UNIVERSITY OF LONDON
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2003

BEng Honours Degree in Computing Part I
MEng Honours Degrees in Computing Part I
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 C130

# DATABASES I

Wednesday 7 May 2003, 16:00
Duration: 90 minutes
(Reading time 5 minutes)

*Answer THREE questions*

Paper contains 4 questions
Calculators not required

1a A hospital has a number of wards. Each ward contains a number of patients who are tended by the nurses associated with that ward. The ward's nursing staff are supervised by a Ward Sister, who is also a member of the nursing staff, and who has responsibility for only one ward. Patients do not move to other wards during their stay in hospital. Nurses work in only one ward.

The hospital has a number of doctors and each is associated with one ward and is responsible for the treatment of several of the patients within the ward. Several doctors would normally be allocated to a ward; however any one patient would be treated by only one doctor who may prescribe several drugs to help improve the patient's condition.

Draw an entity – relationship diagram illustrating the relationships between the entities: Hospital, Ward, Doctor, Patient, Drug, Nurse and Ward-Sister. The information to be modeled should include the date of admission of a patient to the ward, the date a patient was last seen by his doctor, the date a given drug was first prescribed to a given patient by his doctor, and the date a given doctor was first appointed to the ward in which he works.

b The relations R (A,B,C,D) and S (W, X, Y, Z) have been defined where R is defined over the attributes A,B,C,D and S is defined over the attributes W,X,Y,Z.

The relational algebra operator UNION (U) is to be applied to R and S.

i) Discuss the necessary condition for the operation R U S to be valid.

ii) Give the relational algebra expression required if the result of R U S is to be held in relation T with attributes K,L,M,N.

c The following relations have been defined:

Suppliers (<u>Sno</u>, Sname, Address)

Parts (<u>Pno</u>, Pname, Colour)

Catalogue (<u>Sno</u>, <u>Pno</u>, Price)

The key fields are underlined. The Catalogue relation lists the price charged for a given part by a given supplier.

Express the following queries in *relational algebra*:

i) Find the Supplier name (Sname) for those suppliers who supply some red part.

ii) Find the Supplier number (Sno) for those suppliers who supply every part.

iii) Find the Part name (Pname) for those parts that are supplied by at least two different suppliers.

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

2a i) Given a relation R, a set of attributes A of R, and a set of functional dependencies F that hold for R, explain what is meant by the *closure* $A^+$ of A under F.

ii) A *superkey* for a relation R can be defined as a set of attributes K of R that includes some candidate key of R as a subset.

Give an alternative definition of a superkey expressed in terms of the closure $K^+$ of K under a given set of functional dependencies.

b The set of dependencies

$$F_c = \{A \rightarrow B, B \rightarrow C, A \rightarrow D\}$$

is said to be an *irreducible cover* for the set of dependencies

$$F = \{A \rightarrow BC, B \rightarrow C, A \rightarrow B, AB \rightarrow C, AC \rightarrow D\}$$

i) What are the properties of $F_c$ that make it an irreducible cover for F?

ii) Using only Armstrong's Axioms (Reflexivity, Augmentation, Transitivity) derive from $F_c$ each of the three dependencies that appears in F but not in $F_c$.

c A relation Y is defined over the attributes S, B, D. A tuple of this relation denotes that sailor S has reserved boat B on date D.

A sailor can reserve a given boat for at most one day and on any given day at most one boat can be reserved.

i) Indicate the functional dependencies implied by the above information.

ii) Determine in which Normal Form this relation is currently held and, if necessary, decompose the relation into Boyce-Codd Normal Form. Discuss whether your decomposition is *non-loss* and *dependency preserving.*

*The three parts carry, respectively, 15%, 40% and 45% of the marks*

3a Define what is meant by a *view* in the context of relational databases.

A relation M is defined over attributes <u>A</u>,B,C,D and a relation N is defined over attributes <u>E</u>,F,G where the relevant key attributes are underlined.

A view, V, is to be defined on the two base relations M and N and is to include the attributes C,D,E.

Discuss the problems associated with adding tuples to the view V.

b i) Explain what is meant by a *schedule* in the context of concurrent running of transactions and distinguish clearly between a *serial* schedule and a *serialisable* schedule.

ii) Explain what is meant by the *two-phase locking protocol* (2PL) and distinguish between *Basic* 2PL and *Strict* 2PL.

c Distinguish between *write* locks and *read* locks.

Construct a serialisable schedule for the two transactions T1, T2 given below incorporating write/read locks as appropriate. Give the serial sequence of transactions to which the constructed schedule is equivalent.

| <u>T1</u> | <u>T2</u> |
|---|---|
| begin trans | begin trans |
| **read** (A) | **read** (A) |
| **read** (B) | **read** (B) |
| **read** (Z) | **display** (A + B) |
| . | commit |
| . | |
| . | |
| **write** (A) | |
| commit | |

A,B,..., Z are data items. The operation **write** (A) records the value of A on the database, **display** (A) displays the current value of A on a terminal.

d It has been suggested that if read locks could be upgraded to write locks and write locks downgraded to read locks this could increase concurrency.

Adapt the two-phase locking protocol to include this facility and show how use of this facility could increase concurrency in the case of T1 and T2.

*The four parts carry, respectively, 25%,25%,20%, and 30% of the marks.*

4a i) Describe two approaches to deadlock detection giving the advantages/disadvantages of each such approach.

ii) Describe either the *Wait-Die* or the *Wound-Wait* approach to deadlock prevention. State the name of the approach that you are describing.

iii) The following rule has been suggested for the prevention of deadlocks, where $T_i$ is requesting a lock currently held by $T_j$ :

'If $T_i$ is younger than $T_j$, then $T_i$ waits else $T_i$ aborts'

($T_i$ is *younger* than $T_j$ if it was initially scheduled after $T_j$)

Discuss whether this rule will prevent deadlocks and whether it will prevent cyclic restarts arising.

Contrast the behaviour of this rule with the approach discussed in part (ii).

b i) Explain what is meant by a *checkpoint* and why this is a *useful* feature of the recovery process.

Describe other facilities which are *necessary* to enable recovery of the database to take place.

iii) A transaction T1 updates a relation R1, invokes a transaction T2 to update relation R2, and then continues by updating other relations. A system failure occurs after the successful completion of T2 but before T1 completes all its actions.

Discuss the problems for the recovery process implied by allowing such nesting of transactions.

*The two parts carry, respectively, 55% and 45% of the marks.*