Cloudera Administrator Training for Apache Hadoop

Duration: 4 Days  Course Code: GK3901

Overview:
In this hands-on course, you will be introduced to the basics of Hadoop, Hadoop Distributed File System (HDFS), MapReduce, Hive, Pig, and HBase. You will cover core administration skills, such as cluster deployment, job management, and ongoing Hadoop maintenance and monitoring, as you gain the expertise to support your environments in day-to-day activities. This course covers concepts addressed on the Cloudera Certified Administrator for Apache Hadoop (CCAH) exam and includes a CCAH exam voucher you'll receive at the end of class.

Target Audience:
System administrators looking to understand all of the steps necessary to operate and manage Apache Hadoop clusters

Objectives:
- HDFS and MapReduce
- Configure the FairScheduler to provide service-level agreements for multiple users of a cluster
- Optimal hardware configurations for Hadoop clusters
- Maintain and monitor your cluster
- Network considerations to take into account when building out your cluster
- Load data into the cluster from dynamically generated files using Flume and from relational database management systems using Sqoop
- Configure Hadoop options for best cluster performance
- System administration issues with other Hadoop projects such as Hive, Pig, and HBase

Prerequisites:
- Basic level of Linux system administration experience
- Prior knowledge of Apache Hadoop is not required

Testing and Certification
This course is part of the following programs or tracks:
- CCAH: Cloudera Certified Administrator for Apache Hadoop (CDH3)

Follow-on-Courses:
- Cloudera Training for Apache HBase
- Cloudera Training for Apache Hive and Pig
Content:

Hadoop and HDFS
- Why Hadoop?

HDFS
- MapReduce
- Hive, Pig, HBase, and Other Ecosystem Projects
- Choosing the Right Hardware
- Node Topologies
- Choosing the Right Software
- Using SCM Express for Easy Installation
- Typical Configuration Parameters
- Configuring Rack Awareness
- Using Configuration Management Tools
- FIFO Scheduler
- Fair Scheduler
- Copying Data with Distcp
- Rebalancing Cluster Nodes
- Adding and Removing Cluster Nodes
- Backup and Restore
- Upgrading and Migrating
- NameNode Metadata
- Using the NameNode and JobTracker Web UIs
  - Interpreting Job Logs
  - Monitoring with Ganglia
  - Other Monitoring Tools
  - General Optimization Tips
  - Benchmarking Your Cluster
  - Using Flume
  - Best Practices for Data Ingestion
  - Pig
  - HBase
  - Metastore

Managing and Scheduling Jobs
- Starting and Stopping MapReduce Jobs

HDFS
- MapReduce
- Hive, Pig, HBase, and Other Ecosystem Projects
- Choosing the Right Hardware
- Node Topologies
- Choosing the Right Software
- Using SCM Express for Easy Installation
- Typical Configuration Parameters
- Configuring Rack Awareness
- Using Configuration Management Tools
- FIFO Scheduler
- Fair Scheduler
- Copying Data with Distcp
- Rebalancing Cluster Nodes
- Adding and Removing Cluster Nodes
- Backup and Restore
- Upgrading and Migrating
- NameNode Metadata
- Using the NameNode and JobTracker Web UIs
  - Interpreting Job Logs
  - Monitoring with Ganglia
  - Other Monitoring Tools
  - General Optimization Tips
  - Benchmarking Your Cluster
  - Using Flume
  - Best Practices for Data Ingestion
  - Pig
  - HBase
  - Metastore

HDFS
- MapReduce
- Hive, Pig, HBase, and Other Ecosystem Projects
- Choosing the Right Hardware
- Node Topologies
- Choosing the Right Software
- Using SCM Express for Easy Installation
- Typical Configuration Parameters
- Configuring Rack Awareness
- Using Configuration Management Tools
- FIFO Scheduler
- Fair Scheduler
- Copying Data with Distcp
- Rebalancing Cluster Nodes
- Adding and Removing Cluster Nodes
- Backup and Restore
- Upgrading and Migrating
- NameNode Metadata
- Using the NameNode and JobTracker Web UIs
  - Interpreting Job Logs
  - Monitoring with Ganglia
  - Other Monitoring Tools
  - General Optimization Tips
  - Benchmarking Your Cluster
  - Using Flume
  - Best Practices for Data Ingestion
  - Pig
  - HBase
  - Metastore
General Optimization Tips
Benchmarking Your Cluster
Using Flume
Best Practices for Data Ingestion
Pig
HBase
Metastore

HDFS
MapReduce
Hive, Pig, HBase, and Other Ecosystem Projects
Choosing the Right Hardware
Node Topologies
Choosing the Right Software
Using SCM Express for Easy Installation
Typical Configuration Parameters
Configuring Rack Awareness
Using Configuration Management Tools
FIFO Scheduler
Fair Scheduler
Copying Data with Distcp
Rebalancing Cluster Nodes
Adding and Removing Cluster Nodes
Backup and Restore
Upgrading and Migrating
NameNode Metadata
Using the NameNode and JobTracker Web UIs
Interpreting Job Logs
Monitoring with Ganglia
Other Monitoring Tools
General Optimization Tips
Benchmarking Your Cluster
Using Flume
Best Practices for Data Ingestion
Pig
HBase
Metastore

Planning Your Hadoop Cluster
General Planning Considerations

HDFS
MapReduce
Hive, Pig, HBase, and Other Ecosystem Projects
Choosing the Right Hardware
Node Topologies
Choosing the Right Software
Using SCM Express for Easy Installation
Typical Configuration Parameters
Configuring Rack Awareness
Using Configuration Management Tools
FIFO Scheduler
Fair Scheduler
Copying Data with Distcp
Rebalancing Cluster Nodes
Adding and Removing Cluster Nodes
Backup and Restore
Upgrading and Migrating
NameNode Metadata
Using the NameNode and JobTracker Web UIs
Interpreting Job Logs
Monitoring with Ganglia
Other Monitoring Tools
General Optimization Tips
Benchmarking Your Cluster
Using Flume
Best Practices for Data Ingestion
Pig
HBase
Metastore

HDFS
MapReduce
Hive, Pig, HBase, and Other Ecosystem Projects
Choosing the Right Hardware
Node Topologies
Choosing the Right Software
Using SCM Express for Easy Installation
Typical Configuration Parameters
Configuring Rack Awareness
Using Configuration Management Tools
FIFO Scheduler
Fair Scheduler
Copying Data with Distcp
Rebalancing Cluster Nodes
Adding and Removing Cluster Nodes
Backup and Restore
Upgrading and Migrating
NameNode Metadata
Using the NameNode and JobTracker Web UIs
Interpreting Job Logs
Monitoring with Ganglia
Other Monitoring Tools
General Optimization Tips
Benchmarking Your Cluster
Using Flume
Best Practices for Data Ingestion
Pig
HBase
Metastore

HDFS
MapReduce
Hive, Pig, HBase, and Other Ecosystem Projects
Choosing the Right Hardware
Node Topologies
Choosing the Right Software
Using SCM Express for Easy Installation
Typical Configuration Parameters
Configuring Rack Awareness
Using Configuration Management Tools
FIFO Scheduler
Fair Scheduler
Copying Data with Distcp
Rebalancing Cluster Nodes
Adding and Removing Cluster Nodes
Backup and Restore
Upgrading and Migrating
NameNode Metadata
Using the NameNode and JobTracker Web UIs
Interpreting Job Logs
Monitoring with Ganglia
Other Monitoring Tools
General Optimization Tips
Benchmarking Your Cluster
Using Flume
Best Practices for Data Ingestion
Pig
HBase
Metastore

HDFS
MapReduce
Hive, Pig, HBase, and Other Ecosystem Projects
Choosing the Right Hardware
Node Topologies
Choosing the Right Software
Using SCM Express for Easy Installation
Typical Configuration Parameters
Configuring Rack Awareness
Using Configuration Management Tools
FIFO Scheduler
Fair Scheduler
Copying Data with Distcp
Rebalancing Cluster Nodes
Adding and Removing Cluster Nodes
Backup and Restore
Upgrading and Migrating
NameNode Metadata
Using the NameNode and JobTracker Web UIs
Interpreting Job Logs
Monitoring with Ganglia
Other Monitoring Tools
General Optimization Tips
Benchmarking Your Cluster
Using Flume
Best Practices for Data Ingestion
Pig
HBase
Metastore
Typical Configuration Parameters
Configuring Rack Awareness
Using Configuration Management Tools
FIFO Scheduler
Fair Scheduler
Copying Data with Distcp
Rebalancing Cluster Nodes
Adding and Removing Cluster Nodes
Backup and Restore
Upgrading and Migrating
NameNode Metadata
Using the NameNode and JobTracker Web UIs
Interpreting Job Logs
Monitoring with Ganglia
Other Monitoring Tools
General Optimization Tips
Benchmarking Your Cluster
Using Flume
Best Practices for Data Ingestion
Pig
HBase
Metastore

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.co.uk

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