Delta Sharing

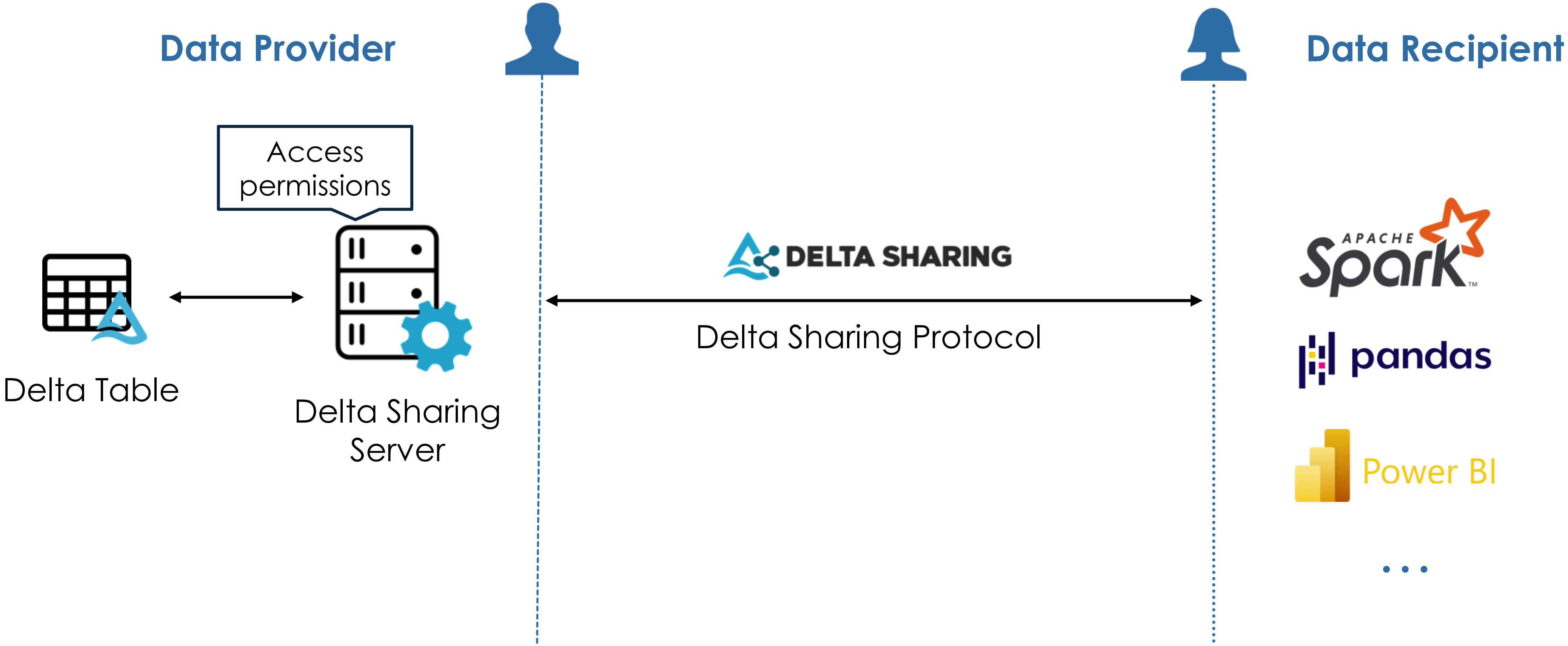
Learning Objectives

- What is Delta Sharing
- How it works under the hood
- Costs and limitations

Delta Sharing

- Open protocol for secure data sharing
- Share live data with organizations regardless of computing platforms

How it Works



Data Provider

Data Recipient

Access permissions

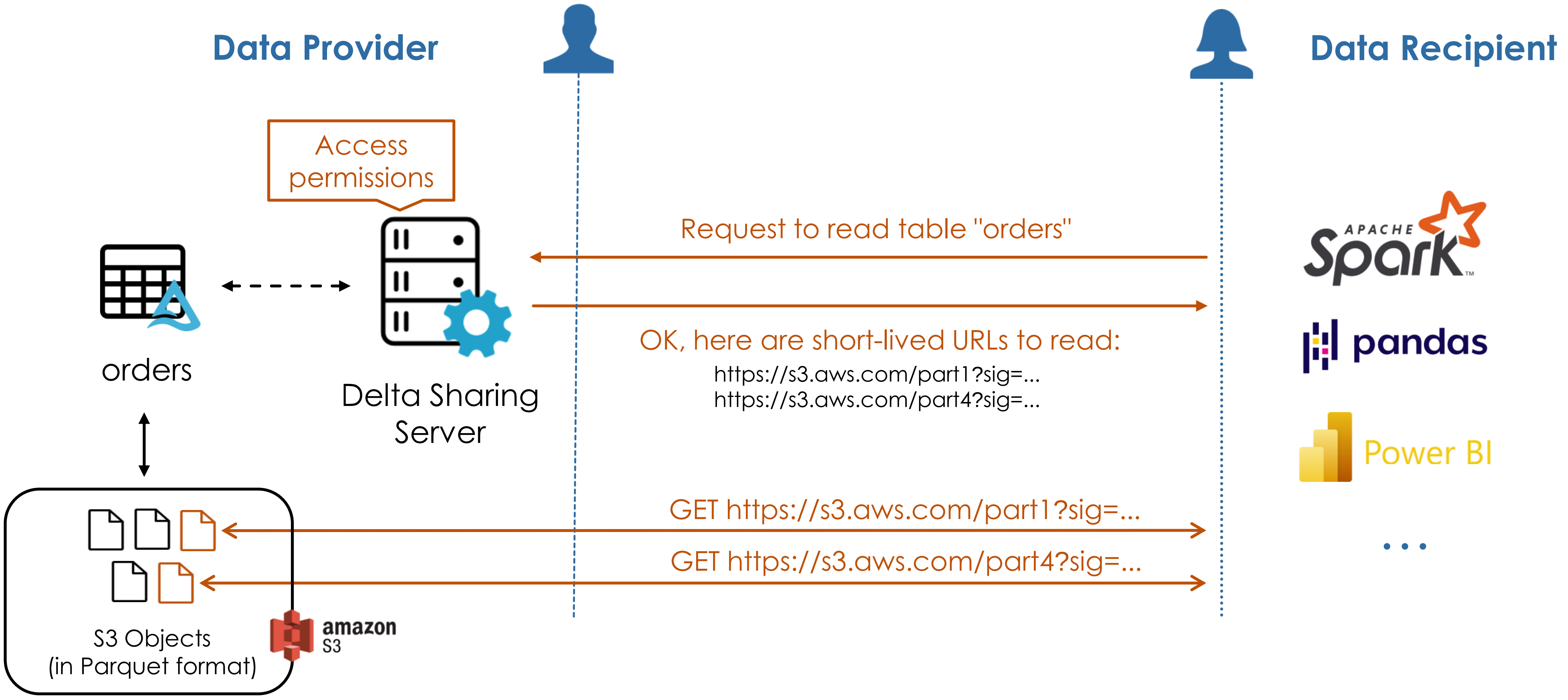
Delta Table

Delta Sharing Server

DELTA SHARING

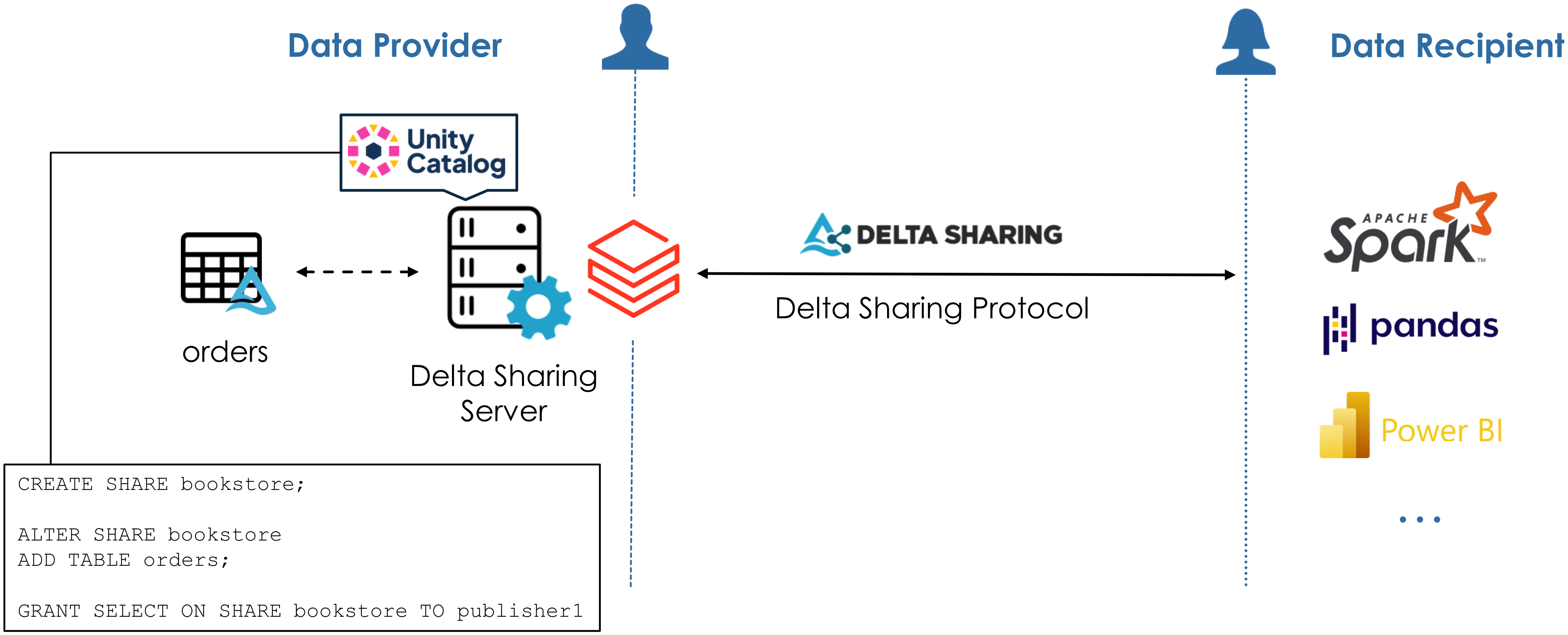
Delta Sharing Protocol

Under the Hood



Data Provider
Data Recipient
Access permissions
orders
Delta Sharing Server
Request to read table "orders"
OK, here are short-lived URLs to read:
https://s3.aws.com/part1?sig=...
https://s3.aws.com/part4?sig=...
GET https://s3.aws.com/part1?sig=...
GET https://s3.aws.com/part4?sig=...
S3 Objects (in Parquet format)
amazon S3
APACHE Spark
pandas
Power BI
. . .

Delta Sharing on Databricks



```
CREATE SHARE bookstore;

ALTER SHARE bookstore
ADD TABLE orders;

GRANT SELECT ON SHARE bookstore TO publisher1
```

Delta Sharing on Databricks

- D2D: The Databricks-to-Databricks sharing protocol
  - Share data between Databricks clients
  - Collection of tables, views, volumes, and notebooks
  - Leverages built-in authentication with no token exchange

- D2O: The Databricks open sharing protocol
  - Share data with users on any computing platform
  - Share tables only
  - Requires authentication via bearer tokens or OpenID Connect (OIDC) Federation

History Sharing

- Databricks-to-Databricks sharing can allow recipients to perform time travel queries and streaming reads by enabling history sharing:
  - `ALTER SHARE <share_name> ADD TABLE <table_name> WITH HISTORY;`
- Allows querying the change data feed of the shared table.
  - CDF must be enabled on the table before you share it WITH HISTORY.
- Improves performance by leveraging temporary cloud storage credentials scoped-down to the root directory of the shared table.
  - Resulting in performance equivalent to direct access to the source table.
  - Performance benefits do not apply to partitioned tables.

Roles

- Users authorized to create and manage Delta Shares:
  - Metastore admins
  - Users with CREATE SHARE privilege on the metastore

Costs

- Delta Sharing does not require data replication

- Egress cost
  - Within a region: no egress cost
  - Cross-clouds or cross-regions: cloud vendor charges data egress fees, instead:
    - Clone the shared data to local regions
    - Share data from Cloudflare R2

Limitations

- Read-Only Access Model
- Data Format Constraints
  - Only Delta tables are supported for sharing