

Mis à jour le 18/08/2023

S'inscrire

Formation Cloudera Administrateur en Production

3 jours (21 heures)

Présentation

Cloudera Administrateur est l'une des plateformes de gestion des données open source les plus puissantes au monde. Basée sur Hadoop il propose des différents outils intégrés et multifonctionnels pour analyser et centraliser les données. Cloudera gère des données en cloud hybride et multicloud qui sont entièrement intégrées à l'ingénierie des données et à la Machine Learning, ce qui permet de réaliser une analyse d'entreprise facile, rapide et sécurisée. Cette plateforme vous offrira la sécurité, la conformité, la migration et la gestion des métadonnées dans tous les environnements. Notre formation Cloudera Administrateur vous enseignera à gérer le cycle de vie des données sur les clouds privés et publics. Vous apprendrez également à appliquer vos compétences d'analyse de données et de business intelligence aux grands outils tels qu'Apache Impala et Apache Hive. À l'issue de cette formation, vous saurez gérer et maintenir un cluster Hadoop grâce à Cloudera Manager, modifier l'infrastructure logicielle de votre entreprise et utiliser les différents services de données comme Data Warehouse, Machine Learning (ML) ou Data Visualisation.

Objectifs

- Maîtriser les bases de l'environnement Hadoop, Spark, MapReduce
- Maîtriser le fonctionnement HDFS et Yarn/MapReduce
- Configurer et gérer des services avec Cloudera Manager
- Développer des solutions Hadoop Cloudera
- Déployer, configurer et gérer un cluster avec Cloudera Manager
- Intégrer le cluster dans le système informatique
- Garantir aux utilisateurs le haut niveau de sécurité et de traçabilité

Public visé

- Administrateurs systèmes
- Responsables informatiques
- Architectes systèmes
- Spécialistes de bases de données
- Développeurs

Pré-requis

- Connaissance des commandes unix/linux
- Connaissance de base en SQL

Programme de notre formation Cloudera Administrateur

Introduction à Hadoop Cloudera

- Qu'est-ce que Cloudera ?
- Avantages de Cloudera
- Installer et configurer MySQL Database pour Cloudera Manager
- Introduction à l'installation du cluster
- Installer et configurer Cloudera Manager

Fonctionnalités de Cloudera

- Créer des data lakes sur Cloud
- Accéder aux data catalogues
- Analyse de données en temps réel
- Opérer sur cloud privé/public
- Modifier l'infrastructure logicielle

Configurer un cluster Cloudera

- Paramètres d'installation et de configuration
- Fichiers de configuration
- Gérer les instances de rôle
- Ajout de nouveaux services
- Suppression des hôtes

Hadoop Distributed File System

- Introduction à HDFS
- Configuration des paramètres
- Modification des logs et de checkpointing
- Tolérance aux pannes et les performances de HDFS
- Lignes de commande HDFS et ses outils

Gestion et maintenance du cluster

- Paramètres de dimensionnement de cluster
- Considération sur le réseau
- Les options de déploiement cloud
- Surveillance de cluster
- Copier les données entre clusters
- Snapshots de répertoires

Configuration avancée du cluster

- Configuration des portes de services
- Paramètres de HDFS
- Paramètres de MapReduce
- activation de la Haute Disponibilité HDFS

Hive et Impala

- Apache Hive
- Apache Impala

Apache Spark

- Introduction à Spark
- Applications d'Apache Spark
- Exécuter des applications sur YARN
- Monitorer des applications

Data Warehousing

- Moteurs SQL
- Déployer des entrepôts de donnée
- Transférer les workloads du site vers le Cloud
- Ingérer des données à grande échelle
- Restitution visuelle de la donnée avec Data Visualisation

MapReduce et YARN

- Introduction à YARN
- Exécution et exploration des applications
- Les logs principaux d'application YARN
- Les applications MapReduce

Sécurité et gouvernance

- Sécurité sur Hadoop
- Sécurisation d'un cluster
- Assurer la sécurité et la conformité des données
- Présentation et fonctionnement de Kerberos
- Migration des métadonnées
- Gestion des métadonnées dans n'importe quel environnement

Pour aller plus loin

Sociétés concernées

Cette formation s'adresse à la fois aux particuliers ainsi qu'aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l'acquisition correcte des compétences.

Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.