

Talend Open Studio for Data Integration

Créer une jointure dans le Job Designer

Agréger et transformer des données à l'aide du composant tMap.

Dans ce tutoriel, vous utiliserez un fichier customers.csv contenant les informations sur des clients américains et un fichier states.txt contenant la liste de tous les Etats américains avec leur identifiant.

Dans ce tutoriel, vous utiliserez les métadonnées créées dans le tutoriel "Créer une métadonnée de connexion à un fichier délimité"

Vous paramétrez le composant tMap pour agréger les données des deux fichiers d'entrée et alimenter le fichier de sortie.

Pré-requis :

Pour suivre ce tutoriel, vous avez besoin d'extraire et d'installer les fichiers customer.csv et state.txt

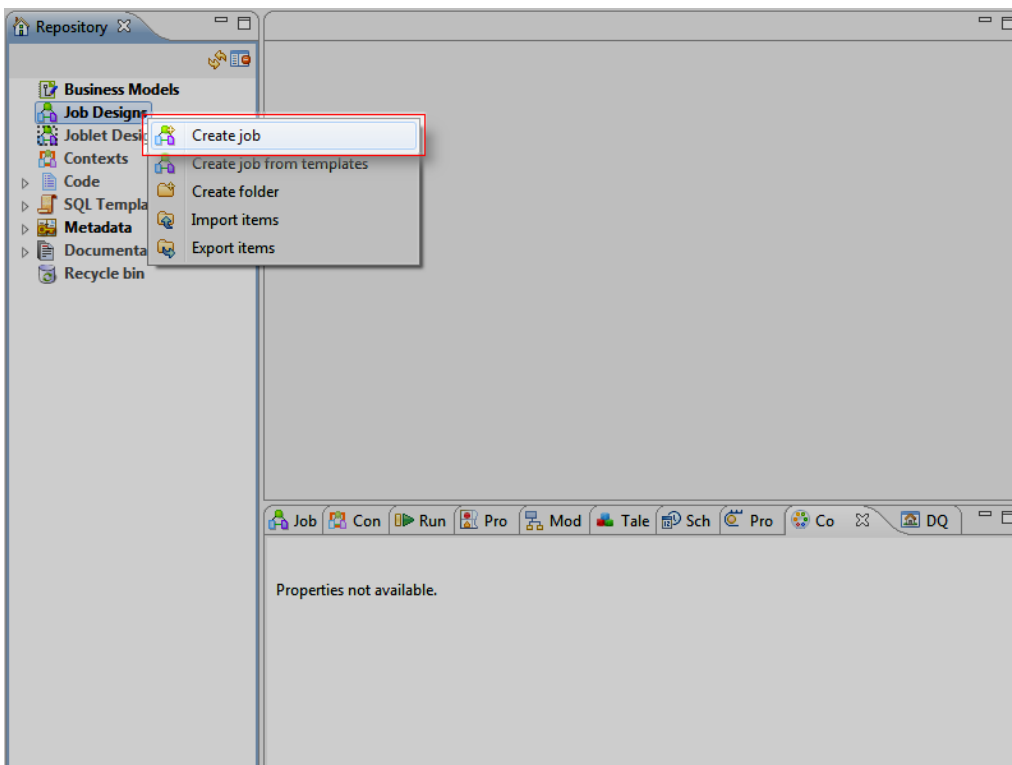


Créer le Job Design

Dans le Repository situé à gauche de Talend Open Studio :

Pour créer un Job, cliquez-droit sur Job Designs.

Dans le menu contextuel, cliquez sur Create Job pour ouvrir l'assistant New Job.



Dans l'assistant New Job

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

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

Le Job Designer présente alors un Job vierge.



Le champ **Name** ne doit pas contenir d'accents, de caractères spéciaux, d'espaces, ni de déb

The screenshot shows a 'New job' dialog box with the following fields and values:

- Name: HowToSetupJoinLink
- Purpose: (empty)
- Description: (empty)
- Author: user@company.com
- Locker: (empty)
- Version: 0.1
- Status: (empty)
- Path: (empty)

The 'Finish' button is highlighted with a red box, and the 'Name' field is also highlighted with a red box.

2

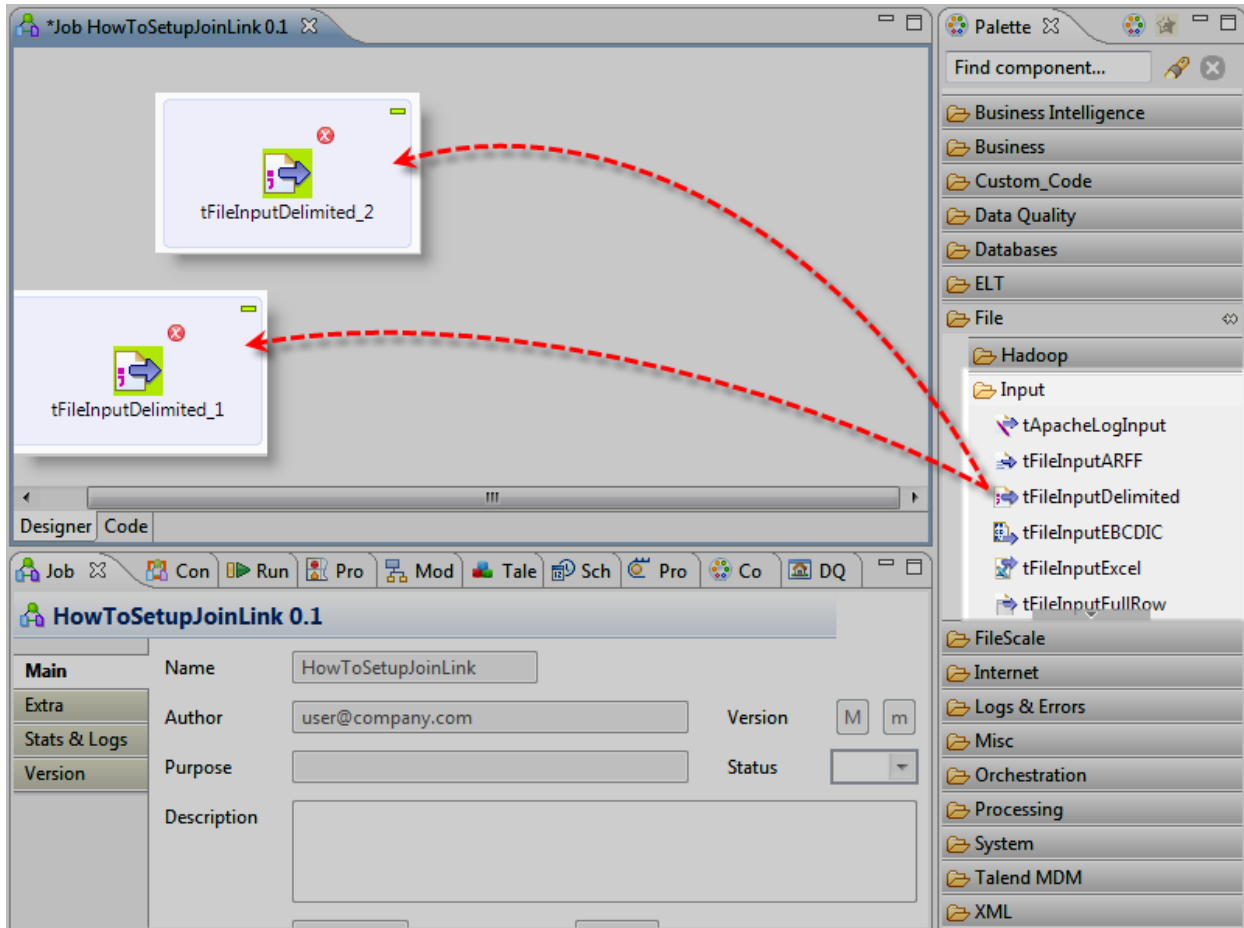
Paramétrer les connecteurs de lecture de fichiers délimités

Dans la Palette située à droite :

Pour ajouter le composant d'entrée, cliquez sur la famille File et sur la sous-famille Input.

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

De la même manière, ajoutez un deuxième composant tFileInputDelimited.



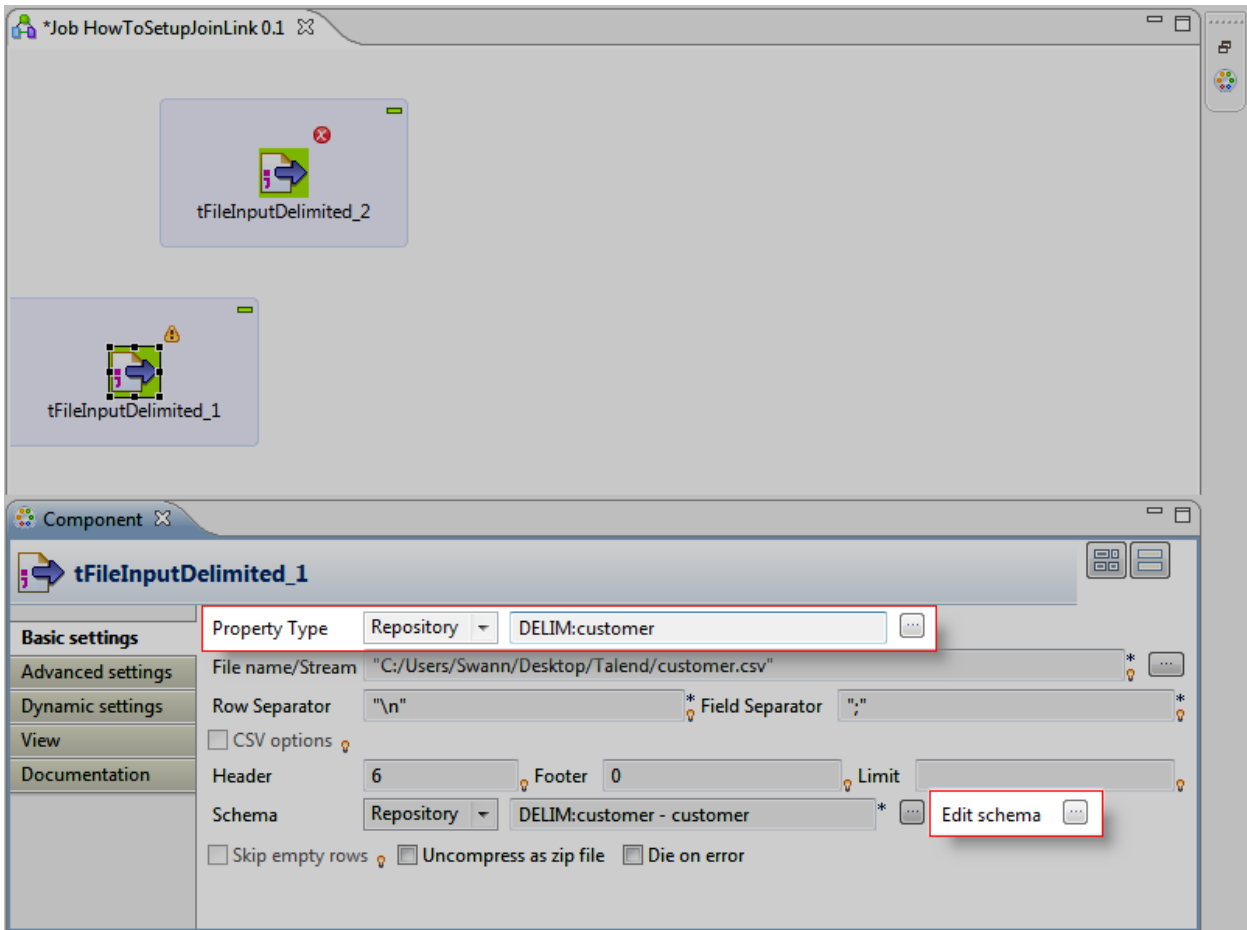
Dans le Job Designer :

Pour paramétrer les propriétés du `tFileInputDelimited_1`, double-cliquez sur le composant. La vue Component correspondante apparaît alors en bas de l'écran.

Dans la vue Component :

Pour spécifier les propriétés du composant, sélectionnez `Repository` dans la liste `Property Type` puis cliquez sur le bouton [...] situé à côté du champ `Edit schema` pour vérifier le schéma du fichier.

L'assistant `Edit parameter using repository` s'ouvre.

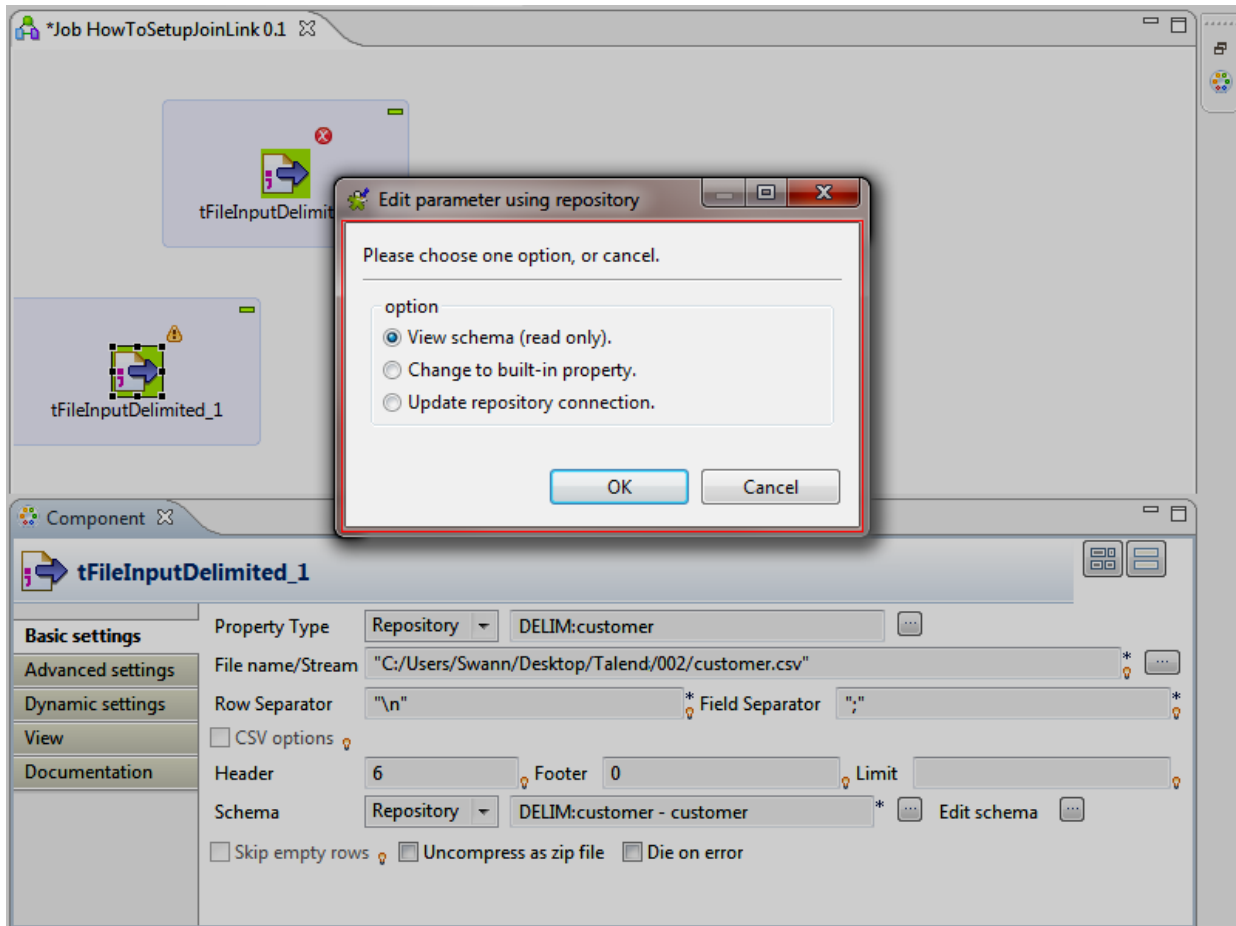


Dans l'assistant Edit parameter using repository :

Sélectionnez View schema (read only) dans la listeoption puisque vous ne souhaitez que consulter le schéma.

Cliquez sur OK.

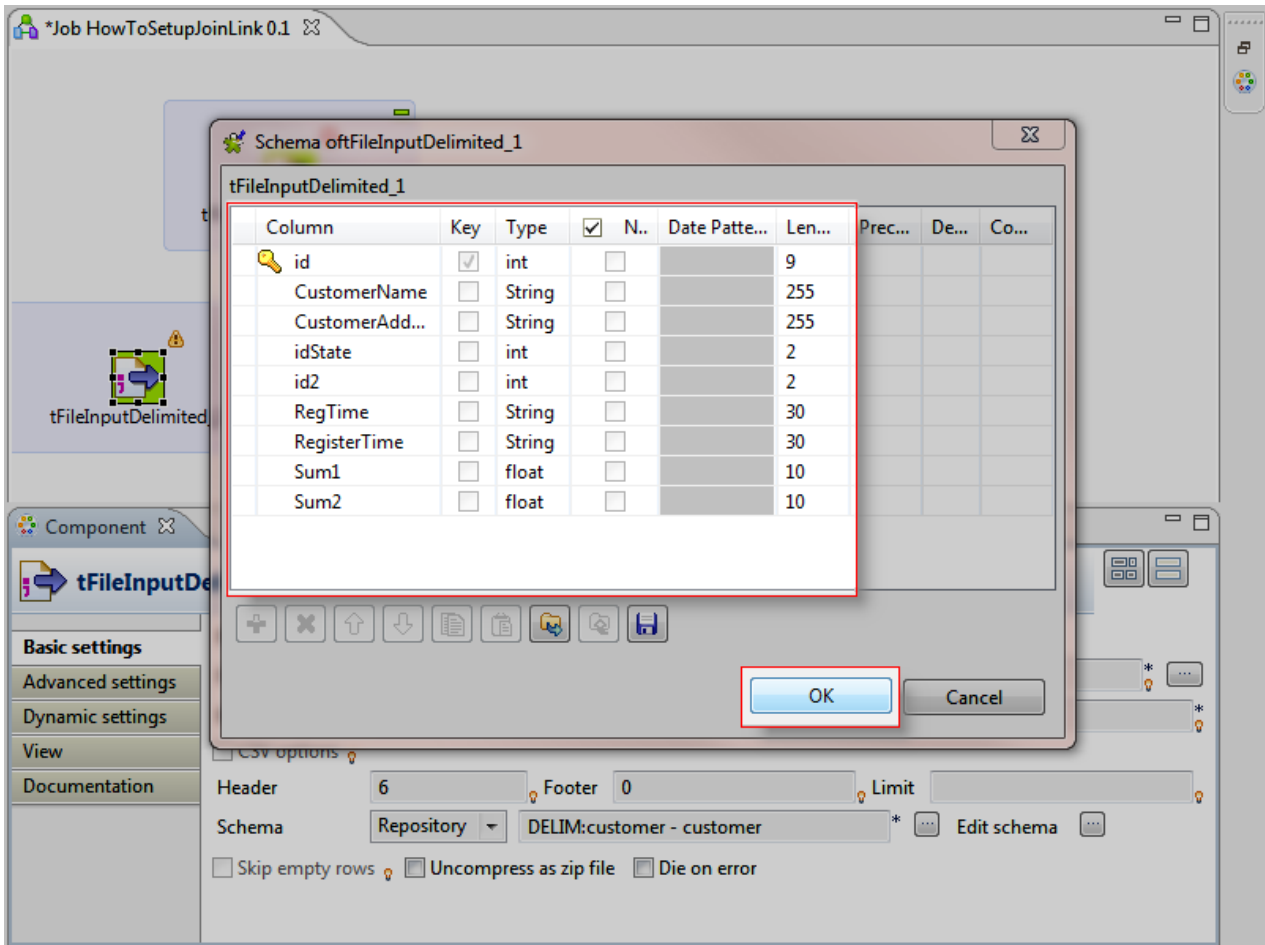
L'assistant Schema of tFileInputDelimited_1 s'ouvre.



Dans l'assistant Schema of tFileInputDelimited_1 :

Le schéma est le même que celui disponible à partir de la vue Repository sous le nœud Metadata > File delimited.

Cliquez sur OK.



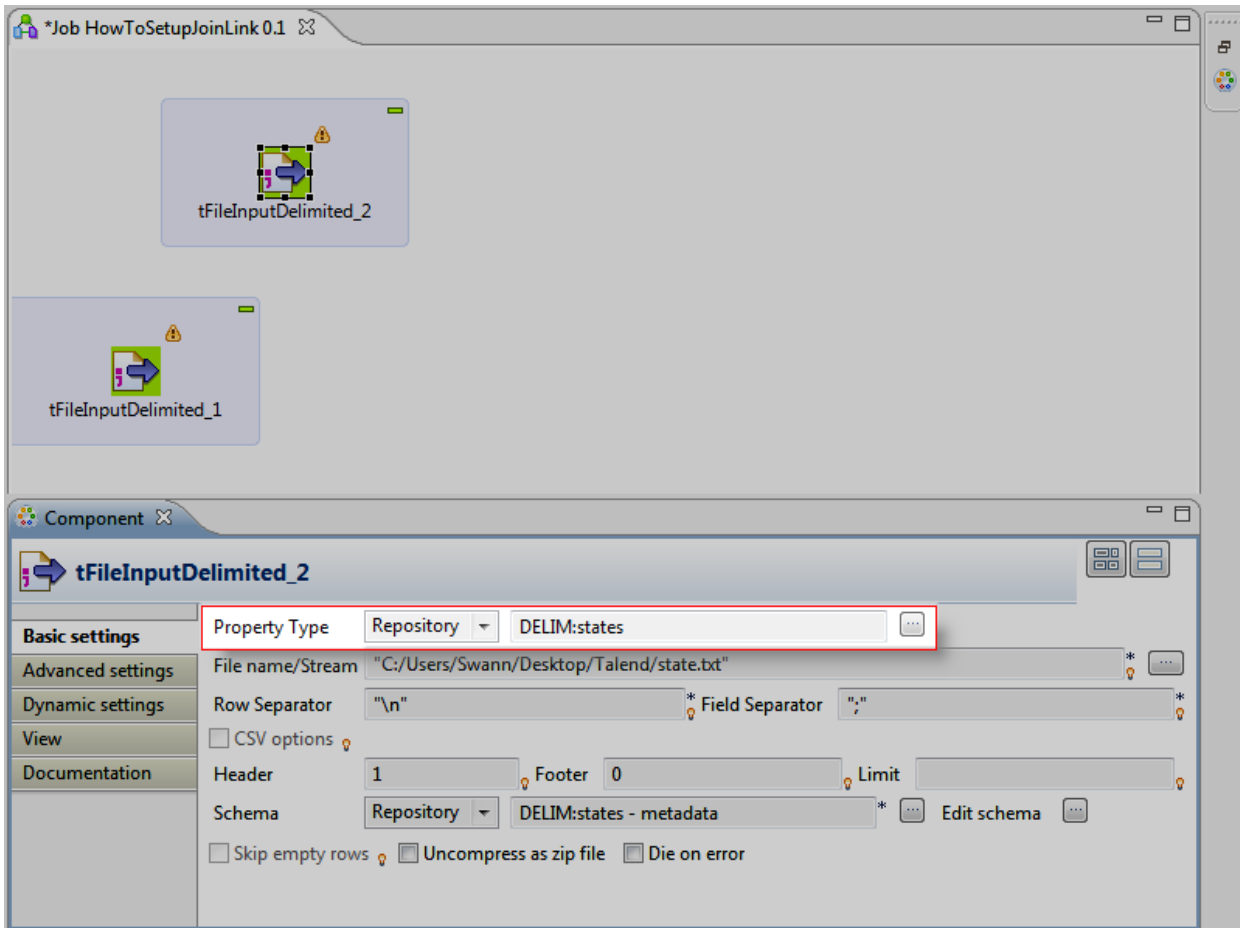
Dans le Job Designer :

Pour paramétrer les propriétés du `tFileInputDelimited_2`, double-cliquez sur le composant. La vue Component correspondante apparaît alors en bas de l'écran.

Dans la vue Component :

Pour spécifier les propriétés du composant, sélectionnez `Repository` dans la liste `Property Type` puis cliquez sur le bouton [...].

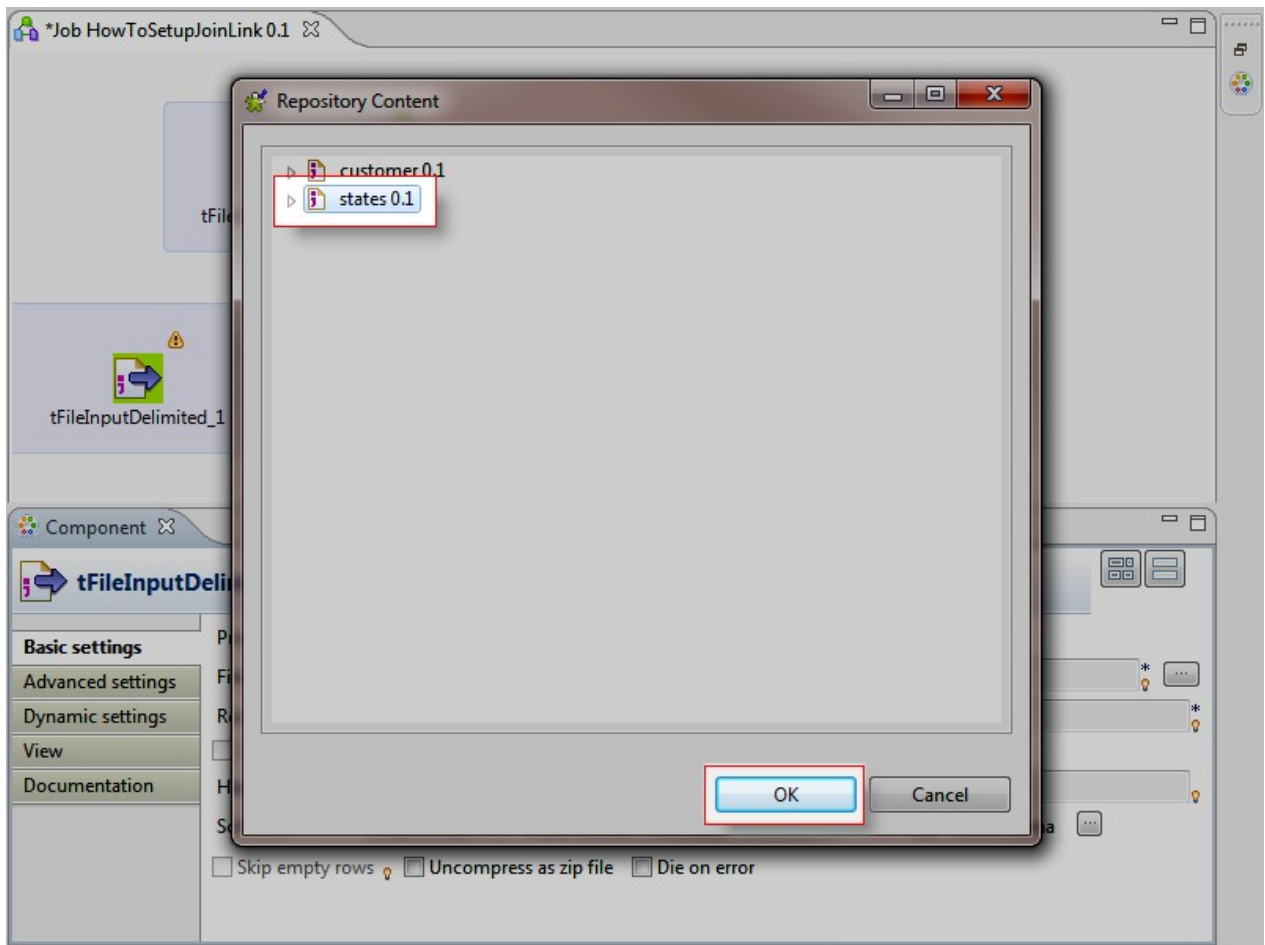
L'assistant `Repository Content` s'ouvre.



Dans l'assistant Repository Content :

Sélectionnez la métadonnée states pour que les propriétés du composant tFileInputDelimited_2 soient automatiquement renseignées avec les propriétés du fichier states.txt.

Cliquez sur OK.



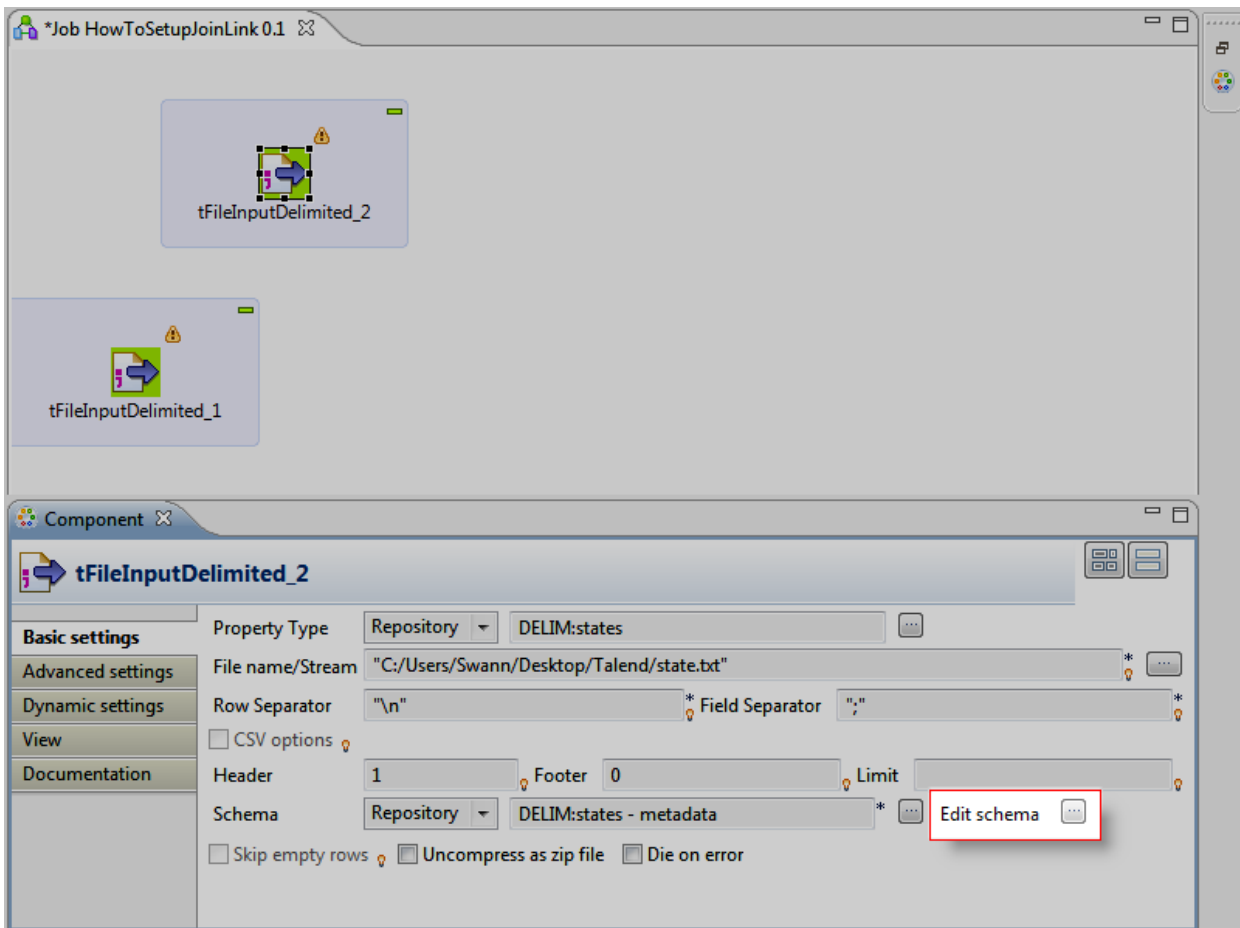
Dans la vue Component :

Cliquez sur le bouton [...] situé à coté du champ Edit schema pour vérifier le schéma de votre fichier.

L'assistant Edit parameter using repository s'ouvre.

Conservez l'option View schema (read only) puisque vous ne souhaitez que consulter le schéma et cliquez sur OK.

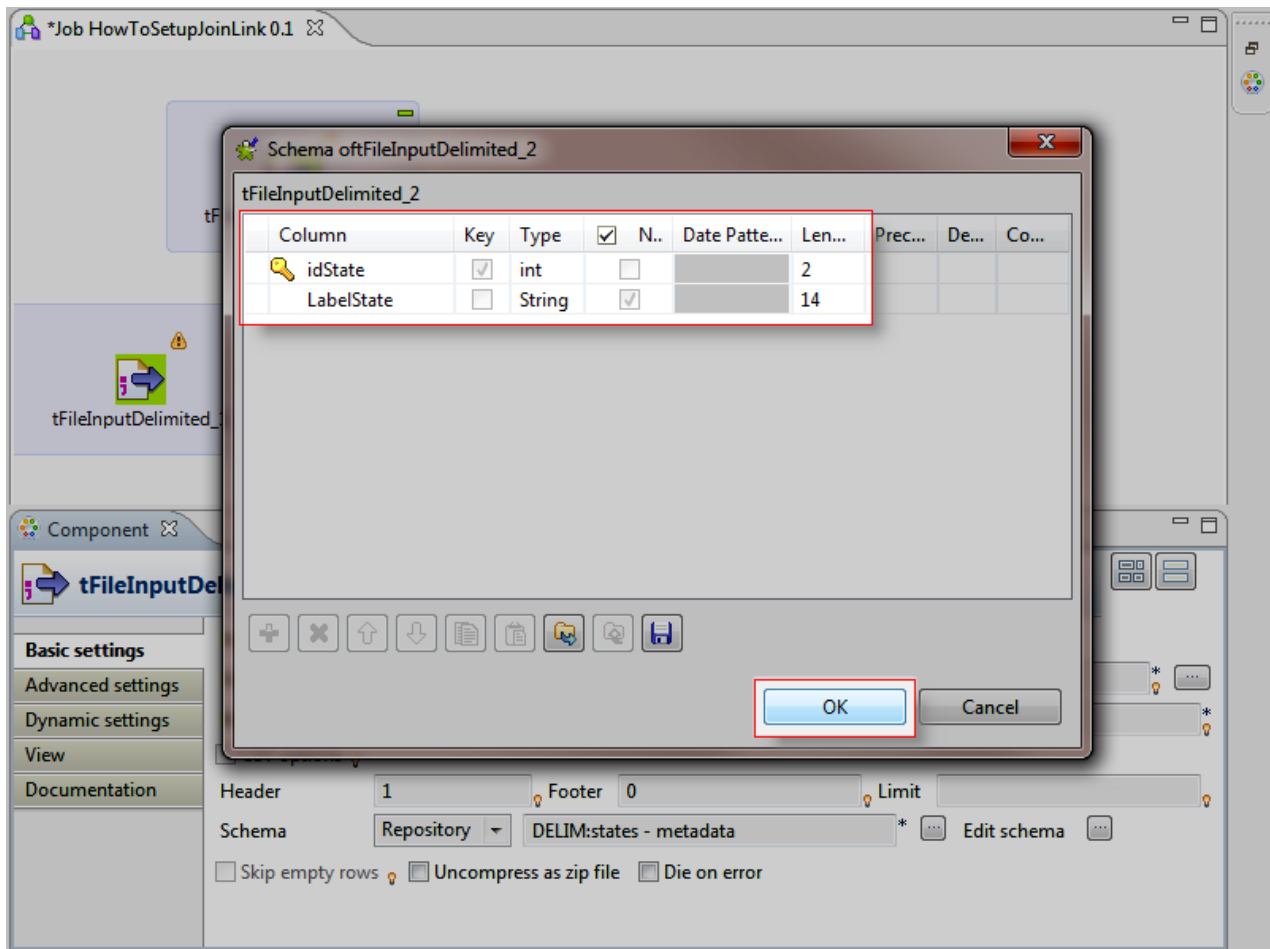
L'assistant Schema of tFileInputDelimited_2 s'ouvre.



Dans l'assistant Schema of tFileInputDelimited_2 :

Le schéma est le même que celui disponible à partir de la vue Repository sous le nœud Metadata > File delimited.

Cliquez sur OK.

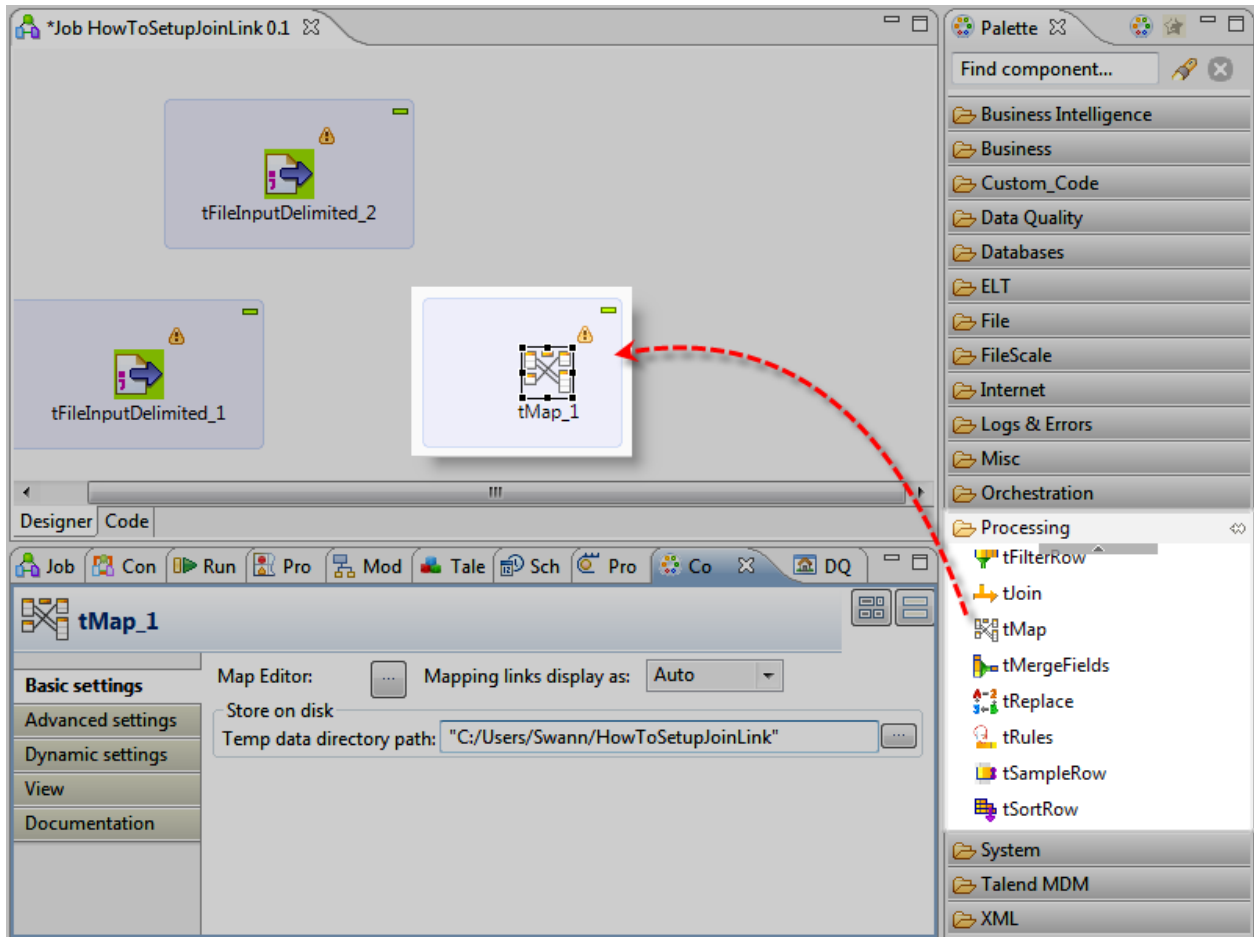


3 Déposer le composant de transformation et le connecteur d'écriture de fichier délimité dans le Job Designer

Dans la Palette située à droite :

Pour ajouter le composant de transformation, cliquez sur la famille Processing.

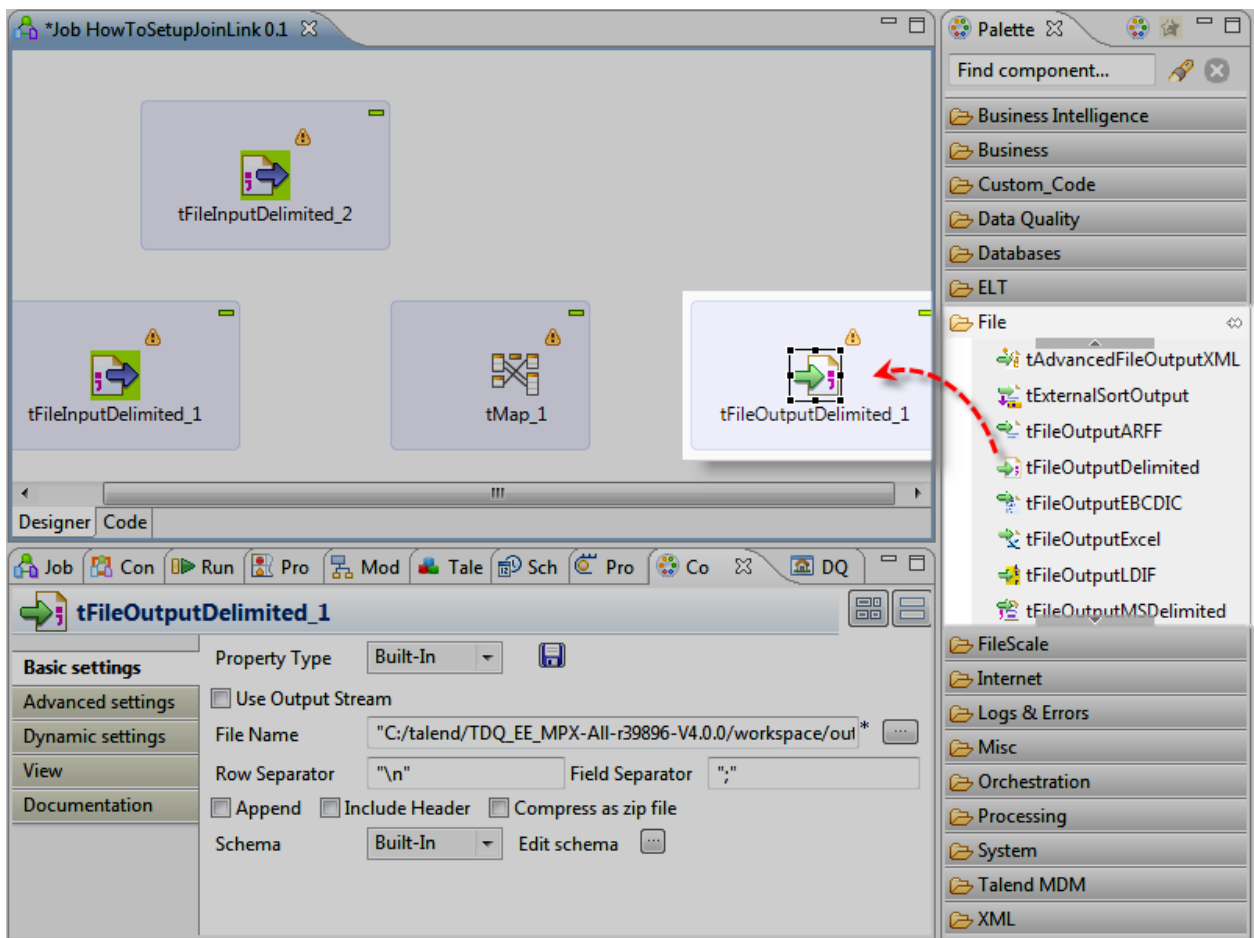
Cliquez sur le composant tMap et déposez-le dans le Job Designer.



Dans la Palette située à droite :

Pour ajouter le composant de sortie, cliquez sur la famille File et sur la sous-famille Output.

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



4

Relier les composants entre eux

Dans le Job Designer :

Cliquez-droit sur le tFileInputDelimited_1 et déplacez-vous jusqu'au tMap.

De la même manière, créez un lien du tFileInputDelimited_2 vers le tMap puis du tMap vers le tFileOutputDelimited.

Dans l'assistant tMap_1 Output, donnez le nom output au lien reliant le tMap au tFileOutputDelimited.



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 displays the Talend Job Designer interface for a job named "HowToSetupJoinLink 0.1". The main workspace shows a flow diagram with three components: **tFileInputDelimited_1**, **tFileInputDelimited_2**, and **tMap_1**. **tFileInputDelimited_1** is connected to **tMap_1** via a link labeled "row1 (Main)". **tFileInputDelimited_2** is connected to **tMap_1** via a link labeled "row2 (Lookup)". **tMap_1** is connected to **tFileOutputDelimited_1** via a link labeled "output (Main)". The **tMap_1** component has a red 'X' icon, indicating an error. The right-hand side shows the "Palette" window with a search bar and a list of components under the "File" category, including **tAdvancedFileOutputXML**, **tExternalSortOutput**, **tFileOutputARFF**, **tFileOutputDelimited**, **tFileOutputEBCDIC**, **tFileOutputExcel**, **tFileOutputLDIF**, and **tFileOutputMSDelimited**. The bottom of the screen shows the "tMap_1" configuration panel with "Basic settings" selected, showing "Map Editor" and "Temp data directory path" set to "C:/Users/Swann/HowToSetupJoinLink".

5 Paramétrer le composant de transformation et le composant d'écriture de fichier délimité

Dans le Job Designer :

Pour paramétrer les propriétés du tMap, double-cliquez sur le composant et l'éditeur du tMap s'ouvre.



Vous pouvez aussi sélectionner le **tMap** dans le Job Designer, cliquez sur la vue **Component** et cliquez sur le bouton [...] situé à côté du champ **Map Editor** pour ouvrir l'éditeur de tMap.

The screenshot displays the Talend Job Designer interface. The main workspace shows a job flow with three components: **tFileInputDelimited_1**, **tMap_1**, and **tFileOutputDelimited_1**. A second input component, **tFileInputDelimited_2**, is connected to **tMap_1** via a **row2 (Lookup)** link. The main flow consists of **row1 (Main)** from the first input to **tMap_1**, and **output (Main)** from **tMap_1** to the output component. The **tMap_1** component is highlighted with a red box. Below the workspace, the **tMap_1** configuration panel is visible, showing the **Map Editor** field with a red box around its ellipsis button. The **Temp data directory path** is set to `"C:/Users/Swann/HowToSetupJoinLink"`. On the right, the **Palette** window is open, showing a list of components under the **File** category, including **tFileOutputDelimited**.

Dans l'éditeur du tMap :

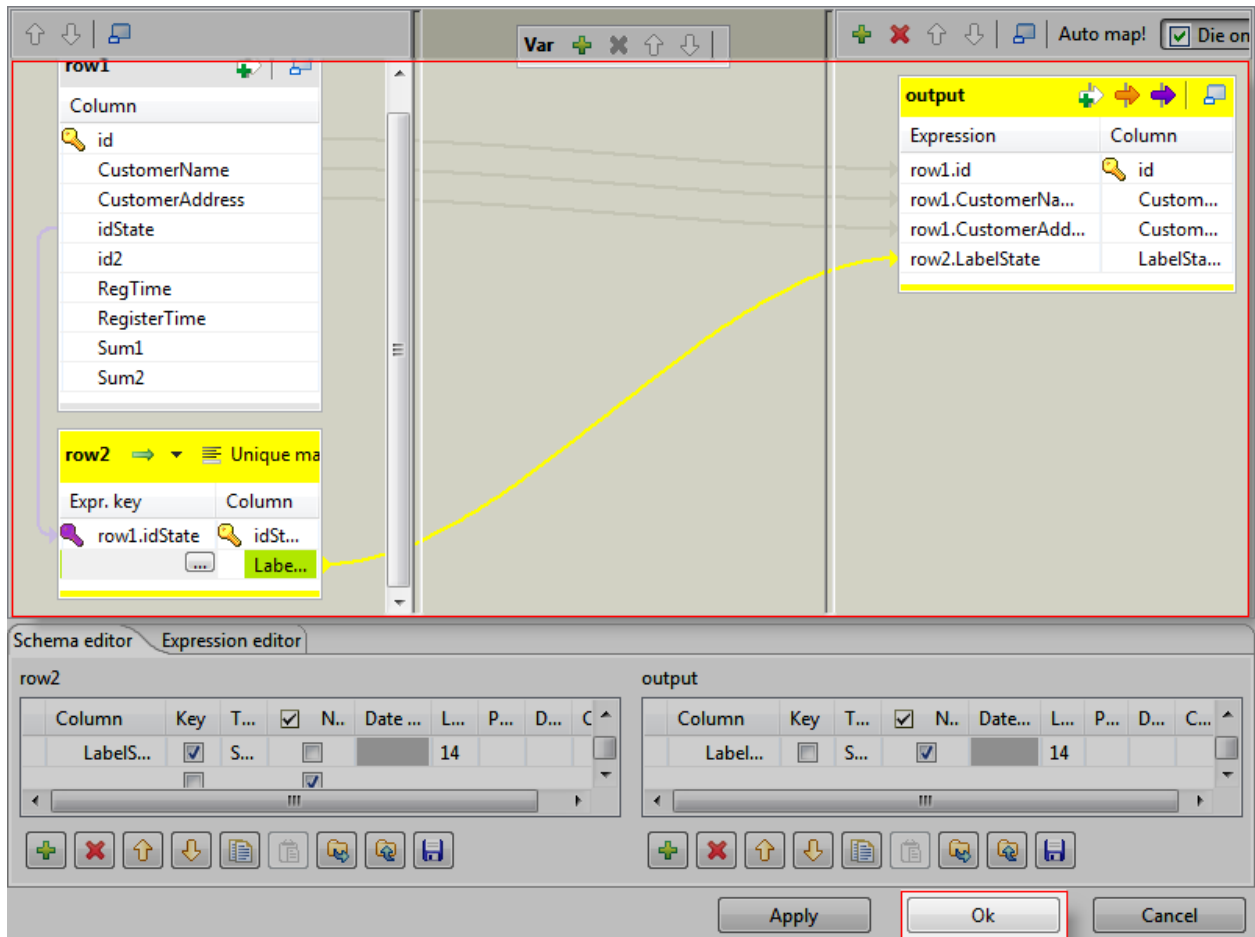
Dans la table row1, sélectionnez les colonnes id, CustomerName et CustomerAddress et glissez-les dans la table output.

Puis sélectionnez la colonne idState et glissez-la dans la colonne idState de la table row2 pour joindre les deux tables.

Dans la table row2, sélectionnez la colonne LabelSate et glissez-la dans la table output.

Cliquez sur OK.

La boîte de dialogue Propagate s'ouvre. Cliquez sur Yes pour propager le schéma défini dans l'éditeur du tMap au composant suivant.



Dans le Job Designer :

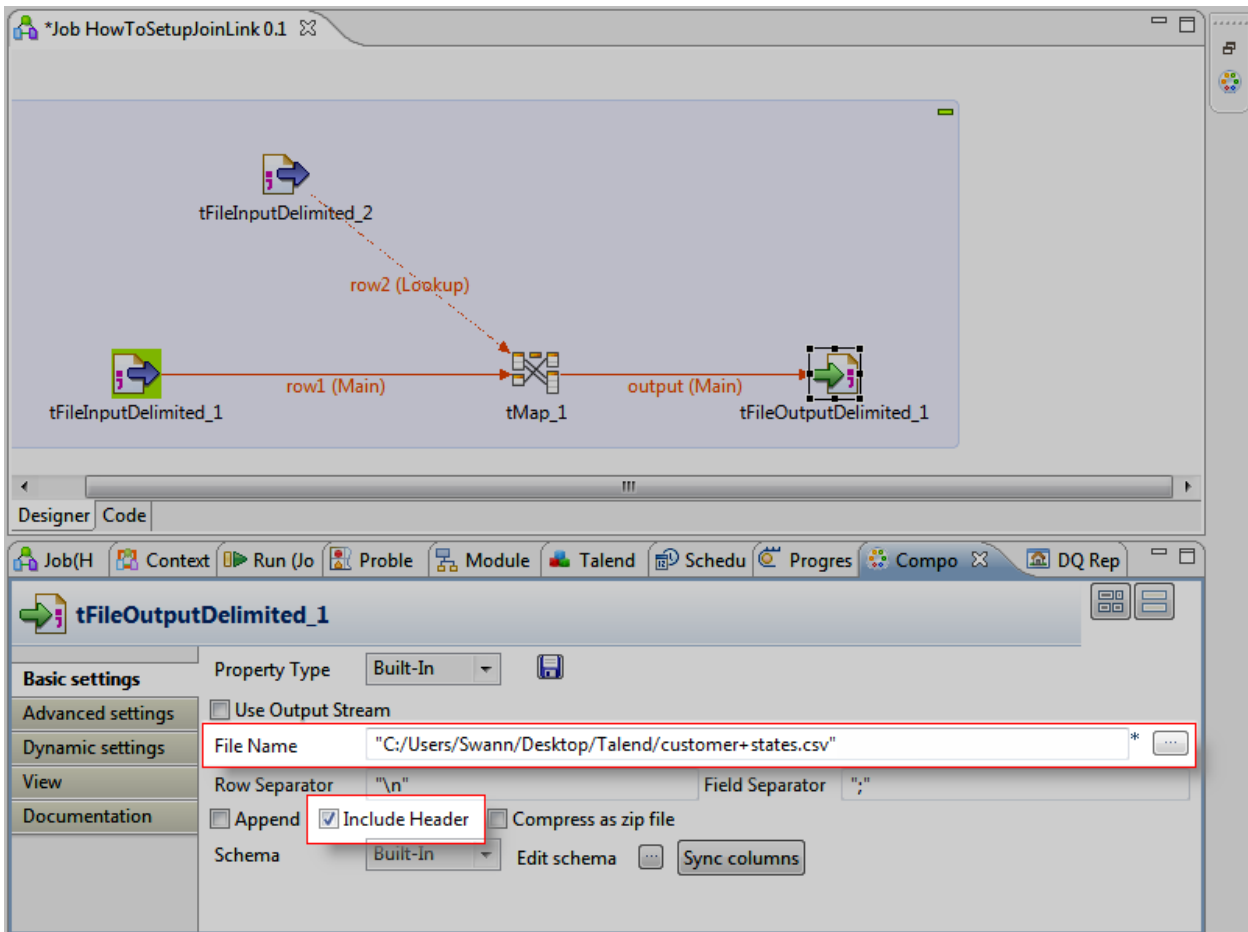
Pour paramétrer les propriétés du `tFileOutputDelimited`, double-cliquez sur ce composant. La vue Component correspondante s'affiche alors en bas de l'écran.

Dans la vue Component :

Pour spécifier le chemin du fichier qui sera créé, cliquez sur le bouton [...] à côté du champ File Name.

Dans l'assistant, définissez le même répertoire que les fichiers customer et state et nommez le fichier `customers+states.csv`.

Cochez la case `Include Header` pour récupérer les en-têtes des fichiers d'entrée.



Dans le Job Designer :

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

Appuyez sur F6 pour lancer l'exécution du Job.

La vue Run s'affiche en bas de Talend Open Studio et la console retrace l'exécution du Job.



Exécutez de nouveau ce Job mais en cochant la case **Statistics** de la vue **Run**.

The screenshot displays the Talend Open Studio interface. The top part shows the Job Designer with a workflow diagram. The workflow consists of three main components: **tFileInputDelimited_1**, **tMap_1**, and **tFileOutputDelimited_1**. The flow is as follows: **tFileInputDelimited_1** feeds into **tMap_1** via a link labeled **row1 (Main)**. **tMap_1** has two outgoing links: **row2 (Lookup)** which feeds into **tFileInputDelimited_2**, and **output (Main)** which feeds into **tFileOutputDelimited_1**. Below the designer, the **Run** tab is active, showing the execution console for the job **Job HowToSetupJoinLink**. The console shows the following log entries:

```
Starting job HowToSetupJoinLink at 00:44 30/12/2010.  
Job HowToSetupJoinLink ended at 00:44 30/12/2010. [exit c...
```

The console also shows a table for context variables:

Name	Value

Le Job HowtoSetupJoinLink est presque terminé !

Il permet d'agréger les données des deux fichiers d'entrée et d'alimenter le fichier de sortie.

Il ne reste plus qu'à le documenter !

Dans le Job Designer :

Pour documenter votre Job, donnez-lui un titre.

Pour cela, cliquez sur la zone bleue entourant votre Job.

Cliquez sur la vue Component.

Pour lui ajouter un titre, cochez la case Show subjob title et dans le champ Title, saisissez le titre correspondant : Set up a Join link (Créer une jointure, en français).

Enregistrez de nouveau votre Job.

The screenshot displays the Talend Job Designer interface. The main workspace shows a job graph titled "Set up a Join Link". The graph consists of three components: "tFileInputDelimited_1" (Main), "tMap_1" (Join), and "tFileOutputDelimited_1" (Main). A second component, "tFileInputDelimited_2" (Lookup), is connected to "tMap_1" via a "row2 (Lookup)" link. The main flow is labeled "row1 (Main)" and "output (Main)".

Below the workspace, the "Subjob" settings panel is visible. Under "Advanced settings", the checkbox "Show subjob title" is checked. The "Title" field contains the text "Set up a Join Link". The "Title color" and "Subjob color" fields are also visible.

Ce tutoriel est maintenant fini.

Le job fonctionne et est documenté.