

Hutchison MRC Research Centre at the University of Cambridge

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. Without a data management system in place, the existing infrastructure at the facility was centred on storage, consisting 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 opportunity

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® Converged Storage System. Mediaflux's policy-based virtualisation would leverage the power of metadata to combine the existing, dispersed data silos.

The challenge

The Hutchison/MRC Research Centre had developed a complex set of needs compared with how SMB is commonly used. In essence, the Centre wanted to use Mediaflux as a pure storage virtualisation technology. This would include using it to:

- 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 increase, and the storage environment to expand, users would need the ability to easily and precisely search and access content on tape to continue to extract a 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

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.

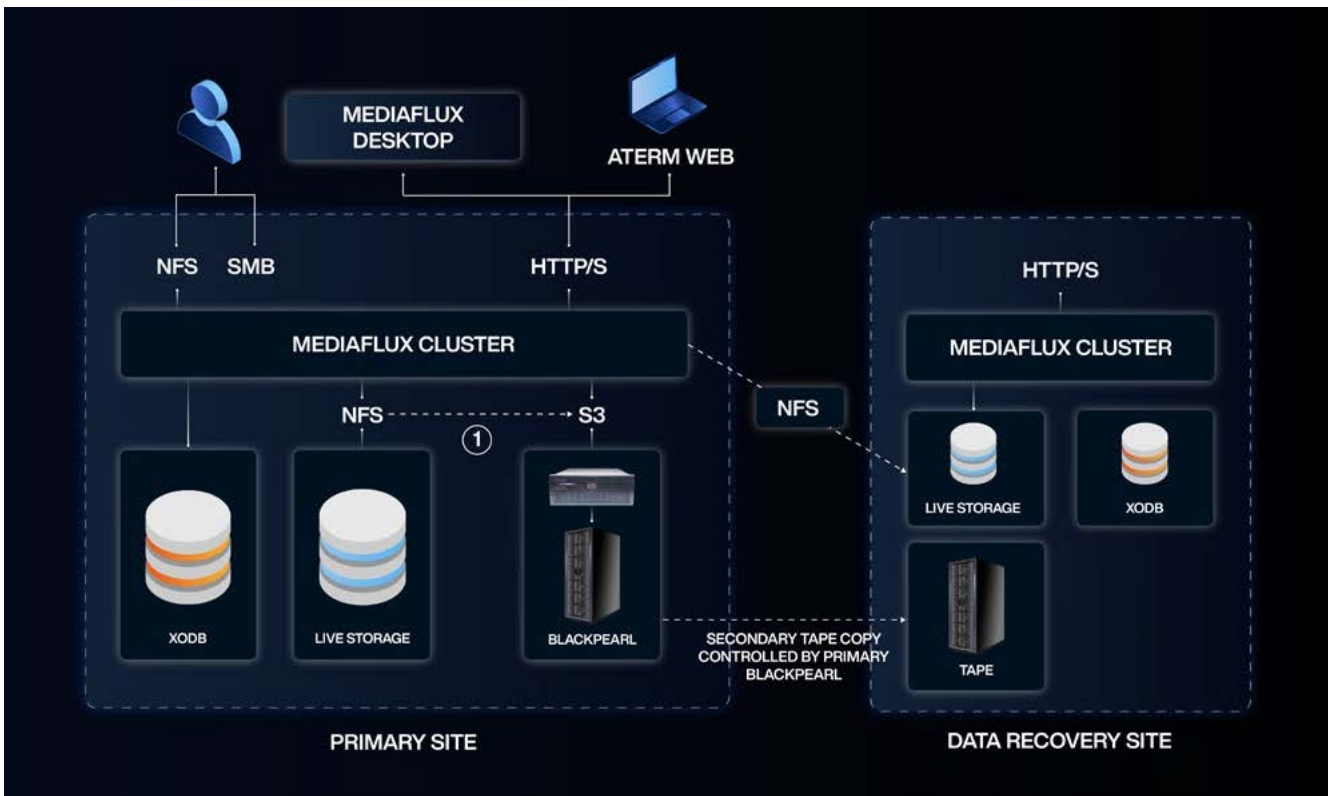
Working automatically in the background, Mediaflux provided analysis workflows and data tiering, and extracted and created metadata to search and act on billions of files, unifying data silos into a secure distributed collaboration environment across a single global namespace. Simultaneously, the scalable storage of BlackPearl, together with Mediaflux, unlocked the most efficient storage targets including online disk, nearline disk, deep storage tape, and even public cloud.



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 any point 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.

“As time was of the essence while new features were introduced, an Arcitecta team member was available in the same timezone for mission-critical functions and to provide rapid access to technical support and resources. The team worked incredibly hard and fast to respond to our immediate challenges. I have never worked with a group of engineers who are so competent at what they do.”

- Mr Stephen Jones (IT Manager, Hutchison/MRC Research Centre)



¹ Old data is moved on to tape storage, triggered by workflow events specific to the MRC Research Centre.

The outcome

Mediaflux and BlackPearl have provided the Hutchison/MRC Research Centre with a storage capacity up to 60% cheaper 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.

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
- Gained ability to leverage scalable storage on demand

On reflection/future uses

At first glance, the MRC Research Centre was unaware of what Mediaflux could do and the number of ways it could be used. Having overcome the immediate need to virtualise its storage, the MRC Research Centre is now exploring other areas for further optimisation, such as leveraging Mediaflux's enhanced search capabilities and best practice metadata management to improve the visibility of the 54,000,000 files currently on the system, and collaboration tools and High-Performance Computing (HPC) task management. Overall, the future with Mediaflux unlocks more time to dedicate to research rather than wrangling data, which will ultimately improve the research outcomes for a wider number of researchers at the the Hutchison/MRC Research Centre.

