

Assignment 1

Task 1: Load Patient Data into Snowflake

- Scenario: You have a CSV file named `patient_records.csv` stored in an AWS S3 bucket. Load this data into the `Patients` table in Snowflake.
- Steps:
 1. Create a file format for CSV:
 2. `CREATE FILE FORMAT my_csv_format TYPE=CSV
FIELD_OPTIONALLY_ENCLOSED_BY='\"';`
 3. Create an external stage for the S3 bucket:
 4. `CREATE OR REPLACE STAGE patient_stage`
 5. `URL='s3://healthcare-data/patient_records'`
 6. `CREDENTIALS=(AWS_KEY_ID='your_key' AWS_SECRET_KEY='your_secret');`
 7. Load the data:
 8. `COPY INTO Patients`
 9. `FROM @patient_stage`
 10. `FILE_FORMAT = (FORMAT_NAME = 'my_csv_format');`

Expected Outcome: The `Patients` table is populated with the data from the CSV file.

Task 2: Create and Query a Materialized View

- Scenario: Generate a report summarizing the average hospital stay duration for each diagnosis using a materialized view.
- Steps:
 1. Create a materialized view:
 2. `CREATE MATERIALIZED VIEW AvgStayByDiagnosis AS`
 3. `SELECT Diagnosis, AVG(DATEDIFF(DAY, AdmissionDate, DischargeDate)) AS AvgStay`
 4. `FROM Patients`
 5. `GROUP BY Diagnosis;`
 6. Query the materialized view:
 7. `SELECT * FROM AvgStayByDiagnosis;`

Expected Outcome: A table displaying each diagnosis and the average hospital stay duration.