

PROJET BUSINESS INTELLIGENCE

L'analyse de la criminalité à USA

Réalisé par :

QASBAJI Zakaria

EZZAHRAOUI Haroun

Encadré par :

Mr. MOUTACHAOUIK HICHAM

Intelligence artificielle & Génie informatique CI-2

2020/2021

L'analyse de la criminalité à USA

Base de données :

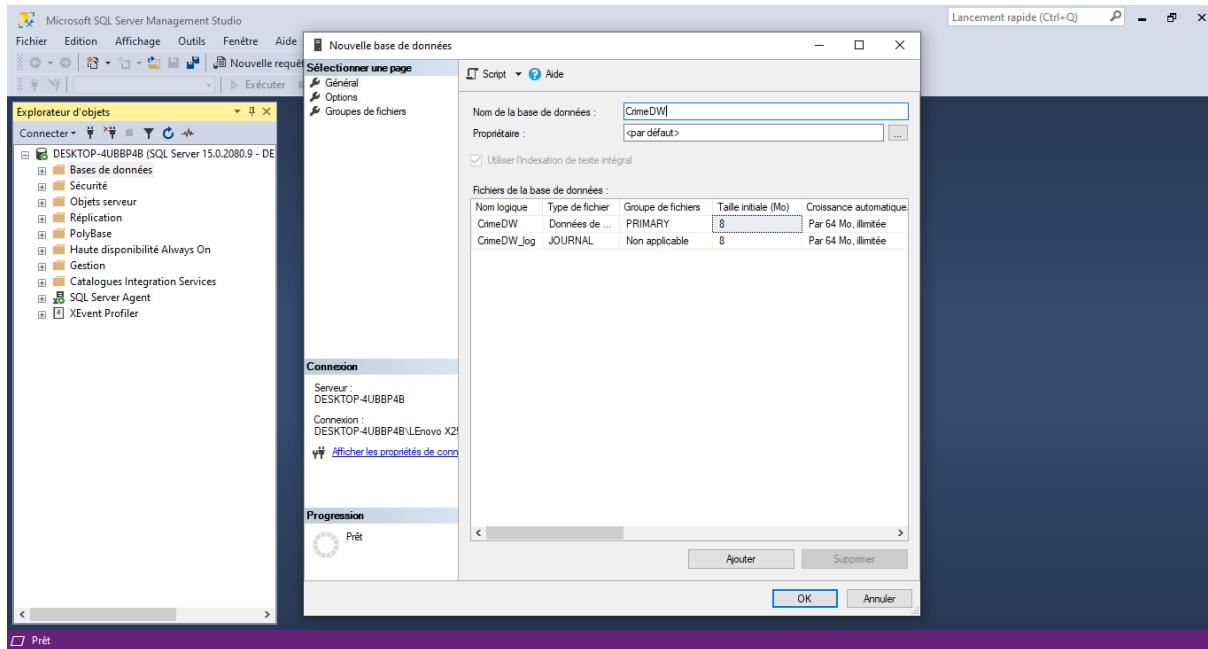
On a récupère une base de données à partir du site web Data.gov qui est le portail open data du gouvernement fédéral américain.

Note base de données prend la forme d'un fichier Csv qui contient des données sur les différents crimes commis en USA durant la période 2006-2008

A	B	C	D	E	F	G	H	I	J	K	L
RowID	State	City	Year	Population	ViolentCrime	MurderAndNonEglicitManslaughter	ForcibleRape	Robbery	AggravatedAssault	PropertyCrime	Burglary
1	115856 Alaska	Anchorage	2007	284142	2405	22	257	453	1673	11107	145
2	115857 Alaska	Bethel	2007	6488	61	2	15	0	44	107	2
3	115858 Alaska	Bristol Bay Borough	2007	1028	8	0	0	0	8	49	1
5	115859 Alaska	Cordova	2007	2322	9	0	0	0	9	14	
6	115860 Alaska	Craig	2007	1186	41	0	0	0	41	18	
7	115861 Alaska	Dillingham	2007	2494	76	1	24	3	48	78	1
8	115862 Alaska	Fairbanks	2007	31287	258	5	43	42	168	1358	23
9	115863 Alaska	Haines	2007	2260	16	0	1	0	15	74	2
10	115864 Alaska	Homer	2007	5629	52	0	2	1	49	220	3
11	115865 Alaska	Houston	2007	1992	3	0	0	0	3	40	
12	115866 Alaska	Juneau	2007	30746	126	1	22	17	86	1364	17
13	115867 Alaska	Kenai	2007	7620	30	1	2	3	24	318	3
14	115868 Alaska	Ketchikan	2007	7384	21	0	9	1	11	476	4
15	115869 Alaska	Kodiak	2007	6242	33	0	5	1	27	209	1
16	115870 Alaska	North Pole	2007	1869	9	0	1	1	7	116	2
17	115871 Alaska	North Slope Borough	2007	6569	70	0	15	7	48	149	5
18	115872 Alaska	Palmer	2007	7931	65	0	4	1	60	282	2
19	115873 Alaska	Petersburg	2007	2890	1	0	0	0	1	117	
20	115874 Alaska	Seward	2007	3054	3	0	1	0	2	146	2
21	115875 Alaska	Sitka	2007	8932	31	0	9	0	22	290	2
22	115876 Alaska	Skagway	2007	827	1	0	0	0	1	14	
23	115877 Alaska	Soldotna	2007	4198	21	0	2	1	18	218	1
24	115878 Alaska	St. Paul	2007	431	6	0	2	0	4	7	
25	115879 Alaska	Unalaska	2007	4155	15	0	4	1	10	74	
26	115880 Alaska	Wasilla	2007	9931	47	0	2	2	45	553	5
27	115881 Alaska	Wrangell	2007	2028	1	0	0	0	1	77	

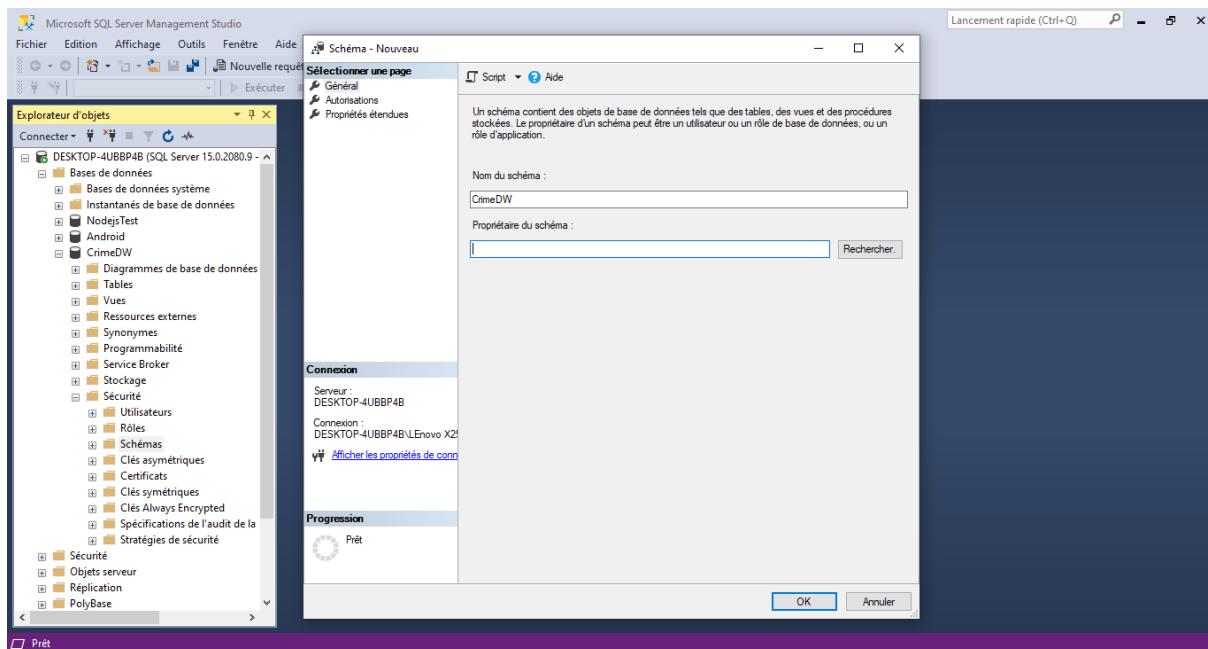
Création de CrimeDW :

Clique droite sur base de données > nouvelle base de données :



Ensute on crée le schéma CrimeDw qui va contenir les objets de la base de données :

Clique droite sur schémas > nouvelle schéma :



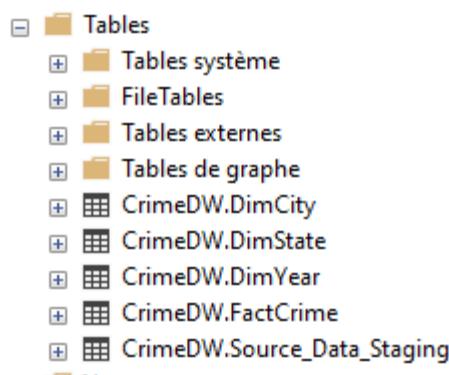
Ensute on crée les tables de notre data Warehouse :

- La table Source_Data_Staging comme son nom l'indique cette table va être utilisée pour préparer les données (elle ne constitue pas la structure de notre data Warehouse)

Pour créer cette table exécuter le script CrimeDWSourceDataStaging.sql

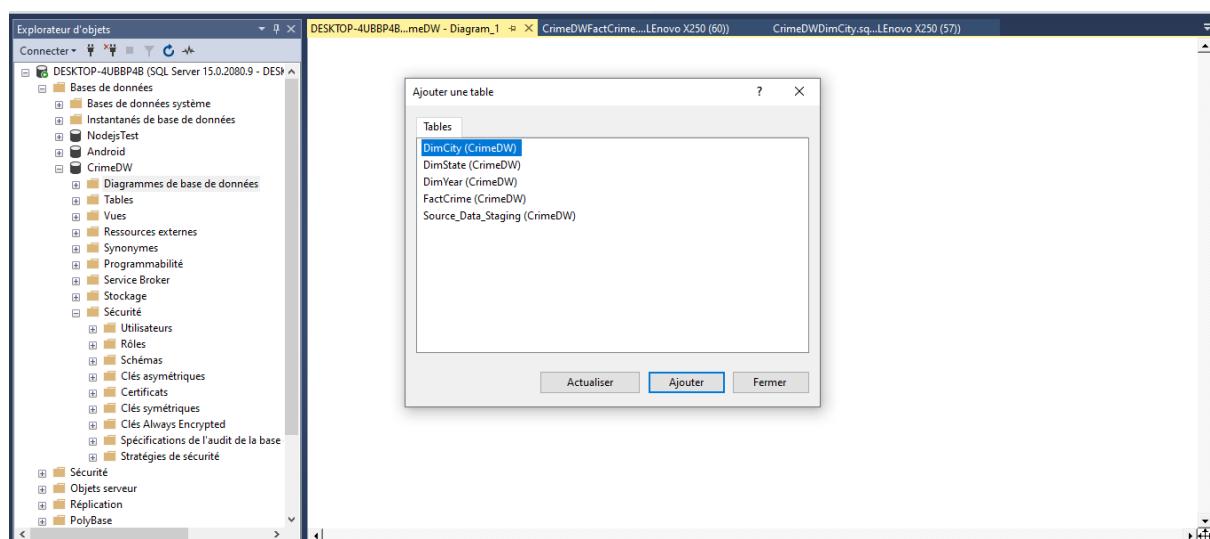
- La table DimYear : on exécute le script CrimeDWDimYear.sql
- La table DimState : on exécute le script CrimeDWDimState.sql
- La table DimCity : on execute le script CrimeDWDimCity.sql
- La table FactCime : on execute le script CrimeDWFactCrime.sql

Les tables sont créées correctement :

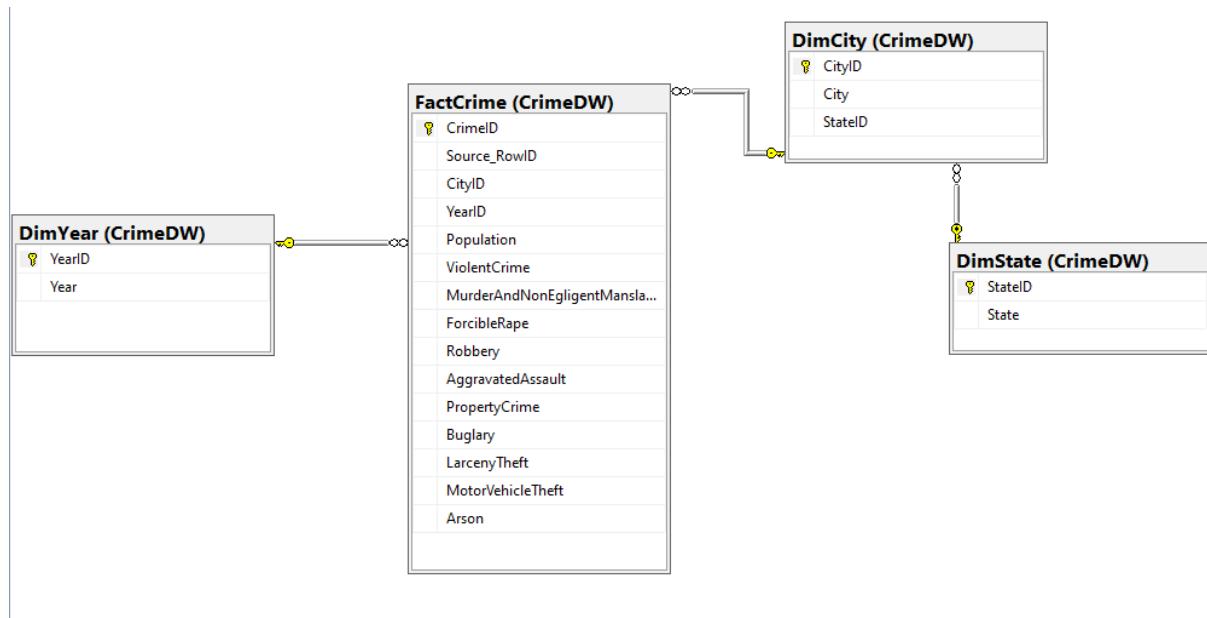


Finalement on crée le diagramme de base de données :

Clique droite de base de données > nouveau schéma de base de données



On ajoute les tables DimCity, DimState, DimYear et FactCrime, on trouve le diagramme suivant :

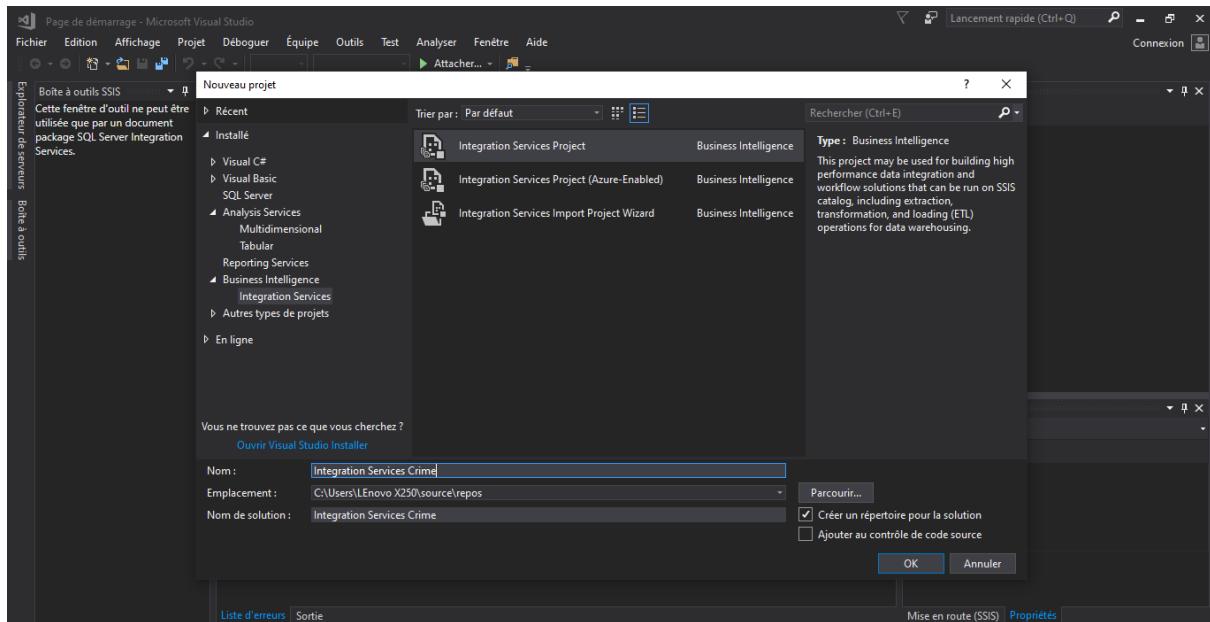


NB : les clés primaires de chaque table vont être générées automatiquement car notre source de données ne contient pas ces champs.

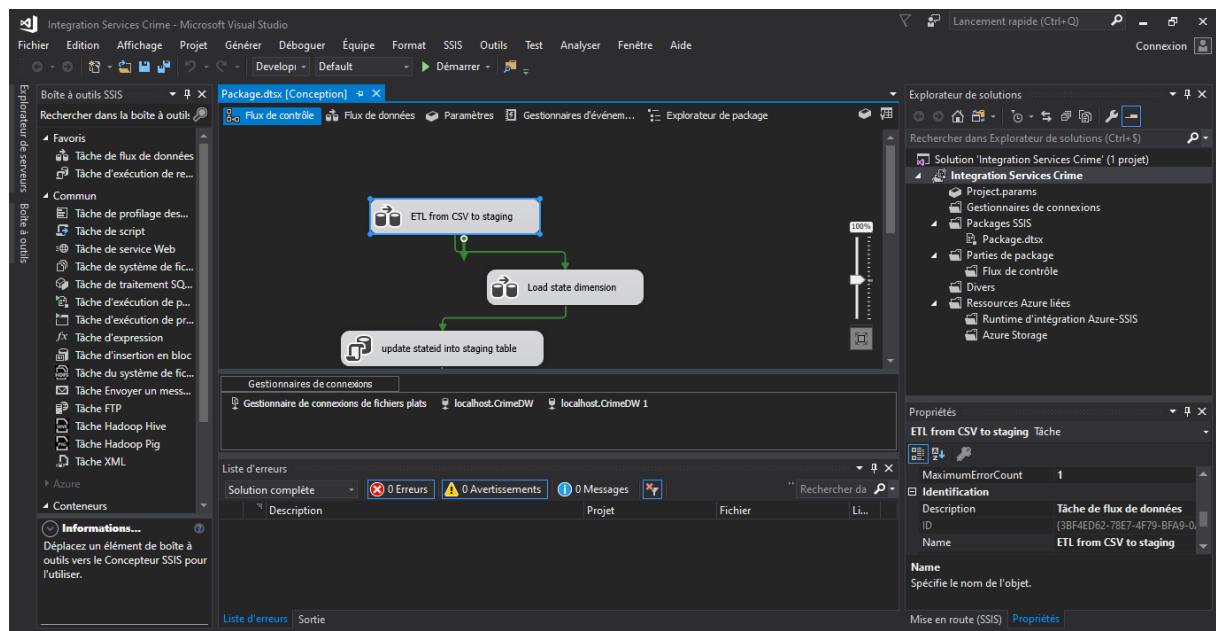
ETL :

On va créer notre projet SSIS nommé Integration Services Crime avec Visual Studio (SSDT) :

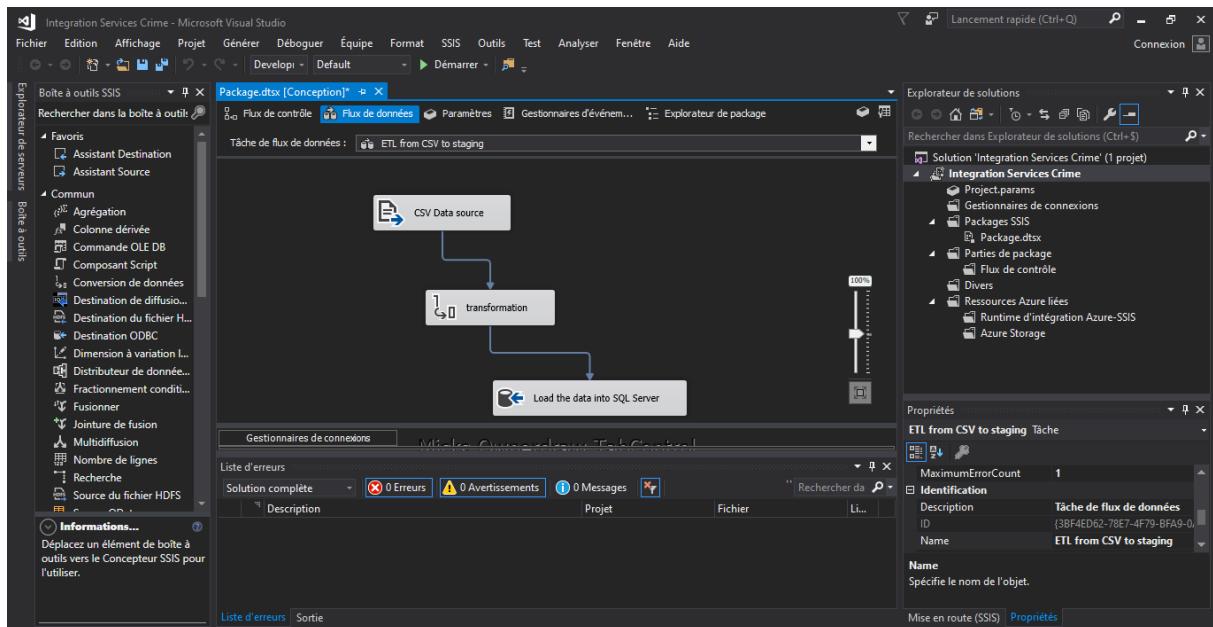
Fichier > nouveau > projet :



Ensute on crée dans le flux de contrôle une tâche de flux de données nommée ETL from CSV to staging :

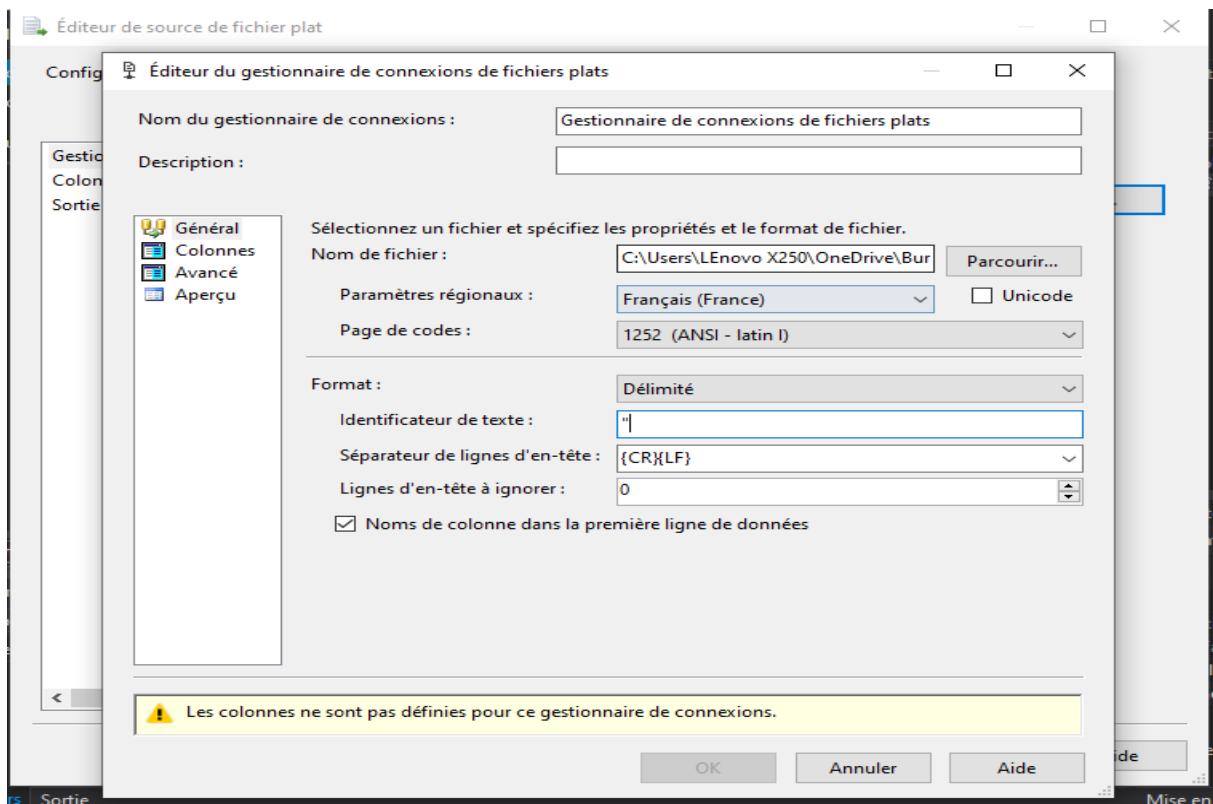


A l'intérieur de cette tâche, on crée les flux de données : source de fichier plat, conversion de données, et destination OLE DB :

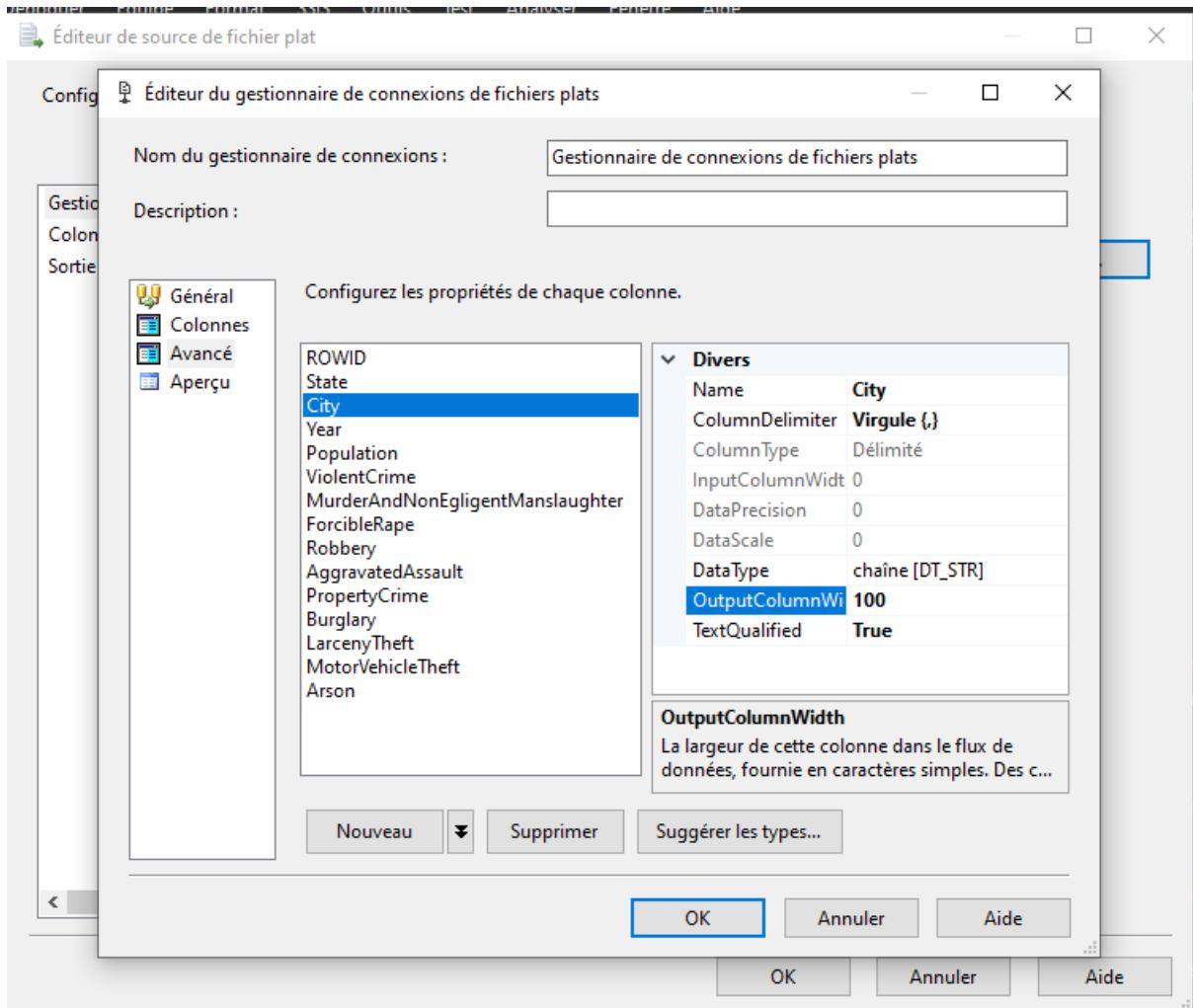


-Dans la source de fichier plat on crée un nouveau gestionnaire de connexion avec les fichiers plats, et on le configure de la manière suivante :

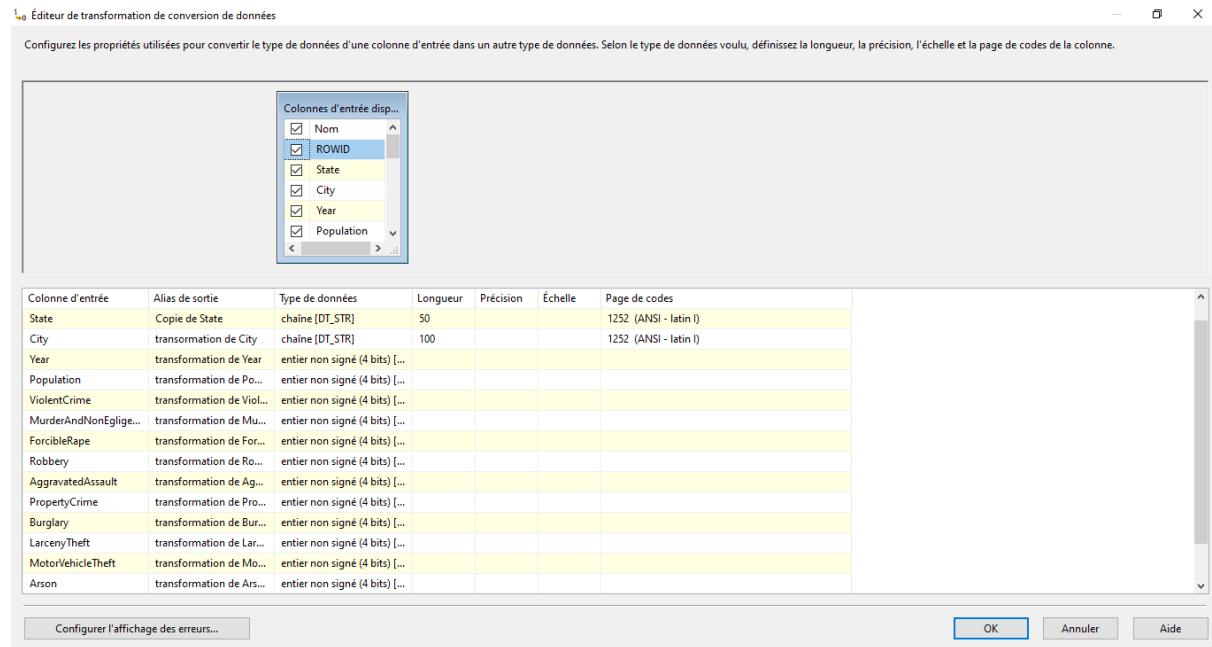
On choisit notre fichier source, et on met l'identificateur de texte pour éviter l'interprétation de « , » dans les champs de données comme un séparateur.



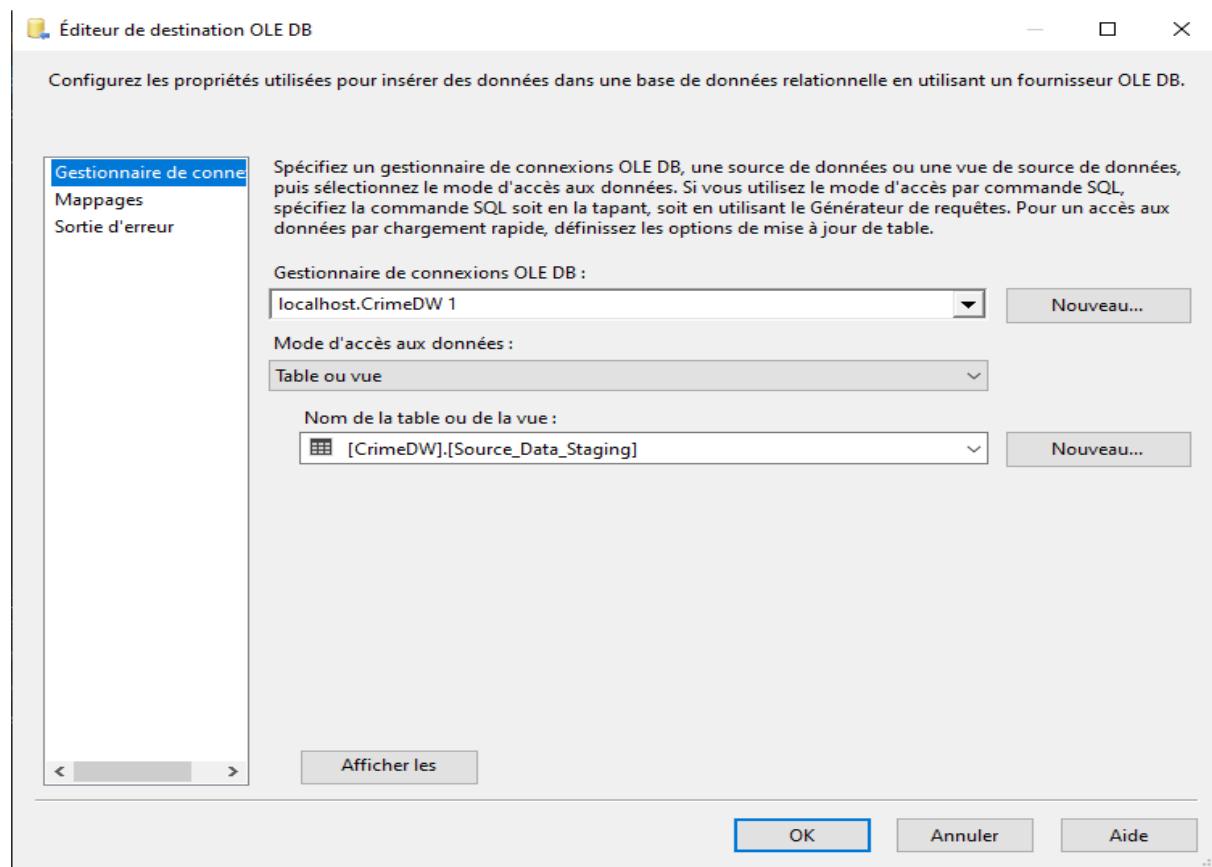
On peut déterminer les tailles des champs par exemple le champ city à 100

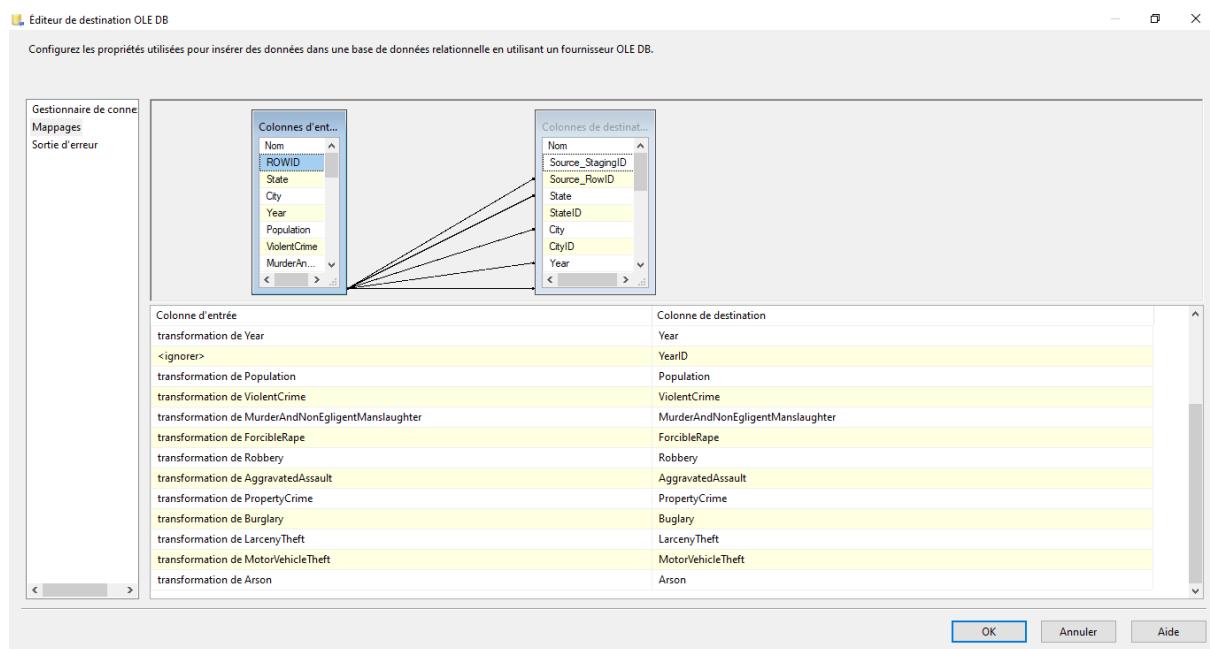
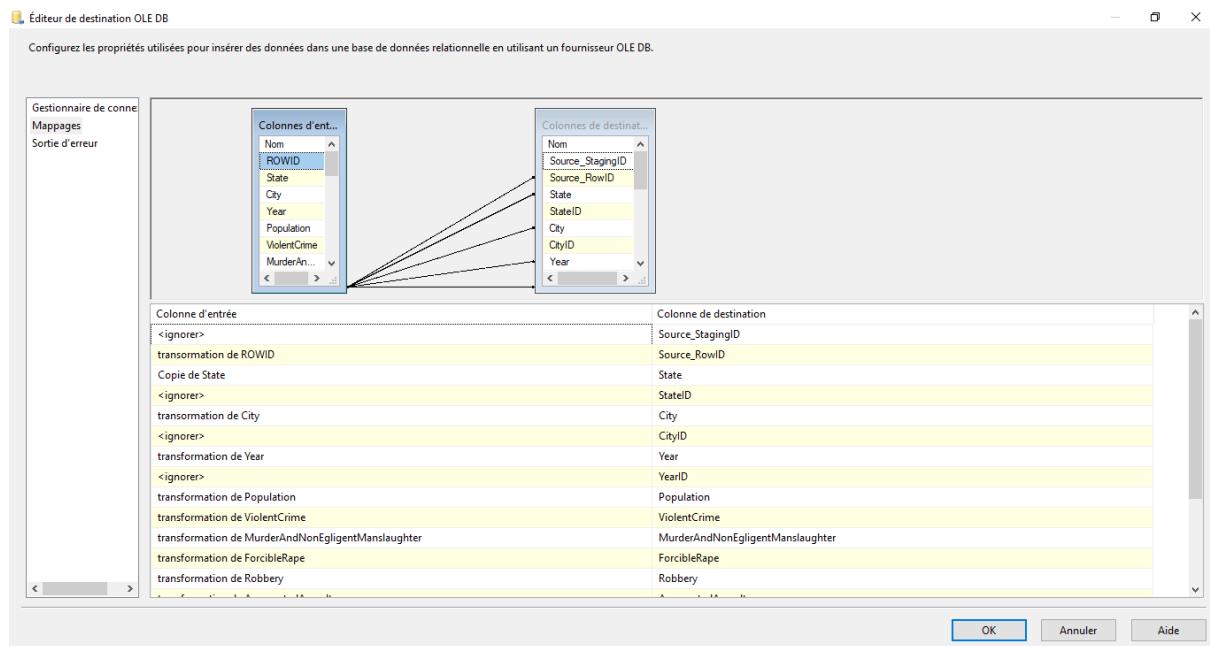


Pour la configuration de conversion de données :

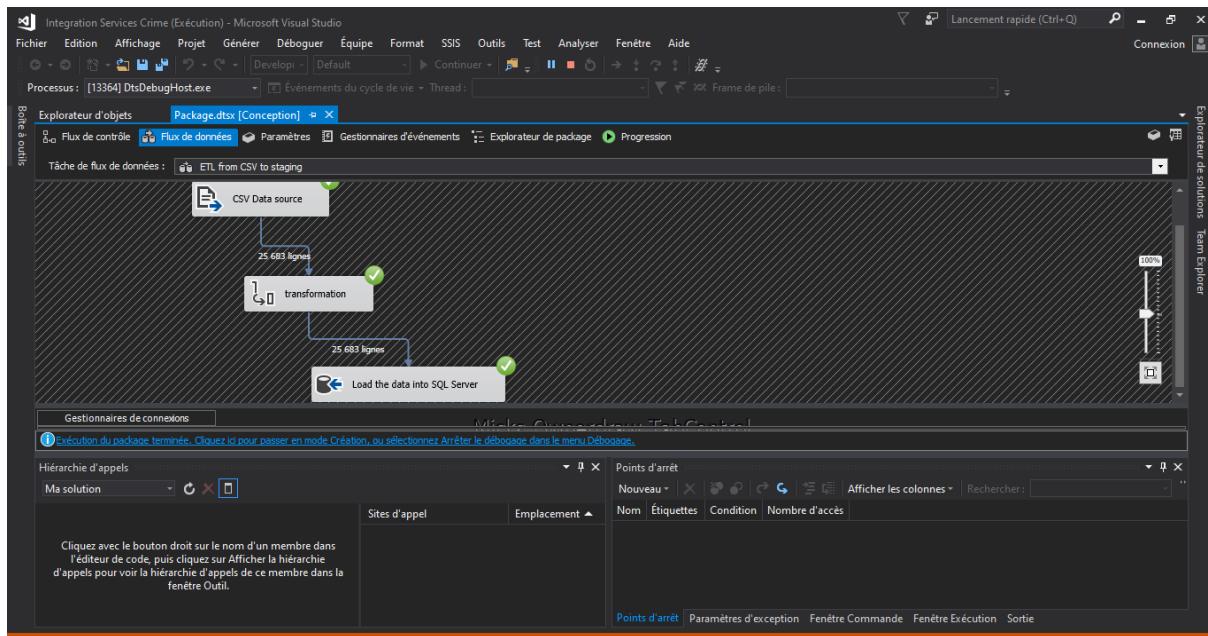


Pour la destination OLE DB on la configure de la manière suivante :

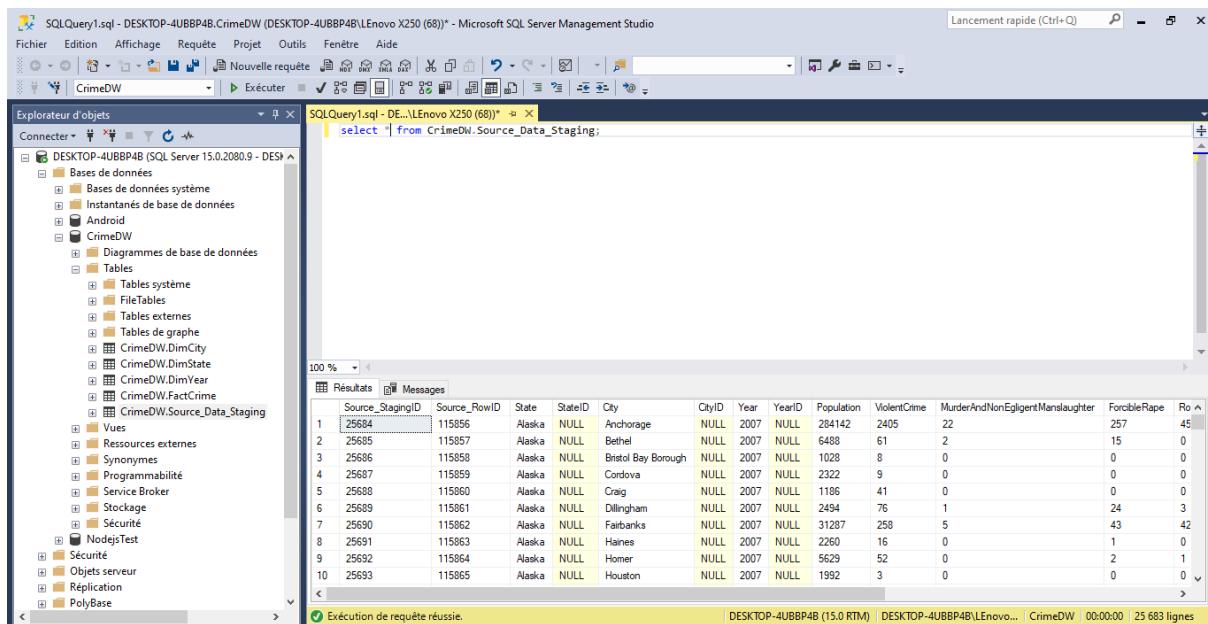




Finalement on exécute la tâche, les données seront chargées :



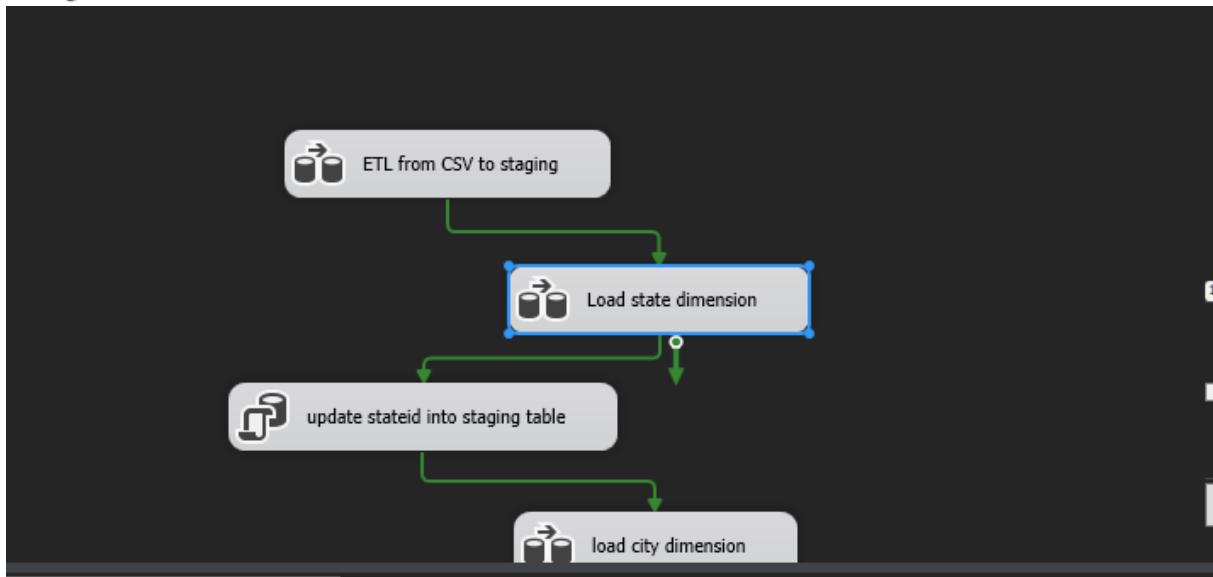
Vérification :



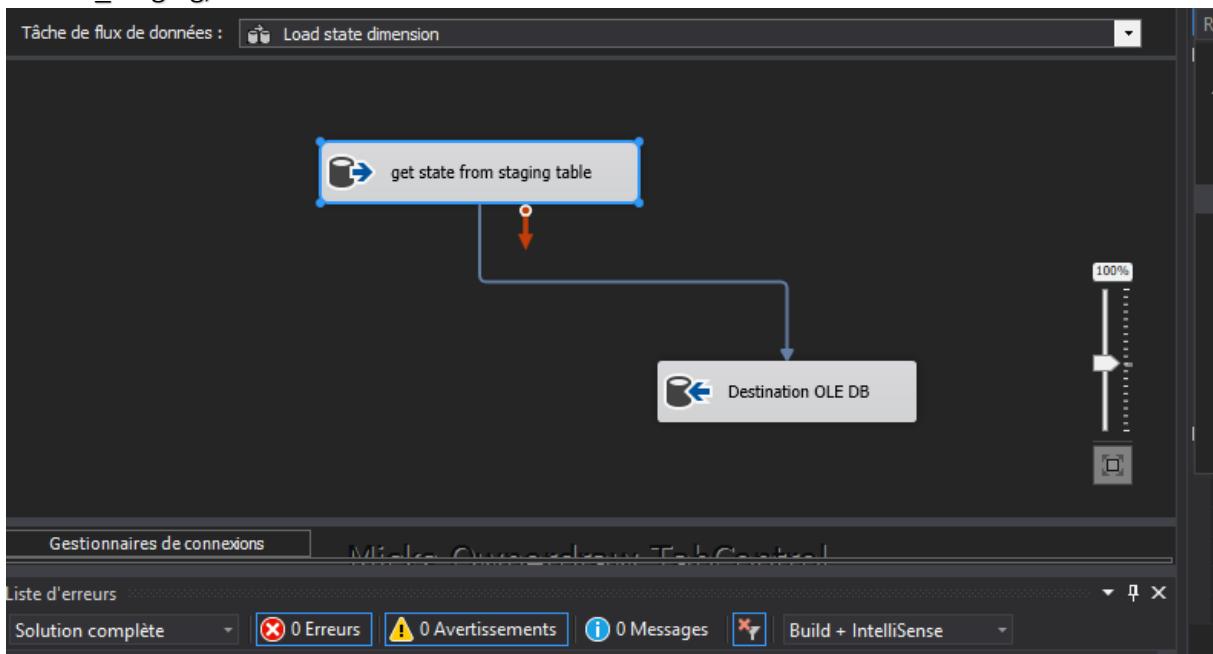
Source_StagingID	Source_RowID	State	StateID	City	CityID	Year	YearID	Population	ViolentCrime	MurderAndNonEglicitManslaughter	ForcibleRape	Robbery
1	115856	Alaska	NULL	Anchorage	NULL	2007	NULL	284142	2405	22	257	45
2	115857	Alaska	NULL	Bethel	NULL	2007	NULL	6488	61	2	15	0
3	115858	Alaska	NULL	Bristol Bay Borough	NULL	2007	NULL	1028	8	0	0	0
4	115859	Alaska	NULL	Cordova	NULL	2007	NULL	2322	9	0	0	0
5	115860	Alaska	NULL	Craig	NULL	2007	NULL	1195	41	0	0	0
6	115861	Alaska	NULL	Dillingham	NULL	2007	NULL	2494	76	1	24	3
7	115862	Alaska	NULL	Fairbanks	NULL	2007	NULL	31287	258	5	43	42
8	115863	Alaska	NULL	Haines	NULL	2007	NULL	2260	16	0	1	0
9	115864	Alaska	NULL	Homer	NULL	2007	NULL	5629	52	0	2	1
10	115865	Alaska	NULL	Houston	NULL	2007	NULL	1992	3	0	0	0

Les clés primaires sont Null car notre source ne contient pas ces données, on va voir par la suite que ces clés seront générées automatiquement.

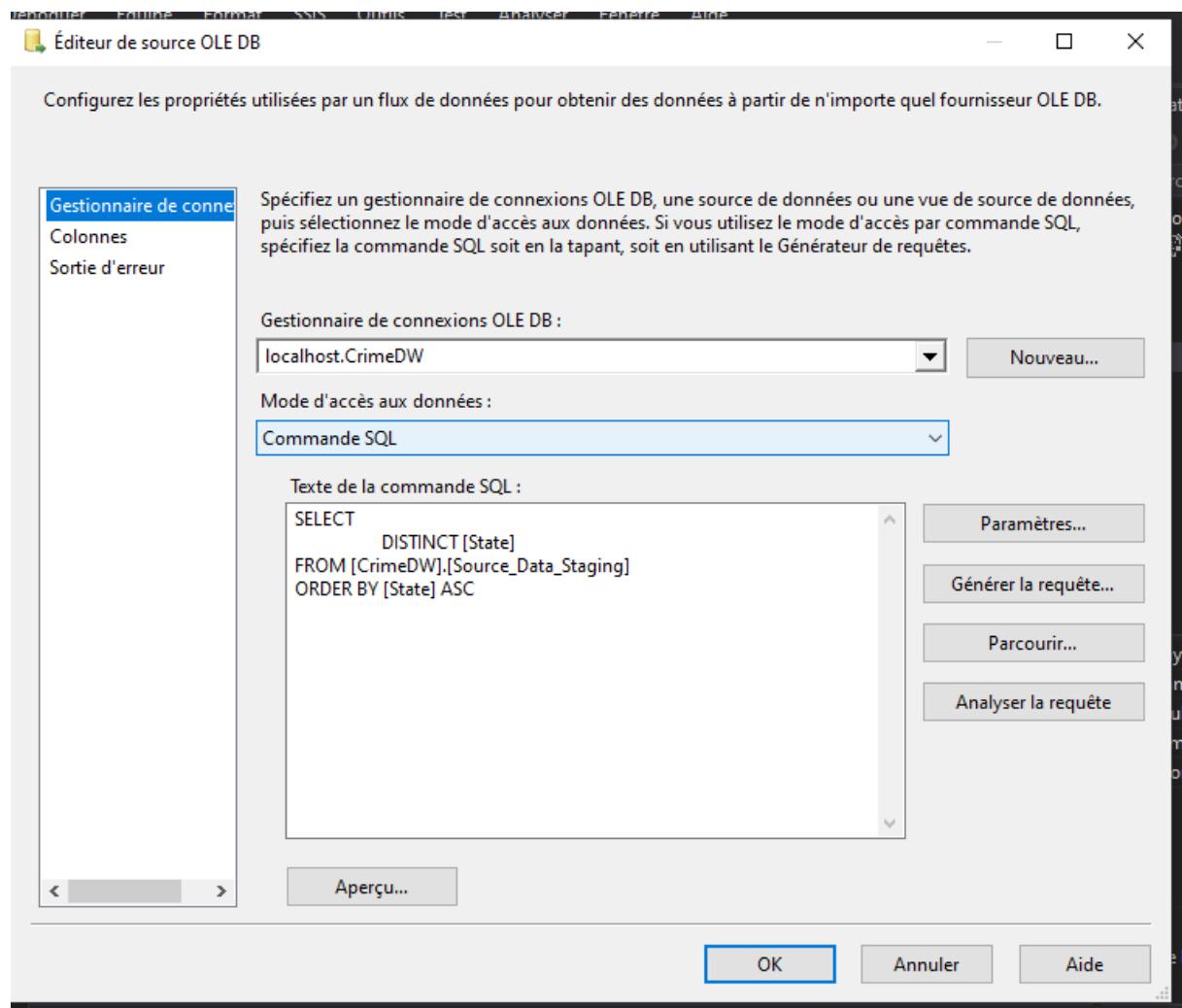
On va créer une nouvelle tâche de flux de données nommée Load state dimension, pour charger les données de la table DimState :



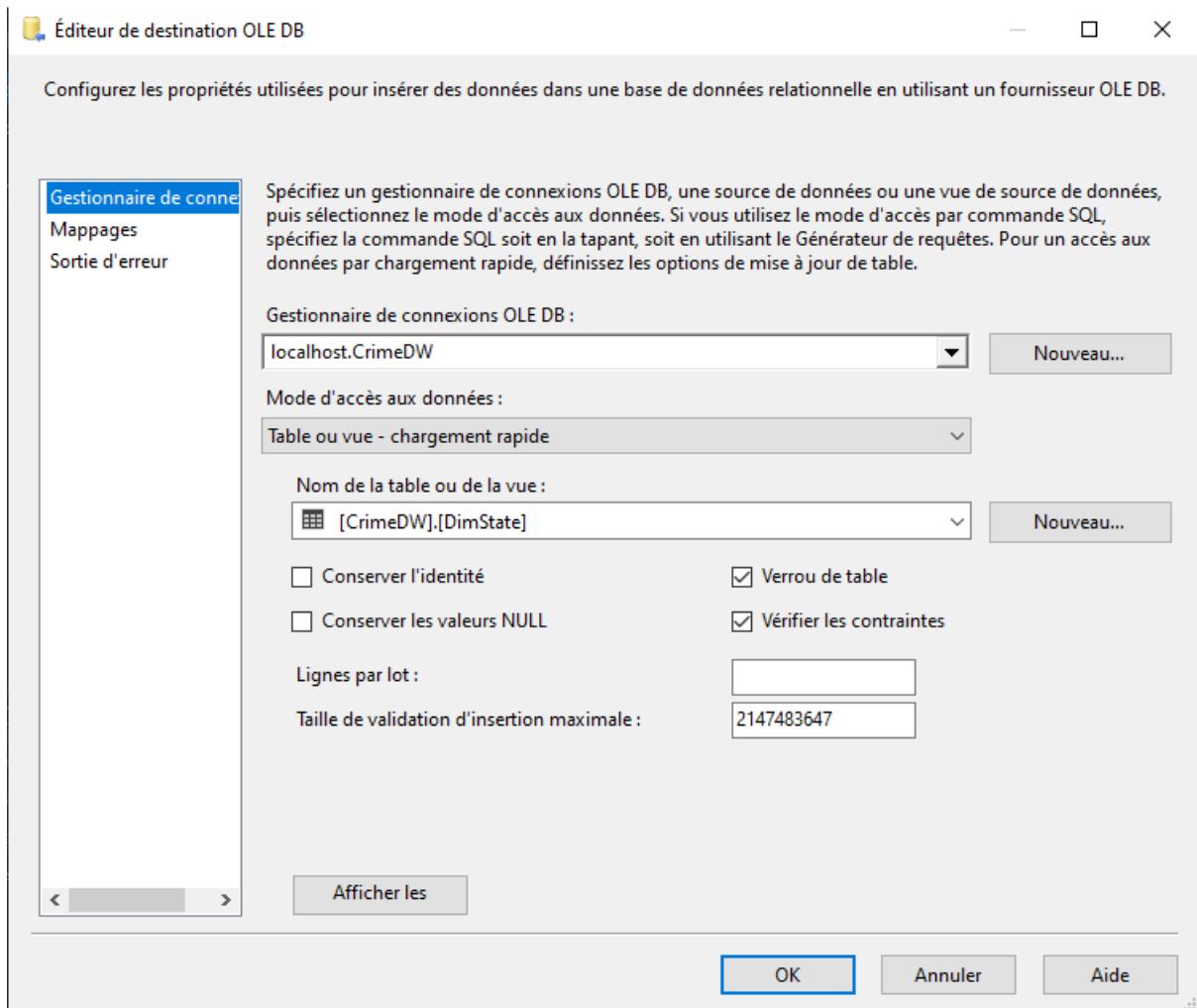
Dans cette tâche on crée une source OLE DB car nos données sont maintenant dans la table Source_Staging, et une destination OLE DB



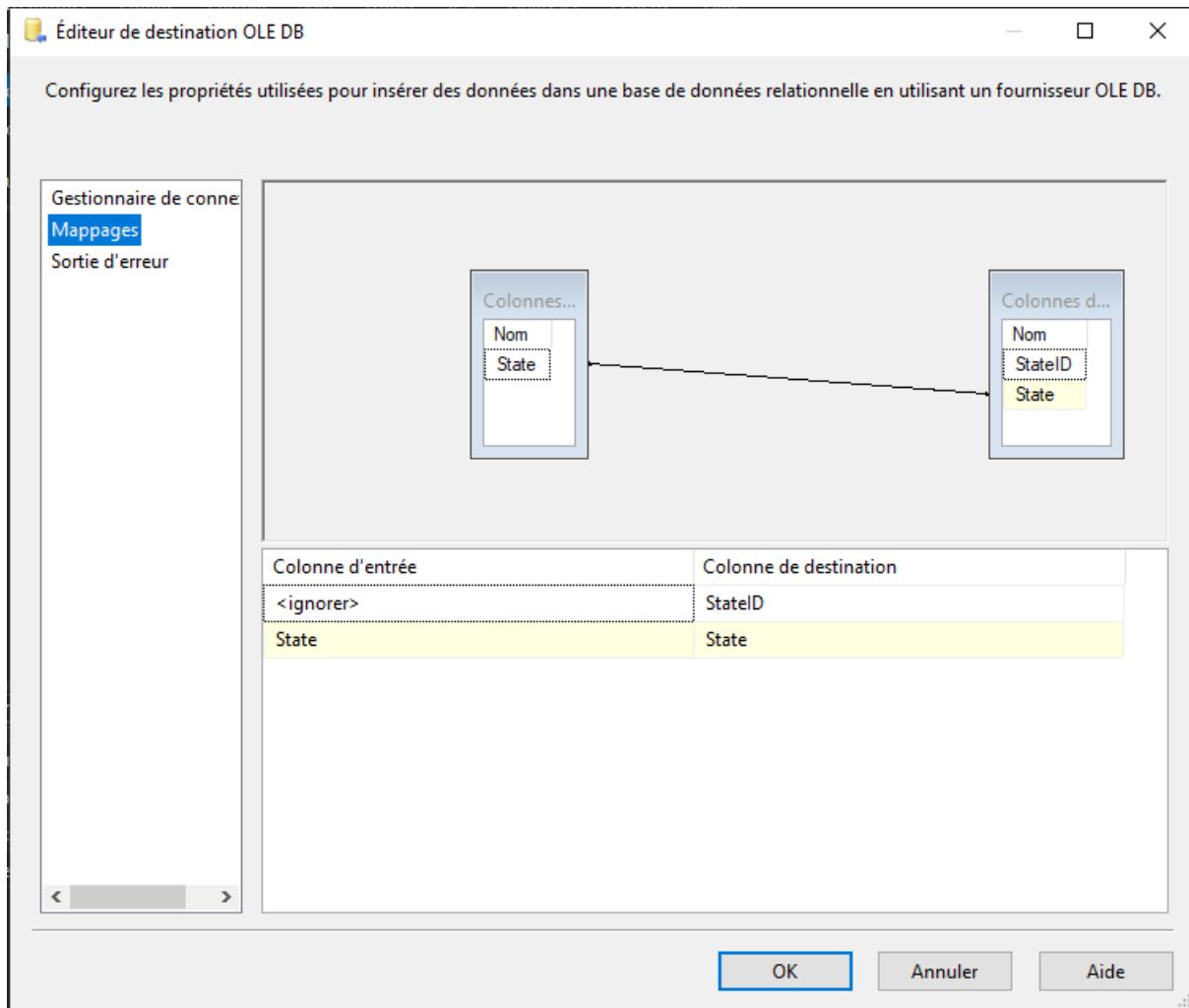
On configure la source OLE DB :



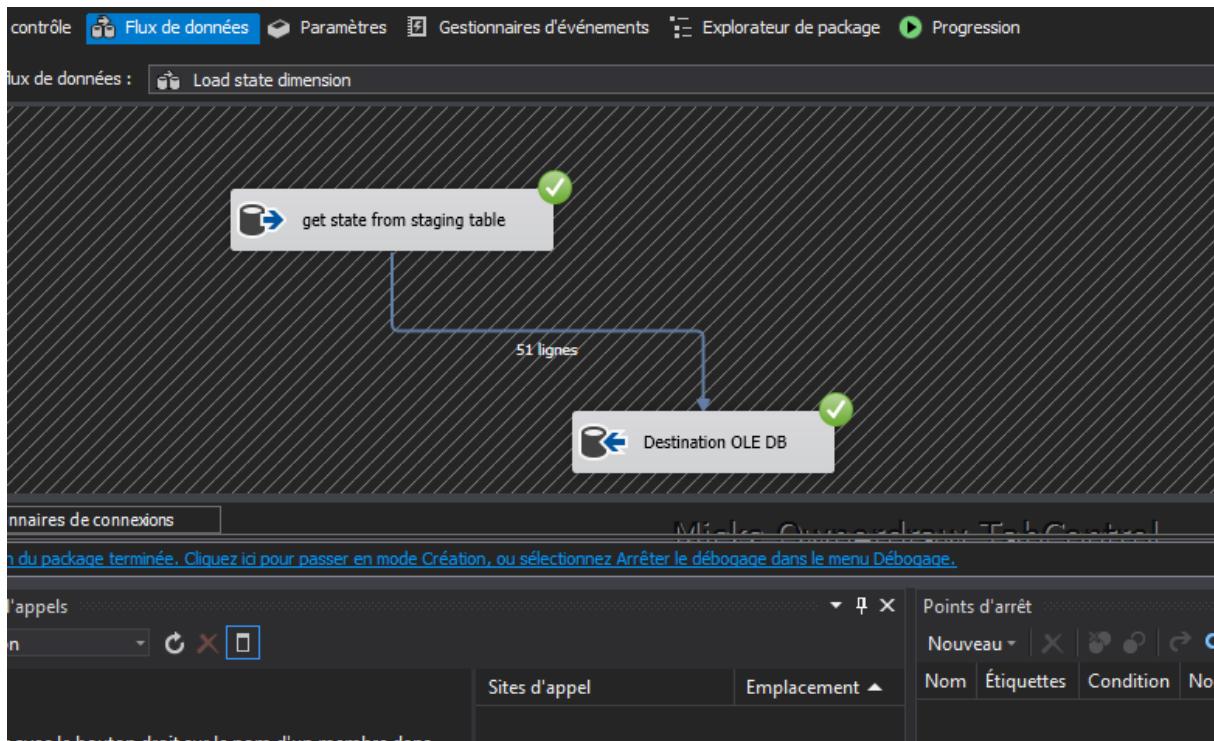
On configure la destination OLE DB :



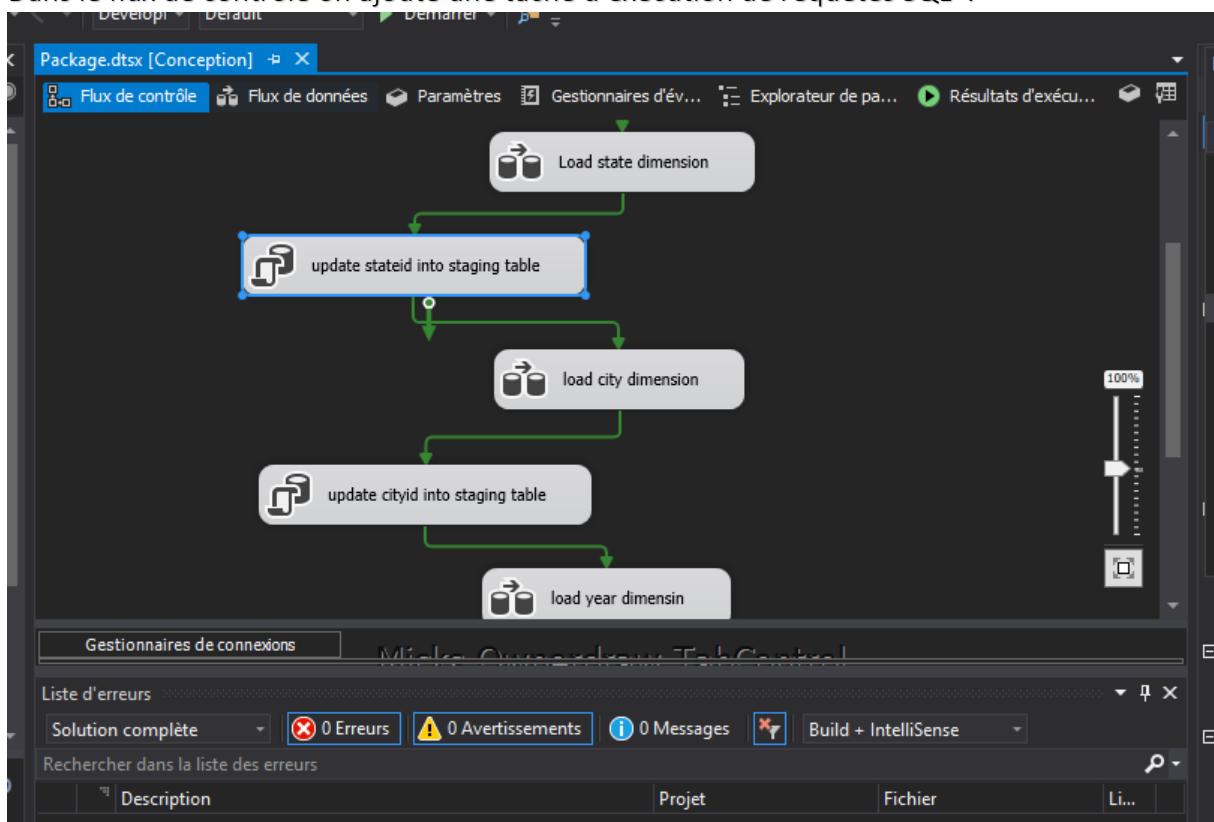
StateID va être généré automatiquement par le système :



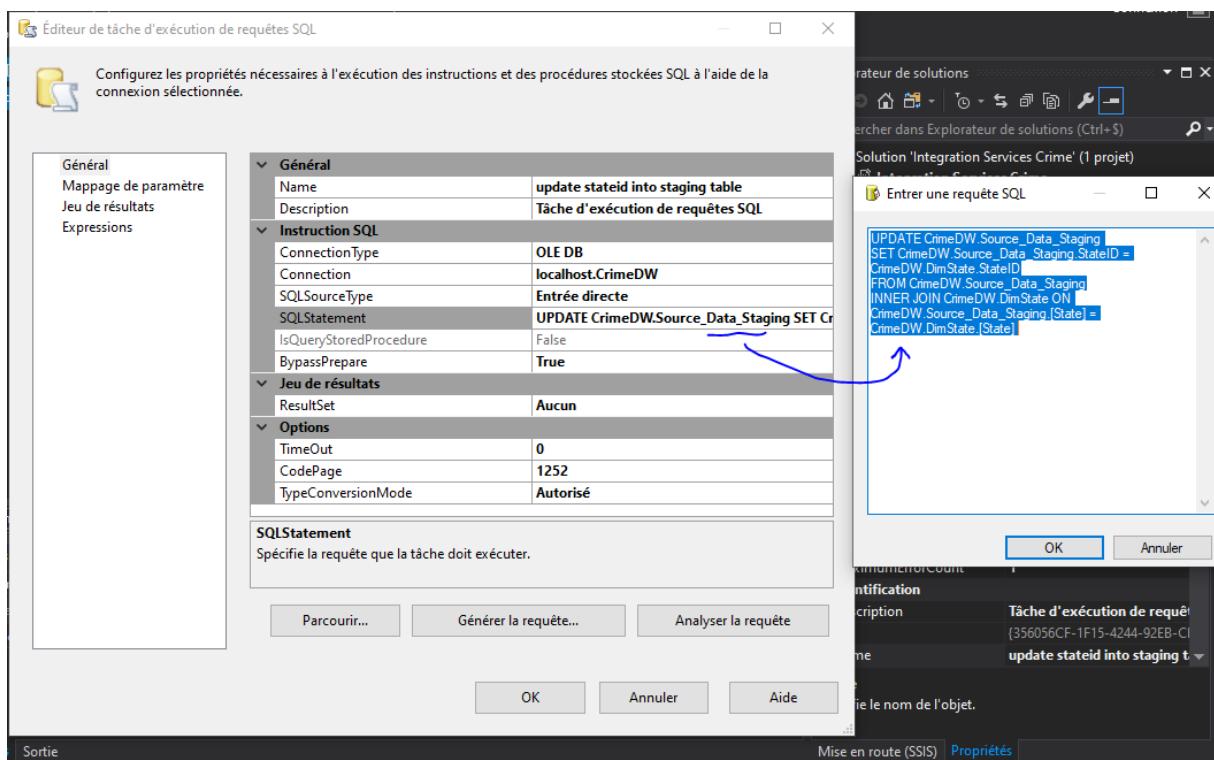
On execute la tâche :



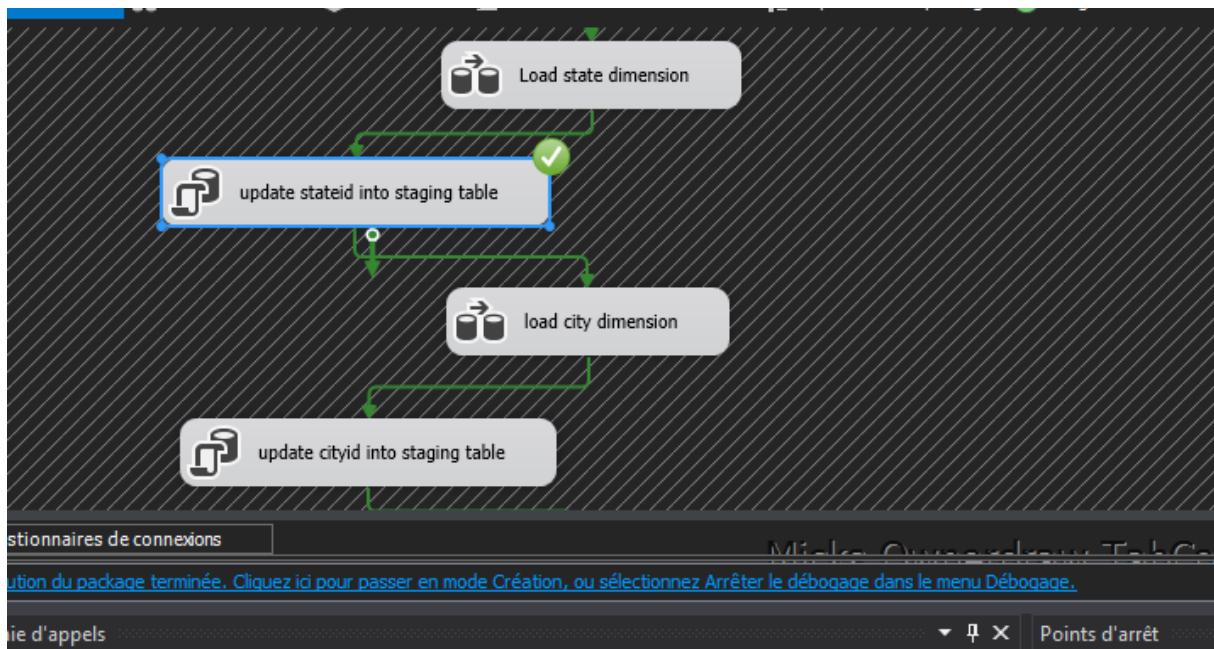
Dans le flux de contrôle on ajoute une tâche d'exécution de requêtes SQL :



On la configures de la manière suivante :

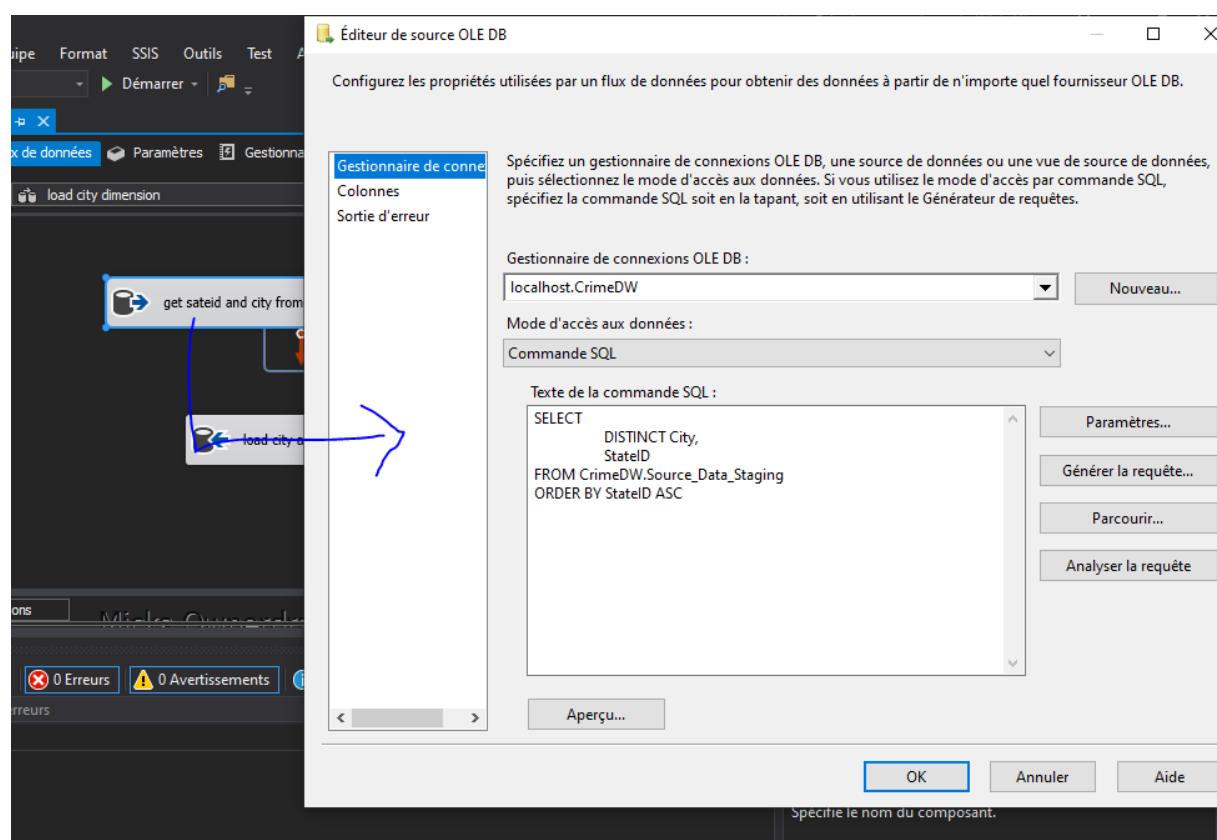
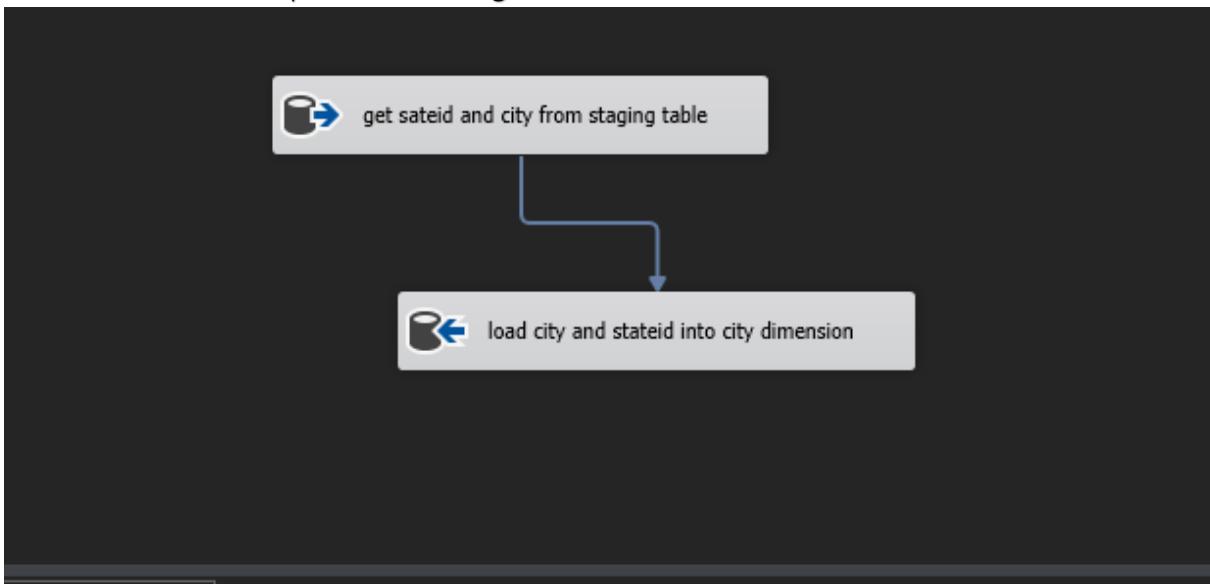


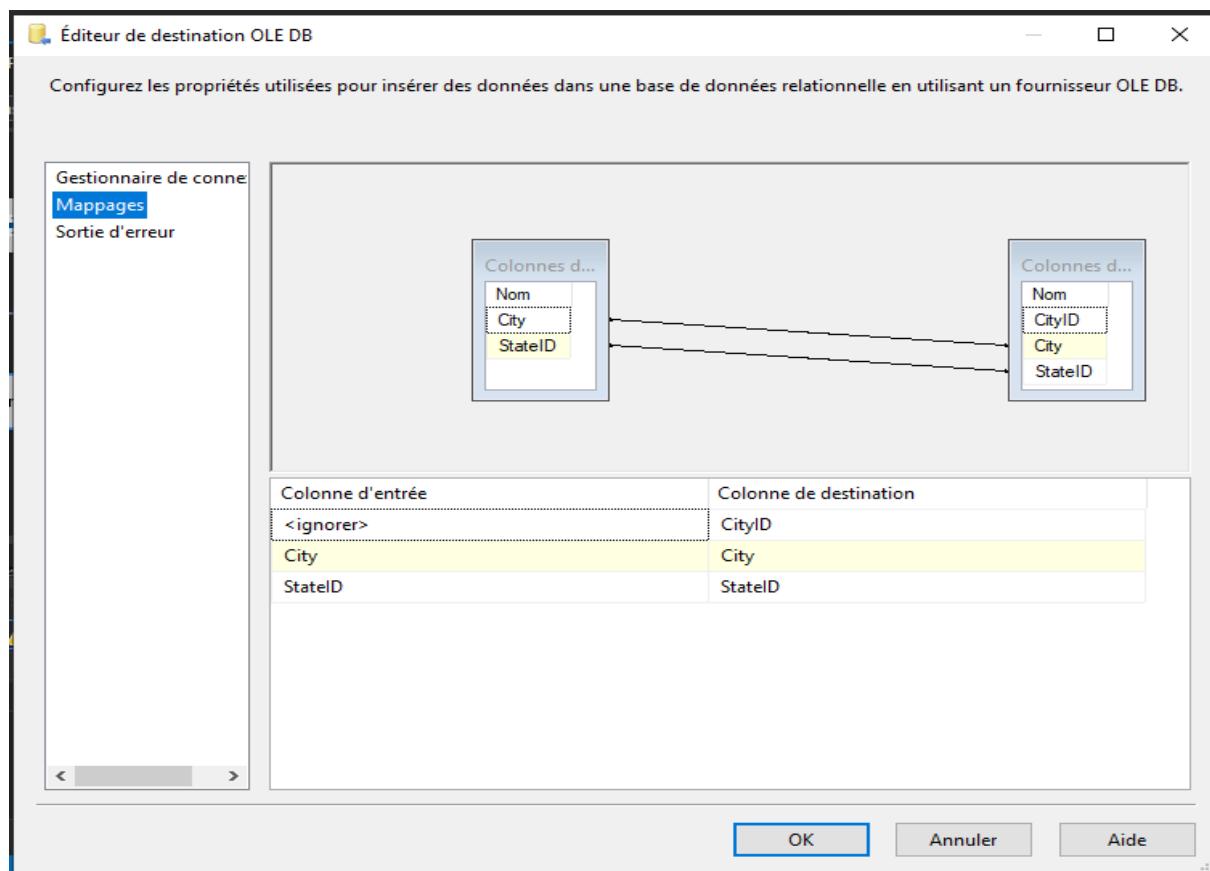
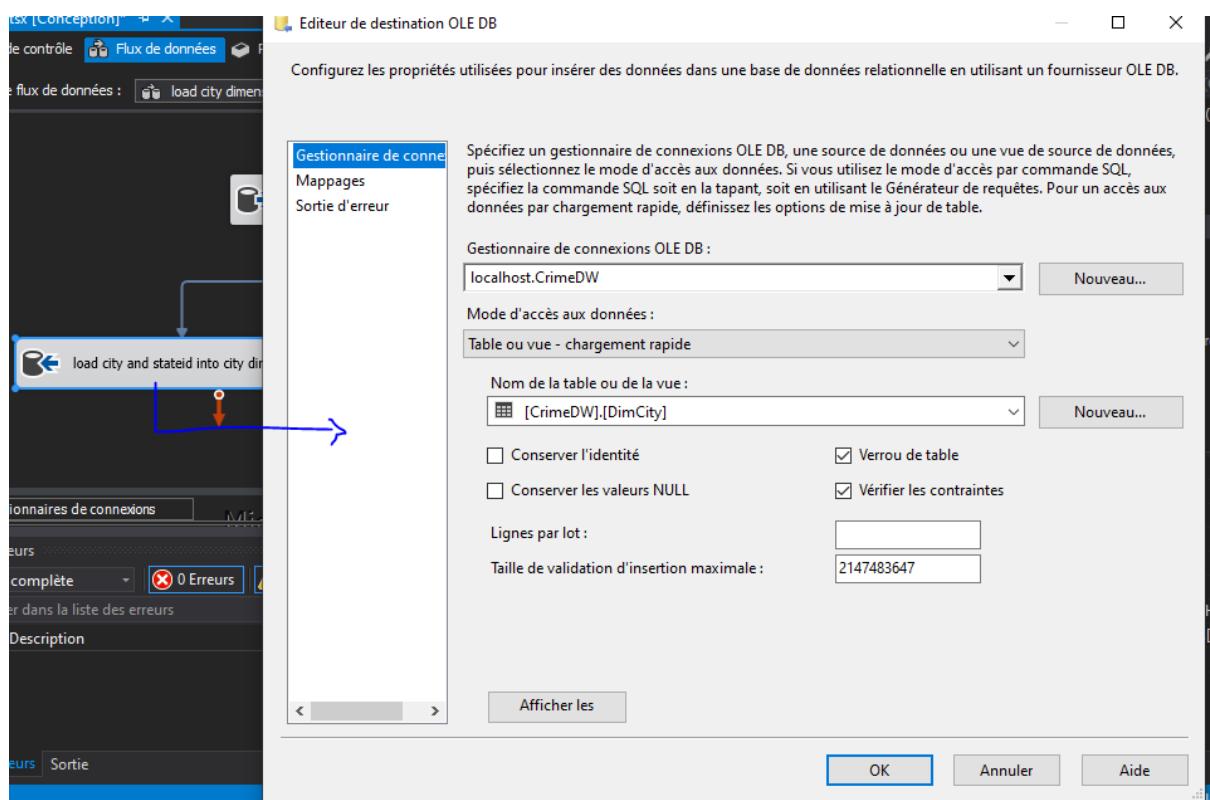
On exécute cette tâche :



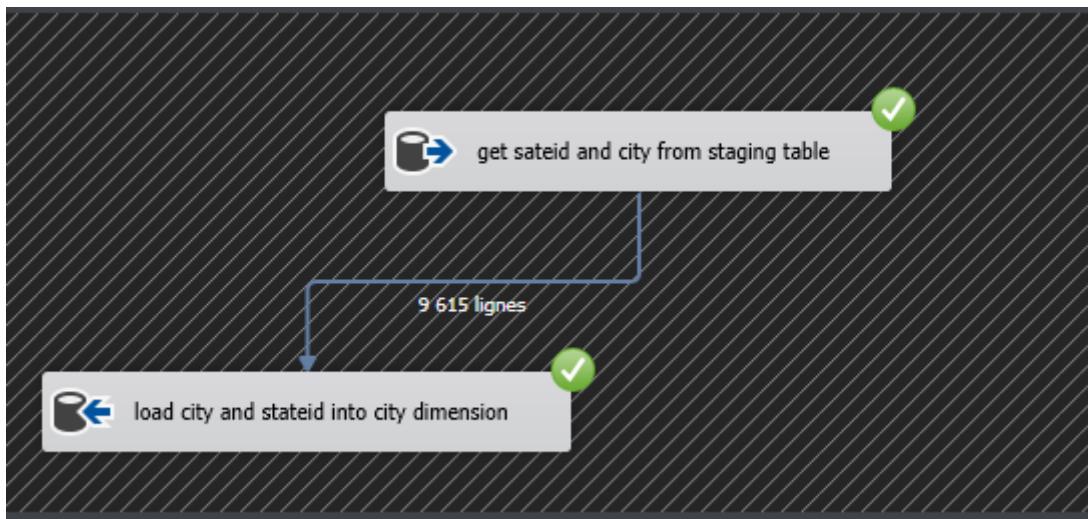
Ensuite on ajoute une tâche de flux de données nommée load city dimension pour charger les données de la table DimCity.

De la même manière que dans le chargement de DimState :

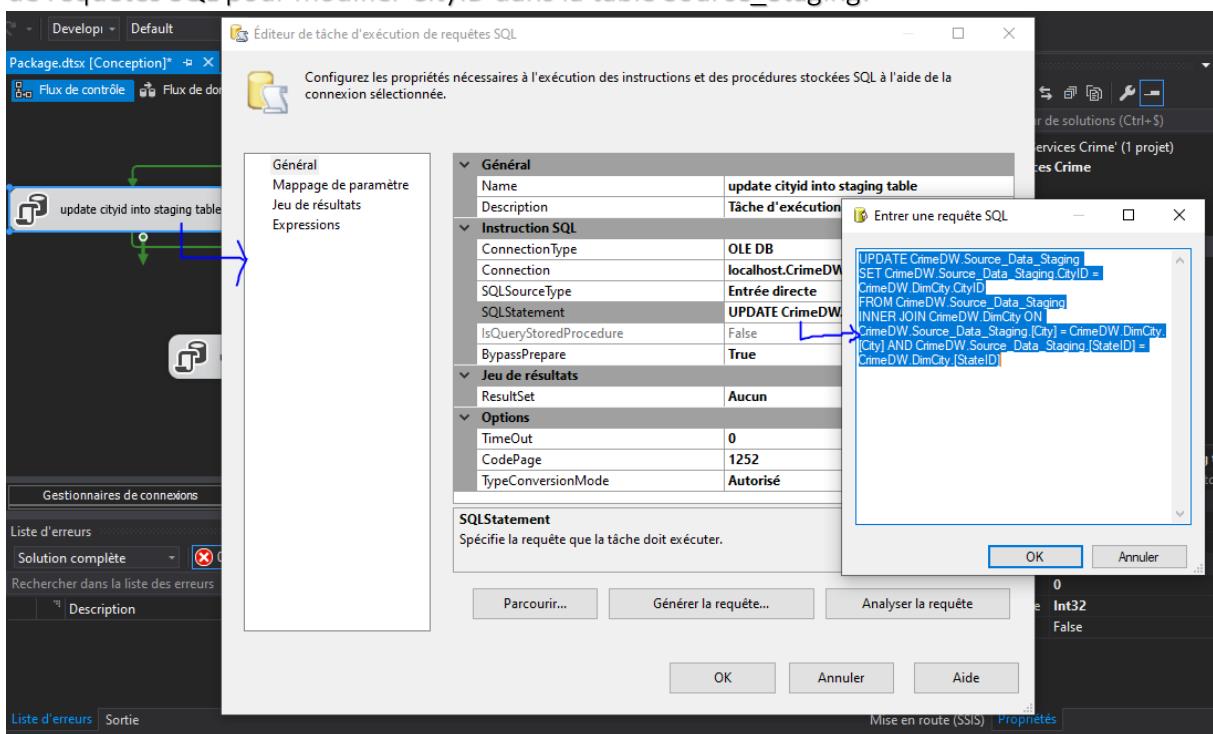


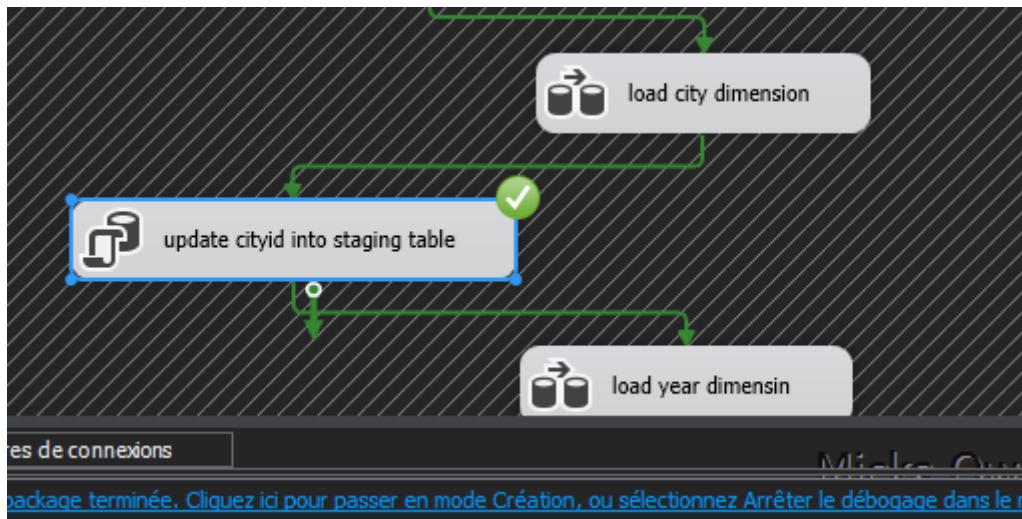


On execute cette tâche :

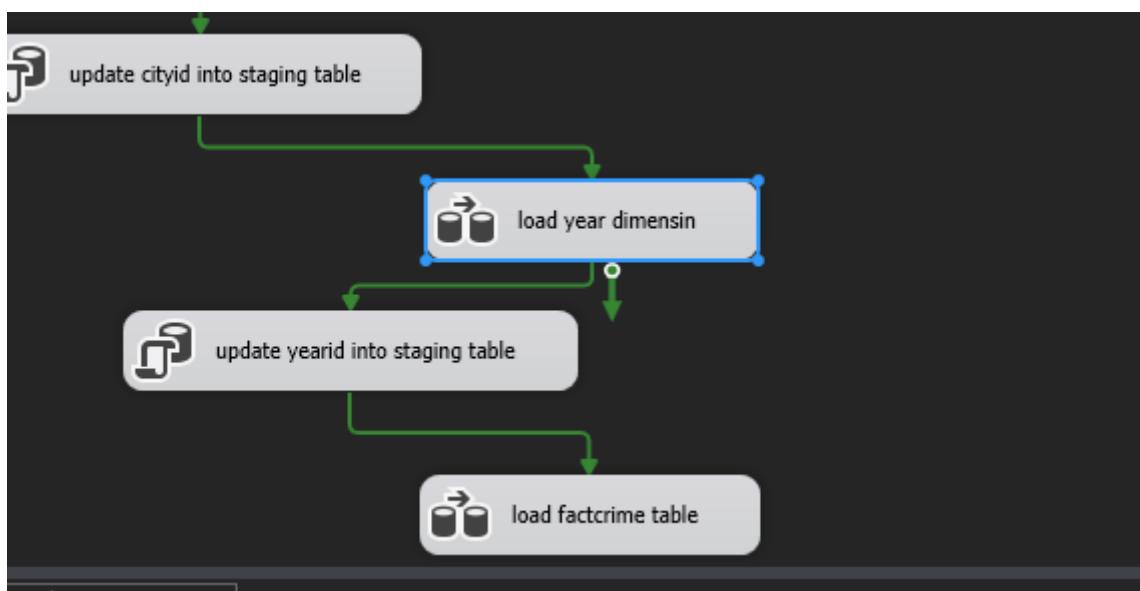


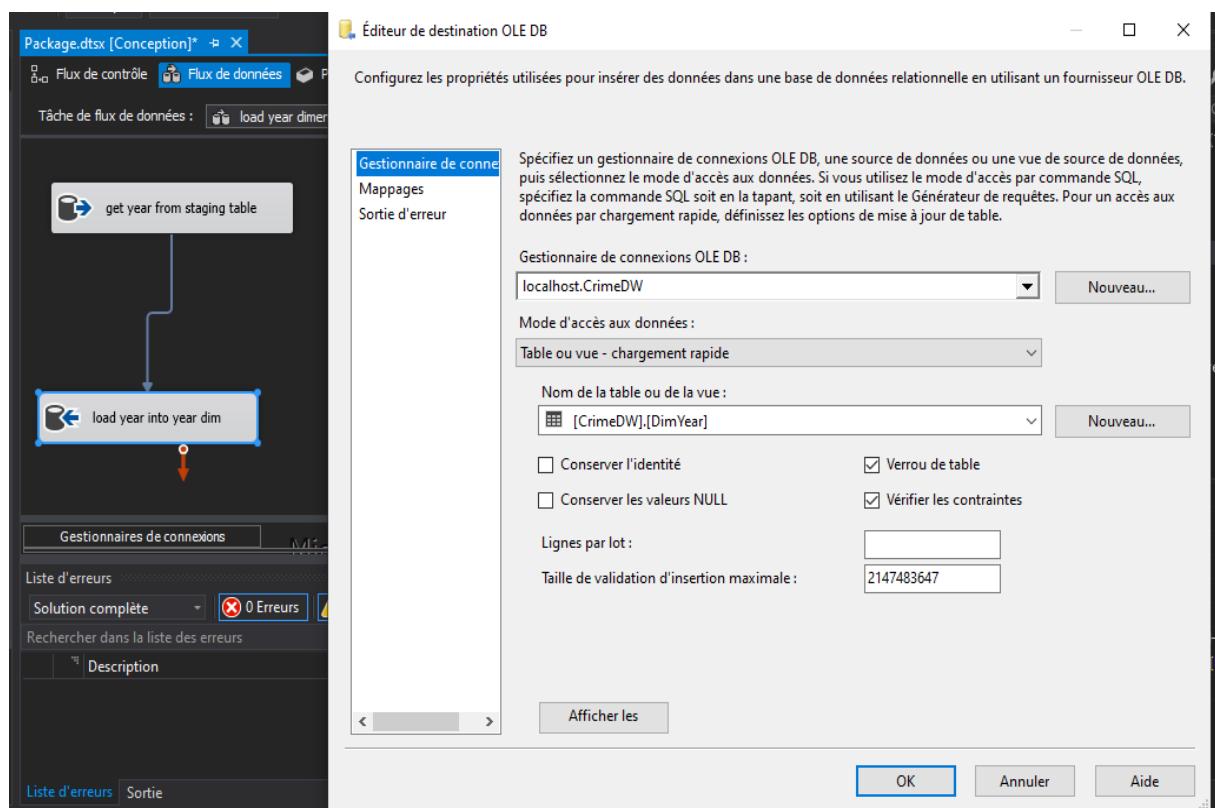
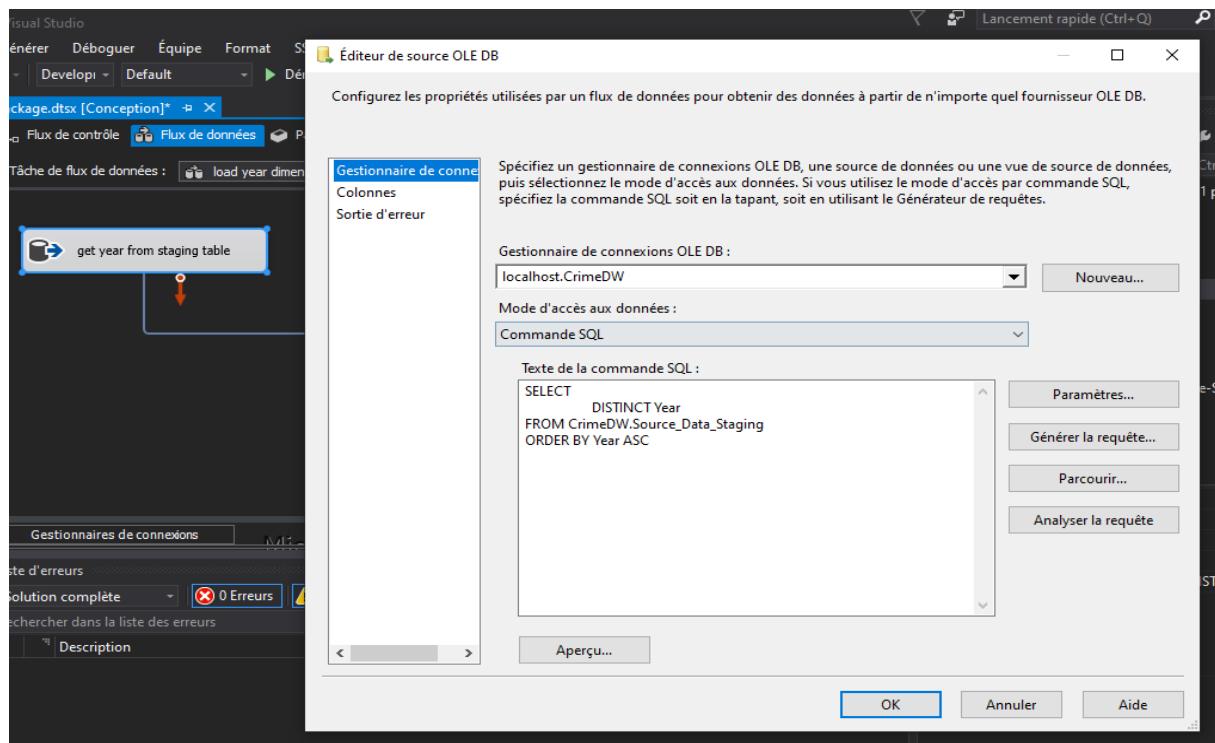
De la même manière qu'auparavant, dans le flux de contrôle on ajoute une tâche d'exécution de requêtes SQL pour modifier CityID dans la table Source_Staging:

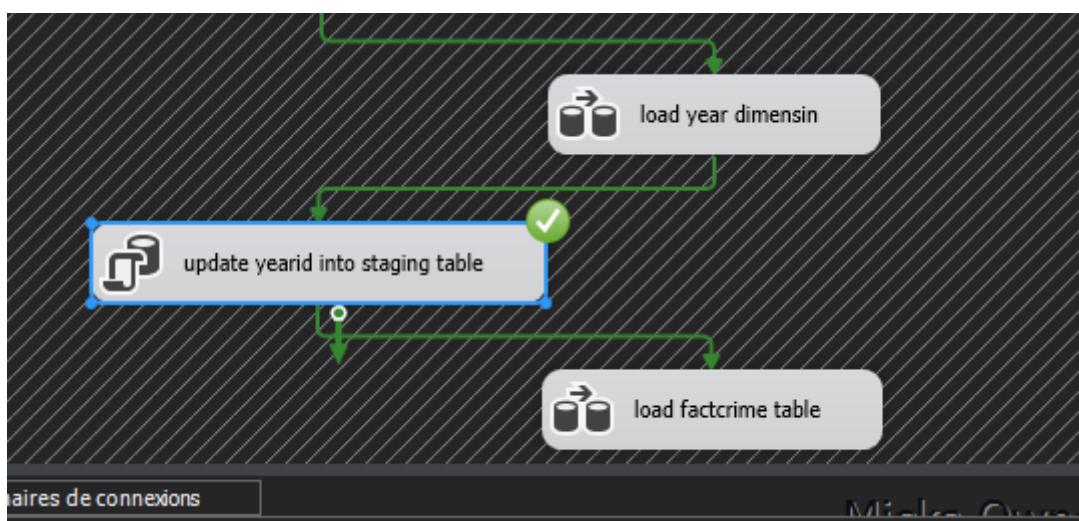
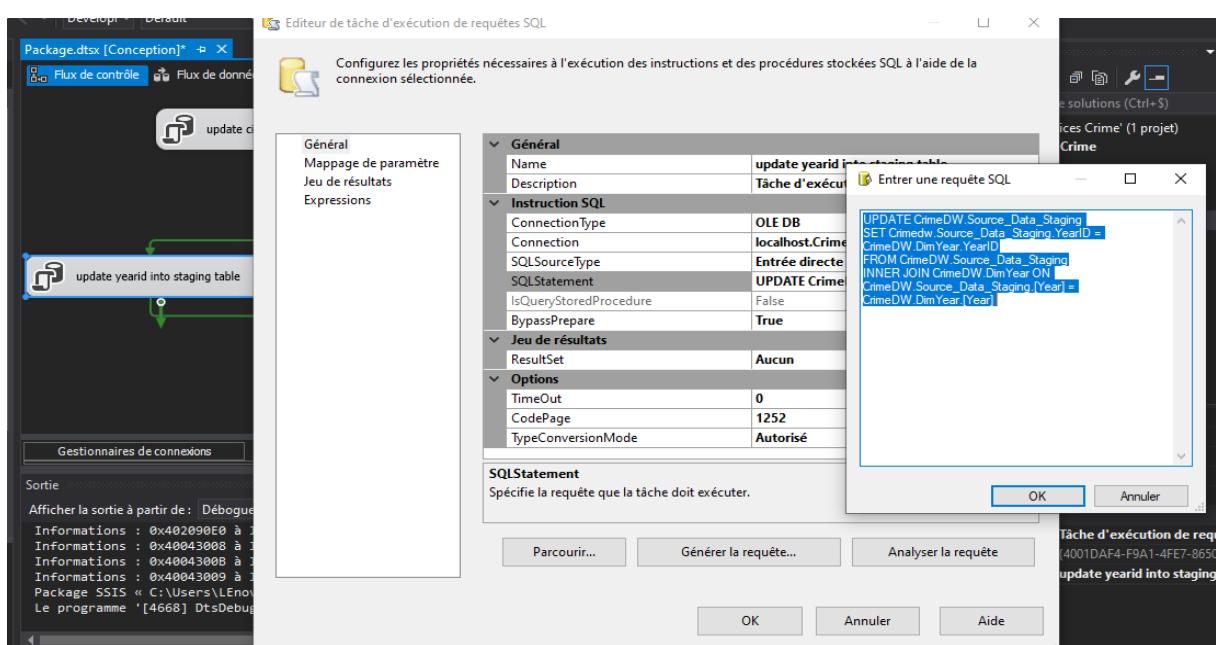
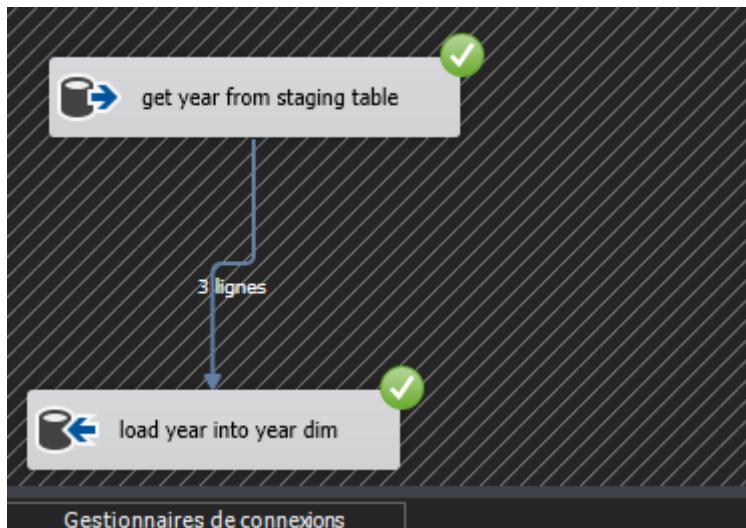




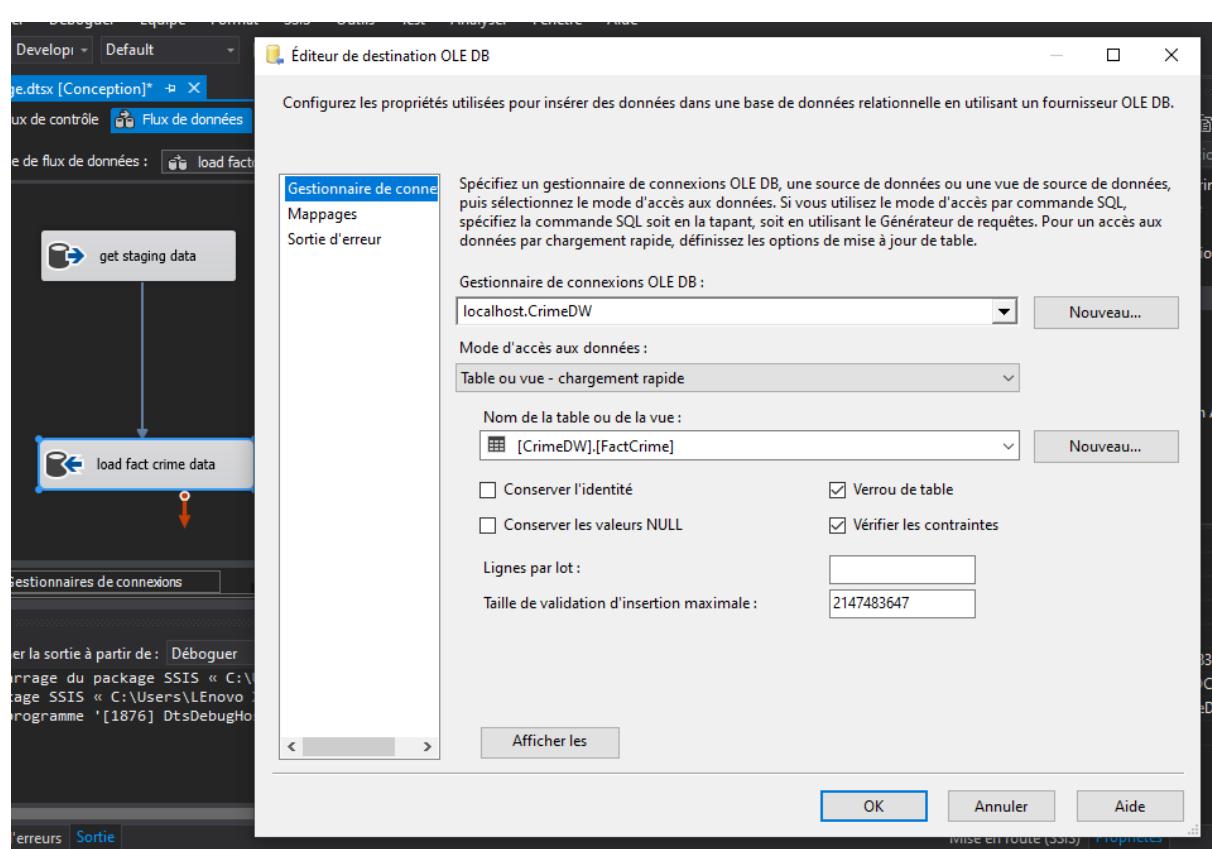
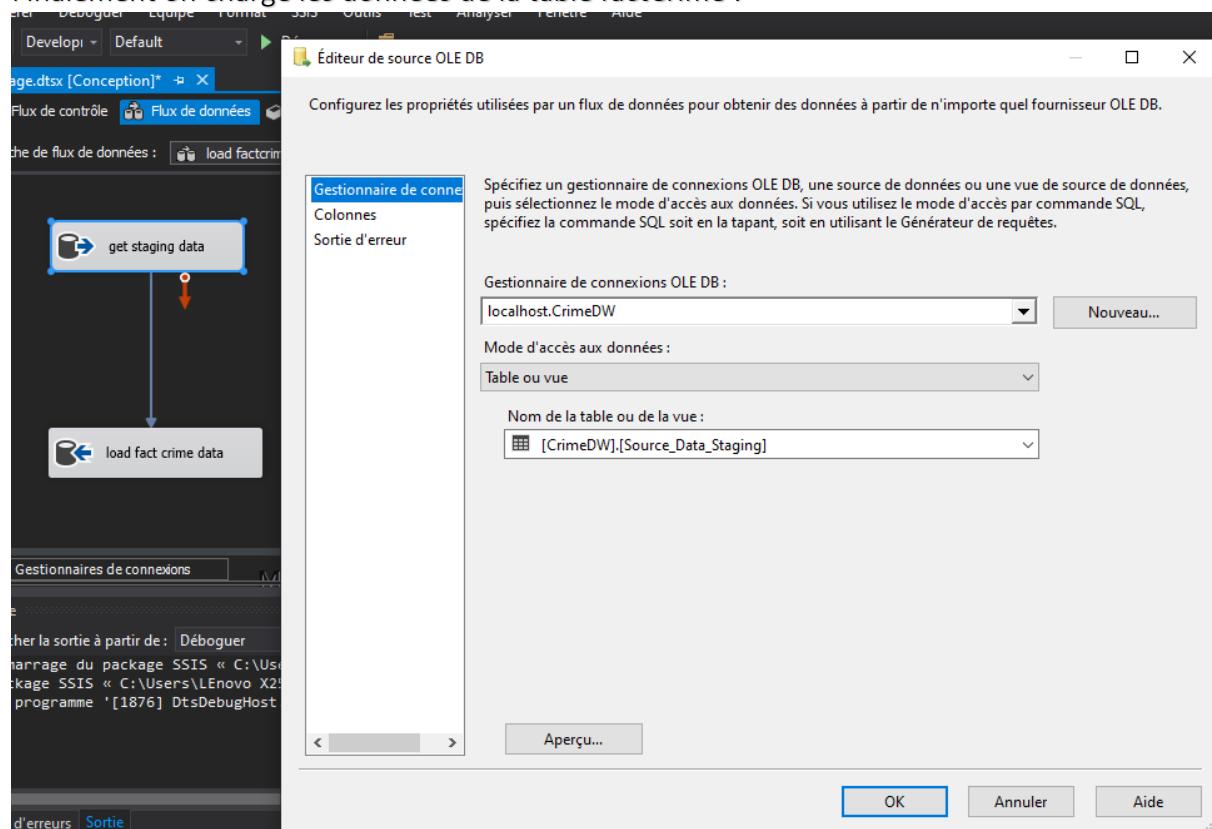
Et de même pour le chargement de la dimension Year :

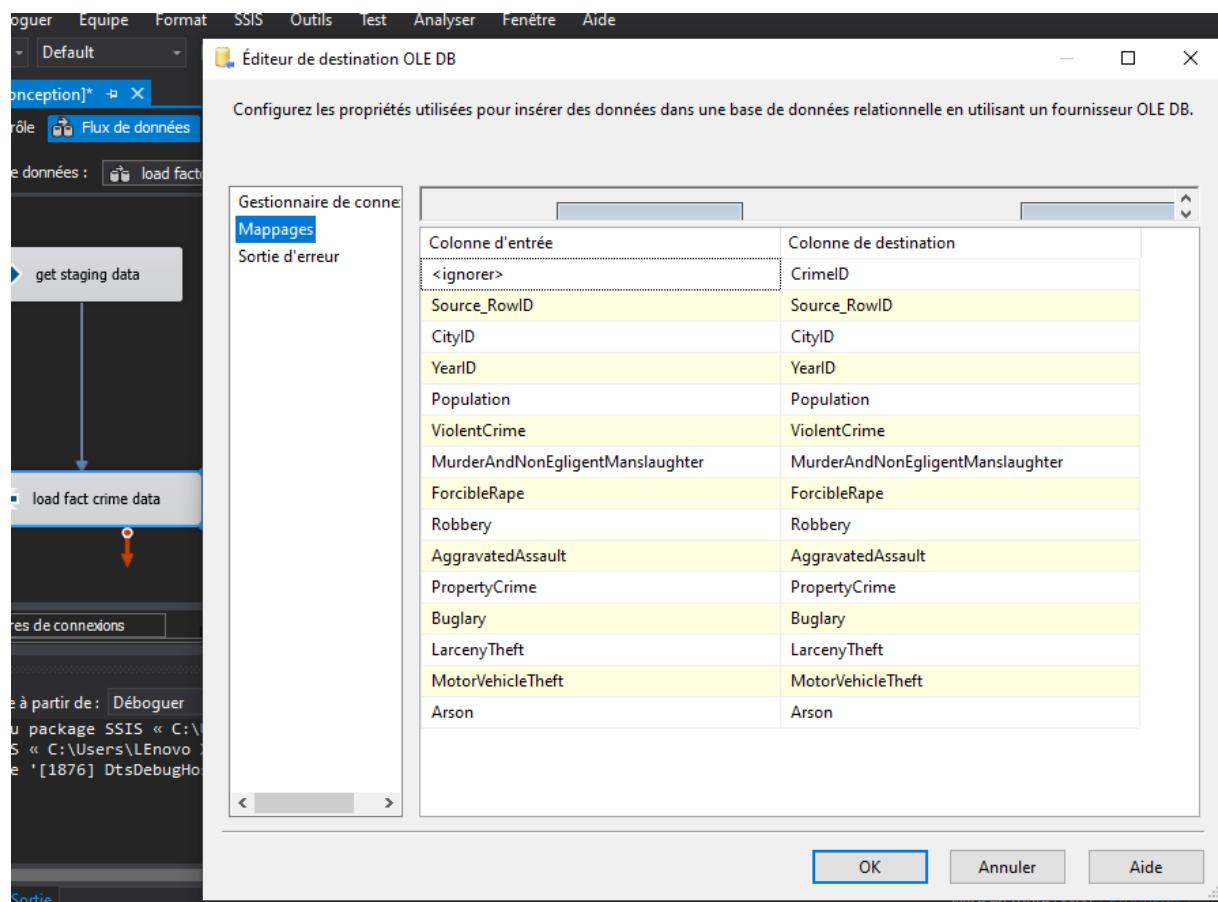






Finalement on charge les données de la table factCrime :





On execute la tâche :



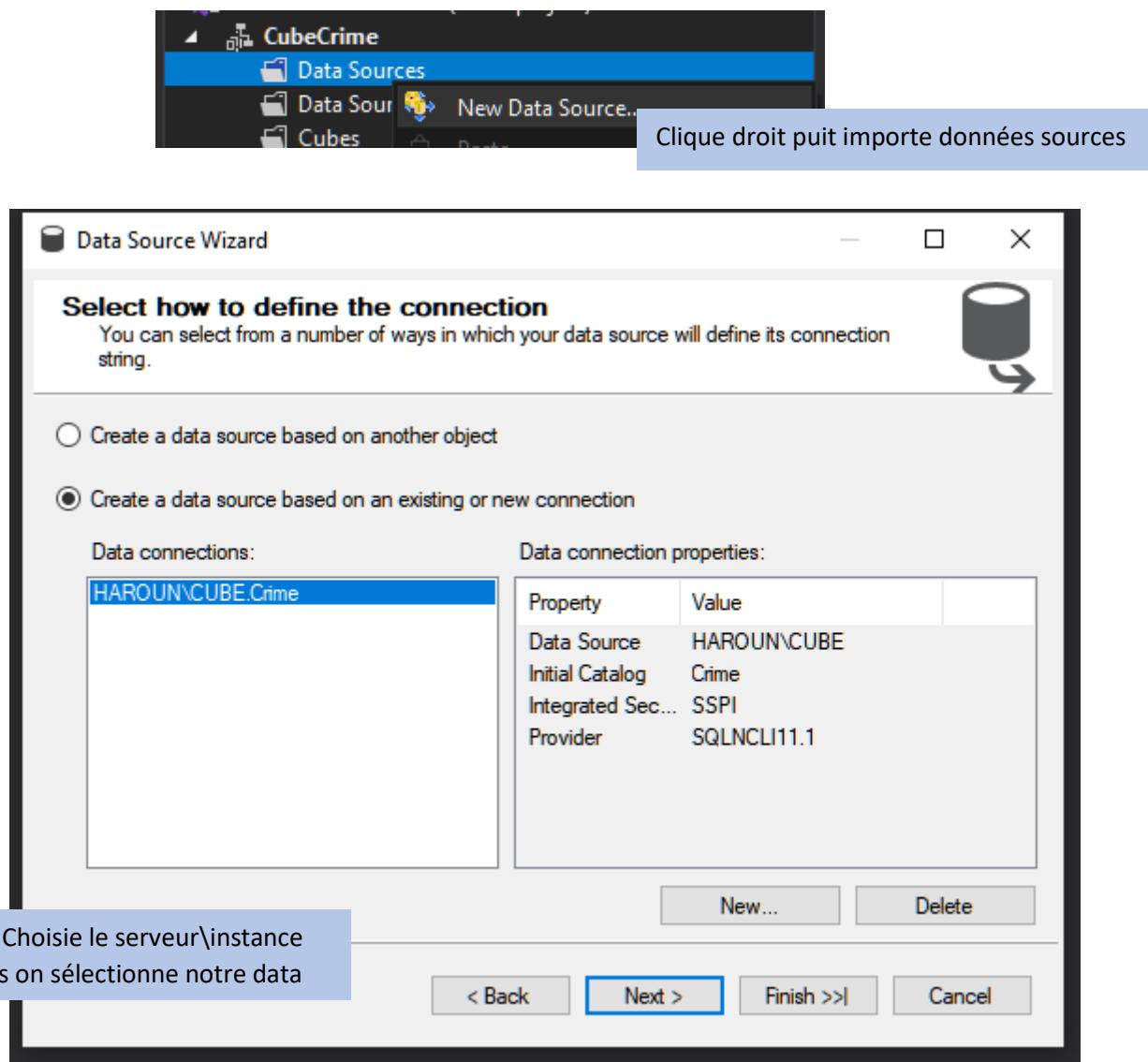
On a terminé la partie ETL et notre datawarehouse est créé, on passe maintenant à la création de notre cube OLAP.

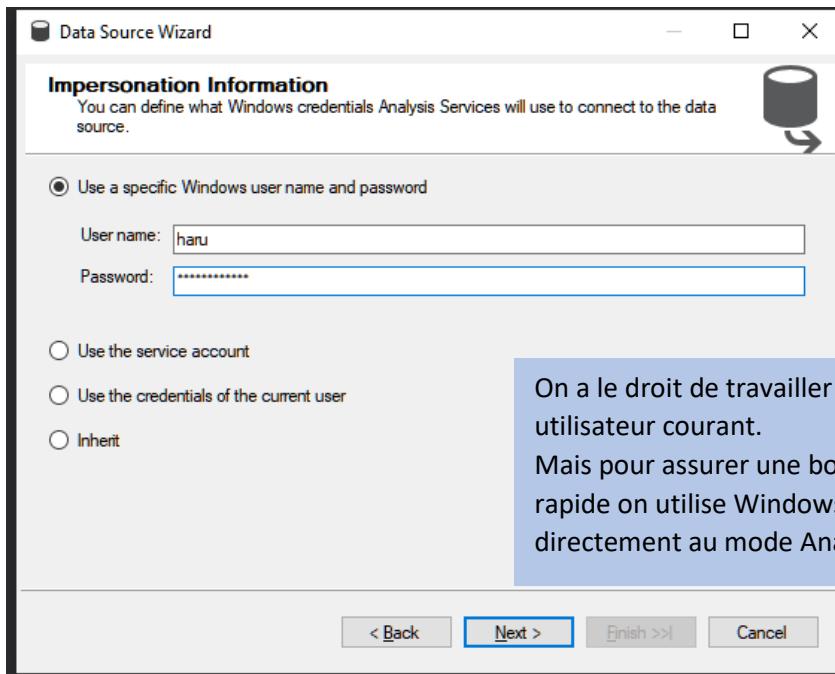
Création de Cube – Multidimensionnel :

Pour commencer, on a créé un nouveau projet Analysis Services sous SQL Server Data Tools.

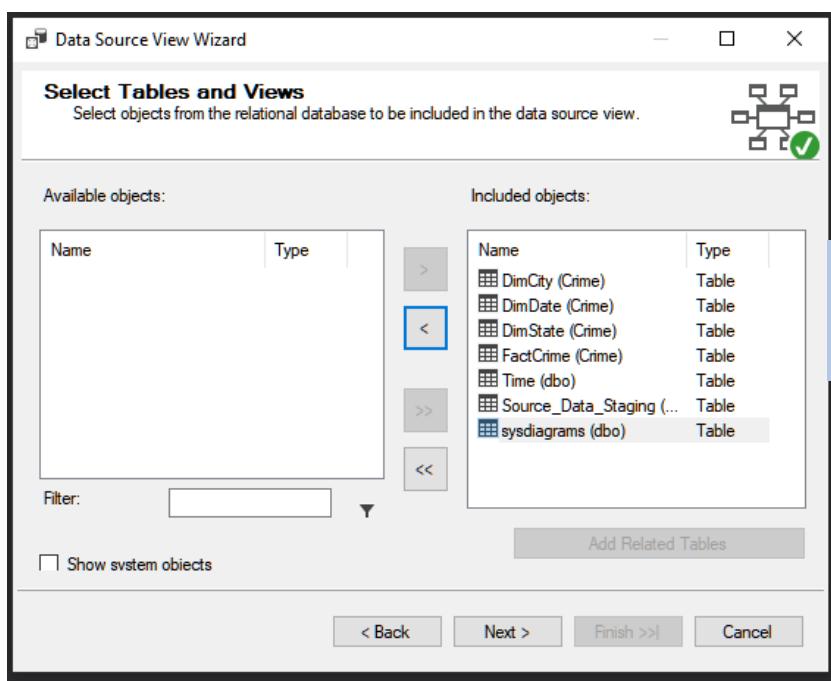
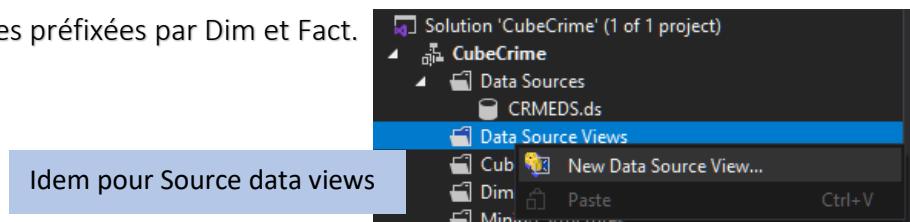
Après avoir créé l'entrepôt de données et le bien modélisé, la création de notre cube a fait en trois étapes ;

Création de la source de données : on utilise la source de données **Crime**, ainsi à la fenêtre de l'Assistant Sources de données, sélectionnez Utiliser le compte de service.

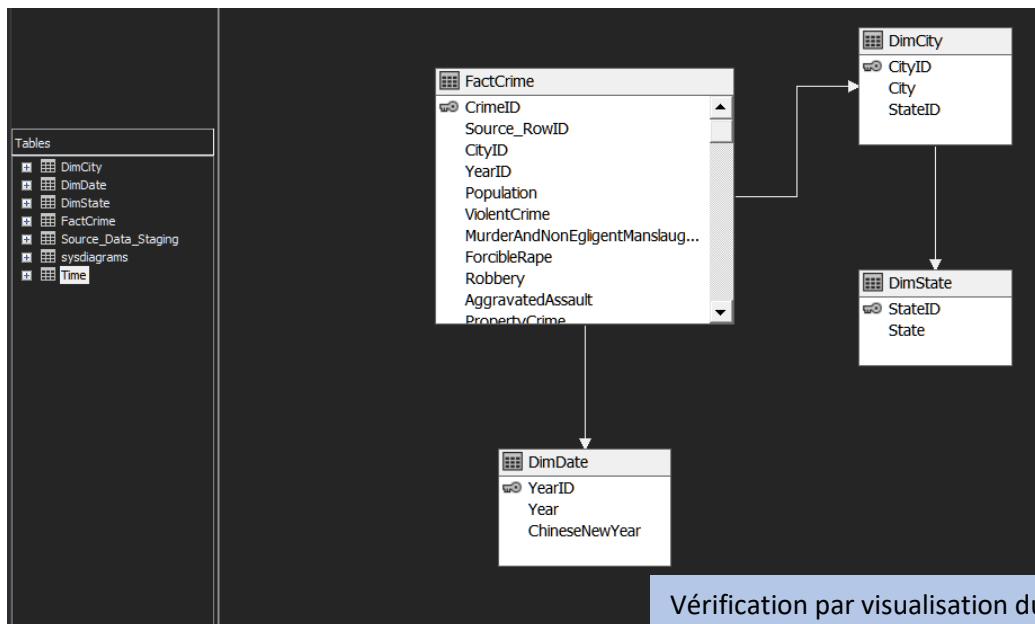




Création de la vue de source de données (DSV) : on utilise la source de données CrimeDS en sélectionnant toutes les tables préfixées par Dim et Fact.

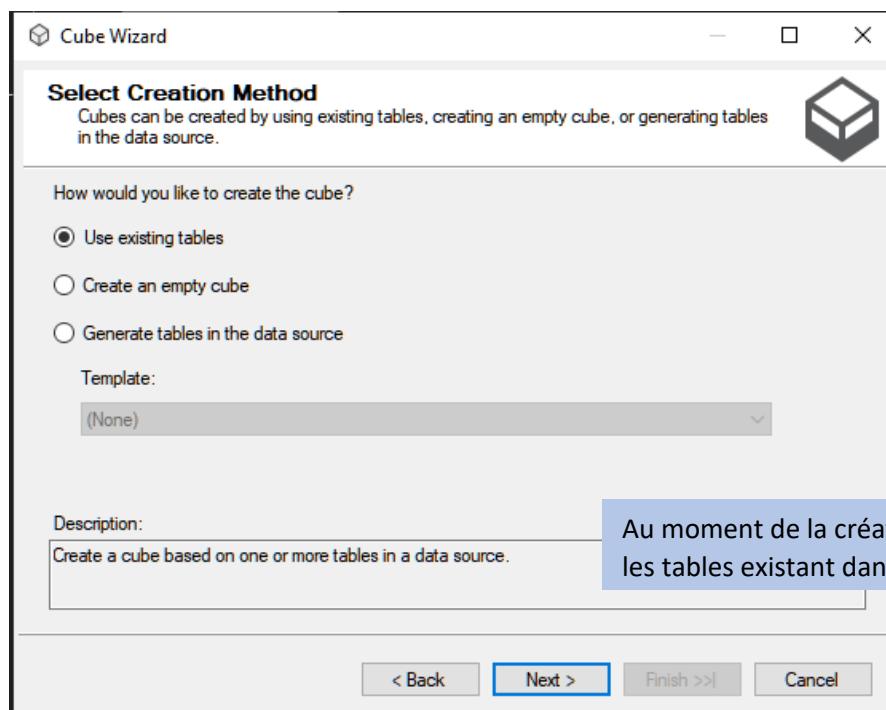


Création du Cube : Avant de créer le cube on doit valider la conformité de notre modélisation du DW.

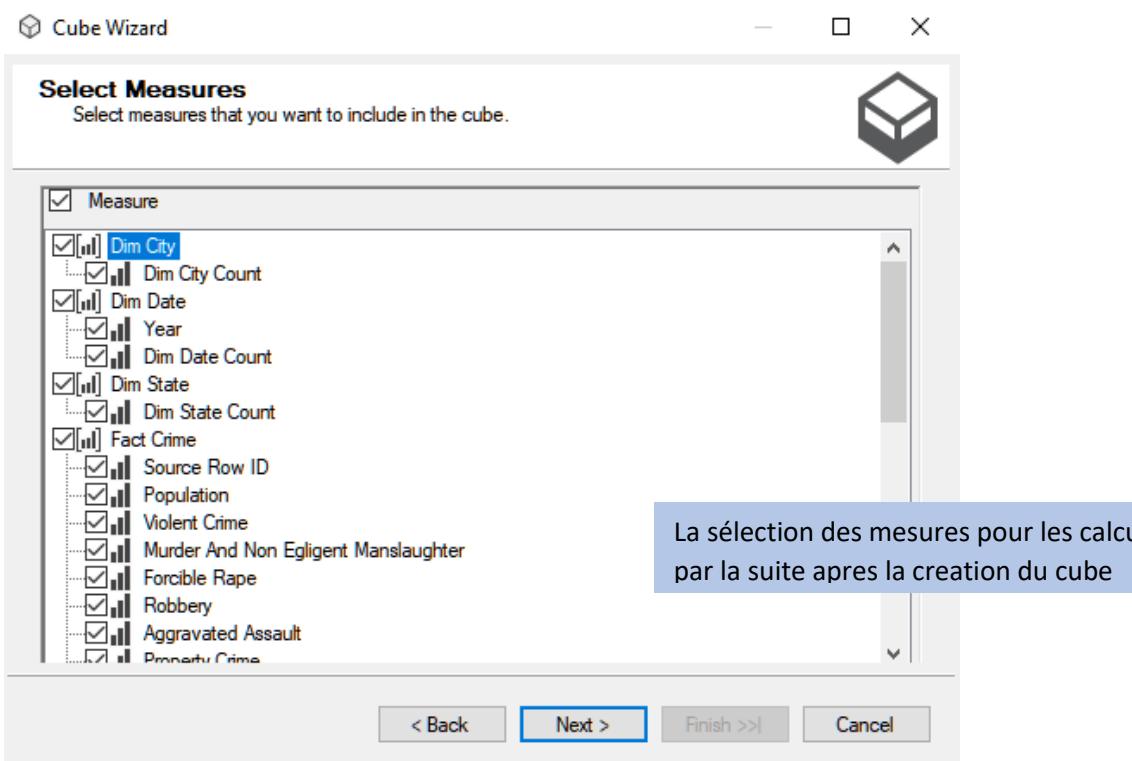
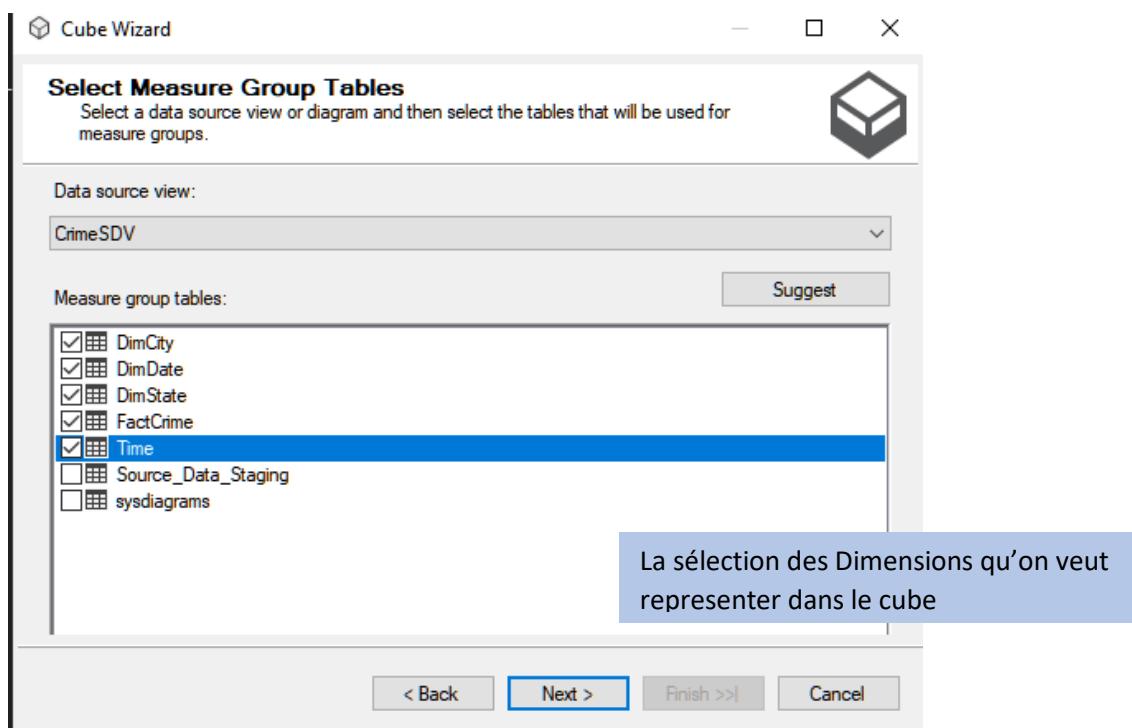


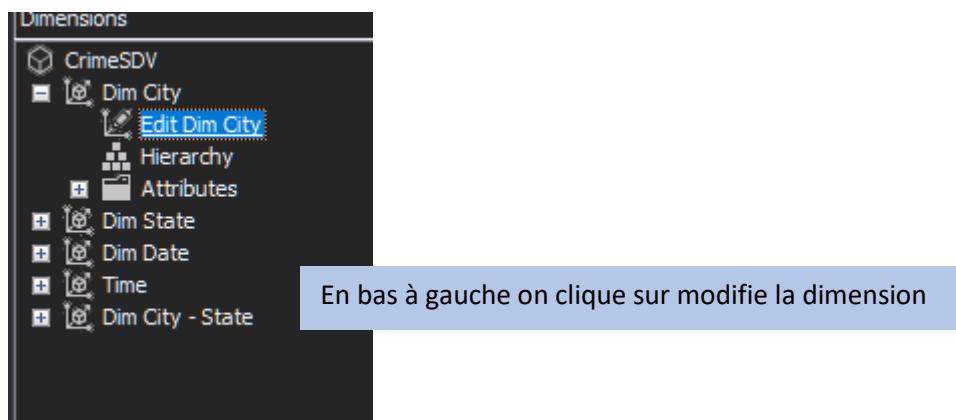
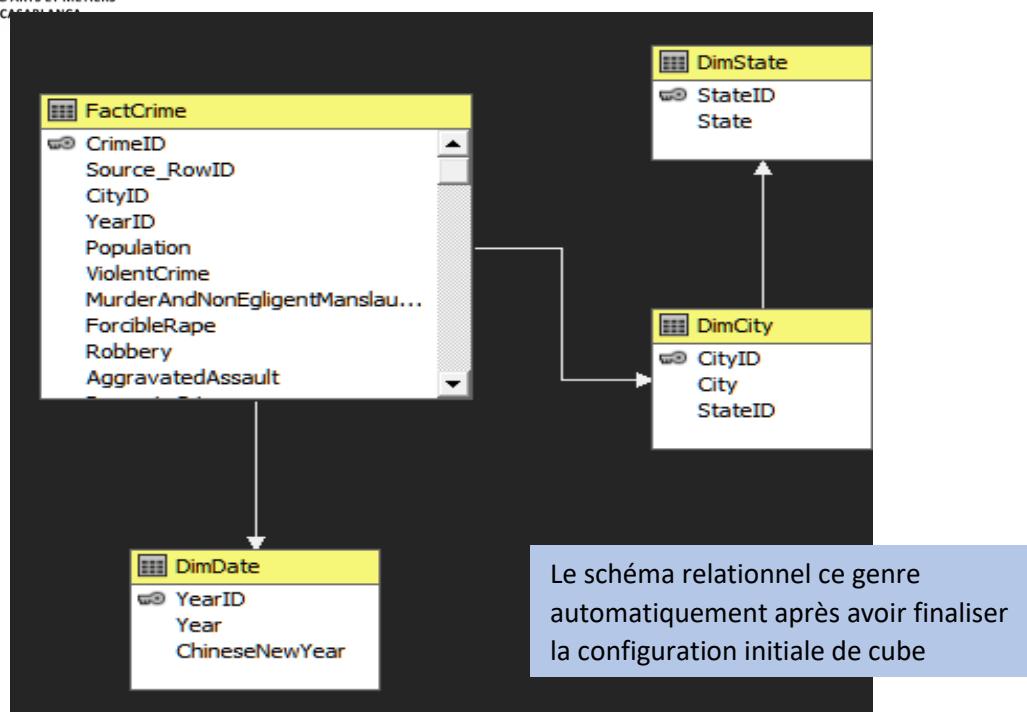
Vérification par visualisation du Schéma de notre Vue source de donnée

Créer un cube par clique droit puis nouveau cube

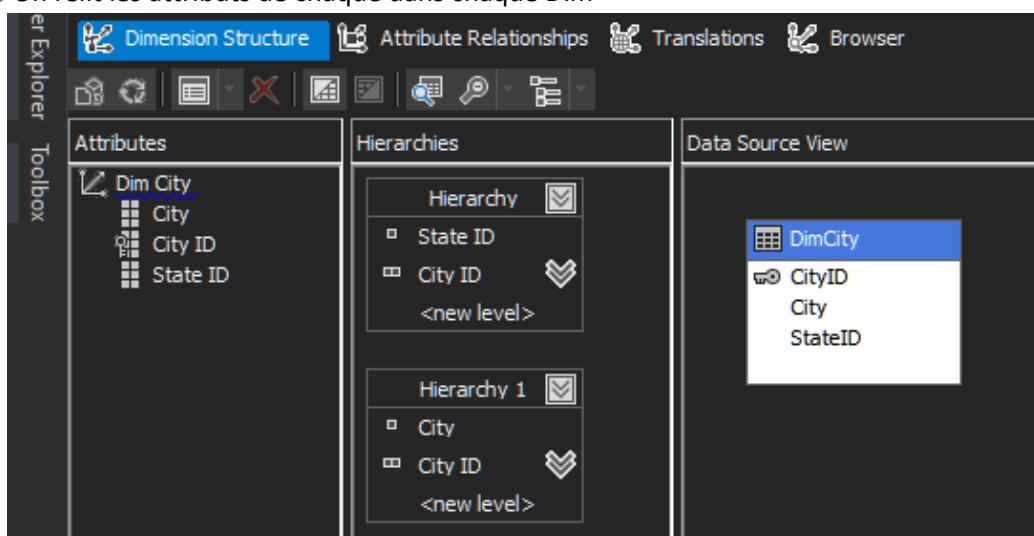


Au moment de la création du cube on choisit les tables existant dans notre data source



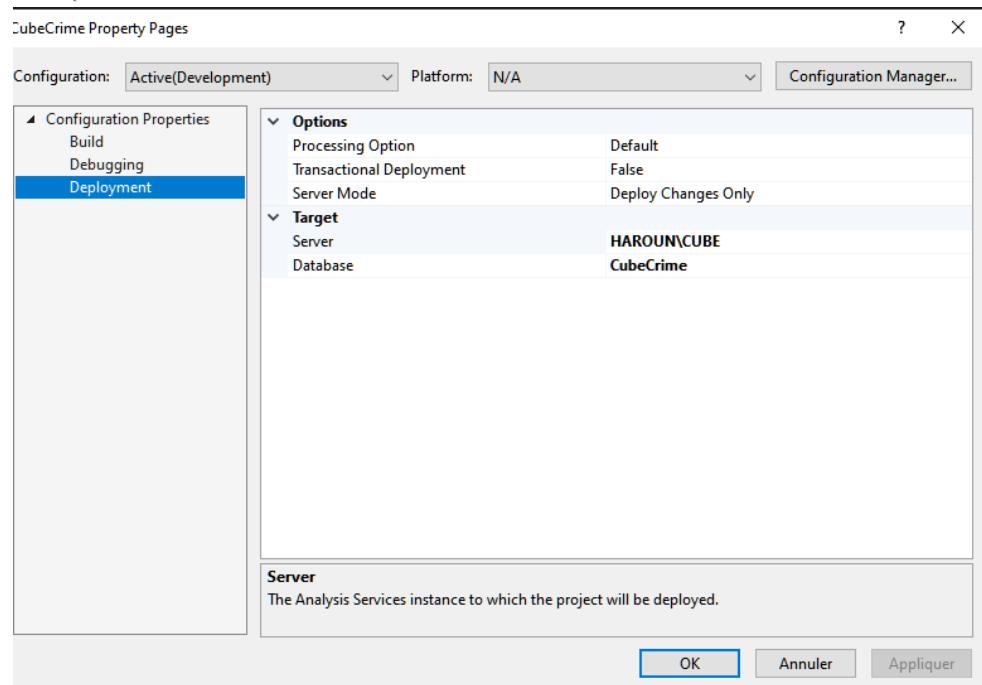


Puis On relit les attributs de chaque dans chaque Dim

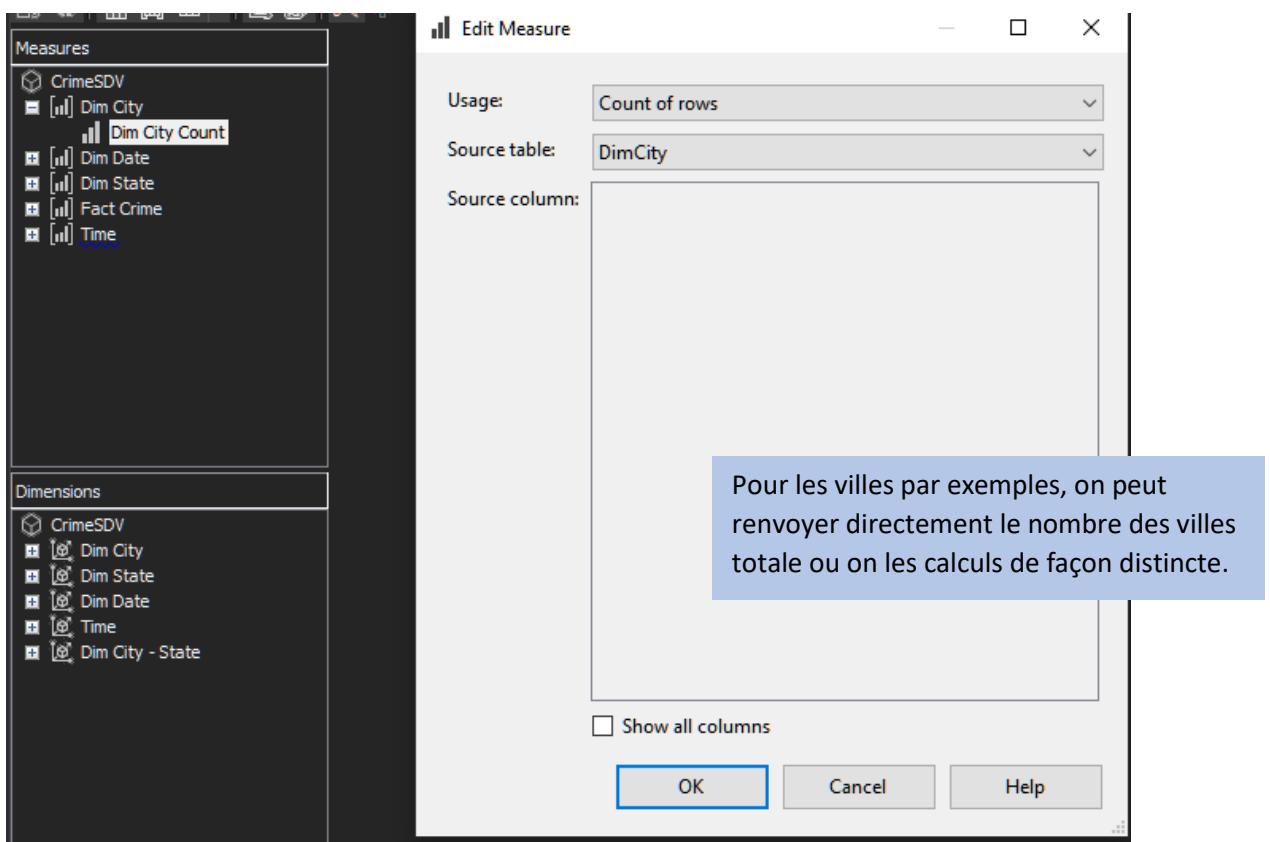


The screenshot shows the SSAS Dimension Structure editor. The top navigation bar includes tabs for Dimension Structure, Attribute Relationships, Translations, and Browser. The left sidebar has buttons for Explorer and Toolbox. The main area is divided into three panes: **Attributes** (listing Dim City, City, City ID, and State ID), **Hierarchies** (showing two hierarchies: State ID and City ID, each with a new level), and **Data Source View** (displaying the **DimCity** table with columns CityID, City, and StateID, where CityID is selected). A callout box in the bottom right corner indicates the selection of an attribute.

Maintenant après avoir créé le cube nous devons le déployer et le traiter si on veut afin d'explorer les données. Pour ce faire clic droit sur CubeCrime/Propriétés/ Deployment, au niveau du champ Serveur, on spécifier le nom et l'instance du serveur Analysis Services CUBE. Puis validez en cliquant sur OK.

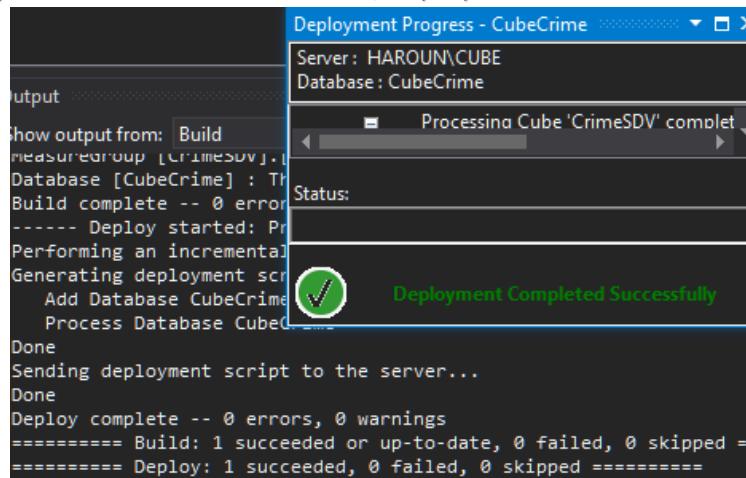


On définit l'ensemble des mesures qu'on veut :

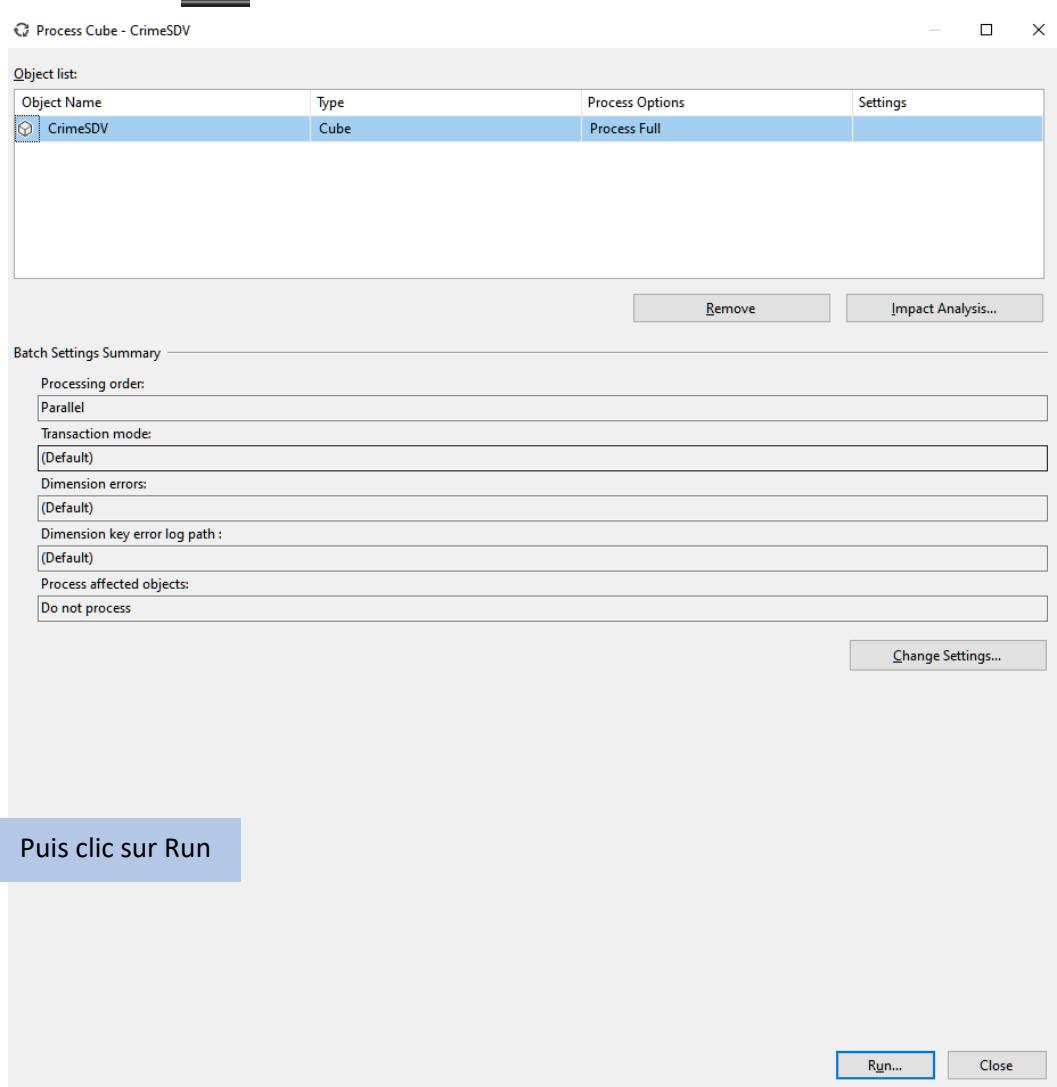


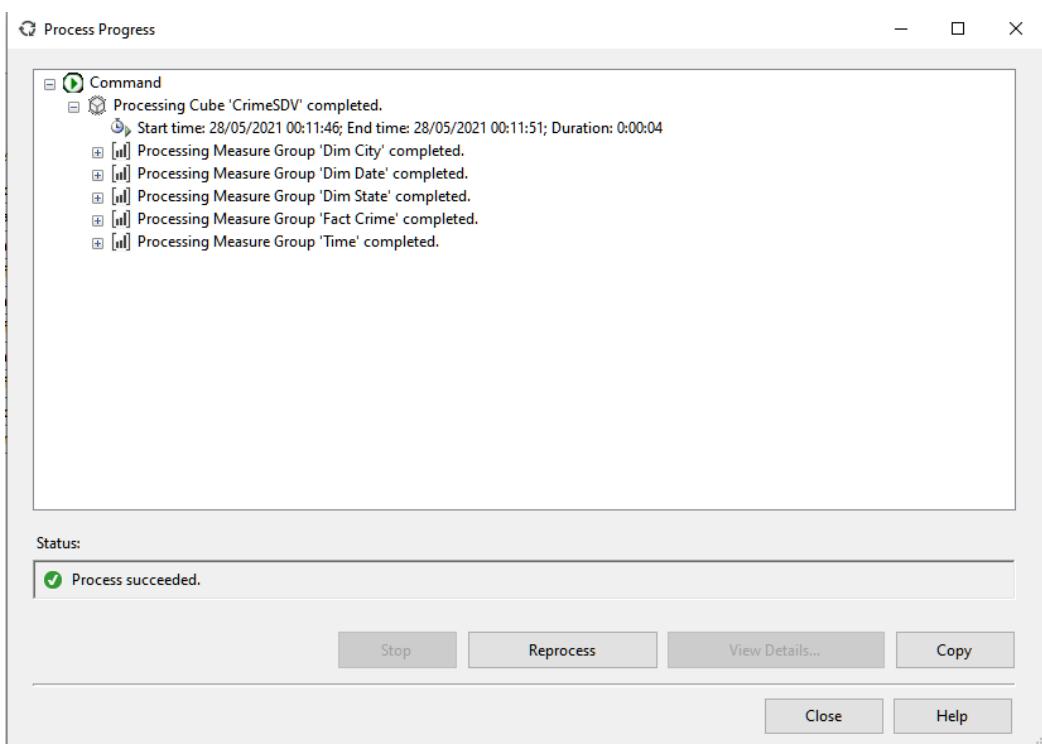
Pour les villes par exemple, on peut renvoyer directement le nombre des villes totale ou on les calcule de façon distincte.

-Pour le déploiement clic sur CubeCrime/Deploy

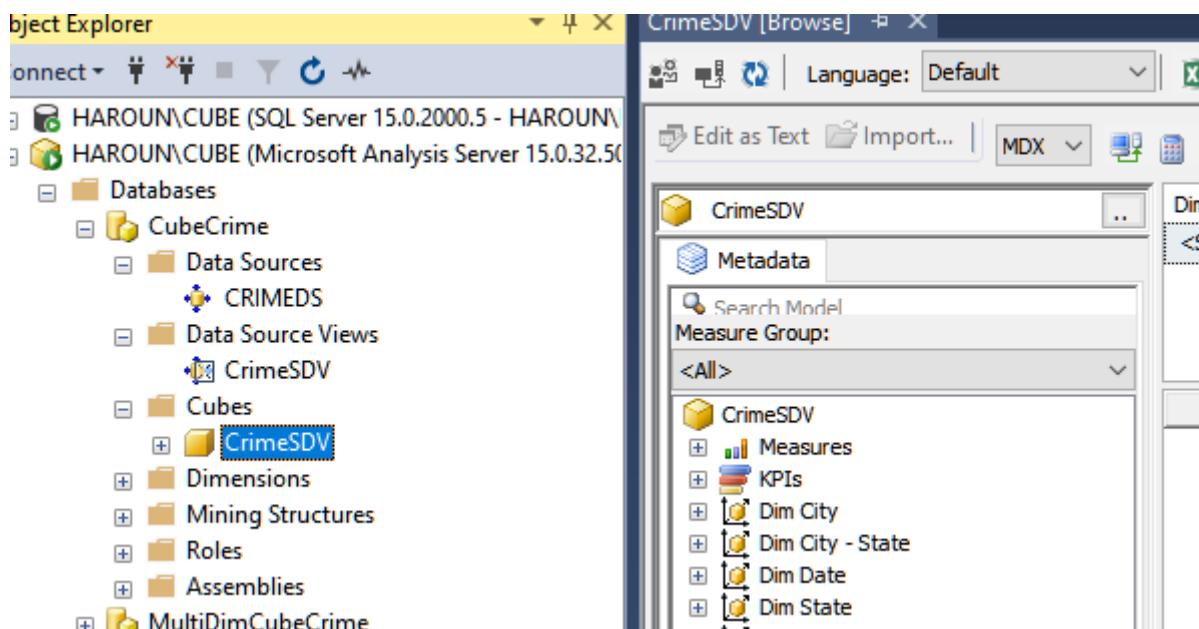


Après chaque modification des mesures pour l'appliquer sur le cube déjà déployer ; cliquons sur Process d'icone





La vérification de déploiement on clique sur Crime Cube puis browse



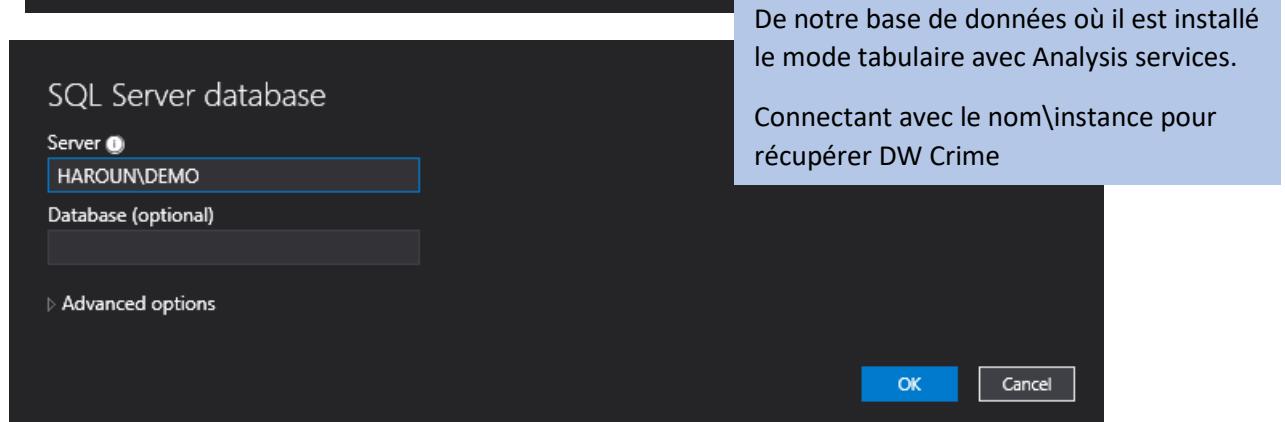
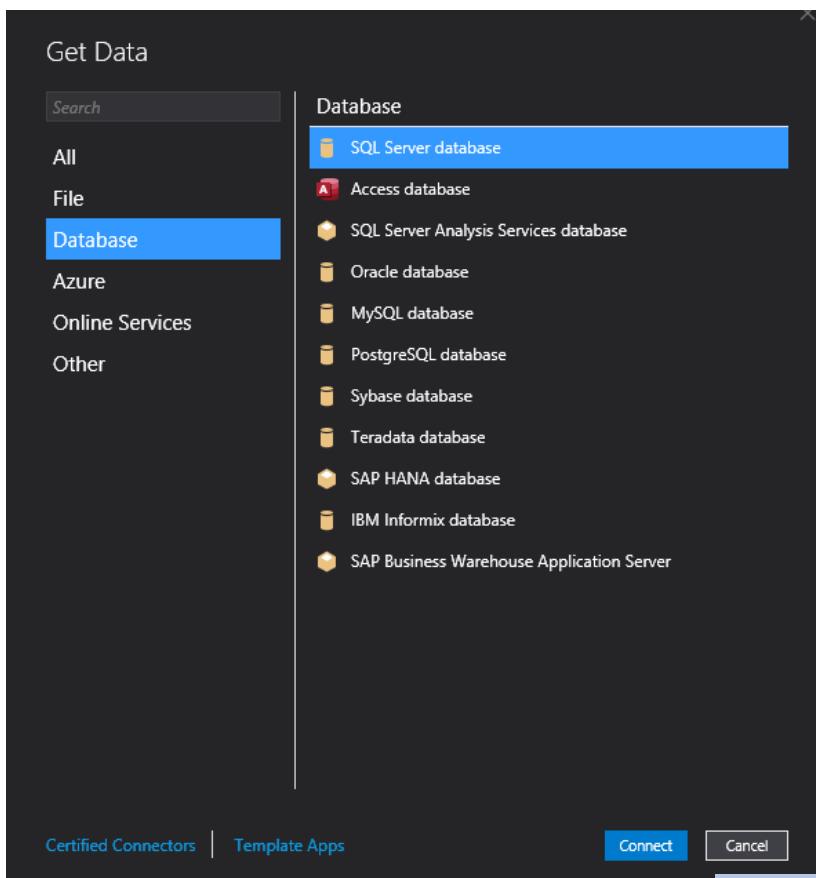
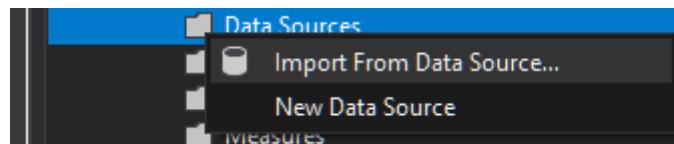
The screenshot shows the SQL Server Management Studio interface with two windows open:

- Object Explorer:** Shows the database structure. A cube named 'CrimeSDV' is selected in the 'Cubes' folder under the 'CubeCrime' database.
- CrimeSDV [Browse]:** This window displays the cube's metadata. The 'Metadata' tab is selected, showing the cube name 'CrimeSDV'. Below it, the 'Measures' tab lists the following dimensions and measures:
 - Dimensions: Dim City, Dim City - State, Dim Date, Dim State
 - Measures: CRIMEDS, CrimeSDV
 - KPIs: KPIs

Création de Cube – Tabular:

La création de notre cube tabulaire Nouveau Projet / Analysis Services Tabluar Project.

Sur source de données on choisit new data source.



On sélectionne par la suite l'ensemble des Dim qu'on veut sur notre Tabular cube

Navigator

Display Options

- Haroun\DEMO: Crime [7]
 - Crime.DimCity
 - Crime.DimState
 - Crime.DimTime
 - Crime.FactCrime**
 - Crime.Source_Data_Staging
 - sysdiagrams
 - fn_diagramobjects

Crime.FactCrime

CrimID	Source_RowID	CityID	YearID	Population	ViolentCrime
1	115856	287	2	284142	
2	115857	288	2	6488	
3	115858	289	2	1028	
4	115859	290	2	2322	
5	115860	291	2	1186	
6	115861	292	2	2494	
7	115862	293	2	31287	
8	115863	294	2	2260	
9	115864	295	2	5629	
10	115865	296	2	1992	
11	115866	297	2	30746	
12	115867	298	2	7620	
13	115868	299	2	7384	
14	115869	300	2	6242	
15	115870	302	2	1869	
16	115871	303	2	6569	
17	115872	304	2	7931	
18	115873	305	2	2890	
19	115874	306	2	3054	
20	115875	307	2	8932	
21	115876	308	2	827	
22	115877	309	2	4198	
23	115878	310	2	431	

Select Related Tables Load Transform Data Cancel

Data Processing

Processing Progress

Processing gets updated data from the original data sources.

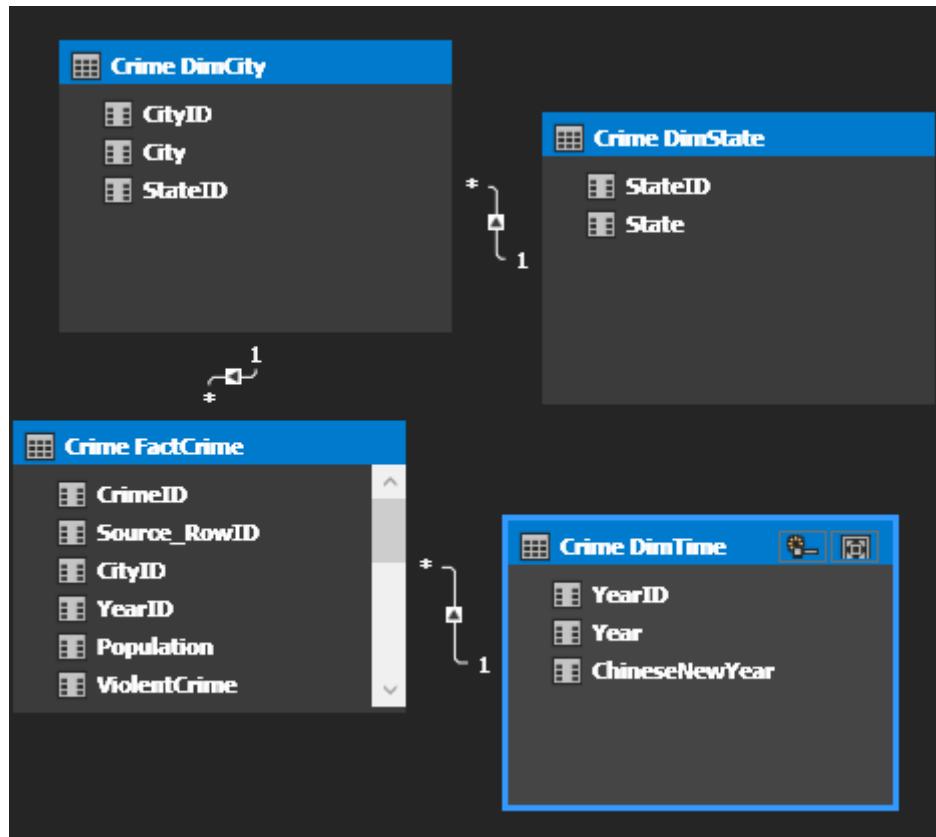
Success

4 Total	0 Cancelled
4 Success	0 Error
Work Item	Status
✓ Crime DimCity	Success, 9 615 rows transferred.
✓ Crime DimState	Success, 51 rows transferred.
✓ Crime FactCrime	Success, 25 683 rows transferred.
✓ Crime DimTime	Success, 3 rows transferred.

[Stop Processing](#) [Close](#)

Le chargement des dimensions est complet

A partir de l'icône Diagramme, voire clairement les relations avec types entre chaque Dim.

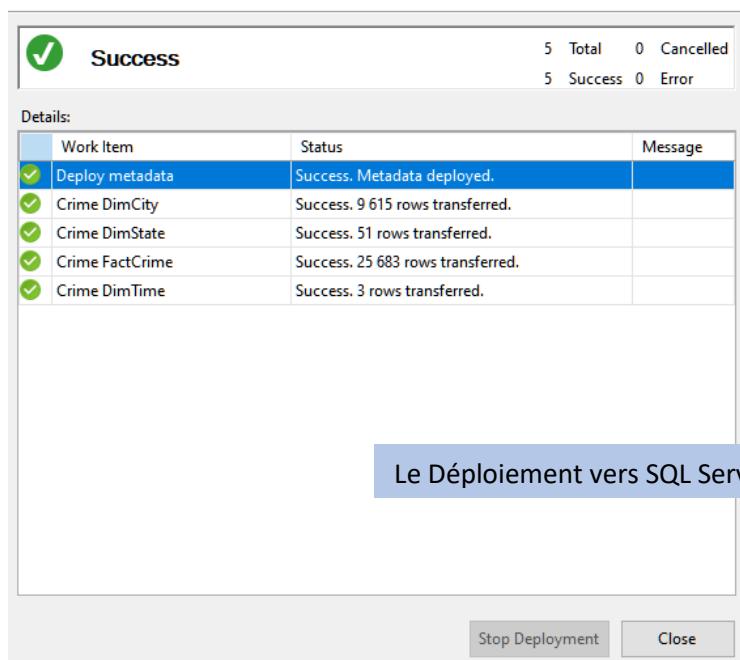


Clic sur le nom tabCube/propriétés/deployment on vérifie notre serveur puis on clique sur Deploy.

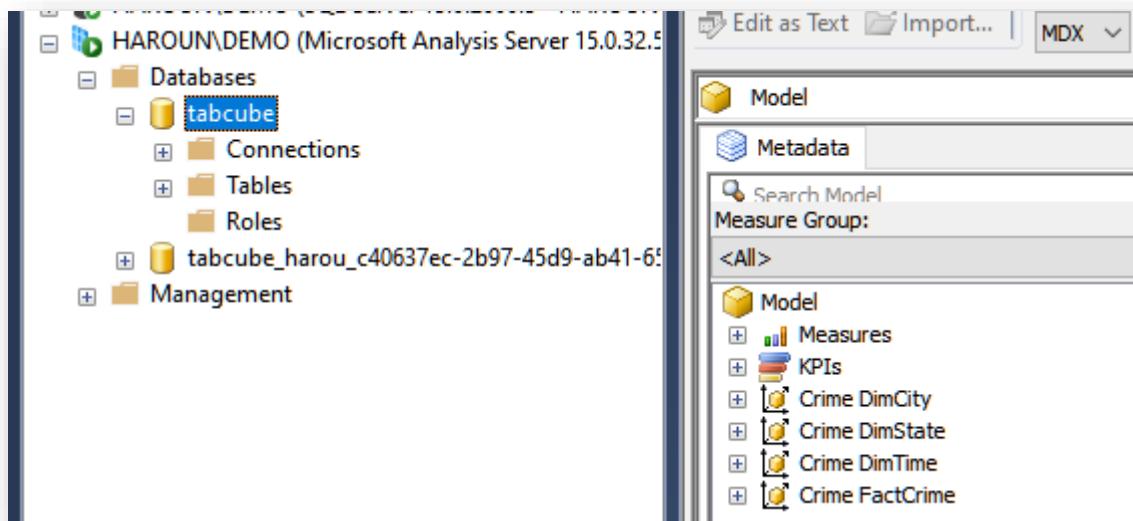
Deploy ? ×

Deploying

The deployment operation may take several minutes to complete.



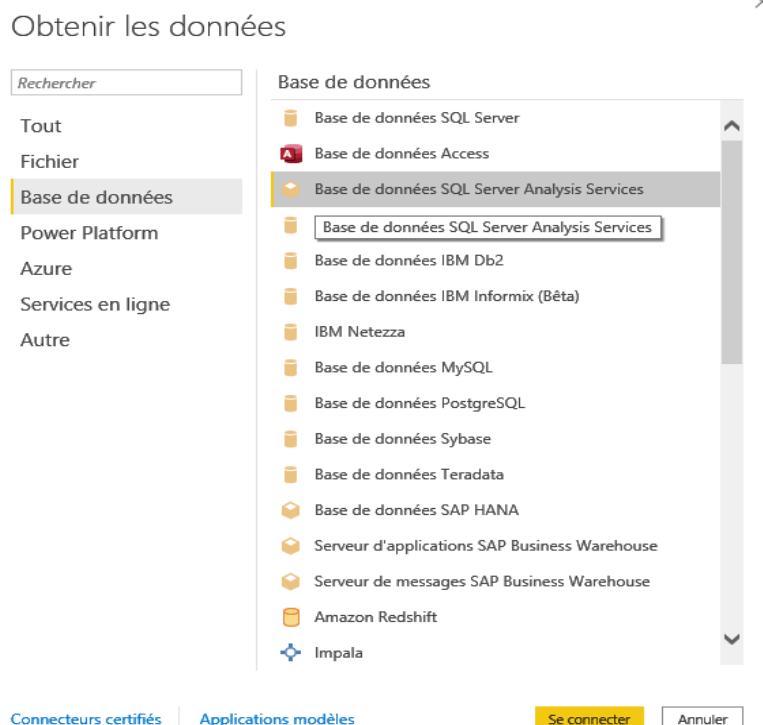
Pour la vérification on Browse notre cube tabcube pour tester les mesures et les Dim.



Visualisation des données sur Power BI

Où on va vérifier qui est la meilleur pour notre data tabular ou multisimensionnel ;

⇒ Commençant par l'obtention de données :



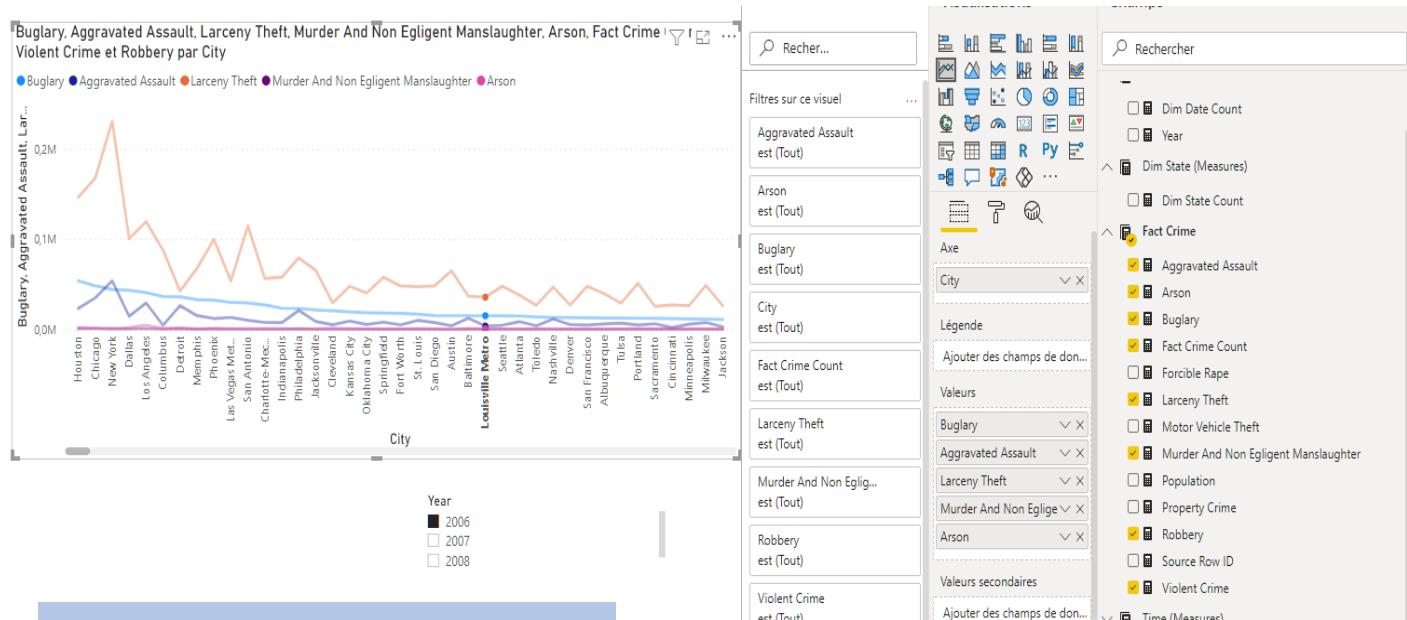
Navigateur

Options d'affichage

- ▲ HAROUN\CUBE [2]
- ▷ CubeCrime [1]
- ▲ MultiDimCubeCrime [1]
- ▲ Crime [1]
- ▲ Crime [10]
- ▷ Dim City [1]
- ▷ Dim Date [2]
- ▷ Dim State [1]
- ▷ Fact Crime [13]
- ▷ Time
- ▷ Dim City [1]
- ▷ Dim City - State [2]
- ▷ Dim Date [3]
- ▷ Dim State [2]
- ▷ Time [15]

Les Dim nécessaire pour la visualisation

Dim State
[Dim State]



⇒ La réalisation finale :

- Page d'accueil Permet de naviguer entre les pages de projets (les crimes dangereux, les crimes de vol, À-propos, Nous).

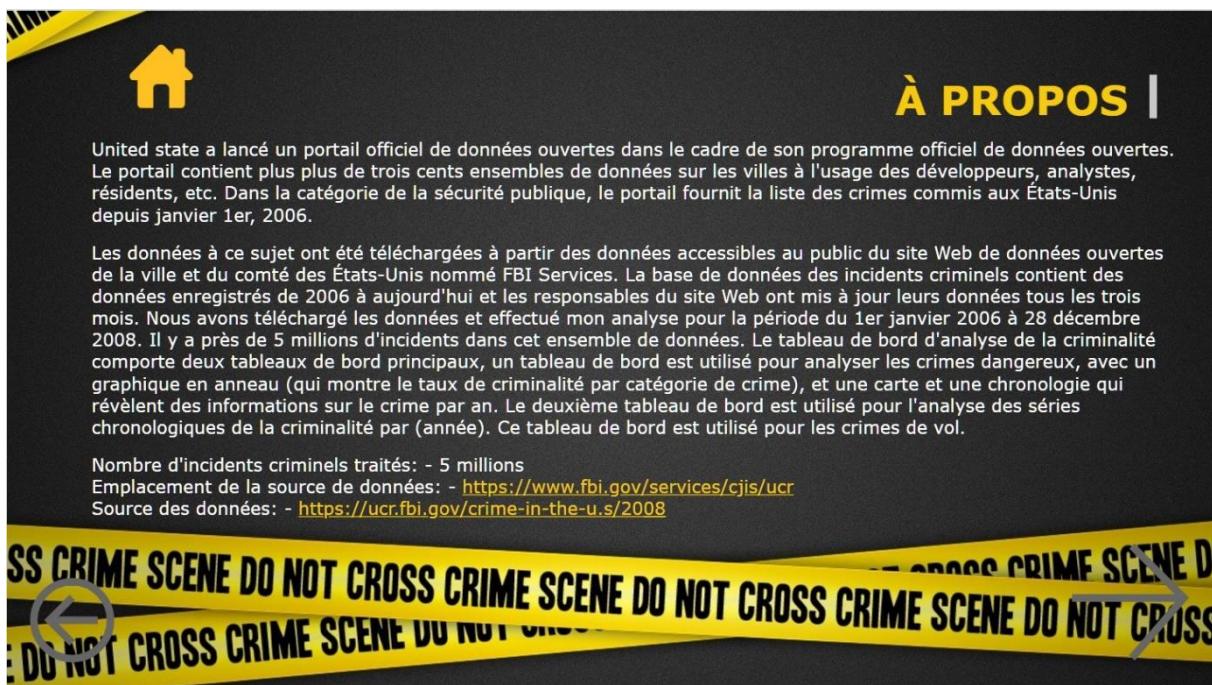
- Page 2, les crimes dangereux :



- Page 3, les crimes de Vol :



- A propos de notre data :



- A propos de nous :



Conclusion :

Dans le but de découvrir le monde d'analyses des crimes, le choix de données est très important pour bien extraire des informations pertinentes en se basant sur des outils de stockage, analyse et reporte, pour notre étude sont SQL SERVER, VISUAL STUDIO et POWER BI. Au but de jouer avec les diagrammes, cartes, filtres sur un tableau de bord qui résume et facilite la lecture de notre rapport. C'est pour cela on a passé par ensembles des étapes ; importation de données création d'ETL création de cube sous les deux modèles. Dans les modèles multidimensionnels, il s'agit d'une analyse par cube OLAP plus classique. À partir des tables de faits, il faut construire les différentes mesures et dimensions ou axes qui vont permettre l'analyse multidimensionnelle, créer les hiérarchies. Mais dans les modèles tabulaires, les tables s'organisent d'une manière équivalente aux bases de données relationnelles. L'avantage d'un modèle tabulaire est qu'il est beaucoup moins dépaysant car son interface ressemble beaucoup à Excel et repose sur le langage DAX.

Références :

- [**https://docs.microsoft.com/en-us/sql/integration-**](https://docs.microsoft.com/en-us/sql/integration-services/ssis-how-to-create-an-etl-package?view=sql-server-ver15)
[**services/ssis-how-to-create-an-etl-package?view=sql-server-**](#)
[**ver15**](#)
- [**https://docs.microsoft.com/en-us/system-center/scsm/olap-**](https://docs.microsoft.com/en-us/system-center/scsm/olap-cubes-overview?view=sc-sm-2019)
[**cubes-overview?view=sc-sm-2019**](#)
- [**https://www.fbi.gov/services/cjis/ucr**](https://www.fbi.gov/services/cjis/ucr)
- [**https://ucr.fbi.gov/crime-in-the-u.s/2008**](https://ucr.fbi.gov/crime-in-the-u.s/2008)