



TIBCO® Data Science Team Studio Release Notes

*Software Release 6.5.0
August 2019*

Important Information

SOME TIBCO SOFTWARE EMBEDS OR BUNDLES OTHER TIBCO SOFTWARE. USE OF SUCH EMBEDDED OR BUNDLED TIBCO SOFTWARE IS SOLELY TO ENABLE THE FUNCTIONALITY (OR PROVIDE LIMITED ADD-ON FUNCTIONALITY) OF THE LICENSED TIBCO SOFTWARE. THE EMBEDDED OR BUNDLED SOFTWARE IS NOT LICENSED TO BE USED OR ACCESSED BY ANY OTHER TIBCO SOFTWARE OR FOR ANY OTHER PURPOSE.

USE OF TIBCO SOFTWARE AND THIS DOCUMENT IS SUBJECT TO THE TERMS AND CONDITIONS OF A LICENSE AGREEMENT FOUND IN EITHER A SEPARATELY EXECUTED SOFTWARE LICENSE AGREEMENT, OR, IF THERE IS NO SUCH SEPARATE AGREEMENT, THE CLICKWRAP END USER LICENSE AGREEMENT WHICH IS DISPLAYED DURING DOWNLOAD OR INSTALLATION OF THE SOFTWARE (AND WHICH IS DUPLICATED IN THE LICENSE FILE) OR IF THERE IS NO SUCH SOFTWARE LICENSE AGREEMENT OR CLICKWRAP END USER LICENSE AGREEMENT, THE LICENSE(S) LOCATED IN THE "LICENSE" FILE(S) OF THE SOFTWARE. USE OF THIS DOCUMENT IS SUBJECT TO THOSE TERMS AND CONDITIONS, AND YOUR USE HEREOF SHALL CONSTITUTE ACCEPTANCE OF AND AN AGREEMENT TO BE BOUND BY THE SAME.

ANY SOFTWARE ITEM IDENTIFIED AS THIRD PARTY LIBRARY IS AVAILABLE UNDER SEPARATE SOFTWARE LICENSE TERMS AND IS NOT PART OF A TIBCO PRODUCT. AS SUCH, THESE SOFTWARE ITEMS ARE NOT COVERED BY THE TERMS OF YOUR AGREEMENT WITH TIBCO, INCLUDING ANY TERMS CONCERNING SUPPORT, MAINTENANCE, WARRANTIES, AND INDEMNITIES. DOWNLOAD AND USE OF THESE ITEMS IS SOLELY AT YOUR OWN DISCRETION AND SUBJECT TO THE LICENSE TERMS APPLICABLE TO THEM. BY PROCEEDING TO DOWNLOAD, INSTALL OR USE ANY OF THESE ITEMS, YOU ACKNOWLEDGE THE FOREGOING DISTINCTIONS BETWEEN THESE ITEMS AND TIBCO PRODUCTS.

This document is subject to U.S. and international copyright laws and treaties. No part of this document may be reproduced in any form without the written authorization of TIBCO Software Inc.

TIBCO, the TIBCO logo, the TIBCO O logo, TIBCO Data Science Team Studio, TIBCO Spotfire, Alpine, and Chorus are either registered trademarks or trademarks of TIBCO Software Inc. in the United States and/or other countries.

Java and all Java based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

All other product and company names and marks mentioned in this document are the property of their respective owners and are mentioned for identification purposes only.

This software may be available on multiple operating systems. However, not all operating system platforms for a specific software version are released at the same time. Please see the readme.txt file for the availability of this software version on a specific operating system platform.

THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT.

THIS DOCUMENT COULD INCLUDE TECHNICAL INACCURACIES OR TYPOGRAPHICAL ERRORS. CHANGES ARE PERIODICALLY ADDED TO THE INFORMATION HEREIN; THESE CHANGES WILL BE INCORPORATED IN NEW EDITIONS OF THIS DOCUMENT. TIBCO SOFTWARE INC. MAY MAKE IMPROVEMENTS AND/OR CHANGES IN THE PRODUCT(S) AND/OR THE PROGRAM(S) DESCRIBED IN THIS DOCUMENT AT ANY TIME.

THE CONTENTS OF THIS DOCUMENT MAY BE MODIFIED AND/OR QUALIFIED, DIRECTLY OR INDIRECTLY, BY OTHER DOCUMENTATION WHICH ACCOMPANIES THIS SOFTWARE, INCLUDING BUT NOT LIMITED TO ANY RELEASE NOTES AND "READ ME" FILES.

This and other products of TIBCO Software Inc. may be covered by registered patents. Please refer to TIBCO's Virtual Patent Marking document (<https://www.tibco.com/patents>) for details.

Copyright © 2019. TIBCO Software Inc. All rights reserved.

Contents

TIBCO Documentation and Support Services	5
TIBCO® Data Science Team Studio Release Notes	6
New Features	6
Deprecated and Removed Features	7
Migration and Compatibility	7
Changes in Platform Support	7
Closed Issues	7
Known Issues	8

TIBCO Documentation and Support Services

How to Access TIBCO Documentation

Documentation for TIBCO products is available on the TIBCO Product Documentation website, mainly in HTML and PDF formats.

The TIBCO Product Documentation website is updated frequently and is more current than any other documentation included with the product. To access the latest documentation, visit <https://docs.tibco.com>.

Product-Specific Documentation

The following documents for this product can be found in the TIBCO Documentation Library.

- *TIBCO® Data Science Team Studio Release Notes*
- *TIBCO® Data Science Team Studio System Requirements*
- *TIBCO® Data Science Team Studio Version and Licensing*
- *TIBCO® Data Science Team Studio Installation and Administration*
- *TIBCO® Data Science Team Studio User's Guide*
- *TIBCO® Data Science Team Studio Development Kit*

How to Contact TIBCO Support

You can contact TIBCO Support in the following ways:

- For an overview of TIBCO Support, visit <http://www.tibco.com/services/support>.
- For accessing the Support Knowledge Base and getting personalized content about products you are interested in, visit the TIBCO Support portal at <https://support.tibco.com>.
- For creating a Support case, you must have a valid maintenance or support contract with TIBCO. You also need a user name and password to log in to <https://support.tibco.com>. If you do not have a user name, you can request one by clicking Register on the website.

System Requirements for TIBCO® Data Science Team Studio

For information about the system requirements for Team Studio, see *TIBCO® Data Science Team Studio System Requirements*.

How to Join TIBCO Community

TIBCO Community is the official channel for TIBCO customers, partners, and employee subject matter experts to share and access their collective experience. TIBCO Community offers access to Q&A forums, product wikis, and best practices. It also offers access to extensions, adapters, solution accelerators, and tools that extend and enable customers to gain full value from TIBCO products. In addition, users can submit and vote on feature requests from within the [TIBCO Ideas Portal](https://community.tibco.com). For a free registration, go to <https://community.tibco.com>.

TIBCO® Data Science Team Studio Release Notes

The release notes for this product version are provided to inform you of new features, known issues, and issues from previous releases that have been closed.

These release notes are for TIBCO® Data Science Team Studio version 6.5.0.

Team Studio is an environment built to facilitate the analytics process at all levels of the business, from defining a problem to driving business action.

New Features

The following new features were added to version 6.5.0 of Team Studio.

New Documentation

Documentation for TIBCO® Data Science Team Studio has been expanded, improved, and moved to the TIBCO documentation portal, at <https://docs.tibco.com/products/tibco-data-science-team-studio>. You can find example data sets on this docsite, also.

Support for data sources

You can now configure Team Studio to connect to the following data sources.

- Azure SQL Data Warehouse
- Google BigQuery
- Oushu 3.1
- Cloudera 6
- Google DataProc
- Greenplum Database 5

Running Python in Jupyter Notebooks for Team Studio on JupyterHub

You can now run the Python engine as a local managed process in Jupyter Notebooks for Team Studio, in addition to running it in a Docker container.

You can configure the runtime environment so the Python engine persists in the environment.

Hive impersonation in Cloudera

If a user logs into the Hive data source, the Resource Manager now reflects the user's name, rather than the generic "alpine user".

Integration with TIBCO® Statistica

Statistica workflow files (*.sdm) are now recognized and filterable through the Workfile browser.

New operators

You can use the following new operators in your workflows. See the help for these operators in the *TIBCO® Data Science Team Studio User's Guide*.

Operator name	Description
Unstack	Takes an HDFS dataset in stacked format and produces an unstacked (wide) HDFS dataset using user-specified grouping and pivot columns.
Line Chart	Produces a line chart for a user-specified X-axis and Y-axis, and aggregates the values in the Y-axis.

Operator name	Description
Generalized Linear Regression Models	Fits a regression model to predict a dependent variable that follows some distribution from the exponential family of distributions.

Deprecated and Removed Features

The following features have been deprecated or removed from version 6.5.0 of Team Studio.

Support for the following products is being deprecated in this release. Support will be removed in a future release.

- Greenplum 4.0.5.x, 4.1.x, 4.2.x, 4.3.1, 4.3.5, 4.3.6, 4.3.7, 4.3.8, 4.3.11.x
- Microsoft SQL Server 2012
- Oracle 10g
- Pivotal HAWQ 1.2
- PostgreSQL 8.4 and 9.3
- Python 2
- Java 7

Migration and Compatibility

If you are migrating from a previous release to version 6.5.0 of Team Studio, contact Team Studio Support for assistance.

There are no migration or compatibility issues between version 6.4.1 and 6.5.0 of Team Studio.

Changes in Platform Support

There are no changes in platform support in version 6.5.0 of Team Studio.

For a complete list of all the currently supported platforms, see the [System Requirements](#).

Closed Issues

Version 6.5.0 of Team Studio contains assorted bug fixes.

Key	Summary
DSC-1056	You can now use default workflow variables (such as <code>@user_name</code>) when using an import or export operator.
DSC-23045	Data type matching issues have been resolved for the Teradata data source.
DSC-23359	Spark AutoTuning now works if the Yarn Web UI uses Spnego to authenticate.
DSC-23365	The ScatterPlot Matrix operator was not using the correct value for the maximum number of scatter points. Now, it correctly defaults to 200, or it pulls the value from the configuration file for <code>alpine.ui.max_scatter_points</code> .

Key	Summary
DSC-23429	Previously, although user permissions were respected under Kerberos impersonation, jobs run under Kerberos impersonation always ran as the Kerberos Principal. They now either run as the logged-in user (if Kerberos impersonation is checked) or are run using the Hadoop Credentials (if Kerberos Impersonation is not checked).
DSC-23543	The contents of the Data Source explorer tab would be displayed as empty if you navigated to it quickly after opening a workflow but before the workflow was fully loaded.
DSC-23623	Previously, Spark-enabled operators would fail to retrieve the Hadoop configuration if the option <code>yarn.resourcemanager.webapp.https.address</code> was enabled. This problem has been fixed.

Known Issues

The table lists known issues in version 6.5.0 of Team Studio.

Key	Summary
DSC-17859	<p>Summary</p> <p>Avro file formats do not support null values and can cause errors when running downstream operators.</p> <p>Workaround</p> <p>Clean your data prior to using the Avro format by using the Null Value Replacement operator.</p>
DSC-22778	<p>Summary</p> <p>In the Hadoop Join operator, clicking a check box to add or remove a column appears to not work.</p> <p>Workaround</p> <p>Click twice to select the check box for adding or removing columns in a Hadoop Join operator. (This problem is due to a limitation in the user interface framework.)</p>
DSC-22859	<p>Summary</p> <p>A Spark normalization run on a very wide data set fails due to an issue in Spark versions prior to 2.3.</p> <p>Workaround</p> <p>Use the MR version on a very wide dataset, or set the option to not use Spark.</p>
DSC-22930	<p>Summary</p> <p>The compression format Snappy is not supported by the file format CSV.</p> <p>Workaround</p> <p>For this file format, use another compression format.</p>

Key	Summary
DSC-23121	<p>Summary</p> <p>The Hadoop Hive 1.1 client libraries that Team Studio is built with do not yet support TLS/SSL encryption when combined with Kerberos authentication.</p> <p>Workaround</p> <p>Do not enable Hive TLS/SSL encryption when Kerberos authentication is also enabled for your Hadoop cluster.</p>
DSC-23585	<p>Summary</p> <p>When you use a Google BigQuery data source in parallel mode, Copy to Database and Copy to Hadoop do not work.</p> <p>Workaround</p> <p>Use simple mode to copy the data to and from a BigQuery data source.</p>
DSC-23587	<p>Summary</p> <p>When you use a Google BigQuery data source, Copy Between Databases does not work with a Greenplum database.</p> <p>Workaround</p> <p>None.</p>
DSC-23629	<p>Summary</p> <p>For BigQuery data sources, using DML issued through JDBC can cause problems because of statement size limitations; therefore, you should avoid using Team Studio operators to copy data into BigQuery.</p> <p>Workaround</p> <p>Google recommends that BigQuery data source users insert data using the Google API. For more information on BigQuery limitations, see https://cloud.google.com/bigquery/quotas.</p>
DSC-23670	<p>Summary</p> <p>Copy to Hadoop and Copy to Database operators do not work in parallel mode between GPDB5 and Hadoop Data Sources.</p> <p>Workaround</p> <p>None</p>
DSC-23804	<p>Summary</p> <p>For the data source format BigQuery, a Naive Bayes analysis fails if a datetime column is included in the data.</p> <p>Workaround</p> <p>None.</p>

Key	Summary
DSC-23858	<p>Summary After you update the email configuration through the user interface, you must restart the server for the configuration changes to take effect.</p> <p>Workaround None.</p>
DSC-341	<p>Summary Hive does not handle new lines in text.</p> <p>Workaround Clean all text data prior to processing.</p>
DSC-342	<p>Summary Non-Spark operators can fail for a newly-added Active Directory user.</p> <p>Workarounds</p> <ul style="list-style-type: none">• The user can use the Spark versions of the operators in the workflow.• The administrator can issue a <code>kinit</code> command for the mapped user on the cluster.