Quiz 1

CSL7750_2023: Distributed Database Systems
Department of Computer Science and Engineering
Indian Institute of Technology, Jodhpur

Q.1 Consider a distributed database system with multiple sites. If one site decides to operate in a mode where it does not share any data with other sites, which feature of distributed databases is it exhibiting ?
a) Data Replication
b) Data Fragmentation
c) Autonomy
d) Data Allocation
Answer: c) Autonomy

Q.2 In a distributed database system, which of the following scenarios would likely result in the highest degree of data inconsistency ?
(a) High data replication with low autonomy
(b) High data fragmentation with high autonomy
(c) Low data replication with high data fragmentation
(d) High data replication with high data fragmentation
Answer: (b) High data fragmentation with high autonomy

Q.3 A distributed database system aims to achieve both high availability and low redundancy. Which technique would be most suitable?
(a) Replicating data at all sites
(b) Fragmenting data without any replication
(c) Controlled replication with selective fragmentation
(d) Centralized data storage
Answer: (c) Controlled replication with selective fragmentation

Q.4 Which of the following best describes the main challenge of query processing in distributed database systems?
(a) Deciding the order of operations in a query
(b) Determining the location of data fragments and the cost of data retrieval
(c) Ensuring data consistency after query execution
(d) Replicating the query results to all sites
Answer: (b) all

Q.5 In a distributed database, which fragmentation technique can potentially lead to the most redundancy?
(a) Horizontal Fragmentation
(b) Vertical Fragmentation
(c) Derived Fragmentation

(d) Hybrid Fragmentation
Answer: (a) Horizontal Fragmentation

Q.6 Which of the following is a challenge associated with data replication in distributed databases?
(a) Ensuring data consistency across replicas
(b) Reducing storage requirements
(c) Improving data availability
(d) Reducing network traffic
Answer: (a) Ensuring data consistency across replicas

Q.7 Which of the following best describes the principle of location transparency in a distributed database?
(a) Data can be moved between sites without affecting the system
(b) Users can access data without knowing its location
(c) All data is stored in a central location
(d) Data is transparently replicated across sites
Answer: (b) Users can access data without knowing its location

Q8. Which of the following is not the transparency type required in the DDB scenario?

1. Data organization transparency
2. Fragmentation transparency
3. Replication transparency
4. Migration Transparency
5. Failure Transparency

Options:

(a) All the given transparencies are required
(b) Option 1, 2, and 3 are correct
(c) Option 4, and 5 are correct
(d) None of the above

Ans: a

Q 9. A multinational corporation wants to enhance its data processing capabilities for its sales data spread across various regions. The company aims to improve query performance and ensure fault tolerance. Which approach would be most suitable for achieving these goals?
A) Distributed Database System
B) Parallel Database System
C) Hybrid Database System
D) Centralized Database System
Answer:
B) Parallel Database System

Explanation: Parallel database systems involve using multiple processors or cores within a single machine to perform tasks simultaneously. This approach improves query performance by distributing tasks among processors. It also ensures fault tolerance through redundant hardware

components. Given the company's goals of enhancing query performance and ensuring fault tolerance, a parallel database system would be the most suitable choice.

Q 10. In complete vertical fragmentation of R, the projection list (L_1, L_2, ..., L_n) satisfies which of the following conditions?

1) $L_1 \cup L_2 \cup \ldots L_n$ = ATTRIBUTE_SET(R).
2) $L_1 \cap L_2 \cap \ldots L_n$ = ATTRIBUTE_SET(R).
3) $L_i \cap L_j$ = PRIMARY_KEY(R) for any $i \neq j$
4) $L_i \cup L_j$ = PRIMARY_KEY(R) for any $i \neq j$

Options:

a) Option a and c are correct
b) Options a and d are correct
c) Options b and c are correct
d) Options b and d are correct

Ans: a

Q 11. What does Execution autonomy refer to?

a) Ability to decide whether and how much to share its functionality
b) the extent to which each node can decide on sharing of information with other nodes.
c) refers to independence of users to act as required.
d) refers to independence of data model usage and transaction management techniques among nodes.

Q 12. What are the types of transparency in Hybrid fragmentation?

1. Location transparency
2. Fragmentation transparency
3. Hazard transparency.
4. Replication transparency

Options:

a. All are correct
b. 1, 2, 3 are correct
c. 1, 2, 4 are correct
d. 1 and 4 are correct

Ans: c

Indian Institute of Technology Jodhpur

Minor_1

Course: DHP7750 - Distributed Database Systems

Date: September 10, 2023; Duration: 2:30-3:30PM; Total Marks: 30

Instructions:

1. Read all questions carefully.
2. Answers must be precise. Verbosity will lead to reduction in scores.
3. All questions are compulsory.
4. All assumptions must be clearly stated as part of the answer.
5. You are encouraged to make reasonable assumptions as part of your answer.
6. You are encouraged to use diagrams to substantiate your answer.

Best Wishes!

Q.1. Briefly discuss the key considerations and challenges associated with data allocation, fragmentation, and replication in the design of a distributed database system. Comment on how these considerations impact the overall performance, consistency, and availability of data?

[3 + 3]

Q2. Consider an airline reservation database, comprising:

1) Flight_description: <flight number, departure and arrival place and time, number of seats available, number of seats reserved, cost (assume only one fare)>
2) Passenger_description: <code, name, address, phone number>
3) Reservation_description: <passenger code, flight number, seat reserved>

Consider the following two applications:

1) A request about flight availability at a terminal => all information about the flight is shown.
2) A request about reservations, which includes:
    a. Checking whether the passenger's data are already available, and if they are not, inserting the passenger's data
    b. Checking whether there are seats available (assume no overbooking)
    c. Inserting the reservation description

A. How would you ensure that application (2) performs correctly?
B. Imagine that you need to distribute the database over four sites which are geographically located at New Delhi, Mumbai, Kolkata, and Chennai.
    a) What user statistics would you consider appropriate for justifying your design?
    b) For the distributed design provide:

1) The definition of the global schema.
2) The definition of the fragmentation schema:
    i) For horizontal fragmentation, provide a complete and minimal set of predicates.
    ii) Design a vertical fragmentation schema such that: site 1 (New Delhi handles departures), site 2 (Mumbai handles arrivals and seat reservations), site 3 (Kolkata manages fare costs), and site 4 (No flight-related data is stored at this site).

C. If a passenger at Site 1 (New Delhi) wants to book a seat on Flight 101 (New Delhi to Mumbai) - How would such a query be resolved?

D. What would your considerations be for defining the allocation schema?

[A:2 + B(a):2 + B(b.1):2 + B(b.2.i):2 + B(b.2.ii):2 + C:2 + D:2]

Q.3

A. Define a “microservice architecture” for an online furniture store that supports local artisans and ships their products all over the country.

B. Comment on “(microservice architecture models) force a substantial amount of data management tasks into the application-level”, through the example you designed in (A).

C. What kind of “cross-microservice queries” would your application, designed in (A), provide for?

D. How would you use a “virtual microservice” to help the application design in (A)?

E. Define “flexible querying” with a simple example - preferably from the example designed in (A).

[2 + 2 + 2 + 2 + 2]

Indian Institute of Technology Jodhpur

Major Examination

Course: DHP7750 - Distributed Database Systems

Date: December 5, 2023; Duration: 6 - 8 PM; Total Marks: 50

Instructions:

1. **Read all questions carefully.**
2. **Answers must be precise. Verbosity will lead to reduction in scores.**
3. **You are encouraged to use neat diagrams to substantiate your answer.**

Best Wishes!

Q1.

Answer the following questions. **[Clearly state all assumptions.]**

a. Design a logic for horizontal fragmentation and vertical fragmentation for the table-snippet shown below.
b. Write SQL queries to create these fragments.
c. State the operators you would use to get back the original table from both fragmentations.

Customer ID	Name	Gender	City	Place ID	Age
101	Rakesh	M	Delhi	D1	23
102	Poonam	F	Pune	M1	24
103	Priya	F	Delhi	D1	29
104	Mahesh	M	Agra	U1	24
105	Sneha	F	Pune	M1	35
...	...	...	...	...	...

[2 + 2 + 2 = 6]

Q2.

For the scenarios given below, suggest which architecture (client-server or peer-to-peer) would be suitable. Explain **very** briefly. **[Clearly state all assumptions]**

a. Case 1: Consider a distributed gaming platform where players interact with each other in real-time.
b. Case 2: Imagine a large e-commerce platform where customers can browse products, add items to their cart, and make purchases.

[2 + 2 = 4]

Q3.

Given the following schema and each of the queries in (a) and (b):

A. Write the relational algebra expressions for the given queries.
B. Draw the optimized operator tree for the expressions.

[Clearly write your assumption of the primary key for each of the tables.]

Instructor (ID, name, dept_name, salary)
Teaches (ID, course_id, sec_id, semester, year)
Course (course_id, title, dept_name, credits)

a. Find the names of all instructors in EE, along with the titles of the courses they are teaching in 2023 and their total credits.
b. Find the names of all instructors in CSE who have taught greater than equal to 8 credits in 2022.

[(1.5 (relational algebra) + 1.5 (optimized tree)) * 2 = 6]

Q4.

Imagine: A banking system with multiple branches (nodes) that need to process concurrent transactions. Each branch maintains its local database of customer accounts. A customer can perform transactions (withdrawal or deposit) at any branch. Answer the following questions for the scenario:

a. Design semantic integrity constraints [Predefined, Precompiled, General].
b. Design efficient enforcement algorithm [Detection or Prevention].

[Use declarative language for specifying integrity constraints and the enforcement algorithm. Clearly state your assumptions for the local database for each of the nodes]

[2 + 2 = 4]

Q5.

Consider a distributed transaction **T** operating under the two-phase commit protocol. Let **N0** be the coordinator node, and **N1**, **N2**, **N3** be the participant nodes. Consider that the following messages were sent:

Time stamp	Messages
T1	**N0** to **N1**: "Phase1: PREPARE"
T2	**N0** to **N2**: "Phase1: PREPARE"
T3	**N1** to **N0**: "OK"
T4	**N0** to **N3**: "Phase1: PREPARE"
T5	**N3** to **N0**: "OK"

Answer the following questions: **[Clearly state your assumptions.]**

a. Which node needs to send a message at **T6** for the transaction to run to completion?
b. To which node should this message be sent?
c. If **N0** received an "ABORT" response from **N3** at time **T5** – What would happen in a 2PC protocol environment?
d. If ➔ **N0** successfully receives all "OK" messages from the participants from the first phase (possibly after **T6**) ➔ It then sends the "Phase2: COMMIT" message to all participants at time **T7** but **N2** crashes before it receives this message. What would the status of **T** be if **N2** comes back on-line at **T8**?

[1.5 * 4 = 6]

Q6.

a. Highlight differences between a single lock manager and distributed lock manager (DLM) approaches for concurrency control.
b. Consider: A distributed social media platform using a DLM experiences communication delays between nodes. Explain how this delay might impact the efficiency of concurrency control and propose a strategy to mitigate this impact. **[Clearly state your assumptions.]**

[2 * 2 = 4]

Q7.

Consider: A distributed e-commerce platform with multiple warehouses handling inventory management. The system allows for concurrent updates to the inventory to reflect new shipments, customer orders, and returns. Answer the following questions: **[Clearly state your assumptions, use diagrams, and demonstrate that your solution works.]**

a. Identify one blocking problem that could occur, and how would you solve the same?
b. Identify one distributed deadlock scenario that could occur, and what would your detection algorithm be?

[5 * 2 = 10]

Q8. Given the use case in **Q.1.** **[Use diagrams and clearly state your assumptions.]**

a. Would the VFAR algorithm improve your vertical fragmentation logic? Briefly explain why / why not.
b. Given the fragments derived from the VFAR algorithm, how would you use the DRNNA algorithm for data distribution? Clearly state the advantages and disadvantages of your design.

[5 * 2 = 10]