UNIVERSITY OF COLOMBO, SRI LANKA

UNIVERSITY OF COLOMBO SCHOOL OF COMPUTING

DEGREE OF BACHELOR OF INFORMATION TECHNOLOGY (EXTERNAL)

Academic Year 2016 – 3rd Year Examination – Semester 6

IT6405: Database Systems II

8th October, 2016

(TWO HOUR)

To be completed by the candidate

BIT Examination Index No: ....

Important Instructions:

- The duration of the paper is **2 (two) hour**.
- The medium of instruction and questions is English.
- This paper has **4 questions** and **14 pages**.
- **Answer all questions.** All questions **do** carry equal marks.
- **Write your answers** in English using the space provided **in this question paper**.
- Do not tear off any part of this answer book.
- Under no circumstances may this book, used or unused, be removed from the Examination Hall by a candidate.
- Note that questions appear on both sides of the paper.
  If a page is not printed, please inform the supervisor immediately.

Questions Answered

Indicate by a cross (×), (e.g. [✗]) the numbers of the questions answered.

	Question numbers			
To be completed by the candidate by marking a cross (×).	1	2	3	4
To be completed by the examiners:				

1) (a) Consider the following two transactions T1 and T2 with respect to an airline reservation system interleaved as given in Table 1 as T1 and T2. Name and briefly explain the problems associated with the given scenario.

(05 marks)

Table 1

T1	T2	Operation	Data Value
Read Flight Information			seats = 15
	Read Flight Information		seats = 15
	Book 2 seats	seats = seats -2	
Book 1 seat		seats = seats -1	
	Write seats		seats = 13
Write seats			seats = 14

ANSWER IN THIS BOX

Called Lost Update [1]

This problem occurs when two transactions, accessing the same data items, have their operations interleaved in such a way, that one transaction will access the data before the other has applied any updates. [2]

This violates the Serialisability property which requires that the results of interleaving must leave the database with the same results as serial processing. It also violates the Isolation property of allowing a transaction to complete without interference from another. [2]

2

(b) Consider the following two transactions T1 and T2 with respect to an airline reservation system interleaved as given in Table 2 as T1 and T2. Name and briefly explain the problems associated with the given scenario.

(05 marks)

Table 2

T1	T2	Operation	Data Value
Read Flight Information			seats = 15
Book 1 seat		seats = seats -1	
Write seats			seats = 14 (Uncommitted)
	Read Flight Information		seats = 14
	Book 2 seats	seats = seats -2	seats = 12
	Write seats		seats = 12
Rollback			seats = 15

ANSWER IN THIS BOX

Uncommitted Data [1]

This problem occurs when one transaction updates a data item, but has not yet

committed the data permanently to the database. Because of failure, the transaction is

rolled back and the data item is returned to its previous value. A second transaction

accesses the updated data item before it is returned to its original value. [2]

This violates the Isolation property of allowing a transaction to complete without

interference from another. [2]

(c) Consider the following three transactions T1, T2, and T3 and its' operations and data values. Here operations are interleaved in such a way that the transaction performing the calculations may read some data items before they are updated and others are shown in Table 3. Note that initially there are no seats booked for the flight F106 and F113. This type of problem is named as what? Briefly explain this problem. What does this violate?

(05 marks)

Table 3

T1	T2	T3	Operation	Data Value
		Read seats from flight F106		F106 seats=12
		Add seats to flight F106 to Total Seats	Total seats = Total seats+12	Total Seats=12
Read Flight F106 Information				F106 seats= 12
Book 1 seat			F106 seats = F106 seats-1	
Write seats				F106 seats= 11
	Read Flight F113 Information			F113 seats= 4
	Book 2 seats		F113 seats = F113 seats-2	
	Write seats			F113 seats= 2
		Read seats from flight F113		F113 Seats= 2
		Add seats to flight F113 to Total Seats	Total seats = Total seats+2	Total Seats= 14

ANSWER IN THIS BOX

Inconsistent Retrievals [1]

This problem occurs when one transaction is calculating summary functions on particular data items while other transactions are updating those data items. The transaction performing the calculations may read some data items before they are updated and others after.

Assume Flight F106 has 12 seats available and Flight F113 has 4 seats available giving a total of 16 available seats. If Transaction 1 were to update Flight F106 by booking 1 seat and Transaction 2 were to update Flight F113 by booking 2 seats, a calculation of the total seats available would yield a result of 13 seats. [2]
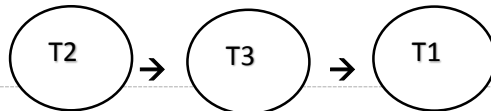
Continued…

4

This also violates the isolation property of allowing a transaction to complete without interference from another. [2]

(d) Draw the precedence graph for the following schedule, where C means "commit".

(03 marks)

```
T1:       R(A) W(Z)                   C
T2:                 R(B) W(Y)         C
T3: W(A)                    W(B)      C
```

ANSWER IN THIS BOX

(e) Consider the following incomplete schedule *S*.

```
T1: R(X) R(Y) W(X)          W(X)
T2:           R(Y)      R(Y)
T3:                W(Y)
```

Determine (by using a precedence graph) whether the schedule is conflict serializable. If the schedule is not conflict serializable modify it to be a serializable schedule.

(05 marks)

ANSWER IN THIS BOX

The precedence graph has an edge, from T1 to T3, because of the conflict between T1:R(Y) and T3:W(Y). It also has an edge, from T2 to T3, because of the conflict between the first T2:R(Y) and T3:W(Y). It also has an edge, from T3 to T2, because of the conflict between T3:W(Y) and the second T2:R(Y).

It is not possible. Since the precedence graph is cyclic, we know that it is not

conflict-serializable.

If we are allowed to add abort actions (which was not mentioned in the question), we could simply

abort either T2 or T3 and the schedule would become conflict-serializable.

(f) For the following schedule, state whether it is conflict-serializable and/or view-serializable. If you cannot decide whether a schedule belongs to either class, explain your reasons briefly. The actions are listed in the order they are scheduled, and prefixed with the transaction name.

```
(i)    T1:W(X) T2:R(Y) T1:R(Y) T2:R(X)
```
(02 marks)

ANSWER IN THIS BOX

conflict-serializable and view serializable

2) (a) Consider the following SQL query for an insurance database where Car(regno, year, model) relation is identified by the primary key regno.

```
SELECT a.regno
FROM   Car a, Car b
WHERE a.year > b.year and b.regno='WP KQ-6894';
```

(i) Write an efficient relational-algebra expression that is equivalent to this query. Justify your choice.

(03 marks)

ANSWER IN THIS BOX

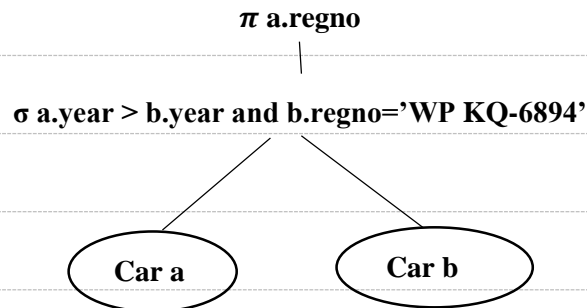**Restrict Car b by regno='WP KQ-894' giving one record named as T1**

Restrict T1 Car a and Car b by a.year> b.year giving matching records as T2

Project regno from Car a for the records of T2 giving Result.

6

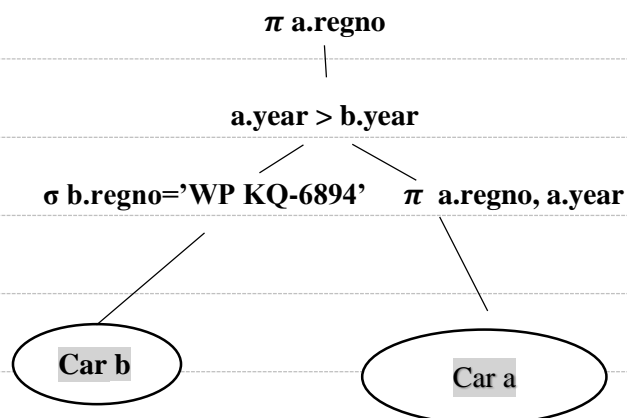(ii) Draw the initial query tree for the above SQL query.

(04 marks)

ANSWER IN THIS BOX

π **a.regno**

σ **a.year > b.year and b.regno='WP KQ-6894'**

Car a   **Car b**

(iii) Draw the optimized query tree for the above SQL query.

(05 marks)

ANSWER IN THIS BOX

π **a.regno**

a.year > b.year

σ **b.regno='WP KQ-6894'** π **a.regno, a.year**

Car b   **Car a**

(iv) Suppose that a B+-tree index on the model is available on relation Car, and that no other index is available. What is the advantage of this index in handling the following expression?.

$\sigma_{model='BMW'}\ car$

(02 marks)

ANSWER IN THIS BOX

If a B+-tree is available on the model then we could filter BMW model using the index file

And quickly locate the car models for BMW. This will expedite the filtering condition and allow

to locate the relevant records quickly   .

7

(v) If the Car relation has 30,000 tuples and 25 tuples fit on 1 block, estimate the number of disk block accesses required if all the records are to be read.

(02 marks)

ANSWER IN THIS BOX

No of blocks occupied for the Car relation = 30000/25 = 1200

Hence 1200 blocks should be read to read the Car relation.

(vi) What type of statistics would be useful to process the query in (a) above?

(02 marks)

ANSWER IN THIS BOX

As regno is unique we would get one record for that. It would be useful if an index is

available for an year. Then the year comparison can be done quickly as well.

(b) Briefly explain the steps involved in query optimization process.

(07 marks)

ANSWER IN THIS BOX

Query process stages involve expressing the SQL Query in a high-level language.

Scanning, Parsing & Validating the SQL query with the database meta-data

Mapping the SQL query to an intermediate form of query as a relational algebra

expression.

Apply Query Optimisation rules and optimize the relational algebra query. Use statistics about the

data as required.

Determine the execution plan.

Generate code for the optimized query.

Execute the query and obtain the results.

Continued…

Send results to the user who executed the query.

3) (a) Discuss the advantages and disadvantages of fragmentation.

(06 marks)

ANSWER IN THIS BOX
Advantages of Fragmentation
Since data is stored close to the site of usage, efficiency of the database system is increased.
Local query optimization techniques are sufficient for most queries since data is locally available.
Since irrelevant data is not available at the sites, security and privacy of the database system can be maintained.
Disadvantages of Fragmentation
When data from different fragments are required, the access speeds may be very high.
In case of recursive fragmentations, the job of reconstruction will need expensive techniques.
Lack of back-up copies of data in different sites may render the database ineffective in case of failure of a site.

(b) Consider the fragmentation strategies for the Customer table of the XYZ Company.

Customer(Cus_Number, Cus_Name, Cus_Address, Cus_Province, Cus_Limit, Cus_Balance, Cus_Rating, Cus_Due)

Customer

Cus_Number	Cus_Name	Cus_Address	Cus_Province	Cus_Limit	Cus_Balance	Cus_Rating	Cus_Due
1	Perera	530, Bagathale Rd, Colombo 03	WP	100,000	80,000	3	5,000
2	Silva	5 Maligawa Rd, Kandy	CP	50,000	40,000	1	0

3	Dias	10, Galle Rd, Colombo 04	WP	75,000	60,000	3	2500
4	Vass	10, Queens Rd, Kandy	CP	30,000	20,000	3	1000
5	De Silva	35, Dickson Rd, Galle	SP	25,000	15,000	2	0

(i)   Suppose XYZ Company's corporate management requires information about its customers in all three provinces, but company locations in each province (WP. CP and SP) require data regarding local customers only. Based on such requirements, suggest the best strategy to distribute the data by province.

(04 marks)

ANSWER IN THIS BOX

Horizontal fragment based on Province for WP, CP and SP. (3)

Having the same attributes in all the sites. (1)

(i)   Suppose the company is divided into two departments: the service department which requires data related to customer location, and the collections department which requires rest of the customer information. Each department is located in a separate building, and as mentioned before each has an interest in only a few of the Customer table's attributes. Based on such requirements, suggest best strategy to distribute the customer data.

(05 marks)

ANSWER IN THIS BOX

Vertically fragment based on Customer location and rest with Cus_Number appearing in both (1)

i.e. Customer_V1(Cus_Number, Cus_Name, Cus_Address, Cus_Province) (2)

** Customer_V2(Cus_Number, Cus_Limit, Cus_Balance, Cus_Rating, Cus_Due) (2)**

(c) Several factors influence the decision to use data replication. Briefly explain these factors.

(07 marks)

ANSWER IN THIS BOX

Database size. The amount of data replicated will have an impact on the storage requirements

Continued…

and data transmission costs. Replicating large amounts of data requires a window of time and high bandwidth that could affect other applications.

Usage frequency. The frequency of data usage determines how frequently the data needs to be updated. Frequently used data needs to be updated more often, for example, than large data sets that are used only every quarter.

Costs, including those for performance, software overhead and management associated with synchronizing transactions and their components vs. fault-tolerance benefits that are associated with replicated data.

Data replication information is stored in the distributed data catalogue (DDC), whose contents are used by the TP to decide which copy of a database fragment to access.

(d) Data allocation describes the process of deciding where to locate data. Briefly describe the data allocation strategies.

(03 marks)

ANSWER IN THIS BOX

With centralized data allocation, the entire database is stored at one site.

With partitioned data allocation, the database is divided into two or more

disjointed parts (fragments) and stored at two or more sites.

With replicated data allocation, copies of one or more database fragments are stored at several

sites.

4) (a) Write a trigger to add 15% VAT for the cost of an item in table *trigger_items* when updating the cost of an item in table. Assume the table name is *trigger_item_cost* and *prices* are stored in *trigger_items*.

(05 marks)

ANSWER IN THIS BOX

```
CREATE TRIGGER `after_update_cost`

    AFTER UPDATE ON `trigger_items_cost` FOR EACH ROW

    BEGIN

        UPDATE trigger_items

        SET price = (NEW.cost * 1.5)

        WHERE item_id = NEW.item_id;

    END
```

(b) Create a **row level** trigger for the Customer table that would fire for INSERT or UPDATE or DELETE operations performed on the Customer table. This trigger should display the salary difference between the old values and new values.

(07 marks)

ANSWER IN THIS BOX

```
CREATE OR REPLACE TRIGGER display_salary_changes

BEFORE DELETE OR INSERT OR UPDATE ON Customers

FOR EACH ROW

WHEN (NEW.ID > 0)

DECLARE sal_diff number;

BEGIN

    sal_diff := :NEW.salary  - :OLD.salary;

    dbms_output.put_line('Old salary: ' || :OLD.salary);

    dbms_output.put_line('New salary: ' || :NEW.salary);
```

Continued…

12

```
    dbms_output.put_line('Salary difference: ' || sal_diff);

END;
```

(c) If the following statements are executed, what would be the results of the trigger in part (b) above?

INSERT INTO CUSTOMERS (ID,NAME,AGE,ADDRESS,SALARY)
VALUES (7, 'Kriti', 22, 'HP', 7500.00);

(03 marks)

ANSWER IN THIS BOX

Old salary:

New salary: 7500

Salary difference:

(d) If the following statements are executed, what would be the results of the trigger in part (b) above?
Assume old salary is 1500.
```
    UPDATE customers
    SET salary = salary + 500
    WHERE id = 2;
```

(03 marks)

ANSWER IN THIS BOX

Old salary: 1500

New salary: 2000

Salary difference: 500

(e) Write a stored procedure to create a procedure '*employer_details*' which gives the details of the employee first and last name and current salary. Assume *emp_tbl* has the relevant data.

(05 marks)

ANSWER IN THIS BOX

CREATE OR REPLACE PROCEDURE employer_details

```
IS

CURSOR emp_cur IS

SELECT first_name, last_name, salary FROM emp_tbl;

emp_rec emp_cur%rowtype;

BEGIN

  FOR emp_rec in emp_cur

  LOOP

    dbms_output.put_line(emp_cur.first_name || ' ' ||emp_cur.last_name

        || ' ' ||emp_cur.salary);

  END LOOP;

END;
```

(f) Write down the syntax to execute a stored procedure.

(02 marks)

ANSWER IN THIS BOX

EXECUTE [or EXEC] procedure_name;

****