B.Sc. (Data Science)
Discipline Specific Elective (DSE)
Semester V
BSDB33502T: Big Data Management

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 40%

Objectives

 To enable students to understand the basic concepts of big data Management basics, classification and machine learning. To know about Hadoop and Map-Reduce programming to tackle big data problems

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER

1. The syllabus prescribed should be strictly adhered to.

2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.

3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.

4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES
   Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section A

UNIT I: Big Data: Introduction, Data Storage and Analysis, Comparison with Other Systems, Rational Database Management System.

UNIT II: Big Data Management: challenges, benefits, Big Data Management Services and vendors.

UNIT III: Classification of Big- Data- Introduction**,** Architecture of Big Data**,** Structured Data, Unstructured Data, Semi Structured Data and their algorithms.

UNIT IV: Processing Big Data- Processing Big data through Acquisition, preprocessing and preparation.

Section B

UNIT V: Streaming Data- Streaming data, benefits of streaming data, examples of streaming data, Challenges in working with streaming data, Streaming of data Processing BigData with Apache SPARK.

UNIT VI: No-SQL Databases- Types of NoSQL databases, Advantages of NoSQL, Use of NoSQL, SQL vs NoSQL, Schema-less Models.

UNIT VII: No-SQL Tools- Introduction to MongoDB, its key features, Core Server tools, Creating and Querying through Indexes, Document-Oriented, principles of schema design, constructing queries on Databases.

UNIT VIII: Hbase – Analyzing big data with twitter – Big data for E-Commerce, Big data for blogs – Review of Basic Data Analytic Methods.

Suggested Readings
1. Mike Frampton, Mastering Apache Spark, Packt Publishing, 2015.
2. Mohammed Guller, Big Data Analytics with Spark, Apress, 2016