



CASE STUDY

Hutchison MRC Research Centre at the University of Cambridge reduces storage infrastructure costs by up to 60%

“ Addressing the needs of customers with reliable data storage lies at the heart of the Spectra and Arcitecta relationship. The joint solution enables customers to better manage their data and metadata by optimizing multiple storage targets, retrieving data efficiently and tracking content and resources. ”

Matt Starr, Chief Technology Officer, Spectra Logic



Spectra BlackPearl Hybrid Storage Platform and Stack Tape Library

About the Hutchison MRC Research Centre

The Hutchison/MRC Research Centre at the University of Cambridge is a cancer research facility. Built in 2001, it is now the leading site for basic and translational cancer research in Cambridge. Located on the Cambridge Biomedical Campus, the Hutchison/MRC Research Centre houses active clinicians as well as basic scientists, enabling the rapid translation of discoveries made at the research bench into clinically viable applications at the patient's bedside.

The Challenge

The Hutchison/MRC Research Centre's existing infrastructure consisted of several network-attached storage (NAS) devices used by more than 200 users. Data storage was reaching its limits, and the equipment was reaching its end of life, requiring additional storage to be purchased on an ad-hoc basis to meet the growing demand. As a result, the Hutchison/ MRC Research Centre data became scattered across several different data stores, making it difficult for researchers to understand where to find the data they needed.

The Hutchison/MRC Research Centre had developed a complex set of needs. This would include: run applications whose databases are accessing files multiple times per second, potentially opening and closing those files every time; be able to securely share research data with collaborators; and limit access to particular pieces of data in particular locations within the folder hierarchies.

As the volume of data continued to expand, and the storage environment continued to expand, users would need the ability to easily and precisely search and access content on tape to continue to extract its value into the future.

The Hutchison/MRC Research Centre also required a full backup of the file system that could operate at scale. This would allow any file or directory in managed storage to roll back to an earlier version if something were mistakenly deleted, overwritten or a new version had been created.

Most critically, the Centre needed an immediate solution that could be seamlessly integrated into the existing environment to maintain stable workflows and solve immediate storage problems, with an evolutionary approach to implementing these other data management capabilities in the future.

The Solution

The Hutchison/MRC Research Centre put out to tender for a solution that would allow it to virtualise its data storage. Arcitecta®, in partnership with Spectra Logic, proposed a new tiered data management system using its Mediaflux® platform, backed by Spectra Logic's BlackPearl® Hybrid Storage Platform. Mediaflux's policy-based virtualisation would leverage the power of metadata to combine the existing, dispersed data silos.

CASE STUDY: Hutchison MRC Research Centre

Mediaflux was deployed by scanning all the existing storage data and virtualising its location into a single global namespace. As a side benefit, this process detected corrupt or duplicate files. Now, the Hutchison/MRC Research Centre's disparate storage devices could be managed similarly as a single asset hierarchy. Users could then view this hierarchy via SMB or NFS so that researchers could access files as though they were in the same place as before, regardless of where the data was now stored.

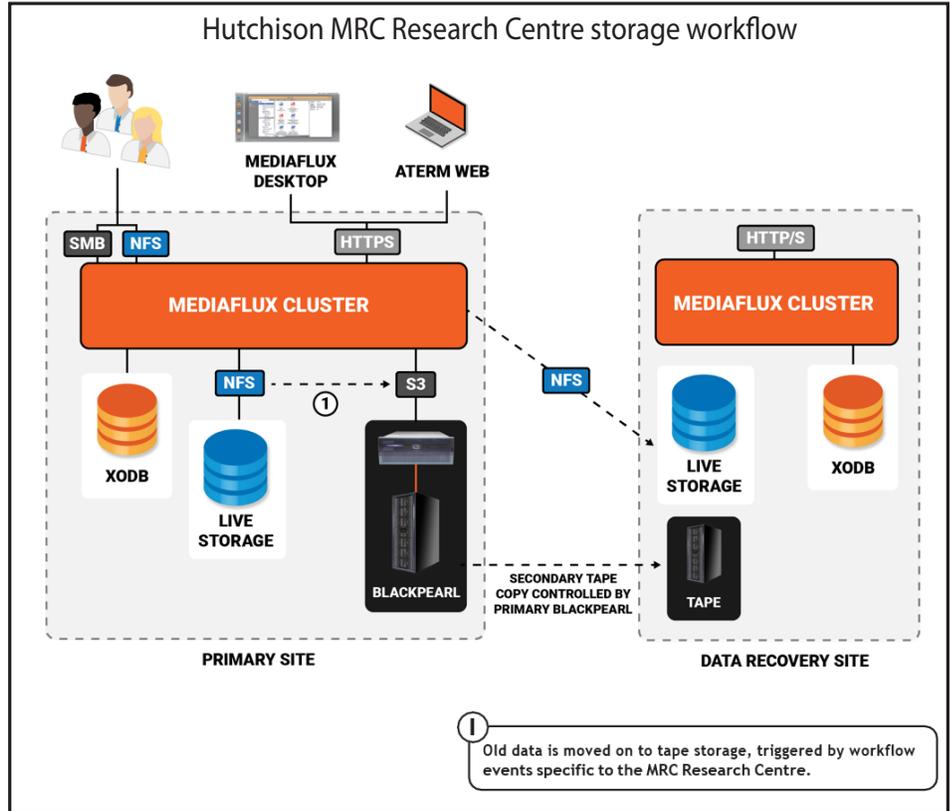
The next major milestone was to integrate Spectra Logic's BlackPearl so the Centre could start moving infrequently used data to the newly created archive tier on tape. Based on usage policies, data was moved from the primary to the archive tier, freeing up expensive disk storage. Meanwhile, this was all transparent to the users, who could still locate their data exactly as in the past.

Working automatically in the background, Mediaflux provided analysis workflows and data tiering, and extracted and created metadata to search and act on billions of files, unifying data silos into a secure distributed collaboration environment across a single global namespace.

Simultaneously, the scalable storage of BlackPearl, together with Mediaflux, unlocked the most efficient storage targets including online disk, nearline disk, deep storage tape, and even public cloud.

Mediaflux and BlackPearl have provided the Hutchison/MRC Research Centre with a storage capacity up to 60% less expensive by leveraging Mediaflux's scalability and agility across distributed storage resources while keeping data readily available to researchers. To the user, data is easier to find and data-intensive workflows are smarter and more flexible, which has significantly reduced the time researchers and administrators spend wrangling their data.

Arcitecta leveraged existing technologies in the Mediaflux stack to develop its 'Point-in-Time' feature to meet the Hutchison/MRC Research Centre's needs for a real-time backup application that could restore files to anypoint in time in their environment. 'Point-in-Time' now works



across all stored data to automate backup and archiving workflows, better enabling the use of cheaper storage technologies, and eliminating unnecessary and duplicated data.

Using Mediaflux and BlackPearl, the Hutchison/MRC Research Centre has reduced storage infrastructure costs by up to 60%; improved data discovery using metadata; virtualised data silos into a single global namespace; and gained the ability to leverage scalable storage on demand.

Environment Snapshot

- Two Spectra® Stack Tape Libraries with LTO-8 Half-Height tape drives
- Spectra BlackPearl Multi-Purpose Hybrid Storage Platform
- Arcitecta Mediaflux platform

Partner Profile:

ARCITECTA®

Arcitecta specializes in data management systems for large-scale distributed data. Its core product, Mediaflux™, is a flexible data + metadata management solution that virtualizes otherwise incompatible data stoves into a secure distributed collaboration environment.

Mediaflux contains powerful multi-site data federation and replication capabilities to optimize storage utilization and rapidly adapt to changing use cases, applications, and security requirements.