Curating and Managing Archival Data: Petalibraries

Data is growing at a limitless rate, roughly doubling every two years, placing an immeasurable burden on the keepers and archivers of that data: corporate libraries, universities, and public institutions. Universities and other higher education environments are driving much of that growth, creating data through research programs and by taking on the responsibilities of archiving and curating stored data in libraries for posterity. Higher education has a unique requirement to maintain, curate, and make available all of this data to researchers, faculty, students, and the public while also being constrained by public and/or grant-related funds. Historically this forced institutions to choose between capacity, scale, and cost. With Qumulo, higher education research environments and libraries no longer have to choose. Qumulo Core delivers a data-aware scale-out storage platform that scales both capacity and performance based on demand when and where needed.

Petalibraries – Data Storage at Exponential growth.

Many universities and other higher education institutions tasked with maintaining limitless digital archives are turning to a new class of digital storage, Petalibraries: massive petabyte-sized digital libraries that contain billions of unique files. Digital documents, media such as audio and video files, and newer types of files such as digital maps and 3D models all take up a tremendous amount of storage space, and universities are struggling to provide cost-effective methods for growing the storage while keeping pace with the data growth.

Qumulo's next generation data-aware scale-out NAS allows universities and other higher education institutions to deploy and manage petalibraries using a pay-as-you-grow architecture, scaling the infrastructure as needed to keep pace with the exponentially growing data demand. Petalibrary storage administrators can simply add new storage nodes as needed and Qumulo Core will transparently grow the storage cluster without impacting existing data. Institutions have control and flexibility of deploying and growing Qumulo Core on their terms, anywhere to meet the demanding research and archival needs of higher education.

Curate Everything: Real-time Visibility Across the Entire File System

In addition to raw scale and capacity, one of the largest challenges facing any institution responsible for managing a petabyte-sized storage infrastructure is visibility: Where does the data reside? Who can access what data? Where is the storage consumed? Who and what is impacting the performance of the storage? Curating a massively large file system is about managing the data rather than managing the storage. The ability to scale the management of billions of files on-demand becomes a requirement which grows in tandem with raw petabytes of storage.

Qumulo Core's integrated file system analytics allows higher education institutions to obtain real-time answers about their data footprint at incredible scale. With metadata built directly into the Qumulo Scalable File System (QSFS), storage administrators have visibility into their data and storage, usage patterns, activity by group or department, and which workloads are impacting performance and capacity. Petalibrary data can be organized and managed in a way that meets the need of higher education without sacrificing the ability to catalog, search, and retrieve data instantly.

Qumulo Features for Higher Education and Petalibraries

- **Massive Scalability:** Scales incrementally, allowing institutions to simply grow storage capacity and performance by adding more nodes as needed

- **Analytics and Visibility:** Provides unmatched real-time visibility into data and usage at the file-level

- **High Performance:** High IOPS and throughput with low latency, allowing multiple groups to access library content without sacrificing performance

- **Multiple File System Architecture:** With support for NFS, SMB, and REST, Qumulo Core can integrate into any existing storage infrastructure and support separate groups with unique storage needs

Qumulo Care Cloud-Based Monitoring

Qumulo Care Cloud-Based Monitoring remotely monitors & reports status of your Qumulo cluster:

- Provides proactive monitoring of your Qumulo cluster

Monitoring covers both software and hardware events like:

- Capacity triggers
- Latency
- Throughput
- IOPS

Manage Archival Data, Not Storage

Qumulo's data-aware scale-out NAS software gives institutions real-time visibility into and manageability of their petalibrary data footprint – critical data that until now was completely invisible and stored away in large, difficult to manage storage containers.

Data is What Matters

Data should be visible and storage should be invisible, a key tenet for large-scale libraries responsible for housing the world's data. With metadata built directly into the Qumulo Scalable File System (QSFS) as an intrinsic element, storage administrators have visibility into their data and storage, usage patterns and which users or workloads are impacting performance and capacity.By supporting both file-level and block-level data access, administrators have complete flexibility in how their unique data is archived and cataloged.

Achieve Cost Containment and Flexibility

Built on Linux, Qumulo's data-aware scale-out NAS is a software-only solution that runs on commodity hardware, on dedicated appliances, or in virtual machines providing the ultimate flexibility and cost advantages while seamlessly scaling to archival library needs.
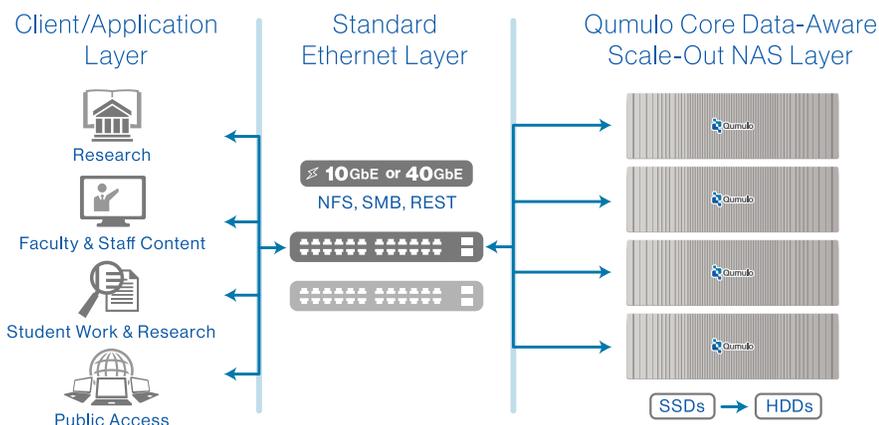
No Compromise Scale-out NAS

Qumulo Core is optimized for the widest range of workloads and file sizes including transactional and sequential access patterns and small and large files. Whether storing billions of small documents, massively large digital movies, or anywhere in between, Qumulo Core offers the same level of cost and capacity performance required for the most demanding archival needs.

Fully Programmable Storage Access and Management

Qumulo Core was designed with flexibility first. Data access, management, and integration are all 100% programmable an interactive REST API allowing storage administrators to integrate Qumulo Core with existing file- or block-based applications. The REST API allows Qumulo Core to plug into existing management tools including custom-designed solutions that can integrate with a suite of development environments.

Qumulo Benefits for Higher Education

- Qumulo Core's integrated file system analytics allows higher education institutions to get real-time answers about their data footprint at incredible scale

- Support integration with existing file and management tools via an open REST API with active GitHub open-source community

- Provide real-time visibility into data and storage solving data management problems created by first generation scale-out NAS

- Agile development supports rapidly bringing new features to market that support the growing needs of higher education and digital archiving

Higher Education Architecture

Client/Application Layer
- Research
- Faculty & Staff Content
- Student Work & Research
- Public Access

Standard Ethernet Layer
10GbE or 40GbE
NFS, SMB, REST

Qumulo Core Data-Aware Scale-Out NAS Layer
Qumulo
SSDs → HDDs

Contact Qumulo
to learn more about our Higher Education solutions:

1-855-478-6856
info@qumulo.com

Qumulo

Q121 0915