

# Spark pour la Data Science



STG-E18-004

**Durée** : 3 jours

**Niveau** : Intermédiaire

**Prix** : 1 770 €

## OBJECTIFS

Spark est un framework de calcul distribué indispensable pour effectuer des analyses complexes à grande échelle. Il peut être utilisé depuis des langages comme Scala, Java, Python ou R.

Les objectifs sont de connaître l'architecture de Spark et les principes de traitement afin de réaliser des traitements distribués. Le stagiaire apprendra à réaliser des pipelines d'apprentissage et de prédiction (prédiction ?) avec des techniques de modélisation distribuées. Il apprendra également le traitement des données en flux et les modes d'industrialisation des traitements.

## SESSION(S) PLANNIFIEE(S)

- 25, 26 et 27 juin 2018 - (*Bordeaux*)

[S'inscrire sur le site Internet de la formation](#) ou par téléphone : 07 64 14 76 12

Formation **disponible toute l'année** à la demande.

## PREREQUIS

- Connaissances de base de la programmation informatique
- Connaissances pratiques en SQL ainsi qu'en Python ou R

## PROGRAMME DETAILLE

### Structures de données

- Resilient Data Sets
- DataFrames et RDD

- Gestion de la distribution des données
- Accumulateurs et variables diffusées

### **Apache Spark**

- Principe de fonctionnement d'Apache Spark
- Historique et évolution
- Architecture et distribution
- Modes de déploiement et modules de Spark
- Créer son environnement Spark

### **Manipuler et modéliser des données**

- Accès aux données avec Spark SQL
- Transtypage des données
- Passage RDD et Dataframes
- Modéliser des données avec MLLib
- Apprentissage supervisé ou non supervisé
- Recodage et création de variables
- Technique d'évaluation des modèles
- Mise en place des pipelines de traitement

### **Spark en production**

- Production
- Industrialisation de traitements

### **Spark et R**

- Exemple de codes sources avec Sparkr et Sparklyr

## **LOGICIELS UTILISES**

Spark et Python (ou R)

Compatible PC & MAC

Logiciels non payants

## **MODALITES PEDAGOGIQUES**

Afin de permettre aux étudiants d'acquérir des compétences opérationnelles en entreprise, cette formation associe des connaissances notionnelles (exposés théoriques) avec des savoir-faire pratiques (applications, cas concrets, travaux pratiques).

**Vous souhaitez obtenir plus de détails sur cette formation ? Rendez-vous sur notre site Internet, [sur la page dédiée à la formation](#).**

**Pour toute question, n'hésitez pas à contacter [conseil@datauniversity.fr](mailto:conseil@datauniversity.fr) et à consulter notre site Internet <https://datauniversity.fr>**