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. ☒) the numbers of the questions answered.

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

1) (a) (i) What is a Sequential File?

(01 mark)

> **ANSWER IN THIS BOX**
>
> **Sequential file is an ordered file where records are kept sorted by values of an**
>
> **ordering field.**

(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) **half the file blocks (b/2)**
>
> (ii) **all blocks (b)**
>
> (iii) **$Log_2\ b$**

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

(02 marks)

> **ANSWER IN THIS BOX**
>
> **A single-level index is an auxiliary file that is used to efficiently search for a record in the**
>
> **data file.**
>
> **The index is usually specified on one field of the file called the indexing field.**
>
> **It contains index entries ordered by the field value with a pointer to its physical block.**

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

(04 marks)

ANSWER IN THIS BOX

Primary index: It is specified on the ordering key field of an ordered file of data

records. E.g. employee file ordered by employee-no key field

Clustering index: It is specified on the ordering non-key field of an ordered file of data

records. E.g. non-key department-no field

Secondary index: It is specified on any non-ordering field of a file of data records.

e.g. non-ordering employee-name field

[only two of above]

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

(01 mark)

ANSWER IN THIS BOX

Sparse or non-dense

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

(02 marks)

ANSWER IN THIS BOX

Database queries and transactions with respect to the files accessed, join conditions

used and other restrict conditions used in the query and its frequency

Update, insert and delete transactions and its frequency of use.

Performance requirements and applicable constraints

(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**
>
> **Whether to index an attribute?**
>
> **What attribute(s) to index on?**
>
> **Whether to set up a clustered index?**
>
> **Whether to use a hash index over a tree index?**
>
> **Whether to use dynamic hashing for the file?**
>
> [any four]

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

(02 marks)

> **ANSWER IN THIS BOX**
>
> **Database tuning is the process of continuing to revise/adjust the physical database**
>
> **design by monitoring resource utilization as well as internal DBMS processing to**
>
> **reveal bottlenecks such as contention for the same data or devices.**
>
> **Goal is to make application run faster; to lower the response time of queries /**
>
> **transactions and improve the overall throughput of transactions.**

(d) 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)

4

ANSWER IN THIS BOX

Clustered index on rental type and number of rooms > 3
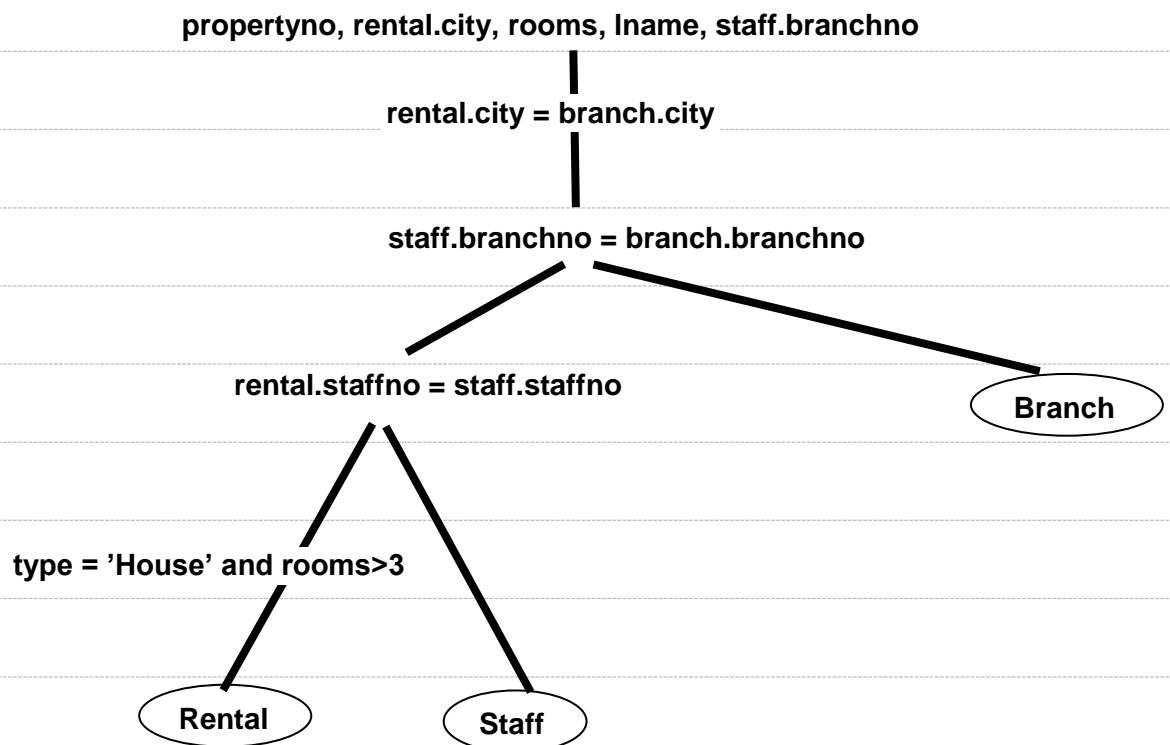
Or secondary indices on rtype and rooms

Filter house properties on both conditions

For selected property, obtain matching staffno & branchno using Foreign Keys

and determine if in same city.

Project required attributes

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

(05 marks)

ANSWER IN THIS BOX

propertyno, rental.city, rooms, lname, staff.branchno

rental.city = branch.city

staff.branchno = branch.branchno

rental.staffno = staff.staffno

Branch

type = 'House' and rooms>3

Rental

Staff

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

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

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) **Non-serializable as it cannot swap conflicting $r_3(a); w_1(a);$**

(i.e. read-write conflict) to form a serial schedule

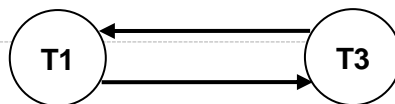(B) **Is Serializable as it can swap non-conflicting $r_2(a); r_3(a);$**

Serial Schedule is

$r_2(a);$ $r_3(a); w_3(a);$ $r_1(a); w_1(a);$ $[T_2, T_3, T_1]$

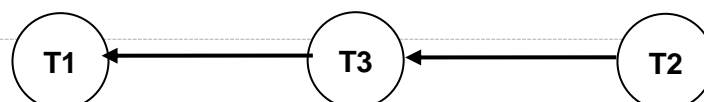(ii) Draw precedence graphs for two of the above two schedules.

(04 marks)

ANSWER IN THIS BOX

(A) **non-serializable schedule**

(B) **serializable schedule**

(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 =  r₁(a), r₂(c), r₁(c), r₃(a), r₃(b), w₁(a), c₁, w₃(b), c₃, r₂(b),
      w₂(c), w₂(b), c₂
```

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)			**S(a)**	
t2		READ(c)		**S(c)**	
t3	READ(c)			**S(c)**	
t4			READ(a)	**S(a)**	
t5			READ(b)	**S(b)**	
t6	WRITE(a)			**Wait for X(a)**	
t7	COMMIT			**wait**	
t8			WRITE(b)	**X(b)**	
t9			COMMIT		**Release S(a) & X(b)**
				Acquire X(a) for T1, Commit T1	
					& T1 release X(a) & S(c)
t10		READ(b)		**S(b)**	
t11		WRITE(c)		**X(c)**	
t12		WRITE(b)		**X(b)**	
t13		COMMIT			**T2 release X(b) & X(c)**

(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

[Start_Transaction, T3]

[Read_Item, T3, C]

[Write_Item, T3, B, 15, 12]

[Start_Transaction, T2]

[Read_Item, T2, B]
[Write_Item, T2, B, 12, 18]
[Start_Transaction, T1]
[Read_Item, T1, D]
[Write_Item, T1, D, 20, 25]
[Read_Item, T2, D]
[Commit T1]
[Write_Item, T1, D, 25, 26]
[Read_Item, T3, A]

(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

T3 has to be rollback because it did not reach the commit point.

As T2 has read a value changed by T3, cascading rollback is required for T2.

Therefore recovery process would redo only T1 using the log entries, subject to

identification of the completion of T1 via log.

T2 and T3 has to be reissued after the recovery process is over.

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)

<u>**ANSWER IN THIS BOX**</u>

Keeping track of data – keeping track of data distribution, fragmentation and

Replication by expanding system catalogue

Distributed query processing – ability to access remote sites and transmit queries

and data among the various sites via a communication network

Distributed transaction management – ability to devise execution strategies for

queries and transactions that access data from more than one site and to

synchronize data

Replicated data management – ability to decide which copy of a replicated data item

to access and to maintain the consistency of copies

Distributed database recovery – ability to recover from individual site crashes and

from new types of failures such as communication links

Security – execution of distributed transactions with proper authorization / access

privileges of users

<u>**ANSWER IN THIS BOX**</u>

(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

Book relation can be horizontally partitioned by Price as

   B_1: $\sigma_{Price \le 2500}$ **(Book),**

   B_2: $\sigma_{Price > 2500 \text{ and } Price < 5000}$ **(Book),**

   B_3: $\sigma_{Price \ge 5000}$ **(Book).**

BookStore relation can be horizontally partitioned by City as

   BS_1: $\sigma_{City = \text{"Colombo"}}$ **(BookStore),**

   BS_2: $\sigma_{City = \text{"Kandy"}}$ **(BookStore),**

   BS_3: $\sigma_{City \text{ not in } (\text{"Colombo"}, \text{"Kandy"})}$ **(BookStore).**

  (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

A derived horizontal partitioning of STOCK would be defined based on StoreNo

and BookNo attributes related to the above City and Price predicates used in

BookStore (BS) and Book (B) partitions respectively.

   $Stock_1$ = Stock $\infty_{StoreNo}$ BS_1 and Stock ∞_{BookNo} B_1

   $Stock_2$ = Stock $\infty_{StoreNo}$ BS_2 and Stock ∞_{BookNo} B_2

   $Stock_3$ = Stock $\infty_{StoreNo}$ BS_3 and Stock ∞_{BookNo} B_3

(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

Mobile computing architecture is distributed where a number of computers generally referred to as Fixed Hosts and Base Stations, are interconnected through high-speed wired network.

The entire database is distributed mainly among the wired components, possibly with full or partial replication. A base station or fixed host manages its own database with a DBMS like functionality, with additional functionality for locating mobile units and additional query and transaction management features to meet the requirements of mobile environment.

Another method is to distribute the database among wired and wireless components. Data management responsibility is shared among base stations or fixed hosts and mobile units.

(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

Text / Documents – Formatted or unformatted long text; Structured documents; books, magazines, HTML files

Images / Graphics – drawings, photographs and illustrations that are encoded

Continued…

---

using descriptive or standard formats (e.g. bitmap, JPEG and MPEG)

Video / Animation – temporal sequence of image or graphic data for presentation

at specified rates

Audio - stored recorded messages, such as speeches, class presentations, or

even surveillance recordings of phone messages or conversations

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

Using a DBMS to store the documents as text and retrieve when queried.

A relational or object-relational DBMS can be used to store while XML document as

text field within the DBMS records or objects.

Using a DBMS to store the document contents as data elements and retrieve when

queried.

A collection of documents that follow a specific XML DTD or XML schema would have

the same structure and hence can design a database to store the leaf-level data

elements within the XML document.

Designing a specialized system for storing native XML data and retrieve when queried.

New type of database system based on hierarchical model could be designed and

implemented.

Creating or publishing customized XML documents from pre-existing relational

Databases when queried. Data in existing databases can be formatted as XML for

exchanging or displaying over the web.

(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**
>
> **course (<u>cno</u>, title)**
>
> **taken_by (<u>cno</u>, <u>sno</u>, grade)**
>
> **student (<u>sno</u>, name)**
>
> **taken_by.sno and taken_by sno are FK; PK underline**

(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.
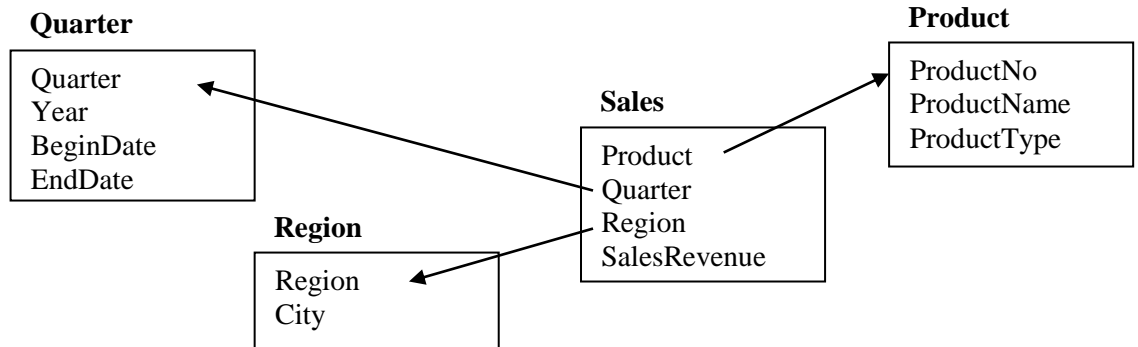
(03 marks)

> **ANSWER IN THIS BOX**
>
> **<course cno="01">**
>
>   **<title>Database</title>**
>
>   **<student sno="111">**
>
>   **<name>Dias</name>**
>
>   **<grade>A</grade>**
>
>   **</student>**
>
> **</course>**

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

Quarter
| Quarter |
| Year |
| BeginDate |
| EndDate |

Product
| ProductNo |
| ProductName |
| ProductType |

Sales
| Product |
| Quarter |
| Region |
| SalesRevenue |

Region
| Region |
| City |

(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)
Quarter	**Sales**
Region	
Product	

(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

Further summarise Quarterly Sales data of Products and obtain Annual Sales

of Products for all Regions.

Drill-down

Increase details Quarterly Product Sales showing by Cities than Region

Slice

Focus on Sales of Products a selected Quarter (e.g. Q1)

Dice

Select Sales for a ProductType (e.g. electronic items) for last two Quarters (e.g.

Q1 and Q2) forming a sub-cube.

(d) Data mining technologies are applied to a large variety of decision-making business applications. If a supermarket chain is to use those data mining technologies to improve its 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)

<u>**ANSWER IN THIS BOX**</u>

Marketing applications would analysis of consumer behaviour based on buying

patterns.

Possible marketing strategies include advertising, store location and targeted

mailing

Based on customer buying patterns can be identified

    how to organize the store,

    where to locate the store,

    what to advertise and

    whom to target etc.

****