



# Talend Open Studio for Big Data

## Release Notes

# 6.4.1

Adapted for v6.4.1. Supersedes previous releases.

Publication date June 29, 2017

## 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

System Requirements .....	1
Big Data: New Features .....	2
1. Enhancements of Hadoop connection wizard .....	2
2. Upgraded support for Hadoop .....	2
3. Enhanced Azure support .....	2
4. Compiler Compliance Level configuration for code generation .....	3
5. Tutorial videos upon Studio start-up .....	3
6. System routines .....	3
7. Component enhancements .....	3
8. Version upgrades .....	5
Big Data: Deprecated Items .....	7
Big Data: Bug Fixes / Change Log .....	8
1. Bug Fixes .....	8
Big Data: Known Issues .....	9
1. Out of memory error when using Java 8 .....	9
2. tSalesforceInput Limitations .....	9
3. Java PSU not supported .....	9
4. Collation limitation with Access .....	10
5. OAuth mode in the Salesforce wizard .....	10
6. Installation on Mac OS X Sierra .....	10
Documentation .....	11
1. Talend Help Center .....	11
2. Open issues .....	11

# System Requirements

Users should refer to the *Talend Installation and Upgrade Guide* on the **Talend Help Center** (<http://help.talend.com>) for more information on Installation and System Requirements.

---

# Big Data: New Features

## 1. Enhancements of Hadoop connection wizard

- Context-smart Hadoop connections:

When defining this connection, you can contextualize the connection parameters using values from different Hadoop environments such as from a test and a production environments, in order to adjust the connection and the Jobs using the connection to the proper environments at runtime by only one click.

## 2. Upgraded support for Hadoop

1. New versions of the following Hadoop distributions are supported:

- Dataproc 1.1 (via the tGoolgeDataprocManage component)
- Cloudera CDH V5.10
- EMR V5.4/5.5
- HDInsights 3.6
- Hortonworks Data Platform V2.6

## 3. Enhanced Azure support

- Talend Studio now supports retrieving/writing data from/into an Azure Storage table via the two new components tAzureStorageInputTable and tAzureStorageOutputTable.
- A new Azure Storage Blob component tAzureStorageContainerList, which allows you to list all containers in a given Azure Storage account, has been added into Talend Studio.
- Talend Studio now supports managing Azure Storage queues and messages via the seven new components tAzureStorageQueueCreate, tAzureStorageQueueDelete, tAzureStorageQueueInput, tAzureStorageQueueInputLoop, tAzureStorageQueueList, tAzureStorageQueueOutput and tAzureStorageQueuePurge.
- All existing and new Azure Storage components now support using a shared access signature to access the storage resources without the need for the account key.
- A new Azure Storage metadata wizard has been created, which allows you to centralize the Azure Storage connection and retrieve the schema of the interested containers, queues and tables.
- Talend Studio now allows you to manage data in Azure SQL Data Warehouse via the seven new components tSQLDWHClose, tSQLDWHCommit, tSQLDWHConnection, tSQLDWHInput, tSQLDWHOutput, tSQLDWHRollback and tSQLDWHRow.

---

## 4. Compiler Compliance Level configuration for code generation

The Studio now allows you to quickly switch the JDK compiler compliance level between Java 7 and Java 8 through project settings.

## 5. Tutorial videos upon Studio start-up

The Studio now comes with a series of embedded tutorial videos to help users get started. For some operating systems, you may need to install an MP4 decoder/player to play the videos.

## 6. System routines

The following system routines have been added into Talend Studio:

1. **Relational:** *isNull*
2. **StringHandling:** *SUBSTR, LTRIM, RTRIM, LPAD, RPAD, INSTR, and TO\_CHAR*
3. **TalendDate:** *TO\_DATE, ADD\_TO\_DATE, and TO\_CHAR*
4. **TalendStringUtil:** *DECODE*

## 7. Component enhancements

### 7.1. Smart tMap

1. You can now quickly search an input or output schema column or multiple columns among hundreds of them in one go.
2. You can now set an expression on multiple output columns in a batch manner.

### 7.2. Enhanced NetSuite support

1. The NetSuite components have been switched to the new framework to improve user experience.
2. A new component, *tNetsuiteConnection*, has been created to allow NetSuite connection reuse.
3. The record type selector of *tNetsuiteInput* and *tNetsuiteOutput* has been improved to provide access to NetSuite's record type views.
4. The NetSuite components are now able to retrieve customizations from NetSuite - custom record types and custom fields.

5. The tNetsuiteOutput component now supports the native Upsert operation provided by NetSuite.
6. The tNetsuiteOutput component now supports batch processing of records.

## 7.3. Amazon SQS support

1. Two new SQS components, tSQSMessageDelete which allows you to delete a message from a queue, and tSQSMessageChangeVisibility which allows you to change the visibility timeout of a message, have been added into Talend Studio.
2. tSQSInput now allows you to specify the maximum number of messages to return.

## 7.4. Marketo support

1. The existing Marketo components, tMarketoInput, tMarketoListOperation and tMarketoOutput, have been switched into the new framework and the user interface has been slightly changed.
2. Two new components, tMarketoConnection which allows you to open a connection to Marketo, and tMarketoBulkExec which allows you to import leads or custom objects into Marketo from a local file in the REST API mode, have been created in the new framework.
3. The tMarketoOutput component now allows you to delete leads.
4. The tMarketoInput and tMarketoOutput components now support retrieving and writing custom objects data from/into Marketo.
5. The tMarketoListOperation and tMarketoOutput components support collecting the rows on error using a Row > Reject connection when not in batch mode.
6. A new Marketo metadata wizard has been created, which allows you to centralize the Marketo connection and retrieve the schema of the interested custom objects using REST API.

## 7.5. Microsoft CRM support

1. tMicrosoftCrmInput and tMicrosoftCrmOutput now support ONLINE v2016.
2. tMicrosoftCrmInput and tMicrosoftCrmOutput now allow you to specify the maximum number of reconnect attempts to the Microsoft CRM database after the token is expired for ONLINE v2011 and v2016(odata).

## 7.6. Salesforce support

1. tSalesforceConnection has been improved by adding the function of saving or reusing the connection session file.
2. tSalesforceInput now supports generating the schema columns based on the specified query and generating the query based on the specified module and schema columns. For the limitations you should be aware of when using this feature, please refer to [tSalesforceInput Limitations](#).
3. The tSalesforceInput component now supports PK (Primary Key) chunking when extracting a large number of records or when the query consistently times out.

4. tSalesforceOutput has been improved to allow Salesforce.com to return Salesforce ID for both inserted and updated records.

## 7.7. Snowflake support

1. Four new Snowflake components, tSnowflakeConnection, tSnowflakeInput, tSnowflakeOutput and tSnowflakeClose, which allow you to read/write data from/into a Snowflake table, have been created in the new framework.
2. A new Snowflake metadata wizard has been created, which allows you to centralize the Snowflake connection and retrieve the schema of the interested tables.

## 7.8. JDBC ELT and SCD support

1. The ELT JDBC components, tELTJDBCInput, tELTJDBCMap, and tELTJDBCOutput now support specifying the metadata mapping file for the database to be used. The supported databases are Exasol, Netezza and Redshift. Other databases may work, but have not been tested.
2. The tELTJDBCOutput component now supports performing an action on the specified table.
3. A new component tJDBCSCDELTA, which allows you to track data changes in a source database table using Type 1 and/or Type 2 Slowly Changing Dimensions (SCD) and write the dimension data into a specified SCD table, has been created.

## 7.9. Others

1. tFileOutputDelimited now supports throwing an exception if the specified output file already exists.
2. The DB Version drop-down list is moved from tELTSybaseOutput to tELTSybaseMap.
3. The tJMSInput component now supports looking up a destination with the JNDI (Java Naming and Directory Interface) name.

# 8. Version upgrades

## 8.1. EXASOL

EXASOL v6.0 is now supported.

## 8.2. Excel

Excel 2016 is now supported.



## **8.3. Microsoft CRM**

Microsoft CRM 2016 and Dynamics 365 are now supported.

## **8.4. Salesforce**

Salesforce has been upgraded to v39.

## **8.5. Sybase**

Sybase v16.0 is now supported.

## **8.6. NetSuite**

The NetSuite components now support NetSuite 2016.

# Big Data: Deprecated Items

The following feature is deprecated in the current release. Deprecated features are no longer recommended for use and may cease to exist in future versions of the product.

- The tAlfrescoOutput component
- The HL7 metadata wizard and the tHL7Input and tHL7Output components
- The Component Designer perspective
- The Axis Webservice (WAR) and Axis Webservice (ZIP) build job types
- The Enable commandline server option used to add a remote CommanLine server

The support for the following Hadoop distributions

- From Hortonwork Data Platform V1.2.0 to Hortonwork Data Platform V2.3.2
- Apache 1.0.0
- From Cloudera CDH 4 to Cloudera 5.4
- From MapR 2.0.0 to MapR 4.1.0
- From Amazon EMR 1.0.3 to Amazon EMR 4.0.0
- From Pivotal HD 1.0.1 to Pivotal HD 2.0
- From Microsoft HD Insight 3.1 to Microsoft HD Insight 3.2

# Big Data: Bug Fixes / Change Log

## 1. Bug Fixes

In addition to the above new features a number of minor improvements within the entire product and significant bug fixes have been made.

See the corresponding Change Log on our bug tracking system for more details on the individual issues:

- <https://jira.talendforge.org/issues/?filter=25944>.

# Big Data: Known Issues

We encourage you to consult the JIRA bug tracking tool for a full list of open issues:

<https://jira.talendforge.org/issues/?filter=25945>.

Note that this list shows issues from both Talend's Community and Subscription products.

## 1. Out of memory error when using Java 8

When using Java 8, the out of memory error may occur when a Job builds, especially when the Job has a lot of child Jobs. A workaround is to increase the JVM memory for the Studio.

## 2. tSalesforceInput Limitations

The following SOQL queries are not supported for the guess schema and guess query features of tSalesforceInput:

- the SOQL query using the aggregate function
- the SOQL query using the TYPEOF expression
- the SOQL query using alias
- the prefix in the SOQL query is the same as the module name after FROM

For example, `SELECT Account.Name, (SELECT Contact.LastName FROM Account.Contacts) FROM Account` is not supported, while `SELECT Name, (SELECT LastName FROM Contacts) FROM Account` that performs the same function is supported.

- the SOQL query with multiple modules after FROM

For example, `SELECT Contact.Firstname, Account.Name FROM Contact, Contact.Account` is not supported. However, the child-to-parent and parent-to-child relationship queries are supported, so this query can be improved to `SELECT Firstname, Account.Name FROM Contact`.

- the SOQL query using toLabel()
- the field with the underscore character '\_' in its name

Besides, the underscore character '\_' in any custom field name needs to be manually replaced with the dot character '.' in the generated SOQL query.

## 3. Java PSU not supported

Talend is only supported on Java CPS Versions, not Java PSU.

Some users reported problems using Talend Studio with Java PSU versions (such as jdk8\_u92), but when reverting to a Java CPS Version (such as jdk8\_u91) the problem was resolved.

See also: <http://www.oracle.com/technetwork/java/javase/cpu-psu-explained-2331472.html>.

## 4. Collation limitation with Access

Since 6.0, the Access components in *Talend Studio* support only the General collation mode.

## 5. OAuth mode in the Salesforce wizard

To establish a Salesforce connection using OAuth2 authentication, you need first to execute a Talend Job to obtain your token file and then initialize it in the wizard.

## 6. Installation on Mac OS X Sierra

By default, Mac OS X Sierra sets all downloaded files to "quarantine". This prevents you running the Studio.

To launch Talend Studio on Mac OS X Sierra, do the following:

1. Unzip the Studio as usual.
2. Open a **Terminal** window on the folder where you unzipped your Studio.
3. Execute either of the following commands.

```
xattr -d com.apple.quarantine *
```

or

```
xattr -c *
```

You can now launch the Studio as usual.

# Documentation

## 1. Talend Help Center

Find out more about how to get the most out of your Talend products on the Talend Help Center: <http://help.talend.com>.

Talend Help Center has undergone a major refurbishment for this release.

## 2. Open issues

We encourage you to consult the JIRA bug tracking tool for a full list of open issues:

<https://jira.talendforge.org/issues/?filter=18375>