CASE STUDY

SPECTRA®
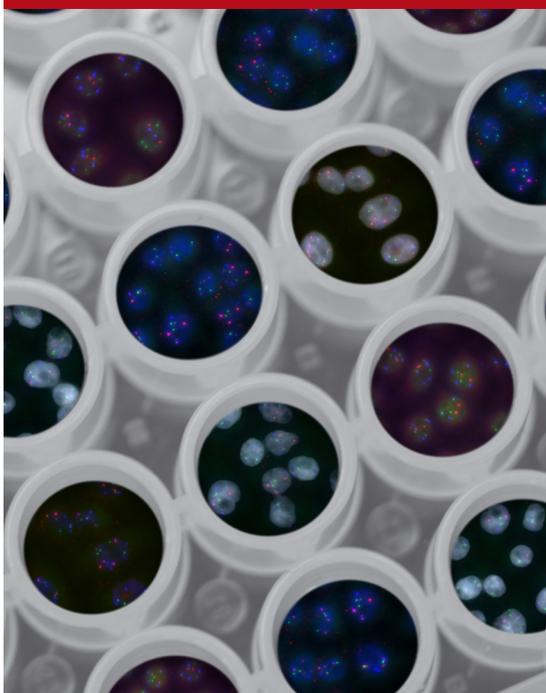
International Research Institute creates active archive for global disease research data with Spectra StorCycle

> " Our goal was to free up space on the primary storage by moving unused data to an archive infrastructure. Spectra's StorCycle provided insight into our data that we needed to make better storage management decisions, helping us scale our storage capacity on a lower-cost tier while keeping familiar access to files. "

End user at International Research Institute

Spectra BlackPearl and Spectra Stack Tape Library

The Challenge

This international research institute promotes global collaboration in disease research, coordinating research across countries and organizations. The institute is interdisciplinary, bringing together skills in diverse areas of scientific expertise to identify the causes of disease so that prevention measures may be adopted and the burden of disease and associated suffering reduced.

The international research institute has approximately 350 staff and students onsite. This institute's HPC cluster is a Linux-based cluster built in-house with open source software. They use a BeeGFS cluster file system to provide primary storage and deliver a parallel file system through the BeeGFS client. The cluster is mainly used by scientists working with genetics datasets and statistics. In addition, it is also used to store large sequencing data generated in-house and by external partners.

The institute's primary storage is structured with two HPE Apollo 4510 servers running BeeGFS services and presenting a 407TB volume to the cluster nodes (head and compute) through BeeGFS client. The BeeGFS volume is also accessible through direct access to files from user workstations (Linux, Mac or windows) via a Server Message Block (SMB) protocol. SMB is a network file-sharing protocol that allows applications on a computer to read and write to files and to request services from server programs in a computer network. At the time when they sought out this solution, the international research institute's primary storage was 93% full. The BeeGFS volume is backed up to a third HPE Apollo 4510 (437 TB) hosted in a second building, which was itself 76% full.

They estimated that over 100TB of data was not being used or accessed and should not reside on the expensive primary storage. They also identified another 70TB of data that should be moved to a long-term archive. In order to lower their overall cost of data ownership, the international research institute went looking for an active archive solution including both storage hardware and storage management software to move data from primary to lower tier storage based on user-defined policies.
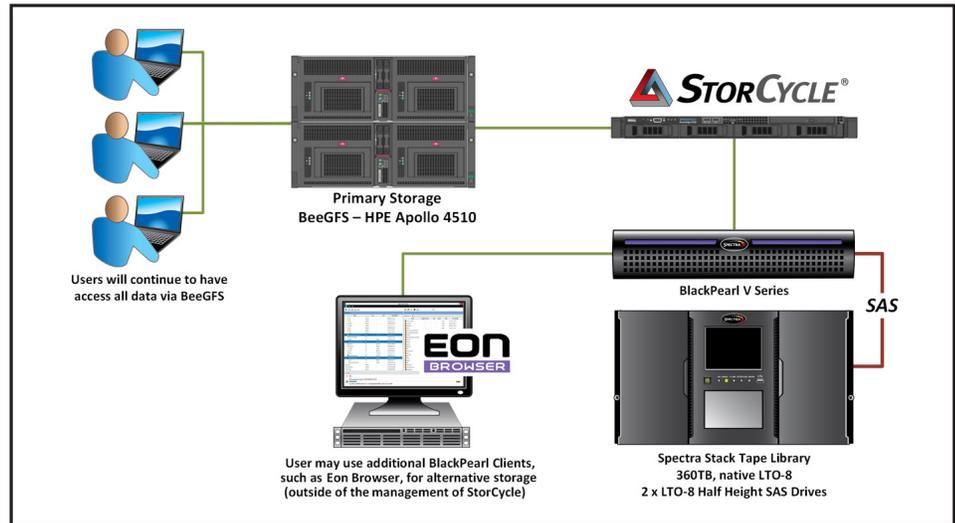
The Solution

The international research institute deployed a 2U Spectra® BlackPearl® Converged Storage System, a Spectra® Stack Tape Library with LTO-8 Half-Height tape drives and the Linux version of Spectra StorCycle® storage lifecycle management software. True to an active archive, the solution provides a combination of open system applications and hardware that intelligently monitor and migrate data across multiple storage devices while maintaining fast access for users. The object-storage-based BlackPearl platform provides a portal to multiple storage targets in a Perpetual Tier of storage, which complements the institute's existing Primary Tier of storage by providing lower-cost storage options.

The international research institute is using StorCycle to scan and identify inactive data in their primary storage based on user-defined criteria, providing insight into the scanned data to enable data managers to create adapted rules to move data. The software will then make a single copy of that data onto the Spectra Stack Tape Library, protecting it long-term on tape media that is encrypted at rest. StorCycle also enables users to autonomously restore their own files by leaving a link for familiar access, meaning end users will continue to have access to their data via BeeGFS. Finally, the institute also installed Spectra's free, open-source Eon Browser, which provides a simple interface to move files between local storage and BlackPearl, to handle data movement into the archive for storage outside of the management of StorCycle.

The international research institute's new solution offers the scalability they need by allowing them to easily add more tape drives and slots to the Spectra Stack Tape Library as their archive grows. The Spectra solution fulfilled the institute's purpose of increasing available storage on their HPC primary storage system, freeing up space on the primary storage by moving unused data to the archive infrastructure.

Primary Storage
BeeGFS – HPE Apollo 4510

Users will continue to have access all data via BeeGFS

BlackPearl V Series

SAS

User may use additional BlackPearl Clients, such as Eon Browser, for alternative storage (outside of the management of StorCycle)

Spectra Stack Tape Library
360TB, native LTO-8
2 x LTO-8 Half Height SAS Drives

Environment Snapshot

- Spectra StorCycle Storage Lifecycle Management Software
- Spectra BlackPearl Converged Storage System, 2U
- Spectra Stack Tape Library
- LTO-8 Half-Height tape drives
- BlueScale® Software with standard encryption
- EON Browser

Why Spectra?

- User accessibility and performance
- Insight into data for improved management
- Familiar access to archived data
- Open standard technology
- Scalability and integrity of object storage tape
- Effortless data migration and preservation with StorCycle

Solution Recap

Spectra StorCycle Storage Lifecycle Management Software – StorCycle is a storage lifecycle management software that automatically identifies and moves inactive data from primary storage to a lower cost tier that includes cloud, object storage disk, network-attached storage and object storage tape. Users can manually archive project-based directories and make additional copies for data protection. Providing seamless, familiar access and recovery of single files or entire projects, StorCycle enables organizations to manage, share and preserve their data for unequaled long-term protection and access.

Spectra BlackPearl Converged Storage System – BlackPearl provides a modern, simple portal to a multi-tier storage architecture that breaks away from legacy solutions based on capacity licensing models that drive greater cost when scaling. BlackPearl's simple RESTful API and integrated advanced bucket management enable users to create a storage architecture that is cost effective, easy to manage and scalable to exabytes of data.

Spectra Stack Tape Library – Built with maximum flexibility at its core, the Spectra Stack Tape Library meets backup, archive and perpetual storage requirements – helping to offload expensive primary storage tiers and preserve content long-term. Scalable up to 560 tape slots and 42 tape drives, the affordable tape library is designed to be easily installed, expanded and managed. Spectra gives users the ability to protect their investment by migrating existing tape drives and media into the Spectra Stack Tape Library.