

# Parcours certifiant Réaliser la collecte, le stockage et la mise à disposition des données d'un projet en intelligence artificielle

## Bloc de compétences du titre RNCP 37828

Titre RNCP de 28 jours - 196h

Réf : ZRI - Prix 2024 : 13 900€ HT

Le prix pour les dates de sessions 2025 pourra être révisé

Ce parcours de formation représente le premier bloc de compétences du titre RNCP de niveau 6 (Bac +3) « Développeur en intelligence artificielle » reconnu par l'État. Il vous apprendra à automatiser l'extraction de données, développer des requêtes SQL, agréger des données, créer des bases de données conformes au RGPD, et développer des API pour partager les données. Ces compétences sont cruciales dans un monde où la gestion efficace des données est essentielle pour la prise de décision et la conformité réglementaire.

### Ce cycle est composé de :

- RGPD, sensibilisation à la réglementation sur la protection des données (Réf. TPD, 1 jour)
- SQL pour PostgreSQL (Réf. SGS, 3 jours)
- MongoDB, prise en main et développement (Réf. MNO, 3 jours)
- Talend Open Studio, mettre en œuvre l'intégration de données (Réf. TOT, 3 jours)
- Développer un site Web, synthèse pratique (Réf. DSW, 4 jours)
- Web Scraping, récolter des données sur le web avec Python (Réf. WPY, 4 jours)
- Python Data Science, manipuler et visualiser les données (Réf. IYT, 4 jours)
- Python, développer des Web Services REST (Réf. CZT, 2 jours)
- Spark Python, développer des applications pour le big data (Réf. QNC, 3 jours)
- Certification réaliser la collecte, le stockage et la mise à disposition des données d'un projet en intelligence artificielle (Réf. ZDD, ½ journée)

### OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

Automatiser l'extraction de données

Développer des requêtes de type SQL d'extraction des données depuis un SGBD et un système big data

Développer des règles d'agrégation de données issues de différentes sources

Créer une base de données dans le respect du RGPD

Développer une API mettant à disposition le jeu de données

### CERTIFICATION

Le bloc de compétences est validé à travers une 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 basée sur la correction du rapport professionnel et une soutenance orale individuelle

### PARTICIPANTS

Toute personne souhaitant réaliser la collecte, le stockage et la mise à disposition des données d'un projet en intelligence artificielle.

### PRÉREQUIS

Être titulaire d'un diplôme de niveau 5 (Bac +2), avoir des connaissances en programmation objet et en SQL. Si ce n'est pas le cas, être titulaire d'un niveau 4 (BAC) et 3 ans d'expérience en développement d'application, sous réserve de la validation du dossier VAP par le certificateur.

### COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

### MODALITÉS D'ÉVALUATION

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

### MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les stages pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
- À l'issue de chaque stage ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
- Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

### MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

### ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

# LE PROGRAMME

---

dernière mise à jour : 05/2024

## 1) RGPD, sensibilisation à la réglementation sur la protection des données

- Appréhender les concepts de base et les composants de la protection des données.
- Comprendre le contenu de la réglementation générale de la protection des données.
- Connaître le rôle des autorités de protection.
- Connaître le cadre juridique du RGPD et son champ d'application.
- Connaître les différentes étapes de mise en conformité et les démarches à suivre.
- Identifier les outils de mise en conformité.
- Etre en mesure de mettre en place un plan d'action de mise en conformité.

## 2) SQL pour PostgreSQL

- Connaître la vision d'ensemble des SGBD.
- Appréhender la base de données PostgreSQL.
- Créer des requêtes simples et complexes.
- Manipuler des jointures internes et externes.
- Utiliser des expressions régulières.
- Connaître les fonctions à fenêtre.

## 3) MongoDB, prise en main et développement

- Installer le SGBD MongoDB.
- Configurer le SGBD MongoDB.
- Manipuler les objets et les données dans MongoDB.
- Implémenter une application sous MongoDB.
- Améliorer les performances.

## 4) Talend Open Studio, mettre en œuvre l'intégration de données

- Concevoir et développer des jobs dans l'application ETL Talend.
- Optimiser les jobs développés par l'utilisation de contextes et jeux de données.
- Réaliser des transformations plus complexes en utilisant variables, expressions et jointures.
- Exécuter et déboguer un job, tracer les statistiques d'exécution.

## 5) Développer un site Web, synthèse pratique

- Comprendre les fondamentaux du Web.
- Maîtriser l'environnement technique d'un site Web.
- Réaliser un site Web ergonomique, accessible et bien référencé.
- Accéder aux données d'une base relationnelle.
- Administrer un site Web.

## 6) Web Scraping, récolter des données sur le web avec Python

- Maîtriser les bases du langage Python.
- Connaître des éléments de programmation avancée en Python.
- Posséder une vue d'ensemble des principales bibliothèques Python disponibles pour gérer tous types de données de sites.
- Sélectionner la bonne bibliothèque Python pour votre projet de web scraping et être capable de la mettre en œuvre.
- Savoir automatiser des récoltes d'envergure (large-scale web scraping) avec des scripts.

## 7) Python Data Science, manipuler et visualiser les données

- Posséder une vue d'ensemble de l'écosystème scientifique de Python.
- Connaître les bibliothèques scientifiques incontournables pour la science des données.

- Être capable de manipuler des données volumineuses avec Python.
- Comprendre l'intérêt de la datavisualisation.
- Savoir visualiser des données avec Python.

#### 8) Python, développer des Web Services REST

- Appréhender les principes des web services REST.
- Manipuler des données JSON.
- Développer des APIs REST avec Django REST Framework.
- Sécuriser des services Web.

#### 9) Spark Python, développer des applications pour le big data

- Découvrir les concepts fondamentaux de Spark.
- Utiliser le concept des RDD de Spark.
- Exploiter des données avec Spark SQL.
- Effectuer de l'analyse en temps réel avec Spark Streaming.
- Utiliser Spark avec les notebooks Jupyter, manipuler les données avec Pyspark comme avec Pandas.
- Aborder le machine learning avec Spark;

## LES DATES

---

Ce parcours est composé d'un ensemble de modules. Les dates indiquées ci-dessous correspondent aux premières sessions possibles du parcours.

### CLASSE À DISTANCE

2024 : 17 oct., 12 déc.

2025 : 01 avr., 03 juil., 16 oct., 18  
déc.

### PARIS

2024 : 05 déc.

2025 : 25 mars, 26 juin, 09 oct.,  
18 déc.