

Data Engineer - 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
Bloc n°1 Collecte et stockage des données			
<p>A1.1 Constitution d'un inventaire des données de l'organisation</p> <p>A1.2 Définition d'une stratégie de collecte et de stockage</p> <p>A1.3 Industrialisation et automatisation de la collecte et du stockage</p>	<p>C1.1 Animer un atelier d'expression de besoins, (questionner les besoins métier et évaluer les besoins futurs) afin de maquetter une base de données relationnelle</p> <p>C1.2 Modéliser et initialiser une base de données relationnelle SQL à partir d'une maquette</p> <p>C1.3 Stocker des données sur un système d'hébergement de données en ligne décentralisé (plateforme cloud), en utilisant les paramètres de sécurité suivant les besoins</p> <p>C1.4 Automatiser la collecte et</p>	<p><u>Mise en situation professionnelle :</u> Mise en place d'une collecte de données et de son stockage_</p> <p><u>Cadre de mise en œuvre et de déroulement de l'évaluation</u> Les candidats sont évalués lors de la réalisation d'une mise en situation au cours de laquelle ils doivent comprendre le besoin métier, créer un lac de données, modéliser une base de données cohérente, et mettre en place un flux de données automatique d'alimentation.</p> <p>Evaluation individuelle des candidats par un examinateur</p>	<ul style="list-style-type: none"> - Les besoins métiers sont correctement interprétés et décrits par le candidat au travers d'un diagramme - Une base de données relationnelle est initialisée, ainsi que les tables présentes sur le diagramme - Les fichiers demandés sont correctement téléchargés depuis et vers le système d'hébergement cloud, avec un accès restreint aux paramètres de sécurité demandés

	<p>l'ingestion de données avec des technologies standard</p> <p>C1.5 Optimiser les usages d'une collecte de données avec un service cloud et mettre en place des procédures de contrôle automatique afin de limiter le risque de surcoûts</p> <p>C1.6 Mettre en place un lac de données</p> <p>C1.7 Contrôler la conformité de la collecte de données dans le cadre du RGPD, et notamment l'emplacement géographique des serveurs</p>		<ul style="list-style-type: none"> - Une collecte de données est automatisée en respect des spécifications - Une collecte de données est automatisée depuis un service cloud en respectant les principes de sécurité - Un lac de donnée est initialisé et alimenté (par exemple avec Hadoop) - La conformité RGPD est détaillée dans la documentation
--	---	--	---

Bloc n°2 Préparation et traitement des données

<p>A2.1 Centralisation des définitions des données de l'organisation</p>	<p>C2.1 Lister et organiser de la manière la plus exhaustive possible les traitements de données à automatiser</p>	<p><u>Mise en situation professionnelle</u> Mise en place d'un traitement automatique de données</p>	<ul style="list-style-type: none"> - La liste des traitements est présente et cohérente par rapport au besoin
<p>A2.2 Mise en place d'un processus de transformation automatisée des données</p>	<p>C2.2 Centraliser les référentiels et les définitions des données en partenariat avec les équipes métier (Data Management et Data gouvernance)</p>	<p><u>Cadre de mise en œuvre et de déroulement de l'évaluation</u> Les candidats sont évalués lors de la réalisation d'une mise en situation au cours de laquelle ils doivent</p>	<ul style="list-style-type: none"> - Les bonnes pratiques de data gouvernance sont appliquées afin d'aligner toutes les définitions métier dans un référentiel
<p>A2.3 Supervision du processus de transformation et des risques associés</p>	<p>C2.3 Etablir une spécification technique d'un processus d'extraction, de transformation et de chargement des données (ETL)</p>	<p>comprendre le besoin métier, établir la spécification technique, mettre en place les traitements et en contrôler le bon fonctionnement.</p>	<ul style="list-style-type: none"> - La spécification est complète et rigoureusement établie. Des diagrammes sont présents pour clarifier les flux de données.
	<p>C2.4 Transformer les données en respect de la spécification avec une architecture décentralisée (par exemple avec Apache Spark)</p>	<p>Evaluation individuelle des candidats par un examinateur</p>	<ul style="list-style-type: none"> - La transformation de données est implémentée, et correspond à la spécification technique.
	<p>C2.5 Utiliser un service en ligne décentralisé (cloud) pour automatiser la transformation de données et mettre en place des procédures de contrôle automatique afin de limiter le risque de surcoûts</p>		<ul style="list-style-type: none"> - La transformation de données est déployée sur un service cloud, et correspond à la spécification technique.

	<p>C2.6 Superviser le processus de transformation de données, et mettre en place des alertes pour se prémunir du risque de sécurité</p> <p>C2.7 Anonymiser, pseudonymiser et agréger les données personnelles afin de respecter le cadre du RGPD</p>		<ul style="list-style-type: none"> - Des alertes sont présentes tout au long du process de transformation de données. L'alerte est activée notamment si le format de données initial est incorrect. - Le cadre du RGPD est respecté (anonymisation, pseudonymisation)
--	--	--	---

Bloc n°3 Mise à disposition et accessibilité			
<p>A3.1 Déploiement et mise à disposition des données aux utilisateurs finaux</p> <p>A3.2 Déploiement et mise à disposition des données aux autres équipes techniques</p> <p>A3.3 Communication et accompagnement des équipes métier</p>	<p>C3.1 Rendre disponible un entrepôt de données en respectant les règles de sécurité</p> <p>C3.2 Analyser les données très volumineuses ("big data") grâce au calcul distribué</p> <p>C3.3 Réaliser des tableaux de bords alimentés automatiquement grâce à un service cloud</p>	<p><u>Mise en situation professionnelle</u> Mise à disposition de données de manière sécurisée</p> <p><u>Descriptif de l'évaluation</u> Les candidats sont évalués lors de la réalisation d'une mise en situation au cours de laquelle ils doivent rendre accessible un tableau de bord connecté à une base de données, de manière sécurisée.</p> <p>Evaluation individuelle des candidats par un examinateur</p>	<ul style="list-style-type: none"> - L'entrepôt de données est accessible, avec les identifiants sécurisés fournis par le candidat. Les données sont inaccessibles sans ces identifiants. - Un calcul de type "Map-Reduce" est implémenté, et permet un calcul distribué sur plusieurs machines. L'analyse se base sur le résultat de ce calcul - Un tableau de bord est disponible, respectant les bonnes pratiques de visualisation de données. Les

	<p>C3.4 Rendre accessible un tableau de bord en prenant en compte les handicaps visuels</p> <p>C3.5 Déployer une interface de programmation (type API REST) accessibles pour les usages des développeurs</p> <p>C3.6 Présenter, vulgariser et expliquer une architecture de données afin de rendre autonomes les utilisateurs</p> <p>C3.7 Sécuriser l'accès aux données et aux tableaux de bord, et mettre en place un suivi (journal, log) des connexions</p>		<p>informations sont automatiquement mises à jour sans nécessité d'intervention humaine.</p> <ul style="list-style-type: none"> - Le tableau de bord prend en compte les principaux handicaps visuels (notamment le code couleur, et les explications sous les graphiques permettant une lecture vocale automatisée). - Une API REST est déployée, et permet un accès aux données pour le développement de programmes informatiques basés sur d'autres langages. - La documentation fournie est claire et concise. L'explication orale est correcte et bien vulgarisée. - L'accès aux données est sécurisé et un journal des connexions est correctement alimenté
--	--	--	---