UNIVERSITY OF LONDON
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 1996

MEng Honours Degrees in Computing Part IV
MSc Degree in Foundations of Advanced Information Technology
for Internal Students of the Imperial College of Science, Technology and Medicine

This paper is also taken for the relevant examinations for the Diploma of Membership of Imperial College Associateship of the City and Guilds of London Institute

PAPER 4.11

DATABASES II

Tuesday, May 14th 1996, 3.00 - 5.00

Answer THREE questions

For admin. only: paper contains
4 questions
4 pages (excluding cover page)

1a In the context of distributed databases, explain what is meant by a *minterm predicate,* and formally define what is meant by a *fragment.*

Discuss the conditions under which a set of predicates would be considered to be:

i) minimal

ii) complete

and explain why it is desirable for predicates to satisfy both of these conditions.

b Distinguish between *location transparency* and *fragmentation transparency.*

Discuss, with the aid of suitable examples, the problems associated with implementing fragmentation transparency in applications which involve regular updating of values in a database which has been fragmented using:

i) hybrid fragmentation

ii) derived fragmentation.

c Explain what is meant by a *join graph* and discuss the advantages of ensuring that join graphs are *simple.*

The following three global relations have been defined:

DOCTOR (DocNum, DocName, Dept)
PATIENT (PatNum, PatName, Dept, DocNum)
CARE (DocNum, PatNum, Drug, StartDate)

The attribute DocNum in PATIENT gives the doctor who is treating the given patient in that department. A patient may be receiving treatment in various departments.

The relation CARE gives details of the drugs prescribed by a given doctor to a given patient and the date the course of drugs was started.

The relation DOCTOR is horizontally fragmented on DocNum.
The relation PATIENT is horizontally fragmented on PatNum.

Two main applications are run using these relations - one requires details about doctors and the patients they are treating with drugs, the other requires details about patients and the drugs that have been prescribed for them.

Discuss the fragmentation process that would be best suited to the relation CARE.

2a Explain what is meant by *transaction atomicity.* Briefly discuss how atomicity can be achieved in a distributed environment using the two-phase commit (2PC) protocol with a hierarchical communication structure.

b Discuss the effect on the hierarchical 2PC protocol if some of the participating sites are ones at which no updates are being undertaken by the transaction in question (i.e. they are read-only sites). Give details of all the messages sent during both phases of the protocol.

c Explain what is meant by a *blocking protocol* and discuss the circumstances under which blocking will occur in the 2PC protocol.

Explain how the addition of a third phase to the 2PC protocol can eliminate the possibility of blocking.

d In the three-phase commit protocol, four of the places at which a process (the coordinator or a participant) could be waiting for a message are:

- a participant is waiting for a PREPARE message
- the coordinator is waiting for a vote (COMMIT/ABORT)
- a participant is waiting for a PRE-COMMIT or ABORT message
- the coordinator is waiting for acknowledgement following the PRE-COMMIT or ABORT message.

Should a process fail to receive a message within a given time interval, a *timeout* is said to have occurred, and appropriate action will need to be taken by that process.

For each of the places mentioned above, discuss the actions that need to be taken by the relevant process in the event of a timeout occurring.

The four parts carry respectively, 20%, 25%, 20% and 35% of the marks.

Turn over

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

Prove that if two histories are conflict equivalent, then their serialisation graphs are identical.

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

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

Determine whether this history is (i) conflict serialisable

(ii) view serialisable

and, where appropriate, give the equivalent serial history.

By considering the operations included in the transactions discuss conditions under which both tests of serialisability will, in general, provide identical results.

b i) Discuss what is meant by *livelock.*

Wound-Wait and *Wait-Die* are timestamp-based approaches for deadlock prevention.

ii) Discuss how it can be guaranteed that timestamps will be unique in a distributed environment.

iii) Compare and contrast the Wound-Wait and Wait-Die approaches and show how livelock is prevented under these approaches.

iv) If timestamps were not guaranteed to be unique, would these two approaches still prevent deadlock and would livelock still be avoided?

4a Let **IN** be the intersection operation and **DF** be the difference operation in relational algebra.

i) Show that R **IN** S = R **DF** (R **DF** S)

ii) The following rule applies in the *algebra of qualified relations*:

[R : q_R] **DF** [S : q_S] => [R **DF** S : q_R]

By considering the relationship between **IN** and **DF** given in (i) derive the corresponding rule for:

[R : q_R] **IN** [S : q_S]

and for

[S : q_S] **IN** [R : q_R]

Comment on the validity of the results obtained.

b i) Define the *semi-join* (**SJ**) operator.

ii) Let R' = R $\mathbf{SJ_A}$ S

The operation R $\mathbf{JN_A}$ S can be expressed as R' $\mathbf{JN_A}$ S.

Explain the benefit of using the semi-join approach (i.e. calculating R' $\mathbf{JN_A}$ S) when evaluating a query as opposed to executing the join directly.

iii) The semi-join approach can be extended to the case where three relations (R, S, T) are to be joined.

In this case R' can be replaced by R **SJ** (S **SJ** T); further reductions can be made by additional chains of semi-joins.

Using relations R, S, T given below, illustrate the effect of the semi-join approach on relation R.

R	A	B
	1	e
	2	f
	3	g

S	B	C
	e	u
	f	v
	g	w

T	C	A
	u	2
	v	3
	w	4

End of paper