

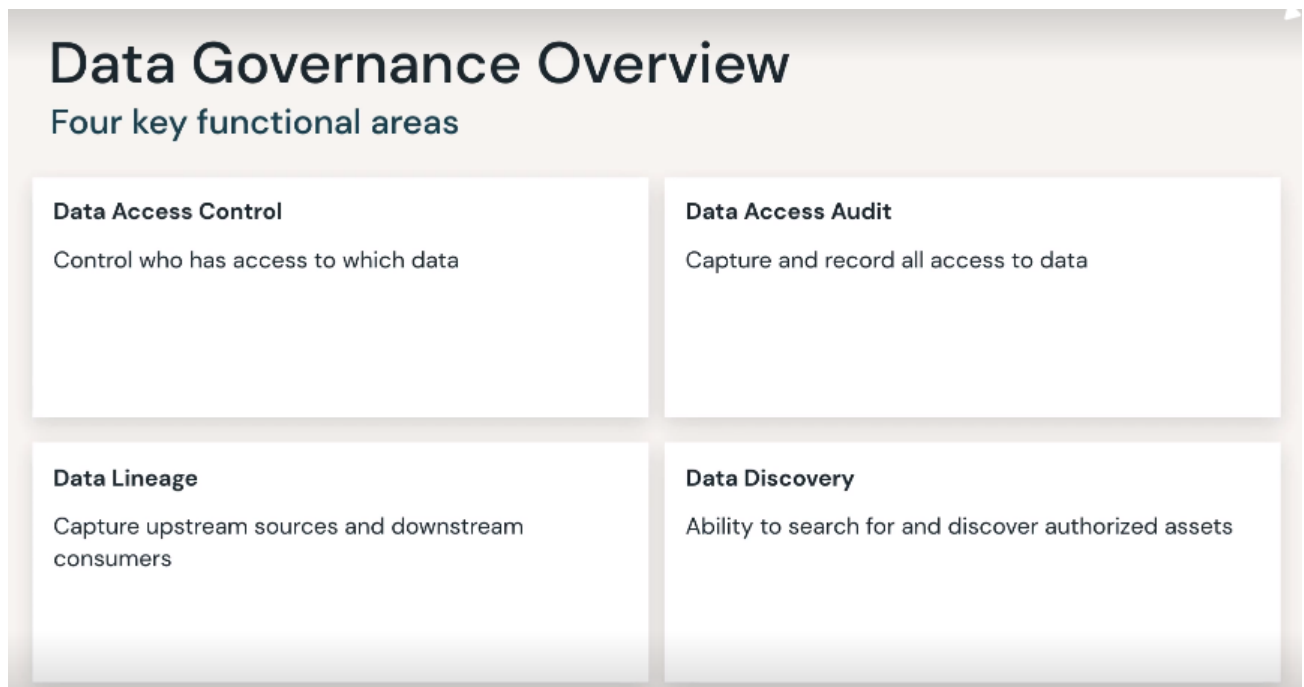
# Data Governance

© Rangel

## Data Access with Unity Catalog

1. Identify one of the four areas of data governance. 1

- Data Access Control
- Data Access Audit
- Data Lineage
- Data Discovery



2. Compare and contrast metastores and catalogs. 1

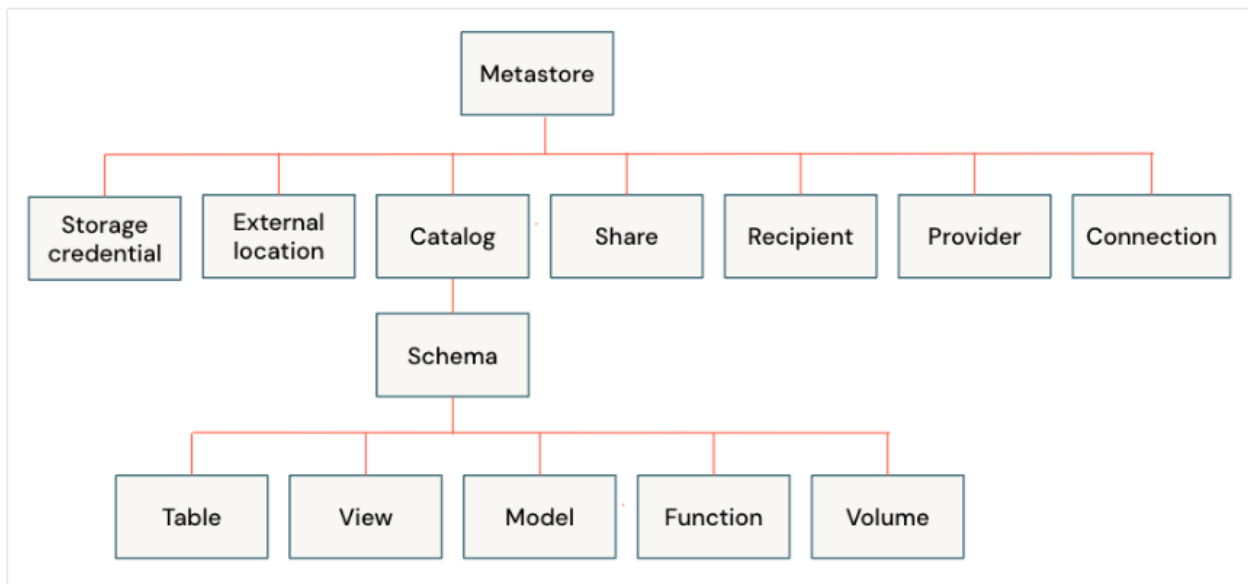
- Metastores are the top level container which has catalogs. By default, we have hive metastore.
- Metastores include **storage credentials**, **external locations**, **shares**, and **recipients**.

3. Identify Unity Catalog securables. 1

- Securables are the things we can grant or revoke privileges onto.

# Securable objects in Unity Catalog

A securable object is an object defined in the Unity Catalog metastore on which privileges can be granted to a principal. Securable objects in Unity Catalog are hierarchical.



4. Define a service principal.

- these are like accounts that can be used for connecting to databricks via API on an external IDE/platform. Similar to service accounts which will have credentials usable by several users.

5. Identify the cluster security modes compatible with Unity Catalog.

- Single user
- User isolation

## Clusters

### Security modes

#### Modes supporting Unity Catalog

##### Single user

Multiple language support, not shareable

##### User isolation

Shareable, Python and SQL, legacy table ACLs

#### Modes not supporting Unity Catalog

##### None

No security

##### Table ACL only

Legacy table ACLs, multiple languages, shareable

##### Passthrough only

Credential passthrough, multiple languages, shareable

6. Create a UC-enabled all-purpose cluster.

- use access mode/security mode to specify the cluster type. Only **single user** and **user isolation** modes are supported by UC.

☐ Multi node ☒ Single node

**Access mode** ⓘ **Single user access** ⓘ

Single user | ▼ jericodev00@gmail.com ▼

Single user  
All languages

No isolation shared  
All languages

7. Create a DBSQL warehouse.

**New SQL Warehouse** ⓘ



Name

Cluster size ⓘ  4 DBU ▼

Auto stop ☒ After  minutes of inactivity.

Scaling ⓘ Min.  Max.  clusters (4 DBU)

Advanced options ▼

Tags ⓘ

Serverless ⓘ ☒  
**Preview**

Unity Catalog ☒

Channel ⓘ ☒ Current ☐ Preview

Cancel

8. Identify how to query a three-layer namespace.

```
SELECT * FROM <catalog>.<schema>.<table>
```

## 9. Implement data object access control

- Go to the **Catalog** tab, and under the catalogs/metastores, you can grant/revoke privileges to the data objects
- For tables/views, use `is_accont_group_member()` function to redact columns or rows using dynamic views

```
# Redact Columns
CREATE OR REPLACE VIEW agg_heartrate AS
SELECT
    CASE WHEN
        is_account_group_member('analysts') THEN 'REDACTED'
    ELSE mrn
    END AS mrn,
    CASE WHEN
        is_account_group_member('analysts') THEN 'REDACTED'
    ELSE name
    END AS name,
    MEAN(heartrate) avg_heartrate,
    DATE_TRUNC("DD", time) date
FROM heartrate_device
GROUP BY mrn, name, DATE_TRUNC("DD", time)

# Redact rows
CREATE OR REPLACE VIEW agg_heartrate AS
SELECT
    mrn,
    time,
    device_id,
    heartrate
FROM heartrate_device
WHERE
    CASE WHEN
        is_account_group_member('developers') THEN device_id < 30
    ELSE TRUE
    END
END
```

## 10. Identify colocating metastores with a workspace as best practice.

- Add the metastore to the same region as the workspace, to enable low latency. Not sure with what the key point is actually referring to lol.

## 11. Identify using service principals for connections as best practice.

- "Databricks recommends using a service principal and its OAuth token or personal access token instead of your Azure Databricks user account and personal access token. Benefits include: \*\*Granting and restricting access

to resources independently of a user\*\*. Enabling users to better protect their own access tokens." [1](https://learn.microsoft.com/en-us/azure/databricks/dev-tools/service-principals)

## 12. Identify the segregation of business units across catalog as best practice

- You can do all governance for several workspaces using Unity Catalog.

