



INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal - 500 043, Hyderabad, Telangana

COURSE CONTENT

BIG DATA-TOOLS AND TECHNIQUES								
VII Semester: IT / CSE (AI&ML) / CSE (CS) /								
Course Code	Category	Hours / Week			Credits	Maximum Marks		
AITD20	Elective	L	T	P	C	CIA	SEE	Total
		3	0	0	3	40	60	100
Contact Classes: 48	Tutorial Classes: Nil	Practical Classes: Nil			Total Classes: 48			
Prerequisite: Basics of databases, programming, and data analysis.								

I. COURSE OVERVIEW:

This course introduces the tools, techniques, and frameworks used in processing and analyzing Big Data. It provides hands-on experience with platforms like Hadoop, Spark, and NoSQL databases while addressing modern data processing techniques and their scalability.

II. COURSES OBJECTIVES:

The students will try to learn

- Explore the modern tools and techniques for handling Big Data.
- Understand the architecture and use of Big Data platforms like Hadoop and Spark.
- Gain familiarity with NoSQL databases for real-time data management.
- apply machine learning techniques in Big Data environments.

III. COURSE OUTCOMES:

After successful completion of the course, students should be able to:

- CO1 Understand and apply Big Data processing frameworks like Hadoop and Spark.
- CO2 Describe the advantages and challenges of NoSQL databases in Big Data
- CO3 Perform data cleaning, processing, and storage using modern tools.
- CO4 Implement scalable machine learning solutions for Big Data problems.
- CO5 Analyze case studies to understand the real-world implementation of Big Data tools
- CO6 Evaluate the performance and scalability of Big Data processing techniques.

IV. COURSE CONTENT

MODULE – I: INTRODUCTION TO BIG DATA (10)

Introduction to Big data: Characteristics of Data, Evolution of Big Data, Definition of Big Data, Challenges with Big Data, Applications of Big Data Across Industries, Traditional Business Intelligence (BI) versus Big Data, Overview of Big Data Tools and Techniques. Big data analytics: Classification of Analytics, Terminologies Used in Big Data Environments, The Big Data Technology Landscape.

MODULE – II: HADOOP ECOSYSTEM (09)

Introducing Hadoop, RDBMS versus Hadoop, Distributed Computing Challenges, History and overview of Hadoop, Hadoop Distributors, Processing Data with Hadoop, Interacting with Hadoop Ecosystem. HDFS (Hadoop Distributed File System): Design, architecture, and operations, Introduction to YARN for resource management.

MODULE – III: UNDERSTANDING MAP REDUCE FUNDAMENTALS (09)

Map Reduce Framework: Exploring the features of Map Reduce, Working of Map Reduce, Exploring Map and Reduce Functions, Techniques to optimize MapReduce jobs, Uses of MapReduce, Controlling MapReduce

Execution with Input Format, Reading Data with Custom Record Reader, -Reader, Writer, Combiner, Partitioners, MapReduce Phases, Developing simple MapReduce Application.

MODULE –IV: INTRODUCTION TO PIG , HIVE AND TOOLS FOR BIG DATA PROCESSING (10)

Introducing Pig: Pig architecture, Benefits, Installing Pig, Properties of Pig, Running Pig, and Getting started with Pig Latin, working with operators in Pig, Working with functions in Pig. Introducing Hive: Getting started with Hive, Hive Services, Data types in Hive, Built-in functions in Hive, Hive DDL

MODULE –V: BIG DATA USE CASES AND BEST PRACTICES (10)

Industry case studies: E-commerce, healthcare, IoT, and finance, Security, privacy, and ethical challenges in Big Data. Performance optimization techniques for Big Data platforms, Future trends in Big Data: Cloud integration and edge analytics.

V. TEXT BOOKS:

1. Tom White, Hadoop: The Definitive Guide, O'Reilly Media.
2. Alex Holmes, Hadoop in Practice, Manning Publications.

VI. REFERENCE BOOKS:

1. Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman, Mining of Massive Datasets.
2. Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia, Learning Spark: Lightning-Fast Big Data Analysis.
3. Seema Acharya and Subhasini Chellappan, Big Data and Analytics.

VII. WEB REFERENCES:

1. <https://in.indeed.com/career-advice/career-development/big-data-tools>
2. <https://neurosys.com/blog/big-data-tools>.
3. <https://github.com/haifengl/bigdata>.

VIII. MATERIALS ONLINE

1. Course template
2. Tutorial question bank
3. Tech-talk topics
4. Open-ended experiments
5. Definitions and terminology
6. Assignments
7. Model question paper – I
8. Model question paper – II
9. Lecture notes
10. PowerPoint presentation
11. E-Learning Readiness Videos (ELRV)