Group Assignment 2
CSC 343 - Fall 2019
University of Toronto Mississauga
Due: Thursday November 14th, 2019 by 11:59pm

Michael Liut

November 2, 2019

This assignment will reflect the entity-relationship diagram for the *Ministry of Health* (*MoH*). Your task is to perform some simple data analytics. On Quercus you will find the following:

a. `createTables.ddl`, a DDL file for CREATE TABLE statements.

b. `loadData.ddl`, a DDL file for INSERT TABLE statements.

c. `asg2ER.pdf`, a simplified schema depicted as an entity-relationship diagram.

d. `skeleton.sql`, an SQL skeleton file where you will put the solutions to Part I.

Please execute scripts `createTables.ddl` and `loadData.ddl` on your MySQL database on the UTM server, as you will use this schema for the questions below.

I. SQL (80 marks)

Written below are 11 questions, totaling 18 queries. Write and provide SQL statements for the following queries. These queries are to be written where indicated in the skeleton file (see d. above). Execute each of your SQL queries against your `MoH` database, and give the result of each query.

(q1) [3 marks] Identify all hospitals (name, city) with an annual budget greater than $3 million. Order the results in descending order of annual budget.

(q2) [4 marks] Identify all patients (their first name, last name, gender, date of birth) who are 40 years of age or less, are from the city of Toronto, and have been diagnosed with some type of Cancer. Do not include duplicates.
Hint: the `DATE` and `LIKE` function may be used, if needed.

(q3) a) [4 marks] Find the average salary for all physicians in each medical specialty.

b) [4 marks] Revise your query in part (a) to report the average salary for only those physicians working in Toronto or Hamilton. Report the average salary only for the medical specialties with at least 5 physicians in the group.

c) [4 marks] Report the average salary for nurses according to their years of service. Display your results such that the most experienced nurses are listed first.

(q4) [5 marks] How many patients were admitted to each hospital during August 5 to August 10, 2017? Show your results per hospital.

(q5) a) [4 marks] Which departments (return the department name) exist in all hospitals? (For example, we might expect that every hospital has a Cardiology department.)

b) [4 marks] Which department and hospital has the largest number of staff (physicians and nurses) working in that department?

c) [4 marks] Which department(s) are unique among all hospitals?
For example, the paediatric oncology department exists only in hospital A. Note that a department is unique if its count is = 1.

(q6) a) [4 marks] Identify all nurses (their first name, last name) who cared for no more than 3 patients (i.e., 0, 1, or 2 patients). Show your results in alphabetical order by last name.

b) [4 marks] Identify all patients with a poor prognosis and was cared for by a nurse in part (a).

(q7) [5 marks] On which date did the Hamilton General Hospital experience the largest number of patient admissions?

(q8) [6 marks] Each time a drug is prescribed in a prescription, and the patient fills the prescription, a sale is generated for the drug according to its cost price. Find the drug (drugcode, name) with the largest sales revenue. Also report the total sales amount for this drug.
Note: simply adding the drug costs, ignoring the dosage, is sufficient!

(q9) [6 marks] Find all patients (ID, first name, last name, gender) that have been diagnosed with Diabetes but have not yet taken a Red Blood Cell nor a Lymphocytes medical test.

(q10) a) [4 marks] For each physician in the Intensive Care Unit (department) at University of Toronto Medical Centre, return the disease and prognosis of each of their patients. Do not return duplicates.

b) [4 marks] For each of the patients in part (a), report the patient ID and the total cost the patient has spent in medical tests (regardless of the physician treating them). Display your results in decreasing order of total cost.

c) [4 marks] For each of the patients in part (a), report the patient ID and the total cost the patient has spent on medical drugs via prescriptions. Display your results in decreasing order of total cost.

(q11) [7 marks] Identify those patients (ID, first name, last name) that have been admitted to exactly two hospitals with an urgent or standard admissions category.

II. Relational Algebra (36 marks)

Given the SQL queries $1, 2, 3, 4, 10$, and 11, above, translate them into their Relational Algebra counterpart. [6 marks per relational algebra expression]

III. Indexes (10 marks)

This section is all about improving performance through the utilization of indexes. From the workload of the given queries in Part I., many display poor performance (i.e., increasing response times). Your task is to improve the performance of these queries by defining two indexes that should be implemented on these tables. For each index, state:

1. The attribute(s) the index is defined on.

2. The properties of the index (e.g., type of index, clustered/unclustered, etc...)

3. Which queries (q1 - q11) you think this index will help, and why.

Specifications

Grading

This is a group assignment to be completed in the pairs (i.e. a team of 2 people); selected in Assignment 0 unless advanced written approval has been given by the Course Instructor. This assignment is worth 10.0% of your final grade in this course and will be graded out of 126-points.

Files that do not execute and/or compile will receive a grade of 0. Files that are submitted of the incorrect type will receive a grade of 0.

Plagiarism

Please refer to the course outline and introduction slides. To serve as a reminder: Turnitin will be used for all written work and MOSS for all code submissions. UTM's policy on Academic Integrity: http://academicintegrity.utoronto.ca/

Submission

All files are to be submitted using the MarkUs platform (https://mcsmark.utm.utoronto.ca/csc343f19/). Only one person from each group is required to submit the files. You may submit as many times as you like, this is a repository for you and your partner to work in. Please ensure your <u>answers are typed</u> (yes, the Relational Algebra too; try LaTeX) and submissions are clearly legible.

Include your, and your partner's, full name and student ID number in <u>all files</u>. Upload four files with the indicated file extensions (no compression based `.tar`, `.zip`, `.rar` files).

- For Part I: Submit your SQL statements in a script file called `queries.sql`. Rename the `skeleton.sql` to match this; your queries are to be put in this file.

  - Ensure your SQL statements are syntactically correct and that they are executable on the MCS's MySQL server. Non-executable files will not be marked.
  - Queries <u>will not be</u> individually run, rather the entire file as a whole will be executed.

  In addition, for Part I: Submit the corresponding query results in a file called `queries.results`.

  - This is an ASCII file, filled with text, that has been renamed for the purposes of this assignment.
  - Clearly label and comment which query corresponds to which result tuples.

- For Part II: Submit your relational algebra expressions in a file named `ra.pdf`.

  - Your relational algebra must be typed. No exceptions.

- For Part III: Submit your index recommendations in a file named `index.pdf`.

  - Your responses must be typed. No exceptions.

**This means you will submit a total of four files: <u>ra.pdf</u>, <u>queries.sql</u>,
<u>queries.results</u> and <u>index.pdf</u>.**

Please note that late assignments will be docked 20% per calendar day of lateness and after three (3) calendar days, the assignment will no longer be accepted.