Total No. of Questions : 12]

SEAT No. :

**P1815**

[Total No. of Pages :4

# [4859]-218
# B.E. (Computer Engineering)
# d-ADVANCED DATABASES
# (2008 Course) (Elective-III) (Semester-II)

*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 the methods of data storage in distributed system. **[8]**

OR

***P.T.O.***

***Q4)*** a) Discuss different kinds of failure in the distributed system and how to handle it? **[8]**

b) What are transparency in the distributed system? Discuss the relative advantages of centralized & distributed databases. **[8]**

***Q5)*** a) Why do we have the XML DTD? What is well- formed documents? Explain with an example. **[10]**

b) Why do we need to maintain state at the middle tier? What are cookies and how does a browser handle the cookies? **[8]**

OR

***Q6)*** Write short note on the following. **[18]**

a) XQUERY.

b) XPATH

c) Thin & Thick Client

d) 3 tier architecture.

## SECTION-II

***Q7)*** a) Explain multidimensional data models in details. **[8]**

b) Explain the following **[8]**

i) OLAP

ii) Data Cube.

OR

***Q8)*** a) Explain CUBE and ROLL-UP extended aggregation with suitable example. **[8]**

b) Explain the architecture of data warehouse and also explain different indexing technique used in data warehouse. **[8]**

***Q9)*** a) What are different data cleaning methods? Explain binning & outlier analysis. **[6]**

b) State and explain K-MEANS algorithm for clustering. **[6]**

c) Consider the following data set. **[6]**

| 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

OR

***Q10)***a) What is Best split? Explain ID3 algorithm to create decision tree. **[6]**

b) Explain the following: **[6]**

i) Text Mining.

ii) GINI Index.

iii) Information gain.

c) Find the association rule for the given dataset which satify following requirements. **[6]**

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 | |

***Q11)***a) What you mean by relevance ranking? Explain different methods of relevance ranking for the boolean & ranked query. **[8]**

b) Explain the following: **[8]**

i) Inverted Index

ii) Ontologies

iii) Stop words

OR

***Q12)***a) What is page ranking and popularity ranking? Explain in brief. **[6]**

b) Explain the following terms. **[10]**

i) Web crawlers.

ii) Homonyms.

iii) Vector space model.

iv) Zipfian distribution.