UNIVERSITY OF COLOMBO, SRI LANKA

UNIVERSITY OF COLOMBO SCHOOL OF COMPUTING

DEGREE OF BACHELOR OF INFORMATION TECHNOLOGY (EXTERNAL)

Academic Year 2013/2014 –3rd Year Examination – Semester 6

IT6404 - Database Systems II

Structured Question Paper

2nd August, 2014
(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 **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 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) (i) What is a Sequential File?

(01 mark)

ANSWER IN THIS BOX

(b) If b is the number of blocks in a file, write down, on average, how many file blocks should be accessed to search in the case of

(i) a specific record from a **Heap File**?

(ii) several records from a **Heap File**?

(iii) a record from a **Sequential File**?

(03 mark)

ANSWER IN THIS BOX

(i)

(ii)

(iii)

(c) (i) Briefly describe a single-level index.

(02 marks)

ANSWER IN THIS BOX

(ii) Name and briefly describe two types of single-level ordered indexes. Give a suitable example for each.

(04 marks)

ANSWER IN THIS BOX

(iii) What is the term used to refer to a primary index that includes an entry for every record?

(01 mark)

ANSWER IN THIS BOX

(d) (i) List the factors that influence the physical database design.

(02 marks)

ANSWER IN THIS BOX

(ii) Design decisions on indices play an important role in the physical database design. List four design decisions on indices.

(02 marks)

ANSWER IN THIS BOX

(iii) Describe what database tuning is and what its goals are.

(02 marks)

ANSWER IN THIS BOX

(e) Consider the following query issued on the warehouse database consisting of several relations including the following three relations where primary keys are underlined and foreign keys are in italics.

```
branch(branchno, street, city, postalcode);
staff(staffno, fname, lname, position, sex, dob, salary, branchno);
rental(propertyno, street, city, rtype, rooms, rent, staffno);

SELECT propertyno, rental.city,rooms,lname, staff.branchno
FROM rental,staff,branch
WHERE rental.staffno = staff.staffno
   AND staff.branchno = branch.branchno
   AND branch.city = rental.city
   AND rooms > 3
   AND rtype = 'House';
```

(i) Suggest suitable indices to improve the above query and discuss how this query would be processed by the query optimizer based on the suggested indices if query optimization rules are applied.

(03 marks)

ANSWER IN THIS BOX

(ii) Draw an optimized query tree for the query given in 1(e).

(05 marks)

ANSWER IN THIS BOX

2) (a) (i) Identify if the following two schedules are (conflict) serializable or not.

(A) $r_1(a);\ r_3(a);\ w_1(a);\ r_2(a);\ w_3(a);$
(B) $r_3(a);\ r_2(a);\ w_3(a);\ r_1(a);\ w_1(a);\ .$

If any of the above is a serializable schedule determine the equivalent serial schedules and if not serializable, indicate why it is non-serializable and identify the type of conflict. Note that r_i and w_i denote respectively the read and write operations of transaction T_i for data item a.

(04 marks)

ANSWER IN THIS BOX

(A)

(B)

(ii) Draw precedence graphs for two of the above two schedules.

(04 marks)

ANSWER IN THIS BOX

(A)

(B)

(b) Consider the following schedule S1 given below. Please note that r_i and w_i denote respectively the read and write operations of transaction T_i and a, b, c are data items.

S1 = $r_1(a)$, $r_2(c)$, $r_1(c)$, $r_3(a)$, $r_3(b)$, $w_1(a)$, c_1, $w_3(b)$, c_3, $r_2(b)$, $w_2(c)$, $w_2(b)$, c_2

Writ down the locks acquired, released or changed (i.e. **Release S(A)** for release of shared lock for A) including any waiting for locks, commits or deadlocks at each of the times t1 to t13.

(09 marks)

ANSWER IN THIS BOX

Time	T1	T2	T3	Acquire Locks	Release or Change Locks
t1	READ(a)				
t2		READ(c)			
t3	READ(c)				
t4			READ(a)		
t5			READ(b)		
t6	WRITE(a)				
t7	COMMIT				
t8			WRITE(b)		
t9			COMMIT		
t10		READ(b)			
t11		WRITE(c)			
t12		WRITE(b)			
t13		COMMIT			

(c) Consider the following three interleaved transaction T1, T2 and T3. Here both T1 and T2 have reached the end of the transaction while T3 is yet to complete. Initial database values of A, B, C and D were as 30, 15, 40 and 20 respectively.

T1	T2	T3
		READ(C)
		B = 12
		WRITE(B)
	READ(B)	
	B = B + 6	
	WRITE(B)	
READ(A)		
READ(D)		
D = D + 5		
WRITE(D)		
	READ(D)	
	D = D + 1	
	WRITE(D)	
COMMIT		
		READ(A)
… System Crash …		

(i) Assuming the schedule is executed under immediate update, give essential log entries for the above schedule to enable it to recover.

(05 marks)

ANSWER IN THIS BOX

Continued…

(ii) Explain how the recovery process would work for the above schedule. Indicate which transactions would be rollback and which can be recovered.

(03 marks)

ANSWER IN THIS BOX

3) (a) Distribution leads to increased complexity in the system design and implementation. What are the additional functions a Distributed DBMS should provide to those of a centralized DBMS?

(06 marks)

ANSWER IN THIS BOX

(b) Consider the following relations:

Books(BookNo, PrimaryAuthor, Subject, Price)
BookStore(StoreNo, City, Province, Zip, InventoryValue)
Stock(StoreNo, BookNo, Qty)

(i) Give three example predicates that would be meaningful for the Book and BookStore relations, each to be horizontally partitioned based on Price and City respectively.

(04 marks)

ANSWER IN THIS BOX

(ii) How would a derived horizontal partitioning of Stock be defined based on the partitioning of BookStore in (i) above?

(04 marks)

ANSWER IN THIS BOX

(c) Recent advances in portable and wireless technology have led to mobile computing, a new dimension in data communication and processing. From a data management standpoint, mobile computing may be considered a variation of distributed computing. Identify the components of a mobile platform and briefly describe possible methods to distribute mobile databases among the identified components.

(05 marks)

ANSWER IN THIS BOX

(d) List and briefly describe four (04) types of multimedia data that are available in the most current Multimedia DBMSs.

(06 marks)

ANSWER IN THIS BOX

Continued…

4) (a) There are several approaches in organizing the content of XML documents to facilitate their subsequent queries and retrieval. Briefly describe two such approaches.

(04 marks)

ANSWER IN THIS BOX

(b) Consider the following Document Type Definition (DTD) that describes a part of a university database:

```
<!DOCTYPE courses [
  <!ELEMENT courses (course*)>
  <!ELEMENT course (title, taken_by)>
    <!ATTLIST course cno CDATA #REQUIRED>
  <!ELEMENT title (#PCDATA)>
  <!ELEMENT taken_by (student*)>
  <!ELEMENT student (name, grade)>
    <!ATTLIST student sno CDATA #REQUIRED>
  <!ELEMENT name (#PCDATA)>
  <!ELEMENT grade (#PCDATA)>
]>
```

(i) Express the above DTD in terms of normalized relations of a relational database with its keys.

(03 marks)

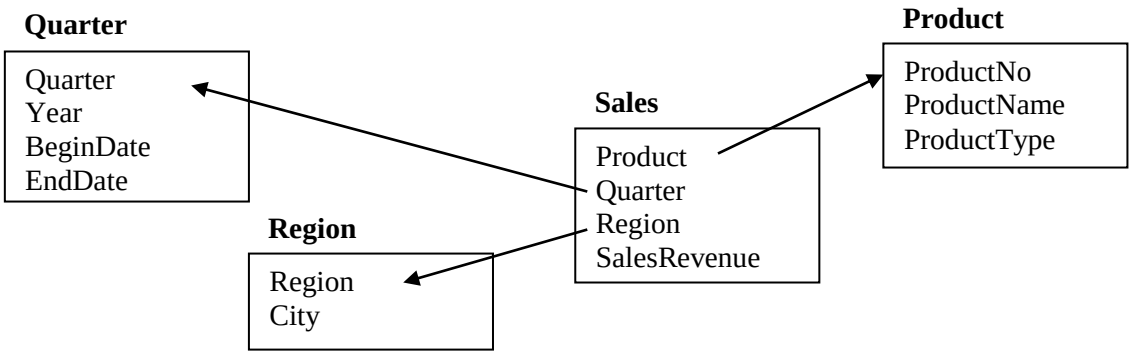
ANSWER IN THIS BOX

(ii) Give an XML document to record a student (name = Dias & sno = 111) who has obtained an 'A' grade for the database course (cno = 01) based on the given DTD.

(03 marks)

ANSWER IN THIS BOX

(c) Consider the following diagram representing a start schema of a sales data warehouse.



(i) Identify the dimension and fact table(s) for the above data warehouse.

(04 marks)

ANSWER IN THIS BOX

Dimension table(s) Fact table(s)

(ii) The above sales data warehouse is to be used for pre-programme functionality such as Roll-up, Drill-down, Slice and Dice. If a data cube had been defined for the warehouse given in (c) above, give an example for each of the four functionalities.

(08 marks)

ANSWER IN THIS BOX

Roll-up

Drill-down

Continued…

Slice

Dice

(d) Data mining technologies are applied to a large variety of decision-making business applications. If a supermarket chain is to use this data mining technologies to improve their marketing strategies, what type of data analysis could be performed? What type of marketing strategies could be expected based on the outcome of the analysis?

(03 marks)

ANSWER IN THIS BOX

****