Cloudera University’s three-day administrator training course for Apache Hadoop provides System Administrators a comprehensive understanding of all the steps necessary to operate and manage Hadoop clusters. From installation and configuration, through load balancing and tuning your cluster, Cloudera’s Administration course has you covered.

Through lecture and interactive, hands-on exercises, attendees will cover topics such as

• Introduction to Apache Hadoop and HDFS
• Apache Hadoop architecture
• Proper cluster configuration and deployment
• Populating HDFS using Apache Sqoop
• Management and monitoring tools
• Job scheduling
• Best practices for maintaining Apache Hadoop in Production
• Installing and managing other Apache Hadoop projects
• Diagnosing, tuning and solving Apache Hadoop issues

**Audience**

This course is designed for people with at least a basic level of Linux system administration experience. Prior knowledge of Hadoop is not required.
### Course Outline: Cloudera Administrator Training for Apache Hadoop

<table>
<thead>
<tr>
<th>Section</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>• An Introduction To Hadoop And HDFS</td>
<td></td>
</tr>
</tbody>
</table>
  - Why Hadoop?  
  - HDFS  
  - MapReduce  
  - Hive, Pig, HBase and other ecosystem projects  
  - Hands-On Exercise: Installing a pseudo-distributed cluster |
| • Planning Your Hadoop Cluster |  
  - General Planning Considerations  
  - Choosing The Right Hardware  
  - Node Topologies  
  - Choosing The Right Software |
| • Deploying Your Cluster |  
  - Installing Hadoop  
  - Using SCM Express for easy installation  
  - Typical Configuration Parameters  
  - Configuring Rack Awareness  
  - Using Configuration Management Tools  
  - Hands-On Exercise: Installing a Hadoop Cluster |
| • Managing and Scheduling Jobs |  
  - Starting and stopping MapReduce jobs  
  - Hands-On Exercise: Managing jobs  
  - The FIFO Scheduler  
  - The Fair Scheduler  
  - Hands-On Exercise: Using the FairScheduler |
| • Cluster Maintenance |  
  - Checking HDFS with fsck  
  - Hands-On Exercise: Breaking the Cluster  
  - Copying data with distcp |
| • Cluster Maintenance (continued) |  
  - Rebalancing cluster nodes  
  - Adding and removing cluster nodes  
  - Hands-On Exercise: Verifying the Cluster's Self-Healing Features  
  - Backup And Restore  
  - Upgrading and Migrating  
  - Hands-On Exercise: Backing Up and Restoring the NameNode Metadata |
| • Cluster Monitoring, Troubleshooting and Optimizing |  
  - Hadoop Log Files  
  - Using the NameNode and JobTracker Web UIs  
  - Interpreting Job Logs  
  - Monitoring with Ganglia  
  - Other monitoring tools  
  - General Optimization Tips  
  - Benchmarking Your Cluster |
| • Populating HDFS From External Sources |  
  - Using Sqoop  
  - Using Flume  
  - Best Practices for Data Ingestion |
| • Installing And Managing Other Hadoop Projects |  
  - Hive  
  - Pig  
  - HBase  
  - Hands-On Exercise: Configuring the Hive Shared Metastore |
| • Cluster Monitoring, Troubleshooting and Optimizing |  
  - Hadoop Log Files  
  - Using the NameNode and JobTracker Web UIs  
  - Interpreting Job Logs  
  - Monitoring with Ganglia  
  - Other monitoring tools  
  - General Optimization Tips  
  - Benchmarking Your Cluster |
| • Cloudera Certified Administrator Exam |  
  - Hive  
  - Pig  
  - HBase  
  - Hands-On Exercise: Configuring the Hive Shared Metastore |

---

### Cloudera Certified Administrator for Apache Hadoop (CCAH)

Establish yourself as a trusted and valuable resource by completing the online certification exam for Apache Hadoop Administrators. The exam is demanding and is designed to test your fluency with concepts and terminology in the following areas:

- **Apache Hadoop Cluster Overview**
  Daemons and normal operation of an Apache Hadoop cluster, both in data storage and in data processing. The current features of computing systems that motivate a system like Apache Hadoop

- **Apache Hadoop Cluster Planning**
  Principal points to consider in choosing the hardware and operating systems to host an Apache Hadoop cluster

- **Apache Hadoop Cluster Management**
  Cluster handling of disk and machine failures. Regular tools for monitoring and managing the Apache Hadoop file system

- **Job Scheduling**
  How the default scheduler and the fair scheduler handle the tasks in a mix of jobs running on a cluster

- **Monitor and Logging**
  Basic functionality and features of Apache Hadoop's logging and monitoring systems