



Talend Open Studio for Big Data

Release Notes

6.2.2

Adapted for v6.2.2. Supersedes previous releases.

Publication date January 23, 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. Upgraded support for Hadoop, NoSQL and Neo4j	2
2. MapR ticket security	2
3. MongoDB GridFS support	2
4. Enhanced Kafka support in the Data Integration Jobs	3
5. Support for AWS DynamoDB	3
6. Rolling update to incorporate the most recent Hadoop distributions	3
7. Enhanced Salesforce support	3
8. No need to restart Studio upon installing Jars	3
9. Enhanced database support	3
10. AWS support	4
11. Others	5
Big Data: Bug Fixes / Change Log	7
1. Bug Fixes	7
Big Data: Known Issues	8
1. Java PSU not supported	8
2. Installation on Mac OS X Sierra	8
Documentation	9
1. Talend Help Center	9
2. Videos	9
3. Revised documents	9
4. Removed/Deprecated documents	9
5. Open issues	9

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. Upgraded support for Hadoop, NoSQL and Neo4j

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

- Cloudera CDH V5.6.0/5.7.0
- Hortonworks Data Platform V2.4.0
- MapR 5.1
- Microsoft HD Insight 3.4 (YARN)
- EMR 4.5.0/4.6.0

2. New versions of the following NoSQL or graph databases are supported:

NoSQL databases

- Cassandra V3.0/3.1/3.2/3.3/3.4
- MongoDB V3.2.X

Graph database

- Neo4j V2.3

2. MapR ticket security

The support of MapR ticket security has been added to MapR configurations, whether the MapR cluster to be used is Kerberos-enabled or not.

3. MongoDB GridFS support

The Studio now includes MongoDB GridFS support via the following new components:

- tMongoDBGridFSGet
- tMongoDBGridFSList
- tMongoDBGridFSPut
- tMongoDBGridFSProperties
- tMongoDBGridFSDelete

4. Enhanced Kafka support in the Data Integration Jobs

1. The support of Kafka 0.9.0.1 has been added.
2. The SSL/TLS security and the Kerberos security are both supported by the Kafka components when using Kafka 0.9.0.1

5. Support for AWS DynamoDB

Talend Studio now allows you to manage data with DynamoDB using the following components:

- tDynamoDBOutput
- tDynamoDBInput

6. Rolling update to incorporate the most recent Hadoop distributions

If support for the Hadoop distribution users want to use is not yet available in Talend Studio, it may be available via an update.

7. Enhanced Salesforce support

- The Salesforce connection wizard and components have been redesigned based on the new framework.
- tSalesforceOutput now supports upserting a lookup field with NULL input values by adding the new Lookup field name column in the Relationship mapping for upsert table in the Advanced settings view.

8. No need to restart Studio upon installing Jars

Now users are no longer required to restart the Studio after downloading and installing .jar files such as third-party libraries and database drivers.

9. Enhanced database support

- tVerticaBulkExec now allows you to specify the character to be escaped when loading data into Vertica.

- tEXABulkExec now allows you to import data from different sources: a local file, a remote file, an EXASol database, an Oracle database, or a JDBC-compliant database.
- tSybaseIQBulkExec and tSybaseIQOutputBulkExec now can store zero-length varchars as NULLs instead of blanks in the database during the load operation.

10. AWS support

10.1. AWS Redshift components, driver, and metadata update

The Studio now allows you to specify additional JDBC parameters when creating a Redshift database connection.

10.2. AWS Cluster Resizing (EMR/Redshift)

- The new component tAmazonEMRResize has been created to allow you to add or resize a task instance group in a cluster on Amazon EMR.
- tAmazonRedshiftManage now supports resizing an existing Amazon Redshift cluster.

10.3. S3 client-side encryption support

- tRedshiftOutputBulk and tRedshiftOutputBulkExec now support generating and uploading the data file to Amazon S3 using client-side encryption.
- With the client-side encryption feature enabled, the data can now be encrypted before sending to Amazon S3 via tS3Connection, tS3Get and tS3Put.

10.4. Other AWS support enhancements

- tAmazonEMRManage now can start either a transient cluster or a long-running cluster, and the job flow step(s) can be invoked on the cluster after its launch.

This component also allows users to specify the identifier of the Amazon VPC (Virtual Private Cloud) subnet, the bootstrap actions, and the classification and property information supplied to the configuration object of the EMR cluster to be created.

- A new component tS3Copy has been created to allow you to copy an Amazon S3 object from one bucket to another.
- A new component tAmazonEMRListInstances has been created to allow you to list instance groups in a cluster on Amazon EMR.
- The S3 components now support obtaining AWS security credentials from Amazon EC2 instance metadata.

- tRedshiftInput now allows you to store the logging information in a specified log file.
- tRedshiftBulkExec now allows you to load data into an Amazon Redshift table from an Amazon DynamoDB table or from data files located in an Amazon S3 bucket, an Amazon EMR cluster, or a remote host that is accessed using an SSH connection.
- The Amazon components on the Palette have been reorganized and the Amazon Aurora, Amazon MySQL, Amazon Oracle, Amazon Redshift components are now available in both the Cloud family and the Databases family.

10.5. New components

10.5.1. Splunk support

The Studio now allows you to send the event data to Splunk using a new component tSplunkEventCollector.

10.5.2. JIRA support

The Studio now allows you to read/write the issue and project information from/into JIRA using new components tJIRAInput and tJIRAOutput.

11. Others

11.1. Upgraded XML validation support

tXSDValidator now allows you to enable feature(s) on the underlying parser.

11.2. Upgraded HTTP request support

tHttpRequest now allows you to specify the Connect timeout and Read timeout for a request.

11.3. Updates of Marketo components

The default API mode of the Marketo components has been set to the REST API.

11.4. Upgrade of messaging components

ActiveMQ 5.13 and WebSphere MQ 8.0 are now supported by the JMS and MOM components.

11.5. Upgraded email support

tPOP now supports retrieving a specified number of the most recent email messages and returning the messages in chronological order by adding a "Newer email first" check box.

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/jira/secure/ReleaseNote.jspa?projectId=11430&version=20903>.

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=24974>.

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

1. 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>.

2. 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>.

2. Videos

From within the Studio, you can directly access the videos hosted in Talend Help Center relating to new features.

- For components, related videos are available from the F1 help view of this component.
- For other features, related videos are available in the help view of the wizard for the feature.

Note that not every feature has related videos in Talend Help Center and for this reason, not all the features have this type of access.

3. Revised documents

In addition to updates to the content across the documentation set, the following specific documentation changes have been made:

- The Getting Started Guides have undergone a major refurbishment and now feature new use-cases built around a fictitious movie rental firm.
- The Installation Guides are now split by operating system, and they also underwent major structural changes, including the fact that they now document the installation using the installer in greater detail.

4. Removed/Deprecated documents

The following documents have been removed:

- Talend ESB Hands-on Guide.

5. 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>