ENSIAS

Université Mohammed V de Rabat

MÉMOIRE DE PROJET DE DATAWAREHOUSE

Création et Mise en œuvre d'un Datawarehouse pour l'Analyse des données de vente sur Internet

Réalisé par :

Derrouich Jaouhar

Génie logiciel (GL1)

Encadré par

Mme. El Asri Bouchra

ANNÉE UNIVERSITAIRE 2020 – 2021

Table des matières

INTRODUCTION **4**

1. Tâche 1 : Création du datawarehouse : **5**
1.1. Création de LightAdventureWorksDW : 5
1.2. Création des dimensions : 5

2. Tâche 2 : Load LightAdventureWorksDW **8**
2.1. Chargement de dimension Customers : 8
2.2. Chargement de dimension Products : 10
2.3. Chargement de dimension Dates : 11
2.4. Chargement de dimension InternetSales : 11
2.5. Exécution et début de chargement des données dans l'entrepôt: 13

3. Tâche 3 : Création du cube **13**
3.1. Dimension Dates 13
3.2. Dimension Customers 14
3.3. Dimension Products 14
3.4. Data source view 15

4. Tâche 4 : Exploration des données avec SSRS, EXCEL et le web **16**
4.1. SSRS 16
4.2. EXCEL 18

Conclusion : **20**

Table des figures

Figure 1 Schéma en étoile 7
Figure 2 Slowly Changing Dimension Customer 8
Figure 3 Slowly Changing Dimention Columns for Customers 9
Figure 4 Slowly Changing Dimension Products 10
Figure 5 Slowly Changing Dimention Columns for Products 10
Figure 6 Load Dates 11
Figure 7 Slowly Changing Dimension InternetSales 11
Figure 8 Slowly Changing Dimention Columns for InternetSales 12
Figure 9 Loading data 13
Figure 10 Dates attributes Relationships 13
Figure 11Customers attributes Relationships 14
Figure 12Products attributes Relationships 14
Figure 13 Data source view 15
Figure 14 Internet Sales Count per Country 16
Figure 15 Sales Count for each country in a year per product 17
Figure 16 Sales count per Year 17
Figure 17 Sales count per Country 18
Figure 18 Graph of internet Sales Count per country 18
Figure 19 graph of internet Sales count per year 19
Figure 20 Graph of sales count per country 19
Figure 21 Graph of sales count for each country in year per product 20

INTRODUCTION

L’informatique décisionnelle est l'informatique à l'usage des décideurs et des dirigeants d'entreprises. Elle désigne les moyens, les outils et les méthodes qui permettent de collecter, consolider, modéliser et restituer les données, matérielles ou immatérielles, d'une entreprise en vue d'offrir une aide à la décision et de permettre à un décideur d’avoir une vue d’ensemble de l’activité traitée.

Les données opérationnelles sont extraites périodiquement de sources hétérogènes : fichiers plats, fichiers Excel, base de données (DB2, Oracle, SQL Server, etc.), service web, données massives et stockées dans un entrepôt de données. Cet entrepôt porte le nom de datawarehouse. Un datawarehouse représente une base de données : les données étant intégrées, non volatiles, et historisées. Grâce à la plateforme SQL Server Integration Services (SSIS), cet entrepôt central sera rempli. Mais avant, il est indispensable de définir sa structure. Avant de remplir le datawarehouse dans SSIS, la conception de celui-ci s’impose. Viennent donc les notions de table de faits et table de dimension.

Pour notre cas ici on va créer un datawarehouse pour l’analyse des données de vente sur Internet en passant par plusieurs tâches.

1. Tâche 1 : Création du datawarehouse :

1.1. Création de LightAdventureWorksDW :

Dans cette étape il s'agit de mettre en place un entrepôt de données LightAdventureWorksDW. Et cela implique la création du schéma ainsi que les dimensions et les tables de faits.

Pour la créer il faut d'abord Lancer SSMS et se connecter à votre instance de SQL Server. Ouvrir une nouvelle fenêtre de requête en cliquant sur le bouton Nouvelle requête. Puis d'après le contexte de la base de données master, on va créer une nouvelle base de données appelée LightAdventureWorksDW utilisons script suivante :

```
CREATE DATABASE [LIGHTADVENTUREWORKSDW]
 CONTAINMENT = NONE
 ON  PRIMARY
( NAME = N'LIGHTADVENTUREWORKSDW', FILENAME = N'C:\Program Files\Microsoft SQL
Server\MSSQL15.MSSQLSERVER\MSSQL\DATA\LIGHTADVENTUREWORKSDW.mdf' ,
                                    SIZE = 307200KB , MAXSIZE = UNLIMITED, FILEGROWTH = 10240KB
)
 LOG ON
( NAME = N'LIGHTADVENTUREWORKSDW_log', FILENAME = N'C:\Program Files\Microsoft SQL
Server\MSSQL15.MSSQLSERVER\MSSQL\DATA\LIGHTADVENTUREWORKSDW_log.ldf' ,
                                    SIZE = 3094592KB , MAXSIZE = 2048GB , FILEGROWTH = 10%)
 WITH CATALOG_COLLATION = DATABASE_DEFAULT
GO
IF OBJECT_ID('dbo.SeqCustomerDwKey','SO') IS NOT NULL
DROP SEQUENCE dbo.SeqCustomerDwKey;
GO
CREATE SEQUENCE dbo.SeqCustomerDwKey AS INT
START WITH 1
INCREMENT BY 1;
GO
```

1.2. Création des dimensions :

Pour créer les dimensions on a exécuté les requêtes suivant :

- La dimension Customers :

```
USE LIGHTADVENTUREWORKSDW;
GO
CREATE TABLE dbo.Customers
(
CustomerDwKey INT NOT NULL,
CustomerKey INT NOT NULL,
FullName NVARCHAR(150) NULL,
EmailAddress NVARCHAR(50) NULL,
BirthDate DATE NULL,
MaritalStatus NCHAR(1) NULL,
Gender NCHAR(1) NULL,
Education NVARCHAR(40) NULL,
Occupation NVARCHAR(100) NULL,
City NVARCHAR(30) NULL,
StateProvince NVARCHAR(50) NULL,
```

```
CountryRegion NVARCHAR(50) NULL,
Age AS
CASE
WHEN DATEDIFF(yy, BirthDate, CURRENT_TIMESTAMP) <= 40
THEN 'Younger'
WHEN DATEDIFF(yy, BirthDate, CURRENT_TIMESTAMP) > 50
THEN 'Older'
ELSE 'Middle Age'
END,
CurrentFlag BIT NOT NULL DEFAULT 1,
CONSTRAINT PK_Customers PRIMARY KEY (CustomerDwKey)
);
GO
```

- La dimension Products :

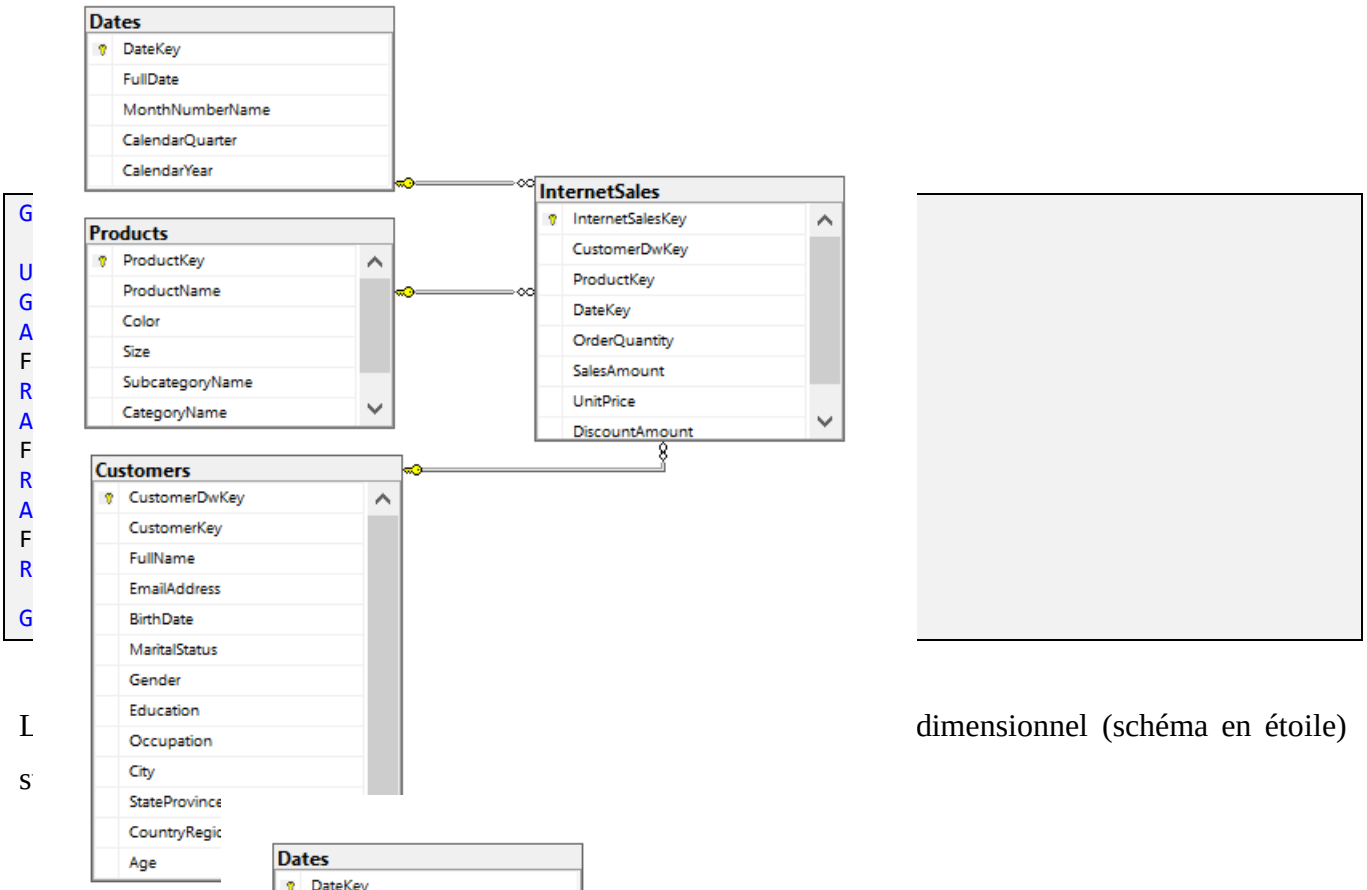
```
USE LIGHTADVENTUREWORKSDW;
GO
CREATE TABLE dbo.Products
(
ProductKey INT NOT NULL,
ProductName NVARCHAR(50) NULL,
Color NVARCHAR(15) NULL,
Size NVARCHAR(50) NULL,
SubcategoryName NVARCHAR(50) NULL,
CategoryName NVARCHAR(50) NULL,
CONSTRAINT PK_Products PRIMARY KEY (ProductKey)
);
GO
```

- La dimension Dates

```
CREATE TABLE dbo.Dates
(
DateKey INT NOT NULL identity(1,1),
FullDate DATE NOT NULL,
MonthNumberName NVARCHAR(15) NULL,
CalendarQuarter TINYINT NULL,
CalendarYear SMALLINT NULL,
CONSTRAINT PK_Dates PRIMARY KEY (DateKey)
);
GO
```

- La dimension InternetSales

```
USE LIGHTADVENTUREWORKSDW;
GO
CREATE TABLE dbo.InternetSales
(
InternetSalesKey INT NOT NULL IDENTITY(1,1),
CustomerDwKey INT NOT NULL,
ProductKey INT NOT NULL,
DateKey INT NOT NULL,
OrderQuantity SMALLINT NOT NULL DEFAULT 0,
SalesAmount MONEY NOT NULL DEFAULT 0,
UnitPrice MONEY NOT NULL DEFAULT 0,
DiscountAmount FLOAT NOT NULL DEFAULT 0,
CONSTRAINT PK_InternetSales
PRIMARY KEY (InternetSalesKey)
);
```



L dimensionnel (schéma en étoile)

s

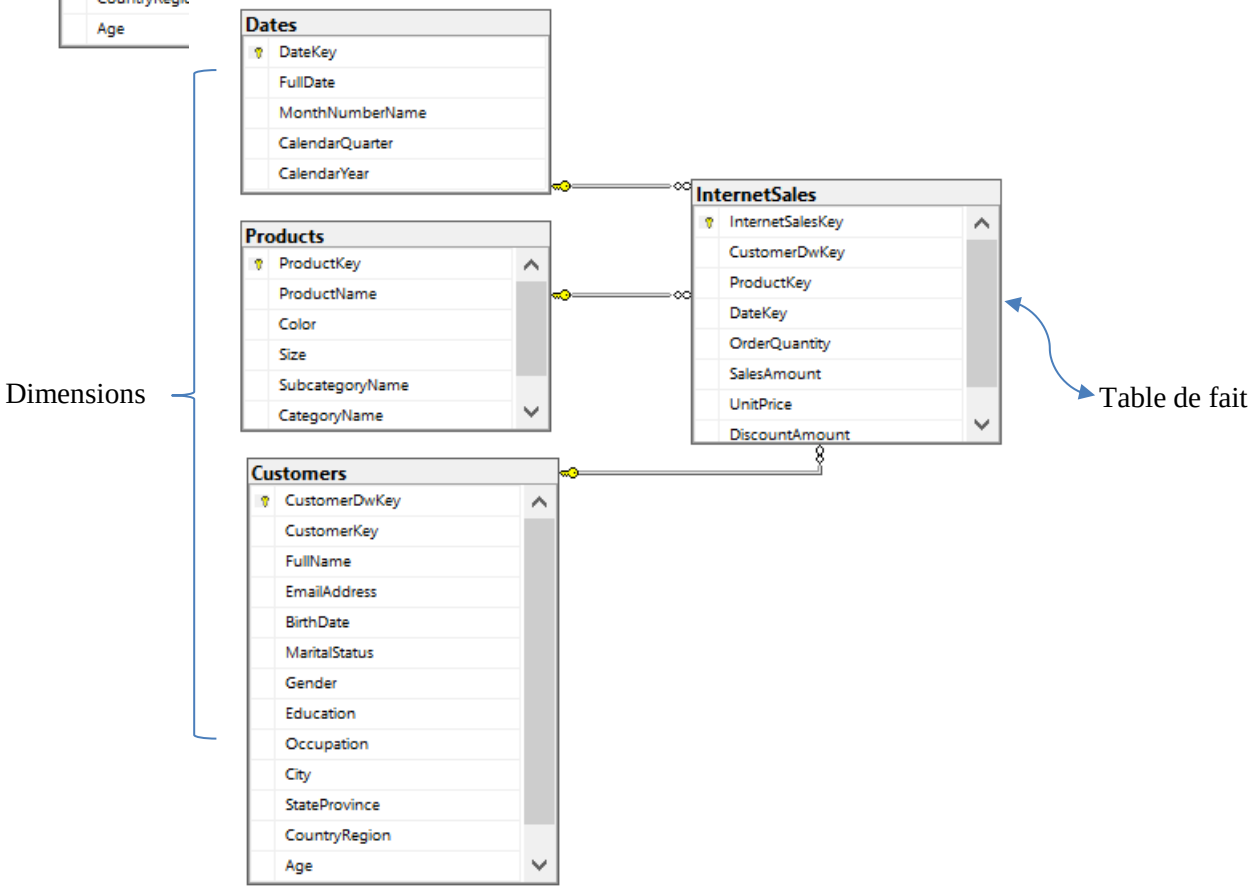


Figure 1 Schéma en étoile

En effet, la table de faits contient les données (faits) qui représentent la valeur des mesures sur lesquelles va porter l'analyse, quant aux dimensions, elles présentent les axes avec lesquels on veut faire l'analyse. On trouve une dimension client, une dimension produit, et date.

2. Tâche 2 : Load LightAdventureWorksDW

Dans cette tâche il s'agit d'alimenter le datawarehouse LightAdventureWorksDW depuis AdventureWorks2012. Pour cela on va utiliser le Slowly Changing Dimensions car on alimente depuis une base de données transactionnelle.

2.1. Chargement de dimension Customers :

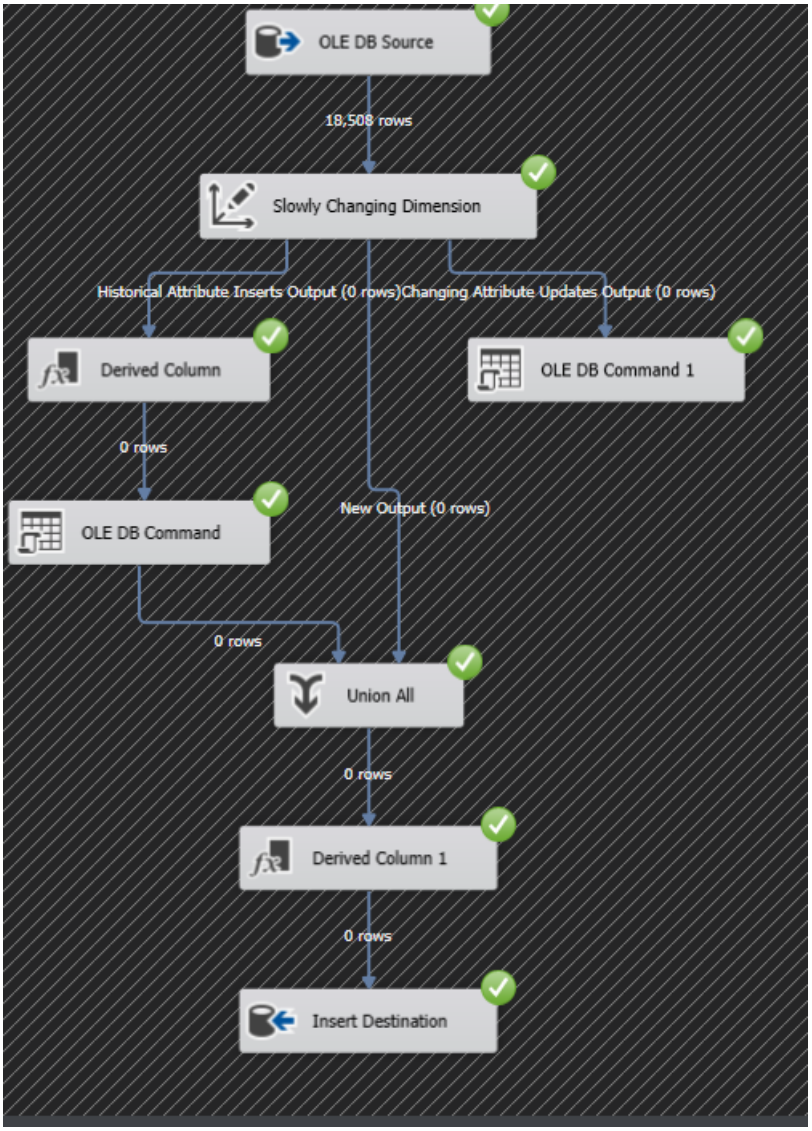


Figure 2 Slowly Changing Dimension Customer

On a utilisé dans la source OLE DB pour récupérer les données de AdventureWorks2012 la requête SQL suivante :

```
SELECT NEXT VALUE FOR LIGHTADVENTUREWORKSDW.dbo.SeqCustomerDwKey AS CustomerDwKey, C.CustomerID
AS CustomerKey ,
              CONCAT(P.FirstName,P.MiddleName,P.LastName) AS FullName ,
              E.EmailAddress AS EmailAddress ,
              ADDR.City AS City,
              SP.Name AS StateProvince,
              CR.Name AS CountryRegion

FROM   AdventureWorks2012.Sales.Customer C,
              AdventureWorks2012.Person.Person P,
              AdventureWorks2012.Person.EmailAddress E ,
              AdventureWorks2012.Person.BusinessEntityAddress A ,
              AdventureWorks2012.Person.Address ADDR,
              AdventureWorks2012.Person.StateProvince SP,
              AdventureWorks2012.Person.CountryRegion CR

WHERE  C.PersonID = P.BusinessEntityID
       AND P.BusinessEntityID = E.BusinessEntityID
       AND P.BusinessEntityID = A.BusinessEntityID
       AND A.AddressID = ADDR.AddressID
       AND ADDR.StateProvinceID = SP.StateProvinceID
       AND SP.CountryRegionCode = CR.CountryRegionCode
```

Et pour les attributs, j'ai choisi comme *Historical attribute* les colonnes City, CountryRegion, EmailAdress et StateProvince puisque pour avoir une trace du la version précédente du client lorsque l'une de ces colonnes a changé (on rappelle que l'objectif est d'analyser les ventes donc on est besoins de savoir par exemple les lieux des clients).

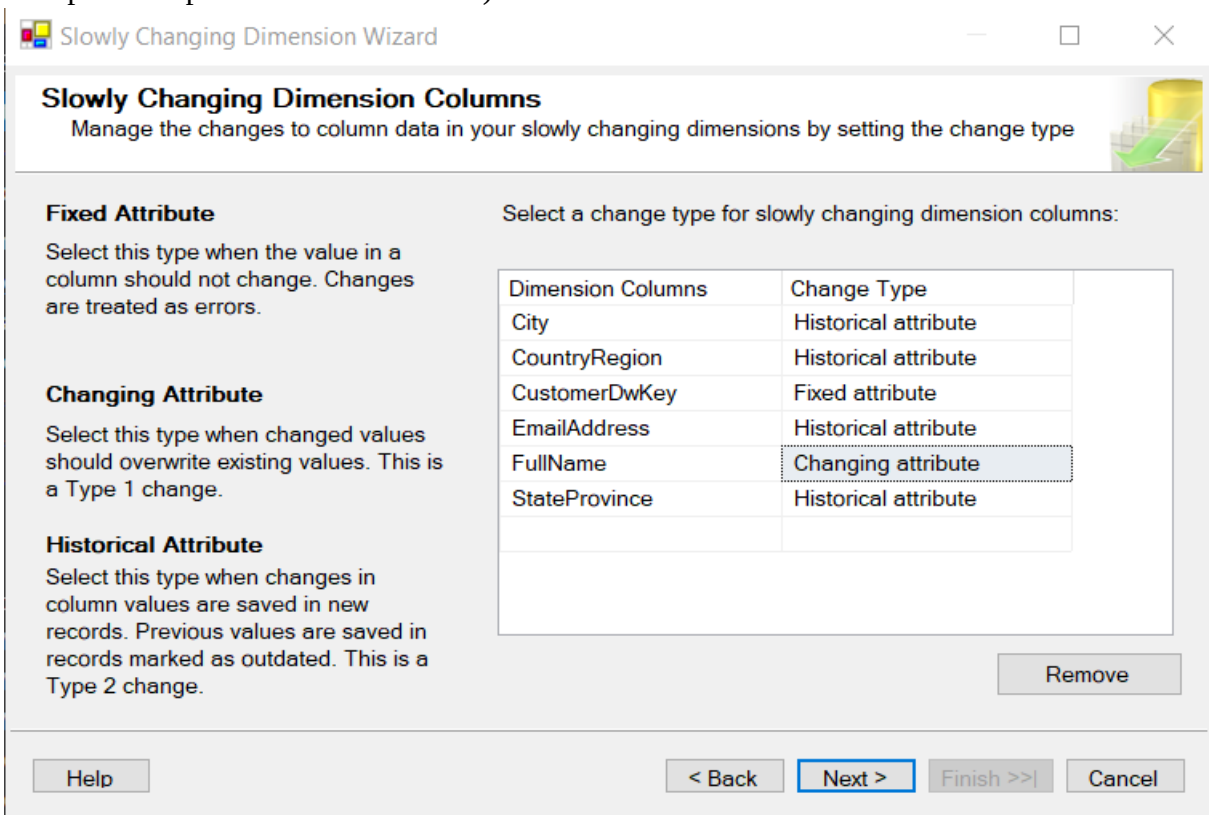


Figure 3 Slowly Changing Dimention Columns for Customers

2.2. Chargement de dimension Products :

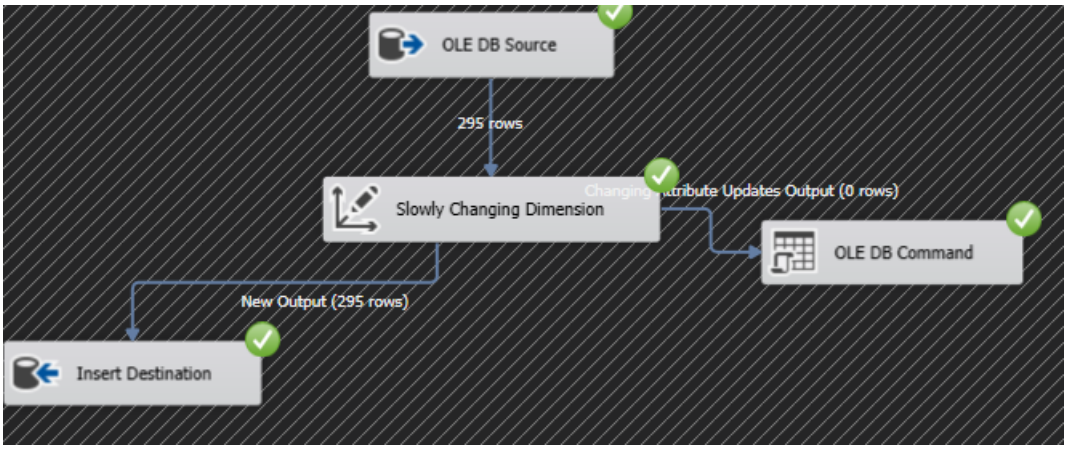


Figure 4 Slowly Changing Dimension Products

On a utilisé dans la source OLE DB pour récupérer les données de AdventureWorks2012 la requête SQL suivante :

```
SELECT p.ProductID as ProductKey, p.Name as ProductName, p.Color as Color, p.Size as Size,
s.Name as SubcategoryName, c.Name as CategoryName

FROM AdventureWorks2012.Production.Product p, AdventureWorks2012.Production.ProductCategory c,
AdventureWorks2012.Production.ProductSubcategory s

WHERE p.ProductSubcategoryID = s.ProductSubcategoryID and s.ProductCategoryID =
c.ProductCategoryID;
```

Et pour les attributs, j'ai choisi tous les colonnes comme *Changing attribute* puisque pour moi une fois l'une de ces colonnes a changé ça sert à rien d'historisé la version précédente.

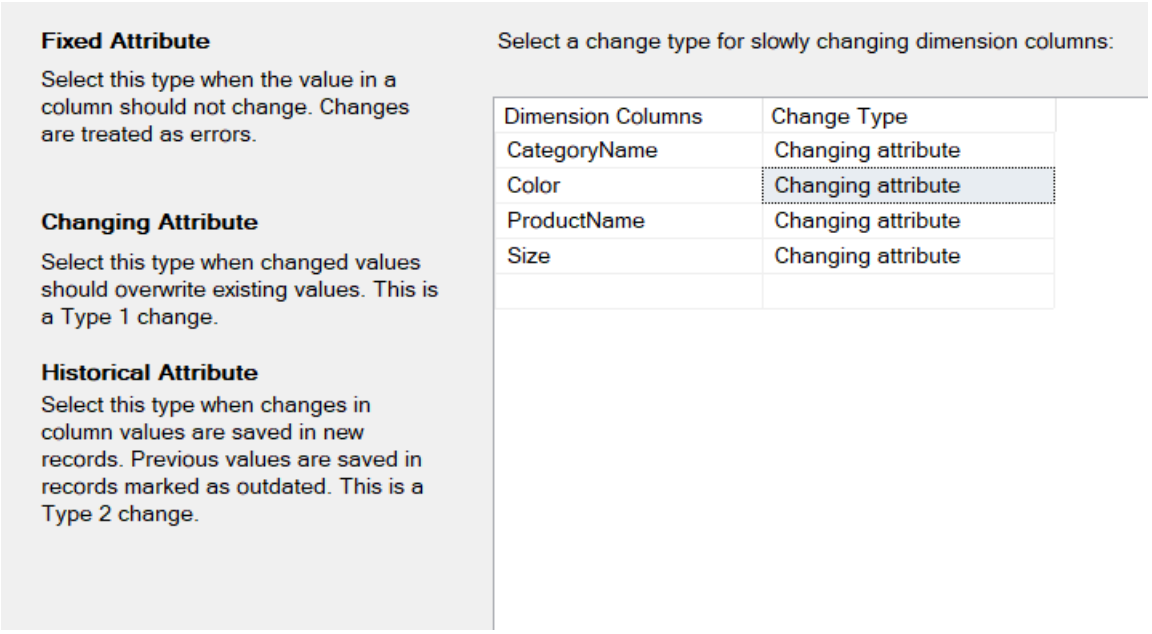


Figure 5 Slowly Changing Dimention Columns for Products

2.3. Chargement de dimension Dates :

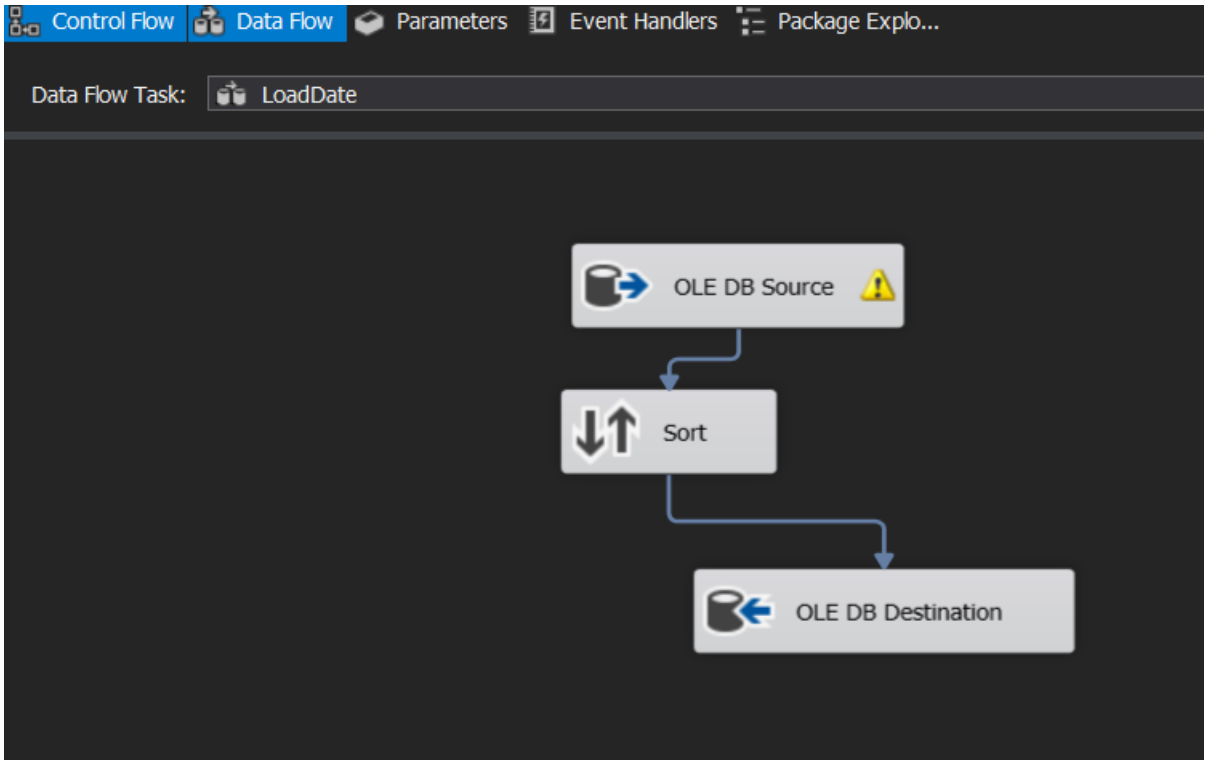


Figure 6 Load Dates

On a utilisé dans la source OLE DB pour récupérer les données de AdventureWorks2012 la requête SQL suivante :

```
SELECT SalesOrderID as DateKey ,CONVERT(date, OrderDate) as FullDate,
SUBSTRING(CONVERT(CHAR(8), OrderDate, 112), 5, 2)
+ ' ' + DATENAME(month,OrderDate) as MonthNumberName,
DATEPART(QUARTER, OrderDate ) as CalendarQuarter,YEAR(OrderDate) as CalendarYear
FROM AdventureWorks2012.Sales.SalesOrderHeader
```

2.4. Chargement de dimension InternetSales :

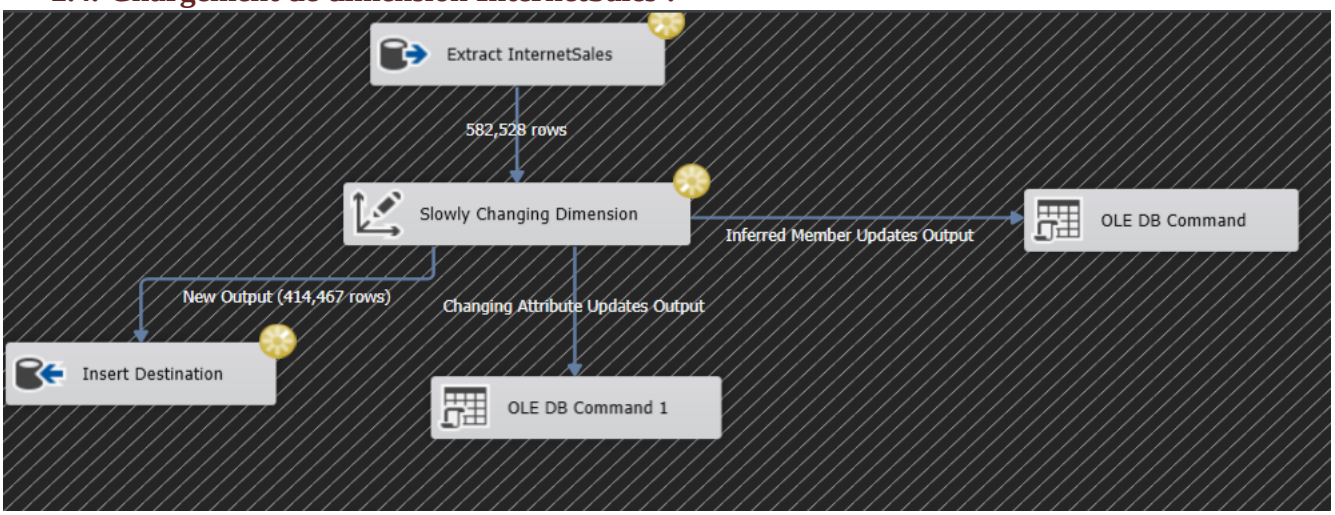


Figure 7 Slowly Changing Dimension InternetSales

On a utilisé dans la source OLE DB pour récupérer les données de AdventureWorks2012 la requête SQL suivante :

```
SELECT
C.CustomerDwKey as CustomerID,
SD.ProductID,
D.Datekey,
SD.OrderQty as OrderQuantity,
SD.OrderQty*SD.UnitPrice AS SalesAmount ,
SD.UnitPrice,
convert(Float,SD.UnitPriceDiscount*convert(Float,SD.OrderQty)) AS DiscountAmount
FROM
AdventureWorks2012.Sales.SalesOrderDetail AS SD,
AdventureWorks2012.Sales.SalesOrderHeader AS SH,
LIGHTADVENTUREWORKSDW.dbo.Customers AS C,
LIGHTADVENTUREWORKSDW.dbo.Products AS P,
LIGHTADVENTUREWORKSDW.dbo.Dates AS D
WHERE
SD.SalesOrderID = SH.SalesOrderID
AND C.CustomerKey = SH.CustomerID
AND P.ProductKey = SD.ProductID
AND D.Fulldate = SH.OrderDate;
```

Et pour les attributs, j'ai choisi comme *Changing attribute* toutes les colonnes sauf les foreign keys.

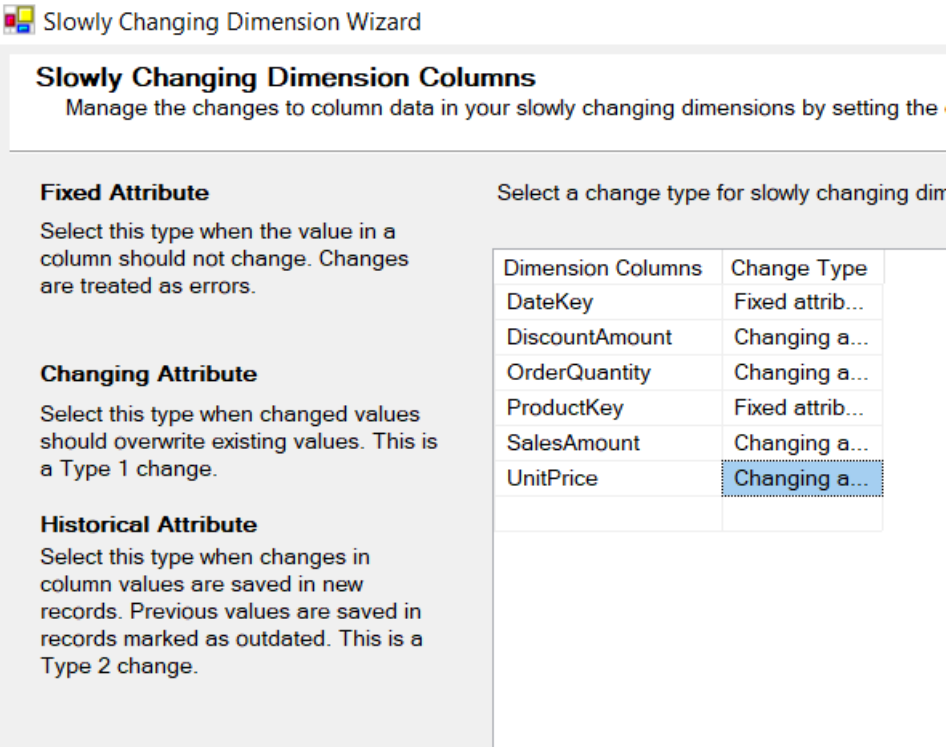


Figure 8 Slowly Changing Dimention Columns for InternetSales

2.5. Exécution et début de chargement des données dans l'entrepôt:

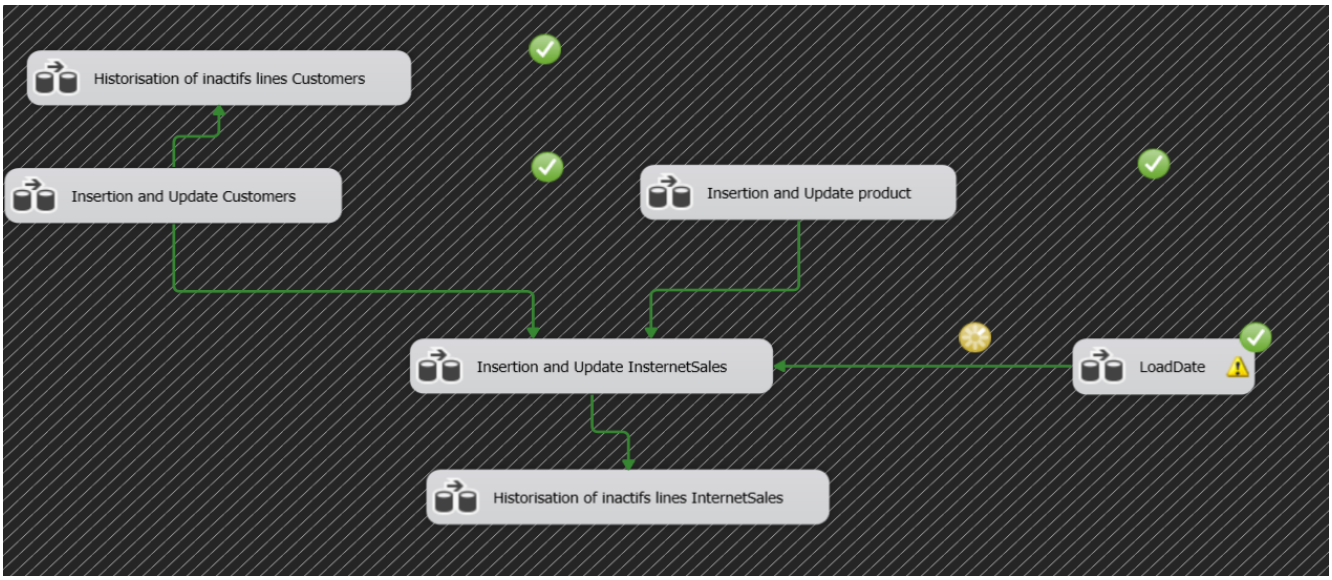


Figure 9 Loading data

Il est évident qu'on a des Historisation puisqu'on a des attributs historisés.

3. Tâche 3 : Création du cube

A l'aide de SSAS, on va créer le cube OLAP correspondant au sujet Vente. En optimisant la construction du cube.

3.1. Dimension Dates

On a créé la hiérarchie suivante :

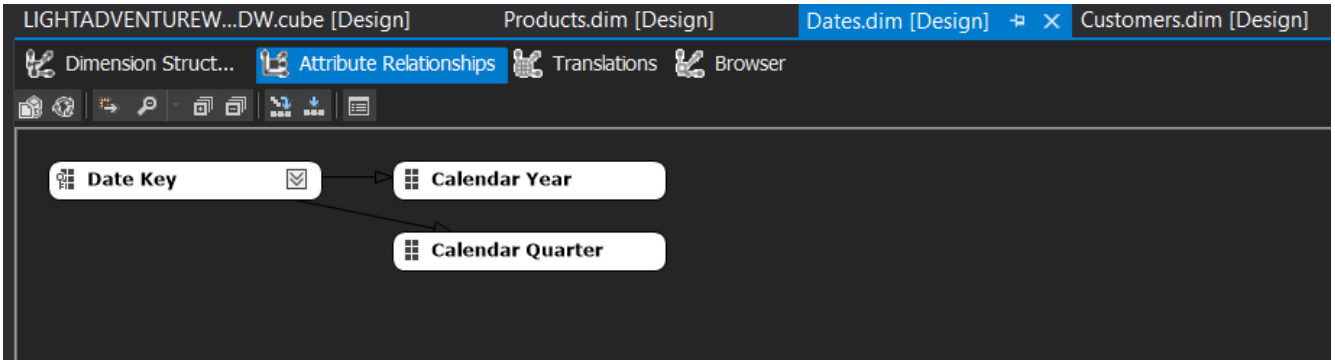


Figure 10 Dates attributes Relationships

3.2. Dimension Customers

On a créé la hiérarchie suivante :

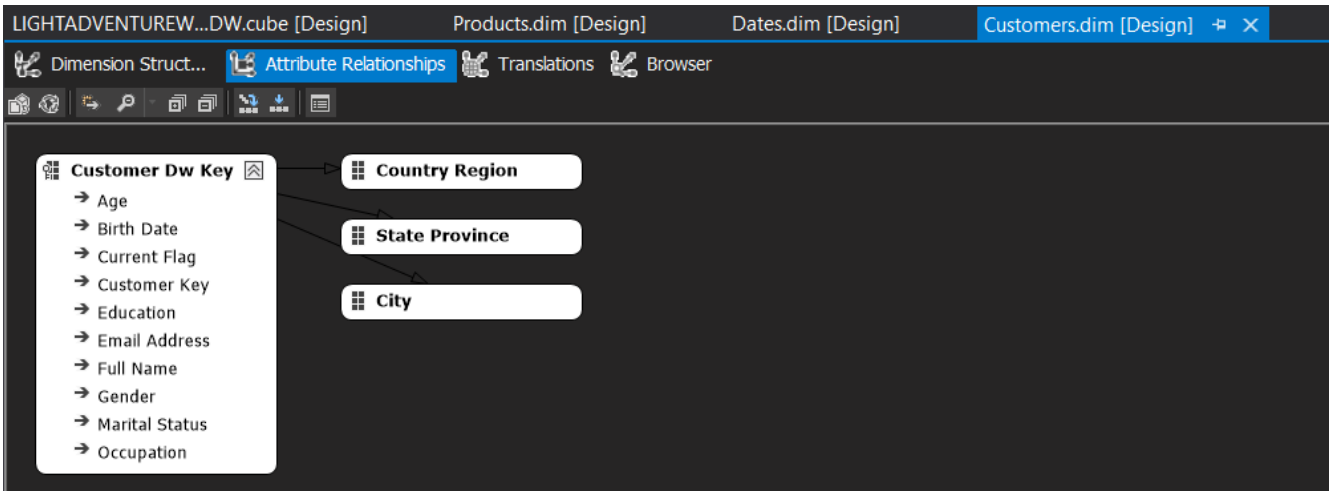


Figure 11Customers attributes Relationships

3.3. Dimension Products

On a créé la hiérarchie suivante :

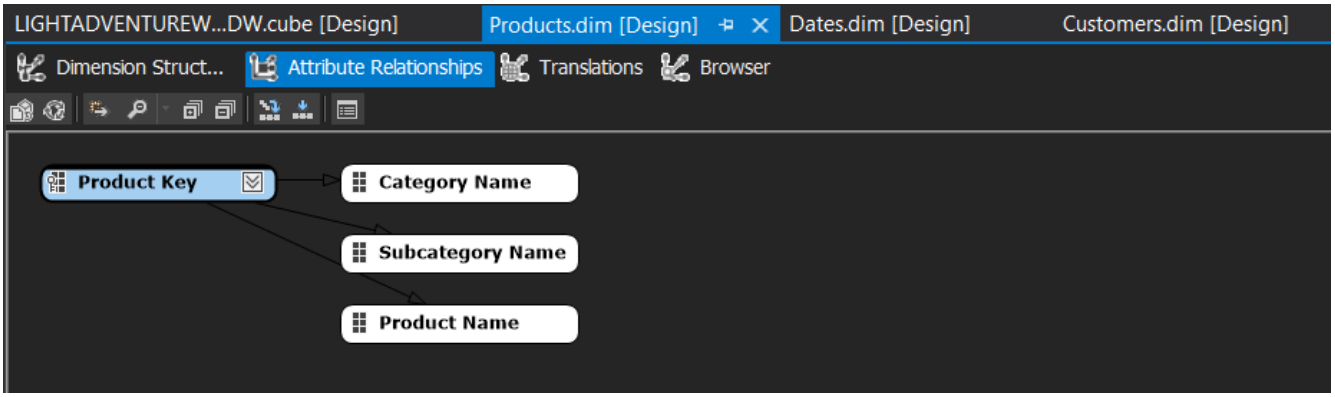


Figure 12Products attributes Relationships

3.4. Data source view

Donc on obtient le résultat suivant :

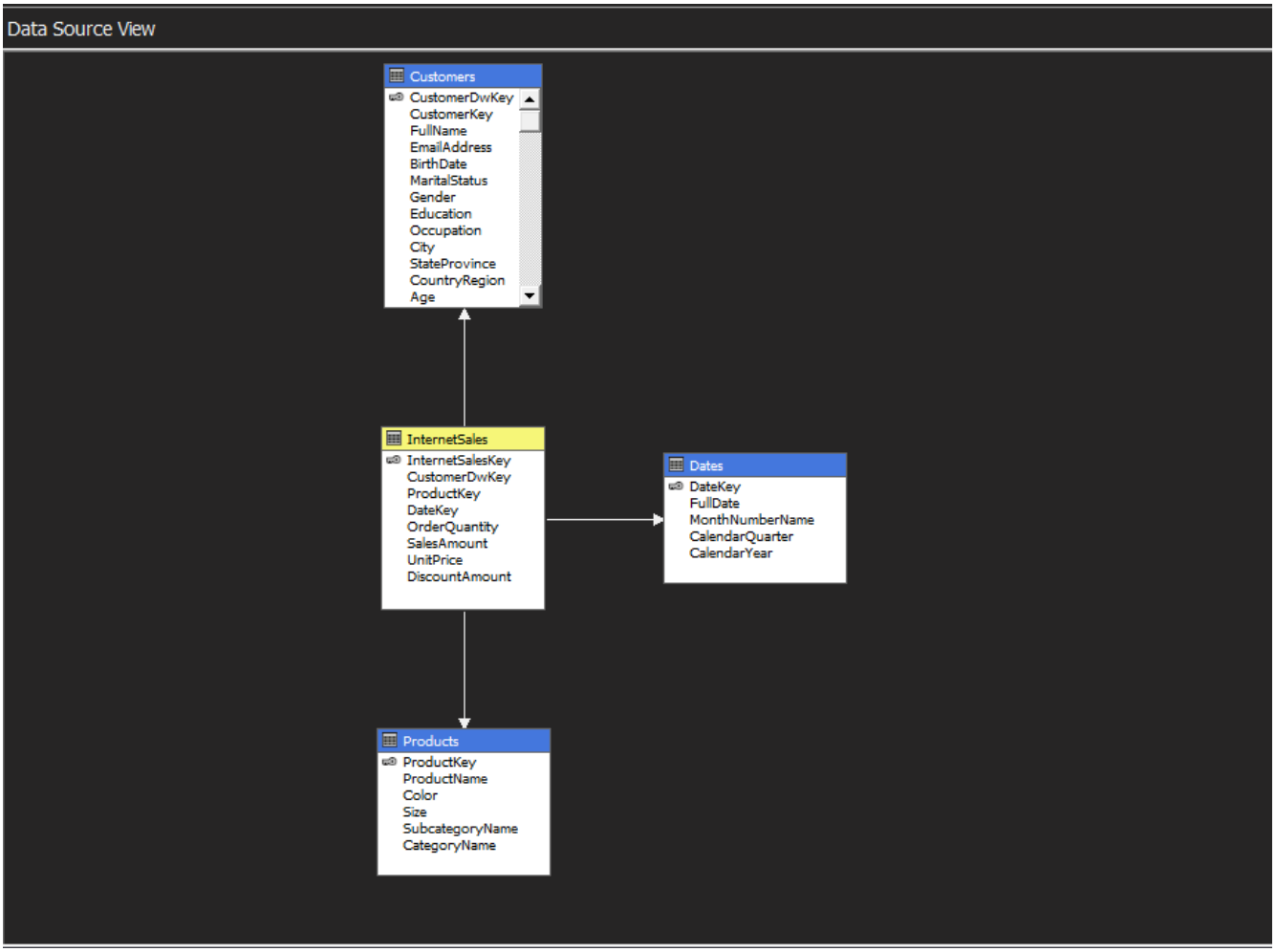


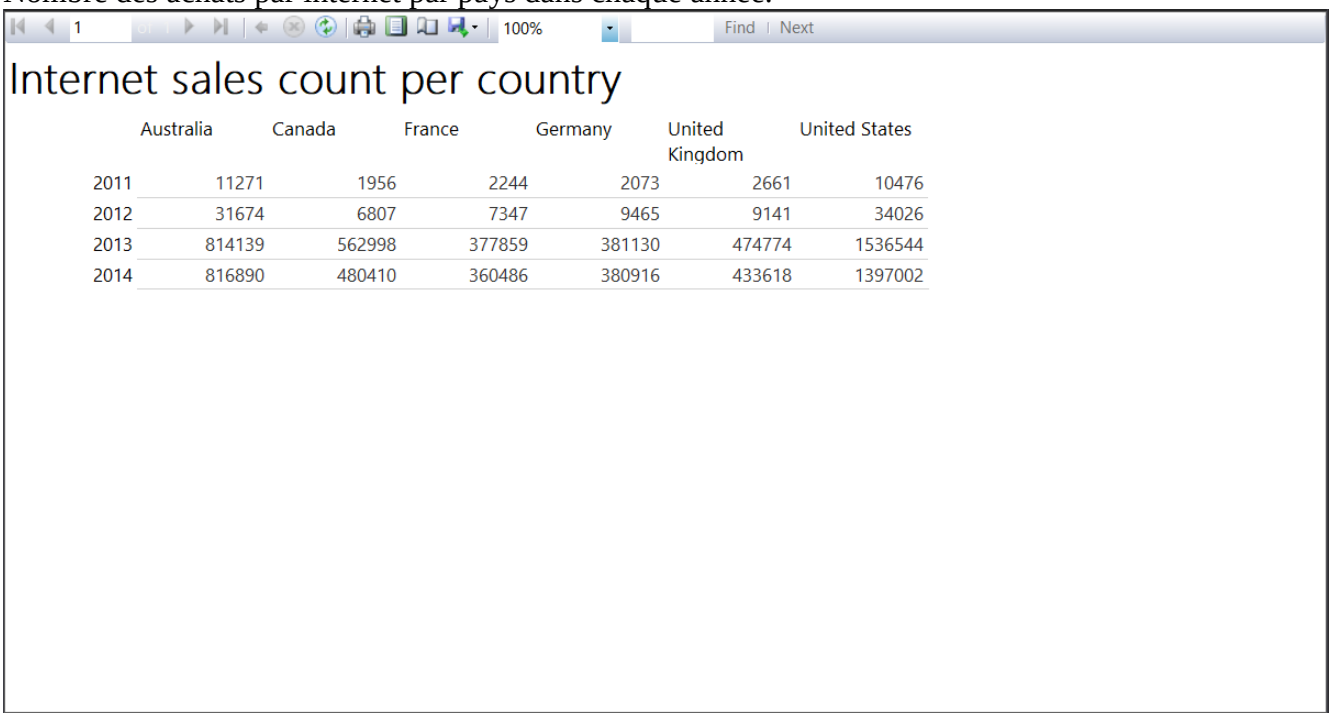
Figure 13 Data source view

4. Tâche 4 : Exploration des données avec SSRS, EXCEL et le web

Dans cette étape, on va créer des rapports types matriciels avec SSRS bien documentés, des graphes avec Excel et une sélection de données à mettre sur le web.

4.1. SSRS

Sous la SSRS on a générer les matrices suivant :
Nombre des achats par internet par pays dans chaque année.



Internet sales count per country

	Australia	Canada	France	Germany	United Kingdom	United States
2011	11271	1956	2244	2073	2661	10476
2012	31674	6807	7347	9465	9141	34026
2013	814139	562998	377859	381130	474774	1536544
2014	816890	480410	360486	380916	433618	1397002

Figure 14 Internet Sales Count per Country

1 of 27 50% Find | Next

Sales Count for each country in year

		2011 Australia	Canada	France	Germany	United Kingdom	United States	2012 Australia	Canada	France	Germany	United Kingdom	United States	2013 Australia	Canada	France	Germany	United Kingdom	United Stat
Accessories	All-Purpose Bike Stand													5485	1730	1511	2186	2221	8
	Bike Wash - Dissolver													14892	12830	3943	4725	5029	30
	Fender Set - Mountain													23054	33103	6940	10935	9377	75
	Hitch Rack - 4-Bike													2294	4841	1559	1845	2714	10
	HL Mountain Tire													17565	20168	5407	5150	5392	49
	HL Road Tire													14127	14555	6038	3260	4677	33
	Hydration Pack - 70 oz.													12200	9274	4986	2782	3504	21
	LL Mountain Tire													14163	11173	5105	7319	5903	25
	LL Road Tire													16420	6542	9128	11575	13528	22
	ML Mountain Tire													15316	19746	7136	5508	7579	38
	ML Road Tire													18664	6754	9279	8117	9602	17
	Mountain Bottle Cage													19582	28339	11360	12521	13191	61
	Mountain Tire Tube													39303	54953	14082	16677	17303	108
	Patch Kit/8 Patches													50077	38525	21013	18234	25254	97
	Road Bottle Cage													31688	8106	13619	13820	16923	38
	Road Tire Tub													37873	22293	20511	23238	30525	56
	Sport-100 Helmet, Black													27689	23712	14138	17150	18920	57
	Sport-100 Helmet, Blue													28319	19397	12908	14936	18962	55
	Sport-100 Helmet, Red													32857	20052	16899	17817	20438	57
	Touring Tire													8392	5502	12477	11299	12962	20
	Touring Tire Tube													15611	10772	22551	14667	23578	27
	Water Bottle - 30 oz.													52725	47955	24895	29078	36376	124

Figure 15 Sales Count for each country in a year per product

Nombre des achats par an :

1 of 1

Sales count per year

Calendar Year	Internet Sales Count
2011	
	30681
2012	
	98460
2013	
	4147444
2014	
	3869322

Figure 16 Sales count per Year

Nombre des achats par pays :

1 100% Find Next

Sales count per Country

Country Region	Internet Sales Count
Australia	
	1673974
Canada	
	1052171
France	
	747936
Germany	
	773584
United Kingdom	
	920194
United States	
	2978048

Figure 17 Sales count per Country

4.2. EXCEL

Sous Excel on a générer les graphes suivant :

Nombre des achats par internet par pays :

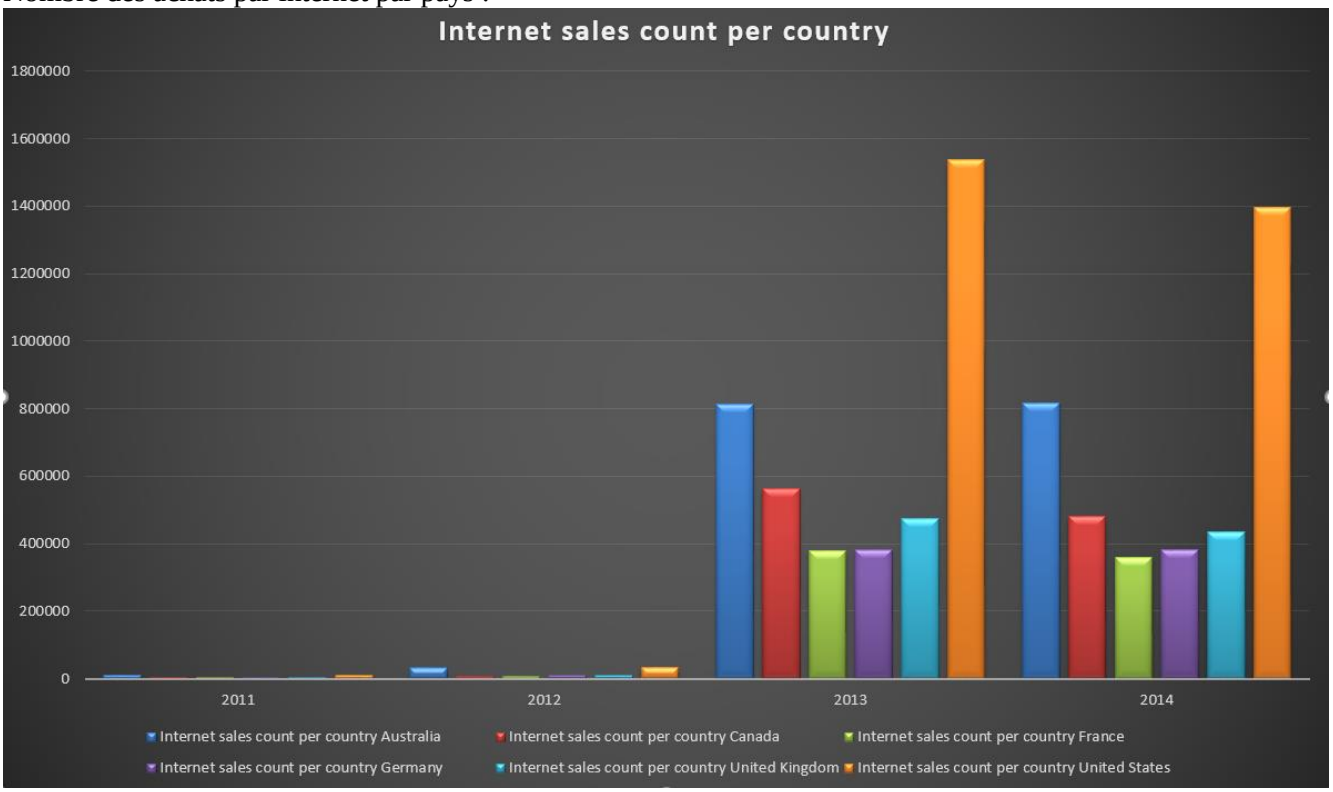


Figure 18 Graph of internet Sales Count per country

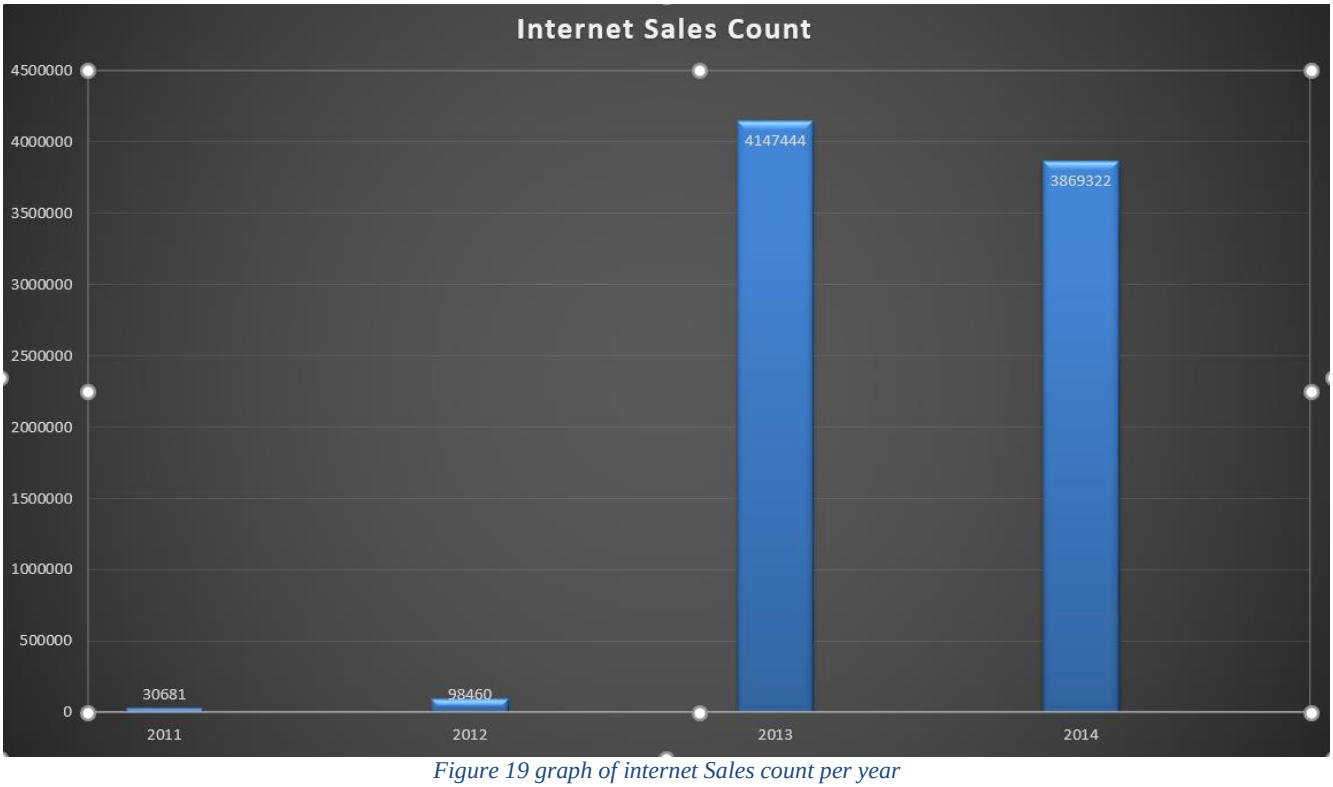


Figure 19 graph of internet Sales count per year

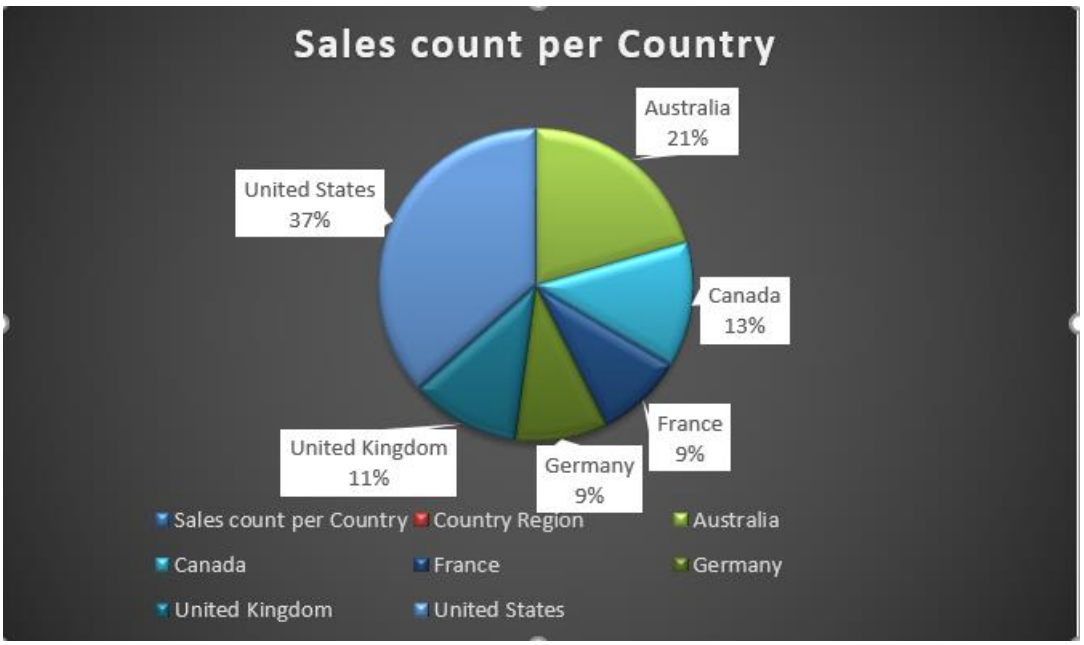


Figure 20 Graph of sales count per country

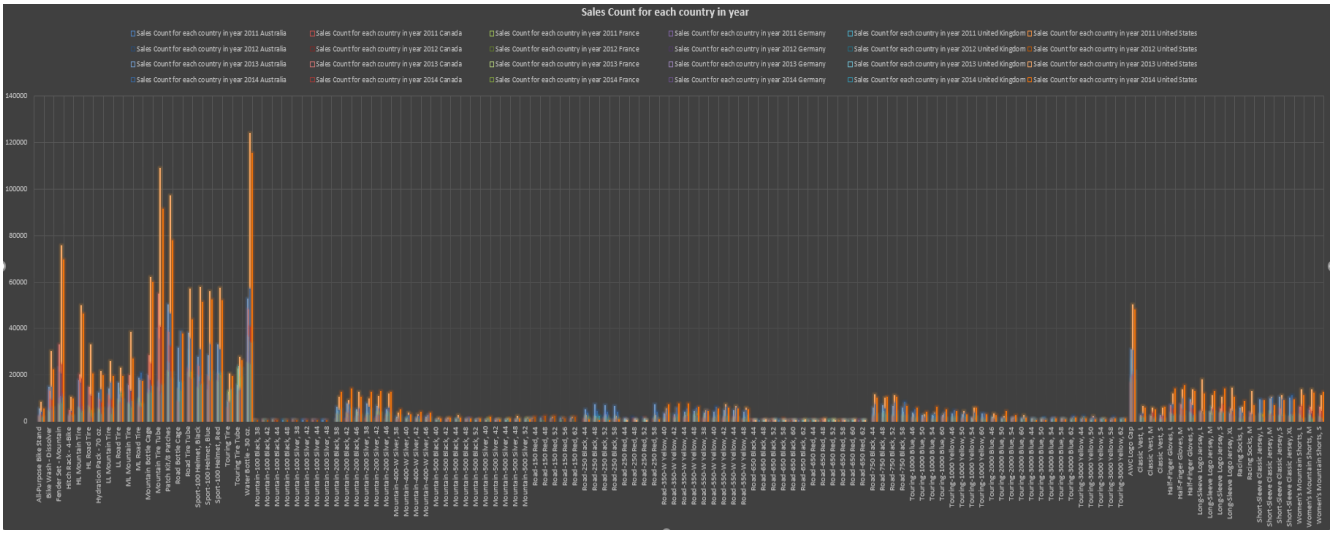


Figure 21 Graph of sales count for each country in year per product

Conclusion :

L’utilisation des Datawarehouse est très demandé aujourd’hui dans le monde de l’entreprise, demande qui répond au besoin toujours plus présent des utilisateurs d’avoir les bonnes informations au bon moment Il entre dans le cadre de l'informatique décisionnelle ; son but est de fournir un ensemble de données servant de référence unique, utilisée pour la prise de décisions dans l'entreprise par le biais de statistiques et de rapports réalisés via des outils de Reporting. D'un point de vue technique, il sert surtout à 'délester' les bases de données opérationnelles des requêtes pouvant nuire à leurs performances