## Third-party data licensing with Delta Sharing

Delta Sharing enables companies to monetize their data product simply and with necessary governance.

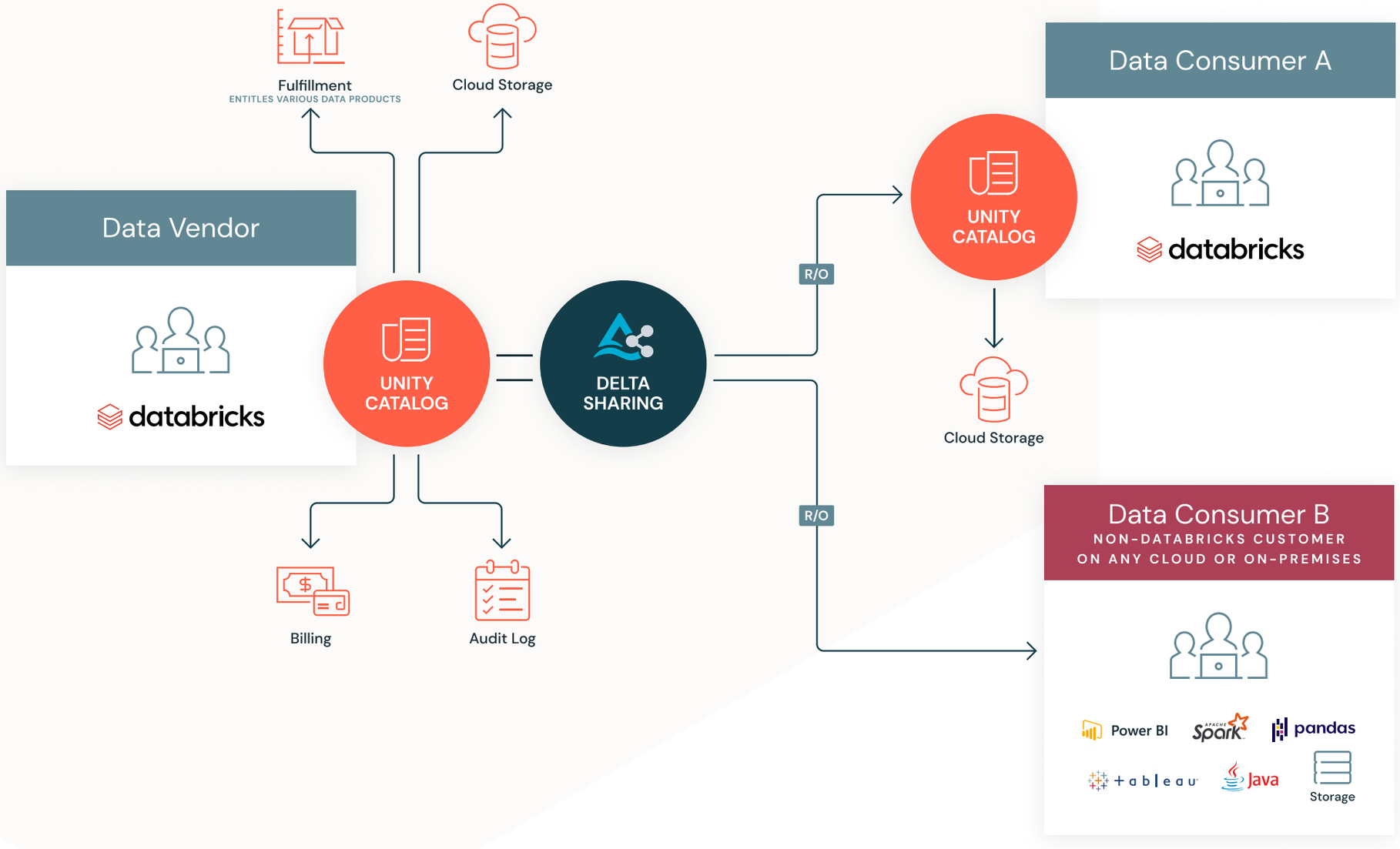
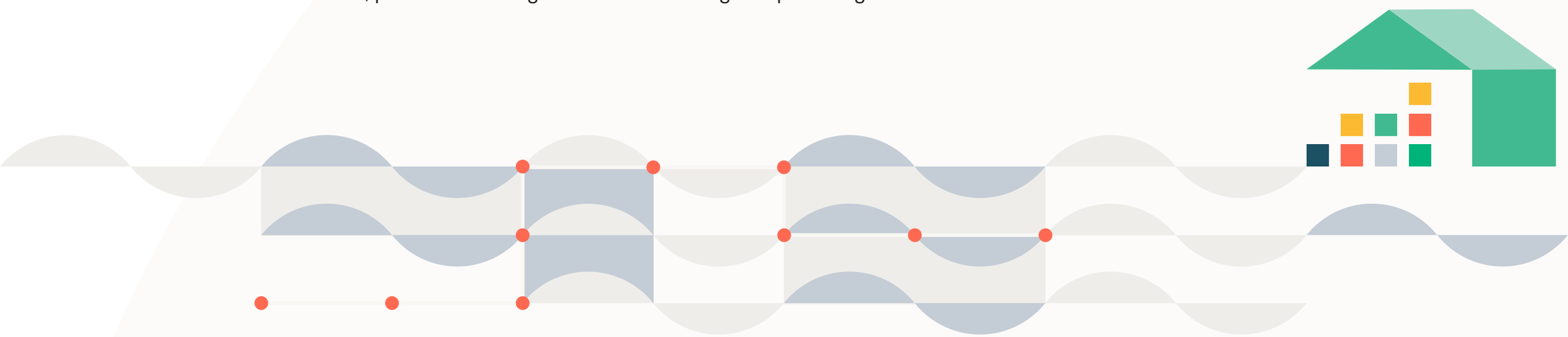



**Figure 6:**
Data monetization with Delta Sharing

Delta Sharing has significantly advanced the capabilities of data providers to license and monetize third-party data and AI models. It allows providers to seamlessly share large datasets without the scalability issues traditionally associated with SFTP servers. Unlike API services that require a dedicated service for each data product, Delta Sharing simplifies the process by enabling providers to grant and manage access to data recipients without replicating the data. There's no need to make multiple copies of data while sharing. Because there's no data movement while sharing, data providers avoid storage duplication and ensure that consumers get timely access to fresh, up-to-date data without delays. Any data exiting ELT/ETL pipelines can become a potential data product. Previously, data providers had to build complex integrations with a variety of platforms to reach all of their partners and customers. With cross-cloud and cross-platform sharing, data providers can expand their market reach to consumers across clouds, platforms and regions without needing complex integrations.



Databricks Marketplace further enhances these capabilities. Databricks Marketplace acts as an open forum for exchanging data and AI products, leveraging Delta Sharing to provide secure sharing and easy access for data consumers. Providers can list datasets, AI models, notebooks and Solution Accelerators on Databricks Marketplace, making them accessible to a broader audience. This platform supports both public listings and private exchanges, where listings are shared only with approved users.

By integrating with Databricks Marketplace, Delta Sharing enables providers to reach new buyers and accelerate sales cycles by reducing time to insight for consumers. The robust Databricks Marketplace infrastructure allows providers to showcase their offerings effectively while ensuring compliance with security and governance standards through Delta Sharing's open protocol.

To mitigate cost concerns, Delta Sharing maintains an audit log that tracks any permitted access to the data. Data providers can use this information to determine the costs associated with any of the data products and evaluate if such products are commercially viable and sensible. Additionally, Delta Sharing minimizes costs by eliminating the need for data replication. The protocol also supports integration with Cloudflare R2, a storage solution that incurs no egress fees, further lowering costs for cross-region or cross-cloud data sharing.

# Chapter 4
# How Delta Sharing Works

Delta Sharing is designed to be simple, scalable, nonproprietary and cost-effective for organizations that are serious about getting more from their data. Delta Sharing is natively integrated with Unity Catalog, which enables customers to add fine-grained governance and security controls, making it easy and safe to share data internally or externally.

Delta Sharing is a simple REST protocol that securely grants temporary access to part of a cloud dataset. It leverages modern cloud storage systems — such as AWS S3, Azure ADLS or Google GCS — to reliably grant read-only access to large datasets. Here's how it works for data providers and data recipients.

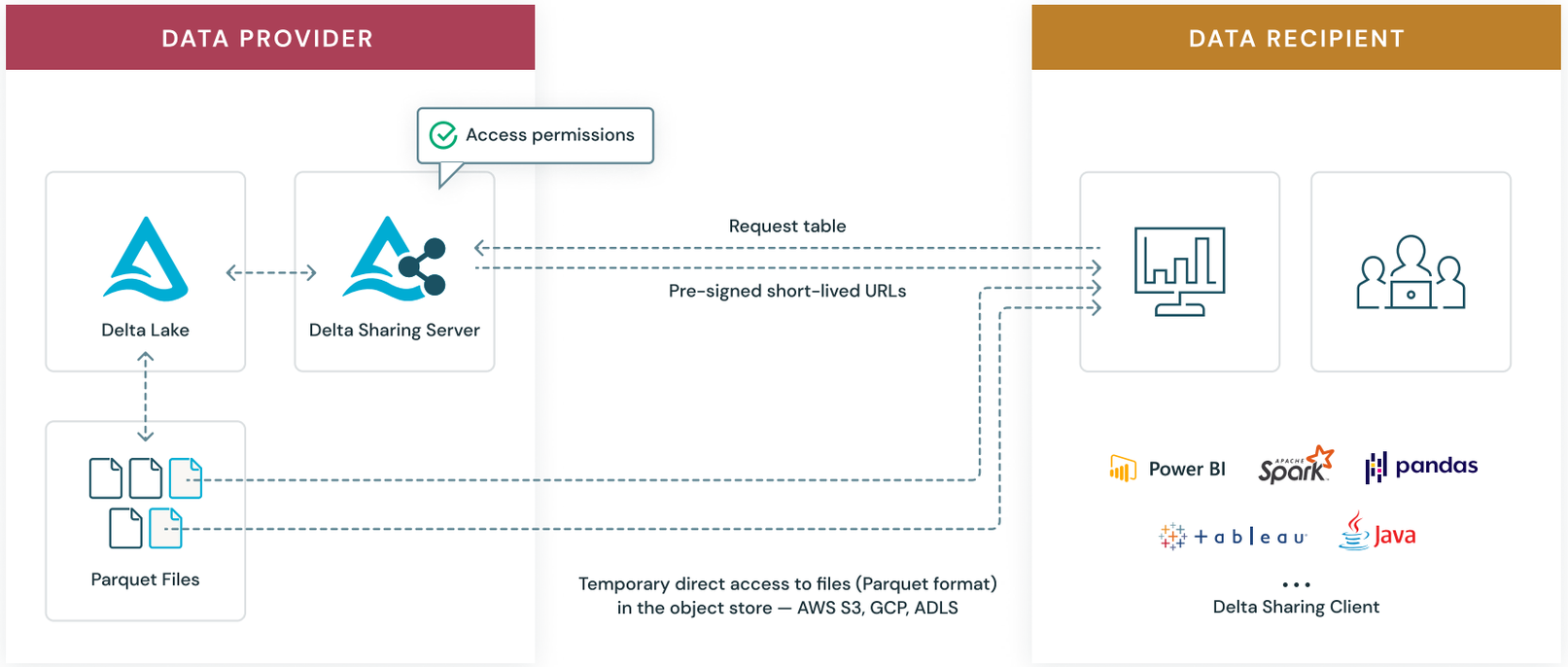



**Figure 7:** How Delta Sharing works connecting data providers and data recipients

## Data providers

The data provider shares existing tables or parts thereof (such as specific table versions or partitions) stored on the cloud data lake in Delta Lake format. The provider decides what data they want to share and runs a sharing server in front of it that implements the Delta Sharing protocol and manages recipient access. To manage shares and recipients, you can use SQL commands, the Unity Catalog CLI or the intuitive user interface.

## Data recipients

The data recipient only needs one of the many Delta Sharing clients that support the protocol. Databricks has released open source connectors for pandas, Apache Spark, Java and Python, and is working with partners on many more.

## The data exchange

The Delta Sharing data exchange follows three efficient steps:

1. The recipient's client authenticates to the sharing server and asks to query a specific table. The client can also provide filters on the data (for example, "country=US") as a hint to read just a subset of the data.
2. The server verifies whether the client is allowed to access the data, logs the request and then determines which data to send back. This will be a subset of the data objects in cloud storage systems that make up the table.
3. To allow temporary access to the data, the server generates short-lived presigned URLs that allow the client to read Parquet files directly from the cloud provider. This allows read-only access in parallel at massive bandwidth without streaming through the sharing server.

# Chapter 5
# Introducing Databricks Marketplace

Enterprises need open collaboration for data and AI. Data sharing — within an organization or externally — allows companies to collaborate with partners, establish new partnerships and generate new revenue streams with data monetization.

The demand for generative AI is driving disruption across industries, increasing the urgency for technical teams to build generative AI models and large language models (LLMs) on top of their own data to differentiate their offerings.

Traditional data marketplaces are restricted and offer only data or simple applications, which limits their value to data consumers. They also don't offer tools to evaluate the data assets beyond basic descriptions or examples. Finally, data delivery is limited, often requiring ETL or a proprietary delivery mechanism.

Enterprises need a better way to share data and AI that is flexible, secure and unlocks business value. An ecosystem makes data sharing and collaboration powerful.

**Data marketplaces present many challenges, and collaboration can be complex for both data consumers and data providers.**

| Data Consumers | Data Providers |
| --- | --- |
| Focus on data only or simple applications | Limited opportunities to monetize new types of assets |
| Lengthy discovery and evaluation | Difficulty reaching more users |
| Delayed time to insights with vendor lock-in | Lack of secure technology and unified governance |