+33 (0)1 48 10 09 46
formation@ambient-it.net
140 Avenue Jean Lolive 93500 Pantin - France

Mis à jour le 23/09/2024

S'inscrire

Formation Apache Airflow

3 jours (21 heures)

Présentation

Notre formation Apache Airflow vous permettra de superviser, monitorer et planifier des centaines de pipelines. Ecrit en Python, Airflow est une alternative aux outils de workflow utilisant des langages descriptif (XML , JSON , YAML …) difficile à lire et à écrire.

À la fin de cette formation, vous pourrez rendre vos workflows dynamiques, automatisés et scalables. Vous apprendrez également à créer vos propres plug-ins, opérateurs et exécuteurs afin de coller au mieux à votre architecture big data.

Vous découvrirez les différents concepts qui composent Airflow, qu’est-ce qu’un DAG, une tâche, un exécuteur, un opérateur, etc. Nous ferons un tour de l’interface utilisateur ainsi que de l’interface en ligne de commande afin de très rapidement développer notre premier workflow interagissant avec différents systèmes de données.

De plus, vous apprendrez également les concepts avancés. Nous verrons avec de la mise en pratique et à travers de multiples exemples : comment rendre Airflow scalable à travers l'utilisation de différents Executors. Un programme qui vous permettra d'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 2.10.

Objectifs

- Découvrir les fonctionnalités d’Apache Airflow
- Savoir concevoir, planifier et monitorer ses workflows
- Connaître le fonctionnement interne de Airflow
- Savoir comment scaler Airflow afin de gérer un nombre conséquent de pipelines
- Connaître les best practices et limitations de Airflow

Public visé

- Ingénieur Big data
- Développeur
- Tech Lead
- Architecte Technique

Pré-requis

- Connaissance en langage Python
- Tester Mes Connaissances

Pré-requis logiciel

- La dernière version de Docker installée
- La dernière version de Docker Compose installée
- La dernière version de Python installée
- La dernière version d'Airflow installée
- Minimum 8GO de RAM et un SSD

Recommandations de lecture avant et après la formation

- Un article informatif et bien structuré sur les bonnes pratiques à utiliser sur Apache Airflow
- Un tutoriel pour automatiser vos pipelines de données, idéal pour renforcer votre productivité !
- 5 minutes de lecture pour comprendre les tests sur Apache Airflow
- Un article génial pour utiliser Airflow de la bonne manière et éviter les erreurs les plus fréquentes

Programme de notre Formation Apache Airflow

Introduction

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

Découverte de Apache Airflow

- Pourquoi l’utiliser et pour quels besoins
- Les concurrents de Airflow
- L’architecture de Airflow
- Comprendre le fonctionnement de Airflow et comment les workflows sont planifiés
- [Pratique] Installer Airflow
- [Pratique] Découverte de l’interface utilisateur
- [Pratique] Découverte de la CLI et HTTP REST API

- Quiz

Coder son premier Workflow avec Airflow

- Définir ce qu’est un DAG (Directed Acyclic Graph)
- [Pratique] Coder notre premier DAG
- Les différent type d’operators
- [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

Exécuteurs et leurs backend

- Les différents exécuteurs et backend
- Leurs avantages et faiblesses
- [Pratique] Configurer un LocalExcutor et run un DAG
- [Pratique] Configurer un CeleryExecutor et run un DAG
- Quiz

Les concepts dynamiques de Airflow

- Minimiser les patterns répétitifs d’un DAG avec les SubDags
- [Pratique] Simplifier un DAG avec les SubDags
- Les XCOMS
- [Pratique] Échanger des données en utilisant les XCOMS
- Comment exécuter différentes tâches en fonction de critères dynamiques
- [Pratique] Créer nos premières tâches conditionnelles
- Contrôler les tâches contenu dans un DAG avec les SLAs
- [Pratique] Définition des SLAs pour nos tâches
- Les airflow variables
- [Pratique] utiliser les variables depuis notre DAG
- Les airflow connections
- [Pratique] utiliser les connections depuis notre DAG
- Jinja Templating, Macros et Variables
- [Pratique] Ajouter des données au run time avec les Macros et les templates
- Les Pools
- Quiz

Les Airflow providers pour interagir avec le reste du monde

- Les providers existant et leur installation
- Savoir choisir les opérateurs à utiliser en fonction des executors et du scaling voulu
- Personnaliser un opérateur existant
- [Pratique] Installer un provider et en personnaliser un opérateur
- Quiz

Administrer Airflow: Sécurité, RBAC, Metriques et Logging

- Sécuriser 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
- [Pratique] Monitorer Airflow et metrics à surveiller
- Quiz

MODULE COMPLÉMENTAIRE (+1 JOUR) : APACHE AIRFLOW AVANCÉ

Création d'un plugin Airflow pour intéragir 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

Administration de Airflow Avancé

- [Pratique] Utilisation de librairie Crypto pour sécuriser Airflow
- [Pratique] Utilisation de InfluxDB et Telegraph pour le monitoring the Airflow
- Quiz

Sociétés concernées

Cette formation s’adresse à la fois aux particuliers ainsi qu’aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

Positionnement à l'entrée en formation

Le positionnement à l'entrée en formation respecte les critères qualité Qualiopi. Dès son inscription définitive, l'apprenant reçoit un questionnaire d’auto-évaluation nous permettant d’apprécier son niveau estimé sur différents types de technologies, ses attentes et objectifs personnels quant à la formation à venir, dans les limites imposées par le format sélectionné. Ce questionnaire nous permet également d’anticiper certaines difficultés de connexion ou de sécurité interne en entreprise (intraentreprise ou classe virtuelle) qui pourraient être problématiques pour le suivi et le bon déroulement de la session de formation.

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 multiples permet de vérifier l'acquisition correcte des compétences.

Sanction

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

Organisme de formation enregistré sous le numéro 11 75 54743 75. Cet enregistrement ne vaut pas agrément de l'État.
© Ambient IT 2015-2024. Tous droits réservés. Paris, France - Suisse - Belgique - Luxembourg