

Programme de la formation

Data Engineer IA — Concevoir et industrialiser des pipelines de données avec Python, SQL, le Cloud et l'IA Formation tutorée à distance 135 heures Tout public, Salarié en poste, Demandeur d'emploi, Particulier, Indépendant et dirigeant, Etudiant, Jeune en insertion professionnelle · De 1 à 5 participants À partir de 3 990 € Net de taxe par participant Éligible au CPF (Code : <https://www.francecompetences.fr/recherche/rncp/40792>) Formation certifiante [RNCP40792](#) gérée par France Compétences**Description courte**

Ce parcours opérationnel forme des Data Engineers juniors capables de concevoir, développer, tester, déployer et maintenir une application data : ingestion de données, persistance relationnelle et non relationnelle, transformations automatisées, exposition sécurisée par API et exploitation sur le cloud.

La stack pédagogique principale est volontairement cohérente : Python et FastAPI pour le développement applicatif, PostgreSQL, SQLAlchemy pour la persistance, MongoDB pour les données semi-structurées, Docker pour la portabilité, Airflow et dbt Core pour l'orchestration et la transformation, BigQuery et Google Cloud Storage pour le cloud, GitHub Actions pour l'intégration continue, pytest et les tests dbt pour la qualité, OpenAPI/Swagger pour la documentation. L'IA générative est utilisée comme assistant de développement sous contrôle humain.

Le fil rouge de la formation est **LogiFlow**, une entreprise de logistique qui modernise son système d'information. Le participant intervient comme Data Engineer junior : il modélise les données, construit un pipeline fiable, développe une API, déploie une solution cloud, diagnostique des incidents et fait évoluer l'application. Chaque module produit une preuve exploitable dans le portfolio final.

La formation prépare au bloc **RNCP40792BC03 — Conception et développement d'applications**, avec un projet centré sur une application data maintenable et sur les compétences de conception, développement, gestion des données, tests, documentation et maintenance corrective ou évolutive.

Profil des participants

Cette formation s'adresse en priorité aux :

- **Data Analysts** souhaitant évoluer vers un rôle plus technique.
- **Développeurs logiciels** souhaitant se spécialiser dans les applications orientées données et les pipelines.
- **Demandeurs d'emploi** visant un poste de Data Engineer junior, développeur back-end data ou ingénieur data cloud junior.

- **Salariés en poste** issus des bases de données, du développement, du DevOps ou des systèmes d'information.
- **Professionnels en reconversion** disposant déjà de bases en programmation ou en systèmes d'information.

Prérequis

- Bases en programmation, idéalement en Python ; une consolidation structurée est prévue au Module 1.
- Notions de bases de données relationnelles et de SQL recommandées.
- Aisance avec l'environnement informatique : installation de logiciels, fichiers, terminal et outils collaboratifs.
- Capacité à lire une documentation technique simple en anglais.
- Projet professionnel cohérent avec le développement, les systèmes d'information ou l'ingénierie de la donnée.
- Matériel informatique et connexion internet permettant l'utilisation d'environnements de développement et de services cloud.

Objectifs

À l'issue du parcours, les participants seront capables de :

- Analyser un besoin métier et concevoir l'architecture d'une application data maintenable.
- Développer des composants Python et une API FastAPI versionnés avec Git.
- Modéliser, interroger, administrer et faire évoluer des bases PostgreSQL et MongoDB.
- Construire des pipelines ETL/ELT idempotents et orchestrés avec Airflow et dbt Core.
- Déployer une chaîne data conteneurisée sur Google Cloud Storage et BigQuery.
- Mettre en place des tests unitaires, d'intégration et de qualité des données ainsi qu'une chaîne CI avec GitHub Actions.
- Diagnostiquer un incident, réaliser une maintenance corrective ou évolutive et mettre à jour les tests et la documentation.
- Documenter l'application, l'API, les modèles de données, le lineage et les procédures d'exploitation.
- Utiliser l'IA générative pour assister le développement et la documentation en vérifiant systématiquement ses productions.
- Présenter et justifier une solution applicative data devant un évaluateur.

Aptitudes et compétences visées / attestées

À l'issue de la formation, le candidat sera capable de :

- **Concevoir une solution applicative data** : analyser le besoin et le contexte, modéliser les données, définir une architecture, sélectionner des composants et formaliser les choix techniques.
- **Développer les composants d'une application** : programmer en Python, utiliser FastAPI, organiser le code, gérer les échanges web et intégrer les composants dans un dépôt Git.
- **Gérer les données** : concevoir ou adapter une base relationnelle ou non relationnelle, utiliser les langages de requête, administrer et déployer une base, gérer les migrations, transactions, droits, sauvegardes et restaurations.
- **Industrialiser des pipelines** : ingérer, transformer et charger des données de façon idempotente, gérer les chargements complets et incrémentaux, les reprises, la déduplication, l'évolution de

schéma et les contrats de données.

- **Garantir la qualité de la solution** : préparer les environnements, écrire et exécuter des tests unitaires, d'intégration et de qualité de données, puis automatiser les contrôles avec l'intégration continue.
- **Déployer et exploiter l'application** : conteneuriser les services, déployer les données et traitements sur GCP, gérer les secrets, les journaux, les droits d'accès et les coûts essentiels.
- **Assurer la maintenance corrective et évolutive** : qualifier un incident, diagnostiquer sa cause, corriger ou faire évoluer l'application et son schéma, réaliser les tests de non-régression et actualiser la documentation.
- **Documenter et communiquer** : produire une documentation technique et utilisateur, une spécification OpenAPI, des schémas, un lineage, des procédures d'exploitation et présenter les choix réalisés.
- **Mobiliser l'IA de manière responsable** : utiliser un assistant IA pour accélérer l'analyse, le code, les tests et la documentation, sans déléguer la validation technique, la sécurité ni la décision finale.

Contenu

La formation repose sur un accompagnement individuel et se déroule sur 5 mois. Elle combine **112 heures de formation tutorée à distance et d'activités pédagogiques** avec **23 heures de mentorat individuel** en visioconférence. Le rythme indicatif est de 7 heures par semaine, adaptable dans le respect du calendrier contractualisé.

Un projet fil rouge: le participant conçoit puis industrialise une application data pour une entreprise de logistique. Chaque livrable alimente un portfolio final démontrant la conception, le développement, la gestion des données, les tests, le déploiement, la documentation et la maintenance.

MODULE 1 — Fondations du Développement Applicatif Data : Python, Git, API et Environnement

MODULE 2 — Persistance et Gestion des Données : PostgreSQL, SQLAlchemy et MongoDB

MODULE 3 — Pipelines ETL/ELT, Orchestration et Transformation

MODULE 4 — Cloud Data, Déploiement et Sécurisation sur GCP

MODULE 5 — Industrialisation, Qualité, Maintenance et IA Appliquée au Data Engineering

MODULE 6 — Projet Final et Préparation au Bloc RNCP40792BC03

Pédagogie et organisation

Ce dispositif garantit une montée en compétences progressive, structurée et directement applicable aux enjeux des entreprises.

La formation repose sur une **pédagogie par la pratique**. Le projet fil rouge relie les apprentissages : le participant analyse un besoin, développe les composants, gère les données, automatise les traitements, déploie la solution, traite un incident et réalise une évolution. Chaque production est revue individuellement et intégrée au portfolio.

Elle est proposée **100 % à distance** et comprend **112 heures de formation tutorée à distance et d'activités pédagogiques** : ressources structurées, démonstrations, exercices, études de cas, travaux

pratiques, évaluations et réalisation des projets. Elle comprend également **23 heures de mentorat individuel** en visioconférence avec un expert Data Engineering, réparties sur les 5 mois.

La formation intègre en bonus un accompagnement à l'employabilité : structuration du portfolio, présentation des projets, optimisation du profil professionnel et préparation aux entretiens techniques.

Positionnement et Modalités d'évaluation des acquis

- **Positionnement** : Entretien de positionnement obligatoire avant l'entrée en formation pour vérifier les attentes et les prérequis, valider le profil et définir le parcours individualisé.
- **Évaluations formatives** : QCM et auto-évaluations par module, exercices corrigés avec feedback personnalisé du mentor.
- **Évaluations sommatives** : Mini-projets avec soutenance orale à l'issue de chaque module.
- **Certification** : La formation prépare au bloc de compétences RNCP40792BC03 — Conception et développement d'applications. L'évaluation du bloc est organisée par l'organisme certificateur. Le candidat peut se présenter en candidat libre, selon les modalités définies par le certificateur. Une procédure détaillée est fournie en fin de formation pour accompagner l'apprenant dans son inscription.

Mode de validation

- **Attestation de fin de formation** délivrée par MyDataMentor à l'issue du parcours.
- **Projet final** présenté en soutenance devant le formateur.
- La formation prépare au bloc de compétences RNCP40792BC03 — Conception et développement d'applications
- L'évaluation du bloc est organisée par l'organisme certificateur (Ministère de l'enseignement supérieur et de la recherche).
- Le candidat peut se présenter à la certification en tant que **candidat libre**. Une procédure détaillée est transmise en fin de formation.
- Seul le Ministère de l'enseignement supérieur et de la recherche est habilité à délivrer le bloc de compétences.

Financements possibles

Cette formation est éligible à plusieurs dispositifs de financement, sous réserve des critères propres à chaque organisme :

- **Financement public** : CPF (Mon Compte Formation), OPCO (selon branche professionnelle — particulièrement adapté au financement via plan de développement des compétences entreprise), AGEFICE (TNS du commerce), FIF-PL (professions libérales), France Travail (demandeurs d'emploi).
- **Financement entreprise** : Prise en charge possible dans le cadre du plan de développement des compétences de l'employeur — devis sur demande pour les entreprises souhaitant former plusieurs collaborateurs.
- **Financement personnel** : Possibilité d'échelonnement et de facilités de paiement pour les apprenants finançant la formation à titre individuel..

Modalités et délai d'accès

- **Inscriptions** : Ouvertes en continu tout au long de l'année.

Siège social 144 rue Paul Bellamy – 44000 NANTES

SAS au capital de 100 000 € - SIRET 452 558 893 00049 – 02 51 84 95 55 – contact@cadresenmission.com
Numéro d'enregistrement de formation : 52 44 06368 44 auprès du Préfet de la Région Pays de la Loire

- **Duree & Rythme** : 135 heures en distanciel sur 5 mois, soit un rythme indicatif de 7 heures par semaine.
- **Délais d'accès** : Démarrage possible sous 14 jours après validation administrative et entretien de positionnement. Planning adaptable et individualisé dès l'entrée en formation.

Référent handicap

PERSONNES EN SITUATION DE HANDICAP : (*mise en place de compensation...*)

Contact : Référent handicap 02 51 84 95 55 / 06 28 70 45 28 / nadiahadjeri@cadresenmission.com

Accessibilité aux personnes en situation de handicap

Personnes en situation de handicap : On s'engage pour l'inclusion de tous les apprenants. Un protocole d'adaptation aux modalités pédagogiques et de passation à la certification est mis en place pour chaque situation individuelle (aménagement du temps, supports adaptés, accompagnement renforcé).

Contactez notre référent handicap pour toute demande d'aménagement avant l'entrée en formation.

Intervenant

Manel Boumaiza — Ingénieure experte en Data Engineering, Data Analytics, Business Intelligence et Intelligence Artificielle appliquée au pilotage de la performance. Forte de plus de 15 ans d'expérience en pilotage de projets et optimisation des performances en entreprise, ancienne responsable innovation en R&D, elle conduit des initiatives stratégiques mêlant agilité, data et amélioration des processus. Formatrice et mentore, elle accompagne professionnels et organisations dans le développement de compétences en data, pilotage et management orienté données, avec une approche structurée, opérationnelle et orientée résultats.