Big-picture Storage Approach

Re-thinking data strategies is critical to keeping up

Jason Hick

Increasingly sophisticated supercomputers and networks are changing the way science is done, whether allowing researchers to scrutinize the last 13 billion years of cosmic expansion or to better understanding subatomic particles. Meanwhile, advancements in high-performance networks are facilitating remarkable levels of scientific collaborations. Every day, thousands of physicists from around the world begin work on their individual computers, yet can all access the same data in near real-time thanks to research networks like the Department of Energy's (DOE) Energy Sciences Network (ESnet). These trends are leading to unprecedented scientific productivity and forcing supercomputing centers to re-evaluate how they manage data.

DOE's National Energy Research Scientific Computing Center (NERSC) is well aware of this need to re-think data strategies. Over the last decade, NERSC's user community has grown from 2,000 to 3,500. In the same span, every new system installed at NERSC has generated approximately 50 percent more data than its predecessor. This trend was especially pronounced when the center upgraded its flagship machine from a 10 teraflop/s IBM SP, called Seaborg, to the current 352 teraflop/s Cray XT4, called Franklin. During the Seaborg period, NERSC's archived data grew from 350 terabytes to 2 petabytes. Since Franklin went into production, the archive has grown to 5.5 petabytes, and a variety of science projects ranging from climate modeling to astrophysics are expecting to produce multi-petabyte datasets.

In 2010, NERSC will deploy its next major supercomputer and several smaller systems providing a complement of simulation and data analysis capabilities and increasing the center's science productivity by several times. In anticipation of this, NERSC decided to refocus its approach to storage. Historically, each new computing platform came with its own distinct storage hardware and software, creating isolated storage "islands." Users typically had different filesystem quotas on each machine and were responsible for keeping track of their data, from determining where the data was to transferring files between machines.

In recent years, NERSC began moving from this model toward global storage resources. The center

was among the first to implement a shared filesystem called the NERSC Global Filesystem (NGF), which is accessible from all of the center's major compute platforms. The goal is to facilitate file sharing between platforms and research collaborations. For years, NERSC provided a filesystem called "global project" for sharing files across systems. Recently, the center deployed two more shared filesystems called "global homes" and "global common," and will soon deploy another called "global scratch," that allow users to access nearly all their data regardless of the

system they log into. With these tools, users spend less time moving data between systems and more time using it.

This also allows NERSC staff to focus on their area of expertise. Under the island model, user consultants and system engineers were required to help identify and solve input/output (I/O) bottlenecks. Data experts were divided into two groups, one focusing on the different data systems and another on handling mass storage. This often led to considerable demands on the limited budget allocated for storage. In the wake of increasing scientific productivity, this strategy ultimately led to short-term fixes for handling data. As part of its global strategy, NERSC merged all storage experts into one group to handle all storage-related issues, allowing the center's data experts to deal with the big picture of how best to manage data, and to tailor storage strategies to fit scientific needs.

Another idea that emerged from NERSC's global data focus was investment in "science gateways" nodes on the NGF system. This allows users to serve science data directly to the Web by placing files in a particular directory. In an early project, astronomers at Lawrence Berkeley National Laboratory's Computational Cosmology Center worked closely with NERSC to set up an astronomical archive called "Deep Sky." For the past decade, automated sky-scanning systems have sent approximately 3,000 astronomical images to NERSC nightly for archiving. By comparing images, astronomers can identify relatively rare transient events like supernovae.

In the island storage model, filesystem quotas were typically calculated using a strict formula. Because researchers don't have equal needs, this approach left some users with unused quotas while others required more project space. Because of these quotas, celestial images in the NERSC archive could not be consolidated onto one machine for true comparison. But, with the help of NERSC's new data focus, 10 years of observations could finally be stored on an NGF science gateway node. The astronomers then developed a user-friendly database and interface to instantly serve up high-resolution cosmic reference images to astronomers around the globe. These tools have helped astronomers discover as many as 36 supernovae within the span of a week.

Deep Sky's success shows a big-picture approach to storage is not only critical for keeping up with the explosive growth in data, but that it is also a critical component for advancing scientific discovery. HPC

Jason Hick heads NERSC's Storage Systems Group. He may be reached at editor@ScientificComputing.com