

Expert big data engineer (MS)

RNCP40234 · RNCP · Publiée

L'essentiel

Niveau de qualification	Niveau 7
Code(s) NSF	326
Formacode(s)	31023, 31006, 31032
Code(s) ROME	M1802, M1801, M1805, M1803, M1806
Type d'enregistrement	Enregistrement sur demande
Date d'enregistrement	28/02/2025
Date d'échéance	28/02/2030

Certificateur(s)

UNIVERSITE DE TECHNOLOGIE DE TROYES (UTT) · SIRET : 19101060200032

Résumé de la certification

OBJECTIFS ET CONTEXTE

L'Expert Big Data Engineer est un professionnel qui a un rôle stratégique dans la manipulation, la gestion et l'analyse des données au sein des organisations. Ce rôle implique une variété de responsabilités, allant de la conception des infrastructures de données à leur optimisation pour soutenir les décisions stratégiques de l'entreprise.

ACTIVITÉS VISÉES

Analyse des besoins et des problématiques
 Identification des données de l'organisation
 Formalisation du besoin et des solutions techniques
 Préconisations de stratégies en matière de possession, gestion et transformation des données
 Définition de l'architecture de stockage
 Tests et mise en place de protocoles
 Extraction d'information
 Mise en œuvre de langages et outils de programmation
 Réalisation d'une visualisation dynamique
 Analyse des données et conseil auprès du commanditaire
 Utilisation de données en grande dimension
 Déploiement d'un modèle d'apprentissage en profondeur
 Traitements de volumes de données

COMPÉTENCES ATTESTÉES

Réaliser un état des lieux de l'outil informatique au sein d'une organisation privée/publique en cartographiant en profondeur l'intégralité du système d'information, en identifiant les éléments clés et les points de vulnérabilité, et en définissant les priorités, afin d'optimiser les performances et la sécurité du système tout en alignant les ressources technologiques avec les objectifs stratégiques de l'organisation
 Analyser les problématiques de structuration de la donnée au sein de d'une organisation privée/publique en mettant en place un déploiement d'une architecture ou d'une structure, en interagissant avec des interlocuteurs différents et concernés par la problématique, afin d'identifier les besoins de l'organisation auxquels il convient d'apporter une réponse
 Animer une réunion avec tous les acteurs principaux et secondaires concernés par une problématique relevant de la donnée, existante ou nouvelle, en tenant compte de l'avis des différents acteurs, en évoquant les pistes de solutions,

et en identifiant les besoins d'adaptation de la démarche en raison d'une situation de handicap, afin s'immerger dans l'écosystème de la structure dans lequel on intervient

Formaliser le besoin émanant de la problématique au sein d'un cahier des charges dans le respect du budget alloué afin de proposer des solutions réalisables et réalistes avec les données et ressources propres possédées

Concevoir une stratégie Data en réponse aux problématiques identifiées, en proposant une solution technique adaptée aux caractéristiques de l'organisation en termes de serveurs, stockage, mise en réseau, logiciels, et les risques liés à la donnée, afin d'assurer une gestion et une transformation des données sécurisée, conforme aux réglementations et alignée avec les objectifs stratégiques de l'organisation

Présenter un rapport détaillé de faisabilité en tenant compte des paramètres extérieurs tels que le développement durable, la réglementation en vigueur liée à la confidentialité des données, l'éthique afin de les rendre compatibles avec les contraintes inhérentes à l'organisation

Préparer la mise en œuvre du projet en identifiant les ressources humaines, matérielles et financières à disposition, en déterminant les besoins manquants, en constituant une équipe projet et en fixant un rétroplanning, afin d'assurer une planification efficace et une exécution réussie du projet

Identifier la façon dont l'organisation assure le stockage et le format des données internes, structurées ou non structurées, en tenant compte des données manquantes, erronées ou perdues, des différents processus en fonction des pays, des entités qui vont produire les flux de données et de la réglementation en vigueur sur la confidentialité des données (RGPD) afin d'avoir un jeu de données adaptées (ex : Data Frame Python, R/Tibble)

Réaliser des tests sur les web services, en appliquant des tests de montée en charge de l'outil et sa sécurité, en évaluant la performance et la fiabilité des services, et en analysant le besoin pour déterminer la nécessité d'une API, afin de valider le fonctionnement de l'outil

Rédiger des protocoles à destination des utilisateurs des web services, et en rédigeant des tutoriels et documents adaptés et en organisant leur diffusion, afin de permettre aux équipes d'utiliser les outils de manière optimale et autonome

Mettre en œuvre des techniques d'extraction d'information intrinsèque ou par recoupement d'interconnexion en utilisant des outils de visualisation de tendances en grandes dimensions, en mettant en relation différentes bases de données et en actionnant ces bases, afin d'extraire des données pertinentes dans les flux de données

Analyser les données de manière approfondie en mettant en œuvre des outils de modélisation (régression, réseaux de neurones), des outils de classification (supervisés ou non) et des outils de réduction de dimension (analyse factorielle), afin de tirer des conclusions exploitables pour la prise de décision

Programmer en utilisant des langages de programmation adaptés à la donnée récoltée au sein de l'organisation et de la donnée provenant de sources extérieures afin de mettre en place des algorithmes gérant le volume important des données

Réaliser une visualisation dynamique, en choisissant et en utilisant les outils de visualisation avancés adaptés, en respectant les méthodes de représentation graphique, et en prenant les dispositions nécessaires pour permettre l'accessibilité de ces données, notamment en cas de handicap, afin de faciliter la compréhension et l'analyse des données par l'ensemble des utilisateurs

Réaliser l'analyse et la visualisation spatio-temporelles des données en utilisant des outils de Scalable MultiMedia Analytics, afin de découvrir des phénomènes inconnus, confirmer ou infirmer des hypothèses de travail, et interpréter les graphiques obtenus pour fournir des informations précises et exploitables

Formaliser des rapports et des conseils stratégiques, économiques et organisationnels en analysant les données visualisées et en interprétant les résultats obtenus, et en établissant des recommandations pertinentes et adaptées aux utilisateurs dédiés, pour optimiser la prise de décision

Mettre en place un système de gestion des données en anticipant le traitement des données, en agrégeant les données provenant d'autres bases, en normalisant les données et en gérant les valeurs manquantes, afin de permettre une extraction efficace des connaissances sur des jeux de données massives et d'implémenter des outils de classification des données

Entraîner un algorithme de machine learning en choisissant l'algorithme approprié et en utilisant des jeux de données pertinents, afin de développer des modèles prédictifs et analytiques robustes, capables de fournir des insights précieux pour l'organisation, par exemple en identifiant les liaisons ou les patterns (motifs) entre les nombreuses bases de données relationnelles

Déployer des modèles de Deep Learning en mettant en place des applications de Deep Learning, en utilisant une interface de programmation (API) dans le web, et en prenant en compte la consommation d'énergie dans le choix des outils ou des prestataires, afin de développer des solutions d'intelligence artificielle efficaces et durables

Présenter le travail de modélisation en réalisant un tableau de bord, en rédigeant un document écrit et en effectuant

une présentation orale, afin de garantir une compréhension claire et approfondie de la démarche de modélisation réalisée

Assurer le traitement de volumes (gros ou non) de données en sélectionnant et en utilisant un langage de programmation/logiciel dédié gérant des algorithmes, et en optimisant le fonctionnement des algorithmes, afin de permettre à la structure d'opérer plus vite

Blocs de compétences

RNCP40234BC01 — Analyser l'existant et proposer des solutions liées à la data

LISTE DE COMPÉTENCES

Réaliser un état des lieux de l'outil informatique au sein d'une organisation privée/publique en cartographiant en profondeur l'intégralité du système d'information, en identifiant les éléments clés et les points de vulnérabilité, et en définissant les priorités, afin d'optimiser les performances et la sécurité du système tout en alignant les ressources technologiques avec les objectifs stratégiques de l'organisation

Analyser les problématiques de structuration de la donnée au sein de d'une organisation privée/publique en mettant en place un déploiement d'une architecture ou d'une structure, en interagissant avec des interlocuteurs différents et concernés par la problématique, afin d'identifier les besoins de l'organisation auxquels il convient d'apporter une réponse

Animer une réunion avec tous les acteurs principaux et secondaires concernés par une problématique relevant de la donnée, existante ou nouvelle, en tenant compte de l'avis des différents acteurs, en évoquant les pistes de solutions, et en identifiant les besoins d'adaptation de la démarche en raison d'une situation de handicap, afin s'immerger dans l'écosystème de la structure dans lequel on intervient

Formaliser le besoin émanant de la problématique au sein d'un cahier des charges dans le respect du budget alloué afin de proposer des solutions réalisables et réalistes avec les données et ressources propres possédées

Concevoir une stratégie Data en réponse aux problématiques identifiées, en proposant une solution technique adaptée aux caractéristiques de l'organisation en termes de serveurs, stockage, mise en réseau, logiciels, et les risques liés à la donnée, afin d'assurer une gestion et une transformation des données sécurisée, conforme aux réglementations et alignée avec les objectifs stratégiques de l'organisation

Présenter un rapport détaillé de faisabilité en tenant compte des paramètres extérieurs tels que le développement durable, la réglementation en vigueur liée à la confidentialité des données, l'éthique afin de les rendre compatibles avec les contraintes inhérentes à l'organisation

Préparer la mise en œuvre du projet en identifiant les ressources humaines, matérielles et financières à disposition, en déterminant les besoins manquants, en constituant une équipe projet et en fixant un rétroplanning, afin d'assurer une planification efficace et une exécution réussie du projet

MODALITÉS D'ÉVALUATION

Réalisation de Projets - Questionnaires - Exercices - Mise en situation professionnelle - Étude de cas

RNCP40234BC02 — Déployer des infrastructures informatiques d'analyse de la data

LISTE DE COMPÉTENCES

Identifier la façon dont l'organisation assure le stockage et le format des données internes, structurées ou non structurées, en tenant compte des données manquantes, erronées ou perdues, des différents processus en fonction des pays, des entités qui vont produire les flux de données et de la réglementation en vigueur sur la confidentialité des données (RGPD) afin d'avoir un jeu de données adaptées (ex : Data Frame Python, R/Tibble)

Réaliser des tests sur les web services, en appliquant des tests de montée en charge de l'outil et sa sécurité, en évaluant la performance et la fiabilité des services, et en analysant le besoin pour déterminer la nécessité d'une API, afin de valider le fonctionnement de l'outil

Rédiger des protocoles à destination des utilisateurs des web services, et en rédigeant des tutoriels et documents adaptés et en organisant leur diffusion, afin de permettre aux équipes d'utiliser les outils de manière optimale et autonome

Mettre en œuvre des techniques d'extraction d'information intrinsèque ou par recoupement d'interconnexion en utilisant des outils de visualisation de tendances en grandes dimensions, en mettant en relation différentes bases de données et en actionnant ces bases, afin d'extraire des données pertinentes dans les flux de données

Analyser les données de manière approfondie en mettant en œuvre des outils de modélisation (régression, réseaux de neurones), des outils de classification (supervisés ou non) et des outils de réduction de dimension (analyse

factorielle), afin de tirer des conclusions exploitables pour la prise de décision

MODALITÉS D'ÉVALUATION

Réalisation de Projets -

Rapport professionnel - Soutenance orale

RNCP40234BC03 — Traiter et visualiser de la data

LISTE DE COMPÉTENCES

Programmer en utilisant des langages de programmation adaptés à la donnée récoltée au sein de l'organisation et de la donnée provenant de sources extérieures afin de mettre en place des algorithmes gérant le volume important des données

Réaliser une visualisation dynamique, en choisissant et en utilisant les outils de visualisation avancés adaptés, en respectant les méthodes de représentation graphique, et en prenant les dispositions nécessaires pour permettre l'accessibilité de ces données, notamment en cas de handicap, afin de faciliter la compréhension et l'analyse des données par l'ensemble des utilisateurs

Réaliser l'analyse et la visualisation spatio-temporelles des données en utilisant des outils de Scalable MultiMedia Analytics, afin de découvrir des phénomènes inconnus, confirmer ou infirmer des hypothèses de travail, et interpréter les graphiques obtenus pour fournir des informations précises et exploitables

Formaliser des rapports et des conseils stratégiques, économiques et organisationnels en analysant les données visualisées et en interprétant les résultats obtenus, et en établissant des recommandations pertinentes et adaptées aux utilisateurs dédiés, pour optimiser la prise de décision

MODALITÉS D'ÉVALUATION

Réalisation de Projets -

Rapport professionnel - Soutenance orale

RNCP40234BC04 — Analyser et modéliser la data

LISTE DE COMPÉTENCES

Mettre en place un système de gestion des données en anticipant le traitement des données, en agrégeant les données provenant d'autres bases, en normalisant les données et en gérant les valeurs manquantes, afin de permettre une extraction efficace des connaissances sur des jeux de données massives et d'implémenter des outils de classification des données

Entraîner un algorithme de machine learning en choisissant l'algorithme approprié et en utilisant des jeux de données pertinents, afin de développer des modèles prédictifs et analytiques robustes, capables de fournir des insights précieux pour l'organisation, par exemple en identifiant les liaisons ou les patterns (motifs) entre les nombreuses bases de données relationnelles

Déployer des modèles de Deep Learning en mettant en place des applications de Deep Learning, en utilisant une interface de programmation (API) dans le web, et en prenant en compte la consommation d'énergie dans le choix des outils ou des prestataires, afin de développer des solutions d'intelligence artificielle efficaces et durables

Présenter le travail de modélisation en réalisant un tableau de bord, en rédigeant un document écrit et en effectuant une présentation orale, afin de garantir une compréhension claire et approfondie de la démarche de modélisation réalisée

Assurer le traitement de volumes (gros ou non) de données en sélectionnant et en utilisant un langage de programmation/logiciel dédié gérant des algorithmes, et en optimisant le fonctionnement des algorithmes, afin de permettre à la structure d'opérer plus vite

MODALITÉS D'ÉVALUATION

Réalisation de Projets -

Rapport professionnel - Soutenance orale

Secteur d'activité et type d'emploi

SECTEURS D'ACTIVITÉS

Compte tenu de la transformation digitale de la société et de l'industrie qui va en s'accélération, tous les types de secteur d'activités générant et manipulant les données sont concernés et en particulier : Assurance et banque - Industrie X.0 en rapport avec la numérisation des processus (mise en valeur des données de production, de maintenance, de logistique ou financières) - Services publics (mise en valeur des données de parcours individuels, de

maintenance, de logistique ou financières - Le réseau d'énergie - Les médias, le marketing, la communication - La (cyber)sécurité - La sûreté

De même toutes les tailles d'entreprises ou de service employeurs sont potentiellement concernés : grands groupes - startups, PME / TPE – industrie - sociétés de services et ESN - agence web et agence de communication – associations - établissements publics & collectivités territoriales

L'Expert Big Data Engineer (MS) peut également exercer en tant que consultant indépendant.

TYPE D'EMPLOIS ACCESSIBLES

Les intitulés de poste identifiés peuvent être regroupés en trois catégories principales : Data Analyst, Data Architect et Data Scientist. Cette classification permet de mettre en évidence les différentes spécialisations visées par la certification, en cohérence avec les besoins du marché. Voici la répartition des intitulés de poste

Data Analyst / Data analyste : Consultant Data Analytics - Data Manager - Analyste digital customer – Economiste -data analyst - Analyste de données

Data Architect : Data Engineer - Cloud Data Engineer - Consultant junior Data Engineer - Consultant Data/API Management - Consultant data engineer - Expert Oracle PLSQL - Ingénieur data - Data product owner - Big data product manager - MOA (Maîtrise d'Ouvrage) - Ingénieur en technologie de l'information

Data Scientist : Data Scientist - Data Consultant /Consultant Data - Ingénieur d'études

RÈGLEMENTATIONS D'ACTIVITÉ

Le métier n'est pas réglementé, néanmoins dans l'exercice professionnel de son activité, l'expert Big Data Engineer doit respecter et tenir compte des droits, obligations et normes en matière de sécurisation et de gestion de données massives pour garantir sa mise en conformité. Il doit prendre en compte les obligations réglementaires et normatives en vigueur notamment le Règlement Général sur la Protection des Données (RGPD) et veiller à ce que le stockage, le traitement ainsi que la visualisation des données répondent au besoin. Il joue aussi un rôle de conseil en matière d'exploitation ou de monétisation de la donnée

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)	Élevée (4/5) Moyenne pondérée par l'emploi ; les familles ci-dessous s'échelonnent de 3 à 5 sur 5.
Facteurs principaux	Manque de main-d'œuvre disponible, Lien formation-emploi
Salaire médian (mensuel)	2 912 €
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) M2X92 — Responsables et cadres de la production, de l'exploitation et de la maintenance informatique et télécom · tension 3/5 (Modérée) M2X93 — Experts et consultants en systèmes d'information · tension 5/5 (Très élevée)
Projets de recrutement (BMO 2026)	43 955
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
M2X92 — Responsables et cadres de la production, de l'exploitation et de la maintenance informatique et télécom
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

L'ensemble des formations mastère spécialisé est ouvert aux étudiants titulaires de l'un des diplômes suivants :

Titre d'ingénieur diplômé conférant le grade de master (formations évaluées par la Commission des titres d'ingénieur, liste publiée au journal officiel)

Diplôme d'une école de management, privée ou consulaire, conférant le grade de master (formations évaluées par la CEFDG, liste publiée au bulletin officiel du MESR)

;

Diplôme de 3e cycle habilité par les autorités administratives (Diplôme national de master, DEA, DESS...) ou diplôme professionnel de niveau Bac+5 ;

Diplôme ou attestation de validation d'un niveau équivalent M1 pour des candidats ayant au moins trois années d'expérience professionnelle en lien avec la formation visée ;

Titre inscrit au Répertoire national des certifications professionnelles (RNCP) niveau 7

Diplôme étranger équivalent aux diplômes français exigés ci-dessus.

Conditions d'accès dérogatoires

a) Dans la limite de 40 % maximum de l'effectif de la promotion suivant la formation Mastère Spécialisé concernée, sont recevables, après une procédure de Validation des acquis personnels et professionnels (VAPP), les candidatures de personnes, justifiant a minima de 5 années d'expérience professionnelle pour lesquelles les activités exercées ont un lien avéré avec les compétences professionnelles visées par la formation.

b) Par dérogation pour 30 % maximum du nombre d'apprenants suivant la formation Mastère Spécialisé concernée, sont recevables les candidatures de titulaires d'un des diplômes suivants :

? Diplôme ou attestation de validation d'un niveau équivalent M1 sans expérience professionnelle ou ayant moins de trois ans d'expérience professionnelle en lien avec la formation visée.

? Diplôme de licence (L3) ou grade de Licence ou titre inscrit au RNCP niveau 6 justifiant d'une expérience professionnelle de 3 ans minimum, en lien avec la formation visée

Le pourcentage total des dérogations prévues au a) et au b) ci-dessus ne doit pas excéder 40%.

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

Le jury de délivrance de la certification sera composé 4 membres, dont au moins deux membres professionnels et extérieurs au dispositif de formation.

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	8 813 €
Selon la branche	7 697 € – 9 800 €
Prise en charge	Branches concernées

8 813 €	Valeur par défaut · 188 branches
9 800 €	CPNE Casinos, CPNE autoroutes, 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
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
8 500 €	CPNE Branche Caoutchouc, CPNEFP de la métallurgie
9 005 €	CPNE FP Branche des magasins du Bricolage
9 000 €	CPNE Banque
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
7 697 €	CPNEFP de la blanchisserie, laverie, location de linge, nettoyage à sec, pressing et teinturerie

Base légale

Date de décision	28/02/2025
Durée de l'enregistrement	5 ans
Date d'échéance de l'enregistrement	28/02/2030
Date de dernière délivrance possible	28/02/2034

Statistiques

Année	Certifiés	Dont VAE	Insertion 6 mois	Métier 2 ans
2022	31	0	81 %	90 %
2021	23	0	95 %	100 %
2020	17	0	75 %	86 %

Liens avec d'autres certifications

ANCIENNES VERSIONS

RNCP36398

CORRESPONDANCES (PAR BLOCS)

RNCP34262, RNCP36398, RNCP36921, RNCP37624, RNCP37638, RNCP37763

Pour plus d'informations

Fiche officielle (France Compétences)

<https://www.francecompetences.fr/recherche/rncp/40234/>

Où se former

Formations référencées	3
Organismes	1
Régions couvertes	2

Prix indicatif	7 000 € – 10 500 €
Durée	1062 h

Offre de formation — MonCompteFormation (Caisse des Dépôts), licence ouverte.

Entrées et sorties de formation

Entrées en formation	2
Parcours menés à terme	0
Taux d'achèvement	0 %
Période	09/2025

Entrées et sorties — Caisse des Dépôts (open data), tous organismes confondus.

Demande CPF — formations engagées

Dossiers CPF engagés	2
Montant total engagé	14 500 €
Coût moyen par dossier	7 250 €
Durée moyenne	1062 h
Période	08/2025 – 09/2025
Régions	52 %
France compétences	30 %
Reste à charge titulaire	18 %

Formations engagées via le CPF — Caisse des Dépôts (open data), tous organismes confondus.