UNIVERSITY OF LONDON
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 1998

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 Royal College of Science
Associateship of the City and Guilds of London Institute

PAPER 2.2 / MC2.2

DATABASES

Tuesday, May 5th 1998, 4.00 - 5.30

Answer THREE questions

For admin. only: paper contains 4 questions

1a From Armstrong's axioms

$$\begin{array}{lll} \text{A1:} & \models X \rightarrow X & \textit{reflexivity} \\ \text{A2:} & \left| \begin{array}{l} X \rightarrow Y \models XZ \rightarrow Y \\ X \rightarrow Y \models XZ \rightarrow YZ \end{array} \right. & \textit{augmentation} \\ \text{A3:} & (X \rightarrow Y) \ \& \ (Y \rightarrow Z) \models X \rightarrow Z & \textit{transitivity} \end{array}$$

derive the axioms

$$\begin{array}{lll} \text{A5:} & \left| \begin{array}{l} X \rightarrow YZ \models X \rightarrow Y \\ X \rightarrow YZ \models X \rightarrow Z \end{array} \right. & \textit{projectivity} \end{array}$$

b Given a relation with attributes R of which K is some subset, state the conditions under which K is a primary key for R.

An additional constraint normally imposed upon a primary key is the so-called entity integrity rule. What is this rule and what motivates it?

c A 1NF database is formed initially with the single relation scheme

R (Id, Name, Dept, Fees, Nat, Email)

with primary key {Id, Name}, and the content of R is

Id	Name	Dept	Fees	Nat	Email
10196uk	A. Smith	doc	h	uk	as1c
10397uk	B. White	maths	h	uk	bwm
10397uk	B. Davis	maths	h	uk	bdm
20296irn	C. Emani	doc	os	iran	cec
10596uk	D. Jones	eee	h	uk	dje
12297uk	A. Smith	doc	h	uk	as2c
30496fr	M. Lauren	doc	eec	france	mlc
20197ch	L. Chan	eee	os	china	lce

Additionally the following functional dependencies are asserted:

Nat → Fees   Id → Nat   Email → Dept

i) Derive the dependency Id → {Fees, Nat}.
Explain why the database is not 2NF and then convert it to 2NF, exploiting only this dependency and the knowledge of the primary key.
Note: present just the new schemes—there is no need to reproduce the data—and identify their primary keys.

Present and explain an example of an update that would present difficulties with the original database but would be viable with the 2NF form.

ii) Convert the database further into 3NF, explaining the steps involved.
Note: present just the new schemes—there is no need to reproduce the data—and identify their primary keys.

Present and explain an example of an update that would present difficulties with the 2NF form but would be viable with the 3NF form.

The three parts carry, respectively, 20%, 20% and 60% of the marks.

2 ***For this question you will need to refer to the Supplementary Sheet showing the Suppliers-and-Parts database.***

a Formulate in relational algebra the following queries, outlining their evaluations and stating the answers obtained:

i) Find the names of those suppliers who supply a quantity (QTY) of less than 300 of any particular part.

ii) Find the colours of those parts made in no cities in which any suppliers of screws are based.

iii) Find the names of those suppliers who supply all blue parts.

Note—this algebra allows just the following nine operations on relations A, B:

$A \cup B$	$A[\ldots]$
$A \cap B$	A $\underline{\textit{where}}$ *condition*
$A \times B$	$A \bowtie B$
$A \div B$	$A - B$
	$\underline{\textit{rename}}$ A $\underline{\textit{as}}$ B

b Reformulate in tuple calculus the query in part aiii), declaring the ranges of your variables.

Hint—you will need three variables ranging over the three relations, and the query's <u>*where*</u> *formula will have the form* $(\forall\ldots)((\exists\ldots)\ \ldots \leftarrow \ldots)$.

The two parts carry, respectively, 80% and 20% of the marks.

Turn over ...

3a Suppose relation *ShipData* has attributes *Ship_name, Year, Captain, Engineer, Tonnage, Port_of_reg, Country_of_reg* and the following functional dependencies (FDs):

Ship_name, Year —> Captain, Engineer

Captain, Year —> Ship_name

Engineer, Year —> Ship_name

Ship_name —> Tonnage, Port_of_reg, Country_of_reg

Port_of_reg —> Country_of_reg

(Captains and Engineers are assigned to ships annually.)

i) Define the Boyce-Codd normal form (BCNF). It is not necessary to explain what is meant by functional dependency and candidate key but you should define any other terms you use.

ii) Show that the relation *ShipData* is not in BCNF.

iii) Construct a (non-loss) decomposition of *ShipData* into a set of BCNF relations. Document each step briefly. Show that your decomposition is non-loss, stating any standard results you need.

b Consider the following relation:

ShipRoutes (Ship, From, To, Journey_time, Class, Price)

with functional dependencies

Ship, From, To —> Journey_time

Ship, From, To, Class —> Price

i) What is a *multi-valued dependency* (MVD)? Your definition need not be formal but must be precise.

ii) With reference to *ShipRoutes*, any given ship can have several routes (*From*, *To* pairs) and several available *Class*'s(1st, 2nd, 3rd, etc). The available *Class*'s for a ship are independent of the route.
Write down two non-trivial, non-FD MVDs that hold in relation *ShipRoutes*.

iii) Write down a (non-loss) decomposition of *ShipRoutes* into a set of 4NF relations. It is not necessary to document the steps of the decomposition.

iv) Justify *briefly* why your answer to part b(iv) is in 4NF. It is not necessary to state the definition of 4NF itself.

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

4a Give brief (one or two sentence) explanations of each of the following.

i) What is a *schedule* of transactions?

ii) What is a *serial* schedule of transactions?

iii) What is a *serialisation* of a schedule of transactions?

iv) What is a *serialisable* schedule of transactions?

b Let transactions T1, T2, T3 have the following effects:

T1: *adds £200 pounds to account* A;

T2: *moves £100 pounds from account* A *to account* B;

T3: *moves 50% of account* B *to account* A.

(The internal details of T1, T2, T3 are shown in part c. They are not required for this part of the question.)

Suppose transactions T1, T2, T3 run concurrently. If the initial values of A and B are *a* and *b*, respectively, give all the possible correct values of A, B, and the total A+B after all three transactions have completed.

c Consider the following schedule S for the transactions T1, T2, T3.

T1	T2	T3
		READ(B, v)
		WRITE(B, 0.5*v)
READ(A, x)		
	READ(A, y)	
	READ(B, z)	
WRITE(A, x+200)		
	WRITE(A, y–100)	
	WRITE(B, z+100)	
		READ(A, w)
		WRITE(A, w + 0.5*v)

WRITE(X, *exp*) means evaluate expression *exp* and write the result to X.

i) Show that S is not conflict serialisable by constructing its precedence graph.

ii) Show that S is not serialisable.

iii) Let S' be the schedule of T2 and T3 that is obtained from S by deleting all operations of T1. Show that S' is serialisable but not conflict serialisable.

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

End of paper

Supplementary Sheet for Paper 2.2=MC2.2: 1997-98

Relation S

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

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

Relation P

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

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