



Talend Open Studio for Big Data

Release Notes

6.0.0

Adapted for v6.0.0. Supersedes previous releases.

Publication date July 2, 2015

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 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.

Table of Contents

System Requirements	1
Big Data: New Features	2
1. Discovery and reusability in Hadoop cluster metadata	2
2. Upgraded support for NoSQL and Hadoop distributions	2
3. Tez framework	2
4. New messaging components	3
5. Hadoop components enhancements	3
6. Enhanced look and feel and user experience	3
7. Java 8 support	4
8. Memory Test Execution	4
9. Components Updates	4
Big Data: Deprecated Items	7
1. Deprecated Items	7
Big Data: Bug Fixes / Change Log	8
1. Bug Fixes	8
Big Data: Known Issues	9
Big Data: Hints and Notes	10
1. JDK Required to Build Jobs/Routes	10
Documentation	11
1. Talend Help Center	11
2. Videos	11
3. New documents	11
4. Revised documents	11
5. Known issues	12
6. Open issues	12

System Requirements

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

Big Data: New Features

1. Discovery and reusability in Hadoop cluster metadata

1. An import wizard for Hadoop configuration has been created to import the configurations of various Hadoop services directly from different sources.
 - This wizard supports importing configurations directly from the Ambari service or the Cloudera manager service.
 - When users import configurations from local files, this wizard supports all of the distributions already officially supported by Talend.
2. Users can create environmental context out of the metadata of a connection to a Hadoop service such as HDFS or Hive. Then this context can be used to define the parameters of the Hadoop connection to be created.

2. Upgraded support for NoSQL and Hadoop distributions

1. New versions of the following Hadoop distributions are supported:
 - Cloudera 5.4 (YARN)
 - MapR 4.1.0 (YARN)
 - Hortonworks Data Platform V2.2
 - EMR Apache 2.4.0
2. Cassandra:

Cassandra CQL3 and the Datastax API are now available to the Cassandra components.
3. MongoDB:
 - MongoDB 3.0.X. is supported
 - The MongoDB components now support more operations such as "Bulk write" or "Write concern" and have been enabled to work with different authentication mechanisms including Kerberos and SCRAM-SHA1.

3. Tez framework

The support of the Tez framework has been added to the Hive and the Pig components.

4. New messaging components

In the Data Integration Jobs, the following Kafka components have been created to help users visually create a topic and consume or publish messages:

- tKafkaCreateTopic
- tKafkaInput
- tKafkaOutput

5. Hadoop components enhancements

1. With HiveServer 2 in the Standalone mode, the Hive components now support the Kerberos authentication.
2. The tHDFSOutputRaw component has been created to write data into more formats than the Text and the Sequence formats.
3. From the Hive components, the SSL connection is now available to HiveServer2 in the **Standalone** mode of the following distributions:
 - Hortonworks Data Platform 2.0 +
 - Cloudera CDH4 +
 - Pivotal HD 2.0 +
4. Support for Kerberos has been added to the HBase components and the HBase connection in the Hive and Pig components.

6. Enhanced look and feel and user experience

The Studio has undergone a number of changes to enhance the look and feel and user experience.

1. The Studio user interface and component icons have been redesigned for enhanced look and feel.
2. The Studio login process has been streamlined:
 - For the first startup, fewer steps are required.
 - After the first start up, users can directly open the same project without seeing the login dialog.
3. Users can now drag and drop an output component from an input one on the design workspace.
4. Enhanced component search:
 - Users can now search for a component by providing a descriptive phrase as the search keywords, without having to know the component name. The maximum number of entries in the search result is configurable.
 - The Palette now shows recently used components to ease component search in Job design. The maximum number of components shown on the **Recently Used** list is configurable.

7. Java 8 support

The Studio now fully supports Java 8 while remaining compatible with Java 7.

8. Memory Test Execution

Users can now monitor the JVM memory consumption and CPU usage via curve graphs during Job execution.

9. Components Updates

9.1. Databases

9.1.1. Batch mode

The following components now support batch mode when using an existing connection:

tAS400Output, tAmazonAuroraOutput, tAmazonMysqlOutput, tAmazonOracleOutput, tDB2Output, tFirebirdOutput, tGreenplumOutput, tInformixOutput, tIngresOutput, tInterbaseOutput, tJDBCOutput, tMSSqlOutput, tMysqlOutput, tNetezzaOutput, tOracleOutput, tParAccelOutput, tPostgresPlusOutput, tPostgresqlOutput, tRedshiftOutput, tSQLiteOutput, tSybaseOutput, tTeradataOutput, tVectorWiseOutput, tVerticaOutput.

9.1.2. Teradata SCD

Teradata SCD (Slowly Changing Dimensions) is now supported through the new component tTeradataSCD.

9.1.3. Vertica

VerticaELT and SQL Templates

1. Vertica ELT is now supported through the following new components:

- tELTVerticaInput
- tELTVerticaMap
- tELTVerticaOutput

2. SQL Templates for Vertica are now supported.

VerticaBulk

tVerticaBulkExec and tVerticaOutputBulkExec now support bulk-loading data to a Vertica database based on the defined schema columns.

9.1.4. PostgreSQL

1. PostgreSQL v9.x.x is now supported.

9.2. Cloud components

9.2.1. Salesforce

1. tSalesforceWaveBulkExec and tSalesforceWaveOutputBulkExec, which let you load data to Salesforce Analytics Cloud, have been added to the Studio.
2. tSalesforceWaveBulkExec and tSalesforceWaveOutputBulkExec support reusing an existing connection and retrieving the upload status.

9.2.2. Netsuite

1. tNetsuiteInput and tNetsuiteOutput, which let you read/write data from/to NetSuite, have been added to the Studio.
2. tNetsuiteOutput supports retrieving data rows with errors via a Reject row connection.

9.2.3. Marketo

1. tMarketoInput, which supports three lead selector types when retrieving lead records from Marketo.
2. tMarketoOutput, which supports providing the status (CREATED, UPDATED, FAILED) and the optional error text.

9.2.4. Others

1. Four tRedshift* load/unload components, which let you load/unload data from/to Amazon S3.
2. Seven tGoogleDrive* components, which let you create, upload, download, copy, delete, move and list files and folders on Google Drive.

3. Six tAmazonAurora* components, which let you read/write data from/to Amazon Aurora databases.

9.3. Other Messaging components

1. Static transport discovery mechanism is now supported for the ActiveMQ server in tMomXXX components.
2. tMomInput and tMomRollback now support backouting the messages to backout queue when the backout count reaches the configured threshold.

Big Data: Deprecated Items

1. 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.

- Support of Cassandra 1.1.2 and 1.2.2.
- MongoDB 2.1, 2.2 and 2.4
- Hive Server 1

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/secure/ReleaseNote.jspa?projectId=11430&version=16505>.

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=20549>

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

Big Data: Hints and Notes

1. JDK Required to Build Jobs/Routes

To build Jobs or Routes, you need to install a full JDK rather than a JRE. If the system fails to initialize automatically with a JDK, you can define which JDK to use by selecting Window > Preferences > Java > Installed JREs.

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 now 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. New documents

The following new documents have been added in this release:

- Talend Open Studio for Data Integration Getting Started Guide, Talend Open Studio for Data Quality Getting Started Guide, Talend Open Studio for ESB Getting Started Guide, and Talend Open Studio for MDM Getting Started Guide.

These Getting Started Guides walk users through their first steps with the product, including launching the Studio and setting up their first projects, and provide some basic examples.

- Talend Open Studio for Big Data Getting Started Guide now has completely new content, and its existing content has been moved to the Talend Open Studio for Big Data User Guide.

4. Revised documents

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

- The contents of the previous Talend Open Studio for Big Data Getting Started Guide are now included in the Talend Open Studio for Big Data User Guide.
- The Talend ESB Getting Started Guide that was available in previous versions has been renamed Talend ESB Hands-on Guide.

5. Known issues

The documentation contains the following known issues:

- The Talend Studio User Guide and the Installation and Upgrade Guide do not mention that, in order to build Jobs or Routes, users need to install a full JDK rather than a JRE.

6. 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>