

# HADOOP

**Duration: 2-3 Months**

## **Introduction To Hadoop**

- What is Enterprise BIGDATA
- What is Hadoop?
- History of Hadoop
- Hadoop Eco-System
- Hadoop Framework
- Hadoop vs RDBMS
- Hadoop vs SAP Hana vs Teradata
- How ETL tools works in Hadoop
- Hadoop Requirements and supported versions
- Case Studies: Hadoop and Hive at Yahoo, Facebook etc...

## **Hadoop Distributed File Systems**

- Installation of Ubuntu 13.04 \*
- Basic Unix Commands \*
- Hadoop Commands
- HDFS & Job Tracker Access URLs & ports.
- HDFS design
- Hadoop file systems
- Master and Slave node architecture
- Filesystem API – Java
- Serialization in Hadoop – Reading and writing data from/to Hadoop URL

## Administering Hadoop

- Cluster specification
- Hadoop cluster setup and installation
- Standalone
- Pseudo-distributed mode
- Fully distributed mode
- fs, fsck, distcp, archive, ---
- dfsadmin, balancer, jobtracker, tasktracker, namenode---
- Step-by-step multi-node installation
- Hadoop Configuration
- Namenode and datanode directory structure
- User commands
- Administration commands
- Monitoring
- Benchmarking a Hadoop cluster

## Mapreduce

- Map/Reduce Overview and Architecture
- Developing Map/Red Jobs
- Mapreduce Data types
- Custom DataTypes/Writables
- Input File Formats
- Text Input File Format
- Zip File Input Format
- LZO Compression & LZO Input Format
- XML Input Format
- JSON Input Format
- Packaging, Launching, Debugging jobs
- Hash Partitioner
- Custom Partitioner
- Capacity Scheduler
- Fair Scheduler
- Output Formats
- Job Configuration
- Job Submission
- Mapreduce workflows
- Practicing Map Reduce Programs

- Combiner
- Partitioner
- Search
- Sorting
- Secondary Sorting
- Distributed Cache
- Chain Mapping/Reducing
- Scheduling
- One Example for Each Concept\*
- Practical Examples execution on Local, HDFS and Using Eclipse Plugins\* too.

## **HIVE**

- Hive concepts
- Hive installation
- Hive configuration, hive services & metastore
- Hive datatypes – primitive and complex types
- Hive operators
- Hive Builtin functions
- Hive Tables
- creating tables
- External Table
- Internal Table
- Partitions and buckets
- Browsing tables and partitions
- Storage formats
- Loading data
- Joins
- Aggregations and sorting
- Insert into local files
- Altering, dropping tables
- Importing data

## PIG

- Why pig
- Pig and Pig latin
- Pig installation
- Pig latin command
- Pig latin relational operators
- Pig latin diagnostic operators
- Data types and Expressions
- Builtin functions
- Data processing in pig
- load and store
- Filtering the data
- Grouping the data
- Joining the data
- Sorting the data

## Sqoop

- Sqoop installation
- Sqoop commands
- Sqoop connectors
- Importing the data from mysql
- Exporting the data
- Creating hive tables by importing data

## HBase

- HBase Introduction.
- HBase Installation
- HBase Architecture
- Zoo Keeper
- Keys & Column families
- Integration with MapReduce
- Integration with Hive