

Course Code	Course Name	Group	Teaching Scheme	Credits
			Lectures	
RJVSCDSAI351	Big Data	VSC	2	2
<u>Course Outcome:</u> 1. To provide knowledge of basic and advanced methods of big data technology and tools. 2. To provide the knowledge of MapReduce, Hadoop and its ecosystem. 3. To provide hands-on training that enable effective participation in big data projects. <u>Learning Outcome:</u> Upon completion of this course, the student should be able to 1. Apply Hadoop ecosystem components. 2. Build and maintain reliable, scalable and distributed systems with Apache Hadoop. 3. Apply big data concepts to various use cases. 4. Develop application using Zookeeper and Monitoring the cluster				

UNIT	TOPICS
UNIT - I	INTRODUCTION TO BIG DATA Introduction: Characteristics of Data, and Big Data Evolution of Big Data, Definition of Big Data, Challenges with big data Distributed file system, Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications. Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce. INTRODUCTION HADOOP Big Data: Apache Hadoop & Hadoop Ecosystem, Moving Data in and out of Hadoop, Understanding inputs and outputs of MapReduce, Data Serialization.
UNIT - II	HADOOP ARCHITECTURE Hadoop Architecture, Hadoop Storage: HDFS, Common Hadoop Shell commands, Anatomy of File Write and Read., Name Node, Secondary Name Node, and Data Node, Hadoop MapReduce paradigm, Map and Reduce tasks, Job, Task trackers, Cluster Setup, SSH & Hadoop Configuration, HDFS Administering, Monitoring & Maintenance.
UNIT - III	HADOOP ECOSYSTEM AND YARN Hadoop ecosystem components: Schedulers, Fair and Capacity, Hadoop 2.0 New Features Name Node High Availability, HDFS Federation, MRv2, YARN, Running MRv1 in YARN.
UNIT - IV	Hive and HiveQL, HBase Hive Architecture and Installation, Comparison with Traditional Database. HiveQL Querying Data, Sorting and Aggregating, Map Reduce Scripts, Joins & Subqueries. HBase concepts

	Advanced Usage, Schema Design, Advance Indexing, PIG, Zookeeper, how it helps in monitoring a cluster, HBase uses Zookeeper and how to Build Applications with Zookeeper.
References 1. “Professional Hadoop Solutions”, Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, Wiley, ISBN: 9788126551071, 2015. 2. “Understanding Big data”, Chris Eaton, Dirk deroos et al, McGraw Hill, 2012. 3. “HADOOP: The definitive Guide”, Tom White, O Reilly 2012. 6 IT2015 SRM(E&T). 4. “Big Data Analytics with R and Hadoop”, Vignesh Prajapati, Packet Publishing 2013. 5. “Oracle Big Data Handbook”, Tom Plunkett, Brian Macdonald et al, Oracle Press, 2014. 6. “Big Data and Business analytics”, Jy Liebowitz, CRC press, 2013. 7. http://www.bigdatauniversity.com/	