



Talend Open Studio for Big Data

Installation and Upgrade Guide

6.1.2

Adapted for v6.1.2. Supersedes previous releases.

Publication date: September 13, 2016

Copyright

This documentation is provided under the terms of the Creative Commons Public License (CCPL).

For more information about what you can and cannot do with this documentation in accordance with the CCPL, please read: <http://creativecommons.org/licenses/by-nc-sa/2.0/>

Notices

Talend is a trademark of Talend, Inc.

All brands, product names, company names, trademarks and service marks are the properties of their respective owners.

License Agreement

The software described in this documentation is licensed under the Apache License, Version 2.0 (the "License"); you may not use this software except in compliance with the License. You may obtain a copy of the License at <http://www.apache.org/licenses/LICENSE-2.0.html>. Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

This product includes software developed at AOP Alliance (Java/J2EE AOP standards), ASM, Amazon, AntLR, Apache ActiveMQ, Apache Ant, Apache Avro, Apache Axiom, Apache Axis, Apache Axis 2, Apache Batik, Apache CXF, Apache Cassandra, Apache Chemistry, Apache Common Http Client, Apache Common Http Core, Apache Commons, Apache Commons Bcel, Apache Commons JXPath, Apache Commons Lang, Apache DataFu, Apache Derby Database Engine and Embedded JDBC Driver, Apache Geronimo, Apache HCatalog, Apache Hadoop, Apache Hbase, Apache Hive, Apache HttpClient, Apache HttpComponents Client, Apache JAMES, Apache Log4j, Apache Lucene Core, Apache Neethi, Apache Oozie, Apache POI, Apache Parquet, Apache Pig, Apache PiggyBank, Apache ServiceMix, Apache Squirrel, Apache Thrift, Apache Tomcat, Apache Velocity, Apache WSS4J, Apache WebServices Common Utilities, Apache Xml-RPC, Apache Zookeeper, Box Java SDK (V2), CSV Tools, Cloudera HTrace, ConcurrentLinkedHashMap for Java, Couchbase Client, DataNucleus, DataStax Java Driver for Apache Cassandra, Ehcache, Ezmorph, Ganymed SSH-2 for Java, Google APIs Client Library for Java, Google Gson, Groovy, Guava: Google Core Libraries for Java, H2 Embedded Database and JDBC Driver, Hector: A high level Java client for Apache Cassandra, Hibernate BeanValidation API, Hibernate Validator, HighScale Lib, HsqlDB, Ini4j, JClouds, JDO-API, JLine, JSON, JSR 305: Annotations for Software Defect Detection in Java, JUnit, Jackson Java JSON-processor, Java API for RESTful Services, Java Agent for Memory Measurements, Jaxb, Jaxen, JetS3T, Jettison, Jetty, Joda-Time, Json Simple, LZ4: Extremely Fast Compression algorithm, LightCouch, MetaStuff, Metrics API, Metrics Reporter Config, Microsoft Azure SDK for Java, Mondrian, MongoDB Java Driver, Netty, Ning Compression codec for LZ4 encoding, OpenSAML, Paracel JDBC Driver, Parboiled, PostgreSQL JDBC Driver, Protocol Buffers - Google's data interchange format, Resty: A simple HTTP REST client for Java, Riak Client, Rocoto, SDSU Java Library, SL4J: Simple Logging Facade for Java, SQLite JDBC Driver, Scala Lang, Simple API for CSS, Snappy for Java a fast compressor/decompressor, SpyMemCached, SshJ, StAX API, StAXON - JSON via StAX, Super SCV, The Castor Project, The Legion of the Bouncy Castle, Twitter4J, Uuid, W3C, Windows Azure Storage libraries for Java, Woden, Woodstox: High-performance XML processor, Xalan-J, Xerces2, XmlBeans, XmlSchema Core, Xmlsec - Apache Santuario, YAML parser and emitter for Java, Zip4J, atinject, dropbox-sdk-java: Java library for the Dropbox Core API, google-guice. Licensed under their respective license.

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. Prior to installing the Talend products	1
1.1. Installation requirements	2
1.2. Studio specific prerequisites	3
1.3. Compatible Platforms and Java environments	3
Chapter 2. Installing Talend Studio for the first time	5
2.1. Downloading and installing Talend Studio	6
2.2. Installing the XULRunner package (for Linux users)	6
2.3. Installing the MDM server	7
2.3.1. Installing MDM modules using the jar file	7
2.4. Launching Talend Studio	11
2.4.1. Launching the Studio	11
2.5. Installing external modules	12
2.5.1. Identify required external modules	12
2.5.2. Install external modules	15
Chapter 3. Upgrading your Talend products	17
3.1. Backing up the environment	18
3.2. Upgrading the Talend projects in the Studio	18
Appendix A. Supported Third-Party System/Database/Business Application Versions	19
A.1. Supported systems, databases and business applications by Talend components	20
A.2. Supported Hadoop distribution versions	22

Preface

1. General information

1.1. Purpose

This Installation Guide explains how to install configure and upgrade the *Talend* modules and related applications. For detailed explanation on how to use and fine-tune the *Talend* applications, please refer to the appropriate Administrator or User Guides of the *Talend* solutions.

Information presented in this document applies to *Talend* products **6.1.2**.

1.2. Audience



This guide is for administrators and users of the *Talend* products.



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. Prior to installing the Talend products

This chapter provides useful information on software and hardware prerequisites you should be aware of, prior to starting the installation of the *Talend* modules.



In the following documentation:

- recommended: designates an environment recommended by *Talend* based on our experiences and customer usage;
- supported: designates a supported environment for use with the listed component or service;
- supported with limitations: designates an environment that is supported by *Talend* but with certain conditions explained in notes.

1.1. Installation requirements

To make the most out of the *Talend* products, please consider the following hardware and software requirements.

Memory usage heavily depends on the size and nature of your *Talend* projects. However, in summary, if your Jobs include many transformation components, you should consider upgrading the total amount of memory allocated to your servers, based on the following recommendations.

Table 1.1. Memory usage

Product	Client/Server	Recommended alloc. memory
<i>Studio</i>	Client	3GB minimum, 4 GB recommended

The same requirements also apply for disk usage. It also depends on your projects but can be summarized as:

Table 1.2. Disk usage

Product	Client/Server	Required disk space for installation	Required disk space for use
<i>Studio</i>	Client	3GB	3+ GB

Ulimit settings on Unix systems

- To make the most out of the *Talend* server modules and improve performance on Unix systems, you should set the limit of system resources (ulimit) to unlimited.

Environment variable configuration: on Windows

Prior to installing your *Talend* solutions, you have to set the `JAVA_HOME` Environment variable:

- Define your `JAVA_HOME` environment variable so that they point to the JRE directory.

Example: if the JRE path is `C:\Java\JREx.x.x\bin`, you must set the `JAVA_HOME` environment variable to point to: `C:\Java\JREx.x.x`.



It is highly recommended that the full path to the server installation directory is as short as possible and does not contain any space character. If you already have a suitable JRE installed in a path with a space, you simply need to put quotes around the path when setting the values for the environment variable.

Environment variable configuration: on Linux

Prior to installing your *Talend* solutions, you have to set the `JAVA_HOME` and `Path` environment variables:

- Find the JRE home dir and export it in the `JAVA_HOME` environment variable.

Example:

```
export JAVA_HOME=/usr/lib/jvm/jre1.8.0_65
export PATH=$JAVA_HOME/bin:$PATH
```

- Add these lines at the end of the global profiles in the `/etc/profile` file or in the user profiles in the `~/.profile` file.

Note that after changing one of these files you have to log on again.

For more information on how to set the `JAVA_HOME` variables on Unix and Windows systems, see the online [Oracle documentation](#).

1.2. Studio specific prerequisites

Some bulk components, like Oracle, Sybase, Informix or Ingres, require database client software to run properly.

Install the following utilities according to your needs:

- OracleBulkExec uses the *sqlldr* external utility. This utility is available in Oracle clients that must be installed on the computer.
- Informix uses the *dbload* external utility.
- Ingres uses the *sql* external utility.
- Sybase uses the *bcp.exe* external utility. This utility is asked for in the Sybase bulk components' **Basic Settings** view. For more information, see *tSybaseBulkExec*, *tSybaseOutputBulk* and *tSybaseOutputBulkExec* components on the appropriate *Talend Components Reference Guide*.

1.3. Compatible Platforms and Java environments

Please refer to the following grids for a summary of supported OS and Java Runtime environments.

Table 1.3. Java environments

Support type	Vendor	Version
Recommended	Oracle Java	8

For more information on Java specificities (version, OS compatibility), please see the Knowledge Base articles on [Talend Help Center](#).

Table 1.4. Talend Studio

Support type	OS	Version	Processor	
Recommended	Unix	Linux Ubuntu	12.04	64-bit
	Windows	Microsoft Windows Professional	7	64-bit
Supported	Unix	Linux Ubuntu	16.04	64-bit
		Linux Ubuntu	14.04	64-bit
		Redhat Linux Enterprise Server Edition/ CentOS	7.1	64-bit
		Redhat Linux Enterprise Server Edition/ CentOS	6.5/6.6	64-bit
		SUSE SLES	11	64-bit
	Windows	Microsoft Windows	10	64-bit
		Microsoft Windows	8.1	64-bit
		Microsoft Windows	8	64-bit
	Mac OS	OS X	El Capitan/10.11	64-bit
		OS X	Yosemite/10.10	64-bit
OS X		Mavericks/10.9	64-bit	
Deprecated	Mac OS	OS X	Mountain Lion/10.8	64-bit
		OS X	Lion/10.7	64-bit



Chapter 2. Installing Talend Studio for the first time

We strongly encourage you to read the [Prior to installing the Talend products](#) before starting this chapter.

This chapter details the procedures required to install Talend Studio.

2.1. Downloading and installing Talend Studio

Download

1. Download the product on [this page](#).

Note that the `.zip` file contains binaries for ALL platforms (Linux/Unix, Windows and MacOS).

2. Once the download is complete, extract the archive file on your hard drive.



It is recommended to avoid spaces and long names in the target installation directory path.

Configure the memory settings

- If you want to tune the memory allocation for your JVM, you only need to edit the `.ini` file corresponding to your executable file. For example:

- For *Talend Studio* on 32bit-Windows, edit the file: `TOS_BD-win32-x86.ini`;
- For *Talend Studio* on Linux, edit the file: `TOS_BD-linux-gtk-x86.ini`.

The default values are:

```
-vmargs -Xms40m -Xmx500m -XX:MaxMetaspaceSize=128m
```

If you only have 512MB of memory on your computer, you can specify the memory allocation as following, for example:

```
-vmargs -Xms40m -Xmx256m -XX:MaxMetaspaceSize=64m
```

Learn more on <http://www.oracle.com/technetwork/java/hotspotfaq-138619.html>

2.2. Installing the XULRunner package (for Linux users)

On Linux, the XULRunner package is required to run the Studio.

The XULRunner package version that is recommended is XULRunner v1.9.2.28.

The supported versions are v1.8.x - 1.9.x and v3.6.x.

1. Download XULRunner version from [this location](#).
2. Unpack the archive file in the same directory where you unpacked the studio archive, but do not unpack it within the Studio folder.
3. Add the following line at the end of the Studio `.ini` file that corresponds to your Linux architecture:

```
-Dorg.eclipse.swt.browser.XULRunnerPath=</usr/lib/xulrunner>
```

where `</usr/lib/xulrunner>` is the *xulrunner* installation path.

For example, if you have unpacked the Studio in a directory under your user home directory `/home/<user>/Talend/`, you need to add the following to the `.ini` file:

```
-Dorg.eclipse.swt.browser.XULRunnerPath=/home/<user>/Talend/xulrunner/
```

2.3. Installing the MDM server

You can install the MDM server using a graphical installer.



In the following documentation, `<TomcatPath>` designates the path where Tomcat has been installed, and `<$INSTALLDIR>` specifies the path where the MDM server has been installed.

2.3.1. Installing MDM modules using the jar file

The jar file allows you to launch a cross-platform MDM-dedicated graphical installer to install the Tomcat server (optional) and deploy the MDM Server in simple click-next steps.

Using the jar file provided by **Talend**, you can install the MDM modules in three different modes:

- A cross-platform graphical installer to help you install Tomcat (optional) and deploy the MDM Server in simple click-next steps. For further information, see [Installing in GUI mode](#).
- Otherwise, open your command-line and use the command: `java -jar <jar name>.jar -console`, and then follow the instructions to complete the installation of the MDM server and the stewardship console. For further information, see [Installing in Command/Console mode](#).
- You can also generate a script through selecting the corresponding button in the last panel of the graphical installer. Script generation requires running the graphical installer, but the generated script can be used from the command-line. The installation can be run by passing the installation script as a parameter to the installer. For further information, see [Installing in Silent mode](#).

The sections below explain in detail these three installation modes.

2.3.1.1. Installing in GUI mode

Talend Studio and *Talend MDM Web User Interface* require that you install an MDM server.

Prerequisite(s):

- JRE 1.8.0 or higher must be installed. You should also make sure that the `JAVA_HOME` environment variable is set to point to the JRE directory.

For example, if the path is `C:\Java\JREx.x.x\bin`, you must set the `JAVA_HOME` environment variable to point to: `C:\Java\JREx.x.x`.

- Make sure you have read and write access to the folder where the MDM server will be installed (`C:\Talend` for example).
- (Linux only) A Windows Manager must be installed.



- It is highly recommended that the full path to the server installation directory is as short as possible and does not contain any space character.

- If you already have a suitable JRE installed in a path with a space, you need to put quotes around the path when setting the values for the environmental variable.

To install the MDM server using a `.jar` file, follow the procedures below.

Launching the installer

1. Right-click the *.jar* file and select with your Java runtime environment to run the installer.

A language selection pop-up opens.

2. From the language selection pop-up, select an installation language from the list and click **OK** to close the pop-up and proceed to the next step.
3. On the **Talend MDM** welcome page, click **Next** to proceed to the next step.
4. Read the license agreement and select the accept option. Click **Next** to proceed to the next step.
5. Read the license information and click **Next** to proceed to the next step.

Selecting the packs

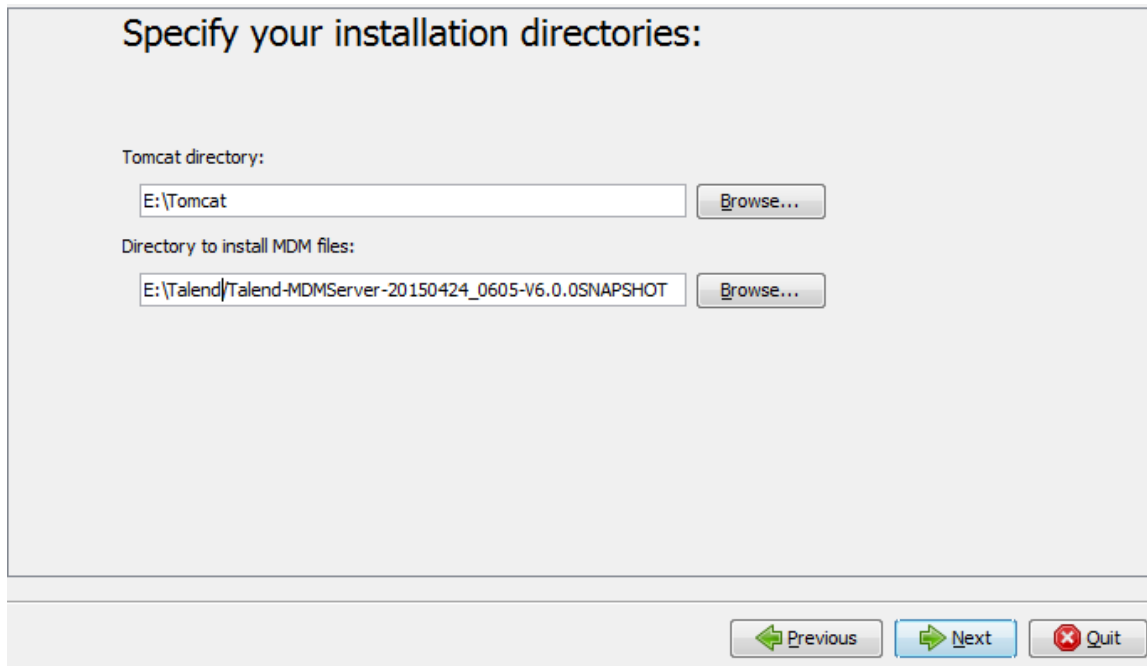
1. Select the check boxes of the packs you want to install, and then click **Next** to proceed to the next step.

If you have a Tomcat application server already installed on your machine and you do not want to re-install it, clear the **Tomcat** check box.

The check boxes of required packs are already selected and unavailable: **Talend MDM Application**.

2. Once you have made your choice, click **Next** to proceed to the next step.
3. Browse to where you want to install the selected packs, and then click **Next** to proceed to the next step.

If you cleared the Tomcat check box, you also need to specify the directory where Tomcat is installed.



A message is displayed to inform you that the target directory will be created.

4. Click **OK** to close the message and proceed to the next step.

Selecting the MDM server ports configuration

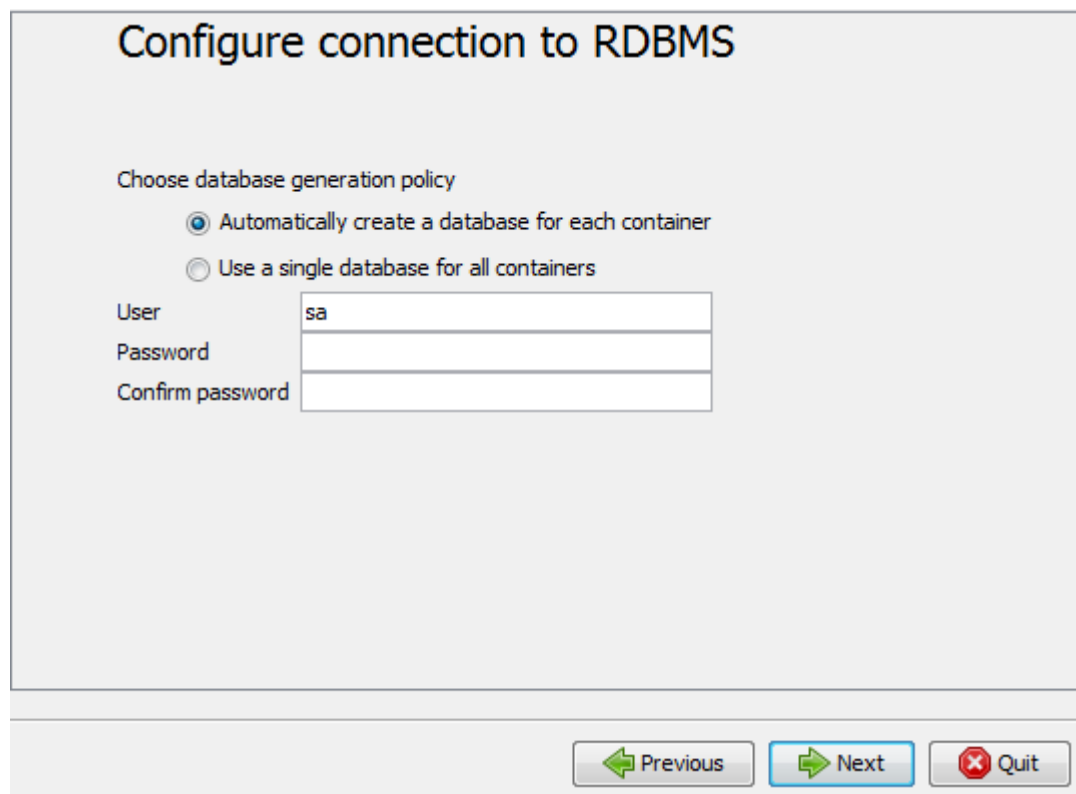
- From the **Ports** list, select the set of ports to be used by the MDM server, and then click **Next**. By default, **ports-1 (8180)** is selected.

Selecting and configuring the database

1. Select the type of database you want to use from the **Database Type** drop-down list and then click **Next** to proceed to the next step.
2. If you selected an RDBMS database, configure its connection parameters:

When H2 Embedded is selected:

1. Choose the database generation policy. You have a choice between creating a database for each container, or using a single database for all containers.
2. Enter the **User name** and **Password** to use for connecting to the database.



Configure connection to RDBMS

Choose database generation policy

Automatically create a database for each container

Use a single database for all containers

User

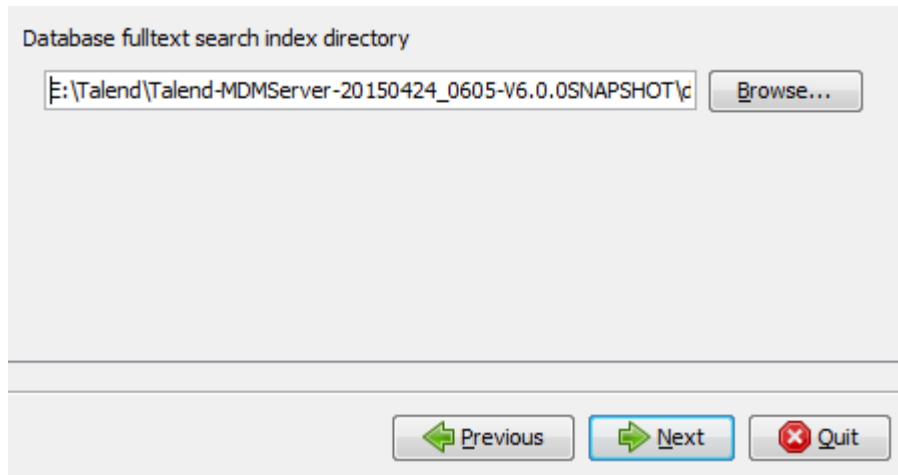
Password

Confirm password

3. Click **Next** to proceed to the next step.

Specifying the location for the fulltext search index directory

1. Browse to the directory where you want to store the database fulltext search index.



2. Click **Next** to proceed to the next step.

A message is displayed to inform you that the target directory will be created.

The installation settings you defined in the previous steps will be displayed in the wizard.

1. Click **Next** to proceed to the next step and start the installation.

Two progress bars indicate how much of the installation has been completed.

2. When the progress bars indicate the end of the installation, click **Next** to have a confirmation message that the installation is completed successfully.



Click **Generate an automatic installation script** to generate a script for MDM server installation.

You can use this script to install the MDM server on any platform without using the graphical installer.

3. Click **Done** to close the installer.

The MDM server is installed.

To run the MDM server, execute *startup.bat* (Windows) or *startup.sh* (Linux) in the `<TomcatPath>/bin` folder.

To shut the MDM server down, press **Ctrl + C** in the console window, or run *bin/shutdown.bat* or *bin/shutdown.sh*.

2.3.1.2. Installing in Command/Console mode

You can install the MDM server in a non-GUI mode using the command-line.

Prerequisite(s):

- JRE 1.8.0 or higher must be installed. You should also make sure that the `JAVA_HOME` environment variable is set to point to the JRE directory.

For example, if the path is `C:\Java\JREx.x.x\bin`, you must set the `JAVA_HOME` environment variable to point to: `C:\Java\JREx.x.x`.

- Make sure you have read and write access to the folder where the MDM server will be installed (`C:\Talend` for example).
- (Only Linux) A windows manager must be installed.



- It is highly recommended that the full path to the server installation directory is as short as possible and does not contain any space character.

- If you already have a suitable JRE installed in a path with a space, you simply need to put quotes around the path when setting the values for environmental variable.

To use the command-line capabilities to install the MDM server:

1. Unzip the .jar server file provided by *Talend*.
2. Open your console server depending on the platform you have.
3. Enter the below command, and then press the **Enter** key on your keyboard to launch the installation procedure through this text-only interface.

```
java -jar <jar name>.jar -console
```

4. Follow the instructions to install the MDM server.

2.3.1.3. Installing in Silent mode

When installing the MDM server using the graphical installer as outlined in [Installing in GUI mode](#), you can generate an installation script at the last step in the graphical installer. This script is an XML file that contains the installation configuration settings.

You can later use this generated script from your console to install the MDM server. The installation can be run by passing the installation script as a parameter to the installer as the following:

```
java -jar <jar name>.jar -<script name>
```

2.4. Launching Talend Studio

The following procedures describe how to launch the Studio, how to log in the Studio and how to create your first project.

2.4.1. Launching the Studio

Launch the Studio

- On Windows, double-click the executable file to launch *Talend Studio*.

On Unix-like systems, add execution rights on the desired *TOS_BD-** binary before launching it.

On a standard Linux distribution, the command is:

```
$ chmod +x TOS_BD-linux-gtk-x86.sh
$ ./TOS_BD-linux-gtk-x86.sh
```

On Mac OS X, launch the following file:

```
TOS_BD-macosx-cocoa.app/Contents/MacOS/TOS_BD-macosx-cocoa
```

Public license

- First screen is a license screen. In the **[License]** window that appears, read and accept the terms of the license agreement to proceed to the next step.

Login and first project

1. As first time user, you need to set up a new project or you can also import a Demo project which gathers numerous job samples.

Select what you want to do next:

Create a new project:

Import a demo project

Import an existing project

Always ask me at startup

2. To import a demo project, select **Import a demo project** and click **Finish**. In the dialog box that opens, select the project you want to import.

To create a new project, select **Create a new project** and enter the name of your project in the corresponding field.

3. Click **Finish** when complete, to open a welcome window and launch the Studio.

2.5. Installing external modules

Talend Studio requires specific third-party Java libraries or database drivers (*.jar* files) to be installed to connect to sources and targets. Those libraries or drivers, known as external modules, can be required by some of *Talend* components. Due to license restrictions, *Talend* may not be able to ship certain external modules within *Talend Studio*.

2.5.1. Identify required external modules

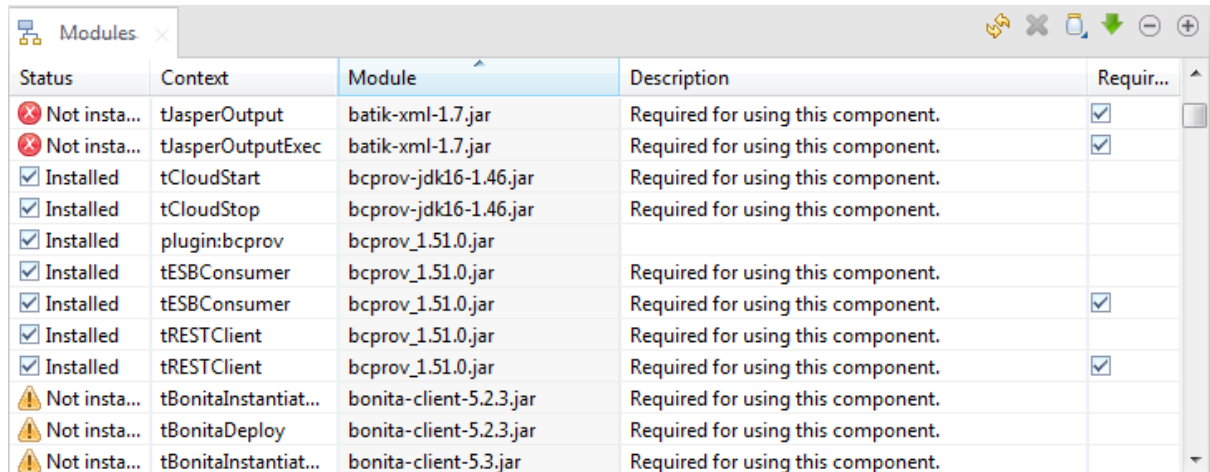
When you launch *Talend Studio* or select **Help > Install Additional Packages** in the *Talend Studio* menu, if any external modules are found missing for any features, the **[Additional Talend packages]** wizard opens, showing the **Optional** and **Required third-party libraries** check boxes. Make sure these check boxes are selected and click **Finish** to open the **[Download external modules]** dialog box, which lists all the available external modules, displays the license terms under which the external modules are provided, and lets you install all the modules at a single click. For more information, see [Install external modules](#).

On your design workspace, if a component requires the installation of external modules before it can work properly, a red error indicator appears on the component. With your mouse pointer over the error indicator, you can see a tooltip message showing which external modules are required for that component to work.

When you open the **Basic settings** or **Advanced settings** view of a component for which one or more external modules are required, you will see a piece of highlighted information about external modules, followed by an **Install** button. Clicking the **Install** button opens a wizard that will show you the external modules to be installed.




The **Modules** view lists all the modules required to use the components embedded in the Studio, including those Java libraries and drivers that you must install to get the relevant components working.

If the **Modules** view is not shown under your design workspace, go to **Window > Show View... > Talend** and then select **Modules** from the list.




Status	Context	Module	Description	Requir...
Not insta...	tJasperOutput	batik-xml-1.7.jar	Required for using this component.	<input checked="" type="checkbox"/>
Not insta...	tJasperOutputExec	batik-xml-1.7.jar	Required for using this component.	<input checked="" type="checkbox"/>
Installed	tCloudStart	bcprov-jdk16-1.46.jar	Required for using this component.	<input type="checkbox"/>
Installed	tCloudStop	bcprov-jdk16-1.46.jar	Required for using this component.	<input type="checkbox"/>
Installed	plugin:bcprov	bcprov_1.51.0.jar		<input type="checkbox"/>
Installed	tESBConsumer	bcprov_1.51.0.jar	Required for using this component.	<input type="checkbox"/>
Installed	tESBConsumer	bcprov_1.51.0.jar	Required for using this component.	<input checked="" type="checkbox"/>
Installed	tRESTClient	bcprov_1.51.0.jar	Required for using this component.	<input type="checkbox"/>
Installed	tRESTClient	bcprov_1.51.0.jar	Required for using this component.	<input checked="" type="checkbox"/>
Not insta...	tBonitaInstantiat...	bonita-client-5.2.3.jar	Required for using this component.	<input type="checkbox"/>
Not insta...	tBonitaDeploy	bonita-client-5.2.3.jar	Required for using this component.	<input type="checkbox"/>
Not insta...	tBonitaInstantiat...	bonita-client-5.3.jar	Required for using this component.	<input type="checkbox"/>

The table below describes the information presented in the **Modules** view.

Column	Description
Status	points out if a module is installed or not installed on your system. The  icon indicates that the module is not necessarily required for the corresponding component listed in the Context column. The  icon indicates that the module is absolutely required for the corresponding component.
Context	lists the name of Talend component using the module. If this column is empty, the module is then required for the general use of <i>Talend Studio</i> .  This column lists any external libraries added to the routines you create and save in the Studio library folder. For more information, see the <i>Talend Studio User Guide</i> .
Module	lists the module exact name.
Description	explains why the module/library is required.
Required	the selected check box indicates that the module is required.

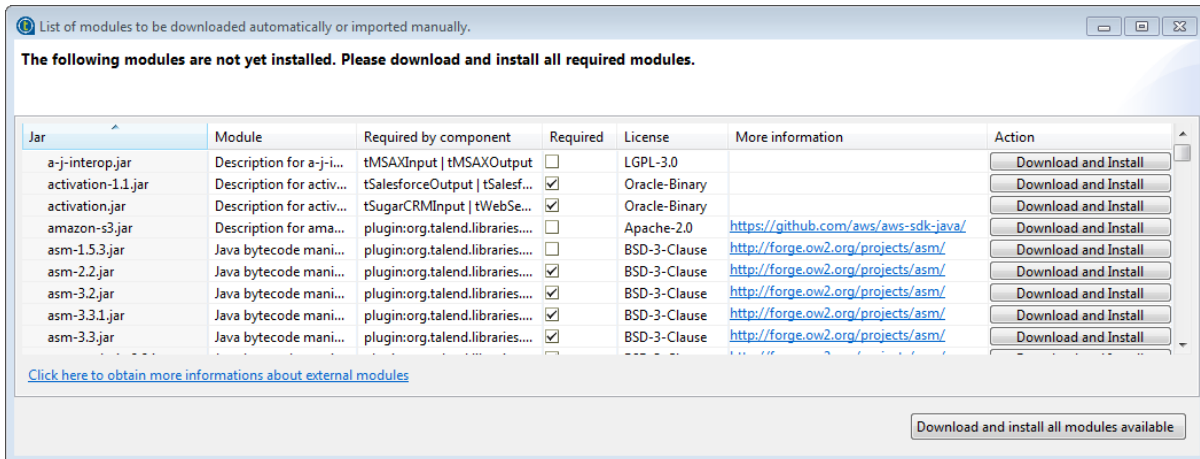
In addition to the **Modules** view, the Studio provides a mechanism that enables you to easily identify, download and install most of the required third-party modules from the **Talend** website and directs you to valid websites for the rest.

A Jar installation wizard appears whenever any required external module is found missing for any feature in the Studio, including when you:

- drop a component from the **Palette** if one or more external modules required for that component to work are missing in the Studio, or
- click the **Guess schema** button in the **Component** view of a component if one or more external modules required for that component to work are missing in the Studio,
- click **Install** on the top of the **Basic settings** or **Advanced settings** view of a component for which one or more required external modules are missing,
- run a Job that involves components for which one or more required external modules are missing, or
- click the  button in the **Modules** view.



When you click this button, the wizard that appears will list all the required external modules that are not integrated in the Studio.



The table below describes the information presented in the wizard.

Item	Description
Jar	The file name of the external module.
Module	A short description about the nature of the module.
Required by component	Lists the components that require the external module.
Required	The selected check box indicates that the module is required.
License	The license under which the module is provided.
More information	Provides the URL of the valid website where you can find more information about this module and download the module manually.
Action	<p>: Click to open the [Download external modules] dialog box to download and install the module, which is available on the Talend website;</p> <p>Open in browser : Click the link to open the valid website to download the module, which is not available on the Talend website, and then click the jar button to import the downloaded module into your studio;</p> <p>: You need to find and download the module yourself and click the jar button to import it into the your studio.</p>
	Click to open the [Download external modules] dialog box to download and install all the required modules that are available on the Talend website.
Do not show again	<p>Select to prevent the wizard from appearing again unless you click the button in the Modules tab view.</p> <p>This check box shows only when you drop a component, or guess the schema of a database, that requires an external module, or click the Install button on the Component tab of a component that requires an external module.</p>
Click here to obtain more information about external modules	Click to go to Talend online documentation on installing third-party modules.

This wizard lists the external modules to be installed, the licenses under which they are provided, and the URLs of the valid websites where they are downloadable, and allows you to download and install automatically all the modules available on the Talend website and download those not available on the **Talend** website by following the links provided in the **Action** column and then install them into your Studio manually.

When you drop a component, or guess the schema of a database, that requires an external module for which neither the Jar file nor its download URL information is available on the **Talend** website, the Jar installation wizard does

not appear, but the **Error Log** view will present an error message informing you that the download URL for that module is not available. You can try to find and download it by yourself, and then install it manually into the Studio.



To show the **Error Log** view on the tab system, go to **Window > Show views**, then expand the **General** node and select **Error Log**.

2.5.2. Install external modules

To download and install modules in the Studio

To download and install external modules automatically, do the following:

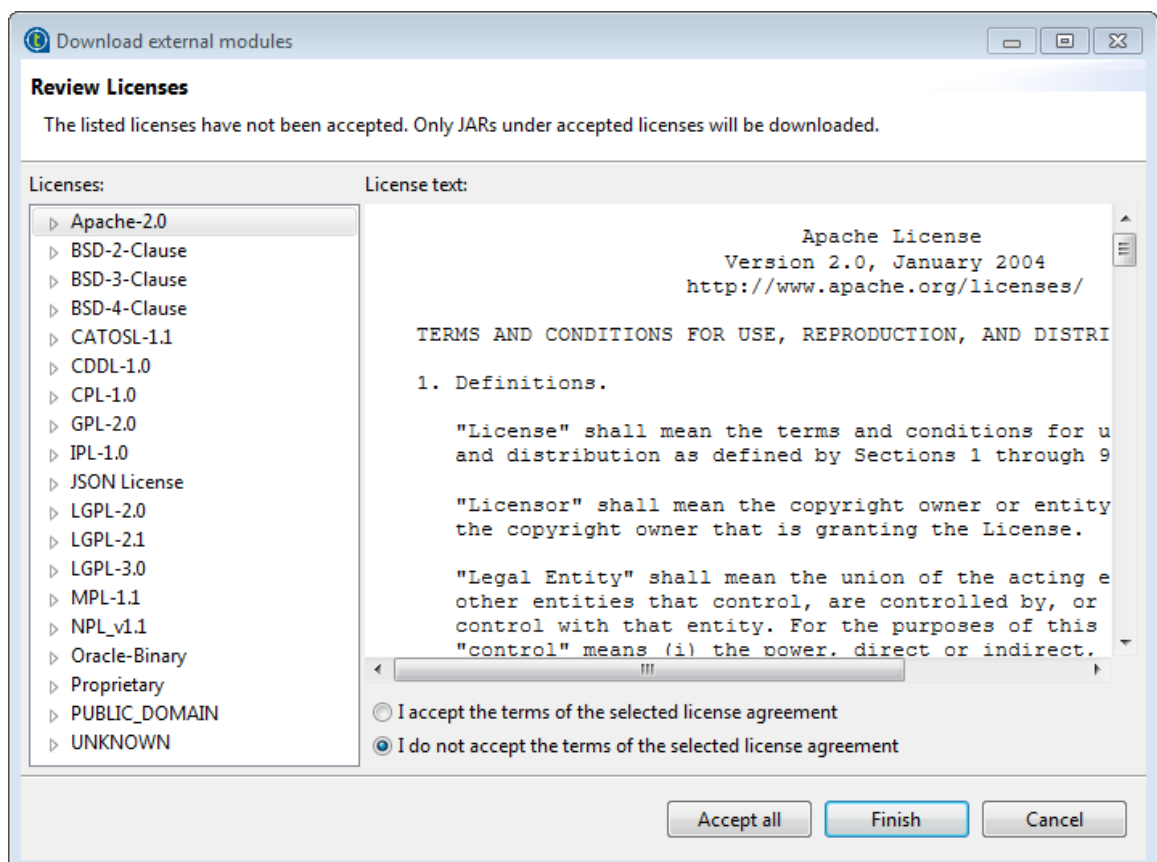


*If you are working behind a network proxy, make sure you have correctly set up your proxy before you can download and install external modules in your Studio. To access the proxy settings, select **Window > Preferences** from the menu to open the [Preferences] window, then expand the **General** node and click **Network Connections**.*

1. In the Jar installation wizard, click the **Download and Install** button to install a particular module, or click the **Download and install all modules available** button to install all the available modules, or select **Help > Install Additional Packages** from the menu to open the [Additional Talend Packages] wizard. From this wizard, make sure the Optional and Required third-party libraries check boxes are selected and click **Finish**. The [Download external modules] dialog box opens.



This [Additional Talend Packages] wizard appears automatically when you launch *Talend Studio* if any additional packages, including external modules, need to be installed for any features to function in the Studio.



2. To download and install the external module(s) provided under a particular license, select that license from the **Licenses** pane, review the license terms, select the **I accept the terms of the license agreement** option, and click **Finish** to start the download and installation process.


To download and install all external modules provided under all the listed licenses, click the **Accept all** button to start the download and installation process.

Upon installation of the chosen external module or modules, a dialog box appears to notify you about the number of modules successfully installed and/or about the modules failed to install, if any.

To install manually an external module you already have in your local file system, do the following:



Talend Open Studio for Big Data does not come with the JDBC drivers for Oracle databases due to Apache license restrictions. For Oracle9i, the required JDBC driver downloadable from Oracle website is named *ojdbc14.jar*, the same as that for Oracle 10g. To enable the JDBC driver for Oracle9i you have downloaded to work in **Talend Open Studio for Big Data**, you have to change the file name to *ojdbc14-9i.jar* before installing it into the Studio.

1. Click the  button in the upper right corner of the **Modules** view or in Jar installation wizard to browse your local file system.

If the **Modules** view is not shown under your design workspace, go to **Window > Show View... > Talend** and then select **Modules** from the list.

2. In the **[Open]** dialog box of your file system, browse to the module you want to install, double-click the *.jar* file, or select it and then click **Open** to install it.

The dialog box closes and the selected module is installed in the library folder of the current Studio.

You can now use the component dependent on this module in any of your Job designs.

To install modules downloaded from external websites

Some modules are not available on the **Talend** website but can be downloaded directly from external websites. Once downloaded, these modules must be placed in specific folders.

- For the studio, the downloaded modules must be placed in the following folder:

<StudioPath>/configuration/.m2



Chapter 3. Upgrading your Talend products

This chapter describes the various operations required to migrate version of the *Talend* solutions.

We assume that you have installed and configured these solutions as described in the [Installing Talend Studio for the first time](#).

The migration and upgrade process includes the following mandatory steps:



These steps usually need to be completed in the following order.

1. Backing up the environment, see the [Backing up the environment](#).
2. Upgrading the Talend projects in the Studio, see the [Upgrading the Talend projects in the Studio](#).

3.1. Backing up the environment

Before you start migrating your *Talend* solutions, make sure your environment is correctly backed up.


The environment backup process includes the following mandatory steps:



These steps usually need to be completed in the following order.

1. Saving the local projects, see [Saving the local projects](#).

Saving the local projects

1. Launch the Studio.
2. Click the  icon and export your local projects to an archive file.

3.2. Upgrading the Talend projects in the Studio

Importing your local projects

1. Launch the new Studio you have just installed.
2. In the login window, select **Import**, then import the archive file containing your local projects.

The local projects are displayed in the **Project** list and appear on the Studio **Repository** view.



For more information on how to export local projects to an archive file, see [Saving the local projects](#).



Appendix A. Supported Third-Party System/ Database/Business Application Versions

This document provides the information about the versions of the systems or databases or business applications supported by Talend Studio.

A.1. Supported systems, databases and business applications by Talend components

The access to these systems, databases and business applications varies depending on the Studio you are using.

Systems/Databases	Versions	OS	Available with...
Alfresco	2.1	N/A ¹	All Talend products
Amazon Redshift	Initial release of Amazon Redshift	N/A ¹	All Talend products
AS/400	V5R2 to V5R4 V5R3 to V6R1 V6R1 to V7R2	N/A ¹	All Talend products
Access ²	2003 2007	Windows	Talend products with Data Integration (DI), Master Data Management (MDM), Enterprise Service Bus (ESB) or Big Data
Bonita	5.2.3 5.3.1 5.6.1 5.10.1 6.5.2	N/A ¹	All Talend products
Cassandra	2.0.0 (Deprecated versions: 1.1.2/1.2.2)	Windows + Linux	Talend products with Big Data
CouchBase	2.0	Windows	Talend products with Big Data
CouchDB	1.0.2	Windows	Talend products with Big Data
DB Generic	ODBC	Windows	All Talend products
DB2	10.5 10.1	Windows + Linux	Talend components with all products. Talend products with MDM or ESB.
EXASolution	4	Windows	Talend products with DI, MDM, ESB or Big Data
FireBird	2.1	Windows + Linux	Talend products with DI, MDM, ESB or Big Data
Greenplum	4.2.1.0	Windows (client only) + Linux	Talend products with DI, MDM, ESB or Big Data
HSQLDb	1.8.0	N/A ¹	Talend products with DI, MDM, ESB or Big Data
Informix	11.50	Windows + Linux	All Talend products
Ingres	9.2	Windows + Linux	All Talend products
Interbase	7 and above	N/A ¹	Talend products with DI, MDM, ESB or Big Data
JavaDB	6	Windows + Linux	Talend products with DI, MDM, ESB or Big Data
LDAP	No version limitation	Windows + Linux	All Talend products
Microsoft AX	Dynamics AX 4.0 Dynamics AX 2012	N/A ¹	All Talend products
Microsoft CRM	2011 2015	N/A ¹	All Talend products





Systems/Databases	Versions	OS	Available with...
MS SQL Server	2000 2003 2005 2008 2012	Windows + Linux	All Talend products
MaxDB	7.6	N/A ¹	Talend products with DI, MDM, ESB or Big Data
MongoDB	2.5.X 2.6.X 3.0.X	Windows + Linux	Talend products with Big Data
MySQL	Mysql4 Mysql5 MariaDB	Windows + Linux	All Talend products
Netezza	7.2	Windows + Linux	All Talend products
Neo4j	1.X.X 2.X.X 2.2.X	Linux	Talend products with Big Data
OleDb	2000 2003 2005 2007 2010	N/A ¹	All Talend products
Oracle	Oracle 8i/9i/10g/11g/11g (11.6)/12c	Windows + Linux	All Talend products
ParAccel	3.1 3.5	N/A ¹	Talend products with DI, MDM, ESB or Big Data
PostgreSQL	Prior to v9/v9.X	Windows + Linux	All Talend products
PostgresPlus	Prior to v9/v9.X	Windows + Linux	Talend products with DI, MDM, ESB or Big Data
Salesforce	V34 and earlier	Windows + Linux	All Talend products
SAP	4.6	Windows	All Talend products
SAP Hana	1.0	Windows	All Talend products
SQLite	3.6.7	Windows + Linux	All Talend products
Sybase	12.5 12.7 15.2 15.5 15.7	Windows + Linux	All Talend products
SybaseIQ	12.5 12.7 15.2	Windows + Linux	All Talend products

















































































Systems/Databases	Versions	OS	Available with...
Teradata	12 13 14 15	Windows + Linux	All Talend products
VectorWise	2	Windows + Linux	Talend products with DI, MDM, ESB or Big Data
Vertica	3 3.5 4 4.1 5.0 5.1 6.0 6.1.X 7.0.X	Windows + Linux	Talend products with DI, MDM, ESB or Big Data
VtigerCRM	Vtiger 5.0 Vtiger 5.1	N/A ¹	All Talend products

1. The test information is not available yet.

2. When working with Java 8, only the General collation mode is supported.

A.2. Supported Hadoop distribution versions

- : officially supported.
- : not officially supported.
- : the Kerberos kinit option is supported by the Studio.
- : the Kerberos kinit option and the Kerberos keytab option are both supported by the Studio.

		HBase	HCatalog	HDFS	Hive		Oozie	Pig	Spark		Sqoop	Talend MapReduce
					Hive1 ²	Hive2 ²			Standalone	YARN		
HDP	V1.2.0 (Bimota)											
												
	V1.3.0 (Condor)											
												
	V2.0.0 (BigWheel)											
												
	V2.1.0 (Baikal) ³											
												

		HBase	HCatalog	HDFS	Hive		Oozie	Pig	Spark		Sqoop	Talend MapReduce
					Hive1 ²	Hive2 ²			Standalone	YARN		
	V2.2.0 ³	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✗ ■■	✗ ■■	✓ ■■	✓ ■■
	V2.3.0	✓ ■■	✓ ■■	✓ ■■	✗ ■■	✓ ■■	✓ ■■	✓ ■■	✗ ■■	✓ ■■	✓ ■■	✓ ■■
Apache	1.0.0	✓ ■■	✗ ■■	✓ ■■	✓ ■■	✗ ■■	✗ ■■	✓ ■■	✗ ■■	✗ ■■	✗ ■■	✓ ■■
	0.92.0	■■	■■	■■	0.9.0	■■	■■	■■	■■	■■	■■	■■
Cloudera	CDH4 (from 4.0 to 4.2)	✓ ■■	✗ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✗ ■■	✗ ■■	✓ ■■	✓ ■■
	CDH4.3 + (YARN mode)	✓ ■■	✗ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✗ ■■	✗ ■■	✓ ■■	✓ ■■
	CDH 5.0 (YARN mode)	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✗ ■■	✗ ■■	✓ ■■	✓ ■■
	CDH 5.1 (MR mode)	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✗ ■■	✗ ■■	✓ ■■	✓ ■■
	CDH 5.1 (YARN mode)	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✗ ■■	✗ ■■	✓ ■■	✓ ■■
	CDH 5.4	✓ ■■	✓ ■■	✓ ■■	✗ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■
	CDH 5.5	✓ ■■	✓ ■■	✓ ■■	✗ ■■	✓ ■■	✗ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■
MapR	2.0.0	✓ ■■	✗ ■■	✓ ■■	✓ ■■	✗ ■■	✓ ■■	✓ ■■	✗ ■■	✗ ■■	✓ ■■	✓ ■■
	2.1.2	✓ ■■	✗ ■■	✓ ■■	✓ ■■	✗ ■■	✓ ■■	✓ ■■	✗ ■■	✗ ■■	✓ ■■	✓ ■■
	2.1.3	✓ ■■	✗ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✗ ■■	✗ ■■	✓ ■■	✓ ■■
	3.0.1	✓ ■■	✗ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✗ ■■	✗ ■■	✓ ■■	✓ ■■
	3.1.0	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✗ ■■	✗ ■■	✓ ■■	✓ ■■
	4.0.1 (YARN mode)	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✗ ■■	✗ ■■	✓ ■■	✓ ■■
	4.1.0 (YARN mode)	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■
5.0.0	✓ ■■	✓ ■■	✓ ■■	✗ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	✓ ■■	
Amazon EMR	Apache 1.0.3	✓ ■■	✗ ■■	✓ ■■	✓ ■■	✗ ■■	✗ ■■	✓ ■■	✗ ■■	✗ ■■	✗ ■■	✓ ■■

		HBase	HCatalog	HDFS	Hive		Oozie	Pig	Spark		Sqoop	Talend MapReduce
					Hive1 ²	Hive2 ₂			Standalone	YARN		
	Apache 2.4.0	✓	✗	✓	✓	✗	✗	✓	✗	✗	✓	✓
	EMR 4.0.0 (Apache 2.6.0)	✗	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓
Pivotal HD	1.0.1	✓	✗	✓	✓	✗	✗	✓	✗	✗	✓	✓
	2.0	✓ ☐☐	✗	✓ ☐☐	✓ ☐☐	✓ ☐☐	✓	✓ ☐☐	✗	✗	✓ ☐☐	✓ ☐☐
Microsoft HD Insight	3.1	✗	✗	✗	✓	✓	✓	✓	✗	✗	✗	✓
	3.2	✗	✗	✗	✓	✓	✗	✓	✓	✗	✗	✓
Custom ¹												

1. This enables the connection between the Studio and a custom Hadoop distribution not yet officially supported in the Studio. For further information, see the sections describing how to connect to a custom Hadoop distribution of the Talend Big Data Getting Started Guide or the documentation of any related component that creates the connection to a Hadoop distribution, such as tHDFSConnection.

2. In the Standalone mode, Hive 1 does not support the Kerberos security.

3. This option also allows you to connect to a Teradata Hadoop platform. For further information about the version of the Hortonworks Data Platform used in the Teradata platform you are using, see Teradata's related documentation.

For further information about the versions of all the supported third-party systems/databases, see [Supported systems, databases and business applications by Talend components](#).