BIG DATA AND BUSINESS ANALYTICS

VII Semester: CSE / IT

Course Code	Category	Hours / Week			Credits	Maximum Marks		
		L	T	P	C	CIA	SEE	Total
ACS012	Core	3	1	-	4	30	70	100
Contact Classes: 45	Tutorial Classes: 15	Practical Classes: Nil				Total Classes: 60		

I. COURSE OVERVIEW:

This course provides a clear understanding on concepts of sources of big data, characteristics, storing and processing components, and analytics applications. This course emphasizes on potential impact of big data challenges, open research issues, and various tools associated with it. This course includes the introduction and processing big data with an overview of Hadoop technology and its components such as pig, hive, etc.

II. COURSE OBJECTIVES:

The course should enable the students to:

I	The scope and essentiality of Big Data and Business Analytics.
II	The technologies used to store, manage, and analyze big data in a Hadoopecosystem.
III	The techniques and principles in big data analytics with scalability and streamingcapability.
IV	The hypothesis on the optimized business decisions in solving complex real-worldproblems.

III. COURSE OUTCOMES:

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

CO 1	**Explain** the evolution of big data and big data analytics along with its characteristics and challenges included in traditional business intelligence.	Understand
CO 2	**Make use of** appropriate components for processing, scheduling and knowledge extraction from large volumes the applications for handling huge volume of data	Apply
CO 3	**Develop** a Map Reduce application for optimizing the jobs.	Apply
CO 4	**Develop** the applications for handling huge volume of data usingPig Latin.	Apply
CO 5	**Explain** the importance of big data framework HIVE and its built-infunctions, data types and services like DDL in Hadoop distributed file system.	Understand
CO 6	**Extend** the big data technologies used to process and querying the big data in Hadoop, Map Reduce, Pig and Hive.	Analyze

IV. SYLLABUS:

UNIT-I	INTRODUCTION TO BIG DATA	Classes: 08

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.

UNIT-II	INTRODUCTION TO HADOOP	Classes: 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

UNIT-III	THE HADOOP DISTRIBUTED FILESYSTEM	Classes: 09

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

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

UNIT-IV	UNDERSTANDING MAP REDUCE FUNDAMENTALS	Classes: 09

Map Reduce Framework: Exploring the features of Map Reduce, Working of Map Reduce, Exploring Map and Reduce Functions, Techniques to optimize Map Reduce jobs, Uses of Map Reduce.
Controlling MapReduce Execution with InputFormat, Reading Data with custom RecordReader,-Reader, Writer, Combiner, Partitioners, Map Reduce Phases, Developing simple MapReduce Application.

UNIT-V	INTRODUCTION TO PIG and HIVE	Classes: 10

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.

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.

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.

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

E-Text Books:

1. https://www.books.google.co.in/books?id=rkWPojgfeM8C&printsec=frontcover&dq=HIGH+PERFORMANCE+COMPUTING.
2. http://www.datameer.com/pdf/big-data-analytics-ebook.pdf?mkt_tok.

Course Home Page: