CSE 111 – DATABASE SYSTEMS

Lab 5: SQL Queries

In this lab session you have to write 15 SQL queries for the TPCH database created and populated in the previous labs. The queries are the following (1 point per query):

1. How many customers and suppliers are in every nation from `ASIA`?
2. For every order priority, count the number of line items ordered in `1998` and received (`l_receiptdate`) earlier than the commit date (`l_commitdate`).
3. How many customers from every region have placed at least one order and have more than the average account balance?
4. Count the number of distinct suppliers that supply parts whose type contains `BRUSHED` and have size equal to any of `10, 20, 30, or 40`.
5. Find the highest value line item(s) (`l_extendedprice*(1-l_discount)`) shipped after `October 6, 1994`. Print the name of the part corresponding to these line item(s).
6. For parts whose type contains `NICKEL`, return the name of the supplier from `AFRICA` that can supply them at maximum cost (`ps_supplycost`), for every part size. Print the supplier name together with the part size and the maximum cost.
7. Print the name of the parts supplied by suppliers from `ARGENTINA` that have total value in the `top 10%` total values across all the supplied parts. The total value is `ps_supplycost*ps_availqty`. Hint: Use the `LIMIT` keyword with a `SELECT` subquery.
8. Based on the available quantity of items, who is the manufacturer `p_mfgr` of the most popular item (the more popular an item is, the less available it is in `ps_availqty`) from `Supplier#000000084`?
9. How many suppliers in every region have more balance in their account than the average account balance of their own region?
10. How many customers are not from `AMERICA` or `EUROPE`?
11. What is the total supply cost (ps_supplycost) for parts less expensive than `$1500` (p_retailprice) shipped in `1996` (`l_shipdate`) by suppliers who did not supply any line item with an extended price less than `1800` in `1998`?
12. Count the number of orders made in `1995` in which at least one line item was received (`l_receiptdate`) by a customer later than its commit date (`l_commitdate`). List the count of such orders for every order priority.
13. For any two regions, find the gross discounted revenue (`l_extendedprice*(1-l_discount)`) derived from line items in which parts are shipped from a supplier in the first region to a customer in the second region in `1996` and `1997`. List the supplier region, the customer region, the year (`l_shipdate`), and the revenue from shipments that took place in that year.
14. The market share for a given nation within a given region is defined as the fraction of the revenue from the line items ordered by customers in the given region that are supplied by suppliers from the given nation. The revenue of a line item is defined as `l_extendedprice*(1-l_discount)`. Determine the market share of `GERMANY` in `ASIA` in `1995` (`l_shipdate`).
15. For the line items ordered in `February 1996` (o_orderdate), find the largest discount that is smaller than the average discount among all the orders.

In order to complete the lab you have to perform the following tasks:

1. Write the `SQL` statement corresponding to every query in the file `test/x.sql`, where `x` is the number of the query above. Every query goes into its separate file. These are the only files you have to modify and submit in this assignment.

2. The format of the expected output for every query is available in `output/x.out`. The included results are only samples. They are not the correct results. So, make sure you match the format, not the exact results.

3. The submission consists of a compressed `zip` file that contains the files in the `test` folder. The name of the file has to be `lab-5.zip`. When you create the file, include the folder `test` into the compression, not every file `test/x.sql` separately.