

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la recherche Scientifique
Université des Sciences et de la Technologie Houari Boumediene
Faculté d'Informatique



Rapport de Projet

Analyse de la base de données Northwind

Module : Business Intelligence

Spécialité : Big Data Analytics

Réaliser par :

BENYAMINA ALI

Introduction :

Ce projet a pour objectif de réaliser une solution BI qui permet d'analyser l'action de l'import et l'export des produits alimentaires de la société fictive Northwind Traders.

Les données sont stockées dans deux bases de données relationnelles différentes, la base Northwind sous Ms SQL Server et les fichiers txt.

On doit utiliser :

- SQL Server pour l'alimentation.
- Talend Open Studio pour l'intégration de données.
- Tableau pour la création de tableau de bord.

Réalisation :

1. La création de la base de donnée **Northwind** dans SQL Server et la création de la DWH par rapport a la requête

« Nombre de commandes livrées et nombre de commandes non encore livrées par client (ID client, Nom client, Ville, Pays), Employé (ID employé, nom et prénom, ville, pays), Mois et Année. »

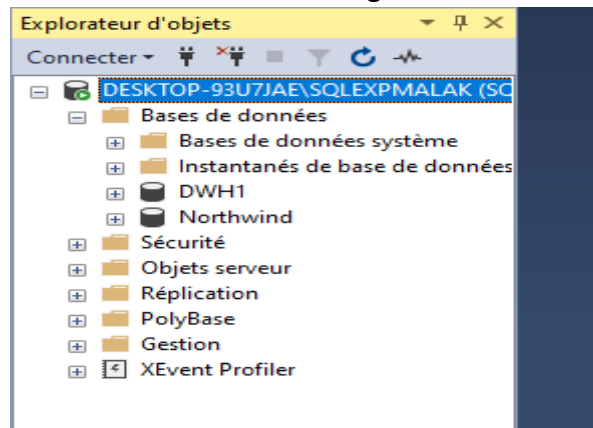
La requête

```
select c.CustomerID as client,c.ContactName, c.CompanyName,c.City,c.Country,
e.EmployeeID,e.FirstName, e.LastName,e.City as cityemp,e.Country as Countryemp,
FORMAT(o.OrderDate,'MM') as ordermonth,FORMAT(o.OrderDate,'yyyy') as orderyear,
count( case when o.ShippedDate IS NULL then 1 else null end ) as
nbr_commande_nonlivrees,
count( case when o.ShippedDate IS not NULL then 1 else null end ) as
nbr_commande_livrees
from Customers c, Employees e, Orders o
where c.CustomerID=o.CustomerID
and o.EmployeeID=e.EmployeeID
Group by c.CustomerID, c.CompanyName,c.ContactName,c.City,c.Country,
e.EmployeeID,e.FirstName,
e.LastName,e.City,e.Country,FORMAT(o.OrderDate,'MM')
,FORMAT(o.OrderDate,'yyyy') ;
```

résultat:

	client	ContactName	CompanyName	City	Country	EmployeeID	FirstName	LastName	cityemp	Countryemp	ordermonth	orderyear	nbr_commande_nonlivrees	nbr_commande_livrees
1	ALFKI	Maria Anders	Alfreds Futterkiste	Berlin	Germany	1	Nancy	Davolio	Seattle	USA	01	1998	0	1
2	ALFKI	Maria Anders	Alfreds Futterkiste	Berlin	Germany	1	Nancy	Davolio	Seattle	USA	03	1998	0	1
3	ALFKI	Maria Anders	Alfreds Futterkiste	Berlin	Germany	3	Janet	Leverling	Kirkland	USA	04	1998	0	1
4	ALFKI	Maria Anders	Alfreds Futterkiste	Berlin	Germany	4	Margaret	Peacock	Redmond	USA	10	1997	0	2
5	ALFKI	Maria Anders	Alfreds Futterkiste	Berlin	Germany	6	Michael	Suyama	London	UK	08	1997	0	1
6	ANATR	Ana Trujillo	Ana Trujillo Emparedados y helados	México D.F.	Mexico	3	Janet	Leverling	Kirkland	USA	08	1997	0	1
7	ANATR	Ana Trujillo	Ana Trujillo Emparedados y helados	México D.F.	Mexico	3	Janet	Leverling	Kirkland	USA	11	1997	0	1
8	ANATR	Ana Trujillo	Ana Trujillo Emparedados y helados	México D.F.	Mexico	4	Margaret	Peacock	Redmond	USA	03	1998	0	1
9	ANATR	Ana Trujillo	Ana Trujillo Emparedados y helados	México D.F.	Mexico	7	Robert	King	London	UK	09	1996	0	1
10	ANTON	Antonio Moreno	Antonio Moreno Taquería	México D.F.	Mexico	1	Nancy	Davolio	Seattle	USA	09	1997	0	1

- L'importation de la base de Northwind : on exécute le script complet de la création de la base Northwind dans SQL server management studio.



✓ La création de DWH :

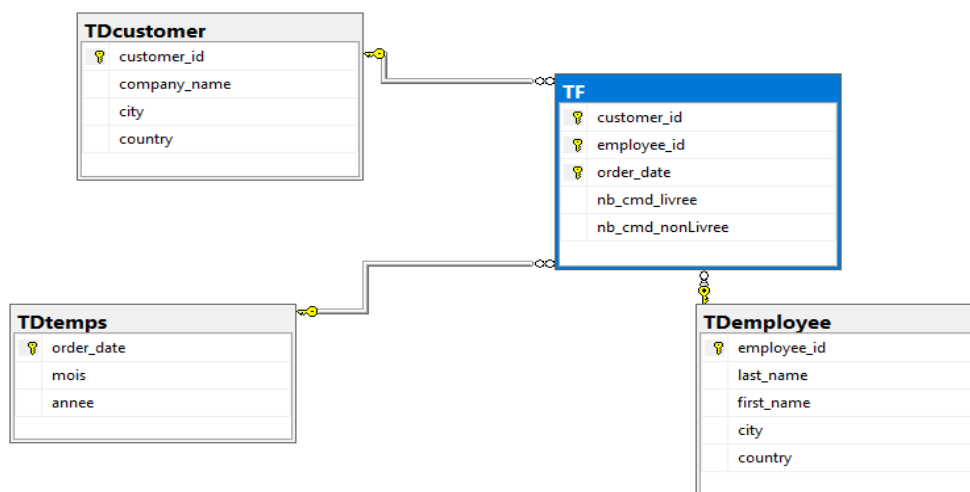
- Table de dimension :

TDcusmter, TDempolyee ,TDtemps

- Table de fait

TF

Notre Datawarehouse



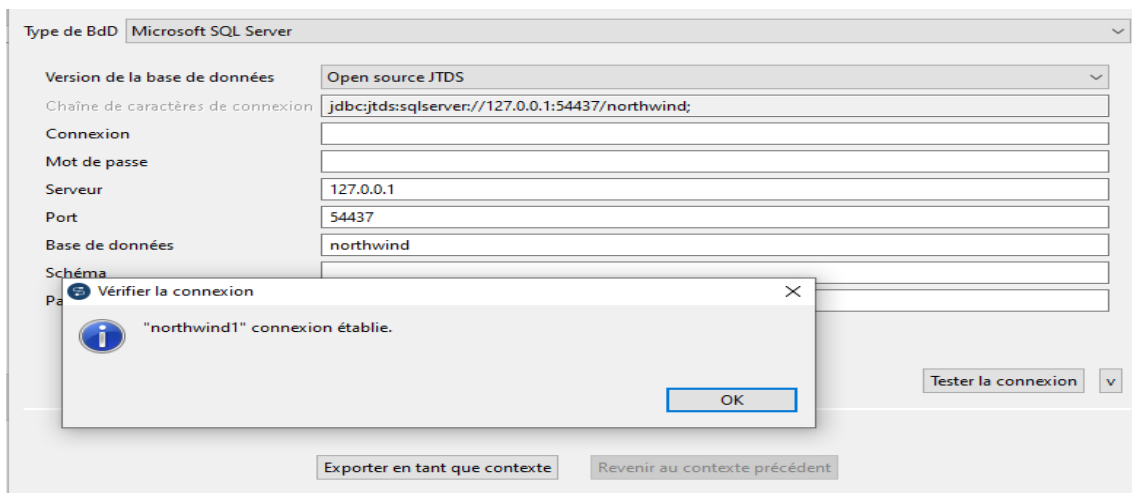
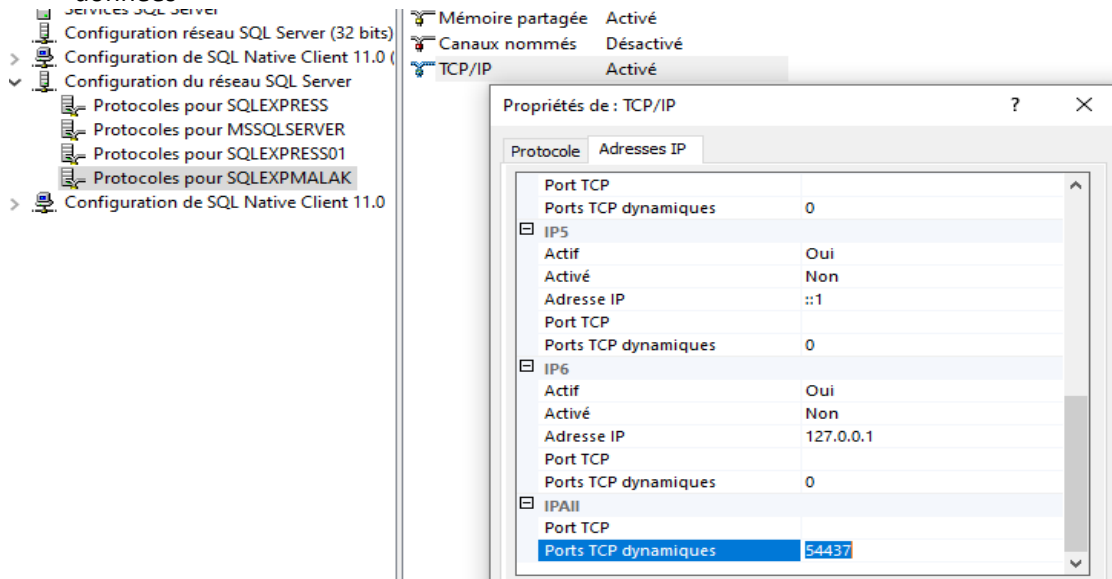
2. La connexion entre Talend et SQLServer :

Après l'ouverture de TOS, la première tâche à effectuer est de connecter notre outil aux différentes bases de données dont on a besoin pour réaliser notre projet BI, à savoir : la bases Northwind sous SQL server et Northwind, les fiches plate (client & commande) et la DWH1 qu'on a créé dans SQL server va représenter notre base de destination (l'entrepôt de donnée).

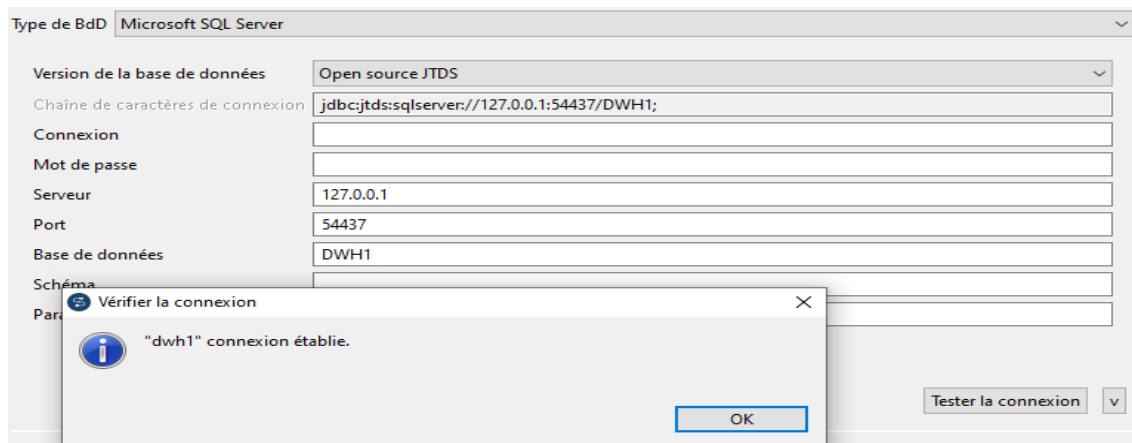
Pour créer une connexion de type base de données SQL Server sur *Talend*, on a besoin de télécharger un *jar* « **jtds-1.3.1.jar** », puis copier l'exécutable qui se trouve dans le répertoire *SSO* et le coller dans le dossier *bin* de *JAVA*, puis on passe à activer le protocole *TCP/IP* dans SQL Server afin de communiquer les données avec *Talend*. Pour créer une connexion.

- **Northwind :**

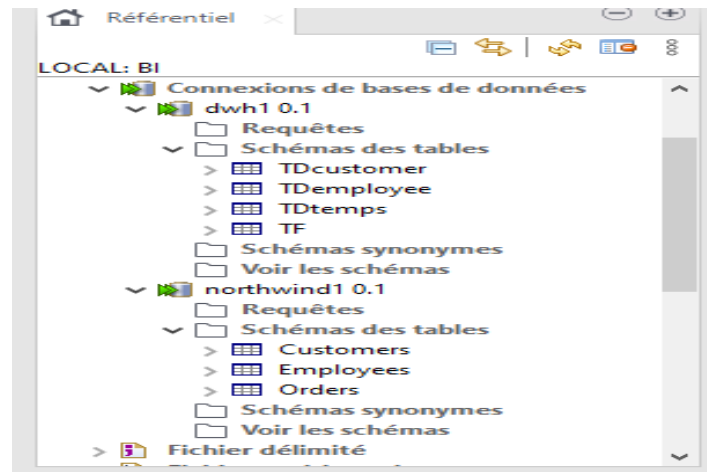
On récupère numéro de port et le nom de serveur et en écrivant le nom de notre base de données



- **DWH**



Après l'établissement de connexion avec la base de données, on pourra récupérer les schémas. Pour aboutir à cela on clique un clic droit sur la connexion et on choisit « récupérer le schéma », on sélectionne seulement les tables quand doit l'utiliser dans notre cas (customers, employees, orders)



- **Fichier plat « client & commande »**

- **Client :**

Fichier - étape 3 de 3

Mettre à jour la métadonnée Fichier dans le référentiel
Définir les paramètres d'analyse du fichier



Paramètres du fichier Encodage: <input type="text" value="US-ASCII"/> Séparateur de champs: <input type="text" value="Custom UTF"/> Custom UTF8 <input type="text" value=";"/> Séparateur de lignes: <input type="text" value="Standard EO"/> Caractère correspondant <input type="text" value="\\n"/>		Lignes à ignorer Si des lignes doivent être ignorées, spécifiez les paramètres suivants En-tête: <input checked="" type="checkbox"/> 1 Pied de page: <input type="checkbox"/> <input type="checkbox"/> Ignorer les lignes vides	
Paramètres du caractère d'échappement <input type="radio"/> CSV <input checked="" type="radio"/> Délimité Caractère d'échappement: <input type="text" value="Vider"/> Entourage du texte: <input type="text" value="Vider"/> <input type="checkbox"/> Scinder la ligne avant le champ		Limite de lignes Si le nombre de lignes doit être limité, spécifiez ce nombre Limite: <input type="text"/>	

Aperçu | Sortie

☒ Définir la ligne d'en-tête comme nom de colonnes

IDClient	wilaya	
MIN	Blida	
ACV	Alger	
ACM	Tipaza	

- **Commande :**



Paramètres du fichier

Encodage UTF-8

Séparateur de champs Custom UTF Custom UTF8 "#"

Séparateur de lignes Standard EO Caractère correspondant "\n"

Paramètres du caractère d'échappement

☐ CSV ☒ Délimité

Caractère d'échappement Vider

Entourage du texte Vider

☐ Scinder la ligne avant le champ

Lignes à ignorer

Si des lignes doivent être ignorées, spécifiez les paramètres suivants

En-tête ☒ 1

Pied de page ☐

☐ Ignorer les lignes vides

Limite de lignes

Si le nombre de lignes doit être limité, spécifiez ce nombre

Limite ☒ 3

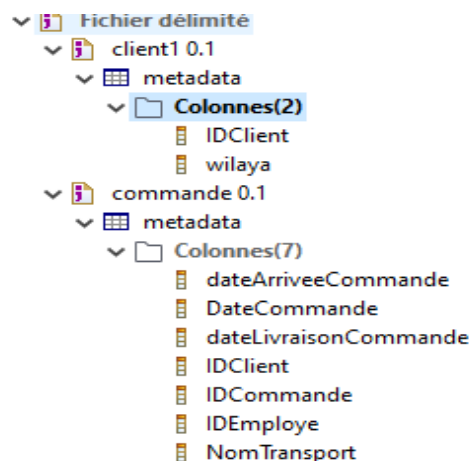
Aperçu Sortie

☒ Définir la ligne d'en-tête comme nom de colonnes Actualiser l'aperçu

IDCommande	IDClient	IDEmployé	DateCommande	dateArrivéeCommande	dateLivraisonCommande	NomTransport
1	MIN	3	2022-01-05	2022-02-04	2022-01-15	TranMohamed
2	ACV	9	2022-02-10	2022-02-20	2022-02-10	TKamel
3	ACM	7	2022-05-15	2022-05-25	2022-05-10	TranHmd

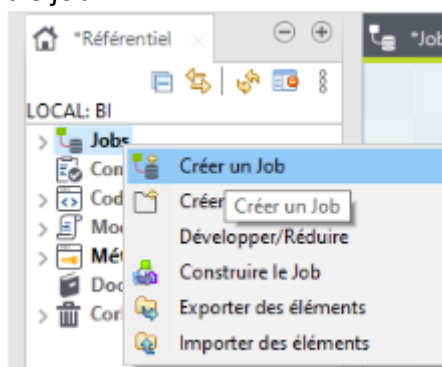
Exporter en tant que contexte Revenir au contexte précédent

- Voici le résultat :

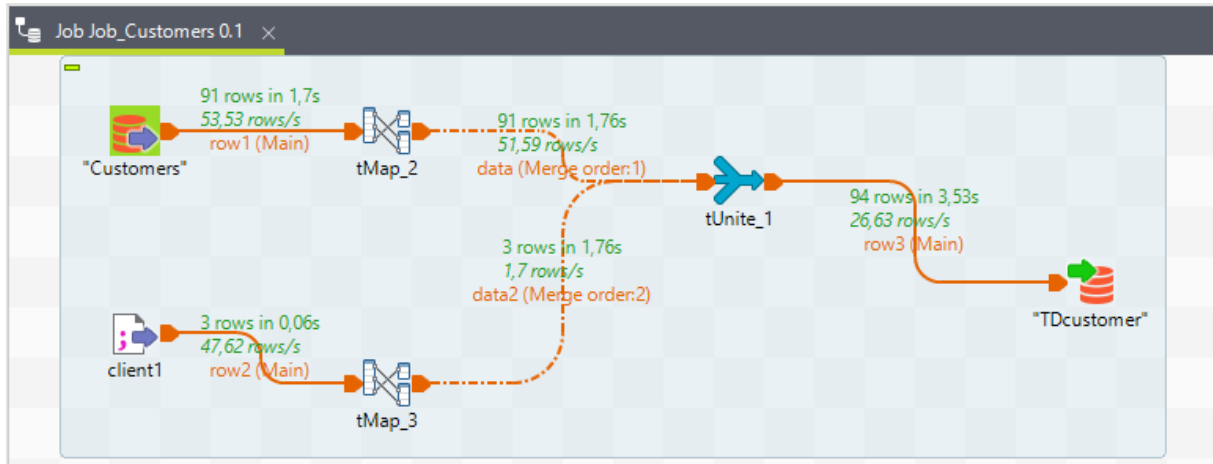


3. L'alimentation des dimensions et la table de fait :

→ La création des Job : on click droit sur JOB après o click sur Créer un job, ensuite donne le un nom pour notre job

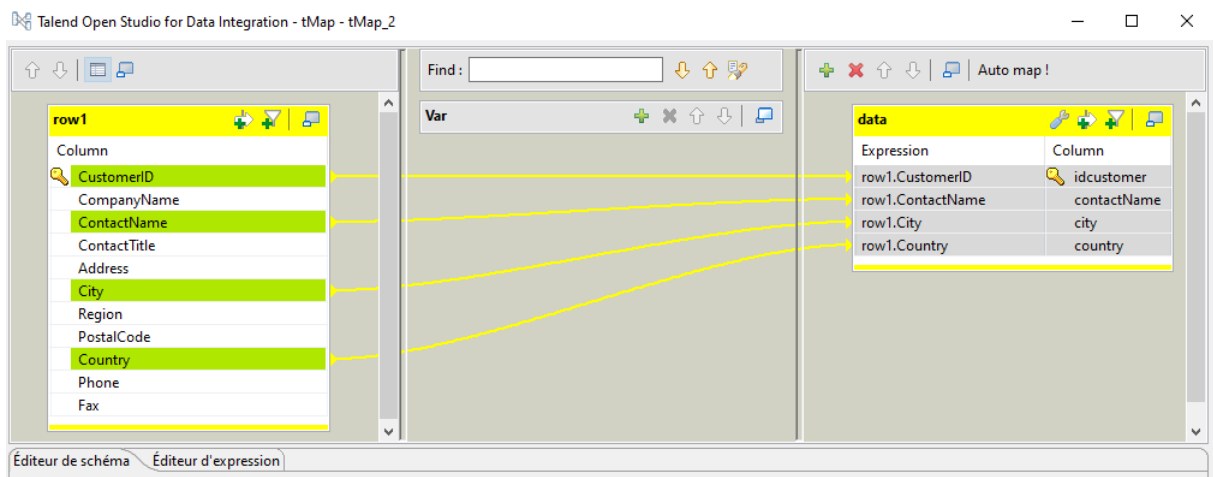


- **TDcustomer** : On crée notre *Job_customers*, on doit utiliser deux sources de donnée la 1^{ère} est la table Customers de Northwind et la 2^{ème} le fichier plat client et on relie chaque donne avec un Tmap pour faire l'extraction et la sélection des attributs nécessaires. On établit une jointure entre les deux sources sortantes des deux tMap dans tUnite. Après on met sa sortie dans la TDcustomer.

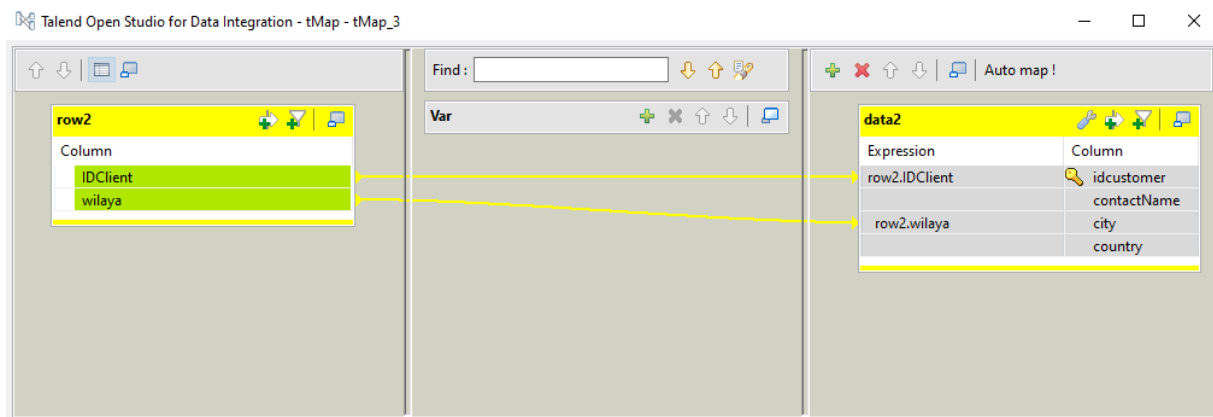


Paramètre des différentes composantes :

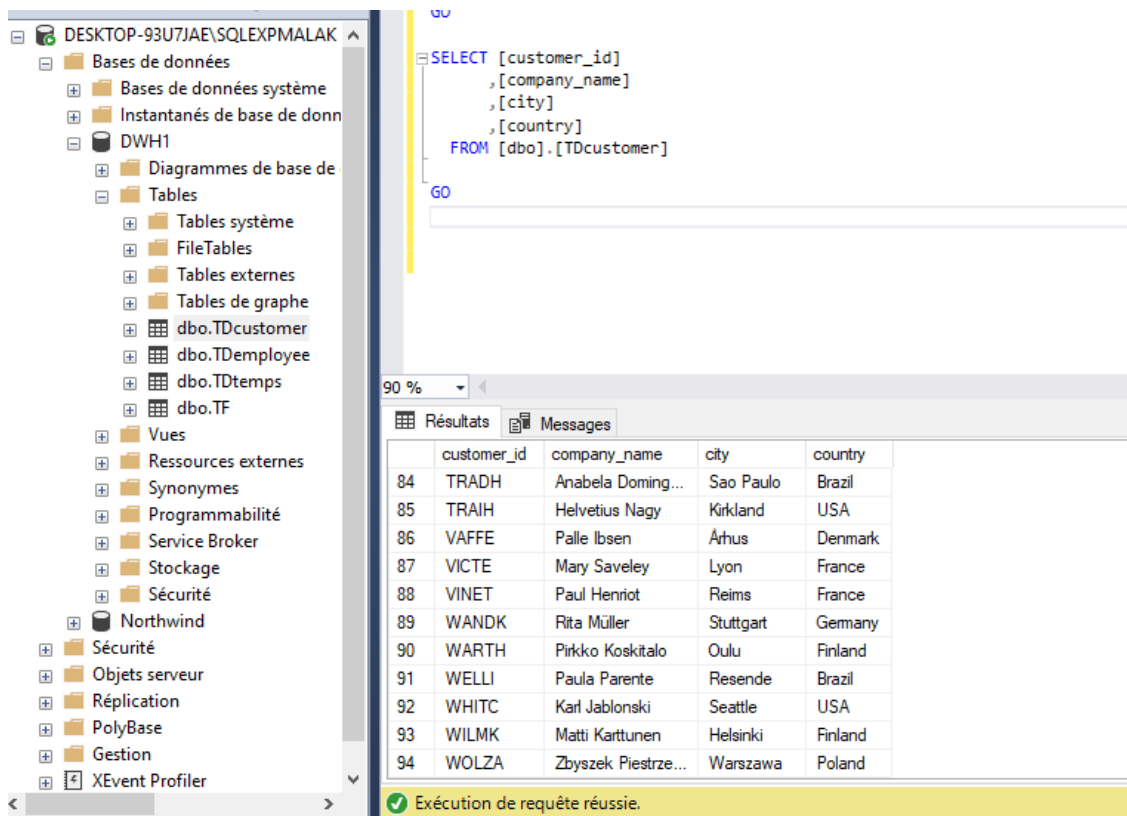
- tMap_2 :



- TMap_3



Vérification du remplissage de la dimension dans SSMS :



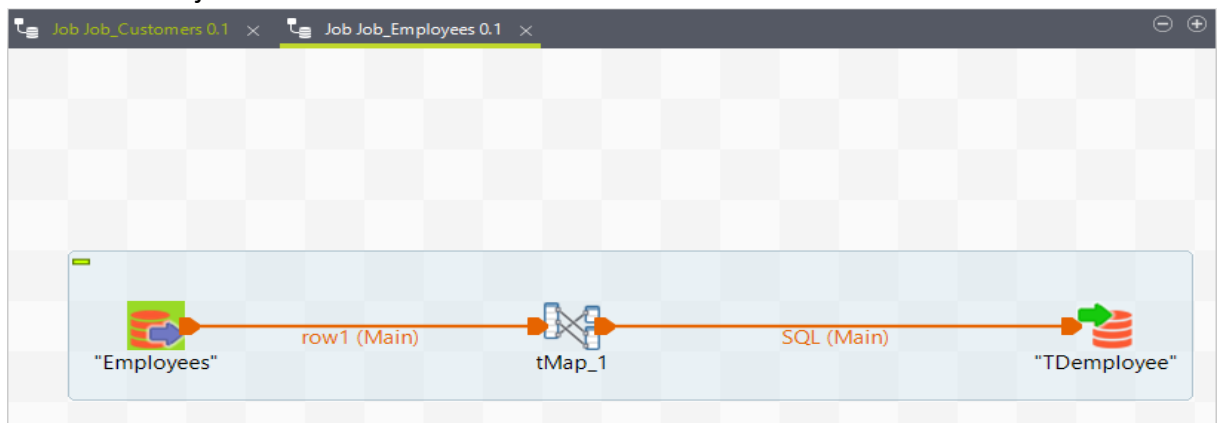
```

SELECT [customer_id]
      ,[company_name]
      ,[city]
      ,[country]
FROM [dbo].[TDcustomer]
GO
  
```

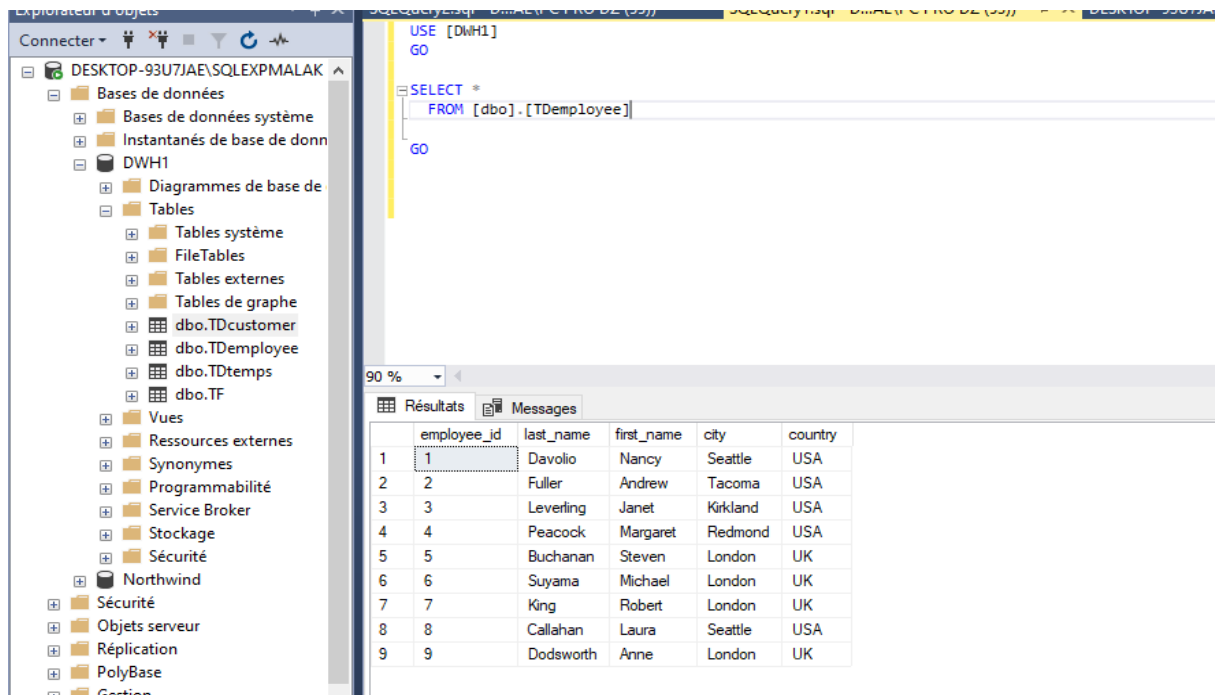
customer_id	company_name	city	country
84	TRADH	Anabela Doming...	Sao Paulo
85	TRAH	Helvetius Nagy	Kirkland
86	VAFFE	Palle Ibsen	Arhus
87	VICTE	Mary Saveley	Lyon
88	VINET	Paul Henriot	Reims
89	WANDK	Rita Müller	Stuttgart
90	WARTH	Pirkko Koskitalo	Oulu
91	WELLI	Paula Parente	Resende
92	WHITC	Karl Jablonski	Seattle
93	WILMK	Matti Karttunen	Helsinki
94	WOLZA	Zbyszek Piestrze...	Warszawa

Exécution de requête réussie.

- **TDemployee** : on crée le Job_employees après on relie les données source de Northwind 'la table employees' avec un Tmap pour faire l'extraction et la sélection des attributs nécessaires. Apres on le relie avec TDemployee de notre DWH 'sortant'. On exécute notre job.

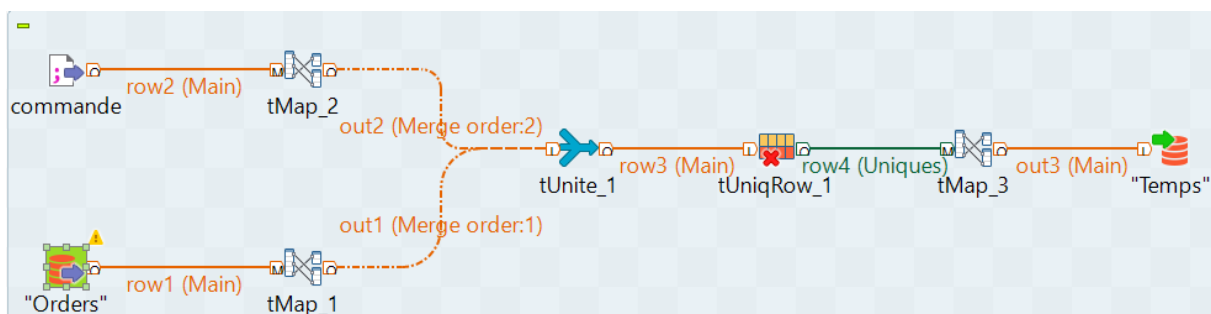


Vérification du remplissage de la dimension dans SSMS :

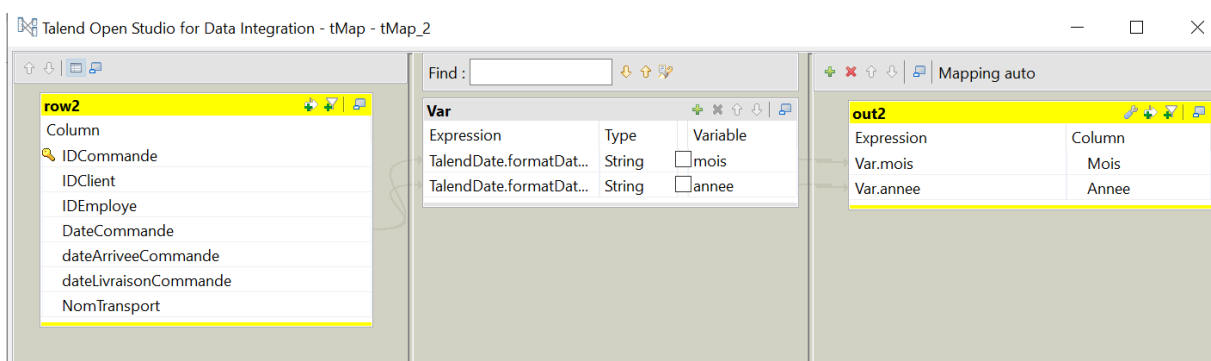


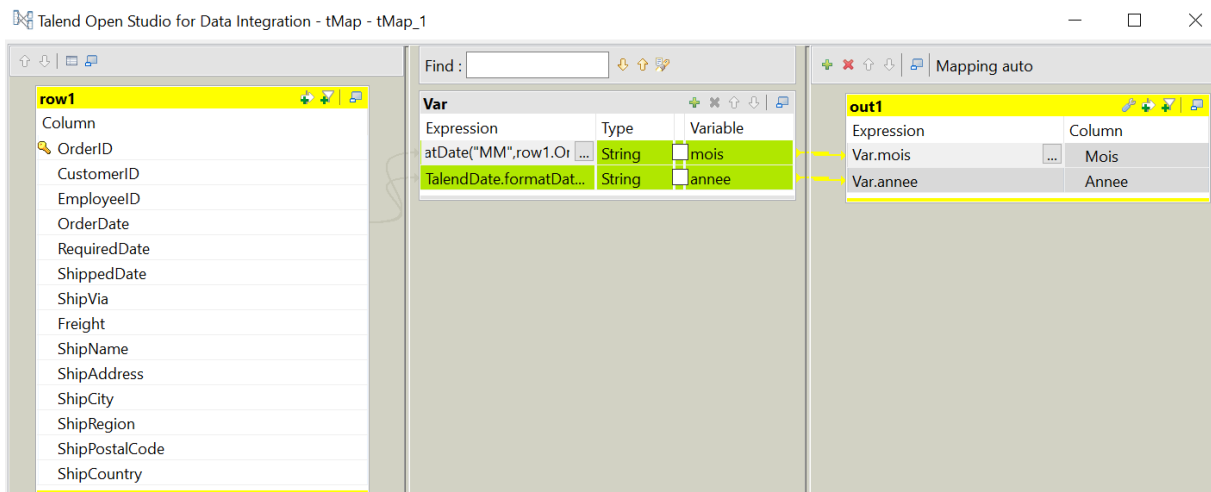
- **Tdtemps** : on crée notre job_temps ensuite :

Pour tUnit 1 : on va récupérer les dates de commande livre de 2 sources de données (ordres de Northwind et le fichier commande) et on relie les sources avec un Tmap chacune afin de faire l'extraction et la sélection des attributs nécessaires. Après on relie la sortie de TUnit1 avec la composante appelée « TuniqRow » pour supprimer les redondances.

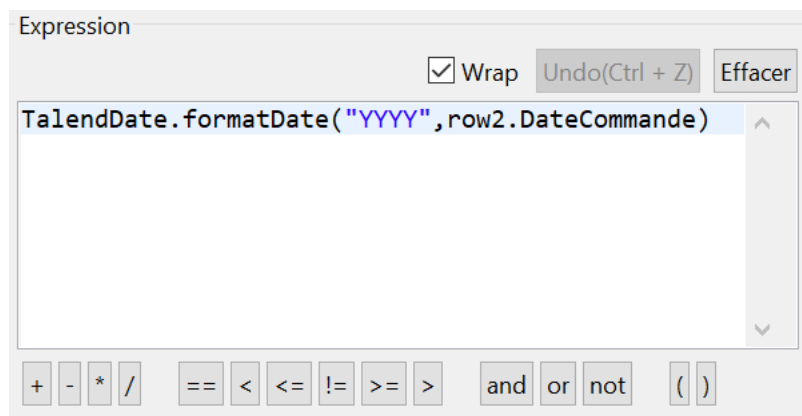


Paramètre des différentes composantes :

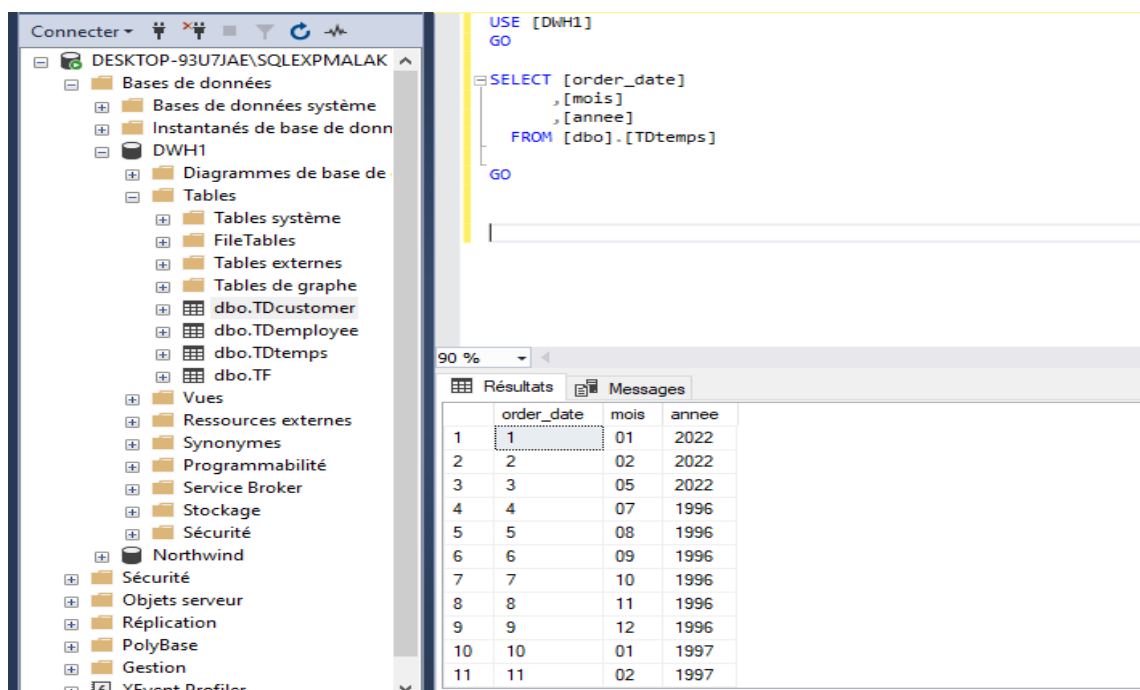




- Exemple d'extraction de l'année apartir de la table Orders :



Vérification du remplissage de la dimension dans SSMS :



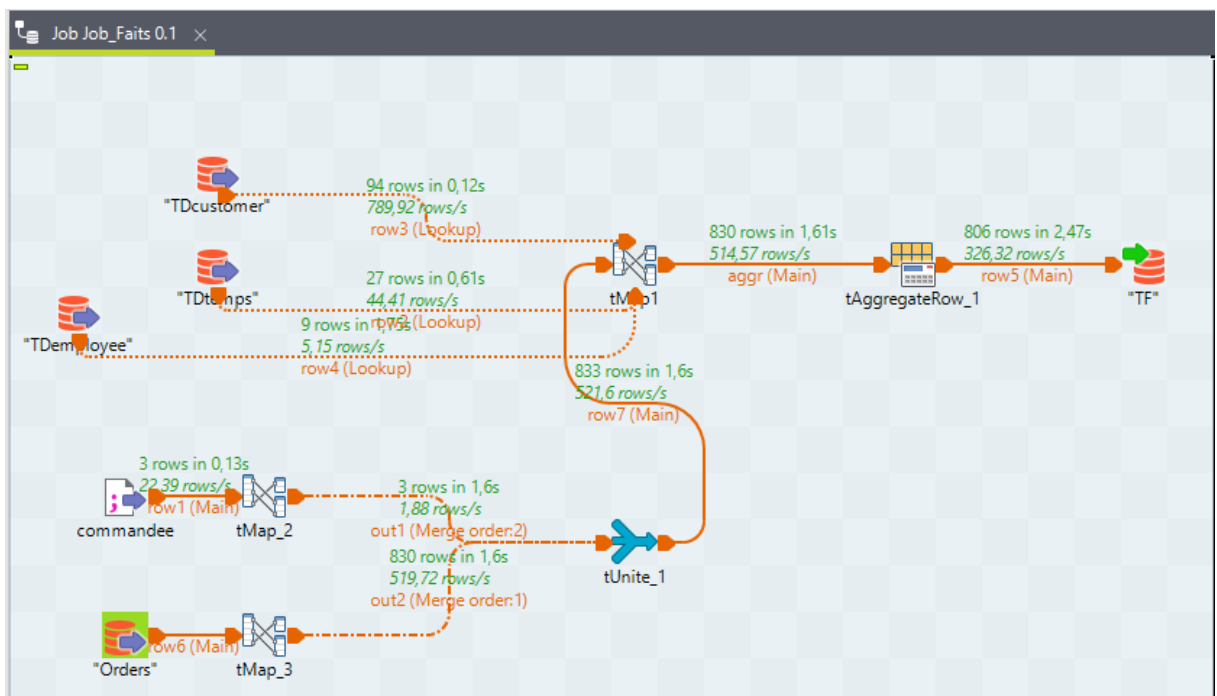
- **TF** : on crée notre job_Fait ensuite :

Pour tUnit2 : on doit récupérer les *dates de commande livre* et les *data de commande* de la source de donnée commande et on relie la source avec un Tmap2.

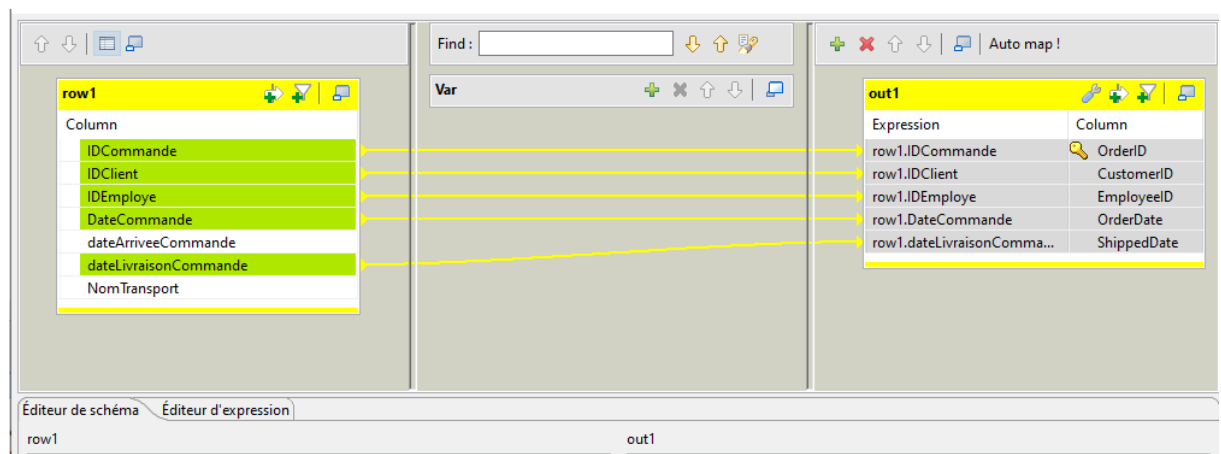
On doit récupérer les *orders dats* et les *shpperdate* de la source de donnée *Orders* de *Northwind* et on relie la source avec un Tmap3 afin de faire l'extraction et la sélection des attributs nécessaires. On établit une jointure entre les deux sorties de tMap2 et tMap3 dans tUnit1.

Après on relie notre dimension (TDCustomer, TDEmployee, TDtemps) et la jointure tUnit1 avec tMap1, afin de faire l'extraction et la sélection des attributs nécessaires.

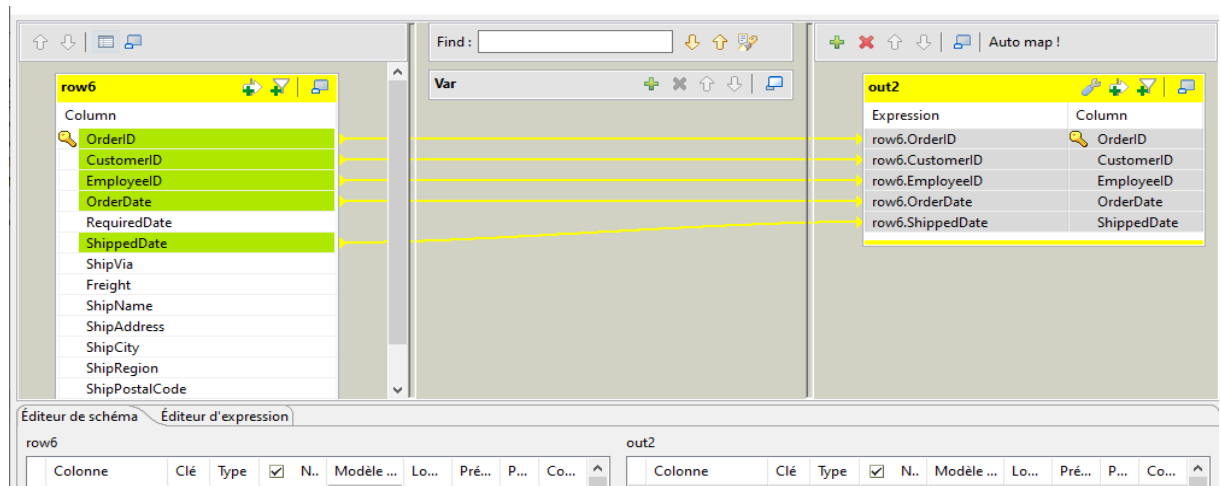
Après on relie la sortie de tMap avec la composante appelée «tAggregateRow» pour faire l'agrégation. À la fin on met sa sortie dans la table de fait TF.



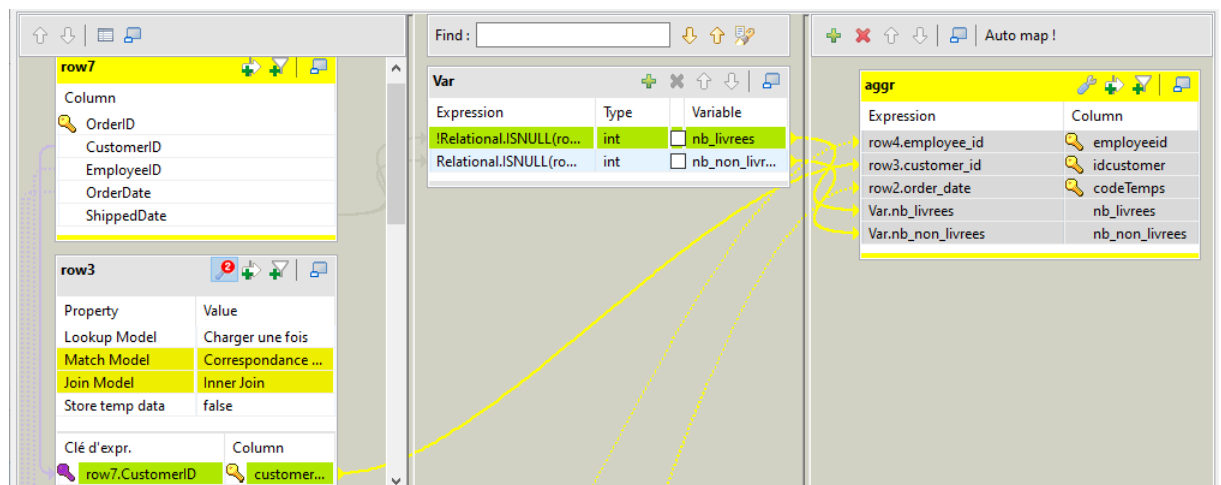
- TMap_2 :



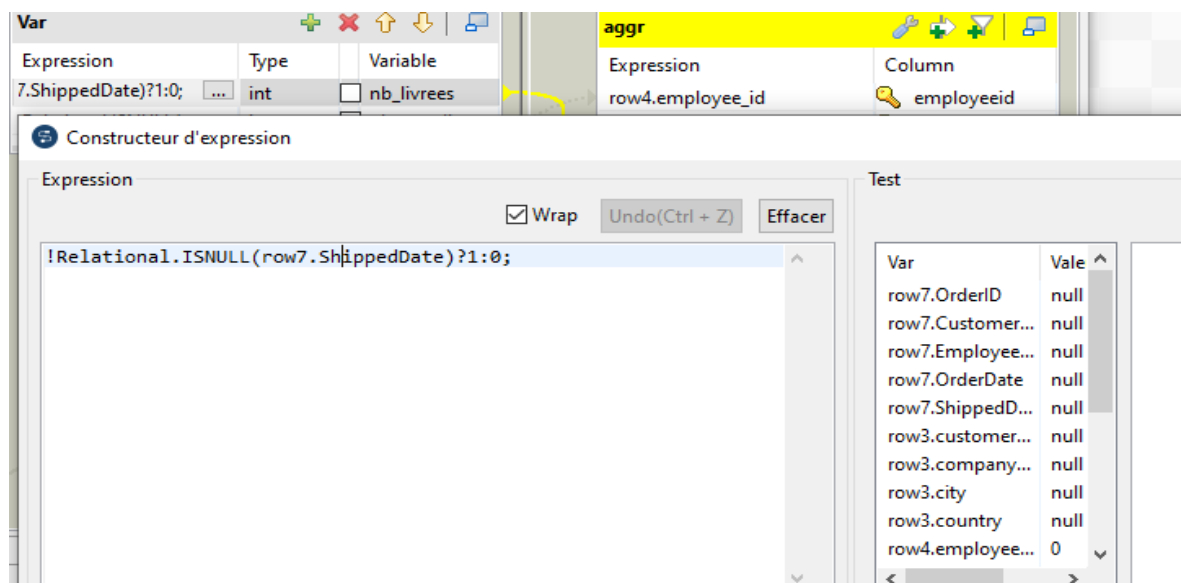
- TMap_3 :



- TMap_1 :



Expression pour nb livrées :



Expression pour nb non livrees :

The screenshot shows the 'Constructeur d'expression' (Expression Builder) window in SQL Server Data Tools. The 'Var' pane on the left lists variables: 'nb_livrees' and 'nb_non_livrees'. The 'Expression' pane in the center shows the expression 'Relational.ISNULL(row7.ShippedDate)?1:0;'. The 'Test' pane on the right shows a list of variables: 'row7.customer_id', 'row3.customer_id', 'row7.employee_id', 'row7.order_date', 'row7.customer_id', 'row3.customer_id', and 'row3.order_date'.

Composant tAggregateRow : on a utilisé la somme dans l'opération de group by pour sommer les 1 qu'on a affectées.

The screenshot shows the 'tAggregateRow_1' component configuration in SQL Server Data Tools. The 'Paramètres simples' (Simple Parameters) tab is selected. The 'Group by' section shows columns 'employee_id', 'customer_id', and 'order_date'. The 'Opérations' (Operations) section shows two rows: 'nb_cmd_livree' with 'somme' (sum) and 'nb_cmd_nonLivree' with 'somme' (sum).

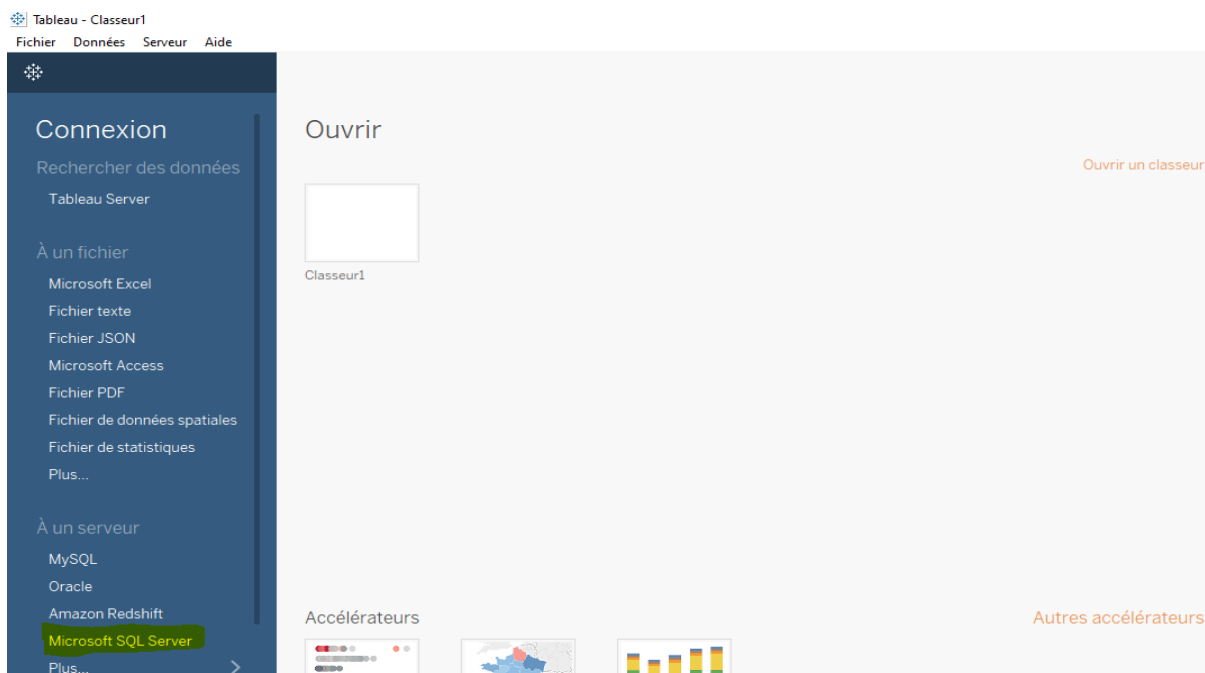
Vérification du remplissage de la dimension dans SSMS :

The screenshot shows the 'Northwind' database in SQL Server Enterprise Manager (SSMS). The 'Tables' folder is expanded, showing 'dbo.TDcustomer', 'dbo.TDemployee', 'dbo.TDtemps', and 'dbo.TF'. The 'Results' pane shows a table with columns: customer_id, employee_id, order_date, nb_cmd_livree, and nb_cmd_nonLivree. The table contains 10 rows of data.

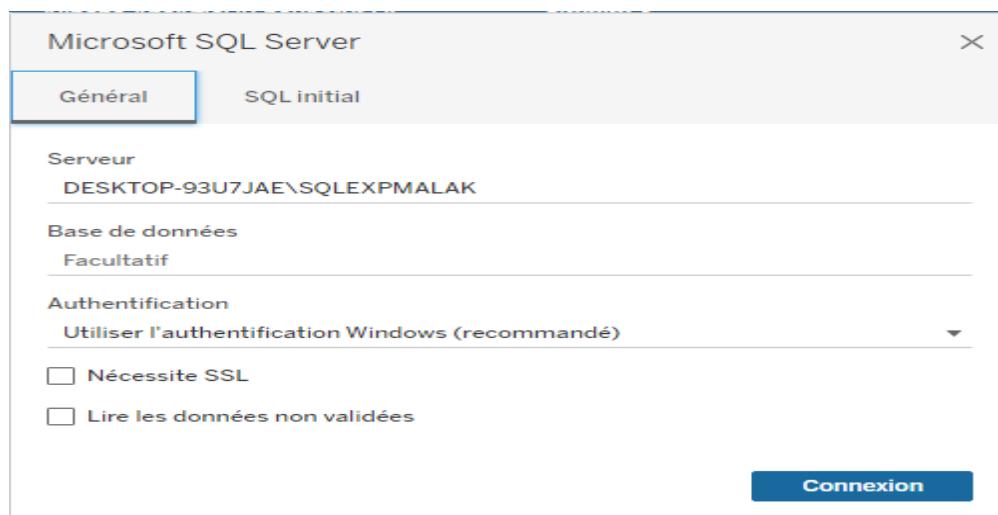
customer_id	employee_id	order_date	nb_cmd_livree	nb_cmd_nonLivree	
799	WILMK	2	25	1	0
800	WILMK	3	23	1	0
801	WILMK	4	23	1	0
802	WILMK	7	19	1	0
803	WOLZA	1	9	1	0
804	WOLZA	1	21	1	0
805	WOLZA	4	23	1	0
806	WOLZA	4	25	1	0
807	WOLZA	5	23	1	0
808	WOLZA	6	16	1	0
809	WOLZA	8	25	1	0

4. Connexion Tableau avec SQL serveur

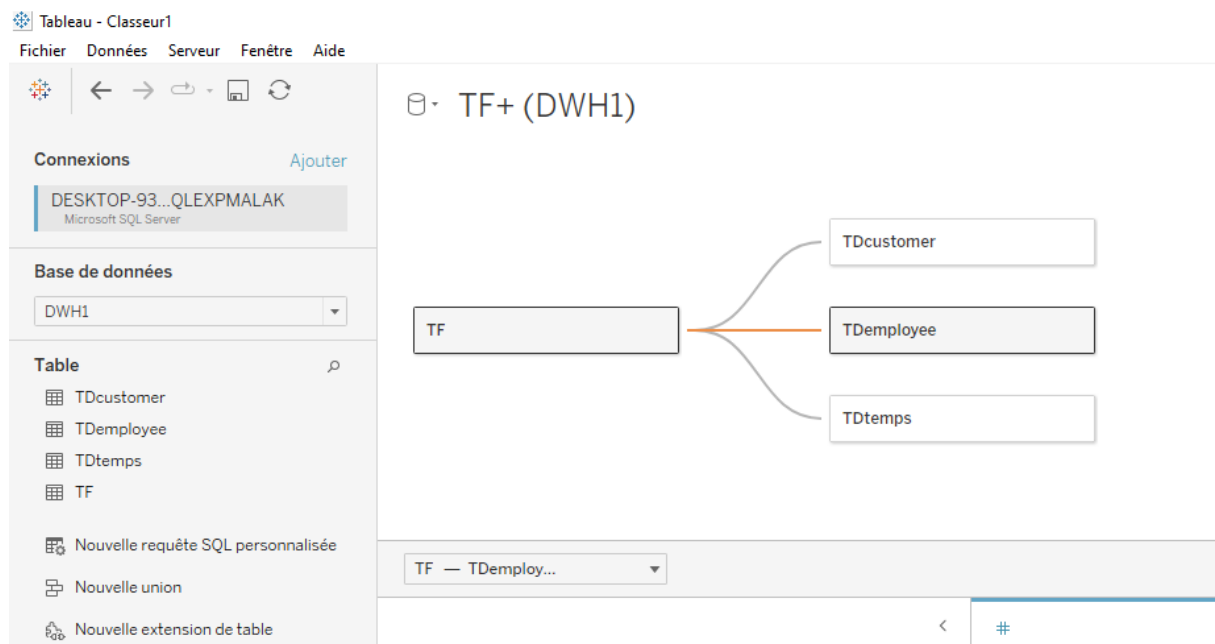
1. On ouvre Tableau après on clique sur CONNEXION Microsoft SQL.



Une fenêtre apparaît indiquant qu'on doit remplir le nom de notre SQL serveur après pour récupérer les bases de données.

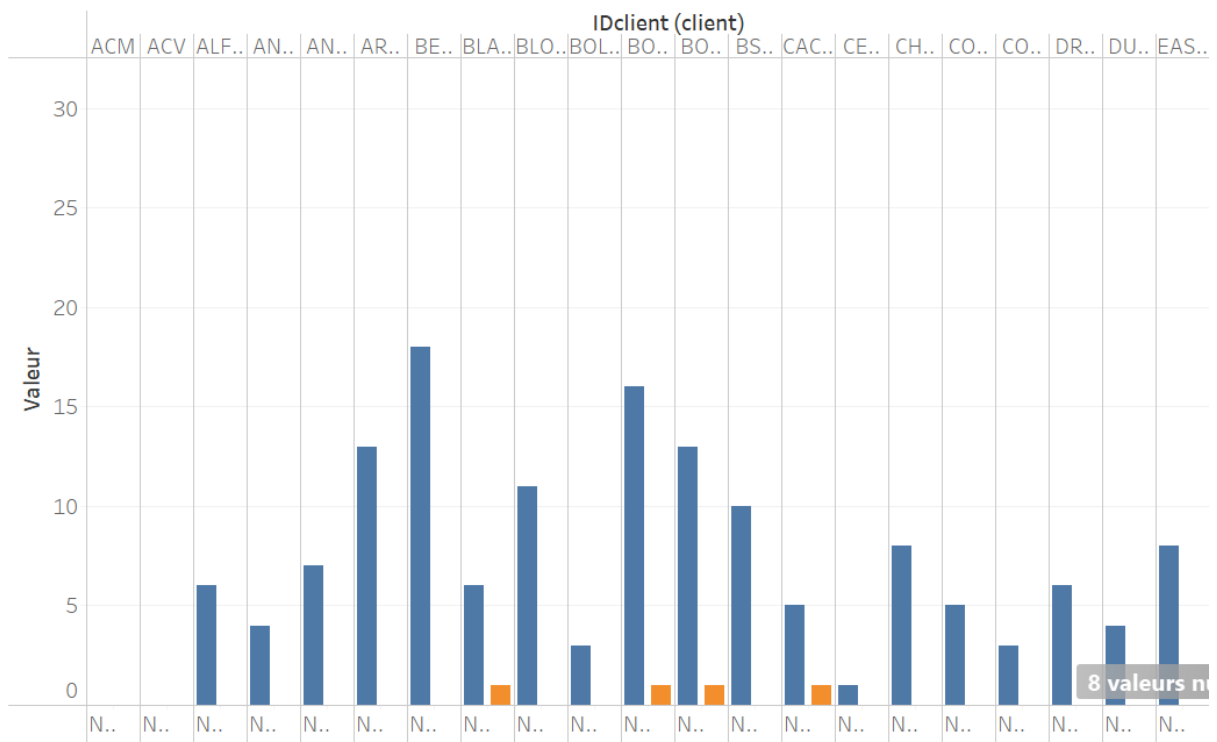


2. On glisse les 4 tables de notre DataWarehouse pour les importer

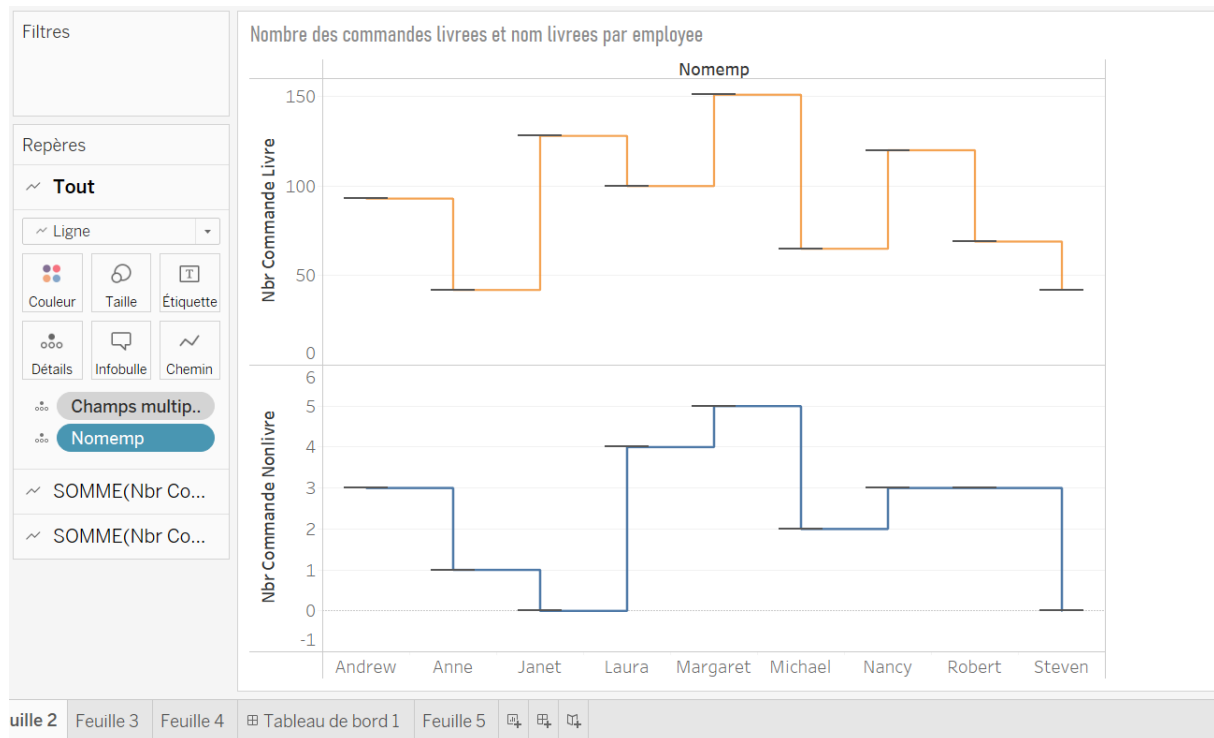


- **Feuille1** : nombre de commande livrée et non livrée pour chaque client

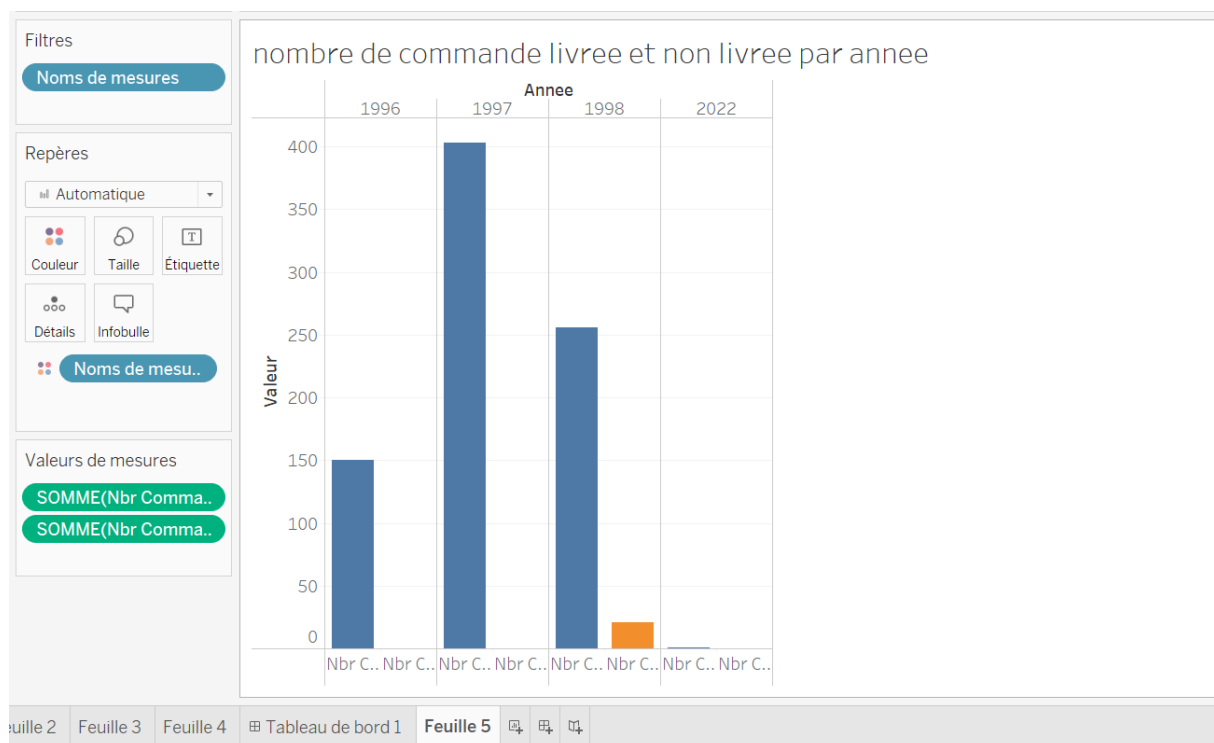
nombre de commande livree et non livree pour chaque client



- **Feuille2** : Nombre des commandes livrées et nom livrées par employée



- **Feuille3** : Nombre de commande livrée et non livrée par année



A la fin on crée un tableau de bord et on ajoute les feuilles qu’on a créé.

