

Department of CSE, I I T Madras
CS3700: Introduction to Database Systems
Take-Home, Open Book Examination, Aug-Nov 2020.
Points: 100, Examination Duration: Dec 21st, 2020 - Dec 22nd, 2020

Note: Please be concise in your answers. Ensure that your hand-writing is legible and not too small to read. Write your roll 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.

Task #0.

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

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.

- a) 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.

CS17Bxxx: Compute $\text{xxx} \bmod 3$. If that is d , choose description $(d+1)$.

CS18Byyy: If yyy is even, choose (2) else choose (3).

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 sheets.) Indicate the keys, role names(if any) appropriately. (12 points)

- b) Using the standard method of deriving relational schema from ER models, derive the relational schema for the ER model you have designed in part(a) above. Clearly state the key(s) and foreign key(s) for each relation and also draw the arrows to show the references being made by the foreign keys. (8 points)

Task #2.

Use the Academic Institute DB Scheme discussed in the class for this task. DO NOT make use of any examples mentioned in the course slides/tutorials/previous question papers, your own assignment submissions, for this task.

- a) Prepare and write down a data retrieval request that involves the use of a universal quantifier and the implication operator, when expressed using TRC. (Like, for example, the query that asked for students who have completed *all* the courses taught by Prof Ramanujam, discussed in the class). For the query prepared by you, write down the TRC expression and also write the equivalent relational algebra expression. Briefly explain your query expressions. (6 points)
- b) Prepare a data retrieval request that requires at least two joins and checking of at least 2 selection conditions. Write down the SQL query for this. Use the principles of heuristic optimization techniques discussed in the course to obtain the relational algebra expression tree that can be used to execute the above SQL query. Draw the intermediate trees and mention the heuristics used for the transformation of the tree. (8 points)
- c) Prepare a data retrieval request that requires two levels of data aggregation (that is aggregation needs to be done on aggregated data) and needs the use of a view. Write down the query and the needed view definition and the SQL query. Briefly explain your approach. (6 points)

Task #3.

The spirit of this task is to demonstrate your understanding of the indexing structures used in disk file systems:

- a) Consider a B⁺-tree that has order of the internal-node as 3 and that of a leaf-node also as 3. Devise a sequence of insertions of elements (integers) into an initially empty tree so that it grows to have 2 levels (including the root) in addition to the leaf level. Draw the tree at those stages where splits occur. (10 points)
- b) Consider the linear hashing technique for organising dynamic files, using a family of hash functions, discussed in the class. Assume that the family of hash-functions is as given in the slides of the course. Study the technique and write pseudocode for inserting a given hash field value into the structure so that splitting of blocks, switching of hash functions are all taken care of. You can organise your pseudocode into a suitable number of functions. (10 points)

Task #4.

- a) Write down the relational scheme of the database developed by your group as part of the course assignment submission. (If necessary, you are allowed to make few modifications so that the DB scheme is more comprehensive and apt for this task). Give a brief description of the intended meaning of attributes. Now, write down the functional dependencies that hold among the attributes of your scheme. For each FD, write a brief justification as to why it should hold. For each relation in your scheme, identify what is the highest normal form it satisfies wrt to the FDs applicable for that relation. (10 points)
- b) Consider the algorithm, (call it Algorithm L-Test), discussed in the class to test if a given decomposition of a relation scheme is lossless or not wrt to a given set of FDs. Attempt a proof for the following: If the matrix operated on by L-Test doesn't have a row of all "a" symbols at the end of its running then the given decomposition is lossy wrt to the given FDs. (10 points)

Task #5.

- a) Develop a proof for the following statement: A schedule of operations of a transaction set is conflict-serializable if and only if the precedence graph of the schedule is acyclic. Ensure that the proof works for cases where the cycle length (# of edges) may be more than 2 also. (6 points)
- b) Supply your own examples of schedules (involving at least 3 transactions) for the following: (Render the schedules as tables as shown in the course slides) (8 points)
 - 1) A non-recoverable but conflict-serializable schedule.
 - 2) A recoverable schedule that is not conflict-serializable.
 - 3) A recoverable, conflict-serializable schedule that is not cascadeless.
 - 4) A recoverable, conflict-serializable and cascadeless schedule.
- c) Design an example (consisting of operations of at least 2 transactions) to illustrate error-recovery from a crash when REDO logging method is employed. Show the log entries seen after the system restarts and explain the recovery operations. The example you construct should clearly demonstrate why it is important to read the log in the forward direction. (6 points)

Appendix:

Domain Descriptions for Task #1.

1. This domain is to do with meetings in a business enterprise. The enterprise members participate in a number of meetings. The enterprise seeks to develop a system that keeps track of the associated information. Every meeting has a name and set of members who are supposed to attend the meeting. A meeting is convened (or called) by a particular employee. The members who have actually attended the meetings also need to be recorded. A meeting is scheduled to take place at a venue on a certain date and at a specific time and these details need to be recorded. Meetings discuss agenda items and record resolutions for each item discussed. The resolutions of a meeting become part of the minutes of the meeting. The minutes of a meeting can become agenda items for another meeting. A meeting could be a one-off meeting or could be repetitive meeting – here the sequence number of the meeting needs to be recorded. An item of discussion has a short title and an associated description. Appropriate details of the attendees of the meeting, such as name, email address, employee Id and designation need to be recorded.
2. In a library, information about books, borrowers, suppliers and locations inside the library, where books are placed, need to be recorded in an information system. A book copy is borrowed by a user/borrower and is usually issued out for a specific period of time. These periods can be short-term, medium-term and semester-long. Details of book – title, authors, ISBN number, accession number, subject category, number of copies – need to be recorded. A book is placed in a specific location of the library, when it is not lent. A room number and a rack number identify a location. A borrower has a unique Id, name, email address, phone number, department and academic program (if any). Also, there are four types of borrowers – course students, research students, staff and faculty members. A number of suppliers supply books to the library and the library needs to keep track as to who supplied what book to the library and at what price. Supplier details

– name, address, email, phone number, bank, and account number – need to be recorded.

3. 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.