

PROJET BI SQL SERVER



RÉALISÉ PAR :
EL MAZHARI AYA
BOUDDINE ABDELMOUGHIT

Supervisé par:
M.MOUTACHAOUIK Hicham

PLAN

❖ CONTEXTE

❖ ETUDE THEORIQUE

➤ Introduction-Définitions

➤ Table d'analyse

➤ Schéma en étoile dimensionnel du Datawarehouse

❖ ETUDE PRATIQUE

➤ Création de la base de données

➤ Création du schéma du Data Warehouse

➤ ETL

➤ Création du cube-OLAP

➤ Dashboard Power BI

Contexte

Pour beaucoup d'entreprises, les achats sont une activité cruciale. L'achat efficace d'un produit au bon prix, en vue de sa revente, est évidemment important pour des distributeurs tel qu'une chaîne de distribution alimentaire. Les achats ont aussi un effet majeur sur les résultats de toute organisation qui achète des produits en tant que matières premières servant à la fabrication.

Il s'agit de modéliser Le Datawarehouse des achats d'une chaîne de distribution alimentaire. Cette dernière achète des produits regroupés par familles. Une transaction d'achat est effectuée par un employé du service d'achat et correspond à un produit. Le Datawarehouse doit pouvoir fournir le montant de la transaction d'achat et le nombre d'unités achetées d'un produit par date, fournisseur, et employé. Les objets du Datawarehouse sont les suivants :

Produit, caractérisé par: code_produit,nom_produit,Cost,original sale price,discount,current price,taxesetc ;

SalesPerson caractérisée par: SalesPersonId,name;

customer, caractérisée par : code_client,nom_client ;

location, caractérisé par: LocationID ,Name ,County ,State,Code ,State ,Type ,Latitude ,Longitude ,Population ,Households ;

Time, caractérisée par: date ,day ,Month ,year ,quarter;

Achat caractérisée par : oder_id,product_id,SalesPerson_Id,costumer_id,purchase_date,montant de transaction,nombre d'unités.

ETUDE THEORIQUE

I. Introduction- Définitions

ETL

ETL signifie Extract, Transform et Load. Ce sont trois fonctions de base de données qui sont combinées en un seul outil pour extraire des données d'une base de données, les modifier et les placer dans une autre base de données.

Plus précisément, le processus d'extraction de données consiste à lire des données à partir d'une base de données. La transformation se produit lorsque les données sont converties - en utilisant des règles, des tables de recherche ou en les combinant avec d'autres - en données qui répondent aux exigences établies avec le client, puis en les chargeant dans une nouvelle base de données ou un nouvel entrepôt de données.

L'utilisation d'ETL garantit que les données sont pertinentes et utiles pour le client, qu'il soit précis, de haute qualité et facilement accessible afin que l'entrepôt de données soit utilisé de manière efficace et efficiente par les utilisateurs finaux.

SSIS

SSIS signifie SQL Server Integration Services. SSIS fait partie du logiciel de données Microsoft SQL Server, utilisé pour de nombreuses tâches de migration de données. Il s'agit essentiellement d'un outil ETL qui fait partie de la suite de Business Intelligence de Microsoft et est principalement utilisé pour réaliser l'intégration de données.

Cette plate-forme est conçue pour résoudre les problèmes liés à l'intégration de données et aux applications de flux de travail. Il dispose d'un outil de stockage pour ETL.

SSIS suit les étapes suivantes pour réaliser l'intégration :

- Cela commence par un entrepôt de données opérationnel, une base de données conçue pour intégrer des données provenant de plusieurs sources pour des opérations supplémentaires sur les données.
 - Le processus d'extraction, de transformation et de chargement (ETL) est effectué.
- L'entrepôt de données capture les données de diverses sources pour un accès et une utilisation utiles.
- Les données sont stockées dans l'entrepôt de données pour rassembler et gérer les données provenant de diverses sources afin de répondre aux questions commerciales. Par conséquent, elles aident à la prise de décision.

En plus d'ETL, SSIS permet d'autres processus tels que le nettoyage, l'agrégation et la fusion des données, entre autres. Il facilite le transfert de données d'une base de données à une autre et peut extraire des données d'une grande variété de sources telles que les bases de données SQL Server, les fichiers Excel, les bases de données Oracle et DB2, etc.

SSIS comprend également des outils graphiques et des assistants pour exécuter des fonctions de flux de travail telles qu'envoi d'e-mails, opérations FTP, sources de données et destinations.

Il est donc clair qu'il ne serait pas rigoureux de parler de la différence entre ETL et SSIS, puisque le nom ETL fait référence à un concept, alors que SSIS est un outil Microsoft développé pour fonctionner avec le concept ETL.

SSAS

Grâce à SSIS, vos données sont stockées dans une base de données isolée des outils qui les produisent.

Mais cela ne suffit pas, car pour un utilisateur lambda la donnée brute reste inexploitable. Il faut la transformer en agrégat, y associer des dimensions pour que l'on arrive à quelque chose d'utile. Un exemple d'agrégat et des dimensions associées : à partir des logs de tracking de votre web analytics, en sortir un nombre de visites par période, par univers du site, par typologie de clients. Et ne pas passer 3 heures à attendre que les données brutes moulinent. Et bien appliquer des règles de calcul et en restituer rapidement les données sont les fonctionnalités de SSAS.

Pour cela, SSAS génère des cubes, qui précalculent quand vous dormez vos indicateurs et leurs dimensions associées.

Terminologie importante du SSAS

Maintenant, dans ce tutoriel sur le modèle tabulaire SSAS, nous allons apprendre quelques terminologies importantes de SSAS :

- La source de données
- Vue de la source de données
- cube
- Tableau des dimensions
- Dimension
- Niveau
- Table de faits
- Mesure
- Schéma

✓ La source de données :

La source de données est une sorte de chaîne de connexion. Il établit une connexion entre la base de données d'analyse et le SGBDR.

✓ Vue de la source de données :

La vue de la source de données est un modèle logique de la base de données

✓ Cube :

Un cube est une unité de stockage de base. Il s'agit d'un ensemble de données qui ont été agrégées pour permettre aux requêtes de renvoyer rapidement des données.

✓ MOLAP :

Le MOLAP est constitué d'un cube de données qui contient des mesures et des dimensions. Il comprend tous les membres qui peuvent être dans une relation hiérarchique.

Il s'agit d'un ensemble de règles spécifiques qui vous aide à déterminer comment certaines cellules sont calculées dans un cube clairesmé et à mesurer les valeurs regroupées à l'intérieur de ces hiérarchies.

✓ Tableau des dimensions :

- Une table de dimension contient les dimensions d'un fait.
- Ils sont joints à la table de faits à l'aide d'une clé étrangère.
- Les tableaux de dimensions sont des tableaux dénormalisés.
- Les dimensions offrent des caractéristiques des faits à l'aide de leurs attributs.
- Ne propose pas de limite définie pour un nombre donné de dimensions
- La dimension contient une ou plusieurs relations hiérarchiques.

✓ Dimension :

Dimension offre le contexte entourant un événement de processus métier. En termes simples, ils donnent qui, quoi, où d'un fait. Dans le processus de gestion des ventes, pour le numéro de vente de fait, les dimensions seraient les noms des clients.

✓ Niveau :

Chaque type de résumé qui peut être récupéré à partir de la dimension unique est appelé étiquette.

✓ Tableau de faits :

Une table de faits est la table la plus importante d'un modèle dimensionnel. Une table de faits contient des mesures / faits et une clé étrangère de la table de dimension. Par exemple, les opérations de paie.

✓ Mesure :

Chaque table de faits contient une ou plusieurs tables qui doivent être analysées. Par exemple, un livre vend une table d'information. Cela peut être un profit ou une perte pour le nombre de livres vendus.

✓ Schéma :

Le schéma de base de données d'un système de base de données et sa structure décrits dans un langage formel. Il prend en charge le système de gestion de base de données. Le terme «schéma» signifie l'organisation des données en tant que modèle de la manière dont la base de données est construite.

OLE DB

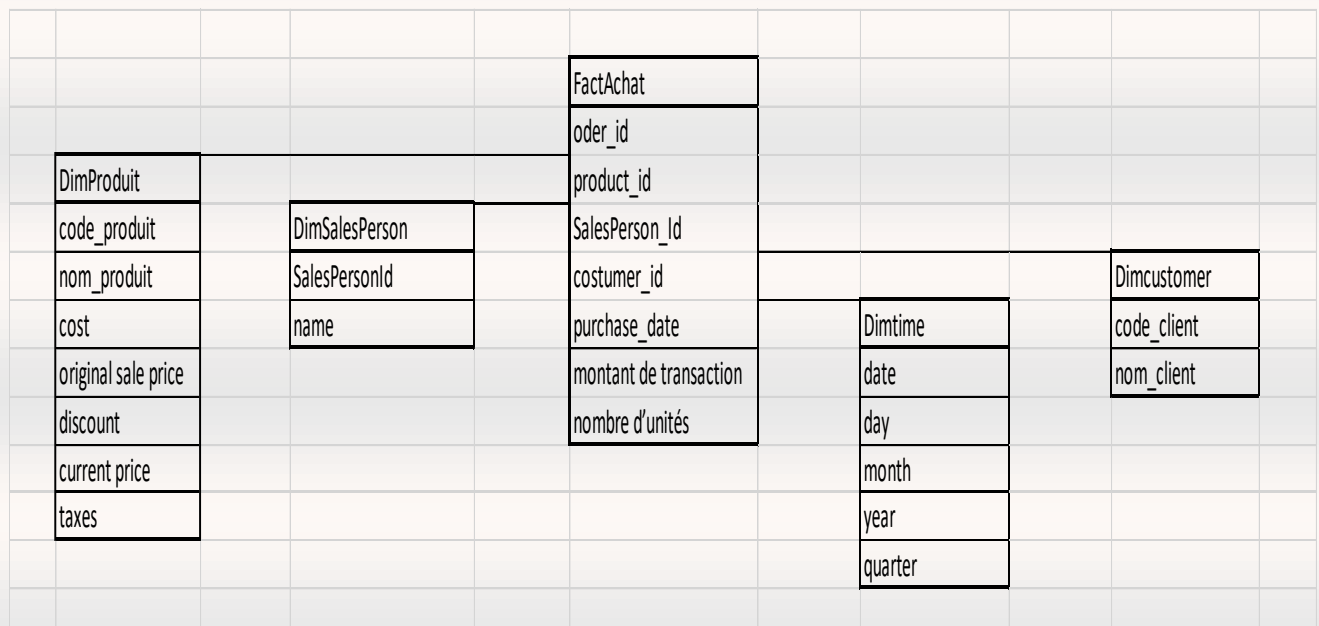
La destination SSIS OLE DB est utilisée pour charger des données dans diverses tables de base de données, vues ou commandes SQL. L'éditeur de destination OLE DB nous offre le choix de sélectionner la ou les tables existantes, les vues, ou vous pouvez créer une nouvelle table.

II. Table d'analyse

Suivant le contexte nous proposons la table d'analyse ci-dessous

	Produit	SalesPerson	customer	location	time
	code_produit	SalesPersonId	code_client	Location ID	date
	nom_produit	name	nom_client	Name	day
	cost			County	month
	original sale price			State Code	year
	discount			State	quarter
	current price			Type	
	taxes			Latitude	
				Longitude	
				Population	
				Households	
Analyse: achat					
Unity price;					
Quantity					

III. Schéma en étoile dimensionnel du Datawarehouse



ETUDE PRATIQUE

I. Création de la base de données :

La 1^{ère} étape de ce processus est la création de la base de données « TP_BI » ainsi que les tables correspondantes qui sont déjà mentionnées dans la partie théorique et divisées en 2 types. Il existe les tables de dimensions : « DimCustomer », « DimTime », « DimSalesPeople », « DimLocation », « DimProduct » ; et la table de faits : « FactSales ».(Figure1)

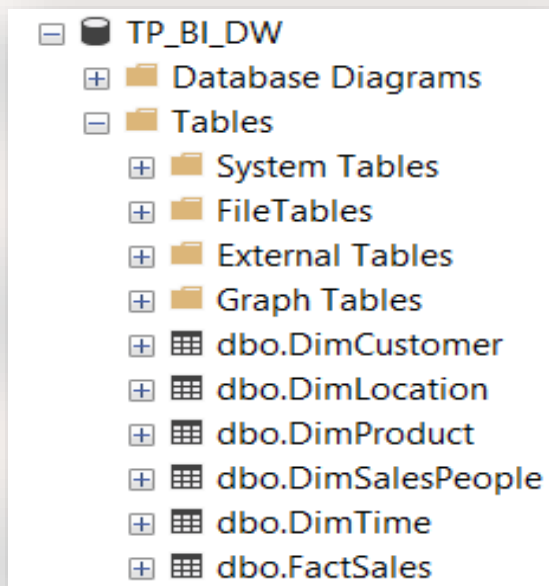


Figure1 : Résultat dans le SSMS

II. Création du schéma du DataWarehouse :

C'est l'équivalent du schéma en étoile déjà présenté dans la partie théorique, en liant la table de fait avec les autres tables de dimension grâce aux clés primaires et secondaires, ce qui nous donne le schéma sur SQL Server (figure 2)

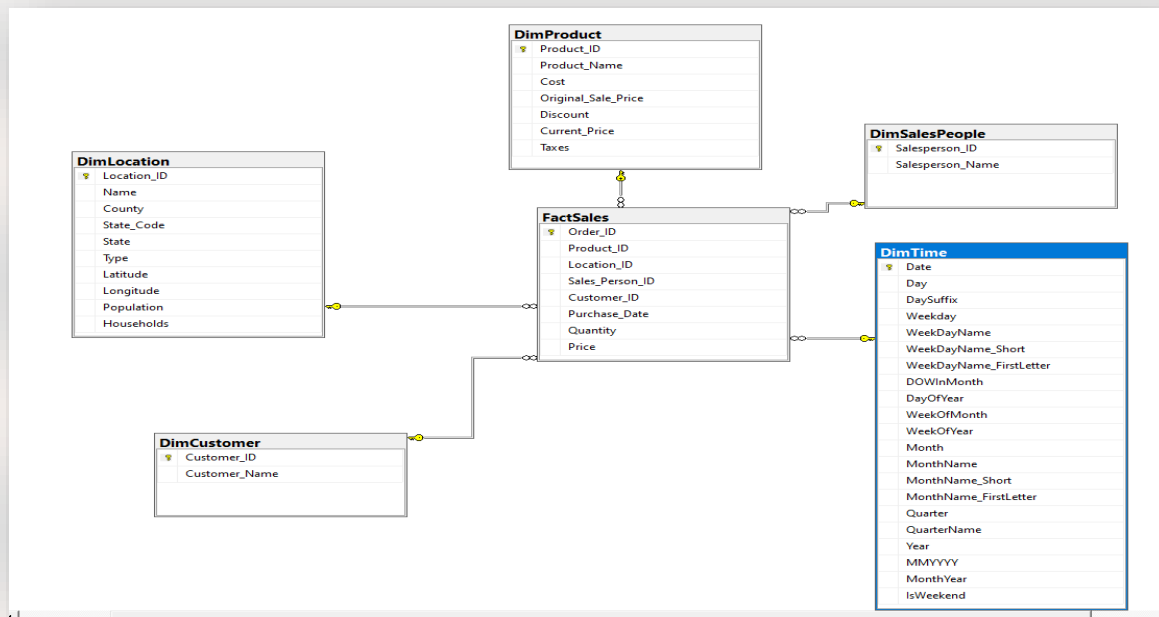


Figure 2 : schéma sur SQL Server

N.B : Les liaisons sont créées automatiquement.

III. ETL :

Pour cette partie, nous avons installé les deux services nécessaires de SQL Server Data Tools : SSIS « SQL Server Integration Services » et SSAS « SQL Server Analysis Services ».

Pour l'extraction et le chargement des données, nous avons créé un 'Data Flow Task' en premier, après il y avait la création en dedans des différents « Flat File Source » qui seront la source des données, c-à-d ils seront liés aux fichiers de données, ainsi « OLE DB Destination » correspondants à ces sources, et qui seront les responsables de remplir la base de données sur SQL Server par les données.

La scène du Data Flow Task (Figure3) apparaîtrait en premier comme ça :

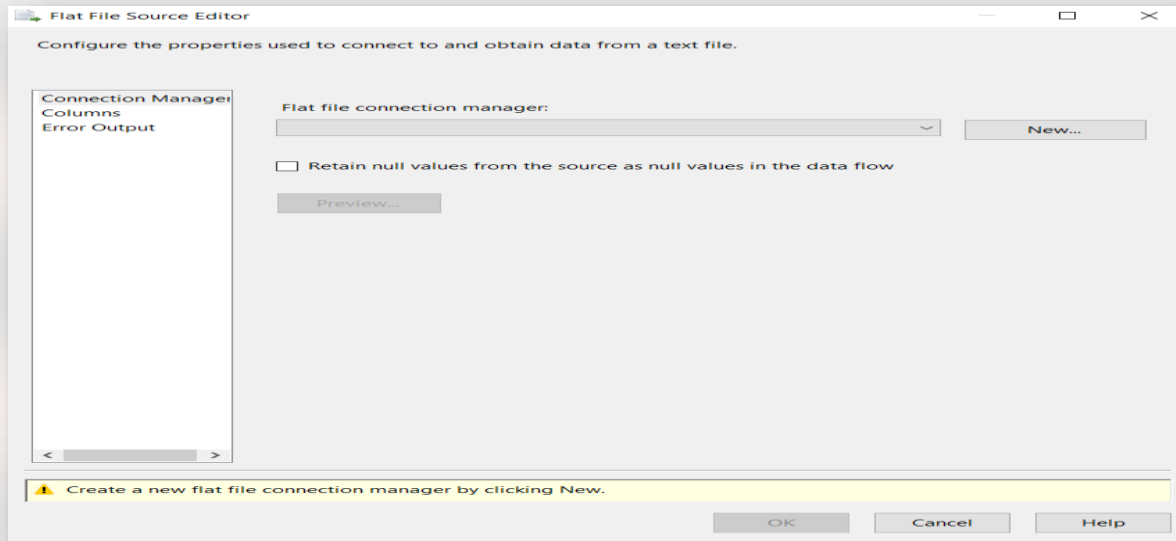


Figure 4 : Extraction file Customer

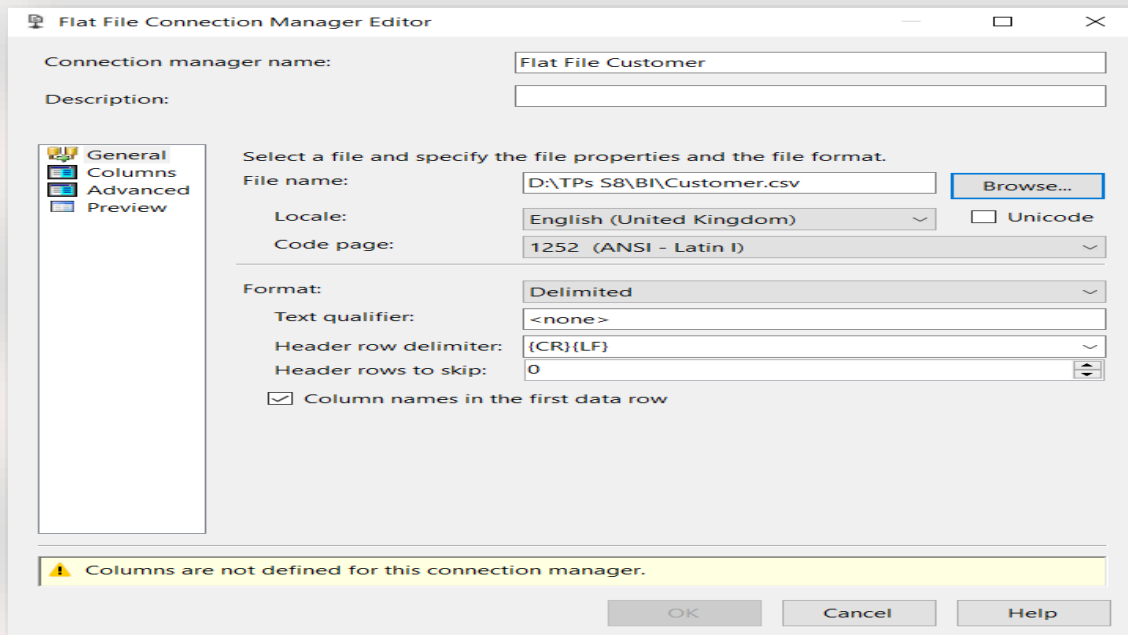


Figure 5: connection manager

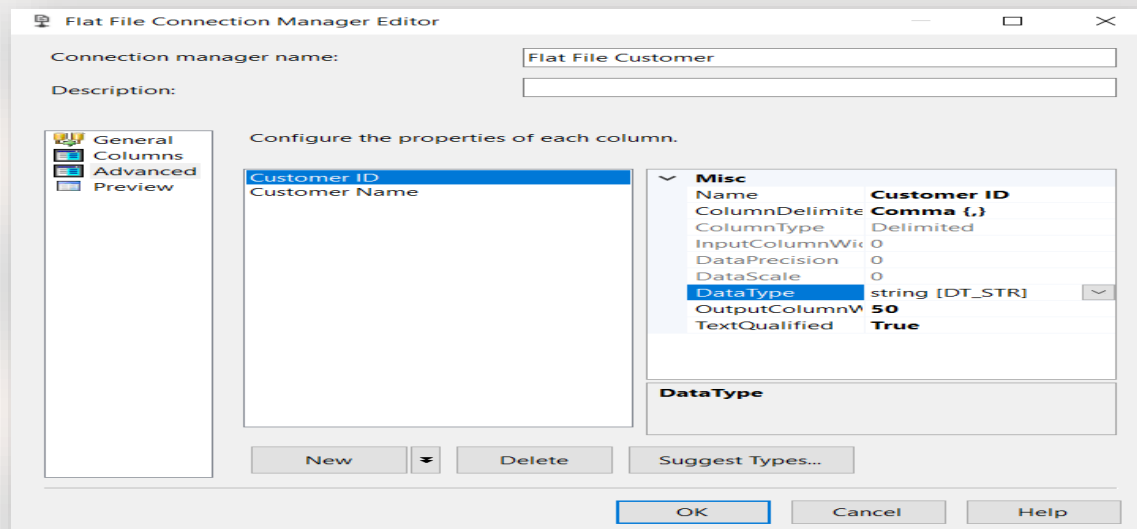


Figure 6: connection manager (advanced)

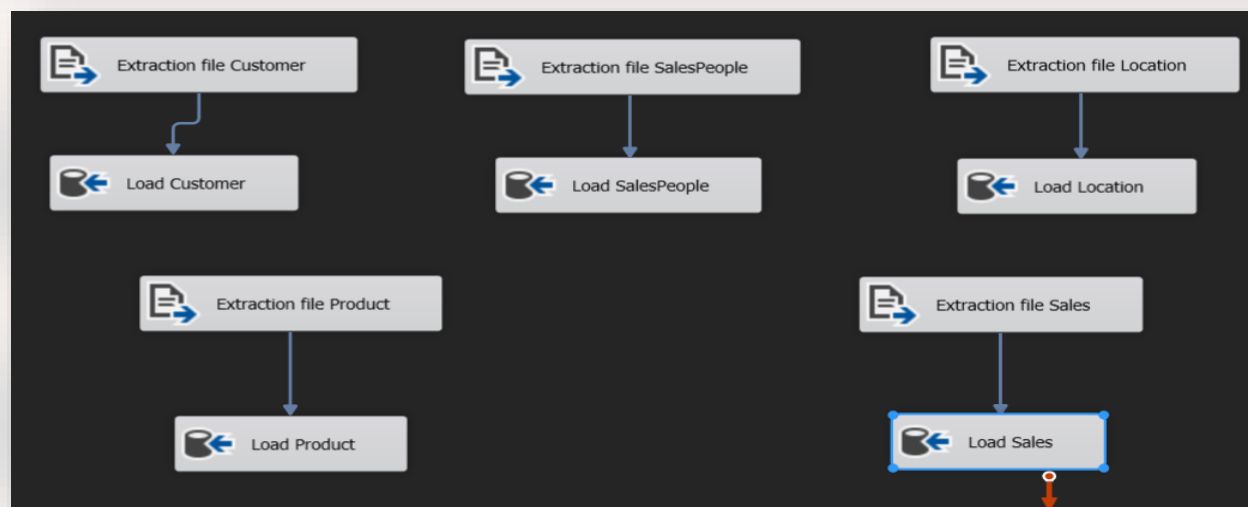


figure7 : liaison Flat File Source avec OLE DB Destination

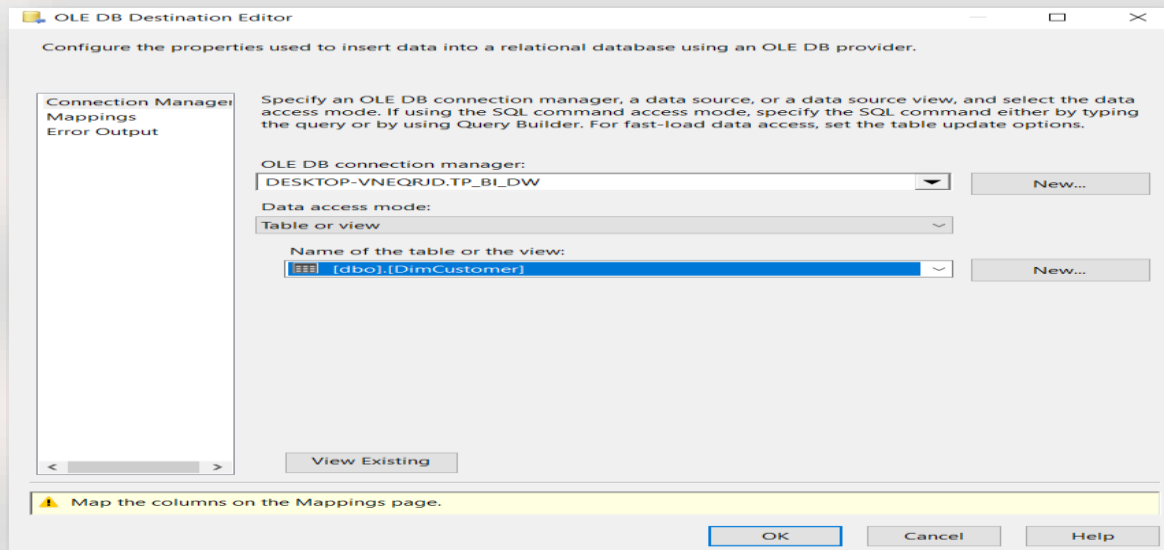


Figure 8 : configuration de la destination.

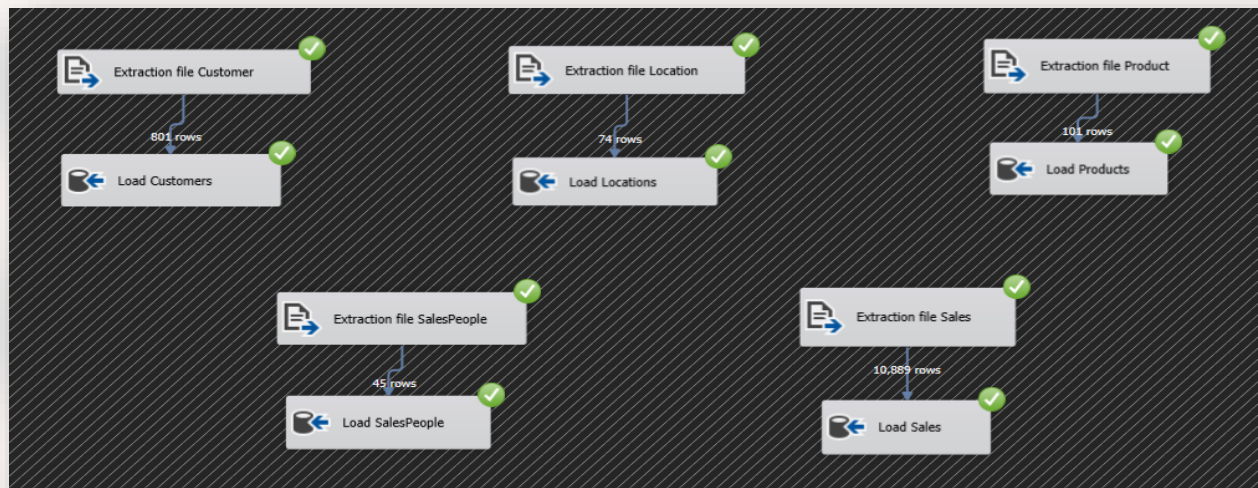


Figure 9 : chargement des données réussi

IV. Création du cube-OLAP :

Pour cette partie, nous aurons besoin d'utiliser le deuxième service déjà téléchargé SSAS « SQL Server Analysis Services » dans lequel nous allons créer un nouveau projet pour manipuler le cube OLAP (figure 10).

Voici les composants de la solution du projet créé :

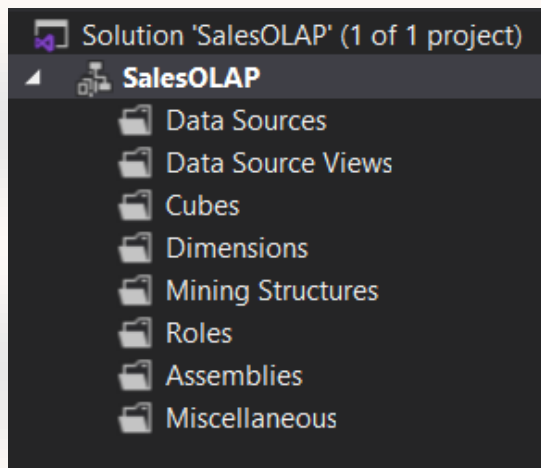


Figure 10

- a) Création de la source de données en utilisant connexion existante à partir de la base (Figure 11) « TP_BI_DW » et le compte du service « SSAS ». (Figure 12)

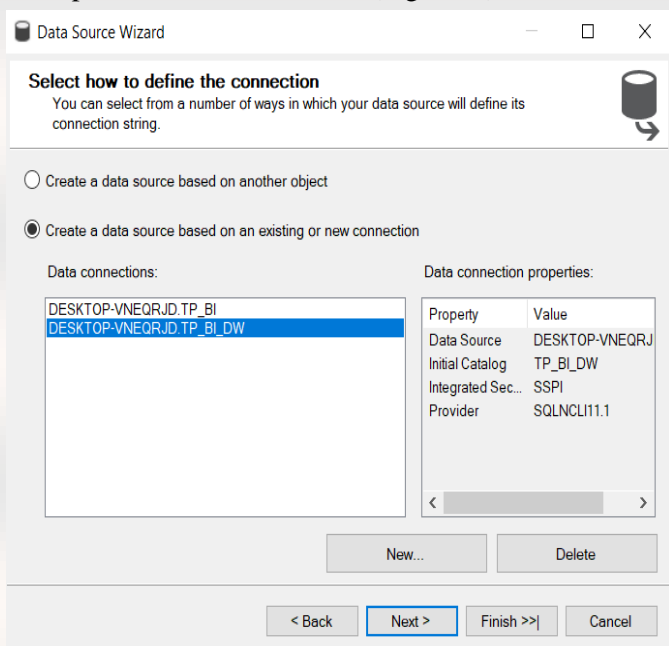


Figure 11 : connexion existante à partir de la bd

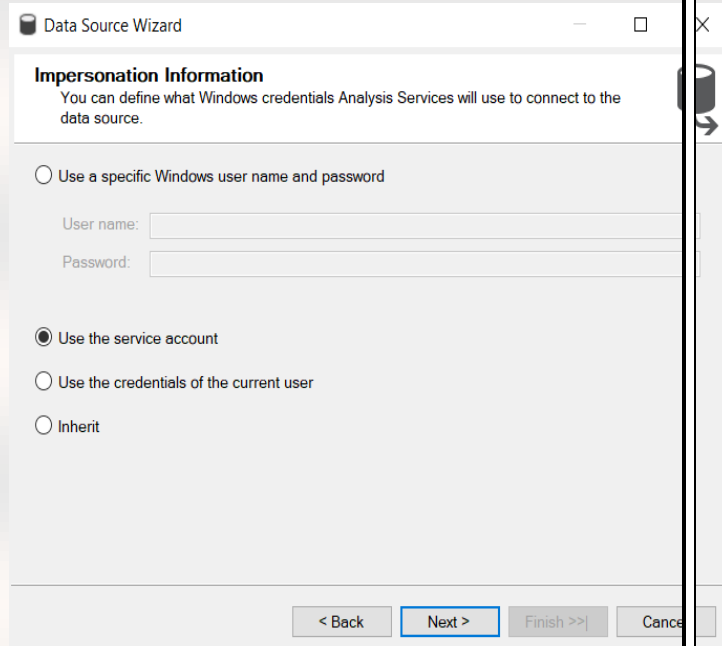


Figure 12 : choix du compte du service

b) Création de la vue de source de données (figure13) où nous sélectionnons les tables qui seront traitées(figure14).

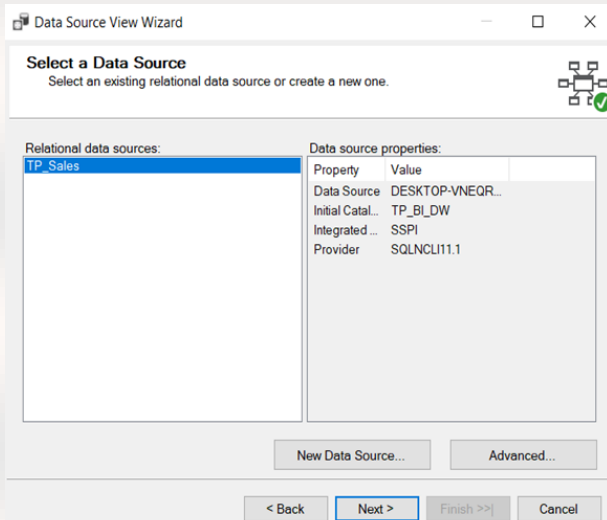


Figure13 : vue de la source de données

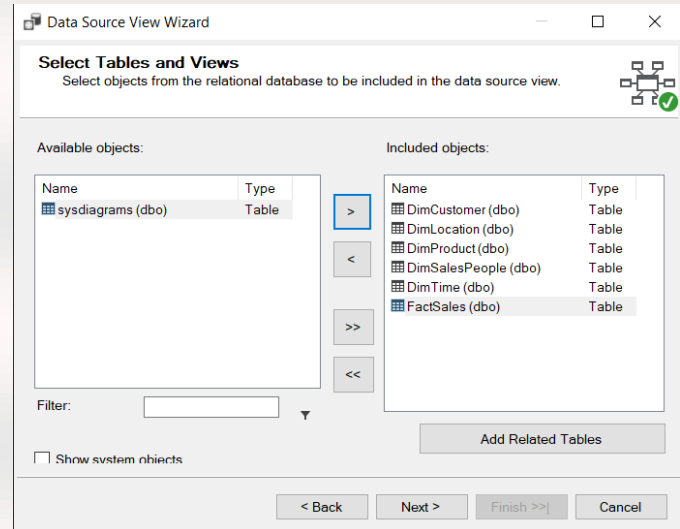


Figure14 : sélection des tables

c) Cette étape est très importante car elle concerne la création du cube dont nous précisons les tables de dimensions et table de fait. Les différentes étapes de cette création sont résumées dans les images suivantes :

- ✓ Sélection de la table de faits et les mesures (figure 15-16) : (Les autres mesures seront calculées après dans Power BI)
- ✓ Sélection des tables de dimensions(figure17) :

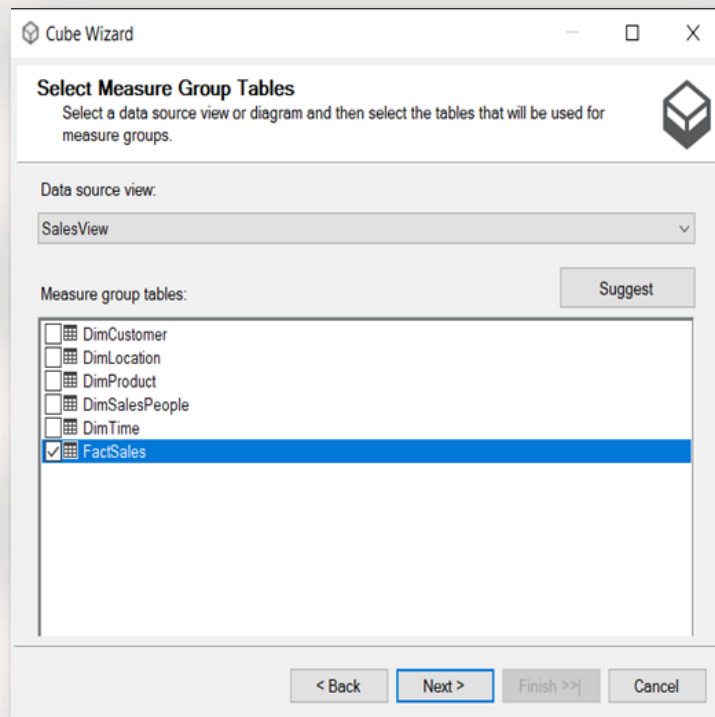


Figure15

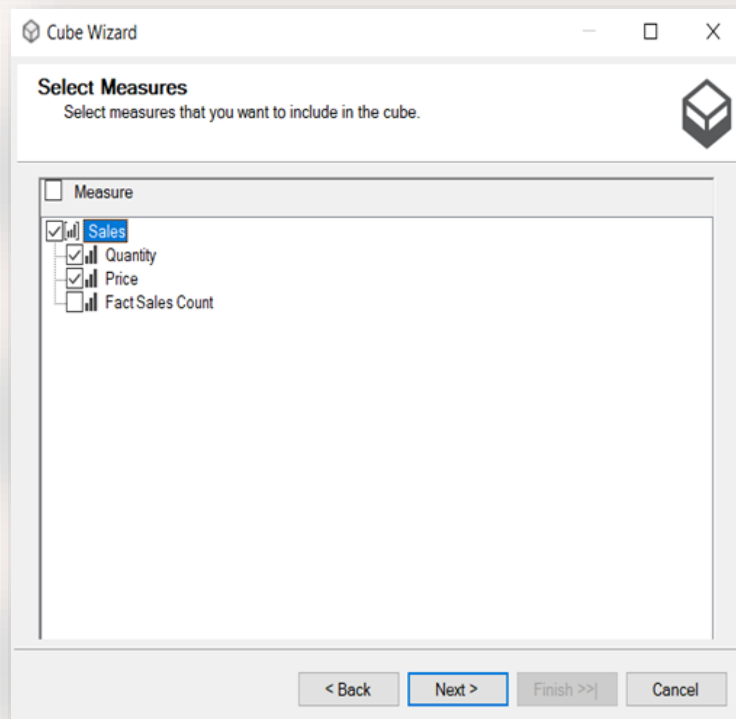


Figure16

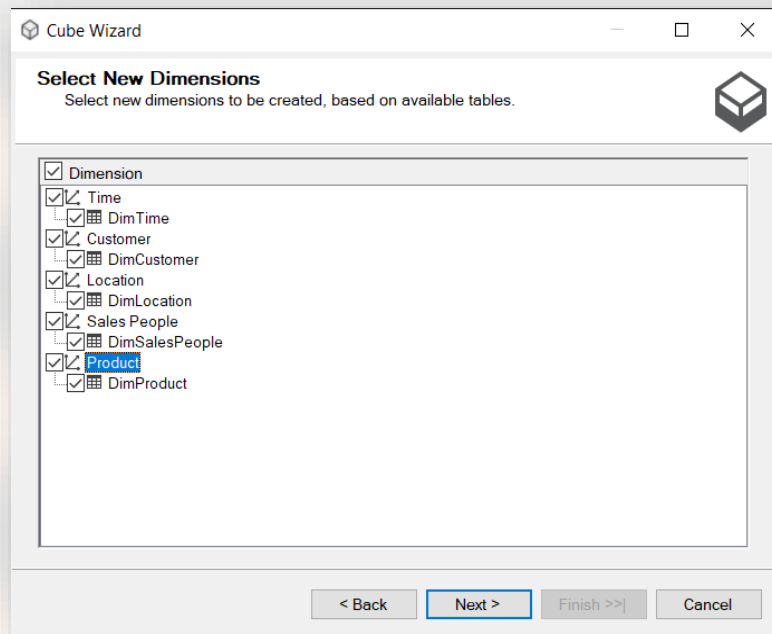


Figure17

- d) Déploiement et traitement du cube (figure19) nous permet de créer une structure de données et l'alimenter avec des agrégats, à l'aide d'un clique-droit sur le cube et puis choisir l'option « Process » (figure18).

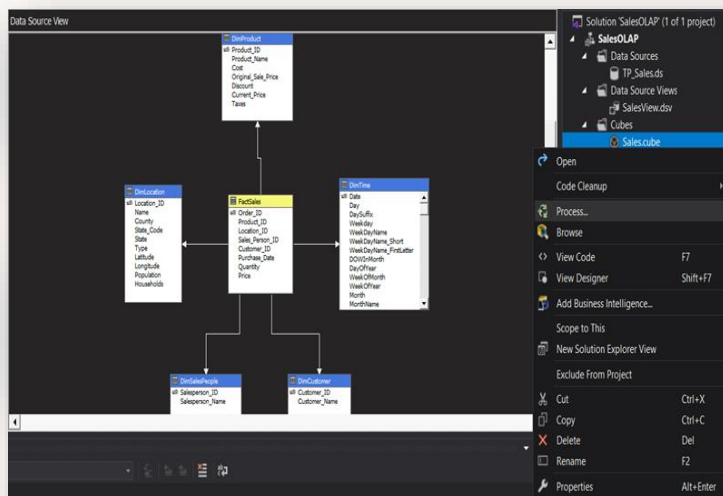


Figure18

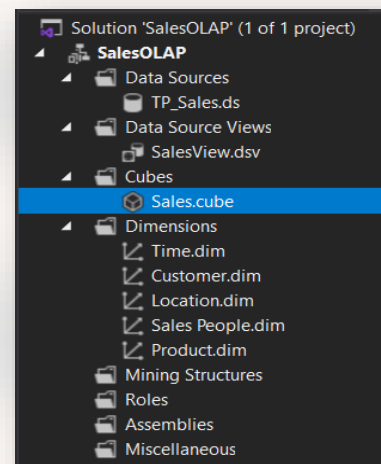
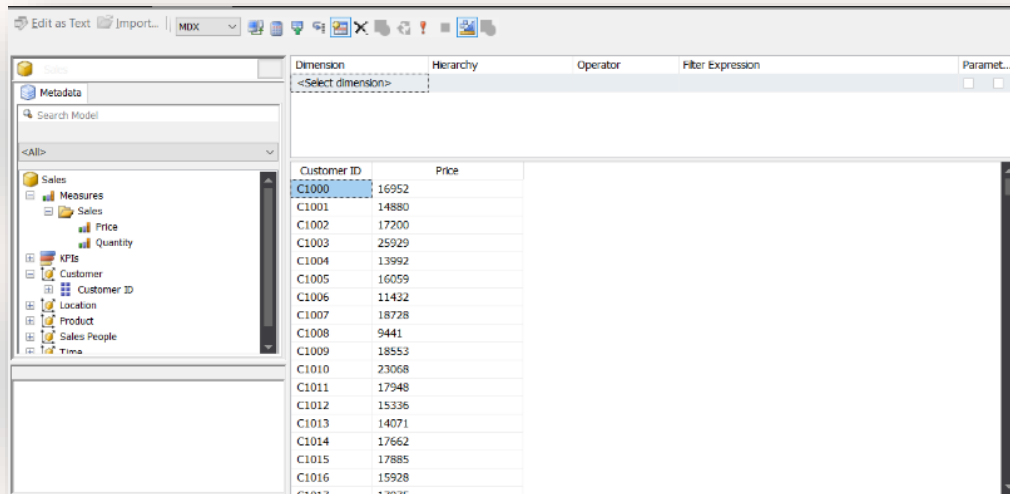


Figure 19

e) L'exploration du cube sous l'onglet Browser :

Nous prenons l'exemple de la mesure « Prix » par la dimension « Customer ID », ce qui nous résulte le tableau suivant (figure20) :



Customer ID	Price
C1000	16952
C1001	14880
C1002	17200
C1003	25929
C1004	13992
C1005	16059
C1006	11432
C1007	18728
C1008	9441
C1009	18553
C1010	23068
C1011	17948
C1012	15336
C1013	14071
C1014	17662
C1015	17885
C1016	15928

Figure 20

V. Peaufiner le cube :

Cette partie est spéciale, car elle permet de sélectionner les attributs à importer avec les tableaux de dimension et les nouvelles mesures à calculer.

Nous avons créé dans la partie précédente un cube basé sur l'entrepôt de données « SalesOLAP ». La création et la génération du cube a été rapide. Mais il est apparu clairement que quelques finitions étaient nécessaires pour le rendre accessible aux utilisateurs.

Dimensions : attributs

Une des finitions les plus évidentes est la nécessité de retravailler les dimensions. Dans notre travail de finition, pour chaque dimension, nous allons devoir identifier les attributs à afficher et quand c'est possible chercher à les organiser en hiérarchie.

La succession des étapes est bien comme suivant : Sous l'onglet « Dimensions », nous cliquons sur la dimension souhaitée, puis « Éditer la dimension ». Nous glissons les attributs depuis l'onglet « Data Source View » vers l'onglet « Attributs », puis obligatoirement modifier la propriété 'NameColumn' des attributs. A la fin, afin de s'assurer que les attributs choisis sont bien intégrés, nous allons sous l'onglet « Browser » où nous pouvons étendre un attribut pour visualiser ses valeurs.

Exemple : Dimension 'Customer' (figures 21-24)

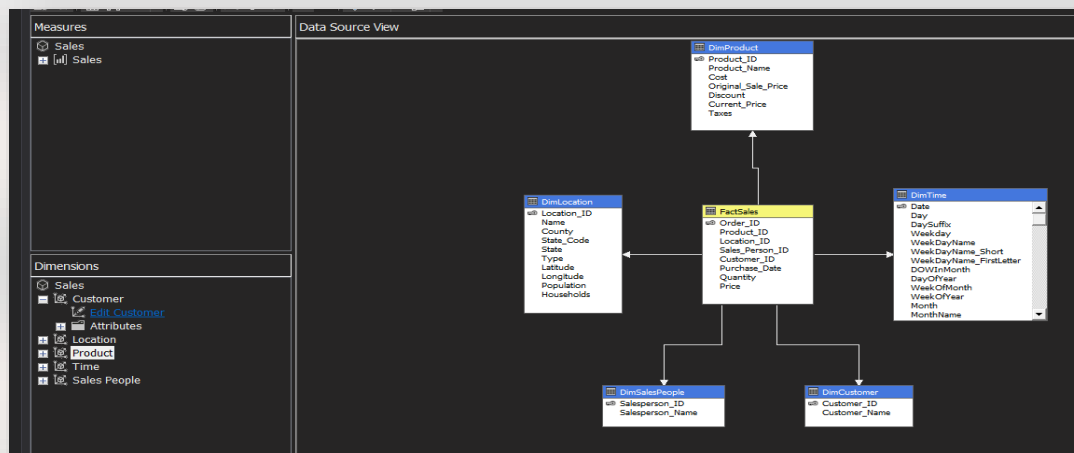


Figure21 :

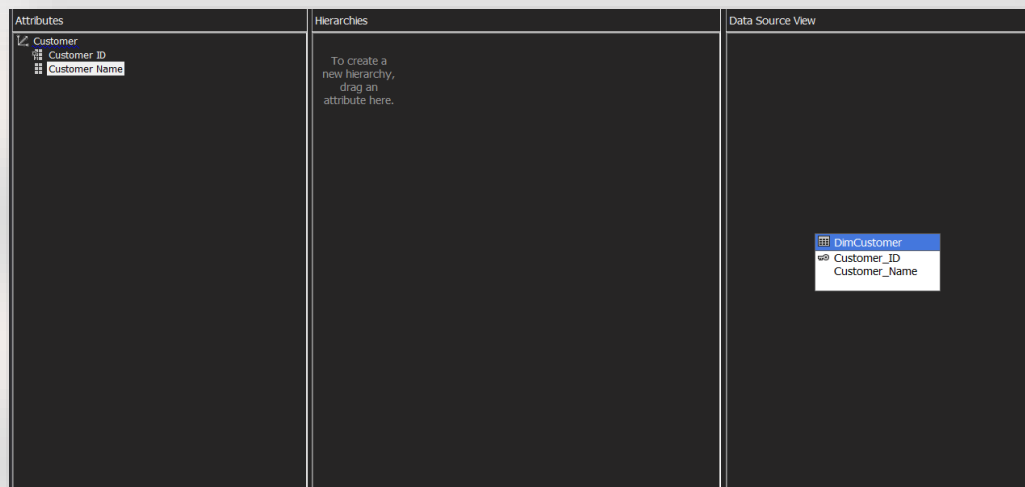


figure22

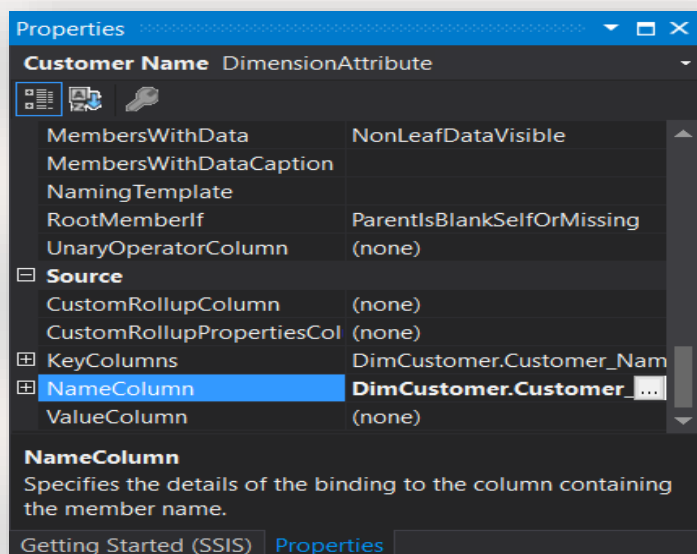
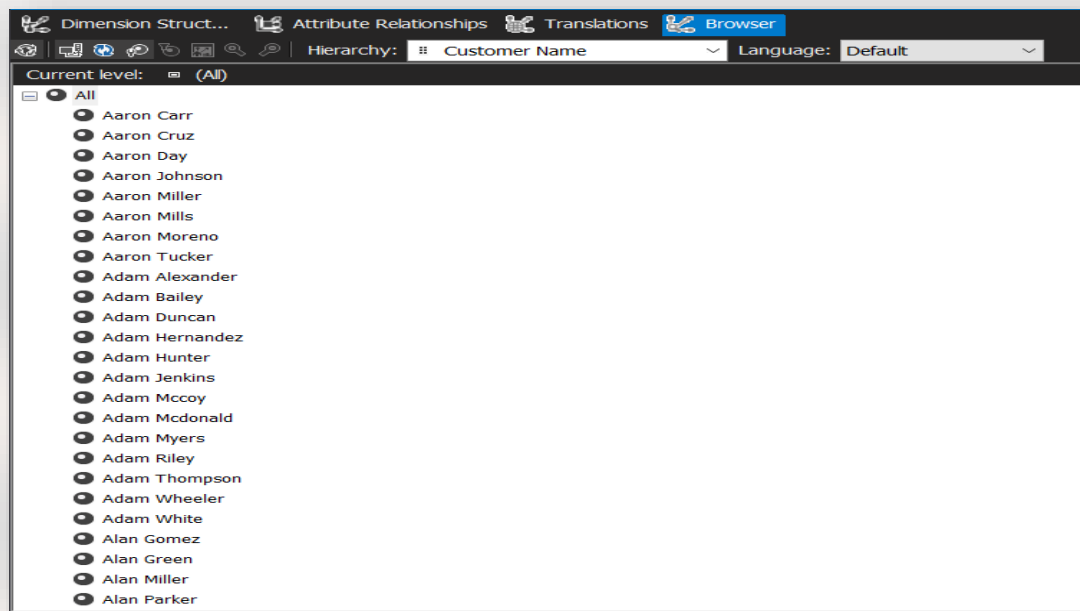


figure23



✓ Les valeurs de l'attribut Customer Name de la dimension Customer
Figure24

Mise en forme des mesures :

Lors de cette partie, l'utilisateur doit calculer les mesures qu'il veut bien ajouter à la table de faits. Néanmoins, vu que ce calcul doit être en DMX, jamais étudié, nous avons opté de les calculer après la liaison entre le cube et le Power BI pour bénéficier du langage DAX.

VI. Dashboard Power BI :

Dans ce présent projet, nous avons utilisé le logiciel Power BI afin de faciliter pour les employés de faciliter l'analyse et prendre la décision au sein de l'entreprise d'une façon rapide et efficace.

Nous proposons deux tableaux de bord pour étudier les données de ventes au détail des différentes mesures par rapport à plusieurs métriques :

Page de garde :

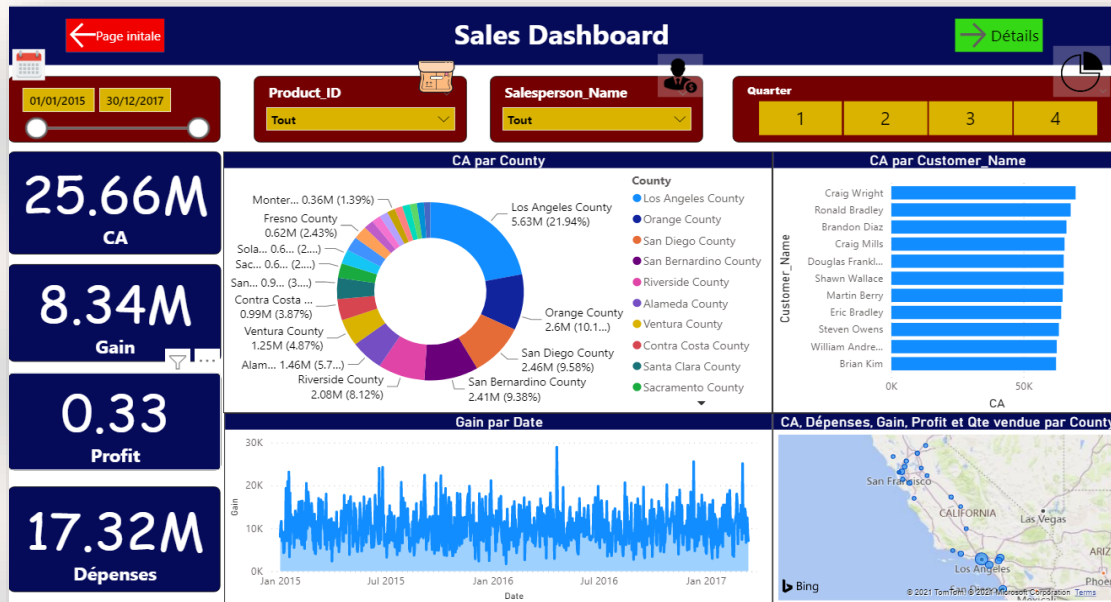


Les interprétations :

Commençons par la 1^{ère} page de ce projet Power Bi qui sera bien sûr la page de garde contenant deux composants principaux :

- Un bouton titré « Rapport des ventes » qui permettra de nous diriger vers le premier tableau de bord.
- Un autre bouton de type informations titré « Détails » permettant de nous diriger vers un autre tableau de bord présentant quelques autres détails à propos de ce rapport de ventes.

Tableau de bord I :



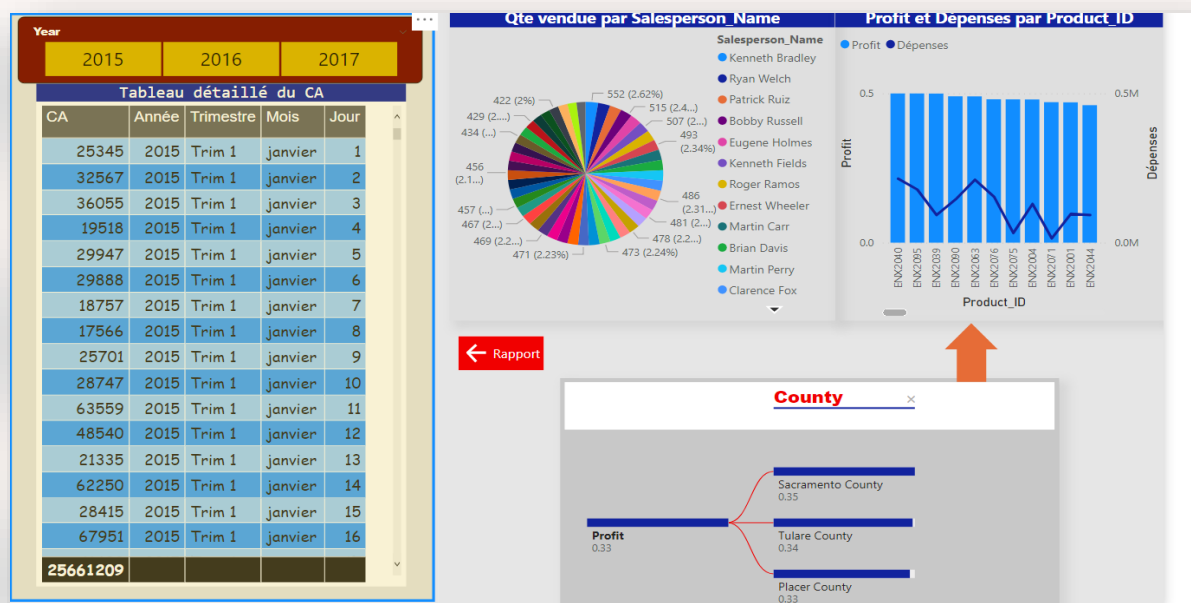
Voici le tableau de bord de gestion des ventes que nous avons créé pour comparer et analyser les différentes relations entre les métriques d'une manière interactive et dynamique dans Power BI.

Les interprétations

- ✓ Nous intégrons également des informations sur les ventes pour vérifier combien nous avons vendu, le chiffre d'affaires, le gain et les dépenses de l'entreprise...comme indiqué ci-dessus à gauche du tableau de bord.
- ✓ Notre tableau comporte en haut deux boutons permettant de se déplacer vers les autres pages. Un bouton pour se diriger vers la page de garde, et l'autre pour se diriger vers la deuxième page de détails.
- ✓ Au-dessous, nous avons introduit des attributs venant des différentes tables de dimensions pour nous permettre de constater les variations des mesures selon eux.
- ✓ Graphique en anneau montre le chiffre d'affaires par région dans l'état de Californie.
- ✓ Graphique à barres groupées indique les clients en se basant sur leur part du chiffre d'affaires.

- ✓ Graphe en courbe illustre la variation du gain par rapport à 3 ans (Nous pouvons préciser la durée à l'aide du slicer en haut)
- ✓ Carte géographique de Californie intégrant toutes les mesures (CA, profit, gain, dépenses et quantité vendue) visualise des cercles pointant sur les villes de l'état Californie avec des différentes tailles selon l'importance des ventes dans chaque ville.

Tableau de garde II :



- ✓ Ce tableau de bord consiste à fournir des détails supplémentaires à propos de CA durant les 3 années, jour par jour, ainsi que d'autres informations et comparaisons entre le profit et les dépenses et la quantité vendue selon différentes métriques.