UNIVERSITY OF COLOMBO, SRI LANKA

UNIVERSITY OF COLOMBO SCHOOL OF COMPUTING

DEGREE OF BACHELOR OF INFORMATION TECHNOLOGY (EXTERNAL)

Academic Year 2005/2006 –3rd Year Examination – Semester 6

IT6402: Advanced Database Management Systems

Structured Question Paper

20th August, 2006
(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 **16 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 **4** 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) Name three primary file organisations which determine how the records of a file are physically placed on the disk. Indicate how records are placed and accessed with respect to the file organisation techniques which you have named.

(03 marks)

ANSWER IN THIS BOX
Three of the following:
Heap file to append records at the end of the file and searching is linear.
Sorted file to keep records ordered by a value of a field and accessed using binary search
Hash file to determine arbitrary placement of a record by a hash function and access through hash function
B-Tree and files of mixed records use relationships among records to organising and access data
RAID uses independent disks and data distribute data across disks for improved performance.

(b) (i) Give the most commonly used structure for a high-level SQL query statement.

(02 marks)

ANSWER IN THIS BOX
SELECT output attributes
FROM list of tables
WHERE conditions

(ii) Describe briefly the process of formulating an initial query tree from the query statement of (b)(i).

(02 marks)

ANSWER IN THIS BOX
Map query output attributes to the root node of the tree as a projection.
Introduce the query conditions at the next level of the tree as a restrict operation.
Include the list of tables as the leaf nodes of the tree.
Join the leaf nodes to the tree using Cartesian product.

(iii) Transformation rules are used to optimise a query. List the main transformation rules used in the query optimisation process.

(03 marks)

ANSWER IN THIS BOX

Break query conditions into individual operations to enable one to move them down the tree.

Move selected query conditions down the tree and apply them to allow restricting rows retrieved from tables.

Project attributes that are needed for the join operations and for the output results. It reduces the record length of tables.

Where applicable, replace Cartesian operations using join attribute to perform join operations.

(c) The examination branch uses a relational database to record and process student examination results. The following are some of the relations of this student database. Here, Student relation records student data and Subject relation records subject data. Actual marks gained by the students for respective subjects are recorded in the Marks relation along with a grade.

```
Student(index no, name, address)
Subject(subject code, subject_name, lecturer)
Marks(index no, subject code, mark, grade)
```

(i) Write an SQL statement to list all the students taking the subject called "Database Systems" giving the index no, name and grade of each.

(03 marks)

ANSWER IN THIS BOX

```
SELECT c.index_no, c.name, m.grade
FROM Marks AS m, Subject AS s, Student As c
WHERE (s.subject_code = m.subject_code) AND
(c.index_no = m.index_no) AND
(s.subject_name = ' Database Systems');
```

(ii) Applying the more restrictive operators first, express the query of (b)(i) above in relational algebra.

(03 marks)

ANSWER IN THIS BOX

Database = $\sigma_{\text{subject\_name = 'Database Systems'}}$ (Subject)

DB = $\pi_{\text{subject\_code}}$ (Database)

St = $\pi_{\text{index\_no, name}}$ (Student)

DB_Marks = DB $\infty_{\text{db.subject\_code = marks.subject\_code}}$ Marks

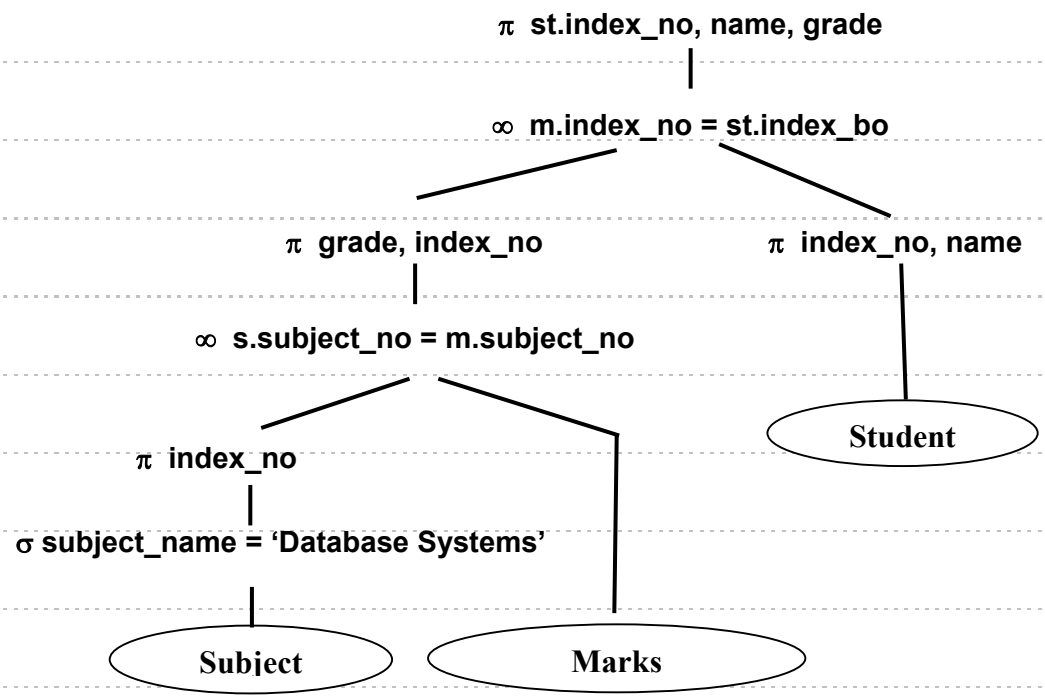
STDB_Marks = St $\infty_{\text{st.index\_no = db\_marks.index\_no}}$ DB_Marks

Result = $\pi_{\text{index\_no, name, grade}}$ (STDB_Marks)

(iii) Draw the optimised query tree for the above query.

(06 marks)

ANSWER IN THIS BOX



(d) What areas of queries, applications and transactions have to be analysed to identify the factors which influence physical database design?

(03 marks)

ANSWER IN THIS BOX

Analyse files which would be accessed or updated.

Type of operations performed on files

Attributes on which selection conditions for query are specified

Attributes whose values would be changed

Analyse expected frequency of invocation to measure usage.

Analyse time constraints of queries to measure performance.

Analyse expected frequencies of update operations.

Analyse the uniqueness constraints on attributes to determine access paths.

2) (a) Use of several isolation levels is possible when implementing database transactions. When executing simultaneous transactions, one or more violations may occur under most isolation levels.

(i) Describe briefly these violations.

(04 marks)

ANSWER IN THIS BOX

Dirty read: A transaction T1 may read the update of a transaction T2, which has not yet committed.

If T2 fails and is aborted, then T1 would have read a value that does not exist and is incorrect.

Non-repeatable read: A transaction T1 may read a given value from a table.

If another transaction T2 later updates that value and T1 reads that value again, T1 will see a different value.

Phantoms: A transaction T1 may read a set of records based on some specified condition.

Continued…

If transaction T2 inserts a new record satisfying the same condition used in T1 into the same table used by T1, and if

T1 is repeated, then T1 will see a phantom, a row that previously did not exist.

(ii) Specify the type of violations for isolation levels which you identified in (a)(i) above.

(03 marks)

ANSWER IN THIS BOX

	Type of Violation		
Isolation Level	**Dirty Read**	**Non-repeatable read**	**Phantom**
Read Uncommitted	**yes**	**yes**	**yes**
Read Committed	**no**	**yes**	**yes**
Repeatable Read	**no**	**no**	**yes**
Serializable	**no**	**no**	**no**

(b) Consider the following two transactions T1 and T2 with the database value for X as 500.

T1	T2
READ(X)	READ(X)
Y = X + 100	Y = X – 100
WRITE(Y)	WRITE(Y)
COMMIT	COMMIT

(i) Without considering the locking technique, write a possible serial schedule for T1 and T2. What is the final database value of Y?

(03 marks)

ANSWER IN THIS BOX

READ(X)

Y = X + 100 or Y = X – 100

WRITE(Y)

COMMIT

Continued…

READ(X)

Y = X – 100 or Y = X + 100

WRITE(Y)

COMMIT

Y is 400 or 600

(ii) Without considering the locking technique, write a possible non-serial schedule for T1 and T2 that would yield a correct result. What is the final database value of Y?

(03 marks)

ANSWER IN THIS BOX

READ(X)	
	READ(X)
Y = X + 100	
	Y = X - 100
WRITE(Y)	
COMMIT	
	WRITE(Y)
	COMMIT

Y is 400 or 600

(iii) Write a database log for the above schedule of (b)(ii).

(03 marks)

ANSWER IN THIS BOX

<T1, Begin>

<T2, Begin>

<T1, Y, null, 600>

<T1, Commit>

<T2, Y, null, 400>

<T2, Commit>

(iv) Assume that the last checkpoint record is just before the commencement of the schedule given in (b)(ii) above and the database is using immediate update technique. If the schedule for T1 and T2 fails prior to the very last commit statement, explain the database recovery actions which would take place. Indicate what changes took place in the database and what has to be done after the database is recovered to ensure the complete execution of the expected schedule.

(05 marks)

ANSWER IN THIS BOX

Recovery process prepares two lists, namely redo and undo.

i.e., Redo <T1> and Undo <T1, T2>

Recovery process will rollback the undo list.

Automatically execute redo list and complete the recovery process.

Due to the recovery process the database changes Y from 600 to 400.

After the recovery, to complete the schedule, T2 has to be executed again.

(v) If binary lock concept is used, write a possible non-serial schedule for T1 and T2 that would yield a correct result. Show the locks acquired and released.

(04 marks)

ANSWER IN THIS BOX

Lock(X)	
Read(X)	
Y = X + 100	
Unlock(X)	
	Lock(X)
	Read(X)
Lock(Y)	
Write(Y)	
	Y = X – 100
Unlock(Y)	
	Lock(Y)
COMMIT	
	Write(Y)

Continued…

Unlock(Y)

COMMIT

Note: Alternate solutions exist.

3) (a) (i) Legal, ethical and policy issues control the right to access information. Using examples identify them.

(03 marks)

ANSWER IN THIS BOX

Laws governing privacy prevent organisations recording data without obtaining permission from the owner.

Policies at different level would control which data would be made public.

E.g. name of the patient admitted to the hospital but not his medical records.

(ii) System level security can be encorded to control access to a database system. Identify what they are.

(02 marks)

ANSWER IN THIS BOX

Physical access to hardware level

Access to the operating system level

Access to the database management system level

(iii) Some organisations classify data into multiple security levels. Using examples, identify what they are.

(03 marks)

ANSWER IN THIS BOX

Top secret – e.g. medical insurance claims

Secret – e.g. salary details of employee

Confidential – e.g. date of birth of employee

Unclassified – e.g. name of employee

(b) The following two relations are part of a Univerity examinations database.

Course(coursecode, coursename, lecturername, departmentname)
Marks(coursecode, studentid, mark)

The University has provided all heads of departments (e.g. users H1, H2) full rights to change data in Marks relation which was entered by their teaching staff (e.g. users S1, S2). University has given only insert rights to all teaching staff to enable them to insert Marks data for their courses.

To facilitate the above functionality, two roles named as head and staff are to be defined. Login access is to be provided to all users with appropriate previleges.

(i) Write (a) SQL statement(s) to retrieve the data accessible by a paticular teaching staff member. You may assume that user account names tally with lecturer name of Course relation.

(02 marks)

ANSWER IN THIS BOX

```
CREATE VIEW Staff_Marks AS
SELECT m.*, c.lecturer FROM Marks m, Course c
WHERE c.coursecode=m.coursecode

SELECT coursecode, studentid, mark
FROM Staff_Marks WHERE lecturer=$USER
```

(ii) Write (a) SQL statement(s) to retrieve the data accessible by a paticular head of a department. You may assume that user account names tally with department name of Course relation.

(02 marks)

ANSWER IN THIS BOX

```
CREATE VIEW Head_Marks AS

SELECT m.* FROM Marks m, Course c

WHERE c.coursecode=m.coursecode



SELECT coursecode, studentid, mark

FROM Staff_Marks WHERE department=$USER
```

(iii) Create roles for each user group and assign previleges to manipulate authorised relations.

(06 marks)

ANSWER IN THIS BOX

```
CREATE ROLE STAFF

GRANT SELECT, INSERT ON Staff_Marks TO STAFF

GRANT STAFF TO S1, S2




CREATE ROLE HEAD

GRANT ALL ON Head_Marks TO HEAD

GRANT HEAD TO H1, H2
```

(c) (i) What actions chould be taken to protect the confidentiality of sensitive data when such data are transmitted over a network?

(02 marks)

ANSWER IN THIS BOX

Encrypt the data

Use secure data connection: SSL, VPN

(ii) Audit trail is used to keep track of database activities. Identify the type of information that should be recorded in a database log file to assit in tracing back database changes.

(02 marks)

ANSWER IN THIS BOX

User, accessed data, changes to the data / structure

(iii) Identify possible useful activites to monitor to enable one to detect irregular database activities.

(03 marks)

ANSWER IN THIS BOX

Shutdown/re-boot

Login / access failures

Attempts with non-existing users

Attempts to access at unusual hours

Multiple users accessing from same terminal

Same user access from multiple terminals

4) (a) (i) In a distributed database, data replication is one of the possible activities. What are the advantages in replicating the data of a database?

(02 marks)

ANSWER IN THIS BOX

Replication is done for better availability, performance and reliability.

(ii) Data in a distributed database can be replicated using snapshots or replicated master. Describe what it is and its main purpose.

(02 marks)

ANSWER IN THIS BOX

It generates a copy of the data and allows update of data.

(iii) Several types of transparencies are possible in a distributed database. Name and briefly explain them.

(05 marks)

ANSWER IN THIS BOX

Distribution or network transparency – freedom from the operational details - May be divided into location and naming transparency

Location transparency – Commands used are independent of the location of the data.

Naming transparency – Name of objects can be accessed unambiguously without additional specifications.

Replication transparency – User is unaware of the existence of multiple copies.

Continued…

Fragmentation transparency – two types of fragmentation: horizontal and vertical.

Horizontal fragmentation distributes a relation into sets of tuples (rows).

Vertical fragmentation distributes a relation into sub-relations where each sub-relation is defined by a set of columns of the original relation.

Through a global relation these fragmentations and the existence of these fragments are hidden from the user.

(b) Explain which relational database problems can be solved by using XML databases. Use an example to illustrate the identified problems.

(03 marks)

ANSWER IN THIS BOX

Relational databases manage structured data and are unable to manage semi-structures and unstructured data effectively.

e.g. In relational databases multi-valued attributes have to be separated to form a uniform structure.

In XML, such data can be defined and kept together as it allows any structure.

(c) Data warehouse facilitates complex, data-intensive and frequent ad hoc queries. Briefly describe the typical functions available in a data warehouse to perform these queries.

(04 marks)

ANSWER IN THIS BOX

Roll-up: Data is summarised with increasing generalisation, e.g. weekly to quarterly to annually.

Roll-down: increase levels of details, e.g. annual to quarterly to weekly

Pivot: Cross tabulation

Slice and dice: performing projection operations on the dimensions

Sorting

Selection

Derived attributed

(d) Type constructors have been added to specify complex objects. Using an example, describe how one could define a construct for the address of an employee.

(03 marks)

ANSWER IN THIS BOX

```
CREATE TYPE Address AS (
  Street VARCHAR(30),
  City VARCHAR(30),
  Postal VARCHAR(20) );
```

(e) Database systems allow the management of extremely large objects like video, audio and text documents. Identify the new data types available to support these requirements.

(03 marks)

ANSWER IN THIS BOX

Three of the following

BLOB – binary large objects

CLOB – character large objects

BFILE – binary file stored outside the database

NCLOB – fixed width multi-byte CLOB

(f) Describe briefly the goal of clustering the data.

(03 marks)

ANSWER IN THIS BOX

Clustering is to place records into groups, such that records in a group are similar to each other and dissimilar to records in other groups. i.e. disjoint groups of data.

Then when accessing similar data they are all together.

****