

Mis à jour le 14/03/2024

S'inscrire

Formation Apache Airflow avancé

1 jour (7 heures)

Présentation

Notre formation Apache Airflow avancé vous permettra de maîtriser l'un des outils leader du MLOps. Ainsi, vous pourrez gérer efficacement vos workflows data.

Notre cours débutera en mettant l'emphase sur la génération dynamique de DAGs et ses actions associées. Vous apprendrez les fondamentaux du scheduling et les techniques avancées de templating avec Jinja.

Nous poursuivrons avec l'apprentissage des opérateurs personnalisés, la manipulation des hooks et l'intégration des APIs. Les divers opérateurs tels [PythonVirtualenvOperator](#) ou [KubernetesPodOperator](#) n'auront plus de secret pour vous.

Enfin, nous terminerons par la présentation d'études de cas, du ModelOps ainsi que les bonnes pratiques à connaître.

Comme pour toutes nos formations, nous vous enseignerons la dernière version stable : [Apache Airflow 2.8](#).

Objectifs

- Créer des DAGs dynamiques en utilisant des configurations externes
- Personnaliser les opérateurs et les capteurs pour des besoins spécifiques
- Intégrer des tâches avec des environnements externes tels que PythonVirtualenv, Docker, et Kubernetes
- Appliquer les meilleures pratiques en ModelOps et en gestion des ressources pour l'ingénierie des données
- Comprendre les limitations d'Airflow

Public visé

- Ingénieur Big data
- Développeur
- Tech Lead
- Architecte Technique
- Data engineer
- MLOps
- Data Scientist

Pré-requis

- Connaissance du langage Python
- 2 ans d'expérience minimum en data science
- Compétences de base sur Airflow
- De préférence, avoir suivi notre [formation Airflow](#)

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

Programme de notre formation Apache Airflow avancé

Modèles avancés de DAG

- Génération dynamique de DAG
 - Groupes de tâches
 - Déclenchement de l'exécution du DAG
 - Mappage dynamique des tâches
 - Opérateur différé
 - Installation
- Pourquoi ne pas utiliser les SubDAGs ?
- Datasets et data aware scheduling
 - Rappel sur le scheduling classique et les timetables
- Templating avancé avec Jinja
- Générer des DAGs dynamiquement basés sur des configurations externes
- Pratique :
 - Les groupes de tâches
 - Le déclenchement de l'exécution du DAG
 - Le mappage dynamique des tâches
 - Les opérateurs différés pour organiser des flux de travail complexes

Opérateurs personnalisés

- Rappel sur les opérateurs Airflow disponibles
- Création d'opérateurs et de capteurs personnalisés
- Création de hooks pour des systèmes externes
- Créer un opérateur personnalisé pour interagir avec une API
- Créer un capteur qui attend une condition spécifique

Intégration avec des environnements externes

- ExternalPythonOperator
- PythonVirtualenvOperator
- DockerOperator
- KubernetesPodOperator
- Implémenter une tâche en utilisant le PythonVirtualenvOperator

Cas pratiques, meilleures pratiques et limitations

- ModelOps
- Allocation dynamique des ressources
- Workflow de l'ingénierie des données
- Meilleures pratiques
 - Idempotence
 - Évitez le code de haut niveau
 - Décidez où exécuter les tâches de traitement des données
 - Ne pas passer de grandes quantités de données à travers Xcom
- Les limites
- Ressources pour améliorer ses connaissances

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.