No. of Pages: 4   **B**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M.TECH DEGREE EXAMINATION, DECEMBER 2017

Branch: *Computer Science & Engineering*

Stream(s): *Computer Science & Engineering*

01CS6103 : *Topics in Database Technology*

Answer *any two full* questions from *each* part

Limit answers to the required points.

Max. Marks: 60   Duration: 3 hours

PART A

1. a. The total cost of hash join is $3(b_S + b_R) + b_{RES}$, where b_S, b_R and b_{RES} are number   (2)
        of disk blocks in relation R, relation S and resultant relation, respectively.
        Justify the statement

    b. Consider the following relations

        1. Emp (eid :integer, ename ; char(30), did : integer, sal : integer, hobby
            : char(20))

        2. Dept (did :integer, dname : char (20), mgrId : integer, phone :
            char(10))

        3. Project (pid : integer, pname : char (20), budget : real)

        4. WorksOn(eid: integer, did : integer, pid : integer, hoursWorked
            :integer)

        Find the optimized relation algebraic expression and the optimized Query
        Tree for the following SQL query
                                                                                    (5)
        SELECT D.dname, P.pname, E. ename
        FROM Emp E, Dept D, Project P, WorksOn W
        WHERE E.eid = W.eid AND D.did = W.did AND P.pid =W.pid AND
        E.sal >= 50000 AND E.hobby = 'Gardening'

    c. What is meant by SQL injection? Illustrate with an example.   (3)

2. a. Describe the wait-die protocol for deadlock detection   (2)

1

b. Consider the following statistics about a relational table, EMPLOYEE(ENO, Name, Salary, DNO). There are 12000 records in 3000 blocks with a blocking factor of 4. There is a secondary index on non key attribute DNO with $X_{DNO} = 2$. There are 150 different departments and hence selection cardinality is 80. We have another relation, DEPARTMENT (DNUMBER, DName, MGRSSN). There are 150 rows in this table, stored in 10 disk blocks. There exists a primary index on DNUMBER with $XDNUMBER = 1$. Join selectivity between EMPLOYEE and DEPARTMENT relation is 1/150 and blocking factor for resulting relation after the join is 4 records. Estimate the cost of join operation (EMPLOYEE $\bowtie$ DEPARTENT) using i) nested loop join ii) single loop join and which one is best choice. (6)

c. Check whether the following schedule is conflict serializable or not (2)

$r_1(A)\ w_2(A)\ r_1(B)\ w_3(A)\ r_1(A)\ r_3(B)$

3. a. How speedup differ from scaleup? Suppose a transaction is written in C with embedded SQL, and about 80% of time is spend in the SQL code, with remaining 20% in C code. How much speedup can one hope to attain if parallelism is used only for SQL code (4)

b. "Transaction log for a recovery system with deferred database modification does not have store old values of data items" Justify this statement (2)

c. Suppose we have a relation with 500000 record and each records requires 10 bytes. Let the disk-block size be 4,096 bytes. Let the number of buffers (n_B) provided is 4. Assume we would sort the records by Sort-Merge algorithm
   i. How many sorted subfiles will there be after the initial pass of the sort?
   ii. How many passes (including the initial pass just considered) are required to sort this file? http://www.ktuonline.com
   iii. Find the total number of block transfers required for the sorting (4)

PART B

4 a. Consider a parallel system with shared-nothing architecture. Assume that there are four processors P_0, P_1, P_2, and P_3 with associated disks D_0, D_1, D_2 and D_3. Assume that the records with the following keys are partitioned among the disks such that record with key, k goes to disk D_i if k mod 4 = i:

26, 16, 27, 15, 30, 18, 14, 25, 1, 9, 8, 24, 10, 12, 23, 17, 21, 13, 3, 7, 20

Assume parallel external merge-sort is used to sort the records. What will be the contents of disks before sorting? Assume that the same partition vector [8, 17, and 22] is used in the merge phase. Show snapshots of data distribution after local sorting, range partitioning and merging (5)

b. Illustrate how *semi-join* leads to efficient query processing in distributed databases (5)

5. a. Explain 2 phase commit protocol in a distributed environment . What actions would be taken when a *coordinator fails*? (5)

    b. Discuss how efficient access of spatial data is possible with k-d trees and Quad-trees. (5)

6. a. What form of parallelism (inter-query, interoperation or intra-operation) is likely to be most important for the following tasks (3)

        i) Increasing the throughput of a system with many small queries

        ii) Increasing throughput of a system with a few large queries, when the number of disks and processors is large

    b. Explain how deadlocks are handled using local/global wait-for graph in a distributed environment? What is false cycles? (4)

    c. Suppose you have a relation containing the x, y coordinates and names of restaurants. Suppose also that only queries that will be asked are of the following form; The query specifies a point, and asks if there is a restaurant exactly at that point. Which type of index would be preferable, R-tree or B-tree? Why? (3)

PART C

7. a. Discuss the concept of encapsulation and tell how it is used to create complex object structure (4)

    b. It is necessary to represent the basic details about books. Each book has a title, one or more authors, publisher and price. The book has chapters, each with a chapter number and name. Each chapter can have sections. A section has name and number. A section can have other sections. The book also has appendices each with appendix number and name. The structure of appendix is same as that of a chapter. Design a DTD for this book structure. (6)

8. a. Consider the following the relations: FACULTY relation with attributes FNO, NAME, DOB, SALARY, DNUM and DEPARTMENT relation with attributes DNO, DNAME, DLOCATION, HOD. All attributes are predefined basic types, which you may suitably assume. FNO and DNO are the primary keys of FACULTY and DEPARTMENT relations. DNUM is a foreign key that identifies the department to which an faculty belongs. HOD is a foreign key identifying the faculty who is acting as the Head of Department. Write required declarations in ORACLE to connect these tables using object references. (7)

    b. How structured, semistructured and unstructured data differ? (3)

9. a. Outline the methods in OODBMS makes objects persistent (3)

b. Use the DTD below to answer the following queries. Assume that all undeclared elements are of type PCDATA:

```
<!DOCTYPE songs[
        <!ELEMENT songs (song+)>
        <!ELEMENT song(song_id, sname, year, director, genre, singer+)>
        <!ATTLIST song language CDATA# REQUIRED>
        <!ELEMENT singer(name, nickname, gender)>
        .......
]>
```

i) Names of Malayalam songs (use XPath)

ii) Names of songs by a singer whose nick name is 'KJ' (use XPath)

iii) List names of songs and corresponding singers (use XQuery)

(7)