TACC Ranch technology upgrade improves valuable data storage

May 22 2019

New upgrade to hardware and software brings good news to over 52,000 users of TACC's Ranch long-term mass storage system. Credit: TACC

There's a joke by comedian Steven Wright that goes, "You can't have everything. Where would you put it?"

Users of advanced computing can likely relate to this. The exponential growth of data poses a steep challenge to efforts for its reliable storage.

For over 12 years, the Ranch system at the Texas Advanced Computing Center (TACC) has provided long-term archiving of research data ranging from astrophysics to climate science to particle physics, and more. A new upgrade to hardware and software brings good news to over 52,000 users of TACC's Ranch long-term mass data storage system.

The archives are valuable to scientists who want to use the data to help reproduce the measurements and results of prior research. Computational reproducibility is one piece of the larger concept of scientific reproducibility, which forms a cornerstone of the scientific method.

TACC strives to comprehensively support the data needs of scientists. The local compute systems such as Stampede2 and Lonestar provide a high-speed scratch space dedicated for temporary storage of data. Next up the ladder are the front-facing data collection systems of Stockyard and Corral, which provide a combined storage of 50 petabytes directly accessible through high speed web connections or the iRODS data grid. Ranch, on the other hand, allows long-term archiving of data for months to years.
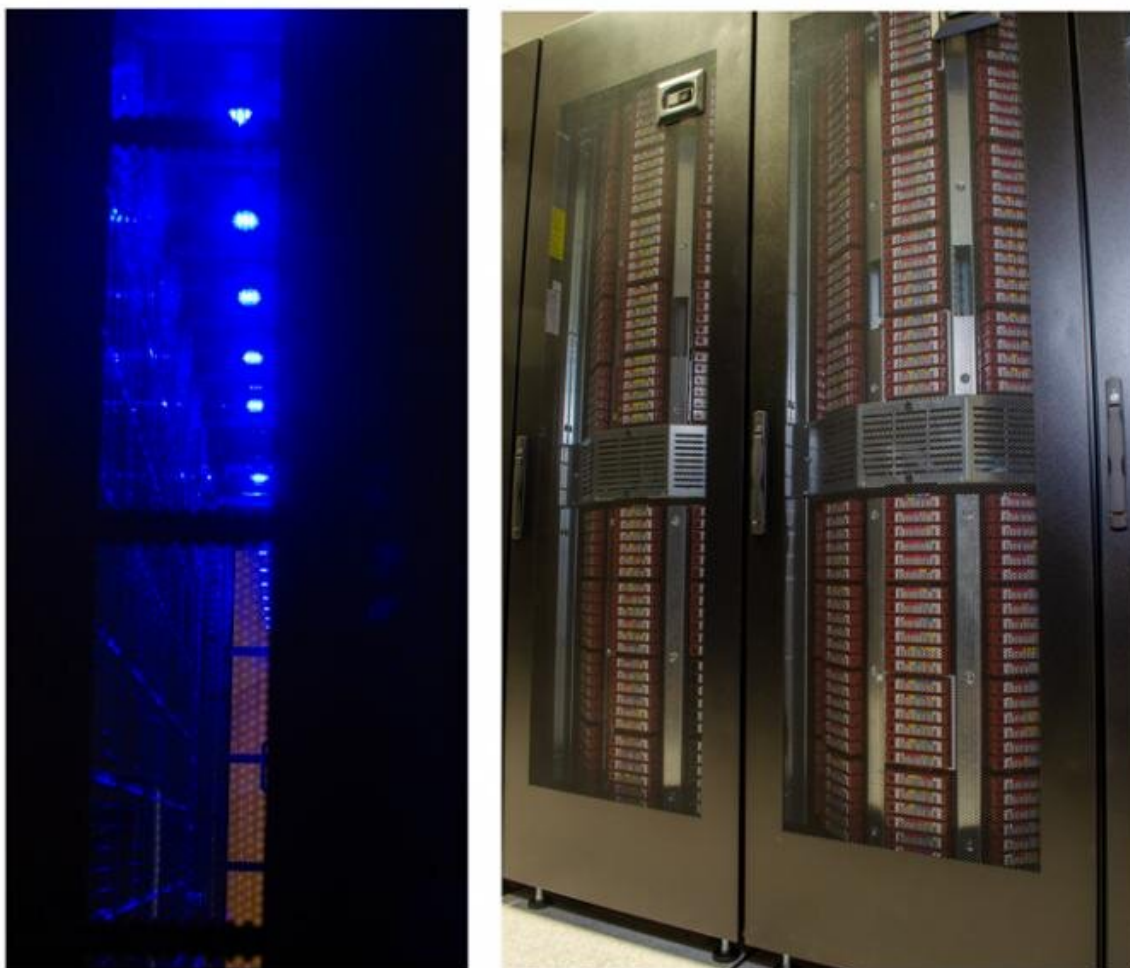
TACC's Ranch archival system includes a Quantum Scalar i6000 tape library with the StorNext archival file system that coordinates both disk and tape storage. Inside view (left) and outside (right). Credit: TACC

As of April 2019, Ranch stores over 70 petabytes, or 70 million gigabytes of scientific data. Over 52,000 users have uploaded close to 1.7 billion computer files on the old library of Ranch that's being upgraded. Hypothetically, the new upgrades to Ranch could expand its storage to reach a mind-boggling exabyte, or 1,000 petabytes.

"For users, more data will be more readily available, with 15 times more disk cache than what we had on the previous Ranch system," said Tommy Minyard, director of Advanced Computing Systems at TACC. Fresh data that's generated from TACC supercomputers such as Stampede2, Lonestar, or Maverick is staged first on Ranch's spinning disk and flash drives, then later moved to tapes.

Ranch has been upgraded with a block storage system supplied by DataDirect Networks, the DDN SFA14K DCR, which provides 30 petabytes of spinning disk cache, versus just two on the replaced hardware. "That means that we'll be able to keep a lot more data staged on disks so that it's more quickly retrievable and you don't have to recall it from tape," Minyard added.

Another welcome change to Ranch comes from the company Quantum, which provided a Scalar i6000 tape library with the Quantum StorNext archival file system that coordinates both disk and tape storage. "We chose the Quantum system based on its capability and flexibility," said Junseong Heo, senior systems administrator and manager, Large Scale Systems of TACC's Advanced Computing Systems group.

Ranch upgraded to the DDN SFA14K DCR block storage system, supplied by

"Specifically, Quantum provides a quota control mechanism that enables TACC to provide an allocation-based project area for users," Heo added. That means that users can have much easier access to data and oversee the usage of resources. "The project-based quota has been at the top of the wish list from users in the past," Heo said.

"TACC's focus on constant innovation creates an environment that places tremendous stress on storage, and Quantum has long been at the forefront in managing solutions that meet the most extreme reliability, accessibility and massive scalability requirements," said "Eric Bassier, Senior Director of Product Marketing, Quantum. "Combining Scalar tape with StorNext data management capabilities creates an HSM (Hierarchical storage management) solution that is capable of delivering under the demanding conditions of the TACC environment."

"The new system provides some additional capabilities for us to be able to handle project data and tiers of storage better than the old system did," Minyard added. This is especially good news for heavy Ranch users, whose allocation might change frequently. The heaviest 100 users combined have more than 20 petabytes on the current archive system.

One of the big changes users will notice is the adoption of the Community Enterprise Operating System (CentOS) environment, which replaced the previous Solaris environment. As users navigate CentOS to migrate their files to the new Ranch libraries, TACC is providing 12 months of read-only access to data in the old library until the end of March 31, 2020. Users should inspect and migrate data in favorable sizes for the tape archives, preferably bigger than 100 gigabytes in individual

file size.

"Archival data isn't glamorous, but it's required," said Minyard. "I don't know how many times we've had somebody panic about accidentally deleting a file, where TACC was able to recall the file for them from Ranch. From their point of view, Ranch is a life-saver," Minyard said.

Provided by University of Texas at Austin