DAWAN

Plan de la formation

www.dawan.fr

Formation Hadoop : Deploiement + Administration

Durée :	5 jours (35 heures)
Tarifs inter-entreprise :	3 875,00 CHF HT (standard) 3 100,00 CHF HT (remisé)
Public :	Tous
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 - Ecrire des requêtes avec HiveQL - Installer les services d'un nœud Hadoop - Assembler plusieurs nœuds Hadoop - Déployer une nouvelle application sur un cluster existant - Effectuer une restauration de données suite à une reprise sur incident
Modalités pédagogiques, techniques et d'encadrement :	• Formation synchrone en présentiel et distanciel. • Méthodologie basée sur l'Active Learning : 75 % de pratique minimum. • Un PC par participant en présentiel, possibilité de mettre à disposition en bureau à distance un PC et l'environnement adéquat. • Un formateur expert.
Modalités d'évaluation :	• Définition des besoins et attentes des apprenants en amont de la formation. • Auto-positionnement à l'entrée et la sortie de la formation. • Suivi continu par les formateurs durant les ateliers pratiques. • Évaluation à chaud de l'adéquation au besoin professionnel des apprenants le dernier jour de formation.
Sanction :	Attestation de fin de formation mentionnant le résultat des acquis
Référence :	BUS101732-F
Note de satisfaction des participants:	Pas de données disponibles
Contacts :	commercial@dawan.fr - 09 72 37 73 73

Modalités d'accès :	Possibilité de faire un devis en ligne (www.dawan.fr, moncompteformation.gouv.fr, maformation.fr, etc.) ou en appelant au standard.
Délais d'accès :	Variable selon le type de financement.
Accessibilité :	Si vous êtes en situation de handicap, nous sommes en mesure de vous accueillir, n'hésitez pas à nous contacter à referenthandicap@dawan.fr, nous étudierons ensemble vos besoins

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

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 perfomances

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

Administration de Hadoop

Présentation d’un nœud existant
Organisation des services et étude du séquencement avec YARN

Atelier : modifier la taille des blocs HDFS pour diminuer le nombre de Map/Reduce

Mettre Hadoop en place

Relation entre la plateforme installée et les framework de développement
Proposer de frameworks indépendants pour assurer la compatibilité : Spring Data

Atelier : déployer une application d’accès à HBase au travers d’un mapping O/R Spring Data

Travailler avec MapReduce

Déployer un programme Map/Reduce sur un cluster de nœuds Hadoop
Recherche des logs
Remonter les anomalies aux développeurs
Proposer l’usage de file Kafka

Atelier : utilisation de file d’entrée sortie pour un programme Map/Reduce

Routage de données

Définition de routes logicielles
Mettre en place un cas de calcul où les données déclenchent les programmes

Atelier : faire un routage de données depuis un répertoire HDFS vers une file Kafka qui est l’entrée d’un programme Map/Reduce

Utilisation des vues

Utilisation des vues Ambari
Visualisation de l’état des nœuds d’un cluster
Importer/exporter des fichiers de configuration

Atelier : relancer une grappe de services, utilisation des vues YARN et Tez

Gestion des droits

Gestion des comptes utilisateurs
Gestion des droits de fichier sur un système de fichier distribué
Utilisation de certificat

Atelier : configurer les services Knox et Ranger