

# PRÉSENTATION HADOOP

**Durée**

**2 jours**

**Référence Formation**

**4-IT-ISTQB**

## 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

Accueil des stagiaires dans une salle dédiée à la formation équipée d'un vidéo projecteur, tableau blanc et paperboard ainsi qu'un ordinateur par participant pour les formations informatiques.  
Positionnement préalable oral ou écrit sous forme de tests d'évaluation, feuille de présence signée en demi-journée, évaluation des acquis tout au long de la formation.  
En fin de stage : QCM, exercices pratiques ou mises en situation professionnelle, questionnaire de satisfaction, attestation de stage, support de cours remis à chaque participant.  
Formateur expert dans son domaine d'intervention  
Apports théoriques et exercices pratiques du formateur  
Utilisation de cas concrets issus de l'expérience professionnelle des participants  
Réflexion de groupe et travail d'échanges avec les participants  
Pour les formations à distance : Classe virtuelle organisée principalement avec l'outil ZOOM.  
Assistance technique et pédagogique : envoi des coordonnées du formateur par mail avant le début de la formation pour accompagner le bénéficiaire dans le déroulement de son parcours à distance.

## 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

## CAP ÉLAN FORMATION

www.capelanformation.fr - Tél : 04.86.01.20.50  
Mail : contact@capelanformation.fr  
Organisme enregistré sous le N° 76 34 0908834  
[version 2023]

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 ?