

(3 Hours)**[Total Marks : 80]**

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

1. Write a short note on following (any Four)
 - (a) Challenges handling Big Data (5)
 - (b) Role of Name node and Data node in HDFS (5)
 - (c) Business drivers for NOSQL (5)
 - (d) Pig's execution modes (5)
 - (e) Transformation in Spark (5)
 - (f) Goals of data visualization (5)
2. (a) Compare and contrast Traditional Data Processing and Big Data Processing. (10)
- (b) Discuss the Read and Write operations in HBase with a proper workflow (10)
3. (a) Write and explain the algorithm for the following Matrix (A) and Vector (x) Multiplication using Map Reduce. (10)

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad x = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$$
- (b) What is the Hive Warehouse Directory and Meta-store? Explain their role in Hive. (10)
4. (a) Describe how Kafka ensures fault tolerance and scalability in a distributed environment. (10)
- (b) Explain the concept of data visualization and its importance in data analysis. Provide examples of different types of visualizations (10)
5. (a) What are the different transformations and operations in Pig Latin? Explain with examples. (10)
- (b) Compare different approaches to big data visualization. How do they differ in terms of scalability and efficiency? (10)
6. (a) Describe the role of YARN in Hadoop 2.0 and its key components with suitable diagram (10)
- (b) Discuss the difference between RDD, DataFrame, and Dataset in Spark. When should you use each? (10)
