

Talend Open Studio for Data Integration

Présentation de Talend Open Studio for Data Integration

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.



Ouvrir Talend Open Studio

Double-cliquez sur le raccourci Talend Open Studio pour ouvrir son assistant de connexion.

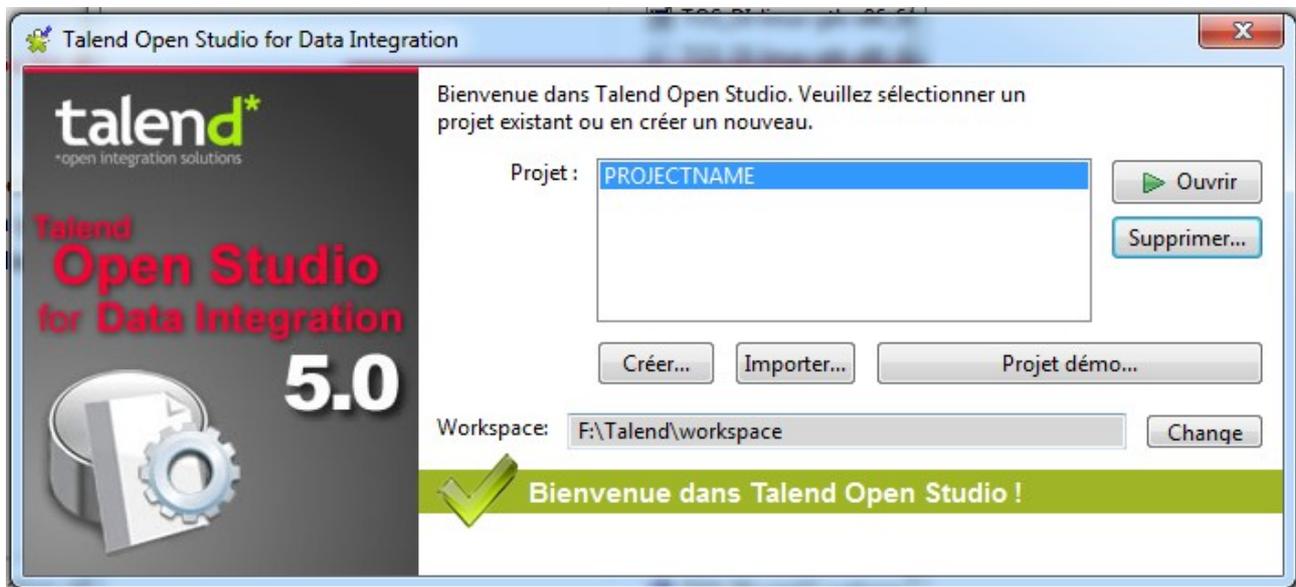


2

Créer et ouvrir un projet

Dans l'assistant Talend Open Studio :

Cliquez sur Créer pour ouvrir l'assistant de création.



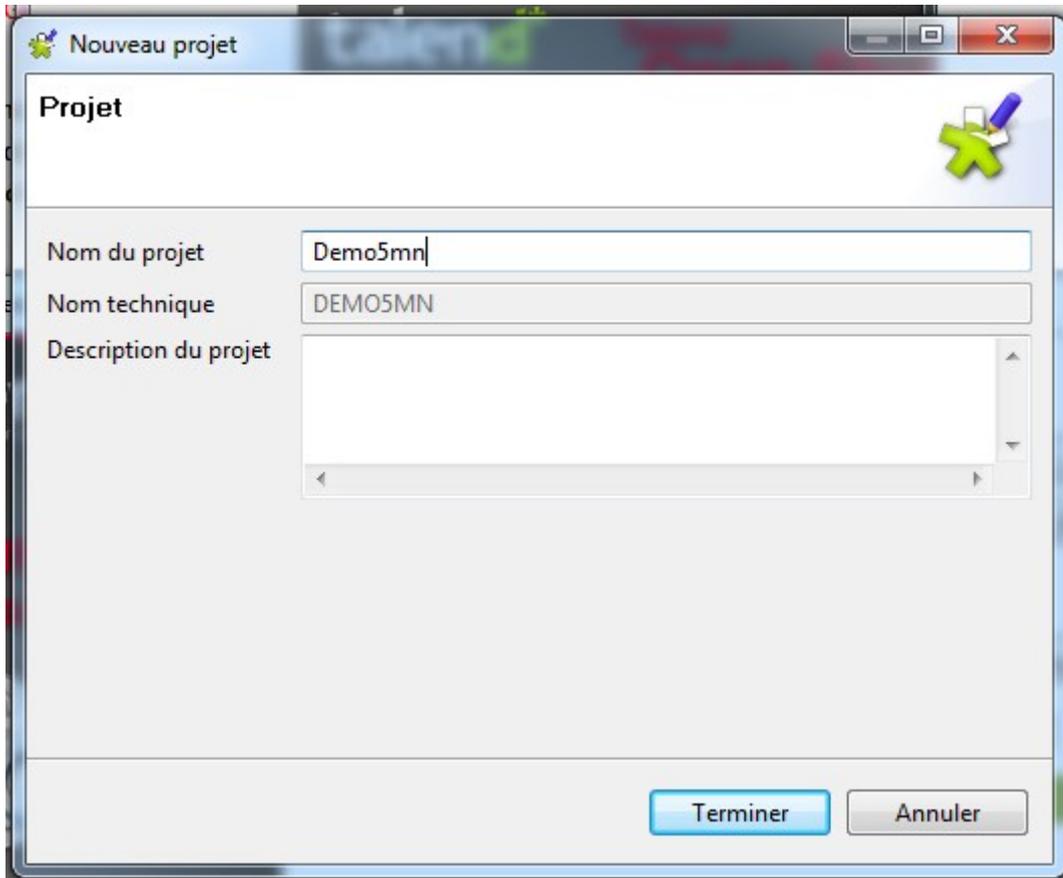
Dans l'assistant Nouveau projet :

Dans le champ Nom du projet, saisissez le nom du projet : Demo5mn.

Cliquez sur Terminer.



Le champ **Nom** ne doit pas contenir d'accents, de caractères spéciaux, d'espace



Nouveau projet

Projet

Nom du projet Demo5mn

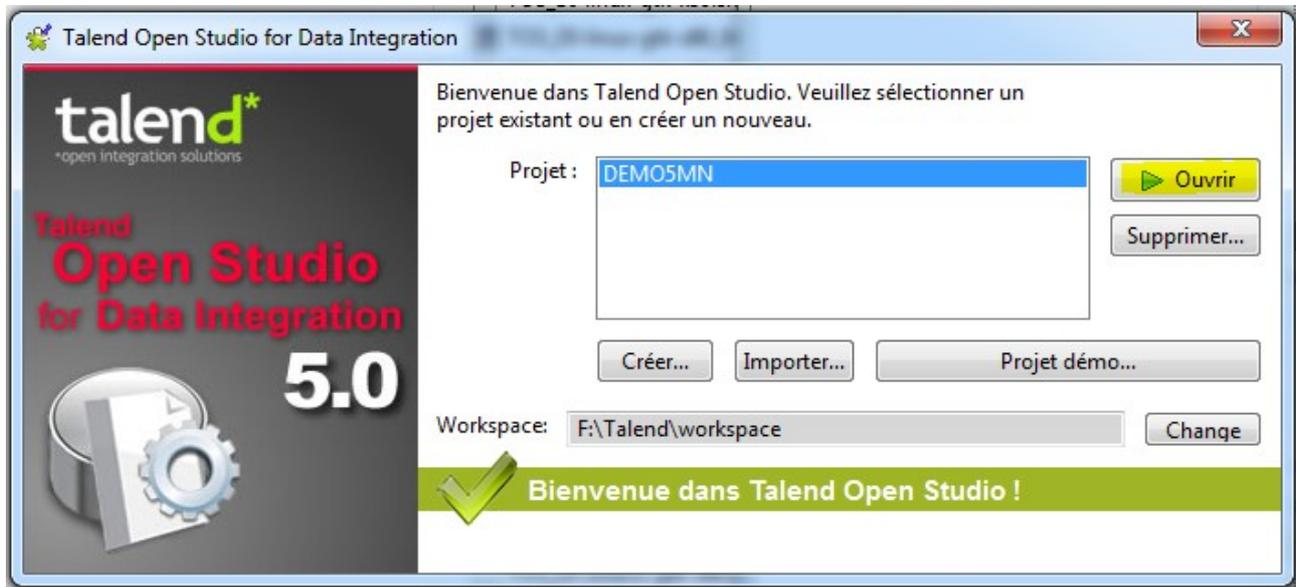
Nom technique DEMO5MN

Description du projet

Terminer Annuler

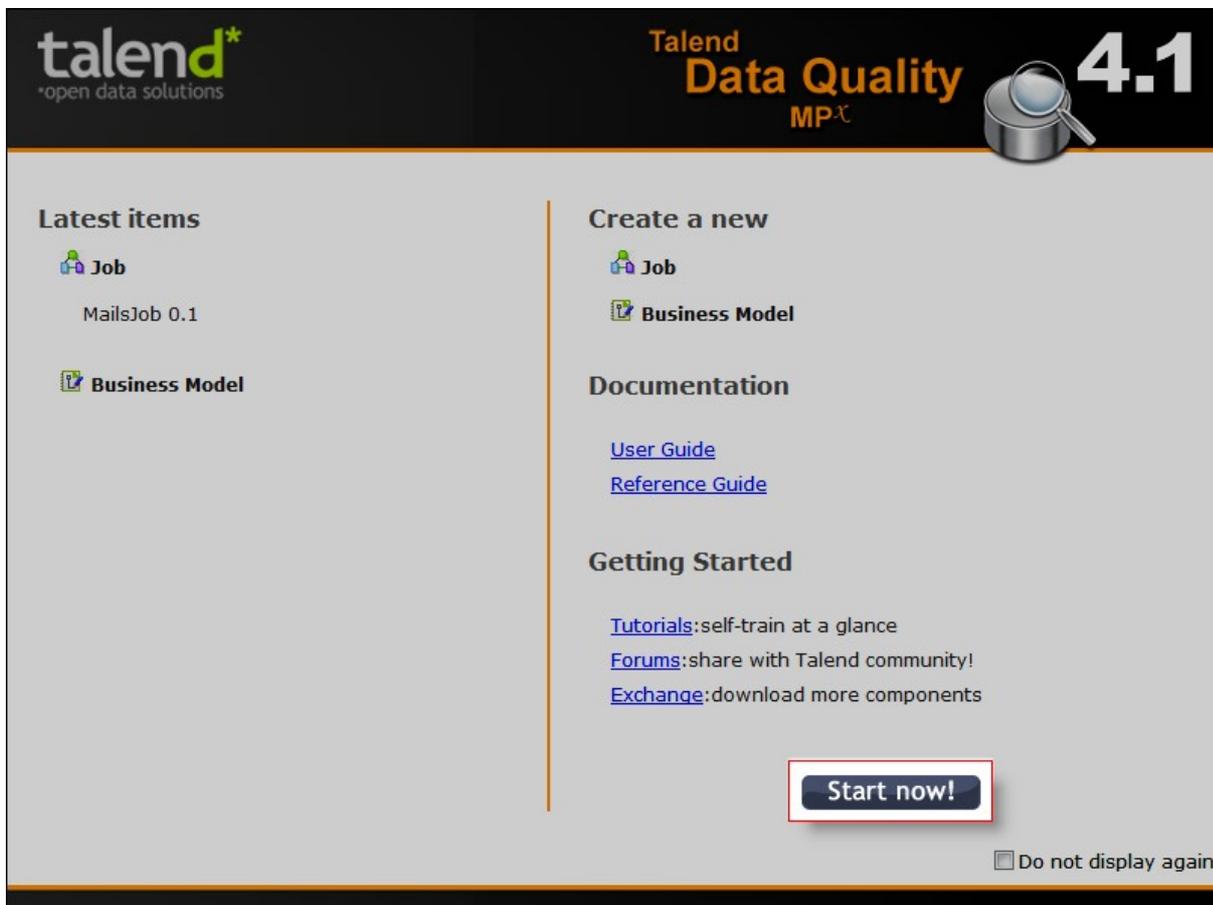
Dans l'assistant Talend Open Studio :

Dans la zone Project, sélectionnez le projet Demo5mn que vous venez de créer dans la liste et cliquez sur Ouvrir.



Lorsque vous ouvrez Talend Open Studio pour la première fois, un onglet affiche des informations utiles, notamment des liens vers le Guide d'installation (Installation guide), le Guide Utilisateur (User Guide) et la Communauté (Community).

Cliquez sur Start now! pour fermer cet onglet et accéder à Talend Open Studio.



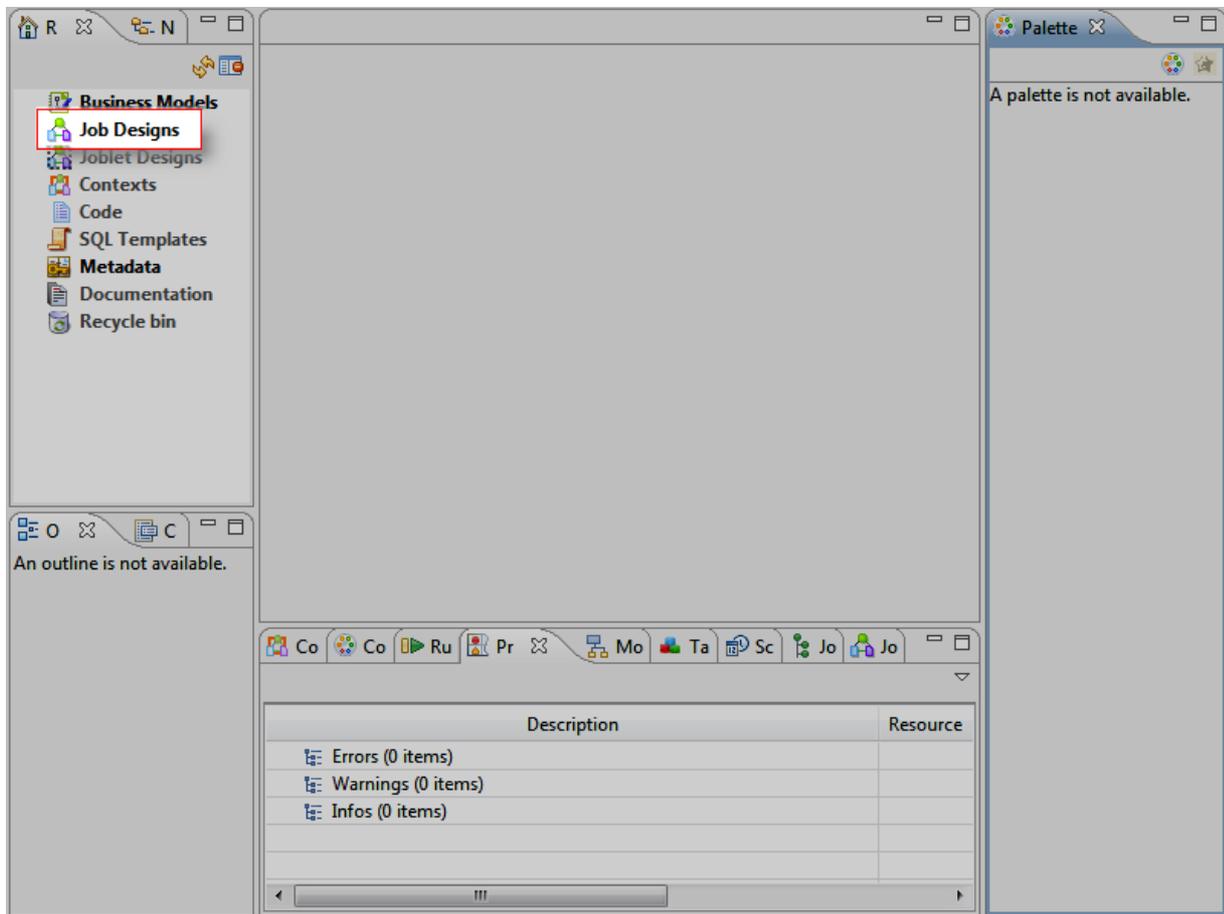
3

Créer un Job Design

Dans le Repository situé à gauche de Talend Open Studio :

Cliquez-droit sur Job Designs.

Dans le menu contextuel, cliquez sur Create Job. L'assistant New job s'ouvre.



Dans l'assistant New job :

Dans le champ Name, saisissez le nom du Job :Demo5mn.

Cliquez sur Finish pour fermer l'assistant et créer le Job.

Le Job Designer présente alors un Job vierge.



Le champ **Name** ne doit pas contenir d'accents, de caractères spéciaux, d'espaces, ni débiter par un chiffre.

The screenshot shows the 'New job' dialog box in the Talend Studio interface. The dialog is titled 'New job' and contains the following fields:

- Name: Demo5mn
- Purpose: (empty)
- Description: (empty text area)
- Author: svichot@talend.com
- Locker: (empty)
- Version: 0.1
- Status: (empty)
- Path: (empty)

The 'Finish' button is highlighted with a red box, indicating the next step in the process.

4

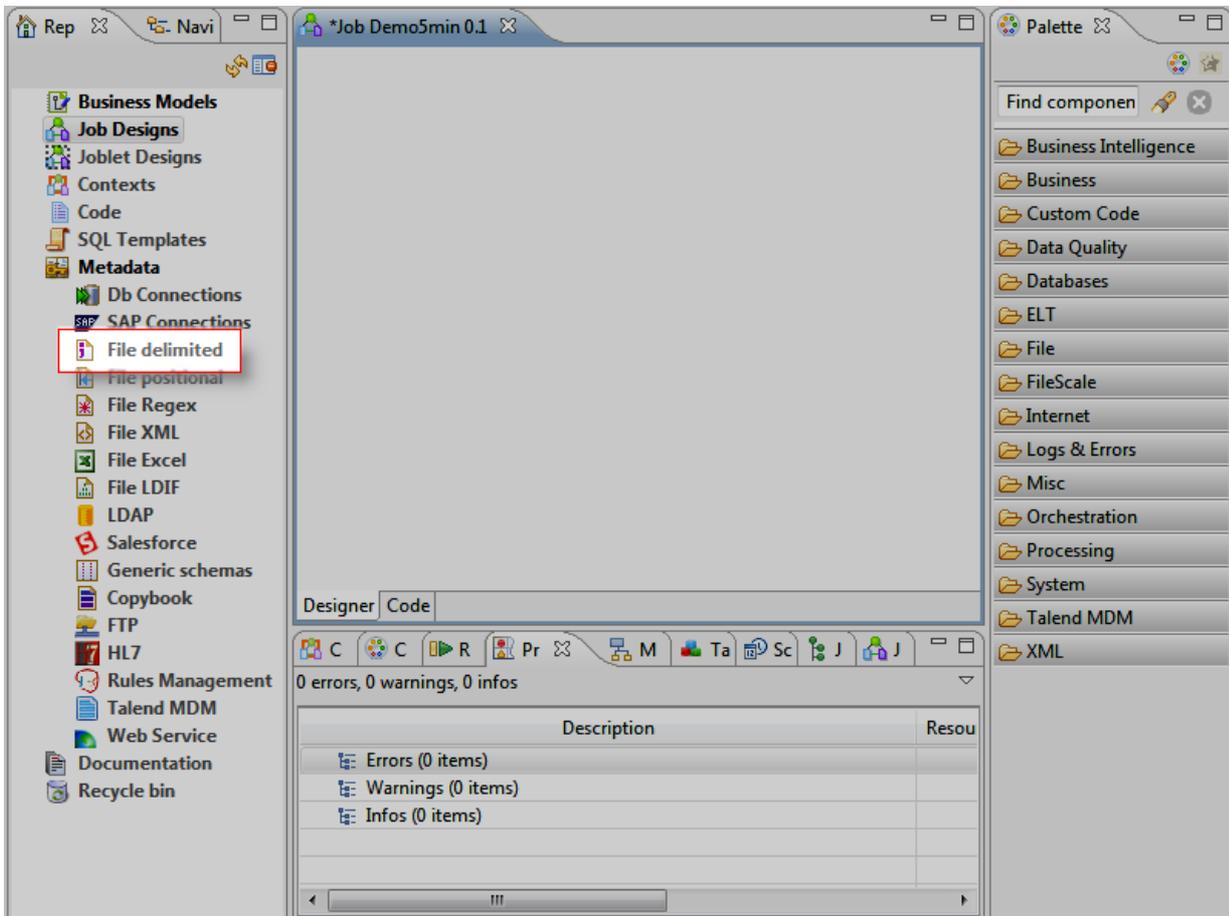
Créer la métadonnée

Dans le Repository situé à gauche de Talend Open Studio :

Développez le nœud Metadata.

Cliquez-droit sur File delimited.

Dans le menu contextuel, cliquez sur Create file delimited. L'assistant New Delimited File s'ouvre.



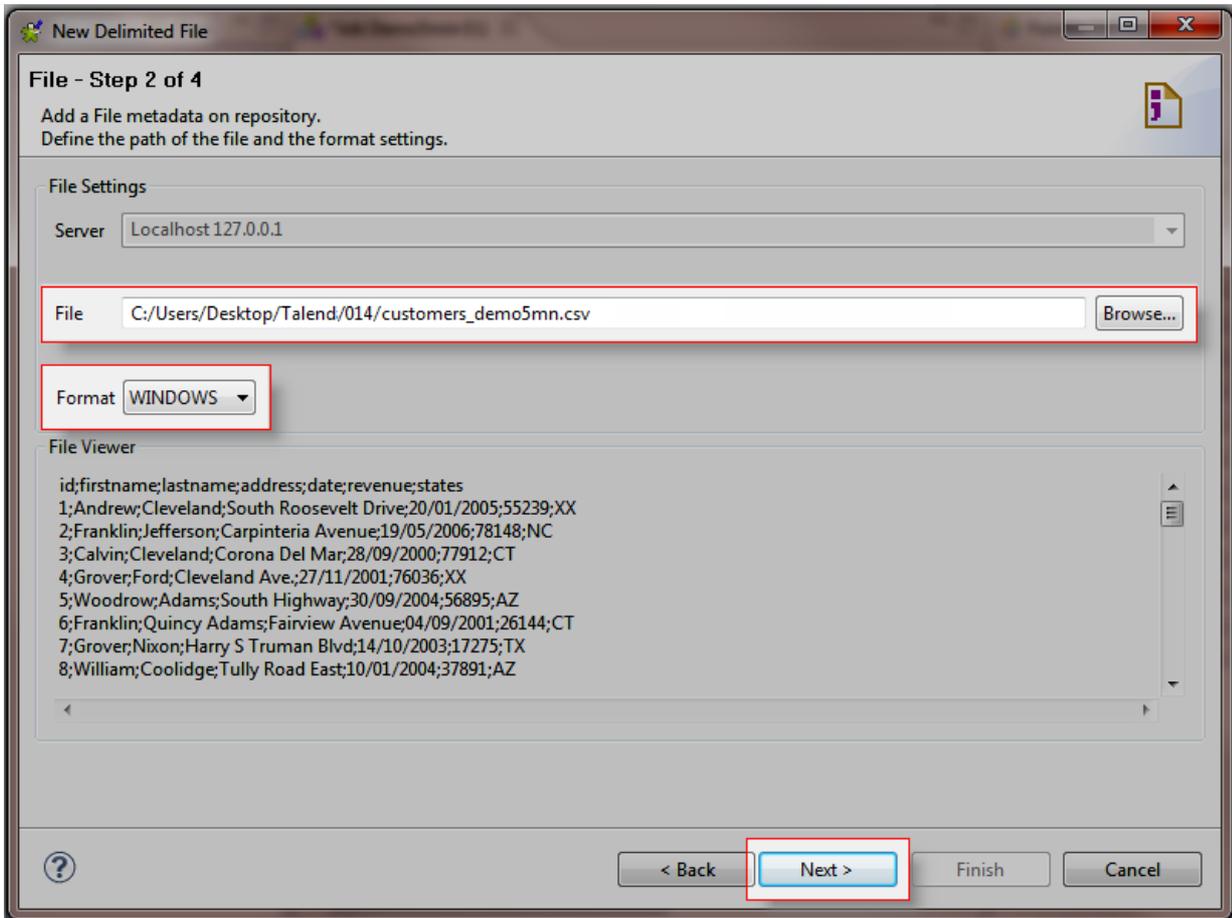
Dans l'assistant New Delimited File :

Dans le champ Name, saisissez le nom de la métadonnée : Customers.

Cliquez sur Next pour continuer.

The screenshot shows a window titled "New Delimited File" with a subtitle "File - Step 1 of 4". Below the subtitle is the instruction "Add a File metadata on repository. Define the properties." The form contains several fields: "Name" (containing "Customers"), "Purpose", "Description", "Author" (containing "svichot@talend.com"), "Locker", "Version" (containing "0.1"), "Status", and "Path". At the bottom, there are four buttons: "< Back", "Next >", "Finish", and "Cancel". The "Next >" button is highlighted with a red box.

Cliquez sur Browse... et sélectionnez customers_demo5mn.csv dans l'assistant qui s'ouvre.
Dans la liste Format, sélectionnez le système d'exploitation de votre ordinateur.
Cliquez sur Next.



Dans la zone Preview en bas de l'assistant, cochez la case Set heading row as column names pour récupérer le nom des colonnes du fichier.

Cliquez sur Refresh Preview pour mettre à jour l'aperçu de la structure et des données.

Cliquez sur Next.



Lorsque vous cochez la case **Set heading row as column names**, le nombre que vous avez saisi dans le champ **Header** est incrémenté.

File - Step 3 of 4
Add a File metadata on repository. Define the setting of the parsed job.

File Settings
Encoding: US-ASCII
Field Separator: Semicolon Corresponding Character: ";"
Row Separator: Standard EOL Corresponding Character: "\n"
Escape Char Settings: Delimited, Escape Char: Empty

Rows To Skip
If any rows must be ignored, specify the following param
Header: 1
Footer:

Limit Of Rows
If the number of lines must be limited, specify this number
Limit:

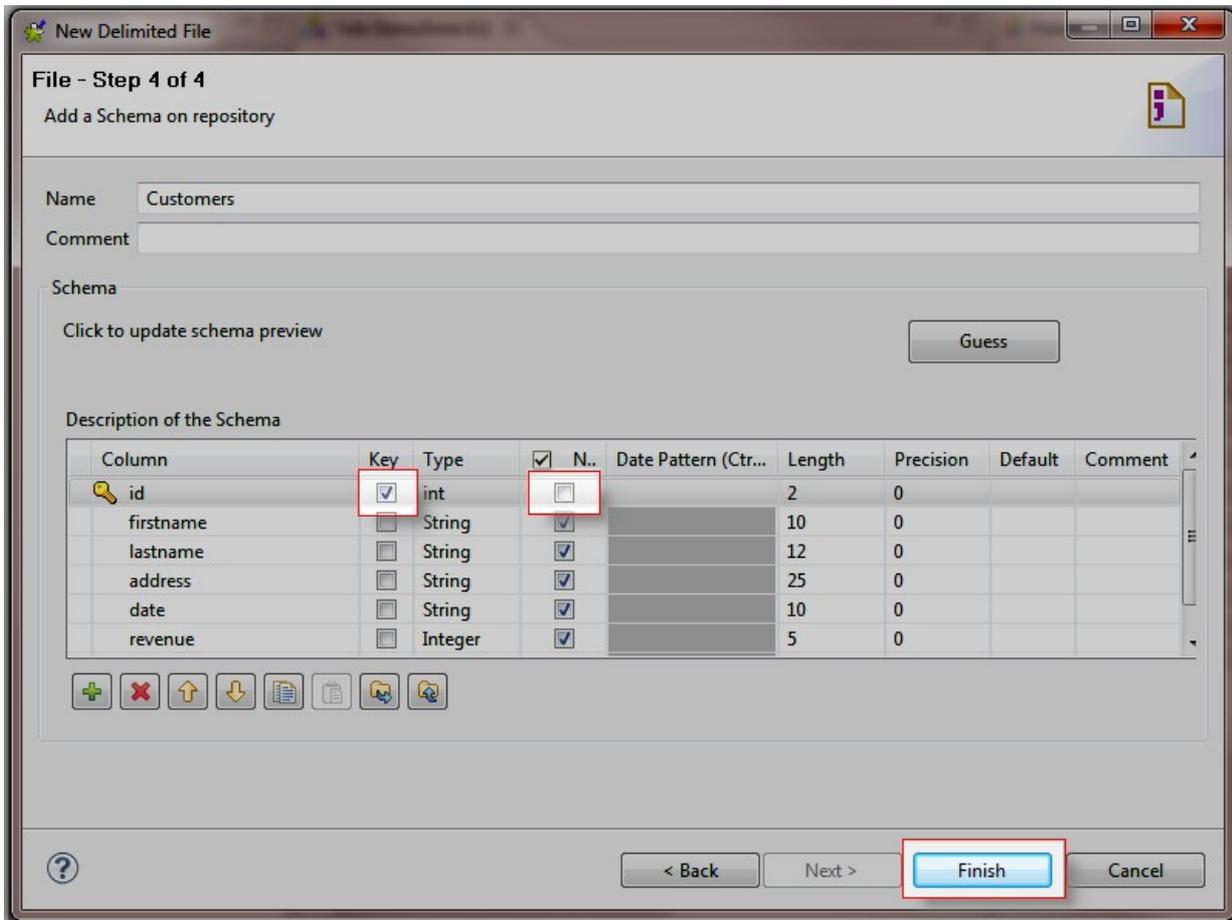
Preview | Output

Set heading row as column names

id	firstname	lastname	address	date	revenue	states
1	Andrew	Cleveland	South Roosevelt Drive	20/01/2005	55239	XX
2	Franklin	Jefferson	Carpinteria Avenue	19/05/2006	78148	NC

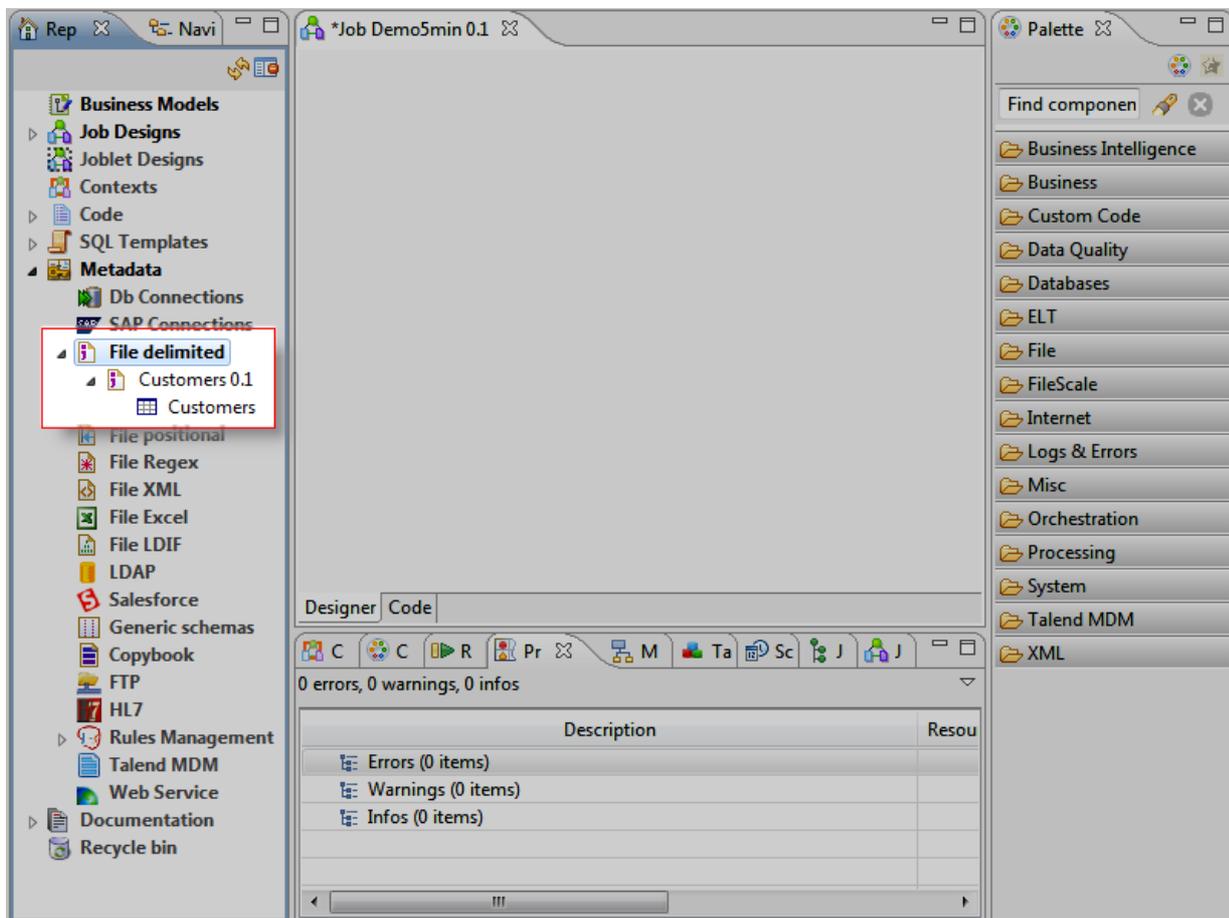
Dans la table Description of the Schema, vous pouvez modifier la description de la structure des données.

Cochez la case Key et décochez la case Nullable de la colonne id pour la définir comme clé primaire.



Dans le Repository à gauche :

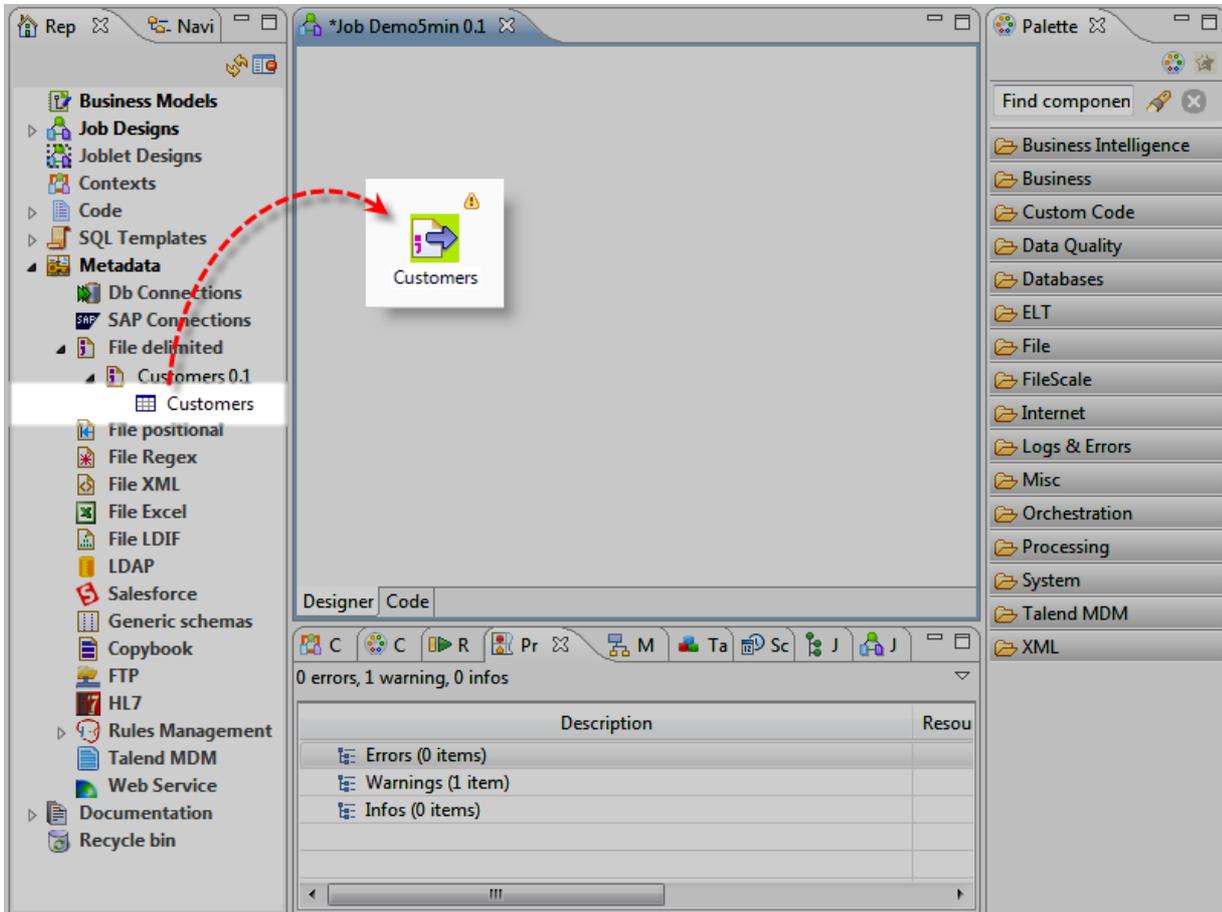
La métadonnée Customers apparaît sous le nœud Metadata > File delimited.



Dans le Repository situé à gauche :

Pour ajouter le composant d'entrée, appuyez sur Ctrl+Shift tout en glissant-déposant la métadonnée customers dans le Job Designer.

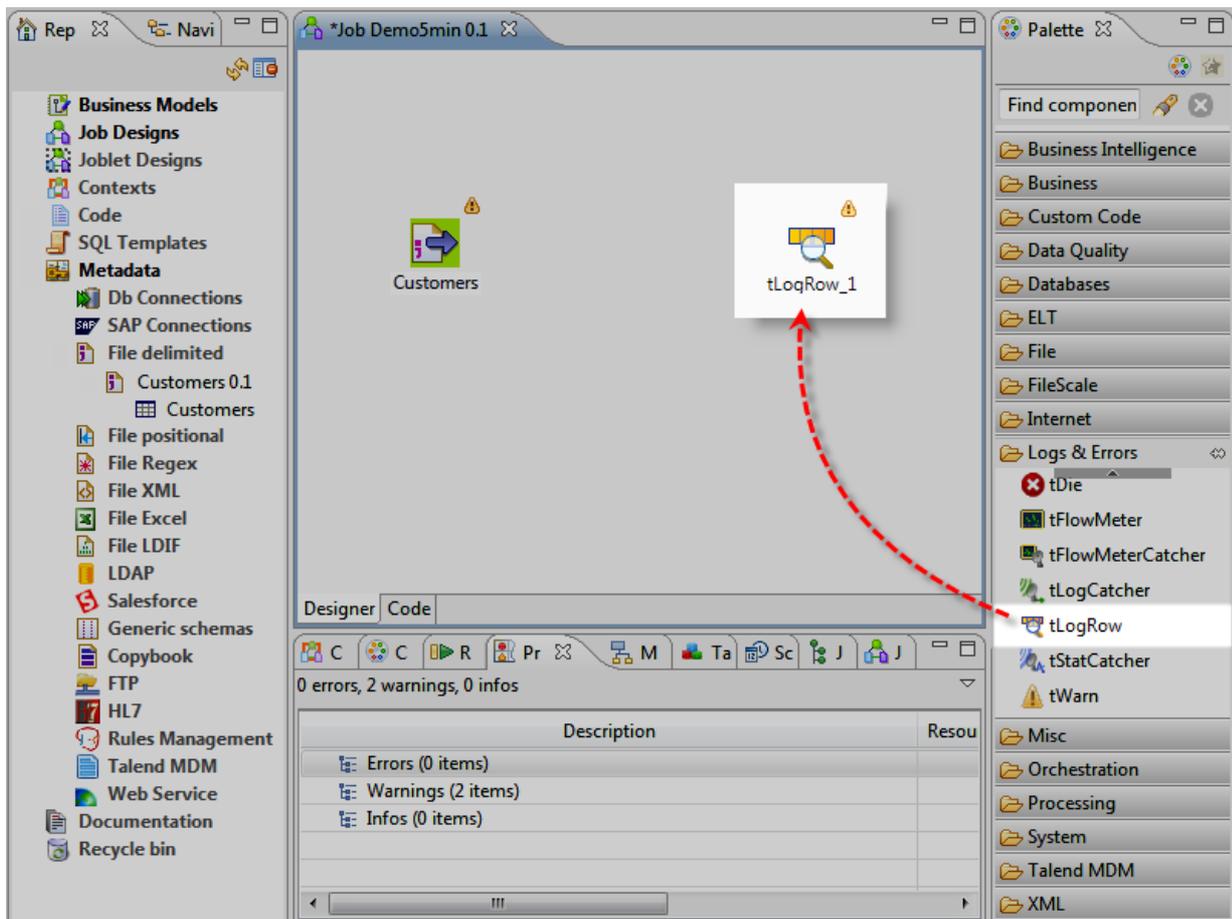
Le composant tFileInputDelimited est automatiquement créé dans le Job Designer.



Dans la Palette située à droite :

Cliquez sur la famille Logs & Errors pour ajouter le composant de sortie.

Cliquez sur le composant tLogRow et déposez-le dans le Job Designer.



Dans le Job Designer :

Pour relier les composants entre eux, cliquez-droit sur le `tFileInputDelimited` nommé Customers et déplacez-vous jusqu'au `tLogRow`.



Vous pouvez aussi créer ce lien en cliquant-droit sur le composant, en sélectionnant **Row** > **Main** dans le menu contextuel et en cliquant sur le composant de sortie.

The screenshot shows the Talend Job Designer interface for a job named '*Job Demo5min 0.1'. The central workspace displays a job design with two components: 'Customers' (a green icon with a right-pointing arrow) and 'tLogRow_1' (a yellow icon with a magnifying glass). A red dashed arrow labeled 'row1 (Main)' connects the 'Customers' component to the 'tLogRow_1' component. The interface includes a left-hand navigation pane with categories like Business Models, Job Designs, and Metadata. The right-hand side features a 'Palette' with various components such as tDie, tFlowMeter, tLogRow, and tWarn. At the bottom, there is a status bar showing '0 errors, 0 warnings, 0 infos' and a table with columns 'Description' and 'Resou'.

Description	Resou
Errors (0 items)	
Warnings (0 items)	
Infos (0 items)	

Dans le Job Designer :

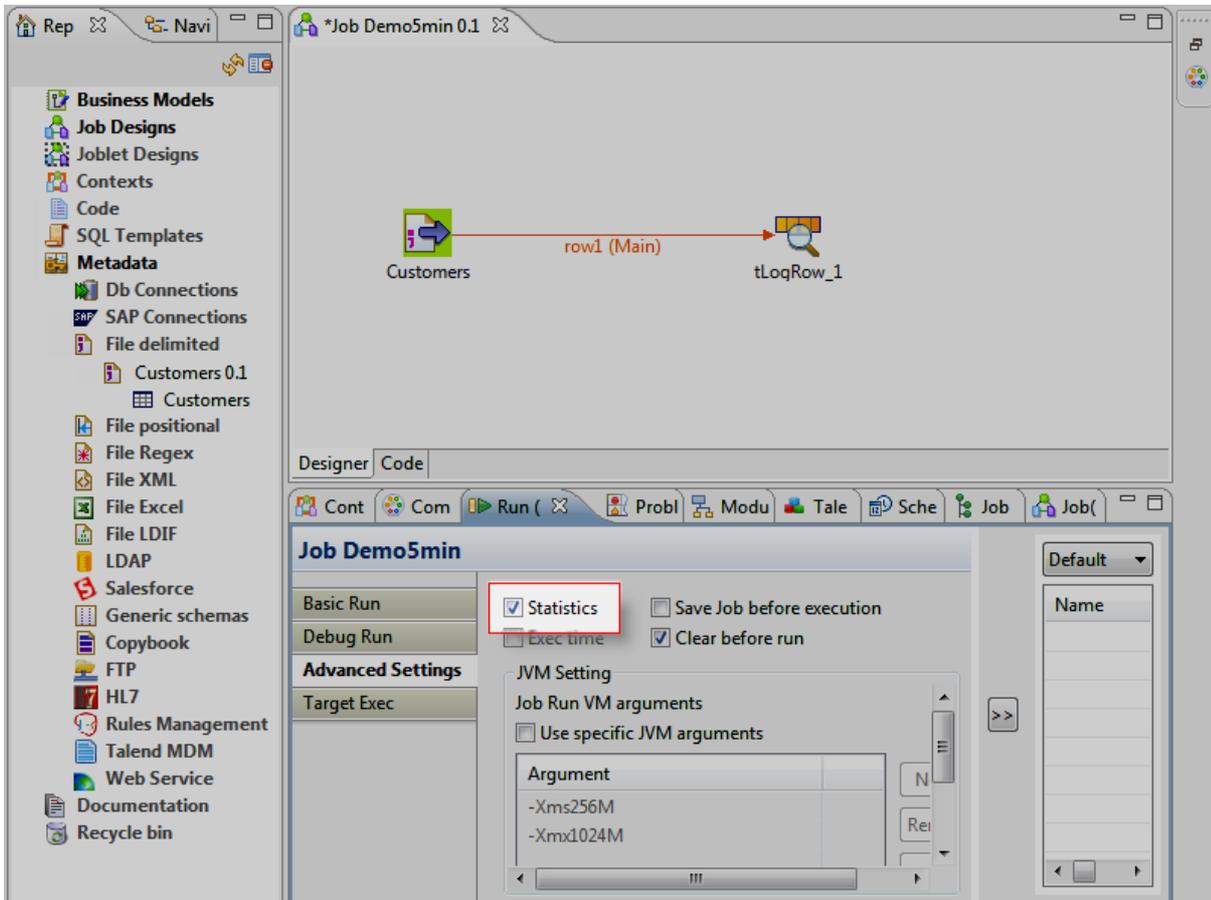
Avant d'exécuter votre Job, enregistrez-la via Ctrl+S.

Cliquez sur la vue Run.

Dans la vue Run :

Cliquez sur l'onglet Advanced Settings.

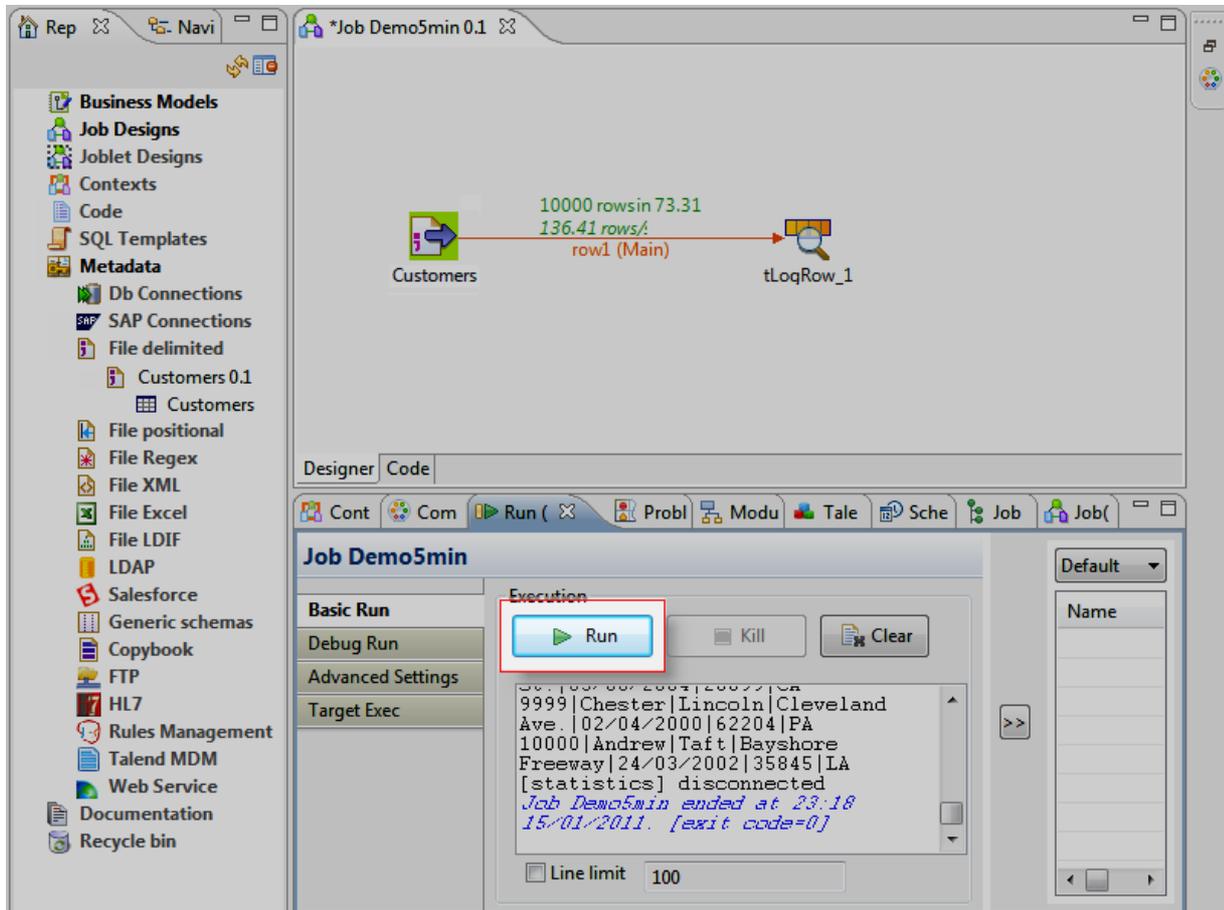
Cochez la case Statistics.



Dans la vue Run :

Cliquez sur l'onglet Basic Run.

Cliquez sur le bouton Run pour lancer l'exécution du Job en mode Statistics.



Ce premier Job est maintenant terminé !

Il lit un fichier délimité et affiche les données dans la console.

Maintenant, nous allons modifier les données à l'aide du composant de transformation et afficher le résultat dans la console.