Quantum

CASE STUDY

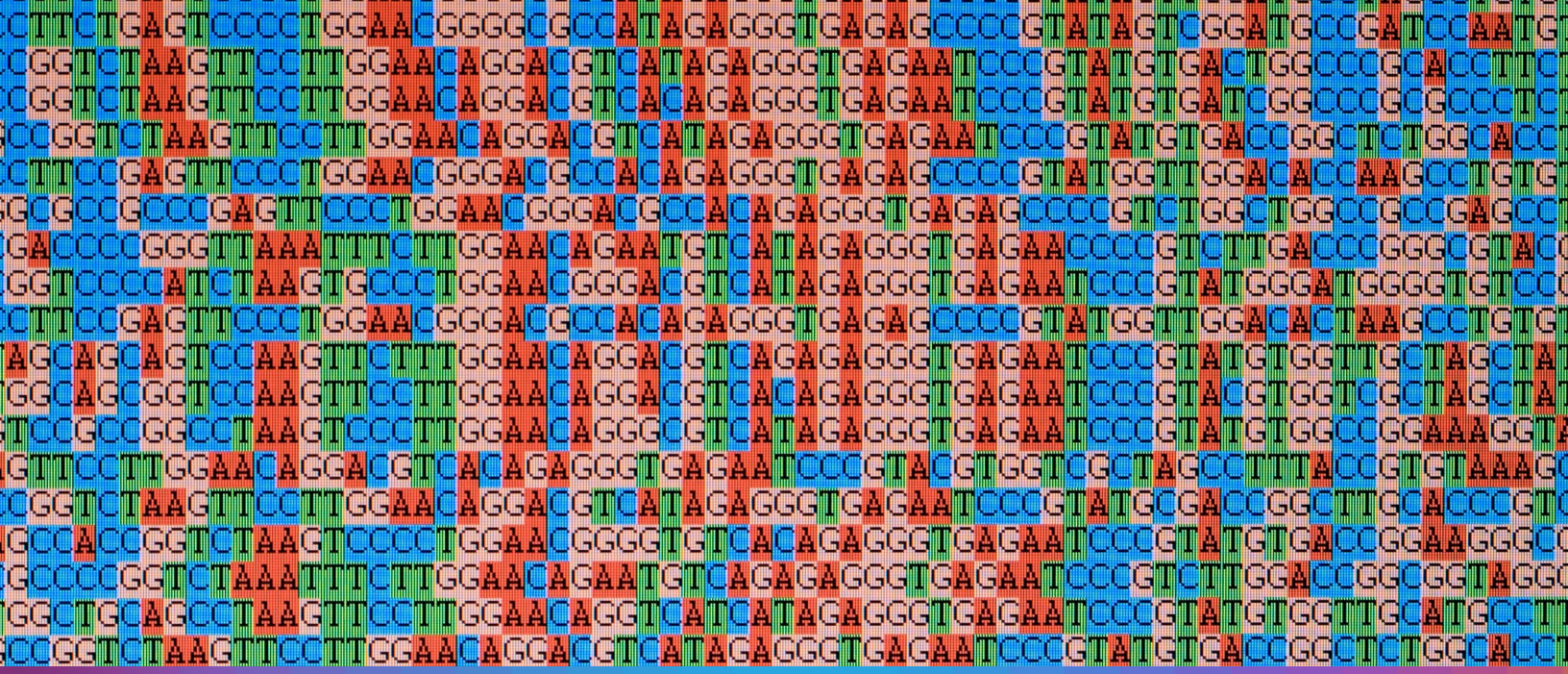
Advanced Genomics Sequencing Provider Preserves Unique Data with StorNext File System Solution

As a provider of genomics sequencing solutions to support advanced cancer therapies and treatments, it is a business-critical requirement to safely store data while maintaining access to it over long periods of time. As data growth accelerates with advancements in sequencing technologies, finding ways to preserve data cost effectively becomes most paramount. Quantum's disk and tape active archive, powered by StorNext File System, delivers a scalable, flexible, and cost-effective solution for a provider of advanced genomic sequencing and analytics solutions, helping the company preserve its unique data at 10% of primary NAS costs without sacrificing data resiliency and accessibility.

FEATURED PRODUCTS

StorNext®

Quantum's disk and tape active archive, powered by StorNext File System, delivers a scalable, flexible, and cost-effective solution for a provider of advanced genomic sequencing and analytics solutions.



The StorNext File System platform is helping the genomics sequencing solution provider preserve its unique data at 10% of primary NAS costs without sacrificing data resiliency and accessibility.

SOLUTION OVERVIEW

- Quantum StorNext® File System
- Disk cache as landing space
- Tape library with 12 LTO drives

KEY BENEFITS

- StorNext File System enabled the company to push more than 100 TB per day in and out of the archive.
- Data stored on tape is available via NFS and SMB mounts; the customer gains the benefit of tape infrastructure at scale without having to manage robots, drives, and media.
- Though data is stored on tape, it remains immediately accessible through StorNext File System.
- Capacity can be scaled by adding slots in the library with additional media.
- To future-proof the investment, the company can upgrade tape drives for greater capacity and performance—an incremental cost that decreases the cost per GB of data stored.
- The company maintains a 10 PB archive at 10% of primary infrastructure costs.

In April 2003, the human genome was sequenced in its entirety for the first time. It was the beginning of a new era in medical research. The possibilities of identifying and treating diseases based on a person's individual genome opened our minds to immunotherapies and individualized medicine. Today, anyone can have their genome sequenced for $150. In addition to answering many questions about our ancestry and identifying perpetrators in crimes, the sequenced genomes can provide insights into the formation of mutations that cause disease.

A biopharmaceutical research company is providing genomic sequencing solutions to support the development of advanced cancer therapies and treatments. Their DNA-sequencing research generates petabytes of data that is processed, analyzed, and used to inform cancer treatments. All data eventually lands in an archive powered by the Quantum StorNext platform.

CHALLENGES

Gene sequencers are continuously generating terabytes of data. This raw data needs to be processed and analyzed, then the results are shared with researchers, doctors, and patients. The company's services depend on an infrastructure that supports:

- High-throughput data ingest and movement across storage platforms. Once data is generated by the sequencers, it lands on a small NAS where the company's software does its analysis. Results are then posted to a separate NAS.
- Raw data, results data, and the associated software versions and configuration files are sent to an archive as a TAR file. Up to 160 TB of data has been ingested into the archive in a single day, requiring sustained throughput to be in the 5 GB/s range.
- Today, the archive holds 700 million files and consumes 10 PB of capacity. Storing this much data on disk drives added costs such as cooling, power, and floor space. The archive had to deliver density, low power

draw, and low cost without sacrificing data accessibility and protection.

- Cloud was considered in support of archiving, but the cost of egress was too high. Data stored in the archive is often retrieved back to the NAS for reprocessing with newer versions of the company's software. The frequent access to data in the archive made cloud an expensive option.

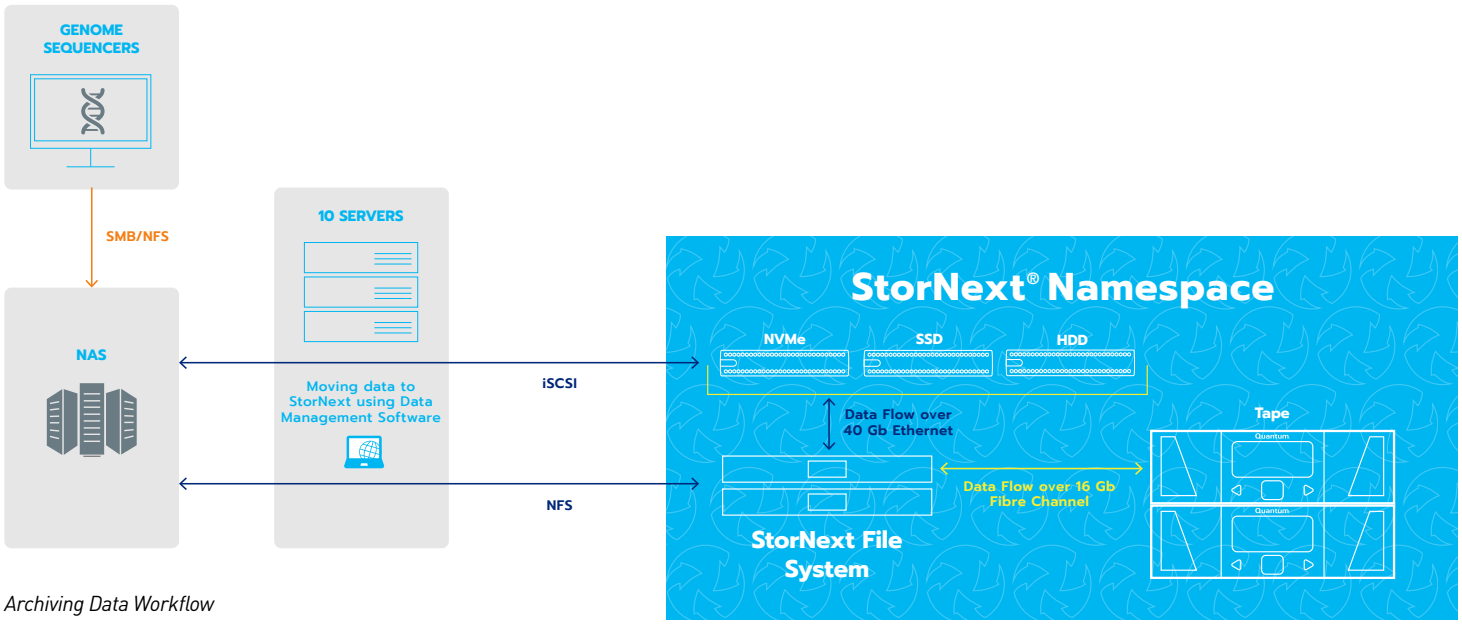
The company's business relies on its ability to generate, analyze, share, and retain data. The company approached Quantum for its archive solution that combined the StorNext platform's comprehensive lifecycle management that included tape.

SOLUTION

Storage systems are designed to deliver three fundamental services: performance, capacity, and data protection. It is well understood that higher-performance drives increase the cost per increment of capacity. Though the value of data may not change over time, access patterns and performance demands do. The industry views the two ends of a performance continuum as high performance on one end and archive on the other. Few organizations require enough throughput to ingest 100 TB of data in one day when speaking of archiving. Though several systems can support 2 GB/s throughput, the cost of doing so can be prohibitive. To resolve the need for throughput against the need for cost efficiency, the company selected the Quantum StorNext File System platform.

- An old HPC cluster consisting of 10 servers was used to push data off the NAS into the archive. StorNext File System parallelized the streams for maximum throughput. The target storage environment consists of 231 TB SSD. Twenty minutes after landing on disk, data is moved to tape.
- Once data lands on disk cache, StorNext executes a policy to either keep the data on disk for a set period, or immediately move it to tape. Once on tape, two copies of the data would be made; one copy remains in the library while the second copy is sent off site.
- Data often has to be retrieved from the archive for additional analysis or to rerun the data against a newer version of analytics software. There is enough throughput available on the tape drives to support, concurrently, ingest and retrieval requirements. Over 400 TB of data move through the archive on a weekly basis.

As a result, the provider's StorNext archive is home to 700 million files over 10 PB of storage capacity. Looking ahead, the company is planning on expanding its archive capacity by licensing additional slots in the library.



Archiving Data Workflow

- Using LTO-8 drives, up to 12 TB of data can be stored on a single cartridge. Due to the nature of the data being stored, data is encrypted. The density of the tape solution, together with security and lower draw on power and cooling, made it an attractive option. After detailed analysis of media costs, the company chose to move forward with LTO-7 M8 media to provide the lowest-cost media option.

BENEFITS AND RESULTS

As a result, the provider's StorNext archive is home to 700 million files over 10 PB of storage capacity. Looking ahead, the company is planning on expanding its archive capacity by licensing additional slots in the library. Adding SSD-based storage systems will increase the throughput capabilities of the archive. As the amount of data generated by gene sequencers increases, the company is confident in its ability to store data securely and efficiently.