CS 1555/2055 – Database Management Systems (Fall 2019)
Dept. of Computer Science, University of Pittsburgh

Assignment #4: SQL DML

Release: Sept. 26, 2019 Due: 8PM, Wednesday, Oct. 2, 2019

Goal

Gain familiarity with SQL DML and practice SQL DML in the PostgreSQL and Oracle servers.

Description

- Assume the relational database schema that records information related to US forests used in HW2 and shown below:
  - FOREST = (Forest_No, Name, Area, Acid_Level, MBR_XMin, MBR_XMax, MBR_YMin, MBR_YMax)
  - STATE = (Name, Abbreviation, Area, Population)
  - COVERAGE = (Forest_No, State, Percentage, Area)
  - ROAD = (Road_No, Name, Length)
  - INTERSECTION = (Forest_No, Road_No)
  - SENSOR = (Sensor_Id, X, Y, Last_Charged, Maintainer)
  - REPORT = (Sensor_Id, Report_Time, Temperature)
  - WORKER = (SSN, Name, Age, Rank)

  Assumptions: A forest can extend to more than one state. MBR stands for Minimum Boundary Rectangle, which is a rectangle that contains the forest with minimum size. The corners with coordinates (XMin, YMin) and (XMax, YMax) define the size of MBR. Area in FOREST table is the total area of the forest. Area in STATE table is the total area of the state. Area in COVERAGE table is the area of a forest in a state. Percentage in COVERAGE table is the ratio of the area of a forest in a state over the total area of the forest. A sensor is assumed to be in a forest if it is in the MBR of that forest. Please see `hw4-db.sql` for more details about the schema.

- Answer the following questions [for a total of 100 points]:

  1. [95 points total] Use `hw4-db.sql` to create the tables and populate the tables with sample data. Then express the following queries in PostgreSQL. Do not use any views.

     (a) List the names of all forests that have acid_level between 0.65 and 0.85 inclusive.

     (b) Retrieve the names of all forests, each of which has at least 50% of its area in PA.

     (c) Find the names of all roads in the forest whose name is "Allegheny National Forest".

(d) List all the sensors along with the name of the workers who maintain them. The sensors without maintainers should also be listed.

(e) List the pairs of states that share at least one forest (i.e., cover parts of the same forests). Each pair should be listed only once, e.g., if the tuple (PA, OH) is already listed, then the tuple (OH, PA) should not be listed.

(f) For each forest, find its average temperature and number of sensors. Display the result in descending order of the average temperatures.

(g) Find the locations of sensors that reported the highest temperature.

(h) Find the states whose forests cover more than 30% of the state's area (assuming that forests do not overlap).

(i) Find the states that have higher area of forest than the forest area in Ohio.

(j) Find the names of all forests such that no sensors in those forests reported anything between Jan. 9, 2019 and Jan. 11, 2019 (Please use **Set Membership** or **Set Comparisons** operators for this question).

2. [5 points total] Repeat the queries from Question 1 in the Oracle server using DataGrip or SQL Plus (Oracle's interactive terminal). You can find the guide for connecting to the Oracle server on the class web server at `DatabaseEnvironment.pdf` and `OracleEnvironment_1555.pdf` for DataGrip and SQL Plus respectively. Use `hw4-db-oracle.sql` to create the tables and populate the tables with sample data.

What to submit

You are required to submit **four** files under your **pitt_user_name** (e.g., pitt01).

- **pitt_user_name-query.sql** (e.g., pitt01-query.sql)
  In this file, please submit the answers to question 1.
  In addition to providing the answers, you are expected to:
  - **include your name and pitt user name at the top of the file** using SQL comments
  - identify the question number before each answer using SQL comments
  - write simple queries at the beginning of this file to list the content of all the tables.

  The entire text file should be composed of **valid SQL statements**.

- **pitt_user_name-output.txt** (e.g., pitt01-output.txt)
  In this file, please submit the output after running the pitt_user_name-query.sql file. You should use the PostgreSQL interactive terminal `psql` to generate the output. You can find the guide to start and use `psql` on class server at `QuickGuidePostgres_1555.pdf`. Briefly speaking, you need to add `psql` to your system's `PATH`, run your query file in `psql` and output the query output to a text file.
  In addition to providing the query output, you are expected to:
  - **include your name and pitt user name at the top of the file**
  - identify the question number before each answer

- **pitt_user_name-query-oracle.sql** (e.g., pitt01-query-oracle.sql)
  In this file, please submit the answers to question 2.
  In addition to providing the answers, you are expected to:
  * **include your name and pitt user name at the top of the file** using SQL comments
  * identify the question number before each answer using SQL comments
  * write simple queries at the beginning of this file to list the content of all the tables.

  The entire text file should be composed of **valid SQL statements**.

- **pitt_user_name-output-oracle.txt** (e.g., pitt01-output-oracle.txt)
  In this file, please submit the output after running the pitt_user_name-query-oracle.sql file. You need to use the Oracle interactive terminal `sqlplus` to generate the output. In `sqlplus`, you could use the command `spool` to begin a recording session and the command `start` to run your query file. More details about how to run and use `sqlplus` can be found at: `db.cs.pitt.edu/courses/cs1555/current.term/handouts/OracleEnvironment_1555.pdf`.

How to submit your assignment

1. Submit your assignment (the 4 files described above) through the Web-base submission interface (i.e., go to the class web page `http://db.cs.pitt.edu/courses/cs1555/current.term` and click the Submit button). **It is your responsibility to make sure the assignment was properly submitted.**
2. Submit your assignment by the due date (8PM, Wednesday, Oct. 2, 2019). There is no late submission.

Academic Honesty

The work in this assignment is to be done *independently*. Discussions with other students on the assignment should be limited to understanding the statement of the problem. Cheating in any way, including giving your work to someone else will result in an F for the course and a report to the appropriate University authority.