Etude de cas

**Objectif :**

Travaille un Etude de cas qui sairt a créer un processus ETL pour extraire, transformer et charger ces données dans un entrepôt de données afin de faciliter l'analyse des performances des ventes et des tendances.

**Sources de données (deux fichiers): *Commandes.csv*** , ***Clients.csv***

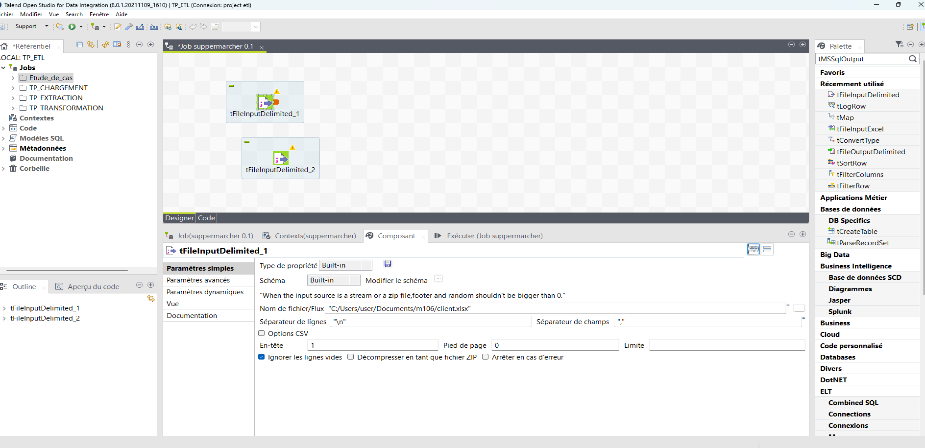
**Partie 1 :**

**Phase 1 : Extraction des données**

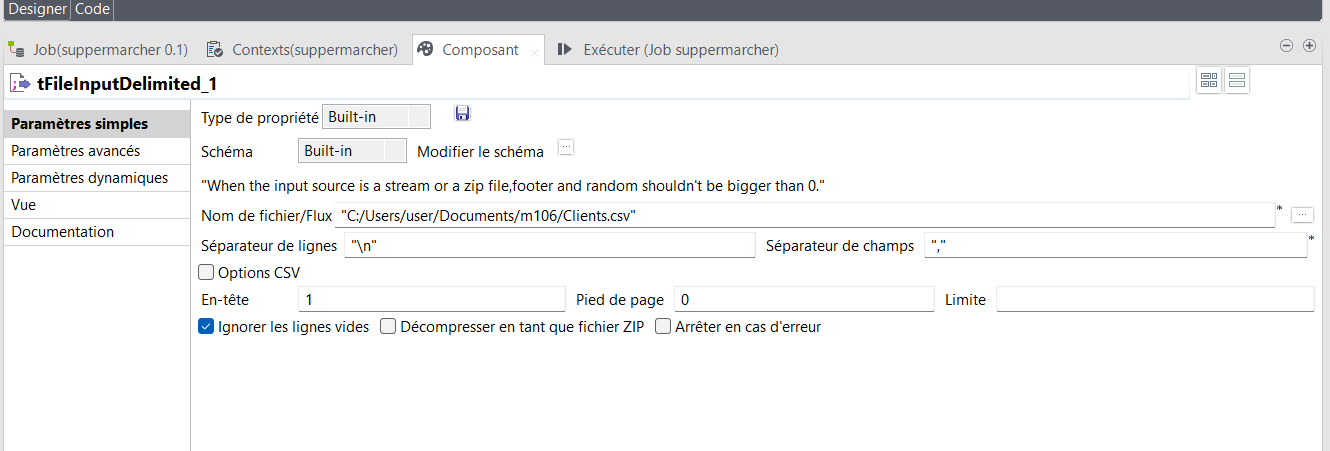
1. Quel est le composant Talend qui permet d’extraire le fichier **Commandes.csv ?**

* **Tfileinputdelimited**

1. Quels sont les champs à remplir pour configurer l’extraction des deux fichiers **Commandes.csv** et **Clients.csv ?** (Cochez les bonnes réponses) 
   * Chemin du fichier
   * Supprimer les données manquantes
   * Caractère de séparation
   * Définir l’entète
   * Le type de donnée dans chaque colonne
2. Ajouter donc ce composant à Talend pour extraire les données des deux fichiers **Commandes.csv** et **Clients.csv ?**

On acced a talend puis en cree un dossier ‘’etud\_de\_cas’’, apres cree une job ‘‘supermarché’’, apres en glisse deux **tfileinputdelimited** de la palette vers l’espace de modilisation pour extraire les donnees 

1. Pour la configuration on definie le Chemin du fichier puis en modifie la schema apres en definie l’entete .



1-modifier le schema

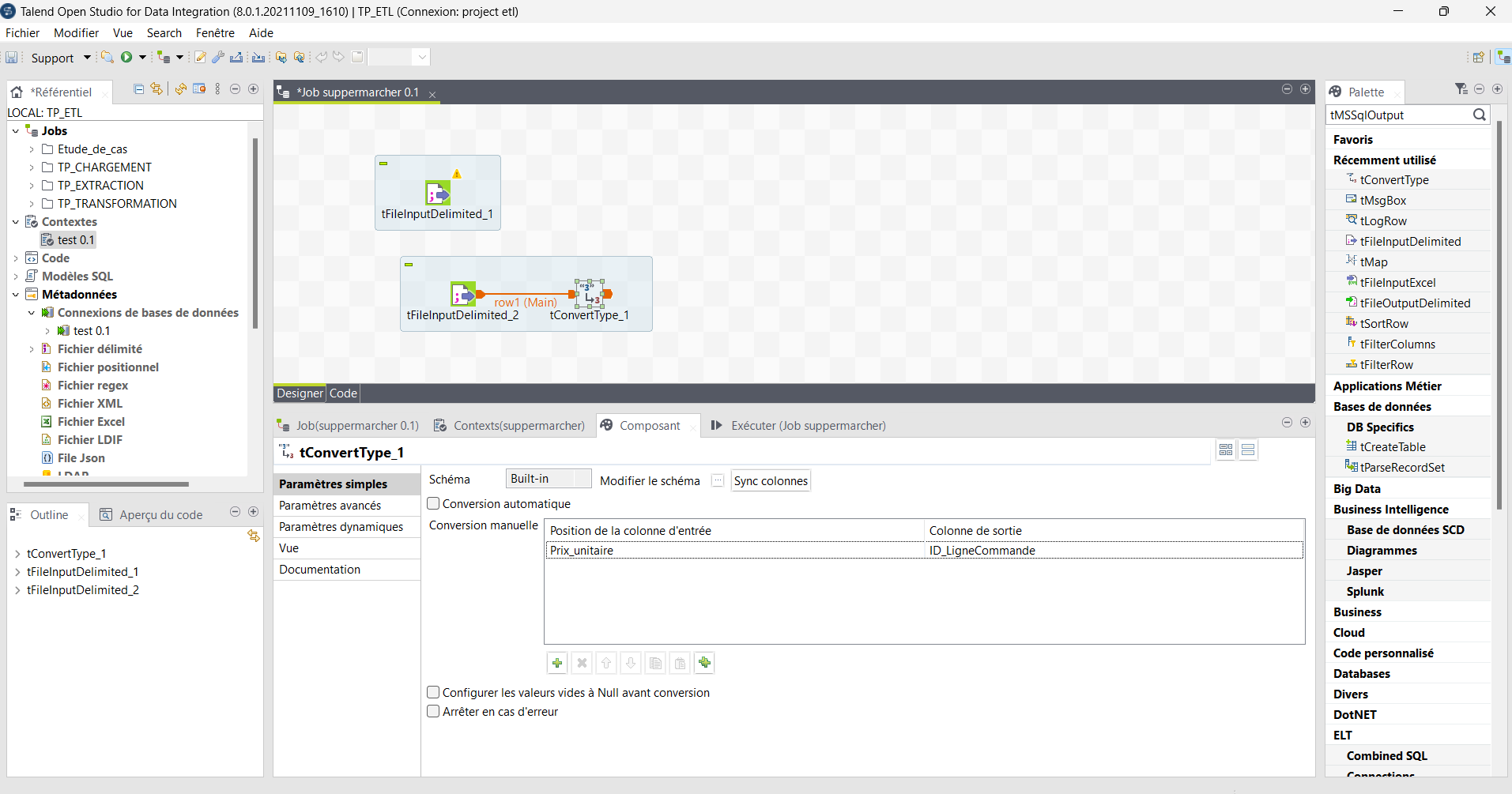
2- nom de fichier

3-En-tete

**Phase 2 : Transformation des données**

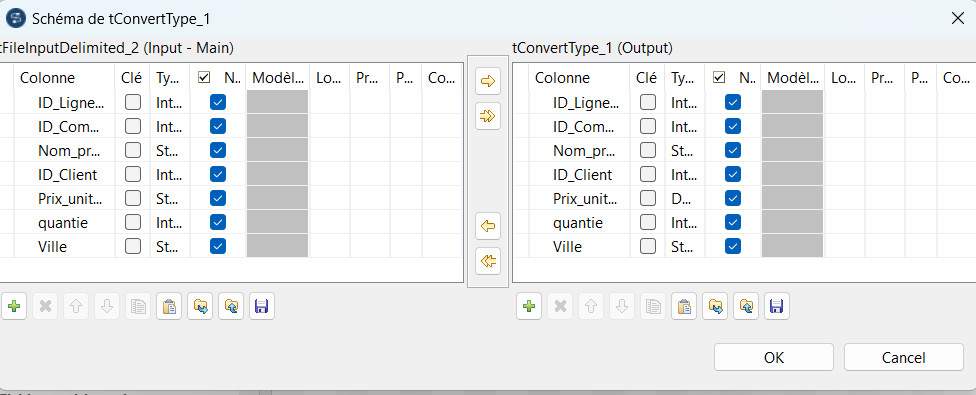
**Tache 1: Conversion de type**

1. Dans la table **Commandes.csv,**  la colonne prix unitaire nécessitant une conversion de type.
2. Dans Talen Open Studio, le composant convenable qui permet de convertir cette colonne est **tconvertType**.
3. Ajouter donc ce composant à Talend et faites la configuratoin neccessaire comme suit :

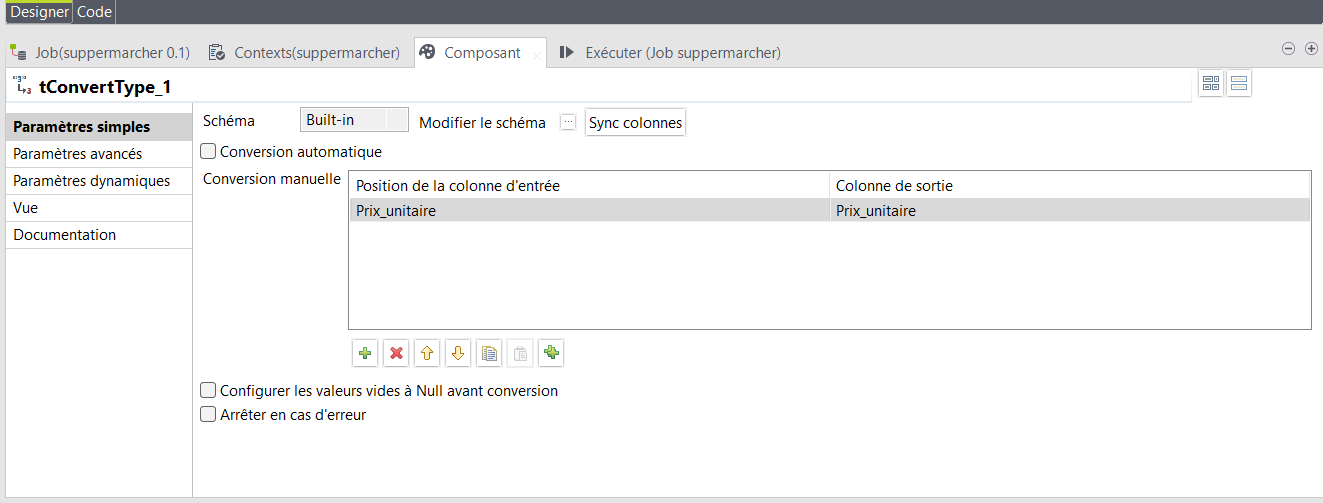


on glisse le composant tconvertType de la palette vers l’espace de modilisation apres on glisse la ligne principal de tFileinputDelimited vers tconvertType ,apres on va configurer le composent .

1-modifier le schema



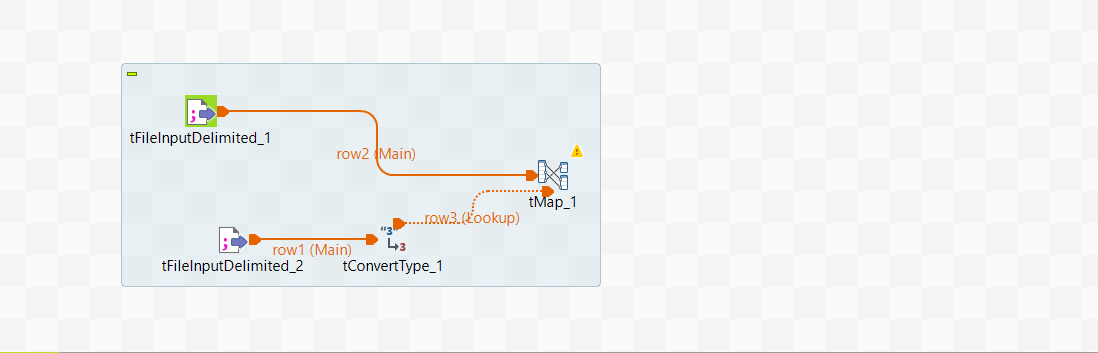
On change le type de Prix\_unitaire vers double 1. click sur sync colonnes pour recuperer data



2. on ajoute une ligne 3. on selectionne **Prix\_unitaire.**

* **Tache 3: Jointure des données**

1. A l’aide de la colonne ID\_client on peut appliquer la jointure des deux tableaux **Clients** et **Commandes.**
2. Le composant tMap est le plus convenable qui permet d’appliquer la jointure entre les tables d’entrées.



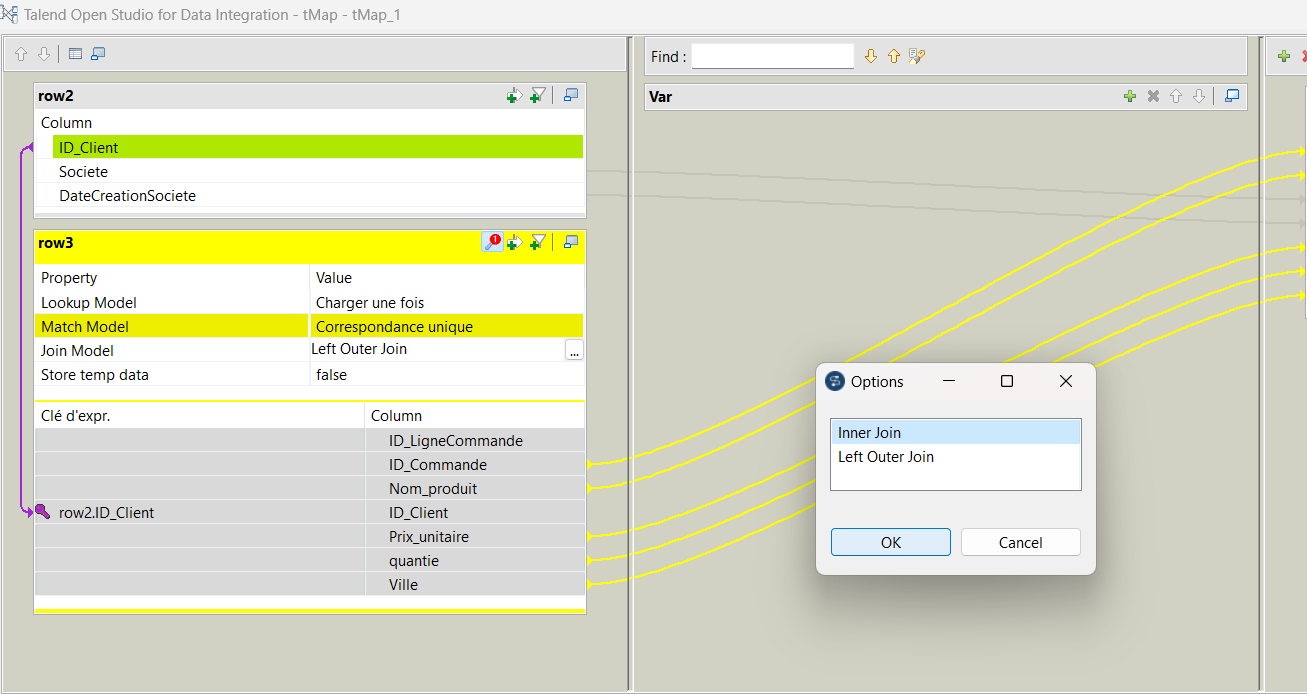
On ajoute tmap et on le connecter avec les deux composant precedant comme dans la photo apres pour la confugiration on click deux fois sur tmap .

en ajoute un table de sortie



Pour copier les colonnes coller les colonne

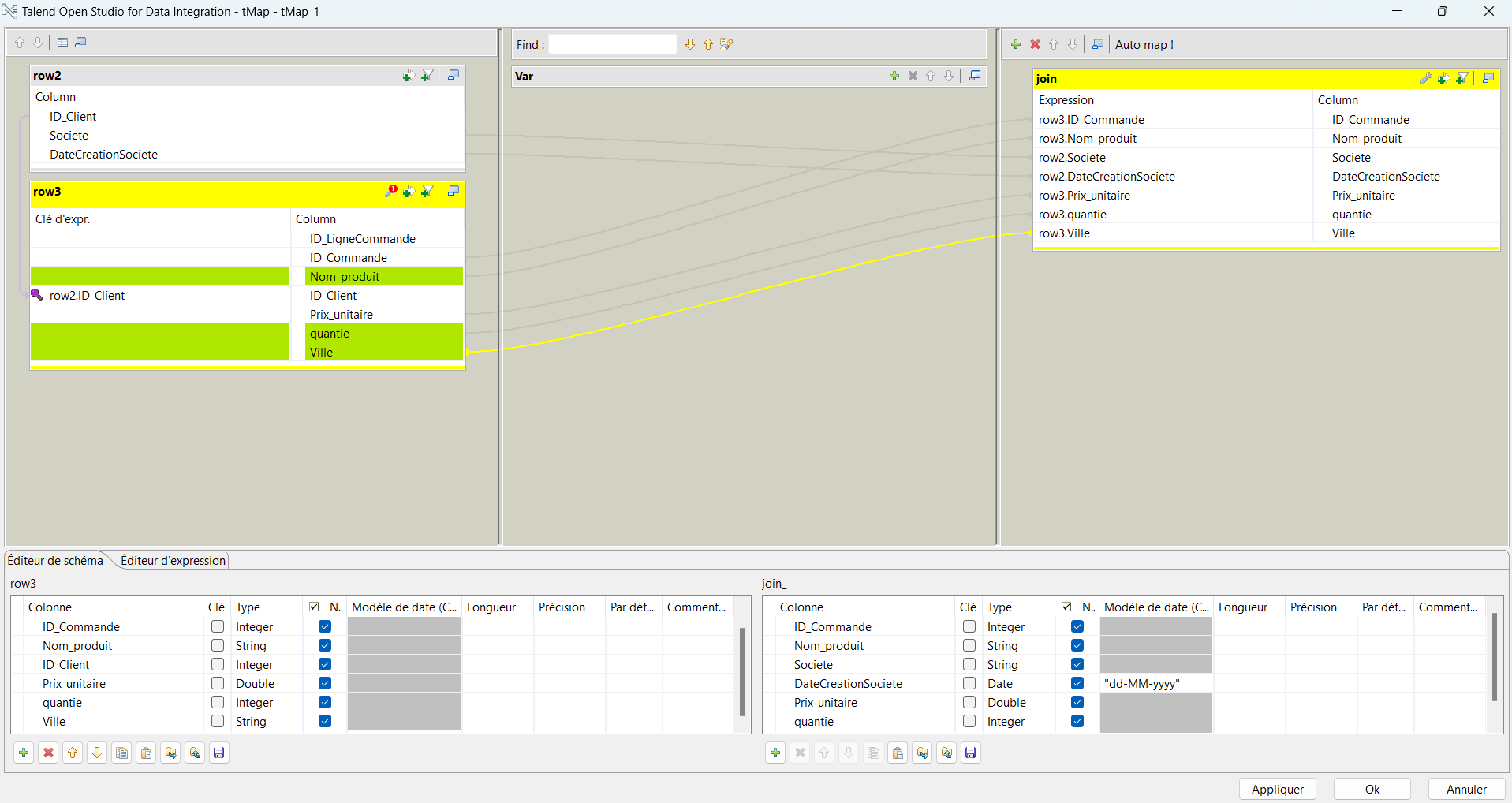
apres sa on fait la jointure entre les tables a l’aide de la colonne ID\_client



Click pour changer l’options de jointure changer a inner join click sur ok

On glisse la colonne ID\_client de row2 vers row3 dans la même colonne pour faire la jointure

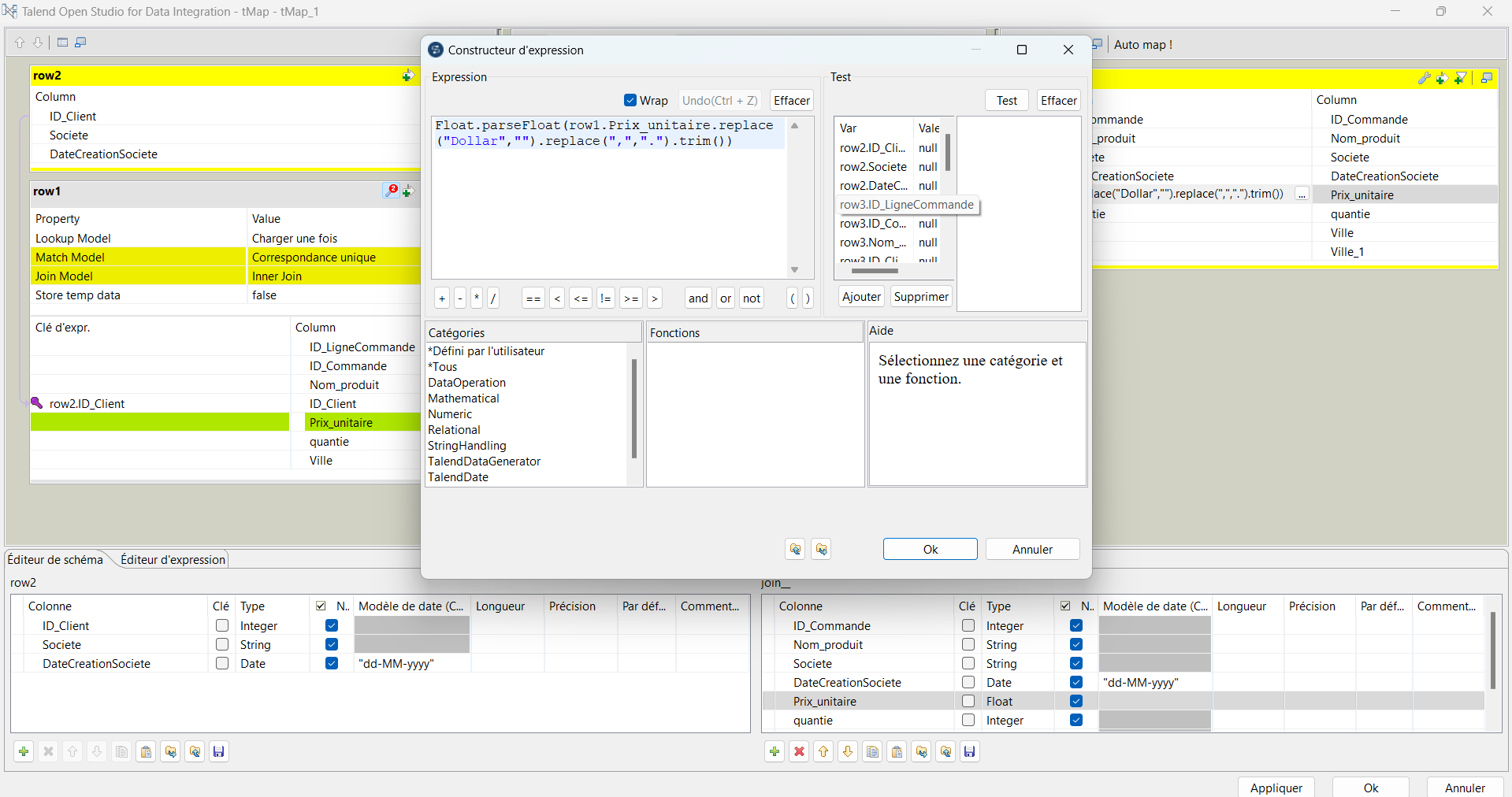
La même chose pour les colonne d’entrées et les colonne de sortie



Pour appliquer tout les modification afin davoir une resultat bien definie

1.click pour afficher le constructeur.

2.On ajoute la fonction suivante pour changer le type de prix\_unitaire de string vers float.



4.appliquer tous les modification sur la schémas.

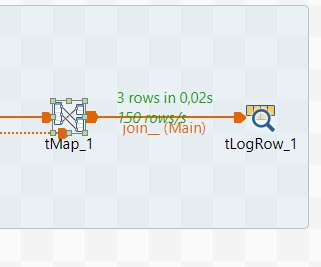
3.click sur ok pour enregistrer la fonction.

1. C’est le moment pour tester mon Job Talend et ses composants préalablement ajoutés,tlogRow est le composant qui permet d’afficher les données sur la console sous forme de texte/tables .

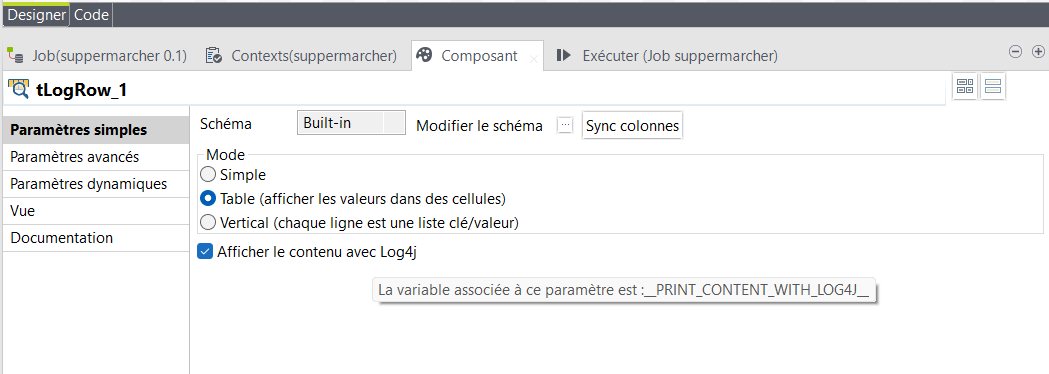
On ajoute **tlogRow** sur l’espace de modilisation on glisse la ligne de **tMap** vers **tlogRow**

La ligne qui fait la relation.

Le composant tlogrow pour l’output



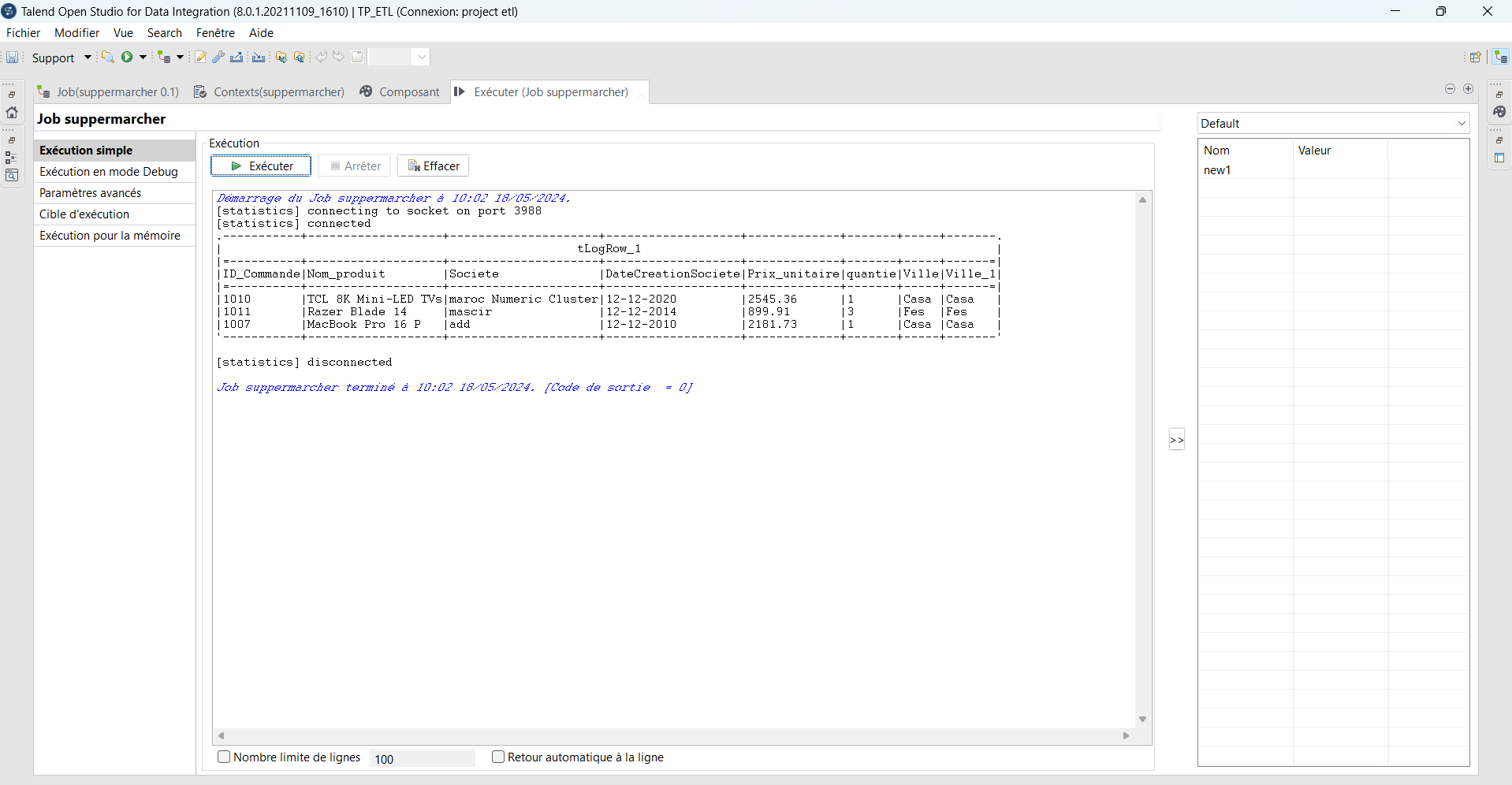
Pour la confuguration de se composant c’est de click sur table et sync colonnes



1. click sur table.

2. sync colonnes. 3. click sur la fenetre de executer.

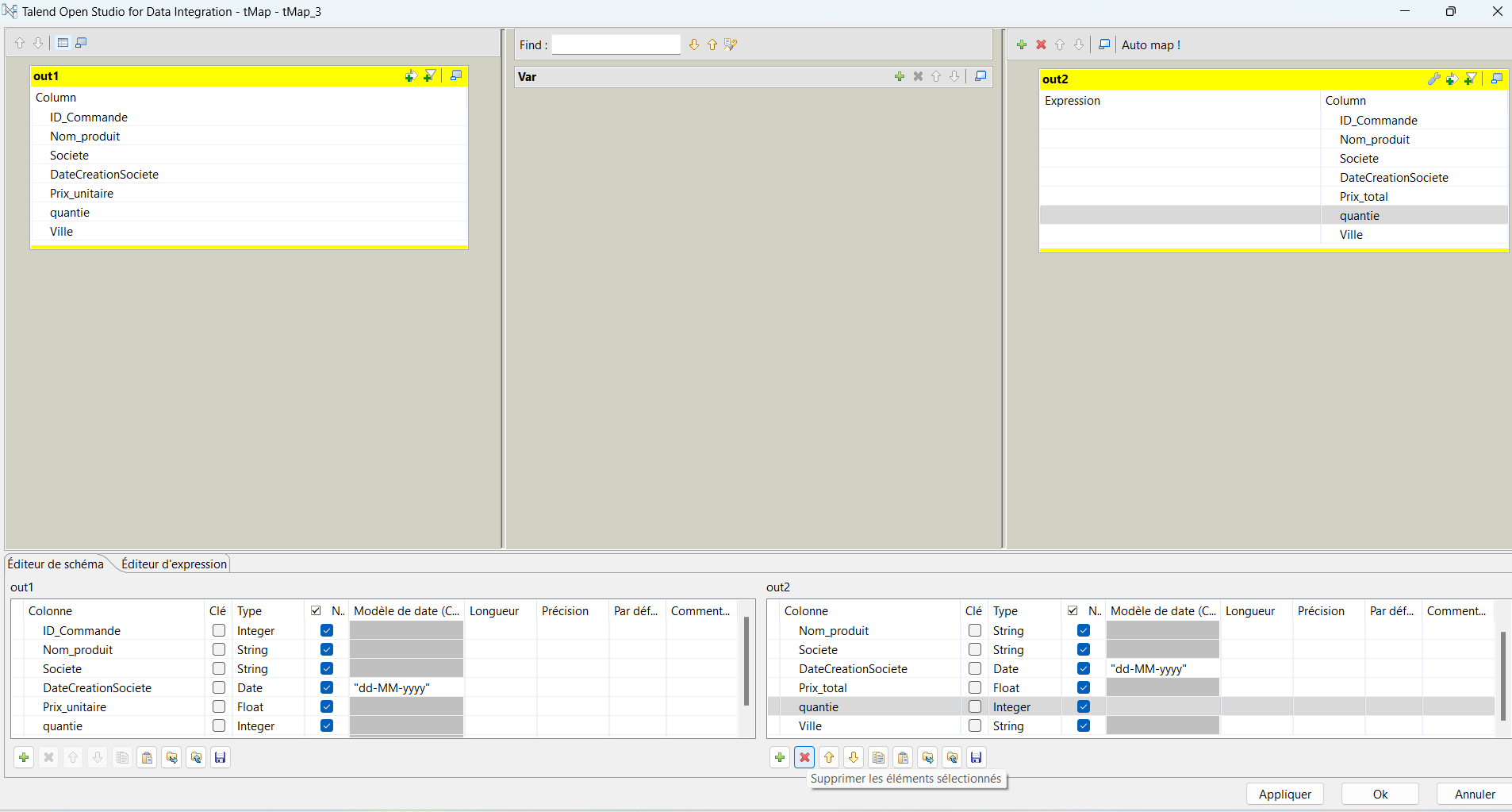
Click sur Executer.



9. Nous aimerons perfectionner le Job Talend, au lieu de garder les deux colonnes Prix\_Unitaire et Quantite, nous aimerons avoir une colonne PrixTotal, dont l’expression conditionnelle est PrixTotal = Prix\_Unitaire \* Quantite, nous pouvons résoudre ce problème au niveau de Talend a l’aide de composant **tMap** comme suit :

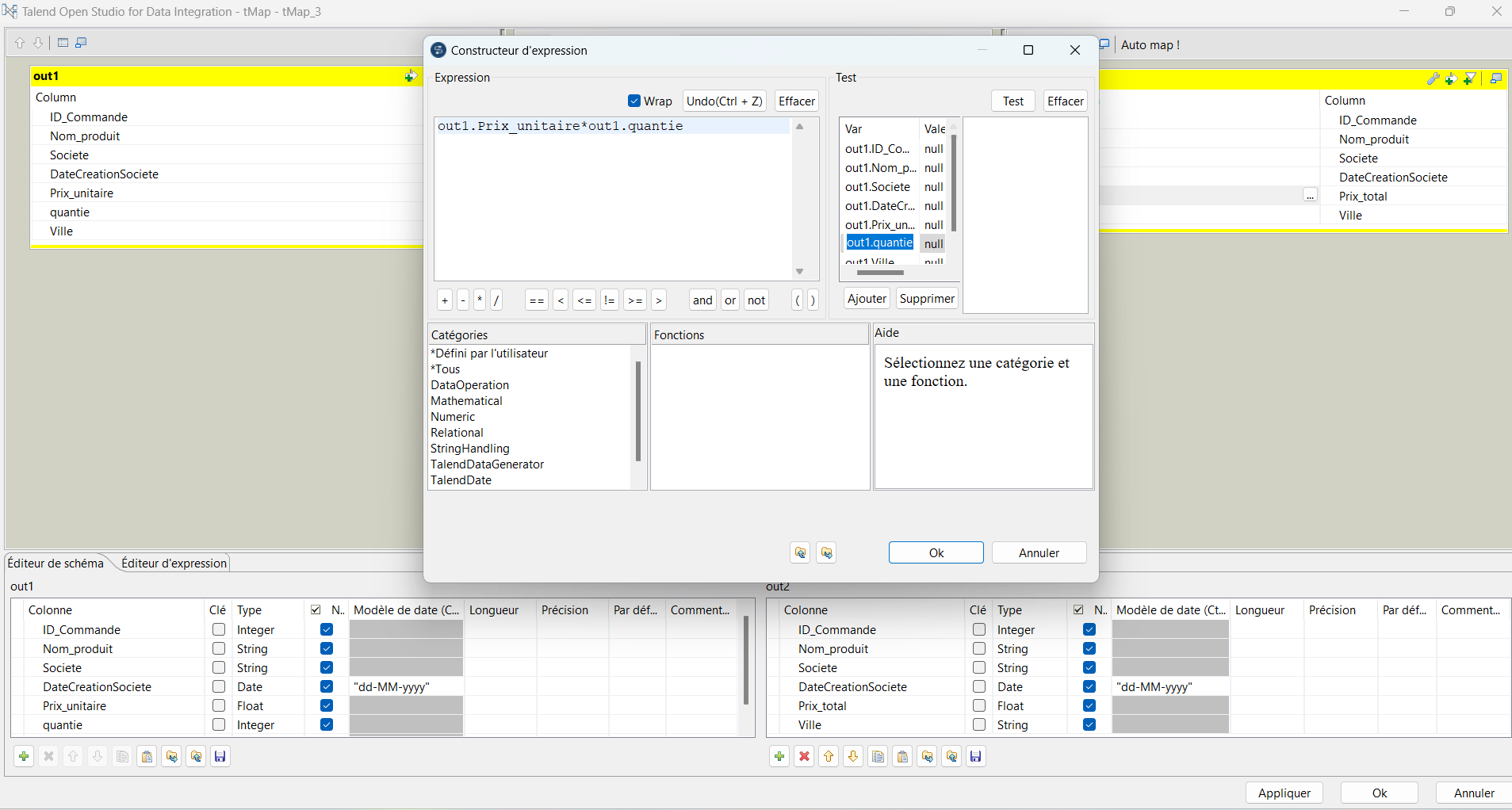
Dans le composant tmap : 3. Ajoute une variable dans le constructeur.

2.on ajoute prix\_total



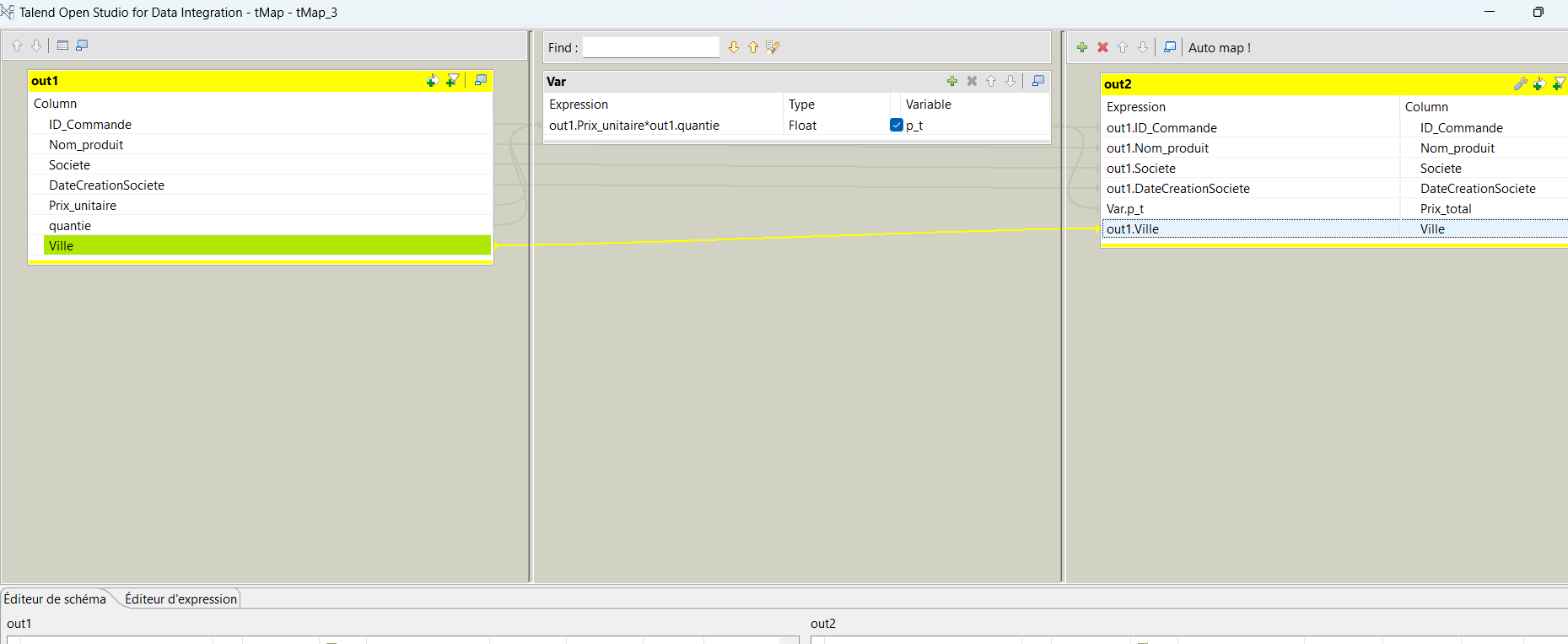
1. On supprime les deux colonnes quantite et prix\_unitaire.

Dans le constructeur.



1. on ajoute la fonction suivant pour cree la valeur de prix\_total. 2.click ok.

La resultat



* **Tache 4: Aggrégation des données**

Dans cette tache, vous devez calculer et afficher la **somme** des **prix totals** par **Société**.

1. Quel est le composant qui permet d’appliquer l’aggrégation**?**

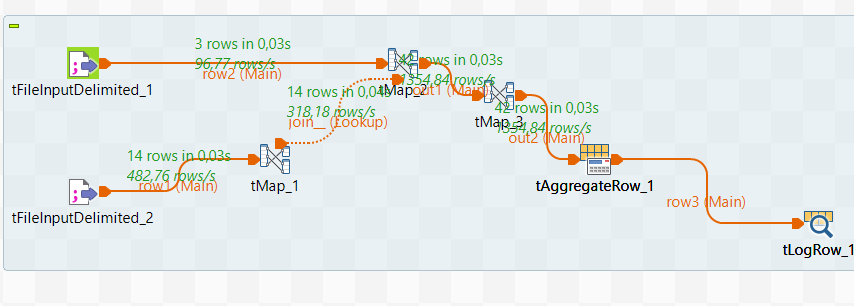
**taggregateRow**

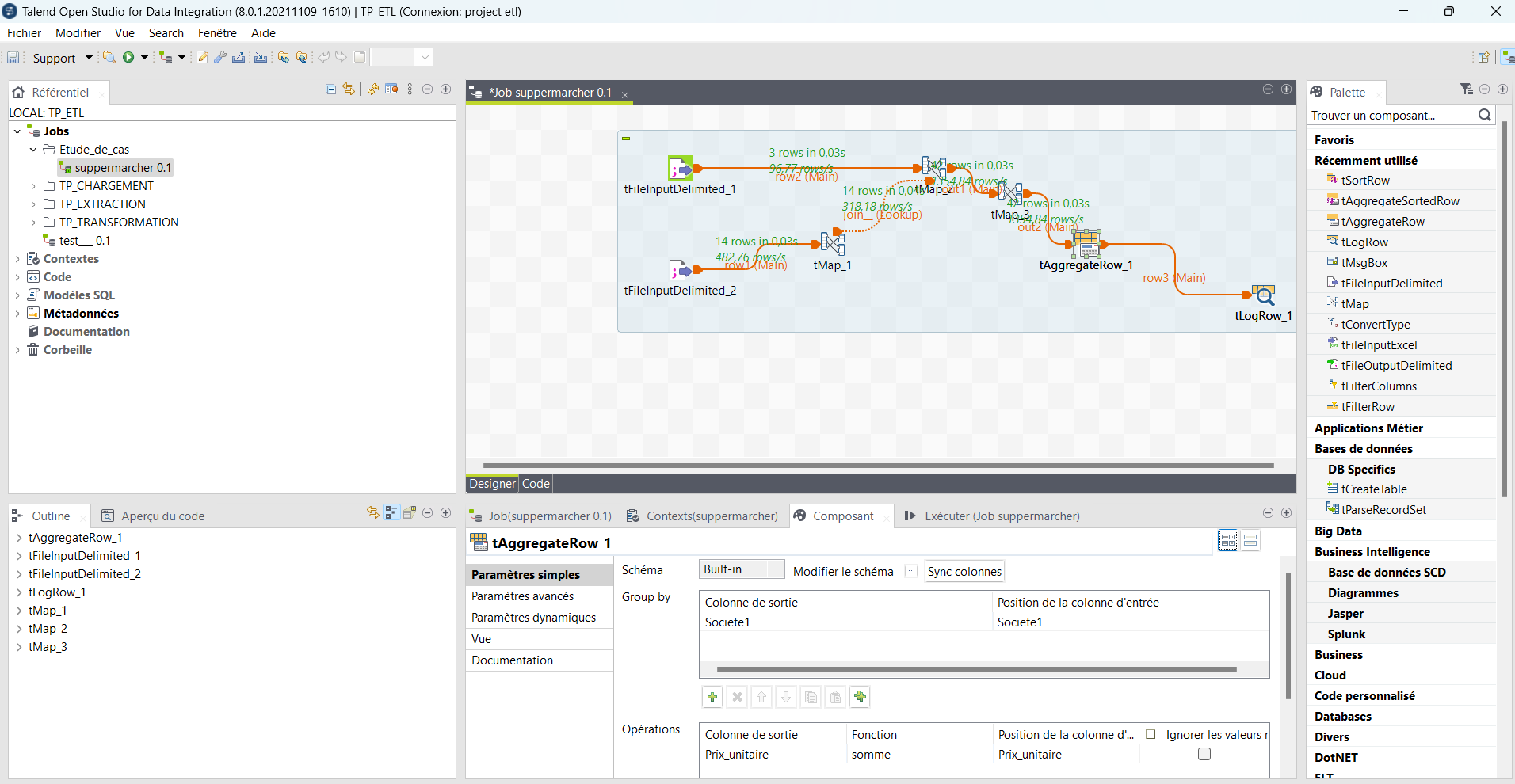
1. A la sortie de ce composant d’aggrégation, vous devez garder combien de colonnes ?

**2 colonnes**

1. Ajouter et configurer ce composant pour calculer et afficher la **somme des ventes** par **Société**.

Le composant souhaiter

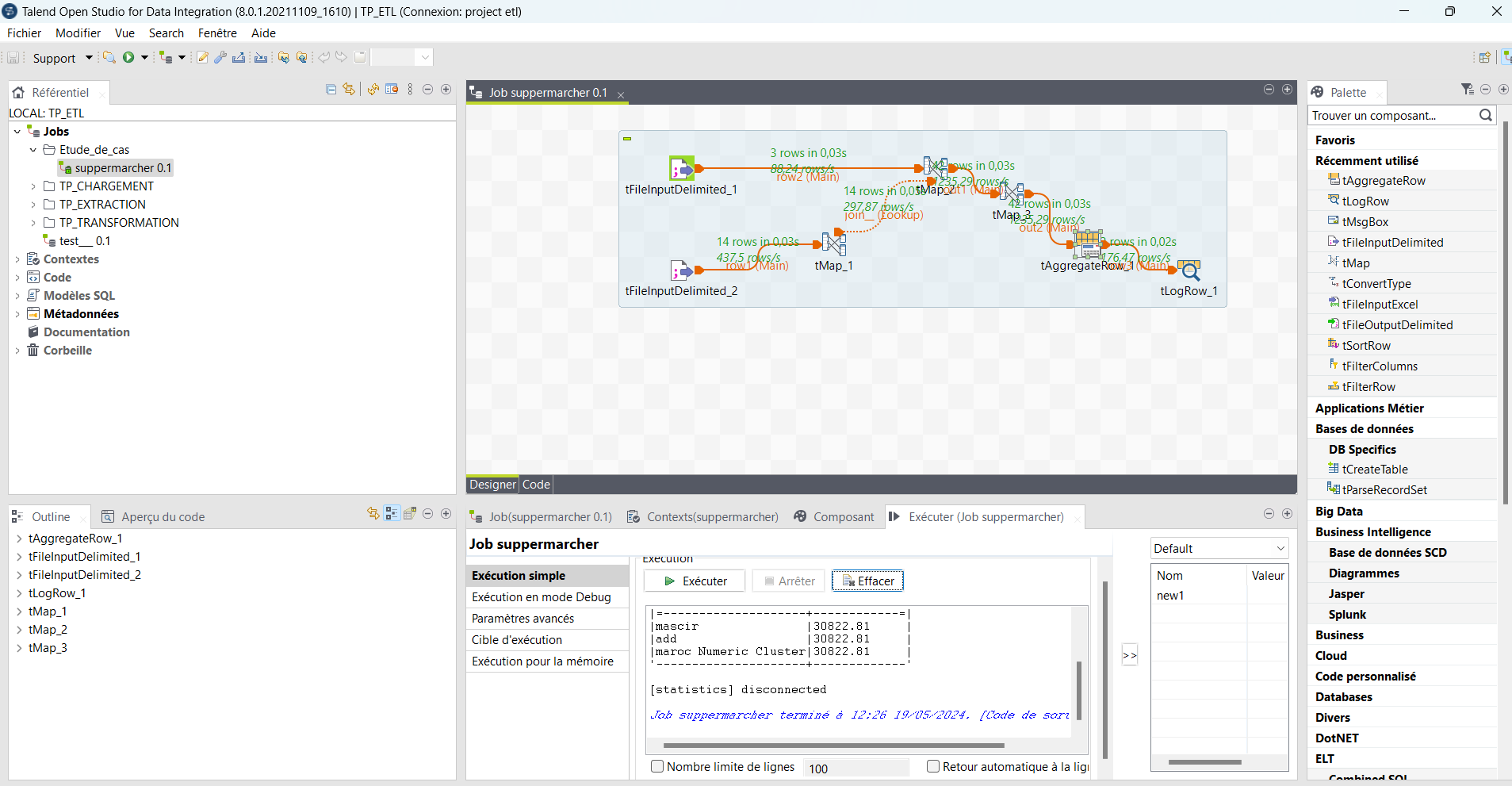




Somme\_vente

1.On selectionne la colonne qui va grouper 2. On va calculer dans la colonne somme\_vente 3. On va selectionne la colonne qu’on va calculer ,et la methode de calcule.

Apres l’execution :



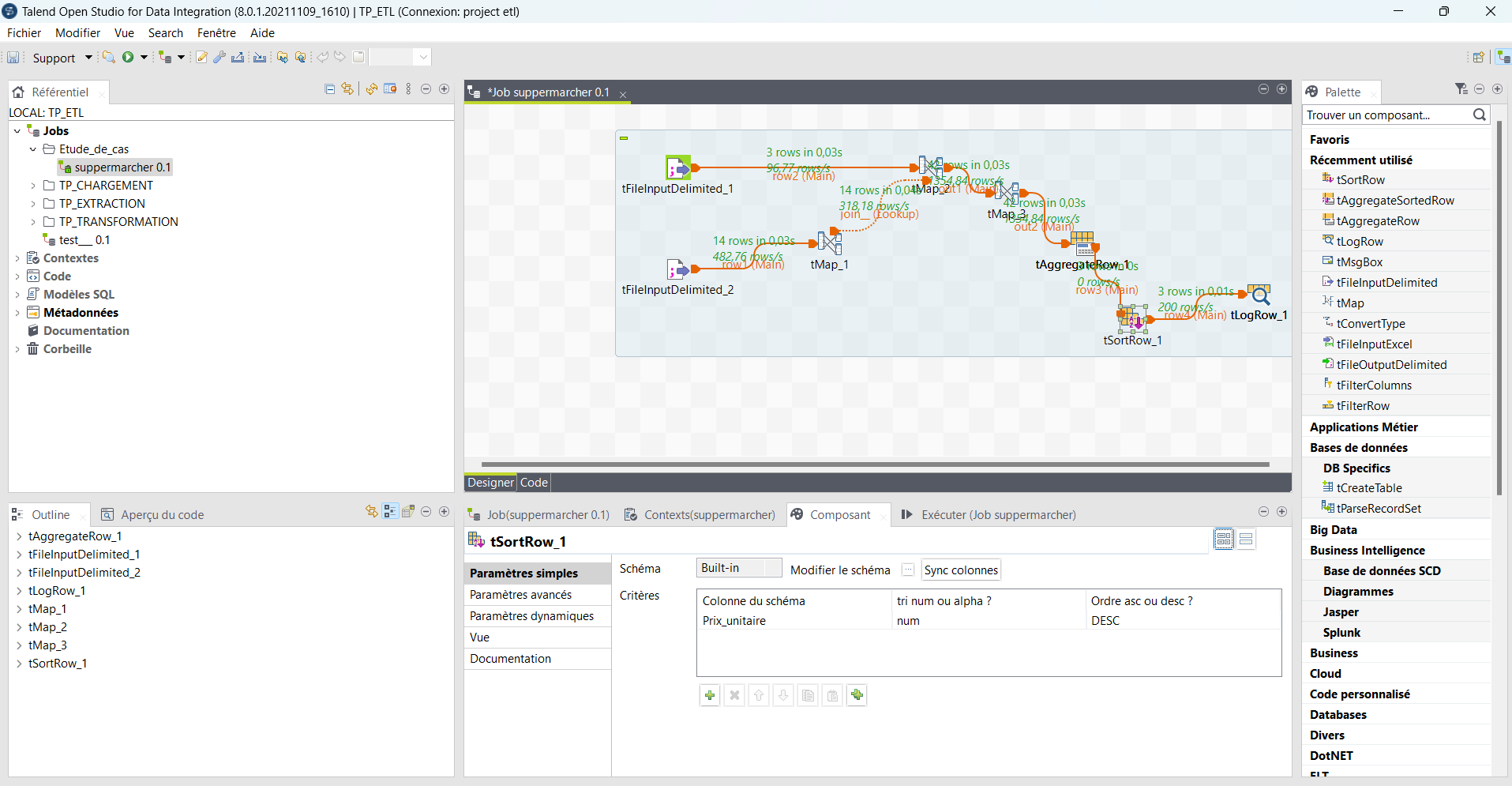
* **Tache 5: Trier les données agrégées.**

Dans cette tache, vous devez trier par order **décroissant** la **somme** des **prix totals** par **Société**.

1. Quel est le composant qui permet d’appliquer le tri sur une colonne

”tsortRow"

1. Ajouter et configurer ce composant dans Talend pour trier par ordre décroissant la somme des prix totals par Société. Note: la table de sortie contiendra seulement deux colonne: **Société | SommeVentes 1. le composant** tsortRow

****

**2. on selectionne la colonne qui va etre trier ,et num ,et la methode " DESC"**

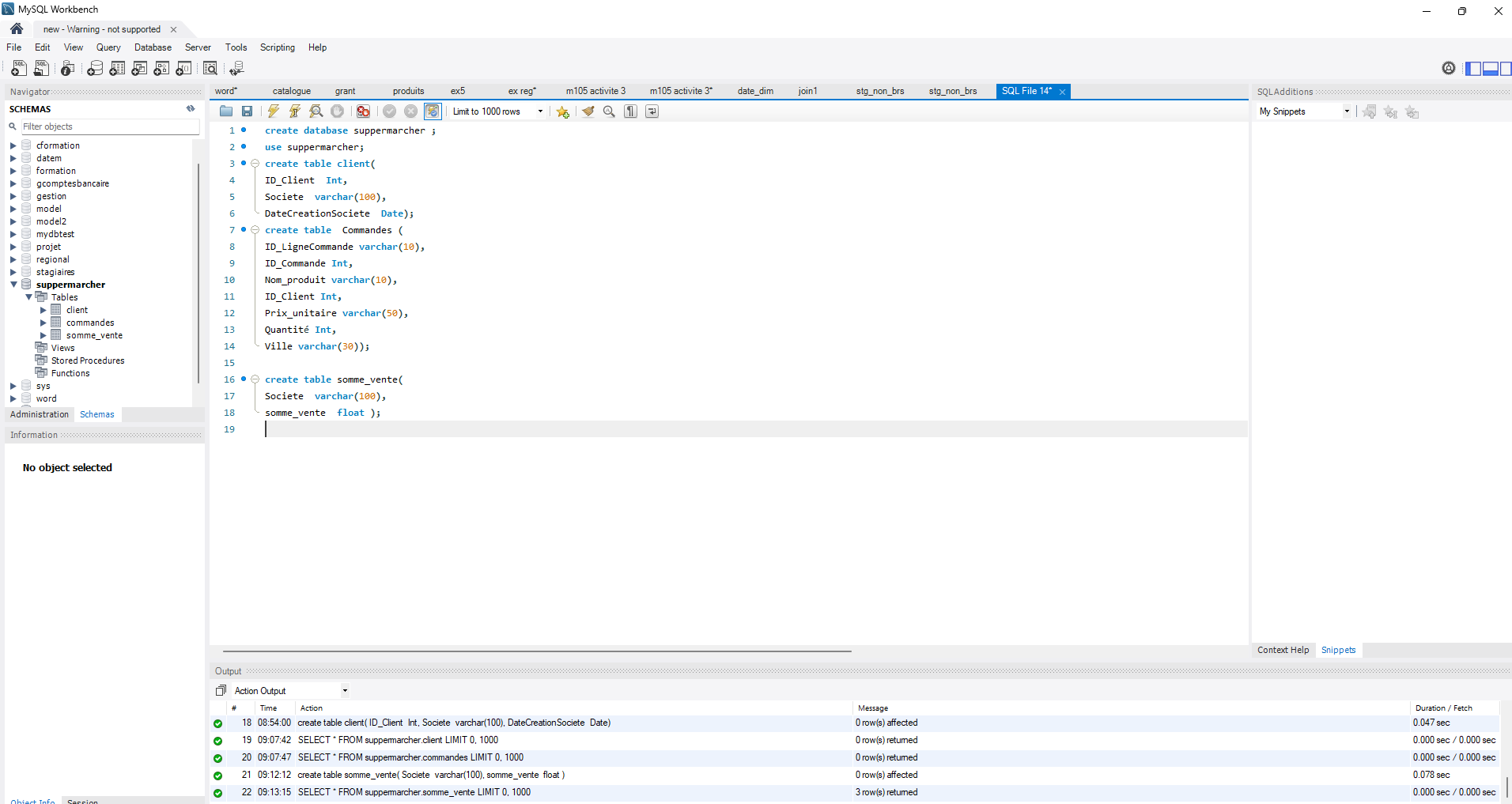
**Phase 3 : Chargement des données vers une BD**

Aprés les transformations effectuées, vous devez charger la table résultante dans une base de données MySql.

1. Quel est le composant qui permet de charger les données dans une base de données Mysql

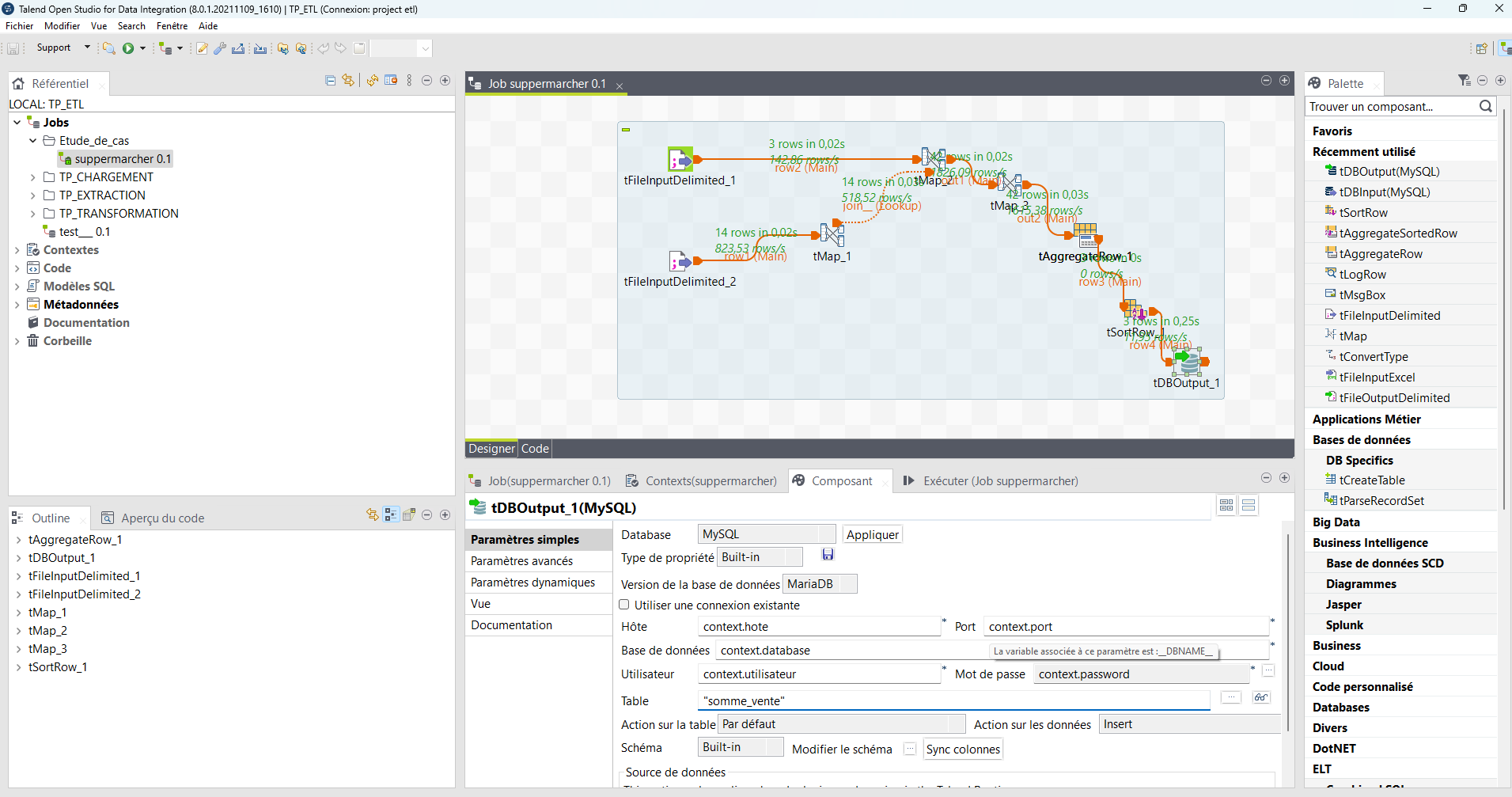
« tMySqlOutput »

1. Ajouter donc le composant sur Talend et toute la configuration neccessaire
   * + 1. On va cree une base de donner sur Mysql workbench contien la base de donner et les table qu’on va traiter, et on cree une table « somme\_vente » contient deux colonne societe ,somme\_vente.



2. pour le chargement de donner dans cette table on accede a talend et a l’aide du composant « tMySqlOutput »

2- on va remplire tout les donner par « context."identifent dans le groupe de context" ». 1-choisir MariaDB



3- pour inserer la resultat dans la table.

On cree un groupe de contexte contient les information pour acceder a MySql

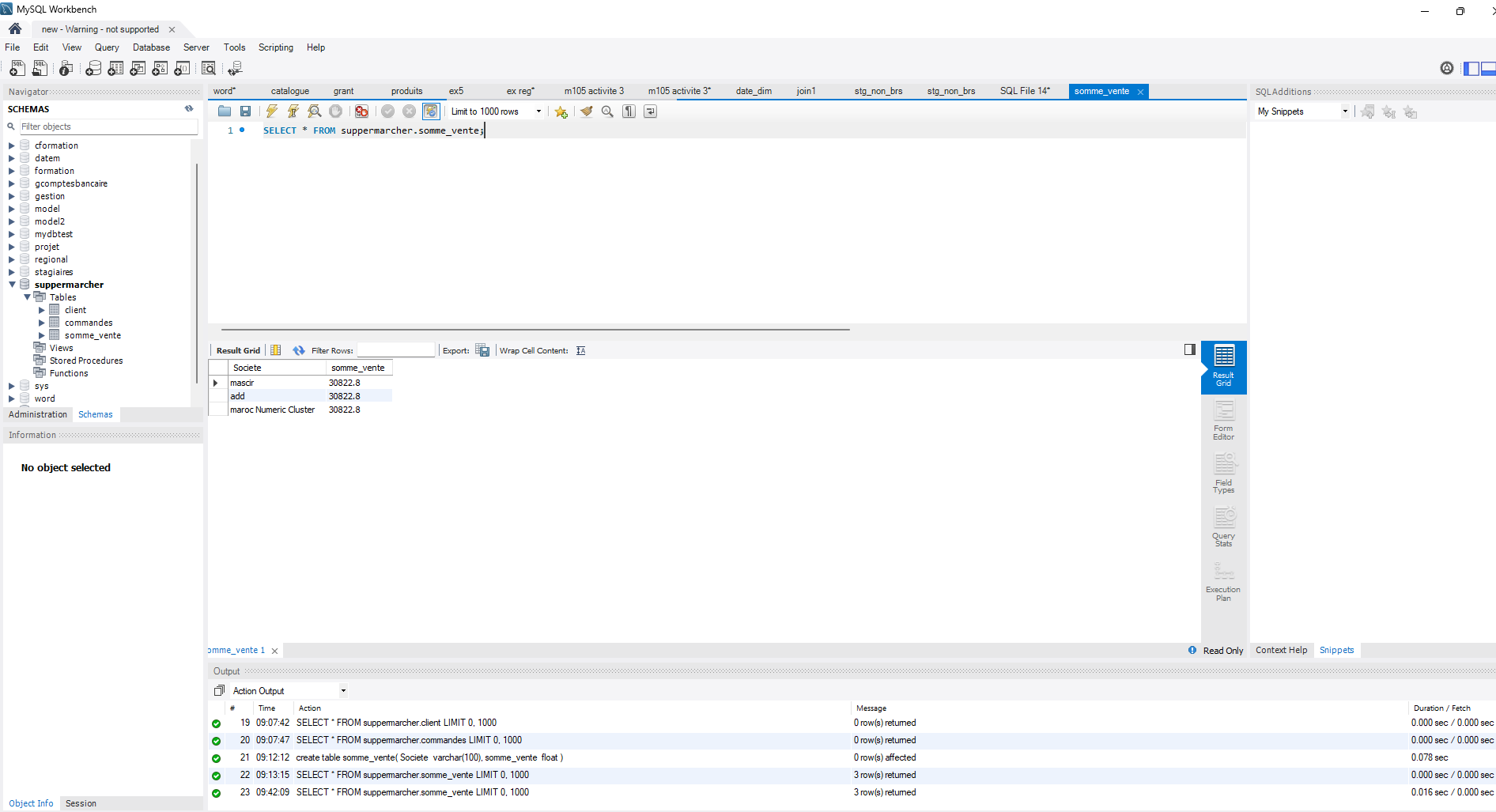
1. l’identifient de chaque valeur. 2. Les valeur pour acceder a la base de donner.



On va executer le jobe suppermarcher



On va acceder a workbench pour verifier la resultat par la requete suivant " SELECT \* FROM suppermarcher.somme\_vente; "



Le job a bien executer .

**Partie 2**

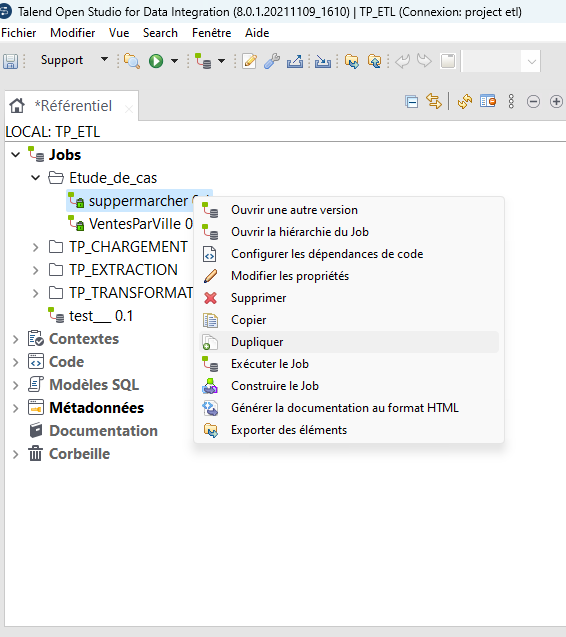
Dans ce qui suit nous aimerons faire sortir la somme des ventes par rapport à plusieurs dimensions à savoir :

1. La somme des ventes par Ville
2. La somme des ventes par Produit
3. La somme des ventes par Commande et Société
4. La somme des ventes par Ville et Produit

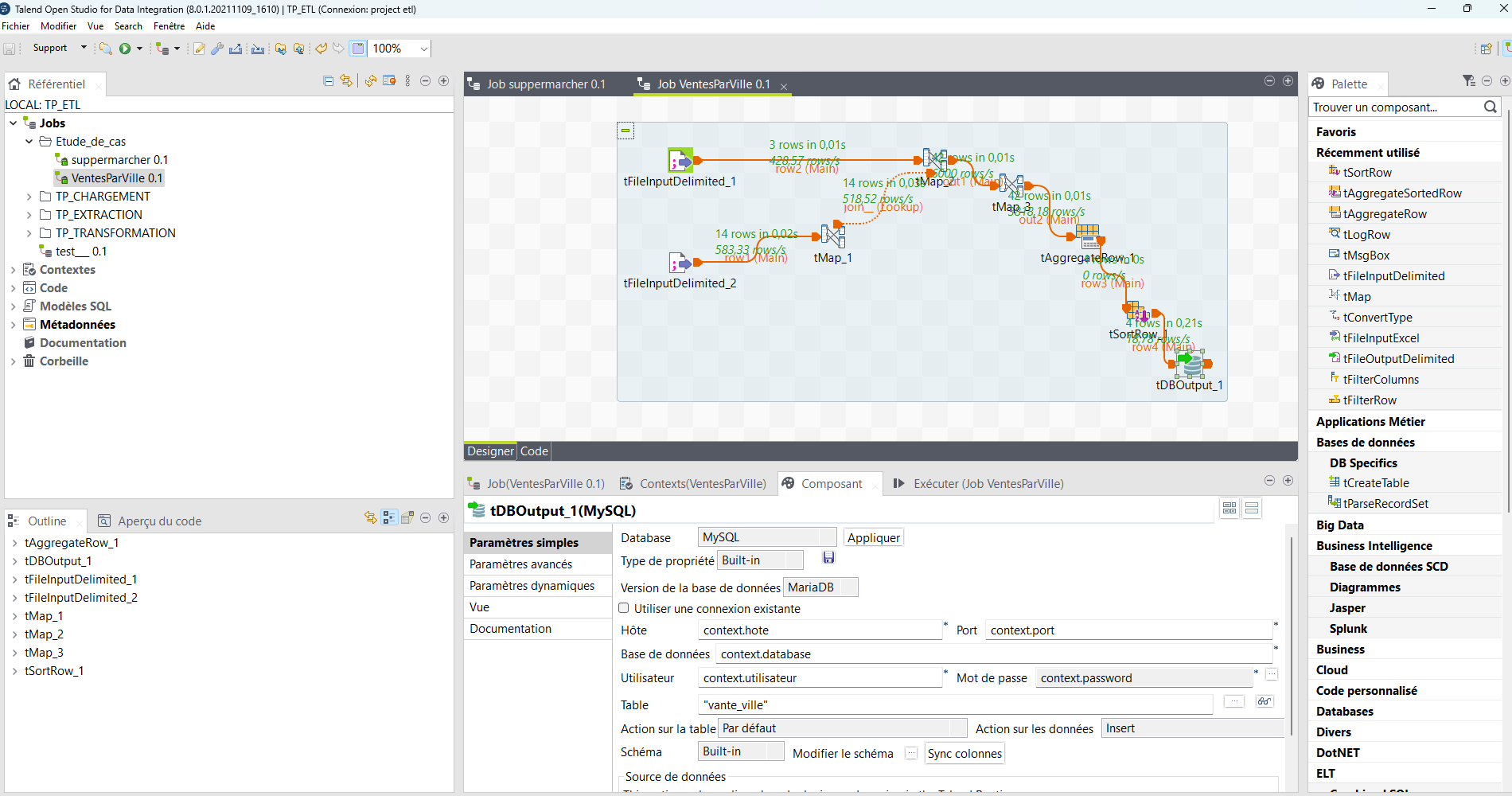
**La somme des ventes par Ville :**

Pour faire cela, suivez ces étapes :

1. Dupliquer le job de la partie 1 de l’étude de cas et nommer le job **VentesParVille**



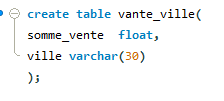
Apres la duplication et le nomage on accede a la nouvelle job **VentesParVille.**



1. Faites les adaptions neccessaire sur le Job

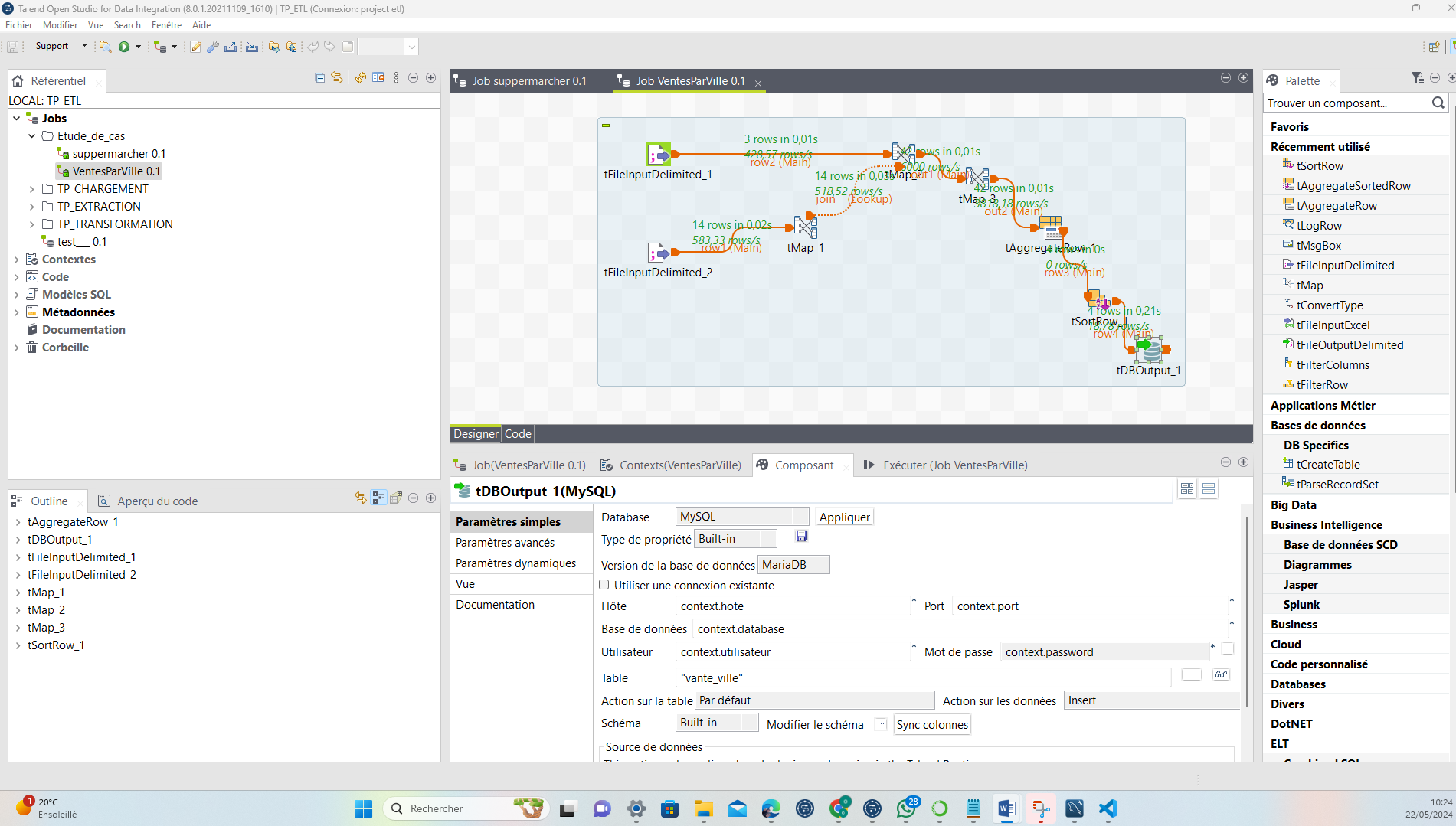
1-On va changer la colonne qui groupe "ville" dans le composant **tAggregateRow** 

2- on va cree une autre table sur workbench "vente\_ville"



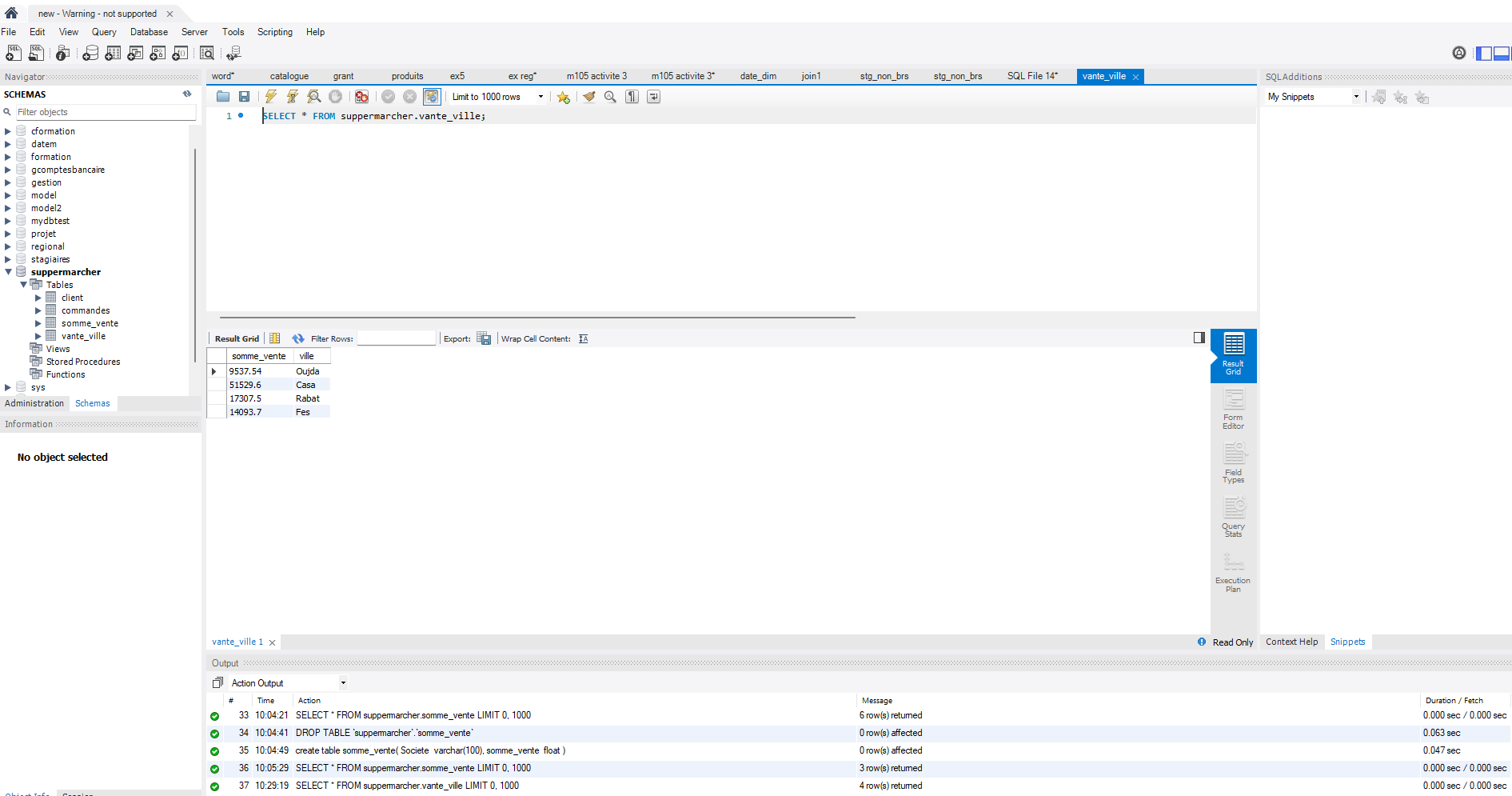
-sur talend avec le composent **tMySqlOutput**:

On change le nom du table vers "vente\_ville"



1. Tester le job

En execute le job **VentesParVille,**sur workbenchon execite la requete " SELECT \* FROM suppermarcher.vante\_ville; ".

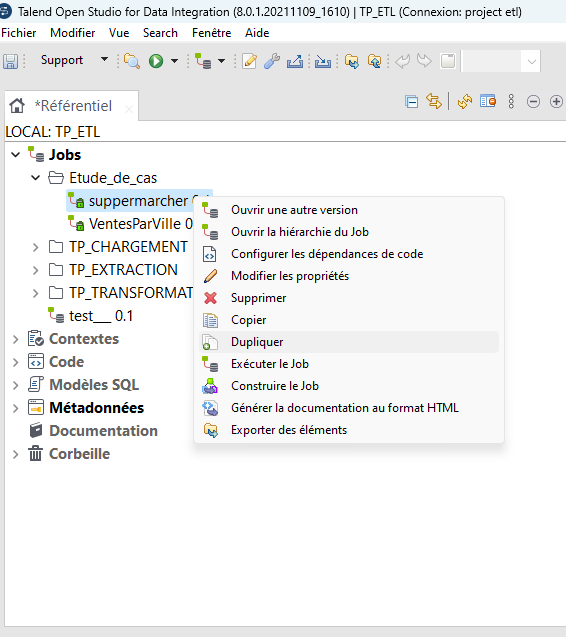
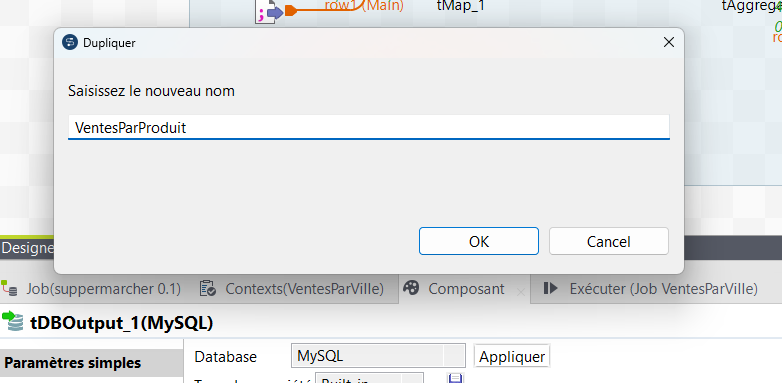


Job est bien executer.

**La somme des ventes par Produit :**

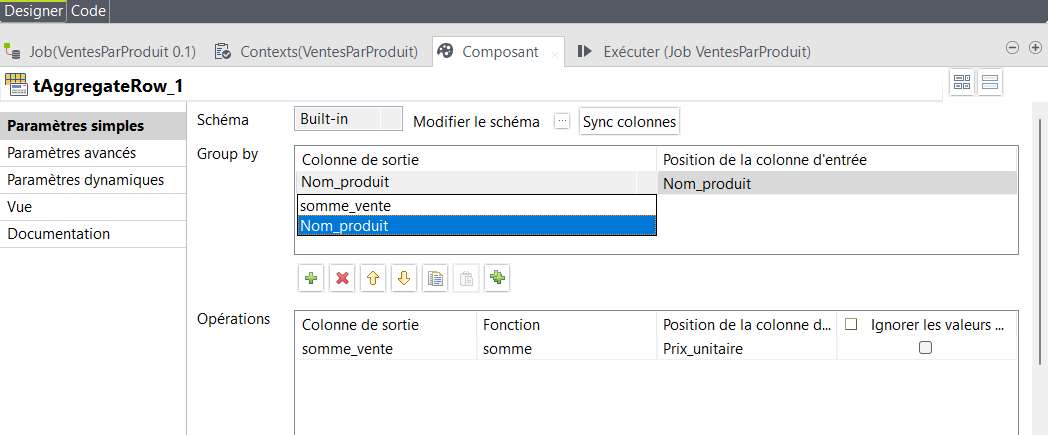
Pour faire cela, suivez ces étapes :

1. Dupliquer le job de la partie 1 de l’étude de cas et nommer le job **VentesParProduit**

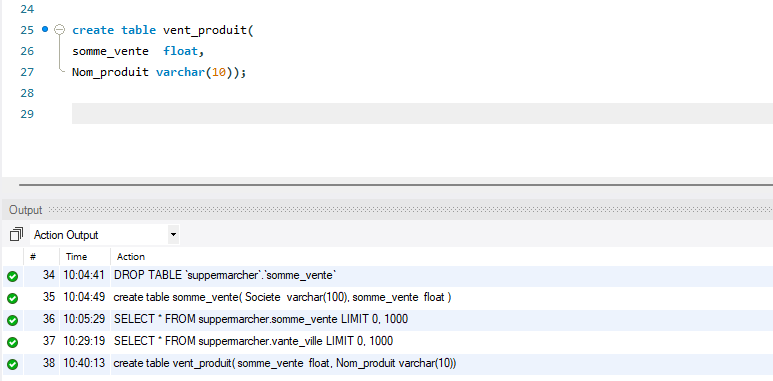
 

1. Faites les adaptions neccessaire sur le Job

1-On va changer la colonne qui groupe "nom\_produit" dans le composant **tAggregateRow**

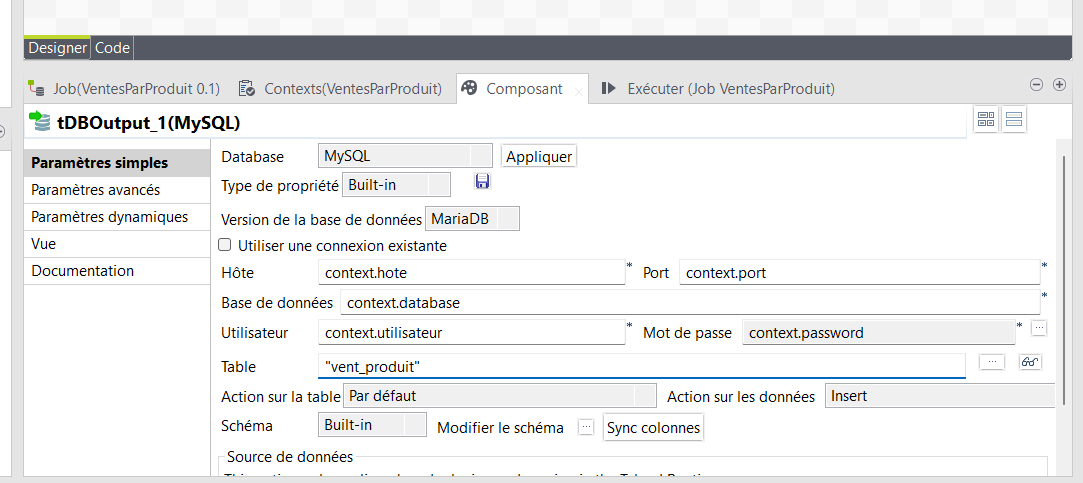


2- on va cree une autre table sur workbench" vent\_produit "



sur talend avec le composent **tMySqlOutput**:

On change le nom du table vers " vent\_produit"

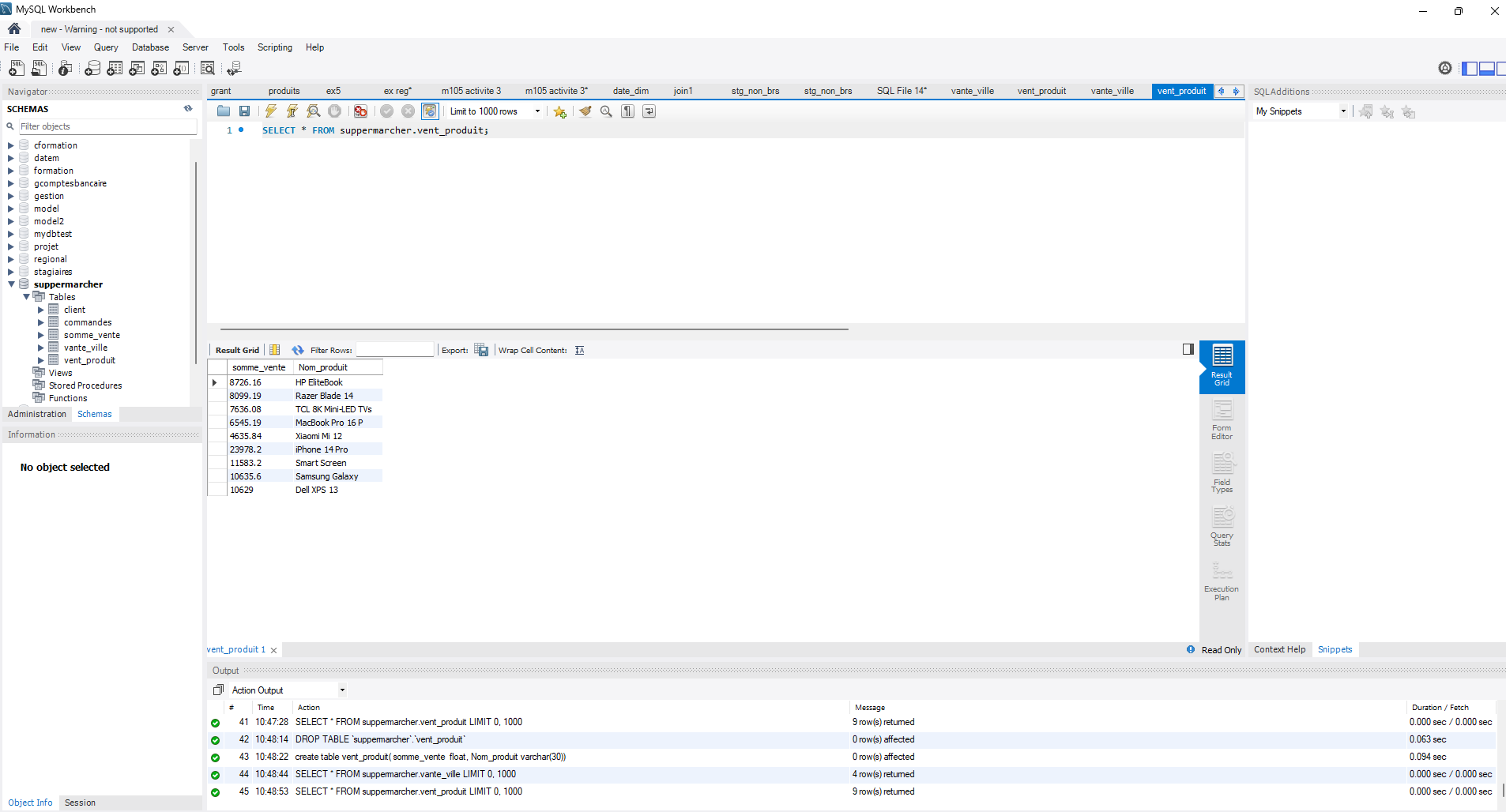


1. Tester le job

En execute le job



Sur workbenchon execite la requete " SELECT \* FROM suppermarcher. vent\_produit; ".

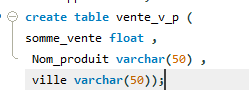


**La somme des ventes par Commande et Société :**

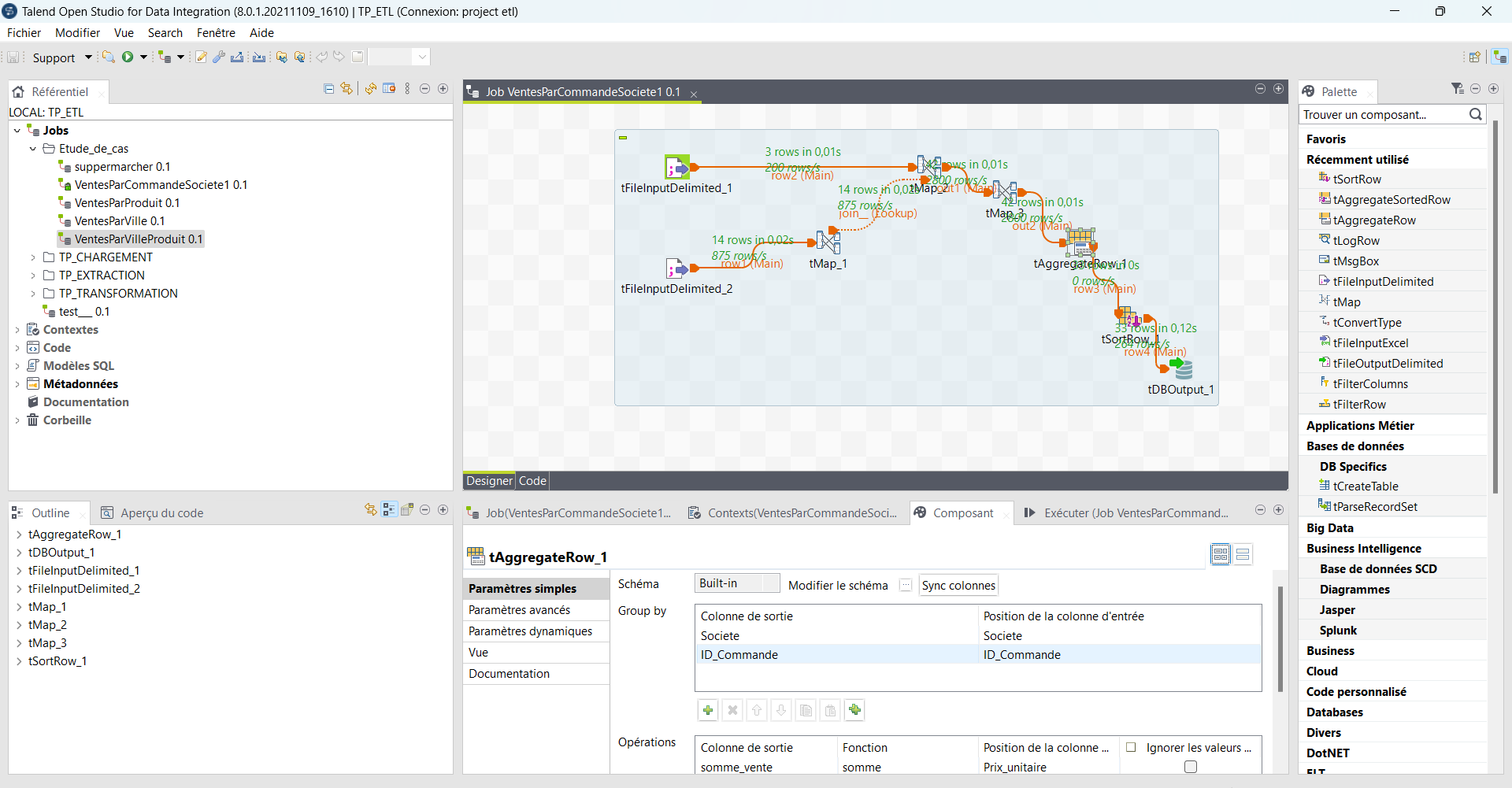
Pour faire cela, suivez ces étapes :

1. Dupliquer le job de la partie 1 de l’étude de cas et nommer le job **VentesParCommandeSociete**
2. Faites les adaptions neccessaire sur le Job

Cree table dans orkbench :

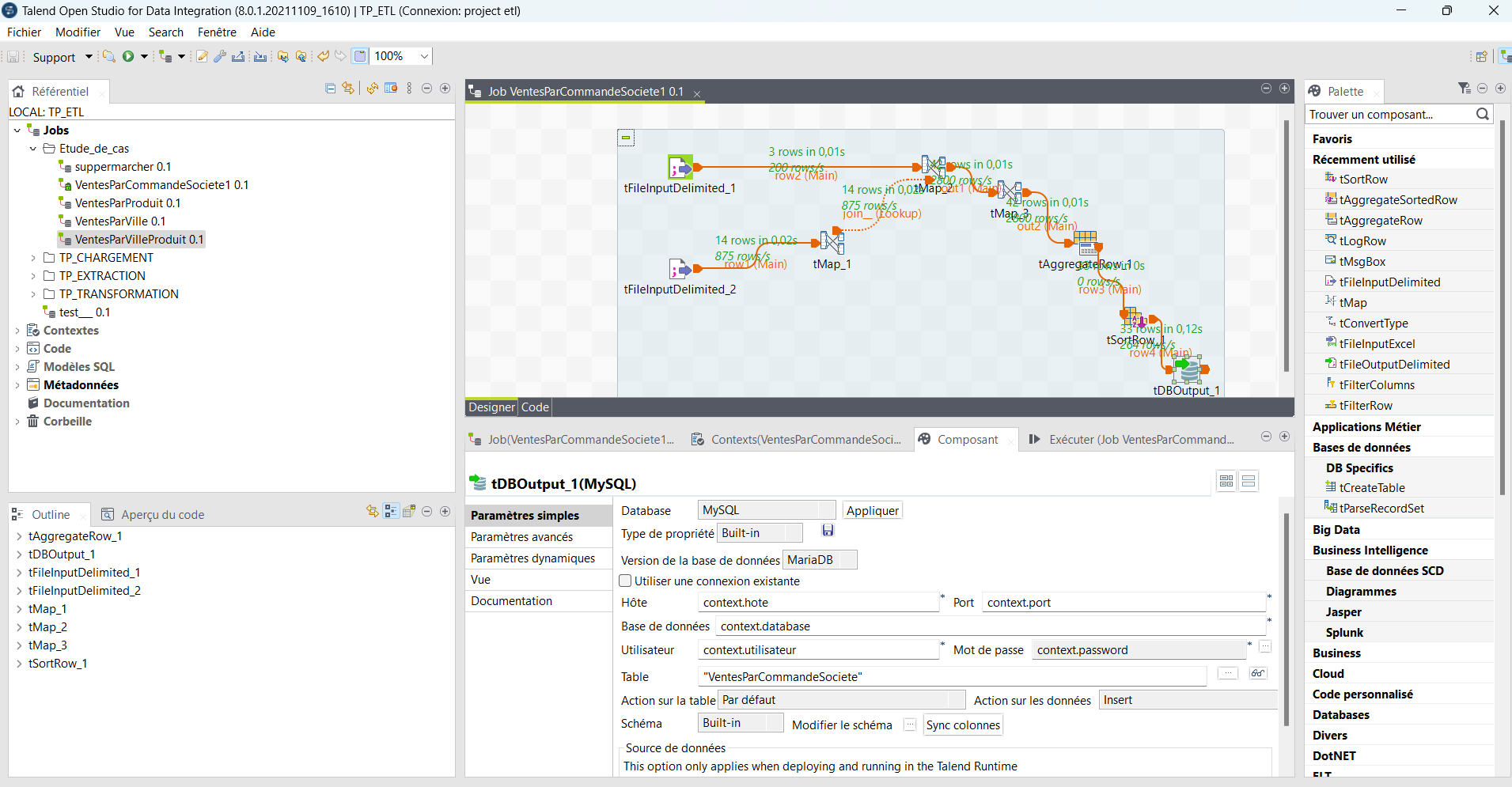


Changer dans le conposant taggregateRow et ajouter ID\_commande :

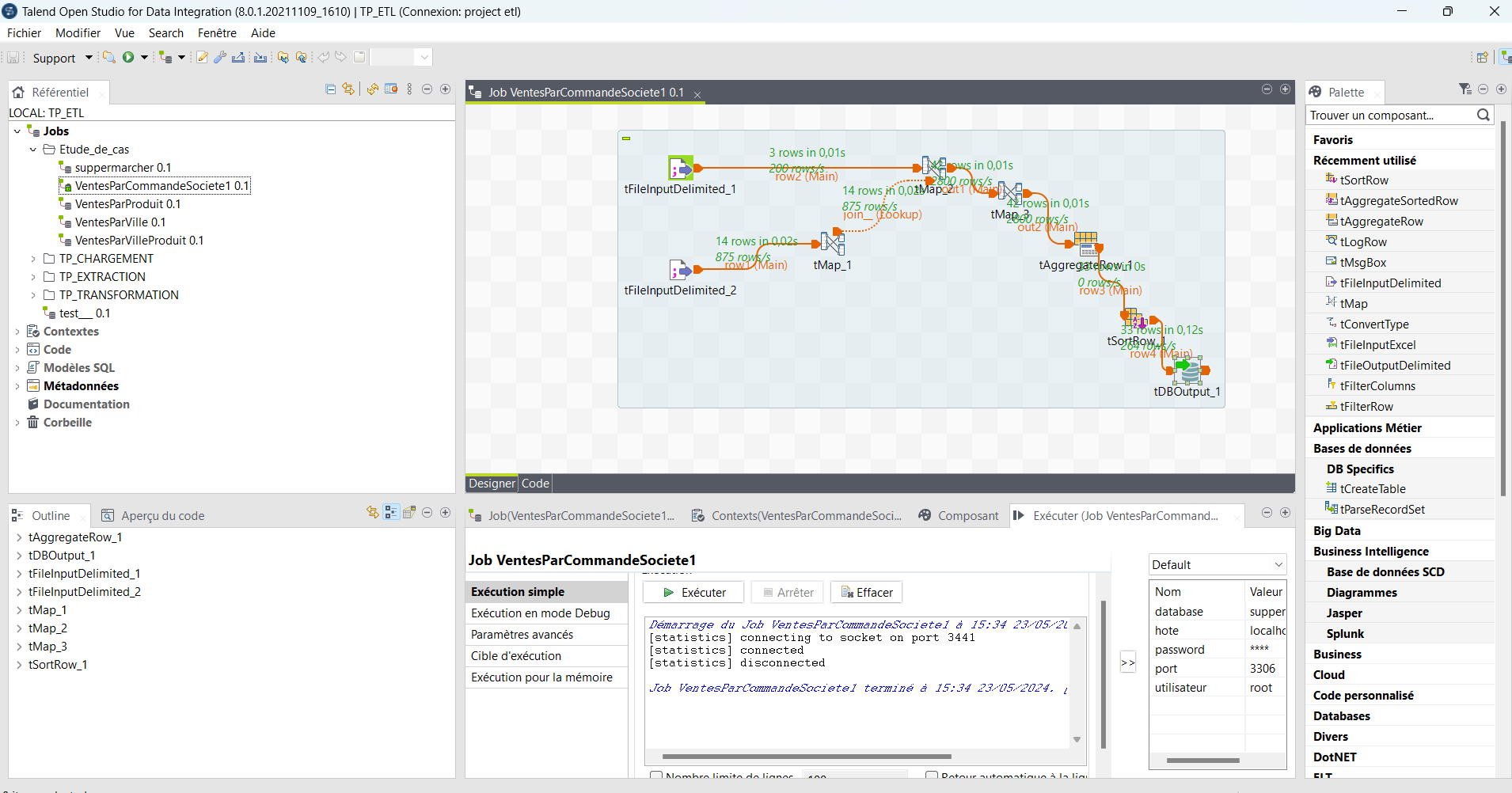


sur talend avec le composent **tMySqlOutput**:

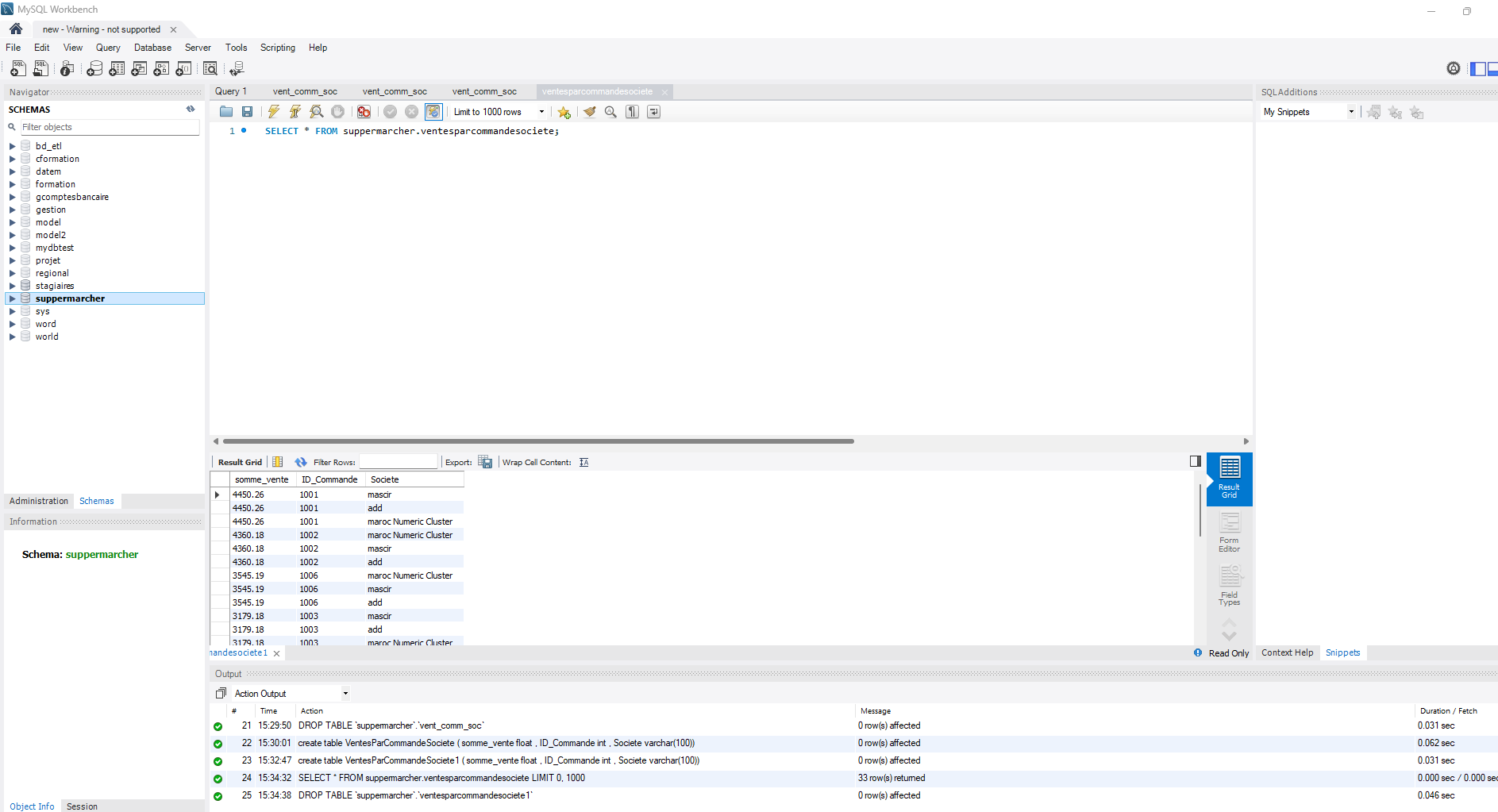
On change le nom du table vers " VentesParCommandeSociete" :



1. Tester le job



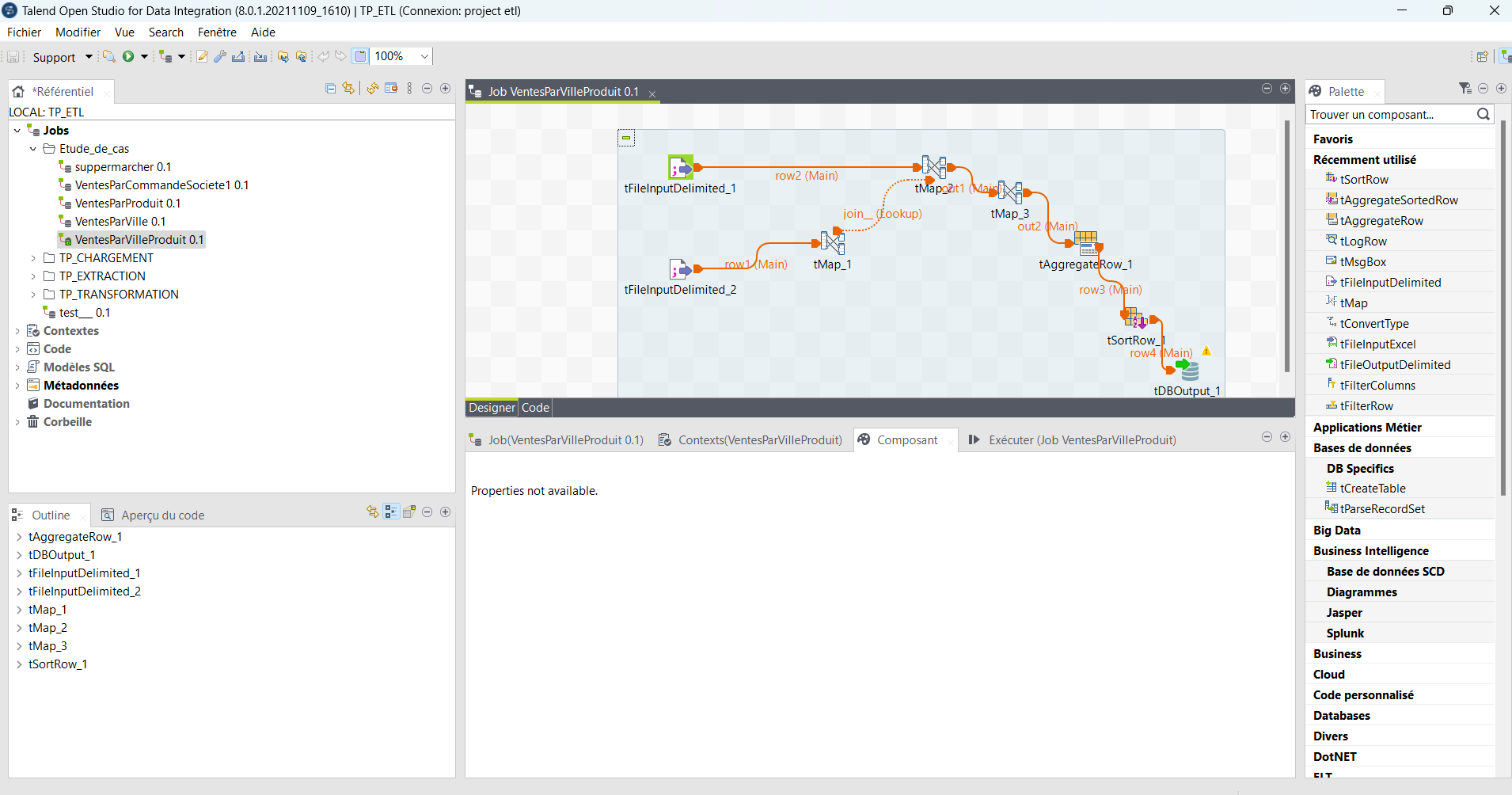
Sur workbenchon execite la requete " SELECT \* FROM suppermarcher.ventesparcommandesociete;".



**La somme des ventes par Ville et Produit :**

Pour faire cela, suivez ces étapes :

1. Dupliquer le job de la partie 1 de l’étude de cas et nommer le job **VentesParVilleProduit**

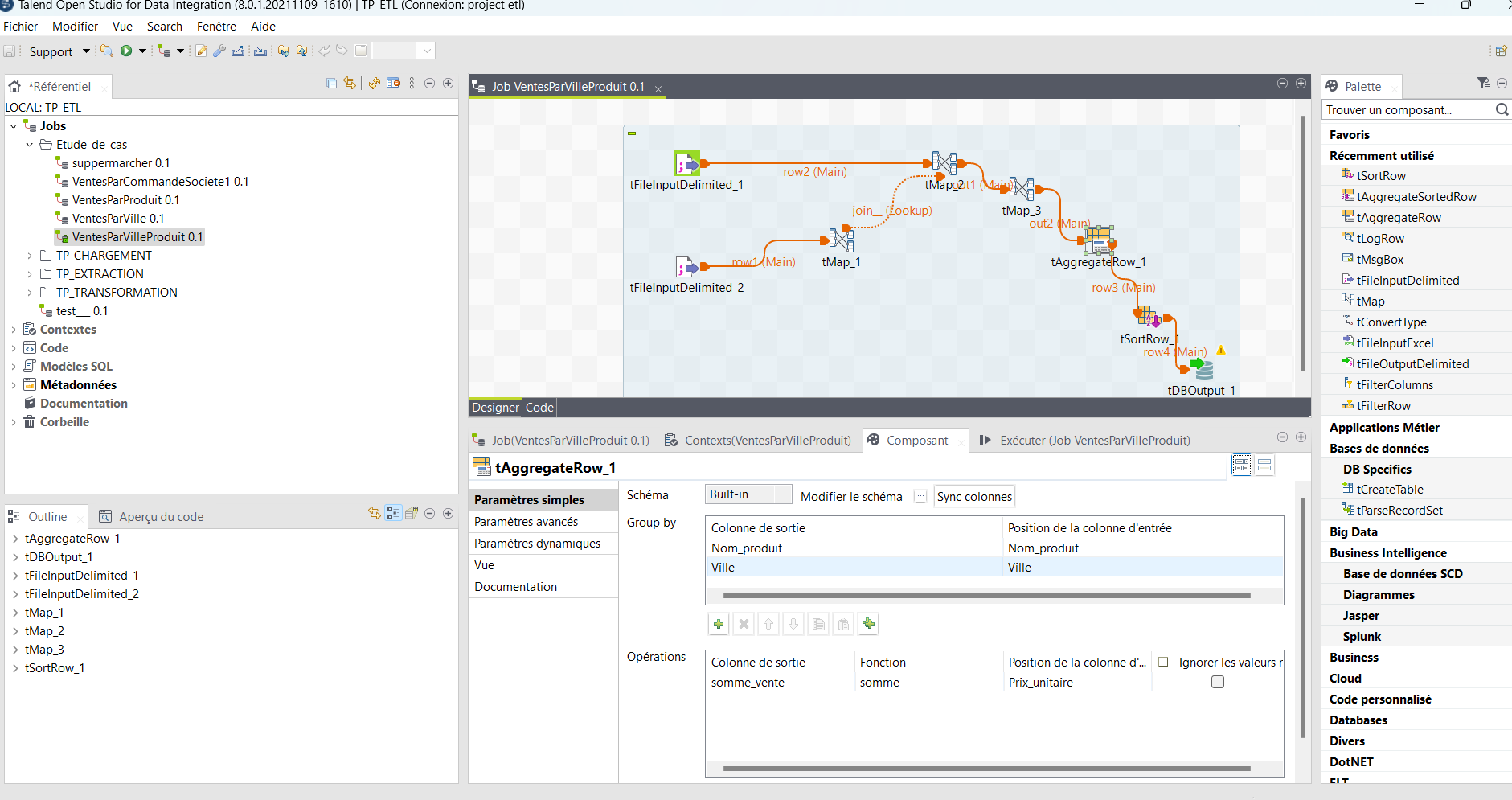


1. Faites les adaptions neccessaire sur le Job

On cree une table

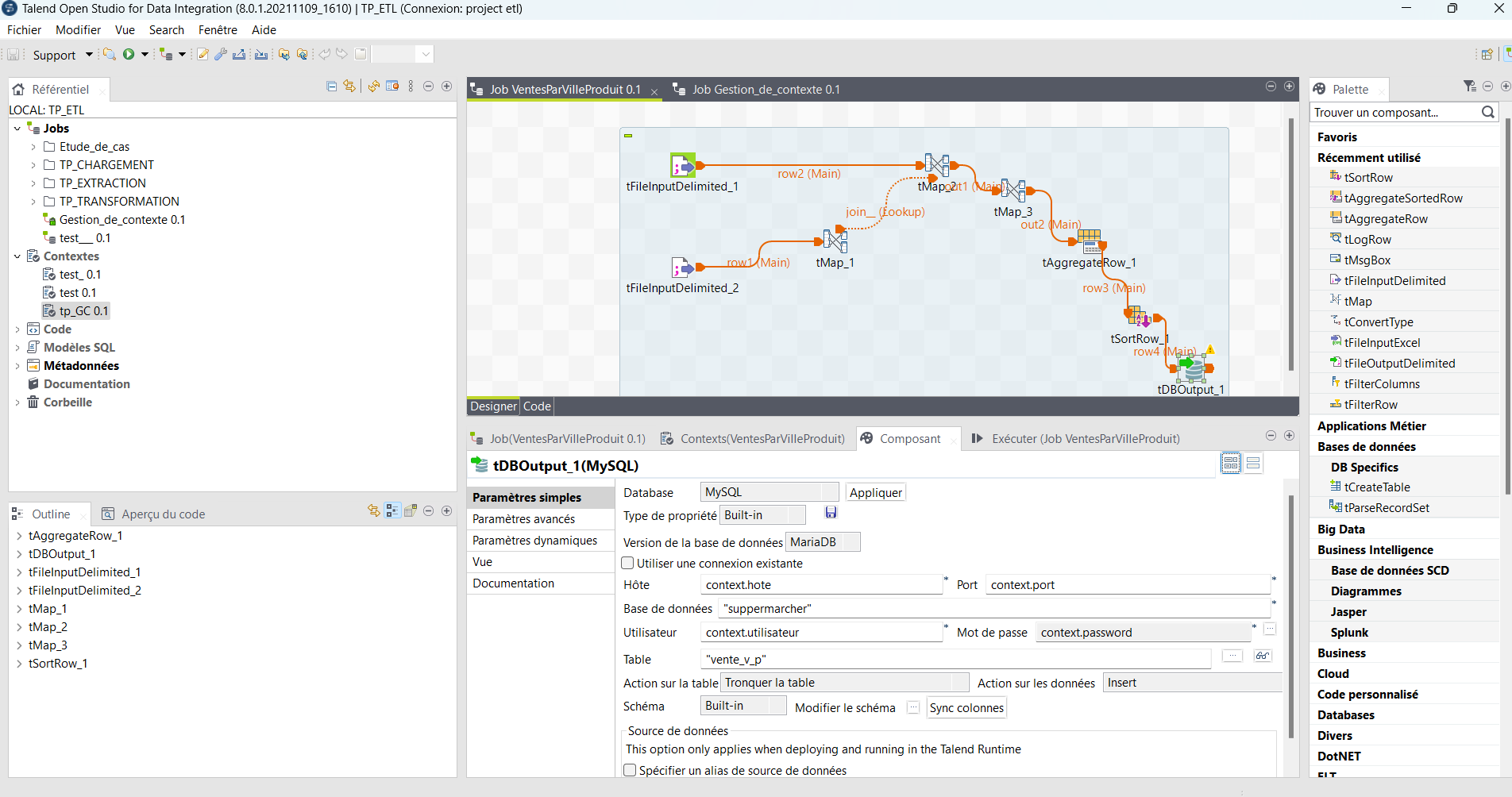


1. -On va changer la colonne qui groupe "ville" et "nom\_produit" dans le composant **tAggregateRow**

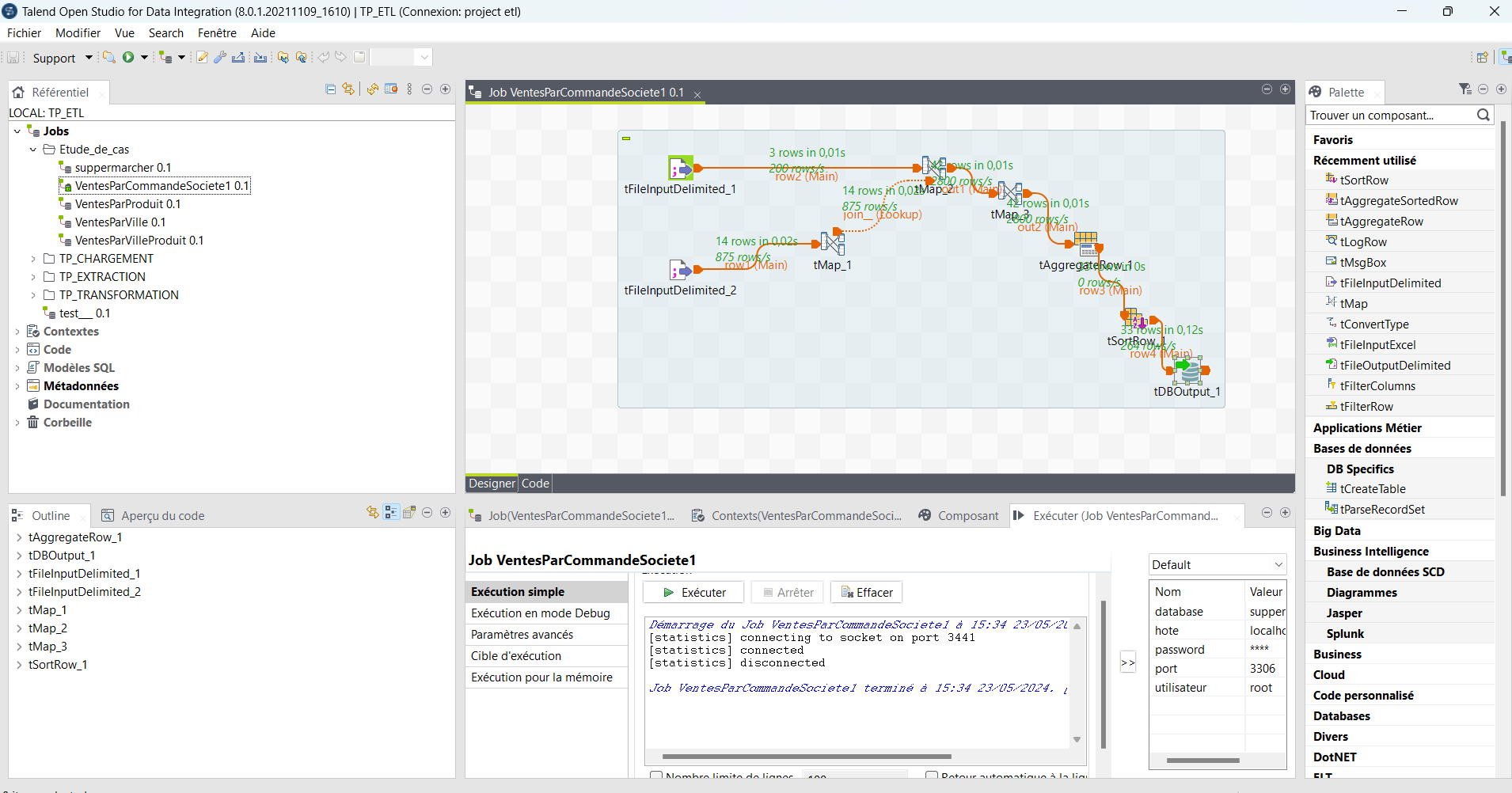


sur talend avec le composent **tMySqlOutput**:

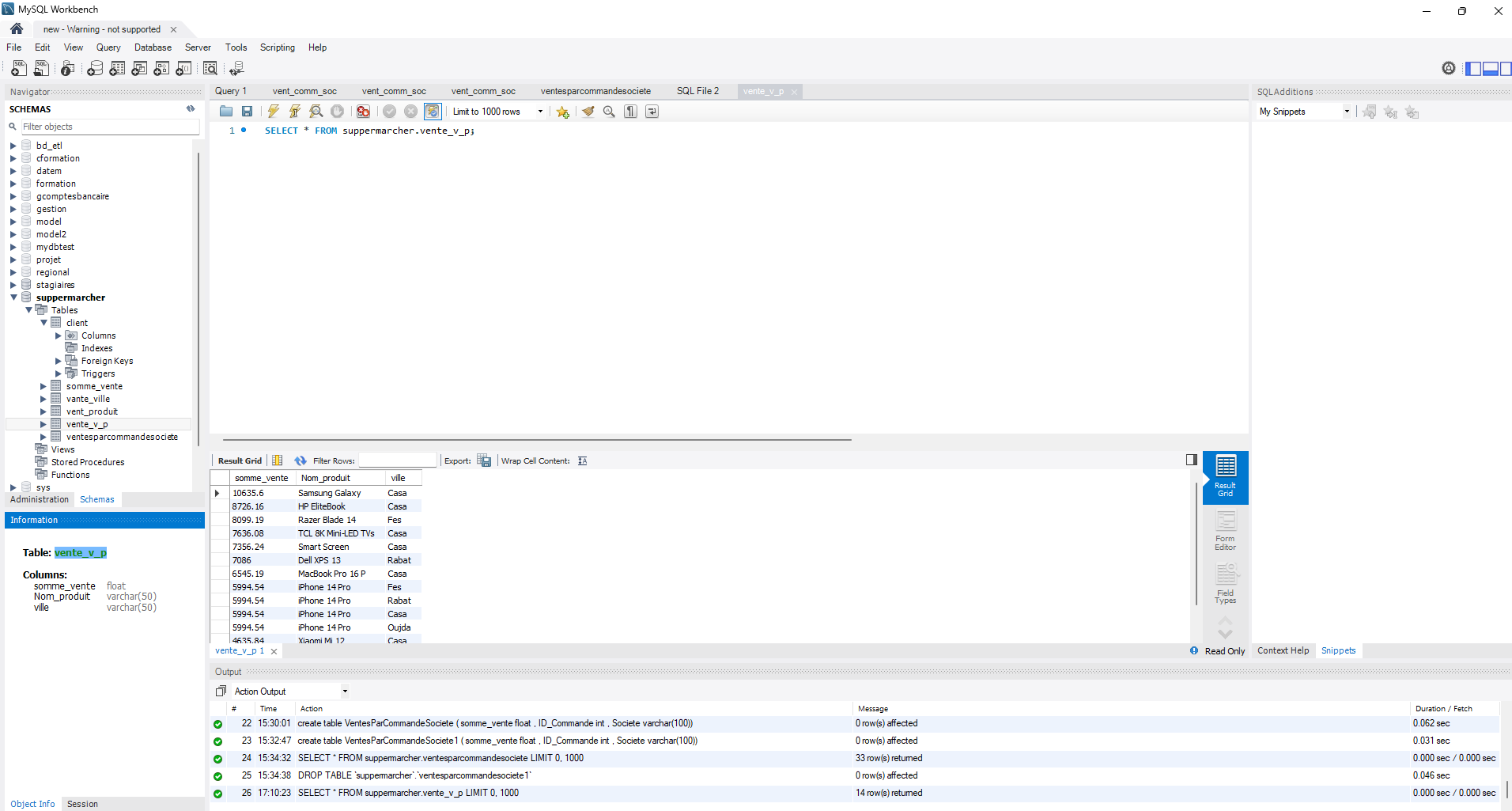
On change le nom du table vers " **vente\_v\_p** "



1. Tester le job



Dans workbench execute la requete “SELECT \* FROM suppermarcher.vente\_v\_p;”



Conclusion:

en fin, ce etude de cas nous a permet de configurer et gerer le processus ETL .

Fin cas d’etude