

NOUVEAU POSSIBLE EN INTRA

ORCHESTRATION DE PIPELINES AVEC APACHE AIRFLOW

La multiplication des traitements et des flux de données rend nécessaire l'automatisation et la supervision de leur enchaînement. Apache Airflow permet d'orchestrer des workflows complexes, de planifier les traitements, de gérer leurs dépendances et d'en assurer le suivi.



1 360 € HT



2 JOURS (14 H.)



NOUS CONSULTER
POUR LES DATES DE SESSION

COMPÉTENCE PRINCIPALE VISÉE

Concevoir, orchestrer et fiabiliser des pipelines de données avec Apache Airflow afin d'automatiser leur exécution et leur exploitation en production

LES + DE LA FORMATION

60 % du temps est consacré à des travaux pratiques, permettant aux apprenants de mettre immédiatement en œuvre les notions abordées et de construire progressivement un pipeline complet avec Apache Airflow.



PUBLIC

- Data Engineers, Data Scientists et développeurs
- Ingénieurs DevOps et architectes Data



PRÉREQUIS

- Connaissances de base en Python et en environnement Linux
- Connaissances des principes fondamentaux des bases de données et du langage SQL



OBJECTIFS PÉDAGOGIQUES

- Concevoir et planifier des workflows de données sous forme de DAGs
- Orchestrer et intégrer des traitements en gérant leurs dépendances, leur parallélisation et leurs interactions avec des systèmes externes
- Superviser et fiabiliser les pipelines afin d'assurer leur bonne exécution et leur reprise en cas d'incident.

CONTENU

PARTIE 1 - Introduction à Apache Airflow

- Enjeux de l'orchestration des traitements
- Positionnement et cas d'usage d'Airflow
- Concepts fondamentaux : DAG, Task, DAG Run, Task Instance
- Architecture : Scheduler, DAG Processor, Executor/Workers, Metadata Database, interface Web

TRAVAUX PRATIQUES - Prise en main d'Airflow

- Découverte de l'environnement
- Création d'un premier DAG
- Correction d'une erreur de parsing
- Déclenchement et suivi d'une exécution
- Consultation et analyse des logs

PARTIE 2 - Concevoir et planifier un DAG

- Structure d'un DAG
- Operators et tâches
- Dépendances entre tâches
- Bonnes pratiques : atomicité et idempotence
- Scheduling et expressions Cron
- Logical Date et Data Interval
- Catchup et Backfill

TRAVAUX PRATIQUES - Concevoir et planifier un workflow

- Création d'un DAG composé de plusieurs tâches
- Mise en œuvre de dépendances
- Configuration de différentes planifications
- Manipulation des Data Intervals
- Expérimentation du Catchup et du Backfill

PARTIE 3 - Construire des workflows complexes

- Traitements séquentiels et parallèles
- Fan-Out / Fan-In
- Contexte d'exécution
- XCom et échanges entre tâches
- Branching et Trigger Rules

TRAVAUX PRATIQUES - Construire un pipeline dynamique

- Parallélisation de traitements
- Exploitation du contexte d'exécution
- Échange d'informations avec XCom
- Sélection conditionnelle d'un traitement

PARTIE 4 - Intégrer Airflow aux systèmes externes

- Principe des Providers
- Operators et Hooks
- Airflow Connections

97,2%
de clients
satisfaits*

* enquête réalisée auprès
de nos clients en
septembre 2025

- Connexion aux bases de données
- Gestion des informations de connexion

TRAVAUX PRATIQUES - Intégration avec PostgreSQL

- Configuration d'une Connexion
- Utilisation du Provider PostgreSQL
- Exécution de requêtes SQL depuis un DAG
- Manipulation d'une connexion avec un Hook
- Chargement de données dans PostgreSQL

PARTIE 5 - Déclencher des traitements à partir des données

- Principe des Sensors
- Attente d'une ressource externe
- Assets et orchestration orientée données
- Relations Producer / Consumer entre DAGs

TRAVAUX PRATIQUES - Construire un pipeline piloté par les données

- Attente de la disponibilité d'un fichier
- Déclenchement d'un traitement
- Publication d'un Asset
- Déclenchement automatique d'un DAG consommateur

PARTIE 6 - Fiabiliser et superviser les pipelines

- Gestion des erreurs
- Retries et Retry Delay
- Timeouts
- Idempotence et reprise après incident
- Analyse des logs
- Relance des traitements
- Bonnes pratiques d'exploitation

TRAVAUX PRATIQUES - Fiabiliser un pipeline

- Simulation d'une erreur
- Configuration des retries
- Analyse et diagnostic à partir des logs
- Correction et reprise du traitement

ÉQUIPE PÉDAGOGIQUE

L'équipe enseignante est composée de plusieurs enseignants-chercheurs de l'INSA Lyon, membres du laboratoire LIRIS, avec une activité de recherche reconnue en sciences des données. Cette équipe dispose d'une forte expertise technico-scientifique, obtenue via des nombreux projets de recherche partenariale, avec des grands groupes français mais également avec des start-up, PME ou ETI.

MOYENS ET MÉTHODES PÉDAGOGIQUES

Alternance d'exposés, illustrations, études de cas, travaux pratiques et échanges avec des experts du domaine Un support de cours sera remis à chacun des participants

PROCHAINE SESSION

L'ouverture de la session est conditionnée par un nombre minimum de participants. Nous consulter pour d'autres dates.

ÉVALUATION ET RÉSULTATS

Évaluation des acquis de la formation

Évaluation des acquis des apprenants par auto-examen. 94.3% des apprenants ont acquis la compétence principale visée. (sur 177 apprenants évalués sur cette thématique depuis 2020)

Évaluation de la satisfaction des participants en fin de formation (Niveau 1 KIRKPATRICK)

4.3 par les participants. (sur 199 participants ayant suivi une formation dans la thématique depuis 2020)



RENSEIGNEMENTS ET INSCRIPTION

Tel : +33 (0)4 72 43 83 93

Fax : +33 (0)4 72 44 34 24

mail : formation@insavalor.fr

Préinscription sur formation.insavalor.fr

Accueil des personnes en situation de handicap nécessitant un besoin spécifique d'accompagnement : nous contacter à l'inscription. Nos locaux sont accessibles aux personnes à mobilité réduite.

Actualisée le 01/10/2026