COMP5112
Data Structures and Database Systems
Department of Computing
The Hong Kong Polytechnic University
Quiz 2 - Solution Guide
Subject Lecturer: Dr. Kevin Yuen
Wednesday, 27 Nov 2024

Question 1 [40 marks]

Part of the project selection database schema contains two relations as below.

- Project (ProjectID, InstructorID, Title, Abstract)
- Instructor (InstructorID, FirstName, LastName, Phone, Email)

The *InstructorID* attribute in the *Project* relation is a foreign key to the *Instructor* relation. All data fields are in string format. *Email* and *Phone* are unique in *instructor* relation. Answer the questions below.

1(a) (20 marks)

At the current timestamp, all project relation data is given in Table 1a. **Write** the relational algebra expressions and the corresponding SQL statements, respectively, that will result in Table 1b. The display order of tuples listed in the table can be omitted.

ProjectID	InstructorID	Title	Abstract
P1	N1	T1	A1
P2	N1	T2	A2
P3	N2	T3	A3
P5	N2	T5	A5
P7	N4	T7	A7

Table 1a: *Project* relation data (given)

ProjectID	InstructorID	Title	Abstract
P1	N1	T1	A1
P2	N1	T2	A2
P3	N2	T3	A3
P4	N3	T4	A4
P5	N2	T5	A5
P6	N3	T6	A6

Table 1b: *Project* relation data (final)

Solution:

relational algebra expressions	SQL statements
$Project \leftarrow Project - \sigma_{projectID='P7'}(Project)$ $Project \leftarrow Project \cup \{ ('P3','N2','T3','A3') \}$ $Project \leftarrow Project \cup \{ ('P6','N3','T6','A6') \}$	DELETE FROM Project WHERE ProjectID='P7' INSERT INTO Project VALUES ('P4','N3','T4','A4'); INSERT INTO Project VALUES ('P6','N3','T6','A6');

- The order for insert and delete is not important.
- Accept the other solutions.

1(b) (10 marks)

Write an SQL statement to find the number of projects offered by each instructor.

Solution

```
select InstructorID, count(ProjectID) from Project
group by InstructorID;
```

- if "InstructorID" in "select InstructorID" is missing, 2 marks are reduced.

1(c) **(10 marks)**

Write a relational algebra expression to find the last name, phone and email address of each instructor, who does not offer any projects.

Solution

There are no more than one instructors who have the same phone and email address.

$$\Pi_{lastName,phone,email}(instructor) - \Pi_{lastName,phone,email}(instructor \bowtie Project)$$

- Accept the other solutions. For example, using instructionID in the algebra expression.

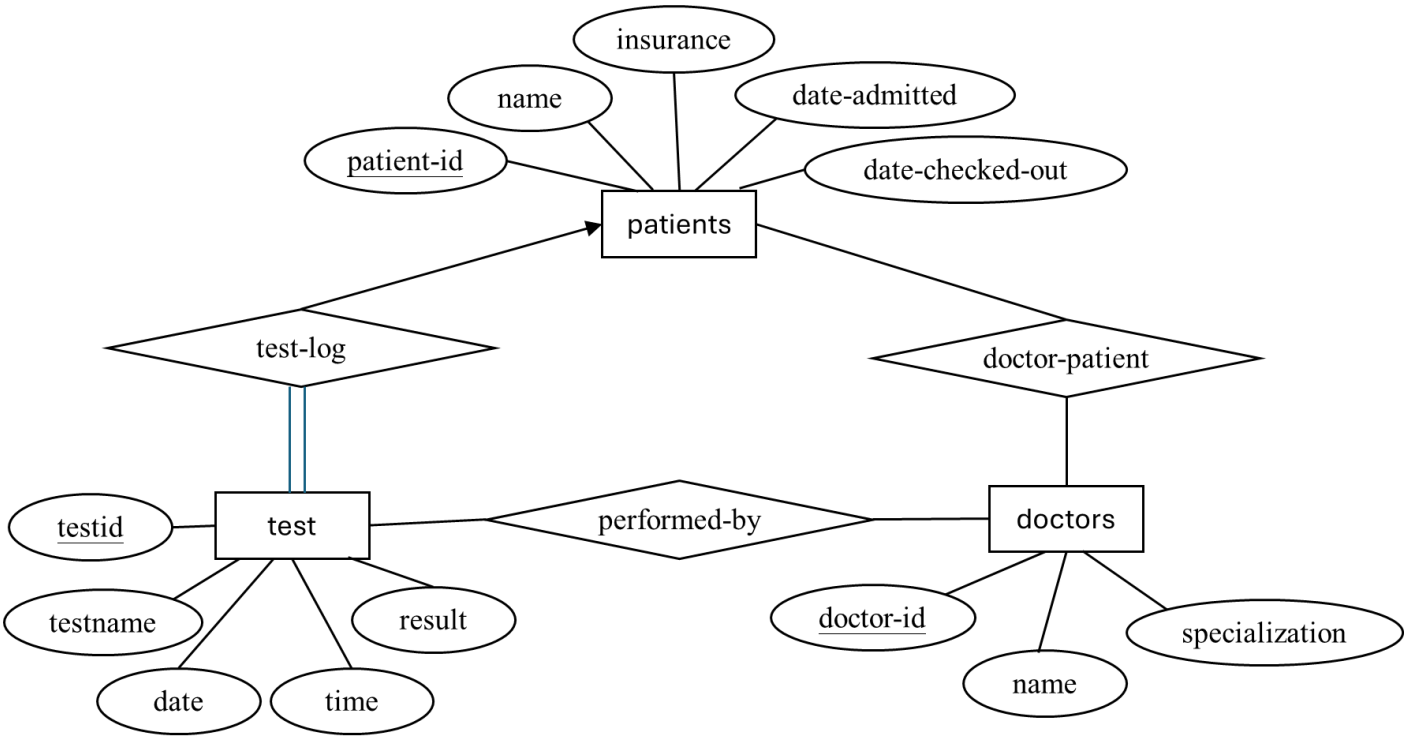
Question 2 [30 marks]

A database schema of a hospital contains the following relation schemas.

- patients (patient-id, name, insurance, date-admitted, date-checked-out)
- doctors (doctor-id, name, specialization)
- test (testid, testname, date, time, result)
- doctor-patient (patient-id, doctor-id)
- test-log (testid, patient-id)
- performed-by (testid, doctor-id)

Construct an E-R diagram for a hospital with a set of patients and a set of medical doctors. Each patient is associated with a log of the various tests and examinations conducted.

Solution



Question 3 **[30 marks]**

You are given a relational schema R= (A, B, C, D, E, F, G, H) and the following set $\mathcal{F}$ of functional dependencies.

- A → C
- B → G H
- C → A
- A → D
- B → E F
- H → B
- G → B
- B → F

3(a) **(15 marks)**

Find a candidate key for the relational schema *R*. **Show** that it is a candidate key.

Solution

To test if AB is a candidate key, we need to test $(AB)^+$ contains ABCDEFGH and A^+ or B^+ does not contain ABCDEFGH.

By using A → C, $(AB)^+$ contains ABC;
By using A → D, $(AB)^+$ contains ABCD;
By using B → EF, $(AB)^+$ contains ABCDEF;
By using B → GH, $(AB)^+$ contains ABCDEFGH;
Thus, AB is the super key.

A^+ is CD and not a super/candidate key;
B^+ is EFGH and not a super/candidate key;
Thus, AB is a candidate key.

- Accept the other solutions to reach the conclusions. For example, CB, AH, or AG is also a candidate key.

3(b) **(15 marks)**

Find the canonical cover for $\mathcal{F}$. **Show** your steps.

Solution

A → CD can replace A → D and A → C
Because B → E F includes B → F, which can be removed.
B → EFGH can replace B → EF and B → EF.

Finally, the canonical cover for $\mathcal{F}$ is
{A → CD, B → EFGH, C → A, H → B, G → B}

- Accept the appropriate approach to reach the conclusion.