2910205 Databases

Examination paper: Zones A and B

Time allowed: three hours

This paper is divided into two parts. Candidates should attempt **three** questions from Part A and **three** questions from Part B. All questions carry equal 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.

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.

PART A

1. (a) What is the basic difference between an internal and an external sorting method?

[2]

(b) A colleague claims that Quicksort is the "best" internal sorting algorithm available. How might this claim be justified?

[4]

(c) Describe fully the external sorting algorithm which uses natural selection for creating partitions.

[10]

(d) Three methods of performing an external sort are called internal sorting, replacement selection and natural selection. What are the respective advantages and disadvantages of each of these methods?

[4]

(e) A random access file is indexed using a B^+-tree. Why is there no need for records in the file or the keys in the index to be physically stored in order? Would there be any advantage if they were sorted into order?

[5]

2. (a) Describe in general terms the process of hashing in the context of file storage.

[3]

(b) An employee's identity number is composed of six symbols which may be either letters or numbers. Describe a possible hashing algorithm which uses the identity for allocating employee records into a set of N buckets. Carefully explain any consideration you would take into account in designing your hashing algorithm.

[5]

(c) It is recommended that hashed files operate with a packing density of about 70%. Why is this so? What would you do if you found that the hashed file you were working with had a packing density of 90%?

[7]

(d) Two problems associated with hashed files are termed primary and secondary clustering. Explain what is meant by each of these terms.

[5]

(e) Two methods of handling overflow in a hashed file are open and closed addressing. Explain what each is and whether it is susceptible to primary or secondary clustering.

[5]

3. (a) You are distributing to customers a database of information on CD-ROM. You need to provide and efficient search and retrieval program to access data on the CD-ROM. Discuss the circumstances under which you would use hashing, static indexes or B-trees.

[4]

(b) Records in a file are accessed by using a multi-level index. Explain very carefully the physical process by which a record is retrieved. Your answer should clearly define what a key is, what is actually stored in an index record, where it is stored, how it used and what is physically pointed at by index records.

[8]

(c) Outline the structure of a B-tree index. Give an example of a B-tree index and demonstrate how insertion of keys is achieved. Also explain why a B-tree is said to be "self-balancing".

[9]

(d) How does a B^+-tree differ from a B-tree? What are the advantages and disadvantages of using a B^+-tree in preference to a B-tree?

[4]

4. (a) A museum maintains an electronic catalogue of all the items in its collection. For each item, there is a record which contains the following information.

(i) A unique catalogue number.
(ii) A description of the item.
(iii) The location of the item in the museum.
(iv) A single keyword description of the item, (e.g. VASE, PICTURE, STATUE, etc.)
(v) The country of origin, (e.g. EGYPT, ENGLAND, CHINA, etc.)
(vi) A number indicating the century in which the item was made, (e.g. 19 for 19th. century, 5 for 5th. century, -2 for 2nd. century BC, etc.)

This information is also printed on a card which is physically attached or kept with the item itself.
The catalogue for the museum is kept in one file. Records will need to be retrieved from the file in several ways and for several purposes. The system must be capable of the following functions.

(i) When the location of the item is changed. the record will need to be updated and the card for the item reprinted.
(ii) Lists of all items from a particular country, (e.g. a list of all Egyptian items,) or of the same type, (e.g. all pictures,) or from the same century, (e.g. all 5th. century items) will need to be produced.
(iii) Lists of all items with two or more attributes will need to be able to be produced, (e.g. a list of all 5th. century Italian items, a list of all 2nd. century BC Greek vases, etc.)

Describe a suitable record structure and procedures so that all the above uses of the catalogue can be achieved. (Note, all operations must be capable of operating simultaneously on just one file.)

[17]

Outline how your systems would produce lists of items such as the following

All Greek vases.
All 5th. and 6th. century Italian statues.

Describe any feature you would add to your system to make the production of such lists efficient.

[4]

Outline what procedures would be required for the insertion and deletion of items from the catalogue.

[4]

PART B

5. Considering the relation PART in the following table:

P#	PNAME	WEIGHT	MADE
P1	Nut	12	London
P2	Bolt	17	Paris
P3	Screw	17	Rome
P4	Screw	14	London
P5	Cam	12	Paris
P6	Cog	19	London

(a) Give an example from the above table to explain the following terms. [4]
(i) tuple
(ii) entity
(iii) attribute
(iv) entity-type

(b) Which attribute (attributes) cannot be the prime key in the relation? [3]

(c) Is this relation in Third Normal Form? If yes, give your reason. If no, recast it as a set of relations in Third Normal Form. [8]

(d) Under what circumstances might this relation contain *null* values? Under what circumstances would null values not be allowed? Give your reasons. [5]

(e) List *three* functional dependencies that hold in the relation PART. [5]

6. Scenario:

A library consists of four departments: reader service, book collection, security and computer support. Employees normally work in their own department. However, some employees in the reader service department may work as security guards in the security department and some employees in the book collection department may work in the reader service department if necessary. We need to store, for each employee, the ID#, name, sex, birth-date, salary, telephone-number (telephone-numbers), the department (departments) in which he may work and the locations of the departments.

Some employees are members of the coffee club of the library and the payment is collected regularly. We need to keep track of the names of the coffee-club-members and the outstanding-pay of each member.

Some employees are married couples and are entitled to housing benefit. We need to keep a record of one name for each married couple that is entitled to housing benefit, the amount of the housing-benefit-payment and the partner's information (for example, name and the working department(s)).

(a) Identify at least four entities in the above scenario. [5]

(b) Draw an Entity-Relationship Diagram to illustrate the relationship among the entities. [10]

(c) Decide entity types for each entity, and relationship types for each relationship. Prepare a normalised relational schema for a preliminary database. [10]

7. Three relations about the computer usage in a small research group are STUDENT, COMPUTER and USAGE as follows.

STUDENT

SName	FName	Age	Sex
Smith	John	22	M
Wong	Franklin	36	M
Evans	Alicia	24	F
English	Joyce	41	F
Smith	James	24	M

COMPUTER

CName	CType
AAA	Workstation
BBB	IPC
CCC	PC

USAGE

SName	FName	Date (ddmmyy)	From	To	CName
Wong	Franklin	231196	900	1200	CCC
Smith	John	231196	1145	1700	BBB
Evans	Alicia	231196	1310	1800	CCC
Smith	James	231196	1310	1520	AAA
Smith	John	231196	1925	2200	BBB
Wong	Franklin	241196	700	1030	BBB
Evans	Alicia	241196	930	1100	AAA
English	Joyce	241196	1900	100	AAA

(a) Give a brief definition and an example of each of the following terms:
(1) relation [2]
(2) domain [2]

(b) Is relation USAGE in Third Normal Form? If yes, explain why it is. If not, recast it as a set of relations in Third Normal Form. [4]

(c) For each relation, indicate whether it is an *entity relation* or a *relationship relation*. [4]

Using the relational operators, e.g. PROJECT, SELECT, JOIN, UNION, DIFFERENCE and INTERSECT, derive expressions in the relational algebra to generate new relations as follows:
(i) A full name list of male students in MALELIST; [3]
(ii) A first name list of the students who are under thirty years of age in FIRSTNAMEYOUNG; [5]
(iii) A full name list of the students who used the workstation during 900-1200 every day in STUDENTUSEDA. [5]

8. Consider a taxi control scenario as follows:

Several taxi control desks have been set up spreading over a city. Each control desk receives telephone requests for a taxi and sends taxis to the customers. Each taxi is controlled by its unique individual control desk.

The requests from customers are kept in the database until the customer has been served. In other words, a request record including the customer's name will be removed from the database after a taxi is sent to the location of the call. Note this implies that a customer may ring the control desk and book taxies many times, but only one call is answered by the control desk at any given moment.

Every telephone request that exists in the database is bound to the control desk that responded to it. One does not associate any call with more customers than the caller. Note that there may be control desks that are not, at a given moment, in charge of any call. This characteristic applies also to taxis (some of them may not be, at a given moment, on their way to answer a call) and calls (some of them may not yet have had a taxi routed to them by a command of the control desk.).

Once it finds a taxi available, the control desk will send the taxi to the customer. The same taxi can be routed to serve many customers at their call locations by a single command of the control desk.

(a) Draw an entity-relationship diagram for the taxi control system. [12]

(b) Give a normalised relational schema from the entity-relationship diagram you drew in (a).

Suggest a suitable primary key for each relation in your schema. [13]