

Formation Apache Airflow

Durée

3 jours (21 heures)

S'inscrire

Présentation

Apache Airflow est un outil d'orchestration vous permettant de superviser, monitorer et planifier des centaines de pipelines de données plus ou moins complexes le tout par le biais de la programmation. Fini le format XML difficile à lire et à écrire, avec Airflow tous vos pipelines sont codés en Python vous offrant une flexibilité inégalée.

Airflow a rapidement gagné en popularité du fait de sa simplicité d'utilisation grâce à son interface utilisateur et de sa capacité à rendre vos workflows dynamiques. Airflow est aussi extrêmement modulaire vous permettant de créer vos propres plug-ins, opérateurs et exécuteurs afin de coller au mieux à votre architecture big data. Sachez enfin que vous pouvez tout aussi bien profiter des créations la communauté qui est très active.

Dans cette formation vous allez découvrir les différents concepts de bases qui composent Airflow, qu'est-ce qu'un DAG, une tâche, un exécuteur, un opérateur, etc. Nous ferons un rapide tour de l'interface utilisateur ainsi que de l'interface en ligne de commande afin de très rapidement développer notre premier workflow interagissant notamment avec HDFS et Hive. Les concepts avancés seront aussi abordés et mis en pratiques à travers de multiples exemples et nous verrons aussi comment rendre Airflow scalable à travers l'utilisation de Celery et RabbitMQ. Un programme chargé donc, mais qui vous permettra d'appréhender sereinement Airflow et de pouvoir orchestrer vos pipelines de données de manière simple et robuste.

Comme dans toutes nos formations nous utiliserons la dernière version stable des logiciels [Apache Airflow 2.1](#).

Objectifs

- Découvrir les fonctionnalités d'Apache Airflow.
- Savoir concevoir, planifier et monitorer ses workflows.
- Connaître les mécanismes avancés de l'outil.
- Savoir comment scaler Airflow afin de gérer un nombre illimité de tâches à exécuter.
- Connaître les best practices et limitations de Airflow.

Public visé

Ingénieur Big data ,Développeur, Tech Lead, Architecte Technique

Pré-requis

Connaissance en langage Python

Pour aller plus loin

Nos formations sur le Big data

- [Big data : l'état de l'art](#)
- Notre formation [TensorFlow](#)
- Notre formation [Kafka](#)

Programme de notre Formation Apache Airflow

Introduction

- Les objectifs
- Organisation de la formation
- Qui suis-je ?
- Présentation de l'environnement de développement (VM ou installation des outils)

Découverte de Apache Airflow

- Pourquoi l'utiliser et pour quels besoins.
- Qu'est-ce que Airflow et quels sont ses composants.
- Comprendre son fonctionnement et comment les workflows sont planifiés.
- [Pratique] Installer Airflow
- [Pratique] Découverte de l'interface utilisateur
- [Pratique] Découverte de l'interface en ligne de commande
- Quiz

Coder son premier Workflow avec Airflow

- Définir ce qu'est un DAG (Directed Acyclic Graph)
- [Pratique] Coder notre premier DAG
- Les operators : SensorOperator, ActionOperator et TransferOperator
- [Pratique] Ajout des trois principaux opérateurs dans notre DAG
- Les dépendances entre opérateurs et la bitshift composition
- [Pratique] Définition de l'ordre d'exécution des tâches dans notre DAG
- Comment le planificateur de Airflow fonctionne précisément.
- Définition d'un Workflow

- Quiz

Databases et exécuteurs

- Les exécuteurs séquentiels avec SQLite
- Les exécuteurs locaux avec PostgreSQL
- [Pratique] Configurer un DAG avec les exécuteurs locaux et PostgreSQL
- Les exécuteurs Celery avec PostgreSQL et RabbitMQ
- [Pratique] Configurer un DAG avec les exécuteurs Celery, PostgreSQL et RabbitMQ
- Quiz

Utilisation des concepts avancés de Airflow

- Minimiser les patterns répétitifs d'un DAG avec les SubDags
- [Pratique] Simplifier un DAG avec les SubDags
- Comment interagir avec des sources externes en utilisant les Hooks
- [Pratique] Récupérer les résultats d'une base de données PostgreSQL avec un Hook
- Comment échanger des données entre les tâches en utilisant les XCOMS
- [Pratique] Échanger des données en utilisant les XCOMS
- Comment exécuter différentes tâches en fonction de critères dynamiques
- [Pratique] Créer nos première tâches conditionnelles
- Contrôler les tâches contenu dans un DAG avec les SLAs
- [Pratique] Définition des SLAs pour nos tâches
- Comment utiliser et sauvegarder nos données avec les variables.
- [Pratique] Manipuler les variables depuis notre DAG
- Jinja Templating, Macros et Variables
- [Pratique] Ajouter des données au run time avec les Macros et les templates
- Quiz

Créer un plugin Airflow pour interagir avec Elasticsearch et PostgreSQL

- Ajouter des fonctionnalités dans Apache Airflow
- [Pratique] Créer un Hook pour interagir avec Elasticsearch
- [Pratique] Créer un TransferOperator pour transférer des données depuis PostgreSQL vers ElasticSearch
- [Pratique] Ajouter une vue à l'interface utilisateur de Airflow
- Quiz

Administrer Airflow: Sécurité, RBAC, Metriques et Logging

- Sécuriser ses connexions et données sur Airflow
- [Pratique] Utilisation de librairie Crypto pour sécuriser Airflow
- Utiliser Airflow en SSL derrière un proxy inversé
- [Pratique] Configuration du proxy inversé pour être utilisé avec Airflow
- La création d'utilisateurs et l'authentification
- [Pratique] Revue des différents moyens d'authentification

- Découverte du RBAC (Role-based Access Control)
- [Pratique] Création de différents utilisateurs et rôles
- Découverte du système de log et sa configuration
- [Pratique] Configuration des logs
- Monitorer Airflow et metrics à surveiller
- [Pratique] Utilisation de InfluxDB, Telegraph et Grafana pour le monitoring the Airflow.
- Quiz

Sociétés concernées

Cette formation s'adresse aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée.

Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

Validation

À la fin de la session, un questionnaire à choix multiple permet de vérifier l'acquisition correcte des compétences.

Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.