

Code: CS7T1

**IV B.Tech - I Semester –Regular / Supplementary Examinations  
MARCH - 2021**

**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) Explain briefly about Voluntary Computing.
- b) What are the two types of nodes that control the job execution process?
- c) Explain about Blocks in HDFS.
- d) Which Command is used to list the files in Hadoop?
- e) Write about HAR files briefly.
- f) Compare and contrast NoSQL vs. Relational Databases.
- g) What is serialization? Why is it needed?
- h) What is the mechanism used by HDFS to ensure data integrity?
- i) Explain data ingest with flume.
- j) Briefly explain MapReduce data flow.
- k) Why is a Block in HDFS so Large?

PART – B

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

3 x 16 = 48 M

2. a) Explain with example how the Data will be analyzed in Hadoop? 8 M
- b) With neat sketch explain about MapReduce logical data flow. 8 M
3. a) Given a text file containing multiple words, we need to count how many times each word appears in the file using MapReduce. Explain shuffle and sort in detail for this word count problem using a diagram. 8 M
- b) Explain about Hadoop file system and Anatomy of a File Read. 8 M
4. a) Describe Hadoop configuration API. 8 M
- b) Write an application to find the maximum temperature. 8 M
5. a) Explain with neat sketch how Hadoop runs a MapReduce job. 8 M

b) Draw the diagram which shows the relationship of the Streaming and Pipes executable to the task tracker and its child.

8 M

6. Explain about:

i.) Input Splits and Records

ii.) InputFormat class hierarchy

iii.) Preventing splitting

iv.) Processing a whole file as a record

16 M