The University Of New South Wales

Final Exam

June 1999

COMP9311/3311

Database Systems

Time allowed: **2 hours**
Total number of questions: **5**
Total number of marks: **100**

Textbooks, lecture notes, etc. are **not** permitted.
Calculators may **not** be used.
All questions are worth equal marks.
Answer **all** questions.
Answers must be written in ink.
You can answer the questions in any order.
Answer these questions in the script book provided.
Start each question on a new page.
If you use more than one script book,
fill in your details on the front of *each* book.
You may **not** take this question paper out of the exam.

Family Name:

Other Names:

Signature:

Student Number:

Question 1

Which of the following statements are correct? (Answer true or false)

(1) ODL can model a "n-way" relationship where $n \geq 3$.   *(2 marks)*

(2) In ER, a relationship type must have a key.   *(2 marks)*

(3) In the relational data model, a primary key is also a candidate key.   *(2 marks)*

(4) An equi-join is a theta join.   *(2 marks)*

(5) In SQL, SUBSTR ('XUEMIN', 1, 3) returns the string 'Xue'.   *(2 marks)*

(6) In the standard relational model, every relation is in 1NF.   *(2 marks)*

(7) A relation schema in 2NF is also in 3NF.   *(2 marks)*

(8) The *isolation* property means that a transaction is either performed in its entirety or not performed at all.   *(2 marks)*

(9) The time-stamp ordering method does not produce deadlocks.   *(2 marks)*

(10) A locking mechanism does not always guarantee the serializability of a transaction processing schedule.   *(2 marks)*

Question 2

a) Suppose that we wish to keep a genealogy. We shall have one class, *Person*. The information we wish to record about persons includes their names, birth dates, and the following relationships:

- The Mother relationship specifies the mother of a person.
- The Father relationship specifies the father of a person.
- The Parent relationship specifies the parents of a person.

Give an ODL design for the *Person* class; and be sure to indicate the inverse of the relationships. *(6 marks)*

b) A database is being constructed to keep track of the teams and games of a basketball league. Each Team is identified by its name, and is described by its players and the head coach. Each game is identified by its number, and is described by the location and the time. We assume that, for each game, the two participating teams are specified as a home team and a foreign team. With respect to each game, both the home team and foreign team keep the records of their total scores, the number of rebounds, and the number of steals.

    i) Draw an ER diagram to represent your schema for this database.   *(8 marks)*

    ii) Derive a relational database schema from your ER-diagram.   *(6 marks)*

Question 3

a) In a relational database, there are three main integrity constructs:

- Key constraints,
- Entity constraints,
- Referential constraints.

Indicate which of these constraints should be checked when an update modifies an attribute that is part of

    i) a primary key,   *(2 marks)*

    ii) a foreign key (but not part of a primary key),   *(2 marks)*

    iii) neither.   *(2 marks)*

b) Let $F = \{A \to B, AB \to D, CE \to G, C \to H\}$ where $\{A, B, C, D, E, G, H\}$ is a set of attributes. Determine whether or not $F \models AC \to G$. (Justify your answer.)   *(5 marks)*

c) Let $F = \{A \to C, BC \to D\}$ be a set of FDs. Determine the highest normal form of the table (A, B, C, D) with respect to F. (Justify your answer.)   *(5 marks)*

d) Let $F = \{Y \to X, Z \to XYW\}$. Find a minimum cover for F.   *(4 marks)*

Question 4

a) Explain the difference between each of the following:

    i) Primary vs. secondary indexes.   *(3 marks)*

    ii) Dense vs. sparse indexes.   *(3 marks)*

    iii) Clustered vs. unclustered indexes.   *(3 marks)*

b) Is it a good idea to create a sparse index that is not clustered? Why?   *(5 marks)*

c) Consider a relation stored as a randomly ordered file, for which the only index is an unclustered index on a field called *age*. If you want to retrieve all records with $age < 60$, is using the index always the best alternative? Explain.   *(6 marks)*

Question 5

a) Use an example to illustrate the incorrect summary problem in transaction processing.
   (4 marks)

b) Consider the following two transactions:

T1	T2
READ (A)	READ (B)
A $\longleftarrow$ 10 A + 4	B $\longleftarrow$ 2*B + 3
WRITE (A)	WRITE (B)
READ (B)	READ (A)
B $\longleftarrow$ 3*B	A $\longleftarrow$ 100 - A
WRITE (B)	WRITE (A)

   i) Use the two-phase locking protocol to construct a serializable schedule for processing the two transactions. *(4 marks)*

   ii) Is there a non-serial but serializable schedule for T_1 and T_2? Why? *(4 marks)*

   iii) Can a schedule for $T1$ and $T2$ result in deadlock? If so, please show your schedule. If not, why not? *(4 marks)*

c) Determine whether the following schedule is serializable:

$$S \quad = \quad [R_2(Y), W_2(Y), R_3(Y), R_1(X), W_1(X), W_1(Z)$$
$$W_3(Y), R_2(X), R_1(Y), W_1(Y), W_2(X), R_3(W), W_3(W)]$$

   where $R_i(Z)/W_i(Z)$ indicates a read/write by transaction i on data item Z. Describe your method. *(4 marks)*