PRESENTATION NOTES
RMACC 2023

The Impact of Data Management on HPC Workflows

Current HPC Workflows

More numerous and higher-resolution experiments and simulations are producing more and more data, putting stress on HPC workflows.



Tiered Storage Architecture

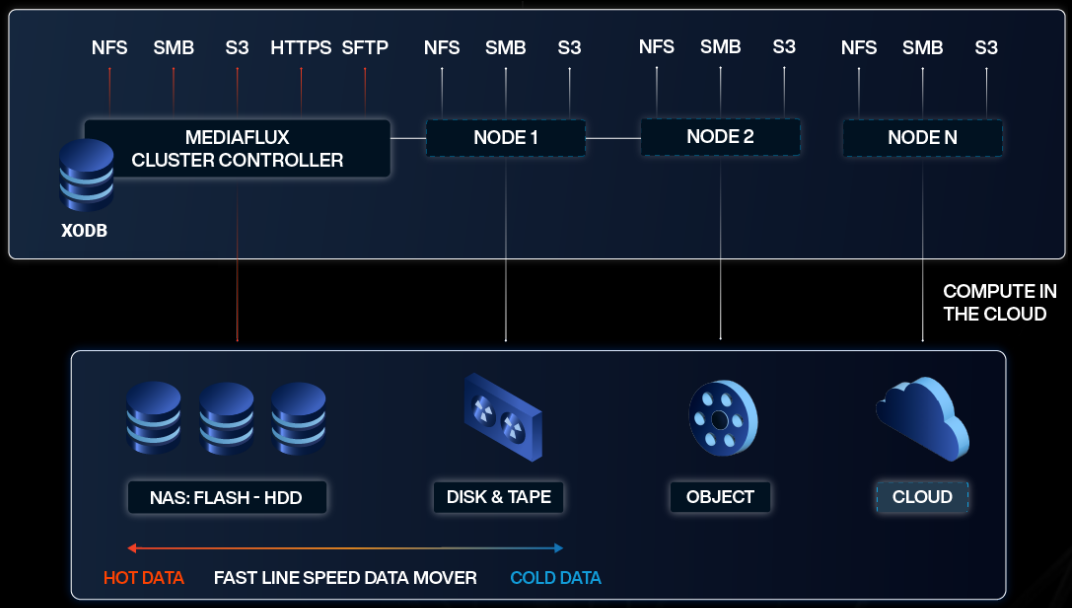
Tiered storage architectures solve the problem for HPC workflows. It optimizes performance and cuts cost.

The simple storage abstraction hides the underlying storage complexity, moves input data to a fast storage tier, moves final outputs to persistent storage for analysis, and archives to the lowest cost media available.



ARCITECTA
OPERATING SYSTEMS FOR META+DATA

Arcitecta®, Mediaflux® and XODB® are registered trademarks of Arcitecta IP Pty. Ltd. in the USA and trademarks of Arcitecta IP Pty. Ltd. in Australia
© 2023 Arcitecta IP Pty. Ltd.
www.arcitecta.com
talk@arcitecta.acom



VIRTUALIZED STORAGE - SINGLE GLOBAL NAMESPACE

MEDIAFLUX®
MADE BY ARCITECTA®

Mediaflux® automates movement and placement of datasets consumed by HPC workflows.

- Single global namespace view of data, no matter where stored (100s of Billions)
- Multi-protocol support enables data to be accessed by any application
- Intelligent data placement and movement (tiering and migration) so data is in the right place on the right technology at the right cost (forever)
- Replication for Disaster Recovery (DR)
- Extensive metadata harvesting, annotation, and cataloging
- Global, distributed access
- High-speed WAN file transfer
- Multi-factor authentication access control, approval workflows and administrative actions
- File and file system versioning ensures provenance and easy data recovery
- End-user self-service tools free IT from routine data recovery tasks

© Copyright 2023 Arcitecta Inc.
All rights reserved

Case Study: Cancer Center

Research Computing Data Management System

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.

The Problem

- Managing NAS server sprawl across over 30 ZFS servers
- 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



© Copyright 2023 Arcitecta Inc.
All rights reserved

Case Study: Cancer Center

The Solution



1. Creating a Global Namespace across ZFS Servers

Front-ended the ZFS storage servers with Mediaflux

Unified the view of all data stored across all the different storage servers (6PB, 2B files)

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

2. Replication to DR Site

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, and keeping research going.

3. Automatic Archiving to Low-Cost AWS S3 Deep Archive

Automatic Archiving to Low-Cost AWS S3 Deep Archive

Automatic, namespace expansion to AWS S3 Deep-Archive

The Center saves on storage costs with high durability and availability

There is no need to manually manage the storage of data or data retention policies

© Copyright 2023 Arcitecta Inc.
All rights reserved

Case Study: Cancer Center

The Result

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

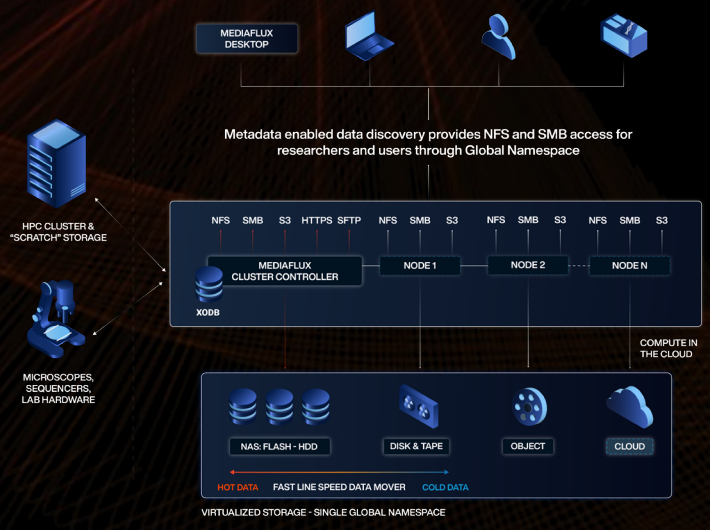
Infrastructure easily scales to respond to new research

HPC Scratch

Mediaflux copies input data to HPC Scratch, and once ready

Have Mediaflux initiate the HPC job

Mediafluxcan be integrated with batch scheduling systems, such as PBS and SLURM to automate the process



© Copyright 2023 Arcitecta Inc.
All rights reserved

What Makes Arcitecta Different?

We are a “First Principles” Company.

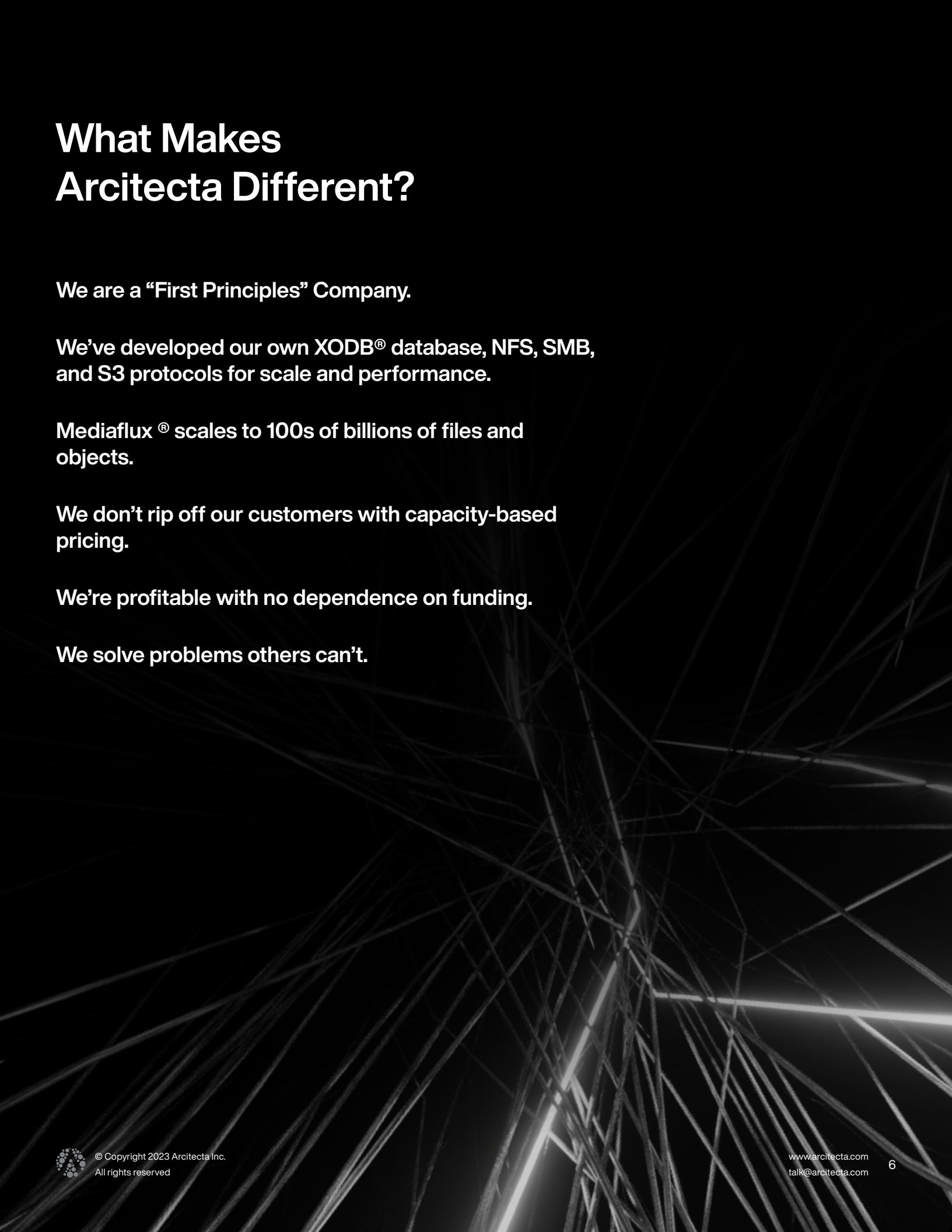
We’ve developed our own XODB® database, NFS, SMB, and S3 protocols for scale and performance.

Mediaflux ® scales to 100s of billions of files and objects.

We don’t rip off our customers with capacity-based pricing.

We’re profitable with no dependence on funding.

We solve problems others can’t.



© Copyright 2023 Arcitecta Inc.
All rights reserved