Devenir Data Engineer

Document mis à jour le 25/11/2026

Dispositif de formation :	**Préparation Opérationnelle à l'Emploi (POE)**
Durée de la formation :	**85 jours (595h)**
Date de formation :	**Du 22 avril au 11 septembre 2024**
Lieu(x) de formation :	**Villeneuve d'Ascq**
Réunion d'Information :	**Le 8 avril 2024**
Date limite d'envoi de candidature :	**Le 4 avril 2024**

Le **Data Engineer** définit, développe, met en place et maintient les outils et infrastructures adéquats à l'analyse de la donnée par les équipes de Data Science. Il veille à créer une solution permettant le traitement de volumes importants de données tout en garantissant la sécurité de celles-ci.

Il représente le premier maillon de la chaîne de traitement de données.

Cette fonction exige un vrai sens de l'organisation et de la méthode ainsi qu'un sens aigu du travail en équipe.

Lors d'une formation de 85 jours, en présentiel à Villeneuve d'Ascq, M2i Formation en partenariat avec Invest in Digital People (dispositif France Travail) et les entreprises partenaires, vous transmettront les compétences techniques essentielles du métier.

En amont de la formation, les candidats participeront à un job dating le 11 Avril 2024.

Votre formation se déroulera du 22 Avril au 06 Septembre 2024 (période de congé incluse du 03 au 18 Août 2024).

Nous recrutons, pour ce projet, des candidats prêts à rejoindre durablement le secteur de l'informatique pour les accompagner dans le cadre d'une Préparation Opérationnelle à l'Emploi (POE).

Votre candidature sera étudiée selon les prérequis du poste, les profils retenus seront obligatoirement ceux en adéquation avec la formation envisagée.

OBJECTIFS DE FORMATION

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

- d'apporter une expertise en Big Data permettant la manipulation des données
- de concevoir les plateformes permettant de traiter des volumes importants données
- de mettre en place des bases de données
- de veiller à ce que les pipelines de données soient sécurisés et clairs pour être analysés et transformés

Ainsi, en tant que **Data Engineer** vous définirez, développerez, mettrez en place et maintiendrez les outils et infrastructures adéquats à l'analyse de la donnée par les équipes de Data Science. Vous veillerez à créer une solution permettant le traitement de volumes importants de données tout en garantissant la sécurité de celles-ci.
Vous représenterez le premier maillon de la chaîne de traitement de données.

PRÉREQUIS

– Vous êtes titulaire d'un BAC +3 minimum scientifique ou informatique
– Vous avez des notions en Algorithmie, Python, SQL et/ou Linux
– 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.

COMPÉTENCES ET TECHNOLOGIES ABORDÉES

Planning non contractuel pouvant être soumis à modifications :

Module	Durée (Jours)	Durée (Heures)
Présentation du cursus, des plateformes pédagogiques et du fil rouge (DataViz + GCP) : Ice Breaking, tour de table - Avoir une description du cursus et des choix d'orientation - Comprendre l'utilisation des différentes plateformes d'apprentissage du cursus - Découvrir le projet "fil-rouge".	0,5	3,5
Apprendre à apprendre : Comprendre comment fonctionne le cerveau pour mémoriser plus efficacement - Mettre en place des outils et méthodes d'apprentissage et de mémorisation - Comprendre l'utilisation des différentes plateformes d'apprentissage du cursus - Avoir une description du cursus et des choix d'orientation.	0,5	3,5
Adopter la posture gagnante du consultant en ESN : Maîtriser le contexte des ESN - Maîtriser les différents types de prestations - Comprendre les cycles de ventes - Comprendre le métier de consultant - Adapter votre posture - Identifier vos axes de progressions.	1	7
Algorithmique et programmation structurée orientée Python : Disposer des connaissances nécessaires à l'apprentissage d'un langage de développement - Connaître les structures de base de la programmation (boucles, conditions) - Savoir quelles sont les grands paradigmes de programmation (procédural, objet) - Comprendre la notion d'objet et les concepts associés - Identifier les apports de la modélisation UML - Disposer d'un premier point de vue sur les approches Python - Découvrir les variables et le typage des données.	4	28
Python et Python Orienté Objet dans Visual Studio Code : Identifier les usages courants du langage - Installer la prise en charge de Python dans Visual Studio - Mettre en pratique le scripting en Python - Structurer votre code en fonction, classes et modules - Utiliser des modules existants - Décrire la programmation réseau avec Python - Expérimenter la programmation objet en Python.	5	35
Mises en pratique et Validation des acquis - Python : Etudes de cas - Quiz - Exercices - Projet groupé	2	14
Linux - les fondamentaux : Distinguer les logiciels libres et l'Open Source - Identifier les principes fondamentaux du système d'exploitation - Utiliser interactivement le Shell et connaître les commandes essentielles - Gérer les fichiers et les dossiers - Editer un fichier - Reconnaître les métacaractères et les expressions régulières.	4	28
Linux - Programmation Shell Bash : Décrire les notions de bases de Linux, de manière approfondie - Enrichir votre corpus de commandes - Développer des scripts Shell.	4	28
Mises en pratique et Validation des acquis - Infrastructure Linux : Etudes de cas - Quiz - Exercices - Projet groupé	2	14
Langage SQL : Fondamentaux : Décrire les principaux concepts des SGDBR (Système de Gestion des Bases de Données Relationnelles) et d'algèbre relationnelle utilisés dans le langage SQL - Interroger une base de données avec la clause SQL SELECT - Utiliser les commandes SQL de mise à jour des données - Identifier les commandes SQL de début et fin de transaction BEGIN, COMMIT et ROLLBACK - Présenter les concepts de gestion des privilèges systèmes et objets avec les commandes SQL GRANT et REVOKE - Créer, modifier et supprimer certaines catégories d'objets (table, index, vues...) avec CREATE, ALTER et DROP.	5	35

Agilité et Pilotage de projets : Gestion de projet : prédictive vs agile - Pourquoi l'agilité ? - Apprendre les pratiques agiles - Maîtriser la méthode agile Scrum .	2	14
Maitriser Talend: Mettre en place et paramétrer un environnement de développement stable sur la base de la version gratuite de Talend - Modéliser vos besoins - Utiliser la bibliothèque de composants - Implémenter, déboguer et déployer vos jobs..	3	21
Big Data -Enjeux et perspectives : Identifier l'ensemble des enjeux et facteurs à prendre en compte pour réussir l'intégration du Big Data dans la vision large du SI - Evaluer et sélectionner les outils appropriés dans le cadre d'un plan de mise en oeuvre du Big Data	2	14
Fondamentaux du cloud : Présenter la terminologie et les concepts de base du Cloud - Enumérer les services de base de la plateforme - Tester la navigation dans une console Cockpit - Gérer des ressources et des coûts sur le Cloud	2	14
Base de données NoSQL : Définir les caractéristiques techniques des bases de données NoSQL et les différentes solutions disponibles - Identifier les critères de choix.	3	21
Machine Learning avec Python : Expliquer et mettre en place un processus complet de Machine Learning - Explorer et préparer les données - Choisir et appliquer le bon algorithme -Entraîner et améliorer votre modèle et le déployer	5	35
Journée immersive en entreprise ou technique de recherche d'emploi : Mise en situation en milieu professionnel coachée par un consultant senior et aide à la recherche d'emploi	2	14
Introduction à l'IA générative :	1	7
Power BI - Les fondamentaux à l'expertise : Décrire le cycle de création d'un rapport Power BI - Vous connecter à des sources de données - Transformer, nettoyer et combiner des sources - Structurer un modèle de données - Créer des indicateurs - Afficher les indicateurs dans des rapports PBI - Partager des rapports - Mettre en place une solution de Business Intelligence avec les outils avancés Power BI - Les fonctions avancées du DAX - Partager en ligne vos tableaux de bords et rapports - Utiliser des visualisations interactives.	5	35
Qlik Sense - Du designer au développeur : Décrire les avantages et les fonctionnalités de Qlik Sense - Naviguer dans l'application Qlik Sense - Gérer les feuilles dans une application Qlik Sense - Manipuler les objets de feuille pour choisir ou restituer des données - Organiser les objets sur la feuille - Créer un graphique pour mettre en valeur les données - Charger simplement des données dans Qlik Sense - Exploiter l'outil de développement - Vous connecter aux sources de données - Nettoyer, manipuler et transformer des données - Réaliser des modélisations associatives et des optimisations - Utiliser des fichiers QVD - Réaliser des calculs avancés (filtres, agrégations, totaux et sous-totaux...) - Utiliser les éléments principaux.	5	35
Azure Data Engineering : Expérimenter les options de calcul et de stockage pour les workloads de Data Engineering dans Azure - Exécuter des requêtes interactives à l'aide de pools SQL sans serveur - Effectuer l'exploration et la transformation des données dans Azure Databricks - Explorer, transformer et charger des données dans le Data Warehouse à l'aide d'Apache Spark - Ingérer et charger des données dans le Data Warehouse - Transformer les données avec Azure Data Factory ou Azure Synapse Pipelines - Intégrer les données à partir des notebooks avec Azure Data Factory ou Azure Synapse Pipelines - Prendre en charge le traitement analytique transactionnel hybride (HTAP) avec Azure Synapse Link - Assurer la sécurité end-to-end avec Azure Synapse Analytics -Effectuer un traitement de flux en temps réel avec Stream Analytics - Créer une solution de traitement de flux avec Event Hubs et Azure Databricks.	5	35
Mises en pratique et Validation des acquis - Data Visualisation et Data Science : Etudes de cas - Quiz - Exercices - Projet groupé	2	14
Fondamentaux Devops : Décrire les principes de la démarche DevOps - Démontrer l'impact de la démarche DevOps dans les infrastructures et notamment l'Infrastructure as Code - Reconnaître l'impact de la démarche DevOps, des containeurs et du CaaS - Entamer une démarche vers une organisation DevOps.	2	14
Web scrapping: Réaliser du scraping de données - Faire les actions d'ingestion nécessaires pour alimenter un Data Lake.	3	21
Python Data Analysis : Utiliser Scikitlearn pour créer des modèles d'apprentissage machine - Concevoir des expériences et interpréter les résultats des tests A/B - Visualiser l'analyse de clustering et de régression en Python à l'aide de Matplotlib - Produire des recommandations automatisées de produit ou de contenu avec des techniques de filtrage collaboratif - Appliquer les meilleures pratiques en matière de nettoyage et de préparation de vos données avant l'analyse.	2	14

Docker pour Linux : Expliquer les avantages et inconvénients de la conteneurisation - Déployer et administrer la plateforme Docker - Configurer et utiliser le moteur Docker - Décrire la création des images Docker et "Dockerfile"- Intéragir avec le Docker Hub et registry privés - Mettre en oeuvre et configurer des conteneurs - Déployer des applications dans les conteneurs - Organiser la gestion des réseaux et du stockage - Maintenir et surveiller une infrastructure de conteneurs en production.	3	21
Kubernetes - Orchestrer ses conteneurs: Décrire les principes de l'orchestration de conteneurs - Manipuler les ressources de base Kubernetes - Déployer des applications et les mettre à disposition	2	14
Gérer le versionnning avec Git : Décrire les principes d'un gestionnaire de versions distribué - Identifier par la pratique, la philosophie de Git et ses apports.	2	14
Elastic-search: Situer Elasticsearch dans un écosystème Big Data - Identifier les enjeux et les cas d'utilisation d'un moteur de recherche - Indexer des données - Faire des recherches simples et complexes - Manipuler les agrégations et des recherches évoluées.	3	21
Projet final et Présentation des choix techniques et des solutions proposées en binome : Les apprenants vont restituer à l'oral les grandes lignes du projet fil rouge, ainsi que la manière dont il se sont appropriés le sujet et ont réalisés leurs recherches et leurs travaux.	4	28

À PROPOS DES CERTIFICATIONS

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.

LES PLUS DE M2I

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 ACADEMIIC pour :

- Acquérir les connaissances prérequises 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 dispensée à distance par des formateurs professionnels : apports théoriques, exercices de mise en situation professionnelle.

Les formateurs alternent entre la méthode affirmative (présentations, démonstrations), interrogative (tests, quiz...) et active (jeu de rôles, cas pratique, apprentissage par les pairs ...). Ils s'appuient sur les 4 piliers de l'apprentissage issus des dernières découvertes en neurosciences. Les journées sont rythmées par l'alternance entre théorie et pratique via un jeu d'exercices : exercices progressifs, exercices récapitulatifs et de validation des acquis.

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 Politique Handicap.

POURQUOI CHOISIR M2I ?

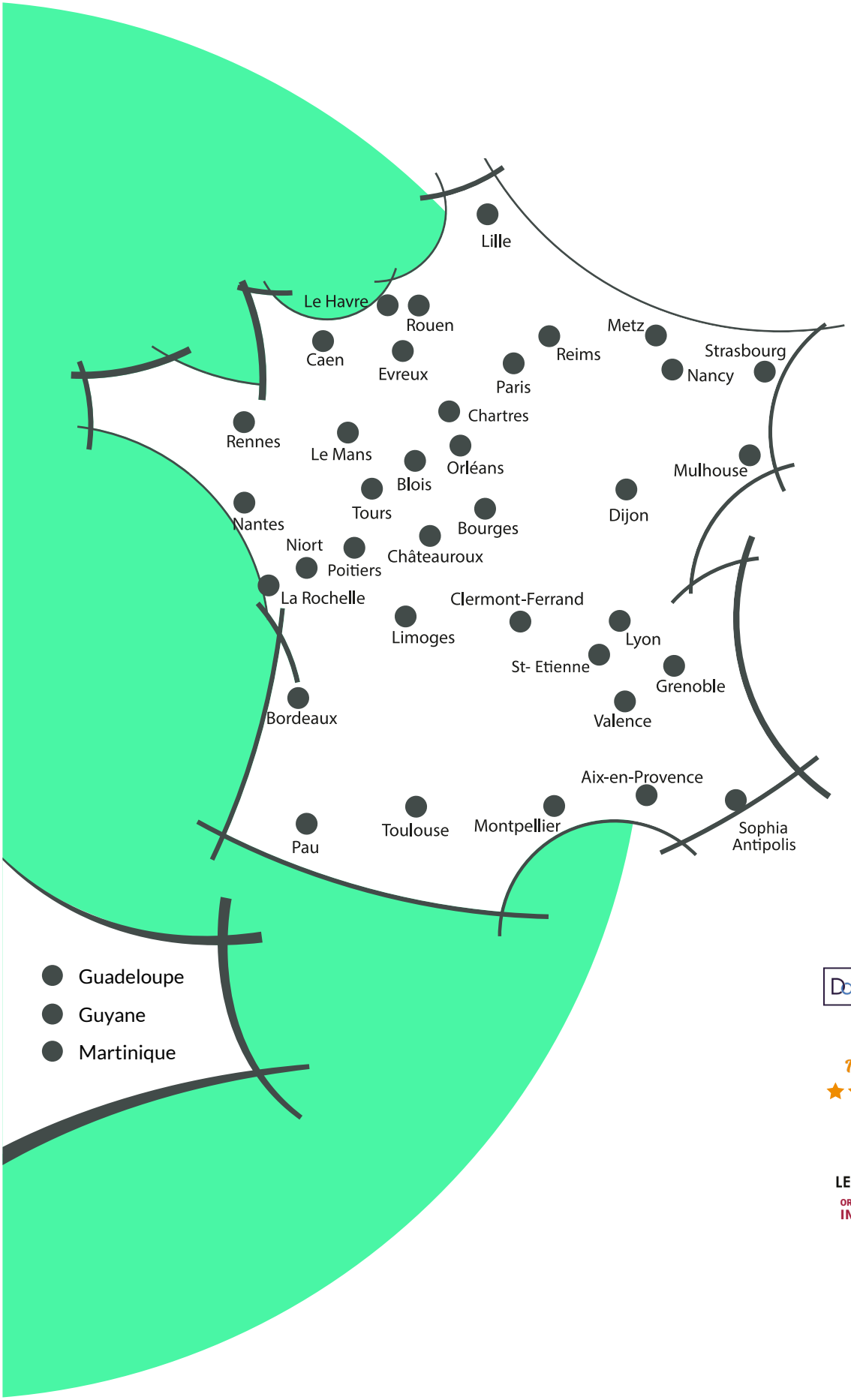
- Un apprentissage métier proactif axé sur l'emploi et basé sur le faire avec l'accompagnement de nos formateurs tout au long du parcours.
- Un accès à des experts : bénéficiez de l'expertise de nos formateurs.
- En présentiel ou à distance : accès individuel aux ressources de formation et progression 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

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.

MaCarrière

m2i Formation



Datadock

Avis Vérifiés 4,6/5

LEADERS LEAGUE
ORGANISME DE FORMATION INCONTOURNABLE

© Ma Carrière M2i Formation - Crédits photos : Getty Images - Document non contractuel - Les informations contenues dans ce document peuvent être modifiées sans préavis - NE PAS JETER SUR LA VOIE PUBLIQUE