

Birkbeck
(University of London)

MSc and MRes Examination for Internal Students

MSc in E-Business

MSc in Advanced Information Systems

MSc in Information and Web Technologies

MRes in Computer Science

Department of Computer Science and Information Systems

Advances in Data Management (COIY025H7)

CREDIT VALUE: 15 credits

13 May 2010 (2.30pm – 4.30pm)

There are six questions on this paper. Candidates should attempt any FOUR of them. Calculators are not permitted.

1. (a) Describe the structure of a B+ tree. Explain how the insertion and deletion operations leave a B+ tree balanced. (10 marks)

- (b) Consider a B+ tree of order 2 which consists of a root node:

20, 25, 30, 35

and five leaf nodes:

15, 16, 17, 18 20, 21, 22, 23 26, 27 31, 32 36, 37

Describe what happens to this B+ tree during the following series of operations: insert 19, insert 24, delete 27, delete 20, delete 19.

(15 marks)

(25 marks)

2. (a) Describe what is the purpose of the *transaction manager* of a DBMS and what is meant by the *ACID* properties of transactions. (8 marks)

- (b) Describe the *strict two-phase locking* method of concurrency control in DBMSs. (9 marks)

- (c) Some DBMSs allow programmers to control the isolation level of transactions. What is meant by *isolation level* and why is this kind of control desirable? What isolation levels are supported in SQL? (8 marks)

(8 marks)

(25 marks)

3. (a) Briefly describe the main steps involved in the optimisation of SQL queries. (7 marks)

- (b) A relational database contains two tables:

DVDs(*DVDNumber*, *title*, *rating*)

DVDRentals(*DVDNumber*, *fromDate*, *returnDate*, *customerID*, *rentalPrice*)

which contain information about the DVDs stocked by a DVD Rentals company and the DVDs that have been rented by customers.

Each DVD is uniquely identified by its *DVDNumber*, and it also has a *title* and a *rating* (the possible values of the *rating* attribute are 'U', 'PG', '12', '15', '18').

Customers are uniquely identified by their *customerID*. The *DVDRentals* table contains a row for each DVD that has been rented. The combination of *DVDNumber* and *fromDate* uniquely identifies rows in in this table. Also stored is the *customerID* of the customer who rented the DVD, the rental price paid (in pounds), and the date that the DVD was returned.

Consider the following query:

```
SELECT  R.customerID
FROM    DVDs D, DVDRentals R
WHERE   D.rating = 'U' AND D.DVDNumber = R.DVDNumber
```

- (i) Translate this query into the relational algebra, using the operators π , σ and \times . (4 marks)

- (ii) Write down 2 possible query plans for this query, assuming that there is a clustered hash index on the *rating* column of the *DVDs* table. (8 marks)

(8 marks)

- (iii) Estimate the cost of each of your 2 query plans, assuming that:
- The *DVDs* table consists of 100 data pages, with 50 records per page.
 - One fifth of DVDs have *rating* = 'U'.
 - The *DVDRentals* table consists of 10,000 data pages, with 50 records per page.
- State any further assumptions that you make.

(6 marks)

(25 marks)

4. A relational database contains two tables

DVDs(*DVDNumber*, *title*, *rating*)

DVDRentals(*DVDNumber*, *fromDate*, *returnDate*, *customerID*, *rentalPrice*)

The information held in these tables is as described in Question 3.

- (a) Write a PL/SQL function that takes a *rating* value and a *customerID* and returns the number of times this customer has rented a DVD with this rating.
- (b) What is the purpose of *triggers* in databases, and what kinds of application functionalities might be implemented using triggers?
- (c) Write an Oracle BEFORE trigger which whenever a new row is inserted into the *DVDRentals* table first checks if the *customerID* of the row is also present in the table *NewCustomers*(*customerID*), in which case the rental price of the row to be inserted is reduced by 1 pound.

(10 marks)

(5 marks)

(10 marks)

(25 marks)

5. (a) Consider the following table `Albums` storing music recordings in an XML-enabled relational database system (the attribute `Album` is of type XML):

`Albums`

ProdID	Title	Album
1	Thriller	<pre><?xml version="1.0"> <music year="1982"> <artist>Michael Jackson</artist> <label>Epic Records</label> <media>record</media> <content> <track number="1"> <title> Wanna Be Startin&apos;Somethin&apos; </title> <length> <min>6</min> <sec>2</sec> </length> </track> <track number="2"> ... </track> ... </content> </music></pre>
2	Help!	<pre><?xml version="1.0"> <music year="1965"> ...</pre>

Formulate the following queries in SQL/XML. Unless specified otherwise, returning XML and/or relational data is fine.

- (i) For each album, sort the tracks according to their length (longest track first). Output the name of each track and its length. (4 marks)
- (ii) List the albums published under the two labels EMI and Sony. Output the title of each album together with the number of tracks it has. (5 marks)
- (iii) Output a relational table containing a list of all available tracks together with their lengths (in minutes and seconds). List each track only once, i.e. eliminate any duplicates.

Track	Min	Sec
Wanna Be Startin'Somethin'	6	2
...		

(6 marks)

- (b) (i) When applying the function `distinct-values` to a sequence of nodes, it returns a sequence of values rather than a sequence of nodes. Briefly explain which feature of the XQuery data model makes it difficult to eliminate duplicate values directly on a sequence of nodes. (3 marks)

(3 marks)

- (ii) How can you turn this sequence of values back into a sequence of nodes? (3 marks)

(3 marks)

- (c) What is the difference between the following two XQuery statements?

```
<album>                                document {
  <title>Load</title>                    <album>
  <artist>Metallica</artist>            <title>Load</title>
</album>                                <artist>Metallica</artist>
                                        </album>
                                        }
```

(4 marks)

(25 marks)

6. (a) A relational schema includes the two tables:

DVDs(*DVDNumber*, *title*, *rating*)

DVDRentals(*DVDNumber*, *fromDate*, *returnDate*, *customerID*, *rentalPrice*)

The information held in these tables is as described in Question 3.

What is the AutoMed schema corresponding to the *DVDRentals* table ?

(5 marks)

- (b) Write an IQL query on the AutoMed schema of *DVDRentals* which returns for each customer the DVDs that have been rented by that customer.

(5 marks)

- (c) Write an IQL query on the AutoMed schemas of the *DVDRentals* and *DVDs* tables which returns the number of rentals that have been made for each DVD that has rating 'U'.

(8 marks)

- (d) Discuss some of the challenges posed by distributed heterogeneous databases with respect to query processing.

(7 marks)

(25 marks)