



# **Talend Open Studio for Big Data Installation and Upgrade Guide for Windows**

**7.1.1**

# Contents

<b>Copyleft.....</b>	<b>3</b>
<b>Talend Open Studio for Big Data: Prerequisites.....</b>	<b>5</b>
Preparing your installation.....	5
Hardware requirements.....	5
Software requirements.....	6
Setting up JAVA_HOME.....	8
<b>Installing your Talend Open Studio for Big Data manually.....</b>	<b>9</b>
Installing and configuring your Talend Studio.....	9
<b>Upgrading your Talend products.....</b>	<b>16</b>
Backing up the environment.....	16
Upgrading the Talend projects in Talend Studio.....	16
<b>Appendices.....</b>	<b>17</b>
Supported Third-Party System/Database/Business Application Versions.....	17

# Copyleft

Adapted for 7.1.1. Supersedes previous releases.

Publication date: October 15, 2019

The content of this document is correct at the time of publication.

However, more recent updates may be available in the online version that can be found on [Talend Help Center](#).

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 Sqoop, 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.

# Talend Open Studio for Big Data: Prerequisites

## Preparing your installation

### Software packages

This page details the software packages you need to download to install your Talend product.

In this page:

- YYYYMMDD\_HHmm corresponds to the package timestamp
- A.B.C. corresponds to package version number (Major. Minor. Patch.)

The software modules must be all in the same versions/revisions. This means that both YYYYMMDD\_HHmm and A.B.C must match on both client side and server side.

### Manual installation software packages

File name	Description
Talend-Studio-YYYYMMDD_HHmm-VA.B.C.zip	Studio IDE (GUI) To download it, go to <a href="#">this page</a>

## Community and Support

There are several ways to get help and support for your Talend installation:

- [Official Talend Documentation](#). Here you can find everything to help you install and use your Talend product.
- [Talend Community](#). This is the place where you can ask questions to the community, and get answers.

## Hardware requirements

Before installing your Talend product, make sure the machines you are using meet the following hardware requirements recommended by Talend.

Memory and disk 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.

### Memory usage

Product	Client/Server	Memory requirements (minimum-recommended)	Note
Talend Studio	Client	3GB – 4GB	

**Note:** Depending on the number of executed processes running on a module, you may need to increase the available memory. If you have several products installed on the same host, Talend recommends to use an i7 CPU with 8 logical processors.

## Disk space requirements

Product	Client/Server	Required disk space for installation	Required disk space for use
Talend Studio	Client	3GB	3+ GB

## Software requirements

### Compatible Operating Systems

This page details the recommended and supported Operating Systems for Talend products.

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.

### Talend Studio

Support type	Operating System (64-bit)	
Recommended	Linux	Ubuntu 18.04 LTS
	Windows	Microsoft Windows 10
Supported	Linux	Ubuntu 16.04 LTS
		Red Hat Enterprise Linux Server/ CentOS 7.5
		Red Hat Enterprise Linux Server/ CentOS 7.4
		Red Hat Enterprise Linux Server/ CentOS 7.3
		Red Hat Enterprise Linux Server/ CentOS 7.2
		Red Hat Enterprise Linux Server/ CentOS 7.1
		Red Hat Enterprise Linux Server/ CentOS 6.9

Support type	Operating System (64-bit)	
		Red Hat Enterprise Linux Server/ CentOS 6.8
	Windows	Microsoft Windows Professional 7
		Microsoft Windows Server 2016 RTM
		Microsoft Windows Server 2012 RTM
	Windows Server on AWS	Microsoft Windows Server 2016 RTM
		Microsoft Windows Server 2012 RTM
	Mac	Apple macOS 10.14/Mojave
		Apple macOS 10.13/High Sierra
		Apple macOS 10.12/Sierra
Deprecated	Mac	Apple OS X 10.11/El Capitan

## Compatible Java Environments

The following tables provide information on the recommended Java Environment you should download and install to use your Talend product.

The Compiler Compliance Level corresponds to the Java version used for the Job code generation. This option can be changed in the Studio preferences. For more information, see the Talend Studio User Guide.

**Note:** All Talend products and associated third-party applications, such as the Hadoop cluster, should use the same Java version for compliance. Before you install or upgrade any associated third-party application, Talend recommends that you check which Java version they support.

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.

### Studio Java environments

Support type	JRE Version	Note
Recommended	OpenJDK 8	Recommended distribution: <a href="#">Zulu</a>
Recommended	Oracle 8	Studio JDK Compiler Compliance Level 1.8 (default)

## Setting up JAVA\_HOME

In order for your Talend product to use the Java environment installed on your machine, you must set the JAVA\_HOME environment variable.

### Procedure

1. Find the folder where Java is installed.

For example:

- C:\Program Files\Java\JREx.x.x
- C:\Program Files\Zulu

2. Open the **Start** menu and type `Environment variable` in the search bar to open the Environment variable properties.
3. Click **Environment Variables...**
4. Under **System Variables**, click **New...** to create a variable. Name the variable `JAVA_HOME`, enter the path to your Java environment, and click **OK**.
5. Under **System Variables**, select the **Path** variable, click **Edit...** and add the following variable at the end of the Path variable value: `;%JAVA_HOME%\bin`



# Installing your Talend Open Studio for Big Data manually

## Installing and configuring your Talend Studio

### Unzip the archive

#### Procedure

1. Download your product from [this page](#).
2. Unzip it using 7-zip.

### Editing the memory and JVM settings

To gain in performance at runtime and when launching Talend Studio, proceed as follows: you can edit the memory settings in the `.ini`.

#### Procedure

1. Edit the `TOS_BD-win-x86_64.ini` file.
2. Edit the memory attributes. For example:

```
-vmargs -Xms512m -Xmx1536m -XX:MaxMetaspaceSize=512m
```

**Tip:** For big projects, you may need to increase Xmx to 4096m.

For more details, see <http://www.oracle.com/technetwork/java/hotspotfaq-138619.html>.

### Launching your Talend Studio

#### Procedure

Double-click the `TOS_BD-win-x86_64.exe` executable to launch your Talend Studio.

### 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 database drivers, known as external modules, may be required by some of Talend components or by some connection wizards or by both. Due to license restrictions, Talend may not be able to ship some of these external modules within Talend Studio. You need to install them for your Studio to functional properly.

#### When to install external modules

Your Talend Studio will let you know when you need to install external modules and what external modules you need to install.

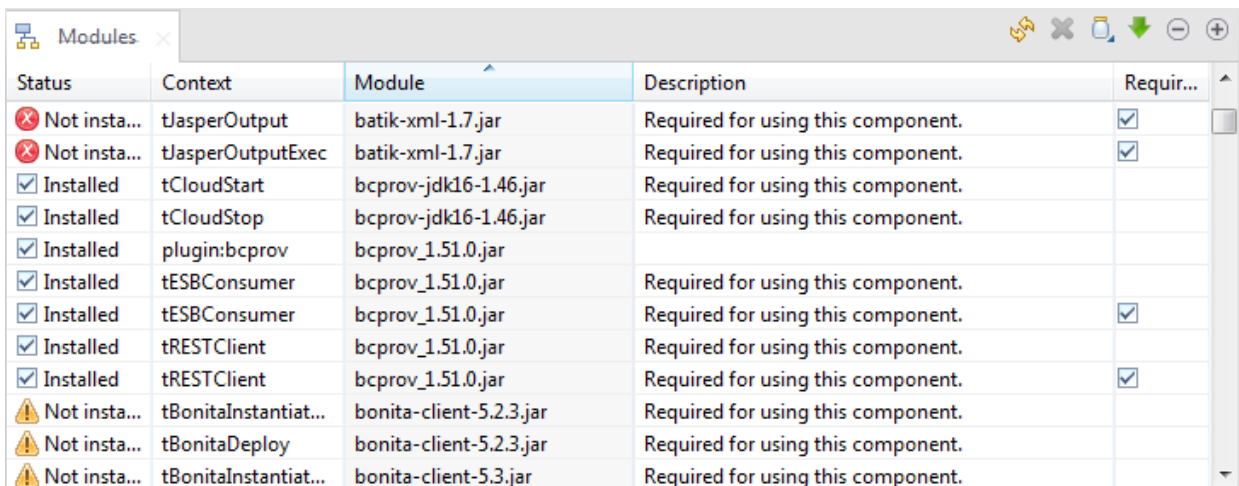
Your Talend Studio notify you about required external modules in several ways.

- The **Additional Talend packages** wizard opens when you launch your Talend Studio if any additional packages, including external modules, need to be installed for any features to function in the Studio.

**Tip:** The **Additional Talend packages** wizard also opens when you select **Help > Install Additional Packages** from the Studio menu.

- 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 for the Studio to work properly, including those Java libraries and drivers that you must install.

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

In this view:

### 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 or Metadata connection listed in this column.

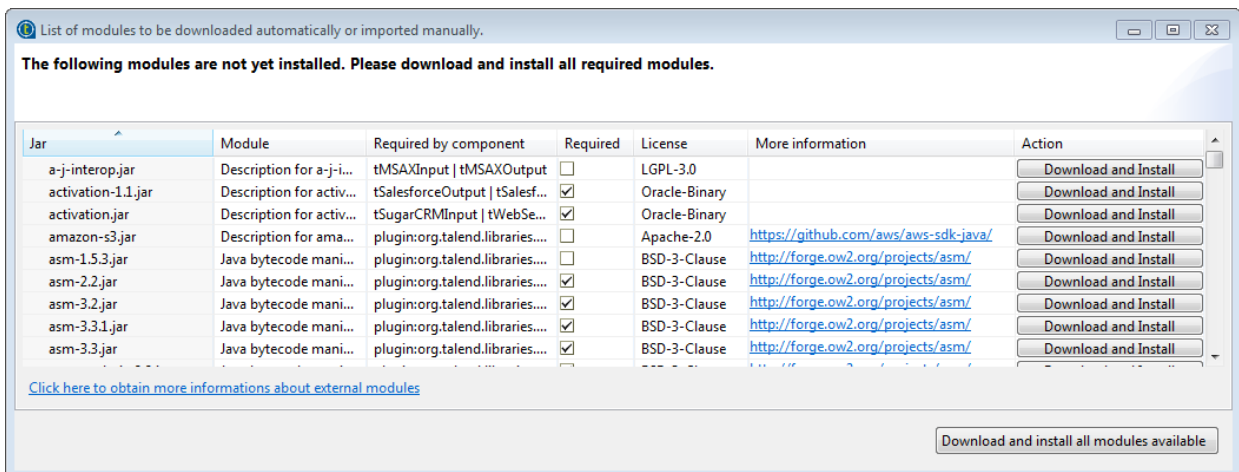
The  icon indicates that the module is absolutely required for the corresponding component or Metadata connection.

### Context

gives the name of the component or Metadata connection using the module. If this column is empty, the module is then required for the general use of your Talend Studio.

<b>Module</b>	gives the exact name of the module.
<b>Description</b>	explains why the module/library is required.
<b>Required</b>	the selected check box indicates that the module is required.
	refreshes this view to reflect the latest module installation status.  In case of collaborative work, once a required module is installed in one user's studio, the other users can simply refresh their <b>Modules</b> view to add this module to their own studio(s).
	allows you to install an already downloaded external module into your Studio. For details, see <a href="#">Installing external modules manually using the Modules view</a> on page 14
	opens the Jar download and installation wizard, which will list all the required external modules that are not integrated in the Studio.

- A Jar installation wizard appears 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.
  - click the **Check** button in a Metadata connection setup wizard in the Studio if one or more external modules required for the connection are missing in the Studio.
  - 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 or Metadata connections for which one or more required external modules are missing.
  - click the  button in the **Modules** view.



This wizard:

- lists the external modules to be installed and the licenses under which they are provided,
- provides the URLs of the valid websites where they are downloadable,
- lets you download and install automatically all the modules available on the Talend website,
- allows you to 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, set up a connection, 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.

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

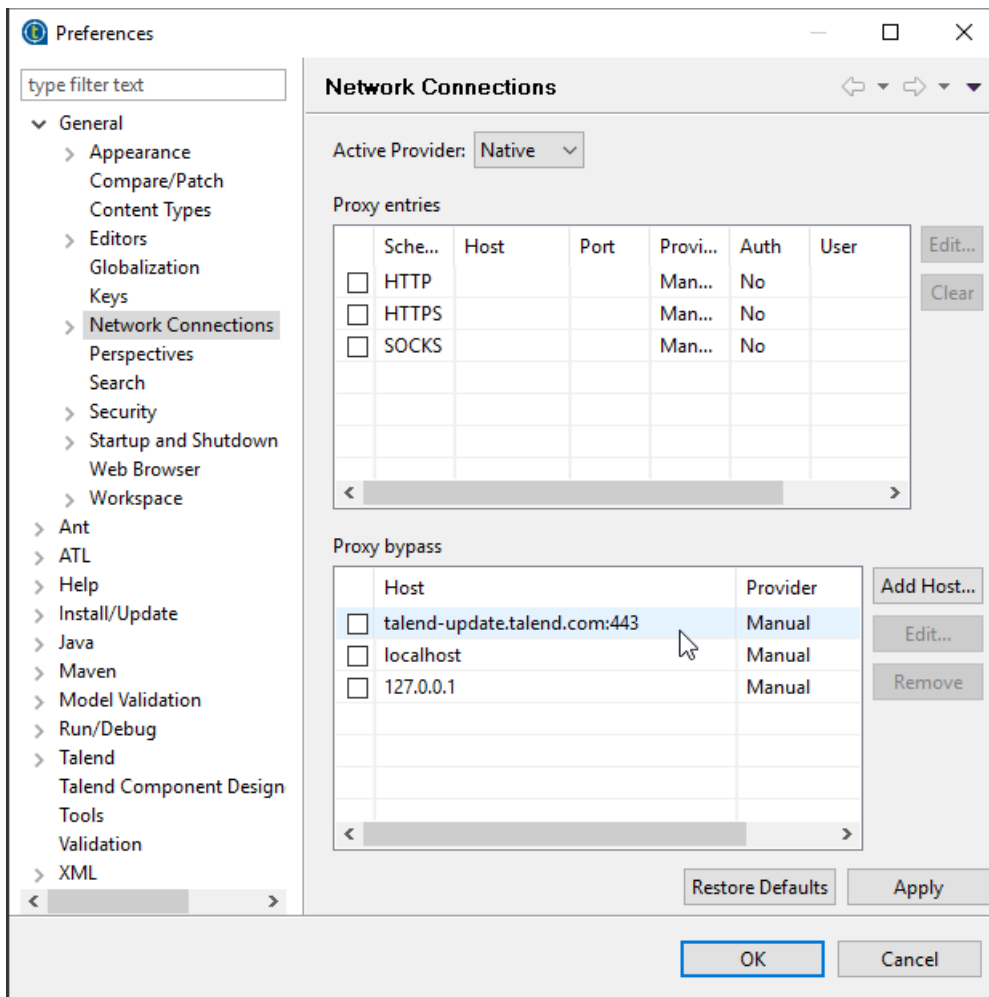
### Installing external modules from within the Studio

You can download and automatically install most external modules using the wizard provided by your Talend Studio.

#### Before you begin

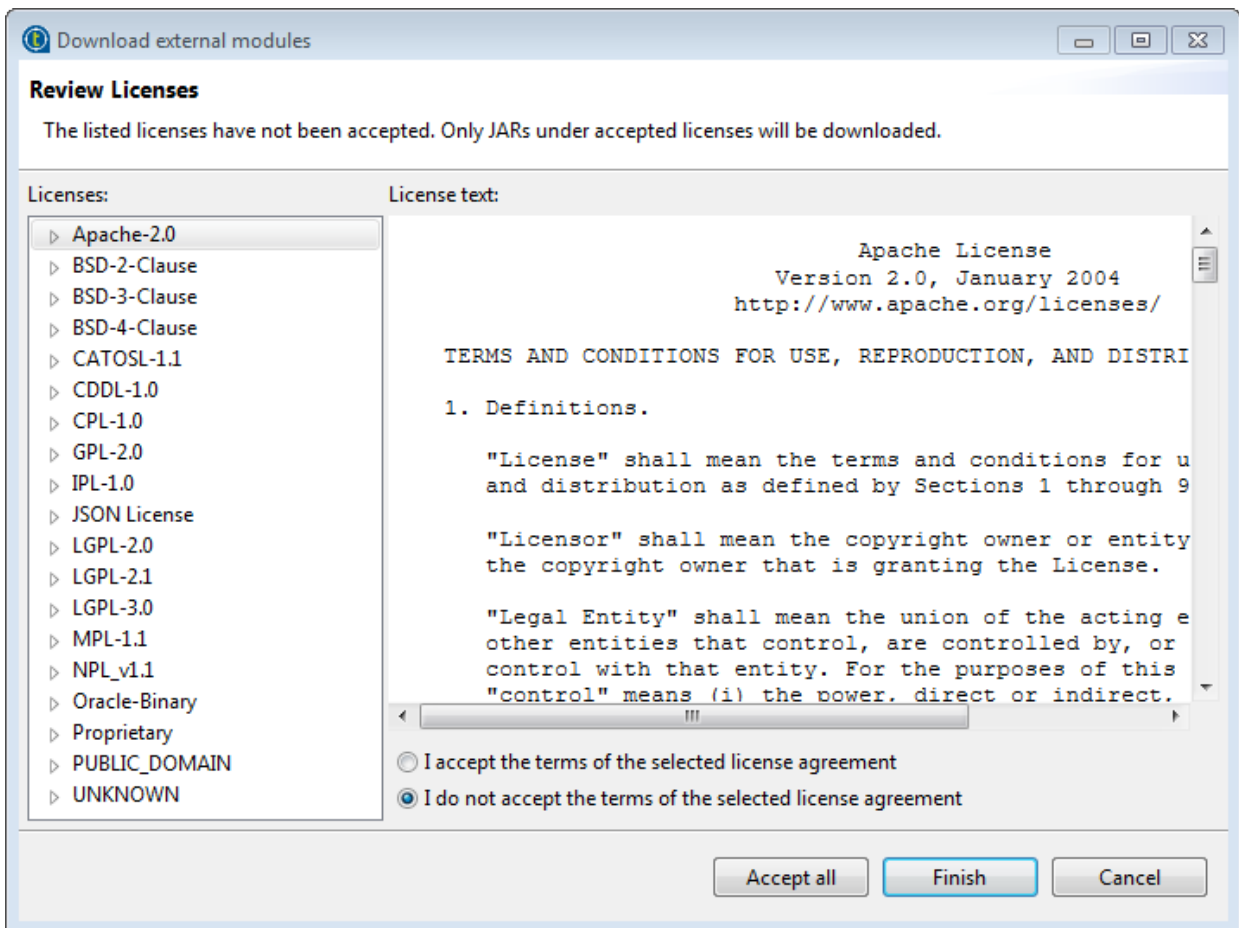
Make sure your Talend Studio has a secure Internet connection.

If you are working behind a network proxy, make sure you have correctly set up your proxy and add the web site `http://talend-update.talend.com` and the port 443 to your whitelist. 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**.



## Procedure

- Do the following to open the **Download external modules** dialog box:
  - In the **Additional Talend Packages** wizard, select the **Required third-party libraries** and/or **Optional third-party libraries** check boxes and click **Finish**.
  - 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.



## 2. Accept the license terms and start the download and installation process:

- 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 download and install all external modules provided under all the listed licenses, click the **Accept all** button.

## Results

When the installation process is completed, the chosen external module or modules are installed into your Talend Studio, and you can use Talend Studio features that depend on these modules.


## Installing external modules manually using the Modules view

If you have already downloaded external modules, you can install them manually into your Talend Studio.

## Before you begin

If you are going to install the JDBC driver for Oracle 9i into your Talend Studio, change the file name from `ojdbc14.jar` to `ojdbc14-9i.jar` first.

## Procedure

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

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 into your Talend Studio.

## Results

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

## Installing external modules manually for Talend Web applications

Some modules required for a Talend Web application to work 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.

## Procedure

- For the Talend MDM Server, place the downloaded JDBC drivers for the Oracle and MySQL databases in the following folder:

```
<TomcatPath>/webapps/talendmdm/WEB-INF/lib
```

- For Talend Administration Center, place the downloaded modules in the following folder:

```
<TomcatPath>/webapps/org.talend.administrator/WEB-INF/lib
```

# Upgrading your Talend products

## 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:

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

1. Saving the local projects, see [Saving the local projects](#) on page 16.

### Saving the local projects

#### Procedure

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

## Upgrading the Talend projects in Talend Studio

### Importing your local projects

#### Procedure

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

#### Results

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

For more information on how to export local projects to an archive file, see [Saving the local projects](#) on page 16.



# Appendices

## 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.

### 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	Note
Access	2003 2007	Windows	When working with Java 8, only the General collation mode is supported.
Amazon Aurora	Amazon Aurora MySQL edition v5 (MySQL 5.6/5.7)		
Amazon RDS for Microsoft SQL Server	N/A		
Amazon Redshift	Initial release of Amazon Redshift	N/A	
AS/400	V6R1 to V7R2 (Deprecated versions: V5R2 to V5R4/ V5R3 to V6R1)	N/A	
Bonita	6.5.2 7.2.4 (Deprecated versions: 5.2.3/5.3.1/5.6.1/5.10.1)	N/A	
Cassandra	3.0/3.1/3.2/3.3/3.4 (Deprecated versions: 1.1.2/1.2.2/2.0.0)	Windows + Linux	
CouchBase	5.x 6.0 (Deprecated versions: 2.0/4.x)	Windows	
CouchDB	1.0.2	Windows	
DB Generic	ODBC	Windows	
DynamoDB	No specified version	N/A	

Systems/Databases	Versions	OS	Note
Elasticsearch	2.3.x 5.6.x (Deprecated version: 1.7.x)	N/A	
EXASolution	6.0 and earlier	Windows	
Excel	N/A	N/A	
eXist-db	1.4.0	N/A	
FireBird	2.1	Windows + Linux	
FTP	N/A		
Greenplum	4.3.x 5.x (Deprecated version: 4.2.1.0)	Windows (client only) + Linux	
Hbase	N/A		
HDFS	N/A		
Hive	N/A		
HSQldb	1.8.0	N/A	
IBM DB2 and IBM DB2 Z/OS	10.5 11.1 (Deprecated version: 10.1)	Windows + Linux	
Impala	N/A		
Informix	11.50	Windows + Linux	
Ingres	10.2 11 (Deprecated version: 9.2)	Windows + Linux	
Interbase	(Deprecated versions: 7 and above)		
JavaDB	6	Windows + Linux	
JDBC	N/A		
JSON	N/A		

Systems/Databases	Versions	OS	Note
Kafka	0.8.2.0 0.9.0.1 0.10.0.1 1.1.0	Windows + Linux	The Kerberos kinit option and the Kerberos keytab option are both supported by Talend Studio. For information about the security options supported by the Kafka components, see <a href="#">Talend Help Center</a> .
LDAP	No version limitation	Windows + Linux	
MapRDB	N/A		
MarkLogic	V9	N/A	
MaxDB	7.6	N/A	
Microsoft Azure Blob Storage			
Microsoft Azure SQL Data Warehouse			
Microsoft AX	Dynamics AX 4.0 Dynamics AX 2012	N/A	
Microsoft CRM	2011 2015 2016	N/A	
Microsoft CRM Online	2011 2016	N/A	
Microsoft SQL Server	No version limitation	Windows + Linux	Microsoft SQL Server support is provided through the Microsoft SQL JDBC driver. For more information, see the <a href="#">Download Microsoft JDBC Driver for SQL Server</a> page.
MongoDB	3.4.x 3.6.x 4.0.x (Deprecated versions: 2.5.x/2.6.x/3.0.x/3.2.x)	Windows + Linux	

Systems/Databases	Versions	OS	Note
MySQL	MySQL 5.x MySQL 8.x MariaDB Google Cloud SQL (Deprecated version: MySQL 4)	Windows + Linux	
MOM	N/A		
Neo4j	1.x.x 2.x.x/2.2.x/2.3 3.2.x	Linux	
Netezza	7.0.x 7.1.x 7.2.x	Windows + Linux	
NetSuite	2018 (Deprecated versions: 2014/2016)	Windows + Linux	
OleDb	2000 2003 2005 2007 2010	N/A	
Oracle	Oracle 12c Release 1 Oracle 12c Release 2 Oracle 18c (Deprecated versions: Oracle 8i/Oracle 9i/Oracle 10g/Oracle 11g)	Windows + Linux	
ParAccel	3.1 3.5	N/A	
PostgreSQL	Prior to 9/9.x 9.x 10.x Google Cloud SQL	Windows + Linux	
PostgresPlus	Prior to 9/9.x 9.x	Windows + Linux	
Red Hat BRMS	6.1	Windows + Linux	
REST Service	N/A	Windows + Linux	
Sage X3	N/A		

Systems/Databases	Versions	OS	Note
Salesforce	V44 and earlier	Windows + Linux	
SAP	4.6		
SAP Business Suite (ERP)	Netweaver: From 7.3 to 7.5 ERP6.0, From EhP6 to EhP8	Windows	
SAP Business Warehouse (BW)	Netweaver: From 7.31 to 7.5	Windows	
SAP HANA	1.0 2.x	Windows	Supported through SAP JDBC driver
SAS	9.1 9.2	Windows + Linux	
SOAP Service	N/A		
SQLite	3.6.7	Windows + Linux	
Sqoop	N/A		
SugarCRM	5.2	Windows + Linux	
Sybase	12.5 12.7 15.2 15.5 15.7 16.0	Windows + Linux	
SybaseIQ	12.5 12.7 15.2 16.0	Windows + Linux	
Teradata	12 13 14 15 16	Windows + Linux	
VectorWise	2	Windows + Linux	
Vertica	9.0.x (Deprecated versions: 3/3.5/4/4.1/5. 0/5.1/6.0/6.1.x/7.0.x/7.1.x)	Windows + Linux	
VtigerCRM	Vtiger 5.0 Vtiger 5.1	N/A	

## Messaging brokers supported by Talend messaging components

Component	Supported messaging brokers / standards
tJMSInput tJMSOutput	JMS standard 1.1
tMicrosoftMQInput tMicrosoftMQOutput	MicrosoftMQ 3.0
tMomInput tMomOutput	JBoss Messaging 1.4.4 WebSphere MQ 8.0 ActiveMQ 5.13.2