BLACkLIGHT & THE DATA SUPERCeLL

The world’s largest shared-memory supercomputer, PSC’s Blacklight, has helped to open XSEDE resources to many non-traditional HPC projects.

As a leading partner in XSEDE, the most powerful collection of integrated digital resources and services in the world, PSC helps to shape the vision and progress of U.S. science and engineering.

The Data Supercell

PSC this year deployed a disk-based file repository and data-management system, the Data Supercell (DSC). This innovative technology provides significant advantages over tape-based archiving. The PSC development team – led by Nowoczynski, Jared Yanovich, Zhilai Zhang, Jason Sommerfield, J. Ray Scott, and Michael Levine – exploited increasing cost-effectiveness of commodity disk technologies, and adapted sophisticated PSC filesystem software (called SLASH2) to use with DSC. A patent application is under review.

“The Data Supercell is a unique technology, building on the cost-effectiveness of disk and the capabilities of PSC’s SLASH2 file system,” said Michael Levine and Ralph Roskies, PSC scientific directors. “It enables more efficient, flexible analyses of very large-scale datasets.”

Intended especially to serve users of large scientific datasets, such as many XSEDE researchers, the DSC’s initial capacity, four petabytes, can be expanded as needed. In comparison with tape-based archiving, DSC facilitates very fast data transfer (latency 10,000 times less than and bandwidth many times more than tape), while it also incorporates high reliability and security.

Departments at the University of Pittsburgh, Carnegie Mellon and Drexel are now using DSC, and researchers with large genomic datasets, produced through Galaxy, a web-based platform for bioinformatics at Penn State, are currently using 470 terabytes of DSC storage.

Creating National Cyberinfrastructure

Times are changing for high-performance computing (HPC) research, as a field of study that haven’t traditionally used HPC have begun taking advantage of these powerful tools. This is especially true for PSC’s Blacklight, an SGI® Altix® UV 1000 system acquired in July 2010, with help from a $2.8 million award from the National Science Foundation. As the largest shared-memory system in the world, Blacklight has opened new capability for U.S. scientists and engineers.

“Blacklight has opened new doors to high-performance computation,” said PSC scientific directors Michael Levine and Ralph Roskies, “and rapidly become a force across a wide and interesting spectrum of fields.”

This was part of the plan for the NSF’s XSEDE (Extreme Science and Engineering Discovery Environment) program, which launched in July 2011. The program this year took large steps toward this objective, with a number of nontraditional projects – the common denominator being the need to process and analyze large amounts of data – using XSEDE resources, especially Blacklight, to arrive at new insights.

Among those, described in this booklet, are work that analyzes huge quantities of finance-trading data to arrive at important new findings concerning non-beneficial effects of computer trading of stocks (see pp. 22-25). Several projects in assembly and analysis of “next-generation” sequence data have found that Blacklight’s shared memory is uniquely well suited to advance work in this field (see pp. 18-21).

Shared memory offers a large advantage for many data-intensive applications because all of the system’s memory can be directly accessed from all of its processors, as opposed to distributed memory (in which each processor’s memory is directly accessed only by that processor). Because all processors share a single view of data, a shared-memory system is, relatively speaking, easy to program and use.

Jim Kasdorf, who joined PSC scientific directors Michael Levine and Ralph Roskies in writing the proposal that established PSC, is a PSC’s director of special projects, involved in planning and coordination of many PSC initiatives.