SHRIMATHI DEVKUNVAR NANALAL BHATT VAISHNAV COLLEGE FOR WOMEN
(AUTONOMOUS)
(Affiliated to the University of Madras and Re-accredited with 'A+' Grade by NAAC)
Chromepet, Chennai - 600 044.
M.Sc.Computer Science - END SEMESTER EXAMINATIONS - NOV'2024
SEMESTER - III
23PCSCT3008 - Big Data Analytics

Total Duration : 2 Hrs. 30 Mins. Total Marks : 60

Section B

Answer any **SIX** questions $(6 \times 5 = 30$ Marks)

1. Explain the Distributed File System (DFS) and its key characteristics. How does a DFS differ from a traditional file system?
2. Discuss the key components of Hive architecture and explain how they interact to process queries in a Hadoop environment.
3. Describe the HBase architecture, including RegionServers, HMaster, and HDFS integration.
4. Compare and contrast SQL and NoSQL databases in terms of structure, scalability, and use cases.
5. Examine the basic syntax and components of MongoDB Query Language (MQL).
6. Analyze the different methods for importing and saving data in MongoDB with examples.
7. Describe the key components of the Hadoop environment and explain how they interact to process large datasets.
8. Design a MapReduce algorithm to count the frequency of words in a given large text file.

Section C

I - Answer any **TWO** questions $(2 \times 10 = 20$ Marks)

9. Explain the key features of Apache Hadoop and describe how its architecture supports the storage and processing of large datasets.
10. Analyze the role of Apache ZooKeeper in distributed systems. Discuss its key features, including its architecture and data management capabilities.
11. Design a Spark application in Scala that performs data analysis on a large dataset. Outline the steps involved in reading data, processing it using transformations and actions, and writing the results back to storage.

Contd...

12. Construct Queries on DB's and documents in MongoDB with examples.

II - Compulsory question $(1 \times 10 = 10$ Marks)

13. Design a comprehensive solution for implementing Apache Hive to manage and analyze large datasets. Include specific practice examples that demonstrate the creation of a Hive table, data insertion, querying, and optimization techniques.

******