

Any Thing &
Every Thing in
Data Science

## BIG DATA HADOOP Syllabus (Associate)

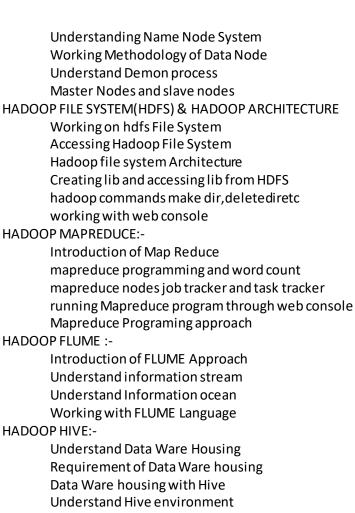
## **Duration: 40 Hours**

## CONTENT

INTRODUCTION OF BIG DATA:-**Basics of Big data Big Data generation Big Data Introduction Big Data Architecture** Understand Bigdata Problem Big data Management Approach Traditional and Current Data storing approach Understand various data formats and data units **Big data With Industry Requirements Big Data Challenges Big Data Complexity Big Data Storage** Big Data performance optimization HADOOP ENVIRONMENT:-**Understand Hadoop Environment Requirement of Hadoop** Importance of Data Analytics Setting up HadoopEnvironment Hadoop advantages over RDMS Understanding RDBMS Approach HADOOP CLUSTER & FILE SYSTEM:-Explaining Various file systems Hdfs GFS, POSIX, GPFS explain clustering methodology Advantages of IBM BIGINSIGHTS over other tools IBM BIGINSIGHTS Approach **RACK Awareness** 



Any Thing & Every Thing in Data Science



HADOOP FLUME :-HADOOP HIVE:working with Hive Query Language Perform DDL approach Through Hive Perform DML approach through Hive

## HADOOP PIG:-

Introduction of PIG **Requirement of Pig** Working with pig Script **Running and managing Pig Script** Perform Streaming Data Analytics through PIG Pig Advantages and Disadvantages

