

# Big Data, découvrir la plateforme Hadoop et son écosystème

## tutorat en option

Formation en ligne - 1h15

Réf : 4TX - Prix 2024 : 95CHF HT

Ce cours en ligne a pour objectif de vous présenter le big data en général et plus spécifiquement la plateforme Hadoop et ses concepts clés, afin de comprendre son fonctionnement et ses possibilités. Il s'adresse à tous les informaticiens qui veulent avoir un premier aperçu d'Hadoop. La pédagogie s'appuie sur un auto-apprentissage séquencé par actions de l'utilisateur sur l'environnement à maîtriser. Une option de tutorat vient renforcer l'apprentissage.

### OBJECTIFS PÉDAGOGIQUES

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

Connaître les notions de NoSQL et de Big Data

Installer un environnement Hadoop

Utiliser les distributions Hortonworks, MapR et Cloudera

Réaliser des algorithmes de calcul distribué avec MapReduce

Comprendre les concepts principaux autour du stockage et du traitement des données dans Hadoop

Prendre en main l'écosystème basé sur Hadoop (HBase, les outils de traitement Pig, Hive et Spark)

### PÉDAGOGIE ET PRATIQUES

Une évaluation tout au long de la formation grâce à une pédagogie active mixant théorie, exercice, partage de pratique et gamification. Un service technique est dédié au support de l'apprenant. La formation est diffusée au format SCORM (1.2) et accessible en illimité pendant 1 an.

### ACTIVITÉS DIGITALES

Démonstrations, cours enregistrés, partages de bonnes pratiques, quiz, fiches de synthèse.

## LE PROGRAMME

dernière mise à jour : 10/2023

### 1) Origines d'Hadoop

- Qu'est-ce que le NoSQL ?
- Définition du big data.
- Histoire d'Hadoop.

### 2) Installation d'un environnement Hadoop de base

- Écosystème complexe.
- Distributions Hadoop.
- Introduction à CDH : Cloudera Hadoop.
- Démarrage d'une QuickStart VM.

### 3) Stockage de fichiers : HDFS (Hadoop Distributed File System)

- Présentation de l'HDFS.
- Manipulation de fichiers en ligne de commande.

### 4) Paradigme MapReduce

- Principe général.

### PARTICIPANTS

Informaticiens qui veulent avoir un premier aperçu d'Hadoop.

### PRÉREQUIS

Aucune connaissance particulière.

### COMPÉTENCES DU FORMATEUR

Les experts qui ont conçu la formation et qui accompagnent les apprenants dans le cadre d'un tutorat sont des spécialistes des sujets traités. 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. 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

La progression de l'apprenant est évaluée tout au long de sa formation au moyen de QCM, d'exercices pratiques, de tests ou d'échanges pédagogiques. Sa satisfaction est aussi évaluée à l'issue de sa formation grâce à un questionnaire.

### MOYENS PÉDAGOGIQUES ET TECHNIQUES

Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : documentation et support de cours, exercices pratiques d'application et corrigés des exercices, études de cas ou présentation de cas réels. ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques. Une attestation de fin de formation est fournie si l'apprenant a bien suivi la totalité de la formation.

### 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.

- Fonction map.
- Fonction reduce.

### 5) Développement d'un premier MapReduce

- Mapper.
- Reducer.
- Lanceur.
- Lancement de votre premier MapReduce.
- Combineur.

### 6) Écosystème Hadoop

- Hbase.
- Hive.
- Pig.
- Zookeeper.
- Sqoop.
- Oozie.
- Flume.
- Kafka.
- Spark.