



Développeur Python

Web scraping avec Python

3 jours (21h00) | ★★★★★ 4,6/5 | BIG-COLL | Évaluation qualitative de fin de stage |
Formation délivrée en présentiel ou distanciel

Formations Informatique › Langages et développement › Développeur Python

Contenu mis à jour le 13/10/2023. Document téléchargé le 23/05/2024.

Objectifs de formation

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

- Réaliser du scraping de données
- Faire les actions d'ingestion nécessaires pour alimenter un Data Lake.

Modalités, méthodes et moyens 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

Prérequis

Avoir des connaissances de base sur Python et en algorithmique.

Public concerné

Développeurs, intégrateurs, chefs de projets, consultants BI.

Cette formation :

- Est animée par un consultant-formateur dont les compétences techniques, professionnelles et pédagogiques ont été validées par des diplômes et/ou testées et approuvées par l'éditeur et/ou par M2i Formation
- Bénéficie d'un suivi de son exécution par une feuille de présence élargie par demi-journée par les stagiaires et le formateur.

Programme

Les bases du langage Python

- Les caractéristiques du langage Python
- Pourquoi choisir Python pour l'analyse de données ?
- Types de bases
- Les instructions de bases
- Les procédures et fonctions

L'ingestion avec Python

- Utiliser la librairie Pandas pour manipuler les données
- Introduction du concept de DataFrame
- Les structures :
 - Interrogation
 - Indexation
- Traitement de "données manquantes"
- Fusion de DataFrames
- Manipulation des dates
- Application de mesures statistiques variées sur les DataFrames
- Bonne compréhension des problèmes d'échelle de mesure, de normalisation
- Création de métriques d'analyse

Scraping de données

- Qu'est-ce que le scraping ?
- Définition du scraping et de ses différents niveaux de difficulté sur plusieurs supports
 - Depuis le Web
 - Depuis du papier
 - Depuis des PDF
- Exemples de projets réalisés grâce au scraping
- L'architecture d'Internet
- Qu'est-ce qu'un "client" ? Qu'est-ce qu'un "serveur" ? Pourquoi est-ce important ?
- Comment HTTP et HTML impactent-ils nos scrapers ?
- Qu'est-ce qu'une balise HTML ? Un attribut ?
- Comment identifier certains éléments avec une "class" ou un "id" ?

Python comme solution ETL

- Les formats de données structurées : CSV, flux XML et JSON
- Lecture et écriture de fichiers
- Exploitation des données de fichiers de différentes sources
- Fonctions d'accès et de chargement de données en blocs de lignes
- Outils spécifiquement dédiés au scraping :
 - BeautifulSoup
 - CSS Select

Mise en oeuvre d'un scraper

- Un scraper simple (requêtes GET, pages séquencées)
- Identifier la stratégie à adopter pour naviguer sur le site
- Coder le scraper
- Un scraper complexe : envoyer des données à un site Internet pour obtenir des résultats plus complexes
- Qu'est ce qu'une requête POST et une requête GET ?
- Parcourir un site pour trouver les données
- Identifier la stratégie à adopter
- Coder le scraper

Le contenu de ce programme peut faire l'objet d'adaptation selon les niveaux, prérequis et besoins des apprenants.

Modalités d'évaluation des acquis

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation

Accessibilité de la formation

Le groupe M2i s'engage pour faciliter l'accessibilité de ses formations. Les détails de l'accueil des personnes en situation de handicap sont consultables sur la page Accueil et Handicap.

Modalités et délais d'accès à la formation

Les formations M2i sont disponibles selon les modalités proposées sur la page programme. Les inscriptions sont possibles jusqu'à 48 heures ouvrées avant le début de la formation. Dans le cas d'une formation financée par le CPF, ce délai est porté à 11 jours ouvrés.