

**Faculty of Science & Technology**

**Seventh Semester B.Tech. Computer Science & Engineering (IT) (C.B.C.S.) Examination**

**BIG DATA ANALYTICS**

**Prog. Ele.–V**

Time : Three Hours]

[Maximum Marks : 70

**INSTRUCTIONS TO CANDIDATES**

- (1) All questions carry marks as indicated.
- (2) Solve Question No. 1 **OR** Question No. 2.
- (3) Solve Question No. 3 **OR** Question No. 4.
- (4) Solve Question No. 5 **OR** Question No. 6.
- (5) Solve Question No. 7 **OR** Question No. 8.
- (6) Solve Question No. 9 **OR** Question No. 10.
- (7) Due credit will be given to neatness and adequate dimensions.
- (8) Assume suitable data wherever necessary.

1. (a) What is Big Data, and how does it differ from traditional data ? 7
- (b) Explain the challenges of handling unstructured data in Big Data analytics. 7

**OR**

2. (a) Explain how image recognition and computer vision technologies work in processing unstructured image data. 7
- (b) List and explain the five main characteristics of Big Data. 7
3. (a) How do privacy and compliance issues pose challenges to Big Data analytics ? 7
- (b) Why do organizations need Big Data analytics to gain a competitive edge ? 7

**OR**

4. (a) Explain the concept of clustering in data analysis. What are its applications ? 7
- (b) Provide an example of a real-world application of classification in analytics. 7
5. (a) How does Hadoop 3.0 handle resource allocation and management in a multi-tenant environment ? 7
- (b) Discuss the role and significance of Yet Another Resource Negotiator (YARN) in Hadoop 3.0. 7

**OR**

6. (a) How does Apache Oozie help in orchestrating and managing Hadoop workflows ? 7  
(b) Discuss the role of Apache Storm in real-time stream processing within Hadoop ecosystems. 7
7. (a) Describe the primary phases of data processing in a MapReduce job. 7  
(b) How does the Resource Manager allocate resources and manage jobs in YARN cluster ? 7

**OR**

8. (a) What is YARN (Yet Another Resource Negotiator), and why was it introduced in Hadoop 2.0 ? 7  
(b) What are the key advantages of using MapReduce for processing large datasets ? 7
9. (a) Explain the concept of tabular stores in NoSQL databases. 7  
(b) Discuss the role of JSON and BSON in document-oriented databases. 7

**OR**

10. (a) Explain the concept of schema-less models in NoSQL databases. 7  
(b) How does Hive sharding enhance data availability and performance in distributed systems ? 7