Qumulo

Accelerating Biomedical Breakthroughs with Qumulo

ANY DATA.
ANY LOCATION.
TOTAL CONTROL.

OVERVIEW

Cryo-Electron Microscopy (CryoEM) is revolutionizing life sciences and pharmaceuticals by enabling atomic-level 3D visualization of proteins, viruses, and molecular complexes. This process generates massive, complex datasets that demand a high-performance, scalable, and secure data infrastructure.

Qumulo's Cloud Data Fabric (CDF) delivers a unified, cloud-agnostic platform for ingesting, processing, and sharing CryoEM data, accelerating drug discovery and vaccine development across hybrid and multi-cloud environments.

CHALLENGES FACING LIFE SCIENCES

Data Explosion and Ingest Bottlenecks Are Slowing Science
CryoEM workflows produce tens to hundreds of terabytes per project, with thousands of high-resolution electron micrographs and I/O-intensive pipelines.

Hybrid Architecture Needs
Data moves between wet labs (on-premises) and dry labs (cloud/HPC), requiring seamless integration and real-time availability.

Workflow Complexity
Traditional workflows involve manual replication, data duplication, and complex scripts to move data between microscopes, processing sites, and scientific teams.

Performance Bottlenecks
GPU-heavy computation and the need to support diverse protocols (SMB, NFS, S3, POSIX) place massive demands on storage and throughput, often resulting in performance bottlenecks.

Collaboration & Compliance
Achieving secure, compliant sharing across research teams and institutions often requires a trade-off, as implementing it without sacrificing essential performance or flexibility is challenging.

KEY BENEFITS

Accelerated Time-To-Results
Real-time data movement and processing reduce delays and enable rapid scientific insights.

Simplified File-Object Operations
Eliminates the need to copy data in and out of cloud object storage for processing, reducing latency and eliminating the need for additional ingestion or transfer tools.

Scalability
Handles billions of files and multi-terabyte objects, supporting growth in data volume and future expansion.

Edge-to-Core-to-Cloud
Deployable at the edge, across hyperscalers (Azure, AWS, GCP, OCI), or on-premises — enabling hybrid and multi-cloud research workflows.
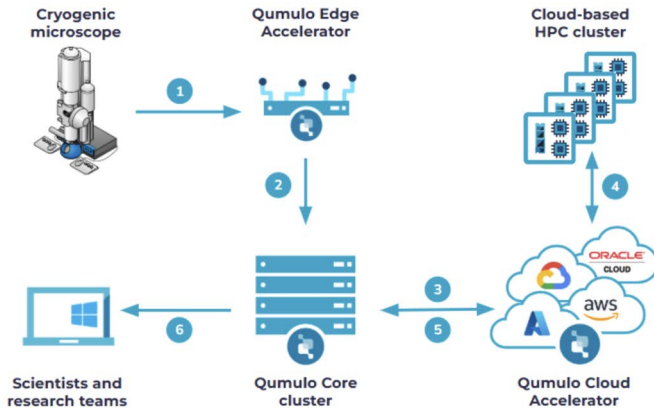
Cost Efficiency
Eliminates duplicated storage and manual data movement, lowering infrastructure and operational costs.

Compliance & Security
Meets regulatory requirements for clinical research, with robust security and auditability features.

SOLUTION OVERVIEW

Cloud Data Fabric's unique architecture connects edge labs, core data centers, and cloud environments into a single globally visible namespace, ensuring strictly consistent data access and low-latency data performance across all stages of CryoEM workflows.

1. **Edge (Lab) Ingest:** A Qumulo Edge Accelerator appliance provides a local ingest point for microscope output at line speeds, guaranteeing uninterrupted instrument operation.

2. **Core (Data Center or Private Cloud):** Cloud Data Fabric consolidates ingested and modified data into a single globally accessible dataset in which all data and metadata operations are visible in real time from any location.

3. **Cloud Acceleration:** Qumulo Cloud Accelerators can be deployed on AWS, Azure, GCP, and OCI, pulling ingested image files at line speed from the on-premises Qumulo cluster and delivering them in real time directly to a cloud-based CryoSparc HPC engine without object storage staging. CDF provides native support for SMB, NFS, and S3 protocols, enabling workflow compatibility across hybrid environments.

4. **Real-Time Visibility:** Within seconds of an image being uploaded from the microscope via an Edge Accelerator, the CryoSparc engine's GPUs are able to read it, analyze it, and generate new 3D models that it writes back to the local Cloud Accelerator appliance.

5. **Real-Time Updates:** Processed results from CryoSparc are immediately propagated from the Cloud Accelerator back to the on-premises Qumulo cluster.

6. **Real-Time Collaboration:** From the on-premises Qumulo cluster, the new 3D imagery is available to scientific teams on their preferred platforms—Windows, Mac, or Linux—without having prestaged or replicated any imaging data. The entire process, from initial ingest to review, occurs in real time and at line speed. Snapshots and project-level isolation allow secure sharing with external collaborators while maintaining compliance.

SOLUTION BENEFITS

Qumulo's Cloud Data Fabric unifies CryoEM workflows from sample acquisition to 3D reconstruction:

Real-Time Data Availability: After being uploaded to any Qumulo endpoint, data from microscopes is immediately visible and accessible for downstream processing—whether on-premises or in the cloud. CDF eliminates the need to replicate data between sites, as well as the associated costs, delays, and complexity.

Scalable, High-Performance Storage: Designed for high performance and massive scale, Qumulo supports CryoEM's image datasets (10–200 TB per project), millions of small files, and GPU-heavy 3D modeling and analysis computations.

Faster Time-to-Result: Scientists and IT teams all benefit from CDF's simplified workflows, reduced time-to-result, and lower storage costs—without the need for data duplication or specialized transfer tools.

Platform-Agnostic Data Services: Cloud Native Qumulo extends seamlessly to AWS, Azure, Google Cloud, or OCI, enabling real-time data accessibility from on-premises to GPU resources in virtually any location for advanced AI/ML workloads. Unlike other storage solutions, CDF eliminates the need to stage Cryo-EM data into and out of object storage: reducing latency, lowering costs, and minimizing complexity.

ABOUT QUMULO

Qumulo is the leading provider of cloud file data services, providing real-time visibility, massive scale and API control of data across the data center, private, and public clouds. Organizations of all sizes and in all sectors trust Qumulo to solve their most complex and mission-critical data challenges quickly, efficiently and securely. Learn more at www.qumulo.com.