Hadoop Course Content

What is Hadoop?

Hadoop is a software framework for storing and processing Big Data. It is an open-source tool build on java platform and focuses on improved performance in terms of data processing on clusters of commodity hardware.

Hadoop comprises of multiple concepts and modules like HDFS, Map-Reduce, HBASE, PIG, HIVE, SQOOP and ZOOKEEPER to perform the easy and fast processing of huge data

Hadoop conceptually different from Relational databases and can process the high volume, high velocity and high variety of data to generate value

Hadoop is a free, Java-based programming framework which supports the processing of large data sets in a distributed computing environment. It is part of the Apache project sponsored by the Apache Software Foundation.

Course Overview

This course is designed for information workers, administrator and developers who would be managing Bigdata through within the Hadoop® framework.

Course Duration

This Course is designed to be competed in 20-24 hours for corporate trainings and 1 month for the regular class trainings

Course Objectives

This Hadoop & Bigdata Training program will empower the Participants to be able to:

- Understand the basic concepts of Hadoop and Hadoop Distributed File System
- Write Map reduce programs
- Understand using Hive and Hbase for data management and Pig for data analysis

Course Contents

The Motivation for Hadooop

- Problems with traditional large-scale systems
- Problems with traditional large-scale systems

Hadoop Basic Concepts

- Overview of Hadoop
- The Hadoop Distributed File System
- How MapReduce Works
- Anatomy of a Hadoop Cluster
- Other Hadoop Ecosystem Components

Writing a MapReduce Programs

- The MapReduce Flow
- Examining a MapReduce Program
- The Mapper
- The Reducer

Introduction to Hive

- What is Hive?
- Hive Schema and Data Storage
- Comparing Hive to Traditional Database
- Interaction with Hive

HiveQL: Data Definition

- Databases in Hive
- Data Types
- Alter Database
- Create – Alter – Drop tables
- Partitioned, Managed Tables

HiveQL: Data Management

- Running Hive Queries in Shell and Hue
- Loading data into Hive tables
- Querying Hive using HiveQL
- Common Built-in Functions

Introduction to Pig

- What is Pig?
- Pig Features
- Interacting with Pig

Basic Data Analysis Using Pig

- Pig Latin Syntax
- Loading Data
- Simple Data Types
- Field Definitions
- Data Output
- Viewing the Schema
- Filtering and Sorting Data
- Commonly-Used Functions

Introduction to other projects from ecosystem and their relevance

- Sqoop
- ZooKeeper
- Oozie
- Hbase

Hadoop Course
Content

Hands On Table

S.no	Hands on Table Hadoop	Comments
1	Installing and Configuring hadoop in psedo mode	
2	Installing and Configuring Hadoop in distributed mode	Depending on Infra availibility
3	Playing with hadoop Commands	
4	Hive Queries and Data Modelling	
5	Hbase and Pig Hands On	
6	Hands on Scoop	
7	Learning analytic tools of your choice for hadoop	Depending on Infra availibility
8	Scripts for automating work.	