

Data engineer

RNCP37624 · RNCP · Publiée

L'essentiel

Niveau de qualification	Niveau 7
Code(s) NSF	326
Formacode(s)	31052
Code(s) ROME	M1803, M1806, M1805
Type d'enregistrement	Enregistrement sur demande
Date d'enregistrement	31/05/2023
Date d'échéance	31/05/2026

Certificateur(s)

LUNALOGIC GROUP · SIRET : 80037898600015

STEPHENSON FORMATION · SIRET : 30975404200039

Partenaires habilités

10 organismes habilités, d'après la liste publiée par France compétences : 9 à former et organiser l'évaluation et 1 à former.

Organisme	SIRET	Habilitation
AUTUNOIS MORVAN DEVELOPPEMENT FORMATION	51520053300021	Former et organiser l'évaluation
AXI TECHNOLOGIES	88762499700011	Former et organiser l'évaluation
ENSUP	35119949200114	Former et organiser l'évaluation
ENSUP	35119949200122	Former et organiser l'évaluation
ENSUP MARSEILLE	97851728200015	Former et organiser l'évaluation
FORMASUP NORD-PAS DE CALAIS	42813525500050	Former
GROUPE PARIS ECOLE DE MANAGEMENT	48888299400046	Former et organiser l'évaluation
JUNIA XP	89085431800015	Former et organiser l'évaluation
JUNIA XP	89085431800049	Former et organiser l'évaluation
LE CAMPUS NUMERIQUE IN THE ALPS.	83186635500017	Former et organiser l'évaluation

Résumé de la certification

OBJECTIFS ET CONTEXTE

La data est devenue un sujet stratégique avec le développement des outils permettant de « dématérialiser » les données et ainsi de les rendre plus facilement « partageables ». L'exemple le plus basique est le logiciel de bureautique qui a permis de passer de tableaux de données papier d'une expérience, à un tableau de données numériques de cette même expérience. La donnée est ainsi devenue une data : elle a été numérisée. En une petite quinzaine d'années, à peine, le développement constant d'outils informatiques permettant de numériser toute sorte de données (images, sons, textes, vidéos, ...) et l'avènement d'internet ont non seulement permis d'accélérer la numérisation des données, mais aussi leur production en masse ; les activités économiques et sociales de toute

nature (vente en ligne, services, formation), ayant elles aussi migré vers l'internet. Ainsi l'accroissement de la quantité de données produites par les entreprises, les particuliers, les organisations, ... est exponentiel. Les chiffres de 2019 et les projections faites par les experts, le démontrent. Ainsi, « Selon les estimations publiées dans le Digital Economy Compass 2019 de Statista, le volume annuel de données numériques créées à l'échelle mondiale a été multiplié par plus de vingt au cours de la dernière décennie et devrait s'approcher de 50 zettaoctets cette année ». On parle désormais de big data, ou données massives, les volumes générés de data étant gigantesques.

Par ailleurs, parallèlement à l'augmentation de la production de données numériques, et pour pouvoir en faire l'exploitation, de nouveaux moyens et de nouvelles solutions ont été mis au point pour pouvoir les stocker de manière massive, rapide et économique. C'est l'émergence des data centers et des solutions « clouds ».

Comme le souligne le syndicat français du numérique, Syntec, « Ces données sont devenues le langage de nos activités. Pour les faire parler, il faut les recueillir, les formater, les analyser, les consolider, les stocker, les diffuser... Autant d'opérations regroupées dans ce que l'on appelle le big data. » et de poursuivre que « Tous les métiers du numérique, conseil, édition de logiciels, services, ingénierie, hébergement Cloud, sont concernés. Et tous les secteurs le sont aussi car les données sont devenues la matière première ou la production de toutes les activités. »

Ainsi les données produites, collectées et stockées sont ensuite exploitées pour produire des connaissances, des produits ou des services, dans tous les secteurs : santé, banque, assurance, automobile, formation, etc. Elles permettent d'analyser les comportements de clients, pour le marketing, de suivre l'évolution d'apprenants sur des plateformes et d'orienter la suite de leurs parcours, ou encore de faire des prévisions.

En pénétrant tous les secteurs et toutes les activités, le big data nécessite des compétences spécifiques et face aux besoins exponentiels des structures, qu'elles soient entreprises - de toute taille - ou encore services publics, les métiers de la data sont en explosion depuis quelques années. Le besoin est tel que le gouvernement français a créé en août 2020 une « Mission numérique des grands groupes » qui « vise à faire émerger des projets communs et concrets, qui bénéficieront à l'ensemble du tissu économique français, avec l'ambition de le faire accéder à un nouveau palier de transformation numérique, dont la crise a révélé l'urgente nécessité. » Pour y parvenir, le gouvernement lui a fixé cinq chantiers prioritaires dont « soutenir la transformation des compétences et des formations ».

Dans ce contexte le Data Engineer est

l'architecte des données massives et du développement de l'intelligence artificielle, qui sont des conditions de réussite sine qua none de la transformation numérique de notre société : le développement de ce métier clef constitue donc une opportunité.

ACTIVITÉS VISÉES

Etude et développement, ou choix et adaptation d'une solution sécurisée pour collecter, nettoyer, traiter, stocker et exploiter des données

Management de la transition data de l'entreprise

Déploiement ou industrialisation, maintenance d'une solution opérationnelle de gestion de données

Conception,

mise en œuvre et optimisation d'un modèle d'Intelligence Artificielle/Machine Learning (ML)

Exploitation de la donnée pour piloter l'activité de l'entreprise et créer de nouvelles opportunités

COMPÉTENCES ATTESTÉES

Analyser les besoins en intégrant les ressources disponibles, les contraintes, les risques afin de proposer une solution correspondant à la demande du client

Analyser le besoin de protection des données en mesurant leur sensibilité afin de définir une politique de sécurité et une gouvernance du Système d'Information (SI)

Identifier et mesurer les risques liés à la protection juridique de l'entreprise et de ses parties prenantes en respectant la législation des pays concernés en matière de chiffrement, détention, transit, domiciliation des prestataires afin de diminuer les risques de conflit avec une juridiction ou entre deux juridictions

Rédiger les cahiers des charges fonctionnel et technique de manière accessible en se basant sur l'analyse des besoins validée par le client, afin de spécifier les caractéristiques de la solution à développer ou à adapter

Concevoir l'architecture sécurisée de la solution à développer ou à adapter en se basant sur les différents cahiers des charges en vue de proposer un système opérationnel de gestion de données massives

Conduire et réaliser le développement ou l'adaptation d'une solution en ayant recours aux langages de programmation et aux technologies de base de données les plus adaptés et en garantissant la conformité aux cahiers des charges afin de livrer un système opérationnel, sécurisé, accessible, destiné à collecter, nettoyer, traiter, stocker et exploiter des données

Réunir et coordonner des compétences spécialisées nécessaires au développement de la solution en mettant en oeuvre les bonnes pratiques de gestion d'équipe, de gestion de projet et de prise en compte des éventuelles situations de handicap

Réaliser des tests en se basant sur des jeux de données et valider le bon fonctionnement et la sécurité de la solution développée ou adaptée en conformité avec les attentes du client

Contribuer à la rédaction d'une charte éthique (accessible à l'ensemble des collaborateurs) relative aux données manipulées, en déclinant les exigences sociétales et environnementales fixées par l'entreprise afin de respecter les engagements de celle-ci dans ces domaines

Collaborer à l'élaboration de la stratégie de transition data de l'entreprise, respectueuse de la charte éthique, en identifiant ses métiers et activités concernés, en définissant le périmètre du projet de transition, ses objectifs, ses étapes, ses délais, le budget requis et les indicateurs de réussite

afin de soutenir le développement de l'entreprise

Proposer des évolutions du système de données de l'entreprise, en modifiant sa structuration et ses modes d'accès, en définissant des indicateurs de mesure, pour améliorer sa mise à disposition et son exploitation par les utilisateurs

Piloter et mettre en place dans son champ d'expertise la transition data de l'entreprise en faisant évoluer des process, des métiers, des activités, afin d'optimiser leurs fonctionnements

Planifier et gérer les projets relevant de la transition en définissant les objectifs, les indicateurs, les jalons, et les livrables, afin de respecter les délais convenus avec le client

Concevoir et mettre en exploitation des outils de suivi inclusifs à partir des données qu'il aura identifiées afin de piloter un projet pour détecter et corriger les écarts, les risques

Manager et assurer l'accompagnement des métiers et des acteurs de l'entreprise dans leur évolution en organisant ou contribuant à des actions de formation et d'acculturation accessibles afin de passer à une culture de la donnée respectueuse de la sécurité, du cadre juridique et de l'éthique

Évaluer le niveau d'adhésion des acteurs de l'entreprise à la démarche de transition data en mettant en place des indicateurs d'acceptation afin d'identifier d'éventuelles difficultés ou réticences

Mettre en place un dispositif de veille juridique et technologique expert portant sur les champs de la Data, du Big Data, de l'IA et de la cybersécurité en identifiant les sources d'informations les plus pertinentes afin de faire évoluer le Système d'Information (SI) et les solutions Data mises en place

Exploiter les informations issues de la veille juridique et technologique afin de proposer des évolutions du SI et de faire évoluer les solutions Data dont il a la responsabilité

Mobiliser et coordonner les expertises techniques en utilisant les méthodes de conduite de projet et d'animation d'équipe, en prenant en compte des éventuelles situations de handicap, afin de mettre en oeuvre le déploiement ou l'industrialisation et la maintenance d'une solution de gestion de données

Déployer ou industrialiser une solution dans le respect du cahier des charges en mobilisant les ressources techniques, budgétaires, humaines et en tenant compte des contraintes pour mettre à disposition un système de collecte maintenable, robuste, scalable et pouvant gérer des données massives

Mettre en place et superviser avec les services informatiques de l'entreprise, la protection de la solution et des données, en déployant les outils de sécurité, d'anonymisation, et en instaurant un mécanisme de supervision afin de consolider l'intégrité, la disponibilité, l'accessibilité, la confidentialité et la lisibilité des données

Mettre en exploitation la solution en monitorant la charge et la performance pour permettre la collecte, le nettoyage, le traitement, le stockage et l'exploitation des données

Attribuer et configurer les droits d'accès au système et aux données mises à disposition pour restreindre leur accessibilité aux personnes autorisées en se référant au cahier des charges du projet et à la politique de sécurité de l'entreprise

Evaluer la performance du système en effectuant des retours sur expérience par la collecte et la synthèse des informations relatives aux insuffisances et aux points d'amélioration potentiels en vue de son évolution

Conduire les évolutions du système en réponse aux points d'amélioration et insuffisances détectées lors de l'évaluation de ses performances afin de garantir un fonctionnement optimisé

Planifier et mettre en place une maintenance préventive du système de gestion de données afin d'assurer son bon fonctionnement et la disponibilité des données

Analyser expérimentalement les différents modèles d'Intelligence Artificielle/Machine Learning à travers des tests et des analyses statistiques afin de répondre aux besoins du client

Affiner le modèle d'Intelligence Artificielle/Machine Learning en cherchant les paramètres optimaux et en évaluant l'adéquation des modèles d'apprentissage avec le besoin exprimé afin de garantir ses capacités de prédiction

Déployer ou Industrialiser le modèle d'Intelligence Artificielle/Machine Learning à travers la mise en place des

pipelines de données et dans le respect du cadre réglementaire pour garantir la qualité, la fiabilité, la rapidité et la scalabilité de la solution

Évaluer la performance du modèle d'Intelligence Artificielle/Machine Learning déployé ou industrialisé en se basant sur les métriques de référence afin d'optimiser ce modèle

Communiquer ses résultats au client (interne/externe) de façon accessible (datavisualisation) pour démontrer la performance du modèle d'Intelligence Artificielle/Machine Learning

Recueillir les besoins des utilisateurs/décideurs par un processus de questionnement afin d'identifier les données internes à collecter et exploiter

Identifier les données externes au moyen d'un système de veille afin de les intégrer ou de les croiser aux données internes.

Collecter des données structurées, ou non structurées, en mobilisant des sources variées en prévision de leur stockage et de leur exploitation

Stocker des données structurées, ou non structurées, dans une base de données, un data lake, ou un entrepôt de données, en garantissant leur sécurité afin de les mettre à disposition des différents utilisateurs/décideurs

Extraire les données identifiées d'une base de données, d'un data lake ou d'un entrepôt de données, à l'aide d'outils adaptés afin de les mettre à disposition des différents utilisateurs/décideurs dans un format exploitable.

Nettoyer les données contenues dans une base de données, un data lake, ou un entrepôt de données en s'assurant de leur qualité et de leur utilisabilité pour les restituer en un ensemble exploitable, homogène ou hétérogène.

Créer des outils de pilotage accessibles (tableaux de bord automatisés) en identifiant les indicateurs clés pour permettre aux utilisateurs/décideurs de faire face aux évolutions de l'activité et prendre les décisions adaptées.

Présenter le fonctionnement des outils de pilotage et les indicateurs clés identifiés, aux parties prenantes de façon accessible (datavisualisation) afin d'en obtenir la validation et l'appropriation

Présenter et défendre ses recommandations en se basant sur l'analyse des informations issues des outils de pilotage, afin de permettre aux clients internes une prise de décision éclairée (data gouvernance) et de créer de nouvelles opportunités

Blocs de compétences

RNCP37624BC01 — Conduire et réaliser le développement ou l'adaptation d'une solution pour collecter, nettoyer, traiter, stocker et exploiter des données

LISTE DE COMPÉTENCES

Analyser les besoins en intégrant les ressources disponibles, les contraintes, les risques afin de proposer une solution correspondant à la demande du client

Analyser le besoin de protection des données en mesurant leur sensibilité afin de définir une politique de sécurité et une gouvernance du Système d'Information (SI)

Identifier et mesurer les risques liés à la protection juridique de l'entreprise et de ses parties prenantes en respectant la législation des pays concernés en matière de chiffrement, détention, transit, domiciliation des prestataires afin de diminuer les risques de conflit avec une juridiction ou entre deux juridictions

Rédiger les cahiers des charges fonctionnel et technique de manière accessible en se basant sur l'analyse des besoins validée par le client, afin de spécifier les caractéristiques de la solution à développer ou à adapter

Concevoir l'architecture sécurisée de la solution à développer ou à adapter en se basant sur les différents cahiers des charges en vue de proposer un système opérationnel de gestion de données massives

Conduire et réaliser le développement ou l'adaptation d'une solution en ayant recours aux langages de programmation et aux technologies de base de données les plus adaptés et en garantissant la conformité aux cahiers des charges afin de livrer un système opérationnel, sécurisé, accessible, destiné à collecter, nettoyer, traiter, stocker et exploiter des données

Réunir et coordonner des compétences spécialisées nécessaires au développement de la solution en mettant en oeuvre les bonnes pratiques de gestion d'équipe, de gestion de projet et de prise en compte des éventuelles situations de handicap

Réaliser des tests en se basant sur des jeux de données et valider le bon fonctionnement et la sécurité de la solution développée ou adaptée en conformité avec les attentes du client

MODALITÉS D'ÉVALUATION

Modalité d'évaluation des compétences 1 à 4 : Mise en situation d'analyse des besoins du client et rédaction du cahier des charges fonctionnel et du cahier des charges technique.

Le candidat (seul) doit, à partir d'une demande émanant d'un client réel ou fictif et présentant :

- l'entreprise,
- le contexte,
- la problématique à résoudre,
- les ressources disponibles,
- ...

fournir un dossier dans lequel il devra :

- 1 - Lister et hiérarchiser dans un document, les besoins, les ressources et les contraintes du client
- 2- Identifier les menaces/risques (techniques et juridiques) et les besoins de protection des données
- 3- Proposer une évolution de la politique de sécurité et de la gouvernance du SI
- 4- Rédiger en respectant les normes d'accessibilité les cahiers des charges fonctionnel et technique correspondant aux besoins et contraintes du client

Modalité d'évaluation des compétences 5 à 8 : Mise en situation de développement ou d'adaptation d'une solution.

Le candidat (en équipe) doit à partir d'un cahier des charges technique fourni par un client réel ou fictif :

- 1- Définir le schéma général de la solution permettant de collecter, nettoyer, traiter, stocker et exploiter des données de façon sécurisée
- 2 - Définir et mettre en oeuvre la méthodologie de projet et les modalités de pilotage
- 3- Définir un plan d'action incluant les jalons et répartir les missions en fonction des profils des membres de l'équipe en prenant en compte les éventuelles situations de handicap
- 4 - Développer ou adapter la solution préconisée dans le cahier des charges technique en utilisant les langages et les technologies de bases de données adaptés (la solution est respectueuse des normes d'accessibilité de type RGAA).
- 5 - Réaliser des tests pour vérifier et valider les fonctionnalités et la sécurité de la solution retenue correspondant aux spécifications du cahier des charges technique.

Chaque membre de l'équipe rédige un rapport qu'il présente individuellement lors d'une soutenance devant un jury de deux personnes représentant le client. Le rapport présente les éléments suivants, leur justification et leur analyse critique et personnelle :

- le schéma général de la solution sécurisée
- la méthodologie de projet choisie,
- les modalités de pilotage et les décisions majeures prises tout au long du projet, les difficultés rencontrées et les solutions mises en oeuvre
- le plan d'action et sa mise en oeuvre,
- la solution développée ou adaptée,
- les langages et technologies utilisés,
- les résultats des tests,

La présentation du rapport devant le jury intègre un temps de présentation et de démonstration du fonctionnement de la solution.

A la fin du rapport, le candidat réalise un bilan qui évalue la solution, son fonctionnement, sa sécurité au regard des tests et du cahier des charges.

RNCP37624BC02 — Manager la transition data de l'entreprise

LISTE DE COMPÉTENCES

Contribuer à la rédaction d'une charte éthique (accessible à l'ensemble des collaborateurs) relative aux données manipulées, en déclinant les exigences sociétales et environnementales fixées par l'entreprise afin de respecter les engagements de celle-ci dans ces domaines

Collaborer à l'élaboration de la stratégie de transition data de l'entreprise, respectueuse de la charte éthique, en identifiant ses métiers et activités concernés, en définissant le périmètre du projet de transition, ses objectifs, ses étapes, ses délais, le budget requis et les indicateurs de réussite afin de soutenir le développement de l'entreprise

Proposer des évolutions du système de données de l'entreprise, en modifiant sa structuration et ses modes d'accès, en définissant des indicateurs de mesure, pour améliorer sa mise à disposition et son exploitation par les utilisateurs

Piloter et mettre en place dans son champ d'expertise la transition data de l'entreprise en faisant évoluer des process, des métiers, des activités, afin d'optimiser leurs fonctionnements

Planifier et gérer les projets relevant de la transition en définissant les objectifs, les indicateurs, les jalons, et les livrables, afin de respecter les délais convenus avec le client

Concevoir et mettre en exploitation des outils de suivi inclusifs à partir des données qu'il aura identifiées afin de piloter un projet pour détecter et corriger les écarts, les risques

Manager et assurer l'accompagnement des métiers et des acteurs de l'entreprise dans leur évolution en organisant ou contribuant à des actions de formation et d'acculturation accessibles afin de passer à une culture de la donnée respectueuse de la sécurité, du cadre juridique et de l'éthique

Évaluer le niveau d'adhésion des acteurs de l'entreprise à la démarche de transition data en mettant en place des indicateurs d'acceptation afin d'identifier d'éventuelles difficultés ou réticences

Mettre en place un dispositif de veille juridique et technologique expert portant sur les champs de la Data, du Big Data, de l'IA et de la cybersécurité en identifiant les sources d'informations les plus pertinentes afin de faire évoluer le Système d'Information (SI) et les solutions Data mises en place

Exploiter les informations issues de la veille juridique et technologique afin de proposer des évolutions du SI et de faire évoluer les solutions Data dont il a la responsabilité

MODALITÉS D'ÉVALUATION

Pour des raisons de faisabilité les différentes évaluations porteront sur la transition d'une activité ou d'un métier appartenant à une entreprise

Modalité d'évaluation des compétences 1 à 3 : Mission de conseil de transition data d'une activité ou d'un métier.

Le candidat (travaillant seul) doit à partir d'un cas réel ou fictif rendre un dossier proposant une stratégie de transition data d'une activité/d'un métier contenant :

- 1- La mise à jour de la charte éthique (accessible) de l'entreprise
- 2 - Les différents éléments de la stratégie de transition de l'entreprise portant sur une activité/un métier
- 3 - La proposition d'évolution coordonnée d'un système de données de l'entreprise pour améliorer sa mise à disposition et son exploitation dans le cadre de l'activité/du métier visé

Modalité d'évaluation des compétences 4 à 6 : Étude de cas consistant à planifier et piloter un projet de transition data.

Le candidat (travaillant seul) doit, à partir d'un cas réel ou fictif de transition d'une entreprise portant sur une activité ou un métier, rendre un dossier contenant :

- 1 - Le plan d'action de la transformation data de l'activité ou du métier,
- 2 - Les éléments essentiels du projet (contexte, objectifs, indicateurs, jalons, et livrables),
- 3 - Une planification,
- 4 - Les outils de suivi (tableaux de bord et autres) accessibles aux acteurs (adaptés aux personnes en situation de handicap)

Modalité d'évaluation des compétences 7 et 8 : Étude de cas portant sur l'accompagnement et l'évaluation de l'adhésion des métiers et des acteurs à la transition data de l'entreprise.

Le candidat (travaillant seul) doit, à partir d'un cas réel ou fictif de transition de l'entreprise portant sur des métiers en lien avec une même activité, rendre un dossier contenant :

- 1- L'identification, la hiérarchisation des difficultés/réticences potentielles par métier
- 2 - Le plan d'accompagnement des métiers en lien avec la transition intégrant les délais, les budgets et l'accessibilité
- 3 - La présentation et l'explication des indicateurs choisis pour mesurer l'adhésion à la démarche de transition des métiers

Modalité d'évaluation des compétences 9 et 10 : Mise en situation portant sur la mise en place et l'exploitation d'un système de veille portant sur la Data, le Big Data, l'IA, la cybersécurité

Le candidat (travaillant seul) doit rendre un dossier contenant :

- 1 - La présentation de son système de veille et de recueil de connaissances de pointe et la présentation de ses sources d'informations
- 2 - La sélection des informations récentes en lien avec les évolutions du SI et des solutions Data
- 3 - Leur hiérarchisation en fonction de leur impact sur le SI et les solutions data,
- 4 - Des préconisations d'évolutions du SI et des solutions Data basées sur ces informations

RNCP37624BC03 — Organiser et mettre en oeuvre le déploiement ou l'industrialisation et la maintenance d'une solution opérationnelle de gestion de données

LISTE DE COMPÉTENCES

Mobiliser et coordonner les expertises techniques en utilisant les méthodes de conduite de projet et d'animation d'équipe, en prenant en compte des éventuelles situations de handicap, afin de mettre en oeuvre le déploiement ou l'industrialisation et la maintenance d'une solution de gestion de données

Déployer ou industrialiser une solution dans le respect du cahier des charges en mobilisant les ressources techniques, budgétaires, humaines et en tenant compte des contraintes pour mettre à disposition un système de collecte maintenable, robuste, scalable et pouvant gérer des données massives

Mettre en place et superviser avec les services informatiques de l'entreprise, la protection de la solution et des données, en déployant les outils de sécurité, d'anonymisation, et en instaurant un mécanisme de supervision afin de consolider l'intégrité, la disponibilité, l'accessibilité, la confidentialité et la lisibilité des données

Mettre en exploitation la solution en monitorant la charge et la performance pour permettre la collecte, le nettoyage, le traitement, le stockage et l'exploitation des données

Attribuer et configurer les droits d'accès au système et aux données mises à disposition pour restreindre leur accessibilité aux personnes autorisées en se référant au cahier des charges du projet et à la politique de sécurité de l'entreprise

Evaluer la performance du système en effectuant des retours sur expérience par la collecte et la synthèse des informations relatives aux insuffisances et aux points d'amélioration potentiels en vue de son évolution

Conduire les évolutions du système en réponse aux points d'amélioration et insuffisances détectées lors de l'évaluation de ses performances afin de garantir un fonctionnement optimisé

Planifier et mettre en place une maintenance préventive du système de gestion de données afin d'assurer son bon fonctionnement et la disponibilité des données

MODALITÉS D'ÉVALUATION

Modalité d'évaluation des compétences 1 à 8 : Mise en situation portant sur le déploiement ou l'industrialisation et la maintenance d'une solution opérationnelle de gestion de données.

Le candidat (en équipe) doit à partir d'une solution de gestion de données (à déployer/industrialiser) tirée d'un cas réel ou fictif, d'un jeu de données, et du cahier des charges technique

:

1 - Définir un plan d'action incluant les modalités de pilotage du déploiement/industrialisation, une planification et une répartition des missions en fonction des expertises techniques des membres de l'équipe et en prenant en compte les éventuelles situations de handicap,

2 - Déployer la solution sur un serveur et y charger les données,

3 - Mettre en place des outils de sécurité de la solution et d'anonymisation des données et instaurer un mécanisme de supervision,

4 - Vérifier le bon fonctionnement de la solution et l'optimiser en prévision d'une montée en charge dans le respect des attentes du client

5 - Attribuer

les droits d'accès en fonction des besoins des utilisateurs,

6 - Configurer les droits d'accès au système et aux données des personnes autorisées,

7 - Mesurer la performance du système et restituer visuellement les écarts positifs et/ou négatifs,

8 - Décider et mettre en oeuvre des améliorations du système basées sur la mesure des écarts de performance et les prescriptions du cahier des charges,

9 - Mettre en place une maintenance en installant et paramétrant des outils de surveillance et de gestion d'alerte et en choisissant des seuils d'alerte pertinents,

10 - Planifier un cycle de maintenance préventive, en indiquant le protocole de maintenance

Le candidat individuellement réalise une présentation documentée, argumentée et personnelle :

- du plan d'action et des éléments qui le composent
- du système de gestion de données,
- de son fonctionnement,

des éléments de sécurité mis en place, des droits d'accès, des performances du système, des améliorations mises en oeuvre, et du système de maintenance

à un jury de deux personnes représentant le client.

Le candidat conclut sa présentation par un bilan critique du projet.

RNCP37624BC04 — Concevoir, mettre en œuvre et optimiser un modèle d'Intelligence Artificielle/Machine Learning

LISTE DE COMPÉTENCES

Analyser expérimentalement les différents modèles d'Intelligence Artificielle/Machine Learning à travers des tests et des analyses statistiques afin de répondre aux besoins du client

Affiner le modèle d'Intelligence Artificielle/Machine Learning en cherchant les paramètres optimaux et en évaluant

l'adéquation des modèles d'apprentissage avec le besoin exprimé afin de garantir ses capacités de prédiction
Déployer ou Industrialiser le modèle d'Intelligence Artificielle/Machine Learning à travers la mise en place des pipelines de données et dans le respect du cadre réglementaire pour garantir la qualité, la fiabilité, la rapidité et la scalabilité de la solution

Évaluer la performance du modèle d'Intelligence Artificielle/Machine Learning déployé ou industrialisé en se basant sur les métriques de référence afin d'optimiser ce modèle

Communiquer ses résultats au client (interne/externe) de façon accessible (datavisualisation) pour démontrer la performance du modèle d'Intelligence Artificielle/Machine Learning

MODALITÉS D'ÉVALUATION

Modalité d'évaluation des compétences 1 à 5 : Mise en situation portant sur la conception et la mise en œuvre d'un modèle d'IA/ML.

Le candidat (seul), doit à partir d'une problématique, de données de test, et d'une série d'exigences, fournies par une entreprise fictive ou réelle :

- 1 - Choisir le modèle d'IA/ML le plus adapté à la problématique en se basant sur des tests et des analyses statistiques et justifier son choix,
- 2 - Paramétrer le modèle pour qu'il traite le plus efficacement possible la problématique,
- 3 - Mettre en place la solution la plus pertinente et respectueuse du cadre réglementaire dans une perspective d'industrialisation,
- 4 - Mesurer la performance du modèle retenu et optimiser le modèle en fonction des résultats obtenus et des métriques de référence fournies par le client (performance du modèle, métriques métiers),
- 5 - Communiquer sur la solution de façon accessible à l'aide d'une datavisualisation et justifier ses choix à travers une présentation,

Le candidat présente sa solution de ML :

à un jury de deux personnes représentant le client (data visualisation et présentation) et justifie ses choix techniques par rapport aux besoins du client

RNCP37624BC05 — Exploiter la donnée pour piloter l'activité de l'entreprise et créer de nouvelles opportunités

LISTE DE COMPÉTENCES

Recueillir les besoins des utilisateurs/décideurs par un processus de questionnement afin d'identifier les données internes à collecter et exploiter

Identifier les données externes au moyen d'un système de veille afin de les intégrer ou de les croiser aux données internes.

Collecter des données structurées, ou non structurées, en mobilisant des sources variées en prévision de leur stockage et de leur exploitation

Stocker des données structurées, ou non structurées, dans une base de données, un data lake, ou un entrepôt de données, en garantissant leur sécurité afin de les mettre à disposition des différents utilisateurs/décideurs

Extraire les données identifiées d'une base de données, d'un data lake ou d'un entrepôt de données, à l'aide d'outils adaptés afin de les mettre à disposition des différents utilisateurs/décideurs dans un format exploitable.

Nettoyer les données contenues dans une base de données, un data lake, ou un entrepôt de données en s'assurant de leur qualité et de leur utilisabilité pour les restituer en un ensemble exploitable, homogène ou hétérogène.

Créer des outils de pilotage accessibles (tableaux de bord automatisés) en identifiant les indicateurs clés pour permettre aux utilisateurs/décideurs de faire face aux évolutions de l'activité et prendre les décisions adaptées.

Présenter le fonctionnement des outils de pilotage et les indicateurs clés identifiés, aux parties prenantes de façon accessible (datavisualisation) afin d'en obtenir la validation et l'appropriation

Présenter et défendre ses recommandations en se basant sur l'analyse des informations issues des outils de pilotage, afin de permettre aux clients internes une prise de décision éclairée (data gouvernance) et de créer de nouvelles opportunités

MODALITÉS D'ÉVALUATION

Modalité d'évaluation des compétences 1 à 6 : Mise en situation de mise en place du cycle de l'information nécessaire à la création d'outils de pilotage.

Le candidat travaillant seul, doit, à partir d'une problématique proposée par une entreprise (réelle ou fictive) et de plusieurs sources de données :

- 1 - Formaliser le besoin exprimé par le client en termes de pilotage d'activité et identifier les données à exploiter (internes et externes),

- 2 - Collecter et stocker, dans des conditions sécurisées, les données qui seront exploitées pour piloter l'activité,
- 3 - Extraire et nettoyer les données nécessaires au pilotage de l'activité du client

Le candidat présente :

à un jury de deux personnes représentant le client son analyse des besoins et les résultats de son travail de collecte, de stockage, d'extraction et de nettoyage de données. Il justifie ses choix et ses actions par rapport aux besoins du client.

Modalité d'évaluation des compétences 7 à 9 : Mise en situation de création et présentation des outils de pilotage, analyse de données et préconisations.

Le candidat travaillant seul, doit, à partir d'une problématique proposée par une entreprise (réelle ou fictive) :

- 1 - Créer les outils de pilotage accessibles correspondant aux besoins du client et identifier les indicateurs clés,
- 2 - Présenter de façon accessible les outils de pilotage et les indicateurs clés identifiés,
- 3 - Produire une analyse des informations issues des outils de pilotage et en tirer des préconisations en termes de prise de décision,
- 4 - Identifier et présenter à partir des outils de pilotage de nouvelles opportunités d'activités

Le candidat présente :

Ses outils de pilotage, ses indicateurs, son analyse et ses préconisations à un jury de deux personnes représentant le client. Il justifie ses choix par rapport aux besoins du client.

Secteur d'activité et type d'emploi

SECTEURS D'ACTIVITÉS

Le Data Engineer est amené à travailler au sein d'entreprises de toutes tailles, de la PME aux très grandes entreprises, et des secteurs privés comme publics. Ainsi il peut intervenir auprès de : Collectivité territoriale ; Entreprise ; Entreprise de Services Numériques - ESN ; Entreprise publique/établissement public ; Société de conseil ; Administration / Services de l'État ; Armée ; Constructeur de matériel informatique ; Informatique et télécommunications ; ...

Au sein de ces structures, et en fonction de leur organisation et de leur maturité en terme d'exploitation de la data, le Data Engineer peut être intégré dans des "services data", "services informatiques", "services BI", ou dans des entités métiers de type "service marketing", "service analyses financières", ...

TYPE D'EMPLOIS ACCESSIBLES

Le Data Engineer est connu sous différentes appellations en fonction des caractéristiques de entreprise qui l'emploie et de son niveau d'expérience.

Après l'obtention de la certification, le candidat peut aller sur des postes de type Data Engineer, et peut aussi se spécialiser dans des missions intégrant du Machine Learning ; il sera alors recruté sous l'appellation Machine Learning Engineer

Il peut aussi se spécialiser dans la gestion de données massives (Big Data) ; il sera alors recruté sous l'appellation Ingénieur Data/Big data / Ingénieur en développement Big Data.

RÈGLEMENTATIONS D'ACTIVITÉ

Le Data Engineer comme tout professionnel de la Data fonctionne avec plusieurs cadres juridiques et réglementaires :

Le cadre Français :

- La loi no 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, plus connue sous le nom de loi informatique et libertés (dans sa rédaction en vigueur depuis 2019)
- Le décret n° 2019-536 du 29 mai 2019 qui détaille les modalités d'application de la Loi Informatique et Liberté et de la RGPD
- La loi n° 2004-575 du 21 juin 2004 pour La Confiance en l'Économie Numérique (LCEN)
- Le Décret n°99-199 du 17 mars 1999 définissant les catégories de moyens et de prestations de cryptologie

Le cadre Européen :

- Le règlement UE 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, dit règlement général sur la protection des données (RGPD)
- Le règlement sur la gouvernance des données (Data Governance Act) - voté en mai 2022, et applicable en septembre 2023 qui régit le partage des données personnelles et non personnelles au niveau européen.

- La proposition de règlement sur les données (Data Act) - du 23 février 2022 - propose de gérer la répartition de la valeur issue de l'utilisation des données personnelles et non personnelles entre les acteurs de l'économie de la donnée (utilisation des objets connectés et développement de l'Internet des objets).
- Le cadre International et notamment pour les USA le Clarifying Lawful Overseas Use of Data Act ou CLOUD Act (H.R. 4943)- 2018

Marché du travail

Indicateurs au niveau des familles professionnelles (FAP) des métiers visés, agrégés au niveau national — France Travail / DARES (open data).

Tension de recrutement (DARES 2024)	Très élevée (5/5) Moyenne pondérée par l'emploi ; les familles ci-dessous s'échelonnent de 4 à 5 sur 5.
Facteurs principaux	Manque de main-d'œuvre disponible, Lien formation-emploi
Salaire médian (mensuel)	2 892 €
Diplôme le plus fréquent	Diplôme de niveau bac+5 ou plus (doctorat, master ou équivalent) en communication et information (41 % des actifs en emploi)
Familles professionnelles (FAP)	M1X80 — Techniciens d'étude et de développement en informatique · tension 4/5 (Élevée) M1X81 — Techniciens de production, d'exploitation, d'installation, et de maintenance, support et services aux utilisateurs en informatique · tension 4/5 (Élevée) M2X90 — Ingénieurs et cadres d'étude, recherche et développement en informatique et télécom · tension 5/5 (Très élevée) M2X91 — Chefs de projet et directeurs de service informatique · tension 4/5 (Élevée) M2X93 — Experts et consultants en systèmes d'information · tension 5/5 (Très élevée)
Projets de recrutement (BMO 2026)	42 073
Jugés difficiles à recruter	50 %
Emplois saisonniers	2 %
Familles professionnelles (FAP)	M1X80 — Techniciens d'étude et de développement en informatique M1X81 — Techniciens de production, d'exploitation, d'installation, et de maintenance, support et services aux utilisateurs en informatique M2X90 — Ingénieurs et cadres d'étude, recherche et développement en informatique et télécom M2X91 — Chefs de projet et directeurs de service informatique M2X93 — Experts et consultants en systèmes d'information

Voies d'accès

MODALITÉS D'ACCÈS

Formation initiale, Formation continue, Apprentissage, Contrat de professionnalisation, VAE

PRÉREQUIS À L'ENTRÉE EN FORMATION

Pour la voie formation le niveau initial requis pour être candidat à cette certification est un niveau 6.

Il est possible de candidater avec un niveau inférieur au niveau 6, si le candidat répond aux exigences suivantes :

- il a validé à minima un niveau 5 en lien avec la certification
- il peut se prévaloir d'une expérience professionnelle significative (plus de 3 ans) en lien avec les activités et les compétences relevant de la certification

PRÉREQUIS À LA VALIDATION

Pour la voie formation :

La certification est obtenue sous les conditions suivantes :

- avoir validé les 5 blocs de compétences,
 - avoir réalisé une période en entreprise de 6 mois (à temps plein) évaluée par un professionnel de la data,
 - avoir rédigé et soutenu un mémoire professionnel. La moyenne des deux notes obtenues par le candidat doit être à minima égale à 10 sur 20 : (note mémoire + note soutenance) / 2
- ou = à 10 sur 20.

Pour la voie VAE :

La certification est obtenue sous réserve :

- d'avoir rédigé un livret de preuves,
- d'avoir passé un entretien avec le Jury de VAE

Le jury évalue la maîtrise des compétences à travers le livret de preuves et l'entretien avec le candidat.

COMPOSITION DU JURY — FORMATION INITIALE, FORMATION CONTINUE, APPRENTISSAGE, CONTRAT DE PROFESSIONNALISATION

Les membres du jury sont :

Le Directeur du centre de formation ou son représentant

Le responsable pédagogique de la formation

2 professionnels du métier représentant les entreprises (dont une personne est désignée comme Président du jury.

Le Président du jury tranche en cas de désaccord).

COMPOSITION DU JURY — VAE

Le Directeur du centre de formation ou son représentant

Le responsable pédagogique en charge de la VAE

2 professionnels du métier représentant les entreprises (dont une personne est désignée comme Président du jury.

Le Président du jury tranche en cas de désaccord).

Le jury est composé à 50% de membres extérieurs à l'autorité délivrant la certification. Ces membres extérieurs sont composés à 50% d'employeurs et à 50% de salariés en activité (soit, 25% du total pour chaque catégorie). Ces représentants sont choisis de façon à concourir à une représentation équilibrée entre les hommes et les femmes.

Financement en alternance (NPEC)

Niveau de prise en charge d'un contrat d'apprentissage par les OPCO, fixé par France compétences par branche professionnelle (CPNE).

Prise en charge usuelle	9 157 €
Selon la branche	8 000 € – 12 000 €
Prise en charge	Branches concernées
9 157 €	Valeur par défaut · 180 branches
12 000 €	CPNE Entreprises des activités du déchet et de la propreté urbaine, CPNE autoroutes, CPNE de l'édition de livres, de l'édition phonographique et de l'édition de musique, CPNEF de la Publicité, CPNEF de la presse et des agences de presse
8 696 €	CPNEFP des commerces et services de l'audiovisuel, de l'électronique et de l'équipement ménager, CPNEFP des jardineries et graineteries, CPNEFP du Négoce de l'ameublement
10 683 €	CPNEFP de l'enseignement privé indépendant, CPNEFP de la Restauration rapide
10 164 €	CPNE Casinos, CPNE de la branche de la Distribution Directe
10 000 €	CPNE des exploitations frigorifiques, CPNE des industries de santé (CPNEIS)
9 992 €	CPNEF branche retraite complémentaire et prévoyance, CPNEFP Mutualité
11 447 €	CPNE des Télécoms
11 446 €	CPNE Travail Temporaire
10 500 €	CPNEFP des industries des tuiles et briques

10 163 €	CPNEF de l'Exploitation cinématographique et de la Distribution de films
9 842 €	CPNEFP Grands Magasins/ Magasins Populaires
9 800 €	CPNEF de l'audiovisuel
9 000 €	CPNEFP de la branche des prestataires de services du secteur tertiaire
8 775 €	CPNEFP de la branche des entreprises du bureau et du numérique (commerces et services)
8 701 €	CPNE FP Branche des magasins du Bricolage
8 570 €	CPNEFP de la Branche des Caisses régionales de Crédit Agricole et autres organismes
8 500 €	CPNEFP de la métallurgie
8 000 €	CPNE chimie

Base légale

Date de décision	31/05/2023
Durée de l'enregistrement	3 ans
Date d'échéance de l'enregistrement	31/05/2026
Date de dernière délivrance possible	31/05/2030

Liens avec d'autres certifications

CORRESPONDANCES (PAR BLOCS)

RNCP34262, RNCP34567, RNCP35197, RNCP36398, RNCP37172, RNCP37422, RNCP37431

Pour plus d'informations

LIEN VERS LE DESCRIPTIF

Pour Lunalogic :

<https://www.lunalogic.com/certification-professionnelle-de-data-engineer-rncp7/>

Pour Stephenson Formation :

<https://stephenson-formation.fr/formation-alternance/titre-data-engineer/>

Fiche officielle (France Compétences)

<https://www.francecompetences.fr/recherche/rncp/37624/>

Se former en apprentissage

Formations en apprentissage	2
Organismes gestionnaires (CFA)	1
Organismes formateurs	1
Régions couvertes	1
Départements couverts	2
Organismes certifiés Qualiopi	1
Sessions	de 09/2025 à 08/2027

Catalogue des formations en apprentissage — réseau des Carif-Oref. Sans frais pour l'alternant : formation financée par l'OPCO de la branche (NPEC).