

Branch Name:	Artificial Intelligence & Data Science
Program Code:	Semester VI/ Third Year
Subject Title:	Distributed Computing & Big Data Analytics
Subject Code:	1ET1080603
Pre-requisite:	Programming skills

Course Objective:

1. To provides practical foundation level training that enables immediate and effective participation in big data projects.
2. To understand grounding in basic and advanced methods to big data technology and tools, including Map Reduce and Hadoop and its ecosystem.

Teaching Scheme (Hours per week)				Evaluation Scheme (Marks)				
Lecture	Tutorial	Practical	Credit	Theory		Practical		Total
				University Assessment	Continuous Assessment	University Assessment	Continuous Assessment	
03	-	02	04	70	30	30	20	150

Subject Contents					
Sr. No	Topic			Total Hours	Weight (%)
1	Distributed Computing: What is Distributed computing, Characteristics, Advantages, Distributed computing models			7	10
2	Synchronization: Synchronization: Clock synchronization, Logical clocks, Global state, Mutual exclusion, Election algorithms: Bully algorithm, Ring algorithm, Leader election in rings, anonymous rings, Asynchronous rings, synchronous rings, election in wireless networks, Deadlocks in Distributed systems, Deadlocks in Message communication			4	15
3	Interprocess and Remote Communication: Message Passing and its features, IPC message format, IPC synchronization, Buffering, multi datagram messaging, process addressing techniques, failure handling, RPC basics, RPC implementation, RPC Communication and Other issues, Sun RPC			7	30
4	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			6	10
5	Big Data – Apache Hadoop & Hadoop Ecosystem – Moving Data in and out of Hadoop – Understanding inputs and outputs of Map Reduce - Data Serialization. Hadoop Architecture, Hadoop Storage: HDFS, Common Hadoop Shell commands, Anatomy of File Write and Read., Name Node, Secondary Name Node, and Data Node, Hadoop Map Reduce paradigm, Map and Reduce tasks			5	10

6	NoSQL: What is NoSQL? NoSQL business drivers, NoSQL data architecture patterns: Key-value stores, Graph stores, Column family (Bigtable) stores, Document stores, Variations of NoSQL architectural patterns; Using NoSQL to manage big data: What is a big data NoSQL solution? Understanding the types of big data problems; Analyzing big data with a shared-nothing architecture; Choosing distribution models: master-slave versus peer-to-peer; Four ways that NoSQL systems handle big data problems	4	25
---	---	---	----

Course Outcome:

Upon successful completion of this course the student would be able to:

1. Learn tips and tricks for Big Data use cases and solutions
2. Learn to build and maintain reliable, scalable, distributed systems with Apache Hadoop
3. Apply Hadoop ecosystem components

List of References:

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

E-Resources:

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

List of Experiments:

1. Master the concepts of HDFS and MapReduce framework.
2. Understand Hadoop 2.x Architecture.
3. Setup Hadoop Cluster and write Complex MapReduce programs.
4. Learn data loading techniques using Sqoop and Flume.
5. Perform data analytics using Pig, Hive and YARN.
6. Implement HBase and MapReduce integration.
7. Implement Advanced Usage and Indexing.
8. Implement best practices for Hadoop development.
9. Study about different types of noSQL Databases.
10. Write a program to draw a Graph using NoSQL database.

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

CLO	Description	Bloom's Taxonomy Level
CLO1	To get knowledge in distributed architecture, naming, synchronization, consistency and replication, fault tolerance, security, and distributed file systems	2 Understanding
CLO2	Understanding the concept Synchronization, Mutual exclusion, and Deadlock	2 Understanding
CLO3	Gain insights into various Election algorithms: Bully algorithm, Ring algorithm, Leader election in rings, anonymous rings, Asynchronous rings, synchronous rings	3 Applying
CLO4	Learn depth understanding of Hadoop, Architecture, storage and Hadoop Eco-System	4 Analyze
CLO5	Demonstrate competency in describing how NoSQL databases differ from relational databases from a theoretical perspective	2 Understanding 3 Applying

Mapping of CLOs with POs & PSOs

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	M	L	L	H	M	M			L				H	L		
CLO2		L	M	H	L	L					L	L	L	L		
CLO3	L	M	L	L	M	H		M		L			M	L		
CLO4	H	M	M	L	M	H		M	L				L	L		
CLO5	M	L	M	L	M	H		M		L			H	L		

H: High, M: Medium, L: Low