

Expert en ingénierie des données

RNCP36739 · RNCP · Publiée

L'essentiel

Niveau de qualification	Niveau 7
Code(s) NSF	114, 326
Formacode(s)	31025, 11052, 31028
Code(s) ROME	M1806
Type d'enregistrement	Enregistrement sur demande
Date d'enregistrement	20/07/2022
Date d'échéance	20/07/2025

Certificateur(s)

EFREI PARIS · SIRET : 39889833800023

Partenaires habilités

1 organisme habilité, d'après la liste publiée par France compétences : 1 à former et organiser l'évaluation.

Organisme	SIRET	Habilitation
DRS SCHOOL	85398915000018	Former et organiser l'évaluation

Résumé de la certification

OBJECTIFS ET CONTEXTE

La numérisation et la mémorisation des activités des individus en tant que citoyens, employés, ou clients ont permis l'éclosion d'une nouvelle ère centrée autour de la notion de données.

Selon l'IDC (International Data Corporation), l'univers du digital ne devrait cesser de s'accroître. Sa vitesse de croissance estimée devrait lui permettre de doubler tous les deux ans. Dans ce contexte, les entreprises font face à une grande quantité de données pouvant provenir de leur propre système d'information, mais aussi de multiples sources ouvertes suivant la mouvance de l'Open data et du Big data. L'exploration des données offre potentiellement une forte valeur ajoutée pour les entreprises (notamment dans l'amélioration de leur performance et de leur compétitivité) mais cette plus-value ne peut être extraite sans des compétences techniques qui requièrent des certifications spécifiques à cet effet. La capacité de porter des projets orientés autour de la donnée est devenue une priorité pour 26% des directions métiers en France (IDC 2019). À titre d'exemple, une bonne maîtrise des processus et des pipelines de gestion des données clients pour le stockage, la collecte, le traitement, l'analyse, la restitution des données et surtout la prise de décision est un élément crucial pour la stratégie d'entreprise.

D'autre part le marché du cloud public connaît une croissance très importante, depuis quelques années. Les solutions cloud (SaaS, 'Software as a Service', PaaS 'Platform as a Service', IaaS 'Infrastructure as a service') apportent des avantages incontestables pour accélérer la transformation digitale des entreprises (exemples : modèle de tarification flexible (Pas As You Go), Go-To-Market rapid, une scalabilité naturelle). Fort de ce constat, les entreprises sont de plus en plus à la recherche de compétences cloud et la connaissance et la maîtrise de services proposés par les fournisseurs de cloud les plus connus (Microsoft Azure, Google Cloud Platform, Amazon Web Services) deviennent plus que jamais importantes.

Le métier d' Expert en ingénierie des données : l'Expert en ingénierie des données organise, en tant que développeur informatique, la collecte et la mise à disposition des données en lien avec les équipes métiers. Il est aussi en charge d'industrialiser et mettre en production des traitements sur les données en lien avec les équipes métiers et les

équipes qui les analysent. Le métier d'Expert en ingénierie des données se retrouve sur l'ensemble du cycle de vie de la donnée. Il doit répondre à des besoins hétérogènes, tout en garantissant une uniformité et une cohérence à l'intérieur de l'entreprise. Ses principales missions sont :

- Analyser le besoin et concevoir l'infrastructure de traitement :

L'Expert en ingénierie des données analyse et développe une solution permettant le traitement de volumes importants de données, tout en garantissant la sécurité de celles-ci. Il préconise l'architecture technique nécessaire aux traitements des algorithmes conçus avec le Data Scientiste (ex. : puissance de calcul, stockage, flux de données, utilisation des solutions cloud et développements spécifiques). Il co-anime l'équipe de traitement des données avec le Data Scientiste et contribue tout au long du traitement des données.

- Déployer et maintenir la solution sur son cycle de vie :

Il est le premier échelon de la chaîne de traitement de données et conçoit des solutions permettant le traitement de volumes importants de "pipelines" de données. Il veille à entretenir une solution permettant le traitement massif de données en utilisant ses diverses expertises. Il anticipe les évolutions dans les données (ex. : variations de volumes ou d'usages) et prépare l'architecture de la plateforme à y répondre techniquement.

- Intégrer les technologies du Big Data dans son activité :

Il maîtrise les diverses technologies du Big Data permettant le traitement et la manipulation de données (Hadoop, Spark, Kafka...) ainsi que les langages (JavaScript, Scala, Python...), des systèmes d'exploitation et des solutions de base de données (SQL, NO SQL).

Les évolutions et besoins du marché : L'émergence du Big Data a fait évoluer les besoins en compétences de manière importante. L'étude de Data Transition Numérique sur Les 7 métiers porteurs du Big Data en 2019-2020, définit le Data engineer comme le premier acteur du processus de traitement de la donnée. Son travail intervient en amont de celui du Data scientiste, directement après que l'infrastructure technique ait été mise en œuvre par les architectes et les administrateurs. Métier émergent en tension sur le recrutement, selon l'OPIIEC (Observatoire des métiers du numérique, de l'ingénierie, du conseil et de l'évènement), 1 193 offres d'emploi sur un an ont été publiées, avec une augmentation de plus de 5% par an. Le site Linkedin.com proposait décembre 2021 plus de 3000 offres juste sur le mot clé « Data engineer ». Et cette tendance est confirmée par

l'APEC pour qui " les entreprises ont besoin de mobiliser des compétences spécifiques pour collecter les données, les stocker, les modéliser ou encore les protéger" : depuis 2017, le nombre d'emploi de "Data engineer" sur le site de l'APEC a ainsi augmenté de 76%. Notons que même si les profils de Data Engineer sont de plus en plus recherchés, leurs missions ne sont pas toujours très claires. Maillons incontournables dans la chaîne de la donnée permettant l'accès à une Data propre et valorisable, ce sont des profils clefs dont les fonctions peuvent beaucoup varier : de la collecte des données à la maintenance des data bases et des Framework, en passant par la construction de pipelines de données.

Les entreprises qui recrutent des Data Engineer : toutes les entreprises qui disposent de données (donnée clients, RH, activités, produits, etc.) recrutent de Data engineers. Le rôle de ces derniers est de créer et de développer les outils, les infrastructures, les architectures et les services pour collecter, transformer et traiter les données. Sans les ingénieurs data, les spécialistes et les analystes de données ne peuvent pas accomplir leurs tâches. Le besoin de recrutement est pressant dans tous les secteurs : santé, télécommunication, business, transport, énergie, finance, assurance, numérique etc.

- Santé (Sanofi, la sécurité sociale, Ameli, Bioserentity, etc.)
- Energie (Engie, EDF, etc.)
- Télécommunication (Bouygues, Orange, SFR, etc.)
- Transport (SNCF, RATP, etc.)
- E-commerce (La Redoute, C- discount, vente privée, etc.)
- Finance (BNP, BPCE, Société Générale, etc.)
- Assurance (Axa, La Mutuelle Générale, Allianz, Directe Assurance, etc.)
- Numérique
- Intelligence artificielle (Axi technologies)
- Telecom (Bouygues)
- Banque (CGI)

Compétences recherchées :

- Un langage de programmation (Python, R, JavaScript, Java, Scala, Go, Rust, etc.)
- L'architecture Lambda (bonus: architecture Kappa)
- Stream processing (Kafka, Spark, Flume)
- Maîtrise des différentes techniques d'ETL et de la gestion de gros volumes

- Bases de données distribuées
- Mathématiques/statistiques
- Frameworks Big Data
- Maîtrise de SQL et des bases de données NoSQL
- Cloud computing

Dans ce contexte la certification d'« Expert en ingénierie des données » d'EFREI Paris s'appuie sur le programme du Mastère Data Engineering de l'école. L'originalité de cette certification repose sur son interdisciplinarité centrée autour de la gestion des données en couplant cela avec les compétences du cloud les plus recherchées. Il s'agit donc de couvrir un ensemble de fondamentaux techniques, mais aussi et surtout de former les candidats aux multiples enjeux associés aux données. Ces enjeux devant être connus et appréhendés par les décideurs de demain concernent la maîtrise de la chaîne de la collecte, l'analyse, la protection et la gestion du stockage de données, mais aussi l'utilisation de ces données pour en extraire de la valeur ajoutée grâce à l'utilisation de l'intelligence artificielle. La certification intègre donc les problématiques associées à la gestion des données sur le cloud dans toute leur complexité et leur multidisciplinarité. Ainsi, les aspects d'ingénierie associés aux données seront le socle de cette certification, mais en complément les problématiques liées à l'administration des données, à l'économétrie, au marketing analytique, à la prise de décision, aux contraintes juridiques auront un rôle tout aussi crucial.

ACTIVITÉS VISÉES

Les activités du métier d'« Expert en ingénierie des données » consistent de manière non exhaustive à utiliser les outils de Gestion de données, d'Intelligence Artificielle, de Machine Learning et de Cloud pour le/la :

Conception d'architecture de stockage de données, Développement d'une architecture de stockage de données, Construction de lac de données Traitement et Analyse de données

Conception de solution de traitement de données, Développement de solution données, Déploiement d'une solution de traitement des données massives, Transformation de données issues de sources différentes, Optimisation de pipelines,

Application de systèmes appropriés en réponse à une demande, Création et automatisation de tests

Conception de solutions de traitement de données, Mise en œuvre de solutions de stockage de données dans le cloud, Développement de solution de traitement de données, Exploitation de données, Mise à disposition de données à des fins d'utilisation, Sécurisation de données, Stratégie et Politique de sécurisation de données, Conformité réglementaire, Protection des données et des accès utilisateurs, Optimisation des coûts

Extraction de données depuis différentes sources, Analyse exploitation de données avec outils appropriés, Transformation de données selon les besoins définis, Construction de tableaux de bord, reporting, Communication de données et d'indicateurs en temps réel, Implémentation d'algorithmes de machine Learning, Analyse et comparaison d'outils de (machine Learning), Choix d'outils appropriés

Définition de stratégie de gestion de données, Application de politique et de standards appropriés pour la gestion des données, Définition de responsabilité et de rôles dans la gestion des données, Veille et amélioration des pratiques en gestion de données par rapport à la réglementation en vigueur, Vérification de la qualité et la fiabilité des données, Valorisation de l'utilisation de la donnée, Elaboration d'une stratégie et une sécurisation des données, politique d'accès

COMPÉTENCES ATTESTÉES

Les compétences sont réparties dans 4 bocs :

1. Concevoir et développer une architecture de stockage de données
2. Concevoir, développer et déployer une solution de traitement des données massives
3. Implémenter et optimiser des solutions de stockage et de traitement de données sur le cloud
4. Implémenter des méthodes d'intelligence artificielle pour modéliser et prédire de nouveaux comportements et usages.
5. Concevoir une stratégie de management et de gouvernance de données pour transformer les données en informations créatrices de valeur

Blocs de compétences

RNCP36739BC01 — Concevoir et développer une architecture de stockage de données

LISTE DE COMPÉTENCES

Concevoir et développer une base de données relationnelle en réponse aux besoins d'un client en vue de la mise à disposition de ses données structurées pour un tout en utilisant les technologies et les langages de requêtes adaptés aux développements envisagés

Concevoir et développer une base de données non-relationnelle en vue de la mise à disposition des données semis-structurées et non-structurées pour un traitement analytique ou d'intelligence artificielle tout en utilisant les technologies et les langages de requêtes adaptés

Concevoir et construire un lac de données «?Datalake?» en choisissant les architectures, les indicateurs de performance et les solutions de stockage appropriées afin d'intégrer les données provenant des systèmes d'information multivariés : Données non structurées, semi-structurées, structurées et des solutions de stockage de données massives ; redondantes.

Créer une API en utilisant les technologies qui permettent de rendre accessibles les données en augmentant l'efficacité et la praticité des applications et des services

MODALITÉS D'ÉVALUATION

Travaux pratiques?: Évaluation : Rapport de TP rédigé par étudiant et examiné par un jury d'évaluation.

Étude de cas?réalisée en amont du passage devant le jury Création d'une base de données

Présentation orale et démonstration par binôme devant un jury d'évaluation

Mise en situation?: Challenge (hackathon)?: construction d'un datalake qui permet de stocker tout type de données

RNCP36739BC02 — Concevoir, développer et déployer une solution de traitement des données massives

LISTE DE COMPÉTENCES

Concevoir en s'appuyant sur une veille technologique et mettre en œuvre une architecture distribuée répondant au besoin du client pour traiter les données massives en entreprise en utilisant les technologies de traitement

Implémenter un système distribué en utilisant des technologies de streaming identifiées à partir d'une veille pour traiter des données sur une période précise ou en temps quasi réel

Transformer les données provenant de différentes sources en prenant en compte la variété de données pour faire de l'analytique à échelle (intégration, formatage, manipulation, stockage données multidimensionnelles)

Optimiser la performance des pipelines en utilisant les techniques d'intégration et de mise en scène adéquates pour le traitement des données massives

Automatiser la création, les tests, l'intégration et le déploiement des pipelines de données en s'appuyant sur une veille technologique qui permet d'identifier et de mobiliser les solutions pour maximiser l'efficacité et réduire le 'time to market' tout en utilisant les technologies de containerisation et d'ordonnancement.

MODALITÉS D'ÉVALUATION

Étude de cas

réalisée en amont: Conception d'une application de traitement distribué.

Présentation de l'architecture par binôme devant un jury et les autres apprenants.

Étude de cas réalisée en amont: Conception d'une solution d'intégration et de transformation de données.

Étude de cas réalisée en amont: Automatiser la création, les tests, l'intégration et le déploiement des pipelines de données

RNCP36739BC03 — Implémenter et optimiser des solutions de stockage et de traitement de données sur le cloud

LISTE DE COMPÉTENCES

Mettre en œuvre des solutions de stockage de données dans le cloud pour permettre aux entreprises d'explorer leurs données en implémentant des techniques et stratégies adaptées à leur utilisation et en choisissant les types de stockage et d'archivage appropriés

Concevoir et développer des pipelines et des solutions de traitement de données dans le cloud en chargeant les données en les transformant et en les mettant à disposition des utilisateurs pour permettre leur exploitation

Mettre en place une politique de sécurité des données dans le cloud en développant une stratégie de chiffrement et de gestion des identités et des accès et protéger les données en étant conforme aux règlements en vigueur

Optimiser les solutions de stockage et de traitement des données dans le cloud en définissant des indicateurs de performance pour assurer la disponibilité des services et optimiser les coûts

MODALITÉS D'ÉVALUATION

Certification externe par un fournisseur du cloud public (e.g. Azure, Aws, Google)

Étude de cas?: réalisé en amont: Conception et déploiement d'une solution de stockage et de traitement sur le cloud.

Soutenance Orale et démonstration de la solution par binôme devant un jury et les autres apprenants. Rapport écrit en binôme examiné par un jury.

Étude de cas?: Mise en place d'une politique de sécurité de données dans le cloud.

Rapport écrit par binôme évalué par un jury.

RNCP36739BC04 — Implémenter des méthodes d'intelligence artificielle pour modéliser et prédire de nouveaux comportements et usages.

LISTE DE COMPÉTENCES

Extraire des données en provenance de systèmes d'information multivariés pour les exploiter et les analyser en utilisant les outils professionnels courants

Préparer les données en les transformant et en les nettoyant pour faire l'analyse et le reporting selon les besoins des différents métiers?

Elaborer une communication infographique visuelle en construisant des tableaux de bord interactifs afin de communiquer les résultats d'analyses et d'assurer l'extraction de connaissances en temps réel.

Développer un modèle prédictif pour identifier de nouveaux comportements et usages en implémentant des algorithmes d'apprentissage automatiques supervisés

Évaluer la performance d'un modèle de Machine Learning en analysant ses résultats et en comparant avec d'autres modèles afin d'implémenter la solution qui convient le mieux à un cas d'usage

MODALITÉS D'ÉVALUATION

Étude de cas?: réalisée en amont: Environnement logiciel

Soutenance Orale par binôme devant un jury et les autres apprenants.

Rapport de TP écrit en binôme examiné par un jury.

Étude de cas?: réalisé en amont : Mise en place d'une application interactive avec de tableaux de bord

Présentation du tableau de bord effectué par binôme examiné par un jury d'évaluation

Mise en situation?: Proposé via une plateforme de compétition data science de type kaggle

Soumission des résultats de prédiction sur un ensemble de données test et obtenir un score du modèle proposé.

Soumission d'un rapport expliquant la stratégie suivie et le choix du modèle. Le rapport sera examiné par un jury d'évaluation comprenant un représentant opérationnel professionnel et un enseignant spécialisé

RNCP36739BC05 — Concevoir une stratégie de management et de gouvernance de données pour transformer les données en informations créatrices de valeur

LISTE DE COMPÉTENCES

Définir une gouvernance de données en mettant en place les politiques et les standards afin d'établir les rôles, les responsabilités et la propriété des données

Mettre en place les bonnes pratiques de gestion de données en appliquant les réglementations en vigueur pour protéger les données et respecter la vie privée dans une logique de transparence.

S'assurer que la qualité de la donnée permet d'atteindre les objectifs business en faisant un audit de qualité et en proposant des plans de prévention et de remédiation en cas de non-qualité

Identifier l'approche stratégique pour faire de la donnée un actif central ancré dans la culture de l'entreprise et intégrer l'approche dans un plan de communication diffusé aux concernés

Élaborer une stratégie de sécurité afin de protéger les données en utilisant les technologies adéquates permettant de parer aux menaces et en mettant en place une politique d'accès aux données

MODALITÉS D'ÉVALUATION

Études de cas?: réalisées en amont

Soutenance Orale et démonstration de la solution par binôme devant un jury et les autres apprenants.

Rapport écrit en binôme examiné par un jury.

Étude de cas?: réalisée en amont: mise en place de bonnes pratiques et de conformité à la réglementation en vigueur à partir d'un cas d'usage

Secteur d'activité et type d'emploi

SECTEURS D'ACTIVITÉS

Edition de logiciel - Marketing digital - Web E-Commerce - Technologie et numérique - Santé (Sanofi, la sécurité sociale, Ameli, Bioserentity, etc.) - Energie (Engie, EDF, etc.) - Télécommunication (Bouygues, Orange, SFR, etc.) - Transport (SNCF, RATP, etc.) - E-commerce (La Redoute, C- discount, vente privée, etc.) - Finance (BNP, BPCE, Société Générale, etc.) - Assurance (Axa, La Mutuelle Générale, Allianz, Directe Assurance, etc.) - Intelligence artificielle (Axi technologies) - Telecom (Bouygues) - Banque (CGI)

TYPE D'EMPLOIS ACCESSIBLES

Spécialiste base de données - spécialiste blockchain - spécialiste test et validation - ingénieur machine learning - consultant architecte technique - spécialiste infrastructure Data

RÉGLEMENTATIONS D'ACTIVITÉ

Si le métier d'Expert en ingénierie des données n'est pas réglementé en tant que tel, ses activités s'exercent dans le respect des réglementations et contraintes liées notamment à la sécurité des données (notamment dans le cloud), à la propriété/protection des données et au respect de la vie privée dans le cadre de la stratégie et de la gouvernance des données entre autres.

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)
Facteurs principaux	Lien formation-emploi, Manque de main-d'œuvre disponible
Salaire médian (mensuel)	3 115 €
Diplôme le plus fréquent	Diplôme de niveau bac+5 ou plus (doctorat, master ou équivalent) en communication et information (35 % des actifs en emploi)
Familles professionnelles (FAP)	M2X93 — Experts et consultants en systèmes d'information · tension 5/5 (Très élevée)
Projets de recrutement (BMO 2026)	4 758
Jugés difficiles à recruter	45 %
Emplois saisonniers	0 %
Familles professionnelles (FAP)	M2X93 — Experts et consultants en systèmes d'information

Voies d'accès

PRÉREQUIS À L'ENTRÉE EN FORMATION

L'accès au dispositif menant à la certification d'« expert en ingénierie des données » est soumis au fait de détenir une certification de niveau 6 dans le domaine visé et enregistrée au RNCP. Le dispositif s'adresse à des candidats détenant à minima ce niveau 6.

PRÉREQUIS À LA VALIDATION

La certification s'acquiert par la validation de l'ensemble des blocs de compétences qui la composent.

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

Les membres du jury sont choisis en fonction de leur expertise sur la certification concernée et nommés par la direction de l'Efrei

Le jury est constitué de cinq (5) personnes

- 2 personnels administratifs d'Efrei extérieurs au processus de formation menant à la certification dont un président désigné par la direction d'Efrei;
- 3 professionnels qualifiés extérieurs à l'autorité de certification

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	9 050 €
Selon la branche	8 423 € – 9 800 €
Prise en charge	Branches concernées
9 050 €	Valeur par défaut · 186 branches
9 800 €	CPNE Casinos, CPNE de l'ingénierie, des services informatiques et du conseil, CPNE de l'édition de livres, de l'édition phonographique et de l'édition de musique, CPNE des Télécoms, CPNE des industries de santé (CPNEIS)
8 696 €	CPNE FP Commerce de Détail de l'Horlogerie Bijouterie, CPNEFP Commerce succursaliste de la Chaussure, CPNEFP Commerce à distance, CPNEFP Grands Magasins/ Magasins Populaires, CPNEFP Import-Export
9 276 €	CPNEF de l'Exploitation cinématographique et de la Distribution de films
9 094 €	CPNEFP de la blanchisserie, laverie, location de linge, nettoyage à sec, pressing et teinturerie
9 005 €	CPNE FP Branche des magasins du Bricolage
8 813 €	CPNEFP Professions de la Photographie
8 775 €	CPNEFP de la branche des entreprises du bureau et du numérique (commerces et services)
8 600 €	CPNEFP de l'habillement, mercerie, chaussure et jouet
8 500 €	CPNEFP de la métallurgie
8 423 €	CPNEFP de la Branche des Caisses régionales de Crédit Agricole et autres organismes

Base légale

Date de décision	20/07/2022
Durée de l'enregistrement	3 ans
Date d'échéance de l'enregistrement	20/07/2025

Liens avec d'autres certifications

NOUVELLE(S) CERTIFICATION(S)
RNCP40875

CORRESPONDANCES (PAR BLOCS)
RNCP35450, RNCP36921, RNCP37137, RNCP37172, RNCP38584, RNCP38587

Pour plus d'informations

LIEN VERS LE DESCRIPTIF
<https://www.efrei.fr/wp-content/uploads/2021/12/Mastere-Data-Engineering-1.pdf>
<https://www.efrei.fr/programmes-experts/msc-data-engineering/>

Fiche officielle (France Compétences)

<https://www.francecompetences.fr/recherche/rncp/36739/>