PVP 14

Code: CS7T1

IV B.Tech - I Semester – Regular Examinations – October - 2017

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) Compare RDBMS with MapReduce approach?
b) What is Grid Computing?
c) State the necessity of MapReduce approach.
d) Define HDFS.
e) Classify the different Data ingesting methods in HDFS.
f) Draw the Transition diagram of an Oozie workflow.
g) What is GenericOptionsParser class?
h) Name the four independent entities in classic MapReduce of a job run.
i) How the Capacity Scheduler approach differs with Fair Scheduler?
j) Write the advantage of Lazy output.
k) What is the use of MultipleOutputs?

PART – B

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

2. a) Write & explain the classes in Java MapReduce API. 8 M

   b) Describe MapReduce dataflow with Multiple reduce tasks. 8 M

3. a) Explain Hadoop File system in detail. 10 M

   b) How to Read and Write the data into/from HDFS using Java API, explain with an example. 6 M

4. a) What is the use of MR unit? Explain MR unit test approach in MapReduce application. 8 M

   b) Explain different properties of Configuration class. 8 M

5. a) With a neat sketch explain how Hadoop runs a job using a Classic MapReduce. 8 M

   b) Explain the Task Execution of a job in the MapReduce System. 8 M

6. a) Write the OutputFormat class hierarchy in Hadoop. 8 M

b) Explain MapReduce types in Hadoop? 8 M