UNIVERSITY OF LONDON
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2003

MSc in Computing Science
BEng Honours Degree in Information Systems Engineering Part III
MEng Honours Degree in Information Systems Engineering 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

PAPER M311=I3.2

DATABASES

Friday 16 May 2003, 14:30
Duration: 120 minutes

Answer THREE questions

Paper contains 4 questions
Calculators not required

Section A (Use a separate answer book for this section)

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

$A \to C$	$CD \to EF$
$B \to D$	$AE \to F$

i) Give all the candidate keys of R.

ii) Explain why R is not in 3NF.

iii) An attempt has been made to normalise R into 3NF relations resulting in relations R1 and R2:
R1(A,B,F) R2(C,D,E,F)

Is relation R2 in 3NF? Explain your answer.
Using appropriate instances of relations R, R1, R2 show in detail why this attempt at normalisation is incorrect.

b

i) Give Armstrong's axioms for functional dependencies in relational databases.

ii) Using Armstrong's axioms show $X \to YZ \models X \to Y$.

iii) Give an irreducible cover for the following set of functional dependencies and justify your answer.

$AB \to CD$	$D \to E$	$AE \to F$
$FG \to H$	$AEG \to H$	$AB \to CE$
$B \to DA$	$B \to F$	

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

2a Consider the following relation schemes

R1(<u>A</u>,B,C) R2(<u>E</u>,F,G) R3(<u>A</u>,<u>E</u>,K)

and the query

R1X.A, R1X.B where ∀R2X ∃R3X (R3X.E=R2X.E and R3X.A=R1X.A)

formulated in relational calculus, where R1X, R2X, R3X range over relations R1, R2, R3.

Give a formulation of the query in relational algebra.

b Consider the following relation schemes dealing with the administration of a library. The primary keys are underlined.

Books(<u>BID</u>, Title, Author)
Users(<u>UID</u>, Name, Postcode)
Loans(<u>BID</u>, UID, Date_Due, Overdue_fine)

Tuples in *Books* and *Users* give details of the library books and the library users, respectively. Different copies of the same book have different BIDs. Tuples in *Loans* give details of the books on loan to users. The attribute *Overdue_fine* is initially 0 (zero) for all loans and gets updated automatically when books become overdue.

Consider the following queries.

Q1) List the titles of all books of which the library has more than one copy

Q2) List the names and postcodes of all the users who have at least one book out on loan and who have non-zero overdue fines on all the books that they have out.

Q3) List the names of all the users who either have no books out, or have at least one book out, but none of the books they have out is by Tolkein.

i) Formalise Q1, Q2, Q3 in relational algebra.

ii) Formalise Q2, Q3 in relational calculus.

Parts a, b(i), b(ii) carry, respectively, 20%, 45%, 35% of the marks.

Section B (use a separate answer book for this section)

3a Distinguish between *conflict equivalent* and *view equivalent* histories.

The following complete history has been suggested for the concurrent running of three transactions T_1, T_2, T_3:

$$H = W_1[x]W_2[x]W_2[y]C_2W_1[y]W_3[x]W_3[y]C_3W_1[z]C_1$$

Determine whether this history is i) conflict serialisable

ii) view serialisable

and, where appropriate, give the equivalent serial history.

b Explain what is meant by the *two-phase locking protocol* (2PL) and distinguish between *Basic* 2PL and *Strict* 2PL. Explain, with the aid of an example, why Strict 2PL is preferable to Basic 2PL.

c Describe the Timestamp Ordering scheme for scheduling concurrent transactions.

Show that if H is a history representing an execution produced by a Timestamp Ordering scheduler, then H is serialisable.

Transactions T4,T5,T6 are to be run concurrently. The following table gives details of the proposed interleaving of read/write operations and the time when each such operation is to be scheduled.

Time	T4	T5	T6
1	Read (A)		
2		Read (A)	
3			Read (D)
4			Write (D)
5			Write (A)
6		Read (C)	
7	Write (B)		
8		Write (B)	

Determine whether the proposed sequence of operations will lead to a serialisable execution if concurrency is to be controlled using:

i) Strict two-phase locking

ii) Timestamp ordering.

The three parts carry, respectively, 30%, 20% and 50% of the marks.

4 The process of recovery sometimes involves *undoing* the effect of certain transactions and *redoing* the effect of other transactions.

a Explain why the *undo* and *redo* operations are said to be *idempotent.*

Explain carefully the process whereby a database can be restored to a consistent state following a system failure as a result of which the contents of main memory and the input/output buffers have been lost whilst leaving the database itself intact. Identify any features of this process which may be redundant.

b Discuss how recovery from system failure could be guaranteed

i) without ever having to *undo* any transactions;

ii) without ever having to either *undo* or *redo* any transactions.

c A travel agent making a <booking> for a customer would typically be involved in three activities

- booking an airline seat;
- making a hotel reservation;
- arranging a car rental.

Discuss the advantages to be gained in the areas of concurrency and recovery if each activity within <booking> is considered to be a separate transaction as opposed to considering <booking> to be one large transaction.

The three parts carry, respectively, 35%, 35% and 30% of the marks.