

Référentiels d'activités, de compétences et d'évaluation

Certification professionnelle de niveau 7

Data engineer

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. »

La certification à finalité professionnelle « Data Engineer » est constituée de quatre (4) blocs de compétences. Chaque bloc de compétence est valable à vie. Le candidat à la certification a le choix de son parcours, il n'y a pas d'ordre de passage de blocs de compétences défini.

Chaque bloc de compétence est validé par une ou plusieurs modalités d'évaluation adaptées aux référentiels de compétences et d'activités.

Pour obtenir le titre à finalité professionnelle le candidat doit valider les quatre (4) blocs de compétences constituant la certification.

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>	RÉFÉRENTIEL D'ÉVALUATION <i>Définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITES D'EVALUATION	CRITÈRES D'ÉVALUATION
Bloc de compétences 1 : Piloter la conduite d'un projet data			
<u>A1. Encadrement de la conception d'un projet data au sein d'une organisation</u> - Analyse du besoin exprimé, selon le formalisme de l'acheteur : DCE, brief interne, clauses d'un contrat privé, etc.	C1. Analyser l'expression d'un besoin de projet data dans une étude de faisabilité en explorant, à l'aune des enjeux stratégiques de l'organisation, le besoin métier avec les parties prenantes pour valider les orientations et sélectionner les	E1. Etude de cas. (C1) L'évaluation doit se faire dans le cadre d'une étude de cas réelle ou fictive sur la base de l'expression d'un besoin data, de l'organigramme de l'organisation et des activités métiers associées.	Les grilles d'entretiens questionnent les activités métiers impliquées dans le projet data Les grilles d'entretiens questionnent la caractérisation des données impliquées, des métadonnées, des accès, des

<ul style="list-style-type: none"> - Analyse du besoin face aux enjeux relatifs à l'éco-responsabilité : analyse du cycle de vie simplifié - Questionnement de l'expression de besoin, et notamment sur les aspects : objectifs du projet, périmètre fonctionnel du projet, moyens humains, techniques et financiers mobilisables. - Identification de tous les métiers concernés par le périmètre du projet data suivant : sources et génération des données, collecte et stockage des données, exploitation et analyse des données. - Audition des usagers potentiels du projet data pour compléter l'expression du besoin sur les aspects précis d'analyse et d'exploitation des données du projet data - Anticipation de l'effort nécessaire et des éventuels freins à lever pour la mise en conformité du projet avec le RGPD, au regard de l'expression de besoin. - Inventaire des éventuels besoins spécifiques auprès des futurs usagers du projet data. - Anticipation de l'effort d'accessibilité à mettre en oeuvre tout au long du projet 	<p>hypothèses techniques du projet avec le ou les commanditaire(s).</p>	<p>Le but de cette étude de cas est de donner à voir à travers les outils que sont les grilles d'entretien, la démarche générale d'analyse de besoin menée par le candidat</p> <p><u>Livrables</u> : les grilles d'entretien</p> <p><u>Évaluation</u> : Correction des grilles et Présentation orale individuelle</p> <p>E2. Mise en situation (C1, C2, C3, C4, C6)</p> <p>L'évaluation doit se faire dans un contexte de projet data réel ou fictif d'une organisation et des éléments de preuves de la réalisation du projet.</p> <p>La mise en situation a pour but de confronter le ou la candidat/e à un besoin data rencontré par une organisation à laquelle il ou elle devra répondre dans son intégralité. Selon la situation choisie, cette évaluation pourra intégrer les éléments d'infrastructure technique associés (fusion alors avec E6) ou se limiter à la dimension de pilotage du projet.</p> <p><u>Livrable</u> : rapport professionnel individuel</p> <p><u>Évaluation</u> :</p>	<p>stockages et des traitements appliqués.</p> <p>La note de synthèse rend compte du questionnement et de l'analyse du besoin, du périmètre fonctionnel du projet, des moyens disponibles, de la faisabilité et de la gouvernance de la donnée liée au projet.</p> <p>La note de synthèse est organisée, par exemple selon le plan suivant :</p> <ol style="list-style-type: none"> 1. intro : rappel du contexte, enjeux du projet, reformulation de l'expression de besoin initiale, annonce du plan 2. objectifs et périmètre fonctionnel du projet 3. étude d'opportunités : <ol style="list-style-type: none"> a. synthèse des informations issues des entretiens métier au regard des objectifs du projet : opportunités ? Contraintes ? b. benchmark des solutions et projets existants dans le périmètre fonctionnel visé 4. étude de la faisabilité : rapport objectifs, qualité, coûts, délais, moyens mobilisables. 5. conclusions : analyse RICE*
---	---	--	---

<p>- Conclure sur une recommandation de cadrage du projet data.</p> <p>- Communiquer le cadrage du projet data au(x) commanditaire(s)</p>		<p>- Correction du rapport professionnel - Soutenance orale individuelle</p> <p>E3. Jeu de rôle “lancement d’un projet data” (C5, C6, C7)</p> <p>Le/la candidat(e) est mis en situation de documenter et d’animer l’introduction de la réunion de lancement d’un projet data réel ou fictif.</p> <p>Le jeu de rôle a pour but de mettre le candidat dans la posture de chef de projet lors d’une étape clé de cette activité. Cette simulation permettra au candidat ou à la candidate de présenter les documents ressources pour la coordination de projet, leurs usages, ainsi que de montrer au jury la mise en oeuvre du travail de collaboration et de communication (interne et externe).</p> <p><u>Livrables</u> : Le support de la présentation et les documents associés (par exemple : avant-projet, feuille de route, calendrier, stratégie de communication)</p> <p><u>Évaluation</u> : - correction des livrables - jeu de rôle. Simulation de l’introduction de la réunion de lancement.</p>	<p>Les objectifs du projet sont rédigés selon la méthode SMART*</p> <p>Le cadrage projet reprend et complète la note de synthèse : hypothèses et préconisations macro de solutions techniques, évaluation des aménagements à implémenter pour l’accessibilité du projet dans sa réalisation et dans l’utilisation future de ses produits, les actions pour la mise en conformité avec le RGPD.</p> <p>L’effort d’accessibilité tout au long du projet est anticipé dans l’avant-projet, concernant :</p> <ul style="list-style-type: none"> - la réalisation du projet : adaptation des postes de travail de l’équipe technique, - les utilisateurs finaux : adaptation des interfaces homme-machine des outils et des supports techniques (documentations, communications...). <p>L’avant-projet liste les actions techniques et non-techniques à mener tout au long du projet pour assurer la conformité du projet avec le RGPD.</p>
<p>- Inventaire des sources (applications, IoT, ventes, etc.) et des domaines de données utiles au projet data.</p>	<p>C2. Cartographier les données disponibles en référençant les usages, les sources, les métadonnées et les</p>		<p>La topographie des données est complète, respectant la structuration suivante en quatre parties :</p>

<ul style="list-style-type: none"> - Inventaire des gestionnaires ou responsables des différents domaines (clients, produits, fournisseurs. - Inventaire des données et métadonnées nécessaires au projet et déjà disponibles, et de leurs structures. - Inventaire des modalités d'accès, d'autorisation ou encore d'interconnexion logiciel aux sources de données identifiées. - Identification des données manquantes pour le projet data. - Formalisation d'une topographie des données 	<p>données, afin de valider les hypothèses techniques du projet data</p>		<ul style="list-style-type: none"> - La sémantique : les métadonnées des données et les objets métier propres à l'organisation dans un glossaire métier. - Les modèles de données : la façon dont les données sont modélisées et stockées dans les différents systèmes (structurées, semi structurées ou non structurées). - les traitements et les flux de données : les informations importantes sur les méthodes de transformation, de manipulation et de traitement des données à travers les différents SI de l'organisation. - la mise à disposition, les accès et les conditions d'utilisation des données.
<ul style="list-style-type: none"> - Qualification des flux de données : fréquence et volume des flux. - Qualification des composants du système d'information existants impliqués dans le cadre technique d'exploitation des données envisagé. - Définition des processus de mise en conformité avec le RGPD. 	<p>C3. Concevoir un cadre technique d'exploitation des données en analysant les contraintes techniques, de moyens et la cartographie des données afin de définir une réponse technique adaptée aux ressources mobilisables dans le respect du RGPD et d'une démarche éco-responsable.</p>		<p>L'étude technique rend compte de l'ensemble des éléments suivants :</p> <ul style="list-style-type: none"> - L'analyse fonctionnelle : Que fait le système décrit ? Quelles sont les contraintes métiers qui auront un impact sur l'architecture ? - Les besoins non-fonctionnels (outils et contraintes techniques) - La représentation fonctionnelle - La représentation applicative

<ul style="list-style-type: none"> - Définition de la stratégie d'éco-responsabilité sur l'activité data - Identification des moyens humains, financiers et logistiques disponibles. - Choix d'une (ou des) orientation(s) technique(s) pour la collecte, le stockage, le traitement et la mise à disposition des données, - Rédaction des spécifications techniques d'architecture 			<ul style="list-style-type: none"> - La représentation d'infrastructure - La représentation opérationnelle - Les décisions d'architecture : documentation des choix techniques effectués - Les processus de mise en conformité RGPD (registre des données personnelles, traitements des tris et suppression, etc.) - La stratégie d'éco-responsabilité selon le référentiel général d'éco-conception des services numériques du gouvernement¹ - Les risques et les coûts <p>Les choix techniques d'architecture prennent en compte et respectent les objectifs du projet data les moyens mobilisables.</p> <p>L'accessibilité des livrables du projet est anticipée et des solutions d'adaptation de poste de travail sont proposées quand cela est nécessaire.</p> <p>La représentation applicative contient une matrice des flux.</p> <p>La matrice des flux permet la qualification des flux de données</p>
---	--	--	---

¹ <https://ecoresponsable.numerique.gouv.fr/publications/referentiel-general-ecoconception/>

			<p>concernées par le projet.</p> <p>Les choix dans l'étude technique d'architecture s'appuient sur les ressources techniques et matériels du SI directement mobilisables pour le projet.</p> <p>Le choix des éventuels prestataires techniques intègre le paramètre d'adoption d'une démarche d'éco-responsabilité chez le ou les prestataires,</p> <p>Le choix des outils et services techniques concourent au respect de la stratégie d'éco-conception du projet</p> <p>L'étude techniques d'architectures définit et illustre les processus nécessaires à implémenter lors de la réalisation du projet pour la conformité avec le RGPD</p> <p>Les processus de conformité définis par l'étude technique d'architecture sont conformes au RGPD</p>
<ul style="list-style-type: none"> - Définition de la ou des thématiques de veille. - Planification des temps dédiés à la veille. - Choix d'un outil d'agrégation des flux d'informations et d'actualités. 	<p>C4. Réaliser une veille technique et réglementaire en sélectionnant des sources et en collectant et traitant les informations collectées afin de formuler des recommandations projet toujours en phase avec l'état de l'art.</p>		<p>La thématique de veille choisie porte sur un outil et/ou une réglementation mobilisée dans la mise en situation.</p> <p>Les temps de veille sont planifiés régulièrement (à minima une récurrence d'une heure hebdomadaire)</p>

<ul style="list-style-type: none"> - Choix d'un outil de partage ou communication des synthèses des informations collectées. - Identification des sources et des flux d'informations utiles à la veille thématique visée. - Qualification de la fiabilité des sources et des flux identifiés. - Configuration des outils d'agrégation selon les flux et sources et la ou les thématiques de veille. - Rédaction des synthèses des informations collectées. - Communications des synthèses aux parties prenantes du projet et aux éventuels intéressés (à intervalles réguliers ou à l'opportunité par exemple). 			<p>le choix des l'outils d'agrégation est cohérent avec les sources d'informations visées et le budget disponible (flux RSS, flux réseaux sociaux, agrégation newsletter, etc)</p> <p>Les synthèses sont communiqués aux parties prenantes dans un format qui respecte les recommandations d'accessibilité (par exemples celles de l'association Valentin Haüy²ou de Atalan - AcceDe³).</p> <p>Les informations partagées dans la synthèse répondent à la thématique de veille choisie.</p> <p>Les sources et flux identifiés répondent aux critères de fiabilité :</p> <ul style="list-style-type: none"> - L'auteur de la page est identifié - Des informations sur l'auteur sont disponibles et confirment ses compétences, sa notoriété et l'absence d'intérêts personnels - l'analyse du contenu est valable (date de publication récente, sources de l'information indiquées, niveau de langue correct), - la source (site) ou le document
---	--	--	--

² <https://www.avh.asso.fr/fr/favoriser-laccessibilite/accessibilite-numerique/accessibilite-des-documents-et-des-courriels-0>

³ <https://www.accede-web.com/wp-content/uploads/accessibility-guidelines-editors-template.pdf>

			<p>est structuré</p> <ul style="list-style-type: none"> - les sources (sites) ou documents respectant les normes d'accessibilités sont privilégiés. - l'information peut être confirmée par d'autres sites de confiance
<p><u>A2 Coordination de la réalisation d'un projet data</u></p> <ul style="list-style-type: none"> - Composition de l'équipe projet - Identification des grandes étapes de la <i>feuille de route</i>* du projet (par exemple : Roadmap, PERT et chemin critique, etc.) - Affectation du budget - Association d'un ou plusieurs responsables pour chaque grande étape de la feuille de route - Identification des livrables attendues pour chaque grande étape de la feuille de route - Identification des besoins de validation externes éventuels à chaque grande étape de la feuille de route (recette par le client par exemple). - Evaluation des risques pour chaque grandes étapes de la feuille de route du 	<p>C5. Planifier la réalisation d'un projet data en attribuant les moyens nécessaires et en définissant les étapes de réalisation et les méthodes de suivi du projet afin d'organiser sa mise en oeuvre</p>		<ul style="list-style-type: none"> - La préconisation de composition d'équipe couvre en compétences les besoins nécessaires à la réalisation du projet. <p>L'équipe projet est composée selon les compétences nécessaires à la réalisation du projet et prend en compte le budget affecté</p> <p>les moyens financiers alloués correspondent au budget prévisionnel du projet</p> <p>La feuille de route est découpée en grandes étapes de réalisation du projet</p> <p>Les grandes étapes de la feuille de route respectent les ensembles fonctionnels du projet</p> <p>Le calendrier rend compte</p> <ul style="list-style-type: none"> - des tâches et livrables attendus et dates d'échéance associées - de l'attribution des ressources - de la pondération des efforts

<p>projet : besoins spécifiques, difficultés organisationnelles ou techniques, etc.</p> <ul style="list-style-type: none"> - Définition des indicateurs d'avancement - Configuration des outils de suivi (par exemple : <i>Scrum burndown chart*</i>, diagram de Gantt*, etc.) - Découpage de chaque grande étape de la feuille de route en tâches (par exemple : Scrum backlog, diagram de Gantt, etc.) - Évaluation avec les membres de l'équipe technique de la durée ou de la difficulté pour chaque tâche (poker planning, date au plus tôt et au plus tard de Gantt, etc.) - Assignation ou organisation de l'assignation des tâches entre les membres de l'équipe technique - Planification de la réalisation des tâches (Scrum board, Kanban, etc.) - Communication des éléments de planification (feuille de route et planification de la production) 			<p>nécessaires à la réalisation des tâches</p> <ul style="list-style-type: none"> - des rituels d'animation du travail collaboratif de l'équipe projet <p>Le suivi du calendrier permet d'atteindre les objectifs du projet en respectant les contraintes.</p> <p>L'attribution des ressources est cohérente avec la répartition des compétences au sein des membres de l'équipe</p> <p>La pondération est réalisée selon une méthode choisie et partagée avec l'équipe (poker planning, méthode de l'unité équivalente, etc)</p> <p>Le paramétrage des outils de suivi est cohérent avec les délais et les attributions de missions du planning.</p> <p>Les outils de suivi intègrent les indicateurs de suivi.</p> <p>Les éléments de planification sont communiqués à l'équipe dans un format qui respecte les recommandations d'accessibilité (par exemples celles de l'association Valentin Haüy⁴ ou de Atalan - AcceDe⁵)</p>
---	--	--	---

⁴ <https://www.avh.asso.fr/fr/favoriser-laccessibilite/accessibilite-numerique/accessibilite-des-documents-et-des-courriels-0>

⁵ <https://www.accede-web.com/wp-content/uploads/accessibility-guidelines-editors-template.pdf>

			<p>L'enchaînement des tâches permet la réalisation de chacune d'entre elles.</p>
<ul style="list-style-type: none"> - Animation des échanges entre les membres de l'équipe technique en charge de la réalisation du projet, - mise en place des outils et des rituels nécessaires au suivi de l'avancement dans les réalisations des tâches, au regard du planning prévisionnel, - arbitrage et résolution des problèmes bloquant tout au long de la réalisation du projet. - Suivi des indicateurs d'avancement - Suivi du budget - encadrement des prestataires externes 	<p>C6. Superviser la réalisation d'un projet data en organisant les méthodes, les outils de travail et la communication entre les parties prenantes, afin d'accompagner les membres de l'équipes dans la réussite du projet</p>		<p>L'animation des échanges est adaptée à l'information à transmettre : les supports et les modalités d'animation sont adaptés aux besoins de communication.</p> <p>Toutes les personnes concernées sont impliquées dans les échanges.</p> <p>Le contenu et le discours sont adaptés au public et au contexte des échanges.</p> <p>Les outils de suivi sont configurés et accessibles à toutes les parties prenantes.</p> <p>Les outils de suivi respectant les normes d'accessibilité (par exemple : RGAA) sont privilégiés.</p> <p>Les rituels sont documentés : règles de participation et d'organisation aux rituels, planification, etc.</p> <p>Les indicateurs de suivi sont mis à jour tout au long du projet, lors des rituels de suivi de l'avancement du projet.</p> <p>Les dépenses sont imputées au budget prévisionnel tout au long du projet.</p>

			<p>Les commandes communiqués aux prestataires externes couvrent les objectifs, la méthodologies de travail, les outils à utiliser, les livrables et les critères d'acceptance.</p>
<ul style="list-style-type: none"> - Planification des étapes de communication du projet à toutes les parties prenantes. - Réalisation du ou des supports de communication - Communication à chaque étape du plan de communication, selon le ou les support réalisés - Encadrement de la production de la documentation utilisateur des livrables techniques du projet - Organisation des temps d'accompagnement à la prise en main des livrables par les utilisateurs finaux. - Justification des orientations choisies tout au long du projet : technique, planification, etc. - Recueil des retours des parties prenantes tout au long de la 	<p>C7. Communiquer tout au long de la réalisation du projet data sur les orientations, les réalisations et leurs impacts en élaborant la stratégie et les supports de communication afin d'informer toutes les parties prenantes des évolutions ou des opportunités internes comme externes, portés par le projet</p>		<p>Toutes les étapes de communication du projet sont planifiées : au lancement, à chaque jalon de la feuille de route, pour les démonstrations, à la livraison du projet.</p> <p>Les supports de communication sont accessibles à toutes les parties prenantes.</p> <p>Les supports de communications respectent les préconisations de mise en page sur des critères d'accessibilités (par exemples celles de l'association Valentin Haüy⁶ ou de Atalan - AcceDe⁷).</p> <p>Toutes les personnes concernées sont impliquées dans les échanges.</p> <p>Le contenu et le discours sont adaptés au public et au contexte des échanges.</p> <p>Les communications présentent les orientations choisies et arbitrages menés pour la réalisation du projet.</p>

⁶ <https://www.avh.asso.fr/fr/favoriser-laccessibilite/accessibilite-numerique/accessibilite-des-documents-et-des-courriels-0>

⁷ <https://www.accede-web.com/wp-content/uploads/accessibility-guidelines-editors-template.pdf>

<p>communication sur le projet et sur son avancement</p>			<p>Les tâches de production de la documentation utilisateurs sont planifiées et réparties entre les membres de l'équipe.</p> <p>Les temps d'accompagnement des utilisateurs finaux sont planifiés.</p> <p>Les missions des utilisateurs invités aux temps d'accompagnement sont cohérentes avec la prise en main du livrable visée.</p> <p>Le recueil des retours des parties prenantes et leur traitement suit un processus qui est intégré à la stratégie de communication du projet.</p>
<p>Bloc de compétences 2 : Réaliser la collecte, le stockage et la mise à disposition des données d'un projet data</p>			
<p><u>A3 : Programmation de la collecte de données depuis plusieurs sources pour un projet data</u></p> <ul style="list-style-type: none"> - Identification des contraintes techniques propres aux sources de données (consulter la documentation, les règles de confidentialité, etc.). - Rédaction des spécifications techniques pour l'extraction des données. 	<p>C8. Automatiser l'extraction de données depuis un service web, une page web (scraping), un fichier de données, une base de données et un système big data en programmant le script adapté afin de pérenniser la collecte des données nécessaires au projet.</p>	<p>E4. Mise en situation (C8, C9, C10, C11, C12)</p> <p>L'évaluation doit se faire dans un contexte de réalisation d'un service numérique réel ou fictif basé sur l'usage de données, à partir du cadrage pour la réalisation d'un service numérique (spécifications fonctionnelles et techniques par exemple).</p> <p>Le projet évalué a pour but d'optimiser,</p>	<p>La présentation du projet et de son contexte est complète : acteurs, objectifs fonctionnels et techniques, environnements et contraintes techniques, budget, organisation du travail et planification.</p> <p>Les spécifications techniques précisent : les technologies et outils, les services externes, les exigences de</p>

<ul style="list-style-type: none"> - Construction des requêtes HTTP pour la récupération des données depuis un service web (REST). - Lecture d'un fichier de données dans un script (Python, R, etc.). - Téléchargement de l'HTML d'une ou plusieurs pages web visées par une action de <i>scraping</i>. - Connexion programmatique (Python, R, etc.) à un système de gestion de base de données et à un système big data (Hive, Apache Impala, etc.). - Programmation des filtres/parsing des données utiles dans les résultats obtenus depuis : <ul style="list-style-type: none"> - une API, - depuis la lecture d'un ou de plusieurs fichiers de données, - et depuis l'HTML collecté d'un site web (<i>scraping</i>). - Exécution programmatique des requêtes d'extraction de type SQL. - Exécution programmatique des requêtes d'extraction depuis un système big data. 		<p>d'automatiser, de pérenniser et de mettre à disposition les flux de données et les données, utiles et nécessaires à la réalisation du service numérique, par les équipes techniques (par exemple en analyse statistique, en business intelligence, en machine learning ou encore en intelligence artificielle).</p> <p><u>Livrable</u> : rapport professionnel individuel</p> <p><u>Évaluation</u> :</p> <ul style="list-style-type: none"> - Correction du rapport professionnel - Soutenance orale individuelle 	<p>programmation (langages), l'accessibilité (disponibilité, accès).</p> <p>Le périmètre des spécifications techniques est complet : il couvre l'ensemble des moyens techniques à mettre en œuvre pour l'extraction et l'agrégation des données en un jeu de données brutes final.</p> <p>Le script d'extraction des données est fonctionnel : toutes les données visées sont effectivement récupérées à l'issue de l'exécution du script.</p> <p>Le script comprend un point de lancement, l'initialisation des dépendances et des connexions externes, les règles logiques de traitement, la gestion des erreurs et des exceptions, la fin du traitement et la sauvegarde des résultats.</p> <p>Le script d'extraction des données est versionné* et accessible depuis un dépôt Git*.</p> <p>L'extraction des données est faite depuis un mix entre au moins les sources suivantes : un service web (API REST), un fichier de données, un scraping, une base de données et un système big data.</p>
--	--	--	---

<p>- Ecriture des requêtes d'extraction de type SQL de récupération de données</p>	<p>C9. Développer des requêtes de type SQL* d'extraction des données depuis un</p>		<p>Les requêtes de type SQL pour la collecte de données sont fonctionnelles : les</p>

<p>stockées en base de données et depuis un système big data (Hive, Spark, etc).</p> <ul style="list-style-type: none"> - Documentation des requêtes d'extraction. 	<p>système de gestion de base de données et un système big data en appliquant le langage de requête propre au système afin de préparer la collecte des données nécessaires au projet.</p>		<p>données visées sont effectivement extraites suites à l'exécution des requêtes.</p> <p>La documentation des requêtes met en lumière choix de sélections, filtrages, conditions, jointures, etc., en fonction des objectifs de collecte.</p> <p>La documentation explicite les optimisations appliquées aux requêtes .</p>
<ul style="list-style-type: none"> - Rédaction des spécifications techniques pour l'agrégation des données. - Programmation des règles d'agrégation des données collectées depuis chaque source en un jeu de données brutes unique. - Programmation de l'identification des entrées corrompues dans le jeu de données (par exemple, données partielles et/ou manquantes). - Programmation de la suppression des entrées corrompues. - Programmation de l'identification des entrées au format non normalisé. - Programmation de la l'homogénéisation des formats de données (par exemple, format des dates, des unités, etc.). 	<p>C10. Développer des règles d'agrégation de données issues de différentes sources en programmant, sous forme de script, la suppression des entrées corrompues et en programmant l'homogénéisation des formats des données afin de préparer le stockage du jeu de données final.</p>		<p>Le script d'agrégation des données est fonctionnel : les données sont effectivement agrégées, nettoyées et normalisées en un seul jeu de données à l'issue de l'exécution du script.</p> <p>Le script d'agrégation des données est versionné et accessible depuis un dépôt Git.</p> <p>La documentation du script d'agrégation est complète : dépendances, commandes, les enchaînements logiques de l'algorithme, les choix de nettoyage et d'homogénéisation des formats données.</p>

<ul style="list-style-type: none"> - Versionnement* des scripts avec Git et un dépôt Git en ligne. - Documentation des scripts. 			
<p><u>A4 : Développement de la mise à disposition technique des données collectées pour un projet data</u></p> <ul style="list-style-type: none"> - Rédaction des spécifications techniques pour le stockage des données. - Modélisation de la structure des données de la base de données selon la méthode MERISE*. - Choix du système de gestion de base de données. - Création de la base de données dans le système de gestion de base de données. - Documentation de la procédure d'installation du système de gestion de base de données. - Rédaction ou mise à jour du registre des traitements de données personnelles en vue de la mise en conformité de la base de données avec le RGPD. - Rédaction des procédures de tri des données personnelles stockées dans la 	<p>C11. Créer une base de données dans le respect du RGPD en élaborant les modèles conceptuels et physiques des données à partir des données préparées et en programmant leur import afin de stocker le jeu de données du projet.</p>		<p>Les modélisations des données respectent la méthode et le formalisme MERISE.</p> <p>Le modèle physique des données est fonctionnel : il est intégré avec succès lors de la création de la base de données, sans erreur.</p> <p>La base de données est choisie au regard de la modélisation des données et des contraintes du projet.</p> <p>La reproduction des procédures d'installation décrites (base de données et API) a pour résultat un système conforme aux objets techniques attendus..</p> <p>Le script d'import fourni est fonctionnel : il permet l'insertion des données dans le système mis en place.</p> <p>La documentation technique du script d'import est versionné à la racine du même dépôt Git que celui utilisé pour le script d'import.</p>

<p>base de données pour la mise en conformité avec le RGPD (par exemple : détection et suppression des données personnelles inutiles, trop anciennes, etc.).</p> <ul style="list-style-type: none"> - Programmation du script d'import des données en base de données. - Documentation du script d'import. 			<p>Les documentations techniques des script couvrent les parties suivantes :</p> <ul style="list-style-type: none"> - les dépendances nécessaires pour la réutilisation des scripts (langages, dépendances externes, etc) - les commandes pour l'exécution des scripts. <p>Le registre des traitements de données personnelles intègre l'ensemble des traitements de données personnelles impliqués dans la base de données.</p> <p>Les procédures de tri des données personnelles pour la mise en conformité de la base de données avec le RGPD sont rédigées.</p> <p>Les procédures de tri détaillent les traitements de conformité (automatisés ou non) à appliquer ainsi que leur fréquence d'exécution.</p>
<ul style="list-style-type: none"> - Rédaction des spécifications techniques des moyens de mise à disposition et d'accès aux données du projet : API (REST) et accès direct à la base de données. - Développement des points de terminaison de l'API REST. - Développement des règles d'autorisation et d'accès aux points de terminaison de 	<p>C12. Partager le jeu de données en configurant des interfaces logicielles et en créant des interfaces programmables afin de mettre à disposition le jeu de données pour le développement du projet.</p>		<p>La documentation technique de l'API (REST) couvre tous les points de terminaisons (<i>end points*</i>)</p> <p>La documentation technique couvre les règles d'authentification et/ou d'autorisation de l'API</p> <p>La documentation technique respecte les standards du modèle choisi (par exemple <i>Open API*</i>)</p>

<p>l'API REST.</p> <ul style="list-style-type: none"> - Rédaction de la documentation technique de l'API REST. - Configuration des accès à la base de données. - Rédaction de la documentation technique d'accès à la base de données adaptée aux environnements techniques du projet. 			<p>L'API REST est fonctionnelle pour l'accès aux données du projet : elle restreint par une autorisation (ou authentification) l'accès aux données,</p> <p>L'API REST est fonctionnelle pour la mise à disposition : elle permet la récupération de l'ensemble des données nécessaires au projet.</p>
<p>Bloc de compétences 3 : Élaborer et maintenir un entrepôt de données* (data warehouse)</p>			
<p><u>A5 : Mise en place de l'entrepôt de données</u></p> <ul style="list-style-type: none"> - Explicitation des données attendues selon le besoin d'analyse. - Production des schémas des structures de données en faits* et dimensions* selon les besoins d'analyse. - Définition des relations entre les faits et les dimensions dans les schémas de conception. 	<p>C13. Modéliser la structure des données d'un entrepôt de données en s'appuyant sur les dimensions* et les faits* afin d'optimiser l'organisation des données pour les requêtes analytiques.</p>	<p>E5. Mise en situation (C13, C14, C15)</p> <p>L'évaluation doit se faire dans un contexte de réalisation d'un projet fictif proposé par l'équipe pédagogique ou d'un projet professionnel réalisé en poste. Le projet évalué s'appuie sur le cadre technique de l'organisation et sur le cadre d'exploitation des données.</p> <p>Le projet a pour but de couvrir toutes les étapes de mise en place d'un entrepôt de données, de sa modélisation à son usage fonctionnel (réponse au besoin d'analyse)</p>	<p>Les données nécessaires aux analyses sont listées.</p> <p>La liste des données nécessaires aux analyses est exhaustive.</p> <p>Les modélisations logiques et les modélisations physiques sont explicités sans erreur d'interprétation.</p> <p>Les modélisations appliquent les pratiques de modélisation d'entrepôt de données : en flocon, en étoile, en constellation.</p> <p>L'approche - <i>top-down*</i> ou <i>bottom-up*</i> de</p>

		<p><u>livrable</u> : rapport professionnel individuel</p> <p><u>Évaluation</u> :</p> <ul style="list-style-type: none"> - Correction du rapport professionnel - Soutenance orale individuelle 	<p>création et modélisation de l'entrepôt de données est justifiée en fonction des caractéristiques du projet, par exemple : volume de données, fréquences des mise à jour, nature(s) des analyses...</p>
<ul style="list-style-type: none"> - Configuration des outils. - Configuration des data marts*. - Configuration des accès. - Documentation de l'architecture technique de l'entrepôt de données. - Documentation de la procédure d'installation et de configuration de l'entrepôt de données. - Test du bon fonctionnement de l'entrepôt de données. 	<p>C14. Créer un entrepôt de données à partir des paramètres du projet, des contraintes techniques et matérielles et de la modélisation de la structure des données afin de soutenir l'analyse de l'activité et l'aide à la décision stratégique de l'organisation</p>	<p>E6. Etude de cas (C16, C17)</p> <p>L'évaluation doit se faire dans le cadre d'une situation professionnelle fictive, à partir d'un entrepôt de données en place et d'un besoin d'évolution de celui-ci (technique, évolution dans le schéma des données sources etc.) en environnement de test.</p> <p>Le projet évalué s'appuie sur le cadre technique de l'organisation et sur le cadre d'exploitation des données.</p> <p>Lors de cette étude de cas, le candidat ou la candidate rend compte de sa capacité à maintenir un entrepôt de données en conditions opérationnelles, qu'il s'agisse aussi bien d'évolutions techniques que d'évolutions du besoin d'analyse.</p> <p><u>Livrable</u> : rapport professionnel individuel</p> <p><u>Évaluation</u> :</p> <ul style="list-style-type: none"> - Correction du rapport professionnel - Echanges de questions - réponses individuel 	<p>L'entrepôt de données remplit les fonctionnalités attendues.</p> <p>Les configurations principales appliquées sont explicitées.</p> <p>Les accès aux données opérationnelles sources sont correctement configurés.</p> <p>Les accès à l'entrepôt de données et/ou <i>datamarts</i> pour les équipes analytiques sont correctement configurés.</p> <p>La procédure de test est présentée.</p> <p>La procédure de test couvre l'ensemble du spectre technique et fonctionnel de l'entrepôt de données.</p> <p>La documentation technique détaille l'architecture technique et couvre la procédure d'installation et de configuration de l'entrepôt de données.</p> <p>La documentation respecte une structure permettant d'y rechercher rapidement une information spécifique.</p>

			<p>La documentation respecte les règles d'accessibilités.</p> <p>Un retour d'expérience est fait concernant la pile technique utilisée au regard des besoins d'analyse et du volume de données géré.</p>
<ul style="list-style-type: none"> - Identification des formats et du volume des sources de données, - Intégration des sources, - Identification des formats des zones de sortie, - Configuration des sorties, - Intégration des règles de transformation et de traitement nécessaires ; par exemple : passage de table-relations à faits-dimensions. 	<p>C15. Intégrer les ETL* nécessaires en entrée et en sortie d'un entrepôt de données afin de garantir la qualité et le bon formatage des données en accord avec les modélisations logiques et physiques préalablement établies</p>		<p>Les formats et le volume des données sont connus et expliqués.</p> <p>Les ETL sont alimentés avec les données identifiées.</p> <p>Les formats des zones de sortie sont connus et expliqués.</p> <p>Les données en sortie respectent le format attendu.</p> <p>Les ETL appliquent les traitements nécessaires pour la mise en conformité avec les schémas physiques de données des zones de sortie.</p> <p>Les ETL appliquent les traitements de nettoyage des données utiles et nécessaires à la qualité des jeux de données en sortie : unicité des formats et des unités, détection des doublons, etc...,</p> <p>Le fonctionnement général et les règles de traitement de chacun des ETL sont clairement explicités, sans ambiguïté.</p>

<p><u>Activité 6 : Maintenance et évolution de l'entrepôt de données</u></p> <ul style="list-style-type: none"> - Organisation de l'activité de maintenance de l'entrepôt de données : répartition des tâches et priorisation. - Suivi de l'activité de maintenance : indicateurs de service en fonction des SLA* à l'aide d'un tableau de bord. - Maintenance en condition opérationnelle des outils et composants techniques de l'entrepôt de données - Intégration des évolutions de l'entrepôt de données : ajout datamart, scalabilité, etc. - Gestion des évolutions des accès à l'entrepôt de données et des accès aux données opérationnelles sources. - Rédaction ou mise à jour du registre des traitements de données personnelles en vue de la mise en conformité de l'entrepôt de données avec le RGPD. - Rédaction des procédures de tri des données personnelles stockées dans l'entrepôt de données pour la mise en conformité avec le RGPD (par exemple : détection et suppression des données personnelles inutiles, trop anciennes, 	<p>C16. Gérer l'entrepôt de données à l'aide des outils d'administration et de supervision dans le respect du RGPD, afin de garantir les bons accès, l'intégration des évolutions structurelles et son maintien en condition opérationnelle dans le temps.</p>		<p>Une journalisation de l'activité de l'entrepôt de données est mise en place.</p> <p>La journalisation catégorise à minima les alertes et les erreurs.</p> <p>Un système d'alerte (e-mail, sms, notification..) est mis en place et activé en cas d'erreur notifiée dans les journaux.</p> <p>Les tâches de maintenance sont priorisées selon les objectifs et les exigences de maintenance.</p> <p>Les tâches de maintenance sont assignées entre les membres de l'équipe de maintenance.</p> <p>Les indicateurs de service se base sur les SLA.</p> <p>Le tableau de bord permet de rendre compte de l'ensemble des indicateurs de service.</p> <p>Des tâches planifiées de <i>backup</i> partiel et de <i>backup</i> complet du <i>datawarehouse</i> sont programmées et configurées.</p> <p>Les tâches planifiées produisent les résultats attendus.</p> <p>La documentation couvre les principaux cas d'usage de gestion de l'entrepôt :</p>
---	---	--	---

<p>etc.).</p>			<p>l'intégration de nouvelles sources de données, la création de nouveaux accès à l'entrepôt de données, espace de stockage, <i>datamarts</i>, capacité de calcul...</p> <p>La documentation est structurée par cas d'usage et explicite la mise en œuvre des procédures concernées.</p> <p>Les nouvelles sources de données sont correctement configurées et ajoutées au processus d'alimentation de l'entrepôt de données.</p> <p>Les <i>ETL</i> sont mis correctement à jour en fonction.</p> <p>Les nouveaux accès à l'entrepôt de données sont configurés conformément au besoin.</p> <p>Le registre des traitements de données personnelles intègre l'ensemble des traitements de données personnelles impliqués dans le projet d'entrepôt de données.</p> <p>Les procédures de tri des données personnelles pour la mise en conformité de l'entrepôt de données avec le RGPD sont rédigées.</p> <p>Les procédures de tri détaillent les traitements de conformité (automatisés</p>
---------------	--	--	---

			ou non) à appliquer ainsi que leur fréquence d'exécution.
<ul style="list-style-type: none"> - Qualification des variations à intégrer - Intégration des variations à l'entrepôt de données et aux <i>ETL</i> - Mise à jour de la documentation technique, notamment les modèles logiques des données 	<p>C17. Implémenter des variations dans les dimensions de l'entrepôt de données</p> <p>en appliquant la méthode adaptée en fonction du type de changement demandé afin d'historiser les évolutions de l'activité de l'organisation et maintenir ainsi une bonne capacité d'analyse.</p>		<p>La modélisation des variations intègre pleinement les changements dans les données sources.</p> <p>La modélisation des variations permet d'historiser les changements dans les données sources.</p> <p>Les variations sont intégrées à l'entrepôt de données.</p> <p>L'intégration des variations respecte la modélisation initiale.</p> <p>Les ETL sont mis à jour en fonction des besoins liés aux variations.</p> <p>La documentation est à jour, avec les variations.</p>
<p>Bloc de compétences 4 : Encadrer la collecte massive et la mise à disposition des données issues de l'activité de l'organisation avec un <i>data lake</i>*</p>			

<p><u>A7 : Mise en place du data lake</u></p> <ul style="list-style-type: none"> - Sélection des systèmes compatibles avec l'infrastructure et la politique IT de l'organisation - Sélection des outils <i>batch</i>* nécessaire à la collecte et l'alimentation en données - Sélection des outils streaming d'acquisition nécessaire à la collecte et l'alimentation en données - Sélection de l'outil de catalogue* de données adapté au modèle de données et à l'infrastructure du <i>data lake</i> 	<p>C18. Concevoir l'architecture du <i>data lake</i> en sélectionnant les technologies appropriées en fonction de la volumétrie, de la variété et de la vitesse des données dans le but de définir l'architecture technique optimale à intégrer.</p>	<p>E7. Mise en situation (C18, C19, C20, C21)</p> <p>L'évaluation doit se faire dans un contexte de réalisation d'un projet fictif ou d'un projet professionnel réalisé en poste. Le projet évalué s'appuie sur le cadre technique de l'organisation et sur le cadre d'exploitation des données.</p> <p>Le projet a pour but de couvrir toutes les étapes collecte et de mise à disposition de données massives (big data) avec un data lake</p> <p><u>Livrable</u> : rapport professionnel individuel</p> <p><u>Évaluation</u> :</p> <ul style="list-style-type: none"> - Correction du rapport professionnel - Soutenance orale individuelle 	<p>Les propositions techniques sont cohérentes avec le cadre d'exploitation.</p> <p>Le schéma d'architecture tient compte des contraintes de volume, de vitesse et/ou de variété.</p> <p>Le schéma d'architecture est lisible et utilise un formalisme approprié.</p> <p>Plusieurs catalogues sont proposés et comparés au regard des contraintes.</p> <p>L'outil de catalogue sélectionné répond aux contraintes d'exploitabilité et de droits d'accès.</p>
<ul style="list-style-type: none"> - Configuration du système de stockage. - Configuration des outils <i>batch</i> et temps réel. - Installation de l'outil de catalogue de données. - Documentation des procédures de configuration et d'installation. 	<p>C19. Intégrer les différents composants d'infrastructure du <i>data lake</i> en appliquant la procédure adaptée, afin d'assurer l'acquisition, le stockage et la mise à disposition du catalogue de données</p>		<p>L'ensemble des éléments de documentation nécessaires à la mise en œuvre de la procédure d'installation sont présentés.</p> <p>La procédure d'installation se déroule sans erreur dans un environnement de test.</p> <p>Le système de stockage est installé et fonctionnel en environnement de test.</p>

			<p>Les outils <i>batch</i>* et temps réel sont fonctionnels et connectés au système de stockage.</p> <p>Le catalogue est connecté au système de stockage.</p> <p>La documentation couvre la procédure d'installation et de configuration du système de stockage, des outils batch et de l'outil de catalogue.</p>
<p><u>A8 Gestion de l'accès aux données</u></p> <ul style="list-style-type: none"> - Choix de la ou des méthodes d'alimentation du <i>data lake</i>. - Programmation des procédures d'alimentation automatisées. - Programmation des procédures de suppression automatique liées au cycle de vie. - Maintien en condition opérationnelle : mise à jour, intégration des évolutions, etc. - Rédaction ou mise à jour du registre des traitements de données personnelles en vue de la mise en conformité du <i>data lake</i> avec le RGPD. - Rédaction des procédures de tri des 	<p>C20. Gérer le catalogue des données en tenant compte de leur nature, de leurs sources d'alimentation et de leur cycle de vie, dans le respect du RGPD, afin de garantir les fonctionnalités du service.</p>		<p>Les choix des méthodes d'alimentation sont justifiés et appropriés à chaque source de données.</p> <p>Les scripts d'alimentation s'exécutent sans erreur.</p> <p>Les données sont importées correctement dans le système de stockage.</p> <p>Les métadonnées sont intégrées dans le catalogue.</p> <p>Les procédures de suppression sont conformes aux contraintes d'accès (notamment réglementaires) et aux contraintes opérationnelles.</p> <p>Le monitoring permet le suivi des conditions matérielles et applicatives.</p>

<p>données personnelles stockées dans le <i>data lake</i> pour la mise en conformité avec le RGPD (par exemple : détection et suppression des données personnelles inutiles, trop anciennes, etc.).</p>			<p>Le monitoring génère une alerte lors d'une rupture de service.</p> <p>Le registre des traitements de données personnelles intègre l'ensemble des traitements de données personnelles impliqués dans le projet d'entrepôt de données.</p> <p>Les procédures de tri des données personnelles pour la mise en conformité de l'entrepôt de données avec le RGPD sont rédigées.</p> <p>Les procédures de tri détaillent les traitements de conformité (automatisés ou non) à appliquer ainsi que leur fréquence d'exécution.</p>
<ul style="list-style-type: none"> - Paramétrage des droits pour l'alimentation en données. - Paramétrage des droits pour la recherche sur l'index des données. - Documentation des accès et des règles de partage. 	<p>C21. Implémenter les règles de gouvernance des données en sécurisant la recherche, la récupération et l'ajout de données afin de respecter les règles de gouvernance des données de l'organisation, dans le cadre du déploiement de la politique de protection des données.</p>		<p>Les droits sont appliqués à des groupes et non à des individus dès que possible.</p> <p>Les accès fournis répondent aux besoins des groupes concernés.</p> <p>Les accès fournis sont limités aux ressources nécessaires aux usages des groupes concernés.</p> <p>Les accès fournis sont conformes à la réglementation (RGPD).</p> <p>La documentation couvre les groupes d'accès et les droits associés ainsi que</p>

			les procédures de mise à jour des règles.
--	--	--	---

Glossaire

- **RICE** : Acronyme qui désigne une méthode de priorisation des tâches en agilité. L'ordre de priorité est inversement proportionnel au score obtenu avec cette méthode (prio 1 correspond au score RICE le plus élevé). La calcul du score prend en compte les 4 éléments de priorisation qui sont : la portée ("Reach"), l'impact ("Impact"), la confiance ("Confident") et l'effort ("Effort").
- **SMART** : L'acronyme SMART peut correspondre à des termes différents selon les caractéristiques de l'objectif que l'on souhaite définir. Nous proposons ici : Spécifique, Mesurable, Acceptable (et Ambitieux), Réaliste, Temporellement défini⁸
- **Feuille de route** : c'est une technique de planification stratégique qui place les objectifs d'un projet et les principaux livrables (tâches, jalons) sur une chronologie, tous regroupés dans une seule représentation visuelle ou graphique
- **Scrum burndown chart** : Un burndown chart ou BDC (en français, graphique d'avancement) est une représentation graphique de l'évolution de quantité de travail restante par rapport au temps sur une période de temps donnée⁹
- **diagram de Gantt** : Le diagramme de Gantt est un outil utilisé en ordonnancement et en gestion de projet et permettant de visualiser dans le temps les diverses tâches composant un projet.¹⁰
- **Versionner** : Garder la trace d'un ou plusieurs fichiers et de l'historique de leurs modifications grâce à un système de gestion de versions.
- **Dépôt Git** : Un dépôt Git est un entrepôt virtuel du projet. Il permet d'enregistrer les versions de code et d'y accéder au besoin.
- **SQL** : (sigle de "Structured Query Language", en français : langage de requête structurée) Il s'agit d'un langage de programmation permettant d'écrire des instructions reconnues par la grande majorité des systèmes de gestion de bases de données relationnelles (abrégié SGBDR) du marché. Ces instructions couvrent 4 domaines :
 - Langage de définition de données,
 - Langage de manipulation de données,
 - Langage de contrôle de données,
 - Langage de contrôle des transactions.

⁸ Source : https://fr.wikipedia.org/wiki/Objectifs_et_indicateurs_SMART

⁹ Source : https://fr.wikipedia.org/wiki/Burndown_chart

¹⁰ source : https://fr.wikipedia.org/wiki/Diagramme_de_Gantt

- **Méthode MERISE** : Merise est une méthode d'analyse, de conception et de gestion de projet informatique¹¹.
- **endpoints** : Les endpoints sont les points de communication avec une API, ouvert publiquement ou nécessitant une authentification.
- **open API** : C'est une spécification pour une norme pour la définition d'interface programmable (ou API) lisible par machine pour décrire, produire, consommer et explorer des services Web RESTful¹² (API REST).
- **Entrepôt de données** : base de données permettant l'analyse et la prise de décisions, alimentée à partir des données opérationnelles de l'organisation.
- **Encodage** : C'est l'action de transformation d'une donnée pour lui appliquer une nature ou une forme souhaitée.
- **Embedding** : Dans un contexte de base de données NoSQL, l'embedding consiste à stocker les données liée à une autre dans la même entité (document, collection...). Dans un contexte SQL, cela aurait donné lieu à une joint 1-n, par exemple.
- **Referencing** : De la même manière que l'embedding, le referencing est une technique pour lier des données entre elles en environnement NoSQL. Cette technique se rapproche de systèmes des jointures en SQL mais reste tout de même différente, notamment dans l'exploitation du lien réalisé entre les données.
- **Approche orientée requêtes** : C'est une méthode de modélisation des données, principalement utilisée dans un contexte NoSQL et / ou de Data Warehouse.
- **NoSQL** : "Famille" de base de données non structurées, notamment performantes dans la gestion de gros volumes de données non structurées.
- **MDM** : Master Data Management (en français : gestion des données de référence) : ensemble des méthodes, outils, concepts et processus permettant de s'assurer que les données de référence soient identifiées et exploitables.
- **RGPD** : Règlement Général sur la Protection des Données. Le RGPD encadre le traitement des données personnelles sur le territoire de l'Union européenne.
- **Fait** : élément issus de données opérationnelles stocké dans une table d'un entrepôt de données. On parle aussi de métrique, pour signifier l'élément de mesure de l'activité étudiée.
- **Dimension** : axe d'analyse stocké dans une table d'un entrepôt de données
- **datamart** : sous-ensemble d'un entrepôt de données destiné à répondre à un besoin spécifique
- **top-down** : approche de conception et modélisation de l'entrepôt de données dans laquelle chaque datamart doit respecter un modèle normalisé pour l'ensemble de l'entrepôt de données
- **bottom-up** (approche de création et modélisation de l'entrepôt de données) : approche de conception et modélisation de l'entrepôt de données dans laquelle chaque l'entrepôt n'est que la

¹¹ Source : [https://fr.wikipedia.org/wiki/Merise_\(informatique\)](https://fr.wikipedia.org/wiki/Merise_(informatique))

¹² Source : https://en.wikipedia.org/wiki/OpenAPI_Specification

combinaison de datamarts indépendants.

- **ETL** (Extract Transform Load) : logiciel permettant de collecter des données en provenance de sources multiples et de les transformer afin de les intégrer à un système de stockage
- **procédure de backup** : sauvegarde des données, généralement sur un système externe, en cas de défaillance matérielle
- **SLA** : le "service-level agreement" (SLA) ou "accord de niveau de service" est un document qui définit la qualité de service, prestation prescrite entre un fournisseur de service et un client.
- **les variations de dimension de type 1, 2 ou 3 de Ralph Kimball** : approches de gestion du changement des dimensions des données via l'écrasement des valeurs (type 1), l'ajout de lignes (type 2) ou l'ajout de colonnes (type 3).
- **data lake** : espace de stockage global des données relationnelles et non relationnelles d'une organisation permettant leur traitement en fonction de besoins non nécessairement connus au moment de l'insertion
- **batch** : méthode de traitement reposant sur des la mise en lots. Cela permet principalement de réduire le nombre de traitements informatiques nécessaires à la réalisation d'une tâche de même nature. Il s'agira par exemple de regrouper plusieurs traitements d'import de données de petit volume en un seul traitement de volume supérieur.
- **catalogue (de données)** : inventaire organisé des données de l'organisation
- **VM (Machine Virtuelle)** : environnement qui reproduit un système informatique complet (processeur, mémoire, réseau, stockage) sur une machine qui l'héberge.
- **conteneur** : environnement qui isole les processus d'une application des autres processus d'un système d'exploitation.