

Formation Apache Airflow

3 jours (21 heures)



Présentation

En participant à cette formation, vous allez acquérir une compréhension approfondie des concepts fondamentaux d'Airflow et de ses fonctionnalités avancées.

Vous apprendrez à créer, planifier et surveiller des workflows avec une efficacité accrue, en utilisant des techniques simples et intuitives.

Notre approche pédagogique interactive vous permettra d'explorer de manière pratique chaque aspect d'Airflow, vous donnant ainsi les compétences nécessaires pour optimiser vos processus.

Cette formation vous fournira les outils et les connaissances dont vous avez besoin pour maximiser son potentiel dans votre organisation.

Objectifs

- Comprendre l'historique et le concept d'Airflow ainsi que ses applications courantes
- Maîtriser l'architecture et les composants clés d'Airflow
- Savoir configurer et déployer Airflow, y compris le passage à l'échelle
- Utiliser l'interface utilisateur et la CLI pour la navigation et la gestion des workflows
- Planifier, exécuter et surveiller les workflows avec gestion des erreurs et modifications
- Personnaliser les workflows avec des plugins, hooks, et macros pour une meilleure efficacité

Public visé

Profils techniques : Data Engineer, MLOps, BI Engineer, Data Analyst, Data Scientist, DevOps, Software Engineer, etc.

Pré-requis

- Connaissance de base en développement et en gestion des données
- Familiarité avec le langage de programmation Python
- Expérience préalable avec Docker et la gestion de conteneurs est un plus



Programme détaillé

Jour 1

Introduction à Airflow

- Historique d'Airflow
- Qu'est-ce que Airflow ?
- Pourquoi utiliser Airflow ?
- Applications courantes d'Airflow

Architecture d'Airflow et composants clés

- Les composants d'Airflow : DAG, tâches, opérateurs, workflows, etc.
- L'architecture d'Airflow : webserver, scheduler, workers, etc.
- Relations entre les différents composants

Configuration et installation d'Airflow

- Déploiement d'Airflow
- Les différents types d'executors
- Passage à l'échelle d'un cluster Airflow
- Vérification du bon fonctionnement

TP : Utiliser docker-compose pour déployer une instance personnalisée

Présentation de l'interface utilisateur

- Navigation dans l'interface
- Affichage des DAGs et de leur statut
- Historique de l'exécution des workflows
- Gestion des erreurs et des tâches manquantes
- Présentation de la CLI

TP : Navigation dans l'interface et utilisation de la CLI

Jour 2

Planification et exécution des workflows

- Planification des workflows
- Exécution manuelle ou automatique
- Utilisation des datasets
- Gestion des erreurs
- Suivi de l'exécution et modification

TP : Planifier l'exécution de dags par plusieurs moyens

Gestion des variables et des connexions

- Utilisation des variables
- Gestion des connexions aux bases de données et aux services externes
- Sécurisation des informations sensibles



TP : Construire un dag qui utilise les variables et les connexions

Personnalisation des workflows avec les plugins, les hooks, les macros

- Ne pas réinventer la roue
- Utilisation des hooks pour personnaliser les tâches
- Utilisation des macros pour réutiliser du code et des variables

TP : Utilisation des macros pour construire une dag factory

Jour 3

Gestion de la donnée avec les pools, les queues et les XComs

- Échanger de la donnée entre les tâches
- Limiter le parallélisme avec les pools
- Définir des priorités entre les tâches

TP : Échanger de la donnée entre les tâches et limiter le nombre d'appel concurrents à une base de données

Monitoring et sécurité

- Gestion des SLA
- Configuration des alertes
- Gestion des rôles et des utilisateurs
- Monitorer son cluster Airflow avec Grafana
- Appliquer des politiques à notre cluster

TP : Mise en place d'une suite Prometheus Grafana pour monitorer le cluster et mise en place de cluster policies

Conclusion et perspectives

- Perspectives d'utilisation d'Airflow
- Présentation des solutions concurrentes
- Réponses aux questions et ressources supplémentaires

Organisation

- La formation alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions et de cas pratiques.
- Support de la formation distribué au format numérique à tous les participants.

Prix : 1590€ HT / participant

Tarif dégressif appliqué à partir de 5 participants, dans la limite de 25 participants maximum par session de formation.

Contact : +33 (0)7 45 41 66 05 / contact@ai-training.fr

