Exercise: Mini Project Using Unity Catalog and Data Governance

SaiPrabath Chowdary S

```sql
-- Part 1: Setting Up the Environment

-- Task 1: Create a Metastore

-- Task 2: Create Department-Specific Catalogs

CREATE CATALOG marketing_data;
CREATE CATALOG engineering_data;
CREATE CATALOG operations_data;

-- Task 3: Create Schemas for Each Department

-- Marketing
CREATE SCHEMA marketing_data.ads_data;
CREATE SCHEMA marketing_data.customer_data;

-- Engineering
CREATE SCHEMA engineering_data.projects;
CREATE SCHEMA engineering_data.development_data;

-- Operations
CREATE SCHEMA operations_data.logistics;
CREATE SCHEMA operations_data.supply_chain;


-- Part 2: Loading Data and Creating Tables

-- Task 4: Prepare Datasets and 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 FLOAT
);

INSERT INTO marketing_data.ads_data.ads
VALUES
  (1, 1000, 50, 0.75),
  (2, 1500, 75, 0.60),
```

```sql
    (3, 2000, 90, 0.85);
-- Engineering projects_data
CREATE TABLE engineering_data.projects.projects_data (
  project_id INT,
  project_name STRING,
  start_date DATE,
  end_date DATE
);

INSERT INTO engineering_data.projects.projects_data
VALUES
    (101, 'Project Alpha', '2024-01-01', '2024-06-01'),
    (102, 'Project Beta', '2024-02-01', '2024-07-01'),
    (103, 'Project Gamma', '2024-03-01', '2024-08-01');

-- Operations logistics_data
CREATE TABLE operations_data.logistics.logistics_data(
  shipment_id INT,
  origin STRING,
  destination STRING,
  status STRING
);

INSERT INTO operations_data.logistics.logistics_data
VALUES
    (1001, 'New York', 'Los Angeles', 'In Transit'),
    (1002, 'Chicago', 'Houston', 'Delivered'),
    (1003, 'San Francisco', 'Seattle', 'Pending');



-- Part 3: Data Governance Capabilities
-- Data Access Control

-- Task 6: Create Roles and Grant Access

GRANT USAGE ON CATALOG marketing_data TO `marketing_group`;
GRANT SELECT ON SCHEMA marketing_data.ads_data TO `marketing_group`;
GRANT SELECT ON TABLE marketing_data.ads_data.ads TO `marketing_group`;



-- Task 7: Configure Fine-Grained Access Control

-- users in the marketing department can only access marketing data
GRANT SELECT ON TABLE marketing_data.ads_data.ads TO `user@example.com`;
```

```sql
-- engineers can only access project data.
REVOKE SELECT ON TABLE engineering_data.projects.projects_data FROM
`user@example.com`;


-- operations_role full access to logistics data
GRANT ALL PRIVILEGES ON TABLE operations_data.logistics.logistics_data TO
`user@example.com`;
```

Task 8 and 9 – Databricks

```sql
-- Data Discovery
-- Task 10: Explore Metadata in Unity Catalog

-- View the schema of a table
DESCRIBE TABLE marketing_data.ads_data.ads;

-- Get extended metadata about the table
DESCRIBE EXTENDED engineering_data.projects.projects_data;

-- View the number of rows and other statistics
ANALYZE TABLE operations_data.logistics.logistics_data COMPUTE STATISTICS;



-- description to a catalog
COMMENT ON CATALOG marketing_data IS 'Catalog for marketing-related
data.';

-- Add a comment to the ads_data schema
COMMENT ON SCHEMA marketing_data.ads_data IS 'This schema contains ads
performance data for marketing.';

-- description to a table
COMMENT ON TABLE marketing_data.ads_data.ads IS 'This table stores ad
performance data for marketing campaigns.';
```