

Big Data - Python avancé pour Data Scientist

Référence : **DELY010**

Durée : **4 jours (28 heures)**

Certification : **Aucune**

Connaissances préalables

- Bonne connaissance de la programmation Python

Profil des stagiaires

- Développeurs en Python, Développeurs de logiciels, Programmeurs, Data analysts, Data scientists

Objectifs

- Savoir utiliser les principaux outils de traitement et d'analyse de données pour Python
- Savoir appliquer les pratiques optimales en matière de nettoyage et de préparation des données avant l'analyse
- Être capable d'extraire des données d'un fichier
- Comprendre les mécanismes d'interconnexion aux bases de données
- Comprendre les principaux outils de traitement et d'analyse de données pour Python

Certification préparée

- Aucune

Méthodes pédagogiques

- 6 à 12 personnes maximum par cours, 1 poste de travail par stagiaire
- Remise d'une documentation pédagogique papier ou numérique pendant le stage
- La formation est constituée d'apports théoriques, d'exercices pratiques et de réflexions

Formateur·rice

- Consultant-Formateur expert Bigdata

Méthodes d'évaluation des acquis

- Auto-évaluation des acquis par le stagiaire via un questionnaire
- Attestation des compétences acquises envoyée au stagiaire
- Attestation de fin de stage adressée avec la facture

Contenu du cours

1. Positionnement Python dans l'analyse de données

- Besoins des data-scientists : calculs, analyse d'images, machine learning, interface avec les bases de données
- Apports de python : grande variété d'outils, expertise dans le domaine du calcul scientifique
- Tour d'horizon des outils : pandas, pyarrow, agate, bokeh, scikit-learn, pybrain, tensorflow, keras, mxnet, caffe, Pytorch

2. Calculs et graphiques

- NumPy : Base du calcul sur des tableaux
- SciPy : Scientific Tools for Python, couche scientifique
- Manipulation de tableaux, fonctions mathématiques
- Représentation graphique avec basemap et matplotlib
- Exemple de Travaux Pratiques : Mise en oeuvre de SciPy/NumPy (manipulation d'images, détection de contours)

3. Être capable d'extraire des données d'un fichier

- Pandas : manipulation de tables de données. Notion de dataframe
- Manipulation de données relationnelles
- Tableaux avec Pandas : indexation, opérations, algèbre relationnelle
- Stockage dans des fichiers : CSV, JSON
- Comparaison et performances Pandas / pyarrow / NumPy
- Exemple de Travaux Pratiques : Construction d'ETL de base entre json et csvkaggl.com

4. Comprendre les mécanismes d'interconnexion aux bases de données

- Définitions : pilotes, connexions, curseurs, CRUD, transactions
- Les pilotes : postgresql, mysql, mariadb, ... Présentation de sql-alchemy
- Opérations : gestion du curseur, chargement de données, insertion et modification d'enregistrements
- Exemple de Travaux Pratiques : Mise en oeuvre avec postgresql. Construction d'ETL SQL/json

5. Comprendre les principaux outils de traitement et d'analyse de données pour Python

- Présentation des outils d'apprentissage Python : scikit-learn, pybrain, TensorFlow/keras, mxnet, caffe
- Exemple de Travaux Pratiques : Mise en oeuvre de scikit-learn

6. Créer des sélections et des classements dans de grands volumes de données pour dégager des tendances

- Présentation de pyspark
- Machine learning et deep learning : les solutions Python
- TensorFlow : principe de fonctionnement, plateformes supportées, distribution

7. Sites de références data-sciences

- Ressources d'apprentissage, datasets, modèles de données pré-entraînés, etc.
- Présentation de : kaggle.com, data-puzzles.com, huggingface.co

8. Optimisation des développements

- Tour d'horizon des outils actuels et futurs
- Jupyter notebook, Aide à la vérification de code, respect des recommandations PEP8 : exemples avec pydocstyle, Pylint, Black
- Analyse et production de code informatique avec une IA
- Génération de code avec OpenAI : démonstrations ChatGPT, apports, bonnes pratiques.
- Exemple de Travaux Pratiques : Utilisation de la génération de code et de snippets Python avec ChatGPT

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