

Data Governance with Unity Catalog



Databricks Academy

2023

Meet your instructor

◁Add Name>, ▷Add title▷



- Team: ▷Add team▷
- Time at Databricks: ▷Add time▷
- Fun fact: ▷Add fun fact▷

Meet your classmates

- Where is everyone joining us from today (city, country)?

Meet your classmates

- How long have you been working with Unity Catalog?

Meet your classmates

- What are you hoping to get out of this class?

Getting Started with the Course

Course goals

- 1 Describe fundamental concepts about using Unity Catalog for data governance
- 2 Describe data access patterns in Databricks
- 3 Define data access rules and manage data ownership
- 4 Secure access to external storage
- 5 Upgrade legacy data assets to Unity Catalog



Course topics

Data Governance with Unity Catalog

- Overview of Data Governance
- Unity Catalog Key Concepts
- Unity Catalog Architecture
- Unity Catalog Roles & Identities
- Data Administration Fundamentals
- Data Access Control
- External Data Storage
- Data Segregation



Overview of Data Governance

80% of organizations seeking to scale digital business **will fail** because they do not take a modern approach to **data and analytics governance**

Source: [Gartner](#)

Data Governance

Four key functional areas

Data Access Control

Control who has access to which data

Data Access Audit

Capture and record all access to data

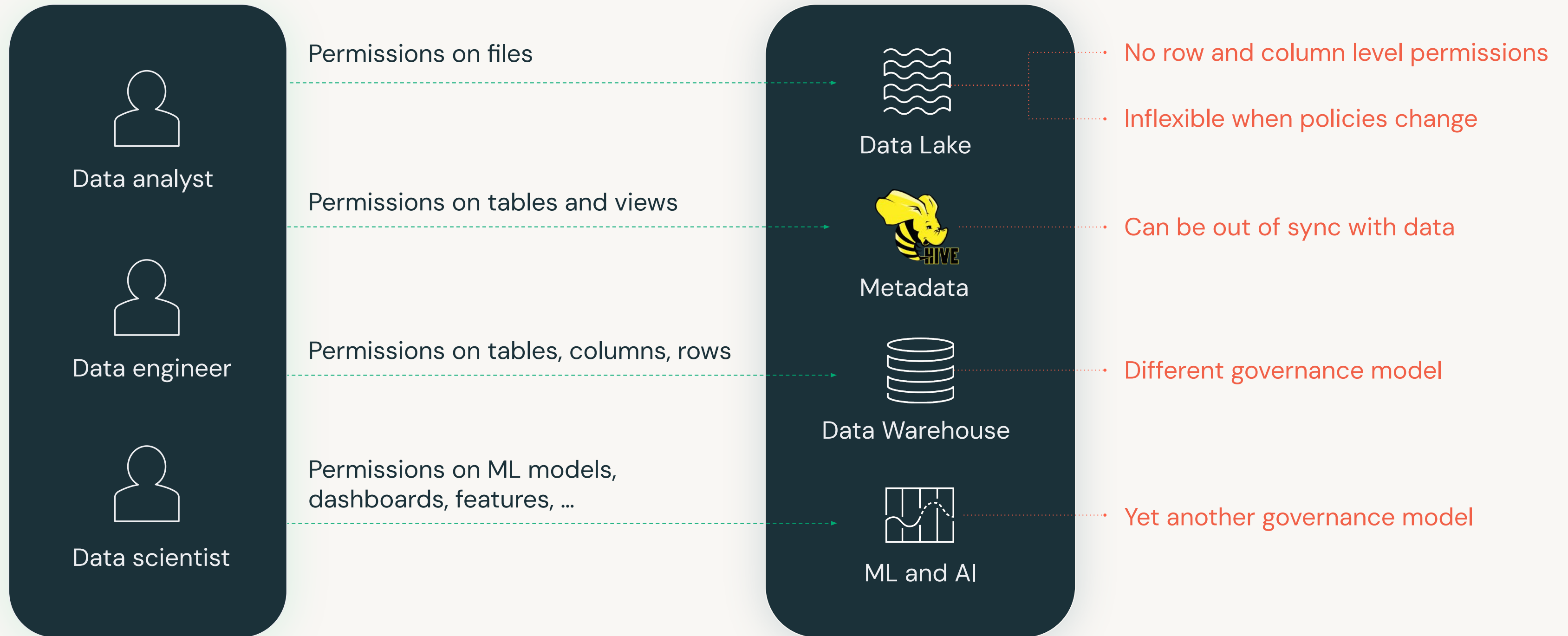
Data Lineage

Capture upstream sources and downstream consumers

Data Discovery

Ability to search for and discover authorized assets

Governance for data, analytics and AI is complex



Databricks Unity Catalog

Unified governance for data, analytics and AI



Unity Catalog

Overview



Unified governance across clouds

Fine-grained governance for data lakes across clouds – based on open standard ANSI SQL.

1



Unified data and AI assets

Centrally share, audit, secure and manage all data types with one simple interface.

2



Unified existing catalogs

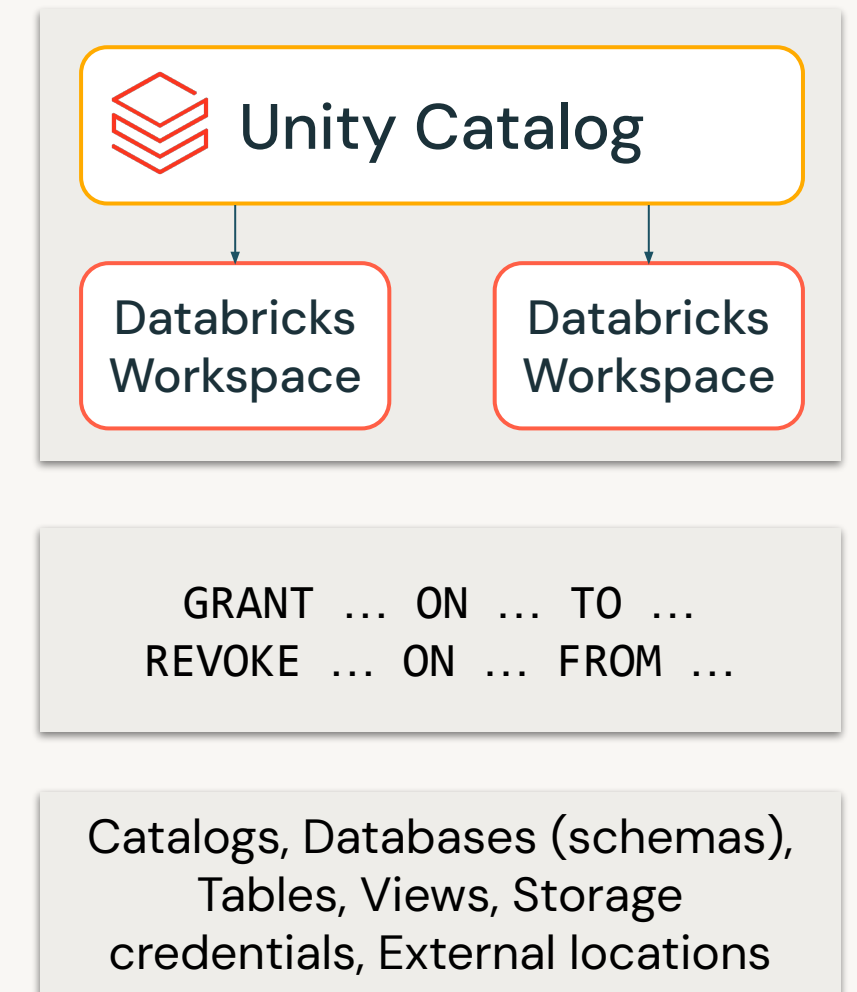
Works in concert with existing data, storage, and catalogs – no hard migration required.

3

Unity Catalog

Key Capabilities

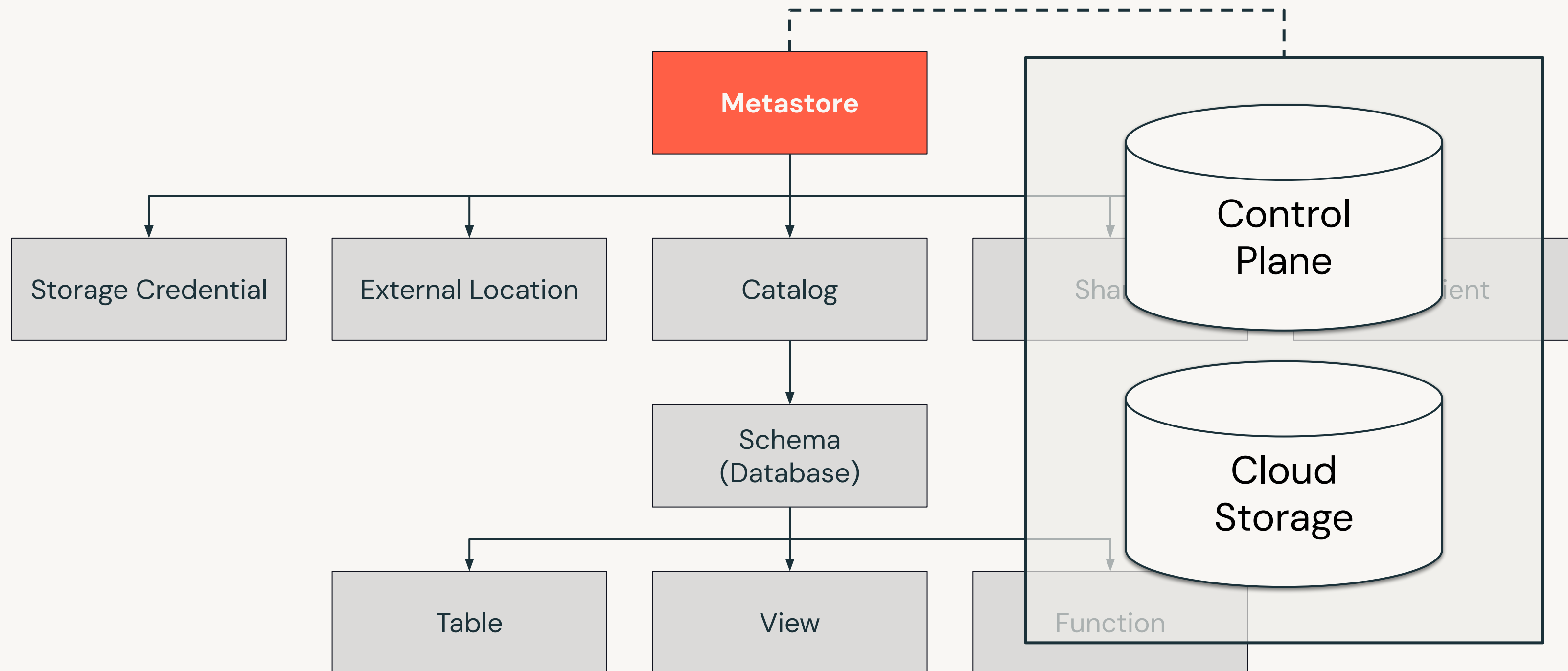
- Centralized metadata and user management
- Centralized data access controls
- Data access auditing
- Data lineage
- Data search and discovery
- Secure data sharing with Delta Sharing



Unity Catalog Key Concepts

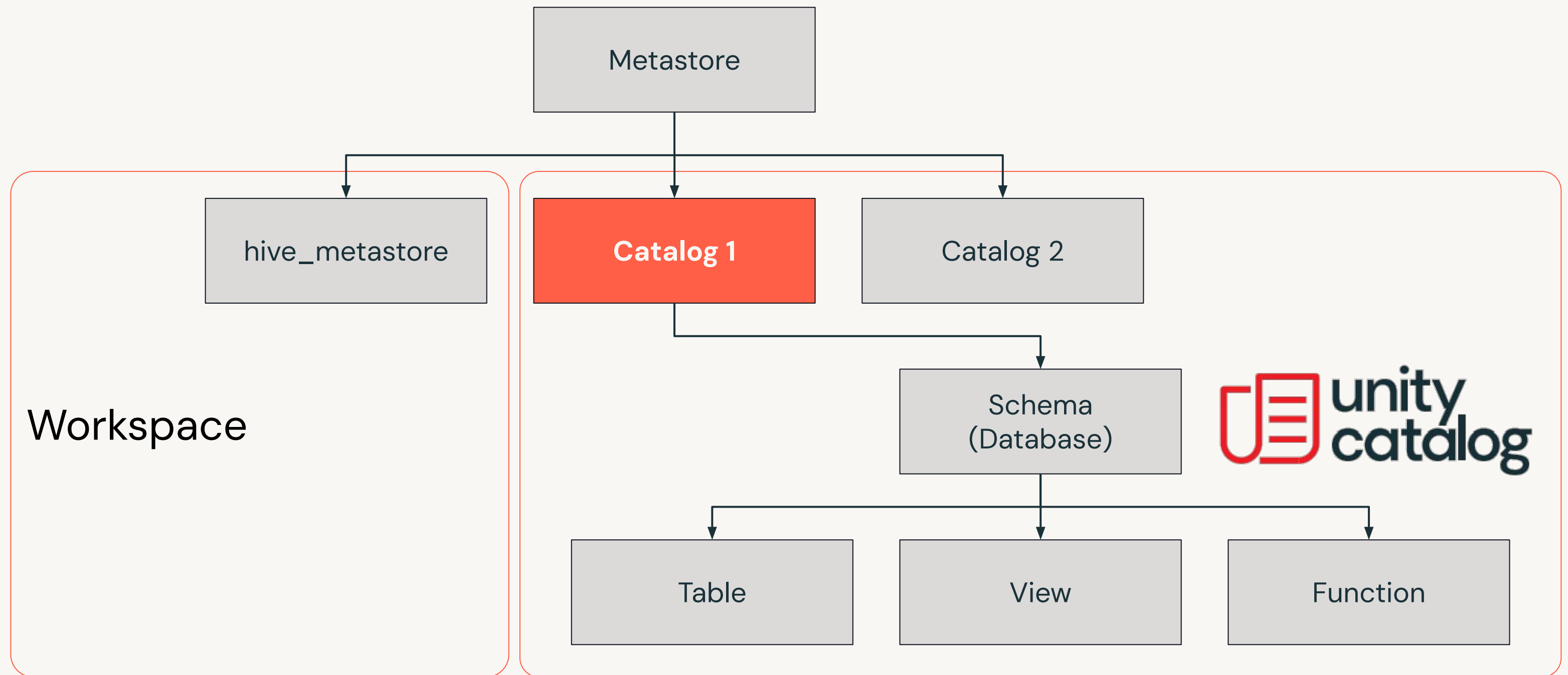
Metastore

Unity Catalog metastore elements



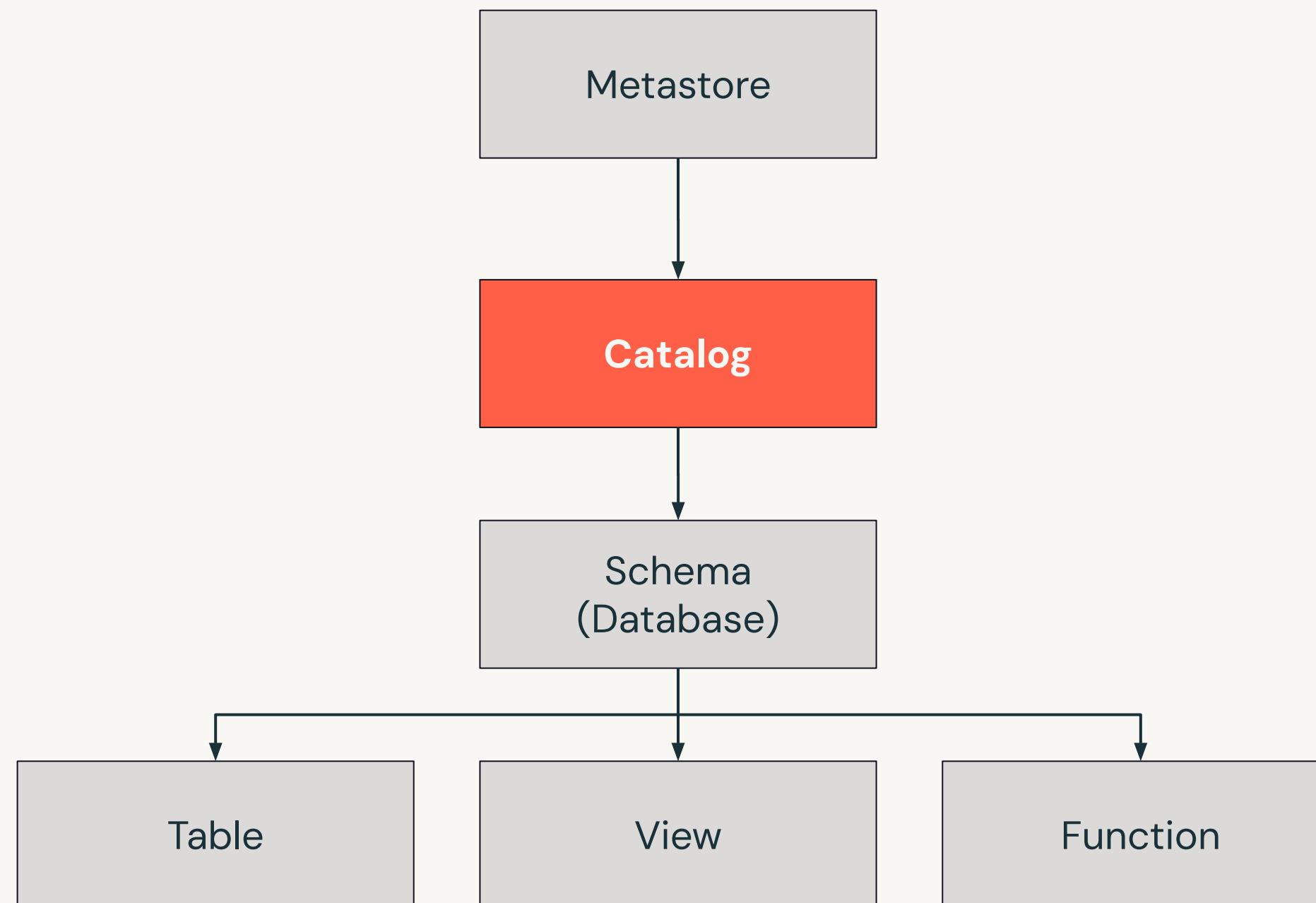
Metastore

Accessing legacy Hive metastore



Catalog

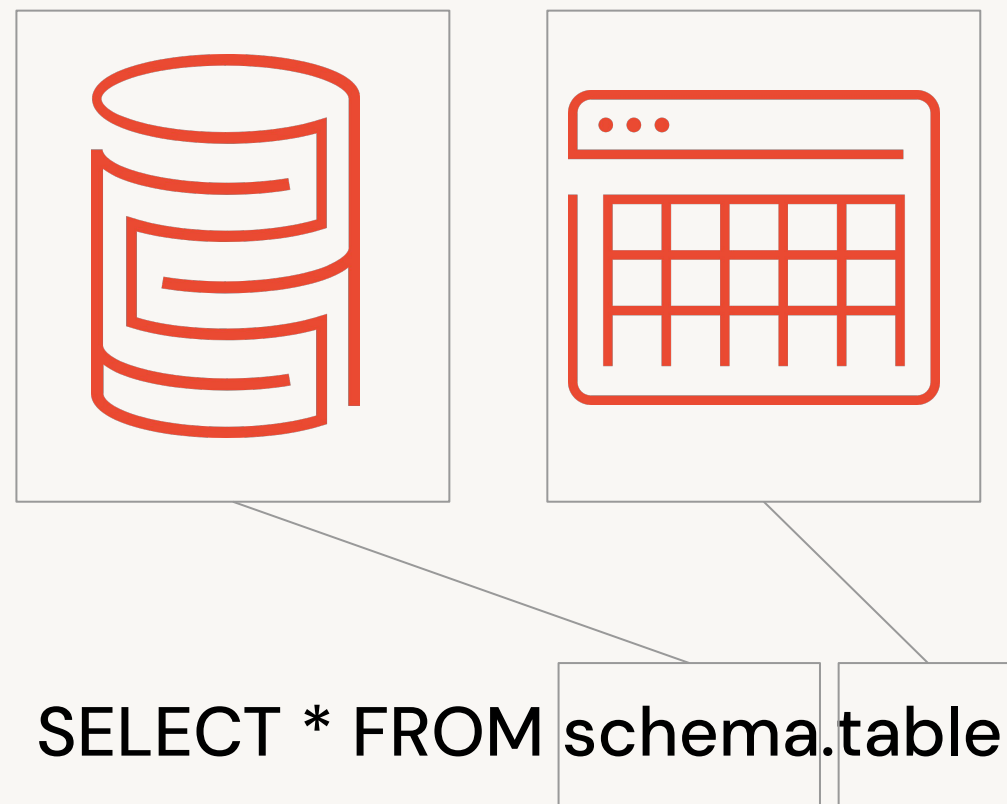
Top-level container for data objects



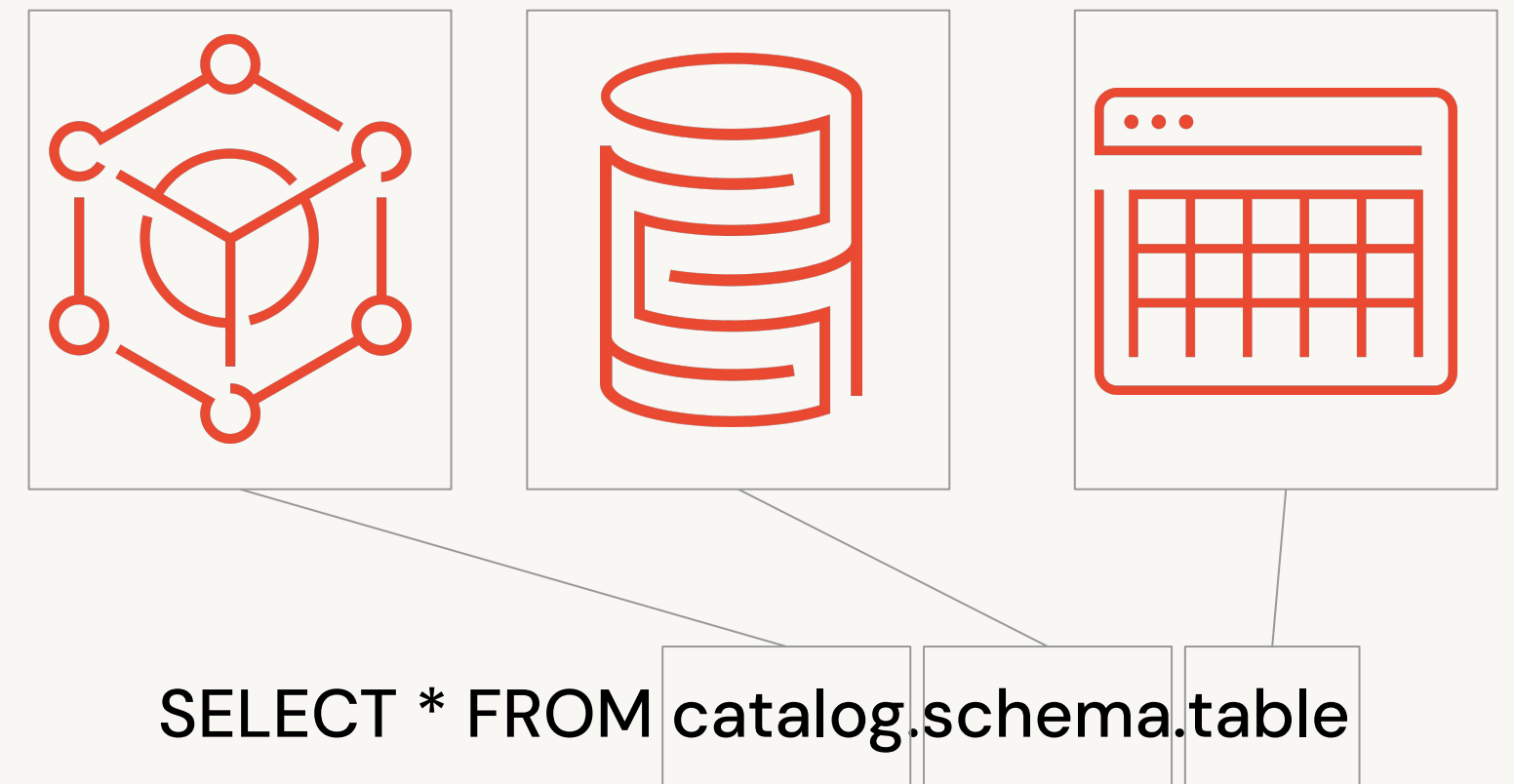
Catalog

Three-level namespace

Traditional SQL two-level namespace

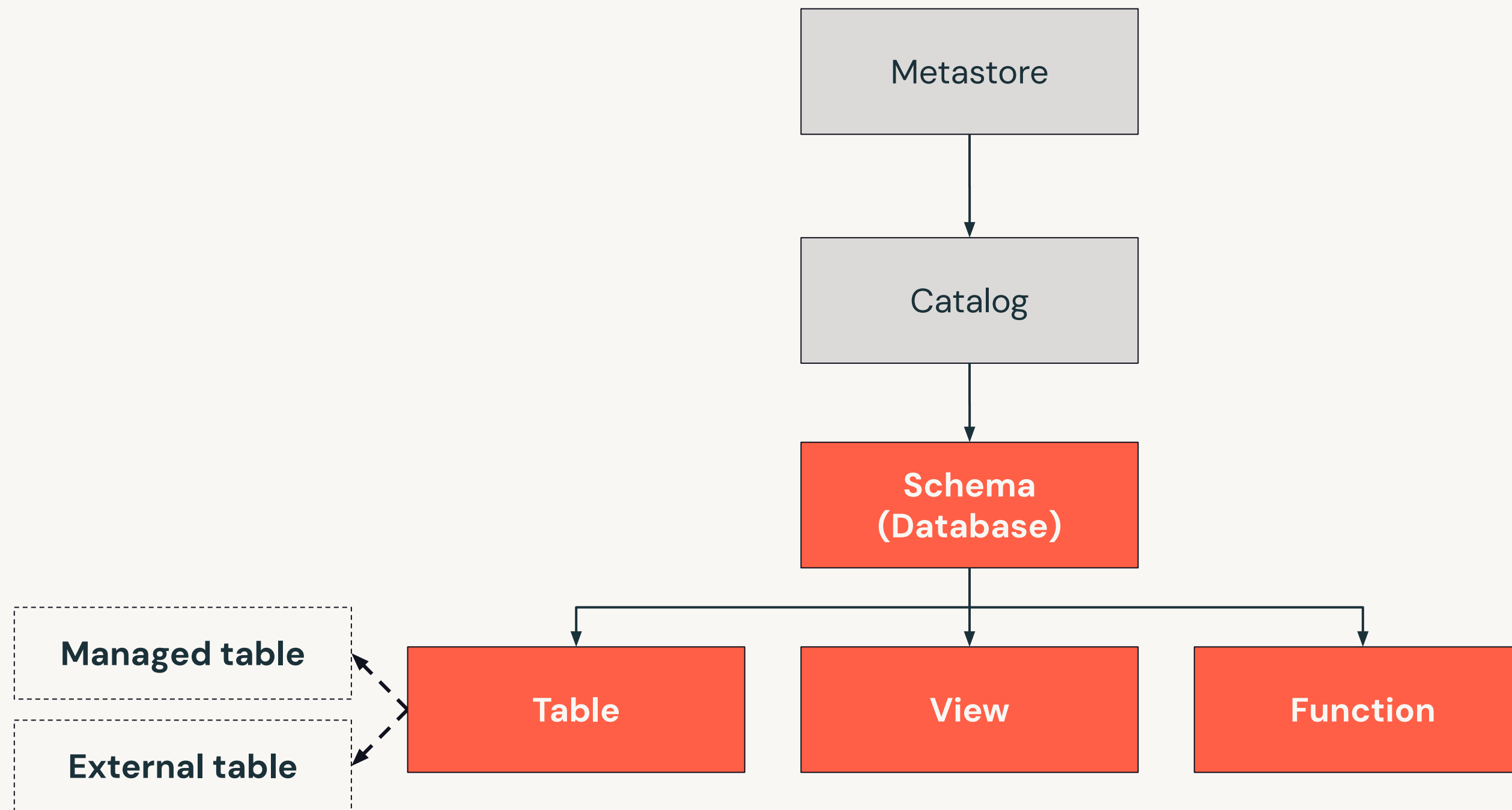


Unity Catalog three-level namespace



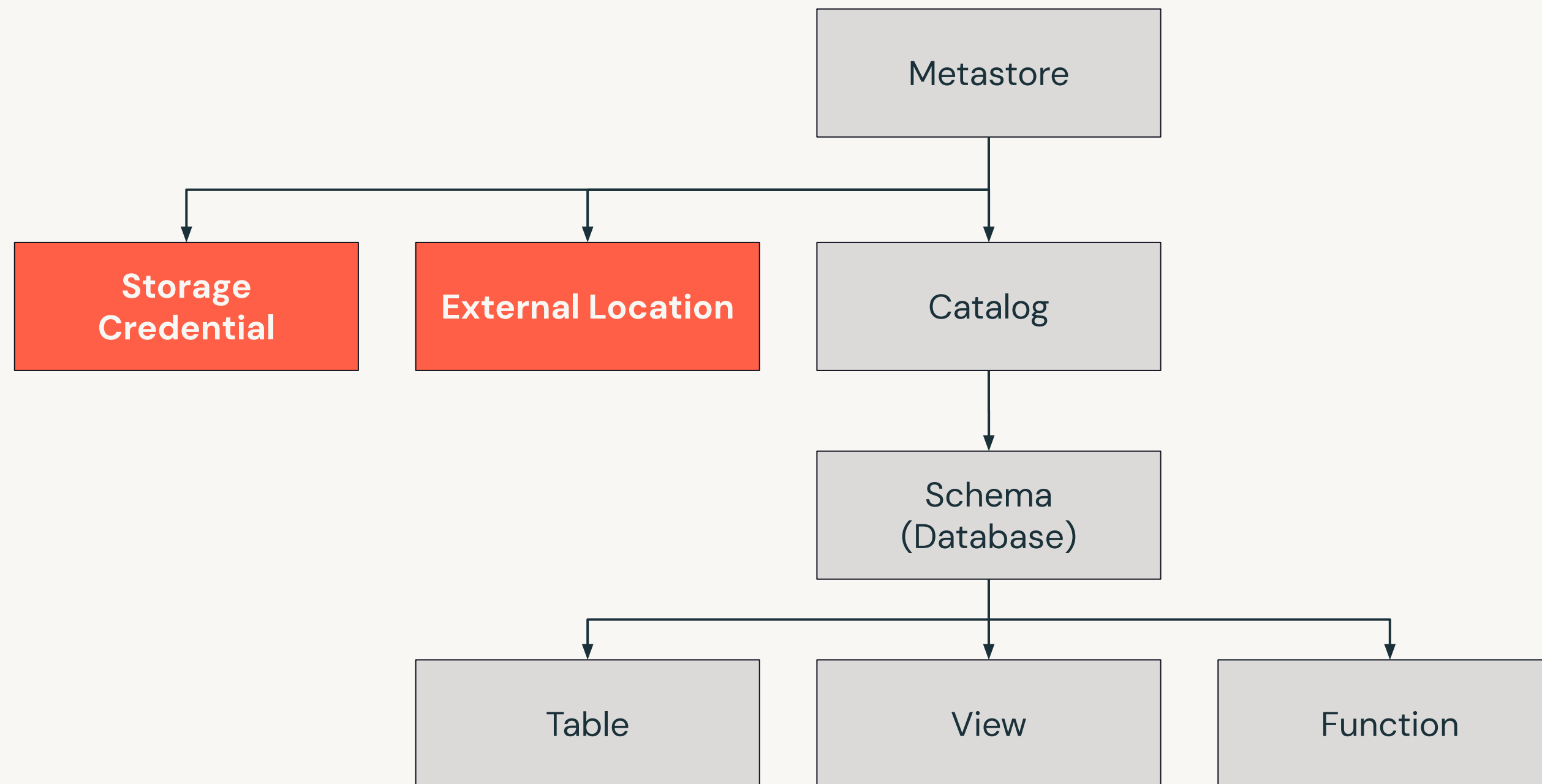
Data Objects

Schema (database), tables, views, functions



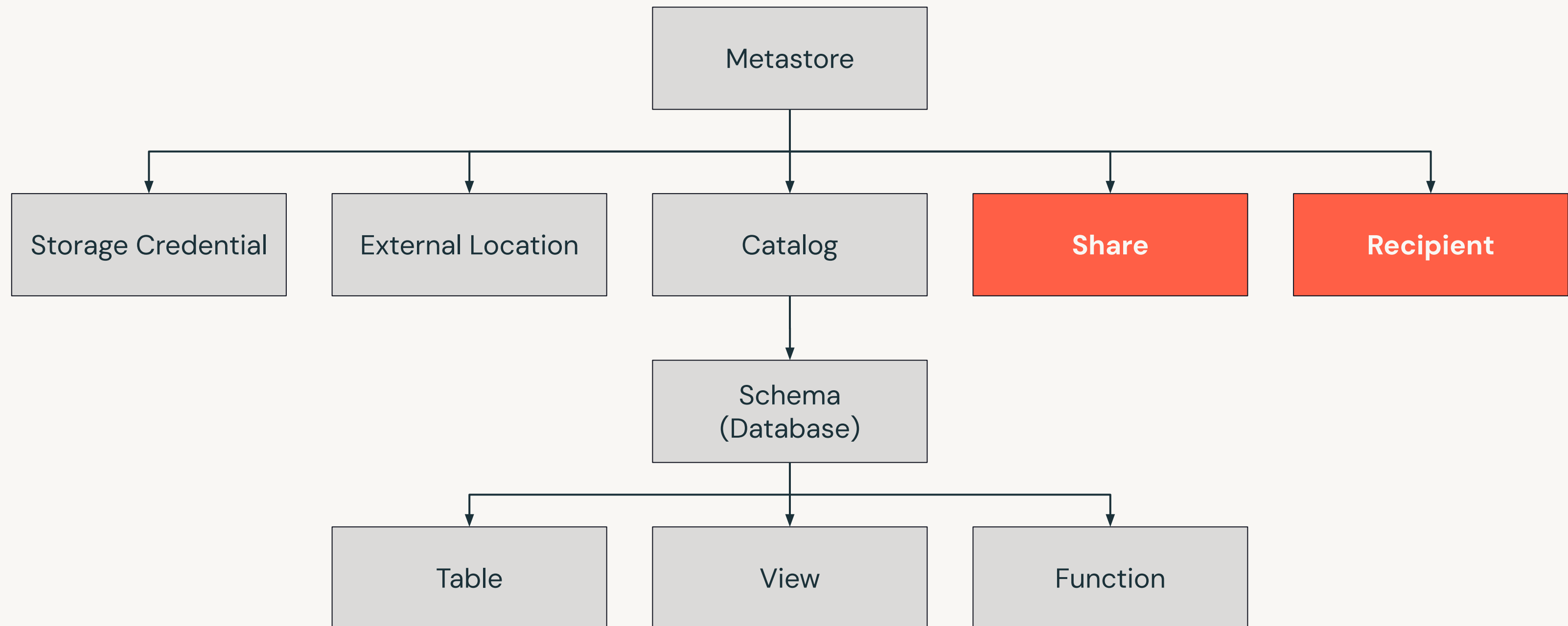
External Storage

Storage credentials and external locations



Delta Sharing

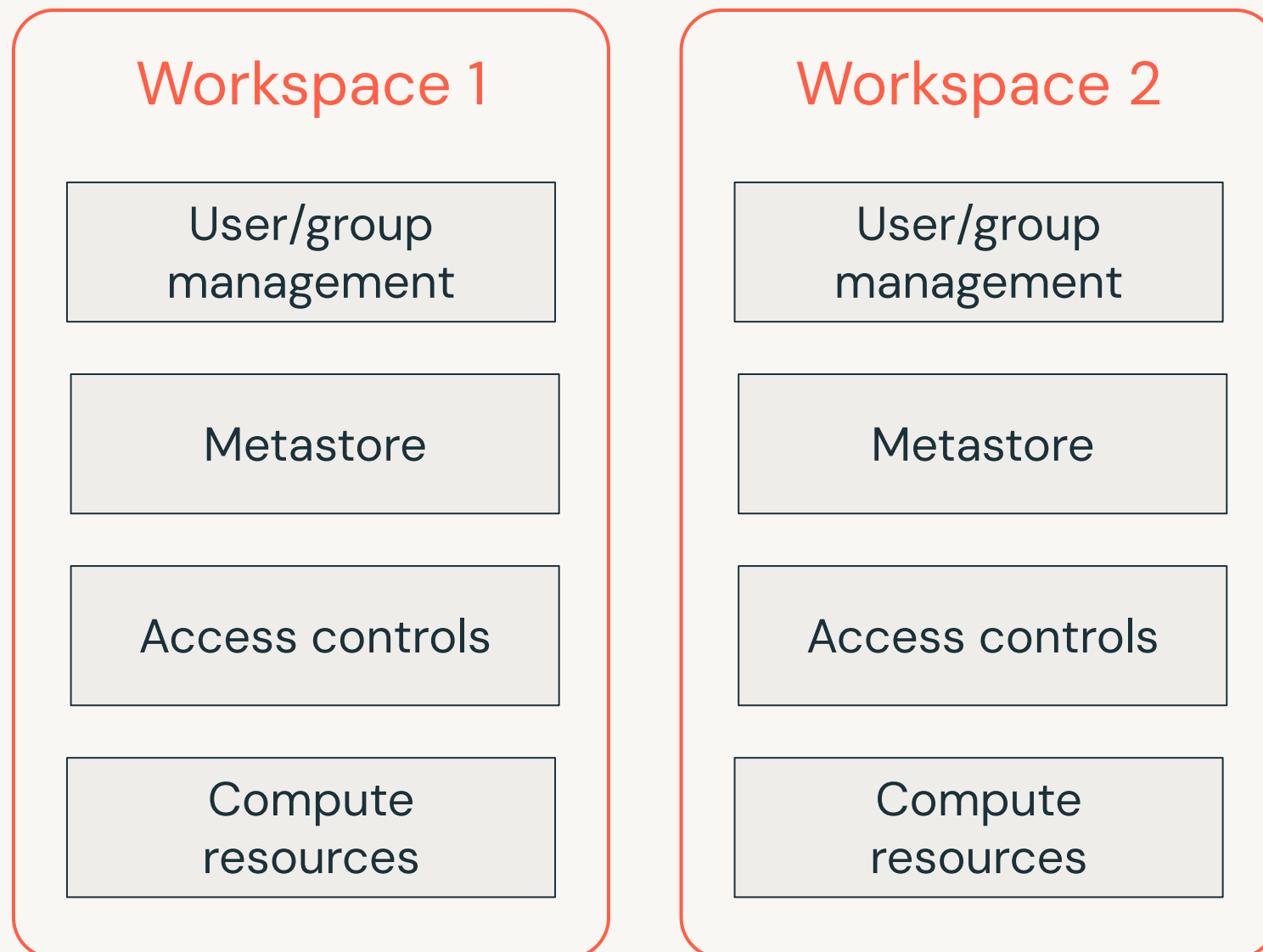
Shares and recipients



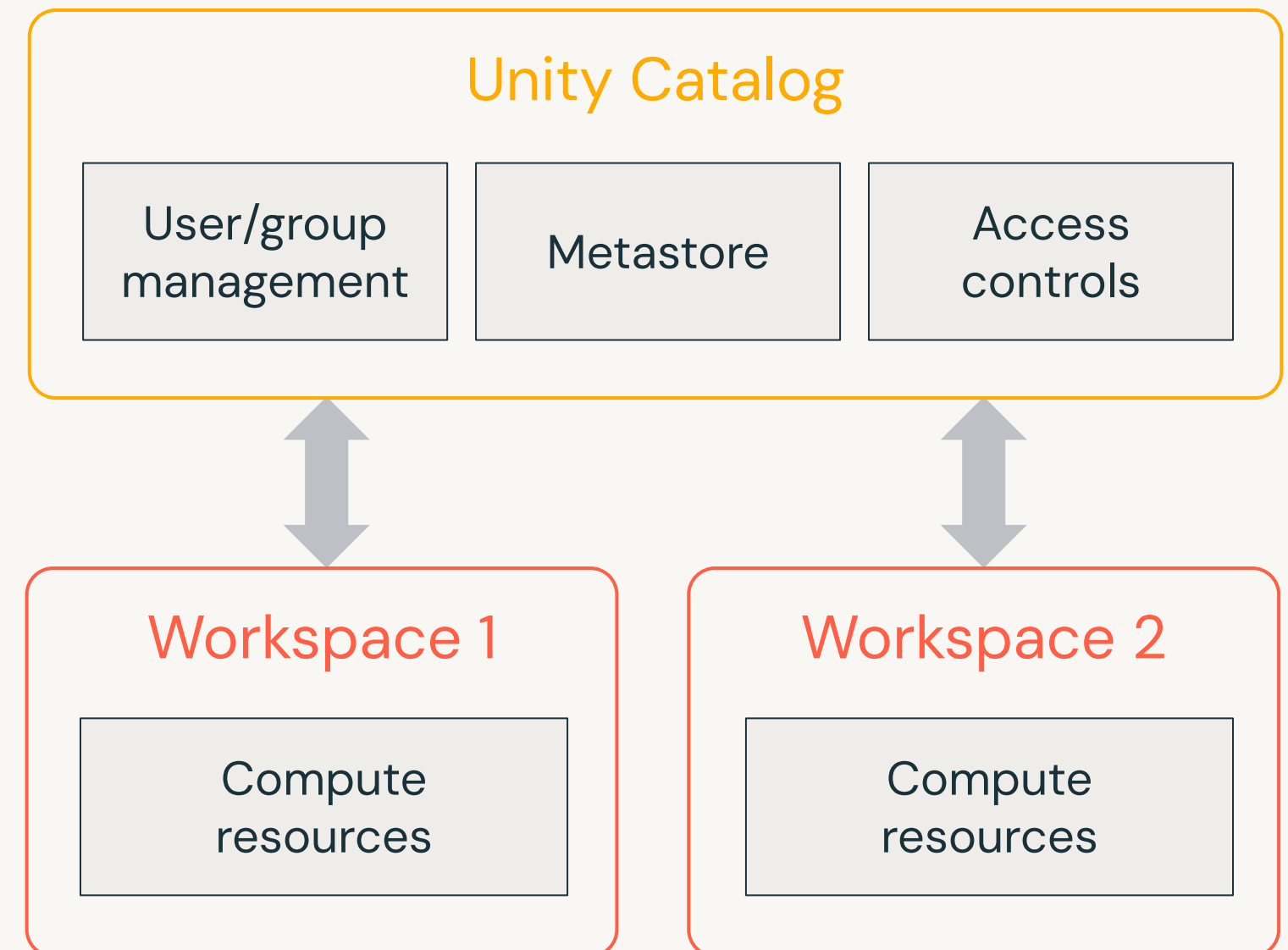
Unity Catalog Architecture

Architecture

Before Unity Catalog

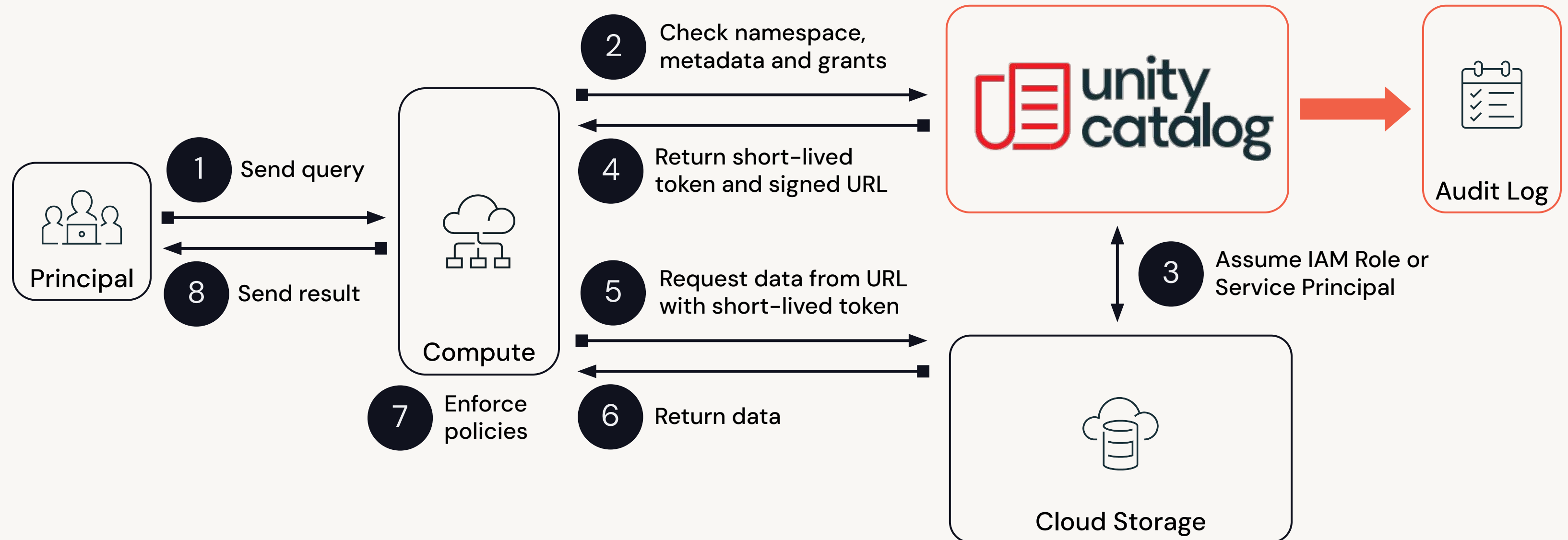


With Unity Catalog



Query Lifecycle

Unity Catalog Security Model



Roles and Identities in Unity Catalog

Unity Catalog

Roles

Cloud Admin

Identity Admin

Account Admin

Metastore Admin

Data Owner

Workspace Admin

Cloud Admin

- Manage underlying cloud resources
 - Storage accounts/buckets
 - IAM role/service principals/managed identities

Identity Admin

- Manage users and groups in the identity provider (IdP)
- Provision into account (with account admin)

Unity Catalog

Roles

Cloud Admin

Identity Admin

Account Admin

Metastore Admin

Data Owner

Workspace Admin

Account Admin

- Create or delete metastores, assign metastores to workspaces
- Manage users and groups, integrate with IdP
- Full access to all data objects

Metastore Admin

- Create or drop, grant privileges on, and change ownership of catalogs and other data objects

Data Owner – owns data objects they created

- Create nested objects, grant privileges on, and change ownership of owned objects

Unity Catalog

Roles

Cloud Admin

Identity Admin

Account Admin

Metastore Admin

Data Owner

Workspace Admin

Workspace Admin

- Manages permissions on workspace assets
- Restricts access to cluster creation
- Adds or removes users
- Elevates users permissions
- Grant privileges to others
- Change job ownership

Unity Catalog

Identities

- User
- Account Administrator
- Service Principal
- Service Principal with administrative privileges

user01@domain.com

First name

First name

Last name

Last name

Password

●●●●●●●●

Admin role

☒

terraform

App ID

UUID

Name

terraform

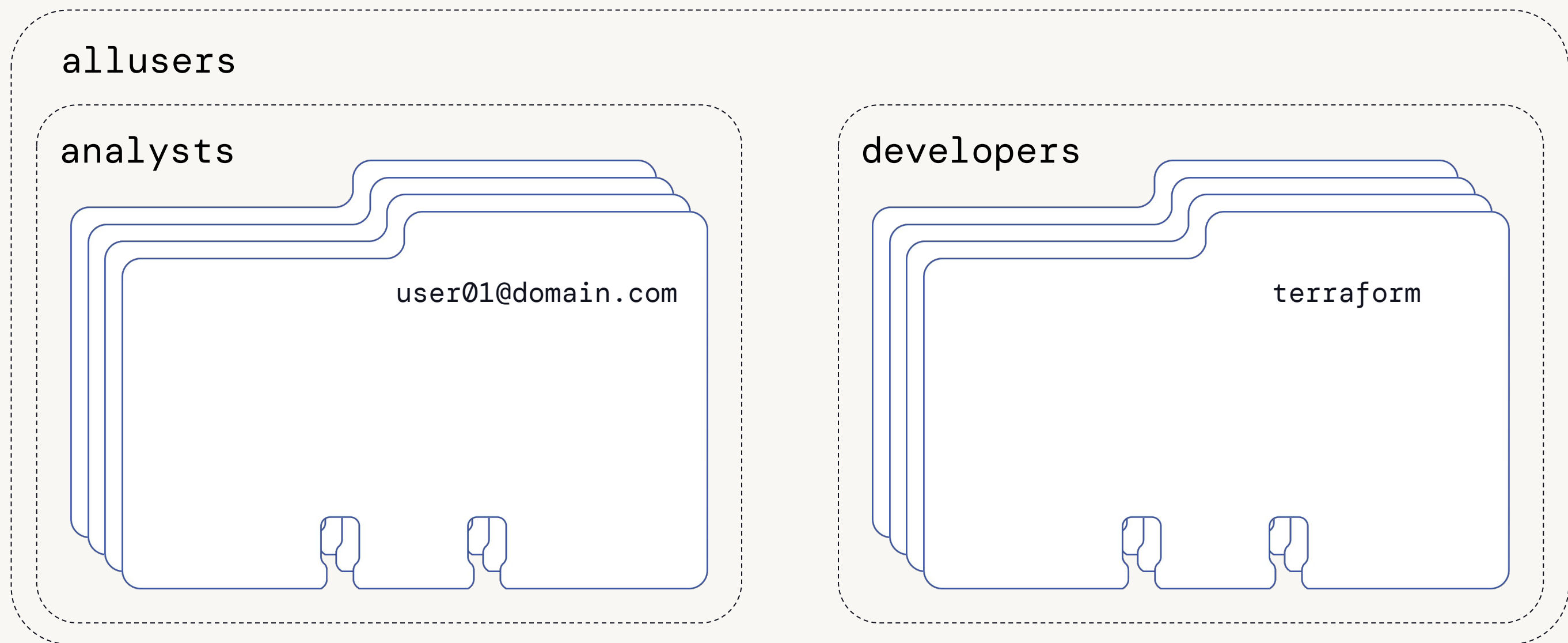
Admin role

☒

Unity Catalog

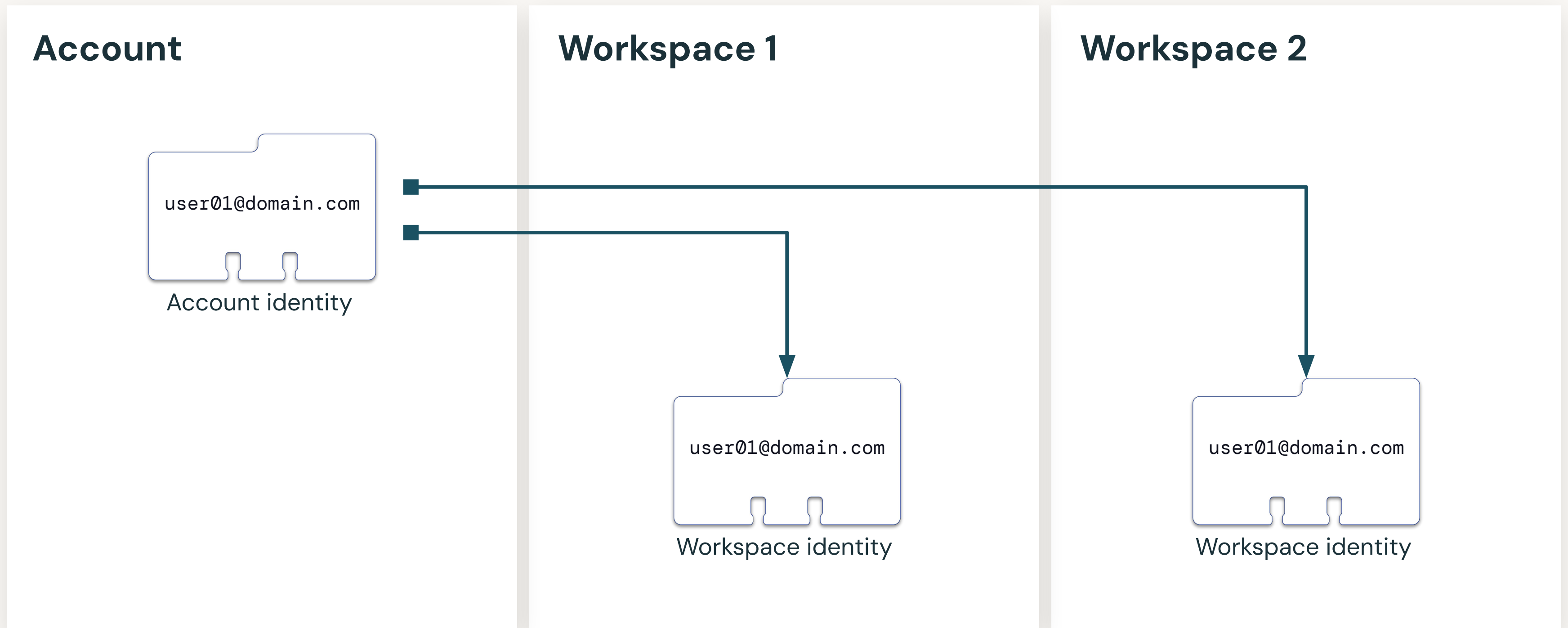
Identities

- Groups



Unity Catalog

Identity Federation





Data Administration Fundamentals



Lab: Populating the Metastore



Learning Objectives

In this lab, you will learn how to:

- Use the Unity Catalog three-level namespace
- Use SQL to create and explore data objects
- Use the Data explorer to explore data objects and lineage
-



Data Security Model



Data Security Model

Access control lists (ACLs)



Data Security Model

Access control lists (ACLs)

Table **t**



Data Security Model

Access control lists (ACLs)

Table **t**

| Principal | Privilege |
|-----------------|---------------|
| analysts | SELECT |
| | |



Data Security Model

Access control lists (ACLs)

Table `t`

| Principal | Privilege |
|-----------------|-----------------------|
| analysts | SELECT |
| dbadmins | ALL PRIVILEGES |
| | |



Data Security Model

Access control lists (ACLs)

Table `t`

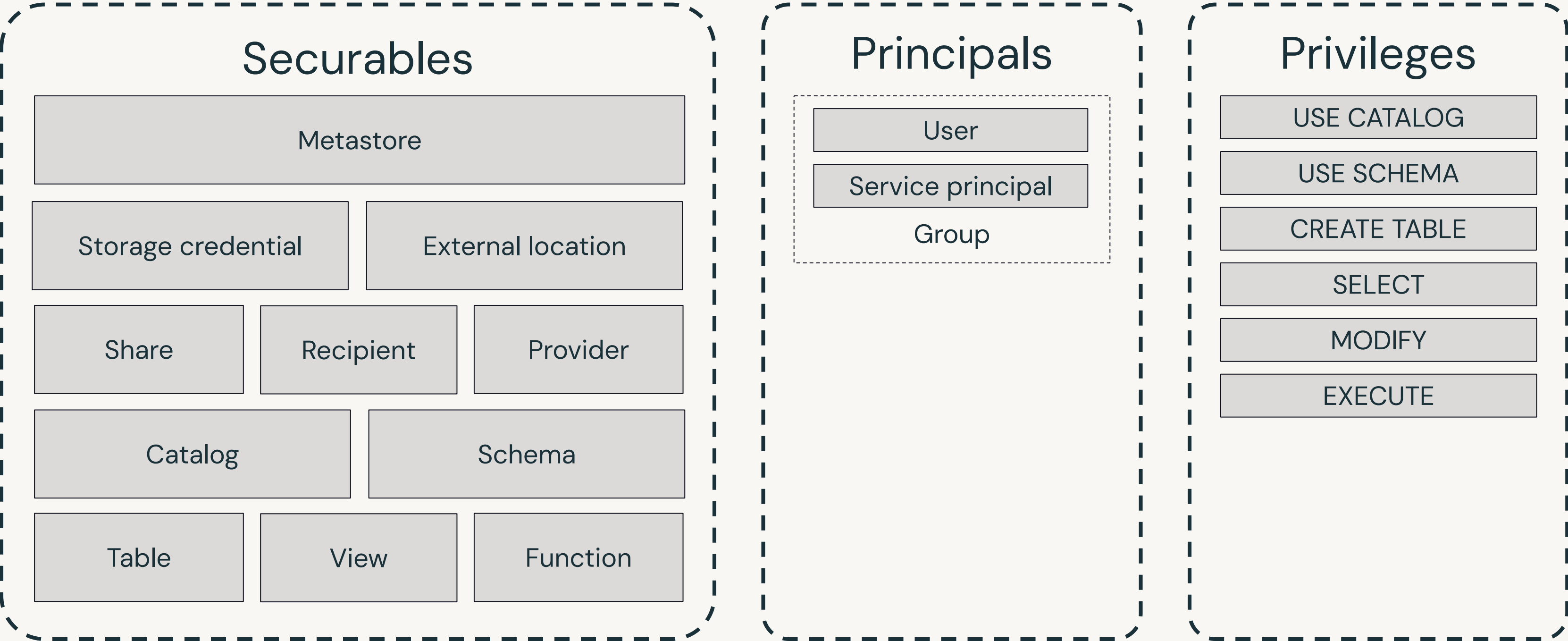
Owned by `jdope@company.com`

| Principal | Privilege |
|-----------|----------------|
| analysts | SELECT |
| dbadmins | ALL PRIVILEGES |
| | |



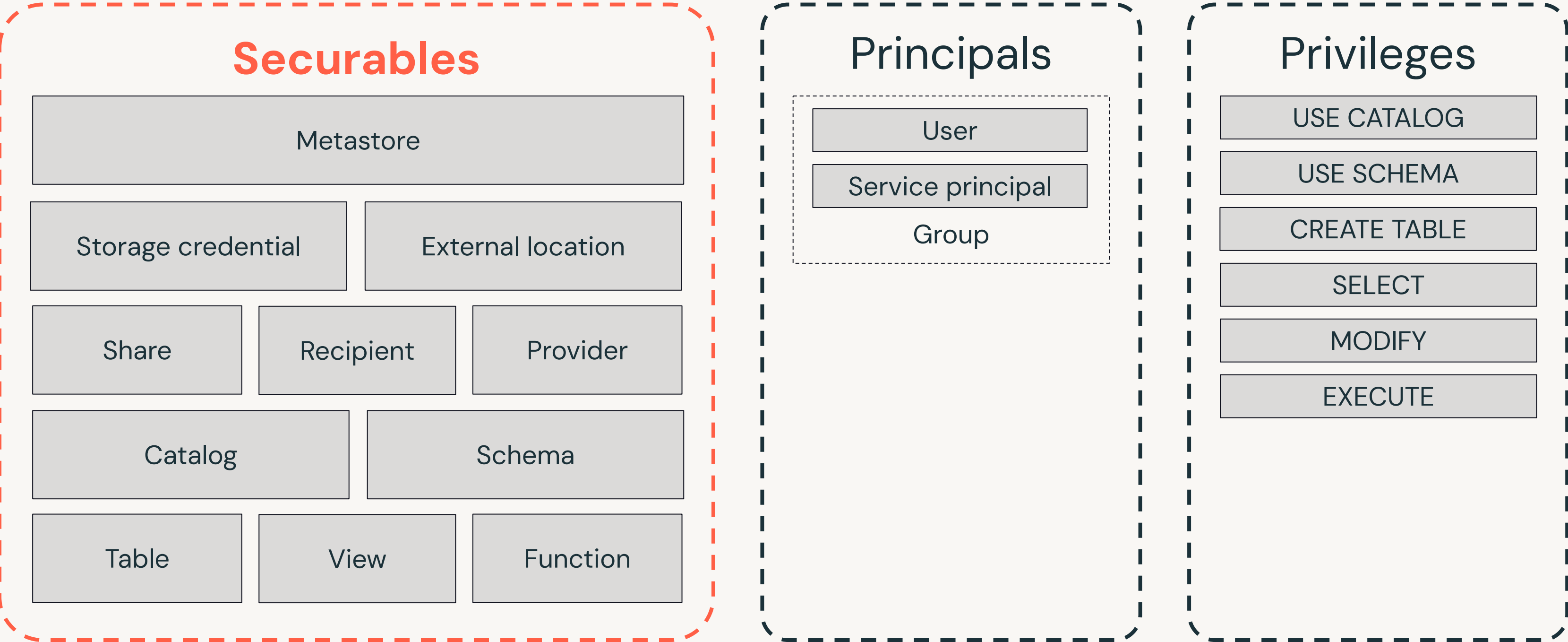
Data Security Model

Access control lists (ACLs)



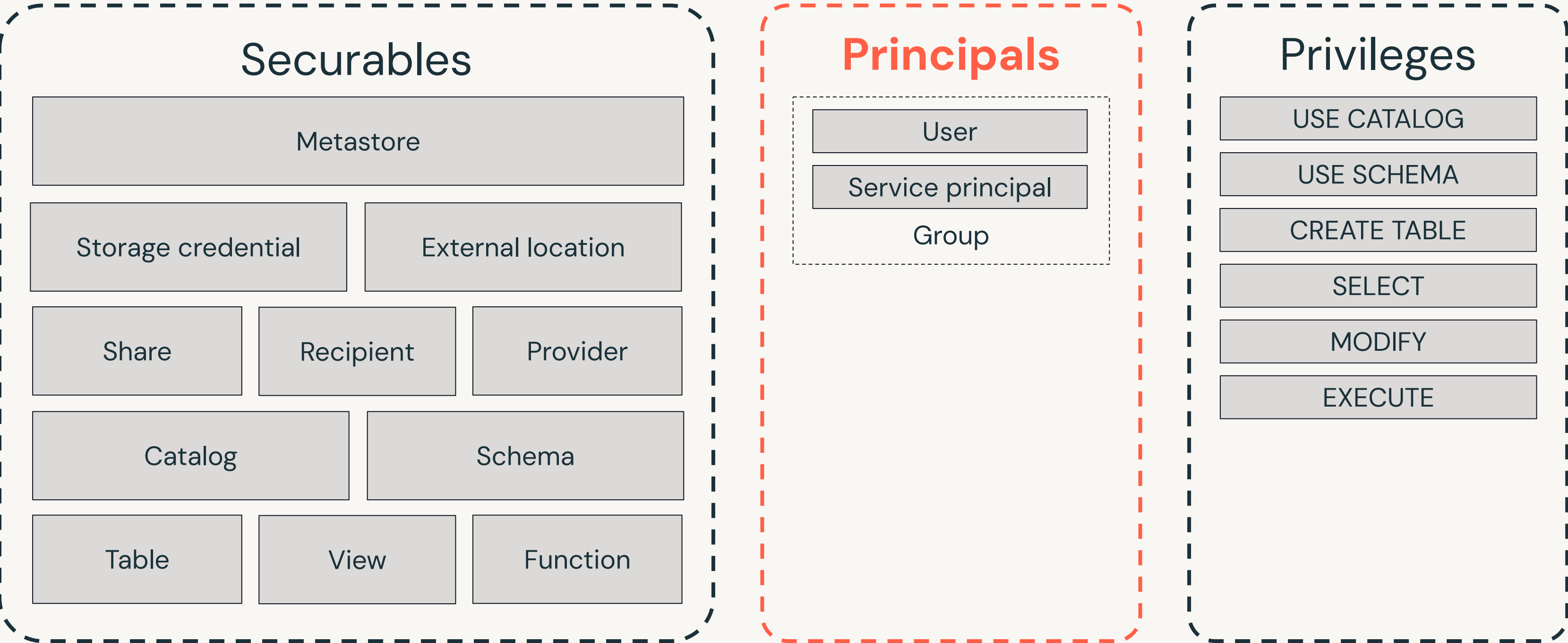
Data Security Model

Access control lists (ACLs)



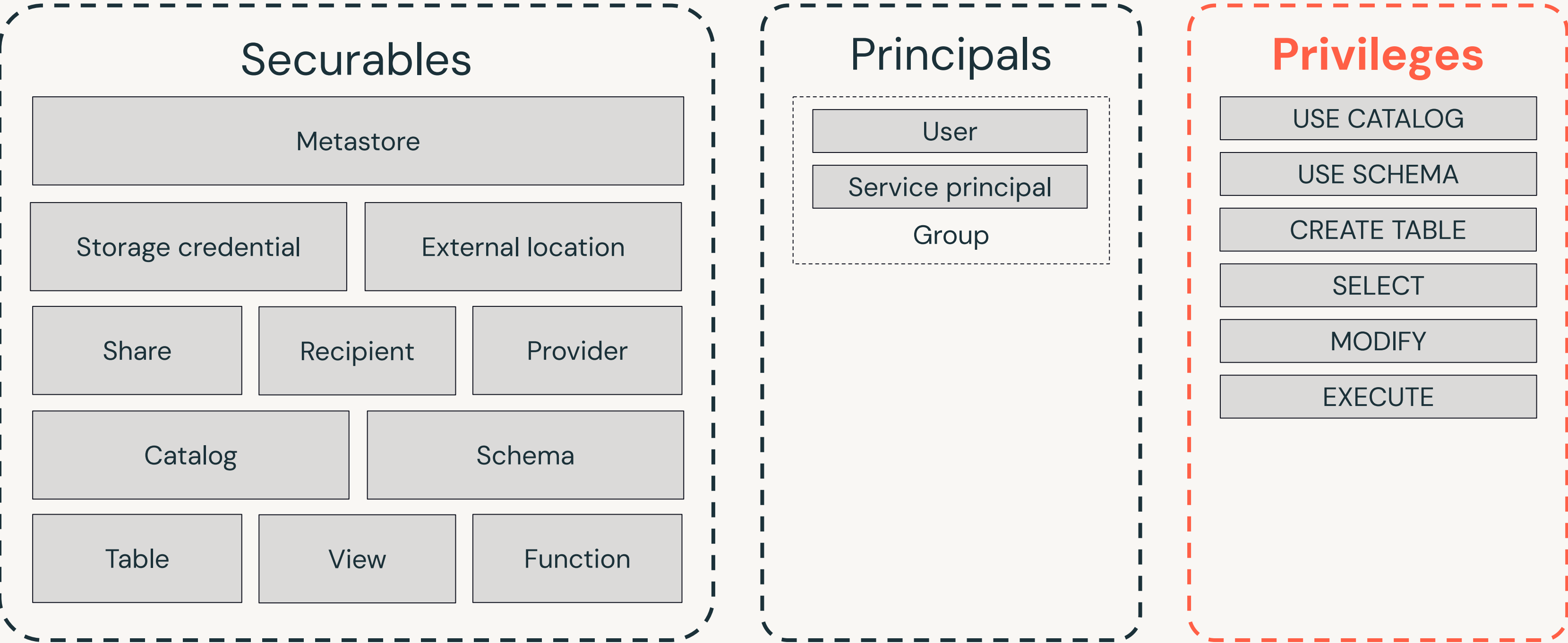
Data Security Model

Access control lists (ACLs)



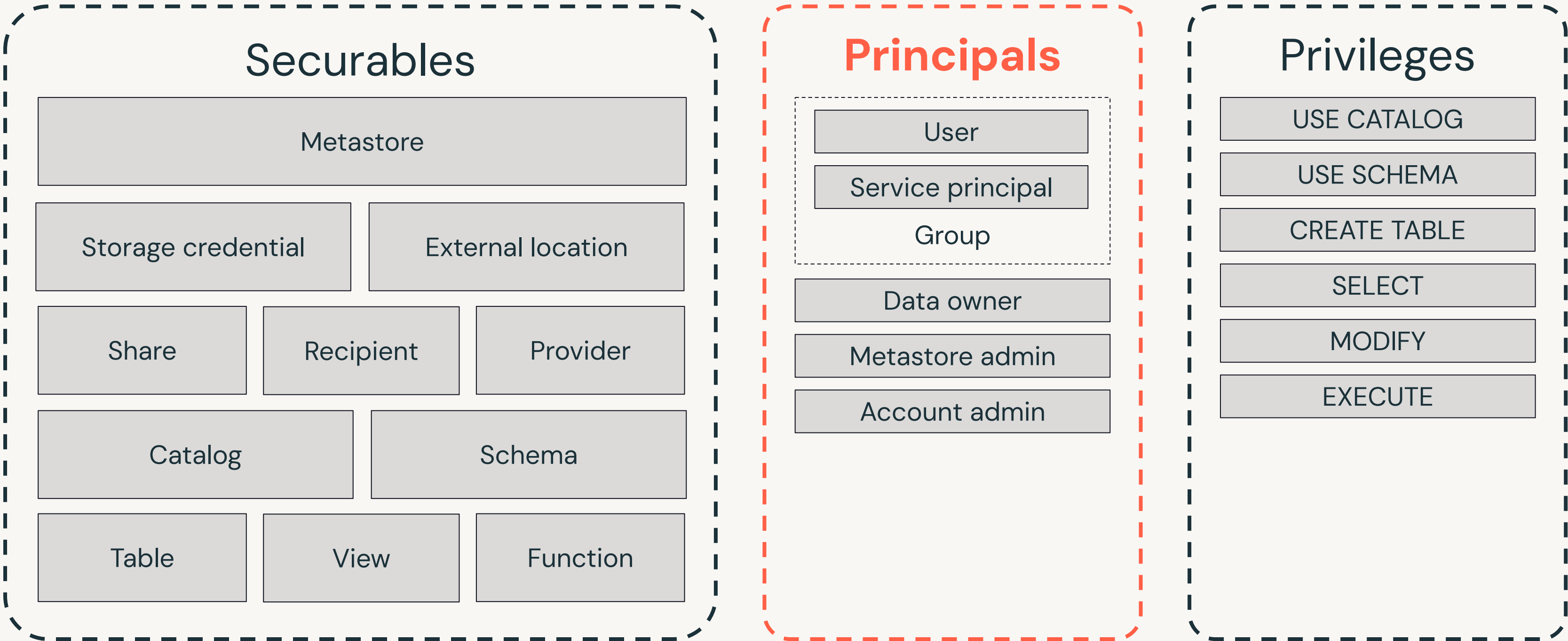
Data Security Model

Access control lists (ACLs)



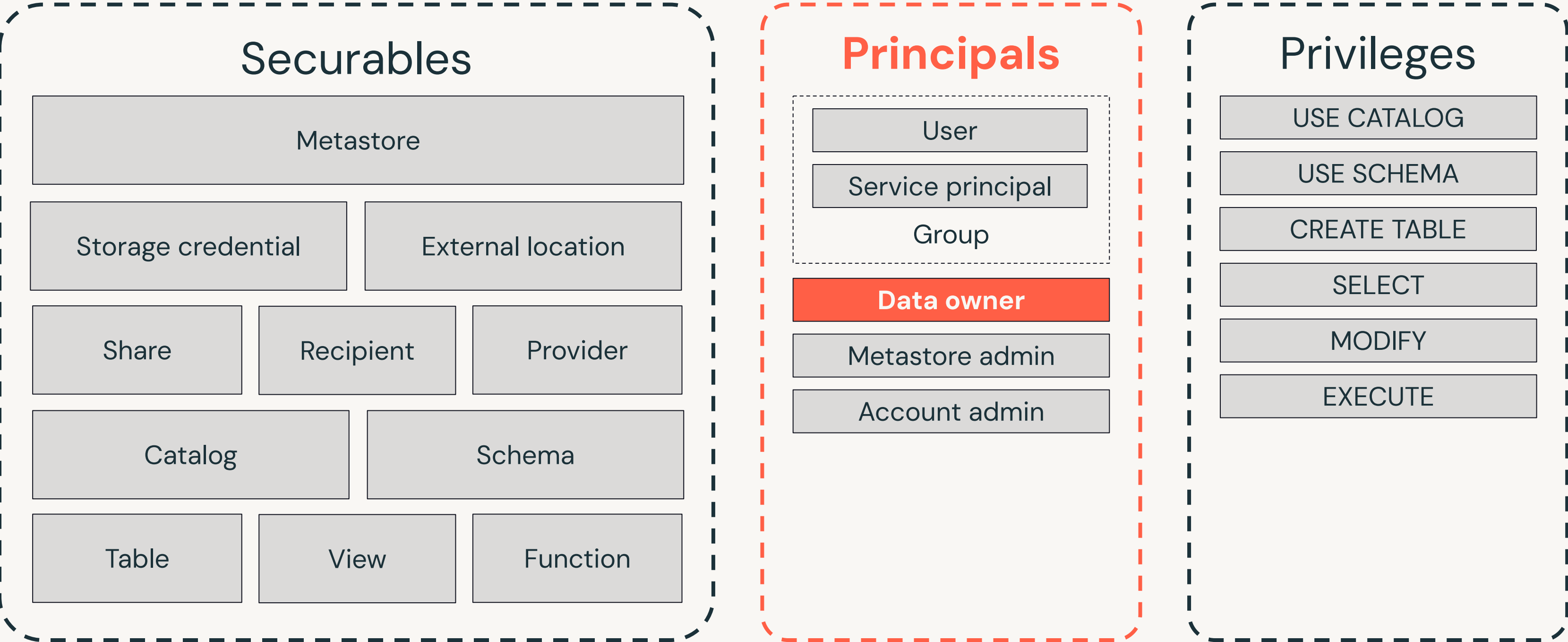
Data Security Model

Access control lists (ACLs)



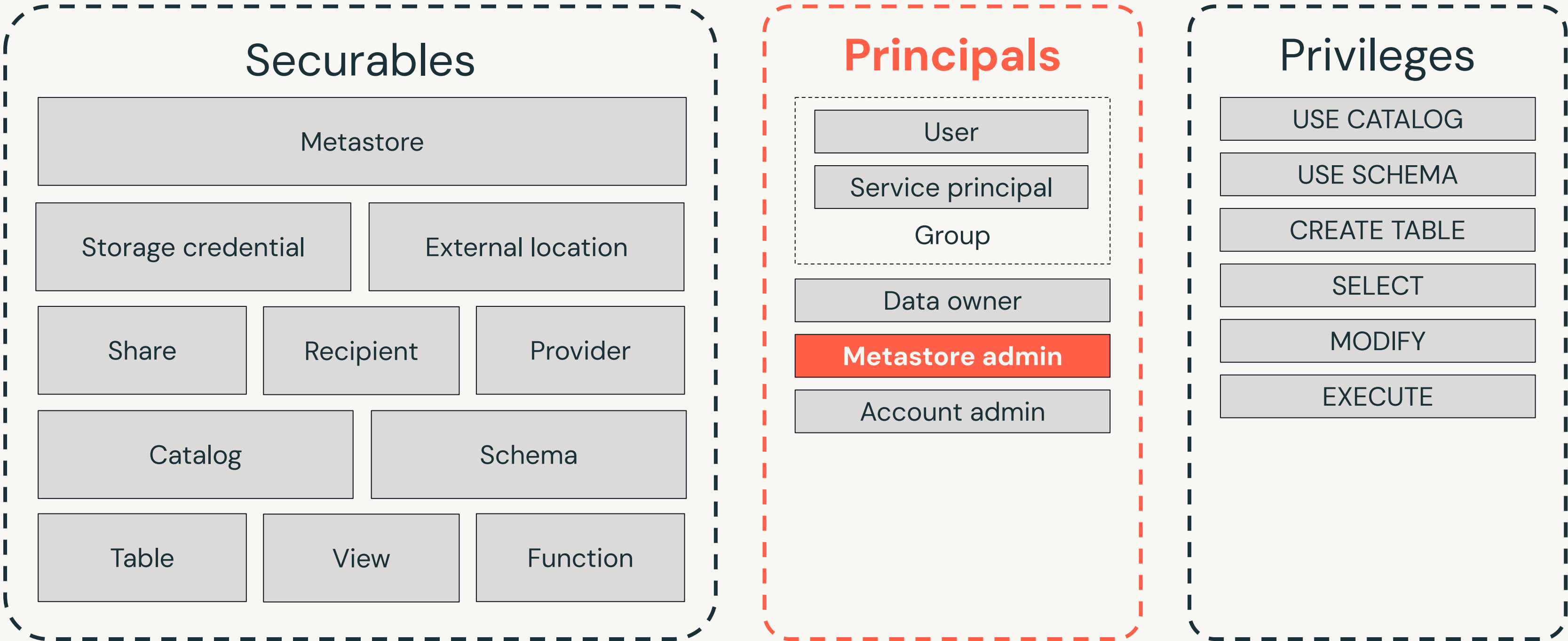
Data Security Model

Access control lists (ACLs)



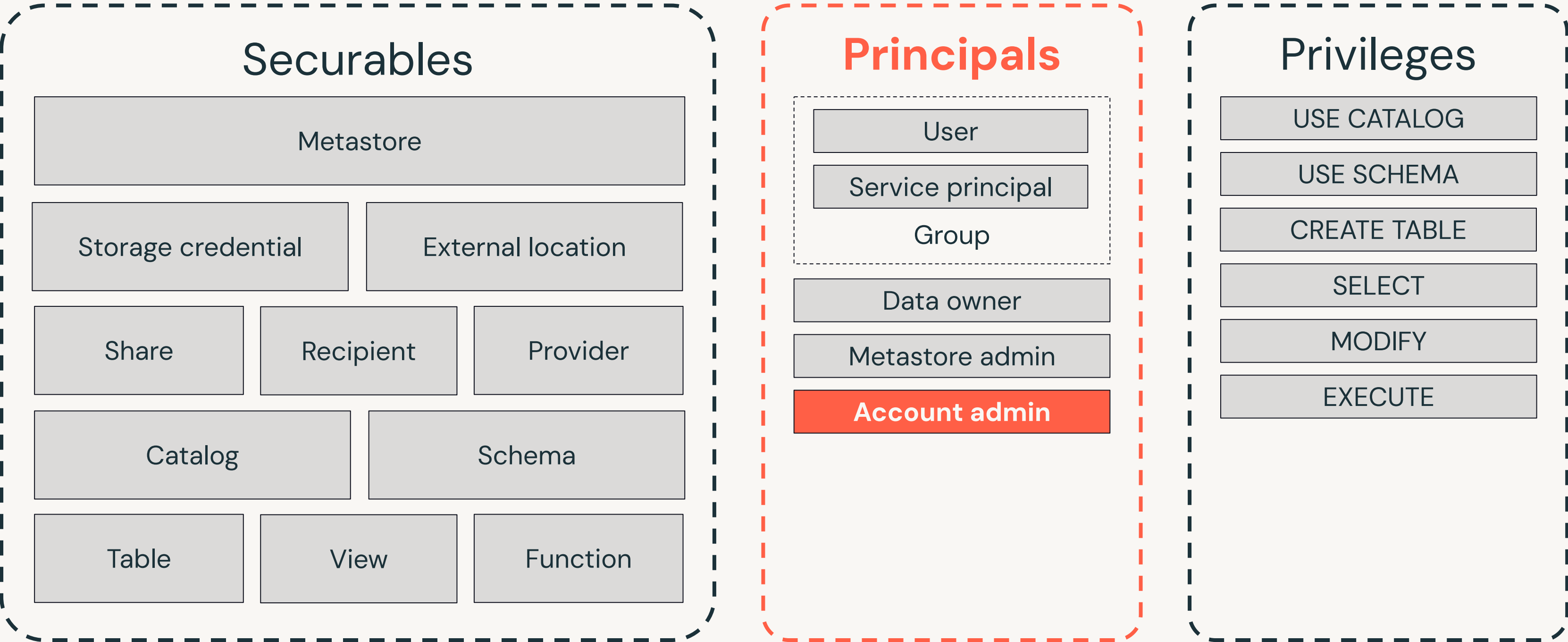
Data Security Model

Access control lists (ACLs)



Data Security Model

Access control lists (ACLs)



Data Security Model

Managing ACLs



Data Security Model

Managing ACLs

1. SQL

GRANT SELECT ON TABLE t TO analysts

REVOKE SELECT ON TABLE t FROM analysts



Data Security Model

Managing ACLs

1. SQL

GRANT SELECT ON TABLE t TO analysts
REVOKE SELECT ON TABLE t FROM analysts



Notebook

Databricks SQL



Data Security Model

Managing ACLs

1. SQL

GRANT SELECT ON TABLE t TO analysts

REVOKE SELECT ON TABLE t FROM analysts

2. Data explorer

Data Science & Engineering workspace or DBSQL



Data Security Model

Managing ACLs

1. SQL

GRANT SELECT ON TABLE `t` TO analysts

REVOKE SELECT ON TABLE `t` FROM analysts

2. Data explorer

Data Science & Engineering workspace or DBSQL

3. Programmatically

Databricks CLI, Terraform, and REST APIs

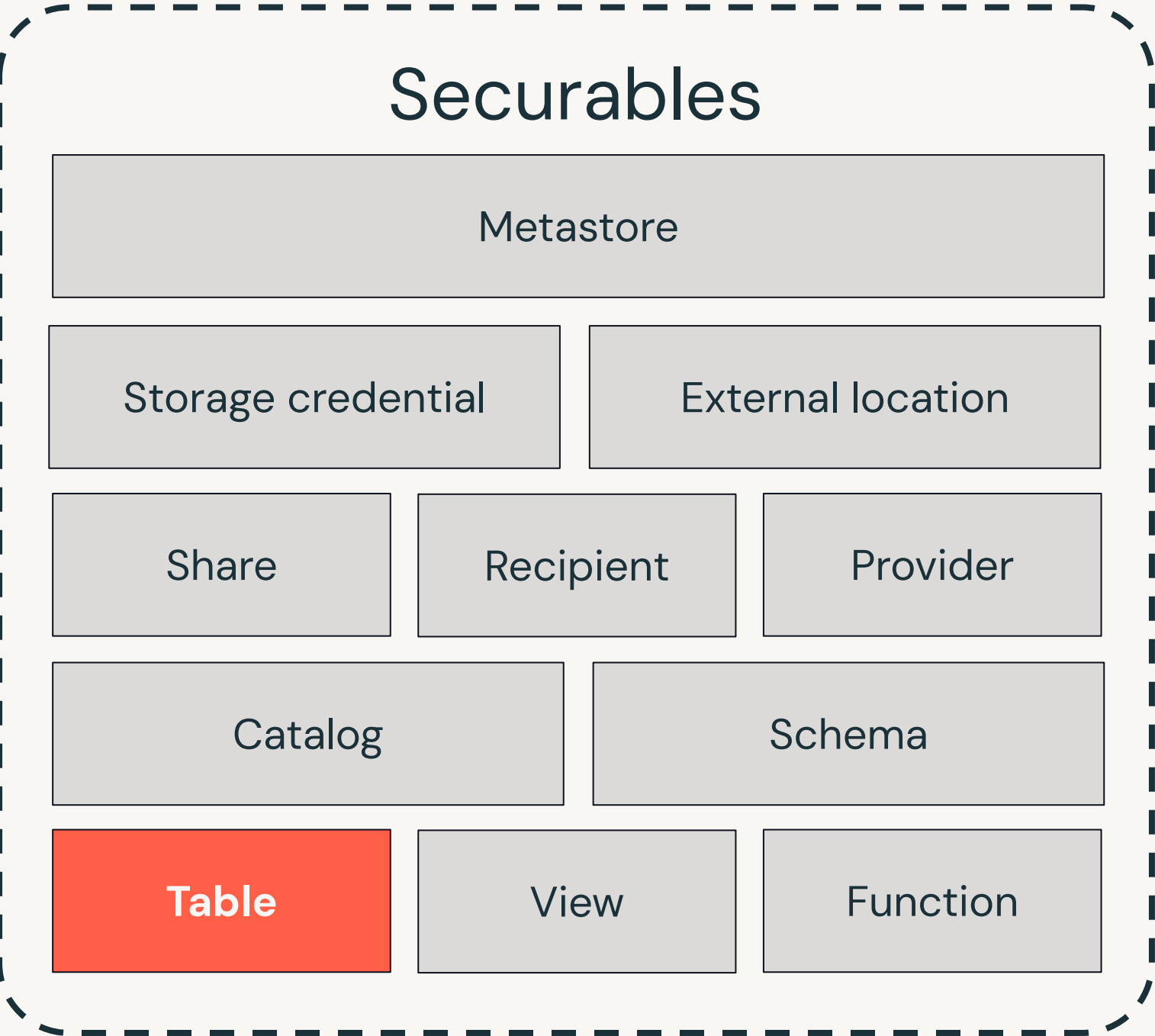


Data Access Control

Privilege Types

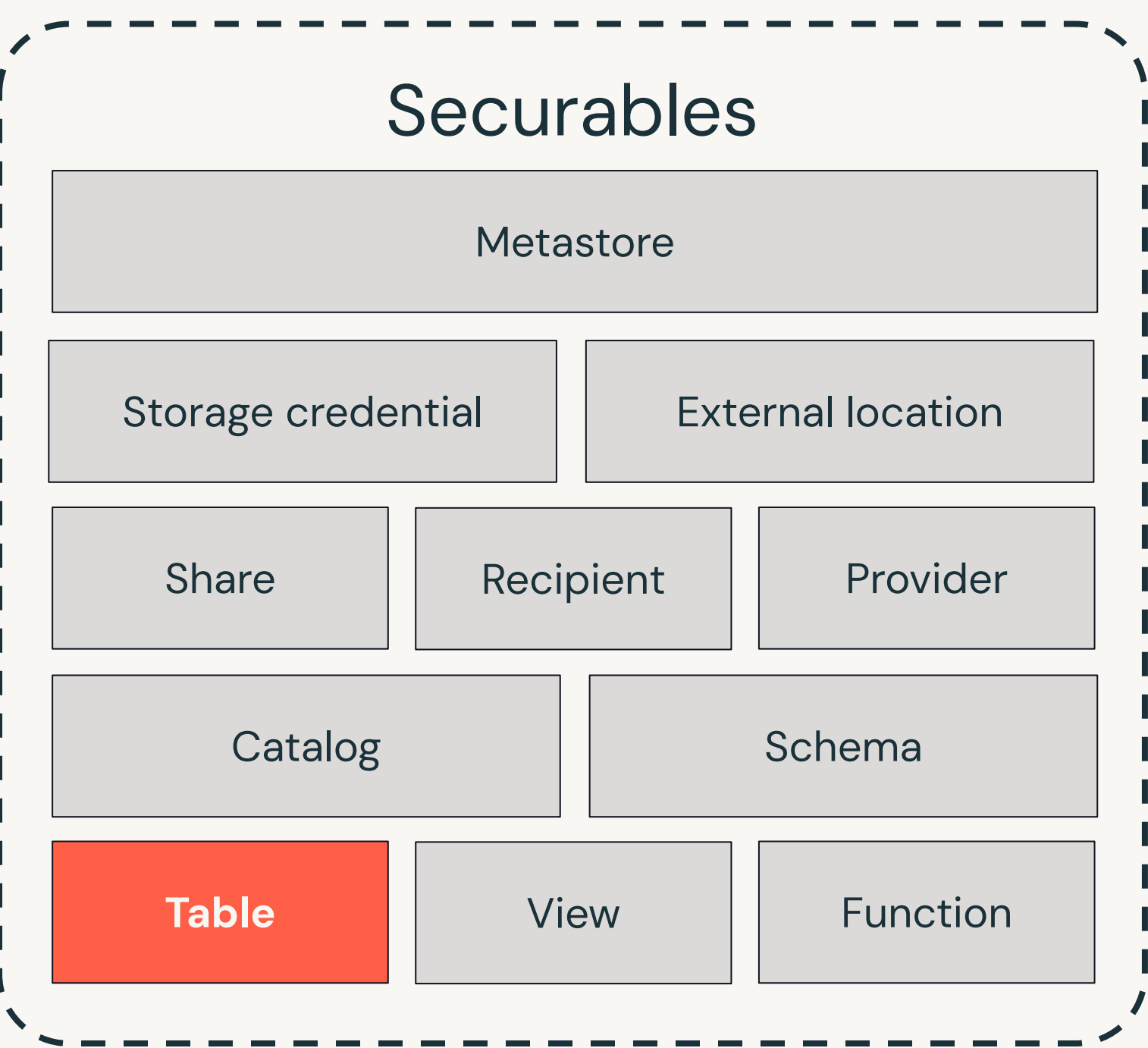
Privilege Types

Table



Privilege Types

Table



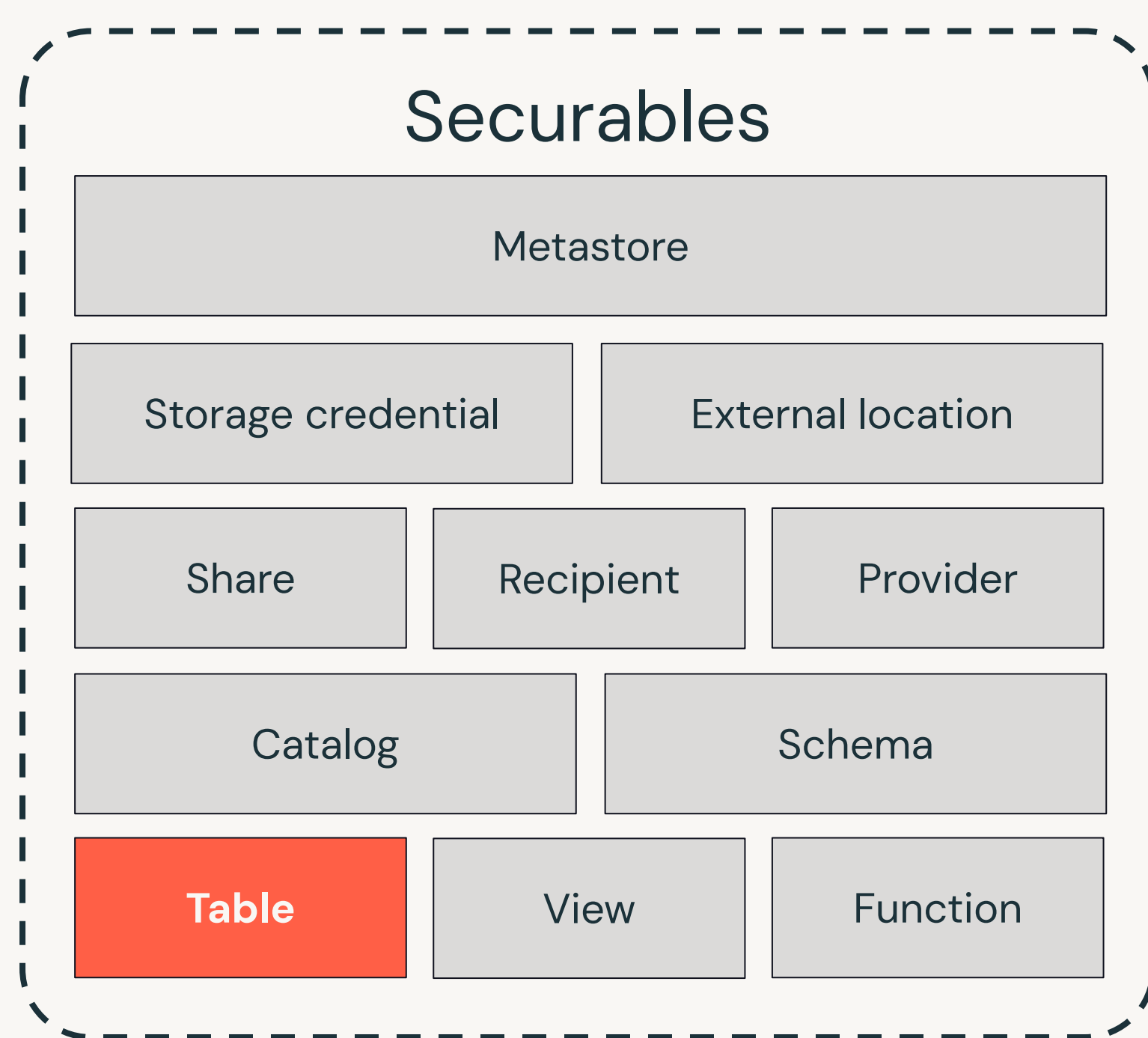
Privileges

SELECT
Read from a table (SELECT)



Privilege Types

Table



Privileges

SELECT

Read from a table (SELECT)

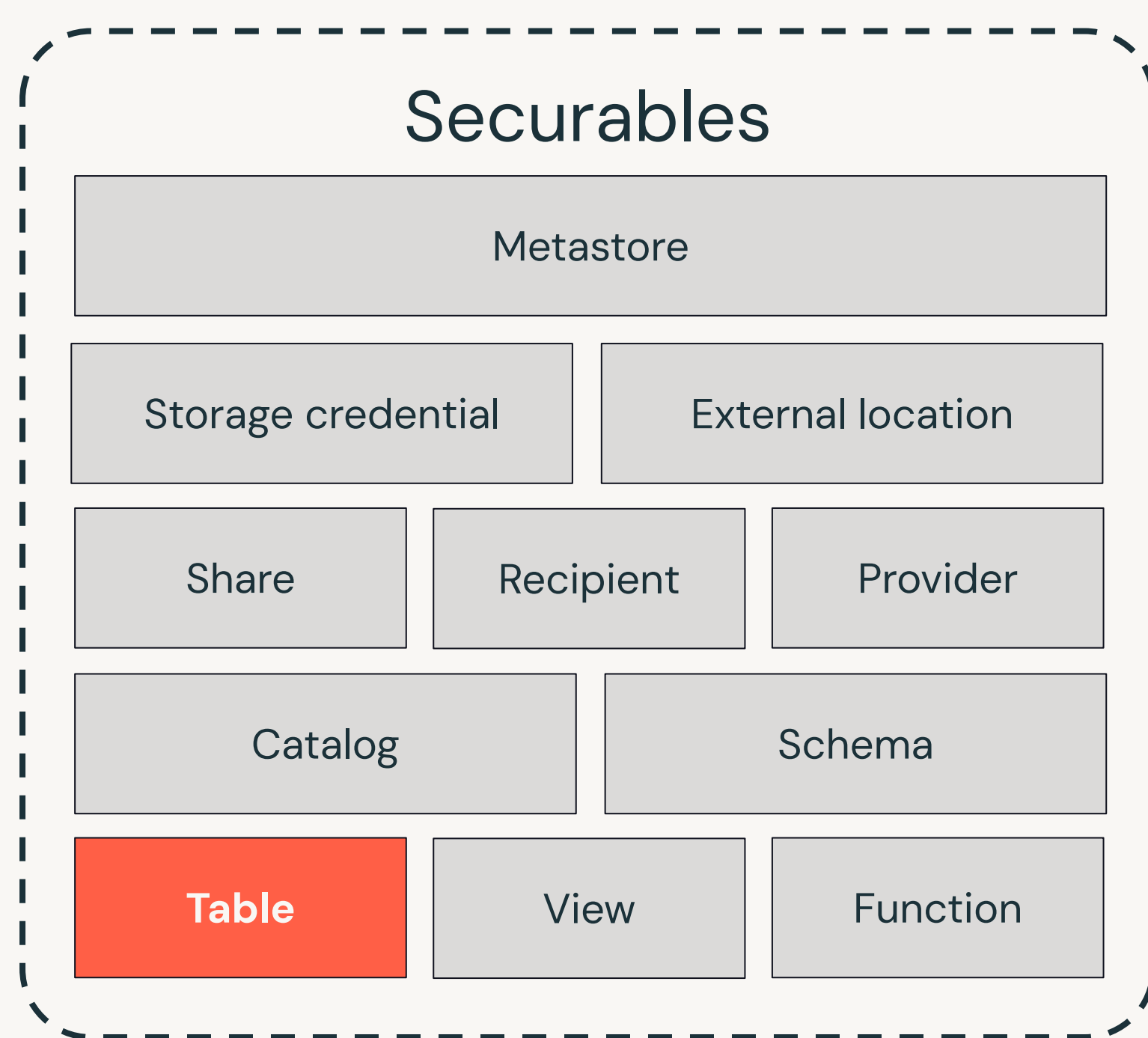
MODIFY

Modify table data (UPDATE) or metadata (ALTER)



Privilege Types

Table



Privileges

SELECT

Read from a table (SELECT)

MODIFY

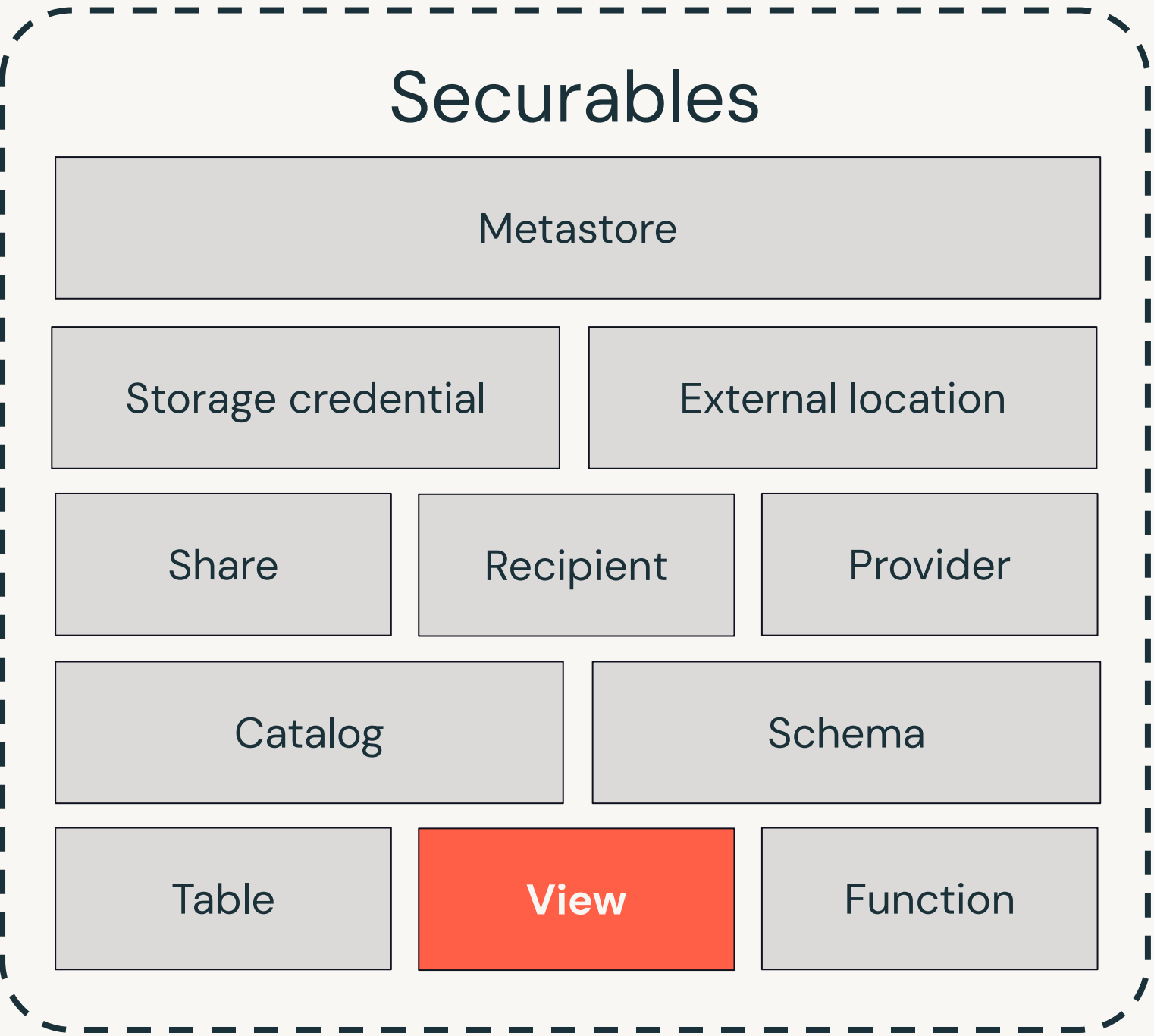
Modify table data (UPDATE) or metadata (ALTER)

ALL PRIVILEGES



Privilege Types

View



Privileges

SELECT

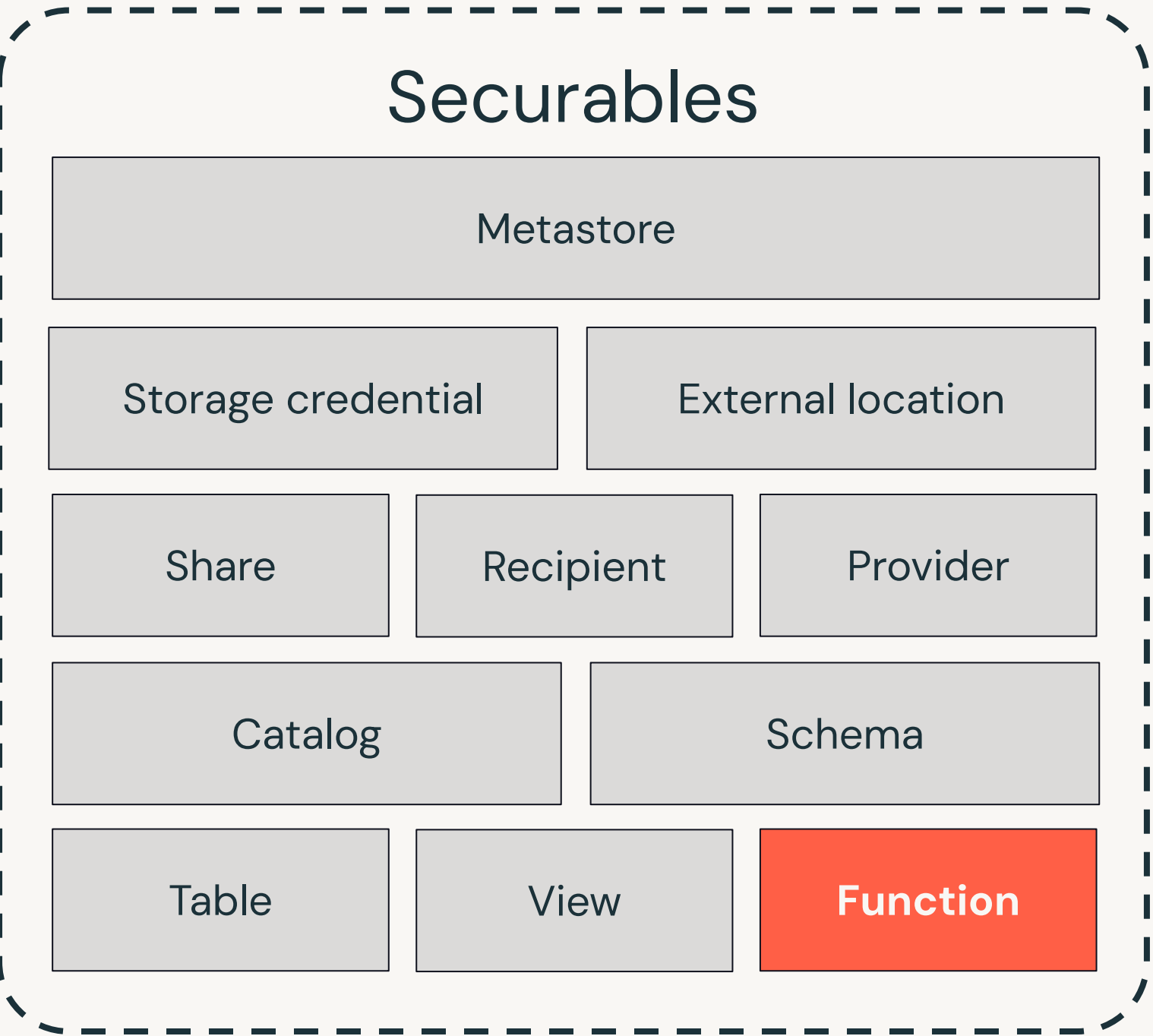
Query view (SELECT)

ALL PRIVILEGES



Privilege Types

Function



Privileges

EXECUTE

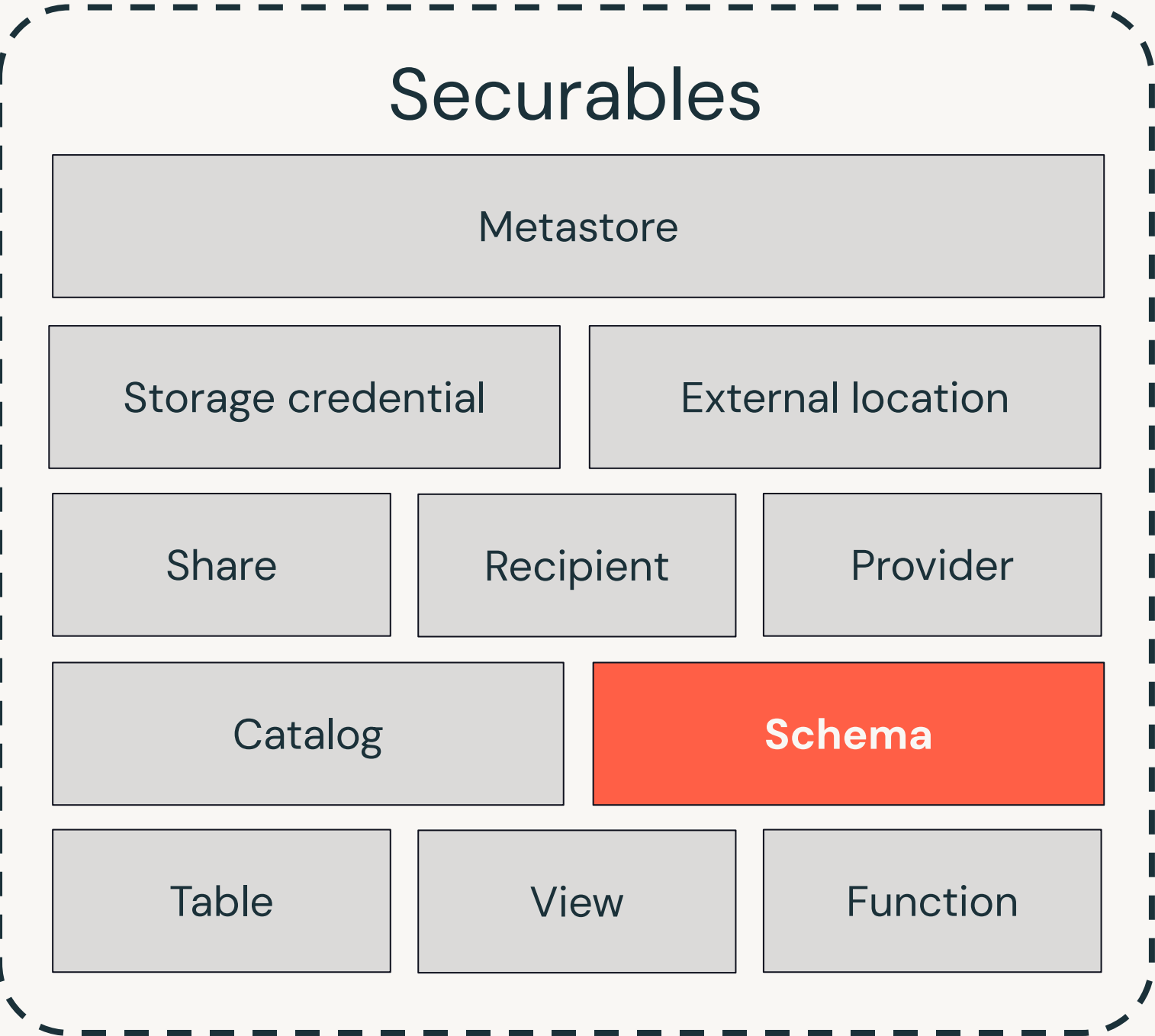
Use function in a query

ALL PRIVILEGES



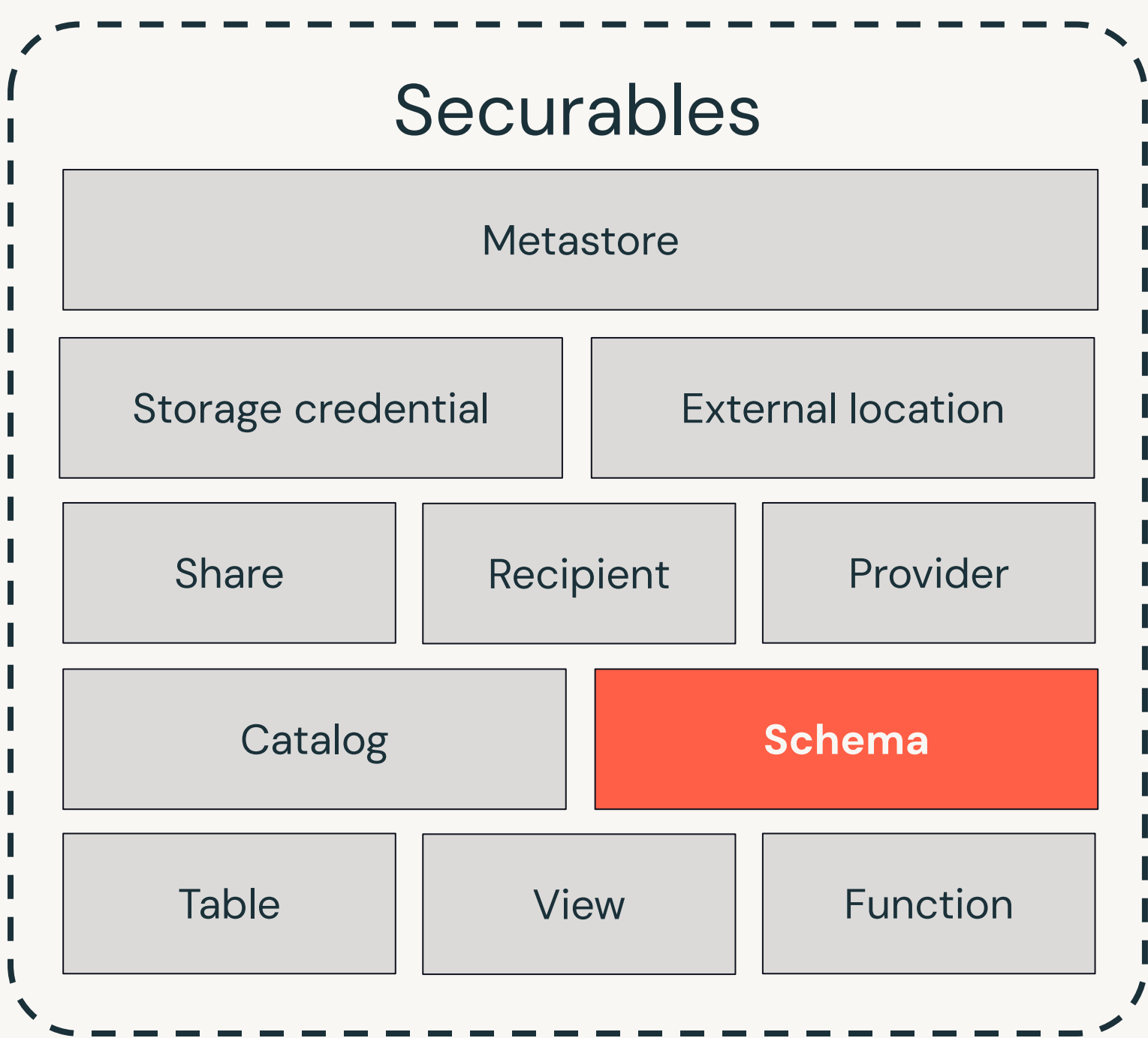
Privilege Types

Schema



Privilege Types

Schema



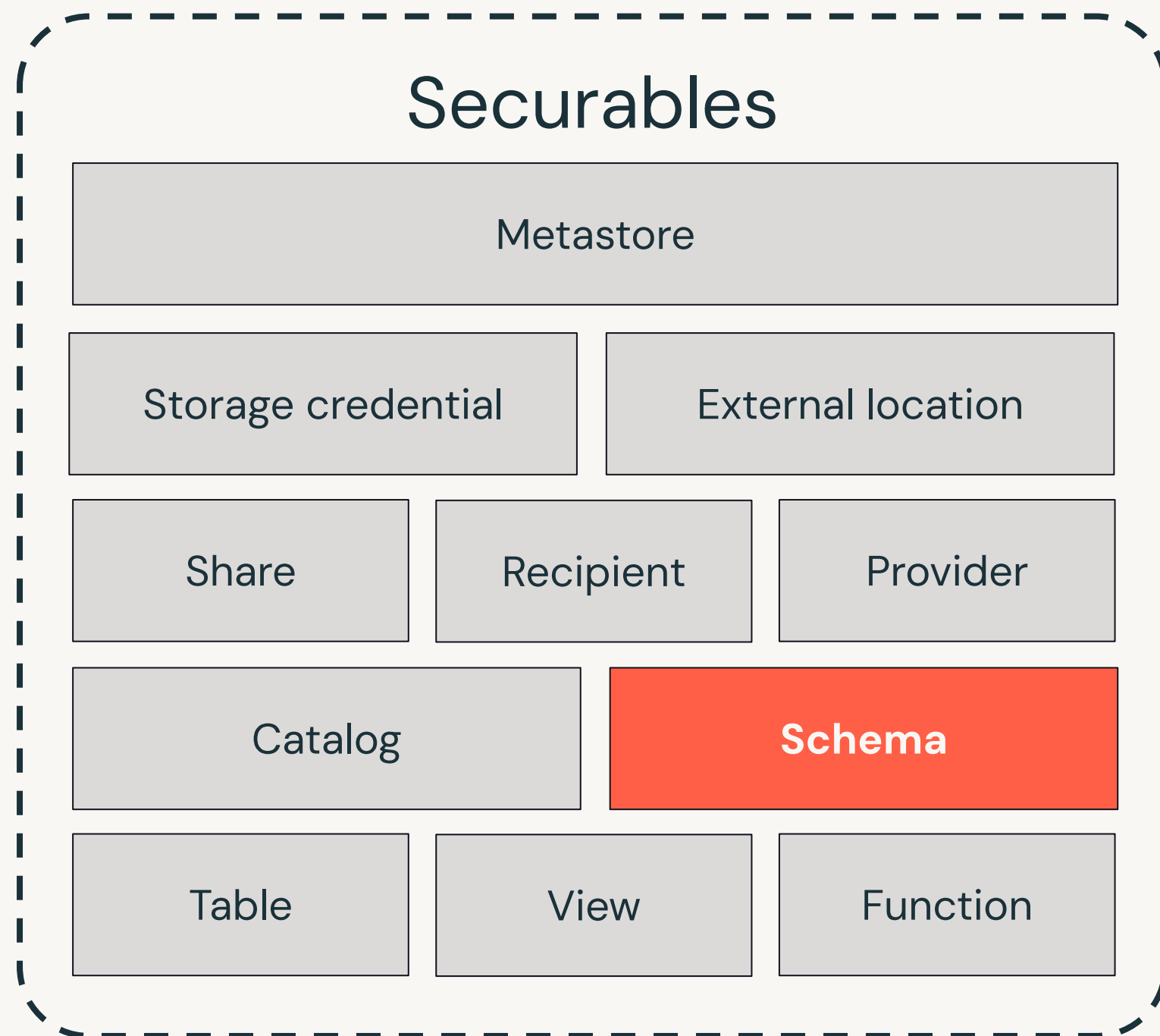
Privileges

USE SCHEMA
Access objects in schema



Privilege Types

Schema



Privileges

USE SCHEMA

Access objects in schema

CREATE TABLE

Create table or view in schema

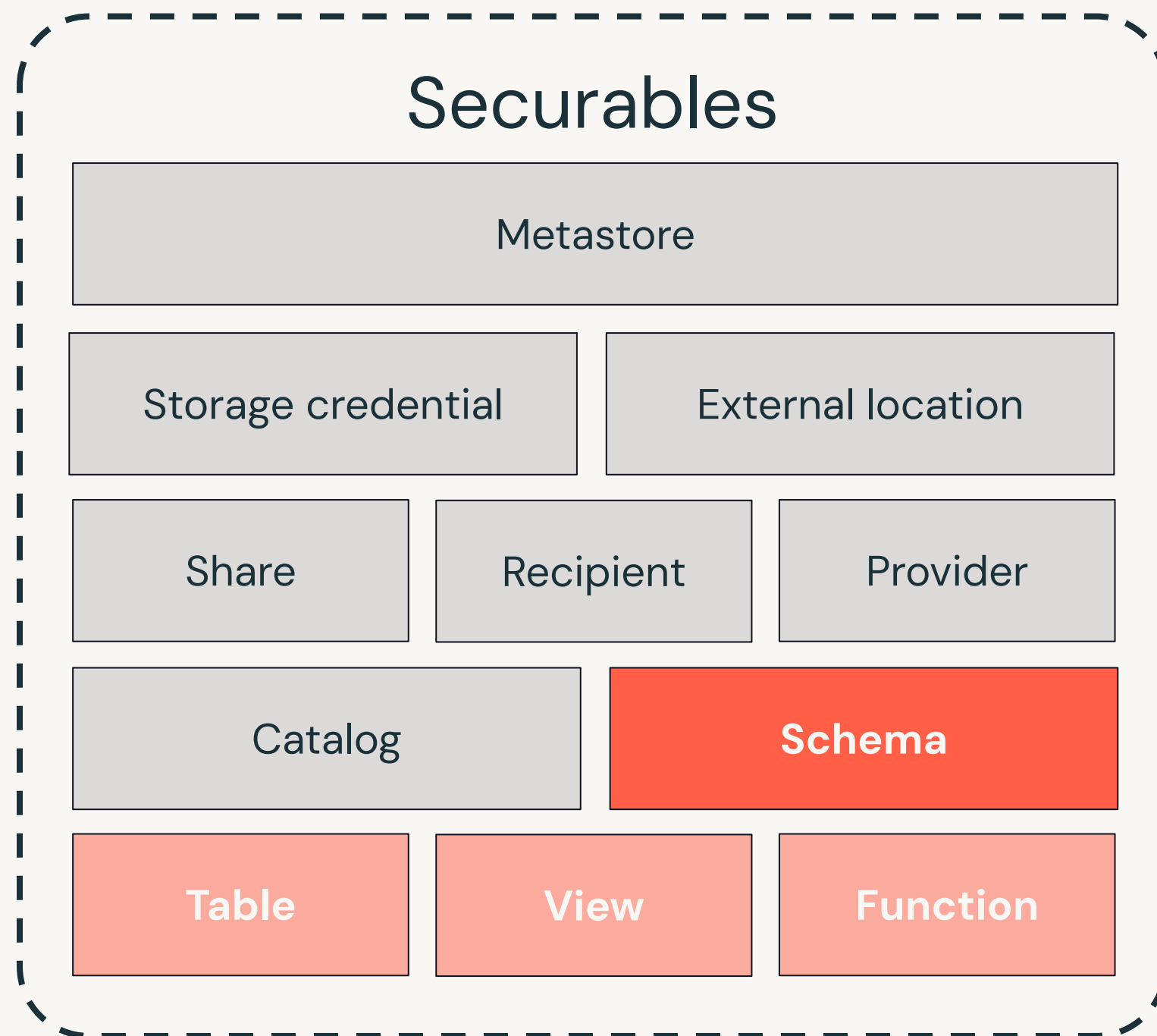
CREATE FUNCTION

Create user-defined function in schema



Privilege Types

Schema privilege inheritance



Bequeathed privileges

SELECT

Read from any table or view in schema

MODIFY

Modify data or metadata of any table in schema

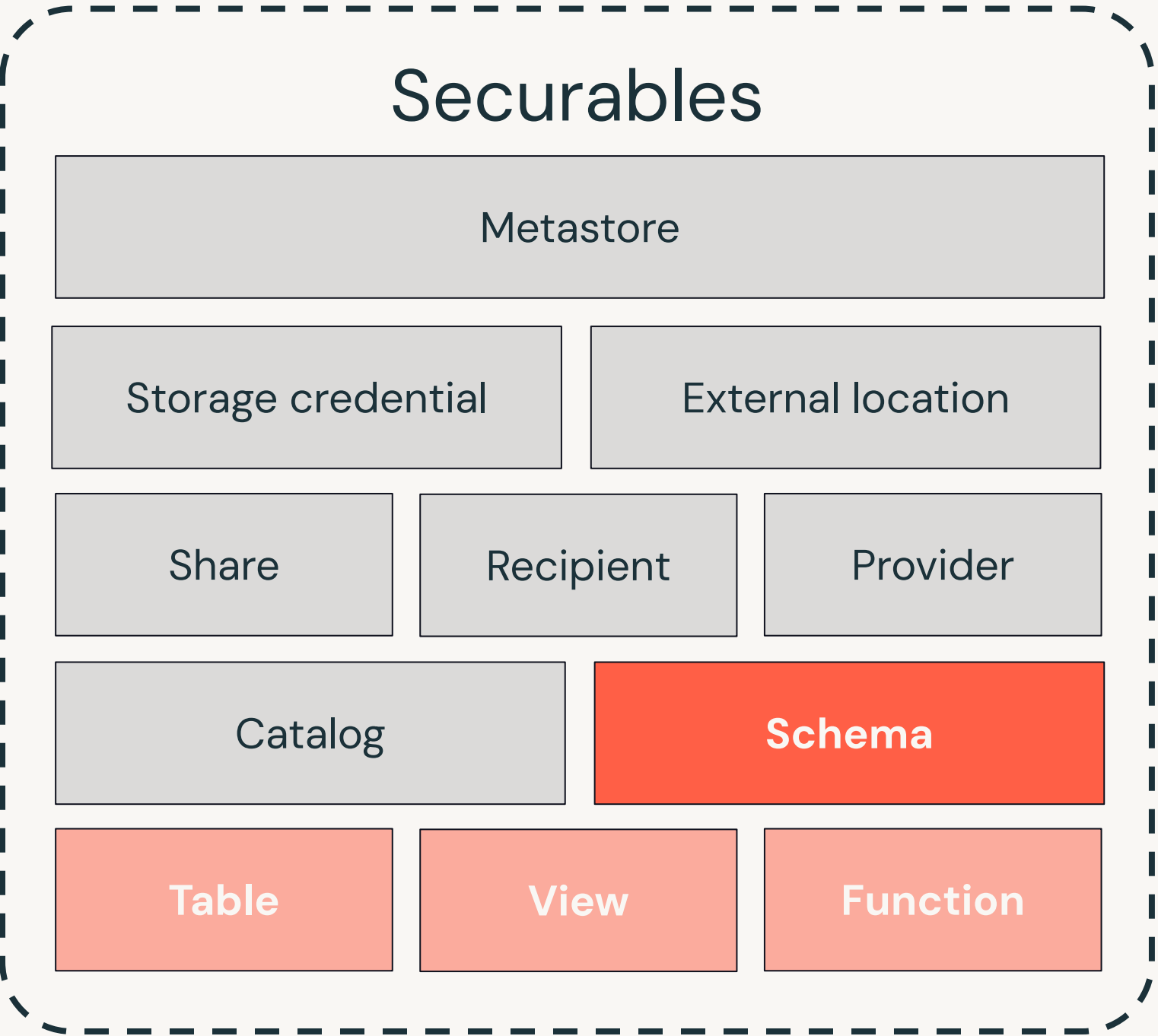
EXECUTE

Use any function in schema



Privilege Types

Schema



ALL PRIVILEGES

USE SCHEMA

CREATE TABLE

CREATE FUNCTION

SELECT

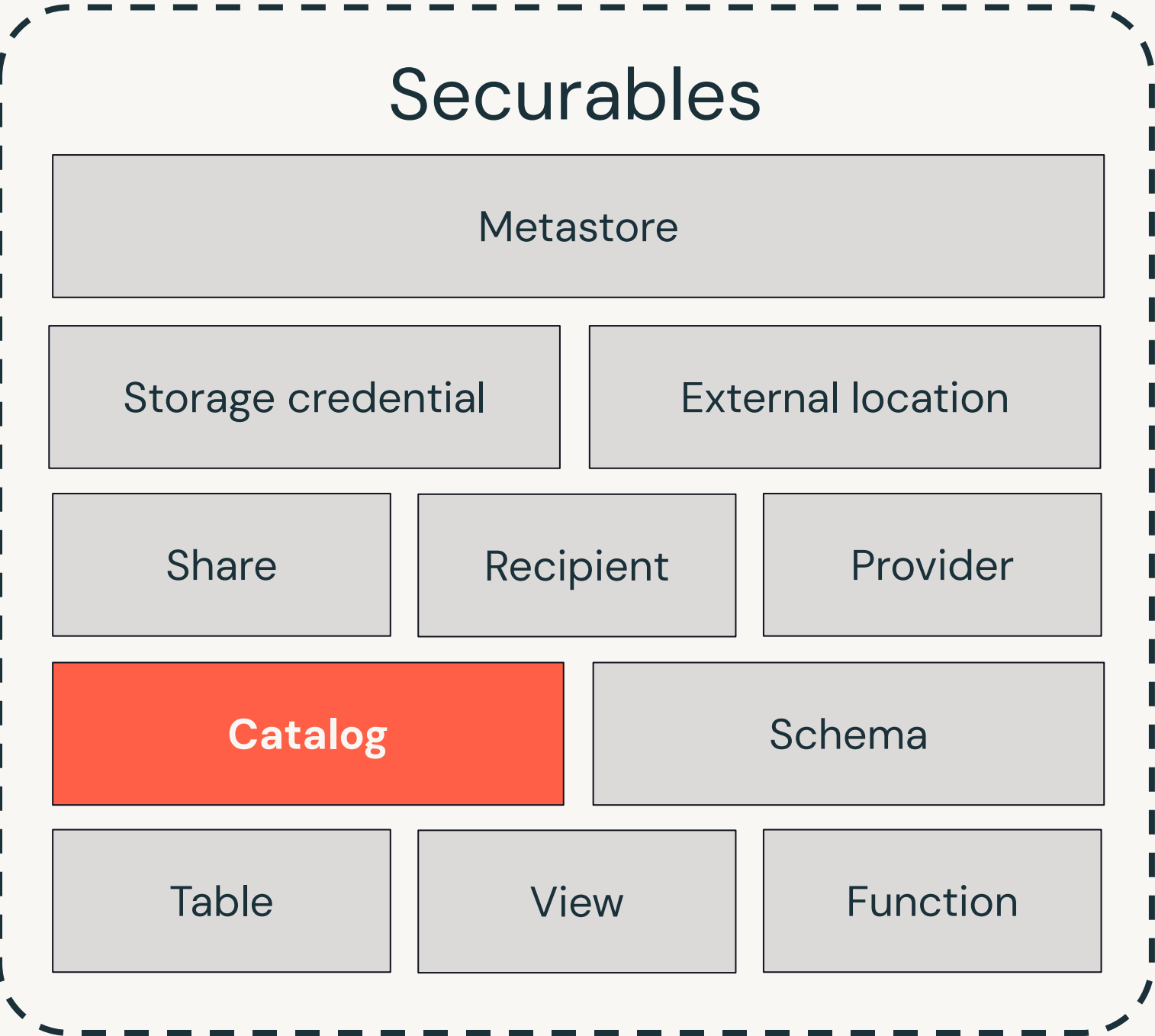
MODIFY

EXECUTE



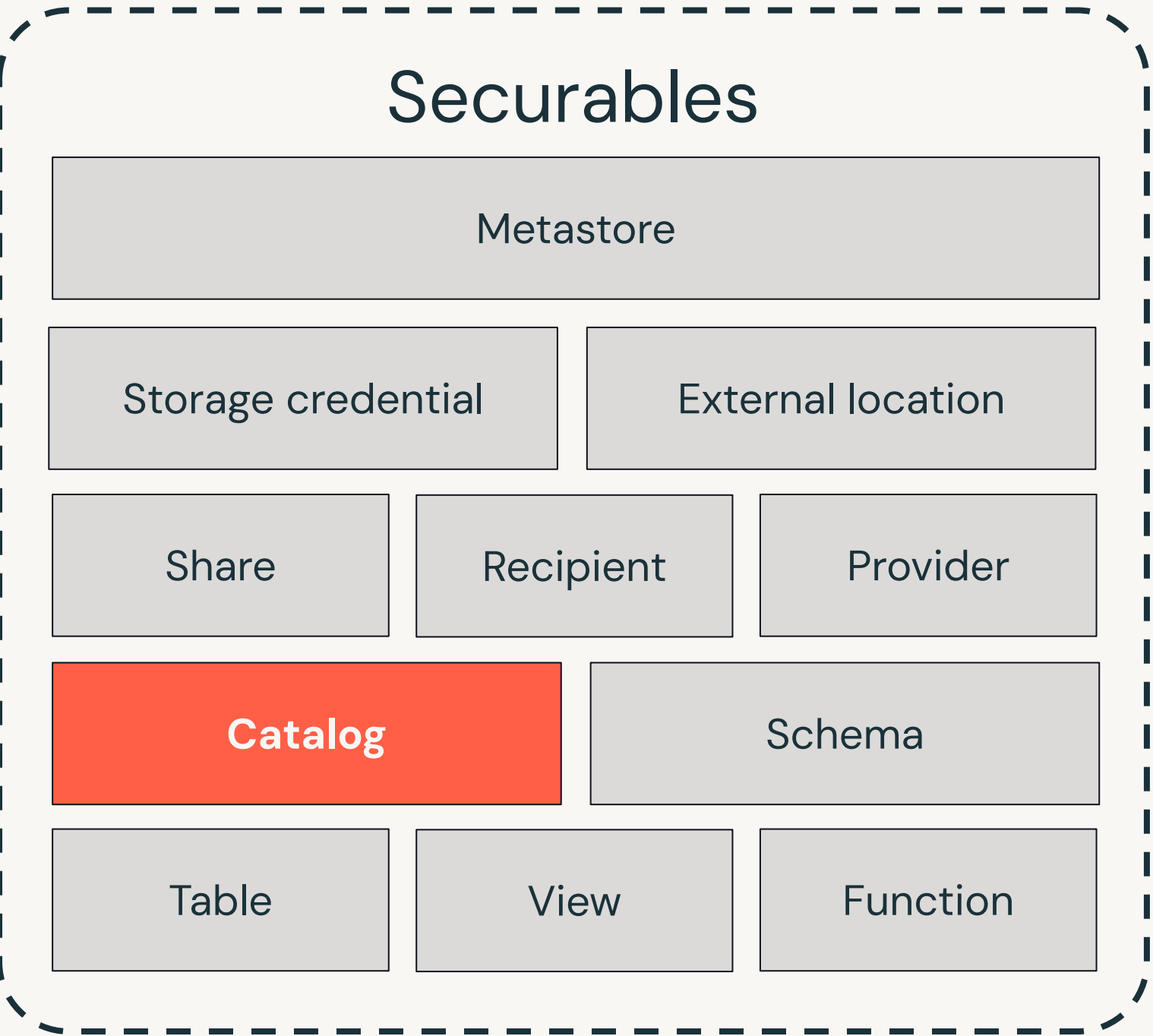
Privilege Types

Catalog



Privilege Types

Catalog



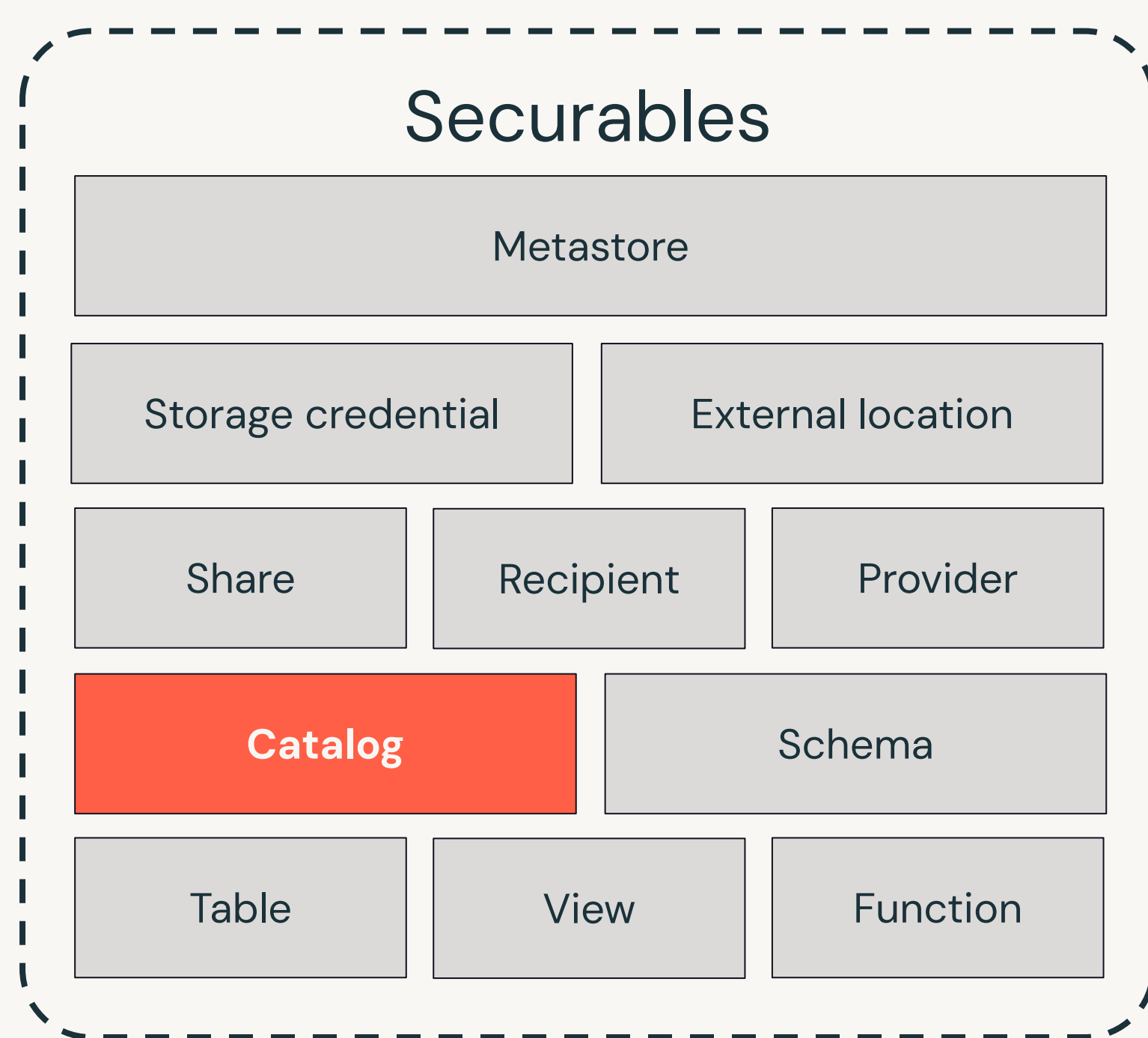
Privileges

USE CATALOG
Access schemas in catalog



Privilege Types

Catalog



Privileges

USE CATALOG

Access schemas in catalog

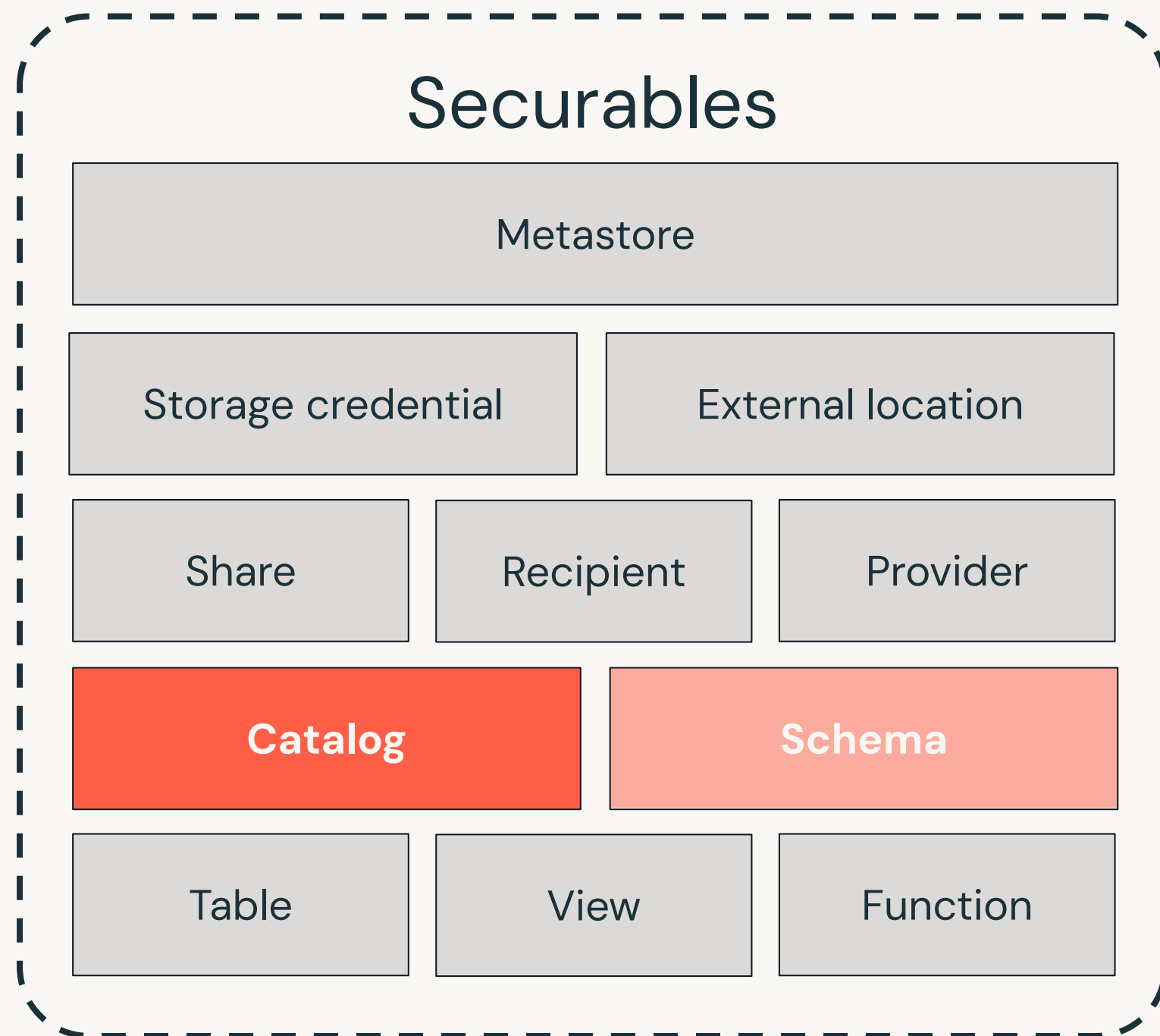
CREATE SCHEMA

Create schema in catalog



Privilege Types

Catalog privilege inheritance



Bequeathed privileges

CREATE TABLE

Create table or view in any schema

CREATE FUNCTION

Create user-defined function in any schema

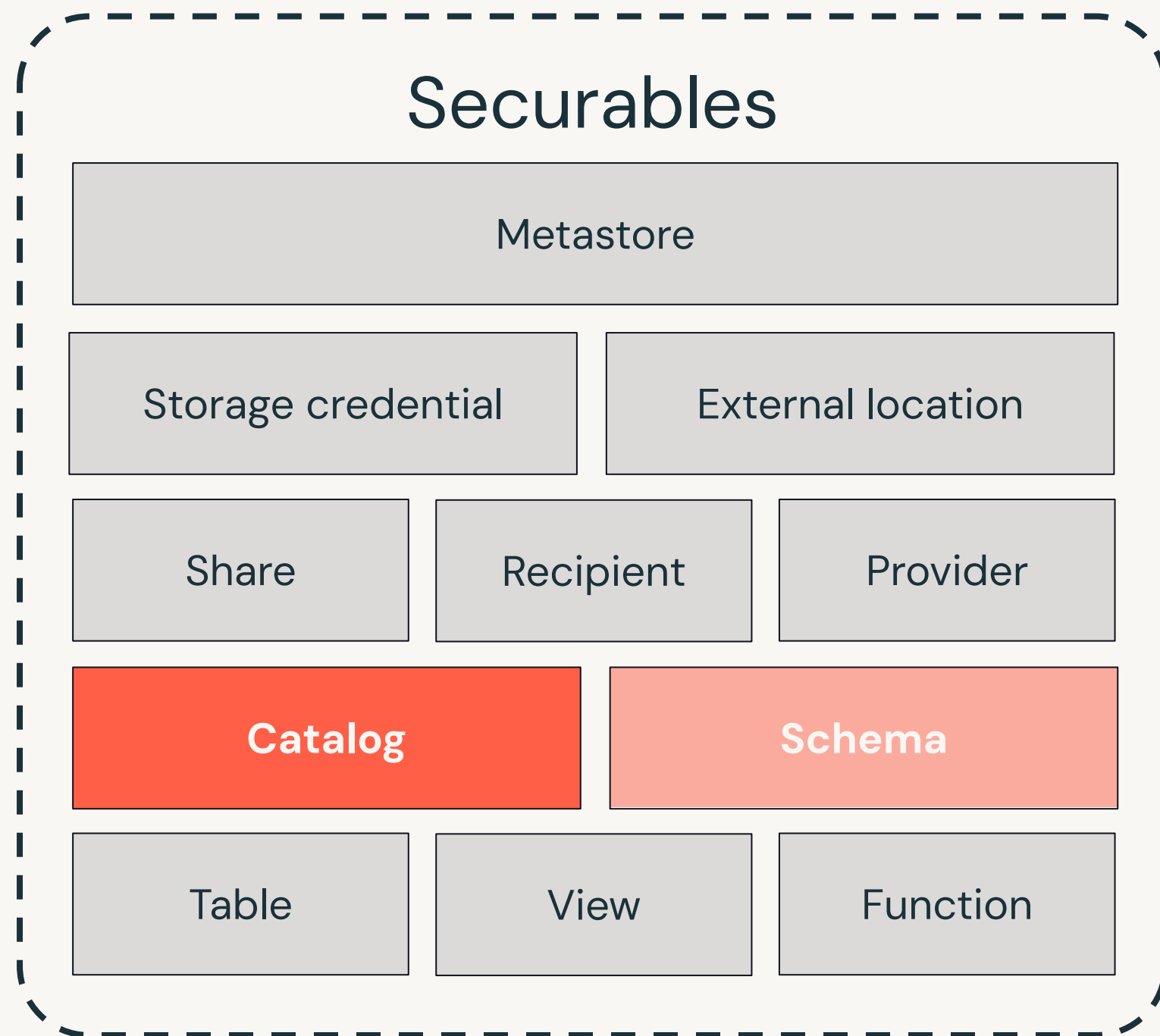
USE SCHEMA

Access objects in any schema



Privilege Types

Catalog privilege inheritance



Bequeathed privileges

CREATE TABLE

Create table or view in any schema

CREATE FUNCTION

Create user-defined function in any schema

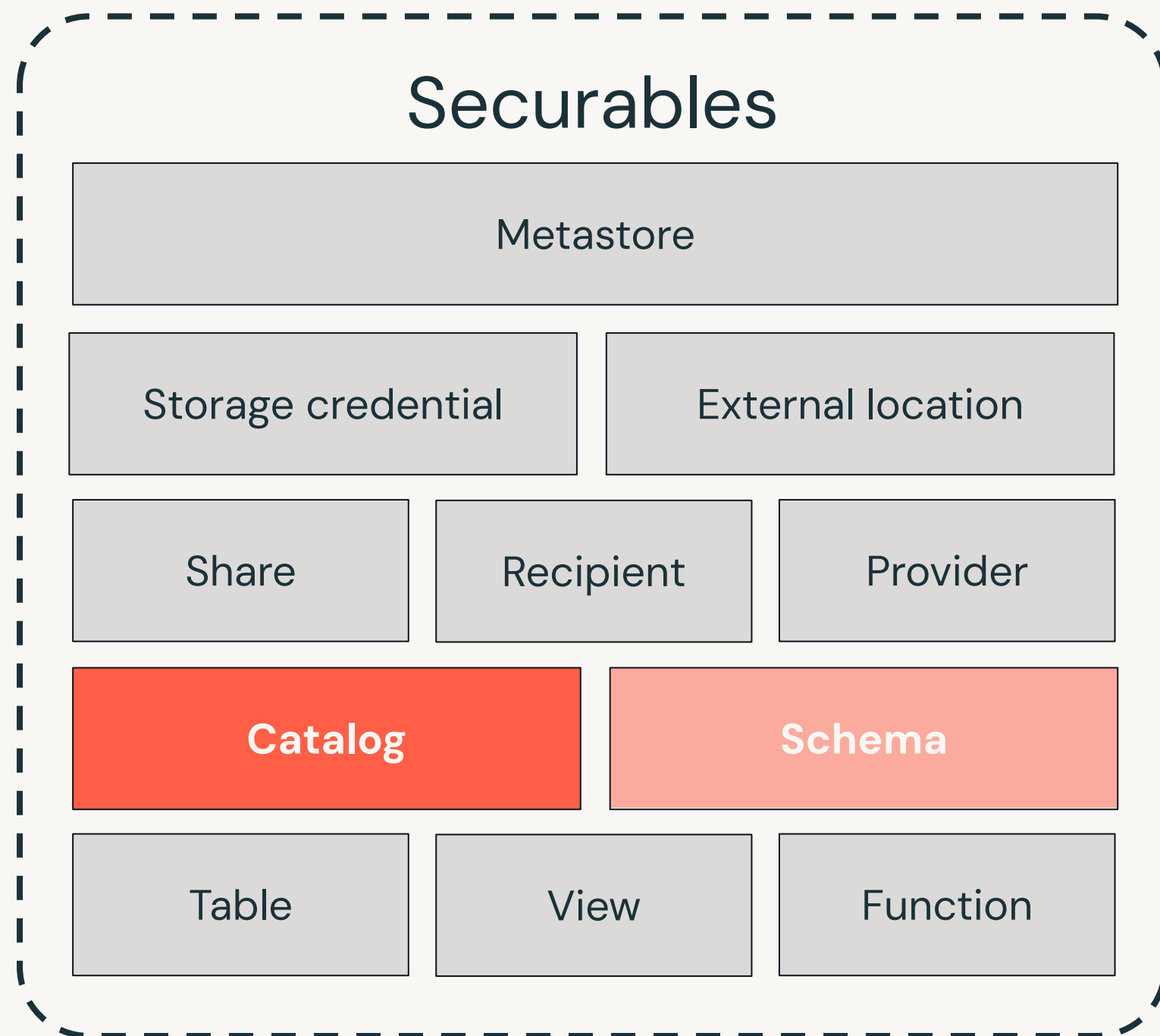
USE SCHEMA

Access objects in any schema



Privilege Types

Catalog privilege inheritance



Bequeathed privileges

CREATE TABLE

Create table or view in any schema

CREATE FUNCTION

Create user-defined function in any schema

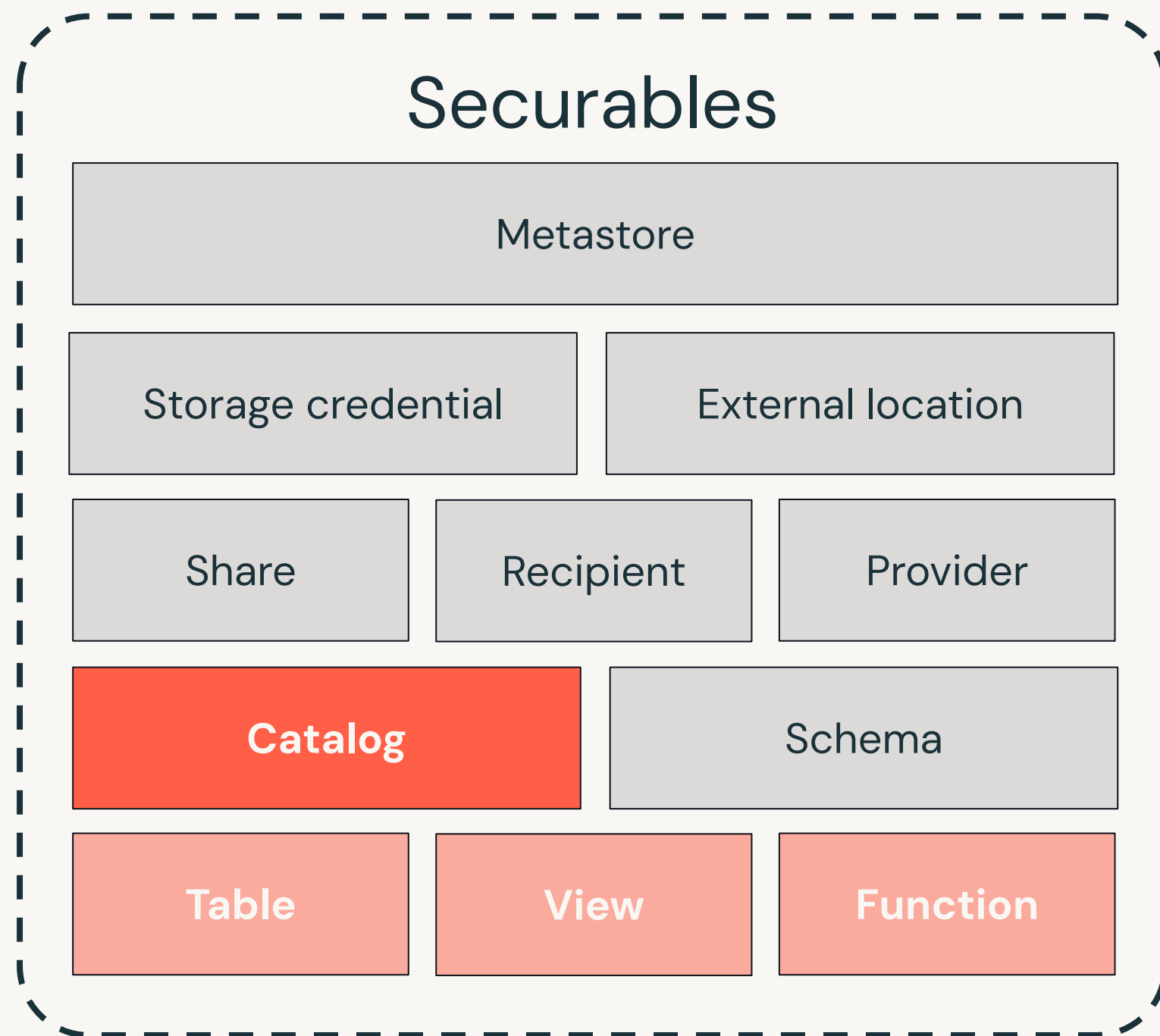
USE SCHEMA

Access objects in any schema



Privilege Types

Catalog privilege inheritance



Bequeathed privileges

SELECT

Read from any table or view from any schema

MODIFY

Modify data or metadata of any table in any schema

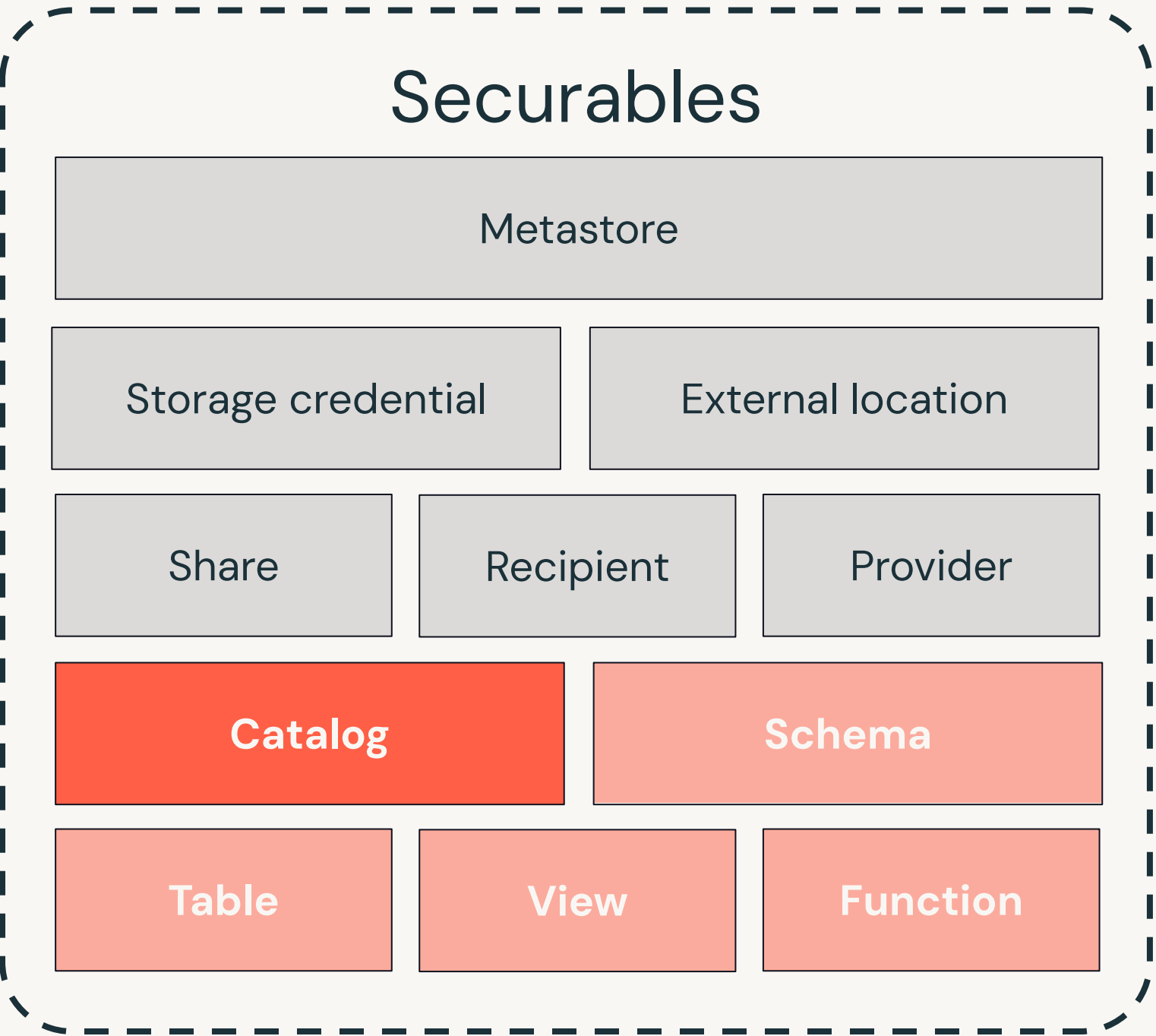
EXECUTE

Use any function from any schema



Privilege Types

Catalog



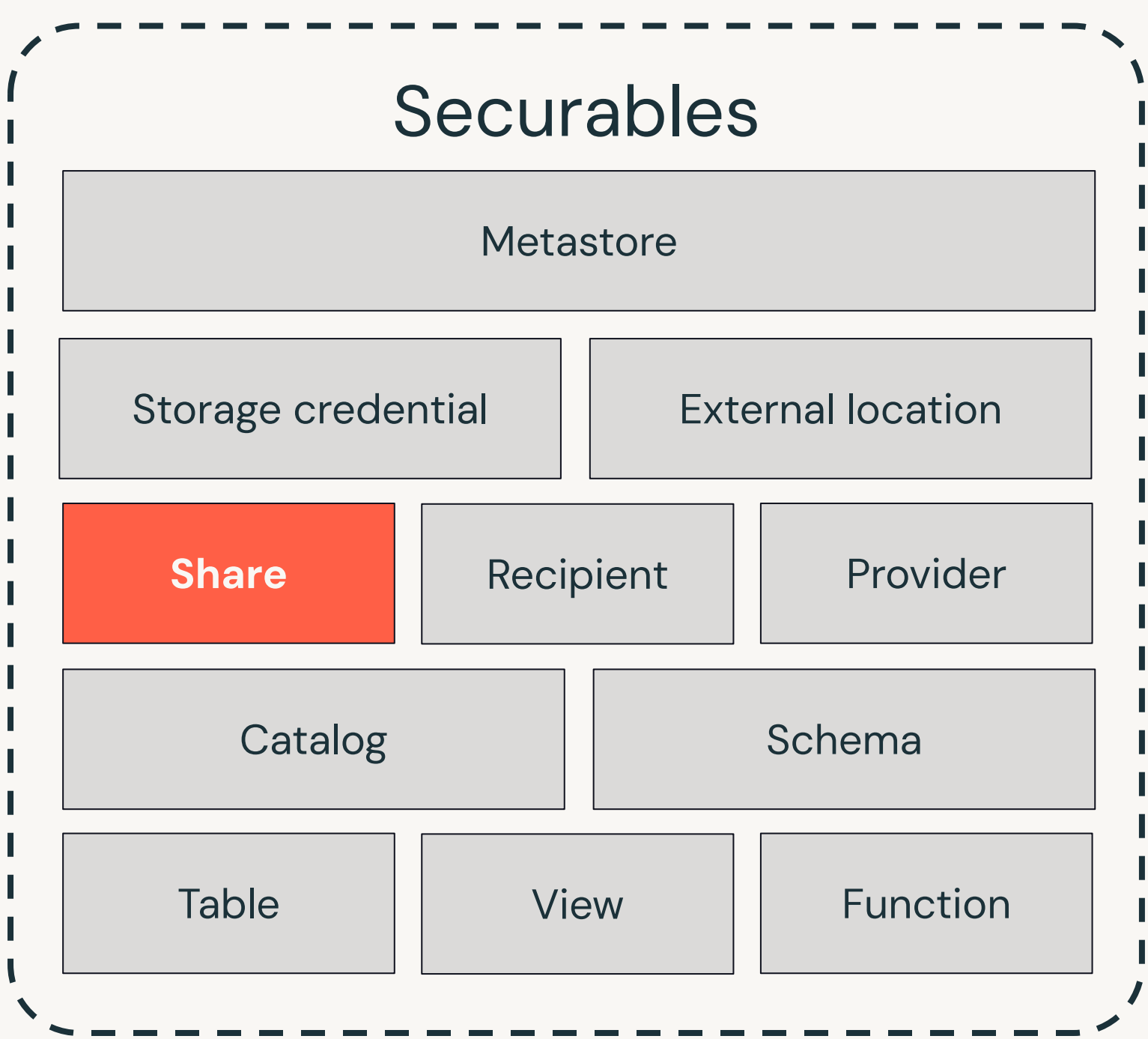
ALL PRIVILEGES

| | |
|-----------------|------------|
| USE CATALOG | USE SCHEMA |
| CREATE SCHEMA | SELECT |
| CREATE TABLE | MODIFY |
| CREATE FUNCTION | EXECUTE |



Privilege Types

Delta sharing



Privileges

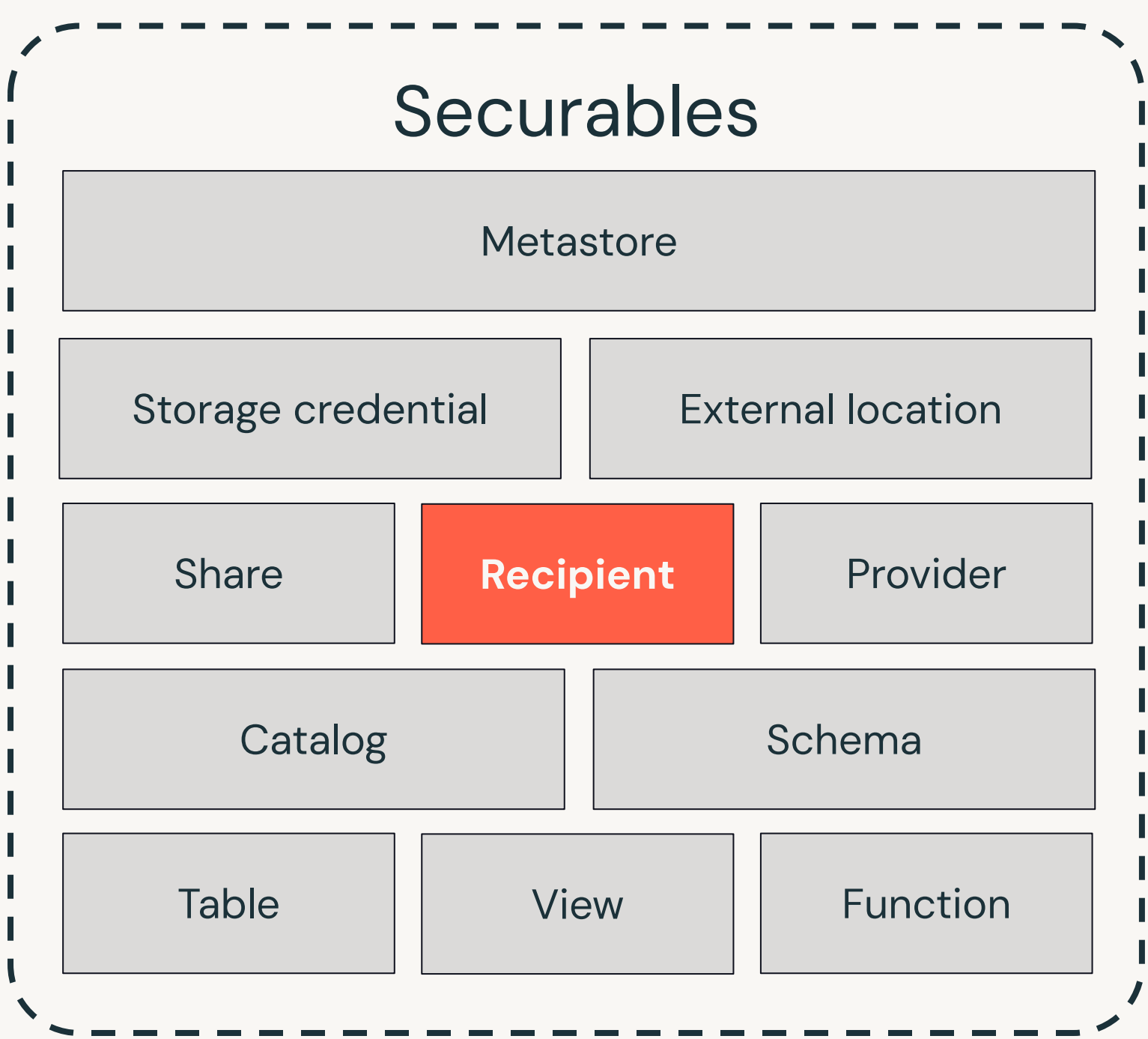
SELECT

Provide read ability to a recipient



Privilege Types

Delta sharing



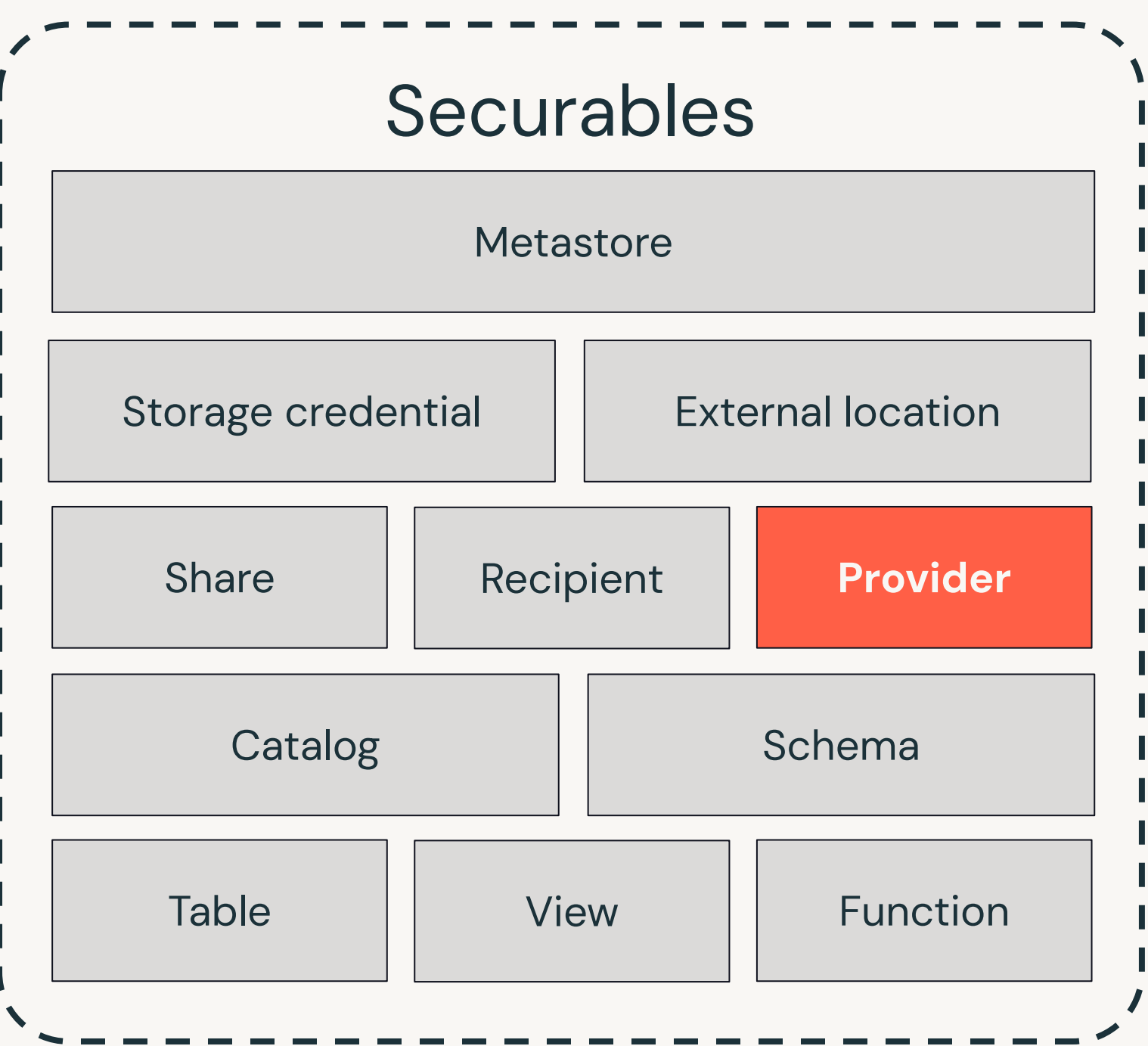
Privileges

No privileges



Privilege Types

Delta sharing



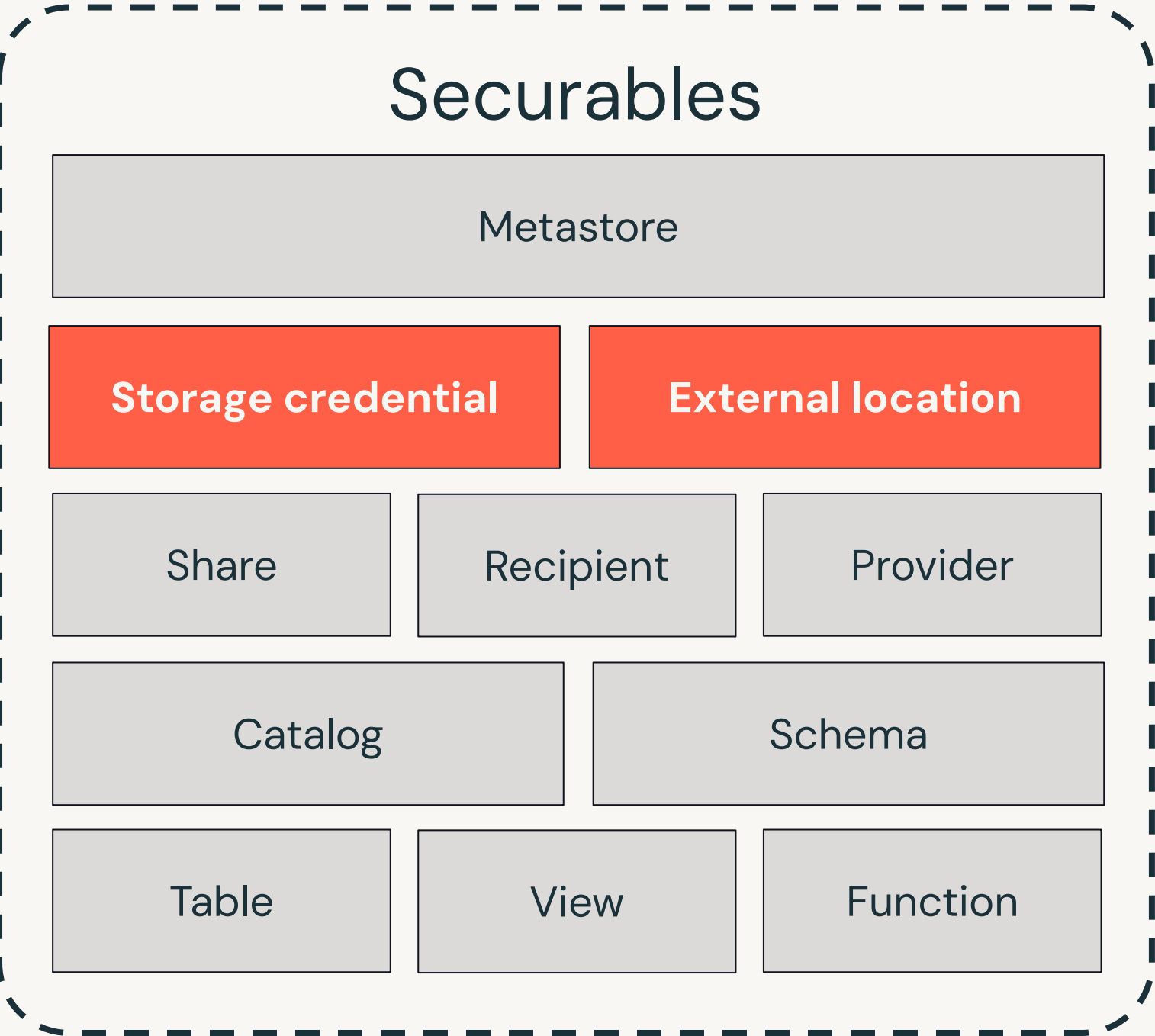
Privileges

No privileges



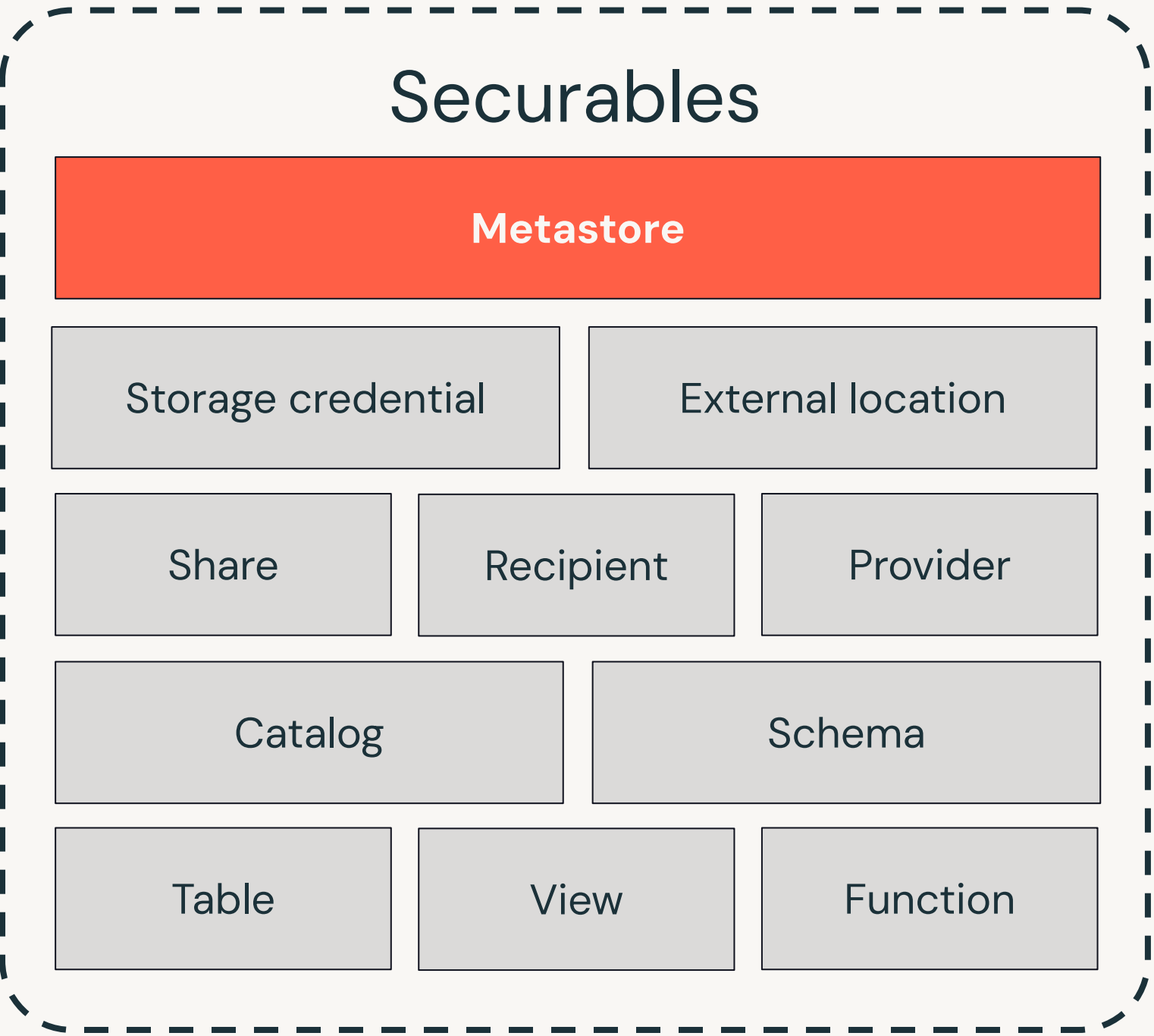
Privilege Types

External storage management



Privilege Types

Metastore



Privileges

- CREATE CATALOG
- CREATE EXTERNAL LOCATION
- CREATE RECIPIENT
- CREATE SHARE
- CREATE PROVIDER



Privilege Scenarios

Querying a table

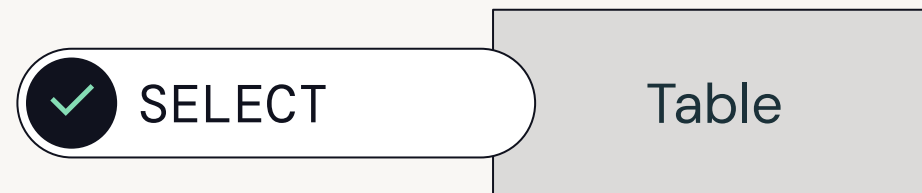


Table



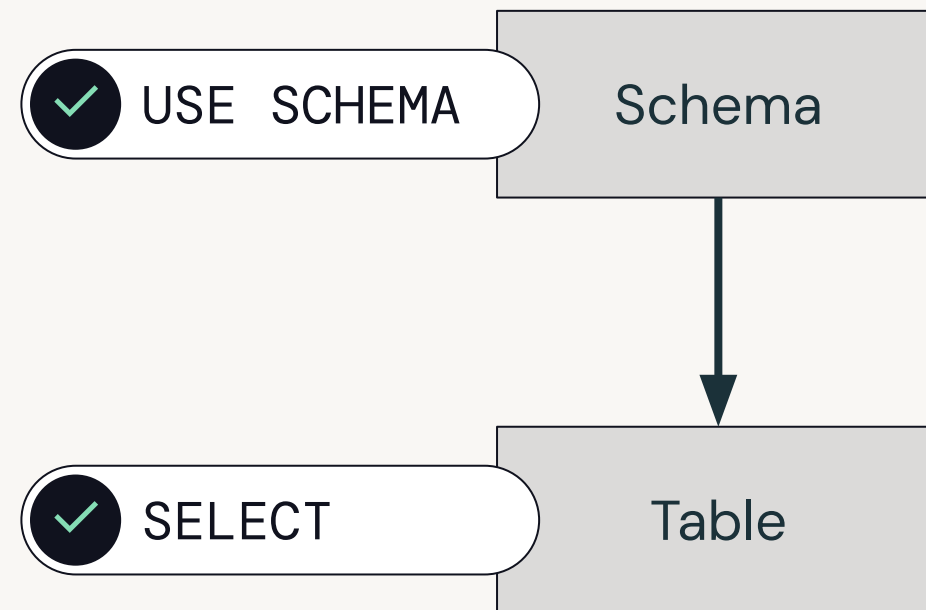
Privilege Scenarios

Querying a table



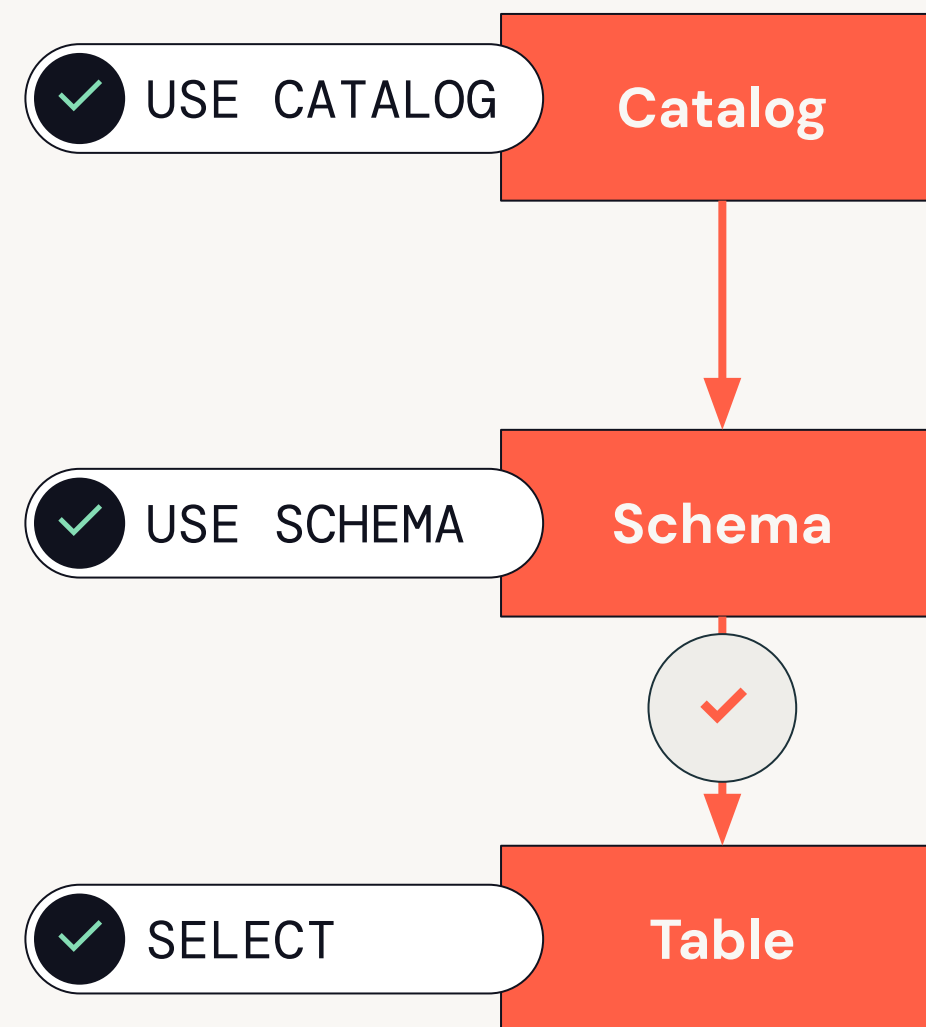
Privilege Scenarios

Querying a table



Privilege Scenarios

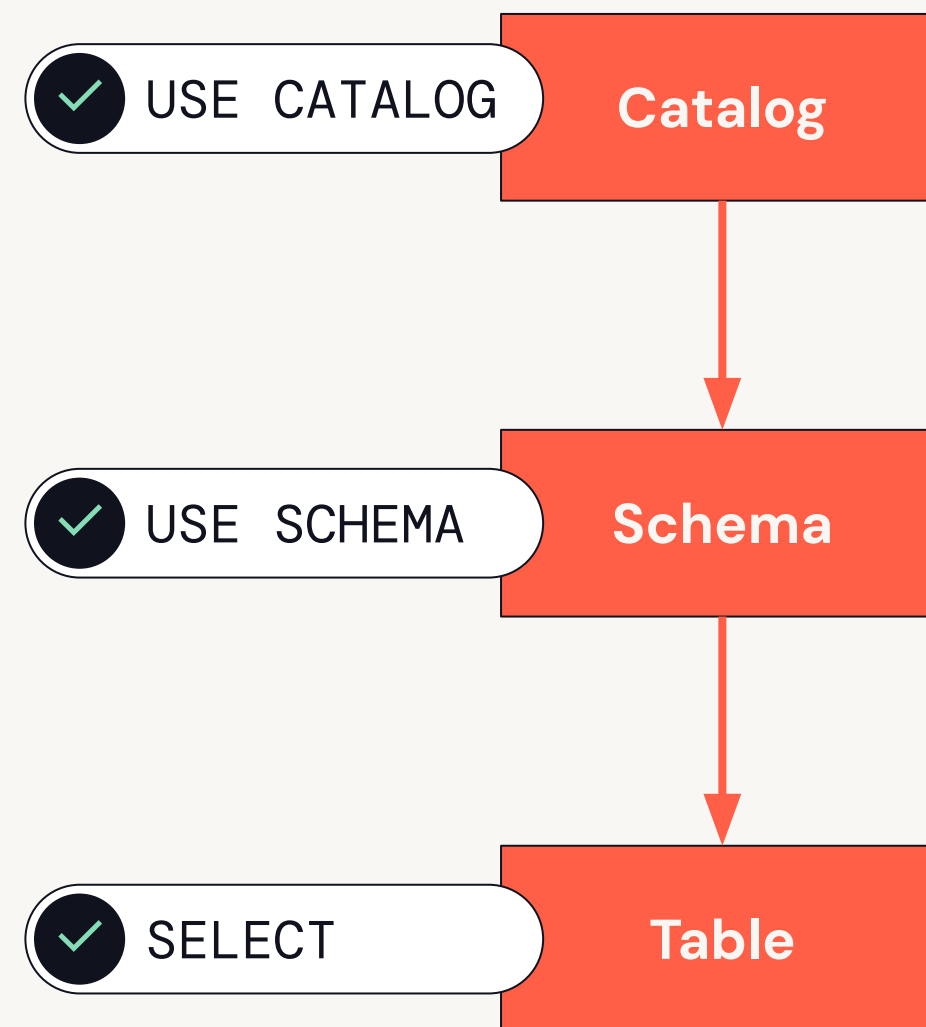
Querying a table



Privilege Scenarios

Querying a table

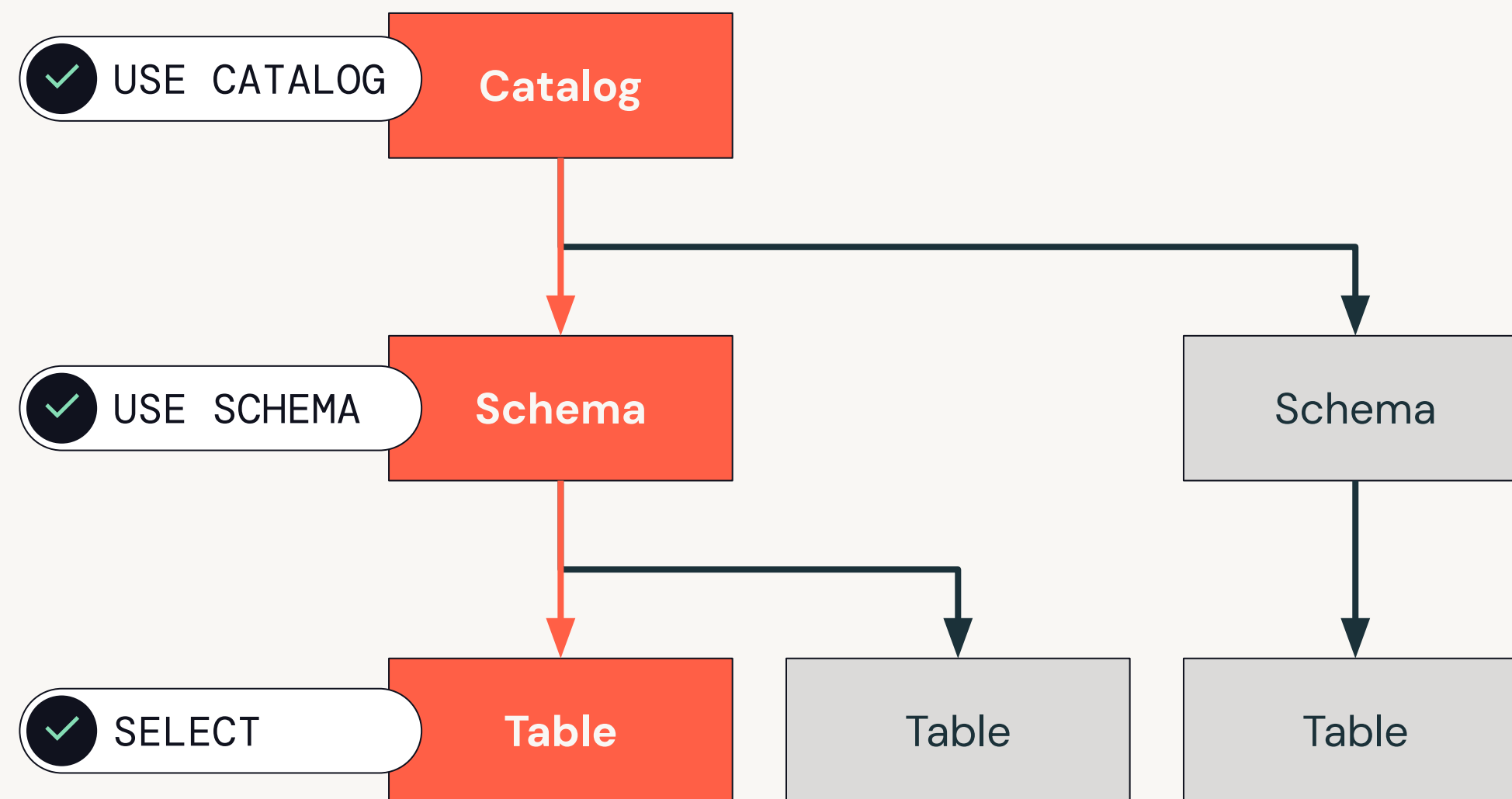
Explicit



Privilege Scenarios

Querying a table

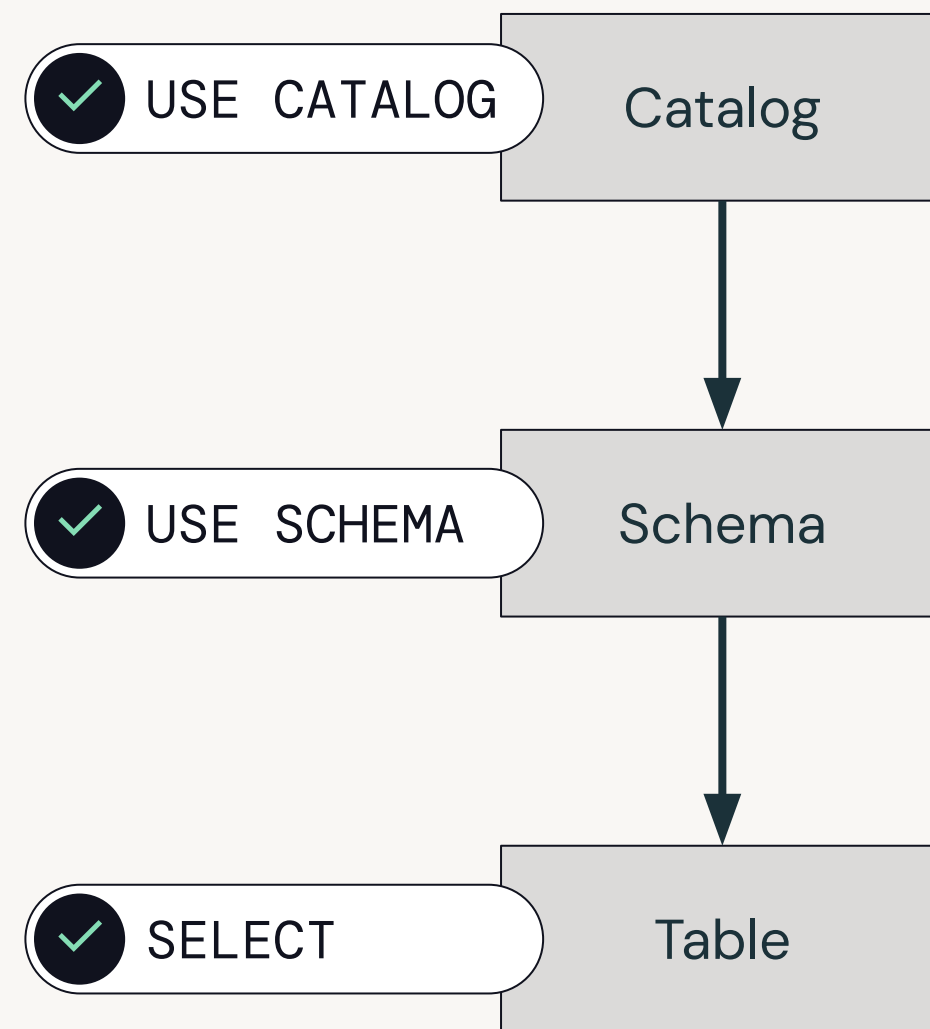
Explicit



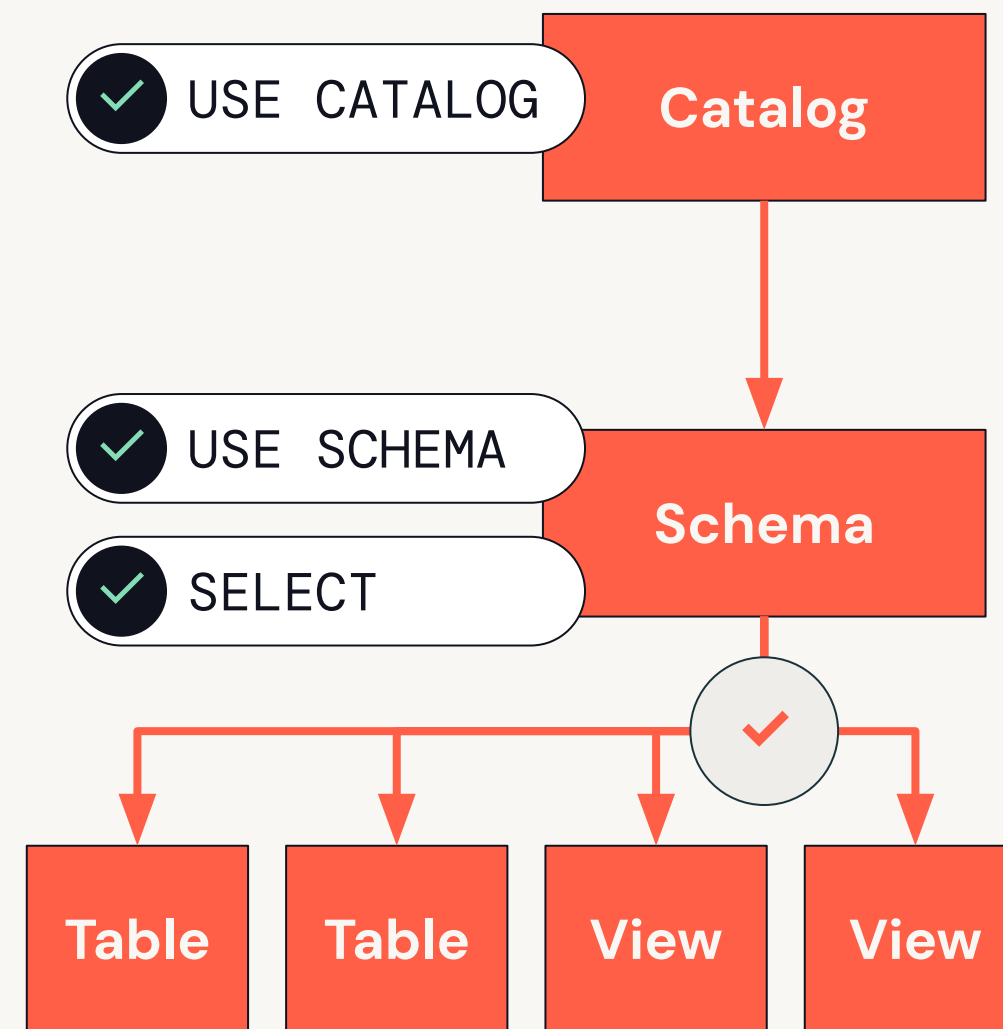
Privilege Scenarios

Querying a table

Explicit



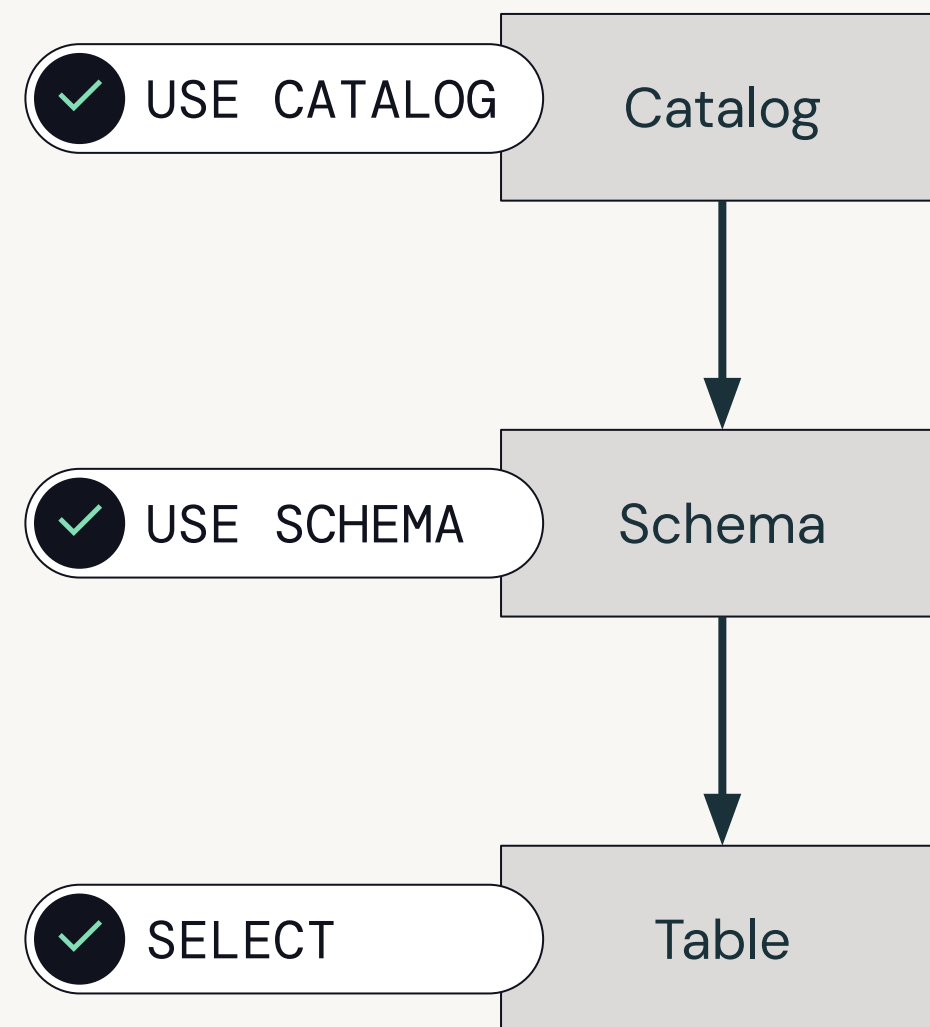
Inherited from schema



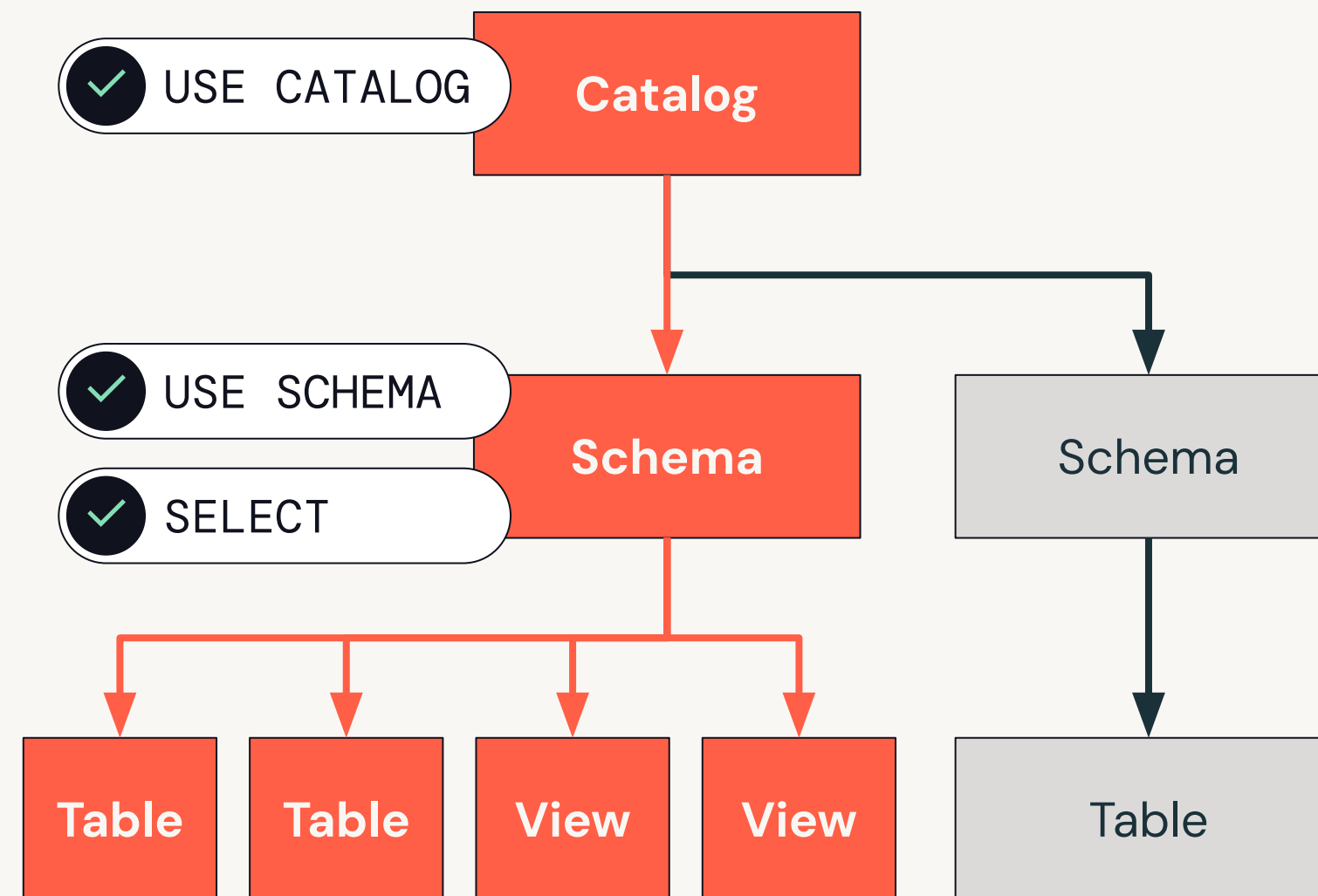
Privilege Scenarios

Querying a table

Explicit



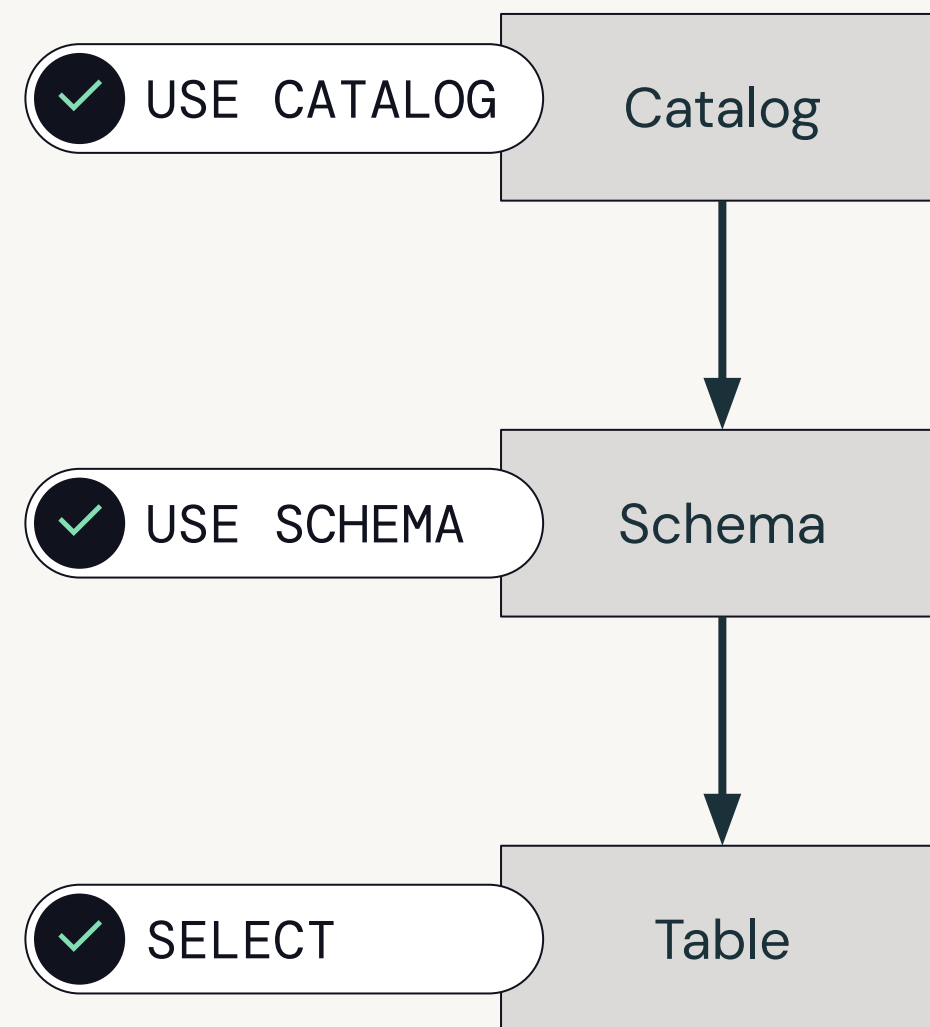
Inherited from schema



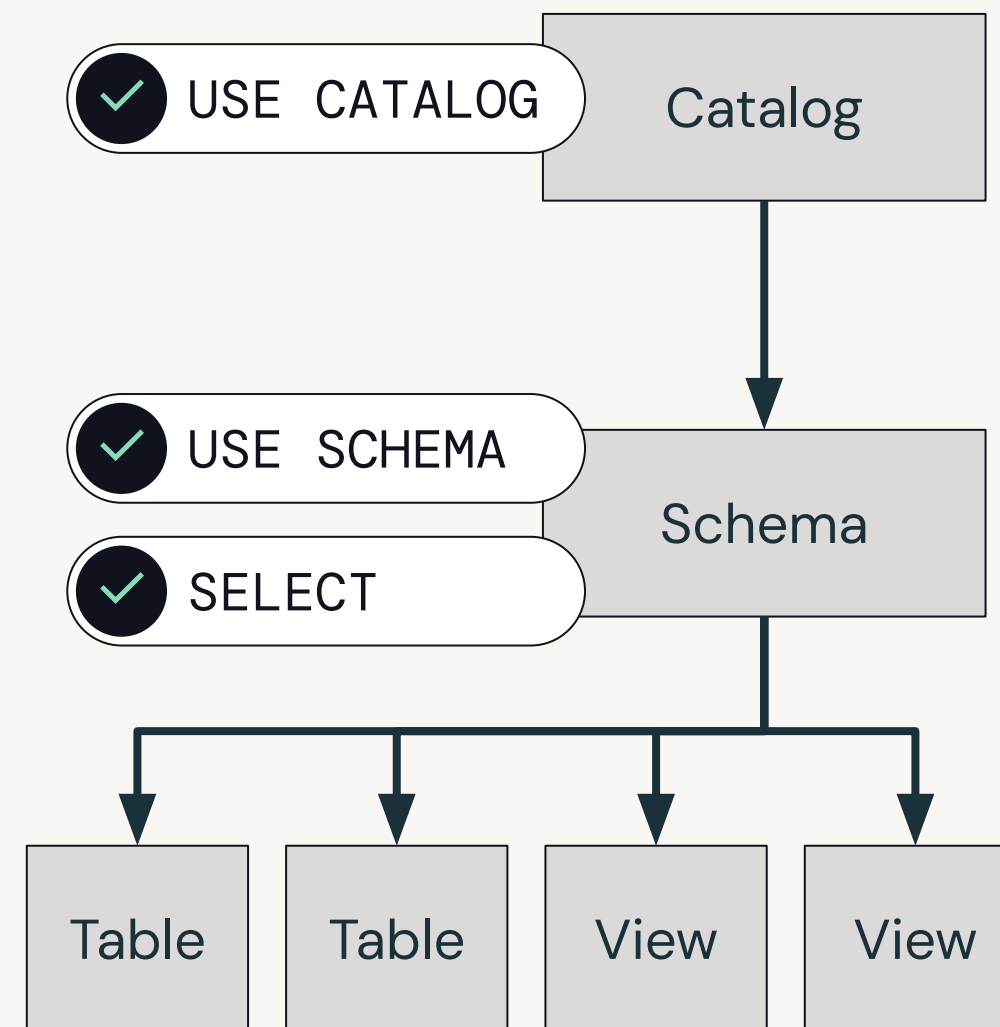
Privilege Scenarios

Querying a table

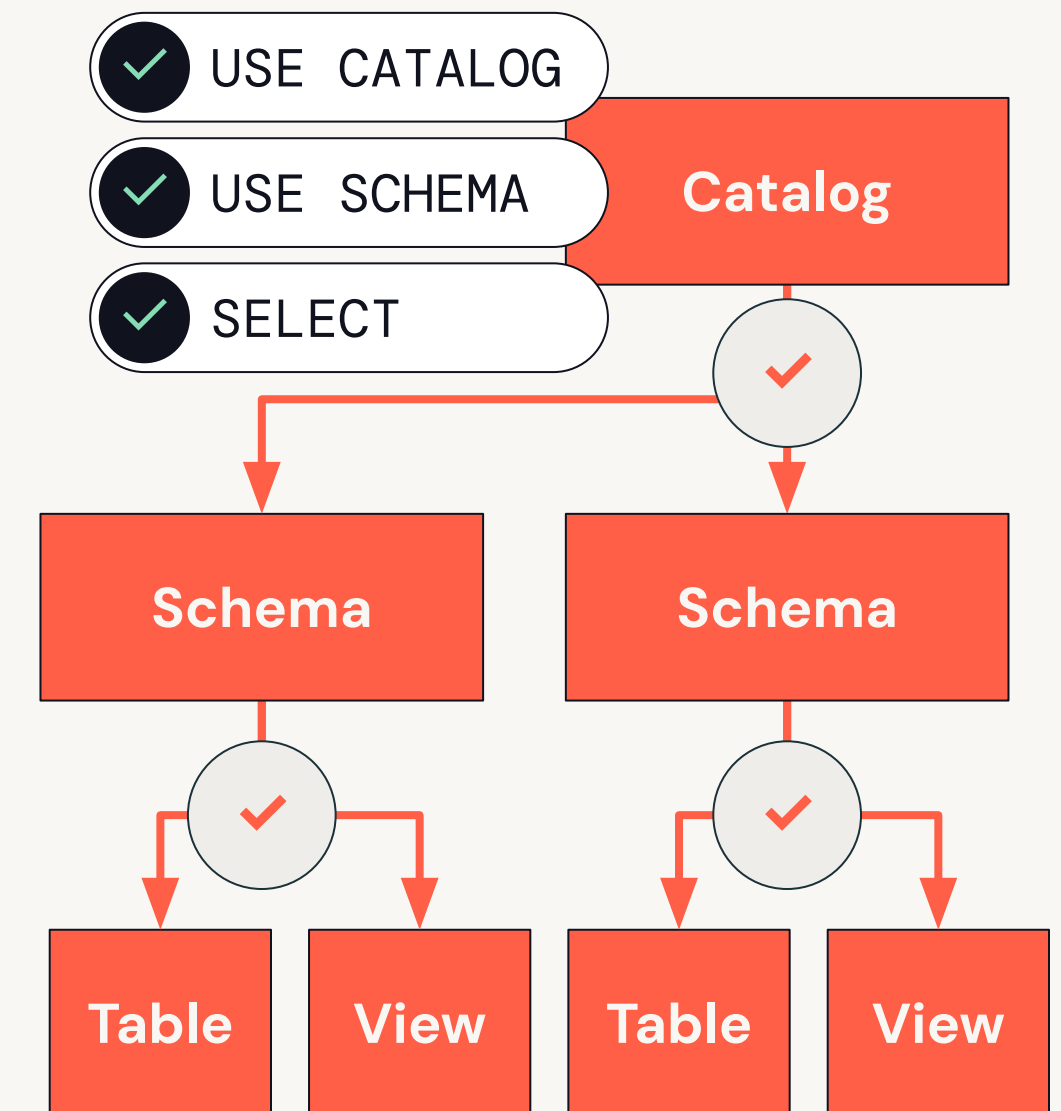
Explicit



Inherited from schema



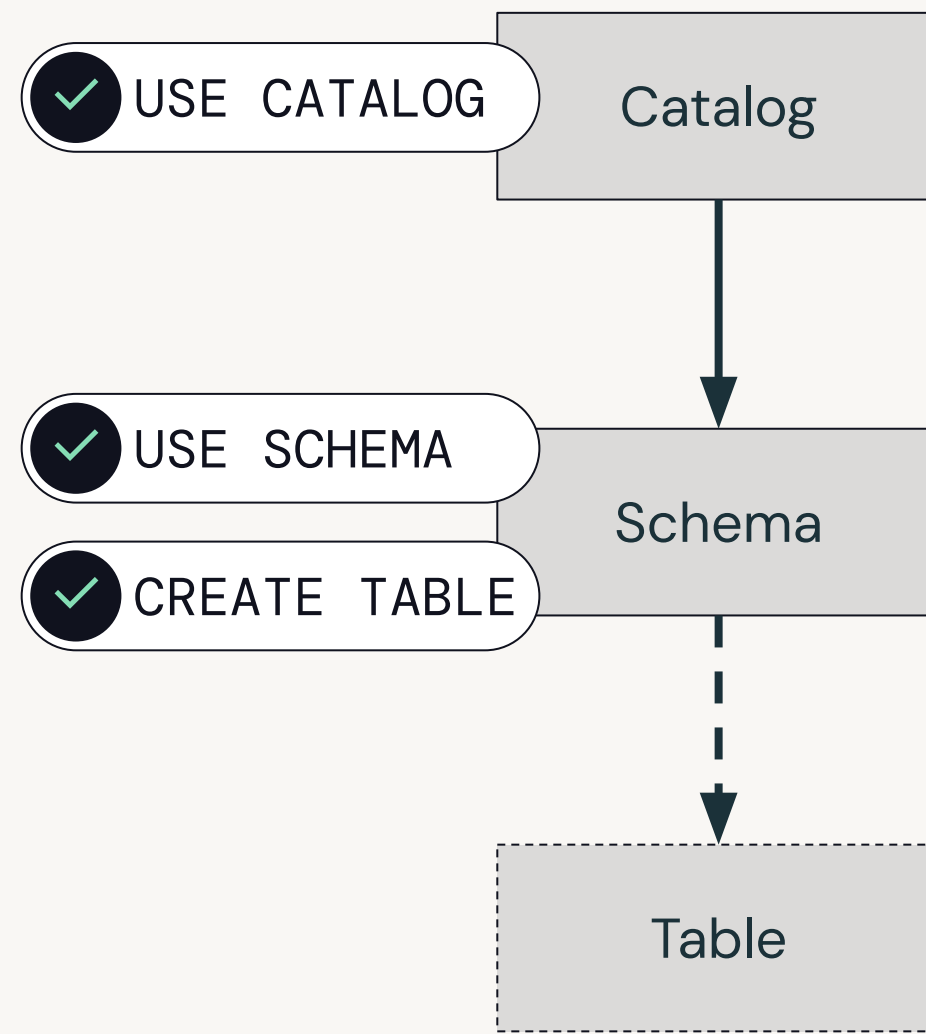
Inherited from catalog



Privilege Scenarios

Creating a table

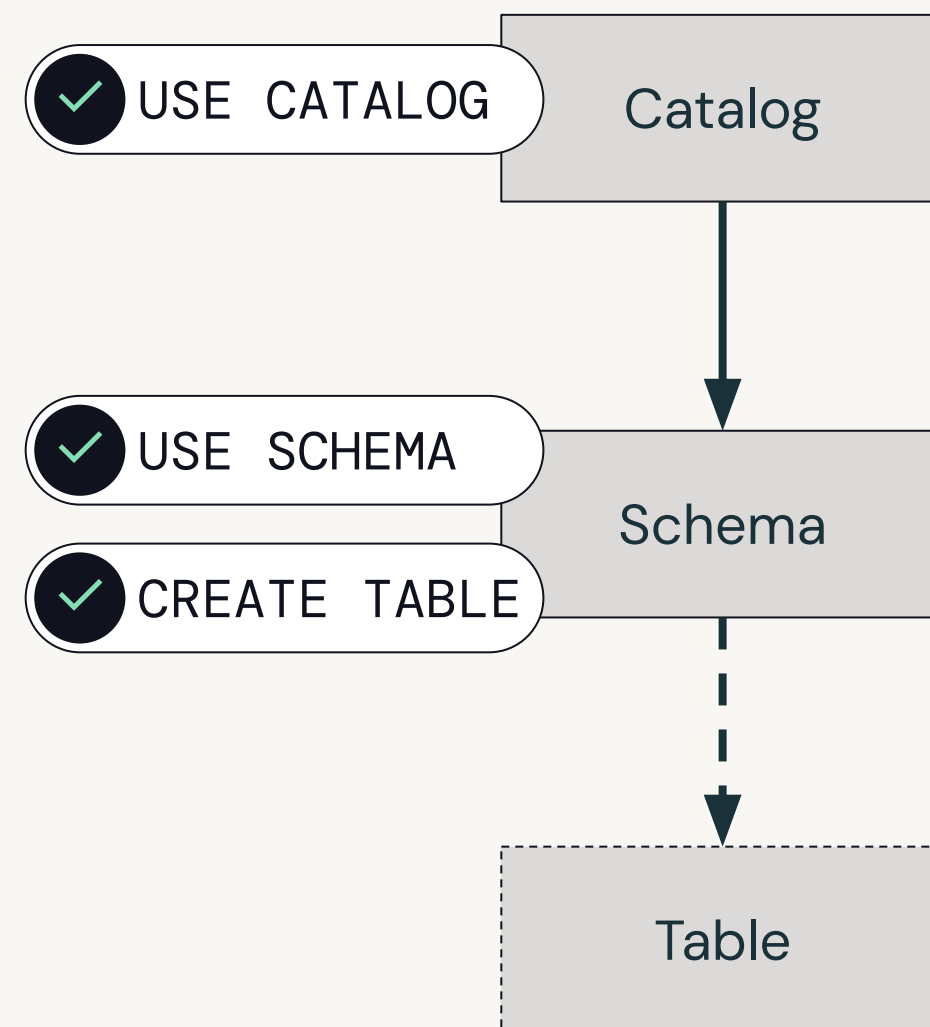
Explicit



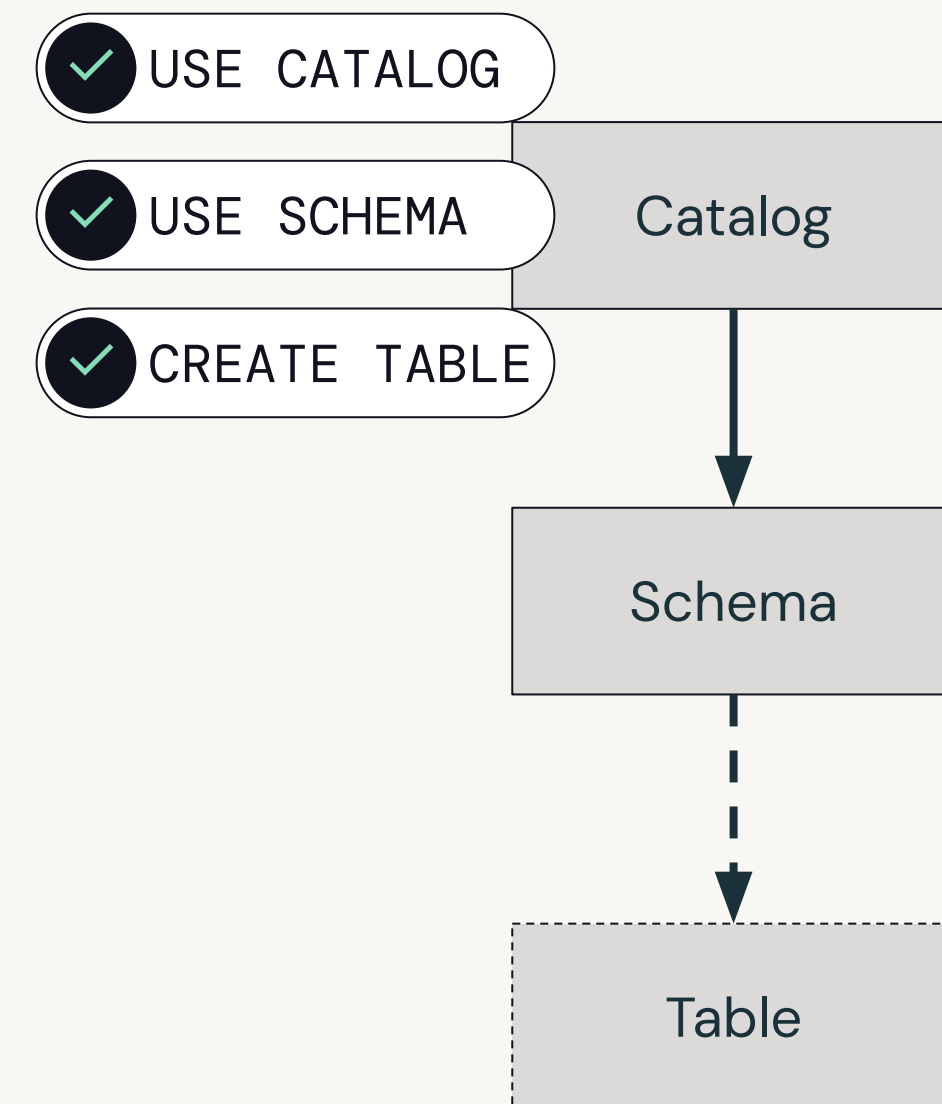
Privilege Scenarios

Creating a table

Explicit



Inherited from catalog



Lab: Controlling Access to Data



Learning Objectives

In this lab, you will learn how to:

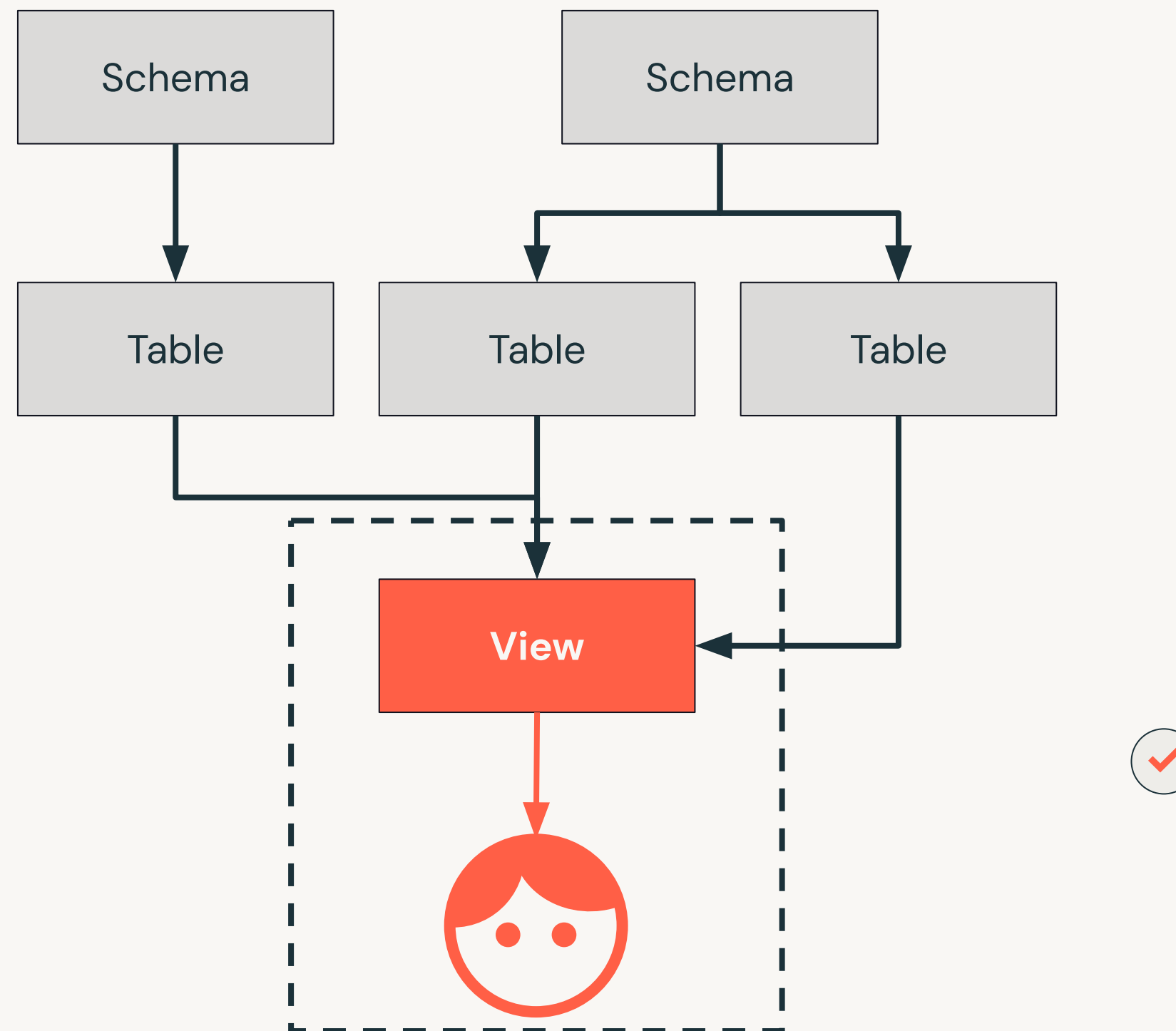
- Grant and revoke access to data objects
- Validate access
- Understand the difference between inherited and explicit grants



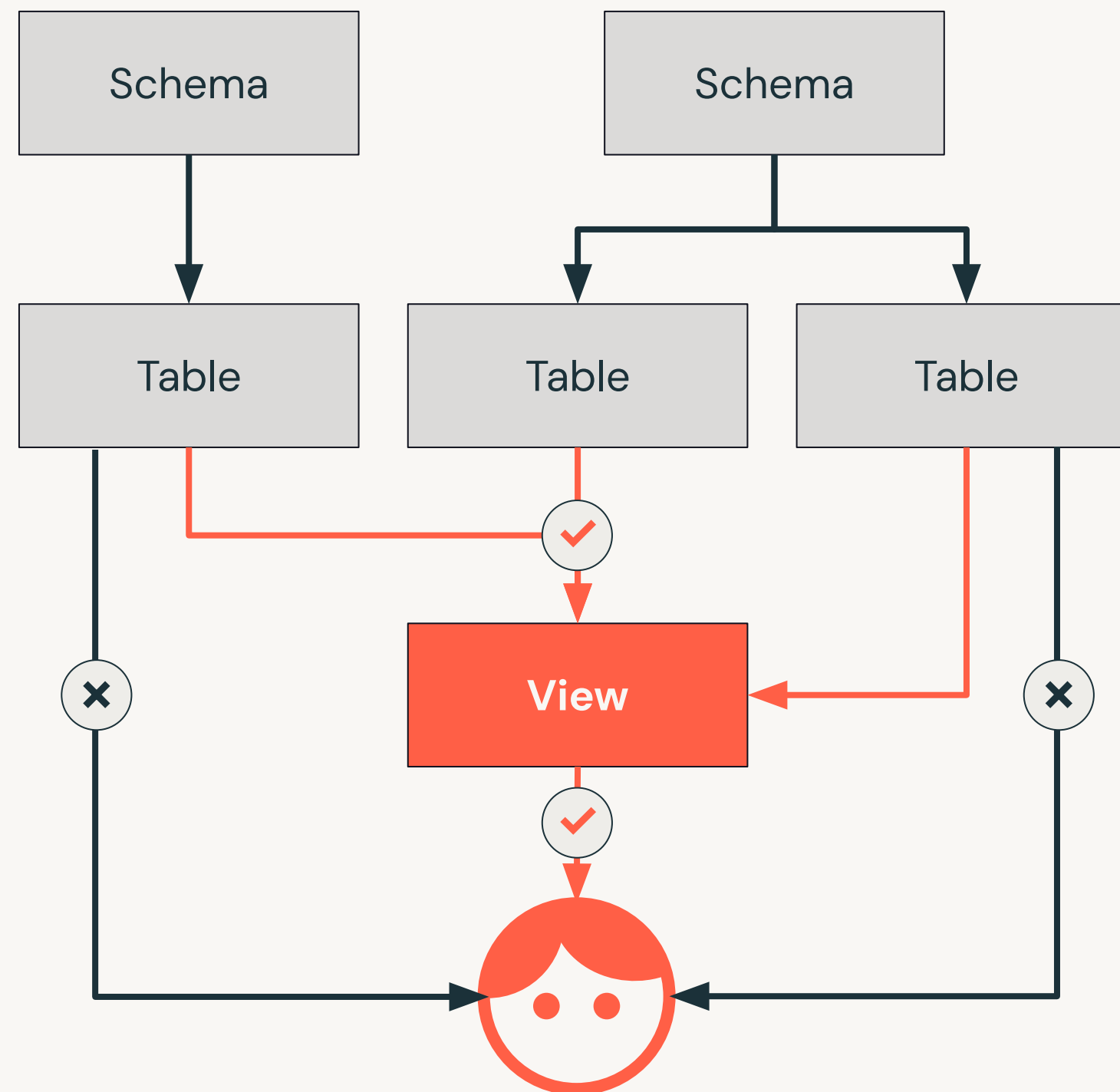
Dynamic Views



Dynamic Views



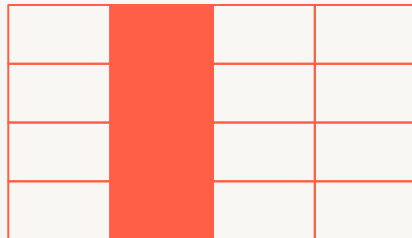
Dynamic Views



Dynamic Views

Three common use cases

Control access to columns



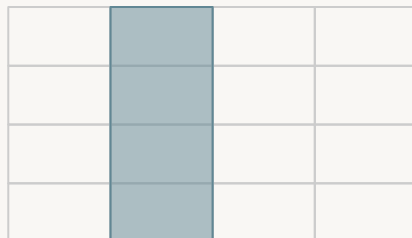
| | | | |
|--|--|--|--|
| | | | |
| | | | |
| | | | |
| | | | |



Dynamic Views

Three common use cases

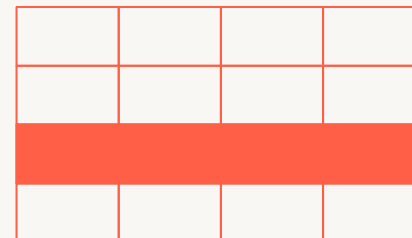
Control access to columns



A 4x4 grid representing a table. The second column from the left is highlighted in a darker blue, indicating that access is controlled at the column level.

| | | | |
|--|--|--|--|
| | | | |
| | | | |
| | | | |
| | | | |

Control access to rows



A 4x4 grid representing a table. The third row from the top is highlighted in a darker orange, indicating that access is controlled at the row level.

| | | | |
|--|--|--|--|
| | | | |
| | | | |
| | | | |
| | | | |



Dynamic Views

Three common use cases

Control access to columns

| | | | |
|--|--|--|--|
| | | | |
| | | | |
| | | | |
| | | | |

Control access to rows

| | | | |
|--|--|--|--|
| | | | |
| | | | |
| | | | |
| | | | |

Mask data

.....@databricks.com



Demo: Protecting Columns and Rows



Learning Objectives

In this lab, you will learn how to:

- Redact columns
- Filter rows
- Mask data



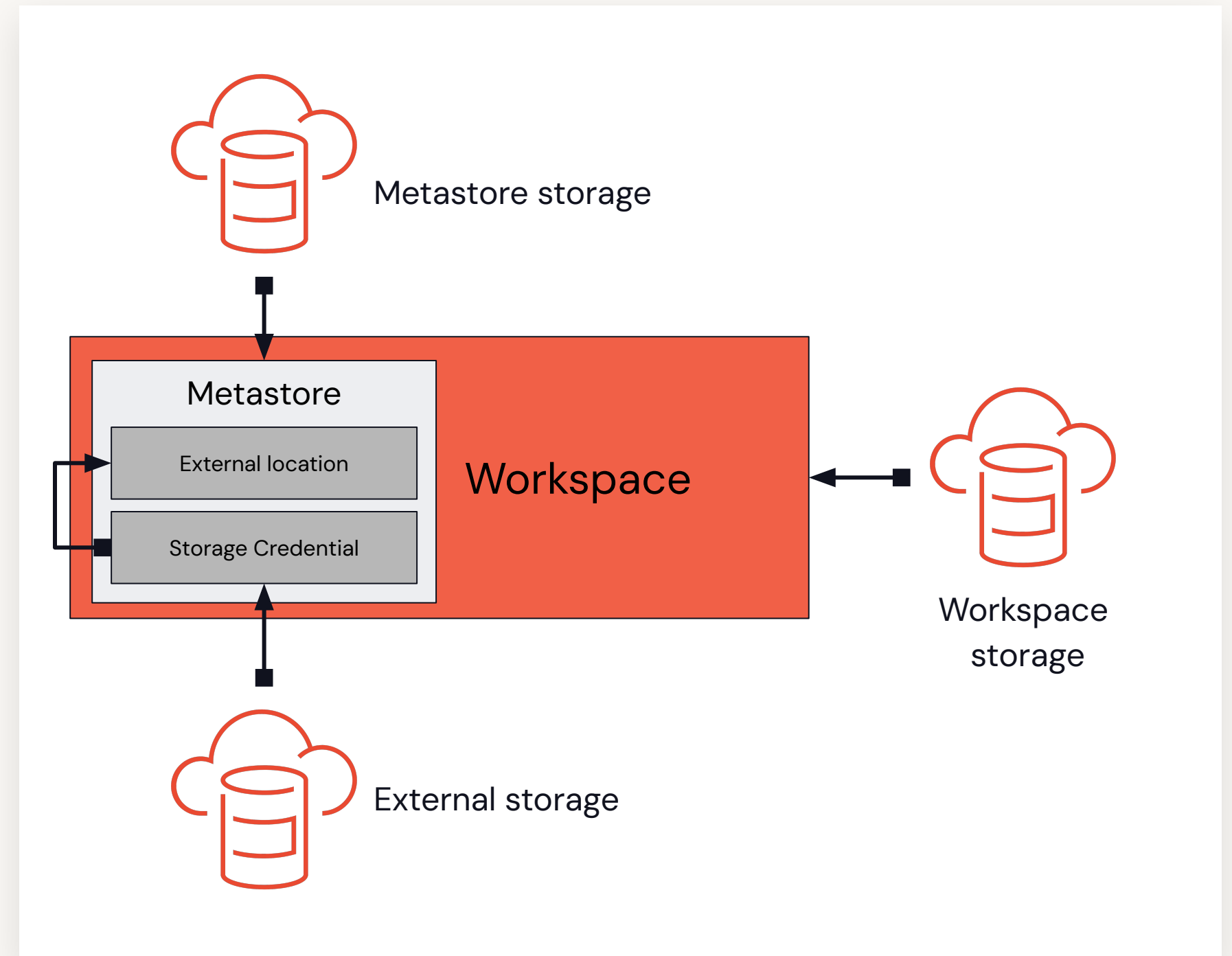
External Data Storage

External Data Storage

Integrates arbitrary cloud storage containers to address important use cases

- Data exploration
- Data ingestion
- External tables

Can be done any time

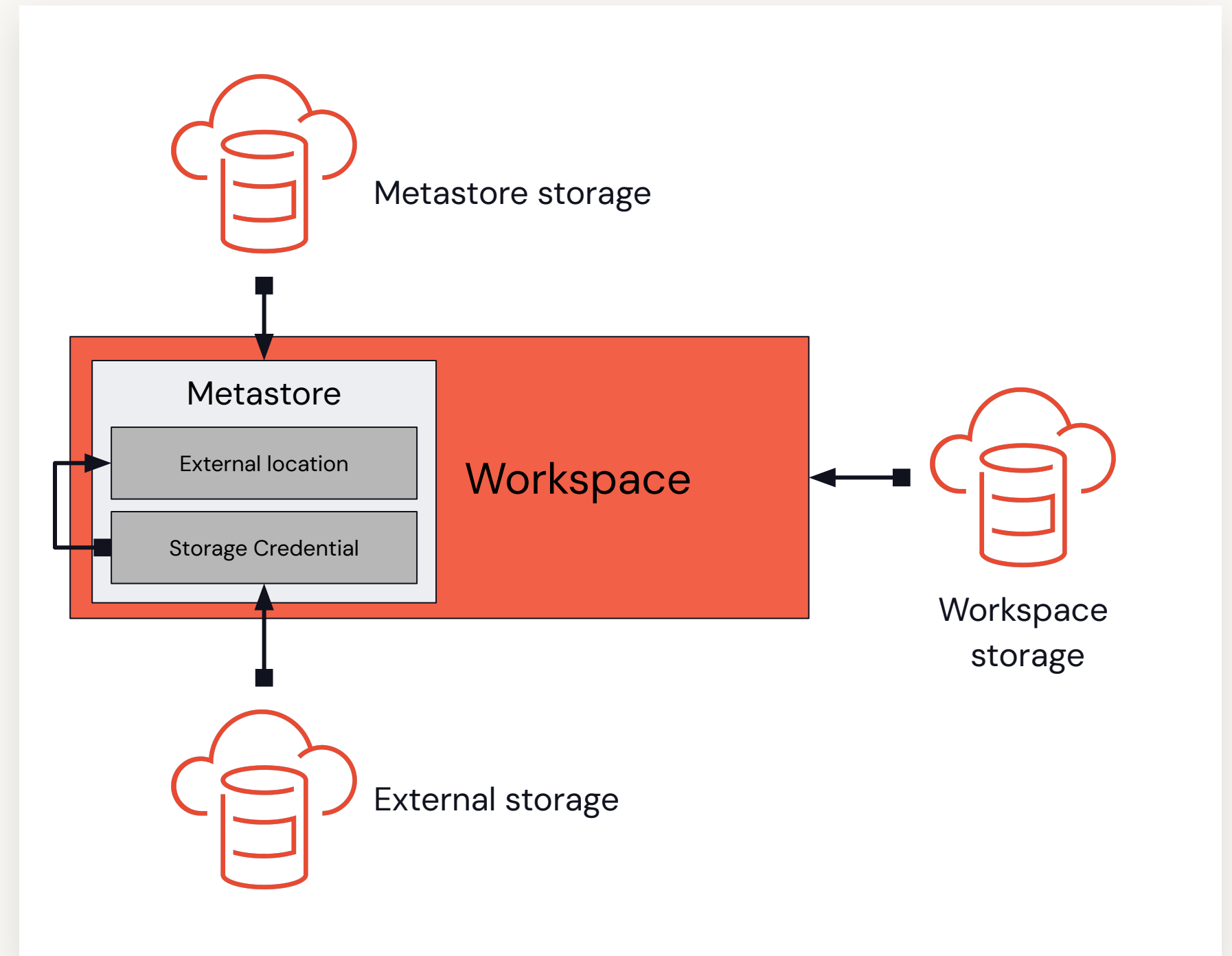


External Data Storage

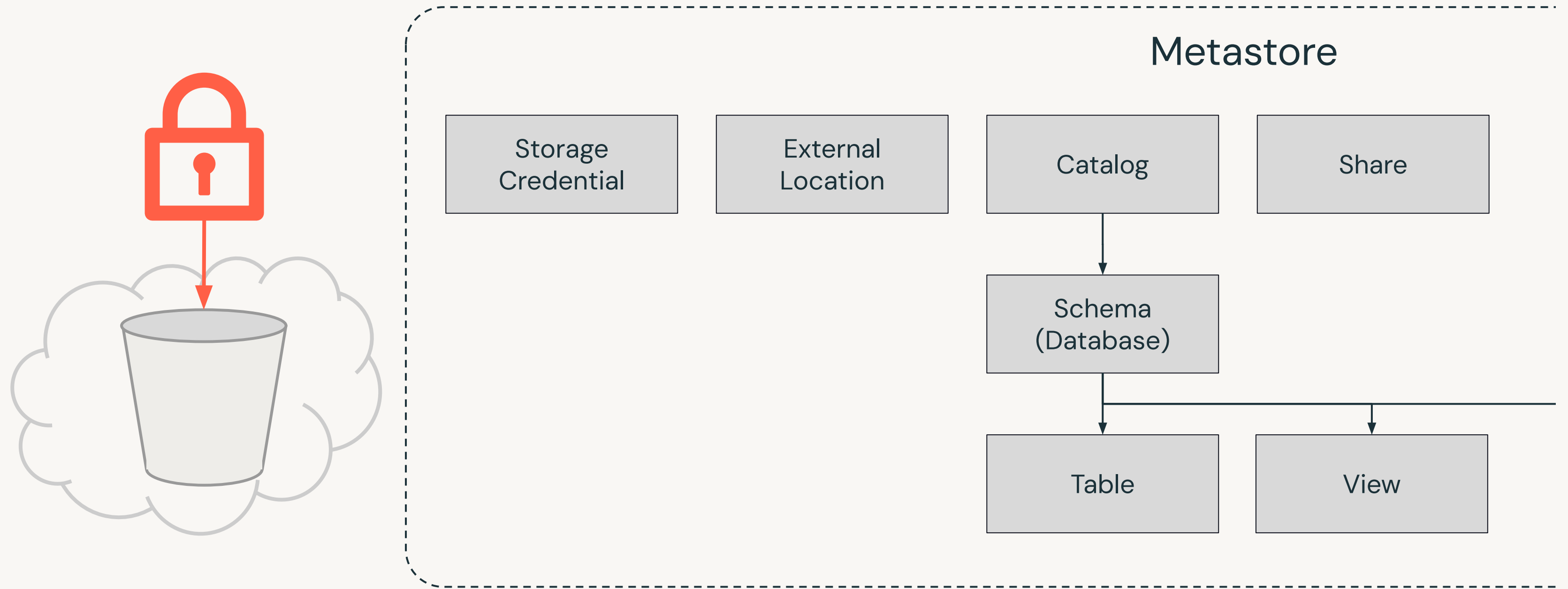
Integrates arbitrary cloud storage containers to address important use cases

- Data exploration
- Data ingestion
- External tables

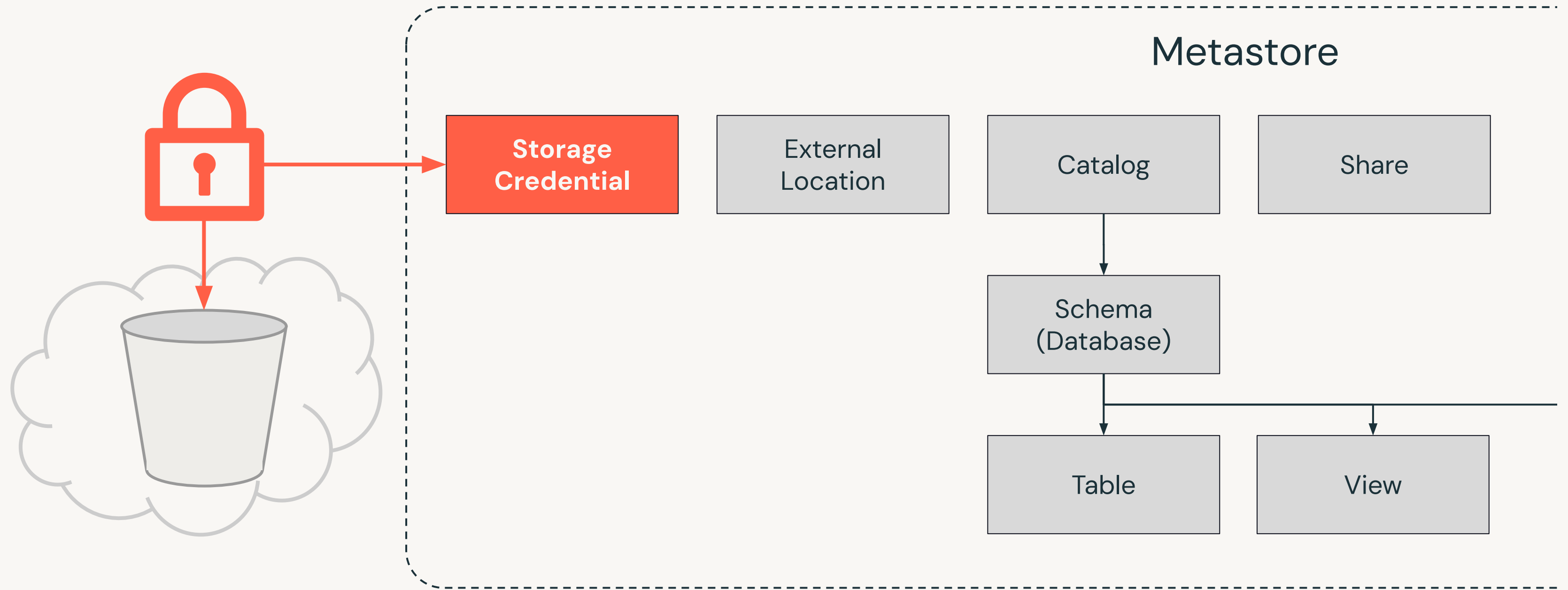
Can be done any time



External Data Storage

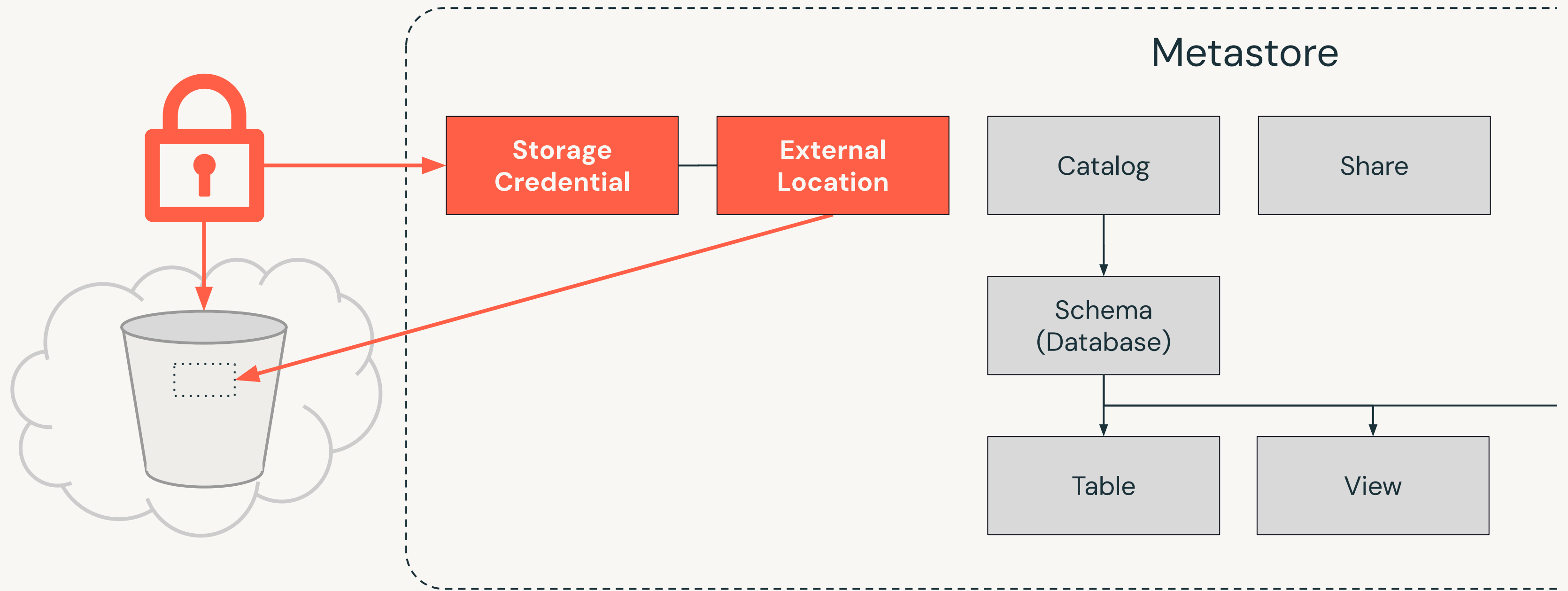


Storage credential



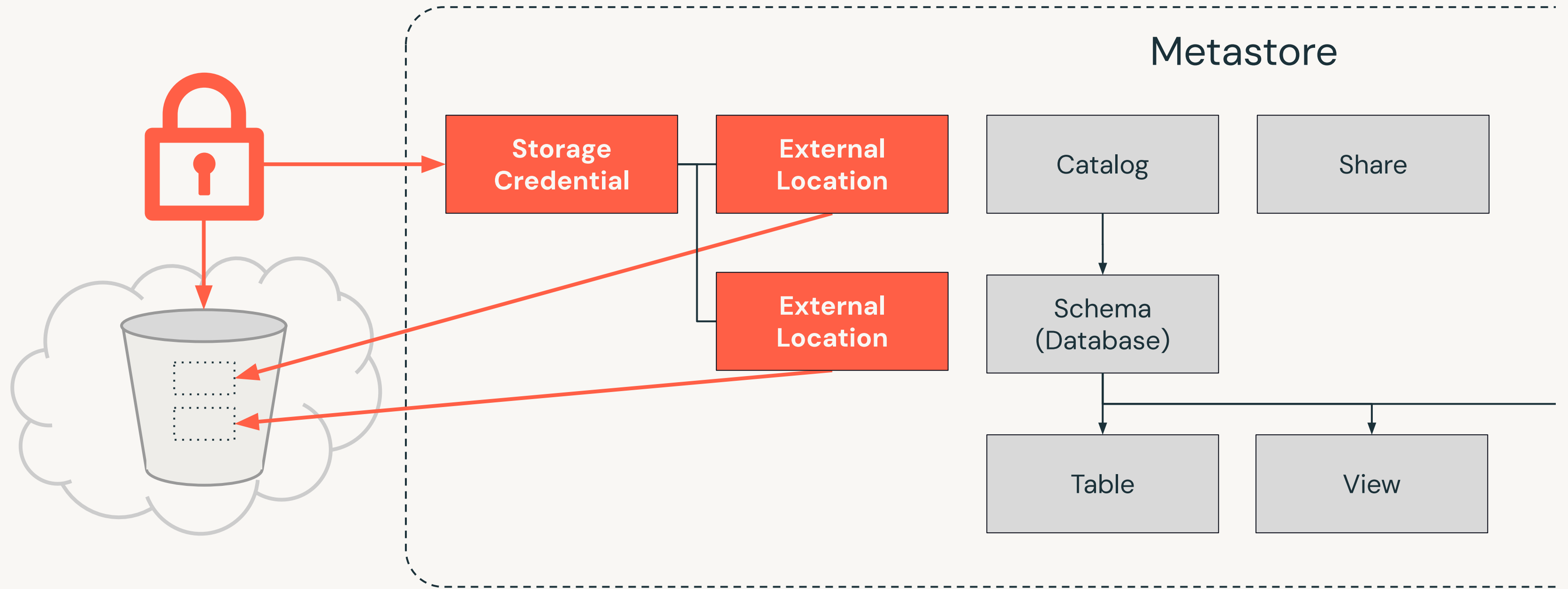
External Data Storage

External location



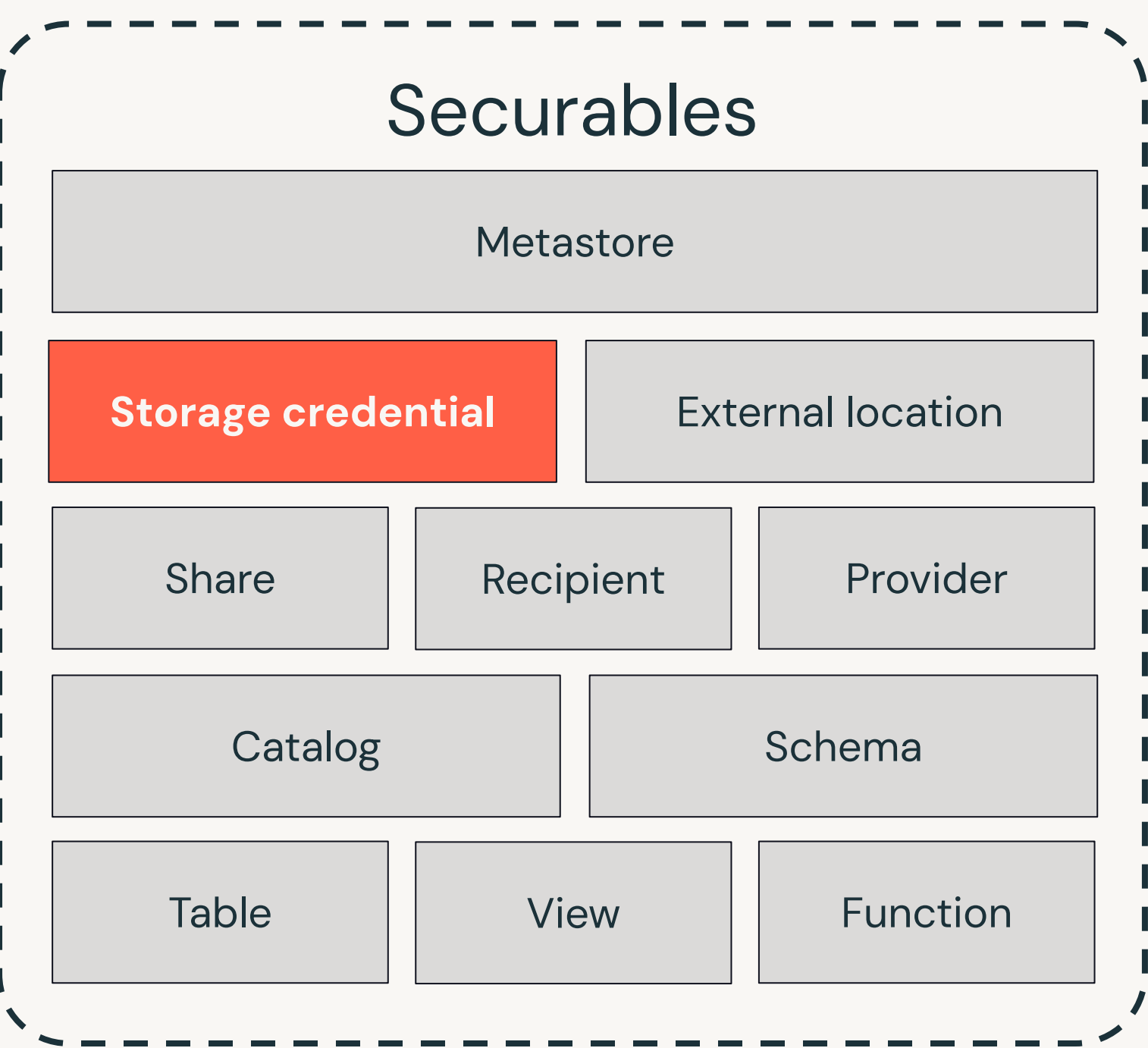
External Data Storage

External location



External Data Storage

Privilege types



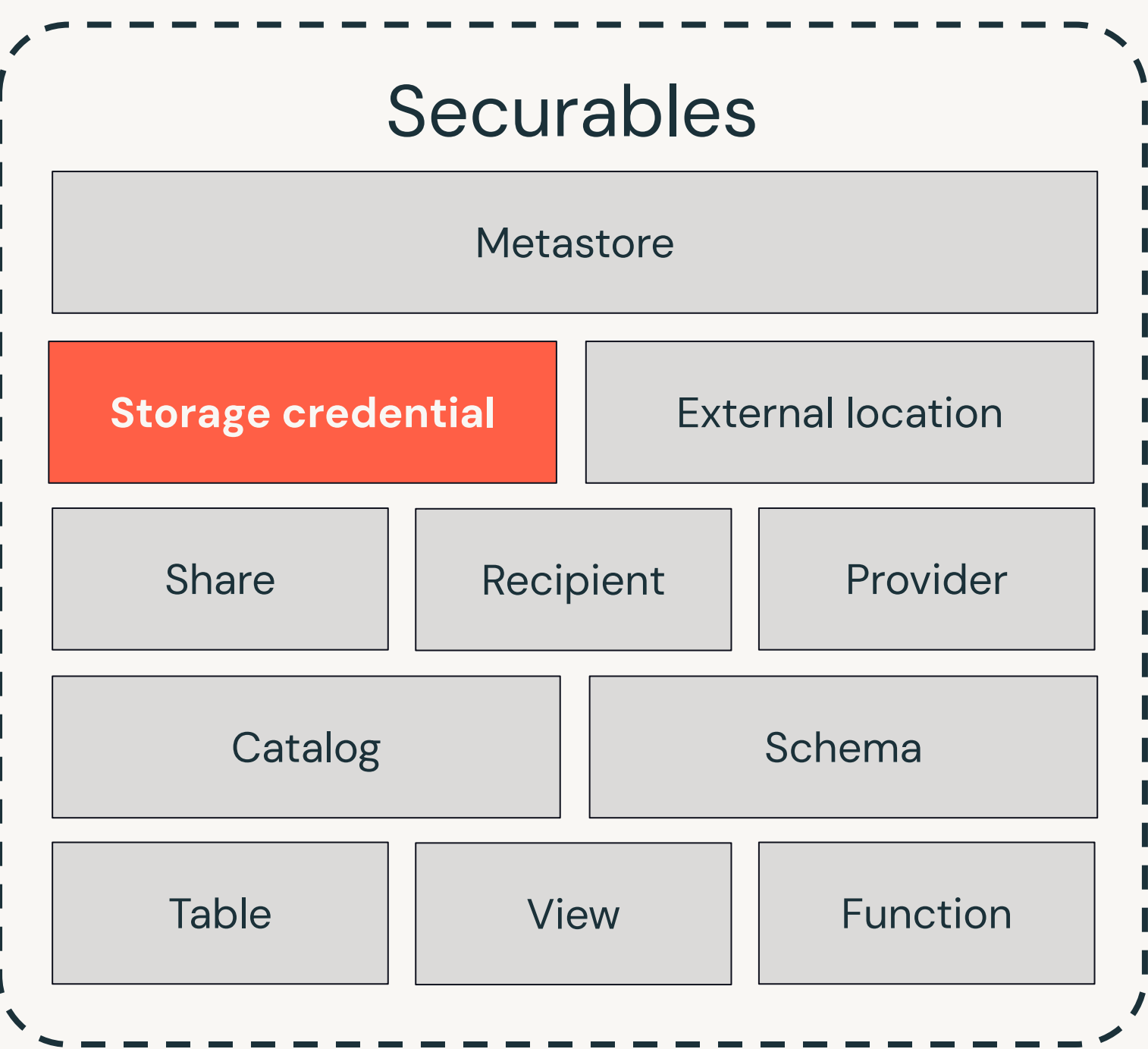
Privileges

CREATE EXTERNAL LOCATION
Create referencing external location



External Data Storage

Privilege types



Privileges

CREATE EXTERNAL LOCATION

Create referencing external location

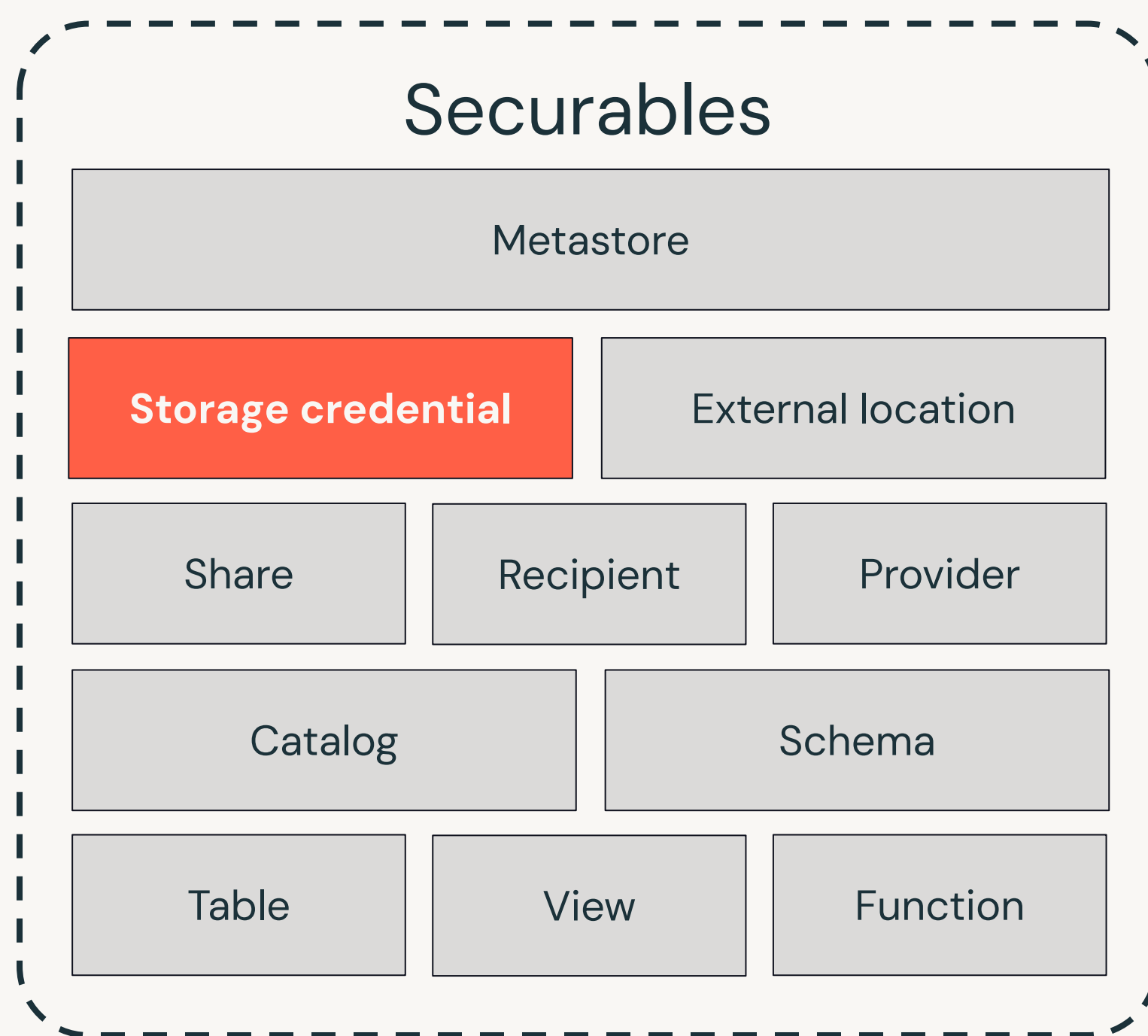
CREATE EXTERNAL TABLE

Create table based on contained files



External Data Storage

Privilege types



Privileges

CREATE EXTERNAL LOCATION

Create referencing external location

CREATE EXTERNAL TABLE

Create table based on contained files

READ FILES

Read contained files

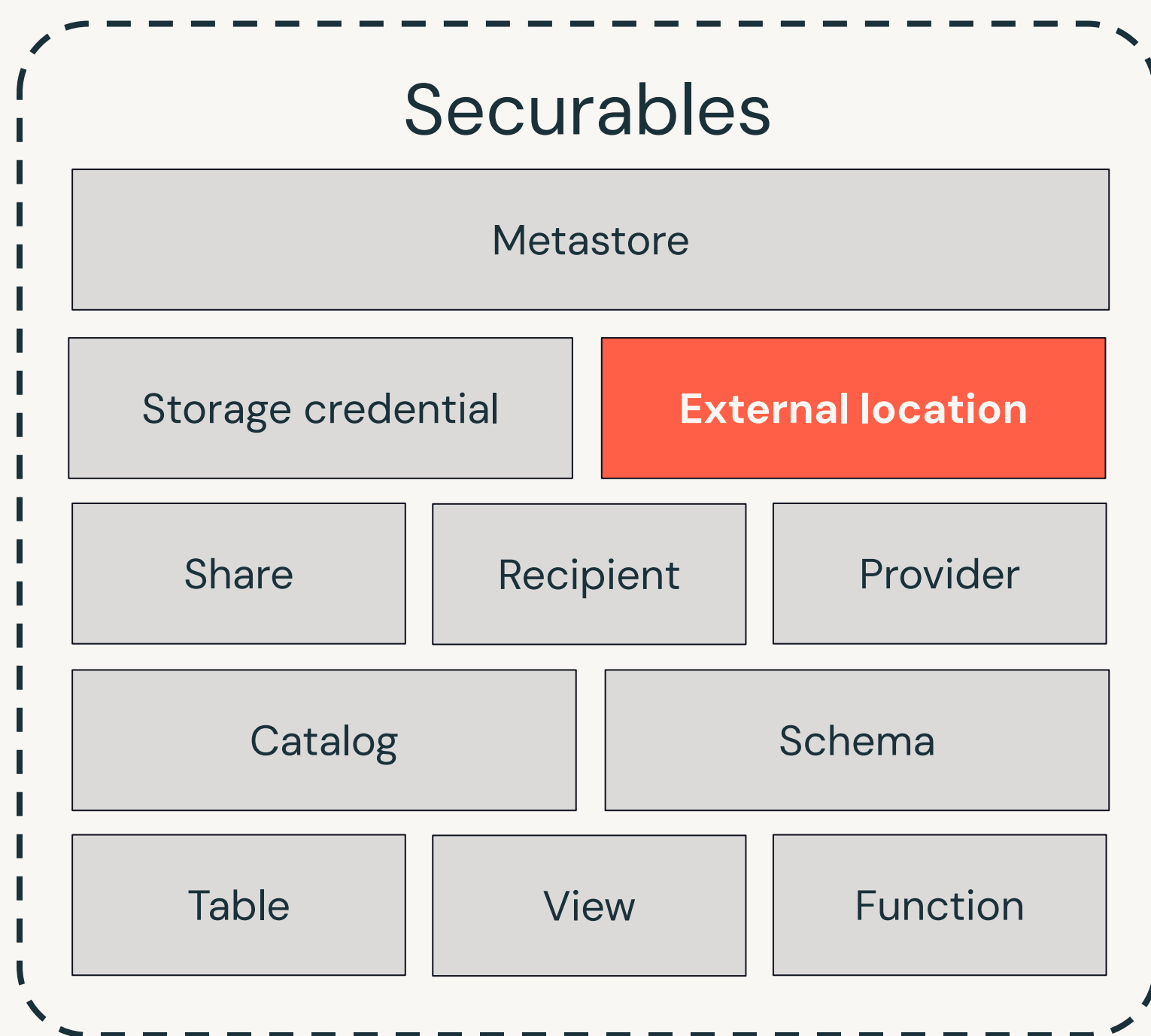
WRITE FILES

Modify contained files



External Data Storage

Privilege types



Privileges

CREATE EXTERNAL TABLE

Create table based on contained files

READ FILES

Read contained files

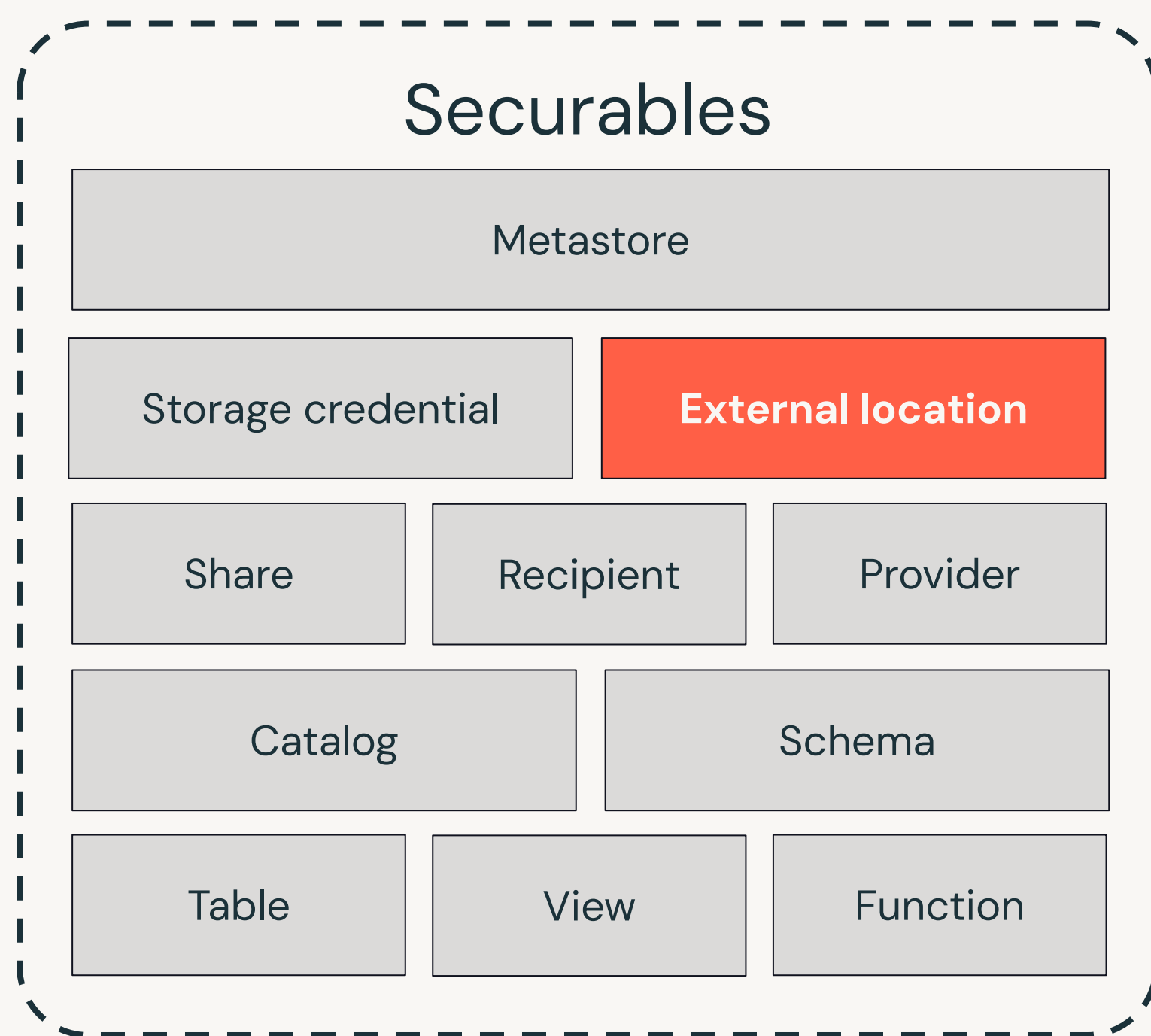
WRITE FILES

Modify contained files



External Data Storage

Privilege types



Privileges

CREATE EXTERNAL TABLE

Create table based on contained files

READ FILES

Read contained files

WRITE FILES

Modify contained files

CREATE MANAGED STORAGE

Create referencing external location



External Data Storage

Storage Credential

Enables Unity Catalog to access cloud data storage

Securable object

- ACL applies to entire storage container

External Location

Combines storage credential and cloud storage path

Securable object

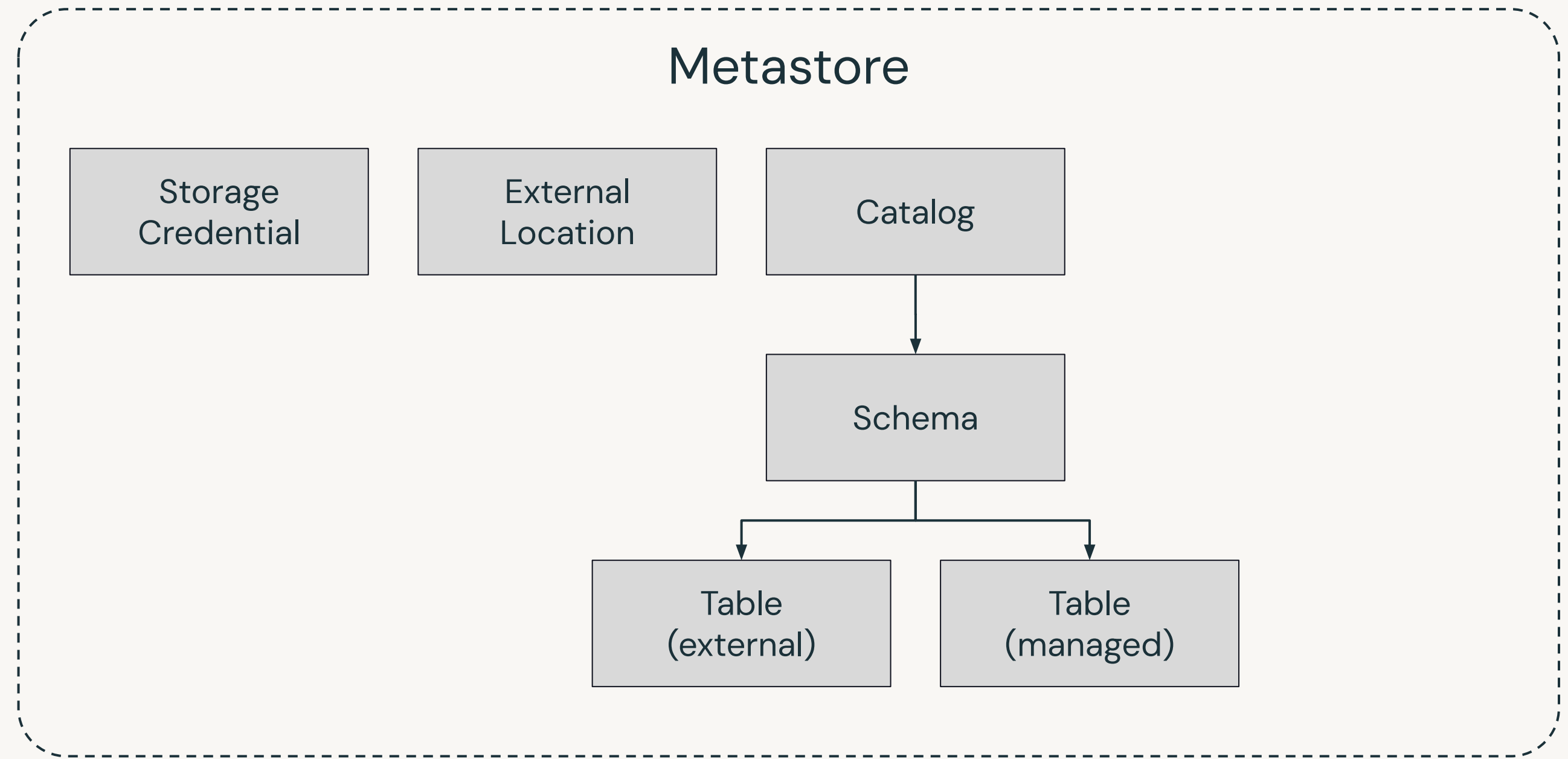
- ACL applies to the portion of the storage container represented by the path



Managed and External Tables

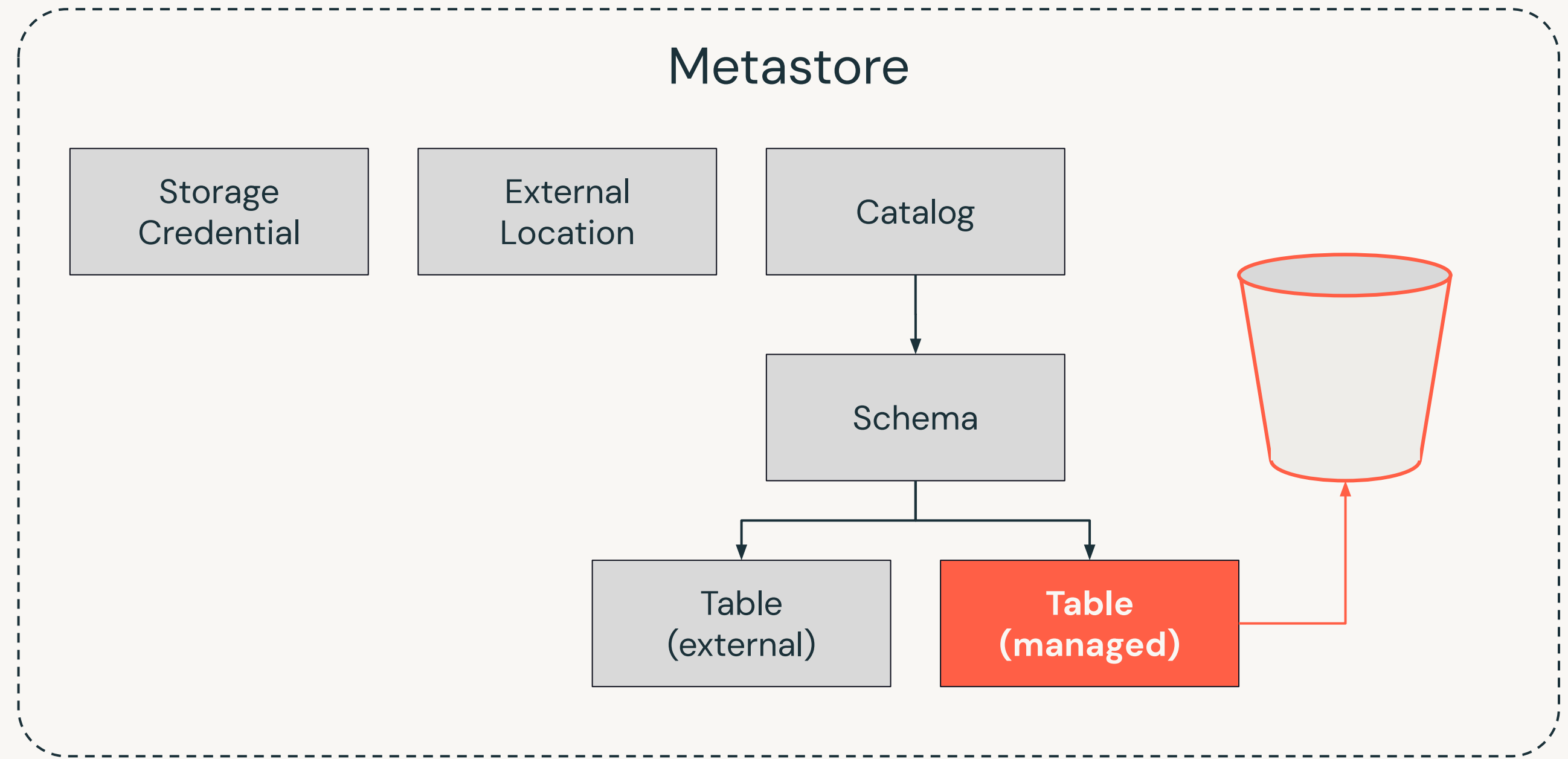
Managed and External Tables

Tables



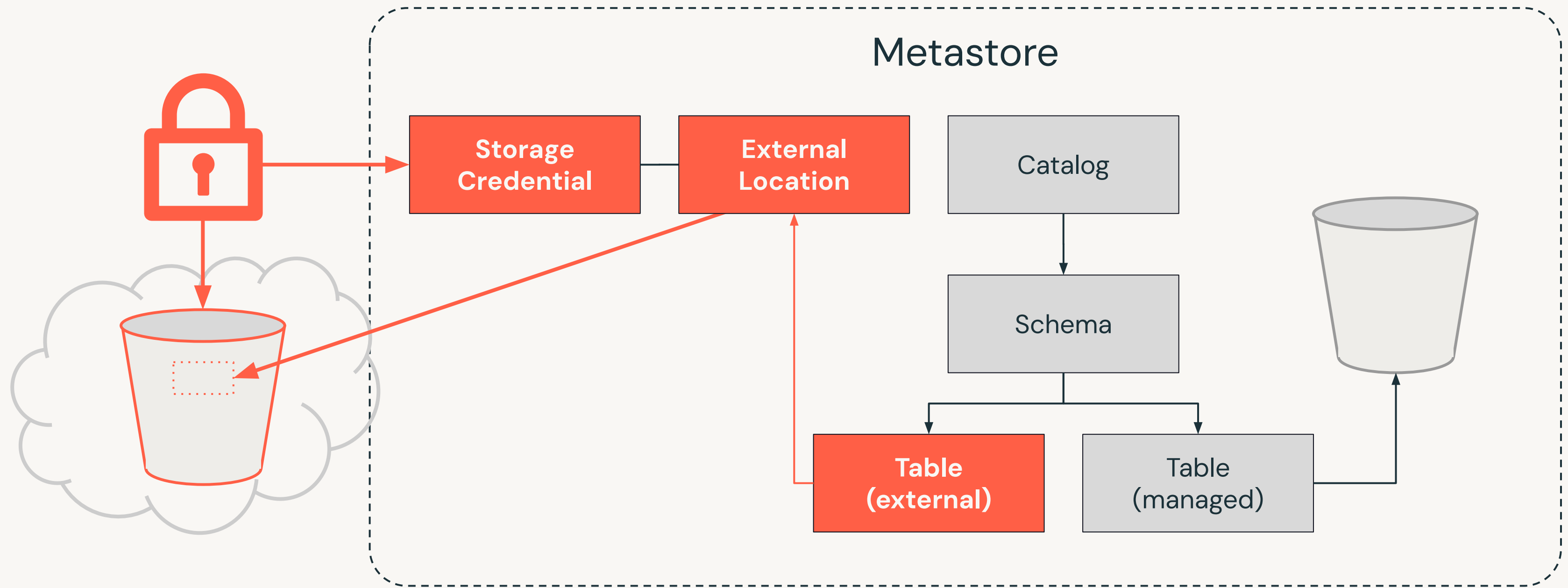
Managed and External Tables

Managed tables



Managed and External Tables

External tables



Managed and External Tables

Recap

Managed Tables

Metadata lives in control plane

Data lives in metastore managed storage location

DROP discards metadata and data

Delta format only

New product features supported first

External Tables

Metadata lives in control plane

Data lives in user-provided storage location

DROP discards metadata only

Several formats supported



Managed and External Tables

Recap

Managed Tables

Metadata lives in control plane

Data lives in metastore managed storage location

DROP discards metadata and data

Delta format only

New product features supported first

External Tables

Metadata lives in control plane

Data lives in user-provided storage location

DROP discards metadata only

Several formats supported



Managed and External Tables

When to use external tables

Quick and easy upgrade from external table in Hive metastore



Managed and External Tables

When to use external tables

Quick and easy upgrade from external table in Hive metastore

External writers



Managed and External Tables

When to use external tables

Quick and easy upgrade from external table in Hive metastore

External writers

Non-Delta support requirement



Managed and External Tables

When to use external tables

Quick and easy upgrade from external table in Hive metastore

External writers

Non-Delta support requirement

Requirement for specific storage naming or hierarchy

Infrastructure-level isolation requirements



Demo: Managing External Storage



Demo: Managing External Tables



Learning Objectives

In this lab, you will learn how to:

- Create and and manage external tables
- Compare external and managed table behaviour

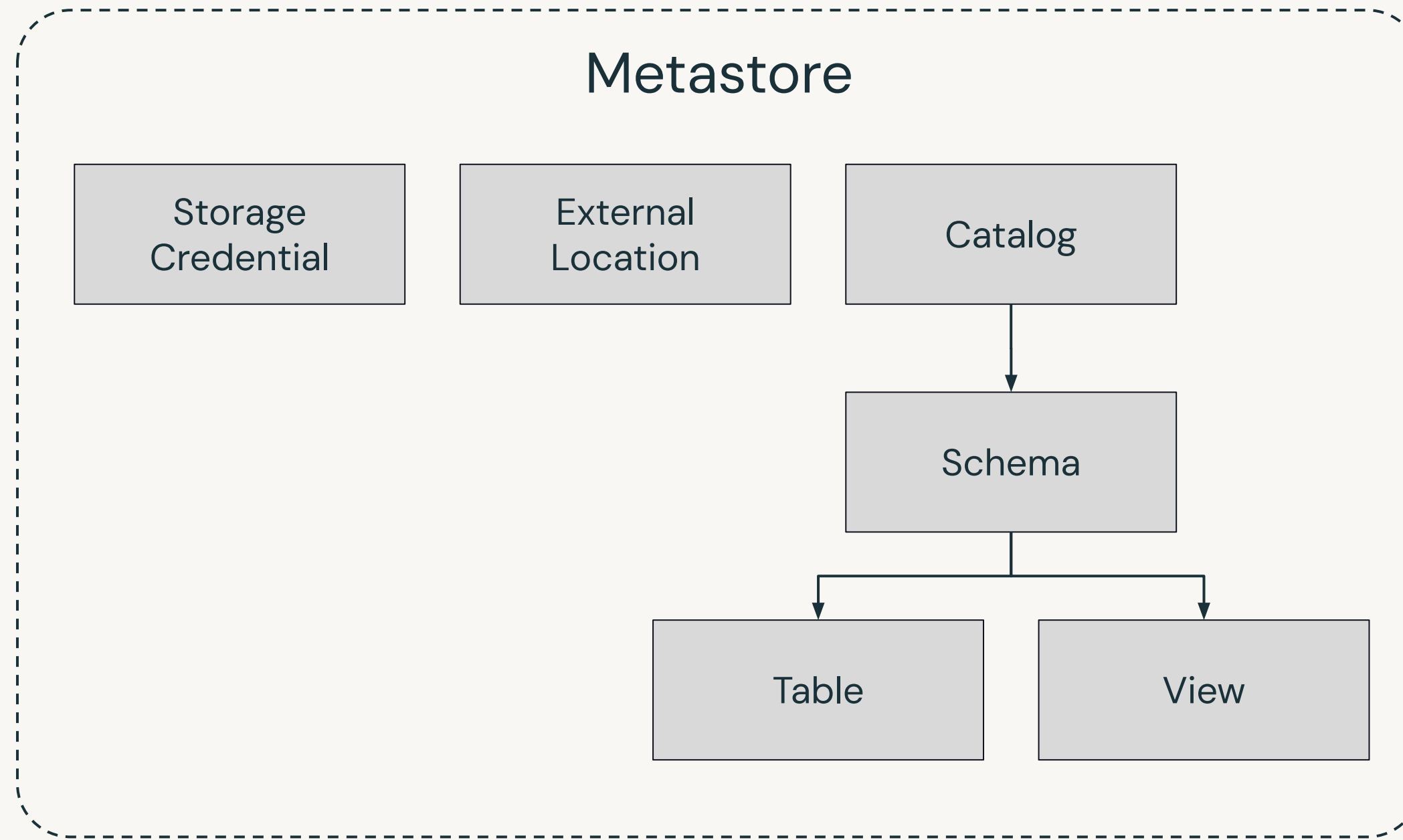


Additional Topics

Data Segregation

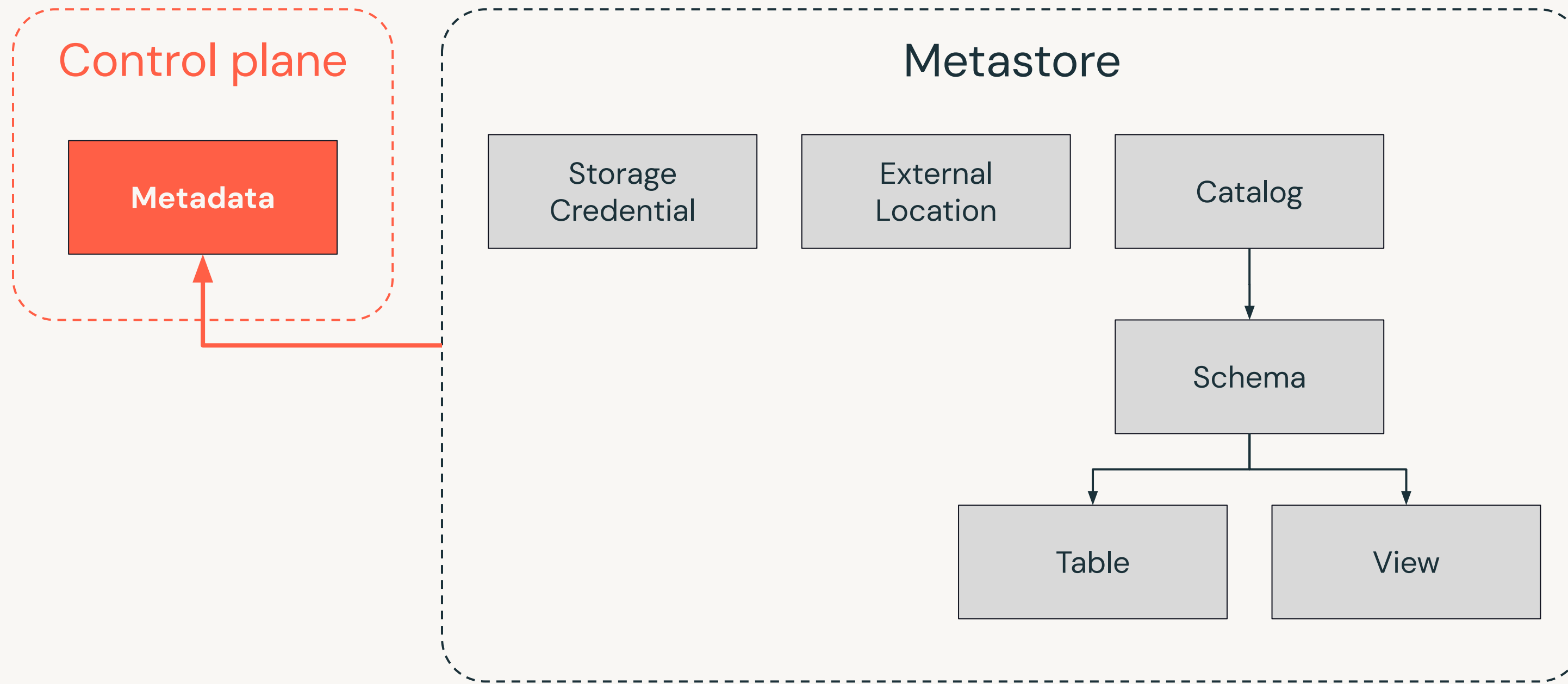


Data Segregation



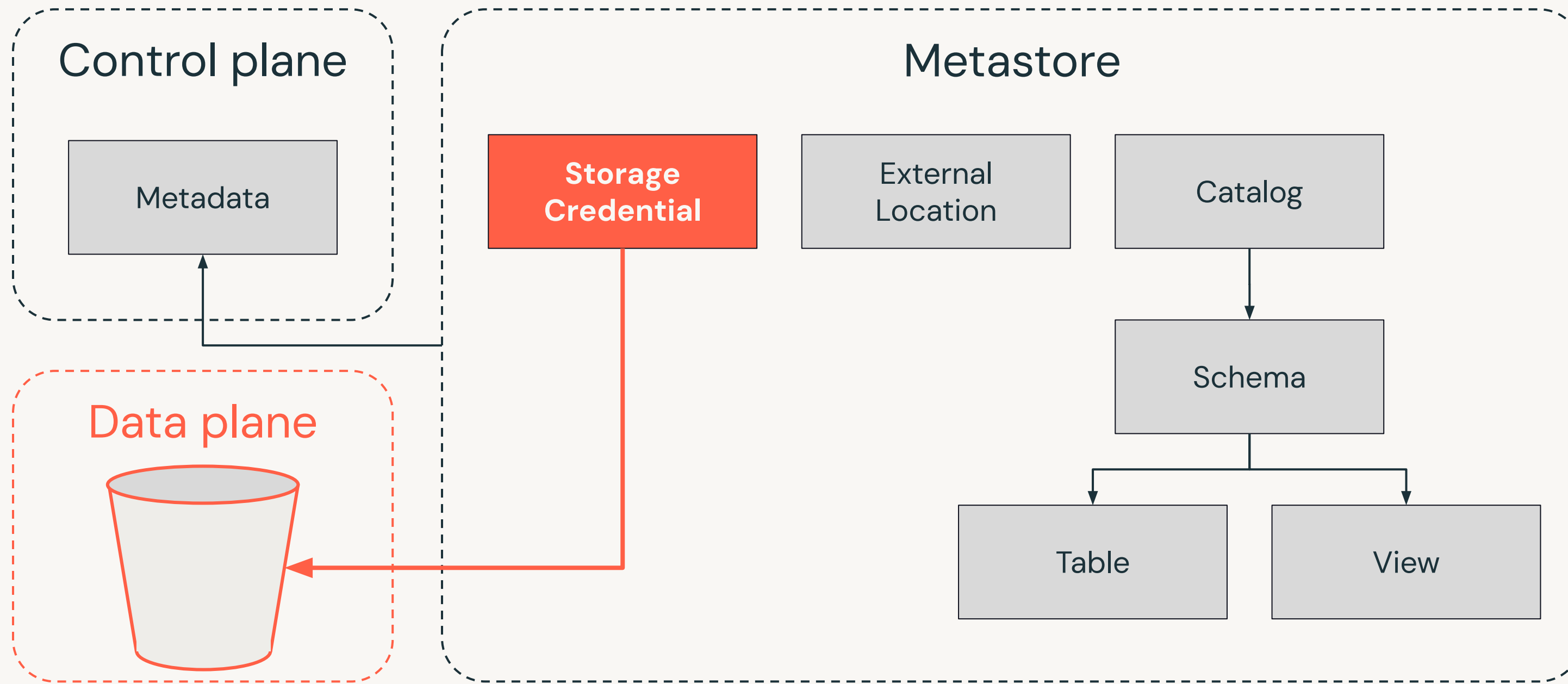
Data Segregation

Metastore requirements



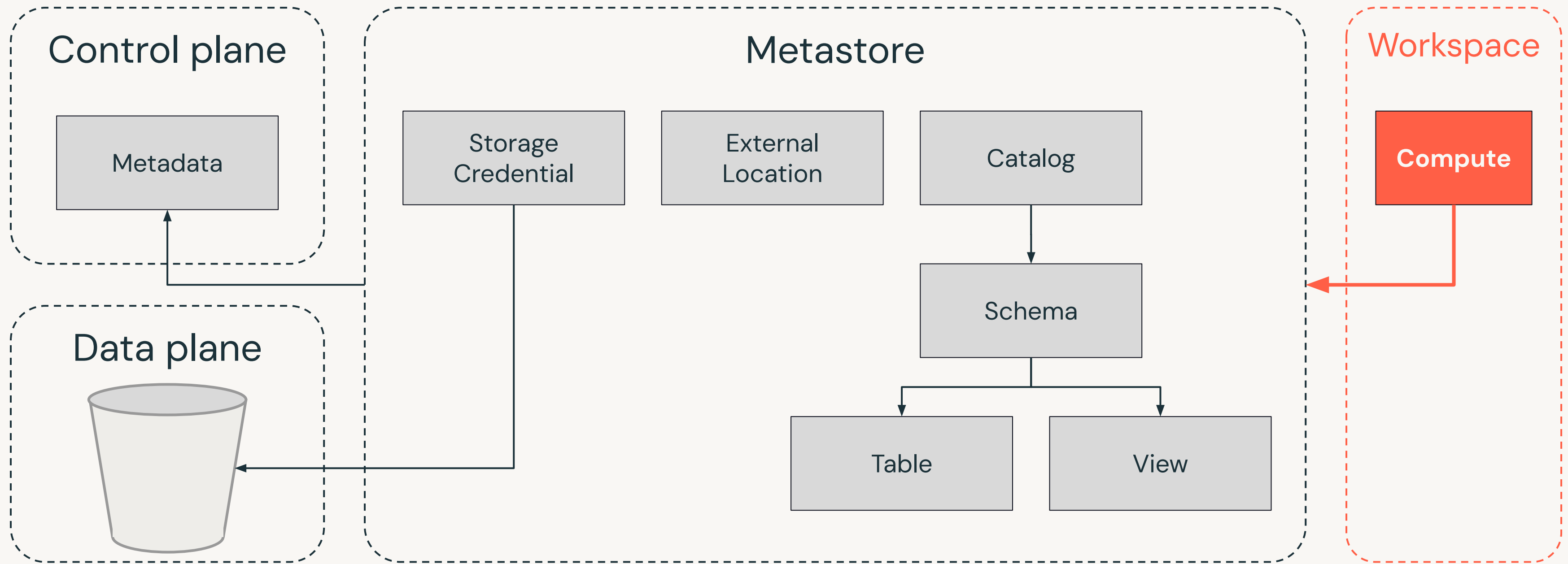
Data Segregation

Metastore requirements



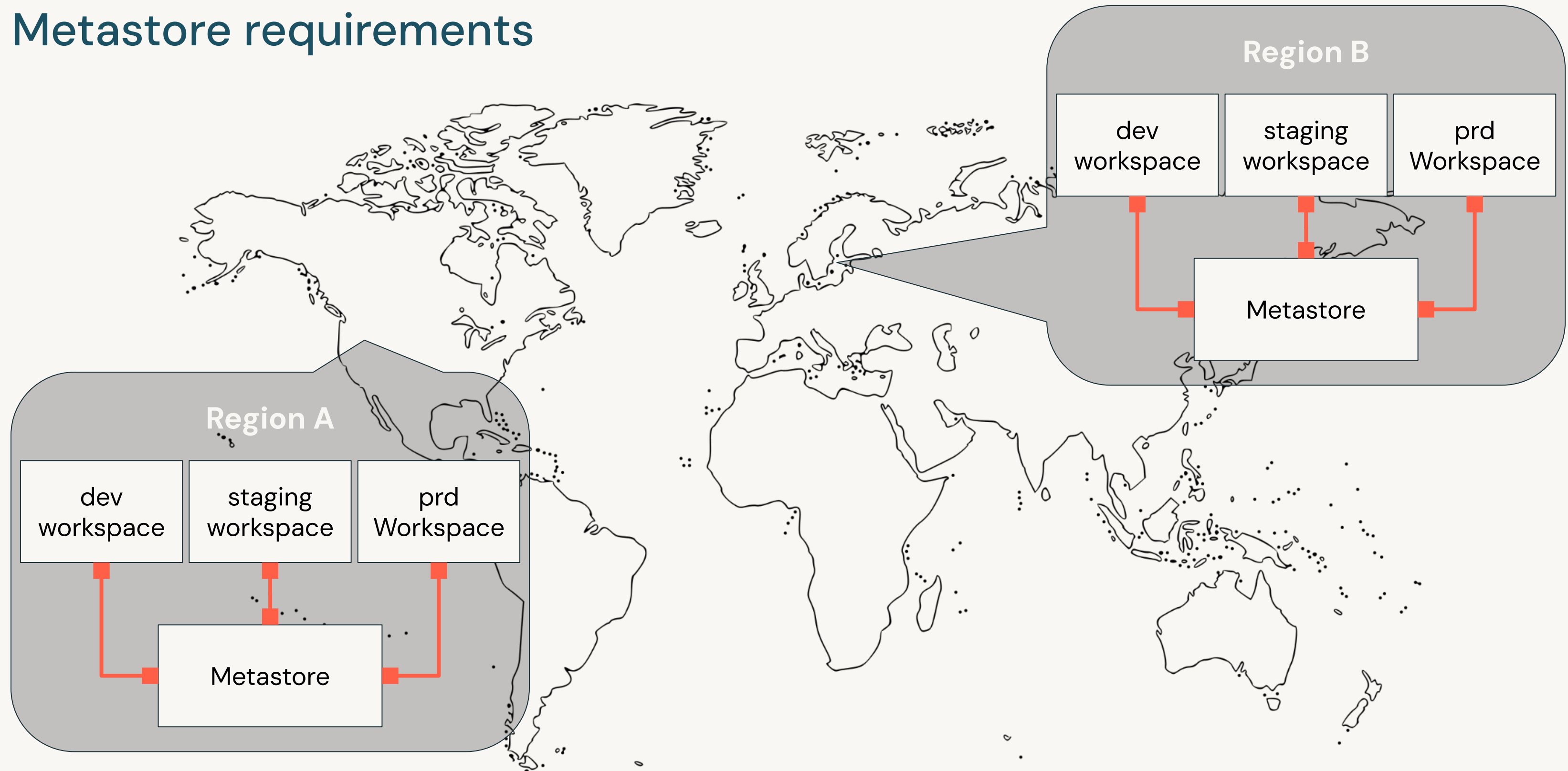
Data Segregation

Metastore requirements



Data Segregation

Metastore requirements



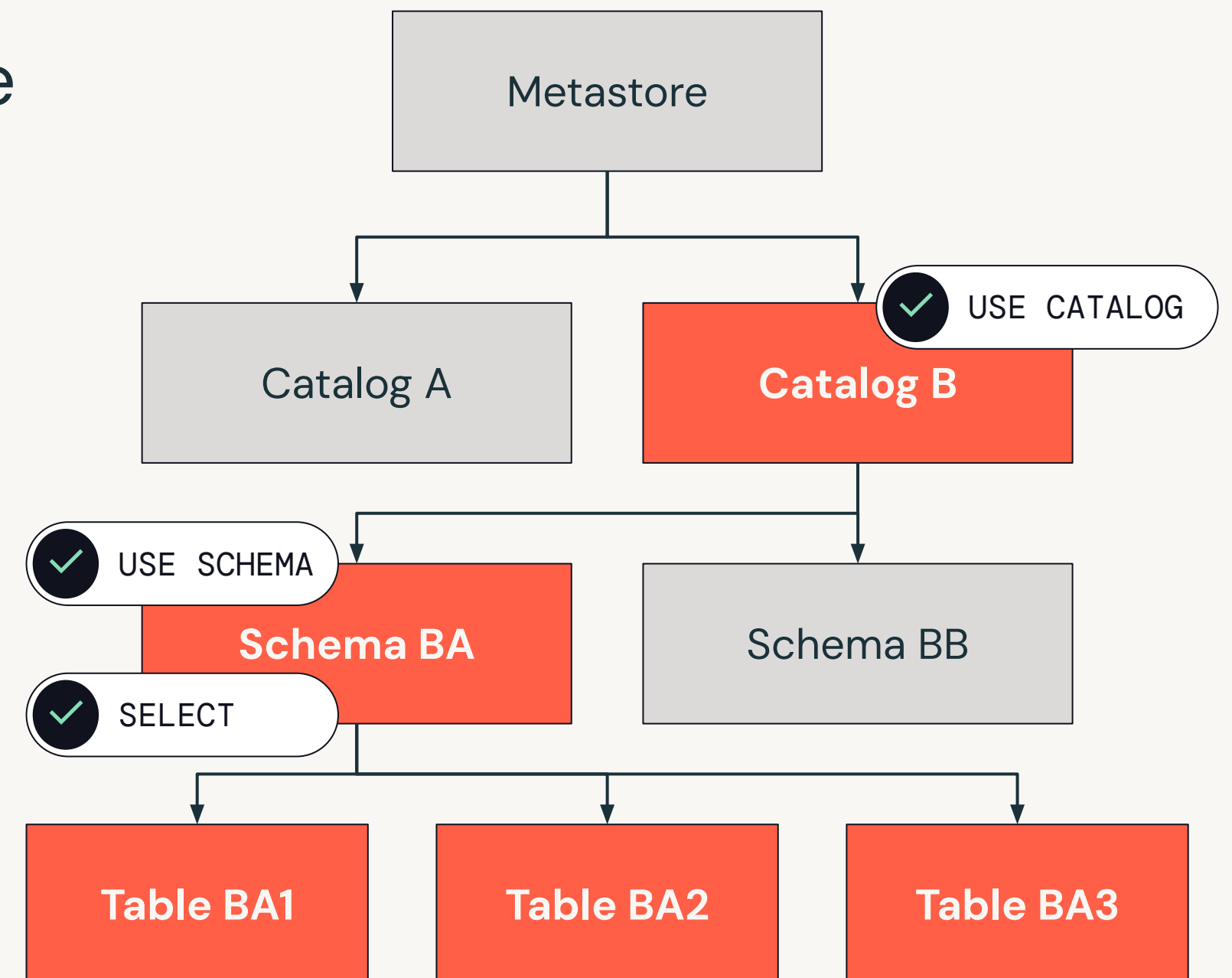
Data Segregation

Logical separation

Use catalogs and/or schemas to segregate data

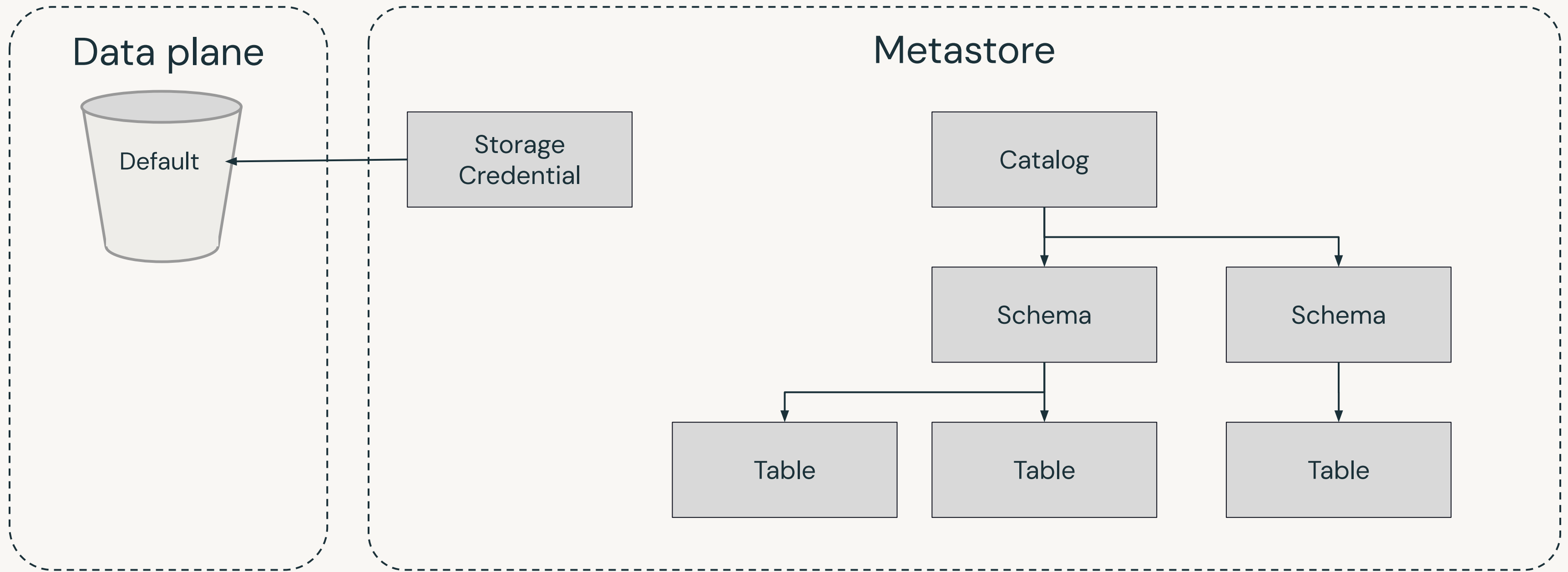
Apply permissions appropriately

- **USE CATALOG** on catalog B
- **USE SCHEMA** on schema BA
- **SELECT** on schema BA



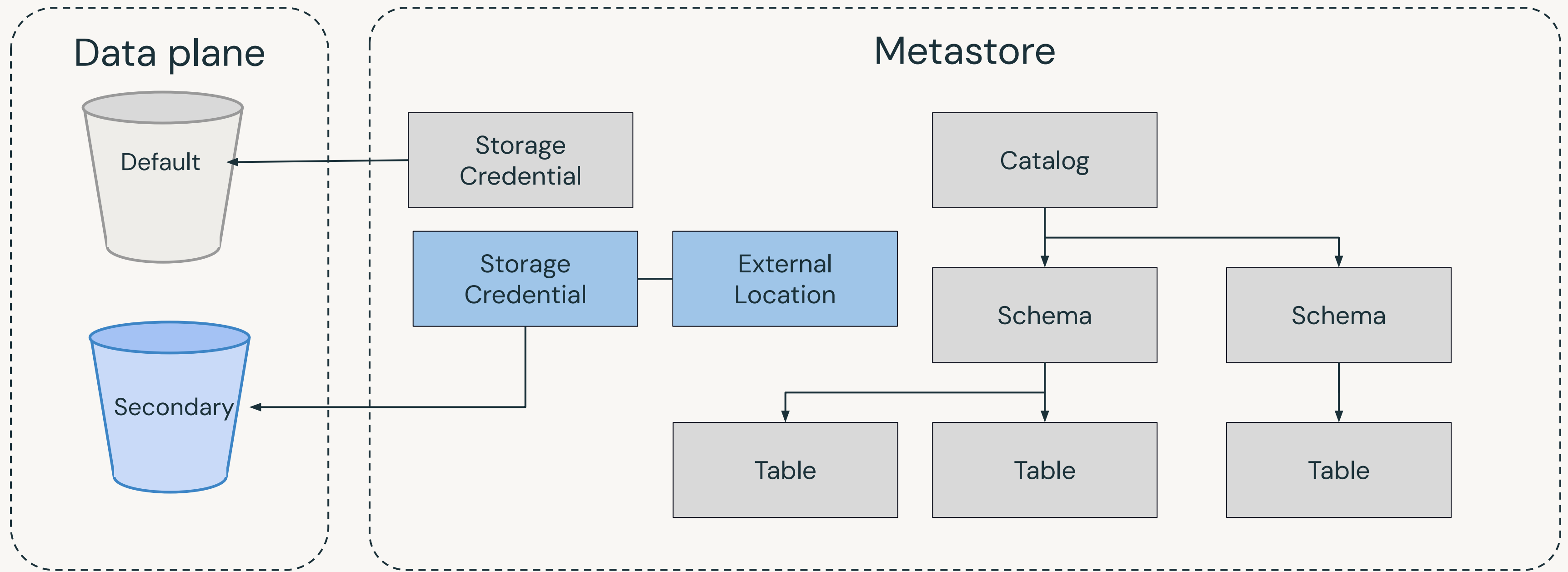
Data Segregation

Physical separation



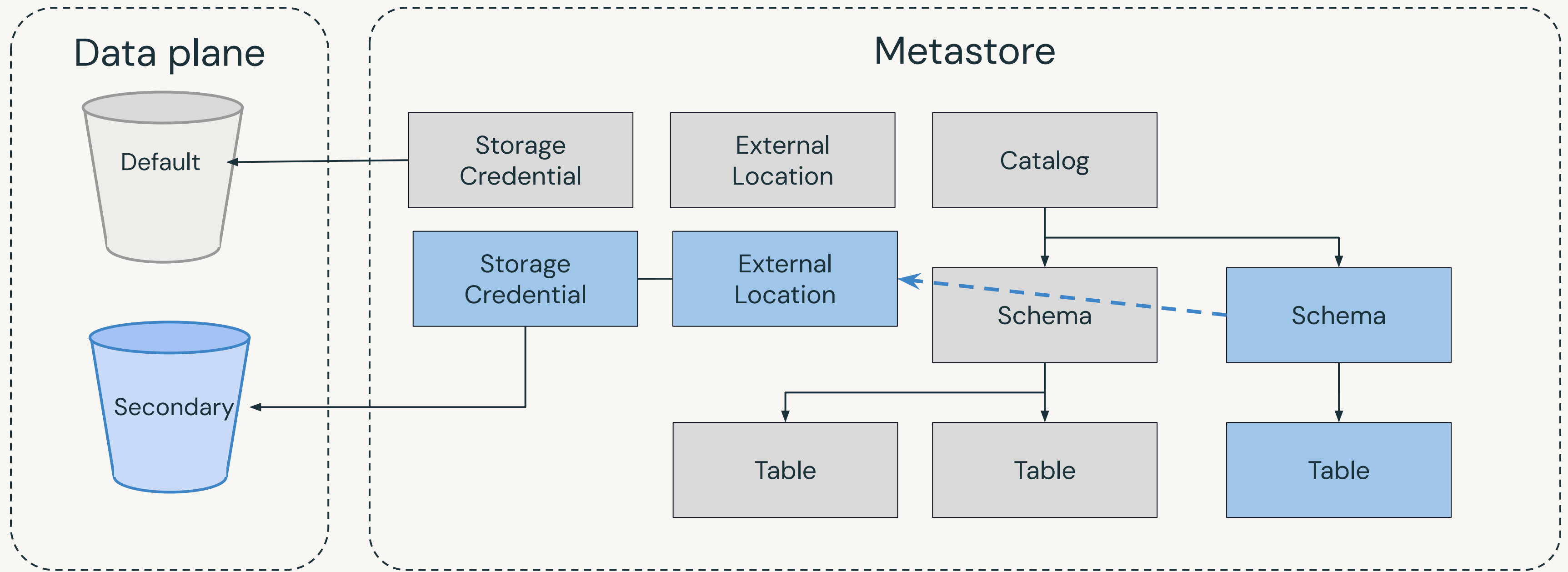
Data Segregation

Physical separation



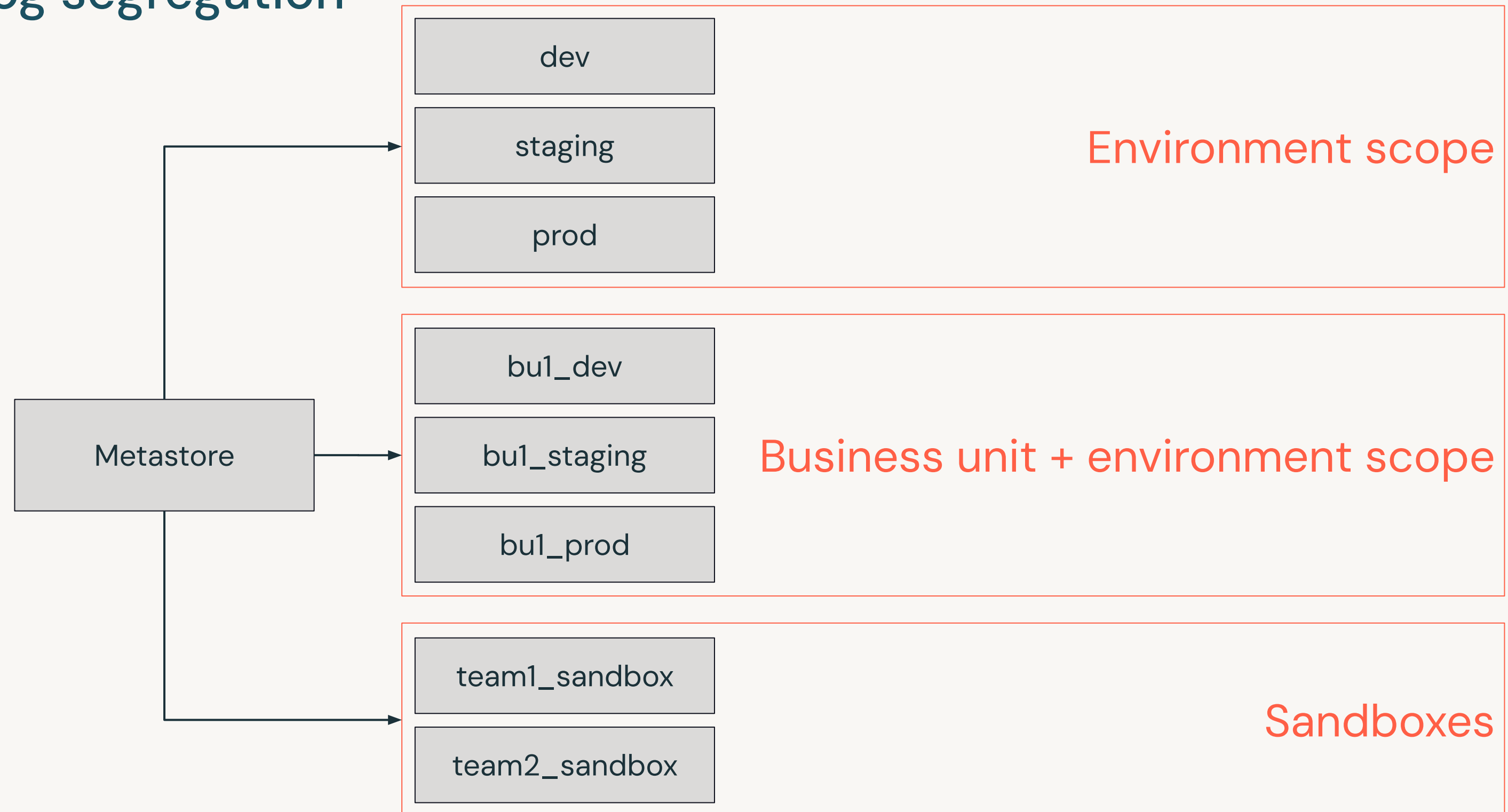
Data Segregation

Physical separation



Data Segregation

Catalog segregation



Demo: Segregating Data



Learning Objectives

In this lab, you will learn how to:

- Redirect schemas and catalogs to external storage



Demo: Upgrading Tables to Unity Catalog



