3 Hours / 70 Marks   Seat No.

Instructions :   (1)   All Questions are *compulsory.*

(2)   Illustrate your answers with neat sketches wherever necessary.

(3)   Figures to the right indicate full marks.

(4)   Assume suitable data, if necessary.

Marks

1.   **Attempt any FIVE of the following :**   **10**

(a)   Define Big data.

(b)   State the significance of Big Data Analytics.

(c)   Define Data Science.

(d)   State the importance of Hadoop.

(e)   Explain the role of Name Node in HDFS.

(f)   State the use of Apache Spark.

(g)   Define RDD.

2.   **Attempt any THREE of the following :**   **12**

(a)   Explain the challenges with Big Data.

(b)   Describe the responsibilities of a Data Scientist.

(c)   Explain analytics flow for Big data.

(d)   Describe HDFS.

3. **Attempt any THREE of the following :** 12

   (a) Explain terminologies used in Big Data Environment.

   (b) Describe Mapping analytics flow to Big data stack.

   (c) Describe case study of Weather Data Analysis.

   (d) How Hadoop addresses the challenges of processing and analyzing large scale data ?

4. **Attempt any THREE of the following :** 12

   (a) Describe the user defined function.

   (b) Describe Hive Query Language.

   (c) Explain Spark Real Time use case.

   (d) Describe data frame operation.

   (e) How does Mapreduce work in Hadoop ? What role does it play in distributed data processing ?

5. **Attempt any TWO of the following :** 12

   (a) List the various domain specific of Big Data & explain any one with example of Big Data.

   (b) Give the key advantages of Hadoop. Also compare RDBMS & Hadoop.

   (c) Write a code for building spark SQL application with SBT.

6. **Attempt any TWO of the following :** 12

   (a) What is Hive ? Describe Hive Architecture.

   (b) Explain SERDE & RC file implementation.

   (c) State RDD. Explain RDD operation & creating a RDD.

   _____