

Traitement de données distribuées avec Spark en Scala

Description

Cette formation permet aux profils techniques de concevoir, développer et optimiser des traitements de données distribuées avec Apache Spark en utilisant le langage Scala. Elle met l'accent sur la compréhension de l'architecture Spark, les mécanismes d'exécution distribuée et les bonnes pratiques de data engineering.

Durée

3.00 jours (21.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data engineers
- Développeurs
- Équipes SI

Prérequis :

- Bases en programmation (Scala ou Java) et notions de data

Objectifs pédagogiques

- Comprendre l'architecture et le fonctionnement interne de Spark
- Développer des traitements distribués en Scala
- Utiliser efficacement RDD, DataFrames et Spark SQL
- Optimiser les performances et la consommation de ressources
- Analyser et corriger des problématiques en production

Programme détaillé

- Jour 1 – Architecture Spark et fondamentaux Scala
 - Présentation de l'écosystème Spark
 - Architecture distribuée : driver, executors, cluster manager
 - Rappels Scala utiles pour Spark
 - Premiers traitements distribués
- Jour 2 – APIs Spark et traitements distribués
 - RDD : transformations et actions
 - DataFrames et Spark SQL
 - Lecture et écriture de données à grande échelle
 - Gestion du partitionnement
- Jour 3 – Optimisation et mise en production
 - Plans d'exécution et Catalyst Optimizer
 - Gestion de la mémoire et des performances

Stat4decision

37-39 avenue Ledru Rollin Paris 75012

Tel. +33172254082 | E-mail : info@stat4decision.com | www.stat4decision.com

Numéro SIRET: 81048985600015 | Numéro de déclaration d'activité: 11755352275 (auprès du préfet de région de: Ile de France)

Organisme de formation certifié Qualiopi pour ses actions de formation

- Debugging et analyse des jobs Spark
- Bonnes pratiques de mise en production

Organisation de la formation

Moyens pédagogiques et techniques

- Supports pédagogiques numériques
- Environnements Spark et Scala
- Études de cas issues de contextes réels

Dispositif de suivi de l'exécution et d'évaluation des objectifs de la formation

- Évaluation continue via exercices pratiques

Stat4decision

37-39 avenue Ledru Rollin Paris 75012

Tel. +33172254082 | E-mail : info@stat4decision.com | www.stat4decision.com

Numéro SIRET: 81048985600015 | Numéro de déclaration d'activité: 11755352275 (auprès du préfet de région de: Ile de France)

Organisme de formation certifié Qualiopi pour ses actions de formation