

SDSC's 'Comet' supercomputer enters early operations phase

April 30 2015



SDSC's Comet Launch Team. SDSC Director Mike Norman (yellow shirt) is Comet's PI; SDSC Deputy Director Richard Moore (behind Mike) is co-PI of the petascale supercomputer project. Photos by Ben Tolo/SDSC

Comet, a new petascale supercomputer designed to transform advanced scientific computing by expanding access and capacity among traditional as well as non-traditional research domains, has transitioned into an early

operations phase at the San Diego Supercomputer Center (SDSC) at the University of California, San Diego.

Comet is the result of a National Science Foundation Award currently valued at \$21.6 million including hardware and operating funds. The new cluster is capable of an overall peak performance of more than two petaflops, or two quadrillion operations per second.

"Comet is really all about providing high-performance computing to a much larger research community - what we call 'HPC for the 99 percent' - and serving as a gateway to discovery," said SDSC Director Michael Norman, the project's principal investigator. "Comet has been specifically configured to meet the needs of researchers in domains that have not traditionally relied on supercomputers to solve their problems."

Comet joins SDSC's Gordon supercomputer as another key resource within the NSF's XSEDE (eXtreme Science and Engineering Discovery Environment) repertoire, which comprises the most advanced collection of integrated digital resources and services in the world. Researchers may request allocations on Comet via XSEDE. More information is available [here](#).

Comet was designed to provide a solution for emerging research requirements often referred to as the 'long tail' of science, which describes the idea that the large number of modest-sized, computationally-based research projects still represents, in aggregate, a tremendous amount of research and resulting scientific impact and advance.

"One of our key strategies for Comet has been to support modest-scale users across the entire spectrum of NSF communities, while welcoming research communities that are not typically users of more traditional HPC systems, such as genomics, the social sciences, and economics,"

said SDSC Deputy Director Richard Moore, a co-PI for the new system.

"Based on Comet's early performance, we are confident in its ability to address a broader set of research topics and communities than past systems," said Irene Qualters, division director for Advanced Cyberinfrastructure at NSF. "We congratulate SDSC and UC San Diego on their accomplishment and look forward to Comet's full deployment."

A key strategy for Comet is to reach large communities of users via Science Gateways. A Science Gateway is a community-developed set of tools, applications, and data that is integrated through a web-based portal or a suite of applications. Gateways provide scientists access to many of the tools used in cutting-edge research - telescopes, seismic shake tables, supercomputers, sky surveys, undersea sensors, and more - and connect often diverse resources in easily accessible ways that save researchers and institutions both time and money. Moreover, researchers can focus on their scientific goals without having to know how supercomputers and other data cyberinfrastructures work.

"The variety of hardware and support for complex, customized software environments will be of particular benefit to Science Gateway developers," said Nancy Wilkins-Diehr, an associate director of SDSC and co-director of XSEDE's Extended Collaborative Support Services. "We now have more than 30 such Science Gateways running on XSEDE, each designed to address the computational needs of a particular community such as computational chemistry, atmospheric science or the social sciences."

Comet is a Dell-integrated cluster using Intel's Xeon® Processor E5-2600 v3 family, with two processors per node and 12 cores per processor running at 2.5GHz. Each compute node has 128 GB (gigabytes) of traditional DRAM and 320 GB of local flash memory. Since Comet is designed to optimize capacity for modest-scale jobs,

each rack of 72 nodes (1,728 cores) has a full bisection InfiniBand FDR interconnect from Mellanox, with a 4:1 over-subscription across the racks. There are 27 racks of these compute nodes, totaling 1,944 nodes or 46,656 cores.

"Comet will greatly expand the accessibility and impact of high-performance computing to the nation's open science researchers by offering a comprehensive set of capabilities in one integrated system, bringing research capacity and access to critical technology and to the long tail of science," said John Mullen, vice president and general manager of Dell's North America Commercial Sales.

In addition, Comet has 36 GPU nodes, each with four NVIDIA GPUs (graphic processing units) and two Intel processors, and will soon have four large-memory nodes, each with four Intel processors and 1.5 TB of memory. The GPUs and large-memory nodes are for specific applications such as visualizations, molecular dynamics simulations, and de novo genome assembly.

Comet users will also have access to 7.6 PB (petabytes) of Lustre-based high-performance storage, with 200 GB/s bandwidth to the cluster. It is based on an evolution of SDSC's Data Oasis storage system, with Aeon Computing as the primary storage vendor. This system is split between a scratch file system, and an allocated file system for persistent storage. There are significant improvements in these second-generation Data Oasis file systems, beginning with a ground-up design based on ZFS-backed storage for both performance and data integrity. ZFS continually monitors and repairs low-level blocks of data stored on disk, avoiding the silent data corruption that can occur with storage as large as Comet's. Comet will have a second level of data reliability as well, since the first-generation Data Oasis servers are being consolidated and re-deployed to create a 'nearline' replica of the active file systems.

"With its latest Lustre file system SDSC is leading the way" said Jeff Johnson, co-founder of Aeon Computing. "SDSC and Aeon Computing collaborated on the design of this new Lustre file system and it leads as one of the first large-scale Lustre file systems that make full use of ZFS direct to disk drives without any hardware RAID technology."

Comet will feature a new 100 Gbps (Gigabits per second) connectivity to Internet2 and ESNet, allowing users to rapidly move data to SDSC for analysis and data sharing, and to return data to their institutions for local use.

Comet replaces Trestles, which entered production in early 2011 under an earlier NSF grant to provide researchers not only significant computing capabilities, but to allow them to be more computationally productive. Trestles and Gordon are the leading Science Gateway systems in the XSEDE portfolio, with more than 1,200 users per month accessing those systems through the popular CIPRES phylogenetics gateway alone.

"Trestles users have spanned a wide range of domains, including astronomy, biophysics, climate sciences, computational chemistry, and material sciences, and we expect that Comet will attract researchers from many more domains," added Moore.

Provided by University of California - San Diego

Citation: SDSC's 'Comet' supercomputer enters early operations phase (2015, April 30) retrieved 3 May 2024 from <https://phys.org/news/2015-04-sdsc-comet-supercomputer-early-phase.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>
--