

UC San Diego

UC San Diego News Center

August 24, 2015 | By Jan Zverina

SDSC Awarded One-Year Extension for Gordon Supercomputer

National Science Foundation extends five-year term of data-intensive cluster

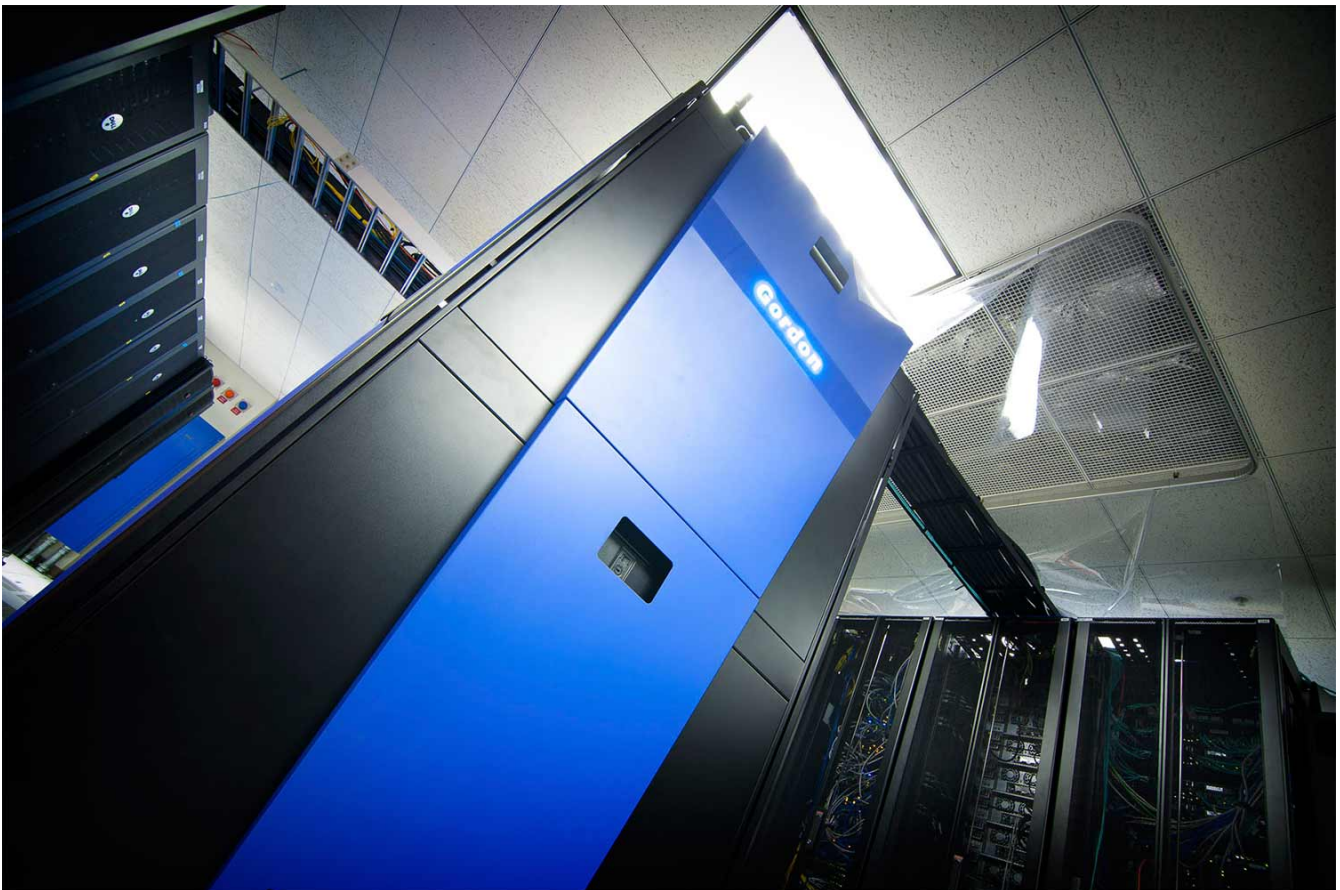


Photo by Erik Jepsen/UC San Diego Publications

The National Science Foundation (NSF) has awarded the San Diego Supercomputer Center (SDSC) at the University of California, San Diego, a one-year extension to continue operating its *Gordon* supercomputer, providing continued access to the cluster for a wide range of researchers with data-intensive projects.

The result of a five-year, \$20 million NSF grant awarded in late 2009, *Gordon* entered operations in early 2012 as one of the 50 fastest supercomputers in the world at the time – and the only one to employ massive amounts of flash-based memory, making it many times faster than conventional high-performance computing (HPC) systems for some data-intensive analyses.

With 300 trillion bytes of flash memory and 64 I/O nodes, *Gordon* is ideal for researchers who need to sift through tremendous amounts of data. In effect, *Gordon* was designed to do for scientific research what Google does for web searches. Full specifications of SDSC's *Gordon* [can be found here](#). By the end of 2014, 1,098 research projects using *Gordon* were awarded among 762 principal investigators at numerous academic institutions across the U.S. including UC San Diego and the UC system.

“We are delighted that *Gordon* will be available to researchers for another year, especially to those who are working in data-intensive domains such as social and political science, earth sciences, molecular biology and chemistry, genomics, and experimental high energy physics,” said SDSC Director Michael Norman, who also is the principal investigator for the *Gordon* supercomputer project. “*Gordon* has served the research community very well, and the NSF’s decision provides scientists continued access to what remains a state-of-the-art system.”

The one-year extension, until August 31, 2016, is being made possible through a combination of unspent funds and additional funding from the NSF. The extension will benefit researchers, the majority of whom access the system through the NSF’s [XSEDE \(eXtreme Science and Engineering Discovery Environment\) program](#), which comprises the most advanced collection of integrated digital resources and services in the world.

“Extending *Gordon* provides a number of important benefits to XSEDE and its research community,” said Bob Sinkovits, director of the Scientific Applications Group in SDSC’s Data-Enabled Scientific Computing Division. “*Gordon* is a primary resource for several widely used Science Gateways that serve large communities of users. These include [CIPRES](#), for phylogenetic inference; [GridChem](#) for computational chemistry; and [UltraScan](#), for analysis of ultracentrifuge data. *Gordon* also hosts a number of novel projects ranging from the biomedical sciences to analysis of the Internet, which require persistent access to flash storage and are not currently handled by other XSEDE resources.”

Gordon to Support Dedicated I/O Node Projects

“Since *Gordon* was initially proposed and then configured as a data-intensive system rather than a traditional compute-intensive resource, we had significant latitude to explore alternate modes of allocation,” said Wayne Pfeiffer, SDSC Distinguished Scientist and *Gordon* Project Manager. “One of the most successful alternatives was to carve out small fractions of the system and offer them as dedicated resources to projects that could make effective use of *Gordon*’s solid state drives, or SSDs.”

Specifically, a typical allocation would involve a single I/O node plus some fraction of the compute nodes connected to the same switch. Below are two examples of how *Gordon*’s unique configuration enabled advances in significant large-scale, data-rich projects.

Particle Physics and the Search for Dark Matter

One dedicated I/O node project was also one of *Gordon*’s most data-intensive tasks: rapidly processing raw data from almost half a billion particle collisions as part of a project to help define the future research agenda for the Large Hadron Collider (LHC). After discovering the Higgs boson at the LHC, physicists now hope to find new particles that herald the presence of dark matter. To do so, they recently increased by several fold the data capture rate of the Compact Muon Solenoid (CMS), one of two large detectors attached to the LHC.

To help reduce the data backlog, 125 TB (terabytes) of data representing 400 million events were transferred to SDSC’s *Gordon* during a four-week period in 2013 and processed using 1.5 million core hours. The analysis pipeline was complex, involving compute nodes for event processing, a dedicated I/O node for serving 750 GB (gigabytes) of calibration data from flash memory, SDSC’s *Data Oasis* parallel file system running Lustre to stage the input and output, and a high-speed network to move data between SDSC, CMS, and the Open Science Grid.

Topographical Mapping and Analysis

The NSF-funded OpenTopography Facility provides efficient online access to high-resolution LIDAR (Light Detection And Ranging) topography data, online processing tools, and derivative products. Users interact via the OpenTopography portal to sub-select LIDAR point cloud data and run processing algorithms on their area of interest to produce custom derivative products, such as digital elevation models (DEMs).

A performance benchmark using the local binning algorithm and a test case of 208 million LIDAR returns was run on a dedicated *Gordon* I/O node to compare performance with OpenTopography's commodity clusters where the algorithm previously ran. Using this dedicated I/O node reduced run times of massive concurrent out-of-core processing jobs by a factor of 3x for single job access to a given data set, and by 20x for multiple concurrent jobs accessing that data set.

An XSEDE allocation of a dedicated *Gordon* I/O node and 16 compute nodes enabled the integration of compute-intensive algorithms such as the TauDEM hydrology suite into the OpenTopography workflow. "This capability has now been deployed on OpenTopography's XSEDE-allocated resources on *Gordon* and is available to the wider earth sciences community via the OpenTopography portal, paving the way for new types of analysis on large topographic data sets," said SDSC's Pfeiffer.

MEDIA CONTACT

Jan Zverina, 858-534-5111, jzverina@sdsc.edu

UC San Diego's [Studio Ten 300](#) offers radio and television connections for media interviews with our faculty, which can be coordinated via studio@ucsd.edu. To connect with a UC San Diego faculty expert on relevant issues and trending news stories, visit <https://ucsdnews.ucsd.edu/media-resources/faculty-experts>.