

[Time: 3 Hours]

[Total Marks: 80]

- N.B:** 1) Question No.1 is **compulsory**.  
2) Attempt any **three** from the remaining **five** questions.

Write a short note on following (any Four)

1. (a) Yarn Daemon (5)  
(b) Visualization using Bar Chart (5)  
(c) Dataset Vs DataFrames (5)  
(d) Joins in Hive (5)  
(e) HDFS Vs HBase (5)  
(f) DataTypes of Big Data (5)
2. (a) Illustrate the characteristics of social media which make it suitable for Big Data Analytics. (10)  
(b) Explain data visualization. List out and Illustrate the various Data Visualization Tools. (10)
3. (a) Explain the significance of Pig in Hadoop Ecosystem. Explain the Data Model in Pig Latin with suitable example. (10)  
(b) Explain the CAP Theorem and NoSQL Data architectural pattern with suitable example. (10)
4. (a) Discuss Matrix Multiplication by MapReduce. Perform Matrix Multiplication with 1-step method (10)  

Matrix A:	Matrix B:
1    2	5    6
3    4	2    3
- (b) Illustrate Hive. Explain Hive architecture with suitable example. (10)
5. (a) Explain Apache Spark. What are the advantages of Apache spark over Map Reduce. (10)  
(b) Illustrate NoSQL data architecture patterns with suitable example. (10)
6. (a) Explain the distributed storage system of Hadoop with the help of a neat diagram. (10)  
(b) Illustrate the Apache Kafka. Explain with suitable example the streaming of real time data with respect to Apache Kafka. (10)