



## **UNIT-1 DATA WAREHOUSING**

1. What are the uses of multifeature cubes? (Nov/Dec 2007)
2. Compare OLTP and OLAP Systems. (Apr/May 2008), (May/June 2010)
3. What is data warehouse metadata? (Apr/May 2008)
4. Explain the differences between star and snowflake schema. (Nov/Dec 2008)
5. In the context of data warehousing what is data transformation? (May/June 2009)
6. Define Slice and Dice operation. (May/ June 2009)
7. List the characteristics of a data ware house. (Nov/Dec 2009)
8. What are the various sources for data warehouse? (Nov/Dec 2009)
9. What is bitmap indexing? (Nov/Dec 2009)
10. What is data warehouse? (May/June 2010)
11. Differentiate fact table and dimension table. (May/June 2010)
12. Briefly discuss the schemas for multidimensional databases. (May/June 2010)
13. How is a data warehouse different from a database? How are they similar? (Nov/Dec 2007,Nov/Dec 2010)
14. What is descriptive and predictive data mining? (Nov/Dec 2010)
15. List out the functions of OLAP servers in the data warehouse architecture. (Nov/Dec 2010)
16. Differentiate data mining and data warehousing. (Nov/Dec 2011)
17. What do you understand about knowledge discovery? (Nov/Dec 2011)

### **PART B**

1. Describe in detail about access tools types and the overall architecture of data warehouse?
2. Demonstrate in detail about Data marts and data warehouse administration?
3. Explain the functional blocks needed to build a data warehouse?
4. Describe in detail about Mapping the Data warehouse to a multiprocessor architecture?
5. what are the information needed to support DBMS schemas for Decision support?
6. Discuss in detail about Vendor approaches and access to legacy data?
7. Describe in detail about data extraction and transformation tools?
8. How would you explain Metadata implementation with examples ?
9. Describe in detail about Bitmapped indexing and STARjoin and index?



## **UNIT-2**

### **BUSINESS ANALYSIS**

1. What is the need for preprocessing the data? (Nov/Dec 2007)
2. What is parallel mining of concept description? (Nov/Dec 2007) (OR) What is conceptdescription? (Apr/May 2008)
3. What is dimensionality reduction? (Apr/May 2008)
4. Mention the various tasks to be accomplished as part of data pre-processing. (Nov/ Dec 2008)
5. What is data cleaning? (May/June 2009)
6. Define Data mining. (Nov/Dec 2008)
7. What are the types of concept hierarchies? (Nov/Dec 2009)
8. List the three important issues that have to be addressed during data integration.(May/June 2009) (OR) List the issues to be considered during data integration. (May/June 2010)
9. Write the strategies for data reduction. (May/June 2010)
10. Why is it important to have data mining query language? (May/June 2010)
11. List the five primitives for specifying a data mining task. (Nov/Dec 2010)
12. What is data generalization? (Nov/Dec 2010)
13. How concept hierarchies are useful in data mining? (Nov/Dec 2010)
14. How do you clean the data? (Nov/Dec 2011)
15. What is need of GUI? (Nov/Dec 2011)

### **PART B**

1. How would you explain in detail about reporting query tools and applications?
2. Compare in detail about tool categories in query tools?
3. How would you show your ing about the need for OLAP?
4. Discuss in detail about the OLAP tools?
5. Describe in detail about state of the market?
6. Examine the approaches used in multidimensional and multirelational OLAP?
7. Discuss in detail about the OLAP guidelines?
8. Describe in detail about the applications of business analytics?
9. Examine in detail about various categories of OLAP tools
10. Compose in detail about OLAP tool and its application in the internet?



### UNIT-3

#### DATA MINING

1. Define frequent set and border set. (Nov/Dec 2007)
2. How are association rules mined from large databases? (Nov/Dec 2007)
3. List two interesting measures for association rules. (April/May 2008) (OR)
4. What are Iceberg queries? (April/May 2008)
5. What is over fitting and what can you do to prevent it? (Nov/Dec 2008)
6. In classification trees, what are surrogate splits, and how are they used? (Nov/Dec 2008)
7. Explain the market basket analysis problem. (May/June 2009)
8. Give the difference between Boolean association rule and quantitative association rule.  
(Nov/Dec 2009)
9. Give the difference between operational database and informational database. (Nov/Dec 2009)
10. List the techniques to improve the efficiency of Apriori algorithm. (May/June 2010)
11. Define support and confidence in Association rule mining.(May/June 2010) (Nov/Dec 2010)
12. What is FP growth? (May/June 2010)
13. How Meta rules are useful in constraint based association mining. (May/June 2010)
14. Mention few approaches to mining Multilevel Association Rules. (Nov/Dec 2010)
15. How rules do help in mining? (Nov/Dec 2011)
16. What is transactional database? (Nov/Dec 2011)

#### PART B

1. Demonstrate in detail about data mining steps in the process of knowledge discovery?
2. Explain in detail about data mining functionalities?
3. What approach would you designed to mine interestingness of patterns?
4. Distinguish various data mining task primitives?
5. Explain in detail about the classification of data mining systems?
6. Demonstrate in detail about integration of data mining system with a data warehouse?
7. Describe in detail about data preprocessing?
8. Describe in detail about data preprocessing?
9. Describe in detail about data integrity and reduction in data preprocessing?
10. Discuss in detail about various data transformation and data discretization techniques?



## **UNIT-4**

### **ASSOCIATION RULE MINING AND CLASSIFICATION**

1. What is tree pruning? (Nov/Dec 2007)
2. List the requirements of clustering in data mining. (Nov/Dec 2007)
3. What is classification? (April/May 2008) (May/June 2009)
4. What is the objective function of the K-means algorithm?
5. The naïve Bayes classifier makes what assumption that motivates its name?
6. What is an outlier? (May/June 2009) (OR) Define outliers. List various outlier detection approaches. (May/June 2010)
7. Compare clustering and classification. (Nov/Dec 2009)
8. What is meant by hierarchical clustering? (Nov/Dec 2009)
9. What is Bayesian theorem? (May/June 2010)
10. What is Association based classification? (Nov/Dec 2010)
11. Why tree pruning useful in decision tree induction? (May/June 2010) (Nov/Dec 2010)
12. Compare the advantages of and disadvantages of eager classification (e.g., decision tree) versus lazy classification (k-nearest neighbor) (Nov/Dec 2010)
13. What is called Bayesian classification? (Nov/Dec 2011)

### **PART B**

1. Can you explain in detail about mining frequent pattern association and correlations?
2. How would you summarize in detail about mining methods and various kinds of association rules?
3. Describe in detail about what you learned to develop constraint and correlation based association mining?
4. What approach would you use to describe decision tree induction?
5. What are the features of Bayesian classification explain in detail?
6. What is the function of support vector machine?
7. Examine in detail about Lazy learners with examples?
8. Describe in detail about frequent pattern classification?
9. Demonstrate in detail about Back propagation?
10. Can you formulate a theory for Bayes and rule based classification techniques?

## **UNIT-5**

### **CLUSTERING AND APPLICATIONS AND TRENDS IN DATA MINING**

1. What do you go for clustering analysis? (Nov/Dec 2011)
2. What are the requirements of cluster analysis? (Nov/Dec 2010)



3. What is mean by cluster analysis? (April/May 2008)

A cluster analysis is the process of analyzing the various clusters to organize the different objects into meaningful and descriptive objects.

4. Define CLARANS.

It overcomes the problem of scalability that K-Medoids suffers from.

5. Define BIRCH, ROCK and CURE.

6. What is meant by web usage mining? (Nov/Dec 2007)(April/May 2008)(Nov/Dec 2009)  
(May/June 2010)

7. What is mean by audio data mining? (Nov/Dec 2007)

8. Define visual data mining. (April/May 2008)

9. What is mean by the frequency item set property? (Nov/Dec 2008)

10. Mention the advantages of hierarchical clustering. (Nov/Dec 2008)

11. Define time series analysis. (May/June 2009)

12. What is mean by web content mining? (May/June 2009)

13. Write down some applications of data mining. (Nov/Dec 2009)

14. List out the methods for information retrieval. (May/June 2010)

15. What is the categorical variable? (Nov/Dec 2010)

16. What is the difference between row scalability and column scalability? (Nov/Dec 2010)

17. What are the major challenges faced in bringing data mining research to market?  
(Nov/Dec 2010)

18. What is mean by multimedia database? (Nov/Dec 2011)

19. Define DB miner. (Nov/Dec 2011)

### **PART B**

1. Based on what you know how will you explain cluster analysis?

2. Describe in detail about categorization of major clustering methods?

3. Describe in detail about the features of K-means partitioning method?

4. Explain in detail about hierarchical and density based methods?

5. What is grid based method and examine about proximity based approach?

6. Demonstrate in detail about model based clustering methods?

7. Describe in detail about clustering high dimensional data?

8. How would you discuss the outlier analysis and detection techniques?

9. Will you explain in detail about data mining applications?

10. Design statistical approaches in outlier detection with neat design and with examples?