B. Tech.

(SEM. VIII) EXAMINATION, 2006-07
DATA MINING & WAREHOUSING

Time : 3 Hours] [Total Marks : 100

Note : Attempt all the questions.

1 Attempt any four parts : 5x4=20
   (a) Compare the two-tier and three-tier client server architecture.
   (b) What are the main software modules of a distributed DBMS? Discuss the main functions of such modules in the context of client server architecture.
   (c) What is a data warehouse? Explain various data warehousing architecture.
   (d) What is meant by data allocation and data replication in distributed database design?
   (e) How did Relational DBMS evolve from the centralized architecture to the client-server architecture?
   (f) Write short notes on the following:
      (i) Cluster system
      (ii) Distribution transparency and fragmentation transparency.

VB-1042] 1 [Contd...
2 Attempt any four parts: \[5 \times 4 = 20\]

(a) Explain multidimensional data base structure and relational database structure for data warehouse application.

(b) What do you mean by Metadata?

(c) Illustrate various database management system schemes for decision support system development.

(d) Compare and contrast two basic approaches to building a data warehouse.

(e) Explain mapping the data warehouse to multiprocessor architecture.

(f) What do you understand by data extraction? Also explain various transformation tools involved in data warehousing.

3 Attempt any four parts: \[5 \times 4 = 20\]

(a) Explain the term OLAP. Briefly explain the following OLAP operations:
   (i) Slicing a cube
   (ii) Drill down.

(b) What are the two basic approaches used to analyse a data warehouse database?

(c) Give the arguments in favour of multidimensional on-line analytical processing.

(d) What do you understand by virtual data warehouse? Also give the idea about data-marts.

(e) What kind of database can OLAP software use? What types of output can OLAP software provide to its user?

(f) Explain why intelligent system technology is useful in DSS, with its advantages and disadvantages.
4 Attempt any two parts: $10 \times 2 = 20$

(a) What are the various goals of data mining? And explain the various tools and techniques of data mining briefly. What are supervised and unsupervised learning methods? How do they help in data mining?

(b) Describe the working of the DBSCAN algorithm. Explain the concept of a duster as used in DBSCAN and ROCK and compare them, with the help of suitable example.

(c) Describe the essential features in a decision tree. How is it useful to classify data? What are the disadvantages of decision tree over other classification techniques?

5 Attempt any two parts: $10 \times 2 = 20$

(a) What is data visualization? How it helps in DSS model? Give the methods for visualization of generalized data?

(b) Discuss the role of data warehouse managers in data warehousing. Write down Codd's guidelines for OLAP and explain them.

(c) Explain the concept of 'Big data better Returns'. Also discuss the advantages and disadvantages of 'Big data-better return' concept.