

***Durée : 3 jours soit 21h***

## COURSE SUMMARY

This three-day hands-on training course is for system administrators and others responsible for managing Apache Hadoop clusters in production or development environments.

## YOU WILL LEARN :

- How the Hadoop Distributed File System and MapReduce work
- What hardware configurations are optimal for Hadoop clusters
- What network considerations to take into account when building out your cluster
- How to configure Hadoop's options for best cluster performance
- How to configure the FairScheduler to provide service-level agreements for multiple users of a cluster
- How to maintain and monitor your cluster
- How to load data into the cluster from dynamically-generated files using Flume, and from relational database management systems using Sqoop
- What system administration issues exist with other Hadoop projects such as Hive, Pig, and HBase

## HANDS-ON EXERCISES

Throughout the course, hands-on labs help students build their knowledge and apply the concepts being discussed.

## PROGRAM

### Introduction

#### An Introduction To Hadoop And HDFS

- Why Hadoop?
- HDFS
- MapReduce
- Hive, Pig, HBase, and Other Ecosystem Projects
- Hands-On Exercise:

#### Planning Your Hadoop Cluster

- General Planning Considerations
- Choosing The Right Hardware
- Choosing The Right Hardware
- Network Considerations
- Configuring Nodes

#### Configuring and Deploying Your Cluster

- Deployment Types
- Installing Hadoop



**Durée : 3 jours soit 21h**

- Using Cloudera Manager for Easy Installation
- Typical Configuration Parameters
- Configuring Rack Awareness
- Using Configuration Management Tools
- Hands-On Exercise

#### Managing and Scheduling Jobs

- Deployment Types
- Installing Hadoop
- Using Cloudera Manager for Easy Installation
- Typical Configuration Parameters
- Configuring Rack Awareness
- Using Configuration Management Tools
- Hands-On Exercise

#### Cluster Maintenance

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

#### Cluster Monitoring and Troubleshooting

- General System Monitoring
- Managing Hadoop's Log Files
- Using the NameNode and JobTracker Web UI
- 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

#### Conclusion

Cloudera Certified Administrator for Apache Hadoop Exam