

Talend Open Studio for Data Integration

Créer une jointure

Familiarisez-vous avec Talend Open Studio for Data Integration.

Ce document présente les principales fonctionnalités de Talend Open Studio for Data Integration.

Dans ce tutoriel, découvrez les fonctionnalités du tMap en créant une jointure entre un fichier d'entrée et la table d'une base de données. Les données de ces deux fichiers sont ensuite transformées vers une autre table de cette même base de données.

Prérequis :

Pour effectuer ce tutoriel, vous devez télécharger et extraire le fichier exampleFile.zip.

Une fois dézippé, vous obtenez deux nouveaux fichiers :

- customers_demo5mn.csv, le fichier d'entrée utilisé pour la jointure.
- demo5mn.sql, le fichier de création de la base de données.

Créez une base de données « demo5mn » et exécutez le script « demo5mn.sql » fourni de manière à créer 2 tables : customers et states.

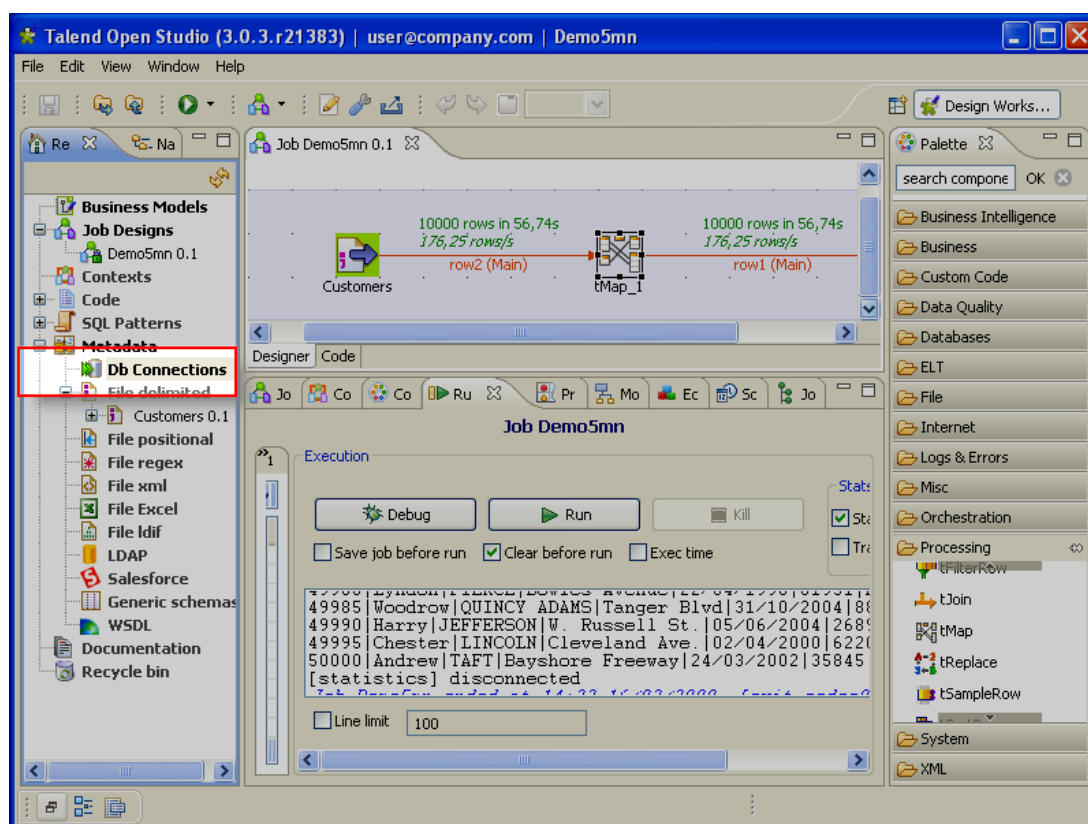


Créer et utiliser une métadonnée Db Connection

Dans le Repository à gauche :

Cliquez-droit sur Connections.

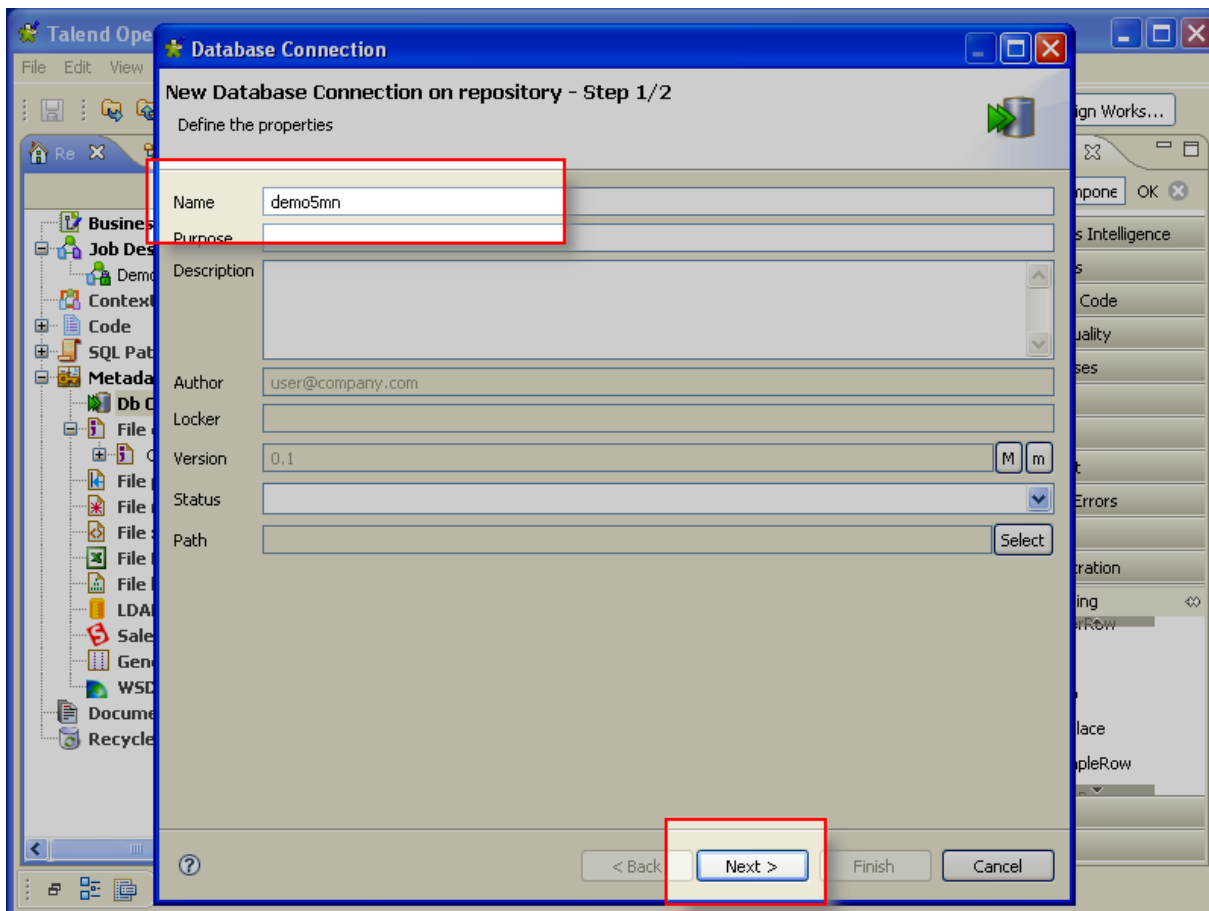
Dans le menu contextuel, cliquez sur Create connection. L'assistant Database connection s'ouvre.



Dans l'assistant Database connection :

Dans le champ Name, saisissez le nom de la connexion :demo5mn.

Cliquez sur Next pour continuer.



Dans la zone Database Settings :

Dans la liste DB Type, sélectionnez le type de base de données à laquelle vous souhaitez vous connecter.

Dans les champs Login et Password, renseignez l'identifiant et le mot de passe de connexion à la base de données.

Dans le champ Server, renseignez le nom ou l'adresse IP du serveur sur lequel est stocké la base de données.

Dans le champ Port, renseignez le numéro du port d'écoute du serveur.

Dans le champ Database, renseignez le nom de la base de données à laquelle vous souhaitez vous connecter.

Cliquez sur Check pour vérifier la connexion à la base de données.

Connexion à la base de données

Nouvelle connexion à la base de données dans le référentiel - Etape 2/2

Vous devez cliquer sur le bouton Vérifier afin de vérifier les paramètres de la connexion

Paramètres de la base de données

DB Type: MySQL

Version de la base de données: MySQL 5

Chaîne de caractères de connexion: jdbc:mysql://localhost:3306/demo5mn?noDatetimeStringSync

Identifiant: root

Mot de passe:

Serveur: localhost

Port: 3306

Base de données: demo5mn

Additional parameters: noDatetimeStringSync=true

Vérifier

Propriétés de la base de données

Syntaxe SQL: SQL 92

Séparateur de chaîne de caractères: "

Caractère Null: 000

Exporter en tant que contexte

Revenir au contexte précédent

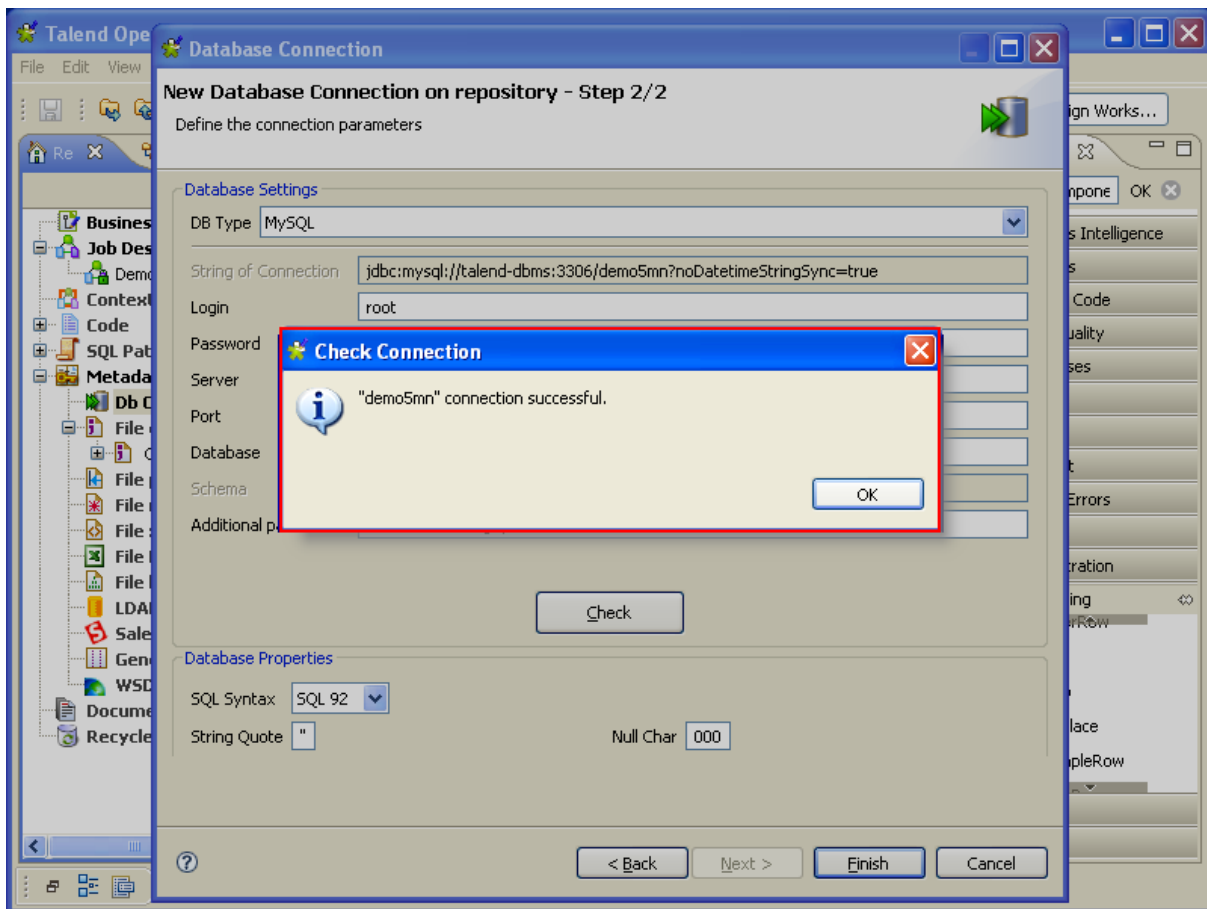
< Retour

Suivant >

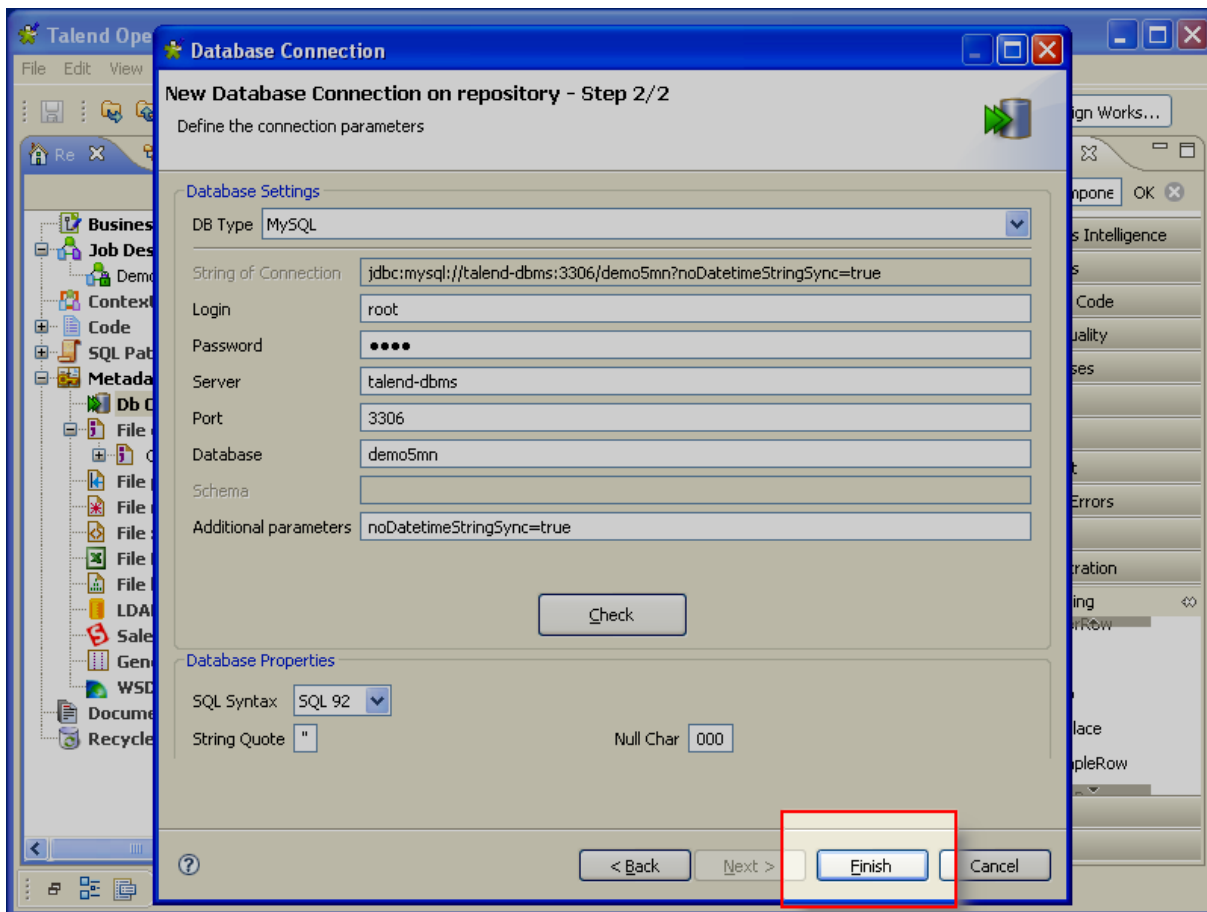
Terminer

Annuler

La boîte de dialogue Check Connection vous indique si la connexion fonctionne.
Cliquez sur OK.



Si la connexion fonctionne, cliquez sur Finish pour fermer l'assistant et valider la nouvelle connexion.
Si la connexion a échoué, modifiez les paramètres de connexion à la base de données et cliquez à nouveau sur le bouton Check.

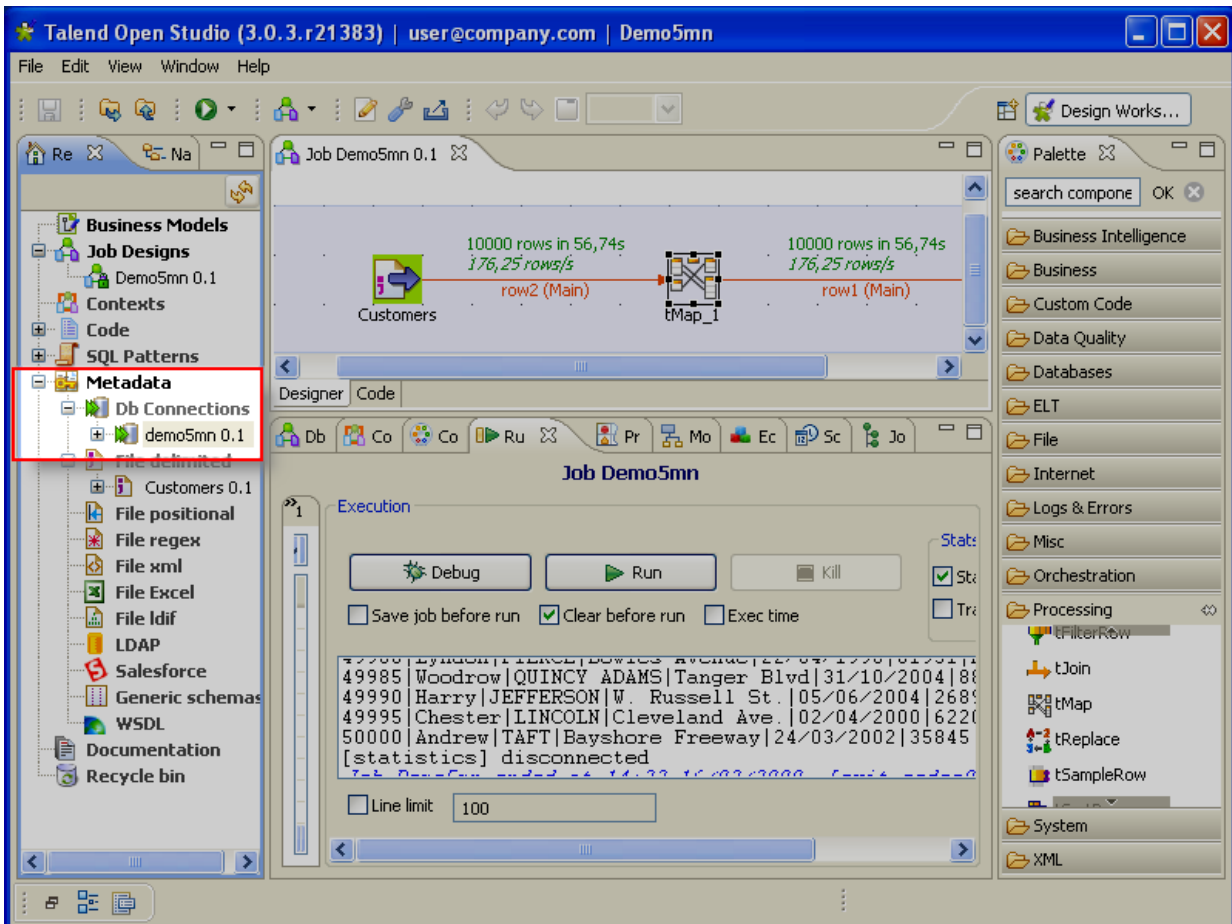


Dans le Repository situé à gauche :

La nouvelle connexion apparaît sous le nœud Metadata> Db Connections.

Cliquez-droit sur la connexion.

Dans le menu contextuel, cliquez sur Retrieve Schema. L'assistant Schema s'ouvre.



Dans l'assistant Schema :

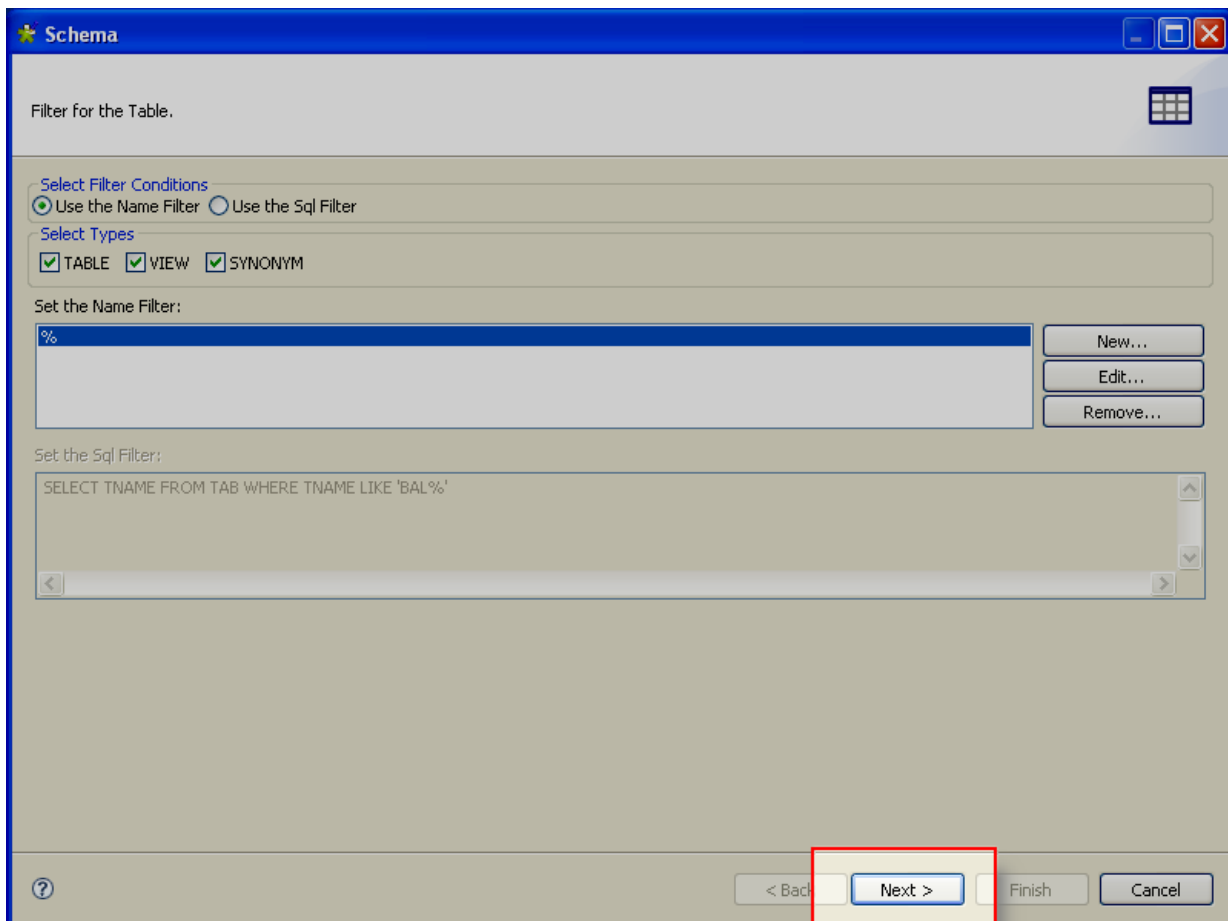
Dans la zone Select Filter Conditions, sélectionnez Use the Name Filter ou Use the Sql Filter en fonction de si vous souhaitez faire une recherche en fonction du nom d'une table ou d'une requête.

Dans la zone Select Types, sélectionnez le type d'éléments recherché : table, view et/ou synonym.

Dans les zones Set the Name Filter ou Set the Sql Filter, configurez les filtres à utiliser.

Ici, utilisez les filtres par défaut.

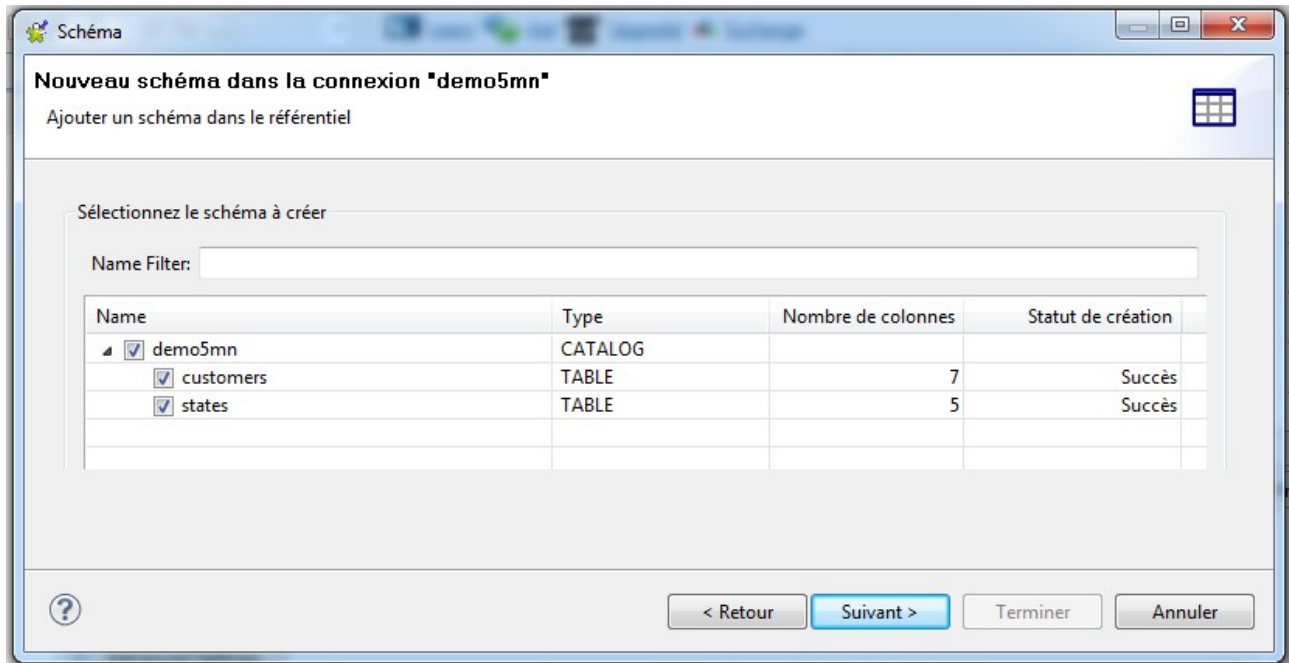
Cliquez sur Next.



Dans la zone Select Schema to create, saisissez le nom de la table que vous souhaitez sélectionner dans le champ Name Filter et cochez sa case dans la listeName, ou cochez simplement la case correspondante dans la liste.

Ici, sélectionnez les deux tables customers et states.

Cliquez sur Next.



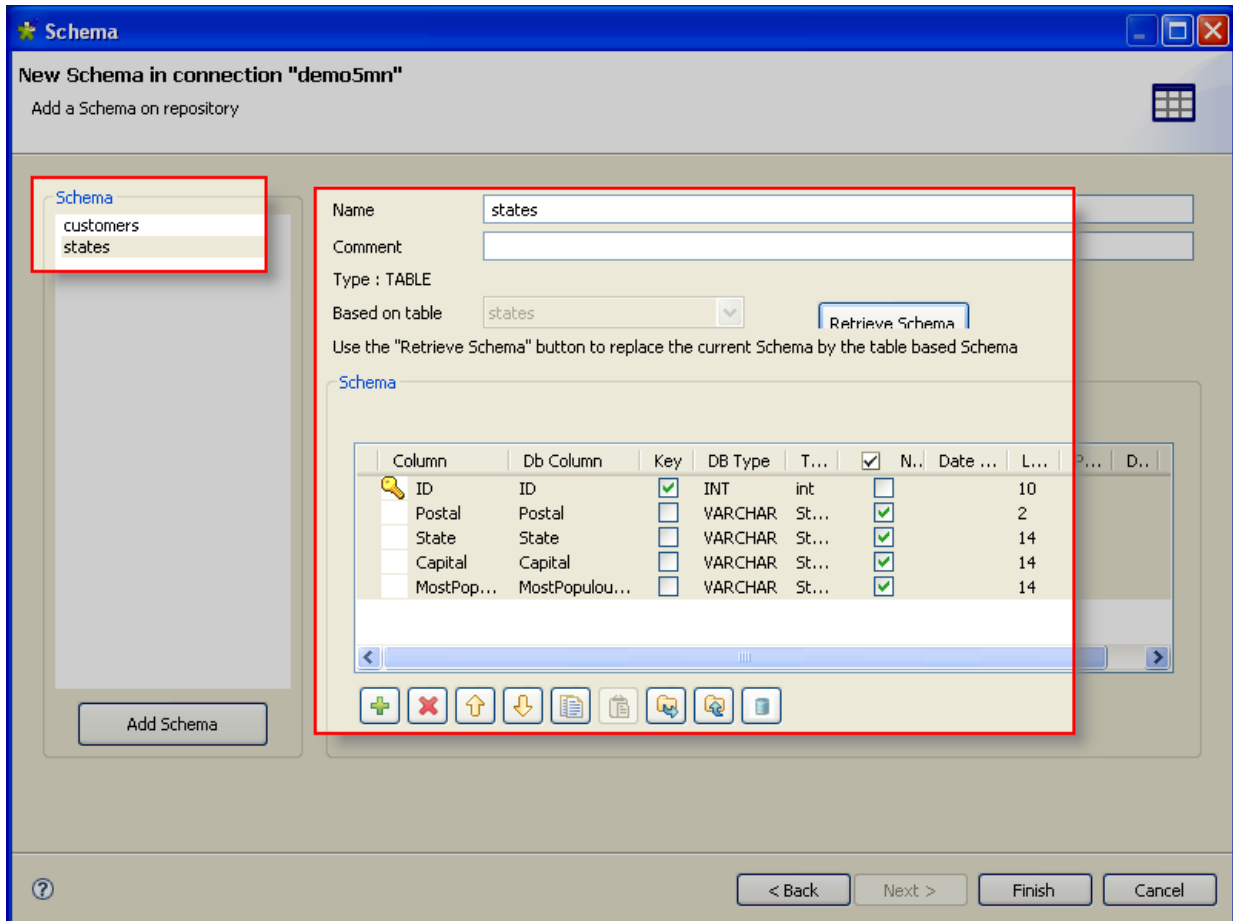
Dans la zone Schema à gauche de l'assistant, sélectionnez la table.

Une fois sélectionnée, son nom apparaîtra dans le champ Name en haut de l'assistant. Vous pouvez le modifier.

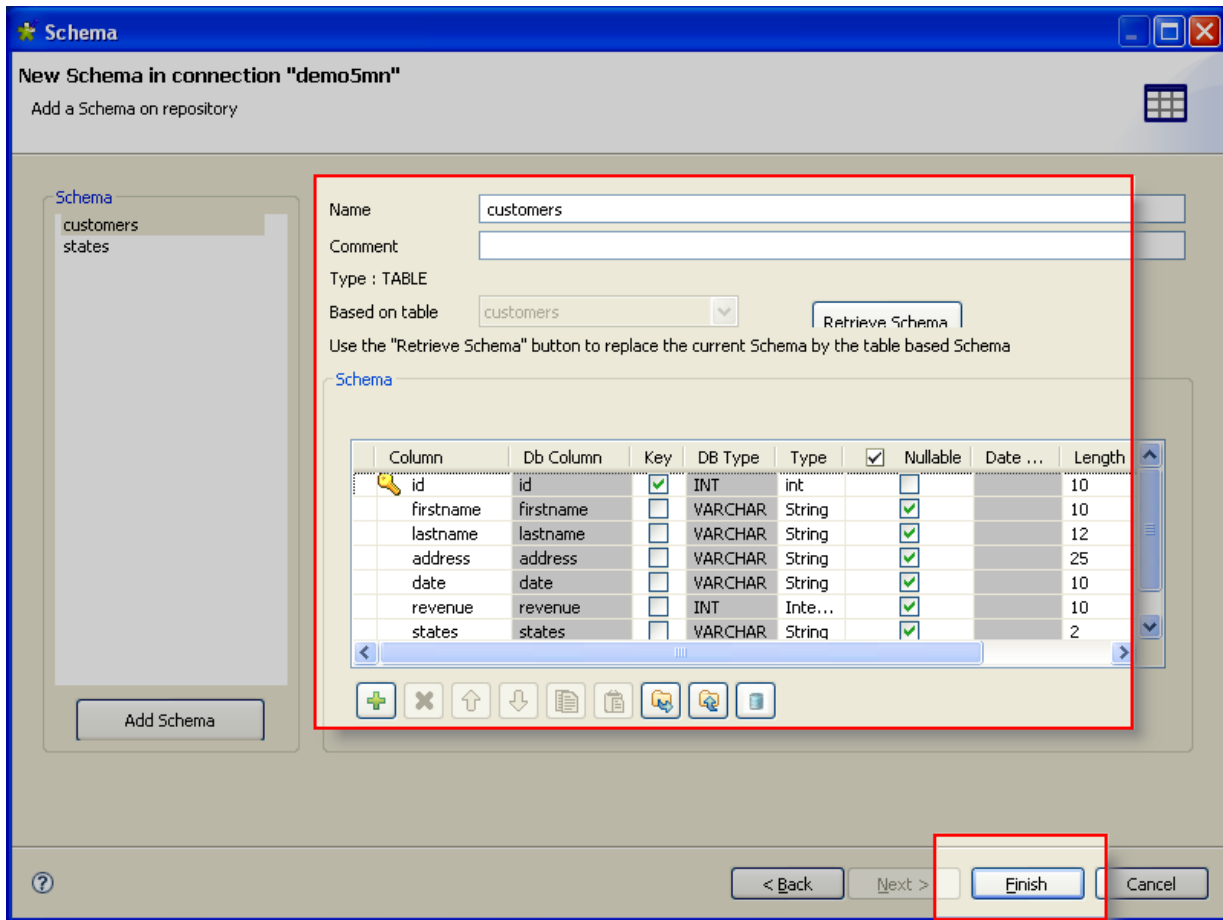
Sa structure apparaît dans la zone Schema en bas de l'assistant. Vous pouvez la modifier à l'aide de la barre d'outils en dessous de la table.

Pour ce tutoriel, ne modifiez pas la structure.

Dans la zone Schema à gauche, cliquez sur la deuxième table.



Le champ Name affiche le nom de la table.
La zone Schema en bas affiche sa structure.
Cliquez sur Finish.

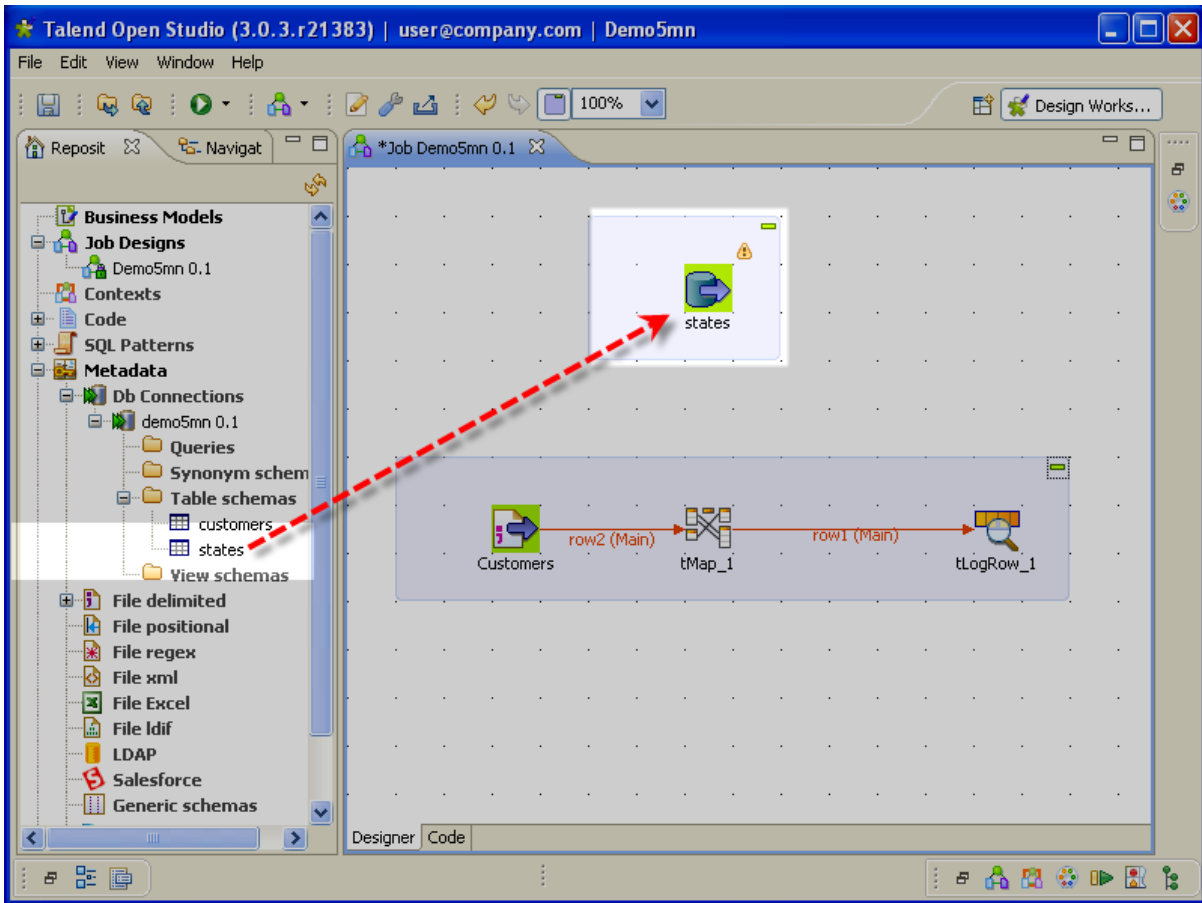


Dans le Repository situé à gauche :

Le nouveau schéma apparaît sous le nœud Metadata > Db Connections > demo5mn > Table schemas.

Pour ajouter le composant d'entrée, appuyez sur Ctrl+Shift tout en glissant-déposant la métadonnées dans le Job Designer au dessus du tMap.

Le composant tMySQLInput est automatiquement créé dans le Job Designer.

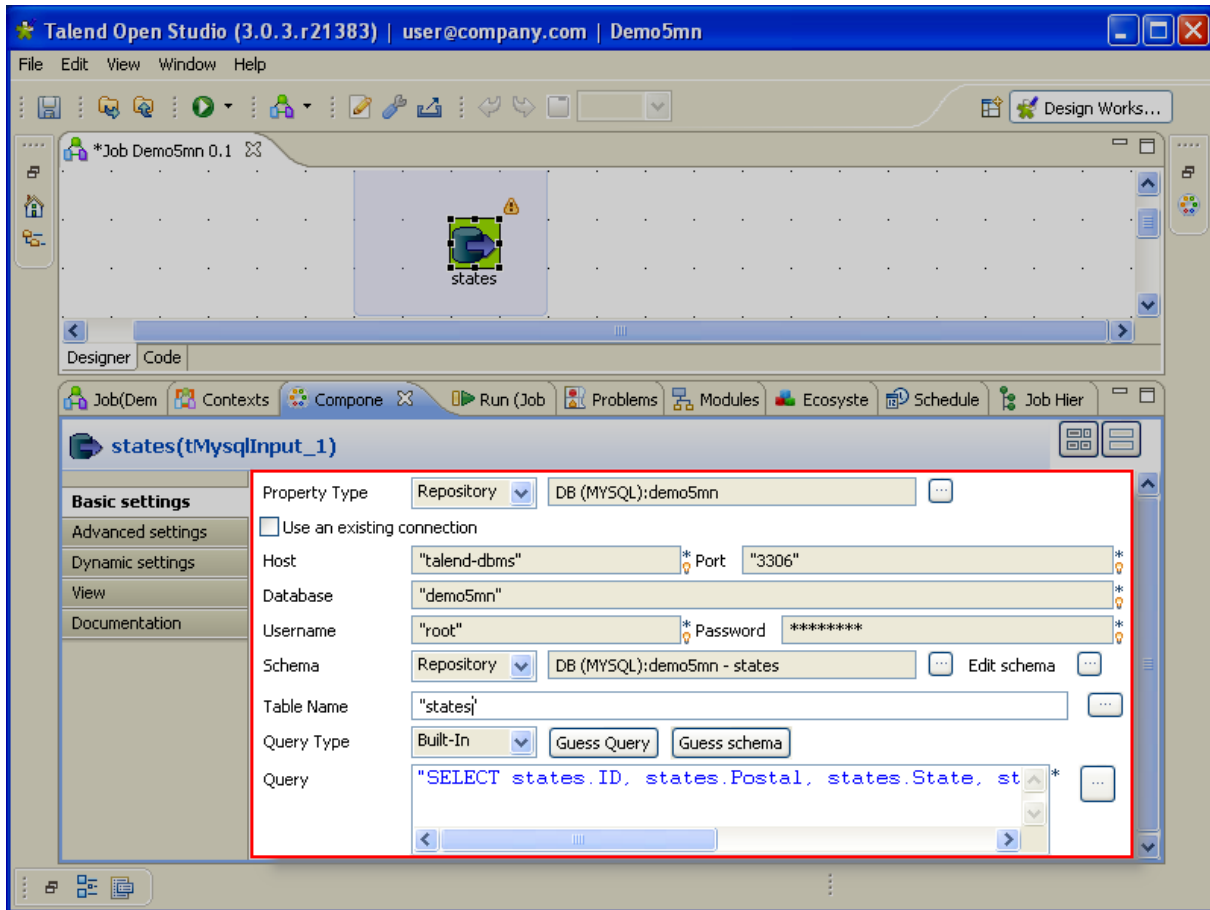


Dans le Job Designer :

Pour paramétrer les propriétés du tMySQLInput, double-cliquez sur le composant. La vue Componentcorrespondante apparaît alors en bas de l'écran.

Dans la vue Component :

Cliquez sur [...] à côté du champ Query. L'assistant SQL Builder s'ouvre.



Dans l'assistant SQL Builder :

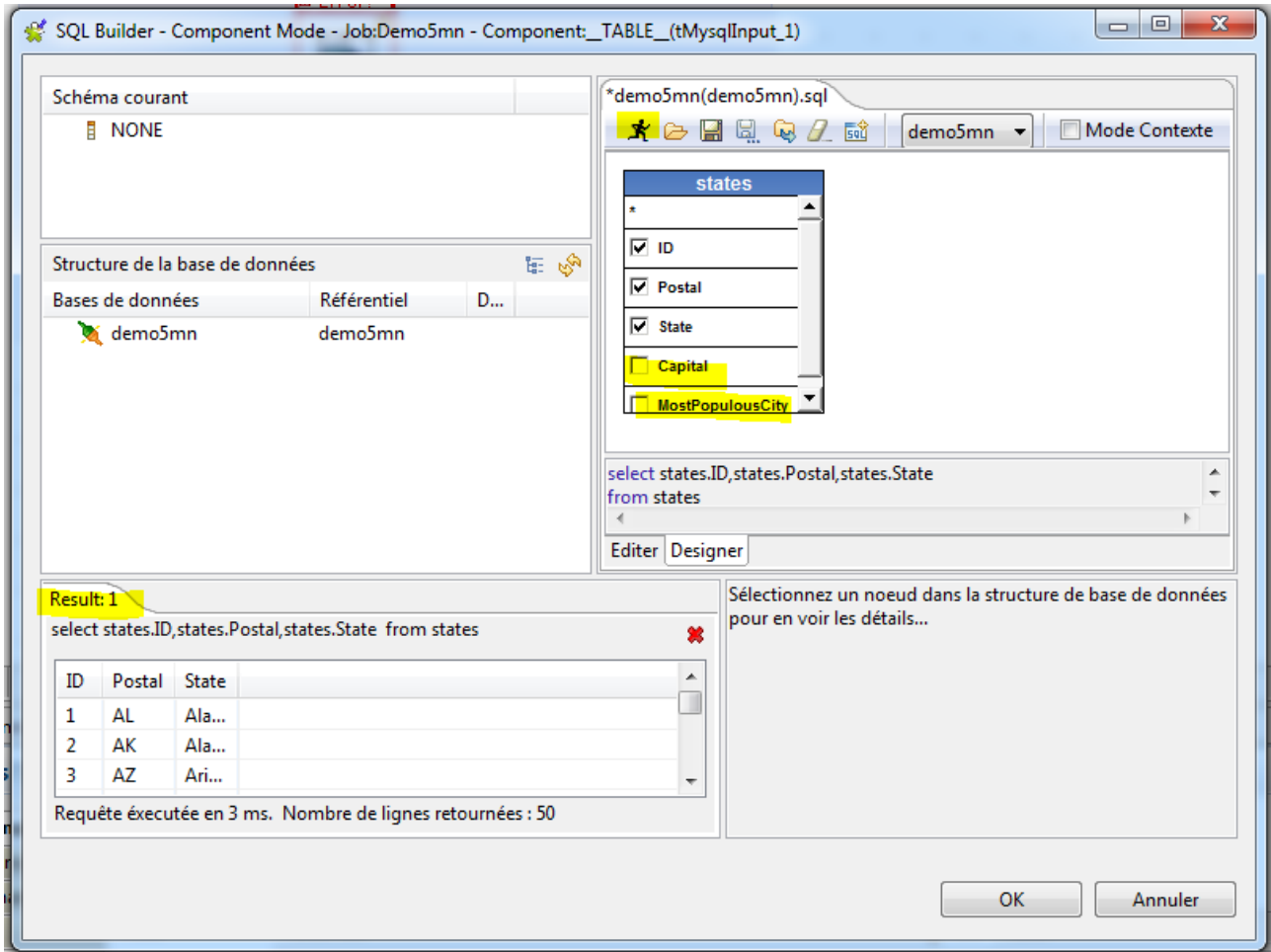
Cliquez sur l'onglet Designer en bas de la vue demo5mn(demo5mn).sql.

Décochez les cases Capital et MostPopulousCity.

Cliquez sur l'icône de l'homme qui court.

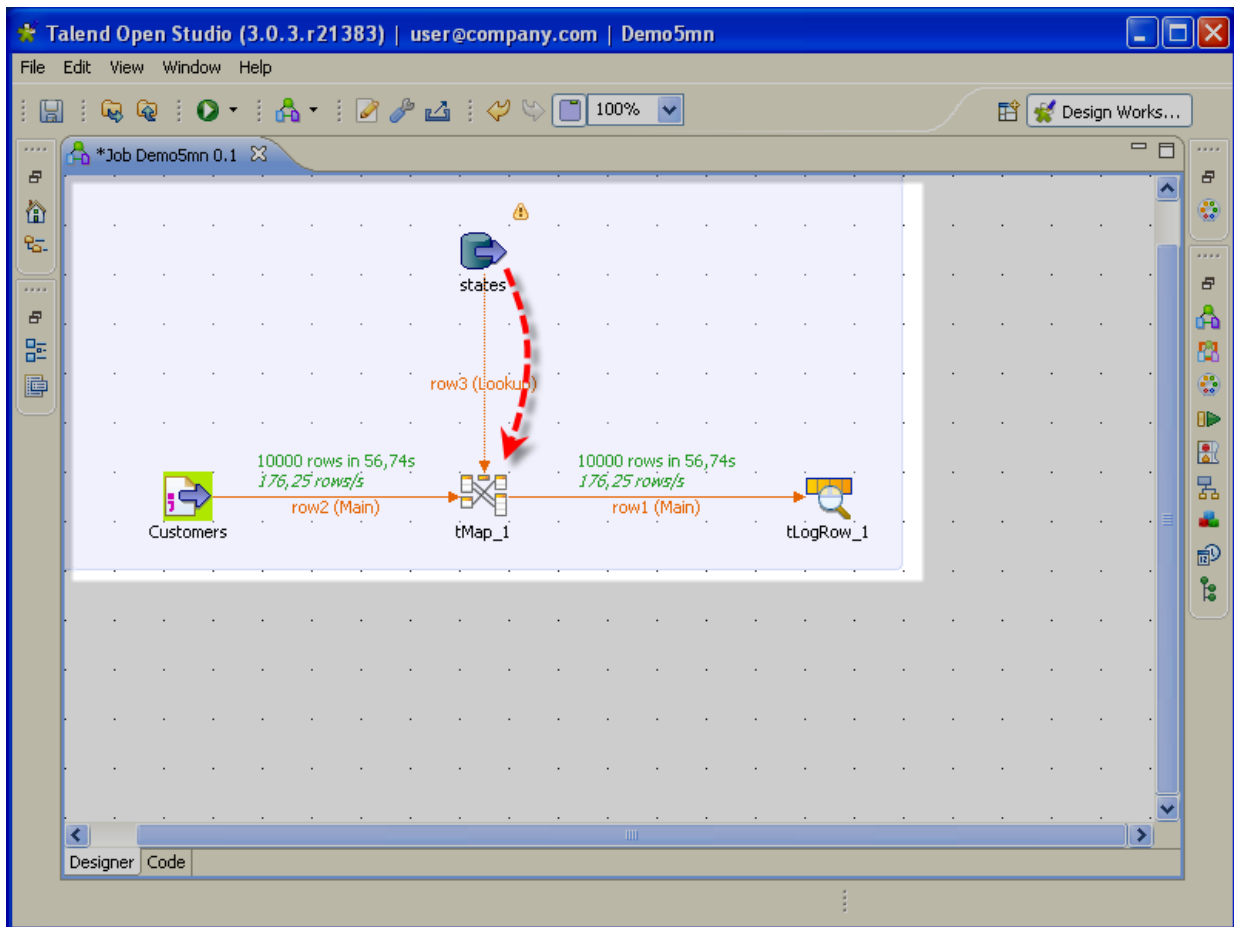
Le résultat de la requête apparaît dans la vue Result: 1 en bas de l'assistant.

Cliquez sur OK.



Dans le Job Designer :

Pour relier les composants entre eux, cliquez-droit sur le tMySQLInput, et déplacez-vous jusqu'au tMap.

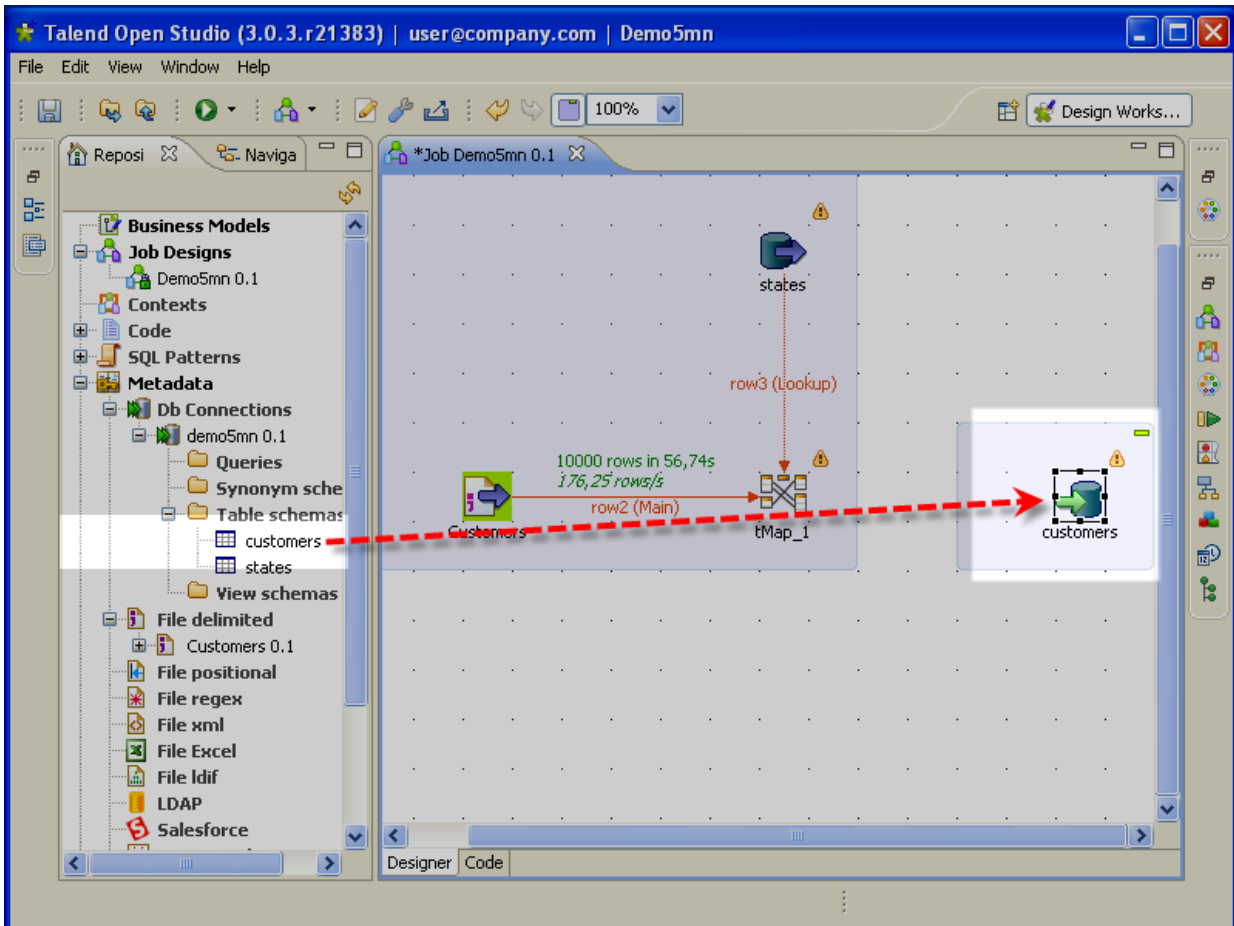


Dans le Job Designer :

Cliquez-droit sur le tLogRow. Dans le menu contextuel, cliquez sur Delete.

Dans le Repository à gauche :

Pour ajouter le composant de sortie, appuyez sur Ctrl tout en glissant-déposant la métadonnée customers dans le Job Designer, à la place du tLogRow.

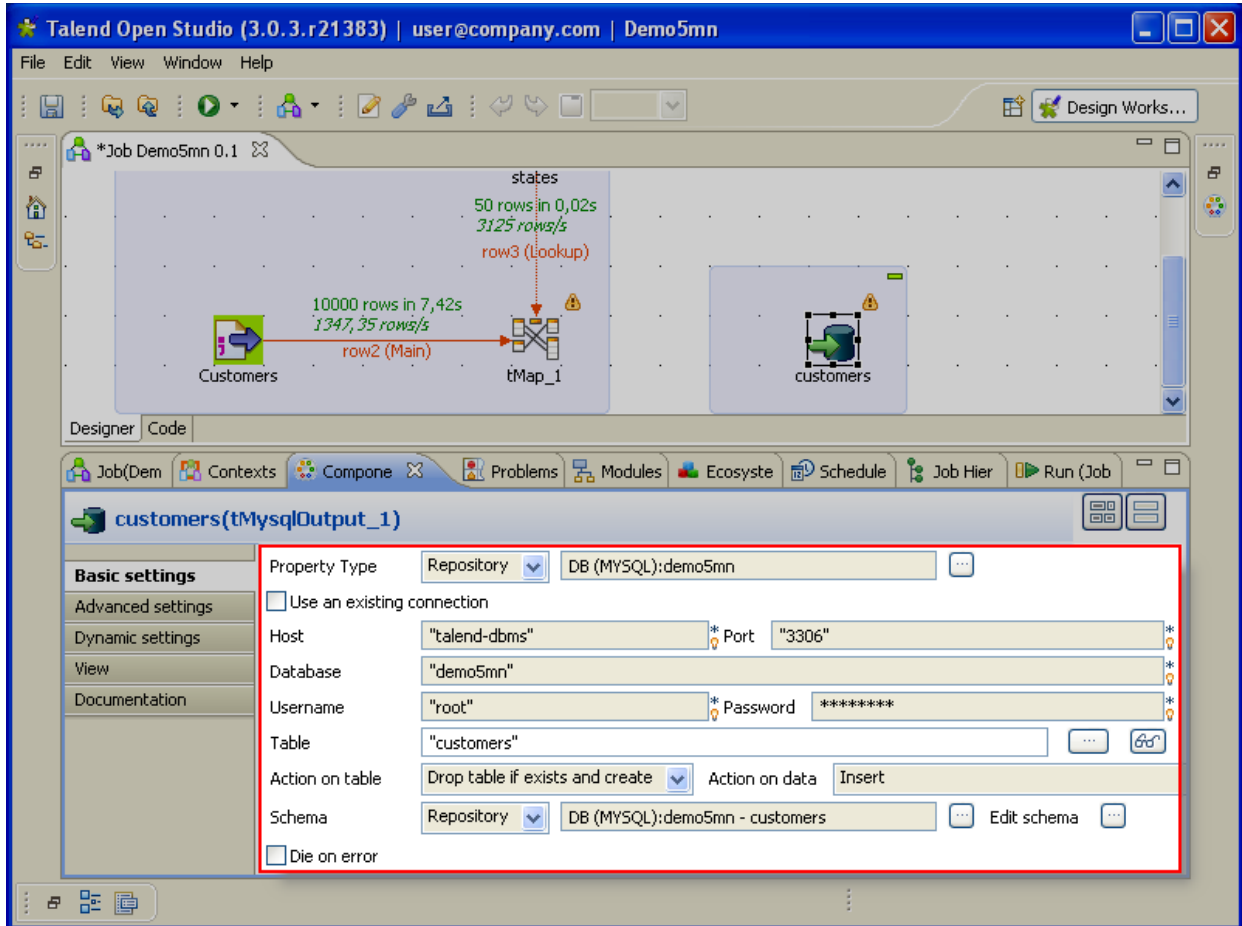


Dans le Job Designer :

Pour paramétrer les propriétés du tMySQLOutput, double-cliquez sur le composant. La vue Component correspondante apparaît alors en bas de l'écran.

Dans la vue Component :

Dans la liste Action on table, cliquez sur Drop table if exists and create. pour remplacer l'ancienne version de la table par la nouvelle.



Dans le Job Designer :

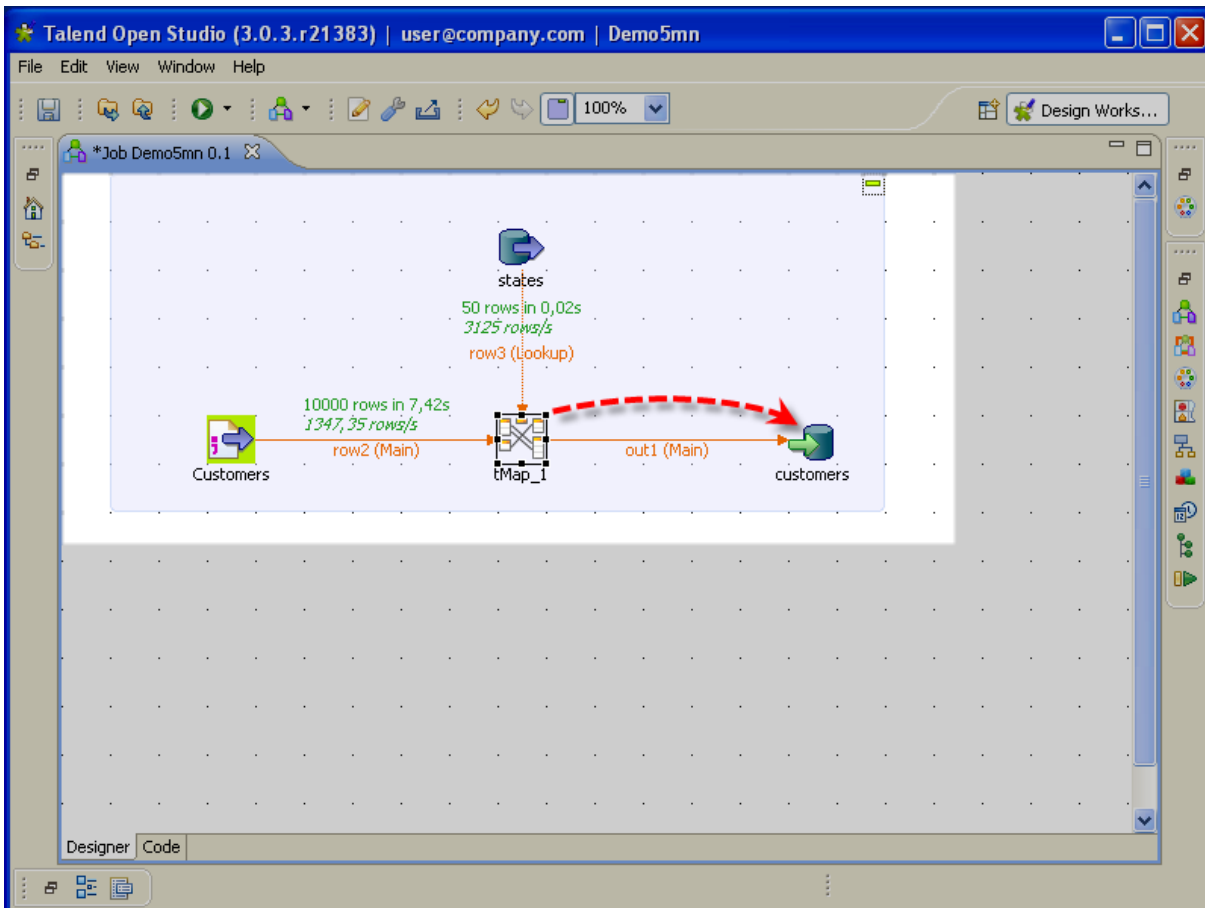
Pour relier les composants entre eux, cliquez-droit sur le tMap, et déplacez-vous jusqu'au tMysqlOutput.

Dans l'assistant tMap_1 Output, nommez le lien reliant le tMap au tMysqlOutput : out1.

Répondre « Oui » à la question « Récupérer le schéma du composant cible ».



Double-cliquez sur le tMap. L'éditeur du tMap s'ouvre.



2

Paramétrer le composant de transformation

Dans l'éditeur du tMap :

Cliquez sur la table row1 de droite et cliquez sur [x] pour la supprimer.

Dans la table row2, sélectionnez toutes les colonnes sauf la colonne states et glissez-les dans la table out1.

Puis sélectionnez la colonne states et glissez-la dans la colonne Postal de la table row3 pour créer la jointure.

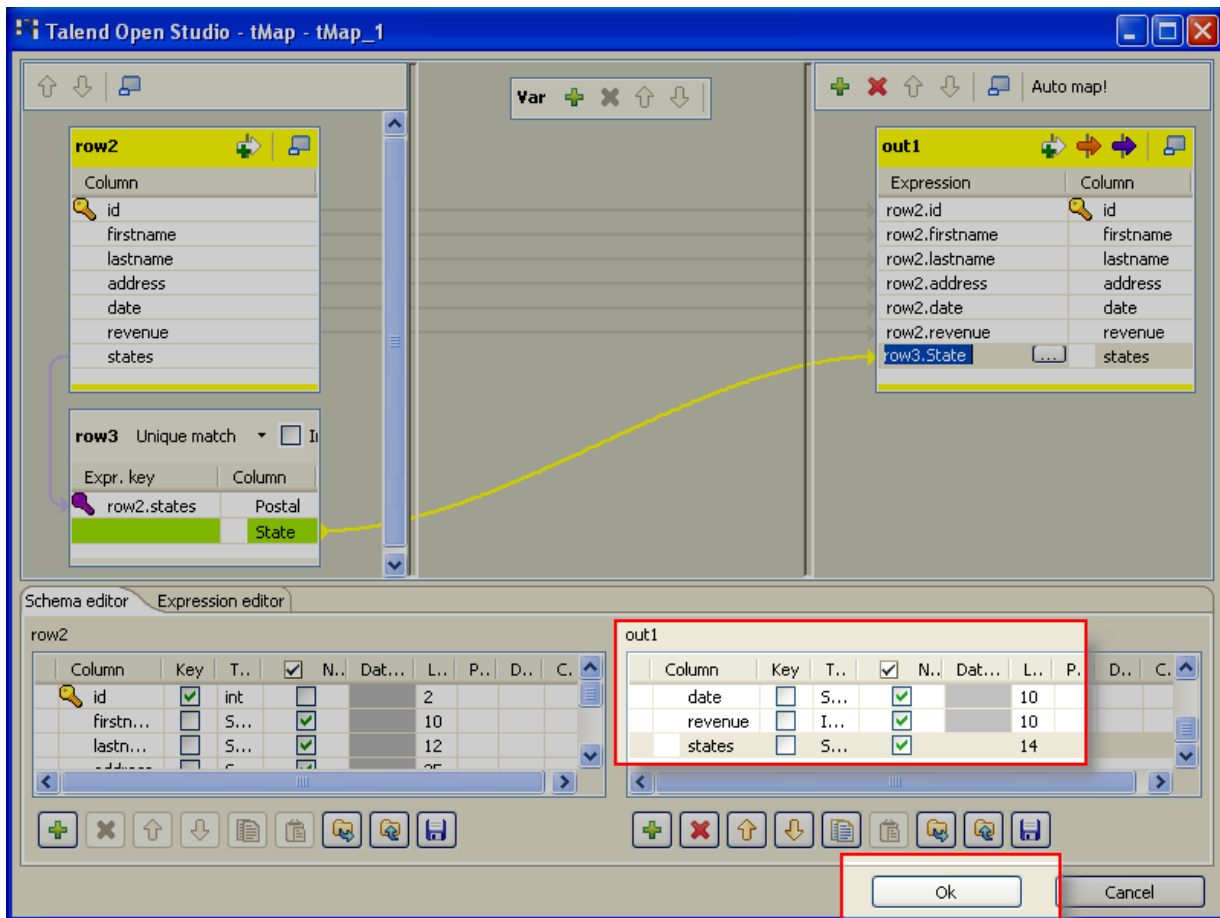
Dans la table row3, sélectionnez la colonne State et glissez-la à la fin de la table out1.

The screenshot shows the Talend Open Studio tMap editor. The main workspace is divided into three panes. The left pane shows the 'row2' table with columns: id, firstname, lastname, address, date, revenue, and states. The right pane shows the 'out1' table with columns: id, firstname, lastname, address, date, revenue, and states. The bottom pane shows the 'row3' table with columns: Expr. key, Column, and State. A yellow arrow points from the 'row2.states' column to the 'Postal' column in 'row3'. Another yellow arrow points from the 'row3.State' column to the 'states' column in 'out1'. The bottom pane also shows the 'Schema editor' and 'Expression editor' tabs, with the 'Schema editor' displaying the column definitions for 'row2' and 'out1'.

Column	Key	T..	✓	N..	Dat...	L..	P..	D..	C..
id	✓	int	✓			2			
firstn...		S...	✓			10			
lastn...		S...	✓			12			
add...		S...	✓			25			

Column	Key	T..	✓	N..	Dat...	L..	P..	D..	C..
date		S...	✓			10			
revenue		I...	✓			10			
states		S...	✓			14			

Dans la table out1, sélectionnez la colonne row3.State. Dans la table out1 de la zone Schema editor, paramétrez sa longueur à 14 caractères. Cliquez sur OK.



3

Exécuter le Job

Dans le Job Designer :

Avant d'exécuter votre Job, enregistrez-le via Ctrl+S.

Cliquez sur la vue Run.

Cochez les cases Statistics .

Cliquez sur « Debug des Traces » dans « Exécution Debug ».

Dans le Job Designer :

Les tables Traces affichent les données traitées.

Dans la vue Run :

Cliquez sur Kill pour arrêter l'exécution du Job.

The screenshot displays the Talend Open Studio interface during a job execution. The main workspace shows a job design with a 'Customers' component connected to a 'tMap_1' component, which is then connected to an 'out1 (Main)' component. The 'Customers' component is processing 24 rows at a rate of 1.03 rows/s. The 'out1 (Main)' component is also processing 24 rows at the same rate. Below the job design, two data tables are visible, both with a current row of 24. The first table, 'row2 (Main)', shows data for a customer with id 24, first name William, last name Coolidge, address W. Russell St., date 13/07/1998, revenue 57326, and states RI. The second table, 'out1 (Main)', shows the same data but with the states field set to Rhode Island. A red box highlights the 'revenue' and 'states' fields in both tables. At the bottom of the interface, the 'Run (Job Demo5mn)' window is open, showing execution controls (Debug, Pause, Kill) and a console window with the following output:

```
Starting job Demo5mn at 11:12 23/02/2009.  
[statistics] connecting to socket on port 3639  
[statistics] connected  
[trace] connecting to socket on port 5160
```

Dans la vue Run :

Cliquez à nouveau sur Run en « Execution simple ».

The screenshot displays the Talend Open Studio interface. The top window shows the job design in the 'Designer' view, featuring a 'tMap_1' component with two input rows: 'row2 (Main)' and 'row3 (Lookup)'. The 'row2 (Main)' input is connected to a 'Customers' component, and the 'row3 (Lookup)' input is connected to a 'states' component. The output of 'tMap_1' is connected to an 'out1 (Main)' component, which is in turn connected to another 'customers' component. The 'row2 (Main)' input shows a throughput of '10000 rows in 7,01s' at '1426,74 rows/s'. The 'out1 (Main)' output shows a throughput of '10000 rows in 7,01s' at '1426,74 rows/s'. The 'row3 (Lookup)' input shows '50 rows in 0s' at '0 rows/s'. The bottom window shows the 'Run (Job Demo5mn)' view, which is highlighted with a red border. This view includes a 'Execution' section with buttons for 'Debug', 'Run', and 'Kill', and checkboxes for 'Save job before run', 'Clear before run', and 'Exec time'. The 'Stats & Traces' section has checkboxes for 'Statistics' (checked) and 'Traces', along with a 'Clear' button. The execution log shows the following output:

```
[statistics] connecting to socket on port 4224
[statistics] connected
[statistics] disconnected
Job Demo5mn ended at 11:13 23/02/2009. [exit code=0]
```

Le Job est maintenant terminé !

Il crée une jointure entre deux fichiers d'entrée et alimente une base de données avec les données agrégées.

Nous souhaitons maintenant exporter le Job.