

Big Data - Administration et infrastructure

Cloudera - Administration d'une plateforme

3 jours (21h00) | ★★★★★ 4,6/5 | BIG-CLDR | Évaluation qualitative de fin de stage |
Formation délivrée en présentiel ou distanciel ⁽¹⁾

Formations Informatique > Big Data > Big Data - Administration et infrastructure



À l'issue de ce stage vous serez capable de :

- Mettre en oeuvre l'administration des clusters Cloudera
- Configurer et monitorer le cluster avec Cloudera Manager
- Explorer les technologies de Cloudera.

Niveau requis

Maîtriser les commandes Shell et avoir des connaissances de base autour des requêtes SQL.

Public concerné

Administrateurs système et consultants SI.

Cette formation :

- Est animée par un consultant-formateur dont les compétences techniques, professionnelles et pédagogiques ont été validées par des diplômes et/ou testées et approuvées par l'éditeur et/ou par M2i Formation
- Bénéficie d'un suivi de son exécution par une feuille de présence émarginée par demi-journée par les stagiaires et le formateur.

(1) Modalité et moyens pédagogique :

Formation délivrée en présentiel ou distanciel * (e-learning, classe virtuelle, présentiel à distance). Le formateur alterne entre méthodes ** démonstrative, interrogative et active (via des travaux pratiques et/ou des mises en situation). La validation des acquis peut se faire via des études de cas, des quiz et/ou une certification.

Les moyens pédagogiques mis en oeuvre (variables suivant les formations) sont : ordinateurs Mac ou PC (sauf pour les cours de l'offre Management), connexion internet fibre, tableau blanc ou paperboard, vidéoprojecteur ou écran tactile interactif (pour le distanciel). Environnements de formation installés sur les postes de travail ou en ligne. Supports de cours et exercices.

* Nous consulter pour la faisabilité en distanciel. ** Ratio variable selon le cours suivi.

Programme

Jour 1

Introduction

- Introduction au Big Data
- Présentation d'Hadoop et de son écosystème
- Architecture solution
- NoSQL
- Présentation de :
 - Cloudera Data Platform (CDP)
 - Cloudera Manager

Installation d'un cluster CDP

- Cluster : vue d'ensemble
- Justification d'une solution de gestion de cluster
- Caractéristiques de Cloudera Manager
- Prérequis pour l'installation d'un cluster
- Installation de CDP
- Installation de Cloudera Manager

Exemples de travaux pratiques (à titre indicatif)

- Installation de Cloudera Manager et de Hadoop (CDH)
- Préparation des environnements

Jour 2

Gestion des utilisateurs

- Gestion des utilisateurs et des groupes avec Cloudera Manager
- Gestion des privilèges pour les utilisateurs locaux et LDAP

Configuration d'un cluster Cloudera

- Vue d'ensemble
- Paramètres de configuration
- Fichiers de configuration
- Gérer les instances de rôle
- Ajouter de nouveaux services
- Ajouter et supprimer des hôtes

Exemple de travaux pratiques (à titre indicatif)

- Mise en situation d'une configuration

Le système de stockage

- Fonctionnalités HDFS
- Ecriture et lecture de fichiers
- Considérations relatives à la mémoire NameNode
- Vue d'ensemble de la sécurité HDFS
- Gestion du stockage HDFS
- Utilisation de Hadoop File Shell
- Hbase
- Cloudera Storage

Exemple de travaux pratiques (à titre indicatif)

- Manipulation des fichiers via HDFS

Jour 3

Ingestion des données HDFS

- Ingestion de données à partir de :
 - Importation de données relationnelles avec Sqoop
 - Intégration de données avec NiFi
 - Apache Kafka
- Bonnes pratiques pour l'importation de données

Accès et analyse des données

- Apache
 - Hive
 - Impala
 - Spark
- Exécution de frameworks de calcul sur YARN
- Hue

Sécurité Hadoop

- Importance de la sécurité Hadoop
- Concepts du système de sécurité de Hadoop
- Présentation et fonctionnement de Kerberos
- Sécurisation d'un cluster Hadoop avec Kerberos
- Autres concepts de sécurité

Gestion des ressources

- Configuration des Cgroups avec les pools de services statiques
- Planification des ressources
- Configuration des pools de ressources dynamiques
- Mémoire YARN et paramètres du CPU
- Planification de requêtes Impala

Entretien du cluster

- Vérification de l'état HDFS
- Copier les données entre clusters
- Ajout et suppression de noeud de cluster
- Rééquilibrage du cluster snapshots de répertoires
- Mise à niveau du cluster

Surveillance et dépannage du cluster

- Fonctionnalités de surveillance du Cloudera Manager
- Tests de santé
- Surveillance des clusters Hadoop
- Dépannage des clusters Hadoop

Modalités d'évaluation des acquis

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation ou une certification (M2i ou éditeur)