BIG DATA ANALYTICS
[As per Choice Based Credit System (CBCS) scheme]
(Effective from the academic year 2016 -2017)
SEMESTER – VIII

Subject Code	15CS82	IA Marks	20
Number of Lecture Hours/Week	4	Exam Marks	80
Total Number of Lecture Hours	50	Exam Hours	03

CREDITS – 04

Course objectives: This course will enable students to

- Understand Hadoop Distributed File system and examine MapReduce Programming
- Explore Hadoop tools and manage Hadoop with Ambari
- Appraise the role of Business intelligence and its applications across industries
- Assess core data mining techniques for data analytics
- Identify various Text Mining techniques

Module – 1	**Teaching Hours**
Hadoop Distributed File System Basics, Running Example Programs and Benchmarks, Hadoop MapReduce Framework, MapReduce Programming	**10 Hours**
Module – 2	
Essential Hadoop Tools, Hadoop YARN Applications, Managing Hadoop with Apache Ambari, Basic Hadoop Administration Procedures	**10 Hours**
Module – 3	
Business Intelligence Concepts and Application, Data Warehousing, Data Mining, Data Visualization	**10 Hours**
Module – 4	
Decision Trees, Regression, Artificial Neural Networks, Cluster Analysis, Association Rule Mining	**10 Hours**
Module – 5	
Text Mining, Naïve-Bayes Analysis, Support Vector Machines, Web Mining, Social Network Analysis	**10 Hours**

Course outcomes: The students should be able to:

- Master the concepts of HDFS and MapReduce framework
- Investigate Hadoop related tools for Big Data Analytics and perform basic Hadoop Administration
- Recognize the role of Business Intelligence, Data warehousing and Visualization in decision making
- Infer the importance of core data mining techniques for data analytics
- Compare and contrast different Text Mining Techniques

Question paper pattern:
The question paper will have ten questions.
There will be 2 questions from each module.
Each question will have questions covering all the topics under a module.
The students will have to answer 5 full questions, selecting one full question from each module.

Text Books:

1. Douglas Eadline,**"Hadoop 2 Quick-Start Guide: Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem",** 1st Edition, Pearson Education, 2016. ISBN-13: 978-9332570351

2. Anil Maheshwari, **"Data Analytics",** 1st Edition, McGraw Hill Education, 2017. ISBN-13: 978-9352604180

Reference Books:

1) Tom White, **"Hadoop: The Definitive Guide",** 4th Edition, O'Reilly Media, 2015.ISBN-13: 978-9352130672
2) Boris Lublinsky, Kevin T.Smith, Alexey Yakubovich,**"Professional Hadoop Solutions",** 1stEdition, Wrox Press, 2014ISBN-13: 978-8126551071
3) Eric Sammer,**"Hadoop Operations: A Guide for Developers and Administrators",**1stEdition, O'Reilly Media, 2012.ISBN-13: 978-9350239261