CSCI 585- Database Systems

Spring 2015

Homework Assignment 2

Deadline: March 12th 11:50 PM

During assignment 1, you came up with an EER for the MyTube++ system. *Inspired* by MyTube++, 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.

When asked to return a 'user', unless explicitly asked to return e.g. the name of the user, it is fine to return only the user_id (PK) and returning all the fields is not mandatory.

1. List all users who registered during May 2012.

2. Average age of users who have more than 4 videos in their public channel. Please return the floor of the average value. E.g. floor of 26.7 is 26.

3. List the three hottest videos that are uploaded in December 2013 along with their View-Count and their Artist of the video. In case of a tie (i.e., multiple videos having equal number of view counts, just mentioning three videos is enough). Hottest videos are those visited the most.

4. Given the videos table, we are interested to get some insights about the trend in watching different *types* of videos in different *years*. Hence, write a query that returns the total view count across all videos belonging to each type uploaded in each year. Each row in the result set would include three columns: upload year, type, and total aggregated view count of videos.

5. Find a channel that its owner is the youngest member of the website.

6. List the personal information of user U-2 that is accessible to the user U-17. Note that the users can only see the other person's information if they have already established a friendship in the site.

7. For each user, write how many friend requests s/he has made (A), together with the count of users who accepted (B), rejected (C), and not yet decided (D) on the friendship request. Apparently, A = B+C+D

8. Which friend has been announced to be the Close-family by the most number of the users in MyTube++.

Submission Guidelines

1. Your submission should include one createdb.sql file, and one .sql file including all queries. Please make sure to write your name and student ID as a comment line in both files.

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 allowed but not required towards completing the assignment.
- There are a couple of ways to implement columns that include multiple values (e.g., 'v1, v2, v3'). We first explain different approaches and finally mention which one is accepted in this assignment. 1) use a string (and break it into multiple parts using functions when it comes to writing a query); 2) use an array type (if supported by the database as some databases do and some do not); 3) introduce another table to keep the multiple values. For consistency in this assignment, everyone is expected to use the first approach and use on string to represent the list of values.
- 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.
- 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: In query 8, all videos in favorite channel of u-2 belongs to u-19, but u-19 and u-2 are not even friends. Is it a mistake? No. It is okay that all 2 videos in the favorite channel of U-2 belongs to U-19 and one of them belongs to himself.

Also it is okay if someone that is not a friend of someone else at MyTube++, would share videos from the people that are not friends.

- Can a foreign key be composite? Yes, when it is pointing to a composite primary key.