

Roll No. ....

Total Pages : 3

46215

BT-6/M-24

BIG DATA ANALYTICS

Paper : PE-IT-S-306A

Time : Three Hours]

[Maximum Marks : 75

Note : Attempt *five* questions in all, selecting at least *one* question from each unit. All questions carry equal marks.

#### UNIT-I

1. (a) Write short note on 4 Vs in Big Data and explain the Matrix Vector Multiplication by Map Reduce. (7)
- (b) What is Big Data? Why do we need the Drivers of Big Data in Hadoop? (8)
2. (a) Why do we need a Big Data Analytics in the Business? What is the Role of Drivers in Big Data? Justify with Example. (7)
- (b) What do you mean by Distributed File System? Why do we need to analyze Big Data? List out the characteristics of big data and challenges in handling big data. (8)

#### UNIT-II

3. (a) Who takes care of replication consistency in a Hadoop cluster and what do under/over-replicated blocks mean? Draw the Architecture of the Hadoop Cluster. (7)

46215/100/KD/1025



[P.T.O.  
19/6

- (b) Explain the Programming model of working of MapReduce. Explain Moving data in and out of Hadoop with suitable example. (8)
4. (a) What do you understand by Apache Hadoop? Draw the neat and clean Apache Hadoop Architecture. (7)
- (b) What do you mean by Apache Hadoop Ecosystem? Draw the Architecture of Hadoop. (8)

### UNIT-III

5. (a) Explain with a neat diagram the workflow of the Map reduce model. Also, write a mapper and a reducer code to find the number of occurrences of the word "PIET" in the "File.txt". (5)
- (b) Write HDFS Commands for the following operations :
- (i) to find the version of HDFS.
  - (ii) download and install Hadoop.
  - (iii) to run postmap services.
  - (iv) to run mapper and reducer with input file.
  - (v) to start and stop name node and data node.
- (2×5=10)
6. (a) What is NoSQL? Explain different NoSQL data models. Compare and contrast NoSQL vs. Relational Databases. Explain. (7)

- (b) An enterprise has 500 GB of unstructured data to be processed. Suggest a suitable architecture for this work. Explain with a diagram all the components of the file system supported by your solutions. (8)

#### UNIT-IV

7. (a) How do you run Map Reduce and Message Passing Interface (MPI) on YARN architecture? Discuss. (7)
- (b) Compare the Hadoop 1.0 and Hadoop 2.0. Which version most suitable in the Big Data Analytics. Justify. (8)
8. (a) What is Scheduler? What do you understand by YARN? What is the Role of Resource Manager and Node Manager in the YARN Framework? (7)
- (b) Explain the following Hadoop components :  
(i) Fair and Capacity (ii) MRV1 (iii) MRV2. (3+3+2= 8)
-