Compte Rendu de Projet Entrepôt de Données

Date : 21 Avril 2024

Nom du Projet : Analyse de Marché pour l'Investissement de Mr Smith

Membres de l'équipe : Miri Rayen , Kouki Ayhem , Elee Ben Ameur

# 1. Introduction

Notre équipe a entamé la phase de modélisation de l'entrepôt de données en appliquant le modèle en étoile. Cette étape est cruciale pour la mise en place d'un système décisionnel efficace pour notre client, Mr Smith.

# 2. Avancement du Projet

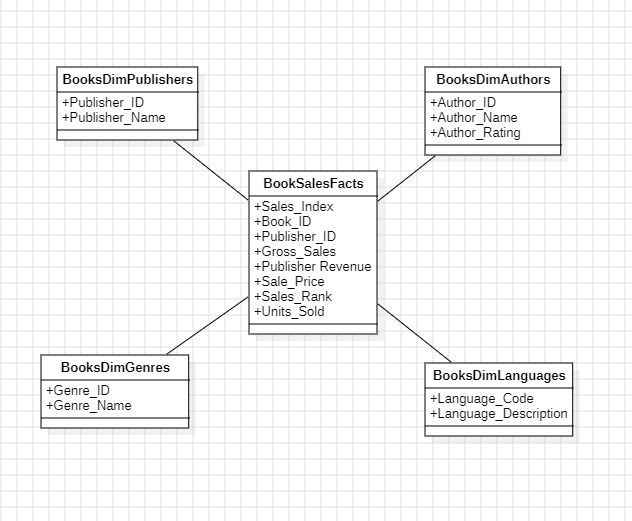
Feedback du Professeur sur le Data Cleaning :

Nous avons pris en compte le feedback du professeur concernant le data cleaning en Python et avons implémenté les ajustements recommandés dans SSIS en utilisant les fonctions d'agrégation et les colonnes dérivées.

Application du Modèle en Étoile :

Nous avons structuré l'entrepôt de données en suivant le modèle en étoile, qui sépare les données en dimensions et faits. Cette structure est optimisée pour les requêtes analytiques et prépare le terrain pour des analyses multidimensionnelles.

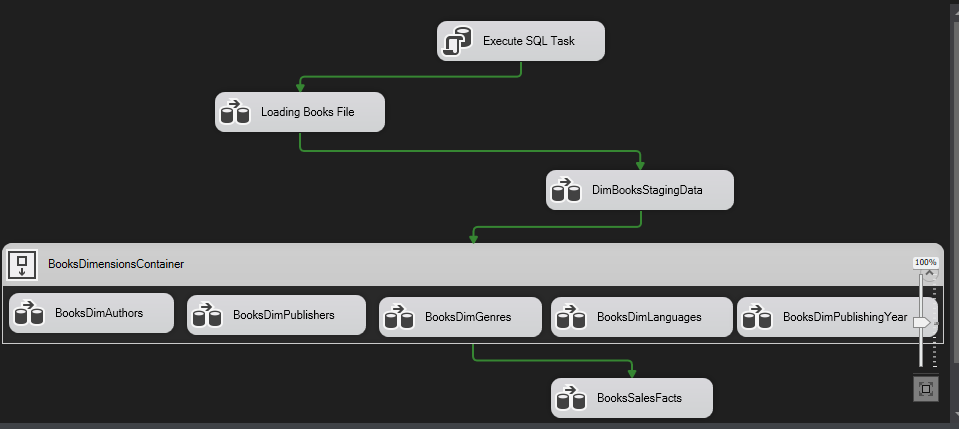
Example du conception du StarSchema des Livres :

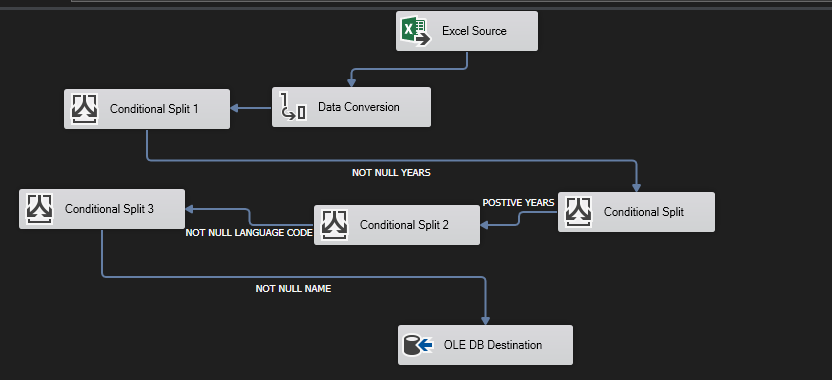


Conception des Tables de Dimensions et de Faits :

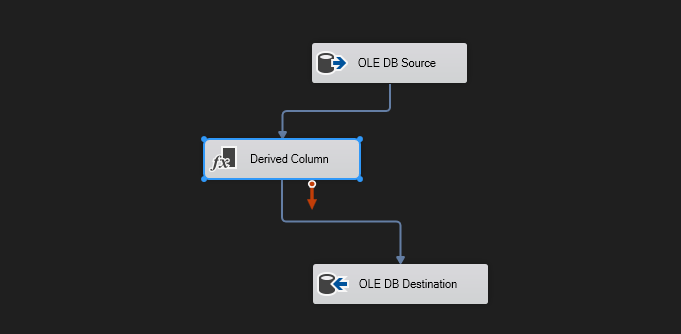
Les tables de dimensions ont été conçues pour contenir des attributs descriptifs, tandis que les tables de faits enregistrent les performances des ventes, les évaluations et d'autres mesures quantitatives. Cette distinction permet une analyse plus rapide et plus intuitive.

Application dans le Visual Studio :

Package des Livres

Loading BooksFiles DataFlow   


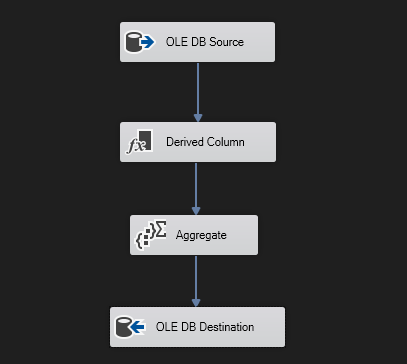
#### BooksDimAuthors

BooksDimGenres

L’expression du colonne derive utilisé est :

Genre\_Name == "genre fiction" ? "genre" : Genre\_Name

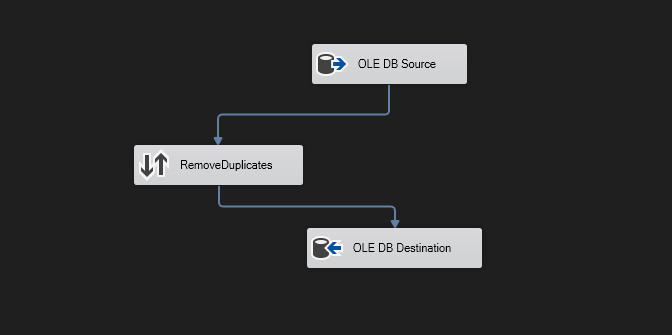
#### BooksDimLanguages :



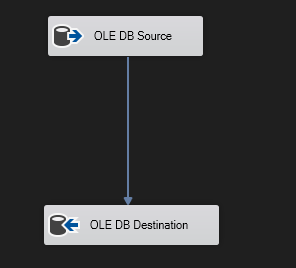
L’expression du colonne derive utilisé est :

SUBSTRING(Language\_Description,1,2) == "en" ? "en" : Language\_Description

#### BoksDimPublishingYear :



BooksSalesFacts :



Pour l’extraction des données j’ai utilisé une requete de jointure dans la source :

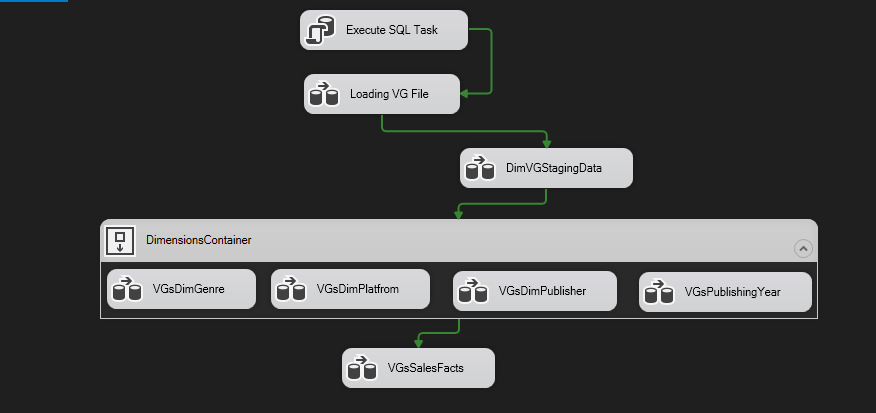
SELECT BDB.Book\_ID ,BDimP.Publisher\_ID,Publishing\_Year\_ID ,BDB.Gross\_Sales , BDB.Publisher\_Revenue , BDB.Sale\_Price , BDB.Sales\_Rank , BDB.Units\_Sold

FROM BooksDimension as BDim , BooksDB as BDB , BooksDimPublishers as BDimP , BooksDimPublishingYear as BDY

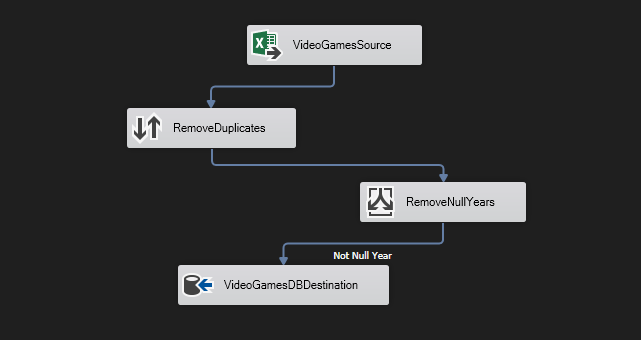
WHERE BDB.publisher = BDimP.Publisher\_Name

AND BDB.publishing\_year = BDY.Publishing\_Year

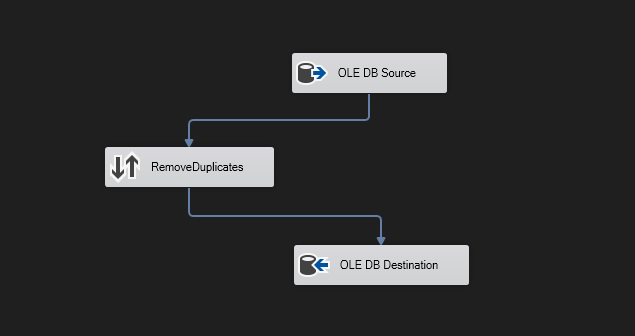
AND BDim.Book\_ID = BDB.Book\_ID order by Gross\_Sales desc

Package des Jeux de Videos

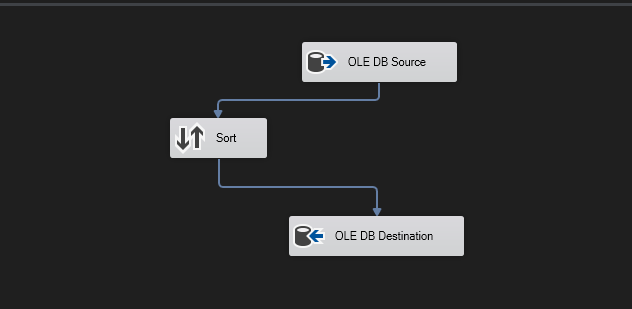
LoadingVGFile :



Toues les Dimensions du Video Games



VGsSalesFacts



Pour l’extraction des données j’ai utilisé une requete de jointure dans la source :

SELECT VG\_ID,VG\_RANK,Platform\_ID,Publishing\_Year\_ID,Genre\_ID,Pub.Publisher\_ID,V.na\_sales,EU\_Sales,JP\_Sales,Other\_Sales,Global\_Sales

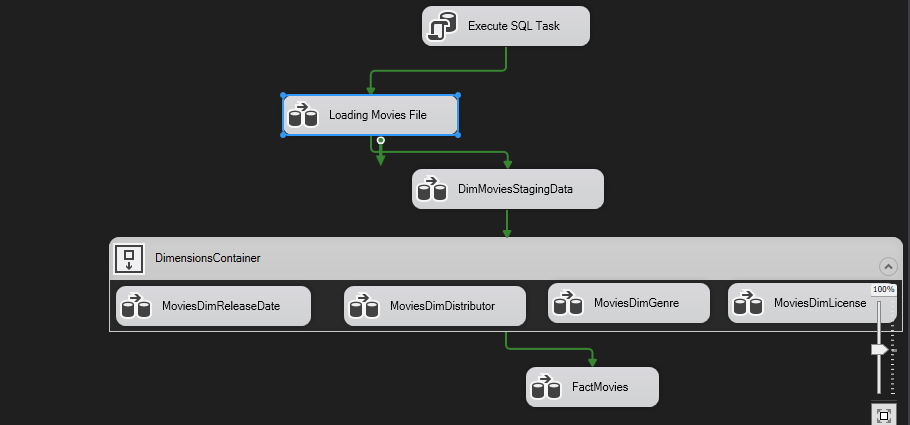
FROM VGDB as V , VGsDimGenre as G , VGsDimPlatform as P , VGsDimPublisher as Pub , VGsDimPublishing\_Year as VY

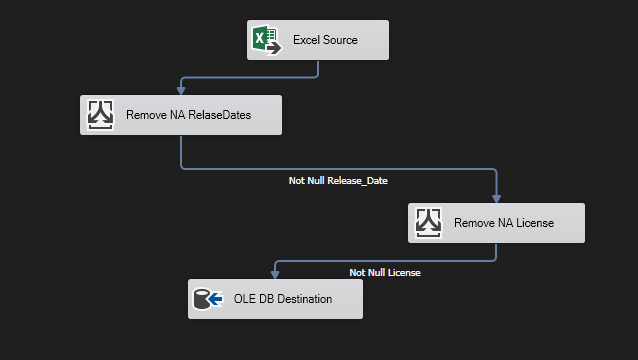
WHERE Pub.Publisher = V.publisher

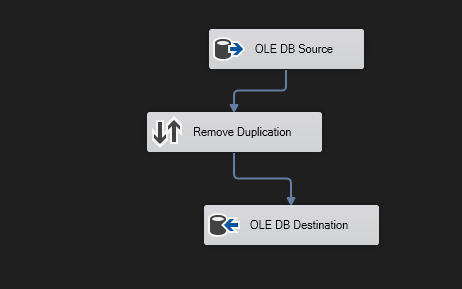
AND P.VG\_Platform = V.VG\_latform

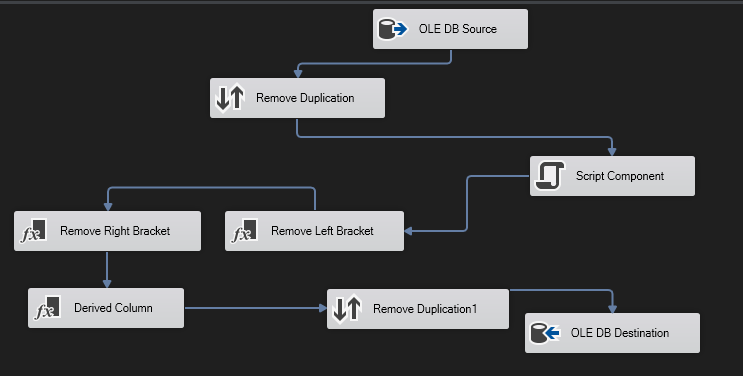
AND G.Genre = V.genre

AND V.VG\_Publishing\_Year = VY.Publishing\_Year

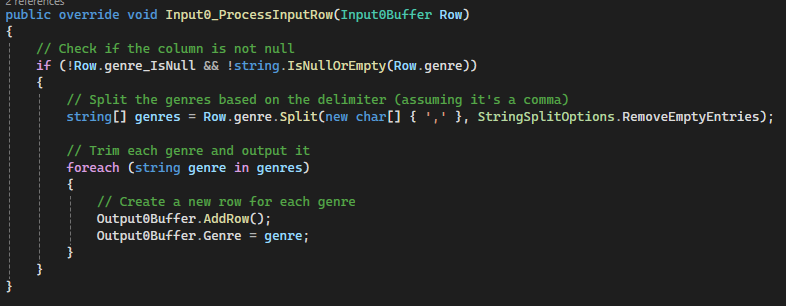
Package des Films

Loading Movies File

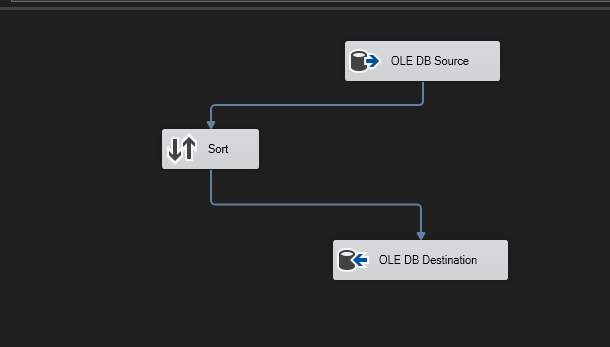
MoviesDimReleaseDate + MoviesDimDistributor

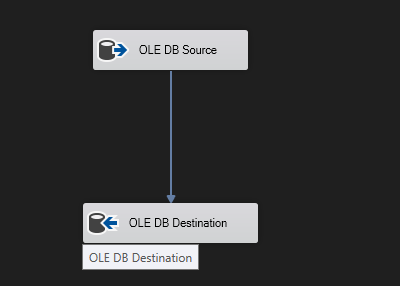
MoviesDimGenre

Pour le Script Component , le SSIS permet de utiliser le C# pour créer des columns ou faire une operation complexe c’est pour ce la j’ai l’utiliser , ce ci est un script en C# pour extraire chaque genre separament :

Voici un extrait du main function du code :

Pour les derived columns c’est just pour avoir une genre sans les « ‘[ ]’ » pour d’avoir une Clean String

MoviesDimLicense

FactMovies : 

Pour l’extraction des données j’ai utilisé une requete de jointure dans la source :

SELECT \* FROM (SELECT MG.Genre, M.Title, Movie\_ID,Distributor\_ID,Release\_Date\_ID,License\_ID,Genre\_ID,Domestic\_Sales,International\_Sales,World\_Sales

FROM MoviesDB as M , MoviesDimDistributor as MD , MoviesDimReleaseDate as MRD , MoviesDimGenre as MG , MoviesDimLicense as ML

WHERE M.Distributor = MD.Distributor

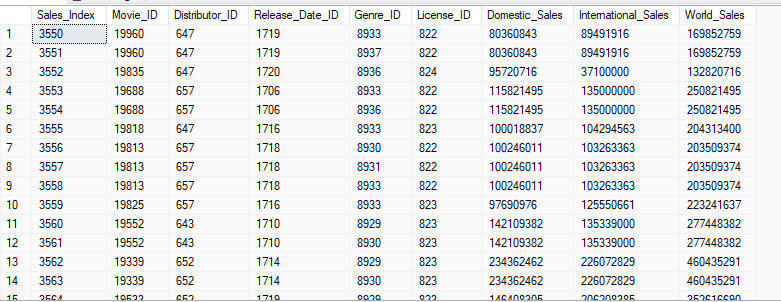
AND YEAR(M.Release\_Date) = MRD.Release\_Date

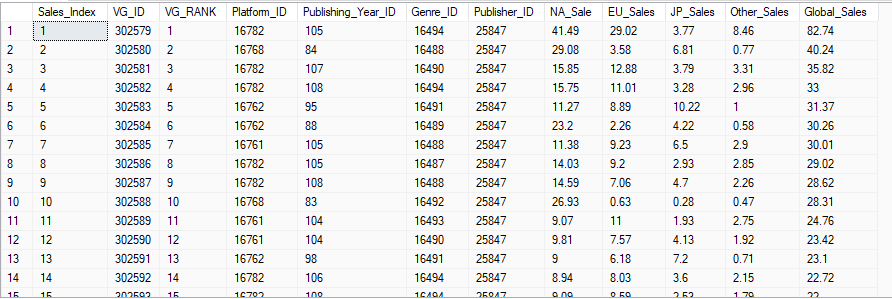
AND M.Genre LIKE '%'+MG.Genre+'%'

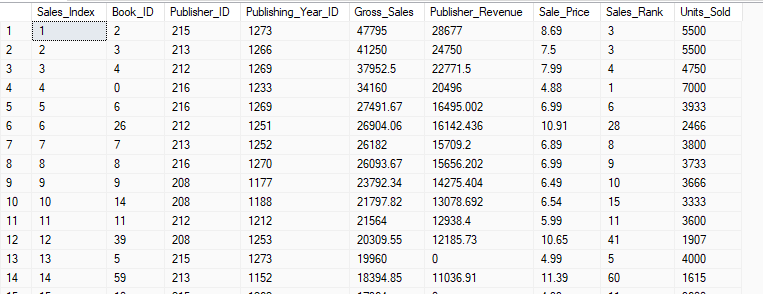
AND M.License = ML.License ) AS T1 ORDER BY T1.Title DESC

# 3. Résultats et Observations

Le modèle en étoile a été appliqué avec succès, et l'entrepôt de données est prêt pour l'analyse. La structure actuelle soutient notre objectif de fournir des insights précis et pertinents à Mr Smith.

Table Facts des Jeux Videos après l’execution

Table Facts des Films après l’execution

Table Facts des Livres après l’execution

# 4. Conclusion

L'application du modèle en étoile est une réussite pour notre projet. Nous sommes désormais bien positionnés pour entamer l'analyse des données et avancer vers la création d'un outil décisionnel pour Mr Smith.