B

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M.TECH DEGREE EXAMINATION, DECEMBER 2018

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. Consider the tables R (A, B, C), T(D,E,F), S(G, H) and U(A,D, G, I) where A, D and G   (7)
        in U are foreign keys referring to the primary keys with the same names. Show an
        initial query tree for the following query and optimize it using the rules of heuristics:

        select B, E, G, H, I from R, T, S, U where R.A = U.A

        and T.D = U.D and S.G = U.G and R.C = 'TEXT' and U.I > 20 and T.E = 25

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

2. a. "In the recovery of concurrent transactions under the immediate database   (4)
        modification scheme, undo operations must be done before redo operations." Justify
        this statement with the support of a small sample log.

    b. What is a cascade-less schedule? Give an example   (4)

    c. Suppose that a task, which has been running in 2 hours, now runs in 3.8 hours when   (2)
        the system resources are doubled. Compute the *speed-up*.

3. a. There are two relations R(A,B,C) and S(C,D,E) with the following statistics: R   (6)
        has 1,000 tuples of size 15 bytes and R2 has 40,000 tuples of size 15 bytes.
        Estimate the number of block transfers required using each of the following
        join strategies for R*S assuming a block size of 300 bytes:-

        i.   Nested-loop join with 10 buffers

        ii.   ii. Simple hash join with a uniform hash function (assume that the
              smaller relation fits completely into memory)

    Ignore the cost of result output.

    b. Assume that in two-phase locking, read locks are granted and released as and   (4)
        when needed, but write locks are granted in the growing phase and released
        in the shrinking phase only. Will this new protocol ensure serializability? If
        yes, justify. If no, give a counter example.

1

PART B

4 a. Suppose that in the two-phase commit protocol for distributed databases, phase 1 is (5) modified as follows: when receiving the <prepare, T> message, each participating site which is ready to commit sends a <ready,T> message but does not write a <ready, T> record to the log. Will this modified protocol work correctly always? Justify your answer.

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

5. a. List out important data partitioning schemes in parallel databases. (2)

   b. Suppose that we have 4x4 matrix of processors labeled as P11, P12, P13, P14, (8) P21, P22, P23, P24, P31, P32, P33, P34, P41, P42, P43, P44 where Pij represents the processor at row i, column j. There are two relations R with 4000 records and S with 400 records. R is divided into four equal partitions and stored in P11, P21, P31, P41. Similarly, S is partitioned into four equal partitions and stored in P11, P12, P13, P14. Compute the communication cost for asymmetric and symmetric versions of fragment-and-replicate join. Communication cost is the total number of records communicated between processors as part of join. Assume that the relations are not re-partitioned as part of the join process. Ignore the cost of result storage, transfer etc. List out any assumptions made.

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

   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 (4) environment? What is false cycles?

   c. Write a brief note on spatial data management (3)

PART C

7. a. What is the difference between mutable and immutable objects? (3)

   b. Write INFORMIX DDL statements to create a table with row objects to store details of (7) students which include rollno, name, age, courses_taken. All except courses_taken are simple attributes. courses_taken is to be maintained as a nested table of course details. For each course, the following details are to be cno, cname, credits all of which are simple attributes.

   You may assume suitable types for all the simple attributes.

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 do structured, semi-structured and unstructured data differ? (3)

9. a. Identify the any two different DTDs that can match the following XML document: <A><B><C>xyz</C>< </B></A> (3)

   b. Use the DTD below to answer the following queries. Assume that all undeclared elements are of type PCDATA: (7)

   <!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 English or Malayalam songs (use XPath)

   ii) Names of songs by a singers who sang in the year 2018 (use XPath)

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