SWE2011	Big Data Analytics	L	T	P	J	C
		3	0	0	4	4
Pre-requisite	**SWE1004**	**Syllabus version**				
		v.1.0				

Course Objectives:

1. To introduce fundamental concepts of big data analytics.
2. To elucidate different data learning techniques.
3. To explore various data analytic and visualization tools.

Expected Course Outcome:

1. Understand characteristics and sources of big data.
2. Recognise of various data analytical techniques and approaches for handling big data.
3. Apply data analytic methodologies in streaming data.
4. Familiar with diverse learning models and clustering techniques.
5. Use visualization techniques and tools in big data analystics
6. Campare the different types of frameworks and tools for big data analytics
7. Analyze Big Data in various forums like Social Networks, e-Commerce etc
8. Illustrate the phases of Big Data Analytics with the help of Data Sets from various domains and presenting the results.

Student Learning Outcomes (SLO)	**2, 4,14, 17**

2.Having a clear understanding of the subject related concepts and of contemporary issues
4.Having Sense-Making Skills of creating unique insights in what is being seen or observed (Higher level thinking skills which cannot be codified)
14.Having an ability to design and conduct experiments, as well as to analyze and interpret data
17.Having an ability to use techniques, skills and modern engineering tools necessary for engineering practice

Module:1	**Introduction to Big Data**	**7 hours**

Analytics – Nuances of big data – Value – Issues – Case for Big data – Big data options Team challenge – Big data sources – Acquisition – Nuts and Bolts of Big data. Features of Big Data - Security, Compliance, auditing and protection - Evolution of Big data – Best Practices for Big data Analytics - Big data characteristics - Volume, Veracity, Velocity, Variety

Module:2	**Data Analysis and Approaches**	**7 Hours**

Evolution of analytic scalability – Convergence – parallel processing systems – analytic data sets – Analytic methods - Analysis approaches – Statistical significance – business approaches – Analytic innovation – Traditional approaches – Iterative

Module:3	**Stream Data Mining**	**5 hours**

Introduction to Streams Concepts – Stream data model and architecture - Stream Computing, Sampling data in a stream – Filtering streams – Counting distinct elements in a stream – Estimating moments – Counting oneness in a window – Decaying window – Real time Analytics Platform(RTAP) applications.

Module:4	**Predictive Analytics**	**8 hours**

Predictive Analytics – Supervised – Unsupervised learning – Neural networks – Kohonen models – Normal – Deviations from normal patterns – Normal behaviors – Expert options – Variable entry - Mining Frequent itemsets - Market based model – Apriori Algorithm – Handling large data sets in Main memory – Limited Pass algorithm – Counting frequent itemsets in a stream – Clustering Techniques – Hierarchical – K- Means.		
Module:5	**Visualizations**	**5 hours**
Clustering high dimensional data Visualizations - Visual data analysis techniques, interaction techniques; Systems and applications.		
Module:6	**Framework for implementation**	**6 hours**
Map Reduce Framework - Hadoop – Hive - – Sharding – NoSQL Databases - S3 -Hadoop Distributed file systems – Hbase – Impala.		
Module:7	**Big Data for E-Commerce**	**5 hours**
Analyzing big data with twitter – Big data for E-commerce – Big data for blogs.		
Module:8	**Contemporary issues:** Applications of Big Data Analytics in Industry	**2 hours**
	Total Lecture hours:	**45 hours**

Text Book(s)

1. Vignesh Prajapati, Big data analytics with R and Hadoop, SPD 2013.

Reference Books

1. Tom White, "Hadoop: The Definitive Guide", Third Edition, O'Reilley, 2012.
2. Eric Sammer, "Hadoop Operations", O'Reilley, 2012.
3. E. Capriolo, D. Wampler, and J. Rutherglen, "Programming Hive", O'Reilley, 2012.
4. Lars George, "HBase: The Definitive Guide", O'Reilley, 2011.

Recommended by Board of Studies	5-3-2016		
Approved by Academic Council	No. 40th	Date	18-3-2016