

Atelier 3 - Préparation des données

Dans la continuité des deux précédents ateliers, vous amorcer une préparation des données (nettoyage, consolidation, ...) selon une architecture en médaillon avec dbt. Vous pouvez vous appuyer sur les commits de l'ébauche de projet <https://github.com/labasse/dbt-park-bx>.

Etape 1 - Isolation des données brutes

Créer un schéma pour y mettre les données brutes importées `CREATE SCHEMA IF NOT EXISTS raw;` (ou autre nom de votre choix).

Modifier l'ingestion ou déplacer les données importées dans ce schéma.

Etape 2 - Initialisation du projet dbt

Créer un projet dbt dans le répertoire de votre projet avec `dbt init`.

Créer le fichier `profiles.yml` un fichier `models/sources.yml` référençant la ou les tables de raw.

Versionnez l'étape.

Etape 3 - Modèles

Créer les modèles "bronze" pour chaque table du schéma de données brutes.

Créer quelques modèles "silver" dont au moins un avec jointure d'un modèle existant.

Dans un fichier `silver/models.yml`, imposez des tests sur l'existence et l'unicité des clés primaires de ces modèles.

Vérifier avec `dbt build` et `dbt run` la bonne exécution de vos transformations.

Versionnez l'étape.

Etape 4 - Documentation

Renseignez les descriptions de vos sources et modèles et visualisez la documentation grâce aux commandes `dbt docs build` et `dbt docs serve`.

Versionnez l'étape.

Livrable

Pdf avec une copie d'écran de la documentation et du graphe de documentation.