CSE 514 Homework 3: Advanced SQL

Objectives: To practice advanced SQL. To get familiar with commercial database management systems (SQL Server) and using a database management system in the cloud (SQL Azure).

Assignment tools: SQL Server on Windows Azure through SQL Azure. SQL Server Management Studio has been installed on the CSE lab and VDI machines if you would like to use that instead of Azure's web interface. Azure setup guide.

What to turn in: hw3-q1.sql, hw3-q2.sql, etc. (see below)

Where to turn in: Gradescope

Assignment Details

SQL Queries (90 points):

For each question below, write a single SQL query to answer that question and save your submission in individual files hw3-q1.sql, hw3-q2.sql, etc. Each .sql file should include:

- The SQL query that once executed returns the expected result.
- A comment that indicated the number of rows your query returns
- A comment that indicates how long the query took, and
- A comment that contains the first 20 rows of the result (if the result has fewer than 20 rows, output all of them).
    - You can simply copy and paste the first rows into the comment.

Now answer the following questions:

1. For each origin city, find the destination city (or cities) with the longest direct flight. By direct flight, we mean a flight with no intermediate stops. Judge the longest flight in time, not distance. (15 points)
   Name the output columns **origin_city**, **dest_city**, and **time** representing the flight time between them. Do not include duplicates of the same origin/destination city pair. Order the result by origin_city and then dest_city (ascending, i.e. alphabetically).
   [Output relation cardinality: 334 rows]

2. Find all origin cities that only serve flights shorter than 3 hours. You should not include canceled flights in your determination. (15 points)
   Name the output column **city** and sort them in ascending order alphabetically. List each

city only once in the result.
[Output relation cardinality: 109]

3. For each origin city, find the percentage of departing flights shorter than 3 hours.You should not include canceled flights in your determination. (15 points)
   Name the output columns **origin_city** and **percentage**. Order by percentage value, then city, ascending. Be careful to handle cities without any flights shorter than 3 hours. You should return 0 as the result for these cities, not NULL (which is shown as a blank cell in Azure) Report percentages as percentages not decimals (e.g., report 75.2534 rather than 0.752534). Do not round the percentages.
   [Output relation cardinality: 327]

4. List all cities that can be reached from Seattle through one stop (i.e., with any two flights that go through an intermediate city) but **cannot** be reached through a direct flight. **Do not include Seattle as one of these destinations (even though you could get back with two flights)**. (15 points)
   Name the output column **city**. Order the output ascending by city.
   [Output relation cardinality: 256]

5. List all cities that cannot be reached from Seattle through a direct flight nor with one stop (i.e., with any two flights that go through an intermediate city). Warning: this query might take a while to execute. We will learn about how to speed this up in lecture. (15 points)
   Name the output column **city**. Order the output ascending by city.
   (You can assume all cities to be the collection of all origin_city or all dest_city)
   (Note: Do not worry if this query takes a while to execute. We are mostly concerned with the results)
   [Output relation cardinality: 3 or 4, depending on what you consider to be the set of all cities]

6. List the names of carriers that operate flights from Seattle to San Francisco, CA. Return each carrier's name only once. Use a nested query to answer this question. (7 points)
   Name the output column **carrier**. Order the output ascending by carrier.
   [Output relation cardinality: 4]

7. Express the same query as above, but do so without using a nested query. Again, name the output column **carrier** and order ascending by carrier. (8 points)
   [Output relation cardinality: 4]

D. Using a Cloud Service (10 points)

The DBMS that we use in this assignment is running somewhere in one of Microsoft's data centers. What do you think about the idea of offering a DBMS as a service in a public cloud? What do you consider the benefits or drawbacks to this approach compared to running the database locally?

Save your answer in a file called hw3-d.txt in the submission directory.

Submission Instructions

Please make sure that

- You are submitting the script files directly to Gradescope
- Your file names match the expected file names (hw3-q1.sql, hw3-q2.sql, …, hw3-q7.sql, hw3-d.txt)