

THE OPEN UNIVERSITY OF SRI LANKA
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING



FINAL EXAMINATION 2011
BACHELOR OF SOFTWARE ENGINEERING – LEVEL 5

ECI5266 – ADVANCE DATABASE SYSTEMS

DATE: MARCH 30, 2011

TIME: 1400-1700 HRS

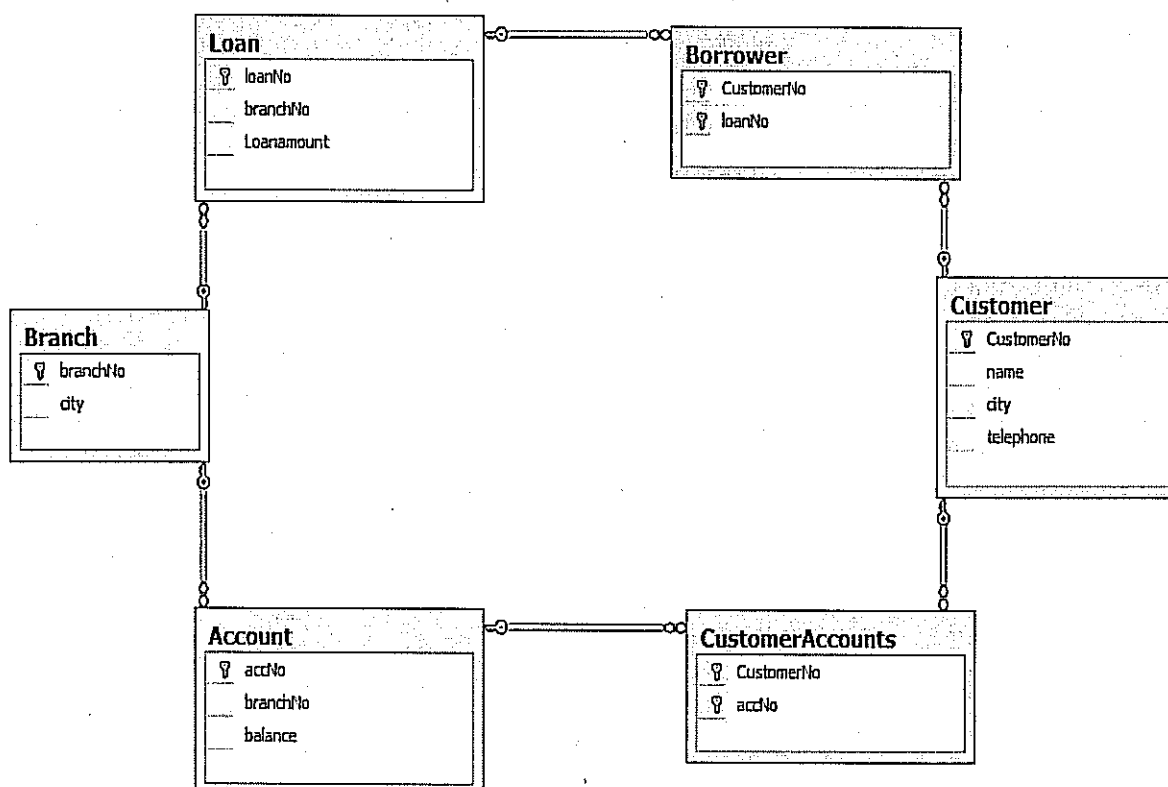
This question paper consists of eight questions.

Answer FIVE questions ONLY.

Question 1 – Database Programming

(20 marks)

Consider the following logical database schema which contains information about customers (*Customer* relation), branch (*Branch* relation), accounts (*Account* relation), Accounts assigned for customers (*CustomerAccounts* relation), loans (*Loans* relation) and Borrower (*Borrower* relation).



- (a) For all customers who have taken a loan, list the customer number and total amount borrowed. Some customers may have taken few loans. Write an SQL query to obtain details. (3 marks)
- (b) List the names of all customers who have taken loans from the bank but do not have an account in the same bank branch. Write an SQL query to obtain details. (5 marks)
- (c) Write a stored procedure to return number of bank loans for a given customer. (6 marks)
- (d) Write a trigger to ensure that total loan amount not more than 5,000,000 for each customer. (6 marks)

Question 2 – Relational Algebra

(20 marks)

Consider the following schema:

Employee (SSN, Fname, Lname, BDate, addr, salary, DNum)
 Department (DNum, Dname, mgrSSN, mgrStartDate)
 Works_On (ESSN, Pno, Hours)
 Project (Pno, Pname, PLocation, DNum)
 Dependent (ESSN, Depend name, BDate, relationship)

Write the following queries in **relational algebra**.

- (a) For every project located in 'Colombo', list the project number, the controlling department number, and the department manager's last name, address, and birth date. (5 marks)
- (b) List the names of all employees who have two or more dependents. (5 marks)
- (c) List the names of managers who do not have projects assigned yet. Assume that department manager become a manager of all the projects which are assigned to the department. (5 marks)
- (d) List the SSN of employees who works more than 50 hours in all projects assigned to them but salary less than 50000/=. (5 marks)

Question 3 – Disk, Files and Indexes

(20 marks)

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

(6 marks)

(b) Consider a disk with sector size of 512 bytes, 4000 tracks per surface, 20 sectors per track, and five single-sided platters. Each disk block spans two sectors. The disk platters rotate at 6000 rpm (revolutions per minute).

i. What is the capacity of the disk in bytes?

(3 marks)

ii. What is the average access time for a disk block if the average seek time is 100 msec (0.1 seconds)? Assume $\frac{1}{2}$ revolution for average rotational delay. You can transfer 20 sectors in a single revolution.

Note: Disk Access Time = Seek Time + Rotational Delay + Transfer Time

(5 marks)

(c) Which of the three basic file organizations would you choose for a file where the most frequent operations are as follows?

- i. Search for records based on a range of field values.
- ii. Perform inserts and scans where the order of records does not matter
- iii. Search for records based on a particular field value

(3 marks)

(d) Consider a relation stored as a randomly ordered file for which the only index is an unclustered index on a field called *sal*. If you want to retrieve all records with *sal* > 20, is using the index always the best alternative? Explain.

(3 marks)

Question 4 – Query Processing

(20 marks)

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

(3 marks)

(b.) Consider the following schema:

Sailors (sid, sname, rating, age)
 Boats (bid, bname, color)
 Reserves (sid, bid, day)

Consider the following query:

```
SELECT S.sname
FROM Sailors S, Reserves R, Boats B
WHERE S.sid = R.sid AND R.bid = B.bid AND B.color = 'red'
```

- i. Write the relational algebra expression which gives the best performance for the above query.
(3 marks)
- ii. Draw the graphical representation (Query Tree) of the relational algebra expression.
(3 marks)
- iii. Recommend indexes for higher performance of the above query and briefly explain.
(6 marks)
- iv. Both *Boats* and *Reserves* tables have clustered B+ tree indexes on *bid*. Assume that the indexes on *Boats* and *Reserves* have 100 and 1000 leaf pages respectively and the heights of their B+ trees are 2 and 3 respectively.

Estimate the cost of performing a sort merge join, briefly explaining the steps in your calculation. (Note that a direct data page scan is not guaranteed to retrieve all the records in sorted order of the index search key.)

(5 marks)

Question 5 – Transactions and Concurrency Control

(20 marks)

- (a.) List the properties of a transaction. Briefly explain each of them.
(4 marks)
- (b.) Explain briefly with an example the term *unrecoverable schedule*.
(3 marks)
- (c.) Briefly explain rules of the Strict two Phase Locking (Strict 2PL) Protocol.
(3 marks)
- (d.) Consider the following sequence of actions, listed in the order they are submitted to the DBMS:

Sequence: $T_1:R(X)$, $T_2:W(X)$, $T_2:W(Y)$, $T_3:W(Y)$, $T_1:W(Y)$, $T_1:Commit$, $T_2:Commit$, $T_3:Commit$.

Considering Strict 2 Phase Locking Protocol, draw the schedule for the above sequence.

(2 marks)

- (e.) Consider the following schedule. The transactions are involved in a deadlock. 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).

T ₁	T ₂	T ₃	T ₄
S(A)			
R(A)			
	X(B)		
	W(B)		
S(B)			
		S(C)	
		R(C)	
	X(C)		
			X(B)
		X(A)	

- i. Draw a wait-for-graph to show availability of the deadlock. (4 marks)
- ii. Draw the schedules considering the following deadlock prevention algorithm:
Wait-die approach (4 marks)

Question 6 – Data Mining and Warehousing (20 marks)

- (a.) Name two approaches to build the data warehouse and briefly explain each. (2 marks)
- (b.) What is “Star Schema”? What are the main two table types Star Schema has? Briefly explain them (6 marks)
- (c.) What is a Materialized view and what are the advantages of the materialized view over the other views? (3 marks)
- (d.) Consider the following transactions with regard to a computer store that sells computers and related equipments.

TID	ITEMS
100	PC, Notebook, Printer
200	Notebook, Speakers, Scanners
300	PC, Notebook, Printer
400	Speakers, PC, Printer

- Find all associative rules with a minimum support of 60% and minimum confidence of 70%.
(Note - Use Apriori algorithm) (9 marks)

Question 7 – XML Data**(20 marks)**

(a.) Briefly explain main approaches to store XML data.

(4 marks)

(b.) Consider the *catalog.xml* file and translate the following queries into FLWOR expressions.

Note: answers should contain FLWOR expression and expected output of it.

```

<catalog>
  <product dept="WMN">
    <number>557</number>
    <name language="en">Fleece Pullover</name>
    <colorChoices>navy black</colorChoices>
  </product>
  <product dept="ACC">
    <number>563</number>
    <name language="en">Floppy Sun Hat</name>
  </product>
  <product dept="ACC">
    <number>443</number>
    <name language="en">Deluxe Travel Bag</name>
  </product>
  <product dept="MEN">
    <number>784</number>
    <name language="en">Cotton Dress Shirt</name>
    <colorChoices>white gray</colorChoices>
    <desc>Our favorite shirt!</desc>
  </product>
</catalog>

```

- i. Return the names of all products in the "ACC" department. (4 marks)
- ii. Return the name elements of all products that are ordered on the product name. (4 marks)
- iii. Return the name element of all products in which the number is less than 500. (4 marks)
- iv. Return all product elements in which the number is less than 500 and the department is ACC. (4 marks)

Question 8 – Trends in Databases**(20 marks)**

- i. Briefly explain three characteristics of data warehouses. (5 marks)
- ii. Data mining technologies can be used in variety of decision-making contexts in organizations. What are the knowledge discovery areas in data mining? (5 marks)
- iii. Explain major impacts high speed networks have on the design of distributed database systems. (5 marks)
- iv. Electronic profiling process tracks and store visitor website behavior mainly to provide a personalized customer website experience. Explain 3 techniques that can be used to avoid unethical access to data collected in such databases. (5 marks)

-All Rights Reserved-