

Expert en ingénierie des données

RNCP42157 · RNCP · Publiée

L'essentiel

Niveau de qualification	Niveau 7
Code(s) NSF	326
Formacode(s)	31028, 31023, 31094
Code(s) ROME	M1802, M1805
Type d'enregistrement	Enregistrement sur demande
Date d'enregistrement	30/04/2026
Date d'échéance	30/04/2029

Certificateur(s)

ASSOCIATION INTERNATIONALE POUR LA FORMATION · SIRET : 33037752400047

ASSOCIATION POUR LA FORMATION A L'INFORMATIQUE ET AU NUMERIQUE · SIRET : 39350481600355

Partenaires habilités

14 organismes habilités, d'après la liste publiée par France compétences : 14 à former et organiser l'évaluation.

Organisme	SIRET	Habilitation
ASSOCIATION POUR LA FORMATION A L'INFORMATIQUE ET AU NUMERIQUE	39350481600330	Former et organiser l'évaluation
ASSOCIATION POUR LA FORMATION A L'INFORMATIQUE ET AU NUMERIQUE	39350481600355	Former et organiser l'évaluation
ASSOCIATION POUR LA FORMATION A L'INFORMATIQUE ET AU NUMERIQUE (AFINUM)	39350481600090	Former et organiser l'évaluation
ASSOCIATION POUR LA FORMATION A L'INFORMATIQUE ET AU NUMERIQUE (AFINUM)	39350481600215	Former et organiser l'évaluation
ASSOCIATION POUR LA FORMATION A L'INFORMATIQUE ET AU NUMERIQUE (AFINUM)	39350481600223	Former et organiser l'évaluation
ASSOCIATION POUR LA FORMATION A L'INFORMATIQUE ET AU NUMERIQUE (AFINUM)	39350481600249	Former et organiser l'évaluation
ASSOCIATION POUR LA FORMATION A L'INFORMATIQUE ET AU NUMERIQUE (AFINUM)	39350481600272	Former et organiser l'évaluation
ASSOCIATION POUR LA FORMATION A L'INFORMATIQUE ET AU NUMERIQUE (AFINUM)	39350481600306	Former et organiser l'évaluation
ASSOCIATION POUR LA FORMATION A L'INFORMATIQUE ET AU NUMERIQUE (AFINUM)	39350481600314	Former et organiser l'évaluation
ASSOCIATION POUR LA FORMATION A L'INFORMATIQUE ET AU NUMERIQUE (AFINUM)	39350481600322	Former et organiser l'évaluation

Liste complète des 14 partenaires habilités : <https://data.aktaconseil.fr/certifications/rncp42157-expert-en-ingenierie-des-donnees/partenaires>

Résumé de la certification

OBJECTIFS ET CONTEXTE

Le secteur de la donnée connaît une mutation sans précédent : la numérisation généralisée, la prolifération des usages connectés et l'émergence de technologies de pointe ont fait passer la donnée du rang de simple ressource technique à celui d'actif stratégique et moteur de compétitivité. Face à l'explosion des données produites et consommées, à l'essor de l'intelligence artificielle et à des exigences légales, réglementaires et normatives toujours plus strictes, l'ingénierie des données s'impose comme un pilier incontournable de la transformation numérique. La valorisation des données devient un levier de développement décisif : véritable matière première, la donnée nourrit la prise de décision, réduit l'incertitude et ouvre de nouveaux horizons économiques.

Les organisations, qu'elles soient publiques ou privées, doivent désormais relever des défis stratégiques et techniques : absorber et traiter en temps réel des volumes massifs d'informations, garantir la qualité, la traçabilité et la sécurité, respecter une mosaïque de réglementations, tout en maîtrisant l'empreinte environnementale de leurs infrastructures numériques.

Dans ce contexte, la certification professionnelle « Expert en Ingénierie des Données » a pour objectif de doter les candidats des compétences nécessaires pour concevoir et faire évoluer des architectures de systèmes de données capables de traiter des flux massifs et hétérogènes. Elle prépare au pilotage de l'optimisation des pipelines de traitement, qu'ils soient en mode batch ou temps réel, afin de garantir la performance et la fiabilité des systèmes ; elle atteste de la capacité à assurer la qualité, la résilience et la disponibilité des données sur l'ensemble de leur cycle de vie, tout en pilotant leur gouvernance et leur conformité réglementaire, légale et normative, notamment en matière d'archivage, de purge et d'anonymisation. Enfin, elle certifie l'aptitude à identifier et à valoriser les gisements de données à fort potentiel, transformant l'information brute en un levier d'innovation et de compétitivité durable.

ACTIVITÉS VISÉES

Modélisation des architectures data orientées usages métiers
Pilotage de la sécurisation et de la mise en conformité des données
Développement des pipelines de traitement de données
Intégration des sources de données hétérogènes
Gestion des infrastructures de stockage et de traitement des données
Optimisation des coûts et des performances des infrastructures
Conduite du projet de développement d'une plateforme Big Data
Coordination de l'équipe projet
Définition de la stratégie de gouvernance des données
Définition et implémentation d'une stratégie de valorisation des données

COMPÉTENCES ATTESTÉES

Diagnostiquer les besoins « Data » des directions métiers de l'organisation, en réalisant des enquêtes ou interviews et en analysant la structure de l'organisation et ses implications sur la circulation des données, afin de concevoir la/les plateformes de données adaptées aux besoins et attentes des directions métiers et de formaliser les éventuelles contraintes associées (organisationnelles, techniques, budgétaires, réglementaires, ...).

Concevoir une architecture hybride de collecte et de restitution des données, en identifiant les sources de données appropriées, en structurant, schématisant et en documentant les processus de transformation et de chargement, afin d'en garantir la fiabilité et l'évolutivité de l'architecture.

Concevoir des modèles de données, en identifiant les entités et leurs relations, en définissant les attributs et les types de données, et en appliquant les règles de normalisation, afin de garantir l'intégrité, la cohérence et l'efficacité des systèmes d'information (SI) et des usages analytiques.

Concevoir un Data Lake en intégrant des solutions de stockage évolutives, responsables, éco-responsables et sécurisées, et en mettant en place des pipelines d'ingestion et de traitement des données massives, afin de rendre accessibles et interopérables des volumes importants de données hétérogènes, et de permettre des analyses avancées exploitable par différents services de l'organisation.

Elaborer des plans de migration de données, en s'appuyant sur une analyse approfondie des environnements sources et cibles, en définissant une stratégie de migration adaptée aux contraintes techniques, organisationnelles et réglementaires et en mobilisant des solutions d'automatisation, afin de garantir l'intégrité des données.

Superviser les dispositifs de sécurité des données conformément aux orientations définies par le RSSI, en intégrant des protocoles de chiffrement adaptés à la sensibilité des données et aux exigences réglementaires, et des mécanismes avancés de contrôle d'accès, afin de garantir la confidentialité des données dans un environnement

organisationnel complexe.

Élaborer une stratégie de sauvegarde et de récupération des données, en identifiant les exigences de disponibilité, de criticité et de conformité des données, en sélectionnant les technologies et outils de sauvegarde adaptés au contexte opérationnel, et en définissant des procédures de récupération robustes, afin de garantir la continuité d'activité.

Concevoir des modèles prédictifs et des algorithmes de machine learning en mobilisant des techniques avancées de traitement des données et des outils d'analyse statistique, afin d'identifier les anomalies et d'optimiser la qualité et la fiabilité des jeux des données.

Structurer un dispositif de veille technique, réglementaire et technologique centré sur les innovations liées à la gestion et à l'exploitation des données, en sélectionnant des sources d'informations en français et en anglais, afin de formuler des recommandations et de permettre la prise de décisions techniques relatives aux projets d'ingénierie des données.

Concevoir des pipelines ETL robustes, en modélisant les flux de données, en utilisant des outils d'orchestration et des technologies de traitement par lots ou en temps réel, afin d'assurer la cohérence et la disponibilité des données pour les analyses et les processus métiers.

Implémenter un pipeline de traitement de données, en intégrant des outils de collecte et de transformation par lot ou en streaming, en configurant des systèmes d'ingestion pour les données structurées et non structurées, afin d'assurer l'efficacité, la sécurité et la scalabilité des flux de données conformément à l'architecture définie.

Optimiser les performances des pipelines de données, en identifiant les goulots d'étranglement et en implémentant des solutions techniques telles que la parallélisation des tâches, l'optimisation des requêtes SQL et l'utilisation de technologies de traitement distribué, afin de garantir l'efficacité, la scalabilité et la réactivité des processus de traitement des données.

Développer des tests de surveillance des pipelines de données, en identifiant les points de contrôle stratégiques et en déployant des outils de monitoring

automatisés, afin de garantir l'intégrité et la traçabilité des données tout au long du cycle de vie des pipelines.

Configurer des pipelines MLOps en intégrant des outils d'automatisation et en paramétrant les environnements de déploiement, afin d'assurer la mise en production fiable, reproductible et scalable des modèles d'IA.

Concevoir une architecture de migration inter-systèmes et multi-environnements, en utilisant un outil de gestion des données adapté aux besoins spécifiques du projet, afin de garantir l'intégration harmonieuse de données hétérogènes et l'interopérabilité optimale des différentes sources de données.

Elaborer des architectures de traitement des données en temps réel, en intégrant des technologies de streaming et en configurant des pipelines de données, afin de garantir la réactivité, la scalabilité et la robustesse des systèmes de gestion des données.

Développer des pipelines de traitement de données, en intégrant les services et fonctionnalités des architectures Data Lake, en appliquant les pratiques de gestion du cycle de vie des données, afin de maîtriser le parcours de la donnée et de faciliter l'exploitation des données par les différents usages analytiques et opérationnels.

Organiser les sources de données, en utilisant des outils adaptés de data visualisation, en sélectionnant les indicateurs clés (KPI) et en structurant les informations de manière concise, afin de faciliter la prise de décision des directions métiers.

Documenter les sources de données et les processus d'intégration, en élaborant des schémas détaillés et des descriptions précises des flux de données et en utilisant des outils de modélisation, afin de maintenir la traçabilité et la transparence des données.

Concevoir une architecture de stockage intégrant des systèmes de gestion de bases de données optimisés au regard des besoins du projet, en respectant les normes de sécurité définies, et en utilisant des techniques de nettoyage et d'amélioration des données, afin de faciliter l'exploitation, l'analyse et la valorisation des données.

Déployer une architecture de stockage distribuée, en configurant et en administrant des clusters de nœuds et en optimisant les ressources matérielles et logicielles, afin de garantir la continuité de service dans un environnement distribué.

Optimiser la configuration de la plateforme de stockage de données, en ajustant les paramètres de performance, en dimensionnant la capacité de stockage et en maximisant la bande passante, afin de garantir une disponibilité et une efficacité optimales des ressources pour les utilisateurs.

Mettre en place un système d'auto scaling des ressources en configurant des outils de surveillance et des algorithmes d'optimisation afin d'anticiper les montées en charge.

Assurer la maintenance et la mise à jour des infrastructures de stockage et de traitement des données, en utilisant des outils de monitoring et de diagnostic, en appliquant les correctifs et les mises à jour nécessaires, afin de soutenir durablement les besoins métiers et techniques de l'organisation, tout en optimisant l'efficacité énergétique et en

réduisant l'empreinte environnementale des infrastructures.

Analyser les coûts et les performances de stockage des données, en utilisant des outils de monitoring, des indicateurs clés de performance et en tenant compte des licences, des supports et de l'évolutivité, afin de garantir la pérennité de la solution de stockage des données.

Déployer des infrastructures de cloud computing, en configurant les environnements de calcul et de stockage, en utilisant des outils de gestion et d'orchestration, afin de répondre aux impératifs de performance des environnements cloud de l'organisation.

Développer des scripts d'automatisation des tâches répétitives en utilisant des langages de programmation adaptés et des outils de gestion des flux de travail afin d'améliorer la productivité des processus de traitement des données et de réduire les tâches manuelles.

Effectuer des audits des systèmes de gestion des données en recourant à des outils d'analyse et des indicateurs de performance, afin de répondre aux standards de conformité réglementaires et normatifs et de formuler des recommandations aux parties prenantes.

Évaluer l'empreinte environnementale des processus de traitement des données à l'aide d'outils de mesure et d'indicateurs, afin d'optimiser l'usage des ressources et de réduire l'empreinte environnementale des solutions déployées.

Planifier le projet de développement de plateforme Big Data, en s'appuyant sur les besoins utilisateurs collectés et en répartissant les activités en fonction des ressources humaines, techniques et financières, afin d'identifier l'ensemble des étapes de réalisation et d'organiser le projet en tâches et livrables.

Gérer un projet, en utilisant les méthodes agiles et outils adaptés, dans le but d'optimiser les délais de livraison et d'assurer la bonne exécution du projet de solution Big Data.

Réaliser le reporting du projet auprès du commanditaire, en élaborant des tableaux de bord de suivi des performances, dans le but d'analyser l'utilisation des ressources à chaque étape, d'anticiper et de corriger les écarts en temps réel et de limiter les contraintes et les retards dans l'exécution du projet.

Piloter les prestataires extérieurs gérant les ressources informatiques du SI existant (listées dans la cartographie établie), en présentant le projet, les rôles et objectifs de l'ensemble des parties prenantes et en s'assurant de la pertinence des contrats existants, afin d'identifier les KPI, de calibrer le suivi et de sécuriser la mise en œuvre technique.

Conduire l'équipe projet, en favorisant l'inclusion des personnes en situation de handicap et en adaptant les conditions de travail face aux contraintes de temps et aux incertitudes, afin de renforcer la motivation, la résilience et l'équilibre entre vie professionnelle et personnelle.

Mettre en œuvre un processus de communication inclusif régulier au sein de l'équipe, en adaptant les modes de communication selon le contexte, les cultures et la langue des collaborateurs, afin de synchroniser les activités quotidiennes et de garantir l'intégration de tous les membres de l'équipe.

Proposer des solutions innovantes, en organisant des séances de brainstorming, en recueillant des retours réguliers et en encourageant l'autonomie ou la prise d'initiative, afin de favoriser les interactions au sein de l'équipe, de maintenir une dynamique de groupe et de renforcer l'esprit d'équipe au service du projet.

Établir un processus de partage d'information, en identifiant les besoins des parties prenantes, en sélectionnant l'outil adapté et en établissant des protocoles et procédures de partage, afin de faciliter l'inclusion, la collaboration entre les membres de l'équipe projet et de sécuriser l'information liée au projet.

Analyser la stratégie de gouvernance des données de l'organisation et son fonctionnement interne, en mobilisant des outils de diagnostic, des indicateurs de performance et des techniques d'analyse exploratoire des données, afin de produire un diagnostic du système de gouvernance des données et de formuler des recommandations stratégiques adaptées.

Analyser les sources de données et les traitements associés, en mobilisant des outils de profilage, de nettoyage, de data science et de visualisation, afin d'évaluer la complétude, la cohérence, et la conformité des données, et de produire un rapport d'audit détaillé.

Cartographier les flux, les sources, les zones critiques et les données sensibles, à l'aide d'outils de data visualisation et d'analyse des risques liés à la gestion et à l'exploitation des données, afin d'évaluer le patrimoine Data de l'organisation, d'identifier les risques et de proposer des mesures correctives ou préventives adaptées.

Élaborer un cadre de gouvernance durable des infrastructures de données, en définissant les règles de gestion des données de référence et en mettant en place des processus de validation et de contrôle de qualité afin d'assurer la pérennité des infrastructures de données, de garantir la cohérence et la qualité des informations utilisées par l'organisation et de soutenir une exploitation optimale et sécurisée du patrimoine Data.

Élaborer des indicateurs de performance Green IT, en analysant les impacts environnementaux et sociaux des

ressources informatiques, en utilisant des outils de mesure et des méthodologies adaptés, afin de favoriser une utilisation responsable et durable des ressources numériques et de soutenir la performance globale de l'organisation dans sa transition écologique.

Élaborer une stratégie Big Data alignée sur les objectifs de l'organisation, en collaboration avec la DSI et en définissant les orientations en matière de gouvernance des données, d'intégration des analyses dans les processus décisionnels et de pilotage de la performance, afin de structurer une gouvernance data-driven de l'organisation.

Organiser la mise à disposition des données auprès des directions métiers, en assurant leur accessibilité, leur contextualisation et leur alignement avec les besoins opérationnels et stratégiques, afin de renforcer l'autonomie décisionnelle des utilisateurs et de faciliter l'appropriation d'une culture data-driven.

Animer des sessions de formation sur les outils d'analyse des données, de Data Literacy, et les pratiques Green IT associées, en élaborant des supports pédagogiques accessibles et en utilisant des méthodes interactives, afin de développer l'autonomie des utilisateurs et d'optimiser l'exploitation des outils.

Blocs de compétences

RNCP42157BC01 — Elaborer et concevoir l'architecture des systèmes de données

LISTE DE COMPÉTENCES

Diagnostiquer les besoins « Data » des directions métiers de l'organisation, en réalisant des enquêtes ou interviews et en analysant la structure de l'organisation et ses implications sur la circulation des données, afin de concevoir la/les plateformes de données adaptées aux besoins et attentes des directions métiers et de formaliser les éventuelles contraintes associées (organisationnelles, techniques, budgétaires, réglementaires, ...).

Concevoir une architecture hybride de collecte et de restitution des données, en identifiant les sources de données appropriées, en structurant, schématisant et en documentant les processus de transformation et de chargement, afin d'en garantir la fiabilité et l'évolutivité de l'architecture.

Concevoir des modèles de données, en identifiant les entités et leurs relations, en définissant les attributs et les types de données, et en appliquant les règles de normalisation, afin de garantir l'intégrité, la cohérence et l'efficacité des systèmes d'information (SI) et des usages analytiques.

Concevoir un Data Lake en intégrant des solutions de stockage évolutives, responsables, éco-responsables et sécurisées, et en mettant en place des pipelines d'ingestion et de traitement des données massives, afin de rendre accessibles et interopérables des volumes importants de données hétérogènes, et de permettre des analyses avancées exploitable par différents services de l'organisation.

Elaborer des plans de migration de données, en s'appuyant sur une analyse approfondie des environnements sources et cibles, en définissant une stratégie de migration adaptée aux contraintes techniques, organisationnelles et réglementaires et en mobilisant des solutions d'automatisation, afin de garantir l'intégrité des données.

Superviser les dispositifs de sécurité des données conformément aux orientations définies par le RSSI, en intégrant des protocoles de chiffrement adaptés à la sensibilité des données et aux exigences réglementaires, et des mécanismes avancés de contrôle d'accès, afin de garantir la confidentialité des données dans un environnement organisationnel complexe.

Élaborer une stratégie de sauvegarde et de récupération des données, en identifiant les exigences de disponibilité, de criticité et de conformité des données, en sélectionnant les technologies et outils de sauvegarde adaptés au contexte opérationnel, et en définissant des procédures de récupération robustes, afin de garantir la continuité d'activité.

Concevoir des modèles prédictifs et des algorithmes de machine learning en mobilisant des techniques avancées de traitement des données et des outils d'analyse statistique, afin d'identifier les anomalies et d'optimiser la qualité et la fiabilité des jeux des données.

Structurer un dispositif de veille technique, réglementaire et technologique centré sur les innovations liées à la gestion et à l'exploitation des données, en sélectionnant des sources d'informations en français et en anglais, afin de formuler des recommandations et de permettre la prise de décisions techniques relatives aux projets d'ingénierie des données.

MODALITÉS D'ÉVALUATION

Mise en situation professionnelle reconstituée

Conception d'une architecture de système de données sur la base d'un cahier des charges issu d'un cas réel ou fictif.

Phase 1 : Préparation de la conception de l'architecture

Préparation tutorée par groupe de projet.

Phase 2 : Production écrite en groupe de projet

Le candidat devra produire un dossier rédigé comprenant :

- Une évaluation documentée des besoins data de l'organisation ;
- La présentation de l'architecture de collecte et de restitution des données ;
- La présentation des modèles de données conçus ;
- La description argumentée du Data Lake créé ;
- La présentation du plan de migration proposé ;
- La description des actions de sécurisation et de mise en conformité réalisées.

Phase 3 : Présentation orale de la production

Soutenance orale en groupe de projet sous forme d'une démonstration technique avec questionnement individuel par le jury.

Évaluation des compétences C1 à C9.

RNCP42157BC02 — Superviser le développement et l'automatisation des flux de données

LISTE DE COMPÉTENCES

Concevoir des pipelines ETL robustes, en modélisant les flux de données, en utilisant des outils d'orchestration et des technologies de traitement par lots ou en temps réel, afin d'assurer la cohérence et la disponibilité des données pour les analyses et les processus métiers.

Implémenter un pipeline de traitement de données, en intégrant des outils de collecte et de transformation par lot ou en streaming, en configurant des systèmes d'ingestion pour les données structurées et non structurées, afin d'assurer l'efficacité, la sécurité et la scalabilité des flux de données conformément à l'architecture définie.

Optimiser les performances des pipelines de données, en identifiant les goulots d'étranglement et en implémentant des solutions techniques telles que la parallélisation des tâches, l'optimisation des requêtes SQL et l'utilisation de technologies de traitement distribué, afin de garantir l'efficacité, la scalabilité et la réactivité des processus de traitement des données.

Développer des tests de surveillance des pipelines de données, en identifiant les points de contrôle stratégiques et en déployant des outils de monitoring automatisés, afin de garantir l'intégrité et la traçabilité des données tout au long du cycle de vie des pipelines.

Configurer des pipelines MLOps en intégrant des outils d'automatisation et en paramétrant les environnements de déploiement, afin d'assurer la mise en production fiable, reproductible et scalable des modèles d'IA.

Concevoir une architecture de migration inter-systèmes et multi-environnements, en utilisant un outil de gestion des données adapté aux besoins spécifiques du projet, afin de garantir l'intégration harmonieuse de données hétérogènes et l'interopérabilité optimale des différentes sources de données.

Elaborer des architectures de traitement des données en temps réel, en intégrant des technologies de streaming et en configurant des pipelines de données, afin de garantir la réactivité, la scalabilité et la robustesse des systèmes de gestion des données.

Développer des pipelines de traitement de données, en intégrant les services et fonctionnalités des architectures Data Lake, en appliquant les pratiques de gestion du cycle de vie des données, afin de maîtriser le parcours de la donnée et de faciliter l'exploitation des données par les différents usages analytiques et opérationnels.

Organiser les sources de données, en utilisant des outils adaptés de data visualisation, en sélectionnant les indicateurs clés (KPI) et en structurant les informations de manière concise, afin de faciliter la prise de décision des directions métiers.

Documenter les sources de données et les processus d'intégration, en élaborant des schémas détaillés et des descriptions précises des flux de données et en utilisant des outils de modélisation, afin de maintenir la traçabilité et la transparence des données.

MODALITÉS D'ÉVALUATION

Mise en situation professionnelle reconstituée

Construction de l'infrastructure d'une plateforme Big Data sécurisée sur la base d'un cahier des charges issu d'un cas réel ou fictif.

Phase 1 : Préparation de la construction de l'infrastructure d'une plateforme Big Data

Préparation tutorée par groupe de projet.

Phase 2 : Production écrite en groupe de projet

Le candidat devra produire un dossier rédigé comprenant :

- Une présentation complète des pipeline ETL et des traitements de données développés ;
- Une description argumentée des tests et contrôles de performance réalisés ;
- Une présentation détaillée des sources de données utilisées (sources, typologie, usages, ...) ;

- Une présentation de l'architecture conçue.

Phase 3 : Présentation de la production

Soutenance orale en groupe projet sous forme d'une démonstration technique avec questionnement individuel par le jury.

Évaluation des compétences C10 à C19.

RNCP42157BC03 — Piloter et optimiser les infrastructures de stockage et de traitement des données

LISTE DE COMPÉTENCES

Concevoir une architecture de stockage intégrant des systèmes de gestion de bases de données optimisés au regard des besoins du projet, en respectant les normes de sécurité définies, et en utilisant des techniques de nettoyage et d'amélioration des données, afin de faciliter l'exploitation, l'analyse et la valorisation des données.

Déployer une architecture de stockage distribuée, en configurant et en administrant des clusters de nœuds et en optimisant les ressources matérielles et logicielles, afin de garantir la continuité de service dans un environnement distribué.

Optimiser la configuration de la plateforme de stockage de données, en ajustant les paramètres de performance, en dimensionnant la capacité de stockage et en maximisant la bande passante, afin de garantir une disponibilité et une efficacité optimales des ressources pour les utilisateurs.

Mettre en place un système d'auto scaling des ressources en configurant des outils de surveillance et des algorithmes d'optimisation afin d'anticiper les montées en charge.

Assurer la maintenance et la mise à jour des infrastructures de stockage et de traitement des données, en utilisant des outils de monitoring et de diagnostic, en appliquant les correctifs et les mises à jour nécessaires, afin de soutenir durablement les besoins métiers et techniques de l'organisation, tout en optimisant l'efficacité énergétique et en réduisant l'empreinte environnementale des infrastructures.

Analyser les coûts et les performances de stockage des données, en utilisant des outils de monitoring, des indicateurs clés de performance et en tenant compte des licences, des supports et de l'évolutivité, afin de garantir la pérennité de la solution de stockage des données.

Déployer des infrastructures de cloud computing, en configurant les environnements de calcul et de stockage, en utilisant des outils de gestion et d'orchestration, afin de répondre aux impératifs de performance des environnements cloud de l'organisation.

Développer des scripts d'automatisation des tâches répétitives en utilisant des langages de programmation adaptés et des outils de gestion des flux de travail afin d'améliorer la productivité des processus de traitement des données et de réduire les tâches manuelles.

Effectuer des audits des systèmes de gestion des données en recourant à des outils d'analyse et des indicateurs de performance, afin de répondre aux standards de conformité réglementaires et normatifs et de formuler des recommandations aux parties prenantes.

Évaluer l'empreinte environnementale des processus de traitement des données à l'aide d'outils de mesure et d'indicateurs, afin d'optimiser l'usage des ressources et de réduire l'empreinte environnementale des solutions déployées.

MODALITÉS D'ÉVALUATION

Mise en situation professionnelle reconstituée

Administration et supervision d'une plateforme Big Data sécurisée sur la base d'un cahier des charges issu d'un cas réel ou fictif.

Phase 1 : Préparation de l'administration de la plateforme Big Data

Préparation tutorée par groupe de projet.

Phase 2 : Production écrite en groupe de projet.

Le candidat devra produire un dossier rédigé comprenant :

- Une présentation complète de l'architecture de stockage conçue incluant les optimisations réalisées, les tests réalisés et outils de surveillance mis en place ;
- Un rapport d'analyse des coûts et de la performance du système de stockage de données incluant une analyse de l'impact environnemental ;
- Une présentation des solutions cloud et d'automatisation mises en place ;
- Une présentation argumentée des recommandations formulées.

Phase 3 : Présentation de la production

Soutenance orale en groupe de projet sous forme d'une démonstration technique avec questionnement individuel par le jury.

RNCP42157BC04 — Manager des projets stratégiques en lien avec la gouvernance des données

LISTE DE COMPÉTENCES

Planifier le projet de développement de plateforme Big Data, en s'appuyant sur les besoins utilisateurs collectés et en répartissant les activités en fonction des ressources humaines, techniques et financières, afin d'identifier l'ensemble des étapes de réalisation et d'organiser le projet en tâches et livrables.

Gérer un projet, en utilisant les méthodes agiles et outils adaptés, dans le but d'optimiser les délais de livraison et d'assurer la bonne exécution du projet de solution Big Data.

Réaliser le reporting du projet auprès du commanditaire, en élaborant des tableaux de bord de suivi des performances, dans le but d'analyser l'utilisation des ressources à chaque étape, d'anticiper et de corriger les écarts en temps réel et de limiter les contraintes et les retards dans l'exécution du projet.

Piloter les prestataires extérieurs gérant les ressources informatiques du SI existant (listées dans la cartographie établie), en présentant le projet, les rôles et objectifs de l'ensemble des parties prenantes et en s'assurant de la pertinence des contrats existants, afin d'identifier les KPI, de calibrer le suivi et de sécuriser la mise en œuvre technique.

Conduire l'équipe projet, en favorisant l'inclusion des personnes en situation de handicap et en adaptant les conditions de travail face aux contraintes de temps et aux incertitudes, afin de renforcer la motivation, la résilience et l'équilibre entre vie professionnelle et personnelle.

Mettre en œuvre un processus de communication inclusif régulier au sein de l'équipe, en adaptant les modes de communication selon le contexte, les cultures et la langue des collaborateurs, afin de synchroniser les activités quotidiennes et de garantir l'intégration de tous les membres de l'équipe.

Proposer des solutions innovantes, en organisant des séances de brainstorming, en recueillant des retours réguliers et en encourageant l'autonomie ou la prise d'initiative, afin de favoriser les interactions au sein de l'équipe, de maintenir une dynamique de groupe et de renforcer l'esprit d'équipe au service du projet.

Établir un processus de partage d'information, en identifiant les besoins des parties prenantes, en sélectionnant l'outil adapté et en établissant des protocoles et procédures de partage, afin de faciliter l'inclusion, la collaboration entre les membres de l'équipe projet et de sécuriser l'information liée au projet.

MODALITÉS D'ÉVALUATION

Mise en situation professionnelle reconstituée

Gestion d'un projet de gouvernance des données stratégiques en lien avec la gouvernance des données

Phase 1 : Préparation et cadrage du projet

Dossier écrit en groupe de projet.

A partir d'un cas précisant les éléments de contexte, le candidat devra fournir les éléments suivants :

- La planification du projet ;
- Une présentation détaillée et argumentée de la méthodologie choisie et ses modalités de mise en œuvre ;
- Les outils de suivi de projet permettant de piloter les indicateurs définis ;
- Des tableaux de bord de suivi ;
- Une présentation de l'équipe projet ;
- Une présentation des outils de communication mis en place ;
- Une proposition de solutions innovantes pour des situations de rencontres et d'échanges ;
- Une présentation des outils collaboratifs choisis ;
- Des ordres du jour de réunion d'animation de l'équipe projet ;
- La présentation d'une stratégie d'organisation du partage d'informations en utilisant des outils numériques.

Phase 2 : Présentation de la production

Le dossier écrit sera présenté à l'oral en groupe de projet avec questionnement individuel par le jury.

Évaluation des compétences C30 à C37.

RNCP42157BC05 — Définir une stratégie de valorisation de la donnée

LISTE DE COMPÉTENCES

Analyser la stratégie de gouvernance des données de l'organisation et son fonctionnement interne, en mobilisant des outils de diagnostic, des indicateurs de performance et des techniques d'analyse exploratoire des données, afin de produire un diagnostic du système de gouvernance des données et de formuler des recommandations stratégiques adaptées.

Analyser les sources de données et les traitements associés, en mobilisant des outils de profilage, de nettoyage, de

data science et de visualisation, afin d'évaluer la complétude, la cohérence, et la conformité des données, et de produire un rapport d'audit détaillé.

Cartographier les flux, les sources, les zones critiques et les données sensibles, à l'aide d'outils de data visualisation et d'analyse des risques liés à la gestion et à l'exploitation des données, afin d'évaluer le patrimoine Data de l'organisation, d'identifier les risques et de proposer des mesures correctives ou préventives adaptées.

Élaborer un cadre de gouvernance durable des infrastructures de données, en définissant les règles de gestion des données de référence et en mettant en place des processus de validation et de contrôle de qualité afin d'assurer la pérennité des infrastructures de données, de garantir la cohérence et la qualité des informations utilisées par l'organisation et de soutenir une exploitation optimale et sécurisée du patrimoine Data.

Élaborer des indicateurs de performance Green IT, en analysant les impacts environnementaux et sociaux des ressources informatiques, en utilisant des outils de mesure et des méthodologies adaptés, afin de favoriser une utilisation responsable et durable des ressources numériques et de soutenir la performance globale de l'organisation dans sa transition écologique.

Élaborer une stratégie Big Data alignée sur les objectifs de l'organisation, en collaboration avec la DSI et en définissant les orientations en matière de gouvernance des données, d'intégration des analyses dans les processus décisionnels et de pilotage de la performance, afin de structurer une gouvernance data-driven de l'organisation.

Organiser la mise à disposition des données auprès des directions métiers, en assurant leur accessibilité, leur contextualisation et leur alignement avec les besoins opérationnels et stratégiques, afin de renforcer l'autonomie décisionnelle des utilisateurs et de faciliter l'appropriation d'une culture data-driven.

Animer des sessions de formation sur les outils d'analyse des données, de Data Literacy, et les pratiques Green IT associées, en élaborant des supports pédagogiques accessibles et en utilisant des méthodes interactives, afin de développer l'autonomie des utilisateurs et d'optimiser l'exploitation des outils.

MODALITÉS D'ÉVALUATION

Mise en situation professionnelle reconstituée

Élaboration d'une stratégie Big Data à partir d'une situation réelle ou reconstituée proposée par le certificateur

Phase 1 : Préparation de la stratégie Big Data

Préparation tutorée en groupe de projet.

Phase 2 : Rédaction du dossier

Production écrite en groupe de projet.

Le candidat devra produire un dossier rédigé comprenant :

- Un rapport d'audit des données de l'organisation ;
- Une présentation de la stratégie Big Data et de la stratégie de valorisation des données est proposée et les indicateurs et outils de pilotage associés ;
- Un support de formation permettant la prise en main des outils d'analyse de données.

Phase 3 : Présentation de la production

Soutenance orale sous forme d'une démonstration technique devant un jury de professionnel avec questionnaire individuel par le jury.

Évaluation des compétences C38 à C45.

Secteur d'activité et type d'emploi

SECTEURS D'ACTIVITÉS

L'Expert en Ingénierie des Données peut exercer ses activités dans des entreprises de toutes tailles : PME-PMI, start-up, grandes entreprises ou grands groupes. Dans la pratique, il peut être embauché directement par une entreprise, une administration publique ou une association. Il peut également travailler pour une société de services ou un cabinet de conseil qui le missionnera auprès de différents clients.

Plusieurs catégories d'organisations concentrent la majorité des recrutements d'Experts en Ingénierie des Données – Data Engineer, notamment les grandes entreprises privées opérant dans des secteurs où la donnée est stratégique. Les entreprises du CAC 40 et du SBF 120, ainsi que les domaines de la télécommunication, de la finance, de l'e-commerce, de la santé, de l'aéronautique ou de l'énergie, recherchent massivement ce type de profil.

L'Expert en Ingénierie des Données peut également exercer au sein de structures telles que les opérateurs d'importance vitale (OIV) ayant de fortes exigences en matière de données, les cabinets de conseil en data et analytics, les start-ups et scale-ups spécialisées en Big Data ou IA, le secteur public et parapublic (collectivités, établissements de recherche, services gouvernementaux), les sociétés de services numériques (ESN) et intégrateurs

cloud, les assureurs, mutuelles et organismes financiers.

TYPE D'EMPLOIS ACCESSIBLES

Les appellations courantes associées au métier d'Expert en ingénierie des données varient selon le secteur, la taille de l'entreprise, la maturité numérique des organisations et le niveau d'expertise du professionnel. Le champ d'action s'étend du développement de pipelines de données à la conception d'architectures complexes, jusqu'à la supervision de projets stratégiques de gouvernance ou de transformation numérique.

Les fonctions visées et types d'emplois accessibles directement en sortie de formation/certification (premier emploi / junior) sont : Data Engineer / Ingénieur des Données ; Data & Automation Engineer ; Ingénieur Big Data ; Intégrateur de Données (ETL) ; Administrateur de Bases de Données (DBA) ; Spécialiste Data Pipeline.

Les fonctions visées à moyen et long terme (3 à 10 ans d'expérience / confirmé ou senior) seront : Expert en Ingénierie des Données / Expert data ; Data Architect / Architecte Data ; Lead data engineer / Responsable Technique Data ; Expert Big Data ; Chief Data Officer ; Responsable Plateforme de Données ; Responsable Pipelines ou infrastructures Data ; Consultant Data Engineering ; Ingénieur MLOps / Responsable Data Industrialisation ; Responsable Gouvernance et Qualité des Données (technique) ; Directeur de l'ingénierie des Données.

RÈGLEMENTATIONS D'ACTIVITÉ

Le métier d'Expert en ingénierie des données – Data Engineer n'est pas une profession réglementée au sens strict du terme. Toutefois, certaines réglementations spécifiques doivent être maîtrisées par ce professionnel, dans le cadre de ses activités. On peut citer notamment :

- Le règlement général sur la protection des données (RGPD) ou règlement UE 2016/679, adopté par le Parlement européen le 27 avril 2016. Il établit les règles et principes visant à accroître et garantir la protection des personnes physiques à l'égard du traitement qui est réservé à leurs données personnelles, mais aussi à responsabiliser les acteurs de ce traitement (entreprises, institutions, services publics...). Son application est entrée en vigueur le 25 mai 2018 au sein de toute l'Union européenne ;
- La directive NIS2 (en français : sécurité des réseaux et des systèmes d'information) vise à renforcer le niveau de cybersécurité des tissus économiques et administratifs ;
- La loi pour « réduire l'empreinte environnementale du numérique en France » (dite REEN), du 15 novembre 2021, vise à développer les principes d'écoconception des services numériques et à lutter contre l'obsolescence du matériel et des logiciels informatiques, afin de faire converger transition écologique et transition numérique ;
- La loi du 23 décembre 2021 « visant à renforcer la régulation environnementale du numérique par l'Autorité de régulation des communications électroniques, des postes et de la distribution de la presse (ARCEP) » complète et mobilise les moyens de mise en œuvre de certaines dispositions de la loi REEN. Elle confère à l'Arcep un pouvoir de collecte de données relatives à l'impact environnemental du numérique. Elle prévoit également qu'un bilan de l'empreinte environnementale du secteur des communications électroniques, des terminaux et des centres de données soit désormais intégré au rapport annuel sur l'état de l'internet de l'Arcep ;
- La loi européenne sur l'Intelligence Artificielle, publiée le 13 juin 2024, aussi appelé IA Act. Le règlement sur l'IA est le tout premier cadre juridique complet sur l'IA dans le monde. Il établit des règles harmonisées en matière d'Intelligence Artificielle. L'objectif de ces règles est de favoriser une IA digne de confiance en Europe ;
- Le règlement du Parlement européen sur la gouvernance européenne des données, entré en vigueur le 24 septembre 2023, appelé aussi Data Governance Act établit un cadre européen structuré pour faciliter le partage sécurisé des données à travers l'Union européenne ;
- La norme ISO/IEC 27001 spécifie les exigences pour établir, mettre en œuvre, maintenir et améliorer un système de management de la sécurité de l'information ;
- La norme ISO/IEC 27701 fournit des lignes directrices pour la mise en œuvre d'un système de management de la protection de la vie privée ;
- La norme ISO/IEC 25012 établit un modèle général de qualité des données pour celles conservées dans un format structuré au sein d'un système informatique ;
- La norme ISO/IEC 14001 fournit aux organisations des lignes directrices pour mettre en œuvre des mesures proactives pour réduire leur empreinte environnementale, se conformer aux exigences légales en vigueur et atteindre leurs objectifs environnementaux.

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)	2 942 €
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) M2X90 — Ingénieurs et cadres d'étude, recherche et développement en informatique et télécom · tension 5/5 (Très élevée) M2X93 — Experts et consultants en systèmes d'information · tension 5/5 (Très élevée)
Projets de recrutement (BMO 2026)	32 216
Jugés difficiles à recruter	51 %
Emplois saisonniers	1 %
Familles professionnelles (FAP)	M1X80 — Techniciens d'étude et de développement en informatique M2X90 — Ingénieurs et cadres d'étude, recherche et développement en 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

Le dispositif s'adresse aux titulaires d'un baccalauréat ou certification de niveau 4 pour un parcours en 5 ans.

Dans le cadre d'admission parallèle :

- Admission en 3^e année : être titulaire d'un diplôme ou titre de niveau 5 dans le domaine de l'informatique pour un parcours en 3 ans ;
- Admission en 4^e année : être titulaire d'un diplôme ou titre de niveau 6 dans le domaine de l'informatique pour un parcours en 2 ans.

Dans le cas où un candidat ne disposerait pas des prérequis définis, il a la possibilité de déposer un dossier de candidature qui sera examiné par une commission réunissant la direction pédagogique d'EPSI et le responsable en charge du recrutement des candidats en vue d'une admission « exceptionnelle ». Le dossier de candidature fera l'objet d'une VAPP (Validation des Acquis Professionnels et Personnels).

Pour tous les candidats, la copie du dernier diplôme obtenu doit être fournie.

PRÉREQUIS À LA VALIDATION

Pour être convoqués aux épreuves d'évaluation, les candidats doivent avoir rempli les conditions suivantes (prérequis à la validation de la certification) :

- Suivi au minimum une proportion du programme de formation indiquée dans le règlement des études ;
- Avoir rendu les livrables attendus et spécifiés dans les modalités d'évaluation de chaque bloc de compétences dans les délais impartis et selon le formalisme attendu ;
- Avoir réalisé la période d'application professionnelle en entreprise selon la voie d'accès à la formation :
 - o En contrat d'apprentissage / contrat de professionnalisation / formation initiale (hors alternance) : 24 semaines ;
 - o Après un parcours de formation continue : stage conseillé mais non obligatoire pour les personnes ayant déjà une expérience significative dans le domaine visé ;
 - o Par expérience (VAE) : pas de stage exigé.

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

Le jury de délivrance de la certification est composé de trois membres :

- Un représentant de l'organisme certificateur AFINUM-EPSI : membre de l'équipe pédagogique (direction pédagogique ou son représentant) n'ayant pas participé à la préparation des candidats
- Deux professionnels du secteur informatique experts dans le métier visé, externes à l'organisme de formation, l'un d'eux assurant la présidence du jury.

COMPOSITION DU JURY — VAE

Le jury de délivrance de la certification est composé de trois membres :

- Un représentant de l'organisme certificateur AFINUM-EPSI : membre de l'équipe pédagogique (direction pédagogique ou son représentant) n'ayant pas participé à la préparation des candidats et assurant la présidence du jury
- Deux professionnels du secteur informatique experts dans le métier visé, externes à l'organisme de formation.

Base légale

Date de décision	30/04/2026
Durée de l'enregistrement	3 ans
Date d'échéance de l'enregistrement	30/04/2029
Date de dernière délivrance possible	30/04/2033

Statistiques

Année	Certifiés	Dont VAE	Insertion 6 mois	Métier 2 ans
2024	20	0	100 %	72 %
2023	24	0	100 %	82 %

Liens avec d'autres certifications

ANCIENNES VERSIONS

RNCP36921

CORRESPONDANCES (PAR BLOCS)

RNCP37750, RNCP38919, RNCP39775

Pour plus d'informations

Fiche officielle (France Compétences)

<https://www.francecompetences.fr/recherche/rncp/42157/>