CENTRAL UNIVERSITY OF KERALA DEPARTMENT OF COMPUTER SCIENCE M.Sc. COMPUTER SCIENCE – PROGRAMME STRUCTURE					
COURSE CODE	**COURSE TITLE**	**CONTACT HRS/WEEK**			**CREDITS**
		LEC	**LAB**	**TUT**	
SEMESTER III					
CSC5301	Big Data Analytics	2	2	1	4

This is an experimental, problem solving, **skill development and employability based course.**

Course Objective
  The objective of the course is to provide theoretical and practical aspects of big data analytics.

By completing this course, students will obtain the following course outcomes:
1. Knowledge gained:
    (i)   State-of-art Big Data Analytics techniques and algorithms
2. Skill gained:
    (ii) Critically Analyze and perform big data analysis using Hadoop and MapReduce technologies
    (iii) Ability to identify the characteristics of data sets and compare the trivial and big data for various applications.
    (iv) Ability to solve problems associated with batch learning and online learning
    (v)   Effectually handling big data characteristics such as high dimensionality, dynamically growing data and scalability issues
3. Competency gained:
    (vi) Implement real world big data applications

Prerequisites: Basic knowledge in data mining.

Grading:
Lab experiments and implementation	– 15%
Mini project (individual)	– 10%
Class Test	- 10%
Assignment/Quiz/presentation	– 5%
Final Exam	– 60%

CSC5301 – Big Data Analytics

Module 1
Introduction to Big Data, challenges of conventional systems, characteristics of Big Data-Volume, Variety, Velocity, Veracity, etc., Big Data analytics, Big Data applications. Introduction to enabling technologies for Big Data, introduction to Big Data stack, introduction to some Big Data distribution packages

Module 2
Introduction to Big Data platforms, overview of Apache Spark, YARN, Hadoop. Hadoop distributed file system, components of Hadoop, Hadoop architecture, analysing the data with Hadoop, introduction to MapReduce, MapReduce programming model, MapReduce examples.

Module 3
Introduction to Big Data storage platforms for large scale data storage, introduction to Big Data streaming platforms for fast data. Introduction to Big Data applications (Machine Learning), overview of Big Data Machine Learning, Mahout introduction, Big Data Machine Learning algorithms in Mahout- kmeans, Naïve Bayes etc.

Module 4
Predictive Analytics-Simple linear regression, multiple linear regression, interpretation of regression coefficients. Visualizations - Visual data analysis techniques, interaction techniques-systems and applications.

References:
1. Dirk Deroos et al., Hadoop for Dummies, Dreamtech Press, 2014, ISBN: 978-1-118-60755-8(pbk), 978-1-118-65220-6(ebk), 978-1-118-70503-2(ebk).
2. Chuck Lam, Hadoop in Action, December, 2010, Manning Publications, ISBN: 9781935182191
3. Jiawei Han, Micheline Kamber "Data Mining Concepts and Techniques", 2nd Edition, Elsevier, Reprinted 2008, ISBN 978-0-12-381479-1.
4. J. Leskovec, A. Rajaraman, J.D. Ullman, Mining of Massive Datasets, Cambridge University Press, ISBN: 978-1-107-07723-2., 9781108476348, 2020
5. Arshdeep Bahga, Vijay Madisetti, "Big Data Science & Analytics: A Hands On Approach", VPT, 2016, ISBN: 978-0996025539