

**B.Sc. (Data Science)**  
**Skill Enhancement Course (SEC)**  
**Semester IV**  
**BSDB32404T: Introduction of Analytics and AI**

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

**Objective**

To provide an overview of an exciting growing field of big data analytics and introduce the tools required to manage and analyze big data like Hadoop, No-Sql Map-Reduce. This course will enable students to have skills that will help them to solve complex real-world problems of AI for decision support

**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 Analytics:** Overview and Evolution of Big data, characteristics Big Data and its importance, Drivers for Big data, Big data analytics, Role of Big data in Industry

**Unit II: Big Data Applications:** Characteristics of Big Data Applications, Perception and Quantification of Value, Understanding the Big Data Storage.

**Unit III: Map Reduce Framework:** General Overview of High-Performance Architecture, HDFS, Map-Reduce and YARN – Map Reduce Programming Model.

**Unit IV: Introduction to Hadoop:** Hadoop Architecture, Apache Hadoop Eco-System, Moving Data in and out of Hadoop, Understanding inputs and outputs of Map Reduce, shuffling, sorting, grouping, Data Serialization.

## Section B

**Unit V: Introduction to AI,** Problem Solving, Applications of AI using Big Data, Basic Concepts.

**Unit VI: Problem-solving through Search:** forward and backward, state-space, blind, heuristic, problem-reduction, neural, and evolutionary search algorithms.

**Unit VII: Machine Learning and Knowledge Acquisition:** learning from memorization, examples, explanation, and exploration. Learning nearest neighbor, naive Bayes, and decision tree classifiers.

**Unit VIII: Analytics and AI Strategy-for Business Transfer Re-engineering Business to think AI and Analytics , Robust Data Monetization Strategy , Accelerated Decision-making with Real-Time Analytics**

### Suggested Readings

1. Tom White, Hadoop, the Definitive guide, O'Reilly Media, 2009
2. Donald Miner, Map Reduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems, O'Reilly Media, 2012
3. Nathan Marz, Big Data: Principles and best practices of scalable real-time data systems, Manning Publications, 2015