BIG DATA : DEVELOPPEMENT DE SOLUTIONS AVEC HADOOP®

Public	Aux développeurs, architectes et testeurs désireux de posséder une expérience pratique d'écriture de code pour Hadoop. Elle peut aussi être utile aux managers techniques intéressés par les processus de développement.
Durée	4 jours - 28 heures
Pré-requis	Vous devez avoir une expérience de Java
Objectifs	La gestion des ensembles de données volumineux offre aux entreprises de toutes tailles de nouvelles opportunités et de nouveaux défis à relever. Au cours de cette formation, vous allez acquérir les compétences pratiques de programmation nécessaires pour développer des solutions compatibles avec la plateforme Hadoop d'Apache grâce auxquelles vous pourrez traiter efficacement différents types de Big Data.
Méthodes pédagogiques	Pour bien préparer la formation, le stagiaire remplit une évaluation de positionnement et fixe ses objectifs à travers un questionnaire. La formation est délivrée en présentiel ou distanciel (e-learning, classe virtuelle, présentiel et à distance). Le formateur alterne entre méthodes démonstratives, interrogatives et actives (via des travaux pratiques et/ou des mises en situation). La validation des acquis peut se faire via des études de cas, des quiz et/ou une certification. Cette formation est animée par un consultant-formateur dont les compétences techniques, professionnelles et pédagogiques ont été validées par des diplômes et/ou testées et approuvées par l'éditeur et/ou par Audit Conseil Formation.
Moyens techniques	1 poste de travail complet par personne De nombreux exercices d'application Mise en place d'ateliers pratiques Remise d'un support de cours Passage de certification(s) dans le cadre du CPF Remise d'une attestation de stage
Modalité d'évaluation des acquis	Evaluation des besoins et objectifs en pré et post formation Evaluation technique des connaissances en pré et post formation Evaluation générale du stage
Délai d'accès	L'inscription à cette formation est possible jusqu'à 5 jours ouvrés avant le début de la session
Accessibilité handicapés	Au centre d'affaires ELITE partenaire d'ACF à 20 m. Guide d'accessibilité à l'accueil.

PRÉSENTATION D'HADOOP

- Évaluer la valeur que peut apporter Hadoop à l'entreprise
- Examiner l'écosystème d'Hadoop
- Choisir un modèle de distribution adapté

DÉFIER LA COMPLEXITÉ DE LA PROGRAMMATION PARALLÈLE

- Examiner les difficultés liées à l'exécution de programmes parallèles : algorithmes, échange des données
- Évaluer le mode de stockage et la complexité du Big Data

PROGRAMMATION PARALLÈLE AVEC MAPREDUCE

- Fragmenter et résoudre les problèmes à grande échelle
- Découvrir les tâches compatibles avec MapReduce
- Résoudre des problèmes métier courants

APPLIQUER LE PARADIGME HADOOP MAPREDUCE

- Configurer l'environnement de développement
- Examiner la distribution Hadoop
- Étudier les démons Hadoop
- Créer les différents composants des tâches MapReduce
- Analyser les différentes étapes de traitement MapReduce : fractionnement, mappage, lecture aléatoire et réduction

CRÉER DES TÂCHES MAPREDUCE COMPLEXES

- Choisir et utiliser plusieurs outils de mappage et de réduction, exploiter les partitionneurs et les fonctions map et reduce intégrées, coordonner les tâches avec le planificateur de workflows Oozie, rationaliser les tâches dans différents langages de programmation

RÉSOUDRE LES PROBLÈMES DE MANIPULATION DES DONNÉES

- Exécuter les algorithmes : tris, jointures et recherches parallèles, analyser les fichiers journaux, les données des média sociaux et les courriels

MISE EN œUVRE DES PARTITIONNEURS ET DES COMPARATEURS

- Identifier les algorithmes parallèles liés au réseau, au processeur et aux E/S de disque
- Répartir la charge de travail avec les partitionneurs
- Contrôler l'ordre de groupement et de tri avec les comparateurs
- Mesurer les performances avec les compteurs

BIEN-FONDÉ DES DONNÉES DISTRIBUÉES

- Optimiser les performances du débit des données
- Utiliser la redondance pour récupérer les données

INTERFACER AVEC LE SYSTÈME DE FICHIERS DISTRIBUÉ HADOOP

- Analyser la structure et l'organisation du HDFS
- Charger des données brutes et récupérer le résultat
- Lire et écrire des données avec un programme
- Manipuler les types SequenceFile d'Hadoop
- Partager des données de référence avec DistributedCache

STRUCTURER LES DONNÉES AVEC HBASE

- Passer du stockage structuré au stockage non structuré
- Appliquer les principes NoSQL avec une application de modèle à la lecture, se connecter à HBase à partir des tâches MapReduce, comparer HBase avec d'autres types de magasins de données NoSQL

EXPLOITER LA PUISSANCE DE SQL AVEC HIVE

- Structurer bases de données, les tables, les vues et les partitions
- Extraire, transformer et charger les données
- Lancer des requêtes avec HiveQL
- Accéder aux servers Hive via IDBC, ajouter des fonctionnalités à HiveQL avec les fonctions définies par l'utilisateur

EXÉCUTER DES WORKFLOWS AVEC PIG

- Développer des scripts Pig Latin pour consolider les workflows, intégrer des requêtes Pig à Java
- Interagir avec les données par le biais de la console Grunt
- Étendre Pig avec les fonctions définies par l'utilisateur

TESTER ET DÉBOGUER LE CODE HADOOP

- Enregistrer des événements importants à auditer et à déboguer
- Valider les spécifications avec MRUnit
- Déboguer en mode local

DÉPLOYER, SURVEILLER ET AFFINER LES PERFORMANCES

- Déployer la solution sur un cluster de production, utiliser des outils d'administration pour optimiser les performances, surveiller l'exécution des tâches via les interfaces utilisateur web

NOUS CONTACTER

Siège social
16, ALLÉE FRANÇOIS VILLON
38130 ÉCHIROLLES

Téléphone
04 76 23 20 50 - 06 81 73 19 35

Suivez-nous sur les réseaux sociaux, rejoignez la communauté !

ACF Audit Conseil Formation

@ACF_Formation

Dernière mise à jour : 03/04/2020