

# ANALYSE DE DONNÉES ET MACHINE LEARNING AVEC SPARK

IIA20

3 jours (21h) 2 280,00 €<sup>HT</sup>

## Objectifs

Utiliser Spark SQL et Spark ML pour réaliser des analyses traditionnelles des données, de l'analyse statistique et appliquer des algorithmes de Machine Learning au Big Data

## Profil Stagiaire(s)

analyste de données, Data Scientist et développeur ayant une appétence pour le Machine Learning

## Pré-requis

La connaissance de Python, R, Scala ou SQL est préférable

La connaissance de java et spark n'est pas requise

Cette formation se focalise sur l'utilisation de Spark dans le cadre de l'Analyse de Données et du Machine Learning et n'est pas une formation dédiée au fonctionnement interne de Spark d'un point de vue développeur (pour une formation 100% dédiée au développement sur Spark et donc moins de focus sur ses APIs d'analyse de données voir la formation certifiante CCA IBD72)

## Profil animateur(s)

Consultant-formateur spécialiste Data Science, IA et Big Data



## Contactez-nous

Conseiller formation  
09 88 66 10 00  
inscriptions@demofr

## Nos sessions

6 - 8 sept. 2021 : Paris / A Distance

29 nov. - 1 déc. 2021 : Paris / A Distance

## PROGRAMME

### Analyse de données structurées avec Spark DataFrames

#### Introduction à Spark sur Hadoop

RDD (Resilient Distributed Datasets)

Interagir avec HDFS

Soumettre une requête Python

#### Hands-On : DataFrame

L'API de base

Chargement de données structurées

Manipulations basiques

Types de données

Opérations et statistiques

#### Hands-On : Manipulations avancées des DataFrames

Aggregations

Jointures

Window aggregations

Chargement et sauvegarde pour différents formats

SQL sur Spark

#### Hands-On : Bonnes pratiques et performances

Format des données et compression

Fonctionnement de Catalyst

Partition discovery

Python, R ou Scala pour Spark, que choisir ?

### Machine Learning avec Spark ML

# ANALYSE DE DONNÉES ET MACHINE LEARNING AVEC SPARK

IIA20

3 jours (21h) 2 280,00 €<sup>HT</sup>

## Rappels sur la Machine Learning Spark ML – MLib

Structure de l'API & Concepts clés

Données d'entrée

Exemples supervisé et non supervisé

Limites de l'API

## Hands-On : Spark ML

Structure de l'API & Concepts clés

Transformer / Estimator

Paramètres

Évaluation de modèles (classification, régression)

## Hands-On : Feature Engineering

Indexing

Encoding

Text Processing

Scaling

## Hands-On : Tuning de paramètres

Train-Validation Split

Grid-Search

Cross Validation

## Hands-On : Pipeline

Construction de Pipeline de Machine Learning

## Hands-On : Autres applications

Réduction de Dimension

Clustering

Systèmes de Recommandation

# ANALYSE DE DONNÉES ET MACHINE LEARNING AVEC SPARK

IIA20

3 jours (21h) 2 280,00 €<sup>HT</sup>

## Méthode pédagogique

Au travers de discussions et d'exercices interactifs, les participants travailleront sur les différentes fonctionnalités de Spark

1/3 de pratique et 2/3 de théorie

La partie pratique s'articulera autour d'un projet fil rouge tout au long de la formation. Dans un premier temps, l'accent sera mis sur la manipulation avancée de données sur les tables à disposition. Sur la partie Machine Learning, les participants travailleront sur l'élaboration d'un premier modèle supervisé simple et son évaluation, pour ensuite l'améliorer en le complexifiant avec les différentes notions vues pendant la formation

Pour les formations "A distance", elles sont réalisées avec un outil de visioconférence de type Teams ou Zoom selon les cas, permettant au formateur d'adapter sa pédagogie.

Retrouvez sur notre site internet toutes les précisions sur les sessions à distance ou les classes virtuelles.

## Moyens pédagogiques et techniques de mise en œuvre

Nos formateurs DEMOS sont recrutés conformément aux spécifications mentionnées pour chaque formation. Ce sont des professionnels en activité et/ou des experts dans leur domaine.

Ils utilisent des méthodes et outils appropriés aux formations qu'ils dispensent et adaptent leur pédagogie au public accueilli.

Par ailleurs, nos centres de formation DEMOS sont tous équipés :

- Salles de formation lumineuses, spacieuses, design, ergonomiques, mobiles et équipées d'écrans plasma, de PC portables/fixes si nécessaire à la formation suivie.
- Accès au wifi haut débit sur l'ensemble des lieux
- Espace de co-learning et webcorners
- Cafés, collations et rafraîchissement sont offerts

Et tous sont accessibles aux personnes à mobilité réduite

## Le dispositif de suivi et d'évaluation

Pour assurer un suivi individuel, Demos a mis en place 2 types d'évaluation :

- Des évaluations des acquis en cours et en fin de formation

Elles peuvent être faites de différente manière selon le contenu de la formation suivie :

Quiz, exercice pratique, étude de cas, jeu de rôles, mise en situation, soutenance devant un jury pour les formations à finalité certifiante.

- Une évaluation de la satisfaction de chaque stagiaire est réalisée en ligne. Cette évaluation est complétée par l'appréciation du formateur à l'issue de chaque session.

## En complément

**Avant cette formation vous pouvez suivre :**

- Programmation en Python
- Scala : Fondamentaux du Langage Appliqués au Big Data