

Data Engineering

Data Engineering avec Kafka, Cassandra et Spark

4 jours (28h00) | ★★★★★ 4,6/5 | BIG-ENG | Évaluation qualitative de fin de stage |
Formation délivrée en présentiel ou distanciel ⁽¹⁾

Formations Informatique > Big Data > Data Engineering



À l'issue de ce stage vous serez capable de :

- Maîtriser les concepts fondamentaux de Spark, Cassandra et Kafka
- Développer une application orientée use case avec Spark, Cassandra et Kafka
- Vous immerger dans le Data Engineering
- Approfondir vos connaissances en architecture Big Data.

Niveau requis

Avoir des connaissances sur Shell, Python et/ou Scala.

Public concerné

Développeurs, architectes, Data Analysts et/ou Data Engineers.

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 M2i Formation
- Bénéficie d'un suivi de son exécution par une feuille de présence élargée par demi-journée par les stagiaires et le formateur.

(1) Modalité et moyens pédagogique :

Formation délivrée en présentiel ou distanciel * (e-learning, classe virtuelle, présentiel à distance). Le formateur alterne entre méthodes ** démonstrative, interrogative et active (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.

Les moyens pédagogiques mis en oeuvre (variables suivant les formations) sont : ordinateurs Mac ou PC (sauf pour les cours de l'offre Management), connexion internet fibre, tableau blanc ou paperboard, vidéoprojecteur ou écran tactile interactif (pour le distanciel). Environnements de formation installés sur les postes de travail ou en ligne. Supports de cours et exercices.

* Nous consulter pour la faisabilité en distanciel. ** Ratio variable selon le cours suivi.

Programme

Jour 1

Introduction à Kafka

- Kafka : une plateforme de données de flux
- Un aperçu de Kafka et de son efficacité
- Producers
- Brokers
- Consumers

Installation et architecture

- Installation et options avancées
- Fichiers journaux de Kafka
- Réplication et fiabilité
- Chemins d'écriture et de lecture de Kafka
- Partitions, consumer groups et scalability

Développement et mise en oeuvre

- Cas d'utilisation de Kafka Streams
- Kafka Streams
- Etude d'une application et usage de Kafka Streams

Jour 2

Introduction à Cassandra

- Introduction au NoSQL
- Prérequis de Cassandra
- Installation
- Fichier de configuration : conf/cassandra.yaml
- L'architecture
- Les répliquions
- *Création de colonnes*
- *Insertion*
- *Modification recherche*
- *Définition des colonnes et des types de données*
- *Définition d'une clé primaire*
- *Reconnaître une clé de partition*
- *Le CQL : Cassandra Query Language*
- *Exécution de scripts*
- *Data Modeling*
- *Créer une application*
- *Ajout des noeuds*

Exemples de travaux pratiques (à titre indicatif)

- *Commandes de base*
 - *Connexion au système de base de données*

Jour 3

Apache Spark

- Introduction à Spark
- Vue d'ensemble de Spark
- Ecosystème Spark
- Spark vs Hadoop
- Installation et configuration
- *Spark :*
 - *RDD (Resilient Distributed Dataset)*
 - *DataFrame*
 - *SQL*
 - *Streaming*
 - *Mllib*
 - *GraphX*
- *Création d'applications Spark*

Exemples de travaux pratiques (à titre indicatif)

- *Remise à niveau Python / Scala*

Jour 4

- Projet avec la mise en oeuvre d'une solution bout en bout (ingestion, stockage et calcul) via :
 - Kafka
 - Cassandra
 - Spark

Modalités d'évaluation des acquis

L'évaluation des acquis se fait :

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation ou une certification (M2i ou éditeur)