Total No. of Questions : 12] [Total No. of Pages : 7

P1319

[3864]-417

B.E. (IT)

ADVANCED DATABASE MANAGEMENT

(2003 Course) (414442)

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

Instructions to the candidates :

1) *Answers to the two sections should be written in separate books.*
2) *Neat diagrams must be drawn wherever necessary.*
3) *Assume suitable data, if necessary.*
4) *Section I : Q. 1 or Q. 2, Q. 3 or Q. 4, Q. 5 or Q. 6.*
5) *Section II : Q. 7 or Q. 8, Q. 9 or Q. 10, Q. 11 or Q. 12.*

SECTION - I

Q1) a) Explain Parallel Hash Join with suitable example. **[6]**

b) Histogram are used for constructing load balanced range partition. **[6]**

i) Suppose a histogram where values are between 1 and 100, and are partitioned into 10 ranges, 1-10, 11-20, ..., 91-100, with frequencies 15, 5, 20, 10, 10, 5, 5, 20, 5 and 5, respectively. Express a load balanced range partitioning function to divide the values into 5 partition.

ii) Write an algorithm for computing a balanced range partition with p partitions, given a histogram of frequency distributions containing n ranges.

c) Explain a nonuniform memory architecture (NUMA). **[5]**

OR

Q2) a) Describe interoperation parallelism, left-deep trees versus bushy trees, and query cost estimation. **[6]**

P.T.O.

b) Evaluate how well partitioning techniques support the following types of data access. [6]

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.

c) Explain cache-coherency protocol. [5]

Q3) a) State different types of failures in distributed systems and explain failure handling in distributed database using 2 Phase Commit Protocol. [5]

b) Consider the relations : [7]

Employee (name, address, salary, plant_number)

Machine (machine_number, type, plant_number)

Assume that the employee relation is fragmented horizontally by plant_number, and that each fragment is stored locally at its corresponding plant site. Assume that machine relation is stored in its entirely at the Armonk site. Describe a good strategy for processing each of the following queries.

i) Find all employees at the plant that contains machine number 101.

ii) Find all machines at the Almaden plant.

iii) Find employee ∞ machine.

c) Explain the technique that the database system you are using provides for dealing with inconsistent states that can be reached with lazy propagation of updates. [5]

OR

Q4) a) Define semi-join. Compute semi-join $r \alpha s$ for the relations r and s. [5]

Relation r			Relation s		
A	B	C	C	D	E
1	2	3	3	4	5
4	5	6	3	6	8
1	2	4	2	3	2
5	3	2	1	4	1
8	9	7	1	2	3

b) Consider multiple-granularity locking protocol. In distributed databases, the site containing the root object in the hierarchy can become a bottleneck. Modify the protocol to allow only intension locks on the root and implicitly grant all possible intension locks to every transaction. [7]

i) Explain why this modification works correctly, in that transactions continue to be able to set locks on desired parts of the hierarchy.

ii) Explain how it reduce the demand on the root.

c) Explain how LDAP can be used to provide multiple hierarchical view of data, without replicating the base-level data. [5]

Q5) a) Consider following DTD for bibliography. [12]

<!ELEMENT bib (book*)>

<!ELEMENT book (title, (author+ | editor+), publisher, price)>

<!ATTLIST book year CDATA #REQUIRED>

<!ELEMENT author (last, first)>

<!ELEMENT editor (last, first, affiliation)>

<!ELEMENT title (#PCDATA)>

<!ELEMENT last (#PCDATA)>

<!ELEMENT first (#PCDATA)>

<!ELEMENT affiliation (#PCDATA)>

<!ELEMENT publisher (#PCDATA)>

<!ELEMENT price (#PCDATA)>

Create XML document, XML Schemas and solve the following queries in XQuery on the bibliography fragment.

i) List books published by Addison-Wesley after 1991, including their year and title.

ii) Find pairs of books that have different titles but the same set of authors (possibly in a different order).

iii) For each book in the bibliography, list the title and authors, grouped inside a "result" element.

b) Describe the various issues for efficient evaluation of XML Queries. [4]

OR

Q6) a) <?xml version = "1.0" encoding = "UTF-8"?> [8]

<!ELEMENT bids (bid_tuple*)>

<!ELEMENT bid_tuple (userid, itemno, bid, bid_date)>

<!ELEMENT userid (#PCDATA)>

<!ELEMENT itemno (#PCDATA)>

<!ELEMENT bid (#PCDATA)>

<!ELEMENT bid_date (#PCDATA)>

Create XML document, XML Schemas and solve the following queries in XQuery on the bibliography fragment.

i) List the item number and description of the item(s) that received the largest number of bids, and the number of bids it (or they) received.

ii) List item numbers and average bids for items that have received three or more bids, in descending order by average bid.

b) Explain XML schemas restrictions and facets. [4]

c) Write a short note on SOAP. [4]

SECTION - II

Q7) a) Explain Binning method and Regression method to handle noisy data in Data Warehouse. [6]

b) Explain Data Reduction strategies in Data Warehouse. [6]

c) Write a short note on Materialized view. [5]

OR

Q8) a) Design conceptual model for Financial Services data warehouse. [6]

b) List various features of fact data and explain the guidelines to be followed while determining facts from dimensions. [6]

c) Explain how meta data can be used for data transformation and loading, and query generation in data warehouse. [5]

Q9) a) Consider following training set : [10]

Outlook	Temperature	Humidity	Wind	Class Attribute
Sunny	Hot	High	False	N
Sunny	Hot	High	True	N
Overcast	Hot	High	False	P
Rain	Mild	High	False	P
Rain	Cool	Normal	False	P
Rain	Cool	Normal	True	N
Overcast	Cool	Normal	True	P
Sunny	Mild	High	False	N
Sunny	Cool	Normal	False	P
Rain	Mild	Normal	False	P
Sunny	Mild	Normal	True	P
Overcast	Mild	High	True	P
Overcast	Hot	Normal	False	P
Rain	Mild	High	True	N

Write ID3 Classification algorithm. Construct a decision tree based on above training set using ID3.

b) Explain how to handle candidate item sets using hash tree with suitable example. [7]

OR

Q10) a) Consider following data set : [9]

Object	Attribute 1	Attribute 2	Attribute 3
A	1	1	2
B	3	2	4
C	3	4	6
D	4	6	3

Write K-means clustering algorithm. Find the cluster for the objects in data set with K = 2.

b) Consider following training data set : [8]

Age	Income	Student	Credit_rating	Buys_Computer
<= 30	high	No	Fair	no
<= 30	high	No	Excellent	no
31...40	high	No	Fair	yes
> 40	medium	No	Fair	yes
> 40	low	Yes	Fair	yes
> 40	low	Yes	Excellent	no
31...40	low	Yes	Excellent	yes
<= 30	medium	No	Fair	no
<= 30	low	Yes	Fair	yes
> 40	medium	Yes	Fair	yes
<= 30	medium	Yes	Excellent	yes
31...40	medium	No	Excellent	yes
31...40	high	Yes	Fair	yes
> 40	medium	No	Excellent	no

Write Naïve Bayesian Classifier algorithm. Consider Buys_Computer as a Class Attribute with values yes and no classes. Find the class label for data sample.

X = (age <= 30, Income = medium, Student = yes Credit_rating = Fair) using Naïve Bayesian Classifier.

Q11) a) Define Information Retrieval System. Describe how it is differ from database system. [6]

b) Write short notes on Signature Files. [5]

c) Explain the following terms in Information Retrieval with suitable example. [5]

i) Synonyms

ii) Homonyms

iii) Proximity

iv) TF-IDF

OR

Q12) a) Explain any two techniques that support the evaluation of Boolean and Ranked queries. [6]

b) Write short notes on : [10]

i) Web Crawler.

ii) Document Indexing.

❑❑❑❑