



Cloudera Administration

Course ID: ITG-BIG-210

ITG Software Engineering

Cloudera Administration

ITG-BIG-210

Course Overview:

This 3 day course introduces content for system administrators and others responsible for managing a Hadoop Cluster in production or development environments. Attendees will be introduced to the Apache Hadoop architecture, HDFS, Yarn, MapReduce and Hadoop configuration. You will also explore Hadoop clients, cluster maintenance and scheduling.

Prerequisites:

Basic Linux knowledge and experience as a systems administrator.

Who Should attend this course?

Those who have an interest in learning and managing the Cloudera framework.

Topics:

• Apache Hadoop Overview	• HDFS
• Yarn	• MapRequest
• Hadoop Clusters	• Hadoop Configurations
• Hadoop Clients	• Hadoop Security
• Cloudera Manager	• Cluster Maintenance
• Scheduling Jobs	• Managing Jobs
• Flume	• Sqoop
• Hlve	• Pig
• Impala	• HUE
• Fairscheduler	• Troubleshooting & Monitoring

Module 01: Introduction to Hadoop

- What is Hadoop
- Hadoop Components
- Hadoop's main concepts
- Hadoop Architecture

Module 02: HDFS

- Writing & Reading Files
- NameNode Memory Considerations
- NameNode Web UI
- Using the Hadoop File Shell

Module 03: Loading Data into HDFS

- Ingesting Data from external resources (Flume)
- Ingesting Data from Relational databases with Sqoop
- Rest Interfaces
- Best Practices for importing data

Module 04: Using Yarn & MapReduce

- MapReduce Overview
- Yarn Clusters
- Recovering from failures
- Yarn web interface

Module 05: Planning Hadoop Clusters

- Hardware choices
- Network requirements
- Node configuration
- Managing clusters

Module 06: Installation & Configuration

- Deployment Types
- Hadoop Installation
- Hadoop Configuration
- HDFS, YARN, MapReduce configurations

Module 07: Configuration in Depth

- Hadoop logging
- Setting up Ecosystem products (Hive, Pig, Impala)
- Hive overview
- Pig & Impala overview

Module 08: Hadoop Clients

- What is an Hadoop client
- Setting up and configuring Hadoop clients
- HUE: Installation & configuration
- HUE: Authentication & Authorization

Module 09: Cloudera Manager

- Features
- Express & Enterprise versions
- Topology
- Installation & Administrative tasks

Module 10: Cluster Configuration

- Advanced configuration parameters
- Configuring Hadoop ports
- Rack awareness
- High availability

Module 11: Hadoop Security

- Hadoop Security Overview
- Security related system concepts
- Securing Hadoop clusters
- Kerberos

Module 12: Job Scheduling

- Manage running jobs
- Schedule jobs in Hadoop
- FairScheduler
- Scheduling Impala Queries

Module 13: Cluster Maintenance

- HDFS Status
- Copying data from one cluster to another
- Adding & deleting clusters
- Rebalancing clusters & upgrading clusters

Module 14: Troubleshooting & Monitoring

- System monitoring overview
- Monitoring Hadoop clusters
- Troubleshooting basics
- Most common configuration errors