Q.1  
   a. What makes a pattern interesting? Can a data mining system generate all of the interesting patterns?  
   b. “Data Warehouse is an environment, not a product” Comment.  
   c. Explain the Drill-Down and Roll-up operations of OLAP.  
   d. Differentiate Agglomerative and Divisive Hierarchical Clustering.  
   e. What do you mean by Constraint-Based Association Mining?  
   f. Explain the importance of Data Cleaning.  
   g. Write four data mining applications for retail industry. (7 × 4)

Q.2  
   a. Discuss in details three main reasons why data warehouse modeling requires modeling techniques other than OLTP database modelling. (9)  
   b. What is a data mart? Differentiate between dependent and independent data marts. (5)  
   c. Every data structure in a data warehouse contains the time element. Why? (4)  

Q.3  
   a. What is the difference between the three main types of data warehouse usage: information processing, analytical processing, and data mining? (9)  
   b. Discuss the motivation behind OLAP mining (OLAM). With the help of a clean diagram discuss architecture of OLAM. (9)  

Q.4  
   a. In real-world data, tuples with missing values for some attributes are a common occurrence. Describe any five methods for handling this problem. (5)  
   b. What is Apriori property? Why it is used? Discuss the Apriori algorithm for discovering frequent itemsets for mining Boolean association rules. (8)
c. Association rule mining often generates a large number of rules. Discuss effective methods that can be used to reduce the number of rules generated while still preserving most of the interesting rules.

Q.5  
  a. Discuss two common techniques for assessing classifier accuracy, based on randomly-sampled partitions of the given data. Are there general techniques for improving classifier accuracy?

  b. What are the fields in which clustering techniques are used? Mention any four fields. Discuss basic requirements of cluster analysis.

  c. Why is outlier mining important? Briefly describe the different approaches behind statistical based outlier detection and distance based outlier detection.

Q.6  
  a. Why is decision tree induction popular? Discuss over-fitting of an induced tree and two approaches to avoid over-fitting using suitable example/diagrams.

  b. How can you use the Web as a data source for your data warehouse? What types of information can you get from the Web?

Q.7  
  a. Explain how data mining is used in banking industry.

  b. Name the major phases of a data mining operation. Out of these phases, pick two and describe the types of activities in these two phases.

  c. Explain data granularity and how it is applicable to the data warehouse.