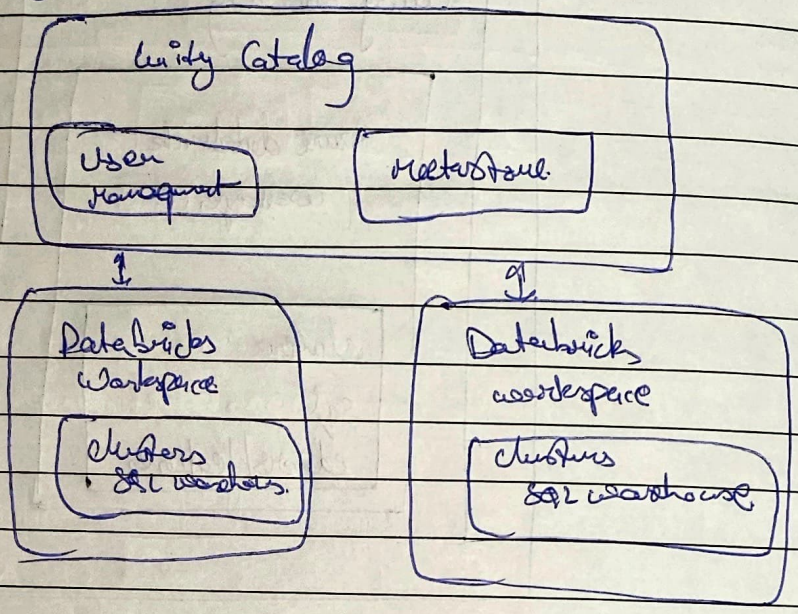
(17)

Unity Catalog

⇒ It provides centralized access control, auditing, lineage, and data discovery capabilities across azure databricks workspaces.

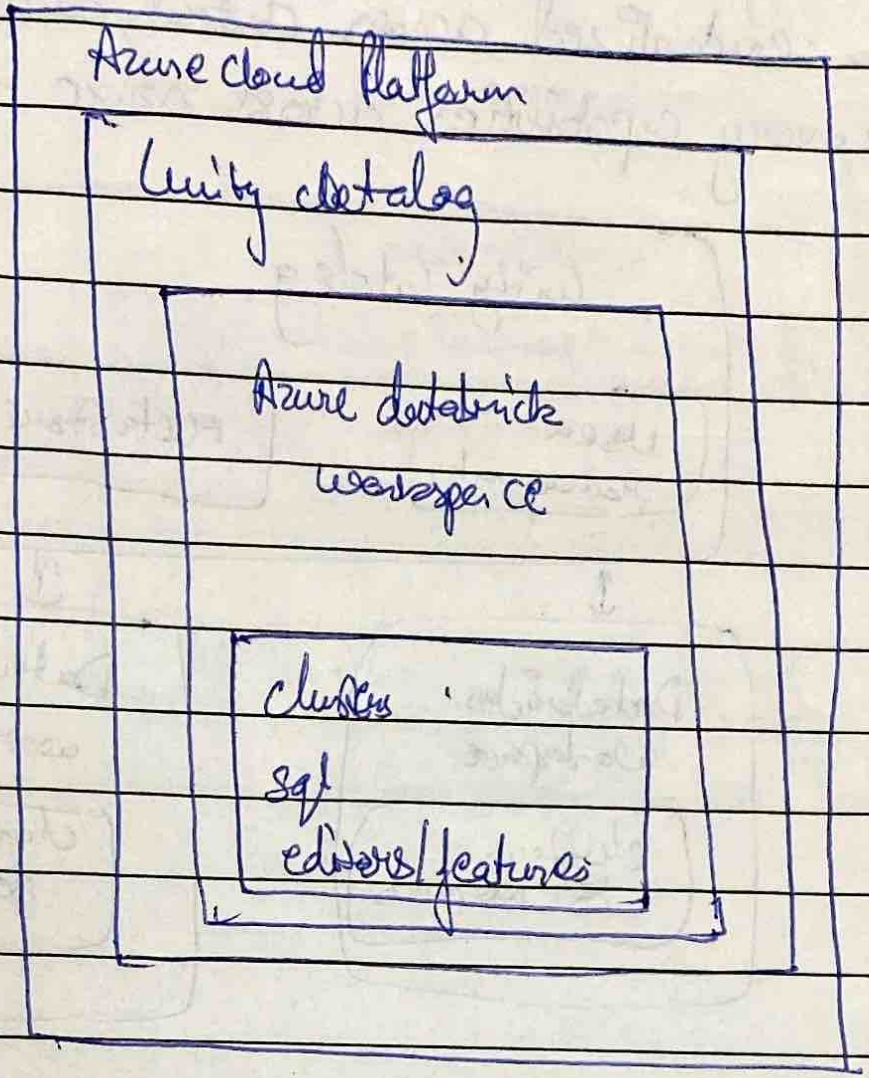


Features of Unity Catalog.

1. Define once, secure everywhere :- Offers a single place to administer data access policies that apply across all workspaces.
2. Standards-Compliant Security Model :- UC's security model is based on standard ANSI SQL.
3. Built-in auditing and lineage :- UC automatically captures user-level audit logs that record access to your data.
4. Data Discovery :- UC lets you tag & document data assets and provides a search interface to help data consumers find data.
5. System Tables (Public Preview) :- UC lets you easily access and query your account's operational data, including audit logs, billable usage and lineage.

Date
Page

Architecture Diagram of Unity Catalog :-



(3)

Unity Catalog Object Model.

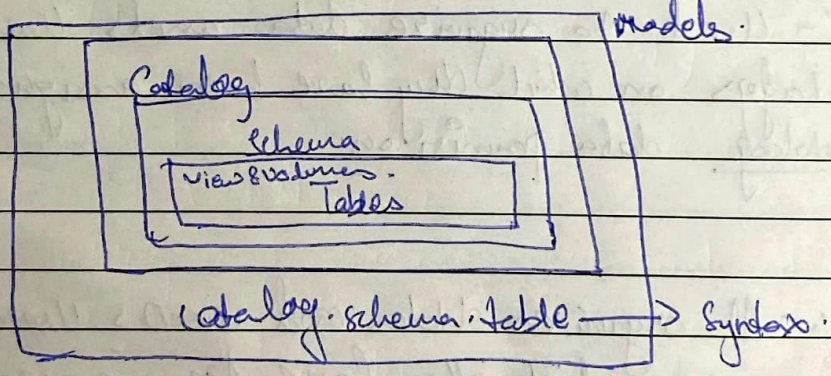
(1) Catalog :- First layer of object hierarchy, used to organize your data assets.

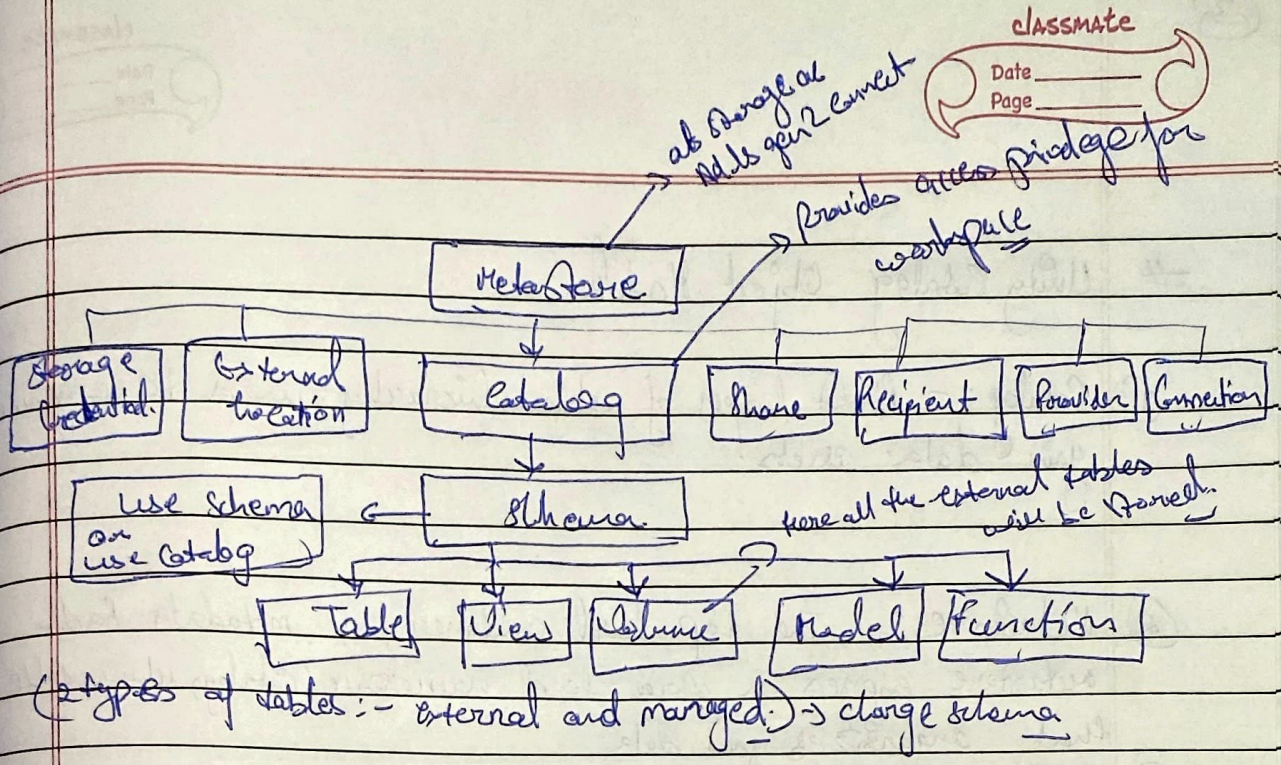
(2) Metastore :- The top-level container for metadata. Each metastore exposes a three-level namespace (catalog, schema, table) that organizes your data.

(3) Tables :- Schema :- Also known as datasets, they are the 2nd layer of the object hierarchy and contain tables and views.

(4) Table Views, Volumes :- Lowest level in the data object hierarchy are tables, views and volumes. Volumes store non-tabular data.

(5) Model :- Data Assets, registered models, can also be managed in Unity Catalog and reside at the lowest level in the object hierarchy.





(2 types of tables :- external and managed) -> change schema

Metastore :- It registers the 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.

Catalog :- Used to organize data assets. Users can see all catalogs on which they have been assigned the USE Catalog data permission.

Schemas :- It organizes tables and views. Users can see all schemas on which they have been assigned. using USE schema.

Tables :- Contains rows of data, to create a tables users must have CREATE and USE schema permissions. A table can be managed or external.

8

Managed Tables.

=> Default way to create tables in Unity Catalog are managed table.

=> By default they are stored in the root storage location that you configure when you create a metastore.

External Tables

=> Tables whose data lifecycle & creating, adding, updating, deleting and file layout are not managed by Unity Catalog.

=> They are used to register large amounts of existing data in Unity Catalog.

File Formats that can be used for External Tables :-

- DELTA
- CSV
- JSON
- AVRO
- PARQUET
- ORC
- TEXT