



HADOOP Apache pour les administrateurs (certification Cloudera)

Déscription :

Cette formation de 3 jours s'adresse aux administrateurs systèmes et aux personnes qui gèrent et maintiennent les clusters Apache Hadoop en production ou qui évoluent dans un environnement de développement.

Objectifs

- Reposer les bases de l'environnement Hadoop et HDFS
- Poser les bases de Hadoop Cluster : choix du matériel, "configuring nodes"
- Configurer, déployer et gérer des Clusters : utiliser "Cloudera Manager for Easy Installation", "Configuration Management Tools" et Ganglia,
- Rajouter et retirer des « Cluster Nodes », rééquilibrer un Cluster et optimiser un Cluster
- Gérer des tâches : Le "FIFO Scheduler", Le "FairScheduler"
- Gérer les "Log files" d'Hadoop
- Utiliser le "NameNode" et "Job Tracker Web UI"
- Troubleshooting
- Savoir "Benchmarking" un Cluster

Publics

Cette formation s'adresse aux administrateurs systèmes et aux personnes qui gèrent et maintiennent les clusters Apache Hadoop en production ou qui évoluent dans un environnement de développement.

Durée

3 jours

Pré-requis

Expérience dans l'administration système Linux.

Programme de cette formation

Une introduction à Hadoop et HDFS

- Pourquoi Hadoop
- HDFS

-
- MapReduce
- Hive, Pig, HBase, et d'autres écosystèmes projet
- Exercices pratiques

Planning Your Hadoop Cluster

- Choisir le bon matériel
- Réseau
- "Configuring Nodes"

Configurer et déployer un Cluster

- Types de déploiement
- Installer Hadoop
- Utiliser "Cloudera Manager for Easy Installation"
- Les paramètres de configuration typiques
- Utiliser "Configuration Management Tools"
- Exercices pratiques

Gérer les tâches

- Gérer les tâches en cours
- Exercices pratiques
- Le "FIFO Scheduler"
- Le "FairScheduler"
- Exercices pratiques

Gestion de Clusters

- HDFS
- Exercices pratiques : copier des données entre clusters
- Rajouter et retirer des « Cluster Nodes »
- Rééquilibrer un Cluster
- Exercices pratiques : Optimiser un Cluster

Cluster Monitoring et Troubleshooting

- Gérer les "Log files" d'Hadoop
- Utiliser le "NameNode" et "Job Tracker Web UI"
- Exercices pratiques
- Gestion de Cluster avec Ganglia
- Troubleshooting
- Benchmarking un Cluster

Conclusion

Cloudera Certified Administrator for Apache Hadoop Exam