

Sessions garanties disponibles

Formation dbt : industrialiser vos transformations de données en SQL

Réf. SF34618

Passez d'une collection de scripts SQL éparés à une chaîne de transformations versionnée, testée et documentée, suivant les pratiques de l'ingénierie logicielle.

2 j Durée	1 700 € HT Inter / participant	3 800 € HT Intra (≤12 pers.)	13 Sessions à venir
--------------	-----------------------------------	---------------------------------	------------------------

Objectifs pédagogiques

- ✓ Expliquer le rôle de dbt dans le Modern Data Stack et la logique ELT
- ✓ Structurer un projet dbt selon les conventions staging / intermédiaire / marts
- ✓ Développer des modèles SQL dynamiques avec Jinja et les macros
- ✓ Choisir la matérialisation adaptée (view, table, incremental, ephemeral) selon le cas d'usage
- ✓ Mettre en œuvre des snapshots pour historiser les évolutions de données
- ✓ Implémenter des tests automatisés et générer la documentation (lineage) d'un projet dbt
- ✓ Intégrer les packages de l'écosystème dbt (`dbt\_utils`, `codegen`...)
- ✓ Appliquer les bonnes pratiques de passage en production (environnements dev/prod, introduction à la CI/CD)

Public visé & prérequis

Public visé Business analysts, Data analysts, Data scientists, Ingénieurs Big Data

Prérequis- Maîtrise solide de SQL (jointures, agrégations, CTE, sous-requêtes) - Familiarité avec un data warehouse (BigQuery, Snowflake, PostgreSQL ou Redshift) - Bases de Git (clone, commit, branche)

Programme détaillé

Fondamentaux de dbt et du Modern Data Stack

- Modern Data Stack : positionnement de dbt dans la chaîne data (ingestion, stockage, transformation, restitution)
- ELT vs ETL : pourquoi la transformation se déplace dans le data warehouse
- dbt Core vs dbt Cloud : différences d'usage, arbitrage selon le contexte projet
- Rôle de l'Analytics Engineer : pratiques issues de l'ingénierie logicielle appliquées au SQL

Activités pratiques :

- Tour d'horizon des composants du Modern Data Stack sur un exemple concret
- Comparaison d'un pipeline ETL classique et d'un pipeline ELT avec dbt

Mise en route d'un projet dbt

- Installation et configuration de l'environnement de développement
- `profiles.yml` et `dbt\_project.yml` : configurer la connexion et les paramètres du projet
- Arborescence d'un projet dbt : dossiers `models`, `seeds`, `snapshots`, `macros`, `tests`
- Commandes de base : `dbt debug`, `dbt run`, `dbt compile`, `dbt build`
- Connexion à un data warehouse (BigQuery, Snowflake, PostgreSQL ou Redshift)

Activités pratiques :

- Création d'un projet dbt connecté à un data warehouse
- Exécution d'un premier modèle de staging sur une table brute
- Exploration des artefacts générés (`target/`, `manifest.json`)

Modèles, sources et matérialisations

- Modèles dbt : écriture en SQL, bonnes pratiques de nommage
- Sources : déclaration YAML des tables brutes, `{{ source() }}` pour les référencer
- Références entre modèles : `{{ ref() }}` et gestion automatique des dépendances (DAG)
- Matérialisations : view, table, incremental, ephemeral — choisir selon le cas d'usage
- Conventions de structuration : staging / intermediate / marts
- Seeds : intégrer des données de référence versionnées dans le projet

Activités pratiques :

- Construction d'un flux staging → intermediate → marts sur un jeu de données CRM (clients, commandes, produits)
- Comparaison des matérialisations view / table / incremental sur un même modèle
- Ajout d'un seed de référentiel et intégration dans un modèle

Transformations dynamiques avec Jinja et macros

- Syntaxe Jinja dans dbt : variables, blocs, filtres, expressions conditionnelles
- Modèles dynamiques : générer du SQL paramétré (filtrage conditionnel, pivots, boucles)
- Macros : créer des fonctions SQL réutilisables à l'échelle du projet
- Variables de projet : paramétrer un pipeline selon l'environnement
- Hooks : exécuter du code avant ou après un `dbt run`

Activités pratiques :

- Écriture d'un modèle dynamique avec conditions Jinja
- Création d'une macro pour un calcul transverse (normalisation, cast, formatage)
- Mise en place d'un pre-hook ou post-hook sur un modèle

Fiabilisation : snapshots, tests et documentation

- Snapshots : historiser les évolutions (SCD type 2) sur des dimensions à évolution lente
- Tests built-in : `not_null`, `unique`, `relationships`, `accepted_values`
- Tests personnalisés : écrire des assertions SQL adaptées au métier
- Documentation YAML : décrire modèles, colonnes, sources
- Lineage graph : visualiser les dépendances avec `dbt docs generate` et `dbt docs serve`

Activités pratiques :

- Mise en place d'un snapshot sur une table de dimensions (ex. clients)
- Ajout d'une suite de tests couvrant les modèles d'un flux complet
- Génération et exploration de la documentation et du lineage du projet

Écosystème et passage en production

- Packages dbt Hub : installer et utiliser `dbt_utils`, `codegen`, packages tiers
- Organisation d'un projet à l'échelle : conventions de nommage, structuration des dossiers
- Environnements multiples : configuration dev / prod, targets, schemas isolés
- Introduction à dbt Cloud : interface, scheduler, déclencheurs
- Introduction à la CI/CD : exécution de `dbt run` et `dbt test` sur pull request (GitHub Actions)

Activités pratiques :

- Installation et exploitation de `dbt_utils` pour simplifier des modèles existants
- Configuration de deux environnements (dev / prod) avec des targets distincts
- Mise en place d'un workflow GitHub Actions exécutant `dbt build` sur pull request

Prochaines sessions

DATE	LIEU	STATUT
30/11/2026	à distance	Sur confirmation
28/01/2027	Paris	Sur confirmation
28/01/2027	Lille	Sur confirmation

DATE	LIEU	STATUT
28/01/2027	Nantes	Sur confirmation
28/01/2027	à distance	Sur confirmation
03/05/2027	Lyon	Sur confirmation

Pourquoi choisir Sparks Formation ?

20 ans d'expertise IT, des formateurs experts et des sessions garanties contre l'annulation.

20 ans
d'expertise

1 068+
formations

11 178
apprenants formés/an

+67
Net Promoter Score

Avis des participants

4,9/5 — 13 avis

Arnaud M. — That was really useful for me, even though I am a PDA. The trainer was very patient with me (I needed more explanation as I was the only...

Thierry R. — Formation idéale pour appréhender la technologie nosql Bon support de formation. Bonne pédagogie

Votre formateur

Maxime L. — Formateur Data Engineering & Big Data

Maxime intervient sur des formations portant sur le data engineering et le Big Data. Parmi les formations qui lui sont confiées figurent notamment « Airflow » et « Le rôle...

Ils nous font confiance

Cette année notre filière Data Engineering & Big Data a accompagné plus de 73 entreprises dont :

SOPRA STERIA CAPGEMINI TECHNOLOGY SERVICES AMADEUS SAS GROUPAMA SUPPORTS & SERVICES AUBAY FRAIKIN FRANCE LBC FRANCE
ORANGE BUSINESS SERVICES SA ACATUS INFORMATIQUE ACENSI SANA +63 autres

Financement & certifications

Cette formation est éligible aux financements OPCO et FNE-Formation. Nos équipes vous accompagnent dans le montage de votre dossier.



Réservez votre place dès maintenant

20 ans d'expertise · 1 000+ formations · NPS +67 · réponse sous 24h ouvrées

0 805 950 800

demande@sparks-formation.com