

Formation NoSQL : Fondamentaux

Durée :	3 jours
Public :	Toute personne souhaitant profiter des nombreux avantages liés aux technologies dédiées au Big Data
Pré-requis :	Avoir des connaissances pratiques de la plateforme Microsoft Windows. Des notions de programmation sont utiles sans toutefois être obligatoires - Intégrer les composants Big Data pour créer un Data Lake approprié - Sélectionner des entrepôts de Big Data adaptés pour gérer plusieurs ensembles de données - Traiter des ensembles de données volumineux avec Hadoop pour faciliter la prise de décisions techniques et métier - Interroger des ensembles de données volumineux en temps réel
Objectifs :	
Sanction :	Attestation de fin de stage mentionnant le résultat des acquis
Taux de retour à l'emploi:	Aucune donnée disponible
Référence:	BAS100301-F
Note de satisfaction des participants:	Pas de données disponibles

Définition

Les quatre dimensions du Big Data : volume, vitesse, variété, véridité Présentation de l'ensemble MapReduce, stockage et requêtes

Améliorer les résultats de l'entreprise grâce au Big Data

Mesurer l'importance du Big Data au sein d'une entreprise
Réussir à extraire des données utiles
Intégrer le Big Data aux données traditionnelles

Analyser les caractéristiques de vos données

Sélectionner les sources de données à analyser
Supprimer les doublons
Définir le rôle de NoSQL

Présentation des entrepôts de Big Data

Modèles de données : valeur clé, graphique, document, famille de colonnes Système de fichiers distribué Hadoop (HDFS)
HBase
Hive

Cassandra
Hypertable
Amazon S3
BigTable
DynamoDB
MongoDB
Redis
Riak
Neo4J

Choisir un entrepôt de Big Data

Choisir un entrepôt de données en fonction des caractéristiques de vos données
Injecter du code dans les données, mettre en œuvre des solutions de stockage des données multilingues
Choisir un entrepôt de données capable de s'aligner avec les objectifs de l'entreprise

Intégrer différents entrepôts de données

Mapper les données avec le framework de programmation, se connecter aux données et les extraire de l'entrepôt de stockage, transformer les données à traiter
Fractionner les données pour Hadoop MapReduce

Utiliser Hadoop MapReduce

Créer les composants des tâches Hadoop MapReduce
Distribuer le traitement des données entre plusieurs fermes de serveurs, exécuter les tâches Hadoop MapReduce
Surveiller l'avancement des flux de tâches

Principes fondamentaux de Hadoop MapReduce

Identifier les démons Hadoop
Examiner le système de fichiers distribué Hadoop (HDFS)
Choisir le mode d'exécution : local, pseudo-distribué, entièrement distribué

Gérer des données de streaming

Comparer des modèles de traitement en temps réel
Exploiter Storm pour extraire des événements en direct
Traitement rapide avec Spark et Shark

Synthétiser les tâches Hadoop MapReduce avec Pig

Communiquer avec Hadoop en Pig Latin
Exécuter des commandes avec le shell Grunt
Rationaliser les traitements de haut niveau

Lancer des requêtes ad hoc sur le Big Data avec Hive

Assurer la persistance des données dans le Hive MegaStore
Lancer des requêtes avec HiveQL
Examiner le format des fichiers Hive

Extraire des données donnant de la valeur à l'entreprise

Analyser les données avec Mahout, utiliser des outils de génération d'états pour afficher le résultat du traitement
Interroger en temps réel avec Impala

Élaborer une stratégie dédiée au Big Data

Définir les besoins en matière de Big Data
Atteindre les objectifs grâce à la pertinence des données
Évaluer les différents outils du marché dédiés au Big Data
Répondre aux attentes du personnel de l'entreprise

Une méthode analytique innovante

Identifier l'importance des traitements métier
Cerner le problème
Choisir les bons outils
Obtenir des résultats exploitables

Mettre en œuvre une solution Big Data

Bien choisir les fournisseurs et options d'hébergement
Trouver le juste équilibre entre les coûts engendrés et la valeur apportée à l'entreprise
Garder une longueur d'avance