UNIVERSITY OF COLOMBO, SRI LANKA

UNIVERSITY OF COLOMBO SCHOOL OF COMPUTING

DEGREE OF BACHELOR OF INFORMATION TECHNOLOGY (EXTERNAL)

Academic Year 2011/2012 –3rd Year Examination – Semester 6

IT6404 - Database Systems II

Structured Question Paper

3rd August, 2012
(TWO HOURS)

To be completed by the candidate

BIT Examination Index No: .....

Important Instructions:

- The duration of the paper is **2 (two) hours**.
- The medium of instruction and questions is English.
- This paper has **4 questions** and **14 pages**.
- **Answer all questions (**25 marks each**)**.
- **Write your answers** in English using the space provided **in this question paper**.
- Do not tear off any part of this answer book.
- Under no circumstances may this book, used or unused, be removed from the Examination Hall by a candidate.
- Note that questions appear on both sides of the paper.
  If a page is not printed, please inform the supervisor immediately.

Questions Answered

Indicate by a cross (×), (e.g. [X]) the numbers of the questions answered.

	Question numbers				
To be completed by the candidate by marking a cross (×).	1	2	3	4	
To be completed by the examiners:					

Index No: ……...

1) (a) (i) Define the following terms.
Seek Time, Rotational Delay (Latency), Transfer Time and Access Time.

(04 marks)

ANSWER IN THIS BOX
Seek Time - **Average time to move the read-write head to the correct cylinder.**
Rotational Delay - **Average time for the sector to move under the read-write head.**
Transfer Time - **Time to read a sector and transfer the data to memory.**
Access Time - **Access time = seek time + rotational delay + transfer time**

(ii) Processing time is measured in nanoseconds, milliseconds or seconds. Give the most suitable time measurement unit for Seek time, Rotational delay and CPU cycle time respectively.

(02 marks)

ANSWER IN THIS BOX	
Seek time - **milliseconds**	Rotational delay - **seconds**
CPU cycle time - **nanoseconds**	

(b) What are the differences between sequential file organization and direct file organization?

(03 marks)

ANSWER IN THIS BOX	
Sequential file organization	Direct file organization
File accessed in order, one record at a	**File is accessed in any order, by record**
time, from first to last.	**number.**
Each record can be of varying length.	**Each record must be of identical length.**

(c) Eight records are entered in the given order as Record-C, Record-G, Record-B, Record-A, Record-H, Record-E, Record-F and Record-D respectively, where A-H are key values. Following three figures represent examples of how these records are stored on disk.

Record-C
Record-G
Record-B
Record-A
Record-H
Record-E
Record-F
Record-D

Figure 1a

Record-A
Record-B
Record-C
Record-D
Record-E
Record-F
Record-G
Record-H

Figure 1b

Key	Record No
A	3
B	2
C	0
D	7
E	5
F	6
G	1
H	4

Figure 1c

(i) What type of file organisation does each figure represent?
(ii) Explain how each of these files would have been constructed.

(06 marks)

ANSWER IN THIS BOX

Figure-1a – **Serial file (Unordered file)**

Unordered files are constructed by appending new records to the end of the file.

Figure-1b – **Sequential file (Ordered file)**

Ordered files are maintained by physically rearranging the file in the order of the key values.

Figure-1c – **Index sequential file**

Index sequential file maintains a full index to locate its unordered file records. (e.g. figure 1a).

Index entries are ordered by key values and corresponding record identification is maintained as in figure 1c.

(d) Consider the following SQL query.

```
SELECT E.Ename, D.Dname, W.Salary
FROM Employee E, Department D, Worker W
WHERE E.Designation='%Engineer' AND E.DeptNo=D.DeptNo AND
      E.EmpNo=W.EmpNo AND W.Salary >=50000;
```

(i) What would the above query retrieve?

(02 marks)

ANSWER IN THIS BOX

The above query would retrieve all employees' names, department and salary where designations end with Engineer and being paid a salary of 50,000 or above.

(ii) Briefly explain how the above SQL statement would initially be represented internally using relational operators prior to the query optimisation process.

(03 marks)

ANSWER IN THIS BOX

Three relations are identified and joined either as cartesian products or using join conditions starting with the Employee relation.

This will construct the leaf nodes of the internal tree structure.

Thereafter apply the filtering (where clause) conditions and to restrict data based on designation and Salary (and join conditions if initial join was based on Cartesian products).

This will connect all leaf nodes.

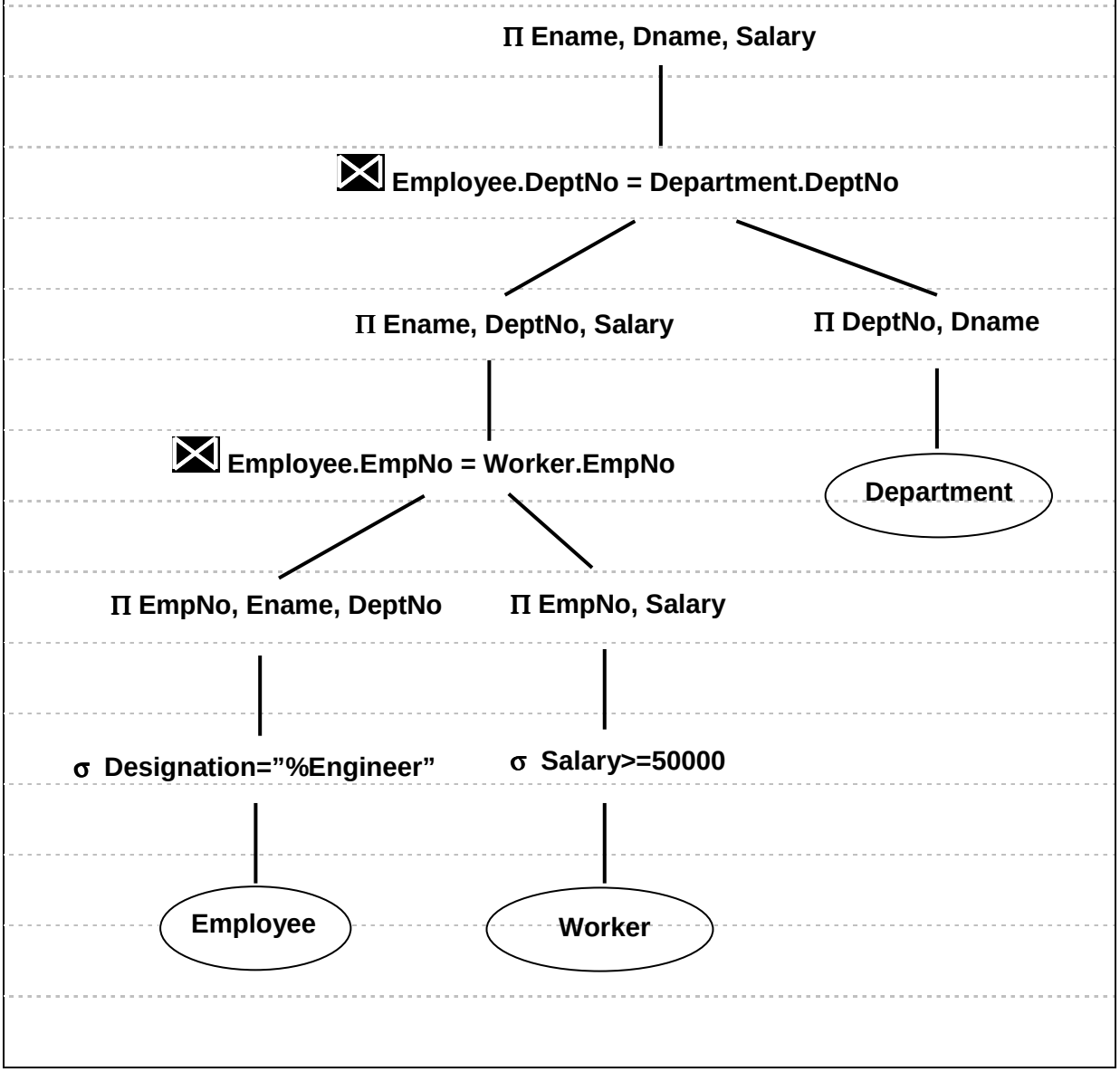
Finally the output is projected to give employee name, department and salary.

This will construct the root node.

(iii) Draw the optimized query tree for the query in (d).

(05 marks)

ANSWER IN THIS BOX



2) (a) In respect of the following statements fill in the blank with the most suitable word(s).

(i) ____**A**_____ implies that once a transaction is completed successfully, the changes made by the transaction persist in the database, even if the system fails.

(ii) The interleaving of the operations of transactions in such a way that the final output is the same as that of some serial schedule of those transactions is known as _____**B**_____.

(iii) A schedule, which is conflict equivalent to some serial schedule is known as _____**C**_____.

(iv) A situation in which failure of a single transaction leads to a series of transaction rollbacks is called _____**D**__.

(04 marks)

ANSWER IN THIS BOX

A – Durability **B – Serializable schedule**

C - Conflict serializable **D - Cascading rollback**

(b) Consider the following set of actions by three transactions T1, T2 and T3.

(i) T1:R(X), T2:W(X), T3:W(X)

(ii) T1:R(X), T2:R(X), T3:R(X)

(iii) T1:R(X), T2:W(Y), T3:R(X)

(iv) T1:R(X), T2:R(X), T3: R(Y), W(Y)

For each of the above sets of actions, indicate giving reasons, whether they are conflicting or.

(04 marks)

ANSWER IN THIS BOX

(i) is conflicting as action on X belongs to different transactions and involves write operation.

(ii) is not conflicting as no write operation is involved.

(iii) is not conflicting as write operation involved is not on the object accessed by the other two transactions.

(iv) is not conflicting as write operation involved (Y) is accessed by the same transaction.

(c) Two employees at a small merchandising company access the corporate database relation `Items(ItemtNo, Name, UnitPrice, Quantity)` at the same time. The first person is the company's Sales Manager. The second is the Accountant.

The Sales Manager wants to increase the price of a shirt sold by their firm by Rs. 10%, but is having a little trouble with the syntax of the SQL language (i.e. uses UnitPrice = UnitPrice*0.10 instead of UnitPrice = UnitedPrice*1.10). At the same time, unknown to the Sales Manager, the Accountant is trying to calculate the retail value of the current inventory to be included in a report that he volunteered to bring to the next management meeting (i.e. SUM(Quantity*UnitPrice)). Sales Manager having retrieved the new unit prices of shirts realized the error he had made and corrects it through a rollback followed by a correct execution of the update statement to increase the unit price of a shirt.

(i) Give a schedule for the above scenario identifying the full SQL statements issued by the two transactions T1 (Sales Manager) and T2 (Accountant).

(06 marks)

ANSWER IN THIS BOX

T1: UPDATE Items SET UnitPrice = UnitPrice*0.10 WHERE Name=”shirt”;

T2: SELECT SUM(Quantity*UnitPrice) FROM Items;

T1: SELECT * FROM Items WHERE Name=”shirt”;

T1: ROLLBACK;

T1: UPDATE Items SET UnitPrice = UnitPrice*1.10 WHERE Name=”shirt”;

T1: SELECT * FROM Items WHERE Name=”shirt”;

T1: COMMIT;

(ii) What is the inconsistency error illustrated in (i) above called?

(02 marks)

ANSWER IN THIS BOX

Dirty Read

(iii) What is the minimal isolation level that the Accountant should have worked with in order to avoid the above inconsistency?

(02 marks)

ANSWER IN THIS BOX

Committed Isolation

(d) Consider the three transactions T1, T2, T3.

(i) Indicate the action taken in terms of locks acquired or released (i.e. S(A) for shared lock for A) including any waiting for locks or deadlocks at each of the times t1 to t16.

(05 marks)

ANSWER IN THIS BOX

Time	T1	T2	T3	Acquire Locks	Release or Change Locks
t1	READ(A)			**S(A)**	
t2	READ(B)			**S(B)**	
t3		READ(C)		**S(C)**	
t4		WRITE(C)		**X(C)**	**S(C)**
t5	WRITE(B)			**X(B)**	**S(B)**
t6	WRITE(A)			**X(A)**	**S(A)**
t7			READ(B)	**Wait for S(B)**	
t8		READ(A)		**Wait for (S(A)**	
t9	COMMIT			**T2-S(A), T3-S(B)**	**X(A), X(B)**
t10		WRITE(A)		**X(A)**	**S(A)**
t11			READ(A)	**Wait for S(A)**	
t12		COMMIT		**T3-S(A)**	**X(A), X(C)**
t13			WRITE(A)	**X(A)**	
t14			READ(C)	**S(C)**	
t15			WRITE(C)	**X(C)**	
t16			COMMIT		**X(A), X(B), X(C)**

(ii) Is the given schedule serializable? If so what is its serial order of the transactions? If not why is the schedule non-serializable?

(02 marks)

ANSWER IN THIS BOX

Yes. Serial order is T1, T2, T3.

3) (a) Briefly describe ODMG and its' primary goal.

(03 marks)

ANSWER IN THIS BOX

ODMG – Object Data Management Group is a set of specifications to allow a developer to write portable applications for object database (ODBMS - which stores objects directly) and object-to-database mapping (ODM - which convert and store the objects in relational or other systems) products.

It facilitates persistence of object-oriented programming language objects in databases where its data schema, programming language bindings, and data manipulation and query languages are portable.

(b) (i) What are the major components of ODMG 3.0?

(02 marks)

ANSWER IN THIS BOX

Object Model

Object Query Language

Object Specification Languages

C++, Smalltalk, Java Language Binding

(ii) Briefly describe one (01) of the major components identified in (i) above.

(03 marks)

ANSWER IN THIS BOX

Object Model specifies the kinds of semantics that can be defined explicitly to an ODMS. e.g. objects, literals, types, operations, properties, attributes, relationships

Object Specification Languages are used to define the object types that conform to the ODMG Object Model.

Object Query Language (OQL) is a declarative (nonprocedural) language for query and updating.

Programming Language Binding defines the binding between the ODMG Object definition Language (ODL) and the programming language.

[one of the above]

(c) (i) Collection data types let you store and manipulate collections of data within a single row of a table. List three collection data types supported by object-relational databases.

(03 marks)

ANSWER IN THIS BOX

Any three of the following

Array **List**

Set **Multiset**

(ii) What are the other collection data types supported by the ODMG Object Model?

(02 marks)

ANSWER IN THIS BOX

Bag **Dictionary**

(d) List three (03) object-oriented structures supported by SQL.

(03 marks)

ANSWER IN THIS BOX

Any three of the following

User-defined types (ADTs, named row types, and distinct types)

Type constructors for row types and reference types
Type constructors for collection types (sets, lists, and multi sets)

User-defined functions and procedures

Support for large objects (BLOBs and CLOBs)

(e) Type constructor *row* is used to specify complex types known as user-defined types.

(i) Give an example to specify a row type for an address that would include street, city and zip.

(03 marks)

ANSWER IN THIS BOX

```
CREATE ROW TYPE Address (
 Street VARCHAR(25),
 City   VARCHAR(20),
 Zip    VARCHAR(9) );
```

(ii) Show how this row type address can be used to create a Customer table having customer-id, customer-name, customer-address and phone number.

(03 marks)

ANSWER IN THIS BOX

```
CREATE TABLE Customer (
Customer-ID          CHAR(10),
Customer-Name        VARCHAR(25),
Customer-Address     Address,
Phone                VARCHAR(15));
```

(f) In the table given below, **Column I** contains typical functions of a Data Warehouse. **Column II** lists description of some data warehouse functionality.

	Column I		**Column II**
A	Roll-up	1	Data is available by value or range.
B	Drill-down	2	Performing projection operations on the dimensions
C	Pivot	3	Cross tabulation is performed.
D	Slice and dice	4	Data is summarised with increasing generalization.
E	Selection	5	Increasing levels of details are revealed.
		6	Attributed are computed by operations on stored values.
		7	Data is sorted by ordinal value.

Match each function from **Column I** with the most appropriate description in **Column II**. Write your answer in the box given below.

(03 marks)

ANSWER IN THIS BOX

A – **4** B – **5** C – **3**

D – **2** E – **1**

4) (a) What is a mobile database?

(02 marks)

ANSWER IN THIS BOX

A mobile database is a database that can be connected to via a mobile computing device over a mobile network.

(b) What is a multimedia database?

(02 marks)

ANSWER IN THIS BOX

A multimedia database is a database that hosts one or more primary media file types such as .txt (documents), .jpg (images), .swf (videos), .mp3 (audio), etc.

(c) What are the two main structuring concepts used to construct an XML document? Using an example show how the two concepts can be used in an XML document.

(07 marks)

ANSWER IN THIS BOX

Elements and Attributes

e.g.

```
<bookstore>
  <book category="CHILDREN">
    <title>Harry Potter</title>
    <author>J K. Rowling</author>
```

Continued...

```
  <year>2005</year>
  <price>29.99</price>
 </book>
</bookstore>
```

In the example above, <bookstore> and <book> have element contents, because they contain other elements.

<book> also has an attribute (category="CHILDREN").

<title>, <author>, <year>, and <price> have text content because they contain text.

(d) List the six (06) phases of a Knowledge Discovery process.

(03 marks)

ANSWER IN THIS BOX

(1) data selection

(2) data cleansing

(3) enrichment

(4) data transformation or encoding

(5) data mining

(6) the reporting and display of discovered information

(e) A variety of distributed database options exist. Briefly explain the range of distributed database environments.

(08 marks)

ANSWER IN THIS BOX

Organizations may use homogeneous or heterogeneous environments.

Continued...

Homogeneous - The same DBMS is used at each node.

It can be Autonomous where each DBMS works independently,

passing messages back and forth to share data updates OR

Non-autonomous where a central or master DBMS coordinates database access

and updates across the nodes

Heterogeneous - Potentially different DBMSs are used at each node.

Distributed systems support some (partial Multi-database) or all of the functionality

of one logical database.

Federated databases support local databases for unique data requests which are

loosely (many schemas exist, for each local database) or

tightly (one global schema) integrated.

(f) What are the basic strategies for distributing data among the sites (or nodes) of a network?

(03 marks)

ANSWER IN THIS BOX

Data replication

Horizontal partitioning

Vertical partitioning

Combinations of the above

****