

# **POSSESSION OF MOBILES IN EXAM IS UFM PRACTICE.**

Name \_\_\_\_\_

Enrolment No. \_\_\_\_\_

Jaypee Institute of Information Technology, Noida

T1 Examination, Even 2023

B.Tech (CSE/IT), VI Semester

**Course Title:** Introduction to Large Scale Database Systems  
**Course Code:** 21B12CS314

**Maximum Time : 1 hr**  
**Maximum Marks : 20**

CO1: Infer the background processes in queries & transactions, explain how these impact database design.

CO2: Choose appropriate ways of storing data and optimizing queries

CO3: Explain challenge of big data and demonstrate the comparison of relational database systems with NoSQL databases

CO4: Compare & discover the suitability of large databases to store, query & analyse various forms of data

CO5: Apply techniques of data fragmentation, replication and allocation to design a distributed or parallel database system.

**Q1[CO2, Marks 4]** Consider a file of 16384 records where each record is 32 bytes long. The key field is of size 6 bytes and the file organization is unspanned (i.e., record of a file is stored inside the block only if it can be stored completely inside it). The file system has a block size of 1024 bytes and the size of a block pointer is 10 bytes. Find the average number of blocks to search for a record with indexing for the following 2 cases:

- The file is ordered on the key field.  1024 / 32
- The file is ordered on a non-key field where size of the non-key field is 8 bytes and there are 12028 unique values for the field.

**Q2[CO1, Marks 4]** For the schedule below:

S: R1(A), W1(A), W2(A), R3(A), W2(A), W1(A), W3(A)

Is the schedule S view serializable? What is the necessary condition for any schedule to be view serializable?

**Q3[CO1, Marks 5]** a) Insert lock and unlock requests in the schedule S' and check if it is allowed by 2PL.

S': R1(X), R2(X), R3(X), R1(Z), R2(Y), R3(Y), W1(X), W2(Z), W3(Y), W2(Y)

What is the disadvantage of simple locking which does not follow 2PL?

b) Is the following schedule recoverable, strict or cascadeless?

S'': W1(A) W2(A) a2 c1 R3(A) R3(B) c3

(Note that a denotes abort and c denotes commit of a transaction)

**Q4[CO3, Marks 3]** As a database architect, you need to select a database management system for a video game development team. From player profiles to telemetry to leaderboards, data is crucial to make games work. The games are multiplayer, online and keep evolving. Will you select a relational or non-relational database for the video game data? Justify your choice.

**Q5 [CO1, Marks 4]** Apply timestamp ordering based concurrency control to the following schedule (Note st denotes start of a transaction):

S'': st1, st2, st3, R1(A), R2(B), W1(C), R3(B), R3(C), W2(B), W3(A)

If Thomas Write Rule is applied, what will happen in the execution?

**Course Title:** Introduction to Large Scale Database Systems  
**Course Code:** 21B12CS314**Maximum Time:** 1Hr  
**Maximum Marks:** 20

After Pursuing this course, students will be able to

- C331-3.1** Infer the background processes involved in queries and transactions, and explain how these impact on database operation and design
- C331-3.2** Choose appropriate ways of storing data and optimize queries.
- C331-3.3** Explain the concept and challenge of big data and demonstrate the comparison of relational database systems with NoSQL databases
- C331-3.4** Compare and discover the suitability of appropriate large databases to manage, store, query, and analyze various forms of big data
- C331-3.5** Apply techniques for data fragmentation, replication, and allocation to design a distributed or parallel database system

Attempt all questions.

Q1. Consider the following relational schema and SQL query. The database stores information about employees, departments, and company finances.

Emp(eid, did, salary, hobby);  
Dept(did, name, floor, phone);  
Finance(did, budget, sales, expenses);

```
SELECT d.name, f.budget
FROM Emp e, Dept d, Finance f
WHERE e.did = d.did AND d.did = f.did
AND d.floor = 1
AND e.salary > 59000
AND e.hobby = 'swimming';
```

(a) Translate this query to relational algebra. Show the initial query tree corresponding to your algebra query just above.

(b) Depict another query tree for the above query by showing how the initial query tree from part (a) can be optimized using query optimization methods.

(c) Suppose that the following information is available:

- There are primary (clustering) indexes on all primary keys.
- There are secondary (non-clustering) indexes on Emp(did), Emp(salary), and Dept(floor).
- There are a total of 50,000 employees and 5,000 departments (each with corresponding financial information) in the database.
- The system's statistics indicate that employee salaries range from 10,001 to 60,000 (inclusive), employees enjoy 200 different hobbies, and the company has two floors in the building.

For each of the query's base tables (Emp, Dept and Finance), estimate the number of rows that would be initially selected if all selection predicates were pushed down as much as possible.

**Q2.** Zincamin is a newly formed software firm by Anjana, Sumit and Chandan, which helps farmers about the weather prediction. They provide weather service where farmers can call any one of them up and receive a weather report that is updated every 15 minutes. Since they are keeping costs low they decide to all work out of their own homes. Anjana talks to local Meteorologist and negotiates a contract where he calls Anjana every 15 minutes with the latest weather report at which point she calls Sumit and Chandan to update them. They open their service and things are going great. However, one day there is a bad snow storm and it knocks out the phone line between Anjana and Sumit and Chandan. They continue to get phone calls from farmers and Anjana continues to get weather updates but now farmers get different reports if they call Anjana than if they call Sumit or Chandan. Depict from the scenario which type of systems is being followed here and do suggest what to do if this happens again. [CO3] [3]

A. P

**Q3.** You need to store the following data in MongoDB about a restaurant collection which consists of documents containing restaurant id, name, address (building, street, Zipcode), cuisine, grade (date, grade and score). Give the command to create the required collection and write the queries to find

restaurant

- a. To find document from ~~emp~~ collection where name starts with s and ends with k.
- b. Display documents where in collection, street field not have Ob, Mob
- c. Write a MongoDB query to display the next 5 restaurants after skipping first 5 which are in the borough Bronx. street
- d. Find the restaurants which do not prepare any cuisine of American and achieved a score more than 70 and located longitude value less than a given number.
- e. Update zipcode to '585' for all the restaurants located at street Medina.
- f. List limitations of creating and using indexes.

[CO4] [7]

**Q4.** What do you mean by Meta data in HDFS? List the files associated with metadata. Also state the problem in having lots of small files in HDFS? [CO4] [3]

**POSSESSION OF MOBILES IN EXAM IS UFM PRACTICE.**

Name

Enrollment No

**Jaypee Institute of Information Technology, Noida  
End Term Examination, EVEN 2022  
B.Tech CSE/IT, VI Semester**

**Course Title : Introduction to Large Scale Database Systems  
Course Code : 21B12CS314**

**Maximum Time : 2 Hrs  
Maximum Marks : 35**

After Pursuing this course, students will be able to

**C331-3.1 Infer the background processes involved in queries and Transactions and explain how these impact on database operation and design.**

**C331-3.2 Choose appropriate ways of storing data and optimize queries**

**C331-3.3 Explain the concept and challenge of big data and demonstrate the comparison of relational database systems with NoSQL databases.**

**C331-3.4 Compare and discover the suitability of appropriate large databases to manage, store, query and analyze various forms of big data.**

**C331-3.5 Apply techniques for data fragmentation, replication and allocation to design a distributed or parallel database systems.**

**Q1 [CO1]** Consider a variant of tree protocol called forest protocol. The database is organized as a forest of rooted trees. Each transaction  $T_i$  must follow the following rules:

- The first lock in each tree may be on any data item.
  - The second and all subsequent locks in a tree may be requested only if the parent of the requested node is currently locked.
  - Data items may be unlocked at any time.
  - A data item may not be relocked by  $T_i$  after it has been unlocked by  $T_i$ .
- Show that the forest protocol does not ensure serializability. [2]

**Q2 [CO2]** Let relations  $R1(A,B,C)$  and  $R2(C,D,E)$  have the following properties:  $R1$  has 20000 tuples and  $R2$  has 45000 tuples. 25 tuples of  $R1$  fit on one block and 30 tuples of  $R2$  fit on one block. Estimate the number of block transfers and seeks required using Nested loop join and Block nested loop join for  $R1$  join  $R2$ . [4]

**Q3 [CO4]** Create a Student collection in MongoDB for the schema

{ Address(Street, Building), city, state, results (Date, grade, score), name, student\_id}. Write queries to answer the following questions.

- Find the student details that belong to Uttar Pradesh', achieved a grade point 'A' and not belong to the 'South Street'. The document must be displayed according to the state in descending order.
  - Find the student id, name for those students who gained grade 'A' and score '9' on date '02-06-2022'.
  - Find the student id, name , address and city for those students whose name begins with letter 'Aka' or who belongs to any state except 'Himachal Pradesh' and 'Uttar Pradesh'.
  - Find the student id, name, city and state for those students who contain 'Pri' as three letters somewhere in their name.
- [6]

**Q4 [CO4]** Consider the following relation schema that stores rainfall data for 2016-2021 for four cities of Haryana, Gujarat, Himachal Pradesh and Uttar Pradesh along with the set of applications that are frequently accessing the relation:

Rain\_Stats (City, state, year, rainfall\_rcvd)

Application1: Find the rainfall received for a particular state.

Application2: Find the rainfall received from year 2019-2021.

Apply and show all the steps for primary horizontal fragmentation to fragment the given relation according to the requirement and check for correctness of the fragments.

[6]

**Q5 [CO5]** What form of parallelism ( Interquery, interoperation or intraoperation) is likely to be the most important for each of the following tasks?

- Increasing the throughput of a system with many small queries.
- Increasing the throughput of a system with a few large queries, when the number of disks and processors are large. [3]

**Q6 [CO5]** Perform vertical partitioning by using Bong Energy Algorithms to obtain a vertical set of attributes in table A. Assume A1 is the primary key. The frequency of execution of queries (Q1, Q2, Q3, Q4) at three different site has been given in table B. [8]

Table A

|    | A1 | A2 | A3 | A4 |
|----|----|----|----|----|
| Q1 | 0  | 1  | 1  | 0  |
| Q2 | 1  | 1  | 1  | 0  |
| Q3 | 1  | 0  | 0  | 1  |
| Q4 | 0  | 0  | 1  | 0  |

Table B

|    | S1 | S2 | S3 |
|----|----|----|----|
| Q1 | 10 | 20 | 0  |
| Q2 | 5  | 0  | 10 |
| Q3 | 0  | 35 | 5  |
| Q4 | 0  | 10 | 0  |

**Q7[CO5]** Social networking site Facebook maintains a feature which indicates 'list of friends'. When you visit someone's profile, you see a list of friends that you have in common. As they serve millions of requests every day, it has been decided to pre-compute calculations to reduce the processing time of requests. Use map reduce to calculate everyone's common friends from the friends list (Person -> [List of friends]) given below and show the results.

[6]

A -> B C D  
B --> A C D E  
C-> A B D E  
D -> A B C E  
E -> B C D

**Jaypee Institute of Information Technology, NOIDA**

T1 Examination, Odd Semester 2019

B.Tech VI semester

**Course Title: Introduction to Large Scale Database Systems    Max Time: 1 Hr.****Course Code: 21B12CS314****Max Marks: 20****Note: Answer all questions.**

After pursuing this course the students will be able to:

- C331-3.1** Infer the background processes involved in queries and transactions, and explain how these impact on database operation and design
- C331-3.2** Choose appropriate ways of storing data and optimize queries.
- C331-3.3** Explain the concept and challenge of big data and demonstrate the comparison of relational database systems with NoSQL databases
- C331-3.4** Compare and discover the suitability of appropriate large databases to manage, store, query, and analyze various form of big data
- C331-3.5** Apply techniques for data fragmentation, replication, and allocation to design a distributed or parallel database system

| <b>Q.1</b> | Consider the following two transactions:<br>T1: read(A);<br>read(B);<br>if A=0 then B := B + 1;<br>write(B).<br>T2: read(B);<br>read(A);<br>if B=0 then A := A + 1;<br>write(A).  | <b>Marks</b><br><b>[3]</b><br><b>CO1</b>   |           |      |   |      |       |   |      |       |   |      |         |   |      |       |  |
|------------|---|--|-----------|------|---|------|-------|---|------|-------|---|------|---------|---|------|-------|--|
| <b>Q.2</b> | Add lock and unlock instructions to transactions T1 and T2, so that they observe the two-phase locking protocol. Can the execution of these transactions result in a deadlock?<br><br>Consider a Hard Disk in which Block Size =1024 Bytes, each record is of size =100 Bytes. If total no. of records are 30000 and the data entered in Hard Disk are in the sorted order with respect to Enroll_No is given in the student table:   | <b>Marks</b><br><b>[3+3]</b><br><b>CO2</b> |           |      |   |      |       |   |      |       |   |      |         |   |      |       |  |
| <b>Q.3</b> | <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">Enroll_No</th> <th style="text-align: left;">Aadhar_No</th> <th style="text-align: left;">Name</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>1010</td> <td>Veena</td> </tr> <tr> <td>2</td> <td>1001</td> <td>Arjun</td> </tr> <tr> <td>3</td> <td>2021</td> <td>Geetika</td> </tr> <tr> <td>4</td> <td>1012</td> <td>Arjun</td> </tr> </tbody> </table> <p>If we want to perform below queries:</p> <ol style="list-style-type: none"> <li>1. Select * from student_table where Enroll_No =2</li> <li>2. Select * from student_table where Aadhar_No= 2021</li> </ol> <p>Then which indexing would be applied for 1 and 2 query and determine the Number of block access required for both cases. Consider the Index table entry is of 15Bytes</p> | Enroll_No                                  | Aadhar_No | Name | 1 | 1010 | Veena | 2 | 1001 | Arjun | 3 | 2021 | Geetika | 4 | 1012 | Arjun | <b>Marks</b><br><b>[1+2+2]</b><br><b>CO1</b> |
| Enroll_No  | Aadhar_No   | Name                                       |           |      |   |      |       |   |      |       |   |      |         |   |      |       |  |
| 1          | 1010  | Veena                                      |           |      |   |      |       |   |      |       |   |      |         |   |      |       |  |
| 2          | 1001  | Arjun                                      |           |      |   |      |       |   |      |       |   |      |         |   |      |       |  |
| 3          | 2021  | Geetika                                    |           |      |   |      |       |   |      |       |   |      |         |   |      |       |  |
| 4          | 1012  | Arjun                                      |           |      |   |      |       |   |      |       |   |      |         |   |      |       |  |

Q.4

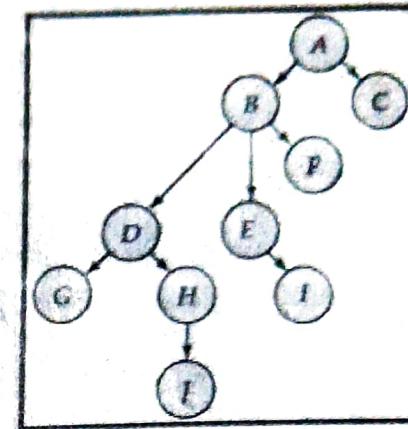
- (a) For the given schedule, indicate if it is strict schedule, in case of abort, is it recoverable and which transactions will be rollback. Write1(A), read1(C), write1(C), write2(A), write1(B), A2, C1, read 3(A), read 3(B), C3.

- (b) Consider the below schedule and determine which transaction will be rolled back if time stamp ordering protocol is applied and why. Assume that transactions are coming in the order T1 < T2 < T3

S1: RI(A) R2(B) W1(C) R3(B) RI(C) W2(B) W3(A)

- (c) Consider the below schedule and the tree depicting the partial ordering of data items. Determine whether the given schedule is following the rules of graph based protocol or not.

|    | T <sub>1</sub> | T <sub>2</sub> | T <sub>3</sub> |
|----|----------------|----------------|----------------|
| 1  | Lock-X(A)      |                |                |
| 2  | Lock-X(B)      |                |                |
| 3  |                | Lock-X(D)      |                |
| 4  |                | Lock-X(H)      |                |
| 5  |                | Unlock-X(D)    |                |
| 6  | Lock-X(E)      |                |                |
| 7  | Lock-X(D)      |                |                |
| 8  | Unlock-X(B)    |                |                |
| 9  | Unlock-X(E)    |                |                |
| 11 |                |                | Lock-X(B)      |
| 12 |                |                | Lock-X(E)      |
| 13 |                | Unlock-X(H)    |                |
| 14 | Lock-X(B)      |                |                |
| 15 | Lock-X(G)      |                |                |
| 17 | Unlock-X(D)    |                | Unlock-X(E)    |
| 18 |                |                | Unlock-X(B)    |
|    |                |                | Unlock-X(G)    |



Marks  
[1+3  
+2]  
CO1

**POSSESSION OF MOBILES IN EXAM IS UFM PRACTICE**

Name \_\_\_\_\_

Enrollment No. \_\_\_\_\_

**Jaypee Institute of Information Technology, Noida**

**T2 Examination, April 2023  
B.Tech VI Semester**

**Course Title : Introduction to Large Scale Database Systems Maximum Time : 1 Hour  
Course Code : 21B12CS314 Maximum Marks : 20**

**CO1: Infer the background processes in queries and transactions & explain how these impact database design**

**CO2: Choose appropriate ways of storing data and optimize queries**

**CO3: Explain concept & challenge of big data & compare relational database system with NoSQL databases**

**CO4: Compare & discover suitability of appropriate large databases to manage, store, query & analyse forms of big data**

**CO5: Apply techniques for data fragmentation, replication and allocation to design distributed & parallel database system**

**Q1.[CO2, Marks 4] For the schema and query**

**Artist( Aid, name, age, country)**

**Painting( Pid, title, medium)**

**Sold( Aid, Pid, price)**

**SELECT A.name**

**FROM Artist A, Painting P, Sold S**

**WHERE A.Aid= S.Aid AND S.Pid= P.Pid AND A.country='USA' AND P.medium='oil'**

Convert the SQL query into Relational algebra assuming no indexes. Show a physical query plan for this query. Suggest an alternate query plan for an optimized query. Will any index(es) be required for the optimization?

**Q2. [CO2, Marks 7] a) Let R and S be two relations with the following schema**

**R(P,Q,R1,R2,R3) and S(P,Q,S1,S2)**

Where (P,Q) is the key for both relations. Which of the following expressions are equivalent:

- (i)  $\Pi_P(R \setminus S)$
- (ii)  $\Pi_P(\Pi_{P,Q}(R) \cap \Pi_{P,Q}(S))$
- (iii)  $\Pi_P(\Pi_{P,Q}(R) - (\Pi_{P,Q}(R) - \Pi_{P,Q}(S)))$

b) Give an instance of relations X and Y (each having A as one of the attributes) that shows whether the expressions are/are not equivalent

$$\Pi_A(X-Y) \text{ and } \Pi_A(X)-\Pi_A(Y)$$

c) If an index is available on the attribute branch\_city for the relation Branch, describe how the following will be executed, where  $\neg$  is the negation

$$\sigma_{\neg(\text{branch\_city} < \text{"Chennai"})}(\text{Branch})$$

**Q3. [CO2, Marks 4]** Consider query  $\sigma_{(A < 12 \text{ AND } C \geq 10)}(X \Delta Y)$  on relations X(A,B) and Y(B,C). Assume that  $T(R)$  is the number of tuples in a relation R and  $\text{DOM}(R,A)$  is the domain of values of attribute A in a relation (Note data values are uniformly and independently distributed in all columns)

$$T(X)=3000, T(Y)=2000$$

$$\text{DOM}(X,A)=30 \text{ (integers from 0 inclusive to 30 exclusive)}$$

$$\text{DOM}(X,B)=20 \text{ (integers from 0 inclusive to 20 exclusive)}$$

$$\text{DOM}(X,A) = 10 \text{ (integers from 5 inclusive to 15 exclusive)}$$

$$\text{DOM}(Y,B) = 20 \text{ (integers from 0 inclusive to 20 exclusive)}$$

Estimate the statistics of  $T(X \Delta Y)$ ,  $\text{DOM}(X \Delta Y, A)$ ,  $\text{DOM}(X \Delta Y, B)$  and  $\text{DOM}(X \Delta Y, C)$

**Q4.. [CO4, Marks 5]** Create a Restaurants collection in MongoDB, Insert documents with appropriate fields and values. Write Queries to:

- display all documents in the collection
- display fields: restaurant\_id, name and cuisine but exclude the field \_id for all documents in the collection
- How many restaurants have grade A
- Find restaurants serving cuisine 'Chinese' and having score more than 90
- Find the restaurants who do not prepare the cuisine 'American' and are located in India