Quantum.

Photo by Will Beeler, RENCI

RENCI Relies on StorNext Multi-Tier Storage for Data-Intensive Genome Sequencing and Coastal Modeling

The Renaissance Computing Institute (RENCI) supports cutting-edge research in life, earth, and data sciences. The organization uses StorNext® multi-tier storage to enable its high performance computing (HPC) cluster and to archive large amounts of precious scientific data for years to come.

renci
RESEARCH \ ENGAGEMENT \ INNOVATION

FEATURED PRODUCTS

StorNext®

"StorNext provides a single, multi-tier platform that simplifies administration and helps ensure researchers have the resources they need right away."

Erik Scott
Senior research software developer,
RENCI

![renci - RESEARCH \ ENGAGEMENT \ INNOVATION]

SOLUTION OVERVIEW

- StorNext® Scale-out Storage, including:
  - StorNext M662 Metadata Appliances
  - StorNext G302 Gateway Appliances
- Scalar® i6000 Tape Libraries

KEY BENEFITS

- **Manages hundreds of terabytes of data** across a multi-tier storage system
- **Provides a large, scalable tape archive** for long-term preservation of valuable data
- **Offers simple, reliable access to all data** through a single storage file system
- **Enables fast multi-user HPC data access** with dedicated scratch disk space

Based at the University of North Carolina (UNC) at Chapel Hill, RENCI supports groundbreaking research. Scientists from UNC, North Carolina State University, Duke University, and other universities investigate the genetic causes of disease, create detailed storm surge models, facilitate the use of big data for government and industry, and more.

Groundbreaking research requires vast amounts of data. "Earth science researchers collect data from aerial LIDAR surveys and other sources to produce high-resolution 3D models of coastal areas," says Erik Scott, senior research software developer at RENCI. "Those models help forecasters predict storm surges. As Hurricane Sandy approached, for example, high-resolution models helped the Coast Guard prepare for the storm."

ARCHIVING LARGE VOLUMES OF GENOMIC AND COASTAL MODELING DATA

RENCI offers scientists robust compute, storage, networking, and visualization resources to support research. For instance, scientists take advantage of the organization's 5,168-core HPC cluster—named Hatteras—to analyze genome sequences and create storm surge models.

Providing sufficient storage is also critical. "Genome sequencing can produce 100GB or more of data per person," says Chris Bizon, senior informatics scientist at RENCI. "In addition to keeping the final products, we store temporary data as it's generated—if something goes wrong, researchers can look at data without having to start at the beginning."

The IT group at RENCI must also keep valuable data available for research over the long term. "Genomic data is expensive and time-consuming to collect," says Bizon. "We want to make sure we preserve data so researchers can conduct new analyses in the future."

CREATING A MULTI-TIER ENVIRONMENT FOR HIGH PERFORMANCE COMPUTING

Several years ago, the IT group at RENCI implemented a multi-tier storage environment powered by StorNext data management

software to streamline their workflows and remove impediments for scientific projects. Currently, the StorNext multi-tier environment includes StorNext metadata appliances, StorNext gateway appliances, and Quantum Scalar i6000 tape libraries as well as disk storage from other vendors.

"Before StorNext, we were constantly monitoring data consumption and moving data among multiple systems," says Marcin Sliwowski, IT manager at RENCI. "StorNext provides a single, multi-tier platform that simplifies administration and helps ensure researchers have the resources they need right away."

The tape libraries provide a cost-effective way to archive large amounts of scientific data. "Tape is much less expensive than disk," says Sliwowski. "In addition to avoiding hardware costs, we can reduce power, cooling, and real estate expenses. The tape libraries allow us to store large data volumes efficiently, in a small space."

PROVIDING EASY ACCESS TO RESEARCH DATA TO ENABLE SCIENTIFIC COLLABORATION

StorNext data management software provides a single, transparent namespace for the entire multi-tier environment, making it easy for researchers to access the data they need. "The research projects conducted through RENCI are often collaborative projects that include scientists from other institutes," says Bizon. "With StorNext, we can provide access to all research data, so scientists can analyze information without moving data off-site."

"StorNext allows researchers to access data as if it's on their devices," adds Scott. "They have a huge repository of files right at their fingertips."

SCALING TO ACCOMMODATE CONTINUOUS GROWTH IN RESEARCH DATA

The StorNext tape archive at RENCI currently offers nearly 3PB of capacity, giving researchers plenty of room to preserve genomics data and earth science data for years to come. "The StorNext archive can scale to a very large size, so researchers don't have to worry about how much data they're saving," says Scott.

The research at RENCI can fluctuate, so the ability to easily scale capacity is key to supporting researchers. "At one point, we were storing 500 simulated storms," says Scott. "By having access to all that data, researchers can help government agencies better mitigate the effects of serious storms."

IMPROVING HPC SPEED WITH DEDICATED SCRATCH DISK BUILT ON STORNEXT

A few years ago, the RENCI IT group expanded its use of StorNext to support a new scratch disk environment for its Hatteras HPC cluster. "We had been using a shared storage resource for the HPC cluster, but we decided to create a dedicated scratch disk space to improve input/output for research computations," says Sliwowski.

Researchers use the StorNext scratch disk environment as they analyze genomic data, create coastal models, and run other workloads. The StorNext distributed LAN client (DLC) enables fast access to scratch data over Ethernet.

> "With StorNext, we can provide access to all research data, so scientists can analyze information without moving data off-site."
>
> **Chris Bizon,**
> Senior informatics scientist, RENCI
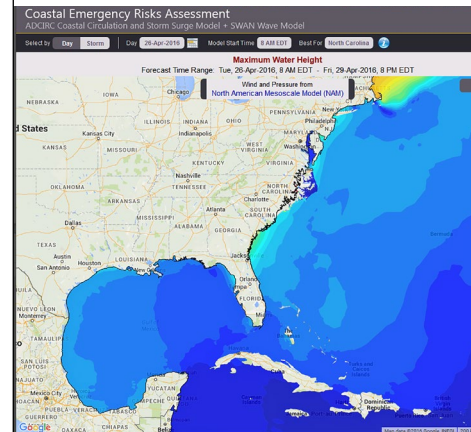
Source: Coastal Emergency Risks Assessment, CERA (http://nc-cera.renci.org/)

ABOUT THE RENAISSANCE COMPUTING INSTITUTE

The Renaissance Computing Institute (RENCI) develops and deploys advanced technologies to enable research discoveries and practical innovations. RENCI partners with researchers, government, and industry to engage and solve the problems that affect North Carolina, the United States, and the world. An institute of the University of North Carolina at Chapel Hill, RENCI was launched in 2004 as a collaboration involving UNC Chapel Hill, Duke University, and North Carolina State University.

Photo by Josh Coyle, RENCI
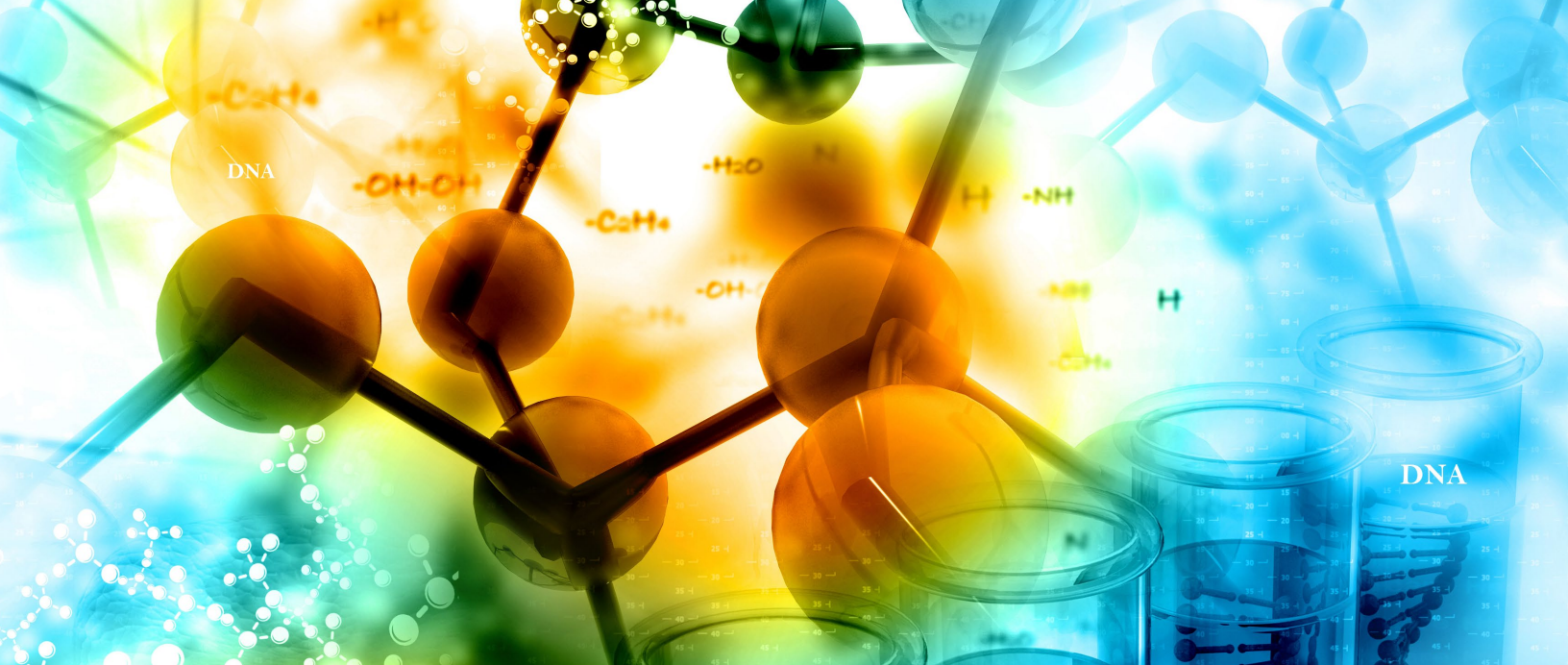
Researchers have noticed performance improvements from the dedicated scratch space. "The previous scratch storage solution slowed when multiple users tried to stream data at once—it was laptop-class performance," says Scott. "With StorNext, that's not a problem anymore. Multiple users can collaborate on data without performance issues."

The IT group also capitalizes on StorNext policy-based tiering to move data from disk to the archive. "If files aren't accessed for a certain number of days, StorNext automatically archives the data to tape based on the policies we've set," says Sliwowski.

ENHANCING RELIABILITY AND EASING IT ADMINISTRATION BURDENS

With StorNext, the RENCI IT group now has a robust, reliable data management platform to efficiently support researchers. "Researchers use the file system to get what they need, when they need it, without our assistance," says Sliwowski. "StorNext keeps them focused on the science."

Quantum.

www.quantum.com

CS00386A-v02