

Aller au-delà de pandas : manipuler de grandes données avec Polars, DuckDB et Arrow

Description

Cette formation de 2 jours permet d'aller au-delà de pandas pour la manipulation de grandes données en Python, sans recourir à une infrastructure Big Data.

Les participants découvrent et mettent en pratique Polars, DuckDB et Apache Arrow afin de construire des pipelines data performants, lisibles et reproductibles.

Destinée aux data analysts et data scientists maîtrisant pandas, la formation privilégie des cas concrets et des volumes de données réalistes, en lien direct avec les usages professionnels.

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data analysts, data scientists, développeurs Python orientés data.

Prérequis :

- Bonne maîtrise de Python et pratique régulière de pandas (DataFrame, filtres, groupby, jointures).

Objectifs pédagogiques

- Comprendre les limites de pandas pour les volumes de données importants
- Manipuler efficacement de grandes données sans infrastructure Big Data
- Utiliser Polars pour des transformations performantes
- Exploiter DuckDB pour l'analyse SQL de fichiers volumineux
- Comprendre le rôle d'Apache Arrow dans l'écosystème data
- Construire des pipelines data performants et reproductibles

Programme détaillé

- Polars : le DataFrame moderne
 - Limites de pandas et comparaison avec Polars
 - Fondamentaux de Polars et lecture de données volumineuses
 - Manipulation avancée, jointures et agrégations
 - Lazy execution et optimisations
 - Atelier : migration d'un pipeline pandas vers Polars
- DuckDB, Arrow et pipelines modernes
 - Apache Arrow : format columnar et interopérabilité
 - DuckDB : SQL analytique embarqué
 - Workflows hybrides Polars / DuckDB

Stat4decision

37-39 avenue Ledru Rollin Paris 75012

Tel. +33172254082 | E-mail : info@stat4decision.com | www.stat4decision.com

Numéro SIRET: 81048985600015 | Numéro de déclaration d'activité: 11755352275 (auprès du préfet de région de: Ile de France)

Organisme de formation certifié Qualiopi pour ses actions de formation

- Structuration de pipelines data performants
- Atelier fil rouge : pipeline complet et analyse des performances

Organisation de la formation

Moyens pédagogiques et techniques

- Supports de cours numériques, notebooks Jupyter, jeux de données volumineux, environnement Python (Polars, DuckDB, pandas). Alternance d'apports théoriques, démonstrations et ateliers pratiques.

Dispositif de suivi de l'exécution et d'évaluation des objectifs de la formation

- Évaluation continue à travers les ateliers pratiques, questions orales et validation des exercices réalisés.

Stat4decision

37-39 avenue Ledru Rollin Paris 75012

Tel. +33172254082 | E-mail : info@stat4decision.com | www.stat4decision.com

Numéro SIRET: 81048985600015 | Numéro de déclaration d'activité: 11755352275 (auprès du préfet de région de: Ile de France)

Organisme de formation certifié Qualiopi pour ses actions de formation