Life Sciences

Challenge

The customer generates massive amounts of data of genome sequences, up to 6TB per operation. They needed a storage subsystem that could grow as bioinformatics apps analyze current sequences and ingest more data. Traditional RAID and NAS storage systems could not easily scale within the genomics performance requirements.

Solution

WARP Mechanics installed the WARP 38000-S Storage Matrix, starting with as little as 80TB of capacity and scaling to 2PB per rack. Its unlimited filesystem size and advanced data protection features make adding new capacity simple and reliable.

Results

The customer has effectively replaced their legacy storage system with the WARP 38000-S. They are able to keep pace with the bioinformatics application without worrying about running out of space. As their data footprint grows with their company, they have no practical limit to the amount of sequence information they can analyze.

Next generation sequencing (NGS) of genomes qualifies as big data. A single Illumina HiSeq 2000 machine sequences many patient samples simultaneously. The aggregated genome sequences can be as large as 6TB and grows as bioinformatics apps analyze the sequences.

Even though 4TB hard drives are prevalent in the market, and not all of the gene sequencing data need be retained, a storage system must easily grow to add capacity. WARP Mechanics' **WARP 38000-S** Storage Matrix simplifies capacity growth and slows server and rack sprawl in the datacenter. The file system has no practical maximum size which avoids multiple storage heads and prevents forcing applications to refactor to use additional storage targets. The customer doubled their capacity from 50TB to 100TB within the first six months of implementation, growing without downtime or costly data migrations.

Sequencing runs are expensive in consumables (biochemicals), sequencer utilization, and lab preparation time. Data loss negatively impacts productivity severely, needlessly burning through samples and costing time and money to reset the operation. Through its advanced data protection technology, WARP Mechanic's reliability is the best in the industry using multi-level block checksums, transaction logging, copy-on-write, scrubbing, multi-level parity, and OEM-quality hardware.

A compute grid is needed to run the bioinformatics apps because the algorithms can take hours to execute and multiple cores reduce the lapsed time. Whether input data is staged to a compute node or not, fast storage can reduce I/O wait times during analysis. The WARP 38000-S provides several high-performance interface options to the compute grid, including 10Gb/40Gb Ethernet and QDR/FDR Infiniband. In addition, simplified storage expansion allows the customer to add spindles for increased bandwidth, or embed low-latency NVRAM and SSD modules should future workloads demand it.

Often bioinformatics apps rely on index files (many small files containing sequence metadata) which are read/written throughout the analytics job. This results in random access I/O patterns that can strain HDD storage and slow down the job. Therefore, SSD is an important option to shadow compute node buffer caches. The WARP Mechanics appliance architecture allows any mix of rotational or solid state drives within the same filesystem as main storage or to accelerate reads and writes as needed. As the company grows their business, having such seamless and flexible options allows them to focus on the science and not future-proofing their storage.

The customer is a NGS startup in Silicon Valley developing an assay for early detection of cancer. They needed a unified centralized storage appliance to meet several requirements:

1. Illumina sequencer output (CIFS).
2. Ingest to bioinformatics compute grid (NFS).
3. Results computed by its bioinformatics apps (NFS).
4. Sapio Sciences' Exemplar LIMS and its database on VMware datastores (NFS).
5. Collaboration folders for employee Macs (CIFS).
6. Snapshots to guard against accidental file erasure/corruption and provide a consistent backup source.

The customer evaluated proprietary and open source object stores (with/without erasure coding) and file systems. The WARP Mechanics 38000-S appliance was purchased because it delivered all of the above at an industry-leading $/GB along with elite pre-/post-sales support. WARP Mechanics has delivered on all counts and more.

Illumina HiSeq 2000
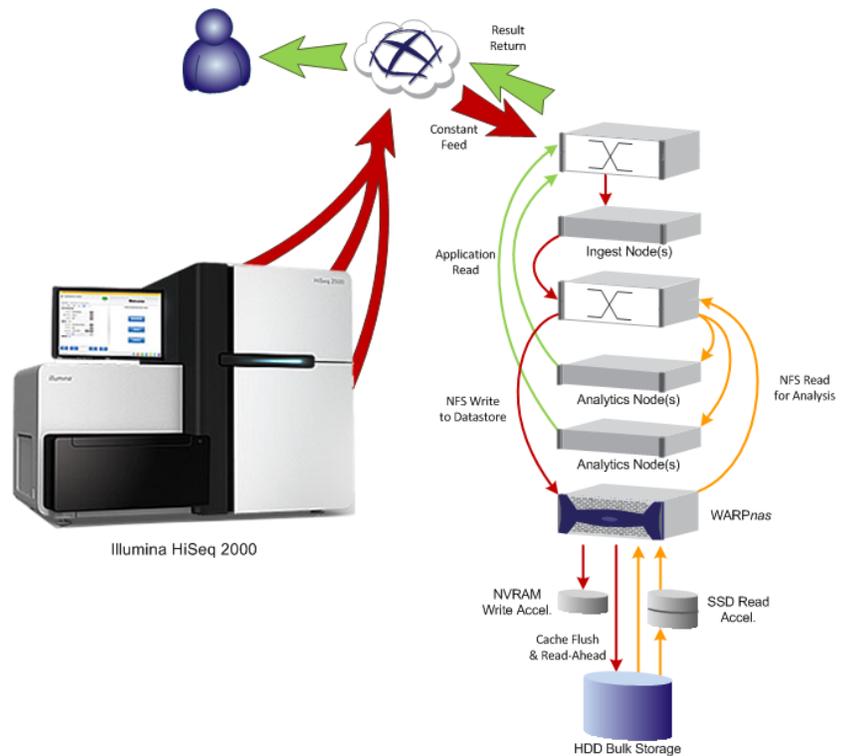
High Quality Data Trumps Big Data

The customer was founded by a group of individuals who thought that next-generation sequencing artfully applied could create fundamental shifts in the quality of healthcare, especially for cancer. They realized that exponential decreases in sequencing costs (although seemingly disruptive) were not alone sufficient to help solve problems that have continued to plague cancer care. Using their expertise as pioneers in the fields of digital signal processing, next-generation sequencing and rare genomics, they developed their core technology with an architecture that could continually maximally leverage any advances in next-gen sequencing and next-gen bioinformatics to ensure the highest performance possible today as well as tomorrow.

www.WARPmech.com

WARP MECHANICS®

Corporate U.S. Headquarters
1288 Columbus Ave #176
San Francisco, CA 94133
Phone: 888-WARPMECH
+1.888.927.7632
info@WARPmech.com