

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. [5]
b) Explain design issues in Parallel Database System. [6]
c) Explain any two Parallel Database Architectures. [6]

OR

- Q2) a) Evaluate how well partitioning techniques support the following types of data access [9]
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.
b) Write a short note on [8]
i) Cash Coherency Protocol.
ii) Parallel Query Optimization.

- Q3) a) If we are to ensure atomicity, all the sites in which a transaction T executed must agree on the final outcome of the execution T must either commit at all sites, or it must abort at all sites. Describe the technique or protocol used to ensure this property in detail. [7]
b) Write a short note on Persistent Messaging in Distributed Transaction Processing. [5]
c) Explain Heterogeneous distributed databases. [5]

OR

Q4) a) Define semi-join. Compute semi-join $r \bowtie s$ for the relations r and s. [7]

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) Describe the voting and read-any-write-all approaches to synchronous replication. [5]
- c) Explain Optimistic methods for Distributed Concurrency Control. [5]

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

```
<?xml version="1.0" encoding="UTF-8"?>
<!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.

- List books published by Addison-Wesley after 1991, including their year and title.
 - Find pairs of books that have different titles but the same set of authors (possibly in a different order).
 - For each book in the bibliography, list the title and authors, grouped inside a "result" element.
- b) Explain advantages and disadvantages of the Web-DBMS approach. [4]

OR

- Q6) a)** Consider following DTD for bid
 <?xml version = "1.0" encoding = "UTF-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.
- 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.
 - List item numbers and average bids for items that have received three or more bids, in descending order by average bid. [8]
- b) Describe the various issues for efficient evaluation of XML Queries. [8]

SECTION - II

- Q7) a)** Explain Data Reduction and Data Decretization preprocessing in Data Warehouse. [5]
- b) Explain cube and rollup extended aggregation operation with suitable example. [5]
- c) Explain Kimball's nine steps design for Data Warehouse. [7]

OR

- Q8) a)** Explain different conceptual schemas design for data warehouse with suitable example. [10]
- b) Explain different indexing techniques in Data Warehouse. [7]

- Q9) a)** Consider following training set [12]

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) Write a short note on Text Mining. [5]

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 $X = (\text{age} \leq 30, \text{Income} = \text{medium}, \text{Student} = \text{yes}, \text{Credit_rating} = \text{Fair})$ using Naïve Bayesian Classifier.

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

- b) Write short notes on [10]
 i) Signature Files.
 ii) Ranking Document Similarity.

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) Precision and Recall.

