Course Code	Course Title	L	T	P	J	C
ITA6008	Big Data Analytics	3	0	0	4	4
Pre-requisite	ITA5008	Syllabus version				
		1.0				

Course Objectives:

- Understand the fundamentals of various big data analysis techniques
- Analyze the big data analytic techniques for useful business applications
- Perform map-reduce analytics using Hadoop and related tools

Expected Course Outcome:

On completion of this course, student should be able to

- Analyze Big data, create statistical models, identify insights that can lead to actionable results
- Perform map-reduce analytics using Hadoop
- Implement software tools such as R and Hadoop for big data analytics

Student Learning Outcomes (SLO): 3, 7, 14

Module:1	**Introduction to Big Data Analytics**	**6 hours**	**SLO: 3**

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**	**SLO: 14**

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

Module:3	**Map Reduce**	**6 hours**	**SLO: 7**

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**	**SLO: 14**

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**	**SLO: 3**

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**	**SLO: 7**

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**	**SLO: 7**

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	40^{th}	Date	18-03-2016