

Data Engineering avec Kafka, Cassandra et Spark

4 jours (28 heures)

Perfectionnement

Présentiel ou classe à distance

à partir de

3 140 € HT par personne (TVA en vigueur) Repas inclus en présentiel

Contenu mis à jour le 25/09/2026. Document téléchargé le 04/10/2026.

Objectifs

A l'issue de cette formation, vous serez capable de :

- Gérer les outils Spark, Cassandra et Kafka
- Intégrer ses outils dans un même environnement
- Développer une solution de Data Engineering de bout en bout avec Spark, Cassandra et Kafka
- Identifier les points critiques et les leviers d'amélioration de la performance.

Les compétences visées par cette formation font l'objet d'une évaluation.

À qui s'adresse cette formation ?

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

Prérequis

Avoir des connaissances sur Shell et Python.

Programme

1 Pendant votre formation

Data Engineering avec Kafka, Cassandra et Spark

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 réplications

Exemples de travaux pratiques (à titre indicatif)

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

Jour 3

Apache Spark

- Introduction à Spark
- Vue d'ensemble de Spark
- Ecosystème Spark
- Spark vs Hadoop
- Installation et configuration

Exemples de travaux pratiques (à titre indicatif)

- *Remise à niveau Python / Scala*
- *Spark :*
 - *RDD (Resilient Distributed Dataset)*
 - *DataFrame*
 - *SQL*
 - *Streaming*
 - *MLlib*
 - *GraphX*
- *Création d'applications Spark*

Jour 4

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

Financement

Plusieurs dispositifs de financement sont accessibles via les OPCO (Opérateurs de Compétences), organismes agréés par le ministère du Travail dont le rôle est d'accompagner, collecter et gérer les contributions des entreprises au titre du financement de la formation professionnelle.

Pour plus d'information, une équipe de gestionnaires SKOLAE Formation spécialisée vous accompagne dans le choix de vos formations et la gestion administrative.

[Nous contacter pour en savoir plus →](#)

Modalités pédagogiques

Formation délivrée en présentiel ou distanciel* (blended-learning, e-learning, classe virtuelle, présentiel à distance).

Le formateur alterne entre méthode** démonstrative, interrogative et active (via des travaux pratiques et/ou des mises en situation).

Variables suivant les formations, les moyens pédagogiques mis en oeuvre sont :

- Ordinateurs Mac ou PC (sauf pour certains 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

En cas de formation intra sur site externe à M2I, le client s'assure et s'engage également à avoir toutes les ressources matérielles pédagogiques nécessaires (équipements informatiques...) au bon déroulement de l'action de formation visée conformément aux prérequis indiqués dans le programme de formation communiqué.

* nous consulter pour la faisabilité en distanciel

** ratio variable selon le cours suivi

[Accessibilité aux personnes en situation de handicap](#)

Détails de la session

Du 17 novembre 2026 au 20 novembre 2026 (4 jours / 28 heures)

Toulouse

Du 17 novembre 2026 au 20 novembre 2026

Toulouse

- 17 novembre 2026 9h00 – 15h30
- 18 novembre 2026 9h00 – 15h30
- 19 novembre 2026 9h00 – 15h30
- 20 novembre 2026 9h00 – 15h30