BIG DATA LES TECHNIQUES D'ANALYSE ET DE VISUALISATION

Public	Data Engineers, Data Scientists, Data Analysts, ingénieurs DevOps, développeurs, architectes Big Data, chefs de projets.
Durée	5 jours - 35 heures
Pré-requis	Avoir des connaissances en langage Python pour l'analyse de données via Spark. Maîtriser le langage de requêtage SQL. Avoir des connaissances en ETL et en BI est un plus.
Objectifs	Charger et analyser des données de diverses sources avec Hadoop, NiFi, Kafka et Spark Restituer ces données avec Kibana.
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
Planning	Du 16/06/2025 au 20/06/2025 Du 15/09/2025 au 19/09/2025 Du 17/11/2025 au 21/11/2025
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.

VUE D'ENSEMBLE DU BIG DATA

- Introduction au Big Data : de quoi s'agit-il ?
- Perspectives offertes par le Big Data
- Les acteurs du Big Data
- Exemples pratiques
- Démystification du Big Data
- Big Data et Cloud SaaS, PaaS et FaaS
- Les différents métiers du Big Data
- Data Lakes vs Data Warehouses vs Lakehouses
- TCO (Total Cost of Ownership)
- Opportunity Cost (coût d'opportunité)
- ROI (retour sur investissement)
- Travaux pratiques sur Hadoop
- Aspects réglementaires (RGPD) et éthiques

ARCHITECTURE HADOOP

- Les outils de stockage
- Les exigences de stockage
- Le théorème de CAP
- Le NoSQL
- HDFS
- MapReduce
- Tez
- Les outils d'ingestion et d'analyse de données (Spark, Pig, Hive, Impala)
- La mise en qualité
- Le stockage de données

FONDAMENTAUX DU DATA ENGINEERING

- Définition du Data Engineering
- Cycle de vie et workflow du Data Engineering
- Technologies du Data Engineering
- Evolution du métier de Data Engineer
- Compétences, activités et responsabilités du Data Engineer : Type A vs Type B
- Le Data Engineer au sein d'une Data Team
- Parties prenantes : producteurs upstream, consommateurs downstream
- Data Maturity d'une organisation
- Data Engineering, Agilité, DevOps et DataOps
- Gouvernance et qualité des données
- Nouvelles tendances du Data Engineering: Lakehouse, orchestration, Pipelines as Code

NIFI

- Ingestion des données avec NiFi
- Différence entre données froides et données chaudes
- Présentation des outils et technologies du Big Data
- Installation et configuration de NiFi
- Vue d'ensemble de l'architecture NiFi
- Approches de développement
- Outils de développement d'applications et état d'esprit
- Extraction, transformation et chargement (ETL) des outils et de la mentalité

VISUALISATION DES DONNÉES AVEC KIBANA

- Préparer les données pour Elasticsearch
- Construire un tableau de bord Kibana
- Créer des visualisations

STREAMING DES DONNÉES AVEC KAFKA

- Comprendre comment Kafka utilise des logs
- Les "Topics"
- Les producteurs Kafka
- Les "Brokers"
- Les consommateurs Kafka
- Différences entre traitement batch et traitement en streaming

TRAITEMENT DES DONNÉES AVEC SPARK

- Introduction à Apache Spark
- Architecture de Spark
- Apache Spark MLlib
- Apache Spark Streaming
- Apache Spark SQL
- Apache Spark GraphX

NOUS CONTACTER

Siège social
16, ALLÉE FRANÇOIS VILLON
38130 ÉCHIROLLES

Centre de formation
87, RUE GÉNÉRAL MANGIN
38000 GRENOBLE

Téléphone
04 76 23 20 50 - 06 81 73 19 35

E-mail
contact@audit-conseil-formation.com

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

ACF Audit Conseil Formation @ACF_Formation ACFauditconseilformation