Lab 5: Lather, Rinse, Repeat

Due date: There is no due date. The lab is to be completed during the October 25, 2007 lab period.

This lab does not involve an outright submission. It is designed to give you the opportunity to practice and the opportunity to fix previously existing errors in your datasets.

Lab Assignment

Fixing things up

This part of the assignemnt is to be done by each student individually, although I allow for consultations with me, or with your classmates, in case questions arise.

Your first task for this lab is to ensure that your work for Labs 2 and 3 (and Lab 4 if necessary) is complete. In particular, completness means the following:

- You have a working collection of SQL scripts that creates and populates each of the five databases: BAKERY, CARS, CSU, MARATHON and STUDENTS. Here "working" means *that when the scripts are run in a proper sequence (setup, populate, test), no errors ocur and expected behavior (tables are created, all insertions are made, all contents of the tables are displayed)* ***indeed occurs***.
- Your CREATE TABLE statements include all relevant constraints present in the database. In particular
  1. all relations **must** have primary keys.
  2. all foreign key relationships must be declared.

  You can optionally elect to specify various UNIQUE and CHECK constraints, but for the most part they are not required.

Please inspect your files, make any necessary changes, and keep a copy handy, as you would need to use it in a number of followup labs.

If you have any questions, talk to me during the lab period.

Trying out complex queries

This assignment can be done individually, but I *encourage* you to do it in pairs.

Your goal is to write SQL queries for each information need below. While no submission is required, some of these, or similar queries may appear in followup lab assignments. Therefore, I strongly recommend that each student (i) attempts to complete as many of these queries as possible (with the added benefit of being able to do so in collaboration with another student) and (ii) keeps a copy of (an) SQL script(s) for own records.

We will use just one dataset for this exercise: CSU. The list of queries is provided below.

1. Report the average, the maximum and the minimum enrollment among the CSU campuses in 2003.
2. Report the number of years in which enrollment in `'California Polytechnic State University - San Luis Obispo'` has exceeded 15000.
3. Report the latest year in which a CSU campus was founded.
4. Report the maximum total (gradutate and undergraduate) enrollment in `'Computer and Info. Sciences'` discipline in 2004.
5. Report the maximum number of degrees granted by a CSU campus from `'Los Angelos'` county after year 2000.
6. Report the total number of CSU campuses for which enrollment information exists for the year 1951.
7. For each campus report the average enrollment/per year between 1995 and now. Output the full name of the campus, and the average enrollment.
8. For each campus report the total amount of fees a student would pay between years 1996 and 2000.
9. For each campus report the highest recorded faculty FTE.
10. For each year report the highest undergraduate enrollment in a CSU campus.
11. For each discipline report the number of schools that had undergraduate students studying it in 2004.
12. For each campus for which data exists for more than 20 years, report the average and the maximum enrollment (for all years).

13. For each campus that averaged less than $2500 in fees/year report the average number of degrees granted during the period for which the fee data exists.
14. For each discipline taught in more than 12 campuses in 2004, report the number of campuses it is taught at.
15. For each campus which has taught `'Engineering'` discipline in 2004 report the number of faculty that year.