

Web scraping avec Python

3 jours (21 heures)

Perfectionnement

Présentiel ou classe à distance

à partir de

2 085 € HT par personne (TVA en vigueur) Repas inclus en présentiel

Contenu mis à jour le 25/09/2026. Document téléchargé le 02/10/2026.

Objectifs

A l'issue de cette formation, vous serez capable de :

- Décrire le fonctionnement et les applications des Web scrapers
- Développer un Web scraper avec des features avancées (formulaires, login, JavaScript, API)
- Activer le levier des architectures multi-thread et multi-processus pour du scraping à large échelle.

Les compétences visées par cette formation font l'objet d'une évaluation.

À qui s'adresse cette formation ?

Développeurs, Data scientists, Data analysts, ingénieurs Data, intégrateurs, chefs de projets, consultants BI/Big Data.

Prérequis

Avoir une expérience en développement, de préférence avec Python.

Programme

1 Pendant votre formation

Web scraping avec Python

Jour 1

Les fondamentaux du Web scraping

- Rappels sur Internet : modèle à sept couches, HTML, CSS, JavaScript, outils du développeur Web
- Aspects réglementaires et éthiques du Web scraping : marques déposées, copyrights, brevets...
- Applications du Web scraping : marketing, ventes, entraînement d'IA, développement de produits, recherche académique...
- Bien démarrer un projet de Web scraping

Exemple de travaux pratiques (à titre indicatif)

- Création d'un premier Web scraper avec Jupyter et la librairie BeautifulSoup

Analyse HTML avancée

- Les objets de BeautifulSoup
- Naviguer à travers des structures arborescentes
- Gérer les expressions régulières
- Accéder à des attributs
- Expressions Lambda
- Gérer différentes mises en page de sites
- Structurer des crawlers : crawlers à base de recherche, à base de liens, par type de page

Exemple de travaux pratiques (à titre indicatif)

- Développement d'un Web crawler mono-domaine ou multi-domaine

Introduction à Scrapy

- Exemple de scraper Scrapy
- Utilisation des règles
- Création d'items
- Items d'output
- Pipeline d'items
- Gestion des logs avec Scrapy

Exemple de travaux pratiques (à titre indicatif)

- Création d'un Web scraper avec Scrapy

Jour 2

Scraping avancé

- Lecture de documents texte, CSV, PDF, MS Word
- Nettoyage de texte avec Pandas
- Travailler avec du texte normalisé
- Lire et résumer du langage naturel avec NLTK

- Analyse statistique et lexicographique avec NLTK

Exemple de travaux pratiques (à titre indicatif)

- Création d'un Web scraper avancé capable de traiter le langage naturel

Crawling à travers des formulaires et interfaces de login

- Présentation de la librairie Requests
- Exemple de soumission d'un formulaire
- Gérer les boutons radio, les cases à cocher, et autres inputs
- Soumission de fichiers et images
- Gérer les logins et cookies

Exemple de travaux pratiques (à titre indicatif)

- Amélioration du Web scraper en lui ajoutant des features liées aux formulaires et logins

Scraping sur du code JavaScript

- Rappels sur JavaScript, Ajax, et le HTML dynamique
- Exécuter des scripts JavaScript dans Python avec Selenium
- Autres Selenium WebDrivers
- Gérer les redirections

Jour 3

Scraping sur des API

- Rappels sur les API
- Parsing de JSON
- Documenter des API non-documentées

Gérer les CAPTCHA avec la reconnaissance d'images et de texte

- Bibliothèques disponibles pour la gestion des CAPTCHA : Pillow, Tesseract, NumPy
- Exemple de reconnaissance d'images et de texte
- Entraînement de Tesseract pour la lecture de CAPTCHA

Exemple de travaux pratiques (à titre indicatif)

- Amélioration du Web scraper en lui ajoutant des features de gestion de CAPTCHA

Parallélisation du Web scraping

- Rappel sur les processus et threads
- Exemple de crawling multi-thread
- Exemple de crawling multi-processus

Exemple de travaux pratiques (à titre indicatif)

- Amélioration du Web scraper en lui ajoutant des features de gestion de code Javascript et d'API

Financement

Plusieurs dispositifs de financement sont accessibles via les OPCO (Opérateurs de Compétences), organismes agréés par le ministère du Travail dont le rôle est d'accompagner, collecter et gérer les contributions des entreprises au titre du financement de la formation professionnelle.

Pour plus d'information, une équipe de gestionnaires SKOLAE Formation spécialisée vous accompagne dans le choix de vos formations et la gestion administrative.

[Nous contacter pour en savoir plus →](#)

Modalités pédagogiques

Formation délivrée en présentiel ou distanciel* (blended-learning, e-learning, classe virtuelle, présentiel à distance).

Le formateur alterne entre méthode** démonstrative, interrogative et active (via des travaux pratiques et/ou des mises en situation).

Variables suivant les formations, les moyens pédagogiques mis en oeuvre sont :

- Ordinateurs Mac ou PC (sauf pour certains cours de l'offre Management), connexion internet fibre, tableau blanc ou paperboard, vidéoprojecteur ou écran tactile interactif (pour le distanciel)
- Environnements de formation installés sur les postes de travail ou en ligne
- Supports de cours et exercices

En cas de formation intra sur site externe à M2I, le client s'assure et s'engage également à avoir toutes les ressources matérielles pédagogiques nécessaires (équipements informatiques...) au bon déroulement de l'action de formation visée conformément aux prérequis indiqués dans le programme de formation communiqué.

* nous consulter pour la faisabilité en distanciel

** ratio variable selon le cours suivi

[Accessibilité aux personnes en situation de handicap](#)

Détails de la session

Du 18 novembre 2026 au 20 novembre 2026 (3 jours / 21 heures)

Mulhouse

Du 18 novembre 2026 au 20 novembre 2026

Mulhouse

- 18 novembre 2026 9h00 – 16h00
- 19 novembre 2026 9h00 – 16h00
- 20 novembre 2026 9h00 – 16h00