COMP2400/6240 - Relational Databases

Assignment 2 (Database Theory)

Due date: 23:59, 12 October 2020

Instructions:

- **This assignment must be done individually (no group work)**.
- This assignment will count for 15% of the final grade. Marks are assigned for the process of finding a solution, not only for the result. Hence, include all essential ideas and steps that are necessary to derive a solution.
- You must submit a single PDF file named as "u1234567.pdf" (replace u1234567 with your UID). Make sure you only upload a PDF file, not a Word or text file.
- You should try your best to type the solutions. The scanned images of handwritten texts and equations can be unreadable for marking. As for the EER diagram, you are highly recommended to export a JPEG file from TerraER and include it in the PDF file.
- Late submission is not granted under any circumstance. You will be marked on whatever you have submitted at the time of the deadline. Please take careful note of deadlines and adhere to them. Of course, if you find yourself in a situation beyond your control that you believe significantly affects an assessment, you should follow the ANU's special consideration process (http://www.anu.edu.au/students/program-administration/assessments-exams/special-assessment-consideration).
- **Plagiarism will attract academic penalties in accordance with the ANU guidelines. A student in this course is expected to be able to explain and defend any submitted assessment item. The course convener can conduct or initiate an additional interview about any submitted assessment item for any student. If there is a significant discrepancy between the two forms oexpressionsf assessment, it will be automatically treated as a case of suspected academic misconduct.**

Question 1 **4 Marks**

A sports center aims to setup a database to maintain the following information.

The database stores the name, the address and a unique ID of each employee. The employees are classified into three categories: trainers, receptionists and administrators. Each trainer has one or more specialties (e.g., swimming, tennis and squash etc). Each receptionist can be reached through a landline phone number. Each receptionist must be supervised by one administrator and a administrator may supervise multiple receptionists. Each member of this sports center is identified by their membership id and the database also stores their names and emails. Each week a trainer may offer regular training classes that members can participate in. Each training class should focus on one of the specialities of the corresponding trainer and must have a class size no more than 10. Each training class from a trainer is referred to by the day of the week (i.e., Monday, Tuesday, ...) because each trainer offers at most one training class per day. After a member participated in a training class, the member may submit feedback to a receptionist and the feedback contains a rating from 1 (very unsatisfied) to 5 (very satisfied), a date and detailed description.

Your task is to design an Enhanced Entity Relationship (EER) diagram for the above database, which should include entities, relationships, attributes and constraints wherever appropriate (you can make more assumptions if necessary).

You also need to identify the requirements that cannot be captured in an EER-diagram.

Question 2 **3 Marks**

Consider the relation schema R={A, B, C, D, E} and the following set Σ of FDs:

- A → B
- AB → C
- BC → AE
- DE → AB

2.1 What are the candidate keys of R? Justify your answer (i.e., include the main steps used for finding the candidate keys). (1 Mark)

2.2 Find a minimal cover of Σ. Justify your answer (i.e., include the main steps used for finding a minimal cover). (2 Mark)

Question 3 **2 Marks**

Consider the relation schema APPOINTMENT={Customer, Branch, Date, Time, Staff, Room} and the following set Σ of FDs:

- Customer, Branch, Date, Time → Staff, Room
- Branch, Date, Time, Room → Customer
- Branch, Date, Room → Staff
- Staff, Date → Branch, Room
- Staff → Branch

Is the above relation schema APPOINTMENT in BCNF? If not, identify a BCNF decomposition for APPOINTMENT. You need to include the main steps used for identifying the BCNF decomposition. Check if this BCNF decomposition is dependency preserving. (2 Mark)

Question 4 **6 Marks**

The following table contains the relational algebra operators covered in our course:

$\sigma_{\varphi}R$	Selection by condition φ
$\pi_{A_1,\ldots,A_n}R$	Projection onto the set of attributes $\{A_1 \ldots, A_n\}$
$\rho_{R'(A_1,\ldots,A_n)}R$	Renaming the relation name to $R^{'}$ and attribute names to $A_1, \ldots, A_n$
$\rho_{R'}R$	Renaming the relation name to $R^{'}$
$\rho_{(A_1,\ldots,A_n)}R$	Renaming the attribute names to $A_1, \ldots, A_n$
$R_1 \cup R_2$	Union of two relations R_1 and R_2
$R_1 \cap R_2$	Intersection of two relations R_1 and R_2
$R_1 - R_2$	Difference of two relations R_1 and R_2
$R_1 \times R_2$	Cartesian product of two relations R_1 and R_2
$R_1 \bowtie_{\varphi} R_2$	Join of two relations R_1 and R_2 with the join condition φ
$R_1 \bowtie R_2$	Natural join of two relations R_1 and R_2
$\varphi_1 \bigwedge \varphi_2$	condition φ_1 AND condition φ_2
$\varphi_1 \bigvee \varphi_2$	condition φ_1 OR condition φ_2

Consider the following relation schemas:

STUDENT={SID, Name, College, Address, Phone} with the primary key {SID},

COURSE={CourseNo, CourseName, Semester} with the primary key {CourseNo, Semester},

LECTURER={LID, Name, College, Email, CourseNo, Semester} with the primary key {LID, CourseNo, Semester} and the foreign key: [CourseNo,Semester]⊆COURSE[CourseNo,Semester],

TUTOR={TID, Email, CourseNo, Semester} with the primary key {TID, CourseNo, Semester} and the foreign keys: [CourseNo,Semester]⊆COURSE[CourseNo,Semester] and [TID]⊆STUDENT[SID],

ENROL={SID, CourseNo, Semester, Unit, Status} with the primary key {SID, CourseNo, Semester} and the foreign keys: [CourseNo,Semester]⊆COURSE[CourseNo,Semester] and [SID]⊆STUDENT[SID].

4.1 Answer the following questions using relational algebra queries. You should only use the relational algebra operators in the above table. You are encouraged to use relational algebra expressions to represent intermediate results if needed. (3 Mark)

[a] List the CourseNo of courses without any tutors in 'S2 2020'? (1 Mark)

[b] List the names of students who have been enrolled in a same course for at least two times (i.e., in different semesters). (1 Mark)

[c] List the emails of lecturers who had taught exactly two courses in 'S1 2020'. (1 Mark)

4.2 Optimize the following relational algebra queries (Your marks will depend on how well you present the key ideas of query optimization in your answer). In addition to this, draw the query trees correspond to queries before and after your optimisation. (3 Mark)

[a] $\pi_{LID,SID}(\sigma_{(Student.Name=Lecturer.Name)\wedge(Student.College='CECS')}(\text{LECTURER} \times \text{STUDENT}))$ (1.5 Mark)

[b] $\pi_{SID}(\sigma_{(College='CECS')\vee(Status='withdrawn')}(\text{COURSE} \bowtie \text{ENROL} \bowtie \text{STUDENT}))$ (1.5 Mark)

+++++