

# Les fondamentaux d'Hadoop

## DESCRIPTION

Cette formation est une initiation aux fondamentaux d'Hadoop. Elle donne aux participants une connaissance théorique et pratique de la plateforme, au travers de plusieurs exercices pratiques appliqués à des cas réels. A l'issue de la session, les participants seront en capacité d'utiliser les outils de l'écosystème Hadoop pour explorer des données stockées sur un entrepôt Big Data.

## OBJECTIFS PÉDAGOGIQUES

Appréhender le fonctionnement d'Hadoop  
Identifier l'écosystème : quels outils pour quels usages ?  
Manipuler les principales commandes shell d'interaction avec Hadoop  
Emettre des requêtes SQL avec Hive et HCatalog  
Créer des traitements de données avec Pig

## PUBLIC CIBLE

Analyste  
Data scientist  
Architecte  
Développeur

## PRÉ-REQUIS

- Connaissances de base en programmation ou en scripting.

## MÉTHODE PÉDAGOGIQUE

Apports théoriques préalables présentant la plateforme Hadoop, son fonctionnement et son écosystème.  
Travaux pratiques pour une mise en application immédiate.  
Les travaux pratiques s'effectueront sur un cluster Apache Hadoop 2.x.

## PROFILS DES INTERVENANTS

Toutes nos formations sont animées par des consultants-formateurs expérimentés et reconnus par leurs pairs.

## MODALITÉS D'ÉVALUATION

L'évaluation des acquis se fait tout au long de la session au travers des ateliers et des mises en pratique. Une évaluation à chaud est réalisée systématiquement en fin de session.

## POUR ALLER PLUS LOIN :

- Livre blanc "Hadoop, feuille de route"
- Toutes nos formations Hadoop
- Formation officielle Hortonworks "Administrer la plateforme Hadoop 2.X Hortonworks 1" (HDP Operations: Hadoop Administration 1) (HWADM)
- Formation officielle Hortonworks "Administrer la plateforme Hadoop 2.X Hortonworks : niveau avancé" (HDP Administrator: Hadoop Administration 2) (HWAD2)

## Stage pratique HADOOP HORTONWORKS

Code :  
**HDP01**

Durée :  
**2 jours (14 heures)**

Exposés :  
**60%**

Cas pratiques :  
**30%**

Échanges d'expérience :  
**10%**

### Sessions à venir :

16 - 17 nov. 2017  
Genève / 1 900 CHF

14 - 15 déc. 2017  
Paris / 1 580 € HT

14 - 15 mai 2018  
Paris / 1 590 € HT

27 - 28 sept. 2018  
Paris / 1 590 € HT

Tarif & dates intra :  
**Sur demande**

- Formation officielle Hortonworks " Administrer la plateforme Hadoop 2.X Hortonworks : sécurité " (HDP Operations: Security) (HWADS)
- Formation officielle Hortonworks " Analyse de données pour Hadoop 2.X Hortonworks avec Pig, Hive et Spark " (HDP Developer: Apache Pig and Hive) (HWAPH)
- Formation officielle Hortonworks " Développer des applications pour Apache Spark avec Python ou Scala " (HDP Developer: Enterprise Spark 1) (HWSPK)
- Toutes nos formations Data Science
- Formation " Fondamentaux de la Data Science " (DSFDX)
- Formation " Data Science : niveau avancé " (DSNVA)

## Programme pédagogique détaillé par journée

### Jour 1

#### INTRODUCTION AU BIG DATA

- Qu'est-ce que Big Data ?
- Les grands enjeux métier
- Paysage technologique : les architectures Big Data

#### INTRODUCTION À HADOOP

- Historique succinct
- Le cœur de la plateforme : HDFS et YARN
- L'écosystème Hadoop
  - Frameworks et algorithmes
  - Bases de données
  - Traitements des données
  - Intégration
- Hadoop et la sécurité des données

#### MANIPULER LA LIGNE DE COMMANDE HADOOP

- Présentation des principales commandes
- Mise en pratique "Manipulation et transfert de fichiers en ligne de commande"

#### UNE INTERFACE UTILISATEUR POUR HADOOP : HUE

- Présentation de Hue et de ses modules
- Mise en pratique "Manipulation interactive de données"

#### INTERROGER HADOOP AVEC DU SQL : HIVE

- Présentation de Hive
- Mise en pratique "Manipulation de données avec SQL"
  - Créer un modèle de données
  - Importer des fichiers sources
  - Requêter les données

### Jour 2

#### TRANSFORMER DES DONNÉES : LE LANGAGE PIG

- Présentation de Pig
- Mise en pratique "Transformation de données avec Pig"
  - Charger des données semi-structurées
  - Croiser avec des données Hive
  - Sauvegarder le résultat dans HDFS

#### ECRITURE DE TRAITEMENTS AVANCÉS

- Présentation du framework Hadoop Streaming
- Présentation express du langage Python et du squelette de programme pour l'exercice
- Mise en pratique "Ecriture d'un programme de manipulation complexe"

## COMPOSITION ET ORDONNANCEMENT DE TRAITEMENTS

- Présentation d'Oozie
- Mise en pratique "Création d'un pipeline de traitement de données"