The Future of Distributed Data: An Introduction to Cloud Data Fabric

Architecture White Paper

March 2025

This document describes how Cloud Data Fabric (CDF) enables organizations to transcend geographical limitations and access their data from anywhere. CDF provides a consistent view of data across core, edge, and clouds. Its architecture supports many use cases, from high-performance computing and artificial intelligence to efficient data archival. Its design streamlines data management and reduces an organization's total cost of ownership.

Executive Summary

As unstructured data continues its explosive growth and global workforces expand, businesses must rethink their data strategies to eliminate silos, inefficiencies, and performance constraints. Organizations pursuing a data-driven approach must overcome obstacles such as lacking performance, operational inefficiencies, and the complexity of managing diverse storage systems. These factors hinder collaboration, reduce productivity, and create ripple effects that impact overall business performance.

Qumulo's Cloud Data Fabric redefines distributed data management by delivering a unified, globally consistent file system that seamlessly integrates on-premises, cloud, and edge environments. By eliminating data silos, it enables organizations to access and collaborate on data from any location without disruption. At the heart of this architecture, Cloud Data Fabric (CDF) uses portals to span directories across multiple Qumulo instances.

Each portal originates from a single source directory, called a "hub," on a single Qumulo instance. The directory can then be expressed as a "spoke" onto one or more additional Qumulo instances, which can be hosted on-premises, in the cloud, or at the edge. A portal can be configured as read-only or read-write and changed as needed, determining the access mode at the spoke endpoints[1]. The hub will always have write capability.

CDF's Global Namespace was developed with a design tenet of *strict data consistency*. The CAP Theorem states that when developing a distributed system, you can only have two out of the three following characteristics: consistency, availability, or partition tolerance. Qumulo chose consistency to eliminate the risk of data loss or corruption, differentiating Cloud Data Fabric from the other solutions on the market.

[1] A CDF endpoint is either a hub directory, or a directory mounted as a spoke in a portal.

Qumulo Cloud Data Fabric Benefits

This white paper provides a comprehensive overview of Qumulo's Cloud Data Fabric (CDF), which enables a single shared dataset to span on-premises, cloud, and edge locations, with all modifications and updates immediately visible to applications or users who access the portal from any of its endpoints.

Customers who use Cloud Data Fabric in their hybrid and multi-cloud operations will realize the following benefits:

- **Strict Consistency**: With CDF, strict data consistency is ensured, providing users with immediate access to the most current data, regardless of their location.
- **Accelerated Workflows**: Access and work with data across geographically dispersed locations with minimal latency, leading to faster project completion times.
- **Increased Collaboration:** CDF enables teams to create, access, and modify data across locations in real time.
- **Improved Agility**: Enables organizations to quickly adapt to changing business needs by providing a flexible and scalable solution for data access and collaboration.
- **Streamlined Management:** Qumulo simplifies the management of data-access permissions and data protection under a single management model that works in all locations.
- **Optimized Costs**: CDF reduces the cost associated with data transfer and storage by efficiently caching and managing only requested data across locations rather than wholesale dataset replication.

Table of Contents

Introducing Qumulo's Cloud Data Fabric

Consider the following scenarios:

- A medium-sized animation studio wants to move its burst rendering workflows to the cloud, but the required media asset files are stored in an on-premises storage cluster.
- A cancer research lab wants to leverage an AWS-based AI engine to identify tumors in mammogram images, but the dataset consists of 500 TB of on-premises files.
- An energy company has 20 PB of old seismic data that they want to move from on-premises storage to the cloud, but their on-premises analytics algorithm will continue to need intermittent access to data from the archive.

In each of these scenarios, data that is stored in one location is needed in another. The legacy solution would require that the organization initiate a wholesale replication of data from on-premises to the cloud – or from the cloud back to the data center – before any work can begin on the selected dataset.

Replication takes time: even with a 9Gbps connection between the data center and the cloud, the cancer research lab would have to wait a minimum of 123 hours for the entire 500TB dataset to be available to the AWS-based AI engine. Without knowing which specific files and folders from the 20 PB archive are required for on-premises use, the energy company might need a dedicated WAN link just for their analytics engine to pull the needed files from the archive datasets. As for the animation studio, if any changes are made to their base assets after replication begins, the cloud-based rendering engine will waste time and money rendering obsolete data. Large-scale replication between locations and platforms leads to complexity, cost, and waste.

CDF addresses the challenges of data consistency and real-time access through a combination of distributed logging, data locking, and the use of a coherent cache. Combined, these mechanisms ensure that even globally distributed users and applications always access the most recent version of their shared data, synchronizing updates in real-time across locations and eliminating the risk of stale or conflicting files.

Qumulo Cloud Data Platform

Qumulo customers have flexibility when choosing how they want to run their Qumulo instances. With Qumulo's software-defined "run anywhere" architecture allows them to use the vendor, platform, or cloud of their choice to run Qumulo instances on. Qumulo Nexus provides real-time data visibility, allowing organizations to streamline data management, optimize resource utilization, enhance agility, and seamlessly integrate with existing IT infrastructure.

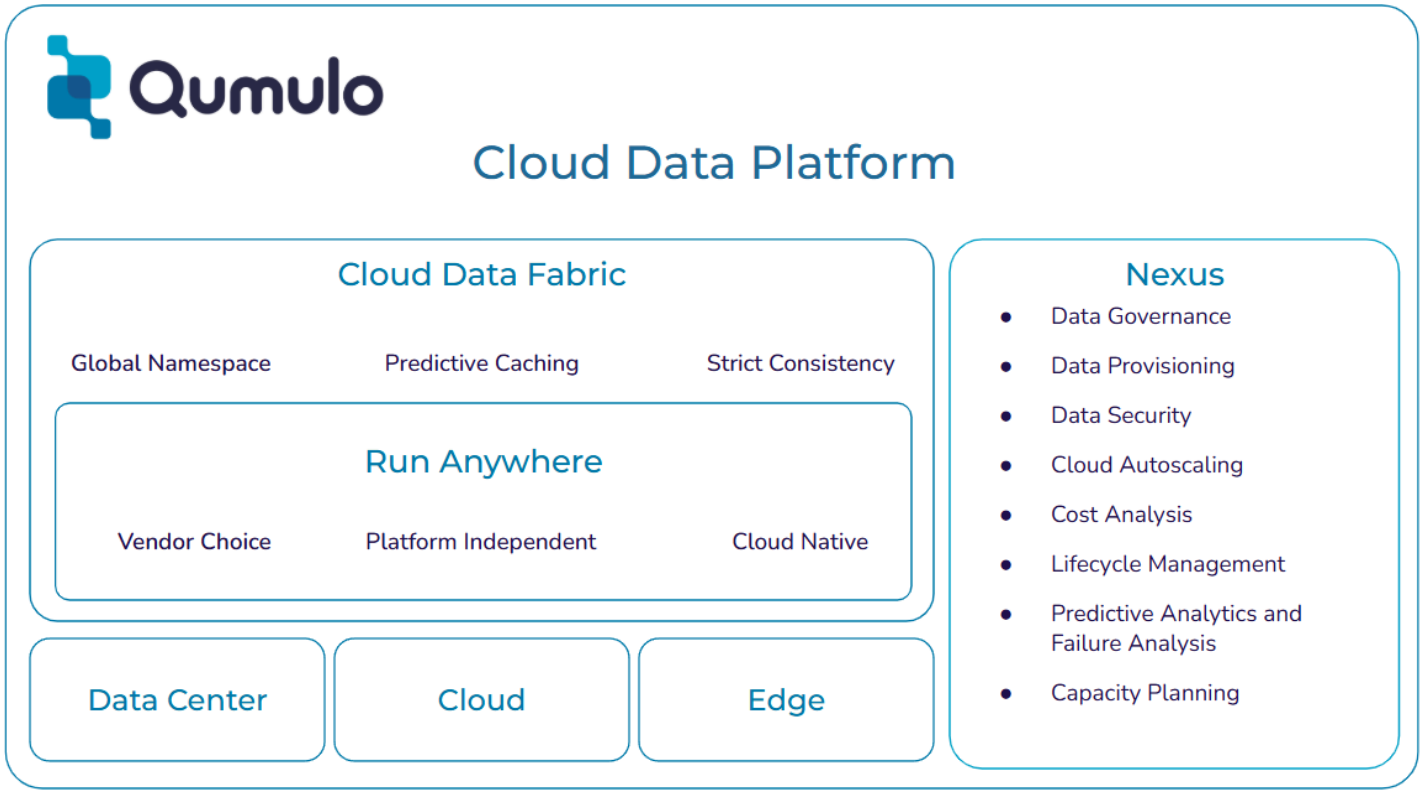

Figure 1: Cloud Data Platform

Cloud Data Fabric leverages "portals" to connect a single directory on the hub Qumulo instance to directories on one or more other Qumulo instances. In the context of a portal the directory on the other Qumulo instances is called the spoke. Files and sub-directories within the hub directory are

made consistently available on all spoke directories, and appear as if they are local to each of them respectively.

Unlike traditional solutions that require preloading or full data replication, CDF will fetch just the data segments needed to fulfill client read requests, along with data predicted to be read in the near future by the predictive cache engine. When the predictive cache detects a recognizable access pattern, it will proactively and aggressively cache data from the hub to optimize performance on the spoke. By aggressively caching data in anticipation of being read, CDF substantially improves the read performance on the spoke clusters.

For writes, CDF includes a write-back cache mechanism that logs any write on a hub or spoke endpoint in a manner that ensures data consistency, even in the event of a network outage that isolates one of the endpoints.

Qumulo Cloud Data Fabric

- **Global Namespace:** CDF's global namespace provides a consistent single view of data across disparate locations, ensuring that users at both ends of the portal see the same files and content in real time, whether reading, writing, or modifying data.
- **Predictive Cache:** Predictive Cache is a key feature of Qumulo core that anticipates which data blocks are most likely to be requested next based on several inputs, including historical patterns, in order to minimize read latency.
- **Strict Consistency**: CDF guarantees that all reads reflect the most recent write, preventing stale or conflicting data across distributed locations, even during network disruptions.

Run Anywhere

- **Vendor Choice:** Deploy Qumulo in the data center using a wide range of industry-standard hardware from virtually any hardware vendor, with your choice of Intel or AMD architectures. Run Anywhere gives you the freedom to choose the best platform for your needs which meets your specific enterprise standards.
- **Platform Independent:** Run Qumulo on your preferred platform – bare metal, virtualized environments, or containers – and integrate it with your existing IT infrastructure across various architectures.
- **Public Cloud**: You can provision a Qumulo instance in several public clouds, including Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP), and Oracle

Cloud Infrastructure (OCI). Qumulo cloud instances are built using native cloud primitives, and can quickly scale out, or in, to meet workload performance demand.

Deployment Options

To support a variety of use cases, Qumulo supports multiple deployment scenarios, letting customers choose which specific deployment and operational models align with their operational and strategic requirements, as shown in Figure 2 below.

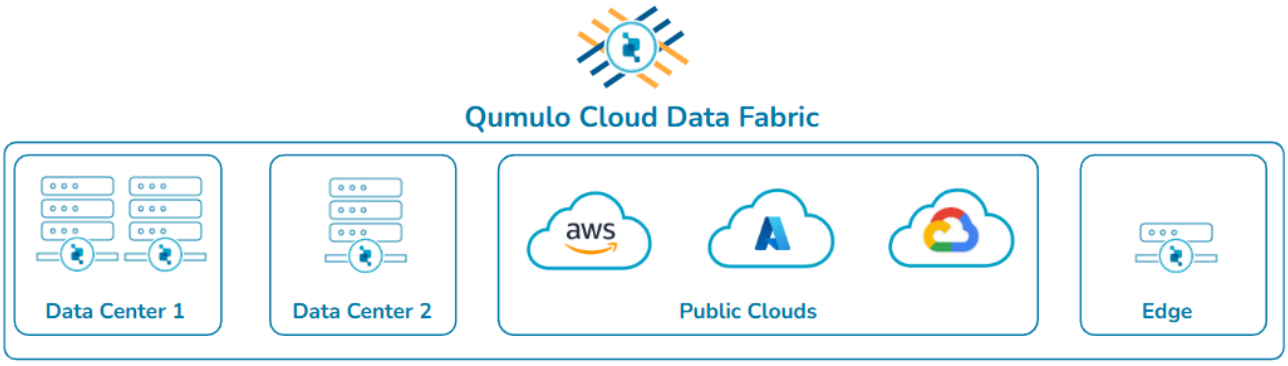

Figure 2: Qumulo Deployment Versatility

On-Premises Deployment

- A Qumulo instance deployed in the customer's own data center (or private cloud) using customer-owned and managed commodity hardware appliances from the customer's preferred hardware vendor.
- Customers can increase both capacity and performance by adding more hardware nodes to each on-premises deployment.

Cloud Native Qumulo

Cloud Native Qumulo (CNQ) is a customer-managed solution deployable in AWS, Azure, GCP or OCI, using customer-managed cloud resources to provide maximum control over configuration and cost management. CNQ's flexible model allows tailored deployments and non-disruptive scaling of both capacity and throughput to meet specific performance and business needs.

- Deployable in customer-managed cloud accounts.
- Leverages cloud compute, network, and storage resources natively to enable full customer control over configuration and cost management.
- Flexible deployment model lets organizations customize their CNQ deployments to their specific business and performance needs.
- Non-disruptive elastic scaling capabilities for both performance and capacity.

Azure Native Qumulo

Azure Native Qumulo (ANQ) is a fully-managed Qumulo service available exclusively in Microsoft Azure, delivering seamless scalability and high performance without requiring customers to manage the underlying infrastructure. While Qumulo and Microsoft handle service management, customers retain full control over their data, configurations, and features. ANQ can be deployed effortlessly via the Azure Portal, API, or CLI, making it an ideal solution for organizations seeking a hands-off yet flexible cloud file system.

Edge Deployment

- Compact and versatile, supporting a single-node virtual machine or bare metal hardware form-factor.
- Can act as a read-only or read-write cached edge as part of a Qumulo Cloud Data Fabric portal.
- Ideal for remote or space-constrained environments with smaller-scale performance and access needs.

Cloud Data Fabric Use Cases

Cloud Data Fabric (CDF) empowers organizations to reimagine their workflows, introducing innovative capabilities that can improve existing workflows or address previously insurmountable challenges, ultimately driving greater efficiency, agility, and innovation. As illustrated in Figure 3 below, any multi-site organization can benefit from CDF's unified data experience across on-premises, edge, and cloud environments.

CDF facilitates a wide range of applications, including (but not limited to): finance, media production, analytics, and AI. The following scenarios demonstrate how customers can leverage CDF to drive efficiency and innovation across different sectors.

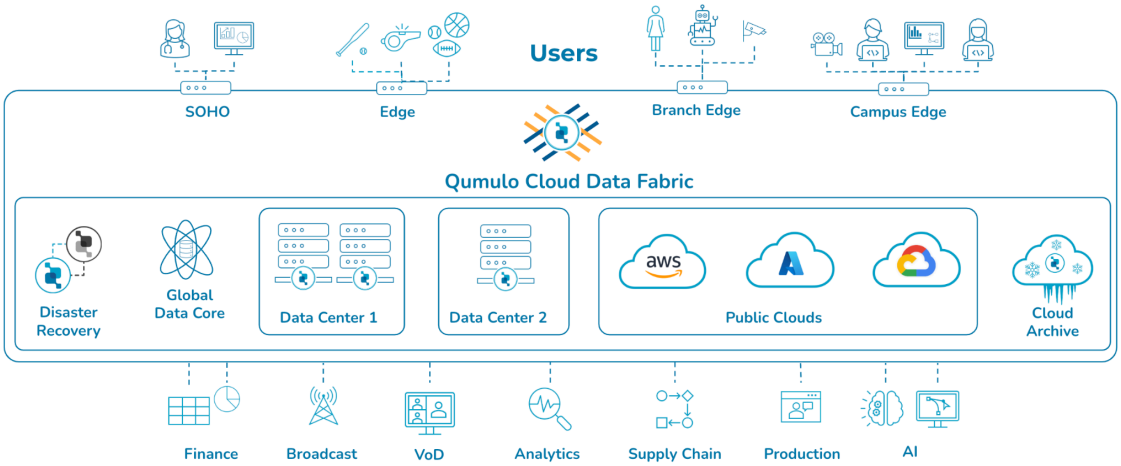

Figure 3: Cloud Data Fabric Use Cases

Use Case: Seamless Hybrid Cloud Rendering for Media & Entertainment

VFX studios and animation houses face escalating pressure to deliver high-resolution content under tight deadlines that often exceed the ability of their on-premises render farms, resulting in long queues, time crunches, and potential deadline misses. While cloud rendering offers flexibility, traditional hybrid solutions based on data replication suffer from slow transfer times, hindering instant burst capacity utilization, and can lead to wasted time and money when stale assets result in a worthless render.

Qumulo's Cloud Data Fabric (CDF) addresses this by expressing directories from the on-premises clusters to CNQ cloud instances. This approach is simple, seamless, secure, and efficient; especially when compared to full dataset replication. CDF eliminates the risk of wasted compute time due to de-synchronized assets. As rendering jobs progress, studios can immediately see results that have been saved in the portal directory from any endpoint throughout the portal, including the on-premises hub.

CDF transforms cloud rendering by removing both barriers and bottlenecks, providing the speed, flexibility, and efficiency required to meet deadlines and scale effortlessly. CDF accelerates production timelines, lowers costs, and increases agility.

Independent Scaling of Spoke Clusters

Spoke clusters operate as autonomous file systems, allowing their performance to scale independently from hub clusters.

As shown in Figure 4 below, a media & entertainment (M&E) burst rendering workflow benefits from a spoke cluster that can deliver many times the IOPS and throughput of the hub, supporting workloads with 100,000+ CPU cores on the spoke cluster. By extending a single directory from a moderately sized on-premises cluster to a high-performance spoke cluster, customers can unlock virtually unlimited compute power without being constrained by on-site compute, storage or network limitations.

With this approach, studios can accelerate rendering, shorten project timelines, and reduce costs by avoiding overinvestment in local infrastructure.

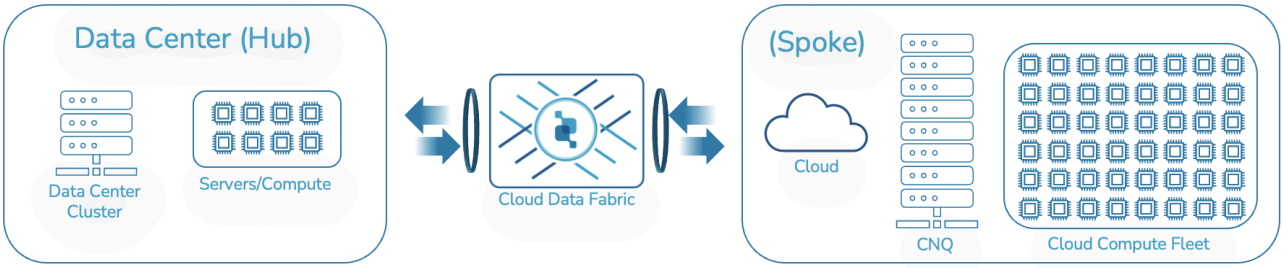

Figure 4: Example: Cloud Bursting to a High-Performance Spoke Cluster

Use Case: Collaborative Video Editorial for Media & Entertainment

Video editing teams working with terabytes of 4K footage across multiple studios face challenges beyond basic file transfer, even with high-bandwidth connections. Studio pipeline and IT teams are often charged with orchestrating data movement via manual scripts and complex workflows, leaving less time for core administrative tasks. Replication also leads to unnecessary duplication, straining storage and requiring meticulous management of very large datasets across multiple sites. Precise dependency tracking is critical, as imperfections cause missing assets and workflow disruptions. Data replication incurs costs in human capital and specialized solutions.

Cloud Data Fabric addresses these inefficiencies by connecting disparate locations into a single unified namespace, enabling real-time access and editing of high-resolution content from anywhere. Instead of file transfers and version control struggles, editors can rely on CDF's

pre-fetching algorithm to keep the cache full of useful data, slashing latency and eliminating frustrating waits. This results in faster editorial workflows, enhanced real-time collaboration, and increased productivity, allowing teams to focus on creative work rather than logistical hurdles.

Use Case: Accelerating Genomic Analysis with AI for Life Sciences

Life sciences organizations are increasingly leveraging the power of AI to analyze genomic data, driving breakthroughs in drug discovery and personalized medicine. This AI-driven analysis often involves petabyte-scale datasets and computationally demanding workloads. Large genomic datasets can be difficult to move to the cloud, where powerful AI tools and scalable compute resources reside. AI applications, particularly those utilizing GPUs, require low-latency access to data to maximize performance and avoid costly GPU idle time. This necessitates a solution that can bridge the gap between on-premises data and cloud-based AI.

Qumulo's Cloud Data Fabric (CDF) provides a unified data experience across on-premises, edge, and cloud environments. It allows the whole human genome files existing on their on-premises storage to be visible and accessible in their cloud of choice, without the wait or complexity associated with large-scale data migrations.

Organizations that use CDF can leverage cloud resources for compute while keeping their data on-premises and maintaining data sovereignty. CDF streamlines data pipelines, allowing researchers to focus on analysis and discovery rather than data logistics. CDF empowers life sciences organizations to accelerate genomic analysis, leading to faster breakthroughs and improved patient outcomes.

Use Case: Real-Time Reservoir Simulation for Oil and Gas (Energy)

Reservoir engineers in the oil and gas industry utilize computational models to simulate fluid movement within subsurface formations, integrating diverse data sources like seismic imaging, well logs, and production data for optimized well placement and hydrocarbon recovery. These data-intensive simulations, involving large-scale, multi-disciplinary datasets along with high-performance computing environments, present significant data challenges. These challenges include managing overwhelming data volumes, ensuring efficient data integration across disciplines, maintaining real-time data synchronization for frequent model updates, and handling version control across global teams.

Cloud Data Fabric’s unified namespace addresses these challenges by consolidating all reservoir-related datasets into a single, accessible storage platform, eliminating silos and enabling

seamless collaboration. Its high-throughput capabilities support parallel processing of simulation data, ensuring efficient computational model execution. Real-time data synchronization and versioning allow engineers to track changes and maintain consistency across multiple simulation iterations. Additionally, with hybrid cloud integration, teams can dynamically scale storage based on computational demands to support the large-scale needs of reservoir modeling and simulation.

Use Case: Seismic Imaging & Exploration for Oil and Gas (Energy)

Seismic imaging, a critical process in oil and gas exploration, involves geophysicists utilizing 3D and 4D seismic surveys to analyze subsurface structures and locate hydrocarbon reserves. These surveys require high-performance computing (HPC) and extensive storage resources to generate petabyte-scale datasets.

Managing and storing seismic data is highly complex due to its sheer volume and the need for rapid processing. Real-time interpretation necessitates instant access to large datasets, which traditional storage solutions struggle to handle efficiently. Exploration teams often work in remote locations, making data transfer to central processing sites expensive and bandwidth-intensive. Data consistency across different survey methods and historical archives further complicates integration and accessibility.

Cloud Data Fabric provides a scalable, high-performance file storage system capable of managing massive seismic datasets efficiently, while Qumulo's real-time analytics allow exploration teams to track and manage storage usage dynamically. By supporting multi-cloud and hybrid storage, Qumulo ensures that seismic data can be processed both on-premises and in the cloud, reducing dependency on local infrastructure. The ability to replicate and tier data seamlessly between exploration sites and central repositories optimizes bandwidth usage, ensuring that large seismic datasets are available where they are needed without unnecessary delays.

Cloud Data Fabric - Architectural Example

Figure 5 below depicts a potential Cloud Data Fabric (CDF) portal topology.

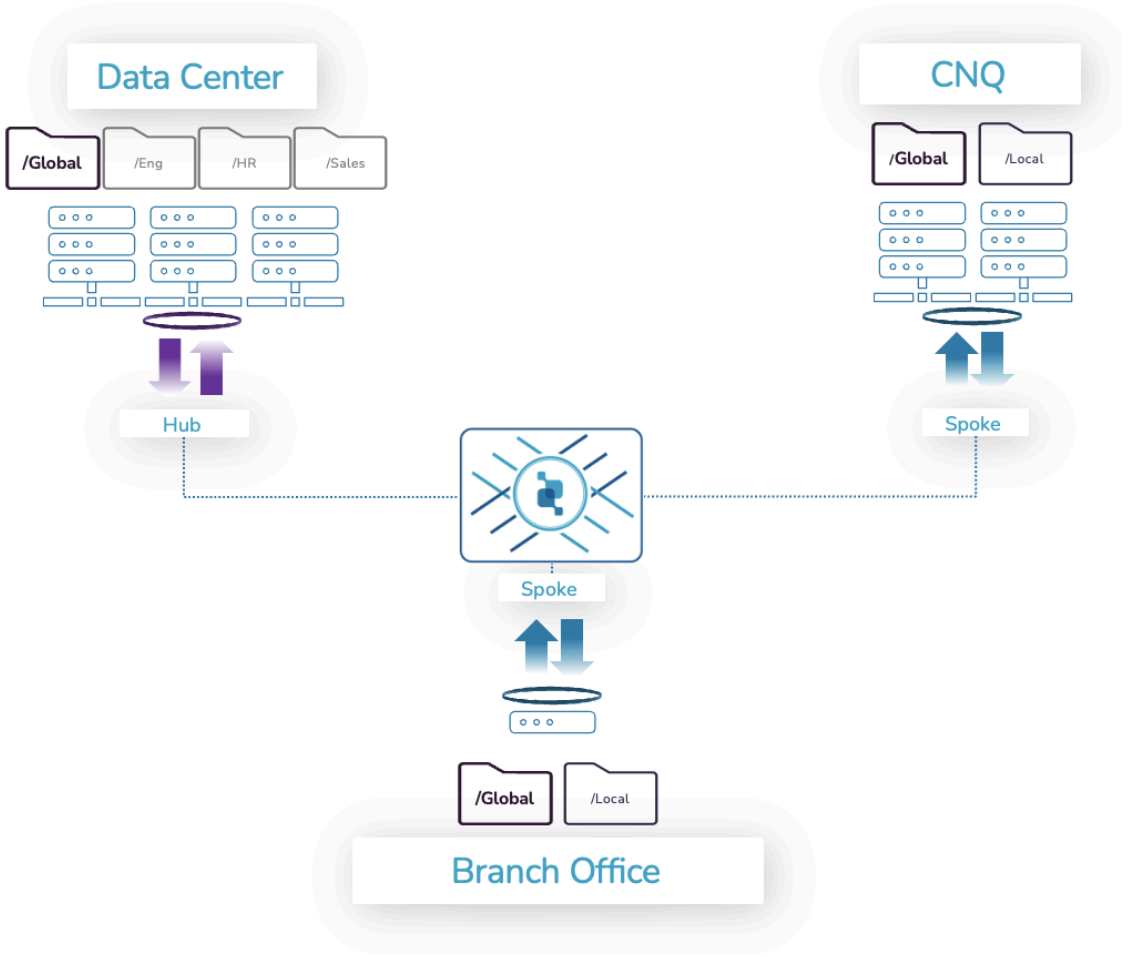

Figure 5: Cloud Data Fabric Global Design

On-Premises

In the top left corner of Figure 5 is a Qumulo instance deployed on-premises within a corporate data center. Besides its own site-specific shares such as Engineering, Human Resources, and Sales, this cluster also hosts the /Global share which is made available to the CNQ and branch office / edge endpoints via a CDF portal. Users and applications can access the contents of the "/Global" folder via their own local Qumulo instance, whether from the corporate data center, AWS, Azure, or a branch office.

Cloud Native Qumulo

The upper right corner of the diagram shows a Cloud Native Qumulo (CNQ) cluster, which can be deployed in AWS, Azure, and other cloud environments. Azure Native Qumulo (ANQ), a managed service available to Azure customers, can serve as a CDF endpoint as well. This illustrates CDF's

ability to extend the data fabrics across multiple cloud providers, offering flexibility and choice. CNQ allows for rapid deployment and scalability within these cloud environments.

Branch Office

At the bottom of the diagram, "Branch Office" represents edge locations: remote offices or facilities that need to enable data access across distributed geographies. These locations often have limited bandwidth and rely on locally stored datasets to ensure smooth operations.

Distributed Data Access

CDF uses a hub-and-spoke model for each portal. In this structure, a central 'hub' folder *(*which can be located on-premises, in the cloud, or at the edge) acts as the central authority for all portal endpoints. 'Spoke' folders, hosted on remote endpoints in multiple sites, connect to the hub and provide local access points to the directory shared through the portal.

For instance, as shown in Figure 5, the on-premises data center cluster serves as the hub and hosts the "/Global" share. The CNQ clusters and edge deployments are spokes, each with the same consistent access to the "/Global" share. This configuration allows users and applications at each spoke location to interact locally with the data, and ensures that all data remains consistent.

First Remote Read

When a file is accessed in a spoke directory for the first time, the requested blocks are transferred to the spoke endpoint's cache, and potentially prefetch additional data to minimize latency for future requests. Once cached, the data is immediately available to any client of the spoke endpoint. Over time, CDF dynamically adapts to access patterns and automatically evicts less frequently used data from the cache.

Inside the Architecture

Qumulo's Cloud Data Fabric (CDF) is more than just a global namespace. It's built on a foundation of advanced features that work together to deliver a smooth and high-performance experience to end-users and applications.

Figure 6 (below) illustrates the layered architecture of the Qumulo core file system: the front-end protocols, file system, locking, transaction system, cluster quorum, and protection layers. Unlike an external add-on or overlay, CDF is integrated directly into these layers, forming the backbone of data consistency, synchronization, and fault tolerance across the fabric. This tight integration ensures that every file operation – whether a read, write, or modification – is accurately tracked and seamlessly transferred, reinforcing CDF as a foundational component of Qumulo's architecture.

The coherent caching system keeps data current, no matter where it's accessed from. A distributed lock manager ensures that only a single portal endpoint is authorized to make changes to a file in a given time frame. Lastly, every IO is appended to a logical log in a manner that ensures consistency across the portal. Additionally, CDF includes predictive caching which leverages a dynamic heat scoring mechanism based on a multitude of inputs to keep high-value data in cache, and evict the less valuable data.

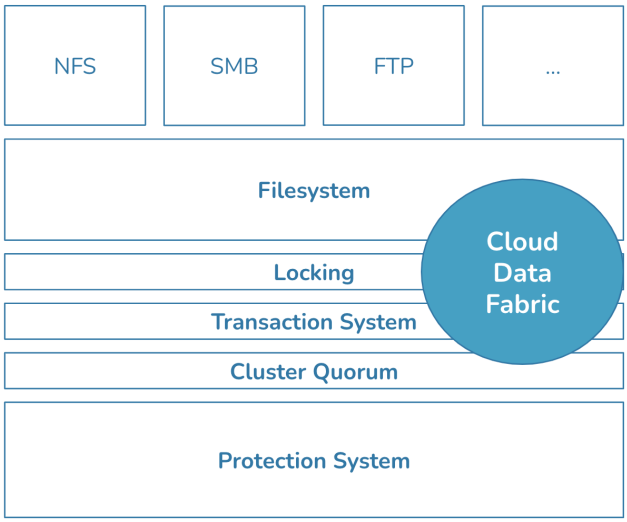

Figure 6: GNS is the File System

Coherent Cache

Let's explore the components of a Cloud Data Fabric portal to see how they all work together to deliver a high performance coherent cache. The three core components are CacheFS, Logical Logging, and Distributed Locking.

CacheFS: Cloud Data Fabric's CacheFS functions as a sparse, partial file system backing each spoke directory, acting as a local cache for data stored on the hub. This allows for rapid access to frequently used data while minimizing latency. Because the spoke acts as a cache, there needs to be a mechanism that maintains consistency with the hub.

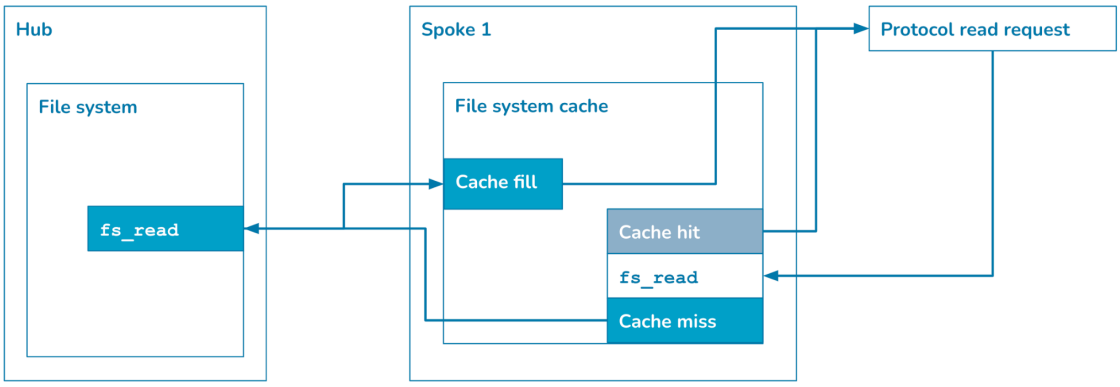

Figure 7: CDF CacheFS Operations Path

As shown in Figure 7 above, when a client reads a file over a protocol (SMB, NFS, FTP, S3, etc.) the system first checks if the data is available in the local cache. If it is (cache hit), the data is returned immediately. If not (cache miss), the system retrieves the data from the hub. The spoke is then updated with this fresh content, and the consistent data is returned to the client.

CacheFS is a separate instance of the Qumulo file system, independently addressable. This allows for flexible caching capabilities, enabling the system to cache anything from portions of a file to entire directories. This is particularly useful when working with extremely large files or directories containing hundreds of millions of entries, as it allows only the necessary segments to be transferred. Being a separate filesystem allows us to cache every attribute of the file or directory exactly. From a client's perspective, data in the cache in a spoke directory is indistinguishable from the data on the hub. File ids, permissions, ctime, mtime, everything is the same.

Because CacheFS operates independently of the primary file system, a cluster can instantly discard its cache when a portal is administratively disconnected from CDF. This ensures that cached data is no longer accessible on the spoke, an essential capability for maintaining security and compliance.

The CacheFS is necessary for portal operation but not sufficient. We need a mechanism to invalidate and update the remote cache when writes happen in either the hub or spoke directories. That mechanism is a logical log.

Logical Time and Logical Logging: Cloud Data Fabric (CDF) uses a distributed, sharded log that tracks all changes to data at both the hub and the spokes. This log is essential for cache invalidation and updates. The log must guarantee the correct ordering of events and also maintain high performance. Qumulo has developed a new logical time system that spans the entire distributed system, ensuring proper ordering of log entries even across multiple clusters. This enables all participants in the system to generate log events independently from each other, and still have the events be ordered in the correct order, without bottlenecks on locking or any other resources.

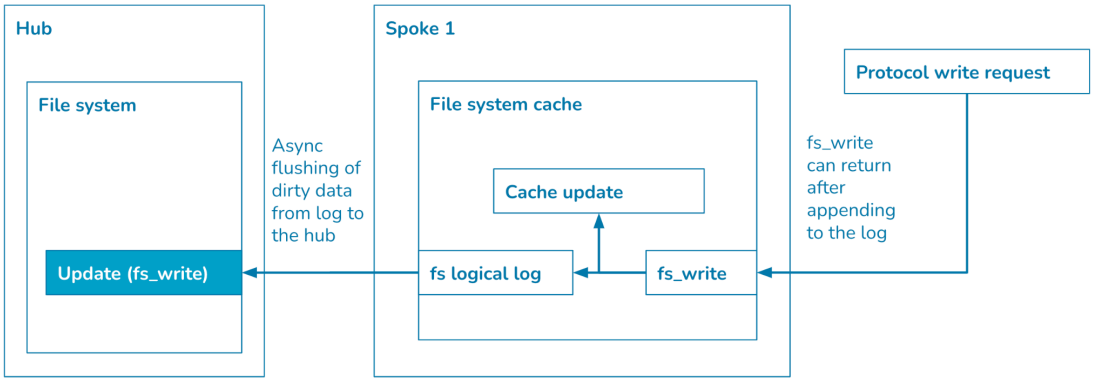

Figure 8: CDFs Logging Execution Sequence

As shown in Figure 8, a new protocol write request calls into fs_write which runs a transaction that both updates the spoke cache and appends to the logical log. The fact that the log entry and data in the filesystem is written transactionally is critical for correctness. If either is written without the other, you have data inconsistency. Once the data is safely written to the log, we can immediately acknowledge the write to the client. This write-back cache with fast acknowledgements to clients is critical for write performance. The common alternative of write-through caching requires a WAN round trip for every write, severely limiting performance and increasing write latency seen by the clients. The log is then asynchronously flushed across the portal, ensuring durability for any future read, anywhere across the fabric.

CacheFS and logging alone do not guarantee strict consistency. We need a mechanism to ensure logs are applied before a client is allowed to access a file to ensure consistency. We do that with distributed locking.

Distributed Locking: Figure 9 below shows the entire flow for a read from a spoke directory. The same diagram could just as easily describe a read from the hub – the architecture is symmetrical.

When a read request comes in, it first takes a lock from the cross cluster distributed lock manager. That lock manager will force the relevant logs to be flushed before granting the lock, ensuring data consistency.

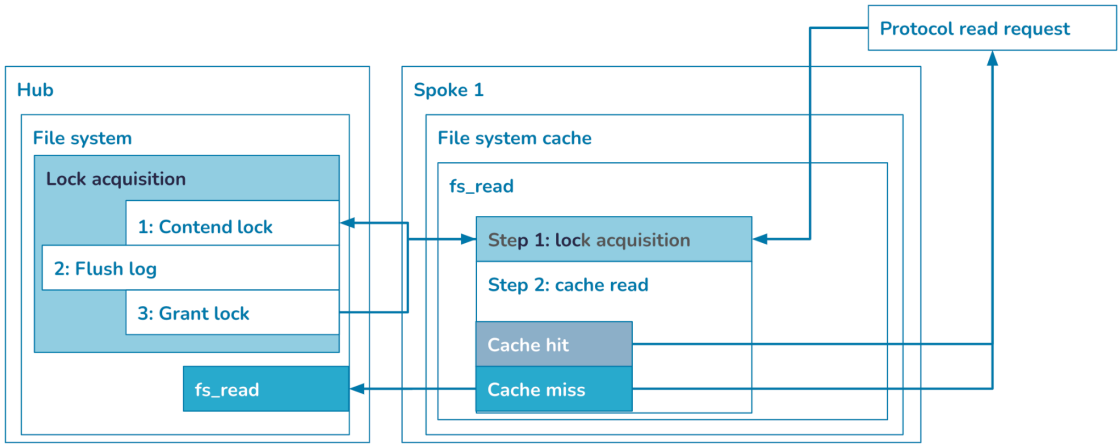

Figure 9: Cloud Data Fabric Locking Workflow

Cross-cluster locking coordinates data modifications across clusters, preventing conflicts. This system has two key features that contribute to its efficiency:

Lock caching: When a cluster acquires a lock on a file or directory, it retains that lock and the lock is cached. If there are subsequent lock requests for that file or directory, the cluster does not need to re-request the lock. This caching mechanism reduces latency and improves performance by avoiding unnecessary network round trips.

Log flushing before lock granting: Before granting a lock, the system ensures that all outstanding log entries related to the data being locked have been processed and flushed. This guarantees that when a cluster acquires a lock, it has the most up-to-date and consistent view of the data. For example, if a spoke cluster holds a write lock on a file and has made local modifications (recorded in its log), any other cluster requesting a read lock on that file will first trigger a replay of the spoke's log entries. This ensures that the reading cluster receives the latest data, and any stale cached data is invalidated.

Strict Data Consistency

CDF is designed to maintain consistent data access and prevent data corruption across distributed clusters, even during network interruptions or other disruptions. The system's built-in log replay mechanism ensures a smooth reconnection process for disconnected clusters, by applying any missed entries in the logs.

As mentioned above, CDF logs write operations, and transmits these log entries across the network. This means that when a disconnected cluster reconnects, the system can simply resume replaying the log from where it left off, ensuring that any changes missed during the disconnection are properly applied. This optimized change replay minimizes data transfer and avoids having to re-scan the cache for stale data, leading to rapid recovery and efficient resource utilization.

The lock manager ensures that all write operations are properly coordinated across the entire fabric, preventing conflicts and ensuring that all clients see a consistent view of the data at all times. By combining distributed logging with distributed locking, Qumulo delivers a strictly consistent file system that provides both high performance and data integrity.

Predictive Prefetch

Cloud Native Qumulo's Predictive Prefetch mechanism analyzes client data access patterns in real-time, using advanced algorithms and real-world usage data to anticipate future requests. The prefetcher intelligently guides the Predictive Cache, ensuring that the most relevant data is proactively stored in cache, optimizing client performance and reducing access latency.

While Predictive Prefetching is a key feature in both Cloud Native Qumulo (CNQ) and Cloud Data Fabric (CDF), each uses different caching strategies optimized for its unique architecture and storage environment. These purpose-built approaches ensure efficient data access and high performance, whether in a cloud-native deployment or a globally distributed file system. In CDF, Predictive Prefetch works with the CacheFS to proactively store entire files or specific data ranges within a file before they are requested. Driven by client access patterns at the edge, this approach optimizes performance and minimizes latency for seamless data access.

To learn more about Predictive Prefetch, see the Qumulo CNQ Architecture White Paper.

Conclusion

Qumulo's Cloud Data Fabric (CDF) enables organizations to leverage data as a strategic asset, driving business growth and outperforming competitors. By unifying data across various locations into a single, globally coherent file system, it overcomes the limitations of traditional storage architectures. This innovative approach delivers the performance, scalability, and efficiency required by modern enterprises.

With an unwavering commitment to data consistency, Qumulo ensures that users and applications always access the most current data. This fosters collaboration and innovation, regardless of location. CDF provides a future-proof foundation that allows organizations to harness their growing data, drive innovation, and maintain a competitive edge.

More than just a technology, Qumulo's Cloud Data Fabric acts as a catalyst for transformation, unlocking opportunities to maximize the value of data. This groundbreaking capability is revolutionizing how businesses access and interact with their data, accelerating innovation and operational agility. For the first time, organizations can access files within a file system, using familiar file system semantics, from any location on the planet and beyond. This empowers them to respond swiftly to changing demands with a flexible, scalable solution that adapts to evolving data needs.

Contributors

This article is maintained by Qumulo. The following contributors originally wrote it.

Kevin McDonald (KMac) |Principal Technical Marketing Engineer at Qumulo

Related Resources

Qumulo Cloud Data Fabric

Qumulo Cloud Data Fabric Solutions Brief

Qumulo CNQ Architecture White Paper

Qumulo Nexus