



Data et Big Data

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

3 jours (21h00) | 9 4,6/5 | IA-ARCML | Évaluation qualitative de fin de stage | Formation délivrée en présentiel ou distanciel

Formations Intelligence Artificielle > IA par Métier > Data et Big Data

Contenu mis à jour le 13/10/2023. Document téléchargé le 27/07/2024.

## Objectifs de formation

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

- Concevoir une plateforme "Cloud-native" ou hybride pour le Machine Learning et l'IA (Intelligence Artificielle)
- 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 capacités prédictives et prescriptives aux capacités descriptives de la BI (Business Intelligence).

## Modalités, méthodes et moyens 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

## Prérequis

Avoir des connaissances en SI (Systèmes d'Information) et en bases de données.

## Public concerné

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

## 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 émargée par demi-journée par les stagiaires et le formateur.

# Programme

## Jour 1

### Fondamentaux de l'IA et des architectures Data

- Qu'est-ce que l'IA ?
- Similarités et différences entre IA et le développement d'applications traditionnelles
- Machine Learning vs Deep Learning
- Batch vs temps réel vs streaming
- De la BI (Business Intelligence) au Big Data aux Business Analytics
- Evolution des plateformes d'analytics unifiées : Data Warehouse vs Data Lake vs lakehouse vs Data Mesh
- Les différents types d'apprentissage de modèles
- Les grands modèles de langage
- De GPT-1 à GPT-4
- Etudes de cas d'utilisation des grands modèles de langage et des modèles génératifs
- Avantages et limites
- Vue d'ensemble de l'offre IA sur le Cloud Azure
- Les différents services :
  - D'Azure Cognitive Services
  - D'Azure OpenAI

### Concevoir une plateforme d'analytiques moderne

- Cycle de vie de la Data
- Gouvernance et sécurité
- Apports des offres Cloud
- Offres Cloud classiques (Azure, AWS, GCP) et offres multi-Cloud (Databricks, Snowflake, Fivetran, dbt)
- Moderniser les workflows Data

### Exemples de travaux pratiques (à titre indicatif)

- *Prise en main d'Azure Cognitives Services, création d'une application de reconnaissance d'images*
- *Prise en main d'Azure OpenAI, création d'une application à base de GPT-4*

## Jour 2

### 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

### Exemples de travaux pratiques (à titre indicatif)

- Etudes de cas de conception de plateformes d'analytics mettant en oeuvre les principes abordés

### 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

### Exemples de travaux pratiques (à titre indicatif)

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

## Jour 3

### Architecture d'un Data Warehouse

- Spécificités d'un Data Warehouse
- Variantes architecturales d'un Data Warehouse (ETL, intégrateurs, magasins de données...)
- Etapes d'un projet de Data Warehouse
- 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

### Exemples de travaux pratiques (à titre indicatif)

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

### Architecture d'un lakehouse

- Spécificités d'un Data Lakehouse : le meilleur de deux mondes

- 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

### **Exemples de travaux pratiques (à titre indicatif)**

- Déploiement d'un lakehouse sur le Cloud et exécution d'analytics

### **Architectures de streaming**

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

Le contenu de ce programme peut faire l'objet d'adaptation selon les niveaux, prérequis et besoins des apprenants.

### **Modalités d'évaluation des acquis**

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation

### **Accessibilité de la formation**

Le groupe M2i s'engage pour faciliter l'accessibilité de ses formations. Les détails de l'accueil des personnes en situation de handicap sont consultables sur la page Accueil et Handicap.

### **Modalités et délais d'accès à la formation**

Les formations M2i sont disponibles selon les modalités proposées sur la page programme. Les inscriptions sont possibles jusqu'à 48 heures ouvrées avant le début de la formation. Dans le cas d'une formation financée par le CPF, ce délai est porté à 11 jours ouvrés.