

Formation Programmer Hadoop en Java : Approfondissement

Durée :	2 jours
Public :	Développeurs Java, Administrateurs (DBA ou systèmes)
Pré-requis :	Avoir suivi Programmer Hadoop en Java Initiation ou notions équivalentes
Objectifs :	Charger et récupérer des données de HBase et du système de fichiers distribué Hadoop (HDFS) - User Defined Functions de Hive et Pig
Sanction :	Attestation de fin de stage mentionnant le résultat des acquis
Taux de retour à l'emploi:	Aucune donnée disponible
Référence:	BUS102000-F
Note de satisfaction des participants:	Pas de données disponibles

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
Intégrer des travaux MapReduce avec des requêtes Hive
Lancer des requêtes avec HiveQL
Accéder aux servers Hive via JDBC, ajouter des fonctionnalités à HiveQL 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