Pipelines de données avec Python, Airflow et dbt

Concevoir et automatiser des chaînes de traitement de données fiables : collecte, transformation et chargement dans un entrepôt, orchestration avec Airflow, modélisation avec dbt, tests et supervision.

En pratique

DURÉE
21 h (3 jours)

MODALITÉS
Présentiel · Distanciel · Intra-entreprise · Inter-entreprises

LIEU
Orléans (45), dans vos locaux ou en classe virtuelle

EFFECTIF
3 à 8 participants

TARIF
Inter : 1 790 € HT par participant ; intra : sur devis

DÉLAI D'ACCÈS
30 jours ouvrés entre la demande et le début de la formation, sous réserve d'un nombre suffisant de participants

VALIDATION
Attestation de fin de formation

NIVEAU
Développeur

Objectifs pédagogiques

À l'issue de la formation, le stagiaire sera capable de :

- Concevoir l'architecture d'une chaîne de données (ETL ou ELT)
- Collecter des données depuis des fichiers, des API et des bases
- Modéliser et transformer les données dans un entrepôt avec dbt
- Orchestrer, planifier et superviser les traitements avec Airflow
- Tester la qualité des données et documenter la chaîne

Public visé

- Data engineers, data analysts et développeurs
- Équipes qui veulent remplacer des consolidations manuelles par des traitements automatisés

Prérequis

- Pratique de Python et de SQL
- Notions de Docker souhaitables

Programme

Jour 1 — Architecture et collecte

Architectures de données

- Base, entrepôt, data lake ; ETL et ELT
- Panorama : PostgreSQL, Snowflake, outils cloud et open source

Collecter

- Extraction depuis fichiers, API et bases de données en Python
- Chargements complets et incrémentaux

Jour 2 — Transformer avec dbt

Modéliser l'entrepôt

- Couches brute, intermédiaire et restituée
- Modélisation dimensionnelle

dbt

- Modèles SQL, dépendances, matérialisations
- Tests de qualité, documentation, lignage

Jour 3 — Orchestrer et superviser

Airflow

- DAG, tâches, dépendances, planification
- Reprises sur erreur, alertes

Exploiter

- Conteneurisation, environnements, secrets
- Supervision et coûts
- Projet final

Méthodes pédagogiques

- Possibilité de travailler sur les données des participants (en intra, après accord)
- Supports de cours, codes source et corrigés remis aux participants
- Formation dispensée par un consultant data et IA qui conçoit et met en production ce qu'il enseigne

Modalités d'évaluation

- Questionnaire de positionnement en début de formation
- Exercices pratiques corrigés tout au long de la formation
- Projet final : chaîne de données complète, de la source au tableau de bord
- Questionnaire de satisfaction à chaud

Accessibilité

Formation accessible aux personnes en situation de handicap. Contactez-nous en amont (contact@dcid-consulting.fr) afin d'étudier ensemble les adaptations nécessaires.

Financement

DCID Consulting n'est pas certifié Qualiopi à ce jour : cette formation est financée directement par l'employeur et n'est éligible ni au CPF ni à une prise en charge par un OPCO.

Contact et inscription

contact@dcid-consulting.fr · 02 38 25 33 90 ·dcid-consulting.fr/formations/pipelines-donnees-airflow-dbt/