

# Formation Spark, Les fondamentaux du Big Data

## Présentation

Cette formation vous permettra de maîtriser les concepts fondamentaux de Spark.

Durée : 14,00 heures (2 jours)

Tarif INTRA : Nous consulter

## Objectifs de la formation

- Comprendre les fondamentaux de Spark
- Découvrir les différents modules de Spark
- Manipuler des données avec Spark SQL

## Prérequis

- Bonnes connaissances du langage Python

## Public

- Développeurs
- Architectes

## Programme de la formation

### Présentation d'Apache Spark

- Introduction et historique
- Les langages de programmation supportés par Spark (Scala, Python et Java)
- Comparaison avec l'environnement Apache Hadoop
- Les modules de Spark (Spark SQL, Spark Streaming, MLlib, GraphX )

### Resilient Distributed Dataset (RDD)

- C'est quoi un RDD ?
- Créer, manipuler et réutiliser des RDD
- Accumulateurs et variables broadcastées
- Utiliser des partitions



**Manipuler des données structurées avec Spark SQL**

- SQL, DataFrames et Datasets
- Les différents types de sources de données
- Interopérabilité avec les RDD
- Performance de Spark SQL
- JDBC/ODBC server et Spark SQL CLI

**Spark sur un cluster**

- Les différents types d'architectures : Standalone, Apache Mesos ou Hadoop YARN
- Configurer un cluster en mode Standalone
- Packager une application avec ses dépendances
- Déployer des applications avec Spark-submit
- Dimensionner un cluster

## 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.

**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.