

CERTIFICATION PROFESSIONNELLE

Accueil > Trouver une certification > Répertoire national des certifications professionnelles > Expert en infrastructures de données massives

Expert en infrastructures de données massives

Code de la fiche :
RNCP37638

Etat :
Active

 [Télécharger la fiche](#)  [Aide en ligne](#)  [Supplément Europass : FR - EN](#)

L'essentiel

	Nomenclature du niveau de qualification	Niveau 7
	Code(s) NSF	326 : Informatique, traitement de l'information, réseaux de transmission
	Formacode(s)	31036 : Administration base de données 31052 : Data Warehouse 24218 : Transmission numérique 71654 : Logiciel système gestion bases données
	Date d'échéance de l'enregistrement	31-05-2026

Certificateur(s)	Résumé de la certification	Blocs de compétences	Secteur d'activité et type d'emploi	Voie d'ac
Liens avec d'autres certifications professionnelles, certifications ou habilitations		Base légale	Pour plus d'informations	

Certificateur(s)

Nom légal	Siret	Nom commercial	Site internet
SIMPLON.CO	79279132900016	-	https://simplon.co/

Résumé de la certification

Objectifs et contexte de la certification :

Aujourd'hui, la donnée est partout, permettant aux entreprises d'accroître le suivi de leur performance via des indicateurs toujours plus nombreux, d'analyser des comportements clients et par là même d'identifier des opportunités de marché. Les métiers de l'ingénierie de la Data se retrouvent sur l'ensemble du cycle de vie de la donnée, car, dans ce contexte, les entreprises ont besoin de structurer et mettre en oeuvre une stratégie autour des données de l'entreprise. Cela nécessite d'un part de solliciter des professionnels capables d'accompagner cette maîtrise d'ouvrage en conseillant le commanditaire, d'autre part de mobiliser des compétences spécifiques pour piloter la mise en oeuvre de la stratégie, collecter des données, les stocker, les modéliser ou encore les protéger. A l'heure du Big Data et de l'intelligence artificielle, le métier d'expert en infrastructures de données massives connaît un très fort développement car il répond à un enjeu stratégique pour le futur de l'économie et des entreprises. Les besoins en matière de recrutement sont très élevés face à une pénurie de profils adaptés.

Les 4 blocs de compétences de la certification adressent les différentes granularités du métier : la mission de consultance, de conseil et de pilotage ainsi que la préparation des données pour un projet bien défini, jusqu'à la formalisation d'un référentiel de données à l'échelle de l'entreprise, que ce soit dans le cadre de l'informatique décisionnelle (et ses entrepôts de données) ou avec les technologies du Big Data. Ces blocs, en traduisant l'évolution des compétences de ce métier, contribuent à sécuriser le parcours professionnel de cet expert. Ils permettent également de développer des connaissances pointues sur le management des données et une maîtrise fine des techniques et infrastructures associées.

Activités visées :

Encadrer la conception d'un projet data au sein d'une organisation

Coordonner la réalisation d'un projet data au sein d'une organisation

Programmer la collecte de données depuis plusieurs sources pour un projet data au sein d'une organisation

Développer la mise à disposition technique des données collectées pour un projet data au sein d'une organisation

Mettre en place l'entrepôt de données

Assurer la maintenance et l'évolution de l'entrepôt de données

Mettre en place un data-lake

Assurer la gestion de l'accès aux données

Compétences attestées :

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 hypothèses techniques du projet avec le ou les commanditaire(s).

Cartographier les données disponibles en référençant les usages, les sources, les métadonnées et les données afin de valider les hypothèses techniques du projet data.

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.

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.

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 de d'organiser sa mise en oeuvre

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'équipe dans la réussite du projet

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.

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.

Développer des requêtes de type SQL d'extraction des données depuis un 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.

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.

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.

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.

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.

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

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 respectant les modélisations logiques et physiques préalablement établies

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.

Implémenter des variations dans les dimensions de l'entrepôt de données 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.

Modalités d'évaluation :

- Etude de cas
- Mise en situation professionnelle
- Soutenance orale avec le jury de certification
- Rendu de livrables en amont de la soutenance orale avec le jury

Blocs de compétences

RNCP37638BC01 - Piloter la conduite d'un projet data au sein d'une organisation

Liste de compétences	Modalités d'évaluation
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 hypothèses techniques du projet avec le ou les commanditaire(s).	Etude de cas. 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. 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

Liste de compétences	Modalités d'évaluation
Cartographier les données disponibles en référençant les usages, les sources, les métadonnées et les données afin de valider les hypothèses techniques du projet data. 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. 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. 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 de d'organiser sa mise en oeuvre 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'équipe dans la réussite du projet 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.	<i>Livrables</i> : les grilles d'entretien - <i>Évaluation</i> : Correction des grilles et Présentation orale individuelle Mise en situation 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. La mise en situation a pour but de confronter le candidat à un besoin data rencontré par une organisation à laquelle il devra répondre dans son intégralité. Selon la situation choisie, cette évaluation pourra intégrer les éléments d'infrastructure technique associés ou se limiter à la dimension de pilotage du projet. <i>Livrable</i> : rapport professionnel individuel - <i>Évaluation</i> : correction du rapport professionnel et soutenance orale individuelle Jeu de rôle "lancement d'un projet data" Le candidat est mis en situation de documenter et d'animer l'introduction de la réunion de lancement d'un projet data réel ou fictif. 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 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). <i>Livrables</i> : le support de la présentation et les documents associés (par exemple : avant-projet, feuille de route, calendrier, stratégie de communication) <i>Évaluation</i> : correction des livrables et jeu de rôle, simulation de l'introduction de la réunion de lancement.

RNCP37638BC02 - Réaliser la collecte, le stockage et la mise à disposition des données d'un projet data au sein d'une organisation

Liste de compétences	Modalités d'évaluation
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. Développer des requêtes de type SQL d'extraction des données depuis un 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. 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.	Mise en situation 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). Le projet évalué a pour but d'optimiser, 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). <i>Livrable</i> : rapport professionnel individuel - <i>Évaluation</i> : correction du rapport professionnel et soutenance orale individuelle

Liste de compétences	Modalités d'évaluation
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. 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.	

RNCP37638BC03 - Elaborer et maintenir un entrepôt de données (data warehouse)

Liste de compétences	Modalités d'évaluation
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. 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. 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 respectant les modélisations logiques et physiques préalablement établies. 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. Implémenter des variations dans les dimensions de l'entrepôt de données 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.	Mise en situation 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. 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). <i>Livrable</i> : rapport professionnel individuel - <i>Évaluation</i> : correction du rapport professionnel et soutenance orale individuelle Etude de cas L'évaluation doit se faire dans le cadre d'une situation professionnelle fictive réalisée durant le parcours de formation, 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. Le projet évalué s'appuie sur le cadre technique de l'organisation et sur le cadre d'exploitation des données. Lors de cette étude de cas, le candidat 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. <i>Livrable</i> : rapport professionnel individuel - <i>Évaluation</i> : correction du rapport professionnel et échanges de questions réponses individuel

RNCP37638BC04 - Encadrer la collecte massive et la mise à disposition des données issues de l'activité de l'organisation grâce à un data lake

Liste de compétences	Modalités d'évaluation
Concevoir l'architecture du data lake 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. Intégrer les différents composants d'infrastructure du data lake en appliquant la procédure adaptée, afin d'assurer l'acquisition, le stockage et la mise à disposition du catalogue de données 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. 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.	Mise en situation 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. Le projet a pour but de couvrir toutes les étapes de mise en conditions opérationnelles d'un data lake* de son architecture à la gestion de tous les outils permettant son usage. <i>Livrable</i> : rapport professionnel individuel - <i>Évaluation</i> : correction du rapport professionnel et soutenance orale individuelle

Description des modalités d'acquisition de la certification par capitalisation des blocs de compétences et/ou par correspondance :

Chaque bloc de compétence peut être validé de façon autonome, indépendamment des autres. Pour obtenir le titre à finalité professionnelle, le candidat doit valider les quatre blocs de compétences constituant la certification professionnelle.

Secteur d'activité et type d'emploi

Secteurs d'activités :

L'Expert en infrastructures de données massives intervient dans trois contextes professionnels principaux :

au sein d'entreprises de services du numérique (ESN) et d'entreprises de services spécialisées dans l'intégration, en sous-traitance et pour le compte de leurs clients, d'infrastructures data et big data ;

au sein d'entreprises et d'opérateurs qui mettent en œuvre des infrastructures data et/ou big data dans le cadre de leur pilotage d'activité, de leur stratégie ou encore dans le cadre du développement de nouveaux produits et services numériques.

au sein d'équipes R&D d'entreprises et d'opérateurs qui innovent pour imaginer les cas d'usages de la data et/ou de l'IA et développer des prototypes testables avant industrialisation.

Type d'emplois accessibles :

- Expert en infrastructures de données massives
- Data Base Engineer
- Consultant big data

Code(s) ROME :

- M1801 - Administration de systèmes d'information
- M1805 - Études et développement informatique
- M1806 - Conseil et maîtrise d'ouvrage en systèmes d'information
- M1803 - Direction des systèmes d'information
- M1802 - Expertise et support en systèmes d'information

Références juridiques des réglementations d'activité :

L'exercice du métier d'Expert en infrastructures de données massives ne nécessite aucune habilitation particulière. Toutefois, cet expert doit exercer ses activités dans le respect des règlements et recommandations en matière de gestion des données personnelles, d'accessibilité et d'éco-responsabilité dans l'intégration des systèmes de gestion des données.

Dans la réglementation forte qu'il convient de respecter autour du traitement et du stockage appliqué aux données numériques dites "sensibles" (données de santé, données bancaires, données personnelles, etc), on retrouve notamment la Loi française de 1978 Informatique et Libertés, mise à jour en 2018 et en 2019 au regard du règlement général sur la protection des données (RGPD) de 2018 qui encadre une nouvelle fois le traitement et le stockage des données personnelles, informatique et non informatique, à l'échelle Européenne.

Voie d'accès

Le cas échant, prérequis à l'entrée en formation :

Afin d'intégrer le dispositif de formation ou pour suivre une formation, le candidat doit :

- soit être titulaire d'un titre ou d'un diplôme de niveau 5 en informatique avec une expérience professionnelle d'un an minimum dans le secteur des systèmes et réseaux informatiques.
- soit être titulaire d'un titre ou d'un diplôme de niveau 6 en informatique.

Dans le cas où un candidat ne disposerait pas des prérequis définis, il a la possibilité de déposer un dossier de candidature qui sera examiné par une commission en vue d'une admission « exceptionnelle ».

Le cas échant, prérequis à la validation de la certification :

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

Pré-requis distincts pour les blocs de compétences :

Non

Voie d'accès à la certification	Oui	Non	Composition des jurys	Date de dernière modification
Après un parcours de formation sous statut d'élève ou d'étudiant	X		Le jury de certification est composé : d'au moins deux professionnels avec une expérience avérée de 2 ans dans le secteur de la gestion et du traitement des données. L'un d'entre eux a spécifiquement une expérience de Data Engineer, Data Miner ou encore Développeur Data ; la ou les personnes supplémentaires peuvent être un ou des professionnels qui maîtrisent les pratiques de recrutement en entreprise. Au moins 50% des jurés sont extérieurs à l'organisme certificateur et à l'organisme évaluateur. Les membres du jury ne doivent pas entretenir, ou avoir entretenu, un lien professionnel ou personnel avec le candidat. Ils ne doivent pas non plus avoir participé au processus de formation ou à la préparation du candidat à la certification.	05-06-2023
En contrat d'apprentissage	X		Le jury de certification est composé : d'au moins deux professionnels avec une expérience avérée de 2 ans dans le secteur de la gestion et du traitement des données. L'un d'entre eux a spécifiquement une expérience de Data Engineer, Data Miner ou encore Développeur Data ;	05-06-2023

Voie d'accès à la certification	Oui	Non	Composition des jurys	Date de dernière modification
			la ou les personnes supplémentaires peuvent être un ou des professionnels qui maîtrisent les pratiques de recrutement en entreprise. Au moins 50% des jurés sont extérieurs à l'organisme certificateur et à l'organisme évaluateur. Les membres du jury ne doivent pas entretenir, ou avoir entretenu, un lien professionnel ou personnel avec le candidat. Ils ne doivent pas non plus avoir participé au processus de formation ou à la préparation du candidat à la certification.	
Après un parcours de formation continue	X		Le jury de certification est composé : d'au moins deux professionnels avec une expérience avérée de 2 ans dans le secteur de la gestion et du traitement des données. L'un d'entre eux a spécifiquement une expérience de Data Engineer, Data Miner ou encore Développeur Data ; la ou les personnes supplémentaires peuvent être un ou des professionnels qui maîtrisent les pratiques de recrutement en entreprise. Au moins 50% des jurés sont extérieurs à l'organisme certificateur et à l'organisme évaluateur. Les membres du jury ne doivent pas entretenir, ou avoir entretenu, un lien professionnel ou personnel avec le candidat. Ils ne doivent pas non plus avoir participé au processus de formation ou à la préparation du candidat à la certification.	05-06-2023
En contrat de professionnalisation	X		Le jury de certification est composé : d'au moins deux professionnels avec une expérience avérée de 2 ans dans le secteur de la gestion et du traitement des données. L'un d'entre eux a spécifiquement une expérience de Data Engineer, Data Miner ou encore Développeur Data ; la ou les personnes supplémentaires peuvent être un ou des professionnels qui maîtrisent les pratiques de recrutement en entreprise. Au moins 50% des jurés sont extérieurs à l'organisme certificateur et à l'organisme évaluateur. Les membres du jury ne doivent pas entretenir, ou avoir entretenu, un lien professionnel ou personnel avec le candidat. Ils ne doivent pas non plus avoir participé au processus de formation ou à la préparation du candidat à la certification.	05-06-2023
Par candidature individuelle		X	-	-
Par expérience	X		Le jury de certification est composé : d'au moins deux professionnels avec une expérience avérée de 2 ans dans le secteur de la gestion et du traitement des données. L'un d'entre eux a spécifiquement une expérience de Data Engineer, Data Miner ou encore Développeur Data ;	05-06-2023

Voie d'accès à la certification	Oui	Non	Composition des jurys	Date de dernière modification
			la ou les personnes supplémentaires peuvent être un ou des professionnels qui maîtrisent les pratiques de recrutement en entreprise. Au moins 50% des jurés sont extérieurs à l'organisme certificateur et à l'organisme évaluateur. Les membres du jury ne doivent pas entretenir, ou avoir entretenu, un lien professionnel ou personnel avec le candidat. Ils ne doivent pas non plus avoir participé au processus de formation ou à la préparation du candidat à la certification.	

	Oui	Non
Inscrite au cadre de la Nouvelle Calédonie		X
Inscrite au cadre de la Polynésie française		X

Liens avec d'autres certifications professionnelles, certifications ou habilitations

Aucune correspondance

Base légale

Date de décision	31-05-2023
Durée de l'enregistrement en années	3
Date d'échéance de l'enregistrement	31-05-2026
Date de dernière délivrance possible de la certification	31-05-2030

Pour plus d'informations

Statistiques :

Lien internet vers le descriptif de la certification :

<https://simplon.co/formation/certification-data-engineer/745>

Liste des organismes préparant à la certification :

[Liste des organismes préparant à la certification](#)

Référentiel d'activité, de compétences et d'évaluation :

Référentiel d'activité, de compétences et d'évaluation