National University of Computer and Emerging Sciences

Lahore Campus

Advance Database Concepts (CS4064)

Date: Tue, 27 May 2025

Course Instructor(s)

Muhammad Ishaq Raza

Muhammad Naveed

Final Exam

Total Time (Hrs.): 3

Total Marks: 60

Total Questions: 6

Roll No _____ Section _____ Student Signature _____

Note: Please ensure that you attempt all questions and their respective parts in the given order.

CLO # 3: To develop a solution for given scenario/challenging problem in the domain of DB systems.

Q. No 1: Consider the following database that keeps track of the student representatives of different departments. Each student representative is selected for a period of one year and works under the supervision of a faculty member i.e. Advisor. The attributes AdvisorID and StudentID are the foreign keys from faculty and student relations respectively. [10]

Student (SID, SName, SGender, Birthdate, DegreeProgram, Batch, SDept)
Faculty (FacultyID, FName, FGender, MaritalStatus, FDept)
StudentRep (StudentID, Year, AdvisorID)

```
SELECT SID, SName AS StudentRepName, SDept, Year, AdvisorID
FROM Student JOIN StudentRep ON SID=StudentID JOIN Faculty ON AdvisorID=FacultyID
WHERE DegreeProgram= 'CS' AND Birthdate>= '01-Jan-2005';
```

$\pi_{SID, SName, SDept, Year, AdvisorID}(\sigma_{DegreeProgram='CS' \wedge Birthdate='01\text{-}Jan\text{-}2005'}(Student \bowtie_{SID=StudentID} StudentRep \bowtie_{AdvisorID=FacultyID} Faculty))$

Your task is to optimize the above query and draw the best possible query tree. Take appropriate database statistics to support your answer.

CLO # 3: To develop a solution for given scenario/challenging problem in the domain of DB systems.

Q. No 2: [6+4= 10]

a. Consider the above student representatives database schema and query. Assume that the frequency of access of this query is very high. Identify the attributes that are more appropriate to create indexes to improve the performance of this query and mention the type of each index (B-tree, Hash, or Bitmap).

b. Consider the above student representatives database schema and assume that *Student, Teacher* and *StudentRep* tables have *40000, 2000* and *10000* rows respectively. Estimate the potential join selectivity (js), join cardinality (jc), join selection factor of Student table, and join selection factor of StudentRep table for the query: ***Student*** $\bowtie_{SID=StudentID}$ ***StudentRep***. Justify your answer.

2.5×10^{-5} 10 000 0.25 1

CLO # 3: To develop a solution for given scenario/challenging problem in the domain of DB systems.

Q. No 3: Consider a disk with block size *B=512* bytes. A block pointer is *P=8* bytes long, and a record pointer is P_R=9 bytes long. A table has *r=40,000* Student records of fixed length and ordered on SID (i.e. student ID) key-attribute. Record length R is *100* bytes long and *DegreeProgram* attribute is 12 bytes long. Suppose there is a secondary index (multi-level) on non-key attribute *DegreeProgram* using option with an extra level of indirection that stores record pointers. Assume there are *50* distinct values of *DegreeProgram*, and that the Student records are evenly distributed among these values. For each of the following selection queries, estimate the I/O cost of the best possible solution. Take any valid assumption where needed. Show your working. [10]

a. SELECT * FROM student WHERE DegreeProgram= 'CS'; 817
b. SELECT * FROM student WHERE SID IN (1, 5, 13); 3x3 = 39

CLO # 2: Apply the models and approaches to become enabled to select and apply appropriate methods for a particular case.

Q. No 4:

a. Consider the following schedule: [3+3= 6]

S: r1(X); r2(Y); w1(X); w2(Y); r3(X); w3(X); r4(Y); w4(Y).

Draw the serializability (precedence) graph for this schedule. State whether this schedule is conflict-serializable (correct) or not. If the schedule is conflict-serializable, write down the equivalent serial schedule(s) otherwise explain why it is not. *Also state whether this schedule is view-serializable or not.*

b. Determine whether each schedule is strict, cascadeless, recoverable, or non-recoverable. Provide proper reason.

S1: r1(X); w1(X); r2(Y); w2(Y); c1; r3(Z); w3(Z); c2; r4(X); w4(X); c3; r5(Y); w5(Y); c4; c5. Strict
S2: r1(X); w1(X); r2(X); w2(Y); c2; c1; r3(Y); w3(Y); c3; r4(Z); w4(Z); c4; r5(X); w5(X); c5. none
S3: r1(X); w1(X); r2(X); w2(X); r3(Y); w3(Y); r4(Z); w4(Z); r5(X); w5(X); c1; c2; c3; c4; c5. recoverable

CLO # 2: Apply the models and approaches to become enabled to select and apply appropriate methods for a particular case.

Q. No 5: Consider the following schedule of actions: [9]

S: r1(X); r2(Y); w1(X); w2(Y); r3(X); w3(X); r4(Y); w4(Y), c1, c2, c3, c4.

For each of the following concurrency control mechanisms, describe how the concurrency control mechanism handles the schedule. Assume that the timestamp of transaction *Ti* is *i*. For lock-based concurrency control mechanisms, add lock and unlock requests to the above schedule of actions as per the locking protocol. The DBMS processes actions in the order shown. If a transaction is blocked, assume that all its actions are queued until it is resumed; the DBMS continues with the next action (according to the listed schedule) of an unblocked transaction.

a. Rigorous 2PL with timestamps used for deadlock avoidance (Use wound-wait policy)
b. Basic Timestamp Ordering (Assume T1 < T2 < T3)
c. Optimistic concurrency control technique (Use defer the validation until a later time when the conflicting transactions have finished.)

CLO # 1: Understanding advance data models, technologies, and approaches for building DB systems.

Q. No 6: [15]

a. List the main categories of NOSQL systems. MongoDB document-based Column wide. KeyValue graph based XML

b. Why is data replication useful in distributed databases? What typical units of data are replicated?

c. What are intra-query and inter-query parallelisms? Which one is harder to achieve in the shared-nothing architecture?

d. Describe the write-ahead logging protocol.

e. Describe the three phases of the ARIES recovery method.