Quantum

CASE STUDY

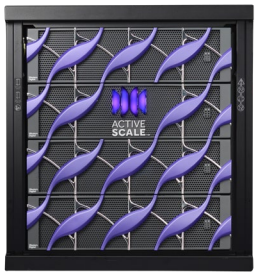
Genomics England Stores, Protects, and Provides Access to Petabytes of Genomic Data with Quantum ActiveScale

Genomics England was established in 2013 by the UK's Department of Health & Social Care to support the 100,000 Genomes Project—a groundbreaking effort to sequence whole genomes from a vast number of patients with rare diseases and common cancers. In 2018, the project was significantly expanded, requiring Genomics England to scale its data storage infrastructure. Using Quantum ActiveScale™ object storage as part of a single, integrated environment enables the organization to store, protect, and provide access to hundreds of petabytes of genomic data.

Genomics england

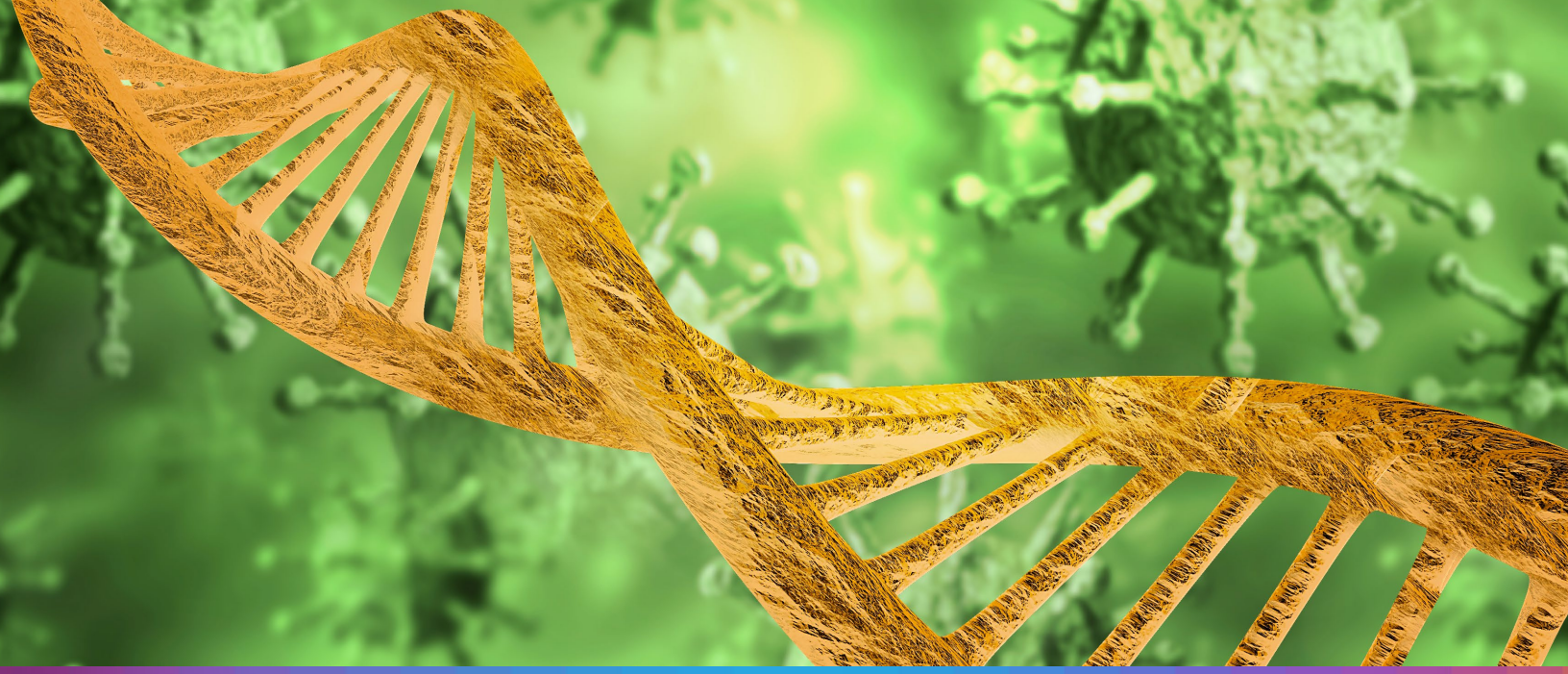
FEATURED PRODUCTS

ActiveScale Object Storage



> What we love about the ActiveScale system is that its inherent architecture is underpinned by its RAID replacement technology, the intelligent, dynamic placement of erasure-coded data.

David Ardley
Head of Technical Delivery, Genomics England



> We needed something that's much more scalable than existing NAS solutions—an infrastructure that could grow to hundreds of petabytes.

David Ardley - Head of Technical Delivery, Genomics England

Genomics england

SOLUTION OVERVIEW

- Quantum ActiveScale object storage
- WekaIO distributed file system
- Mellanox high-speed networking

KEY BENEFITS

- **Overcame capacity limits of the previous storage solution,** gaining the scalability for fast-growing volumes of genomic data.
- **Improved agility for taking on new, urgent challenges,** including the need to sequence genomes of people with COVID-19.
- **Enhanced protection and resiliency for critical data** with a geo-dispersed environment spread across three data centers.
- **Reduced the costs and complexity of storing data** with an efficient environment that is simple to use.

Genomics England has ambitious aims. The organization was created to sequence the genomes of an enormous number of people, generating new insights that can help improve treatments—while also accelerating the development of the UK genomics industry.

Within a few years of its inauguration, the organization proved that it could reach its initial goal of sequencing 100,000 genomes. And it was rewarded in 2018 with a new goal of sequencing up to five million genomes over five years.

Unfortunately, the existing network-attached storage (NAS) solution used for storing genomic data was not up for the task. The NAS, which was storing 21 PB of data, could not be scaled further—it had reached its node-scaling limit.

As the NAS approached capacity, performance suffered. Meanwhile, the organization could not ensure resilience for all this genomic information. Though data was backed up to tape, Genomics England lacked a robust disaster recovery strategy that could restore data rapidly in the event of a disaster or another significant disruption.

The organization needed a solution that could store and protect fast-growing data volumes. "We needed something that's much more scalable than existing NAS solutions—an infrastructure that could grow to hundreds of petabytes," says David Ardley, head of technical delivery at Genomics England. A new solution also had to facilitate simple, flexible access to data by more than 3,000 researchers around the world.

USING QUANTUM ACTIVESCALE OBJECT STORAGE

Genomics England consulted with Nephos Technologies, an independent UK-based data services organization to design and implement a new storage solution. Together, teams from Nephos and Genomics England reviewed current requirements, projected future needs, and then designed and implemented a new, multi-faceted solution.

The storage solution incorporates a WekaIO high-performance file system, Mellanox high-speed networking, and Quantum ActiveScale object storage. It creates a two-tier architecture that combines flash storage plus ActiveScale object storage, which serves as a long-term data lake repository. The initial deployment included 1.4 PB of flash and 40 PB of object storage.

The two storage tiers—each of which can be scaled independently—present as a single hybrid storage environment. As a result, researchers have the flexibility to query data in a highly randomized fashion.

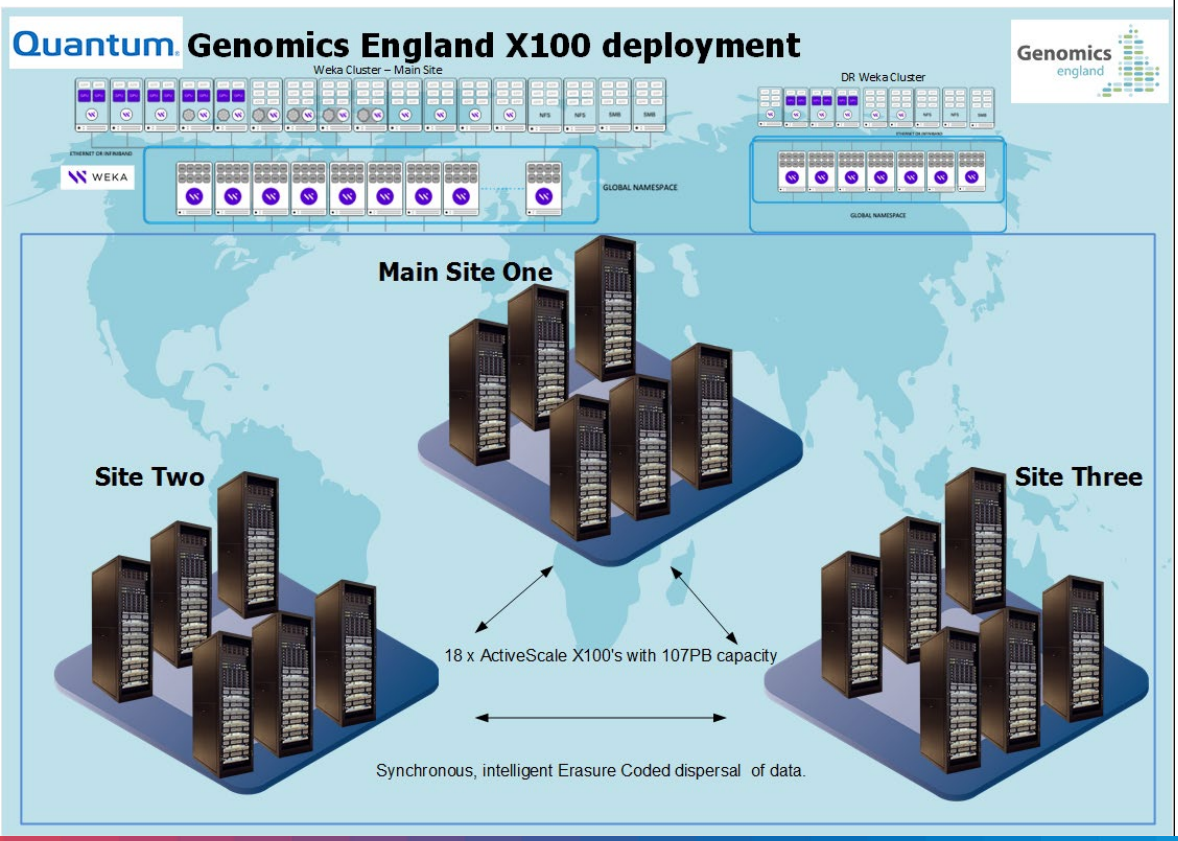
TAKING ON NEW CHALLENGES DURING THE COVID-19 PANDEMIC

Within a few years of deploying the new storage environment, Genomics England needed to expand again. The emergence of the COVID-19 pandemic in early 2020 presented new, urgent challenges for the global medical-scientific community, and Genomics England was in a prime position to help better understand who is susceptible to the virus. The organization committed to sequencing the genomes of up to 20,000 intensive care patients with COVID-19 plus up to 15,000 people with the virus who are experiencing only mild symptoms. In addition, the organization launched a next-generation genomic research platform to enable researchers and drug developers to access and use genomic data in their COVID-19 work.

Around the same time that Genomics England was ramping up participation in COVID-19 research, the ActiveScale solution platform was acquired by Quantum. A Quantum team facilitated a smooth transition for Genomics England.

With full confidence in Quantum and the ActiveScale system, Genomics England moved forward with expanding its object storage environment to accommodate COVID-19 work. The organization scaled to more than 100 PB of capacity.

Scaling was seamless with RAID (Redundant Array of Independent Disks). "What we love about the ActiveScale system is that its inherent architecture is underpinned by its RAID replacement technology, the intelligent, dynamic placement of erasure-coded data," says Ardley. That placement of data eliminates the need for system rebalancing, which can compromise performance and availability.



"We love the simplicity, ease of use, and architecture of the ActiveScale systems. These systems are truly designed to scale seamlessly to exabyte-scale deployments of on-premises and hybrid data storage solutions."

David Ardley,
Head of Technical Delivery,
Genomics England

ABOUT GENOMICS ENGLAND

Genomics England was created in 2013 to support the 100,000 Genomes Project—an initiative to sequence 100,000 whole genomes from patients with rare diseases, and their families, as well as patients with common cancers. The project—which subsequently expanded to up to five million genomes—is funded by the UK's National Institute for Health Research, NHS England (part of the UK's National Health Service), and other organizations. Owned by the Department of Health & Social Care, Genomics England aims to enable faster, deeper genomic research and bring genomic healthcare to all who need it.

In 2020, Genomics England began collaborating with a consortium of medical doctors and scientists to sequence whole genomes of people who have had COVID-19. The goal is to better understand the genetic susceptibility to the virus.

PROTECTING VITAL GENOMIC DATA

ActiveScale object storage protects data and provides the data resiliency that Genomics England needs for its critical work. The organization takes advantage of the geo-distributed capability of ActiveScale. With ActiveScale, the organization distributes data across three data centers, for full data protection against a major disaster such as site loss. Data can continue to be accessed for reading and writing at the remaining sites and withstand additional hardware failures, offering 19x9s data durability.

In addition, the technology helps ensure high data integrity by enabling a self-healing process through automatic data scrubbing, verification, repair, and optimization.

GAINING SCALABILITY WHILE CONTROLLING COSTS AND COMPLEXITY

With ActiveScale, Genomics England no longer faces the capacity limits of its previous NAS solution. The organization has been able to expand its object storage to support more genomic analysis and even take on additional COVID-19 work without a major storage overhaul. In the future, Genomics England can easily integrate ActiveScale object storage with Amazon S3-compliant public cloud environments for additional protection and scaling flexibility.

This scalable storage environment is also helping to reduce costs. According to Nephos, the Genomics England team decreased storage costs by 75 percent per genome compared with the previous environment. The organization is expected to reduce costs by 96 percent by 2023.

Just as important, the Genomics England team has been able to experience these benefits without adding complexity. "We love the simplicity, ease of use, and architecture of the ActiveScale systems," says Ardley. "These systems are truly designed to scale seamlessly to exabyte-scale deployments of on-premises and hybrid data storage solutions."