

Hadoop Administration

Duration: 4 Days

Description

In this Hadoop Administration training class, students learn all about working with Hadoop and HDFS.

Delegates will learn

- the fundamental concepts of Hadoop.
- to plan your Hadoop cluster.
- HDFS features.
- how to get data into HDFS.
- to work with MapReduce.
- installation and configuration of Hadoop.
- cluster maintenance.

Outline

Hadoop Overview

- What is Big Data?
- How did we get to this point?
- How does Hadoop compare to a relational database system?
- Big Data Introduction
- History
- Comparison to Relational Databases
- Hadoop Ecosystem

HDFS

- Architecture/Concepts
- Access
- Namenodes
- Filesystem Shell
- Accessing HDFS with Java

- Reading/Writing/Browsing file system

HBASE

- Overview
- Architecture
- Data Model
- Installation and Shell
- Access via Java API
- Administration access via Java
- Scan API
- Filters
- Storage Model
- Table Design

Map Reduce on YARN

- Introduction
- Processing Model
- Command line tools
- MapReduce framework
- Submitting MapReduce Jobs
- Writing MapReduce jobs in Java
- MapReduce Theory
- Distributive Cache
- Speculative Executin
- YARN Components
- Counters
- Details of MapReduce Job Execution

Hadoop Streaming

- Implementing a streaming job
- Counters in streaming jobs
- Contrast with Java Jobs

MapReduce Workflows

- Problem decomposition into MapReduce Jobs
- Coding workflows
- Using the JobControl Class

Oozie

- Oozie Installation
- Writing Oozie workflows
- Deploying and running Oozie jobs

Pig

- Installation
- Pig Latin
- Writing Pig Scripts
- User Defined functions
- Data set joins

Hive

- Installation
- Table creation and deletion
- Partitioning
- Loading data into Hive
- Joins
- Bucketing

Prerequisites

There is no prerequisite