## **Assignment 2 Spec**

### **CMPUT 291 - Introduction to File and Database Management Fall 2018**

# **Assignment Two**

## **(individual assignment)**

### **Due date: *Oct 12 at 5pm* ([submission details](https://eclass.srv.ualberta.ca/mod/page/view.php?id=3105792" \l "submission))**

### **Clarifications:**

*No clarification will be posted here after 5pm on Oct 11th.*

* **Oct 7.** [Here](https://eclass.srv.ualberta.ca/pluginfile.php/4443941/mod_page/content/23/a2-rubric.txt) is a marking rubric for this assignment.

### **Introduction**

The goal of this assignment is to improve your skills of writing declarative queries on a relational database, in general, and also to improve your fluency in SQL (and SQLite).

You have been provided with the following relational schema.

* *members(email, name, phone)*
* *cars(cno, make, model, year, seats, owner)*
* *locations(lcode, city, prov, address)*
* *rides(rno, price, rdate, seats, lugDesc, src, dst, driver, cno)*
* *bookings(bno, email, rno, cost, seats, pickup, dropoff)*
* *enroute(rno, lcode)*
* *requests(rid, email, rdate, pickup, dropoff, amount)*

The tables are derived from the specification of Assignment 1 and the names of the tables and columns should give the semantics, except minor differences which are explicit in table definitions, insert statements or queries.

### **Creating the database**

Using [the SQL statements provided](https://eclass.srv.ualberta.ca/pluginfile.php/4443941/mod_page/content/23/a2-tables.sql), create the above tables in SQLite3 on Lab machines with some data. Here is [a small initial data](https://eclass.srv.ualberta.ca/pluginfile.php/4443941/mod_page/content/23/a2-initial-data.sql) to get you started.

### **(90 marks) Queries**

Write down the following queries in SQL and run them in SQLite3 over the database created. You will be writing ONE SQL statement for every query (here One SQL statement starts with a SELECT and ends with a semicolon but may include multiple select statements combined in the form of subqueries and/or using set operations). Your SQL queries for questions 1-3 cannot use any of aggregation, grouping, or nesting (set operations are ok).

1. Find (name and email of) members who have at least two cars and at least one car is associated with a ride.
2. Find (name and email of) members who have cars and bookings but have not offered any ride.
3. Find (email of) members who are booked on a ride from *Edmonton* to *Calgary* in *November 2018*. *Hint*: Check out the date and time functions in SQLite.
4. Find requests that are served by the rides offered. A request is served by a ride if the dates are the same, the ride price is not greater than the requested amount, and the pickup and the drop off cities and provinces are also the same (the street addresses may not be the same). For each qualifying request, list the rid of the request, the requester's email, the location codes for the requested pickup and drop-off locations, and the rno of the rides that serve it.
5. Find top 3 destination cities with the largest number of rides. For the qualifying cities, list both the province and the city. *Hint*: Check out the *limit* clause for SQLite.
6. For every city, list the city, the province it is in, the number of rides from, the number of rides to, and the number of rides enroute that city. Include every city where at least one of the counts is non-zero. *Hint*: you may find outer join useful.
7. Find (the rno of) the cheapest ride(s) from *Edmonton* to *Calgary* in *October 2018* with available seats (meaning not all seats are booked).
8. Find (email of) members who have offered rides to more than half of the locations in *Alberta* and all their rides take place in *2016* or after.
9. Create a view called *ride\_info* with columns *rno, booked, available, rdate, price, src*, and *dst*. The view includes for each ride in future, respectively the rno of the ride, the number of seats booked, the number of seats available, the date, the price, the source city and the destination city.
10. Using the view created in Q9, find all ride(s) from *Edmonton* to *Calgary* in *December 2018* with available seats (meaning not all seats are booked). For each qualifying ride, list all information in the view, the email of the member offering the ride, and the number of days from the date of the ride to *Jan 1, 2019*. Sort the result on price from the cheapest to more expensive ones.

### **(upto 5 bonus marks for the first 3 people ) Preparing test data**

Written queries should be tested for correctness and bug fixes, very much like programs written in any programming language. For testing, you need to have enough data in your tables such that all your queries are meaningful and non-trivial (e.g. the returned answers are not empty). You are encouraged to share your data with your classmates or use data prepared by them. *To make this collaboration happen, there will be up to 5 bonus marks (at the instructor's discretion) to the first 3 people who prepare a test data and share it with the rest of the class. Make sure your data is correct and meets the expectation of the assignment. If you are sharing your test data, please post it to the course discussion forum.* Put all your *insert* statements in a file called *a2-data.sql*. Make sure to put down your name, email and a date when it is published or revised at the beginning of the file as a comment line (e.g. -- Data prepared by <firstname lastname>, <email address>, and published on <date>). If you are using data prepared by someone else, leave the identification line unchanged.

### **(10 marks) Testing and report**

Starting from scratch, create your database as

*sqlite3 a2.db <a2-tables.sql*

and populate your tables using data file a2-data.sql (prepared in the previous step) as

*sqlite3 a2.db <a2-data.sql*

Put all your SQL queries in a file named a2-queries.sql; Add the following line at the beginning of the file

*.echo on*

and the following line before each SQL query (replacing X with the query number).

--Question X

Run your queries on your data file as

*sqlite3 a2.db <a2-queries.sql >a2-script.txt*

You will be submitting both a2-data.sql and a2-script.txt electronically as described in the instructions for submissions.

### **Instructions for Submissions**

We will make use of some automated tools in testing your queries. Thus it is important that you follow the following instructions closely.

1. Your queries will be tested under a TA account with the provided tables. Do not use any table or column names other than those provided.
2. Write each query in a separate file. Your solution must have **one SQL statement for each query**. In other words, you cannot use views or temporary tables unless you are explicitly asked to do so. The first query must be saved in a file named 1.sql, the second query in a file named 2.sql, and so on until the tenth query, which is to be saved in a file called 10.sql (**the names are important!**).
3. The first line of each query file must have the command:

.print Question *X* - CC*ID*

1. where *X* is the number of the query and CCI*D* is your CCID. For example, the first line of the third query file for the user with ccid 'drafiei' will be:

.print Question 3 - drafiei

1. The rest of each file must contain the SQL query you are submitting and nothing else.
2. Include with your submission a README.txt file that has your name, ccid, lab section, and the list of people you collaborated with (as much as it is allowed within the course policy) or the line "I declare that I did not collaborate with anyone in this assignment". A submission without a README.txt file or with missing information will lose 5% of the total mark.
3. Bundle all your queries, insert statements (a2-data.sql) and scripts (a2-script.txt) into a single tarfile by executing the Unix command (everything should be on one line):

tar -czf a2.tgz README.txt a2-data.sql a2-script.txt 1.sql 2.sql 3.sql 4.sql 5.sql 6.sql 7.sql 8.sql 9.sql 10.sql

1. Submit the file *a2.tgz* at the [submission page](https://eclass.srv.ualberta.ca/mod/assign/view.php?id=3105811) after logging into eclass.

Eclass does not support versioning (unfortunately) and each new submission replaces your previous one. This makes last minute submissions somewhat risky. Avoid last minute submissions as much as you can, and check your submissions after an upload to make sure the right content is uploaded. A common mistake is to use a wrong tar command and submit a corrupt file.