YOUCODE

Rapport de Projet

Optimisation de la Gestion des Données pour une Plateforme E-commerce

YOUCODE

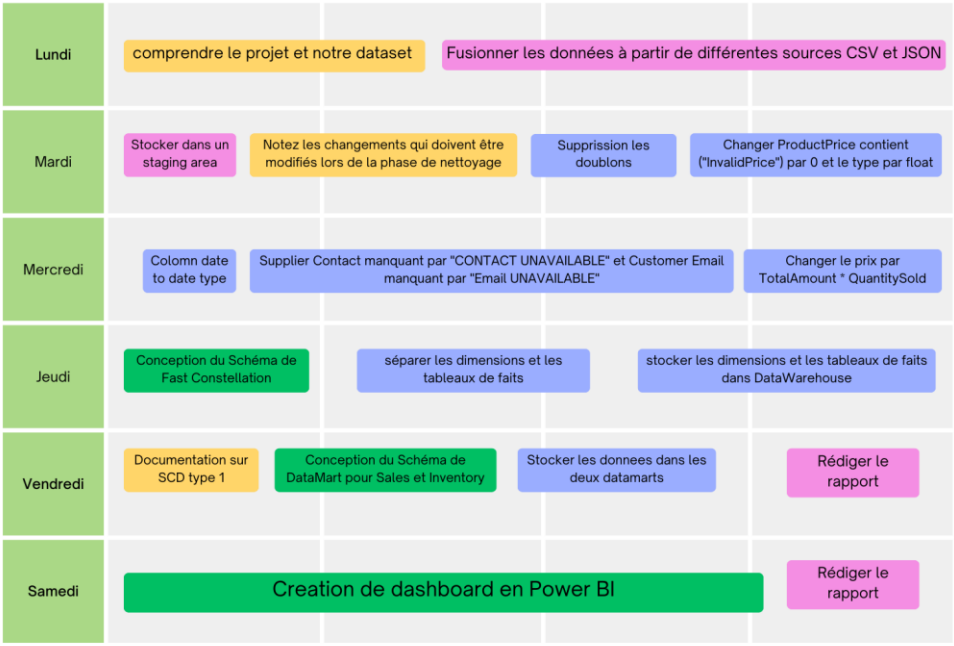
09 octobre 2023

Créé par : Yassine Harrati

Introduction:

Le projet avait pour objectif d'optimiser la gestion des données d'une plateforme e-commerce en utilisant des technologies telles que SQL Server, Talend pour l'ETL, et Power BI pour l'analyse. Le processus a été divisé en étapes clés, allant de l'extraction des données à leur visualisation en passant par leur transformation, leur stockage dans un entrepôt de données, et l'optimisation pour une utilisation plus efficace. Les aspects de sécurité et de conformité au RGPD ont également été pris en compte.

Planification de projet :



Compréhension du Projet et du Dataset:

Avant de commencer, nous avons pris le temps de comprendre en profondeur le projet et le dataset à traiter. Cela a permis d'établir une vision claire des objectifs à atteindre et des données avec lesquelles travailler.

Préparation des Données :

1- Division des Fichiers CSV et JSON :

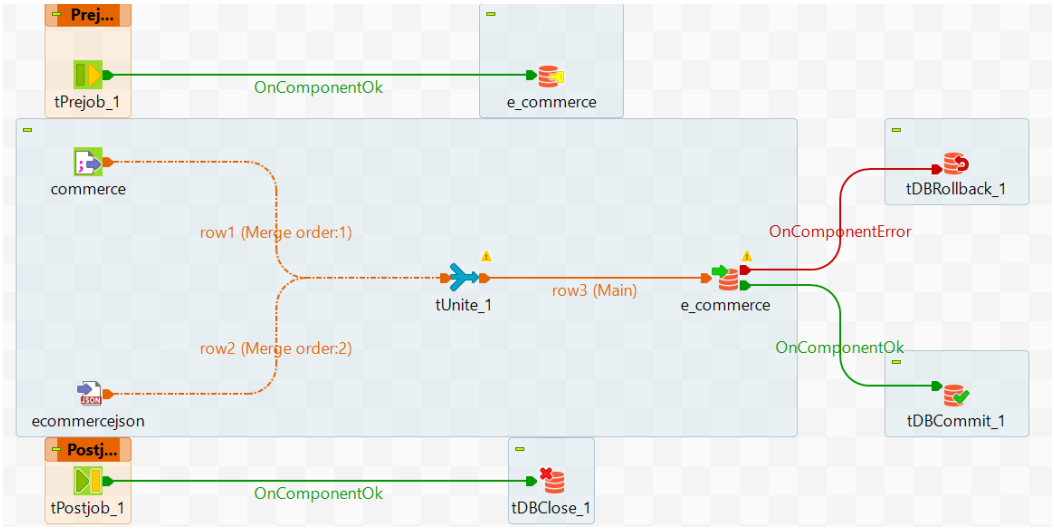
Nous avons commencé par diviser les fichiers CSV en deux parties, en format CSV et JSON, pour une utilisation future.

2- Fusion des Données :

Nous avons fusionné les données provenant de différentes sources CSV et JSON pour les rendre prêtes à être traitées.

3- Stockage dans une Staging Area:

Les données ont été stockées dans une zone de transit (staging area) pour permettre une manipulation plus fluide et optimisée.



4- Suppression des Doublons :

Les doublons ont été supprimés pour garantir des données propres et fiables.

Clé unique

Column	Attribut de clé	Sensible à la casse
Date	☑	☐
ProductName	☑	☐
ProductCategory	☑	☐
ProductSubCategory	☑	☐
ProductPrice	☑	☐
CustomerName	☑	☐
CustomerEmail	☑	☐
CustomerAddress	☑	☐
CustomerPhone	☑	☐
CustomerSegment	☑	☐
SupplierName	☑	☐
SupplierLocation	☑	☐
SupplierContact	☑	☐
ShipperName	☑	☐
ShippingMethod	☑	☐
QuantitySold	☑	☐

5- Nettoyage des Types de Colonnes :

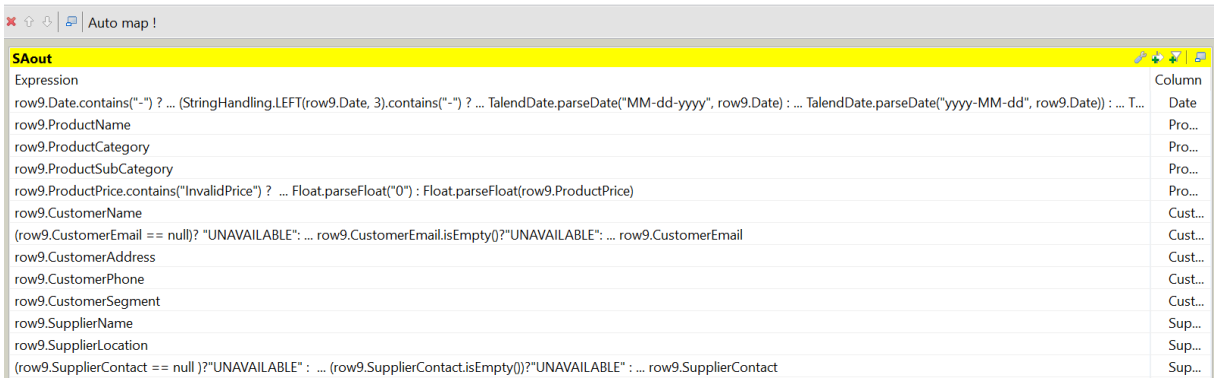
Les types de colonnes ont été ajustés pour s'assurer de la cohérence des données et de leur conformité avec le schéma.

6- Correction des Valeurs Invalides :

Certaines valeurs incorrectes ont été corrigées, par exemple, en remplaçant "InvalidPrice" par 0 dans la colonne ProductPrice. Et après j'ai remplaçais ProductPrice par TotalAmount / QuantitySold

7- Remplissage des Données Manquantes :

Les données manquantes ont été remplacées par "UNAVAILABLE" pour les contacts et les emails de fournisseurs.



8- Cryptage des Données Sensibles :

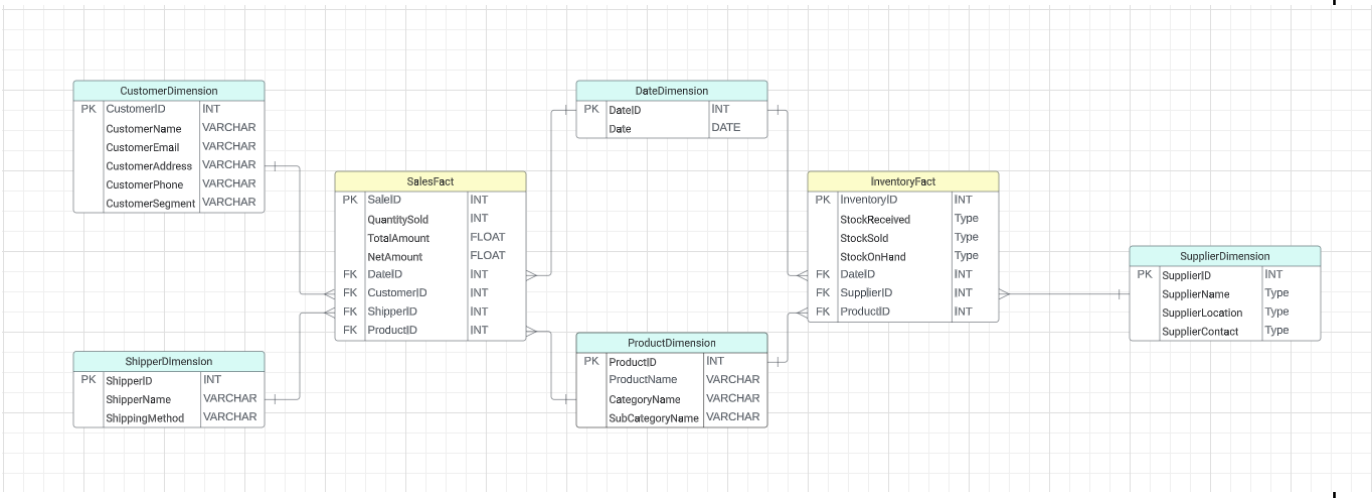
Les données sensibles, conformément au RGPD, ont été cryptées pour garantir la confidentialité des informations personnelles comme les emails et les contacts des clients.

```
StandardPBEStringEncryptor encryptor = new StandardPBEStringEncryptor();

encryptor.setAlgorithm("PBEWithMD5AndDES");

encryptor.setPassword("yaserrati");

//Code généré selon les schémas d'entrée et de sortie
output_row.Date = input_row.Date;
output_row.ProductName = input_row.ProductName;
output_row.ProductCategory = input_row.ProductCategory;
output_row.ProductSubCategory = input_row.ProductSubCategory;
output_row.ProductPrice = input_row.ProductPrice;
output_row.CustomerName = input_row.CustomerName;
output_row.CustomerEmail = encryptor.encrypt(input_row.CustomerEmail);
output_row.CustomerAddress = encryptor.encrypt(input_row.CustomerAddress);
output_row.CustomerPhone = encryptor.encrypt(input_row.CustomerPhone);
output_row.CustomerSegment = input_row.CustomerSegment;
output_row.SupplierName = input_row.SupplierName;
output_row.SupplierLocation = input_row.SupplierLocation;
output_row.SupplierContact = input_row.SupplierContact;
output_row.ShipperName = input_row.ShipperName;
output_row.ShippingMethod = input_row.ShippingMethod;
output_row.QuantitySold = input_row.QuantitySold;
output_row.TotalAmount = input_row.TotalAmount;
output_row.DiscountAmount = input_row.DiscountAmount;
output_row.NetAmount = input_row.NetAmount;
output_row.StockReceived = input_row.StockReceived;
output_row.StockSold = input_row.StockSold;
output_row.StockOnHand = input_row.StockOnHand;
```

Transformation et Chargement :

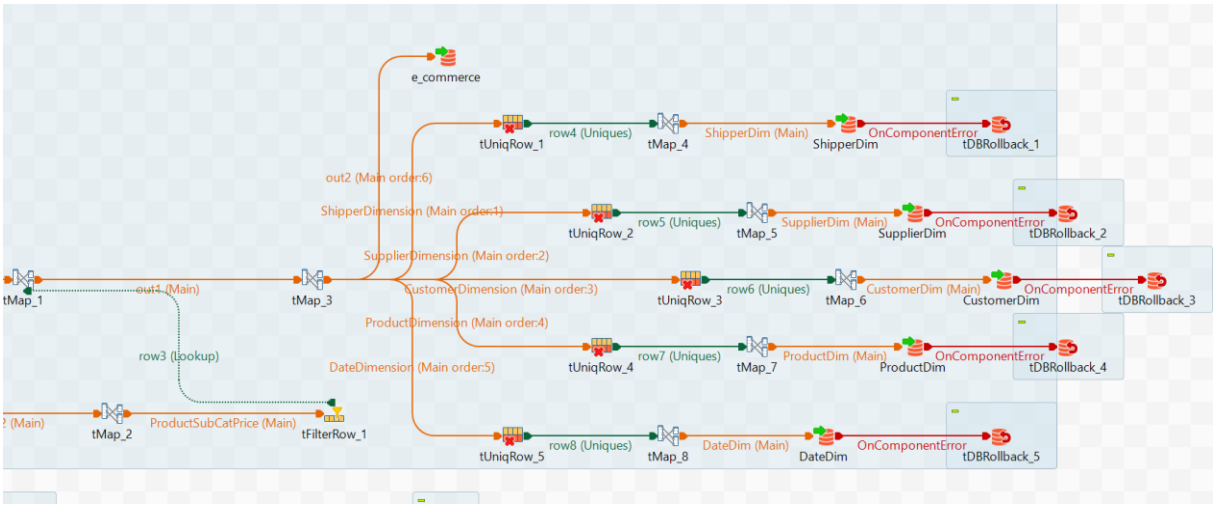
1- Conception du Schéma de Fast Constellation :

Les tables de dimensions et de faits ont été conçues selon un schéma de constellation, permettant une structuration efficace des données.



2- Stockage dans le Data Warehouse:

Les données ont été stockées dans le data warehouse en respectant le schéma et en garantissant la qualité des données.



3- Optimisation :

L'optimisation a été effectuée en créant des index et en partitionnant les données pour des requêtes plus rapides et efficaces.

Sécurité et Conformité :

1- Mise en Place des Mesures RGPD :

Des mesures ont été mises en place pour assurer la conformité au RGPD, incluant le cryptage des données et les droits des personnes concernées.

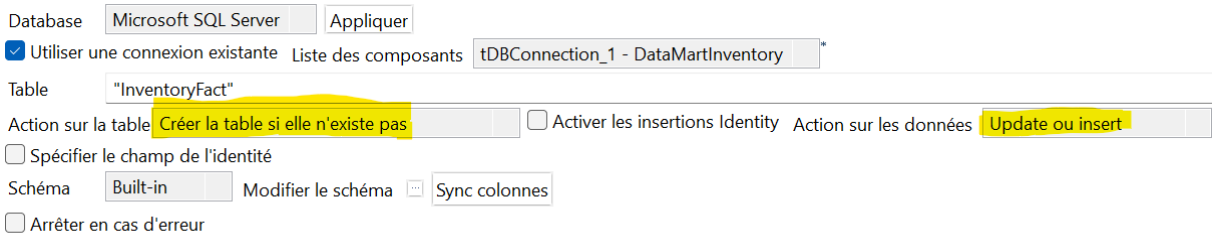
2- Gestion des Utilisateurs et des Rôles :

Les utilisateurs de la base de données ont été créés et attribués à des rôles pour garantir un accès sécurisé et approprié aux données.

Utilisation de SCD type 1

Alors dans cette étape et pour applique le SCD type 1, on doit faire ‘Update ou Insert’

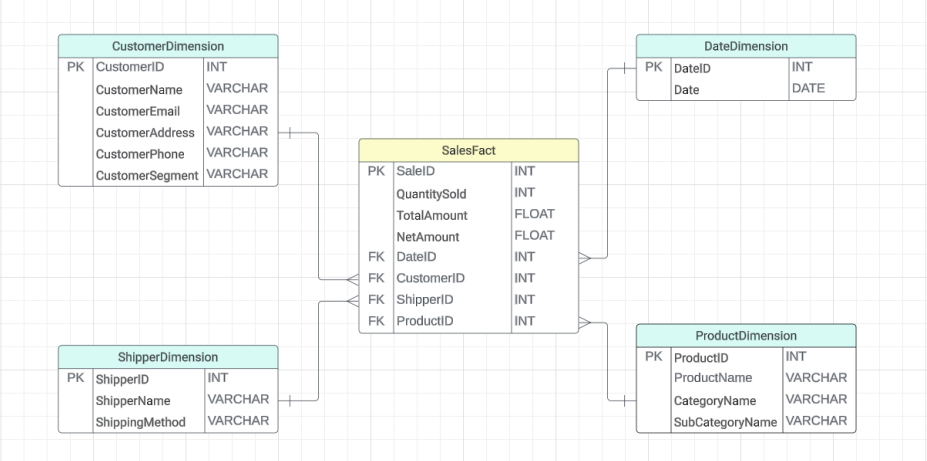
Dans Action sur les données, après ‘Créer la table si elle n’existe pas’ dans Action sur la table.
Par exemple le screenshoot suivant, et la même chose pour tous les table

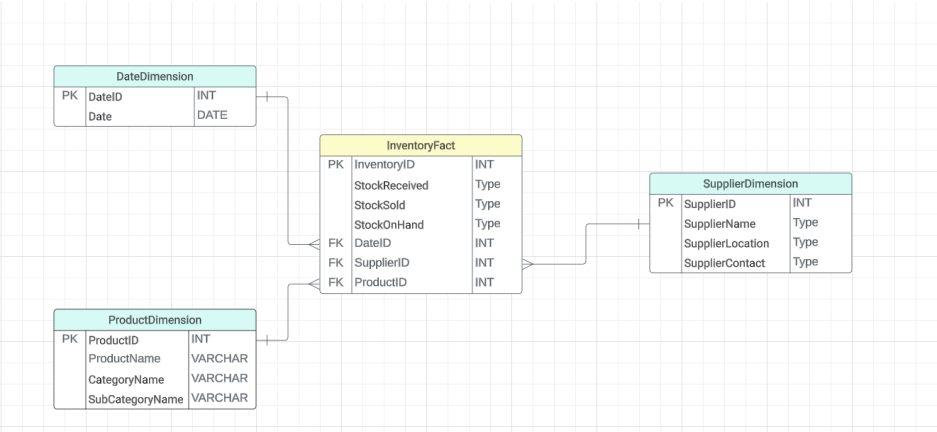


Conception des DataMarts:

1- Conception des DataMarts :

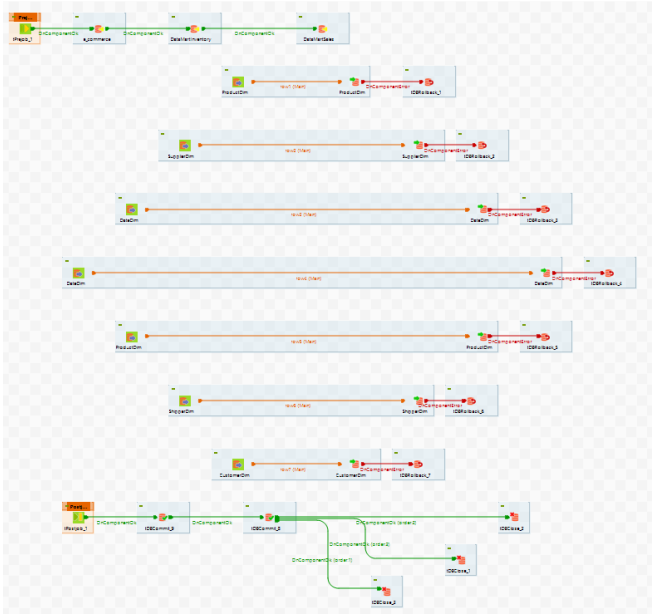
Des data marts ont été créés pour les ventes et l'inventaire, contenant des tables liées aux transactions spécifiques.

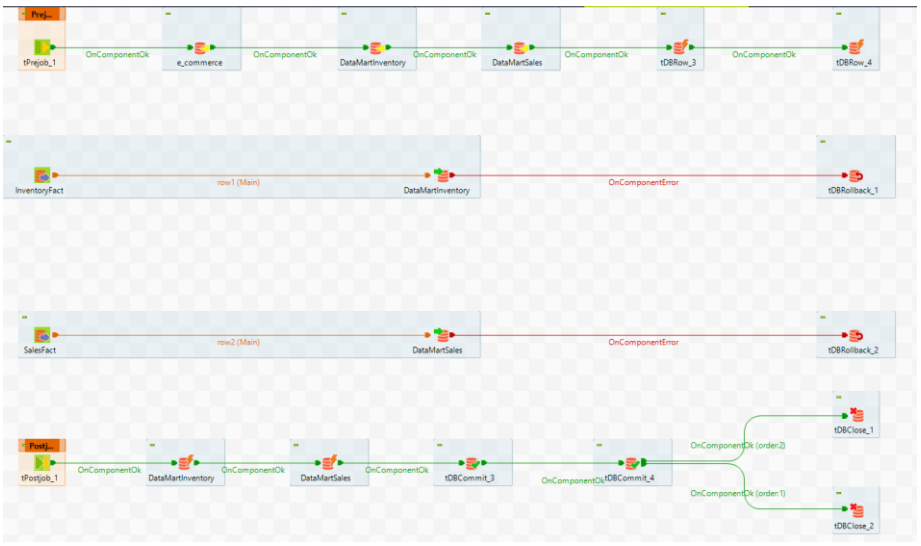




2- Stockage des Données dans les DataMarts :

Les données ont été stockées dans les data marts respectifs pour faciliter l'analyse et l'accès aux informations.





Analytique avec Power BI :

1- **Création de Dashboard Power BI** : fichier Power BI joint à ce rapport

Des tableaux de bord informatifs ont été conçus, fournissant des analyses approfondies sur les ventes et l'inventaire, permettant une meilleure compréhension des tendances du marché et des performances.

Conclusion :

Ce projet a été une opportunité passionnante de mettre en pratique nos connaissances en gestion de données, en ETL, en schéma de constellation, en optimisation et en sécurité des données. À travers une approche méthodique et structurée, nous avons réussi à atteindre les objectifs définis, en garantissant la qualité, la sécurité et la conformité des données. Les analyses générées via Power BI offrent des insights précieux pour une meilleure prise de décision. Ce projet a renforcé notre compréhension des bonnes pratiques et des défis inhérents à la gestion des données dans un contexte d'entreprise moderne.