



# **Projet DataWarehouse**

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

### **Réalisé par :**

EL ARFAOUI Ikrame (GL2)

EL AZHAR Asmaa (GL2)

### **Encadré par :**

Mme EL ASRI Bouchra



## Table des Matières

<b>1. Introduction générale</b>	5
<b>2. Création du datawarehouse</b>	5
<b>3. Load LightAdventureWorksDW</b>	6
3.1. Table Customer	7
3.2. Table Produit	10
3.3. Table Date	12
3.4. Table de fait : Internet Sales	13
<b>4. Création du cube</b>	14
4.1. Hiérarchies	14
4.2. Le cube	16
<b>5. Génération du rapport</b>	17
5.1. SSRS	17
5.2. Excel	19
<b>6. Conclusion</b>	20

## Table des Figures

Figure 1 : schéma en étoile .....	6
Figure 2 : flux de control global .....	7
Figure 3 : script loadCustomers .....	7
Figure 4 : Customers data flow .....	8
Figure 5 : Colonnes de dimensions à variation lente .....	9
Figure 6 : Customers : Data Mapping .....	10
Figure 7 : Script Load Products .....	10
Figure 8 : Products Data Flow .....	11
Figure 9 : Colonnes de dimensions à variation lente .....	11
Figure 10 : Script LoadDates .....	12
Figure 11 : Dates Data Flow .....	12
Figure 12 : Script Load InternetSales .....	13
Figure 13 : Colonnes de dimensions à variation lente .....	13
Figure 14 : succès de l'alimentation .....	14
Figure 15 : Customers Hierarchy .....	15
Figure 16 : Products Hierarchy .....	15
Figure 17 : Création du cube .....	16
Figure 18 : visualisation des données .....	17
Figure 19 : SalesAmount/CountryRegion .....	17
Figure 20 : Rapport détaillé de chaque année .....	18
Figure 21 : SalesAmount/ year .....	18
Figure 22 : Reporting with excel .....	19
Figure 23 : exemple de graph avec Excel .....	19

## **1. Introduction générale**

L'informatique décisionnelle (BI) est basée sur des données présentes dans le système d'information de l'entreprise et en externe (Internet, bases de données externes).

Ces données sont ensuite stockées dans un datawarehouse pour être analysées.

La première étape d'un projet BI est de créer un entrepôt central pour avoir une vision globale des données de chaque service. 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. Les principaux objets de la base de données sont des dimensions et des tables de faits.

Ce projet consiste à créer, déployer, analyser et explorer un datawarehouse des données de vente sur Internet.

## **2. Création du datawarehouse**

Cette étape consiste à créer le datawarehouse LightAdventureWorksDW. Donc la création du schéma, les dimensions et les tables de faits.

Les caractéristiques du DatawareHouse LightAdventureWorksDW :

- Table de fait : InternetSales

- Dimensions :

  - Customers

  - Products

  - Dates

➤ Schéma global en étoile :

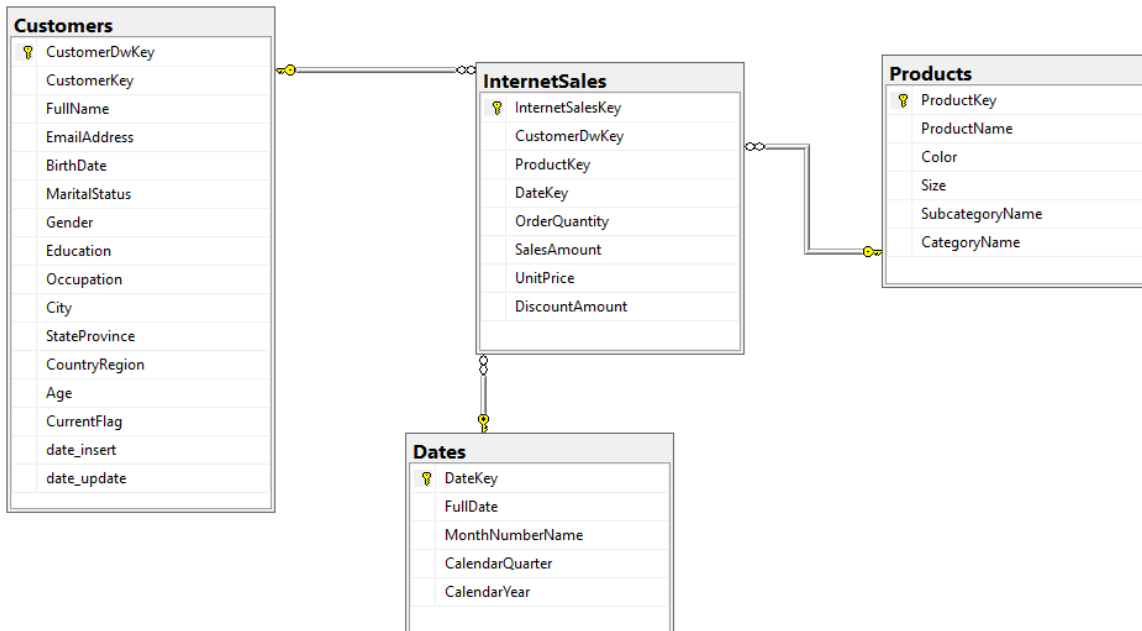


Figure 1 : schéma en étoile

### 3. Load LightAdventureWorksDW

Dans cette étape on alimente l'entrepôt de données LightAdventureWorksDW à travers la base de données transactionnelle AdventureWorks2012, en utilisant un projet SSIS (SQL Server integration services).

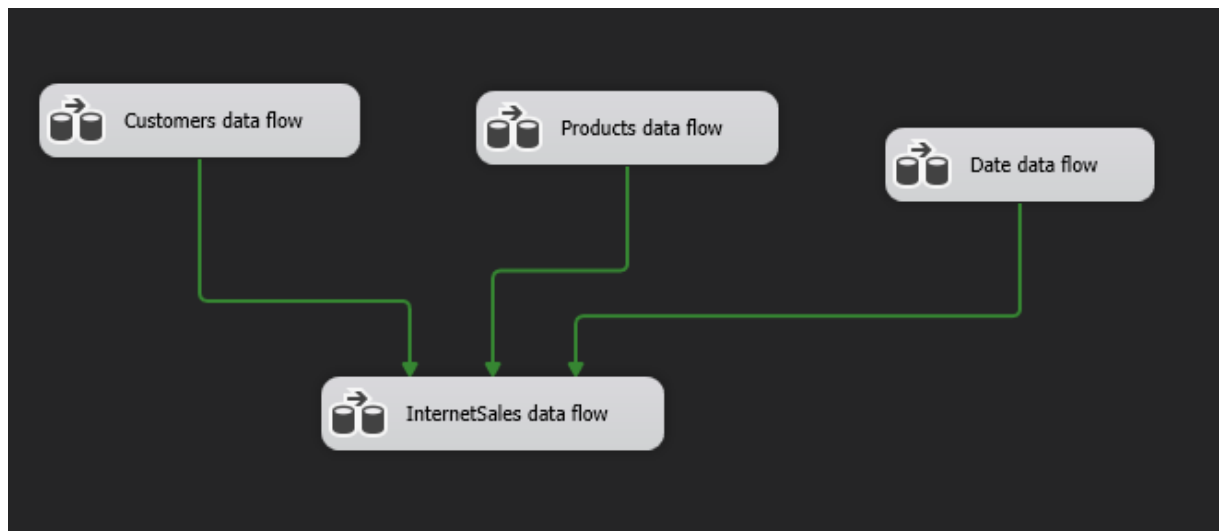


Figure 2 : flux de control global

### 3.1. Table Customer

On utilise le script ci-dessous pour extraire les données de la dimension customers :

```
--dim_customer
SELECT Sales.Customer.CustomerID, Person.Person.FirstName + ' ' + Person.Person.LastName AS Fullname,
       Person.Address.City, Person.StateProvince.Name AS StateProvince,
       Person.CountryRegion.Name AS CountryRegion
FROM   Person.StateProvince INNER JOIN
       Person.CountryRegion ON Person.StateProvince.CountryRegionCode = Person.CountryRegion.CountryRegionCode AND
       Person.StateProvince.CountryRegionCode = Person.CountryRegion.CountryRegionCode AND
       Person.StateProvince.CountryRegionCode = Person.CountryRegion.CountryRegionCode INNER JOIN
       Person.Address ON Person.StateProvince.StateProvinceID = Person.Address.StateProvinceID INNER JOIN
       Sales.SalesOrderHeader INNER JOIN
       Sales.Customer ON Sales.SalesOrderHeader.CustomerID = Sales.Customer.CustomerID AND
       Sales.SalesOrderHeader.CustomerID = Sales.Customer.CustomerID INNER JOIN
       Person.Person ON Sales.Customer.PersonID = Person.Person.BusinessEntityID ON Person.Address.AddressID = Sales.SalesOrderHeader.BillToAddressID AND
       Person.Address.AddressID = Sales.SalesOrderHeader.ShipToAddressID AND Person.Address.AddressID = Sales.SalesOrderHeader.BillToAddressID AND
       Person.Address.AddressID = Sales.SalesOrderHeader.ShipToAddressID AND Person.Address.AddressID = Sales.SalesOrderHeader.BillToAddressID AND
       Person.Address.AddressID = Sales.SalesOrderHeader.ShipToAddressID INNER JOIN
       Person.EmailAddress ON Person.Person.BusinessEntityID = Person.EmailAddress.BusinessEntityID AND
       Person.Person.BusinessEntityID = Person.EmailAddress.BusinessEntityID AND Person.Person.BusinessEntityID = Person.EmailAddress.BusinessEntityID AND
       Person.Person.BusinessEntityID = Person.EmailAddress.BusinessEntityID
GROUP BY Sales.Customer.CustomerID, Person.Person.FirstName, Person.Person.LastName, Person.Address.City, Person.StateProvince.Name, Person.CountryRegion.Name
```

Figure 3 : script loadCustomers

➤ Description du <<data flow Customers>>

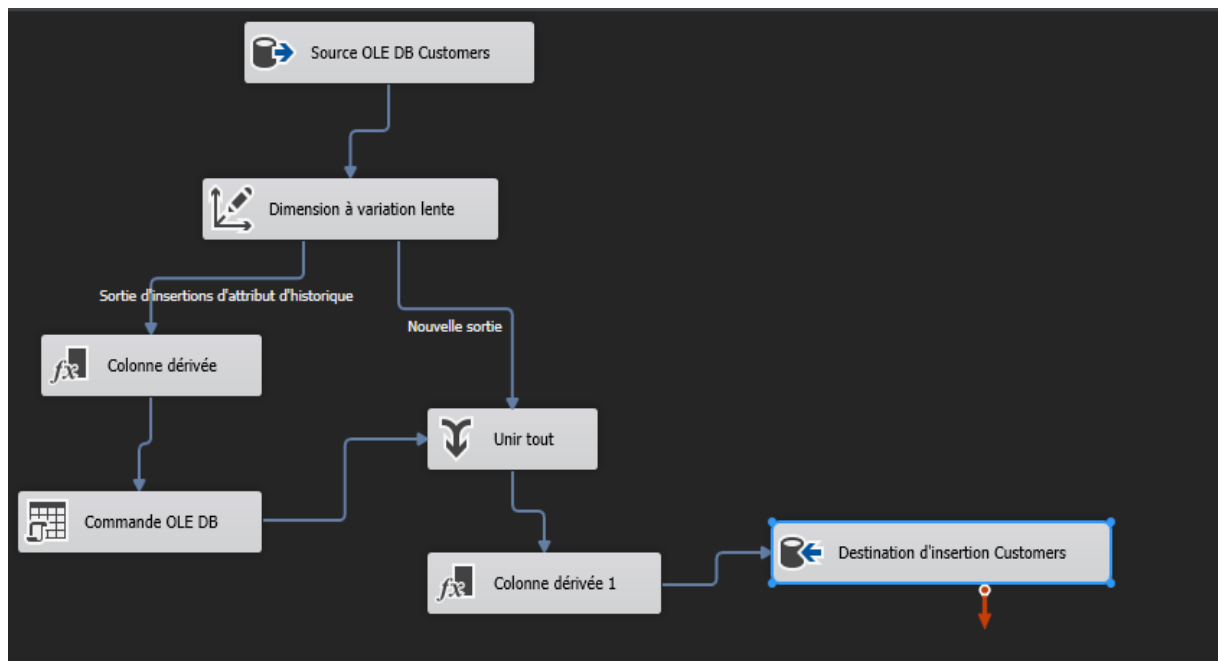


Figure 4 : Customers data flow



## ➤ Spécification des types d'attributs

Assistant Dimension à variation lente

### Colonnes de dimensions à variation lente

Gérez les modifications des données des colonnes dans les dimensions à variation lente en définissant le type de changement pour les colonnes de dimensions.

**Attribut fixe**  
Sélectionnez ce type lorsque la valeur dans une colonne ne doit pas varier. Les modifications sont traitées comme des erreurs.

**Modification d'attribut**  
Sélectionnez ce type lorsque les valeurs modifiées doivent remplacer les valeurs existantes. Il s'agit d'une

**Attribut d'historique**  
Sélectionnez ce type lorsque des modifications dans des valeurs de colonnes sont enregistrées dans de nouveaux enregistrements. Les valeurs précédentes sont enregistrées dans les enregistrements marqués comme obsolètes. Il s'agit d'une

Sélectionnez un type de modification pour les colonnes de dimension à variation lente :

Colonnes de dimension	Modifier le type
City	Attribut d'his...
FullName	Attribut fixe
StateProvince	Attribut d'his...
CountryRegion	Attribut fixe

Supprimer

Aide < Précédent Suivant > Terminer >> Annuler

Figure 5 : Colonnes de dimensions à variation lente

- City : si le client change sa ville, on garde obligatoirement l'historique des ventes de l'ancienne ville.
- FullName : le nom d'un client doit être fixe.
- StateProvince : on doit historiser toutes les données en cas de changement concernant la province du client.

## ➤ Insertion de données

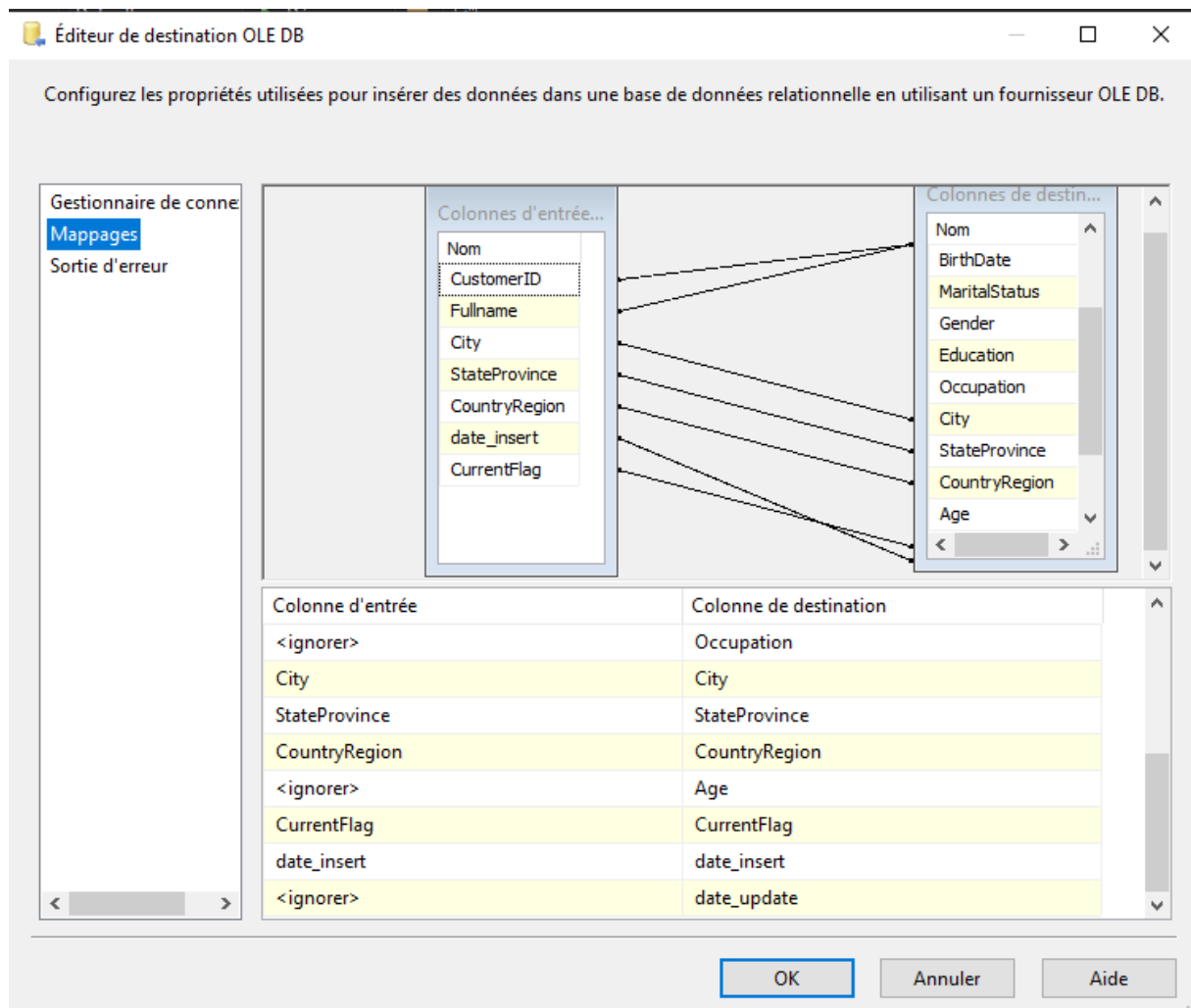


Figure 6 : Customers : Data Mapping

## 3.2. Table Produit

Extraction des données de la base de données adventureworks2012.

```
-- Dim_produit
SELECT distinct Production.Product.ProductID, Production.Product.Name,
                Production.Product.Size, Production.Product.Color,
                Production.ProductSubcategory.Name AS SubcategoryName,
                Production.ProductCategory.Name AS CategoryName
FROM      Production.Product INNER JOIN
          Production.ProductSubcategory ON Production.Product.ProductSubcategoryID = Production.ProductSubcategory.ProductSubcategoryID INNER JOIN
          Production.ProductCategory ON Production.ProductSubcategory.ProductCategoryID = Production.ProductCategory.ProductCategoryID
```

Figure 7 : Script Load Products

## ➤ Description du <<Products Data Flow>>

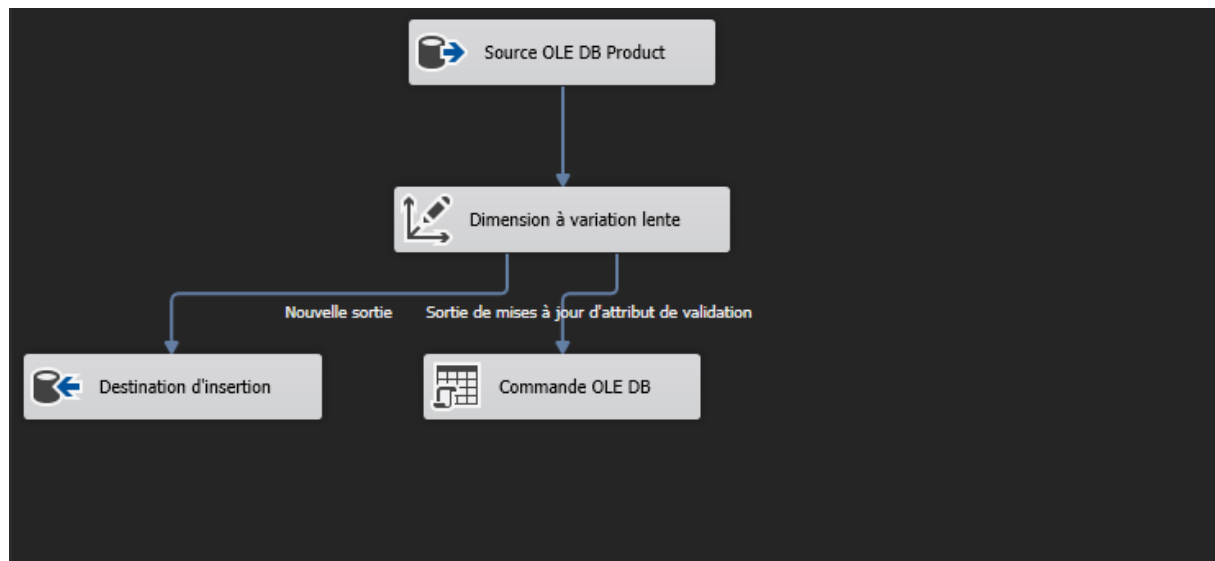


Figure 8 : Products Data Flow

## ➤ Type des attributs

Assistant Dimension à variation lente

### Colonnes de dimensions à variation lente

Gérez les modifications des données des colonnes dans les dimensions à variation lente en définissant le type de changement pour les colonnes de dimensions.

**Attribut fixe**  
Sélectionnez ce type lorsque la valeur dans une colonne ne doit pas varier. Les modifications sont traitées comme des erreurs.

**Modification d'attribut**  
Sélectionnez ce type lorsque les valeurs modifiées doivent remplacer les valeurs existantes. Il s'agit d'une

**Attribut d'historique**  
Sélectionnez ce type lorsque des modifications dans des valeurs de colonnes sont enregistrées dans de nouveaux enregistrements. Les valeurs précédentes sont enregistrées dans les enregistrements marqués comme obsolètes. Il s'agit d'une

Sélectionnez un type de modification pour les colonnes de dimension à variation lente :

Colonnes de dimension	Modifier le type
CategoryName	Attribut fixe
Color	Modification ...
ProductName	Modification ...
Size	Modification ...
SubcategoryName	Attribut fixe

Supprimer

Aide < Précédent Suivant > Terminer >> Annuler

Figure 9 : Colonnes de dimensions à variation lente

- CategoryName et subCategoryName : fixes parce que la catégorie du produit est fixe.
- Color, size, productName : sont des attributs modifiables.

### 3.3. Table Date

Extraction des données de la base de données adventureworks2012.

```
-- dim_date
use Adventureworks2012
SELECT distinct CAST(CONVERT(CHAR(8),CAST(OrderDate as DATETIME),112) as INT) as datekey,
                OrderDate as fulldate,DATENAME(MONTH,OrderDate) as MonthNumberName ,YEAR(OrderDate) as year,
                DATENAME(QUARTER,CAST( OrderDate as datetime)) as quarter
FROM      Sales.SalesOrderHeader
```

Figure 10 : Script LoadDates

#### ➤ Description du <<Dates Data Flow>>

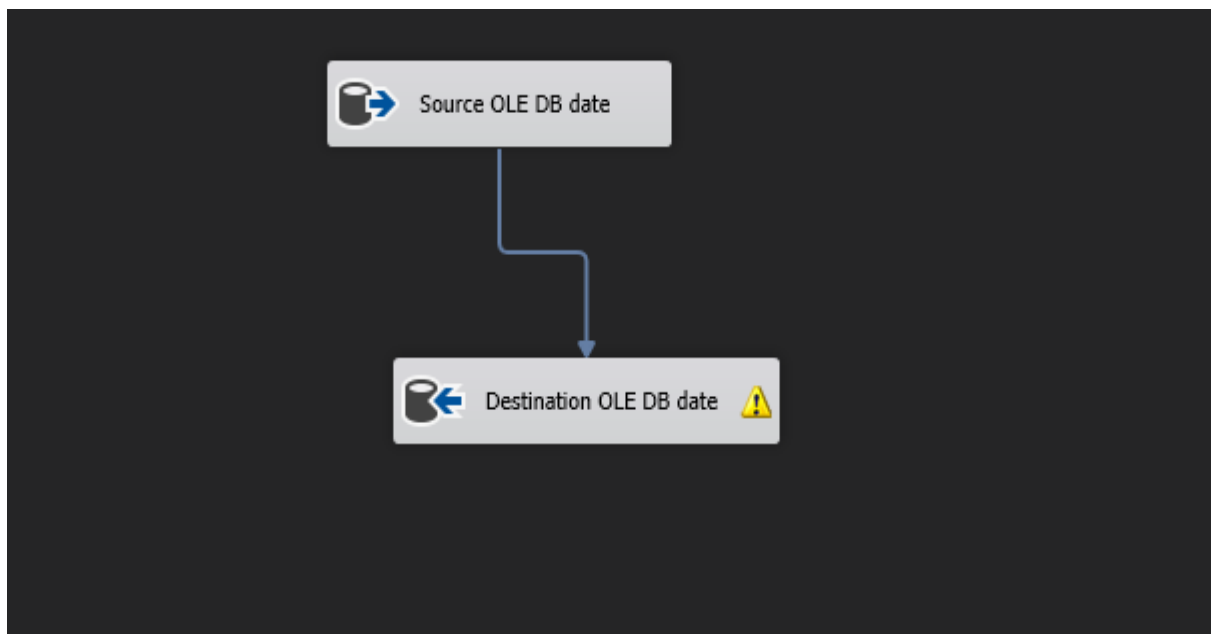


Figure 11 : Dates Data Flow

### 3.4. Table de fait : Internet Sales

Extraction des données de la base de données adventureworks2012.

```
--table de fait
SELECT distinct SD.ProductID,
    C.CustomerDwKey as CustomerID ,
    CAST(CONVERT(CHAR(8),CAST(SH.OrderDate as DATETIME),112) as INT) as Datekey,
    SD.OrderQty as orderQuantity,
    SD.UnitPrice,
    convert( float, SD.UnitPriceDiscount*convert(float,SD.OrderQty)) as DiscountAmount,
    SD.UnitPrice*SD.OrderQty as SalesAmount
FROM AdventureWorks2012.Sales.SalesOrderDetail as SD,
    AdventureWorks2012.Sales.SalesOrderHeader as SH,
    LIGHTADVENTUREWORKSDW.dbo.Customers as C,
    LIGHTADVENTUREWORKSDW.dbo.Products as P
where SD.SalesOrderDetailID = SH.SalesOrderID
and C.CustomerKey = SH.CustomerID
and P.ProductKey = SD.ProductID;
```

Figure 12 : Script Load InternetSales

#### ➤ Type des attributs

**Colonnes de dimensions à variation lente**  
Gérez les modifications des données des colonnes dans les dimensions à variation lente en définissant le type de changement pour les colonnes de dimensions.

**Attribut fixe**  
Sélectionnez ce type lorsque la valeur dans une colonne ne doit pas varier. Les modifications sont traitées comme des erreurs.

**Modification d'attribut**  
Sélectionnez ce type lorsque les valeurs modifiées doivent remplacer les valeurs existantes. Il s'agit d'une

**Attribut d'historique**  
Sélectionnez ce type lorsque des modifications dans des valeurs de colonnes sont enregistrées dans de nouveaux enregistrements. Les valeurs précédentes sont enregistrées dans les enregistrements marqués comme obsolètes. Il s'agit d'une

Sélectionnez un type de modification pour les colonnes de dimension à variation lente :

Colonnes de dimension	Modifier le type
DiscountAmount	Attribut fixe
OrderQuantity	Attribut fixe
SalesAmount	Attribut fixe
UnitPrice	Modification ...

Supprimer

Aide < Précédent Suivant > Terminer >> Annuler

Figure 13 : Colonnes de dimensions à variation lente

- UnitPrice : attribut modifiable.

Ensuite, tous les données se chargent dans leur destination avec succès.

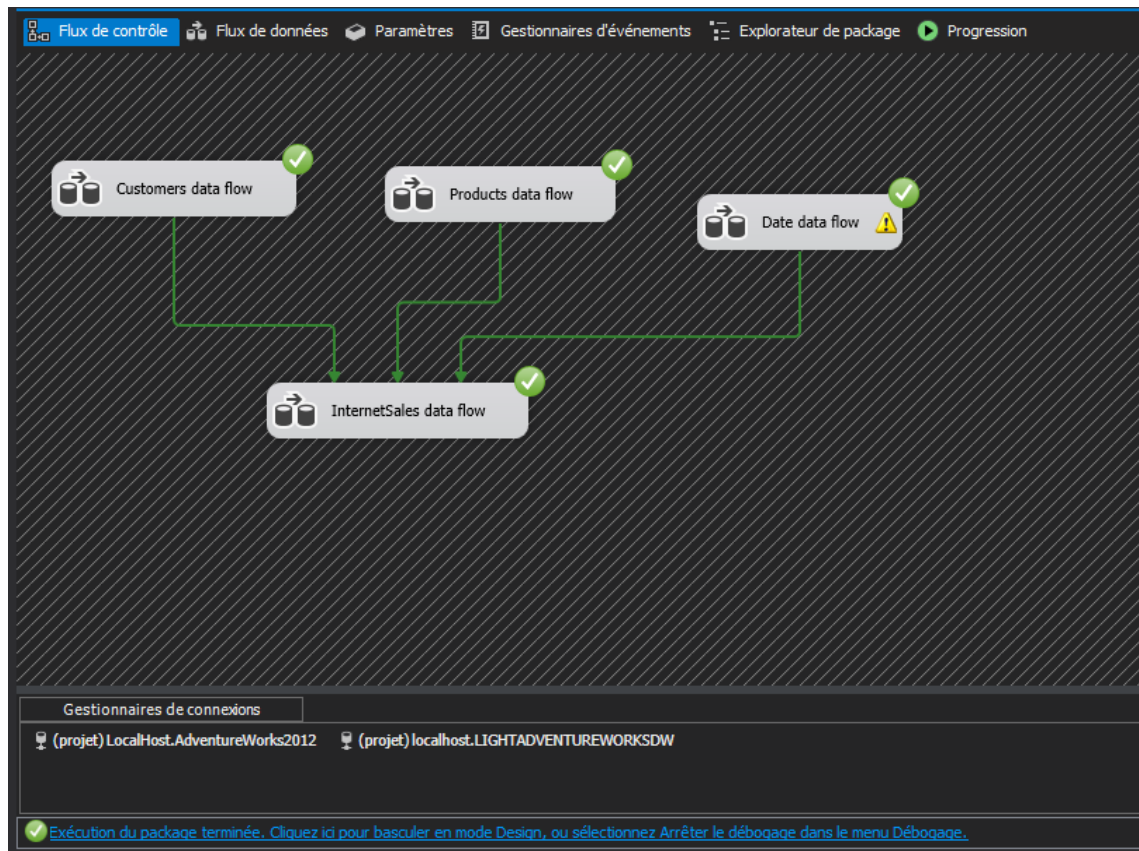


Figure 14 : succès de l'alimentation

## 4. Création du cube

Cette étape consiste à décrire les différentes dimensions et les hiérarchies déployées.

### 4.1. Hiérarchies

On choisit un ensemble de colonnes pour créer nos hiérarchies, en suite on crée des relations rigides entre les attributs étant un best practice dans la création des hiérarchies.

## Dim Customer

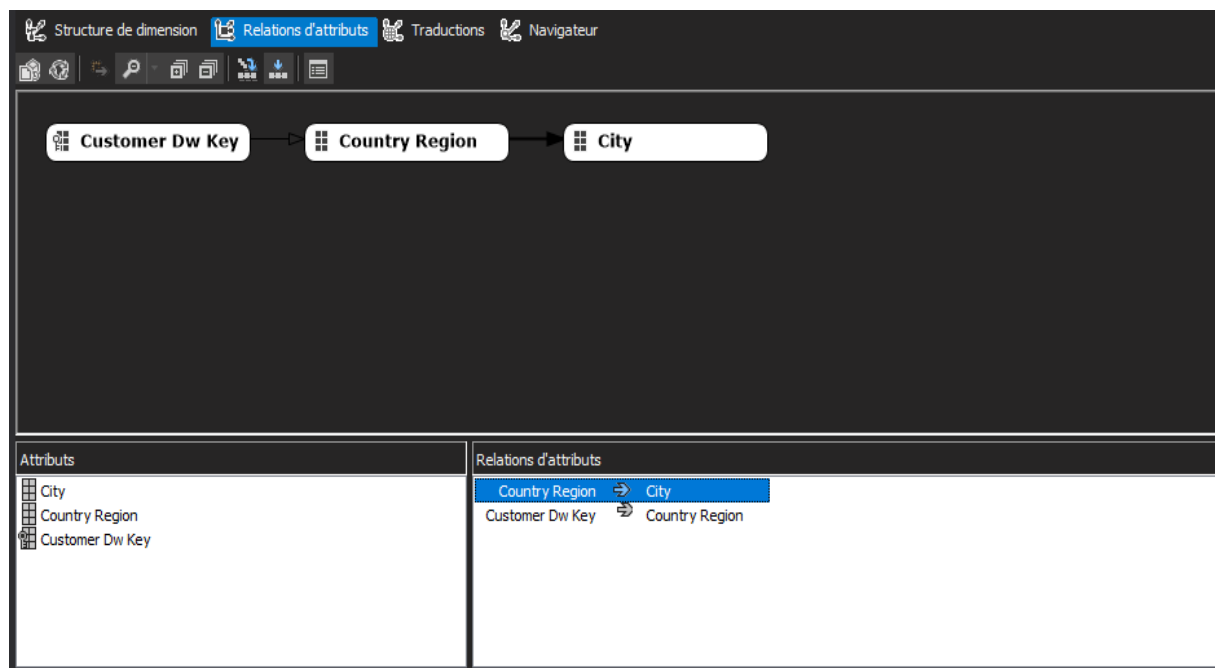


Figure 15 : Customers Hierarchy

## Dim Produit

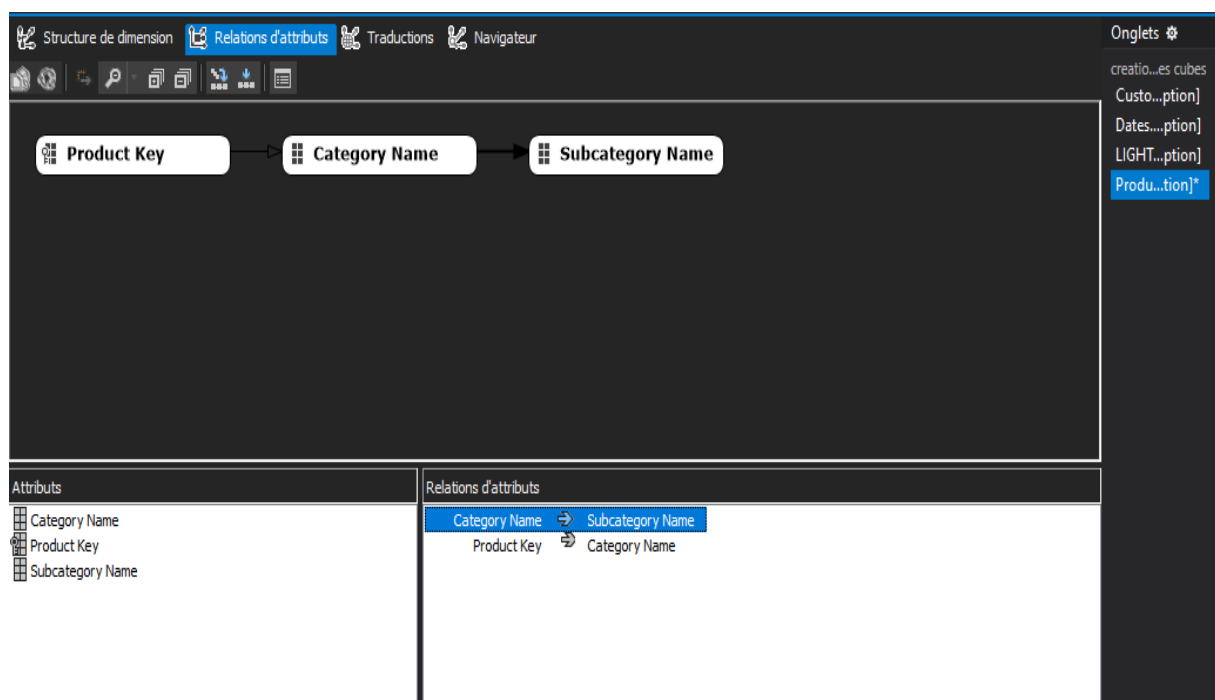
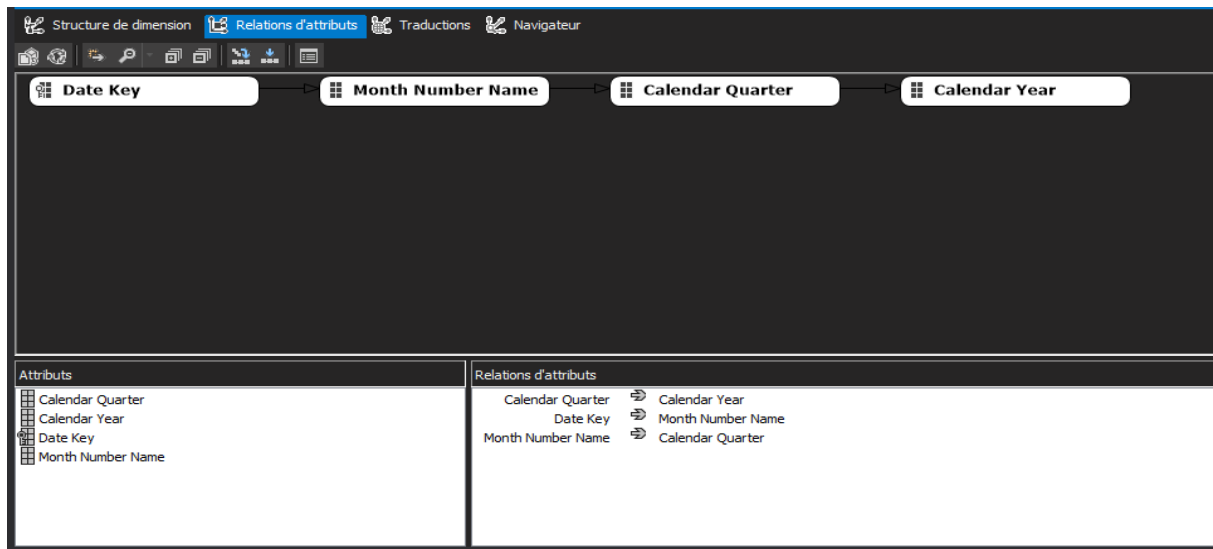


Figure 16 : Products Hierarchy

## Dim Date



### 4.2. Le cube

On arrive alors à la création de notre cube à travers les trois dimensions :

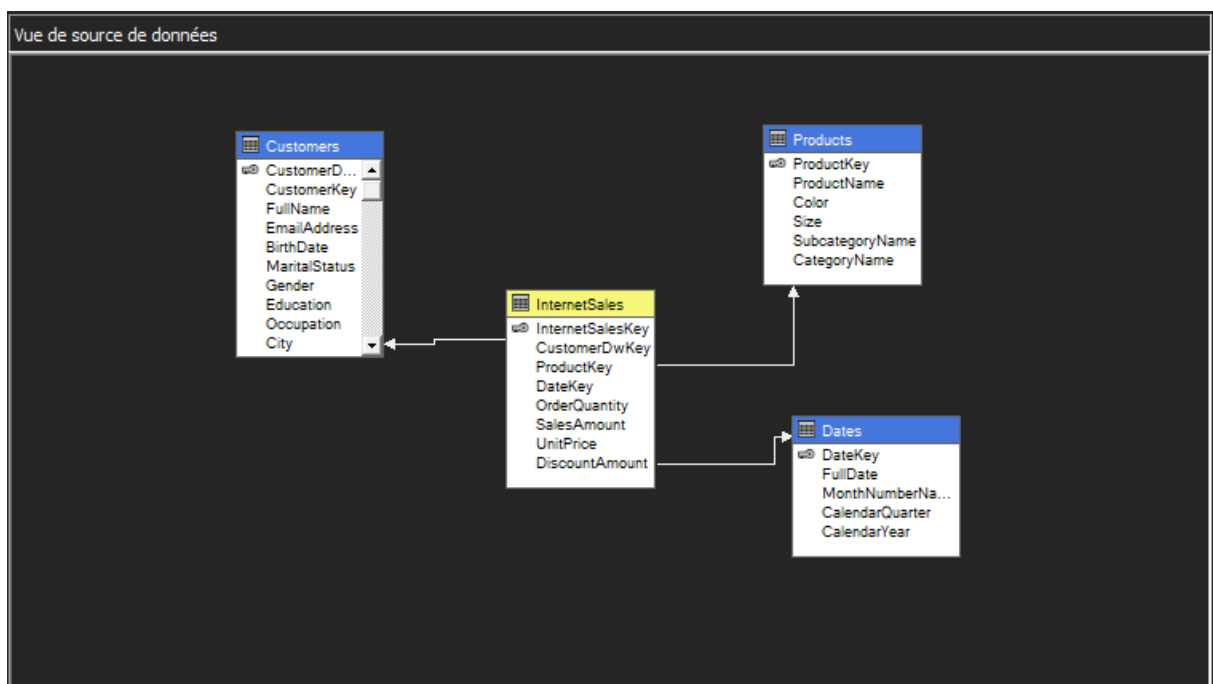


Figure 17 : Création du cube

On arrive finalement à visualiser nos données :



Calendar Year	Country Region	City	Category Name	Subcategory Name	Sales Amount
2005	Australia	Be...	Accessories	Bottles and Cages	9,99
2005	Australia	Be...	Accessories	Cleaners	7,95
2005	Australia	Be...	Accessories	Helmets	69,98
2005	Australia	Be...	Accessories	Tires and Tubes	50,24
2005	Australia	Be...	Bikes	Mountain Bikes	2294,99
2005	Australia	Be...	Clothing	Caps	8,99
2005	Australia	Bri...	Accessories	Bottles and Cages	4,99
2005	Australia	Bri...	Accessories	Fenders	21,98
2005	Australia	Bri...	Accessories	Helmets	69,98
2005	Australia	Bri...	Accessories	Tires and Tubes	66,25
2005	Australia	Bri...	Bikes	Mountain Bikes	2319,99
2005	Australia	Cal...	Accessories	Bottles and Cages	14,98
2005	Australia	Cal...	Accessories	Fenders	21,98
2005	Australia	Cal...	Accessories	Helmets	34,99
2005	Australia	Cal...	Accessories	Tires and Tubes	33,98
2005	Australia	Cal...	Bikes	Mountain Bikes	2319,99
2005	Australia	Cal...	Clothing	Caps	8,99
2005	Australia	Cal...	Clothing	Gloves	24,49
2005	Australia	Cal...	Clothing	Shorts	69,99
2005	Australia	Clo...	Accessories	Bottles and Cages	4,99

Figure 18 : visualisation des données

## 5. Génération du rapport

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

### 5.1. SSRS

Voilà un ensemble de rapport d'analyse qu'on a arrivé à générer :

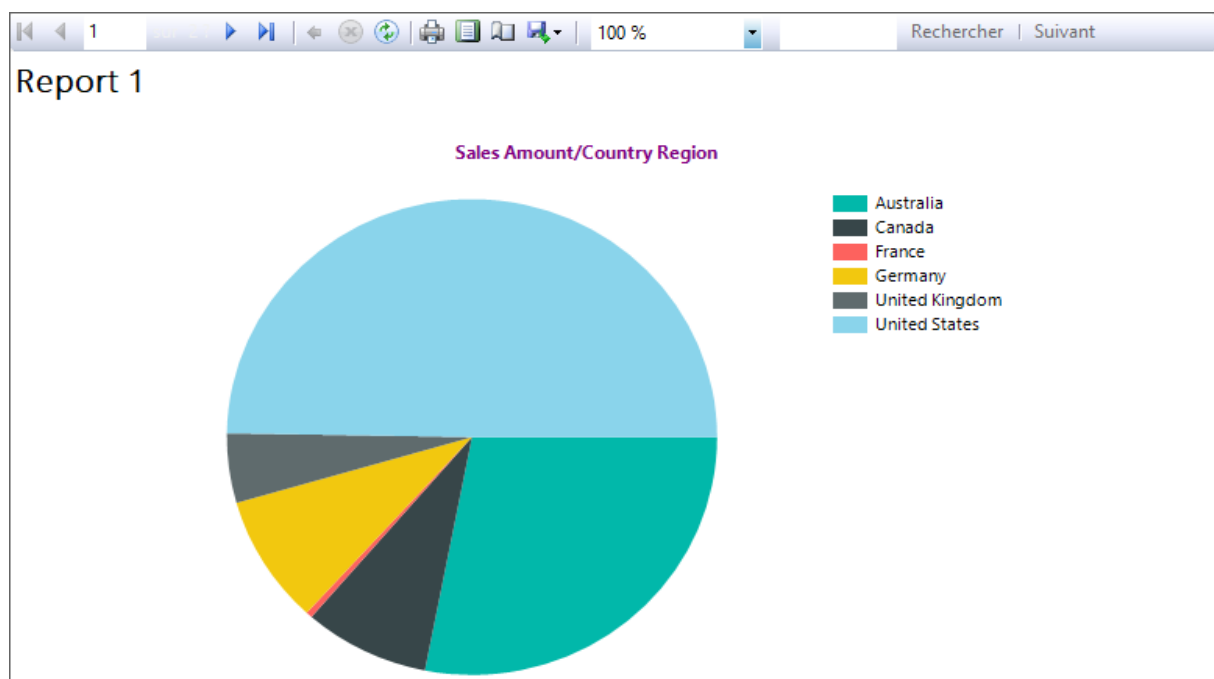


Figure 19 : SalesAmount/CountryRegion



Figure 21 : SalesAmount/ year

Report3

Année : 2005

		Australia	Canada	France	Germany	United Kingdom	United States
Accessories	Bike Racks	360			120		240
	Bike Stands	318	159		159		636
	Bottles and Cages	378,44	182,76	58,91	106,86	92,88	627,1
	Cleaners	31,8	15,9	15,9		23,85	71,55
	Fenders	351,68	263,76	21,98	21,98	43,96	527,52
	Helmets	1504,57	594,83	209,94	209,94	349,9	2344,33
	Hydration Packs	494,91			54,99	109,98	219,96
Bikes	Tires and Tubes	2115,37	548,47	377,81	362,35	502,26	3123,77
	Mountain Bikes	32703,83	9628,93		9794,95	2294,99	51761,69
	Road Bikes	31184,06	6304,3		7087,17	7262,93	57953,28
Clothing	Touring Bikes	10369,89	5982,99		7197,84	2384,07	23572,29
	Caps	179,8	44,95	62,93	17,98	26,97	197,78
	Gloves	195,92	122,45		48,98	48,98	416,33
	Jerseys	827,84	261,95	311,94	157,97	527,9	1805,65
	Shorts	699,9		139,98	139,98		1119,84
	Socks	44,95	17,98	17,98		35,96	89,9
	Vests	127			63,5	63,5	317,5

Figure 20 : Rapport détaillé de chaque année

## 5.2. Excel

On utilise le pivotTable tools à travers excel, et on arrive à visualiser les données comme suit et générer des différents graphs :

	A	B	C	D	E	F
1						
2						
3	Sales Amount	Column Labels				
4	Row Labels	2005	2006	2007	2008	Grand Total
5	Australia	81887,96	599469,3843	1768881,417	1794069,181	4244307,942
6	Accessories	5554,77	13315,8243	39684,3096	35512,8698	94067,7737
7	Bike Racks	360	5151,6	12328,8	5838	23678,4
8	Bike Stands	318	318	2385	2226	5247
9	Bottles and Cages	378,44	483,8488	1805,274	2226,5544	4894,1172
10	Cleaners	31,8	206,064	615,7275	706,914	1560,5055
11	Fenders	351,68	439,6	1538,6	1626,52	3956,4
12	Helmets	1504,57	3017,8875	10621,5644	11655,8688	26799,8907
13	Hydration Packs	494,91	1616,706	3048,0957	3219,1146	8378,8263
14	Tires and Tubes	2115,37	2082,118	7341,248	8013,898	19552,634
15	Bikes	74257,78	443166,7517	1494178,248	1478831,84	3490434,619
16	Clothing	2075,41	18322,0819	41747,0586	41497,8697	103642,4202
17	Components		124664,7264	193271,8008	238226,6016	556163,1288
18	Canada	24128,27	344312,9999	856584,0331	1165856,355	2390881,658
19	Accessories	1764,72	5550,4489	23780,6576	24916,271	56012,0975
20	Bike Racks		1008	6784,8	6168	13960,8
21	Bike Stands	159	159		636	954
22	Bottles and Cages	182,76	269,8295	1303,7804	1412,6704	3169,0403
23	Cleaners	15,9	139,443	344,712	352,98	853,035
24	Fenders	263,76	131,88	967,12	1077,02	2439,78
25	Helmets	594,83	2063,0104	6877,9843	7199,5424	16735,3671

Figure 22 : Reporting with excel

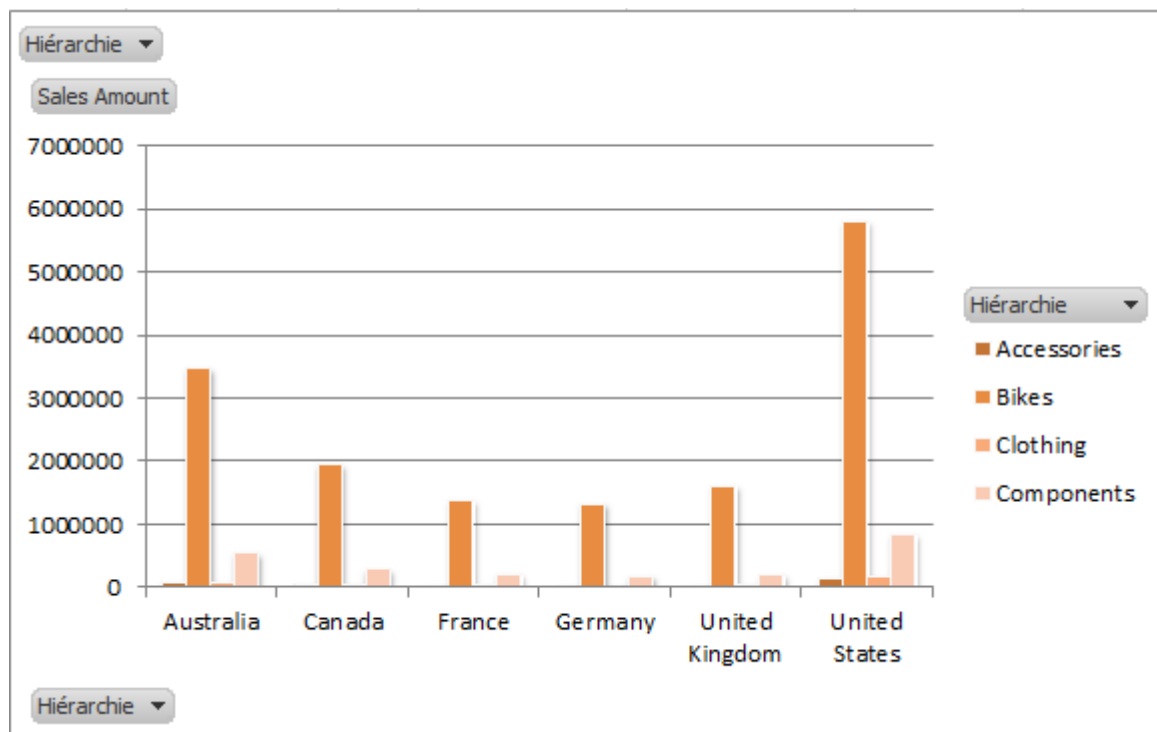


Figure 23 : Exemple de graph avec Excel

## **6. Conclusion**

L'utilisation d'un entrepôt de données présente un grand nombre d'avantages. Un Data Warehouse est un entrepôt informatisé dans lequel l'information est stockée. L'organisation qui possède cette information peut l'analyser afin de trouver des modèles historiques ou des tendances qui lui permettront de prendre des décisions. En effet il permet une meilleure analyse de la situation, une meilleure accessibilité aux données, et l'enregistrement des modifications pour construire l'historique. D'un point de vue technique, il sert surtout à 'délester' les bases de données opérationnelles des requêtes pouvant nuire à leurs performances.