22684

12425

3 Hours / 70 Marks Seat No. | | | | | | | | |

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

(2) Assume suitable data, if necessary.

Marks

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

(a) Define Big data analytics.

(b) List the terminology used in big data environments.

(c) State the various raw data sources.

(d) Define Hadoop.

(e) List advantages of Hadoop.

(f) List all hive data types.

(g) State spark shell.

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

(a) State the responsibilities of data scientist.

(b) Explain challenges with big data.

(c) State the different big-data stack.

(d) Demonstrate any one domain specific example of big data.

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

(a) Classify the analytics process.

(b) Describe mapping analytics flow to big data stack.

(c) Classify analytics flow for big data.

(d) List the features of Hadoop.

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

(a) Describe HDFS in detail.

(b) Describe Hive file format.

(c) Sketch Hive architecture.

(d) Explain spark core RDD operations.

(e) Explain spark real time use case for data analytics project architecture.

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

(a) Explain big data of wheather analysis.

(b) Explain Hadoop in detail.

(c) Execute RC file implementation.

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

(a) Execute various commands to create Hive table.

(b) Implement code for building SPARK SQL application with SBT.

(c) Explain Apache Spark Architecture.
