

INT312:BIG DATA FUNDAMENTALS

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

Course Outcomes: Through this course students should be able to

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

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

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

CO4 :: evaluate how hadoop solves big data problems

CO5 :: apply tools and techniques to analyze Big data

CO6 :: examine 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, word count on command line

Unit III

Map Reduce and YARN : 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, Java code for word count, sum of even numbers, palindrome, factorial, Armstrong number

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, Filtering in HBase-Prefix and Single value column, TTL for columns in HBase

Unit VI

Introduction to Apache Cassandra : Installation of Apache Cassandra, Cassandra Architecture – peer-to-peer architecture, gossip protocol, replication and consistency levels, read and write paths, Cassandra Data Model – keyspace, table, primary key, partition key, clustering columns, wide rows; CQL (Cassandra Query Language) – data types, creating keyspaces and tables, insert, update, delete, select, filtering,, indexing; Cassandra Administration - compaction.

List of Practicals / Experiments:

Practical List

- Execution of hdfs commands
- Wordcount problem solving using apache hadoop
- Java code required to handle the Mapper class, the Reducer class, and the program driver needed to access MapReduce
- Perform partitioning in Hive
- Perform bucketing in Hive
- Execution of Cassandra Commands
- Execution of HBase shell commands

Text Books: 1. BIG DATA by ANIL MAHESHWARI, MCGRAW HILL EDUCATION

References: 1. BIG DATA AND ANALYTICS by SEEMA ACHARYA, SUBHASHINI CHELLAPPAN, WILEY

