UNIVERSITY OF COLOMBO, SRI LANKA

UNIVERSITY OF COLOMBO SCHOOL OF COMPUTING

DEGREE OF BACHELOR OF INFORMATION TECHNOLOGY (EXTERNAL)

Academic Year 2007/2008 –3rd Year Examination – Semester 6

IT6402 - Advanced Database Management Systems

Structured Question Paper

31st August, 2008

(THREE HOURS)

To be completed by the candidate

BIT Examination Index No: .....

Important Instructions:

- The duration of the paper is **3 (three) hours**.
- The medium of instruction and questions is English.
- This paper has **4 questions** and **15 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.
- **Non-programmable Calculators may be used.**

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:					

1) (a) What kind of operations must a file in a database support and how does the file support each such operation?

(04 marks)

ANSWER IN THIS BOX

A file must support insertion/deletion/modification of records.

It must also support fetching a particular record and scanning all the records.

In order to support these operations, a file must keep track of the pages in a file,

free space on pages and records on a page.

(b) If one is to create an index on a relation, what indexing must be chosen if the most frequent operations are

(i) searching for records based on a range of non-key field values?

(03 marks)

ANSWER IN THIS BOX

Clustered B+ tree index using necessary fields as search keys

(ii) searching for a record based on equality of a particular field value?

(03 marks)

ANSWER IN THIS BOX

Hash index using the necessary field as the search key

(c) Consider the following `Employee` and `Department` relations with the following characteristics.

```
Employee(EmpNo, eName, Designation, DeptNo)
Department(DeptNo, dName, Manager, Telephone)
```

Characteristics: Primary keys are underlined; `DeptNo` of `Employee` is a foreign key of `Department`; `eName` is a non-clustered index and `DeptNo` of `Employee` is a clustered index.

Assume that both relations have large volumes of data (hypothetical situation) and the following represent parts of selected data.

Employee

EmpNo	eName	Designation	DeptNo
0101	Perera	Software Engineer	IT
0201	Silva	Manager	Fin
0203	De Silva	Accounts Clerk	Fin
0102	Soysa	Manager	IT

Department

DeptNo	dName	Manager	Telephone
IT	Information Technology	Soysa	112581245
Fin	Finance	Silva	112589123

Discuss how each of the following queries is expected to be processed by a typical DBMS with respect to the following aspects:

- tables accessed (Full Scan or Index Scan or None at all),
- indexes used,
- how the data is selected (selecting from table or from index only) and
- operations applied

Assume that only indexes could influence the queries (i.e. effects from cache, database statistics are ignored). Query operations include table accessing, use of nested loops, hash-join, merge-joins and filters.

(i) `SELECT dName FROM Department;`

(02 marks)

ANSWER IN THIS BOX

Full table scan of *Department* as it is not indexed by *dName*

No Indexes used

Select data from table

(ii) `SELECT * FROM Employee WHERE DeptNo="IT";`

(02 marks)

ANSWER IN THIS BOX

Index scan of *Employee* using index of *DeptNo*

Use *DeptNo* clustered index

Select data from table

(iii) `SELECT eName FROM Employee;`

(02 marks)

ANSWER IN THIS BOX

No table scan as index contain the required information

Use eName non-clustered index

Select data from Index

(iv) `SELECT DeptNo, COUNT(*) FROM Employee GROUP BY DeptNo;`

(03 marks)

ANSWER IN THIS BOX

No table scan as index contain the required information

Use DeptNo clustered index

Count the number of data entries in the index

(v)
```
SELECT e.*, d.* FROM Employee e, Department d
WHERE e.DeptNo =d.DeptNo;
```

(03 marks)

ANSWER IN THIS BOX

Full scan of both tables as all data of both tables are required

Join tables using matching DeptNo of *Employee* with *DeptNo* use Hash-Join

Merge results

(vi)
```
SELECT d.dName, e.eName FROM Department d, Employee e
   WHERE d.DeptNo=e.DeptNo and e.DeptNo="IT";
```

(03 marks)

ANSWER IN THIS BOX

Access *Department* table using Index Scan of *DeptNo*

Access *Employee* table using Index Scan of *DeptNo*

Filter *Department* data using Index

Select *dName* of *DeptNo*="IT"

Select *eName* from *Employee* "IT" using clustered index

2) (a) Data fragmentation is one of the considerations when designing a distributed database.

(i) What is data fragmentation?

(02 marks)

ANSWER IN THIS BOX

Data fragmentation is the process of dividing a relation (table) into a number of

sub-relations for the purpose of allocating them across different sites.

(ii) Why should one fragment data? Give reasons.

(06 marks)

ANSWER IN THIS BOX

Usage – Database applications usually work with sub-sets of relations (views)

rather than entire relations. Hence data distribution could resemble its usage.

Efficiency – Most frequently used data can be stored nearby (locally) and

processed quickly as they are smaller in size.

Continued...

Security - Data that is not required by local applications will not be stored locally.

Hence it is not available to unauthorized users of the local site.

Parallelism - A transaction can be divided into several sub-queries which operate on fragments, thereby allowing transactions to execute them in parallel.

(iii) Fragmentation process should fulfil the correctness rules. Describe these rules and their purposes.

(06 marks)

ANSWER IN THIS BOX

Completeness - When a relation R is decomposed into fragments R1, R2, . . . Rn, each data item that can be found in R must appear in at least one of the fragments R1, R2, . . . Rn.

This rule is necessary to ensure that there is no loss of data during fragmentation.

Reconstruction - It must be possible to define a relational operation that will reconstruct the relation R from the fragments.

This rule ensures that functional dependencies are preserved.

Continued...

Disjointness - A data item in a fragment should not appear in any other fragment,

except in the case of the primary key of a vertical fragmentation

which is needed to allow reconstruction.

This rule ensures minimal redundancy.

(b) The Government of Sri Lanka monitors projects using the following three tables of the Project database. The primary keys of the tables are underlined.

```
Employee (empno, name, designation, telephone, salary)
Project (projectid, pname, Budget, location)
Assignment (projectid, empno, responsibility)
```

In this system employees are assigned to several projects each with a specific role in each project. For project monitoring, the Government pays more attention to project budgets of those carried out in two districts and hence the predicates `location="Jaffna"`, `location="Hambantota"` appear in majority of queries involving projects. Besides, the Government pays attention to salaries paid to employees and hence the predicates with `salary<10000`, `salary>50000` appear in majority of queries involving employees. From among the attributes of `Employee` relation, only `designation` appears in majority of queries as it is associated with the salary.

(i) Fragment the `Employee` relation for the benefit of the above needs. Justify your answer.

(06 marks)

ANSWER IN THIS BOX

Employee mixed fragmentation through vertical and horizontal fragmentations

First vertical fragmentation of Employee to separate designation and salary

from others as Emp1 and Emp2.

Emp1(empno, designation, salary) = $\pi_{empno,\ designation,\ salary}$ Employee

Emp2(empno, name, telephone) = $\pi_{empno,\ name,\ telephone}$ Employee

Then horizontal fragmentation of Emp1 on salary to store

salaries <10000 and >50000 in separate fragments and

others together as Emp3, Emp4 and Emp5

Continued...

Emp3(empno, designation, salary) = $\sigma_{salary<10000}$ Emp1

Emp4(empno, designation, salary) = $\sigma_{salary>50000}$ Emp1

Emp5(empno, designation, salary) = $\sigma_{salary>=10000 \text{ and } salary<=50000}$ Emp1

(ii) Fragment the `Project` relation for the benefit the above needs. Justify your answer.

(05 marks)

ANSWER IN THIS BOX

Horizontal fragmentation of Project to separate projects in Jaffna and Hambantota from others as Proj1, Proj2 and Proj3

No vertical fragments as all attributes are needed

Proj1 = $\sigma_{location="Jaffna"}$ Project

Proj2 = $\sigma_{location="Hambantota"}$ Project

Proj3 = $\sigma_{location<>"Jaffna" \text{ and } location<>"Hambantota"}$ Project

3) (a) What is database auditing? During an audit, what type of questions has to be answered?

(05 marks)

ANSWER IN THIS BOX

Database auditing is a surveillance mechanism that watches over access to all sensitive information contained within the database to detect unauthorised activity.

Type of questions include the following:

- **Who accessed the data?**
- **When?**
- **Using what computer program or client software?**
- **From what location on the network?**
- **What was the SQL query that accessed the data?**
- **Was it successful; and if so, how many rows of data were retrieved?**

(b) Discuss how a Data Warehouse is different from a Database System.

(05 marks)

ANSWER IN THIS BOX

Data Warehouse	Database System
Subject-oriented data	Application/Process-oriented data
De-normalised data repository	Normalised data repository
Non volatile data (Append only)	Volatile data (Real-time updates)
Historical data (years/time-variant)	Current data (days and months)
Summarised data	Detailed data
Diverse sources including external data	Internal application data
Analyse data for decision support	Analyse for operational support
Low frequency of access	High frequency of access
Process in large amounts of data	Process in small amounts of data

(c) What is a cube in a data warehouse? Explain how it is constructed and used.

(05 marks)

ANSWER IN THIS BOX

A cube is a pre-computed multidimensional data structure that contains

dimensions (time), hierarchies (calendar and fiscal),

levels (year, quarter) and measures (count, price, cost).

Each individual point in a cube is referred to as a cell.

Operational data is aggregated into cubes which are cross joined by dimensions.

A Data warehouse stores pre-computed cubes and supports a quick search and

analyses through operations such as drill down, roll up, slice, dice and pivot.

(d) Object-relational databases address four major weaknesses of conventional relational systems. Identify what they are and discuss how SQL standards such as SQL3 (SQL/99) support to achieve them giving SQL syntax as an example for each weakness address.

(10 marks)

ANSWER IN THIS BOX

Major weaknesses addressed in object-relational databases are:

Support for base type extension (beyond char, int, date)

Support for complex objects

Support for inheritance

Support for active rules

Allows defining new base types (address) and uses them in a way similar to built-in types

CREATE TYPE Addr (street VARCHAR(30), city VARCHAR(20), zip VARCHAR(7))

CREATE TABLE Customer (name VARCHAR(20), address Addr, Phone CHAR(10))

Continued...

Allows defining structured types with non-atomic values as complex objects consist

of aggregation of values of other types

Could use them to introduce new types, user defined functions to operate them

CREATE TYPE Person AS OBJECT (nic CHAR(10), name VARCHAR(20))

Allows re-use of complex objects and user-defined functions which operate on them by defining subtypes from existing types

The new type will inherit the data and the functions of its supertype.

CREATE TYPE Student UNDER Person (deptid CHAR(5), degree VARCHAR(20))

Allows additional integrity constraints in forms of active rules

Each rule is associated with an event and is carried out when the event occurs.

CREATE TRIGGER cascade_updates

AFTER UPDATE OF deptno ON Dept FOR EACH ROW

UPDATE Emp SET Emp.deptno = :new.deptno WHERE Emp.deptno = :old.deptno;

4) (a) (i) Database transactions use schedules. What is a schedule?

(02 marks)

ANSWER IN THIS BOX

Sequence of instructions from concurrent transactions indicating the chronological order in which these instructions are executed

(ii) Which protocol is commonly used to ensure conflict serializable schedules?

(02 marks)

ANSWER IN THIS BOX

Two-phase locking protocol

(iii) What is stored in a database lock table?

(02 marks)

ANSWER IN THIS BOX

Database lock table stores granted locks and pending requests for locks.

(iv) What are the two different approaches for log-based recovery?

(02 marks)

ANSWER IN THIS BOX

Deferred database modifications and

Immediate database modifications

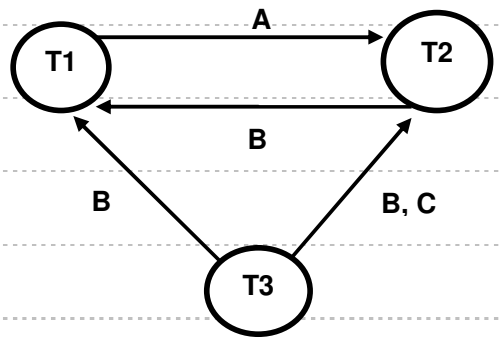
(b) Consider the following three transactions T1-T3 which are to be executed concurrently.

T1	T2	T3
	READ(C) READ(B) B:=B+10 WRITE(B)	
		READ(B) READ(C)
READ(A) A:=A+20 WRITE(A)		
		B:=B+C WRITE(B) C:=B-C WRITE(C)
	READ(A) A:=A-10	
READ(B) B:=B-A WRITE(B)		
	WRITE(A)	

(i) Draw the precedence graph of the above schedule.

(03 marks)

ANSWER IN THIS BOX



(ii) Using the precedence graph produced in (i) above, discuss whether the given schedule is conflict serializable?

(03 marks)

ANSWER IN THIS BOX

There is a cycle in the precedence graph.

i.e. T1 $\xrightarrow{A}$ T2 and T2 $\xrightarrow{B}$ T1.

Hence, the schedule is not conflict serializable.

(iii) Hence, discuss whether the given schedule is view serializable?

(03 marks)

ANSWER IN THIS BOX

Result of this schedule is not equivalent to any of the possible six serial schedules of T1-T3

(i.e. T1T2T3, T1T3T2, T2T1T3, T2T3T1, T3T1T2 or T3T2T1) as values of A and B are incorrectly overwritten by transactions T1 and T2. Therefore, the given schedule is not view serializable.

(c) The three transactions T1-T3 of (b) above have been executed concurrently as follows.

T1	T2	T3
		READ(B)
		READ(C)
READ(A)		
A:=A+20		
WRITE(A)		
		B:=B+C
		WRITE(B)
		C:=B-C
		WRITE(C)
	READ(C)	
READ(B)		
B:=B-A		
WRITE(B)		
	READ(B)	
	B:=B+10	
	WRITE(B)	
	READ(A)	
	A:=A-10	
	WRITE(A)	

(i) Discuss conflict serializability and view serializability of the schedule in part (c). Justify your answer.

(05 marks)

ANSWER IN THIS BOX

All conflicting operations are interleaved to achieve serializability

The schedule is conflict serializable to the serial schedule T3, T1, T2.

Hence it is also view serializable

(ii) Is it possible to design a schedule for the above transactions under the timestamp ordering protocol? Justify your answer.

(03 marks)

ANSWER IN THIS BOX

No. There are conflicting read and write operations on B in T1 and T2

Such operations cannot be executed in timestamp order.

****