

# Supercomputer project to study solid-state drives under data-intensive workloads

November 11 2015

---



Seagate solid state drives being used in SDSC's Comet supercomputer will be analyzed to see how they are used in domains that rely on advanced computation and storage, including genomics and the social sciences. Credit: Ben Tolo, SDSC.

The addition of the 800GB (gigabyte) Seagate SAS SSDs will significantly boost Comet's data analytics capability by expanding its node-local storage capacity for data-intensive workloads. Pairs of the drives will be added to all 72 compute nodes in one rack of Comet, alongside the existing SSDs. This will bring the flash storage in a single node to almost 2TB (terabytes), with total rack capacity at more than 138TB.

The installation process of the Seagate drives began in October under a donation arrangement with SDSC and its Center for Large Scale Data Systems Research, an industry-university collaboration that focuses on issues including "big data" system architectures and software, analytics, and performance, with the goal of understanding the full value that can be extracted from voluminous amounts of data now becoming available to organizations. User access to the drives will begin before January 2016.

Under the partnership, Seagate and SDSC/CLDS are deploying a lightweight framework for extracting metrics suitable for the analysis of data I/O patterns and overall drive performance. These results and other metrics will be used to further develop best practices and reference HPC (high-performance computing) architectures, while leading to more precise analyses of SSD performance in operational HPC workloads.

"The addition of the Seagate solid state drives on Comet continues the work we pioneered in the area of system versatility around flash-based SSDs on our [Gordon supercomputer](#)," said SDSC Director Michael Norman, also principal investigator for the Comet project. "It complements Comet's other dimensions, such as its fast Lustre parallel file storage systems and large memory nodes. With such a wide range of workflows in both traditional and emerging science domains such as genomics, the greater research community will benefit from these heterogeneous but integrated capabilities."

The new drives will also extend the abilities of Comet's upcoming virtualized HPC clusters. "Currently, some virtual machines on Comet can have large disk images that take advantage of the fast local storage on the compute nodes hosting them," said Rick Wagner, SDSC's manager of HPC systems. "Groups using virtual machines will be able store more data inside of their [virtual machines](#), closer to their custom application stacks."

Comet is capable of an overall peak performance of almost two petaflops, or two million billion operations per second. It has the ability to perform 10,000 research jobs simultaneously. Like the tail of a comet, SDSC's newest HPC cluster is intended to serve what's called the 'long tail' of science: the idea that the large number of modest-sized computationally-based research projects represent, in aggregate, a tremendous amount of research that can yield scientific discovery.

"The goal of our partnership with SDSC is to inform the wider HPC community via papers and workshops on how to select the most appropriate, high-performance components suitable to their architectures and workloads, while gaining insight into how Seagate SSDs are used in domains that are relying on advanced computation and storage, such as genomics and the social sciences," said Tony Afshary, director of ecosystem solutions and marketing for flash at Seagate.

The result of a National Science Foundation grant worth nearly \$24 million including hardware and operating funds, Comet is available for use by U.S. academic researchers through the NSF's eXtreme Science and Engineering Discovery Environment (XSEDE) program, a national collection advanced, integrated digital resources and services.

Provided by University of California - San Diego

Citation: Supercomputer project to study solid-state drives under data-intensive workloads (2015, November 11) retrieved 3 May 2024 from <https://phys.org/news/2015-11-supercomputer-solid-state-data-intensive-workloads.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.