

Branch Name:	Artificial Intelligence and Data Science
Program Code	ET108
Semester/Year:	Semester-VII/Fourth Year
Course Name:	Distributed Computing and Big Data Analytics Practical
Course Code:	2ET1080702P
Pre-requisite:	Knowledge of Data Structures, DBMS, Operating Systems, and Programming (Java/Python)

Course Objective:

1. To understand concepts of distributed computing, scalability, and fault tolerance.
2. To introduce distributed file systems and computing frameworks such as Hadoop and Spark.
3. To learn various techniques for Big Data storage, processing, and analytics.
4. To develop practical skills in managing and analyzing large datasets using distributed systems.

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.	
1	-	-	2	2	-	-	-	-	10	25	10	25	50

Sr. No	List of Practical	Content Hours
1.	Study of Distributed System Architectures.	2
2.	Implementation of Message Passing in Distributed Systems.	2
3.	Installation and Configuration of Hadoop Cluster and Working with HDFS	2
4.	Implementing MapReduce word count and sorting.	2
5.	Data processing using Pig/Hive.	2
6.	Study of different types of NoSQL databases.	2
7.	Data Processing using Apache Spark (RDDs and DataFrames).	2
8.	Streaming analytics using Spark Streaming and Kafka.	2
9.	Mini-project: End-to-end Big Data pipeline (data ingestion → storage → processing → visualization).	2

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/>

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

CLO	Description	Bloom's Taxonomy Level
CLO1	Explain the principles and architecture of distributed computing systems	2 Understanding
CLO2	Implement distributed data storage and processing using Hadoop ecosystem.	2 Understanding 3 Applying
CLO3	Apply Big Data analytics tools such as Spark for large-scale data analysis.	3 Applying
CLO4	Demonstrate proficiency in handling streaming and real-time data analytics.	4 Analyze
CLO5	Evaluate performance and fault tolerance aspects in distributed environments.	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	M	M	M	L						M	H	M	M	
CLO2	L	H	H	M	H					L		M		H	L	
CLO3	L	H	M	M	H	L								M		M
CLO4	M	M	H	M	M			L					H		M	H
CLO5	H	M	H		H				L			M	M		M	H

H: High, M: Medium, L: Low