No. of Pages: 3

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. Assume we have a ***Department*** relation with 50 records and an ***Employee*** relation with 6000 records stored in 10 and 2000 disk blocks respectively. There exists secondary indexes on attributes ***ENO of Employee*** and ***MGRSSN of Department*** with number of index levels $X_{ENO} = 4$ and $X_{MGRSSN} = 2$, respectively. Assume, every Department record has a matching Employee record. When the query,

DEPARTMENT $\bowtie_{MGRSSN=ENO}$ **EMPLOYEE**

is to be executed with a single loop join, show the importance of Join Selection factor by computing the number of block accesses required in each case (choice of relation in outer loop). (4)

b. Consider the tables FACULTY (FID, FNAME, DESIGNATION, PHONE, EMAIL,DNUMBER), COURSE(CNO, CNAME, CREDITS), STUDENT(ROLLNO, NAME, ADDRESS, AGE) and ENROLMENT(FID, CNO, ROLLNO, GRADE) where FID, CNO and ROLLNO in ENROLMENT 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 FNAME, CNAME, ROLLNO, NAME
FROM FACULTY F, COURSE C, STUDENT S, ENROLMENT E
WHERE F.FID=E.FID
and C.CNO=E.CNO
and S.ROLLNO = E.ROLLNO
and DESIGNATION = 'Professor'
and GRADE = 'D'
and CNAME = 'DBS'
```

(4)

c. Justify the statement: The worst case performance of External Sort-Merge of degree 2 is $(2 * b) + (2 * (b * (\log_2(b)))$ (2)

2. a. Consider the relations r_1(A, B,C), r_2(C, D, E), and r_3(E, F), with primary keys A, C, and E, respectively. Assume that r_1 has 1000 tuples, r_2 has 1500 tuples, and r_3 has 750 tuples. Estimate the size of $r_1 \bowtie r_2 \bowtie r_3$ (2)

b. Consider the following statistics about a relational table, **STUDENT**. There are 20000 rows, each of size 200 bytes. There exists a clustering index on **GRADE** attribute with X_{GRADE} = 3, secondary index on attribute **DNO** with X_{DNO} = 2, and a secondary index on **SEX**, X_{SEX} = 1. Half the students are males. There are only five departments and student distribution to the departments is almost equal. 50% of students are having grade > 5. Estimate the minimum cost of the following query assuming blocking factor is 4.

$\exists$GRADE > 5 and SEX = 'male' and DNO = 5 **(STUDENT)** (5)

c. Assume T_1, T_2, T_3 are three transactions having its time stamps 5, 10 and 20. If T_1 requests an item held by T_2, what action would be taken on T_1 and T_2 if the deadlock prevention scheme is i) *wait-die* ii) *wound-wait* (3)

3. a. Consider the following instances of a multi-level schema, R

Name		Salary		JobPerf		TC
Mahesh	U	4000	C	Fair	S	S
Gaurav	C	8000	S	Good	C	S

i) How a user with security classification S would see the relation?

ii) How it would be, if security classification of the user is C ? (2)

b. List out the different architectural models for parallel systems (3)

c. Suppose we have a relation with 1,000,000 record and each records requires 10 bytes. Let the disk-block size be 4,096 bytes. Let the minimum number of buffers (n_B)to be provided is cubic root of b, where b is number of blocks required to hold the relation. Assume we would sort the records by Sort-Merge algorithm

i. How many sorted sub-files 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?

iii. Find the total number of block transfers required for the sorting (5)

PART B

4. a. How hash partitioning is different from range partitioning? What factors could result in a skew when relation is partitioned by i) Hash partitioning ii) Range partitioning (3)

b. Illustrate and explain ***asymmetric fragment-and-replicate join*** and ***symmetric fragment-and-replicate join.*** (4)

c. Will functional dependencies be preserved if a relation is converted to a temporal relation by adding a time attribute? How is this problem handled in a temporal database? (3)

5. a. Explain 2 phase commit protocol in a distributed environment . What actions would be taken when a ***site recovers from failure***? (5)

b. Explain the R-tree index structure for spatial database. 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? (5)

6. a. List out any three advantages of distributed systems (2)

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

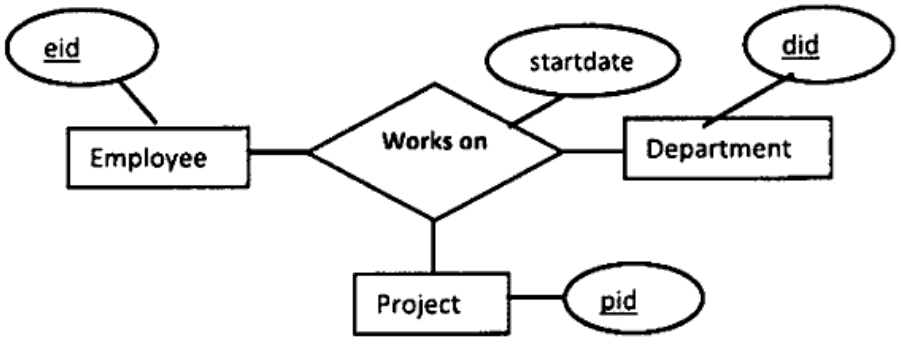
c. How does a *nearness query* differ from a *region query*? (3)

PART C

7. a. What is meant by object persistence? How is it achieved in object databases? (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. What is the purpose of object references in ORACLE? For the following ER diagram, where all the relationships are many-many, write the statements that form object tables with necessary references. (8)



b. What primary characteristics should an Object Identifier possess? (2)

9. a. How do regular inheritance, multiple inheritance and selective inheritance differ (5)

b. How XML schema differ from XML DTD. Illustrate with an example (5)