

HADOOP - CLUDERA POUR ARCHITECTES ET ADMINISTRATEURS

3 jours en présentiel (21 heures)

Objectifs pédagogiques

Cette formation Hadoop - Cloudera pour architectes et administrateurs vous permettra de savoir :
Comment élaborer une architecture Hadoop, comment Hadoop Distributed File System (HDFS) et YARN/MapReduce fonctionnent, comment exploiter et gérer un cluster Hadoop, comment configurer Hadoop et ses outils, comment mettre en œuvre un cluster Hadoop, comment gérer la maintenance des nœuds dans un cluster, comment déplacer et gérer les données dans un cluster, comment administrer un cluster Hadoop.

Population visée

Cette formation Hadoop - Cloudera pour architectes et administrateurs est destinée aux administrateurs systèmes et aux personnes qui gèrent et maintiennent les clusters Apache Hadoop en production ou qui évoluent dans un environnement de dev

Pré-requis

Cette formation Hadoop - Cloudera pour architectes et administrateurs nécessite de savoir naviguer et modifier des fichiers dans Linux. Une connaissance préliminaire d'Hadoop n'est pas exigée.

Méthodes pédagogiques

1 poste et 1 support par stagiaire 8 à 10 stagiaires par salle Remise d'une documentation pédagogique papier ou numérique pendant le stage La formation est constituée d'apports théoriques, d'exercices pratiques, de réflexions et de retours d'expérience

Formateur

Formateur consultant expert en BIG DATA

Modalités de validation des acquis

Auto-évaluation des acquis par le stagiaire via un questionnaire en ligne Attestation de fin de stage remise au stagiaire

Contenu

INTRODUCTION

- Présentation générale d'Hadoop
- Exemples d'utilisation dans différents secteurs
- Historique et chiffres clés : Quand parle-t-on de Big Data ?
- Les possibilités d'implantation de l'infrastructure et les coûts associés

L'ECOSYSTEME D'HADOOP

- Le système de fichier HDFS
- Le paradigme MapReduce et l'utilisation à travers YARN
- Le garant des informations : Zookeeper
- Le système de stockage de la donnée basé sur le socle : HBase
- La présentation de l'architecture globale : où installer les services ?
- La définition du dimensionnement pour un cluster Hadoop
- Les particularités de MapR

INSTALLATION DU SOCLE D'HADOOP

- Installation et configuration manuelle pas-à-pas sur plusieurs machines de : HDFS, YARN, Zookeeper, HBase

MANIPULATION DES DONNEES DANS UN CLUSTER HADOOP

- Architecture, Objectif et fonctionnement de : Hue, Hive, Pig, Oozie, Sqoop, Flume, Spark
- Mise en pratique sur les différents outils

EXPLOITATION D'UN CLUSTER HADOOP

- Architecture, objectif et fonctionnement du Manager : MapR Control System
- Mise en pratique

CONCEPTS D'ARCHITECTURE COMPLEXE

- Architecture en rack : les avantages et fonctionnement et paramétrage du rack awareness
- Scalabilité et performances : le Namenode Federation
- High Availability des serveurs maîtres
- Utilisation des snapshots : sauvegarde et restauration