Data Engineering Batch 1 **Day 21:19/02/2024**

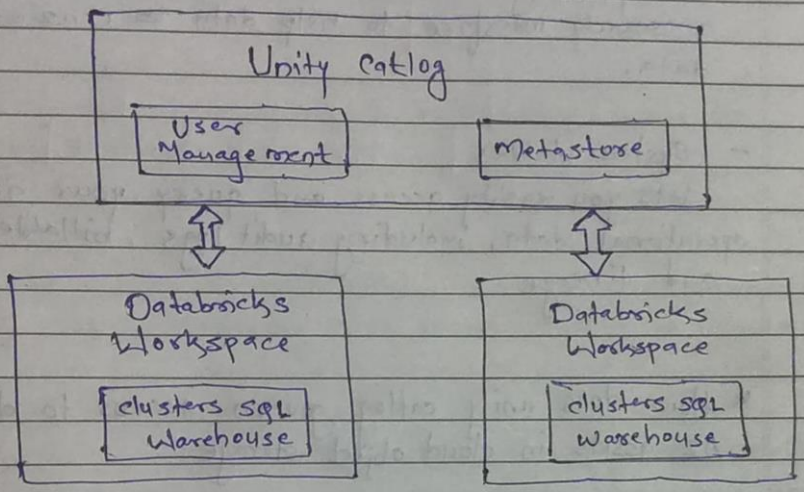
Name: Pradip Bochare

Notes

Page No.
Date

* Unity Catlog.

unity catlog provides centralized access control, building, auditing, lineage and data discovery capabilities across Azure Databricks coorkspaces.



* key features.

- Define once, secure everywhere
offers single place to administer data access policies that apply across all coorkspaces.

- Standards-compliant security model
security model is based on standard ANSI SQL and allows administrators to grant permissions.

- Built-in auditing and lineage
  automatically captures user-level audit logs that records access to your data.

- Data discovery
  lets you tag and document data assets and provides a search interface to help data consumers find data.

- System tables
  lets you easily access and query your accounts operational data, including audit logs, billable usage and lineage.

* How does unity catlog govern access to data and AI assets in cloud object storage

- storage credentials
- External locations
- Managed storage locations
- Volumes
- Tables

Page No.
Date

* Unity catalog object model

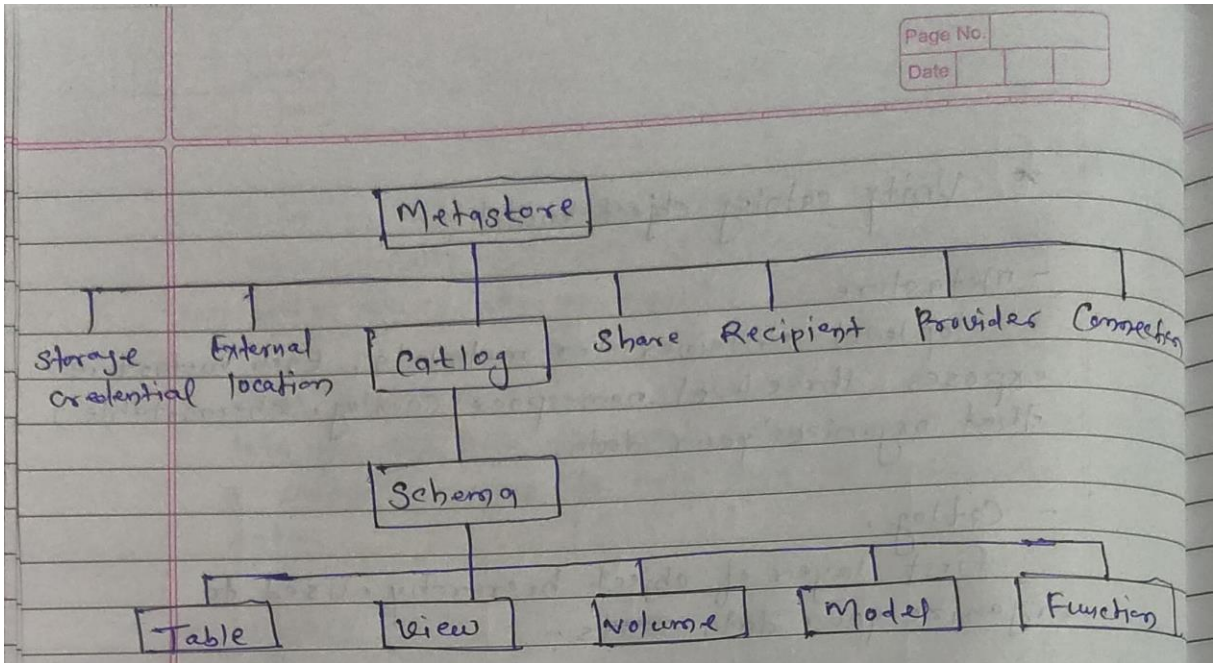
- Metastore
  top level container for metadata. Each metastore exposes three level namespace (catlog.schema.table) that organizes your data.

- Catlog.
  first layer of object hierarchy. used to organize your data assets.

- Schema.
  Also known as databases, schemas are second layer of object hierarchy and contain tables & views.

- Tables, views & volume
  lowest level in data object hierarchy are tables, views, & volumes. Volumes provide governance for non-tabular data.

- Models.
  Registered models can also be managed in unity catlog and reside at lowest level in object hierarchy.


Metastore
Storage credential
External location
Catlog
Share
Recipient
Providers
Schema
Table
View
Volume
Model
Function

❖ Metastores

- A metastore is the top-level container of objects in Unity Catalog. It registers metadata about data and AI assets and the permissions that govern access to them. Azure Databricks account admins should create one metastore for each region in which they operate and assign them to Azure Databricks workspaces in the same region. For a workspace to use Unity Catalog, it must have a Unity Catalog metastore attached.
- A metastore can optionally be configured with a managed storage location in an Azure Data Lake Storage Gen2 container in your Azure account.

❖ Catalogs

- A catalog is the first layer of Unity Catalog's three-level namespace. It's used to organize your data assets. Users can see all catalogs on which they have been assigned the USE CATALOG data permission.
- Depending on how your workspace was created and enabled for Unity Catalog, your users may have default permissions on automatically provisioned catalogs, including either the main catalog or the workspace catalog (<workspace-name>).

❖ Schemas

- A schema (also called a database) is the second layer of Unity Catalog's three-level namespace. A schema organizes tables and views. Users can see all schemas on which they have been assigned the USE SCHEMA permission, along with the USE CATALOG permission on the schema's parent catalog.

- To access or list a table or view in a schema, users must also have SELECT permission on the table or view. If your workspace was enabled for Unity Catalog manually, it includes a default schema named default in the main catalog that is accessible to all users in your workspace.

❖ Tables

- A table resides in the third layer of Unity Catalog's three-level namespace. It contains rows of data. To create a table, users must have CREATE and USE SCHEMA permissions on the schema, and they must have the USE CATALOG permission on its parent catalog.

- To query a table, users must have the SELECT permission on the table, the USE SCHEMA permission on its parent schema, and the USE CATALOG permission on its parent catalog.

❖ Views

- A view is a read-only object created from one or more tables and views in a metastore. It resides in the third layer of Unity Catalog's three-level namespace. A view can be created from tables and other views in multiple schemas and catalogs. You can create dynamic views to enable row- and column-level permissions.

❖ Volumes

- A volume resides in the third layer of Unity Catalog's three-level namespace. Volumes are siblings to tables, views, and other objects organized under a schema in Unity Catalog.

- Volumes contain directories and files for data stored in any format. Volumes provide non-tabular access to data, meaning that files in volumes cannot be registered as tables.