THIS PAPER IS NOT TO BE REMOVED FROM THE EXAMINATION HALLS

UNIVERSITY OF LONDON **291 0205**

BSc Examination
for External Students

COMPUTING AND INFORMATION SYSTEMS

Databases

Dateline: Tuesday 12 May 2009 : 2.30 – 5.30 pm

Duration: 3 hours

This paper is divided into two parts, **Part A** and **Part B**. Candidates should attempt **THREE** questions from Part A and **THREE** questions from Part B. Each question carries 25 marks and full marks can be obtained for complete answers to **SIX** questions. Questions involving a description or explanations should, wherever possible, be accompanied by an appropriate example.

Candidates should write their answers to Part A and Part B in separate answer booklets.

Calculators are NOT allowed.

PART A

1. (a) What is meant by saying that the complexity of insertion sort is $O(n^2)$?

(3 *marks*)

(b) What is the difference between an internal sorting algorithm and an external sorting algorithm? What factors determine which type of sorting algorithm would be appropriate in an application?

(3 *marks*)

(c) Why is it not sensible to compare an internal sorting algorithm and an external sorting algorithm by looking at their "big-O" values?

(4 *marks*)

(d) Give full details of one external sorting algorithm.

(10 *marks*)

(e) In an external sorting algorithm it is generally accepted that the longer the partitions produced the more efficient the algorithm. Explain why this should be so.

(5 *marks*)

2. (a) Describe the general idea of indexing a file and how an index would be used to locate a record in a file given the value of its key field.

(8 *marks*)

(b) What factors determine the size of the index blocks of an indexed file?

(3 *marks*)

(c) Describe the structure and operational features of a B-tree index. In particular, say what the B stands for.

(8 *marks*)

(d) How does a B^+-tree index differ from a B-tree index? What are the advantages and disadvantages of using a B^+-tree as opposed to a static index?

(6 *marks*)

3. (a) In the context of record storage and retrieval define what is meant by hashing.

(4 *marks*)

(b) Under what circumstances would hashing a file be considered?

(2 *marks*)

(c) Describe a hashing technique which involves the use of buckets. Carefully explain the insertion procedure and what happens if the home bucket is full and the retrieval procedure when the home bucket is found not to contain the required record.

(9 *marks*)

(d) What are primary and secondary clustering in the context of hashing? Describe procedures which may avoid primary and secondary clustering.

(6 *marks*)

(e) Is it possible to both index and hash a file at the same time? Justify your answer.

(4 *marks*)

4. (a) What is a threaded file? What is an inverted file? For each give a description and an illustrative example.

(11 *marks*)

(b) Is it possible for a file to be hashed, indexed and inverted all at the same time? If you believe it cannot be achieved explain the limitations that prevent it. If you believe it can be achieved, explain how and detail the difficulties that would be involved in maintaining the file.

(14 *marks*)

PART B

5. **(a)** Write brief definitions of **five** of the following terms, illustrating your answer with reference to an example where possible.

(1) Determinant
(2) Entity integrity
(3) Degree
(4) Functional dependency
(5) Candidate key
(6) Composite (or concatenated) Key

[10 marks]

(b) The Ferrous Metals Company organizes the training of its new workers in the following way. Each Trainee-Worker is apprenticed to a Master-Craftsman, who will be in charge of several apprentice Trainee-Workers. Each Trainee-Worker will be assigned to work on a single Training Project. There will be several Trainee-Workers working on a given Training Project at any one time. Each Training-Project will be supervised by a single Project Manager, who may be in charge of several Training Projects at any given time.

(1) Draw an Entity-Relationship Diagram to illustrate the relationship among **Trainee-Workers**, **Master-Craftsmen**, **Training Projects**, and **Project Managers**.

[9 marks]

(2) Prepare a relational schema which can record the *relationships* illustrated by your answer to part **(1).** Assume that Trainee-Workers, Master-Craftsmen, and Project Managers are identified by *employee-numbers*; and that Training Projects are identified by *Project-Codes*. Be sure to indicate the **key** of each relation.

[6 marks]

6. **(a)** The Temerario financial advice consultancy employs several dozen analysts whose job it is to monitor the progress of major firms listed on the stock market. Information about the responsibility of each analyst is held in a relation called **FA**, whose primary key is the attribute FIRM

A partial "snapshot" of this relation might look like the following:

FIRM	ANALYST
Glaxo-Wellcome	Larry Ladron
Biotrace	Roger Ratero
Health Services	Larry Ladron
Winston Marketing	Curt Bolsas
Sandemann	Curt Bolsas

Construct expressions in relational algebra to answer the following queries:

(1) List the names of all of our analysts.

(2) List the names of all firms which are analysed by Larry Ladron.

(3) List the names of all firms which are analysed by Larry Ladron or Curt Bolsas.

(4) List the names of all firms which are **not** analysed by Larry Ladron or Curt Bolsas.

[8 marks]

(b) The Investment Consultancy referred to in **(a)** has the following additional relation named **FS** in its database. This relation lists all of the firms listed on the stock exchange, and the industrial sector under which they are classified. A firm will be classified under only one sector.

A partial "snapshot" of this relation might look like the following:

FIRM	SECTOR
Glaxo-Wellcome	Health Care
Grannie's Pizzas	Retail Food
Gundersens	Automotive
Harrod's	Retail
Henderson	Health Care

(question continues on next page)

You will need to use the relation **FA** mentioned in Part **A** as well as **FS** (the relation shown in this question).

(1) Which attribute or attributes make up the **key** of **FS**?

Construct expressions in **either** relational algebra **or** SQL to answer the following queries.

(2) Which analysts cover firms in the Health Care sector?

(3) Which sectors are currently not covered by any analyst?

[12 marks]

(c) Someone has proposed that the relations **FA** and **FS** could be combined into one relation. Could this be done, retaining all information currently held by the two relations, and would it be a good idea to do it? Briefly explain your answer.

[5 marks]

7. An aircraft maintenance company holds information on the servicing of its clients' airplanes in a single table. The table records which mechanic has worked on which aircraft, and the date that the mechanic worked on it. Mechanics are identified by their employee numbers (such as M0001), and aircraft are identified by unique registration numbers, such as B832. This table also records, for each aircraft, what model it is, and how many passengers that model holds (its 'capacity').

A partial "snapshot" of this relation might look like the following:

MECHANIC	AIRCRAFT	DATE	MODEL	CAPACITY
M0001	B832	21-12-98	Apache	4
M0001	C720	21-12-98	Aerostar	6
M0001	B502	21-12-98	Cub	2
M5672	B832	03-06-99	Apache	4
M8341	D924	12-07-99	Aerostar	6
M8341	B432	13-07-99	Hawk	2

(a) State two problems which could arise, in the normal course of adding to, modifying and deleting from this table. Use an example in each case to illustrate your point.

[8 marks]

(b) Re-cast the table above into a set of equivalent tables in Third Normal Form.

[12 marks]

(c) For each model of aircraft, there are certain pilots who are qualified to fly it, and also certain mechanics who are qualified to perform maintenance procedures on it. This information is currently stored in the following relation, whose key is MODEL+QUALIFIED PILOT + QUALIFIED MECHANIC. For example, the Aerostar can be piloted by three different pilots, and there are two mechanics qualified to work on it.

MODEL	QUALIFIED PILOT	QUALIFIED MECHANIC
Aerostar	P2999	M0001
Aerostar	P3428	M8341
Aerostar	P4451	M8341
Apache	P1045	M0001
Apache	P1045	M5672
Cub	P2999	M0001

Is this table fully normalized? If so, state why it is. If not, propose an alternate schema which can hold the same information but which is fully normalized. **[5 marks]**

8. **(a)** An aircraft repair depot records the availability of parts from competing suppliers, including the price each supplier charges for that part. This information is held in the relation **PSP**.

PSP

KEY: PART+SUPPLIER
MEANING: The given PART *is available from* the given SUPPLIER *at* the given PRICE.

PART	SUPPLIER	PRICE
HP-234/K	BB009	34.56
HP-234/K	RK324	33.89
LK-887/A	BB009	103.45
MA-991/B	PR984	67.87
MA-991/B	RK324	66.78
MB-385/U	PP883	9.56

The repair depot also records the location of each supplier's warehouses, in case an aircraft needs emergency repairs when it is away from its home base. This information is kept in the relation **SW**.

SW

KEY: SUPPLIER + LOCATION
MEANING: The given SUPPLIER *has a warehouse in* the given LOCATION.

SUPPLIER	LOCATION
BB009	Kansas City, USA
BB009	Boston, USA
BB009	Tokyo, Japan
CA001	Boston, USA
CA001	London, UK
PR984	Hong Kong, PRC
PR984	Taiwan, ROC
PR984	Seoul, ROK
PR984	Hanoi, Vietnam

(question continues on next page)

Consider the following relational algebra operations on the relations **PSP**, **SW** and relations derived from them. (In these operations, *RESTRICT* has the same meaning as *SELECT* in older textbooks. The *RESTRICT* of a relation yields a new relation with the same attributes as the old relation, but with only those tuples specified by the accompanying expression. If we *RESTRICT* **SW** *WHERE* LOCATION=London, we get a new relation with the same attributes as **SW**, but with only those tuples where the LOCATION is London. If we then *PROJECT* this new relation *OVER* SUPPLIER, we get a relation whose *meaning* is all SUPPLIERs with a LOCATION in London. Note that in SQL, the operations of *RESTRICT* and *PROJECT* are combined in the expression SELECT: using SQL, we would find all SUPPLIERs with a London LOCATION by the following expression: SELECT SUPPLIER FROM **SW** WHERE LOCATION=London.)

R1 ← *PROJECT* **PSP** *OVER* PART

R2 ← *RESTRICT* **PSP** *WHERE* SUPPLIER = BH005
R3 ← *PROJECT* **R2** *OVER* PART

R4 ← *RESTRICT* **PSP** *WHERE* SUPPLIER = AK047
R5 ← *PROJECT* **R4** *OVER* PART

R6 ← *UNION* (**R3**, **R5**)
R7 ← *INTERSECT* (**R3**, **R5**)
R8 ← *DIFFERENCE* (**R3**, **R5**)

R9 ← *RESTRICT* **PSP** *WHERE* SUPPLIER <> BN988
R10 ← *PROJECT* **R9** *OVER* PART

R11 ← *DIFFERENCE* (**R1**, **R3**)

R12 ← *RESTRICT* **SW** *WHERE* LOCATION = London, UK
R13 ← *JOIN* **R12**, **PSP** *ON* SUPPLIER
R14 ← *PROJECT* **R13** *OVER* PART

R15 ← *DIFFERENCE* (**R1**, **R10**)

Give the **meanings** of

(1) R3
(2) R6
(3) R7
(4) R8
(5) R10
(6) R11
(7) R14
(8) R15

Note: a 'meaning' is not a re-statement of the relational algebra expression, nor is it a list of tuples derived from an expression. It is a statement in natural language of the information the derived relation holds.

[16 marks]

(question continues on next page)

(b) A Beauty College holds information on its students and courses in the following tables: STUDENT-MASTER, COURSE-MASTER, and ENROLMENTS.

STUDENT-MASTER

KEY: STUDENT-NUMBER
MEANING: Details of each student. Each student will be represented in this table.

STUDENT-NUMBER	SURNAME	FIRST NAME
S001	Soong	Yuan
S004	Hong	Park
S009	Ismail	Ahmed
S010	Flanders	Ned
S013	Patel	Anwar

COURSE-MASTER

KEY: COURSE-CODE
MEANING: Details of each subject. Each subject will be represented in this table.

COURSE-CODE	SUBJECTNAME	LENGTH
I01	Introduction to Hair Maintenance	10 sessions
I02	Financial Record Keeping	15 sessions
B01	Diseases of the Scalp	10 sessions
B03	Basic Hair-dyeing	20 sessions

ENROLMENTS

KEY: STUDENT + COURSE
MEANING: Details of each which student is studying which subject.

STUDENT	COURSE	START-DATE
S134	I01	14/06/2008
S134	B01	28/06/2008
S203	B02	05/05/2008
S286	I01	14/06/2008

(question continues on next page)

Your answers to the following questions should be illustrated with examples taken from the relations shown above.

(1) What type of integrity constraint will be automatically enforced by the relational software when data is entered into these tables, regardless of decisions made by the database designer?

[3 marks]

(2) Comment on two other types of integrity constraint which the database designer would want to incorporate into this database when these tables are created.

[6 marks]

END OF EXAMINATION