

Référence	<b>4-IT-HADO</b>
Durée	<b>2 jours (14 heures)</b>
Éligible CPF	<b>NON</b>
Mise à jour	<b>27/11/2023</b>

## Présentation Hadoop



### OBJECTIFS PÉDAGOGIQUES

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



### PUBLIC CONCERNÉ

Administrateurs de cluster Hadoop, développeurs, data scientist



### PRÉREQUIS

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.



### MODALITÉS D'ÉVALUATION

- Feuille de présence signée en demi-journée,
- Evaluation des acquis tout au long de la formation,
- Questionnaire de satisfaction,
- Attestation de stage à chaque apprenant,
- Positionnement préalable oral ou écrit,
- Evaluation formative tout au long de la formation,
- Evaluation sommative faite par le formateur ou à l'aide des certifications disponibles



### MOYENS TECHNIQUES EN PRÉSENTIEL

Accueil des stagiaires dans une salle dédiée à la formation équipée à minima d'un vidéo projecteur et d'un tableau blanc et/ou paperboard.

Pour les formations nécessitant un ordinateur, un PC est mis à disposition de chaque participant.



### MOYENS TECHNIQUES EN DISTANCIEL

A l'aide d'un logiciel (Teams, Zoom...), d'un micro et éventuellement d'une caméra les apprenants interagissent et communiquent entre eux et avec le formateur.

Sessions organisées en inter comme en intra entreprise.

L'accès à l'environnement d'apprentissage ainsi qu'aux preuves de suivi et d'assiduité (émargement, évaluation) est assuré.

Pour toute question avant et pendant le parcours, assistance technique à disposition au 04 67 13 45 45.



### ORGANISATION

Délai d'accès : 5 jours ouvrés  
(délai variable en fonction du financeur)

Les cours ont lieu de 9h à 12h30 et de 13h30 à 17h



### ACCESSIBILITÉ

Les personnes en situation de handicap sont invitées à nous contacter directement, afin d'étudier ensemble les possibilités de suivre la formation.

Pour tout renseignement, notre référent handicap reste à votre disposition : mteyssedou@ait.fr



### PROFIL FORMATEUR

Nos formateurs sont des experts dans leurs domaines d'intervention

Leur expérience de terrain et leurs qualités pédagogiques constituent un gage de qualité.



### CERTIFICATION POSSIBLE

Aucune

# Présentation Hadoop

## 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 ?