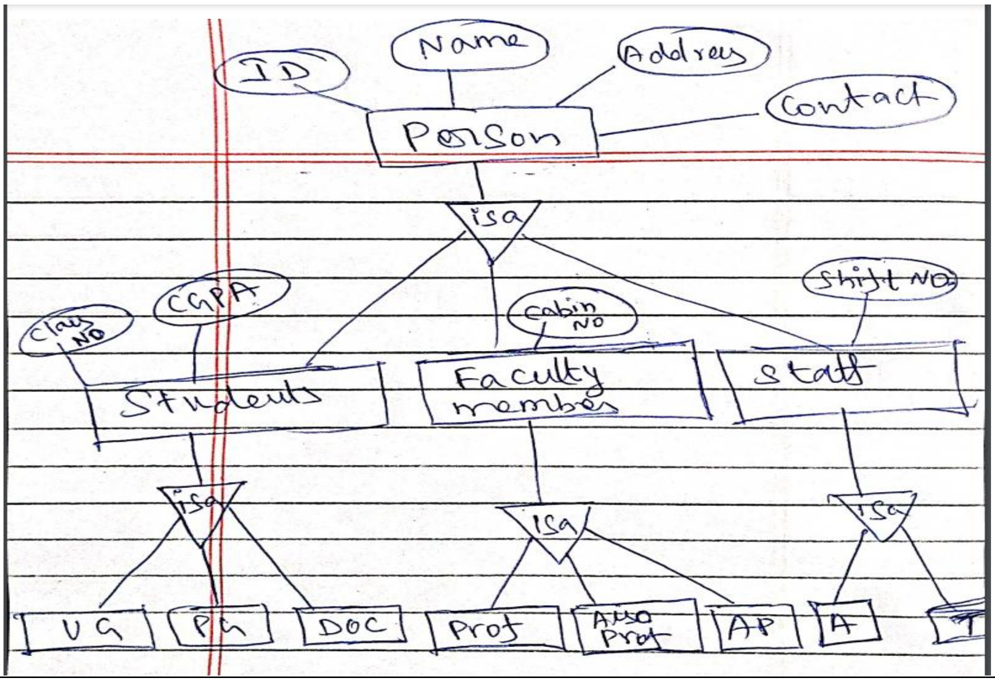
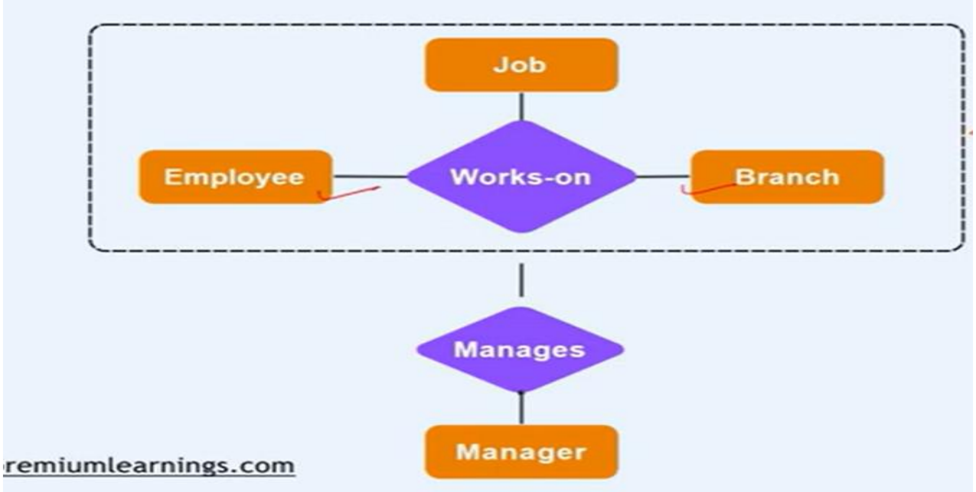
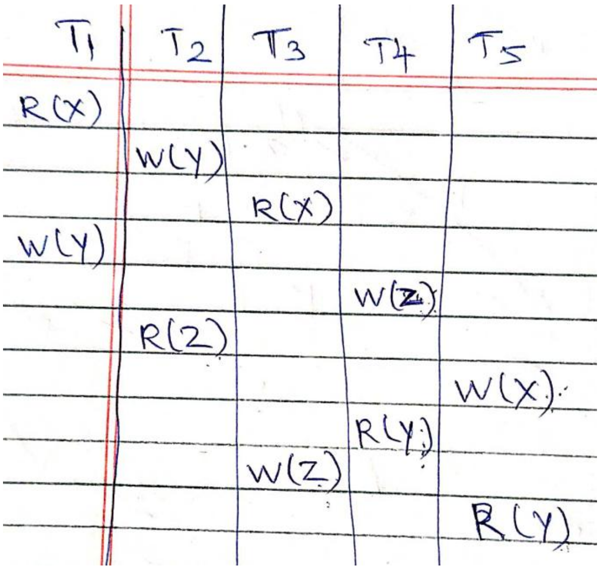
Answer Key

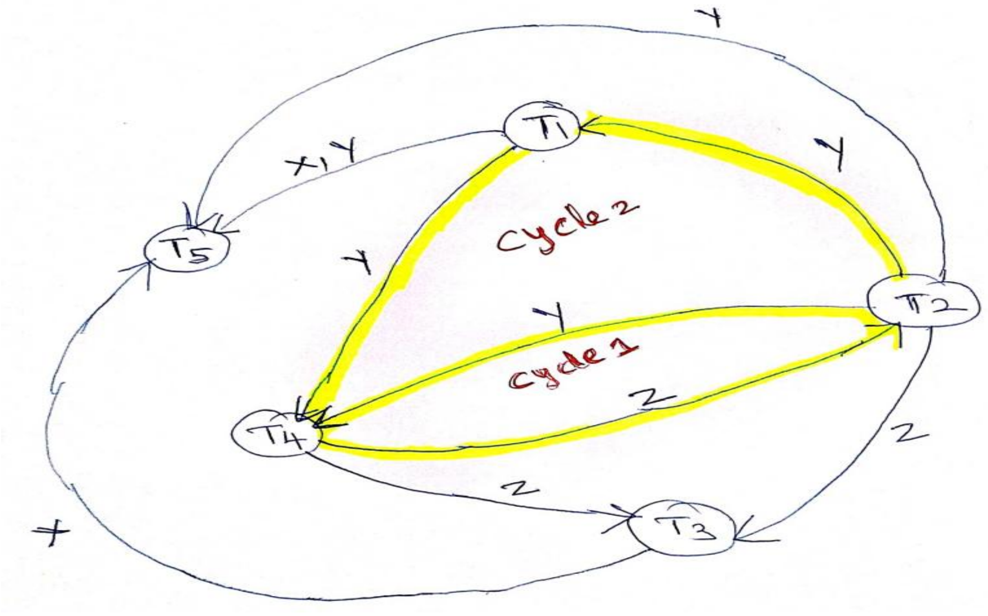
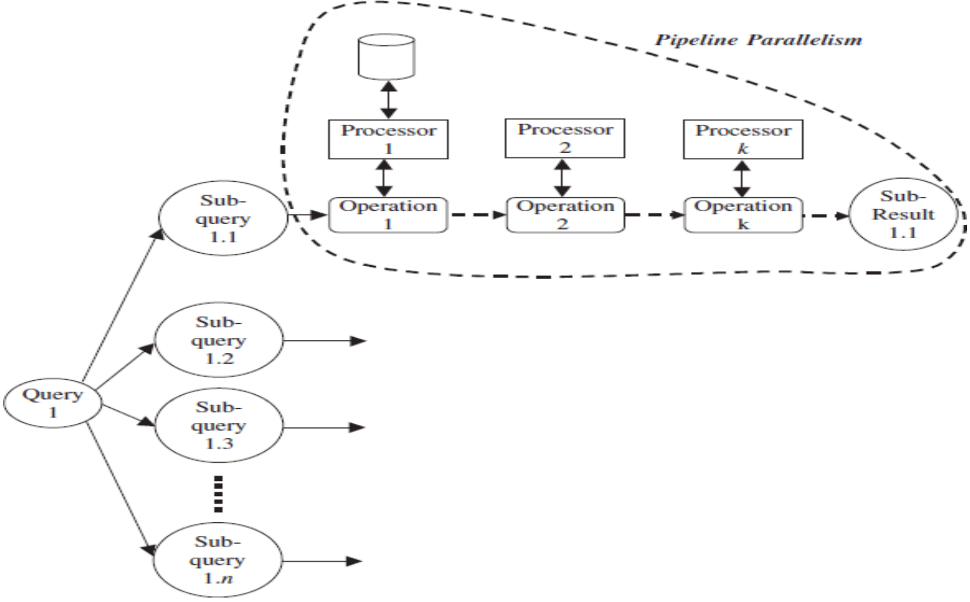
CSI2004 and Advanced Database Management Systems

WINTER SEMESTER 2024-2025

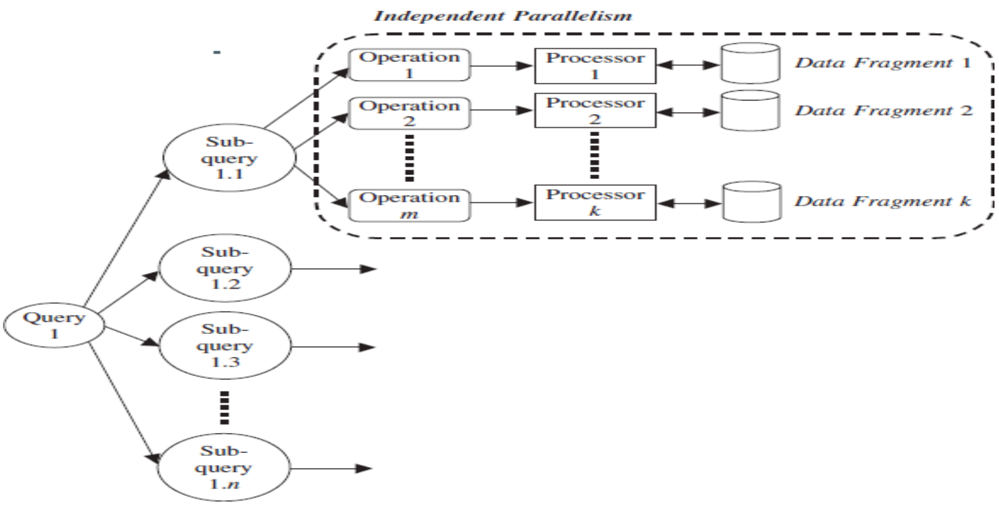
Q. No	Question	M	CO	BL
1.	a) Design a generalization–specialization hierarchy for a university database. The university has students, faculty member and staff. Students can be under_graduate, post_graduate, and doctoral. Faculty member can be professors, associate professors, and assistant professors. Staff can be administrative and technical. Justify your placement of attributes at each level of the hierarchy and explain the diagram in detail. **Answer: Here, I have added sample attributes. Students can add their own related attributes also. But, they have to follow this hierarchy.** **Diagram – 3 marks, explanation – 2 marks** 	5		
	b) Draw an Entity-Relationship diagram for an organization's job management system using an aggregation concept. The system includes employees, branches, and jobs. Employees work on specific jobs in various branches, Now, the manager has assigned a task that need to manage or supervise the above said relationship. Justify the use of aggregation to show how employees, jobs, and branches are interrelated under the supervision of a manager. **Answer: Diagram – 3 marks, Justification – 2 marks**	5		

<table>
<tr>
<td></td>
<td>


</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>
Consider the following schedule:
R1(X), W2(Y), R3(X), W1(Y), W4(Z), R2(Z), W5(X), R4(Y), W3(Z), R5(Y)
Draw the precedence graph for the given schedule with proper justifications and determine whether it is conflict serializable schedule or not?
Answer: 2 cycles are there. Cycle 1. T4-T2-T4, Cycle 2. T1-T4-T2-T1. Since cycle is there, it is not a conflict serializable.
Precedence graph with proper justification: 8 marks
Finding 2 cycles and statement of “Not a conflict serializable”: 2 marks

<table>
<tr><th>T1</th><th>T2</th><th>T3</th><th>T4</th><th>T5</th></tr>
<tr><td>R(X)</td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>W(Y)</td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td>R(X)</td><td></td><td></td></tr>
<tr><td>W(Y)</td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td>W(Z)</td><td></td></tr>
<tr><td></td><td>R(Z)</td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td>W(X)</td></tr>
<tr><td></td><td></td><td></td><td>R(Y)</td><td></td></tr>
<tr><td></td><td></td><td>W(Z)</td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td>R(Y)</td></tr>
</table>
</td>
<td>10</td>
<td></td>
<td></td>
</tr>
</table>

				
3	In a parallel database system, tasks are divided across multiple processors to optimize query execution. Especially, Interoperation parallelism involves achieving parallelism by executing multiple operations of the same query or transaction concurrently. Identify the two different forms of Interoperation parallelism to execute a single query and also draw the architecture of executing a single query for each form and explain in detail. Answer: Identification of forms: 1. Pipelined, 2. Independent parallelism – 2 marks Architecture of Pipelined with explanation – 4 marks 	10		

Architecture of Independent parallelism with explanation – 4 marks



4. Distributed database systems are preferred to enhance scalability, performance and few other factors by distributing data across multiple sites. However, their implementation introduces challenges such as data consistency, communication overhead, and complexity in query optimization. Discuss any five advantages of distributed databases in order to achieve the scalability and performance improvements. Also, discuss any five disadvantages and challenges in the management and maintenance of distributed system environment.

10

Answer:
Any 5 advantages: 5 marks

- Transparent management of distributed, fragmented, and replicated data
- Data Independence
- Network Transparency
- Replication Transparency
- Fragmentation Transparency
- Increased reliability and availability
- Improved performance
- Easier expansion (scalability)

Any 5 disadvantages: 5 marks

- Complex Software
- Increased Processing overhead
- Data integrity
- Different data formats might be used
- Deadlock

5 Star Electronics and Appliances company maintains a database of various electronic products. The data includes details such as the product_ID, product_name and the price of the product. To optimize the processing and distribution of data across a parallel database system, the company has set up 4 processors (W, X, Y & Z) and 4 disks (D0, D1, D2 & D3). The price of the product is used as the partition attribute and the company has defined a range vector v is [1000, 5000, 10000]. Apply various partitioning strategies on the given table to distribute the different records into the various disks and show the final disk status for each partitioning strategy.

	Product_ID	**Product_Name**	**Price**
Record 1	P001	Smartphone	15470
Record 2	P002	Laptop	48000
Record 3	P003	Headphones	900
Record 4	P004	TV	7350
Record 5	P005	Microwave	8755
Record 6	P006	Blender	2557
Record 7	P007	Iron box	4000
Record 8	P008	Refrigerator	9501

Answer:
Final disk status of Round robin : 3 marks
Final disk status of hash : 4 marks
Final disk status of range : 3 marks

10

1) Round Robin

Total Records = 8

Total Disks = 4

Formulae = $i \bmod n$

Record 1 mod 4 = 1

2 mod 4 = 2

3 mod 4 = 3

4 mod 4 = 0

5 mod 4 = 1

6 mod 4 = 2

7 mod 4 = 3

8 mod 4 = 0

D0	D1	D2	D3
Record 4	R1	R2	R3
Record 8	R5	R6	R7

2) Hash Partitioning

Partition Attribute = price

Formulae = h(price) = (price mod n)

R1 : h(15490 mod 4) = 2

R2 : h(48000 mod 4) = 0

R3 : h(900 mod 4) = 0

R4 : h(7350 mod 4) = 2

R5 : h(8755 mod 4) = 3

R6 : h(2557 mod 4) = 1

R7 : h(4000 mod 4) = 0

R8 : h(9501 mod 4) = 1

D0	D1	D2	D3
R2	R6	R1	R5
R3	R8	R4	
R7			

3) Range Partitioning

Range vectors V = [1000, 5000, 10000]

Disk 0 = <= 1000

Disk 1 = > 1000 and <= 5000

Disk 2 = > 5000 and <= 10,000

Disk 3 = > 10,000

D0	D1	D2	D3
R3	R6	R4	R1
	R7	R5	R2
		R8	

****