Course Name: Big Data Analytics
Course Code: TE7552
Faculty: Engineering
Course Credit: 3
Course Level: 4
Sub-Committee (Specialization): Artificial Intelligence and Machine Learning
Learning Objectives:

Students will be able to1. To optimize business decisions and create competitive advantage with Big Data analytics
2. To explore the fundamental concepts of big data analytics.
3. To learn to analyze the big data using intelligent techniques.
4. To understand the various search methods and visualization techniques.
5. To learn to use various techniques for mining data stream.
6. To understand the applications using Map Reduce Concepts.
7. To introduce programming tools PIG & HIVE in Hadoop echo system

Books Recommended:

Book	Author	Publisher
Big Data Analytics with R and Haoop	Vignesh Prajapati	Packet Publishing 2013
Big Data and Business analytics	JyLiebowitz	CRC press, 2013
HADOOP: The definitive Guide	Tom White	O Reilly 2012
Oracle Big Data Handbook	Tom Plunkett, Brian Macdonald et al	Oracle Press, 2014
Professional Hadoop Solutions	1. Boris lublinsky, Kevin t. Smith, Alexey Yakubovich	Wiley, ISBN: 9788126551071, 2015
Understanding Big data	Chris Eaton, Dirk deroos et al	McGraw Hill, 2012

Course Outline:

Sr. No.	Topic	Actual Teaching Hours	Contact Hours Equivalence
1	**Introduction to Big Data:Big Data Fundamentals and Big Data Analytics. Structured Data, unstructured Data and semi Structured Data. Introduction of Big Data and Hadoop Overview and Evolution of Big-Data Hadoop, Architecture/Framework, HDFS Architecture/Framework, Map reduce, Hadoop Environment Setup, Distributed File System(s)**	**6**	**6**
2	**Big Data Analytics and Big Data Analytics Techniques:Big Data and its Importance, Drivers for Big data, Optimization techniques, Dimensionality Reduction techniques, Time series Forecasting, Social Media Mining and Social Network Analysis and its Application, Big Data analysis using Hadoop, Pig, Hive, Mongodb, Spark and Mahout, Data analysis techniques like Discriminant Analysis and Cluster Analysis, Introduction to NOSQL (Neo4j) and MongoDB, Hive Architecture, HBase concepts, PIG, Zookeeper - how it helps in monitoring a cluster, HBase uses Zookeeper and how to Build Applications with Zookeeper, No SQL databases: Cassandra and HBase (columnar), MongoDB and Elastic Search (document-based), Neo4j (graph based)**	**12**	**12**

3	**Hadoop Architecture, Hadoop StorageHDFS, Common Hadoop Shell commands, Anatomy of File Write and Read., NameNode, Secondary NameNode, and DataNode, Hadoop MapReduce paradigm, Algorithms using Map Reduce, Understanding inputs and outputs of MapReduce, Map and Reduce tasks, Job, Task trackers ,Cluster Setup, SSH and Hadoop Configuration, HDFS Administering ,Monitoring and Maintenance Moving Data in and out of Hadoop, Data Serialization**	**10**	**10**
4	**Big Data and High Dimensional Data AnalysisIntroduction to Spark, Framework and comparisons between Spark and Hadoop Frameworks. Apache Spark (using Scala, Java, Python), Mining streaming data,Apache Kafka, Spark MLlib, Infrastructure for Big Data, Big Data Management and Frameworks. Big Data Search, Big Data as a Service.**	**10**	**10**
5	**Big Data Analytics Applications/UsecasesAnd Visualization of Big DataBig Data Analytics in E-Governance & Society, Applications in Science, Engineering, Healthcare, Visualization, Business etc. Case Study of Existing Big Data Analytics Systems.Big Data visualization with the tools like D3, Kibana, and Grafana, Scala and Python for Data Visualization**	**7**	**7**
	Total	**45**	**45**

Pre Requisites:

Data mining fundamentals

Evaluation:

A) Continuous Assessment (30 marks)

1. Essential

a) Quizzes b) Assignments c) Tests

Pedagogy:

1. Classroom teaching
2. Hands on Lab exercises
3. Case studies
4. Project-based learning

Expert:

Dr. Shraddha Phansalkar,HOD, CS/IT department,SIT