

# CAP457:INTRODUCTION TO BIG DATA - LABORATORY

L:0 T:0 P:2 Credits:1

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

CO1 :: learn the fundamental Big Data analytics concepts

CO2 :: understand how to use Cloudera with virtual box to get the benefits of virtualization in real time

CO3 :: use the Big Data framework to generate and obtain data for analysis

CO4 :: explore the newest tools for Big Data that are available

## List of Practicals / Experiments:

### Introduction to Big Data Tools

- Introduction about Tools used in Big Data Analysis

### Introduction to Virtual Box

- learning usage of virtualboxinstalling and working with virtualbox
- Installation of cloudera (for HADOOP, Hive installation)

### Introduction to Hadoop Administration

- Hadoop - Environment SetupHadoop - HDFS Overview
- File Management in HADOOP
- Understanding the layers of HADOOP HDFS, YARN and Map Reduce

### Working with HIVE

- Introduction to Data Types in Hive , Creating Data tables, Alter Table, Drop Table, inserting data into the tables, creating views, Apply various functions.

### MAP REDUCE Programming

- Understand the concept of Map Reduce, Execute the basic programs using Map Reduce such as Basic Word Count, Matrix Multiplication, Finding average age of Male and Female died in Titanic Disaster and similar ones.

### Introduction to No-SQL with MongoDB (basic Commands)

- Create database / select, Checking database, See all your databases, Show your collections / db's, Delete your collection, Drop selected database, Create collection, Capped option \*(for create collection), Insert Data, Updating Data, Removing data, Applying Projection and Limit, Sorting data

**Text Books:** 1. BIG DATA FUNDAMENTALS by THOMAS ERL, PEARSON

**References:** 1. MINING OF MASSIVE DATASETS by JURE LESKOVEC, ANAND RAJARAMAN, JEFF ULLMAN, STANFORD UNIVERSITY PRESS