Hadoop Online training-Complete Course Details

HDFS and MAPREDUCE

- ✧ Introduction to BIG DATA and Its characteristics
- ✧ 4 V's of BIG DATA(IBM Definition of BIG DATA)
- ✧ What is Hadoop?
- ✧ Why Hadoop?
- ✧ Core Components of Hadoop
- ✧ Intro to HDFS and its Architecture
- ✧ Difference b/w Code Locality and Data Locality
- ✧ HDFS commands
- ✧ Name Node's Safe Mode
- ✧ Different Modes of Hadoop
- ✧ Intro to MAPREDUCE
- ✧ Versions of HADOOP
- ✧ What is Daemon?
- ✧ Hadoop Daemons?
- ✧ What is Name Node?
- ✧ What is Data Node?
- ✧ What is Secondary name Node?
- ✧ What is Job Tracker?
- ✧ What is Task Tracker?
- ✧ What is Edge computer in Hadoop Cluster and Its role
- ✧ Read/Write operations in HDFS
- ✧ Complete Overview of Hadoop1.x and Its architecture
- ✧ Rack awareness
- ✧ Introduction to Block size
- ✧ Introduction to Replication Factor(R.F)

- ✧ Introduction to HeartBeat Signal/Pulse
- ✧ Introduction to Block report
- ✧ MAPREDUCE Architecture
- ✧ What is Mapper phase?
- ✧ What is shuffle and sort phase?
- ✧ What is Reducer phase?
- ✧ What is split?
- ✧ Difference between Block and split
- ✧ Intro to first Word Count program using MAPREDUCE
- ✧ Different classes for running MAPREDUCE program using Java
- ✧ Mapper class
- ✧ Reducer Class and Its role
- ✧ Driver class
- ✧ Submitting the Word Count MAPREDUCE program
- ✧ Going through the Jobs system output
- ✧ Intro to Partitioner with example
- ✧ Intro to Combiner with example
- ✧ Intro to Counters and its types
- ✧ Different types of counters
- ✧ Different types of input/output formats in HADOOP
- ✧ Use cases for HDFS & MapReduce programs using Java
- ✧ Single Node cluster Installation
- ✧ Multi Node cluster Installation
- ✧ Introduction to Configuration files in Hadoop and Its Imp.
- ✧ Complete Overview of Hadoop2.x and Its architecture
- ✧ Introduction to YARN
- ✧ Resource Manager

- ✧ Node Manager
- ✧ Application Master(AM)
- ✧ Applications Manager(AsM)
- ✧ Journal Nodes
- ✧ Difference Between Hadoop1.x and Hadoop2.x
- ✧ High Availability(HA)
- ✧ Hadoop Federation

PIG

- ✧ Intro to PIG
- ✧ Why PIG?
- ✧ The difference between MAPREDUCE and PIG
- ✧ When to go with MAPREDUCE?
- ✧ When to go with PIG?
- ✧ PIG data types
- ✧ What is field in PIG?
- ✧ What is tuple in PIG?
- ✧ What is Bag in PIG?
- ✧ Intro to Grunt shell?
- ✧ Different modes in PIG
- ✧ Local Mode
- ✧ MAPREDUCE mode
- ✧ Running PIG programs
- ✧ PIG Script
- ✧ Intro to PIG UDFs
- ✧ Writing PIG UDF using Java
- ✧ Registering PIG UDF
- ✧ Running PIG UDF

- ✧ Different types of UDFs in PIG
- ✧ Word Count program using PIG script
- ✧ Use cases for PIG scripts

HIVE

- ✧ Intro to HIVE
- ✧ Why HIVE?
- ✧ History of HIVE
- ✧ Difference between PIG and HIVE
- ✧ HIVE data types
- ✧ Complex data types
- ✧ What is Metastore and its importance?
- ✧ Different types of tables in HIVE
- ✧ Managed tables
- ✧ External tables
- ✧ Running HIVE queries
- ✧ Intro to HIVE partitions
- ✧ Intro to HIVE Buckets
- ✧ How to perform the JOINS using HIVE queries
- ✧ Intro to HIVE UDFs
- ✧ Different types of UDFs in HIVE
- ✧ Running HIVE queries for Word Count example
- ✧ Use cases for HIVE

HBASE

- ✧ Intro to HBASE
- ✧ Intro to NoSQL database
- ✧ Sparse and dense Concept in RDBMS
- ✧ Intro to columnar/column oriented database
- ✧ Core architecture of HBase
- ✧ Why Hbase?
- ✧ HDFS vs HBase
- ✧ Intro to Regions, Region server and Hmaster
- ✧ Limitations of Hbase
- ✧ Integration with Hive and Hbase
- ✧ Hbase commands
- ✧ Use cases for HBASE

FLUME

- ✧ Intro to Flume
- ✧ Intro to Sink, Source, Flume Master and Flume agents
- ✧ Importance of Flume agents
- ✧ Live Demo on copying LOG DATA into HDFS

SQOOP

- ✧ Intro to Sqoop
- ✧ Importing and exporting the RDBMS into HDFS
- ✧ Intro to incremental imports and its types
- ✧ Use cases to import the Mysql data into HDFS

ZOOKEEPER

- ✧ Intro to Zookeeper
- ✧ Zookeeper operations

OOZIE

- ✧ Intro to Oozie
- ✧ What is Job.properties
- ✧ What is workflow.xml
- ✧ Scheduling the jobs in Oozie
- ✧ Scheduling MapReduce, HIVE,PIG jobs/Programs using Oozie.
- ✧ Setting up the VMware for Hadoop
- ✧ Installing all Hadoop Components
- ✧ Intro to Hadoop Distributions
- ✧ Intro to Cloudera and its major components

SCALA

- ✧ Getting started With Scala.
- ✧ Scala Background, Scala Vs Java
- ✧ Introduction to Scala – REPL
- ✧ Scala data types, variables, simple functions.
- ✧ Intro to Scala compiler
- ✧ Installing Scala on Linux
- ✧ Intro to Functional Programming Language
- ✧ Differences between OOPS and FPP
- ✧ Word count pgm, file handling
- ✧ Running Scala script
- ✧ Intro to Maps, Sets, groupBy, Options, flatten, flatMap and more

SPARK

- ✧ What is Spark Ecosystem
- ✧ Batch vs real time data processing
- ✧ Intro to Spark Architecture
- ✧ Installing Scala on Linux

- ✧ Scala utility in Spark
- ✧ Spark Cluster Managers
- ✧ Spark -Standalone mode Installation
- ✧ Spark on YARN
- ✧ Spark on MESOS
- ✧ What is SparkContext
- ✧ Intro to RDDs
- ✧ Intro to DAG
- ✧ RDD's lineage
- ✧ How to work on RDD in Spark
- ✧ What is transformations and Actions
- ✧ Intro to Spark Streaming(SS)
- ✧ Intro to Discretized Streams RDD
- ✧ Applying Transformations and Actions on Streaming data
- ✧ Intro to Spark Streaming Architecture
- ✧ Applying transformations and Actions on SS data
- ✧ How to run a Spark Cluster
- ✧ Comparison of MapReduce vs Spark
- ✧ Integration of Hadoop and Spark

TABLEAU

- ✧ Tableau Fundamentals
- ✧ Tableau Analytics
- ✧ Visual Analytics
- ✧ Creating different types of WorkSheets,Dashboards and Stories.
- ✧ Connecting with different data sources
- ✧ Hadoop Integration with Tableau

Note:- Various Hands on exercises, Realtime use cases and Assignments on each & every Eco-System and Project Guidance.