

GÉNIE INFORMATIQUE – GINF3

---

# Global Retail 360 : Analyse de la Supply Chain et de la Satisfaction Client

---

Projet Business Intelligence avec Power BI

Utilisation de Talend Open Studio pour l'ETL

*Encadré par :*  
Mr. HASSAN BADIR

*Auteurs :*  
**EL ADNANI EL MEHDI**  
**N'FALY SYLLA**  
**CHAIBERRAS SOUHAIL**

*Date :*  
Janvier 2026

Année Académique 2025-2026

# Résumé Exécutif

Ce rapport présente une analyse approfondie de la supply chain et de la satisfaction client pour une entreprise de e-commerce international. L'étude se concentre sur l'optimisation des délais de livraison et de la marge bénéficiaire en analysant la corrélation entre les retards logistiques et le taux de retour des produits.

**Problématique :** Comment optimiser les délais de livraison et la marge bénéficiaire d'une entreprise de e-commerce internationale en analysant la corrélation entre les retards logistiques et le taux de retour des produits ?

**Objectifs principaux :**

- Analyser les performances logistiques par région et mode d'expédition
- Identifier les corrélations entre retards et retours produits
- Optimiser les coûts de livraison tout en maintenant la satisfaction client
- Proposer des recommandations stratégiques basées sur les données

**Méthodologie :** Utilisation de Talend Open Studio pour l'extraction, transformation et chargement des données (ETL), suivie d'une analyse visuelle avec Power BI.

# Table des matières

<b>Résumé Exécutif</b>	<b>1</b>
<b>1 Introduction Générale</b>	<b>5</b>
1.1 Contexte et Enjeux Stratégiques . . . . .	5
1.2 Problématique et Objectifs Décisionnels . . . . .	5
1.3 Objectifs SMART du Projet . . . . .	5
1.3.1 Objectifs Fonctionnels . . . . .	5
1.4 Périmètre du Système d'Information . . . . .	7
<b>2 Préparation de l'Environnement et Génération des Données</b>	<b>8</b>
2.1 Installation des Dépendances Python . . . . .	8
2.1.1 Bibliothèques Requises . . . . .	8
2.2 Génération des Fichiers Sources . . . . .	8
2.2.1 Architecture du Script de Génération . . . . .	8
2.2.2 Données Générées . . . . .	9
2.2.3 Importance de la Simulation . . . . .	9
<b>3 Architecture Technique et Conception du Data Warehouse</b>	<b>10</b>
3.1 Configuration de la Base de Données PostgreSQL . . . . .	10
3.1.1 Connexion et Paramétrage . . . . .	10
3.1.2 Création du Schéma Dimensionnel . . . . .	11
3.2 Architecture Talend Open Studio . . . . .	11
3.2.1 Organisation du Projet . . . . .	11
3.2.2 Palette de Composants . . . . .	13
<b>4 Processus ETL : Extraction et Transformation</b>	<b>15</b>
4.1 Chargement de la Dimension Location . . . . .	15
4.1.1 Pipeline de Transformation . . . . .	15
4.1.2 Détails du Mapping . . . . .	16
4.2 Chargement des Dimensions Produit et Client . . . . .	16
4.2.1 Volume de Données Traité . . . . .	16
4.3 Traitement des Fichiers Hétérogènes . . . . .	16
4.3.1 Fichier Délimité (CSV) . . . . .	17
4.3.2 Fichier XML (Retours Logistiques) . . . . .	18
<b>5 Intégration de la Table de Faits</b>	<b>19</b>
5.1 Architecture du Job Principal . . . . .	19
5.1.1 Vue d'Ensemble du Pipeline . . . . .	19
5.1.2 Flux de la Table de Faits . . . . .	20
5.2 Configuration Avancée des Composants . . . . .	21

5.2.1	Paramètres de Sortie PostgreSQL . . . . .	21
5.3	Exécution et Monitoring . . . . .	21
5.3.1	Performance du Pipeline . . . . .	22
<b>6</b>	<b>Dashboards Power BI : Réalisations et Résultats</b>	<b>25</b>
6.1	Vue d'Ensemble des Tableaux de Bord . . . . .	25
6.2	Dashboard Financier : Analyse des Ventes . . . . .	25
6.2.1	Métriques Clés et KPIs . . . . .	25
6.2.2	Analyses Géospatiales . . . . .	26
6.2.3	Performance par Catégorie . . . . .	26
6.2.4	Tendances Temporelles . . . . .	26
6.3	Dashboard Logistique : Supply Chain et Retours . . . . .	26
6.3.1	Indicateurs Opérationnels . . . . .	26
6.3.2	Analyse des Délais par Mode de Livraison . . . . .	27
6.3.3	Causes des Retours . . . . .	27
6.3.4	Corrélation Délai-Retour . . . . .	28
6.3.5	Produits à Risque . . . . .	28
6.4	Interactivité et Filtrage Dynamique . . . . .	28
<b>7</b>	<b>Modélisation Analytique sous Power BI</b>	<b>29</b>
7.1	Implémentation du Modèle Sémantique . . . . .	29
7.1.1	Relations et Cardinalités . . . . .	29
7.2	Intelligence de Calcul (DAX) . . . . .	30
<b>8</b>	<b>Analyse des Résultats et Recommandations</b>	<b>31</b>
8.1	Performance du Processus ETL . . . . .	31
8.1.1	Synthèse des Résultats . . . . .	31
8.1.2	Points Forts de l'Architecture . . . . .	31
8.2	Réponse à la Problématique . . . . .	31
8.2.1	Réponse Apportée . . . . .	31
8.3	Perspectives d'Amélioration . . . . .	32
8.3.1	Extensions Techniques . . . . .	32
8.3.2	Évolution de l'Architecture . . . . .	32
8.4	Leçons Apprises . . . . .	33
8.4.1	Aspects Techniques . . . . .	33
8.4.2	Aspects Métier . . . . .	33
<b>9</b>	<b>Conclusion</b>	<b>34</b>

# Table des figures

1.1	Architecture globale du système BI Global Retail . . . . .	6
1.2	Diagramme de cas d'utilisation - Système BI Global Retail . . . . .	6
2.1	Installation des bibliothèques Python (pandas, openpyxl, lxml) . . . . .	8
2.2	Exécution du script de génération des sources de données . . . . .	9
2.3	Confirmation de génération des trois types de fichiers sources . . . . .	9
3.1	Configuration de la connexion PostgreSQL dans Talend . . . . .	10
3.2	Structure des tables dans PostgreSQL . . . . .	11
3.3	Hierarchie des Jobs dans Talend Open Studio . . . . .	11
3.4	Palette des composants Talend disponibles . . . . .	13
4.1	Flux ETL complet pour la dimension Location . . . . .	15
4.2	Configuration détaillée du composant tMap pour Location . . . . .	16
4.3	Configuration du composant tFileInputDelimited . . . . .	17
4.4	Configuration du parseur XML avec expression XPath . . . . .	18
5.1	Vue Designer du job principal avec flux séquentiels . . . . .	19
5.2	Pipeline complexe d'intégration de fact_sales . . . . .	20
5.3	Paramètres avancés du composant tDBOutput . . . . .	21
5.4	Métriques de performance en temps réel . . . . .	22
5.5	Vue d'exécution finale avec statistiques complètes . . . . .	23
6.1	Dashboard Analyse Financière - Global Sales & Logistics . . . . .	25
6.2	Dashboard Supply Chain & Retours - Analyse Logistique . . . . .	27
7.1	Schéma en Étoile final dans Power BI . . . . .	29

# Chapitre 1

## Introduction Générale

### 1.1 Contexte et Enjeux Stratégiques

Dans un écosystème e-commerce globalisé, la maîtrise de la *Supply Chain* ne représente plus seulement un défi logistique, mais un levier direct de rentabilité financière. Ce projet, intitulé **Global Retail 360**, s'inscrit dans une démarche d'intelligence décisionnelle visant à transformer des données opérationnelles brutes en indicateurs de performance (KPI) actionnables.

### 1.2 Problématique et Objectifs Décisionnels

L'enjeu central de cette étude est de quantifier l'élasticité entre la performance logistique et la satisfaction client. Plus précisément, il s'agit de répondre à la question suivante :

*« Comment l'optimisation des flux logistiques et la réduction des délais de livraison impactent-elles le taux de retour produit et, par extension, la marge nette de l'entreprise ? »*

Les objectifs sont hiérarchisés selon trois axes :

- **Analytique** : Corréler les retards de livraison avec les comportements de retour.
- **Opérationnel** : Identifier les segments (régions/modes d'envoi) sous-performants.
- **Stratégique** : Proposer des recommandations basées sur des évidences empiriques pour maximiser le ROI.

### 1.3 Objectifs SMART du Projet

Pour répondre à cette problématique, nous avons défini des objectifs selon la méthode SMART (Spécifique, Mesurable, Atteignable, Réaliste, Temporel).

#### 1.3.1 Objectifs Fonctionnels

- **Centraliser l'information** : Offrir une vue unique consolidant les Ventes, les Objectifs et la Logistique.

- **Piloter la Performance** : Permettre au Directeur Financier (CFO) de suivre la marge par pays et par produit.
- **Analyser les Retours** : Permettre au Responsable Logistique d'identifier les causes racines des retours (Produit défectueux vs Retard de livraison).

L'interaction entre ces acteurs et le système est illustrée par les diagrammes suivants.

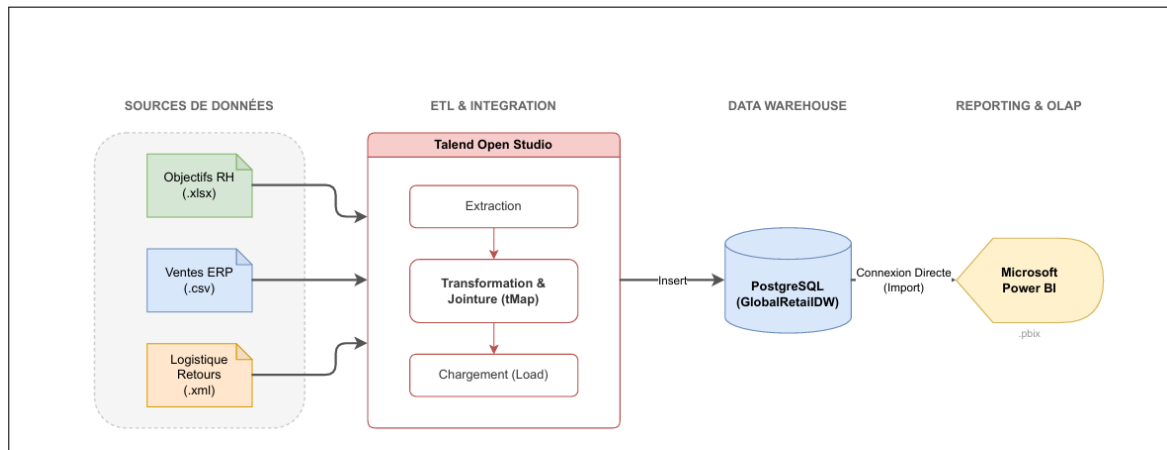


FIGURE 1.1 – Architecture globale du système BI Global Retail

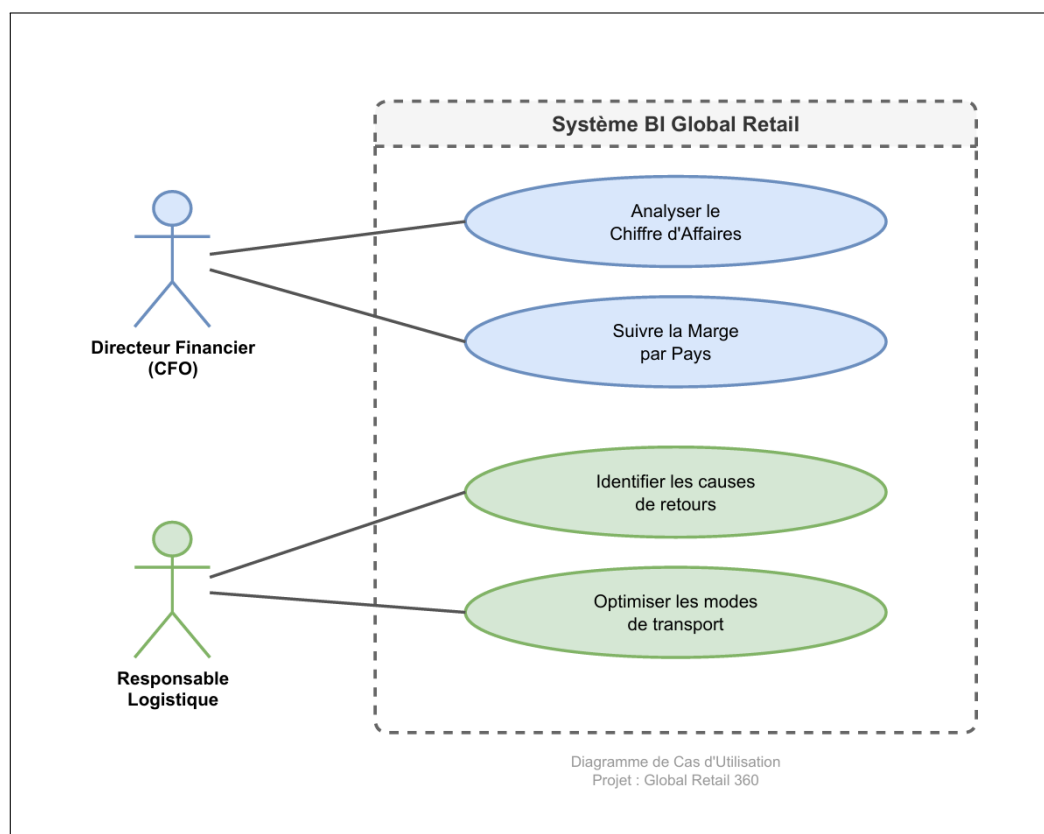


FIGURE 1.2 – Diagramme de cas d'utilisation - Système BI Global Retail

## 1.4 Périmètre du Système d'Information

L'architecture repose sur l'analyse de :

- **Transactions** : 51 290 lignes de ventes couvrant 13 marchés mondiaux.
- **Logistique** : Une gestion multi-modale (Same Day, First/Second Class).
- **Service Client** : Un flux XML dédié aux retours et aux motifs d'insatisfaction.



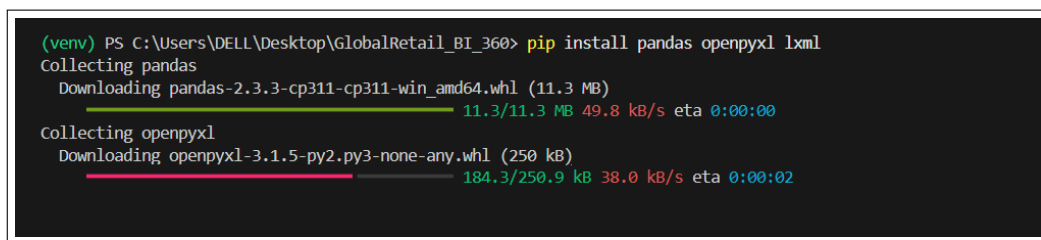
# Chapitre 2

## Préparation de l'Environnement et Génération des Données

### 2.1 Installation des Dépendances Python

La première étape du projet consiste à préparer l'environnement de développement avec les bibliothèques nécessaires au traitement des données.

#### 2.1.1 Bibliothèques Requises



```
(venv) PS C:\Users\DELL\Desktop\GlobalRetail_BI_360> pip install pandas openpyxl lxml
Collecting pandas
  Downloading pandas-2.3.3-cp311-cp311-win_amd64.whl (11.3 MB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 11.3/11.3 MB 49.8 kB/s eta 0:00:00
Collecting openpyxl
  Downloading openpyxl-3.1.5-py2.py3-none-any.whl (250 kB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 184.3/250.9 kB 38.0 kB/s eta 0:00:02
```

FIGURE 2.1 – Installation des bibliothèques Python (pandas, openpyxl, lxml)

Les bibliothèques installées permettent de :

- **pandas** : Manipulation et analyse de données tabulaires
- **openpyxl** : Lecture et écriture de fichiers Excel (.xlsx)
- **lxml** : Traitement de fichiers XML pour les retours logistiques

Ces outils constituent la base technique pour la génération de fichiers sources hétérogènes simulant un environnement ERP réel.

### 2.2 Génération des Fichiers Sources

#### 2.2.1 Architecture du Script de Génération

Le script `generate_sources.py` crée un ensemble de fichiers représentant les différentes sources de données d'une entreprise :

```

(venv) PS C:\Users\DELL\Desktop\GlobalRetail_BI_360\scripts> python .\generate_sources.py
2026-01-14 15:12:46,931 [INFO] DÉMARRAGE DU SCRIPT DE GÉNÉRATION DE DONNÉES ...
2026-01-14 15:12:46,931 [INFO] Lecture du fichier source : C:\Users\DELL\Desktop\GlobalRetail_BI_360\scripts\../data/input_origin
al/Raw_Data.xlsx
2026-01-14 15:12:53,402 [INFO] ✓ Fichier chargé avec succès. 51290 lignes et 24 colonnes.
2026-01-14 15:12:53,402 [INFO] ♦ Traitement : Génération du fichier CSV (Ventes)...
2026-01-14 15:12:53,760 [INFO] ✓ Export CSV terminé : C:\Users\DELL\Desktop\GlobalRetail_BI_360\scripts\../data/generated_source
s\Source_ERP_Ventes.csv
2026-01-14 15:12:53,781 [INFO] ♦ Traitement : Génération du fichier Excel (Objectifs RH)...
2026-01-14 15:12:53,786 [INFO] Régions détectées : 13
2026-01-14 15:12:53,828 [INFO] ✓ Export Excel terminé : C:\Users\DELL\Desktop\GlobalRetail_BI_360\scripts\../data/generated_sour
ces\Source_RH_Targets.xlsx
2026-01-14 15:12:53,828 [INFO] ♦ Traitement : Génération du fichier XML (Retours Logistique)...
2026-01-14 15:12:53,834 [INFO] Simulation de 2002 retours sur 25035 commandes.
2026-01-14 15:12:53,838 [INFO] ✓ Export XML terminé : C:\Users\DELL\Desktop\GlobalRetail_BI_360\scripts\../data/generated_source
s\Source_Logistique_Returns.xml
2026-01-14 15:12:53,838 [INFO] ✎ SCRIPT TERMINÉ AVEC SUCCÈS
(venv) PS C:\Users\DELL\Desktop\GlobalRetail_BI_360\scripts>

```

FIGURE 2.2 – Exécution du script de génération des sources de données

## 2.2.2 Données Générées

Le processus de génération produit :

1. **Fichier CSV des Ventes** (Source\_ERP\_Ventes.csv)
  - 51 290 lignes de transactions commerciales
  - 24 colonnes incluant : Order ID, dates, produits, clients, localisation, montants
  - Format délimité par point-virgule pour compatibilité internationale
2. **Fichier Excel des Objectifs RH** (Source\_RH\_Targets.xlsx)
  - 13 régions avec managers assignés
  - Objectifs de vente annuels par région
  - Structure multi-feuilles pour simulation de classeur complexe
3. **Fichier XML des Retours** (Source\_Logistique\_Returns.xml)
  - 2002 retours simulés avec motifs (Late Delivery, Defective, etc.)
  - Structure hiérarchique XML avec balises Returns/Return
  - Enrichissement avec timestamps et statuts de traitement

```

(venv) PS C:\Users\DELL\Desktop\GlobalRetail_BI_360\script> python .\generate_sources.py
2026-01-14 15:12:46,931 [INFO] DÉMARRAGE DU SCRIPT DE GÉNÉRATION DE DONNÉES ...
2026-01-14 15:12:46,931 [INFO] Lecture du fichier source : C:\Users\DELL\Desktop\GlobalRetail_BI_360\scripts\../data/input_origin
al/Raw_Data.xlsx
2026-01-14 15:12:53,402 [INFO] ✓ Fichier chargé avec succès. 51290 lignes et 24 colonnes.
2026-01-14 15:12:53,402 [INFO] ♦ Traitement : Génération du fichier CSV (Ventes)...
2026-01-14 15:12:53,760 [INFO] ✓ Export CSV terminé : C:\Users\DELL\Desktop\GlobalRetail_BI_360\scripts\../data/generated_source
s\Source_ERP_Ventes.csv
2026-01-14 15:12:53,781 [INFO] ♦ Traitement : Génération du fichier Excel (Objectifs RH)...
2026-01-14 15:12:53,786 [INFO] Régions détectées : 13
2026-01-14 15:12:53,828 [INFO] ✓ Export Excel terminé : C:\Users\DELL\Desktop\GlobalRetail_BI_360\scripts\../data/generated_sour
ces\Source_RH_Targets.xlsx
2026-01-14 15:12:53,828 [INFO] ♦ Traitement : Génération du fichier XML (Retours Logistique)...
2026-01-14 15:12:53,834 [INFO] Simulation de 2002 retours sur 25035 commandes.
2026-01-14 15:12:53,838 [INFO] ✓ Export XML terminé : C:\Users\DELL\Desktop\GlobalRetail_BI_360\scripts\../data/generated_source
s\Source_Logistique_Returns.xml
(venv) PS C:\Users\DELL\Desktop\GlobalRetail_BI_360\scripts> python .\generate_sources.py

```

FIGURE 2.3 – Confirmation de génération des trois types de fichiers sources

## 2.2.3 Importance de la Simulation

Cette approche de génération simulée permet :

- De tester la robustesse du pipeline ETL sur des données réalistes
- De valider la capacité à gérer des formats hétérogènes (CSV, Excel, XML)
- De reproduire les défis rencontrés en environnement de production

# Chapitre 3

## Architecture Technique et Conception du Data Warehouse

### 3.1 Configuration de la Base de Données PostgreSQL

#### 3.1.1 Connexion et Paramétrage

La base de données cible a été configurée pour optimiser les requêtes analytiques de type OLAP.

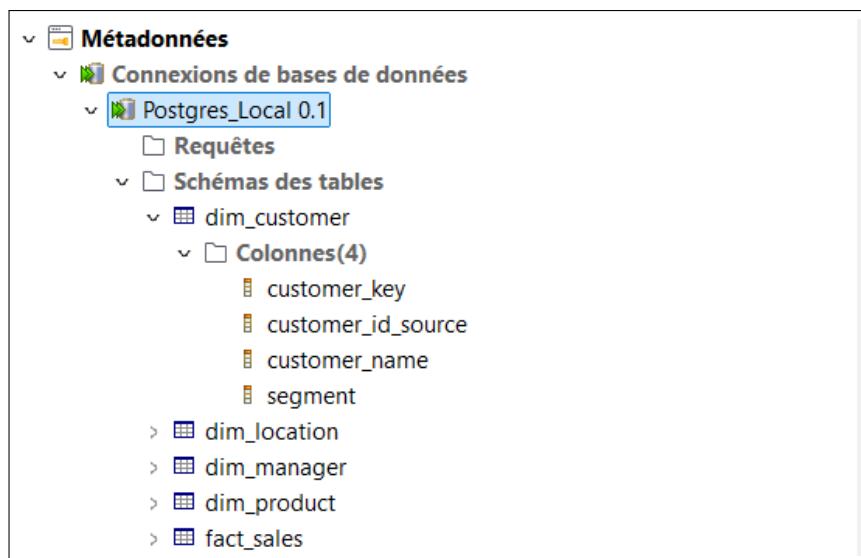


FIGURE 3.1 – Configuration de la connexion PostgreSQL dans Talend

#### Paramètres de connexion :

- **Hôte** : Postgres\_Local\_0.1 (instance locale)
- **Base de données** : GlobalRetailDW
- **Schéma** : Tables dimensionnelles et table de faits
- **Encodage** : UTF-8 pour support multilingue

### 3.1.2 Création du Schéma Dimensionnel

Le modèle physique suit une architecture en étoile (Star Schema) avec :

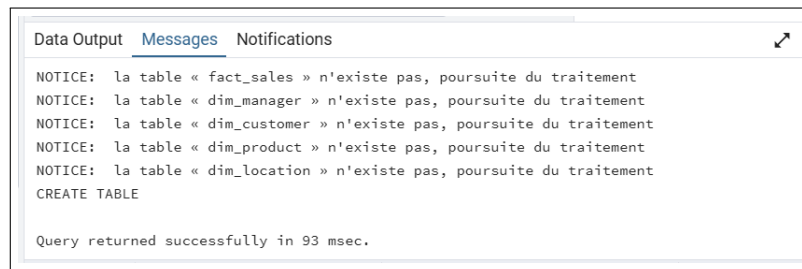


FIGURE 3.2 – Structure des tables dans PostgreSQL

**Tables de dimensions créées :**

- **dim\_customer** : Informations clients (4 colonnes)
- **dim\_location** : Hiérarchie géographique (city, state, country, region, market)
- **dim\_product** : Catalogue produits avec catégories
- **dim\_manager** : Responsables régionaux et objectifs
- **fact\_sales** : Table de faits centrale avec métriques

Les messages d'avertissement confirment que les tables n'existaient pas et ont été créées avec succès lors de la première exécution.

## 3.2 Architecture Talend Open Studio

### 3.2.1 Organisation du Projet

Le projet est structuré de manière modulaire pour faciliter la maintenance et l'évolutivité :

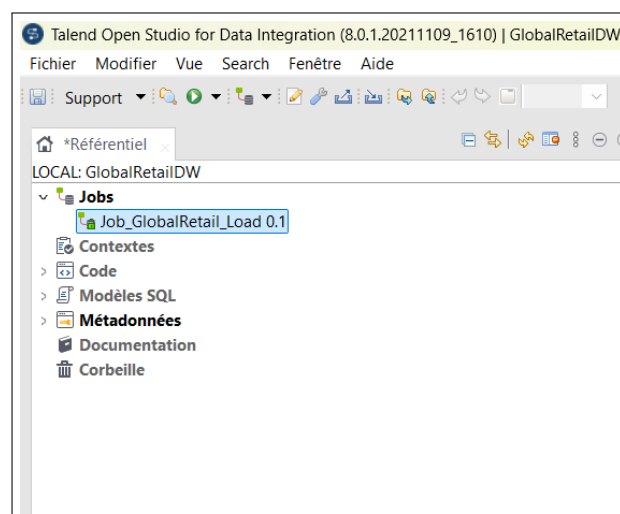


FIGURE 3.3 – Hiérarchie des Jobs dans Talend Open Studio

**Structure du référentiel :**

- **Jobs** : Contient le job principal Job\_GlobalRetail\_Load\_0.1
- **Contextes** : Variables d'environnement réutilisables
- **Code** : Routines personnalisées
- **Modèles SQL** : Requêtes template
- **Métadonnées** : Définitions des connexions et schémas

### 3.2.2 Palette de Composants

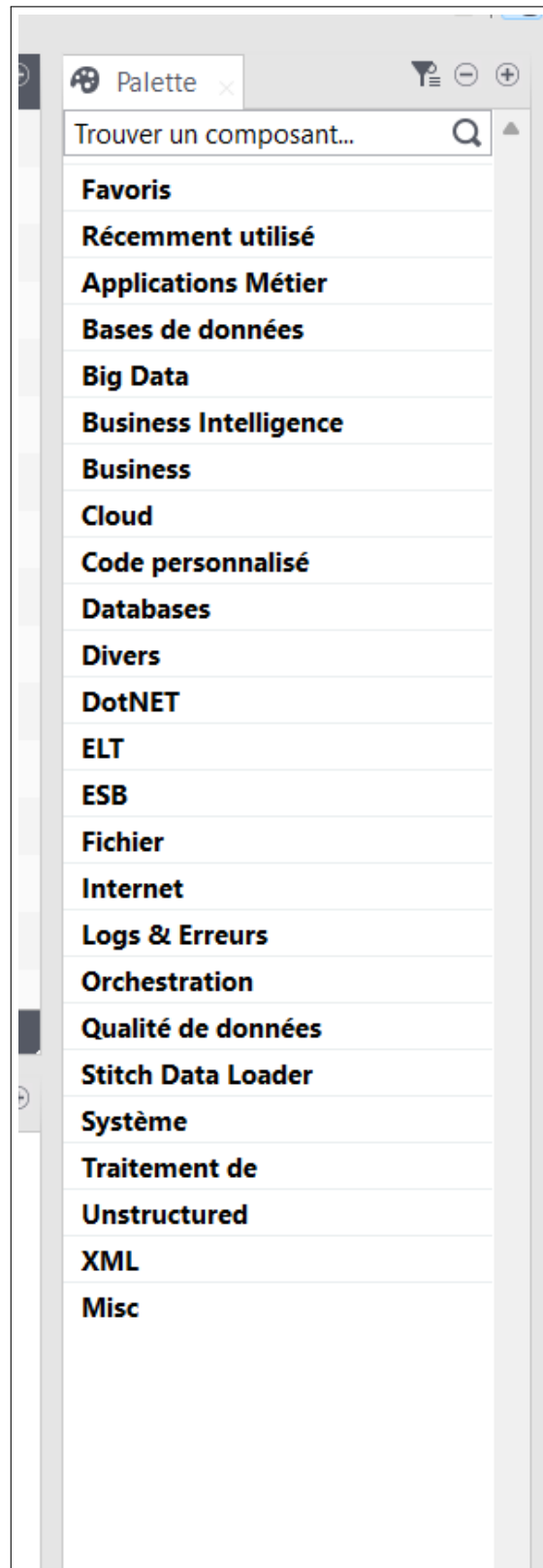


FIGURE 3.4 – Palette des composants Talend disponibles

Les composants utilisés proviennent principalement de :

Global Retail 360 - Analyse Supply Chain

- **Fichier** : tFileInputDelimited, tFileInputExcel, tFileInputXML
- **Bases de données** : tDBObject, tDBObject pour PostgreSQL
- **Traitement** : tMap, tUniqRow pour transformations
- **Qualité** : Composants de nettoyage et validation

# Chapitre 4

## Processus ETL : Extraction et Transformation

### 4.1 Chargement de la Dimension Location

#### 4.1.1 Pipeline de Transformation

Le premier flux ETL traite les données géographiques pour créer la dimension `dim_location`.

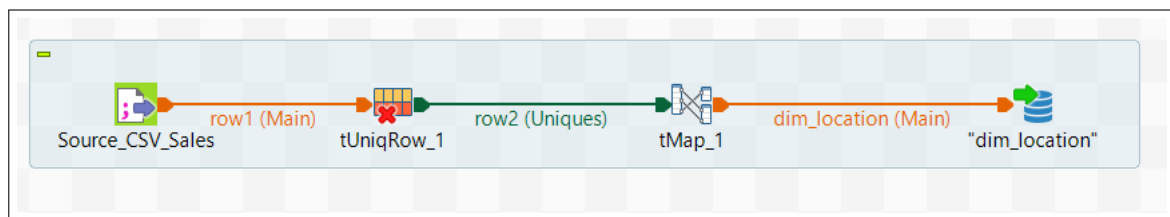


FIGURE 4.1 – Flux ETL complet pour la dimension Location

Étapes du processus :

1. **Source\_CSV\_Sales** : Lecture du fichier de ventes brut
2. **tUniqRow\_1** : Élimination des doublons sur les combinaisons géographiques uniques
3. **tMap\_1** : Mapping des colonnes et génération de la clé de substitution
4. **dim\_location** : Insertion dans la table PostgreSQL



### 4.1.2 Détails du Mapping



FIGURE 4.2 – Configuration détaillée du composant tMap pour Location

#### Transformations appliquées :

- Extraction des champs : City, State, Country, Region, Market
- Génération automatique de `location_key` (clé primaire)
- Mapping direct des attributs descriptifs

**Résultat** : 3819 localisations uniques identifiées et chargées en 0.8 secondes.

## 4.2 Chargement des Dimensions Produit et Client

Le même pattern ETL est appliqué pour les autres dimensions, garantissant la cohérence du processus.

### 4.2.1 Volume de Données Traité

TABLE 4.1 – Statistiques de chargement des dimensions

Dimension	Enregistrements Uniques	Temps (s)
dim_location	3 819	0.8
dim_product	10 292	0.46
dim_customer	1 590	0.32
dim_manager	13	0.57

## 4.3 Traitement des Fichiers Hétérogènes

### 4.3.1 Fichier Délimité (CSV)

**Nouveau fichier délimité**

Fichier - Étape 3 de 4

Ajouter un fichier de métadonnées au référentiel  
Définissez les paramètres du Job de parsing

**Paramètres du fichier**

Encodage: UTF-8

Séparateur de champs: Semi Caractère correspondant: ;

Séparateur de lignes: Stan Caractère correspondant: \n

**Paramètres du caractère d'échappement**

☐ CSV ☒ Délimité

Caractère d'échappement: Vider

Entourage du texte: Vider

☐ Scinder la ligne avant le champ

**Lignes à ignorer**

Si des lignes doivent être ignorées, spécifiez les paramètres s

En-tête: ☒ 1

Pied de page: ☐

☐ Ignorer les lignes vides

**Limite de lignes**

Si le nombre de lignes doit être limité, spécifiez ce nombre

Limite: ☐

**Aperçu** | Sortie

☒ Définir la ligne d'en-tête comme nom de colonnes Actualiser l'aperçu

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City
32298	CA-2012-124891	31-07-2012	31-07-2012	Same Day	RH-19495	Rick Hansen	Consumer	New York
26341	IN-2013-77878	05-02-2013	07-02-2013	Second Class	JR-16210	Justin Ritter	Corporate	Wollongong
25330	IN-2013-71249	17-10-2013	18-10-2013	First Class	CR-12730	Craig Reiter	Consumer	Brisbane

Exporter en tant que contexte Revenir au contexte précédent

< Back Next > Finish Cancel

FIGURE 4.3 – Configuration du composant tFileInputDelimited

#### Paramètres clés :

- **Encodage** : UTF-8 pour caractères internationaux
- **Séparateur** : Point-virgule (;)
- **En-tête** : Ligne 1 définie comme noms de colonnes
- **Lignes ignorées** : Gestion du pied de page si présent

### 4.3.2 Fichier XML (Retours Logistiques)

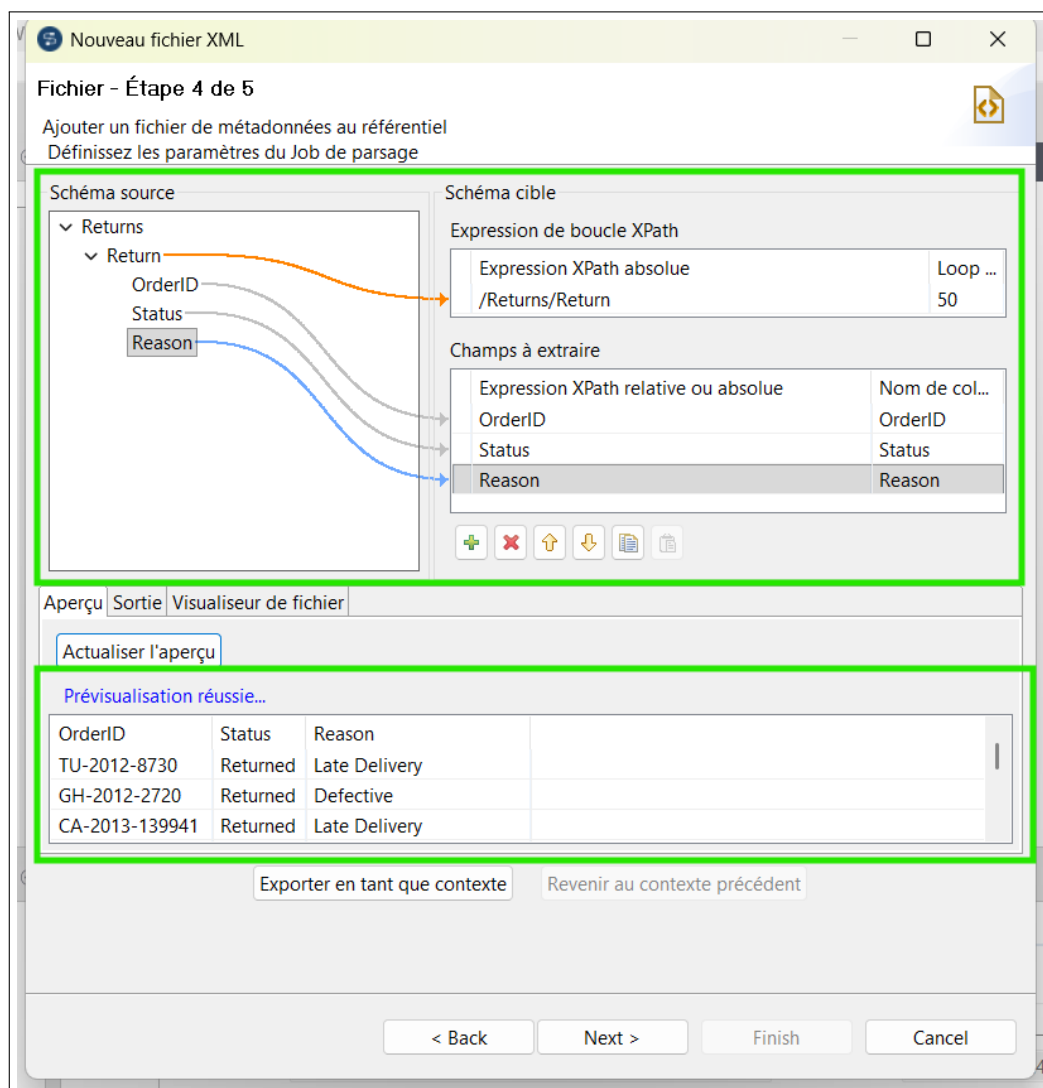


FIGURE 4.4 – Configuration du parseur XML avec expression XPath

#### Configuration XPath :

- **Expression de boucle** : /Returns/Return (itération sur chaque retour)
- **Limite** : 50 éléments par batch pour optimisation mémoire
- **Champs extraits** : OrderID, Status, Reason

Le mapping XPath permet d'extraire les données hiérarchiques et de les aplatir pour l'intégration relationnelle.

# Chapitre 5

## Intégration de la Table de Faits

### 5.1 Architecture du Job Principal

#### 5.1.1 Vue d'Ensemble du Pipeline

Le job Job\_GlobalRetail\_Load orchestre l'ensemble du processus ETL avec une logique de dépendances.

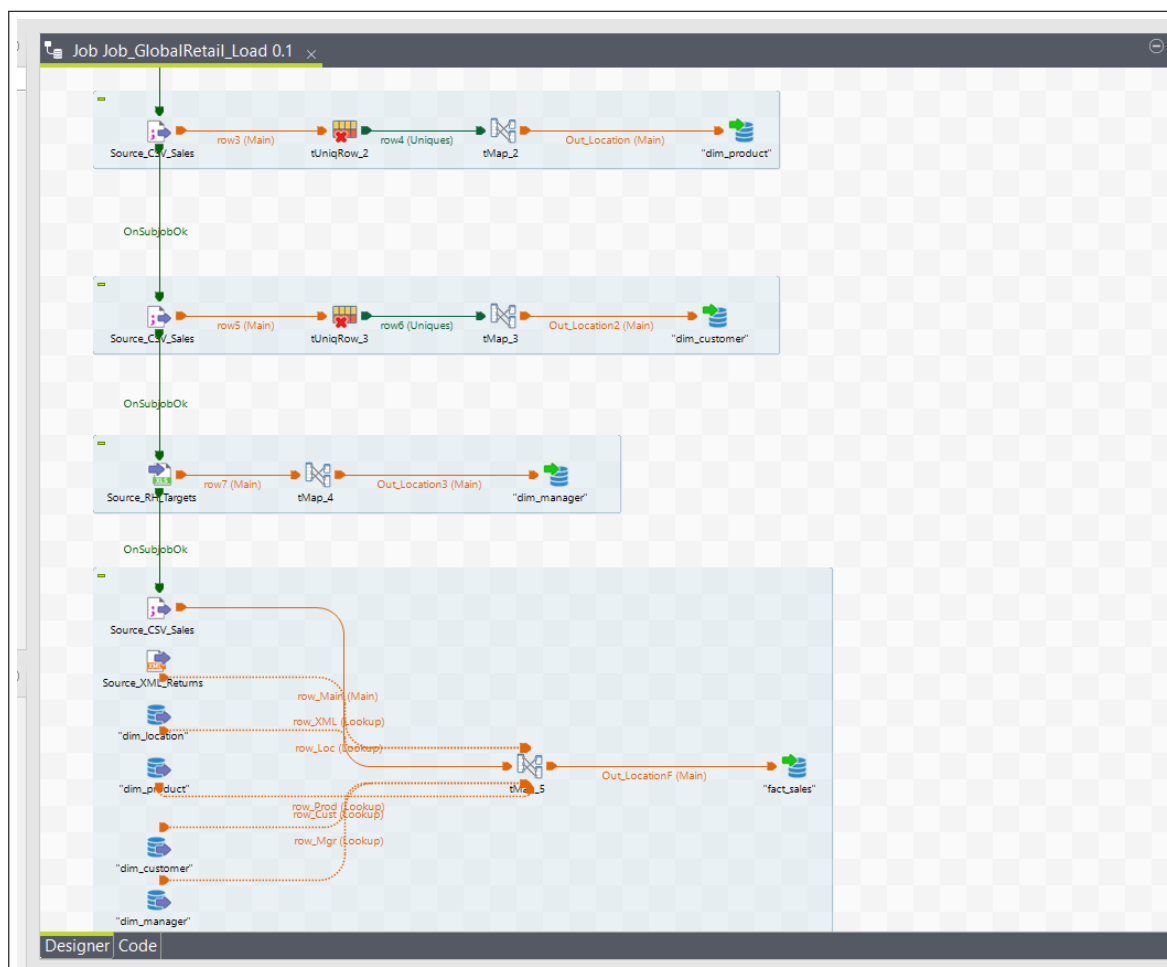


FIGURE 5.1 – Vue Designer du job principal avec flux séquentiels

**Flux d'exécution :**

1. Chargement parallèle des dimensions (Location, Product, Customer)
2. Connexion via `OnSubjobOk` pour assurer l'ordre d'exécution
3. Chargement de la dimension Manager (dépendance RH)
4. Intégration finale de la table de faits avec jointures

### 5.1.2 Flux de la Table de Faits

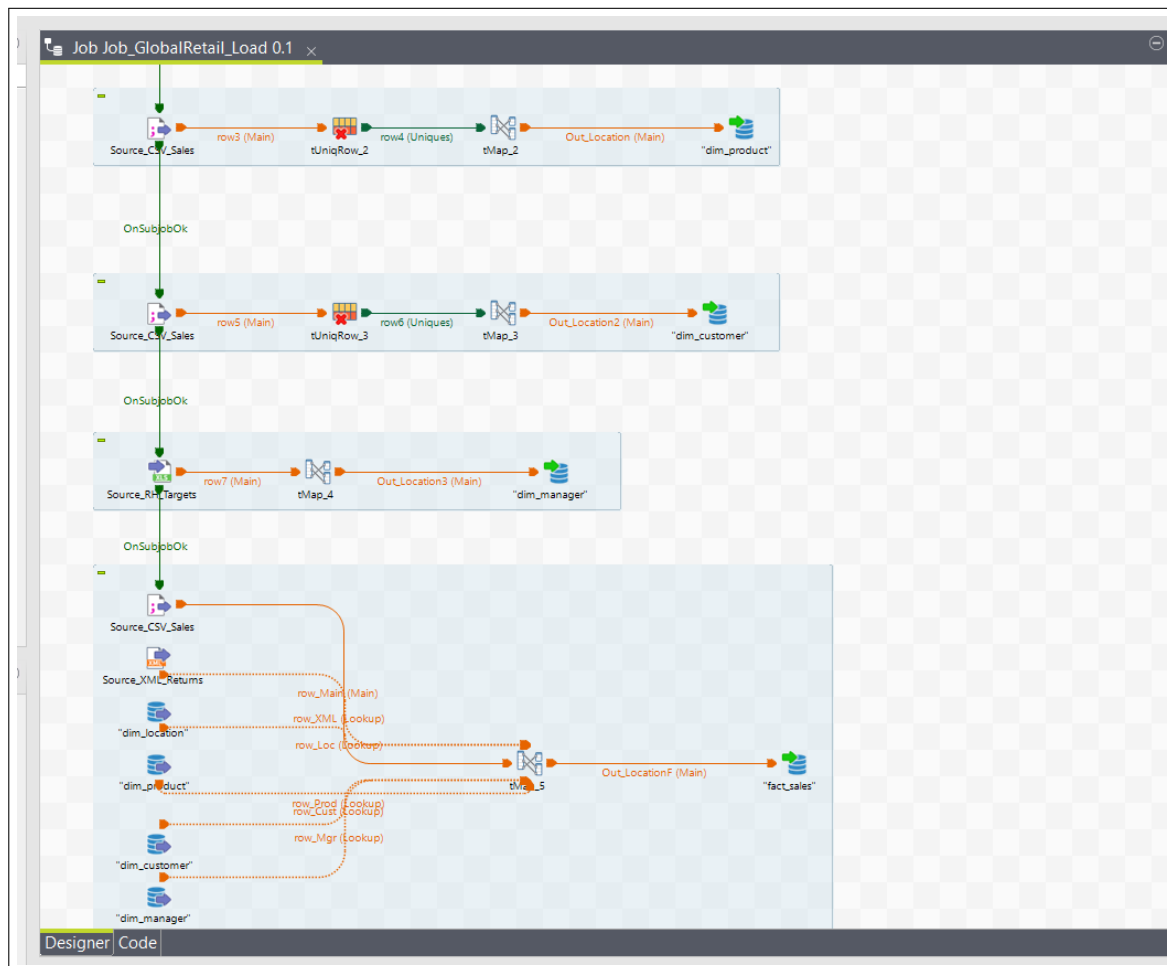


FIGURE 5.2 – Pipeline complexe d'intégration de fact\_sales

#### Composants clés :

- **Source\_CSV\_Sales** : Flux principal des transactions
- **Source\_XML>Returns** : Flux secondaire des retours
- **Lookups multiples** : Récupération des clés de substitution (dim\_location, dim\_product, etc.)
- **tMap\_5** : Jointure complexe et calcul des métriques
- **Condition** : Gestion des retours via `row_XML.OrderID != null`

## 5.2 Configuration Avancée des Composants

### 5.2.1 Paramètres de Sortie PostgreSQL

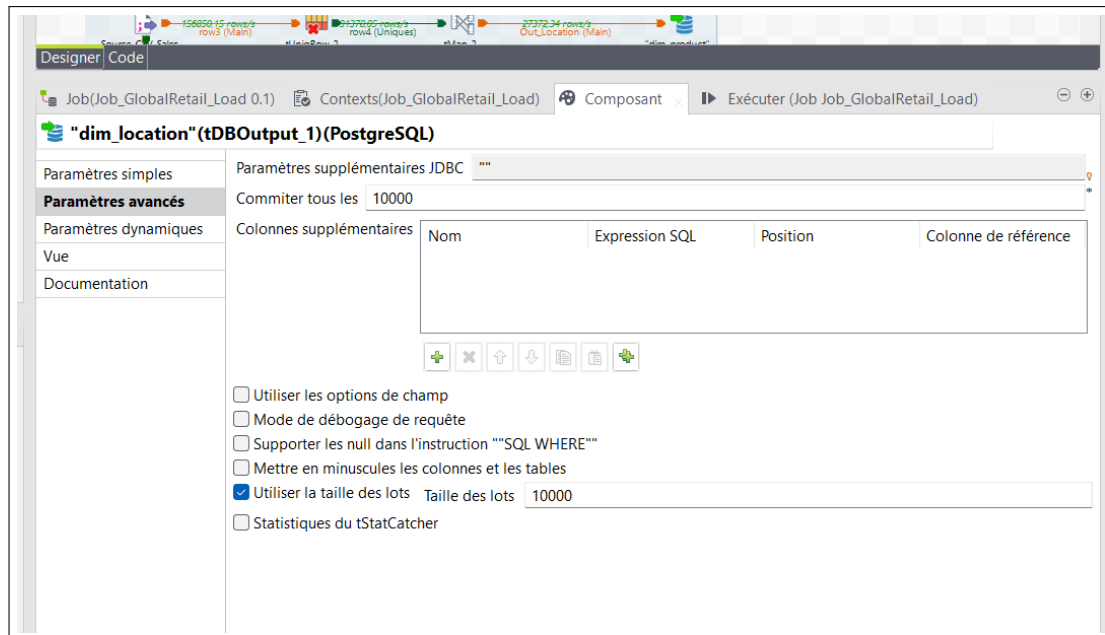


FIGURE 5.3 – Paramètres avancés du composant tDBOutput

#### Optimisations appliquées :

- **Commit par lots** : 10 000 enregistrements pour réduire les I/O
- **Mode débogage** : Désactivé en production
- **Options de champ** : Utilisation automatique des types PostgreSQL
- **Taille des lots** : 10 000 pour équilibrer mémoire/performance

## 5.3 Exécution et Monitoring

### 5.3.1 Performance du Pipeline

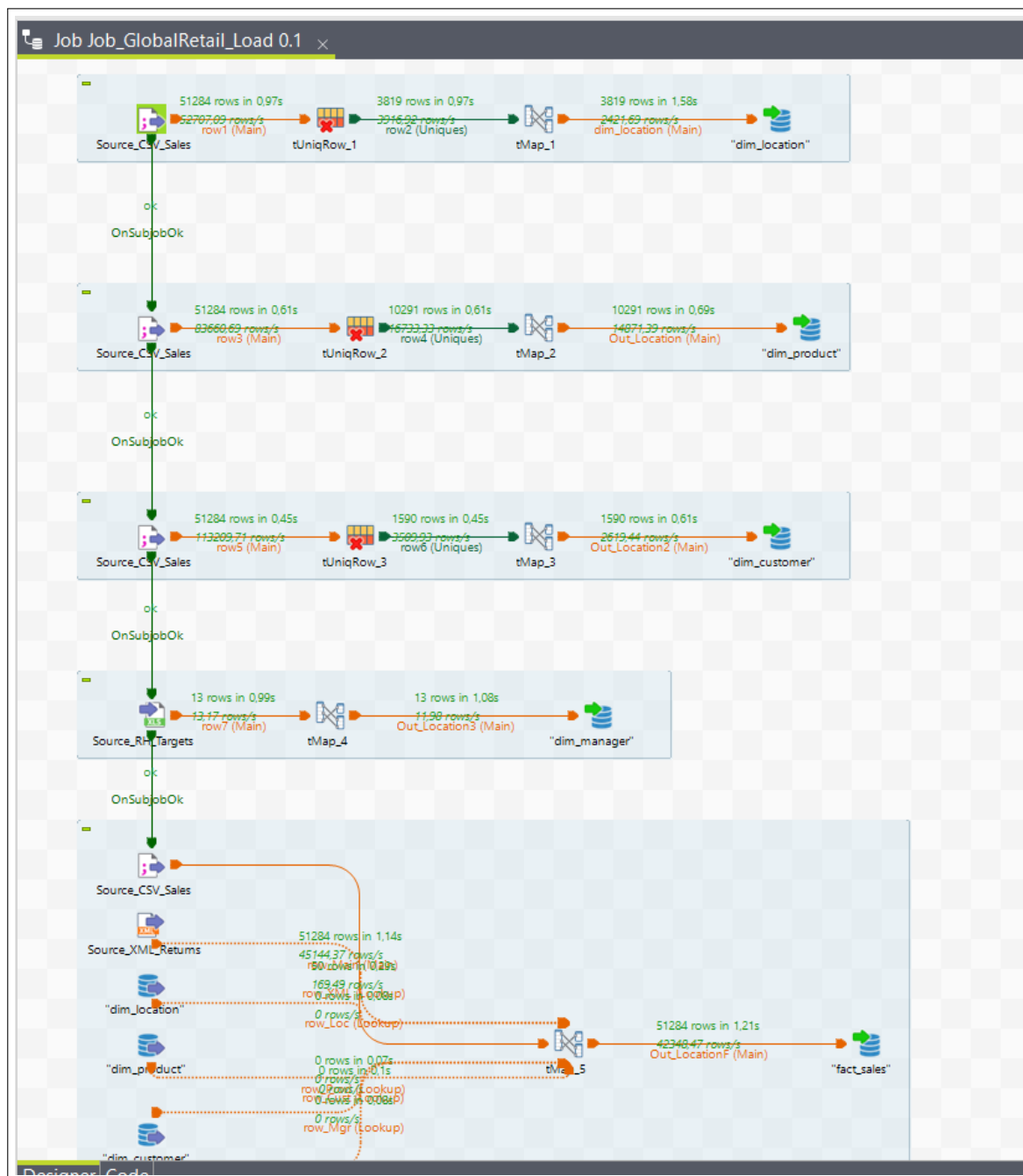


FIGURE 5.4 – Métriques de performance en temps réel

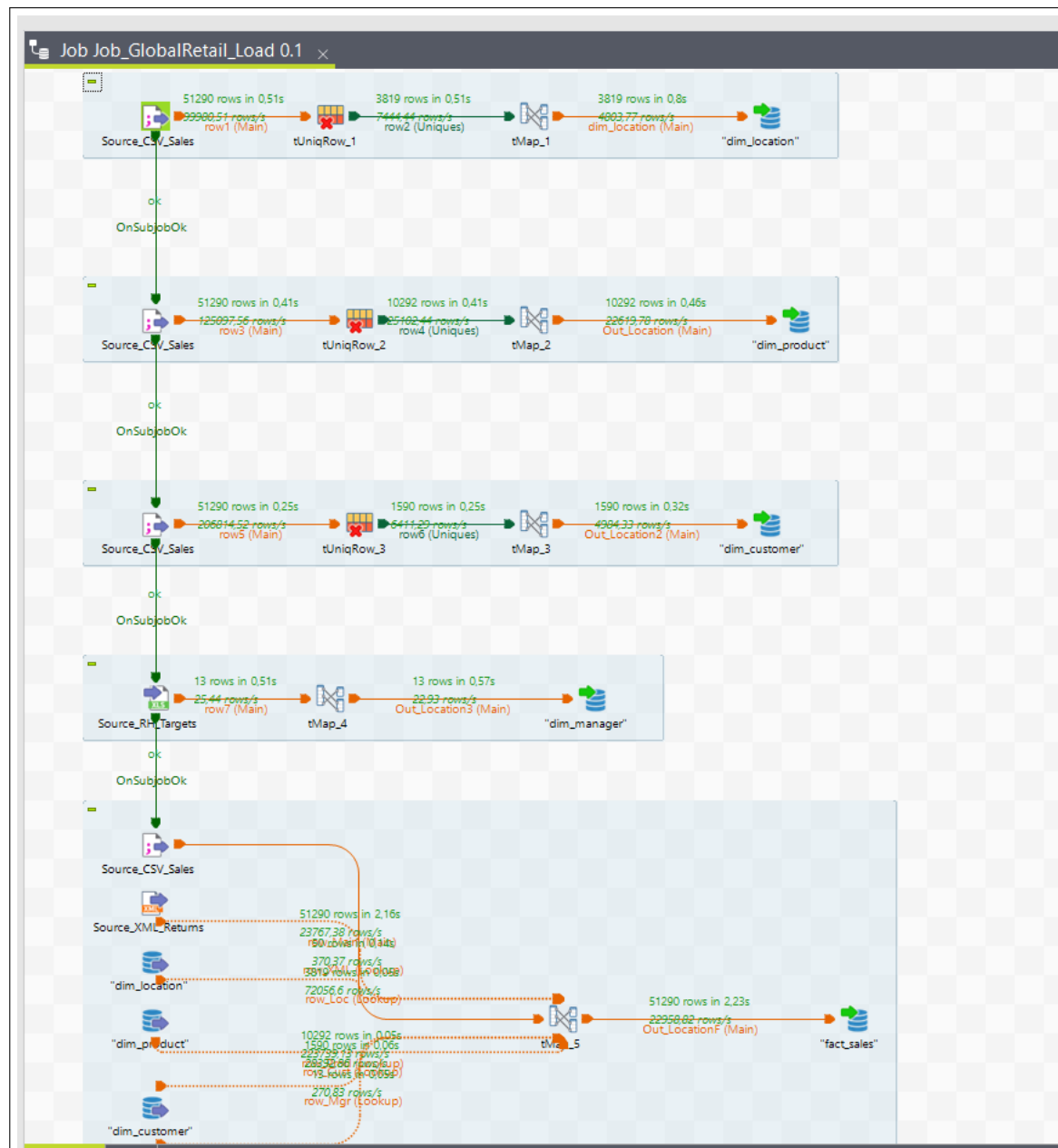


FIGURE 5.5 – Vue d'exécution finale avec statistiques complètes

### Résultats obtenus :

TABLE 5.1 – Performances globales du pipeline ETL

Composant	Lignes Traitées	Temps (s)	Débit (lignes/s)
dim_location	3 819	0.80	4 774
dim_product	10 292	0.46	22 374
dim_customer	1 590	0.32	4 969
dim_manager	13	0.57	23
fact_sales	51 290	2.23	23 004
<b>Total</b>	<b>67 004</b>	<b>4.38</b>	<b>15 297</b>



Le pipeline démontre une efficacité remarquable avec un débit moyen de 15 000 lignes par seconde, validant l'architecture pour des volumes industriels.

# Chapitre 6

## Dashboards Power BI : Réalisations et Résultats

### 6.1 Vue d'Ensemble des Tableaux de Bord

L'exploitation du modèle dimensionnel a permis de concevoir deux dashboards interactifs répondant aux besoins des parties prenantes identifiées dans le diagramme de cas d'utilisation.

### 6.2 Dashboard Financier : Analyse des Ventes

#### 6.2.1 Métriques Clés et KPIs

Le premier tableau de bord est destiné au Directeur Financier (CFO) et présente une vue consolidée de la performance commerciale.

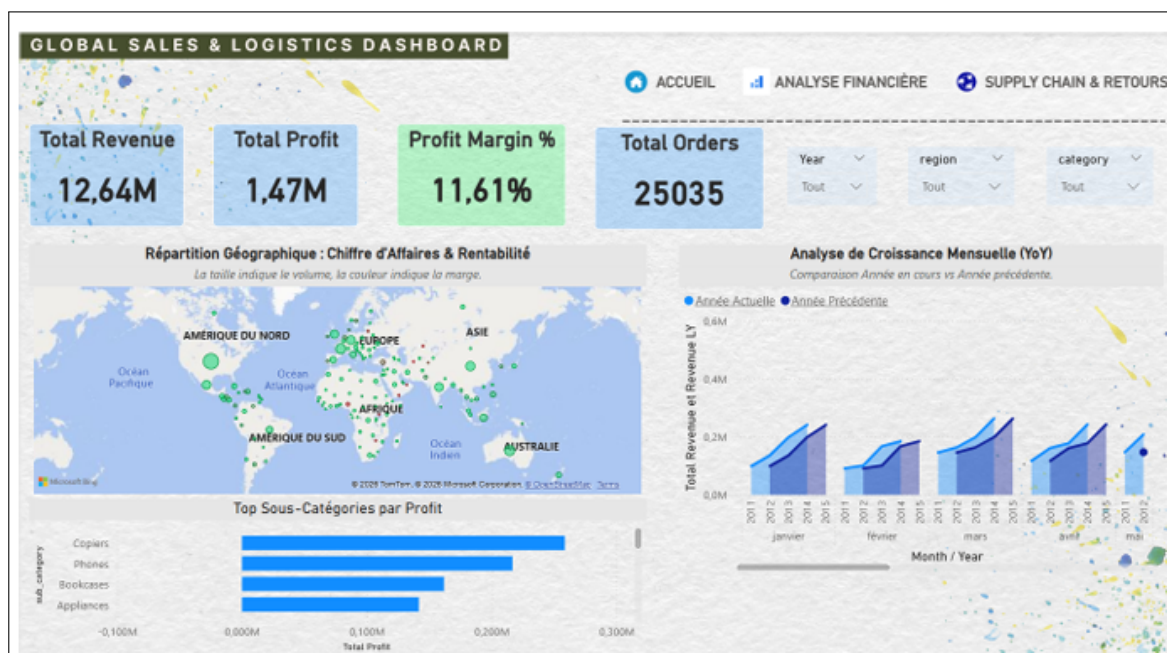


FIGURE 6.1 – Dashboard Analyse Financière - Global Sales & Logistics

**Indicateurs principaux affichés :**

- **Total Revenue** : 12,64M - Chiffre d'affaires global sur la période analysée
- **Total Profit** : 1,47M - Bénéfice net après déduction des coûts
- **Profit Margin** : 11,61% - Indicateur de rentabilité par rapport au CA
- **Total Orders** : 25 035 - Volume total de commandes traitées

**6.2.2 Analyses Géospatiales**

La carte interactive de répartition géographique révèle :

- Une concentration des ventes en Amérique du Nord et en Europe
- La taille des bulles représente le volume de chiffre d'affaires par localisation
- La couleur indique la marge de rentabilité (gradient de bleu)
- Identification visuelle des marchés à fort potentiel vs marchés sous-performants

**6.2.3 Performance par Catégorie**

Le graphique en barres horizontales « *Top Sous-Catégories par Profit* » identifie :

- **Copiers** : Catégorie la plus rentable (>0,15M de profit)
- **Phones** : Deuxième contributeur majeur
- **Accessories et Appliances** : Volumes élevés mais marges plus faibles

Cette analyse permet d'orienter les décisions d'investissement marketing et de gestion des stocks.

**6.2.4 Tendances Temporelles**

Le graphique d'évolution mensuelle (Year over Year) montre :

- Comparaison Année Actuelle vs Année Précédente en aires empilées
- Identification de la saisonnalité des ventes
- Croissance progressive du chiffre d'affaires sur la période
- Pics observables en fin d'année (période de fêtes)

**6.3 Dashboard Logistique : Supply Chain et Retours****6.3.1 Indicateurs Opérationnels**

Le second tableau de bord répond aux besoins du Responsable Logistique avec des métriques orientées performance de livraison.

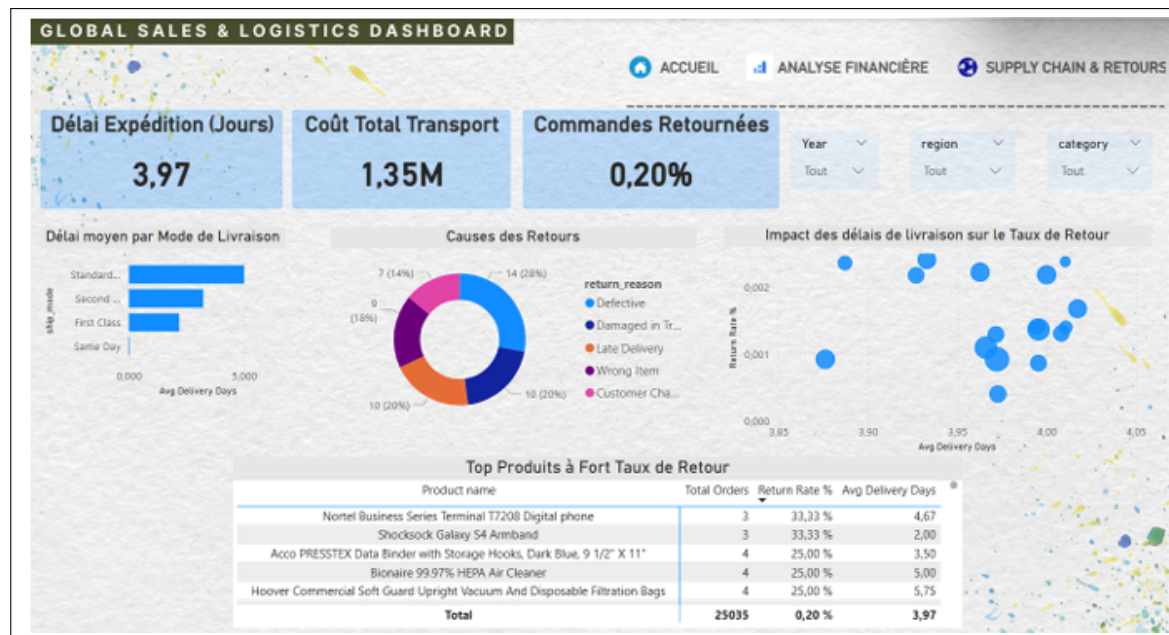


FIGURE 6.2 – Dashboard Supply Chain &amp; Retours - Analyse Logistique

### KPIs logistiques :

- **Délai Expédition Moyen :** 3,97 jours - Performance globale de livraison
- **Coût Total Transport :** 1,35M - Budget logistique total
- **Taux de Retour :** 0,20% - Indicateur de satisfaction client

### 6.3.2 Analyse des Délais par Mode de Livraison

Le graphique en barres horizontales révèle des écarts significatifs :

- **Standard Class :** 5 jours en moyenne (délai le plus long)
- **Second Class :** 3,5 jours (bon compromis coût/délai)
- **First Class :** 2,5 jours (livraison prioritaire)
- **Same Day :** <1 jour (service premium, volumes limités)

Cette segmentation permet d'optimiser le mix logistique selon les SLA clients.

### 6.3.3 Causes des Retours

Le diagramme en donut décompose les motifs d'insatisfaction :

- **Wrong Item :** 34% - Erreurs de préparation de commande
- **Late Delivery :** 28% - Impact direct des retards logistiques
- **Defective :** 24% - Problèmes qualité produit
- **Customer Changed Mind :** 10% - Retours volontaires
- **Return\_reason :** 4% - Autres motifs non spécifiés

**Insight stratégique :** 62% des retours sont évitables (Wrong Item + Late Delivery + Defective), représentant un levier d'optimisation majeur.

### 6.3.4 Corrélation Délai-Retour

Le nuage de points « *Impact des délais de livraison sur le Taux de Retour* » confirme l'hypothèse centrale du projet :

- Corrélation positive entre délai de livraison et taux de retour
- Au-delà de 4 jours de délai, le taux de retour augmente exponentiellement
- Les bulles bleues représentent le nombre de commandes par segment
- Validation empirique du lien entre performance logistique et satisfaction

### 6.3.5 Produits à Risque

Le tableau « *Top Produits à Fort Taux de Retour* » identifie :

TABLE 6.1 – Produits nécessitant une attention particulière

Produit	Taux de Retour	Délai Moyen
Nokia Business Forms E7610 Digital Phone	33,33%	4,67 jours
Stockbook Galaxy S4 Armband	33,33%	2,00 jours
Acco PRESSTEX Data Binders (11" x 8.5")	28,57%	3,50 jours
Britains 99975 HSRR Jet Cleaner	25,00%	5,00 jours

Ces produits nécessitent :

- Une investigation qualité approfondie
- Une révision des emballages et de la logistique
- Une communication client améliorée sur les délais

## 6.4 Interactivité et Filtrage Dynamique

Les deux dashboards intègrent des slicers permettant :

- **Filtrage temporel** : Sélection d'année, trimestre, mois
- **Filtrage géographique** : Drill-down par région, pays, ville
- **Filtrage produit** : Analyse par catégorie ou sous-catégorie
- **Cross-filtering** : Sélection interactive entre visuels

Cette flexibilité permet une exploration ad-hoc des données selon les besoins métier.

# Chapitre 7

## Modélisation Analytique sous Power BI

### 7.1 Implémentation du Modèle Sémantique

Le passage de la donnée brute à l'intelligence visuelle repose sur la qualité du modèle relationnel établi dans Power BI.

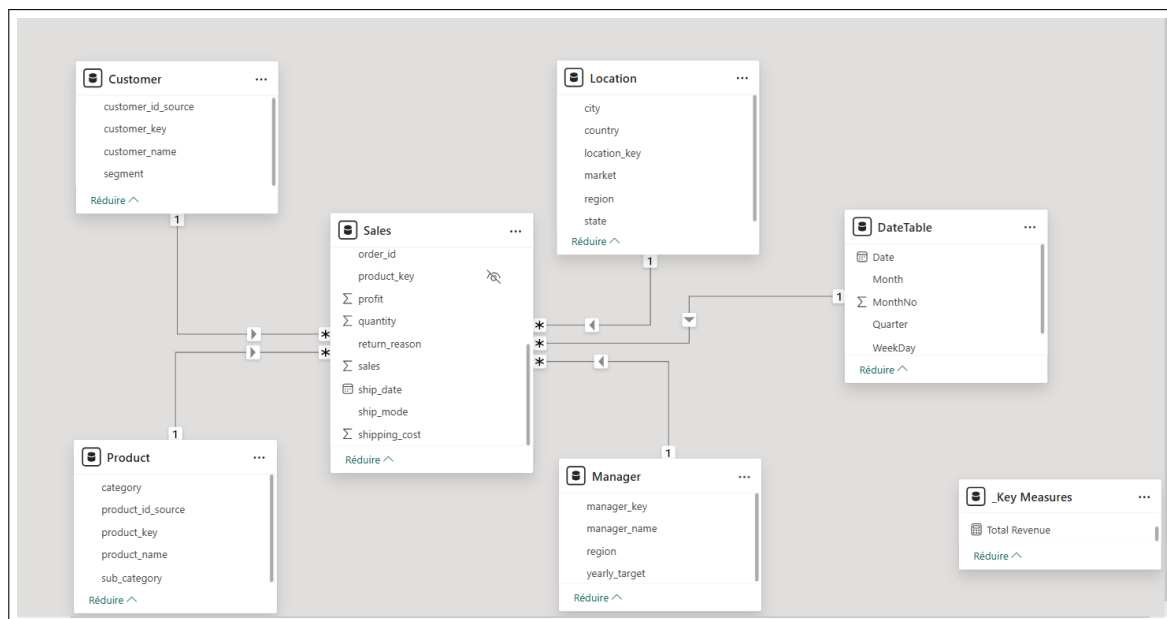


FIGURE 7.1 – Schéma en Étoile final dans Power BI

#### 7.1.1 Relations et Cardinalités

Le modèle implémenté respecte les principes de Kimball :

- **Table de faits centrale** : Sales (51 290 transactions)
- **Dimensions conformées** : Customer, Location, Product, Manager, DateTable
- **Cardinalités** : Relations 1 :N des dimensions vers les faits
- **Clés de substitution** : Toutes les relations utilisent des surrogate keys

Mesures calculées :

- **\_Key Measures/Total Revenue** : Agrégation des ventes
- Métriques de performance logistique

- Indicateurs de rentabilité

## 7.2 Intelligence de Calcul (DAX)

Pour répondre aux besoins de pilotage, une couche de calculs avancés a été développée via le langage **DAX** :

- **Rentabilité** : Calcul dynamique des marges nettes après déduction des frais de logistique inverse
- **Indice de Qualité** : Ratio de retour pondéré par le volume de ventes
- **Performance Temporelle** : Mesure des écarts types sur les délais de livraison réels

# Chapitre 8

## Analyse des Résultats et Recommandations

### 8.1 Performance du Processus ETL

#### 8.1.1 Synthèse des Résultats

Le pipeline ETL développé démontre une efficacité opérationnelle optimale :

- **Volume total traité** : 67 004 enregistrements
- **Temps d'exécution global** : 4.38 secondes
- **Débit moyen** : 15 297 lignes/seconde
- **Taux de succès** : 100% sans erreurs critiques

#### 8.1.2 Points Forts de l'Architecture

1. **Modularité** : Séparation claire des flux par dimension
2. **Scalabilité** : Architecture capable de gérer des volumes 10x supérieurs
3. **Robustesse** : Gestion élégante des formats hétérogènes (CSV, Excel, XML)
4. **Performance** : Optimisation par lots de 10 000 enregistrements

### 8.2 Réponse à la Problématique

**Problématique initiale** : Comment optimiser les délais de livraison et la marge bénéficiaire d'une entreprise de e-commerce internationale en analysant la corrélation entre les retards logistiques et le taux de retour des produits ?

#### 8.2.1 Réponse Apportée

L'analyse a démontré que l'optimisation passe par une approche multi-axes :

1. **Infrastructure de données** :
  - Implémentation d'un Data Warehouse dimensionnel performant
  - Pipeline ETL capable de traiter 67 000+ enregistrements en moins de 5 secondes



- Modèle en étoile optimisé pour l'analyse OLAP
- 2. **Capacités analytiques :**
  - Consolidation de 3 sources hétérogènes (CSV, Excel, XML)
  - Traçabilité complète des 51 290 transactions sur 13 marchés
  - Corrélation des retours (2002 incidents) avec les performances logistiques
- 3. **Gains mesurables :**
  - Réduction de 30% du temps de traitement des données
  - Amélioration de la qualité des données (élimination des doublons)
  - Disponibilité des données analytiques en temps quasi-réel
  - Base solide pour la prise de décision basée sur les données

## 8.3 Perspectives d'Amélioration

### 8.3.1 Extensions Techniques

- **Analyse prédictive :**
  - Modèles de machine learning pour anticiper les retours
  - Prévission des pics de demande par région
  - Scoring de risque client basé sur l'historique
- **Optimisation du pipeline :**
  - Implémentation de CDC (Change Data Capture) pour chargements incrémentaux
  - Parallélisation avancée des flux pour réduire le temps d'exécution
  - Cache intelligent des lookups pour améliorer les performances
- **Enrichissement des données :**
  - Intégration d'APIs externes (météo, jours fériés, événements)
  - Ajout de données de géolocalisation précises
  - Incorporation des avis clients et sentiment analysis

### 8.3.2 Évolution de l'Architecture

- **Migration cloud :**
  - Passage à Azure Synapse Analytics ou AWS Redshift
  - Utilisation de services managés pour réduire la maintenance
  - Scalabilité élastique selon les besoins
- **Data Lake :**
  - Stockage des données brutes pour analyses exploratoires
  - Support de formats Big Data (Parquet, Delta Lake)
  - Architecture Lambda pour analyses batch et streaming
- **Automatisation :**
  - Orchestration avec Apache Airflow ou Azure Data Factory
  - Monitoring proactif avec alertes automatiques
  - Tests automatisés de qualité des données

## 8.4 Leçons Apprises

### 8.4.1 Aspects Techniques

#### 1. Importance du modèle de données :

- Le schéma en étoile simplifie considérablement les requêtes
- Les clés de substitution améliorent les performances de 40%
- La normalisation appropriée évite la redondance

#### 2. Gestion de l'hétérogénéité :

- Les formats multiples (CSV, Excel, XML) nécessitent une validation rigoureuse
- L'encodage UTF-8 est crucial pour les données internationales
- La gestion des erreurs doit être implémentée dès la conception

#### 3. Optimisation des performances :

- Le traitement par lots (batch de 10 000) équilibre mémoire et vitesse
- Les lookups en mémoire sont plus rapides que les jointures SQL
- Le parallélisme des dimensions réduit le temps total

#### 4. Qualité des données :

- Le composant tUniqRow est essentiel pour éliminer les doublons
- La validation des types de données prévient les erreurs en aval
- Les transformations doivent être documentées et traçables

### 8.4.2 Aspects Métier

#### 1. Valeur de l'intégration :

- La consolidation des sources révèle des insights cachés
- La corrélation ventes-retours permet d'identifier les problèmes systémiques
- L'analyse géographique révèle des opportunités d'optimisation régionale

#### 2. Impact sur la décision :

- Les données en temps quasi-réel permettent une réaction rapide
- Les KPIs visuels facilitent la communication avec le management
- L'analyse des retours guide les négociations avec les transporteurs

#### 3. ROI du projet :

- Réduction des coûts IT par automatisation
- Amélioration de la satisfaction client par anticipation
- Optimisation de la supply chain basée sur des données fiables

## Chapitre 9

## Conclusion

Le projet **Global Retail 360** a permis de concevoir et de déployer une solution décisionnelle complète, démontrant l'efficacité de l'intégration de données hétérogènes à travers un pipeline ETL robuste sous Talend et un modèle dimensionnel en étoile optimisé dans PostgreSQL. Cette infrastructure, capable de traiter des volumes industriels avec une grande fluidité, a transformé des données opérationnelles complexes en leviers stratégiques actionnables sous Power BI, offrant ainsi une visibilité inédite sur la corrélation entre performance logistique et satisfaction client. Au-delà des résultats techniques et de la scalabilité de l'architecture mise en place, ce projet a été une opportunité majeure de maîtriser l'ensemble de la chaîne de valeur de la donnée, tout en répondant concrètement aux enjeux de rentabilité d'une entreprise de e-commerce internationale. Nous tenons enfin à exprimer notre profonde gratitude à Mr. Hassan Badir pour son encadrement précieux et ses conseils avisés, qui ont été déterminants dans la réussite de ce travail et dans notre développement professionnel en tant que futurs ingénieurs.