Python, extraction de données

Lien : https://innov-systems.com/formation/python-extraction-de-donnees

DURÉE
4 jours (28h)

RÉFÉRENCE
DEV89

CATÉGORIE
Python, C# et .NET

OBJECTIFS DE LA FORMATION

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

- Etre capable de réaliser du scraping de données
- Faire les actions d'ingestion nécessaires pour alimenter un Data Lake

POUR QUI ?

- Développeurs
- Intégrateurs
- Chefs de projets

Innov Systems

Programme détaillé

1 / Rappels sur les bases du langage Python

- Présentation des possibilités du langage Python
- Python pour l'analyse de données : pourquoi ce choix ?
- Les variables & les typages
- Les différents types d'instructions
- Les procédures et fonctions

2 / L'ingestion avec Python

- Analyse de données avec Pandas
- DataFrame : Présentation du concept
- Les structures : Interrogation, Indexation
- Gérer des données manquantes
- Fusion de DataFrames
- Manipuler les dates
- Mesures statistiques variées sur les DataFrames
- Comprendre les problèmes d'échelle de mesure, de normalisation
- Créer les métriques d'analyse

3 / Scraping de données

- Définition du scraping
- Présentation de ses différents niveaux de difficulté sur plusieurs supports :
- Depuis le Web
- Depuis du papier
- Depuis des PDF
- Présentation de quelques projets réalisés grâce au scraping
- L'architecture d'Internet

Innov Systems

- L'impact d'HTTP et d'HTML sur les scrapers
- Identifier certains éléments avec une "class" ou un "id"

4 / Python comme solution ETL

- Différentes formats de données structurées : CSV, flux XML et JSON
- Lecture et écriture de fichiers
- Traiter des données de fichiers de différentes sources
- Fonctions d'accès et de chargement de données en blocs de lignes
- Outils pour le scraping : Beautiful Soup, CSS Select

5 / Mettre en oeuvre un scraper

- Un scraper simple (requêtes GET, pages séquencées)
- Identification de la stratégie à adopter pour naviguer sur le site
- Codage du scraper
- Un scraper complexe : envoyer des données à un site Internet pour obtenir des résultats plus complexes
- Requête POST et GET
- Parcourir un site pour trouver les données
- Présentation de la stratégie à adopter

Approche pédagogique

- Support Ecrit et Projection
- Exposés Interactifs, Podcasts et Vidéos
- Brainstorming et Jeux de Rôle
- Cas Pratiques et Labs inclus pour leur impact opérationnel
- Test de Validation des Acquis des Connaissances

Inov Systems

Prochaines dates programmées

12 au 15 Jan. 2027	Présentiel - Casablanca
09 au 12 Mars 2027	Présentiel - Casablanca
04 au 07 Mai 2027	Présentiel - Casablanca
29 Juin au 02 Juil. 2027	Présentiel - Casablanca
31 Août au 03 Sep. 2027	Présentiel - Casablanca

Autres dates possibles sur demande. Contactez-nous pour organiser une session intra-entreprise.

Réservation & Renseignements

Téléphone : +212 522 247 210
Email : contact@innov-systems.com
Web : https//www.innov-systems.com

Innov Systems — Tous droits réservés

Innov Systems