COMP 23111

Two hours

SECTION A AND SECTION B ARE COMPULSORY

UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE

Fundamentals of Databases

Date: Tuesday 17th January 2017

Time: 09:45 - 11.45

The Paper is in THREE Sections

You must answer Section A, worth 10 marks
You must answer Section B, worth 10 marks
You must answer ONE of the two questions in Section C, worth 20 marks.

Use a SEPARATE answerbook for each QUESTION.

This is a CLOSED book examination

The use of electronic calculators is NOT permitted

[PTO]

Section A

This section contains one question, worth 10 marks. You MUST answer it using a SEPARATE answer book.

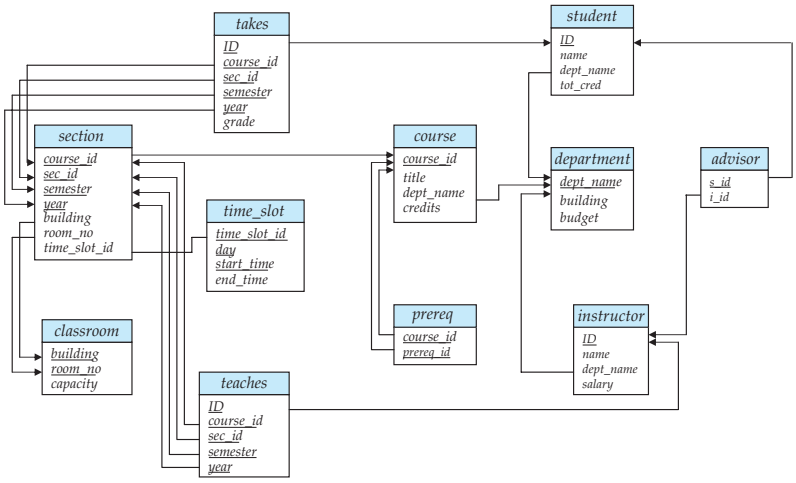
1. a) Consider the entity-relationship (ER) approach to conceptual modelling as studied in this course unit.

   i) Briefly contrast the notions of key, strong key and weak key in ER modelling. (1 mark)

   ii) Briefly explain the (conceptual) modelling consequences of a weak key in an ER model (i.e., what necessarily follows, in an ER model, from the realization that an attribute is a weak key). (2 marks)

   iii) Consider an ER model with the following three entity types: `Book` (with attributes `ID` and `name`), `Author` (with attributes `ID` and `name`), and `Chapter` (with attributes `Number` and `Length`). Using your normal intuition, state which attribute is a weak key. Briefly explain your answer and assumptions. (1 mark)

   b) Consider the following University database schema represented as graph where nodes denote relations (relations names in the shaded area, columns names listed below, primary key components underlined) and a directed edge from attribute f of relation S to an attribute p of relation R denotes that f is a foreign key in S referencing p in R.



   i) Code a solution to the following problem as a relational-algebraic expression against the University database schema: *Find the course ID of courses that use a classroom with a capacity greater than 100.* (2 marks)

   ii) Code a solution to the following problem as an SQL statement against the University database schema: *Retrieve the largest number of sections that an instructor teaches.* Note that the required solution must effectively return a number (i.e., a single-column, single-row table). (4 marks)

Section B

This section contains one question, worth 10 marks. You MUST answer it using a SEPARATE answer book.

2. a) Consider the following relation and its extent. Which of dependencies (i) to (v) (below) may hold in the relation? If the dependency cannot hold, explain why by specifying the tuples that cause the violation.

TUPLENUM	A	B	C
1	10	b1	c1
2	10	b2	c2
3	11	b4	c1
4	12	b3	c4
5	13	b1	c1
6	14	b3	c4

i. A $\rightarrow$ B
ii. B $\rightarrow$ C
iii. C $\rightarrow$ B
iv. B $\rightarrow$ A
v. C $\rightarrow$ A

(5 marks)

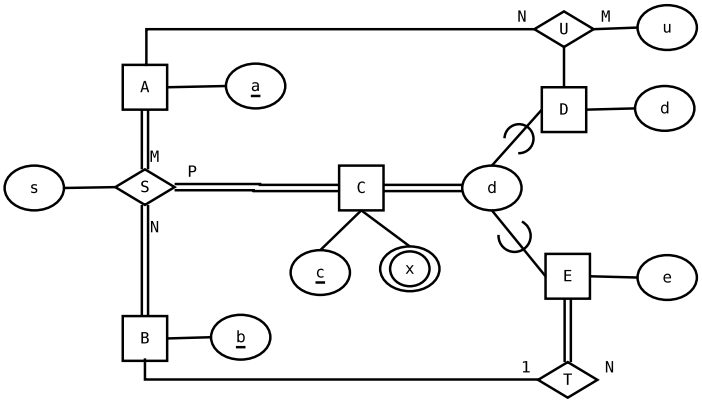
b) Does the above relation have a potential candidate key? If it does, what is it? If it does not, why not? (2 marks)

c) State which are the three most common problems caused by concurrency in database management systems and explain, in each case, what happens when the problem manifests itself and no concurrency control protocol is in use. (3 marks)

Section C

This section contains two questions, each worth 20 marks. You MUST choose ONE of them and answer it using a SEPARATE answer book.

3. a) i) Map the following ER conceptual model into a relational logical model. In your answer, present the logical model as a list of relation schemas with primary key components underlined. You do not need to indicate references from foreign keys because every attribute name in the conceptual model happens to be unique, so no ambiguity arises. (9 marks)



b) i) Translate the SQL statement below into a relational-algebraic operator tree: (4 marks)

```
select count(distinct S.x)
from   R, S
where  R.i = S.j and R.y > 2 or S.y < 3;
```

ii) Translate the assignment sequence below into an SQL query. (4 marks)

$$
\begin{aligned}
I_1 &\leftarrow P \bowtie Q \\
I_2 &\leftarrow I_1 \bowtie R \\
I_3 &\leftarrow \sigma_{a>b}(I_2) \\
I_4 &\leftarrow \pi_{a,b,c}(I_3) \\
R &\leftarrow \delta(I_4)
\end{aligned}
$$

c) i) Briefly draw a contrast between speed-up and scale-up in the context of parallel DBMS architectures. (1 mark)

ii) Your technical manager is concerned that the workload of the database is consistently taking longer to process than what is required (currently, the system is taking 8.25 hours to process the workload). She has decided on applying a speed-up strategy. She has asked you to compute how much larger the new system needs to be to reduce the processing time to 7.5 hours. Compute the increase factor. (1 mark)

iii) Assume that the actual measured speed-up was slightly smaller than predicted. Your manager has asked you what may be the cause of that. Briefly answer her question. (1 mark)

4. a) Consider a database that stores results of medical tests performed on patients of a hospital, where there is a constraint on the class of tests that is available to a patient, depending on the patient's age group. The data describes information about the laboratory in which each test was performed, including a score that describes how well regarded the laboratory is in the healthcare community, as well as information about patients and any tests each has been submitted to.

   Does the table below represent a good design? Illustrate your answer with an example of a table update problem that might result from the current design of the table. (2 marks)

```
CREATE TABLE patient_results (
             lab_id SMALLINT NOT NULL,
             lab_name VARCHAR(15),
             lab_address VARCHAR(45),
             lab_reputation_score DECIMAL(2,1),
             lab_test_id SMALLINT NOT NULL,
             lab_test_name VARCHAR(15),
             lab_test_class SMALLINT,
             lab_test_age_group SMALLINT,
             lab_test_year_month_day DATE NOT NULL,
             lab_test_time_day TIME NOT NULL,
             lab_test_result VARCHAR(145),
             patient_id SMALLINT NOT NULL,
             patient_name VARCHAR(25),
             patient_address VARCHAR(45),
             patient_age_group SMALLINT,
             patient_age SMALLINT,
             patient_blood_type VARCHAR(4),
PRIMARY KEY (lab_id, patient_id, lab_test_id)
);
```

   b) Explain that the above schema is not in 3NF (and hence not in BCNF) arguing from the respective definitions and then normalize it step-by-step to 3NF and, if possible, to BCNF. (6 marks)

c) Consider the following `CREATE TABLE` statements. Note the table named `number_of_bookings`, which contains aggregates (i.e., the number of bookings for a given flight) that are derived from the `booking` table. To keep the `number_of_bookings` table up to date, the `booking` table must be monitored, so that these changes can be reflected on the `number_of_bookings` table. Write one trigger that respond to insert events on the `booking` table and propagate changes to `number_of_bookings` accordingly. (7 marks)

```
CREATE TABLE booking (
   passenger_id VARCHAR(10)
      REFERENCES passenger(passenger_id) ON DELETE CASCADE,
   flight_number VARCHAR(10)
      REFERENCES flight(flight_number),
   year_month_day_time  DATE,
   PRIMARY KEY (passenger_id, flight_number, year_month_day_time));

CREATE TABLE number_of_bookings (
   flight_number VARCHAR(10)
      REFERENCES flight(flight_number),
   year_month_day_time  DATE,
   num NUMBER,
   PRIMARY KEY (flight_number, year_month_day_time));
```

d) Consider transactions `T1`, `T2` and `T3` and schedules `S1` and `S2`, describing the three transactions running concurrently, as shown below, where `r1(X)`, e.g., denotes one read of object `X` from disk into memory, and `w3(Z)` denotes one write of object `Z` back to disk from memory.

```
T1: r1(X); r1(Z); w1(X); w1(Z)
T2: r2(Y); r2(Z); w2(Z)
T3: r3(Y); r3(X); w3(Y)

S1: r1(X); r3(Y); r3(X); r2(Y); r2(Z);
    w3(Y); w2(Z); r1(Z); w1(X); w1(Z)
S2: r1(X); r3(Y); r2(Y); r3(X); r1(Z);
    r2(Z); w3(Y); w1(X); w2(Z); w1(Z)
```

Are both schedules conflict-serializable? Why or why not? If not both, is any one of schedules conflict-serializable? Why or why not? (5 marks)

END OF EXAMINATION