

**B.Tech. 8<sup>th</sup> Semester**  
**(Computer Science And Information Technology)**  
**(G-Scheme), Examination, December-2022**  
**BIG DATA ANALYSIS**  
**Paper -PCC-CSE-404-G/PCC-IT-403-G**

*Time allowed : 3 hours]*

*[Maximum marks : 75*

*Note: Attempt five questions in all, selecting at least one from each section. Question no.1 is compulsory. All question carry equal marks.*

1. Explain the following : 6×2.5=15
- (a) Map Reduce
  - (b) Big Data Platform
  - (c) HDFS
  - (d) Data Lakes
  - (e) Six V in big data
  - (f) Mapper class

**Section-A**

2. Discuss big data and its characteristics. Explain application of Big Data. 15
3. Explain various steps used in Data Science Process. 15

**Section-B**

4. What is NoSQL? Explain various types of NoSQL data types. 15
5. Discuss various Big Data Processing Tools. 15

**Section-C**

6. What is Cluster? Explain the setting up Hadoop cluster. 15
7. Explain the following terms: 3×5=15
  - (i) Data Quality
  - (ii) Data Ingestion
  - (iii) Scalability

**Section-D**

8. Explain some analytical operations used in Big Data Processing pipelines. 15
9. Write in detail the concept of developing the Map Reduce Application. 15