

BIG DATA AND BUSINESS ANALYTICS

VII Semester: CSE/IT								
Course Code	Category	Hours / Week			Credits	Maximum Marks		
ACS012	Core	L	T	P	C	CIA	SEE	Total
		3	1	-	4	30	70	100
Contact Classes: 45		Tutorial Classes: 15		Practical Classes: Nil			Total Classes: 60	
<p>COURSE OBJECTIVES: The course should enable the students to:</p> <ol style="list-style-type: none"> I. Optimize business decisions and create competitive advantage with Big data analytics II. Understand several key big data technologies used for storage, analysis and manipulation of data. III. Recognize the key concepts of Hadoop framework, map reduce. IV. Demonstrate the concepts in Hadoop for application development. <p>COURSE OBJECTIVES:</p> <ol style="list-style-type: none"> 1. Understand the key issues in big data analytics and its associated applications in business analytics. 2. Illustrate different types of big data technologies in Hadoop parallel world. 3. Interpret disparate data storing in Hadoop Distributed File Systems (HDFS). 4. Explore map reduce framework and optimize its jobs. 5. Explain the basic methodologies of pig and hive. <p>COURSE LEARNING OUTCOMES (CLOs):</p> <ol style="list-style-type: none"> 1. Understand what Big Data, importance and various sources of data. 2. Describe the elements of big data-volume, variety, velocity and veracity. 3. Understand the importance and challenges of big data. 4. Define big data analytics advantages and its applications. 5. Define distributed and parallel computing for big data. 6. Analyze the core components of hadoop with basic commands. 7. Explain the key features of hadoop in processing big data.. 8. Understand hadoop ecosystem with its animal planet. 9. Explain the basic terminology of Hadoop Distributed File Systems (HDFS). 10. Describe in detail about Distributed file system. 11. Understand the concept of Hadoop cluster architecture. 12. Explain a file in HDFS and represent the anatomy of file read and write. 13. Understand Map Reduce and its qualities and retain advanced Map Reduce thoughts 14. Understand the architecture of Map Reduce framework. 15. Demonstrate the techniques to optimize Map Reduce jobs. 16. Understand the typical use occasions of input and output forms of Map Reduce. 17. Demonstrate an ability to use frameworks like pig and hive to process Big Data and Analytics. 18. Design the architecture of pig with its data types and operations. 19. Explain the architecture of hive with different operations. 20. Design and implement different technologies for processing big data in pig and hive. 								
UNIT-I	INTRODUCTION TO BIG DATA						Classes: 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.								

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: 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.		
Text Books:		
<ol style="list-style-type: none"> 1. Seema Acharya, Subhashini Chellappan, -Big Data and Analytics, Wiley Publications, 2nd Edition,2014DT Editorial Services, -Big Data, Dream Tech Press, 2nd Edition, 2015. 2. Tom White, -Hadoop: The Definitive Guide, O'Reilly, 3rd Edition, 2012. 3. Black Book Big Data, dreamtech publications , 1st Edition, 2017. 		
Reference Books:		
<ol style="list-style-type: none"> 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:		
<ol style="list-style-type: none"> 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:		
<ol style="list-style-type: none"> 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. 		