

Branch Name:	Information and Communication Technology
Program Code:	ET208
Course Name:	Big Data Analytics
Course Code:	2ET2080202T
Pre-requisite Course:	-

Course Objective:

1. Understand the Big Data Platform and its Use cases.
2. Students will understand basic and advanced methods to big data technology and tools, including Map Reduce and Hadoop and its ecosystem.
3. To improve practical foundation level training that enables immediate and effective participation in big data projects.

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

Subject Contents			
Sr. No	Topic	Total Hours	Weight (%)
1	Introduction to Big Data Introduction to Big Data, Big Data characteristics, Challenges of Conventional System, Types of Big Data, Intelligent data analysis, Traditional vs. Big Data business approach, Case Study of Big Data Solutions	3	16

2	Hadoop Ecosystem & Architecture Hadoop Ecosystem, Hadoop Components – HDFS, MapReduce, YARN, Analyzing the Data with Hadoop, Scaling Out, Hadoop Streaming, Job Scheduling, Hadoop environment. Setting up a Hadoop Cluster, Cluster specification, Cluster Setup and Installation, Hadoop Configuration, security in Hadoop, Administering Hadoop, Monitoring-Maintenance, Hadoop benchmarks	7	30
3	NoSQL What is NoSQL? NoSQL business drivers; NoSQL case studies; NoSQL data architecture patterns: Key-value stores, Graph stores, Column family stores, Document stores, MongoDB Basics	4	25
4	Spark Introduction to Data Analysis with Spark, In-Memory Computing with Spark, Spark Basics, Interactive Spark with PySpark, Writing Spark Applications, Spark Installation & Cluster	5	16
5	Hive and Hiveql, Hbase Hive Architecture and Installation, Comparison with Traditional Database, HiveQL - Querying Data - Sorting And Aggregating.	5	13

Text Books:

1. Chris Eaton, Dirk deRoos et al., “Understanding Big data”, McGraw Hill, 2012.

Reference Books:

1. Boris Imlinskiy, Kevin t. Smith, Alexey Yakubovich, “Professional Hadoop Solutions”, Wiley, 2015.
2. Tom White, “HADOOP: The definitive Guide”, O Reilly 2012

List of Open Source Software/learning website:

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

Course Learning Outcomes (CLO): On completion of this course, the students will be able to:

CLO	Description	Bloom's Taxonomy Level
CLO1	Students will to build and maintain reliable, scalable, distributed systems with Apache Hadoop	2 Understanding, 6 Creating
CLO2	Able to write Map-Reduce based Applications	1 Remembering, 2 Understanding
CLO3	Able to design and build MongoDB based Big data Applications and learn MongoDB query language	3 Applying, 2 Understanding, 6 Creating

CLO4	Learn difference between conventional SQL query language and NoSQL basic concepts	3 Applying, 2 Understanding
CLO5	Learn tips and tricks for Big Data use cases and solutions.	5 Evaluate, 3 Applying, 4 Analyze

Mapping of CLOs with POs & PSOs

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

H:High, M:Medium, L:Low