ITA6008	Big Data Analytics	L	T	P	J	C
		3	0	0	4	4
Pre-requisite	**ITA5008**	**Syllabus version**				
		v. 1.0				

Course Objectives:

1. To understand the big data platform and its use cases.
2. To impart knowledge in applying skills and tools to manage and analyze the big data.
3. To apply analytics on structured and unstructured data.

Expected Course Outcomes:

1. Demonstrate knowledge of the fundamental elements and concepts related to big data.
2. Analyze the core architectural concepts to meet the challenges in implementing big data systems.
3. Design and develop a Big Data Environment according to the benchmarks
4. Setup a Big Data Environment and implement security techniques.
5. Evaluate the use of data through cleansing, warehousing, analytics, and visualization to the ultimate business decision.
6. Analyze the data using various statistical methods.
7. Develop applications using large scale analytics tools to solve open big data problems

Student Learning Outcomes (SLO)	**2,7,14**

Module:1	**Introduction to Big Data Analytics**	**6 hours**

Big Data Overview, State of practice in analytics, Role of Data Scientists, Examples of Big Data Analytics, Data Analytics Lifecycle

Module:2	**Introduction to Big Data Analytics**	**6 hours**

Components of Hadoop, Analyzing Big data with Hadoop, Design of HDFS, Developing a Map reduce Application

Module:3	**Map Reduce**	**6 hours**

Distributed File System(DFS), Map Reduce, Algorithms using Map Reduce, Communication cost Model, Graph Model for Map Reduce Problem

Module:4	**Hadoop Environment**	**7 hours**

Setting up a Hadoop Cluster, Hadoop Configuration, Security in Hadoop, Administering Hadoop, Hadoop Benchmarks, Hadoop in the cloud.

Module:5	**Big Data Analytics Methods using R**	**6 hours**

Introduction to R-Attributes, R Graphical user interfaces, Data import and export, attribute and Data Types, Descriptive Statistics, Exploratory Data Analysis.

Module:6	**Statistical methods for evaluation**	**6 hours**

Hypothesis Testing, Difference of Means, Wilcoxon Rank-Sum Test, Type I and Type II errors, power and sample size, ANOVA

Module:7	**Advanced Analytics - technologies and tools**	**6 hours**
Analytics for unstructured data, The Hadoop ecosystem – pig – Hive- HBase- Mahout- NoSQL		
Module:8	**Contemporary issues**	**2 hours**
Expert Talk		
	Total Lecture Hours:	**45 hours**

Text Book(s)	
1.	Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data by EMC Education Services, 2015, publishing.
Reference Books	
1.	Anand Raja Raman and Jeffrey David Ullman, Mining of Massive Datasets, 2012, Cambridge University Press.
2.	Tom White, Hadoop: The Definitive Guide, 3rd Edition, O'Reilly Media

Recommended by Board of Studies	05-03-2016		
Approved by Academic Council	40th	Date	18-03-2016