# Table of Contents

**Preface**  ........................................................................................................... v

1. General information  ........................................................................................... v
   1.1. Purpose ......................................................................................................... v
   1.2. Audience ...................................................................................................... v
   1.3. Typographical conventions ......................................................................... v

2. Feedback and Support  ......................................................................................... v

Chapter 1. Getting Started with Talend Studio ....................................................... 1

1.1. Launching Talend Studio .................................................................................. 2
   1.1.1. How to launch the Studio for the first time ........................................... 2
   1.1.2. How to connect to TalendForge ......................................................... 3
   1.1.3. How to access a Repository .................................................................. 5
   1.1.4. How to set up a project ....................................................................... 6

1.2. Working with different workspace directories ................................................. 7
   1.2.1. How to create a new workspace directory ............................................. 7
   1.2.2. How to connect to a different workspace directory .............................. 8

1.3. Working with projects ..................................................................................... 10
   1.3.1. How to create a project ....................................................................... 10
   1.3.2. How to import a demo project ........................................................... 11
   1.3.3. How to open a project ........................................................................ 13

Chapter 2. Working in Talend Studio - basic data integration Job examples .......... 15

2.1. Getting started with a basic Job ..................................................................... 16
   2.1.1. Creating a Job ................................................................................... 16
   2.1.2. Adding components to the Job ......................................................... 17
   2.1.3. Connecting the components together ................................................. 20
   2.1.4. Configuring the components ............................................................. 21
   2.1.5. Executing the Job ............................................................................. 23

2.2. Use cases ....................................................................................................... 24
   2.2.1. Updating data in a database table ...................................................... 24
   2.2.2. Mapping data using a filter and a simple explicit join ......................... 27

Appendix A. Glossary  ............................................................................................ 33
Preface

1. General information

1.1. Purpose

This guide aims at helping users get started with the Talend Open Studio for Data Integration quickly. For detailed explanations on features and functions of the Talend Open Studio for Data Integration, see the other documentation delivered with the Talend Open Studio for Data Integration.

Information presented in this document applies to Talend Open Studio for Data Integration 6.1.2.

1.2. Audience

This guide is for users and administrators of Talend Open Studio for Data Integration.

💡 The layout of GUI screens provided in this document may vary slightly from your actual GUI.

1.3. Typographical conventions

This guide uses the following typographical conventions:

- text in **bold**: window and dialog box buttons and fields, keyboard keys, menus, and menu and options,
- text in [bold]: window, wizard, and dialog box titles,
- text in **courier**: system parameters typed in by the user,
- text in *italics*: file, schema, column, row, and variable names,

💡 The 📝 icon indicates an item that provides additional information about an important point. It is also used to add comments related to a table or a figure,

⚠️ The ⚠ icon indicates a message that gives information about the execution requirements or recommendation type. It is also used to refer to situations or information the end-user needs to be aware of or pay special attention to.

Any command is highlighted with a grey background or code typeface.

2. Feedback and Support

Your feedback is valuable. Do not hesitate to give your input, make suggestions or requests regarding this documentation or product and find support from the Talend team, on Talend's Forum website at:
http://talendforge.org/forum
Chapter 1. Getting Started with Talend Studio

This chapter provides basic information required to get started with Talend Studio, including launching Talend Studio and creating projects.

After this chapter, you are encouraged to try your own Job designs in your Talend Studio by following the examples given in this guide, and read the Talend Studio User Guide to learn more about your Talend Studio.
1.1. Launching Talend Studio

This section guides you through the basics for launching Talend Studio for the first time and opening your first project in the Studio, and provides information on setting up a project.

1.1.1. How to launch the Studio for the first time

To open Talend Studio for the first time, complete the following:

1. Uncompress the Talend Studio zip file and, in the folder, double-click the executable file corresponding to your operating system.

   The Studio zip archive contains binaries for several platforms including Mac OS X and Linux/Unix.

2. In the [User License Agreement] dialog box that opens, read and accept the terms of the end user license agreement to proceed.

3. In the Talend Studio login window, select an option to define your project that will hold all Jobs and Business models designed in the Studio.

   - Select Create a new project, specify a project name and click Finish to create a new project. For more information, see How to create a project.

   - Select Import a demo project and click Finish to import a demo project that includes numerous samples of ready-to-use Jobs. This Demo project can help you understand the functionalities of different Talend components. For more information, see How to import a demo project.

   - Select Import an existing project and click Finish to import an existing projects. For more information, see your Talend Studio User Guide.
• If you want to modify the default repository connection, click **Manage Connections** to set up your connection before setting up a project. For further information about connecting to a repository, see *How to access a Repository*.

As the purpose of this procedure is to create a new project, select **Create a new project**, fill in a project name in the text field, and click **Finish**.

4. Depending on the Studio product you are using, you will see either of the following:

   • A Quick Tour to **Talend Studio**. Click **Next** to go the next slide of the presentation, or click **Close** to end the presentation and display the main window of your **Talend Studio**.

      This presentation automatically starts at the Studio initial launch. To open it manually later, go to **Help > Studio Quick Tour** from the Studio menu bar.

   • The **[Welcome]** window, which provides direct links to Demo projects, user documentation, tutorials, **Talend** forum, **Talend** on-demand training and **Talend** latest news. Click **Start now!** to open **Talend Studio** main window, which displays a page that provides useful tips for beginners on how to get started with the Studio. Clicking an underlined link brings you to the corresponding tab view or opens the corresponding dialog box.

When the **[Additional Talend Packages]** wizard opens, install additional packages such as language packs if needed. For more information, see the section about installing additional packages in the **Talend Installation and Upgrade Guide**.

5. You can skip this installation step and close the wizard by clicking **Cancel**.

   This wizard appears each time you launch the studio if any additional package is available for installation unless you select the **Do not show this again** check box. You can also display this wizard by selecting **Help > Install Additional Packages** from the menu bar.

**1.1.2. How to connect to TalendForge**

Every fourth time you launch **Talend Studio**, until you are connected to the **Talend** Community, the **[Connect to TalendForge]** dialog box opens, inviting you to connect to the **Talend** Community so that you can check, download, install external components and upload your own components to the **Talend** Community to share with other **Talend** users directly in the **Exchange** view of your Job designer in the Studio.

To learn more about the **Talend** Community, click the **TalendForge Terms of Use** link. For more information on using and sharing community components, see the section on how to download/upload **Talend** community components of your Studio User Guide.

If you want to connect to the **Talend** Community later, click **Skip this Step** to continue launching the Studio without setting up a connection to the **Talend** Community.

1. By default, the Studio will automatically collect product usage data and send the data periodically to servers hosted by **Talend** for product usage analysis and sharing purposes only. If you do not want the Studio to do so, clear the **I want to help to improve Talend by sharing anonymous usage statistics** check box.

   You can also turn on or off usage data collection from the **[Preferences]** dialog box (**Talend > Usage Data Collector**). For more information, see the section on setting **Talend Studio** preferences of your Studio User Guide.

2. Fill in the required information, select the **I Agree to the TalendForge Terms of Use** check box, and click **CREATE ACCOUNT** to create your account and connect to the **Talend** Community automatically and continue launching the Studio.
Be assured that any personal information you may provide to **Talend** will never be transmitted to third parties nor used for any purpose other than joining and logging in to the **Talend** Community and being informed of **Talend** latest updates.

If you already have created an account at [http://www.talendforge.org](http://www.talendforge.org), click **Connect to Existing Account**, fill in your user name and password, and click **CONNECT TO MY ACCOUNT** to sign in the **Talend** Community and continue launching the Studio.
1.1.3. How to access a Repository

When launching Talend Studio, you can connect to a local repository where you store the data for your projects, including Jobs and business models, metadata, routines, etc. You can also connect to a remote repository where you store the same type of data to work collaboratively on projects.

1.1.3.1. How to connect to a local repository

To set a connection to a local repository, do the following:

1. On the login window of Talend Studio, click the Manage Connections button to open the repository connection setup dialog box.
1.1.4. How to set up a project

To open Talend Studio, you must first set up a project.

You can set up a project by:

- creating a new project. For more information, see How to create a project.
- importing one or more projects you already created in other sessions of Talend Studio. For more information, see Talend Studio User Guide.

Depending on the Studio product you are using, the product information displayed in your Studio may differ slightly from what is shown above.

2. If needed, type in a name and a description for your connection in the relevant fields.

3. In the **User E-mail** field, type in the email address that will be used as your user login. This field is compulsory to be able to use Talend Studio.

   Be aware that the email entered is never used for purposes other than logging in.

4. By default, the **Workspace** field shows the path to the current workspace directory which contains all of the folders belonging to the project created. To change the workspace directory, type in the name of an existing directory or click the [...] button next to the **Workspace** field and browse to your preferred workspace directory. Upon changing your workspace directory, unless it is the first startup, you need to restart your Talend Studio by clicking the **Restart** button back on the login window for your change to take effect.

   For more information about workspace directories, see Working with different workspace directories.

5. Click **OK** to validate your changes and return to the login window.
• importing the Demo project. For more information, see How to import a demo project.

1.2. Working with different workspace directories

*Talend Studio* makes it possible to create many workspace directories and connect to a workspace different from the one you are currently working on, if necessary.

This flexibility enables you to store these directories wherever you want and give the same project name to two or more different projects as long as you store the projects in different directories.

1.2.1. How to create a new workspace directory

*Talend Studio* is delivered with a default workspace directory. However, you can create as many new directories as you want and store your project folders in them according to your preferences.

1. If you have already started the Studio, select *File > Switch Project or Workspace* from the menu bar to restart the Studio.

2. On the login window, click *Manage Connections* to open the connection setup dialog box.

3. On the connection setup dialog box, click the [...] button next to the *Workspace* field.

![Talend Data Fabric Connection Setup](image)

4. In the [Browse For Folder] dialog box, browse to the parent directory under which you want to create a new workspace directory, click *Make New Folder*, and enter the name of your new workspace directory. Then click *OK* to validate directory creation and close the dialog box.
How to connect to a different workspace directory

In Talend Studio, you can select the workspace directory you want to store your project folders in according to your preferences.

1. If you have already started the Studio, select File > Switch Project or Workspace from the menu bar to restart the Studio.

2. On the login window, click the Manage Connections button to open the connection setup dialog box.

3. On the connection setup dialog box, click the [...] button next to the Workspace field.

5. Click OK to validate your connection setup and go back to the login window.

6. Back on the login window, click the Restart button to restart your Talend Studio for the change to take effect.
4. In the [Browse For Folder] dialog box, browse to your preferred folder to use as the new workspace directory, and click OK to validate your directory selection and close the dialog box.

5. Click OK to validate your connection setup and go back to the login window.
6. Back on the login window, click the Restart button to restart your Talend Studio for the change to take effect.

### 1.3. Working with projects

In *Talend Studio*, the highest physical structure for storing all different types of data integration Jobs, metadata, routines, and so on is the "project".

This section will guide you through the basic steps to manage projects before starting your work on Business Models, Jobs, and so on in your *Talend Studio*. For more information on project management, see your *Talend Studio User Guide*.

#### 1.3.1. How to create a project

To create a project at the initial startup of the Studio, do the following:

1. Launch *Talend Studio*.

2. On the login window, select the **Create a new project** option and enter a project name in the field.

3. Click **Finish** to create the project and open it in the Studio.

To create a new project after the initial startup of the Studio, do the following:

1. On the login window, select the **Create a new project** option and enter a project name in the field.
How to import a demo project

2. Click **Create** to create the project. The newly created project is displayed on the list of existing projects.

3. Select the project on the list and click **Finish** to open the project in the Studio.

Later, if you want to switch between projects, on the Studio menu bar, use the combination **File > Switch Project or Workspace**.

### 1.3.2. How to import a demo project

You can import one or more demo projects that include numerous samples of ready to use Jobs into your *Talend Studio* to help you understand the functionalities of different *Talend* components.

To import a demo project, proceed as follows:
1. When launching your Talend Studio, select the **Import a demo project** option on the Studio login window and click **Select**, or click the **Demos** link on the welcome window, to open the **[Import demo project]** dialog box.

After launching the Studio, click button on the toolbar, or select **Help > Welcome** from the Studio menu bar to open the welcome window and then click the **Demos** link, to open the **[Import demo project]** dialog box.

2. In the **[Import Demo Project]** dialog box, select the demo project you want to import and view the description on the right panel.

   *The demo projects available in the dialog box may vary depending on the product you are using.*

   ![Import demo project dialog box](image)

3. Click **Finish** to close the dialog box.

4. In the new dialog box that opens, type in a new project name and description information if needed.
5. Click **Finish** to create the project.

All the samples of the demo project are imported into the newly created project, and the name of the new project is displayed in the **Project** list on the login screen.

6. To open the imported demo project in *Talend Studio*, back on the login window, select it from the **Project** list and then click **Finish**.

The Job samples in the open demo project are automatically imported into your workspace directory and made available in the **Repository** tree view under the **Job Designs** folder.

### 1.3.3. How to open a project

*When you launch Talend Studio for the first time, no project names are displayed on the **Project** list. First you need to create a project or import a Demo project in order to populate the **Project** list with the corresponding project names that you can then open in the Studio.***

To open a project in *Talend Studio*:

On the Studio login screen, select the project of interest from the project list and click **Finish**.
A progress bar appears. Wait until the task is complete and the *Talend Studio* main window opens.

When you open a project imported from a previous version of the Studio, an information window pops up to list a short description of the successful migration tasks.
Chapter 2. Working in *Talend Studio* - basic data integration Job examples

This chapter provides basic data integration Job examples to help users get started with *Talend Studio*. For more real-life examples, see the *Theory into practice* chapter of your *Talend Studio User Guide*. 
2.1. Getting started with a basic Job

This section provides a continuous example that will help you create, add components to, configure, and execute a simple Job. This Job will be named A_Basic_Job and will read a text file, display its content on the Run console, and then write the data into another text file.

2.1.1. Creating a Job

To create the example Job described in this section, proceed as follows:

1. In the Repository tree view of the Integration perspective, right-click the Job Designs node and select Create Job from the contextual menu.

   The [New Job] wizard opens to help you define the main properties of the new Job.

   ![New Job wizard](image)

   Fill the Job properties as shown in the previous screenshot.

   The fields correspond to the following properties:

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>the name of the new Job. Note that a message comes up if you enter prohibited characters.</td>
</tr>
<tr>
<td>Purpose</td>
<td>Job purpose or any useful information regarding the Job use.</td>
</tr>
<tr>
<td>Description</td>
<td>Job description containing any information that helps you describe what the Job does and how it does it.</td>
</tr>
<tr>
<td>Author</td>
<td>a read-only field that shows by default the current user login.</td>
</tr>
</tbody>
</table>
### Field Description

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locker</td>
<td>A read-only field that shows by default the login of the user who owns the lock on the current Job. This field is empty when you are creating a Job and has data only when you are editing the properties of an existing Job.</td>
</tr>
<tr>
<td>Version</td>
<td>A read-only field. You can manually increment the version using the M and m buttons.</td>
</tr>
<tr>
<td>Status</td>
<td>A list to select from the status of the Job you are creating.</td>
</tr>
<tr>
<td>Path</td>
<td>A list to select from the folder in which the Job will be created.</td>
</tr>
</tbody>
</table>

3. An empty design workspace opens up showing the name of the Job as a tab label.

The Job you created is now listed under the **Job Designs** node in the **Repository** tree view.

You can open one or more of the created Jobs by simply double-clicking the Job label in the **Repository** tree view.

Related topics:

- Classify the Jobs you created by creating folders. For more information, see your *Talend Studio* User Guide.
- Create a data integration Job. For more information, see your *Talend Studio* User Guide.
- Customize the workspace. For more information, see your *Talend Studio* User Guide.

### 2.1.2. Adding components to the Job

Now that the Job is created, components have to be added to the design workspace, a **tFileInputDelimited**, a **tLogRow**, and a **tFileOutputDelimited** in this example.

There are several ways to add a component onto the design workspace. You can:

- find your component on the **Palette** by typing the search keyword(s) in the search field of the **Palette** and drop it onto the design workspace.
- add a component by directly typing your search keyword(s) on the design workspace.
- add an output component by dragging from an input component already existing on the design workspace.
- drag and drop a centralized metadata item from the **Metadata** node onto the design workspace, and then select the component of interest from the **Components** dialog box.
This section describes the first three methods. For details about how to drop a component from the Metadata node, see your Talend Studio User Guide.

### 2.1.2.1. Dropping the first component from the Palette

The first component of this example will be added from the Palette. This component defines the first task executed by the Job. In this example, as you first want to read a text file, you will use the `tFileInputDelimited` component.

For more information regarding components and their functions, see Talend Open Studio Components Reference Guide.

To drop a component from the Palette, proceed as follows:

1. Enter the search keyword(s) in the search field of the Palette and press Enter to validate your search.

   The keyword(s) can be the partial or full name of the component, or a phrase describing its functionality if you don’t know its name, for example, `tfileinputdelimited`, `fileinput`, or `read file row by row`.

   To use a descriptive phrase as keywords for a fuzzy search, make sure the Also search from Help when performing a component searching check box is selected on the Preferences > Palette Settings view. For more information, see your Talend Studio User Guide.

2. Select the component you want to use and click on the design workspace where you want to drop the component.

Each newly-added component is shown in a blue box to show that it as an individual Subjob.
2.1.2.2. Adding the second component by typing on the design workspace

The second component of our Job will be added by typing its name directly on the workspace, instead of dropping it from the Palette or from the Metadata node.

**Prerequisite**: Make sure you have selected the Enable Component Creation Assistant check box in the Studio preferences. For more information, see your Talend Studio User Guide.

To add a component directly on the workspace, proceed as follows:

1. Click where you want to add the component on the design workspace, and type your keywords, which can be the full or partial name of the component, or a phrase describing its functionality if you don't know its name. In our example, start typing `tlog`.

   To use a descriptive phrase as keywords for a fuzzy search, make sure the Also search from Help when performing a component searching check box is selected on the Preferences > Palette Settings view. For more information, see your Talend Studio User Guide.

A list box appears below the text field displaying all the matching components in alphabetical order.

2. Double-click the desired component to add it on the workspace, `tLogRow` in our example.

2.1.2.3. Adding an output component by dragging from an input one

Now you will add the third component, a `tFileOutputDelimited`, to write the data read from the source file into another text file. We will add the component by dragging from the `tLogRow` component, which serves as an input component to the new one to be added.

1. Click the `tLogRow` component to show the `o` icon docked to it.

2. Drag and drop the `o` icon where you want to add a new component.

   A text field and a component list appear. The component list shows all the components that can be connected with the input component.
3. To narrow the search, type in the text field the name of the component you want to add or part of it, or a phrase describing the component’s functionality if you don’t know its name, and then double-click the component of interest, **tFileOutputDelimited** in this example, on the component list to add it onto the design workspace. The new component is automatically connected with the input component **tLogRow**, using a Row > Main connection.

   ![Diagram showing component connections]

   To use a descriptive phrase as keywords for a fuzzy search, make sure the Also search from Help when performing a component searching check box is selected on the Preferences > Palette Settings view. For more information, see your Talend Studio User Guide.

2.1.3. Connecting the components together

Now that the components have been added on the workspace, they have to be connected together. Components connected together form a subjob. Jobs are composed of one or several subjobs carrying out various processes.

In this example, as the **tLogRow** and **tFileOutputDelimited** components are already connected, you only need to connect the **tFileInputDelimited** to the **tLogRow** component.
To connect the components together, proceed as follows:

1. Right-click the source component, **tFileInputDelimited** in this example.

2. In the contextual menu that opens, select the type of connection you want to use to link the components, **Row > Main** in this example.

3. Click the target component to create the link, **tLogRow** in this example.

Note that a black crossed circle is displayed if the target component is not compatible with the link.

According to the nature and the role of the components you want to link together, several types of link are available. Only the authorized connections are listed in the contextual menu.

### 2.1.4. Configuring the components

Now that the components are linked, their properties should be defined.

**Configuring the tFileInputDelimited component**

1. Double-click the **tFileInputDelimited** component to open its **Basic settings** view.
2. Click the [...] button next to the **File Name/Stream** field.

3. Browse your system or enter the path to the input file, *customers.txt* in this example.

4. In the **Header** field, enter *1*.

5. Click the [...] button next to **Edit schema**.

6. In the Schema Editor that opens, click three times the [+] button to add three columns.

7. Name the three columns *id*, *CustomerName* and *CustomerAddress* respectively and click **OK** to close the editor.

![Schema of tFileInputDelimited_1](image)

8. In the pop-up that opens, click **OK** accept the propagation of the changes.

This allows you to copy the schema you created to the next component, **tLogRow** in this example.

![Propagate](image)

**Configuring the tLogRow component**

1. Double-click the **tLogRow** component to open its **Basic settings** view.

2. In the **Mode** area, select **Table (print values in cells of a table)**.

   By doing so, the contents of the *customers.txt* file will be printed in a table and therefore more readable.
Executing the Job

Configuring the tFileOutputDelimited component

1. Double-click the tFileOutputDelimited component to open its Basic settings view.

2. Click the [...] button next to the File Name field.

3. Browse your system or enter the path to the output file, customers.csv in this example.

4. Select the Include Header check box.

5. If needed, click the Sync columns button to retrieve the schema from the input component.

2.1.5. Executing the Job

Now that components are configured, the Job can be executed.

To do so, proceed as follows:

1. Press Ctrl+S to save the Job.

2. Go to Run tab, and click on Run to execute the Job.

The file is read row by row and the extracted fields are displayed on the Run console and written to the specified output file.
2.2. Use cases

Let's get started with Job designs in Talend Studio by following a few basic end-to-end examples.

### 2.2.1. Updating data in a database table

This example describes a two-component Job that updates data in a MySQL table according to that in a delimited file.

**Dropping and link components**

1. Drop `tFileInputDelimited` and `tMysqlOutput` from the Palette onto the design workspace.
2. Connect the two components together using a Row Main link.
Configuring the input component

1. Double-click `tFileInputDelimited` to display its **Basic settings** view and define the component properties.

2. From the **Property Type** list, select **Repository** if you have already stored the metadata of the delimited file in the **Metadata** node in the **Repository** tree view. Otherwise, select **Built-In** to define manually the metadata of the delimited file.

   For more information about storing metadata, see **Talend Studio User Guide**.

3. In the **File Name** field, click the three-dot button and browse to the source delimited file that contains the modifications to propagate in the MySQL table.

   In this example, we use the `customer_update` file that holds four columns: `id`, `CustomerName`, `CustomerAddress` and `idState`. Some of the data in these four columns is different from that in the MySQL table.

   
<table>
<thead>
<tr>
<th>id;CustomerName;CustomerAddress;idState</th>
</tr>
</thead>
<tbody>
<tr>
<td>858;Froggy's Gourmet Catering;1831 Beverly Place #9-1D;4</td>
</tr>
<tr>
<td>859;Dependable Plumbing and Sewer;1550 Ridge Rd.;25</td>
</tr>
<tr>
<td>860;Lickmen Restoration;1235 Easton Rd.;40</td>
</tr>
<tr>
<td>861;Acturial Enterprises Ltd.;3148 Cottonwood Ct.;18</td>
</tr>
<tr>
<td>862;Rythmics Ltd.;857 Woodbine Rd.;30</td>
</tr>
<tr>
<td>863;Acturial Enterprises Ltd.;1482 Concorde Circle;48</td>
</tr>
<tr>
<td>864;Crossbars Car Wash;218 Oakridge Ave.;39</td>
</tr>
<tr>
<td>865;Meonits &amp; Mogogni Inc.; 616 Cobblestone Cir.;17</td>
</tr>
<tr>
<td>866;Foy Aviation;2220 Grant Blvd.;50</td>
</tr>
<tr>
<td>867;Ebert Music Center;12 Broadview Lane;29</td>
</tr>
<tr>
<td>868;janice Mann Accounting Service;1660 Park Ave.;9</td>
</tr>
<tr>
<td>869;Johnson, Erico &amp; Co CPA's;2922 Twin Oaks Drive;40</td>
</tr>
<tr>
<td>870;Corbins;Rodriguez, &amp; Savocchi;115 Pleasant Ave.;18</td>
</tr>
<tr>
<td>871;Nina's Snow Plowing;3385 University Ave.;20</td>
</tr>
<tr>
<td>872;Darcy Frame and Matting Service;1101 Deerfield Place;47</td>
</tr>
<tr>
<td>873;Marks, Kaplan and Jones Ltd.;1949 Cloverdale Rd.;9</td>
</tr>
</tbody>
</table>

4. Define the row and field separators used in the source file in the corresponding fields.

5. If needed, set **Header**, **Footer** and **Limit**.

   In this example, **Header** is set to 1 since the first row holds the names of columns, therefore it should be ignored. Also, the number of processed lines is limited to 2000.

6. Click the [...] button next to **Edit Schema** to open a dialog box where you can describe the data structure of the source delimited file that you want to pass to the component that follows.
7. Select the **Key** check box(es) next to the column name(s) you want to define as key column(s).

![Table](image)

- It is necessary to define at least one column as a key column for the Job to be executed correctly. Otherwise, the Job is automatically interrupted and an error message displays on the console.

### Configuring the output component

1. In the design workspace, double-click **tMysqlOutput** to open its **Basic settings** view where you can define its properties.

2. Click **Sync columns** to retrieve the schema of the preceding component. If needed, click the three-dot button next to **Edit schema** to open a dialog box where you can check the retrieved schema.

3. From the **Property Type** list, select **Repository** if you have already stored the connection metadata in the **Metadata** node in the **Repository** tree view. Otherwise, select **Built-In** to define manually the connection information in the corresponding fields: **Host**, **Port**, **Database**, **Username** and **Password**.

   For more information about storing metadata, see **Talend Studio User Guide**.

4. In the **Table** field, enter the name of the table to update.

5. From the **Action on table** list, select the operation you want to perform, **Default** in this example since the table already exists.

6. From the **Action on data** list, select the operation you want to perform on the data, **Update** in this example.

### Saving and executing the Job

1. Press **Ctrl+S** to save your Job.

2. Press **F6** or click **Run** on the **Run** tab to execute the Job.
2.2.2. Mapping data using a filter and a simple explicit join

The Job described below aims at reading data from a csv file, looking up at a reference file, and then extracting data from these two files based on a defined filter to an output file and reject files.

Adding and linking the components

1. Add two `tFileInputDelimited` components, a `tMap` and three `tFileOutputDelimited` components onto the design workspace.

2. Rename the two `tFileInputDelimited` components as `Cars` and `Owners`, either by double-clicking the label in the design workspace or via the View tab of the Component view.

3. Connect the two input components to `tMap` using Row > Main connections and label the connections as `Cars_data` and `Owners_data` respectively.

4. Connect `tMap` to the three output components using Row > New Output (Main) connections and name the output connections as `Insured`, `Reject_NoInsur` and `Reject_OwnerID` respectively.

The MySQL table `customers` has been modified according to the delimited file.
Configuring the input components

1. Double-click the `tFileInputDelimited` component labelled `Cars` to display its Basic settings view.

![Basic settings view of tFileInputDelimited component](image)

2. Select Repository from the Property type list and select the component's schema, `cars` in this scenario, from the [Repository Content] dialog box. The rest fields are automatically filled.

   ![Repository Content dialog box](image)

   This scenario assumes that the metadata of the input files is stored in the Metadata node of the Repository tree view for easy retrieval. For further information regarding metadata creation in the Repository, see Talend Studio User Guide.

If you do not have the metadata of your input files centralized in the Repository, you need to set the property type to Built-In and specify file path and define the file schema manually. Below is an abstract of the input file `cars.scv`:

```
ID_Owner;Registration;Make;Color;ID_Reseller
1;WZG 555;Ford;red;22
2;HYZ 472;Lexus;red;39
3;VYZ 862;Lexus;blue;21
4;ZYZ 350;Audi;red;31
5;EDZ 99;Audi;green;62
6;ZZX 845;Citroen;black;75
7;PBS 410;Renault;grey;11
8;JFO 929;Citroen;white;86
9;DPG 217;Lexus;black;13
```
3. Double-click the component labelled Owners and repeat the setting operation. Select the appropriate metadata entry, owners in this scenario.

If you do not have the metadata of your input files centralized in the Repository, you need to set the property type to Built-In and specify file path and define the file schema manually. Below is an abstract of the reference input file owners.csv:

<table>
<thead>
<tr>
<th>ID_Owner</th>
<th>Name</th>
<th>ID_Insurance</th>
<th>Children_Nr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>George EISENHOWER</td>
<td>108</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>James LINCOLN</td>
<td>35</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>William TAFT</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>Harry WILSON</td>
<td>134</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>Woodrow HOOVER</td>
<td>45</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>Chester TAYLOR</td>
<td>148</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>John REAGAN</td>
<td>31</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>Dwight POLK</td>
<td>105</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>William PIERCE</td>
<td>177</td>
<td>2</td>
</tr>
</tbody>
</table>

Configuring the mapping component

1. Double-click the tMap component to open the Map Editor.

   Note that the input area is already filled with the defined input tables and that the top table is the main input table, and the respective row connection labels are displayed on the top bar of the table.

2. Create a join between the two tables on the ID_Owner column by simply dropping the ID_Owner column from the Cars_data table onto the ID_Owner column in the Owners_data table.

3. Define this join as an inner join by clicking the tMap settings button, clicking in the Value field for Join Model, clicking the small button that appears in the field, and selecting Inner Join from the [Options] dialog box.

4. Drag all the columns of the Cars_data table to the Insured table.
5. Drag the ID_Owner, Registration, and ID_Reseller columns of the Cars_data table and the Name column of the Owners_data table to the Reject_NoInsur table.

6. Drag all the columns of the Cars_data table to the Reject_OwnerID table.

   For more information regarding data mapping, see Talend Studio User Guide.

7. Click the plus arrow button at the top of the Insured table to add a filter row.

   Drag the ID_Insurance column of the Owners_data table to the filter condition area and enter the formula meaning 'not undefined': Owners_data.ID_Insurance != null.

   With this filter, the Insured table will gather all the records that include an insurance ID.

8. Click the tMap settings button at the top of the Reject_NoInsur table and set Catch output reject to true to define the table as a standard reject output flow to gather the records that do not include an insurance ID.

9. Click the tMap settings button at the top of the Reject_OwnerID table and set Catch lookup inner join reject to true so that this output table will gather the records from the Cars_data flow with missing or unmatched owner IDs.
Click **OK** to validate the mappings and close the **Map Editor**.

**Configuring the output components**

1. Double-click each of the output components, one after the other, to define their properties. If you want a new file to be created, browse to the destination output folder, and type in a file name including the extension.

2. Select the **Include header** check box to reuse the column labels from the schema as header row in the output file.

**Executing the Job**

1. Press **Ctrl + S** to save your Job.

2. Press **F6** to run the Job.

   The output files are created, which contain the relevant data as defined.
Mapping data using a filter and a simple explicit join

<table>
<thead>
<tr>
<th>ID_OWNER;Registration;ID_Reseller;Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:DPG 217;13:William PIERCE</td>
</tr>
<tr>
<td>16;ZZY 702;48:Rutherford HOOVER</td>
</tr>
<tr>
<td>26;ZZT 904;37:Millard WASHINGTON</td>
</tr>
<tr>
<td>34;RZI 397;97:Theodore WASHINGTON</td>
</tr>
<tr>
<td>38;GZT 196;43:Andrew ADAMS</td>
</tr>
<tr>
<td>44;ADL 67;59:Franklin MADISON</td>
</tr>
<tr>
<td>49;DCZ 760;52:Benjamin QUINCY</td>
</tr>
<tr>
<td>52;KMQ 503;25:Martin COOLIDGE</td>
</tr>
<tr>
<td>68;SNH 801;101:Richard MONROE</td>
</tr>
<tr>
<td>70;ZII 771;51:Ronald JEFFERSON</td>
</tr>
<tr>
<td>72;TEU 459;26:Lyndon EISENHOWER</td>
</tr>
</tbody>
</table>
# Appendix A. Glossary

When working with *Talend Studio* and in order to understand its functional mechanism, it is important to understand some basic vocabulary.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>component</td>
<td>A component is an executable part of a Job or Route used to connect to an external source or perform a specific data integration operation, no matter what data sources you are integrating: databases, applications, flat files, Web services, etc. A component can minimize the amount of hand-coding required to work on data from multiple, heterogeneous sources. Components are grouped in families according to their usage and displayed in the Palette of the Integration perspective of <em>Talend Studio</em>. For detailed information about components types and what they can be used for, see <em>Talend Open Studio Components Reference Guide</em>.</td>
</tr>
<tr>
<td>item</td>
<td>An item is the fundamental technical unit in a project. Items are grouped, according to their types, as: Job Design, Business model, Context, Code, Metadata, etc. One item can include other items. For example, the business models and the Jobs you design are items, metadata and routines you use inside your Jobs are items as well.</td>
</tr>
<tr>
<td>Job</td>
<td>A Job is a graphical design, of one or more components connected together, that allows you to set up and run dataflow management processes. It translates business needs into code, routines and programs. Jobs address all of the different sources and targets that you need for data integration processes and all other related processes.</td>
</tr>
<tr>
<td>metadata</td>
<td>Metadata is information that describes the characteristics of any data object, such as its name, type, location, author, date created, size, and so on, together with relationships with other data objects that the enterprise has to manage or that an IT tool may generate. Metadata can be created manually or automatically by a system.</td>
</tr>
<tr>
<td>project</td>
<td>Projects are structured collections of items and their associated metadata. All of the Jobs and business models you design are organized in Projects.</td>
</tr>
<tr>
<td>repository</td>
<td>A repository is the storage location <em>Talend Studio</em> uses to gather data related to all of the technical items that you use either to describe business models or to design Jobs.</td>
</tr>
<tr>
<td>workspace</td>
<td>A workspace is the directory where you store all your project folders. You need to have one workspace directory per connection (repository connection). <em>Talend Studio</em> enables you to connect to different workspace directories, if you do not want to use the default one.</td>
</tr>
</tbody>
</table>