

Formation Airflow

Déployer automatiquement des projets data

La formation sur Airflow est conçue pour aider les participants à maîtriser les concepts clés et les fonctionnalités de ce puissant système de gestion de workflow. Les participants apprendront à utiliser Airflow pour créer, planifier et surveiller des workflows efficacement, en utilisant des techniques simples et intuitives. La formation couvre également les concepts fondamentaux de la gestion de workflow, tels que la dépendance entre les tâches et la gestion des erreurs, ce qui permettra aux participants de tirer le meilleur parti d'Airflow pour gérer leurs processus complexes. La formation est conçue pour offrir une expérience pratique, avec des activités et des démonstrations qui permettront aux participants de mettre en pratique les connaissances acquises au fil de la formation. En fin de compte, cette formation est un excellent moyen de se familiariser avec Airflow et de développer les compétences nécessaires pour utiliser efficacement ce système de gestion de workflow dans un environnement professionnel.

Durée

2 jours

Objectifs pédagogiques

- Connaître l'architecture et les composants Airflow
- S'approprier les principales fonctionnalités pour créer, planifier et surveiller des workflows

Public

Ingénieurs Big Data, data scientists, data analysts...

Prérequis

Connaissances en programmation, bases de données, systèmes de fichiers et ligne de commande (scripts shell).

Programme de formation

Introduction à la formation Airflow

Qu'est-ce que Airflow

Pourquoi utiliser Airflow

Applications courantes d'Airflow

Principaux avantages et inconvénients d'Airflow

Architecture d'Airflow et composants clés

Les composants clés d'Airflow : DAG, tâches, opérateurs, workflows, etc.

L'architecture d'Airflow : webserver, scheduler, executors, etc.

Explication des relations entre les différents composants

Exemples d'utilisation de chaque composant

Configuration et installation d'Airflow

- Prérequis pour installer Airflow
- Procédure d'installation d'Airflow
- Configuration de l'environnement d'exécution d'Airflow
- Démarrage et vérification du bon fonctionnement d'Airflow

Présentation de l'interface utilisateur d'Airflow

- Navigation dans l'interface utilisateur d'Airflow
- Affichage des DAGs et de leur statut
- Vérification de l'historique de l'exécution des workflows
- Gestion des erreurs et des tâches manquantes
- Utilisation de la console de commande d'Airflow

Création de DAG et définition de ses tâches

- Création d'un nouveau DAG
- Définition des tâches et de leur ordre d'exécution
- Utilisation des différents types d'opérateurs disponibles (BashOperator, PythonOperator, etc.)
- Paramétrage des tâches (arguments, retours, etc.)
- Enregistrement et activation du DAG

Utilisation des opérateurs courants

- Les différents types d'opérateurs disponibles
- Utilisation des opérateurs les plus couramment utilisés (BashOperator, PythonOperator, etc.)
- Exemples concrets d'utilisation de ces opérateurs
- Paramétrage des opérateurs pour les adapter à vos besoins

Gestion des dépendances entre tâches

- Définition des dépendances entre les tâches
- Utilisation des différents types de dépendances (upstream, downstream, etc.)
- Résolution des dépendances et planification des tâches
- Prise en compte des erreurs et des tâches manquantes dans les dépendances

Planification et exécution des workflows

- Définition de la planification des workflows (heure de départ, fréquence, etc.)
- Exécution manuelle ou automatique des workflows
- Gestion des erreurs et des tâches manquantes
- Suivi de l'exécution et modification des workflows en cours d'exécution

Gestion des variables d'environnement et des connexions

- Utilisation des variables d'environnement pour paramétrer les tâches
- Gestion des connexions aux bases de données et aux services externes
- Création, modification et suppression des connexions
- Sécurisation des informations sensibles (mots de passe, clés API, etc.)

Personnalisation des workflows avec les hooks et les macros

- Utilisation des hooks pour personnaliser les tâches (pré/post-tâche)
- Utilisation des macros pour réutiliser du code et des variables
- Création, modification et suppression des hooks et des macros
- Exemples concrets d'utilisation des hooks et des macros

Monitoring

- Monitoring des workflows en temps réel
- Configuration des alertes pour les erreurs et les tâches manquantes
- Utilisation de l'interface utilisateur pour visualiser les alertes
- Envoi d'alertes par email, SMS, notifications push, etc.

Conclusion et perspectives

- Récapitulation des principaux points de la formation
- Présentation des perspectives d'utilisation d'Airflow
- Discussion sur les évolutions futures d'Airflow
- Réponse aux questions et présentation de ressources supplémentaires

Moyens et méthodes pédagogiques

- La formation alterne entre présentations des concepts théoriques et mises en application à travers d'ateliers et exercices pratiques.
- Les participants bénéficient des retours d'expérience terrains du formateur ou de la formatrice
- Un support de cours numérique est fourni aux stagiaires

Modalités d'évaluation

- **En amont de la session de formation**, un questionnaire d'auto-positionnement est remis aux participants, afin qu'ils situent leurs connaissances et compétences déjà acquises par rapport au thème de la formation.
- **En cours de formation**, l'évaluation se fait sous forme d'ateliers, exercices et travaux pratiques de validation, de retour d'observation et/ou de partage d'expérience.
- **En fin de session**, le formateur évalue les compétences et connaissances acquises par les apprenants grâce à un questionnaire reprenant les mêmes éléments que l'auto-positionnement, permettant ainsi une analyse détaillée de leur progression.