



Formation  
**Cloudera  
Administration**

**DataTipsLearning.com**

**Tarif :  
2000 € HT**

**Durée :  
4 Jours  
(28h)**

**Lieu :  
PARIS/  
Distancielle/  
Votre Local**

**Contact :**

contact@datatipslearning.com

+33 0751386021

Réf. DTPSL-CLDA-2022

[S'inscrire](#)

**Plus de détails :**

Objectifs

Public / Prérequis

Programme détaillé

Méthode pédagogique

Cloudera a été cofondée en 2008. En 2018 Cloudera fusionne avec Hortonworks et s'oriente vers un monde du Cloud. Cloudera met en avant sa plateforme de gestion de données : la Cloudera Data Platform, un mix plus efficace entre l'ancienne plateforme Cloudera CDH et celle d'Hortonworks HDP.

La CDP permet une gestion unifiée, centralisée et hybride des infrastructures Big Data. Orientée aussi bien IT que métiers, elle élargie pour ses clients les capacités d'analyse de la donnée sur tous les cas d'usage Data existant.

# Objectifs

- Les bases de l'environnement Hadoop.
- Planifier le matériel et les logiciels pour un cluster Hadoop.
- Installation, administration et déploiement d'un cluster Hadoop via Cloudera Manager.
- Lancer des applications YARN, y compris MapReduce et Spark.
- Importer des données HDFS à partir de sources externes à l'aide de Sqoop et Flume.
- Installer de Hive et Impala.
- Déployer de clients Hadoop.
- Configurer HDFS pour le high availability.
- Gérer des ressources : le FIFO Scheduler et le FairScheduler.
- Gérer les Log files d'Hadoop.
- Utiliser le NameNode et Job Tracker Web UI.
- Dépanner, diagnostiquer, mettre au point et résoudre les problèmes sur Hadoop.

# Public

- Data Engineer
- Data Scientist
- Data Analyste
- Data Architectes

# Prérequis

- Connaissance de l'écosystème Hadoop.
- Quelques notions de scripting shell.

# Programme détaillé

01. Présentation d'Hadoop
02. Les modes d'installation
03. Introduction au Big Data
04. Présentation de Cloudera Data Platform
05. Présentation de Cloudera Manager

06. Installation d'un cluster CDP
07. Installation et Configuration de Hive, Impala, et Pig
08. Installation de Hue
09. Authentification et autorisation de Hue
10. Principe de gestion des nœuds et accès JMX
11. Configuration des ports Hadoop
12. Data locality : Configuration de HDFS
13. Configuration de la haute disponibilité HDFS
14. Gestion des utilisateurs et des groupes avec Cloudera Manager
15. Gestion des privilèges pour les utilisateurs locaux et LDAP
16. Configuration d'un cluster Cloudera
17. Paramètres et fichiers de configuration
18. Ajouter de nouveaux services
19. Ajouter et supprimer des hôtes
20. Le système de stockage HDFS
21. La mémoire NameNode
22. Hadoop File Shell
23. Cloudera Storage
24. La sécurité HDFS
25. Importation de données relationnelles avec Sqoop
26. Importation de données avec Flume
27. Intégration de données avec NiFi
28. Apache Kafka
29. Accès et analyse des données avec Apache Hive, Impala et Spark
30. Utilisation de Hue
31. Concepts du système de sécurité de Hadoop
32. Fonctionnement de Kerberos
33. Sécurisation d'un cluster Hadoop avec Kerberos
34. Les autres concepts de sécurité
35. Configuration des cgroups avec les pools de services statiques
36. Planification des ressources
37. Configuration des pools de ressources dynamiques
38. Mémoire YARN et paramètres du CPU
39. Planification de requêtes Impala
40. Vérification de l'état HDFS
41. Copier les données entre clusters
42. Ajout et suppression de nœud de cluster
43. Rééquilibrage du cluster snapshots de répertoires
44. Mise à niveau du cluster
45. Fonctionnalités de surveillance du Cloudera Manager
46. Surveillance des clusters Hadoop
47. Dépannage des clusters Hadoop

# Méthode pédagogique

La formation se compose d'une partie théorique, et également une partie pratique représentant 60% de de la formation.

La partie pratique contient plusieurs exercices sous forme de notebook Databricks avec les corrections, avec aussi un projet à la fin de la formation comme simulation d'une prod.

Chaque jour, une évaluation rapide des connaissances est effectuée avant de commencer les nouvelles parties de la formation.

A la fin, une synthèse globale est délivré aux stagiaires, renforcé par un projet prod.

Finalement, une évaluation QCM est proposée.

Un support de cours sera remis à chaque stagiaire comprenant les slides, les exercices et les corrigés et un git du projet prod.

Une feuille de présence est fournie en fin de formation avec une certificat de complétion de formation pour chaque stagiaire.

Le formateur est un Data Engineer expert, qui intervient sur le sujet depuis plusieurs années en formation mais aussi en conseil.