

# INT312:BIG DATA FUNDAMENTALS

L:2 T:0 P:2 Credits:3

**Course Outcomes:** Through this course students should be able to

CO1 :: restate the need and importance of fundamental concepts and principles of Big Data

CO2 :: examine internal functioning of different modules of Big Data and Hadoop

CO3 :: understand the big data ecosystem and appreciate its key components

CO4 :: deduce how Hadoop solves those Big Data problems

CO5 :: apply tools and techniques to analyze Big Data

CO6 :: explore solution for a given problem using suitable Big Data Techniques

## Unit I

**Introduction to Hadoop** : Introduction to Big Data, Types of Data, V's of Big Data, Introduction to Hadoop, Components of Hadoop, Installation of Apache Hadoop

## Unit II

**Hadoop Architecture** : Hadoop Architecture, Hadoop Storage: HDFS, Hadoop MapReduce paradigm, MapReduce Terminology, Hadoop - Namenode, DataNode, Job Tracker and TaskTracker

## Unit III

**Map Reduce and YARN** : the MapReduce model v1, limitations of Hadoop 1 and MapReduce 1, review of the Java code required to handle the Mapper class, the Reducer class, and the program driver needed to access MapReduce, the YARN model, comparison of YARN / Hadoop 2 / MR2 vs Hadoop 1 / MR1

## Unit IV

**Introduction to Apache Hive** : hive installation, hive data types, hive bucketing, hive partitioning, hiveql operations, hive operators

## Unit V

**Introduction to Apache HBase** : Installation of Apache Hbase, Hbase Fundamentals and Hbase Data Model, Hbase Architecture, General Commands in Apache Hbase, Interacting with Hbase using Java API, MapReduce with HBase

## Unit VI

**Introduction to Apache Pig** : apache pig installation, pig latin basics, load and store operations in apache pig, pig diagnostic operators, grouping and joining in pig, combining and splitting in pig, pig filtering, pig sorting, pig latin built in functions

## List of Practicals / Experiments:

### Practical List

- Installation of apache hadoop
- Execution of hdfs commands
- Wordcount problem solving using apache hadoop
- Partitioning and Bucketing in Hive
- Interacting with Hbase using Java API
- Data Analysis using Apache Pig
- Java code required to handle the Mapper class, the Reducer class, and the program driver needed to access MapReduce

**Text Books:** 1. BIG DATA 2E by ANIL MAHESHWARI, Tata McGraw Hill, India

**References:** 1. BIG DATA ANALYTICS by RAJ KAMAL, PREETI SAXENA, Tata McGraw Hill, India

