The University of Nottingham

SCHOOL OF COMPUTER SCIENCE

A LEVEL 1 MODULE, SPRING SEMESTER 2008-2009

DATABASE SYSTEMS

Time allowed TWO hours

Candidates may complete the front cover of their answer book and sign their desk card but must NOT write anything else until the start of the examination period is announced

Answer THREE out of FIVE questions

No calculators are permitted in this examination.

Dictionaries are not allowed with one exception. Those whose first language is not English may use a standard translation dictionary to translate between that language and English provided that neither language is the subject of this examination. Subject specific translation dictionaries are not permitted.

No electronic devices capable of storing and retrieving text, including electronic dictionaries, may be used.

DO NOT turn examination paper over until instructed to do so

1. Consider the following tables:

Book			
ISBN	Author	Title	Year
949	A.Green	Memoirs	2001
287	B.Black	Biology	2002

Publisher	
PublisherName	ISBN
Thompson	949
Elsevier	287

(a) Give a table returned by (1 marks)

```
SELECT *
FROM Book, Publisher
```

(b) Give a table returned by (1 marks)

```
SELECT *
FROM Book CROSS JOIN Publisher
```

(c) Give a table returned by (2 marks)

```
SELECT *
FROM Book NATURAL JOIN Publisher
```

(d) Give a table returned by (2 marks)

```
SELECT *
FROM Book INNER JOIN Publisher USING(ISBN)
```

(e) Rewrite the previous `INNER JOIN` query with `ON` $<$ `condition` $>$ instead of `USING(ISBN)`. (2 marks)

(f) What are `FULL OUTER JOIN`, `LEFT OUTER JOIN` and `RIGHT OUTER JOIN`? Give an example of the three kinds of outer joins on two tables both of which contain dangles. (3 marks)

(g) What does it mean for two relations to be union-compatible? (2 marks)

(h) List relational algebra operators and explain what they do. (6 marks)

(i) Translate the following relational algebra expression into SQL: (3 marks)

$$\pi_{\texttt{Title}}\, \sigma_{\texttt{Year}>2001}\, (\texttt{Book} \times \texttt{Publisher})$$

(j) Translate the following SQL query into relational algebra: (3 marks)

```
SELECT Author, Title
FROM Book, Publisher
WHERE Book.ISBN = Publisher.ISBN AND
      PublisherName = 'Elsevier'
```

2. You are asked to design a database for the following scenario. A research laboratory is running several drug trials on healthy volunteers to check whether drugs have side effects. Each drug has a unique name. Each trial involves exactly one drug and several volunteers (who take the drug and report if they had any side effects). For each volunteer in each trial it needs to be recorded whether the volunteer had any side effects, and if yes, what those side effects were (there could be several side effects experienced by the same person, for example headache, dry mouth, and fever). It is important that side-effects are described using some standard terminology, so that the laboratory can report what proportion of volunteers had the same side effect. For example, the researchers may tell you that headache should always be recorded as 'headache' and not sometimes as 'pain in the head' and sometimes as 'sore head'. There is however no fixed pre-defined set of possible side effects, because new effects can always be discovered (for example the drug may turn people a bright green colour). For simplicity, assume that each volunteer takes part in at most one drug trial. Data stored about volunteers is their National Insurance number, name, age, gender, address and telephone number.

   (a) Draw an entity-relationship diagram for the drug trial scenario. (10 marks)

   (b) Write SQL statements to create the tables (don't forget to specify primary and foreign keys). (10 marks)

   (c) Explain what is referential integrity, and explain it on the example of foreign keys in your answer to question (b). (5 marks)

3. This question refers to the following tables: `Children`, `Playgroups`, `Activities`. The `Children` tables contains data about children (names, ages and addresses of parents) - we assume for simplicity that names are unique, `Playgroups` says which child is in which playgroup and `Activities` says what children in the playgroup did on a certain date (for example, went to a zoo).

Children		
Name	Age	Address

Playgroups	
PlaygroupID	Name

Activities		
PlaygroupID	ADate	Description

Write SQL queries to do the following:

(a) Find a list of names of all children. (1 marks)

(b) Find a list of names of all children aged 4. (2 marks)

(c) Return a list of names and addresses of all children in the playgroup with `PlaygroupID` equal to 1. (3 marks)

(d) Find a list of names and ages of children in a playgroup which went to the zoo on the 21st of February 2009 (check for `Activities.Description` value `Zoo`). (5 marks)

(e) Return a list of playgroup IDs and average age of children for each playgroup. (5 marks)

(f) Find names and addresses of all children who are in the same playgroup as a child called 'Amy Jones'. (5 marks)

(g) Add an extra column to the `Activities` table which is called `Supervisor` and has the type of values `VARCHAR(20)`. (2 marks)

(h) Replace description `Zoo` in the `Activities` table everywhere with `Visit to the Zoo`. (2 marks)

4. (a) Define functional dependency. (2 marks)

   (b) Define 2 NF. (2 marks)

   (c) Define 3 NF. (2 marks)

   (d) Define BCNF. (2 marks)

   (e) What are insertion, deletion and update anomalies? (3 marks)

   (f) Consider a relation Listing with attributes Cinema, Film, Day, Time, Certificate. Sample tuples:

   ```
   Savoy, 'Slumdog Millionaire', Wed, 18:00, 15
   Savoy, 'Slumdog Millionaire', Wed, 20:00, 15
   Cineworld, 'Slumdog Millionaire', Wed, 20:30, 15
   Cineworld, 'Vicky Christina Barcelona', Wed, 20:30, 12A
   ```

   Each film is assigned a certificate by the British Board of Film Classification; certificate 15 means that nobody younger than 15 can see this film in a cinema. The same cinema can show a film on multiple times during a day, and may show different films at the same time (on different screens).

   Does this relation violate the second normal form requirements? Explain your answer. (5 marks)

   (g) Give an example of update anomalies in the Listing relation from part (f). (4 marks)

   (h) Decompose the Listing relation from part (f) to BCNF, and explain why the resulting relations are in BCNF. (5 marks)

5. (a) What is a NULL? Explain A-marks and I-marks. (3 marks)

   (b) How does SQL evaluate conditions for tuples containing NULLS? Explain the general approach. For a tuple {`Name` : `John`, `Age` : `NULL`}, will the following `WHERE` conditions evaluate to true? (6 marks)

```
1. WHERE Age > 18
2. WHERE Age > 18 OR Name = 'John'
3. WHERE Age > 18 OR NOT (Age > 18)
```

   (c) How are arithmetic $(+, -, *)$ operations applied to NULLs evaluated in SQL? (2 marks)

   (d) How are aggregate operations (SUM, AVG, Count) on NULLs evaluated in SQL? (2 marks)

   (e) In response to what problem was the timestamping protocol developed? (2 marks)

   (f) Describe the timestamping protocol. (5 marks)

   (g) Trace the timestamping protocol for the following two transactions `T1` and `T2`, assuming that the statements are executed in a strictly alternating way (first statement of `T1` followed by the first statement of `T2` followed by the second statement of `T1`, and so on) and there are no other transactions. At each step, indicate what the time stamps of `T1` and `T2` are, and what the read and write timestamps of resources `X`, `Y` are. Assume that before `T1` and `T2` are executed the timestamps of resources are 0. Trace until both transactions can commit. (5 marks)

Transaction 1	Transaction 2
Write(X)	Read(Y)
Write(Y)	Write(X)