

## RNCP Data Engineer – Niveau 7 (EU)

Référentiel d'activités, de compétences et d'évaluation

Article L6113-1 En savoir plus sur cet article... Créé par LOI n°2018-771 du 5 septembre 2018 - art. 31 (V)

*« Les certifications professionnelles enregistrées au répertoire national des certifications professionnelles permettent une validation des compétences et des connaissances acquises nécessaires à l'exercice d'activités professionnelles. Elles sont définies notamment par un référentiel d'activités qui décrit les situations de travail et les activités exercées, les métiers ou emplois visés, un référentiel de compétences qui identifie les compétences et les connaissances, y compris transversales, qui en découlent et un référentiel d'évaluation qui définit les critères et les modalités d'évaluation des acquis. »*

### Candidat en situation de handicap :

Dans le cadre du respect du règlement de la certification, tout candidat peut saisir le référent handicap afin d'étudier les possibilités d'aménagement des modalités d'évaluation.

Les supports et le matériel nécessaires à la réalisation des évaluations pourront être adaptés. Sur conseil du référent handicap et dans le respect des spécifications du référentiel de la certification, le format de la modalité pourra être adaptée si nécessaire. L'ingénieur de certification s'engage dans la mesure du possible à élaborer des modalités d'évaluation inclusives permettant une adaptation du format.

# RNCP Data Engineer – Niveau 7 (EU)

Référentiel d'activités, de compétences et d'évaluation

REFERENTIEL D'ACTIVITES <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	REFERENTIEL DE COMPETENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	REFERENTIEL D'EVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p><b>BLOC 1 : ANALYSER LES BESOINS EN MATIERE D'INFRASTRUCTURES DE DONNÉES ADAPTÉS AU PROJET D'INTELLIGENCE ARTIFICIELLE</b></p> <p><b>A1. Mise en œuvre d'un processus de veilles technologique et réglementaire en matière d'intelligence artificielle.</b></p>	<p><b>C1.</b> Mettre en place un processus de veilles technologique et réglementaire en matière d'intelligence artificielle en s'appuyant sur des sources dédiées afin d'anticiper l'évolution des normes et règlements applicables à la certificabilité des systèmes intégrant des intelligences artificielles ou des traitements de données massives.</p> <p><b>C2.</b> Exercer une veille technologique et réglementaire à l'aide de sources dédiées, en collectant, classifiant et en analysant l'information afin de mettre en place de nouvelles technologies issues de la recherche en intelligence artificielle et en Science des données les mieux adaptées au projet de développement des infrastructures de données.</p>	<p><b>EPREUVE N°1 (C1 à C7) : Etude de cas</b> « Analyse des besoins en matière d'infrastructures de données adaptés au projet d'intelligence artificielle »</p> <p>L'objectif est d'évaluer la capacité du candidat à <b>mettre en œuvre et exploiter une veille technologique et réglementaire en matière d'intelligence artificielle et à proposer une cartographie des fonctionnalités attendues.</b></p> <ul style="list-style-type: none"> <li>• <b>Production :</b> A partir d'une étude tirée d'un cas réel, le candidat devra présenter une analyse des besoins en matière d'infrastructures de données adaptés au projet d'intelligence artificielle d'une entreprise donnée.</li> </ul>	<p><b>C1 à C2 : Evaluation des points suivants :</b></p> <ul style="list-style-type: none"> <li>- Les réglementations RGPD, de sécurisation informatique, cybersécurité sont respectées.</li> <li>- Les situations de handicap sont prises en compte dans les évolutions des normes et des règlements.</li> <li>- Le respect des normes et règlements applicables à la certificabilité des systèmes est anticipé.</li> </ul>

## RNCP Data Engineer – Niveau 7 (EU)

Référentiel d'activités, de compétences et d'évaluation

<p><b>A2. Analyse des besoins d'architectures issus des fonctionnalités attendues par les parties prenantes et de la capitalisation de l'expérience acquise.</b></p> <p><b>A3. Définition du projet de développement d'infrastructures de données massives adapté au contexte.</b></p>	<p><b>C3.</b> Analyser les cas d'usage en data et en intelligence artificielle déjà implémentés, en s'appuyant sur la réalisation d'une cartographie des exigences normatives et réglementaires de l'entreprise, pour les évaluer, les enrichir et en créer de nouveaux en capitalisant sur l'expérience acquise.</p> <p><b>C4.</b> Définir le projet de développement d'infrastructures de données massives à réaliser en interaction avec les parties prenantes, en procédant par itérations successives, en vue d'établir une liste exhaustive de ses attentes et prenant en compte les situations de handicap.</p> <p><b>C5.</b> Identifier les outils de Data Science applicables selon le contexte et les objectifs attendus afin de présenter des préconisations et d'intégrer des outils inclusifs IA/Data science adaptés à la problématique métier ainsi qu'à la politique RSE de la structure.</p>	<p>A ce titre, le candidat devra :</p> <ul style="list-style-type: none"> <li>- Présenter un processus de veille technologique et réglementaire ;</li> <li>- Démontrer la prise en compte des situations de handicap ;</li> <li>- Cartographier les fonctionnalités attendues à partir de l'expérience acquise et des exigences normatives et réglementaires.</li> </ul> <p>• <b>Soutenance :</b> La définition du projet de développement d'infrastructures de données massives est présentée lors d'une soutenance orale devant un jury.</p>	<p><b>C3 à C5 : Evaluation des points suivants :</b></p> <ul style="list-style-type: none"> <li>- Les fonctionnalités attendues (flux de données, stockage, puissance de calcul, accessibilité) sont cartographiées.</li> <li>- La capitalisation de l'expérience acquise est restituée.</li> <li>- Les exigences normatives et réglementaires sont cartographiées.</li> <li>- Une matrice SWOT (forces, faiblesses, opportunités et menaces) est présentée.</li> <li>- Les situations de handicap sont prises en compte dans la définition des fonctionnalités.</li> </ul> <p><b>C5 : Evaluation des points suivants :</b></p> <ul style="list-style-type: none"> <li>- Les outils en intelligence artificielle et en Science des données applicables selon le contexte, les objectifs sont identifiés et la politique RSE de l'entreprise sont présentés. La sélection est argumentée.</li> <li>- Le contexte et des objectifs attendus et les situations de handicap sont pris en compte dans le choix des outils de data science.</li> </ul>
--	---	--	--

RNCP Data Engineer – Niveau 7 (EU)  
Référentiel d'activités, de compétences et d'évaluation

<p><b>A4. Evaluation des compétences nécessaires adaptées au projet et aux nouveaux enjeux du secteur de l'intelligence artificielle.</b></p>	<p><b>C6.</b> Evaluer toutes les compétences nécessaires à la réalisation du projet d'intelligence artificielle, s'appuyant sur la cartographie des profils nécessaires au projet et prenant en compte les situations de handicap, en vue de constituer une équipe inclusive de développement spécialisée en intelligence artificielle.</p> <p><b>C7.</b> Intégrer les profils recrutés en favorisant l'inclusion des personnes en situation de handicap, en collaborant avec le référent handicap de l'entreprise et en identifiant les besoins en formation, afin de permettre leur pleine intégration au sein de l'équipe projet.</p>	<p><b>EPREUVE N°2 (C6 à C7) : Mise en situation professionnelle</b> « <i>Evaluation des compétences nécessaires adaptées au projet de développement d'intelligence artificielle</i> »</p> <p>L'objectif est d'évaluer la capacité du candidat à <b>identifier les profils nécessaires au projet d'intelligence artificielle, à prendre en compte les situations de handicap dans le processus de recrutement et</b></p>	<p>La sélection des outils en intelligence artificielle et en Science des données est argumentée. L'outil de data science est intégré selon la problématique métier donnée.</p> <p><b>C1 à C7 : Evaluation des points suivants :</b></p> <ul style="list-style-type: none"> <li>- Les critères d'évaluation de l'étude de cas et de la soutenance sont présentés dans le tableau de synthèse des modalités d'évaluation.</li> </ul> <p><b>C6 – C7 : Evaluation des points suivants :</b></p> <ul style="list-style-type: none"> <li>- Les profils nécessaires au projet sont identifiés et cartographiés.</li> <li>- Les situations de handicap sont prises en compte et l'intégration adaptée.</li> <li>- Une grille d'entretien de recrutement est élaborée et prend en compte les situations de handicap.</li> <li>- Les nouvelles recrues sont intégrées en collaboration avec le référent handicap de l'entreprise.</li> <li>- Les besoins en formation des membres de l'équipe projet</li> </ul>
---	--	---	--

## RNCP Data Engineer – Niveau 7 (EU)

Référentiel d'activités, de compétences et d'évaluation

		<p><b>construire une grille d'analyse des besoins en formation.</b></p> <ul style="list-style-type: none"><li>• <b><u>Production</u></b> : A partir du cas réel ou fictif, le candidat devra évaluer les compétences nécessaires à la réalisation du projet d'intelligence artificielle. A ce titre, il présentera :<ul style="list-style-type: none"><li>- La cartographie des profils nécessaires au projet d'intelligence artificielle.</li><li>- La grille de recrutement des profils prenant en compte les situations de handicap.</li><li>- Le processus d'intégration des nouvelles recrues.</li><li>- Une proposition de grille d'identification des besoins en formation de membres de l'équipe projet prenant en compte toutes les situations de handicap.</li></ul></li></ul>	<p>sont identifiés et consolidés dans une grille spécifique. Les formations proposées prennent en compte les différents types de handicap : moteur, sensoriel (auditif et visuel), mental, cognitif et psychique.</p>
--	--	--	---

**RNCP Data Engineer – Niveau 7 (EU)**  
Référentiel d'activités, de compétences et d'évaluation

<b>REFERENTIEL D'ACTIVITES</b> <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	<b>REFERENTIEL DE COMPETENCES</b> <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	<b>REFERENTIEL D'EVALUATION</b> <i>définit les critères et les modalités d'évaluation des acquis</i>	
		<b>MODALITÉS D'ÉVALUATION</b>	<b>CRITÈRES D'ÉVALUATION</b>
<p><b>BLOC 2 : PILOTER LE PROJET DE DÉVELOPPEMENT DE L'INFRASTRUCTURE DE DONNÉES MASSIVES</b></p> <p><b>A1. Conception de l'architecture de données massives à partir des propositions validées par les parties prenantes.</b></p>	<p><b>C8.</b> Superviser la rédaction du cahier des charges fonctionnels et techniques, en justifiant les choix opérés, et spécifier les livrables afin de déterminer les budgets et les ressources alloués pour la réalisation du projet.</p> <p><b>C9.</b> Vérifier l'adéquation du cahier des charges avec les besoins de l'entreprise, en s'appuyant sur des échanges avec celle-ci et en prenant en compte les spécificités techniques et fonctionnelles (sécurisation des données, handicap), en vue de prévenir tout risque d'écarts en cours de réalisation.</p>	<p><b>EPREUVE N°3 (C8 à C18) : Mise en situation professionnelle.</b> <i>« Pilotage du projet de développement de l'infrastructure de données massives »</i></p> <p>L'objectif est d'évaluer la capacité du candidat à <b>superviser la rédaction du cahier des charges et à en vérifier la cohérence avec les besoins exprimés en amont.</b></p> <p>• <b>Production :</b> A partir d'un cas réel ou fictif, le candidat devra piloter un projet de développement de l'infrastructure de données massives. A ce titre, il devra :</p> <ul style="list-style-type: none"> <li>- Justifier les choix opérés dans les cahiers des charges fonctionnels et techniques ;</li> <li>- Spécifier les livrables et démontrer la prise en compte des normes en</li> </ul>	<p><b>C8 à C9 : Evaluation des points suivants :</b></p> <ul style="list-style-type: none"> <li>- La rédaction des cahiers des charges fonctionnels et techniques est cohérente.</li> <li>- Les spécificités techniques et fonctionnelles sont prises en compte (sécurisation des données, handicap).</li> <li>- La prévention des risques d'écarts en cours de réalisation du projet est présentée de façon argumentée.</li> </ul>

## RNCP Data Engineer – Niveau 7 (EU)

Référentiel d'activités, de compétences et d'évaluation

<p><b>A2. Développement d'une base de données massive.</b></p>	<p><b>C10.</b> Définir la structure générale d'une base de données en ayant recours à la méthodologie de conception et en respectant les cahiers des charges afin de mettre en place une solution d'intelligence artificielle.</p> <p><b>C11.</b> Définir un format des données stockées adapté à la structure de la base de données, en utilisant les méthodologies de conception et d'exploitation des données adaptés, en vue de faciliter la saisie et la recherche d'informations.</p> <p><b>C12.</b> Vérifier la cohérence de la base de données définie et des formats en s'appuyant sur la méthodologie de conception des bases de données dans le but de les rendre exploitables par une solution d'intelligence artificielle.</p>	<p>termes de sécurisation des données et de handicap.</p> <p>Présenter les risques d'écarts en cours de réalisation du projet.</p> <p>Elaborer la structure générale des données collectées à des cahiers des charges validés par les parties prenantes.</p> <p>Définir un format des données stockées adapté à la structure de la base de données.</p> <p>Rédiger les langages de programmation en anglais.</p> <p>Démontrer la cohérence avec la structure de la base de données et ses formats.</p> <p>Démontrer le respect des cahiers des charges dans la définition de la structure générale de la base de données.</p> <p>Proposer des plans d'évolution du système d'information.</p> <p>Définir des indicateurs de performance.</p> <p>Planifier des opérations de tests et des processus de maintenance</p>	<p><b>C10 à C12 : Evaluation des points suivants :</b></p> <ul style="list-style-type: none"> <li>- La structure générale de la base des données collectées est définie.</li> <li>- La méthodologie de conception et exploitation des bases de données est maîtrisée (NoSQL, SQL, mongodb, MySQL, pipeline, merise, data warehouse).</li> <li>- Les cahiers des charges sont respectés.</li> <li>- La cohérence avec la structure de la base de données et ses formats est vérifiée et argumentée.</li> <li>- La maîtrise de l'anglais est démontrée dans les phases de rédaction des langages de programmation.</li> </ul> <p><b>C13 à C15 : Evaluation des points suivants :</b></p> <ul style="list-style-type: none"> <li>- Des opérations de tests et des processus de maintenance</li> </ul>
<p><b>A3. Pilotage de l'implémentation de l'architecture de données massives avec les parties prenantes.</b></p>	<p><b>C13.</b> Définir et mettre en place des indicateurs de mesure de la performance à partir opérations des tests et des processus de maintenance curative/préventive dans le but de monitorer le</p>		

## RNCP Data Engineer – Niveau 7 (EU)

Référentiel d'activités, de compétences et d'évaluation

<p><b>A4. Animation et management de l'équipe projet d'intelligence artificielle.</b></p>	<p>retour sur investissement des projets applicatifs.</p> <p><b>C14.</b> Gérer les données historiques en utilisant des indicateurs de performance afin de garantir la disponibilité, la qualité et l'amélioration continue de la fiabilité des données.</p> <p><b>C15.</b> Etablir un plan d'investissement pluriannuel et mobiliser les financements en se basant sur les plans d'évolution du système d'information afin de réaliser un plan global d'intelligence artificielle et de valorisation des données.</p> <p><b>C16.</b> Mobiliser les techniques et outils de conduite de projets inclusifs en mobilisant la culture DevOps et mettant en place des indicateurs de suivi afin de produire les livrables du projet dans le temps imparti, au niveau de qualité attendu.</p> <p><b>C17.</b> Maîtriser les méthodes de management de projet en utilisant les outils et les techniques de résolution de problèmes s'inscrivant dans la culture DevOps afin de surmonter les difficultés techniques et humaines rencontrées lors d'un projet.</p> <p><b>C18.</b> Motiver et fédérer les équipes projet d'intelligence artificielle en les sensibilisant au processus d'amélioration continue et en construisant les objectifs individuels ou collectifs à travers le dialogue et le consensus prenant en compte les situations de handicap et le développement durable.</p>	<p>curative/préventive. Réaliser un plan d'investissement global IA et valorisation des données.</p> <p>• <b>Soutenance :</b> L'animation et le management de l'équipe projet d'intelligence artificielle sont présentés lors d'une soutenance orale devant un jury.</p>	<p>curative/préventive sont planifiées. Des indicateurs de performance sont définis (KPI). Les plans d'évolution du système d'information sont élaborés. Un plan d'investissement global IA et valorisation des données est réalisé.</p> <p><b>C16 à C18 : Evaluation des points suivants :</b></p> <p>Des techniques de projet (animation d'une démarche agile et innovante) sont présentées et illustrées. Les outils de la culture DevOps sont maîtrisés et mis en œuvre : Git, Jenkins, Gitlab-CI, Docker, Kubernetes, Ansible, Terraform, Helm...</p> <p>Les rôles nécessaires aux objectifs et contraintes de projet sont définis. La collaboration avec les membres d'une équipe, partage des informations et des avis sont mis en œuvre et illustrés avec des exemples concrets. Des actions de</p>
---	--	--	---

RNCP Data Engineer – Niveau 7 (EU)  
Référentiel d'activités, de compétences et d'évaluation

			<p>sensibilisation au processus d'amélioration continue et à la démarche d'assurance qualité (théories de Deming, Burns, Dowton) sont mises en œuvre.</p> <p>Les situations de handicap sont prises en compte au sein de l'équipe projet ainsi que les enjeux autour de la protection de l'Environnement.</p>
--	--	--	---

# RNCP Data Engineer – Niveau 7 (EU)

Référentiel d'activités, de compétences et d'évaluation

REFERENTIEL D'ACTIVITES <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	REFERENTIEL DE COMPETENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	REFERENTIEL D'EVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p><b>BLOC 3 : CONCEVOIR LES ARCHITECTURES BIG DATA VALORISANT LES DONNÉES</b></p> <p><b>A1. Conception et communication de la stratégie de l'architecture d'entrepôt de données massives (Data Lake, Data Warehouse).</b></p>	<p><b>C19.</b> Concevoir une architecture d'entrepôt de données massives en mobilisant une expertise des solutions de manipulation des données (ETL/ELT) pour optimiser le stockage et la protection de données structurées ou non structurées provenant de sources multivariées.</p> <p><b>C20.</b> Communiquer la stratégie de mise en œuvre</p>	<p><b>EPREUVE N°4 (C19 à C29)</b> <b>Mise en situation professionnelle</b> « Conception d'une architecture d'entrepôt de données massives »</p> <p>L'objectif est d'évaluer la capacité du candidat à <b>concevoir une architecture d'entrepôt de données massives en respectant la réglementation en matière de protection de données.</b></p> <p>A partir d'un cas réel ou fictif, le candidat devra concevoir une architecture d'entrepôt de données en sélectionnant et en mettant en place un outil ETL/ELT.</p> <ul style="list-style-type: none"> <li>• <b>Production :</b> A ce titre, il devra : <ul style="list-style-type: none"> <li>- Concevoir une architecture d'entrepôt de données massives ;</li> <li>- Rédiger un support de</li> </ul> </li> </ul>	<p><b>C19 – C20 : Evaluation des points suivants :</b></p> <ul style="list-style-type: none"> <li>- Les solutions de stockage et de manipulation des données ETL/ELT (Airflow, Jenkins, Talent open studio, Pig Latin, Sqoop) sont maîtrisées et correctement mobilisées.</li> <li>- Des data pipelines sont créées et sont adaptées à l'architecture d'entrepôt de données.</li> <li>- La réglementation en matière de protection de données structurées ou non structurées est maîtrisée et explicitée.</li> <li>- La transformation des données de base dans des formats spécifiques permet l'intégration de données provenant de sources multivariées.</li> <li>- La communication de la</li> </ul>

## RNCP Data Engineer – Niveau 7 (EU)

Référentiel d'activités, de compétences et d'évaluation

<p><b>A2. Collecte des ensembles volumineux de données structurées et non structurées provenant de différentes sources.</b></p>	<p>de l'architecture d'entrepôt de données massives et ses résultats aux parties prenantes en respectant le cahier des charges ainsi que la réglementation en matière de sécurité informatique afin de dégager un avantage stratégique et commercial pour l'entreprise.</p> <p><b>C21.</b> Collecter des ensembles volumineux de données structurées et non structurées issues de sources internes et externes de l'entreprise à l'aide de solutions adaptées de manipulation de données afin d'en dégager des résultats visualisables sur des interfaces de visualisation et/ou modélisation.</p>	<p>communication sur la stratégie de mise en œuvre de l'architecture big data.</p> <p>Présenter la collecte des ensembles volumineux de données en mobilisant des technologies de manipulation de données ;</p>	<p>stratégie de mise en œuvre de l'architecture Big data est claire et argumentée.</p> <p><b>C21 : Evaluation des points suivants :</b></p> <ul style="list-style-type: none"> <li>- Les solutions de bases de données SQL et NoSQL sont maîtrisées.</li> <li>- Le choix des technologies de manipulation de données (Hadoop, Spark, Kafka, HBase, Cassandra, Hive, Pig, Oozie) est argumenté.</li> <li>- La visualisation et modélisation de données sont présentées via des outils (Tableau, Microsoft PowerBI, Google Data Studio, QlikView).</li> </ul>
<p><b>A3. Développement des composants de l'Intelligence artificielle.</b></p>	<p><b>C22.</b> Développer les composants de l'intelligence artificielle en langage de programmation approprié en respectant les normes du domaine et les principes déontologiques établis afin de respecter la démarche d'assurance qualité du code produit.</p> <p><b>C23.</b> Développer des algorithmes tels que les forêts aléatoires, les arbres de décision, l'algorithme K-Nearest Neighbors, la régression linéaire, l'algorithme de Naïve</p>	<p>Présenter les composants de l'intelligence artificielle et les justifier ;</p>	<p><b>C22 à C24 : Evaluation des points suivants :</b></p> <ul style="list-style-type: none"> <li>- Les langages de programmation (Python, Scala Java, Shell, VBA, C++, Ruby, Perl) sont maîtrisés.</li> <li>- Le code informatique est rédigé en anglais.</li> <li>- Les modèles statistiques et algorithmes associés</li> </ul>

RNCP Data Engineer – Niveau 7 (EU)  
Référentiel d'activités, de compétences et d'évaluation

	<p>Bayes, la machine à vecteurs de support (SVM), régression logistique et boosting de gradient, reposant sur des modèles statistiques en vue de mettre en œuvre les processus d'apprentissage automatique (« Machine learning »).</p> <p><b>C24.</b> Mettre en œuvre les technologies reposant sur les réseaux de neurones et prenant en compte les principes éthiques selon les normes en vigueur en vue d'intégrer les processus d'apprentissage profond (« Deep learning »).</p>		<p>sont présentés de manière argumentée (forêts aléatoires, les arbres de décision, l'algorithme K-Nearest Neighbors, la régression linéaire, l'algorithme de Naïve Bayes, la machine à vecteurs de support (SVM), régression logistique et boosting de gradient).</p> <p>- Les données stockées sont utilisées pour favoriser l'apprentissage par les machines (« Machine Learning »).</p> <p>- Les principes éthiques selon les normes en vigueur sont pris en compte.</p>
<p><b>A4. Evaluation des solutions de formatage et de stockage des données massives.</b></p>	<p><b>C25.</b> Evaluer différentes solutions de formatage et de stockage en se basant sur les solutions de chiffrements et de pare-feux ainsi qu'en prenant en compte les contraintes éthiques et légales en vue de favoriser leur traitement, leur centralisation et leur sécurisation.</p>	<p>- Présenter une évaluation des solutions de formatage et de stockage des données massives ;</p>	<p><b>C25 : Evaluation des points suivants :</b></p> <p>- Les solutions de chiffrements et de pare-feu sont mises en œuvre et évaluées.</p> <p>- Les contraintes éthiques et légales (RGPD) sont présentées.</p>
<p><b>A5. Optimisation du traitement et de la visualisation des données.</b></p>	<p><b>C26.</b> Créer ou identifier les différents algorithmes d'optimisation du traitement et de la visualisation des données pour évaluer leur capacité à</p>	<p>- Optimiser le traitement et la visualisation des données grâce à</p>	<p><b>C26 à C29 : Evaluation des points suivants :</b></p> <p>- Les différents algorithmes d'optimisation sont</p>

**RNCP Data Engineer – Niveau 7 (EU)**  
Référentiel d'activités, de compétences et d'évaluation

	<p>résoudre des problèmes spécifiques.</p> <p><b>C27.</b> Mettre en œuvre un algorithme répondant au besoin d'optimisation du traitement des données répondant aux besoins des parties prenantes afin de résoudre un problème de décision séquentiel simple.</p> <p><b>C28.</b> Présenter les résultats d'analyses de données massives sous la forme d'infographies en mettant en la configuration des requêtes de manière à faciliter la prise de décision managériale ou opérationnelle.</p> <p><b>C29.</b> Etablir le schéma général des accès à l'application développée prenant en compte le référentiel d'accessibilité (RGAA), en vue d'assurer leur cohérence avec l'organisation de l'entreprise utilisatrice et les différentes parties prenantes.</p>	<p>l'identification ou la création d'algorithmes d'optimisation ;</p>	<p>identifiés ou créés.</p> <p>Un algorithme est mis en œuvre par le candidat et permet de démontrer la résolution d'un problème de décision séquentiel simple.</p> <p>Les requêtes et l'accès à la base de données massive sont configurés correctement.</p> <p>Le schéma général des accès à l'application développée est présenté et prend en compte le référentiel d'accessibilité (RGAA).</p>
--	--	---	--

# RNCP Data Engineer – Niveau 7 (EU)

Référentiel d'activités, de compétences et d'évaluation

REFERENTIEL D'ACTIVITES <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	REFERENTIEL DE COMPETENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	REFERENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p><b>BLOC 4 : PILOTER LA MAINTENANCE DE L'ARCHITECTURE DE DONNÉES MASSIVES</b></p> <p><b>A1. Identification des usages et conditions d'utilisation sur la durée de vie de la solution ou de l'équipement d'intelligence artificielle.</b></p>	<p><b>C30.</b> Valider régulièrement la production de la solution d'intelligence artificielle en s'appuyant sur le retour d'expérience des utilisateurs en interne et en externe afin d'adapter les fonctionnalités et les caractéristiques techniques du projet d'intelligence artificielle.</p> <p><b>C31.</b> Assurer la pérennité de la solution informatique en surveillant l'évolution des données de travail afin d'éviter la dégradation des performances des modèles d'intelligence artificielle mis en production et assurer un niveau de sécurisation des données optimal défini par la DSI.</p>	<p><b>EPREUVE N°5 (C30 à C36)</b> <b>Mise en situation professionnelle</b> « Pilotage de la maintenance de l'infrastructure de données massives »</p> <p>L'objectif est d'évaluer la capacité du candidat à <b>analyser le retour d'expérience des utilisateurs de la solution ou de l'équipement d'intelligence artificielle, à rédiger les protocoles de maintenance et les procédures techniques/technologiques d'exploitation à destination des utilisateurs et à programmer des tests de surveillance.</b></p>	<p><b>C30 à C31 : Evaluation des points suivants :</b></p> <ul style="list-style-type: none"> <li>- Les expériences utilisateurs internes et externes sont recueillies et analysées.</li> <li>- Les fonctionnalités de la solution ou de l'équipement sont évaluées selon des critères définis en amont.</li> <li>- Les évolutions sont validées d'un point de vue technique.</li> <li>- Les performances en production de la solution sont surveillées.</li> <li>- Le niveau de sécurisation des données défini par la DSI est mis en œuvre.</li> </ul>
<p><b>A2. Rédaction d'un protocole de maintenance et la documentation technique/technologique d'exploitation mis à disposition des</b></p>	<p><b>C32.</b> Rédiger les protocoles de maintenance et les procédures techniques/technologiques d'exploitation à destination des entités</p>	<p>Dans le cadre d'une mise en situation le candidat devra démontrer sa capacité à piloter la maintenance de l'infrastructure de données</p>	<p><b>C32 : Evaluation des points suivants :</b></p> <ul style="list-style-type: none"> <li>- Les procédures garantissant la</li> </ul>



## RNCP Data Engineer – Niveau 7 (EU)

Référentiel d'activités, de compétences et d'évaluation

	en se basant sur des indicateurs-clés afin d'appliquer des correctifs dans le cadre de la résolution de bugs techniques et/ou fonctionnels.	<ul style="list-style-type: none"><li>• <b>Soutenance :</b> Pilotage de la maintenance de l'infrastructure de données massives est présentée lors d'une soutenance orale devant un jury.</li></ul>	définis. La dérive des données est détectée et mesurée. Les correctifs appliqués pour résoudre de bugs techniques et/ou fonctionnels sont présentés de façon argumentée.
--	---	--	--

# RNCP Data Engineer – Niveau 7 (EU)

## Référentiel d'activités, de compétences et d'évaluation

La certification professionnelle *Data Engineer* s'articule autour des quatre blocs de compétences suivants :

1. **ANALYSER LES BESOINS EN MATIERE D'INFRASTRUCTURES DE DONNÉES ADAPTÉS AU PROJET D'INTELLIGENCE ARTIFICIELLE**
2. **PILOTER LE PROJET DE DÉVELOPPEMENT DE L'INFRASTRUCTURE DE DONNÉES MASSIVES**
3. **CONCEVOIR LES ARCHITECTURES BIG DATA VALORISANT LES DONNÉES**
4. **PILOTER LA MAINTENANCE DE L'ARCHITECTURE DE DONNÉES MASSIVES**

La certification est délivrée au vu de la validation des quatre blocs de compétences. Les blocs de compétences sont capitalisables. La validation des quatre blocs de compétences est obligatoire pour l'obtention de la certification professionnelle. La validation partielle d'un bloc n'est pas possible.

### Glossaire :

**Apprentissage machine** : en anglais *machine learning* ; cela consiste à enrichir le corpus de connaissances d'une intelligence artificielle pour la rendre plus efficace dans sa prise d'information ou de décision. Cet apprentissage peut être guidé par un opérateur humain ou, autonome.

**Apprentissage profond** : en anglais *deep learning* ; forme d'apprentissage machine qui implique que les connaissances sont stockées sur des couches de plus en plus nombreuses, et donc de plus profondes. Cet apprentissage plus efficace imite la structure cérébrale de l'être humain.

**Données massives (ou mégadonnées)** : en anglais *big data* ; c'est la compilation et l'analyse d'un volume très important de données destinée à identifier des liens entre des actions, des événements, ou des conditions. L'objectif est d'identifier des liens de causalité et d'établir une loi de prédictive.

**Intelligence artificielle ou IA** : en anglais *artificial intelligence (ou AI)* ; il s'agit d'un système informatique qui pilote une action nécessitant une identification et un traitement complexes. Un tel système a par ailleurs la capacité d'apprendre et de s'auto-ajuster.

**Data Lake** : en français « lac de données », permet de stocker et d'analyser les données, et offre une flexibilité à moindre coût. Il permet de dégager de la valeur à partir de n'importe quel type de données. Le principal point fort d'un lac de données est de permettre de centraliser du contenu de différentes sources.

**Data Warehouse** : en français « entrepôt de données », désigne une base de données utilisée pour collecter, organiser, journaliser et stocker des informations provenant de bases de données opérationnelles et fournir ainsi un socle à l'aide à la décision en entreprise.

**ETL / ELT** : en anglais « Extract Transform Load » et « Extract Load Transform » ; il s'agit d'outils qui collectent et traitent (presque en temps réel) des données provenant de diverses sources pour ensuite les envoyer dans un espace de stockage (data Lake, data Warehouse) à des fins d'analyse et de Business Intelligence.