

REFERENTIEL D'ACTIVITES <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	REFERENTIEL DE COMPETENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	REFERENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p><b>BC1. Concevoir et développer une architecture fonctionnelle et matérielle pour assurer l'acquisition, le traitement et le stockage de données massives</b></p> <p><b>Activités 1: Conception d'une architecture et développement des infrastructures supports à l'exploitation des données</b></p> <ul style="list-style-type: none"> <li>Veille scientifique, technologique et réglementaire, identification de l'état de l'art des outils et plateformes logiciels ETL<sup>1</sup></li> </ul>	<p><b>C1. Mettre en œuvre une veille scientifique, technologique et réglementaire concernant les outils et plateformes logiciels d'extraction de traitement et de stockage (ETL) de données massives pour identifier l'état de l'art disponible sur le marché ;</b></p>	<p><b>Webographie – bibliographie professionnelle</b></p> <p>L'établissement d'une webographie et d'une bibliographie a pour objectif d'alimenter une connaissance approfondie de l'écosystème scientifique et technologique de la data.</p> <p>Pour ce faire le candidat doit :</p> <ul style="list-style-type: none"> <li>- Recenser l'ensemble des supports : blogs, forums ou sites internet traitant des technologies de captation de la data ;</li> <li>- Sélectionner les sources d'information professionnelles spécialisées dans le domaine de l'extraction du traitement et du stockages de données ;</li> <li>- Classifier et présenter les différentes sources d'informations ;</li> <li>- Identifier les évolutions règlementaires et les tendances en matière de protection et d'accessibilité aux données ;</li> <li>- Faire ressortir les tendances de l'écosystème au moment de l'exercice.</li> </ul> <p><u>Evaluation</u>: La webographie-bibliographie fait l'objet d'un rendu écrit individuel. Elle est examinée par un jury d'évaluation comprenant un(e) professionnel(le) de la data et un(e) intervenant(e) pédagogique spécialisé(e).</p>	<p><b>Pour la webographie-bibliographie professionnelle</b></p> <ul style="list-style-type: none"> <li>- Les blogs, sites et forums recensés sont récents et permettent d'accéder à des informations sur l'état de l'art scientifique et technologique de l'écosystème de la data ;</li> <li>- La webographie rassemblée couvre les 3 processus : importation, traitement et stockage des données ;</li> <li>- Les revues professionnelles en ligne et hors ligne sont classifiées selon leurs thématiques ;</li> <li>- Un chapitre spécifique est consacré aux réglementations et référentiels concernant l'accessibilité des données (RGAA<sup>2</sup>) et la protection des données personnelles (RGPD) ;</li> <li>- Le lien entre la source d'informations et l'analyse que l'on peut en faire est clairement établi ;</li> <li>- Les sources d'information ont permis de faire ressortir plusieurs tendances du marché de la data et de l'IA.</li> </ul>

<sup>1</sup> Extract – Transform - Load

<sup>2</sup> Référentiel Générale d'Amélioration de l'Accessibilité

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
<ul style="list-style-type: none"> <li>Spécification des besoins et dimensionnement des infrastructures de données massives dans une perspective pluriannuelle</li> <li>Evaluation et sélection d'une solution ETL globale incluant des développements en propre, des plateformes logicielles et des API disponibles sur le marché</li> </ul>	<p>C2. <b>Anticiper les besoins futurs pour concevoir et dimensionner une architecture</b> compatible avec la montée en charge des cas d'usage notamment ceux liés à l'IA ;</p> <p>C3. <b>Mobiliser des savoirs techniques hautement spécialisés en plateformes et outils logiciels</b> pour évaluer différentes solutions ETL assurant la collecte, le nettoyage et l'entreposage de données massives internes ou externes structurées, semi-structurées et non structurées.</p>	<p><b>Etude de cas</b></p> <p>Individuellement, sur la base d'une situation réelle reconstituée sous forme de spécification détaillée de besoin, les candidats doivent :</p> <ul style="list-style-type: none"> <li>Identifier et caractériser les solutions ETL du marché correspondant au besoin spécifié ;</li> <li>Proposer des critères pour évaluer les différentes solutions techniques ;</li> <li>Proposer une recommandation de choix de solution ETL ;</li> <li>Proposer une vision pluriannuelle et évolutive de l'architecture et des infrastructures ;</li> <li>Intégrer la solution dans le schéma directeur informatique fourni dans le cas.</li> </ul> <p><u>Evaluation</u> : Rapport écrit évalué par un jury comprenant un(e) professionnel(le) de la data et un(e) intervenant(e) pédagogique spécialisé(e).</p>	<p><u>Pour l'étude de cas</u></p> <ul style="list-style-type: none"> <li>Au moins 5 fournisseurs majeurs de solutions ETL ont été identifiées sur le marché ;</li> <li>Les critères d'évaluation permettent d'évaluer les différentes solutions : ils sont discriminants, exhaustifs et représentatifs de la performance attendue ;</li> <li>Le choix de solutions proposée est conforme à la spécification du besoin ;</li> <li>La vision pluriannuelle prévoit un nombre croissant des cas d'usage liés à l'intelligence artificielle ;</li> <li>Les possibilités d'évolutions du système sont décrites dans un chapitre spécifique du document : les conclusions du chapitre sont intégrables au schéma directeur informatique.</li> </ul>
<ul style="list-style-type: none"> <li>Développement d'une infrastructure et management des flux et des stocks de données dans le respect du RGAA et du RGPD</li> </ul>	<p>C4. <b>Recueillir les besoins des différents cas d'usage</b> notamment ceux liés à l'intelligence artificielle pour développer une architecture de flux de données en établissant un cahier des charges fonctionnel et technique ;</p> <p>C5. <b>Concevoir une infrastructure fonctionnelle et matérielle</b> pour organiser les flux et les traitements de données depuis des sources multivariées internes et externes jusqu'à un utilisateur final (DG,</p>	<p><b>Projet applicatif tutoré</b></p> <p>Conduit au sein d'une entreprise par groupe de 3 à 4 apprenants, le projet applicatif tutoré consiste à réaliser au moins 5 entretiens avec des scientifiques, des analystes et des ingénieurs de la data pour définir le besoin global de mouvements et de stockage de données au sein de l'entreprise.</p> <p>A partir de la définition du besoin les candidats sont chargés de concevoir un schéma de pipeline de données complet de la captation de la donnée brute à la mise à disposition de la donnée traitée et capable d'intégrer les algorithmes d'IA.</p>	<p><u>Pour le Projet applicatif</u></p> <ul style="list-style-type: none"> <li>Les entretiens ont été menés de manière semi-directive et ont favorisé l'idéation et le développement du pipeline en co-construction avec les interviewés ;</li> <li>Les évolutivités horizontale (augmenter le nombre des serveurs) et verticale (augmenter la taille des serveurs) de l'infrastructure ont été analysées et évaluées ;</li> <li>La mise à disposition des données traitées prévoit des dispositifs d'accès aux données pour les PSH</li> <li>Les principes de conception universelle inclus dans la charte de l'ONU (2007) sont respectés ;</li> </ul>

REFERENTIEL D'ACTIVITES <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	REFERENTIEL DE COMPETENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	REFERENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
	<p>Direction marketing, Data analystes, data scientistes ...);</p> <p><b>C6. Mobiliser les principes de la RSE et le RGPD</b> pour participer à l'établissement des règles de gouvernance et de conservation des données et garantir l'accessibilité des data selon les normes du RGAA.</p>	<p><u>Evaluation</u> : Rapport écrit et présentation orale devant un jury d'évaluation comprenant un(e) professionnel(le) de la data employé par l'entreprise commanditaire et un(e) intervenant(e) pédagogique spécialisé(e).</p>	<ul style="list-style-type: none"> <li>- Les principes du RGPD sont pris en compte dans l'énoncé des principes de gouvernance des données ;</li> <li>- Les indicateurs de performance du pipeline incluent au moins 3 critères RSE dont l'efficacité énergétique ;</li> <li>- Les 13 points de test du référentiel RGAA sont pris en compte dans le rendu du projet ;</li> <li>- Le commanditaire est satisfait du livrable qui répond à son besoin.</li> </ul>
<p><b>BC2. Déployer des outils d'extraction et de stockage pour recueillir des données brutes structurées, non structurées, internes ou externes dans un lac de données</b></p> <p><b>Activités 2 : Déploiement d'outils d'extraction et stockage de données brutes</b></p> <ul style="list-style-type: none"> <li>▪ Identification et sélection des solutions data lake les plus performantes sur le marché</li> </ul>	<p><b>C7. Recenser et caractériser les outils d'extraction et de stockage de données</b> brutes disponibles sur le marché pour sélectionner la solution la plus adaptée aux besoins de l'entreprise ;</p> <p><b>C8. Mobiliser des savoirs techniques approfondis en déploiement de bases de données</b> pour évaluer les mérites et les limites d'un stockage en propre par rapport à une solution déportée sur le cloud.</p>	<p><b>Etudes de cas</b></p> <p>Individuellement, sur la base d'une situation réelle reconstituée sous forme de spécification détaillée de besoin, les candidats doivent :</p> <ul style="list-style-type: none"> <li>- Identifier et caractériser les solutions de data lake disponibles sur le marché ;</li> <li>- Déterminer les avantages et les inconvénients de chaque solution selon des critères de performance : temps, facilité d'accès aux données ... ;</li> <li>- Comparer une solution de stockage déportée sur le cloud par rapport à un stockage en propre</li> <li>- Comparer les modes d'importation de données par batch ou en continue par streaming ;</li> <li>- Proposer une recommandation de choix de solution data lake.</li> </ul> <p><u>Evaluation</u> : Rapport écrit évalué par un jury comprenant un(e) professionnel(le) de la data et un(e) intervenant(e) pédagogique spécialisé(e).</p>	<p><u>Pour l'étude de cas</u></p> <ul style="list-style-type: none"> <li>- Au moins 5 solutions data lake répondant potentiellement au besoin ont été identifiées sur le marché ;</li> <li>- L'évaluation des différentes solutions est réalisée sur la base de critères discriminants, exhaustifs et représentatifs de la performance globale attendue du data lake ;</li> <li>- Les mérites et limites de l'importation par batch et de l'importation en streaming ont été examinés suivant au moins 3 caractéristiques correspondant aux critères de performance ;</li> <li>- Les solutions logicielles data lake préconisées permettent d'ingérer des données structurées, semi-structurées et non-structurées ;</li> <li>- Le choix de solutions est conforme à la spécification du besoin.</li> </ul>
<ul style="list-style-type: none"> <li>▪ Déploiement d'outils d'extraction et de stockage de données brutes en interne ou dans le cloud</li> </ul>	<p><b>C9. Mettre en œuvre un processus complet d'extraction</b> pour recueillir des données issues de sources multiples et les stocker dans une</p>	<p><b>Travaux pratiques</b></p> <p>Les apprenants sont amenés à mobiliser en binôme les outils d'une plateforme logicielle type Microsoft Azure pour :</p>	<p><u>Pour les travaux pratiques</u></p> <ul style="list-style-type: none"> <li>- L'importation des données a été effectué sans perte de données ;</li> </ul>

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
<ul style="list-style-type: none"> <li>▪ Management des opérations de traitement des données brutes avant stockage dans un data lake</li> <li>▪ Supervision de la rédaction des documents supports à l'utilisation des outils et plateformes logicielles</li> <li>▪ Management de la mise en place des bases de données internes et externes dans le respect du RGPD et des critères du RGAA</li> </ul>	<p>base de données brutes (lac de données) centralisée ;</p> <p>C10. <b>Identifier et préparer toutes les données pertinentes</b> afin de faciliter leur traitement et leur stockage ultérieurs dans un entrepôt de données ;</p> <p>C11. <b>Documenter l'utilisation des bases de données externes et internes</b> pour gérer et entretenir le lac de données en respectant le RGPD et en garantissant l'accessibilité des données selon les normes du RGAA.</p>	<ul style="list-style-type: none"> <li>- Importer des données massives semi-structurées et non structurées ;</li> <li>- Stocker des données brutes dans une base de données (lac de données) dans le respect du RGPD ;</li> <li>- Préparer les données avant traitement ;</li> <li>- Etablir des requêtes SQL et NO-SQL ;</li> <li>- Documenter la logique de stockage ;</li> <li>- Veiller à la structuration, la présentation et l'accessibilité selon la RGAA.</li> </ul> <p><u>Evaluation</u> : Rapports écrits de travaux pratiques rédigés par un binôme d'apprenants et évalué par un jury comprenant un(e) professionnel(le) de la data et un(e) intervenant(e) pédagogique spécialisé(e).</p>	<ul style="list-style-type: none"> <li>- Une base de données relationnelle a été mise en place et son bon fonctionnement a été démontré et documenté dans le respect du RGPD ;</li> <li>- Un schéma de données relationnelles pour les données structurées a permis d'effectuer des requêtes SQL ;</li> <li>- Une base de données non-relationnelle a été mise en place et son bon fonctionnement a été prouvé et documenté dans le respect du RGPD ;</li> <li>- La structuration de la base de données de type clé-valeurs a permis d'effectuer des requêtes NO-SQL ;</li> <li>- Le nettoyage des données a été effectué avec succès dans au moins 3 registres : suppression des doublons, identification des données obsolètes et élimination des données incohérentes ou erronées ;</li> <li>- Les possibilités d'évolutions du système et les logiques de stockage sont décrites dans un document spécifique dont le sommaire et le contenu sont conformes aux attendus de la profession ;</li> <li>- La structuration et la présentation des informations élémentaires fait l'objet d'une analyse selon le RGAA (critères 9 et 10).</li> </ul>
<p><b>BC3. Mettre en œuvre des solutions informatiques pour traiter et entreposer des données exploitables par l'intelligence artificielle et l'analyse humaine</b></p>	<p>C12. <b>Synthétiser les besoins des utilisateurs</b> pour concevoir des modèles de données performants et résilients afin de faciliter les travaux de data science et de datavisualisation ;</p> <p>C13. <b>Mobiliser des savoirs techniques avancés en gestion de bases de</b></p>	<p><b>Etudes de cas</b></p> <p>Individuellement, sur la base d'une situation réelle reconstituée sous forme de spécification détaillée de besoin, les candidats doivent :</p> <ul style="list-style-type: none"> <li>- Identifier et caractériser les solutions de traitement automatisé et d'entreposage de données disponibles sur le marché ;</li> </ul>	<p><u>Pour l'étude de cas</u></p> <ul style="list-style-type: none"> <li>- Au moins 5 solutions Data Warehouse répondant potentiellement au besoin ont été identifiées sur le marché ;</li> <li>- L'évaluation des différentes solutions est réalisée sur la base de critères discriminants, exhaustifs et représentatifs de la performance globale</li> </ul>

REFERENTIEL D'ACTIVITES <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	REFERENTIEL DE COMPETENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	REFERENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p><b>Activités 3 : Mise à disposition de données fiables et pertinentes</b></p> <ul style="list-style-type: none"> <li>▪ Evaluation et sélection d'une solution de traitement et d'entreposage des données</li> </ul>	<p><b>données</b> pour sélectionner et mettre en œuvre les solutions de traitement et de stockage massifs adaptées au besoin de l'entreprise ;</p> <p><b>C14.Veiller à l'application des critères du RGAA dans la mise en œuvre des solutions</b> pour garantir l'accessibilité des données aux PSH ;</p>	<ul style="list-style-type: none"> <li>- Déterminer les avantages et les inconvénients de chaque solution selon des critères de performance : accessibilité aux données, temps d'accès, ... ;</li> <li>- Comparer une solution d'entrepôt déporté dans le cloud par rapport à un stockage des données en propre ;</li> <li>- Proposer une recommandation de choix de solution entrepôt de données ;</li> <li>- Documenter la logique de traitement et d'entreposage des données ;</li> <li>- Evaluer les solutions techniques à travers les critères du RGAA.</li> </ul> <p><u>Évaluation</u> : Rapport écrit évalué par un jury comprenant un(e) professionnel(le) de la data et un(e) intervenant(e) pédagogique spécialisé(e).</p>	<p>attendue d'un entrepôt de données : les aspects temps d'accès, accessibilité et qualité globale de la donnée font partie des critères de performance retenus ;</p> <ul style="list-style-type: none"> <li>- Les mérites et limites d'un entrepôt dans le cloud et celles d'un stockage sur site ont été examinés suivant au moins 3 axes caractéristiques correspondant aux critères de performance ;</li> <li>- Le choix de solutions proposées est conforme à la spécification du besoin et sont capables d'intégrer des algorithmes d'intelligence artificielle ;</li> <li>- La documentation support montre qu'une analyse des critères du RGAA a été réalisée : les chapitres applicables au cas d'espèce sont identifiés et des solutions concrètes sont proposées.</li> </ul>
<ul style="list-style-type: none"> <li>▪ Déploiement des processus de traitement et de formatage des données brutes pour entreposage</li> <li>▪ Mise à disposition de données fiables et pertinentes pour les utilisateurs</li> <li>▪ Elaboration d'un guide d'utilisation des plateformes et des API déployées</li> </ul>	<p><b>C15.Trier, nettoyer et structurer des données brutes</b> pour fabriquer des données propres exploitables par les data scientists et les data analystes ;</p> <p><b>C16.Organiser le stockage des données dans un entrepôt</b> de manière à assurer la gestion de la data durant la totalité de son cycle de vie conformément aux directives RGPD ;</p> <p><b>C17.Mettre en place des solutions innovantes</b> de types API pour améliorer le temps de mise à</p>	<p><b>Travaux pratiques</b></p> <p>Les apprenants sont amenés à mobiliser en binôme les outils d'une plateforme logicielle type Amazon Redshift pour :</p> <ul style="list-style-type: none"> <li>- Nettoyer et dédupliquer les données avant chargement dans l'entrepôt ;</li> <li>- Transformer les données et rendre leur format compatible avec le système cible de Business Intelligence ;</li> <li>- Effectuer une mise en containers de données</li> <li>- Créer des <b>interfaces de programmation d'application (API)</b> ;</li> <li>- Effectuer des requêtes manuelles et vérifier l'accessibilité des données ;</li> <li>- Rédiger une documentation support à l'utilisation des API ;</li> </ul>	<p><u>Pour les travaux pratiques</u></p> <ul style="list-style-type: none"> <li>- Le chargement des données dans l'entrepôt a été réalisé avec succès : <ul style="list-style-type: none"> <li>▪ Il a été effectué sans perte de donnée ;</li> <li>▪ La qualité des données après traitement est conforme à l'objectif.</li> </ul> </li> <li>- La conteneurisation des données mise en place réduit le temps d'accès aux données et optimise les consommations de ressources ;</li> <li>- Les API ont été réalisés en C# et en JavaScript conformément aux instructions ;</li> <li>- La syntaxe et la grammaire pour ces 2 langages ont été respectés : le binôme a produit des codes informatiques fonctionnels ;</li> </ul>

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
	disposition de données fiables et exploitables par les data analystes.	<u>Evaluation</u> : Rapports écrits de travaux pratiques rédigés par un binôme d'apprenants et évalué par un jury comprenant un(e) professionnel(le) de la data et un(e) intervenant(e) pédagogique spécialisé(e).	<ul style="list-style-type: none"> <li>- Les API ont permis d'établir des liens entre les données dans l'entrepôt et les outils d'analyse et de business intelligence ;</li> <li>- Les requêtes effectuées en langage SQL ont permis des accès à tous les datamarts de l'entrepôt.</li> </ul>
<p><b>BC4. Piloter des projets pour anticiper les besoins et développer en continu les plateformes logicielles (ETL) et les interfaces de programmation applicatives (API)</b></p> <p><b>Activités 4 : Adaptation en continu des fonctionnalités du pipeline, des plateformes et des interfaces</b></p> <ul style="list-style-type: none"> <li>▪ Elaboration d'un cahier des charges technique et fonctionnel répondant aux besoins des scientifiques et des analystes</li> <li>▪ Pilotage du développement d'API, de requêtes et de routines automatisées avec implication de l'ensemble des parties prenantes</li> </ul>	<p><b>C18.Établir un cahier des charges technique et fonctionnel</b> pour répondre aux besoins évolutifs des data scientifiques, data analystes, partenaires métiers et de la direction générale de l'entreprise ;</p> <p><b>C19.Piloter des projets en mobilisant des équipes pluridisciplinaires</b> (data scientifiques, responsable marketing, direction générale, ...) pour développer des interfaces types API et des requêtes automatisées à destination de cas d'usage multivariés ;</p> <p><b>C20.Mobiliser des connaissances approfondies en programmation</b> pour développer et superviser le développement des routines informatiques en mode DEV/OPS<sup>3</sup> et réduire le temps de mise à disposition de nouvelles fonctionnalités utiles à l'utilisateur des données.</p>	<p><b>Hackathon 1</b> Cette épreuve se déroule pendant 24h nonstop et consiste, par équipe de 3 apprenants, à imaginer des solutions innovantes, établir un cahier des charges pour construire un entrepôt de données et une API qui permet d'accéder aux données.</p> <p><u>Evaluation</u> : Démonstration des performances de l'entrepôt et du prototype d'API devant un jury comprenant un(e) professionnel(le) de la data et un(e) intervenant(e) pédagogique spécialisé(e).</p> <p><b>Travaux pratiques</b> Les apprenants sont amenés à mobiliser en binôme les outils et les méthodes DEV/OPS :</p> <ul style="list-style-type: none"> <li>- Création de lignes de codes à l'aide d'outils logiciels spécifiques ;</li> <li>- Test des codes avec des protocoles et outils spécifiques ;</li> <li>- Applications de la méthode SCRUM pour réaliser un développement informatique CI/CD (intégration et déploiement en continu).</li> </ul> <p><u>Evaluation</u> : Démonstration des fonctionnalités des programmes créés en mode DEV/OPS en CI/CD</p>	<p><u>Pour l'Hackathon 1</u></p> <ul style="list-style-type: none"> <li>- Le cahier des charges comprend une partie fonctionnelle et une partie technique ;</li> <li>- Le prototype d'API livrable est décrit selon au moins 3 perspectives : <ul style="list-style-type: none"> <li>▪ Performance en termes de temps d'accès,</li> <li>▪ Potentiel évolutif</li> <li>▪ Fiabilité et répétabilité</li> </ul> </li> <li>- La démonstration du prototype est convaincante, les fonctionnalités démontrées sont conformes au cahier des charges.</li> </ul> <p><u>Pour les travaux pratiques</u></p> <ul style="list-style-type: none"> <li>- Les candidats ont mobilisé les techniques de développement agile et en particulier la méthode SCRUM et le séquençage du projet en Sprints ;</li> <li>- Les environnements outils et langages spécifiques au DEV/OPS type GitHub GitLab sont maîtrisés ;</li> <li>- Les protocoles et outils de test type JUnit, TestNG ont été mobilisés de manière appropriée ;</li> <li>- Les réponses aux questions lors de l'oral de démonstration ont permis de vérifier que la culture et les techniques de développement et de</li> </ul>

<sup>3</sup> Techniques et outils de développement permettant de développer et déployer en simultanée un programme informatique

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
		rapport de test écrit par un binôme d'apprenants et évalué par un jury comprenant un(e) professionnel(le) de la data et un(e) intervenant(e) pédagogique spécialisé(e).	déploiement simultanés DEV/OPS et CI/CD sont intégrées.
<ul style="list-style-type: none"> <li>▪ Anticipation des besoins utilisateurs et amélioration continue des plateformes logicielles ETL</li> <li>▪ Mise en œuvre d'indicateurs de performance des projets sur l'ensemble de leur cycle de vie</li> </ul>	<p><b>C21. Collaborer avec les chefs de produits, les responsables métiers et la direction générale</b> afin d'anticiper les nouveaux usages de la data et élaborer des nouvelles fonctionnalités ;</p> <p><b>C22. Optimiser en continu les processus ETL pour proposer des solutions accessibles</b> aux PSH et plus efficaces en termes de consommation d'énergie et de ressources ;</p> <p><b>C23. Mettre en place des indicateurs de performance projet</b> pour assurer la production de livrables au niveau de qualité attendu, dans les délais et les budgets fixés par le cahier des charges.</p>	<p><b>Projet applicatif tutoré</b> Conduit au sein d'une entreprise par groupe de 3 à 4 apprenants, le projet applicatif tutoré consiste à :</p> <ul style="list-style-type: none"> <li>- Réaliser au moins 5 entretiens avec des chefs de produits et des Responsables métiers (DG, RH, production, Supply Chain, marketing, finances, ...) ;</li> <li>- Définir les besoins en termes de nouveaux usages de la data notamment à travers l'apprentissage machine ;</li> <li>- Rédiger le cahier des charges fonctionnel et technique de l'API ;</li> <li>- Conduire un projet sous tutorat pédagogique pour créer une nouvelle interface (API) fiable, évolutive et documentée ;</li> <li>- Emettre des recommandations pour améliorer l'accessibilité de l'API dans les situations de handicap.</li> </ul> <p><u>Evaluation</u> : Démonstration des fonctionnalités du prototype de l'API devant le jury d'évaluation avec questions – réponses orales. Rapport écrit documentant les fonctionnalités, l'accessibilité et l'évolutivité de l'API. Le jury d'évaluation comprend un(e) professionnel(le) de la data employé par</p>	<p><u>Pour le Projet applicatif</u></p> <ul style="list-style-type: none"> <li>- Les entretiens ont été menés de manière semi-directive et ont favorisé l'émergence de nouvelles idées pour de nouveaux usages liés à la data ;</li> <li>- Les opportunités et contraintes liées à l'apprentissage machine et aux algorithmes d'IA ont été exposées et discutées avec les responsables métiers ;</li> <li>- Le cahier des charge techniques et fonctionnels est cohérent avec la synthèse de définition du besoin ;</li> <li>- Les indicateurs de performance permettent de piloter la performance projet depuis la transcription du besoin à la production des livrables finaux ;</li> <li>- Le prototype est fonctionnel : au moins 80% des fonctionnalités prévues au cahier des charges sont opérantes ;</li> <li>- L'évolutivité de l'API est démontrée et documentée ;</li> <li>- Le commanditaire est satisfait du livrable qui répond à son besoin.</li> </ul>

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
		l'entreprise commanditaire et un(e) intervenant(e) pédagogique spécialisé(e).	
<p><b>BC5. Manager l'industrialisation des mouvements de data au sein de l'entreprise pour mettre à disposition des utilisateurs des données fiables et pertinentes</b></p> <p><b>Activités 5 : Fiabilisation et qualité des infrastructures, des plateformes et des API</b></p> <ul style="list-style-type: none"> <li>Management des processus d'automatisation et de fiabilisation des infrastructures, des plateformes et des interfaces</li> <li>Contribution à l'amélioration de la qualité des données mises à dispositions des utilisateurs</li> </ul>	<p><b>C24. Automatiser et fiabiliser les processus d'extraction et de traitement</b> pour réduire le temps de mise à disposition de données fiables et pertinentes ;</p> <p><b>C25. Fiabiliser une infrastructure, des plateformes logicielles et des API</b>, pour mettre des données fiables et pertinentes à disposition des utilisateurs en anticipant les opérations de maintenance et de mise à jour ;</p> <p><b>C26. Réaliser des tests d'intégration avant déploiement et organiser des retours sur expérience utilisateurs</b> pour contribuer à l'assurance qualité de l'architecture et de l'infrastructure des mouvements et stockages de données ;</p> <p><b>C27. Mobiliser des savoirs hautement spécialisés pour évaluer et anticiper les risques concernant la sécurité des données en lien avec le RSSI et la protection des données en lien avec le DPO.</b></p>	<p><b>Travaux pratiques</b></p> <p>Les apprenants sont amenés à mobiliser, en binôme, les outils de la plateforme ETL Talend pour réaliser des développements en Java qui visent à :</p> <ul style="list-style-type: none"> <li>Mettre en œuvre des routines automatisées pour transformer, agréger et mettre à jour les données ;</li> <li>Analyser et améliorer en continue la qualité des données ;</li> <li>Combiner de manière automatique des données provenant de différentes sources internes et externes ;</li> <li>Concevoir, développer et mettre en œuvre des protocoles de tests avant déploiement de l'infrastructure ETL ;</li> <li>Mesurer et caractériser les écarts après tests avec les objectifs poursuivis ;</li> <li>Proposer des actions de remédiations pour résorber les écarts ;</li> <li>Prendre en compte la protection et la sécurité des mouvements et stockage de données.</li> </ul> <p><u>Evaluation</u> : Évalué par un jury comprenant un(e) professionnel(le) de la data et un(e) intervenant(e) pédagogique spécialisé(e).</p>	<p><u>Pour les travaux pratiques</u></p> <ul style="list-style-type: none"> <li>La syntaxe et la grammaire du langage Java ont été respectés : le binôme a produit des codes informatiques fonctionnels ;</li> <li>La combinaison des données de différentes sources a permis de faire émerger des données références (Master data) ;</li> <li>Les routines d'analyse de la qualité des données ont été mises en œuvre avec succès ;</li> <li>Les protocoles de tests ont été définis et mis en œuvre avant le déploiement</li> <li>Les écarts après tests ont été identifiés et plusieurs remédiations possibles ont été proposées ;</li> <li>La sécurité et la protection des données ont été abordés dans un document spécifique ;</li> <li>Les principes du RGPD et ceux de la protection des données contre les risques d'intrusions externes sont maîtrisés.</li> </ul>

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
<ul style="list-style-type: none"> <li>Conception et mise en œuvre de Tests d'intégration avant déploiement d'algorithmes d'IA, de nouvelles plateformes logicielles ou d'API</li> </ul>	<p><b>C28. Assurer la fourniture de données propres et pertinentes</b> pour faciliter l'apprentissage machine et la science des données en collaboration avec les data scientistes ;</p> <p><b>C29. Déployer à grande échelle de nouveaux algorithmes</b> conçus en laboratoire par les data scientistes pour mieux modéliser les relations entre les données ;</p> <p><b>C30. Prendre en compte les risques sur le plan de l'éthique et la RSE</b> vis-à-vis des technologies d'intelligence artificielle pour définir les critères qualité d'une infrastructure data fiabilisée et industrialisée.</p>	<p><b>Hackathon 2</b></p> <p>Cette épreuve se déroule pendant 24h nonstop et consiste, par équipe de 3 apprenants, à choisir des algorithmes d'apprentissage, à les tester et les déployer en situation d'apprentissage machine supervisé</p> <p><u>Evaluation</u>: Démonstration des performances d'apprentissage machine en mode supervisé d'un algorithme industrialisé et réflexion éthique et RSE concernant les technologies mises en œuvre devant un jury comprenant un(e) professionnel(le) de la data et un(e) intervenant(e) pédagogique spécialisé(e).</p>	<p><u>Pour l'Hackathon 2</u></p> <ul style="list-style-type: none"> <li>L'industrialisation de l'algorithme a respecté les étapes attendues : test en mode bac à sable, essai-erreur dans le déploiement, analyse du retour sur expérience, déploiement ;</li> <li>Les données de départ ont été annotées pour permettre un apprentissage machine en mode supervisé ;</li> <li>La démonstration de l'algorithme est convaincante : de nouvelles relations apparaissent entre les données ;</li> <li>La réflexion éthique et RSE a permis de caractériser les risques liés à l'apprentissage machine en mode supervisé et non supervisé.</li> </ul>