

# Hadoop, développement

Cours Pratique de 3 jours - 21h

Réf : HDD - Prix 2024 : 2 270€ HT

Ce cours vous apprendra à développer des applications vous permettant de traiter des données distribuées en mode batch. Vous collecterez, stockerez et traiterez des données de formats hétérogènes avec Apache Hadoop afin de mettre en place des chaînes de traitement intégrées à votre système d'information.

## OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

Comprendre l'écosystème Hadoop  
Cloudera/Hortonworks

Présenter les principes du Framework Hadoop

Mettre en œuvre des tâches Hadoop pour  
extraire des éléments pertinents d'ensembles de  
données volumineux

Développer des algorithmes parallèles efficaces  
avec MapReduce

Charger des données non structurées des  
systèmes HDFS et HBase

## MÉTHODES PÉDAGOGIQUES

Cours 30%, travaux pratiques 70%.

Développement d'applications pour le  
Big Data.

## LE PROGRAMME

dernière mise à jour : 11/2022

### 1) Le Big Data

- Définition du périmètre du Big Data.
- Le rôle du projet Hadoop.
- Les concepts de base des projets Big Data.
- Présentation du Cloud Computing.
- Différence entre Cloud Computing privé et public.
- Les architectures Big Data à base de projet Hadoop.
- L'écosystème Hadoop Cloudera/Hortonworks.

*Démonstration* : Usage d'Hadoop.

### 2) Collecte de données et application de Map Reduce

- Analyse des flux de données dans l'entreprise.
- Données structurées et non-structurées.
- Les principes de l'analyse sémantique des données d'entreprise.
- Graphe des tâches à base de MapReduce.
- La granularité de cohérence des données.
- Transfert de données depuis un système de persistance dans Hadoop.
- Transfert de données d'un Cloud dans Hadoop.

*Travaux pratiques* : Gérer la collecte d'informations clientèles par Map Reduce. Configuration de l'implémentation YARN. Développement d'une tâche basée sur Map Reduce.

### 3) Le stockage des données avec HBase

- Plusieurs types de base de données XML.

## PARTICIPANTS

Développeurs, chefs de projets,  
data-scientists, architectes

## PRÉREQUIS

Avoir la connaissance d'un langage  
de programmation objet comme  
Java et du scripting.

## COMPÉTENCES DU FORMATEUR

Les experts qui animent la  
formation sont des spécialistes des  
matières abordées. Ils ont été  
validés par nos équipes  
pédagogiques tant sur le plan des  
connaissances métiers que sur  
celui de la pédagogie, et ce pour  
chaque cours qu'ils enseignent. Ils  
ont au minimum cinq à dix années  
d'expérience dans leur domaine et  
occupent ou ont occupé des postes  
à responsabilité en entreprise.

## MODALITÉS D'ÉVALUATION

Le formateur évalue la progression  
pédagogique du participant tout au  
long de la formation au moyen de  
QCM, mises en situation, travaux  
pratiques...

Le participant complète également  
un test de positionnement en amont  
et en aval pour valider les  
compétences acquises.

## MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Les moyens pédagogiques et les  
méthodes d'enseignement utilisés  
sont principalement : aides  
audiovisuelles, documentation et  
support de cours, exercices  
pratiques d'application et corrigés  
des exercices pour les stages  
pratiques, études de cas ou  
présentation de cas réels pour les  
séminaires de formation.
- À l'issue de chaque stage ou  
séminaire, ORSYS fournit aux  
participants un questionnaire  
d'évaluation du cours qui est  
ensuite analysé par nos équipes  
pédagogiques.
- Une feuille d'émargement par  
demi-journée de présence est  
fournie en fin de formation ainsi  
qu'une attestation de fin de  
formation si le stagiaire a bien  
assisté à la totalité de la session.

## MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24  
heures avant le début de la  
formation.

## ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique  
d'accessibilité ? Contactez Mme  
FOSSE, référente handicap, à  
l'adresse suivante psh-  
accueil@orsys.fr pour étudier au  
mieux votre demande et sa  
faisabilité.

- Patterns d'usages et application au Cloud.
- Application de Hadoop database au sein d'un workflow.
- Utilisation des projets Hive/Pig.
- Utilisation du projet HCatalog.
- L'API Java HBase.

*Travaux pratiques : Gérer les modifications d'un catalogue de données.*

#### 4) Le stockage des données sur HDFS

- Patterns d'usages et application au Cloud.
- Architecture et installation d'un système HDFS, journal, NameNode, DataNode.
- Opérations, commandes et gestion des commandes.
- L'API HDFS Java.
- Analyse de données avec Apache Pig.
- Le langage Pig Latin. Utiliser Apache Pig avec Java.
- Requêtage avec Apache Hive.
- Réplication de données. Partage de données sur une architecture HDFS.

*Travaux pratiques : Administrer un référentiel client partagé sur Hadoop. Utilisation de la console de visualisation.*

## LES DATES

---

CLASSE À DISTANCE  
2024 : 01 juil., 25 sept., 27 nov.

PARIS  
2024 : 19 juin, 18 sept., 20 nov.