CAMBRIDGE UNIVERSITY – DATA INTENSIVE UNIVERSITY RESEARCH FINDS VALUE WITH OPEN STORAGE

Cambridge, United Kingdom
https://www.cam.ac.uk
Education, Research, Scalability

BENEFITS

- Modern SDS is easily expandable as storage grows
- Clear, stable pricing addresses budgetary constraints
- Strong data protection and replication secures critical research

THE CHALLENGE

Much has been written about the hunger for data storage in university research, whether in the humanities, arts, social sciences, genomic studies, astronomy, or seismic research. According to a recent paper by Cambridge University, “Data acquisition rates in fluorescence microscopy are exploding due to the increasing size and sensitivity of detectors, brightness and variety of available fluorophores, and complexity of the experiments and imaging equipment. An increasing number of laboratories are now performing complex imaging experiments that rapidly generate gigabytes and even terabytes of data, but the practice of data storage continues to lag behind data acquisition capabilities.”

Alongside the growing storage needs, universities must balance costs, controlling them to support individual research projects. In addition, there is a decision to be made between on-premises systems and cloud infrastructure, with universities uniquely able to evaluate both architectures. Cambridge University needed to balance those complex demands to efficiently support research needs for modern storage systems.

TrueNAS OPEN STORAGE

UNIVERSITY OF CAMBRIDGE

THE SOLUTION

While cloud storage providers offer universities a computing environment without having to manage a data center, they can greatly differ in how they price and deliver storage solutions. Cloud pricing structures do not typically meet the budgetary requirements of most universities because pricing scales too linearly. As usage levels increase, bulk storage discounts do not deliver the promised savings, significant data movement charges add up on large data volumes moved to and from the cloud, and services typically lack the latency and bandwidth necessary for rapid access to data. These limitations become more prominent when the amount of data stored exceeds a few terabytes.

In contrast, Open Storage has evolved over the past decade and now embracesnext-generation software-define storage (SDS) and hyper-converged storage (HCI), making it a viable option to tame 'data-explosion' challenges facing university IT professionals and academics. The capabilities and maturity of Open Storage have advanced quickly over the past several years as many contributors to code have devoted resources to advanced platforms such as OpenZFS. Features previously only available at a premium with proprietary and cloud storage solutions, such as snapshots, replication, advanced compression, access to APIs, etc., have now become standard items in Open Storage.

Cambridge University's choice to provide storage for its research needs is TrueNAS, which incorporates all of the features of the OpenZFS filesystem while adding a user-friendly web interface for storage configuration and management.
TrueNAS also features a data protection suite that provides the ability to combine RAID protection, replication, snapshots, automatic corruption repair, and optional high availability to deliver enterprise-class data center protection to satisfy mission-critical research environments. The TrueNAS solution has transparent pricing that scales in a linear fashion that helps academic researchers plan accurate budgets for years to come.

CHALLENGES

- Volumes of research data exploding
- Cost control and allocation are important to universities
- Critical research requires safe & secure storage

THE ARCHITECTURE

A typical TrueNAS storage system can either use server-grade hardware or systems with redundant storage nodes for five 9's availability and 24×7 operation available from iXsystems. Either system type can then be expanded by adding expansion shelves with additional disks. In the case of the TrueNAS M60, it can scale to over 20 PB and serve data over ethernet connectivity up to 100 Gigabits (using NFS/SMB/CliFS for file storage, iSCSI for block storage, or an S3 interface for object storage) or fiber channel up to 32Gb/s.

For one of its research projects, Cambridge solved its storage needs for 250 usable terabytes with a server-grade system, expansion shelves, and less than 100 drives. For robust data protection, they used a RAID-Z2 layout which is similar to RAID-6 with dual parity drives for redundancy.

As highlighted by Cambridge University, network connectivity for the Open Storage infrastructure is made using 10 GbE cards that can accommodate either copper wire or optical fiber connectors. The storage server is connected to two networks: a local network between the data storage system and a lightsheet microscope using a 10 Gbps transfer rate and a general-use 1 GbE connection to the rest of the Cambridge University's network. This dual connection ensures that even if the university network is unavailable due to failure or maintenance, administrators can still use TrueNAS to store large datasets from the light-sheet microscope.

For added resiliency and protection of the data from corruption or human error like accidental deletes or overwrites, Cambridge University also makes extensive use of periodic and unlimited snapshots to give them a "save point" to rollback to in case an issue arises. Those snapshots are also replicated to another TrueNAS system to keep backups. TrueNAS also provides the capability to tier the data to the public cloud for a secondary backup or to complement a Disaster Recovery strategy.