UNIVERSITY OF LONDON
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 1997

BEng Honours Degree in Computing Part III
MEng Honours Degrees in Computing Part IV
MEng Honours Degree in Information Systems Engineering Part IV
BSc Honours Degree in Mathematics and Computer Science Part III
MSci Honours Degree in Mathematics and Computer Science Part III
MSc Degree in Advanced Computing
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
Associateship of the Royal College of Science

PAPER 3.11 / 4.11 / I4.4

ADVANCED DATABASES

Monday, April 21st 1997, 2.30 - 4.30

Answer THREE questions

For admin. only: paper contains 4 questions

1a In the context of distributed databases, explain what is meant by a *complete* set of simple predicates and discuss why this is a desirable property for simple predicates to have.

Explain what is meant by *fragmentation transparency.* The information defining the fragmentation to be adopted would often form part of the data dictionary for the database. Discuss how the location of the data dictionary affects the degree of site autonomy that can be achieved.

b The following relations store details of the different departments of a University, the staff employed and courses run by the University and the students attending those courses. The University is split across several sites.

DEPT (Dept Name, Dept Head, Location)

STAFF (Staff No, Staff Name, Dept Name)

STUDENT (Student No, Student Name, Tutor No.)

COURSE (Staff No, Course No, Student No, Start Date)

A department is led by one Director (whose name is given in Dept Head) and employs several members of staff.

The attribute Tutor No (in STUDENT) gives the member of staff (identified by Staff No.) acting as Tutor to the given student. The University most commonly accesses STUDENT information grouped by Tutor.

The relation COURSE gives details of the courses given by members of Staff and the students attending these courses.

The relation DEPT is horizontally fragmented on Location, and STAFF has a fragmentation derived from that of DEPT.

i) From the point of view of database design, explain the benefits of ensuring that a department is not split across two locations. State another similar constraint that should be maintained, and give reasons for imposing this constraint.

ii) Using the semi-join operator, give a relational algebra expression to define the STAFF fragments.

iii) Explain what is meant by *referential integrity*. Give an example of a referential integrity constraint that would need to be maintained for the given relations.

iv) Outline the operations required when a different member of staff becomes Tutor to a given student. (A student's Tutor may be drawn from any department of the University).

v) Two main applications are run using these relations - one requires details about staff and the courses they give, the other requires details about students and the courses they attend.

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

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

©University of London 1997

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 centralised coordinator. Do not consider details of the log entries.

Discuss the changes that could be made to this protocol to reduce the task of the coordinator if each participant site communicates with all other participant sites. Contrast the number of messages required under this approach with the number of messages required where the participants normally only communicate with the coordinator.

b Under 2PC it is a normal requirement for any local decision to be first recorded in the local log before being communicated to other sites. Explain why this is necessary. Consider decisions made at the coordinator site and at participant sites.

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

Suppose the 2PC protocol were changed such that the coordinator were to send all the necessary COMMIT messages *before* writing the corresponding entry in its log. Discuss the circumstances under which this variation of 2PC will be non-blocking.

The three parts carry, respectively, approximately 40%, 25% and 35% of the marks.

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

By choosing an appropriate example, show that a *view serialisable* history may not necessarily be *conflict serialisable.*

ii) Explain what is meant by a *one copy serialisable execution.*

A non-replicated history H_1 contains operations from transactions T_1, T_2, T_3.

The following serialisation graph has been constructed based on the operations included in H_1:

$$
\begin{array}{ccc}
SG_1 = & T_1 & \longrightarrow T_2 \\
 & \downarrow & \\
 & T_3 &
\end{array}
$$

Explain clearly why the execution implied by history H_1 must be (conflict) serialisable.

If the same rules for the construction of a serialisation graph were applied to a *replicated* history H_2 and the resulting serialisation graph was identical to SG_1, explain why the execution implied by history H_2 need not necessarily be serialisable.

b Describe a *centralised* approach to deadlock detection in a distributed environment.

Discuss, with the aid of an example, a *fully distributed* approach to deadlock detection and discuss any advantages it may have over the centralised approach.

One way of detecting the possible existence of deadlock in a database system is with the 'time-out' method whereby a transaction is aborted if a given time interval has elapsed after the transaction has entered a wait state.

Discuss the problems associated with choosing an appropriate time-out interval and explain why the time-out method is only considered an acceptable approach for lightly loaded systems.

©University of London 1997

4a Explain what is meant by a *join graph* and discuss the advantages of ensuring that join graphs are *simple.*

b In the *algebra of qualified relations* one of the rules that applies is:

$$[\, R : q_r \,] \ \mathbf{JN}_F \, [S : q_s] \ \Rightarrow \ [\, R \ \mathbf{JN}_F \ S : q_r \text{ AND } q_s \text{ AND } F \,]$$

Prove this rule stating clearly any further rules used in the proof. (Note that join (JN) can be derived from selection and cartesian product).

c A medical database contains information on doctors and the clinics in which they practice, and on the drugs prescribed by these doctors.

The information is held in the following two global relations :

DOCTOR (Doc No, Doc Name, Clinic)

DRUG (Doc No, Drug, Dosage, Date)

There are currently two clinics in operation; a doctor may practice in more than one clinic.

The following query has been formulated and is to be run against the given relations:

Find the Doc Name of each doctor who has an entry in the DRUG relation.

i) Express the query in relational algebra and draw the corresponding operator tree.

The relation DOCTOR has been horizontally fragmented on Clinic; DRUG has a fragmentation derived from that applied to DOCTOR.

ii) Explain what is meant by a *canonical expression*. Incorporate the relevant canonical expressions in the operator tree produced in part i). By reordering the operations included in the tree now produced, draw an alternative operator tree with equivalent effect.

iii) If doctors were restricted to working in only one clinic, which operator tree would be chosen for the evaluation of the given query? Justify your answer and comment on any implications the approach chosen may have for the storage of the different fragments.

The three parts carry, respectively, approximately 15%, 25% and 60% of the marks.

End of Paper

©University of London 1997