

Devenir Concepteur·trice Développeur·se en science de la donnée

MaCarrière

m²i
Formation

En partenariat avec :



Région
Hauts-de-France

Document mis à jour le 04/07/2025

Dispositif de formation :	Programme Régional de Formation (PRF)
Date de formation :	Du 13 octobre 2025 au 11 septembre 2026
Période de stage :	Du 18 mai au 24 juillet 2026
Lieu(x) de formation :	Villeneuve d'Ascq
Réunion d'Information :	Le 11 septembre 2025 ou le 3 octobre 2025
Date limite d'envoi de candidature :	Le 1^{er} octobre 2025

Vous êtes à l'aise avec les chiffres et la logique ? Vous aimez résoudre des problèmes ? Vous êtes curieux et aimez apprendre ? Vous êtes patient et avez le souci du détail ? Vous souhaitez exercer un rôle clé dans une organisation et aider à une prise de décision éclairée ?

Devenez Concepteur Développeur en science de données !

Depuis l'émergence de l'Intelligence Artificielle et des Big Data, les Concepteurs Développeurs en Science des Données sont recrutés dans de nombreux domaines (finance, santé, recherche, administration, retail...), et par une grande variété d'organisations (de l'industrie en passant par de grandes entreprises comme des start-up).

Le travail d'analyse du concepteur développeur en science de données est à l'origine de l'élaboration de la stratégie d'une organisation.

Tous les secteurs d'activités utilisant de la donnée sont ainsi concernés et quel que soit la taille de l'entreprise.

La région Hauts-de-France vous forme et vous prépare au passage du titre professionnel Concepteur développeur en sciences de données !

Votre formation se déroulera en présentiel dans notre centre de Villeneuve d'Ascq du 13 octobre 2025 au 11 septembre 2026 (périodes de congé incluses).

Votre formation sera animée par des experts métier et terrain qui vous donneront les clés de la réussite dans ce métier.

En parallèle de votre formation, un accompagnement dans votre retour à l'emploi vous sera proposé à travers des ateliers collectifs et individuels.

Les candidatures seront étudiées à partir du lundi 16 juillet 2025.

Cette formation financée par la région Hauts-de-France est à destination des personnes résidant dans les Hauts-de-France.

OBJECTIFS DE FORMATION

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

- Créer des infrastructures de gestion de données robustes et adaptées et de les alimenter,
- Développer des algorithmes d'intelligence artificielle et de les mettre en production,
- Collaborer avec les différentes équipes-métiers métier d'une organisation pour évaluer et adapter les besoins data.

PRÉREQUIS

- Bac +2 scientifique ou technique (mathématiques, économie, informatique, ingénierie, physique...)
- Bonnes bases en algorithmique et logique (vous avez des notions en Algorithmie)
- Appétence pour les environnements Cloud et les technologies de données (Python, SQL et/ou Linux)
- Appétence pour le développement et idéalement connaissance du langage de programmation Python
- Bonne connaissance des outils informatiques
- Capacité à travailler en équipe et à communiquer avec des interlocuteurs techniques et métiers
- Autonomie, rigueur, esprit d'analyse, organisation, efficacité et agilité
- Vous avez un excellent relationnel et êtes capable de vous adapter à différents interlocuteurs : clients, interlocuteurs techniques, etc.
- Vous êtes bon communicant appréciez le travail en équipe et faites preuve d'excellentes capacités rédactionnelles
- La maîtrise de l'anglais professionnel sera fortement appréciée

PUBLIC CONCERNÉ ET FINANCEMENT

Cette formation financée à 100%* est à destination des demandeur-se-s d'emploi inscrit-e-s à France Travail et résidant dans la région Hauts-de-France.

COMPÉTENCES ET TECHNOLOGIES ABORDÉES

Compétences	Module et objectifs opérationnels	Durée (Jours)	Durée (Heures)
Préparation	Intro à la data et à l'environnement technique : Présentation du programme, outils utilisés, prise en main de l'environnement de travail, découverte de l'écosystème data.	1	7
Préparation	Linux Bash et administration : commandes et scripts: Maîtriser les commandes de base du shell Linux, créer des scripts pour automatiser des tâches. Gérer utilisateurs, permissions, services de base, configuration réseau et dépannages courants.	3	21
Préparation	Préparer Docker (installation, volumes, docker-compose) : Introduction aux concepts de la conteneurisation, manipulation de Docker en ligne de commande, volumes, images, docker-compose.	4	28
Préparation	Mini Data Lake local (MinIO + Docker + S3 Python SDK) : Création d'un environnement local simulant un Data Lake à l'aide de MinIO et Docker. Utilisation du SDK boto3 pour l'accès programmatique.	3	21
Préparation	Python bases : variables, types, structures: Bases du langage Python : variables, structures de contrôle, listes, dictionnaires, boucles.	3	21
Préparation	Python avancé : fonctions, modules, fichiers: Modularisation avec fonctions et fichiers, manipulation de fichiers, compréhension des erreurs.	3	21
Préparation	Initiation à Git et Jupyter : Versionnage avec Git, travail collaboratif, utilisation des notebooks Jupyter pour le data analysis.	2	14
Bloc 1	Bases de données relationnelles & modélisation : Comprendre les bases de données relationnelles, modéliser les données avec un schéma relationnel.	2	14
Bloc 1	SQL fondamental (SELECT, WHERE, GROUP BY) : Maîtriser les requêtes SQL de base : sélection, filtrage, regroupement, tri.	3	21
Bloc 1	SQL avancé (JOINS, VUES, AGREGATS) : Requêtes complexes avec jointures, vues, sous-requêtes, agrégats.	3	21
Bloc 1	ETL : fichiers, Excel, API REST : Collecter, Extraire et transformer des données depuis des fichiers structurés, feuilles Excel et APIs.	3	21
Bloc 1	Web scraping avec Python : Utiliser BeautifulSoup ou Scrapy pour collecter automatiquement des données sur le web.	3	21
Bloc 1	NoSQL : MongoDB + DynamoDB AWS: Découverte des bases de données NoSQL avec MongoDB (modèle document) et DynamoDB (clé-valeur) sur AWS.	3	21
Bloc 1	Stockage Cloud avec AWS S3 / MinIO : Apprendre à stocker, organiser et sécuriser des données dans un stockage objet compatible S3 (AWS ou MinIO) en contexte de Data Lake.	3	21
Bloc 1	Éval Bloc 1 - Étude de cas Data Lake & ETL : Étude de cas sur l'intégration et le traitement de données massives via un data lake.	2	14
Bloc 2	ETL distribué avec Spark + Pandas : Conception de pipelines de traitement et nettoyage de données avec Spark et Pandas sur des jeux de données massifs.	3	21
Bloc 2	Manipulation avec Numpy & Pandas : Nettoyage, transformation, fusion de données avec les bibliothèques Pandas et Numpy.	2	14
Bloc 2	Spark SQL & DataFrames distribués : Utilisation de Spark SQL pour interroger de grandes volumétries de données en mémoire distribuée.	2	14
Bloc 2	Nettoyage avec Spark & Pandas : Comparaison et combinaison de nettoyage de données avec Pandas localement et Spark de manière distribuée.	2	14
Bloc 2	Visualisation de données (Matplotlib, Seaborn, Plotly) : Créer des graphiques interactifs et professionnels avec des bibliothèques Python de visualisation de données.	2	14

Bloc 2	Statistiques inférentielles & tests : Réaliser des tests statistiques (student, khi2) pour analyser les distributions.	2	14
Bloc 2	Analyse multivariée : Comprendre les relations entre plusieurs variables avec ACP, corrélations, clustering hiérarchique.	2	14
Bloc 2	Éval Bloc 2 - Analyse exploratoire & Spark : Études de cas avec données massives sur Spark.	2	14
Bloc 3	ML supervisé avec Scikit-learn : Créer des modèles de classification et de régression avec Scikit-learn sur des jeux de données réels.	4	28
Bloc 3	ML avec Spark MLlib : Utilisation de Spark MLlib pour l'apprentissage supervisé et non supervisé à l'échelle Big Data.	3	21
Bloc 3	Optimisation & évaluation des modèles : Métriques de scoring, validation croisée, hyperparamètres.	3	21
Bloc 3	Apprentissage non-supervisé : Découvrir les techniques de segmentation (clustering, réduction de dimension) et leur mise en œuvre.	4	28
Bloc 3	Éval Bloc 3 - ML Scikit vs Spark : Évaluation comparative entre deux pipelines ML sur Scikit-learn et Spark MLlib appliqués à des données volumineuses.	2	14
Bloc 4	Prétraitement texte et NLP : Techniques de nettoyage, vectorisation, stemming et tokenization de texte pour des applications NLP.	4	28
Bloc 4	Deep Learning avec TensorFlow (MLP, CNN) : Construire des modèles de réseaux de neurones profonds avec TensorFlow : perceptrons multicouches et CNN.	5	35
Bloc 4	Modèles avancés et évaluation : Appliquer des techniques de régularisation, dropout, et validation pour fiabiliser les modèles profonds.	3	21
Bloc 4	Éval Bloc 4 - Étude de cas NLP : Étude de cas complète : entraîner un modèle pour analyser des sentiments sur des avis utilisateurs textuels.	2	14
Bloc 5	Dockerisation d'un projet Data : Création d'un environnement dockerisé complet pour encapsuler un projet de data science ou d'analyse.	3	21
Bloc 5	Création d'API Flask / FastAPI : Développer et documenter une API RESTful en Python avec Flask ou FastAPI pour servir des modèles ML.	4	28
Bloc 5	Déploiement via Docker + SageMaker : Déploiement d'API ou de modèle dans des conteneurs Docker localement ou sur SageMaker via endpoints.	3	21
Bloc 5	MLflow, suivi des modèles : Gérer les versions, suivis de performances et déploiements de modèles ML grâce à MLflow.	3	21
Bloc 5	Éval Bloc 5 - Déploiement IA + API : Évaluation complète incluant le déploiement d'un modèle ML dans une API Dockerisée et testée.	2	14
Bloc 6	Gestion de projet Data : Maîtriser les étapes d'un projet data de A à Z : cadrage, conception, gestion des risques, communication.	4	28
Bloc 6	Indicateurs, pilotage et vulgarisation : Élaboration de tableaux de bord de suivi projet et techniques de vulgarisation pour directions non techniques.	4	28
Agilité	Agilité : travailler en méthode agile	2	14
Posture	Posture professionnelle : Savoir-être, communication, autonomie et posture consultant.	2	14
Bloc 6	Projet final data	6	42
Bloc 6	Éval Bloc 6 - Soutenance Projet Data : Soutenance du projet global devant jury.	1	7
TRE	Techniques de recherche d'emploi : Préparer CV, entretiens, LinkedIn, valoriser les projets.	3	21
Bloc 6	Révisions & Examens blancs : Préparer les évaluations finales du titre.	4	28
Autonomie	E-learning/TP	30	210

DIPLÔME

Passage du titre professionnel de niveau 6 : Concepteur développeur en science de la donnée.

À PROPOS DES CERTIFICATIONS

Certification éditeur :

Les certifications éditeurs dépendent des éditeurs uniquement (PEGA, SAP, Salesforce, Microsoft, Red Hat...). Chaque éditeur a sa propre façon d'évaluer son candidat.

Cela se fait souvent par le biais d'un QCM chronométré. Sur cette base, il faut généralement obtenir un pourcentage de succès supérieur à 70%. Merci de prendre le temps de demander à vos interlocuteurs M2i le cadre et les conditions de ce passage.

Validation des acquis M2i :

La validation des acquis M2i se fait soit par un QCM chronométré, soit par un examen de fin de parcours ou les candidats passent par groupes de 2 ou 3 devant un jury qui déterminera vos acquis suite à votre cursus de formation. Généralement cette soutenance est précédée de 3 jours de travaux pratiques en groupe afin de préparer cette soutenance. Les équipes M2i pourront vous guider dans votre projet.

LES PLUS DE M2I FORMATION

Microsoft Teams Education, un outil de suivi et d'animation en présentiel et à distance :

- Espace de stockage collaboratif pour accéder aux différents supports de cours et cahiers d'exercices
- Fonctionnalités pour gérer des sessions à distance

La playlist e-learning : tous les apprenants ont accès avant, pendant et après le cursus à notre plateforme e-learning M2i Learning pour :

- Acquérir les connaissances prérequis avant de démarrer (sous condition de l'existence des modules pour combler les lacunes)
- Utiliser les modules conseillés par les formateurs pour faire de l'ancrage mémoriel sur des sujets abordés pendant la formation
- Revenir sur un sujet après la formation pour continuer à s'auto-former

MODALITÉS, MOYENS ET MÉTHODES PÉDAGOGIQUES

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

Le formateur alterne entre des méthodes démonstratives 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.

MODALITÉS D'ACCÈS

Nos équipes accorderont toute leur attention au traitement de votre candidature et s'engagent à vous faire un premier retour dans un délai de 7 jours.

ADMISSION

- Dossier de candidature
- Test de vérification des prérequis
- Entretien
- Validation et éligibilité France Travail (dans le cas de certains dispositifs)

Le groupe M2i s'engage pour faciliter l'accessibilité de ses formations. Les détails de l'accueil des personnes en situation en handicap sont consultables sur la page [Accueil PSH](#).

POURQUOI CHOISIR M2I FORMATION ?

- Un apprentissage métier proactif axé sur l'emploi et basé sur le faire avec l'accompagnement de nos formateurs tout au long du parcours.
- Un accès à des experts : bénéficiez de l'expertise de nos formateurs.
- En présentiel ou à distance : accès individuel aux ressources de formation et progression personnalisée si besoin.
- Outils de suivi collectif et individuels (espaces d'échanges et de partage en ligne, salles virtuelles si en présentiel ou à distance, supports de cours, TP, exercices).

MODALITÉS D'ÉVALUATION

Exemples de validation des acquis de formation :

- Travaux dirigés dans chaque module
- Mise en situation via des cas pratiques et un mini projet
- Certification (si prévue dans le programme de formation)
- Soutenance devant un jury de 30 à 40 min

