The Open University of Sri Lanka
Faculty of Engineering Technology
Department of Electrical and Computer Engineering

058

| | |
|---|---|
| Study Programme | : Bachelor of Science Honours in Engineering<br>Bachelor of Software Engineering Honours |
| Name of the Examination | : Final Examination |
| Course Code and Title | **: EEX5466- Advanced Database Systems**<br>**EEI5466 - Advanced Database Systems** |
| Academic Year | : 2022/23 |
| Date | : 13, February 2024 |
| Time | : 0930-1230hrs |
| Duration | : **3 hours** |

## General Instructions

1. Read all instructions carefully before answering the questions.
2. This question paper consists of **Eight (8) questions** in **Eight (8)** pages.
3. Answer **Five (05) questions only**. All questions carry equal marks.
5. Answer for each question should commence from a new page.
6. This is a Closed Book Test **(CBT).**
7. Answers should be in clear hand writing.
8. Do not use Red colour pen.

---

## Question 1 – Database Programming (20 marks)

Consider the following **relational database schema** designed for a manufacturing company.

**Customer** (*CID varchar(20), name varchar(50), address varchar(50), contactNo int, credit_Limits int*)

**Manufacturer** (*MID varchar(20), manufacture_name varchar(50), address varchar(50), contact_No int* )

**Product** ( *PID varchar(20), name varchar(50),color: varchar(10), weight: int,* warranty *varchar(10),* unit_Price *int, manu_ID varchar(20), manufactured_Date DatetTime,*)

**Order** (*OID varchar(20),customer_ID varchar(20), reference_No varchar(10), order_Date date* )

**Order_Items** (OID varchar(20), product_ID varchar(20), qty int)

**Customer table** consists of ID (*CID*), *name*, *address*, contact number (*ContactNo*) and credit limit (*credit_limit*) of all customers. **Manufacturer table** consists of the ID (*MID*), name(*manufacturer_name*), *address* and contact number (*contact_No*) of all manufacturers. **Product table** stores information of all the products manufactured. The table stores the ID *(PID)*, *name*, *color, weight* in grams, *warranty* in years and *unit price* of all products along with the ID of the manufacturer(*manu_ID*) and the date the product is manufactured (*manufacutured_Date*). **Order** table stores the ID of the order (*OID*), ID of the customer who has placed the order (*customer_ID*), a reference number(*reference_No*) and the date the order was placed(*order_Date*).**Order_Items table** stores the product ID (*product_ID*) and quantity (*qty*) of the product in an order along with the order ID (OID). All primary keys are underlined.

(a) Write **SQL queries** to answer the following questions.

i. Display the reference number of orders and the name of customer for orders placed by customers with a credit limit over 100.

(4 marks)

ii. Find the product names of products which are not manufactured by the manufacturers from 'France'.

(4 marks)

(b) Create a function that calculates and returns the total price to be paid for an order, given the order ID. Assume that MS SQL Server is installed. Therefore, write T-SQL to address this query.

(6 marks)

(c) Assuming that the orders are delivered as packages, create a trigger to ensure that the weight of the package does not exceed 2 kg.

(6 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 2023. Assume the base rate to be Rs. 100,000.

(3 marks)

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

(a) Illustrate the components of a magnetic disk using a diagram.

(2 marks)

(b) Suppose that a disk has the following parameters:
seek time ST=20 msec; rotational delay RD=10 msec; block transfer time BTT=1 msec; block size B=2048 bytes; Cylinder size CS= 250 blocks.

An EMPLOYEE file has the following fields:
ENO, 10 bytes; LNAME, 20 bytes; FNAME, 20 bytes; MIDNAME, 20 byte; BIRTHDATE, 10 bytes; PHONE, 10 bytes; DEPARTMENT, 10 bytes.

The EMPLOYEE file has r=30000 employee records, fixed-length format, and unspanned blocking.

Write appropriate formulas and calculate the following values for the above EMPLOYEE file:

i. Calculate the record size R in bytes.

(2 marks)

ii. How many records fit onto a block?

(2 marks)

iii. How many blocks are required to store the entire file?

(2 marks)

iv. If the Next Block Concept is used to store the EMPLOTEE file, how many cylinders are needed?

(2 marks)

v. What is the time required to read a data block from the EMPLOYEE file?

(2 marks)

(c) 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 – XML Databases (20 marks)

(a) What are the components of a FLWOR expression and what do them each do?

(3 marks)

Consider the following XML document for the questions given below:

```
<Project location = "Colombo">
   <employee id="135">
       <name>Ramesh</name>
       <phone>0714123123</phone>
       <office>R344</office>
   </employee>
   <employee id="136">
       <name>Prabash</name>
       <phone>0777123123</phone>
       <office>R345</office>
   </employee>
   <employee id="138">
       <name>Lahiru</name>
       <phone>0776123123</phone>
       <office>R501</office>
   </employee>
</Project >
```

Note that the above XML document stored in `Department( id int, projDetails xml)` table created in MS SQL Server and it contains only single record.

(b) What is the output of the following XPath expression:

```
SELECT
projDetails.query('//employee[office="R344" or @id="138"]/phone')
FROM department
```

(2 marks)

(c) What would be the output of the following XPath expression?

```
SELECT projDetails.query('/Project/employee[1]/office')
FROM department
```

(3 marks)

(d) What would be the output of the following XQuery expression?

```
SELECT projDetails.query('
 for $d in /Project
 let $emp:=$d//employee/name
 where $d/@location = "Colombo"
 order by $d/@location
 return
 <EmpList>{$d/@location, $emp}</EmpList>')
FROM department
```

(3 marks)

(e) Write an expression to insert the following employee after employee id 136.

```
<employee id="137">
    <name>Saman</name>
    <phone>0755123123</phone>
    <office>R532</office>
</employee>
```

(5 marks)

(f) Write an expression to delete the phone number of employee id 138.

(4 marks)

## Question 6 – 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 7 – 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

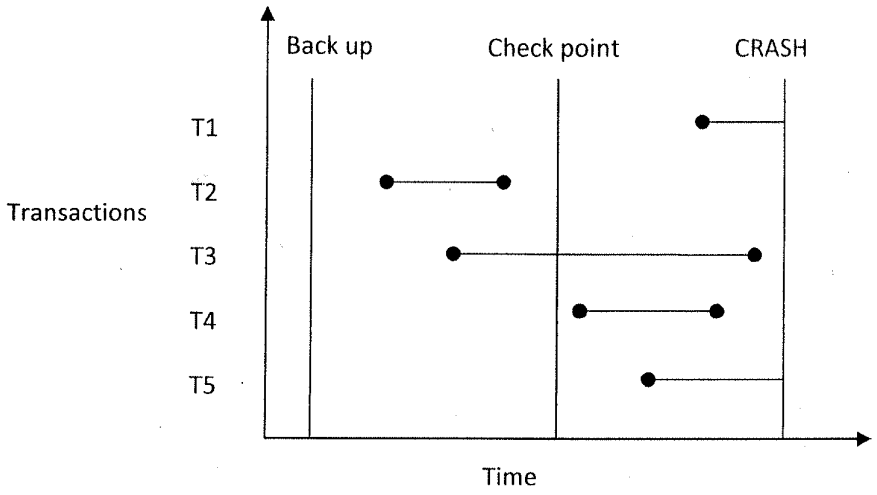
Assume that older transaction has higher priority always.

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)

## 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.) 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)

**End of the Paper**