CSCI 585- Database Systems

Fall 2015

Homework Assignment 2

Please do check class policies posted as a presentation on D2L-> contents -> overview.

Deadline: October 5th 11:50 PM

During assignment 1, you came up with an EER for the *MandysList* system. *Inspired* by *MandysList*, in this assignment you will create a database and populate it with the data in the attached excel file. Please note that the database you create would NOT reflect a complete schema for assignment 1, nor would it follow all the rules described in assignment 1. The excel file has tabs in the bottom. Each tab guides you to a table that you should create in your database. You will be creating your database as well as the queries in sqLite which is a lightweight tiny database. We suggest you use sqlitebrowser. It can be installed on Mac, windows and Linux. (20 points).

Queries on the database (80 points)

Write the following queries and run them on your database developed above. Each query is worth 10 points.

Note: When asked to return an entity, e.g. 'user', unless explicitly asked to return a specific field of the entity, e.g. the name of the user, it is fine to return only the Primary Key, e.g. user_id, and returning all the fields is not mandatory.

1. We want to draw a histogram of year-of-birth for users that were born in or after 1970. Write a query that shows years starting 1970 and the number of users born in that year. You may skip a year for which there is no user born in that year (optional).
2. Given that there are 24 hours in a day, what is the hour that has the highest number of ads created.
3. How many ads were posted to category '250' after user 'lhartj' logged out?
4. What city has the largest number of regions?
5. What is the name of the user whose ad has been liked the most?
6. Mandy'sList popularity: What is the region where the largest number of ads have been posted in?
7. Frequent posters: list top three users who have posted the largest number of times during 2015.
8. What is the title and price of the most recent ad created by user 'bnguyen50'

Submission Guidelines

1. Your submission should include one createdb.sql file, and one queries.sql file including all queries. Please make sure to write your name and student ID as a comment line in both files. Queries in queries.sql may be delimited in any desired way as long as they are identifiable. E.g. you can identify the first query as:

// Q1:

SELECT * FROM ….

2. createdb.sql file should create the required tables, generate primary and foreign keys, etc., and populate all data provided.

Notes:

- It is required to identify primary and foreign keys in your schema.
- Using 'Views' is optional towards completing the assignment.
- Each tab in the dataset.xlsx file represents a table. Please feel free to convert the excel file into a csv file and copy/paste the contents to your createdb script.
- Q: Would it be acceptable if I convert my assignment 1 EER diagram into a database? No. Delivering a different schema that includes different table/column names or data, or even a more comprehensive version that might look like assignment 1 is NOT accepted. Also, introducing extra tables is NOT allowed.
- Q: Can I introduce temporary tables and delete them in the end? No.
- Mac users only: Please use either 'Numbers' or 'Microsoft excel 2011' to open the datasheet. Mac users that use Microsoft excel 2008 might get some date columns with wrong values.
- In case the date format in the data set does not match the date format in your database, e.g. one is YYYY-MM-DD and the other is MM-DD-YYYY, you may use either one but need to be consistent across all your queries. Alternatively, feel free to modify the date formats in excel file to match your needs. Our goal in this assignment is to make sure your queries work properly.
- Can a foreign key be composite? Yes, when it is pointing to a composite primary key.
- In queries that ask for 'highest', 'maximum', etc. in case of a tie, you should return all tuples that satisfy the condition.