

## PRÉSENTATION HADOOP

<b>Durée</b>	<b>2 jours</b>	<b>Référence Formation</b>	<b>4-IT-HADO</b>
--------------	----------------	----------------------------	------------------

### Objectifs

Découvrir les concepts et les enjeux liés à Hadoop  
Comprendre le fonctionnement de la plateforme et de ses composants

### Participants

Administrateurs de cluster Hadoop, développeurs, data scientist

### Pré-requis

Bonnes connaissances de l'administration Linux

### Moyens pédagogiques

Réflexion de groupe et apports théoriques du formateur  
Travail d'échange avec les participants sous forme de réunion-discussion  
Utilisation de cas concrets issus de l'expérience professionnelle  
Validation des acquis par des questionnaires, des tests d'évaluation, des mises en situation et des jeux pédagogiques.  
Remise d'un support de cours.

### PROGRAMME

#### Introduction à Hadoop

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

#### L'écosystème d'Hadoop

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

#### Manipulation des données dans un cluster Hadoop

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

#### Requête sur Hive

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

#### Programmation avec Pig Latin

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

#### Comment et quand l'utiliser ?

- Oozie : Fonctionnement de Oozie

#### 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 ?

#### **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 ?