

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <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>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
BLOC 1 : ANALYSER L'EXISTANT ET PROPOSER DES SOLUTIONS LIÉES A LA DATA			
A.1.1 Analyse des besoins et des problématiques : - référencement et regroupement l'ensemble des sources de données structurées ou non structurées nécessaires et pertinentes à l'analyse - analyse de la problématique liée aux données (<i>data</i>) - conception d'un projet <i>data</i> avec les parties prenantes - prise en compte de l'outil informatique existant (serveurs de stockage).	C.1.1 Réaliser un état des lieux de l'outil informatique au sein d'une organisation privée/publique en cartographiant en profondeur l'intégralité du système d'information, en identifiant les éléments clés et les points de vulnérabilité, et en définissant les priorités, afin d'optimiser les performances et la sécurité du système tout en alignant les ressources technologiques avec les objectifs stratégiques de l'organisation	Réalisation de Projets : en équipe, les candidats devront répondre à une problématique client fictive. En relation avec des experts, ils doivent étudier un système d'information donné, appréhender le volume de données et les risques afin d'apporter des recommandations. Évaluation écrite : Ces évaluations se réalisent sous la forme de questionnaires ou d'exercices. Ils sont d'une durée d'une heure à deux heures au maximum. Il s'agit de questions portant sur la réglementation de la gestion de données massives et sur la méthodologie de l'analyse des besoins et des problématiques.	Le candidat est évalué sur : - la collecte des éléments (l'existant) - l'interaction avec les acteurs du projet - la synthèse de l'existant - l'État des lieux de l'outil informatique permet de cartographier en profondeur l'intégralité du système d'information, les priorités sont définies. - l'évaluation de la pertinence de l'outil informatique en place - l'évaluation des capacités offertes par l'outil informatique existant - les recommandations sur le fait de changer l'outil informatique si nécessaire - La veille technologique ou la veille sur l'Open Data / Données ouvertes (veille gratuite et disponible) réalisée comprend l'actualité des données publiques et de l'administration réglementaire - les données trouvées sont fiables (ex : météo France, sites gouvernementaux...)
	C.1.2 Analyser les problématiques de structuration de la donnée au sein de d'une organisation privée/publique en mettant en place un déploiement d'une architecture ou d'une structure, en interagissant avec des interlocuteurs différents et concernés par la problématique, afin d'identifier les besoins de l'organisation auxquels il convient d'apporter une réponse.		

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <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>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p>A.1.2 Identification des données de l'organisation</p> <ul style="list-style-type: none"> - supervision d'une réunion d'équipe - collecte des données auprès des différents services/responsables métiers - cartographie des données 	<p>C.1.3 Animer une réunion avec tous les acteurs principaux et secondaires concernés par une problématique relevant de la donnée, existante ou nouvelle, en tenant compte de l'avis des différents acteurs, en évoquant les pistes de solutions, et en identifiant les besoins d'adaptation de la démarche en raison d'une situation de handicap, afin de contextualiser l'intervention, d'intégrer le fonctionnement, les caractéristiques et les contraintes de l'organisation.</p>	<p>Mise en situation professionnelle avec un jeu de données en équipe : il s'agit d'un travail en équipe sur des sites, afin de trouver des algorithmes. Ce travail se déroule en fil rouge. Les candidats sont dans une situation d'un travail en équipe, avec un chef d'équipe qui répartit les tâches au sein du groupe (vérification de la base, travail des algorithmes, rédaction d'un rapport...).</p> <p>Le candidat est mis en situation dans un groupe dont il anime une réunion en tant que chef de projet.</p>	<p>Le candidat est évalué sur :</p> <ul style="list-style-type: none"> - la pertinence de ses interrogations - l'écoute active - la reformulation des besoins du client en tenant compte des enjeux - la gestion de difficultés/conflits éventuelles - sa capacité de conviction pour l'obtention des données - l'adhésion des collaborateurs externes au projet par la présentation d'exemples ou la production de résultats préliminaires - l'adaptation aux personnes et son respect du travail effectué au préalable - un rythme adapté au projet - la retranscription des données récupérées en interne au sein d'une cartographie - la vérification de l'anonymisation des données si nécessaire - la cartographie comprenant le volume et la variété des données, la vitesse d'acquisition, les moyens d'accès et de requête automatique des données - le respect de la réglementation en vigueur sur la confidentialité des données (CNIL, RGPD) - la qualité de sa veille sur les réglementations inhérentes à la confidentialité

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <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>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p>A.1.3 Formalisation du besoin et des solutions techniques :</p> <ul style="list-style-type: none"> - formalisation du besoin et des enjeux - analyse de manière qualitative et quantitative le projet dans son ensemble - analyse des données (data) et des données non-structurées - réalisation d'un cahier des charges 	<p>C.1.4 Rédiger le cahier des charges en tenant compte des spécifications techniques et organisationnelles identifiées, dans le respect du budget alloué dans le respect du budget alloué, afin de proposer des solutions réalisables et réalistes avec les données et ressources propres possédées.</p>	<p>Réalisation de projets : réalisation d'un cahier des charges présentant le besoin et les contraintes (budget, délai, technique) et le présenter à l'oral.</p> <p>En plus du rapport de projet, le candidat devra rendre un suivi écrit retraçant les éventuelles difficultés rencontrées et liées à la réalisation de l'objectif du projet.</p>	<p>Le candidat est évalué sur :</p> <ul style="list-style-type: none"> - la réalisation d'un cahier des charges reprenant bien tous les enjeux et objectifs auxquels le commanditaire est soumis. - la réalisation d'un cahier des charges reprenant les fonctions, les critères et niveaux d'appréciation - la prise en compte des contraintes liées au développement - la proposition d'un développement souple des solutions (méthode Agile, outils modernes...) - la réalisation d'un diagramme de Gantt pour planifier de façon optimale ainsi que communiquer sur le planning établi et les choix qu'il impose. <p>La réalisation du diagramme de Gantt permet :</p> <ul style="list-style-type: none"> * la détermination des dates de réalisation d'un projet ; * l'identification des marges existantes sur certaines tâches ; * la visualisation d'un seul coup d'œil le retard ou l'avancement des travaux. <ul style="list-style-type: none"> - Les recommandations et l'utilisation d'outils appropriés pour le développement

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <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>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p>A.1.4 Préconisations de stratégies en matière de possession, gestion et transformation des données :</p> <ul style="list-style-type: none"> - proposition d'une stratégie Data - prise en compte du droit des données - prise en compte du <i>cloud computing</i> (serveurs, stockage, mise en réseau, logiciels) de l'entreprise/collectivité dans sa stratégie - prise en compte des risques liés à la donnée - proposition de serveurs auprès des utilisateurs de l'organisation privée/public. 	<p>C.1.5 Concevoir une stratégie <i>Data</i> en réponse aux problématiques identifiées, en proposant une solution technique adaptée aux caractéristiques de l'organisation en termes de serveurs, stockage, mise en réseau, logiciels, et les risques liés à la donnée, afin d'assurer une gestion et une transformation des données sécurisée, conforme aux réglementations et alignée avec les objectifs stratégiques de l'organisation.</p>	<p>Étude de cas : Il est remis au candidat un scénario, dont il doit faire un rapport détaillé en tenant compte des différents aspects nécessaires au respect de la procédure réglementaire.</p> <p>Évaluations écrites : Elles se réalisent sous la forme de questionnaires ou d'exercices. Elles sont d'une durée d'une heure à deux heures au maximum. Il s'agit de questions portant sur la réglementation de la gestion de données massive et sur les stratégies en matière de possession, gestion et transformation des données.</p>	<p>Le candidat est évalué sur :</p> <ul style="list-style-type: none"> - la pertinence de la stratégie Data proposée - le respect du droit des données dans sa proposition - la nature technique et la modernité des solutions proposées - la prise en compte des différents risques liés à la donnée (risque d'image, risque technique, qualité du matériel, sensibilisation des collaborateurs, crash et pannes, sauvegarde des données...) - la qualité de sa veille technologique - la proposition d'outils en adéquation avec le niveau technique des utilisateurs potentiels
	<p>C.1.6 Présenter un rapport détaillé de faisabilité en tenant compte des paramètres extérieurs tels que le développement durable, la réglementation en vigueur liée à la confidentialité des données, l'éthique afin de les rendre compatibles avec les contraintes inhérentes à l'organisation.</p>		
	<p>C.1.7 Organiser le déploiement opérationnel du projet en identifiant les ressources humaines, matérielles et financières à disposition, en déterminant les besoins manquants, en constituant une équipe projet et en fixant un rétroplanning, afin d'assurer une planification efficace et une exécution réussie du projet.</p>		

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <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>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
BLOC 2 : DÉPLOYER DES INFRASTRUCTURES INFORMATIQUES D'ANALYSE DE LA DATA			
<p>A.2.1 Définition de l'architecture de stockage</p> <ul style="list-style-type: none"> - identification des solutions de stockage possibles des données - utilisation d'un logiciel de stockage des données (Hadoop) - proposition solutions d'entrepôts de données (stockage interne, stockage en ligne) en fonction de l'organisation - explication et documentation du choix de la solution proposée - prise en compte de la législation des données selon la situation des serveurs (UE/hors UE) 	<p>C.2.1 Identifier les modalités d'organisation assure le stockage et le format des données internes, structurées ou non structurées, en tenant compte des données manquantes, erronées ou perdues, des différents processus en fonction des pays, des entités qui vont produire les flux de données et de la réglementation en vigueur sur la confidentialité des données (RGPD) afin de disposer un jeu de données adaptées (ex : Data Frame Python, R/Tibble).</p>	<p>Réalisation de projets : étude sur la mise en place d'un projet data en intégrant les contraintes techniques et législatives. Soutenance orale du projet</p>	<p>Le candidat est évalué sur :</p> <ul style="list-style-type: none"> - la description correcte des différentes solutions de stockage - l'explication de la solution choisie à mettre en place en fonction comparativement aux solutions existantes - la mise en place de connexions solides et la gestion non seulement de la sécurité des périphériques, mais également celle du réseau. - la description précise des avantages et inconvénients inhérents à chaque solution - - la proposition d'un stockage en fonction du budget, des besoins et des capacités de l'entreprise

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <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>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
A.2.2 Tests et mise en place de protocoles - réalisation de tests sur les web services - rédaction des protocoles à destination des utilisateurs de ces outils - communication auprès des équipes sur les protocoles existants et sur leur utilisation	C.2.2 Réaliser des tests sur les web services , en appliquant des tests de montée en charge de l'outil et sa sécurité, en évaluant la performance et la fiabilité des services, et en analysant le besoin pour déterminer la nécessité d'une API, afin de valider le fonctionnement de l'outil	Réalisation de projets : le projet porte sur la rédaction de protocoles / tutoriels. Le candidat doit préparer des exemples types, et fournir des solutions adaptées à ces derniers. La mise en place d'un forum et wiki de discussion autour de l'outil mis en place est aussi requise.	Le candidat est évalué sur : - la réalisation des tests pour la validation du fonctionnement de l'outil - la réalisation de tests de montée en charge de l'outil (volume des données traitées, qualité de requête des données...) - la rédaction de programmes de solutions de l'outil - la capacité à porter un jugement sur la nécessité d'une API (Application Programming Interface) pour l'analyse du besoin - la capacité à créer une API (Application Programming Interface) - la capacité de rédaction de protocoles et tutoriels adaptés à l'outil et aux futurs utilisateurs (ex. : forum, wiki autour de l'outil mis en place) - la mise en place d'exemples types et de solutions inhérentes
	C.2.3 Rédiger des protocoles à destination des utilisateurs des web services, et en concevant des tutoriels et documents adaptés, en prenant les dispositions permettant leur lecture par tous les utilisateurs y compris en situation de handicap, et en organisant leur diffusion, afin de permettre aux équipes d'utiliser les outils de manière optimale et autonome.		

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <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>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p>A.2.3 Extraction d'information (Information Retrieval) :</p> <ul style="list-style-type: none"> - utilisation d'outils de visualisation de tendances en grandes dimensions - mise en place d'enrichissement des données - calcul d'indicateurs - mise en œuvre d'interconnexion et visualisation de graphes sociaux distribués 	<p>C.2.4 Mettre en œuvre des techniques d'extraction d'information intrinsèque ou par recoupement d'interconnexion en utilisant des outils de visualisation de tendances en grandes dimensions, en mettant en relation différentes bases de données et en actionnant ces bases, afin d'extraire des données pertinentes dans les flux de données</p> <p>C.2.5 Analyser les données de manière approfondie en mettant en œuvre des outils de modélisation (régression, réseaux de neurones), des outils de classification (supervisés ou non) et des outils de réduction de dimension (analyse factorielle), afin de tirer des conclusions exploitables pour la prise de décision.</p>	<p>Réalisation de projets : mise en place d'outils de visualisation des données et de calculs d'indicateurs par exemple à l'aide des bibliothèques pandas de Python et tidyverse de R.</p> <p>Le candidat doit rendre rapport et un script (une succession de commandes commentées soit sur Python/R)</p>	<p>Le candidat est évalué sur :</p> <ul style="list-style-type: none"> - la maîtrise des outils de visualisation classique (d3.js, Tableau, Python, R, Infogram, FusionCharts, Power BI, Jupyter...) - la mise en œuvre des outils sur les données potentiellement massives - la capacité de mettre en relation des bases de données (enrichissement des données) et à actionner ces bases existantes internes ou externes libres de droit - la capacité à utiliser des outils de modélisation (Régression, réseau de neurones...), des outils de classification (supervisés ou non) et des outils de réduction de dimension (analyse factorielle...)

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <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>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
BLOC 3 : TRAITER ET VISUALISER DE LA DATA			
A.3.1 Mise en œuvre de langages et outils de programmation : - utilisation d'une programmation pour l'analyse décisionnelle - utilisation de langages et outils de programmation : <i>ElasticSearch Java mapReduce, Python, R, SCALA,</i> - mise en œuvre des algorithmes de traitement de grosses volumétries de données	C.3.1 Programmer en utilisant des langages de programmation adaptés à la donnée récoltée au sein de l'organisation et provenant de sources extérieures afin de mettre en place des algorithmes gérant le volume important des données	Réalisation de projets : le projet est centré sur les aspects techniques de la résolution d'un ou plusieurs problèmes réels. Le projet se fait en petit groupe (chef de projet, programmeur et testeur). Les candidats devront répondre à une problématique donnée à leur groupe. Un retour d'expérience des autres groupes est également proposé.	Le candidat est évalué sur : - le choix de langues et d'outils de programmation adaptés au problème à résoudre - l'écriture du code dans différents langages - l'écriture d'un code correct - la documentation nécessaire sur le code - l'écriture d'un code respectant les conventions d'écriture usuelles du langage (ex. : PEP 8 pour Python)
A.3.2 Réalisation d'une visualisation dynamique : - choix d'une solution de visualisation adaptée (Chart.js DataHero, Dygraphs, InstantAtlas Plotly, Kibana , Raw ...) - utilisation d'outils de visualisation des données	C.3.2 Produire une visualisation dynamique, en choisissant et en utilisant les outils de visualisation avancés adaptés, en respectant les méthodes de représentation graphique, et en prenant les dispositions nécessaires pour permettre l'accessibilité de ces données, notamment en cas de handicap, afin de faciliter la compréhension et l'analyse des données par l'ensemble des utilisateurs.	Réalisation de projets : réalisation et présentation d'une visualisation dynamique dans le cadre de son projet. Les candidats devront trouver par eux-mêmes une base de données qui nécessite la mise en place d'une visualisation approfondie et dynamique. Les candidats en binôme devront rendre un rapport de projet lors de l'évaluation. L'évaluation est faite individuellement.	Le candidat est évalué sur : - le choix d'une solution de visualisation en cohérence avec la nature des données à représenter - le choix d'une solution de visualisation en cohérence avec la masse de données à représenter - le caractère judicieux des outils de visualisation choisis - le caractère fonctionnel des outils de visualisation - les caractéristiques de la visualisation : simple, fonctionnelle et adaptée aux futurs utilisateurs - le respect des mesures d'accessibilité numérique : la solution est conforme au RGAA, la communication est perceptible, utilisable, compréhensible et robuste.

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <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>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
A.3.3 Analyse des données et conseil auprès du commanditaire : - analyse et visualisation spatio-temporelles - utilisation de <i>Scalable MultiMedia Analytics</i> - visualisation des données permettant de découvrir un phénomène non connu / de confirmer ou d'affirmer une information ou une hypothèse de travail - interprétation des graphiques obtenues - formalisation de rapports et de conseils (stratégiques, économique, organisationnelle...) auprès des utilisateurs dédiés	C.3.3 Réaliser l'analyse et la visualisation spatio-temporelles des données en utilisant des outils de Scalable MultiMedia Analytics, afin de découvrir des phénomènes inconnus, confirmer ou infirmer des hypothèses de travail, et interpréter les graphiques obtenus pour fournir des informations précises et exploitables.	Réalisation de projets : Une base de données issue de tous les domaines possibles est donnée aux candidats avec une problématique bien définie et des questions ouvertes. Des groupes de candidats procèdent à l'analyse des données. Il est attendu une présentation orale des résultats obtenus. L'évaluation est individuelle. Cette présentation est suivie d'échanges et de questions avec l'évaluateur.	Le candidat est évalué sur : - la validation des conditions d'utilisation des outils (vérification des hypothèses) - la cohérence des explications sur les résultats numériques et graphiques obtenus - le traitement des questions et des objections sur les résultats obtenus - la connaissance des limitations des outils utilisés - la pertinence des conseils formalisés en fonction de la problématique initiale - la préconisation de conseils cohérents et adaptés aux interlocuteurs, à la structure et aux moyens (technique, fonctionnel, budget) - le respect des règles de présentation orale fluidité, clarté, précision, concision, efficacité et respect au temps imparti - l'accessibilité des informations communiquées, adaptées à tous les publics, y compris en situation de handicap
	C.3.4 Formaliser des rapports et des conseils stratégiques, économiques et organisationnels en analysant les données visualisées et en interprétant les résultats obtenus, et en établissant des recommandations pertinentes et adaptées aux utilisateurs dédiés, pour optimiser la prise de décision.		

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <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>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
BLOC 4 : ANALYSER ET MODELISER LA DATA			
<p>A.4.1 Utilisation de données en grande dimension :</p> <ul style="list-style-type: none"> - utilisation d'un système de gestion de bases de données (<i>Bigtable, HyperTable NoSQL, SQL...</i>) - transformation de la donnée - recherche de cibles (<i>patterns</i>) enfouies dans les données - segmentation des données afin d'évaluer des probabilités - Mise en place d'un algorithme issu du <i>Machine Learning</i> pour prédire ou comprendre et expliquer. 	<p>C.4.1 Mettre en place un système de gestion des données en anticipant le traitement des données, en agrégeant les données provenant d'autres bases, en normalisant les données et en gérant les valeurs manquantes, afin de permettre une extraction efficace des connaissances sur des jeux de données massives et d'implémenter des outils de classification des données.</p>	<p>Réalisation de projets :</p> <p>Le projet demande dans un premier temps d'expliquer pourquoi nous sommes dans une problématique de grande dimension. La deuxième partie réclame la mise en application des techniques de grande dimension. La problématique est tournée soit vers un problème de classification, soit vers un problème de prédiction.</p>	<p>Le candidat est évalué sur :</p> <ul style="list-style-type: none"> - les choix dans mise en place d'un système de gestion de bases de données, selon le volume, la nature et la variété des données - l'anticipation du traitement des données - l'agrégation des données provenant d'autres bases de données - la normalisation des données - la gestion des valeurs manquantes (ex. : informations manquantes ou partiellement remplies)
	<p>C.4.2 Entraîner un algorithme de <i>machine learning</i> en choisissant l'algorithme approprié et en utilisant des jeux de données pertinents, afin de développer des modèles prédictifs et analytiques robustes, capables de fournir des <i>insights</i> précieux pour l'organisation, par exemple en identifiant les liaisons ou les <i>patterns</i> (motifs) entre les nombreuses bases de données relationnelles.</p>	<p>Le rendu du projet se fera sous forme d'une présentation orale devant un jury d'évaluation</p>	<ul style="list-style-type: none"> - l'identification de la nature des valeurs manquantes - la pertinence du traitement des valeurs manquantes proposé par le candidat - la mise en place de l'extraction de connaissances (ex. : règles d'association) sur des jeux de données massives ou en grande dimension - la mise en place d'outils de classification des données (supervisés ou non supervisés) : classification ascendante hiérarchique, nuées dynamiques, analyse discriminante...

			<ul style="list-style-type: none"> - le choix d'un algorithme de Machine Learning adapté à la problématique - la capacité à entraîner correctement un algorithme de Machine Learning (SVM – Support Vector Machine, réseaux de neurones, cartes de Kohonen, algorithmes génétiques, forêts aléatoires...) - le codage correct d'un algorithme de Machine Learning sur un grand jeu de données
--	--	--	--

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <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>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p>A.4.2 Déploiement d'un modèle d'apprentissage en profondeur (Deep learning) :</p> <ul style="list-style-type: none"> - mise en place d'application de <i>Deep learning</i> - déploiement d'un modèle via une interface de programmation (API) dans le web. - réalisation d'un tableau de bord (<i>Dashboard</i>) pour présenter son travail de modélisation. - présentation auprès des utilisateurs dans un document écrit (note méthodologique) 	<p>C.4.3 Déployer des modèles de <i>Deep Learning</i> en mettant en place des applications de <i>Deep Learning</i>, en utilisant une interface de programmation (API) dans le web, et en prenant en compte la consommation d'énergie dans le choix des outils ou des prestataires, afin de développer des solutions d'intelligence artificielle efficaces et durables.</p> <p>C.4.4 Présenter le travail de modélisation élaboré en réalisant un tableau de bord, en rédigeant un document écrit et en effectuant une présentation orale, afin de garantir une compréhension claire et approfondie de la démarche de modélisation réalisée.</p>	<p>Réalisation de projets :</p> <p>Le projet est tourné vers deux axes. Le premier axe concerne la prédiction. Pour qualifier cette prédiction, des indicateurs seront précisés et discutés (leur pertinence). Un deuxième axe est orienté vers la consommation de ressources et l'impact environnemental.</p> <p>Le candidat devra rendre un rapport écrit en ne négligeant pas l'impact environnemental. En effet, il devra convaincre le lecteur du déploiement de son algorithme de <i>Deep learning</i>.</p>	<p>Le candidat est évalué sur :</p> <ul style="list-style-type: none"> - la bonne identification de la structure du réseau de Deep Learning - l'utilisation correcte de la solution de logiciel - l'identification de l'architecture matérielle sur laquelle faire tourner l'algorithme/ l'apprentissage - la création d'un algorithme de prédiction à partir d'un modèle ayant été convenablement entraîné - l'encapsulation du modèle dans une API web permettant son utilisation à distance (entraînement et prédiction) - la mise en place d'un tableau de bord permettant le pilotage du modèle et l'évaluation de ses capacités d'ajustement et prédiction

<p>et à l'oral de la démarche de modélisation réalisée</p> <ul style="list-style-type: none"> - prise en compte de la consommation d'énergie dans son choix des outils ou prestataires 			<ul style="list-style-type: none"> - la qualité de la documentation (annexes techniques adaptées) et de la présentation de la démarche suivie pour choisir la méthodologie applicable - la qualité de la documentation (annexes techniques adaptées) et de la communication pour une utilisation future et facile par les utilisateurs concernés et potentiellement non spécialistes du domaine - le respect des règles de présentation orale fluidité, clarté, précision, concision, efficacité et respect au temps imparti - le choix du modèle de Deep-Learning représentant un bon compromis entre efficacité et coût énergétique.
---	--	--	--

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <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>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p>A.4.3 Traitements de volumes de données :</p> <ul style="list-style-type: none"> - mise en œuvre avec le logiciel des applications sur des machines standards - traitement de grandes quantités de données - prise en compte de la consommation d'énergie dans son choix des outils ou prestataires 	<p>C.4.5 Optimiser le traitement de volumes de données en sélectionnant et en utilisant un langage de programmation/logiciel dédié gérant des algorithmes, et en optimisant le fonctionnement des algorithmes, afin de permettre à la structure d'opérer plus vite.</p>	<p>Réalisation de projets :</p> <p>Le projet est axé principalement sur l'explication détaillée et argumentée par le candidat du traitement de la masse des données. Il ne négligera pas la prise en compte de l'énergie dépensée pour traiter la masse de données. Il devra mettre en concurrence plusieurs</p>	<p>Le candidat est évalué sur :</p> <ul style="list-style-type: none"> - la connaissance des prestataires écologiquement responsables - le choix d'un langage adapté au volume des données (ex. : choix d'un langage traditionnel ou spécifique aux données massives) - l'optimisation du fonctionnement des algorithmes (parallélisation, utilisation de grilles de calcul, Map Reduce,...)

		<p>prestataires pour résoudre son problème d'énergie.</p> <p>Un rapport écrit est demandé avec la mise en pratique du traitement de la masse des données et les bons arguments. Il expliquera comment il a choisi et motivera ses décisions.</p> <p>La présentation se fera à l'orale devant un jury composé du professeur et deux professionnels extérieurs.</p>	<p>- l'utilisation des outils spécifiques du domaine (ex. : Spark, Hadoop, l'algorithme Map Reduce, AWS...)</p> <p>- le choix d'un modèle représentant un bon compromis entre efficacité et coût énergétique</p>
--	--	---	--

Chaque bloc de compétences est validé de façon autonome. L'acquisition du bloc fait l'objet de remise d'un certificat.

Les compétences qui composent les blocs sont évaluées de manière indépendante les unes aux autres. Soit sous la forme de projet, soit sous la forme de mise en situation professionnelle réelle ou simulée et/ou d'évaluation écrite.

Le candidat doit aussi réaliser et valider une mission en entreprise de 4 mois minimum et valider des évaluations finales transverses : la rédaction et la soutenance de la thèse professionnelle devant un jury composé d'académiques et de professionnels

Pour valider la certification professionnelle complète, le candidat doit valider les 4 blocs de compétences, la mission en entreprise et la thèse professionnelle.

Mission en entreprise :

Période de formation en milieu professionnel d'une durée minimale de quatre mois équivalent temps plein durant laquelle l'étudiant réalise les missions et les tâches qui lui sont confiées.

Thèse professionnelle :

Rédaction d'un mémoire de thèse puis, lors d'une soutenance orale, présentation des données, de la problématique, de l'état de l'art réalisé, des choix méthodologiques et des résultats obtenus. Cette présentation est suivie d'une discussion et d'interaction avec les membres du jury.