Assignment 3 (CMPT 354): SQL Programming and Normalization
Due: Nov 25 (Sat), 11:59pm
Total mark: 94
Weight=24%

Submission instruction: (1) This assignment must be done by each student independently. (2) Submission is through coursys.sfu.ca in a single pdf file with maximum file size of 10MB. Late submission will not be accepted as answers will be posted immediately after deadline. (3) The student is responsible for submitting the assignment successfully before the deadline, and only the submission that is in the system before the deadline will be accepted.

This assignment has two parts: Part 1 on SQL programming, and Part 2 on functional dependencies and normalization. The detail of Part 2 will be added later. Since Part 1 involves the setup of DMBS servers and running SQL, you should start working on this part as early as possible.

The note to TA: The deadline for finishing marking is Dec 4, 2023

Part 1: SQL Programming (46 marks)

This assignment will use the following "Department-Store Database". The underlined fields are the attributes of primary keys.

Department Store Database

Employee relation:

eid	name	salary	dept
111	Jane	8000	Household
222	Anderson	8000	Toy
333	Morgan	10000	Cosmetics
444	Lewis	12000	Stationery
555	Nelson	6000	Toy
666	Hoffman	16000	Cosmetics

Sales relation:

dept	item
Stationery	pen
Cosmetics	lipstick
Toy	puzzle
Stationery	ink
Household	disk
Sports	skates
Toy	lipstick

Types relation:

item	color
pen	red
lipstick	red
pen	black
puzzle	black
ink	red
ink	blue

DBMS Environments. All SQL statements in Part 1 will be run on a DBMS server. The choices of DBMS servers are SQLite or MYSQL. Alternatively, you can also use the Microsoft SQL Server 2019 on CYPRESS.csil.sfu.ca - the CSIL SQL Server. A SQL database on Microsoft SQL Server has been set up for each of you (but you need to create the tables and enter data into the tables). To use Microsoft SQL Server, you can log into CSIL computers that already have SQL Server Management Studio installed, and access CSIL SQL Server by following the email instruction sent to you at the beginning of this semester from helpdesk@cs.sfu.ca.

There are a lot of online materials on setting up MySQL and SQLite on your computers. Here are a few but you can find more online:

https://www.kdnuggets.com/2022/09/free-sql-database-course.html (Installation on Windows OS Computers, by Matthew Mayo)

https://dev.mysql.com/doc/refman/8.0/en/ (MySQL 8.0 Reference Manual)

SQLite: https://www.sqlite.org/index.html, https://www.sqlitetutorial.net/

For each task below, write the SQL statement for finding the answer requested by the task. For submission purpose, you need to specify the DBMS environment used (i.e., MySQL, SQLite, Microsoft SQL Server) and include the screenshot of the SQL statements and the answer returned by running the SQL statements.

Task 1 (5 marks): Create the above database schema using CREATE TABLE statements, including primary key constraints, and the constraint that salary is integer in the range [5000,20000]. You can assume CHAR (20) type for all other attributes.

Task 2 (5 marks): Insert the above records into the tables using INSERT statements.

Task 3 (36 marks, 6 marks each): Compute the answers to the following queries using SELECT statements. Your SQL statements should be correct for ALL instances of data, not just for the above instance. For example, to find the departments that have a larger average salary than that of "Stationery" department, we do not accept the SQL that uses 12000 as the average salary of "Stationery" department because it only works for the above instance.

1. Compute the maximum salary for each department that sells at least two distinct items.

2. Compute the names of the employees who work in a department that sells some item in black color

3. For each department that has a larger average salary than that of "Stationery" department, find its average salary.

4. Find the number of the departments that have a smaller average salary than that of "Stationery" department.

5. Which department pays every of its employees at least 7000?

6. Which departments sell all items sold by "Cosmetics" department

Part 2: FD and Normalization (48 marks)

Question 1 (10 marks, 5 marks for correct instances and 5 marks for discussion). Consider a relation R=(S, A, C, D, T), representing that the student (S) has the address (A), takes the course (C) from the teacher (T) who is from the dept (D). Assume the following FDs F hold on R:

C -> T: a course determines its teacher, i.e., each course is taught by only one teacher
S -> A: a student determines its address, i.e., each student has only one address.
T -> D: a teacher determines its department, i.e., each teacher is from only one department.

However, a student can take multiple courses, a teacher can teach multiple courses, and each department can have multiple teachers. Use an instance of R to explain data redundancy, update anomaly, insertion anomaly, and deletion anomaly that may exist for the above schema R and FDs.

Question 2 (38 marks) Continue with the R and FDs F in Question 1.

1. (3 marks) Find all keys of R with respect to F.
2. (3 marks) Test if R in BCNF with respect to F, why?
3. (10 marks) Produce a BCNF decomposition through a series of binary decomposition. For each binary decomposition, tell the FD used for the decomposition and show the FDs holding on the decomposed tables.
4. (3 marks) Explain why the decomposed tables produced in 3 is a better representation than the original single table R.
5. (3 marks) Is the final decomposition in 3 dependency-preserving, why
6. (3 marks) Is the original schema R in 3NF with respect to F, why
7. (10 marks0 If the answer to 6 is no, produce a 3NF decomposition that is lossless and dependency-preserving.
8. (3 marks) Is the decomposition produced in 7 in BCNF?