

EXPERT EN INFRASTRUCTURES DE DONNÉES MASSIVES INTENSIVE + CERTIFICATION MICROSOFT AZURE

SIMPSON GRAND OUEST - NANTES
44300 NANTES

Informations pratiques

Du 10/6/2025 au 09/9/2025

- en centre : 448 heures
- en entreprise : 448 heures
- Temps plein
- Cours de jour

Formation entièrement présentielle

[Le détail de la session sur le site de l'organisme](#)

Simplon Grand Ouest - Nantes

Immeuble Le Newton 50 boulevard
Albert Einstein
44300 Nantes
00-00-00-00-00
fsoler@simplon.co
<https://simplon.co/>

Réunions d'information collective

Non renseigné

Portes ouvertes

Non renseigné

Inscriptions

Ouvert

Pré-requis

Niveau d'entrée : Niveau 6 (Licence, Licence pro, BUT, Titres ou équivalents)

- Etre demandeur·se d'emploi inscrit à France Travail
- Avoir une très forte motivation, à démontrer lors de la candidature
- Titre ou diplôme niv 6 (Bac+3/4) en Informatique
- OU Titre ou diplôme niv 5 (Bac+3) en Informatique + 1 an d'expérience professionnelle
- OU Une expérience professionnelle significative

Lieu de la formation

Simplon Grand Ouest - Nantes
Immeuble Le Newton 50 boulevard
Albert Einstein 44300 Nantes
00-00-00-00-00

fsoler@simplon.co

Financements possibles

- France Travail
- Non renseigné
- Non renseigné

La formation

Objectif Général

Certification

Objectifs

- 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.
- Piloter la conduite d'un projet data au sein d'une organisation
- Réaliser la collecte, le stockage et la mise à disposition des données d'un projet data au sein d'une organisation
- Élaborer et maintenir un entrepôt de données (data warehouse)
- Encadrer la collecte massive et la mise à disposition des données issues de l'activité de l'organisation grâce à un data lake

Blocs de compétences

> RNCP37638BC01 (CPP Piloter la conduite d'un projet data au sein d'une organisation)

• 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.

• 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 Livrables : les grilles d'entretien - Évaluation : 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. Livrable : rapport professionnel individuel - Évaluation : 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). Livrables : le support de la présentation et les documents associés (par exemple : avant-projet, feuille de route, calendrier, stratégie de communication) Évaluation : correction des livrables et jeu de rôle, simulation de l'introduction de la réunion de lancement.

> RNCP37638BC02 (CPP Réaliser la collecte, le stockage et la mise à disposition des données d'un projet data

au sein d'une organisation)

• 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.

• 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). Livrable : rapport professionnel individuel - Évaluation : correction du rapport professionnel et soutenance orale individuelle

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

• 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). Livrable : rapport professionnel individuel - Évaluation : 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. Livrable : rapport professionnel individuel - Évaluation : correction du rapport professionnel et échanges de questions réponses individuel

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

• 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. Livrable : rapport professionnel individuel - Évaluation : correction du rapport professionnel et soutenance orale individuelle

Résultats attendus

Certification Microsoft Azure Certified : [Certification Microsoft AZ-900](#), [Certification Microsoft DP-900](#)

Certification Microsoft Azure Certified: [Certification Microsoft DP - 203](#)

Titre professionnel : RNCP 37638 - Expert·e en infrastructures de données massives (Data Engineer) :

<https://www.francecompetences.fr/recherche/rncp/37638/>

Niveau d'entrée

Niveau 6 (Licence, Licence pro, BUT, Titres ou équivalents)

Organisation pédagogique

- > Modalité d'enseignement :
 - Formation entièrement présentielles

En savoir plus

Exemples de métiers accessibles avec cette formation

- > [Administration de systèmes d'information](#)
- > [Expertise et support en systèmes d'information](#)
- > [Direction des systèmes d'information](#)
- > [Études et développement informatique](#)
- > [Conseil et maîtrise d'ouvrage en systèmes d'information](#)

Source : Cariforef des Pays de la Loire - 250370 - Code établissement : 56632



CHOISIR MON MÉTIER, BONJOUR

Et vous êtes déjà **sur la bonne voix !**

La plateforme téléphonique d'information
sur la formation professionnelle et l'apprentissage
en Pays de la Loire.



**Des chargé.e.s
d'information
à votre écoute**

0 800 200 303

Service & appel
gratuits