UNIVERSITY OF LONDON

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 1999

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
MSc Degree in Computing Science
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 M 3.11 / I 3.2

DATABASES

Tuesday, May 4th 1999, 10.00 – 12.00

Answer THREE questions

For admin. only:
paper contains 4 questions

1a Let A, B, C, D be sets of attributes. Which of the following statements about functional and multi-valued dependencies is (are) true, and which is (are) false?

(S1) If $A \rightarrow BC$ then $A \rightarrow\!\!\rightarrow B$.
(S2) If $A \rightarrow B$ & $BC \rightarrow D$ then $AC \rightarrow D$.
(S3) If $A \rightarrow B$ & $BC \rightarrow D$ then $A \rightarrow D$ & $C \rightarrow D$.

For all of those that are false give a counter-example, and for all that are true give a proof using only Armstrong's axioms, and, if necessary, the concept of multi-valued dependencies. Explain every step of your proof(s) by stating the axioms or concepts used.

b Define the fourth normal form (4NF).

c Consider a relation scheme R(A, B, C, D, E, F) with the following set of functional dependencies.

$AB \rightarrow C$
$CD \rightarrow E$
$A \rightarrow E$
$BE \rightarrow F$
$F \rightarrow CA$

i) Give two candidate keys for R.

ii) Is R in 4NF? Give a precise and detailed explanation for your answer.

iii) Consider a decomposition of R into two relation schema R1(C,D,E) and R2(A,B,C,D,F). Discuss the advantages and the disadvantages of this decomposition compared with R in terms of possible data redundancies, update problems and dependency preservation.

The three parts carry 45%, 10%, 45% of the marks, respectively.

2a Let relation schema R1 and R2 be defined as R1(A,B,C) and R2(A,B,D). Give a tuple relational calculus expression that is equivalent to the following relational algebra expression

(R1[A,B] – R2[A,B]) JOIN R1.

b An estate agent maintains the following relation schema. The primary key of each scheme is underlined.

PROP(Code, AskingPrice)
CUST(Surname, Initial, Phone, MaxToSpend)
VIEWED(Surname, Initial, Code)
OFFERED(Surname, Initial, Code, Amount)

A tuple <*c, a*> in *PROP* indicates that the asking price for the property identified by code *c* is *a* pounds. A tuple <*s, i, p, m*> in *CUST* indicates that the customer identified by surname, *s* and initial *i*, has phone number *p* and has a maximum of *m* pounds to spend on a property. A tuple <*s,i,c*> in *VIEWED* indicates that the customer whose surname is *s* and whose initial is *i* has viewed the property identified by code *c*. A tuple <*s,i,c,a*> in *OFFERED* indicates that the customer whose surname is *s* and whose initial is *i* has offered amount *a* (*a* pounds) for the property identified by code *c*.

You will not need to use relation *PROP*. It is given for clarity only.

Consider the following queries to the relations above.

(Q1) Find the surnames and phone numbers of all the customers who have viewed more than one property.
(Q2) Find the surnames, initials and phone numbers of all the customers who have made an offer on at least one property they have not viewed.
(Q3) Find the surnames, initials and phone numbers of all the customers who have made an offer on at least all the properties they have viewed.

i) Formulate all the queries above in relational algebra.
ii) Formulate queries (Q2) and (Q3) in tuple relational calculus.
iii) Formulate query (Q1) in SQL.

Parts a, b(i), b(ii), b(iii) carry 15%, 45%, 30%, 10% of the marks, respectively.

Turn over ...

3 This question is concerned with entity-relationship (ER) modeling and the relational model.

Santa Claus would like to keep data about the houses he has to visit at Christmas, the presents he has to drop off and his team of reindeers that have to transport him. His records have to consist of the following.

For each house he has to record the house number, the street and the number of children who live there. The house number and the street together identify the house. Each house has any number of children or none living there, and each child lives in exactly one house. For each child, Santa has to record the child's name, which identifies the child, age and whether he/she has been good or bad. For a good child Santa records the name of one person or animal the child has been good to, and for a bad child he records the name of one person or animal the child has been bad to.

All children receive a greeting card from Santa, one card per child, but only good children receive presents, at least one present per child. Each greeting card is identified by a special magic number, and each present is identified by a code and has a name.

Santa delivers the presents to the houses with the help of his active reindeers. He needs at least three active reindeers for each delivery, and needs to keep a record of which reindeer helps with the delivery of which present to which house. The active reindeers help deliver any number of presents to any one house. Santa also has some retired reindeers which do not work, but get presents themselves from Santa, one present for each retired reindeer. Each retired reindeer also has a retirement stable identified by a name.

Santa identifies each reindeer, active or retired, by a name and records its number of years of service. Each retired reindeer has at least twenty years of service behind it. The active reindeers have holiday entitlements, and Santa records the total holiday length entitlement of each one of them. No active reindeer is allowed a holiday in the last week of December.

a There are three items of data above that cannot be represented in an ER model. Which are they?

b Represent the rest of the data in an ER diagram. Be careful to represent all the existence constraints and the relationship cardinalities.

c Translate the ER model into the relational model, giving the relation schema and the primary and foreign keys. You do not need to give any foreign key rules, but for each foreign key state whether or not it can accept NULL values.

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

4 a i) Using the concept of foreign keys explain the significance of transaction atomicity in the integrity of a database.

ii) There may be high overheads associated with copying a log record to stable storage as soon as the record is created. To reduce these overheads, log records can be created and kept in volatile storage and copied to stable storage in batch. How can it be ensured that such a scheme does not lead to a violation of transaction atomicity and that all committed transactions can be performed in the database?

b i) Explain when two schedules of the same transactions are equivalent. Explain when two schedules of the same transactions are conflict equivalent.

ii) Is it possible for two schedules of the same transactions to be equivalent but not conflict equivalent? If not then give an informal proof, and if yes give an example.

c i) Explain what is meant by a serialisable schedule of transactions.

ii) Consider the following locking protocol. Shared (S) locks are acquired for reading, and exclusive (X) locks for writing. Once an object is read in a transaction the rest of the transaction is scanned and if there is no command for writing to that object the S lock is released immediately. If there is a command for writing to that object in the rest of transaction then the S lock is not released and is updated to an X lock when the writing occurs.

Does this protocol ensure serialisability? Give a justification for your answer.

The three parts carry 25%, 45%, 30% of the marks, respectively.

End of paper