Course Code	**40119206**			
Category	**Core Subject**			
Course Title	**Big Data Tools (BDT)**			
Scheme and Credits	**Theory**	**Tutorial**	**Lab**	**Credits**
	1	**0**	**4**	**3**
Pre-requisites (if any)	Basic knowledge of programming language and database concepts.			

1. Course Objectives:

1	To understand basics of Big Data.
2	Perform critical analysis of Big Data applications using special purpose tools and software like MongoDB.
3	To understand the big data frameworks like Hadoop.
4	To gain knowledge on Hadoop related tools such as Pig for big data analytics.

2. Course Contents

Unit	Course Content	Weightage
Unit I	**Introduction to Big Data and MongoDB:** **Big Data:** Introduction to Big Data and Analytics, Classification of digital Data, Structured and Unstructured Data - Introduction to Big Data. Why Big Data Traditional Business Intelligence versus BigData **Introduction to MongoDB:** Introduction: What is MongoDB? Why MongoDB? Mongo shell basic commands, Install MongoDB on Windows, MongoDB create database, Add MongoDB array using insert(), MongoDB primary key , MongoDB query document(using find() method with examples MongoDB sort() & limit() , skip(), MongoDB count() & remove() functions, MongoDB update() document ,MongoDB	**30%**

	regular expression, MongoDB Vs. SQL	
Unit II	**Unit 2: Introduction to Technology NoSQL and MapReduce:** **NoSQL:** What is a NoSQL Database? Brief history of NoSQL databases, NoSQL database features, Types of NoSQL database (Document databases, Key-value databases, Wide-column stores, and Graph databases), Difference between RDBMS and NoSQL, Why NoSQL? When should NoSQL be used? NoSQLdatabase misconceptions. **MapReduce:** What is Map Reduce programming? How does Map Reduce works? Map Reduce Word Count example. About Map Reduce ,Understanding block and input splits, MapReduce data types ,Understanding Writable ,Data Flow in MapReduce Application , Understanding MapReduce problem on datasets , MapReduce and functional programming , Writing MapReduce application , Understanding Mapper function , Understanding Reducer Function , Usage of Combiner	**20%**
Unit III	**Unit 3: HDFS(Hadoop Distributed File System)** **Hadoop:** Introducing Hadoop, File System - Concepts Blocks, Replication Factor, Version File ,Safe mode, Namespace IDs ,Purpose of Name Node , Purpose of Data Node, Purpose of Secondary Name Node, Purpose of Job Tracker , Purpose of Task Tracker , HDFS shell commands – copy, delete, create directories etc. , Reading and Writing in HDFS , Difference of Unix Commands and HDFS commands , Hadoop Admin Commands , Hands on exercise with Unix and HDFS commands	**20%**
Unit IV	**Unit 4: Hadoop Eco System** **Pig:** Introduction to PIG, Execution modes of Pig, Comparison of Pig with databases, Grunt, Pig Latin, , load and store , group and joining, combining and splitting, filtering, sorting, built in function, Data processing operators. **Hive:** Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, Querying Data and User Defined Functions.	**30%**

1. **Desirable:**

2. **Text Book(s):**

1) Seema Acharya, Subhashini Chellappan, " Big Data and Analytics", WileyIndia Pvt. Ltd.,2015
2) Matei Zaharia, Patrick Wendell, Andy Konwinski, Holden Karau, "LearningSpark", O'Reilly Media,2015
3) Zachary Radtka and Donald Miner, "Hadoop with Python'', O'ReillyMedia,2016

3. **Reference Books:**

- Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer, 2007.
- Jay Liebowitz, "Big Data and Business Analytics" Auerbach Publications, CRC press (2013)
- Tom Plunkett, Mark Hornick, "Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle R Enterprise and Oracle R Connector for Hadoop", McGraw-Hill/Osborne Media (2013), Oracle press.
- Anand Rajaraman and Jef rey David Ulman, "Mining of Massive Datasets", Cambridge University Press, 2012.
- Bill Franks, "Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streamswith Advanced Analytics", John Wiley & sons, 2012.
- Glen J. Myat, "Making Sense of Data", John Wiley & Sons, 2007
- Paul Zikopoulos ,Dirk DeRoos , Krishnan Parasuraman , Thomas Deutsch , James Giles , David Corigan , "Harness the Power of Big Data The IBM Big Data Platform ", Tata McGrawHill Publications, 2012.

4. **Web Resources:**

1) http://www.mongodb.com
2) http://hadoop.apache.org/
3) https://www.ibm.com/cloud/learn/nosql-databases
4) https://www.coursera.org/lecture/nosql-databases/introduction-to-nosql-VdRNp
5) https://www.geeksforgeeks.org/introduction-to-nosql/
6) https://www.javatpoint.com/nosql-database

5. **Accomplishment of the student after completing the course:**

After completion of the course, students should become capable of to understand the concepts, technology and usage of Big Data