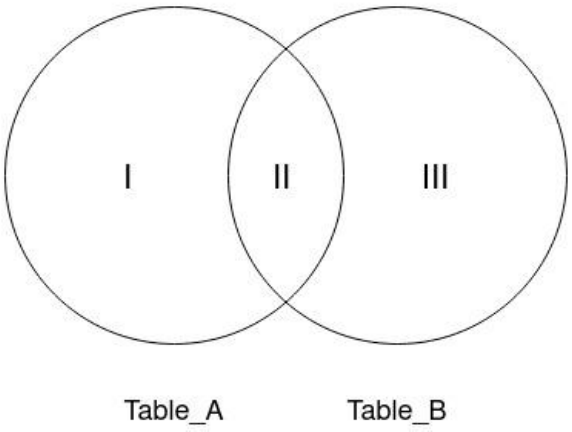
Section 1

Consider the following venn diagram for two tables which can be considered as a visual representation of the result of join of the tables:



For each of the following SQL statements, mention all the regions (among I, II and III) that will be included as a part of the result of the statement and **justify** your answer (Answers without justification will be given 0 marks): (1 + 2 + 2)

1. `SELECT <select_list> FROM Table_A A INNER JOIN Table_B B ON A.Key = B.Key;`

   Region: II

2. `SELECT <select_list> FROM Table_A A FULL OUTER JOIN Table_B B ON A.Key = B.Key WHERE A.Key IS NULL OR B.Key IS NULL;`

   Region: I, III

3. `SELECT <select_list> FROM Table_A A RIGHT JOIN Table_B B ON A.Key = B.Key WHERE A.Key IS NULL;`

   Region: III

Section 2

Consider the following scenarios. For each of the scenarios, mention if any informal database design principle is being violated. If being violated, suggest a way to remedy the violation. (3 + 3 + 3)

1. The following schema is used in a database to store student details:

   STUDENT (StudentID, Name, DOB, CGPA, Major, AdvisorID)

   However, it has been observed that only 10% of the students actually have advisors. All the other students are single degree students who don't have any advisors.

   Lot of NULL values in the table, We can make an additional table R2(Student_ID, AdvisorID) which will contain only those students who have advisor and original table can be R(StudentID, Name, DOB, CGPA, Major)

2. The following schema is used to store the details of movie in a database:

   Movie (MovieID, Title, DirectorName, DirectorContact, Year, Genre, Rating)

   The tuple where MovieID was '1MN2F' was deleted because of some legal conflict. It happens to be the only tuple which consisted of the Director 'Brad Davis'. Now, the information about 'Brad Davis' is also lost.

   Deletion Anomaly, Should store director details in separate table

3. Consider the following schema of transaction table:

   TRANSACTION (OrderID, TransactionID, Amount, Discount, TransactionDate, PaymentMethod)

   It is decomposed into two tables as follows:

   TRANSACTION_1 (OrderID, TransactionID, TransactionDate)
   TRANSACTION_2 (TransactionDate, Amount, Discount, PaymentMethod)

Whenever information about the payment method of a transaction is needed, a JOIN is performed between the two tables TRANSACTION_1 & TRANSACTION_2.

Spurious tuple generation
Should use Transaction ID in second table as joining attribute

Section 3

Consider the following database schema for a flight management system.

- **Flight** (`fid: varchar, origin_airport_id: int, destination_airport_id: int, departure: datetime, arrival: datetime, capacity: int, price: float`)
  - Primary key: `fid`
  - Foreign Keys: `origin_airport_id` and `destination_airport_id` reference Airport's `aid`
- **Airport** (`aid: int, name: varchar, city: varchar, country: varchar`)
  - Primary key: `aid`
- **Passenger** (`pid: int, name: varchar, passport: varchar, dob: date, gender: char`)
  - Primary key: `pid`
- **Ticket** (`tid: int, passenger_id: int, flight_id: varchar, seat: varchar, status: varchar, date: date`)
  - Primary key: `tid`
  - Foreign keys: `passenger_id` references Passenger's `pid`, `flight_id` references Flight's `fid`
- **Crew** (`cid: int, name: varchar, role: varchar, salary: float, flight_id: varchar`)
  - Primary key: `cid`
  - Foreign key: `flight_id` references Flight's `fid`

1. Write a SQL query to find the name and role of the crew members who are working on the flight with the highest capacity. (2)

```
SELECT C.name, C.role FROM (Crew C INNER JOIN Flight F ON C.flight_id
= F.fid) WHERE F.capacity = (SELECT MAX(capacity) FROM Flight);
```

2. Write a SQL query to find the total number of tickets sold for each flight (Display flight id and total number of tickets sold for that flight). (2)

```
SELECT F.fid, COUNT(*) AS tickets_sold FROM (Flight F LEFT JOIN
Ticket T ON F.fid = T.flight_id) GROUP BY F.fid;
```

3. Write a SQL query to find the id and departure time of the flights that depart later than any flight from the city 'Hyderabad'. (3)

```
SELECT F.fid, F.departure FROM Flight F WHERE F.departure > (SELECT
MAX(departure) FROM (Flight INNER JOIN Airport ON
origin_airport_id=aid) WHERE city='Hyderabad');
```

OR (depending upon the interpretation of 'any')

```
SELECT F.fid, F.departure FROM Flight F WHERE F.departure > (SELECT
MIN(departure) FROM (Flight INNER JOIN Airport ON
origin_airport_id=aid) WHERE city='Hyderabad');
```

4. Write a SQL query to find the count of each gender who have booked a ticket for a flight that has has an id starting with 'IN'. (3)

```
SELECT P.gender, COUNT(P.pid) AS gender_count FROM ((Passenger P
INNER JOIN Ticket T ON P.pid = T.passenger_id) INNER JOIN Flight F ON
T.flight_id = F.fid) WHERE F.fid LIKE 'IN%' GROUP BY P.gender;
```

Section 4

1. Let R(A, B, C, D, E, F) be a relational scheme with the following functional dependencies:

- C → F
- E → A
- EC → D
- A → B

Prove or disprove: EC is a key for R. (2)

Proof:

EC is a key for R. Closure of EC = {ECF} (C → F) = {ECFA} (E → A) = {ECFAD} (EC → D) = {ECFADB} (A → B).

All the attributes are functionally dependent on EC. Hence, EC is a key for R.

2. Let R(A, B, C, D, E) be a relational scheme with the following functional dependencies:

- A → B
- A → C
- CD → E
- B → D
- E → A

State True or False for the following statements with explanation (Answers without any explanation with receive 0 marks): (2 + 2)

- CD → AC is implied from the above set of FDs.
- BD → CD is implied from the above set of FDs.

- CD → AC is true as CD → E and E → A
- BD → CD is not true as it cannot be derived from the given set of FDs.

CS4.301: Data and Applications (Monsoon 2023)
Quiz - 2

Max Marks: 30 Time: 45 mins

Roll No:
Branch:

General Instructions:

- This question paper consists of 4 sections. All sections are **compulsory**.
- There are **specific instructions** for each section. Please read those instructions carefully before answering the questions.
- **Clearly specify the attributes, entities that you are using in your queries, any ambiguity may lead to deduction of marks.**
- Papers without roll numbers will be **awarded 0 marks** for this quiz.

Section - A

Instructions:

1. You have to write SQL statements for the given 4 queries.
2. **No partial marks** will be given in this section. (i.e. either full for the part or 0).

Consider the schema following two tables:

Order

Column Name	**Data Type**	**Constraints**
OrderID	INT	Primary Key
CustomerID	INT	
OrderDate	Date	
Price	INT	

Customer

Column Name	Data Type	Constraints
CustomerID	INT	Primary Key
CustomerName	Varchar(30)	

1. Set the default value of the **Price** column in the **Order** table as 10. (1)

2. Make the CustomerID column in the **Order** table a foreign key which references the CustomerID column in the **Customer** table. (2)

3. Change the data type of **Price** column in **Order** table so that decimal numbers can be inserted with a precision of 5 decimal places. (2)

4. Add a positive value constraint on the **Price** column in the **Order** table. (i.e. the price should always be > 0). (2)

Section - B

Instructions:

1. You have to write SQL statements for the given 3 queries.
2. Along with the SQL statement, you have to provide an **explanation** of your query.
3. It is compulsory to use the **JOIN** clause. Answers **without JOIN clause will be given 0 marks.**
4. If your query is correct you will receive full (4 marks) for that part. However, if your query is incorrect, then **you can get up to 2 marks for correct explanation/thought process.**

Consider the following database for an e-commerce platform:

- Customers (`customer_id, customer_name, email_id, address`);
- Products (`product_id, product_name, cat_id, price, stock_quantity`);
- Categories (`category_id, category_name`);
- Orders (`order_id, c_id, order_date, total_amount`);
- OrderDetails (`order_detail_id, o_id, p_id, quantity, subtotal`);

Referential Integrity Constraints:

- The cat_id column in Products refers to the category_id in Categories.
- The c_id column in Orders refers to the customer_id in Customers.
- The o_id column in OrderDetails refers to the order_id in Orders.
- The p_id in OrderDetails refers to the product_id in Products.

1. Write a query to retrieve the total number of orders placed by each customer (the query should display customer_id, customer_name and number of orders placed for each customer). (4)

2. Write a query to find the top 3 categories with the highest total sales (sum of subtotal from OrderDetails) and display the category names and total sales. (4)

3. Write a query to find the top 5 best-selling products by the total quantity sold and show their ids, names and total quantity sold. (4)

Section -C

Instructions:

1. You have to write the number of tuples expected in the result of the given relational algebra expression.
2. Along with the answer, you have to provide an **explanation**.
3. Answers without an explanation will be **awarded 0 marks**.

Consider the following relations A, B and C.

A		
ID	Name	Age
12	Arun	60
15	Shreya	24
99	Rohit	11

B		
ID	Name	Age
15	Shreya	24
25	Hari	40
98	Rohit	20
99	Rohit	11

C		
ID	Name	Age
25	Hari	40
95	Aryan	18

1. Write the number of tuples expected and justify your answer. (2)

$$\pi_{name}(A \bowtie_{A.ID=B.ID} B) \cap \pi_{name}(B \bowtie_{B.ID=C.ID} C)$$

2. Write the number of tuples expected and justify your answer. (2)

$$\pi_{name}(\sigma_{A.age<=50}(A \bowtie_{A.ID=B.ID} B)) \cup \pi_{name}(\sigma_{C.age>20}(B \bowtie_{B.ID=C.ID} C))$$

Section - D

Instructions:

1. There are **2 subjective questions** relating to database design concepts.
2. Please write **concise and correct** answers.

1. Consider the following schema:

   Customer_Order_Details (`CustomerID, Name, Address, Phone, Email, OrderID, Product, Quantity, Price`)

   What kind of modification anomaly can occur in this schema if we want to change the price of a product that has been ordered by multiple customers? Explain your answer with an example. (4)

2. Consider the following table:

StudentID	Name	Course	Grade
101	Alice	CS101	A
102	Bob	CS101	B
103	Carol	CS102	A
104	Dave	CS102	C
105	Eve	CS103	B
106	Frank	CS103	A

Which of the following functional dependencies may hold given the above state of the database? For the FDs that dont hold, give one counter example that shows its violation. (3)

- StudentID -> Name
- Name -> StudentID
- Course -> Grade
- Grade -> Course
- StudentID -> Course
- Course -> StudentID

CS4.301: Data and Applications (Monsoon ‘23)

End-Semester Examination Rubric

Maximum Marks: 50 Time: 2 hours

Instructions:

- This question paper consists of 7 questions. All questions are **compulsory**.
- If any question is ambiguous, state your assumptions clearly and proceed to answer. No clarifications will be provided during the exam.
- Clearly specify the attributes, entities that you are using in your queries, any ambiguity may lead to deduction of marks.

Q1) Consider the following populated tables which are a part of university management system:

Major

Name	Department
Computer Science	Engineering
Mathematics	Science
Physics	Science
Biology	Science
Chemistry	Science

Student

ID	Name	Age	GPA	Major
101	Alice	19	3.5	Computer Science
102	Bob	20	3.2	Mathematics
103	Charlie	21	3.8	Physics
104	David	18	2.9	Biology

Constraints on the database:

- Major Table:
  - Name: Primary key
  - Department: NOT NULL

- Student table:
  - ID: Primary key
  - Name: NOT NULL
  - Age: Should be > 0
  - GPA: Should be between 0 and 4
  - Major: References Major.Name

Mention **all** the database constraints that are violated when the following statements are executed (Treat each statement as independent): [2 + 2 + 2]

- `INSERT INTO Student VALUES (104, 'Eve', 22, 4.1, 'Chemistry');`
- `UPDATE Student SET Major = 'History' WHERE ID = 103;`
- `DELETE FROM Major WHERE Name = 'Physics';`

Answer:

a) 104 already exists in the Student table (Key constraint violation), GPA = 4.1 cannot be inserted (domain constraint violation) (**1 mark for each violation, Entity Integrity constraint not the same as Key constraint, cut 1 mark if entity integrity instead of key constraint**)

b) Cannot set Major = History because there is no major name called 'History' in Major table (Referential integrity violation) (**2 marks**)

c) Student with ID 103 will not have any corresponding major name in Major table after deleting 'Physics' so action is restricted (referential integrity violation) (**2 marks**)

Q2) Consider the following database schema for a restaurant management system:

- **Customers:** (`c_id (Primary Key), customer_name, phone_number, email`)
- **Restaurants:** (`r_id (Primary Key), restaurant_name, location, cuisine_type`)
- **MenuItems:** (`item_id (Primary Key), item_name, price, restaurant_id (Foreign Key references Restaurants)`)

- **Orders:** (order_id (Primary Key), customer_id (Foreign Key references Customers.c_id), restaurant_id (Foreign Key references Restaurants.r_id), order_date, amount)
- **Chefs:** (chef_id (Primary Key), chef_name, specialty, years_of_experience, restaurant_id (Foreign Key references Restaurants.r_id))

Assumptions:

- The order_date column in Orders table has datatype DATE in the format (YYYY-MM-DD)
- The total_amount column in the Orders table stores the amount in Rupees.

Write SQL statements for the following 5 queries. You can also provide an explanation for your statement which will be evaluated in case your query is incorrect (**there is no partial marking for SQL statement**). You can receive up to 1.5 marks for correct explanation.

Give 3 marks for correct query, no partial marking for query
Give upto 1.5 marks for correct explanation, Give full for LEFT JOIN as well

a) Find the total revenue generated(which is the sum of the amount of all orders placed) by each restaurant in the October month of the current year. Display the results(restaurant name, restaurant location and revenue) in descending order of revenue. [3]

Ans:

```
SELECT r.restaurant_name, r.location, SUM(o.amount) as Revenue
FROM (Restaurants r INNER JOIN Orders o ON r.r_id =
o.restaurant_id) WHERE o.order_date >= '2023-10-01' AND
o.order_date <= '2023-10-31' GROUP BY r.r_id, r.restaurant_name
ORDER BY Revenue DESC;
```

b) Find the customer who has spent the most money on orders. Display the customer's name and the total amount spent. [3]

Ans:

```
SELECT c.customer_name, SUM(o.amount) as total_amount_spent FROM
(Customers c INNER JOIN Orders o ON c.c_id = o.customer_id) GROUP
```

```
BY c.c_id, c.customer_name ORDER BY total_amount_spent DESC LIMIT
1;
```

c) Find the top 3 most frequent customers(based on the number of orders placed) of the restaurant having r_id = 90881. Display the customer's name and number of orders. [3]

Ans:

```
SELECT c.customer_name, COUNT(*) as num_orders_placed FROM
(Customers c INNER JOIN Orders o ON c.c_id = o.customer_id) WHERE
o.restaurant_id = 90881 GROUP BY c.c_id, c.customer_name ORDER BY
num_orders_placed DESC LIMIT 3;
```

d) List the names of restaurants where the average order amount is higher than the overall average order amount. [3]

Ans:

```
SELECT restaurant_name from Restaurants WHERE r_id IN (SELECT
restaurant_id FROM Orders GROUP BY restaurant_id HAVING
AVG(amount) > (SELECT AVG(amount) FROM Orders));
```

e) Find the names of restaurants that have chefs with an average experience of more than 7 years. [3]

Ans:

```
SELECT r.restaurant_name FROM (Restaurants r JOIN Chefs c ON
r.r_id = c.restaurant_id) GROUP BY r.r_id, r.restaurant_name
HAVING AVG(c.years_of_experience ) > 7;
```

Q3) Consider the following three tables:

Students:

student_id	student_name	major	age
1	Alice	Computer Science	21
2	Bob	Physics	22
3	Charlie	Physics	20

4	David	Mathematics	23
5	Emily	Computer Science	22

Courses:

course_id	course_name	instructor
101	Database Design	Dr. Smith
102	Physics I	Prof. Johnson
103	Calculus II	Dr. Davis
104	Genetics	Prof. White
105	Organic Chem	Dr. Brown

Registrations:

registration_id	student_id	course_id	grade
1	1	101	A
2	2	102	B
3	3	103	C
4	4	104	B
5	5	105	A

Show the output of the following queries (You have to write the entire output that you would expect when the query is executed): [2.5 + 2.5]

a)

```
SELECT students.student_id, student_name, course_name, grade FROM
students JOIN registrations ON students.student_id =
registrations.student_id JOIN courses ON registrations.course_id =
courses.course_id WHERE grade = 'A' AND major = 'Computer
Science';
```

Ans:

student_id	student_name	course_name	grade
1	Alice	Database Design	A

5	Emily	Organic Chem	A

1.25 marks for each correct row, no partial marks for a row

b)

```
SELECT students.major, AVG(students.age) as average_age FROM
students
RIGHT OUTER JOIN registrations ON students.student_id =
registrations.student_id
GROUP BY students.major ORDER BY average_age;
```

Ans:

Major	average_age
Physics	21
Computer Science	21.5
Mathematics	23

1 mark each for the first 2 rows, 0.5 for the last row, no partial marking for a row, Note that order of the rows also matters since ORDER BY is used. If the order is wrong, give 0.

Q4) Consider a relational scheme R(A, B, C, D, E) with the following FDs:

- A → DE
- B → A
- D → C

a) Identify the primary key for R. (Show how it is a primary key) [2]

b) What do you understand by the second normal form? Is the above relation in 2NF? If yes, justify your answer. If not, decompose it so that it is in 2NF after decomposition. [2 + 2]

c) What do you understand by the third normal form? Is the above decomposed relation (the answer to Q2) in 3NF? If yes, justify your answer. If not, decompose it so that it is in 3NF after decomposition. [2 + 2]

Answers:

a) Primary key for R is B.
B-> A => B →ADE (A -> DE) => B → ACDE (D → C)
Hence, B determines all the attributes of R. Therefore, its a key for R.

Identifying primary key - 0.5
Showing how it is a primary key - 1.5

If some other key is selected, part b and c's decomposition will be incorrect, give 0

b) A relation schema R is in 2NF if every non prime attribute A in R is fully functionally dependent on the primary key of R. (**2 marks**)
The above relation is in 2NF as no attribute is dependent on a part of primary key**(Primary key is only B)** (**2 marks: 1 mark for explanation, 1 mark for answer**)

Note: There's only one candidate key (which is B) and therefore answers claiming that the relation is not in 2NF are wrong.

c) A relation schema R is in 3NF if it satisfies 2NF and no non prime attribute of R is transitively dependent on the primary key. (**2 marks**)
The above decomposed relation is not in 3NF since D is transitively dependent on **B(primary key)**. (**1 mark**)
3NF Decomposition: R1(A, D, E), R2(D, C), R3(B, A) (**1 mark**)

Q5) Given a relation R(P, Q, R, S, T, U, V, W, X, Y) and Functional Dependency set FD = { PQ → R, PS → VW, QS → TU, P → X, W → Y }, determine whether the given R is in 2NF (Note that the key for R is PQS). If not, convert it into 2 NF. [5]

Answer:
Not in 2NF since R is dependent on a part of the primary key (any other example also works). (**2 marks**)
2NF Decomposed schema:
R1(P, Q, R)
R2(P, S, V, W)
R3(Q, S, T, U)
R4(P, X)
R5(W, Y)
R6(P, Q, S)
(**3 marks for all 6 tables otherwise deduct accordingly)**

OR

R1(P,S,V,W,Y)

R2(Q,S,T,U)
R3(P,X)
R4(P,Q,R)

Q6) Consider the following definition of dependency preserving decomposition:

If we decompose a relation R into relations R1 and R2, all dependencies of R must be part of ***either R1 or R2 or must be derivable from combination of functional dependencies(FD) of R1 and R2***

a) Mention any two advantages & two disadvantages of decomposing a table into smaller tables in a database. [2]

b) Consider a relation scheme R(A, B, C, D) with the following functional dependencies:
A → B, B → C, C → D
Mention whether the following decompositions are dependency preserving and justify your answer: [2 + 2]

- R1(A, B, D), R2(B,)
- R1(A, B), R2(B, C), R3(C, D)

Answers:

0.5 marks for each point
Two advantages: Maintaining data integrity, avoiding anomalies etc.
Two disadvantages: performance overhead, query complexity etc.

a) R1(A, B, D) & R2(B, C) **is not** dependency preserving as C → D cannot be derived from the set of individual dependencies of R1 & R2. (**2 marks**)

b) R1(A, B), R2(B, C), R3(C, D) **is** dependency preserving as all the original FDs can be derived from the set of individual dependencies of R1, R2 & R3. (**2 marks**)

Q7) Suppose there is a ternary relationship called **Teaches** between three entity types: Professor, Course, and Student. The relationship indicates which professor teaches which course to which student. Assume the following constraints:

- A course has to be taught by at least 1 professor and at most 2 professors and should be attended by at least 1 student and at most 100 students.

- A professor can teach no course or at most 3 courses.
- A student can attend no course or at most 6 courses.

Write the (min, max) participation constraint for each of the participating entity types and justify your answer (Please note that you have to write exact numbers instead of using N in your answer). [3]

Answer:
Course => (1, 200)
Prof => (0, 300)
Student => (0, 12)

(**1 mark for each correct (min, max) ratio**)