ISYS2120 Assignment 4 (sem1, 2022)

Due: Sunday 30 October, 11:59pm Sydney time
Value: 5% [4% for the group as a whole, and 1% individual component]
This assignment is done in **groups of up to 5 students** (aim for exactly 4 members, but it may happen that sometimes a group is smaller or larger, eg if there are not enough students in a lab). We ask that all students in a group be attending the same lab session, so you can work together more easily, and show progress to the lab demonstrator each week.

Procedure: In week 11 lab, you should form a group. It is expected that most people will stay with the group they had joined in Asst3 but it is not necessary. There will be separate Canvas groups for this assessment; we will initialize them as copies of what were formed for asst2, but membership can change as described below. If any student wishes to not be with their asst3 team, they can simply remove themselves from the initial group; this should also be done if you were in a group from a class that you will not be attending in future. The lab demonstrator will then work with anyone who is unassigned, to form groups properly. If necessary, the demonstrator or coordinator may also rearrange group membership.
If, during the course of the assignment work, there is a dispute among group members that you can't resolve, or that will impact your group's capacity to complete the task well, you need to inform the unit coordinator, alan.fekete@sydney.edu.au. Make sure that your email names the group, and is explicit about the difficulty; also make sure this email is copied to all the members of the group. We need to know about problems in time to help fix them, so set early deadlines for group members, and deal with non-performance promptly (don't wait till a few days before the work is due, to complain that someone is not doing their share). If necessary, the coordinator will split a group, and leave anyone who doesn't participate effectively in a group by themself.
How to submit your work: produce a PDF file called <groupName>_Asst4.pdf, containing the solutions to the questions given below. This file should be uploaded at the corresponding submission link in Canvas, submitted by only one member of the group.

Weekly progress steps
Week 11: In lab, group membership should be settled, (and communication should be arranged for the rest of the assessment). Also, the members should agree about who will be leading the work on each question or subpart of a question (for subparts, being the leader means writing a complete solution, which other members can then check and suggest improvements). We strongly advise that, whenever a question has multiple subparts, that every member be the leader on at least one subpart, so that everyone gets practice with the topic.

Week 12: each student should bring to lab their initial solutions for the subparts where they are leading (from A, B.1 and C). If a member misses the lab session, they should upload a file with their initial answers. During or after the lab, the group should coordinate to combine their work and produce a single pdf with answers that they all agree to submit, responding to all the questions

The file to be submitted on Canvas: a pdf file that contains the agreed answers to all the questions. Each question or subpart should indicate which member was the leader in answering.

Marking criteria

- Group assessment of each question with subparts (A, B.1, C): 1 – all subparts correct, 0.5 at least 3 subparts correct.
- Group assessment of the questions without subparts (B.2, B.3): 0.5 – correct, 0.25 shows understanding of some of the main ideas, with minor errors.
- Individual component: 0.5 for showing effort and progress in the week 12 lab (or file upload). 0.5 for taking the lead in at least one subpart from each of the questions: A, B.1, and C.

The questions:

A [relational algebra] (1 point) has subparts
Note: if you have difficulty producing relational algebra symbols, you may use text, with SEL_{C} to represent σ_C and PROJ_{D} to represent π_D and X JOIN Y to represent X ⋈ Y.

This question's parts all refer to a relational schema as follows
Patient(AdmitNo, PName, Insurance)
Nurse(NStaffNo, NName, NSalary)
Doctor(DStaffNo, DName, DSalary)
Locate(AdmitNo, Ward)
Supervision(NStaffNo, SupervisorNStaffNo)
Assign(NStaffNo, Ward)
Illness(AdmitNo, Disease)
Expertise(DStaffNo, Disease)
Treatment(AdmitNo, DStaffNo, Disease)

A1(i) Give a relational algebra expression to calculate the table containing the AdmitNo, PName for those patients who have the illness 'COVID'.

A(ii) Give a relational algebra expression to calculate the table containing the DStaffNo, DName of those Doctors who treat a patient who has insurance 'NIB'.

A(iii) Give a relational algebra expression to calculate the table containing the NStaffNo and DStaffNo of a pair of doctor and nurse where the doctor treats a patient who is located on the ward to which the nurse is assigned.

A(v) Give a relational algebra expression to calculate the table containing theAdmitNo and DStaffNo of a pair of patient and doctor, where the patient has an illness in which the doctor has expertise.

A(v) Give a relational algebra expression to calculate the table containing the NStaffNo, NName and DStaffNo, DName for pairs of Doctor and Nurse where the Nurse is assigned to a ward where there is a Patient who is treated by the Doctor for the disease "Anthrax".

B [relational design theory] (2 points)
Consider the following relational design, which collects data about the usage of drugs at several hospitals.
DrugUsage(HospitalName, HospitalNumber, DiseaseCode, DrugName, SizeofDose, Cost, ManufacturerName, ManufacturerAddress)
The following functional dependencies are valid in this schema:

- HospitalNumber → HospitalName
- HospitalName → HospitalNumber
- DrugName → DiseaseCode, ManufacturerName
- DrugName, SizeofDose → Cost
- ManufacturerName → ManufacturerAddress

B.1 (1 point) has subparts
B.1(i) Calculate the attribute closure HospitalNumber+. Show your working.
B.1(ii) Calculate the attribute closure HospitalName+. Show your working.
B.1(iii) Calculate the attribute closure DrugName+. Show your working
B.1(iv) Calculate the attribute closure (DrugName, SizeOfDose)+. Show your working.
B.1(v). Calculate the attribute closure ManufacturerName+. Show your working.

B2. (0.5 point) State whether the relation DrugUsage is in BCNF. Show your working.

B3. (0.5 point) Give a decomposition of the relation DrugUsage, into two or more relations, with the properties that the decomposition is both lossless-join and dependency preserving. Explain how you know that the decomposition has these properties. For each relation in the decomposition, state whether or not that relation is itself in BCNF, and explain your working.

C. [index choices] (1 point) has subparts
Suppose a database has a table T(a,b,c,d) with primary key being the attribute a. For each query below, explain whether the query can be answered efficiently without creating any extra index; if the default structure does not allow efficient answering, you should write a CREATE INDEX statement that will improve performance for this query. You should also note whether or not the index you propose covers the query.

C(i)
```
SELECT *
FROM T
WHERE T.b = 53;
```

C(ii)
```
SELECT T.c
FROM T
WHERE T.a = 'XYZ';
```

C(iii)
```
SELECT T.a, T.c
FROM T
WHERE T.c > 34 AND T.c < 47;
```
C(iv)
```
SELECT T.a
FROM T
WHERE T.b = 53 AND T.d = 81;
```

C(v)
```
SELECT *
FROM T
WHERE T.d = 81 AND T.b BETWEEN 49 AND  61;
```