

### Big Data Hadoop Course Objective

- Complete knowledge of Big Data and Hadoop including HDFS (Hadoop Distributed File System), YARN (Yet Another Resource Negotiator) and MapReduce
- Comprehensive knowledge of various tools that is a part of Hadoop Ecosystem like Pig, Hive, Sqoop, Flume, Oozie, and HBase
- Capability to ingest data in HDFS using Sqoop and Flume, and analyze those large datasets stored in the HDFS.
- The exposure to many real world industry-based projects which will be executed in CloudLab
- Projects which are very diverse i.e. different from each other covering various data sets from multiple domains such as banking, telecommunication, social media, insurance, and e-commerce.

### Why choose Apponix as a Top PowerShell Scripting Training Institute in Bangalore?

- Apponix has highly experienced and qualified Big Data Hadoop instructors.
- Till today we have 100% student satisfaction rate.
- More than 1000 students are rated us as best training institute in Bangalore for Big Data Hadoop.
- Well-equipped lab facility, Decent infrastructure.
- All classrooms are Air-Conditioned.
- All students are provided an individual laptop throughout the course with high-speed WiFi.

**Course Duration: 40 Hours**

### Big Data Hadoop Training Course Content

- **Hadoop Introduction**
  - Introduction to Data and System
  - Types of Data
  - Traditional way of dealing large data and its problems
  - Types of Systems & Scaling
  - What is Big Data
  - Challenges in Big Data
  - Challenges in Traditional Application

- New Requirements
- What is Hadoop? Why Hadoop?
- Brief history of Hadoop
- Features of Hadoop
- Hadoop and RDBMS
- Hadoop Ecosystem's overview
- **Hadoop Installation**
  - Installation in detail
  - Creating Ubuntu image in VMware
  - Downloading Hadoop
  - Installing SSH
  - Configuring Hadoop, HDFS & MapReduce
  - Download, Installation & Configuration Hive
  - Download, Installation & Configuration Pig
  - Download, Installation & Configuration Sqoop
  - Download, Installation & Configuration Hive
  - Configuring Hadoop in Different Modes
- **Hadoop Distribute File System (HDFS)**
  - File System - Concepts
  - Blocks
  - Replication Factor
  - Version File
  - Safe mode
  - Namespace IDs
  - Purpose of Name Node
  - Purpose of Data Node
  - Purpose of Secondary Name Node
  - Purpose of Job Tracker
  - Purpose of Task Tracker
  - HDFS Shell Commands – copy, delete, create directories etc.
  - Reading and Writing in HDFS
  - Difference of Unix Commands and HDFS commands
  - Hadoop Admin Commands
  - Hands on exercise with Unix and HDFS commands
  - Read / Write in HDFS – Internal Process between Client, NameNode & DataNodes
  - Accessing HDFS using Java API
  - Various Ways of Accessing HDFS
  - Understanding HDFS Java classes and methods
  - Commissioning / DeCommissioning DataNode

- Balancer
- Replication Policy
- Network Distance / Topology Script
- **Map Reduce Programming**
  - About MapReduce
  - Understanding block and input splits
  - MapReduce Data types
  - Understanding Writable
  - Data Flow in MapReduce Application
  - Understanding MapReduce problem on datasets
  - MapReduce and Functional Programming
  - Writing MapReduce Application
  - Understanding Mapper function
  - Understanding Reducer Function
  - Understanding Driver
  - Usage of Combiner
  - Usage of Distributed Cache
  - Passing the parameters to mapper and reducer
  - Analysing the Results
  - Log files
  - Input Formats and Output Formats
  - Counters, Skipping Bad and unwanted Records
  - Writing Join's in MapReduce with 2 Input files. Join Types
  - Execute MapReduce Job - Insights
  - Exercise's on MapReduce
- **Hive**
  - Hive concepts
  - Hive architecture
  - Install and configure hive on cluster
  - Different type of tables in hive
  - Hive library functions
  - Buckets
  - Partitions
  - Joins in hive
  - Inner joins
  - Outer Joins
  - Hive UDF
  - Hive Query Language

- **PIG**
  - Pig basics
  - Install and configure PIG on a cluster
  - PIG Library functions
  - Pig Vs Hive
  - Write sample Pig Latin scripts
  - Modes of running PIG
  - Running in Grunt shell
  - Running as Java program
  - PIG UDFs
- **Sqoop**
  - Install and configure Sqoop on cluster
  - Connecting to RDBMS
  - Installing Mysql
  - Import data from Mysql to hive
  - Export data to Mysql
  - Internal mechanism of import/export
- **HBase**
  - HBase concepts
  - HBase architecture
  - Region server architecture
  - File storage architecture
  - HBase basics
  - Column access
  - Scans
  - HBase use cases
  - Install and configure HBase on a multi node cluster
  - Create database, Develop and run sample applications
  - Access data stored in HBase using Java API
  - Map Reduce client to access the HBase data
- **YARN**
  - Resource Manager (RM)
  - Node Manager (NM)
  - Application Master (AM)