Assignment 1: RDBMS

Special Topics in Computer Science: Big Data Systems
CS-UH 3260 Spring 2023
MAX 10 points

In this assignment, you will install MySQL and create a database based on the relational schema that was provided to you during class lab. You will need to populate your database with synthetic content (random values), write SQL queries to answer a series of questions, analyze their performance and propose optimizations.

1. System installation:

Install MySQL (reach out to us if you have trouble with this step). (not graded)

2. Database initialization and loading:

a) Initialization. We are going to use the same schema and an initial seed of data as provided during the lab2_data.sql. Please go ahead and load the data as follows:
mysql -u root -p < lab2_data.sql

Deliverable: The output of the command: show table status from classicmodel;
Grading: **1 point**

b) Load. Further to the previous step, write a python script (or any language of your choice) to generate 1 Million new records to insert in the products table.

Requirements:

- Do not fill the attributes with the same value; instead, use random generators.
- Make sure that quantityInStock is between 0 and 10,000.
- Pay attention to attribute constraints.

Deliverables:

- Script file.
- Data file: assignement_load.sql

Grading:

- After a fresh DBMS initialization, this file should load the data without errors.
  mysql -u root -p < a1_load.sql
- **3 points.**

c) Advanced Load (optional). As a bonus**,** create a script to load random orders (with orderdetails) based on the information available so far i.e., 122 customers and 1,000,110 products. **(1 point 🙌)** *please note that your grade cannot exceed the maximum score.*

3. Database Querying:

Write a SQL query to answer the following requests (1 point each)

1. Retrieve all the records from the "customers" table.
2. Select the order number, order date, and status of all orders with the status “In Process”.
3. Count the number of products in the “products” table.
4. Select the maximum, minimum, and average credit limit of all customer.
5. Select the customer name, order number, and *total cost* of the order for all orders.
6. Retrieve the name of customers who made more than 5 orders.
7. Retrieve the orders of customers with more than 200,000 in credit limit (customer.creditLimit)
8. Return the order numbers of products with at least 8,000 items in stock (products.quantityInStock)

Deliverables:

- SQL file with the queries

Grading:

- Queries run and return the expected results
- **0.5 point each**

4. Query Analysis

Explain Analyze query (8) from. How can you improve the execution?

Deliverables:

- Commands for improving performance
- MySQL plans before and after
- Execution time before and after
- Results discussion

Grading:

- Correct analysis and discussion of the results
- **2 point**