Spectra Logic Tape Storage Beefs Up NCSA Blue Waters Supercomputer

By Kevin Davies

May 22, 2012 | The National Center for Supercomputing Applications (NCSA) has selected T-Finity tape libraries from Boulder, CO-based Spectra Logic to provide hundreds of petabytes of data storage for its upcoming Blue Waters supercomputing system, one of the most powerful supercomputing systems in the world.

Based at the University of Illinois at Urbana-Champaign and funded by the National Science Foundation, Blue Waters will be one of the world's largest active file repositories stored on tape media, scaling to a total capacity of 380 petabytes (PB)—the equivalent of 5,054 years of HDTV video or a stack of books over nine times the distance from the earth to the moon—within the next two years or so.

Scientists will use the supercomputer for a variety of research applications, including hurricane/tornado modeling, the Big Bang, and as a "computational microscope" for biomolecular structure and modeling studies (see below).

Michelle Butler, senior technical program manager for network and storage engineering for the NCSA, struggles to contain her enthusiasm at the behemoth currently being assembled. "It's a fantastic machine," she tells *Bio-IT World*. "I've been here for 22 years, doing storage all along, I just can't wait to get going on this. I've got a 2 Petabyte machine [on the side]. But I want to see the 25 PB [disk storage] run and scream!"

Not Top 500

Blue Waters would almost certainly rank in the top 3 or top 5 supercomputers in the semi-annual Top 500 rankings, says Butler, *if* the NCSA chose to compete. "But that's not a very good use [of our resources]," she says. "It [the computer] is to be used for science and engineering. The code for the Top 500 run is not a good judge of supercomputing. It's just a line in the sand. It doesn't do any I/O or any real scientific work. We propose a scientific app to gauge these very large machines. That's what they're supposed to be running, right?"

Six teams, including that of Klaus Schulten, physics professor at the University of Illinois at Urbana-Champaign, have been using the Early Science System (ESS) for a few months. Blue Waters goes into full production this fall, with 26 research teams currently scheduled for time. "Our scientists say they've gotten more results in 6-8 weeks than the last 3 years," says Butler.

Blue Waters ESS consists of 48 Cray XE6 cabinets (15% of the final total of 303) with 2 PB of online disk storage. Cray took over the project last summer, when IBM pulled out of a contract with the University of Illinois. Within two weeks of the contract being signed with NSF, Butler says, the first machines were already on the floor. "The machine is fully delivered," she says. "All the cabinets are there. We're ramping up—the last truck arrived last Friday."

When complete, Blue Waters will have 25 PB of disk storage. (That storage system —Sonexion—is provided by Xyratex, a partner of Cray.) The disk storage will get the data off to storage in another location, which includes a near-line system that has a 1.2-PB cache. "This is where the Spectra Logic library comes in," says Butler. "The environment will be 380 PB of raw storage. It will have 244 tape drives, starting this summer, and scale to 366 tape drives next summer."

The Blue Waters file system actually exceeds read/write rates of 1 TeraByte/sec, settling at about 2.2 PB/hour, and boasts more than 200,000 cores. "It's 4-5 times faster than any machine in the U.S." claims Butler. NCSA is assembling teams of experts to help the research teams get the most out of this immense resource.

Virus Viewing

The leading life sciences application on Blue Waters so far is managed by the University of Illinois' Klaus Schulten, who leads a center for computational biology and software development. "Our software is used by over 250,000 people. It's particularly successful on the most powerful computers in Japan, China, Europe and America," he says.

Schulten's group is using Blue Waters to perform molecular modeling studies of the HIV capsid. During HIV infection, the virus' capsid shell must open quickly to release its contents. Schulten likens the process to finding the little perforation on a bag of peanuts, which his team has successfully located. Now his group is trying to understand how capsid release is programmed in order to identify new drug targets to combat infection.

"A computer alone can't tell the story—you have to do experiments," says Schulten. In collaboration with Angela Gronenborn (University of Pittsburgh), Schulten's group is performing a clever trick: juxtaposing the detailed crystal structure of the individual capsid protein with lower-resolution electron microscopy results of the native capsid.

"The computer knows the crystal structure of the individual proteins at high resolution, [lined up] like football players singing the national anthem… As the game starts, we can't see the details of the players. So we try to superimpose the structures onto the action of the game."

"We're usually very proud if we can do a 1-million atom simulation," says Schulten. "Now it's 60 million atoms… We literally have 20-100 times more data [than a few years ago]—we have to put [the data] somewhere!"

Spectra Vision

The tape storage RFP (request for proposals) process won by Spectra Logic was very open, says Butler. "We didn't want to narrow the research into what was [technically] possible," she says. NCSA outlined the ideal footprint, the data capacity, the number of media slots, reliability, and requested a performance of 100 Gigabytes/sec. From a pool of ten vendors, four were seriously evaluated over about a year.

According to Bill Kramer, deputy director of the Blue Waters project, the T-Finity solution appealed because of its "high enterprise-level performance, ready data accessibility and massively scalable capacity. We are confident it will provide our user community with fast, reliable access to the massive volumes of critical data stored within Blue Waters' Petascale near-line file repository."

The Spectra tape libraries will enable NCSA to keep all of its near-line data accessible in an active repository. NCSA will begin by deploying four 17-frame T-Finity tape libraries in the first year of operation, followed by two additional libraries the next year.

"We are pleased to partner with NCSA and support one of the most powerful and cutting-edge supercomputers in the world," said Nathan Thompson, Spectra Logic's founder and CEO. "It is gratifying to see tape-based storage play a major role in one of the largest, best practice HPC deployments to date and to help support the important scientific breakthroughs and advancements the Blue Waters project will enable."

The storage solution was architected by storage integrator NET Source, a member of Spectra Logic's SpectraEDGE partner program. The T-Finity will be integrated with IBM's enterprise TS1140 Technology tape drives. "Given Spectra's support of TS1140 Technology and proven storage solutions, Spectra Logic was clearly the ideal solution to meet Blue Waters' high performance, data-intensive storage needs," commented NET Source president Joe Fannin.