

Workflow Orchestration with Cloud Composer

La formation Cloud Composer pour orchestrer vos workflows multi-cloud sur Google Cloud

1 jour / 7h

Présentation du cours

Cloud Composer est un service d'orchestration de flux de travail entièrement géré, basé sur Apache Airflow. Composer vous permet de créer, planifier, superviser et gérer des pipelines de flux de travail répartis entre plusieurs clouds et centres de données sur site.

Dans cette formation, vous découvrirez Apache Airflow et son implémentation via Cloud Composer. Vous apprendrez à provisionner des instances Composer, à créer et gérer des DAGs Airflow sur Composer, et à réaliser des tâches telles que le test, le débogage et la surveillance des DAGs Airflow.

Méthodes mobilisées : Ce cours alterne parties théoriques sous forme de lectures (slides), démos et parties pratiques sous forme de labs dirigés

Objectifs pédagogiques

- Explorer Apache Airflow et Cloud Composer en tant que solutions d'orchestration de workflows.
- Créer et gérer des DAGs Airflow selon les meilleures pratiques.
- Tester et déboguer des DAGs Airflow.
- Surveiller et observer les DAGs Airflow sur Cloud Composer.

Public cible

- Ingénieurs de données souhaitant apprendre à utiliser Apache Airflow et Cloud Composer pour orchestrer leurs workflows d'ingénierie des données.
- Data Engineers, Data Practitioners
- Toute personne intéressée par orchestrer ses workflows sur Google Cloud

Prérequis

- Avoir suivi le cours « Building Batch Data Pipelines on Google Cloud » ou connaissances équivalentes en analyse et ingénierie des données sur Google Cloud.

Programme

Module 01: Introduction à Cloud Composer

Sujets :

- Besoin des ingénieurs de données en matière d'orchestration de workflows
- Introduction à Apache Airflow
- Cloud Composer
- Configuration de l'environnement
- Utilisation de Composer et d'Airflow

Objectifs :

- Explorer Apache Airflow et Cloud Composer
- Provisionner des instances Cloud Composer
- Explorer les interfaces utilisateur d'Airflow et de Composer

Activités :

- Lab : Provisionnement de Cloud Composer

Module 02: Créer et gérer des DAGs

Sujets :

- Structure des DAGs et bonnes pratiques
- Opérateurs courants

- Dépendances, règles de déclenchement et contrôle de flux
- Intégration d'Airflow avec les services Google Cloud

Objectifs :

- Rédiger des DAGs
- Explorer les opérateurs Airflow courants
- Gérer les déclencheurs, les dépendances et le contrôle de flux
- Intégrer Airflow aux services Google Cloud

Activités :

- Lab : Assemblage d'un workflow de traitement de données

Module 03: Techniques Airflow avancées et meilleures pratiques

Sujets :

- Fonctionnalités avancées d'Airflow
- Débogage des DAGs
- Performance et évolutivité
- Sécurité et contrôle d'accès
- Observabilité et surveillance

Objectifs :

- Tirer parti des fonctionnalités avancées d'Airflow
- Déboguer des DAGs
- Observer et surveiller vos DAGs en cours d'exécution

Activités :

- Lab : Extension et surveillance des DAGs