

PYTHON POUR L'INGÉNIERIE DES DONNÉES

Durée : 1 jour (7 heures)

CONNAISSANCES PREALABLES

- Connaissances de base en programmation
- Notions de Python recommandées
- Connaissances élémentaires en SQL
- Compréhension générale des bases de données et des flux de données.

PROFIL DES STAGIAIRES

- Data Engineers débutants
- Développeurs Python souhaitant évoluer vers la Data
- Data Analysts
- Administrateurs de bases de données
- Consultants Data
- Ingénieurs DevOps impliqués dans les pipelines de données
- Architectes Data débutants
- Toute personne souhaitant découvrir l'utilisation de Python dans les projets Data Engineering.

OBJECTIFS

À l'issue de cette formation, les participants seront capables de :

- Comprendre le rôle de Python dans les architectures Data modernes
- Manipuler des données avec Python
- Lire et transformer des données provenant de différentes sources
- Automatiser des traitements de données simples
- Construire un pipeline de traitement de données élémentaire
- Utiliser les principales bibliothèques Python dédiées au Data Engineering
- Comprendre l'intégration de Python dans les plateformes Data modernes.

CERTIFICATION PREPAREE

Aucune

METHODES PEDAGOGIQUES

- Présentations illustrées par des cas concrets
- Démonstrations techniques
- Travaux pratiques guidés
- Exercices de manipulation de données
- Étude de cas fil rouge
- Échanges autour des bonnes pratiques Data Engineering.

FORMATEUR

- Consultant expert Data Engineering, Python et architectures Data modernes, intervenant sur des projets Big Data, Cloud Data et plateformes analytiques..

METHODE D'EVALUATION DES ACQUIS

- Quiz de validation des connaissances
- Exercices pratiques
- Atelier de construction de pipeline
- Étude de cas
- Évaluation continue par le formateur.

CONTENU DU COURS

Module 1 – Python dans l'écosystème Data Engineering (1h)

Objectifs

- Comprendre le rôle de Python dans les architectures de données
- Identifier les principaux cas d'usage

Contenu

- Le métier de Data Engineer
- Python comme langage de référence du Data Engineering
- Les principales missions :
 - Collecte des données
 - Transformation
 - Orchestration
 - Automatisation
- Panorama des bibliothèques Data
- Place de Python dans les architectures modernes

Mise en pratique

- Analyse d'une architecture Data moderne
- Identification des composants utilisant Python

Module 2 – Manipuler et transformer les données avec Python (2h)

Objectifs

- Charger et transformer des données
- Automatiser des traitements simples

Contenu

- Lecture et écriture de fichiers :
 - CSV
 - JSON
 - Excel
- Manipulation avec Pandas
- Nettoyage des données
- Filtrage
- Agrégation
- Transformation des données
- Contrôle de la qualité des données

Mise en pratique

- Chargement d'un jeu de données
- Nettoyage et transformation
- Création d'indicateurs simples

Module 3 – Connecter Python aux bases de données (1h30)**Objectifs**

- Lire et écrire des données depuis une base de données
- Automatiser les échanges de données

Contenu

- Connexion aux bases relationnelles
- Utilisation de SQLAlchemy
- Lecture de données SQL avec Pandas
- Export de données
- Gestion des flux d'alimentation

Mise en pratique

- Connexion à une base PostgreSQL ou MySQL
- Extraction de données
- Chargement automatisé de données

Module 4 – Construire un pipeline de données simple (1h30)**Objectifs**

- Comprendre la logique des pipelines Data
- Mettre en œuvre un traitement automatisé

Contenu

- Définition d'un pipeline Data
- Extraction
- Transformation
- Chargement (ETL)
- Gestion des erreurs
- Journalisation des traitements
- Bonnes pratiques de développement

Mise en pratique

- Construction d'un pipeline ETL simple
- Automatisation d'un traitement de données
- Contrôle des résultats

Module 5 – Python et les plateformes Data modernes (0h30)**Objectifs**

- Comprendre l'intégration de Python dans les écosystèmes Data

Contenu

- Introduction à Apache Spark avec PySpark
- Présentation d'Airflow
- Introduction à Kafka
- Data Engineering dans le Cloud
- Cas d'usage Databricks

Mise en pratique

- Démonstration d'un pipeline moderne
- Analyse d'une architecture Data Engineering

Module 6 – Atelier de synthèse : réaliser un mini pipeline Data (0h30)

Objectifs

- Mettre en œuvre les concepts abordés
- Réaliser un traitement de données complet

Contenu

- Lecture de données
- Transformation
- Contrôle qualité
- Export des résultats

Mise en pratique

- Réalisation d'un mini-projet fil rouge
- Restitution des résultats
- Débriefing collectif

Notre référent handicap se tient à votre disposition au [01.71.19.70.30](tel:01.71.19.70.30) ou par mail à <mailto:referent.handicap@edugroupe.com> pour recueillir vos éventuels besoins d'aménagements, afin de vous offrir la meilleure expérience possible.