

BUSINESS INTELLIGENCE

Construction du pipeline de données dans Snowflake

Mon projet consiste à mettre en place un pipeline de données complet dans Snowflake, depuis l'importation de plusieurs fichiers CSV bruts jusqu'à la production d'un fichier final unique au format.csv.

La création du fichier s'est faite en 9 étapes :

Étape	Objectif
WAREHOUSE	Créer et activer l'entrepôt de calcul (WAREHOUSE_SALES)
DATABASE	Créer la base logique qui contiendra les schémas (SALES)
SCHEMA	Séparer les zones fonctionnelles SALES_OP pour l'opérationnel, SALES_CURATED pour l'analytique.
TABLES	Définir les structures des tables (CUSTOMERS, PRODUCTS, SALES, etc.) et leurs clés/relations.
INGESTION	Charger les fichiers CSV depuis le stage dans les tables (commandes COPY INTO).
TRANSFOORMATION	Nettoyer et transformer les données (typage, calculs comme AnnualIncome, SalesAmount, jointures intermédiaires).
CURATED	Créer les tables finales prêtes pour l'analyse (dimensions + tables de faits dans SALES_CURATED).
PRS	Créer des vues orientées reporting (ex. ventes par année, par pays, par catégorie) pour les outils BI.
EXPORT	Exporter un résultat (table/vue) en CSV vers un stage puis le télécharger (COPY INTO @export_stage/...).

Les premières étapes ont servi à préparer l'environnement : création de l'entrepôt de calcul (WAREHOUSE_SALES), de la base de données (SALES), du schéma opérationnel (SALES_OP) et du stage pour déposer les fichiers CSV.

Les fichiers AdventureWorks (clients, produits, catégories, sous-catégories, territoires et ventes 2015–2017) ont été chargés dans des tables brutes distinctes afin de conserver les données originales. Les données ont ensuite été nettoyées : correction des types (dates, montants), réalignement des colonnes et fusion des trois fichiers de ventes en une seule table consolidée.

Une couche CURATED a été construite dans SALES_CURATED avec des dimensions clients, produits (enrichis des catégories) et territoires, ainsi qu'une table de faits des ventes (FACT_SALES) intégrant quantité, prix unitaire et montant de vente. À partir de cette couche, une table finale unifiée reliant ventes, clients, produits, catégories et territoires a été créée, puis exportée en un fichier CSV unique (FINAL_EXPORT.csv) prêt pour le reporting.