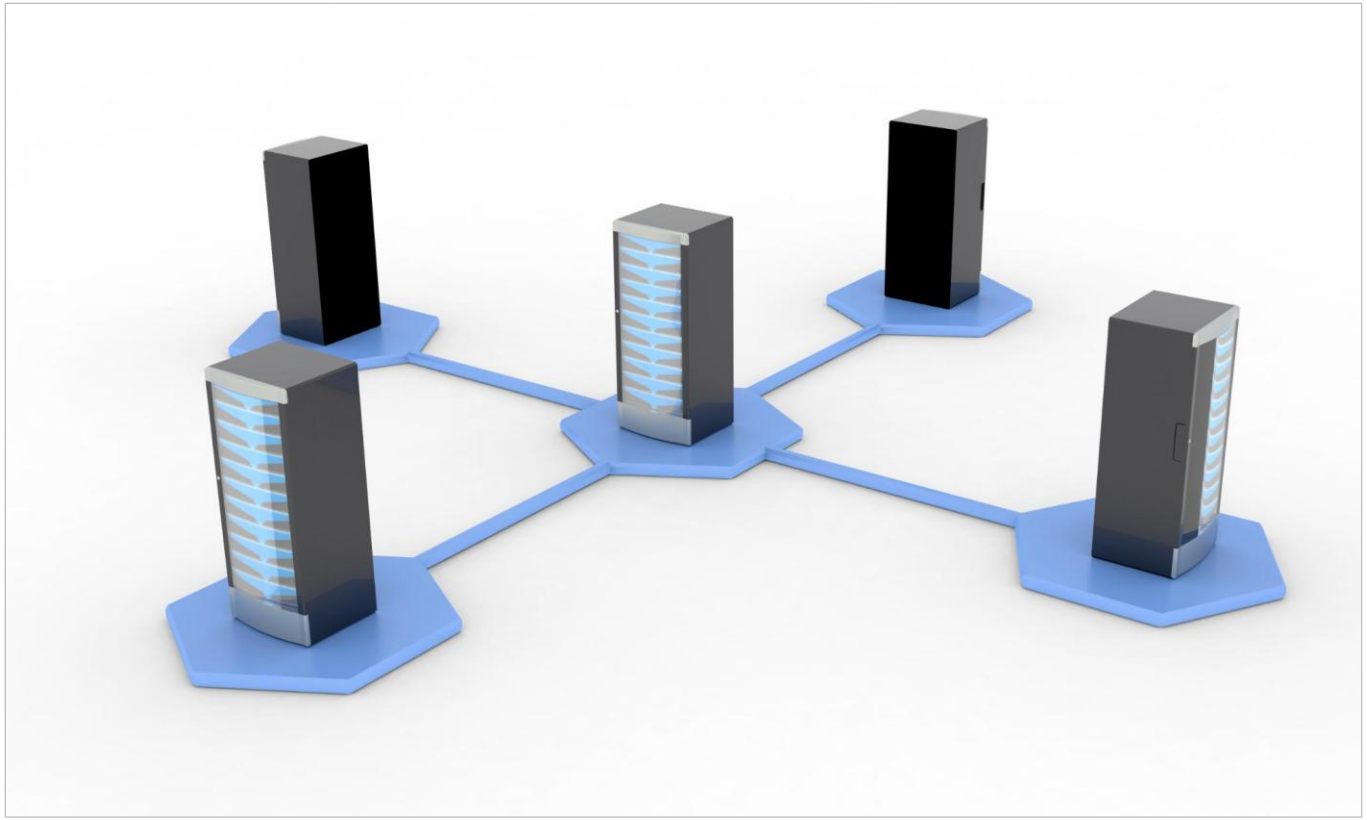
Unity Catalog

Introduction

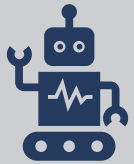
Data Governance: Four Key Functional Areas

- Data Access Control: Manage data access and permissions.
- Data Access Audit: Monitor and track data usage and changes.
- Data Lineage: Understand the origin and flow of data.
- Data Discovery: Find, understand, and use data assets.



Capabilities of Unity Catalog in Azure Databricks

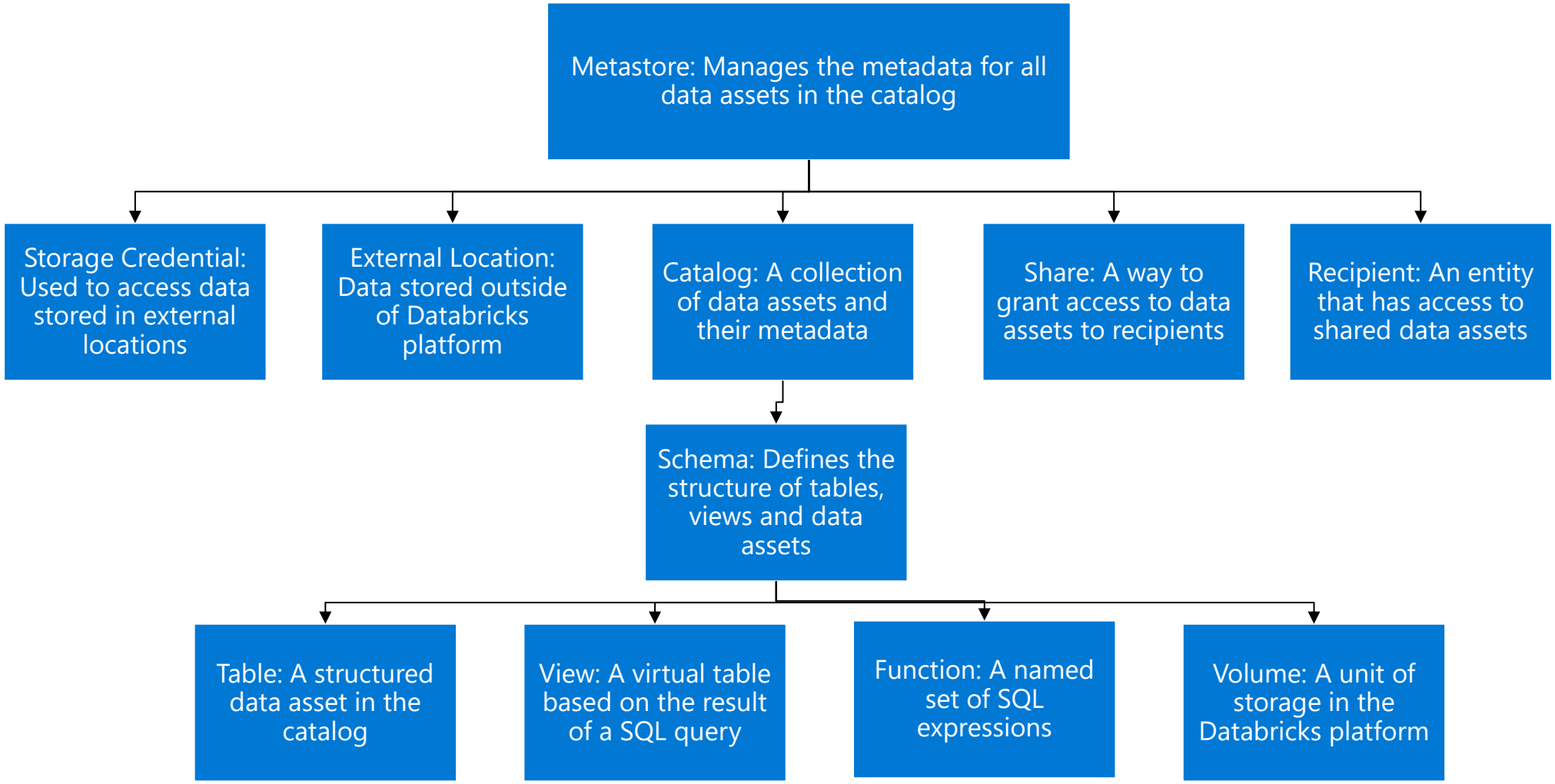
Provides unified governance for data access and security.



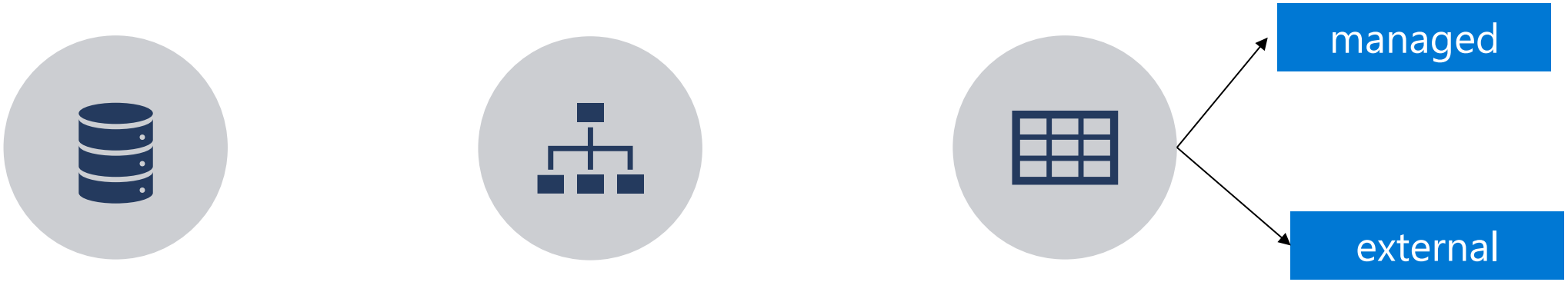
Enables centralized management of data and AI assets.

Allows integration and unification of existing data catalogs.

Unity Catalog Metastore Elements

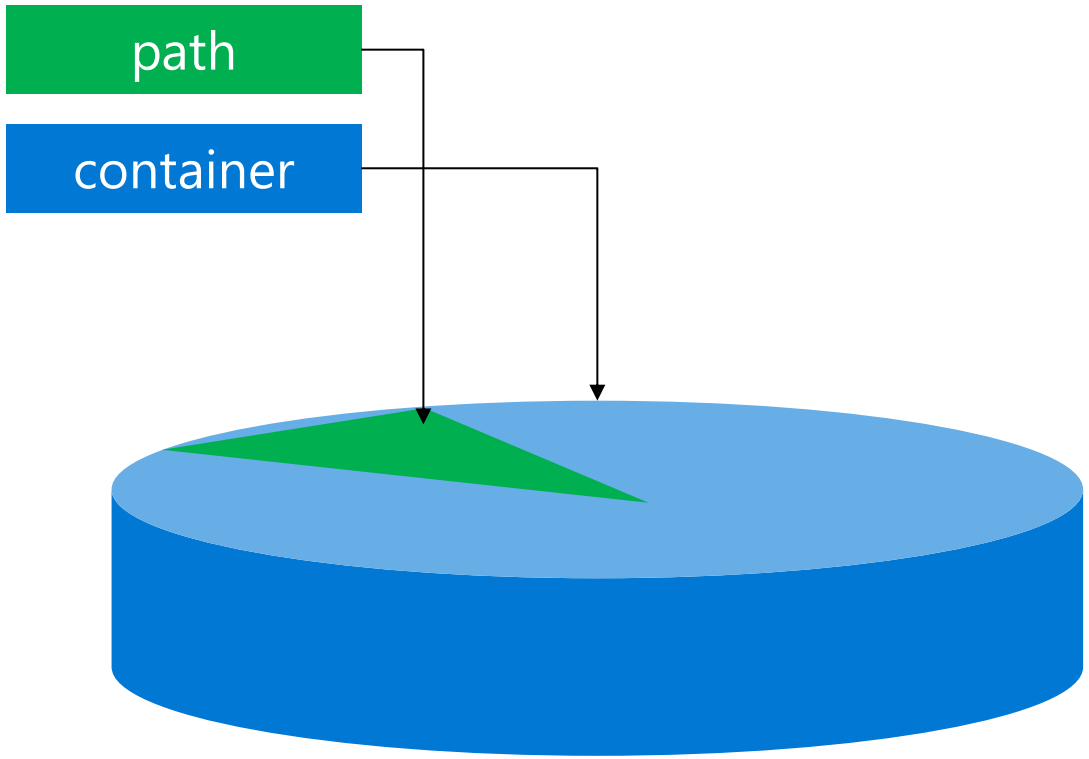


Three Level Namespace of Unity Catalog



Catalog . Schema . Table

Comparision between Storage Credentials and External Location



Storage Credentials allow access to an external location (ADLSG2, Blob Storage, Database etc)

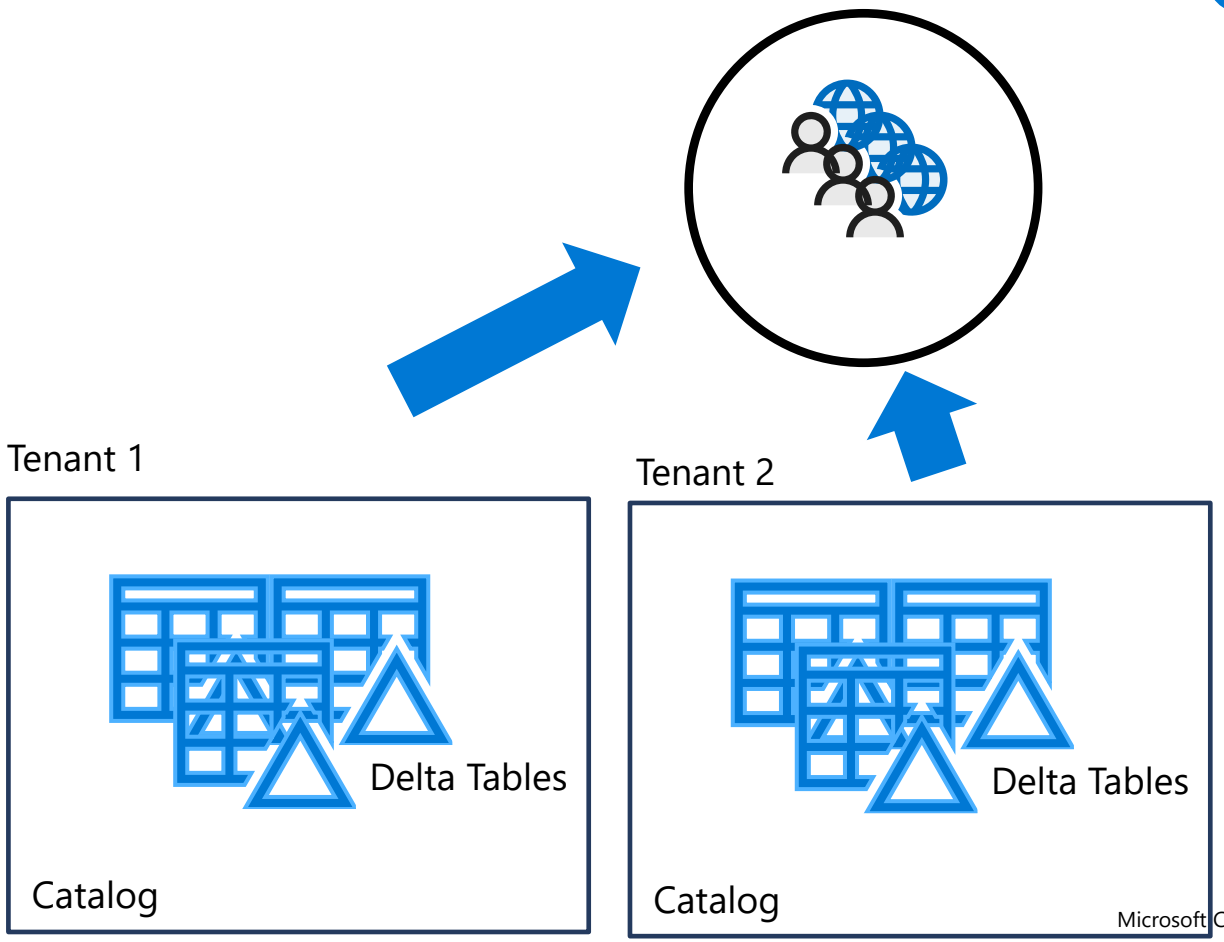
External Location can be a folder or file (more granular than Storage Credential)

Unity Catalog: Share and Recipient

Unity Catalog's Share feature allows users to grant data asset access to recipients.

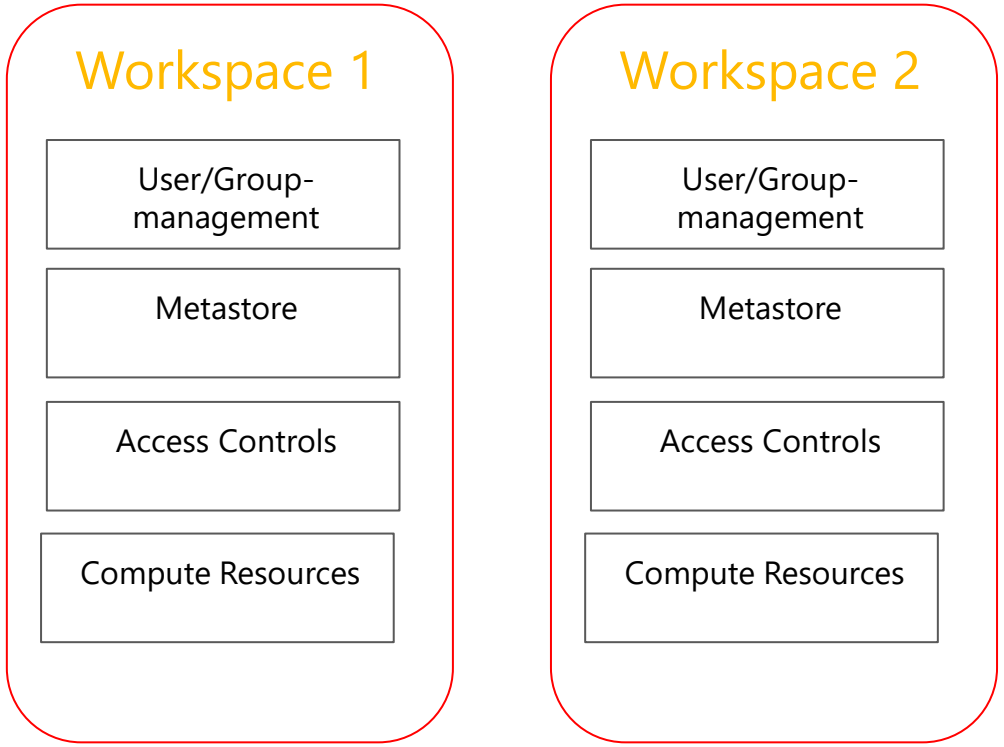
Recipients can be individuals, groups, or roles, allowing for granular access control.

The share feature ensures data governance and secure sharing of data assets.

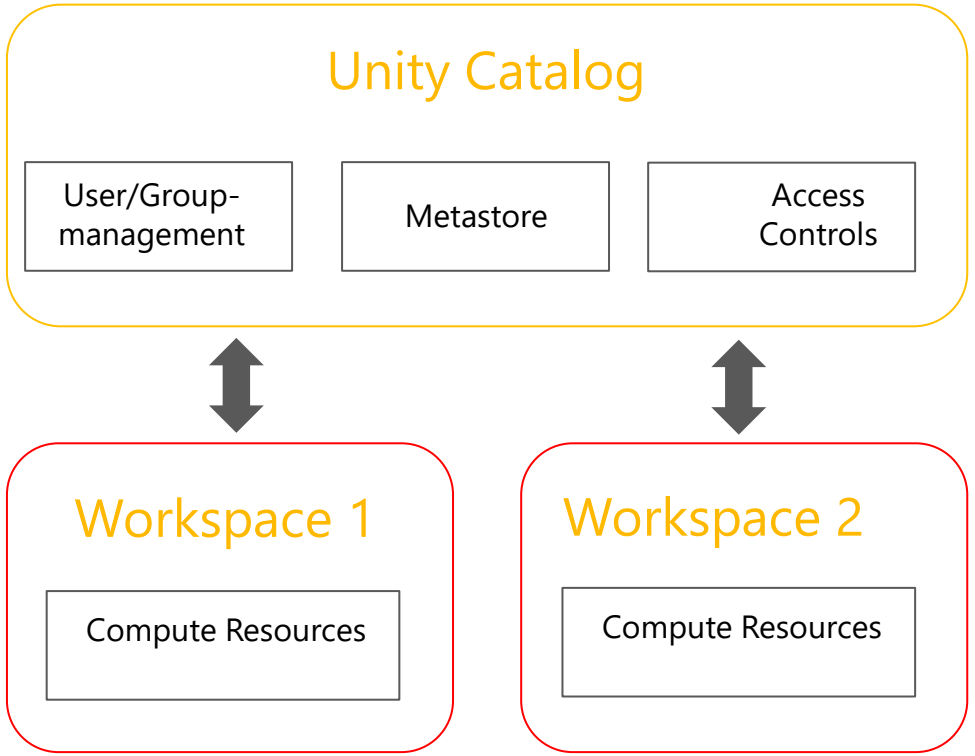


UC - Architecture

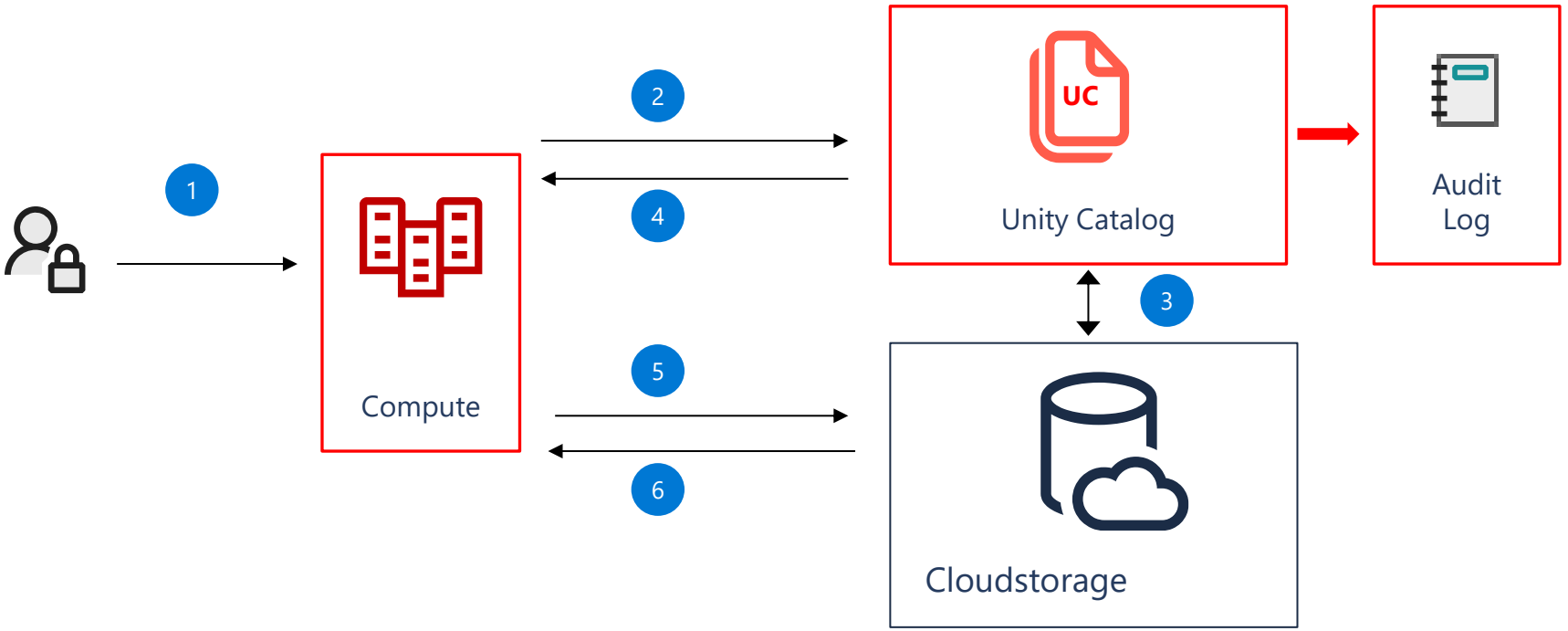
Before Unity Catalog



With Unity Catalog



Query Lifecycle in Unity Catalog



1. Send Query
2. Check Namespace, metadata and grants
3. Assume IAM Role or Service Principal
4. Return Short Lived Token
5. Request Data from URL with Short Lived Token
6. Return Data

Demo

Unity Catalog