

# Chaîne e-learning Big Data

Digital Learning - 06h

Réf : 8TA - Prix 2024 : 190€ HT

Dans un monde où les données sont omniprésentes, le Big Data représente bien plus qu'une simple tendance technologique. C'est une révolution qui transforme la façon dont les entreprises opèrent, prennent des décisions et interagissent avec leurs clients. Notre chaîne spécialisée vous présentera les concepts clés du Big Data et vous apprendra à manipuler de grands volumes de données au travers de différentes méthodes et outils.

## OBJECTIFS PÉDAGOGIQUES

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

Connaître l'histoire du big data.

Comprendre les concepts fondamentaux du big data.

Comprendre les usages du big data.

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).

Définir les notions de NoSQL et de big data.

Installer Hadoop et étudier les différents outils au cœur de la plateforme.

Déployer un cluster sur des machines virtuelles dans le cloud grâce à la distribution Cloudera Data Platform.

Stocker des données dans Hadoop.

Appréhender le paradigme de traitement distribué des données avec MapReduce.

Développer des traitements avec Spark ou en SQL avec Hive.

Connaître le framework distribué Apache Spark.

Installer et prendre en main le framework Spark.

Manipuler Spark pour enrichir des données et faire du Machine Learning.

Utiliser Spark dans une situation d'analyse et de traitement de données.

## PÉDAGOGIE ET PRATIQUES

De nombreux contenus réalisés par des formateurs suivant une démarche pédagogique rigoureuse. Durant chaque cours, des cas opérationnels sont commentés par des experts pour aider les apprenants à mettre en pratique ce qu'ils viennent d'apprendre. Afin de favoriser l'ancrage mémoriel, chaque contenu est découpé en séquences courtes de 3 à 10 minutes. Ce découpage permet un apprentissage dynamique et en toute autonomie pour chaque apprenant.

## ACTIVITÉS DIGITALES

La structure IT : Cours enregistrés, vidéos d'expert et partages de bonnes pratiques.

## PARTICIPANTS

Décideurs et développeurs intéressés par le big data.

## 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.

# LE PROGRAMME

---

dernière mise à jour : 06/2024

## 1) Big data, comprendre sa création, ses concepts et découvrir des exemples

- Le big data : qu'est-ce que c'est ?
- Les données, l'or noir du XXI<sup>e</sup> siècle.
- L'analyse, le monde merveilleux de la data science.
- L'exploitation des données à travers des technologies novatrices.
- Big data : utilisation, dérives et risques.

## 2) Big Data, découvrir la plateforme Hadoop et son écosystème

- Origines d'Hadoop.
- Installation d'un environnement Hadoop de base.
- Stockage de fichiers : HDFS (Hadoop Distributed File System).
- Paradigme MapReduce.
- Paradigme MapReduce.
- Écosystème Hadoop.

## 3) Hadoop, stocker et traiter des données pour le big data

- Présentation des origines de Hadoop.
- Installation d'un environnement Hadoop avec Cloudera.
- Gestion du stockage de données.
- Réalisation de traitements sur les données.

## 4) Spark, connaître le Framework distribué pour le Big Data et le Machine Learning

- Framework Spark et ses rouages.
- Spark pour l'enrichissement de données.
- Spark pour l'apprentissage automatique.