

**Objectives:** To provide knowledge about scalable, reliable and distributed storage environment for huge amount of data. To provide insights of data analysis in distributed environment. To optimize business decisions and create competitive advantage with Big Data analytics. To introduce programming tools PIG & HIVE in Hadoop ecosystem. To be able to reasonably assess the application of big data analytics technologies for different usage scenarios

**Unit I:** Introduction to Big Data: Introduction– distributed file system–Big Data and its importance, Four Drivers for Big data, Big data analytics, Big data applications. Algorithms using map reduce

**Unit II:** Introduction to Hadoop and its Architecture: Big Data – Apache Hadoop & Hadoop Ecosystem, Moving Data in and Hadoop – Understanding inputs and outputs of MapReduce Serialization

**Unit III:** HDFS, HIVE AND HIVEQL, HBASE: HDFS-Overview, Installation and Shell, Java API; Hive Architecture Installation, Comparison with Traditional Database, HiveQL Querying Sorting And Aggregating, Map Reduce Scripts, Joins & Sub queries, concepts, Advanced Usage, Schema Design, Advance Indexing, Zookeeper , how it helps in monitoring a cluster, HBase uses Zookeeper how to Build Applications with Zookeeper

**Unit IV:** SPARK: Introduction to Data Analysis with Spark, Downloading Spark and Started, Programming with RDDs, Machine Learning with MLlib

**Unit V:** NoSQL: Types of NoSQL databases, Advantages of NoSQL, Use of NoSQL in Industry, SQL vs NoSQL, NewSQL

**Unit VI:** Database for the Modern Web: Introduction to MongoDB key features, Core Server tools, MongoDB through the JavaScript's Shell, Creating and Querying through Indexes, Document-Oriented, principles of schema design, Constructing queries Databases, collections and Documents, MongoDB Query Language

### Reference Books

1. Professional Hadoop Solutions by Boris lublinsky, Kevin t. Smith, Alexey Yakubovich Wiley, ISBN-13: 978-1118611937, ISBN-10: 1118611934, 2015
2. Understanding Big data by Chris Eaton and Paul C. Zikopoulos McGraw Hill, 2012
3. Big Data and Analytics by Seema Acharya, Subhashini Chhellappan Willey, Second edition
4. MongoDB in Action by Kyle Banker, Piter Bakkum , Shaun Verch DreamTech, ISBN 9781935182870, Second edition
5. HADOOP: The definitive Guide by Tom White O'Reilly, 2012
6. Big Data Analytics with R and Hadoop by Vignesh Prajapati Packet Publishing, 2013