Assignment 1

January 25, 2018

General Instructions

1. Use Postgres 9.3 for loading the datasets and performing experiments.

2. You have to submit four files – a report (*report.pdf*), two .sql files (*design.sql* and *test_queries.sql*) and one .csv file (*time.csv*). You will lose full points if the file names deviate from what is mentioned here.

3. The assignment is due on 29th January,11:59 PM. For each late day, you will lose 25% of points. We will also not accept partial submissions. For example, you cannot submit Q1 before deadline and Q2 after deadline.

4. The assignment can be done in groups of **at most** three students. The submission should be a .zip file named as **assignment1_<entry no.1>_<entry no.2>_<entry no.3>.zip**. Please make sure you follow this naming convention. Please mention your entry numbers in the file name (of the form 2014CSXXXXX), and **not the Kerberos IDs (like csXXXXXXX)**.

5. Plagiarism will fecth F grade on entire course.

Database Design

1. Consider the ER diagram shown in figure 1 for a student information system: The entities are:

   (a) student
   (b) teacher
   (c) course
   (d) section

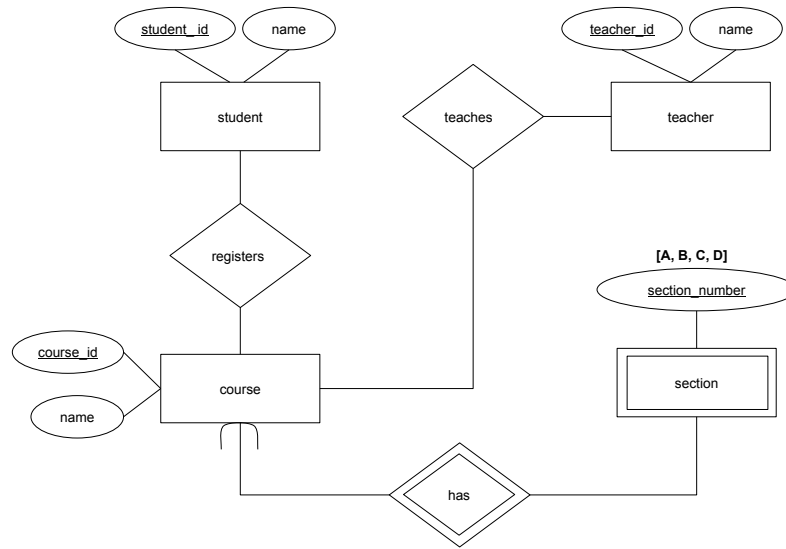   and relationships are:

   (a) registers
   (b) teaches

Figure 1: ER diagram for the student information system: all constraints to be considered are shown in the diagram.

Construct the relations (six tables listed above) according to the ER diagram using sql queries in Postgres. PLEASE MAKE SURE THAT THE TABLE AND COLUMN (ATTRIBUTE) NAMES ARE EXACTLY THE SAME AS THOSE SHOWN IN THE ER DIAGRAM. Specify proper constraints in your design as shown in the figure. Apart from that, enforce the following constraints:

(a) All tables should have a primary key.

(b) Deleting a tuple from "student" table should automatically delete entries from "registers" table.

(c) Deleting a tuple from "course" should automatically delete corresponding entries from "registers" and "teaches" table.

(d) Deleting a tuple from "teacher" table should automatically delete corresponding entries from "teaches" table.

ALL THE ATTRIBUTES (COLUMNS) SHOULD BE SET AS STRINGS UNLESS OTHERWISE STATED.

You should submit the following files:

(a) *design.sql*: It should contain *CREATE TABLE* statements for the six tables. You should name the tables and columns strictly as described in the diagram (including the fact that the names should be all in lowercase), because we will use these names during evaluation.

Any deviation from the naming standards would lead to an automatic grading of zero. No requests would be entertained in this regard.

(b) *test_queries.sql:* This file should contain some sample queries of your choice that you have performed on the tables. These may include (but not limited to) select, alter, insert, and update queries.

Please note that you must put all the constraints shown in the diagram. Marks will be deducted if you miss out any of them.

Your schema will be populated with tuples and evaluated by running sql queries. The queries will be made available after the evaluation. [35 points]

2. For this part of the question, you have to submit a *time.csv* file, which will contain the time taken (in seconds) to populate the *registers* table (around 500,000 tuples) created in the first question, using the following approaches:

(a) bulk load (find out how to do this),

(b) insert statements for each tuple, and

(c) programmatically, using JDBC.

For example, if bulk load takes 10 seconds, insert statements takes 7 seconds and JDBC takes 5 seconds, *time.csv* file will contain a single line like this:

10,7,5

Describe your observations in the pdf submission. [20 points]

You can enclose the three files (two for Q1 and one for Q2) in a directory named **assignment1_<entry no.1>_<entry no.2>_<entry no.3>**, and submit its zip (the zip should be named exactly as mentioned in the instructions).