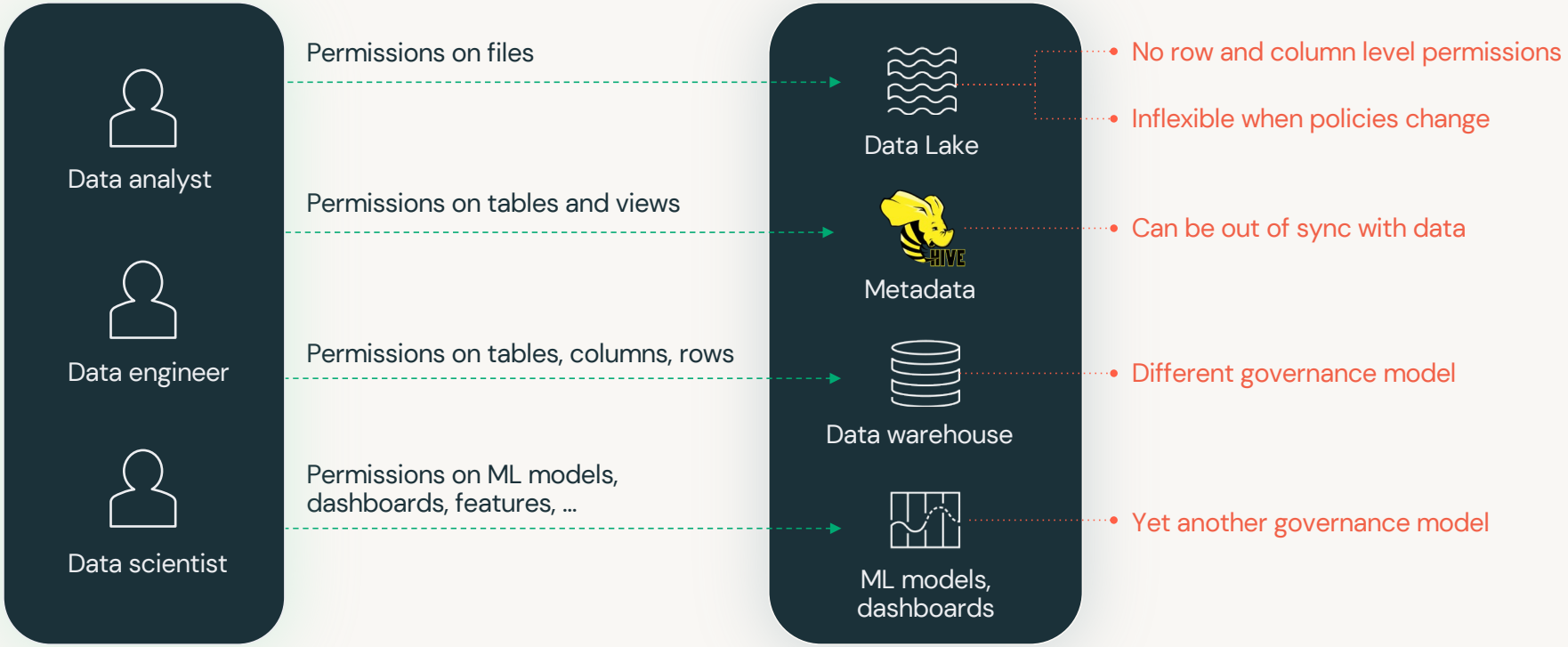
databricks

Unity Catalog

Unified governance for all data and AI assets

©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

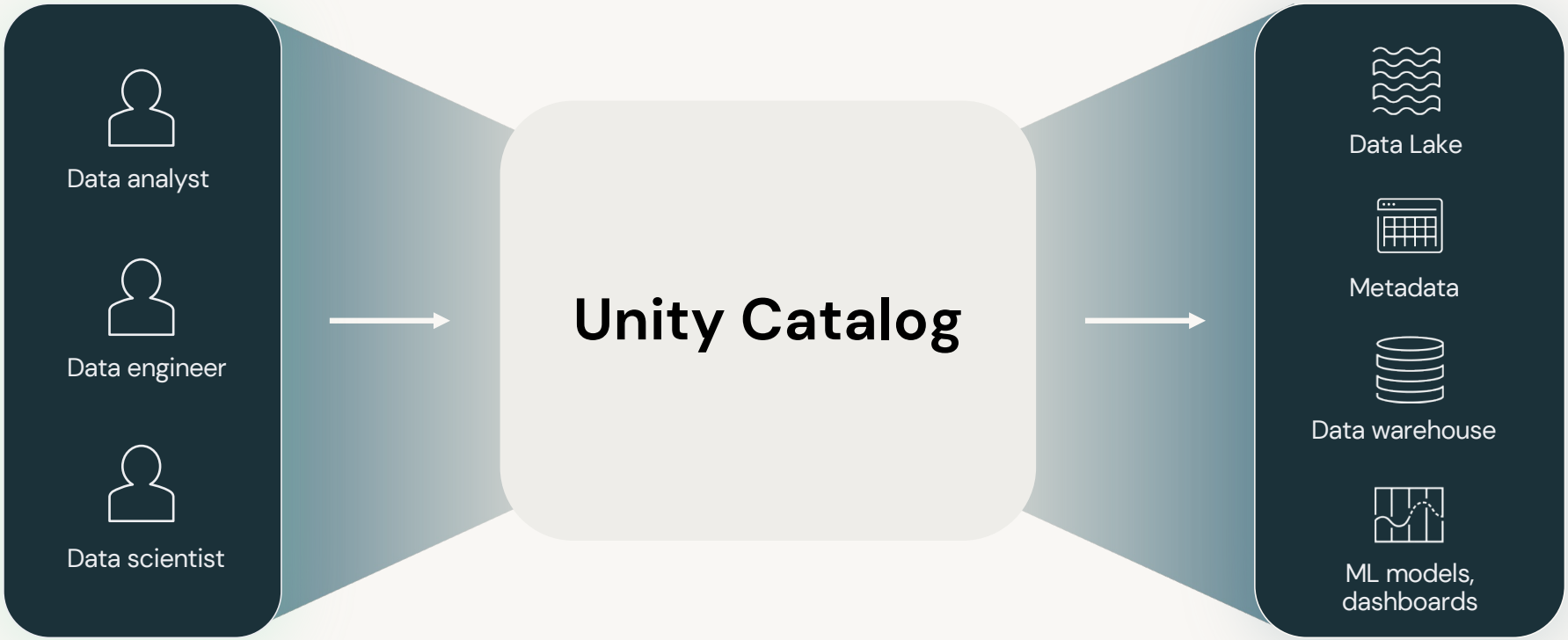
Governance for data and AI is complex



©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

Databricks Unity Catalog

Unified governance for all data and AI assets

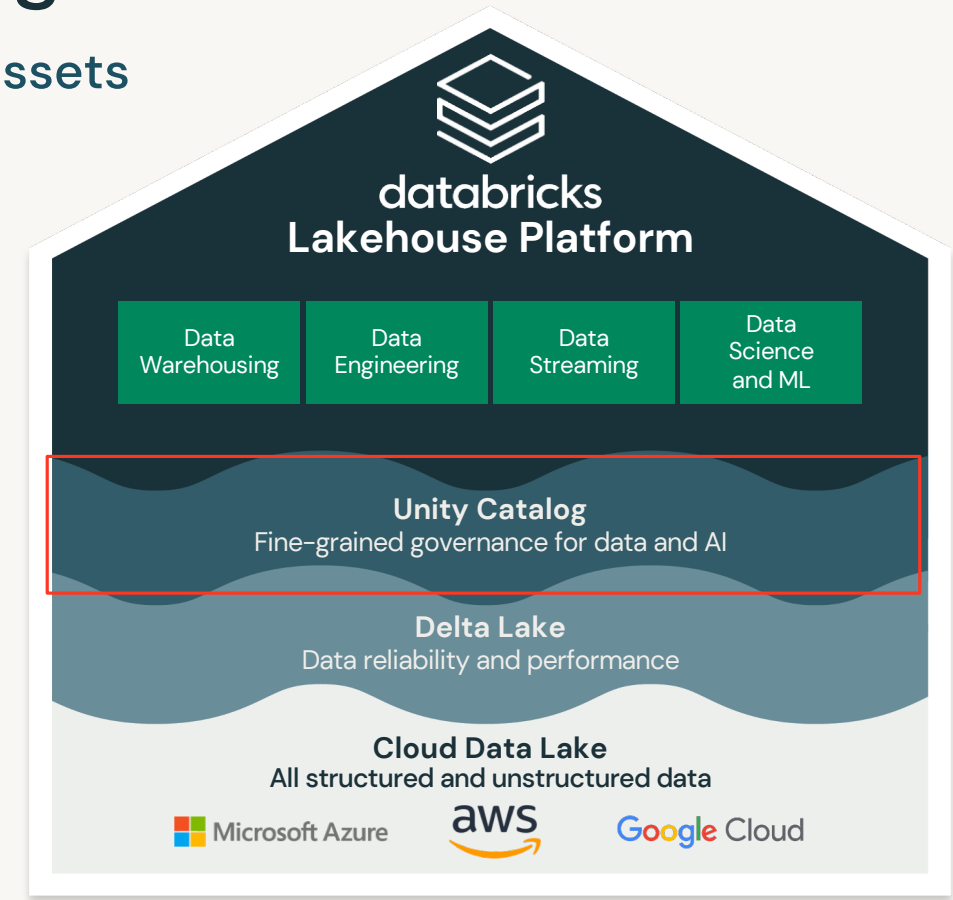


©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

Databricks Unity Catalog

Unified governance for all data and AI assets

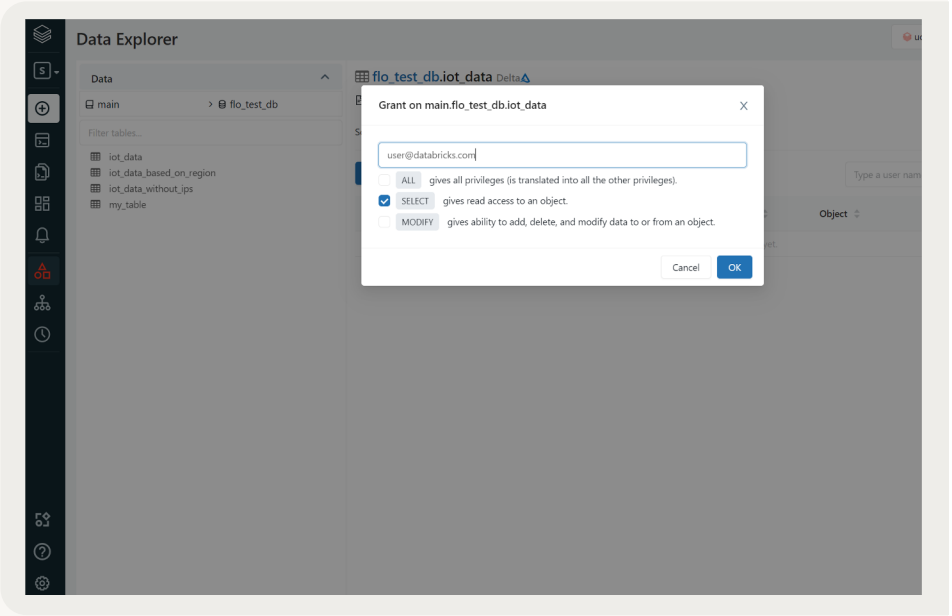
- Centralized governance for data and AI
- Built-in data search and discovery
- Performance and scale
- Automated lineage for all workloads
- Integrated with your existing tools



©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

Centralized governance for data and AI

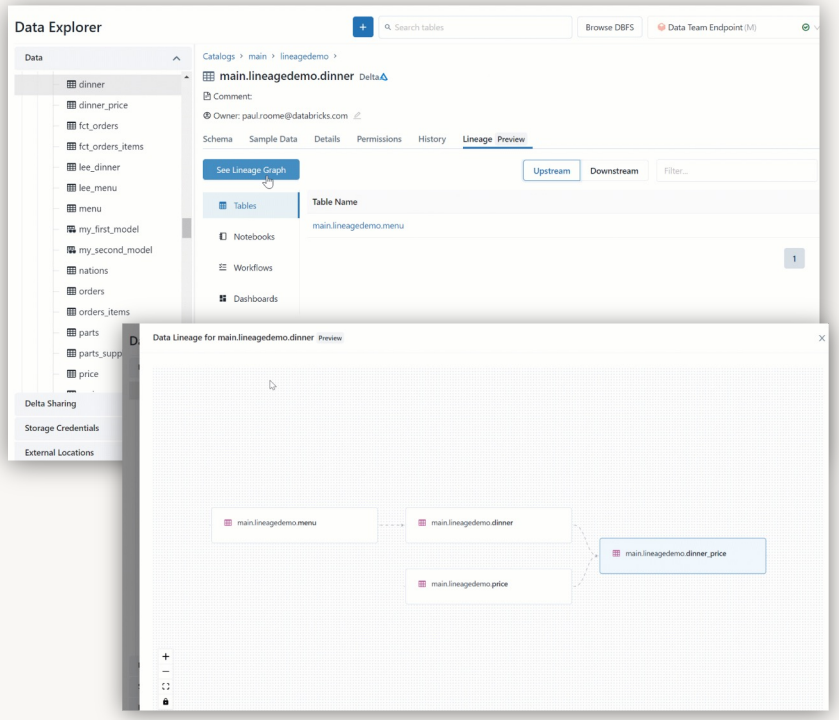
- Create a single source of truth for your data with centralized metadata and user management
- Centrally manage access permissions and audit controls for files, tables across all workspaces and workloads using open standard ANSI SQL
- Enable fine-grained access controls on tables, files, rows, and columns



©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

Automated lineage for all workloads

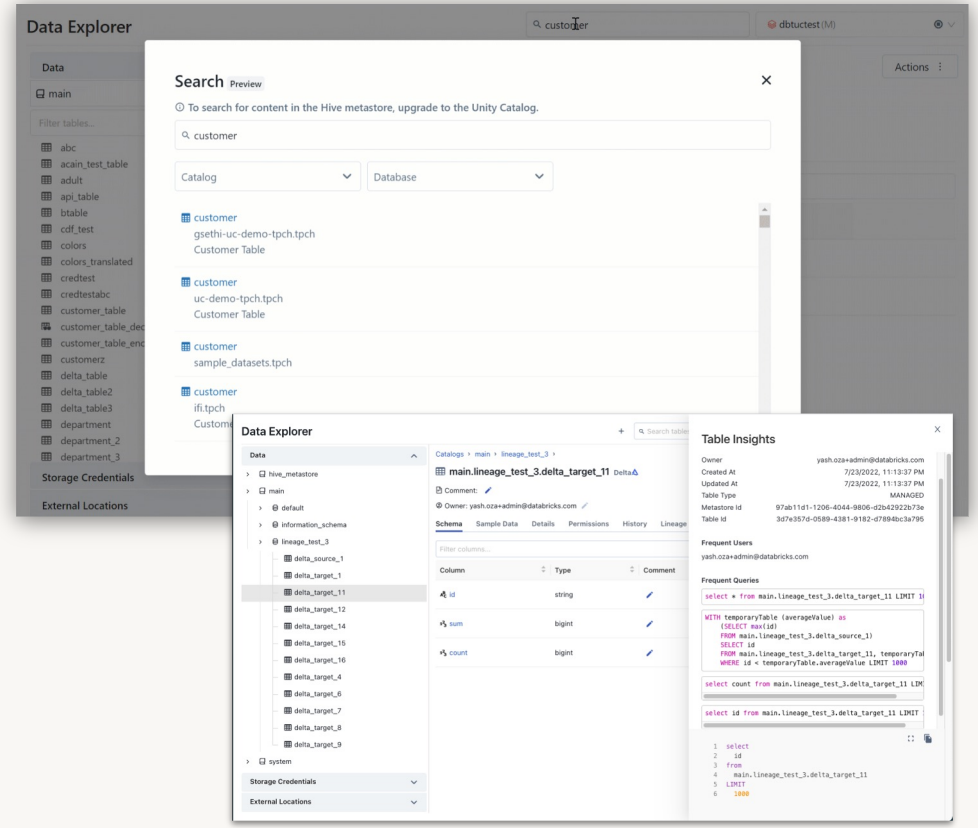
- Get an end-to-end visibility into how data flows in your organization from source to consumption with built-in data lineage
- View lineage across tables, columns, notebooks, workflows, dashboards
- Captured in real time across all workloads in SQL, Python, Scala, and R
- Export API to enable partner integrations



©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

Built-in search and discovery

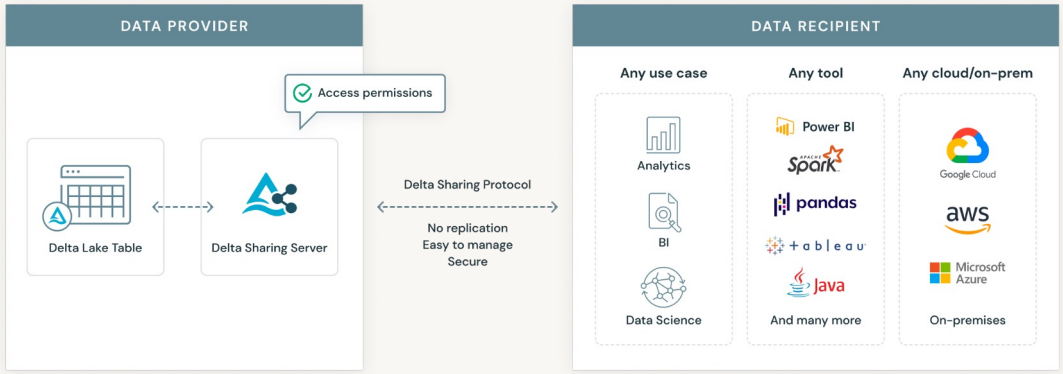
- Quickly find, understand, and reference data from across your data estate with a unified data browsing experience
- Secure by default—leverages common permission model from Unity Catalog
- Auto-generated data insights to understand your data



©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

Secure data sharing and collaboration

- Securely share live data from your lakehouse to any computing platform
- Eliminate unnecessary data replication, and vendor lock-in
- Centrally manage, govern, monitor usage and audit access to shared data with ease

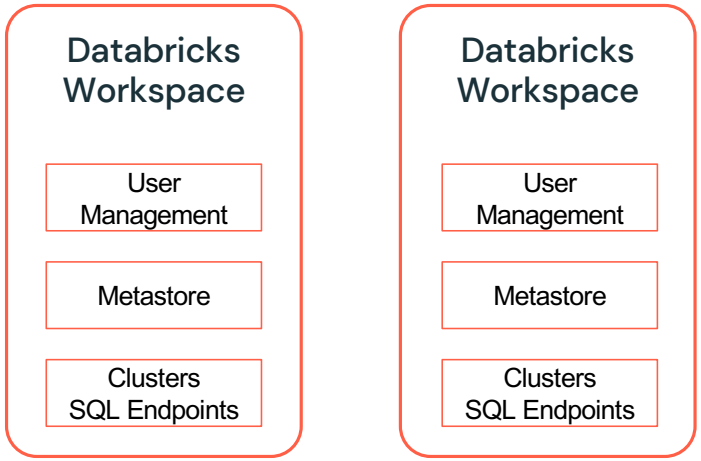


©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

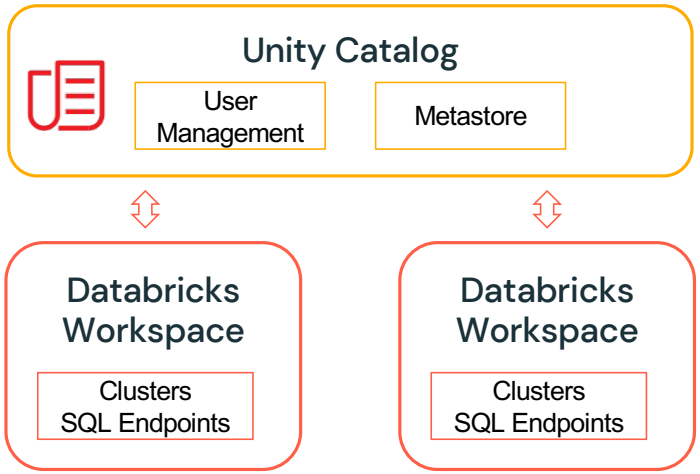
Introducing Unity Catalog

Unified view of your data estate via a centralized metadata and user mgmt

Without Unity Catalog

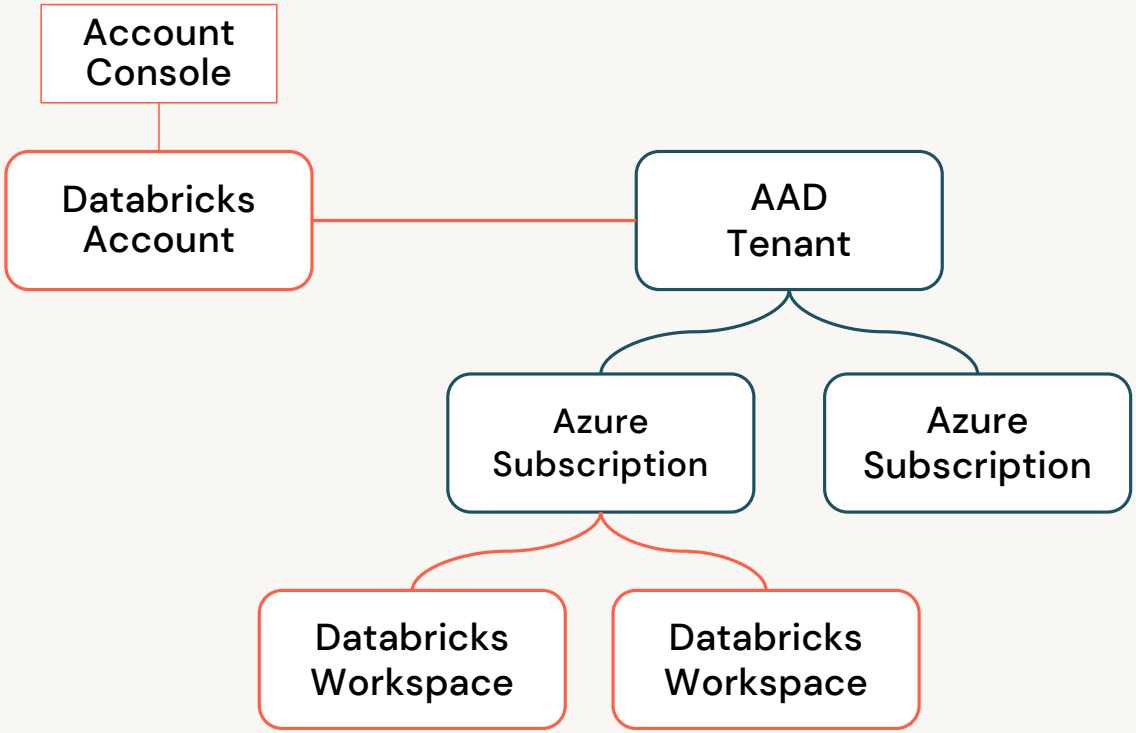


With Unity Catalog



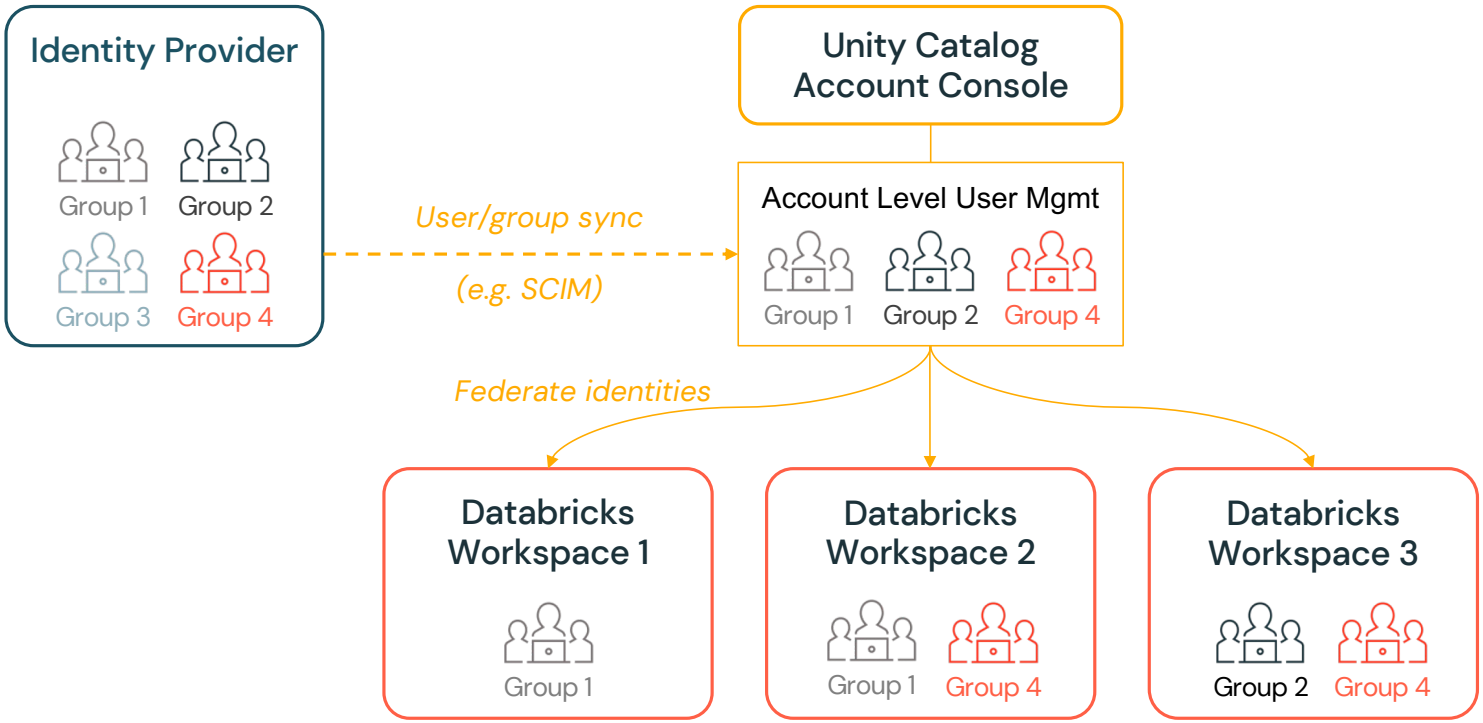
©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

Databricks Accounts and the Cloud Provider Hierarchy



©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

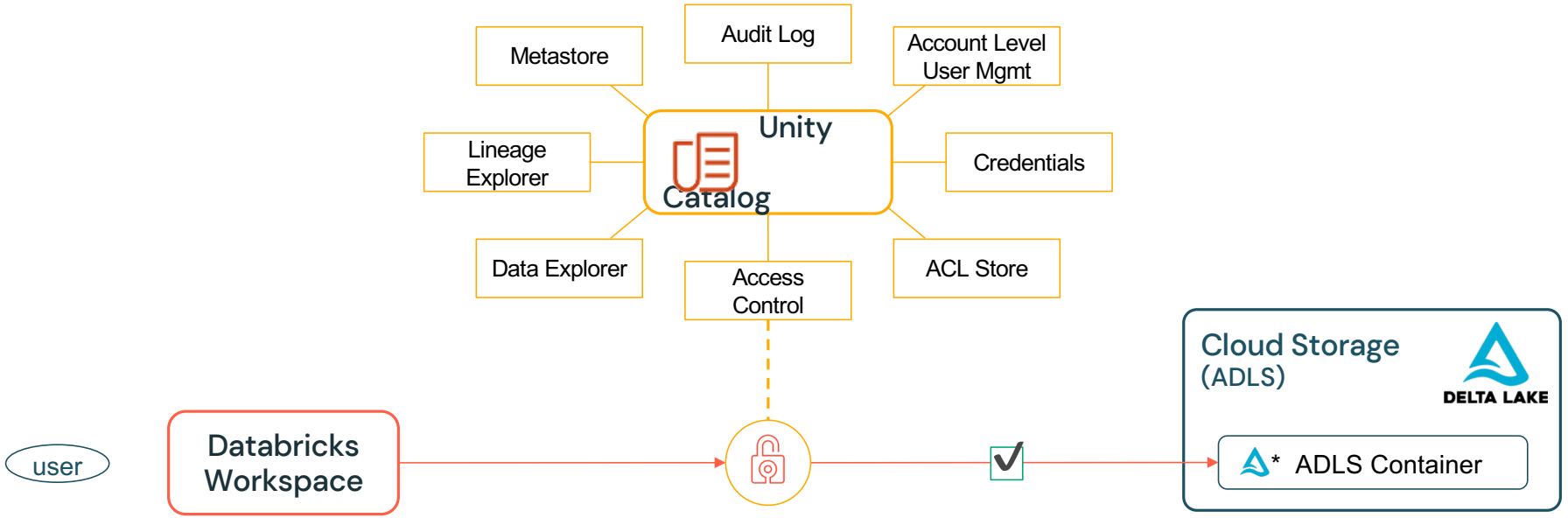
Identity Federation with Unity Catalog



©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

Unity Catalog - Components

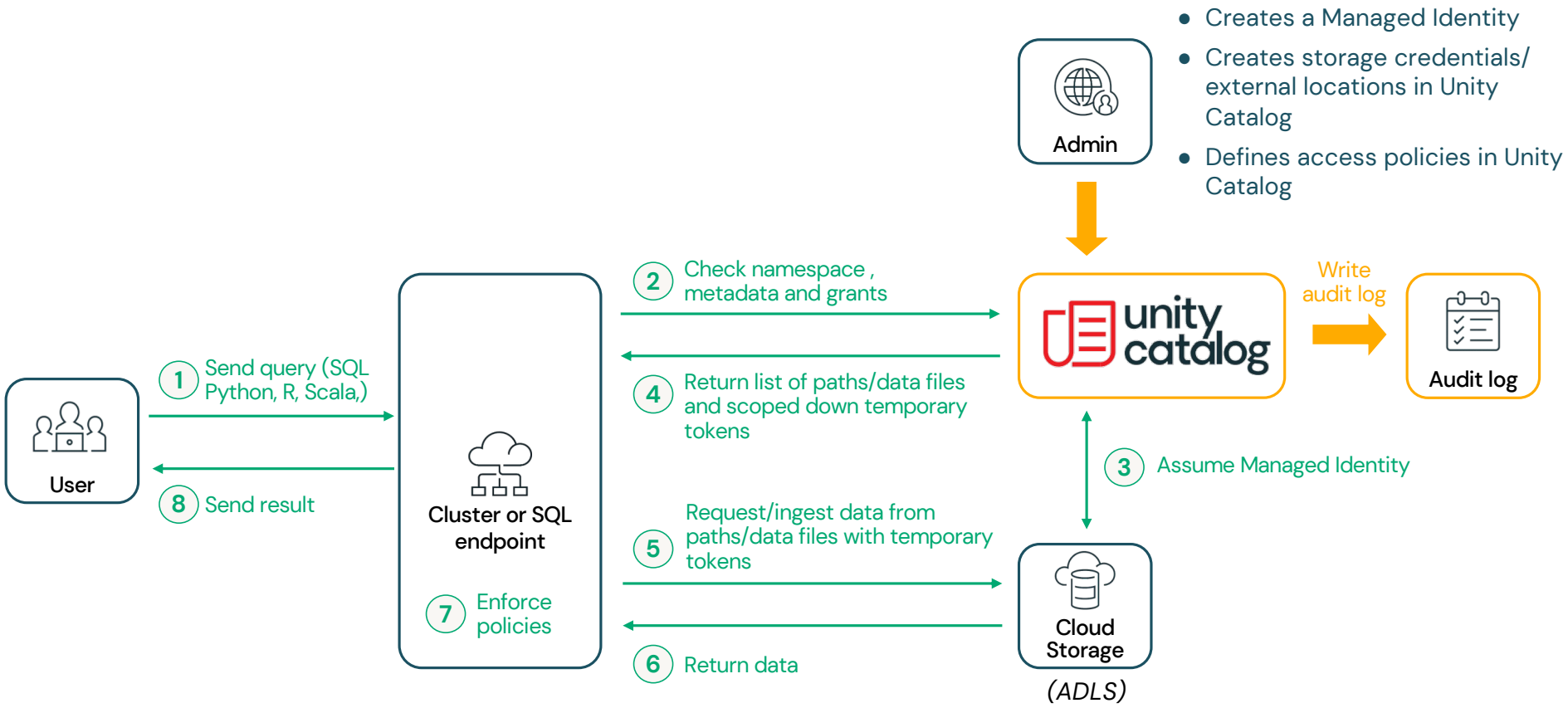
Centralized Governance



©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

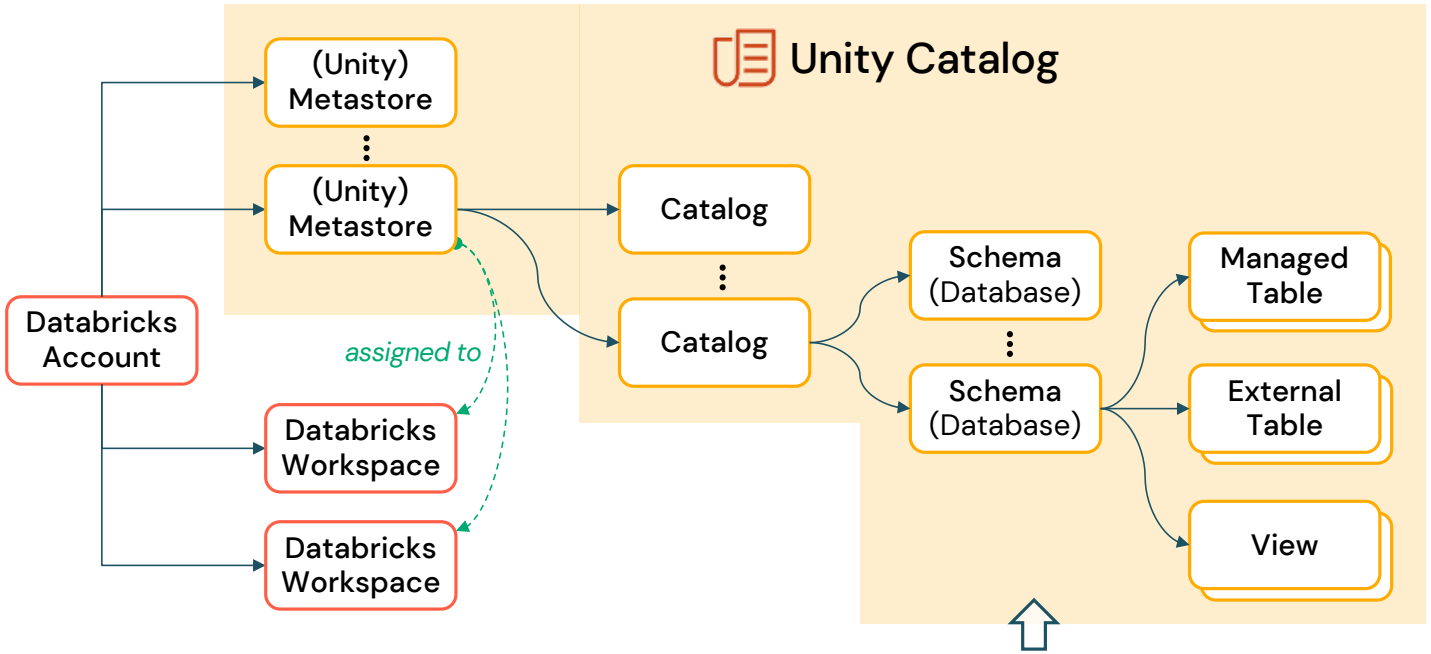
* Unity Catalog will support any data format (table or raw files)

Life of a query with Unity Catalog



©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

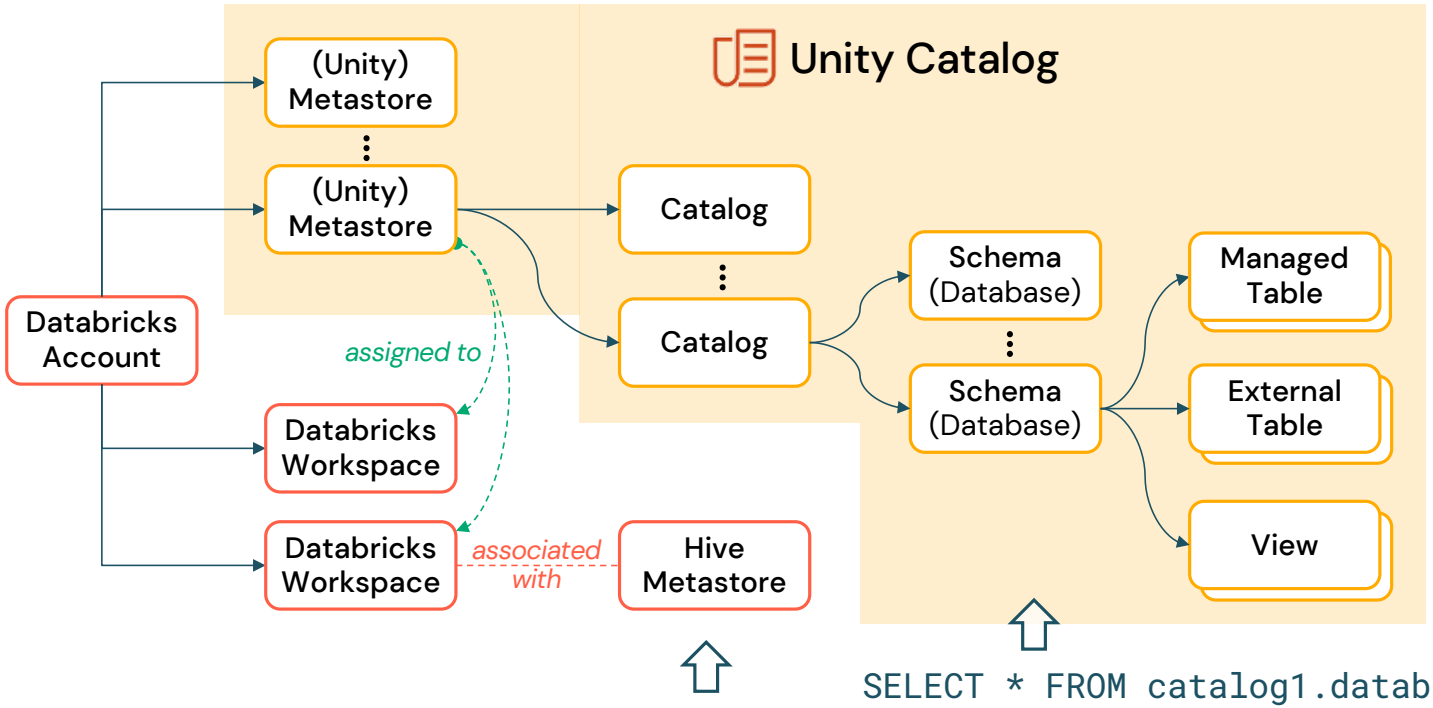
The three level namespace of UC




```
SELECT * FROM catalog1.database1.table1;
```

©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

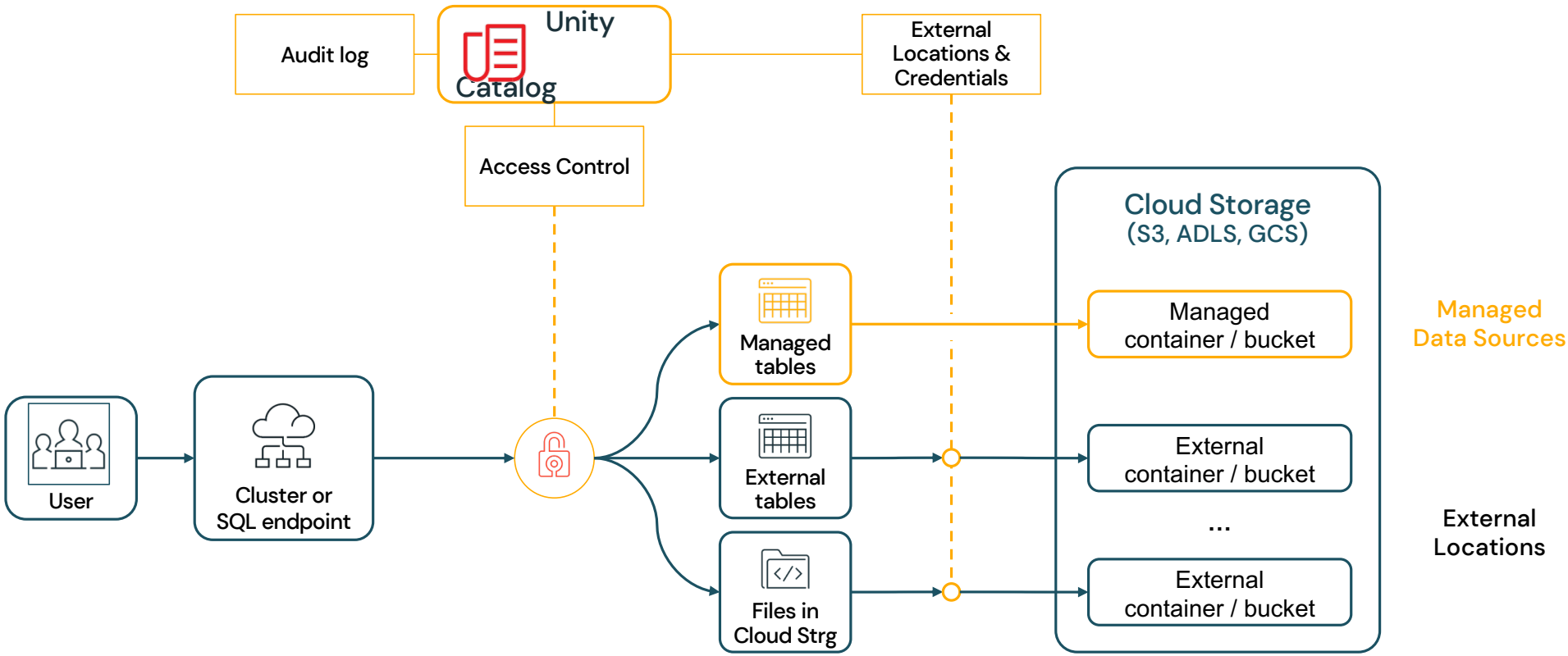
Hive Metastore is integrated into UC




```
SELECT * FROM hive_metastore.database2.table2;
```

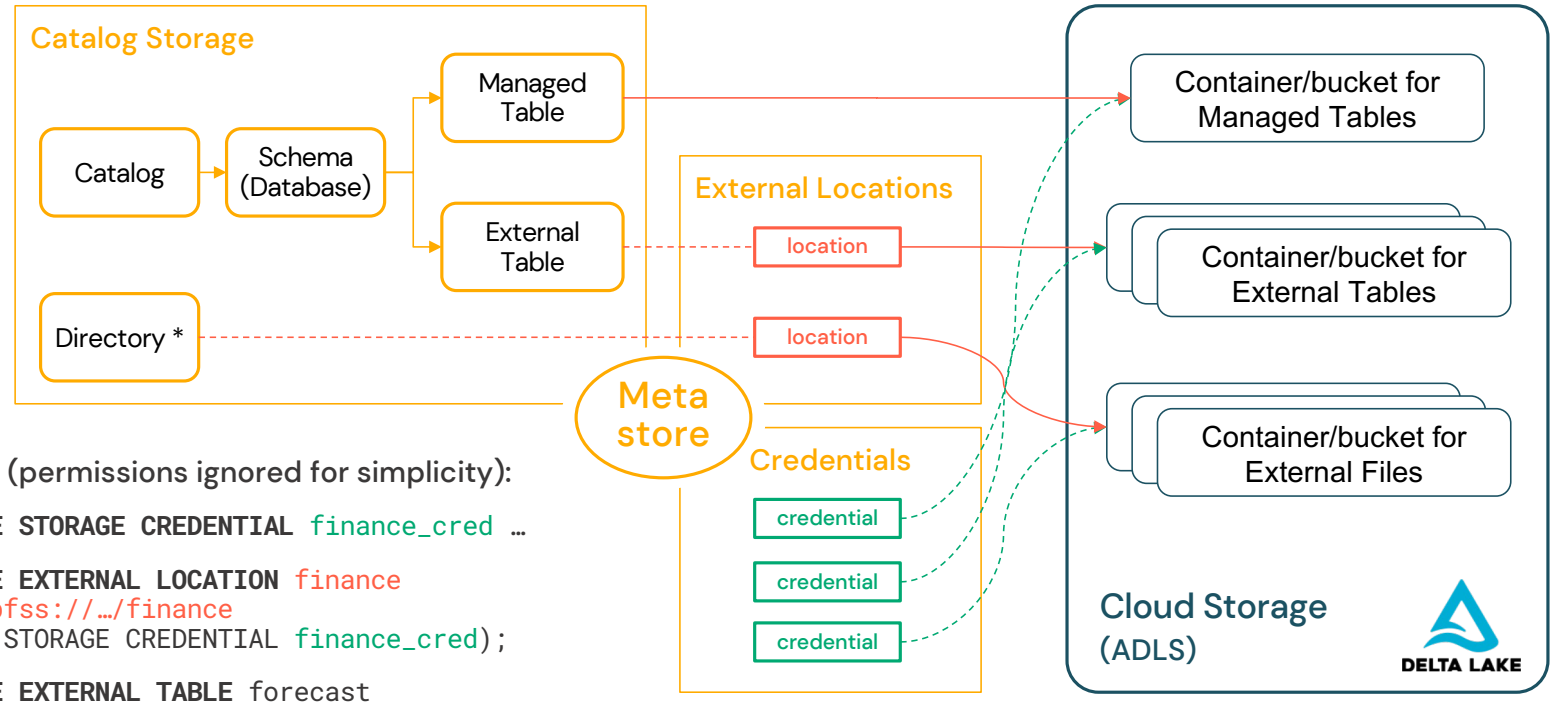
©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

Managed Data Sources & External Locations



©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

Metastore, external locations and credentials



Example (permissions ignored for simplicity):

```
CREATE STORAGE CREDENTIAL finance_cred …

CREATE EXTERNAL LOCATION finance
URL abfss://…/finance
WITH (STORAGE CREDENTIAL finance_cred);

CREATE EXTERNAL TABLE forecast
LOCATION abfss://…/finance/forecast;

* CREATE DIRECTORY eu_invoices
  LOCATION abfss://…/finance/eu/invoices;
```

* coming soon

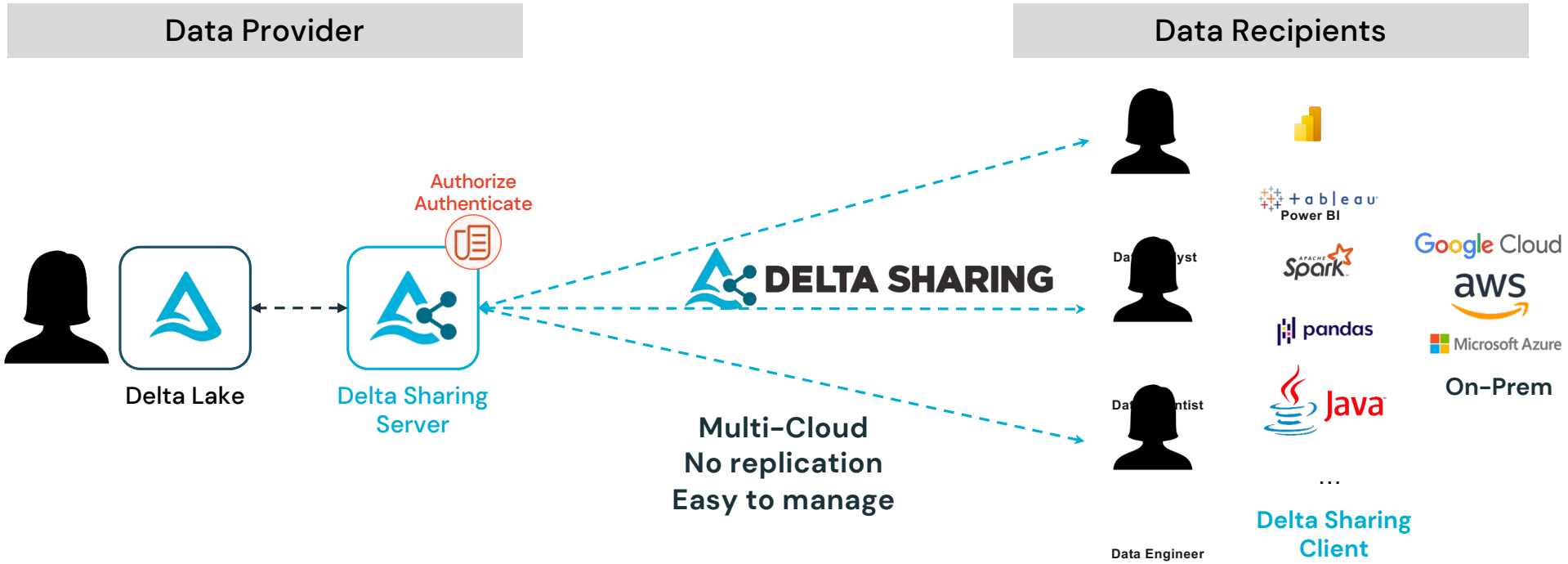
©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

UC and Delta Sharing

©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

A simple, open and easy approach to data sharing

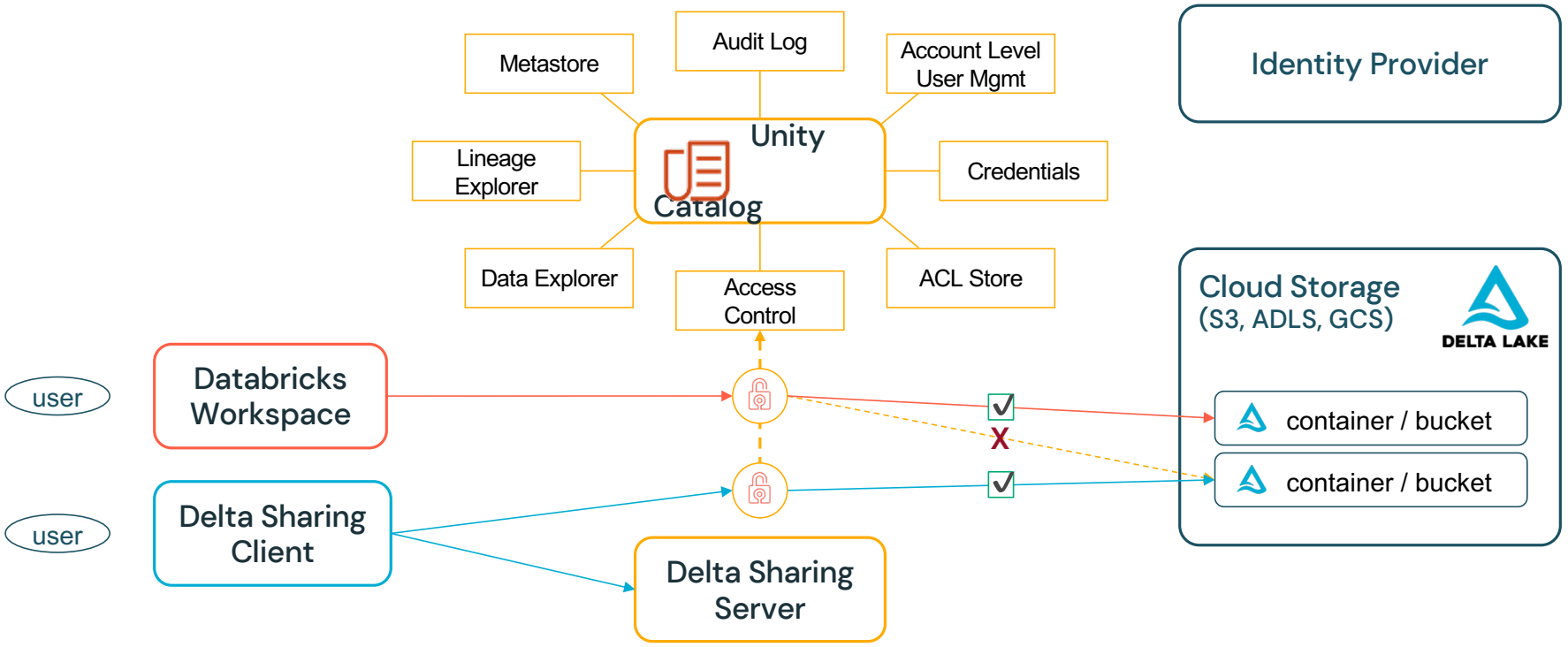
Reduce data sharing and collaboration from days to real-time



©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

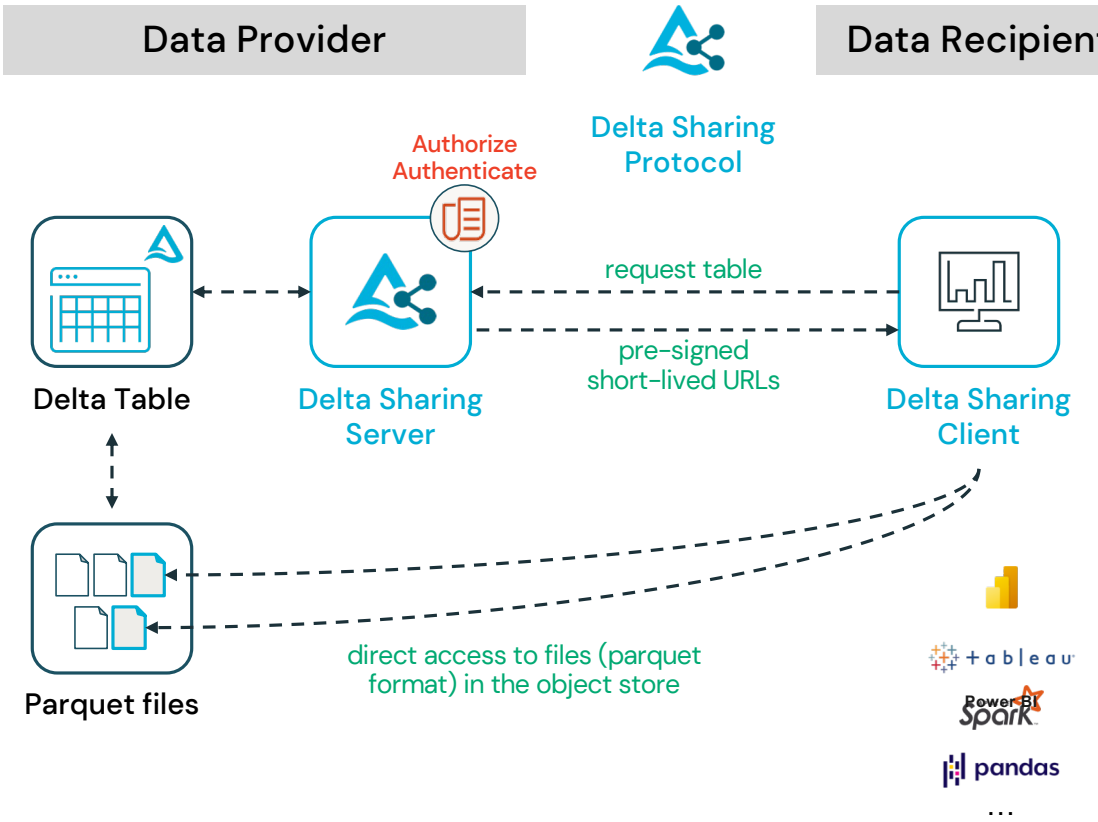
Unity Catalog & Delta Sharing - Components

Centralized Governance



©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

Under the Hood



Delta Sharing Protocol:

- Client authenticates to Sharing Server
- Client requests a table (including filters)
- Server checks access permissions
- Server generates and returns pre-signed short-lived URLs
- Client uses URLs to directly read files from object storage

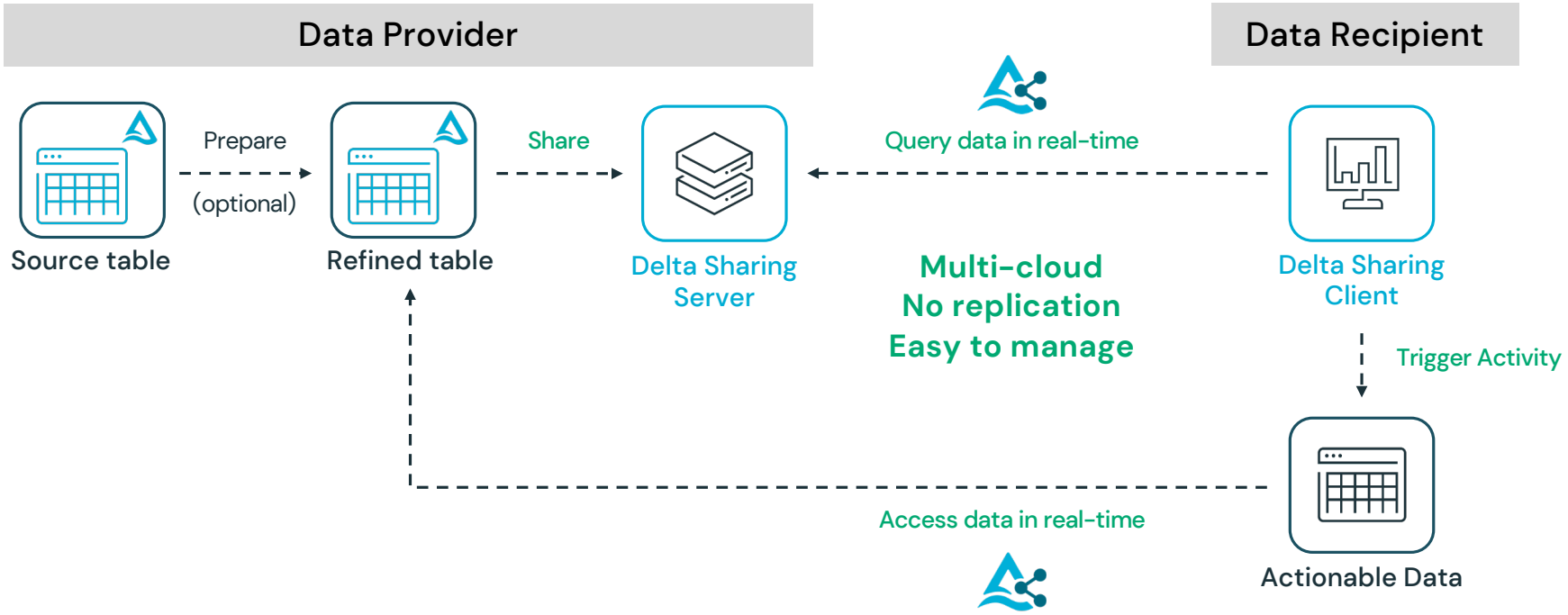
Notes:

- Sharing happens on Delta part files, supporting full tables, partitions, delta versions, ...
- Client is system independent, just needs to be able to read parquet files
- In Databricks Sharing Server and ACL checks are integrated with Unity Catalog

©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

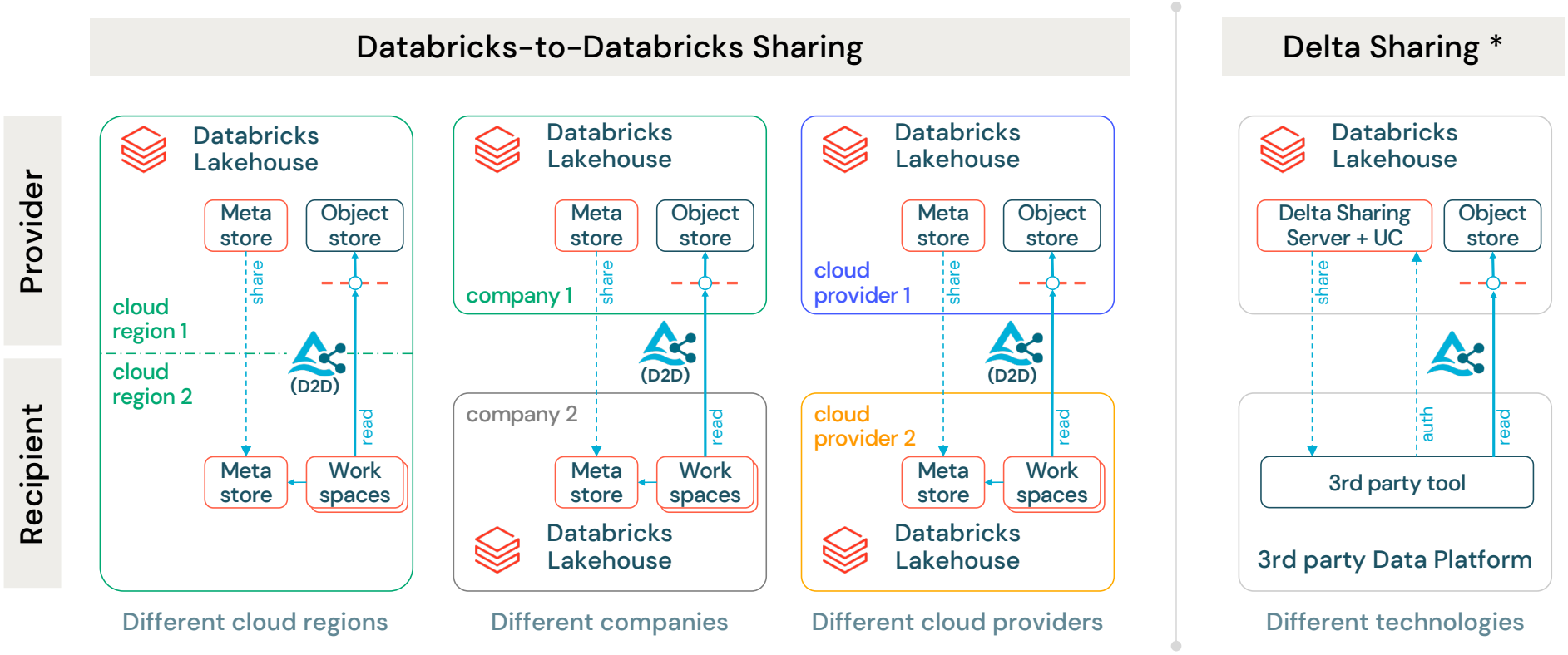
Streamlined Sharing with Delta Sharing

Delta Sharing cuts collaboration time with partners from days to real-time.



©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

Databricks-to-Databricks (D2D) Sharing



©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential

Provide access to the data on a network level

* Implementing the Delta Sharing protocol, 3rd parties can act as a data provider

©2023 Databricks Inc. — All rights reserved
Classified as Microsoft Confidential