00034

(2)

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

Study Programme	: Bachelor of Software Engineering Honours
Name of the Examination	: Final Examination
Course Code and Title	**: EEI5566/ECI5266 Advanced database systems**
Academic Year	: 2017/18
Date	: 12th February 2019
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 **Seven (7)** pages.
3. Answer any **Five (5)** questions only. All questions carry equal marks.
4. Answer for each question should commence from a new page.
5. Relevant charts/ codes are provided.
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 relations in a database created for an online store.

```
Customers (cid: char (4), name: varchar (50), phone: char(10), country: varchar(20))

Employees (eid: char (4), ename: varchar (50), phone: char (4), hiredate: date)

Orders (oid: int, eid: char (4), cid: char (4), orderDate: date, requiredDate: date, shippedDate: date)

OrderDetails (oid: int, productId: char (4), quantity: int, discount: real)

Products (productId: int,productName: varchar(15),UnitPrice: real unitsInStock: int,ROL:int)
```

The database stored information of the customers in the **Customers** table. The table contains customer id (*cid*), *name*, phone number (*phone*) and *country* of each customer. **Employee** table contains employee id (*eid*), name (*ename*), phone number (*phone*) and the date hired (*hiredate*) of each employee. **Orders** table stores an unique id (*oid*), id of the employee responsible for the order (*eid*), date the order is placed (*orderDate*), the date the order is required by the customer (*requiredDate*) and the date the order is shipped (*shippedDate*) for each order. **OrderDetails** table stores order id (*oid*), product id (*productId*), quantity of the product ordered (*quantity*) and any discount obtained for the order (*discount*). **Products** table stores the id of product (*productId*), name of the product (*productName*), unit price of the product (*UnitPrice*) , number of units available in stock (*unitsInStock*) and the reorder level(*ROL*) of each product.

(a) Use SQL queries to answer following questions.

i. Find the names and unit prices of all products which have not been ordered by customers from 'Germany'.

(5 marks)

ii. Find the customer ids and names of the customers who have obtained more than 5% discount for every product in every order they have placed.

(5 marks)

(b) Create a view named incompleteOrders that contains customer name, country and number of times an order is not completed for the orders that cannot be completed. Note that an order cannot be fulfilled when the units in stock is lesser than the quantity ordered for any product in the order.

(5 marks)

(c) Create a function named calcCost to calculate and return the total cost of an order given the order id and the discount. Note that the cost of the order is the total of costs for each product ordered where cost for each product ordered is obtained by multiplying the unit price of the product by the quantity.

(5 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) Describe the role of relational algebra in processing SQL queries in relational DBMS's. (5 marks)

(b) 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)

(c) List the names of all employees with two or more dependents. (5 marks)

(d) List the names of managers who have at least one dependent. (5 marks)

Question 3 – Object Relational Databases (20 marks)

Consider the following object relational database schema for recording the assignments of employees to projects:

Object types:

proj_t (*pno*: char(6), *pname*: varchar2(12), *stdate*: date, *endate*: date)
assignment_t (*project*: ref proj_t, *adate*: date, *hours*: number(2))
assignment_tb: table of *assignment_t*
emp_t (*eno*: number(8), *ename*: varchar2(12), *assignments*: assignment_tb)

Tables:

projects of *proj_t* (pno primary key);
emp of *emp_t* (eno primary key) nested table assignments store as emp_assigntb;

The attributes of *proj_t* are project number (*pno*), project name (*pname*), start date (*stdate*), and end date (*endate*). The attributes of *assignment_t* include ref of *proj_t* (*project*), date on which the employee was assigned to the project (*adate*), and number of hours per week (*hours*). In the type *emp_t*, the attributes are employee number (*eno*), name (*ename*), and the set of project assignments of type *assignment_tb* which is a table of *assignment_t*. Some sample data for this database is shown below.

PROJECTS

PNO	PNAME	STDATE	ENDATE
MARS01	MARS PROBE	01-Jan-2007	31-Aug-2008
LUNA05	LUNAR LANDER	01-Jan-2006	31-Dec-2008
SPCE02	SPACE STATION	01-Jun-2006	30-Jun-2008
COME03	COMET PROBE	05-Sep-2007	31-Dec-2007

EMP

ENO	ENAME	ASSIGNMENTS		
		PROJECT	ADATE	HOURS
12347648	Carol SMART	MARS01	01-Jan-2007	10
		LUNA05	10-Jun-2007	10
		SPCE02	20-Mar-2007	15
14527845	Jim TERRIFIC	MARS01	11-Feb-2007	15
		LUNA05	12-Apr-2007	12
		SPCE02	20-Jun-2007	13

(a) Write Oracle OBJECT SQL statements to answer the following queries. Use columns of REF type instead of joins to link tables.

i. For each employee, get the employee number, name, and the total number of hours per week the employee is assigned to work on projects.

(5 marks)

ii. For each project that has employees assigned to it, find the project number, project name, and the number of employees assigned to it.

(8 marks)

(b) Assuming that the database contains only the given sample data, assign Carol SMART to project number COME03 for 2 hours per week from 5 September 2007.

(7 marks)

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

(a) What is the justification for using I/O costs as the main measure to compare different algorithms for evaluating relational operators? (4 marks)

(b) Briefly explain why it is often advantageous to do selections before joins in a query plan. How do early projections help during the query execution? (4 marks)

(c) Briefly describe what happens in external merge sort in first two passes.

(4 marks)

(d) Briefly describe two main techniques used in RAID. What is the purpose of each technique? (Note: You do not have to explain the RAID levels) (5 marks)

(e) Explain the role of the Disk Space Manager in the database management system architecture?

(3 marks)

Question 5 – XML Databases (20 marks)

Consider the following XML document for the questions given below:

```
<Contact>
        <Names>
                <Name type="Legal">
                        <First>Thomas</First>
                        <Middle>Cruise</Middle>
                        <Last>Mapother</Last>
                </Name>
                <Name type="Stage">
                        <First>Tom</First>
                        <Middle></Middle>
                        <Last>Cruise</Last>
                </Name>
        </Names>
        <Addresses>
                <Address type="Primary">
                        <Street>12345 Main Street</Street>
                        <City>San Diego</City>
                        <State>CA</State>
                        <Zip>92130</Zip>
                </Address>
                <Address type="Other">
                        <Street>6200 Cruise Avenue</Street>
                        <City>San Fernando</City>
                        <State>CA</State>
                        <Zip>92126</Zip>
                </Address>
        </Addresses>
        <Phones>
                <Phone type="Mobile">8085554422</Phone>
                <Phone type="Home">8085553399</Phone>
        </Phones>
</Contact>
```

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

(a) What is the output of the following XPath expression?

```
SELECT conDetails.query('//Phone[@type="Home"]')
FROM Contractors
```

(3 marks)

(b) Display the details of the first address. What would be the XPath expression and its output? (5 marks)

(c) Write an XQuery to print the name (i.e. first name and last name) of both Contractors.

Output:
```
<name>Thomas Mapother</name>
<name>Tom Cruise</name>
```

(6 marks)

(d) Write an expression to insert the following contractor name after the "*Thomas*"

```
<Name type="Legal">
        <First>Saman</First>
        <Middle>Keshara</Middle>
        <Last>Jayaweera</Last>
</Name>
```

(6 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 800 pages using 10 buffer pages. Indicate the different passes and what happens in each pass.

(5 marks)

(c) Assume that we want to join two relations Proj and Work with the following schema and associated statistics:

```
Emp (eno, ename, salary, age) 1000 pages with 10 tuple per page.
Proj (pno, pname, budget) 400 pages with 10 tuples per page.
Work (eno, pno, hours) 3,000 pages with 100 tuples per page.
```

Both Proj and Work tables have clustered B+ tree indexes on *pno*. Assume that the indexes on Proj and Work have 100 and 1000 leaf pages respectively and the heights of their B+ trees are 2 and 3 respectively.

i. 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.)

(4 marks)

ii. If there were no indexes on the two tables, what would be the cost of a sort merge join? Assume that there are enough buffer pages to sort each table in 2 passes.

(8 marks)

Question 7 – Transactions and Concurrency Control (20 marks)

(a) Briefly explain the properties of a transaction.

(3 marks)

(b) Explain the difference between a serial execution and a serializable execution of two transactions in a DBMS.

(3 marks)

(c) What is meant by dirty read in a DBMS? What is the risk in allowing transactions to perform dirty reads?

(3 marks)

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

(3 marks)

(e) Show a schedule that is unrecoverable. Briefly explain the schedule.

(3 marks)

(f) Explain the following about deadlocks.

i. How does a deadlock occur?

(2 marks)

ii. How does the DBMS detect a deadlock?

(2 marks)

iii. How the DBMS resolves deadlock (after it has occurred).

(1 mark)

Question 8 – Crash Recovery (20 marks)

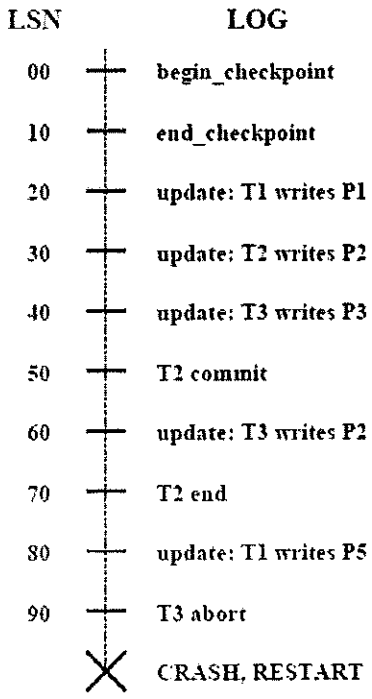
(a) Why is STEAL – NO FORCE approach desirable for DBMSs?

(5 marks)

(b) Why is it difficult to enforce Atomicity and Durability properties in a STEAL - NO FORCE approach?

(5 marks)

(c) Consider the execution shown in Figure and answer the following questions.



i. What is the value of the LSN stored in the master log record?

(3 marks)

ii. What is done during Analysis?

(7 marks)