

# Formation Big Data, Les clefs pour être Data Scientist

## Présentation

Cette formation offre une introduction complète au Big Data et à la Data Science, en couvrant les principes fondamentaux du Big Data, l'écosystème Hadoop, et les différences entre bases de données relationnelles et non relationnelles. Elle inclut également des modules sur la manipulation de données avec Pandas, NumPy et Spark, ainsi que sur l'apprentissage automatique avec Spark MLlib. Enfin, elle aborde les freins à l'usage de l'IA et propose des projets pratiques pour appliquer les connaissances acquises.

Durée : 35,00 heures (5 jours)

Tarif INTRA : Nous consulter

## Objectifs de la formation

- Comprendre les concepts fondamentaux du Big Data et de la data science.
- Apprendre à utiliser Python et Scala pour l'analyse de données.
- Découvrir les frameworks de traitement de données comme Apache Spark.
- Savoir appliquer les techniques de machine learning dans un environnement Big Data.

## Prérequis

- Connaissances de base en informatique : notions fondamentales sur les systèmes d'exploitation.
- Compétences en programmation : connaissance de base en Python.
- Bases de données : familiarité avec les bases de données relationnelles (MySQL) et non relationnelles (Cassandra, MongoDB).
- Notions de base en Machine Learning.
- Big Data : connaissances élémentaires sur les systèmes distribués et les architectures Big Data (Hadoop, HDFS, YARN, MapReduce).
- Data Science : compétences de base pour manipuler et analyser des données.

## Public

- Ingénieurs en informatique



- Data analysts
- Développeurs intéressés par le Big Data
- Toute personne souhaitant devenir Data Scientist

## Programme de la formation

### JOUR 1 : Introduction au Big Data et à la Data Science

#### Module 1 : Introduction au Big Data

- Principes fondamentaux du Big Data.
- Écosystème Hadoop et ses composants (HDFS, YARN, MapReduce).
- Différences entre bases de données relationnelles et non relationnelles (MySQL vs Cassandra, MongoDB).

#### Module 2 : La Data Science et la Transformation de Données

- Introduction à la Data Science.
- Manipulation de données avec Pandas et NumPy.
- Introduction à Spark pour le Big Data.
- Travaux pratiques : manipulation de données simples avec Scala.

### JOUR 2 : Frameworks de Traitement de Données avec Spark

#### Module 3 : Cadres de traitement de données

- Présentation d'Apache Spark.
- Traitement de données massives avec Spark en Python et Scala.
- Notions de RDD, DataFrame et SparkSQL.

#### Travaux pratiques :

- Manipulation de données avec Spark en Python et Scala.
- Création et utilisation de DataFrames et RDDs avec Spark.

#### Objectifs : apprendre à traiter des données volumineuses avec Spark.

### **JOUR 3 : Machine Learning avec Spark (1)**

#### **Module 4 : Apprentissage automatique (1)**

- Concepts de base du Machine Learning.
- Préparation des données pour le Machine Learning.
- Implémentation de modèles de Machine Learning avec Spark MLlib.

#### **Travaux pratiques :**

- Mise en œuvre de modèles de Machine Learning avec Spark MLlib.
- Création et évaluation de modèles avec des jeux de données massifs.

### **JOUR 4 : Machine Learning avec Spark (2) et Freins à l'Usage de l'IA**

#### **Module 5 : Apprentissage automatique (2)**

- Types avancés de Machine Learning.
- Exercices pratiques : apprentissage non supervisé et clustering.

#### **Les freins à l'usage de l'IA et des données :**

- Identification des barrières à l'adoption de l'IA et des données.
- Analyse des enjeux éthiques de l'IA et du RGPD.
- Étude des obstacles techniques, culturels et organisationnels.

#### **Travaux pratiques : discussion en groupe sur les obstacles à l'adoption de l'IA dans différents contextes (entreprises, gouvernements, etc.).**

### **JOUR 5 : Projets Pratiques et Études de Cas**

#### **Module 6 : Projets pratiques et études de cas**

- Analysez un jeu de données massif avec Spark.

#### **Travaux pratiques :**

- Projets en groupe pour analyser un jeu de données massif, appliquer des modèles de Machine Learning et visualiser des rapports d'analyse.
- Présentation des projets en groupe avec retour d'expérience.

## Organisation

### Formateur

Les formateurs de Docaposte Institute sont des experts de leur domaine, disposant d'une expérience terrain qu'ils enrichissent continuellement. Leurs connaissances techniques et pédagogiques sont rigoureusement validées en amont par nos référents internes.

Riches de leur expérience sur le sujet, ils sauront accompagner vos collaborateurs dans leur montée en compétence.

### Moyens pédagogiques et techniques

- Apports des connaissances communes.
- Mises en situation sur le thème de la formation et des cas concrets.
- Méthodologie d'apprentissage attractive, interactive et participative.
- Equilibre théorie / pratique : 60 % / 40 %.
- Supports de cours fournis au format papier et/ou numérique.
- Ressources documentaires en ligne et références mises à disposition par le formateur.
- Pour les formations en présentiel dans les locaux mis à disposition, les apprenants sont accueillis dans une salle de cours équipée d'un réseau Wi-Fi, d'un tableau blanc ou paperboard. Un ordinateur avec les logiciels appropriés est mis à disposition (le cas échéant).

### Dispositif de suivi de l'exécution et de l'évaluation des résultats de la formation

En amont de la formation

- Recueil des besoins des apprenants afin de disposer des informations essentielles au bon déroulé de la formation (profil, niveau, attentes particulières...).
- Auto-positionnement des apprenants afin de mesurer le niveau de départ.

Tout au long de la formation

- Évaluation continue des acquis avec des questions orales, des exercices, des QCM, des cas pratiques ou mises en situation...

A la fin de la formation

- Auto-positionnement des apprenants afin de mesurer l'acquisition des compétences.
- Evaluation par le formateur des compétences acquises par les apprenants.
- Questionnaire de satisfaction à chaud afin de recueillir la satisfaction des apprenants à l'issue de la formation.
- Questionnaire de satisfaction à froid afin d'évaluer les apports ancrés de la formation et leurs mises en application au quotidien.

### **Accessibilité**

Nos formations peuvent être adaptées à certaines conditions de handicap. Nous contacter pour toute information et demande spécifique.