

Formation DBT (Data Build Tool)

2 jours (14 heures)



Présentation

Cette formation vous offre une compréhension approfondie de DBT (Data Build Tool), un outil clé pour transformer et gérer vos données à grande échelle.

Vous apprendrez à créer des modèles SQL réutilisables, à versionner et collaborer avec Git, ainsi qu'à orchestrer vos pipelines de données avec Apache Airflow.

Nous adoptons une approche pratique qui vous permet d'explorer chaque aspect de DBT, Git et Airflow, avec des exercices concrets pour maximiser l'efficacité en production.

Objectifs

- Comprendre le rôle de DBT dans une stack ELT et ses avantages pour la transformation des données.
- Configurer un projet DBT, créer des modèles SQL et les gérer avec Git.
- Intégrer DBT avec Apache Airflow pour orchestrer et automatiser vos pipelines de données.
- Appliquer les meilleures pratiques de versioning et de collaboration sur des projets de données.
- Optimiser et documenter les pipelines de transformation de données.

Public visé

Profils techniques : Data Engineer, BI Engineer, Data Analyst, Data Scientist

Pré-requis

- Bonne maîtrise du SQL et compréhension des concepts ETL/ELT.
- Connaissances de base en Python et Git.
- Expérience avec des outils de gestion de pipeline (comme Airflow) est un plus.



Programme détaillé

Jour 1

Introduction à DBT

- Présentation de DBT : Histoire, concepts et utilisation
- Différences entre DBT Core et DBT Cloud
- Rôle de DBT dans un pipeline ELT

Architecture et composants d'un projet DBT

- Modèles, sources, tests, snapshots
- Structure d'un projet DBT : fichiers et répertoires importants
- Exemple de projet DBT

Configuration de DBT et mise en place du projet

- Installer DBT et les adaptateurs requis
- Configurer le fichier `profiles.yml` pour se connecter à un datawarehouse

TP : Initialisation et configuration d'un projet DBT

Collaboration avec Git

- Introduction à Git : versionning des modèles DBT
- Meilleures pratiques pour travailler avec des branches

TP : Utilisation de Git pour versionner un projet DBT

Jour 2

Modélisation avancée avec DBT

- Création de modèles SQL réutilisables et gestion des dépendances
- Utilisation des ****materializations**** : vue, table, incrémentale
- Tests et documentation des modèles avec DBT

TP : Créer, tester et documenter des modèles complexes

Orchestration avec Apache Airflow

- Introduction à Airflow : architecture et cas d'utilisation
- Intégration de DBT avec Airflow pour automatiser les pipelines
- Planification des tâches DBT dans un DAG Airflow

TP : Orchestration d'un pipeline DBT avec Airflow

Optimisation et déploiement de DBT en production

- Gestion des environnements et variables dans DBT
- Bonnes pratiques pour les commits Git et gestion des branches
- Déploiement continu avec Git et Airflow

TP : Créer un workflow de CI/CD pour un projet DBT



Organisation

- La formation alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions et de cas pratiques.
- Support de la formation distribué au format numérique à tous les participants.

Prix : 1290€ HT / participant

Tarif dégressif appliqué à partir de 5 participants, dans la limite de 25 participants maximum par session de formation.

Contact : +33 (0)7 45 41 66 05 / contact@ai-training.fr

