

Code: CS7T1

**IV B.Tech - I Semester – Regular/Supplementary Examinations
October - 2018**

**BIG DATA CONCEPTS
(COMPUTER SCIENCE & ENGINEERING)**

Duration: 3 hours

Max. Marks: 70

PART – A

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

11 x 2 = 22

1.

- a) Differentiate between MapReduce and Traditional DBMS.
- b) Explain about the Key Value pairs in a Hadoop MapReduce.
- c) What are the two nodes operating in an HDFS cluster?
- d) List the two static methods used for getting a file system instance.
- e) What is MapReduce Application?
- f) What is structured, semi-structured, unstructured data?
- g) State the Relationship between Input Splits and HDFS Blocks.
- h) Define Lazy Output Format with syntax.
- i) List out various MR types.
- j) What are various File Permissions in HDFS?
- k) What is Binary output?

PART – B

Answer any **THREE** questions. All questions carry equal marks.
3 x 16 = 48 M

2. a) Explain about Apache Hadoop Project. 8 M
- b) Write a program for finding the maximum recorded temperature by year from NCDC weather records. 8 M
3. a) Write a java program to Demonstrating file status information from HDFS. 8 M
- b) With neat sketch explain about how data flows between the client interacting with HDFS. 8 M
4. a) Explain about the helper classes in hadoop for making it easier to run jobs from the command line. 8 M
- b) Write about Configuring the Development Environment in Hadoop. 8 M
5. a) With neat diagram explain about How Hadoop runs a MapReduce job? 10 M
- b) What Constitutes Progress in MapReduce? 6 M
6. a) Explain about Input Format class hierarchy. 8 M
- b) Write a program to output Format for writing a whole file as a record. 8 M