



SWARNANDHRA

COLLEGE OF ENGINEERING & TECHNOLOGY (AUTONOMOUS)

Accredited by National Board of Accreditation, AICTE, New Delhi, Accredited by NAAC with 'A' Grade – 3.32 CGPA, Recognized under 2(f) & 12(B) of UGC Act 1956, Approved by AICTE, New Delhi, Permanent Affiliation to JNTUK, Kakinada, Seethampuram, W.G.D.T., Narsapur-534280, (Andhra Pradesh)

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING -DATA SCIENCE

TEACHING PLAN

Course Code	Course Title	Sem	Branch	Contact Periods /Week	Academic Year	Date of commencement of Semester
20AD6T02	BIG DATA ANALYTICS	VI	CSE-DS	6	2023-24	18-11-2024

COURSE OUTCOMES

1	Demonstrate knowledge of Big Data, Data Analytics, challenges and their solutions in Big Data.
2	Analyze Hadoop Framework and eco systems.
3	Compare and work on NewSQL environment and MongoDB and Cassandra.
4	Apply the Big Data using Map-reduce programming in Both Hadoop and Spark framework.
5	Analyze the data Analytics algorithms in Spark

UNIT	CO	Topics No.	Topics/Activity	Text Book / Reference	Contact Hour	Delivery method
------	----	------------	-----------------	-----------------------	--------------	-----------------

UNIT – I : INTRODUCTION TO BIG DATA

I	Demonstrate knowledge of Big Data, Data Analytics, challenges and their solutions in Big Data.	1.1	Data, Characteristics of data	T3	1	Chalk and talk PPT Smart Board
		1.2	Types of digital data: Unstructured, Semi-structured and Structured	T3	2	
		1.3	Sources of data	T3	1	
		1.4	Big Data Evolution - Definition of big data	T3	1	
		1.5	Characteristics and Need of big data	T3	1	
		1.6	Challenges of big data	T3	1	
		1.7	Big data analytics	T3	1	
		1.8	Overview of business intelligence.	T3	1	
		CBS	Data Provenance in Big Data	T3	1	
Total					10	

UNIT – II: BIG DATA TECHNOLOGIES AND DATABASES

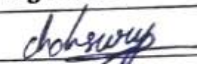



II	Analyze Hadoop Framework	2.1	Hadoop	T1	1	Chalk and talk PPT Smart
		2.2	Requirement of Hadoop Framework	T1	2	
		2.3	Design principle of Hadoop	T1	1	

		2.4	Comparison with other system SQL and RDBMS	T1	2	Board	
		2.5	Hadoop Components – Architecture -Hadoop 1 vs Hadoop 2.	T3	1		
		CBS	Hadoop Ecosystem Beyond Core Components	T3	1		
					Total	8	
UNIT – III: MapReduce and YARN framework							
III	Compare and work on NewSQL environment and MongoDB and Cassandra.	3.1	Introduction to MapReduce	T1	1	Chalk and talk PPT Smart Board	
		3.1.1	Processing data with Hadoop using MapReduce	T1	1		
		3.2	Introduction to YARN	T1	1		
		3.2.1	Architecture	T1	1		
		3.2.2	Managing Resources and Applications with Hadoop YARN	T1	2		
		3.2.3	Big data technologies and Databases	T1	1		
		3.3	Introduction to NoSQL	T1	1		
		3.3.1	Features and Types	T1	1		
		3.3.2	Advantages & Disadvantages	T1	1		
		3.3.3	Application of NoSQL.	T1	1		
		CBS	NoSQL Data Modeling Techniques	T1	1		
							Total
UNIT-IV New SQL							
IV	Apply the Big Data using Map-reduce programming in Both Hadoop and Spark framework	4.1	New SQL: Overview of New SQL	T6	1	Chalk and talk PPT Smart Board	
		4.1.1	Comparing SQL	T6	1		
		4.1.2	NoSQL and NewSQL	T6	1		
		4.2	Introduction to Mongo DB	T7	1		
		4.2.1	Features – Data types	T7	1		
		4.2.2	Mongo DB Query language	T7	1		
		4.2.3	CRUD operations	T7	1		
		4.2.4	Arrays	T7	1		
		4.2.5	Functions	T7	1		
		4.2.6	Count	T7	1		
		4.2.7	Sort	T7	1		
		4.2.8	Limit	T7	1		
		4.2.9	Skip	T7	1		
		4.2.10	Aggregate	T7	1		
		4.2.11	Map Reduce	T7	1		
4.2.12	Cursors	T7	1				
4.2.13	Indexes	T7	1				

		4.2.14	Mongo Import – Mongo Export	T7	1		
		4.3	Cassandra: Introduction	T4	1		
		4.3.1	Features – Data types	T4	1		
		4.3.2	CQLSH – Key spaces	T4	1		
		4.3.3	CRUD operations	T4	1		
		4.3.4	Collections	T4	1		
		4.3.5	Counter	T4	1		
		4.3.6	TTL – Alter commands – Import and Export – Querying System tables.	T4	1		
		CBS	Data Sharding in NoSQL and Distributed Databases	T6	1		
					Total	26	
UNIT-V Big Data Frame Works for Analytics							
V	Analyze the data Analytics algorithms in Spark	5.1	Hadoop Frame Work	T1	1	Chalk and talk PPT Smart Board	
		5.1.1	Map Reduce Programming: I/O formats	T1	2		
		5.1.2	Map side join-Reduce Side Join	T1	1		
		5.1.3	Secondary Sorting	T1	1		
		5.1.4	Pipelining MapReduce jobs Spark Frame Work	T1	1		
		5.2	Introduction to Apache spark	T5	1		
		5.3	How spark works	T5	1		
		5.4	Programming with RDDs	T2	1		
		5.5	Create RDD spark Operations	T2	1		
		5.5.1	Data Frame	T2	1		
		CBS	Resilient Distributed Datasets (RDD) Fault Tolerance in Spark	T5	1		
							Total
CUMULATIVE PROPOSED PERIODS						68	

Text Books:	
S. No.	AUTHORS, BOOK TITLE, EDITION, PUBLISHER, YEAR OF PUBLICATION
1	Tom White , "Hadoop: The Definitive Guide" O'Reilly Media, 2021.
2	Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia "Learning Spark: Lightning-Fast Data Analytics" , O'Reilly Media,2020.
3	Seema Acharya and Subhashini Chellappan, "Big Data and Analytics", Wiley India Pvt. Ltd., 2016.
4	Jeff Carpenter and Eben Hewitt" Cassandra: The Definitive Guide", O'Reilly Media, 2016.
5	Mike Frampton, "Mastering Apache Spark", Packt Publishing, 2015.
6	Tharun V and Rajesh V "NewSQL: A SQL Database for the Modern Web", Wiley India Pvt. Ltd.,2017.

7	Kristina Chodorow and Michael Dirolf, "MongoDB: The Definitive Guide", O'Reilly Media, 2013
Reference Books:	
S. No.	AUTHORS, BOOK TITLE, EDITION, PUBLISHER, YEAR OF PUBLICATION
1	Tom White, "Hadoop: The Definitive Guide", O'Reilly, 4th Edition, 2015.
2	Mohammed Guller, "Big Data Analytics with Spark", Apress, 2015
3	Donald Miner, Adam Shook, "Map Reduce Design Pattern", O'Reilly, 2012.
Web Details	
1	https://www.ibm.com/topics/big-data-analytics
2	https://www.geeksforgeeks.org/what-is-big-data-analytics/
3	https://azure.microsoft.com/en-in/resources/cloud-computing-dictionary/what-is-big-data-analytics
4	https://www.bornfight.com/blog/7-real-world-examples-of-how-brands-are-using-big-data-analytics/
5	https://www.ibm.com/topics/big-data-analytics
Video Links	
1	https://www.youtube.com/watch?v=bY6ZzQmtOzk
2	https://onlinecourses.nptel.ac.in/noc20_cs92/preview
3	https://hadoop.apache.org/

	Name	Signature with Date
i.	Faculty I	Mrs. Ch. Chandrika Surya 
ii.	Course Coordinator	Mr.K.Jai Prakash 
iii.	Module Coordinator	Dr.G.Sudhakar 
iv.	Programme Coordinator	Dr.B.Rama Krishna 


Principal