Total No. of Questions : 12] [Total No. of Pages : 4

P1328

[3664]-350

B.E. (IT)

ADVANCED DATABASE MANAGEMENT

(2003 Course) (414442) Sem I

Time : 3 Hours] *[Max. Marks : 100*

Instructions to the candidates :

1) *Answers to the two sections should be written in seperate books.*
2) *Neat diagrams must be drawn wherever necessary.*
3) *Assume suitable data, if necessary.*
4) *Section I : Q1 or Q2, Q3 or Q4, Q5 or Q6.*
5) *Section II : Q7 or Q8, Q9 or Q10, Q11 or Q12.*

SECTION - I

Q1) a) Explain Transaction Server Process Structure. **[6]**

b) What factors result in skew when a relation is partitioned on one of its attributes by Hash Partitioning and Rang Partitioning? In each case, what can be done to reduce the skew? **[5]**

c) Write a short note on Parallel Hash Join. **[5]**

OR

Q2) a) Describe different approaches to handle cache coherency problem in Parallel Databases. **[8]**

b) Evaluate how well partitioning techniques support the following types of data access.

i) Scanning the entire relation.

ii) Locating tuple associatively.

iii) Locating all tuples such that the value of given attribute lies within a specified range. **[8]**

Q3) a) Define Distributed Databases. Explain types of Distributed Database Management System Architectures. **[6]**

b) State different types of failures in distributed systems and explain failure handling in distributed database using 2Phase Commit protocol. **[6]**

c) Describe how LDAP can be used to provide multiple hierarchical views of data, without replicating the base level data. **[6]**

P.T.O.

OR

Q4) a) Compute semi-join r α s for the relations r and s. **[6]**

Relation r

A	B	C
1	2	3
4	5	6
1	2	4
5	3	2
8	9	7

Relation s

C	D	E
3	4	5
3	6	8
2	3	2
1	4	1
1	2	3

b) Explain the different approaches to detect distributed deadlock which neither site can detect based solely on its local wait-for-graph. **[6]**

c) Explain Optimistic methods for Distributed Concurrency Control. **[6]**

Q5) a) Consider the following nested relational schema.

Emp = (ename, childrenset setoff (children), skillset setoff (skills)

Children = (name, birthday), Birthday = (day, month, year)

Skills = (type, Examset setoff (Exams)), Exams = (year, city).

Answer the following:

i) Write DTD and XML file.

ii) Write a query in XPath to list all skill types in Emp.

iii) Find the names of all the employees who have a child who has a birthday in March.

iv) Find the those employees who took an examination for the skill type "typing" in the city "Pune".

v) List all skill types in Emp. **[10]**

b) Compare and contrast the two-tier, three-tier and n-tier architecture for Web-DBMS. **[6]**

OR

Q6) a) Write short notes on: **[10]**

i) Axes of XPath. ii) SOAP.

b) Explain simple type and Complex type of XML Schemas with suitable example. **[6]**

SECTION - II

Q7) a) Explain different conceptual schemas of Data Warehouse design with suitable example. **[10]**

b) One of the advantages of Data Warehouse is that we can use it to track how the contents of a relation change over time; in contrast, we have only the current snapshot of a relation in a regular DBMS. Discuss how you would maintain history of a relation R; taking into account that 'old' information must somehow be purged to make space for new information. **[6]**

OR

Q8) a) Write short notes on: **[10]**

i) OLAP ii) Meta Data.

b) Explain the various options / steps for designing fact table and summary table for Inventory system. **[6]**

Q9) a) Write steps of Hunt's Algorithm to construct decision tree. Consider following Data set. **[12]**

i) Calculate Information gain of Refund, Marital Status and Taxable Income.

ii) Calculate Gini Index of Refund, Marital Status and Taxable Income.

iii) Calculate Error Rate of Refund, Marital Status and Taxable Income.

iv) Draw Decision Tree.

Instance	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

b) Write a short note on Text Mining. [6]

OR

*Q10)*a) Explain Data Preprocessing in Data Mining. [6]

b) Write K-means algorithm in details and apply this algorithm on the following items to cluster. Assume k = 2. [6]

Object	Attribute 1	Attribute 2
Medicine A	1	1
Medicine B	2	1
Medicine C	4	3
Medicine D	5	4

c) Explain Naïve Bayes Classifier with suitable example. [6]

*Q11)*a) Define Information Retrieval System. Describe how it is differ from database system. [8]

b) Write a short note on:

i) Web Search Engine.

ii) Retrieval Effectiveness. [8]

OR

*Q12)*a) Explain different factors for relevance ranking in information retrieval system. [8]

b) Explain different Indexing of Documents approaches. [8]