CASE STUDY

MEDIAFLUX®
MADE BY ARCITECTA

# Cancer Institute

## Research Computing
## Storage Management System

A world-renowned cancer treatment and research center, known for its innovative research and compassionate patient care, is playing a significant role in advancing cancer treatment and improving patient outcomes by using Mediaflux® for its Research Computing Storage Management System.

# Summary

### Challenges

- Managing NAS server sprawl
- Existing and new scientific instruments overwhelm servers  with exponential data growth
- Scientists spend time moving data between storage silos
- Frequent workflow disruptions when data and accounts are  migrated to new servers

### Results

- Scale-out global namespace across over 30 ZFS NAS servers
- Replication to DR site
- Automatic archiving to AWS S3 Deep-Archive
- Much lower cost than alternatives

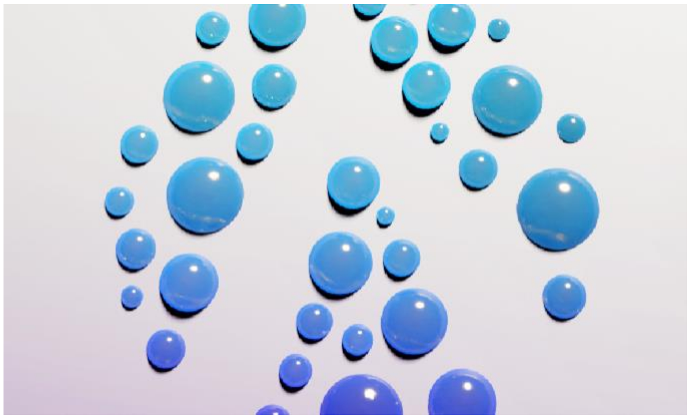
OPERATING SYSTEMS FOR META+DATA

www.arcitecta.com
talk@arcitecta.acom

ARCITECTA

*“Mediaflux on our ZFS cluster delivers better performance, replication and tiering than Scale-Out NAS systems we looked at – with greater flexibility at much lower costs.”*

**Technical Director, Research Computing Services**
**Informatics & Analytics | Computational Solutions**



## The Challenge

The Center was storing their research data from over 200 labs – consisting of large amounts of genomics, CryoEM image, and scientific data – totalling 6 PB and more than 2 billion files on over 30 ZFS Network Attached Storage (NAS) servers. This data supplies the Center’s various analyses and AI pipelines. Each researcher was assigned to an individual server and allocated space. Each server had its own authentication method and filesystem structure, which required scientists to know exactly what particular storage system held their data. When a server reached capacity, research data (and the researcher) had to be moved to another server with available space in a “Tetris-like” manner – a painful process for both IT and researchers that became untenable.

The Center was looking for a way to eliminate managing the different logins on over 30 different servers. Moving to traditional enterprise Scale-Out NAS was determined too expensive and didn’t give them the flexibility required.

## The Solution

The Center front-ended the ZFS storage servers with Mediaflux, providing a scalable load-balanced global namespace for research- ers and instruments with easy management for IT. Mediaflux also replicates data to off-site storage, and automatically archives data to low-cost AWS S3 Deep-Archive.

### Global Namespace Across Over 30 ZFS NAS Servers

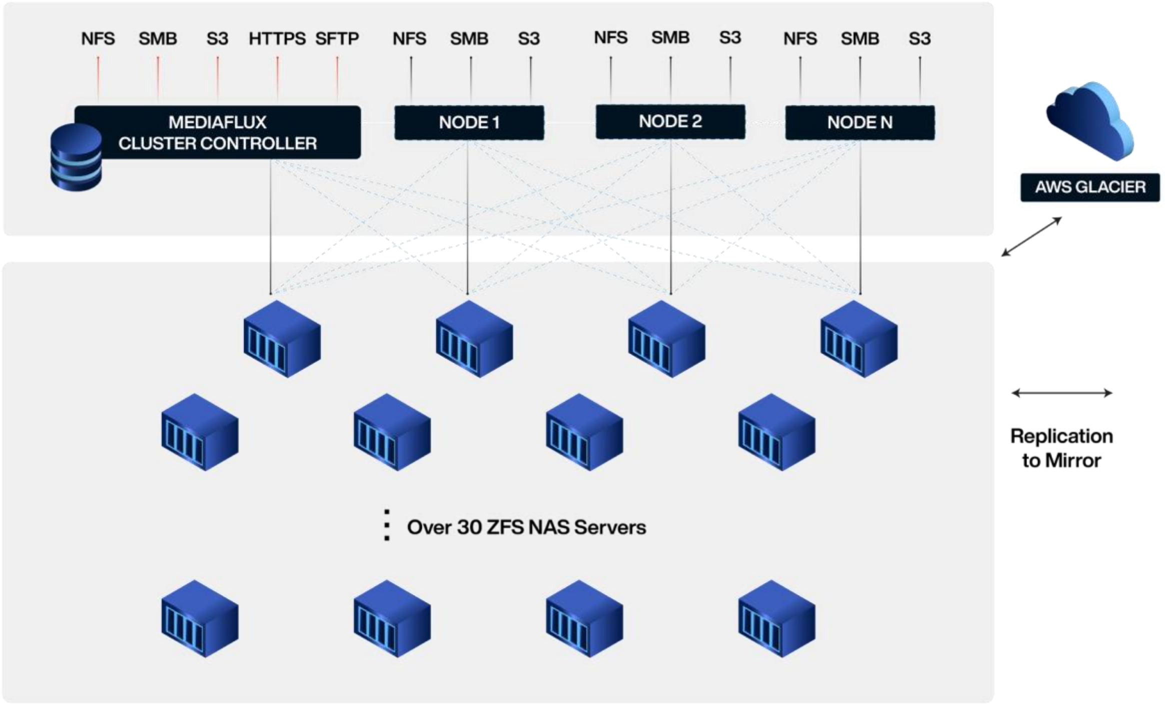
Mediaflux unified the view of all data stored across all the different storage servers. This makes it easier to access research data, regardless of where it is stored. All of the storage servers can be managed as a single entity, which simplifies the administration of the storage environment.

### Replication to Mirrored Storage at DR Site

If one copy of the data is lost, corrupted, or becomes unavailable, there are other copies that can be used, reducing downtime, keeping research going.

### Automatic Archiving to Low-Cost AWS S3 Deep-Archive

The Center saves on storage costs with high durability and availabil- ity. There is no need to manually manage the storage of data or data retention policies.



THE CENTER'S RESEARCH COMPUTING STORAGE MANAGEMENT SYSTEM

## Benefits

- Compelling economics and complete storage visibility and control
- Improved research productivity
- Greatly reduced day-to-day systems administration
- Visibility into each lab's storage usage
- Infrastructure easily scales to respond to new research
- Automatic, namespace expansion to AWS S3 Deep-Archive further lowers storage costs

## The Center's Research Computing Storage Management System

The Center's Research Computing Department is committed to providing reliable and efficient data storage and access solutions to support research initiatives. It also actively surveyed researchers for their requirements and technology vendors for their capabilities.

The Center's Research Computing Storage Management System consolidates all existing storage servers under a single research data namespace, accessible by high-performance scale-out standard SMB and NFS with comprehensive data protection and lifecycle policies that automatically move data between main, replicated, and low-cost AWS Glacier storage. This architecture reduces complexity and confusion, eliminating data silos and making all file assets accessible via file system protocols.

All file assets can be accessed via a single interface, making searching and reporting on file assets more efficient. This system simplifies data management and enables researchers to locate and access the data they need quickly. Overall the Center's system provides a more streamlined and effective data storage solution for research computing, helping to improve research productivity and facilitate new discoveries.

The Center's Research Computing Storage Management system also features redundant automatic data backup and disaster recovery capabilities and off-site data replication options to ensure data safety in the event of hardware or system failures.

Future plans include exploiting Mediaflux to provide role-based access controls, metadata extraction, and tagging to accelerate data discoverability.