

Code: IT8T2D

**IV B.Tech - II Semester - Regular Examinations - March 2018**

**BIG DATA ANALYTICS  
(INFORMATION TECHNOLOGY)**

Duration: 3 hours

Max. Marks: 70

**PART – A**

Answer *all* the questions. All questions carry equal marks

11x 2 = 22 M

1.

- a) List any 4 V's of big data characteristics.
- b) Give an example of semi-structured and unstructured data.
- c) What is distcp?
- d) What do you know about file Reading Anatomy?
- e) List any 4 features of HDFS.
- f) How combiner functions are used in Hadoop?
- g) Define the terms Shuffle and Sort in MapReduce.
- h) Write any 2 output formats.
- i) Define Input Splits and Records in Hadoop.
- j) How Hive is different from Pig?
- k) List the four complex data types of Hive?

## PART – B

Answer any **THREE** questions. All questions carry equal marks.

3 x 16 = 48 M

2. a) Define MapReduce. Compare RDBMS with MapReduce. 8 M
- b) What is Hadoop Ecosystem? List and define any four elements of Hadoop Ecosystem. 8 M
- 3.a) HDFS is a filesystem designed for storing very large files with streaming data access. Examine the statement. 8 M
- b) Explain basic filesystem operations of HDFS. 8 M
- 4.a) Explain the process of Analyzing the Data with Unix tools. 8 M
- b) Explain MapReduce logical data flow with diagram. 8 M
5. Explain the following in Hadoop MapReduce. 16 M
- i) Job Submission    ii) Job Initialization  
iii) Task Assignment    iv) Task Execution
- 6.a) Compare HiveQL with SQL. 8 M
- b) What is Hbasic backdrop? Explain Whirlwind Tour of the Data Model. 8 M