18CSPE702	BIG DATA ANALYTICS	L	T	P	C
		3	0	0	3

Course Objectives:

1.	To introduce basic concepts and challenges of big data (3 V's: volume, velocity, and variety) and methodologies for analyzing structured and unstructured data.
2.	To impart basic concepts about Big Data Environment and Big Data Technology Landscape.
3.	To understand the importance of Hadoop Ecosystem.
4.	To know about Pig, Hive, MongoDB, NoSQL and Cassandra.
5.	To know about Jasper Report using Jasper Soft Studio.

UNIT I	DIGITAL DATA AND BIG DATA	9	+	0

Types of Digital Data- Structured, Semi-structured, Unstructured, Characteristics of data Definition of big Data-Challenges of big data-Why big data-Traditional BI vs. Big data-A typical BI environment- A Big data environment-What is changing in the realms of big data?

UNIT II	BIG DATA ANALYTICS AND TECHNOLOGY LANDSCAPE	9	+	0

Big Data Analytics - Classification of Analytics- Top challenges facing Big Data- Importance of Big Data Analytics- Greatest challenges that prevent businesses from capitalizing on Big Data - kind of technologies to help meet the challenges posed by Big Data- Data Science- Terminologies used in Big Data Environment- Big Data Technology Landscape- NoSQL- Types of NoSQL Databases- Why NoSQL- Advantages of NoSQL - SQL Vs. NoSQL - NewSQL - Comparison of SQL, NoSQL and NewSQL.

UNIT III	HADOOP	9	+	0

Hadoop Overview - Hadoop Components - High Level Architecture of Hadoop - Features of Hadoop- Key advantages of Hadoop- Versions of Hadoop- Hadoop 1.0, Hadoop 2.0- Overview of Hadoop Ecosystems- Interacting with Hadoop EcoSystem- Hive, Pig, HBASE, Sqoop - Hadoop Vs. SQL- High Level Architecture of Hadoop - Hadoop Distributed File System- HDFS Daemons- Special Features of Hadoop- Processing Data With Hadoop- MapReduce Daemons-How Map Reduce Works- Map Reduce Example- Limitation of Hadoop 1.0 - Hadoop 2: HDFS - Hadoop 2: YARN

UNIT IV	NoSQL – MongoDB, CASSANDRA, HIVE , PIG	9	+	0

Introduction to MongoDB- Why MongoDB? –Terms used in RDBMS and MongoDB-Data Types in MongoDB – MongoDB Query Language. Apache Cassandra- An Introduction- Features of Cassandra-CAL Data Types- CQLSH- Keyspaces- CRUD- Collections- Using a Counter- Time to Live(TTL)alter commands- Import and Export-Querying System Tables-Introduction to Hive- Hive Architecture- Hive Data Types- Hive File Format- Hive Query Language (HQL)Introduction to Pig- Key Features of Pig- Anatomy of Pig- Pig on Hadoop – Pig Philosophy-Use Case for Pig- ETL Processing – Pig Latin Overview – Data Types in Pig - Running Pig - Execution Modes of Pig – HDFS Commands- Relational Operators- Eval Functiuon- Complex Data Types - Pig Versus Hive.

UNIT V	JASPER REPORT USING JASPER SOFT STUDIO	9	+	0

Introduction to Jasper Report using Jasper Soft Studio - Reporting using MongoDB - Reporting using Cassandra. Introduction to MAPREDUCE Programming- Mapper - Reducer - Combiner - Partitioner - Searching – Sorting - Compression. Introduction to Machine Learning- Machine Learning Algorithms.

Total (L+T)= 45 Periods

Course Outcomes:

Upon completion of this course, the students will be able to:

CO1	:	Understand the types of digital data and challenges with big data.
CO2	:	Know about the Big data Landscape Technology.
CO3	:	Understand the common Hadoop ecosystem components, Hadoop Architecture, HDFS, Hadoop MapReduce framework and the working of MapReduce on data stored in HDFS.
CO4	:	Learn the concepts of Pig, Hive, MongoDB, NoSQL and Cassandra
CO5	:	Produce Report using Jasper Soft studio.

Text Books:

1.	Seema Acharya, Subhashini Chellappan, “Big Data And Analytics”, Willey ,2015.

Reference Books:

1.	David Loshin,” Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools,Techniques, NoSQL, and Graph”, Morgan Kaufmann Publishers,2013.
2.	Wen-Chen Hu and Naima Kaabouch (eds) ,” Big Data Management, Technologies, and Applications “, IGI Global,2013.
3.	Tom White, “Hadoop: The Definitive Guide”, O`Reilly Publishers, USA, 2012.
4.	Michael Berthold, David J. Hand, “Intelligent Data Analysis”, Springer, 2007