Final F_2023-06-01_key

Q No	Correct
Database Final - Section 1	
1	Invalid
2	Invalid
3	Invalid
4	A
5	B
6	A
7	D
8	C
9	C
10	C
11	B
12	B
13	A
14	A
15	C
16	A
17	C
18	Invalid
19	Invalid
20	D
21	C
22	B
23	C
24	A
25	A
26	A
27	A
28	C
29	A
30	A
31	A
32	B
33	A
34	D
35	B
36	B
37	B
38	C
39	A
40	C
41	B
42	B
43	B
44	D
45	B
46	A
47	A
48	A
49	A
50	C

1. Were you sleeping in the online classes? Hint: You can choose any option other than the option d.
   a. Yes, at times.
   b. To tell you the truth, yes mostly I was sleeping.
   c. No, I listened to your calm soothing voice and … Zzz
   d. What kind of question this is, I will email the HoD

2. Imagine you have finally graduated (I know it seems far-fetched but just assume for the time being). You have bills to pay and suddenly you noticed an overlap between the bills by PTCL and IESCO even having the same deadline … what would you do? Hint: The option d is the correct option.
   a. This is the violation of our Country's policy. I will send an email to the Prime Minister.
   b. Sir? I will email HoD after this paper, for sure.
   c. I will try to pay at 11:59PM and then will send an email to IESCO and PTCL saying I was just one second late. They always accept unlike you, my dear instructors.
   d. This is reality, I will manage my time and pay bills on time.

3. If your course instructor shows the following text in the first lecture of the course, how much he should deduct points once a student is caught cheating?
   All parties involved in any kind of cheating in any exam (Quizzes, Assignments & Projects) will get zero in that evaluation.
   a. A warning and no deduction is sufficient, students didn't think you really meant that. You seem to be so generous (as currently are giving us these easy marks) so why would you make yourself bad
   b. +50%, deduct 50% marks. We will then freely plagiarize after some probability computation to compensate for the cases when one gets caught.
   c. None of the above

4. The mapping between the conceptual schema and the internal schema is done through:
   a. Data independence
   b. Data integrity
   c. Data normalization
   d. Data encapsulation

5. Which schema representation is often defined using a data modeling language such as Entity-Relationship (ER) diagrams or Unified Modeling Language (UML)?
   a. Internal schema
   b. Conceptual schema
   c. Physical schema
   d. Logical schema

6. Which database schema representation is concerned with the specific storage structures, access methods, and physical organization of data on storage media?
   a. Internal schema
   b. Conceptual schema
   c. Physical schema
   d. Logical schema

7. In a database system, which schema representation is closest to the user's view and describes the logical structure of the entire database?
   a. Internal schema
   b. Conceptual schema
   c. Physical schema

d. External schema

8. Which database schema representation focuses on the actual storage and indexing structures used by the database management system?
   a. Internal schema
   b. Conceptual schema
   c. Physical schema
   d. Logical schema
9. What does EER stand for in the context of database modeling?
   a. Enhanced Entity-Relationship
   b. Efficient Entity-Relationship
   c. Extended Entity-Relationship
   d. Exclusive Entity-Relationship
10. Which of the following features is unique to the EER model compared to the traditional ER model?
   a. Attributes
   b. Relationships
   c. Hierarchies
   d. Keys
11. In the EER model, what is a specialization?
   a. A process of merging entities into a single entity
   b. A process of splitting entities into multiple specialized entities
   c. A process of combining relationships into a single relationship
   d. A process of refining attribute definitions within an entity
12. What is a weak entity in the EER model?
   a. An entity that does not have any attributes
   b. An entity that relies on another entity for its existence and has a partial key
   c. An entity that represents a general category in a hierarchical structure
   d. An entity that is not connected to any other entity through a relationship
13. What is a disjoint constraint in the EER model?
   a. It ensures that each entity instance can be a member of only one subclass
   b. It enforces the uniqueness of values within a specific attribute of an entity
   c. It ensures that a relationship can only exist between entities of different types
   d. It defines the maximum number of instances that can participate in a relationship
14. Which of the following is a characteristic of a multivalued attribute in the EER model?
   a. It can have multiple values for each instance of an entity.
   b. It is derived from other attributes within the entity.
   c. It represents a relationship between two or more entities.
   d. It cannot be a part of the primary key of an entity.
15. Data veracity can be improved by:
   a. Increasing data volume
   b. Implementing strong data encryption
   c. Conducting regular data quality checks
   d. Accelerating data velocity
16. Which statement best describes structured data?
   a. It is data that is organized into a predefined format with well-defined relationships between entities.
   b. It is data that does not follow any specific organization or formatting rules.
   c. It is data that is primarily text-based and lacks any predefined structure.

d. It is data that is stored in a distributed and decentralized manner.

17. What is a characteristic of unstructured data?
    a. It is easily queryable and allows for efficient analysis.
    b. It typically resides in relational databases with predefined schemas.
    c. It requires preprocessing and extraction techniques to derive meaning from it.
    d. It is highly structured and follows strict formatting guidelines.

18. The V of big data that emphasizes the quality and reliability of data is:
    c. Volume
    a. Velocity
    b. Variety
    c. Veracity

19. Which V of big data highlights the need for advanced analytics and real-time processing?
    d. Volume
    a. Velocity
    b. Variety
    c. Veracity

20. Veracity issues can lead to:
    a. Increased data privacy concerns
    b. More efficient data storage
    c. Improved data scalability
    d. Inaccurate insights and decision-making

21. Which of the following best describes Big Data?
    a. Data that can be stored and processed using traditional database systems.
    b. Data that is generated at a slow pace and easily managed using standard tools.
    c. Data that is characterized by its volume, variety, and velocity.
    d. Data that is limited to structured formats such as spreadsheets.

22. The JOIN operation in relational algebra is equivalent to which SQL clause?
    a. SELECT
    b. JOIN
    c. GROUP BY
    d. WHERE

23. Given two relations R(A, B) and S(B, C), where R contains 10 tuples and S contains 5 tuples, what will be the maximum number of tuples in the resulting relation after performing the natural join?
    a. 5
    b. 10
    c. 15
    d. 50

24. Given two relations R(A, B) and S(B, C), what will be the result of the natural join R ⋈ S if R has 100 tuples and S has 50 tuples, but there are no common values in attribute B?
    a. 0 tuples
    b. 50 tuples
    c. 100 tuples
    d. 5000 tuples

25. Consider a relation R(A, B, C) and a selection condition σ(B > 10, C = 5). Which of the following relational algebra expressions correctly represents the selection operation?
    a. σ(B > 10 AND C = 5)(R)

b. $\sigma(B > 10 \text{ OR } C = 5)(R)$
c. $\pi(B > 10 \text{ AND } C = 5)(R)$
d. $\pi(B > 10 \text{ OR } C = 5)(R)$

26. Which SQL query will calculate the total sales amount for each product?
Output: ProductID | ProductName | TotalSales
---
1 | Apple | 1000
2 | Banana | 1500
3 | Orange | 800

a) SELECT ProductID, ProductName, SUM(SalesAmount) AS TotalSales FROM Products GROUP BY ProductID;
b) SELECT ProductID, ProductName, COUNT(SalesAmount) AS TotalSales FROM Products GROUP BY ProductID;
c) SELECT ProductID, ProductName, AVG(SalesAmount) AS TotalSales FROM Products GROUP BY ProductID;
d) SELECT ProductID, ProductName, MAX(SalesAmount) AS TotalSales FROM Products GROUP BY ProductID;

27. Which SQL query will retrieve the top 5 customers with the highest total order amounts?
Output: CustomerID | CustomerName | TotalOrderAmount
---
1 | John | 5000
2 | Jane | 4500
3 | Mark | 4000

a) SELECT TOP 5 CustomerID, CustomerName, SUM(OrderAmount) AS TotalOrderAmount FROM Orders GROUP BY CustomerID ORDER BY TotalOrderAmount DESC;
b) SELECT TOP 5 CustomerID, CustomerName, SUM(OrderAmount) AS TotalOrderAmount FROM Orders GROUP BY CustomerID ORDER BY TotalOrderAmount ASC;
c) SELECT CustomerID, CustomerName, SUM(OrderAmount) AS TotalOrderAmount FROM Orders GROUP BY CustomerID ORDER BY TotalOrderAmount DESC LIMIT 5;
d) SELECT CustomerID, CustomerName, SUM(OrderAmount) AS TotalOrderAmount FROM Orders GROUP BY CustomerID ORDER BY TotalOrderAmount ASC LIMIT 5;

28. Which SQL query will retrieve the order details along with the customer name and order date?
Output: OrderID | CustomerName | OrderDate
---
1 | John | 2022-05-10
2 | Jane | 2022-05-12
3 | Mark | 2022-05-15

a) SELECT Orders.OrderID, Customers.CustomerName, Orders.OrderDate FROM Orders, Customers WHERE Orders.CustomerID = Customers.CustomerID;
b) SELECT OrderID, CustomerName, OrderDate FROM Orders LEFT JOIN Customers ON Orders.CustomerID = Customers.CustomerID;
c) SELECT Orders.OrderID, Customers.CustomerName, Orders.OrderDate FROM Orders JOIN Customers ON Orders.CustomerID = Customers.CustomerID;
d) SELECT OrderID, CustomerName, OrderDate FROM Orders RIGHT JOIN Customers ON Orders.CustomerID = Customers.CustomerID;

29. Which SQL query will retrieve the total sales amount for each product along with the product name?
Output: ProductID | ProductName | TotalSales
---
1 | Apple | 1000
2 | Banana | 1500

3 | Orange | 800

a) SELECT ProductID, ProductName, SUM(SalesAmount) AS TotalSales FROM Products JOIN Sales ON Products.ProductID = Sales.ProductID;
b) SELECT ProductID, ProductName, SUM(SalesAmount) AS TotalSales FROM Products LEFT JOIN Sales ON Products.ProductID = Sales.ProductID;
c) SELECT ProductID, ProductName, SUM(SalesAmount) AS TotalSales FROM Products RIGHT JOIN Sales ON Products.ProductID = Sales.ProductID;
d) SELECT ProductID, ProductName, SUM(SalesAmount) AS TotalSales FROM Products, Sales WHERE Products.ProductID = Sales.ProductID;

30. Which SQL query will retrieve the employee details along with their department name, including employees without assigned departments?
Output: EmployeeID | EmployeeName | DepartmentName

---

1 | John | Sales
2 | Jane | HR
3 | Mark | IT
4 | Sarah | NULL

a) SELECT Employees.EmployeeID, Employees.EmployeeName, Departments.DepartmentName FROM Employees LEFT JOIN Departments ON Employees.DepartmentID = Departments.DepartmentID;
b) SELECT EmployeeID, EmployeeName, DepartmentName FROM Employees RIGHT JOIN Departments ON Employees.DepartmentID = Departments.DepartmentID;
c) SELECT Employees.EmployeeID, Employees.EmployeeName, Departments.DepartmentName FROM Employees JOIN Departments ON Employees.DepartmentID = Departments.DepartmentID;
d) SELECT EmployeeID, EmployeeName, DepartmentName FROM Employees, Departments WHERE Employees.DepartmentID = Departments.DepartmentID;

31. Which SQL query will retrieve the details of all customers along with their corresponding orders, even if they have not placed any orders yet?
Output: CustomerID | CustomerName | OrderID | OrderDate

---

1 | John | 1 | 2022-05-10
2 | Jane | 2 | 2022-05-12
3 | Mark | 3 | 2022-05-15
4 | Sarah | NULL | NULL

a) SELECT Customers.CustomerID, Customers.CustomerName, Orders.OrderID, Orders.OrderDate FROM Customers LEFT JOIN Orders ON Customers.CustomerID = Orders.CustomerID;
b) SELECT CustomerID, CustomerName, OrderID, OrderDate FROM Customers RIGHT JOIN Orders ON Customers.CustomerID = Orders.CustomerID;
c) SELECT Customers.CustomerID, Customers.CustomerName, Orders.OrderID, Orders.OrderDate FROM Customers JOIN Orders ON Customers.CustomerID = Orders.CustomerID;
d) SELECT CustomerID, CustomerName, OrderID, OrderDate FROM Customers, Orders WHERE Customers.CustomerID = Orders.CustomerID;

32. Every serializable schedule is conflict-serializable.
a) True
b) False

33. For any schedule S1 and any serial schedule S2, if S1 is conflict equivalent to S2, then S1 is conflict serializable.
a) True
b) False

34. Which of the following is not a feature of TRANSACTION?
a) Users should be able to regard the execution of each transaction as atomic.
b) Each transaction, run by itself.
c) must preserve the consistency of the database
d) dependent on other concurrent transactions

35. While recovering data, which of the following files does a recovery manager examines at first?
a) A system file
b) Log file
c) Data dictionary
d) Metadata

36. Identify the correct statement with respect to normalization.
a) Normalization is a formal technique that can be used only at the starting phase of the database design.
b) Normalization can be used as a top-down standalone database design technique.
c) The process of normalization through decomposition must achieve the lossless join property at any cost whereas the dependency reservation property is sometimes sacrificed.
d) The process of normalization through decomposition must achieve the dependency reservation property at any cost whereas the lossless join property is sometimes sacrificed.

37. Consider the following two transactions:

T1: Read1(X), Write1(X), Read1(Y), Write1(Y)
T2: Read2(X), Write2(X), Read2(Y), Write2(Y)

What is a Conflict Equivalent schedule for the above serial Schedule (T1->T2)
a) Read1(X), Read2(X), Write2(X), Read1(Y), Write1(Y). Write1(X), Read2(Y), Write2(Y)
b) Read1(X), Write1(X), Read1(Y), Write1(Y), Read2(X), Write2(X), Read2(Y), Write2(Y)
c) Read1(Y), Read2(Y), Write2(Y), Read2(X), Write2(X), Read1(X), Write1(X), Write1(Y)
d) Write1(X), Write1(Y), Write2(Y), Read2(X), Write2(X), Read1(X)

38. Consider the following two transactions:

T1: $Read_1(X)$, $Write_1(X)$, $Read_1(Y)$, $Write_1(Y)$
T2: $Read_2(X)$, $Write_2(X)$, $Read_2(Y)$, $Write_2(Y)$

What is a Serializable schedule for the above serial Schedule (T1->T2)
a) $Read_1(X)$, $Read_2(X)$, $Write_2(X)$, $Read_1(Y)$, $Write_1(Y)$. $Write_1(X)$, $Read_2(Y)$, $Write_2(Y)$
b) $Read_1(Y)$, $Read_2(Y)$, $Write_2(Y)$, $Read_2(X)$, $Write_2(X)$, $Read_1(X)$, $Write_1(X)$, $Write_1(Y)$
c) $Read_1(X)$, $Write_1(X)$, $Read_1(Y)$, $Write_1(Y)$, $Read_2(Y)$, $Write_2(Y)$, $Read_2(X)$, $Write_2(X)$
d) $Write_1(X)$, $Write_1(Y)$,$Write_2(Y)$, $Read_2(X)$, $Write_2(X)$, $Read_1(X)$

39. Consider the following two transactions:

T1: $Read_1(X)$, $Write_1(X)$, $Read_1(Y)$, $Write_1(Y)$
T2: $Read_2(X)$, $Write_2(X)$, $Read_2(Y)$, $Write_2(Y)$

Which of the following schedule shows a Lost Update problem?
a) $Read_1(X)$, $Read_2(X)$, $Write_2(X)$, $Read_1(Y)$, $Write_1(Y)$. $Write_1(X)$, $Read_2(Y)$, $Write_2(Y)$

b) $Read_1(X)$, $Write_1(X)$, $Read_1(Y)$, $Write_1(Y)$, $Read_2(X)$, $Write_2(X)$, $Read_2(Y)$, $Write_2(Y)$
c) $Read_1(X)$, $Write_1(X)$, $Read_1(Y)$, $Write_1(Y)$, $Read_2(X)$, $Write_2(X)$, $Read_2(Y)$, $Write_2(Y)$
d) $Read_2(X)$, $Write_2(X)$, $Read_1(X)$, $Write_1(X)$, $Read_2(Y)$, $Write_2(Y)$, $Read_1(Y)$, $Write_1(Y)$

40. Consider the following two transactions:

T1: $Read_1(X)$, $Write_1(X)$, $Commit_1$
T2: $Read_2(X)$, $Write_2(X)$, $Commit_2$

Which of the following schedule shows a Temporary Update (Dirty Read) problem?

a) $Read_1(X)$, $Write_1(X)$, $Read_2(X)$, $Write_2(X)$, $Commit_1$, $Commit_2$
b) $Read_2(X)$, $Write_2(X)$, $Commit_2$, $Read_1(X)$, $Write_1(X)$, $Commit_1$
c) $Read_1(X)$, $Write_1(X)$, $Read_2(X)$, $Write_2(X)$, $Rollback_1$, $Commit_2$
d) $Read_1(X)$, $Write_1(X)$, $Read_2(X)$, $Write_2(X)$, $Rollback_1$, $Rollback_2$

41. Consider the SQL query SELECT * FROM R1 WHERE C1=4 AND C2=10 AND C3=11. Suppose R1 has one million records stored in 100 disk pages, C1 has 10,000 unique values, C2 has 100,000 unique values, and C3 has 1,000 unique values, and values are distributed uniformly. Which relational algebra expression leads to the most efficient query execution plan? [2 Marks]

a) $\sigma C1=4(\sigma C2=10(\sigma C3=11(R1)))$
b) $\sigma C3=11(\sigma C1=4(\sigma C2=10(R1)))$
c) $\sigma C2=10(\sigma C1=4(\sigma C3=11(R1)))$
d) $\sigma C3=11(\sigma C2=10(\sigma C1=4(R1)))$
e) all four are equally efficient

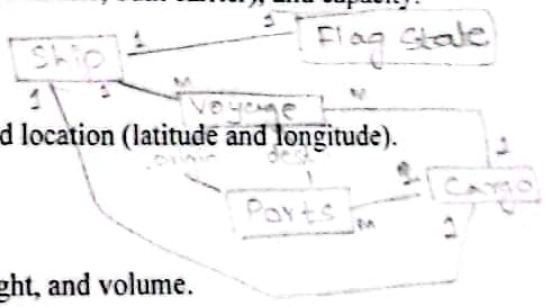
Consider the following case study for questions

Case Study: Maritime Agency

The Maritime Agency wants to develop a database system to manage information related to ships, ports, and cargo. They have provided the following requirements:

Ships:

- Each ship has a unique identification number, name, type (e.g., container, bulk carrier), and capacity.
- Ships can be registered under a specific flag state.
- Each ship can make multiple voyages.



Ports:

- Each port has a unique identification number, name, country, and location (latitude and longitude).
- Ports can handle various types of cargo.
- Ports can be either origin or destination for a voyage.

Cargo:

- Each cargo has a unique identification number, description, weight, and volume.
- Cargo can be transported on multiple voyages.
- Each cargo item is associated with a specific ship and port.

Voyages:

- Each voyage has a unique identification number, start date, end date, and status (e.g., in progress, completed).
- A voyage involves a single ship, an origin port, and a destination port.
- Multiple cargo items can be associated with a voyage.

Flag States:

- Each flag state has a unique identification number, name, and country.

42. Which relationship type represents the association between a ship and its voyages?

a. One-to-One

b) One-to-Many
c. Many-to-One
d. Many-to-Many

43. What is the cardinality of the relationship between a voyage and its associated cargo items?
a. One-to-One
b) One-to-Many
c. Many-to-One
d. Many-to-Many

44. Which of the following attributes would be part of the "Cargo" entity?
a. Voyage ID
b. Port Name
c. Ship Type
d) Cargo Weight

45. Which relationship type represents the association between a port and its handled cargo?
a. One-to-One
b) One-to-Many
c. Many-to-One
d. Many-to-Many

46. Which SQL query can be used to retrieve the names of ships along with their corresponding flag state names?

a) SELECT Ships.name, FlagStates.name FROM Ships JOIN FlagStates ON Ships.flag_state_id = FlagStates.flag_state_id;
b) SELECT Ships.name, FlagStates.name FROM Ships, FlagStates WHERE Ships.flag_state_id = FlagStates.flag_state_id;
c) SELECT Ships.name, FlagStates.name FROM Ships LEFT JOIN FlagStates ON Ships.flag_state_id = FlagStates.flag_state_id;
d) SELECT Ships.name, FlagStates.name FROM Ships RIGHT JOIN FlagStates ON Ships.flag_state_id = FlagStates.flag_state_id;

47. Which SQL query can be used to retrieve the names of ports that have handled at least one cargo item in the past year?

a) SELECT DISTINCT Ports.name FROM Ports JOIN Voyages ON Ports.port_id = Voyages.origin_port_id JOIN Cargo ON Voyages.voyage_id = Cargo.voyage_id WHERE Cargo.date >= DATE_SUB(NOW(), INTERVAL 1 YEAR);
b) SELECT DISTINCT Ports.name FROM Ports JOIN Voyages ON Ports.port_id = Voyages.origin_port_id JOIN Cargo ON Ports.port_id = Cargo.port_id WHERE Cargo.date >= DATE_SUB(NOW(), INTERVAL 1 YEAR);
c) SELECT DISTINCT Ports.name FROM Ports JOIN Voyages ON Ports.port_id = Voyages.origin_port_id JOIN Cargo ON Ports.port_id = Cargo.port_id WHERE Voyages.date >= DATE_SUB(NOW(), INTERVAL 1 YEAR);
d) SELECT DISTINCT Ports.name FROM Ports JOIN Voyages ON Ports.port_id = Voyages.origin_port_id JOIN Cargo ON Voyages.voyage_id = Cargo.voyage_id WHERE Voyages.date >= DATE_SUB(NOW(), INTERVAL 1 YEAR);

48. Which SQL query can be used to retrieve the ship names and the total weight of cargo transported by each ship?

a) SELECT Ships.name, SUM(Cargo.weight) FROM Ships JOIN Voyages ON Ships.ship_id = Voyages.ship_id JOIN Cargo ON Voyages.voyage_id = Cargo.voyage_id GROUP BY Ships.name;

b) SELECT Ships.name, Cargo.weight FROM Ships JOIN Voyages ON Ships.ship_id = Voyages.ship_id JOIN Cargo ON Voyages.voyage_id = Cargo.voyage_id GROUP BY Ships.name;
c) SELECT Ships.name, Cargo.weight FROM Ships LEFT JOIN Voyages ON Ships.ship_id = Voyages.ship_id LEFT JOIN Cargo ON Voyages.voyage_id = Cargo.voyage_id GROUP BY Ships.name;
d) SELECT Ships.name, SUM(Cargo.weight) FROM Ships JOIN Voyages ON Ships.ship_id = Voyages.ship_id JOIN Cargo ON Ships.ship_id = Cargo.ship_id GROUP BY Ships.name;

49. Which SQL query can be used to retrieve the voyages that have transported cargo items weighing more than 100 tons?

a) SELECT Voyages.* FROM Voyages JOIN Cargo ON Voyages.voyage_id = Cargo.voyage_id WHERE Cargo.weight > 100;
b) SELECT Voyages.* FROM Voyages JOIN Cargo ON Voyages.voyage_id = Cargo.voyage_id GROUP BY Voyages.voyage_id HAVING SUM(Cargo.weight) > 100;
c) SELECT Voyages.* FROM Voyages JOIN Cargo ON Voyages.voyage_id = Cargo.voyage_id WHERE SUM(Cargo.weight) > 100;
d) SELECT Voyages.* FROM Voyages JOIN Cargo ON Voyages.voyage_id = Cargo.voyage_id GROUP BY Voyages.voyage_id HAVING Cargo.weight > 100;

50. Which SQL query can be used to retrieve the ships that have not completed any voyages?

a) SELECT Ships.* FROM Ships JOIN Voyages ON Ships.ship_id = Voyages.ship_id WHERE Voyages.status <> 'Completed';
b) SELECT Ships.* FROM Ships JOIN Voyages ON Ships.ship_id = Voyages.ship_id WHERE Voyages.status = 'In Progress';
c) SELECT Ships.* FROM Ships LEFT JOIN Voyages ON Ships.ship_id = Voyages.ship_id WHERE Voyages.voyage_id IS NULL;
d) SELECT Ships.* FROM Ships JOIN Voyages ON Ships.ship_id = Voyages.ship_id WHERE Voyages.status IS NULL;

National University of Computer and Emerging Sciences

FAST School of Computing **Spring-2023** **Islamabad Campus**

Question 1 [20 Marks]

Consider the following shipping manifest, obtained from the customer:

Brokerage Account Summary Statement
Print Date: 01/06/2023

Client Name: Angela
Account Number: 456812
Agent: Smith

Order_No: C01-1234

Stock-ID	Description	Shares	Stock_Price	Total_Stock_Price
T	ATT Common	100	$300	$30,000
VZ	Verizon	200	$100	$20,000
PG	Proctor and Gamble	500	$150	$75,000

Portfolio Grand Total: $125,000

a. Describe all possible functional dependencies that describe the information in this form assuming Order_No as a candidate key. Also consider any multivalued attribute(s) while describing this information **[10 Marks]**

R(ClientName, AccNo, Agent, OrderNo, StockID, Descrip, Shares, Price)

CK = {Order_No}

AccNo → ClientName

StockID → Descrip, Shares, Price

AccNo + OrderNo → Agent

AccNo + OrderNo → StockID + Descrip

3

b. Develop a set of third normal form relational schema from these functional dependencies. Justify that your set of schema are actually in third normal form. [Hint: Perform normalization till 3rd Normal form for the relational schema based on your identified functional dependencies as discussed in the class] **[10 Marks]**

Un-Normalized Form:
1NF:
2NF:
3NF:

Question 2 [15 Marks]

Consider the following relational schema given below.

Student (SID, DID, Enroll_Year, Nationality)
DID references ~~Finance~~ Department
Department (DID, DName, Building_Num, Telephone, FID)
FID references Finance

Finance (FID, Budget, Expenses)

There are students having different nationalities in various departments and hence have different lifestyles (expenses).
Consider the following SQL Query that provides a list of students who are from the USA, enrolled in the Finance department, and have their associated department's budget less than $20,000

You have to write Relational Algebraic expression against this query and also generate query tree for the query given below; [Note: Relational Algebraic Symbol List is given on the last page]

SQL:

```
SELECT Student.SID, Student.DID, Student.Enroll_Year, Student.Nationality
FROM Student
JOIN Department ON Student.DID = Department.DID
JOIN Finance ON Department.FID = Finance.FID
WHERE Student.Nationality = 'USA'
 AND Department.DName = 'Finance'
 AND Finance.Budget <20000;
```

Relational Algebra: [5 Marks]

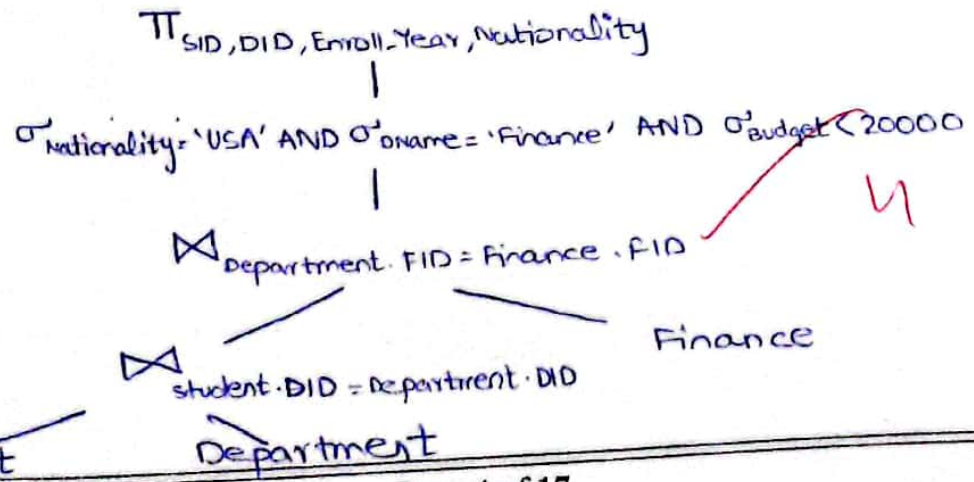
$\Pi_{student.SID, student.DID, student.Enroll\_Year, Student.Nationality}\left(\sigma_{student.Nationality='USA'}(Student)\right.$

$\bowtie_{student.DID = Department.DID}\left(\sigma_{dname='Finance'}(Department)\right)\bigg) \bowtie_{department.FID}$

$= Finance.FID\left(\sigma_{Budget<20000}(Finance)\right)\bigg)$

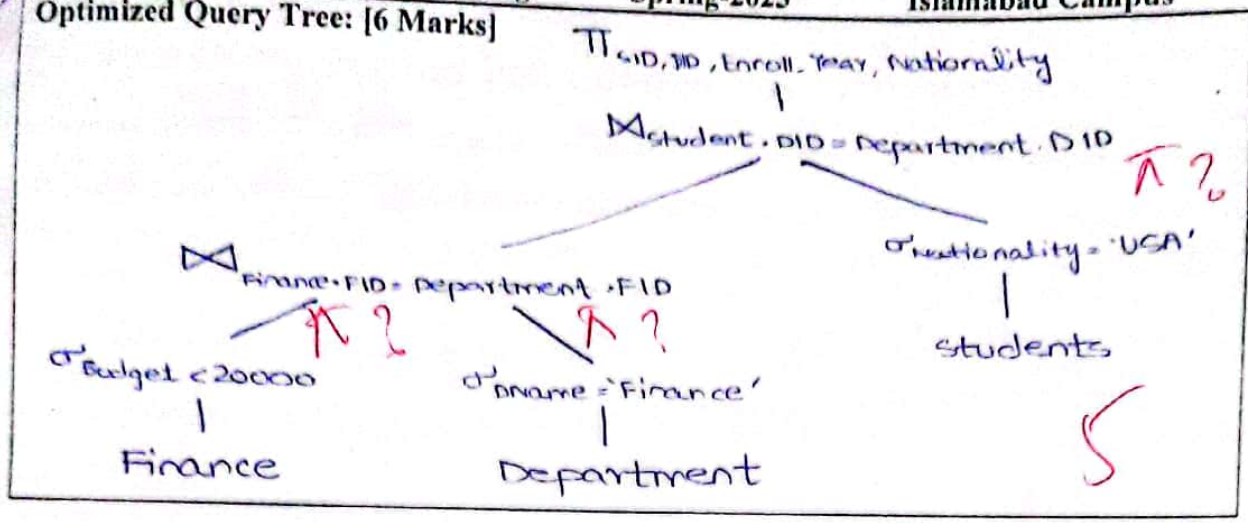
5

Basic Query Tree: [4 Marks]



4

Optimized Query Tree: [6 Marks]



Question 3: [10 Marks]

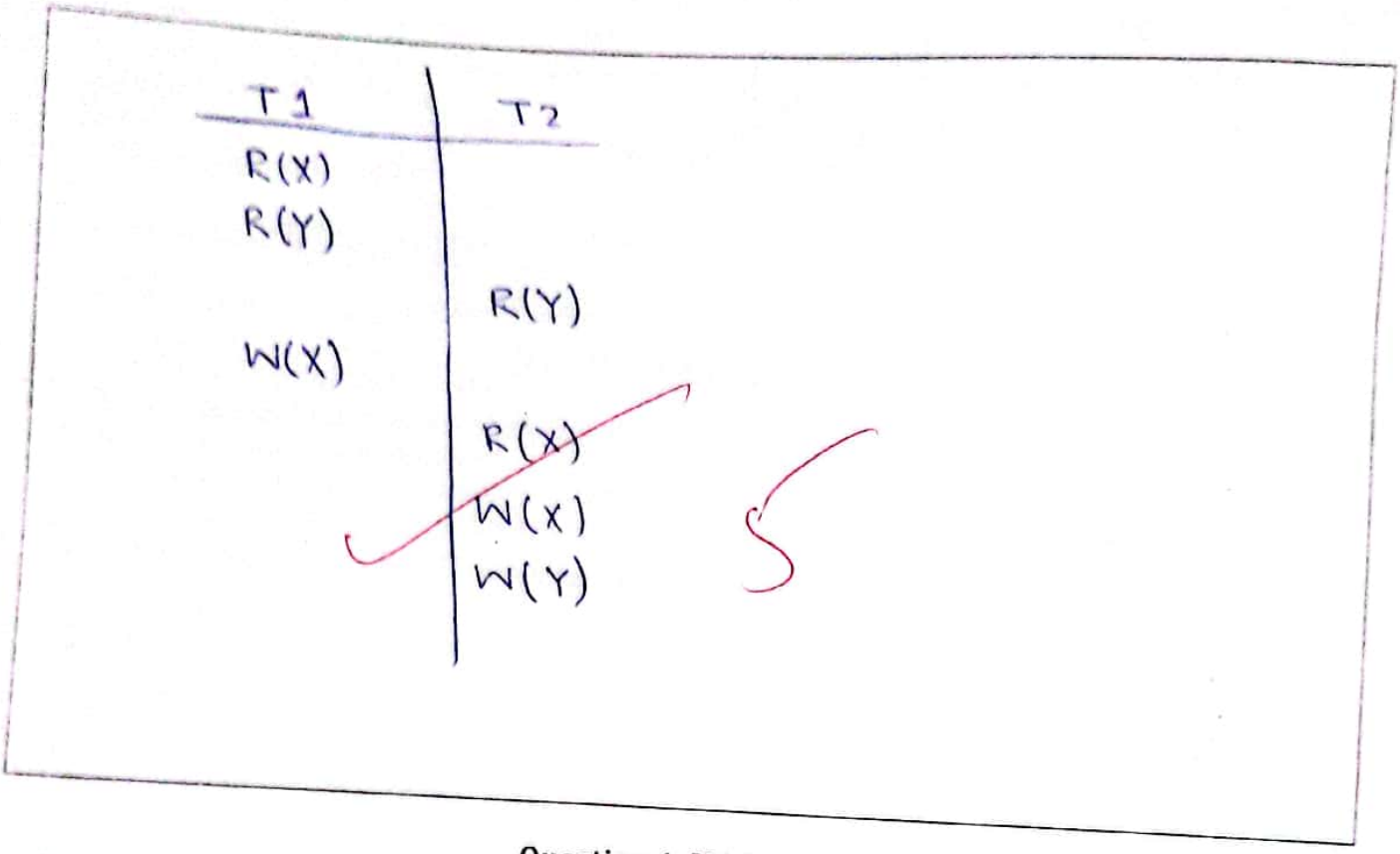
Consider a database with objects X and Y and assume that there are two transactions $T1$ and $T2$. Transaction $T1$ reads objects X and Y and then writes object X. Transaction $T2$ reads objects X and Y and then writes objects X and Y.

1. Give an example of a serial schedule with actions of transactions $T1$ and $T2$ on objects X and Y. **[5 marks]**

T1 : R(X) ; R(Y) ; W(X)
T2 : R(X) ; R(Y) ; W(X) ; W(Y)

T1	T2
R(X)	
R(Y)	
W(X)	
	R(Y)
	R(X)
	W(X)
	W(Y)

2. Give an example of a serializable schedule with actions of transactions $T1$ and $T2$ on objects X and Y. *[5 marks]*

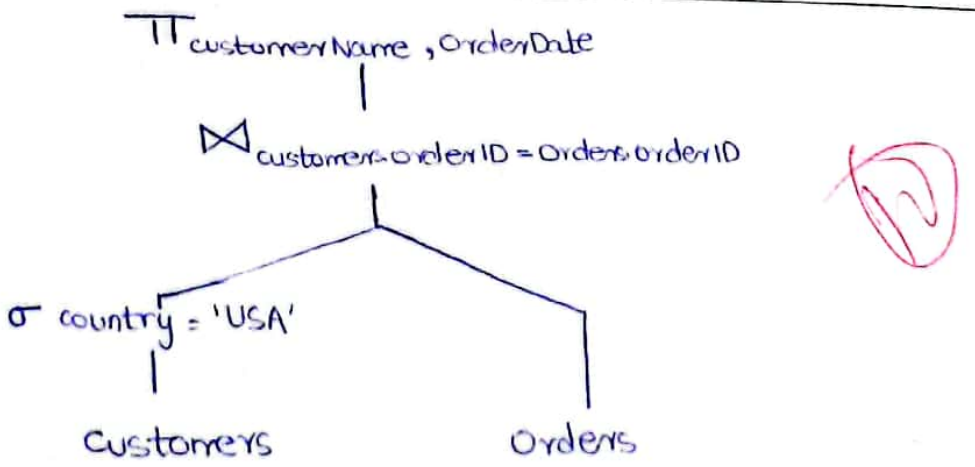


Question 4: [30 Marks]

a) Consider the following relational algebra expression:

π CustomerName, OrderDate (σ Country = 'USA' (Customers ⋈ Orders))

Optimize the above expression using a query tree by rearranging the operations and minimizing the intermediate results. Provide the optimized query tree representation of the expression. [10 marks]



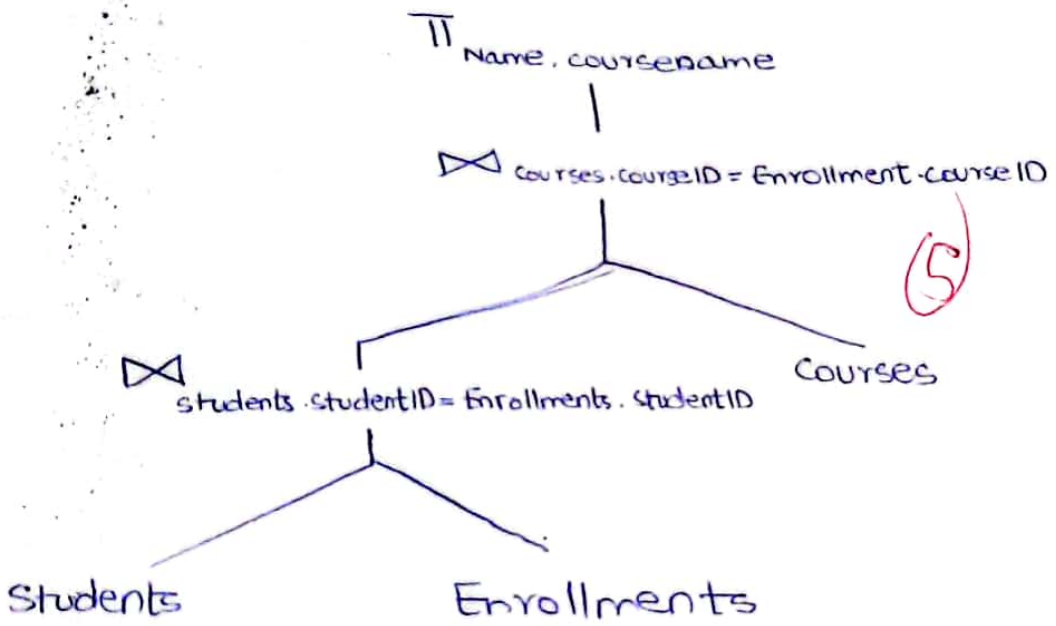
b) Consider the following query

SELECT Students.Name, Courses.CourseName FROM Students INNER JOIN Enrollments ON Students.StudentID = Enrollments.StudentID INNER JOIN Courses ON Enrollments.CourseID = Courses.CourseID;

Provide the relational algebra expression and query tree representation for the given SQL query [5+5].

$\Pi_{students.Name, courses.coursename}((Students \bowtie_{students.studentID = Enrollments.studentID} (Enrollments)) \bowtie_{courses.courseID = Enrollments.courseID} (Courses))$.

5



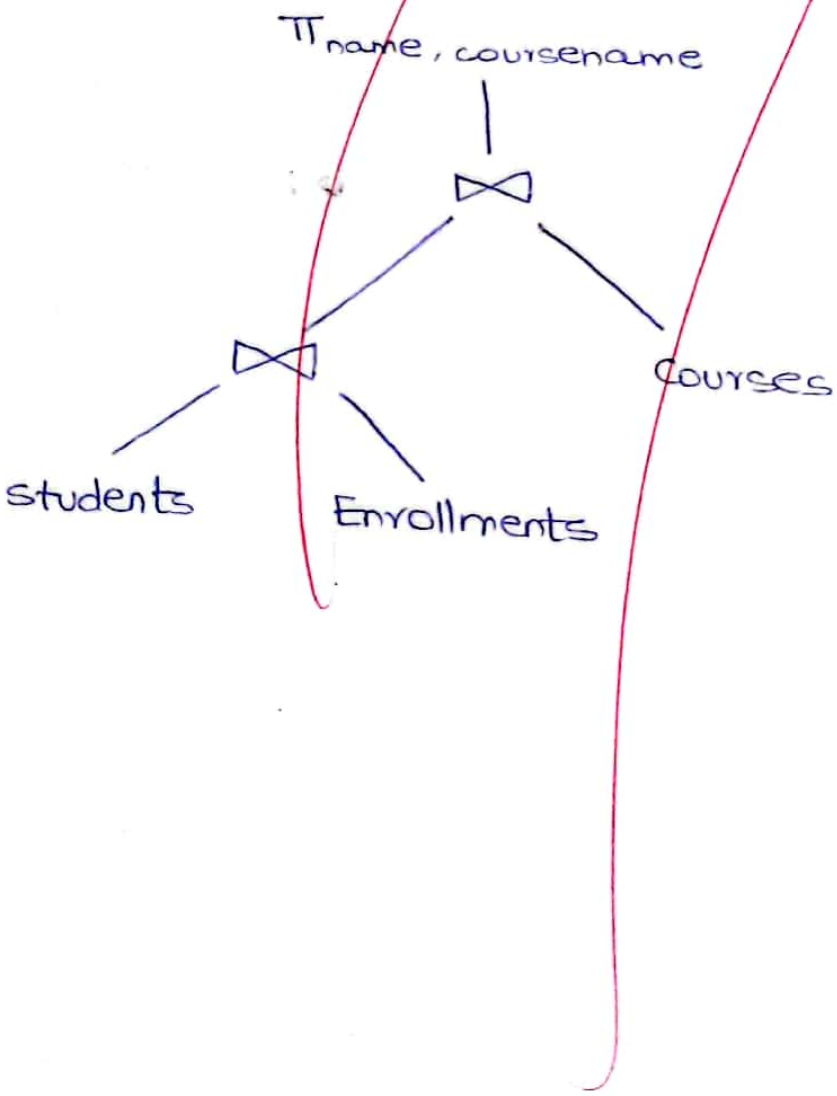
5

c) Consider two relational algebra expressions that perform the same task

Expression 1: π Name, CourseName ((Students $\bowtie$ Enrollments) $\bowtie$ Courses)
Expression 2: π Name, CourseName (Students $\bowtie$ (Enrollments $\bowtie$ Courses))

If we have 20 students, 50 enrollments and 50 courses, which expression will generate more optimized results. Answer precisely and make a query tree for the more optimized expression in this scenario [10 marks]

Expression 1 will generate more optimized results because it will reduce the datasize in first join as compared to that in Expression 2.



Question 5: [20 Marks]

a) Convert the queries below into Relational algebra expression (5+5)

SELECT branch_name, SUM(balance) AS sum_balance FROM account GROUP BY branch_name;

$\Pi_{branch\_name, sum\_balance}(_{branch\_name}\gamma_{sum(balance)\ as\ sum\_balance}(Account))$ 5

SELECT Department, COUNT(*) AS EmployeeCount, AVG(Salary) AS AverageSalary
FROM Employees GROUP BY Department;

$\Pi_{department, Employeecount, Averagesalary}(_{department}\gamma_{count(*)\ as\ Employeecount, AVG(salary)\ as\ Averagesalary}(Employees))$ 5

b) Let's say we have two relations, Employees and Managers, with the following attributes:

Employees (EmpID, Name, Department)
Managers (EmpID, Name, Department)
Create a relational algebra and SQL query for **all the employees who are not managers (5+5)**

$\Pi_{EmpID, Name, Department}(Employees - Managers)$ 5

```
Select * from Employees where EmpID NOT
Exist in (select * from Managers)
```
5

Question 6: [30 Marks]

Part (a) [20 Marks] Imagine you are working with "Summit Seekers," a reputable mountaineering organization that specializes in organizing and managing expeditions to some of the world's most challenging peaks. Summit Seekers collaborates with a trusted company called "Alpine Logistics" to ensure smooth expedition operations. ***Your task is to design a comprehensive data model using Entity-Relationship Diagram (ERD) for Summit Seekers, taking into account the various entities and their relationships.***

CASE STUDY : Summit Seekers

Summit Seekers is responsible for coordinating mountaineering expeditions, providing expert guides, and ensuring the safety and success of participants. They maintain crucial information about mountaineers, expeditions, guides, equipment, base camps, summits, expedition leaders, and the management company.

1. Mountaineer:
   - Represents individuals with a passion for mountaineering.
   - Includes attributes such as Name, Age, Nationality, and Contact Information.
2. Expedition:
   - Signifies a specific mountaineering expedition organized by Summit Seekers.
   - Contains details such as Expedition ID, Start Date, End Date, Difficulty Level, and Location.
3. Guide:
   - Comprises experienced mountaineers hired by Summit Seekers to lead expeditions.
   - Attributes include Guide ID, Name, Experience Level, and Specialties.
4. Equipment:
   - Encompasses the necessary gear and supplies required for successful expeditions.
   - Contains information about Equipment ID, Name, Type, and Availability.
5. Base Camp:
   - Represents the temporary camps set up during expeditions to provide essential facilities and accommodations.
   - Includes attributes such as Camp ID, Location, and Facilities.
6. Summit:
   - Refers to the peaks or summits targeted by the expeditions.
   - Captures details such as Summit ID, Mountain Name, Elevation, and Climbing Route.
7. Expedition Leader:
   - Denotes an experienced mountaineer assigned by Summit Seekers to lead an expedition.
   - Contains attributes such as Leader ID, Name, Experience Level, and Contact Information.
8. Management Company:
   - Represents "Alpine Logistics," the external company engaged by Summit Seekers to handle expedition logistics.
   - Includes information about Company ID, Name, Location, and Contact Information.

Design an ERD for Summit Seekers by incorporating the provided entities and their relevant attributes. Ensure that the relationships between the entities accurately represent the real-world scenarios and align with the mountaineering expedition management process.

In your ERD, ensure that you capture the relationships and associations between the entities accurately. Consider the following guidelines for establishing relationships based on the case study:

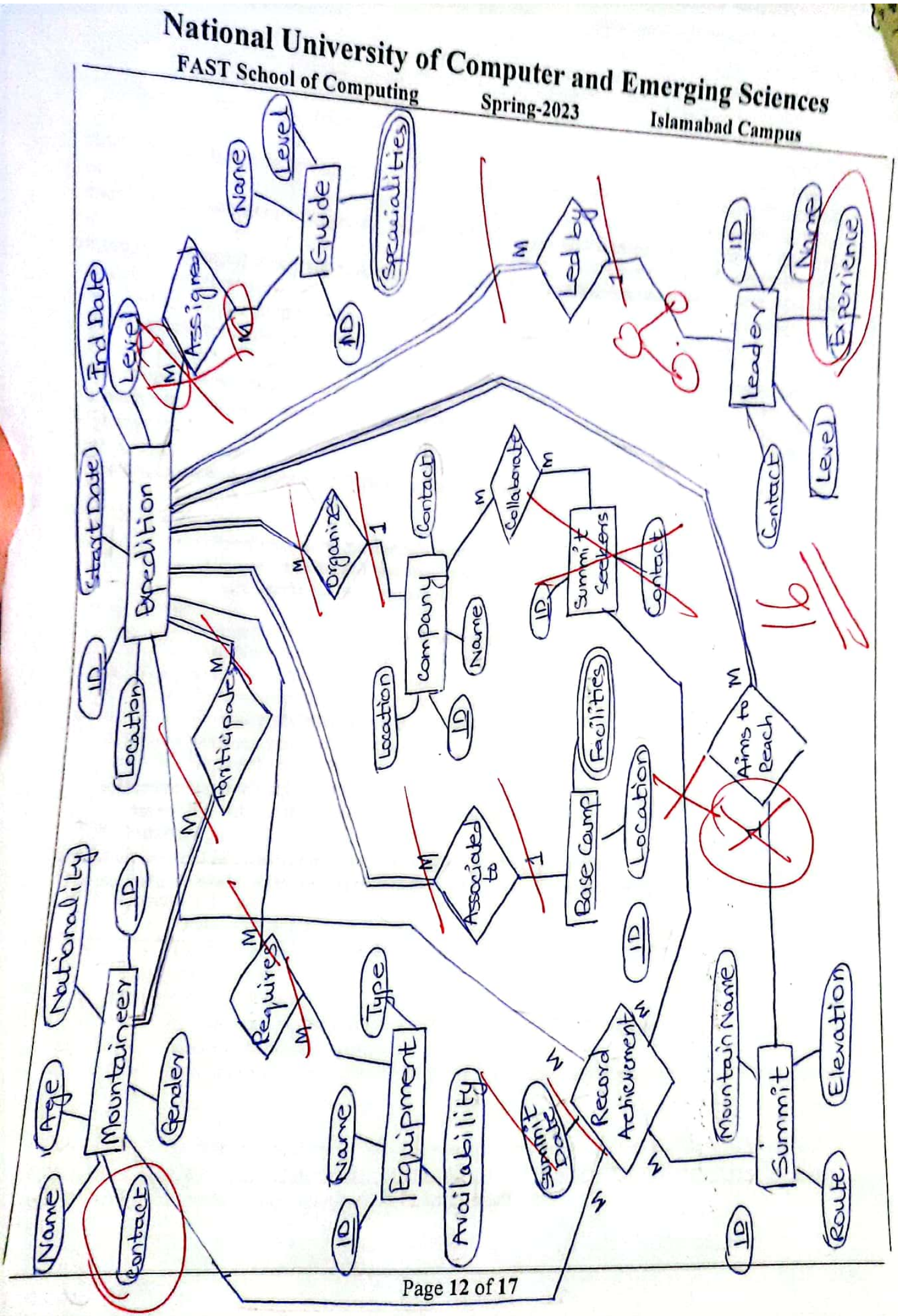
- Mountaineers participate in multiple Expeditions, and each Expedition involves multiple Mountaineers.
- Each Expedition is led by a single Expedition Leader, but an Expedition Leader can lead multiple Expeditions.
- Each Expedition requires various Equipment, and each Equipment can be assigned to multiple Expeditions.
- Each Expedition is associated with a specific Base Camp, where participants can rest and find essential facilities.
- An Expedition aims to reach a particular Summit, and each Summit can be targeted by multiple Expeditions.
- Summit Seekers collaborates with a single Management Company (Alpine Logistics) for expedition logistics. This signifies a one-to-many relationship between Management Company and Expedition.
- Summit Seekers also want to include the ability to track when a mountaineer successfully summits a particular peak as part of a specific expedition. They also want to record the date on which the summit was achieved.

Ensure that you clearly label the entities, attributes, and relationships in your ERD. Use appropriate notation to represent cardinality and connectivity.

Based on the case study, here are the participation constraints for the entities in the ERD:

- Every Mountaineer must participate in at least one Expedition
- An Expedition must have at least one Mountaineer participating
- An Expedition must have exactly one Expedition Leader
- A Guide may or may not be assigned to an Expedition
- Equipment may or may not be assigned to an Expedition
- Each Expedition must have exactly one Base Camp assigned
- Each Expedition must target at least one Summit
- Each Expedition must have exactly one Expedition Leader assigned
- Summit Seekers must have at least one Management Company assigned

These participation constraints specify the minimum and mandatory participation requirements for each entity in the ERD. They define the relationships and ensure that the necessary entities are associated correctly to maintain the integrity and functionality of the mountaineering system. It's important to note that these constraints are based on the information provided in the case study, and they may be subject to further refinement based on any assumptions you make. Make sure to clearly state any assumptions.


Name
Level
Specialities
ID
Guide
Assigned
Start Date
End Date
Level
Expedition
ID
Location
Led by
ID
Name
Experience
Leader
Contact
Level
Organizes
Contact
Company
Location
ID
Name
Collaborate
Summit Seekers
ID
Contact
Participates
Aims to Reach
Associated to
Base Camp
Facilities
Location
ID
Nationality
ID
Mountaineer
Age
Gender
Name
Contact
Requires
Type
Name
ID
Equipment
Availability
Record Achievement
Summit Date
Mountain Name
Summit
Elevation
Route
ID

National University of Computer and Emerging Sciences

FAST School of Computing **Spring-2023** **Islamabad Campus**

Part(b) [10 Marks] Write SQL Queries for the following scenarios.

1. List the expeditions that have the same summit target as the expedition with Expedition ID 123, along with their start dates and difficulty levels. [5 Marks]

Assuming Relationship table

Target

ExpeditionID	SummitID

```
Select ExpeditionID, StartDate, level, SummitID from
Expedition inner join Target on Expedition.ExpeditionID
= Summit.SummitID where ExpeditionID in (1,2,3);
```

1

1. Muhammad Ali Sadpara, a renowned Pakistani mountaineer, successfully summited K2, the second highest mountain in the world with an elevation of 8,611 meters. Muhammad Ali Sadpara achieved this feat as part of an expedition named "K2 Winter Ascent 2021" on the date '2021-01-16'. Retrieve the names of all female Pakistani mountaineers who have successfully summitted an eight thousander peak (a mountain whose height is more than 8000 meters). You must retrieve the information about the expedition, the summit date, the mountain name and the name of the mountaineer. You may include any other relevant information about the summit. [5 Marks]

```
Select Mountaineer.Name from
RecordAchievement inner join Mountaineer on
Mountaineer.ID = RecordAchievement.MID
where Mountaineer.Gender = 'Female' and
```

1