Meeting the **Storage** Challenge of Exponential Data Growth in Genomics

Challenge

- **Budget Constraints.** Need affordable high-density system to implement an enterprise-class storage cloud
- **Collaboration.** Enable research partners across the country or around the world to share data regardless of device platform
- **Scalability/Serviceability.** Deliver non-disruptive capacity and performance scaling, data repairs and upgrades
- **Availability Requirements.** Ensure easy access to data regardless of where it is stored, how old it is or even if a data center goes offline
- **Complexity Sprawl.** Simplify overall deployment and management of storage infrastructure and operations

Solution

- HGST Active Archive System is a complete, easy to install, easy to use object storage solution
- Pay-as-you-grow capacity and performance expansion that does not strain the budget
- Continuous access to data regardless of its physical location or the accessing device type
- Reduced time-to-data and improved collaboration with rich metadata tagging and a single namespace
- Consolidated backup, archive and restore processes for simplified management
- Disaster prevention through the geographic distribution of data

The life sciences and healthcare industries are in the midst of a dramatic transformation that will make personalized medicine common-place. Completing the sequencing of the first human genome in 2003 was a key breakthrough that took more than 10 years and cost over US$3 billion. Since then, the speed at which genomes can be sequenced has more than doubled, easily outpacing Moore's Law. Today's high-throughput sequencing machines can process the human genome in a matter of hours at a cost approaching US$1 thousand. These advancements have allowed researchers to run more sequencing tests in less time, greatly increasing the pace of scientific discovery. Clinicians, too, have begun leveraging genome sequencing to more effectively treat patients with medications tailored to the patient's unique genetic makeup. The result has been an explosion in genomic data, and organizations are now looking for ways to affordably store and manage this avalanche of data.

Traditional Archival Systems Have Limitations

Archival systems are often comprised of complex disk and magnetic tape storage subsystems organized into performance tiers. Disk tiers rely on RAID for data protection and do not scale to levels suitable for life science applications. System performance can be severely impacted when a single multi-terabyte drive fails. The risk of data loss significantly increases if a second drive fails while the data from the first drive is being reconstructed.

As data is used less frequently, it is moved to magnetic tape. Tape archival systems consume little energy and can be cost effective, but the tradeoff is longer retrieval times and added management complexity. Tapes must be stored both onsite and offsite for true disaster recovery, which means a duplicate investment in tape media and ongoing offsite storage costs. While tape systems are designed for long term archival, they offer no means of preventing the natural data degradation that occurs over time, and autoloaders become a single point of failure in the event of a mechanical failure. This means that there is no guarantee that the data will be accessible or readable when it is needed.

Active Archive is an Ideal Solution

Life science involves some of the most complex analyses found in scientific research. It's not surprising that researchers have very unique needs in terms of compute and storage performance, scalability and availability. Clustered file-based storage is frequently used to support the high performance compute clusters used during the analysis phase of the workflow. Utilizing such expensive primary storage for archiving is not practical or cost effective. These architectures must rely on expensive software to replicate data to a remote location in order to provide disaster resilience. Implementing a cloud-scale active archive that does not rely on replication software is a much more cost effective solution.

Up and Running Quickly – Easy to Use

The HGST Active Archive Systems remove the challenges of architecting, purchasing and operating cloud-scale storage solutions. As an integrated rack-level system, it's up and running in minutes. Put it in place, connect the power, configure the network connections and it's online, presenting an S3-compliant object interface and global namespace.

Highest Levels of Data Protection and Availability

The Active Archive System was architected with industry-leading scalability and data durability as primary design goals. At the heart of the system is EasiScale™ object storage software that ensures valuable research data is well protected and always available. Patented technology delivers 15-nines durability, keeping data consistent and accessible even during a full data center outage in a multi-site implementation. Through background data integrity checking, data remains bit-perfect and highly available. The system automatically detects and corrects data degradation without user disruption or IT intervention.

Compelling TCO

Research organizations no longer have to weigh the risks of a public cloud against the cost to build a private cloud or design around the limitations of existing storage architectures. Equally important, there is no need to incur the expense or burden of transferring large data sets to the cloud only to pay a second time to retrieve them. Through innovation and integration, the HGST Active Archive System, delivers cloud-like scale and efficiency at low $/TB and watts/TB.

Replication for disaster recovery is common practice with traditional storage systems. The Active Archive reduces the total storage capacity required by distributing a single copy of data across multiple data centers. Eliminating the need for expensive replica copies and replication software licenses can save up to 60% in extra storage costs. Further savings can be realized through the system's use of patented HGST helium-filled hard drives, which consume 60% savings in power and cooling than air-filled hard drives.

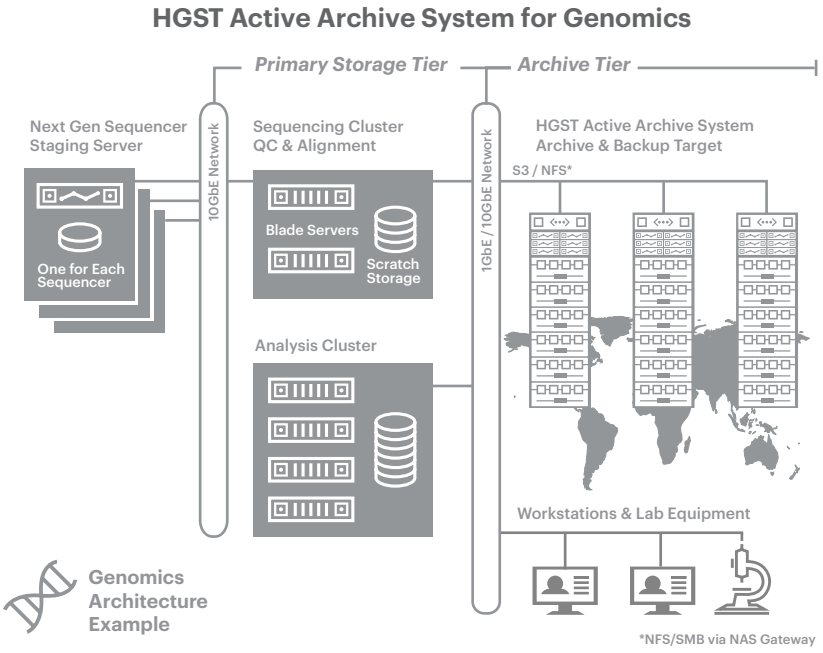

Figure 1. Simplified Enterprise Cloud Storage Infrastructure

The fully integrated design delivers the highest capacity per square foot in the industry, allowing you to optimize expensive data center floor space and delay data center expansion. With 8TB second generation HelioSeal® hard drives, a single rack holds 4.7 petabytes of raw capacity and delivers up to 3.5GB per second total throughput. Performance scales linearly with added capacity to meet the needs of large and growing genomic data sets.

Expensive primary storage should be dedicated to the processing and analysis of genomic data, while archiving and backup and disaster recovery can be consolidated into a single active archive tier. Figure 1 above shows what a simplified two-tier storage infrastructure might look like. By reducing the cost and complexity of managing several tiers of storage and leveraging the pay-as-you-go capacity expansion model, HGST's Active Archive System ensures that more money goes to research than to infrastructure.

To learn more visit http://www.hgst.com/activearchive or contact sales for your region.

Contact Sales

North America: salesna@hgst.com | **Europe/Middle East/Africa (EMEA):** salesemea@hgst.com | **Asia Pacific:** salesap@hgst.com | **Japan:** salesjp@hgst.com

© 2015 HGST, Inc. 3403 Yerba Buena Road, San Jose, CA 95135 USA. Produced in the United States 8/15. Revised 2/16. All rights reserved.

HelioSeal, BitSpread and BitDynamics are registered trademarks, and EasiScale is a trademark of HGST, Inc. and its affiliates in the United States and/or other countries.

HGST trademarks are intended and authorized for use only in countries and jurisdictions in which HGST has obtained the rights to use, market and advertise the brand. Contact HGST for additional information. HGST shall not be liable to third parties for unauthorized use of this document or unauthorized use of its trademarks.

References in this publication to HGST's products, programs, or services do not imply that HGST intends to make these available in all countries in which it operates.

Product specifications provided are sample specifications and do not constitute a warranty. Information is true as of the date of publication and is subject to change. Actual specifications for unique part numbers may vary. Please visit the Support section of our website, www.hgst.com/support, for additional information on product specifications. Photographs may show design models.