SPARK, DÉVELOPPER POUR LE BIG DATA

Formation proposée par BIZOÏ RAZVAN

Spark

Vous développerez des applications en Python, Scala ou R en vue de traiter en temps réel des données issues du Big Data. Vous collecterez, stockerez et traiterez avec Spark des données de formats hétérogènes afin de mettre en place des chaînes de traitement intégrées à votre Système d'Information.

◷DURÉE
5 Jour(s)
▣MODALITÉS
Présentiel / Distanciel
♙PARTICIPANTS
Jusqu'à 10 participants
★LANGUE
Français
LIEU
France entière
♙CERTIFICATION
Attestation de formation
▦MISE À JOUR
02/02/2026

VOTRE FORMATEUR

BR **BIZOÏ RAZVAN**
Formateur professionnel

TARIFS INDICATIFS

Inter-entreprise	**Pas de dates planifiées**
Intra-entreprise	**7950 € HT / groupe**
Individuelle	**Sur devis**
Durée	**5 Jour(s)**

Tarifs indicatifs, sous réserve d'analyse du besoin et de devis.

◎Objectifs pédagogiques

- ✓Maîtriser les concepts fondamentaux de Spark.
- ✓Développer des applications avec Spark Streaming.
- ✓Faire de la programmation parallèle avec Spark sur un cluster.
- ✓Exploiter des données avec Spark SQL.
- ✓Avoir une première approche du Machine Learning.

⌘Compétences visées

Spark

♙Public cible et prérequis

Public cible

- Développeurs, architectes

Prérequis

- Bonnes connaissances du langage Scala, Python ou R.

Programme détaillé

Introduction
- Historique du Framework.
- Comparaison avec l'environnement Apache Hadoop.
- Les différents modules de Spark.
- Langages supportés.

Travaux pratiques
Installation et configuration de Spark. Exécution d'un premier exemple avec le comptage de mots.

Programmer avec les Resilient Distributed Dataset (RDD)
- Présentation des RDD.
- Créer, manipuler et réutiliser des RDD.
- DAG Directed Acyclic Graph.
- Accumulateurs et variables broadcastées.
- Utiliser des partitions.

Travaux pratiques
Manipulation de différents Datasets à l'aide de RDD et utilisation de l'API fournie par Spark.

Manipuler des données structurées avec Spark SQL
- Spark Context vs Spark Session.
- SQL, DataFrames et Datasets.
- Les différents types de sources de données.
- Interopérabilité avec les RDD.
- Utilisation de SQL avec des Dataframes.
- Performance de Spark SQL.
- JDBC/ODBC server et Spark SQL CLI.
- L'analyse du DAG via Spark-UI

Travaux pratiques
Manipulation de Dataframes via des requêtes SQL.

Spark sur un cluster
- Les différents types d'architecture : 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 .

Travaux pratiques
Mise en place d'un cluster Spark.

Manipuler des graphes avec GraphX
- Présentation de GraphX.
- Les différentes opérations.
- Créer des graphes.
- Vertex and Edge RDD.
- Présentation de différents algorithmes.

Travaux pratiques
Manipulation de l'API GraphX à travers différents exemples.

L'architecture Kafka
- L'aperçu de Kafka Producers, Brokers, Consumers.
- Les fichiers journaux de Kafka.
- Les schémas Avro.
- Utilisation de ZooKeeper.

Travaux pratiques
Étude de la configuration de Kafka dans l'architecture KAPPA.

Spark Structured Streaming
- Structured Streaming API.
- Streaming Context.
- Static et Dynamic Datasets.
- Fenêtrage et Agrégation.
- Checkpointing & Watermarking.

- Fiabilité et tolérance aux pannes.
- Intégration avec Kafka.
Travaux pratiques
Mise en œuvre de l’analyse en temps-réel d'un fichier de log.

Introduction au Machine Learning
- Introduction au Machine Learning.
- Les différentes classes d'algorithmes.
- Présentation de MLlib.
- Implémentations des différents algorithmes dans MLlib.
- Clustering : KNN, K-mean
- Régression : Arbre de régression
- Classification : Random Forest, SVM, AUC, Courbe ROC
Travaux pratiques
Mise en œuvre des analyses sur plusieurs jeux de données.

▣Méthodes pédagogiques & évaluation

Méthodes pédagogiques

- ✓ Apports théoriques et retours d'expérience
- ✓ Exercices pratiques et mises en situation
- ✓ Travaux de groupe et échanges entre pairs
- ✓ Études de cas concrets

Modalités d'évaluation

- ✓ Quiz de validation des acquis
- ✓ Mises en situation évaluées
- ✓ Projet fil rouge
- ✓ Feuille de route post-formation

♿Accessibilité

Une formation accessible à tous

Ce programme peut être adapté pour les personnes en situation de handicap. Contactez-nous pour étudier vos besoins spécifiques.