

Formation Hadoop : Déployer du Big Data

Durée :	4 jours
Public :	Administrateurs systèmes
Pré-requis :	Connaissances en administration système, préférablement Java
Objectifs :	Comprendre le Big Data et ses enjeux - Savoir déployer Hadoop et son écosystème - Comprendre HDFS, MapReduce - Structurer les données avec HBase - Écrire des requêtes avec HiveQL - Lancer une analyse avec Pig
Référence :	BUS869-F
Demandeurs d'emploi :	Contactez-nous pour connaître les remises Pôle Emploi

Introduction au Big Data

Qu'est-ce que le Big Data ?
Source des données : l'homme, la machine
La problématique de taille
Position de Hadoop dans le paysage

Introduction à Hadoop

L'origine du projet
Le système de fichiers HDFS
Comprendre l'algorithme MapReduce
L'environnement d'Hadoop : HBase, ZooKeeper, Hive, Pig...
L'API YARN

Mettre Hadoop en place : HDFS

Du mode autonome au mode complètement distribué en cluster
Pré-requis, distributions Hadoop
Cluster Hadoop : NameNode, ResourceManager, DataNode, NodeManager
Les fichiers de configuration
Opérations de base sur le cluster HDFS : formatage, démarrage, arrêt
Atelier pratique : installer Hadoop sur 2 nœuds, formater et manipuler HDFS

Travailler avec MapReduce

L'intérêt de MapReduce
Mappers, reducers, parallélisme et indépendance des traitements
Entrées, sorties
Soumission d'un job à Hadoop
Atelier pratique : exécuter une tâche via MapReduce, avec sortie dans HDFS

Une base de données distribuée : HBase

L'accès aléatoire, temps réel, lecture-écriture au Big Data
Fonctionnalités de HBase, NoSQL

Pré-requis, configuration

Manipulation via le shell HBase

Atelier pratique : mettre en place HBase sur Hadoop, créer et manipuler une table

Et pourquoi pas un peu de SQL avec Hive ?

Présentation de Hive

Gérer le schéma : bases, tables, vues, partitions

Manipulation des données, requêtes et map-reduce avec HiveQL

Audits et journal d'erreurs

Atelier pratique : chargement de données massives dans Hive, requêtes

Analyser les données avec Pig

Présentation, installation du projet Apache Pig

Exécutions de Pig en local, en mode map-reduce

Scripter pour Pig

Le langage Pig Latin

Manipulations de données et stockage avec Pig

Atelier pratique : écrire un script contenant du Pig Latin pour une tâche simple, et l'exécuter en local, puis en mode map reduce.

Aller plus loin avec Hadoop

Gérer les logs et l'audit de tâches Hadoop

Découvrir MRUnit pour les test unitaires dans Hadoop

Débogage en local

Surveillance des performances

Atelier pratique : mise en place d'un job MapReduce plus complexe avec traces et tests unitaires