UNIVERSITY OF LONDON
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2001

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

Friday 18 May 2001, 10:00
Duration: 120 minutes

Answer THREE questions

Paper contains 4 questions
Calculators not required

1a Consider the following statements S1-S4:

S1 Every BCNF relation is 3NF.

S2 Every 3NF relation is BCNF.

S3 Some BCNF relations are 2NF.

S4 Every 2NF relation is BCNF.

For each statement:

i) State whether it is true or false.

ii) For each of those that you decide is false give a counterexample (i.e. an example that demonstrates its falsity).

iii) If you decide that any of the statements is true, then for only one of the true ones give a detailed proof.

b i) What is the entity integrity rule?

ii) Using an example, describe three disadvantages that a 2NF relation scheme which is not 3NF can have compared to the relations that result from any of its 3NF normalisations.

c i) Without referring to the concept of functional dependencies, explain precisely what it means for the decomposition of a relation scheme R into two relation schema R1 and R2 to be lossless.

ii) Consider relation scheme R(A,B,C,D,E,F) with the following functional dependencies:

$AB \rightarrow DE$

$C \rightarrow F$

Suppose R is decomposed into two schema R1(A,B,C,D,E) and R2(A,B,F).

Show that the decomposition is not lossless by constructing a suitable example of instances of the schema R, R1, R2. Note that it is not enough to state sufficient conditions for losslessness.

The three parts carry, respectively, 55%, 20%, 25% of the marks.

2 Consider the following relation schema regarding students, courses and lecturers. The primary keys are underlined.

Student(<u>Name</u>, DegreeYear)

Compulsory(<u>DegreeYear</u>, <u>Course</u>)

Optional(<u>DegreeYear</u>, <u>Course</u>)

Teaches(<u>Course</u>, Lecturer)

Registered(<u>Name</u>, <u>Course</u>)

We abbreviate the schema names to S, C, O, T, R, respectively.

S stores information about the degrees the students are following and the year they are currently in. For example a tuple <jeff, ise3> indicates Jeff is a third year ISE student. C and O store information about which courses are compulsory and optional, respectively, for which degrees and years. T stores information about which lecturer teaches which course. R stores information about which (lecture) courses each student has registered to take.

Consider the following queries to instances of the schema above:

Q1 Find the names of all the students who have not registered for all of their compulsory courses.

Q2 Find the names of all the lecturers who teach more than one optional course for any year on any degree programme (i.e. for any "DegreeYear").

Q3 Find all the courses that are offered as options on a year on a degree, but no student in that year and degree has registered to take them.

Q4 Find all the courses taught by Hopkins that have more than 100 students registered to take them.

a Formulate queries Q1-Q3 in relational algebra.

b Formulate queries Q1-Q3 in relational tuple calculus.

c Formulate query Q4 in SQL.

The three parts carry, respectively, 40%, 40%, 20% of the marks.

3a i) What are the four properties, known as the ACID properties, that all database management systems are required to possess? Give a brief description of each property.

ii) Which ACID properties do locking mechanisms attempt to maintain? Which ACID properties does the database transaction recovery manager attempt to maintain?

b i) What operations are involved in taking a checkpoint, and in what order? Why is the order of some of the operations crucial in the recovery process?

ii) Give the advantages of checkpointing in recovering from system failures.

c Consider the following relation scheme

Personnel(<u>Number</u>, Name, ManagerNumber, Department, Salary)

used to store details of employees consisting of their identifying number, name, department, salary, and the identifying number of their manager.

i) Suggest a useful foreign key to help maintain the integrity of this scheme. Give the foreign key and what it references.

ii) Using abstract syntax, write two active rules for maintaining referential integrity according to your chosen foreign key, based on the following policies:

- Deletions nullify.
- Updates cascade.

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

4a Explain briefly when a schedule of transactions is *conflict serialisable.*

b By drawing the precedence graphs of the following schedules, decide which is conflict serialisable and which is not. For all those that are conflict serialisable give all their serialisations.

i) r1(x), w1(x), w3(x), r2(y), r3(y), w3(y), w1(y), r2(x)

ii) r1(x), r3(y), w1(y), w4(x), w1(t), w5(x), r2(z), r3(z), w2(z), w5(z), r4(t), r5(t)

c Suppose the schedule in b(ii) is run under the two phase locking protocol (2PL) and that locks are released only at the end of each transaction. By drawing the wait-for graphs at various appropriate stages of the schedule:

i) Identify any transactions that are placed in waiting.

ii) Determine whether or not the schedule deadlocks.

d Consider a transaction scheduler that uses 2PL and releases locks only at the end of each transaction. Prove that under this scheduler transactions can be serialised in the order in which they commit. It is enough here to prove it in the case of two transactions.

The four parts carry, respectively, 10%, 50%, 20%, 20% of the marks.