

6. (a) Draw and explain HDFS architecture. Discuss the functions of NameNode and DataNode. What is secondary NameNode? Is it a substitute to NameNode?
 (b) How Hadoop streaming works with mapper and reducer phases?

Unit - III

4. Define Bayesian classification. Discuss the classification using back-propagation algorithm and SVM. Also explain a suitable technique to improve classification accuracy.
 5. Discuss the applications of WEKA tools in data mining. Illustrate, how clustering can be performed for a dataset in data mining. Explain with the help of any two popular clustering algorithms.

Unit - II

3. Discuss frequent pattern analysis with its advantages and disadvantages. Explain frequent pattern analysis for market basket analysis or any other application using Apriori algorithm with suitable small dataset.

- (b) What is data discretization? What are its different categories? How entropy is considered as measure of discretization?

Transaction ID	Item Bought
2000	A, B, C
1000	A, C
4000	A, D
5000	B, E, F

2. (a) Define association rule performance measures. What do you mean by 'confidence' and 'support' in association rule mining. Considering the following example, find the 'support' of frequent item sets {A}, {B}, {C}, {A, C},

Unit - I

- (f) Compare big data analytics with traditional data analytics.
 (g) Write the execution modes of PIG.
 (h) What is collaborative filtering?

I.

Compulsory Question :

- (a) How a method can be designed that mines the complete set of frequent itemsets without candidate generation?
- (b) What is data transformation? Give any two functions for data transformation.
- (c) Name the methods adopted for outlier detection.
- (d) How the quality of cluster can be accessed?
- (e) Describe how map-reduce computation executes.

7. (a) Explain the concepts of blocks and heartbeat messages in HDFS architecture. What are the benefits of block transfer?
- (b) Explain Hadoop architecture and its components with proper diagram.

Unit - IV

8. (a) Write MapReduce code for counting occurrences of specific words in input text files. Also write the commands to compile and run the code.
- (b) Explain how to communicate the outputs of data analysis using R. Also discuss the concept of data import in R language.

9. (a) What is Hadoop Ecosystem? Discuss various components of Hadoop ecosystem.
- (b) What is Hive in big data? Explain its architecture and summarize its features. How does Hive interact with Hadoop?