

Administrateur CDP Private Cloud



A propos de cette formation

Le cours de la formation « Administrateur CDP Private Cloud » à la demande fournit les connaissances fondamentales nécessaires pour effectuer les tâches de planification, d'approvisionnement, de configuration, de surveillance et de gestion requises d'un administrateur pour la Cloudera Data Platform (CDP) en tant que déploiement bare metal ou en tant que déploiement en Private Cloud.

MODALITÉS



Présentiel

Exposés : 65 %
Exercices : 35 %

Objectifs pédagogiques

- Connaître la plateforme CDP
- Gérer un portefeuille de projets data science

Qu'allez-vous apprendre ?

A l'issue de cette formation détaillant l'ensemble des caractéristiques d'un projet de Data Science à l'aide d'exercices interactifs, vous serez en mesure de :

Utilisez Cloudera Manager pour installer un cluster CDP Private Cloud Base.

Installer Cloudera Manager

Configurer et surveiller le cluster à l'aide de Cloudera Manager

Optimiser les performances des clusters

Contrôler et mesurer la performance projet

Effectuer les tâches de maintenance de routine du cluster

Public & Prérequis

Aux administrateurs de systèmes qui ont une expérience de base de Linux. Il n'est pas nécessaire d'avoir une connaissance préalable de la plateforme CDP, ni des plateformes antérieures telles que CDH de Cloudera ou Hortonworks HDP.

Prérequis : Aucun

★ Bonus participants

- Fiches mémoires résumant les points clés de la formation



Programme de la matinée

JOUR 1

Matin [3h] :

M1 - Qu'est-ce que Cloudera CDP ?

1. Cloudera CDP : Pourquoi et comment gérer cet asset ?
2. Hands-On Exercise

M2 - CDP Private Cloud Base Installation

1. - Installation Overview
2. - Cloudera Manager Installation
3. - Hands-On Exercise: Installing Cloudera Manager Server

Programme de l'après-midi

JOUR 1

Après-Midi [4h]

M1 - CDP Private Cloud Base Installation

1. CDP Runtime Overview
2. - Hands-On Exercise: Cluster Installation

M2 - Cluster Configuration

1. Overview et configuration
2. Modifying Service Configurations
3. Configuration Files
4. Managing Role Instances
5. Adding New Services, adding and Removing Hosts
6. Identifier les besoins et le type de données à traiter avec les métiers (use cases)
7. Exemple de cas d'usages classiques Data
8. Retour d'expérience sur quelques cas déjà instruits chez Quantmetry
9. Expression de besoin d'un use case data (Cadrage du besoin avec la méthode SMART)

M3 - Cluster Configuration - HANDS-ON

1. Configuring a Hadoop Cluster



Programme de la matinée

JOUR 2

Matin [3h] :

M1 - Data Storage

1. HDFS Topology and Roles
2. HDFS Performance and Fault Tolerance
3. HDFS and Hadoop Security Overview
4. Working with HDFS
5. HBase Overview
6. Kudu Overview
7. Cloud Storage Overview

M2 - Hands-On Exercise

1. Working with HDFS
2. Storing Data in Amazon S3

Programme de l'après-midi

JOUR 2

Après-Midi [4h]

M1 - Data Ingest Part 1

1. Data Ingest Overview
2. File Formats
3. Ingesting Data using File Transfer or REST Interfaces
4. Importing Data from Relational Databases with Apache Sqoop
5. Hands-On Exercise: Importing Data Using Sqoop

M2 - Data Ingest Part 2

1. Ingesting Data Using NiFi
2. Instructor-Led Demonstration: NiFi User Interface
3. Best Practices for Importing Data
4. Hands-On Exercise: NiFi Verification



Programme de la matinée

JOUR 3

Matin [3h] :

M1 - Data Compute Part 1

1. YARN Overview
2. Running Applications on YARN
3. Viewing YARN Applications
4. YARN Application Logs
5. MapReduce Applications
6. YARN Memory and CPU Settings
7. Hands-On Exercise

M2 - Data Compute Part 2

1. Hive on Tez
2. ACID for Hive
3. Spark Overview
4. How Spark Applications Run on YARN
5. Monitoring Spark Applications
6. Phoenix Overview
7. Hands-On Exercise

Programme de l'après-midi

JOUR 3

Après-Midi [4h]

M1 - Managing Resources

1. Managing Resources Overview
2. Node Labels
3. Configuring cgroups
4. The Capacity Scheduler
5. Managing Queues
6. Impala Query Scheduling
7. Hands-On Exercise

M2 - Planning Your Cluster

1. IGeneral Planning Considerations
2. Choosing the Right Hardware
3. Network Considerations
4. CDP Private Cloud Considerations
5. Configuring Nodes



Programme de la matinée

JOUR 4

Matin [3h] :

M1 - Advanced Cluster Configuration

1. Configuring Service Ports
2. Tuning HDFS and MapReduce
3. Managing Cluster Growth
4. Erasure Coding
5. Enabling High Availability for HDFS and YARN
6. Hands on

M2 - Cluster Maintenance

1. Checking HDFS Status
2. Copying Data Between Clusters
3. Rebalancing Data in HDFS
4. HDFS Directory Snapshots
5. Hands on

Programme de l'après-midi

JOUR 4

Après-Midi [4h]

M1 - Cluster Upgrade

1. Host Maintenance
2. Upgrading a Cluster using CM

M2 - Cluster Monitoring

1. Cloudera Manager Monitoring Features
2. Health Tests
3. Hands-On Exercise: Breaking the Cluster
4. Events and Alerts
5. Charts and Reports
6. Monitoring Recommendations
7. Hands-On Exercise: Confirm Cluster Healing and Configuring Email Alerts



Programme de la matinée

JOUR 5

Matin [3h] :

M1 - Cluster Troubleshooting

1. Overview
2. Troubleshooting Tools
3. Completed
4. Misconfiguration Examples

M2 -Security

1. Data Governance with SDX
2. Hadoop Security Concepts
3. Hadoop Authentication Using Kerberos
4. Hadoop Authorization
5. Hadoop Encryption
6. CM autotls
7. Securing a Hadoop Cluster
8. Apache Ranger
9. Apache Atlas
10. Backup and Recovery

Programme de l'après-midi

JOUR 5

Après-Midi [4h]

M1-CDP Private Cloud Base Installation

1. CDP Overview
2. Private Cloud Capabilities
3. Public Cloud Capabilities
4. What Is Kubernetes?
5. Workload XM Overview
6. Auto-scaling

M2-Quizz test