

Assignment 3 – CIS4301, Fall 2018

Alin Dobra

October 27, 2018

- **Due Date:** Friday November 9, 2017
- **Submit using Canvas before Quiz 3**

The goal of this assignment is to write several SQL queries that will answer questions over the database used by the imaginary Sierra Peak Climbing Club (SPCC), an organization whose members climb the mountain peaks of California's Sierra Nevada mountain range.

The database maintained by the SPCC has four tables:

```
PEAK (NAME, ELEV, DIFF, MAP, REGION)
CLIMBER (NAME, SEX)
PARTICIPATED (TRIP_ID, NAME)
CLIMBED (TRIP_ID, PEAK, WHEN_CLIMBED)
```

The database tables have the following semantics:

- **PEAK** gives information about the mountain peaks that the SPCC is interested in. This table lists the name of each peak, its elevation, its difficulty level for climbers (on a scale of 1 to 5), the map that it is located on, and the region of the Sierra Nevada that it is located in.
- **CLIMBER** lists the SPCC membership, and gives their name and gender.
- **PARTICIPATED** gives the set of climbers who participated in each of the various SPCC-sponsored climbing trips. The number of participants in each trip varies.
- **CLIMBED** tells which peaks were climbed on each of the SPCC-sponsored climbing trips, along with the date that each peak was climbed.

Your assignment is to write SQL queries that answer the following questions, sorted in general on order of difficulty from easiest to most difficult. For questions 1-10, you must also provide the extended relational algebra expression tree as part of the answer.

1. Who has climbed 'Pilot Knob (S)'?
2. Which peaks has Mark climbed?

3. Who has climbed a peak rated class 5 in difficulty?
4. Which peaks have been climbed exactly once?
5. Which peaks exceeding 14,000 feet in height have been climbed by John?
6. Which maps have more than a 2000 foot difference between the highest and lowest peaks?
7. Sort the various maps according to the average height of the peaks on the map, and give the average height of the peaks on each map.
8. Which peaks have been climbed by Mark and Mary?
9. How many peaks remain unclimbed in each region?
10. On which trips did the total elevation gained by all participants exceed 500,000 feet?
11. Compute the average number of peaks scaled by the men in the club and by the women in the club.
12. Which people have climbed every single peak that Maria has climbed?
13. Which region has had the highest fraction of its peaks climbed, and what is that fraction?
14. Which pair of climbers have climbed the most peaks together, and how many peaks is that?
15. Who has climbed more than 20 peaks in some 60 day span?

What To Turn In

Before the quiz on November 9, you need to turn in a digital copy of your solution to the assignment via canvas. For each question, you need to do the following:

- For questions 1-10, You need to provide a extended relational algebra expression tree.
- You need to type/paste the query which answers the question into SQL
- You need to execute the query using SQL
- You need to copy the query and the answer given by SQLite from SQL into a word/text processor

The final solution that you turn in will be a copy of the documents that you create, which should show query/answer pairs for each of the fifteen queries assigned (or as many of the 15 that you choose to do!).

Grading

Each question is worth 10 points. However, the assignment will be graded out of 120 points, so if you answer every question correctly, you can earn up to 25% extra credit. Points will be assigned for each query based on the guidelines below:

For questions 1-10:

For Extended Relational Algebra:

- 0 points: not attempted
- 1 points: attempted but has major flaw
- 3 points: answer given is almost correct, but there is a slight or subtle bug in the tree
- 5 points: completely correct

For SQL Query:

- 0 points: query not attempted, query does not give any results, or it does not compile
- 1 points: query compiles, runs, and is most of the way towards a correct answer, but may have a major bug
- 3 points: the query and answer given are almost correct, but there is a slight or subtle bug in the query
- 5 points: the query is correct and gives the right answer

For questions 11-15:

- 0 points: query not attempted, query does not give any results, or it does not compile
- 1 points: query compiles, runs, and is most of the way towards a correct answer, but may have a major bug
- 3 points: the query and answer given are almost correct, but there is a slight or subtle bug in the query
- 5 points: the query is correct and gives the right answer

Getting Started

You will need to download the SQLite database file that has been posted to canvas. This file is the database that already contains the tables and data needed to complete the assignment.

Online (easiest)

The easiest way to get started with this assignment is to use an online SQLite tool. For example, <https://sqliteonline.com/>. In this tool, you can load the database and perform your queries.

To begin your assignment, navigate to the website and load the database using the File Menu. Once loaded, simply type your query into the tool and select 'Run'.

To check that everything is ok, try the following query:

```
SELECT * FROM CLIMBER;
```

If all of the climbers in the database are spit out to the screen, then you are ready to go!

Locally

If you prefer to work locally (offline) you need to do is to get SQLite. To do this, simply follow the link <https://www.sqlite.org/download.html> and download the appropriate binary for your system. (Note: If you use a Linux distribution, you can use your package manager to install SQLite). If using the CISE servers to complete the assignment, SQLite is already installed.

To begin the assignment, you need to fire up SQLite command line shell (sqlite3), which is a command-line interface to the SQLite database software. The SQLite project provides a simple command-line utility named sqlite3 (or sqlite3.exe on Windows) that allows the user to manually enter and execute SQL statements against an SQLite database. To start the sqlite3 program, navigate to the directory you stored the a3.db and simply type "sqlite3 a3.db" at the command prompt. This will start the command-line interface and open the a3.db so that you can query against it. To check that everything is ok, try the following query:

```
SELECT * FROM CLIMBER;
```

If all of the climbers in the database are spit out to the screen, then you are ready to go!

A Word of Caution

Start early! The average student will need many hours to get all 15 queries to work. Even if you only try the first 12, this assignment will still take a lot of time, so start right now! For example, when a previous TAs attempted this assignment, he spent eight hours completing it (though he spent only five hours on the first 14 questions, and around three hours on the first 12). Obviously, this is not something that can be done in a single sitting for most students.

A Final Note: Using Views

Several of these queries will be nearly impossible to write without breaking them up into pieces. The way to do this is with the SQL `CREATE VIEW` command. A view is nothing more than an "imaginary" database table that is the result of an existing database table, and it can make life a lot easier by allowing you to arrive at an answer step-by-step. For example, to create a list of all of the women in the database, you could use:

```
CREATE VIEW WOMEN AS
SELECT *
FROM CLIMBER
WHERE SEX = 'F';
```

Then, you can query `WOMEN` just like any database table. To kill this view, simply type `DROP VIEW WOMEN;` in SQL. Views are discussed in Section 6.7 of your textbook, and will also be discussed in class.