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

FINAL EXAMINATION 2017
BACHELOR OF TECHNOLOGY – LEVEL 5

ECX5245 –DATABASE MANAGEMENT SYSTEMS

DATE: 5 DECEMBER 2017	TIME: 0930-1230 HRS

This question paper consists of Seven (07) questions. Answer **FIVE** (05) questions **ONLY**.

Question 1 – Database Programming   (20 marks)

Consider the following relations containing airline flight information:

```
Flights (flno, from, to, distance, departs, arrives)
Aircraft (aid, aname, cruisingrange)
Certified (eid, aid)
Employees (eid, ename, salary)
```

Note that the Employees relation describes pilots and other kinds of employees as well; every pilot is certified for some aircraft (otherwise, he or she would not qualify as a pilot.), and only pilots are certified to fly.

Write the following queries in **SQL or T-SQL**.

(a) Find the *eids* of employees who are certified for exactly three aircraft.

(5 marks)

(b) Find the *eids* of employees who are certified for the largest number of aircraft.

(6 marks)

(c) Write a trigger (called *tr_salUpdates*) to monitor salary changes in Employee table.

Your trigger should insert a record to *salChanges* table for each modification of salary of employees in order to track the changes. Assume no primary key in *salChanges* table.

```
salChanges (eid, newSalary, datatime)
```

Note that the salChanges table contains employee identifier (*eid*), new salary (*newSalary*) and change date and time (*datetime*).

(9 marks)

Question 2 – Relational Algebra   (20 marks)

Consider the following relations containing airline flight information:

```
Flights (flno, from, to, distance, depart, arrives)

Aircraft (aid, aname, travelRange)

Certified (eid, aid)

Employee (eid, ename, salary)
```

Note that the Employee relation includes pilots and other kinds of employees as well; every pilot is certified for some aircraft (otherwise, he or she would not qualify as a pilot), and only pilots are certified to fly.

Write the following queries in **Relational Algebra.**

(a) Find the names of the pilots certified for some Boeing (i.e. aname) aircraft.

(3 marks)

(b) Identify and print eid and ename of all the pilots whose salary is more than Rs. 100,000/=.

(3 marks)

(c) Find the aid of all aircraft that can be used on non-stop flights from Madras to Colombo. (Hint: consider travelRange > distance to identify non-stop flights)

(4 marks)

(d) Find *ename* of pilots who have been assigned more than three aircrafts.

(5 marks)

(e) Find *eid* and *ename* of **all** employees and print aircraft information (i.e. *aid*) for pilots.

(5 marks)

Question 3 – Object Relational Databases   (20 marks)

Consider the following object relational database schema for a Movies database:

Object types:
Actor_t (*name*: varchar(15), *gender*:char(1), *rating*: real, *bestMovie*: ref movie_t)
Character_t (*name*: varchar(15), *role*: varchar(10), *played*: ref actor_t)
Characters_nt table of Character_t
Movie_t (*title*: varchar(20), *year*:integer, *length*:float, *filmtype*: varchar(10), *characters*: characters_nt)

Tables:
Actors of actor_t (*name* primary key, *bestMovie* references *movies*)
Movies of movie_t (*title* primary key) Nested table *characters* store as characters_ntb

The tables named Actors and Movies contain tuples for all actors and movies respectively. Attributes of Actors are name, gender, rating, and the best movie of the actor. Attributes of Movies are title, year, length, film type, and the characters of the movie. The attributes of the nested table of characters are the name of the character, role and the actor who played the role. 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 OBJECT SQL statements to answer the following queries (use columns of REF type instead of joins to link tables):

i. Get the title of actress *Gamini Fonseka's* best movie and the name(s) of the character(s) he played in that movie.

(4 marks)

ii. Get the title and year of all movies in which the actor *Joe Abeywickrama* played two or more roles.

(5 marks)

(b) It is required to add a member method called *cost_of_actors* to get an estimate of the total cost of the actors in a movie. For this estimate, assume that the payment for a given role by an actor is obtained by multiplying a base rate in rupees with the actor's rating. The total cost of all the roles gives the cost of actors for the movie. The base rate in rupees will be given as a parameter to the method.

Write Oracle SQL statements to modify the object type *movie_t* by adding this method specification.

(8 marks)

(c) Using the method defined above, write an Oracle SQL statement to display the title and estimated cost of actors, of all feature films (*filmtype* = 'feature') produced in 2017. Assume the base rate to be Rs. 100,000.

(3 marks)

Question 4 – Disk, Files and Indexes   (20 marks)

(a) You have a large file that is frequently accessed sequentially. Briefly explain how you would store the pages in the file on a disk for fast querying.

(2 marks)

(b) Consider a disk with a sector size of 512 bytes, a block size of 1024 bytes, 2000 tracks per surface, 50 sectors per track, 5 double-sided platters, average seek time of 10msec.

i. What is the capacity of a track in blocks?

(1 mark)

ii. What is the capacity of a cylinder in blocks?

(1 mark)

iii. What is the capacity of the disk in blocks?

(1 mark)

iv. If disk platter rotates at 5400 rpm (revolutions per minute), what is the average rotational delay?

(2 marks)

v. What is the access time to read two consecutive disk blocks?

(2 marks)

(c) An index on a database table can be either clustered or unclustered. From the following, choose the statements that are true.

    A. A table can have only one unclustered index but may have any number of clustered indexes.

    B. A table can have only one clustered index but may have any number of unclustered indexes.

    C. Both B+-tree and hash indexes can be clustered.

    D. Only B+-tree indexes can be clustered.

    E. With a clustered index, the corresponding data pages are not always in sorted order.

    F. None of the above.

(3 marks)

(d) Consider the following relation:

    Emp (eid: integer, ename:string, age: integer, salary: float)

On this table, there is a dense clustered B+ tree index on *eid* and an unclustered B+ tree index on *age*.

    i. The data records are stored in a heap file. In what order are the data records of Emp table stored and why?

(1 mark)

    ii. Consider the following query:

```
Select ename, salary
From emp
Where age>18;
```

If 95% of tuples satisfy the selection condition, what would be the best access path for processing this query? Justify your answer.

(3 marks)

    iii. Consider the following query:

```
Select age, avg(salary)
From emp
Group by age;
```

Would a clustered index on <age, salary> be useful? Explain briefly.

(4 marks)

Question 5 – Query Processing   (20 marks)

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

(3 marks)

(b) Estimate the number of I/Os for sorting 1000 pages using 20 buffer pages. Indicate the different passes and what happens in each pass.

(5 marks)

(c) Consider the following schema:

```
Employee (eno, ename, address, salary, pno)
Project (pno, pname, description, budget, head
```

There are 10000 employee records on 100 pages and 1000 project records on 10 pages. *Employee* and *Project* relations have clustered B+ tree indexes on *Employee<pno>* and *Project<pno>* fields respectively. Assume equal sized fields for *Employee* and *Project* relations. Also assume that the employee and project records are in sorted order because of the existing clustered indexes. There are 20 buffer pages available.

Consider the following query:

```
SELECT     e.*
FROM       Employee e, Project p
WHERE      e.pno = p.pno AND p.budget > 25000000
ORDER BY   e.pno
```

Assume that 5% of the tuples in *Project* relation meets the selection condition. Estimate the cost of the best plan possible, assuming that you can create indexe(s) of your choice. Describe the I/O cost for the best plan you have considered.

(12 marks)

Question 6 – Transactions and Concurrency Control   (20 marks)

(a)   Briefly explain the properties of a transaction.

(4 marks)

(b)   Briefly explain, what a *Serializable Schedule* is?

(2 marks)

(c)   Briefly explain the rules in Strict 2 Phase Locking Protocol.

(3 marks)

(d)   Briefly explain what is the *phantom problem*?

(3 marks)

(e)   Consider the following sequence of actions, listed in the order they are submitted to the DBMS. The DBMS processes actions in the order shown. The **Strict 2PL** has used for concurrency control. If a transaction is blocked, assume that all of its actions are queued until it is resumed; the DBMS continues with the next action of an unblocked transaction.

T1:R(X), T2:W(X), T2:W(Y), T3:W(Y), T1:W(Y),T3W(X), T1: Commit, T2: Commit, T3: Commit

Assume that older transaction has higher priority always.

i.   Draw a transaction schedule and briefly explain how above three transactions are running till commit those without deadlocks in schedule. Follow *Wait-Die* policy to deal with deadlock.

(4 marks)

ii.   Follow deadlock detection approach to deal with deadlocks. Draw a transaction schedule and explain how three transactions are running. Show a *waits-for graph* if a deadlock cycle develop.

(4 marks)

(a.) Consider the diagram below:
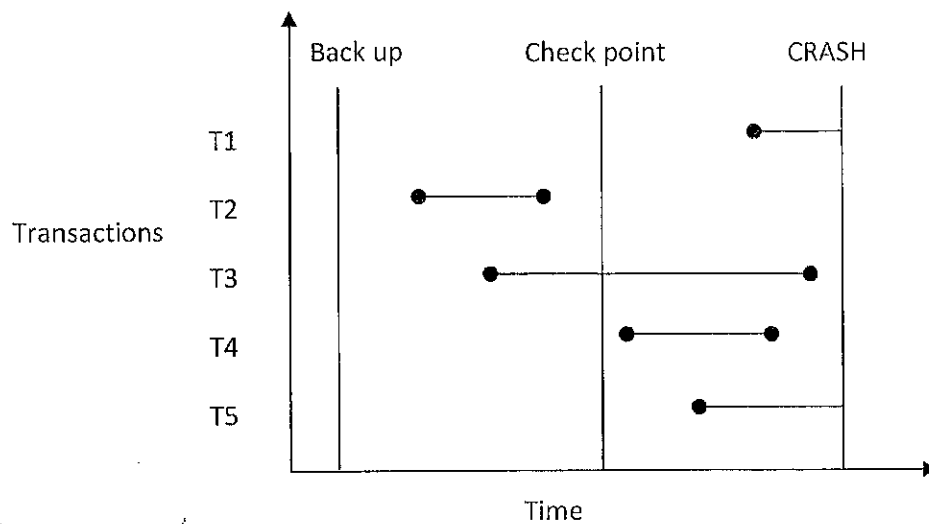
What is the desired state that the database should be in after Crash Recovery? Briefly explain.

(5 marks)

(b.) What is meant by Write-Ahead Logging. Why it is necessary to implement this protocol in a DBMS?

(5 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?

(5 marks)