Department of CSE, I I T Madras
CS3700: Introduction to Database Systems
Take-Home, Open Book Examination, Aug-Nov 2021.
Points: 30, Examination Duration: Nov 20th, 2021 (10:30am-12:30pm)

Note: Please be concise in your answers. Ensure that your hand-writing is legible and not too small to read. Write your roll number and also page number on all the sheets you use. Please use A4 sheets only. At the end of the examination, you are required to scan/photograph the sheets and make a single PDF file with your **roll number** as the **file name**. Upload the file to moodle before the deadline. The PDF file should not have more than 10 pages.

Task #0.

Please write down the following in your first answer sheet, add your signature and name.

I understand that the take-home examination is to test my own understanding of the subject and I will offer an honest attempt to meet this goal. I will not communicate with anyone else while working, answering the examination problems/tasks. I will not post the questions to any public fora and seek answers. I will not help other course-mates taking the examination in any manner. I understand that I can consult books, course slides and videos by the teacher while answering the examination. I also understand that I should not copy materials from books, internet sites etc as part of my answers for the examination questions/tasks. These restrictions apply for the whole examination duration(till the deadline for submission).

Task #1.

Look at the appendix for three descriptions of domains. Use the following method to find the domain description assigned to you based on your roll number.

CS18Bxxx: Compute xxx mod 3. If that is d, choose description (d+1).
CS19Byyy: Compute yyy mod 3. If that is d, choose description (d+1).
Others: ZZxxBsss: If sss is even choose (1) else choose (2).

For the assigned domain description, draw the ER Model diagram and briefly explain the design choices made by you. In the diagram, make use of (min, max) notation to convey the relevant structural constraints and indicate the keys clearly. (If necessary, you are allowed to make minor clarifying assumptions. State these assumptions clearly in your answer sheet.) Indicate the keys, role names (if any) appropriately.

(10 points)

Task #2.

Consider the query frameworks of relational algebra (RA) discussed in the course and also that of the tuple relational calculus (TRC). Develop a formal proof for the statement: Every query expressed in relational algebra can be equivalently written in safe TRC. (You may assume that the underlying relational scheme has relations R_1, R_2, ..., R_k for some finite k and the attributes are named A_1, A_2, ..., A_m for some finite m and no pair of relations have any common attribute names. (You may also make some additional assumptions as long as there is no loss of generality. State the assumptions clearly in your answer sheet.)

(10 points)

Task #3.

1. Consider the following set of FDs:
   $F = \{ ABCD \rightarrow E, E \rightarrow D, A \rightarrow B, AC \rightarrow D \}$
   Compute a minimal cover of F using the algorithm discussed in the class.

2. Show that if we interchange Step 3 (minimizing LHS of FDs) and Step 4 (removing redundant FDs) of the algorithm for minimal covers and run it on the above set of FDs, we do not get the minimal cover.

3. Devise your own example set of FDs F to illustrate the point: To get a minimal cover of a set of FDs F, it is important to remove the redundant attributes on LHS of FDs first and then remove redundant FDs from the resulting set.

4. Devise your own example set of FDs F to illustrate the point: There may be two different minimal covers for a given set of FDs F. Compute and write down the minimal covers.

(10 points)

Appendix:
Domain Descriptions for **Task #1**.

1. The National Physical Address Tracking System (N-PATS) is responsible for the task of maintaining accurate information about where each citizen of the country is located and tracking any changes to the addresses. It issues proof of address (POA) certificates to citizens for use in various other contexts. Each person is assigned a unique citizen Id. The system keeps information about the address of the birthplace, parents of a person and a copy of supporting documents. The current residential address of a person and the date since when that address is applicable is to be recorded. In addition, a person may have several past residential addresses; these and the duration of time for which the addresses are applicable need to be recorded. Each time a change in address is requested, digital copies of the supporting documents and the date of change are to be stored. A residential address can be shared by several persons, typically from the same family. The address has the following details associated with it: the house number, optional house name, optional landmark, street name, locality name(s), city/town name, optional district name, state name, postal code. In the case of flats or apartments, instead of a house number, there will be a flat number and name of the apartment complex. Relevant details about a person also need to be recorded: first name, several middle names, last name or family name, date of birth, gender, blood group. Finally, a person has several contact phone numbers – residential landline numbers, one or more personal cell numbers and these need to be recorded.

2. This domain deals with scientific publications. A manuscript is a document, which is a record of research work and has a title, a list of authors (each having an affiliating address), a set of keywords and the date of creation. A scientific conference has a name, an organizing body and is held once in a year. Each occurrence of such a conference is associated with the year in which it is held, place where it is held and beginning and ending dates. It also results in a book called the proceedings of the conference of that year in which accepted manuscripts are published. A number of manuscripts are submitted by researchers to a particular occurrence of a conference and after due reviews, a subset of the manuscripts are published by the conference in its proceedings. A manuscript might be submitted to many conferences. However, it is published in exactly one conference. Each submission is associated with one or more reviews (each given by a researcher), which are short documents that contain comments/remarks on the manuscript and a recommended decision (accept/reject). An information system is being designed to keep track of all the above details about the domain.

3. A typical relational database management system (RDBMS) such as Oracle, DB2 etc maintains several databases (all relational, of course) and meta information about all of them. Each of these databases has a name and an administrator user who has full access to it and several other users who are allowed to access the database with various permission levels (Say, there are 5 levels). For each such database, the RDBMS needs to keep track of all its meta-data. The meta-data of a DB has all its schema details, such as relation names, attribute names, data-types, keys, foreign keys, RICs etc. In addition, it has details about the number of distinct values of an attribute and optionally histograms associated with attributes. The RDBMS uses an internal relational database to store all this information. We require the ER model of this internal relational database that can hold all the required information about the databases in the system.

-------///------