

Share data products, including AI models, unstructured data, dashboards and notebooks, with greater flexibility. Data providers can choose between sharing an entire table or sharing only a version or specific partitions of a table. However, sharing just tabular data isn't enough to meet today's consumer demands. Databricks Delta Sharing also supports sharing of nontabular data and data derivatives such as data streams, AI models, SQL views, volumes and arbitrary files, enabling increased collaboration and innovation. Volume sharing allows providers to securely share large-scale unstructured data such as images, videos and logs stored in cloud volumes without replication. Data providers can build, package and distribute data products, including datasets, volumes, AI models and notebooks, allowing data recipients to get insights faster. Furthermore, this approach promotes and empowers the exchange of knowledge — not just data — between different organizations. With Databricks Delta Sharing, we're able to achieve both a truly open marketplace and a truly open ecosystem. In contrast, commercial products are mostly limited to sharing raw tabular data and can't be used to share these higher-valued data derivatives.



Share data at a lower cost. Delta Sharing lowers the cost of managing and consuming shares for both data providers and recipients. Providers can share data from their cloud object store without replicating, thereby reducing the cost of storage. Additionally, by integrating with Cloudflare's Bandwidth Alliance and R2 zero-egress fees object storage, Delta Sharing further minimizes costs by eliminating or significantly reducing egress fees — charges incurred when transferring data out of a cloud provider's network.

In contrast, existing data sharing platforms require data providers to first move their data into their platform or store data in proprietary formats in their managed storage, which often costs more and results in data duplication. With Delta Sharing, data providers don't need to set up separate computing environments to share data. Consumers can access shared data directly using their tools of choice without setting up specific consumption ecosystems, thereby reducing costs.



Reduced time to value. Delta Sharing eliminates the need to set up a new ingestion process to consume data. Data recipients can directly access the fresh data and query it using tools of their choice. Recipients can also enrich data with datasets from popular data providers. The Delta Sharing ecosystem of open source and commercial partners is growing every day.

Maximizing the value of data and AI with Delta Sharing

Delta Sharing is already transforming data and AI sharing activities for companies in a wide range of industries.

Given the sheer variety of data available and the technologies that are emerging, it's hard to anticipate all the possible use cases Delta Sharing can address. The Delta Sharing approach is to share any data anytime with anyone easily and securely. In this section we'll explore the building blocks of such an approach and the use cases emerging from these.

"We're excited to continue working with Databricks to enhance our data distribution capabilities through their Delta Sharing platform, which provides an open ecosystem that makes a number of S&P Global content sets more seamlessly accessible and available to our clients. This expands our collaboration with Databricks, which began with S&P Global Capital IQ Workbench, leveraging their technology to create a collaborative analytics notebook environment for our users."

— **David Coluccio**, Head of Distribution Solutions, S&P Global Market Intelligence

"Most data platforms offer closed sharing solutions that restrict our ability to reach all of our customers. We prefer to invest in open solutions which enable us to share data with all of our customers and partners, not only across clouds, but also across platforms."

— **Derek Slager**, CTO and Co-founder, Amperity

"AI21 Labs is pleased that Jamba 1.5 Mini is now on Databricks Marketplace. With Delta Sharing, enterprises can access our Mamba-Transformer architecture, featuring a 256K context window, ensuring exceptional speed and quality for transformative AI solutions."

— **Pankaj Dugar**, SVP and GM, AI21 Labs

"We use reinforcement learning (RL) models in some of our products.

Compared to supervised learning models, RL models have longer training times and many sources of randomness in the training process. These RL models need to be deployed in three workspaces in separate AWS regions. With model sharing we can have one RL model available in multiple workspaces without having to retrain it again or without any cumbersome manual steps to move the model."

— **Mihir Mavalankar**, Machine Learning Engineer, Ripple

"Delta Sharing makes it easy to securely share data with business units and subsidiaries without copying or replication. It enables us to share data without the recipient having an identity in our workspace."

— **Robert Hamlet**, Lead Data Engineer, Cox Automotive

"When we want to launch and grow a product with a partner, such as a point-of-sale consumer loan, the owner of the data needs to send massive datasets on tens of thousands of customers. Before, in the traditional data warehouse approach, this would typically take one to two months to ingest new data sources, as the schema of the sent data would need to be changed in order for our systems to read it. But now we point Databricks at it and it's just two days to value. We used Delta Sharing and we had tables of data showing up in our Databricks workspace in under 10 minutes."

— **Barb MacLean**, SVP and Head of Technology Operations and Implementation, Coastal Community Bank

Internal sharing across business units with Delta Sharing

Internal data sharing is becoming an increasingly important consideration for any modern organization, particularly where data describing the same concepts have been produced in different ways and in different data silos across the organization. So it's important to design systems and platforms that allow governed and intentional federation of data and processes, and at the same time allow easy and seamless integration of said data and processes.

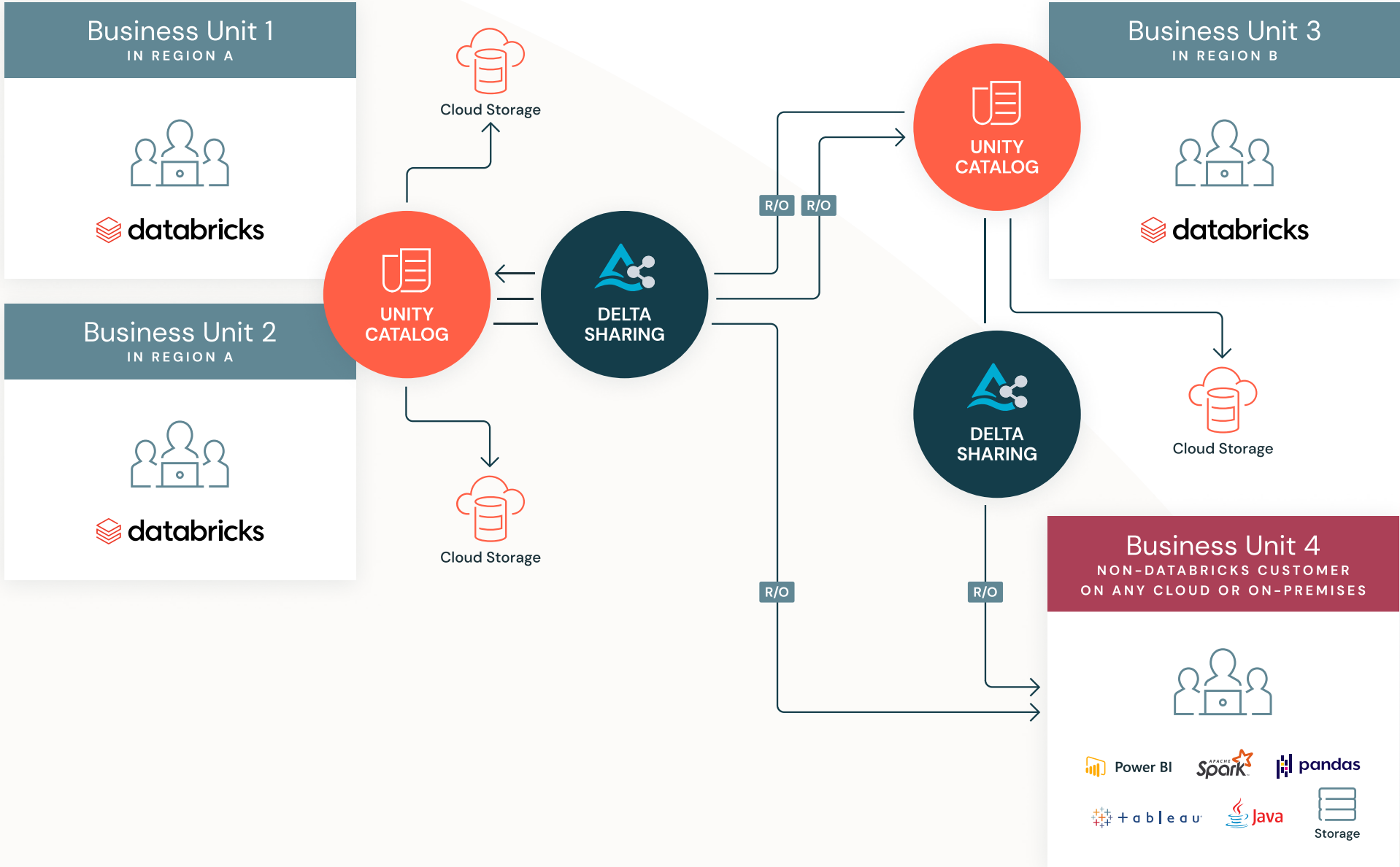


Figure 4:
Internal sharing
across BUs

To make matters even more complicated, organizations can grow through mergers and acquisitions. In such cases we can't assume that the organizations being acquired have followed the same set of rules and standards to define their platforms and to produce their data. Furthermore, we can't even assume that they've used the same cloud providers, nor can we assume the complexity of their data models.

Delta Sharing can simplify and accelerate the efforts of unification and assimilation of newly acquired organizations and their data and processes. Only selected data sources can be exchanged between the different clouds, platforms and regions. This enables teams to move freely between the organizations that are merging without losing their data — if anything, they're empowered to drive insights of higher quality by combining the data of both.

Databricks Marketplace further enhances internal sharing by offering a private exchange capability that enables organizations to securely share data and AI products with specific business units or partners. Unlike public marketplace listings, private exchanges allow providers to control who can discover and access their data and AI products. This capability is built on top of Delta Sharing, ensuring that shared data remains secure without requiring replication.

Unity Catalog plays a pivotal role in streamlining and governing internal data sharing across business units (BUs) within an organization. Unity Catalog provides centralized governance, auditing and access control for all data assets, ensuring that sharing is secure and compliant with organizational policies. It integrates seamlessly with Delta Sharing, which is the open protocol developed by Databricks for secure data sharing across different platforms.

This combination of Unity Catalog with Delta Sharing and Databricks Marketplace private exchanges provides a powerful framework for internal collaboration across BUs while maintaining robust security controls. Teams can move freely between different parts of the organization or newly acquired entities without losing access to critical data — enabling faster insights and better decision-making.

Peer-to-peer sharing with Delta Sharing

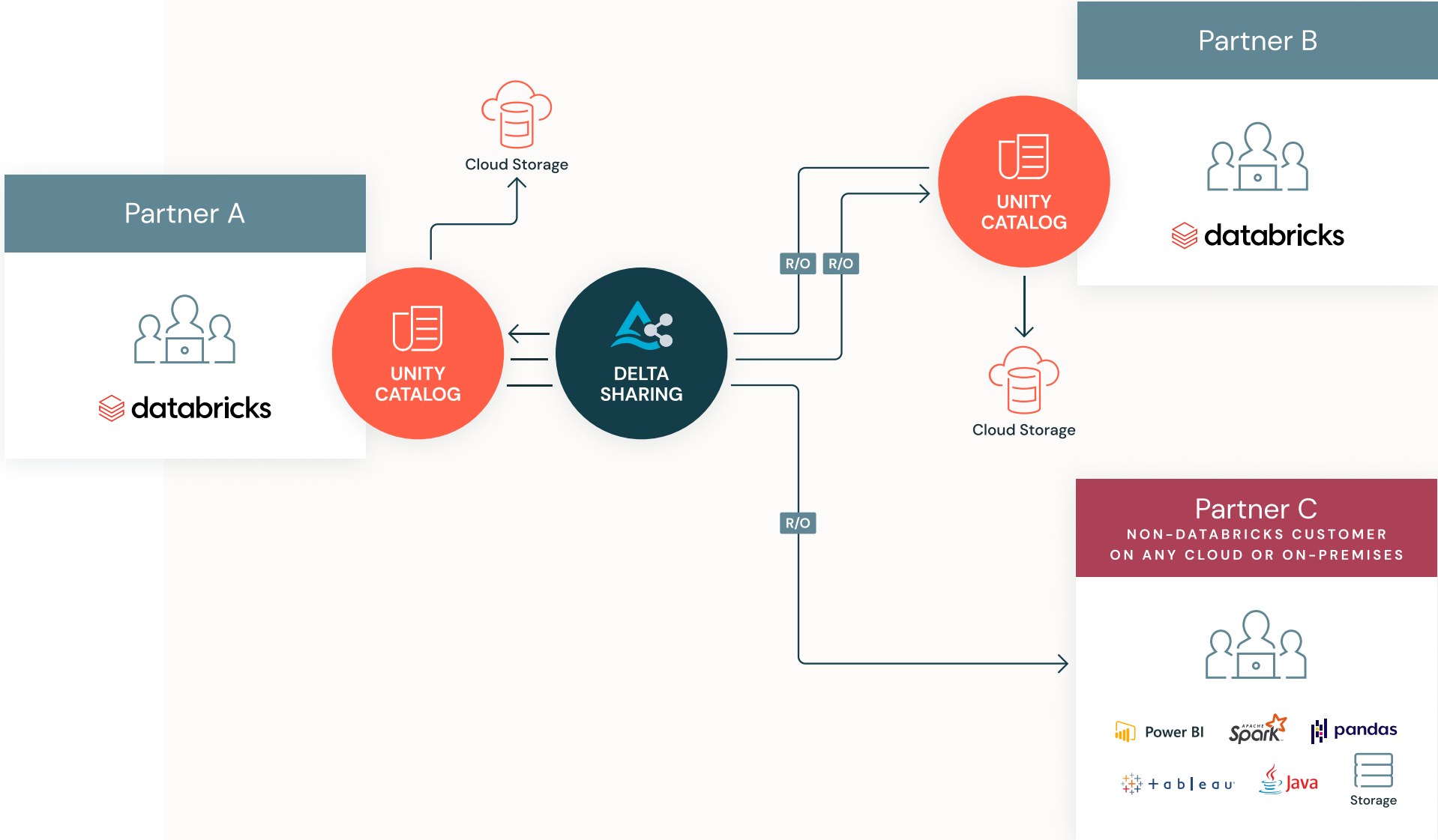


Figure 5:
B2B sharing with
Delta Sharing

Delta Sharing has become a robust solution for bidirectional data exchange, enabling companies to seamlessly integrate partners, customers and suppliers into their workflows.

Traditionally peer-to-peer sharing isn't an easy task. An organization has no control over how their partners are implementing their own data platforms. Different partners may also use various formats, protocols and methods (APIs, CSV, JSON, FTP, HTTP). The complexity further increases when we consider that the partners and suppliers can reside in public cloud, private cloud or on-premises deployed data platforms. The choices of platform and architecture aren't imposed on your partners and suppliers. Delta Sharing addresses these challenges with its open protocol, allowing data to be shared across clouds, platforms and regions without imposing specific architecture choices on partners and suppliers. This flexibility is supported by a wide array of connectors that enable data to land wherever it's needed.

Beyond just data, Delta Sharing also allows sharing of AI models with external parties to add innovative ways to collaborate. You can train your models in one place and deploy them anywhere. The shared models work with Databricks AI capabilities out of the box. Shared models appear in Unity Catalog and customers gain access to AI and governance features to productionize any model. This includes end-to-end model development capabilities, from model serving to fine-tuning, along with Unity Catalog's security and management features.

This means that you can form much more agile data exchange patterns with your partners and suppliers and attain value out of your combined data much quicker than ever before.

