FORMATION · DATA MANAGEMENT

Apache Airflow – Avancé

Durée de la formation : 2 jours
Tarif : 1 790 € HT par participant
Réf : BI180
Délai d'accès : 2 à 4 semaines
Niveau : Avancé

Prérequis

Avoir suivi « Apache Airflow - Orchestrer vos pipelines » ou écrire des DAG en production depuis plusieurs mois. Bonne maîtrise de Python.

Public concerné

Data engineers, architectes data, équipes plateforme

Objectifs pédagogiques

- Concevoir des DAG génériques et dynamiques
- Étendre Airflow avec ses propres opérateurs et hooks
- Exécuter des tâches dans des environnements isolés
- Exploiter et faire évoluer une plateforme Airflow

Objectifs opérationnels

- Générer des DAG à partir de configurations
- Développer un opérateur et un hook réutilisables
- Lancer des tâches dans des conteneurs ou sur Kubernetes
- Mettre en place la supervision et le dimensionnement
- Préparer une montée de version d'Airflow

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 – Avancé

Vos pipelines Airflow tournent déjà ; il s'agit maintenant de les rendre plus génériques, plus fiables et plus simples à exploiter à grande échelle.

Cette formation approfondit la génération dynamique de DAG, la création d'opérateurs, l'intégration avec les environnements d'exécution externes et l'exploitation d'une plateforme Airflow partagée.

Elle s'appuie sur des cas rencontrés par nos équipes sur des plateformes de production.

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

- **DAG avancés**
  - Mapping dynamique de tâches et génération de DAG par configuration
  - Branchements, regroupements et sous-flux
  - Dépendances entre DAG et planification par les données
- **Étendre Airflow**
  - Écrire un hook et un opérateur personnalisés
  - Fournisseurs (providers) et plugins
  - Travaux pratiques : un opérateur réutilisable pour votre entrepôt
- **Environnements d'exécution**
  - Isoler les dépendances : environnements virtuels, conteneurs
  - Exécuter des tâches sur Kubernetes
  - Déclencher des traitements externes (dbt, Spark, services cloud)

Jour 2

- **Exploitation à l'échelle**
  - Exécuteurs et dimensionnement : parallélisme, pools, files
  - Supervision : métriques, journaux centralisés, alertes
  - Sécurité : rôles, secrets et backends de secrets
- **Qualité et cycle de vie**
  - Tests avancés et environnements de recette
  - Montées de version : préparation et vérifications
  - Bonnes pratiques d'une plateforme partagée entre équipes
- **Atelier de synthèse**
  - Revue et refonte d'un projet Airflow existant