Sardar Patel Institute of Technology

Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India
(Autonomous Institute Affiliated to University of Mumbai)

Course Code	Course Name	Teaching Scheme (Hrs/week)			Credits Assigned			
		L	T	P	L	T	P	Total
CE912	Big Data Analytics and Management(BDAM)	4	--	--	4	--	--	4
		Examination Scheme						
		ISE	MSE	ESE				
		10	30	100 (60% Weightage)				

Pre-requisite Course Codes	Core Java, awareness of RDBMS is desirable	
At the end of successful completion of the course, students will be able to		
Course Outcomes	CO1	Understand the basic concepts of Big Data and Hadoop as processing platforms for Big Data
	CO2	Understand the need of Map Reduce and to develop Mapper, Reducer tasks
	CO3	To understand Text Analytics, Recommendation System and Clustering approaches
	CO4	Understand concept of data streams, Link Analysis, Social Mining Graphs and its real life applications
	CO5	Learn about the different options for importing or loading data into HDFS data sources such as relational databases, data warehouses, web server logs

Module No.	Unit No.	Topics	Ref.	Hrs.
1		Introduction to Big Data and Hadoop	2	10
	1.1	Hadoop Ecosystem, Hadoop Architecture(Name Node, Job Tracker, Task Tracker, Data Node, Secondary Name Node), JobTracker functionality , Namenode Backup(SNN)		
	1.2	Apache Hadoop and Hadoop Ecosystem, HDFS Storage,	2	
	1.3	Hadoop File System APIs, Anatomy of a File Read, Anatomy of a File Write, Rack Awareness		
2		Developing Map Reduce	1,2	12
	2.1	Distributed Computing Concept (Map and Reduce), Anatomy of a MapReduce Job Run(MR1), Running on a cluster, Packaging, Launching a Job, The MapReduce Web UI , Retrieving the Results 167, Debugging a Job 169	2	
	2.2	Map Reduce Algorithms, Matrix-Vector Multiplication, Map Reduce and Relational Operators, Matrix Multiplication of Large Matrices, Shuffle and Sort,	1	
	2.3	Hadoop Logs, Remote Debugging, Advanced Map Reduce	2	

		Concepts, Combiner, Partitioner, Distributed Cache(Map Side Join), Reduce Side join		
3		Clustering Approaches, Text Analytics and Recommendation System	1	10
	3.1	CURE Algorithm, Stream-Computing , A Stream-Clustering Algorithm, Initializing & Merging Buckets, Answering Queries		4
	3.2	Introducing text mining, text mining techniques, Understanding Text Mining Process, Sentiment Analysis		3
	3.3	Introduction to RS, content based RS, collaborative RS, hybrid RS. Issues and challenges RS, examples of real word RS, e.g., Amazon, mobile RS, etc.		3
4	**4.1**	Mining Data Streams : Introduction, The Stream Data Model Sampling Data in a Stream : Obtaining a Representative Sample , The General Sampling Problem, Filtering Streams: The Bloom Filter, Analysis., Counting Distinct Elements in a Stream, Counting Ones in a Window:		5
	4.2	Link Analysis : PageRank Definition, Structure of the web, dead ends, Using Page rank in a search engine, Efficient computation of Page Rank: PageRank, Topic sensitive Page Rank, link Spam, Hubs and Authorities.		3
	4.3	Mining Social Nework Graphs : Mining Social-Network Graphs 11.1 Social Networks as Graphs, Clustering of Social-Network Graphs, SimRank		2
5		Managing Big Data	2, 3	10
	5.1	Moving Data into Hadoop • Load Scenarios 1. Understand how to load data at rest, in motion 2. Understand how to load data from common data sources e.g. RDBMS • Using Sqoop 1. Import data from a relational database table into HDFS 2. Use Sqoop import and export command		
	5.2	• Flume Overview 1. Describe Flume and its uses 2. How Flume works • Using Flume 1. List the Flume configuration components 2. Describe how to start and configure a Flume agent		
	5.3	• Introduction to Oozie Workflows 1. Explain the use for Oozie workflows 2. Describe a workflow 3. List some of the workflow elements • Oozie Coordinator		

		1. Explain the use for the Oozie coordinator 2. List some of the coordinator elements 3. Describe how to submit a workflow job and a coordinator job		
			Total	**52**

References:

[1] Jure Leskovec, Anand Rajaraman, Jeffrey Ullman, "Mining Massive Datasets", Cambridge University Press, 2nd Edition.

[2] Tom White, "Hadoop, the Definitive Guide", O'Reilly, Yahoo Press, 3rd Edition.

[3] Tanmay Deshpande, "Hadoop Real-World Solutions Cook Book", Packt Publishing, 2nd Edition.