UC San Diego's Supercomputer Center Boosts Storage Capacity to Mind-Boggling Numbers

Paul K. Mueller

November 3, 2006

If the Industrial Age relied on ore, the Digital Age relies on storage.

None of our now-necessary devices, from the most fearsome research-computing arrays to run-of-the-mill office computers to cell-phones to iPods, can work without storage. That's why Richard Moore, director of Production Systems at the San Diego Supercomputer Center (SDSC), smiles as he ponders the new IBM tape drives being added to the storage "silos" in the center's already crowded computer room.

SDSC already has six storage silos, each of which holds about 6,000 tapes. With the new tape drives and media (IBM System Storage TS1120 tape drives with the new industry-leading 700-gigabyte tape media), Moore and his colleagues can now store 25 petabytes - that's 25 million billion bytes - an upgrade from SDSC's previously phenomenal storage capacity of six petabytes.

That will give SDSC and its host institution, the University of California San Diego, more storage capacity than any other educational institution in the world.

"As an institution which supports not only UC San Diego but also a national community of academic researchers, SDSC has an obligation to provide the increasing amounts of storage capacity these collaborations will require," said Moore. "This latest upgrade with added IBM technology helps us stay ahead of the projections - all of which show an exponentially increasing demand for storage in the future."

For example, Moore said, SDSC is currently serving more than 10,000 researchers at 300 academic, government and industrial institutions in the United States and around the world. "Today, these scientists and engineers increasingly rely on the availability of globally accessible data and the associated cyberinfrastructure tools to drive research and education," he says.

And, he adds, SDSC intends to stay ahead of that curve.

The supercomputer center - now in the midst of an ambitious expansion project on the sunny UC San Diego campus - had an eye-glazing amount of capacity before the upgrade. Now, its storage capacity runs into numbers that are mind-boggling.

Moore uses several analogies to communicate the vast amounts of information SDSC's computers can store and make available. "For reference, the digital equivalent of all the printed materials in the Library of Congress is about 20,000 gigabytes; this represents less than 0.1 percent of our capacity," he says.

For students, he explains it this way: "If every high school student in the U.S. had a gigabyte of music on his or her iPod, all their music - billions of songs - could fit in our archive; although," he says with a smile, "I expect that there's a lot of redundancy in that data."

SDSC operates powerful high-end computing resources led by DataStar, a 15.6 teraflop IBM Power4+ supercomputer with total aggregate memory of 7.3 terabytes. DataStar is ranked among the top supercomputers in the world and is used for large-scale, data-intensive scientific research applications. In addition, SDSC was the first U.S. academic institution to install an IBM Blue Gene system, and will soon triple its size to 17.1 teraflops.

"SDSC is a prime example of the leadership that IBM continues to demonstrate in tape storage," said Kristie Bell, vice president, System Storage Marketing, IBM. "No other vendor can offer the end-to-end solutions that IBM is able to offer, and no one comes close to matching IBM in innovations such as tape virtualization and tape encryption."

SDSC also serves as the data-intensive site lead in the NSF-funded TeraGrid, a multiyear effort to build and deploy the world's first large-scale and production grid infrastructure for open scientific research. SDSC hosts a 4.4-teraflop IA-64 Linux cluster, as well as 220 terabytes of a global file system that is mounted across the TeraGrid computing systems. SDSC is connected to the other national TeraGrid partners by a 10-gigabyte-per-second cross-country high-speed network.

This focus on data cyberinfrastructure, Moore said, "provides a broad and flexible array of integrated technologies to support increasingly challenging, large-scale and collaborative scientific endeavors."

Two examples he cites involve SDSC support for the Southern California Earthquake Center (SCEC) and the National Virtual Observatory (NVO) collection. For the former, supercomputing resources allow researchers to use massive amounts of geological and historical data to simulate, and better understand, major earthquakes - and learn how to mitigate structural damage. For the latter, SDSC resources help the NVO combine over 100 terabytes of data from 50 ground- and space-based telescopes and instruments to create a comprehensive picture of the heavens.

"Our work with IBM to significantly upgrade our storage capacity is vital to a myriad such collaborations," said Moore. "We're confident that this project, and many others, will keep our institutions at the very forefront of data storage and cyberinfrastructure for years to come."

Media Contact: Paul K. Mueller, 858-534-8564

Comment: Richard L. Moore, 858-822-5457