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

FINAL EXAMINATION 2013
BACHELOR OF SOFTWARE ENGINEERING – LEVEL 5

ECI5266 – ADVANCED DATABASE SYSTEMS

DATE: AUGUST 21, 2013 **TIME: 0930 - 1230 HRS**

This question paper consists of eight questions. Answer **FIVE** questions **ONLY.**

Question 1 – Database Programming (20 marks)

Consider the following relations in a movie database.

```
Movie (title:char(25), year:int, length:float, language:char(15), filmtype:char(1))

StarsIn (movieTitle:char(25), movieYear:int, starname:char(15), role:varchar(15))

MovieStar (name: char(15), address:varchar(40), gender:char(1), birthdate: date)
```

The attributes of the `Movie` relation are `title` of the movie, `year` it was made, `length` of the movie, `language`, and `filmtype` which may be 'F' or 'D' for feature or documentary respectively. The `MovieStar` relation has attributes to record the `name`, `address`, `gender` ('M' or 'F'), and `birthday` of stars. `StarsIn` relation associates the movie with stars that acted in them and contains the `role` ('lead', 'support', or 'other') they played. The primary keys of all relations are underlined.

a) For each English language feature movie, get the title, year, and the number of stars who acted in the movie.

(3 marks)

b) Display the names of stars who have acted in 5 or more movies in any year between 2006 and 2011. Display the names without duplicates.

(3 marks)

c) It has been decided to introduce a rating system for the movie stars. A star gets one point for each support role and no points for any other roles. Do the following to implement it.

i. To the MovieStar table, add a new attribute named rating of float type with a default value of zero.

(2 marks)

ii. Write a T-SQL procedure to update the rating of all movie stars.

(6 marks)

iii. Write a trigger to update the rating whenever a new tuple is added to the StarsIn table.

(6 marks)

Question 2 – Object Relational Databases (20 marks)

Consider the following object relational database schema for an insurance policies database:

Object types:
Customer_t (cid: char(6), name: varchar(15), birthdate: date, phone:char(10), address: varchar(50))
Car_t (regno: char(9), make: varchar(12), model: varchar(10), mdate: date, owner: ref customer_t, value: number(8,2))
Claim_t (claimno: char(12), cdate: date, amount: number(8,2), claimant: ref customer)
Claim_ntab table of claim_t
Policy_t (pid: char(7), sdate: date, edate: date, inscar: ref car_t, premium: number(6,2), claims: claim_ntab)

Tables:
Customers of Customer _t (cid primary key)
Cars of car_t (regno primary key, owner references Customers)
Policies of policy_t (pid primary key, inscar references Cars)
Nested table claims store as claims_ntable

- The Customers, Cars, and Policies tables contain tuples for all customers, cars and policies respectively.
- Attributes of Customers are customer id (cid), name, date of birth, phone and address. Attributes of Cars are registration number (regno), make, model, date of manufacture, owner of the car, and insured value of the car.
- Attributes of Policies policy id (pid), starting date, ending date, insured car, annual premium and the claims made.
- Attributes of the nested table of claims are claim number (claimno), claim date (cdate), the amount of claim and the claimant.
- The attribute types are specified in the type descriptions above, as also are the primary keys and referential constraints in the table schema.

a) Write Oracle OR-SQL statements to answer the following query (use columns of REF type instead of joins to link tables).
Find the average insurance premium on cars owned by customers aged between 20 and 25 years. sysdate contains current date and the function month_between(d1,d2) gives the number of months in decimals between the dates d1 and d2 where d1>d2.

(4 marks)

b) Write Oracle OR-SQL to insert a claim against an existing policy that has a pid of SL12354, given the following claim details: claim number: 001, claim date: 12 July 2004, claim amount: 2000, and customer id of claimant: S25431. Assume that the claimant is already present as a customer in the database.

(4 marks)

c) Write Oracle object SQL to calculate the renewal premium for a given policy in the database using the following logic: If the policy had no claims or the total claim is less than 1000, then the new premium is the same as the current premium. If the total claim on a policy is greater than or equal to 1000, then the new premium is to be 20% more than the current.

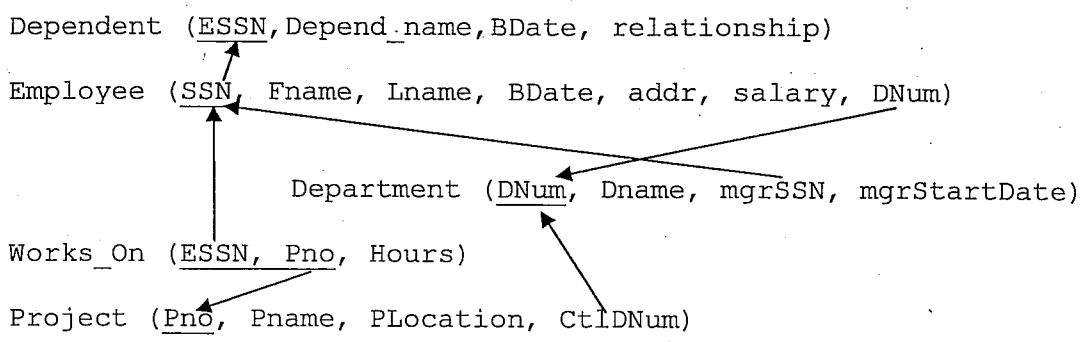
(9 marks)

d) Using the method defined above in (c), get the renewal premium for the policy of a car with registration number SLA984.

(3 marks)

Question 3 – Relational Algebra **(20 marks)**

Consider the following schema:



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.

(4 marks)

b) List the names of managers who have at least one dependent.

(4 marks)

c) List the names of all employees with two or more dependents.

(4 marks)

d) List the names of all employees, working department name and his/her manager name.

(4 marks)

e) List the name of all employees and number of projects they are working. Number of projects should be printed as zero for employees who are not doing any projects.

(4 marks)

Question 4 – Disk and Files **(20 marks)**

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

(5 marks)

b) Consider a disk with sector size of 512 bytes, 2000 tracks per surface, 20 sectors per track, and five single-sided platters. Each disk block spans two sectors. The disk platters rotate at 3000 rpm (revolutions per minute).

i. What is the average access time for a disk block if the average seek time is 100 msec (0.1 seconds)? Assume ½ revolution for average rotational delay. You can transfer 20 sectors in a single revolution.

(4 marks)

ii. What is the capacity of the disk?

(3 marks)

c) In RAID Level 5, the parity blocks are distributed on multiple disks, when compared to RAID Level 4. What is the advantage of doing this?

(4 marks)

d) Operating System does disk space management and buffer management. Briefly explain three advantages of DBMSs' doing their own disk space management and buffer management.

(4 marks)

Question 5 – 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) Briefly explain clustered indexes and un- clustered indexes.

(4 marks)

d) Consider the following schema:

```
Item (i_num:integer; i_name:string, unit_price:float)

Order (o_num:integer, o_date:date, i_num:integer, qty:integer )
```

Consider the following query:

```
SELECT      Item.i_name
FROM        Item, Order
WHERE       Item.i_num = Order.i_num AND Order.o_num = '1234'
```

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. Make sure to do selection before the join operation.

(2 marks)

ii. Draw the graphical representations for the relational algebra expressions given in answers to question 5 d) i.

(2 marks)

iii. What query plan given in part d) ii.of this question would you suggest to execute on a DBMS? Briefly explain the query plan?

(3 marks)

iv. What indexes would you suggest to speed up the query plan considered in part d) (iii.) of this question? Assume that DBMS supports only B+ tree indexes.

(3 marks)

Question 6 – Physical Database Design and Tuning **(20 marks)**

Consider the following relation:

`student (sid:char(10), sname:varchar(100), address:varchar(200), gpa:float)`

The relation has 10000 pages with 100 tuples per page.

a) Consider the following query:

Query 1: `SELECT * FROM Student WHERE sid = 'DIT0001'`

Estimate the cost (in Disk I/Os) of executing Query 1 with the following indexes. The data entries in the index have <k, rid> format. Hash indexes require 1.2 Disk I/Os on average to retrieve a data entry. B+ tree indexes require 3 disk I/Os to retrieve a leaf page.

Estimate the cost (in Disk I/Os) of executing Query 2 with the following indexes.

i. Unclustered hash index on Student <sid>

(3 marks)

ii. Clustered hash index on Student <sid>

(3 marks)

b) Consider the following query:
Query 2: `SELECT * FROM Student WHERE gpa = 3.0;`

Assume that 0.5% of tuples meet the selection criteria (`gpa = 3.0`). The data entries in the index have <k,rid> format and 20% the size of the actual record. Hash indexes require 1.2 Disk I/Os on average to retrieve a data entry. B+ tree indexes require 3 disk I/Os to retrieve a leaf page.

Estimate the cost (in Disk I/Os) of executing Query 2 with the following indexes.

i. Unclustered B+ tree index on Student<gpa>

(5 marks)

ii. Clustered B+ tree index on Student<gpa>

(5 marks)

c) Consider the following query:

```
SELECT      count(*)
FROM        Student
GROUP BY    gpa
```

Estimate the cost (in Disk I/Os) of executing above query without using any indexes.

(4 marks)

Question 7 – Transactions and Concurrency Control **(20 marks)**

a) Briefly explain the properties of a transaction.

(4 marks)

b) Briefly explain the rules in *Strict 2 Phase Locking Protocol.*

(3 marks)

c) Consider a database with objects X and Y and assume that there are two transactions T1 and T2. Transaction T1 reads object X and Y and then writes object X. Transaction T2 reads objects X and Y and then writes objects X and Y. Both T1 and T2 commit after all read and write actions of them.

i. Give an example schedule with actions of transactions T1 and T2 on objects X and Y that results in a write-read conflict.

(2 marks)

ii. Give an example schedule with actions of transactions T1 and T2 on objects X and Y that results in a read-write conflict.

(2 marks)

iii. Give an example schedule with actions of transactions T1 and T2 on objects X and Y that results in a write-write conflict.

(2 marks)

iv. For each of the three schedules, show that Strict 2PL disallows the conflicts.

(2 marks)

d) Consider the following sequence of actions, listed in the order they are submitted to the DBMS. The Strict 2PL has been used for concurrency control.
T1:R(P), T2:W(Q), T2:W(P), T3:R(R), T3:R(Q), T1: W(R)

T_1	T_2	T_3
S(P)		
R(P)		
	X(Q)	
	W(Q)	
	X(P)	
	W(P)	
		S(R)
		R(R)
		S(Q)
		R(Q)
X(R)		
W(R)		

Assume that older transaction has higher priority always. The DBMS processes actions in the order shown.

i. Follow *wound-wait policy* to deal with deadlock in above schedule. Draw schedule again.

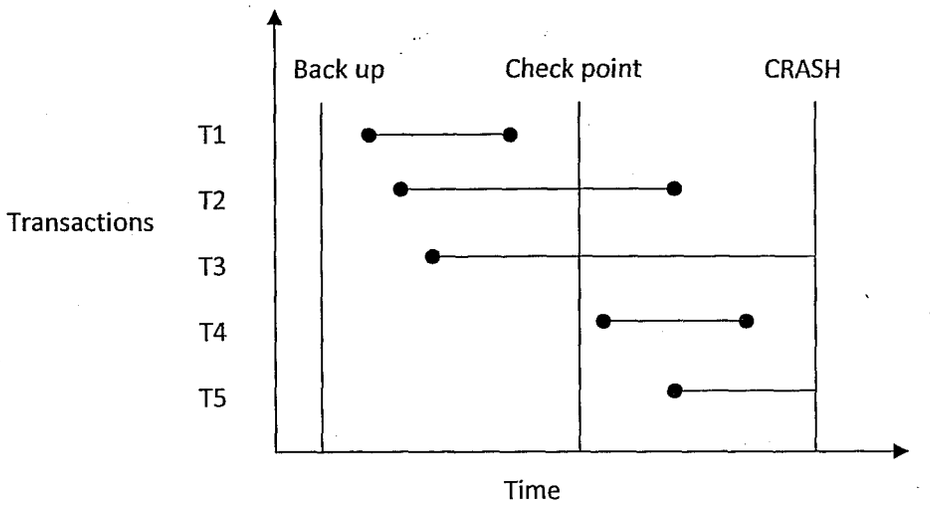
(3 marks)

ii. Follow deadlock detection approach to deal with deadlocks in above schedules. Draw a *wait-for graph* for the schedule given above to detect the deadlock and briefly explain how to identify the deadlock and remove.

(2 marks)

Question 8 – Crash Recovery **(20 marks)**

a) Consider the diagram below:



What is the desired state that the database should be in after Crash Recovery? Briefly explain.

(5 marks)

b) Explain the *Write-Ahead Logging* (WAL) Protocol.

(4 marks)

c) How does WAL protocol assist to ensure *Atomicity* and *Durability* in a STEAL-NO FORCE approach?

(5 marks)

d) What are the roles of Analysis, Redo and Undo phases in ARIES?

(6 marks)