**TP1** : SQL Server

SQL Server BI permet la conception d’une solution décisionnelle suite à l’utilisation des principaux outils

|  |  |
| --- | --- |
| Institut Supérieur d'Informatique  **Auditoire :** 3ème année LSI - ADBD  **Module :** Entrepôts de données  **Année universitaire :** 2024-2025  et de Multimédia de Sfax |  |

de Business Intelligence de Microsoft tels que *Integration Services*, *Analysis Services*, *SQL Server Reporting Services*, *les services « SQL Server R Services » pour la Data Science* et *PowerPivot*.

Dans ce TP, nous allons pratiquer les deux outils d'informatique décisionnelle de Microsoft : *Integration Services* et *Analysis Services*.

1. **Objectifs du TP :**

* Déployer les paquets ETL avec *Integration Services (SSIS).*
* Construire des solutions d’exploration de données en utilisant les Data Cubes avec *Analysis Services (SSAS).*

1. **Modélisation multidimensionnelle :**

La modélisation multidimensionnelle est une méthode de conception logique qui vise à présenter les données sous une forme standardisée et intuitive. Cette modélisation considère les données comme des points dans un espace à plusieurs dimensions. Ces points représentent les centres d’intérêts décisionnels (sujets) analysés en fonction des différents axes d’analyse. Un modèle dimensionnel se compose alors d’un **fait** et d’un ensemble de **dimensions**.

1. **Enoncé de l’exercice :**

Dans ce TP, vous êtes invités à concevoir un schéma multidimensionnel et mettre en place un entrepôt de données pour le compte du magasin de gros.

La base de données Source est constituée des tables et relations suivantes :

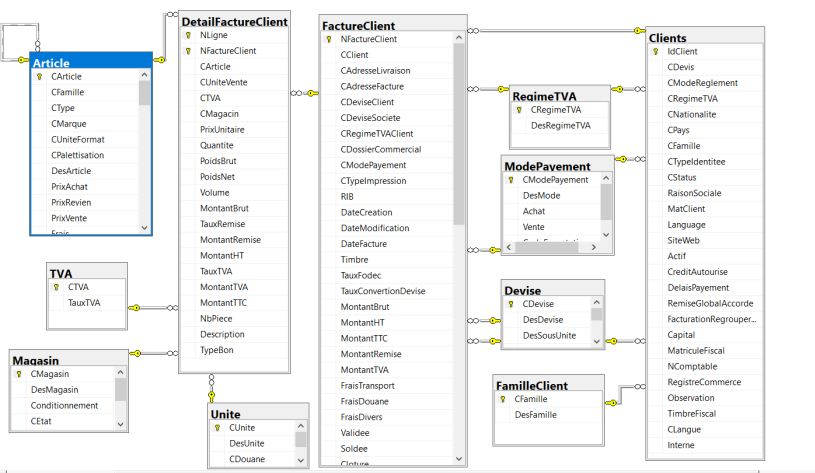


Figure Schéma relationnel de la base de données source

Nous désirons créer l’entrepôt de données permettant d’effectuer l’analyse des ventes réalisées auprès des clients d’un magasin de gros. Plus précisément, pour chaque vente de produit à un client, nous souhaitons disposer les critères d’analyse suivants :

* Le montant Brut de vente par *client*, par *article* et par *jour*.
* La quantité de produits (nombre d’exemplaires) vendue.
* Le montant de remise par article.

Cette analyse devra s’effectuer selon un ensemble d’axes d’analyse :

* L’axe *D\_Date* regroupant les dates des ventes (sous forme d’une hiérarchie).
* L’axe *D\_Article* regroupant la désignation, la catégorie, les sous catégories et la marque des différents articles.
* L’axe *D\_Client* regroupant les clients par raison sociale, par famille et par pays.
* L’axe *D\_Magasin* regroupant les magasins de vente.

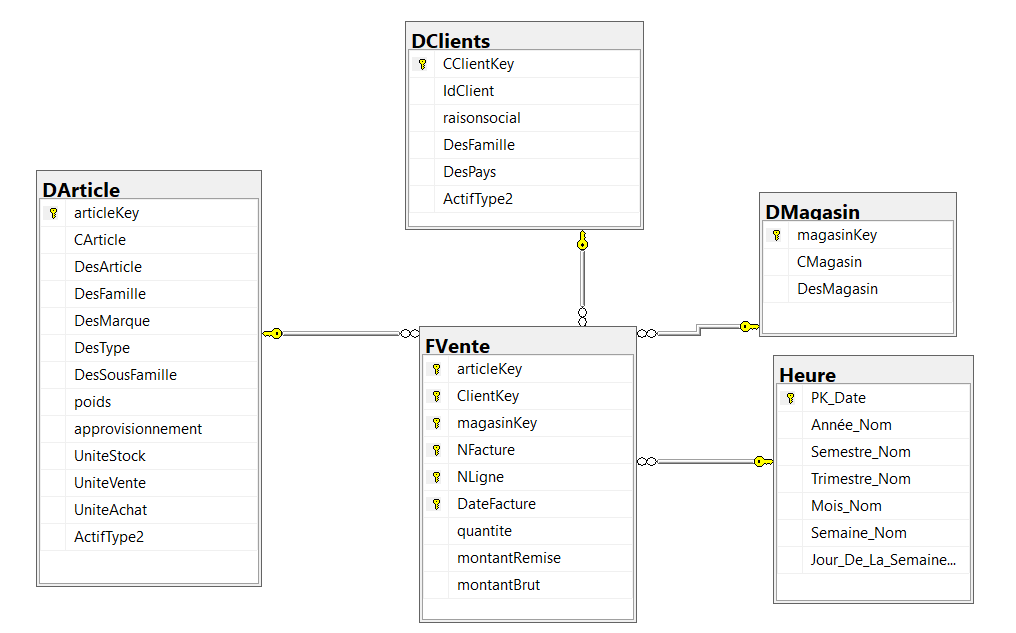
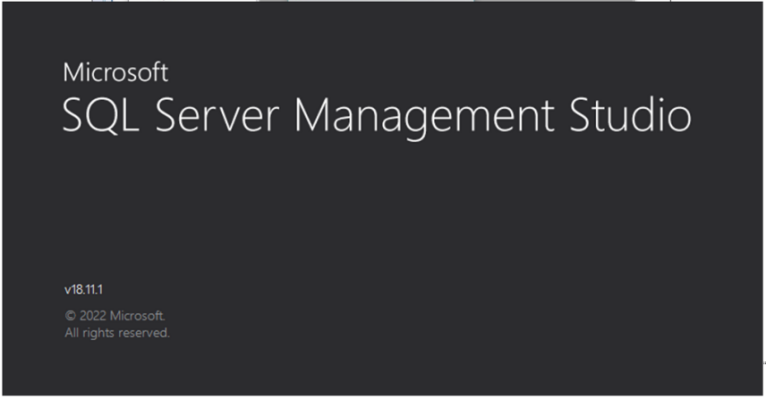
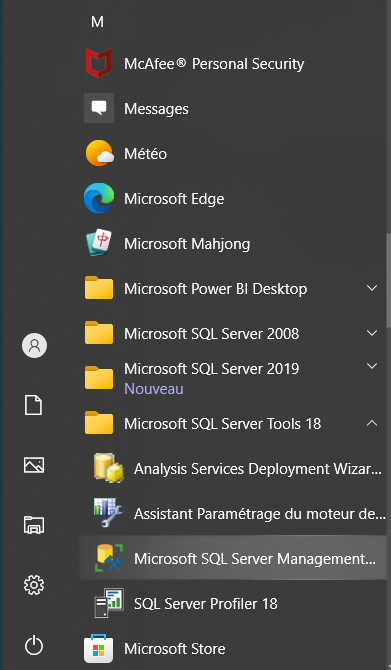


Figure Schéma en étoile de l’entrepôt de données conçu

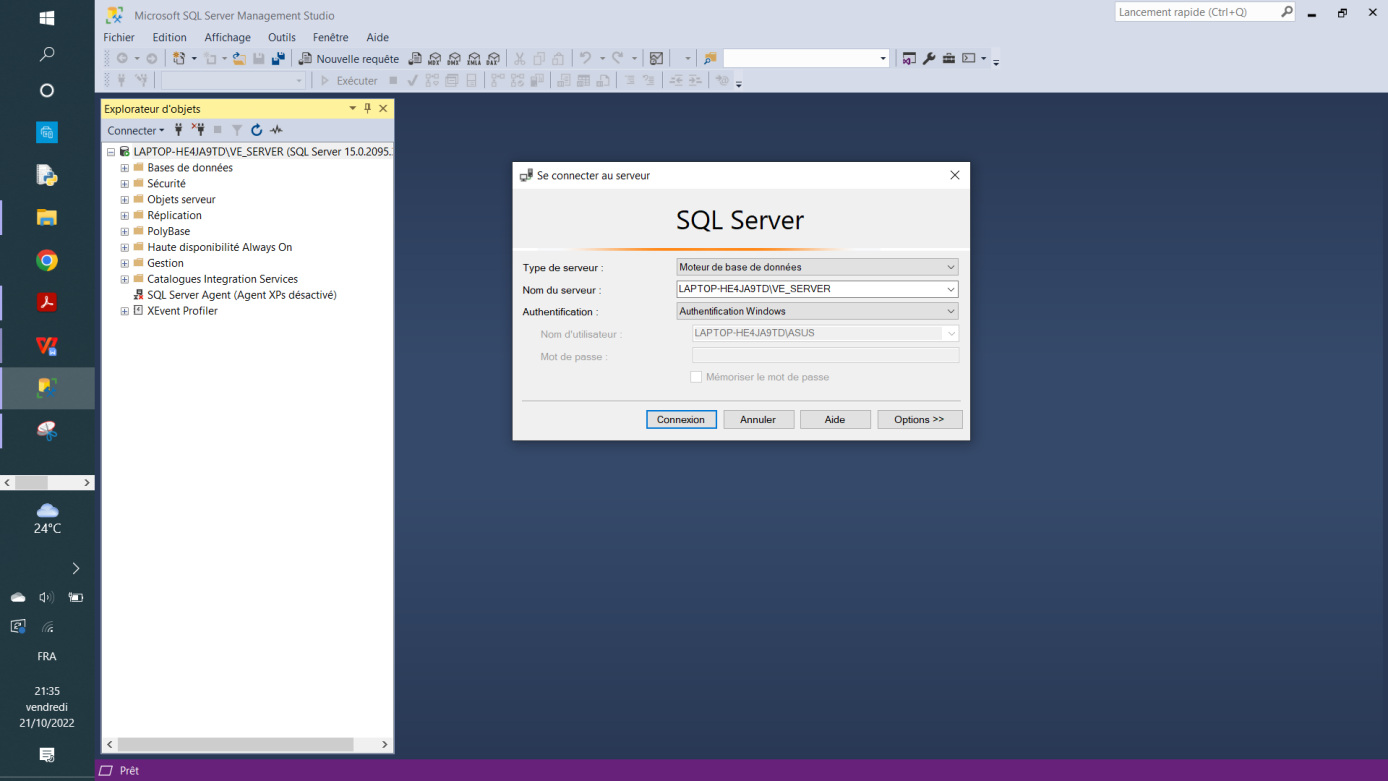
1. Installations nécessaires:

Nous avons besoin d’installer 2 outils :

* SQL Server 2019 : plusieurs composants doivent être installés :
* SQL Integration Services.
* SQL Analyses Services.
* MS Server Management.
* Visual Studio 2019 : deux extensions doivent être ajoutées :
* SQL Server Integration Services projects.
* Microsoft Analysis Services Projects.
* Ouvrir **« SQL Server Management Studio »**



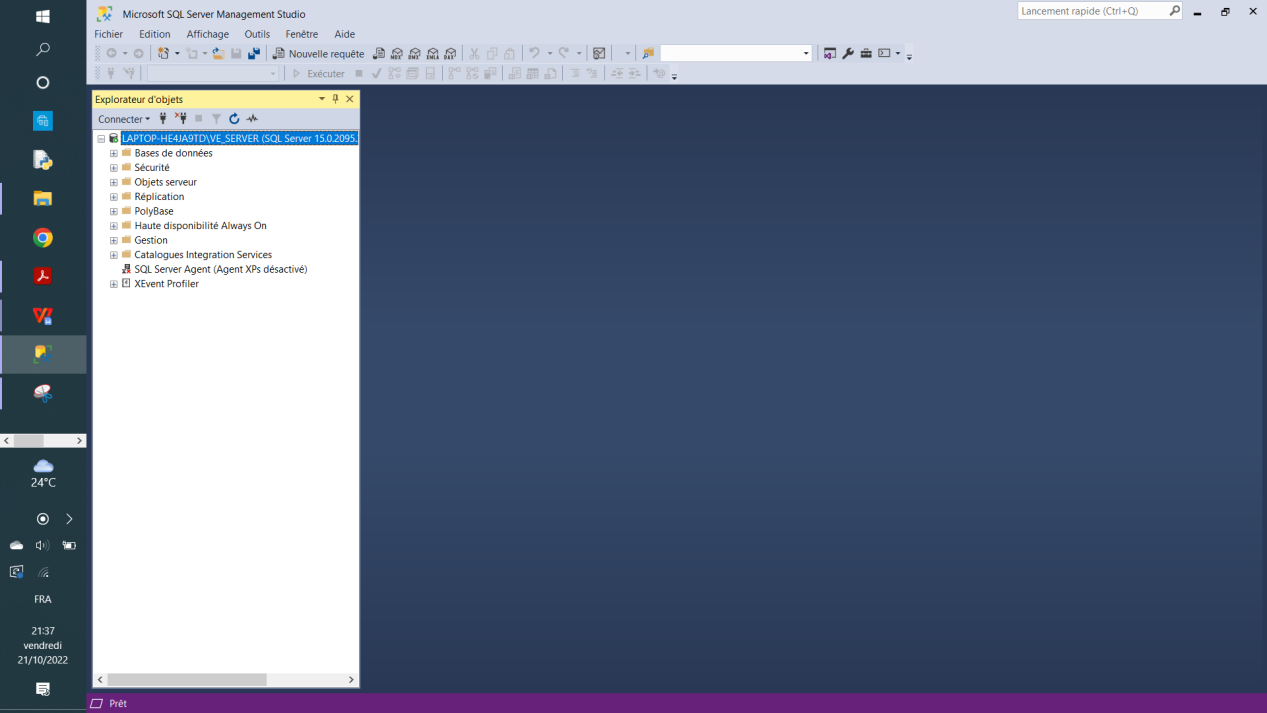
* **Démarrage de Microsoft SQL Server Management Studio**.



Dans la fenêtre de connexion de SQL Server Management nous allons introduire les informations nécessaires pour la connexion. Cette dernière peut être réalisée selon deux modules de connexion :

* La **connexion windows** : qui utilise le compte connecté de Windows.
* La **connexion SQL Server** : basée sur un login et un mot de passe.
  + Dans « Server Name » écrire **« (local) »** ou **« . »** définissant le serveur local sur lequel SQL Server est connecté.
  + Dans « Connexion » écrire **« sa »**.
  + Dans «Mot de passe » écrire **« 123 »**.

**NB.** Ne pas oublier de mémoriser le mot de passe.

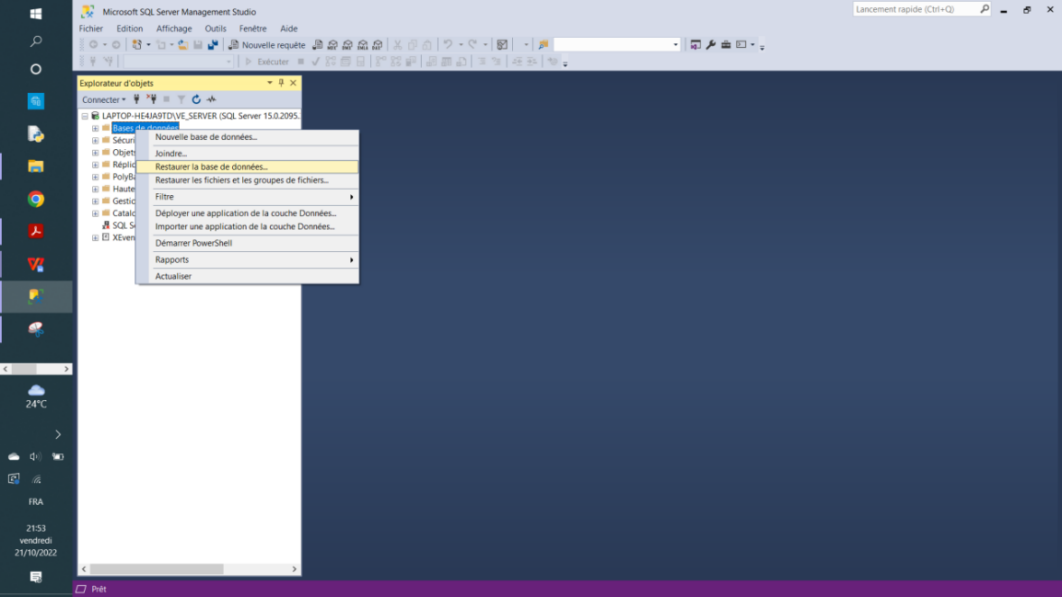


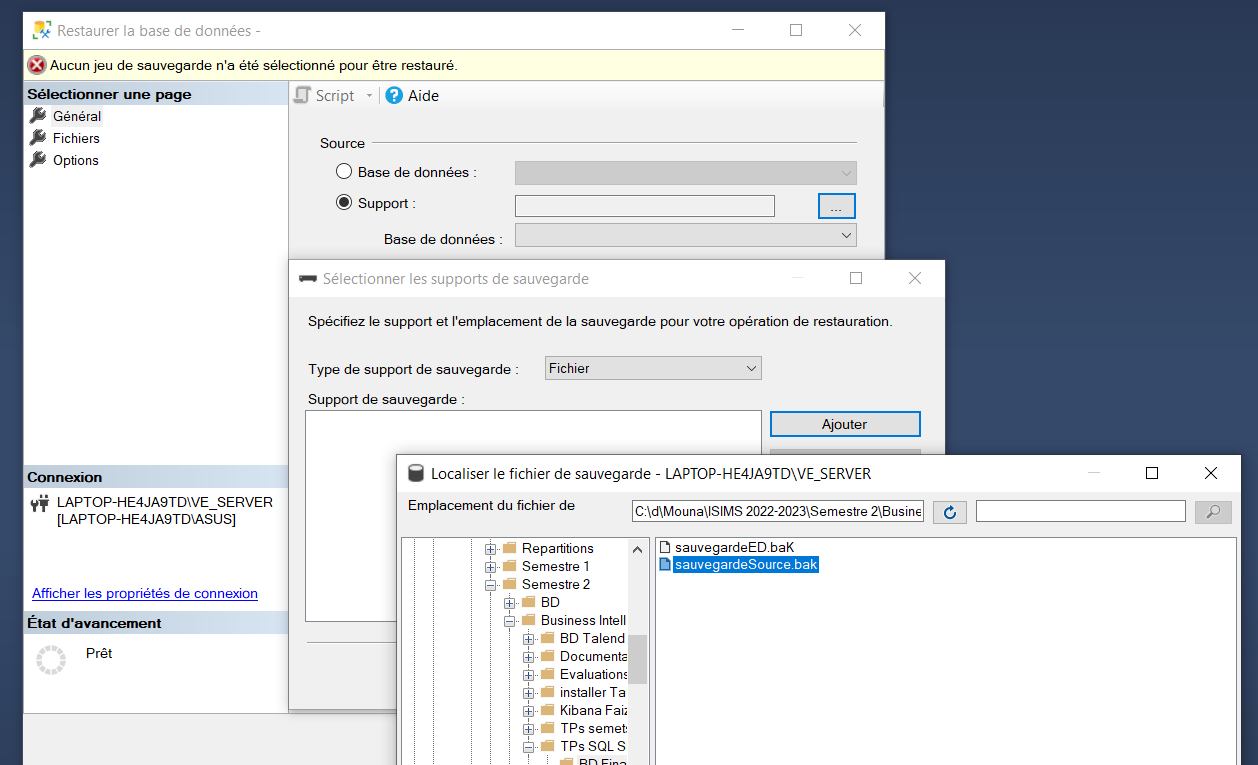
Une fois SQL Server réussi son démarrage, en défilant l’explorateur d’objets nous pouvons remarquer qu’il détient différents objets : *bases de données, sécurité, gesion,* ***SQL Server Agent****, etc.*

* L’objet « **Bases de données** » permet la création ou le téléchargement des bases de données (sources et cibles).
* L’objet **« SQL Server Agent »** désigne le service Microsoft Windows permettant d’exécuter des tâches administratives planifiées dans SQL Server. Cet Agent, à travers des jobs, permet de stocker les informations selon les planifications réalisées par les analystes. Les Jobs sont composés par une ou plusieurs étapes de travail et chaque étape est composée en elle même par ses propres tâches (exp. sauvegarde d'une dimension dans une base de données cible).

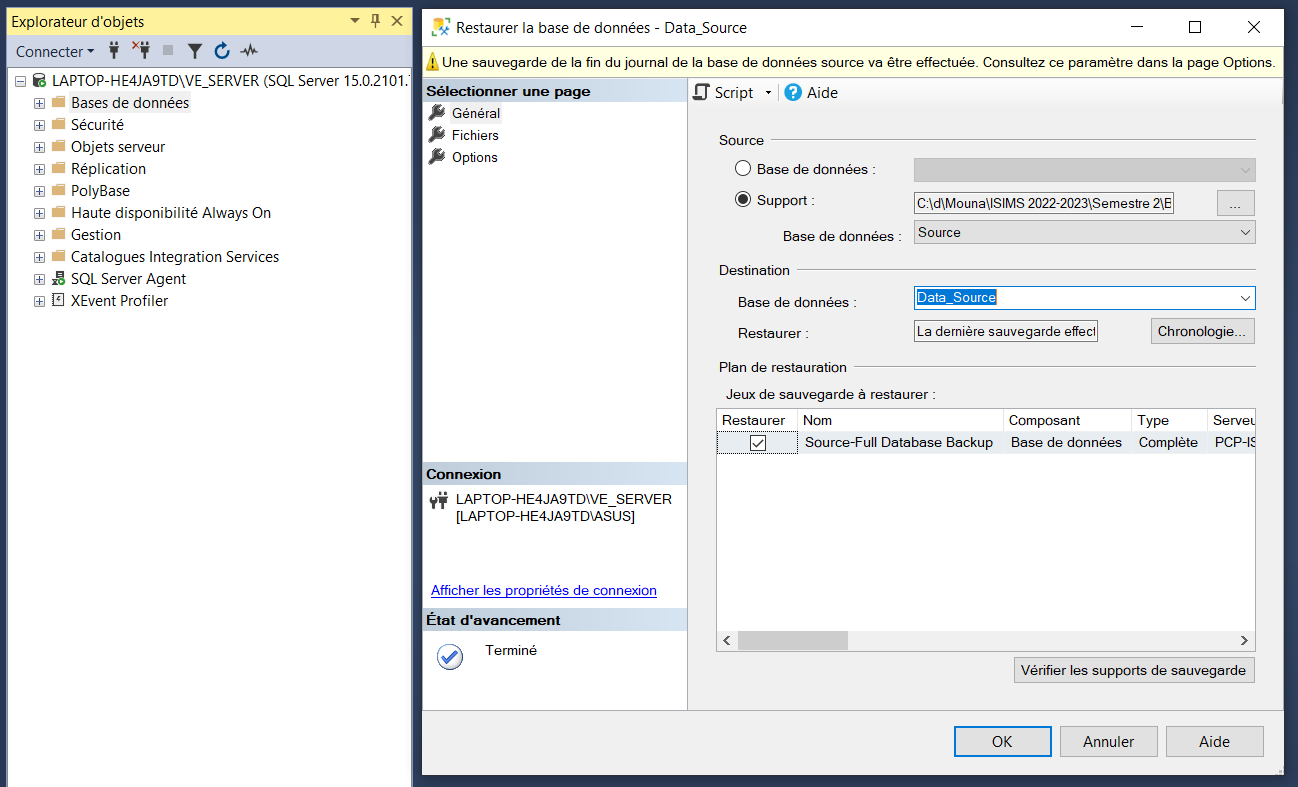
1. Atelier SSIS
   1. **Restauration des bases de données : source et cible**

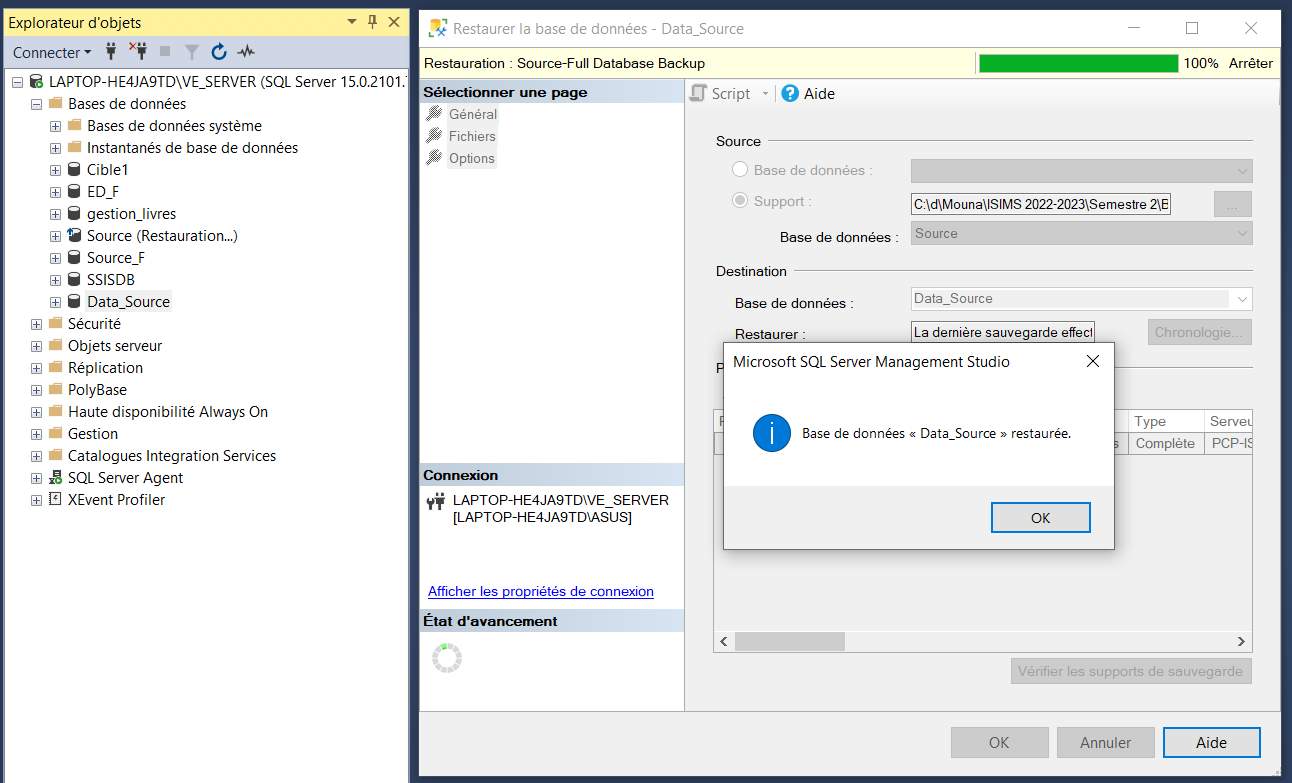
* Dans **« SQL Server Management Studio »** nous allons accéder à deux bases de données : **Source** et **Cible**.
* La première représente, comme l’indique son nom, la source des données.
* La deuxième contient un nombre restreint de tables qui représentent notre entrepôt (par défaut les tables sont vides).
* Cliquer avec le bouton droit sur « **Bases de données**» et choisir **« Restaurer la base données »**





* Ajouter, à partir de votre dossier de travail, la base **« sauvegardeSource.bak »** etnommer là « **Data\_Source** », puis valider par OK.





* Refaire le même travail afin de restaurer la base **« sauvegardeED.bak »** en nommant la base **« Data\_Target ».**

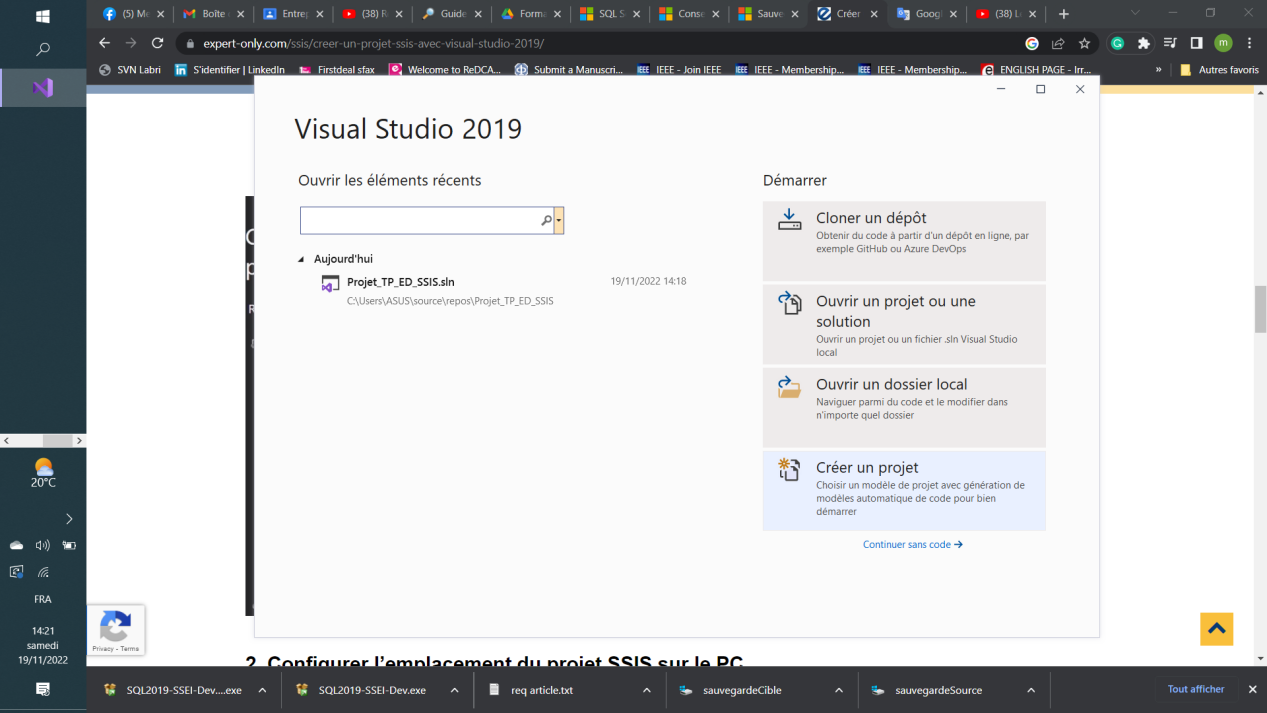
Une fois les bases de données sont restaurées dans « Microsoft SQL Server Management studio 18» nous allons passer à « Visual Studio 2019 ». Cet outil permet la création des solutions analytiques fournissant des informations significatives à visualiser en se référant à différents outils de visualisation de données, tels que *Power Pivot*, *Power BI* et *Excel*.

* 1. **Création de projet SSIS**

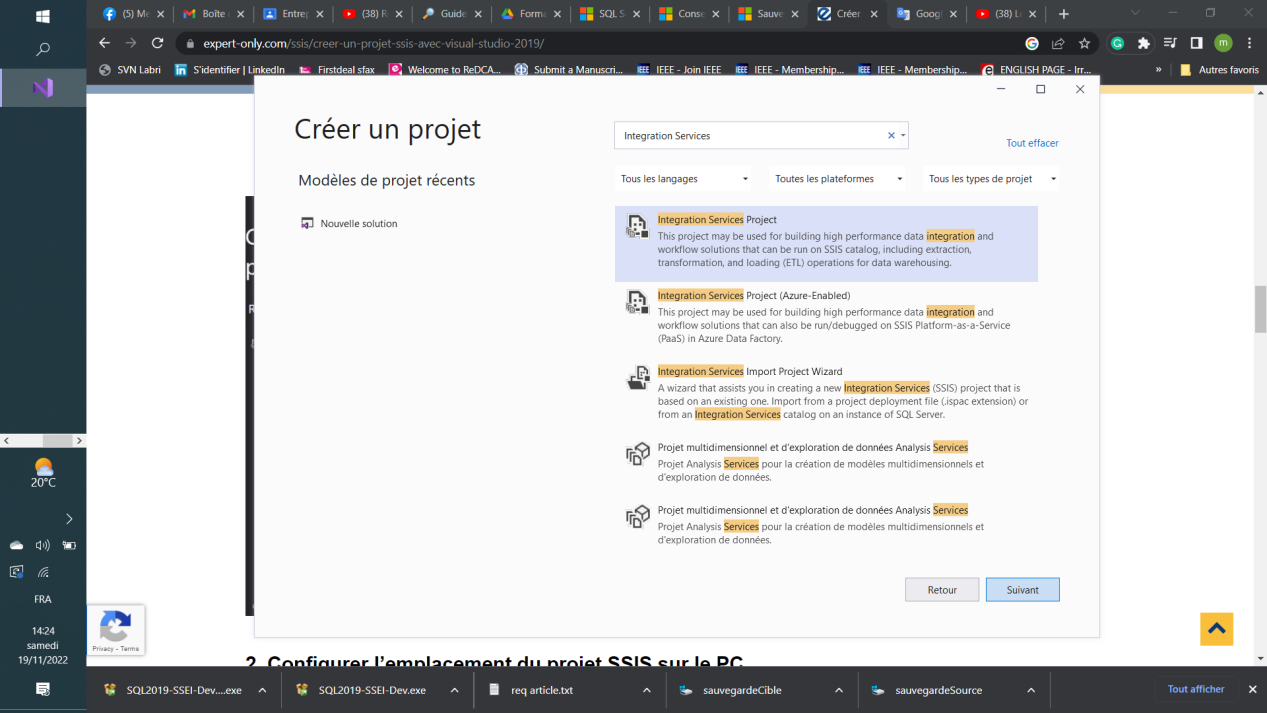
SSIS (SQL Server Integration Services), est une plateforme d'intégration de données développée par Microsoft. SSIS est intégré à l'environnement de développement de Visual Studio et permet aux utilisateurs de concevoir des packages d'intégration de données qui peuvent être utilisés pour automatiser les processus ETL (extraction, transformation, chargement) des données provenant de différentes sources de données, telles que des fichiers plats, des bases de données relationnelles et des sources de données en nuage. Les packages SSIS peuvent également être utilisés pour exécuter des tâches de maintenance et de gestion des données, telles que la sauvegarde et la restauration de bases de données.

Sous Visual Studio, nous allons commencer par la création d’un nouveau projet SSIS.

- Cliquez sur «**Créer un projet**».



- Ecrire dans la zone de recherche « **Integration Services** » et choisir «**Integration Services Project** » comme l’indique la figure suivante puis cliquez sur « **Suivant** ».



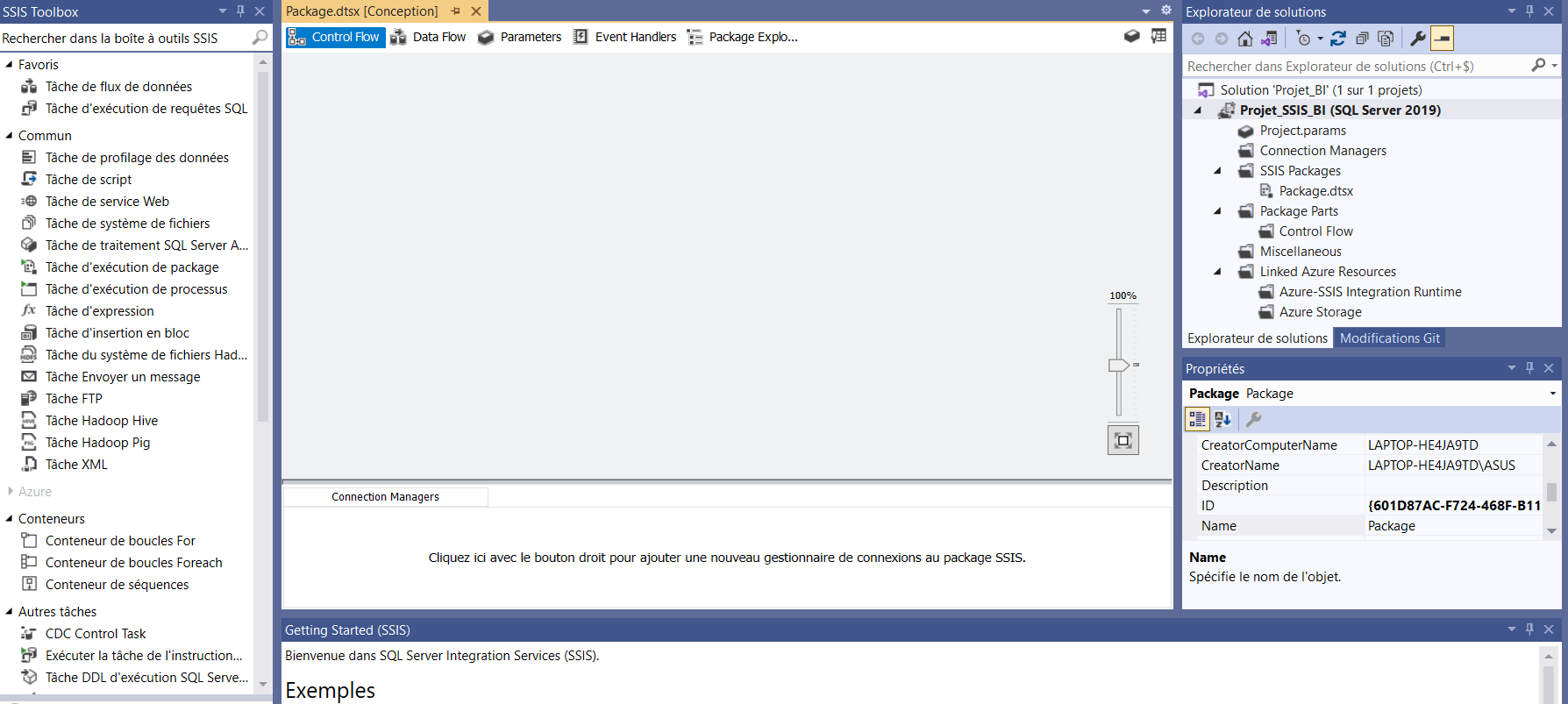
Un projet SSIS, ou un projet Visual Studio est en général, stocké dans une solution. Une solution peut regrouper plusieurs projets de types différents.

- Nommer le projet « **Projet\_SSIS\_BI** » et la solution « **Projet\_BI** ».

## - Configurer l’emplacement de votre projet SSIS sur votre PC.



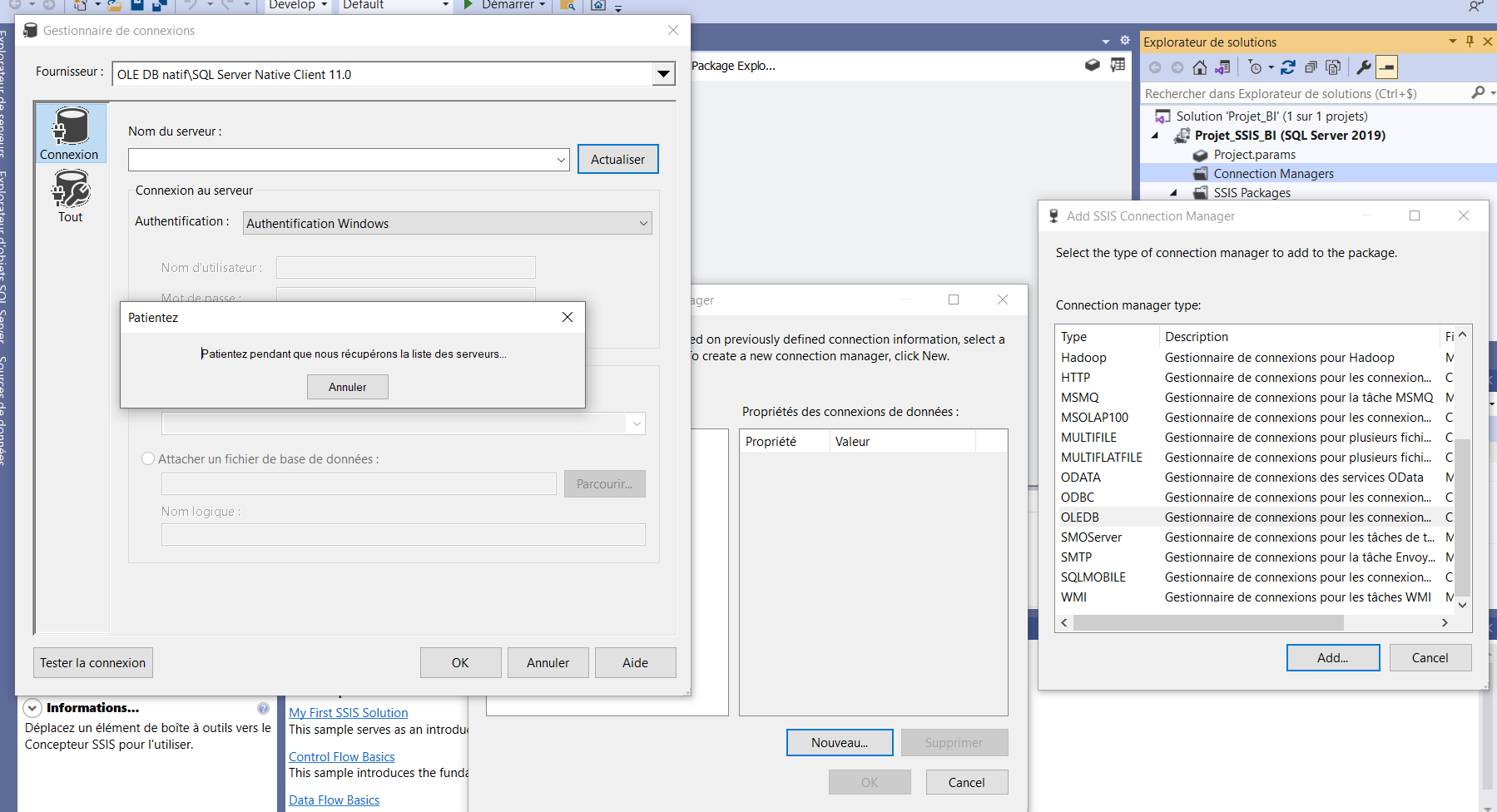
Nous venons de créer alors un nouveau projet d’intégration de service (SSIS).



* 1. **Configuration du projet SSIS**

**Etape 1 : Création des connexions aux bases «Data\_Source» et «Data\_Target»**

Sous le projet « **SSIS** », cliquer avec le bouton droit sur **« Connection Managers »** et accéder à la base de données « **Data\_Source** » déjà restaurée dans **« SQL Server Management Studio »** (voir figures ci-dessous).

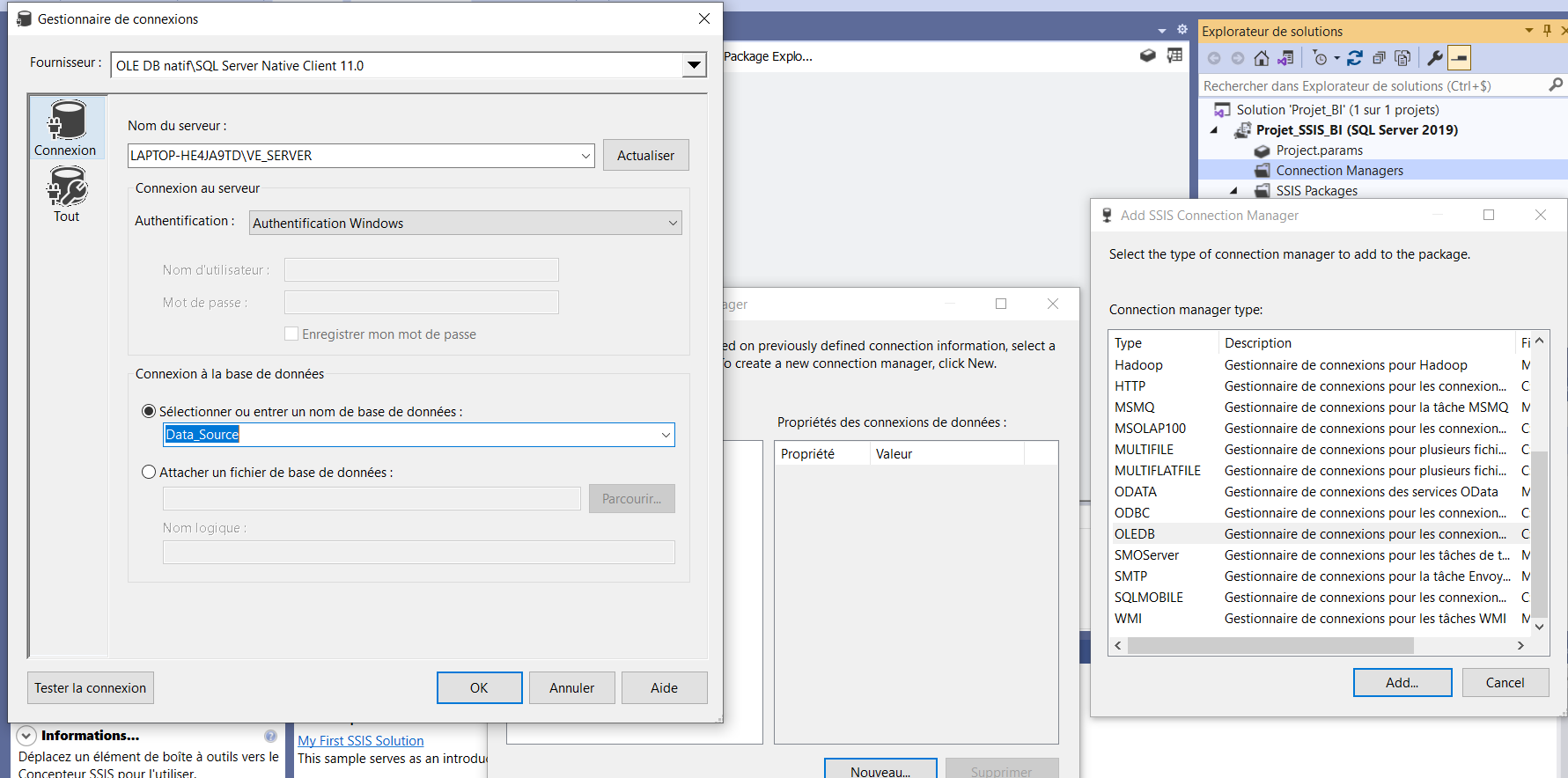


4

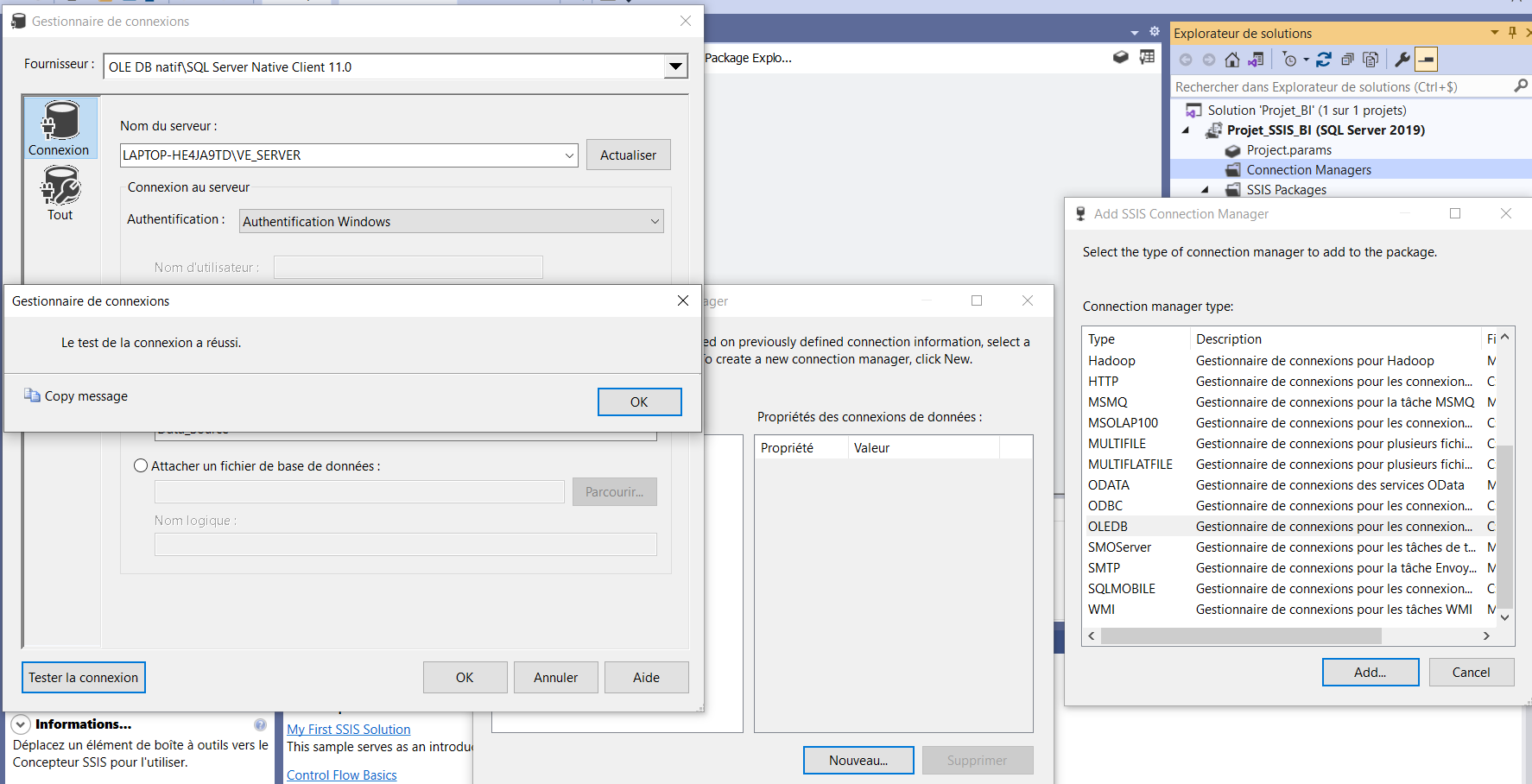
31

21

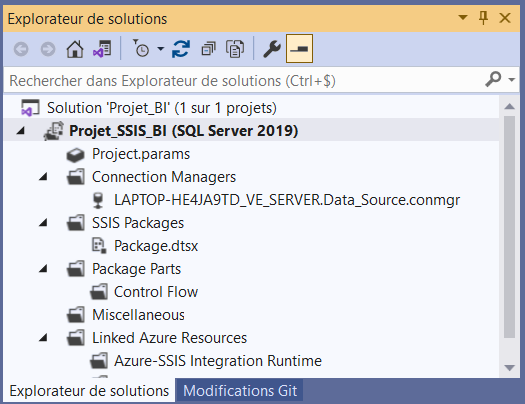
11

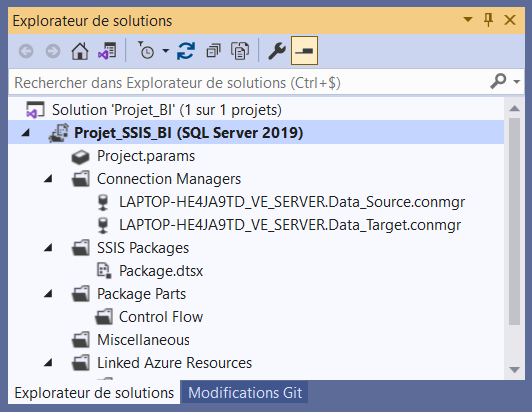


5

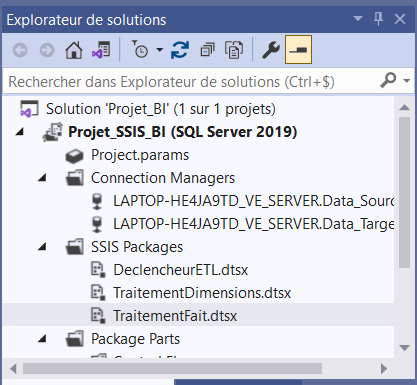


6



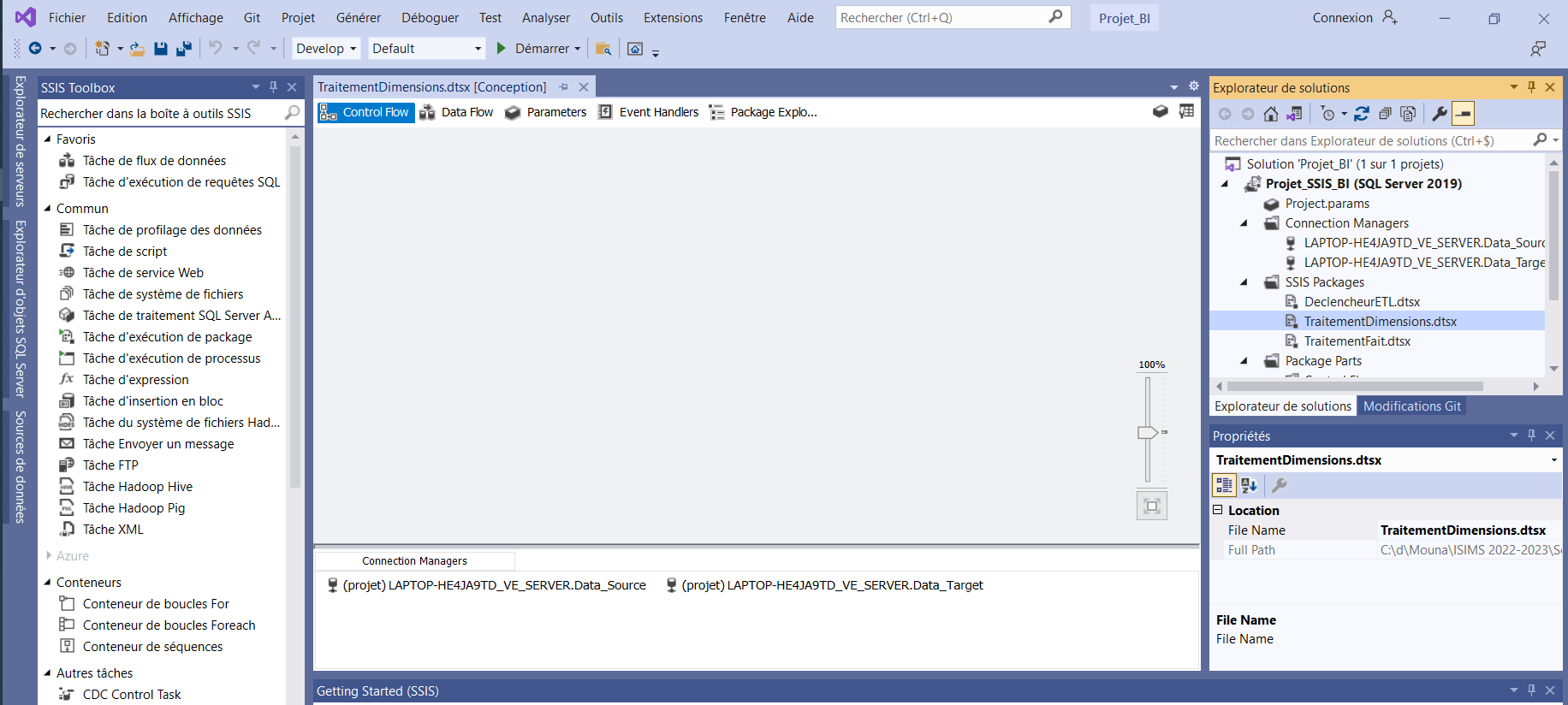


Nous allons refaire le même travail afin d’accéder à la base de données cible «**Data\_Target** » déjà restaurée dans « **SQL Server Management Studio** ».

**Etape 2 : Création des packages d'intégration de données.**

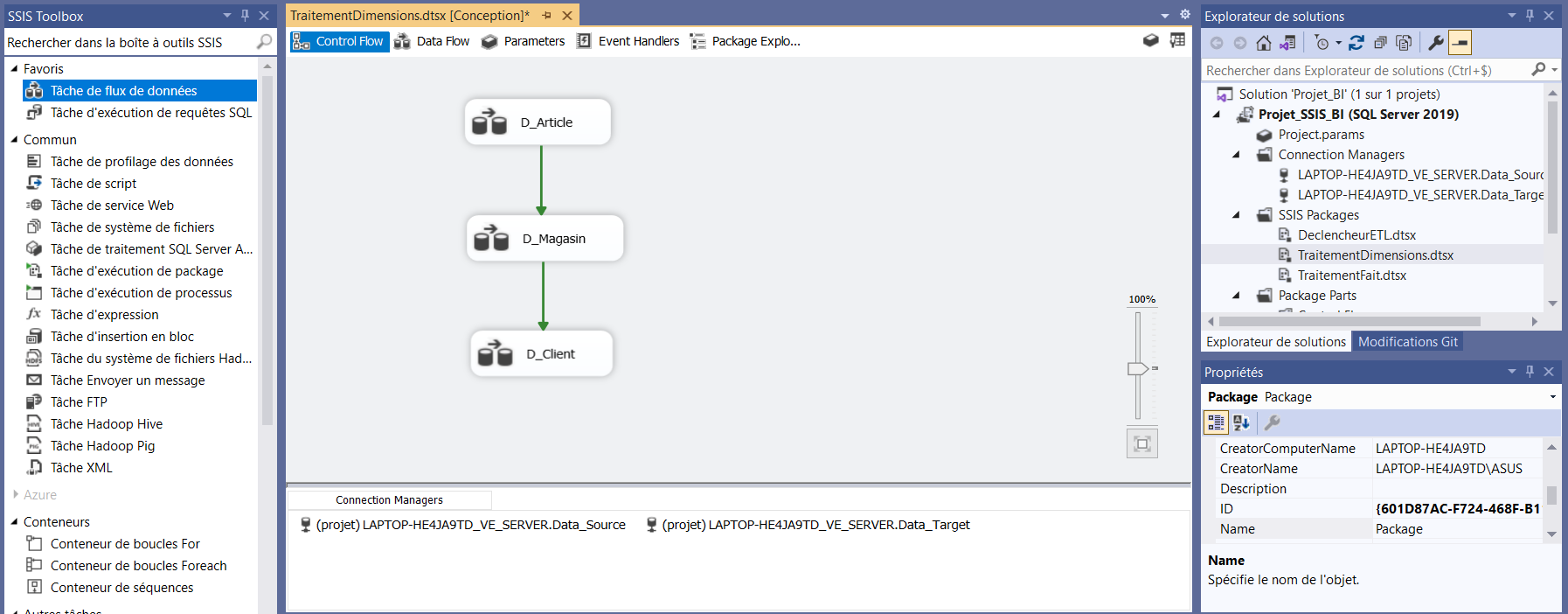
Sous «  **SSIS Packages** » nous allons créer 3 packages :

* Premier package, destiné pour la création de nos *Dimensions* : « **TaitementDimensions.dtsx** ».
* Deuxième package, destiné pour la création de notre table de *Fait* : «**TraitementFait.dtsx** » .
* Troisième package, destiné pour l’exécution successive des deux premiers packages : « **DeclencheurETL.dtsx** ».
* **Traitement des dimensions :**
* Ouvrir le premier package avec un double clique sur **« TaitementDimensions.dtsx** **».**



* Dans l’espace « **Control Flow** », glisser trois fois, à partir de l’onglet « **SSIS Toolbox** » le composant **« Tâche de Flux de données »** et renommer les tâches comme suit : **« D\_Article »**, **« D\_Magasin »** et **« D\_Client ».**

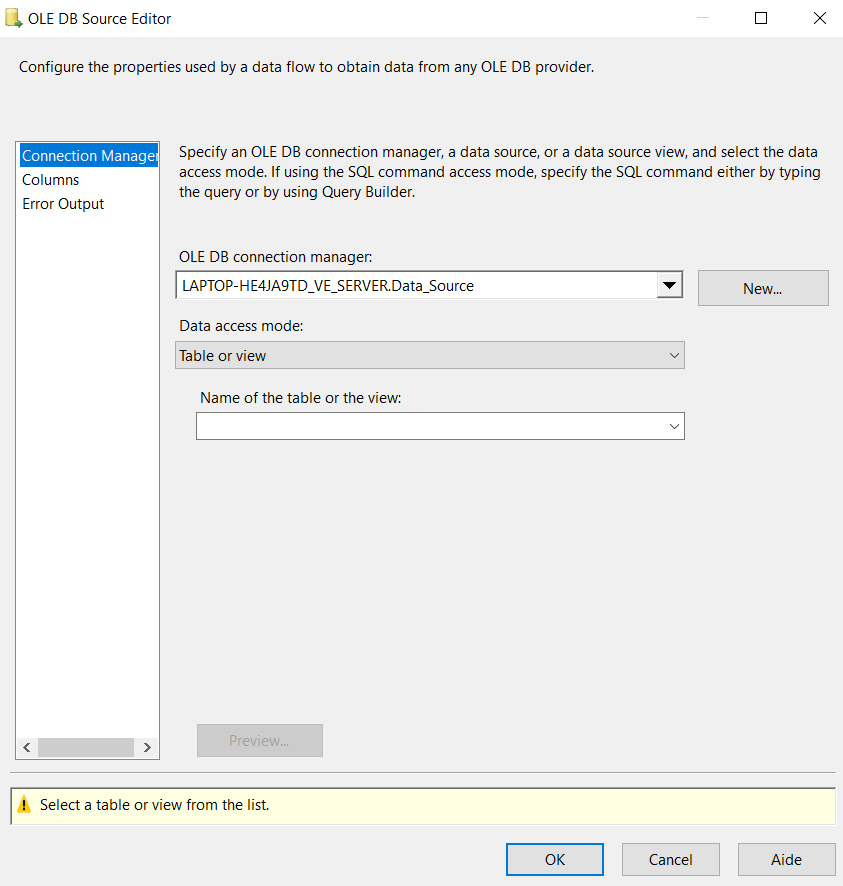
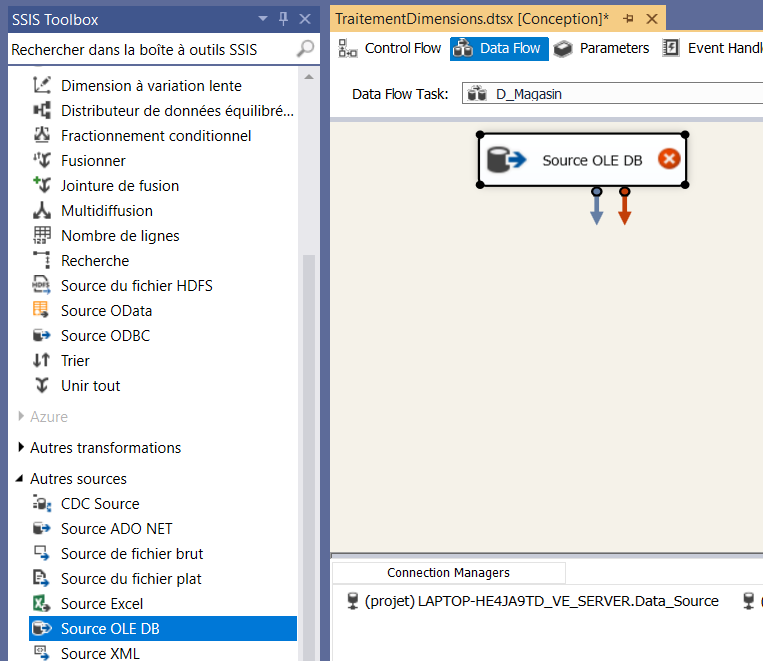
**💣 Ne pas oublier de créer des relations entre les 3 tâches.**

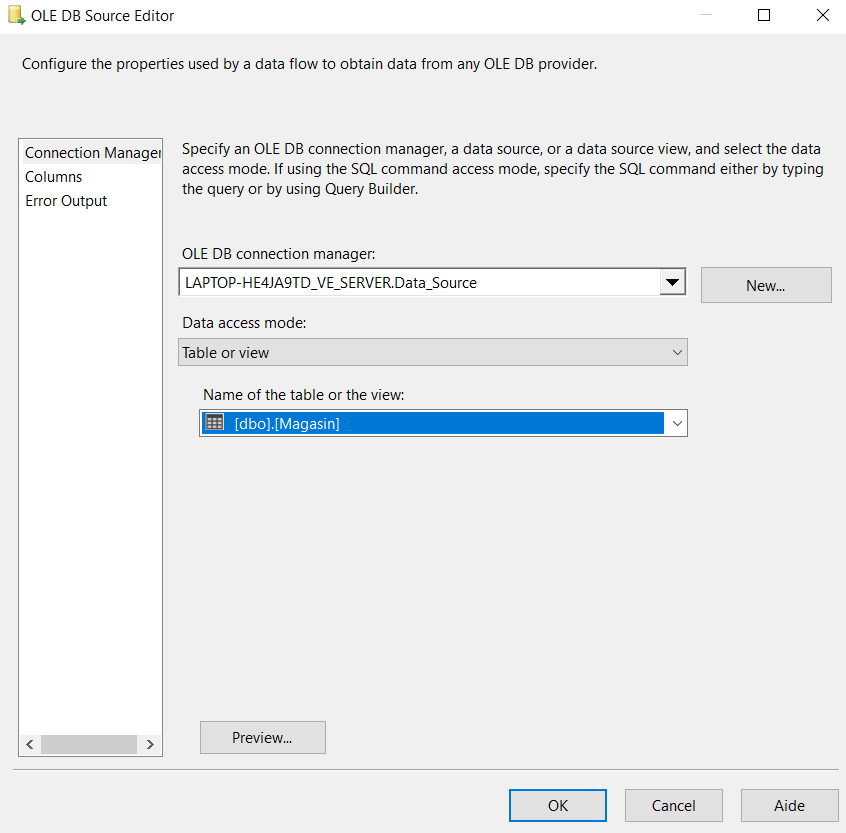
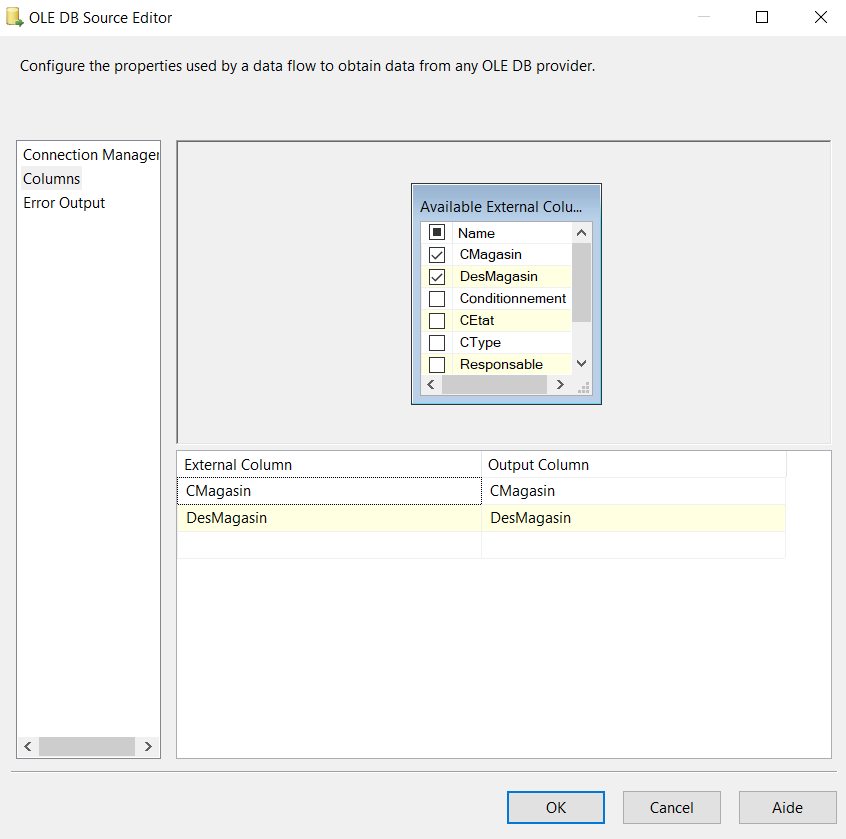


* Nous passons maintenant à la configuration de chaque tâche en spécifiant une **source** de donnée afin de récupérer les données qui vont être acheminées vers **la dimension** désirée**.**

***Tâche 1 : « D\_Magasin »***

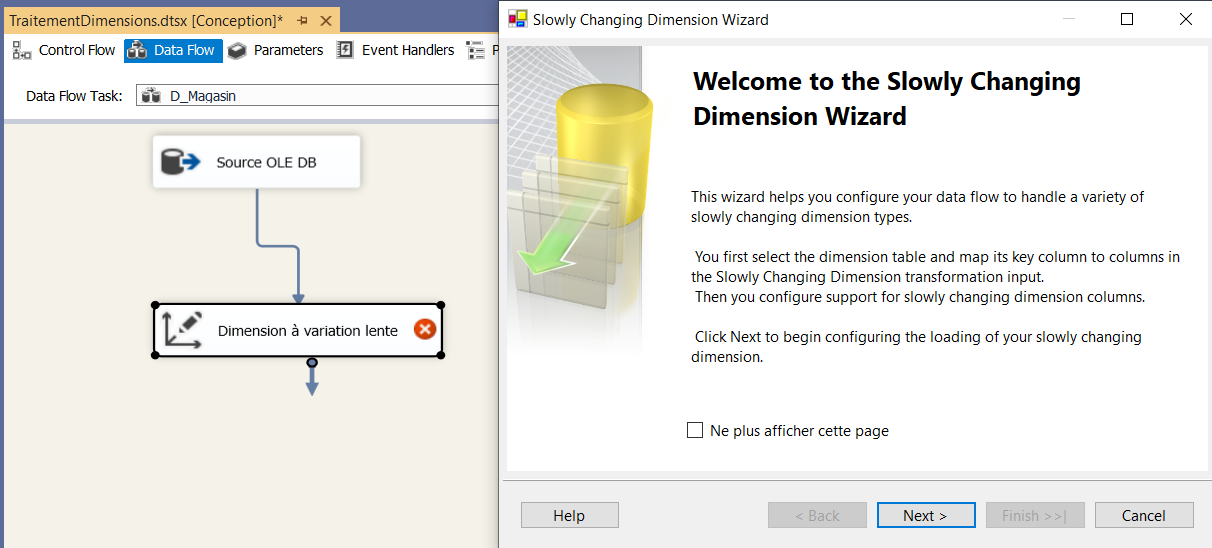
* Double cliquer sur la tâche «**D\_Magasin** », l’espace **« Data Flow »** s’ouvre.
* Glisser dans cet espace le composant «**Source OLE DB**».
* Double cliquer sur «**Source OLE DB**».
* Dans « **OLE DB connection manager** », choisir la connexion à votre base source.
* Dans « **Data access mode** », choisir « **Table ou vue** ».
* Choisir le nom de la table ou de vue « **[dbo].[Magasin]** ».
* Dans l’option « **Columns** » choisir seulement deux champs de la table Magasin : «**CMagasin**» et «**DesMagasin**».
* Valider par OK (voir figures ci-dessous)



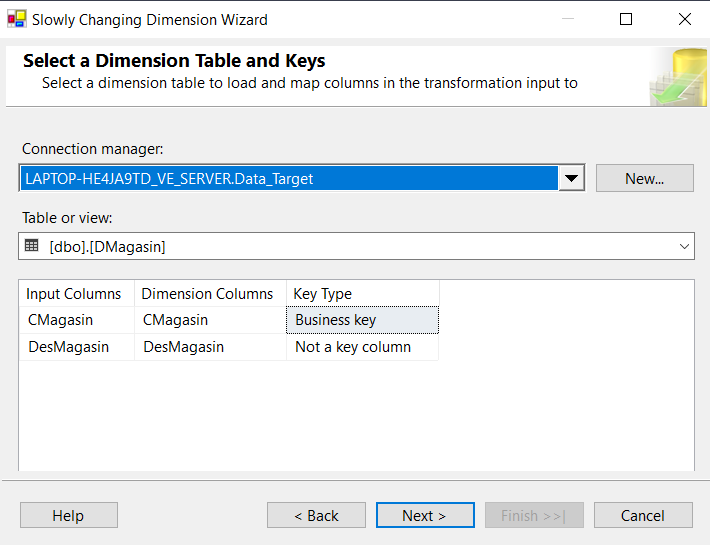
 

1

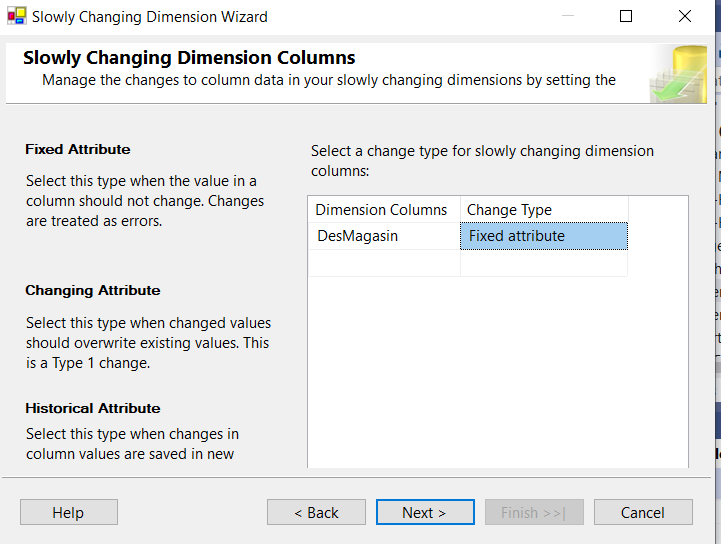
* Glisser ensuite dans l’espace **« Data Flow »** le composant « **Dimension à variation lente** ».
* Acheminer les données à partir de la source de données vers le nouveau composant. Pour cela glisser la flèche bleu à partir du composant « **Source OLE DB** » vers le composant « **Dimension à variation lente** ».
* Double cliquer sur le composant « **Dimension à variation lente** », on arrive à ce niveau :



* Dans « **Connection manager** », choisir la connexion à votre base cible.
* Dans « **Table or view** », choisir « **[dbo].[DMagasin]** ».
* Dans « **Key Type** » affecter « **Business Key** » à la colonne « **CMagasin** ».

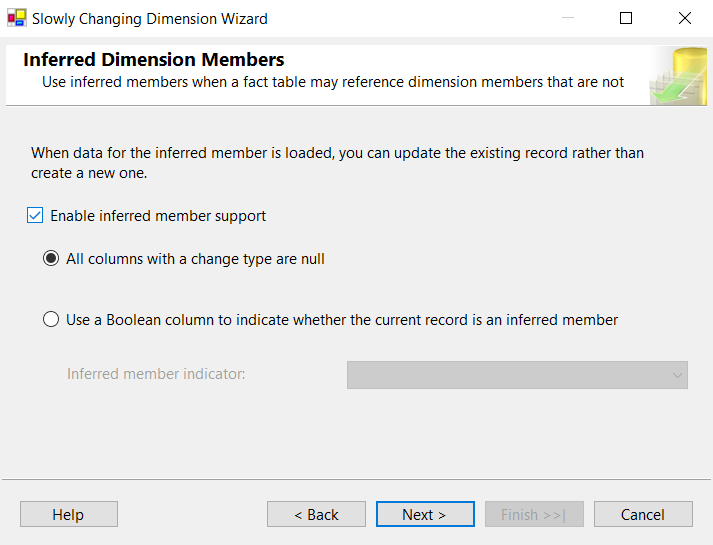
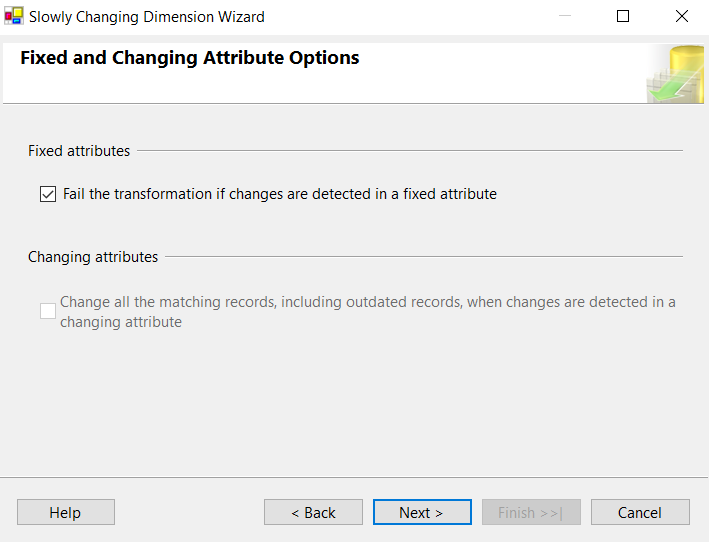


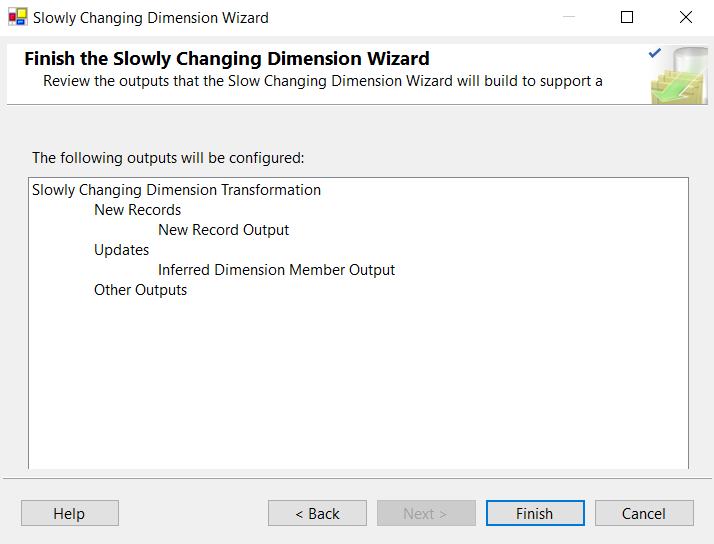
* Arrivant à la fenêtre suivante, on doit choisir le type de chaque colonne à introduire dans notre dimension.

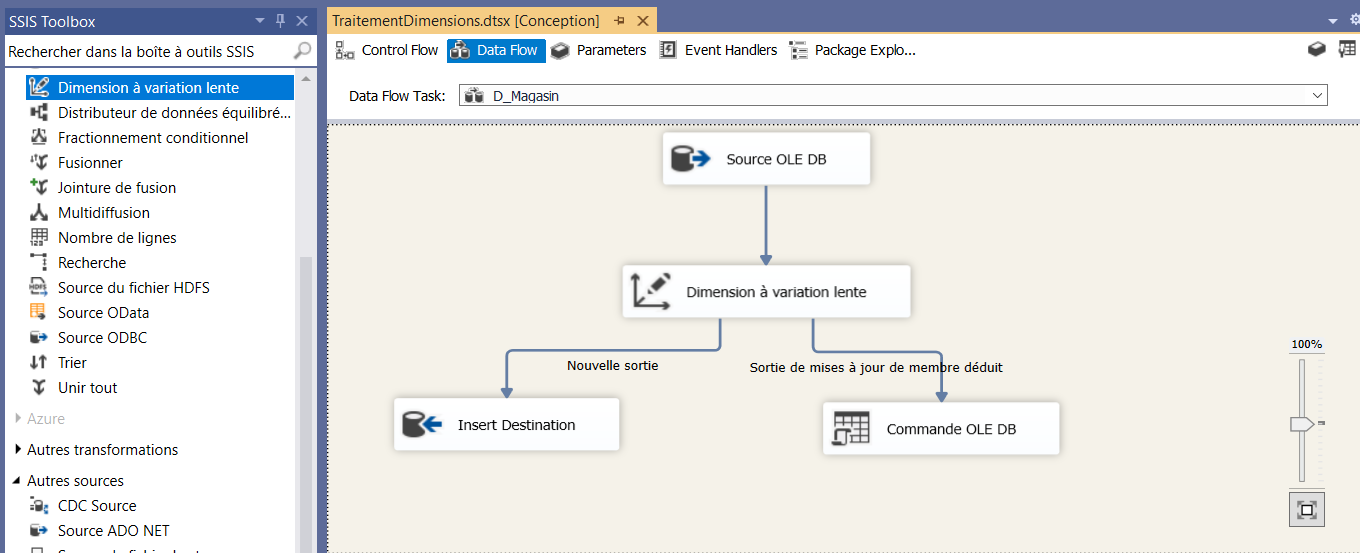


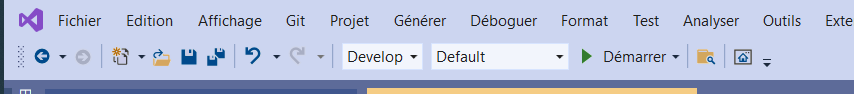
Pour une dimension à variation lente (SCD) on peut affecter à **chaque attribut (colonne)** un **type de changement**. La figure ci-dessus nous montre trois types d’attributs qui sont possibles :

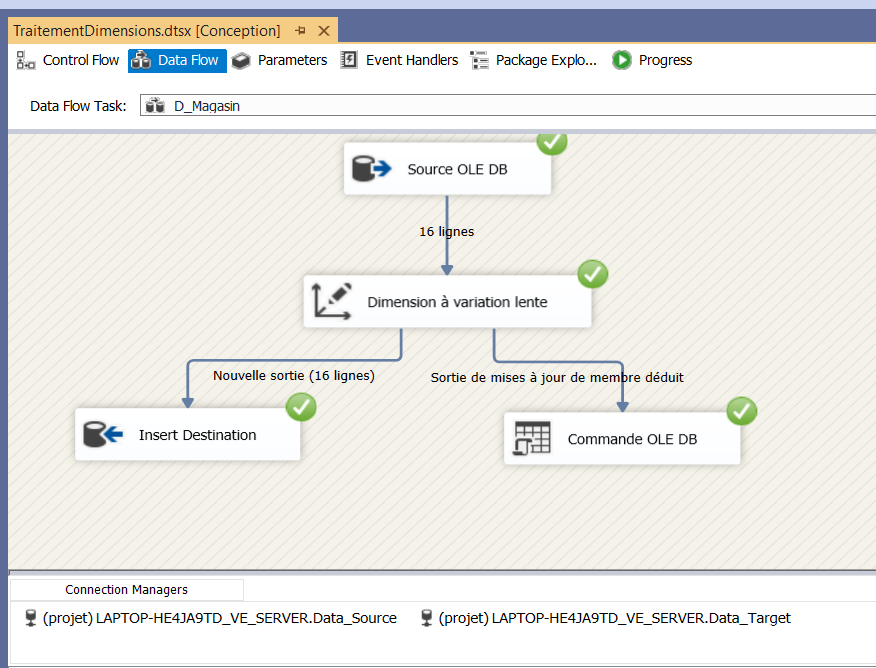
* **Fixed Attribute** : ce type est sélectionné lorsque la valeur dans une colonne ne doit pas varier.
* **Changing Attribute** : ce type est sélectionné lorsque les valeurs modifiées doivent remplacer les valeurs existantes.
* **Historical Attribute** : ce type est sélectionné lorsque les valeurs de colonnes ne doivent pas être supprimées. Donc, pour une même colonne, on trouve les anciennes et les nouvelles valeurs.
* Pour la dimension « **D\_Magasin** » nous allons affecter « **Fixed Attribute**» pour la colonne « **DesMagasin** ».



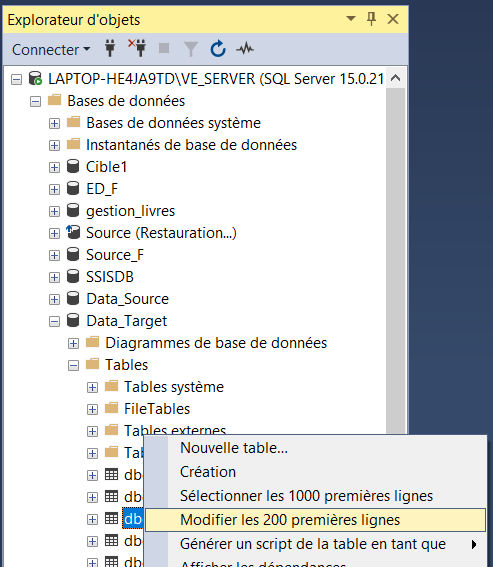
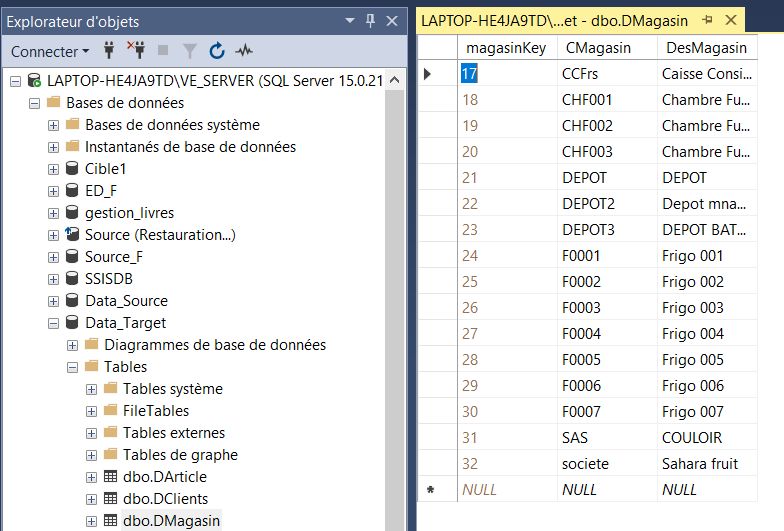




* Après avoir terminé « **Exécuter la tâche** » en cliquant sur le bouton démarrer .



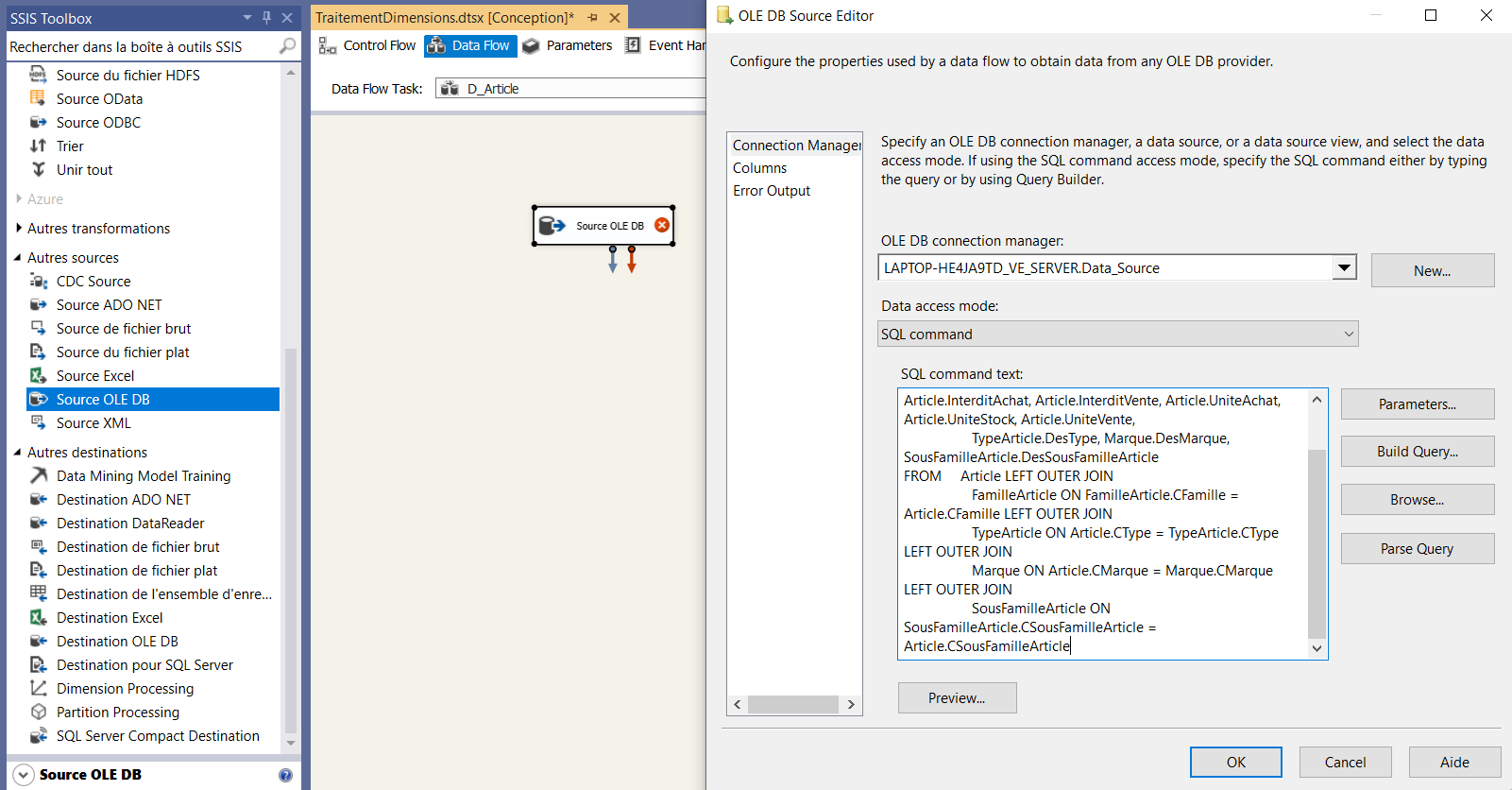
* Consulter **SSMS** pour voir le résultat de chargement de la table « dbo.DMagasin » dans la base « Data\_Target ».

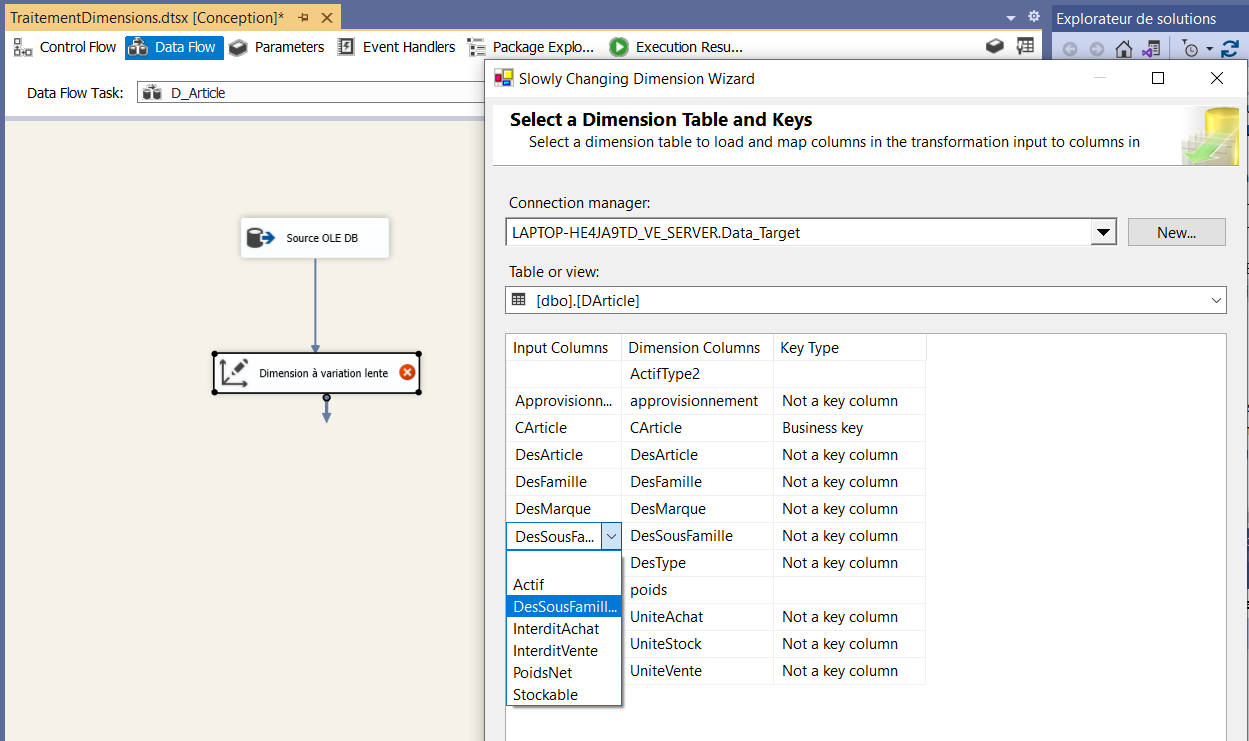
***Tâche2 :*  *« D\_Article »***

Pour la tâche «**D\_Aticles**», les données seront extraites à partir de la base source à travers une requête SQL.

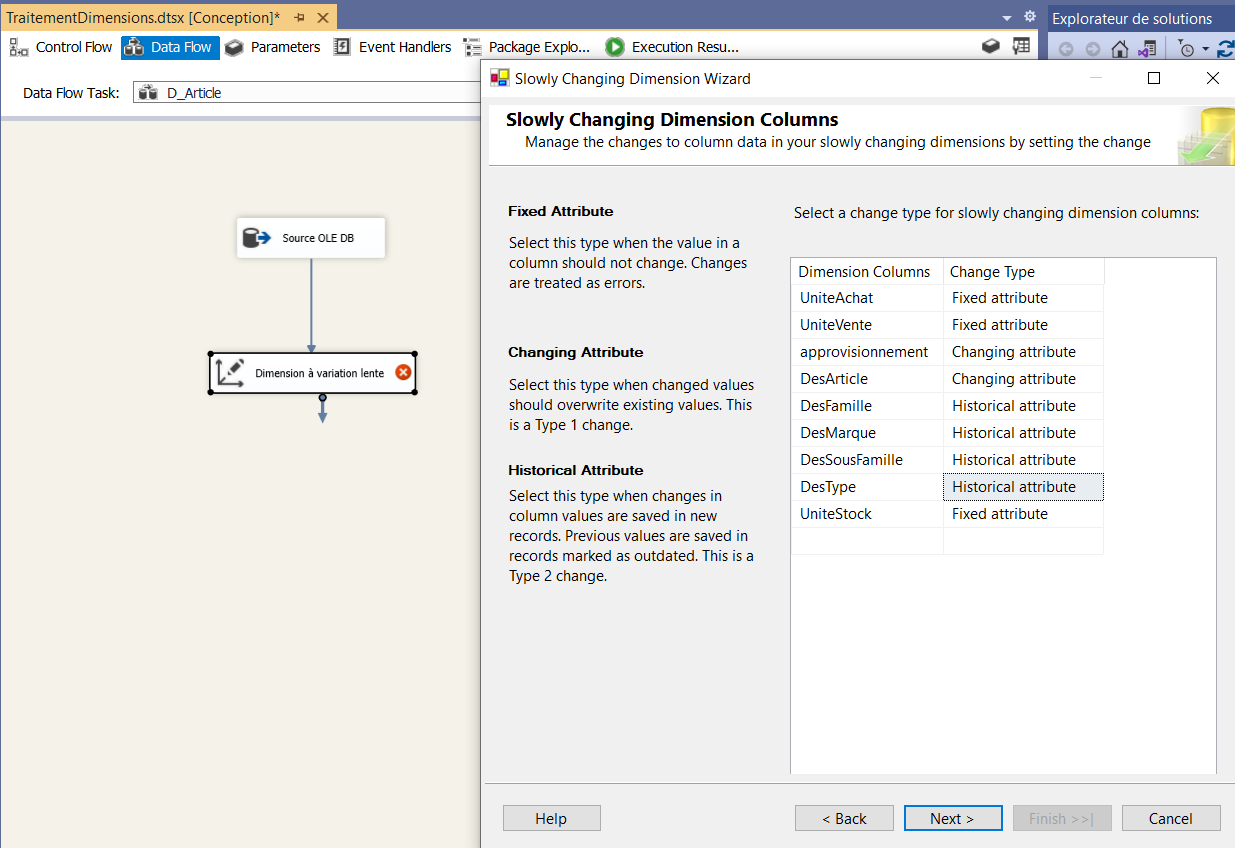
* Double cliquer sur la tâche «**D\_Aticles**», l’espace **« Data Flow »** s’ouvre.
* Glisser dans cet espace le composant « **Source OLE DB** ».
* Double cliquer sur « **Source OLE DB** ».
* Dans « **OLE DB connection manager** », choisir la connexion à votre base source.
* Dans « **Data access mode** », choisir « **SQL Command** ».
* Copier / coller la requête à partir du fichier « **RequeteArticle.txt** ».
* Finir par valider par **OK.**



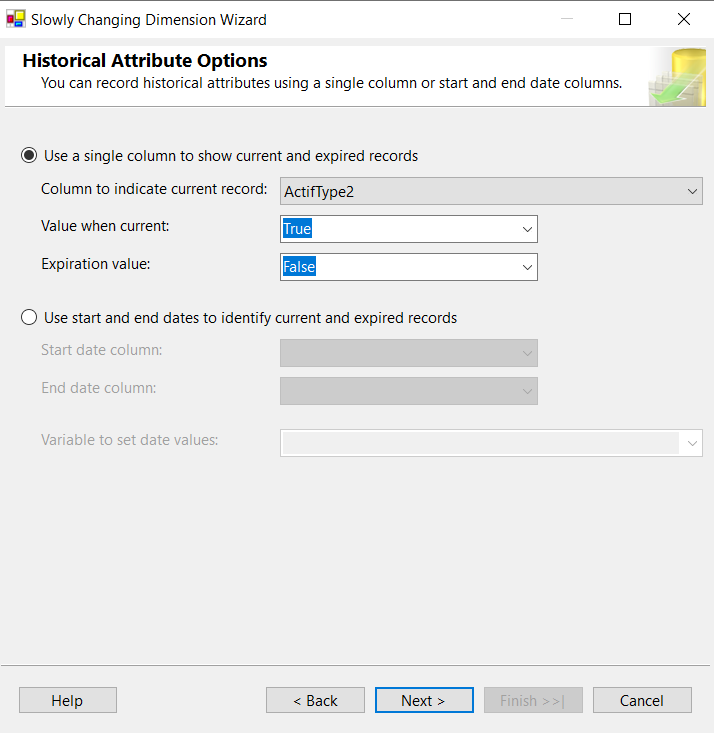
* Glisser dans l’espace Data Flow le composant « **Dimension à variation lente** » et relier le avec le composant « **Source OLE DB** ».
* Double cliquer sur la « **Dimension à variation lente** ».
* Choisir dans « **Connection manager** » la base cible « **Data\_Target** » dans laquelle nous allons enregistrer les données.
* Choisir « **Table or view** ».
* Choisir la table « **[dbo].[DArticle]** »
* Affecter « Business Key » à la colonne **« CArticle ».**
* Remplir les « Input Comumns » vides (DesSousFamille).
* Cliquer sur « Next ».



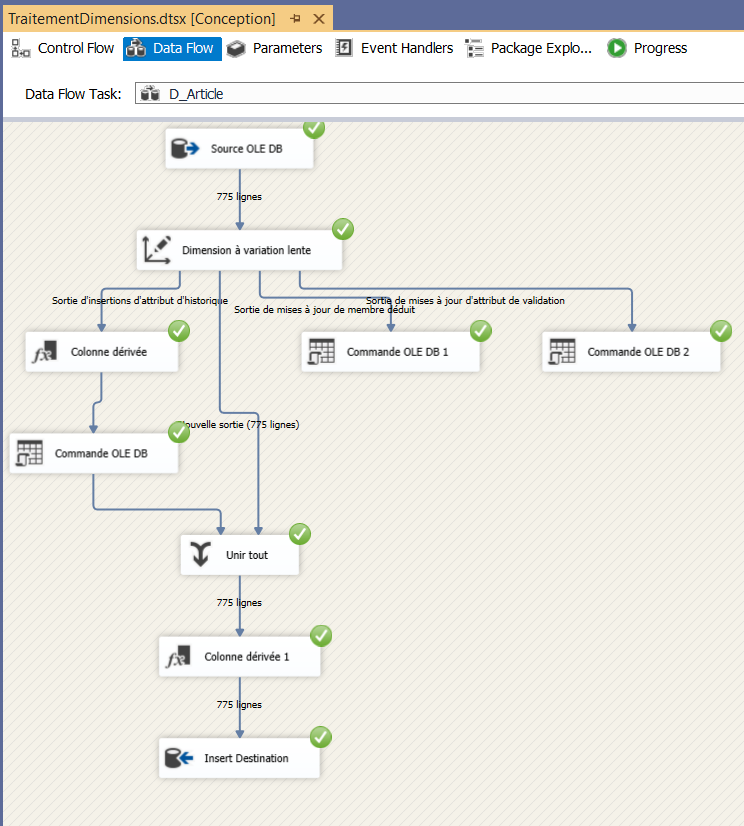
* Pour la dimension « **D\_Article**» nous allons attribuer les types aux diverses colonnes comme suit :
* La colonne « UniteAchat » → « Fixed attribute ».
* La colonne « UniteVente » → « Fixed attribute  ».
* La colonne « approvisionnement » → « Changing attribute ».
* La colonne « DesArticles » → « Changing attribute ».
* La colonne « DesFamilles » → « Historical attribute ».
* La colonne « DesMarques » → « Historical attribute ».
* La colonne « DesSousFamilles » → « Historical attribute ».
* La colonne « DesTypes » → « Historical attribute ».
* La colonne « UniteStock » → « Fixed attribute ».



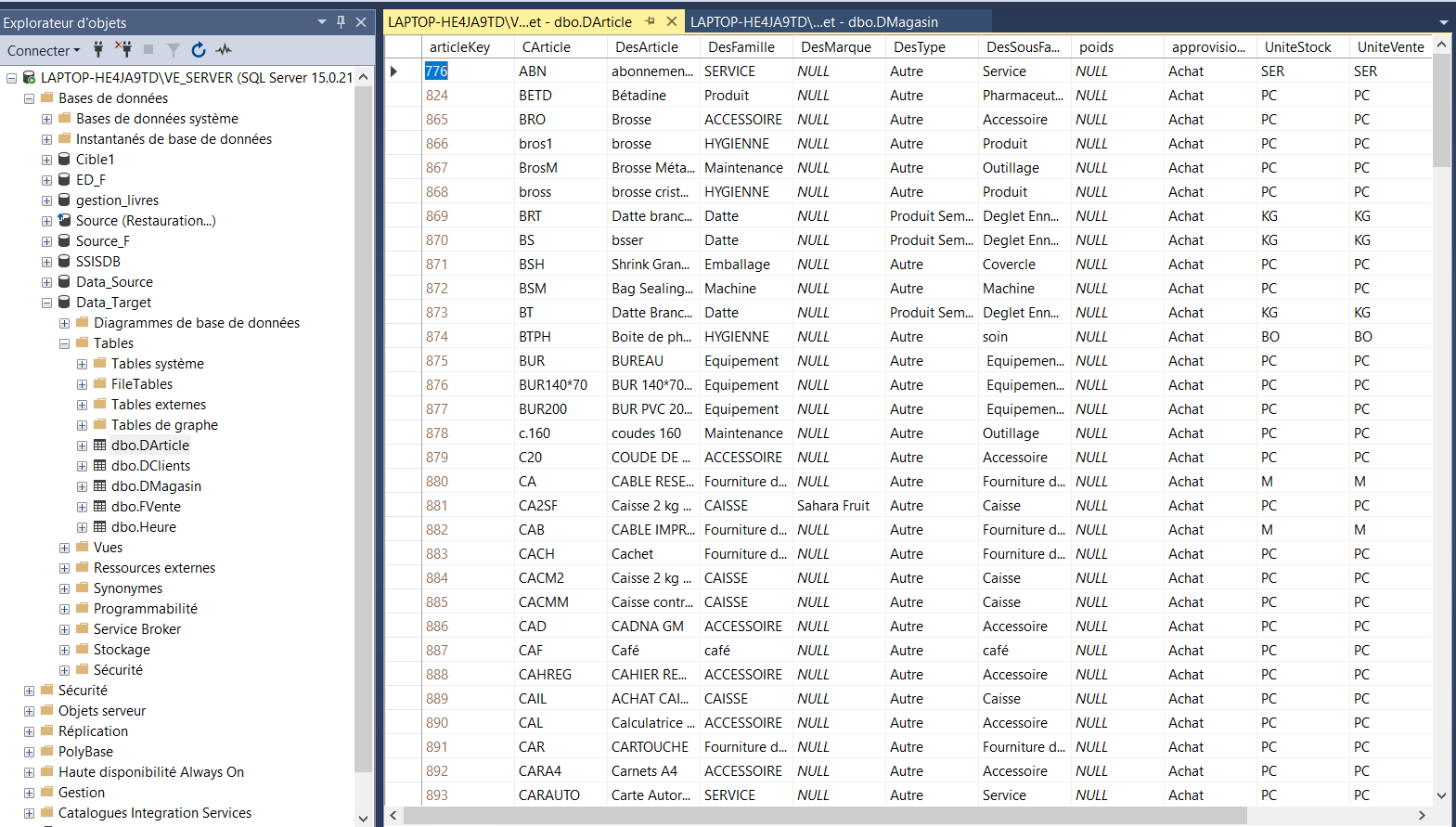
* Cliquer « suivant » deux fois et vous aurez cette fenêtre. Faites les changements nécessaires.



* Dans cette fenêtre, nous configurons dans la dimension la colonne « **ActifType2** » qui sera associée pour chaque colonnes de type « **Historical attribute** ***».***
* Après avoir terminé « Exécuter la tâche ».

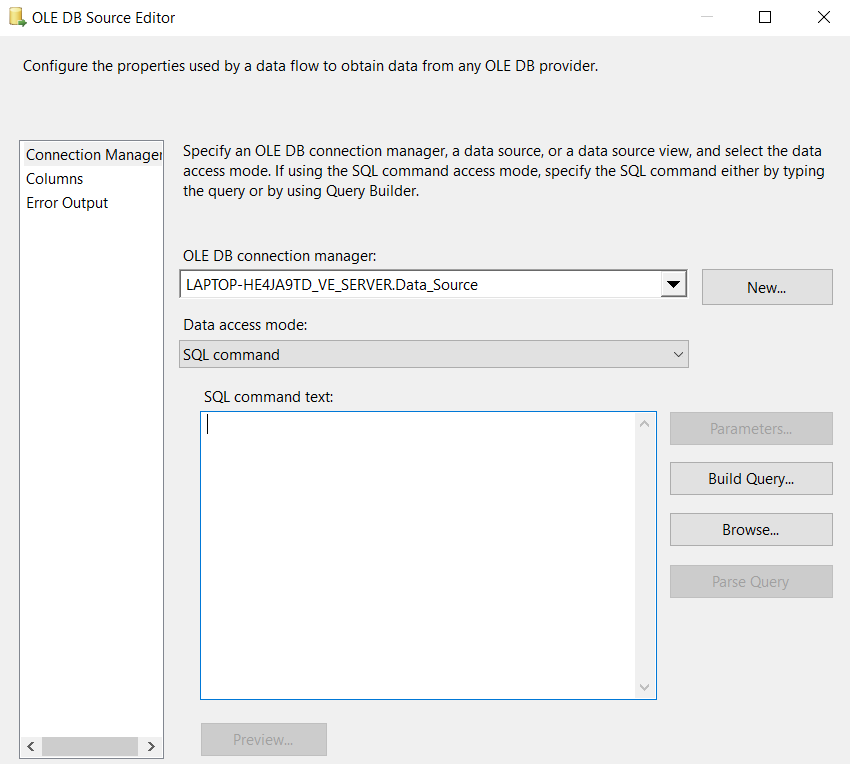


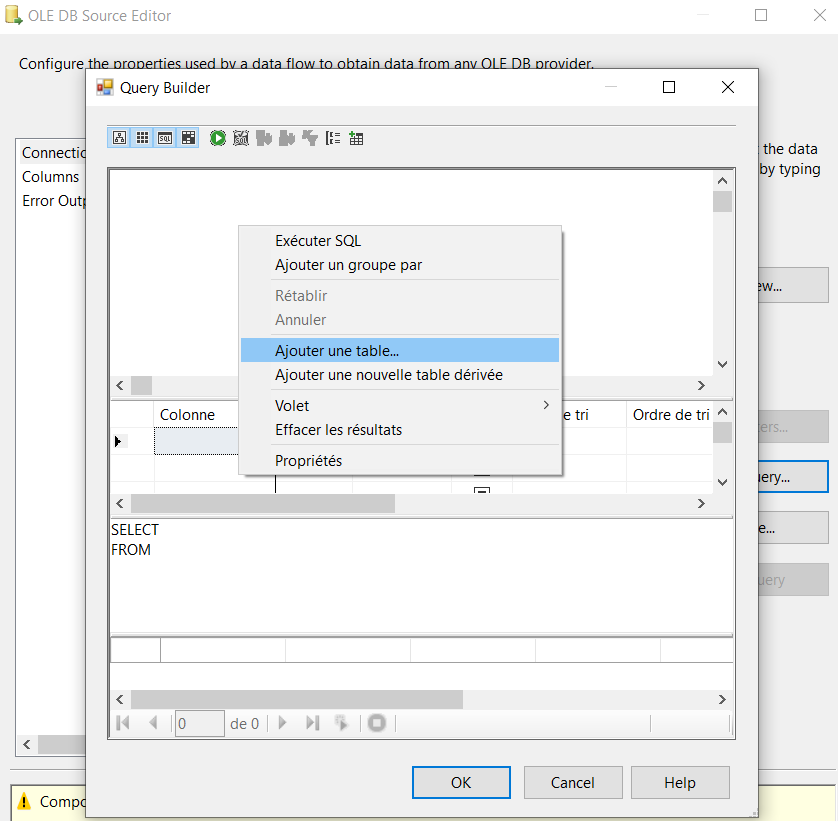
* Consulter **SSMS** pour voir le résultat de chargement de la table « dbo.DArticle » dans la base «Data\_Target ».



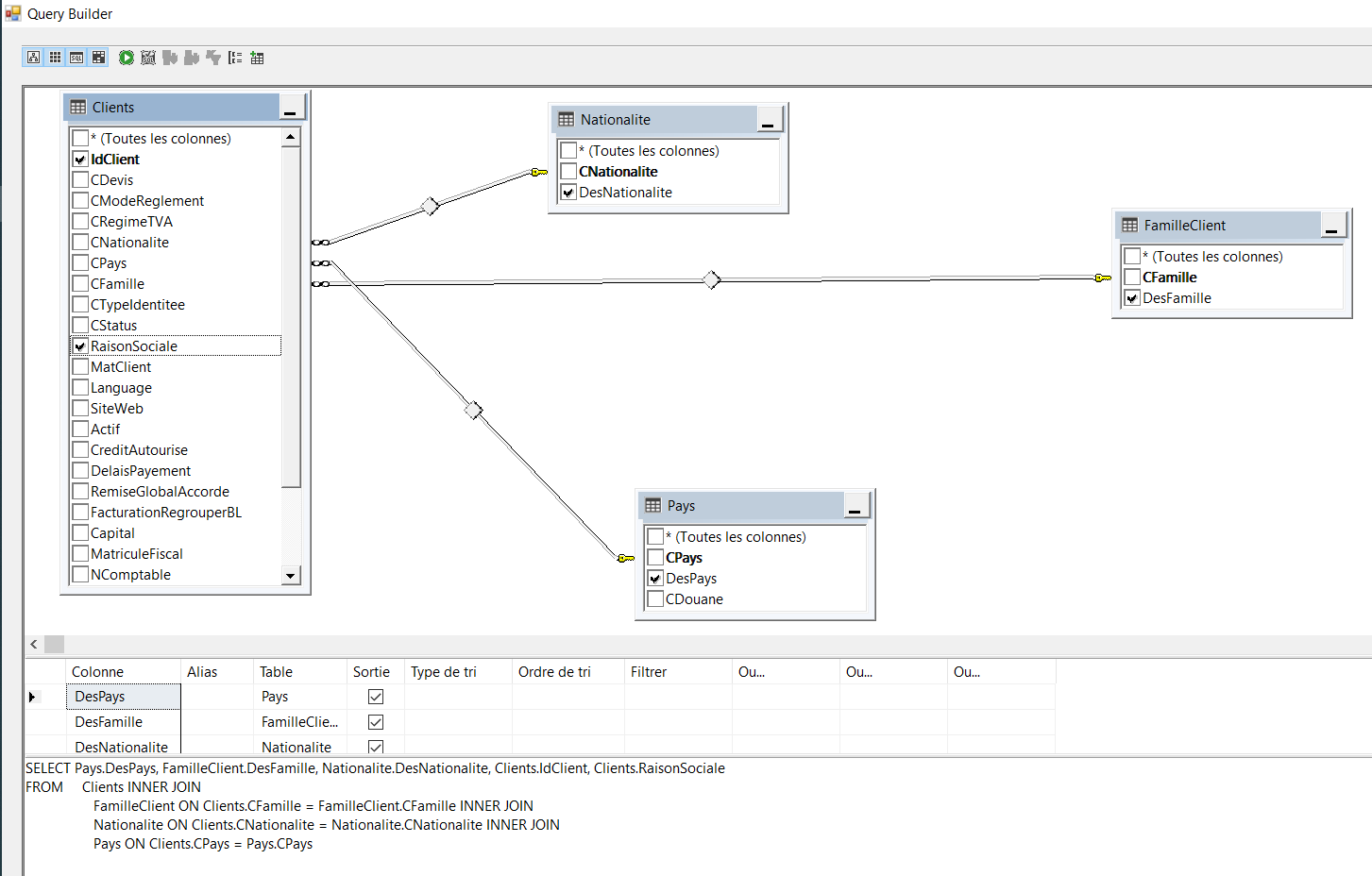
***Tâche 3 : « D\_Client »***

Pour la tâche « D\_Client » les données seront extraites à partir de plusieurs tables.

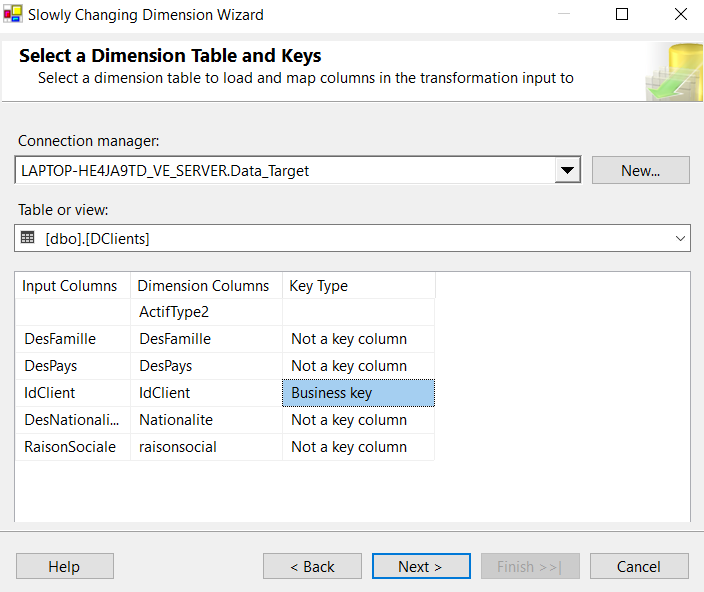
* Double cliquer sur la tâche «**D\_Client**», l’espace **« Data Flow »** s’ouvre.
* Glisser dans cet espace le composant « **Source OLE DB** ».
* Double cliquer sur « **Source OLE DB** ».
* Dans « **OLE DB connection manager** », choisir la connexion à votre base source.
* Dans « **Data access mode** », choisir « **SQL Command** ».
* Cliquer sur « **Build Query** ».
* Cliquer par le bouton droit sur l’espace de Query Builder, puis choisir « Ajouter une table ».



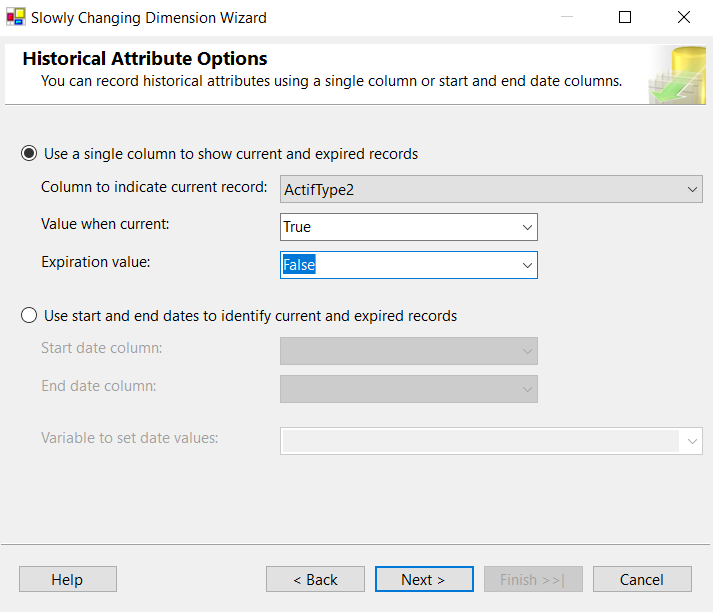
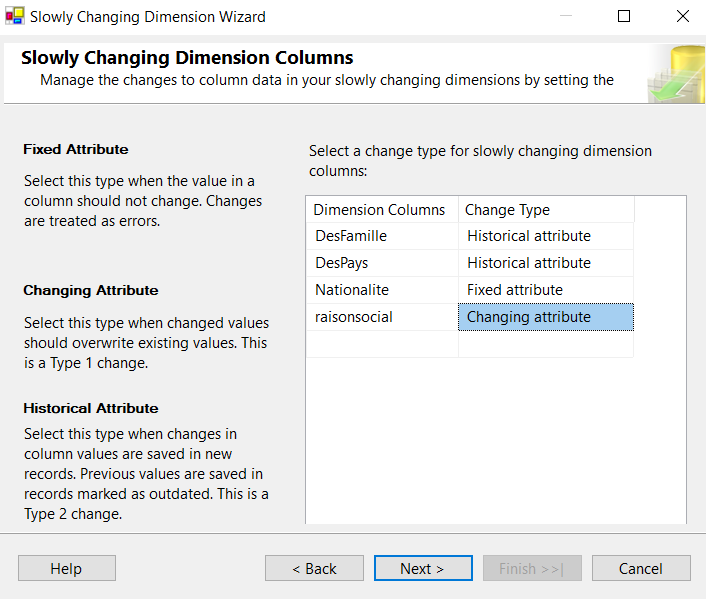
* Ajouter ces quatres tables : « **Clients** », « **FamilleClient** », « **Nationalite** » et «**Pays**».
* Choisir à partir de chaque table les colonnes nécessaires : « **DesPays** », « **DesFamille** », «**DesNationalite**», «**IdClient**» et « **RaisonSociale** »
* La requête SQL sera générée automatiquement.



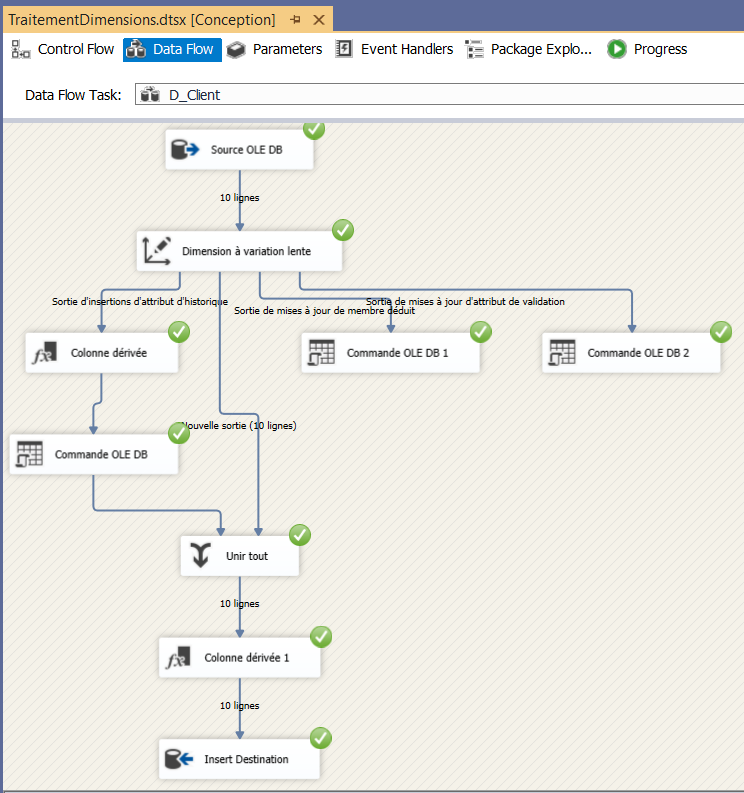
* Glisser dans l’espace Data Flow le composant « **Dimension à variation lente** » et relier le avec le composant « **Source OLE DB** ».
* Double cliquer sur « **Dimension à variation lente** ».
* Choisir dans « **Connection manager** » la base cible « **Data\_Target** » dans laquelle nous allons enregistrer les données.
* Choisir « **Table or view**»
* Choisir la table « **[dbo].[DClients]** »
* Affecter « Business Key » à la colonne **« IdClient »**
* Remplir les « Input Comumns » vides (DesNationalite et RaisonSociale).
* Cliquer sur « Next ».



* Pour la dimension « **D\_Client**» nous allons attribuer les types aux diverses colonnes comme suit :
* La colonne « DesFamille » → « Historical attribute  ».
* La colonne « DesPays » → « Historical attribute  ».
* La colonne « Nationalite» → « Fixed attribute ».
* La colonne « RaisonSociale» → « Changing attribute ».



* Après avoir terminé « Exécuter la tâche ».



* Consulter **SSMS** pour voir le résultat de chargement de la table « dbo.DClients » dans la base « Data\_Target ».

