



Open Source Hadoop Administration

Duration: 4 Days **Course Code: OSHA** **Delivery Method: Virtual Learning**

Overview:

This open source course provides participants with a comprehensive understanding of the steps necessary to install, configure, operate and maintain Hadoop. The course begins with an overview of the Big Data landscape, and then dives into a system administration working view of running Hadoop.

Virtual Learning

This interactive training can be taken from any location, your office or home and is delivered by a trainer. This training does not have any delegates in the class with the instructor, since all delegates are virtually connected. Virtual delegates do not travel to this course, Global Knowledge will send you all the information needed before the start of the course and you can test the logins.

Target Audience:

This course is intended for System administrators, DevOps engineers, and software developers responsible for managing and maintaining Hadoop clusters.

Objectives:

- Upon successful completion of this course, participants should be able to:
 - Describe the fundamental concepts of using Big Data
 - Identify where Hadoop fits into a Big Data strategy
 - Learn to plan your Hadoop cluster.
 - Learn HDFS features.
 - Learn how to get data into HDFS.
 - Learn to work with MapReduce.
 - Learn installation and configuration of Hadoop.
 - Learn cluster maintenance.
-

Content:

Hadoop Introduction

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

Planning Your Hadoop Cluster

- General Planning Considerations
- Choosing Hardware
- Network Considerations
- Configuring Nodes
- Planning for Cluster Management

HDFS

- HDFS Features
- Writing and Reading Files
- NameNode Considerations
- HDFS Security
- Namenode Web UI
- Hadoop File Shell

Getting Data into HDFS

- Pulling data from External Sources with Flume
- Importing Data from Relational Databases with Sqoop
- REST Interfaces
- Best Practices

• MapReduce

- MapReduce overview
- Features of MapReduce
- Architectural Overview
- YARN MapReduce Version 2
- Failure Recovery
- The JobTracker Web UI

Hadoop Installation ; Initial Configuration

- Configuration ; Deployment Types
- Installing Hadoop
- Specifying the Hadoop Configuration
- Initial HDFS ; MapReduce Configuration
- Log Files

Installing/Configuring Hive, Impala, and Pig

- Hive
- Impala
- Pig

Hadoop Clients

- What is a Hadoop Client?
- Installing and Configuring Hadoop Clients
- Installing and Configuring Hue
- Hue Authentication and Configuration

Advanced Cluster Configuration

- Advanced Configuration Parameters
- Configuring Hadoop Ports
- Explicitly Including and Excluding Hosts
- Configuring HDFS for Rack Awareness ; 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

Managing and Scheduling Jobs

- Managing Running Jobs
- Scheduling Hadoop Jobs
- Configuring the FairScheduler

Cluster Maintenance

- Checking HDFS Status
- Copying Data Between Clusters
- Adding/Removing Cluster Nodes
- Rebalancing the Cluster
- NameNode Metadata Backup
- Cluster Upgrades

Cluster Monitoring and Troubleshooting

- General System Monitoring
- Managing Hadoop's Log Files
- Monitoring the Clusters
- Common Troubleshooting Issues

Further Information:

For More information, or to book your course, please call us on Head Office 01189 123456 / Northern Office 0113 242 5931

info@globalknowledge.co.uk

www.globalknowledge.com/en-gb/

Global Knowledge, Mulberry Business Park, Fishponds Road, Wokingham Berkshire RG41 2GY UK