May 2024

B.Tech. (CE(DS)) (Sixth Semester)

Big Data Fundamentals (PCC-DS-603)

Time : 3 Hours]   [*Maximum Marks* : 75

Note : It is compulsory to answer all the questions (1.5 marks each) of Part A in short. Answer any *four* questions from Part B in detail. Different sub-parts of a question are to be attempted adjacent to each other.

Part A

1. (a)   When do you call a data big ?   1.5

    (b)   What is the role of name node ?   1.5

    (c)   What are the challenges of conventional system ?   1.5

    (d)   What are the various dimensions of growth of big data ?   1.5

    (e)   Who is generating big data and what are the ecosystem projects used for processing ?

                                                   1.5

(f) What are the various features of big data ?   1.5

(g) Explain HiveQL.   1.5

(h) What is meant by Job Tracker ?   1.5

(i) Explain Hbase.   1.5

(j) Can Reducers communicate with each other ? Give reason to support your answer.   1.5

Part B

2. (a) Explain Hadoop ecosystem in detail.   7

   (b) What are the steps followed while analysing big data ?   8

3. (a) Why is HDFS preferred over RDBMS ? 5

   (b) Explain the interface of Hadoop file system.   10

4. (a) Discuss the various steps involved in the MapReduce process.   7

   (b) What are the configuration parameters required to be specified in MapReduce ? 8

5. (a) Discuss the various types of Map-reduce. 7

   (b) What are the primary components of the Pig Latin language ?   8

6. (a) Discuss the various file systems available in Hadoop.   7

   (b) What do you understand by NoSQL ? How is it different from SQL ?   8

7. Can you provide a real-world example of when you would use Pig over other frameworks like MapReduce, and describe why it would be more beneficial ? Also, explain the architecture of Apache Pig.   15