

## Hadoop Course



### **Course Curriculum and Duration:**

We provide both Class-room Training and Offline Training.

Duration: 50Hrs.

Week days (Mon- Fri 2Hrs per day)

Weekend batches (Sat-Sun 4Hrs per day)

### **About INNOVATIVE ACADEMY**

INNOVATIVE ACADEMY is a best training institute in Bangalore. We providing classroom led training live instructor-led interactive online training and cooperate training. We cater to professionals and students across the globe in categories like AWS, Angular JS, JAVA, DOT NET, MCSA, CCNA, A+, N+, Databases, VMware, Mobile Technologies, Project Management and Programming.

### **About Course**

Hadoop tutorial will help you to understand the problem with traditional system while processing Big Data and how Hadoop solves it and will provide you a brief idea about HDFS, YARN, PIG, HBASE, HIVE and SPARK using examples and practical demonstration.

## HADOOP SYLLABUS

### 1. Understanding Big Data and Hadoop

**Learning Objectives** - In this module, you will understand Big Data, the limitations of the existing solutions for Big Data problem, how Hadoop solves the Big Data problem, Hadoop ecosystem components, Hadoop Architecture, HDFS, Anatomy of File Write and Read, how MapReduce Framework works.

**Topics** - Big Data, why big data, big data characteristics, types of big data, Examples, applications of big data, challenges with big data, Features, Hadoop Ecosystem, Hadoop 2.x core components, Hadoop Storage: HDFS, Hadoop Processing: MapReduce Framework, Hadoop installation.

**Practical's to be covered**- linux installation and some basic linux commands

### 2. Hadoop Architecture and HDFS

**Learning Objectives** - In this module, you will learn the Hadoop Cluster Architecture, Important Configuration files in a Hadoop Cluster, Data Loading Techniques, how to setup single node hadoop cluster.

**Topics** – HDFS, features of HDFS, HDFS architecture, advantage of hdfs, HDFS operations, HDFS commands.

**Practical's to be covered**- hadoop installation ,hadoop commands, read and write a file from hdfs.

### 3. Hadoop MapReduce Framework

**Learning Objectives** - In this module, you will understand Hadoop MapReduce framework and the working of MapReduce on data stored in HDFS. You will understand concepts like Input Splits in MapReduce, Combiner & Partitioner and Demos on MapReduce using different data sets.

**Topics** - MapReduce, Traditional way Vs MapReduce way, Why MapReduce, Hadoop 2.x MapReduce Architecture, Hadoop 2.x MapReduce Components, YARN MR Application Execution Flow, YARN Workflow, Anatomy of MapReduce Program. Input Splits, Relation between Input Splits and HDFS Blocks, MapReduce: Combiner & Partitioner, Demo on Weather Data set.

**Practical's to be covered**- java installation, program the sample data using MapReduce framework.

#### 4. Advanced MapReduce

**Learning Objectives** - In this module, you will learn Advanced MapReduce concepts such as Counters, Distributed Cache, Reduce Join, Custom Input Format, Sequence Input Format and XML parsing.

**Topics** - Counters, Distributed Cache, Joins in mapreduce, Custom Input Format, Sequence Input Format, Xml file Parsing using MapReduce.

**Practical's to be covered-** interact with map reduce jobs, counter and joins with example .

#### 5. Pig

**Learning Objectives** - In this module, you will learn Pig, types of use case we can use Pig, tight coupling between Pig and MapReduce, and Pig Latin scripting, PIG running modes, PIG UDF, Pig Streaming, Testing PIG Scripts.

**Topics** - About Pig, MapReduce Vs Pig, Pig Use Cases, Programming Structure in Pig, Pig Running Modes, Pig components, Pig Execution, Pig Latin Program, Data Models in Pig, Pig Data Types, Shell and Utility Commands, Pig Latin : Relational Operators, File Loaders, Group Operator, COGROUP Operator, Joins and COGROUP, Union, Diagnostic Operators, Specialized joins in Pig, Built In Functions ( Eval Function, Load and Store Functions, Math function, String Function, Date Function, Pig UDF, Piggybank, Parameter Substitution ( PIG macros and Pig Parameter substitution ), Pig Streaming, Testing Pig scripts with Punit.

**Practical's to be covered-** Pig installation, create first pig program, running pig in local mode and map reduce mode.

#### 6. Hive

**Learning Objectives** - This module will help you in understanding Hive concepts, Hive Data types, Loading and Querying Data in Hive, running hive scripts and Hive UDF.

**Topics** - Hive Background, Hive Use Case, About Hive, Hive Vs Pig, Hive Architecture and Components, Metastore in Hive, Limitations of Hive, Comparison with Traditional Database, Hive Data Types and Data Models, Partitions and Buckets, Hive Tables (Managed Tables and External Tables), Importing Data, Querying Data, Managing Outputs, Hive Script, Hive UDF, Retail use case in Hive.

**Practical's to be covered-** Hive installation, hive commands, how to partition a file and its data, buckets examples and hive queries.

## 7. Advanced Hive and HBase

**Learning Objectives** - In this module, you will understand Advanced Hive concepts such as UDF, Dynamic Partitioning, Hive indexes and views, optimizations in hive. You will also acquire in-depth knowledge of HBase, HBase Architecture, running modes and its components.

**Topics** - Hive QL: Joining Tables, Dynamic Partitioning, Custom Map/Reduce Scripts, Hive Indexes and views Hive query optimizers, Hive : Thrift Server, User Defined Functions, HBase: Introduction to NoSQL Databases and HBase, HBase v/s RDBMS, HBase Components, HBase Architecture, Run Modes & Configuration, HBase Cluster Deployment.

**Practical's to be covered-** Hive installation, hive commands, how to partition a file and its data, buckets examples and hive queries.

## 8. Advanced HBase

**Learning Objectives** - This module will cover Advanced HBase concepts. You will also learn what Zookeeper is all about, how it helps in monitoring a cluster, why HBase uses Zookeeper.

**Topics** - HBase Data Model, HBase Shell, HBase Client API, Data Loading Techniques, ZooKeeper Data Model, Zookeeper Service, Zookeeper, Demos on Getting and Inserting Data, Filters in HBase.

**Practical's to be covered-** Hive installation, hive commands, how to partition a file and its data, buckets examples and hive queries.

## 9. Apache Spark

**Learning Objectives** - In this module you will learn Spark ecosystem and its components.

**Topics** - What is Apache Spark, spark-Architecture, Execution, and Related concepts, RDD operations, functional programming in spark.

**Practical's to be covered-** spark installation, verifying spark installation.

## 10. Apache RDDs in spark

**Learning Objectives** - In this module, you will learn RDD concepts, RDD operations and RDD optimization techniques

**Topics** – Introduction, RDD data types and RDD creation, operations in RDDs, RDD optimization techniques introduction, RDD persistence, Exercise and quiz

**Practical's to be covered-** spark shell and simple RDD program.