THE OPEN UNIVERSITY OF SRI LANKA
Bachelor of Technology-Level 5
Department of Electrical & Computer Engineering
ECX5240 – Information Systems
Final Examination 2011/2012

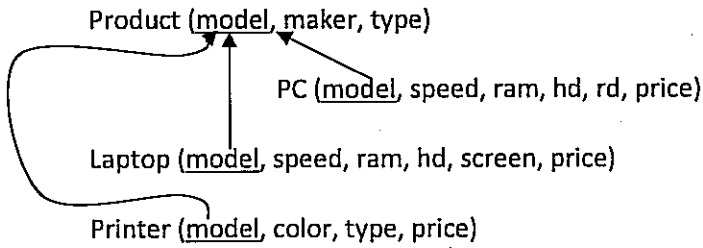
DATE: MARCH 21, 2012 **TIME: 0930-1230 HRS**

This question paper consists of eight questions.

Answer **FIVE** questions **ONLY.**

Question 1 – Database Programming (20 marks)

Consider the following schema:



The **Product** relation gives the manufacturer, model number and type (PC, laptop or printer) of various products.
The **PC** relation gives for each model number, that is a PC, the speed (of the processor, in megahertz), the amount of RAM (in megabytes), the size of the hard disk (in gigabytes), the speed and type of the removable disk (CD or DVD), and the price.
The **Laptop** relation is similar, except that the screen size (in inches) is recorded in place of information about the removal disk.
The **Printer** relation records for each printer model whether the printer produces color output (true, if so), the process type (laser, ink-jet or bubble), and the price.

Write the following queries in SQL:

a) Print the model number (i.e. *model*) and price (i.e. *price*) for all PCs and Laptops made by 'IBM' (i.e. *maker*).
(5 marks)

b) Print the maker who has the most number of Products.
(5 marks)

c) Write a trigger (called *tr_check_HDSize*) to ensure that for each PC its HD size is always higher than its RAM size.
(10 marks)

Question 2 – Relational Algebra

(20 marks)

a) Consider the following relations R1 and R2:

R1

A	B	C
a1	b1	c1
a2	b2	c2
a3	b3	c3

R2

X	Y	Z
c2	y2	z2
c3	y3	z3
c4	y4	z4

Use the above relations to perform the following relational algebra operators and write output relations.

i. $R1 \underset{R1.C = R2.X}{\bowtie} R2$ (3 marks)

ii. $R1 \underset{R1.C = R2.X}{\rhd\!\bowtie} R2$ (3 marks)

iii. $R1 \underset{R1.C = R2.X}{\bowtie\!\lhd} R2$ (3 marks)

iv. $R1 \underset{R1.C = R2.X}{\rhd\!\bowtie\!\lhd} R2$ (3 marks)

b) Consider the following relational schema.

Student (*snum, sname, major, level, age*)
Enrolled (*snum, cname*)
Class (*cname, time, room, lid*)
Lecturer (*lid, fname, deptid*)

Write the following queries in relational algebra.

i. Find the names of all Juniors (*Level = JR*). (3 marks)

ii. Find the names (*cname*) of all classes that were either held in room H501 (*room = H501*) or have hundred or more students enrolled. (5 marks)

Question 3 – Disk a nd Files (20 marks)

a) Explain the terms related to disk access; *Seek time*, *Rotational delay*, *Data transfer time* (3 marks)

b) Consider a disk with the following characteristics: block size 512 bytes, number of blocks per track 20, number of tracks per surface 400. A disk consists of 15 double-sided platters.

i What is the total capacity of a track in Kbytes? (2 marks)

ii. How many cylinders are there? (1 mark)

iii. What is the total capacity of a cylinder in Kbytes? (2 marks)

iv. What is the total capacity of disk in Kbytes? (2 marks)

i. Suppose the disk drive rotates the platters at a speed of 2400 rpm (revolutions per minute); what is the average rotational delay and data transfer time for a disk block in milli seconds (msec)? (2 marks)

ii. Suppose that the average seek time is 30 msec. How long does it take (on average in milliseconds) to locate and transfer a single block given its block address? (2 marks)

iii. Calculate the average time it would take to transfer 20 random blocks and compare it with the time it would take to transfer 20 consecutive blocks in same track. (2 marks)

c) Briefly describe the following record formats giving an example; *Fixed length record*, *Variable length record* (4 marks)

Question 4 – Indexing and Query Processing (20 marks)

a) What are the steps in Query Processing? Explain each step. (4 marks)

b) What is the justification for using I/O costs as the main measure in relational query optimizer? (2 marks)

c) Consider the following schema:

Emp(eid:integer; ename:string, age:integer, pno:integer)

Proj(pno:integer, pname:string, budget:float)

Consider the following query (Q1):

```
SELECT Emp.eid
FROM   Emp, Proj
WHERE Emp.pno = Proj.pno
```

i. Write two relational algebra expressions for the above SQL query in which, natural join is used for one and Cartesian product for the other.

(2 marks)

ii. Draw the graphical representation of the relational algebra expressions in (a.).

(2 marks)

iii. What query plan in (ii) would you suggest to execute on a DBMS and briefly explain the reasons and the query plan you have chosen?

(2 marks)

iv. Re-write the above SQL query (Q1), which prints employees that have been assigned to projects without using a join.

(2 marks)

v. What index would you suggest to speed up the query in (iv)? Assume that DBMS supports only B+ tree indexes.

(2 marks)

vi. Consider the following information.

Employee (Emp) relation consists of 1000 pages with 100 tuples per page. Project (Proj) relation consists of 500 pages with 200 tuples per page.
Assume that 90% of employees have been assigned to projects and all attributes of employee table are equal in size. You are free to assume all null values are placed in together in index file.
Estimate the total cost for the query plan in (v.) ignoring output costs. State any assumptions you made.

(4 marks)

Question 5 – Transactions and Concurrency Control (20 marks)

a) Briefly explain the following terms in database transactions:
Serial Schedule and *Serializable Schedule*. (2 marks)

b) Briefly explain the rules of Strict 2 Phase Locking Protocol. (3 marks)

c) List the possible issues that can occur when the transactions are interleaved. (2 marks)

d) Consider the following part of a transaction schedule.

T_1	T_2	T_3	T_4
S(A)			
R(A)			
	X(B)		
	W(B)		
S(B)			
R(B)			
			S(D)
			R(D)
	X(A)		
	W(A)		
		S(C)	
		R(C)	
			S(C)
			R(C)

Assume that Transaction T_i is higher priority than transaction T_{i+1}(i.e. transaction T_1 has higher priority than T_2; T_2 has higher priority than T_3; and T_3 has higher priority than T_4).

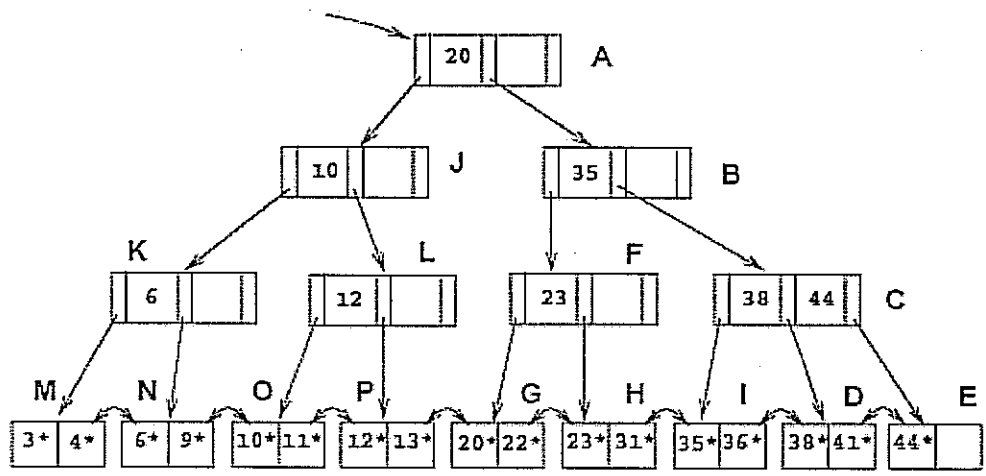
i. How does the DBMS detect a deadlock? (2 marks)

ii. How the DBMS resolves deadlock (after it has occurred). (1 marks)

iii. Draw a wait-for graph for the schedule given above to detect the deadlock. (2 marks)

iv. Briefly explain what the "phantom read problem " is? (2 marks)

e) Consider the tree shown below.



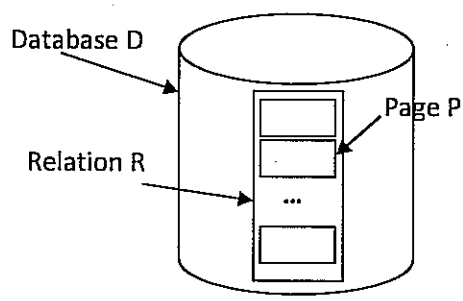
Describe the steps involved in executing each of the following operations according to the *Simple Tree Locking algorithm*, in terms of the order in which nodes are locked, unlocked, read and written. Be specific about what type of locks are obtained and released.

i. Search 40. (1 marks)

ii. Delete 9. (1.5 marks)

iii. Insert 40. (1.5 marks)

f) Consider the following scenario:



Database *D* contains a relation *R*. Relation *R* contains 1000 pages. Assume that multiple granularity locking scheme is used.

Describe the locks acquired when reading all pages in relation R and modify about 5 pages, which can be identified only after reading relation R. (2 marks)

Question 6 – Crash Recovery **(20 marks)**

a) What are the roles of Analysis, Redo and Undo phases in ARIES? (4 marks)

b) Explain the WAL Protocol. (4 marks)

c) Describe the STEAL and NO FORCE policies. (6 marks)

d) Consider the following log file.

LSN	LOG
00,05	begin_checkpoint, end_checkpoint
10	update: T1 writes P5
20	update T2 writes P3
30	T1 abort
40	CLR: Undo T1 LSN 10
45	T1 End
50	update: T3 writes P1
60	update: T2 writes P5
70	T2 Commit
	CRASH, RESTART

Show the log after the Recovery. Assume that ARIES algorithm is used in the recovery process. (6 marks)

Question 7 – Data Mining and Warehousing **(20 marks)**

a) What is a Materialized view and what are the advantages of the materialized view over the other views?

(3 marks)

b) Consider the following star schema of an electrical equipment company used to analyze the sales of its products.

Sales (ProductId, CustomerId, date, amount)
Product (id, name, price)

The primary keys are underlined. Suppose that the management wants to calculate the amount of products and the names of the product that are sold in a particular date. Using SQL, define a (materialized) view named TSOLDAMT that can be used to process this query efficiently.

(4 marks)

c) Using the view you have defined answer the query. Find the total amount sold in each product, between March 1st, 2011 and March 10th, 2011.

(3 marks)

d) Define the terms of Support and the Confident.

(2 marks)

e) Consider the transactions which are in the Table 1.

Transaction ID	Items brought
100	A,B,C
101	A,C
102	A,D
103	B,E,F

Table 1

i. Calculate the support and confidence for the Association Rules below.
A=>C and C=>A (4 marks)

ii. Considering the association rules in part (a), state which association rules meet the minimum support of 50% and minimum confidence of 75%. (1 marks)

ii. Find all frequent Itemsets with a minimum support of 60% and minimum confidence of 70%. (Note - Use Apriori algorithm)

(3 marks)

Question 8 – Data Mining and Warehousing (20 marks)

a) Briefly explain main approaches to store XML data.

(4 marks)

b) Consider the following *bib.xml* file for the following questions given below. Note: answers should contain FLWOR expression and expected output of it.

```
<bib>
    <book year="1994" type="paper">
        <title>TCP/IP Illustrated</title>
        <author>Stevens, W.</author>
        <publisher>Addison-Wesley</publisher>
        <price> 965.95</price>
    </book>

    <book year="1992" type="paper">
        <title>Advanced Programming in the Unix environment</title>
        <author>Stevens, W.</author>
        <publisher>Addison-Wesley</publisher>
        <price>1165.95</price>
    </book>

    <book year="2000" type="paper">
        <title>Data on the Web</title>
        <author>Abiteboul, Serge</author>
        <author>Buneman, Peter</author>
        <author>Suciu, Dan</author>
        <publisher>Morgan Kaufmann Publishers</publisher>
        <price> 39.95</price>
    </book>

    <book year="1999" type="journal">
        <title>The Digital TV</title>
        <editor>Gerbarg, Darcy</editor>
        <publisher>Kluwer Academic Publishers</publisher>
        <price>129.95</price>
    </book>
</bib>
```

i. Find all book titles published after 1995.

(4 marks)

ii. Returns the title element of all books which are published by *Addison-Wesley* and less expensive than 1000.00.

(4 marks)

iii. Find books whose price is larger than average:

(4 marks)

iv. Briefly explain the following FLWOR expression with output.

```
FOR $h IN //book
ORDERBY $h/title
RETURN <holding>
                    $h/title,
                    IF $h/@type = "journal"
                              THEN $h/editor
                    ELSE $h/author
        </holding>
```

(4 marks)