Big Data Analysis	L	T	P	Database Management system
	3	0	2	

Course Objective: Mastering the process of mapping and knowlegde extraction from huge volumes of data.

S. NO	Course Outcomes (CO)
CO1	To learn about distributed file system.
CO2	To understand the working of Apache Hadoop ecosystem.
CO3	To underatand working and commands of Hdoop.
CO4	To study usgaes and design og Hbase concepts.
CO5	To apply big data analytics in real life problem solving .

S. NO	Contents	Contact Hours
UNIT 1	Introduction – distributed file system – Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications. Algorithms using map reduce.	**8**
UNIT 2	Big Data – Apache Hadoop & Hadoop EcoSystem – Moving Data in and out of Hadoop – Understanding inputs and outputs of MapReduce - Data Serialization.	**8**
UNIT 3	Hadoop Architecture, Hadoop Storage: HDFS, Common Hadoop Shell commands , Anatomy of File Write and Read., NameNode, Secondary NameNode, and DataNode, Hadoop MapReduce paradigm, Map and Reduce tasks, Job, Task trackers - Cluster Setup – SSH & Hadoop Configuration – HDFS Administering –Monitoring & Maintenance.	**9**

UNIT 4	HBase concepts- Advanced Usage, Schema Design, Advance Indexing - PIG, Zookeeper - how it helps in monitoring a cluster, HBase uses Zookeeper	**9**
UNIT 5	Data Analytics with R: Machine Learning: Introduction, Supervised Learning, Unsupervised Learning, Collaborative Filtering. Big Data Analytics with BigR.	**8**
	TOTAL	**42**