UNIVERSITY OF LONDON
BSc EXAMINATION 2008
for External Students
COMPUTING AND INFORMATION SYSTEMS
205 Databases

291 0205

Duration: 3 hours

Date and Time: Tuesday 6 May 2008 : 2.30 – 5.30 pm

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 questions carries 25 marks and full marks can be obtained for complete answers to SIX questions. Questions involving a description or explanation should, wherever possible, be accompanied by an appropriate example.

WRITE YOUR ANSWERS TO SECTION A AND SECTION B IN SEPARATE ANSWER BOOKS

A hand held calculator may be used when answering questions on this paper but it must not be pre-programmed or able to display graphics, text or algebraic equations. The make and type of machine must be stated clearly on the front cover of the answer book.

THIS EXAMINATION PAPER MUST NOT BE REMOVED FROM THE EXAMINATION ROOM

PART A

1. (a) What is the difference between internal and external sorting? What factors would determine whether an internal or external sorting technique should be used in a particular situation?

[5]

(b) In the context of external sorting, what is a partition? Describe three techniques for generating sorted partitions that may be used in an external sort routine. Comment on the advantages and disadvantages of each technique.

[8]

(c) It is stated that in an external sorting routine it is more efficient to produce a small number of long partitions rather than a larger number of short partitions. Why is this so?

[7]

(d) Both quicksort and an external sort have a complexity of O(n log n). What does this mean? If you had a choice of using quicksort or an external sort in an application how would you decide which one to use?

[5]

2. (a) In the context of file storage and retrieval, what is an index and what is its purpose?

[5]

(b) Describe the structure of a static index. In particular, describe what is stored in each node, how the nodes are linked and how the index is used in practice.

[11]

(c) What is the difference between a logical and a physical record? What would be the difference in structure and operation between an index of physical records as opposed to an index of logical records? Briefly comment on the advantages and disadvantages of using physical records and using logical records.

[9]

3. (a) In the context of hashing, define the following terms.

Bucket
Hash function
Collision
Linear probe
Primary clustering
Secondary clustering

[10]

(b) Describe techniques for overcoming primary and secondary clustering.

[5]

(c) The key field in a file contains an eight digit number and it is required to hash the file using the value in this field. A commonly described, but rarely used, hashing technique is to add up the eight individual digits in the number. Why is this a poor hashing technique?

[5]

(d) What is meant by packing density when referring to a hashed file? Packing density should either be less than 70% or 100%. Under what circumstances would each packing density be used? Briefly justify your answers.

[5]

4. (a) What is meant by a threaded file? **[3]**

(b) What are the principal advantages of a threaded file and in what sort of applications would its use be considered?

[4]

(c) How does the creation of a threaded file affect indexing the file by other methods in addition to threading? Can a threaded file also be hashed?

[8]

(d) What is an inverted file and what are the benefits of creating an inverted file?

[6]

(e) Suggest a suitable storage and retrieval method for files used in each of the following situations.

(i) A stock market system for tracking price changes of a large number of shares.
(ii) An encyclopaedia on CD-ROM.

[4]

PART B

5. (A) Greyhounds are dogs which are often raced against one another in public arenas. The *Cockney Canines* dog racetrack wishes to record information about the dogs which compete in its races. For each dog taking part in one of its races, it wishes to record which Handler is in charge of it (there will be only one), what date the race will be on, which Kennel the Dog comes from, and who is the Owner of that Kennel.

A Dog can be controlled by different Handlers on different dates, and a Handler can be in charge of more than one Dog. However, a given Dog will not race more than once on the same Date, although a Handler might be involved in more than one race on the same date, if in charge of more than one Dog racing on that date. A Dog comes from just one Kennel (although of course a Kennel will have many Dogs). A Kennel will have just one Owner, but an Owner may own more than one Kennel. The following relation has been proposed to hold the required information. (It is shown here with some sample entries.)

Dog	Date	Handler	Kennel	Owner
White Fang	23.03.2000	Jake Fast	Paradise Acres	Major Smith
Fidelius	23.03.2000	Martin Gray	Dog's Heaven	Nell Gwyn
Rover	23.03.2000	Donald Moore	Paradise Acres	Major Smith
White Fang	19.05.2000	Steve Fry	Paradise Acres	Major Smith
Billy Barker	19.05.2000	Jake Fast	Dog's Heaven	Nell Gwyn
Everfast	19.05.2000	Martin Gray	Whipley Manor	Jeffrey Archer
White Fang	21.06.2000	Jake Fast	Paradise Acres	Major Smith
Steeleye	21.06.2000	Steve Fry	Canus Domus	Major Smith
Rover	21.06.2000	Donald Moore	Paradise Acres	Major Smith
Slicky	21.06.2000	George Jeffries	Canus Domus	Major Smith

Propose a set of equivalent relations which will hold the same information as this relation, but which will be in Boyce-Codd Normal Form.

[9]

(B) A specialist antique broker brings overseas clients to London and takes them on conducted shopping tour of up-market antique dealers. Different clients are interested in different categories of antiques, and the broker records these in a relation called **CI**, shown here with a sample of tuples.

(question continues on next page)

RELATION: **CI**
KEY: CLIENT + CATEGORY
MEANING: The given CLIENT *is interested in* the given CATEGORY.

CLIENT	CATEGORY
C937	Georgian Furniture
C937	Ming Vases
C582	Ming Vases
C994	Ming Vases
C994	Scientific Instruments

Dealers whom the broker may visit with his clients, and the dealers' specialities, are shown in the relation **DS**.

RELATION: **DS**
KEY: CATEGORY + DEALER
MEANING: the given CATEGORY *is offered by* the given DEALER.

CATEGORY	DEALER
Netsuke	Memories
Toby Mugs	Ancient Items
Georgian Furniture	Memories
Georgian Furniture	Henning's Antiques, Ltd
Georgian Furniture	The Treasure Trove
Ming Vases	Memories
Ming Vases	Henning's Antiques, Ltd
Ming Vases	Ancient Items
Scientific Instruments	Memories

Consider the following relational algebra operations on the relations **CI**, **DS** 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 Boolean expression. If we *RESTRICT* **CI** *WHERE* CLIENT = C994, we

(question continues on next page)

get a new relation with the same attributes as **CI**, but with only those tuples where the CLIENT number is C994. If we then *PROJECT* this new relation *OVER* CATEGORY, we get a relation whose *meaning* is all CATEGORYs CLIENT C994 is interested in. Note that in SQL, the operations of *RESTRICT* and *PROJECT* are combined in the expression SELECT: using SQL, we would find all CATEGORYs of interest to C994 with the expression SELECT CATEGORY FROM **CI** WHERE CLIENT=C994.)

R1 ← *PROJECT* **CI** *OVER* CLIENT

R2 ← *RESTRICT* **CI** *WHERE* CATEGORY = Georgian Furniture
R3 ← *PROJECT* **R2** *OVER* CLIENT

R4 ← *RESTRICT* **CI** *WHERE* CATEGORY = Scientific Instruments
R5 ← *PROJECT* **R4** *OVER* CLIENT

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

R9 ← *RESTRICT* **CI** *WHERE* CATEGORY <> Georgian Furniture
R10 ← *PROJECT* **R9** *OVER* CLIENT

R11 ← *DIFFERENCE* (**R1**, **R3**)

R12 ← *RESTRICT* **DS** *WHERE* DEALER = Ancient Items
R13 ← *JOIN* **R12**, **CI** *ON* CATEGORY
R14 ← *PROJECT* **R13** *OVER* CLIENT

R15 ← *DIFFERENCE* (**R1**, **R10**)

Give the **meanings** of

(1) R1
(2) R3
(3) R6
(4) R7
(5) R8
(6) R10
(7) R11
(8) R14
(9) R15

Note: a 'meaning' is not a description of how the computer will carry out the query, nor is it a list of the tuples which match the query. It is a statement in natural language of the information that the derived relation holds.

[16]

6. (A). A company which sells many different **styles** of fitted kitchen kits (such as Swedish Modern, or Old Farmhouse) gets its **components** (such as counter tops, sinks, or shelves) from **wholesalers**, and assembles several different styles of fitted kitchen kits from them. These kits are then sold to **retailers** who market them to the general public. In order not to become dependent on any one wholesaler, the company always makes sure that any given component-type is supplied by at least two separate wholesalers. Each component-type can be used in the assembly of several different styles of kitchen kit. A particular style of kitchen is made up of many different types of component. A given retailer can receive and re-sell many different styles of kitchen kit. No retailer has a monopoly on re-selling any given style of kit.

(1) Draw an Entity-Relationship Diagram to illustrate the relationship among **kitchen kit styles**, **component-types**, **component-wholesalers**, and **retailers**.

[9]

(2) Prepare a relational schema which can record the *relationships* in part **(1).** Assume that kitchen kit styles are identified by *style-codes*, wholesalers by *wholesaler-numbers*, retailers by *retailer-codes*, and component-types by *component-codes*. Be sure to indicate the **key** of each relation. You need not record any information not asked for.

[6]

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

(1) Foreign Key
(2) View
(3) Candidate key
(4) Degree
(5) Primary key
(6) Functional dependency
(7) Attribute
(8) Domain
(9) Cardinality

[10]

7. (A) A community Hobby Centre offers instruction in various hobbies, and employs instructors to teach them. The database of this Hobby Centre includes a relation, called **HobbyTeach**, which holds information about which instructor is qualified to teach which Hobby. (A Hobby can be taught by more than one instructor, and an instructor can be qualified to teach more than one Hobby.)

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

Instructor	Hobby
Bill Logan	Flower Arranging
Bill Logan	Model Aircraft
Adaire Hannah	Model Aircraft
Samuel Ting	Electronic Kit Construction
Thomas Burt	Mirror Grinding
Thomas Burt	Astronomy
Mary Elliott	Flower Arranging
Mary Elliott	Ham Radio

Construct expressions in relational algebra to answer the following queries:

(1) List all the hobbies which can be taught.

(2) What are the hobbies which Thomas Burt can teach?

(3) Which instructors are qualified to teach Ham Radio **or** Model Aircraft?

(4) Who are the instructors who are qualified to teach Mirror Grinding **and** Astronomy?

(5) Which instructors are ***not*** qualified to teach Flower Arranging?

[15]

(question continues on next page)

(B) The Hobby Centre referred to in Part **A** has the following additional relation in its database:

HobbyCost, which lists all of the Hobbies which the Centre has the facilities and equipment to teach, and the weekly charge made for teaching that Hobby. (Some Hobbies in this list may not be currently taught because the Centre has no instructor qualified to teach them.) A Hobby has only one weekly charge associated with it.

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

Hobby	Weekly-Charge
Astronomy	£2.00
Woodcarving	£3.50
Model Aircraft	£3.00
Model Ship Building	£3.50.
Furniture Making	£12.00.
Flower Arranging	£7.50
Ham Radio	£4.00

You will need to use the relation **HobbyTeach** mentioned in Part **A** as well as the relation shown in this part of the question.

(1) Which attribute or attributes make up the **keys** of **HobbyTeach** and **HobbyCost**?

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

(2) Which Hobbies (if any) are currently *not* taught by any instructor?

(3) Which instructors can teach at least one Hobby whose weekly charge is greater than £5.00 per hour?

[10]

8. An Art Class Centre holds annual exhibits of some of its work. Each year, it creates a table to record information about the exhibits for that year. The table includes the Art Style, where the exhibition will be held, the name of each student (called an 'exhibitor') entering for that Art Style, and each exhibitor's birthday.

Exhibitors (who must be at least 18 years old by the day the exhibit begins, for insurance reasons) may take part in exhibiting more than one Art Style. You may assume in answering this question that an exhibitor's full name is unique: that is, there will never be two exhibitors with identical names.

A given Art Style exhibition is always held in one location only. For example, the Still Life display will take place only in Room 12 this year.

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

Art Style	Location	Exhibitor	Birthdate
Still Life	Room 12	Edward Hubble	2 Jan 68
Still Life	Room 12	Percival Lowell	24 Apr 70
Still Life	Room 12	Fred Hoyle	17 Dec 67
Sculpture	Main Hall	Otto Lillienthal	14 Aug 67
Sculpture	Main Hall	Percival Lowell	24 Apr 70
Classical Scenes	Basement	Michael Angelo	30 May 68
Classical Scenes	Basement	Vanessa Bluecoat	21 Aug 71
Battle Scenes	Room 10	Michael Angelo	30 May 68
Battle Scenes	Room 10	John Marconi	27 Dec 73

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

[12]

(question continues on next page)

(B) There is a relation ArtClassCost which is a "master relation" for Art Classes. Before an Art class can be offered at the Art Class Centre, its name and cost must be entered in this relation. With that in mind, describe briefly, using the relations you proposed in part **A** to illustrate your answers, the following integrity constraints which you would expect to be enforced on the relations you proposed in part **A**.

(1) Attribute integrity.
(2) Referential integrity.
(3) Entity integrity.

[9]

(C) The Art Class Centre wants to record, for each Art Class, which books and magazines would be useful for students of that Art Class to have. A given Art Class may have several useful books and magazines, and a given book or magazine may be useful in learning more than one Art Class. (For example, there are three books and two magazines which are useful for learning about Battle Scenes.) The following relation has been proposed, shown here with some sample entries:

Hobby	Book	Magazine
Battle Scenes	Pictures of War	Battle Painting
Battle Scenes	Fighting with the Brush	War Paint Monthly
Battle Scenes	Composition of Troops	
Classical Scenes	Timeless Painting	Old Painters Monthly
Classical Scenes		Master Pictures
Still Life	Technical Maths	Old Painters Monthly

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.

[4]

END OF EXAMINATION