UNIVERSITY OF LONDON
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2001

BEng Honours Degree in Computing Part II
MEng Honours Degrees in Computing Part II
BSc Honours Degree in Mathematics and Computer Science Part II
MSci Honours Degree in Mathematics and Computer Science Part II
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 C230=MC230

DATABASES I

Friday 4 May 2001, 14:00
Duration: 90 minutes
(Reading time 5 minutes)

Answer THREE questions

Paper contains 4 questions
Calculators not required

1 A certain 1NF database consists of a single relation R having the schema

R (A, B, C, D, E)

The following set F of dependencies is asserted to hold for R:

$$F = \{ AB \rightarrow E,\ AB \rightarrow CD,\ C \rightarrow E,\ D \rightarrow E,\ C \rightarrow D,\ B \rightarrow D \}$$

a Construct an *irreducible cover* F_C for F, justifying each step you make. What are the potential benefits of an irreducible cover ?

b i) Justifying your answer, identify a key for R.

ii) Discuss carefully whether F implies that ABC is a key for R.

iii) Transform the database first to 2NF and then to 3NF, explaining clearly the principles and operations involved.

c For each of the following separate update requests

- add a row for which B=1 and D=1
- for all rows in which D=2, change E to 2
- delete all records for which D=3

compare its effect when issued to the database R with its effect when issued to your 3NF database derived in part b.

The three parts carry, respectively, 30%, 40% and 30% of the marks.

2 ***Part b of this question refers to the Supplementary Sheet on Page 3 showing the Suppliers-and-Parts database.***

Note—relational algebra allows just these operations on relations A, B:

$A \cup B$	$A[\ldots]$
$A \cap B$	A *where condition*
$A \times B$	$A \triangleright\!\triangleleft B$
$A \div B$	$A - B$
	rename A as B

a Let R be a relation having the schema (X,Y).

i) Write a single relational algebra expression for a 4-ary relation whose first and fourth columns each contain data identical to that in R's X column and whose second and third columns each contain data identical to that in R's Y column. State the schema of this new relation.

ii) Describe the precondition and the result of the operation R ÷ D where D is some relation. How could the same effect be obtained without using ÷ ? *Hint—first construct (R[X] x D).*

iii) If X is a key for R, under what condition could X also be a key for the relation (R[X] x R[Y])? Justify your answer.

b Formulate in relational algebra the query

"Find the identifiers and colours of parts that are supplied by Clark and made in cities where screws are not made."

Outline the evaluation for the given data and state the answer produced.

The two parts carry, respectively, 60% and 40% of the marks.

Supplementary Sheet for Question 2

Relation S

S#	SNAME	SCITY	STATUS
S1	Smith	London	20
S2	Jones	Paris	10
S3	Blake	Paris	10
S4	Clark	London	20
S5	Adams	Athens	30

For each supplier, gives their name, status and the city in which they are based

Relation P

P#	PNAME	PCITY	COLOUR	WEIGHT
P1	Nut	London	Red	12
P2	Bolt	Paris	Green	17
P3	Screw	Rome	Blue	17
P4	Screw	London	Red	14
P5	Pin	Paris	Blue	12
P6	Brace	London	Red	19

For each part, gives its physical properties and the city in which it is made

Relation SP

S#	P#	QTY
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

For each supplier and part, gives the quantity of that part supplied by that supplier

Section B *(Use a separate answer book for this Section)*

3a Explain what is meant by a *serialisable history* and a *serialisation graph.*

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

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

$H_1 = w_1[x]\ w_2[x]\ w_2[y]\ c_2 w_1[y]\ w_3[x]\ w_3[y]\ c_3 w_1[z]\ c_1$

Determine whether this history is serialisable.

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 Explain what is meant by *locking granularity.*

A bank maintains the following two relations in which details of the sums deposited with it are recorded:

Accounts - with attributes Account No., Branch, Balance - records for each account (identified by Account No.), the branch at which the account is held and the balance of the account.

Assets - with attributes Branch, Total - records for each branch, the sum of all balances held at that Branch.

The following two transactions are typically run against these relations:

T_4 - to insert into Accounts details of new accounts that have been opened and make the appropriate changes to Assets.

T_5 - to compute (from Accounts) the sum of the balances held at a given branch, and compare that figure with the total for that branch (as given in Assets).

Explain, with the aid of an example, why 2PL will not guarantee serialisability if locking is done at the tuple granularity level. Discuss the level of locking granularity required to ensure serialisability.

The three parts carry, respectively, 25%, 25% and 50% of the marks.

4a In the context of database recovery, discuss the information that would typically be held in the *log* and explain what is involved when a *checkpoint* is taken.

A failure has occurred involving the loss of the contents of main memory and the input/output buffers but leaving the database itself intact. A log-based recovery procedure is to be used to restore the database to a consistent state. Explain why it may be necessary to *undo* certain transactions and *redo* other transactions during this procedure.

Describe a recovery procedure that does not require the use of *undo* and *redo* operations.

b A college is organised into several academic departments. Within each department there are several members of staff undertaking teaching duties; each such staff member would be associated with only one department.

Each member of the teaching staff is responsible for the teaching of one or more subjects. It is college policy to employ no more than one member of staff for the teaching of any one subject. Staff are assessed on an annual basis, a separate assessment being made for each subject taught.

Students typically follow five subjects per academic year and would be awarded a grade at the end of the year for each subject studied.

The college wishes to set up a database to hold this information. In particular the following information is to be recorded:

- for each student: student no, name, address, subjects studied
- for each subject: name, timetable information for the current year
- for each staff member: name, address, dept, date of joining dept, subjects taught.

There would typically be several periods per week for each subject taught. For each such period, the day, time and location (room no) of the period would be recorded. The timetable pattern would not vary during the year.

The database would also include all grades/assessments awarded at any time to students/staff, and the dates of those awards. The database would not include any data relating to students/staff who are no longer in the college.

Construct an entity-relationship diagram illustrating the relationships between student, subject, staff and department. Include timetable information and all attributes described. Identify any dominant/subordinate and strong/weak entities.

Transform the entity-relationship diagram into a set of relations, indicating the primary key of each relation.

The two parts of the question are equally weighted.