

SDSC's High-Performance Storage System Achieves Petabyte Milestone

Data storage reaches one million megabytes, continues to grow 10 terabytes per month

May 17, 2004

Greg Lund

The San Diego Supercomputer Center (SDSC) announced today that its High-Performance Storage System (HPSS) has reached the milestone of one petabyte of stored data - the equivalent of 500 million pages of text, which would fill the Library of Congress more than eight times over. The world's largest academic production archival storage system, SDSC's HPSS is capable of warehousing up to six petabytes of data so that users can manage, access, and utilize data-intensive applications at top speed without interruption.

The HPSS installation at SDSC is a centralized file management and storage system that serves as a reliable, long-term file space for SDSC's local and national users. Researchers who use SDSC's computing facilities often require huge amounts of storage space for the data they generate from experiments, computer simulations and field observations. Because storage space is limited on their local computer systems and on most SDSC computing platforms, many researchers transfer their data to SDSC's HPSS.

"The rapid growth of our HPSS program was spawned by SDSC's razor focus on providing data-intensive computing to users across the country," said Bryan Banister, assistant Director of SDSC's High-End Computing group. "As the premier site for the nation's cyber-infrastructure, we store and serve massive amounts of information to researchers via the Internet, high-speed networks such as Internet2, and advanced computational grids such as the TeraGrid."

HPSS uses specialized software modules to send large data files, such as detailed images and massive databases, at top speed between high-performance computers, workstation clusters, and storage libraries. Currently, SDSC's HPSS is capable of transferring more than one gigabyte per second, making it possible to send data-intensive files such as hyper spectral images, classroom videos and simulation data from the Southern California Earthquake Center.

SDSC's HPSS also plays a vital role in supporting massive data archives and data libraries. For example, brain researchers doing positron-emission tomography and functional magnetic resonance imaging of living human subjects can acquire a terabyte -- one thousand gigabytes -- of data in a single one-hour session; storing that information on the HPSS makes it accessible to scientists at many institutions. In addition, The National Virtual Observatory project, a far-reaching effort that is intended to eventually put online almost all astronomical observations, photographs, and catalogs, currently stores approximately 20 terabytes of sky survey data on SDSC's HPSS.

Information is added to the HPSS at an average rate of 10 terabytes per month. Live daily HPSS statistics are available at http://www.sdsc.edu/Storage/statistics/hpss_stats.cgi?1.

About SDSC The mission of the San Diego Supercomputer Center (SDSC) is to innovate, develop, and deploy technology to advance science. SDSC is involved in an extensive set of collaborations and activities at the intersection of technology and science whose purpose is to enable and facilitate the next generation of scientific advances. Founded in 1985 and primarily funded by the National Science Foundation (NSF),

SDSC is an organized research unit of the University of California, San Diego. With a staff of more than 400 scientists, software developers, and support personnel, SDSC is an international leader in data management, grid computing, biosciences, geosciences, and visualization. For more information, see www.sdsc.edu.

Media contact: Greg Lund, SDSC, (858) 534-8314 Ashley Wood, SDSC, (858) 534-8363

