# Exercise: Mini Project Using Unity Catalog and Data Governance

## Objective:

Develop a mini project using Unity Catalog to demonstrate key data governance

capabilities include **Data Discovery**, **Data Audit**, **Data Lineage**, and **Access Control**.

# Part 1: Setting Up the Environment

## Task 1: Create a Metastore

Set up a Unity Catalog metastore - **This will created in databricks catalog.**

## Task 2: Create Department-Specific Catalogs

## Create separate catalogs for the following departments:

## Marketing

**CREATE CATALOG marketing\_data;**

Engineering

**CREATE CATALOG engineering\_data;**

Operations

**CREATE CATALOG operations\_data;**

**Task 3: Create Schemas for Each Department**

For Marketing:

**CREATE SCHEMA marketing\_data.ads\_data;**

**CREATE SCHEMA marketing\_data.customer\_data;**

For Engineering:

**CREATE SCHEMA engineering\_data.projects;**

**CREATE SCHEMA engineering\_data.development\_data;**

For operations:

**CREATE SCHEMA operations\_data.logistics\_data;**

**CREATE SCHEMA operations\_data.supply\_chain;**

**Part 2: Loading Data and Creating Tables**

## Task 4: Prepare Dataset

**Task 5: Create Tables from the Datasets**

**Marketing - Ads Data:**

CREATE TABLE marketing\_data.ads\_data.ads (

ad\_id INT,

impressions INT,

clicks INT,

cost\_per\_click DECIMAL(5, 2)

);

INSERT INTO marketing\_data.ads\_data.ads (ad\_id, impressions, clicks, cost\_per\_click)

VALUES

(1, 1000, 50, 0.25),

(2, 1500, 75, 0.30),

(3, 2000, 100, 0.20),

(4, 2500, 150, 0.35),

(5, 3000, 180, 0.40);

**Engineering - Projects:**

CREATE TABLE engineering\_data.projects.project (

project\_id INT,

project\_name STRING,

start\_date DATE,

end\_date DATE

);

INSERT INTO engineering\_data.projects.project (project\_id, project\_name, start\_date, end\_date)

VALUES

(1, 'AI Development', '2023-01-01', '2023-12-31'),

(2, 'Mobile App Redesign', '2023-02-01', '2023-08-31'),

(3, 'Cloud Migration', '2023-03-01', '2023-10-30'),

(4, 'Data Warehouse Setup', '2023-04-01', '2023-11-30'),

(5, 'Automation Initiative', '2023-05-15', '2023-09-30');

**Operations - Logistics:**

CREATE TABLE operations\_data.logistics\_data.shipment (

shipment\_id INT,

origin STRING,

destination STRING,

status STRING

);

INSERT INTO operations\_data.logistics\_data.shipment (shipment\_id, origin, destination, status)

VALUES

(1, 'New York', 'Los Angeles', 'In Transit'),

(2, 'San Francisco', 'Chicago', 'Delivered'),

(3, 'Houston', 'Miami', 'In Transit'),

(4, 'Seattle', 'Denver', 'Delivered'),

(5, 'Boston', 'Dallas', 'Pending');

# Part 3: Data Governance Capabilities

**Task 6**: **Create Roles and Grant Access**

# CREATE ROLE marketing\_role;

# GRANT USAGE ON CATALOG marketing\_data TO marketing\_role;

# GRANT SELECT ON SCHEMA marketing\_data.ads\_data TO marketing\_role;

# CREATE ROLE engineering\_role;

# GRANT USAGE ON CATALOG engineering\_data TO engineering\_role;

# GRANT SELECT ON SCHEMA engineering\_data.projects TO engineering\_role;

# CREATE ROLE operations\_role;

# GRANT USAGE ON CATALOG operations\_data TO operations\_role;

# GRANT SELECT ON SCHEMA operations\_data.logistics\_data TO operations\_role;

# 

# Task 7: Configure Fine-Grained Access Control

GRANT SELECT ON TABLE marketing\_data.ads\_data.ads TO marketting@gmail.com;

GRANT SELECT ON TABLE engineering\_data.projects.project TO engineering\_role;

## Data Lineage

**Task 8**: **Enable and Explore Data Lineage**

SELECT project\_name, COUNT(\*)

FROM engineering\_data.projects.project

GROUP BY project\_name;

## Data Audit

**Task 9**: **Monitor Data Access and Modifications**

* In the Databricks Admin Console, enable audit logs to monitor data access and changes

## Data Discovery

## Task 10

DESCRIBE TABLE marketing\_data.ads\_data.ads;

**Add descriptions to your tables:**

COMMENT ON TABLE marketing\_data.ads\_data.ads IS 'Stores ad data for marketing department';