



# INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal - 500 043, Hyderabad, Telangana

## COURSE CONTENT

BIG DATA AND ANALYTICS								
<b>V Semester: CSE / IT</b>								
Course Code	Category	Hours / Week			Credits	Maximum Marks		
		L	T	P		C	CIA	SEE
ACSE22	Elective	3	-	-	3	40	60	100
<b>Contact Classes: 48</b>	<b>Tutorial Classes: Nil</b>	<b>Practical Classes: Nil</b>			<b>Total Classes: 48</b>			
<b>Prerequisite: Fundamentals of Computer Networks and Operating Systems</b>								

### I. COURSE OVERVIEW:

This course emphasizes on complex process of examining streaming big data to uncover information such as hidden patterns, correlations, market trends and customer preferences that can transform data into impactful business solutions. It covers analytics, predictive models and different types of tools at storing, processing and monitoring levels such as HDFS, HIVE, PIG and Zookeeper etc. used in cloud applications, mobile applications, social media content, and Internet of Things. This course will help to develop practical skills in building interactive systems, applying machine learning and data analytics.

### II. COURSES OBJECTIVES:

The students will try to learn:

- I. Optimize business decisions and create competitive advantage with big data analytic.
- II. The several key big data technologies used for storage, analysis and manipulation of data.
- III. The key concepts of Hadoop framework and map reduce.
- IV. The concepts of PIG and HIVE for application development.

### III. COURSE OUTCOMES:

**At the end of the course students should be able to:**

- CO1 Explain visual analytics concepts, visualization techniques, and design principles.
- CO2 Analyse and interpret complex datasets using appropriate visual representations.
- CO3 Design effective visualizations considering human perception and usability.
- CO4 Apply interactive visualization techniques for exploratory and explanatory analysis.
- CO5 Evaluate visual analytics solutions for real-world data-driven applications.

#### **IV. COURSE CONTENT:**

##### **MODULE 1: INTRODUCTION TO BIG DATA (09)**

Introduction to Big data: Characteristics of Data, Evolution of Big Data, Definition of Big Data, Challenges with Big Data, Traditional Business Intelligence (BI) versus Big Data.

Big data analytics: Classification of Analytics, Importance and challenges facing big data, Terminologies Used in Big Data Environments, The Big Data Technology Landscape.

##### **MODULE 2: INTRODUCTION TO HADOOP (09)**

Introducing Hadoop, RDBMS versus Hadoop, Distributed Computing Challenges, History and overview of Hadoop, Use Case of Hadoop, Hadoop Distributors, Processing Data with Hadoop, Interacting with Hadoop Ecosystem.

##### **MODULE 3: THE HADOOP DISTRIBUTED FILESYSTEM (09)**

Hadoop Distributed File System (HDFS): The Design of HDFS, HDFS Concepts, Basic Filesystem Operations, HadoopFilesystems.

The Java Interface- Reading Data from a Hadoop URL, Reading Data Using the File system API, Writing Data. Data Flow- Anatomy of a File Read, Anatomy of a File Write, Limitations.

##### **MODULE 4: UNDERSTANDING MAP REDUCE FUNDAMENTALS (09)**

Map Reduce Framework: Exploring the features of Map Reduce, Working of MapReduce, 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 5: INTRODUCTION TO PIG AND HIVE (09)**

Introducing Pig: Pig architecture, Benefits, Installing Pig, Properties of Pig, Running Pig, 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.

#### **V. TEXT BOOKS:**

1. SeemaAcharya, SubhashiniChellappan, Big Data and Analytics, Wiley Publications, 2<sup>nd</sup> Edition, 2014, DT Editorial Services, Big Data, Dream Tech Press, 2<sup>nd</sup> Edition, 2015.
2. Tom White, Hadoop: The Definitive Guide, O 'Reilly, 3<sup>rd</sup> Edition, 2012.
3. Black Book Big Data, Dream tech publications, 1<sup>st</sup> Edition, 2017.

#### **VI. REFERENCE BOOKS:**

1. Michael Minelli, Michele Chambers, AmbigaDhiraj, Big Data, Big Analytics:Emerging Business
2. Intelligence and Analytic Trends for Today 's Business, Wiley CIO Series, 1<sup>st</sup> Edition, 2013.
3. Rajiv Sabherwal, Irma Becerra- Fernandez, "Business Intelligence Practice, Technologies and Management", John Wiley, 1<sup>st</sup> Edition, 2011.
4. ArvindSathi, "Big Data Analytics: Disruptive Technologies for Changing the Game", IBM Corporation, 1<sup>st</sup> Edition, 2012.

#### **VII. ELECTRONIC RESOURCES:**

1. [https://www.sas.com/en\\_us/insights/analytics/big.data.analytics.html](https://www.sas.com/en_us/insights/analytics/big.data.analytics.html)
2. <https://www.searchbusinessanalytics.techtarget.com/definition/big.data.analytics>
3. <https://www.webopedia.com>

#### **VIII. MATERIALS ONLINE**

- 1 Course Outline Description
- 2 Lecture notes
- 3 PowerPoint presentation
- 4 Definitions and Terminology
- 5 Tutorial Question Bank
- 6 Case Studies
- 7 Real life Examples
- 8 Complex Engineering Problems
- 9 Tech Talk Topics
- 10 Concept Video Topics
- 11 Open-ended Exercises
- 12 Assignments
- 13 Model Question Paper – I

- 14 Model Question Paper – II
- 15 GATE Question Bank
- 16 Previous Question Papers and Solutions