

MANGALORE UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

CSS 505: BIG DATA ANALYTICS		
Hours/Week: 4 Credits : 4		I.A. Marks: 30 Exam. Marks: 70
Course Outcomes: CO1: Work with big data platform and explore the big data analytics techniques business applications. CO2: Design efficient algorithms for mining the data from large volumes. CO3: Analyze the HADOOP and Map Reduce technologies associated with big data analytics. CO4: Explore on Big Data applications Using Pig and Hive. CO5: Understand the fundamentals of various big data analytics techniques. CO6: Build a complete business data analytics solution.		
	UNIT-I	12 Hrs.
INTRODUCTION TO BIG DATA: Introduction – distributed file system – Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications. Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce.		
	UNIT-II	12 Hrs.
INTRODUCTION HADOOP: Big Data – Apache Hadoop & Hadoop EcoSystem – Moving Data in and out of Hadoop – Understanding inputs and outputs of MapReduce - Data Serialization. HADOOP ARCHITECTURE: Hadoop Architecture, Hadoop Storage: HDFS, Common Hadoop Shell commands, Anatomy of File Write and Read., NameNode, Secondary NameNode, and DataNode, Hadoop MapReduce paradigm, Map and Reduce tasks, Job, Task trackers - Cluster Setup – SSH & Hadoop Configuration – HDFS Administering –Monitoring & Maintenance.		
	UNIT-III	12 Hrs.
HADOOP ECOSYSTEM AND YARN: Hadoop ecosystem components - Schedulers - Fair and Capacity, Hadoop 2.0 New Features, NameNode High Availability, HDFS Federation, MRv2, YARN, Running MRv1 in YARN.		
	UNIT-IV	12 Hrs.
HIVE AND HIVEQL, HBASE : Introduction to No Query Language, Hive Architecture and Installation, Comparison with Traditional Database, HiveQL - Querying Data - Sorting And Aggregating, Map Reduce Scripts, Joins & Subqueries, HBase concepts Advanced Usage, Schema Design, Advance Indexing - PIG, Zookeeper - how it helps in monitoring a cluster, HBase uses Zookeeper and how to Build Applications with Zookeeper.		

REFERENCE BOOKS:

1. Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, “Professional Hadoop Solutions”, Wiley, ISBN: 9788126551071, 2015.
2. Chris Eaton, Dirk deroos et al. , “Understanding Big data ”, McGraw Hill, 2012.
3. Tom White, “HADOOP: The definitive Guide” , O Reilly, 2012.
4. Vignesh Prajapati, “Big Data Analytics with R and Haoop”, Packet Publishing 2013.
5. Tom Plunkett, Brian Macdonald et al, “Oracle Big Data Handbook”, Oracle Press, 2014.
6. <http://www.bigdatauniversity.com/>
7. Jy Liebowitz, “Big Data and Business analytics”,CRC press, 2013.

