

Formation Apache Airflow : Automatisation et orchestration de flux de données

■ Durée :	2 jours (14 heures)
■ Tarifs inter-entreprise :	1 775,00 € HT (standard) 1 420,00 € HT (remisé)
■ Public :	Développeurs, DevOps, Administrateurs système ou toute personne souhaitant automatiser des workflows
■ Pré-requis :	Notions de base en Python - Maitrise des concepts de workflows et pipelines de données
■ Objectifs :	Comprendre et configurer les workflows avec Apache Airflow - Créer, optimiser et superviser des DAGs complexes - Intégrer Airflow dans un environnement de production sécurisé - Utiliser les meilleures pratiques pour garantir des workflows robustes et maintenables
■ Modalités pédagogiques, techniques et d'encadrement :	• Formation synchrone en présentiel et distanciel. • Méthodologie basée sur l'Active Learning : 75 % de pratique minimum. • Un PC par participant en présentiel, possibilité de mettre à disposition en bureau à distance un PC et l'environnement adéquat. • Un formateur expert.
■ Modalités d'évaluation :	• Définition des besoins et attentes des apprenants en amont de la formation. • Auto-positionnement à l'entrée et la sortie de la formation. • Suivi continu par les formateurs durant les ateliers pratiques. • Évaluation à chaud de l'adéquation au besoin professionnel des apprenants le dernier jour de formation.
■ Sanction :	Attestation de fin de formation mentionnant le résultat des acquis
■ Référence :	OUT102361-F

■ Note de satisfaction des participants :	Pas de données disponibles
■ Contacts :	commercial@dawan.fr - 09 72 37 73 73
■ Modalités d'accès :	Possibilité de faire un devis en ligne (www.dawan.fr, moncompteformation.gouv.fr, maformation.fr, etc.) ou en appelant au standard.
■ Délais d'accès :	Variable selon le type de financement.
■ Accessibilité :	Si vous êtes en situation de handicap, nous sommes en mesure de vous accueillir, n'hésitez pas à nous contacter à referenthandicap@dawan.fr, nous étudierons ensemble vos besoins

Introduction à Apache Airflow

Historique et rôle d'Apache Airflow dans l'orchestration de workflows
 Comparaison avec d'autres orchestrateurs : Apache NiFi, Luigi, etc.
 Cas d'utilisation courants (ETL, automatisation des processus, etc.)

Architecture d'Airflow

Composants principaux : Scheduler, Webserver, Metadata Database, Worker
 Flux de travail interne (DAGs et Tasks)
 Installation et configuration de base (Airflow CLI et configuration initiale)

DAGs : Directed Acyclic Graphs

Structure d'un DAG : définition et configuration
 Comprendre les opérateurs : BashOperator, PythonOperator, DummyOperator
 Programmation d'un premier DAG simple

Exécution des tâches et dépendances

Gestion des dépendances entre tâches
 Définition de schémas d'exécution (sequential, parallel, etc.)
 Gestion du scheduling avec cron et intervalles de temps (TimeDelta, catchup, etc.)

Monitoring et interface web d'Airflow

Exploration des fonctionnalités clés : Gantt chart, log des tâches, exécution manuelle
 Résolution des erreurs et gestion des retards

Atelier Pratique : Création d'un DAG avec des opérateurs multiples

Personnalisation et opérateurs avancés

Développement d'opérateurs personnalisés (PythonOperator avancé)

Utilisation d'opérateurs préconstruits : BranchPythonOperator, SubDagOperator

Introduction aux Sensors pour surveiller des événements (FileSensor, ExternalTaskSensor)

XComs et communication entre tâches

Transfert de données entre tâches via XComs

Meilleures pratiques pour limiter les erreurs

Extensions et intégrations

Connexions avec des bases de données (Postgres, MySQL, etc.)

Intégration avec des outils cloud (AWS, GCP, Azure)

Airflow en production

Configuration avancée : pools, SLA et gestion des priorités

Mise en place de la haute disponibilité

Stratégies de monitoring et alertes

Bonnes pratiques de développement et de gestion des workflows

Structuration des DAGs pour la maintenabilité

Debugging et optimisation des workflows

Gestion des versions et migration

Ateliers pratiques :

- Mise en œuvre d'un projet complet : orchestration d'un pipeline multi-étapes
- Optimisation d'un workflow existant