

— Public	Développeurs, architectes
— Durée	5 jours - 35 heures
— Pré-requis	Bonnes connaissances du langage Scala, Python ou R.
— Objectifs	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.
— Méthodes pédagogiques	Pour bien préparer la formation, le stagiaire remplit une évaluation de positionnement et fixe ses objectifs à travers un questionnaire. La formation est délivrée en présentiel ou distanciel (e-learning, classe virtuelle, présentiel et à distance). Le formateur alterne entre méthodes démonstratives, interrogatives et actives (via des travaux pratiques et/ou des mises en situation). La validation des acquis peut se faire via des études de cas, des quiz et/ou une certification. Cette formation est animée par un consultant-formateur dont les compétences techniques, professionnelles et pédagogiques ont été validées par des diplômes et/ou testées et approuvées par l'éditeur et/ou par Audit Conseil Formation.
— Moyens techniques	1 poste de travail complet par personne De nombreux exercices d'application Mise en place d'ateliers pratiques Remise d'un support de cours Passage de certification(s) dans le cadre du CPF Remise d'une attestation de stage
— Modalité d'évaluation des acquis	Evaluation des besoins et objectifs en pré et post formation Evaluation technique des connaissances en pré et post formation Evaluation générale du stage
— Délai d'accès	L'inscription à cette formation est possible jusqu'à 5 jours ouvrés avant le début de la session
— Accessibilité handicapés	Au centre d'affaires ELITE partenaire d'ACF à 20 m. Guide d'accessibilité à l'accueil.

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.

NOUS CONTACTER

Siège social

16, ALLÉE FRANÇOIS VILLON
38130 ÉCHIROLLES

Téléphone

04 76 23 20 50 - 06 81 73 19 35

Suivez-nous sur les réseaux sociaux, rejoignez la communauté !



ACF Audit Conseil Formation



@ACF_Formation

Dernière mise à jour : 15/04/2026

PROFIL Formateur : Les formateurs sont recrutés selon plusieurs critères :
Expérience, pédagogie, dynamisme et prévoyance.