

CBGS SCHEME

USN

| | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|
| | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|

17CS82

Eighth Semester B.E. Degree Examination, Dec.2023/Jan.2024

Big Data Analytics

Time: 3 hrs.

Max. Marks: 100

Note: Answer any FIVE full questions, choosing ONE full question from each module.

Module-1

- 1 a. What is HDFS? List all components of HDFS and explain any four components. (10 Marks)
b. Explain fault tolerance and speculative execution in hadoop framework. (10 Marks)

OR

- 2 a. Explain the importance of the following :
i) Name node high availability ii) HDFS checkpoints and backups (10 Marks)
b. Explain the following :
i) Terasort test ii) Managing Hadoop MapReduce Jobs (10 Marks)

Module-2

- 3 a. Explain the importance of Apache Sqoop data import and export methods. (10 Marks)
b. Explain YARN application framework with a neat diagram. (10 Marks)

OR

- 4 a. Explain how Apache Flume is used to acquire data streams. (10 Marks)
b. Explain different types of nodes defined in Oozie workflow. (10 Marks)

Module-3

- 5 a. How business intelligence plays a major role in decision making? Explain its types. (10 Marks)
b. What is data warehouse? Explain its architecture with a diagram. (10 Marks)

OR

- 6 a. Why should organization invest in business intelligence solutions? What are potential business intelligence problems? (10 Marks)
b. List and explain different methods used for data cleaning and preparation. (10 Marks)

Module-4

- 7 a. What is neural network? Explain design principles of an ANN. (08 Marks)
b. Data about height and weight for a few volunteers is available. Create a set of clusters for the following data to decide how many size of T-shirt should be ordered.

Data Sample :

| Height | Weight |
|--------|--------|
| 185 | 72 |
| 170 | 56 |
| 168 | 60 |
| 179 | 68 |
| 182 | 72 |
| 188 | 77 |
| 180 | 71 |
| 180 | 70 |
| 183 | 84 |
| 180 | 88 |
| 180 | 67 |
| 177 | 76 |

(12 Marks)

OR

- 8 a. What are association rules? Explain the working of Apriori algorithm. (08 Marks)
 b. Consider the following dataset :

| Student | Test Marks | Grade |
|---------|------------|-------|
| 1 | 95 | 85 |
| 2 | 85 | 95 |
| 3 | 80 | 70 |
| 4 | 70 | 65 |
| 5 | 60 | 70 |

- i) What linear regression equation best predicts statistics performance based on Math aptitude scores?
 ii) If a student made an 80 on aptitude test, what grade would we expect her to make in statistics? (12 Marks)

Module-5

- 9 a. Discuss the problems that can be addressed using text mining. (10 Marks)
 b. What is web mining? Explain its characteristics and three types of web mining. (10 Marks)

OR

- 10 a. Explain Naïve Baye's algorithm. List the advantages and disadvantages of Naïve Baye's algorithm. (10 Marks)
 b. What is SNA? How is it different from other data mining techniques? (10 Marks)
