

Conception et Mise en œuvre d'un Système de Business Intelligence Unifié

Plateforme de Data Warehouse, ETL et Visualisation Multi-Sources

Étude de Cas : Northwind (SQL Server + Access)

Douaoura rassim

232331530909

3ème Année

20 décembre 2025

Table des matières

1	Introduction	2
2	Architecture Globale	3
2.1	Présentation du Système	3
2.2	Flux de Données	3
3	Conception du Processus ETL	4
3.1	Extraction Multi-Sources	4
3.2	Transformation et Unification	4
3.3	Chargement (Loading)	4
4	Modélisation du Data Warehouse	5
4.1	Conception du Schéma en Étoile	5
5	Visualisation et Analytics	6
5.1	Stratégie de Visualisation	6
5.2	Analyse des Indicateurs Clés (KPI)	6
5.3	Analyses Graphiques	6
5.4	Analyse Multidimensionnelle 3D	8
6	Justifications et Résultats	9
6.1	Choix des Technologies	9
6.2	Interprétation des Résultats	9
7	Conclusion	10

Chapitre 1

Introduction

Dans l'économie moderne, la donnée est devenue un actif stratégique. Les systèmes de Business Intelligence (BI) jouent un rôle crucial en transformant des volumes massifs de données transactionnelles en informations exploitables pour soutenir la prise de décision à tous les niveaux.

Ce projet présente la conception et l'implémentation d'un pipeline BI complet basé sur les sources **Northwind**. La particularité de ce système réside dans sa capacité à unifier des données provenant d'environnements hétérogènes : **SQL Server** pour l'historique et **Microsoft Access** pour les données opérationnelles récentes.

La solution couvre l'intégralité de la chaîne décisionnelle, de l'extraction multi-sources à la visualisation interactive 3D, en respectant les standards de l'architecture Lakehouse et les meilleures pratiques de l'ingénierie des données.

Objectifs du Projet

Les principaux objectifs de ce projet sont :

- Concevoir un entrepôt de données (Data Warehouse) unifié et évolutif.
- Implémenter un pipeline ETL fiable gérant l'hétérogénéité des sources.
- Assurer la normalisation et la qualité des données (Data Quality).
- Développer un tableau de bord analytique interactif (KPIs, graphiques 2D/3D).
- Fournir une architecture portable et légère pour la prise de décision.

Chapitre 2

Architecture Globale

2.1 Présentation du Système

Le système BI proposé adopte une architecture en couches, optimisée pour la flexibilité et la portabilité :

- **Sources de Données** : SQL Server (Northwind principal) et MS Access (Northwind 2012).
- **Couche ETL** : Scripts Python (`pyodbc`, `pandas`) pour l'unification des sources.
- **Data Lake / Warehouse** : Fichiers CSV structurés dans le dossier `/data`, prêts pour la consommation.
- **Couche Analytique** : Traitement côté client (JavaScript) et analyses Python statiques.
- **Couche de Présentation** : Dashboard Web interactif et rapports automatisés.

2.2 Flux de Données

Le workflow est organisé selon les étapes suivantes :

1. Extraction parallèle depuis SQL Server et Access.
2. Nettoyage, normalisation des identifiants (préfixage `ACC_`) et réconciliation des schémas.
3. Stockage dans des fichiers CSV structurés.
4. Fusion dynamique en mémoire lors du chargement du tableau de bord.
5. Visualisation des indicateurs de performance (KPIs) en temps réel.

Chapitre 3

Conception du Processus ETL

3.1 Extraction Multi-Sources

L'extraction utilise la bibliothèque `pyodbc` pour se connecter aux bases relationnelles. Les tables extraites incluent les commandes (*Orders*), les détails de commandes, les produits, les clients et les employés de chaque source.

3.2 Transformation et Unification

La phase de transformation est critique pour assurer la cohérence :

- **Normalisation** : Alignement des types de données et des noms de colonnes.
- **Dédoublonnage** : Gestion des conflits d'IDs entre SQL Server et Access.
- **Enrichissement** : Calcul automatique des revenus par ligne ($UnitPrice * Quantity * (1 - Discount)$).
- **Temporalité** : Extraction des dimensions de temps (Année, Mois, Trimestre).

3.3 Chargement (Loading)

Les données transformées sont persistées sous format CSV UTF-8. Ce choix garantit que le système de visualisation peut fonctionner de manière autonome sans dépendre d'un serveur de base de données complexe, tout en conservant une structure relationnelle claire.

Chapitre 4

Modélisation du Data Warehouse

4.1 Conception du Schéma en Étoile

Le système repose sur un **schéma en étoile** virtuel, optimisé pour les requêtes analytiques rapides.

Table de Faits

- **Fact_Sales** : Centralise toutes les transactions (ventes) issues des deux sources unifiées.

Tables de Dimensions

- **Dim_Customers** : Attributs clients et analyse géographique.
- **Dim_Products** : Catalogue produits avec catégories et fournisseurs.
- **Dim_Employees** : Informations sur la force de vente et ancienneté.
- **Dim_Time** : Hiérarchie temporelle pour l'analyse des tendances.

Chapitre 5

Visualisation et Analytics

5.1 Stratégie de Visualisation

Deux approches complémentaires ont été implémentées :

- **Visualisations Statiques (Python)** : Pour le reporting formel et les analyses statistiques poussées.
- **Tableau de Bord Interactif (Web)** : Pour l’exploration dynamique des données via Plotly.js.

5.2 Analyse des Indicateurs Clés (KPI)

Le tableau de bord consolide quatre indicateurs majeurs :

- **Revenu Total** : CA global généré par l’entreprise unifiée.
- **Volume de Commandes** : Fréquence d’achat consolidée.
- **Nombre de Clients** : Taille de la base client active.
- **Panier Moyen** : Valeur moyenne par transaction.

5.3 Analyses Graphiques

Le système fournit plusieurs vues analytiques permettant de transformer les données en intelligence décisionnelle :

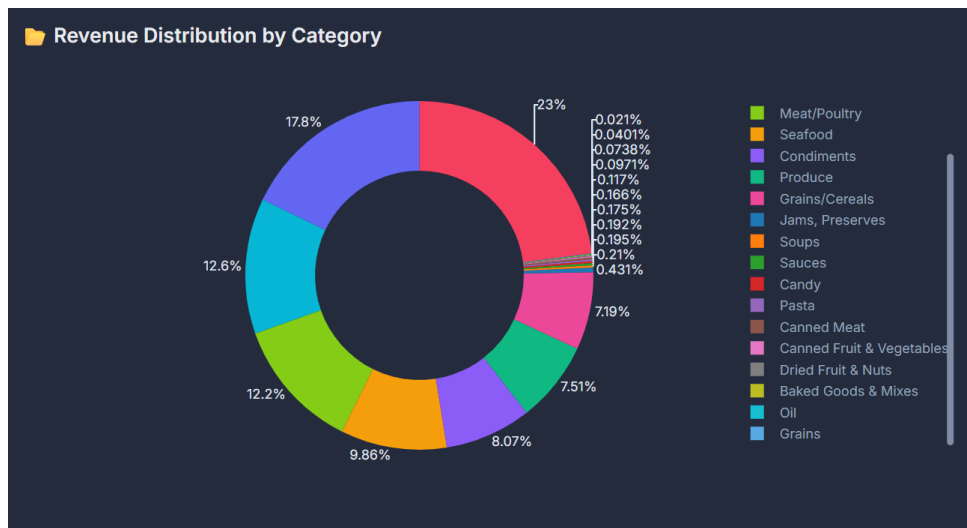


FIGURE 5.1 – Distribution des Revenus par Catégorie (Donut Chart)

- **Distribution des Revenus par Catégorie** (Figure 5.1) : Ce graphique en anneau permet de visualiser la contribution relative de chaque catégorie de produit au chiffre d'affaires total. Il facilite l'identification immédiate des segments de marché les plus rentables.

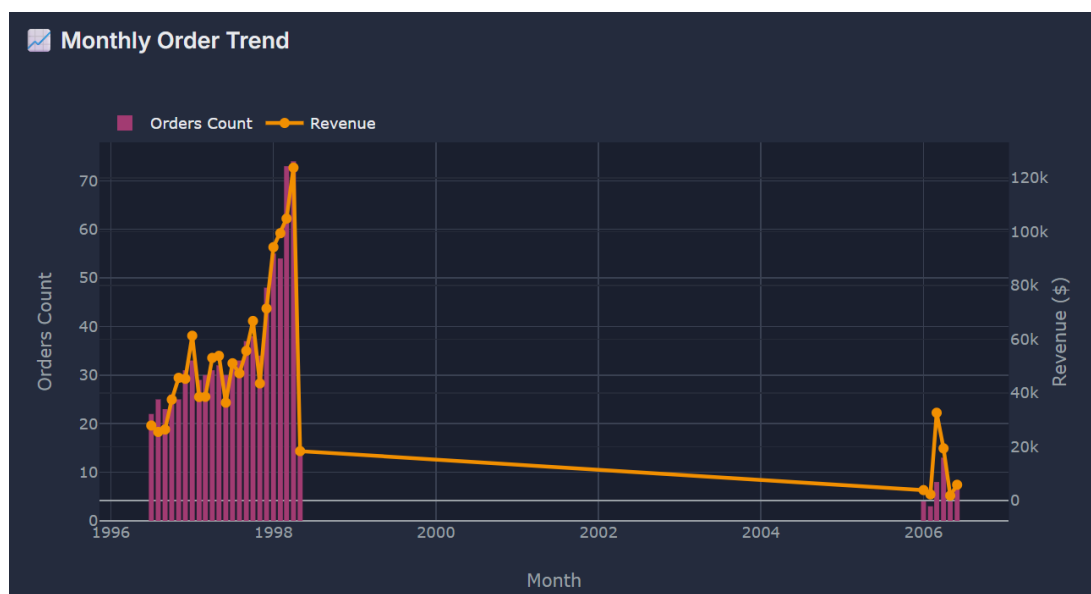


FIGURE 5.2 – Évolution Mensuelle des Commandes et Revenus

- **Tendances Mensuelles** (Figure 5.2) : Cette visualisation combinée (barres et lignes) affiche simultanément le volume des commandes et le revenu mensuel. Elle est essentielle pour détecter la saisonnalité et les pics d'activité sur plusieurs exercices.

5.4 Analyse Multidimensionnelle 3D

Une innovation majeure de ce projet est l'intégration de visualisations interactives 3D, offrant une perspective enrichie sur les données complexes.

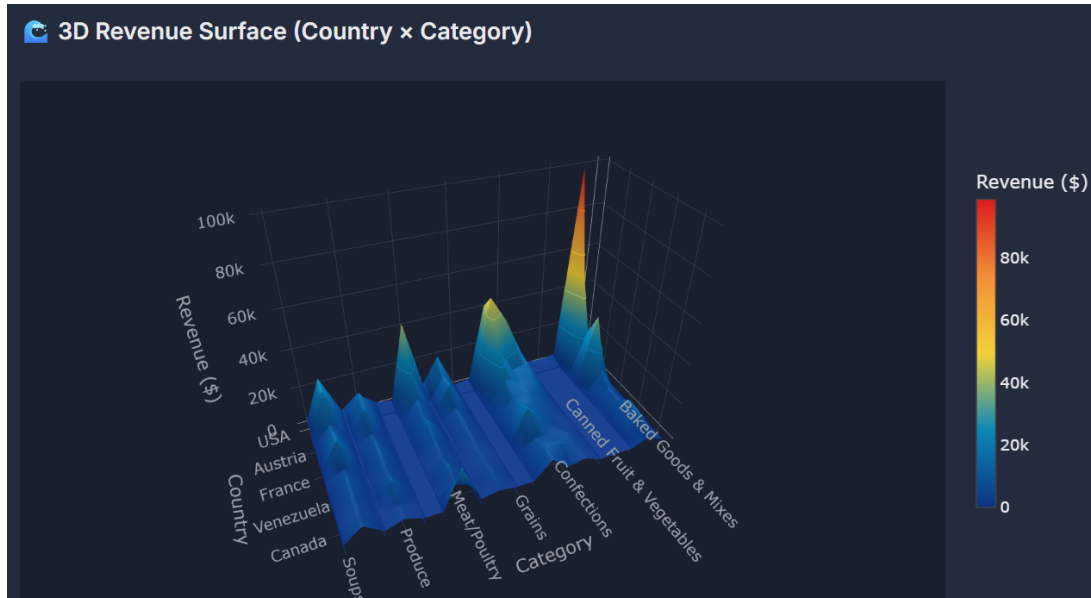


FIGURE 5.3 – Surface des Revenus 3D (Pays \times Catégorie)

- **Surface de Revenu 3D** (Figure 5.3) : Ce graphique modélise l'interaction entre les pays et les catégories de produits. L'axe Z représente le revenu, permettant de repérer visuellement les "sommets" de performance où une catégorie spécifique domine un marché géographique particulier.

Chapitre 6

Justifications et Résultats

6.1 Choix des Technologies

- **Python/Pandas** : Pour la flexibilité totale de l'ETL.
- **Format CSV** : Pour une portabilité maximale "Zero-Server".
- **HTML5/Vanilla JS** : Pour un dashboard léger, responsive et universel.

6.2 Interprétation des Résultats

L'analyse unifiée a permis de dégager des informations stratégiques indispensables, notamment sur la complémentarité des données Access (ventes récentes) et SQL Server (historique profond), offrant ainsi une vue rétrospective et prospective complète de l'activité.

Chapitre 7

Conclusion

Ce projet démontre la mise en œuvre réussie d'un système BI "End-to-End". En combinant ingénierie des données et visualisation avancée, nous avons transformé des sources brutes hétérogènes en un outil de décision robuste. La flexibilité de l'architecture permet d'envisager facilement l'ajout de nouvelles sources ou l'intégration de modèles prédictifs à l'avenir.