



MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL

Paper Code : OEROB801A Big Data Analysis

UPID : 008397

Time Allotted : 3 Hours

Full Marks : 70

The Figures in the margin indicate full marks.

Candidate are required to give their answers in their own words as far as practicable

Group-A (Very Short Answer Type Question)

1. Answer *any ten* of the following :

[1 x 10 = 10]

- (I) What is Pig Script interfaces?
- (II) Which is a technique used to extract meaningful insights from data sets that are too large or complex to be processed by traditional data processing tools?
- (III) What is the process of transforming structured and unstructured data into a format that can be easily analyzed?
- (IV) What do you mean by indexing in HDFS?
- (V) Which is best described as a programming model that is used to construct Hadoop-based applications that can be scaled up and down?
- (VI) What is Pig Latin model?
- (VII) Which is a popular NoSQL database used for Big Data processing?
- (VIII) What is the term used for the ability of a system to handle increasing amounts of data and traffic without compromising performance?
- (IX) What is Job Tracker in building blocks of Hadoop?
- (X) Which is a method for analyzing data in which the data is split into smaller subsets and processed in parallel across multiple servers or nodes?
- (XI) What is the meaning of Hive Clients?
- (XII) What are the common input formats in Hadoop?

Group-B (Short Answer Type Question)

Answer *any three* of the following :

[5 x 3 = 15]

2. Explain the configuration parameters of a MapReduce framework. [5]
3. Explain the Pig Latin application flow. [5]
4. What does a block in the Hadoop Distributed File System (HDFS) mean? Explain briefly. [5]
5. Explain the different features of Hadoop. [5]
6. What is the need for Data Locality in Hadoop? [5]

Group-C (Long Answer Type Question)

Answer *any three* of the following :

[15 x 3 = 45]

7. (a) What is Mapreduce in Hadoop? [3]
 (b) When to use mapreduce with big data? [5]
 (c) Mention the core methods of Reducer. [7]
8. (a) What is a PIG? Explain its Installation Process. [7]
 (b) Explain two execution types or modes in PIG [8]
9. (a) How do you convert unstructured data to structured data? [5]
 (b) What is Data preprocessing? [3]
 (c) What are the steps required for data preparation? [7]
10. (a) What is Hadoop? [3]
 (b) Explain the importance of Hadoop technology in Big data analytics. [6]
 (c) Explain the core components of Hadoop. [6]
11. Explain the complete building blocks of Hadoop. Draw the block diagram and explain all the block separately. [15]