

ÉVALUATION MISE EN SITUATION PROFESSIONNELLE RECONSTITUÉE

Certification professionnelle Expert en ingénierie des données– EID- RNCP N°36921

EVALUATION CERTIFICATIVE

Bloc de compétences n° 3 : Construction d'une plateforme Big Data permettant la collecte, l'assemblage, le traitement et le stockage des données générées par les systèmes d'une entreprise

MSPR TPRE843 - Conception de l'architecture et construction de l'infrastructure d'une plateforme Big Data sécurisée à partir d'une situation réelle ou reconstituée

DATE DE SOUTENANCE DE LA MSPR :

CAMPUS DE :

--	--

CANDIDAT/CANDIDATE

Nom - Prénom

JURYS PROFESSIONNELS

Jury professionnel 1* (Président du Jury) :

- Nom - Prénom

- Entreprise

Jury professionnel 2* :

- Nom - Prénom

- Entreprise

Compétences professionnelles

La Mise En Situation Professionnelle Reconstituée permet l'évaluation des compétences professionnelles listées dans le Bloc de compétences n° 3 : Construction d'une plateforme Big Data permettant la collecte, l'assemblage, le traitement et le stockage des données générées par les systèmes d'une entreprise

Légende :

- Niveau ND : Non démontrée
- Niveau 1 : Niveau de base / Compétence dont la connaissance est partielle – Pratique superficielle
- Niveau 2 : Niveau intermédiaire / Pratique correcte – Maîtrise de l'attitude
- Niveau 3 : Niveau le plus élevé / Pratique approfondie – Possibilité de pouvoir transmettre cette compétence à d'autres

BC 03 – CONSTRUCTION D'UNE PLATEFORME BIG DATA PERMETTANT LA COLLECTE, L'ASSEMBLAGE, LE TRAITEMENT ET LE STOCKAGE DES DONNÉES GÉNÉRÉES PAR LES SYSTÈMES D'UNE ENTREPRISE.

Compétences	Critères d'évaluation	Niveaux de maîtrise des compétences			
		ND	1	2	3
<ul style="list-style-type: none"> • Concevoir une architecture de collecte et de restitution de données robuste, évolutive, sécurisée et utilisant l'intelligence artificielle (machine learning) afin d'améliorer en continu sa capacité à prédire les besoins Data des experts métiers utilisateurs 	<p>Le/la candidate présente une architecture de données sous forme de plan précisant ce qu'il en est de l'importation et de la mise à disposition des données (couche 1- collecte de données et couche 3 – restitution des données)</p> <p>L'architecture de données présentée doit :</p> <ul style="list-style-type: none"> - s'adapter aux changements des sources de données liées aux technologies internet (plateformes de réseaux sociaux, IOT) - nécessiter le minimum de maintenance et d'entretien donc prévoir l'automatisation de l'assimilation et la distribution des données - utiliser le machine learning et l'I.A. pour répondre aux changements, réparer les données erronées et améliorer en continu sa capacité à prédire les besoins Data 				
<ul style="list-style-type: none"> • Concevoir une architecture de stockage de données (data warehouse, data lake...) permettant de répondre aux besoins Data des experts métiers et respectant la politique de sécurité des données définie par le/la RSSI 	<p>Le candidat/la candidate établit une architecture de stockage de données précisant les éléments suivants (Couche 2 – stockage des données) :</p> <ul style="list-style-type: none"> -le type de stockage choisi pour un data lake, un data warehouse et un data mesh ; -son choix technologique de stockage le plus adapté au cahier des charges pour sécuriser l'environnement data (Cloud, Cloud Hybride, Edge Computing...); -les capacités de stockage préconisées ; -les préconisations pour répondre aux indicateurs RSE retenus dans le système de gouvernance de données. <p>Il/elle justifie que les 5V soient respectées : véracité, volumétrie, vitesse, valeur et variété.</p>				
<ul style="list-style-type: none"> • Définir les processus de collecte et d'intégration de données par lot ou en streaming afin de favoriser la vitesse d'intégration et la volumétrie de données dans le respect de l'architecture définie 	<p>Le candidat/la candidate est capable de</p> <ul style="list-style-type: none"> - lister les sources de données, leur type et sélectionner les sources d'entrées et de destinations ; - concevoir des jobs de collecte de données en utilisant les fonctionnalités des outils de type ETL - Il/elle définit les process pour une solution de migration, si nécessaire, inter-systèmes et multi-environnements permettant l'intégration de données diverses, structurées ou non, assurant l'interopérabilité des différentes solutions. 				

Compétences professionnelles

Compétences	Critères d'évaluation	Niveaux de maîtrise des compétences			
		ND	1	2	3
<ul style="list-style-type: none"> Mettre en place un système d'ingestion de données structurées et non structurées afin de permettre la manipulation et l'accès aux données ainsi que l'authentification des utilisateurs 	<p>Le candidat/la candidate est capable de justifier ses choix en matière de gestion de données :</p> <ul style="list-style-type: none"> -en fonction des problèmes à traiter, de la nature des données et de la politique des licences liées à la data gouvernance. -le choix du modèle de datamarts (SSAS, PDI...) pour le mode multidimensionnel, -le choix de stockage pour le mode non structuré.. 				
<ul style="list-style-type: none"> Développer une solution de migration inter-systèmes et multi-environnements à l'aide d'un outil de son choix afin de permettre l'intégration de données diverses et l'interopérabilité des différentes sources de données 	<p>Le candidat/la candidate est capable de développer une solution de migration inter-systèmes et multi-environnements. Il/elle utilise un outil de migration de son choix (Data Migration Toolkit, ETL Integrator, Scriptella, Talend...)</p>				
<ul style="list-style-type: none"> Développer un pipeline de données et/ou un pipeline ETL prenant en compte l'environnement technologique déployé (infrastructure, services, applications...) dans le respect du cahier des charges de la solution proposée 	<p>Le candidat/la candidate est capable de développer un pipeline de données ou un pipeline ETL à partir du cahier de charge établi.</p> <ul style="list-style-type: none"> - Il/elle installe et configure un outil de pipelines de son choix. <p>Ensuite, il/elle développe un pipeline dans le respect du cahier des charges donné.</p>				
<ul style="list-style-type: none"> Créer un lac de données (data lake) afin de collecter des données brutes dans le respect de l'architecture de collecte des données définie dans la solution proposée 	<p>Le candidat/la candidate est capable de créer un Data lake en expliquant les éléments suivants :</p> <ul style="list-style-type: none"> -le serveur de données utilisé, -le catalogue de données structurées -le moteur de requêtes de données -le moteur de recherche élastique de données non structurées ou semi structurées -l'outil d'affichage de données <p>Il/elle utilise les outils de son choix (AWS Glue, Amazon Athena, ES, Quicksight, etc...)</p>				
<ul style="list-style-type: none"> Créer un entrepôt unique à partir du référentiel de données établi pour centraliser les informations stratégiques de l'entreprise et répondre rapidement aux besoins métiers 	<p>Le candidat/la candidate est capable concevoir et d'argumenter le choix du modèle multidimensionnel (en étoile en flocon et grappe)</p> <p>Il/elle déploie son modèle multidimensionnel (un entrepôt ou d'un cube) à l'aide d'un outil de son choix dans une solution BI (Business intelligence)</p>				
<ul style="list-style-type: none"> Déployer un processus de collecte, stockage et traitement de données selon une approche ETL (Extract-Transform-Load) ou une approche ELT afin de permettre l'extraction, le stockage et le traitement des données de manière optimale et adaptée aux besoins utilisateurs métiers. 	<p>Le candidat/la candidate est capable de :</p> <ul style="list-style-type: none"> - développer des jobs de collecte de données en utilisant les fonctionnalités d'un outil de type ETL de son choix ; - développer dans un langage approprié de son choix des routines et des fonctionnalités de transformation ; - programmer des flux de données de bout en bout en utilisant une plateforme à base de flux de type Data Flow. <p>Il/elle justifie du respect des trois phases du processus d'intégration des données et prenant en compte les critères du nombre de connecteurs possibles (en fonction de la diversité des systèmes), de la portabilité et/ou de la compatibilité Cloud (Data Lineage).</p>				

Compétences professionnelles

Bloc de compétences 3 – Construction d'une plateforme Big Data permettant la collecte, l'assemblage, le traitement et le stockage des données générées par les systèmes d'une entreprise.

Mise en Situation Professionnelle Reconstituée : ☐ Validée
☐ Non Validée

Le Jury valide la MSPR sur la base des compétences évaluées et éventuellement des savoirs-être du candidat.

Réalisée dans le cadre d'une évaluation en lien avec les compétences de la certification professionnelle visée : Expert en ingénierie des données- EID- RNCP N°36921

RÉALISÉE DANS LE CADRE D'UN :

- ☐ Parcours continu de formation
☐ Parcours discontinu de formation

COMMENTAIRES/REMARQUES OBLIGATOIRE

Date : / /

Signatures des Jurys :