

No. of Printed Pages : 03

Roll No.

35412

Diploma in Data Science

EXAMINATION, 2025

(Second Semester)

PGDS-23-22

Big Data Analytic Using HADOOP

Time : 3 Hours]

[Maximum Marks : 80

Before answering the question-paper candidates should ensure that they have been supplied to correct and complete question-paper. No complaint, in this regard, will be entertained after the examination.

Note : Attempt *Five* questions in all, selecting *one* question from each Unit. Q. No. **1** is compulsory. All questions carry equal marks.

Unit I

1. (a) Explain the types of digital data with examples. **8**

- (b) Define Big Data. Describe the features and challenges of Big Data Analytics. **8**
- 2. (a) Describe the architecture and core components of Apache Hadoop. **8**
- (b) How can Linux/Unix tools be used to analyze data. **8**

Unit II

- 3. (a) Explain the concept and structure of Hadoop Distributed file system. **8**
- (b) What is Hadoop Archive (HAR) ? How is it used for data management ? **8**
- 4. (a) Explain serialization in Hadoop and how Avro is used for it. **8**
- (b) Explain data flow in a Hadoop system with diagram. **8**

Unit III

- 5. (a) Describe the execution process of a MapReduce job with the help of example. **8**

- (b) Explain the role of shuffle and sort in MapReduce. **8**
- 6. (a) What are the different job scheduling techniques in MapReduce ? **8**
- (b) Define PIG. Explain different Execution mode of PIG. **8**

Unit IV

- 7. (a) What is Hive ? How is it different from a traditional RDBMS ? **8**
- (b) Explain HiveQL with syntax for creating a table and querying it. **8**
- 8. (a) What is HBase ? Compare HBase with RDBMS. **8**
- (b) Explain the concept of Hive Metastore and its role in query execution. **8**

