



# Business Intelligence

AMOA 2

## Rapport Projet NorthWind

Réalisé par  
ZAINI & JERRARI & MOUTALATTIF

Encadré par  
Pr. ELAKKAOUI Zineb

2022

# Sommaire

## 01 Introduction

## 02 Environnement de travail

- Microsoft SQL Server (Developer Edition)
- SQL Server Management Studio
- Microsoft Visual Studio
- SQL Server Integration Services (SSIS)

## 03 Chargement des tables

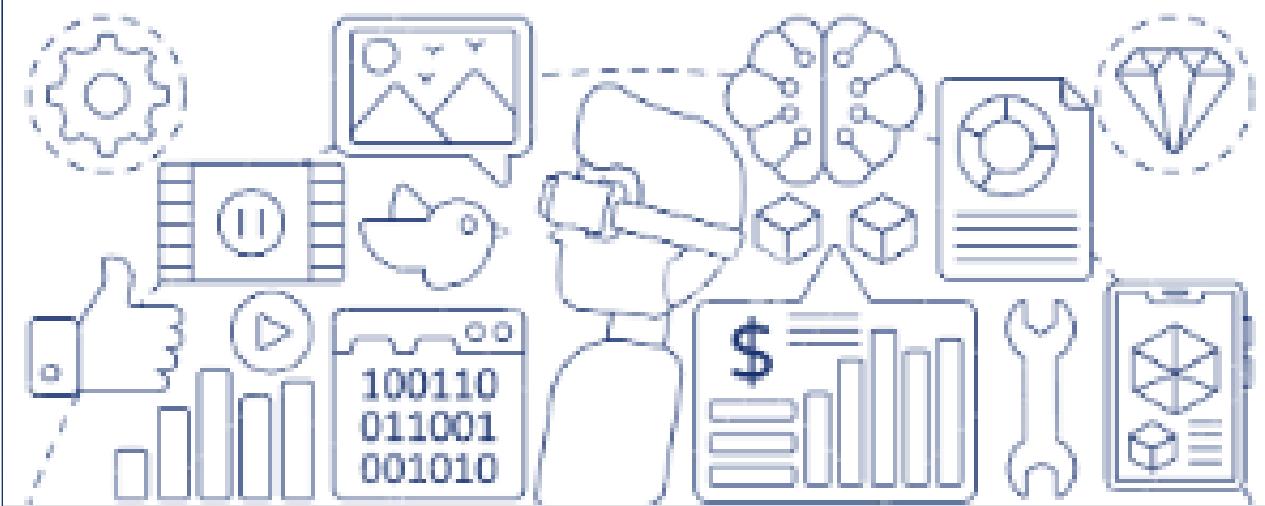
- DimShipper
- DimCategory
- DimEmployees
- DimProduct
- DimArea
- DimState
- DimCountry
- DimGeography
- DimSupplier

## 04 Conclusion

01

# Introduction

- Informatique décisionnelle
  - Processus ETL



# Business Intelligence

La Business Intelligence (BI) est un processus technologique d'analyse des données et de présentation d'informations pour aider les dirigeants, managers et autres utilisateurs finaux de l'entreprise à prendre des décisions business éclairées. La Business Intelligence englobe une grande variété d'outils, d'applications et de méthodologies qui permettent aux organisations de collecter des données à partir de systèmes internes et de sources externes. Ces données sont ensuite préparées pour l'analyse afin de créer des rapports, tableaux de bord et autres outils de Data Viz pour rendre les résultats analytiques disponibles aux décideurs et aux opérations.

Aujourd'hui, les entreprises s'appuient sur les logiciels de Business Intelligence pour identifier et extraire des informations précieuses des grands volumes de données qu'elles stockent. Ces outils permettent d'en tirer des informations tels que des veilles concurrentielles et les tendances du marché, ainsi que des informations internes tel que trouver les raisons des opportunités perdues.

Au cours des dernières années, la Business Intelligence passait d'un domaine qui appartenait principalement à l'IT à une fonction devenue clé dans toutes entreprises, pour détecter des opportunités business, optimiser ses campagnes marketing ou financières, par exemple. Aujourd'hui, cependant, il existe toute une gamme de solutions de Business Intelligence intuitives et accessibles qui permettent aux utilisateurs de toute l'entreprise de se connecter et d'accéder aux tableaux de bord BI et aux analyses qui transforment les données brutes en véritables sources d'informations commerciales.

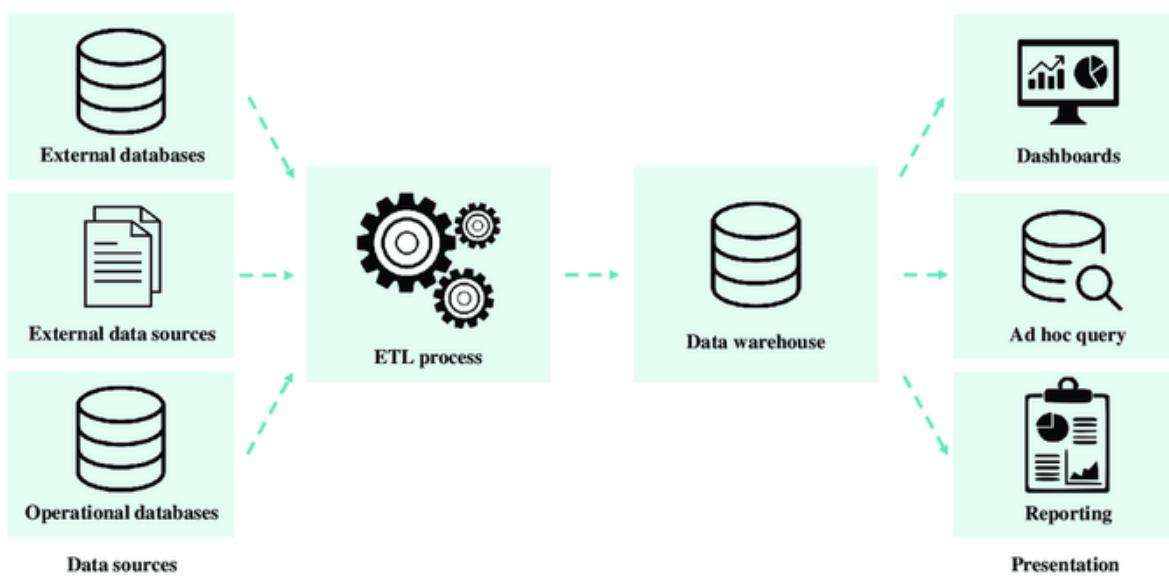
## Le fonctionnement de la business intelligence

Les sources de données utilisées pour enrichir son outil de BI comprennent les systèmes de gestion de la relation client (CRM), les informations sur la chaîne logistique, les tableaux de bord des performances commerciales, les analyses marketing, les données d'appel des call center. Les applications de Business Intelligence aident les entreprises à regrouper toutes ces sources disparates en une seule vue unifiée fournissant des rapports, des tableaux de bord et des analyses en temps réel.

Bien que la business intelligence ne donne pas de directives aux utilisateurs sur les décisions à prendre, ou ce qu'ils doivent faire ou ce qu'il se passera s'ils suivent un certain cours, la Business Intelligence ne se limite pas non plus au reporting. Au contraire, la BI offre aux employés un moyen d'examiner les données de l'entreprise et de comprendre les tendances du marché et d'en tirer des "learnings" pour faire avancer toutes les verticales métiers de leur entreprise.

Les outils d'intelligence d'affaires permettent de factoriser les efforts que les employés doivent déployer pour rechercher, fusionner et interroger les données afin d'obtenir l'information dont ils ont besoin pour prendre de bonnes décisions business.

Les cas d'utilisation potentiels de la Business Intelligence vont au-delà des indicateurs de performance d'entreprise typiques de l'amélioration des ventes et de la réduction des coûts



# Processus ETL

**Extraction, transformation, chargement (ETL)**, un processus automatisé qui prend les données brutes, extrait l'information nécessaire à l'analyse, la transforme en un format qui peut répondre aux besoins opérationnels et la charge dans un Data Warehouse. L'ETL résume généralement les données afin de réduire leur taille et d'améliorer leur performance pour des types d'analyse spécifiques.

Le processus d'extraction des données des systèmes sources et de leur transfert dans l'entrepôt de données est communément appelé **ETL**, qui signifie **extraction, transformation et chargement**. Il est à noter que l'ETL fait référence à un processus général et non à trois étapes bien définies. L'acronyme ETL est peut-être trop simpliste, car il omet la phase de transport et implique que chacune des autres phases du processus est distincte.

## Les tâches suivantes sont les principales actions du processus Extraction des données

La première étape de l'**ETL** est l'extraction. Au cours de l'extraction, les données sont spécifiquement identifiées et ensuite prélevées à de nombreux endroits différent. Ces données peuvent provenir d'une variété de choses, comme des fichiers, des feuilles de calcul, des systèmes de bases de données et des applications, etc. Il n'est généralement pas possible d'identifier le sous-ensemble exact d'intérêt, de sorte que l'on extrait plus de données que nécessaire pour s'assurer qu'il couvre tous les besoins.

Selon les capacités du système source (par exemple, les ressources du système d'exploitation), certaines transformations peuvent avoir lieu pendant ce processus d'extraction. La taille des données extraites varie de quelques centaines de kilo-octets à plusieurs giga-octets, selon le système source et la situation commerciale. C'est également le cas pour la période entre deux extractions ; certaines peuvent varier de jours ou d'heures à presque en temps réel.

## Transport des données

Une fois les données extraites, elles doivent être physiquement transportées vers le système cible ou vers un système intermédiaire pour traitement ultérieur. Selon le mode de transport choisi, certaines transformations peuvent également être effectuées au cours de ce processus.

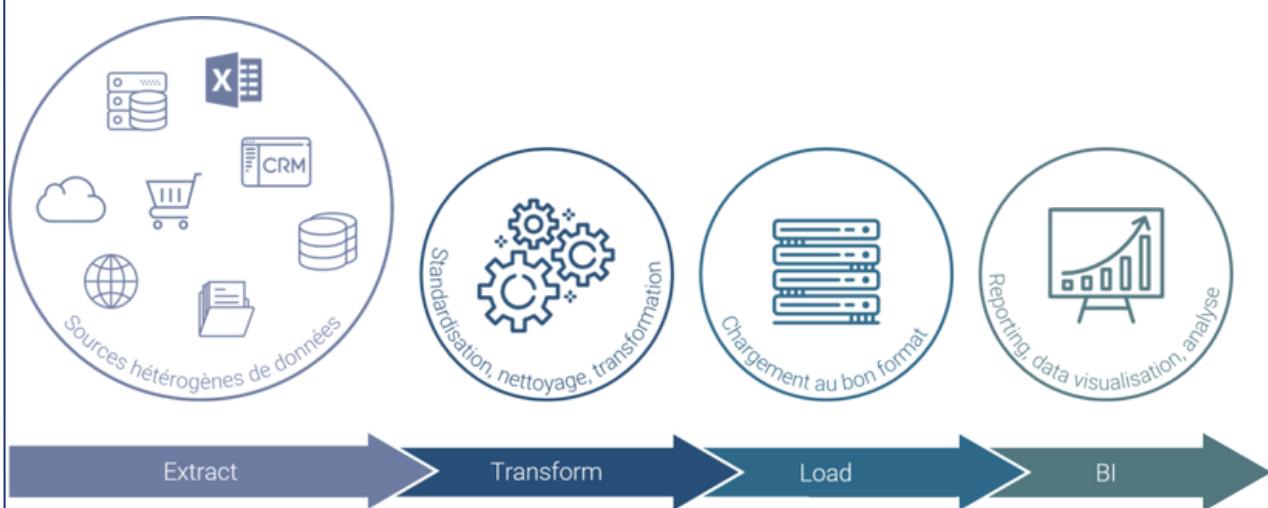
## Transformation des données

La prochaine étape du processus ETL est la transformation. Une fois les données extraites, elles doivent être physiquement transportées vers la destination cible et converties dans le format approprié. Cette transformation de données peut inclure des opérations telles que le nettoyage, l'assemblage et la validation des données.

## Chargement des données

La dernière étape du processus ETL consiste à charger les données transformées dans la cible de destination. Cette cible peut être une bases de données ou un Data Warehouse. Il existe deux méthodes principales pour charger les données dans un entrepôt : chargement complet et chargement incrémentale. La méthode du chargement complet implique un déchargement complet des données qui a lieu la première fois que la source est chargée dans l'entrepôt. La charge incrémentale, par contre, a lieu à intervalles réguliers. Ces intervalles peuvent être des incréments de flux (meilleurs pour de plus petits volumes de données) ou des incréments de lots (meilleurs pour de plus grands volumes de données).

Les équipes de **Business Intelligence (BI)** lancent ensuite des requêtes sur ces données, qui sont ensuite présentées aux utilisateurs finaux ou aux personnes chargées de prendre des décisions commerciales, ou utilisées comme entrées pour des algorithmes de Machine Learning. Un problème courant rencontré ici est que si les résumés OLAP ne peuvent pas supporter le type d'analyse que l'équipe BI veut faire, alors tout le processus doit être relancé, cette fois avec différentes transformations.



# 02

## Environnement de travail

- Microsoft SQL Server (Developer Edition)
- SQL Server Management Studio
- Microsoft Visual Studio
- SQL Server Integration Services



# Microsoft SQL Server

Microsoft SQL Server est un système de gestion de base de données (SGBD) en langage SQL incorporant entre autres un SGBDR (SGBD relationnel »).

Pour notre projet , on utilise l'édition SQL Server Developer qui permet aux développeurs de créer des applications basées sur SQL Server. Il inclut toutes les fonctionnalités de l'édition Enterprise, mais sa licence permet uniquement de l'utiliser comme un système de développement et de test, et non comme un serveur de production. L'édition SQL Server Developer est la solution idéale pour le développement et le test d'applications.



# SQL Server Management Studio

SQL Server Management Studio (SSMS) est un environnement intégré pour la gestion de toute infrastructure SQL, de [SQL Server](#) à Azure SQL Database. SSMS fournit des outils pour configurer, surveiller et administrer des instances de SQL Server et des bases de données. On Utilise SSMS pour déployer, surveiller et mettre à niveau les composants de la couche données utilisés par vos applications, et créer des requêtes et des scripts.



# Microsoft Visual Studio

## SQL Server Data Tools (SSDT)

SQL Server Data Tools (SSDT) est un outil de développement moderne permettant de créer des bases de données relationnelles SQL Server, des bases de données dans Azure SQL, des modèles de données Analysis Services (AS), des packages Integration Services (IS) et des rapports Reporting Services (RS). Avec SSDT, vous pouvez concevoir et déployer n'importe quel type de contenu SQL Server avec la même facilité que vous développeriez une application dans Visual Studio.

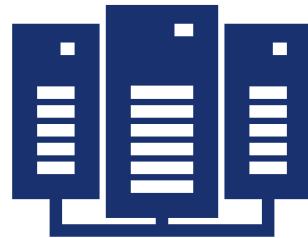


# SQL Server

## Integration Services (SSIS)

SQL Server Integration Services est une plate-forme permettant de créer des solutions d'intégration et de transformation de données au niveau de l'entreprise. On Utilise les Services d'Intégration pour résoudre des problèmes commerciaux complexes en copiant ou en téléchargeant des fichiers, en chargeant des entrepôts de données, en nettoyant et en explorant des données et en gérant des objets et des données SQL Server.



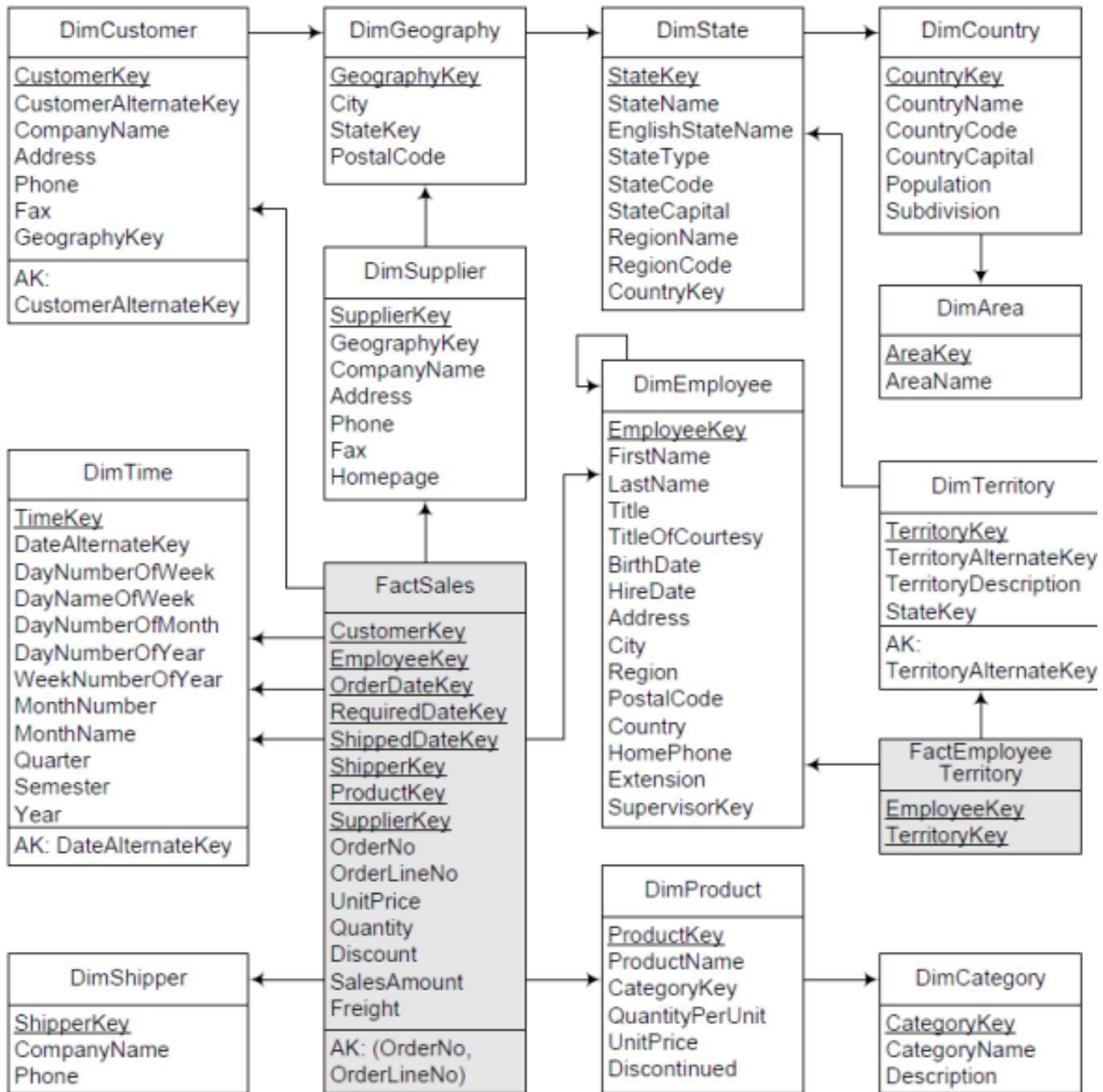


# 03

## Chargement des tables

- DimShipper
- DimCategory
- DimEmployees
- DimProduct
- DimArea
- DimCountry
- DimState
- DimGeography
- DimSupplier

Après la création des deux bases de données Northwind et NorthwindDW en exécutant les scripts sql déjà prêts. On passe au chargement de datawarehouse NorthwindDW en se baser sur le schema suivant:

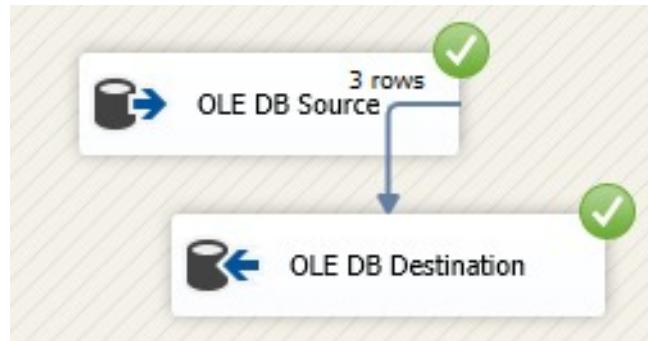


On commence par les tables qui se trouve à l'extrême car il sont plus simples à télécharger puisque qu'elles ne dépendent d'aucune autres tables , puis en passe aux autres tables en se rapprochant de la tables des fait(pincipe de granularité)

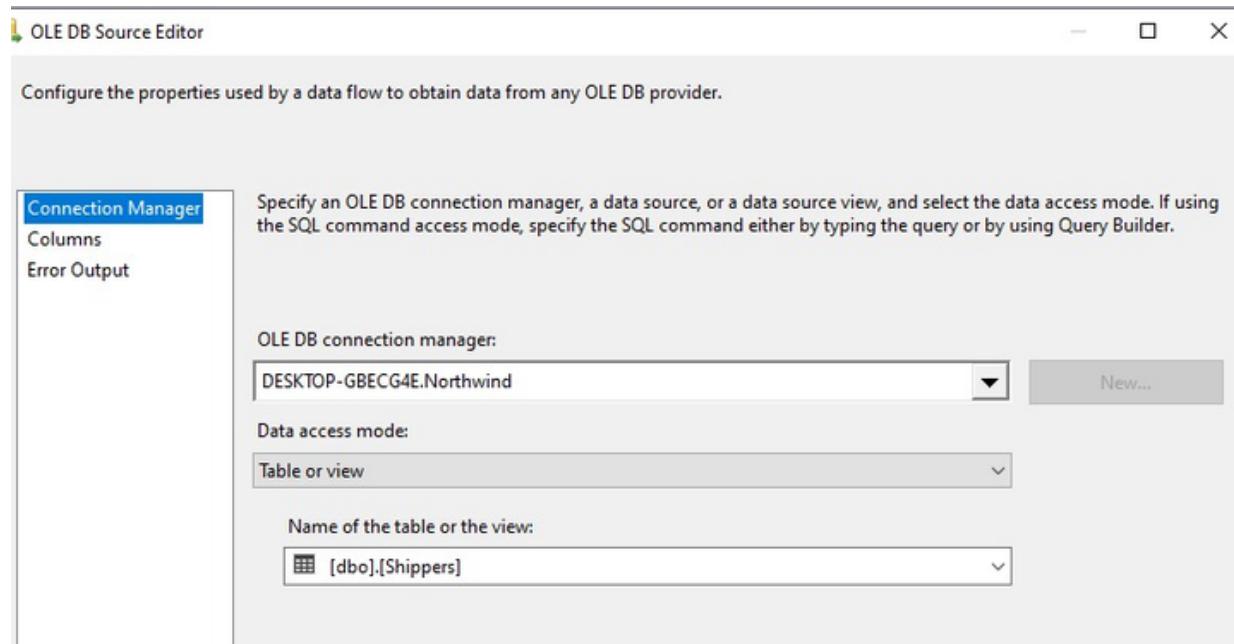
# DimShipper

On commence le chargement des données à partir du tableau "dimshipper" parce qu'il ne dépend d'aucune autre table.

On met en flux de données une OLE DB source et une destination:



Pour la source on prend le tableau "shippers" de la base "Northwind" qui contient les colonnes nécessaires pour remplir la destination:



# DimShipper

Pour la destination on prend le tableau "DimShipper" de la base "NorthwindDW" afin de le remplir par les données chargées à partir de "Shipper". Et on lie les colonnes de la source et la destination:

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

OLE DB connection manager:

DESKTOP-GBECG4E.NorthwindDW

Data access mode:

Table or view - fast load

Name of the table or the view:

[dbo].[DimShipper]

Keep identity  Table lock

Keep nulls  Check constraints

Rows per batch:

Maximum insert commit size: 2147483647

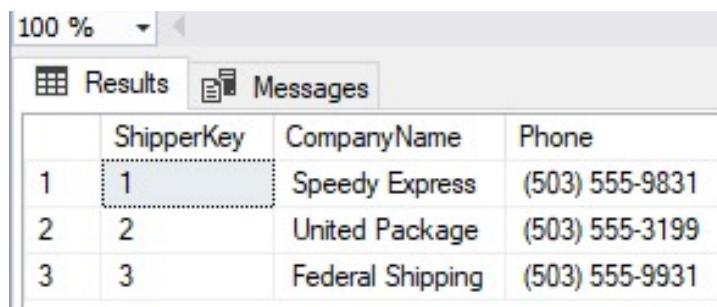
Input Column	Destination Column
ShipperID	ShipperKey
CompanyName	CompanyName
Phone	Phone

Ensuite on vérifie si les données de DimShipper sont bien chargées à partir de sql server:

```
use NorthwindDW
select * from dbo.DimShipper;
```

# DimShipper

Et voilà les données sont bien chargées:



The screenshot shows a Microsoft SQL Server Management Studio (SSMS) window with the 'Results' tab selected. The window title bar says '100 %'. The results grid displays the following data:

	ShipperKey	CompanyName	Phone
1	1	Speedy Express	(503) 555-9831
2	2	United Package	(503) 555-3199
3	3	Federal Shipping	(503) 555-9931

# DimCategory

Après la table DimShipper on charge les données de DimCategory en mettant aussi une source qui contient les données de la table categories de la base Northwind , et une destination qui est liée au tableau de DimCategory de la base NorthwindDW



Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder.

OLE DB connection manager:

DESKTOP-GBECG4E.Northwind

Data access mode:

Table or view

Name of the table or the view:

[dbo].[Categories]

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

OLE DB connection manager:

DESKTOP-GBECG4E.NorthwindDW

Data access mode:

Table or view - fast load

Name of the table or the view:

[dbo].[DimCategory]

Keep identity

Table lock

Keep nulls

Check constraints

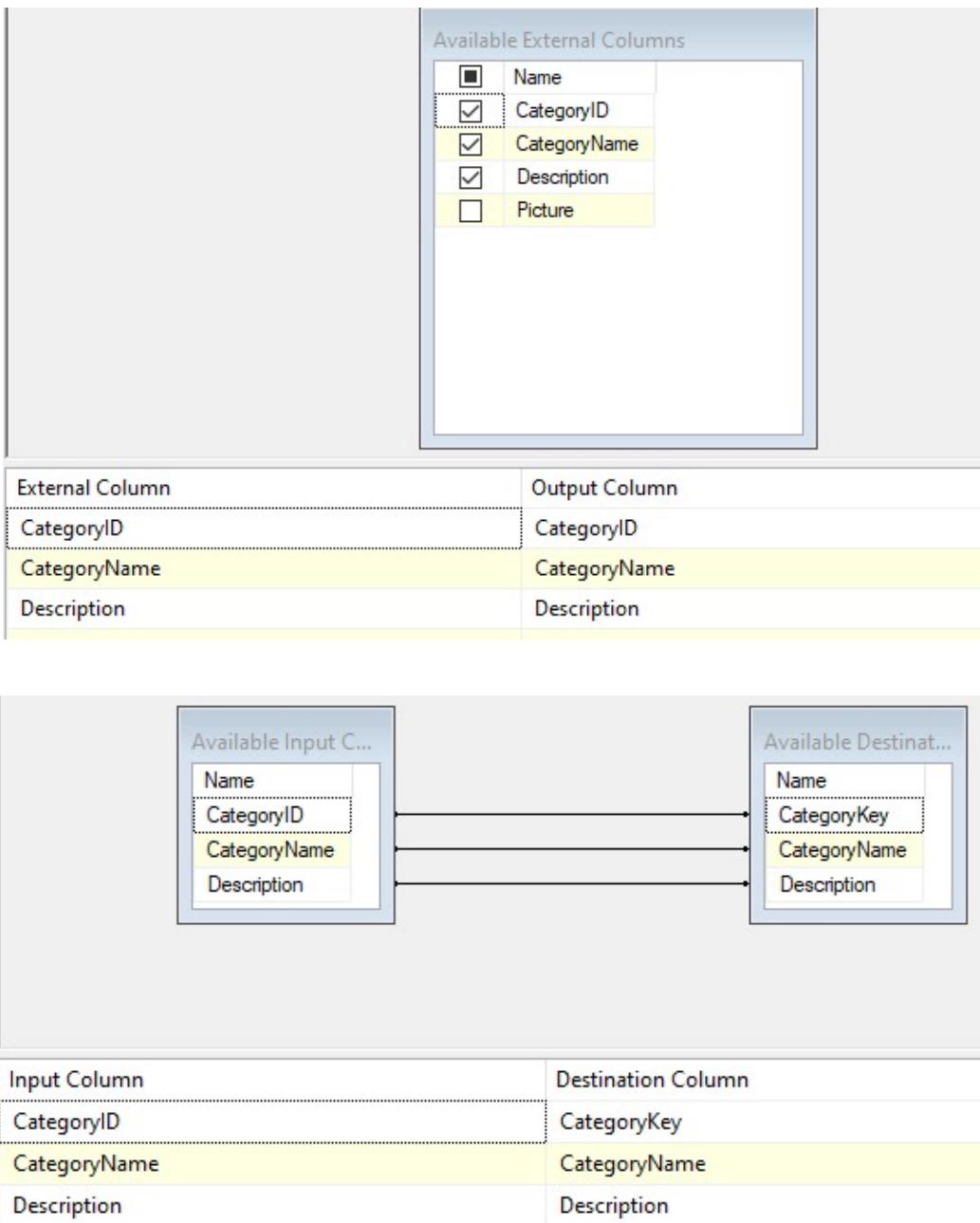
Rows per batch:

Maximum insert commit size:

2147483647

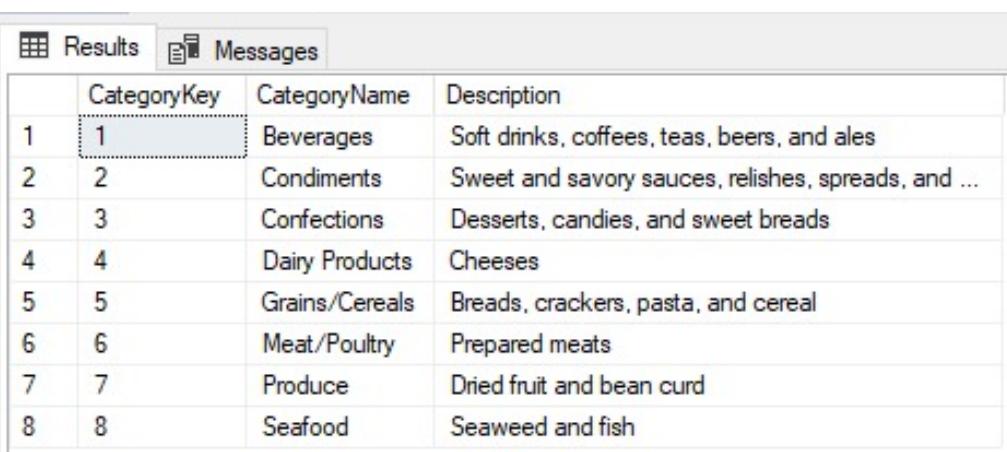
On prend seulement les colonnes qui sont nécessaires pour le remplissage des colonnes de DimCategory et on lie les colonnes de la source avec ceux de la destination:

# DimCategory



Enfin on vérifie si les données sont bien téléchargés:

```
SQLQuery1.sql - DE...GBECG4E\jerra (74)* ⇨ ×
use NorthwindDW
select * from dbo.DimCategory;
```



The table shows the data loaded into the DimCategory table.

	CategoryKey	CategoryName	Description
1	1	Beverages	Soft drinks, coffees, teas, beers, and ales
2	2	Condiments	Sweet and savory sauces, relishes, spreads, and ...
3	3	Confections	Desserts, candies, and sweet breads
4	4	Dairy Products	Cheeses
5	5	Grains/Cereals	Breads, crackers, pasta, and cereal
6	6	Meat/Poultry	Prepared meats
7	7	Produce	Dried fruit and bean curd
8	8	Seafood	Seaweed and fish

# DimEmployees

Pour le remplissage de DimEmployees on est besoin d'une source lié au tableau Employees de Northwind et d'une destination liée au tableau DimEmployees de NorthwindDW



OLE DB connection manager:

DESKTOP-GBECG4E.Northwind New...

Data access mode:

Table or view ▼

Name of the table or the view:

[dbo].[Employees] ▼

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

OLE DB connection manager:

DESKTOP-GBECG4E.NorthwindDW New...

Data access mode:

Table or view - fast load ▼

Name of the table or the view:

[dbo].[DimEmployee] ▼ New...

Keep identity  Table lock

Keep nulls  Check constraints

Rows per batch:

Maximum insert commit size:

View Existing Data...

# DimEmployees

On choisit les colonnes nécessaires pour le remplissage des colonnes de la destination DimEmployees de la base NorthwindDW

Et on lie les colonnes de destination et de source entre eux

	Available External Columns
<input type="checkbox"/>	Name
<input checked="" type="checkbox"/>	EmployeeID
<input checked="" type="checkbox"/>	LastName
<input checked="" type="checkbox"/>	FirstName
<input checked="" type="checkbox"/>	Title
<input checked="" type="checkbox"/>	TitleOfCourtesy
<input checked="" type="checkbox"/>	BirthDate
<input checked="" type="checkbox"/>	HireDate
<input checked="" type="checkbox"/>	Address
<input checked="" type="checkbox"/>	City
<input checked="" type="checkbox"/>	Region
<input checked="" type="checkbox"/>	PostalCode
<input checked="" type="checkbox"/>	Country
<input checked="" type="checkbox"/>	HomePhone
<input checked="" type="checkbox"/>	Extension
<input type="checkbox"/>	Photo
<input type="checkbox"/>	Notes
<input checked="" type="checkbox"/>	ReportsTo
<input type="checkbox"/>	PhotoPath
<input type="checkbox"/>	Salary

Input Column	Destination Column
EmployeeID	EmployeeKey
FirstName	FirstName
LastName	LastName
Title	Title
TitleOfCourtesy	TitleOfCourtesy
BirthDate	BirthDate
HireDate	HireDate
Address	Address
City	City
Region	Region

# DimEmployees

On vérifie le remplissage de DimEmployees avec la commande suivante

```
SQLQuery1.sql - DE...GBECG4E\jerra (74)* ➔ X
use NorthwindDW
select * from dbo.DimEmployee;
```

	EmployeeKey	FirstName	LastName	Title	TitleOfCourtesy	BirthDate	HireDate	Address	City	Region	PostalCode	Country	HomePhone
1	1	Nancy	Davolio	Sales Representative	Ms.	1948-12-08 00:00:00.000	1992-05-01 00:00:00.000	507 - 20th Ave. E Apt. 2A	Seattle	WA	98122	USA	(206) 555-9855
2	2	Andrew	Fuller	Vice President, Sales	Dr.	1952-02-19 00:00:00.000	1992-08-14 00:00:00.000	908 W. Capital Way	Tacoma	WA	98401	USA	(206) 555-9484
3	3	Janet	Leverling	Sales Representative	Ms.	1963-08-30 00:00:00.000	1992-04-01 00:00:00.000	722 Moss Bay Blvd.	Kirkland	WA	98033	USA	(206) 555-3411
4	4	Margaret	Peacock	Sales Representative	Mrs.	1937-09-19 00:00:00.000	1993-05-03 00:00:00.000	4110 Old Redmond Rd.	Redmond	WA	98052	USA	(206) 555-8122
5	5	Steven	Buchanan	Sales Manager	Mr.	1955-03-04 00:00:00.000	1993-10-17 00:00:00.000	14 Garrett Hill	London	NULL	SW1 8JR	UK	(71) 555-4848
6	6	Michael	Suyama	Sales Representative	Mr.	1963-07-02 00:00:00.000	1993-10-17 00:00:00.000	Coventry House Miner Rd.	London	NULL	EC2 7JR	UK	(71) 555-7773
7	7	Robert	King	Sales Representative	Mr.	1960-05-29 00:00:00.000	1994-01-02 00:00:00.000	Edgeham Hollow Winchester Way	London	NULL	RG1 9SP	UK	(71) 555-5598
8	8	Laura	Callahan	Inside Sales Coordinator	Ms.	1958-01-09 00:00:00.000	1994-03-05 00:00:00.000	4726 - 11th Ave. N.E.	Seattle	WA	98105	USA	(206) 555-1181
9	9	Anne	Dodsworth	Sales Representative	Ms.	1966-01-27 00:00:00.000	1994-11-15 00:00:00.000	7 Houndstooth Rd.	London	NULL	WG2 7LT	UK	(71) 555-4444

Et voilà le tableau DimEmployees est bien remplis:

Extension	SupervisorKey
5467	2
3457	NULL
3355	2
5176	2
3453	2
428	5
465	5
2344	2
452	5

# DimProduct

Ensuite on remplit le tableau DimProduct pour cela on a mis dans la source le tableau Product de Northwind et destination contenant le tableau DimProduct de NorthwindDW:



Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder.

OLE DB connection manager: DESKTOP-GBECG4E.Northwind New...

Data access mode: Table or view

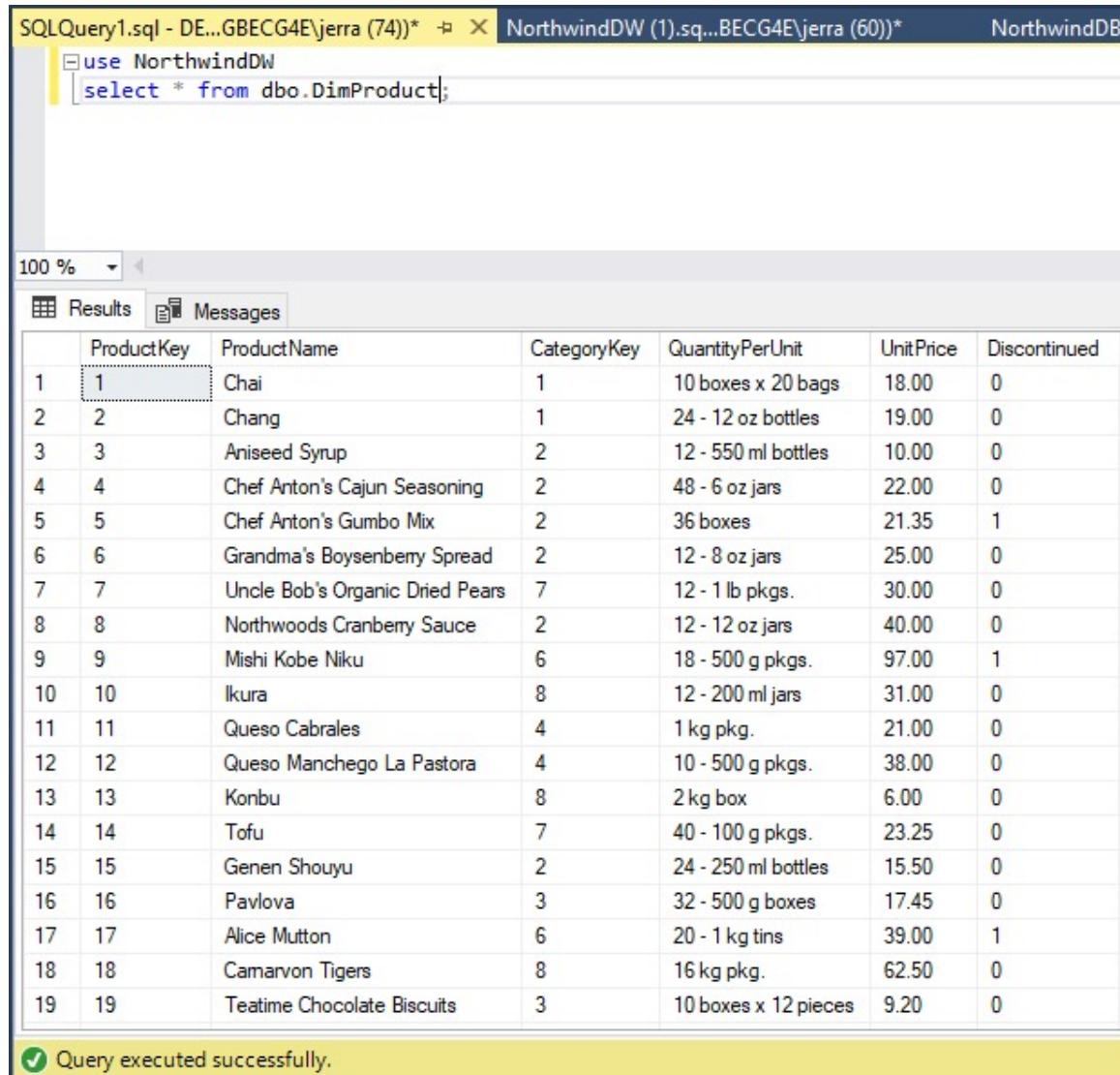
Name of the table or the view: [dbo].[Products]

Available External Columns

External Column	Output Column
ProductID	ProductID
ProductName	ProductName
QuantityPerUnit	QuantityPerUnit

# DimProduct

Enfin on valide le chargement des données



The screenshot shows the SSMS interface with a query window and a results grid. The query window contains the following SQL code:

```
use NorthwindDW
select * from dbo.DimProduct;
```

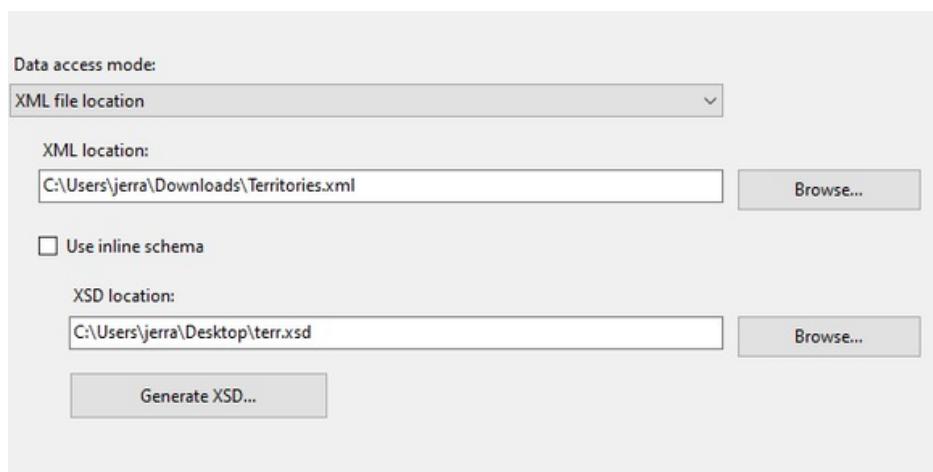
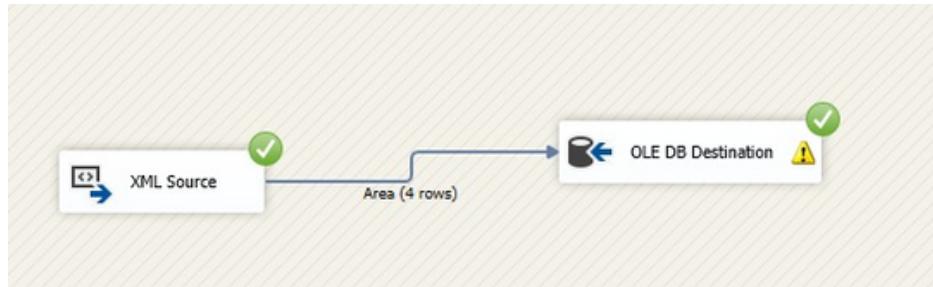
The results grid displays 19 rows of data from the DimProduct table, with columns: ProductKey, ProductName, CategoryKey, QuantityPerUnit, UnitPrice, and Discontinued. The data includes various products like Chai, Chang, Aniseed Syrup, etc., with their respective details. A message at the bottom indicates the query was executed successfully.

	ProductKey	ProductName	CategoryKey	QuantityPerUnit	UnitPrice	Discontinued
1	1	Chai	1	10 boxes x 20 bags	18.00	0
2	2	Chang	1	24 - 12 oz bottles	19.00	0
3	3	Aniseed Syrup	2	12 - 550 ml bottles	10.00	0
4	4	Chef Anton's Cajun Seasoning	2	48 - 6 oz jars	22.00	0
5	5	Chef Anton's Gumbo Mix	2	36 boxes	21.35	1
6	6	Grandma's Boysenberry Spread	2	12 - 8 oz jars	25.00	0
7	7	Uncle Bob's Organic Dried Pears	7	12 - 1 lb pkgs.	30.00	0
8	8	Northwoods Cranberry Sauce	2	12 - 12 oz jars	40.00	0
9	9	Mishi Kobe Niku	6	18 - 500 g pkgs.	97.00	1
10	10	Ikura	8	12 - 200 ml jars	31.00	0
11	11	Queso Cabrales	4	1 kg pkg.	21.00	0
12	12	Queso Manchego La Pastora	4	10 - 500 g pkgs.	38.00	0
13	13	Konbu	8	2 kg box	6.00	0
14	14	Tofu	7	40 - 100 g pkgs.	23.25	0
15	15	Genen Shouyu	2	24 - 250 ml bottles	15.50	0
16	16	Pavlova	3	32 - 500 g boxes	17.45	0
17	17	Alice Mutton	6	20 - 1 kg tins	39.00	1
18	18	Camarvon Tigers	8	16 kg pkg.	62.50	0
19	19	Teatime Chocolate Biscuits	3	10 boxes x 12 pieces	9.20	0

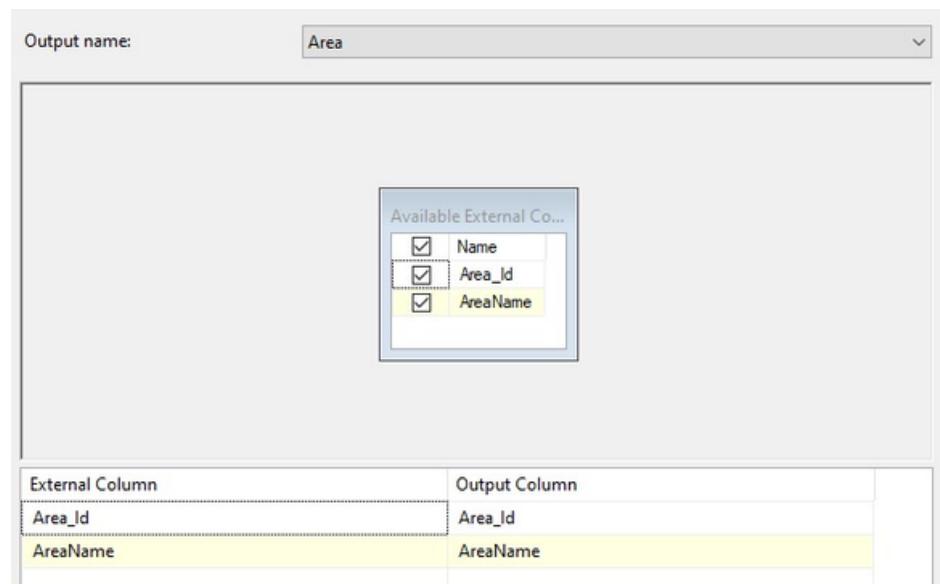
Query executed successfully.

# DimArea

Afin de remplir les colonnes de DimArea on utilise comme source de données le fichier Territories.xml:



Comme output on choisit Area:



# DimArea

Pour la destination on la lie avec le tableau DimArea de NorthwindDW et on lie les colonnes de destination et du fichier xml entre eux:

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

OLE DB connection manager: DESKTOP-GBECG4E.NorthwindDW New...

Data access mode: Table or view - fast load

Name of the table or the view: [dbo].[DimArea] New...

Keep identity  Table lock

Keep nulls  Check constraints

Rows per batch:

Maximum insert commit size: 2147483647

The diagram shows two boxes: 'Available Input...' and 'Available Dest...'. The 'Available Input...' box contains columns 'Name', 'Area\_Id', and 'AreaName'. The 'Available Dest...' box contains columns 'Name', 'AreaKey', and 'AreaName'. Two arrows point from 'Area\_Id' to 'AreaKey' and from 'AreaName' to 'AreaName', indicating a self-join or a copy operation.

Input Column	Destination Column
Area_Id	AreaKey
AreaName	AreaName

On vérifie le téléchargement des données dans DimArea et voilà les données sont bien téléchargées:

SQLQuery1.sql - DE...GBECG4E\jerra (74)\* ↗ X NorthwindDW

```
use NorthwindDW
select * from dbo.DimArea;
```

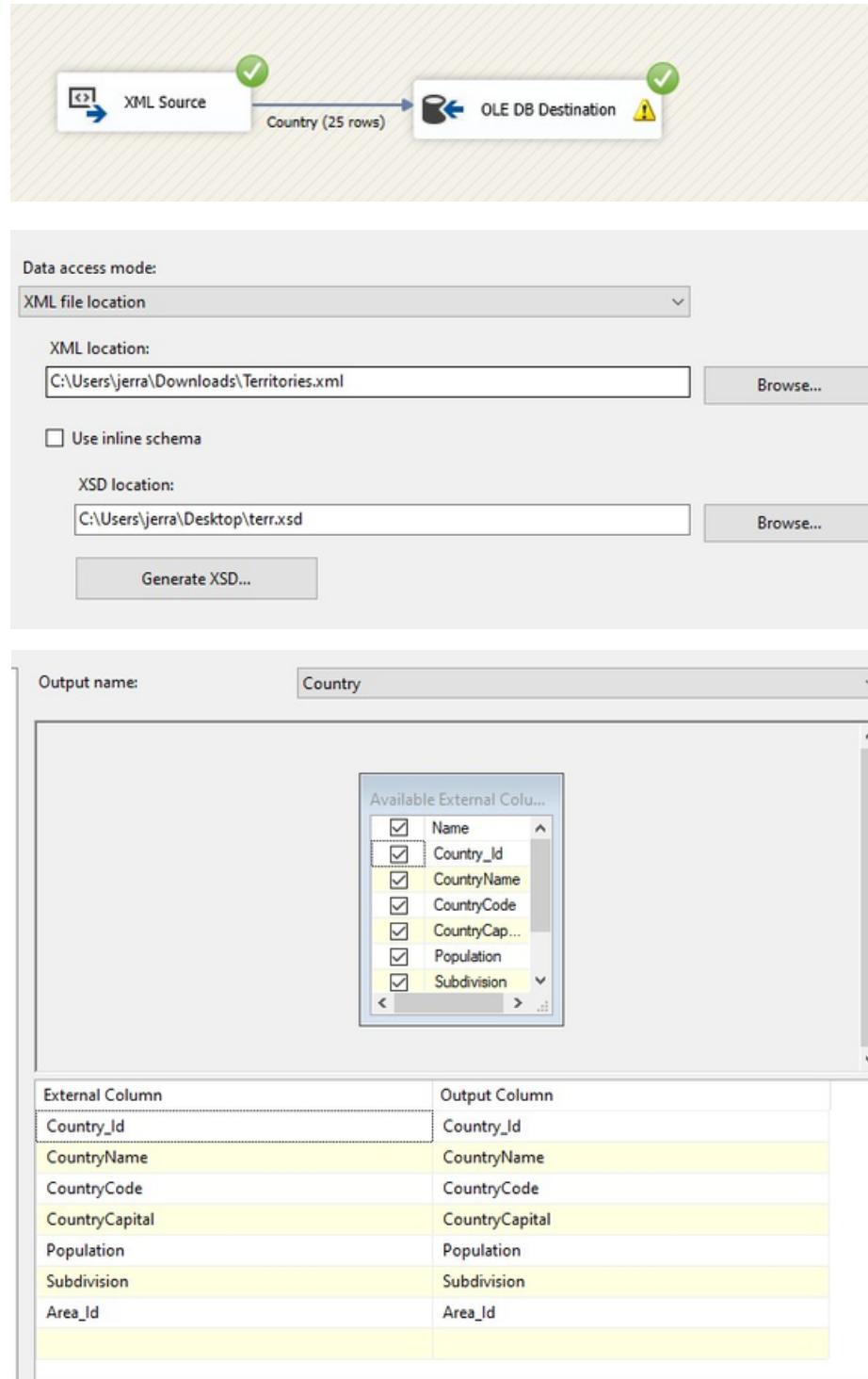
100 %

Results Messages

	AreaKey	AreaName
1	1	Europe
2	3294	North America
3	3734	Pacific
4	4028	South America

# DimCountry

Pour DimCountry on utilise le même fichier que DimArea et on choisit comme output Country



# DimCountry

Pour la destination on la connecte avec NorthwindDW et le tableau DimCountry: et afin de remplir ses colonnes on les lie avec ceux du fichier xml:

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

OLE DB connection manager: DESKTOP-GBECG4E.NorthwindDW New...

Data access mode: Table or view - fast load

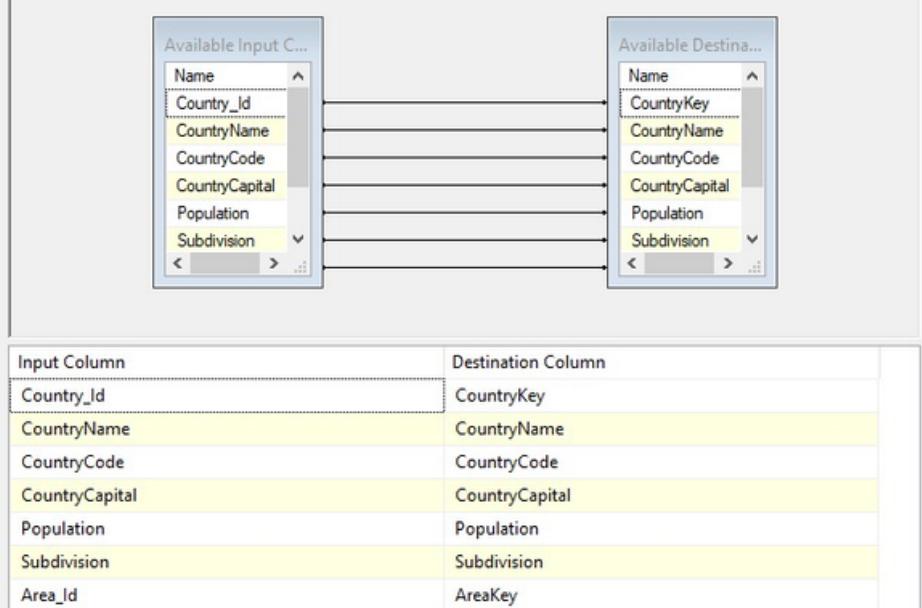
Name of the table or the view: [dbo].[DimCountry] New...

Keep identity  Table lock

Keep nulls  Check constraints

Rows per batch:

Maximum insert commit size: 2147483647

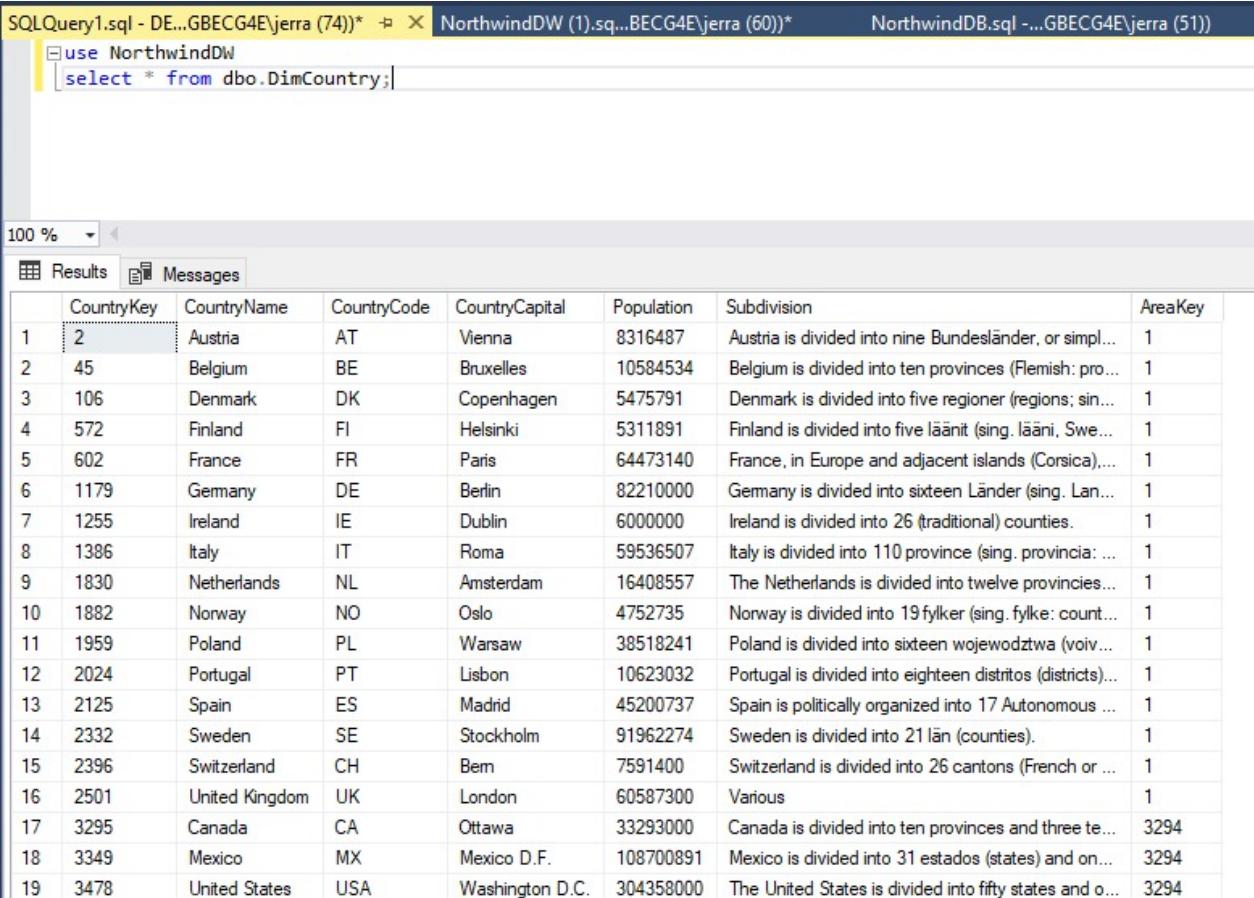


The screenshot shows the 'Available Input Columns' and 'Available Destination Columns' panes. Arrows connect the columns from the input pane to the destination pane. The 'Input Column' and 'Destination Column' mapping table below shows the final mapping.

Input Column	Destination Column
Country_Id	CountryKey
CountryName	CountryName
CountryCode	CountryCode
CountryCapital	CountryCapital
Population	Population
Subdivision	Subdivision
Area_Id	AreaKey

# DimCountry

Voilà les données sont bien téléchargées dans le tableau DimCountry et le vérifie en utilisant la commande suivant dans sql server:



The screenshot shows a SQL Server Management Studio (SSMS) window with three tabs: 'SQLQuery1.sql - DE...GBECG4E\jerra (74)\*', 'NorthwindDW (1).sq...BECG4E\jerra (60)\*', and 'NorthwindDB.sql -...GBECG4E\jerra (51)'. The 'Results' tab is selected, displaying the output of the following query:

```
use NorthwindDW
select * from dbo.DimCountry;
```

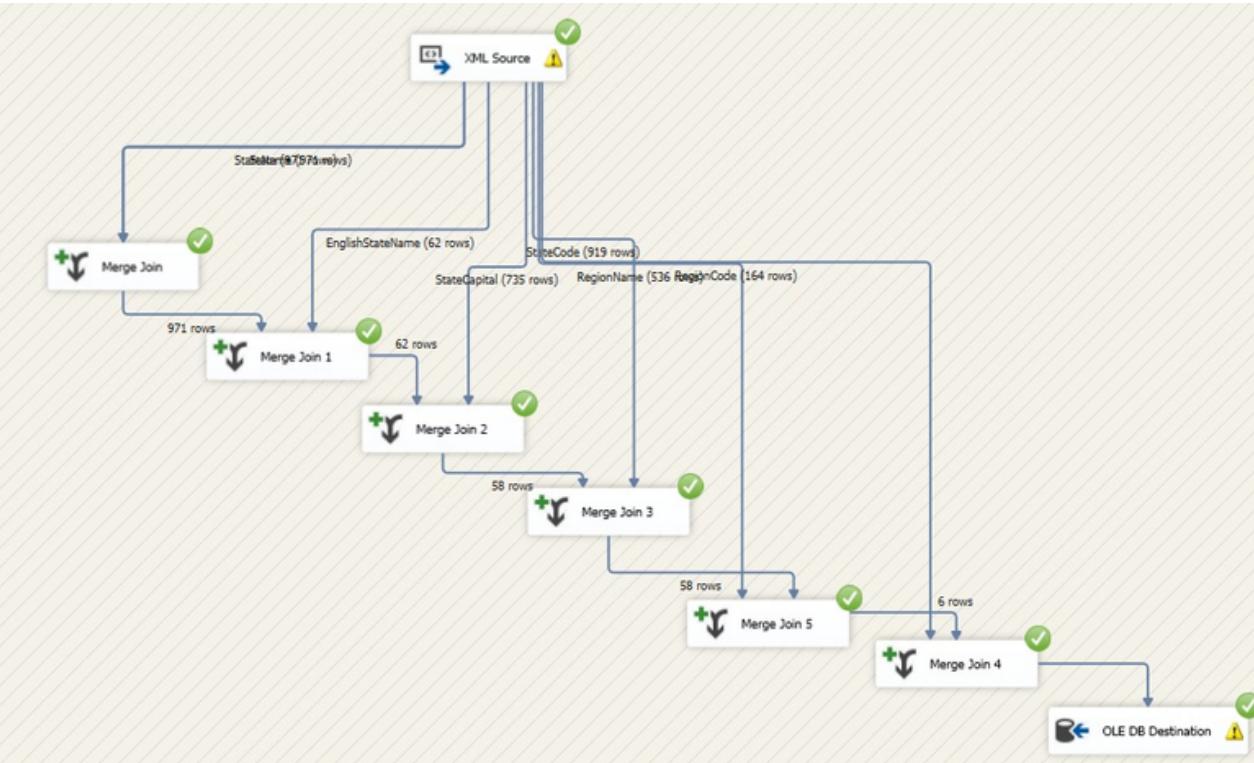
The results table contains 19 rows of data, each representing a country with its key, name, code, capital, population, subdivision, and area key. The data includes entries for Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Italy, Netherlands, Norway, Poland, Portugal, Spain, Sweden, Switzerland, United Kingdom, Canada, Mexico, and the United States.

	CountryKey	CountryName	CountryCode	CountryCapital	Population	Subdivision	AreaKey
1	2	Austria	AT	Vienna	8316487	Austria is divided into nine Bundesländer, or simpl...	1
2	45	Belgium	BE	Bruxelles	10584534	Belgium is divided into ten provinces (Flemish: pro...	1
3	106	Denmark	DK	Copenhagen	5475791	Denmark is divided into five regioner (regions; sin...	1
4	572	Finland	FI	Helsinki	5311891	Finland is divided into five läänit (sing. lääni, Swe...	1
5	602	France	FR	Paris	64473140	France, in Europe and adjacent islands (Corsica)....	1
6	1179	Germany	DE	Berlin	82210000	Germany is divided into sixteen Länder (sing. Lan...	1
7	1255	Ireland	IE	Dublin	6000000	Ireland is divided into 26 (traditional) counties.	1
8	1386	Italy	IT	Roma	59536507	Italy is divided into 110 province (sing. provincia: ...	1
9	1830	Netherlands	NL	Amsterdam	16408557	The Netherlands is divided into twelve provincies...	1
10	1882	Norway	NO	Oslo	4752735	Norway is divided into 19 fylker (sing. fylke: count...	1
11	1959	Poland	PL	Warsaw	38518241	Poland is divided into sixteen województwa (voiv...	1
12	2024	Portugal	PT	Lisbon	10623032	Portugal is divided into eighteen distritos (districts)...	1
13	2125	Spain	ES	Madrid	45200737	Spain is politically organized into 17 Autonomous ...	1
14	2332	Sweden	SE	Stockholm	91962274	Sweden is divided into 21 län (counties).	1
15	2396	Switzerland	CH	Bern	7591400	Switzerland is divided into 26 cantons (French or ...	1
16	2501	United Kingdom	UK	London	60587300	Various	1
17	3295	Canada	CA	Ottawa	33293000	Canada is divided into ten provinces and three ter...	3294
18	3349	Mexico	MX	Mexico D.F.	108700891	Mexico is divided into 31 estados (states) and on...	3294
19	3478	United States	USA	Washington D.C.	304358000	The United States is divided into fifty states and o...	3294

# DimState

Pour DimState c'est différent par rapport aux autres car une seule liaison de xml file et destination ne suffit pas donc on avait besoin d'utiliser plusieurs "Merge join" afin de remplir les colonnes de la destination étape par étape.

Voilà le schéma utilisé:



Pour la source on met le fichier xml Territories:

Data access mode:

XML file location: C:\Users\jerra\Downloads\Territories.xml

XML location: C:\Users\jerra\Downloads\Territories.xml

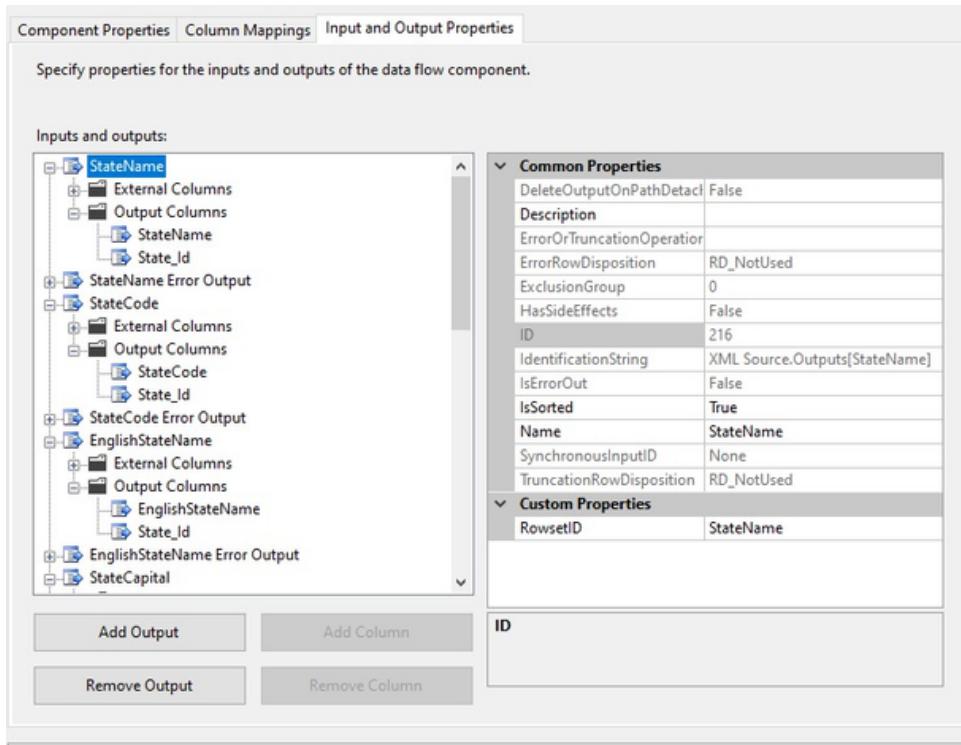
Use inline schema

XSD location: C:\Users\jerra\Desktop\terr.xsd

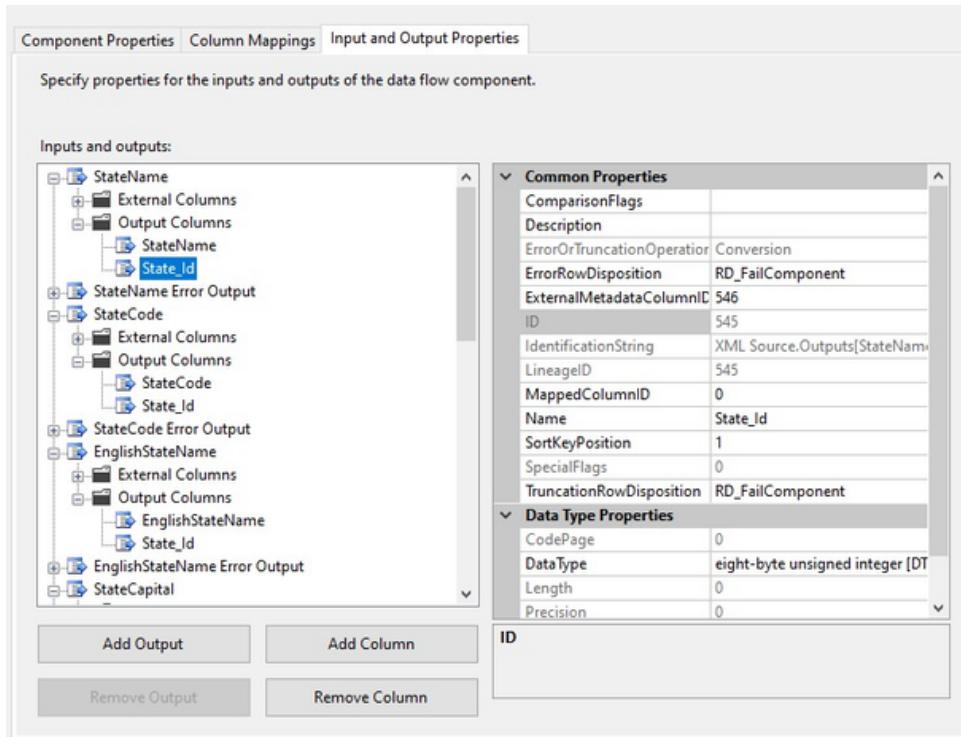
Generate XSD...

# DimState

Là on essaye de trier les données du fichier xml en changeant la valeur de IsSorted de False en True, sinon merge ne fonctionne pas



Là on précise quelle sera la colonne selon laquelle on fera le tri en changeant la valeur de SortKeyPosition. Cela se fait en accédant au "Input and Output Properties" et en cliquant sur State\_Id:



# DimState

## Chaque merge est faite selon State\_Id

Configure the properties used to join two sources of sorted data. Select the join type and then specify the columns to be used as the join key. Join keys must be used in the order specified by the sort-key position of the column.

Join type: Left outer join Swap Inputs

**Merge Join 3**

Name	Order	Join Key
<input checked="" type="checkbox"/> StateCapital	0	<input type="checkbox"/>
<input checked="" type="checkbox"/> EnglishStateName	0	<input type="checkbox"/>
<input checked="" type="checkbox"/> State_Id	1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> type	0	<input type="checkbox"/>
<input checked="" type="checkbox"/> Country_Id	0	<input type="checkbox"/>
<input checked="" type="checkbox"/> StateName	0	<input type="checkbox"/>

**XML Source**

Name	Order	Join Key
<input type="checkbox"/>		
<input checked="" type="checkbox"/> RegionName	0	<input type="checkbox"/>
<input type="checkbox"/>	1	<input checked="" type="checkbox"/>
State_Id		

Input	Input Column	Output Alias
Merge Join 3	StateCapital	StateCapital
Merge Join 3	EnglishStateName	EnglishStateName
Merge Join 3	State_Id	State_Id
Merge Join 3	type	type
Merge Join 3	Country_Id	Country_Id
Merge Join 3	StateName	StateName
Merge Join 3	StateCode	StateCode
XML Source	RegionName	RegionName

La destination finale est liée au tableau DimState de la base NorthwindDW

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

OLE DB connection manager: DESKTOP-GBECG4E.NorthwindDW New...

Data access mode: Table or view - fast load

Name of the table or the view: [dbo].[DimState] New...

Keep identity  Table lock

Keep nulls  Check constraints

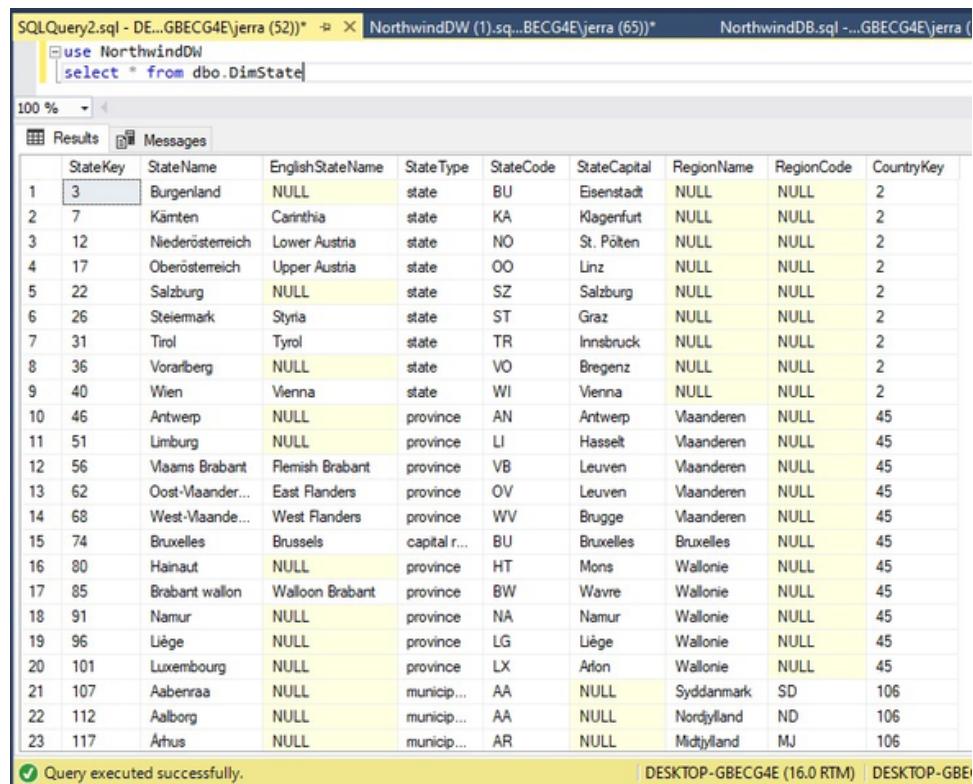
Rows per batch:

Maximum insert commit size:

[View Existing Data...](#)

# DimState

Là on vérifie le téléchargement des données et voilà sont bien téléchargées:



	StateKey	StateName	EnglishStateName	StateType	StateCode	StateCapital	RegionName	RegionCode	CountryKey
1	3	Burgenland	NULL	state	BU	Eisenstadt	NULL	NULL	2
2	7	Kärnten	Carinthia	state	KA	Klagenfurt	NULL	NULL	2
3	12	Niederösterreich	Lower Austria	state	NO	St. Pölten	NULL	NULL	2
4	17	Oberösterreich	Upper Austria	state	OO	Linz	NULL	NULL	2
5	22	Salzburg	NULL	state	SZ	Salzburg	NULL	NULL	2
6	26	Steiermark	Styria	state	ST	Graz	NULL	NULL	2
7	31	Tirol	Tyrol	state	TR	Innsbruck	NULL	NULL	2
8	36	Vorarlberg	NULL	state	VO	Bregenz	NULL	NULL	2
9	40	Wien	Vienna	state	WI	Vienna	NULL	NULL	2
10	46	Antwerp	NULL	province	AN	Antwerp	Vlaanderen	NULL	45
11	51	Limburg	NULL	province	LI	Hasselt	Vlaanderen	NULL	45
12	56	Vlaams Brabant	Flemish Brabant	province	VB	Leuven	Vlaanderen	NULL	45
13	62	Oost-Vlaanderen	East Flanders	province	OV	Leuven	Vlaanderen	NULL	45
14	68	West-Vlaanderen	West Flanders	province	WV	Brugge	Vlaanderen	NULL	45
15	74	Bruxelles	Brussels	capital r...	BU	Bruxelles	Bruxelles	NULL	45
16	80	Hainaut	NULL	province	HT	Mons	Wallonie	NULL	45
17	85	Brabant wallon	Walloon Brabant	province	BW	Wavre	Wallonie	NULL	45
18	91	Namur	NULL	province	NA	Namur	Wallonie	NULL	45
19	96	Liège	NULL	province	LG	Liège	Wallonie	NULL	45
20	101	Luxembourg	NULL	province	LX	Arlon	Wallonie	NULL	45
21	107	Aabenraa	NULL	municip...	AA	NULL	Syddanmark	SD	106
22	112	Aalborg	NULL	municip...	AA	NULL	Nordjylland	ND	106
23	117	Århus	NULL	municip...	AR	NULL	Midtjylland	MJ	106

# DimGeography

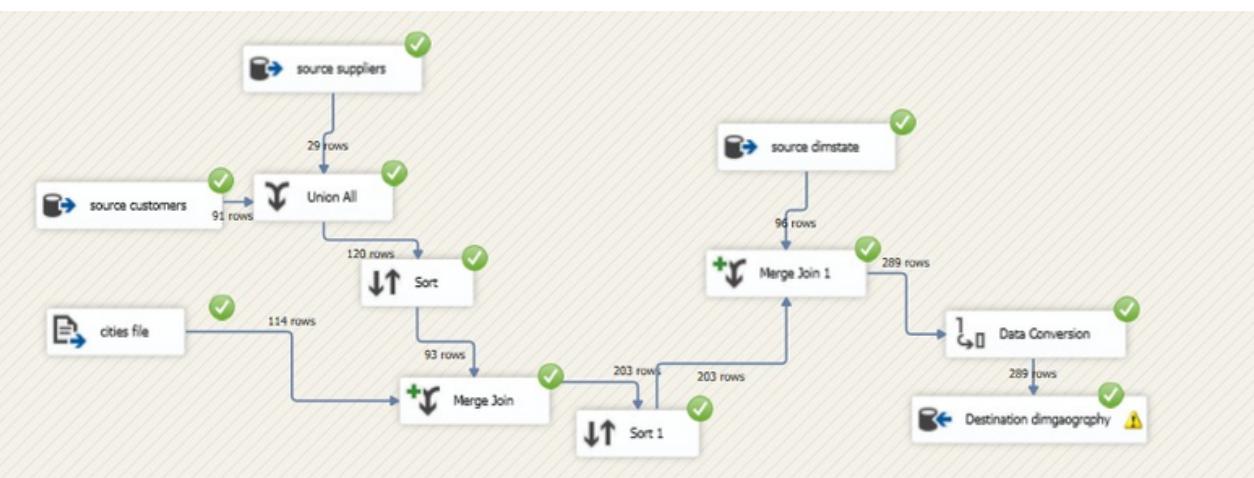
Pour charger la table DimSupplier on s'est servit des deux tables opérationnels customer et supplier pour remplir les attributs city et postal\_code et cela en utilisant la fonction union\_all.

Et afin d'identifier à quel "state" correspond une city on a utilisé le fichier cities.txt en utilisant la fonction merge.

D'une autre part on a aussi utilisé la table DimState pour remplir l'attribut StateKey

Enfin on utilisé la fonction merge join 1 afin de charger la table de destination DimGeography

On a utilisé DataConversion pour corriger une erreur qui nous a rencontré il s'agit de la différence de type de city qui provient de DimStae et city qui provient de cities file



Voilà les étapes de configuration:

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder.

OLE DB connection manager: DESKTOP-GBECG4E.Northwind New...

Data access mode: Table or view

Name of the table or the view: [dbo].[Suppliers]

# DimGeography

Ici on sélectionne seulement les attributs dont on a besoin de la table Suppliers

The screenshot shows the 'Available External Columns' list and the 'External Column' mapping table for the DimGeography dimension.

**Available External Columns:**

Column
Name
ContactName
ContactTitle
Address
<input checked="" type="checkbox"/> City
<input checked="" type="checkbox"/> Region
<input checked="" type="checkbox"/> PostalCode
<input checked="" type="checkbox"/> Country
Phone
Fax

**External Column Mapping:**

External Column	Output Column
City	City
Region	Region
PostalCode	PostalCode
Country	Country

On fait de même pour la table Customers:

The screenshot shows the 'OLE DB connection manager' and 'Data access mode' settings for the DimGeography dimension.

**OLE DB connection manager:** DESKTOP-G8EBCG4E.Northwind

**Data access mode:** Table or view

**Name of the table or the view:** [dbo].[Customers]

The screenshot shows the 'Available External Columns' list and the 'External Column' mapping table for the DimGeography dimension.

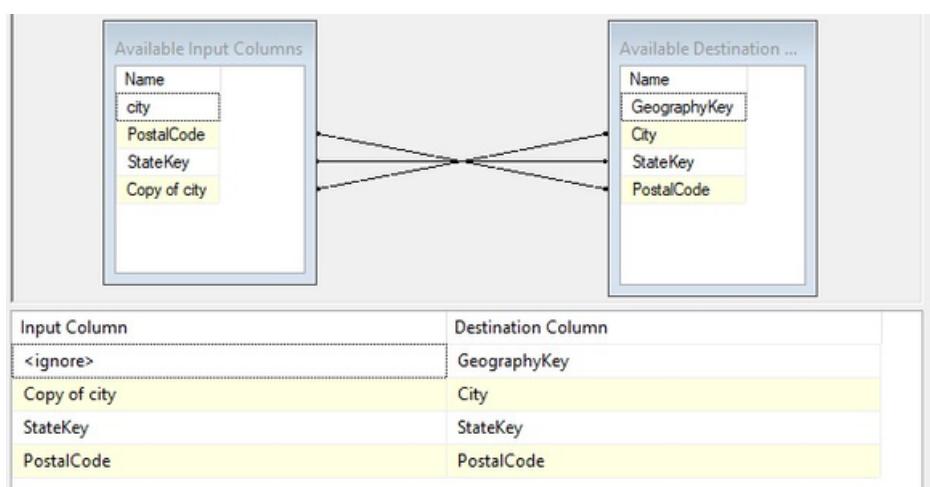
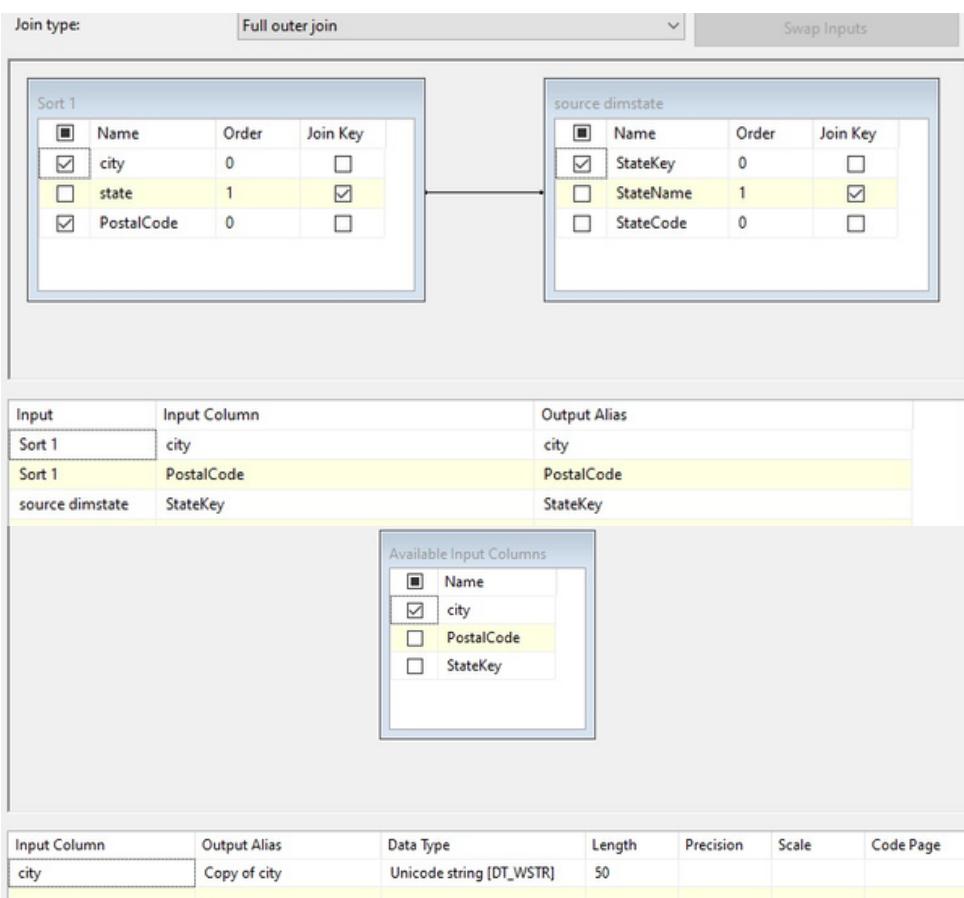
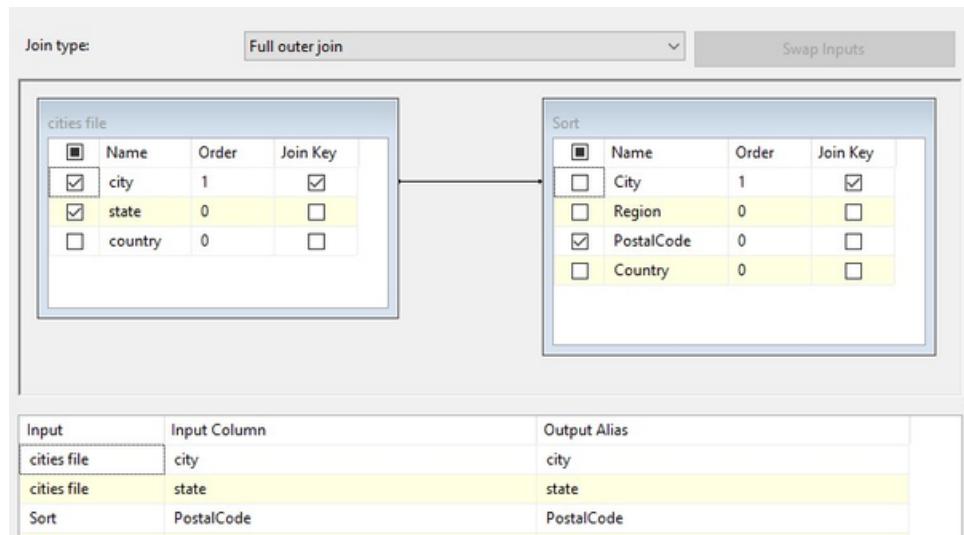
**Available External Columns:**

Column
Name
<input type="checkbox"/> CustomerID
<input type="checkbox"/> CompanyName
<input type="checkbox"/> ContactName
<input type="checkbox"/> ContactTitle
<input type="checkbox"/> Address
<input checked="" type="checkbox"/> City
<input checked="" type="checkbox"/> Region
<input checked="" type="checkbox"/> PostalCode
<input checked="" type="checkbox"/> Country

**External Column Mapping:**

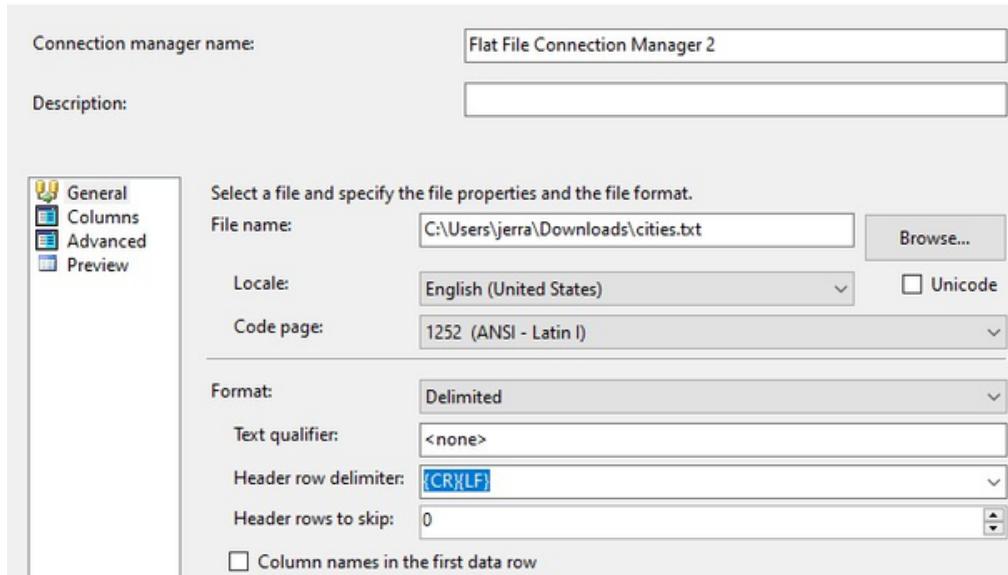
External Column	Output Column
City	City
Region	Region
PostalCode	PostalCode
Country	Country

# DimGeography

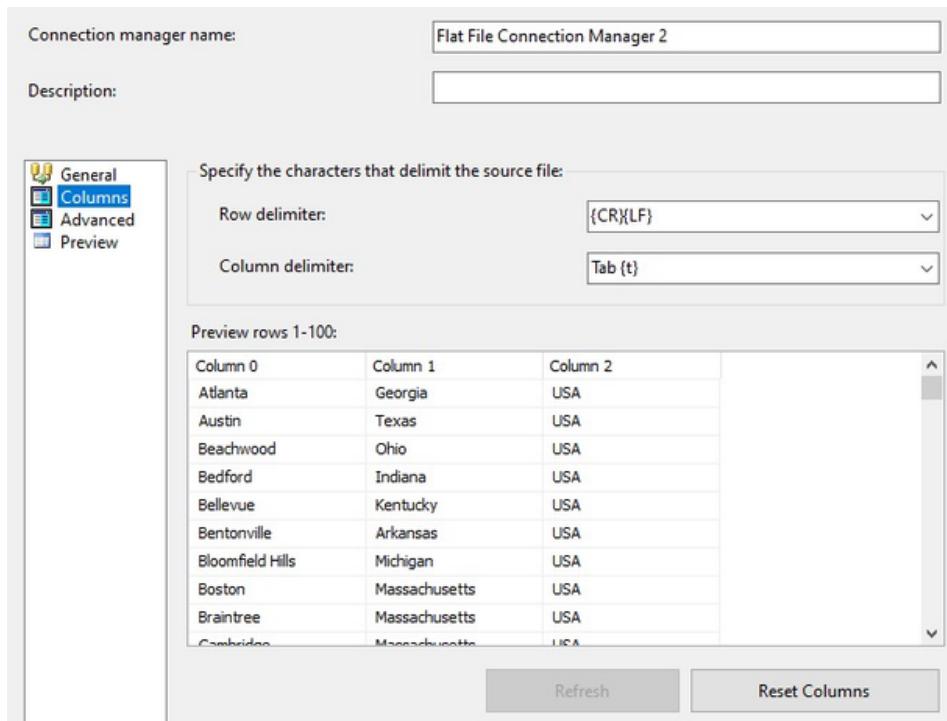


# DimGeography

Là concernant le fichier cities.txt on a utilisé la source flat file  
On décoche la dernière instruction puisque la première ligne de notre fichier ne concerne pas les noms des colonnes



Les trois colonnes sont affichés comme suit:



# DimGeography

Après on change les noms de colonnes:

colonne 0 : correspond à cities

colonne 1 : correspond à state

colonne 2 : correspond à country

Configure the properties of each column.

Column 0	Column 1	Column 2

**Misc**

Name	Column 0
ColumnDelimiter	Tab (t)
ColumnType	Delimited
InputColumnWidth	0
DataPrecision	0
DataScale	0
DataType	string [DT_STR]
OutputColumnWidth	50
TextQualified	True

**Name**

External Column	Output Column
city	city
state	state
country	country

Available External ...

Name
city
state
state
country

External Column	Output Column
city	city
state	state
country	country

# DimGeography

On vérifie le téléchargement des données et voilà le résultat

SQLQuery1.sql - DE...GBECG4E\jerra (59)\*

use NorthwindDW  
select \* from dbo.DimGeography

100 %

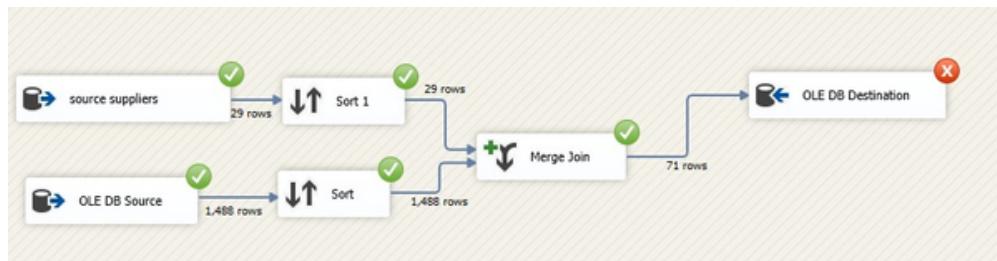
Results Messages

	GeographyKey	City	StateKey	PostalCode
301	301	NULL	933	NULL
302	302	London	NULL	NULL
303	303	NULL	939	NULL
304	304	NULL	945	NULL
305	305	Melville	NULL	NULL
306	306	Cuxhaven	NULL	NULL
307	307	NULL	951	NULL
308	308	Lyngby	NULL	NULL
309	309	Madrid	NULL	NULL
310	310	Sevilla	NULL	NULL
311	311	NULL	957	NULL
312	312	NULL	963	NULL
313	313	Manches...	NULL	NULL
314	314	Reims	969	NULL
315	315	Rockville	NULL	NULL
316	316	Braintree	NULL	NULL
317	317	Boston	NULL	02134
318	318	Westboro	NULL	NULL
319	319	Cambridge	NULL	NULL
320	320	NULL	975	NULL
321	321	Warszawa	NULL	NULL
322	322	NULL	981	NULL
323	323	NULL	987	NULL

# DimSupplier

Afin de remplir la dimension DimSuppliers on utilise la table opérationnel suppliers et l'attribut GeographyKey de la dimension DimGeography en utilisant la fonction merge join qui permet de fusionner ces deux table dans une seule table cible qui est DimSuppliers.

Même si une erreur est affiché sur la destination(DimSupplier), le téléchargement était fait avec succès



Voici la configuration du Merge join utilisé:

Configure the properties used to join two sources of sorted data. Select the join type and then specify the columns to be used as the join key. Join keys must be used in the order specified by the sort-key position of the column.

Join type: Left outer join Swap Inputs

**Merge Join 3**

Name	Order	Join Key
StateCapital	0	<input type="checkbox"/>
EnglishStateName	0	<input type="checkbox"/>
State_Id	1	<input checked="" type="checkbox"/>
type	0	<input type="checkbox"/>
Country_Id	0	<input type="checkbox"/>
StateName	0	<input type="checkbox"/>

**XML Source**

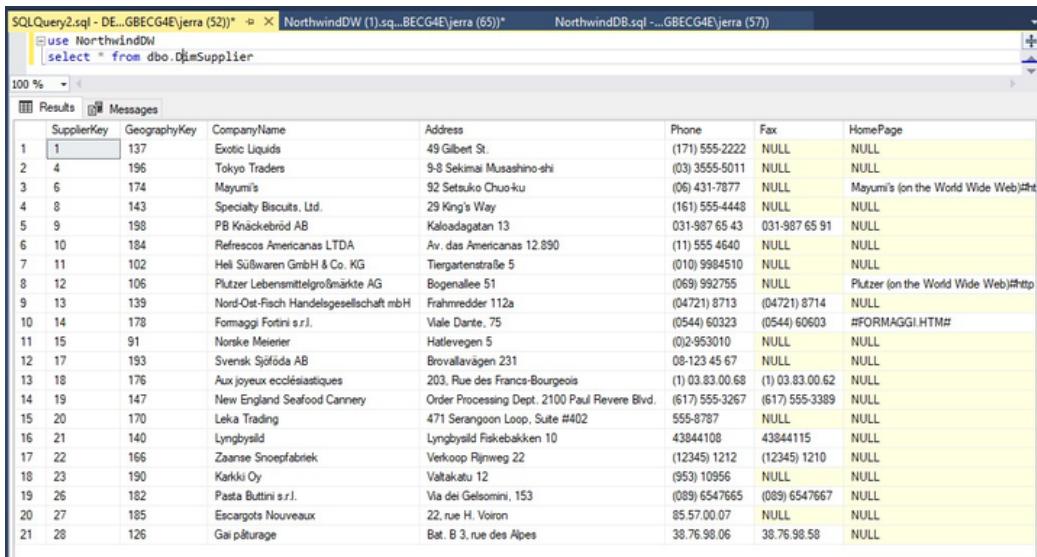
Name	Order	Join Key
RegionName	0	<input type="checkbox"/>
State_Id	1	<input checked="" type="checkbox"/>

**Input Column Mappings**

Input	Input Column	Output Alias
Merge Join 3	StateCapital	StateCapital
Merge Join 3	EnglishStateName	EnglishStateName
Merge Join 3	State_Id	State_Id
Merge Join 3	type	type
Merge Join 3	Country_Id	Country_Id
Merge Join 3	StateName	StateName
Merge Join 3	StateCode	StateCode
XML Source	RegionName	RegionName

# DimSupplier

Et voilà le résultat final:



The screenshot shows a SQL Server Management Studio (SSMS) window with a query results grid. The query is:

```
use NorthwindDW
select * from dbo.DimSupplier
```

The results grid has the following columns:

SupplierKey	GeographyKey	CompanyName	Address	Phone	Fax	HomePage
1	137	Exotic Liquids	49 Gilbert St.	(171) 555-2222	NULL	NULL
2	4	Tokyo Traders	9-8 Sekimai Musashino-shi	(03) 3555-5011	NULL	NULL
3	6	Mayumi's	92 Setsuko Chuo-ku	(06) 431-7877	NULL	Mayumi's (on the World Wide Web)http://www.mayumis.com
4	8	Specialty Biscuits, Ltd.	29 King's Way	(161) 555-4448	NULL	NULL
5	9	PB Knäckebrot AB	Kaloadagatan 13	031-987 65 43	031-987 65 91	NULL
6	10	Refrescos Americanas LTDA	Av. das Americanas 12.890	(11) 555 4640	NULL	NULL
7	11	Hel Süßwaren GmbH & Co. KG	Tiergartenstraße 5	(010) 9984510	NULL	NULL
8	12	Plutzer Lebensmittelgroßmärkte AG	Bogenallee 51	(069) 992755	NULL	Plutzer (on the World Wide Web)http://www.plutzer.de
9	13	Nord-Ost-Fisch Handelsgesellschaft mbH	Frahmredder 112a	(04721) 8713	(04721) 8714	NULL
10	14	Formaggi Fortini s.r.l.	Viale Dante, 75	(0544) 60323	(0544) 60603	#FORMAGGI.HTM#
11	15	Norske Meierier	Høstevagen 5	(02) 953010	NULL	NULL
12	17	Svensk Sjöföda AB	Brovallavägen 231	08-123 45 67	NULL	NULL
13	18	Aux joyeux ecclésiastiques	203, Rue des Francs-Bourgeois	(1) 03.80.00.68	(1) 03.83.00.62	NULL
14	19	New England Seafood Cannery	Order Processing Dept. 2100 Paul Revere Blvd.	(617) 555-3267	(617) 555-3389	NULL
15	20	Leka Trading	471 Serangoon Loop, Suite #402	555-8787	NULL	NULL
16	21	Lyngby Sil	Lyngbysild Fiskebækken 10	43844108	43844115	NULL
17	22	Zaanse Snoepfabriek	Verkoop Rijnweg 22	(12345) 1212	(12345) 1210	NULL
18	23	Karkki Oy	Valtakatu 12	(953) 10956	NULL	NULL
19	26	Pasta Buttini s.r.l.	Via dei Gelsomini, 153	(089) 6547665	(089) 6547667	NULL
20	27	Escargots Nouveaux	22, rue H. Voirin	85.57.00.07	NULL	NULL
21	28	Gai pâturage	Bat. B 3, rue des Alpes	38.76.98.06	38.76.98.58	NULL

# 04

## Conclusion

Ce projet nous a permis d'exploiter les connaissances acquises durant le module de "Fondements de mise en œuvre de système décisionnel" et d'apprendre à utiliser plusieurs outils tels que "SSDT" " SSMS" "SSIS" pour implémenter le processus ETL.

Nous estimons que nous avons pu atteindre l'objectif que le projet vise .

On tient à remercier [Mme. Akkaoui Zineb](#) pour son suivi et son encadrement durant le module de "Fondements de mise en œuvre de système décisionnel".