CS 122A: Introduction to Data Management – Fall 2020

Homework 5: Structured Query Language (SQL) (100 points)

Due Date: **Fri, Nov 5 (6:00 PM PDT)**

Submission

All HW assignments should contain both your student ID and your name and must be submitted online via the HW5 dropbox on **Gradescope**. See the table below for the HW 5 submission opportunities. Note that after 6 PM on Saturday the 6th no further HW 5 submissions will be accepted. (We will be releasing the solution at that time, as usual.) Please strive to get all your work in on time! If possible, try to save the one dropped assignment for the end of the term when you are most likely to want/need it.

Date / Time	**Grade Implications**
Friday, Nov 5 (6:00 PM)	Full credit will be available
Saturday, Nov 6 (6:00 PM)	10 points will be deducted

Structured Query Language (SQL) [100 pts]

Congratulations! SWOOSH's Senior VP of Engineering has been impressed by your database design skills and the expertise that you've demonstrated based on the relational algebra and calculus. Now it's time to get *real* – it's time to use PostgreSQL and its implementation of the SQL query language to write a number of queries that they envision needing for their planned applications.

Schema, Data, and Tools

SWOOSH is pleased with the E-R schema that resulted from HW #1 and the corresponding relations resulting from HW #2. You can refer to the provided solutions to remind yourself of their schemas, *but note that only a subset of these relations and fields are provided for this specific assignment.* You will also be able to see the relations' schemas in PostgreSQL when you are using it for this assignment. (You may also want to tape the E-R schema to the wall above your desk to use as a guide when working on this and future assignments - it's likely to be very helpful as a reference when writing SWOOSH queries.) A new sample data set will be provided for you to use in testing your queries, so please drop your old schema and load the HW #5 dataset using the provided data dump/script. More information about how to load the schema and its associated sample data can be found below. **You are to use PostgreSQL for all of the queries in this assignment and turn in the queries and results based on the provided template.**

Loading Data Steps

1. First, if you haven't done so already, load the sql schema we provided as part of homework 2's solutions. Download the SQL file: https://grape.ics.uci.edu/wiki/asterix/attachment/wiki/cs122a-2021-fall/swooshDDL.sql and use the following commands in the psql command line: **\i path/to/file/swooshDDL.sql**
2. Now you can access the data set folder from this Google Drive link: https://drive.google.com/drive/folders/1ZBGxKi33jBMH7ebR2JKHNWi0ClbrhxJb?usp=sharing. This folder contains the .csv files corresponding to each table's data, as well as the **data.sql** script, which you can run to load all of the csv files at once into the existing database schema. Once you download the folder, run the following commands in psql:

   \cd path/to/HW5Data (this is to navigate **into** the HW5Data folder)

   \i data.sql (this copies the contents of each csv file into its respective table)

   If you would like to load the data through **pgAdmin**, you can run the queries that are located in the data.sql separately in the pgAdmin query tool. As an example, you can run this query:

   COPY swoosh.Users FROM 'path/to/HW5Data/users.csv' DELIMITER ',' CSV HEADER;

   to load data into the Users table, and replicate the same query for all subsequent tables. (Note that for the pgAdmin version, there is no backslash in front of the COPY command!)

You are now ready to run queries against your data!

Query Writing

In this assignment your job is to write, debug, and execute a series of eight SQL queries against the SWOOSH test relations. You will also need to show the result of each query that you wrote. Note that you will ***not*** get any points for providing the result of a query on this assignment if your SQL query is syntactically incorrect (i.e., if it doesn't actually execute to produce the result you are showing). Since you have a "live" system at your disposal, this should not be an issue – you will be able to run and debug each of your queries on PostgreSQL. A good debugging technique is to explore the data using simpler SQL queries and to develop your final query one piece at a time, making sure you have each part working before combining everything into your final answer query and wondering why it seems not to be giving you a correct result. ***Note:*** *Be sure that the results of your queries do not contain duplicate records, as you will lose points for that.*

For some problems, you may find one or more of the following hints helpful: (i) To limit the number of results returned by a query, you can use the LIMIT clause in SQL. (ii) You can put a subquery in the FROM clause of a query and then treat it (in the outer query) as though it were a stored table; you can also use the WITH statement in SQL to accomplish the same thing in a form that you may find more intuitive and/or more readable. To help you assess the correctness of your answers we have provided you with the number of result rows to be expected for each query. (Note that matching this number doesn't mean that your query is right, but at least it's a step in the right direction. Note also that your

queries must be written to work correctly against ***any*** instance of the SWOOSH database. If your query is incorrect, but happens to (perhaps coincidentally) yield the same answer set on the sample data as a correct query, you will not get full credit for your query.

It's query time - PostgreSQL here you come! **The eight queries to be written can be found in the template file** that is linked from HW #5's entry on the course wiki page. Good luck and have fun with your first foray into SQL-land!