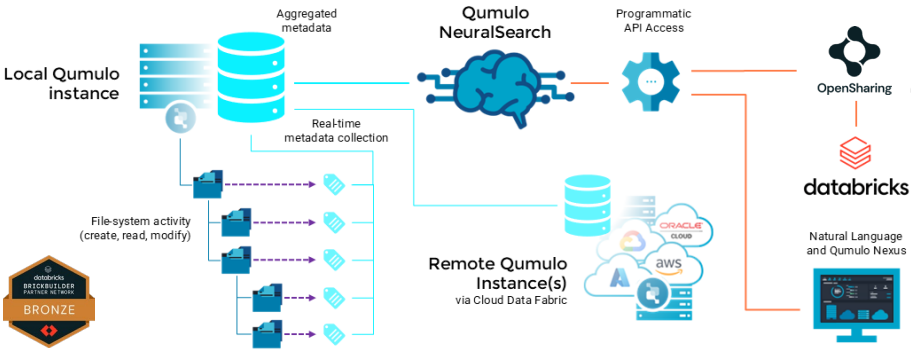
OpenSharing with Databricks: A Smarter Approach to Unstructured Data for AI

ANY DATA
ANY LOCATION
TOTAL CONTROL

EXECUTIVE SUMMARY

Even as organizations find ways to accelerate AI, analytics, and modernization initiatives, most enterprise data remains unstructured—stored across file systems, object stores, edge locations, data centers, and multiple public clouds. Traditional approaches require costly replication, migrations, and duplicate copies of data to make it usable for analytics, creating governance risks, version conflicts, and rising cloud costs.

Qumulo and Databricks solve this challenge by combining OpenSharing, NeuralSearch, and Cloud Data Fabric to provide secure, governed access to unstructured data, without copying or moving it. This enables organizations to access file and object data directly where it resides across core, cloud, and edge environments while maintaining a single source of truth, reducing operational complexity, and accelerating AI and analytics outcomes.



KEY BENEFITS

With Qumulo and Databricks, you can:

- Access data directly where it lives without replication or data movement.
- Reduce costs by eliminating duplicate storage, sync issues, and cloud egress fees.
- Speed AI and analytics by enabling faster access to raw files and training datasets.
- Strengthen governance with centralized access control, auditing, and secure permissions.
- Support multi-cloud operations across AWS, Azure, Google Cloud, and on-premises environments.
- Accelerate time-to-insight by speeding data access and reducing API costs.

THE CHALLENGE

Modern enterprises face significant barriers when trying to operationalize unstructured data for AI and analytics:



Data across files, objects, on-prem, and multiple clouds

Costly replication and migration for analytics

Duplicate copies causing governance and version issues

High cloud egress fees and rising storage costs

Slow discovery from crawlers, ETL, and manual indexing

Limited visibility into dark data across environments

Traditional analytics models were built for structured tables, not for massive volumes of unstructured data such as images, video, genomics, logs, media assets, and AI training datasets.

THE SOLUTION: OpenSharing

Qumulo extends OpenSharing to unstructured data by combining:

Qumulo Cloud Data Fabric
A globally distributed data platform that enables secure access to file and object data across on-premises, cloud, edge, and multi-cloud environments while maintaining strict consistency and a single source of truth.

NeuralSearch
Qumulo's storage-native search platform that turns file system metadata into real-time, searchable intelligence, without crawlers, ETL pipelines, or separate databases.

It supports:

- SQL metadata search
- Natural language queries
- Semantic search
- Constant-time discovery across globally distributed environments

OpenSharing
Databricks' open-source protocol for secure, real-time data sharing across platforms, clouds, and organizations using HTTPS access, temporary credentials, and pre-signed URLs.

Together, these technologies allow teams to securely access data in Databricks directly from where it already resides, without replication, migration, or duplicate copies.

IDEAL USE CASES

AI/ML Training Pipelines
Access raw video, genomics, telemetry, and training datasets instantly without waiting for IT-driven migrations.

Healthcare & Life Sciences
Analyze NFS-based medical images directly without conversion—improving fidelity while reducing time and cost.

Media & Entertainment
Provide governed access to large unstructured media assets across distributed teams and cloud environments.

Enterprise Analytics
Unlock dark data and enable real-time discovery across hybrid and multi-cloud infrastructures.

WHY QUMULO

Native integrations with cloud object storage often require:

- Data movement before AI workflows, like those in Databricks, can begin
- Separate ETL pipelines and crawlers to locate data
- Additional indexing databases and infrastructure
- Limited support for file-based workflows

Qumulo removes these limitations by providing:

- Direct secure projection into Databricks
- Storage-native search without external indexing
- Enterprise-grade support for both file and object data
- Strict global consistency across distributed environments
- A single governed platform across edge, core, and cloud

WHY THIS MATTERS

AI success depends on unified data access, not just GPU/compute access.

Most organizations already have the data they need, but it remains trapped across unstructured data in fragmented storage environments. Qumulo and Databricks remove the barriers created by replication, migration, and siloed infrastructure by enabling governed, real-time access to unstructured data wherever it lives.

The result is faster innovation, lower cost, stronger governance, and a simpler path to modern AI and analytics.

ABOUT QUMULO

Qumulo is the leading provider of cloud file data platforms, offering unrivaled performance, scale, and data management solutions. Qumulo's platform is trusted by Fortune 500 companies and global enterprises to manage petabytes of data, unlocking its value and driving innovation. For more information, visit www.qumulo.com

©2026 Qumulo. All rights reserved. All other brands, products, or service names are or may be trademarks or service marks of their respective owners.