

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

- Alternance d'apports théoriques et de mises en pratique contextualisées.
- Progression pédagogique structurée, de la compréhension à la mise en œuvre opérationnelle.
- Pédagogie immersive favorisant une mise en application immédiate des compétences.
- Formation interactive et participative s'appuyant sur des ateliers pratiques, études de cas et échanges collectifs (+ 60 % du temps de formation.)
- Séquences pédagogiques rythmées et évaluations formatives régulières pour renforcer l'ancrage des apprentissages.
- Supports de cours, ressources documentaires et références mis à disposition par le formateur

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

En amont de la formation *(évaluation en ligne dans l'extranet apprenant)*

- Recueil des besoins et des attentes des participants.
- Auto-positionnement permettant d'évaluer le niveau initial des apprenants.

Tout au long de la formation

- Évaluations continues réalisées par le formateur à travers des questions orales, exercices, QCM, études de cas et mises en situation pratiques.

À l'issue de la formation (*évaluations en ligne dans l'extranet apprenant*)

- Auto-positionnement afin de mesurer l'acquisition des compétences visées.
- Validation de l'acquisition des compétences par le formateur
- Questionnaire de satisfaction « à chaud » à l'issue de la formation
- Questionnaire de satisfaction "à froid", 45 jours après la formation afin d'évaluer l'ancrage des acquis et leur mise en application en situation professionnelle.

Accessibilité

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