Reg No.:_____   Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

First Semester M.Tech Degree Regular and Supplementary Examination December 2023 (2022 Scheme)

COMPUTER SCIENCE AND ENGINEERING

221TCS001 ADVANCED DATABASE MANAGEMENT

Max. Marks: 60   Duration: 2.5 Hours

PART A

Answer all questions. Each question carries 5 marks   Marks

1. Consider the below given SQL query on Online Book database and transform the SQL query into relational algebra expressions. Draw the initial query tree and then the optimized query tree after applying heuristics.   (5)

    SELECT P.P_ID, Pname, Address, Phone

    FROM BOOK B, PUBLISHER P

    WHERE P.P_ID = B.P_ID AND Category = 'Language Book';

2. What is an SQL injection attack? What precautions must be taken to prevent SQL injection attacks?   (5)

3. Differentiate between speed-up and scale-up. Describe any two factors responsible for sub-linear performance of speedup and scaleup.   (5)

4. Assume that the search result for the word "cat" from a set of animal names {"dog", "hen", "goat"} using bloom filters returns true. What could be inferred from this result? Explain the concept of false positives in bloom filters. How can the probability of false positives be reduced?   (5)

5. What is semi-structed data? Illustrate the use of XML for storing semi-structured data?   (5)

PART B

Answer any 5 questions. Each question carries 7 marks

6. Assume that the join operation is to be performed on two relations. If the tuples of the relations are physically sorted by the value of their respective join attributes, which algorithm can be used to perform the join operation? Explain with the help of an example and estimate the cost in terms of block transfers.   (7)

7   Consider the following statistics about a relational table, EMPLOYEE   (7)
    (EMP_ID, NAME, AGE, SALARY, DEPT_NO) There are 10,000 records
    stored in 2000 blocks. Also, following indexes are available:

    - A secondary index on the EMP_ID with 4 levels.
    - A secondary index on non-key attribute DEPT_NO with 2 levels and 4
      first level index blocks. There are 200 distinct values for DEPT_NO.

    The other relation DEPARTMENT (DEPT_ID, DNAME, MGRSSN) has 600
    records. BFR for DEPARTMENT table is 60 and number of blocks are 10.
    There exists a primary key on DEPT_ID with level 1.

    Estimate the cost of the queries given below: (Assume all records are in table)

    i. $\sigma_{DEPT\_NO >20}$ (EMPLOYEE) – Assume Linear search is used

    ii. Join operation EMPLOYEE $|X|_{DEPT\_ID}$ DEPARTMENT (using
         nested-loop join) Assume buffer has only 1 block.

8   Differentiate Mandatory Access control and Role-Based Access Control in the   (7)
    context of Database security.

    Assume a relation R contains the following values as given below and answer
    the questions that follow:

C_Id	CarName	Price(Lakh)		TupleClassification(TC)
1 U	Honda C	15	C	S
2 C	Toyota U	30	S	C
3 U	BMW S	45	TS	C

    (TS, S, C and U are the security classification levels)

    i.   Depict how the relation R will appear to a user with classification level
         C.

    ii. Provide an example of how polyinstantiation can be used in this relation
         R, to prevent a user with lower classification from modifying values at a
         higher classification level.

9   Apply Range Partitioning technique on the below relation Emp_table assuming   (7)
    there are three processors P_0, P_1 and P_2 and three disks D_0, D_1 and D_2. Given the
    partition vector as [2, 4] explain the steps and depict how the relation will be
    partitioned into partition0, partition1 and partition2.

Emp_table		
ENAME	GRADE	DNAME
SMITH	1	RESEARCH
BLAKE	4	SALES
FORD	4	RESEARCH
KING	5	ACCOUNTING
SCOTT	4	RESEARCH
MILLER	2	ACCOUNTING
TURNER	3	SALES
WARD	2	SALES
MARTIN	2	SALES
ADAMS	1	RESEARCH
JONES	4	RESEARCH
JAMES	1	SALES
CLARK	4	ACCOUNTING
ALLEN	3	SALES

10   Differentiate between nearness query and region query. Consider a query to find   (7)
all the restaurants within the geographic boundaries of a given town. Analyse if
this query could be categorized as a nearness query or a region query and
explain why.

11   Construct a valid XML document from the XML DTD given below. (Include all   (7)
the elements mentioned in the DTD)
Interpret how many instances of 'phone' element is possible for a 'person'
element in an XML document that follows this DTD. Also infer how many such
'person' elements could be allowed.

```
<!DOCTYPE company[
  <!ELEMENT company ((person|product)*)>
  <!ELEMENT person(ssn,name,office,phone?)>
  <!ELEMENT ssn (#PCDATA)>
  <!ELEMENT name (#PCDATA)>
  <!ELEMENT office (#PCDATA)>
  <!ELEMENT phone (#PCDATA)>
  <!ELEMENT product (pid, name, description?)>
  <!ELEMENT ssn (#PCDATA)>
  <!ELEMENT pid (#PCDATA)>
  <!ELEMENT description (#PCDATA)>
]>
```

12   Develop the XSD for the XML document given below: (7)

```
<?xml version="1.0"?>
<ORDERDETAILS>
<ORDER>
```

```
    <ORDERID>1234</ORDERID>
    <PRODUCTID>0090</PRODUCTID>
    <ORDERDATE>23-07-2013</ORDERDATE>
    <CUSTOMERID>x4564</CUSTOMERID>
</ORDER>
 <ORDER>
    <ORDERID>1235</ORDERID>
    <PRODUCTID>0030</PRODUCTID>
    <ORDERDATE>23-07-2013</ORDERDATE>
    <CUSTOMERID>x4334</CUSTOMERID>
</ORDER>
</ORDERDETAILS>
```

*****