst architecture already has begun to provide insights into the kind of technologies the ASC program will require over the next decade 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. The machine's expanded DRAM and fast, persistent NVRAM are well suited to a broad range of big data problems including bioinformatics, business analytics, machine learning and natural language processing.

Jonathan Allen, a Lawrence Livermore bioinformatics scientist, is working on new methods to rapidly detect and characterize pathogenic organisms such as viruses, bacteria or fungi in a biological sample.

"We're working on developing scalable analysis tools for next generation sequencing, in particular metagenomic sequencing," Allen said. "By comparing short genetic fragments in a query dataset against a large searchable index of genomes, we can make determinations about the potential threat an organism poses to human health."

Traditional technologies and storage limitations made it challenging to rapidly search a database of reference genomes as more organisms were sequenced and more variants in the population of an organism were included. With Catalyst's unique architecture, Allen and his team are able to store very large reference databases of genomes in memory and

execute expansive analyses with higher resolution.

"We were able to do a metagenomic analysis on a fairly large sample in several hours on a single desktop. With Catalyst, we can process many hundreds of equal size in about the same time."

Catalyst also will serve to host very large models for video analytics and machine learning.

"YouTube claims that 100 hours of video are uploaded to its website every minute," explained Doug Poland, computational engineer working on video analytics. "As the fastest-growing type of content on the Internet, consumer-produced videos are a wealth of information about the world that's essentially untapped."

Yet current tools are unable to search through the richness of video elements such as visual, audio and motion, and associated metadata like semantic tags and geo-coordinates. Poland and his team are looking to build more complex models that consider the sum of those features, and that can be recognized in real-time for user-specific search needs.

"Catalyst allows us to explore entirely new deep learning architectures that could have a huge impact on video analytics as well as broader application to big data analytics."

"Our purpose is to use Catalyst as a test bed to develop optimization strategies for data-intensive computing," Streitz said. "We believe that advancing [big data](#) technology is a key to accelerating the innovation that underpins our economic vitality and global competitiveness."

Provided by Lawrence Livermore National Laboratory

Citation: First-of-a-kind supercomputer at Lawrence Livermore available for collaborative research (2014, May 7) retrieved 17 April 2024 from <https://phys.org/news/2014-05-first-of-a-kind-supercomputer-lawrence-livermore-collaborative.html>

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.