

Scraping : extraire et gérer des données

Cette formation vous apporte les compétences nécessaires pour mener à bien des tâches de scraping web et d'ingestion de données : rappels de HTML et CSS, architecture d'un site Web, scraping de données BeautifulSoup et Selenium, ETL Python et ingestion de données pour alimenter un Data Lake. **Pré-requis** : HTML5, CSS, Python.



Modalité : à distance

Formule au choix : Live Training+ *ou bien*
 Classe virtuelle Teams ou Zoom

Durée totale : 21 H (3 jours)

PLAN DETAILLE

Introduction HTML et CSS : #1

- Le client serveur web
- Les différents langages du web
- HTML : caractéristiques
- Structure d'une page
- Principe des styles CSS
- Balises et règles CSS principales

Les bases du Webscraping : #2

- Librairies Python
- Méthodologie du Web scraping
- L'Objet BeautifulSoup
- Les différents analyseurs
- BS: explorer un document
- Utiliser le module re
- Rechercher, filtrer un objet BS
- find(), find_all(), select(), ...
- Scraping avec navigation
- TP : webscraping avec BS

Mini projet Webscraping: #3

- Scraper des données météo
- Inspecter la page web
- Automatisation avec Python
- Scraper un ensemble de liens
- Rendu des résultats
- Script final

Webscraping avec Selenium: #4

- Introduction
- Inspecter les données à récolter
- Webscraping avec Selenium
- Les librairies nécessaires
- Mise en oeuvre du scraping
- Interagir avec la page web
- Travaux pratiques

Aller plus loin : #5

- Utiliser un proxy
- Utiliser IA pour le scraping
- Web scraping via le No code