FRANCE compétences

REFERENTIELS

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.* »

Etudier en situation de handicap

Au sein de l'Ecole, des efforts sont fournis en permanence pour améliorer la qualité de l'accueil des apprenants en situation de handicap.
En effet, dans le cadre de la démarche sociale et environnementale, l'égalité des chances et l'éducation pour tous sont des droits essentiels que l'Ecole souhaite défendre pour tous ses apprenant(e)s.
La volonté de l'Ecole (et du groupe OMNES Education dont dépend l'Ecole) est non seulement de faciliter l'accès des apprenant(e)s en situation de handicap aux études supérieures, mais également de leur offrir un confort de travail au quotidien leur permettant de se consacrer en toute quiétudes à leurs études et à l'obtention de la certification.
Les apprenants en situation de handicap doivent entrer en contact avec le référent handicap de l'Ecole.
Le référent handicap de l'Ecole est l'interlocuteur privilégié qui permettra la mise en place d'un dispositif d'accompagnement personnalisé pour l'apprenant(e) et obtenir ainsi, la certification visée.
Le programme d'accompagnement développé et mis en place par le référent handicap de l'Ecole repose sur une volonté permanente d'adapter le cadre des études, compte tenu de la typologie du handicap et de ses contraintes tant matérielles qu'humaines.

Dans le cadre de l'obtention de la certification, le référent handicap en accord avec la direction des programmes et de l'Ecole, peut être amené à :

- Aménager le parcours d'admission et d'intégration
- Aménager les modalités de suivi des enseignements (pratiques et théoriques en centre de formation ou en entreprise) en lien avec les intervenants et les tuteurs entreprises
- Aménager les modalités d'évaluation ;
- Obtenir une assistance d'une tierce personne lors de l'évaluation ;
- Adapter les supports et le matériel nécessaires à la réalisation des évaluations.

****

DATA ENGINEER -

DATASCIENTEST

GROUPE OMNES EDUCATION

BLOC 1 : Concevoir un projet d'architecture technique de gestion de données

REFERENTIEL D'ACTIVITES	REFERENTIEL DE COMPETENCES	REFERENTIEL D'EVALUATION	
		MODALITÉS D'ÉVALUATION	**CRITÈRES D'ÉVALUATION**
A1 **Analyse des besoins métier** L'analyse du contexte métier est le socle du travail du Data Engineer. Elle concerne la compréhension des spécificités, des besoins et des enjeux de l'entreprise ou du secteur pour lequel il travaille. Cette compréhension aide à aligner les projets d'architecture sur la stratégie globale de l'entreprise. Le Data Engineer identifie les flux de données. Dans certains contextes, il peut même être nécessaire de cartographier ou de déchiffrer les processus métier pour déterminer comment les données sont générées et utilisées. Cela offre une vision claire des points d'intégration, des silos de données et des zones d'amélioration. En ce sens, le Data Engineer travaille avec les directions et les opérationnels des services impliqués dans l'entreprise pour identifier les besoins métiers. Le Data Engineer peut ainsi appréhender les besoins en données, les problèmes existants et les opportunités potentielles.	C1 **Identifier les besoins en architecture de gestion de données** en analysant la problématique métier, le fonctionnement de l'organisation et l'ensemble de ses flux de données afin de valider l'opportunité de développement d'un projet d'architecture.	**E1** **Mise en situation professionnelle (C1 à C6) :** Dans le cadre d'un projet sur la base d'un dossier documentaire, le candidat devra identifier les besoins en architecture, élaborer un système de veille en lien avec le projet, formaliser son périmètre et émettre des recommandations d'architecture. ● **Livrable écrit :** - Rapport préliminaire (C1) - Système de veille, rapport d'analyse (C2, C3) - Cahier des charges (C4) - Rapport explicitant les recommandations (C5, C6)	C1 - Le candidat doit rédiger un rapport préliminaire intégrant : ● L'analyse exploratoire des flux de données proposée. Elle permet d'identifier comment l'information se transmet et se transforme. Les données et leurs interactions sont schématisées ● La problématique métier identifiée est argumentée au regard des flux de données existants et des services souhaités ● La synthèse de l'existant reprend les éléments essentiels et permet d'identifier l'opportunité de développement d'un projet d'architecture
A2 **Conception et exploitation d'un processus de veille technologique et réglementaire** Au-delà de l'entreprise elle-même, il analyse et maîtrise le secteur d'activité dans son ensemble et ses évolutions. Cela implique de se tenir informé des tendances du marché, des réglementations, de la concurrence et des défis propres au secteur (dont l'impact environnemental).	C2 **Élaborer et exercer un système de veille technologique et réglementaire dédié au domaine du numérique en** recensant des sources vérifiées et en menant une analyse des informations récoltées en vue d'identifier les cas d'usages et les évolutions technologiques et réglementaires à intégrer dans son activité professionnelle.		C2 - Le candidat doit rédiger un rapport de veille : ● La stratégie de veille est argumentée au regard des informations recherchées ● Les sources sont vérifiées et actualisées ● Divers canaux sont exploités (alertes, réseaux sociaux professionnels, flux RSS, moteurs de recherche) ● Les résultats de la veille sont triés et permettent d'identifier les cas d'usages, tendances, outils et méthodologies du domaine ● Le cadre réglementaire impactant le projet est explicité : (CRA, NIS 2, RGAA, RGPD, RSE)

Sur la base de sa veille continue, le Data Engineer pourra définir, anticiper et adapter ses actions et émettre des recommandations répondant aux objectifs de l'entreprise.	C3 **Exploiter la veille, au sein de son organisation**, en remettant en cause ses pratiques à l'aune des évolutions et en diffusant les informations aux interlocuteurs adéquats dans le but d'informer, de sensibiliser et de faire adhérer ses collaborateurs et sa hiérarchie aux bonnes pratiques et au respect des normes réglementaires.		C3 - Le candidat doit rédiger un rapport de veille intégrant : ● La synthèse permet de prendre en compte dans le projet ● La synthèse est structurée et permet ainsi d'identifier les informations clés ● Le candidat démontre la prise en compte du cadre réglementaire en vigueur (CRA, NIS 2, RGPD, RGAA, des enjeux écologiques et d'inclusion des personnes en situation de handicap).
A3 **Cadrage de la solution** Avec une idée claire des besoins et des données à disposition, le Data Engineer détermine alors les spécifications techniques et fonctionnelles de la solution. Ces spécifications détaillent l'architecture à mettre en place, les outils à utiliser, les technologies à privilégier et la méthodologie à suivre. Dans cette phase, il évalue également les opportunités offertes par l'intelligence artificielle générative, en identifiant les cas d'usage pertinents et les bénéfices potentiels pour le projet. Si l'intégration d'une telle approche s'avère pertinente, il sélectionne alors les outils, modèles et méthodes adaptés (frameworks, APIs, plateformes spécialisées, etc.) afin d'optimiser la performance, la valeur ajoutée et la durabilité de la solution. Au-delà des aspects purement techniques, il doit également naviguer dans un environnement réglementaire et éthique complexe. Il prend ainsi en compte les réglementations en vigueur, comme le RGPD pour la protection des données. Le Data Engineer s'assure que la solution proposée respecte l'environnement, dans une démarche d'éco-conception et d'accessibilité à tous, y compris aux personnes en situation de handicap (RGAA). Finalement, avec toutes ces informations en main, le Data Engineer formule des recommandations concrètes à sa hiérarchie. Il les conseille sur la direction à suivre, les ressources à mobiliser, le budget à prévoir et les délais à respecter, s'assurant ainsi que la solution data proposée est non seulement techniquement adaptée, mais aussi alignée sur les aspirations et contraintes de l'entreprise.	C4 **Définir le périmètre du projet de gestion des données** en formalisant les besoins, les objectifs, les contraintes et les risques ainsi qu'en identifiant les sources de données, les spécifications techniques et fonctionnelles, et les enjeux d'accessibilité, réglementaires, écologiques et éthiques impliqués afin d'assurer son inclusivité, limiter son impact écologique et préparer sa mise en œuvre. C5 **Exploiter l'intelligence artificielle générative** en identifiant l'opportunité de l'intégrer dans le projet en gestion de données puis, le cas échéant en sélectionnant les outils et méthodes adaptés afin d'optimiser la solution. C6 **Émettre des recommandations et être force de propositions auprès de sa hiérarchie et d'une équipe pluridisciplinaire**, en présentant les axes de développement de la solution, en argumentant ses propositions et en adaptant sa communication à ses interlocuteurs afin que les propositions soient compréhensibles par l'ensemble des équipes et qu'elles puissent ainsi contribuer au projet.		C4 - Le candidat doit rédiger un cahier des charges comprenant : ● Les objectifs du projet ● Les besoins en architecture ● Les sources de données retenues pour le projet ● Les contraintes (volume de données à analyser en un temps donné, délai de réalisation du projet, disponibilité des données…) : Les éventuelles difficultés sont anticipées ● Les spécifications techniques et fonctionnelles générales de la solution ● Le candidat démontre le respect du cadre réglementaire (CRA, NIS 2, RGPD, RGAA) ● Les propositions prennent en compte les enjeux écologiques et limitent autant que possible les impacts sur l'environnement dans une logique de sobriété du numérique : - L'impact écologique de la solution est estimé - Le cycle de vie des ressources numériques du projet est déterminé (création/retrait/archivage) ● Le cadre réglementaire relatif aux personnes en situation de handicap est pris en compte lors du cadrage en intégrant les principes de conception universelle[1] et d'accessibilité pour : ○ Les utilisateurs de la solution ○ Les parties prenantes du projet C5 - Le candidat doit rédiger un rapport incluant ses recommandations : ● L'intérêt ou la non-pertinence de recourir à l'IA générative est démontrée au regard des objectifs, des données et du contexte du projet Le cas échéant : ● Le candidat identifie et compare plusieurs outils ou API d'IA générative (modèles de langage, services cloud, frameworks open source, etc.) en fonction des besoins techniques du projet. ● Il choisit les méthodes ou architectures d'intégration les plus adaptées (prompt engineering, API embedding, fine-tuning, workflow d'orchestration, etc.). C6 - Le candidat doit faire un rapport incluant ses recommandations : ● Les axes de développement de la solution sont présentés et argumentés au regard de la problématique métier

			• L'argumentaire est structuré • Les propos du candidat sont synthétisés • Le candidat utilise un vocabulaire adapté à son interlocuteur. • La prise de parole en public est fluide

[1] On entend par « conception universelle » la conception de produits, d'équipements, de programmes et de services qui puissent être utilisés par tous, dans toute la mesure possible, sans nécessiter ni adaptation ni conception spéciale. La conception universelle n'exclut pas les appareils et accessoires fonctionnels pour des catégories particulières de personnes handicapées là où ils sont nécessaires. (Convention relative aux droits des personnes handicapées du 30 mars 2007).

BLOC 2 : Elaborer une architecture technique de gestion de données

REFERENTIEL D'ACTIVITES	REFERENTIEL DE COMPETENCES	REFERENTIEL D'EVALUATION	
		MODALITÉS D'ÉVALUATION	**CRITÈRES D'ÉVALUATION**
A4 **Formalisation du processus d'extraction, de traitement et de stockage des données** La formalisation du processus d'extraction, de traitement et de stockage des données est au cœur du métier et vise à garantir que les données sont traitées de manière efficace, sécurisée et conforme aux diverses exigences. Le Data Engineer commence par identifier les données de référence de l'architecture. Ces données de référence servent de fondement à l'ensemble du système et sont essentielles pour assurer la cohérence et la fiabilité des informations traitées. Une fois ces fondations établies, le Data Engineer planifie les différentes étapes du processus. Cela commence généralement par l'extraction des données depuis diverses sources, qu'il s'agisse de bases de données, de flux en temps réel ou d'autres systèmes. Ensuite, ces données passent par des étapes de traitement où elles peuvent être nettoyées, transformées ou enrichies pour répondre aux besoins spécifiques du projet. Enfin, les données sont stockées, que ce soit dans des data warehouses, des data lakes ou d'autres structures de stockage adaptées. Le Data Engineer doit choisir parmi une multitude d'outils disponibles sur le marché, en fonction de leurs spécificités, de leur performance, de leur coût et de leur adaptabilité aux besoins du projet. Le Data Engineer doit également veiller à ce que le processus respecte le cadre réglementaire en vigueur. Dans un monde de plus en plus conscient des enjeux	C7 **Concevoir les procédures d'extraction, de traitement et de stockage des données** de l'architecture en schématisant les données de référence et les interactions attendues entre elles puis en définissant chaque étape d'extraction, de transformation et de chargement tout en assurant l'application de protocoles garantissant la sécurité des données, des réseaux et des systèmes et limitant l'impact écologique de la solution.	**E2** **Etude de cas (C7 à C13)** A partir d'un dossier documentaire, le candidat devra élaborer une architecture technique de gestion de données dans le respect de la réglementation et des exigences en termes de sécurité tout en limitant l'impact écologique de la solution. ● **Livrable écrit :** - Explicitation des procédures d'extraction, de traitement et de stockage (C7) - Base de données implémentée (C8, C9, C10) - Modélisation des données (C11) - ETL et rapport détaillant la configuration et les résultats (C12) - Présentation et justification des algorithmes d'intelligence artificielle intégrés à l'architecture (C13)	C7 - Le candidat doit fournir un rapport intégrant : ● Chaque étape du processus est argumentée au regard des objectifs poursuivis par le projet d'architecture : - Un schéma des données impliquées dans l'architecture est proposé et permet d'identifier les données de référence, les relations entre elles et les flux d'information attendus - Les procédures de nettoyage, de normalisation, d'agrégation permettent de garantir la qualité des données - Une solution de stockage des données est proposée et la procédure associée adaptée aux besoins du projet est présentée - Le choix de la méthode de chargement des données est argumenté, la procédure associée est détaillée - Les procédures de contrôle proposées permettent la vérification de la qualité des données et de l'opérationnalité de la solution - Le processus proposé permet de garantir la sécurité des données, des réseaux et des systèmes - Le choix des technologies et outils est argumenté pour chaque étape du processus proposé *(Airflow, Jenkins, Hadoop, Spark, Kafka, Hbase, Cassandra, langages de programmation Python, Scala, Java, librairies de web scrapping etc.)* - La réglementation applicable en matière de protection des données est explicitée ● La prise en compte des enjeux écologiques est démontrée : - Des mesures permettant de limiter l'empreinte environnementale de la solution sont proposées (cycle de vie des ressources numériques, énergies renouvelables, lutte contre l'obsolescence etc.) - L'impact écologique des procédures proposées est mesuré

environnementaux, le Data Engineer doit aussi considérer les implications écologiques de ses choix. Cela peut concerner la consommation énergétique des centres de données, le choix de technologies plus vertes ou même la réduction des redondances pour minimiser l'utilisation des ressources.			
A5 **Collecte de données structurées et non structurées** La collecte de données structurées et non structurées est une activité nécessairement exercée par le Data Engineer. Cela englobe la récupération, l'organisation et le stockage d'informations provenant de diverses sources en vue de leur exploitation ultérieure. Qu'elles soient structurées, organisées dans des formats classiques comme les tableaux, ou non structurées comme les textes, les images ou les vidéos, le Data Engineer doit pouvoir les récupérer et arbitrer sur les modalités de la collecte. Pour cela, des langages de programmation et des librairies spécifiques comme Python, Beautiful Soup ou Scrapy Python sont exploités. De plus, les API, ou Interfaces de Programmation d'Applications, sont souvent sollicitées pour collecter des données depuis des services ou des plateformes extérieurs. Une fois ces données extraites, elles doivent être organisées et stockées efficacement. Les bases de données relationnelles, généralement gérées par des systèmes comme SQL, sont une solution courante pour le stockage de données structurées. Elles offrent la possibilité d'organiser les informations en tables interconnectées, assurant ainsi une intégrité et une cohérence de l'ensemble. Face à la quantité toujours croissante de données, les bases de données traditionnelles peuvent parfois être insuffisantes. Les technologies NoSQL, sont en ce sens adaptées au stockage de données non structurées ou semi-structurées, offrant une flexibilité en matière de schéma et une capacité à gérer d'importants volumes de données. De plus, des systèmes sont conçus pour traiter et stocker de vastes ensembles de données sur des clusters de machines, garantissant ainsi une grande capacité d'évolutivité et de résilience (Hadoop, HDFS).	C8 **Collecter les données structurées et non structurées** de diverses sources utilisables pour le projet d'architecture via la programmation de scripts et dans le respect du cadre réglementaire et des procédures garantissant la sécurité des données, des réseaux et des systèmes afin de préparer les étapes de transformation et de stockage des données. C9 **Élaborer des solutions de stockage** en créant et structurant les bases de données relationnelles et/ou non- relationnelles (SQL, noSQL) le tout, dans le respect des procédures garantissant la sécurité des données, des réseaux et des systèmes afin de consolider le processus de stockage des données de la solution.		C8 - Le candidat doit fournir la base de données implémentée : ● Les scripts programmés permettent de collecter les données structurées et non structurées nécessaires pour le projet (Python, librairies Beautifulsoup, Selenium, GoogleSearch…) ● Les données sont collectées dans le respect du RGPD C9 - Le candidat doit fournir la base de données implémentée et les scripts associés : ● Les tables SQL sont structurées, la programmation est fonctionnelle *(Python)* ● Une architecture distribuée NoSQL est implémentée à la base de données *(Hadoop, MongoDB, Neo4j etc.)* ● Les protocoles appliqués garantissent la sécurité des données, des réseaux et des systèmes. ● La base de données est chargée ● La base de données est requêtée selon des instructions répondant aux besoins du projet

A6 **Transformation et analyse de données** L'analyse de données menée par le Data Engineer consiste non seulement à préparer les données pour l'analyse, mais également à les exploiter pour en tirer des informations pertinentes. Lorsqu'il s'agit de transformer les données, le Data Engineer utilise des langages de programmation (Python, Java, Scala). Cette transformation implique plusieurs étapes. La première est le nettoyage des données, ensuite, il peut être nécessaire de combiner des données provenant de sources internes et externes pour obtenir un ensemble de données complet et cohérent. Le formatage est également une étape essentielle, car il assure que les données sont présentées de manière uniforme, facilitant ainsi les analyses ultérieures. Après avoir préparé les données, le Data Engineer les exploite pour en extraire des informations significatives. Cela nécessite souvent l'utilisation de méthodes et d'outils d'analyse statistique. De plus, la visualisation de données est un aspect fondamental de cette activité. Elle permet de présenter les résultats d'une manière facilement compréhensible, offrant des insights visuels sur les tendances, les modèles et les relations au sein des données. (Python, Matplotlib, Bokeh, Power Bi, Looker Studio,Tableau).	C10 **Transformer les données en un format approprié pour l'analyse** en nettoyant les jeux de données et en y appliquant des modifications à l'aide d'outils dédiés afin de rendre les données disponibles et exploitables dans leur forme et leur contenu. C11 **Analyser les données** en s'appuyant sur des méthodes et outils d'analyse statistique et de visualisation de données afin d'évaluer l'intégrité et la qualité des données et de présenter les résultats issus de l'analyse aux parties prenantes du projet et aux utilisateurs finaux de la solution.		C10 - Le candidat doit fournir la base de données implémentée : ● Les transformations nécessaires sont appliquées aux données de façon à les rendre exploitables pour la solution *(Python, Java, Scala...)* - formatage - jointures - déduplication - etc. ● Les requêtes mises en place permettent de vérifier l'intégrité et la qualité des données en lien avec les transformations implémentées C11 - Le candidat doit fournir une modélisation des données : ● Un outil de visualisation des données est exploité *(Power Bi, Looker Studio, Tableau*), son rendu permet d'évaluer la qualité et l'intégrité des données et de valider ou d'identifier les écarts à corriger quant aux transformations implémentées ● Le choix de l'outil de visualisation est justifié en lien avec la/les typologie(s) des données. ● La visualisation de données proposée est lisible et compréhensible au grand public (RGAA).

A7 **Création de circuits de flux de données** L'automatisation des flux de données est primordiale pour garantir l'efficacité et la régularité des opérations. Le Data Engineer met ainsi en place un processus ETL automatisant le processus (ETL : Extraction, Transformation et Chargement) Le processus commence par la collecte de données depuis leurs sources originales, qu'elles soient internes ou externes. Une fois collectées, ces données sont transformées et chargées vers les environnements de stockage appropriés, qu'il s'agisse de bases de données relationnelles, de lacs de données ou d'autres systèmes de stockage. Les tests unitaires sont une autre composante de cette activité. Ils garantissent que chaque partie du flux de données fonctionne comme prévu. En examinant chaque unité ou composant individuellement, le Data Engineer peut s'assurer que les données sont correctement collectées, transformées et chargées. Si un test unitaire détecte un problème, il peut être résolu avant que le flux de données ne soit mis en production. Cela contribue à prévenir les erreurs potentielles, les incohérences ou les problèmes de qualité des données.	C12 **Automatiser les circuits de collecte, de traitement et de stockage des données** en s'appuyant sur l'architecture élaborée, en exploitant des outils dédiés, et en testant le processus d'automatisation en vue d'assurer l'opérationnalité de la solution technique.		C12 - Le candidat doit fournir un ETL et un rapport détaillant la configuration et les résultats : ● Les circuits de flux de données sont configurés et permettent d'automatiser la collecte, la transformation et le chargement des données structurées et non structurées vers les environnements de stockage ● Les tests réalisés permettent de valider l'opérationnalité de l'automatisation des flux de données ● Les protocoles appliqués garantissent la sécurité des données, des réseaux et des systèmes. ● Une documentation détaillée sur les circuits automatisés, (étapes, paramétrages, tests, gestion des erreurs) est fournie

A8 **Développement d'algorithmes d'intelligence artificielle** Le développement d'algorithmes d'intelligence artificielle est l'une des missions avancées du Data Engineer. Cela implique de concevoir, développer des modèles pour traiter et interpréter d'importants ensembles de données afin d'obtenir des insights ou d'automatiser des décisions. Le Data Engineer entame un processus détaillé qui débute par le choix du type de modèle le plus adapté au problème en question, qu'il s'agisse d'apprentissage supervisé, non supervisé ou par renforcement en fonction de la nature des données et des objectifs du projet. Il s'appuie, pour cela, sur les expertises métier nécessaires, afin d'assurer la pertinence des approches mises en œuvre au regard du contexte opérationnel. Ensuite, il définit clairement les objectifs de l'algorithme, c'est-à-dire ce qu'il espère accomplir avec ce modèle, et établit les critères qui seront utilisés pour évaluer l'efficacité du modèle. Cela pourrait concerner, par exemple, la précision de la prédiction ou la qualité du regroupement des données. Le choix des outils et des frameworks est une étape essentielle. Avec une variété d'outils disponibles, (TensorFlow, PyTorch, …). Le Data Engineer sélectionne celui qui correspond le mieux aux besoins du projet. Enfin, après avoir développé le modèle, il est essentiel de mesurer ses performances en conditions réelles. Cela implique de tester le modèle sur des données pour s'assurer qu'il fonctionne comme prévu et d'apporter les ajustements nécessaires.	C13 **Développer un algorithme d'intelligence** artificielle en s'appuyant sur les expertises métier nécessaires, en utilisant des méthodes et outils d'apprentissage supervisé et/ou non supervisé dans le respect des principes éthiques et de frugalité[2] en vue d'intégrer des composants d'intelligence artificielle à la solution technique de gestion de données.		C13 - Le candidat doit présenter un algorithme d'intelligence artificielle : ● Le choix du type de modèle est argumenté au regard du besoin métier (supervisé, non supervisé, renforcement) ● Le langage de programmation approprié est exploité, le code fournit démontre sa maîtrise *(Python, Scala, Java, Shell, VBA, C++, Perl, Ruby)* ● Les frameworks appropriés sont exploités *(TensorFlow, Scikit-Learn)* ● L'algorithme est optimisé en termes de ressources tout en restant opérationnel ● La performance du modèle est démontrée via la réalisation de tests, la pertinence des résultats au regard de la problématique métier est justifiée ● Des mesures d'optimisation du modèle sont proposées et argumentées au regard de ses performances et des besoins de l'architecture

[2] une IA durable, respectueuse de l'environnement cf.https://www.healthandtech.eu/fr/tour/news/16181/ia-recommandations-cour-comptes-ameliorer-positionnement-france.html

BLOC 3 : Déployer une solution d'analyse de données massives intégrant l'intelligence artificielle

REFERENTIEL D'ACTIVITES	REFERENTIEL DE COMPETENCES	REFERENTIEL D'EVALUATION	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
A9 **Mise en production de la solution pour l'ensemble des utilisateurs** La mise en production implique de traduire les travaux préliminaires et les prototypes en applications fonctionnelles, fiables et sécurisées accessibles à tous les utilisateurs cibles. La première étape est la conception d'une interface de programmation, qui sert de point d'entrée pour que d'autres systèmes ou applications interagissent avec la solution. Ensuite vient la sélection et l'exploitation des API, qui sont des ensembles de protocoles et d'outils permettant à différentes applications de communiquer entre elles. Pour assurer un déploiement flexible et évolutif des modèles, le Data Engineer peut opter pour une solution de virtualisation comme les containers. L'utilisation de technologies comme Docker permet de créer des environnements isolés, reproduisibles et légers pour exécuter les applications, garantissant ainsi qu'elles fonctionnent de manière cohérente, quelle que soit la machine hôte. Une fois le modèle prêt pour la mise en production, il est implanté dans un environnement dédié, où il sera exécuté en conditions réelles, répondant aux demandes des utilisateurs et interagissant avec d'autres systèmes.	C14 **Concevoir une interface de programmation entre les composants de la solution**, en utilisant les API appropriées dans le respect des exigences de sécurité afin de préparer le déploiement de la solution d'analyse de données intégrant l'intelligence artificielle. C15 **Conteneuriser les composants de l'architecture** en créant des images (virtualisation) pour chacun d'eux, en configurant les conteneurs en prévoyant des mécanismes de gestion d'erreur et les mesures de sécurité appropriées pour protéger les conteneurs contre les vulnérabilités et les accès non autorisés.	**E3** **Etude de cas (C14 à C19) :** A partir d'un cas réel ou fictif, le candidat devra déployer une solution d'analyse de données massives intégrant l'intelligence artificielle. Le candidat devra ensuite orchestrer les services de la solution, contrôler la mise en production et automatiser le déploiement de nouvelles versions. ● **Livrable écrit :** Présentation de chaque étape, justifiant les choix, technique et le respect du cadre réglementaire : - Présentation de la configuration de l'interface de programmation (C14) - Image du conteneur par composant (C15) - Scripts de déploiement et d'orchestration (C16, C17) - Scripts / configuration d'automatisation (CI/CD) (C19) - Script des tests unitaires (C18) - Solution configurée (C15 à C19)	C14 - Le candidat doit fournir la solution configurée et un rapport : ● Les configurations de l'interface de programmation sont explicitées et permettent le bon déploiement de la solution ● Le choix des API et de leur architecture est justifié au regard des caractéristiques de la solution et de la mise en production à effectuer ● L'interface de programmation est paramétrée dans le respect des exigences de sécurité C15 - Le candidat doit fournir une virtualisation descomposants de l'architecture : ● Les images fonctionnelles sont créées pour chaque composant (Docker) ● Les conteneurs sont opérationnels ● Les conteneurs sont connectés au stockage distant ● Les mesures de sécurité protégeant les conteneurs contre les vulnérabilités et les accès non autorisés sont appliquées

L'exploitation de librairies spécifiques, telles que Fast API ou Flask, facilite la création et la gestion d'interfaces web pour les solutions. Ces outils offrent des moyens rapides et efficaces de créer des points d'accès pour les utilisateurs ou d'autres services. Enfin, un aspect crucial est la sécurisation de l'interface de la solution. Avec les préoccupations croissantes concernant la confidentialité des données et les menaces cyber, il est essentiel de s'assurer que la solution est robuste face aux intrusions, aux fuites de données ou à toute autre forme de menace.	C16 **Déployer le modèle dans un environnement de production** en implémentant le modèle dans le système via des librairies dédiées dans l'optique de le rendre opérationnel dans le respect des spécifications fonctionnelles et des bonnes pratiques du domaine.		C16 - Le candidat doit fournir la solution configurée et le script de déploiement : ● Le modèle est implémenté dans l'environnement de production (cloud, serveur ou infrastructure de calcul distribué) sans dégradation des performances du systèmes ou des temps d'arrêts excessifs (Docker, Kubernetes) ● Le modèle s'exécute avec les bonnes versions de dépendance (bibliothèques, packages, modules) sans générer d'erreur (pip Python, npm : Node.js, Maven : Java) ● Le modèle est déployé de manière à utiliser efficacement les ressources : ressources CPU, mémoire et GPU (Prometheus, Grafana) ● Les données sensibles sont correctement cryptées ● L'accès au modèle est restreint aux utilisateurs autorisés
A10 **Orchestration des différents services pour une solution complètement opérationnelle** L'orchestration des différents services par le Data Engineer est une tâche essentielle pour garantir qu'une solution soit fluide, performante et résiliente. Cela revient à coordonner et gérer l'interaction entre différents composants ou services pour assurer le bon fonctionnement d'une application ou d'un système. Dans les architectures modernes, au lieu d'avoir une application monolithique, on décompose les fonctionnalités en services plus petits et indépendants, appelés microservices. Ces services doivent communiquer entre eux de manière efficace pour réaliser une tâche ou fournir une fonctionnalité. Le Data Engineer s'assure que ces liaisons sont non seulement établies mais aussi optimisées pour des performances maximales.	C17 **Orchestrer les services de la solution** en implémentant un processus de gestion et de coordination des composants du système afin de garantir l'exécution fluide et efficace de l'ensemble de l'architecture et de créer une solution évolutive, capable de résoudre des problèmes complexes et de fournir des résultats fiables.		C17 - Le candidat doit fournir la solution configurée et le script d'orchestration : ● Les différents services sont intégrés et interagissent correctement (Kubernetes) ● Les microservices sont coordonnés de façon à exécuter des flux de données complexes en produisant des résultats cohérents ● Le système adapte les ressources en temps réel en fonction de la charge ● Les éventuelles erreurs sont identifiées, les logs fournissent des informations pertinentes pour le dépannage ● L'ajout de nouveaux services ou fonctionnalités ne nécessite pas de modifications majeures du système

Ensuite, la gestion de la charge et la mise à l'échelle sont vitales, surtout dans les environnements où le volume des données ou le nombre de requêtes peut varier considérablement. Le Data Engineer doit s'assurer que le système peut s'adapter dynamiquement à ces variations. Cela peut signifier l'ajout de ressources lors de pics de demandes ou leur réduction pendant les périodes creuses pour optimiser les coûts et les performances. En vue d'assurer la fiabilité et la tolérance aux pannes, le rôle du Data Engineer est d'anticiper ces scénarios et de mettre en place des mécanismes pour que le système puisse récupérer rapidement de ces pannes, minimisant ainsi les interruptions et garantissant une disponibilité continue.			
A11 **Validation et automatisation de la mise en production** La validation et l'automatisation de la mise en production sont des étapes clés du travail du Data Engineer pour assurer que les solutions déployées soient fiables, performantes et conformes aux exigences. Les tests unitaires se concentrent sur des parties isolées (ou "unités") de la solution, vérifiant que chaque composant fonctionne comme prévu. L'étape suivante est la mise en place d'un pipeline CI/CD, qui signifie "intégration continue" et "déploiement continu". Dans un tel système, dès qu'une modification est apportée à la solution, celle-ci est automatiquement testée (intégration continue) pour s'assurer qu'elle n'entraîne pas d'erreurs ou de dysfonctionnements. Si ces tests sont réussis, la solution est alors automatiquement déployée dans l'environnement de production, assurant ainsi que les utilisateurs bénéficient toujours de la version la plus récente et la plus fiable (déploiement continu). En combinant des tests unitaires rigoureux avec un	C18 **Contrôler la mise en production de la solution** grâce à des tests unitaires afin d'assurer l'opérationnalité de la solution et vérifier sa conformité quant aux spécifications établies lors de la constitution du cahier des charges. **C19** **Automatiser le déploiement de nouvelles versions de la solution** et son monitoring notamment concernant l'évolution de ses données à l'aide d'un outil de continuous Integration (CI/CD) permettant de surveiller efficacement l'ensemble du processus, de prévenir toute dégradation des performances et d'assurer la durabilité de la solution.		C18 - Le candidat doit fournir la solution configurée et le rapport de tests : ● Les tests unitaires sont créés et déployés pour l'ensemble des fonctionnalités de l'API ● L'opérationnalité de la solution est évaluée, les éventuels ● Les manquements sont identifiés et caractérisés ● En cas d'anomalie, des actions correctives sont proposées et explicitées C19 - Le candidat doit fournir la solution configurée et les scripts associés ● Les scripts proposés permettent d'automatiser la création et la configuration de serveurs (Jenkins, Gitlab, CI/CD) ● L'automatisation du déploiement de l'infrastructure est fonctionnelle (Ansible, Terraform), les données restent cohérentes et utilisables après chaque mise à jour ● Les métriques de performance et de disponibilité sont collectées et surveillées en temps réel (Prometheus, Grafana) ● Les versions précédentes, les modifications et les mises à jour sont enregistrée (Git)

pipeline CI/CD efficace, le Data Engineer s'assure que les solutions mises en production sont non seulement conformes aux attentes initiales, mais qu'elles sont également constamment mises à jour et améliorées de manière fluide, sans interrompre l'expérience des utilisateurs finaux. Ce processus renforce la confiance dans la solution, minimise les risques associés aux déploiements manuels et augmente la réactivité face aux besoins changeants			

BLOC 4 : Piloter un projet d'architecture technique de gestion de données

REFERENTIEL D'ACTIVITES	REFERENTIEL DE COMPETENCES	REFERENTIEL D'EVALUATION	
		MODALITÉS D'ÉVALUATION	**CRITÈRES D'ÉVALUATION**
A12 **Détermination de l'organisation du projet** Cette étape vise à structurer et à planifier l'ensemble du projet pour assurer son bon déroulement, depuis son lancement jusqu'à sa finalisation. Le Data Engineer cartographie l'ensemble des activités nécessaires pour mener le projet à bien. Cela implique de décomposer le projet en différentes phases ou étapes, allant de la collecte des besoins initiaux à la mise en production finale de la solution. Une fois ces étapes définies, le Data Engineer procède à l'évaluation de la durée estimée de chaque tâche. Il s'agit d'anticiper le temps nécessaire pour chaque activité afin d'établir un calendrier réaliste en fonction des contraintes et des ressources disponibles. L'identification des parties prenantes à intégrer à l'équipe projet est également essentielle. Un projet de data engineering peut nécessiter des compétences variées : du développement à l'analyse de données, en passant par la gestion de projet ou même des experts métier. Il est donc important d'identifier qui, au sein de l'organisation ou en externe, doit être intégré à l'équipe pour apporter son expertise.	C20 **Définir la structure organisationnelle du projet** d'architecture technique de gestion de données en planifiant les différentes étapes du projet et en identifiant les parties prenantes à intégrer, ceci en prenant en compte les situations de handicap afin que le projet bénéficie d'un mode de gouvernance rationalisé et d'un outil de pilotage répondant aux exigences du cahier des charges.	**E4** **Mise en situation professionnelle (C20, C21, C23, C24, C25):** Sur la base d'un projet d'architecture de données, le candidat devra définir la structure organisationnelle du projet. Le candidat coordonnera ses actions avec les autres membres de son équipe en vue de définir les modalités d'organisation. Le candidat devra soumettre les éléments suivants : ● **Livrable écrit :** - Un rapport détaillant les modalités de mise en œuvre et d'optimisation du projet (C20, C21, C24) - Un plan d'accompagnement (C25) ● **Présentation orale** dudit projet via la simulation d'un reporting, et auprès de sa hiérarchie (C21, C23, C25)	C20 - Le candidat doit fournir un rapport détaillant le plan de projet : ● Les différentes étapes sont décrites, listées, priorisées, et leurs durées estimées ● Les parties prenantes à intégrer sont présentées et leurs rôles et tâches déterminés au vu des objectifs poursuivis par le projet ● Les objectifs répondent aux critères des objectifs SMART (spécifique, Mesurable, Acceptable Réaliste, Temporellement défini.) ● Le cadre réglementaire relatif aux personnes en situation de handicap est pris en compte

La prise en compte des parties prenantes en situation de handicap est un élément important pour assurer une inclusion totale. Cela signifie non seulement de veiller à ce que ces personnes puissent participer pleinement au projet, mais aussi d'adapter, si nécessaire, les outils, les méthodes ou l'environnement de travail pour répondre à leurs besoins spécifiques.		**E5** **Etude de cas (C22) :** Sur la base de données d'un projet, le candidat devra mener une analyse des données en vue d'identifier les écarts budgétaires entre les dépenses réelles et le budget prévisionnel, en évaluant les raisons possibles de ces écarts. Le candidat devra identifier et argumenter des mesures correctives permettant de réduire les écarts constatés. **Livrable écrit :** - Analyse financière - Mesures correctives proposées	
A13 **Supervision du projet** La supervision du projet est une dimension essentielle de l'activité du Data Engineer, surtout lorsqu'il occupe un rôle de lead ou de responsable d'équipe. Elle concerne la coordination, le suivi et l'évaluation de l'ensemble des aspects du projet pour assurer son bon déroulement et sa réussite. Dans les méthodologies agiles, l'accent est mis sur la flexibilité, la collaboration et la capacité à s'adapter rapidement aux changements. Le Data Engineer, en tant que superviseur, doit donc s'assurer que l'équipe suit bien les principes et pratiques agiles, comme les sprints, les stand-ups quotidiens ou les revues de sprint. Il veille également à ce que l'équipe reste focalisée sur les objectifs, tout en conservant une dynamique de travail collaborative et réactive. La gestion budgétaire du projet est prise en compte par le Data Engineer. Tout projet comporte des contraintes financières, qu'il s'agisse des coûts liés aux ressources humaines, aux outils, à l'infrastructure ou à d'autres dépenses. Il doit donc surveiller de près le budget alloué au projet, s'assurer que les dépenses restent dans les limites définies et, si nécessaire, procéder à des ajustements pour éviter tout dépassement. Le reporting de l'avancement et des résultats du projet va de pair avec cette activité de supervision. Il s'agit de fournir des mises à jour aux parties prenantes sur l'état d'avancement du projet, les réalisations, les défis	C21 **Encadrer le développement du projet d'architecture de données** via des méthodes de gestion d'équipe et de projet agiles, en prenant en considération les situations de handicap afin de fédérer les acteurs du projet, d'assurer son inclusivité et de monitorer les avancées. C22 **Gérer le budget du projet** en contrôlant régulièrement les dépenses afin d'identifier les écarts avec le budget prévisionnel et, le cas échéant, de prendre des mesures correctives permettant de respecter les contraintes financières du projet. C23 **Réaliser des reportings** auprès des parties prenantes et de sa hiérarchie en communiquant régulièrement les avancements et les résultats rs afin de diffuser l'information permettant la prise de décision et l'implication de l'ensemble des équipes.		C21 - Le candidat doit fournir un rapport détaillant les modalités de mise en œuvre du projet : ● Les méthodes agiles exploitées sont explicitées, leur choix est argumenté au regard du contexte du projet et des acteurs impliqués (Scrum, Kanban), outils de suivi de tâches (Jira, Trello) ● Une matrice RACI est proposée et justifiée en lien avec les objectifs et les acteurs du projet ● Des mesures sont proposées en vue de favoriser l'inclusion des personnes en situation de handicap. La maîtrise de la réglementation en la matière est démontrée (accessibilité universelle, dispositifs d'aide accueil en milieu ordinaire etc.) C22 - Le candidat doit fournir une analyse financière et des propositions d'actions correctives ● Les écarts budgétaires entre le budget prévisionnel et les charges réelles sont identifiés et caractérisés ● Les mesures correctives proposées et argumentées permettent d'optimiser le budget sans remettre en cause la viabilité du projet C23 – Le candidat doit effectuer un reporting : ● Chaque étape du projet est explicitée, l'argumentaire est structuré ● Les propos du candidat sont synthétisés ● Le candidat utilise un vocabulaire adapté à son interlocuteur. ● La prise de parole en public est fluide

rencontrés et les perspectives futures. Cela permet non seulement de garantir une transparence totale, mais également de s'assurer que toutes les parties impliquées sont informées et alignées sur les objectifs et les résultats attendus.			
A14 **Amélioration continue** A travers le pilotage de l'amélioration continue, le Data Engineer veille à la pertinence et à l'efficacité des solutions sur le long terme. La première étape de ce processus est la détermination des indicateurs clés de performance, ou KPI. Ces indicateurs permettent au Data Engineer de mesurer précisément l'efficacité et la pertinence d'une solution ou d'un projet. Une fois les KPI définis, il convient d'évaluer régulièrement les performances. Cela se fait en suivant ces KPI, mais aussi en prenant en compte les retours des utilisateurs. Ces retours peuvent révéler des problèmes non détectés par les KPI ou donner des indications sur des besoins non satisfaits ou de nouvelles fonctionnalités souhaitées. Avec ces informations en main, le Data Engineer peut identifier des axes d'amélioration. Pour assurer l'adoption et le bon usage de la solution par les utilisateurs finaux, le Data Engineer met souvent en place un plan d'accompagnement. Celui-ci est conçu pour répondre aux besoins des utilisateurs et leur faciliter la prise en main de la solution. En formalisant ce plan d'accompagnement, le Data Engineer s'assure que les utilisateurs sont non seulement équipés pour utiliser la solution mais qu'ils la comprennent et peuvent l'intégrer efficacement dans leurs routines de travail.	C24 **Evaluer la performance** du projet de Data Engineering via des métriques adaptées au contexte et des outils d'analyse afin d'identifier des axes d'amélioration basés sur l'analyse des KPI, des retours d'expériences utilisateurs et de garantir ainsi une optimisation continue du projet. C25 **Former les utilisateurs finaux de la solution** en concevant et déployant un plan d'accompagnement anticipant les chantiers de transformation dans le but d'assurer la transition l'adoption et l'utilisation de la solution dans le respect des règles établies par tous les utilisateurs concernés.		C24 – Le candidat doit fournir un rapport détaillant les modalités de mise en œuvre et d'optimisation du projet intégrant : ● Des métriques quantifiables sont déterminées et permettent d'évaluer l'avancement et l'efficacité du projet (KPI) ● La mesure de l'impact environnemental de la solution est pris en compte ● Le niveau de performance du projet est mesuré en lien avec les objectifs initiaux et chaque KPI ● Le niveau de performance et les axes d'amélioration proposés prennent en compte les feedbacks utilisateurs ● Les axes d'améliorations proposés sont argumentés en lien avec les objectifs métiers poursuivis par le projet SMART (spécifiques, Mesurables, Acceptables, Réalistes, Temporellement définis.) C25 – Le candidat doit fournir un plan d'accompagnement : ● Le plan prend en compte des différentes typologies d'utilisateurs et de leurs niveaux de compétence. ● L'ensemble des aspects nécessaires pour la prise en main de la solution sont intégrés au plan ● Le plan est structuré (par exemple, séquencement logique, répartition équilibrée des sujets).