

CLOUDERA APACHE HADOOP ADMINISTRATION COURSE CONTENT

❖ **THE CASE FOR APACHE HADOOP:**

- A Brief History of Hadoop
- Core Hadoop Components
- Fundamental Concepts

❖ **THE HADOOP DISTRIBUTED FILE SYSTEM:**

- HDFS Features
- HDFS Design Assumptions
- Overview of HDFS Architecture
- Writing and Reading Files
- NameNode Considerations
- An Overview of HDFS Security
- Hands-On Exercise

❖ **MAPREDUCE:**

- What Is MapReduce?
- Features of MapReduce
- Basic MapReduce Concepts
- Architectural Overview
- MapReduce Version 2
- Failure Recovery
- Hands-On Exercise

❖ **AN OVERVIEW OF THE HADOOP ECOSYSTEM:**

- What is the Hadoop Ecosystem?
- Integration Tools
- Analysis Tools
- Data Storage and Retrieval Tools

❖ **PLANNING YOUR HADOOP CLUSTER:**

- General planning Considerations
- Choosing the Right Hardware
- Network Considerations
- Configuring Nodes

❖ **HADOOP INSTALLATION:**

- Deployment Types
- Installing Hadoop
- Using Cloudera Manager for Easy Installation
- Basic Configuration Parameters
- Hands-On Exercise

❖ **ADVANCED CONFIGURATION:**

- Advanced Parameters
- Configuring Rack Awareness
- Configuring Federation
- Configuring High Availability
- Using Configuration Management Tools

❖ **HADOOP SECURITY:**

- Why Hadoop Security Is Important
- Hadoop's Security System Concepts
- What Kerberos Is and How it Works
- Configuring Kerberos Security
- Integrating a Secure Cluster with Other Systems

❖ **MANAGING AND SCHEDULING JOBS:**

- Managing Running Jobs
- Hands-On Exercise
- The FIFO Scheduler
- The FairScheduler
- Configuring the FairScheduler
- Hands-On Exercise

❖ **CLUSTER MAINTENANCE:**

- Checking HDFS Status
- Hands-On Exercise
- Copying Data Between Clusters
- Adding and Removing Cluster Nodes
- Rebalancing the Cluster
- Hands-On Exercise
- NameNode Metadata Backup
- Cluster Upgrading

❖ **CLUSTER MONITORING AND TROUBLESHOOTING:**

- General System Monitoring
- Managing Hadoop's Log Files
- Using the NameNode and JobTracker Web UIs
- Hands-On Exercise
- Cluster Monitoring with Ganglia
- Common Troubleshooting Issues
- Benchmarking Your Cluster

❖ **POPULATING HDFS FROM EXTERNAL SOURCES:**

- An Overview of Flume
- Hands-On Exercise
- An Overview of Sqoop
- Best Practices for Importing Data

❖ **INSTALLING AND MANAGING OTHER HADOOP PROJECTS:**

- Hive
- Pig
- HBase