

dbt

Créez des workflows efficaces de gestion de données avec dbt

2 jours / 14h

Objectifs pédagogiques

- Comprendre les concepts clés, les avantages et l'architecture de dbt en tant qu'outil de transformation et de modélisation des données.
- Créer des modèles de données structurés avec dbt, et effectuer des transformations pour traiter et préparer les données en vue de l'analyse.
- Maîtriser les fonctionnalités avancées telles que les macros, les modèles Jinja, les variables et le contrôle de flux.
- Utiliser les instantanés (snapshots) dbt pour suivre les changements au fil du temps et gérer les données historiques, facilitant l'analyse des tendances historiques et les dimensions à changement lent.
- Mettre en œuvre des tests pour garantir la qualité et l'intégrité des données, permettant de valider les résultats des transformations et de détecter les anomalies.

Public cible

- Data Analysts, Data Engineers
- Toute personne intéressée par la transformation des données

Prérequis

- Des connaissances pratiques de SQL équivalentes au cours [SQL les fondamentaux](#)

Programme

Jour 1

Module 1 : Évolution de la pile de données

- Évolution de la pile de données
- Comprendre les différences entre les approches d'intégration de données Extract-Transform-Load (ETL) et Extract-Load-Transform (ELT)
- Introduction à la pile de données moderne

Module 2 : Introduction à dbt

- Aperçu de dbt
- Installation de dbt et configuration de l'environnement de développement
- Créer un projet dbt
- Connexion aux sources de données

Lab : Mise en place d'un projet dbt

Module 3 : Travailler avec des modèles dbt

- Comprendre les modèles dbt
- Comment fonctionnent les modèles de dbt ?
- Options de matérialisation
- Configuration de la matérialisation
- Présentation de la fonctionnalité de tagging pour l'organisation des métadonnées

Atelier : Créer des modèles de données avec dbt

Module 4 : Sources et références dbt

- Introduction aux sources dbt
- Configuration des sources dbt
- Travailler avec des références dbt

Atelier : Configuration des sources dbt, référencement des données externes et gestion des dépendances du modèle

Module 5 : Amorçage de vos modèles de données

- Introduction aux seeds dbt
- Création et remplissage de données seed
- Avantages de l'utilisation de seeds pour l'initialisation des données

- Intégrer des seeds à vos modèles dbt

Atelier : Créer et intégrer des seeds dans vos projets dbt

Jour 2

Module 6 : Snapshots pour gérer les données historiques

- Comprendre les snapshots dans dbt
- Configuration et définition de snapshots
- Exécuter et gérer des snapshots

Atelier : Mise en œuvre d'une stratégie de snapshots

Module 7 : Transformation et contrôle avancés des données

- Comprendre les macros
- Jinja, un langage de modèles
- Utiliser des variables pour gérer la configuration du pipeline de données

Atelier : Transformation et contrôle avancés des données

Module 8 : Packages

- Introduction aux packages dbt
- Explorer le hub dbt
- Installation et utilisation de package dbt

Atelier : Explorer les packages dbt

Module 9 : Tests

- Mettre en évidence les risques potentiels dans le code
- Mise en place de tests automatisés
- Choisir le test approprié
- Implémentation des tests de données

Atelier : Implémentation de tests de données

Module 10 : Documentation

- Documenter les modèles de données
- Utiliser les fonctionnalités de documentation intégrées de dbt pour générer et maintenir une documentation de modèle accessible et à jour
- L'importance du lineage

Laboratoire : Documenter les modèles de DBT

Module 11 : Analyses, hooks et exposition (avancé)

- Effectuer une analyse des données
- Exécuter du code personnalisé avant et après l'exécution de dbt
- Créer des assets de données partageables et accessibles

Atelier : Créer une exposition

Module 12 : Exploiter les artefacts générés (avancé)

- Comprendre manifest.json
- Introduction de run_result.json

Module 13 : Conclusion

- Ressources sur les meilleures pratiques en matière de dbt
- À propos de l'examen de certification DBT