

Concepteur manager des infrastructures de données massives

RNCP37750 · RNCP · Publiée

L'essentiel

Niveau de qualification	Niveau 7
Code(s) NSF	326
Formacode(s)	31009, 31052, 31094, 31006
Code(s) ROME	M1805, M1806, M1803, M1802
Type d'enregistrement	Enregistrement sur demande
Date d'enregistrement	19/07/2023
Date d'échéance	19/07/2026

Certificateur(s)

EXCELIA GROUP · SIRET : 34876850800015<https://www.excelia-group.com>

Partenaires habilités

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

Organisme	SIRET	Habilitation
PSTB-PARIS SCHOOL OF TECHNOLOGY & BUSINESS EN ABREGE P.S.T.B	812095586000 44	Former et organiser l'évaluation

Résumé de la certification

OBJECTIFS ET CONTEXTE

La certification « Concepteur-Manager des infrastructures de données massives » prépare le titulaire à l'exercice d'activités requérant des savoirs et savoir-faire techniques et technologiques hautement spécialisés dans le domaine de la conception et du management de solutions pour l'extraction, le traitement et l'entreposage de données massives. Ces professionnels interviennent en amont des Data scientifiques et des Data analystes avec lesquels ils sont amenés à collaborer activement. Le rôle des titulaires de la certification consiste à concevoir puis développer les infrastructures ou pipeline de données et déployer ensuite des solutions techniques robustes qui permettent aux utilisateurs d'accéder à des données pertinentes et fiables via des plateformes logicielles et des outils d'interface (API). Ces données serviront à éclairer les analyses et les prises de décision métiers (Direction Générale, Marketing, Finances, RH, logistique, ...).

ACTIVITÉS VISÉES

- Veille scientifique, technologique et réglementaire, identification de l'état de l'art des outils et plateformes logiciels ETL
- Spécification des besoins et dimensionnement des infrastructures de données massives dans une perspective pluriannuelle
- Evaluation et sélection d'une solution ETL globale incluant des développements en propre, des plateformes logicielles et des API disponibles sur le marché
- Développement d'une infrastructure et management des flux et des stocks de données dans le respect du RGAA et du RGPD

- Identification et sélection des solutions data lake les plus performantes sur le marché
- Déploiement d'outils d'extraction et de stockage de données brutes en interne ou dans le cloud
- Management des opérations de traitement des données brutes avant stockage dans un data lake
- Supervision de la rédaction des documents supports à l'utilisation des outils et plateformes logicielles
- Management de la mise en place des bases de données internes et externes dans le respect du RGPD et des critères du RGAA
- Evaluation et sélection d'une solution de traitement et d'entreposage des données
- Déploiement des processus de traitement et de formatage des données brutes pour entreposage
- Mise à disposition de données fiables et pertinentes pour les utilisateurs
- Elaboration d'un guide d'utilisation des plateformes et des API déployées
- Elaboration d'un cahier des charges technique et fonctionnel répondant aux besoins des scientifiques et des analystes
- Pilotage du développement d'API, de requêtes et de routines automatisées avec implication de l'ensemble des parties prenantes
- Anticipation des besoins utilisateurs et amélioration continue des plateformes logicielles ETL
- Mise en œuvre d'indicateurs de performance des projets sur l'ensemble de leur cycle de vie
- Management des processus d'automatisation et de fiabilisation des infrastructures, des plateformes et des interfaces
- Contribution à l'amélioration de la qualité des données mises à dispositions des utilisateurs
- Conception et mise en œuvre de Tests d'intégration avant déploiement d'algorithmes d'IA, de nouvelles plateformes logicielles ou d'API

COMPÉTENCES ATTESTÉES

- Mettre en œuvre une veille scientifique, technologique et réglementaire concernant les outils et plateformes logiciels d'extraction de traitement et de stockage (ETL) de données massives;
- Anticiper les besoins futurs pour concevoir et dimensionner une architecture compatible avec la montée en charge des cas d'usage notamment ceux liés à l'IA ;
- Mobiliser des savoirs techniques hautement spécialisés en plateformes et outils logiciels pour évaluer différentes solutions ETL assurant la collecte, le nettoyage et l'entreposage de données massives internes ou externes structurées, semi-structurées et non structurées;
- Recueillir les besoins des différents cas d'usage notamment ceux liés à l'intelligence artificielle pour développer une architecture de flux de données en établissant un cahier des charges fonctionnel et technique ;
- Concevoir une infrastructure fonctionnelle et matérielle pour organiser les flux et les traitements de données depuis des sources multivariées internes et externes jusqu'à un utilisateur final (DG, Direction marketing, Data analystes, data scientifiques ...) ;
- Mobiliser les principes de la RSE et le RGPD pour participer à l'établissement des règles de gouvernance et de conservation des données et garantir l'accessibilité des data selon les normes du RGAA;
- Recenser et caractériser les outils d'extraction et de stockage de données brutes disponibles sur le marché pour sélectionner la solution la plus adaptée aux besoins de l'entreprise ;
- Mobiliser des savoirs techniques approfondis en déploiement de bases de données pour évaluer les mérites et les limites d'un stockage en propre par rapport à une solution déportée sur le cloud;
- Mettre en œuvre un processus complet d'extraction pour recueillir des données issues de sources multiples et les stocker dans une base de données brutes (lac) centralisée ;
- Identifier et préparer toutes les données pertinentes afin de faciliter leur traitement et leur stockage ultérieurs dans un entrepôt de données ;
- Documenter l'utilisation des bases de données externes et internes pour gérer et entretenir le lac de données en respectant le RGPD et en garantissant l'accessibilité des données selon les normes du RGAA;
- Synthétiser les besoins des utilisateurs pour concevoir des modèles de données performants et résilients afin de faciliter les travaux de data science et de datavisualisation ;
- Mobiliser des savoirs techniques avancés en gestion de bases de données pour sélectionner et mettre en œuvre les solutions de traitement et de stockage massifs adaptées au besoin de l'entreprise ;
- Veiller à l'application des critères du RGAA dans la mise en œuvre des solutions pour garantir l'accessibilité des données aux PSH ;
- Trier, nettoyer et structurer des données brutes pour fabriquer des données propres exploitables par les data scientifiques et les data analystes ;

- Organiser le stockage des données dans un entrepôt de manière à assurer la gestion de la data durant la totalité de son cycle de vie conformément aux directives RGPD ;
- Mettre en place des solutions innovantes de types API pour améliorer le temps de mise à disposition de données fiables et exploitables par les data analystes ;
- Etablir un cahier des charges technique et fonctionnel pour répondre aux besoins évolutifs des data scientifiques, data analystes, partenaires métiers et de la direction générale de l'entreprise ;
- Piloter des projets en mobilisant des équipes pluridisciplinaires (data scientifiques, responsable marketing, direction générale, ...) pour développer des interfaces types API et des requêtes automatisées à destination de cas d'usage multivariés ;
- Mobiliser des connaissances approfondies en programmation pour développer et superviser le développement des routines informatiques en mode DEV/OPS[1] et réduire le temps de mise à disposition de nouvelles fonctionnalités utiles à l'utilisateur des données ;
- Collaborer avec les chefs de produits, les responsables métiers et la direction générale afin d'anticiper les nouveaux usages de la data et élaborer des nouvelles fonctionnalités ;
- Optimiser en continu les processus ETL pour proposer des solutions accessibles aux PSH et plus efficaces en termes de consommation d'énergie et de ressources ;
- Mettre en place des indicateurs de performance projet pour assurer la production de livrables au niveau de qualité attendu, dans les délais et les budgets fixés par le cahier des charges ;
- Automatiser et fiabiliser les processus d'extraction et de traitement pour réduire le temps de mise à disposition de données fiables et pertinentes ;
- Fiabiliser une infrastructure, des plateformes logicielles et des API, pour mettre des données fiables et pertinentes à disposition des utilisateurs en anticipant les opérations de maintenance et de mise à jour ;
- Réaliser des tests d'intégration avant déploiement et organiser des retours sur expérience utilisateurs pour contribuer à l'assurance qualité de l'architecture et de l'infrastructure des mouvements et stockages de données ;
- Mobiliser des savoirs hautement spécialisés pour évaluer et anticiper les risques concernant la sécurité des données en lien avec le RSSI et la protection des données en lien avec le DPO ;
- Assurer la fourniture de données propres et pertinentes pour faciliter l'apprentissage machine et la science des données en collaboration avec les data scientifiques ;
- Déployer à grande échelle de nouveaux algorithmes conçus en laboratoire par les data scientifiques pour mieux modéliser les relations entre les données ;
- Prendre en compte les risques sur le plan de l'éthique et la RSE vis-à-vis des technologies d'intelligence artificielle pour définir les critères qualité d'une infrastructure data fiabilisée et industrialisée.

Blocs de compétences

RNCP37750BC01 — Concevoir et développer une architecture fonctionnelle et matérielle pour assurer l'acquisition, le traitement et le stockage de données massives

LISTE DE COMPÉTENCES

Mettre en œuvre une veille scientifique, technologique et réglementaire concernant les outils et plateformes logiciels d'extraction de traitement et de stockage (ETL) de données massives ;

Anticiper les besoins futurs pour concevoir et dimensionner une architecture compatible avec la montée en charge des cas d'usage notamment ceux liés à l'IA ;

Mobiliser des savoirs techniques hautement spécialisés en plateformes et outils logiciels pour évaluer différentes solutions ETL assurant la collecte, le nettoyage et l'entreposage de données massives internes ou externes structurées, semi-structurées et non structurées ;

Recueillir les besoins des différents cas d'usage notamment ceux liés à l'intelligence artificielle pour développer une architecture de flux de données en établissant un cahier des charges fonctionnel et technique ;

Concevoir une infrastructure fonctionnelle et matérielle pour organiser les flux et les traitements de données depuis des sources multivariées internes et externes jusqu'à un utilisateur final (DG, Direction marketing, Data analystes, data scientifiques ...) ;

Mobiliser les principes de la RSE et le RGPD pour participer à l'établissement des règles de gouvernance et de conservation des données et garantir l'accessibilité des data selon les normes du RGAA.

MODALITÉS D'ÉVALUATION

Webographie / bibliographie professionnelle - Etude de cas - Projet applicatif tutoré

RNCP37750BC02 — Déployer des outils d'extraction et de stockage pour recueillir des données brutes structurées, non structurées, internes ou externes dans un lac de données

LISTE DE COMPÉTENCES

Recenser et caractériser les outils d'extraction et de stockage de données brutes disponibles sur le marché pour sélectionner la solution la plus adaptée aux besoins de l'entreprise ;
Mobiliser des savoirs techniques approfondis en déploiement de bases de données pour évaluer les mérites et les limites d'un stockage en propre par rapport à une solution déportée sur le cloud ;
Mettre en œuvre un processus complet d'extraction pour recueillir des données issues de sources multiples et les stocker dans une base de données brutes (lac) centralisée ;
Identifier et préparer toutes les données pertinentes afin de faciliter leur traitement et leur stockage ultérieurs dans un entrepôt de données ;
Documenter l'utilisation des bases de données externes et internes pour gérer et entretenir le lac de données en respectant le RGPD et en garantissant l'accessibilité des données selon les normes du RGAA.

MODALITÉS D'ÉVALUATION

Etude de cas - Travaux pratiques

RNCP37750BC03 — Mettre en œuvre des solutions informatiques pour traiter et entreposer des données exploitables par l'intelligence artificielle et l'analyse humaine

LISTE DE COMPÉTENCES

Synthétiser les besoins des utilisateurs pour concevoir des modèles de données performants et résilients afin de faciliter les travaux de data science et de datavisualisation ;
Mobiliser des savoirs techniques avancés en gestion de bases de données pour sélectionner et mettre en œuvre les solutions de traitement et de stockage massifs adaptées au besoin de l'entreprise ;
Veiller à l'application des critères du RGAA dans la mise en œuvre des solutions pour garantir l'accessibilité des données aux PSH ;
Trier, nettoyer et structurer des données brutes pour fabriquer des données propres exploitables par les data scientists et les data analysts ;
Organiser le stockage des données dans un entrepôt de manière à assurer la gestion de la data durant la totalité de son cycle de vie conformément aux directives RGPD ;
Mettre en place des solutions innovantes de types API pour améliorer le temps de mise à disposition de données fiables et exploitables par les data analysts.

MODALITÉS D'ÉVALUATION

Etude de cas - Travaux pratiques

RNCP37750BC04 — Piloter des projets pour anticiper les besoins et développer en continu les plateformes logicielles (ETL) et les interfaces de programmation applicatives (API)

LISTE DE COMPÉTENCES

Etablir un cahier des charges technique et fonctionnel pour répondre aux besoins évolutifs des data scientists, data analysts, partenaires métiers et de la direction générale de l'entreprise ;
Piloter des projets en mobilisant des équipes pluridisciplinaires (data scientists, responsable marketing, direction générale, ...) pour développer des interfaces types API et des requêtes automatisées à destination de cas d'usage multivariés ;
Mobiliser des connaissances approfondies en programmation pour développer et superviser le développement des routines informatiques en mode DEV/OPS[1]
et réduire le temps de mise à disposition de nouvelles fonctionnalités utiles à l'utilisateur des données ;
Collaborer avec les chefs de produits, les responsables métiers et la direction générale afin d'anticiper les nouveaux usages de la data et élaborer des nouvelles fonctionnalités ;
Optimiser en continu les processus ETL pour proposer des solutions accessibles aux PSH et plus efficaces en termes de consommation d'énergie et de ressources ;
Mettre en place des indicateurs de performance projet pour assurer la production de livrables au niveau de qualité attendu, dans les délais et les budgets fixés par le cahier des charges ;

MODALITÉS D'ÉVALUATION

Hackathon 1 - Travaux pratiques - Projet applicatif tutoré

RNCP37750BC05 — Manager l'industrialisation des mouvements de data au sein de l'entreprise pour mettre à disposition des utilisateurs des données fiables et pertinentes

LISTE DE COMPÉTENCES

Automatiser et fiabiliser les processus d'extraction et de traitement pour réduire le temps de mise à disposition de données fiables et pertinentes ;

Fiabiliser une infrastructure, des plateformes logicielles et des API, pour mettre des données fiables et pertinentes à disposition des utilisateurs en anticipant les opérations de maintenance et de mise à jour ;

Réaliser des tests d'intégration avant déploiement et organiser des retours sur expérience utilisateurs pour contribuer à l'assurance qualité de l'architecture et de l'infrastructure des mouvements et stockages de données ;

Mobiliser des savoirs hautement spécialisés pour évaluer et anticiper les risques concernant la sécurité des données en lien avec le RSSI et la protection des données en lien avec le DPO ;

Assurer la fourniture de données propres et pertinentes pour faciliter l'apprentissage machine et la science des données en collaboration avec les data scientists ;

Déployer à grande échelle de nouveaux algorithmes conçus en laboratoire par les data scientists pour mieux modéliser les relations entre les données ;

Prendre en compte les risques sur le plan de l'éthique et la RSE vis-à-vis des technologies d'intelligence artificielle pour définir les critères qualité d'une infrastructure data fiabilisée et industrialisée.

MODALITÉS D'ÉVALUATION

Travaux pratiques - Hackathon 2

Secteur d'activité et type d'emploi

SECTEURS D'ACTIVITÉS

Initiée au milieu des années 2000 dans les grands groupes du secteur banque et assurances, l'exploitation de données massives à des fins prédictives et prescriptives se développe aujourd'hui au sein de plus petites structures de types ETI et PME dans des secteurs très variés : bancassurance, grande distribution, énergie, industries, transports, services... Le secteur de la data est également structuré à partir d'un écosystème comprenant des startups technologiques (TPE), des agences de conseil en data (TPE et PME), des Entreprises de Services Numériques (ESN) spécialisées en data (PME et ETI) et des éditeurs de plateformes logicielles (grands groupes informatiques internationaux et PME) spécifiques au secteur de la data.

TYPE D'EMPLOIS ACCESSIBLES

Data Engineer - Data Architecte - Expert en infrastructure data - Ingénieur DEV / OPS Data - Ingénieur Big Data

RÉGLEMENTATIONS D'ACTIVITÉ

Il n'existe pas de réglementation spécifique au métier de data engineer

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

Être Titulaire d'une certification de Niveau 6, d'un Bachelor ou d'une licence en développement data et logiciel, informatique, ou formations similaires

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

Au moins quatre membres composent le jury de certification :

- Un président de jury : personnalité qualifiée du monde économique ;
- Au moins deux professionnels issus du secteur du titre visé ;
- Au moins un représentant de la certification.

Au minimum, 50% des membres sont extérieurs à l'autorité délivrant la certification et représentent le secteur professionnel.

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	7 500 € – 12 000 €
Prise en charge	Branches concernées
9 157 €	Valeur par défaut · 184 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
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
9 800 €	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
8 500 €	CPNEFP de la métallurgie
8 270 €	CPNEFP de la Branche des Caisses régionales de Crédit Agricole et autres organismes
8 000 €	CPNE chimie
7 500 €	CPNE des industries de santé (CPNEIS)

Base légale

Date de décision	19/07/2023
Durée de l'enregistrement	3 ans
Date d'échéance de l'enregistrement	19/07/2026
Date de dernière délivrance possible	19/07/2030

Liens avec d'autres certifications

CORRESPONDANCES (PAR BLOCS)

RNCP32123, RNCP36398, RNCP37172, RNCP37422, RNCP37638

Pour plus d'informations

LIEN VERS LE DESCRIPTIF

www.excelia-group.com

Fiche officielle (France Compétences)

<https://www.francecompetences.fr/recherche/rncp/37750/>

Entrées et sorties de formation

Entrées en formation	8
Parcours menés à terme	8
Taux d'achèvement	100 %

Demande CPF — formations engagées

Dossiers CPF engagés	8
Montant total engagé	27 130 €
Coût moyen par dossier	3 391 €
Durée moyenne	162 h
Période	08/2025 – 12/2025
France compétences	84 %
France Travail	13 %
Reste à charge titulaire	3 %

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