

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)

Sale(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] 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) $\Pi_P(R \bowtie S)$

(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 $\sigma_{(A < 12 \text{ AND } C \geq 10)}(X \bowtie Y)$ on relations X(A,B) and Y(B,C). Assume that T(R) is the number of tuples in a relation R and 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)

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

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

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

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

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

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

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