FORMATION · DATA MANAGEMENT

Apache Airflow – Orchestrer vos pipelines

Durée de la formation : 3 jours
Tarif : 2 190 € HT par participant
Réf : BI170
Délai d'accès : 2 à 4 semaines
Niveau : Intermédiaire

Prérequis

Connaissance du langage Python (fonctions, modules, environnements virtuels). Notions de ligne de commande Linux et de bases de données.

Public concerné

Data engineers, développeurs, équipes DevOps

Objectifs pédagogiques

- Comprendre l'architecture d'Airflow et ses composants
- Concevoir des DAG maintenables
- Maîtriser la planification et la gestion des erreurs
- Préparer le déploiement et l'exploitation

Objectifs opérationnels

- Installer un environnement Airflow local
- Écrire des DAG avec les opérateurs et la TaskFlow API
- Échanger des données entre tâches (XCom) et utiliser le templating
- Configurer connexions, variables et reprises automatiques
- Tester un DAG et l'intégrer dans une chaîne d'intégration continue

Méthodes mobilisées et modalités d'évaluation

Apports théoriques, démonstrations et nombreux travaux pratiques sur des cas concrets. À l'issue de chaque module, le formateur évaluera chacun des participants en fonction des cas pratiques et exercices effectués.

Apache Airflow – Orchestrer vos pipelines

Apache Airflow est l'orchestrateur de référence pour planifier, enchaîner et surveiller des traitements de données écrits en Python.

En trois jours, vous apprenez à concevoir des DAG lisibles et robustes, à gérer les dépendances, les reprises sur erreur et la planification, puis à préparer leur passage en production.

Les travaux pratiques s'appuient sur des pipelines réalistes : ingestion de fichiers, appels d'API, chargement dans un entrepôt de données.

Inscription

Contactez-nous au 01 86 95 71 60 ou à l'adresse formation@adbi.fr pour connaître les dates de notre prochaine session. Inscription au plus tard 10 jours avant le début de la session.

Accessibilité

Nos salles de formation sont accessibles aux personnes à mobilité réduite. N'hésitez pas à nous informer si vous avez des besoins particuliers et s'il y a des besoins d'aménagements spécifiques lors de votre inscription.

Modes de participation

Nos formations sont proposées sous 3 formats :

- Chez nous, près de la place d'Italie (Paris 13e)
- Chez vous, dans vos locaux
- À distance, en classe virtuelle

Attestation

Une attestation de fin de formation sera remise à tous les participants à l'issue de leur parcours.

Programme :

Jour 1

- **Les fondamentaux**
  - Orchestration de données : besoins et alternatives
  - Architecture : planificateur, exécuteur, serveur web, base de métadonnées
  - Installation avec Docker Compose et visite de l'interface
- **Premiers DAG**
  - Structure d'un DAG, tâches et dépendances
  - Opérateurs courants : Python, Bash, SQL
  - TaskFlow API et décorateurs
  - Travaux pratiques : un premier pipeline d'ingestion

Jour 2

- **Planification**
  - Intervalles, dates logiques, rattrapage (catchup) et backfill
  - Planification déclenchée par les données (datasets)
  - Capteurs (sensors) et attentes efficaces
- **Données et paramétrage**
  - XCom : bonnes pratiques et limites
  - Templating Jinja et macros
  - Connexions, variables et gestion des secrets
- **Robustesse**
  - Reprises, délais, règles de déclenchement
  - Alertes et notifications
  - Idempotence : rejouer un traitement sans effet de bord

Jour 3

- **Organiser le code**
  - Structure d'un dépôt de DAG, modules partagés
  - Task Groups et DAG dynamiques (mapping de tâches)
- **Tester et déployer**
  - Tests unitaires et tests d'intégrité des DAG
  - Intégration continue et déploiement
  - Choisir un exécuteur : local, Celery, Kubernetes
- **Exploiter**
  - Journaux, métriques et supervision
  - Atelier de synthèse : pipeline complet de l'ingestion à l'entrepôt