INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal - 500 043, Hyderabad, Telangana

COURSE CONTENT

BIG DATA AND BUSINESS ANALYTICS								
VII Semester:								
Course Code	**Category**	**Hours / Week**			**Credits**	**Maximum Marks**		
		L	**T**	**P**	**C**	**CIA**	**SEE**	**Total**
AITC24	**Core**	3	1	0	4	30	70	100
Contact Classes: 45	**Tutorial Classes: 15**	**Practical Classes: Nil**				**Total Classes:60**		
Prerequisites: **Data base management systems**								

I. COURSE OVERVIEW:

Fundamental principles of Big Data Analytics and its role in making better decisions and predictions in the organization. The course also covers the Technology, Infrastructure and Applications of Big Data. Concepts of data identification, data cleansing and integration are also addressed. Software requirements of Big Data are addressed and case studies of Big Data Applications are discussed.

II. COURSE OBJECTIVES:

The students will try to learn:

I. The concepts and challenges of Big Data comparing with traditional business intelligence.
II. The data processing methods with the help of MapReduce abstraction techniques.
III. The applications for handling huge volume of data using Pig Latin.
IV. The use of data analytics using Hadoop component HIVE with its built-in functions and services like DDL.

III. COURSE SYLLABUS:

MODULE – I: BIG DATA ANALYTICS (08)

Introduction to Big Data: Data, characteristics of data and types of digital data: unstructured, semi-structured and structured, sources of data, working with unstructured data, evolution and definition of big data, characteristics and need of big data, challenges of big data, data environment versus big data environment

MODULE – II: INTRODUCTION TO HADOOP (10)

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; Hadoop Distributed File System(HDFS): The Design of HDFS, HDFS concepts, basic file system operations, Hadoop file systems; The Java Interface, reading data from a Hadoop URL, reading data using the files system API, writing data, data flow, anatomy of a file read, anatomy of a file write, limitations.

MODULE – III: BIG DATA TECHNOLOGIES AND DATABASES (08)

Introduction to NoSQL, uses, features and types, need, advantages, disadvantages and application of NoSQL, overview of NewSQL, comparing SQL, NoSQL and NewSQL, introduction to MongoDB and its needs, characteristics of MongoDB.

Introduction of Apache Cassandra, features, CQL, CRUD operations, collections, counter, alter, import and export.

MODULE – IV: UNDERSTANDING MAP REDUCE FUNDAMENTALS (10)

MapReduce Framework: Exploring the features of MapReduce, working of MapReduce, exploring Map and Reduce functions, techniques to optimize MapReduce jobs, uses of Map Reduce; Controlling MapReduce execution with Input Format, reading data with custom Record Reader, reader, writer, combiner, partitioners, MapReduce phases, developing simple Map Reduce application.

MODULE – V: INTRODUCTION TO PIG and HIVE (09)

Introducing Pig: Pig architecture, benefits, installing Pig, properties of Pig, running Pig, Pig Latin, operators in Pig, working with functions in Pig; Introducing Hive, Hive services, datatypes in Hive, built-in functions in Hive, Hive DDL.

IV. TEXT BOOKS:

1. Seema Acharya, Subhashini Chellappan, "Big Data and Analytics", Wiley Publications, 2nd Edition, 2014.
2. Tom White, "Hadoop: The Definitive Guide", O'Reilly, 3rd Edition, 2012.

V. REFERENCE BOOKS:

1. Michael Minelli, Michele Chambers, Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business", Wiley CIO Series, 1st Edition, 2013.
2. Rajiv Sabherwal, Irma Becerra, Fernandez, "Business Intelligence Practice, Technologies and Management", John Wiley, 1st Edition, 2011.
3. Arvind Sathi,"Big Data Analytics: Disruptive Technologies for Changing the Game", IBM Corporation, 1st Edition, 2012.

VI. WEB REFERENCES:

1. 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