CLOUDERA ADMINISTRATOR TRAINING(ADMIN-OD)

The Case for Apache Hadoop

- Why Hadoop?
- Fundamental Concepts
- Core Hadoop Components

Hadoop Cluster Installation

- Rationale for a Cluster Management Solution
- Cloudera Manager Features
- Cloudera Manager Installation
- Hadoop (CDH) Installation

The Hadoop Distributed File System (HDFS)

- HDFS Features
- Writing and Reading Files
- NameNode Memory Considerations
- Overview of HDFS Security
- Web UIs for HDFS
- Using the Hadoop File Shell

MapReduce and Spark on YARN

- The Role of Computational Frameworks
- YARN: The Cluster Resource Manager
- MapReduce Concepts
- Apache Spark Concepts
- Running Computational Frameworks on YARN
- Exploring YARN Applications Through the Web UIs, and the Shell
- YARN Application Logs

Hadoop Configuration and Daemon Logs

- Cloudera Manager Constructs for Managing Configurations
- Locating Configurations and Applying
- Configuration Changes
- Managing Role Instances and Adding Services
- Configuring the HDFS Service
- Configuring Hadoop Daemon Logs
- Configuring the YARN Service

Getting Data Into HDFS

- Ingesting Data From External Sources With Flume
- Ingesting Data From Relational Databases With Sqoop
- REST Interfaces
- Best Practices for Importing Data

Planning Your Hadoop Cluster

- General Planning Considerations
- Choosing the Right Hardware
- Virtualization Options
- Network Considerations
- Configuring Nodes

Installing and Configuring Hive, Impala, and Pig

- Hive
- Impala
- Pig

Hadoop Clients Including Hue

- What Are Hadoop Clients?
- Installing and Configuring Hadoop Clients
- Installing and Configuring Hue
- Hue Authentication and Authorization

Advanced Cluster Configuration

- Advanced Configuration Parameters
- Configuring Hadoop Ports
- Configuring HDFS for Rack Awareness
- Configuring HDFS High Availability

Hadoop Security

- Why Hadoop Security Is Important
- Hadoop's Security System Concepts
- What Kerberos Is and how it Works
- Securing a Hadoop Cluster With Kerberos
- Other Security Concepts

Managing Resources

- Configuring cgroups with Static Service Pools
- The Fair Scheduler
- Configuring Dynamic Resource Pools
- YARN Memory and CPU Settings
- Impala Query Scheduling

Cluster Maintenance

- Checking HDFS Status
- Copying Data Between Clusters
- Adding and Removing Cluster Nodes
- Rebalancing the Cluster
- Directory Snapshots
- Cluster Upgrading

Cluster Monitoring and Troubleshooting

- Cloudera Manager Monitoring Features
- Monitoring Hadoop Clusters
- Troubleshooting Hadoop Clusters
- Common Misconfigurations