Administrer la plateforme CDP Private Cloud Base de Cloudera

Formation officielle Administrator Training: CDP Private Cloud Base

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.

Successeur des distributions Cloudera Distribution of Hadoop (CDH) et Hortonworks Data Platform (HDP), Cloudera Data Platform est une plateforme hybride qui vous permet de gérer les architectures de données modernes, où que se trouvent celles-ci.

Ce module de quatre jours, de Cloudera pour CDP Private Cloud Base, offre aux participants une compréhension complète de toutes les étapes nécessaires pour exploiter et maintenir des clusters sur site à l'aide de Cloudera Manager. De l'installation et de la configuration à l'équilibrage de charge et au réglage, ce cours de formation Cloudera est la meilleure préparation aux défis du monde réel auxquels sont confrontés les administrateurs qui exécutent CDP Private Cloud Base.

OBJECTIFS PEDAGOGIQUES

- Reposer les bases de l'environnement Hadoop, MapReduce, Spark et HDFS
- Gérer un cluster CDP Private Cloud Base 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 NiFi, ou à partir de SGBDR en utilisant Sqoop
- Configurer FairScheduler pour répartir les ressources entre plusieurs utilisateurs d'un cluster
- Préparer et maintenir un cluster en production en utilisant les bonnes pratiques
- Dépanner, diagnostiquer, mettre au point et résoudre les problèmes sur le cluster

Classe à distance
Data Engineering

Code :
CLADM

Durée :
4 jour(s) (28,00 heures)

Exposés : **40.00 %**
Cas pratiques : **50.00 %**
Echanges d'expérience : **10.00 %**

Inter-entreprises :
Prochaines sessions disponibles sur notre site web.
Tarif : 2 700,00 € 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 « CDP 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

CLOUDERA DATA PLATFORM

- L'état de l'art des réponses aux besoins Big Data

- Le défi de devenir Data Driven
- Cloudera Entreprise Data Cloud
- Aperçu de CDP

INSTALLATION DE CDP PRIVATE CLOUD BASE

- Aperçu de l'installation
- Installation de Cloudera Manager
- Aperçu de CDP Runtime
- Introduction à Cloudera Manager

CONFIGURER UN CLUSTER CDP

- Vue d'ensemble
- 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

STOCKAGE DE DONNÉES

- Introduction
- Topologie et rôles HDFS
- 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
- Aperçu de HBase, Kudu et du stockage cloud

Jour 2

INGESTION DE DONNÉES

- 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 relationnelle avec Sqoop
- Ingérer de la donnée avec NiFi
- Les bonnes pratiques d'ingestion de donnée

FLUX DE DONNÉES

- Aperçu de Cloudera Flow Management et de NiFi
- Architecture de NiFi
- Cloudera Edge Flow Management et MiNiFi
- Introduction à Apache Kafka et de son architecture en cluster
- Outils de ligne de commande Apache Kafka

ACCÈS AUX DONNEES

- Apache Hive
- Apache Impala et son optimisation
- Introduction aux service de recherche
- Introduction à Hue, sa gestion et sa configuration
- Introduction à Cloudera Data Science Workbench (CDSW)

LE COMPUTE

- Introduction à YARN
- Exécuter des applications sur YARN
- Explorer les applications YARN
- Les logs d'application YARN
- Les applications MapReduce
- Réglage mémoire et CPU pour YARN
- Présentation de Tez
- Exécution de Hive sur Tez
- Les propriétés ACID et Hive
- Introduction à Apache Spark
- Introduction à Spark
- Comment les applications Spark s'exécutent sur YARN
- Monitoring les applications Spark
- Introduction à Phoenix

Jour 3

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

DIMENSIONNEMENT DU CLUSTER CDP PRIVATE CLOUD BASE

- Considérations générales relatives au dimensionnement

- Choix le bon matériel
- Considérations relatives au réseau
- Le déploiement cloud avec CDP Private Cloud
- Configuration des noeuds

CONFIGURATION AVANCÉE DU CLUSTER

- Configurer les ports de service
- Paramétrer HDFS et MapReduce
- Protection de la donnée avec Erasure Coding
- Comment gérer le passage à l'échelle
- Activer la Haute Disponibilité HDFS

MAINTENANCE DU CLUSTER

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

Jour 4

MONITORING DU CLUSTER

- Fonctionnalités de monitoring dans Cloudera Manager
- Les Health Tests
- Événements et alertes
- Graphiques et rapports
- Les bonnes pratiques de monitoring

DIAGNOSTIC DU CLUSTER

- Introduction
- Outils de diagnostic
- Exemples de mauvaises configurations

SÉCURITÉ

- Gouvernance des données avec SDX
- Les concepts de sécurité sur Hadoop
- Authentification sur Hadoop en utilisant Kerberos
- Les autorisations sur Hadoop
- Le chiffrement sur Hadoop

- Sécuriser un cluster Hadoop
- Introduction à Apache Ranger et Apache Atlas
- Sauvegarde et récupération

PRIVATE CLOUD vs. PUBLIC CLOUD

- Aperçu global de l'offre CDP
- Fonctionnalités de l'offre Private Cloud
- Fonctionnalités de l'offre Public Cloud
- Qu'est-ce que Kubernetes ?
- Aperçu de WXM
- Auto-scaling

Accessibilité

L'inclusion est sujet important pour OCTO Academy.
Nos référent·es sont à votre disposition pour faciliter l'adaptation de votre formation à vos besoins spécifiques.
Pour les contacter : academy.accessibilite@octo.com