Swarnandhra College of Engineering & Technology

Autonomous and recognized under 2(F) and 12(B) by UGC

Recognized by AICTE, permanently affiliated to JNTUK Kakinada

Accredited by NAAC with 'A' Grade (2nd Cycle)

Seetharamapurm, Narsapur – 530280 (Andhra Pradesh)

DEPARTMENT OF INFORMATION TECHNOLOGY

TEACHING PLAN

Course Code	Course Title	Semester	Branch	Contact Periods /Week	Academic Year	Date of commencement
20IT7E05	BIG DATA ANALYTICS	VII	IT	6	2024-25	05-06-2024

COURSE OUTCOMES	
1	List out the basic concepts of Big Data and Big Data Analytics (K1)
2	Analyze the HDFS architecture (K4)
3	Develop the Map Reduce application (K3)
4	Identify the various Hadoop Ecosystem technologies (K3)
5	Outline the Advanced Analytical methods for classification, clustering and Text Analysis (K2)

UNIT	Out Comes / Bloom's Level	Topics No.	Topics/ Activity	Text Book/ Ref	Contact Hour	Delivery Method
I	CO – 1		**Unit-1: Introduction to Big Data**			Chalk & Board Power point presentations Assignment Test
		1.1	Types of Digital Data	T2	1	
		1.2	Characteristics of Data	T2	1	
		1.3	Evolution of Big Data	T2	2	
		1.4	Definition of Big Data – Challenges with Big Data	T2	1	
		1.5	3Vs of Big Data	T2	1	
		1.6	Non Definitional traits of Big Data	T2	2	
		1.7	Business Intelligence vs. Big Data	T2	2	
		1.8	Data warehouse and Hadoop environment – Coexistence	T2	1	
		1.9	Classification of analytics – Data Science – Terminologies in Big Data – CAP Theorem.	T1,T2	2	
Content beyond syllabus		1.9	Principles of Big Data	R1	1	
				Total	14	
II	CO – 2		**Unit-2: Hadoop Distributed File System**			Chalk &
		2.1	Introduction to Hadoop	T1,T2	1	
		2.2	Frame work, features, advantages.	T1,T2	2	
		2.3	HDFS concepts	T1	1	

		2.4	HDFS architecture	T1,T2	1	Board
		2.5	hdfs commands	T1,T2	1	Power point presentations Assignment Test
		2.6	Hadoop file system interfaces	T1,T2	1	
		2.7	Data flow	T1,T2	1	
		2.8	Data Ingest with Flume and Scoop	T2,R2	1	
		2.9	Hadoop archives.	T2,R2	1	
Content beyond syllabus		2.10	Examples of XML files.	R2	1	
				Total	11	
III	CO – 3		**Unit-3: Map Reduce**			
		3.1	Understanding Hadoop API for MapReduce Framework	T1,T2	1	Chalk & Board Power point presentations Assignment Test
		3.2	MapReduce features	T1,T2	1	
		3.3	Techniques to Optimize MapReduce Jobs	T1,T2	2	
		3.4	Mapper – Reducer	T1,T2	1	
		3.5	Combiner – Partitioner	T1,T2	2	
		3.6	Searching – Sorting	T1,T2	1	
		3.7	Compression	T1,T2	1	
		3.8	I/O formats	T1,T2	1	
		3.9	Uses of MapReduce	T1,T2	1	
		3.10	Ingestion layer in big data stack	T1,T2	1	
Content beyond syllabus		3.11	MapReduce Architecture	R1	1	
				Total	13	
IV	CO – 4		**Unit-4: Hadoop Ecosystem**			
		4.1	Serialization: AVRO	T1,T2	1	Chalk & Board Power point presentations Assignment Test
		4.2	Co-ordination: Zookeeper	T1,T2	1	
		4.3	Databases: HBase Concepts	T1,T2	1	
		4.4	Clients, Example, HBase Versus RDBMS	T1,T2	1	
		4.5	Hive-architecture, data types, file formats	T1,T2	1	
		4.6	HQL--Scripting language	T1,T2	1	
		4.7	Scripting language: Pig-features, anatomy	T1,T2	2	
		4.8	pig on Hadoop	T1,T2	2	
		4.9	pig latin over view, data types	T1,T2	1	
		4.10	running pig, execution modes of pig	T1,T2	1	
		4.11	Streaming: Flink, Storm	T1,T2	1	
Content beyond syllabus		4.12	Principles of Analytical Theory	T2	1	
				Total	14	
V	CO – 5		**Unit-5: Advanced Analytics**			

		5.1	Clustering: k-means, k-mode algorithm	T1,T2, R3	2	Chalk & Board Power point presentations Assignment Test
		5.2	Association Rules: Apriori Algorithm, Evaluation of Candidate Rules, Applications of Association Rules	T2,R3	2	
		5.3	Classification: Decision Trees, Naïve Bayes	T2,R3	2	
		5.4	Time Series Analysis: Overview of Time Series Analysis, ARIMA Model	T1,T2, R3	2	
		5.5	Text Analysis: Regular expressions, Collecting Raw Text, Representing Text	T1,T2	2	
		5.6	Term Frequency—Inverse Document Frequency (TFIDF)	T1,T2	1	
Content beyond syllabus		5.7	Hadoop Ecosystem Components	T1, R1	1	
				Total	12	
				Cumulative Proposed Periods	64	

Text Books:

S.No.	AUTHORS, BOOK TITLE, EDITION, PUBLISHER, YEAR OF PUBLICATION
1	Seema Acharya, Subhashini Chellappan, "Big Data and Analytics", Wiley Publication, 2015.
2	Tom White, "Hadoop: The Definitive Guide", O'Reilly, 4th Edition, 2015.
3	Amit Kumar Tyagi,Data Science and Data Analytics: Opportunities and Challenges, 1st Edition, CRC Press, 2021.

Reference Books:

S.No.	AUTHORS, BOOK TITLE, EDITION, PUBLISHER, YEAR OF PUBLICATION
1	Alex Holmes, Hadoop in Practice, 2nd Edition, MANNING Publ.,2014
2	Vignesh Prajapati, Big Data Analytics with R and Hadoop, 2nd Edition, PACKT Publ.,2013

Web Details:

1	Hadoop: http://hadoop.apache.org/
2	Hive: https://cwiki.apache.org/confluence/display/Hive/Home
3	Piglatin: http://pig.apache.org/docs/r0.7.0/tutorial.html

		Name	Signature with Date
i	Faculty	Dr. RVVSV Prasad	
ii	Module Coordinator	Dr. RVVSV Prasad	
iii	Programme Coordinator	Dr. RVVSV Prasad	5/6/21

Principal