UNIVERSITY OF LONDON

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2004

MSc in Computing Science
BEng Honours Degree in Information Systems Engineering Part III
MEng Honours Degree in Information Systems Engineering Part III
BSc Honours Degree in Mathematics and Computer Science Part III
MSci Honours Degree in Mathematics and Computer Science Part III
for Internal Students of the Imperial College of Science, Technology and Medicine

This paper is also taken for the relevant examinations for the Associateship of the City and Guilds of London Institute
This paper is also taken for the relevant examinations for the Associateship of the Royal College of Science

PAPER M311=I3.2

DATABASES

Monday 10 May 2004, 10:00
Duration: 120 minutes

Answer THREE questions

Paper contains 4 questions
Calculators not required

1 a Consider relation scheme R(A,B,C,D,E,F,G) with the following set S of functional dependencies:

$AB \rightarrow CD$	$A \rightarrow G$	$A \rightarrow AC$
$AF \rightarrow E$	$GC \rightarrow F$	
$B \rightarrow D$	$F \rightarrow B$	

i) An attempt to find a canonical cover for S has produced the following set S1 of functional dependencies:

$AB \rightarrow C$	$A \rightarrow GC$
$AF \rightarrow E$	$GC \rightarrow F$
$B \rightarrow D$	$F \rightarrow B$

Is S1 a correct canonical cover for S? If yes, prove it. If no, simplify S1 further to get a correct canonical cover.

ii) Is R in 2NF? Justify your answer.

iii) Suppose R is decomposed into two relations R1 and R2:
R1(A,B,C,E,F,G) R2(B,D,F)
Is this decomposition lossless? Briefly justify your answer.

b

i) Define the Boyce-Codd (BCNF) normal form.

ii) Suppose a relation scheme R is in 3NF and its primary key consists of one attribute only. Is R also in BCNF? Justify your answer in detail.

iii) Consider the relation
Car_Sales(Car#, Date_Sold, Salesman#, Commission%, Discount_amount)
With the following functional dependencies:

Car# $\rightarrow$ Date_Sold
Salesman# $\rightarrow$ Commission%
Date_Sold $\rightarrow$ Discount_amount

Commission% denotes the percentage of the sale price that is paid to the salesman participating in the sale.

State all the candidate keys of *Car_Sales*. What is the highest normal form that *Car_Sales* is in?
Briefly explain one of the problems that the design of relation *Car_Sales* will cause in practice.

Parts a, b each carry 50% of the marks.

2 This question is concerned with entity-relationship (ER) modeling and the relational model.

The Department of Computing wishes to maintain the following information about its structure and activities.

DOC Info:

The department contains several research groups. Each research group is identified by a name, has a group head and has a main research area associated with it.

Each group head is identified by the combination of his/her surname and first name, has a rank and heads one group only. The rank is either professor or senior lecturer.

The department has academics who teach courses and who can be members of research groups. Each academic has a rank and teaches at least one course. Each course is taught by one or more academics. However, if the rank of an academic is professor, then no one else teaches the courses that he/she teaches. Each course is identified by a code and has a title.

An academic is a member of at most one research group, and a research group has one or more academics as members.

All academics and research group heads are departmental staff. Each member of the departmental staff is identified by the combination of his/her surname and first name, and has a salary.

Research assistants (RAs) are also departmental staff. They have contracts of given lengths. They assist the academics in their research and teaching. Each RA assists exactly one academic in his/her research. An academic may have any number (0 or more) RAs assisting in his/her research.

One RA may assist none, one or more academics on the teaching of one course. Each academic is assisted on each course by at most one RA, but one RA may assist an academic on any number of courses.

a Represent the data above in an ER diagram. Be careful to represent all the existence constraints and the relationship cardinalities. Do not include the 'Department of Computing' as an entity, but include all the entities and relationships that are given above as relevant to the department.

Identify any information in DOC info that cannot be represented in an ER diagram.

b Translate the ER model into the relational model, giving all the relation schema and the primary and foreign keys. For each foreign key indicate which relation it refers to and whether or not it can accept NULL values. You do not need to give any further foreign key rules.

Parts a, b, each, carry 50% of the marks.

3a In this question, in each sub-part (i) to (v) below, you are asked to consider two queries $\mathbf{Q_1}$ and $\mathbf{Q_2}$. The queries refer to three relational tables R(a,b), S(b,c,d) and T(b,c,d). The tables may contain duplicate tuples and NULLs. In each case, (i)-(v), determine which of the following is true for all database instances, clearly justifying your answer:

- **Q_1 and Q_2 are the same** i.e. Q_1 and Q_2 produce the same tuples, and the same number of copies of a given tuple. The order of the tuples is not to be considered.
- **Q_1 is contained in Q_2** i.e. Any tuple that is produced by Q_1 is also produced by Q_2. Furthermore, if a tuple appears more than once in the results of Q_1, then it appears at least as many times in Q_2.
- **Q_2 is contained in Q_1.**
- **Q_1 and Q_2 are different** i.e. Q_1 and Q_2 are not the same, and neither one is contained in the other.

i) **Q_1**:
```
(SELECT * FROM T)
INTERSECT ALL
(SELECT * FROM S)
```
Q_2:
```
(SELECT * FROM T)
NATURAL JOIN
(SELECT * FROM S)
```

ii) **Q_1**:
```
SELECT R.a FROM R
WHERE EXISTS
   (SELECT S.c FROM S
    WHERE S.d = R.b);
```
Q_2:
```
SELECT R.a FROM R
WHERE R.b = ANY
   (SELECT S.d FROM S);
```

iii) **Q_1**:
```
SELECT a FROM R
WHERE a >= ALL
   (SELECT b FROM S);
```
Q_2:
```
SELECT a FROM R
WHERE a >= ANY
   (SELECT b FROM S);
```

iv) **Q_1**:
```
SELECT a, MAX(b)
FROM R
GROUP BY a;
```
Q_2:
```
SELECT a, b FROM R r1
WHERE b >= ALL
     (SELECT b
      FROM R r2
      WHERE r1.a = r2.a);
```

v) **Q_1**:
```
SELECT r1.a, r3.a
FROM R r1, R r2, R r3
WHERE r1.b = r2.b AND
      r2.a = r3.a AND
      r2.b <> r3.b;
```
Q_2:
```
SELECT r1.a, r3.a
FROM R r1, R r2, R r3
WHERE r1.a = r2.a AND
      r2.b = r3.b AND
      r1.b <> r2.b;
```

b Consider the execution histories H_1 and H_2. The symbol $r_i(x)$ stands for a read by transaction T_i to item x and $w_i(x)$ stands for a write by T_i to item x. The symbol C_i stands for the commit of transaction T_i.

H_1: $r_1(a)\ w_1(a)\ r_2(a)\ w_1(b)\ r_2(b)\ w_2(a)\ C_2\ C_1$

H_2: $r_4(x)\ r_6(y)\ w_4(y)\ C_6\ r_5(y)\ w_4(z)\ r_4(y)\ C_4\ r_5(z)\ C_5$

i) Determine if the above histories are conflict-serializable and, if so, find an equivalent serial history.

ii) Determine if the above histories are recoverable.

The two parts a and b carry, respectively, 65%, and 35% of the marks.

4a The following SQL queries concern a table Arc(x,y), which you may think of as representing the fact that there is an arc from node x to node y in a certain directed graph. You should not assume that x, y or even x and y together, is a key for Arc, as there can be multiple (parallel) arcs between nodes. You may assume that there are no NULLs in the table. Consider the queries Q_1 and Q_2 below:

Q_1:
```
CREATE VIEW V AS
    SELECT x, COUNT(y) AS ct
    FROM Arc
    GROUP BY x;

SELECT SUM(ct)
FROM Arc, V
WHERE Arc.x = 0 AND
      Arc.y = V.x;
```

Q_2:
```
SELECT COUNT(a2.y)
FROM Arc a1, Arc a2
WHERE a1.x = 0 AND
      a1.y = a2.x;
```

Call v_1 and v_2 the final results of Q_1 and Q_2 respectively. By explaining what is computed by Q_1 and Q_2, determine which of the following relationships is correct.

(1) $v_1 < v_2$ (2) $v_1 = v_2$ (3) $v_1 > v_2$

b Consider a Table T(a, b) where a and b are of type integer, and where NULLs are allowed. In SQL's 3-valued logic, which allows three values True, False, and Unknown, what are the *possible* values for this expression?
(T.a = 5) OR (T.b <> 0) OR (T.a <>5 AND T.b = 0)

c Consider the following four transactions:

T_1: r[x] r[y] w[x] T_2: r[z] r[x] w[x] w[y]
T_3: r[y] w[y] T_4: r[z]

The four transactions were executed concurrently, resulting in a conflict-serializable history H. Without any further knowledge about H, determine the number of serial histories that are equivalent to H.

d Consider the execution histories H_1, H_2 and H_3 below. The symbol $r_i(x)$ stands for a read by transaction T_i to item x and $w_i(x)$ stands for a write by T_i to item x.

H_1: $r_1(a)\ r_1(b)\ w_1(a)\ r_2(a)\ w_1(b)\ r_2(b)\ w_2(b)$
H_2: $r_4(x)\ r_6(s)\ w_4(z)\ r_5(s)\ w_4(x)\ r_5(z)$
H_3: $r_7(a)\ w_8(a)\ w_9(x)\ r_9(a)\ r_7(b)\ r_7(x)\ w_8(b)$

i) Describe what is meant by basic, strict and rigorous two-phase locking.

ii) Explain why strict 2PL is more desirable than basic 2PL and determine which of the above histories are possible under the strict 2PL protocol. Explain your answer in detail.

The four parts carry respectively, 20%, 15%, 20% and 45% of the marks.