IST 4040A FS 2023 Lab

Lab Title: Exploring ETL Process and OLAP Cubes

Introduction:

In this lab, you will explore the Extract, Transform, Load (ETL) process and delve into the creation and querying of OLAP cubes. ETL is a crucial step in data warehousing where data is extracted from various sources, transformed into a suitable format, and loaded into a data warehouse for analysis. OLAP (Online Analytical Processing) cubes provide a multi-dimensional view of data, enabling sophisticated analysis and reporting. In this lab, you will simulate ETL operations and work with OLAP cubes to gain hands-on experience.

Task 1: ETL Process Simulation

Step 1: Extract (E)

Open the provided sales_data.csv file, which contains raw sales data including columns like Date, Product, Quantity, and Revenue.

Study the data to understand its structure and contents.

Step 2: Transform (T)

Implement a Python script named etl_process.py to read sales_data.csv.

Calculate the total revenue for each row (Quantity * Revenue) and add a new column 'Total Revenue' to the dataset.

Save the transformed data into a new CSV file named transformed_sales_data.csv.

Step 3: Load (L)

Set up a database (MySQL, SQLite, etc.) and create a table Transformed_Sales_Data with appropriate columns to match the structure of the transformed data.

Load data from transformed_sales_data.csv into the database table using SQL commands (LOAD DATA INFILE for MySQL or equivalent for other systems).

Task 2: OLAP Cube Creation and Querying

Step 1: Database Setup

Create dimension tables (Date_Dimension, Product_Dimension) and a fact table (Sales_Fact) in your database system. Define appropriate attributes for dimensions and measures.

Step 2: OLAP Cube Definition

Design an OLAP cube schema with dimension keys and aggregated measures.

Create an OLAP cube table in the database (e.g., OLAP_Cube) based on the designed schema.

Use SQL INSERT INTO statements to populate the OLAP cube table with aggregated measures from the Transformed_Sales_Data table.

Step 3: Querying the OLAP Cube

Write SQL queries to perform OLAP operations, such as aggregating total revenue for specific products or time periods.

Execute the queries to retrieve meaningful insights from the OLAP cube.

Submission Instructions on Blackboard:

Submit your etl_process.py Python script for Task 1 along with the transformed_sales_data.csv file.

Provide screenshots or the SQL code used for creating tables and loading data for both Task 1 and Task 2.

Submit the SQL queries used for querying the OLAP cube in Task 2.