

RÈGLEMENT SPÉCIAL DE CERTIFICATION

Le règlement spécial de certification définit les règles spécifiques applicables à la certification **RNCP39586** intitulée « **Ingénieur en science des données spécialisé en infrastructure data ou en apprentissage automatique** » pour l'ensemble des voies d'accès hors VAE.

1. Prérequis à la certification

Le candidat à la certification peut se présenter aux épreuves certificatives sous réserve de remplir les conditions décrites ci-dessous. La vérification de ces conditions est réalisée par le service pédagogique du groupe YNOV ou de l'un de ses centres partenaires habilités à évaluer.

Quelle que soit la voie d'accès, le candidat doit avoir respecté les prérequis d'entrée en formation :

- Être titulaire d'un diplôme de niveau 4 pour un parcours en 5 ans
- Être titulaire d'un diplôme ou titre de niveau 5 ou équivalent dans un domaine en lien avec la certification pour un parcours en 3 ans
- Être titulaire d'un diplôme ou titre de niveau 6 ou équivalent dans un domaine en lien avec la certification pour un parcours en 2 ans
- Avoir obtenu une dérogation d'entrée en formation par la commission spécifique.

2. Modalités d'évaluation

L'évaluation est réalisée conformément aux attendus de la certification. Le référentiel de la certification contenant le référentiel d'évaluation est disponible en téléchargement libre sur le site internet de France Compétences. [RNCP39586 - Ingénieur en science des données spécialisé en infrastructure data ou en apprentissage automatique \(francecompetences.fr\)](https://francecompetences.fr/RNCP39586-Ingénieur-en-science-des-données-spécialisé-en-infrastructure-data-ou-en-apprentissage-automatique)

Un bloc de compétences est considéré comme validé si 50% des compétences du bloc sont acquises et aucune compétence éliminatoire n'est évaluée comme non acquise.

La certification s'obtient par la validation de quatre blocs des compétences : validation des 3 blocs communs (blocs 1, 2 et 3) et validation d'un bloc de spécialisation (bloc 4 ou bloc 5).

Évaluation du Bloc 1 : Collecter, transformer et sécuriser des données

Type d'évaluation : Mise en situation professionnelle réelle ou fictive.

Attendus du candidat : À partir de l'analyse d'une organisation réelle ou fictive de son choix, le candidat propose une stratégie de collecte et de transformation des données.

Livrable attendu : Le candidat remet au jury un dossier écrit de 20 pages maximum (hors annexes et pages de garde) comprenant :

- Une stratégie de collecte de données
- Un exemple de collecte de données
- Une méthode d'automatisation de collecte
- Une stratégie de stockage des données
- Un modèle de données
- Une base de données
- Une solution de stockage Big Data
- Une présentation des outils et des technologies de traitement des données
- Une présentation des données transformées, des méthodes et outils utilisés
- Un processus ETL
- Les solutions utilisés pour l'automatisation et l'orchestration du traitement des données
- Une politique de sécurité des données

- Un schéma d'architecture de sécurité

Évaluateur : Jury d'évaluation composé de 2 membres professionnels externes dans le domaine visé

Modalités de validation : pour chaque compétence, le jury d'évaluation indiquera l'acquisition ou la non-acquisition de la compétence et motivera sa décision par un commentaire en cas de non-acquisition.

Compétences éliminatoires :

- C1.1.2 : Mettre en œuvre des techniques de collecte de données en exploitant des API externes disponibles, des bases de données, des techniques de web scrapping afin de recueillir et rassembler les données ciblées.
- C1.2.2 : Construire une base de données en sélectionnant la technologie (SQL, NoSQL), un système de gestion de base de données (SGBD) et/ou une solution de stockage BIG DATA, en assurant le paramétrage et l'implémentation afin de mettre en œuvre le modèle de données et garantir la disponibilité et l'intégrité des données.
- C1.3.2 : Transformer les données à l'aide de langage de programmation ou d'outils dédiés (ex : outil no code, Excel) afin d'obtenir des données nettoyées exploitables.
- C1.3.3 : Développer un processus ETL en utilisant la technologie ETL préalablement sélectionnée afin d'automatiser l'extraction, la transformation et le chargement de données.

Évaluation du Bloc 2 : Analyser, organiser et valoriser des données

Type d'évaluation : Mise en situation professionnelle réelle ou fictive.

Attendus du candidat : À partir de l'analyse d'une organisation réelle ou fictive de son choix, le candidat propose une stratégie d'analyse, de mise en forme et de valorisation des données.

Livrable attendu : Le candidat remet au jury un dossier écrit de 30 pages maximum (hors annexes et pages de garde) comprenant :

- Une analyse du besoin
- Une présentation d'un plan d'analyse
- Une présentation des requêtes et des résultats
- Une méthodologie de tests statistiques
- La visualisation des résultats de l'analyse
- Une présentation de recommandations
- Un support de formation
- Une documentation technique

Évaluateur : Jury d'évaluation composé de 2 membres professionnels externes dans le domaine visé

Modalités de validation : pour chaque compétence, le jury d'évaluation indiquera l'acquisition ou la non-acquisition de la compétence et motivera sa décision par un commentaire en cas de non-acquisition.

Compétences éliminatoires :

- C2.1.2 : Définir les axes d'analyse et les métriques en identifiant les données à exploiter, celles disponibles et pertinentes pour traduire la problématique d'entreprise énoncée en problème numérique.

- C2.1.3 : Réaliser des requêtes et des calculs en utilisant des outils de dashboarding, des tableurs, des requêtes SQL ou scripts Python afin de produire une analyse des données préalablement collectées.
- C2.2.1 : Représenter les données en choisissant les modèles de représentation les plus adaptés (ex : histogramme, Heat map, nuage de points) et en utilisant des outils de représentation adaptés (ex : Office, power BI) afin de permettre la compréhension et l'exploitation des données par le public visé.

Évaluation du Bloc 3 : Élaborer et piloter un projet data

Type d'évaluation : Mise en situation professionnelle réelle ou fictive.

Attendus du candidat : À partir de l'analyse d'une organisation réelle ou fictive de son choix, le candidat propose la gestion d'un projet Data.

Livrable attendu : Le candidat remet au jury un dossier écrit de 20 pages maximum (hors annexes et pages de garde) comprenant :

- Le cadrage du projet
- Le dimensionnement du projet
- La documentation projet
- Le planning projet
- Un outil de suivi de projet
- Un tableau de bord
- Un plan de développement des compétences
- Les outils de communication et managériaux utilisés
- La présentation d'un cas d'arbitrage rencontré au cours du projet
- Une méthodologie de veille
- Un plan d'actions relatif aux enjeux RSE, de sécurité, d'éthique et de confidentialité

Évaluateur : Jury d'évaluation composé de 2 membres professionnels externes dans le domaine visé

Modalités de validation : pour chaque compétence, le jury d'évaluation indiquera l'acquisition ou la non-acquisition de la compétence et motivera sa décision par un commentaire en cas de non-acquisition.

Compétences éliminatoires :

- C3.1.2 : Dimensionner le projet en évaluant la charge de travail et les ressources nécessaires (humaines, matérielles) au regard des exigences attendues et des contraintes préalablement définies afin d'estimer le temps et le budget nécessaires à la faisabilité du projet.
- C3.2.1 : Planifier l'exécution du projet en organisant la répartition et l'ordonnancement des activités, le planning prévisionnel de réalisation et les ressources nécessaires à son exécution, en prenant en considération les personnes en situation de handicap afin de maîtriser les différentes phases du projet.
- C3.2.2 : Suivre l'avancement du projet en mettant en place un outil de suivi (logiciel de suivi, tableau de bord), en définissant les indicateurs (qualitatifs et/ou quantitatifs) pour chaque jalon défini dans le planning, en réalisant des reportings et des comptes rendus de réunion afin d'optimiser la coordination de l'équipe et anticiper les aléas éventuels

Évaluation du Bloc 4 : (Spécialisation infrastructure data) - Concevoir et opérer une infrastructure data

Type d'évaluation : Mise en situation professionnelle réelle ou fictive.

Attendus du candidat : À partir de l'analyse d'un projet réel ou fictif, le candidat présente la stratégie de conception et de maintien d'une architecture DATA.

Livrable attendu : À l'aide d'un support de présentation de son choix, le candidat présente lors d'une soutenance orale de 45 min (30 min de présentation et 15 de temps d'échange avec le jury) :

- Un rapport d'analyse
- Une présentation des composants de l'architecture DATA
- Un schéma de données
- Des pipelines de traitement de la donnée
- Un pipeline CI/CD
- Une présentation d'un système de supervision
- Une feuille de route d'exploitation
- Une documentation technique
- Un cahier de recettes et de tests
- Une méthodologie d'investigation et de traitement d'un incident.

Évaluateur : Jury d'évaluation composé de 2 membres professionnels externes dans le domaine visé.

Modalités de validation : pour chaque compétence, le jury d'évaluation indiquera l'acquisition ou la non-acquisition de la compétence et motivera sa décision par un commentaire en cas de non-acquisition.

Compétences éliminatoires :

- C4.2.1 : Concevoir une architecture d'entrepôt de données en s'appuyant sur le cahier des charges, en sélectionnant les composants appropriés afin d'optimiser le stockage des données en termes de rapidité, de sécurité et d'accessibilité.
- C4.2.2 : Mettre en place et orchestrer des pipelines DATA temps réel ou asynchrones à l'aide d'outils BIG DATA (ex : DATA brick, plateforme DBT , Snowflake), Cloud et/ou On premise afin d'automatiser la transformation et la transmission des données.
- C4.2.3 : Automatiser l'intégration et le déploiement des composants en utilisant des outils d'intégration et de développement continu, DevOps afin d'industrialiser la mise en production de l'architecture DATA.

Évaluation du Bloc 5 : (Spécialisation apprentissage automatique) - Concevoir et déployer des modèles d'apprentissage automatique

Type d'évaluation : Mise en situation professionnelle réelle ou fictive.

Attendus du candidat : À partir de l'analyse d'un projet réel ou fictif, le candidat présente la stratégie de conception et de déploiement de modèles de machine learning.

Livrable attendu : À l'aide d'un support de présentation de son choix, le candidat présente lors d'une soutenance orale de 45 min (30 min de présentation et 15 de temps d'échange avec le jury) :

- Une analyse du besoin
- Une présentation de la stratégie de résolution du problème
- Une présentation des technologies et des outils sélectionnés
- Un jeu de données exploitables pour le Machine Learning
- Des méthodes de sélection de variables
- Un entraînement d'un modèle d'apprentissage automatique
- Une méthode d'optimisation des modèles d'apprentissage automatique.
- Une méthode de sauvegarde
- Un processus CI/CD

- Un système de monitoring de la performance
- Un système de collecte de données

Évaluateur : Jury d'évaluation composé de 2 membres professionnels externes dans le domaine visé

Modalités de validation : pour chaque compétence, le jury d'évaluation indiquera l'acquisition ou la non-acquisition de la compétence et motivera sa décision par un commentaire en cas de non-acquisition.

Compétences éliminatoires :

- C5.2.1 : Construire des variables en utilisant des langages de programmation (ex : Python, Scala, R, Julia...), en exploitant des bibliothèques d'analyse de données afin de fournir les meilleures variables au modèle Machine Learning.
- C5.2.2 : Sélectionner les variables en identifiant les différentes méthodes de sélection de variables possibles, en utilisant des méthodologies Machine Learning afin d'optimiser la performance du modèle.
- C5.2.3 : Construire et entraîner un modèle d'apprentissage automatique à l'aide de librairies (ex : Scikit-learn XGBoost, TensorFlow, PyTorch) afin d'obtenir des modèles capables de prédictions sur de nouvelles données inconnues.
- C5.2.4 : Optimiser la performance des modèles d'apprentissage automatique en modifiant les hyperparamètres et en analysant les prédictions afin de répondre au mieux à la problématique du commanditaire.