NIRMA UNIVERSITY
INSTITUTE OF TECHNOLOGY
Semester End Examination(IR/RPR), December 2018
M.Tech. in Computer Science & Engineering, Semester – I
3CS1107 – DATABASE SYSTEMS

Roll / Exam No. [   ] Supervisor's Initial with Date [   ]

Time: 3 Hours   Max Marks :100

Instructions: 1. Attempt all questions of Section I and II separately in same Answerbook.
2. Figure to right indicate full marks
3. Draw neat sketches wherever necessary.
4. Assume necessary data wherever required, and indicate clearly.

SECTION – I

Q.1 Do as directed   **[18]**

A) A software developer has released a product, which allow it's operators to define schema (attributes, data types), insert, update, delete, and view records using SQL. Can it be declared as a relational database management system? Justify your answer (4)

B) Does disk architecture play significant role in query performance? Justify your answer using proper example. (4)

C) What are advantages of index based query processing? Why it is not recommended to use index on small table? (4)

D) Consider a relation E (id, name, age, salary). Relation E has 1024 records, which are stored sequentially as per id. V(E,salary)=64. A data block can store 4 records of E. Most frequent queries on this relation are : (6)
   Query 1) Select * from E where id = <id_value>
   Query 2) Select * from E where salary between <value1> and <value2>
   Answer following questions to get best performance of these queries:
   i)   Which field should have index?
   ii) What is type of each index (sparse, dense, or dense on level-1 and sparse on level-2) ?
   iii) What is type of each index (primary or secondary)?

Q.2 Answer the following   **[16]**

A) Compare NSM, DSM, and PAX record formats. Write at-least one suitable application for each of these record formats. (5)

B) Derive formula to compute average seek time of a disk. (5)

OR

OR

B) A disk has following properties: (5)
   i) 8 platters with recording on both surface
   ii) Each surface with 2^{16} tracks,
   ii) Each track has average 2^8 sectors each of 2^{12} bytes
   iii) 10% of each track is used for gaps
   iv) disk rotation speed 7200rpm
   v) time required to move to nth track is $1 + n/1000$ ms.
   Calculate the time required to read a data block of 4096 Bytes. How can it be accelerated?

C) A database system holds two important relations R(A,B,C), and (6) S(A,D,E) that are frequently joined over a common attribute A. Each attribute is 10 bytes long. Relation R and S both have 42000 tuples, and which are stored contiguously as well as unsorted. A block can hold 4200 bytes.

   - Team A proposes row oriented record format.
   - Team B proposes column oriented record format. Hence, Team B propose following relational schema:
     R1(ID_R,A), R2(ID_R,B), R3(ID_R,C),
     S1(ID_S,A), S2(ID_S,D), S3(ID_S,E)
   Relation ids ID_R, and ID_S are of 4 bytes each.
   - The expected number of resulting tuples from R join S is 10.
   - Looking toward statistics, only two pass hash join is to be computed.

   i)   Compute number of I/O required for R join S in row storage format as per Team A.
   ii) Team B needs to perform R1(ID_R,A) join S(ID_S,A), and retrieve respective tuples k for other attributes using result(ID_R,ID_S,A). Compute number of I/O required for this task.
   iii) Which format do you suggest from result (i) and (ii)?

Q.3   Answer the following **[16]**

A) Consider a B+tree (keys=3, pointers=4) by performing following (8) operations sequentially.
   1) Insert 1 to 17 sequentially, and show resulting B+tree
   2) Delete 12,13,14 sequentially, and show resulting B+tree

B) Consider a dynamic hash structure where buckets can hold up to (8) two records. Initially the structure is empty. Then we insert the following records, in the order given below, where we indicate the hashed key in parenthesis (in binary) as

   a [010010]   e [010111]
   b [011001]   f [011011]
   c [101100]   g [101000]
   d [001101]   h [011111]

   How many buffers are allocated in extensible hashing to insert these records?

OR

B) Consider a linear hashing storage system. Each block can hold 3 (8) records. Initially primary buckets M=4. It need to store following elements (a-l):

a [0100]   b[0001]   c[0011]   d[0000]   e[0010]   f[0001]
g[0111]   h[1000]   i[1110]   j[1101]   k[0001]   l[0101]

    i) How many blocks are overflowed?
    ii) How many buffers are required?

SECTION – II

Q.4 **Do as directed** **[18]**

A) Given the following relations: (6)

    Vehicle (reg_no, make, colour)
    Person (eno, name, address)
    Owner (eno, reg_no).

Draw initial logical query plan for following query:

Select name, make, colour
From Person, Owner, Vehicle
Where Person.eno = Owner.eno AND Vehicle.reg_no=Owner.reg_no
AND Person.name = 'Hari' AND make > 2015.

Apply heuristics to make the query plan optimal. Justify your approach of optimization.

B) Relation R is designed to store infrastructure details in terms of (6) rectangular area. R-tree is used to index these objects for efficient access. An internal node of the R-tree is described in following figure:
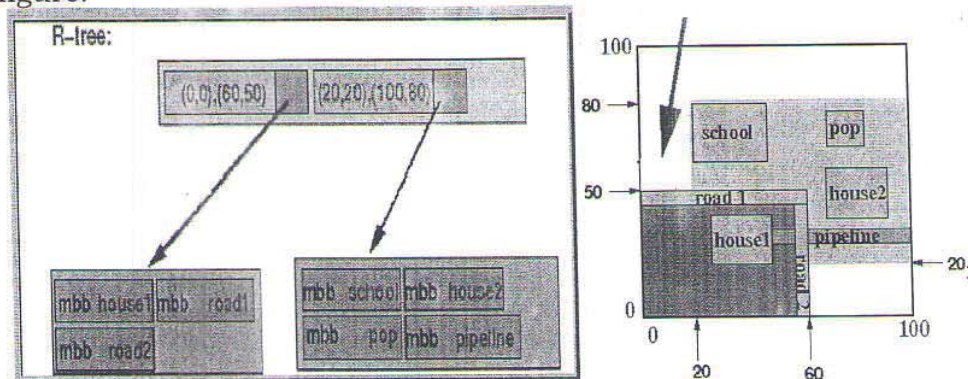
What will be change on R-tree index storage after insertion of two objects: i) House3(50,50,70,70), ii) Complex1(5,70,15,85)

C) Consider a relation R (Age, Salary) containing following 14 records: (6)

    (25, 60) (45, 0) (50, 75) (50,100) (50,120) (70,0) (64, NULL)
    (85,140) (30,260) (25,400) (45,350) (50,275) (60,260) (70, NULL)

Perform following tasks:

    i) Construct bitmap index
    ii) Compress bitmap index using run-length
    iii) Trace execution of query "Select * from R where salary is NULL" on compressed bitmap index

Q.5 **Answer the following**   **[16]**

A) Consider two relations R(A,B,C), and S(B,C,D). Statistics are given   (6)
below:
$$T(R_1) = 100,000 \; ; V(R_1,A) = 100; V(R_1,B) = 50; V(R_1,C) = 20$$
$$T(R_2) = 10,000 \quad ; V(R_2,D) = 10 \; ; V(R_2,B) = 20; V(R_2,C) = 50$$
Estimate number of tuples in result of following operations:

i) $W = (\sigma_{A=3} (R_1)) \; |X| \; (\sigma_{B=5} (R_2))$

ii) $Y = \Pi_{C,D} (\sigma_{B=3 \wedge C>5} (R_2))$

B) Trace two-pass sort-based distinct operator on following dataset.   (6)
Memory has 3 blocks and each block can hold 2 elements.
Dataset = {5,10,20,15,30,18,20,19,32,75,85,34}

OR

B) How many number of blocks need to be Input/Output to perform   (6)
sort-scan on relation of B block size and on system with M memory
block?

C) Describe processing of query "Select * from R minus Select * from   (4)
S", where B(R) > M > B(S)

Q.6 **Answer the following .**   **[16]**

A) Consider a database with two elements, X and Y. The initial values   (10)
of X and Y are both 0. Three transactions T1, T2, and T3 modify
these elements concurrently described below:
    T1: X := 42
    T2: Y := 20, X := 10
    T3: X := 100, Z := 101
Perform following tasks:
(i) Complete the undo log mentioned below by providing the
appropriate values for ??? in line 7 and line 10.
    1. < START T2 >
    2. < START T3 >
    3. < T2;X; 0 >
    4. < T2; Y; 0 >
    5. < COMMIT T2 >
    6. < START CKPT(T3) >
    7. < T3;X; ??? >
    8. < START T1 >
    9. < T3;Z; 0 >
    10. < T1;X; ??? >
(ii) What are the steps to be followed by recovery manager, if the
database crashes immediately after writing the above log entries?

B) With help of precedence graph, check whether following schedule is   (6)
serializable? If yes, specify equivalent serial schedule.
R1(A);W2(B);R2(A);W1(B);W3(B);W1(C);R3(B);W1(A)

OR

B) How do you decide mode of communication(pipeline or materialize)   (6)
between two physical operators of a logical plan?