Two hours

SECTION A AND SECTION B ARE COMPULSORY

UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE

Fundamentals of Databases

Date: Friday 24th January 2014

Time: 14:00 - 16:00

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 SECTION.

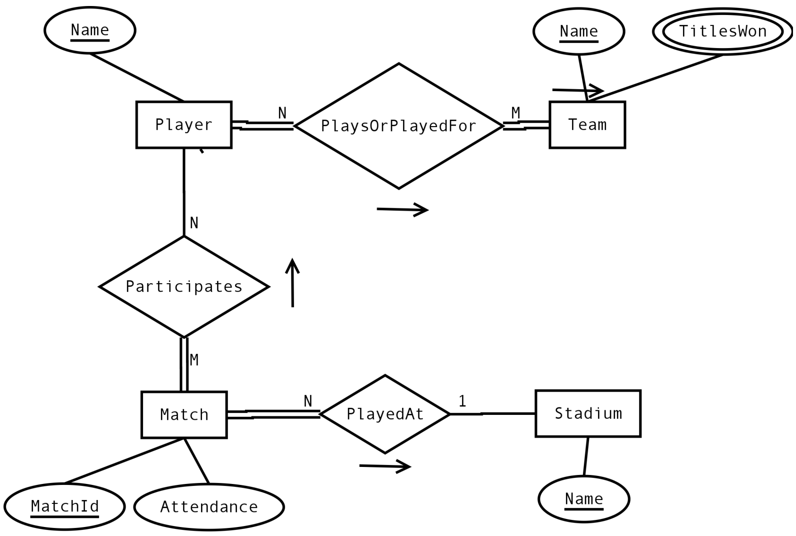
This is a CLOSED book examination

The use of electronic calculators is NOT permitted

Section A

This section contains one question.
You MUST answer it using a SEPARATE answerbook.

1. a) Consider the incomplete ER diagram below. It partially captures the data requirements of a company that provides information to sport media outlets. As the diagram shows, there is an interest in teams and their players, as well as stadia and the matches played.



   i) Suggest one additional attribute for the Player entity type and one for the Participates relationship type that you consider would be of interest. Capture this extension in your revised diagram. (1 mark)

   ii) Suggest one additional attribute for the Stadium entity type that you consider would be of interest and that can be related by some integrity constraint with the Attendance attribute of the Match entity type. Capture this extension in your revised diagram and state what the constraint might be. (2 marks)

   iii) The company is interested in the information as to the period during which a player has played for a team. Capture this requirement with two additional attributes in your revised diagram. (1 mark)

   iv) The company is interested in modelling the fact that each team has a stadium as its home ground. Capture this requirement in your revised diagram. (3 marks)

[Next page →]

b) With respect to the main phases of query processing, give the names of the database management system (DBMS) components that

   i) transform a query into a relational-algebra expression
   ii) transform a relational-algebra expression into an execution plan

(1 mark)

c) Consider a company that is expecting a significant increase in its sales and is worried that their backend processor may not cope with the increase in workload. The goal is to ensure that the company can process the new volumes in the same amount of time it processes the current workload.

   i) When consulted for her technical view, one of your colleagues commented that *"Since there is no need to improve on current processing time, scaling up is the best option."*. Briefly explain whether you agree with your colleague and why. (1 mark)
   ii) Assume that the company uses two servers at the moment. Assume that the new workload is expected to be three times the size of the current one. Making the usual simplification assumptions, work out how many servers will be needed if the new volumes are to be processed in the same amount of time that the current workload is processed. (1 mark)

Section B

This section contains one question.
You MUST answer it using a SEPARATE answerbook.

2. a) Is the following statement true? Why or why not?

   Main memory is the fastest and most expensive storage medium, and often quite large in some server systems, to the extent of being able to store an entire database.

   (2 marks)

   b) Is the following statement true? Why or why not?

   The heap file organization is ideal for supporting high performance applications that involve a large number of updates, including changes to attributes and deletions of tuples.

   (2 marks)

   c) Is the following statement true? Why or why not?

   When designing a trigger, it is mandatory to specify an event that causes the trigger to be checked, as well as a condition that must be satisfied for trigger execution to proceed, and two or more actions to be taken when the trigger executes.

   (2 marks)

   d) Is the following statement true? Why or why not?

   Indexes speed up the response to user queries.

   (2 marks)

   e) Is the following statement true? Why or why not?

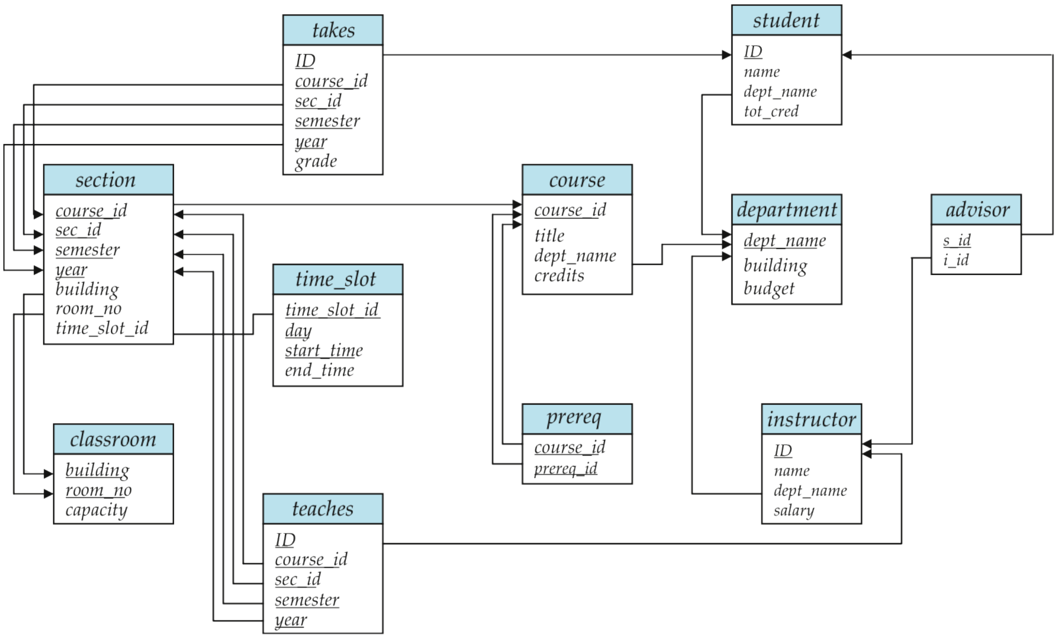
   The recovery mechanism in database management systems depends on information about all transactions that are running on the system at a particular time.

   (2 marks)

Section C

Answer one of the two questions in this section.
Use a SEPARATE answerbook for the question you choose.

3. a) Consider the UNIVERSITY relational database schema represented in diagrammatic form below. Use it to answer the questions that follow.



i) Write an SQL query that retrieves the names of departments that offer a course. (2 marks)

ii) Briefly explain why the following SQL query is invalid:

```
select name, min(salary)
  from instructor;
```

(1 mark)

iii) Write an SQL query that retrieves the name and budget of departments in order, from the largest to the smallest budget. (2 marks)

iv) Recall the notion of direct translation from SQL into relational algebra (e.g., no introduction of joins yet). Write the result of the direct translation of the following SQL query into the corresponding relational-algebraic expression:

```
select s.name, c.course_id, t.grade
  from student s, course c, takes t
 where s.ID = t.ID and c.course_id = t.course_id
   and t.year = 2010
```

(3 marks)

v) Write an SQL query that computes the same result as the following relational-algebraic expression:

$$\pi_{instructor.name, student.name}(instructor \bowtie_{ID} (teaches \bowtie_{course\_id} (takes \bowtie_{ID} student)))$$

(3 marks)

vi) Consider the `advisor` relation in the above diagram. Now, imagine the ER diagram from which this relational schema must have been derived. Briefly state which ER construct (i.e., attribute type or entity type or relationship type) must have given rise to the `advisor` relation. Be as precise and comprehensive as possible in your answer (e.g, if it is a relationship type, give the structural constraints that must have been written for it on the ER diagram).

(3 marks)

b) Given entity types A, B, C, D, and E, draw an EER diagram that captures the following specialization information (expressed, for the purpose of concision, in set-theoretic notation rather than in English):

$$
\begin{array}{llll}
B & \subset & A & \\
C & \subset & A & \\
D & \subset & B & \\
E & \subset & B & \\
B & \cap & C & = \emptyset \\
D & \cap & E & \neq \emptyset \\
B & \cup & C & = A \\
D & \cup & E & \subset B
\end{array}
$$

(4 marks)

c) The following two items are about the views expressed by E. Bertino and R. Sandhu in their paper *Database Security - Concepts, Approaches, and Challenges*, IEEE Transactions on Dependable and Secure Computing (2):1, pp. 2-19, 2005.

i) Security breaches are typically classified into three categories. State two of them. (1 mark)

ii) There are three requirements that a complete solution to data security must meet. State two of them. (1 mark)

4. a) Consider Tables 1-3 in the figure below. Table 1 describes a serial schedule for two transactions, $T1$ (in the left column) and $T2$ (in the right column). The instructions for each transaction appear in chronological order from top to bottom. Both $T1$ and $T2$ are bank transactions. In the table, A and B represent two different bank accounts. Assume that the current values for both accounts A and B are greater than or equal to £1000.

T1	T2
read(A)	
A := A - 50	
write(A)	
read(B)	
B := B + 50	
write(B)	
commit	
	read(A)
	temp := A * 0.1
	A := A – temp
	write(A)
	read(B)
	B := B + temp
	write(B)
	commit

Table 1.

T1	T2
read(A)	
A := A - 50	
	read(A)
	temp := A * 0.1
	A := A - temp
	write(A)
	read(B)
write(A)	
read(B)	
B := B + 50	
write(B)	
commit	
	B := B + temp
	write(B)
	commit

Table 2.

T1	T2
read(A)	
A := A - 50	
write A	
	read(A)
	temp := A * 0.1
	A := A - temp
	write(A)
read(B)	
B := B + 50	
write(B)	
commit	
	read(B)
	B := B + temp
	write(B)
	commit

Table 3.

i) Consider the alternative schedule for transactions $T1$ and $T2$ over bank accounts A and B shown in Table 2. Does it result in a consistent database state? Briefly explain why or why not. (2 marks)

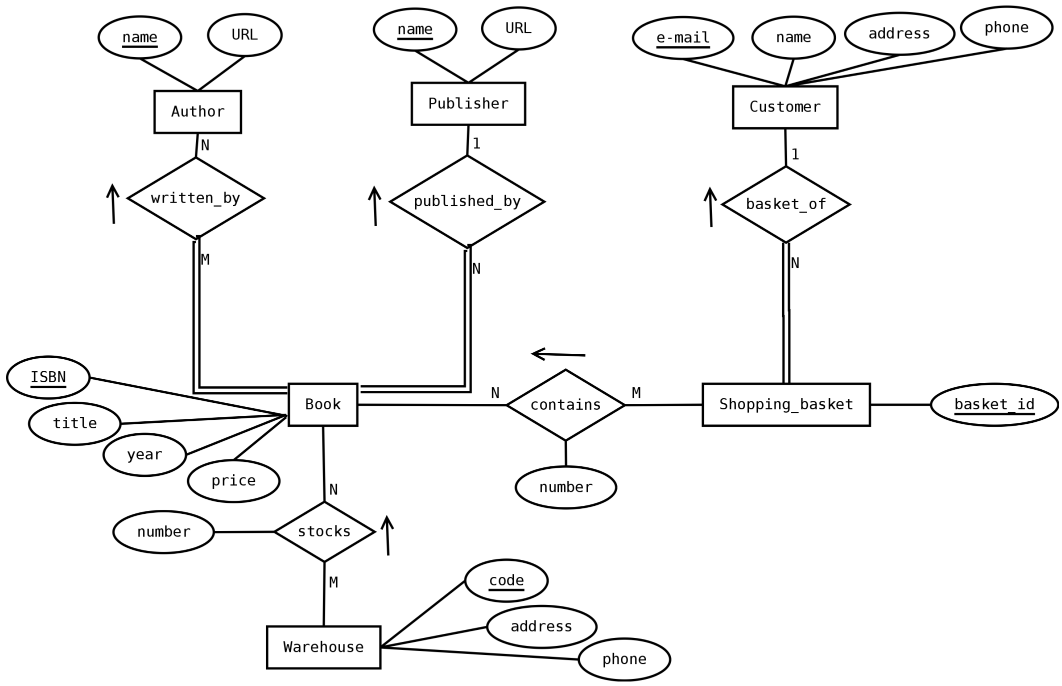
ii) Consider now the alternative schedule for transactions $T1$ and $T2$ over bank accounts A and B, shown in Table 3. Does it result in a consistent database state? Briefly explain why or why not. (3 marks)

iii) Define *serializability* in the context of transaction schedules. (1 mark)

iv) Briefly explain how the two-phase locking protocol implements serializability. Illustrate your explanation with an example showing an inconsistent schedule being made consistent by the implementation of the two-phase locking protocol. If any of the schedules in Tables 2 and 3 is inconsistent, use it in your example. (3 marks)

v) What is the main limitation of the two-phase locking protocol? Illustrate your answer with an example showing two transactions using the two-phase locking protocol running concurrently. (3 marks)

b) Map the entity-relationship (ER) model below into a relational schema, specifying primary and foreign keys. (4 marks)



c) Briefly explain the impact of the quality of the ER model for a relational database on the design of the relational schema for the database, including in your discussion a comparison between normalization applied over a "rough" relational schema and the complexity (or otherwise) of devising a good ER model from scratch. (4 marks)