Roll No. .....

Total Pages : 3

46292

BT-6/M-24

BIG DATA ANALYTICS

Paper : PC-CS-AIDS-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) What is Big Data? What is the significance of the Data ETL process in Big Data Tools? (7)

   (b) Write a short note on 3 Vs in Big Data and Explain the Matrix-Vector Multiplication by Map Reduce. (8)

2. (a) What are the benefits of Big Data? Discuss the Challenges under the Big Data. How Big Data Analytics can be useful in the Development of Smart Cities? (7)

   (b) Why do we need Big Data Analytics in the Business? What is the Role of Drivers in Big Data? Justify with Example. (8)

UNIT–II

3. (a) Define the Data Serialization. What are the types of Big Data Analytics? (7)

(b) What do you mean by Apache Hadoop Ecosystem? Draw the Architecture of Hadoop. (8)

4. (a) Define and differentiate between SQL and NoSQL with respect to Big Data Tools. (7)

(b) 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. (8)

UNIT–III

5. (a) Write Map Reduce Code for Counting occurrences of specific words in the input text file(s). Also, write the commands to compile and run the code. (7)

(b) Explain the following :

(i) Mapper class.

(ii) Reducer class.

(iii) Job.

(iv) Task Tracker. (2+2+2+2=8)

6. (a) Define HDFS. Describe Name Node, Data Node and Block. Explain HDFS operations in detail. (7)

(b) Write 10 different Shell commands with Examples in Hadoop. (8)

UNIT-IV

7. (a) Describe the MapReduce execution steps with neat diagram. (7)

(b) Discuss the Hadoop System and Ecosystem components in Four Layers. (8)

8. (a) Illustrate YARN based execution model and its functions with a neat Diagram. (7)

(b) Illustrate the Hadoop core components with a neat Diagram. (8)
