

Formation Hadoop – Développement

Code : NTE13

Durée : 4 jours

Classe : Présentiel / à distance

Public

- Cette formation s'adresse aux développeurs souhaitant concevoir des applications Big Data basées sur Hadoop administrateurs systèmes, ainsi qu'aux chefs de projets et toute personne désirant comprendre les architectures et le stockage de données dans un environnement Hadoop.

Prérequis

- Compréhension des systèmes distribués
- Bases en gestion de données
- Bonne expérience en développement Java

Objectifs

- Compréhension des systèmes distribués
- Bases en gestion de données
- Bonne expérience en développement Java

Programme détaillé

1- Le big data

- Définition du périmètre du big data
- Le rôle du projet Hadoop
- Les concepts de base des projets big data
- Présentation du cloud computing
- Différence entre cloud computing privé et public
- Les architectures big data à base du projet Hadoop

2- Collecte de données et application de MapReduce

- Analyse des flux de données dans l'entreprise
- Données structurées et non-structurées
- Les principes de l'analyse sémantique des données d'entreprise
- Graphe des tâches à base de MapReduce
- La granularité de cohérence des données
- Transfert de données depuis un système de persistance dans Hadoop
- Transfert de données d'un Cloud dans Hadoop

3- Le stockage des données avec HBase

- Plusieurs types de base de données XML
- Patterns d'usages et application au cloud
- Application de Hadoop database au sein d'un workflow
- Utilisation des projets Hive/Pig
- Utilisation du projet HCatalog
- L'API Java HBase

4- Le stockage des données sur HDFS

- Patterns d'usages et application au cloud
- Architecture et installation d'un système HDFS, journal, NameNode, DataNode
- Opérations, commandes et gestion des commandes
- L'API HDFS Java
- Analyse de données avec Apache Pig
- Le langage Pig Latin et utiliser Apache Pig avec Java
- Requêtage avec Apache Hive
- Réplication de données et partage de données sur une architecture HDFS



Formation Hadoop – Développement

Code : NTE13

Durée : 4 jours

Classe : Présentiel / à distance

5- Spring Data Hadoop

- Introduction à Spring et Spring Data
- Le namespace Hadoop pour Spring
- Utiliser Spring pour simplifier la configuration Hadoop
- Configuration du cache distribué
- Définition des jobs et dépendance entre jobs
- Intégration des outils (Pig, Hive...)

