Talend Open Studio for MDM

Getting Started Guide

6.0.0

Adapted for v6.0.0. Supersedes previous releases.

Publication date: July 2, 2015

Copyleft

Notices

Table of Contents

Preface

1. General information

1.1. Purpose

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

Information presented in this document applies to *Talend Open Studio for MDM* **6.0.0**.

1.2. Audience

This guide is for users and administrators of *Talend Open Studio for MDM*.

💡 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,
- 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.

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

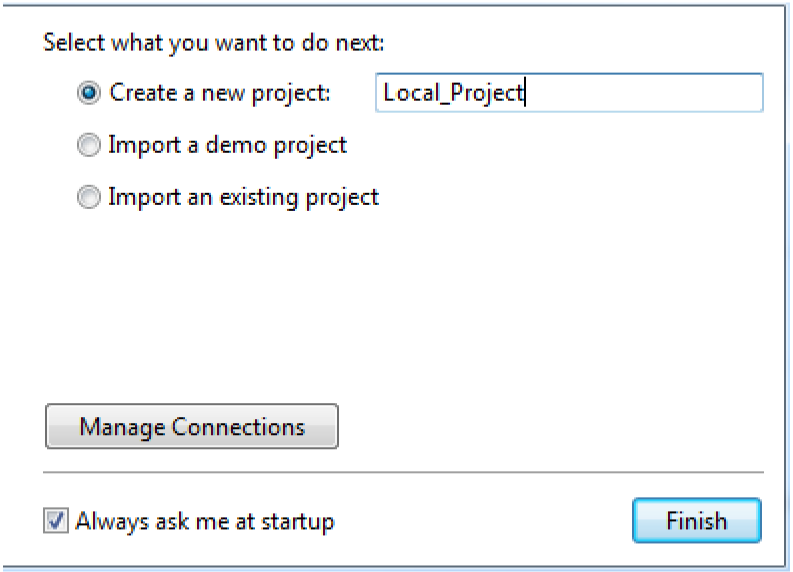
To open *Talend Studio* for the first time, complete the following:

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 what you want to do next:

● Create a new project: Local_Project

○ Import a demo project

○ Import an existing project

☑ Always ask me at startup

   This login window appears only when the Studio is started for the first time. When you launch the Studio again, the normal login window opens, which provides one more option, a connection list box, for subscription-based users to select a repository connection when launching the Studio.

   If you plan to use the same repository connection and / or project at your next Studio launch, you can skip the login window to speed up Studio launch by clearing the **Always ask me at startup** check box. Then, if you want to see the login window again, go to the menu **Window** > **Preferences** to open the **[Preferences]** window, select **Talend**, and select the **Always show project dialog at startup** check box.

   - 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 the demo project*.
   - Select **Import an existing project** and click **Finish** to import an existing projects. For more information, see *How to import projects*.

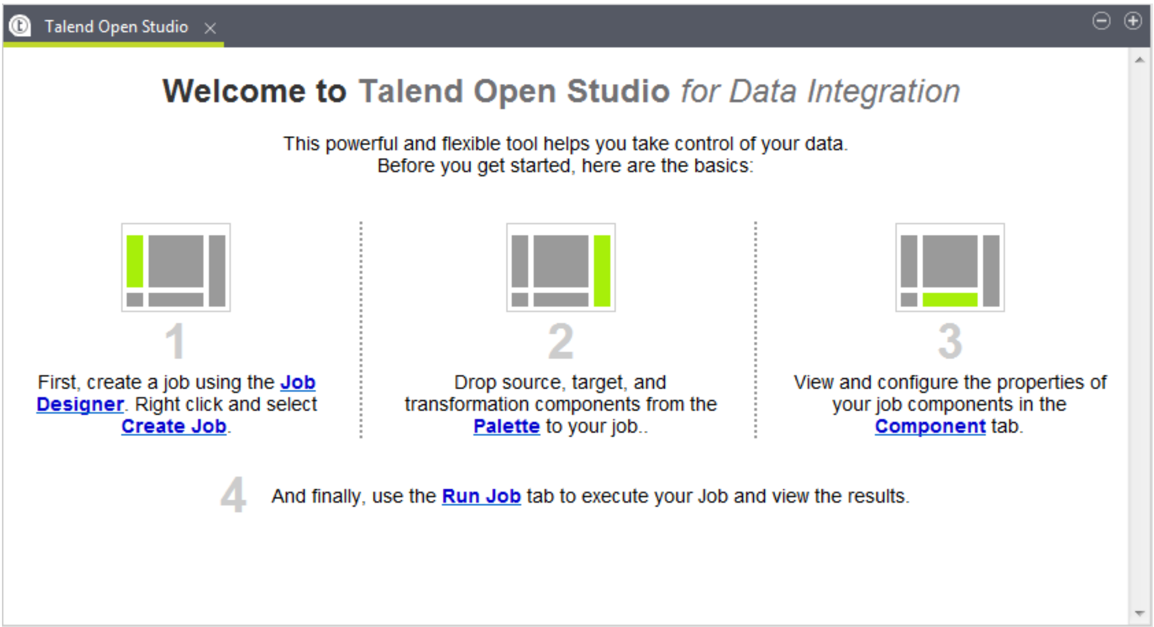
- If you want to modify the default repository connection, click **Manage Connections** to set up your connection before setting up a project. For further information about connecting to a repository, see *How to access a Repository*.

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**.

The **[Welcome]** window opens. From this window you have direct links to Demo projects, user documentation, tutorials, **Talend** forum, **Talend** on-demand training and **Talend** latest news.

4. Click **Start now!** to open *Talend Studio* main window, which displays a welcome 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.

   For more information on how to open a project, see *How to open a project*.



Welcome to Talend Open Studio *for Data Integration*

This powerful and flexible tool helps you take control of your data.
Before you get started, here are the basics:

1 First, create a job using the **Job Designer**. Right click and select **Create Job**.

2 Drop source, target, and transformation components from the **Palette** to your job..

3 View and configure the properties of your job components in the **Component** tab.

4 And finally, use the **Run Job** tab to execute your Job and view the results.

5. 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*.

   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.

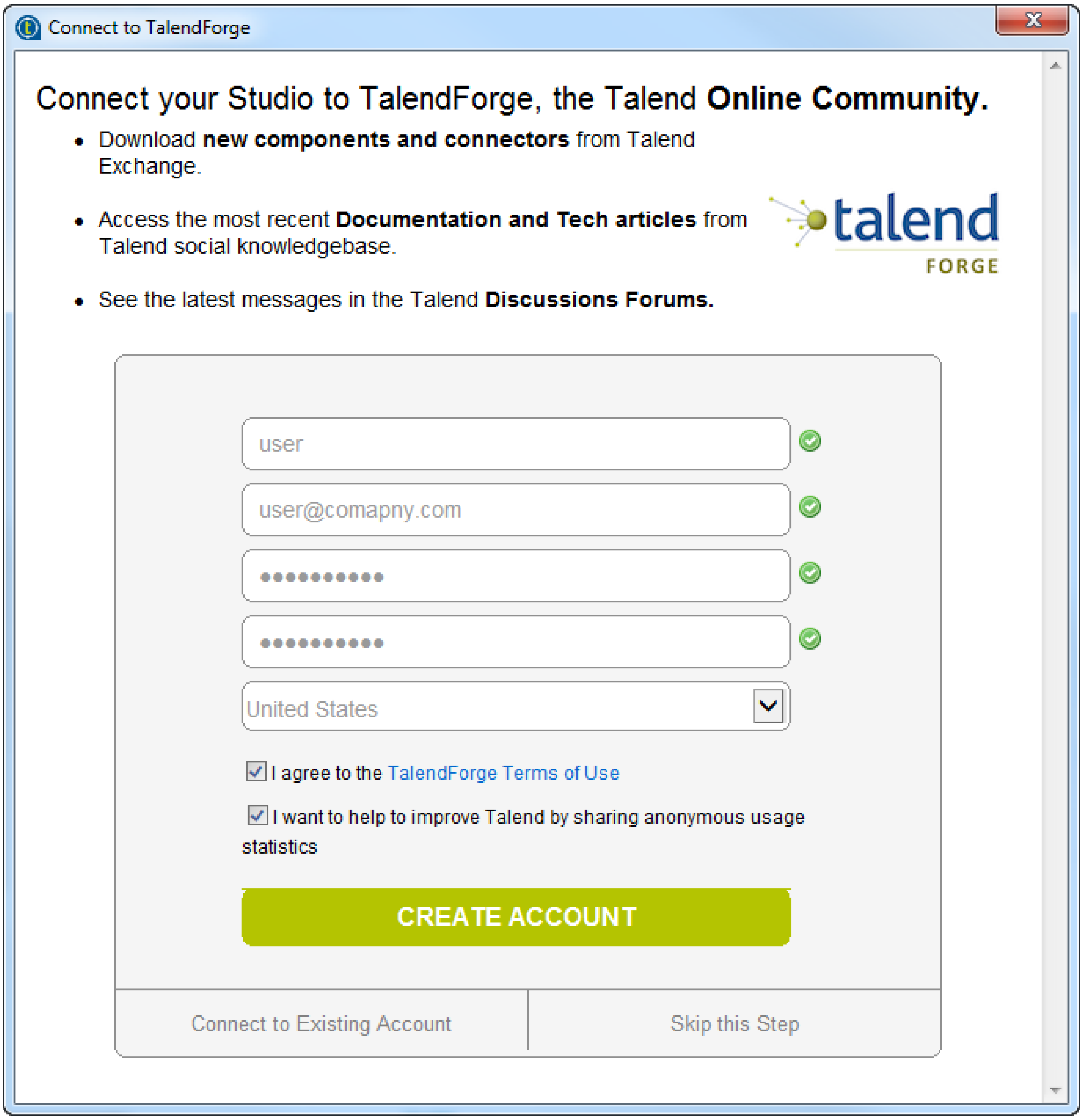
If you want to connect to the **Talend** Community later, click **Skip this Step** to continue launching the Studio without setting up a connection to the **Talend** Community.

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.



Connect to TalendForge

Connect your Studio to TalendForge, the Talend **Online Community.**

- Download **new components and connectors** from Talend Exchange.
- Access the most recent **Documentation and Tech articles** from Talend social knowledgebase.
- See the latest messages in the Talend **Discussions Forums.**

user

user@comapny.com

●●●●●●●●●●

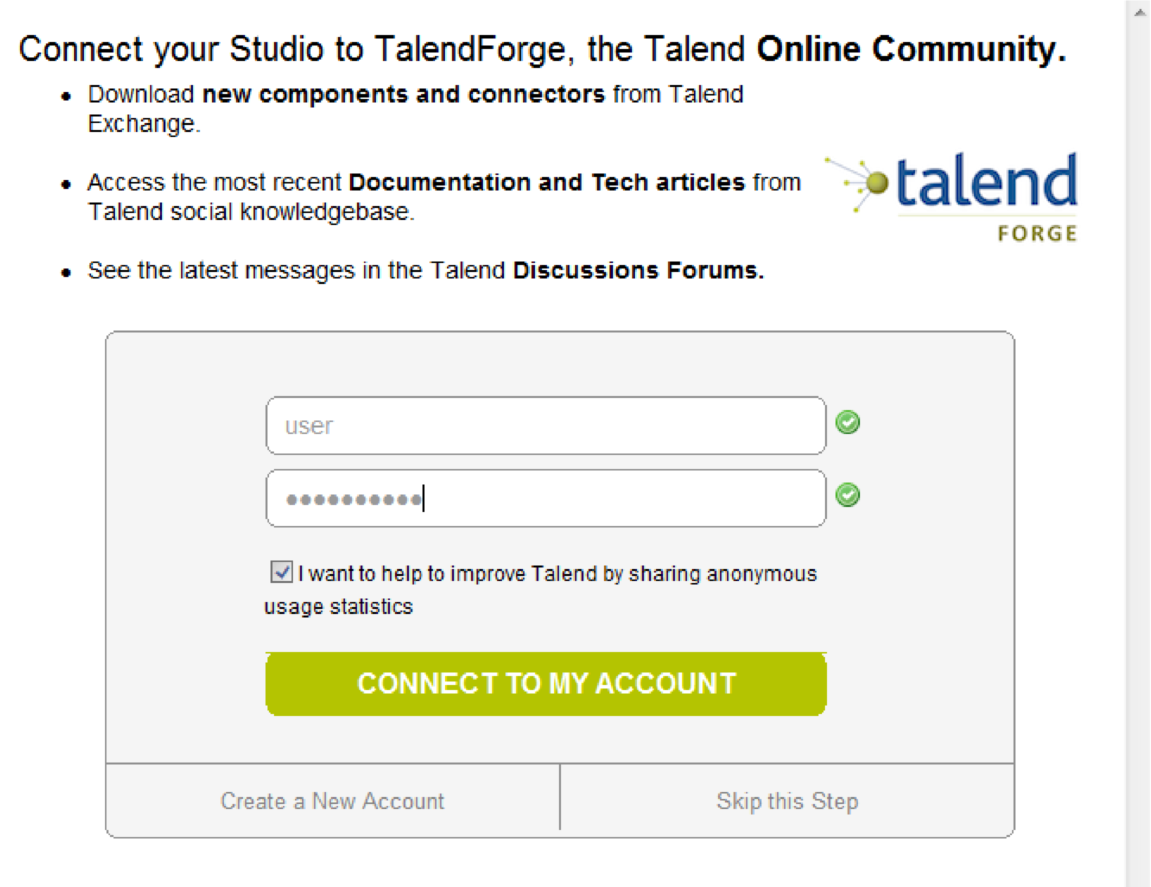
●●●●●●●●●●

United States

☑ I agree to the TalendForge Terms of Use

☑ I want to help to improve Talend by sharing anonymous usage statistics

If you already have created an account at 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.



Connect your Studio to TalendForge, the Talend **Online Community.**

- Download **new components and connectors** from Talend Exchange.
- Access the most recent **Documentation and Tech articles** from Talend social knowledgebase.
- See the latest messages in the Talend **Discussions Forums.**

☑ I want to help to improve Talend by sharing anonymous usage statistics

This page will not appear again when the Studio starts up once you successfully connect to the **Talend** Community. To show this page again, select **Talend** > **Exchange** from the **[Preferences]** dialog box, and click Sign In. For more information, see the section on setting *Talend Studio* preferences of your Studio User Guide.

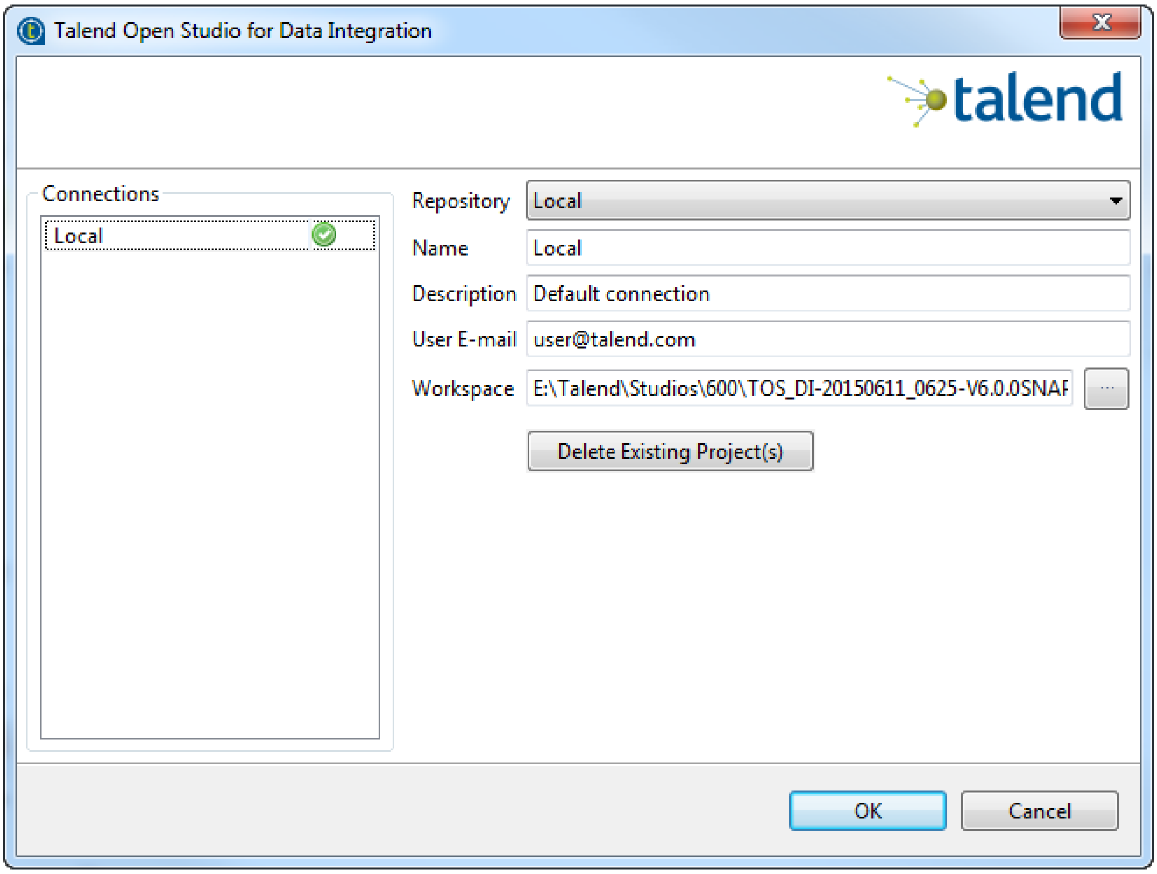
1.1.3. How to access a Repository

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.



Talend Open Studio for Data Integration

Connections

Local

Repository: Local

Name: Local

Description: Default connection

User E-mail: user@talend.com

Workspace: E:\Talend\Studios\600\TOS_DI-20150611_0625-V6.0.0SNAF

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.

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 *How to import projects*.

- importing the Demo project. For more information, see *How to import the demo project*.

1.2. Working with different workspace directories

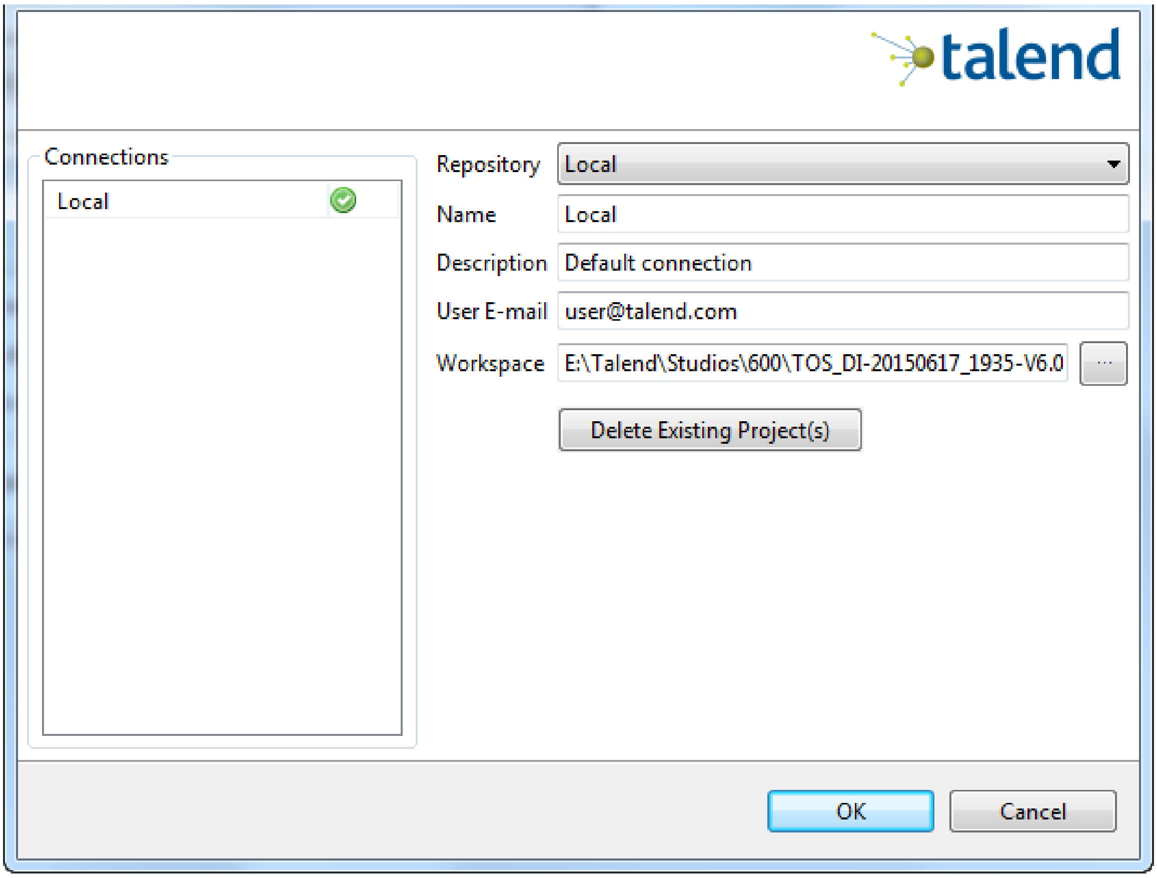
Talend Studio makes it possible to create many workspace directories and connect to a workspace different from the one you are currently working on, if necessary.

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

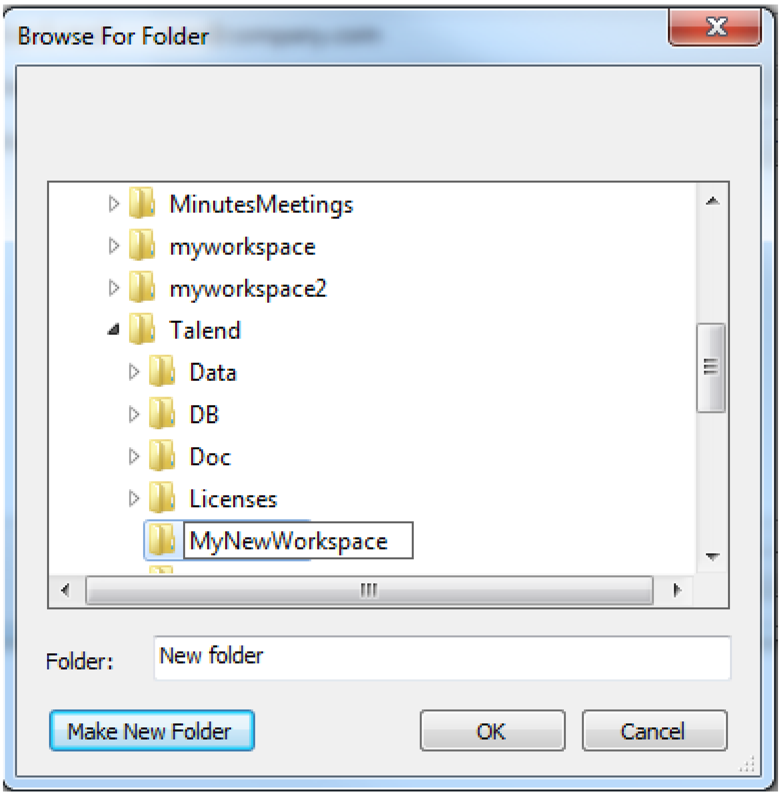
Talend Studio is delivered with a default workspace directory. However, you can create as many new directories as you want and store your project folders in them according to your preferences.

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.



Repository: Local
Name: Local
Description: Default connection
User E-mail: user@talend.com
Workspace: E:\Talend\Studios\600\TOS_DI-20150617_1935-V6.0

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 dialob 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.

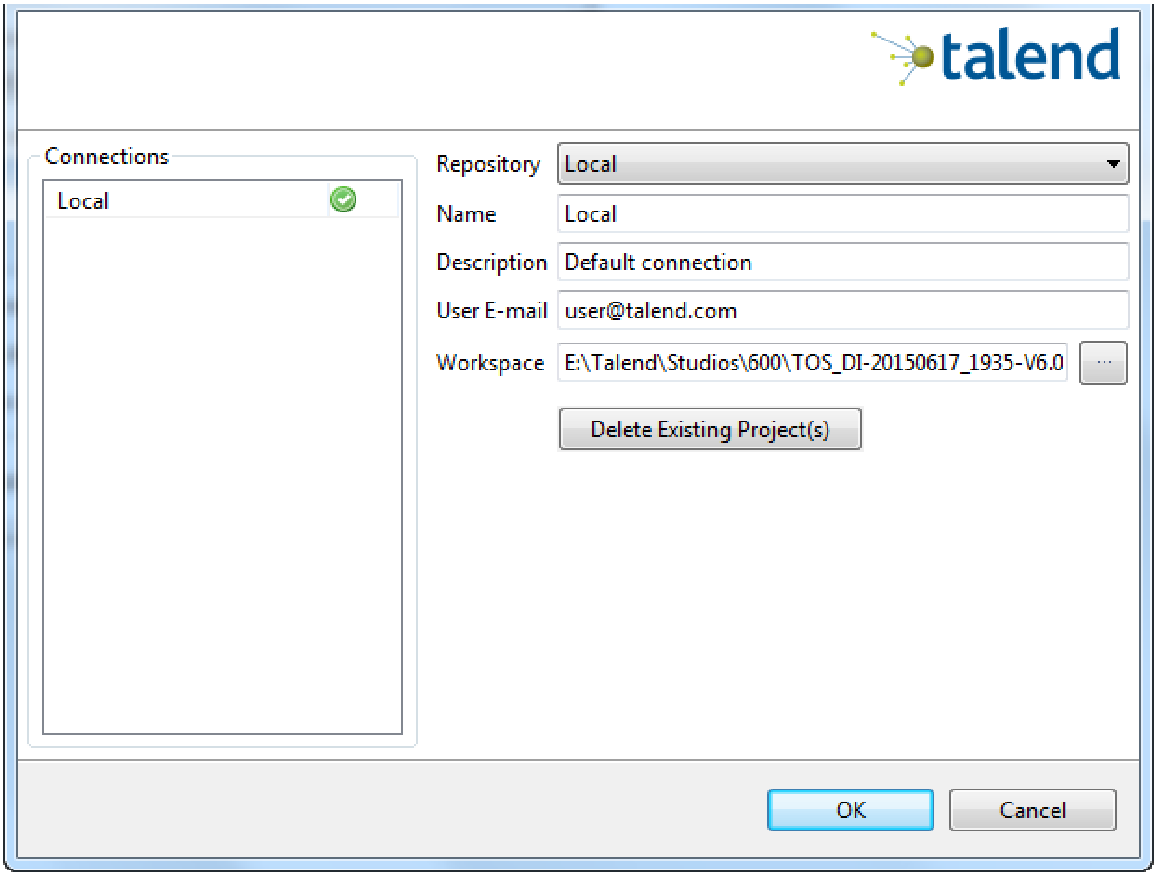
1.2.2. How to connect to a different workspace directory

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.



Repository: Local

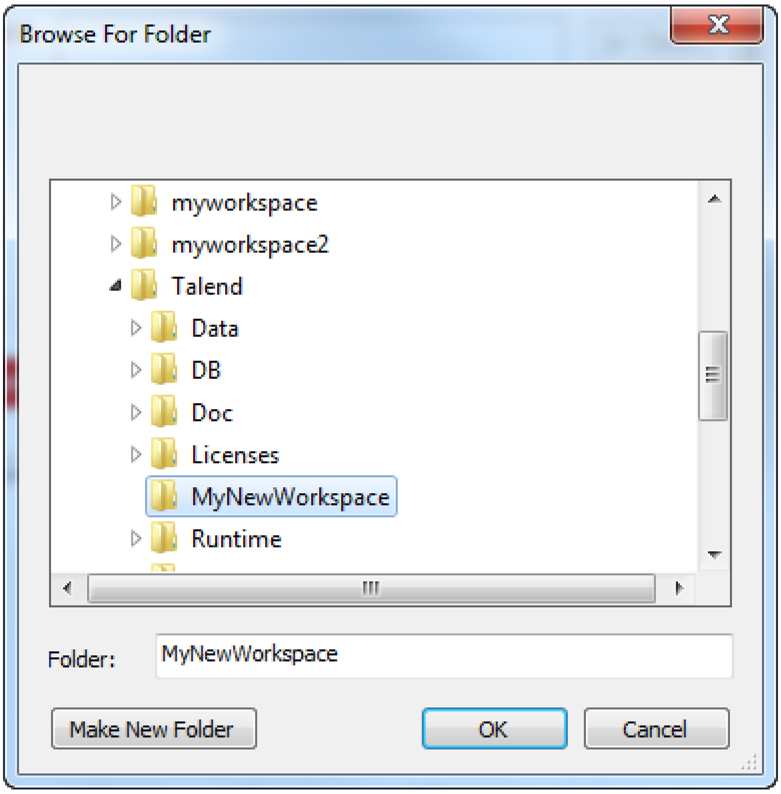
Name: Local

Description: Default connection

User E-mail: user@talend.com

Workspace: E:\Talend\Studios\600\TOS_DI-20150617_1935-V6.0

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.



Folder: MyNewWorkspace

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, etc. is the "project".

From the login window of the Studio, you can:

- create a local project.

  When you launch the Studio for the first time, there are no default projects listed. You need to create a project that will hold all data integration Jobs and business models you design in the current instance of the Studio.

  You can create as many projects as you need to store your data of different instances of your Studio.

  When creating a new project, a tree folder is automatically created in the workspace directory on your repository server. This will correspond to the **Repository** tree view displayed on the main window of the Studio.

  For more information, see *How to create a project*.

- import the Demo project to discover the features of *Talend Studio* based on samples of different ready-to-use Jobs. When you import the Demo project, it is automatically installed in the workspace directory of the current session of the Studio.

  For more information, see *How to import the demo project*.

- import projects you have already created with previous releases of *Talend Studio* into your current *Talend Studio* workspace directory.

  For more information, see *How to import projects*.

- open a project you created or imported in the Studio.

  For more information, see *How to open a project*.

- delete local projects that you already created or imported and that you do not need any longer.

  For more information, see *How to delete a project*.

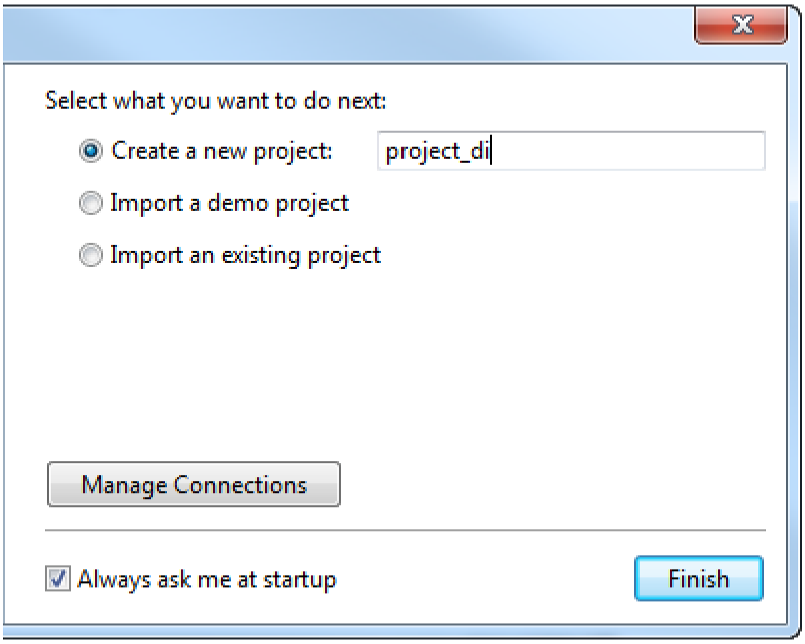
Once you launch *Talend Studio*, you can export the resources of one or more of the created projects in the current instance of the Studio. For more information, see *How to export a project*.

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.



Select what you want to do next:

● Create a new project: project_di

○ Import a demo project

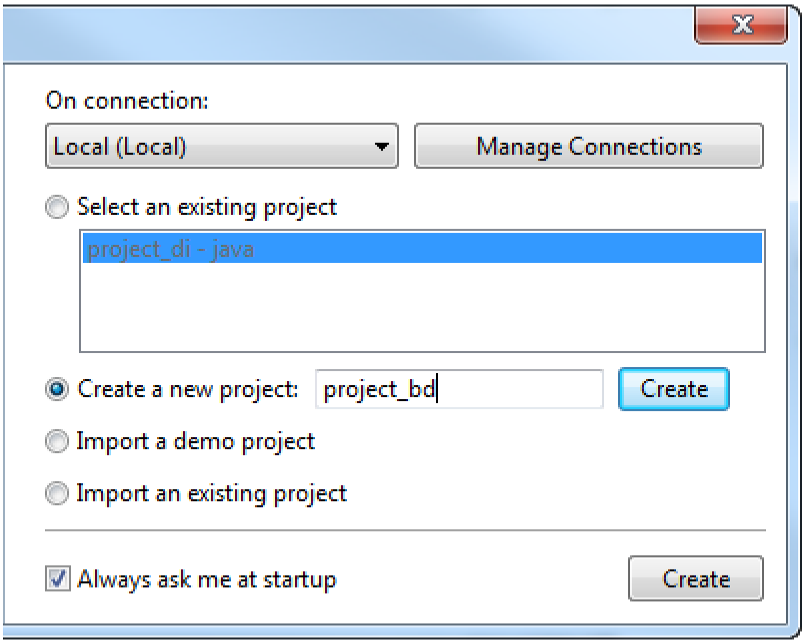
○ Import an existing project

☑ Always ask me at startup

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.



On connection: Local (Local)

○ Select an existing project

project_di - java

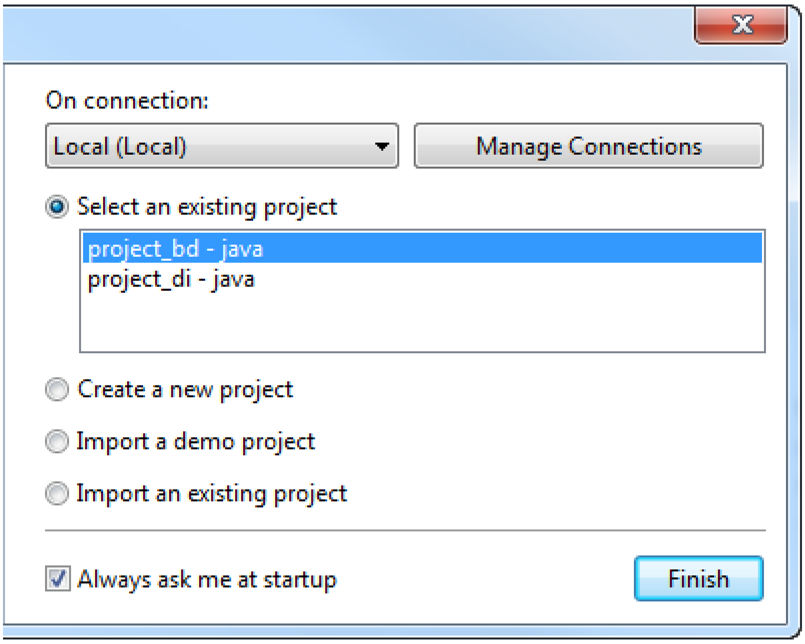
● Create a new project: project_bd

○ Import a demo project

○ Import an existing project

☑ Always ask me at startup

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



On connection:

Local (Local) ▾ | Manage Connections

● Select an existing project

project_bd - java
project_di - java

○ Create a new project

○ Import a demo project

○ Import an existing project

☑ Always ask me at startup

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 the 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.

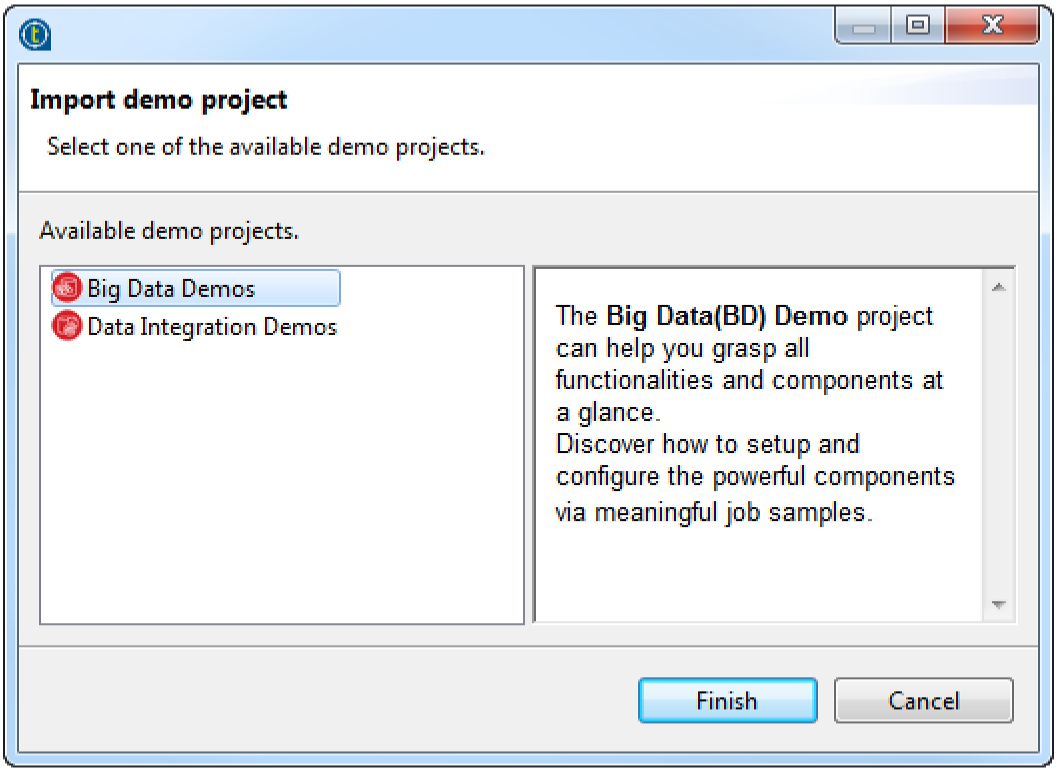
To import a demo project, proceed as follows:

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

Select one of the available demo projects.

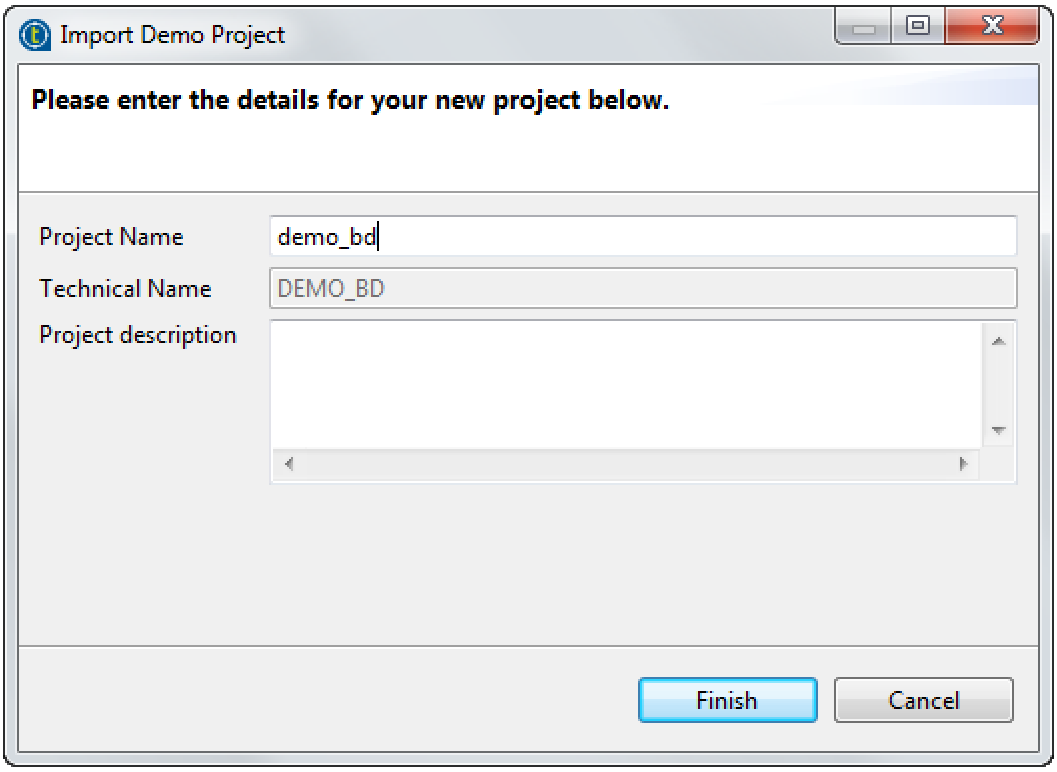
Available demo projects.

- Big Data Demos
- Data Integration Demos

The **Big Data(BD) Demo** project can help you grasp all functionalities and components at a glance. Discover how to setup and configure the powerful components via meaningful job samples.

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.



Import Demo Project

Please enter the details for your new project below.

Project Name: demo_bd

Technical Name: DEMO_BD

Project description:

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**.

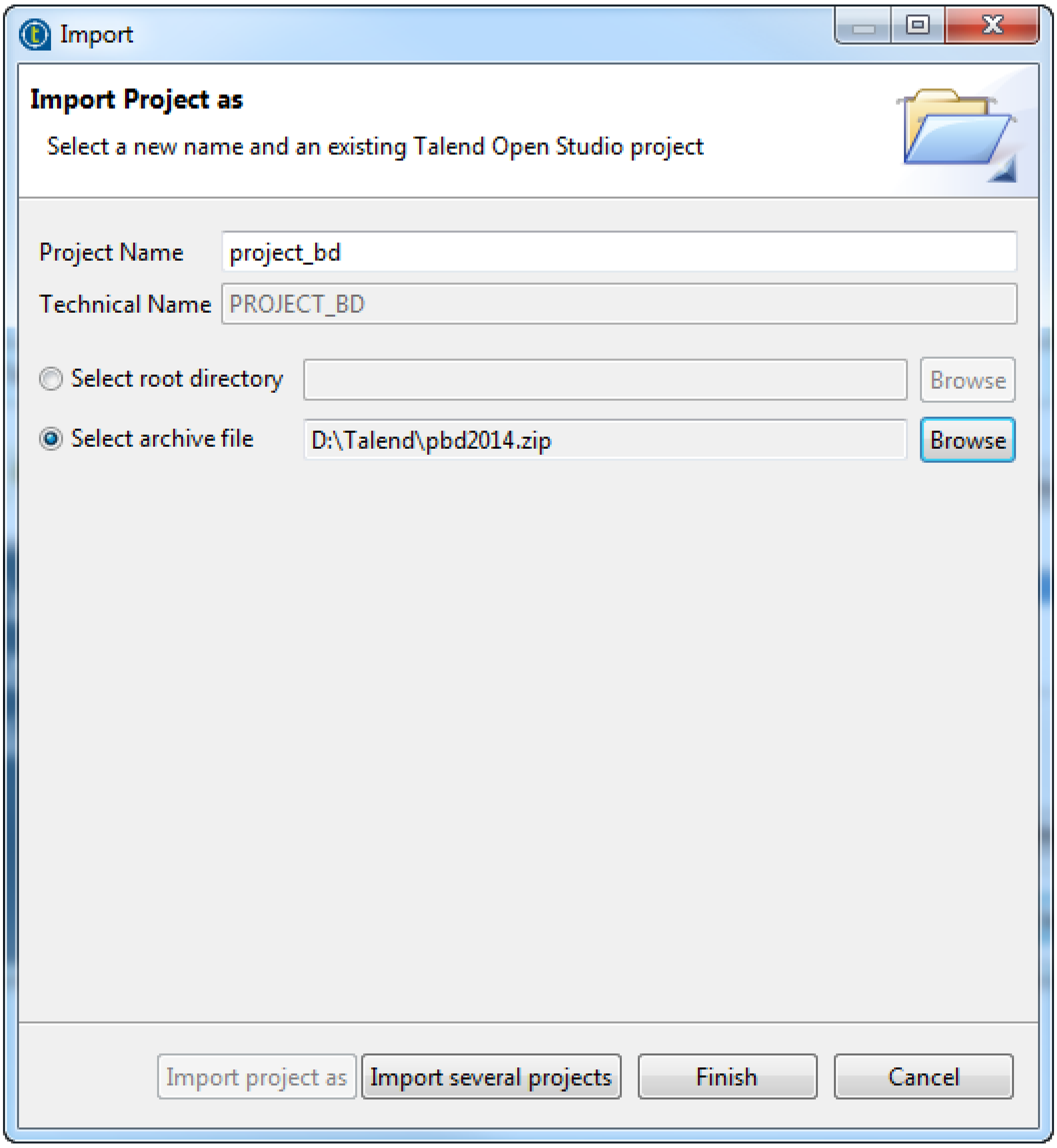
   The Job samples in the open demo project are automatically imported into your workspace directory and made available in the **Repository** tree view under the **Job Designs** folder.

1.3.3. How to import projects

In *Talend Studio*, you can import one or more projects you already created with previous releases of the Studio.

To import a single project, do the following:

1. From the Studio login window, select **Import an existing project** then click **Select** to open the **[Import]** wizard.



Import

Import Project as

Select a new name and an existing Talend Open Studio project

Project Name: project_bd

Technical Name: PROJECT_BD

○ Select root directory

● Select archive file: D:\Talend\pbd2014.zip

2. Click the **Import project as** button and enter a name for your new project in the **Project Name** field.

3. Click **Select root directory** or **Select archive file** depending on the source you want to import from.

4. Click **Browse...** to select the workspace directory/archive file of the specific project folder. By default, the workspace in selection is the current release's one. Browse up to reach the previous release workspace directory or the archive file containing the projects to import.

5. Click **Finish** to validate the operation and return to the login window.

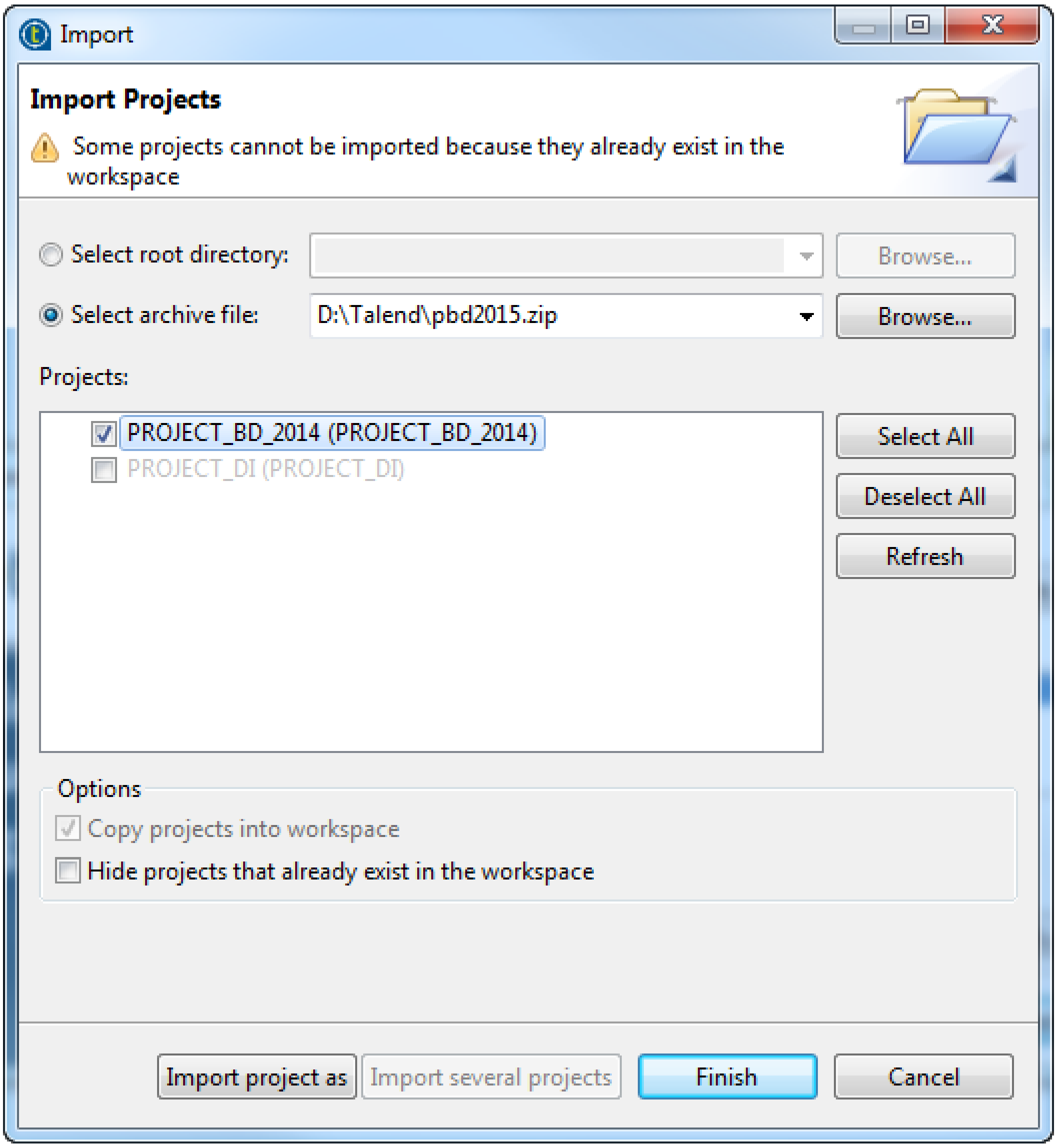
To import several projects simultaneously, do the following:

1. From the Studio login window, select **Import an existing project** then click **Select** to open the **[Import]** wizard.

2. Click **Import several projects**.

3. Click **Select root directory** or **Select archive file** depending on the source you want to import from.

4. Click **Browse...** to select the workspace directory/archive file of the specific project folder. By default, the workspace in selection is the current release's one. Browse up to reach the previous release workspace directory or the archive file containing the projects to import.



Import

Import Projects

⚠ Some projects cannot be imported because they already exist in the workspace

○ Select root directory:

● Select archive file: D:\Talend\pbd2015.zip

Projects:

☑ PROJECT_BD_2014 (PROJECT_BD_2014)

□ PROJECT_DI (PROJECT_DI)

Options

☑ Copy projects into workspace

□ Hide projects that already exist in the workspace

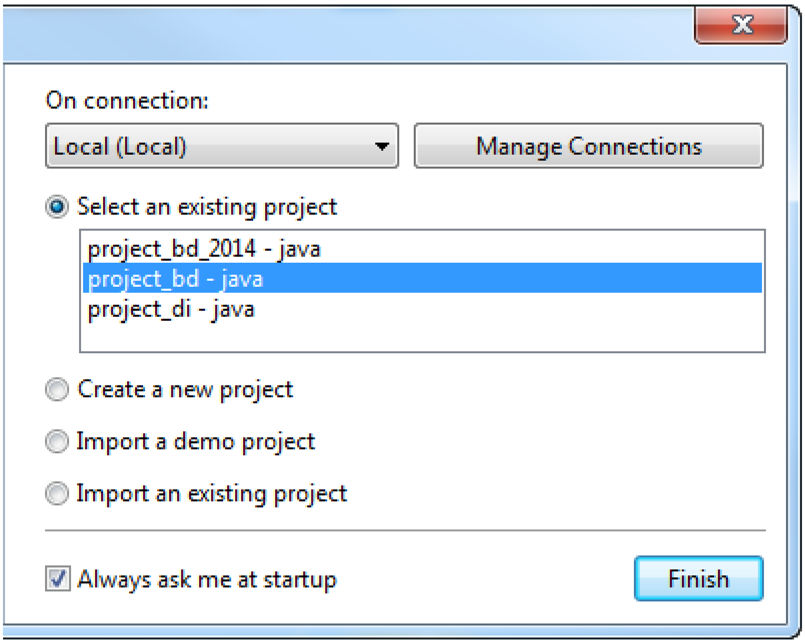
5. Select the **Copy projects into workspace** check box to make a copy of the imported project instead of moving it. This option is available only when you import several projects from a root directory.

If you want to remove the original project folders from the *Talend Studio* workspace directory you import from, clear this check box. But we strongly recommend you to keep it selected for backup purposes.

6. Select the **Hide projects that already exist in the workspace** check box to hide existing projects from the **Projects** list. This option is available only when you import several projects.

7. From the **Projects** list, select the projects to import and click **Finish** to validate the operation.

Upon successful project import, the names of the imported projects are displayed on the **Project** list of the login window.



On connection: Local (Local)

● Select an existing project

project_bd_2014 - java
project_bd - java
project_di - java

○ Create a new project

○ Import a demo project

○ Import an existing project

☑ Always ask me at startup

You can now select the imported project you want to open in *Talend Studio* and click **Finish** to launch the Studio.

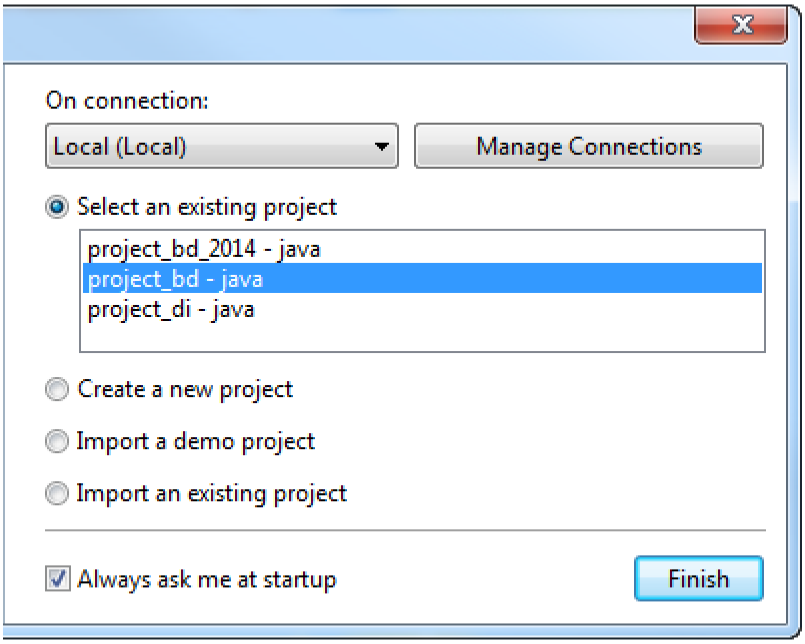
A generation initialization window might come up when launching the application. Wait until the initialization is complete.

1.3.4. 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**.



On connection:

Local (Local)

● Select an existing project

project_bd_2014 - java

project_bd - java

project_di - java

○ Create a new project

○ Import a demo project

○ Import an existing project

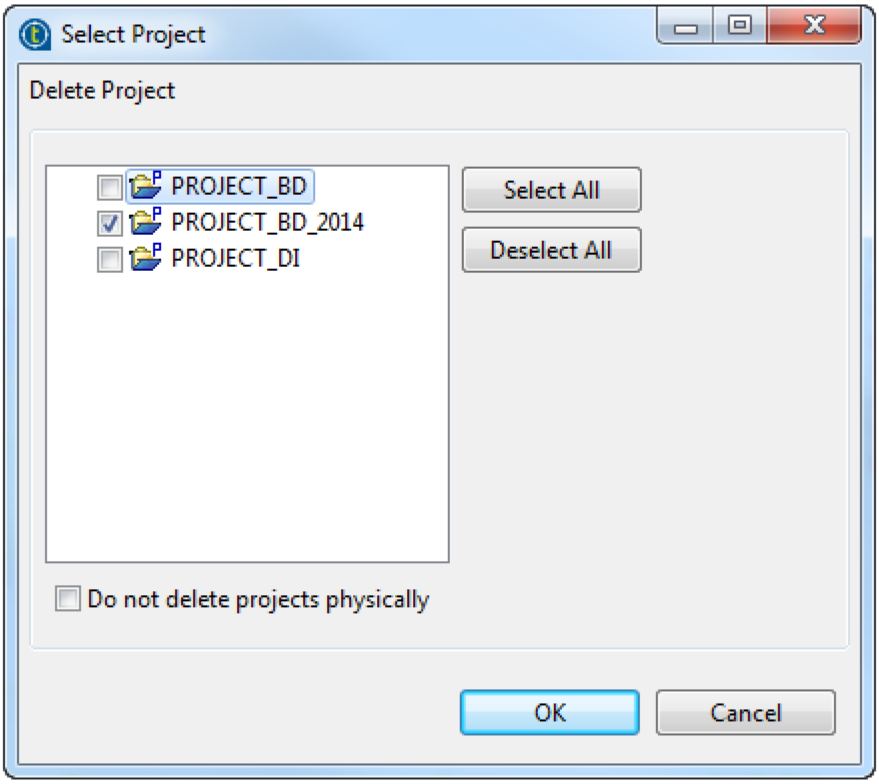
☑ Always ask me at startup

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.

1.3.5. How to delete a project

1. On the login screen, click **Manage Connections**, then on the dialog box that opens click **Delete Existing Project(s)** to open the **[Select Project]** dialog box.



Select Project

Delete Project

☐ PROJECT_BD

☑ PROJECT_BD_2014

☐ PROJECT_DI

☐ Do not delete projects physically

2. Select the check box(es) of the project(s) you want to delete.

3. Click **OK** to validate the deletion.

   The project list on the login window is refreshed accordingly.

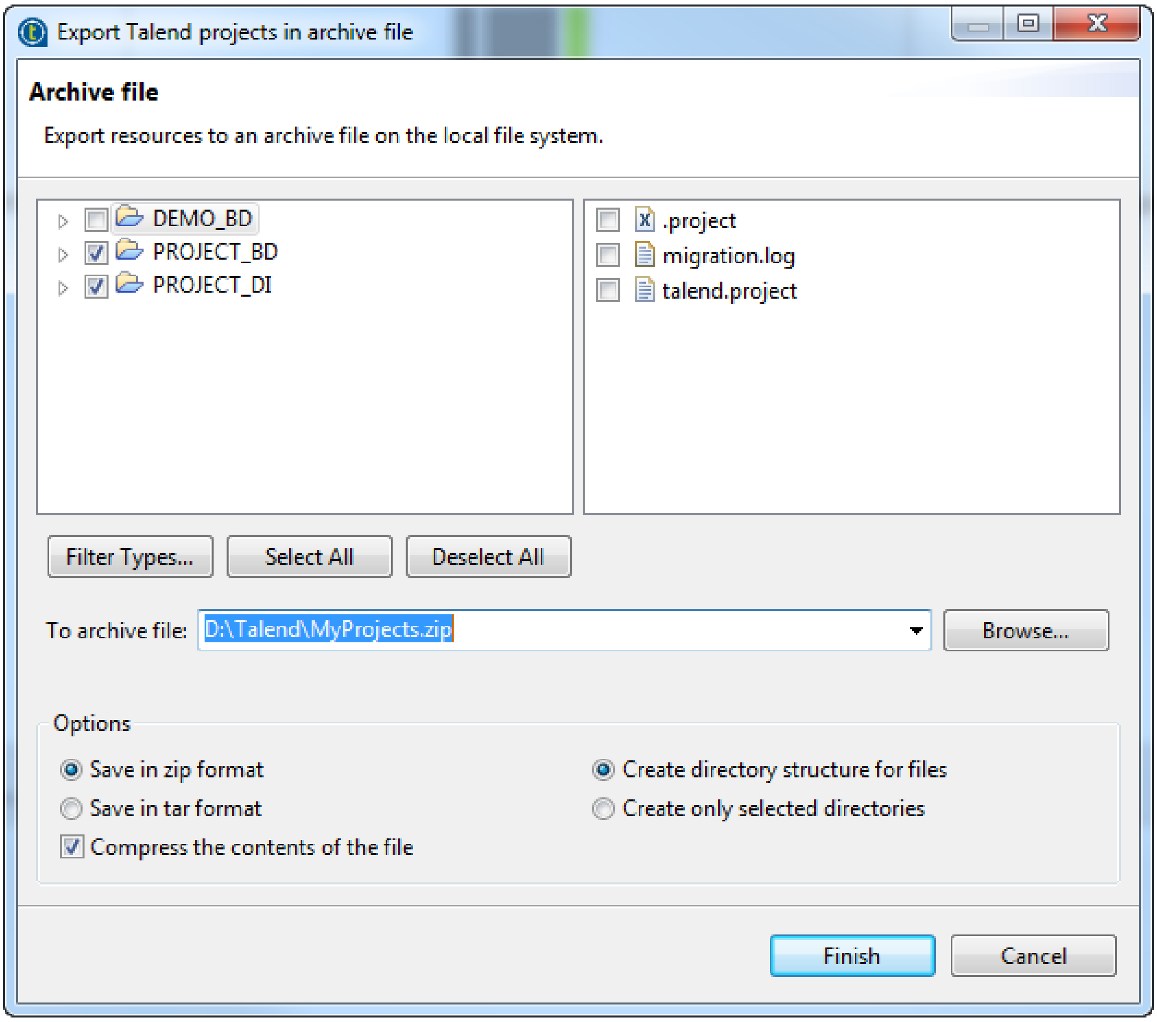
   *Be careful, this action is irreversible. When you click **OK**, there is no way to recuperate the deleted project(s).*

   If you select the **Do not delete projects physically** check box, you can delete the selected project(s) only from the project list and still have it/them in the *workspace* directory of *Talend Studio*. Thus, you can recuperate the deleted project(s) any time using the **Import existing project(s) as local** option on the **Project** list from the login window.

1.3.6. How to export a project

Talend Studio allows you to export projects created or imported in the current instance of *Talend Studio*.

1. On the toolbar of the Studio main window, click to open the **[Export Talend projects in archive file]** dialog box.



Export Talend projects in archive file

Archive file

Export resources to an archive file on the local file system.

☐ DEMO_BD
☑ PROJECT_BD
☑ PROJECT_DI

☐ .project
☐ migration.log
☐ talend.project

To archive file: D:\Talend\MyProjects.zip

Options
● Save in zip format
○ Save in tar format
☑ Compress the contents of the file
● Create directory structure for files
○ Create only selected directories

2. Select the check boxes of the projects you want to export. You can select only parts of the project through the **Filter Types...** link, if need be (for advanced users).

3. In the **To archive file** field, type in the name of or browse to the archive file where you want to export the selected projects.

4. In the **Option** area, select the compression format and the structure type you prefer.

5. Click **Finish** to validate the changes.

The archived file that holds the exported projects is created in the defined place.

1.4. Multi-perspective approach

Talend Studio offers a comprehensive set of tools and functions for all its key capabilities including data and application integration, data profiling and master data management. These tools are all accessible from different perspectives within the studio.

1.4.1. Switching between different perspectives

There are different ways to switch between different perspectives in the studio. They are as follows:

To switch between perspectives using quick access icons, do the following:

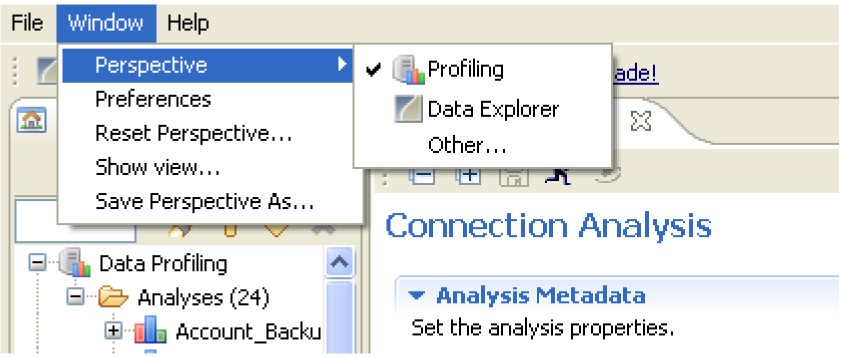
- In the top right corner of the studio, select:

Icon	**to...**
Integration	open the **Integration** perspective where you have access to a set of components and routines dedicated to data integration.
Profiling	open the **Profiling** perspective where you can examine data in different data sources and design data cleansing analyses.
MDM	open the **MDM** perspective where you can build data models and define the rules master data has to follow.
Mediation	open the **Mediation** perspective where you can carry out application integration processes.
BPM	open the **BPM** perspective where you can design business workflows using graphical tools.

- Click the quick access icon in the top left corner of the studio to switch between the perspectives.

Alternatively, you may switch between perspectives using the menu bar:

1. On the menu bar, click **Window** > **Perspective**.



2. Select from the list:

Item	**to...**
Profiling	open the data profiler perspective where you can examine data available in different data sources.
Data Explorer	open the data explorer perspective where you can browse and query analyzed data.

Item	to...
Other...	open a dialog box from which you can select to open different perspectives that extend the studio functionalities.

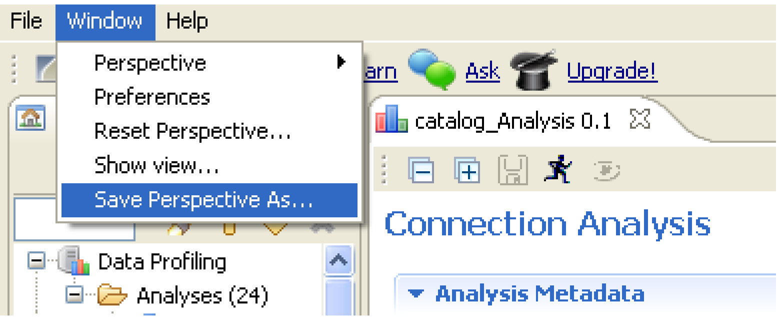
It is also possible, using the **Window - Show view...** combination, to show views from other perspectives in the open perspective.

1.4.2. Saving the configuration of a perspective

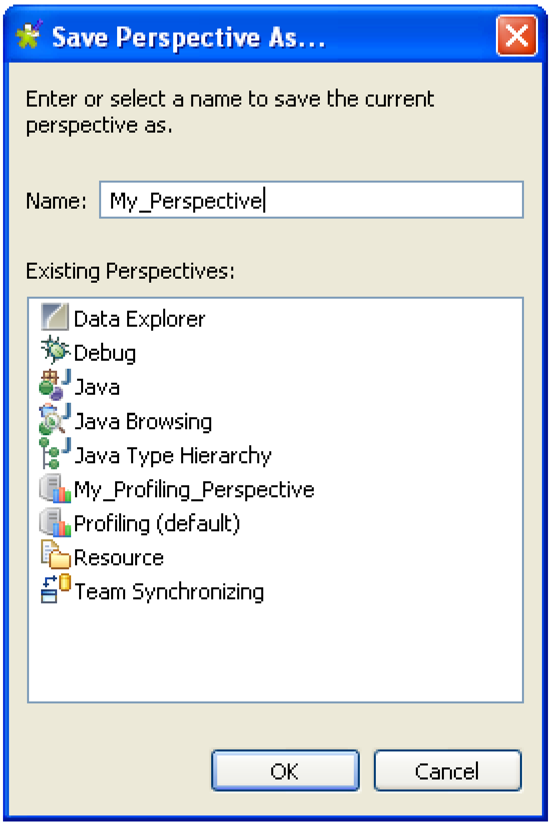
You can save the configuration of your current perspective in order to list it as a new perspective in the perspective dialog box.

To save the configuration of the current perspective, do the following:

1. On the menu bar, click **Window > Save Perspective As....**



2. In the **Name** field, enter a name.



Name: My_Perspective

3. Click **OK** .

The current perspective is saved as a new perspective under the new name.

You can open this perspective any time by selecting it from the **[Open Perspective]** dialog box. For further information, see *Switching between different perspectives*.

Chapter 2. Working in *Talend Studio* - basic Job examples

This chapter provides basic Job examples to help users get started with *Talend Studio*.

2.1. Getting started with a basic Job

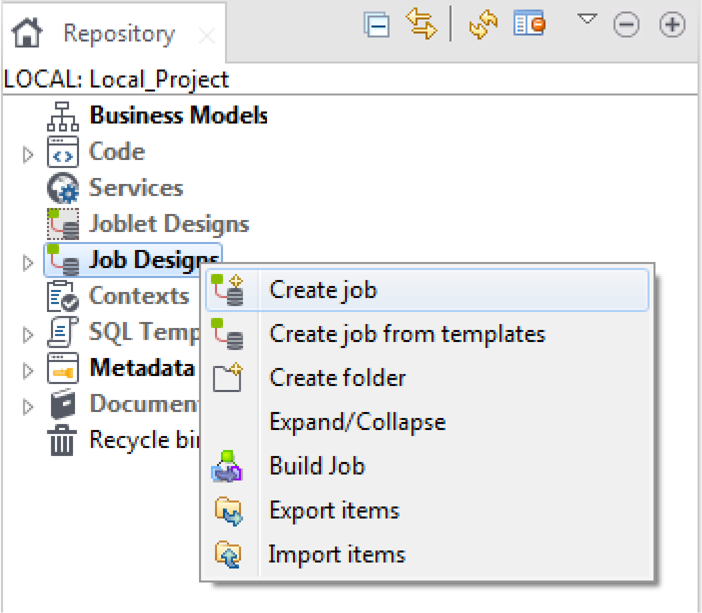
This section provides a continuous example that will help you create, add components to, configure, and execute a simple Job. This Job will be named *A_Basic_Job* and will read a text file, display its content on the **Run** console, and then write the data into another text file.

2.1.1. Creating a Job

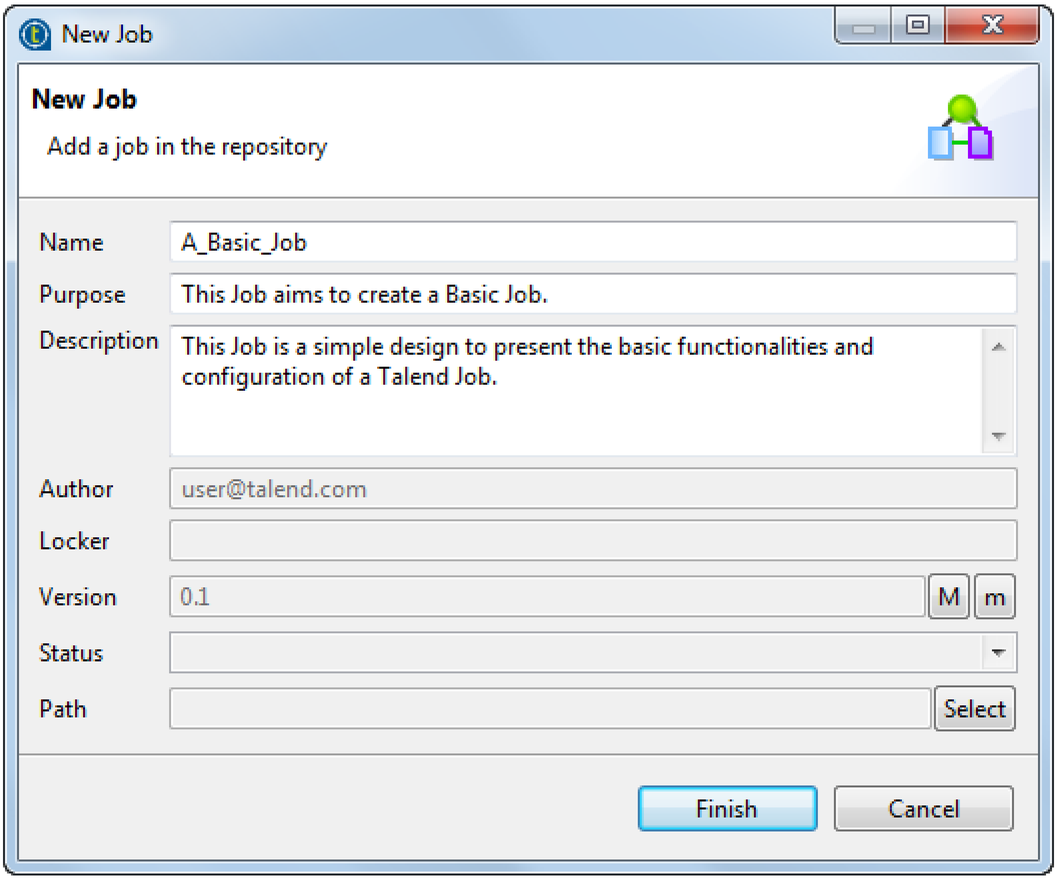
Talend Studio enables you to create a Job by dropping different technical components from the **Palette** onto the design workspace and then connecting these components together.

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

New Job

Add a job in the repository

Name: A_Basic_Job

Purpose: This Job aims to create a Basic Job.

Description: This Job is a simple design to present the basic functionalities and configuration of a Talend Job.

Author: user@talend.com

Locker:

Version: 0.1

Status:

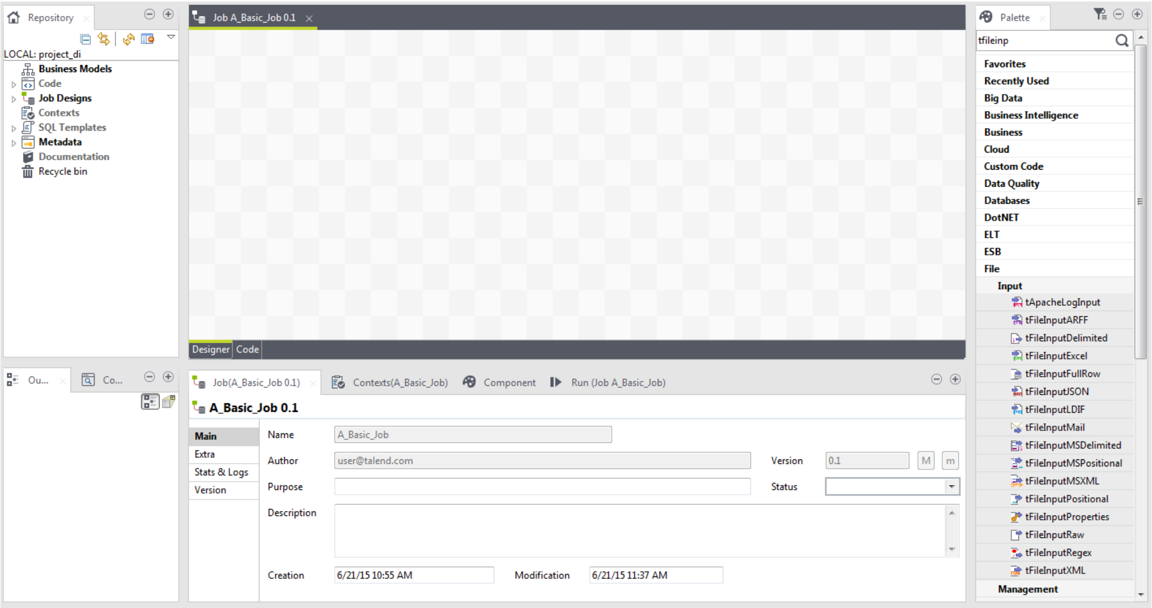
Path:

2. Fill the Job properties as shown in the previous screenshot.

   The fields correspond to the following properties:

Field	**Description**
Name	the name of the new Job. Note that a message comes up if you enter prohibited characters.
Purpose	Job purpose or any useful information regarding the Job use.
Description	Job description containing any information that helps you describe what the Job does and how it does it.
Author	a read-only field that shows by default the current user login.
Locker	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.
Version	a read-only field. You can manually increment the version using the **M** and **m** buttons.
Status	a list to select from the status of the Job you are creating.
Path	a list to select from the folder in which the Job will be created.

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



A_Basic_Job 0.1

Name: A_Basic_Job

Author: user@talend.com

Version: 0.1

Purpose:

Status:

Description:

Creation: 6/21/15 10:55 AM

Modification: 6/21/15 11:37 AM

Palette search: tfileinp

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.

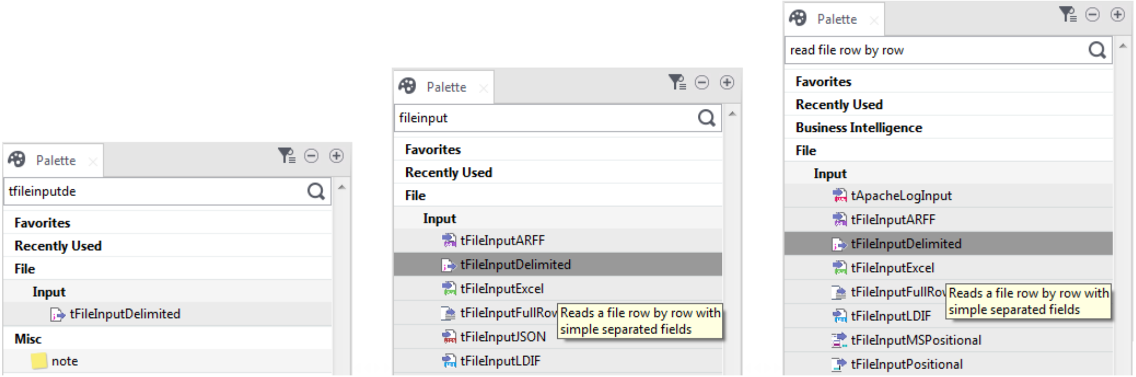
For more information regarding components and their functions, see *Talend Open Studio Components Reference Guide*.

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.

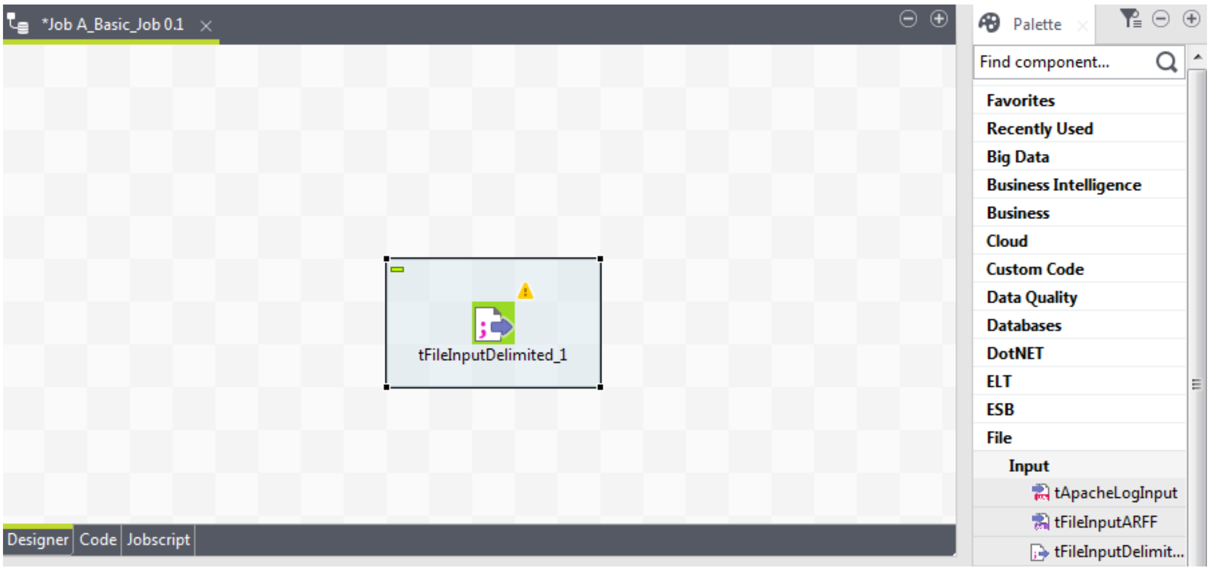
   The keyword(s) can be the partial or full name of the component, or a phrase describing its functionality if you don't know its name, for example, *tfileinputde*, *fileinput*, or *read file row by row*. .

   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.

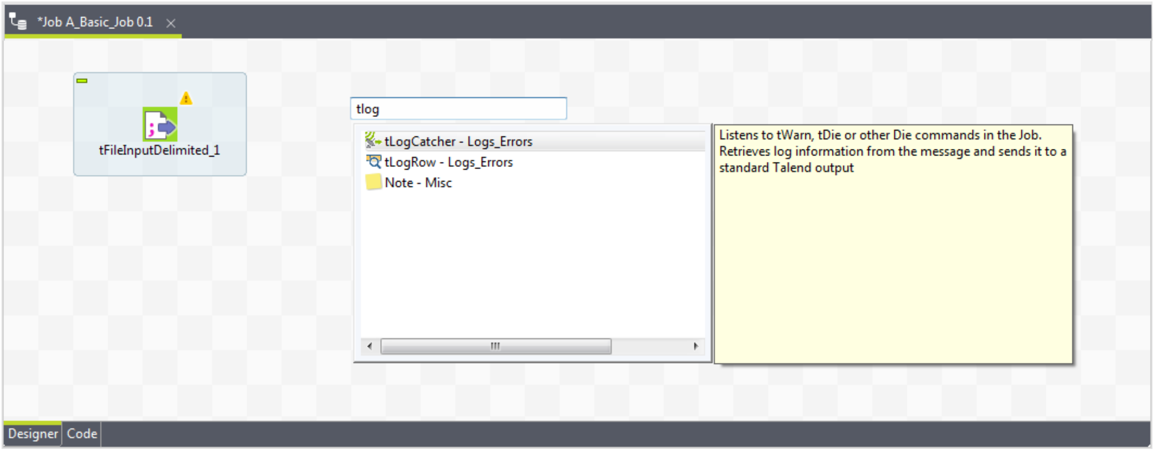
Prerequisite: Make sure you have selected the **Enable Component Creation Assistant** check box in the Studio preferences. For more information, see your *Talend Studio* User Guide.

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.



tlog

Listens to tWarn, tDie or other Die commands in the Job. Retrieves log information from the message and sends it to a standard Talend output

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

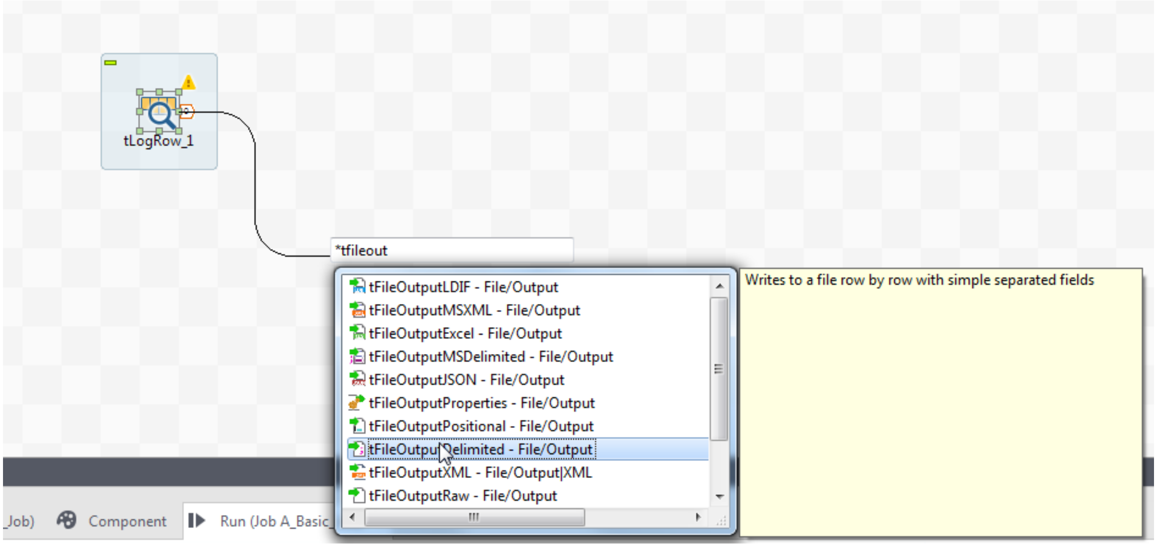
2.1.2.3. Adding an output component by dragging from an input one

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.

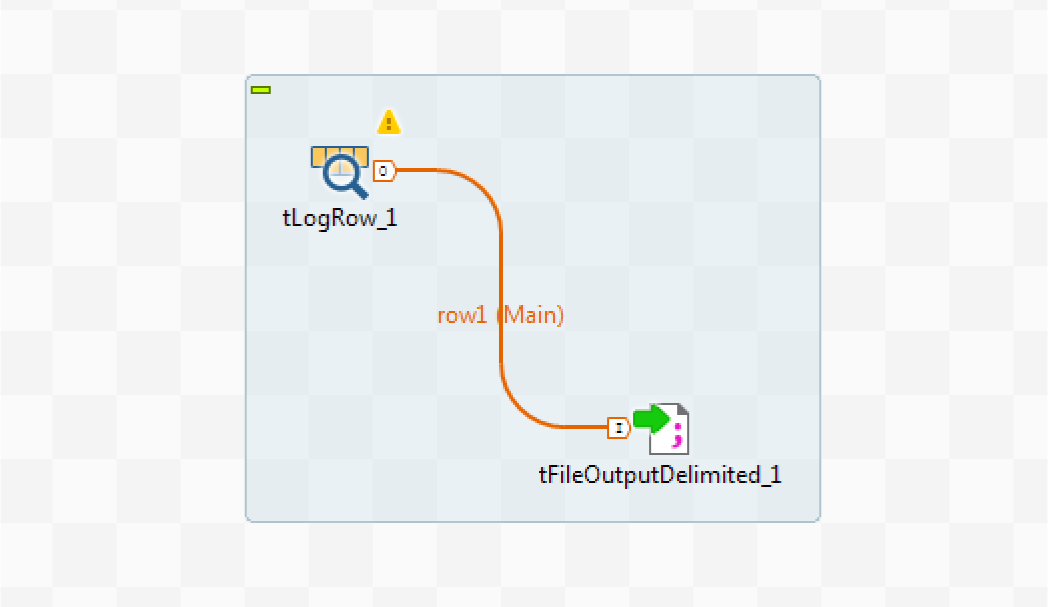
2. Drag and drop the **o** icon where you want to add a new component.

   A text field and a component list appear. The component list shows all the components that can be connected with the input component.


tLogRow_1
*tfileout
tFileOutputLDIF - File/Output
tFileOutputMSXML - File/Output
tFileOutputExcel - File/Output
tFileOutputMSDelimited - File/Output
tFileOutputJSON - File/Output
tFileOutputProperties - File/Output
tFileOutputPositional - File/Output
tFileOutputDelimited - File/Output
tFileOutputXML - File/Output|XML
tFileOutputRaw - File/Output
Writes to a file row by row with simple separated fields
Job)
Component
Run (Job A_Basic

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.

   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.

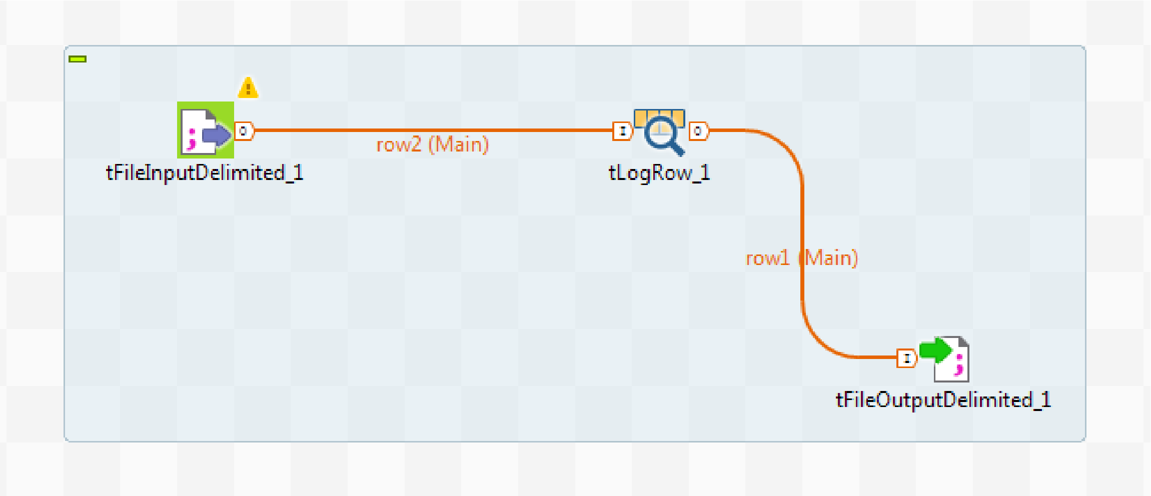
In this example, as the **tLogRow** and **tFileOutputDelimited** components are already connected, you only need to connect the **tFileInputDelimited** to the **tLogRow** component.

To connect the components together, proceed as follows:

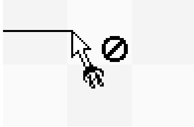
1. Right-click the source component, **tFileInputDelimited** in this example.

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.



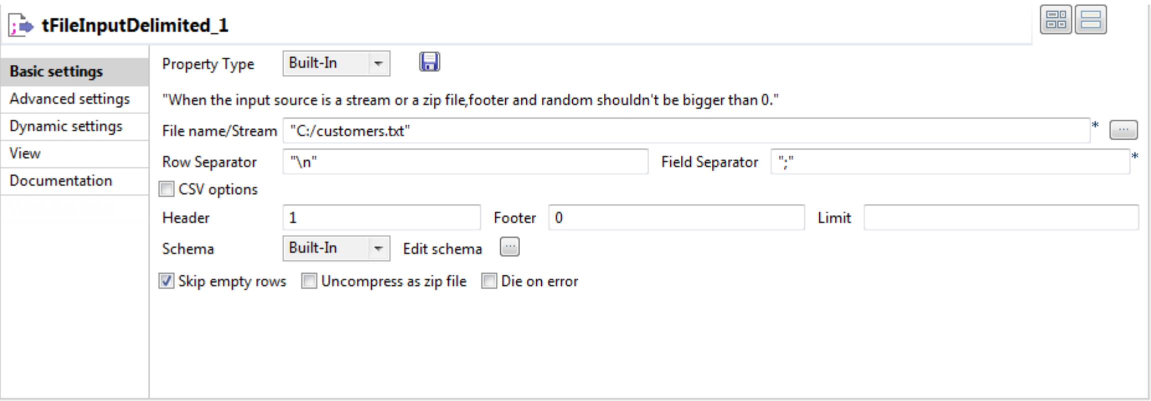
According to the nature and the role of the components you want to link together, several types of link are available. Only the authorized connections are listed in the contextual menu.

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.



tFileInputDelimited_1

Basic settings
Advanced settings
Dynamic settings
View
Documentation

Property Type Built-In

"When the input source is a stream or a zip file,footer and random shouldn't be bigger than 0."

File name/Stream "C:/customers.txt"

Row Separator "\n" Field Separator ";"

□ CSV options

Header 1 Footer 0 Limit

Schema Built-In Edit schema

☑ Skip empty rows □ Uncompress as zip file □ Die on error

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

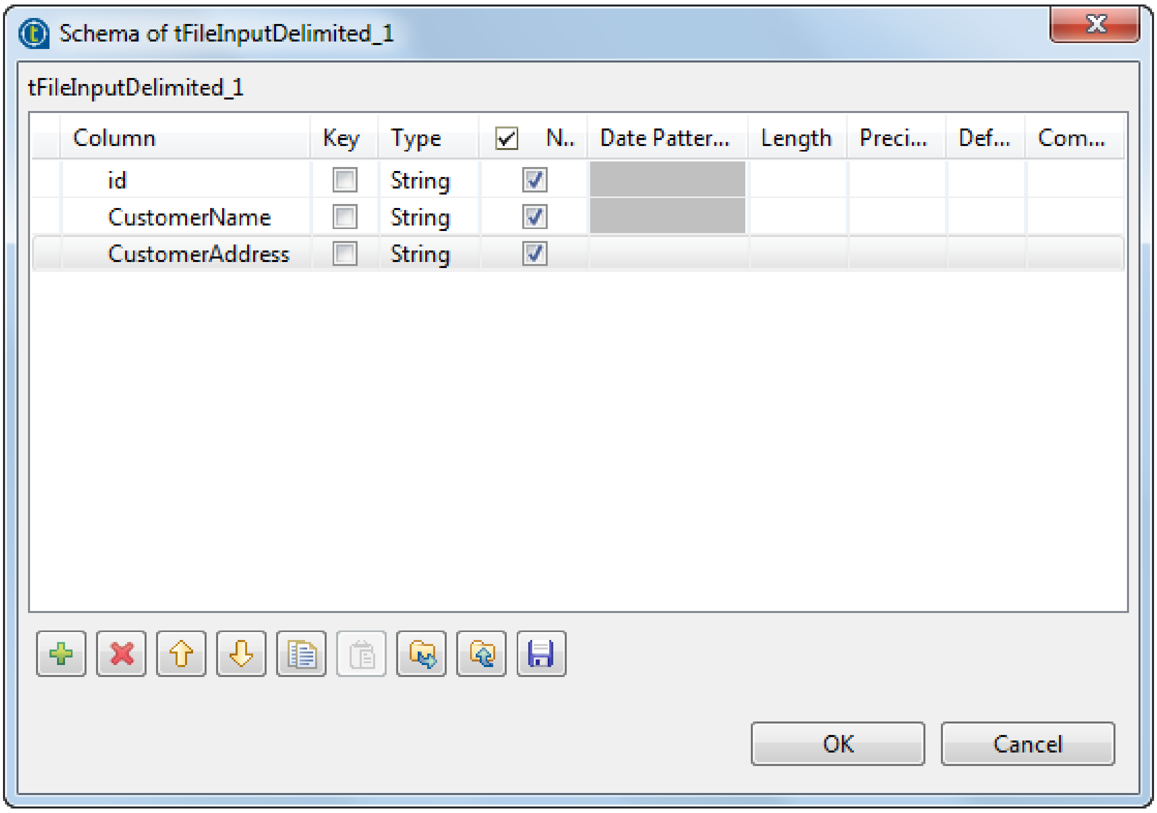
3. Browse your system or enter the path to the input file, *customers.txt* in this example.

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.

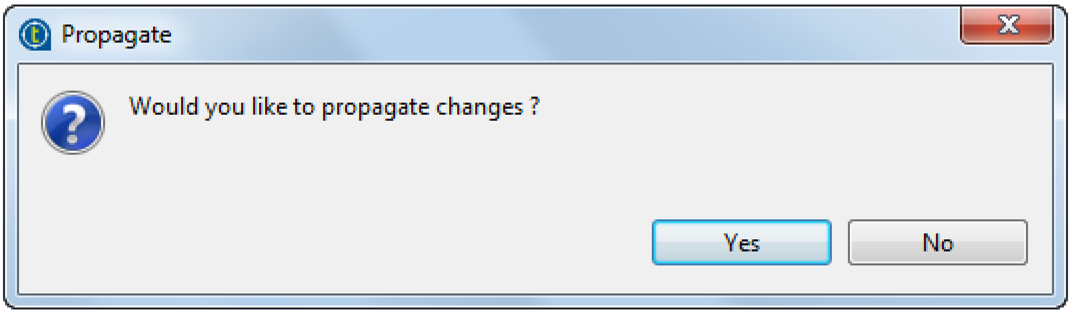
7. Name the three columns *id*, *CustomerName* and *CustomerAddress* respectively and click **OK** to close the editor.



Column	Key	Type	☑ N..	Date Patter...	Length	Preci...	Def...	Com...
id	☐	String	☑					
CustomerName	☐	String	☑					
CustomerAddress	☐	String	☑					

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.



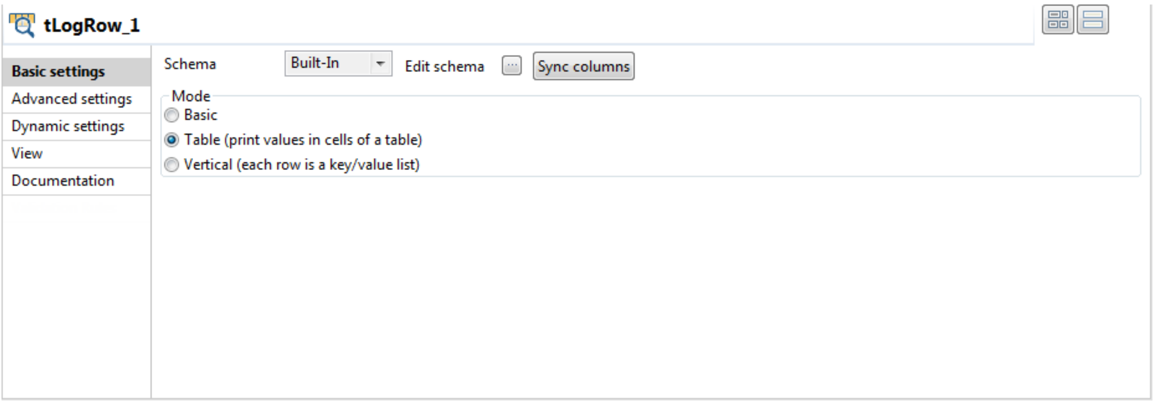
Would you like to propagate changes ?

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.



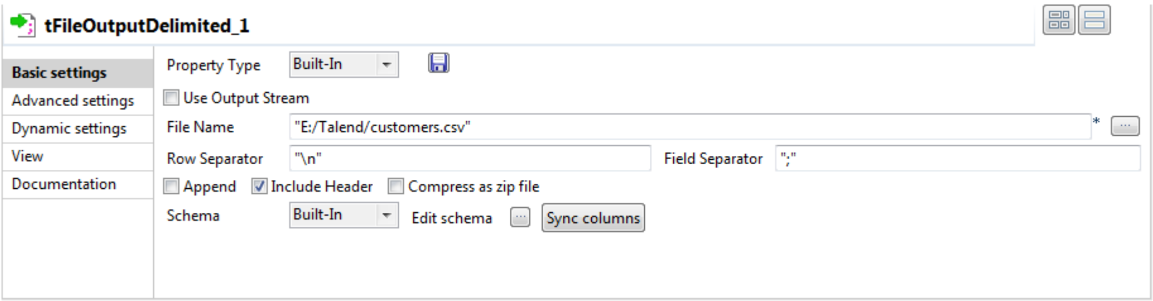
tLogRow_1

Schema: Built-In

Mode
○ Basic
● Table (print values in cells of a table)
○ Vertical (each row is a key/value list)

Configuring the tFileOutputDelimited component

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



tFileOutputDelimited_1

Property Type: Built-In

□ Use Output Stream

File Name: "E:/Talend/customers.csv"

Row Separator: "\n" Field Separator: ";"

□ Append ☑ Include Header □ Compress as zip file

Schema: Built-In

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

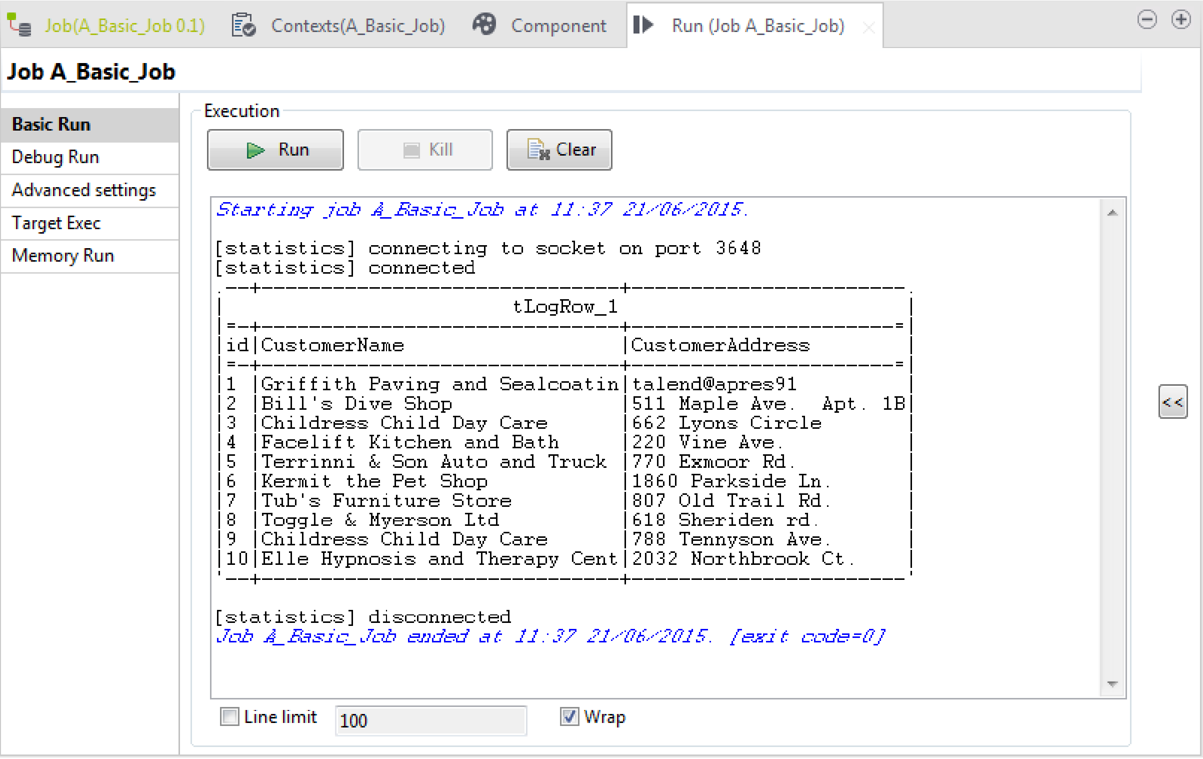
Now that components are configured, the Job can be executed.

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.



```
Starting job A_Basic_Job at 11:37 21/06/2015.

[statistics] connecting to socket on port 3648
[statistics] connected
.--+------------------------------+-----------------------.
|                        tLogRow_1                        |
|=-+------------------------------+----------------------=|
|id|CustomerName                  |CustomerAddress        |
|=-+------------------------------+----------------------=|
|1 |Griffith Paving and Sealcoatin|talend@apres91         |
|2 |Bill's Dive Shop              |511 Maple Ave.  Apt. 1B|
|3 |Childress Child Day Care      |662 Lyons Circle       |
|4 |Facelift Kitchen and Bath     |220 Vine Ave.          |
|5 |Terrinni & Son Auto and Truck |770 Exmoor Rd.         |
|6 |Kermit the Pet Shop           |1860 Parkside Ln.      |
|7 |Tub's Furniture Store         |807 Old Trail Rd.      |
|8 |Toggle & Myerson Ltd          |618 Sheriden rd.       |
|9 |Childress Child Day Care      |788 Tennyson Ave.      |
|10|Elle Hypnosis and Therapy Cent|2032 Northbrook Ct.    |
'--+------------------------------+-----------------------'

[statistics] disconnected
Job A_Basic_Job ended at 11:37 21/06/2015. [exit code=0]
```

□ Line limit 100 ☑ Wrap

Chapter 3. Profiling data

This chapter aims at users of *Talend Data Quality* who seek a real-life use case to help them take full control over data quality products.

It describes how to use the **Profiling** perspective in *Talend Studio* to profile data.

3.1. Profiling customer data

Incorporating appropriate data quality tools in your business processes is vital at the beginning of any project and through the project plan in order to see what type of data quality you have and decide how and what data to resolve.

Suppose, for example, that you want to start a campaign for your sails and marketing groups, or you need to contact customers for billing and payment and your main source to contact appropriate people is email and postal addresses. Having consistent and correct address data is vital in such campaign to be able to reach all people.

This section provides an example of profiling US customer email and postal addresses.

3.1.1. Identifying data anomalies

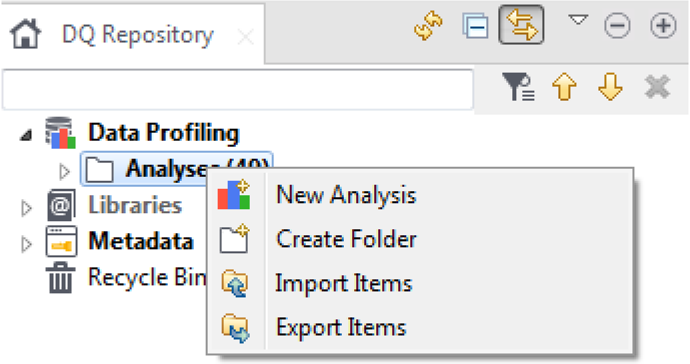
The first step in this example is to profile the customer contact information in a MySQL database. The profiling results provides you with statistics about the values within each column.

3.1.1.1. How to profile address columns

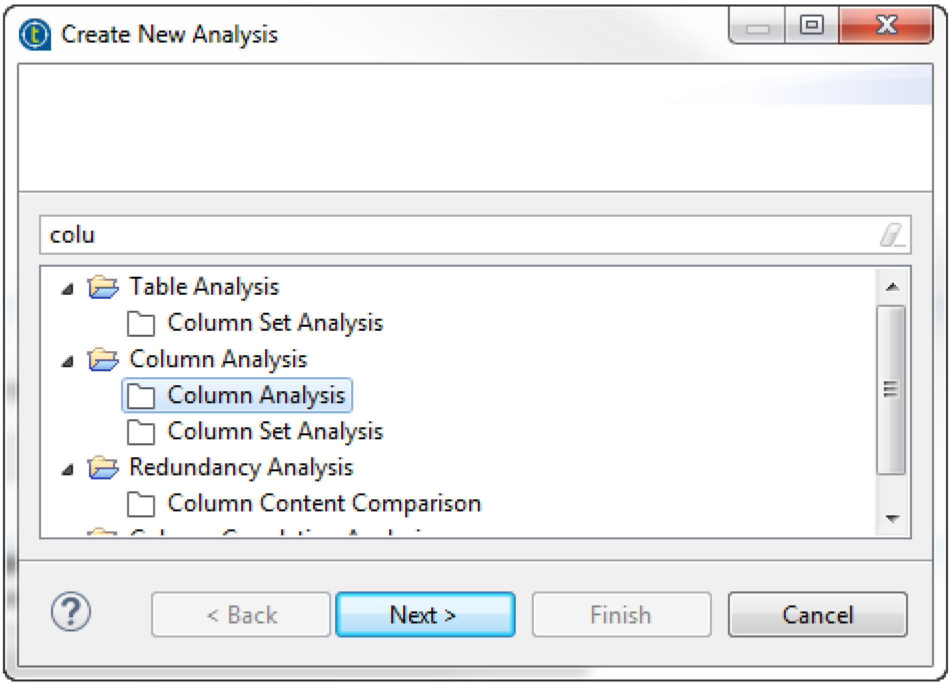
You will use the studio to analyze few customer columns including *email* and *postal*. Using out-of-box indicators and patterns on these columns, you can show in the analysis results the matching and non-matching address data, the number of most frequent records for each distinct pattern and the row, duplicate and blank counts in each column.

Defining the column analysis

1. In the **DQ Repository** tree view, right-click the **Analysis** folder and select **New Analysis**.



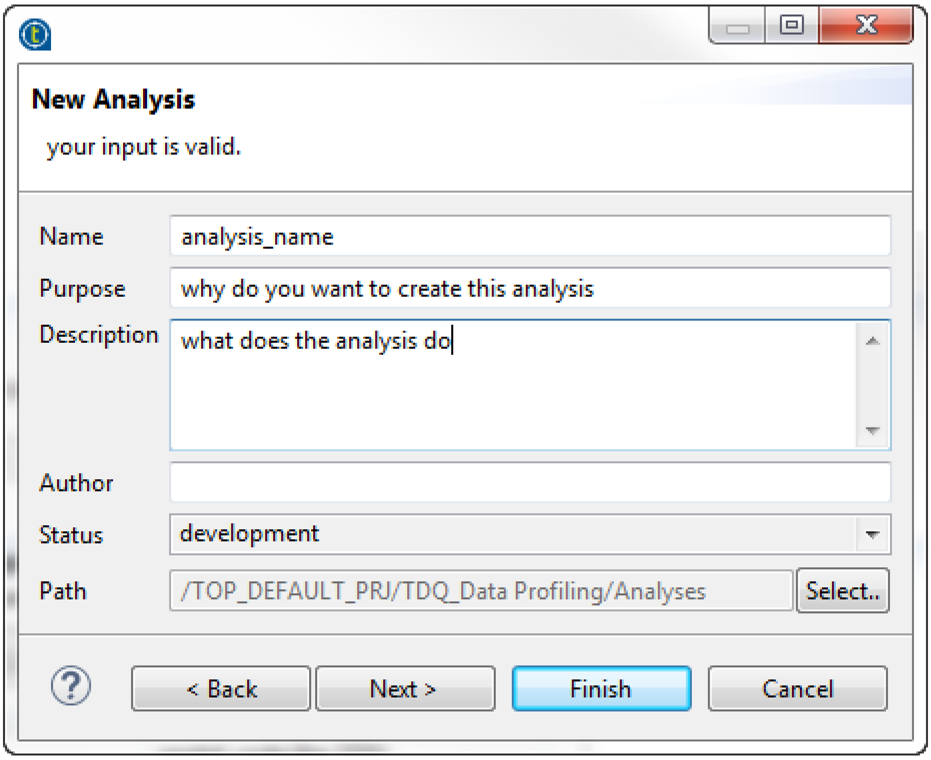
   The **[Create New Analysis]** wizard opens.



Create New Analysis

colu

2. Start typing *column* in the search field, select **Column Analysis** from the list and click **Next**.



New Analysis

your input is valid.

Name: analysis_name

Purpose: why do you want to create this analysis

Description: what does the analysis do

Author:

Status: development

Path: /TOP_DEFAULT_PRJ/TDQ_Data Profiling/Analyses

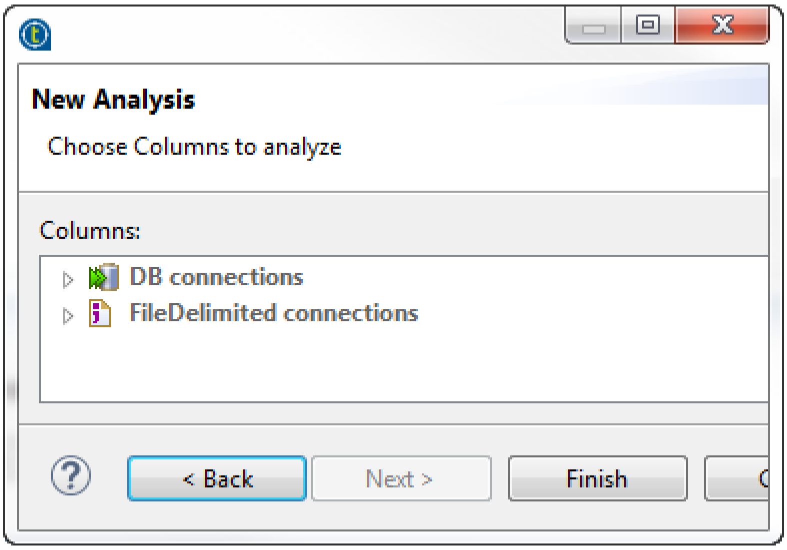
3. In the **Name** field, enter a name for the current column analysis.

Avoid using special characters in the item names including:

"~", "!", "`", "#", "^", "&", "*", "\\", "/", "?", ":", ";", "\"", ".", "(", ")", "'", "¥", "‘", """, "«", "»", "<", ">".

These characters are all replaced with "_" in the file system and you may end up creating duplicate items.

4. Set column analysis metadata (purpose, description and author name) in the corresponding fields and click **Next**.



New Analysis

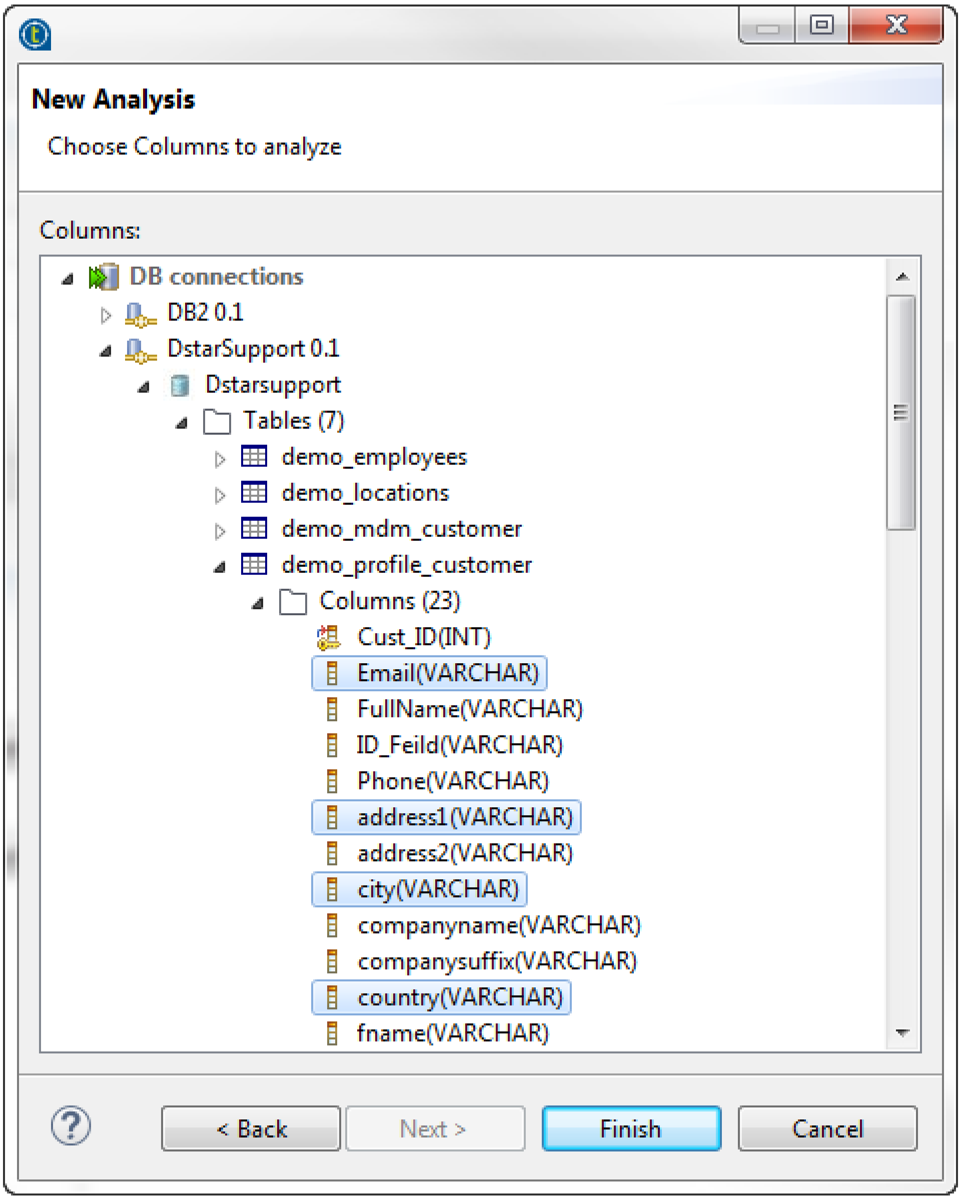
Choose Columns to analyze

Columns:

- DB connections
- FileDelimited connections

Selecting the address columns and setting sample data

1. Expand **DB connections** and browse to the address columns you want to analyze.



New Analysis

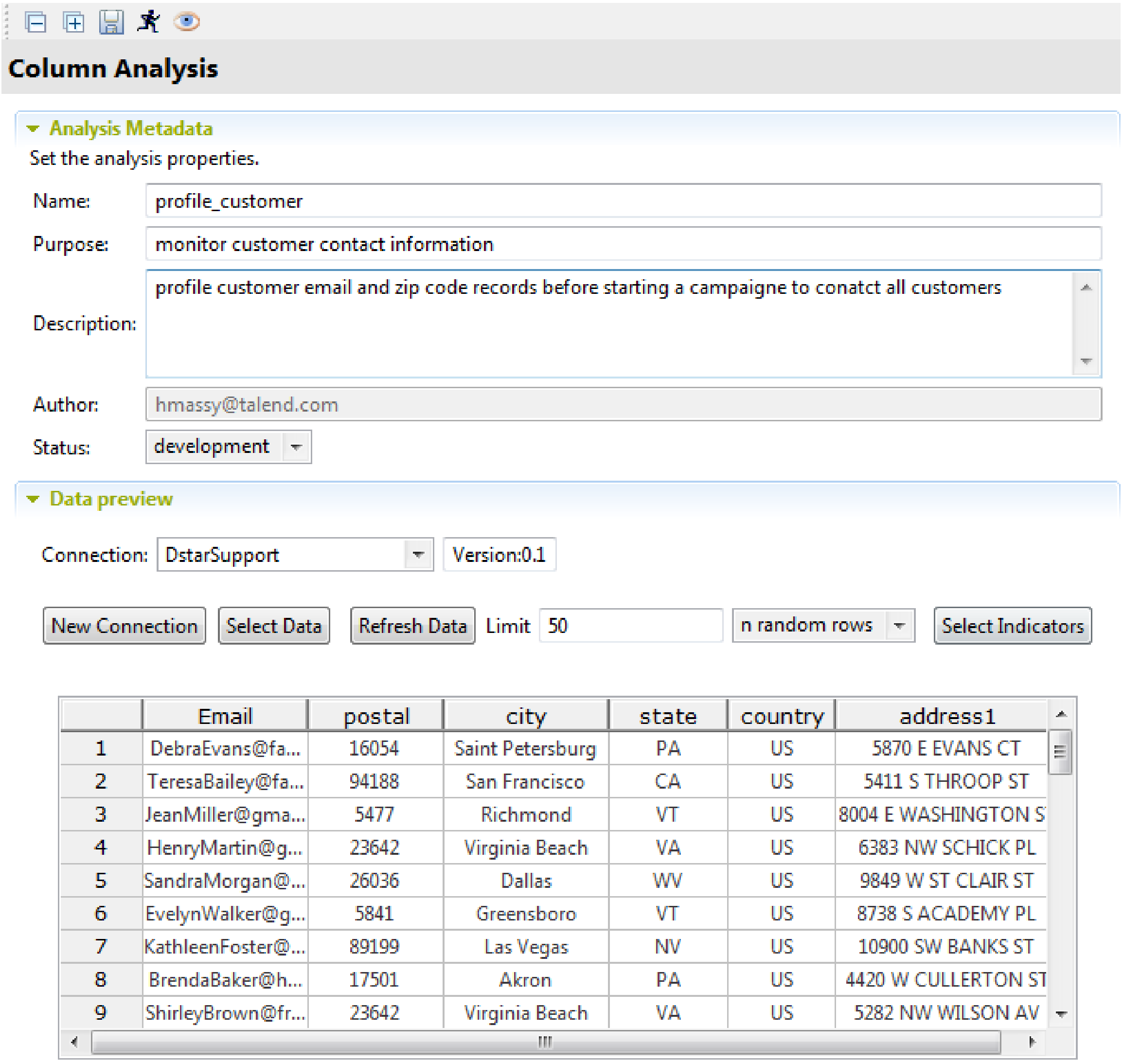
Choose Columns to analyze

Columns:

- DB connections
  - DB2 0.1
  - DstarSupport 0.1
    - Dstarsupport
      - Tables (7)
        - demo_employees
        - demo_locations
        - demo_mdm_customer
        - demo_profile_customer
          - Columns (23)
            - Cust_ID(INT)
            - Email(VARCHAR)
            - FullName(VARCHAR)
            - ID_Feild(VARCHAR)
            - Phone(VARCHAR)
            - address1(VARCHAR)
            - address2(VARCHAR)
            - city(VARCHAR)
            - companyname(VARCHAR)
            - companysuffix(VARCHAR)
            - country(VARCHAR)
            - fname(VARCHAR)

2. Select the columns and click **Finish** to close the wizard.

   A file for the newly created column analysis is listed under the **Analysis** node in the **DQ Repository** tree view, and the analysis editor opens with the analysis metadata.



Column Analysis

Analysis Metadata

Set the analysis properties.

Name: profile_customer

Purpose: monitor customer contact information

Description: profile customer email and zip code records before starting a campaigne to conatct all customers

Author: hmassy@talend.com

Status: development

Data preview

Connection: DstarSupport Version:0.1

Limit 50 n random rows

	Email	postal	city	state	country	address1
1	DebraEvans@fa...	16054	Saint Petersburg	PA	US	5870 E EVANS CT
2	TeresaBailey@fa...	94188	San Francisco	CA	US	5411 S THROOP ST
3	JeanMiller@gma...	5477	Richmond	VT	US	8004 E WASHINGTON S
4	HenryMartin@g...	23642	Virginia Beach	VA	US	6383 NW SCHICK PL
5	SandraMorgan@...	26036	Dallas	WV	US	9849 W ST CLAIR ST
6	EvelynWalker@g...	5841	Greensboro	VT	US	8738 S ACADEMY PL
7	KathleenFoster@...	89199	Las Vegas	NV	US	10900 SW BANKS ST
8	BrendaBaker@h...	17501	Akron	PA	US	4420 W CULLERTON ST
9	ShirleyBrown@fr...	23642	Virginia Beach	VA	US	5282 NW WILSON AV

3. In the **Data preview** view, click **Refresh Data**.

   The data in the selected columns is displayed in the table.

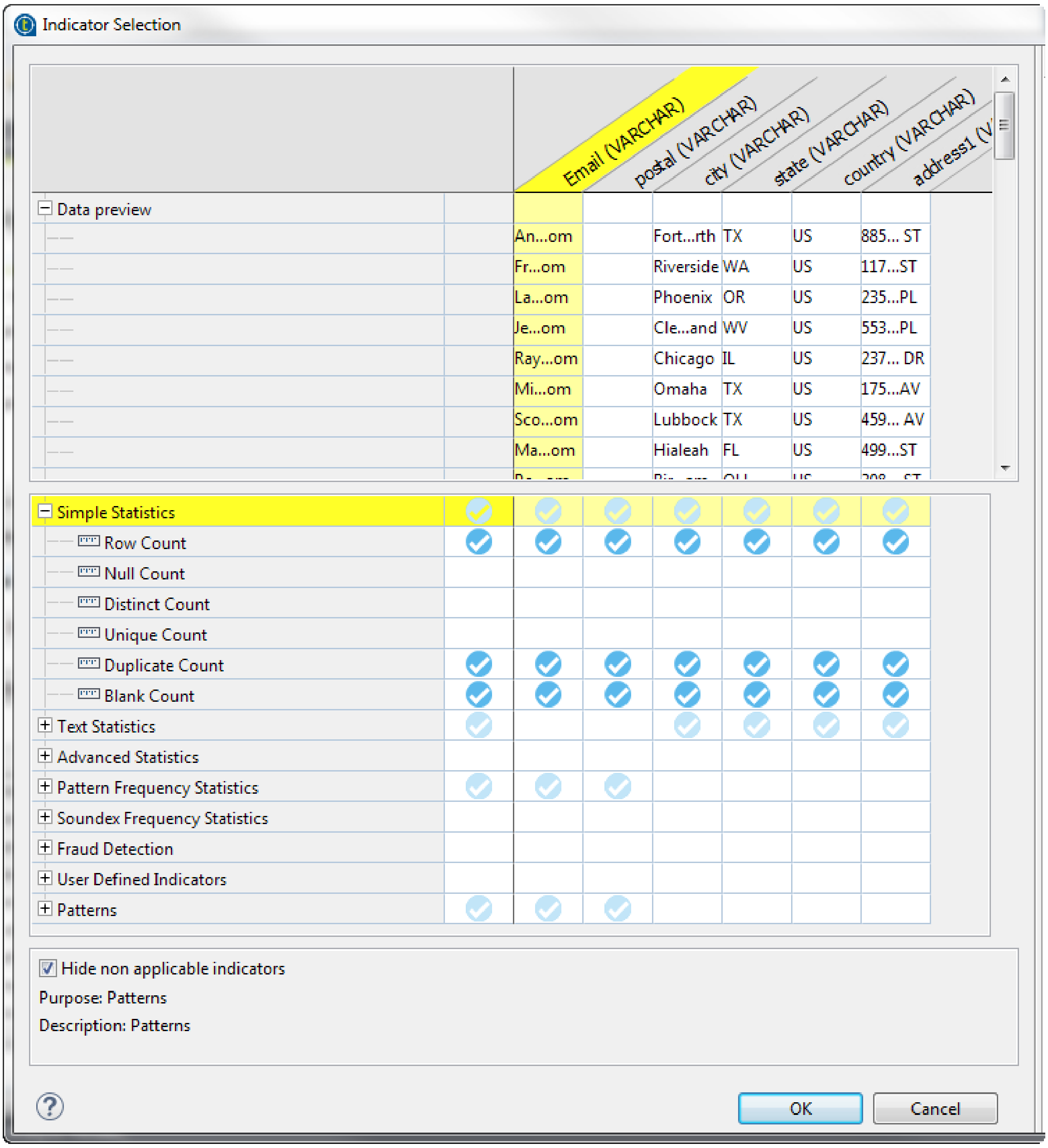
   You can change your data source and your selected columns by using the **New Connection** and **Select Data** buttons respectively.

4. In the **Limit** field, set to *50* the number for the data records you want to display in the table and use as sample data.

5. Select **n random rows** to list *50* random records from the selected columns.

Setting system indicators

1. From the **Data preview** view in the analysis editor, click **Select indicators** to open the **[Indicator Selection]** dialog box.



		Email (VARCHAR)	postal (VARCHAR)	city (VARCHAR)	state (VARCHAR)	country (VARCHAR)	address1 (V
⊟ Data preview							
		An...om		Fort...rth	TX	US	885... ST
		Fr...om		Riverside	WA	US	117...ST
		La...om		Phoenix	OR	US	235...PL
		Je...om		Cle...and	WV	US	553...PL
		Ray...om		Chicago	IL	US	237... DR
		Mi...om		Omaha	TX	US	175...AV
		Sco...om		Lubbock	TX	US	459... AV
		Ma...om		Hialeah	FL	US	499...ST
⊟ Simple Statistics	✓	✓	✓	✓	✓	✓	✓
Row Count	✓	✓	✓	✓	✓	✓	✓
Null Count							
Distinct Count							
Unique Count							
Duplicate Count	✓	✓	✓	✓	✓	✓	✓
Blank Count	✓	✓	✓	✓	✓	✓	✓
⊞ Text Statistics	✓			✓	✓	✓	✓
⊞ Advanced Statistics							
⊞ Pattern Frequency Statistics	✓	✓	✓				
⊞ Soundex Frequency Statistics							
⊞ Fraud Detection							
⊞ User Defined Indicators							
⊞ Patterns	✓	✓	✓				

☑ Hide non applicable indicators

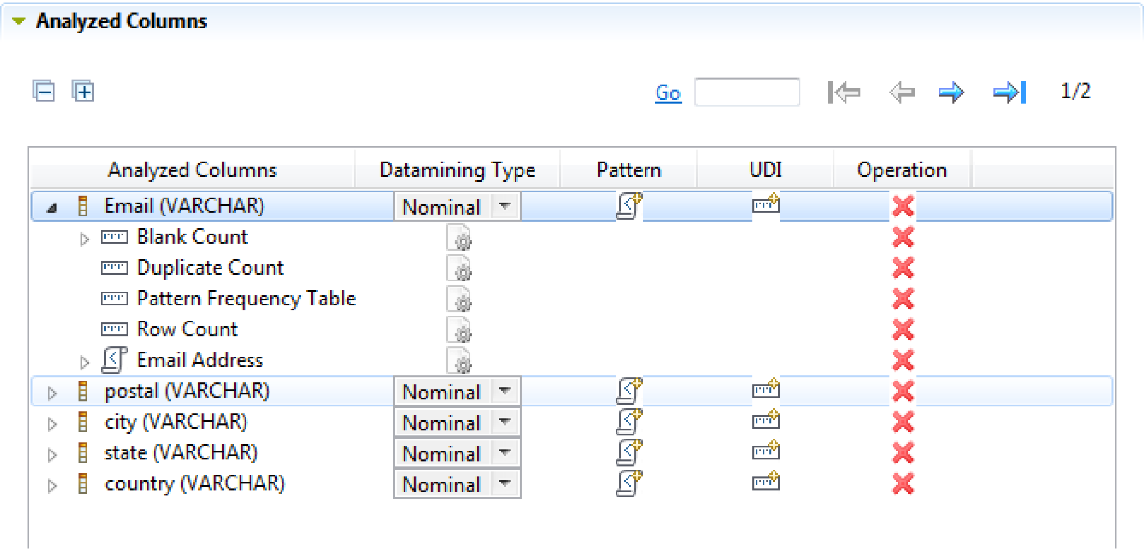
Purpose: Patterns

Description: Patterns

2. Click in the cells next to indicators names to set indicator parameters for the analyzed columns and click **OK**.

   You want to see the row, blank and duplicate counts in all columns to see how consistent the data is. Also you want to use the **Pattern Frequency Table** indicator on the *email* and *postal* columns in order to compute the number of most frequent records for each distinct pattern or value.

   Indicators are added accordingly to the columns in the **Analyzed Columns** view.

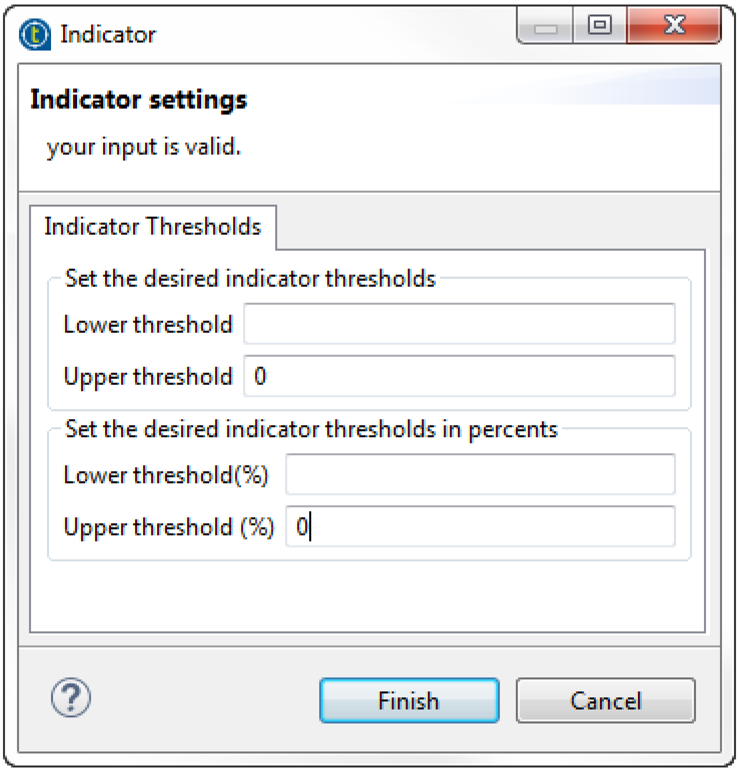


Analyzed Columns

Analyzed Columns	Datamining Type	Pattern	UDI	Operation
Email (VARCHAR)	Nominal			✕
Blank Count				✕
Duplicate Count				✕
Pattern Frequency Table				✕
Row Count				✕
Email Address				✕
postal (VARCHAR)	Nominal			✕
city (VARCHAR)	Nominal			✕
state (VARCHAR)	Nominal			✕
country (VARCHAR)	Nominal			✕

3. Click the option icon next to the **Blank Count** indicator and set *0* in the **Upper threshold** field.

   Defining thresholds on indicators is very helpful as it will write in red the count of the null values in the analysis results.



Indicator

Indicator settings

your input is valid.

Indicator Thresholds

Set the desired indicator thresholds

Lower threshold:

Upper threshold: 0

Set the desired indicator thresholds in percents

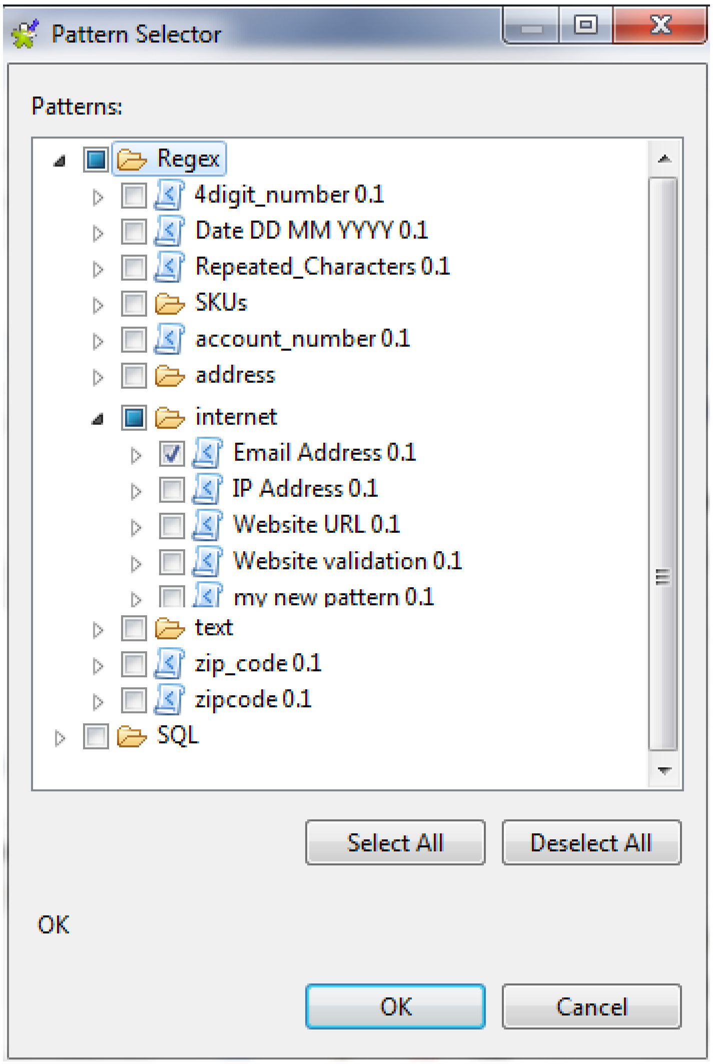
Lower threshold(%):

Upper threshold (%): 0

Setting patterns

You would want now to match the content of the *email* column against a standard email format and the *postal* column against a standard US zip code format. This will define the content, structure and quality of emails and zip codes and give a percentage of the data that match the standard formats and the data that does not match.

1. In the **Analyzed Columns** view, click the icon next to *email*.



2. In the **[Pattern Selector]** dialog box, expand **Regex** and browse to **Email Address** in the **internet** folder, and then click **OK**.

3. Click the option icon next to the **Email Address** indicator and set *98.0* in the **Lower threshold (%)** field.

   If the number of the records that match the pattern is fewer than 98%, it will be written in red in the analysis results.

4. Do the same to add to the *postal* column the **US Zipcode Validation** pattern from the **address** folder.

Executing the analysis and displaying the profiling results

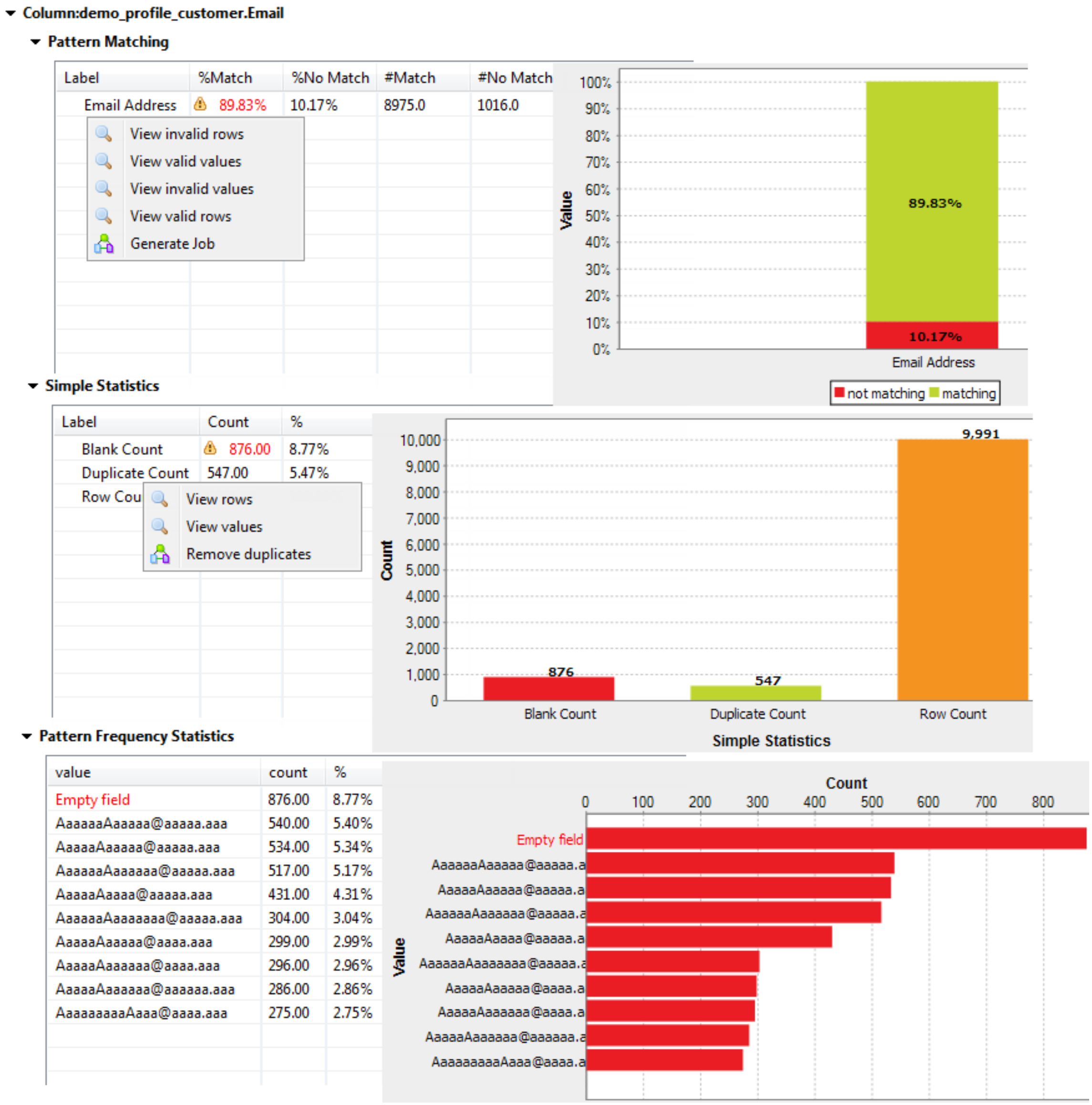
1. Save the column analysis in the analysis editor and then press **F6** to execute it.

   A group of graphics is displayed in the **Graphics** panel to the right of the analysis editor showing the results of the column analysis including those for pattern matching.

2. Click the **Analysis Results** tab at the bottom of the analysis editor to access a more detail result view.

   These results show the generated graphics for the analyzed columns accompanied with tables that detail the statistic and pattern matching results.

   The results for the *email* column look as the following:



Column:demo_profile_customer.Email

Pattern Matching

Label	%Match	%No Match	#Match	#No Match
Email Address	⚠ 89.83%	10.17%	8975.0	1016.0

Simple Statistics

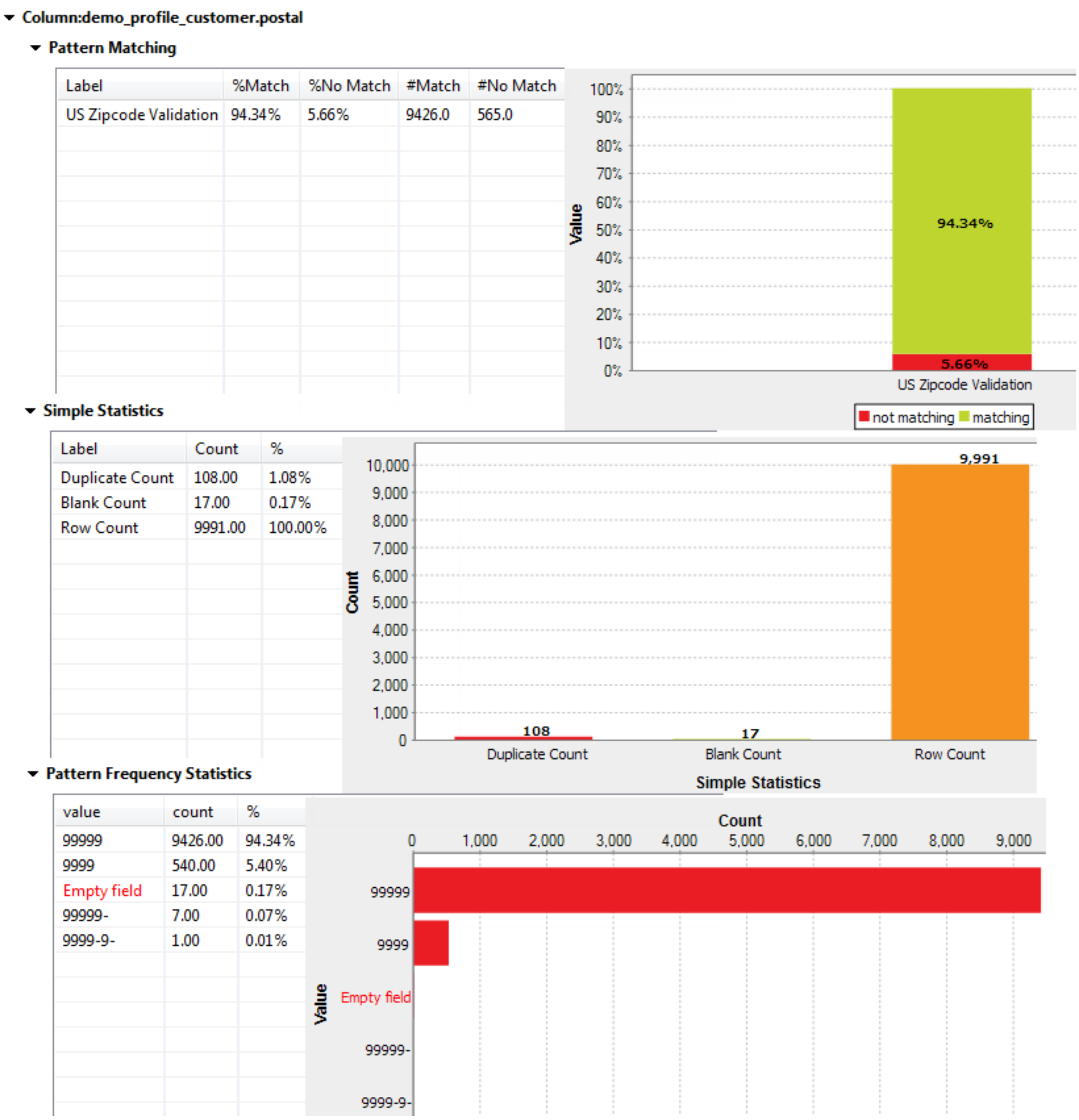
Label	Count	%
Blank Count	⚠ 876.00	8.77%
Duplicate Count	547.00	5.47%
Row Cou		

Pattern Frequency Statistics

value	count	%
Empty field	876.00	8.77%
AaaaaaAaaaaa@aaaaa.aaa	540.00	5.40%
AaaaaAaaaaa@aaaaa.aaa	534.00	5.34%
AaaaaaAaaaaaa@aaaaa.aaa	517.00	5.17%
AaaaaAaaaa@aaaaa.aaa	431.00	4.31%
AaaaaaAaaaaaaa@aaaaa.aaa	304.00	3.04%
AaaaaAaaaaa@aaaa.aaa	299.00	2.99%
AaaaaAaaaaaa@aaaa.aaa	296.00	2.96%
AaaaaAaaaaaa@aaaaaa.aaa	286.00	2.86%
AaaaaaaaaAaaa@aaaa.aaa	275.00	2.75%

The pattern matching results show that about 10% of the email records do not match the standard email pattern. The simple statistic results show that about 8% of the email records are blank and that about 5% are duplicates. And the pattern frequency results give the number of most frequent records for each distinct pattern. This shows that the data is not consistent and you need to correct and cleans the email data before starting your campaign.

The results for the *postal* column look as the following:



Column:demo_profile_customer.postal

Pattern Matching

Label	%Match	%No Match	#Match	#No Match
US Zipcode Validation	94.34%	5.66%	9426.0	565.0

Simple Statistics

Label	Count	%
Duplicate Count	108.00	1.08%
Blank Count	17.00	0.17%
Row Count	9991.00	100.00%

Pattern Frequency Statistics

value	count	%
99999	9426.00	94.34%
9999	540.00	5.40%
Empty field	17.00	0.17%
99999-	7.00	0.07%
9999-9-	1.00	0.01%

The result sets for the *postal* column give the count of the records that match and those that do not match a standard US zip code format. The results sets also give the blank and duplicate counts and the number of most frequent records for each distinct pattern. These results show that the data is not very consistent.

Then some percentage of the customers can not be contacted by either email or US mail service. These results show clearly that your data is not very consistent and that it needs to be corrected.

3.1.1.2. How to view analyzed data

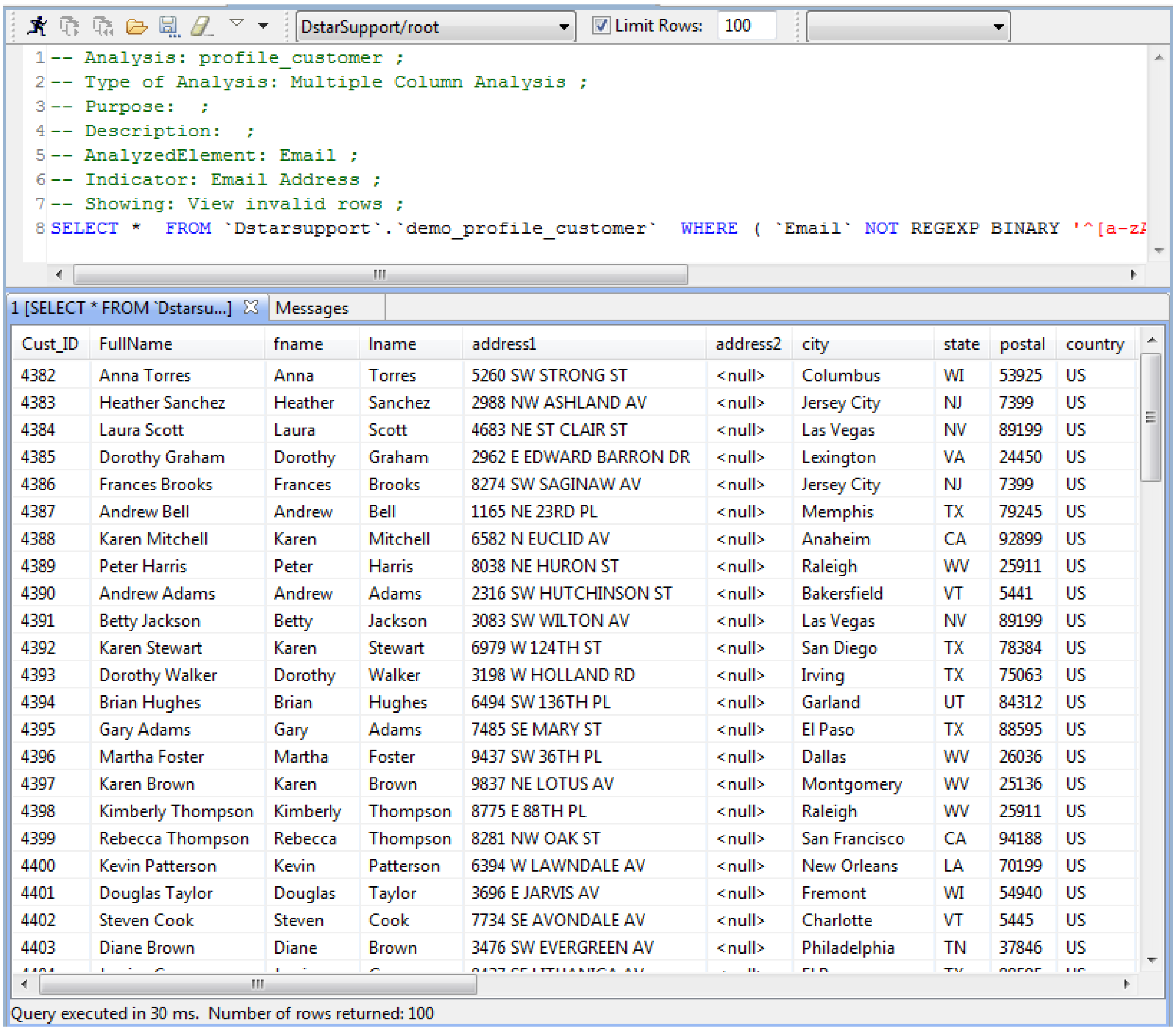
After running the column analysis using the SQL engine and from the **Analysis Results** view of the analysis editor, you can right-click any of the rows/bars in the result tables/charts and access a view of the actual analyzed data. This could be very helpful to see invalid rows for example and start analyzing what needs to be done to clean such data.

To view and export the analyzed data, do the following:

1. At the bottom of the analysis editor, click the **Analysis Results** tab to open a detailed view of the analysis results.

2. Right-click a data row in the statistic results of the *email* column and select **View invalid rows** for example.

The **Data Explorer** perspective opens listing the invalid rows in the *email* column.



```
-- Analysis: profile_customer ;
-- Type of Analysis: Multiple Column Analysis ;
-- Purpose:  ;
-- Description:  ;
-- AnalyzedElement: Email ;
-- Indicator: Email Address ;
-- Showing: View invalid rows ;
SELECT *  FROM `Dstarsupport`.`demo_profile_customer`  WHERE ( `Email` NOT REGEXP BINARY '^[a-z
```

Cust_ID	FullName	fname	lname	address1	address2	city	state	postal	country
4382	Anna Torres	Anna	Torres	5260 SW STRONG ST	<null>	Columbus	WI	53925	US
4383	Heather Sanchez	Heather	Sanchez	2988 NW ASHLAND AV	<null>	Jersey City	NJ	7399	US
4384	Laura Scott	Laura	Scott	4683 NE ST CLAIR ST	<null>	Las Vegas	NV	89199	US
4385	Dorothy Graham	Dorothy	Graham	2962 E EDWARD BARRON DR	<null>	Lexington	VA	24450	US
4386	Frances Brooks	Frances	Brooks	8274 SW SAGINAW AV	<null>	Jersey City	NJ	7399	US
4387	Andrew Bell	Andrew	Bell	1165 NE 23RD PL	<null>	Memphis	TX	79245	US
4388	Karen Mitchell	Karen	Mitchell	6582 N EUCLID AV	<null>	Anaheim	CA	92899	US
4389	Peter Harris	Peter	Harris	8038 NE HURON ST	<null>	Raleigh	WV	25911	US
4390	Andrew Adams	Andrew	Adams	2316 SW HUTCHINSON ST	<null>	Bakersfield	VT	5441	US
4391	Betty Jackson	Betty	Jackson	3083 SW WILTON AV	<null>	Las Vegas	NV	89199	US
4392	Karen Stewart	Karen	Stewart	6979 W 124TH ST	<null>	San Diego	TX	78384	US
4393	Dorothy Walker	Dorothy	Walker	3198 W HOLLAND RD	<null>	Irving	TX	75063	US
4394	Brian Hughes	Brian	Hughes	6494 SW 136TH PL	<null>	Garland	UT	84312	US
4395	Gary Adams	Gary	Adams	7485 SE MARY ST	<null>	El Paso	TX	88595	US
4396	Martha Foster	Martha	Foster	9437 SW 36TH PL	<null>	Dallas	WV	26036	US
4397	Karen Brown	Karen	Brown	9837 NE LOTUS AV	<null>	Montgomery	WV	25136	US
4398	Kimberly Thompson	Kimberly	Thompson	8775 E 88TH PL	<null>	Raleigh	WV	25911	US
4399	Rebecca Thompson	Rebecca	Thompson	8281 NW OAK ST	<null>	San Francisco	CA	94188	US
4400	Kevin Patterson	Kevin	Patterson	6394 W LAWNDALE AV	<null>	New Orleans	LA	70199	US
4401	Douglas Taylor	Douglas	Taylor	3696 E JARVIS AV	<null>	Fremont	WI	54940	US
4402	Steven Cook	Steven	Cook	7734 SE AVONDALE AV	<null>	Charlotte	VT	5445	US
4403	Diane Brown	Diane	Brown	3476 SW EVERGREEN AV	<null>	Philadelphia	TN	37846	US

Query executed in 30 ms. Number of rows returned: 100

Chapter 4. Building a simple MDM project

This chapter takes you through the main steps involved in building a simple MDM project.

4.1. Preparing your project in the Studio

This short scenario walks you through the main steps involved in setting up a simple MDM project. In this example, you recreate some of the content included in the MDM Demo Project.

If you already imported the MDM Demo Project, you will have conflicts with the name of the Data Model and other elements, so make sure you create a new project from scratch.

Firstly, in *Talend Studio*, you define the data model and data container, and you then set up a view that you can use to interact with the data they contain using *Talend MDM Web User Interface*.

More complex actions such as the creation of processes and triggers are beyond the scope of this scenario.

Before you begin this scenario, make sure you have a valid connection to an MDM Server and have created an empty project.

4.1.1. Setting up a data model and creating some business entities

The first step at the beginning of any MDM project involves setting up a data model and creating business entities in this data model.

4.1.1.1. Create a data model

To create a data model, do the following:

1. In the **MDM Repository** tree view, right-click **Data Model** and select **New** from the contextual menu.

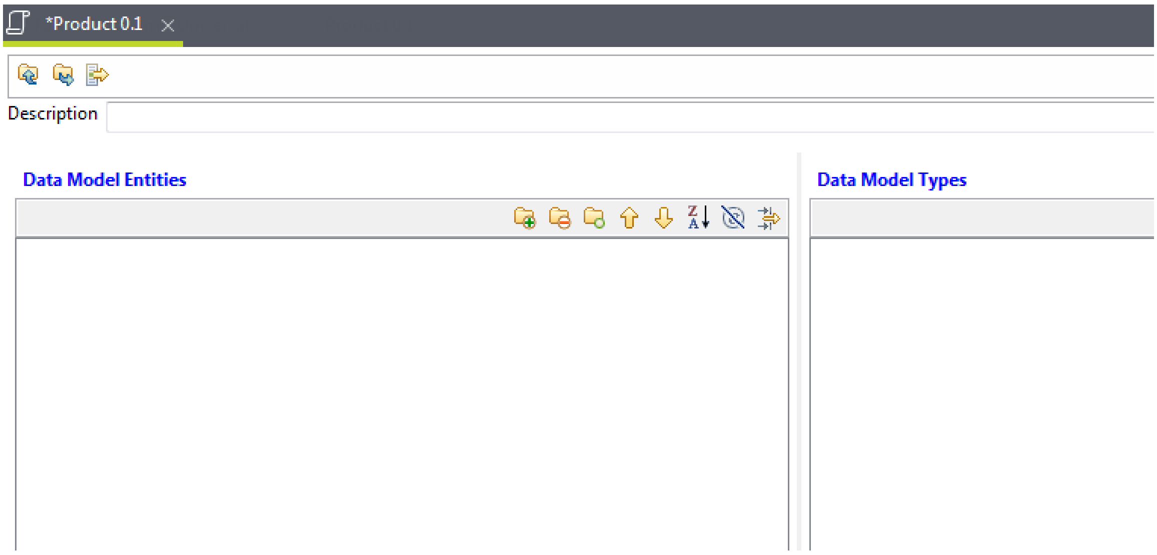
2. Name your data model *Product*, and then click **OK**.

   By default, the **Create the corresponding Data Container at the same time** check box is selected, so that the corresponding data container with the same name will be created.

   If needed, you can clear this check box and create the corresponding data container with the same name later. For more information, see *Defining a data container*.

   A data container and its corresponding data model must have the same name.

   In the Studio workspace, an editor opens where you can define some of the details of your new data model.



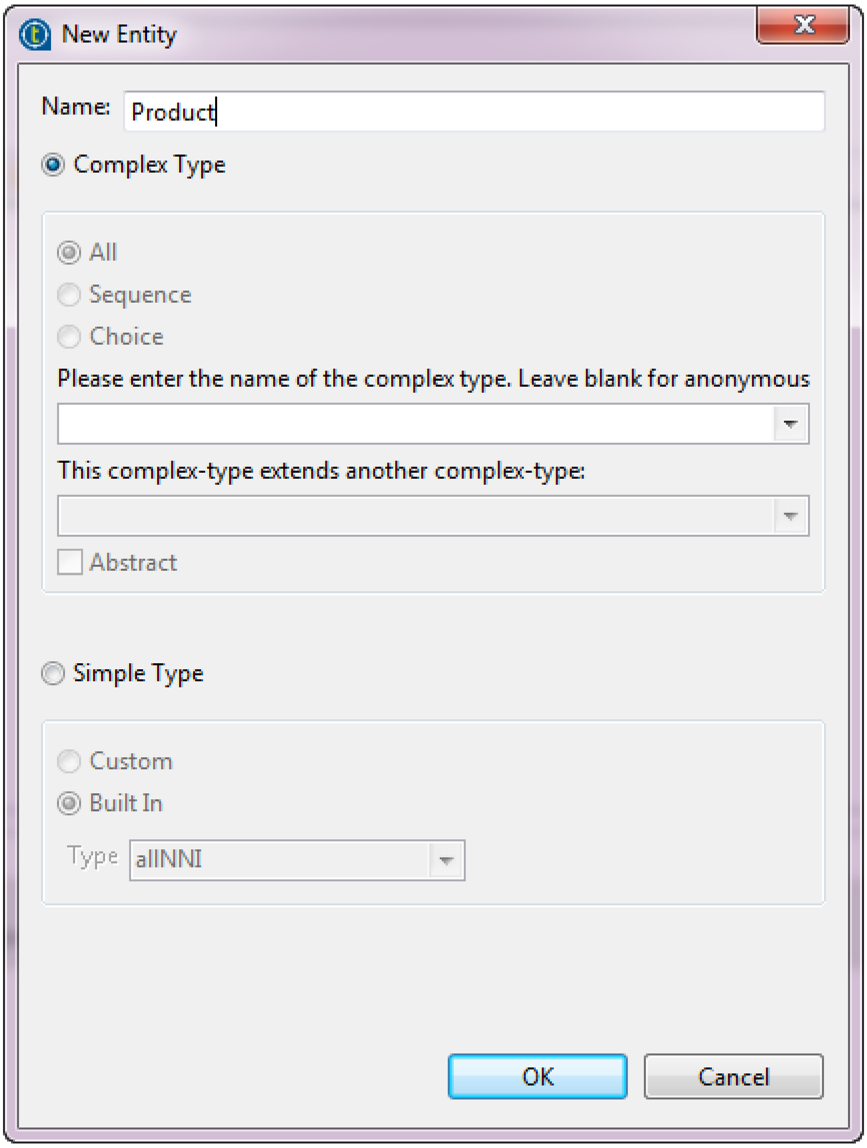
4.1.1.2. Create business entities in the data model

Once you have created your data model, you need to populate it with some business entities.

To create a business entity in your data model, do the following:

1. In the editor, right-click anywhere in the **Data Model Entities** panel, and then click **New Entity**.
2. In the **[New Entity]** dialog box that opens, enter a name for your new entity in the **Name** field: *Product*.
3. Select the **Complex Type** option.

   You use the **Simple type** option if you want to define a single element type such as a phone number or an email address, and the **Complex type** option if you want to define a more complete structure, such as an address or, in this case, the different attributes that describe a product.



New Entity

Name: Product

● Complex Type

● All
○ Sequence
○ Choice

Please enter the name of the complex type. Leave blank for anonymous

This complex-type extends another complex-type:

□ Abstract

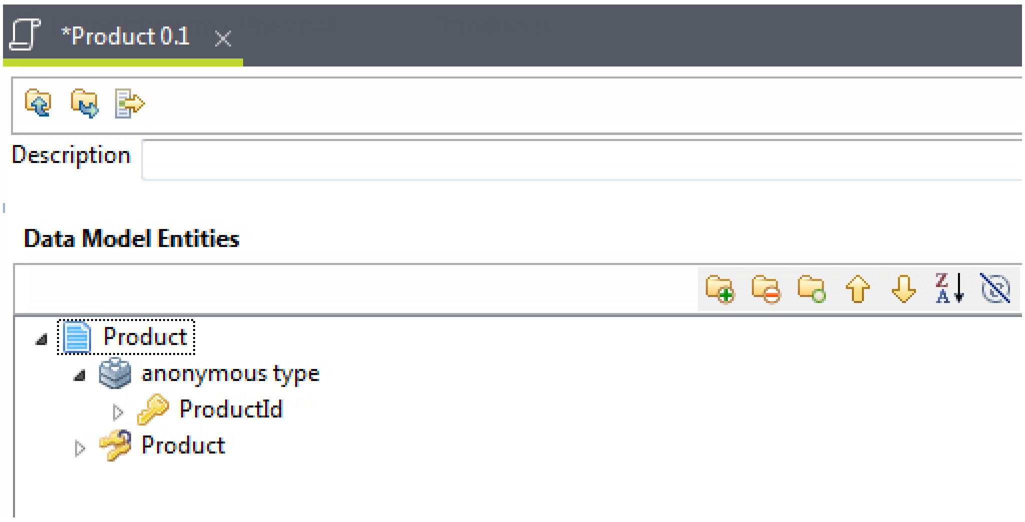
○ Simple Type

○ Custom
● Built In

Type allNNI

4. Leave the other options unchanged and click **OK** to add your new entity to the editor.

   The created business entity is listed in the **Data Model Entities** panel with a by default record, which takes its name from the entity name with the suffix *Id*, and the complex type, if any, is displayed in the **Data Model Types** panel.



*Product 0.1

Description

Data Model Entities

- Product
  - anonymous type
    - ProductId
  - Product

Each time you create a new business entity, a default Primary Key record, which takes its name from the entity name with the suffix *Id*, and a Unique Key record which has the same name as the Entity are automatically created. For example, if you create a new business entity and name it *Agency*, the Primary Key record *AgencyId* will be created automatically.

A Primary Key can be an integer but a Foreign Key must always be a string. The server surrounds Foreign Keys with square brackets to support compound keys.

4.1.1.3. Define attributes

The next step in this scenario involves defining different attributes for the *Product* entity you have just created: its name, description and price.

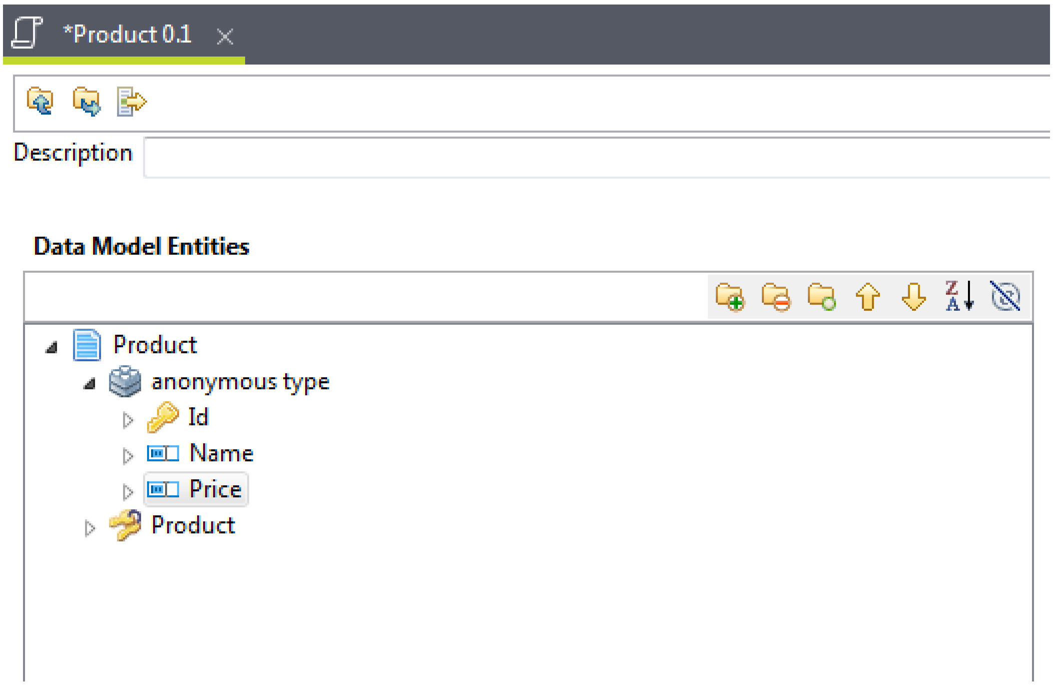
To define attributes for the *Product* business entity, do the following:

1. Expand the *Product* business entity and *anonymous type*, right-click the default Primary Key record and then click **Edit Element** in the contextual menu.

2. Change the name to *Id*, set the minimum and maximum occurrences to *1*, and then click **OK** to close the dialog box.

3. Right-click *Id*, click **Add Element (after)** in the contextual menu, and then add each of the following elements with the characteristics shown in the table below.

Element type	Element name	Minimum occurence	Maximum occurence
String	*Name*	1	1
Decimal	*Price*	1	1



4. Save your changes.

   A **[Validation Result Dialog]** dialog box opens to show the validation result.

5. In the **MDM Repository** tree view, expand the **Data Model** node, right-click the *Product* data model, and then select one of the deployment options to deploy your changes to the MDM Server.

4.1.2. Defining a data container

Once you have set up the basics of your data model, you need to define a data container in the MDM Hub where your master data is to be persisted.

To create a data container, do the following:

1. In the **MDM Repository** tree view, right-click **Data Container**, and then click **New**.

   The **[New Data Container]** dialog box is displayed.

2. Name your new data container *Product*, and then click **OK**.

3. Save your changes and then, in the **MDM Repository** tree view, expand the **Data Container** node, right-click the *Product* data container, and then select one of the deployment options to deploy your changes to the MDM Server.

4.1.3. Creating Views

Before business users are able to extract and query master data in *Talend MDM Web User Interface*, you need to create one or more Views in *Talend Studio*. These Views specify which data records within an entity a business user is allowed to search and view, and may also set conditions that filter which content can be delivered as the result of a search.

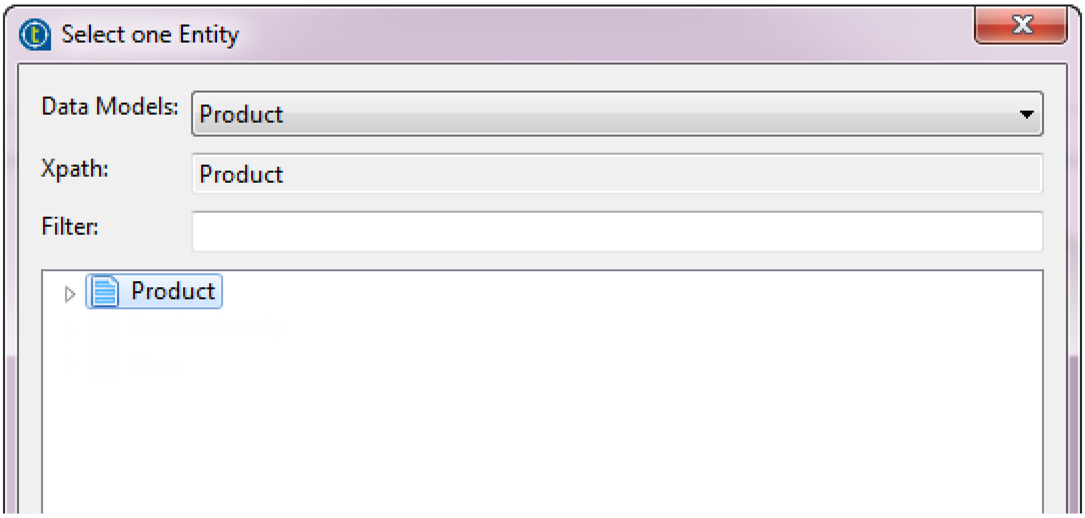
To create a View, do the following:

1. In the **MDM Repository** tree view, right-click **View** and select **New** from the contextual menu.

2. In the **[New View]** dialog box that opens, select **Web filter view**.

3. Click the **[...]** button to specify the entity on which you want to create a view.

4. In the dialog box that opens, select the *Product* data model in the drop-down box, click *Product* to set the appropriate XPath, and then click **Add** to close the dialog box.



Data Models: Product

Xpath: Product

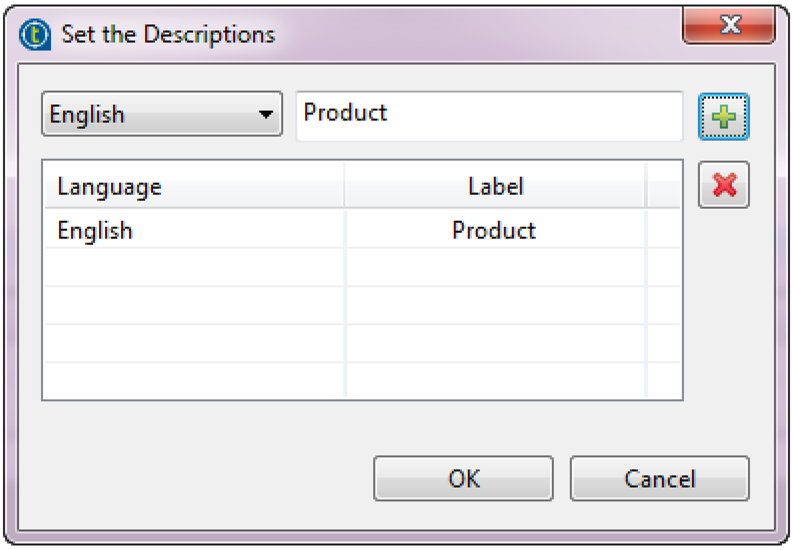
Filter:

5. The name of your View is set automatically as *Browse_items_Product*. Click **OK**.

Your View now exists in the **MDM Repository** tree view, but it has not yet been fully defined. You still need to specify exactly which elements can be viewed and searched.

To define the details of your View, do the following:

1. In the *Browse_items_Product* View which opens in the workspace, click the**[...]** button next to *Description* to define a description for your View in one or more languages (for example, *Product* since it is a View on all the products), and then click **OK** to close the dialog box when you have done so.



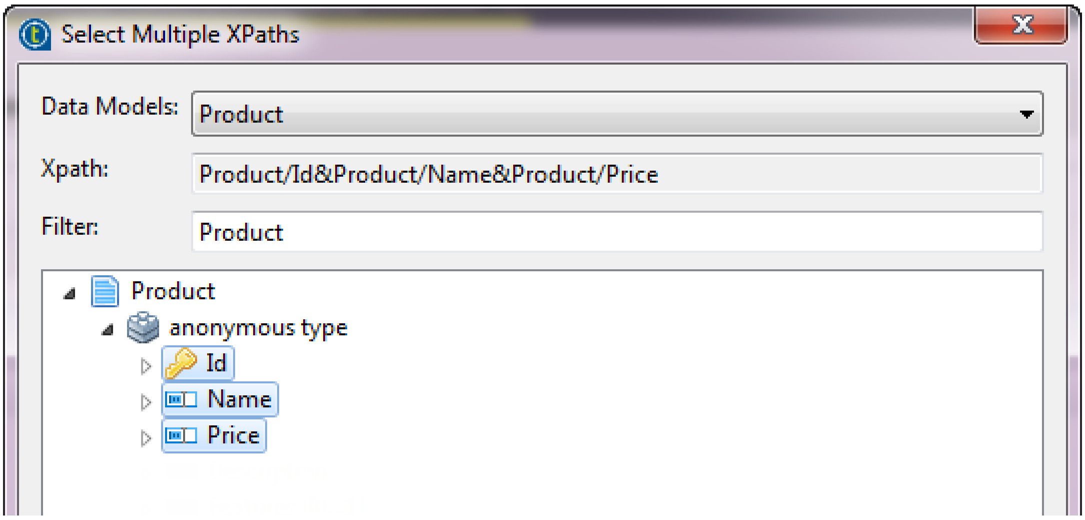
Set the Descriptions

English | Product

Language	Label
English	Product

2. Under **Viewable Business Elements**, click the **Add multiple** button to add multiple elements at the same time.

3. In the dialog box that opens, expand *Product* and *anonymous type*, and then press **Ctrl** and hold it down while you select each of the elements you want to make viewable: *Id*, *Name*, *Price*. Click **Add** to add each of these elements.



Select Multiple XPaths

Data Models: Product

Xpath: Product/Id&Product/Name&Product/Price

Filter: Product

4. Under **Searchable Business Elements**, click the **Add multiple** button to add multiple elements at the same time.

5. In the dialog box that opens, expand *Product* and *anonymous type*, and then press **Ctrl** and hold it down while you select each of the elements you want to make searchable: *Id*, *Name*, *Price*. Click **Add** to add each of these elements.

6. Save your changes and then, in the **MDM Repository** tree view, expand the **View** node, right-click the *Product* View, and then select one of the deployment options to deploy your changes to the MDM Server.

   You have now created the overall structure for your simple MDM project.

4.2. Working with the data in the Web User Interface

Once you have prepared your simple project in *Talend Studio*, you can start to interact with the data in *Talend MDM Web User Interface*, as your business users would do in a real-life project.

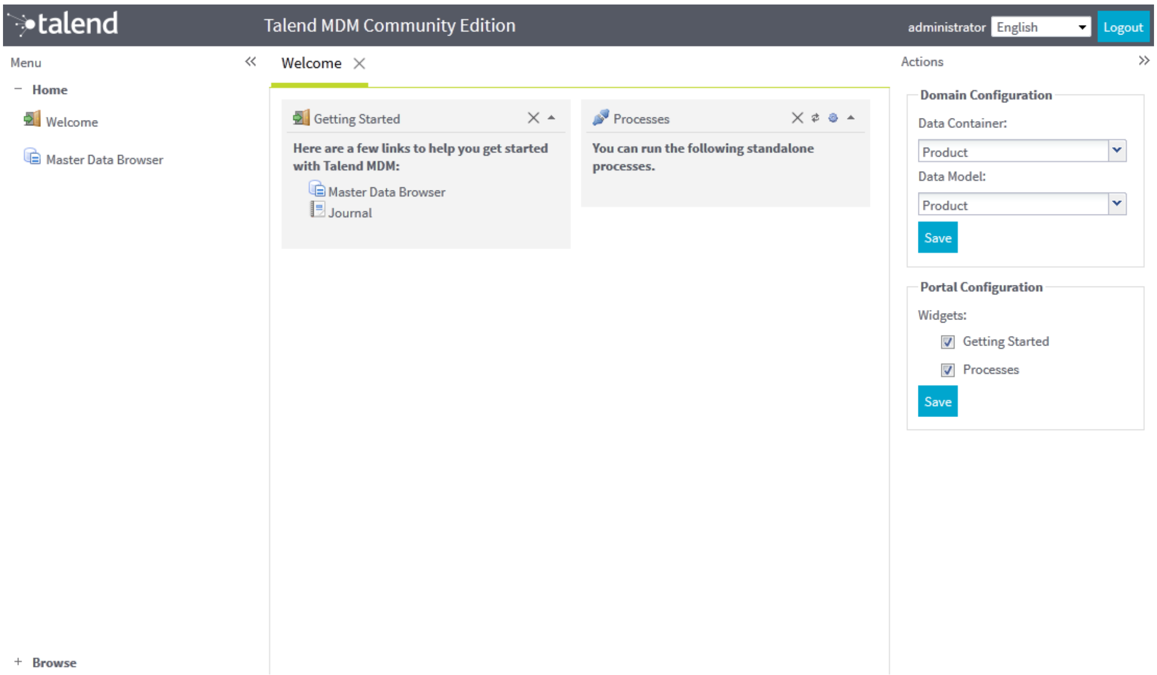
4.2.1. Opening MDM Web User Interface

To open *Talend MDM Web User Interface*, do the following:

1. In a web browser, enter the URL for your MDM Server, for instance `http://localhost:8180/talendmdm/ui`.

2. On the authentication page, enter the default administrator user name and passowrd, *administrator/administrator*, and then click the **Login** button.

   The Welcome Page opens.



Here are a few links to help you get started with Talend MDM:

You can run the following standalone processes.

Data Container: Product

Data Model: Product

Widgets:

☑ Getting Started

☑ Processes

4.2.2. Creating a new data record

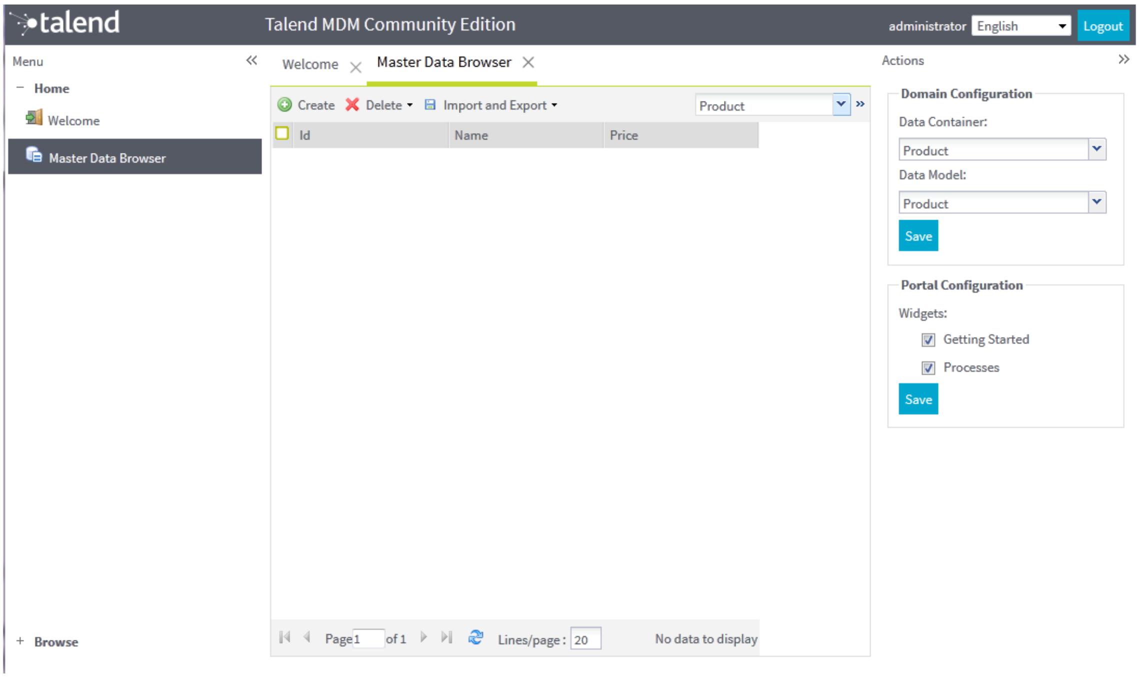
In this scenario, you are going to create a single new data record in the *Product* business entity that corresponds to the criteria you defined earlier in the studio.

Firstly, in the **Domain Configuration** area, select *Product* from the **Data Container** list, and its corresponding data model *Product* is automatically selected from the **Data Model** list. Click **Save** to validate your changes.

To create a new data record, do the following:

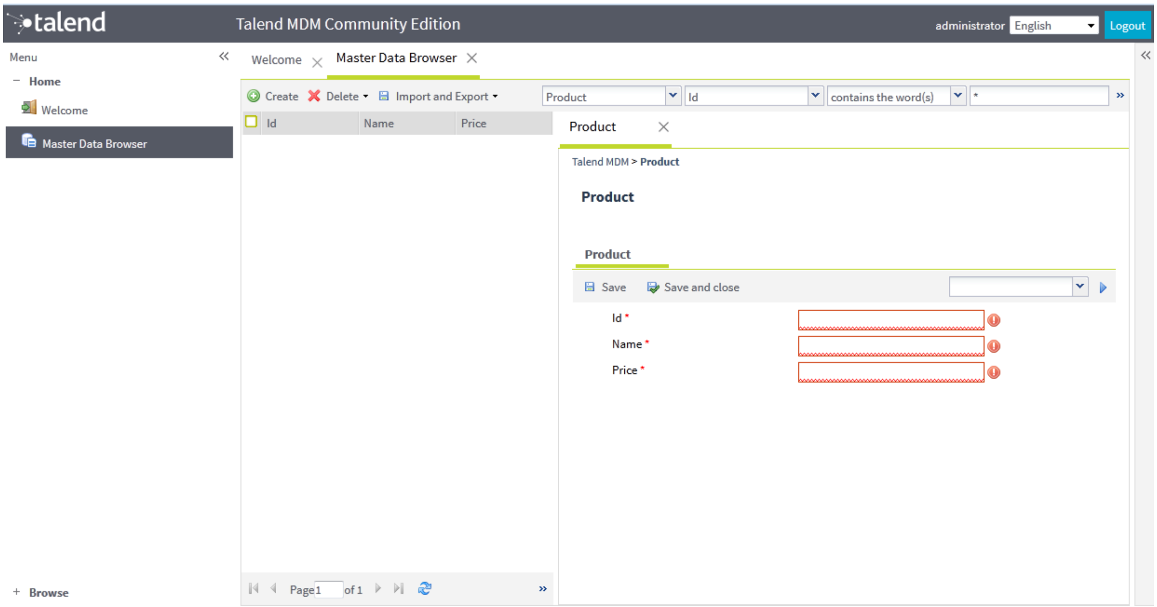
1. In the **Menu** panel, click **Master Data Browser** to open the Master Data Browser.

2. Select the *Product* entity from the **Select an entity** drop-down list.



3. For the moment, no data records are displayed.

Click the **Create** button to create a new empty record.



4. In the new data record that opens, you can see some fields corresponding to the attributes you defined in the studio. Enter any values you like for the *Id* and *Name* fields, and any numerical value you like for the *Price* field, and then click **Save and close** to create your data record.

You have now created a data record that validates against all the criteria you defined in the studio.