B

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M.TECH DEGREE EXAMINATION, DECEMBER 2015

Department: **Computer Science and Engineering**

*Stream: **Computer Science and 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 I

1. a. What is meant by pipelined query evaluation? How is it advantageous over materialized evaluation? (2)

   b. Consider the following tables representing enrollment of students to courses: STUDENT (ROLLNO, NAME, AGE, ADDRESS, EMAIL), COURSE(CNO, CNAME, AREA) and ENROLLEMENT(ROLLNO, CNO,GRADE), where ROLLNO and CNO in ENROLLMENT are foreign keys referring to the primary keys with the same names. (6)

   Show an initial query tree for the following query and optimize it using the rules of heuristics:

```
SELECT CNAME, NAME, EMAIL, GRADE
FROM STUDENT, COURSE, ENROLLMENT
WHERE COURSE.CNO=ENROLLMENT.CNO
and STUDENT.ROLLNO = ENROLLMENT.ROLLNO
and COURSE.CNAME = 'dbs'
and STUDENT.NAME = 'Raju'
```

   Assume that CNAME is a candidate key of COURSE.

   c. Suppose that a certain job was executing on single processor system in 2 hours. Now that the size of the job has been doubled, it is shiftedto a two-processor machine. If it is observed that the new job takes 3.5 hours to run, compute the speed-up achieved. Is the speed-up linear or sub-linear? Why? (2)

2. a. Consider the following statistics about a relational table EMPLOYEE: - There are 12000 rows each of size 200 bytes. There exists a clustering index on SALARY attribute with X_{SALARY} = 3, secondary index on the attribute DNO with X_{DNO}= 2, a secondary index on the attribute SEX with X_{SEX} = 1. Half of the employees are male. On the average 20% of employees belong to the same department. 40% employees have salary >25,000. (5)

   Estimate the minimum execution cost of the following query assuming that the blocking factor is 4.

   $$\sigma_{SALARY>25000\ AND\ SEX='male'\ AND\ DNO=5}(EMPLOYEE)$$

   b. Argue that two-phase locking ensures serializability. (2)

   c. What is the significance of the *-property in the Bell-Lapadula model? Illustrate. (3)

3. a. There are two relational tables BOOK (ISBN, TITLE, PRICE, PUBCODE) and PUBLISHER (PUBCODE, NAME, ADDRESS) where PUBCODE in BOOK is the foreign key referring to the PUBLISHER table. There are 6000 BOOK records each of size 50 bytes and 50 PUBLISHER records each of size 100 bytes. Assuming that there are 6 buffers and that block size is 400 bytes, compute the cost of (i) sorting the BOOK table and (ii) joining the two tables using nested-loop join. (5)

   b. "Transaction log for a recovery system with deferred database modification does not have to store old values of data items." Justify this statement. (3)

   c. In the following multi-level table, if the a user with classification C wants to see all the attribute values in the row with account no CA234, write any *two* possible modified combination of classification attributes for the row. (2)

ACCNO	NAME	BALANCE	TC
SB123 U	REENA C	10101 S	S
CA234 U	LEENA S	20001 U	S
LA353 S	MEENA TS	10001 S	TS

PART II

4. a. There are four sites S1, S2, S3 and S4 in a distributed DB system with weights 1, 2, 3 and 3 respectively. If a data item X is replicated across these sites, find the read quorum and write quorum such that, (8)
   i. majority protocol is followed
   ii. biased protocol is followed

   In each case, argue that the assigned quorum values enable the system to follow the corresponding protocol.

   b. List out *any two* indexing schemes for spatial data. How are they different from each other? (2)

5. a. Give examples of situations, in the context of query processing, where inter-operation and intra-operation parallelism can be used. (1)

b. Suppose that the data records are partitioned across 4 processors P0, P1, ..., P3 in share-nothing system as follows: (6)

Disk	record keys
D0	15, 7, 0, 9, 2,4
D1	24, 16, 3, 7, 13
D2	27, 14, 19, 5, 22
D3	18, 2, 8, 25, 16

Compute the total communication cost involved when the data is sorted using,

i. Parallel External Sort-Merge
ii. Parallel Sort

Assume that partition vector [7, 16, and 21] is used and that each processor can apply external-sorting independently.

Note: - Communication cost is measured as the number of values moving from one processor to another.

c. Briefly explain how time is represented in temporal databases. (3)

6. a. Consider the following relations: (7)

EMPLOYEE(ENO, NAME, ADDRESS, DOB,AGE, SALARY, DNUM)
DEPARTMENT(DNO, DNAME, DLOCATION, DPHONE, MGRENO)

DNUM is a foreign key that identifies the department to which an employee belongs. Assume that every employee belongs to exactly one department and that DNAME is a candidate key.The following is the statistics about relations. EMPLOYEE is at site 1 and has 10,000 records each of size 100 bytes; DEPARTMENT is at site 2 and has 20 records each of size 70 bytes. ENO, ENAME, DNUM, DNO, DNAME are of size 4, 7, 3, 3 and 10 bytes, respectively. Identify the best strategy execute the following query issued at site 1, by computing the communication cost involved.

```
SELECT ENAME, DNAME
FROM EMPLOYEE, DEPARTMENT
WHERE DNUM = DNO and DNAME = 'CSE'
```

b. Write a note on k-d trees. (3)

PART III

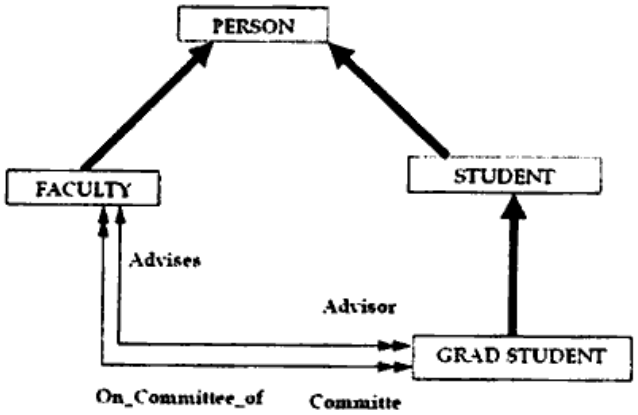
7. a. Suppose that we have to create table of students where each student is represented as an object of type *s_type*. This type contains the following fields: person of somepreviously defined type *p_type* (with fields*name* of type string and age of type*integer*), rollno of type *integer*, email-id of type *string* and year_of_graduation of type *integer*. Write declarative statements in INFORMIXto create the type *s_type* and the corresponding table. (5)

b. Show the INFORMIX SQL statement to insert a student with name 'Leena', age 23, roll no 102, email leena@nomail.com who is expected to graduate in 2016, in the table created in question 7.a. (2)

c. Write a note on RDF. (3)

8. a. Write a note on Object identifiers. (2)

b. Consider the ODMG diagram below and answer express the queries that follow inOQL. Assume that persistent names CS referring to computer science department and GRAD referring to graduate studentsfrom mechanical department exist. (5)



i. Name and salary of faculty from CS department

ii. Advisors of graduate students in Mechanical Department

c. What do you meant by well-formed and valid XML documents? (3)

9. a. Outline the methods in OODMBS make objects persistent. (3)

b. Consider the following relations: EMPLOYEE with attributes ENO, NAME, ADDRESS, DOB,AGE, SALARY, DNUM and DEPARTMENT with attributes DNO, DNAME, DLOCATION, DPHONE, MGRENO. All the attributes are of pre-defined basic types, which you may suitably assume. ENO and DNO are primary keys of EMPLOYEE and DEPARTMENT. DNUM is a foreign key that identifies the department to which an employee belongs. MGRENO is a foreign key identifying the employee who manages the department. (4)
Write required declarations in ORACLE to connect these tables using object references.

c. Show a sample XML file corresponding to the following DTD. Undeclared elements are PCDATA. The XML file should contain enough number of employee elements that demonstrate all the possible combination of elements according to the DTD. (3)

```
<!DOCTYPE employees[
        <!ELEMENT employees (employee+)>
        <!ELEMENT employee (name, age, address, phoneno*, email)>
        <!ELEMENT address (houseno | (housename, streetname))>
]
```