



- CYU – Check Your Understanding
- LOTW – Learning On The Walls
- PPR – Participant Progress Report
- PTS – Post Training Support (Web)

# HADOOP ADMINISTRATION

**Duration: 2 Days**

## INTRODUCTION TO BIG DATA

- ❖ What is Big Data ?
- ❖ Big Data Facts
- ❖ The Three V's of Big Data

## UNDERSTANDING HADOOP

- ❖ What is Hadoop ?
- ❖ Why learn Hadoop ?
- ❖ Relational Databases Vs. Hadoop
- ❖ Motivation for Hadoop
- ❖ 6 Key Hadoop Data Types

## THE HADOOP DISTRIBUTED FILE SYSTEM (HDFS)

- ❖ What is HDFS ?
- ❖ HDFS components
- ❖ Understanding Block storage
- ❖ The Name Node
- ❖ The Data Nodes
- ❖ Data Node Failures
- ❖ HDFS Commands
- ❖ HDFS File Permissions

## THE MAPREDUCE FRAMEWORK

- ❖ Overview of MapReduce
- ❖ Understanding MapReduce
- ❖ The Map Phase
- ❖ The Reduce Phase
- ❖ WordCount in MapReduce
- ❖ Running MapReduce Job

## PLANNING YOUR HADOOP CLUSTER

- ❖ Single Node Cluster Configuration
- ❖ Multi-Node Cluster Configuration

## CLUSTER MAINTENANCE

- ❖ Checking HDFS Status
- ❖ Breaking the cluster
- ❖ Copying Data Between Clusters
- ❖ Adding and Removing Cluster Nodes
- ❖ Rebalancing the cluster
- ❖ Name Node Metadata Backup
- ❖ Cluster Upgrading

## INSTALLING AND MANAGING HADOOP ECOSYSTEM PROJECTS

- ❖ Sqoop
- ❖ Flume
- ❖ Hive
- ❖ Pig



- CYU – Check Your Understanding
- LOTW – Learning On The Walls
- PPR – Participant Progress Report
- PTS – Post Training Support (Web)

- ❖ HBase
- ❖ Oozie

#### MANAGING AND SCHEDULING JOBS

- ❖ Managing Jobs
- ❖ The FIFO Scheduler
- ❖ The Fair Schedule
- ❖ How to stop and start jobs running on the cluster

#### CLUSTER MONITORING, TROUBLESHOOTING, AND OPTIMIZING

- ❖ General System conditions to Monitor
- ❖ Name Node and Job Tracker Web Uis
- ❖ View and Manage Hadoop's Log files
- ❖ Ganglia Monitoring Tool
- ❖ Common cluster issues and their resolutions
- ❖ Benchmark your cluster's performance

#### POPULATING HDFS FROM EXTERNAL SOURCES

- ❖ How to use Sqoop to import data from RDBMSs to HDFS
- ❖ How to gather logs from multiple systems using Flume
- ❖ Features of Hive, Hbase and Pig
- ❖ How to populate HDFS from external Sources

#### INTRODUCTION TO HBASE

- ❖ What Is HBase?
- ❖ Why Use HBase?
- ❖ HBase and RDBMS
- ❖ The Give and Take of HBase

#### HBASE CONCEPTS

- ❖ HBase Concepts
- ❖ Working with HBase
- ❖ Installation of HBase
- ❖ Java Program to create the table in HBase
- ❖ Managing the HBase Table with Hive

#### THE HBASE ADMINISTRATION

- ❖ HBase Shell
- ❖ Administration Calls
- ❖ Monitoring and optimizing HBase deployments
- ❖ Writing to and reading from HBase