

CSE 111 – DATABASE SYSTEMS

Lab 5

In this Lab session you will write 15 SQL queries for the TPCB database. You will execute the queries for the data loaded in Lab 2 and check the results.

Write SQL statements for the following queries on the TPCB database (1 pt. per query):

1. How many customers are not from EUROPE or AFRICA?
2. How many suppliers in each region have more than the average account balance of their own region?
3. For the lineitems ordered in May 1995 (o_orderdate), find the largest discount that is less than the average discount among all the orders.
4. How many customers and suppliers are in each country from EUROPE?
5. For parts of type STEEL, return the name of the supplier from AMERICA that can supply them at minimum cost (ps_supplycost) among the American suppliers. Print the supplier name for each part size together with the minimum cost.
6. Based on the available quantity of items, what is the manufacturer of the most popular item (the more popular, the less available items in ps_availqty) from Supplier#000000053?
7. For each order priority, count the number of parts ordered in 1996 and received earlier (l_receiptdate) than the commit date.
8. Count the number of distinct suppliers that supply items with type MEDIUM POLISHED and size equal to any of 3, 23, 26, and 49.
9. Count the number of supplied parts that have total value (ps_supplycost*ps_availqty) in the top 3% value across all the supplied parts and are supplied by suppliers from CANADA.
10. How many customers from each region have never placed an order and have more than the average account balance?
11. Return the highest value (l_extendedprice*(1-l_discount)) items not shipped as of October 2, 1994.
12. What is the total supply cost for parts less expensive than \$1000 (p_retailprice) shipped in 1996 (l_shipdate) by suppliers who did not ship any part with an extended price less than 1000 in 1995?
13. Count the number of orders made in the fourth quarter of 1996 in which at least one line item was received by the customer later than its committed date. List the count of such orders for each order priority sorted in ascending priority order.
14. For any two regions, find the gross discounted revenue (l_extendedprice*(1-l_discount)) derived from line items in which parts were shipped from a supplier in either region to a customer in the other region in 1995 and 1996. List the supplier region, the customer region, the year (l_shipdate), and the revenue from shipments that took place in that year. Order the answers by supplier region, customer region, and year (all ascending).
15. The market share for a given nation within a given region is defined as the fraction of the revenue, sum of l_extendedprice*(1-l_discount), from the line items in that region that were supplied by suppliers from the given nation. Determine the market share of UNITED STATES in EUROPE in 1996 (l_shipdate).

Instructions. You are required to execute the queries on the TPCB database you populated in Lab 2. We provide you the correct answers for this database. We will check the correctness of your queries on a different database. Folder **Lab5** under **Lab** contains an archive **lab5-results.zip** with the correct answers. There is a file for each query, e.g., **1.out** for query 1, **2.out** for query 2, etc. You are required to turn in an archive containing a file with the SQL statement for every query. The names of the files have to be **1.sql** for query 1, **2.sql** for query 2, and so on. There should be exactly 15 files in the archive file. The name of the archive file has to be **firstname-lastname-lab5.zip** where **firstname** and **lastname** are your first and last names, respectively. The type of the archive file has to be **zip**.