

Rapport sur l'Optimisation de la Gestion des Données e-Commerce

Imad elmasner

January 26, 2025

Contents

1	Introduction	5
1.1	Contexte du Projet	5
2	Objectifs du Projet	7
3	Méthodologie	9
3.1	Extraction des Données	9
3.1.1	Identification des Sources de Données	9
3.1.2	Mise en Place d'un Processus d'Extraction Automatisé	9
3.2	Transformation des Données.....	10
3.2.1	Nettoyage et Normalisation des Données	10
3.2.2	Application de Règles de Transformation.....	10
3.2.3	Implémentation des SCD	10
3.3	Chargement des Données.....	11
3.3.1	Chargement des Données Transformées.....	11
3.3.2	Utilisation des SCD	11
4	Visualisation des Données	13
4.1	Création de Tableaux de Bord Interactifs.....	13
4.2	Analyse des Performances et Tendances des Ventes.....	13
5	Résultats	15
5.1	Amélioration de la Précision des Données.....	15
5.2	Réduction du Temps Nécessaire pour Générer des Rapports.....	15
5.3	Meilleure Prise de Décision Grâce aux Visualisations Claires.....	15
6	Conclusion	17
7	Recommandations	19
7.1	Continuer à Affiner les Processus ETL	19
7.2	Former le Personnel à l'Utilisation Avancée de Power BI.....	19
7.3	Assurer la Sécurité et la Conformité.....	19

Chapter 1

Introduction

Ce rapport présente une analyse approfondie du projet d'optimisation de la gestion des données dans le secteur e-commerce. Il met en œuvre des techniques avancées telles que l'ETL (Extract, Transform, Load), les SCD (Slowly Changing Dimensions) et des outils de visualisation comme Power BI. L'objectif principal est d'améliorer la qualité, la précision et la disponibilité des données afin de faciliter la prise de décision stratégique.

1.1 Contexte du Projet

Le projet vise à structurer les données provenant de différentes sources (CSV, JSON) et à les charger dans un entrepôt de données SQL Server. Ensuite, nous utiliserons Talend pour effectuer les opérations ETL et implémenter les mécanismes de gestion des Slowly Changing Dimensions (SCD). Enfin, Power BI sera utilisé pour créer des tableaux de bord interactifs permettant d'analyser les performances et les tendances des ventes.

Chapter 2

Objectifs du Projet

- Amélioration de la gestion des données : Structurer et organiser les données provenant de différentes sources pour une meilleure utilisation.
- Automatisation des processus ETL : Mettre en place un processus automatisé pour extraire, transformer et charger les données dans un entrepôt de données.
- Gestion des changements temporels : Utiliser les SCD pour suivre et gérer efficacement les modifications des données au fil du temps.
- Visualisation des données : Créer des tableaux de bord interactifs avec Power BI pour analyser les performances et les tendances des ventes.

Chapter 3

Méthodologie

3.1 Extraction des Données

3.1.1 Identification des Sources de Données

Les principales sources de données comprennent :

- Fichiers CSV (ecommerce-data-csv.csv)
- Fichiers JSON (ecommerce-data.json)

Action :

- Analyser les fichiers pour comprendre la structure et les types de données disponibles.
- Identifier les champs pertinents pour l'extraction.

3.1.2 Mise en Place d'un Processus d'Extraction Automatisé

Action :

- Développer un job Talend pour chaque source de données.
- Pour le fichier CSV, utiliser le composant tFileInputDelimited pour lire les données.
- Pour le fichier JSON, utiliser le composant tFileInputJSON.

Listing 3.1: Exemple de code Talend pour lire un fichier CSV

```
// Exemple de code Talend pour lire un fichier CSV  
tFileInputDelimited_1 -> tMap_1
```

Listing 3.2: Exemple de code Talend pour lire un fichier JSON

```
// Exemple de code Talend pour lire un fichier JSON  
tFileInputJSON_1 -> tMap_1
```

3.2 Transformation des Données

3.2.1 Nettoyage et Normalisation des Données

Action :

- Supprimer les valeurs manquantes ou incorrectes dans les colonnes pertinentes.
- Convertir les types de données pour assurer la cohérence (par exemple, convertir des chaînes de caractères en dates).

Listing 3.3: Exemple de nettoyage de données dans Talend

```
// Exemple de nettoyage de données dans Talend  
tMap_1 -> tFilterRow_1
```

3.2.2 Application de Règles de Transformation

Action :

- Harmoniser les formats de données (ex. : uniformisation des noms de produits).
- Ajouter de nouvelles colonnes basées sur des règles métier (ex. : calcul des marges bénéficiaires).

Listing 3.4: Exemple de transformation de données dans Talend

```
// Exemple de transformation de données dans Talend  
tMap_1 -> tAggregateRow_1
```

3.2.3 Implémentation des SCD

Type 1 SCD (Overwrite)

Correction des erreurs de saisie sans conserver l'historique.

- Exemple : correction d'une adresse erronée d'un client.

Listing 3.5: Exemple de Type 1 SCD dans Talend

```
// Exemple de Type 1 SCD dans Talend  
tMap_1 -> tUpdate_1
```


Type 2 SCD (Versioning)

Conservation des changements historiques avec des colonnes de date de début/fin et d'indicateur de ligne active.

- Exemple : suivi des changements de prix d'un produit.

Listing 3.6: Exemple de Type 2 SCD dans Talend

```
// Exemple de Type 2 SCD dans Talend  
tMap_1 -> tSlowChange_1
```

Type 3 SCD (Tracking a Few Changes)

Suivi limité des modifications via des colonnes supplémentaires.

- Exemple : ajout d'une colonne "Valeur Précédente" pour suivre les changements mineurs.

Listing 3.7: Exemple de Type 3 SCD dans Talend

```
// Exemple de Type 3 SCD dans Talend  
tMap_1 -> tMap_2
```

3.3 Chargement des Données

3.3.1 Chargement des Données Transformées

Action :

- Utiliser le composant tJDBCOutput pour charger les données transformées dans SQL Server.
- Vérifier l'intégrité des données chargées.

Listing 3.8: Exemple de chargement de données dans Talend

```
// Exemple de chargement de données dans Talend  
tMap_1 -> tJDBCOutput_1
```

3.3.2 Utilisation des SCD

Action :

- Configurer les tables dimensionnelles dans SQL Server pour prendre en charge les différents types de SCD.
- Mise à jour régulière des dimensions critiques comme CustomerDimension et ProductDimension.

4.1 Création de Tableaux de Bord Interactifs

Action :

- Importer les données du data mart dans Power BI.
- Créer des graphiques variés (lignes, barres, camemberts, tableaux, cartes) pour illustrer les données.

Listing 4.1: Exemple de requête SQL pour extraire les données nécessaires

-- Exemple de requête SQL pour extraire les données nécessaires

```
SELECT * FROM SalesFact  
JOIN TimeDimension ON SalesFact.TimeID = TimeDimension.TimeID  
JOIN ProductDimension ON SalesFact.ProductID = ProductDimension.ProductID  
JOIN CustomerDimension ON SalesFact.CustomerID = CustomerDimension.CustomerID
```

4.2 Analyse des Performances et Tendances des Ventes

Action :

- Analyser les tendances des ventes, segmentation des clients, impact des réductions et performance des transporteurs.
- Prévisions de la demande et analyse des niveaux d'inventaire.

Chapter 4

Visualisation des Données

Chapter 5

Résultats

5.1 Amélioration de la Précision des Données

Résultat :

- Réduction significative des erreurs de données grâce aux processus de nettoyage et de validation.
- Meilleure harmonisation des formats de données entre différentes sources.

5.2 Réduction du Temps Nécessaire pour Générer des Rapports

Résultat :

- Automatisation des processus ETL réduisant le temps nécessaire pour préparer les données.
- Accélération des requêtes grâce à l'indexation et au partitionnement des tables.

5.3 Meilleure Prise de Décision Grâce aux Visualisations Claires

Résultat :

- Tableaux de bord interactifs facilitant la compréhension des tendances et des performances.
- Prises de décisions plus informées et stratégiques basées sur des analyses précises.

Chapter 6

Conclusion

Le projet d'optimisation de la gestion des données e-commerce a prouvé l'importance des techniques ETL et SCD, ainsi que l'efficacité de Power BI pour la visualisation des données. Ces méthodes permettent une gestion plus agile et précise des données, essentielle dans le secteur du commerce électronique.

7.1 Continuer à Affiner les Processus ETL

Recommandation :

- Intégrer de nouvelles sources de données pour enrichir les analyses.
- Optimiser les performances des jobs Talend pour traiter de plus grandes quantités de données.

7.2 Former le Personnel à l'Utilisation Avancée de Power BI

Recommandation :

- Organiser des sessions de formation pour maximiser les bénéfices des visualisations.
- Encourager l'innovation dans la création de nouveaux tableaux de bord et rapports.

7.3 Assurer la Sécurité et la Conformité

Recommandation :

- Renforcer les mesures de sécurité pour protéger les données sensibles.
- S'assurer de la conformité avec les réglementations telles que le RGPD.