COMSATS University
Islamabad
Lahore Campus

Semester Fall 2025

Instructor's Name: **Sardar Waqar Khan** **Subject:** **Advanced Database Systems**
Course Code **Assignment #:** 02
Total Marks 50 **Assigning Date: 31-10-2025**
Due Date: 8-11-2025 **CLO 2**

- Provide clear and complete answers with supporting examples or small diagrams.
- Write your responses in your own words; plagiarism will result in zero marks.
- Your responses should be handwritten and submit during lecture.
- There must be a **cover page** contain university logo and student's complete information.

Scenario:

The MetroBank database maintains the following relations:

Branch(branchNo, branchName, city)
Employee(empID, empName, position, salary, branchNo)
Account(accNo, balance, accType, branchNo, empID)
Customer(custID, custName, city, contact)

The database is distributed across multiple regional servers. As a database engineer, you are asked to analyze and optimize the following SQL query:

```
SELECT e.empName, e.position, c.custName, a.accType, a.balance
FROM Employee e, Branch b, Account a, Customer c
WHERE e.branchNo = b.branchNo
  AND e.empID = a.empID
  AND a.branchNo = b.branchNo
  AND c.city = b.city
  AND a.balance > 100000
  AND e.position = 'Manager'
  AND b.city = 'Islamabad';
```

Tasks:

1. Query Decomposition

a. Perform analysis and build a query tree representation.
b. Convert the WHERE clause into conjunctive normal form (CNF).
c. Identify any semantic or contradictory conditions, if present.
d. Simplify the query by removing redundant predicates.

2. Heuristic Optimization

Using heuristic rules, transform the query tree into an optimized relational algebra tree.
Clearly show how each heuristic is applied, such as:
- Performing SELECT and PROJECT operations early.
- Applying the most restrictive conditions first.
- Minimizing the size of intermediate relations before JOIN operations.

3. Cost-Based Optimization

Assume the following catalog statistics:
- Employee: 2000 tuples
- Branch: 100 tuples
- Account: 3000 tuples
- Customer: 1500 tuples
- Disk block size: 10 tuples/block
- Access cost per block: 10 ms
- Average selectivity for each condition: 10%

Estimate:
a. Total I/O cost for unoptimized query execution (Cartesian product method).
b. Total I/O cost after heuristic optimization.
c. Discuss which plan is optimal and why.

4. Distributed Consideration

If branches in Islamabad are stored on a separate regional server, discuss how this affects communication cost.
Suggest an execution strategy that minimizes communication overhead.

Summarize how both heuristic and cost-based optimization improve query performance in this scenario.