Name : Parth Nandedkar
Date : 19 Feb 2024
Topics : Unity Catalog
Batch : Data Engineering Batch-1

Unity Catalog :

Unity Catalog provides centralized access control, auditing, lineage, and data discovery capabilities across Azure Databricks workspaces.

Key features of Unity Catalog include:

Define once, secure everywhere: Unity Catalog offers a single place to administer data access policies that apply across all workspaces.

Standards-compliant security model: Unity Catalog's security model is based on standard ANSI SQL and allows administrators to grant permissions in their existing data lake using familiar syntax, at the level of catalogs, databases (also called schemas), tables, and views.

Built-in auditing and lineage: Unity Catalog automatically captures user-level audit logs that record access to your data. Unity Catalog also captures lineage data that tracks how data assets are created and used across all languages.
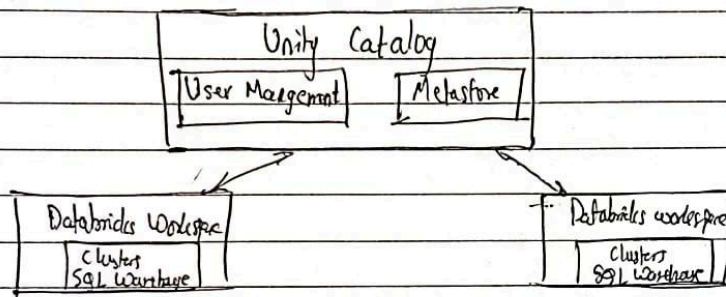
Data discovery: Unity Catalog lets you tag and document data assets, and provides a search interface to help data consumers find data.

System tables (Public Preview): Unity Catalog lets you easily access and query your account's operational data, including audit logs, billable usage, and lineage.

Unity Catalog :

A a unified governance solution for data & AI assets
on the Databricks lakehouse.

```
┌─────────────────────────────────┐
│         Unity  Catalog          │
│  ┌──────────────┐ ┌───────────┐ │
│  │User Management│ │ Metastore │ │
│  └──────────────┘ └───────────┘ │
└─────────────────────────────────┘
      ↙                    ↘
┌──────────────────┐  ┌──────────────────┐
│Databricks Workspace│ │Databricks workspace│
│  ┌────────────┐   │  │  ┌────────────┐  │
│  │ Clusters   │   │  │  │ Clusters   │  │
│  │ SQL Warehouse│ │  │  │ SQL Warehouse│ │
│  └────────────┘   │  │  └────────────┘  │
└──────────────────┘  └──────────────────┘
```

Key features :

- Define once, secure everywhere
- Standards-Compliant security model
- Built-in auditing & lineage
- Data discovery
- System tables (Public Review)

```
┌───────────┐
│ Metastore │
└───────────┘
      │
┌───────────┐
│  Catalog  │
└───────────┘
      │
┌───────────┐
│  Schema   │
└───────────┘
```

Unity catalog is :-

```
┌───────┐┌──────┐┌────────┐┌───────┐┌──────────┐
│ table ││ View ││ Volume ││ Model ││ Function │
└───────┘└──────┘└────────┘└───────┘└──────────┘
```
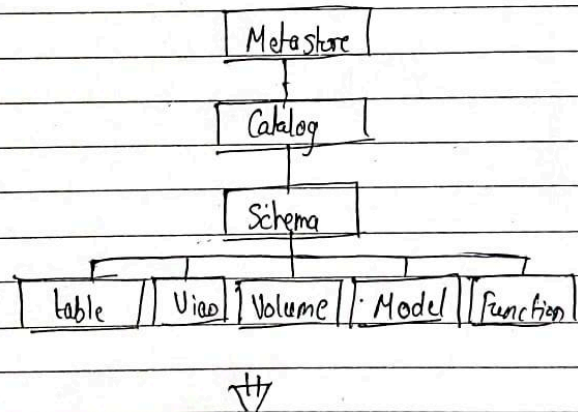
Object Model :

- Metastore
- Catalog
- Schema
- Tables, views, volumes
- Models.

```
┌────────────────────────────────┐
│ ┌──────────────────────────┐   │ Model  contains
│ │                Catalog   │   │
│ │  ┌────────────────────┐  │   │
│ │  │          Schema    │  │   │ Hierarchical
│ │  │ ┌───────────────┐  │  │   │ namespace
│ │  │ │ Views tables, │  │  │   │
│ │  │ │ columns       │  │  │   │
│ │  │ └───────────────┘  │  │   │
│ │  └────────────────────┘  │   │
│ └──────────────────────────┘   │
└────────────────────────────────┘
```

Architecture Diagram:

Unity Catalog

Unity Catalog
- User Management
- Metastore

Databricks Workspace
- Clusters
- SQL Warehouse

Databricks Workspace
- Clusters
- SQL Warehouse

Unity catalog

Key features:
- Define once, secure everywhere
- Standards-compliant security model
- Built-in auditing and lineage
- Data discovery
- System tables (Public Preview):

Unity Catalog provides centralized access control, auditing, lineage, and data discovery capabilities across Azure Databricks workspaces

Azure cloud platform

Unity catalog

Azure databrick

workspace

- clusters
- sql
- editors/features

models

container

Hierarchial namespace

catalogue
schema
views&volumes
Tables

catalog.schema.table

cloud objectstorage -Storage accounts

Unity Catalog

Oject model unity catalog:
- metastore
- catalog
- schema
- table,view,volumes
- Model

Unity Catalog govern access to data in cloud object storage

Metastore

Storage credential | External location | Catalog | Share | Recipient | Provider | Connection

Schema

USE SCHEMA
USE CATALOG

default schema on the main catalog

Table | View | Volume | Model | Function

externally managed data has file layout or fileformats
- DELTA
- CSV
- JSON
- AVRO
- PARQUET
- ORC
- TEXT

models

container

catalogue
schema
views&volumes
Tables

catalog.schema.table

Hierarchial namespace

change schema & catalog ---table of two types(managed&external)

az storage ac
Adls gen2 connect

access privilege for workspace

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.

data lifecycle:creating ,adding ,updating..so on deleting