

Total No. of Questions :12]

SEAT No. :

P3443

[4959]-218

[Total No. of Pages :4

B.E. (Computer Engineering)

ADVANCE DATABASES

(2008 Course) (Semester - II) (Elective - IIID)

Time : 3 Hours]

[Max. Marks :100

Instructions to the candidates:

- 1) *Answer 3 questions from section I and 3 questions from section II.*
- 2) *Answers to the two sections should be written in separate books.*
- 3) *Neat diagrams must be drawn wherever necessary.*
- 4) *Figures to the right indicate full marks.*
- 5) *Assume suitable data, if necessary.*

SECTION-I

Q1) a) What are the different partitioning technique? Give an example of query for which that partitioning technique would provide the fastest response. **[8]**

b) What factors could result in skew when a relation is partitioned on one of its attribute by **[8]**

i) Hash partitioning

ii) Range partitioning

In each case, what can be done to reduce the skew.

OR

Q2) a) What is parallelism? Explain the interquery & Intraquery parallelism. **[8]**

b) Explain partitioned parallel hash join. **[8]**

Q3) a) What are the different approaches for high availability in the distributed system. **[8]**

b) Explain distributed transaction management. **[8]**

OR

P.T.O.

- Q4)** a) Explain the kinds of data storage and failure in distributed system. [8]
b) Explain two phase commit protocol. How three phase commit protocol overcome the disadvantages of the two phase commit protocol. [8]
- Q5)** a) Why do we have the XML DTD? What is well-formed documents? Explain with an example. [8]
b) Why do we need to maintain state at the middle tier? What are cookies and how does a browser handle the cookies? [10]

OR

- Q6)** Write short note on the following: [18]
- a) XQUERY
 - b) XPATH
 - c) Thin & Thick Client
 - d) 3tier architecture

SECTION-II

- Q7)** a) What are different data cleaning methods? Explain outlier analysis. [8]
b) Explain architecture of data warehouse with a neat diagram. [10]
- OR
- Q8)** a) Differentiate between OLAP & OLTP. [6]
b) Explain the following operation on the multidimensional data. [6]
i) Roll up and drill down.
ii) Slicing & dicing
c) What is star schema? With an example design a star schema. [6]

Q9) a) Consider the following data set. **[8]**

Food Item	Protein content	Fat Content
F1	1.1	60
F2	8.2	20
F3	4.2	35
F4	1.5	21
F5	7.6	15
F6	2.0	55
F7	3.9	39

Find the cluster for the object in the dataset by using K-means algorithm, if $k=4$.

b) What is Best split? Explain ID3 algorithm to create decision tree. **[8]**

OR

Q10)a) Find the strong association rule by using Apriori algorithm for the given dataset which satisfy following requirements. **[8]**

- i) Support = 30%
- ii) Confidence = 90%

Customer	Products
C1	S1 S3
C2	S2
C3	S4
C4	S2 S3 S4
C5	S2 S3
C6	S2 S3
C7	S1 S2 S3 S4
C8	S1 S3
C9	S1 S2 S3
C10	S1 S2 S3

- b) Explain the following terms [8]
- i) Closed frequent itemset.
 - ii) Maximal frequent itemset.
 - iii) Outlier analysis.

Q11)a) What you mean by relevance ranking? Explain TF/IDF methods of relevance ranking for the Boolean & ranked query. [8]

- b) Explain the following: [8]
- i) Inverted Index
 - ii) Ontologies
 - iii) Stop words
 - iv) Random walk

OR

Q12)a) What is page ranking and popularity ranking? Explain in brief. [8]

- b) Explain the following terms: [8]
- i) Web crawlers.
 - ii) Homonyms.
 - iii) Vector space model.
 - iv) Synonyms.

EEE