Case Study

MEDICAL RESEARCH

Challenge

- Delivers a very high performance Linux cluster for advanced human genome sequencing using fast 10 Gigabit SAN storage
- Provides a scalable primary storage system for Linux application servers.
- Offers a reliable and scalable storage system for disk to disk backups.

Solution

- EtherDrive HPC Storage
- EtherDrive Primary Storage
- EtherDrive Backup Storage

Benefits

- Very fast performance for HPC Storage requirements
- Reliable Primary Storage
- Build a scaleable storage system for use as disk to disk backups.
- Economical, Scalable Disk to Disk Backup

National Human Genome Research Institute

National Human Genome Research Institute (NHGRI)

Over the past three years, The National Human Genome Research Institute (NHGRI) in Rockville, Maryland has evolved their storage to address the unique and growing challenges of handling complicated image and mathematical analysis workloads. Part of the National Institutes of Health (NIH) , the NHGRI is chartered by the U.S. Department of Health and Human Services to conduct research to understand the human genetic code of DNA using new massively paralleled sequencing technologies. Bioinformatics analyses of these data require unique computing platforms. To accommodate these massive data storage needs the government agency turned to CORAID EtherDrive storage.

Today, the NHGRI storage infrastructure supports approximately 60 Linux-powered compute nodes and stores over 200TB of total disk storage. Attached to their in-house Linux cluster and configured to serve three discrete storage functions, all with high throughput and availability requirements, NHGRI selected a variety of EtherDrive storage systems and deployed them in the following ways:

High Performance Cluster

This data-intense HPC cluster processes complex DNA sequencer images that are captured and fed into a high performance Linux computer cluster over a standard Ethernet connection. The NHGRI required very fast shared storage that would keep up with the exacting performance needs of the cluster and scale to meet demand. The agency had been successfully using CORAID EtherDrive SR1521 storage in their Linux environment for other applications. Finding EtherDrive to be both simple to use and reliable to operate, they turned to CORAID to provide an affordable high performance solution to support the cluster. To achieve the more demanding HPC requirements, the institute selected EtherDrive SR2461 with 10 Gigabit Ethernet connections running between the SAN storage and the HPC cluster.

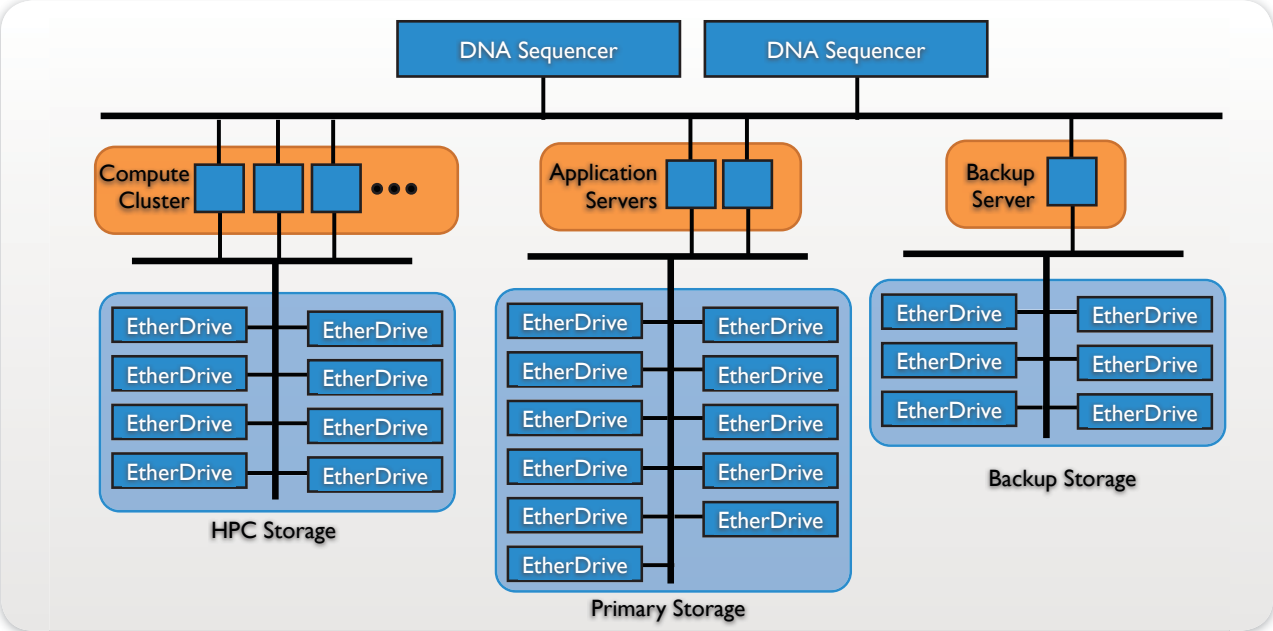
Application Servers

Requiring reliable, Primary Storage for temporary data stored in the HPC cluster and processed into a smaller amount of data to be stored

long-term, the agency chose EtherDrive SR1521 with Gigabit Ethernet connections. For post-processing data across several Linux analysis servers, CORAID delivered a dependable primary storage system that easily scales to the meet the data demands of genomic research.

Disk-to-Disk Backup

Fast backup using disk-to-disk storage is a familiar application for Linux servers. EtherDrive SR2421 was selected as a perfect fit for the institute's Backup Storage needs based on its exceptional scalability and the economic advantage of running over simple Ethernet connections.



The diagram above illustrates the NHGRI storage infrastructure running across a variety of CORAID EtherDrive storage appliances. With the help of EtherDrive performance advantages and storage capacity along with standard features such as disk-to-disk backup, NHGRI researchers can now analyze data faster and store far more data. The system's scalability and adaptability ensures the NHGRI can scale to satisfy research demands for more cost effective storage and I/O performance with minimal management burdens. "We have found the CORAID EtherDrive system to be a good balance between economy and performance. In addition, it gives us a tremendous amount of flexibility in how we configure our storage -- something that is critical in the fast-changing world of next-generation sequencing," said Dr. Elliott Margulies, Investigator, Genome Informatics Section. CORAID EtherDrive provides a tremendously better price to performance value than other storage platforms. With genomic research driving up data demand exponentially, EtherDrive has been an affordable, reliable way to boost research speeds. EtherDrive is simple to administer, and its open standards-based architecture allows users to incrementally scale storage minimizing up front costs, system overhead and maintenance.