Design Laboratory (CS69202)
Spring Semester 2023

Topic: SQL

Date: January 24, 2024

Time: 2.30 PM - 4.00 PM

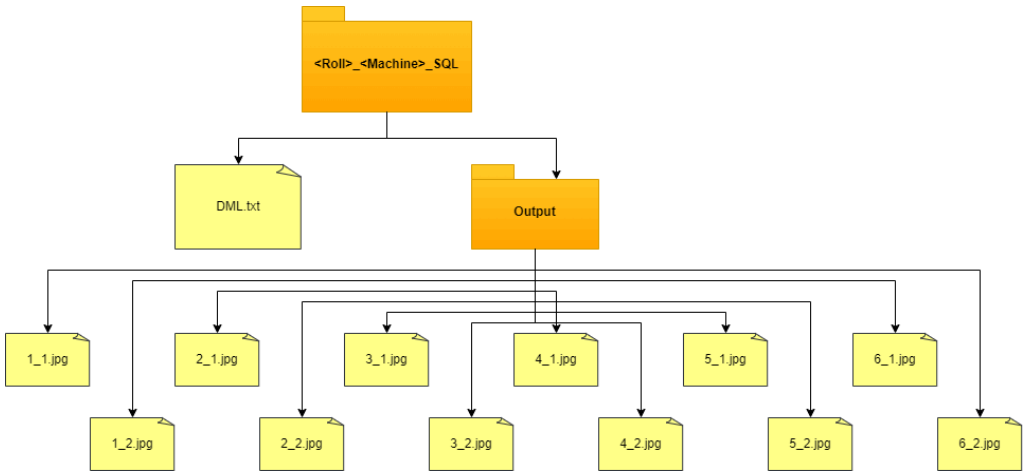
Instructions:

1. There will be no internet access during the exam. Make sure to download any required materials, SQL documentation, or reference guides beforehand.
2. No communication is allowed between students once the exam has begun. Any form of communication will be considered a violation of exam rules.
3. Manage your time wisely. Be aware of the allocated time for the exam to ensure you have sufficient time to complete all questions.
4. We will be using the same database as the previous one. Please make sure the tables and data are in place.
5. Views can be used only for Question Number 3 and Question Number 6. No other solution should contain views in the query. The screenshot of the Solution with views should also showcase the create view command.

Submission Format:

Submit the following files:

- The root folder name should be "<Roll>_<Machine Id>_SQL". For example, if your Roll is 22CS60R01 and you are giving the exam on machine number 20, then the root folder name should be 22CS60R01_20_SQL
- A text file containing SQL queries for all six questions (DML.txt)
- For each query
    - Screenshot of output head that shows query as well (<Ques No>_1.jpg)
    - Screenshot of output tail that shows number of rows (<Ques No>_2.jpg)
- Directory Structure:



Questions:

1. Calculate the total revenue for each supplier ordered by total revenue in descending order.

2. Calculate the total revenue for each month in 2013.

3. Find out the second most popular Product in terms of Quantity ordered. List out the product name and Product ID. Assume the total quantity ordered for TOP 10 items is unique.
   Note: 'LIMIT' cannot be used for the query; on violation, 0 marks will be awarded

4. Identify customers who have consistently increased their order frequency each year (2012, 2013, 2014) to identify our loyal customers. Order frequency is defined as the number of times they have ordered (independent of quantities).

5. List out the top 5 product pairs that are ordered together. List out Product Names and their id along with the frequency with which they are ordered together.

6. Detect potential fraud by identifying orders with unusually very high quantities. List the Product ID, Product Name, Customer Name, and Quantity Ordered. (Assuming orders with quantities that are greater than Fourth Standard Deviations are potential fraud for each product.) We understand that some products may be ordered in high quantities and others in low quantities. So, fraud needs to be calculated for each product's stats.
   Note: STDDEV() function can be used to find Standard Deviation.

   For each product P1, list the order of O1; let P1's mean order quantity is μ_{P_1}, and the standard deviation is σ_{P_1}:
   List out an order if:

   [Quantity Ordered In O1 of P1] > $\mu_{P_1} + 4 \times \sigma_{P_1}$

   For example:

Id	Product	Quantity Ordered	Customer Id
1	Chai Pac	100	-

2	Chai Pac	110	-
3	Ikura	1	-
4	Ikura	1	-
……..			
17	Ikura	1	-
18	Ikura	100	

Here, the order with ID 18 is the only outlier. Whereas orders with ID 1 and 2 are not outliers. This is just an example and not related to the actual answer.
