BIG DATA AND ANALYTICS
(Effective from the academic year 2018 -2019)
SEMESTER – VII

Course Code	18CS72	CIE Marks	40
Number of Contact Hours/Week	4:0:0	SEE Marks	60
Total Number of Contact Hours	50	Exam Hours	03
CREDITS –4			

Course Learning Objectives: This course (18CS72) will enable students to:

- Understand fundamentals of Big Data analytics
- Explore the Hadoop framework and Hadoop Distributed File system
- Illustrate the concepts of NoSQL using MongoDB and Cassandra for Big Data
- Employ MapReduce programming model to process the big data
- Understand various machine learning algorithms for Big Data Analytics, Web Mining and Social Network Analysis.

Module 1	Contact Hours
Introduction to Big Data Analytics: Big Data, Scalability and Parallel Processing, Designing Data Architecture, Data Sources, Quality, Pre-Processing and Storing, Data Storage and Analysis, Big Data Analytics Applications and Case Studies. **Text book 1: Chapter 1: 1.2 -1.7** **RBT: L1, L2, L3**	10
Module 2	
Introduction to Hadoop (T1): Introduction, Hadoop and its Ecosystem, Hadoop Distributed File System, MapReduce Framework and Programming Model, Hadoop Yarn, Hadoop Ecosystem Tools. **Hadoop Distributed File System Basics (T2):** HDFS Design Features, Components, HDFS User Commands. **Essential Hadoop Tools (T2):** Using Apache Pig, Hive, Sqoop, Flume, Oozie, HBase. **Text book 1: Chapter 2 :2.1-2.6** **Text Book 2: Chapter 3** **Text Book 2: Chapter 7 (except walk throughs)** **RBT: L1, L2, L3**	10
Module 3	
NoSQL Big Data Management, MongoDB and Cassandra: Introduction, NoSQL Data Store, NoSQL Data Architecture Patterns, NoSQL to Manage Big Data, Shared-Nothing Architecture for Big Data Tasks, MongoDB, Databases, Cassandra Databases. **Text book 1: Chapter 3: 3.1-3.7** **RBT: L1, L2, L3**	10
Module 4	
MapReduce, Hive and Pig: Introduction, MapReduce Map Tasks, Reduce Tasks and MapReduce Execution, Composing MapReduce for Calculations and Algorithms, Hive, HiveQL, Pig. **Text book 1: Chapter 4: 4.1-4.6** **RBT: L1, L2, L3**	10

Module 5	
Machine Learning Algorithms for Big Data Analytics: Introduction, Estimating the relationships, Outliers, Variances, Probability Distributions, and Correlations, Regression analysis, Finding Similar Items, Similarity of Sets and Collaborative Filtering, Frequent Itemsets and Association Rule Mining. **Text, Web Content, Link, and Social Network Analytics:** Introduction, Text mining, Web Mining, Web Content and Web Usage Analytics, Page Rank, Structure of Web and analyzing a Web Graph, Social Network as Graphs and Social Network Analytics: **Text book 1: Chapter 6: 6.1 to 6.5** **Text book 1: Chapter 9: 9.1 to 9.5**	10

Course Outcomes: The student will be able to:

- Understand fundamentals of Big Data analytics.
- Investigate Hadoop framework and Hadoop Distributed File system.
- Illustrate the concepts of NoSQL using MongoDB and Cassandra for Big Data.
- Demonstrate the MapReduce programming model to process the big data along with Hadoop tools.
- Use Machine Learning algorithms for real world big data.
- Analyze web contents and Social Networks to provide analytics with relevant visualization tools.

Question Paper Pattern:

- The question paper will have ten questions.
- Each full Question consisting of 20 marks
- There will be 2 full questions (with a maximum of four sub questions) from each module.
- Each full question will have sub questions covering all the topics under a module.
- The students will have to answer 5 full questions, selecting one full question from each module.

Textbooks:

1. Raj Kamal and Preeti Saxena, "**Big Data Analytics Introduction to Hadoop, Spark, and Machine-Learning**", McGraw Hill Education, 2018 ISBN: 9789353164966, 9353164966
2. 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

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**", 1st Edition, Wrox Press, 2014ISBN-13: 978-8126551071
3. Eric Sammer, "**Hadoop Operations: A Guide for Developers and Administrators**",1st Edition, O'Reilly Media, 2012.ISBN-13: 978-9350239261
4. Arshdeep Bahga, Vijay Madisetti, "**Big Data Analytics: A Hands-On Approach**", 1st Edition, VPT Publications, 2018. ISBN-13: 978-0996025577

H. O. D.
Dept. Of Computer Science & Engineering
Alva's Institute of Engg. & Technology
Mijar, MOODBIDRI - 574 225