Regulation: 2022

Reg. No.:

BANNARI AMMAN INSTITUTE OF TECHNOLOGY

(An Autonomous Institution Affiliated to Anna University, Chennai)

SATHYAMANGALAM – 638 401

Stay Ahead

PERIODICAL TEST II - JUNE 2024

IV Semester

22AI405 / 22AM405 / 22CD405 / 22CS405 / 22CT405 / 22IS405 / 22IT405 & DATABASE MANAGEMENT SYSTEMS

Degree & Branch: BE / B.Tech & AD, AL, CD, CS, CT, IS, IT

Time: 1:30Hrs. Maximum: 50 Marks

Instructions:

1. Students should not mark/write anything on the Question Paperother than the register number.
2. Section B and C of the Question Paper contains questions for 30 Marks each. Section A contains questions for 15 Marks.
3. Students can attempt answering any two out of three subsections in each section. Maximum mark is limited to 20 in section B&C, and 10 in sections A.

Q.No	Questions		
	SECTION A	COURSE OUTCOME 3	MAXIMUM: 10 MARKS

A1 The online learning platform facilitates seamless engagement between students and instructors. Students register and access a diverse range of courses, featuring interactive materials for flexible learning. Communication tools enable collaboration, clarifications, and project work. Instructors assess progress and provide timely feedback, while students track their performance. Upon completion, students earn certificates for their achievements. The platform ensures accessibility and support for all users, fostering an inclusive learning environment. It serves as a comprehensive hub for educational activities, course management, and skill development, enhancing the learning experiences of both students and instructors.

(i) From the Scenario Identify the primary benefit of using the described online learning platform? Assume that the platform provides limited communication options (e.g., email only).

a) Instructors can easily track student performance.
b) Students have access to a wider variety of courses.
c) The platform offers a secure space for online communication.
d) It facilitates seamless engagement and interaction between students and instructors.

(1 Mark - [Ap/C,2])

(ii) Determine how would you store different types of course content (text, video, quizzes) in the database?

a) Store everything as a single text field.
b) Store all content as BLOBs (Binary Large Objects) without distinction.
c) Use separate tables for different content types (text, video URL, quiz questions/answers).
d) Rely solely on external file storage without database integration.

(1 Mark - [An/C,2])

(iii) Consider a relation Employees with attributes Employee_ID, Name, Department, and Salary.

Table A1.1 Employee relation

Employee_ID	Name	Department	Salary
101	Abi	CSE	10000
102	Sasi	CSD	15000
103	Diya	AIML	20000
104	Farina	IT	5000
106	Hari	AIDS	10000

Identify any three functional dependencies based on the Employee_ID as primary key under Domain Key Normal Form (DKNF).

(3 Mark - [Ap/C,2])

A2 A library database stores information about books (title, author, ISBN) and borrowers (name, address, phone number). The normalized design has separate tables for books and borrowers, linked by a loan table containing book ID and borrower ID.

(i) Identify the potential drawback which might arise from the normalized library database design?

a) Difficulty in adding new book attributes.
b) Increased complexity in deleting borrower information.
c) Slower queries to retrieve information about borrowed books.
d) Inability to track the number of books borrowed by each person.

(1 Mark - [Ap/C,2])

(ii) Determine which of the following is the primary benefit of normalizing the library database?

a) Reduced storage space for book data.
b) Eliminating redundancy of borrower details across loans.
c) Improved data security for borrower information.
d) Faster retrieval of borrowed books by a specific borrower.

(1 Mark - [An/C,2])

(iii) Suppose you're tasked with normalizing a database schema to DKNF for a library database stores information system where books and barrowers information are stored. Re-arrange the steps you would take to analyze the existing schema and refactor it into DKNF while ensuring data consistency and efficiency.

1. Identify dependencies within the schema.
2. Analyze the existing schema.
3. Ensure data efficiency.
4. Ensure data consistency.
5. Normalize the schema to DKNF.
6. Refactor the schema accordingly.

(3 Mark - [An/C,2])

A3 Consider the below Figure A3.1 which shows the steps involved in data base design using CASE tools.

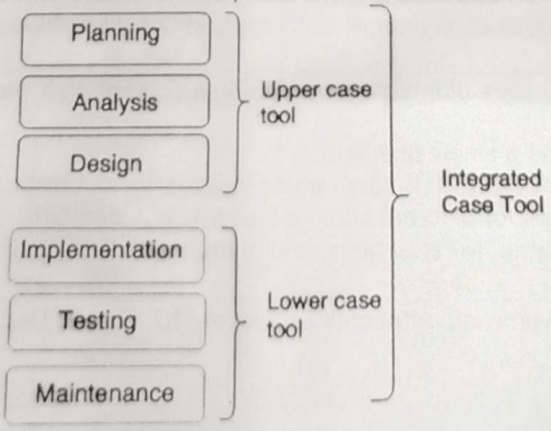


Figure A3.1 Database Design Steps

(i) Consider an E-R model for a university includes an "Enrollment" entity with attributes like student ID, course ID, and semester. A new requirement arises to track grades for each enrollment. How can you modify the model to accommodate this?

a) Add a "Grade" attribute to the "Enrollment" entity.
b) Create a separate "Course Grade" entity with foreign keys to Enrollment and a "Grade" attribute.
c) Convert the "Enrollment" entity to a one-to-one relationship with the "Student" entity.
d) Modify the "Course" entity to include a "Grade" attribute for each enrolled student.

(1 Mark - [An/C,2])

(ii) Identify the primary benefit of using a CASE tool for database design.

a) Improved data security through encryption.
b) Faster and more efficient database development process.
c) Automatic generation of error-free database code.
d) Reduced need for database administration expertise.

(1 Mark - [Ap/C,2])

(iii) Steps to be performed for designing database for complex data relationships is given below. Arrange the steps in correct order.

1. Build, populate, and test the database with sample data and queries.
2. Translate the ERD into a specific DBMS schema with data types and constraints.
3. Visually map entities, attributes, and relationships with an ERD.
4. Optimize physical storage and access patterns based on the DBMS.
5. Document everything and establish procedures for maintenance.
6. Define what data to store, understand entity relationships.

(3 Marks - [Ap/P,2])

SECTION B	COURSE OUTCOME 4	MAXIMUM : 20 MARKS

B1 Consider the below query optimization steps which are given in Figure B1.1.

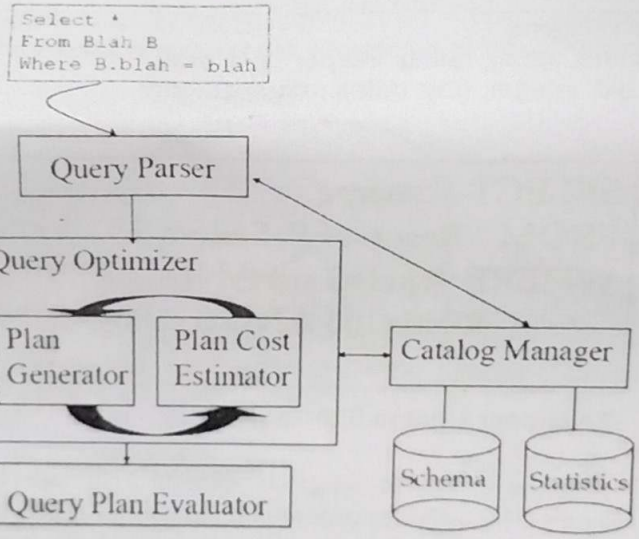


Figure B1.1 Query Optimization

(i) Which of the following statements about database execution plans is TRUE?

a) Execution plans are only generated for complex queries with joins and aggregations.
b) Execution plans provide a visual representation of the exact steps the database takes to execute a query.
c) Execution plans can be used to identify bottlenecks and optimize query performance. (This is the best option)
d) Execution plans are automatically generated and cannot be influenced by the user.

(1 Mark - [Ap/C,2])

(ii) Your company manages a large e-commerce platform with customer data (name, address, order history) and product data (description, specifications, reviews). You're experiencing performance issues with queries that join customer and product data. Which sharding approach would likely be MOST beneficial?

a) Vertical sharding, separating customer and product data to different databases.
b) Horizontal sharding, splitting customer data by region across multiple databases.

c) Horizontal sharding, splitting product data by category across multiple databases.
d) Neither vertical nor horizontal sharding would be effective in this scenario.

(iii) Consider a relation named Products with the following attributes: (1 Mark - [An/C,2])
product_id (integer)
category (varchar)
price (decimal)
in_stock (boolean)

For the below SQL statement identify the correct relational algebra expression.

```
SELECT product_id, category FROM Products WHERE price > 100 AND in_stock = TRUE;
```

a. $\pi_{product\_id,\ category}(\sigma_{price>100 \wedge in\_stock=TRUE}(Products))$

b. $\sigma_{product\_id,\ category}(\pi_{price>100 \wedge in\_stock=TRUE}(Products))$

c. $\pi_{product\_id,\ category}(\bowtie_{price>100 \wedge in\_stock=TRUE}(Products))$

d. $\pi_{product\_id,\ category}(\bowtie_{price<100 \wedge in\_stock=TRUE}(Products))$

(iv) If the statistics is removed from the query optimization process, is it possible to find the optimized execution plan for the SQL query? Justify your answer. (1 Mark - [An/C,2])

(v) Consider the below two relations, (3 Marks - [Ap/C,2])
Sailors (sid: integer, sname: string, rating: integer, age: real)
Reserves (sid: integer, bid: integer, day: dates, rname: string)

```
SELECT S.sname
FROM   Reserves R, Sailors S
WHERE  R.sid=S.sid AND
       R.bid=100 AND S.rating>5
```

Relational Algebra Expression:

$$\pi_{sname}\ (\sigma_{bid=100 \wedge rating>5}\ (Reserves \rhd\lhd_{sid=sid}\ Sailors))$$

Figure B1.2 SQL statement and Relational Algebra Expression

The above figure B1.2 shows the SQL query and its relational algebra expression. Based on that construct the query tree.

(4 Marks - [Ap/C,1])

Consider a situation that Bop is performing the transaction T1, John is performing transaction T2 and David is performing transaction T3 concurrently on the data items B and C. The schedule is given the Figure B2.1 and the precedence graph is shown in Figure B2.2.

T1	T2	T3
		write(C)
	write(B)	
	read(B)	
		read(C)
write(C)		
	read(C)	

Figure B2.1 Transaction Schedule

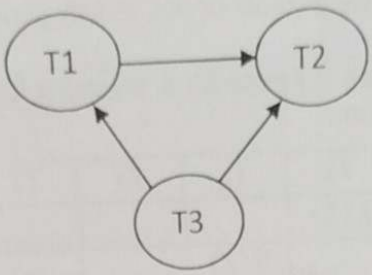


Figure B2.2 Precedence Graph

(i) Assume transaction A holds a share lock on R. If a transaction B also requests for a shared lock on R. Choose the correct option for the given situation.

a) it will result in a deadlock situation
b) it will immediately be granted
c) it will immediately be rejected
d) it will be granted as soon as it is released by A

(1 Mark - [Ap/C,2])

(ii) Concurrency control in RDMBS is important for which of the following reasons.

a) To ensure data integrity when updates occur to the database in a single user environment.
b) To ensure data integrity when read occur to the database in a single user environment
c) To ensure data integrity when updates occur to the database in a multi user environment
d) To ensure data integrity when read occur to the database in a multi user environment

(1 Mark - [Ap/C,2])

(iii) Suppose a database schedule **S** involves transactions T1,T2.....Tn. Construct the precedence graph of with vertices representing the transactions and edges representing the conflicts. Sample precedence graph is given in Figure B2.3.

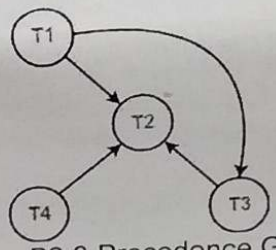


Figure B2.3 Precedence Graph

If S is serializable, which one of the following orderings of the vertices of the precedence graph is guaranteed to yield a serial schedule?

a) Topological order
b) Dept-first order
c) Breath-first order
d) Ascending order of transactions

(1 Mark - [Ap/C,2])

(iv) Consider the below schedule which is shown in Figure B2.4.

T1	T2	T3
	read(C)	
write(C)		
	write(C)	
read(B)		
		write(C)

Figure B2.4 Transaction Schedule

Check is there any deadlock in the above schedule (Figure B2.4) by constructing precedence graph and also check whether the schedule is conflict serializable or not. Justify your answer.

(4 Mark - [Ap/C,2])

(v) Consider the following schedule S shown in Figure B2.4, involving four transactions T1,T2,T3 and T4 and assume that time goes from top to bottom.

T_1	T_2	T_3	T_4
R(X)			
	W(X)		
		W(Z)	
			R(Z)
R(Z)			
W(Z)			

Figure B2.4 Transaction Schedule

Where R(X) denotes read operation on data item X by transaction T_i, W(X) denotes write operation on data item X by transaction T_i Predict the order of execution of all transaction.

(3Marks - [An/C,2])

B3 Consider Emily, who wants to book a flight from New York to Paris for a vacation. She searches for available flights, selects one, and proceeds to make the booking, entering her personal and payment details.

(i) If the system encounters an error while processing Emily's payment information, the transaction will move to the:

a) Active state
b) Committed state
c) Aborted state
d) Partially Committed state

(1 Mark - [Ap/C,2])

(ii) Choose the correct arrangement of the following isolation levels in order of increasing strictness:

1) Read Committed
2) Serializable
3) Repeatable Read
4) Read Uncommitted

a) 2,3,4,1
b) 4,1,3,2
c) 4,1,2,3
d) 2,3,1,4

(1 Mark - [Ap/C,2])

(iii) Two transactions are happening concurrently: Transaction 1 reads the total amount in an account to apply some interest, while Transaction 2 is transferring money to another account. If Transaction 1 applies interest based on the original balance before Transaction 2's transfer is complete, which ACID property is compromised?

a) Atomicity
b) Consistency
c) Isolation
d) Durability

(1 Mark - [Ap/C,2])

(iv) *ACID properties are crucial for maintaining data consistency and integrity, even in seemingly simple applications. Without them, data inconsistencies and errors could occur, leading to unreliable information. Identify the potential error that could violate ACID properties:*

```
BEGIN TRANSACTION;
UPDATE flights SET available_seats = available_seats - 1
WHERE flight_id = 123;
COMMIT TRANSACTION;
UPDATE accounts SET balance = balance - 200
WHERE account_id = 456;
```

(3 Marks -[Ap/C,2

(v) Consider the following two transactions:

```
T31: read(A);
     read(B);
     if A = 0 then B := B + 1;
     write(B).

T32: read(B);
     read(A);
     if B = 0 then A := A + 1;
     write(A).
```

Add lock and unlock instructions to transactions T31 and T32, so that they observe the two-phase locking protocol. Can the execution of these transactions result in a deadlock?

(4 Marks -[An/P,2])

SECTION C	COURSE OUTCOME 5	MAXIMUM: 20 MARKS

C1 Imagine you're tasked with building a social media platform like Twitter, where users post short messages, follow others, and interact with content in real-time. As your user base grows rapidly, you need a database that can handle massive amounts of data, high concurrency, and frequent updates efficiently. However, you also need to maintain data consistency and ensure a smooth user experience.

(i) Consider the above scenario and determine how to optimize user feed performance, what type of database structure could be considered?

a) De-normalize user data with frequently accessed information in the posts table (increases redundancy but improves read performance)
b) Store all user data in a single relational table (complex queries for user feeds)
c) Use a separate database for user data and posts (increases join complexity)
d) Implement complex materialized views for pre-computed user feeds (more maintenance overhead

(1 Marks - [Ap/C

(ii) Considering the massive data volume and real-time updates, what storage approach might be best suited for the core data (posts, user information)?

a) Traditional relational database (may not scale well for massive data and high concurrency)
b) NoSQL document store (flexible schema, good for large datasets)
c) Hierarchical database (not ideal for frequent updates and real-time interactions)
d) Graph database (better suited for social network connections)

(1 Marks - [Ap/C

(iii) To facilitate real-time communication features like notifications, what technology could be integrated with t database?

a) Traditional relational database queries (not ideal for real-time communication)
b) Messaging queue (enables asynchronous communication between servers and clients)
c) File system for storing notification data (less efficient for real-time access)
d) Email system (not designed for real-time communication)

(1 Marks - [Ap/C,2]

(iv) Identify the type of Database (NoSQL or NewSQL) that would be more suitable for the following scenarios.
Scenario 1: If scalability and sheer write performance are your primary concerns, and eventual consistency is acceptable
Scenario 2: If strong data consistency, familiar SQL querying, and horizontal scalability are paramount

(2 Marks - [An/C,2

(v) Consider the below table with missing fields. Identify the missing fields.

Features	SQL	NoSQL	NewSQL
Schema	? Rdbms.	Schema-free	Both
SQL support	? Yes	Depends on the system	Yes, with Enhan fea
ACID	Yes	NO-(BASE)	Yes
OLTP	Partially Support	Not Support	Full Support
Queries	Low Complex Queries	Highly ? Complex Queries	Both.

(4 Marks - [An/C,2

n selecting a NEWSQL database for a mission-critical financial system that handles numerous concurrent sactions, find any two correct features that become a crucial consideration in this process among the wing listed below.

a) Performance & Scalability
b) Reliability & Scalability
c) Availability & Compliance
d) Security & Compliance
e) Community & Reliability
f) Scalability & Reliability

(1 Marks - [An/C,2])

a NoSQL database which follows the principle of key-value store. The key-value store provides ability to me data called a value, inside a key. You can receive this data later only if you know the exact key used to . Redis is a NoSQL database so it facilitates users to store huge amount of data without the limit of a nal database. Redis supports various types of data structures like strings, hashes, lists, sets, sorted sets, s, hyper logs and geospatial indexes with radius queries.

Redis is a flexible, open-source (BSD licensed), in-memory data structure store, used as database, cache, and message broker. Identify the main benefits of using Redis' in-memory storage?

a) High durability and data persistence
b) Fast read and write performance
c) Complex data querying capabilities
d) Automatic data replication

(1 Marks– [An/C,2])

Identify the suitable word which matches the NewSQL databases features definition. Built-in cleaning processes for disks and storage devices.

a) *VoltDB*
b) *NuoDB*
c) *ClustrixDB*
d) *CockroachDB*

(1 Marks– [An/C,2])

viction policies determine what happens when new data needs to be stored, but the in-memory data store aches its capacity (maximum memory setting). To ensure efficient memory usage, Redis employs various iction policies to make space for new data by removing existing entries. What is the purpose of eviction icies in Redis?

a) *To enforce data access control for different user roles.*
b) *To manage memory usage and automatically remove least recently used data when the memory reaches capacity.*
c) *To encrypt data at rest.*
d) *To define data replication rules.*

(1 Marks– [Ap/C,2])

diagram depicts a simplified interaction between a web application (Client), a cache (Main), and a ase (MySQL) A user (Client), a content delivery network (Main), and a content library (MySQL). Fill out issing Parts.

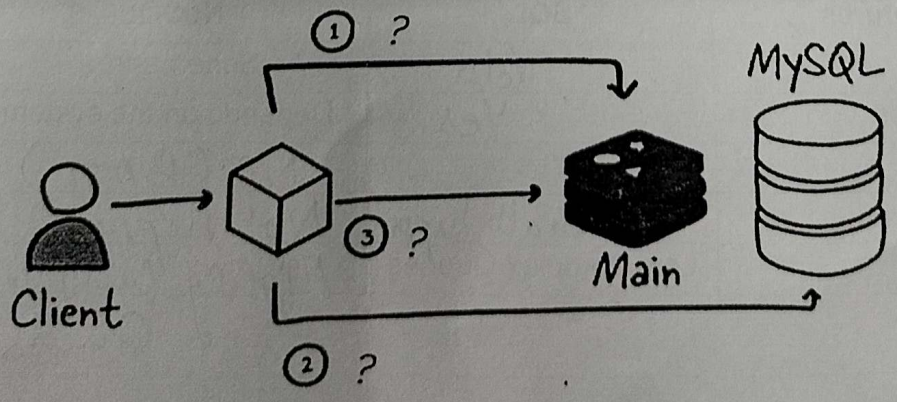
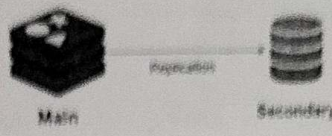
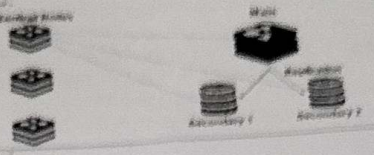
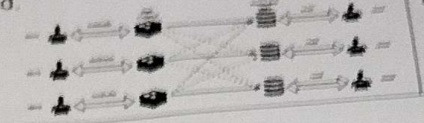


Figure C2.1 Web application with MySQL

(3 Marks - [An/C,2])

(v) Match the Following

1. Redis Sentinel	a. 	a. It provide constant availability of the Redis system
2. Single Redis Instance	b. 	b. It spreads the data we are storing across multiple machines
3.Redis Cluster	c. 	c. In this the Users sets up and runs small instances
4. Redis HA (High Availability)	d. 	d. It Helps in scale reads from Redis

(4 Marks - [An/C.2])

C3 The database server starts and loads the entire dataset (or a designated portion) from persistent storage (usually a disk) into main memory (RAM). This loading process happens once during startup and can be optimized for faster initialization. In-memory databases leverage simpler data structures compared to disk-based databases due to the faster access times of RAM, potentially improving processing efficiency. In-memory caching stores a temporary copy of frequently accessed data from a primary data source (often a disk-based database) to improve performance. In-memory caching improves read performance by reducing the need to access the slower primary data source. Transactions (data modifications) adhere to the ACID (Atomicity, Consistency, Isolation, and Durability) principles. Queries are executed directly against the data in RAM, resulting in significantly faster response times compared to querying disk-based databases. Applications interact with the in-memory database server using standard database protocols and APIs.

(i) Rearrange the process in correct order, Here's a breakdown of the in-memory database process

4,3,2,1,5

1. In-memory databases might perform background tasks (Background Operations)
2. Complex queries can be processed efficiently due to the speed of RAM access and potentially simpler data structures.(Query Processing)
3. All data access and manipulation (reads, writes, updates) occur entirely within RAM. (Data Processing)
4. The database server starts and loads the entire dataset (or a designated portion) from the persistent storage (usually a disk) into the main memory (RAM) (Database Startup)
5. The server processes queries and transactions entirely within RAM, returning results to the client application with minimal latency.(Client Interaction)

(4 Marks - [Ap/C.2])

(ii) An independent database management system that primarily relies on RAM for data storage and retrieval known as In-Memory Database. A layer (often integrated with an application or primary data source) that stores a temporary copy of frequently accessed data in RAM known as In-Memory Caching. Identify the correct and Incorrect statements with respect to the in-memory database and In- memory Caching.

a. In-memory databases store the entire dataset (or a designated portion) in RAM for faster data access. It can leverage simpler data structures due to the speed of RAM access.

A1) (i) d (ii) C (iii) E-ID → Name → dept → Salary

A2) (i) C (ii) B (iii) 2 1 5 4 3 6

A3) (i) B (ii) B (iii) 6 3 2 4 1 5

B1 → c, a, a, (NO), dias

B2 → b, c, a, Precedence graph, T3→T4→T1→T2

B3 → c, b, c, error, commit, local unlock

		b. In-memory caching stores the entire dataset. (It focuses on frequently accessed portions to reduce load on the primary data source).The data is always consistent with the primary data source. (Data consistency between cache and primary source requires update mechanisms). c. In-memory caching stores a temporary copy of frequently accessed data from a primary data source (often a disk-based database) to improve performance. In-memory caching improves read performance by reducing the need to access the slower primary data source. (3 Marks - [An/C,2])
	(iii)	You're the database administrator for an e-commerce platform storing customer information (names, addresses, credit card details) in a relational database. Determine the type of attack is most likely to compromise customer credit card details if a hacker gains access to the database server? a) Denial-of-Service (DoS) attack (Disrupts service availability, not data theft) b) SQL Injection attack (Allows unauthorized access to manipulate data) c) Brute-force password attack (Targets user accounts, not specific data) d) Man-in-the-middle attack (Intercepts communication, less relevant for stored data) (1 Marks - [Ap/C,2])
	(iv)	A social media platform you manage stores user profiles with names, email addresses, and birthdates. A recent data breach reveals that a large portion of user data has been leaked. What security best practice could have potentially mitigated the impact of this breach? a) Storing user passwords in plain text (Makes them easily accessible if breached) b) Implementing strong password policies for users (Reduces brute-force attacks) c) Granting all users full administrative privileges (Increases risk of unauthorized access) d) Storing sensitive data like credit cards (Not applicable to this scenario) (1 Marks - [Ap/C,2])
1-c-a 2-a-c 3-d-b 4-b-d	(v)	You're developing a new mobile application that interacts with a database containing customer purchase history. The application uses a pre-defined username and password hardcoded into the app's source code for database access. What security risk does this approach pose, and how can it be addressed? a) The application is user-friendly and easy to use (Security concern, not a risk) b) The database is secure because only the app can access it (Anyone with the app code can access) c) Hardcoded credentials are easily maintained (Security risk, not a benefit) d) This approach promotes strong password practices (Incorrect, hardcoded credentials are weak) (1 Marks - [Ap/C,2])

End of Question Paper

C1 → A, B, B, NoSql, NewSQL, —, Performance & Scalability

C2 → B, D, B, —, Match. Security & Compliance

C3 → 4,3,2,1,5 | C, In, C | B, D, C

A1 1 1 3

A2 1 1 3

A3 1 1 3

B1 1 1 1 3. 4

B2 1 1 1 4 3

B3 1 1 1 3 4

C1 1 1 1 2 4 1

C2 1 1 1 3 4

C3 4 3 1 1 1