

Devenir Consultant·e Data Engineer

MaCarrière

m²i
Formation

Document mis à jour le 19/05/2025

Dispositif de formation :	Préparation Opérationnelle à l'Emploi (POE)
Date de formation :	Du 30 juin au 5 décembre 2025
Lieu(x) de formation :	Villeneuve d'Ascq
Réunion d'Information :	Le 3 juin 2025 ou le 17 juin 2025
Date limite d'envoi de candidature :	Le 30 mai 2025

L'OPCO Atlas et la Région Hauts-de-France vous forment !

En partenariat avec M2i formation, formez-vous à l'un des métiers phares du monde numérique et devenez Consultant Data Engineer !

Le consultant Data Engineer conçoit, construit, maintient et optimise les architectures de traitement de données massives (Big Data).

Il développe des pipelines de données robustes, gère des infrastructures distribuées, met en œuvre des traitements à grande échelle (batch et temps réel), et exploite les services Cloud.

Il joue un rôle central dans la chaîne de valorisation de la donnée, en collaboration avec les Data Scientists, les Data Analysts et les architectes Cloud.

Ses missions principales sont :

Conception de pipelines de données :

- Développer des flux d'ingestion et de transformation de données avec PySpark, Hadoop et SQL.
- Concevoir des architectures distribuées performantes et résilientes.
- Optimiser les performances des traitements de données massives (batch et stream)

Intégration et exploitation Cloud :

- Déployer et orchestrer les traitements sur Google Cloud Platform (BigQuery, GKE, Dataflow, Cloud Storage).
- Mettre en place des workflows CI/CD pour les projets Data

Visualisation et restitution :

- Créer des dashboards interactifs avec Power BI, Qlik Sense et Looker.
- Travailler en lien avec les métiers pour transformer les données brutes en indicateurs exploitables.
- Garantir la qualité et la fiabilité des restitutions

Sécurité, qualité et conformité :

- Gérer les accès, les rôles (IAM) et les politiques de sécurité sur le Cloud.
- Assurer la conformité avec les standards de gouvernance des données.
- Automatiser les contrôles qualité, les audits de performance et la traçabilité des traitements

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

M2i Formation en partenariat avec l'OPCO Atlas et la région Hauts-de-France vous transmettra les compétences techniques essentielles du métier au cours d'une formation de 97 jours. 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 :

- Comprendre les enjeux du Big Data et de la Data Engineering
- Construire des pipelines robustes avec Hadoop et PySpark
- Interroger, transformer et charger des données avec SQL et BigQuery
- Déployer et gérer des traitements sur GCP avec des outils comme Dataflow ou GKE
- Visualiser efficacement les données avec Power BI, Qlik Sense et Looker
- Automatiser l'infrastructure avec Terraform et orchestrer avec Kubernetes
- Travailler en méthode agile, collaborer avec les Data Analysts, Data Scientists, DevOps et métiers

PRÉREQUIS

- Bac +2 en informatique, mathématiques, statistiques ou équivalent
- 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)
- Capacité à travailler en équipe et à communiquer avec des interlocuteurs techniques et métiers
- Autonomie, rigueur, esprit d'analyse
- 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 apprécié

PUBLIC CONCERNÉ ET FINANCEMENT

Cette formation financée à 100% par l'Opco Atlas et la Région Hauts-de-France est à destination des demandeur-se-s d'emploi inscrit-e-s à France Travail.

Après validation de votre candidature par l'entreprise, elle sera soumise à votre conseiller France Travail pour valider la cohérence entre votre profil et votre projet de formation.

COMPÉTENCES ET TECHNOLOGIES ABORDÉES

Catégorie	Modules et objectifs	Durée (en jours)	Durée (en heures)
Méthode	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
Méthode	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
Fondamentaux	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.	2	14
Programmation	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.	5	35
Programmation	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
Travaux Pratiques	Mises en pratique et Validation des acquis – Python : Etudes de cas - Quiz - Exercices - Projet groupé	2	14
Système	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.	3	21
Scripting	Linux - Programmation Shell Bash : Décrire les notions de bases de Linux, de manière approfondie - Enrichir votre corpus de commandes - Développer des scripts Shell.	3	21
Travaux Pratiques	Mises en pratique et Validation des acquis - Infrastructure Linux : Etudes de cas - Quiz - Exercices - Projet groupé	2	14

Fondamentaux	Langage SQL et NoSql : Fondamentaux : Décrire les principaux concepts des SGDBR et Non Relationnelles (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	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
Big Data	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 œuvre du Big Data.	1	7
Big Data	Big Data - Concevoir et piloter un projet : Bien identifier les challenges dans la mise en place d'un projet Big Data Mesurer l'impact politique et stratégique sur l'entreprise et la pertinence d'un reporting régulier Mettre en place une bonne gestion des conflits et de la conduite du changement.	2	14
Big Data	Hadoop - Introduction à la plateforme et à l'écosystème : Installer un environnement Hadoop - Utiliser les distributions Hortonworks, MapR et Cloudera - Réaliser des algorithmes de calcul distribué avec MapReduce - Comprendre les concepts principaux autour du stockage et du traitement des données dans Hadoop - Appréhender l'écosystème basés sur Hadoop (HBase, les outils de traitement Pig, Hive et Spark).	1	7
Big Data	Hadoop - Développement : Présenter les principes du Framework Hadoop - Utiliser la technologie MapReduce pour paralléliser des calculs sur des volumes importants de données - Identifier les commandes shell courantes pour HDFS.	2	14
Big Data	PySpark - Traitement des données : Décrire le principe de fonctionnement de Spark - Utiliser l'API PySpark pour interagir avec Spark en Python - Mettre en œuvre les méthodes de Machine Learning avec la librairie MLlib de Spark - Traiter les flux de données avec Spark Streaming - Manipuler les données avec Spark SQL.	4	28
Travaux Pratiques	Mises en pratique et Validation des acquis - Big Data : Etudes de cas - Quiz - Exercices - Projet groupé	1	7
Immersion	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
Data Visualisation	Data Visualisation - Rendre visible l'invisible : Analyser, trier, traiter des données brutes pour les présenter sous forme de graphiques, de cartes ou d'organigrammes attrayants - Simplifier la lisibilité et accroître la compréhension.	1	7

Data Visualisation	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.	4	28
Data Visualisation	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
Data Science	Data Science - Les fondamentaux : Identifier les enjeux de la Data Science et de l'IA - Schématiser le cycle d'un projet Data Science - Appliquer les meilleures pratiques en matière de nettoyage et de préparation de vos données avant l'analyse - Modéliser un problème de Data Science.	2	14
Travaux Pratiques	Mises en pratique et Validation des acquis - Data Visualisation et Data Science : Etudes de cas - Quiz - Exercices - Projet groupé	1	7
CI / CD	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 conteneurs et du CaaS - Entamer une démarche vers une organisation DevOps	2	14
CI / CD	Gérer le versionning avec GIT et GitLab et appliquer le TDD avec PyTest : Découverte de la plateforme Github pour le travail collaboratif sur Git - Décrire les principes d'un gestionnaire de versions distribué - Identifier par la pratique, la philosophie de Git et ses apports - Mettre en œuvre les principes fondamentaux et les bonnes pratiques du TDD - Installer le framework PyTest - Gérer les tests logiciels avec PyTest sur un projet Agile.	6	42
Travaux Pratiques	Mises en pratique et Validation des acquis - CI / CD : Etudes de cas - Quiz - Exercices - Projet groupé	1	7
Automatisation et Déploiement	Docker pour Linux - Déploiement de conteneurs virtuels : Identifier les caractéristiques d'un conteneur Linux - Installer et utiliser Docker - Décrire la création des images Docker et "Dockerfile" - Interagir avec le Docker Hub et registry privés - Lister les notions réseaux de Docker (networks, links) - Expliquer la gestion des données avec Docker (volumes) - Utiliser Docker Swarm, Docker Compose et Docker Machine - Intégrer Docker au sein de vos projets - Déployer Docker en production.	4	28

Automatisation et Déploiement	Kubernetes - Orchestrer ses conteneurs : Décrire les principes de l'orchestration de conteneurs Docker - Créer et mettre en œuvre Kubernetes - Orchestrer des conteneurs Docker - Déployer des applications.	4	28
Travaux Pratiques	Mises en pratique et Validation des acquis – Kubernetes : Etudes de cas - Quiz - Exercices - Projet groupé	1	7
Google Cloud Platform	Google Cloud Platform - Core infrastructure - Les fondamentaux : Démontrer l'utilité et les atouts des produits et services Google Cloud Platform - Interagir avec les services Google Cloud Platform - Décrire les manières dont les clients utilisent Google Cloud Platform - Choisir et utiliser des environnements de déploiement applicatif sur Google Cloud Platform : Google App Engine, Google Kubernetes et Google Compute Engine - Gérer les options de stockage Google Cloud Platform : Google Cloud Storage, Google Cloud Filestore, Google Cloud SQL, Google Cloud BigTable et Google Cloud Data Store - Décrire les principes de base d'utilisation de BigQuery, l'entrepôt de données géré par Google et destiné à l'analyse - Présenter les principes de base d'utilisation de Cloud Deployment Manager, l'outil de Google permettant de créer et gérer des ressources Cloud à l'aide de modèles - Identifier les principes de base d'utilisation de Google Stackdriver, le système de surveillance, de journalisation et de diagnostic de Google.	3	21
Google Cloud Platform	Google Kubernetes Engine – Architecting : Décrire le déploiement et la gestion d'applications conteneurisées sur Google Kubernetes Engine (GKE) et les autres services fournis par Google Cloud Platform (GCP) - Explorer et déployer des éléments de solution, notamment des composants d'infrastructure tels que des pods, des conteneurs, des déploiements et des services (ainsi que des réseaux et des services d'application) - Déployer des solutions, notamment la gestion de la sécurité et des accès, la gestion des ressources et la surveillance des ressources - Décrire le fonctionnement des conteneurs de logiciels - Justifier les choix GCP pour les services de stockage managés - Surveiller les applications exécutées dans Kubernetes Engine - Illustrer l'architecture de Kubernetes - Concevoir l'architecture de GCP - Expliquer le fonctionnement de la mise en réseau des pods dans Kubernetes Engine - Créer et gérer des clusters de Kubernetes Engine à l'aide de la console GCP et des commandes gcloud / kubectl - Lancer, annuler et exposer des jobs dans Kubernetes - Gérer le contrôle d'accès à l'aide de Kubernetes RBAC et Google Cloud IAM - Gérer les stratégies de sécurité des pods et des réseaux - Utiliser des "secrets" et des "ConfigMaps" pour isoler les informations d'identification de sécurité et les artefacts de configuration.	2	21
Programmation	Python TensorFlow – Keras : Comprendre les bases du Machine Learning et du Deep Learning. Utiliser TensorFlow et Keras pour créer, entraîner et évaluer des modèles. Manipuler des données avec NumPy, Pandas et scikit-learn pour les préparer à l'entraînement. Résoudre des cas concrets de classification, de régression et de reconnaissance d'image.	3	21

Google Cloud Platform	Big Data and Machine Learning Fundamentals on Google Cloud Platform : Connaissance des produits et services de GCP notamment ceux liés aux traitements des données et au machine learning - Connaissance des produits et services fondamentaux concernant le calcul et le stockage - Connaissance de Cloud SQL et de Dataproc - Connaissance de Datalab et BigQuery - Connaissance de TensorFlow et des APIs machine learning (ML) - Connaissance de Pub/Sub et de Dataflow.	2	14
SQL	BigQuery : Décrire de manière approfondie les techniques du langage SQL afin de mieux tirer profit des possibilités du langage dans la consultation et la manipulation des données.	3	21
Google Cloud Platform	Data Engineering on Google Cloud Platform : Concevoir et développer des systèmes de traitement des données sur Google Cloud - Traiter des données par lot ou par flux en mettant en œuvre des pipelines de données d'autoscaling sur Dataflow - Orchestrer le travail entre les services GCP avec Cloud Composer / Apache Airflow - Obtenir des insights métier à partir d'ensembles de données extrêmement volumineux à l'aide de BigQuery - Exploiter des données non structurées à l'aide de Spark et des interfaces de programmation de ML sur Dataproc - Obtenir des insights immédiats à partir de flux de données - Découvrir les API de machine learning (ML) et BigQuery ML, et apprendre à utiliser Cloud AutoML.	5	35
Google Cloud Platform	Analyse et visualisation des données dans Looker : Définir Looker et les fonctionnalités qu'il fournit pour travailler avec des données - Expliquer les quatre concepts analytiques de base dans Looker (dimensions, mesures, filtres, pivots) - Utilisez des dimensions, des mesures, des filtres et des pivots pour analyser et visualiser les données - Créez instantanément des métriques avancées avec des calculs de table - Créer des tableaux de bord pour combiner et partager des visualisations - Utilisez des dossiers et des tableaux dans Looker pour organiser le contenu pour la navigabilité et la découvrabilité.	3	21
Soutenance	Présentation des choix techniques et des solutions proposées en binôme : 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.	2	14

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 où 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

