

**VELAGAPUDI RAMAKRISHNA
SIDDHARTHA ENGINEERING COLLEGE::VIJAYAWADA
(AUTONOMOUS)
Detailed Syllabus
DEPARTMENT OF IT**

Class	B Tech	Regulation	VR20
Subject Code	20IT6404B	Year & Semester	III/IV B.Tech & VI SEM
Title of the Subject	BIG DATA		

Unit No	Content/Topics Covered (mention Sub Topics as found in books)	Text Book	Chapter/ Section No.	Page Number
Unit I	Introduction to Big Data	[T1]	1.1	3-4
	Big Data-definition		1.1.1	4
	Characteristics of Big Data		1.1.2	5-9
	Data in the warehouse and data in hadoop		1.1.3	9-12
	Why is Big Data Important		1.1.4	15-17
	Patterns in Big Data		1.1.5	17-34
	Introduction to Hadoop	[T2]	1.2	1
	Data		1.2.1	1-3
	Data Storage and Analysis		1.2.2	3-4
	Comparison with Other Systems		1.2.3	4-8
	A Brief History of Hadoop		1.2.4	9-12
	Apache Hadoop and the Hadoop Ecosystem		1.2.5	12-13
	Hadoop Releases		1.2.6	13-14
Unit II	Introduction to NOSQL	[T3]	4.1.1-4.1.2	58-59
	Types of NoSQL Databases		4.1.3	59-60
	Advantages of NoSQL databases		4.1.5	60-61
	SQL versus NoSql		4.1.9	63
	NoSQL databases			
	Introduction to MongoDB		6.1,6.2	110-115
	Data types in MongoDB		6.4	116-117

	MongoDB query language		6.5.1-6.5.5	120-136
	HDFS		2.1	45
	The Hadoop Distributed Filesystem		2.1.1	45
	The Design of HDFS		2.1.2	45-46
	HDFS Concepts		2.1.3	47
	Blocks		2.1.3.1	47-48
	Name nodes and Data nodes		2.1.3.2	48-49
	Basic Filesystem Operations	[T2]	2.1.4	52-53
	Hadoop Filesystems		2.1.5	54-55
	Interfaces		2.1.5.1	55-57
	The Java Interface		2.1.6	57
	Reading Data from a Hadoop URL		2.1.6.1	57-58
	Data Flow		2.1.7	69
	Anatomy of a File Read		2.1.7.1	69-72
	Anatomy of a File Write		2.1.7.2	72-75
	Coherency Model		2.1.7.3	75-76
Unit III	Mapreduce		3	17
	A Weather Dataset		3.1	17
	Data Format		3.1.1	17-19
	Analyzing the Data with UNIX Tools		3.2	19-20
	Analyzing the Data with Hadoop		3.3	20
	Map and Reduce		3.3.1	20-22
	Java MapReduce		3.3.2	22-30
	Scaling Out		3.4	30-37
	Hadoop Streaming		3.5	37-41
	Hadoop Pipes		3.6	41-43
	Pig		3.7	365
	Installation and Running of Pig		3.7.1	366
	Execution Types		3.7.1.1	366-368
	Local Mode		3.7.1.1.1	367
	Mapreduce Mode		3.7.1.1.1	367-368

		[T2]	3.7.1.2	368
	Grunt		3.7.1.3	368-369
	Pig Latin Editors		3.7.1.4	369
	An Example		3.7.2	369-372
	Comparison with Databases		3.7.2.1	372-373
	Pig Latin		3.7.3	373
	Structure		3.7.3.1	373-374
	Schemes		3.7.3.2	382-385
	Functions		3.7.4	386-387
	Eval function		3.7.4.1	386
	Filter function		3.7.4.2	386
	Load function		3.7.4.3	387
	Data Processing Operators		3.7.5	397
	Loading and Storing Data		3.7.5.1	397
	Filtering Data		3.7.5.2	397
Unit IV	HIVE		4	411
	Installing Hive		4.1	412
	The Hive Shell		4.1.1	413-414
	An Example		4.2	414-415
	Running Hive		4.3	415
	Configuration Hive		4.3.1	415-417
	Hive Services	[T2]	4.3.2	417-419
	Comparison with Traditional Databases		4.4	421-422
	HiveQL		4.5	422-427
	Tables		4.6	427-441
	Querying Data		4.7	441-448
	Spark			
	Introduction to data analytics with Spark, Spark Stack		1	16-24
	Programming with RDDS	[R1]	3	46-52
	Working with key/value pairs		4	75-79

	Configuration Hive		4.3.1	415-417
	Hive Services		4.3.2	417-419
	Comparison with Traditional Databases		4.4	421-422
	HiveQL		4.5	422-427
	Tables		4.6	427-441
	Querying Data		4.7	441-448
	Spark			
	Introduction to data analytics with Spark, Spark Stack	[R1]	1	16-24
	Programming with RDDS		3	46-52
	Working with key/value pairs		4	75-79

	Spark SQL		9	214-220
--	-----------	--	---	---------

Text Book(s):

- [1]. Dirk deRoos, Chris Eaton, George Lapis, Paul Zikopoulos, Tom Deutsch, "Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data" 1st Edition, TMH, 2012.
 [2]. Tom White, Hadoop, "The Definitive Guide", 3rd Edition, O'Reilly Publications, 2012.
 [3]. Seema Acharya, Subhashini Chellappan, Big Data and Analytics, Wiley Publishers.

Reference Books:

- [1]. Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia, "Learning Spark: Lightning-Fast Big Data Analysis", O'Reilly Media, Inc.