

Formation Big Data Hadoop – Analyse du Big Data en environnement Hadoop

Code : NTE14**Durée :** 3 jours**Classe :** Présentiel / à distance

Public

- Cette formation s'adresse aux développeurs, analystes de données, ingénieurs IT, administrateurs de bases de données et étudiants en informatique.

Prérequis

- Bases en manipulation de données
- Notions en systèmes d'information
- Motivation pour le Big Data

Objectifs

- Comprendre l'architecture Hadoop
- Gérer les données sur HDFS
- Effectuer des requêtes à l'aide de Hive
- Automatiser les workflows Big Data

Programme détaillé

1- Introduction à Hadoop

- Présentation générale d'Hadoop
- Exemples d'utilisation dans différents secteurs
- Historique et chiffres clés : Quand parle-t-on de Big Data ?

2- L'écosystème d'Hadoop

- Le système de fichier HDFS
- Le paradigme MapReduce et l'utilisation à travers YARN

3- Manipulation des données dans un cluster Hadoop

- Hue : Comment fonctionne cette interface web ?
- Hive : Pourquoi Hive n'est pas une base de données ?

4- Requête sur Hive

- Utilisation de HCatalog
- Utilisation avancée sur Hive
- Utilisation de fonctions utilisateurs
- Paramétrage de requête
- Pig : Fonctionnement de Pig

5- Programmation avec Pig Latin

- Utilisation du mode Local
- Utilisation de fonctions utilisateurs
- Tez : Qu'est-ce que Tez ?

6- Comment et quand l'utiliser ?

- Oozie : Fonctionnement de Oozie

7- Création de Workflows avec Oozie

- Manipulation des workflows
- Ajout d'éléments d'exploitation dans les workflows
- Ajout de conditions d'exécution
- Paramétrage des workflows
- Sqoop : A quoi sert Sqoop ?

8- Chargement des données depuis une base de données relationnelle

- Chargement des données depuis Hadoop
- Utilisation et paramétrage avancée
- Les particularités des distributions : Impala, Hawq
- Quelles sont les bonnes pratiques d'utilisation des différents outils ?