

Architectures de données (data architecture) pour le Machine Learning et l'Intelligence Artificielle (IA)

3 jours (21 heures)

Initiation

Présentiel ou classe à distance

à partir de

2 625 € HT par personne (TVA en vigueur) **Repas inclus en présentiel**

Contenu mis à jour le 25/09/2026. Document téléchargé le 07/10/2026.

Objectifs

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

- Concevoir une plateforme "Cloud-native" ou hybride pour le Machine Learning et l'Intelligence Artificielle (IA)
- Consolider les actifs Data de l'organisation Data-driven dans une seule plateforme
- Démocratiser l'accès aux données et permettre aux équipes métier de produire des modèles de ML (Machine Learning) et d'IA en self-service
- Automatiser la prise de décision dans l'organisation grâce à des pipelines de streaming
- Ajouter des fonctionnalités prédictives et prescriptives aux fonctionnalités descriptives de la BI (Business Intelligence).

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

À qui s'adresse cette formation ?

Architectes Data, développeurs, Data Scientists, Data Analysts, ingénieurs Data, chefs de projets, IA Product Managers, consultants BI / Big Data.

Prérequis

Avoir des connaissances en SI (Systèmes d'Information), en bases de données et en entraînement de modèles de Machine Learning et/ou d'Intelligence Artificielle (IA).

Programme

1 Pendant votre formation

Architectures de données (data architecture) pour le Machine Learning et l'Intelligence Artificielle (IA)

Jour 1 – Matin

Fondamentaux des architectures Data

- Qu'est-ce que le Machine Learning et l'IA ?
- Similarités et différences entre l'IA et le développement d'applications traditionnelles
- Importance des pipelines Data dans la minimisation du Time-to-Value des produits Data
- Rappel des principes de DataOps
- Importance de l'architecture Data dans l'implémentation d'une stratégie DataOps
- De la BI (Business Intelligence) au Big Data aux Analytics Métier
- Architectures Lambda et Kappa
- Evolution des plateformes architecturales d'analytics unifiées : Data Warehouse vs Data Lake vs Lakehouse vs Data Mesh

Concevoir une plateforme d'analytics moderne

- Cycle de vie de la Data
- Gouvernance et sécurité
- Offres logiciels libres en local (On Premise) : DVC, Python SDK, MLflow, Airflow, Spark, dbt (tiers gratuit vs tiers payant), Delta Lake, Iceberg, Hudi
- Offres Cloud classiques (Azure, AWS, GCP) et offres multi-Cloud (Databricks, Snowflake, Fivetran, dbt)
- Avantages et inconvénients des offres locales et des offres Cloud
- Moderniser les workflows Data

Exemples de travaux pratiques (à titre indicatif)

- Démonstration d'outils libres
- Prise en main de Databricks

Jour 1 – Après-midi

Architecture d'un Data Warehouse moderne

- Spécificités d'un Data Warehouse
- Variantes architecturales d'un Data Warehouse moderne
- Etapes d'un projet de Data Warehouse moderne
- Data Warehouse On-Premise vs sur le Cloud : solutions technologiques et avantages / inconvénients
- Intégration du Data Warehouse au SI
- Gouvernance du Data Warehouse

- Le Data Warehouse au service des Data Scientists

Exemple de travaux pratiques (à titre indicatif)

- Déploiement d'un Data Warehouse sur le Cloud et exécution d'analytiques

Jour 2 – Matin

Développer la Data team

- Définir les besoins clés de l'organisation
- Les DADO (Data-Analytics Driven Organizations)
- Les DEDO (Data-Engineering Driven Organizations)
- Les DSDO (Data-Science Driven Organizations)

Frameworks de migration

- La migration en quatre étapes :
 - Prepare / discover
 - Assess / plan
 - Execute
 - Optimize
- Estimation des coûts :
 - Audit de l'existant
 - Devis
 - PoC
 - MVP
- Mettre en place la sécurité et la gouvernance
- Schéma, pipeline, migration des données

Exemple de travaux pratiques (à titre indicatif)

- Etudes de cas de conception de plateformes analytiques, mettant en oeuvre les principes abordés

Jour 2 – Après-midi

Architecture d'un Data Lake

- Spécificités d'un Data Lake
- Etapes d'un projet de Data Lake
- Data Lake On-Premise vs sur le Cloud : solutions technologiques et avantages / inconvénients
- Intégration du Data Lake au SI (Système d'Information)
- Zonage et gouvernance du Data Lake
- Le Data Lake en self-service
- Importance du catalogue de données
- Le Data Lake au service des Data Scientists

Exemple de travaux pratiques (à titre indicatif)

- *Déploiement d'un Data Lake sur le Cloud et exécution d'analytics*

Jour 3 – Matin

Architecture d'un Lakehouse

- Spécificités d'un Data Lakehouse : le meilleur de deux mondes
- Découplage stockage / calcul, propriétés ACID, upserts, Time Travel, dérive de schéma, compaction des fichiers, Z-Ordering, VACUUM
- Variantes architecturales d'un Lakehouse
- Etapes d'un projet de Data Lakehouse
- Offres de Lakehouses sur le Cloud
- Intégration du Lakehouse au SI
- Gouvernance du Lakehouse
- Le Lakehouse au service des Data Scientists

Exemple de travaux pratiques (à titre indicatif)

- *Déploiement d'un Lakehouse sur le Cloud et exécution d'analytiques BI*

Jour 3 – Après-midi

Architectures de streaming

- L'intérêt du streaming
- Streaming ingest (ETL, ELT, Insert, IoT, Sink)
- Tableaux de bord en temps réel, live querying
- Analytiques de traitement de flux de données
- Apprentissage d'un modèle sur des données en streaming

Exemple de travaux pratiques (à titre indicatif)

- *Réalisation d'analytiques sur un flux de données*

Perspectives

- La Data Intelligence, l'IA Générative qui :
 - Gère la gouvernance
 - Enrichit le catalogue de données
 - Permet des requêtes en langage naturel (AI/BI)
 - Optimise l'indexation et les requêtes

Exemple de travaux pratiques (à titre indicatif)

- *Démonstration avec Databricks Data Intelligence Platform*

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 2 février 2027 au 4 février 2027 (3 jours / 21 heures)

À distance

Du 2 février 2027 au 4 février 2027

À distance

- 2 février 2027 9h00 – 16h00
- 3 février 2027 9h00 – 16h00
- 4 février 2027 9h00 – 16h00