Mini Project: Retail Sales Data Governance Platform

**Part 1: Setting Up the Environment**

Task 1: Create a Metastore

1. Create a Unity Catalog metastore from the Databricks admin console.

2. Assign the metastore to your workspace.

Task 2: Create Department-Specific Catalogs

CREATE CATALOG marketing;

CREATE CATALOG engineering;

CREATE CATALOG operations;

Task 3: Create Schemas for Each Department

-- Schemas in the Marketing catalog

CREATE SCHEMA marketing.ads\_data;

CREATE SCHEMA marketing.customer\_data;

-- Schemas in the Engineering catalog

CREATE SCHEMA engineering.projects;

CREATE SCHEMA engineering.development\_data;

-- Schemas in the Operations catalog

CREATE SCHEMA operations.logistics\_data;

CREATE SCHEMA operations.supply\_chain;

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

Task 5: Create Tables from the Datasets

-- Create table for ads data in the marketing catalog

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

ad\_id INT,

impressions INT,

clicks INT,

cost\_per\_click DECIMAL(5, 2)

);

-- Insert sample data into marketing.ads\_data.ad\_performance table

INSERT INTO marketing.ads\_data.ad\_performance

VALUES

(1, 1000, 50, 0.30),

(2, 2000, 100, 0.25),

(3, 1500, 75, 0.28);

-- Create table for customer data in the marketing catalog

CREATE TABLE marketing.customer\_data.customer\_info (

customer\_id INT,

customer\_name STRING,

email STRING,

join\_date DATE

);

-- Insert sample data into marketing.customer\_data.customer\_info table

INSERT INTO marketing.customer\_data.customer\_info

VALUES

(1, 'Alice Johnson', 'alice@example.com', '2023-01-15'),

(2, 'Bob Smith', 'bob@example.com', '2023-02-20'),

(3, 'Charlie Brown', 'charlie@example.com', '2023-03-10');

-- Create table for projects in the engineering catalog

CREATE TABLE engineering.projects.project\_info (

project\_id INT,

project\_name STRING,

start\_date DATE,

end\_date DATE

);

-- Insert sample data into engineering.projects.project\_info table

INSERT INTO engineering.projects.project\_info

VALUES

(1, 'Project Alpha', '2023-01-01', '2023-12-31'),

(2, 'Project Beta', '2024-01-01', '2024-06-30'),

(3, 'Project Gamma', '2024-07-01', '2025-01-01');

-- Create table for development data in the engineering catalog

CREATE TABLE engineering.development\_data.dev\_tasks (

task\_id INT,

task\_name STRING,

assigned\_to STRING,

due\_date DATE

);

-- Insert sample data into engineering.development\_data.dev\_tasks table

INSERT INTO engineering.development\_data.dev\_tasks

VALUES

(1, 'Develop UI', 'John Doe', '2024-02-01'),

(2, 'API Integration', 'Jane Roe', '2024-03-15'),

(3, 'Database Optimization', 'Alan Turing', '2024-04-10');

-- Create table for logistics data in the operations catalog

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

shipment\_id INT,

origin STRING,

destination STRING,

status STRING

);

-- Insert sample data into operations.logistics\_data.shipment\_info table

INSERT INTO operations.logistics\_data.shipment\_info

VALUES

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

(2, 'Chicago', 'Houston', 'Delivered'),

(3, 'Miami', 'Atlanta', 'Pending');

-- Create table for supply chain data in the operations catalog

CREATE TABLE operations.supply\_chain.supply\_orders (

order\_id INT,

supplier\_name STRING,

order\_date DATE,

delivery\_date DATE

);

-- Insert sample data into operations.supply\_chain.supply\_orders table

INSERT INTO operations.supply\_chain.supply\_orders

VALUES

(1, 'Supplier A', '2024-01-05', '2024-01-20'),

(2, 'Supplier B', '2024-02-10', '2024-02-25'),

(3, 'Supplier C', '2024-03-15', '2024-03-30');

**Part 3: Data Governance Capabilities**

Task 6: Create Roles and Grant Access

CREATE ROLE marketing\_role;

CREATE ROLE engineering\_role;

CREATE ROLE operations\_role;

Task 7: Configure Fine-Grained Access Control

-- Grant access to marketing role

GRAtNT SELECT ON TABLE marketing.ads\_data.ad\_performance TO ROLE marketing\_role;

GRANT SELECT ON TABLE marketing.customer\_data.customer\_info TO ROLE marketing\_role;

-- Grant access to engineering role

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

GRANT SELECT ON TABLE engineering.development\_data.dev\_tasks TO ROLE engineering\_role;

-- Grant access to operation role

GRANT SELECT ON TABLE operations.logistics\_data.shipment\_info TO ROLE operations\_role;

GRANT SELECT ON TABLE operations.supply\_chain.supply\_orders TO ROLE operations\_role;

Data Lineage

Task 8: Enable and Explore Data Lineage

Enable data lineage for the tables created in Part 2.

Data Audit

Task 9: Monitor Data Access and Modifications

Set up audit logging to track who is accessing or modifying the datasets.

Access the audit logs to view data access patterns and identify who performed

which actions on the data.

Data Discovery

Task 10: Explore Metadata in Unity Catalog

-- Marketing Catalog - Explore Metadata

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

DESCRIBE TABLE marketing.customer\_data.customer\_info;

-- Engineering Catalog - Explore Metadata

DESCRIBE TABLE engineering.projects.project\_info;

DESCRIBE TABLE engineering.development\_data.dev\_tasks;

-- Operations Catalog - Explore Metadata

DESCRIBE TABLE operations.logistics\_data.shipment\_info;

DESCRIBE TABLE operations.supply\_chain.supply\_orders;