

## Administrer la plateforme Hadoop Cloudera

### Formation officielle Cloudera Administrator Training for Apache Hadoop

#### DESCRIPTION

Vous souhaitez exploiter le potentiel de vos données pour créer de la valeur et développer votre activité. Avec Hadoop et son architecture flexible et évolutive, vous pouvez stocker, traiter et analyser vos données à partir d'une plateforme unique fonctionnant sur du matériel standard.

Dès sa création en 2008, Cloudera a lié son histoire à celle de l'écosystème Hadoop. Avec ses fondations composées à 100 % de logiciels open source et de standards ouverts, la plate-forme Cloudera vous assure un meilleur contrôle des coûts, plus de souplesse et des résultats plus performants pour votre organisation. CDH, la plate-forme open source de Cloudera, est ainsi devenue la distribution la plus populaire de Hadoop.

De l'installation à la configuration en passant par l'équilibrage de charge et le réglage, cette formation de quatre jours fournit aux participants une compréhension complète de toutes les étapes nécessaires pour opérer et maintenir un cluster Hadoop à l'aide de Cloudera Manager.

#### OBJECTIFS PEDAGOGIQUES

- Reposer les bases de l'environnement Hadoop, MapReduce, Spark et HDFS
- Gérer un cluster avec les fonctionnalités de Cloudera Manager
- Déterminer le matériel et l'infrastructure appropriés pour son cluster
- Configurer et déployer correctement le cluster pour l'intégration avec le système d'information
- Charger des données dans le cluster à partir de fichiers générés dynamiquement à l'aide de Flume, ou à partir de SGBDR en utilisant Sqoop
- Configurer FairScheduler pour répartir les ressources entre plusieurs utilisateurs d'un cluster
- Préparer et maintenir Apache Hadoop en production en utilisant les bonnes pratiques
- Dépanner, diagnostiquer, mettre au point et résoudre les problèmes sur Hadoop

#### Classe à distance

Cloudera

Code :

**CLADM**

Durée :

**4 jour(s) (28,00 heures)**

Exposés : **40 %**

Cas pratiques : **50 %**

Echanges d'expérience : **10 %**

#### Inter-entreprises :

Prochaines sessions

disponibles [sur notre site web](#).

Tarif : 2 695,00 € HT /  
participant

#### Intra-entreprise :

Tarifs et dates sur demande.

## **PUBLIC CIBLE**

- Administrateur système
- Responsable informatique
- Architecte système
- Développeur
- Analyste de données
- Administrateur de bases de données

## **PRE-REQUIS**

Connaissance de base de la ligne de commande Linux

## **METHODE PEDAGOGIQUE**

Formation avec apports théoriques, échanges sur les contextes des participants et retours d'expérience pratique du formateur, complétés de travaux pratiques et de mises en situation. Cette formation permet de préparer l'examen associé au titre de la certification « Cloudera Certified Associate Administrator ».

## **PROFIL DES INTERVENANTS**

Toutes nos formations sont animées par des consultants-formateurs expérimentés et reconnus par leurs pairs.

## **MODALITÉS D'ÉVALUATION ET FORMALISATION À L'ISSUE DE LA FORMATION**

L'évaluation des acquis se fait tout au long de la session au travers des ateliers et des mises en pratique. Une évaluation à chaud sur la satisfaction des stagiaires est réalisée systématiquement en fin de session et une attestation de formation est délivrée aux participants mentionnant les objectifs de la formation, la nature, le programme et la durée de l'action de formation ainsi que la formalisation des acquis.

## **PROGRAMME PEDAGOGIQUE DETAILLE**

## Jour 1

### INTRODUCTION

#### CLLOUDERA ENTERPRISE DATA HUB

- Cloudera Enterprise Data Hub
- Introduction au CDH
- Introduction à Cloudera Manager
- Les responsabilités d'un administrateur Hadoop

#### INSTALLATION DE CLOUDERA MANAGER ET DU CDH

- Introduction à l'installation du cluster
- Installation de Cloudera Manager Installation
- Installation du CDH
- Les services du cluster CDH

#### CONFIGURER UN CLUSTER CLOUDERA

- Introduction
- Paramètres de configuration
- Modifier la configuration des services
- Fichiers de configuration
- Gérer les instances de rôle
- Ajouter des nouveaux services
- Ajouter et supprimer des hôtes

#### HADOOP DISTRIBUTED FILE SYSTEM

- Introduction
- Topologie et rôles HDFS
- Modifier les logs et le checkpointing
- La performance HDFS et la tolérance à la panne
- Introduction à la sécurité de HDFS et de Hadoop
- Interfaces utilisateurs web pour HDFS
- Utiliser la ligne de commande HDFS
- Autres outils de ligne de commande

## Jour 2

#### INGESTION DE DONNÉES SUR HDFS

- Introduction à l'ingestion de données
- Formats de fichiers
- Ingérer de la donnée en utilisant File Transfer ou les interfaces REST
- Ingérer de la donnée d'une base de donnée relationnel avec Sqoop
- Ingérer de la donnée d'une source externe avec Flume
- Les bonnes pratiques d'ingestion de donnée

#### HIVE ET IMPALA

- Apache Hive
- Apache Impala

#### YARN ET MAPREDUCE

- Introduction à YARN
- Exécuter des applications sur YARN
- Explorer les applications YARN
- Les logs d'application YARN
- Les applications Map Reduce
- Réglage mémoire et CPU pour YARN

#### APACHE SPARK

- Introduction à Spark
- Les applications Spark
- Comment les applications Spark s'exécutent sur YARN
- Monitorer les applications Spark

### Jour 3

#### DIMENSIONNEMENT DE VOTRE CLUSTER HADOOP

- Considérations générales relatives au dimensionnement
- Choix du matériel
- Considérations sur le réseau
- Options de virtualisation
- Options de déploiement cloud
- Configuration des noeuds

#### CONFIGURATION AVANCÉ DU CLUSTER

- Configurer les ports de service
- Paramétrer HDFS et MapReduce
- Activer la Haute Disponibilité HDFS

#### GESTION DES RESSOURCES

- Configuration de cgroups avec des centres de services statiques
- Le Fair Scheduler
- Configurer la gestion dynamique des ressources
- Planification des requêtes Impala

#### MAINTENANCE DU CLUSTER

- Vérification du statut HDFS
- Copier les données entre clusters
- Rééquilibrage du cluster
- Snapshots de répertoires
- Mise à niveau du cluster

#### Jour 4

#### MONITORING DU CLUSTER

- Fonctionnalités de monitoring de Cloudera Manager
- Tests de santé
- Événements et alertes
- Graphiques et rapports
- Recommandation de monitoring

#### DIAGNOSTIC DU CLUSTER

- Introduction
- Outils de diagnostic
- Exemples de mauvaises configurations

#### INSTALLER ET GÉRER HUE

- Introduction
- Gérer et configurer Hue
- Authentification et autorisation Hue

### SÉCURITÉ

- Les concepts de sécurité sur Hadoop
- Authentification sur Hadoop en utilisant Kerberos
- Authorisation sur Hadoop
- Chiffrement sur Hadoop
- Sécuriser un cluster Hadoop

### CONCLUSION

Suppléments sous réserve de temps disponible

### APACHE KUDU

- Introduction à Kudu
- Architecture
- Installation et Configuration
- Outils de monitoring et de gestion

### APACHE KAFKA

- Qu'est-ce que Apache Kafka ?
- Introduction à Kafka
- Architecture de cluster Kafka
- Outils de ligne de commande Kafka
- Utiliser Kafka avec Flume

### STOCKAGE D'OBJETS DANS LE CLOUD

- Système de stockage d'objet
- Connecter Hadoop et un système de stockage objet

