

CLUDERA ADMINISTRATOR TRAINING FOR APACHE HADOOP(HADOOP-ADMIN)

The Cloudera Enterprise Data Hub

- Cloudera Enterprise Data Hub
- CDH Overview
- Cloudera Manager Overview
- Hadoop Administrator Responsibilities

Installing Cloudera Manager and CDH

- Cluster Installation Overview
- Cloudera Manager Installation
- CDH Installation
- CDH Cluster Services

Configuring a Cloudera Cluster

- Overview
- Configuration Settings
- Modifying Service Configurations
- Configuration Files
- Managing Role Instances
- Adding New Services
- Adding and Removing Hosts

Hadoop Distributed File System

- Overview
- HDFS Topology and Roles
- Edit Logs and Checkpointing
- HDFS Performance and Fault Tolerance
- HDFS and Hadoop Security Overview
- Web User Interfaces for HDFS
- Using the HDFS Command Line Interface
- Other Command Line Utilities

HDFS Data Ingest

- Data Ingest Overview
- File Formats
- Ingesting Data using File Transfer or REST Interfaces
- Importing Data from Relational Databases with Apache Sqoop
- Ingesting Data From External Sources with Apache Flume

- Best Practices for Importing Data

Hive and Impala

- Apache Hive
- Apache Impala

YARN and MapReduce

- YARN Overview
- Running Applications on YARN
- Viewing YARN Applications
- YARN Application Logs
- MapReduce Applications
- YARN Memory and CPU Settings

Apache Spark

- Spark Overview
- Spark Applications
- How Spark Applications Run on YARN
- Monitoring Spark Applications

Planning Your Cluster

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

Advanced Cluster Configuration

- Configuring Service Ports
- Tuning HDFS and MapReduce
- Enabling HDFS High Availability

Managing Resources

- Configuring cgroups with Static Service Pools
- The Fair Scheduler
- Configuring Dynamic Resource Pools
- Impala Query Scheduling

Cluster Maintenance

- Checking HDFS Status
- Copying Data Between Clusters

- Rebalancing Data in HDFS
- HDFS Directory Snapshots
- Upgrading a Cluster

Monitoring Clusters

- Cloudera Manager Monitoring Features
- Health Tests
- Events and Alerts
- Charts and Reports
- Monitoring Recommendations

Cluster Troubleshooting

- Overview
- Troubleshooting Tools
- Misconfiguration Examples
- Essential Points

Installing and Managing Hue

- Overview
- Managing and Configuring Hue
- Hue Authentication and Authorization

Security

- Hadoop Security Concepts
- Hadoop Authentication Using Kerberos
- Hadoop Authorization
- Hadoop Encryption
- Securing a Hadoop Cluster

Apache Kudu

- Kudu Overview
- Architecture
- Installation and Configuration
- Monitoring and Management Tools

Apache Kafka

- What Is Apache Kafka?
- Apache Kafka Overview
- Apache Kafka Cluster Architecture
- Apache Kafka Command Line Tools
- Using Kafka with Flume

Object Storage in the Cloud

- Object Storage
- Connecting Hadoop to Object Storage