Indian Statistical Institute

M.Tech. (CS), Mid-Semester of Second Semester Examination, 2024-25

Database Management Systems

Full Marks: 30 Date: 25-02-2025 Time: 2 Hours

Answer any *three* of the following questions

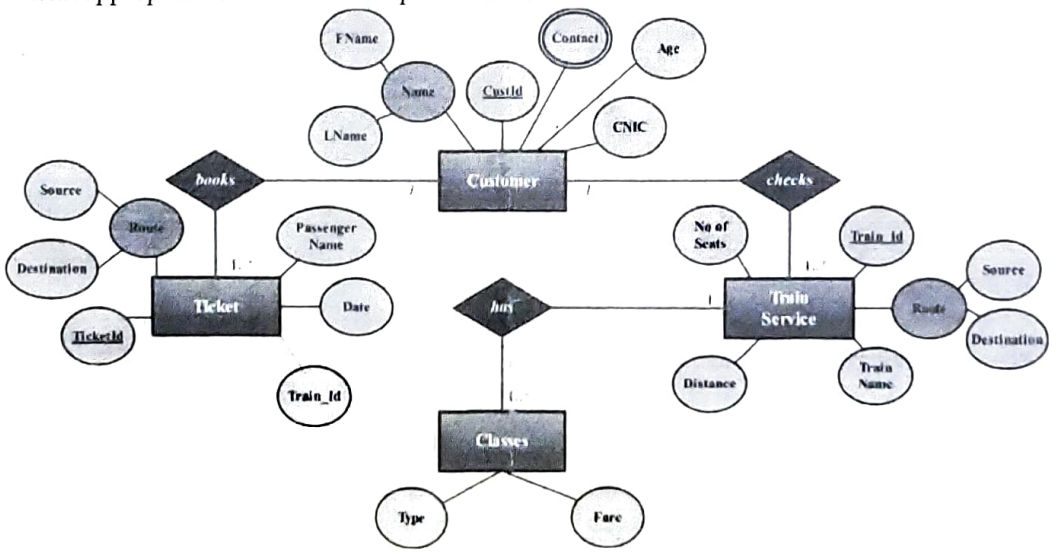
$3 \times 10 = 30$

1. (a) Consider a relation $R = < X, Y, Z >$ having t tuples such that $\text{dom}(X) \subset \text{dom}(Y) \subset \text{dom}(Z)$, where $\text{dom}(A)$ represents the domain of values for the attribute A. What will be the minimum and maximum number of tuples returned by results of the following relational algebra expressions.
   (i) $\pi_{X, Y, Z}(R)$
   (ii) $\pi_{X, Y}(R) \bowtie \pi_{Y, Z}(R)$
   (iii) $R \div \pi_{X, Z}(R)$

   (b) Prove or disprove the statement: "If the natural join and full outer join between a pair of relations R_1 and R_2 results in the same relation, then their semijoin operation is ~~associative~~ commutative, i.e., $R_1 \ltimes R_2 = R_2 \ltimes R_1$". Recall that semijoin is alike natural join with the only exception that attributes in the first relation are returned in the result.

(2+2+2)+4

2. Consider the following Entity Relationship (ER) diagram that depicts a railway reservation system. The customers can avail the train service by checking the train and route details and through booking of tickets. Revise this ER diagram by adding specialization of tickets as "General" and "Tatkal" therein. Finally, convert the revised ER diagram into a set of tables with appropriate constraints incorporated therein.



3+7

3. (a) Consider the following table denoting some employees of a company. Suggest an indexing strategy for the given table. Justify your answer with respect to the different access methods that are applicable on this table.

Name	Age	Salary	Ethnicity	Gender
Suraj	25	40k	Asian	Male
Scott	31	60k	European	Male
Bob	45	70k	American	Male
Shobha	23	110k	Asian	Female
Irwin	28	60k	European	Male
Madhu	30	40k	Asian	Male
Britney	35	60k	European	Female
Chris	38	70k	European	Male

(b) Suppose a relation $R = <A, B, C, D>$ is indexed using a hash table. Justify whether the following operations will be efficient on the relation or not.

(i) $\sigma_{(A > 99) \wedge (A < 999)} (R)$

(ii) $\sigma_{D != 1000} (R)$

6+(2+2)

4. Let there be a relation $R = <U, V, W, X, Y, Z>$ having all the attributes as atomic and holding the functional dependencies FD = $\{X \rightarrow YZ, Y \rightarrow W, U \rightarrow V\}$. Answer all the following questions with proper justifications.

(a) What is the candidate key of R?

(b) What is the highest normal form that R satisfies?

(c) Decompose R such that every decomposed subrelation of R satisfies the Boyce-Codd normal form.

(d) Is your decomposition lossless?

(e) Is your decomposition dependency preserving?

1+1+4+2+2

Indian Statistical Institute

M.Tech. (CS), Final of Second Semester Examination, 2024-25

Database Management Systems

Full Marks: 50 Date: 30-04-2025 Time: 2.5 Hours

Answer any *five* of the following questions

$5 \times 10 = 50$

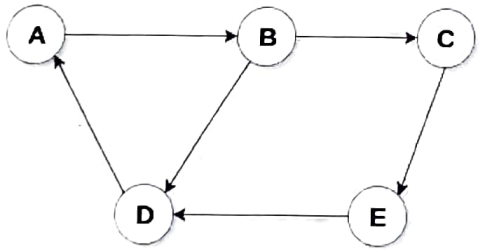
1. (a) Suppose two publishers maintain their databases, say P1 and P2, following the schema given below.

P1TABLE1 = <*id* : *string*, *title* : *string*, *cost* : *integer*, *authors* : *string*, *Edition* : *integer*>

P2TABLE1 = <*ISBN* : *integer*, *title* : *string*, *cost* : *real*, *authors* : *string*>
P2TABLE2 = <*ISBN* : *integer*, *Edition* : *integer*, *Date* : *date*>

The tables under each database P*i* is separately numbered as P*i*TABLE*j*. Note that ISBN denotes the International Standard Book Number. Suggest a method of integrating both the databases into a single database.

(b) A directed graph can be saved in a relation in terms of its adjacency list denoted by the attributes <*source-vertex*, *target-vertex*>. For example, the following graph **G** (shown in left) can be represented as the relation **R** (shown in right).



G

source-vertex	target-vertex
A	B
B	C
B	D
C	E
E	D
D	A

R

A triangle is defined as a directed cycle of length 3. Note that in the graph **G**, (A, B, D) is a triangle, whereas (A, B, C, E, D) is a cycle but not a triangle. Write a relational algebra expression to return all the triangles in any arbitrary **G** from its respective relation **R**.

(c) Can a discriminating attribute of a weak entity set be a foreign key in the same entity set? Justify your answer.

4+4+2

2. Let there be a relational schema R = <U, V, W, X, Y, Z> holding the set of functional dependencies FD = {U → V, VW → X, X → YZ, X → Z}.

(i) Find out the closure of the attribute set UW.

(ii) Derive all other non-trivial functional dependencies that can be obtained from FD.
(iii) Decompose R into any arbitrary subrelations that are lossless-join.
(iv) Decompose R into any arbitrary subrelations that are dependency preserving.

2+2+2+4

3. (a) Consider the relations $R1$ = <A, B, C> and $R2$ = <B, D, E> having 5040 and 24 tuples, respectively. Notably the primary keys in the respective relations are underlined therein. The number of distinct values of B in $R1$ is 7. The range of values of C and D in $R1$ and $R2$ are [50, 120] and [0, 60], respectively. Estimate the size of the following query considering the attribute values are uniformly distributed.

$$\sigma_{(C \leq 100) \wedge (B!=0)}\,(R1) \bowtie \sigma_{(D \geq 100)}\,(R2)$$

(b) Prove the following equivalence relation of relational algebra.

$$\sigma_\theta\,(R_1 - R_2) \equiv \sigma_\theta\,(R_1) - R_2$$

(c) Cite an example where both the sort merge join and hash join are equally good physical plans for performing natural join.

5+3+2

4. (a) Let there be a relation with data items A and B. Derive the *precedence graph* from the following schedule of transactions working on the said relation. Conclude whether the given schedule is *conflict serializable* or not. Justify whether the given schedule is *view serializable* or not.

Transaction T_1	**Transaction T_2**	**Transaction T_3**	**Transaction T_4**
			read(A)
	read(A)		
	A ← A + 12		
		read(A)	
		A ← A * 0.1	
write(B)			
	write(A)		
		read(B)	
	write(B)		
	Commit		

(b) Show the steps to identify the serializability order from the precedence graph that you derived from the schedule given in (a).

(5+2)+3

5. Identify whether any problem will occur in executing each of the following schedules. If so, justify your answer.

(i)

T_1	T_2	T_3	**Timestamp**
lock-X(A)			1
read(A)			2
A = A + 5			3
write(A)			4
	lock-S(B)		5
	read(B)		6
	lock-S(A)		7

lock-X(B)			8
lock-S(B)			9
read(B)			10
B = B – 15			11
		lock-X(B)	12
write(B)			13
unlock(B)			14
		unlock(B)	15
	unlock(B)		16

(ii)

T_1	T_2	T_3	**Timestamp**
lock-X(B)			1
	lock-X(A)		2
	read(A)		3
	A = A + A*0.2		4
	write(A)		5
	unlock(A)		6
lock-X(A)			7
read(A)			8
A = A – 100			9
write(A)			10
unlock(A)			11
		lock-X(A)	12
		read(A)	13
		A = A + 10	14
		write(A)	15
		unlock(A)	

(iii)

T_1	T_2	**Timestamp**
lock-S(B)		1
read(B)		2
B = B + 10		3
	lock-S(B)	4
	read(B)	5
	lock-S(A)	6
	read(A)	7
	A = A * 0.05	8
	write(A)	9
	unlock(A)	10
	unlock(B)	11
upgrade(B)		12
write(B)		13

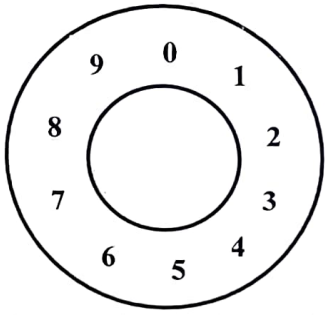
6+3+1

6. (a) Convert the following collection of JSON documents into minimum number of tables.

```
{
        "Roll" : 1,
        "Name" : "Shyam",
        "Centre" : "Kolkata",
        "DOB" : 30-04-2025,
}
{
        "Roll" : 1,
        "Name" : "Somaiya",
        "Centre" : "Bangalore",
        "Rank" : 5,
        "Grade" : 8.5,
}
```

```
{
        "Roll" : 1,
        "Name" : "Malik",
        "Centre" : "Delhi",
        "Rank" : 5,
        "Grade" : 8.5,
}
{
        "Roll" : 1,
        "Name" : "Stefan",
        "Centre" : "Chennai",
        "Rank" : 5,
}
```

(b) Let there be a distributed environment with 7 objects (labeled as 1 to 7) assigned to 4 different nodes (node 1, 2, 3 and 4) implementing consistent hashing. Let the objects are assigned positions on the hash table (as shown below) using the hash function $L\%10$, where the object label is L. On the other side, the nodes are assigned positions using the hash function $(N*2)\%10$, where the node label is N. Now if a new object (labeled as 8) gets hashed into the system after the replacement of one node (node 2 replaced with node 5), what will be the new assignments of objects to different nodes? Is this replacement better or worse in terms of load balancing?



(c) What is consistency a mandatory requirement in both SQL and NoSQL databases but not the others like atomicity, isolation and durability?

4+4+2

7. Show appropriate examples of the following things.
(i) Consistency check in data validation.
(ii) A pair of tables that will return the same result for left and right outer joins.
(iii) Discriminating attribute of a weak entity set.
(iv) A table that satisfies Sixth Normal Form (6NF).
(v) Validation-based protocol for concurrency control.

2+2+2+2+2