

Le Campus Numérique in the Alps

Formation Data Engineer

Objectifs de la formation

Parcours de 10 à 18 mois destiné à des personnes en recherche d'emploi souhaitant se reconvertir vers le métier Data Engineer, et développer des compétences avancées en conception et gestion de systèmes permettant de collecter, stocker, et traiter de grandes quantités de données.

Le data engineer est l'architecte des données massives et du développement de l'intelligence artificielle et a pour activités principales de :

- **Développer des solutions sécurisées pour collecter, nettoyer, traiter, stocker et exploiter des données**
- **Déployer et maintenir une solution opérationnelle de gestion de données (*data warehouse*)**
- **Concevoir, mettre en œuvre et optimiser des modèles d'intelligence artificielle / de machine learning**
- **Piloter la conduite de projets data au sein d'une entreprise.**

Compétences développées

Le programme aborde les technologies et compétences indispensables à la collecte, au stockage et à l'analyse de données : algorithmique, Python, Machine Learning et data mining, bases de données, outils de visualisation des données, outils statistiques… et les soft skills nécessaires en entreprise. Les méthodes d'apprentissage « apprendre à apprendre » permettent aux apprenants de se former en autonomie tout au long de leur carrière pour suivre les évolutions technologiques.

Compétences développées à l'issue de la formation :

- Programmation appliquée à la collecte et la manipulation de données
- Gestion de bases de données
- Architecture des données de différentes sources
- Utilisation d'outils et plateformes big data
- Développement et administration de services cloud
- Mise en place de pratiques de sécurisation des données
- Automatisation des workflows de traitement de données
- Développement de modèles prédictifs d'intelligence artificielle

Méthodes pédagogiques

Le Campus adopte une approche par compétences, les modules de formation étant construits sur la base de compétences à acquérir : savoirs, savoir-faire, savoir-être, outils, méthodologies… Cette approche permet d'assurer une formation professionnalisante avec le développement de compétences opérationnelles correspondant aux exigences des métiers visés.

Le modèle de pédagogie par projet mis en place dans les modules de formation permet de rendre les stagiaires acteurs de la formation et de l'apprentissage. L'autonomie développée dans les modules (e-learning et classes inversées), la transversalité marquée des enseignements et une pratique pédagogique hautement professionnalisante (développement d'applications dès les premières heures passées au Campus et apprentissage en alternance) sont au cœur du projet pédagogique.

Modalités d'évaluation et certification visée

La formation est sanctionnée par l'obtention Certificat de Qualification Professionnelle Statisticien - Data Analyste pour l'industrie (certificat n°2017 0322) de niveau 7 (équivalent bac+5) de l'Union des Industries et Métiers de la Métallurgie (UIMM) (modifié le 06/10/2022). La formation prépare à l'ensemble des blocs de compétence du certificat.

Modalités d'évaluation :

- **Validation des compétences** (auto validation et validation par un formateur professionnel de l'ensemble des compétences associées à chacun des modules de formation)
- **Réalisation de rapports d'activité en entreprise**
- **Réalisation d'un rapport de mission professionnelle**
- **Soutenance devant un jury mixte (représentants de l'organisme de formation et professionnels du métier représentant les entreprises).**

Durée de la formation

La formation se décompose en deux phases :

- **6 mois de formation intensive au Campus (805 heures) permettant d'acquérir le socle de compétences techniques et de travail en équipe**
- **Mission en entreprise de 4 à 12 mois d'élargir et d'approfondir les compétences techniques et les soft skills développées (stage ou alternance).**

Modalités d'accès et prérequis

Tous les candidats de niveau bac+5 possédant une première expérience en développement informatique (formation certifiante ou professionnalisante, connaissance d'au moins un langage de programmation) et étant demandeurs d'emploi inscrits à France Travail peuvent s'inscrire à la formation. Les critères de sélection sont basés sur :

- **Motivation, cohérence et clarté du projet professionnel**
- **Résultats au test de positionnement informatique**
- **Aptitudes métier, qualités professionnelles et aptitudes au travail en équipe**

Processus de sélection :

- **Envoi d'un CV, d'une lettre de motivation et d'une vidéo pitch**
- **Passage de test de positionnement informatique**
- **Projet collectif non numérique (piscine de 2 à 3 jours)**

Délais d'accès : la formation débute en moyenne deux mois après l'ouverture des candidatures.

Prix de la formation

Gratuit grâce aux partenaires financeurs de la formation et les fonds de financement de la formation.

Lieux de formation

Campus de Grenoble

Le Totem
16 boulevard Maréchal Lyautey
38000 GRENOBLE
07 50 66 51 51 / audrey.graffagnino@le-campus-numerique.fr

Campus de Valence

M3 Rovaltain
1 rue Roland Moreno
26300 ALIXAN
06 21 00 66 54 / contact.valence@le-campus-numerique.fr

Campus d'Annecy

Papeteries Image Factory
3 Esplanade Augustin Aussedat – Cran Gevrier
74960 ANNECY
07 69 37 90 37 / carole.ritrovato@le-campus-numerique.fr

Accessibilité aux personnes en situation de handicap

La formation est accessible aux personnes en situation de handicap, avec une adaptation possible du rythme pédagogique et des modalités d'évaluation. Le Campus Numérique est engagé dans la démarche H+ Formation de la Région Auvergne Rhône-Alpes.

Suites de parcours

A visée professionnalisante, la formation vous permet d'exercer le métier de Data Engineer.

Programme Pédagogique

805 heures		
Positionnement	3.5 heures	Bilan, diagnostic et définition du parcours individualisé de formation
Outils numériques pour formation à distance	3.5 heures	Mise en place de l'environnement de formation à distance, formation sur les outils numériques utilisés à distance (installation de logiciels, vérification de connexion internet, tests techniques)
Bloc 1 : Solutions de gestion de données		
Système d'exploitation	21 heures	**Objectifs pédagogiques :** Maîtrise du système d'exploitation open source Linux et construction et manipulation de bases de données. Savoir gérer les processus, la mémoire, les droits et les fichiers, avec une initiation à l'usage de la ligne de commande sous Linux
Introduction à l'écosystème data et architecture des systèmes	42 heures	**Objectifs pédagogiques :** Architecture des systèmes d'information, les flux de données, les types de bases de données (relationnelles et NoSQL), l'infrastructure (serveurs, cloud). Étudier les protocoles de communication des systèmes (REST, SOAP) et les modèles de données.
Python	35 heures	**Objectifs pédagogiques :** Maîtriser les bases de la programmation à l'aide de Python (variables, instructions, conditions, boucles, fonctions, structures de données, programmation orientée objet). Maîtriser la gestion de dépendances en Python (environnements virtuels).
Algorithmique – langage JAVA	35 heures	**Objectifs pédagogiques :** Maîtriser les concepts algorithmiques de la programmation (variables, opérateurs, conditions, boucles et tableaux). Connaître la méthodologie de conception d'une solution informatique. Maîtriser le langage JAVA.
Bases de données	35 heures	**Objectifs pédagogiques :** Maîtrise des systèmes de gestion de base de données (SGBD). Gestion de bases de données relationnelles, modélisation et interrogation via SQL. Comprendre les concepts clés (schéma, tables, relations, jointures, clés) et optimiser les requêtes.
Administration des bases de données et cybersécurité	35 heures	**Objectifs pédagogiques :** Maîtriser les techniques de la cybersécurité pour les bases de données : gestion des accès

		(authentification, autorisation), cryptage des données, sauvegardes sécurisées, détection des intrusions, gestion des risques et systèmes de sécurité.
Intégration de données et orchestration	70 heures	**Objectifs pédagogiques :** Concevoir, développer et déployer des processus ETL en utilisant des outils comme Apache Airflow, Talend, et des scripts Python pour automatiser l'extraction, la transformation et le chargement des données.
RGPD	7 heures	**Objectifs pédagogiques :** Compréhension du RGPD, les différents principes (minimisation de la collecte, finalité, consentement, conservation limitée des données, sécurité et confidentialité des données), les enjeux du RGPD pour l'analyse de données.
Bloc 2 – Big data		
Déploiement et gestion de l'infrastructure data	70 heures	**Objectifs pédagogiques :** Utiliser Docker pour la conteneurisation des applications et Kubernetes pour l'orchestration des conteneurs. Appréhender les services cloud tels que AWS, GCP, Azure pour l'hébergement des solutions.
Big Data et architectures distribuées	70 heures	**Objectifs pédagogiques :** Maîtriser les technologies Big Data comme Hadoop, Spark, et utiliser des architectures Lambda et Kappa pour le traitement en batch et temps réel. Optimiser le stockage avec HDFS et gérer les clusters via YARN.
Industrialisation des pipelines et CI/CD	70 heures	**Objectifs pédagogiques :** Mettre en place des processus CI/CD avec des outils comme Jenkins, GitLab CI, pour automatiser les déploiements des pipelines. Maîtrise de l'orchestration des jobs (airflow, nifi), la scalabilité des transformations (spark, dbt,..) et la sécurité.
Bloc 3 – ML engineering		
Introduction au Machine Learning et aux modèles prédictifs	77 heures	**Objectifs pédagogiques :** Apprendre les bases du Machine Learning supervisé et non supervisé avec des algorithmes classiques (régression linéaire, SVM, k-means). Utiliser scikit-learn pour développer des modèles prédictifs et appliquer des méthodes de validation
Déploiement des modèles	35 heures	**Objectifs pédagogiques :** Apprendre à déployer des modèles en production avec des outils comme MLflow et TensorFlow Serving. Choix du format de

		modèle (onnx, mlmodel,...) en fonction du type de déploiement (cloud, as a service ou embarqué).
IA frugale et Green IT	14 heures	**Objectifs pédagogiques :** Impacts environnementaux et sociétaux du numérique. Leviers de réduction de l'empreinte carbone des projets data (stockage, traitement, cloud...) Les principes de l'IA frugale : optimisation, consommation énergétique, réutilisation des algorithmes.
Bloc 4 - Exploitation des données et compétences transversales		
Analyse avancée des données et visualisation	70 heures	**Objectifs pédagogiques :** Appliquer des techniques d'analyse avancée des données (statistiques multivariées, analyse en composantes principales) et utiliser des outils de visualisation comme Power BI, Tableau et Plotly pour créer des rapports et des dashboards interactifs.
Méthodes agiles	28 heures	**Objectifs pédagogiques :** Utilisation des méthodes de gestion de projet Agile et des outils de travail collaboratifs dans une équipe gestion de projet data.
GIT	7 heures	**Objectifs pédagogiques :** Maîtriser les bases de Git pour versionner, suivre et collaborer efficacement sur des projets de code. Savoir utiliser les commandes essentielles en ligne de commande : init, commit, push, pull, branch, merge, et résoudre des conflits.
TRE	21 heures	**Objectifs pédagogiques :** Maîtriser les techniques de recherche d'emploi, les techniques de création d'un profil professionnel numérique ainsi que les stratégies nécessaires pour s'engager dans un processus actif de recherche d'emploi.
Gestion de relation clients / usagers	11 heures	**Objectifs pédagogiques :** Étude de cas - Maîtriser les étapes et les techniques de la communication client/usager. Identifier les besoins, construire un cahier des charges à partir des demandes client/usager et traduire des résultats d'analyse en recommandations compréhensibles et utiles.
Laïcité, citoyenneté et valeurs de la République	4 heures	**Objectifs pédagogiques :** Comprendre les principes de la laïcité, de la citoyenneté et des valeurs de la République pour adopter une posture professionnelle respectueuse, favoriser le vivre-ensemble et

		réagir de manière adaptée aux situations rencontrées.
Préparation à la certification	35 heures	Préparation à l'examen de certification RNCP 37624 – Data Engineer : format de l'examen, des types de questions, et des attentes spécifiques. Méthodologies de construction de la présentation orale et du rapport professionnel. Soutenances blanches
Bilans	6 heures	**Bilan intermédiaire : point d'étape sur les apprentissages et l'évolution des apprenants**, identification des difficultés éventuelles et ajustement des méthodes pédagogiques. **Bilan final** : **évaluation de l'ensemble du parcours des apprenants et mesure de l'atteinte des objectifs pédagogiques de la formation. Bilan des taux d'insertion dans l'emploi des stagiaires (départs en stage et perspectives d'emploi au terme de la formation).** Recueil des impressions des participants sur le contenu, les méthodes et l'animation de la formation, dans une dynamique d'amélioration continue.

En partenariat avec