

Mis à jour le 05/02/2025

S'inscrire

Formation dbt : Le workflow d'Airflow

2 jours (14 heures)

Présentation

Depuis quelques années, dbt (data build tool) est devenu un outil incontournable dans les architectures data modernes. Gratuit, open source et pensé pour le cloud, il permet aux équipes analytiques de s'inspirer des pratiques de l'ingénierie logicielle afin de rapidement déployer des pipelines de données.

Durant ce cours, vous mettrez en place un environnement de développement spécifique à dbt, ce qui vous permettra de développer et de tester vos premiers modèles.

Vous verrez également comment dbt s'intègre dans les architectures data actuelles, et quelles sont les problématiques auxquelles cet outil peut répondre.

À la suite de cette formation, vous saurez tout des fonctionnalités principales de dbt ainsi que des différentes manières de l'utiliser en production, notamment en utilisant Airflow et Kubernetes. Comme toutes nos formations, celle-ci vous présentera la dernière version en date et ses nouveautés : [dbt Core v1.9](#).

Objectifs

- Mettre en place un projet en partant de zéro
- Savoir utiliser et paramétrer l'IDE dbt
- Savoir déployer une production avec dbt
- Tester et documenter un modèle
- Intérioriser les bonnes pratiques

Public visé

- Data Scientists
- Data Analysts
- Chefs de projet

- Ingénieurs de données
- Business Analysts

PRÉ-REQUIS

- Connaissance des pipelines de données
- Savoir utiliser Docker
- Connaissance de python

PRÉ-REQUIS techniques

- Avoir un compte Gmail
- Avoir un compte Github (lié au compte Gmail)
- Python 3 et Git installés
- Docker et Docker Compose installés
- Un logiciel d'édition de code (ex : Visual Studio)
- Omyzsh

Programme de notre formation dbt

Introduction

- Qu'est-ce que dbt ?
- ETL et ELT : les différences
- dbt Core vs dbt Cloud
- Le rôle de dbt au sein du Modern Data Stack
- Qu'est-ce que n'est pas dbt ?
- Data Warehousing: Data Lakes, Data Warehouses, Data Lakehouse

Mise en place de l'environnement dbt

- Création de l'environnement virtuel Python
- Installation de dbt Core et du plugin adéquat
- Connexion de dbt à l'entrepôt de données
- Fichiers de paramétrage YAML
- Bonnes pratiques en matière de structure de projet : raw data, staging, etc.

Modèles : Développements et tests

- Développement de modèles basiques
- Organisation d'un data warehouse
- Configuration des modèles : schéma, matérialisation, etc.
- Introduction aux différentes commandes dbt et aux flags possibles
- Mise en place des tests et des vérifications de résultats

Modèles : Configurations avancées

- Utilisation des sources et vérification de la freshness
- Ingestion de fichiers plats dans dbt : les seeds et leur utilité
- Matérialisation incrémentale : en append ou en update
- Macros, hooks
- Snapshots

La documentation dans dbt

- Présentation de la documentation au format dbt
- La commande dbt docs et ses différentes utilisations
- Test types: format (ex: generic, singular) + qualitative (ex: structurel, relationnel)
- Test exemples
- Bonnes pratiques en matière de documentation : utilisation de la syntaxe Jinja
- DBT docs et lineage

Les packages

- Le Package Hub
 - dbt_utils
 - codegen
 - re_data

Dbt en production

- Intégration et développement continu (CI/CD)
- Orchestration de dbt à l'aide d'Apache Airflow : BashOperator, KubernetesPodOperator
- Présentation des différentes manières d'exécuter les projets
- Jobs scheduling
 - DBT cloud
 - Airflow

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.