**UNITY CATALOG MINI PROJECT**

**Theophila Murphy S**

**DE-Batch 6**

**UNITY CATALOG**

**Unity Catalog** is Databricks’ **centralized data governance solution** for all your data, AI, and analytics assets.

It provides a **single, unified interface** to manage:

* **Data access** (permissions, row-level security, masking)
* **Data discovery** (search, tags, metadata)
* **Data lineage** (track where data comes from and how it’s used)
* **Audit logging** (track who accessed what and when)

**Key Features**

1. **Centralized Governance Across Workspaces**
   * One **metastore** for all your Databricks workspaces
   * Consistent security & permissions everywhere
2. **Fine-Grained Access Control**
   * Object-level permissions: catalogs, schemas, tables, views
   * Column masking & row-level security with is\_member() functions
3. **Data Discovery**
   * Searchable catalogs, schemas, and tables
   * Metadata & tagging to classify data (e.g., PII, sensitive, gold/silver/bronze)
4. **Data Lineage**
   * Automatic tracking of:
     + **Upstream** sources
     + **Downstream** dependencies
     + Transformations (via SQL, notebooks, jobs)
5. **Audit & Compliance**
   * Audit logs for every action (select, create, update, grant, revoke)
   * Useful for regulatory requirements like GDPR, HIPAA
6. **Multi-Format Support**
   * Works with Delta, Parquet, and other file formats
   * Can register external data sources

**Hierarchy in Unity Catalog**

Metastore → Catalog → Schema → Table/View/Function

* **Metastore:** Top-level container for all catalogs
* **Catalog:** Logical grouping of schemas
* **Schema:** Contains tables, views, and functions
* **Table/View:** Your actual data assets



















































