

Branch Name:	Computer Engineering (Software Engineering)
Program Code:	ET203
Course Title:	Big Data Analytics
Course Code:	2ET2030201T
Pre-requisite Course:	Probability Theory, Linear Algebra, Calculus, And Statistical Analysis

Course Objectives: This course provides practical foundation level training that enables immediate and effective participation in big data projects. The course provides grounding in basic and advanced methods to big data technology and tools, including MapReduce and Hadoop and its ecosystem.

Teaching and Examination Scheme:

Credit	Teaching Scheme Hours per week				Evaluation Scheme (Marks)								
					Theory (Marks)				Practical (Marks)				Total (Marks)
	L	T	P	Total	SEE		CCA		SEE		CCA		
					Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	
3	3	-	-	3	24	60	16	40	-	-	-	-	100

Course Contents:

Unit No	Topics	Total Hours	Weightage (%)
1	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.	02	10
2	HADOOP ECOSYSTEM AND YARN Hadoop ecosystem components - Schedulers - Fair and Capacity, Hadoop 2.0 New Features NameNode High Availability, HDFS Federation, MRv2, YARN.	02	10
3	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	04	18
4	HIVE AND HIVEQL, HBASE Hive Architecture and Installation, Comparison with Traditional Database, HiveQL - Querying Data - Sorting And Aggregating.	04	18

Text Books / Reference Books :

1. Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, “Professional Hadoop Solutions”, Wiley, 2015.
2. Chris Eaton, Dirk deroos et al., “Understanding Big data”, McGraw Hill, 2012.
3. Tom White, “HADOOP: The definitive Guide”, O Reilly 2012

List of Open-Source Software/learning website:

- ☐ <http://bigdatauniversity.com/courses/?cat=124>
- ☐ <http://www.coreservlets.com/hadoop-tutorial/>

Course Learning Outcomes (CLO):

Course Outcomes	Description	Bloom's Taxonomy Level
COL1	learn tips and tricks for Big Data use cases and solutions	Understanding (2) Applying (3)
COL2	learn to build and maintain reliable, scalable, distributed systems with Apache Hadoop	Applying (3)
COL3	apply Hadoop ecosystem components	Applying (3)

Mapping of CLO with PO and PSO

Course Learning Outcomes	Program Outcomes (POs)												Program Specific Outcomes(PSOs)			
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3	PSO4
CLO1	H	M	M		M											
CLO2	H	M			M							L				
CLO3	H	M	M		M											
CLO4	M	M	M		M							L				
CLO5	M	M										L				

H: High, M: Medium, L: Low