

TIBCO® Data Science - Team Studio

Release Notes

Version 7.1.0 | September 2023

Document Updated: December 2023



Contents

Contents	2
New Features	3
Changes in Platform Support	4
Changes in Functionality	5
Deprecated Features	6
Removed Features	7
Migration and Compatibility	8
Closed Issues	9
Known Issues	10
TIBCO Documentation and Support Services	22
Legal and Third-Party Notices	24

New Features

The following features have been added in this release of TIBCO® Data Science - Team Studio.

Compute engine enhancement

Upgraded to Apache Spark 3.3.

Addition of Apache Spark specific operators

- Isolation Forest operator
- Import Excel operator

Performance enhancement

- Integration of TIBCO® Data Virtualization with TIBCO Data Science Team Studio.
- Added Shared Volume Optimized Data Access for Remote Sources method for accessing data in TIBCO Data Virtualization from TIBCO Data Science Team Studio.

Jupyter Notebook enhancement

The rack awareness is disabled for the Jupyter Notebook.

Changes in Platform Support

In this release, platform support for TIBCO® Data Science - Team Studio has changed as follows:

Platform	Status	As of Release
Red Hat Enterprise Linux 8.6	Added	7.1.0
CentOS Stream 9	Added	7.1.0
Debian GNU/Linux 12	Added	7.1.0
Ubuntu 22.04.2 LTS	Added	7.1.0
Oracle Linux Server 8.8	Added	7.1.0
CentOS 8	Removed	7.1.0

For a complete list of supported platforms, see the Readme file.

Changes in Functionality

No functionality changes have been made in this release of TIBCO® Data Science - Team Studio.

Deprecated Features

No features have been deprecated in this release of TIBCO® Data Science - Team Studio.

Removed Features

No features have been removed in this release of TIBCO® Data Science - Team Studio.

Migration and compatibility are not affected in this release of TIBCO® Data Science - Team Studio.

Closed Issues

The following issues have been fixed in this release of TIBCO® Data Science - Team Studio.

Key	Summary
TDS-889	Summary: If you repeatedly run certain workflows or jobs, it causes the TIBCO Data Virtualization server associated with TIBCO Data Science - Team Studio to run out of memory.
TDS-653	Summary: The Join operator displays an error message when selecting columns with the same name but different case letters. The operator compares the columns in a case-sensitive way.
	For example, if a data set has a column name id and another data set has a column name ID , then the Join operator treats them as two different columns.
TDS-430	Summary: All the data sets and operators are run again when you click Run or Step Run in a New Workflow without changing the workflow.
TSDEV- 26091	Summary: When a column in the Excel file has mixed data types, the leading row derives the column data type in the Import Excel operator.
TDS-26086	Summary: When an operator fails while creating a table, the table in TIBCO Data Virtualization is deleted but the table exists in the data source. This results in the failure of operators on every subsequent run.

Known Issues

The following issues exist in this release of TIBCO® Data Science - Team Studio.

Key	Summary and Workaround
TDS-1437	Summary: In Legacy Workflow, Row Filter, Null Value Replacement, and Variable operator may fail with script mode. Workaround: None.
TDS-1345	Summary: The backup and restore from version 6.6.0 to 7.1.0 is not supported. This is due to the restrictions made in the Ruby file that handles the migration of the PostgreSQL database. Workaround: Upgrade 6.6.0 to 7.0.x, and then 7.0.x to 7.1.0.
TDS-1332	Summary: After a workflow finishes running on the cluster, the operators may not display the results. Workaround: Modify the upstream most operator and rerun the workflow, or, you can copy the workflow to a new workflow and run them.
TDS-1331	Summary: The following connection exception error may appear when a workflow is run. java.net.ConnectException: Connection refused (Connection refused) Workaround: Rerun the workflow.
TDS-1326	Summary: When a workflow is run, the Results console may not show the progress of a workflow. Workaround: Refresh your web browser.
TDS-1319	Summary: When you are using Read Dataset using Direct

Connection in a Jupyter Notebook with the shared file system writing back to TIBCO Data Virtualization may fail. Workaround: Use parquet files instead of CSV files. TDS-1305 Summary: When the owner of a workspace is deleted, you cannot change the data sources of the workflow. Workaround: Copy the workflow to the same or a different workspace. TDS-1274 Summary: The output of the Join operator from the Legacy Workflow engine cannot be used as an input for the New Workflow engine if the output column name contains a dot (.) Workaround: None. TDS-1272 Summary: An error may appear while using the Jupyter Notebook with Apache Spark Standalone cluster manager version 3.2. Workaround: Use the Apache Spark version 3.3 or CDP7 clust TDS-1269 Summary: An error may appear while creating the Conda environment due to a difference in the Python components version. Workaround: While creating a custom kernel, use the support	
TDS-1305 Summary: When the owner of a workspace is deleted, you cannot change the data sources of the workflow. Workaround: Copy the workflow to the same or a different workspace. TDS-1274 Summary: The output of the Join operator from the Legacy Workflow engine cannot be used as an input for the New Workflow engine if the output column name contains a dot (.) Workaround: None. TDS-1272 Summary: An error may appear while using the Jupyter Notebook with Apache Spark Standalone cluster manager version 3.2. Workaround: Use the Apache Spark version 3.3 or CDP7 clust on the Python components version.	٦,
cannot change the data sources of the workflow. Workaround: Copy the workflow to the same or a different workspace. TDS-1274 Summary: The output of the Join operator from the Legacy Workflow engine cannot be used as an input for the New Workflow engine if the output column name contains a dot (.) Workaround: None. TDS-1272 Summary: An error may appear while using the Jupyter Notebook with Apache Spark Standalone cluster manager version 3.2. Workaround: Use the Apache Spark version 3.3 or CDP7 clust TDS-1269 Summary: An error may appear while creating the Conda environment due to a difference in the Python components version.	
TDS-1274 Summary: The output of the Join operator from the Legacy Workflow engine cannot be used as an input for the New Workflow engine if the output column name contains a dot (.) Workaround: None. TDS-1272 Summary: An error may appear while using the Jupyter Notebook with Apache Spark Standalone cluster manager version 3.2. Workaround: Use the Apache Spark version 3.3 or CDP7 clust TDS-1269 Summary: An error may appear while creating the Conda environment due to a difference in the Python components version.	
Workflow engine cannot be used as an input for the New Workflow engine if the output column name contains a dot (.) Workaround: None. TDS-1272 Summary: An error may appear while using the Jupyter Notebook with Apache Spark Standalone cluster manager version 3.2. Workaround: Use the Apache Spark version 3.3 or CDP7 clust TDS-1269 Summary: An error may appear while creating the Conda environment due to a difference in the Python components version.	
TDS-1272 Summary: An error may appear while using the Jupyter Notebook with Apache Spark Standalone cluster manager version 3.2. Workaround: Use the Apache Spark version 3.3 or CDP7 clust TDS-1269 Summary: An error may appear while creating the Conda environment due to a difference in the Python components version.	•
Notebook with Apache Spark Standalone cluster manager version 3.2. Workaround: Use the Apache Spark version 3.3 or CDP7 clust TDS-1269 Summary: An error may appear while creating the Conda environment due to a difference in the Python components version.	
TDS-1269 Summary: An error may appear while creating the Conda environment due to a difference in the Python components version.	
environment due to a difference in the Python components version.	er.
Workaround: While creating a custom kernel, use the suppor	
version of the Python components mentioned in the Jupyter Notebook.	ted
TDS-1242 Summary: The Shared Volume Optimized data access for Remote Sources method may fail when using the Shared File System.	
Workaround: The directory of the schema and the directory used by TIBCO Data Virtualization for Shared File System mus be the same.	t
TDS-1215 Summary: The following error may appear when you run a	

Key	Summary and Workaround
	workflow.
	status : 500, message : workflow not found
	Workaround: Rerun the workflow.
TDS-1196	Summary: The Normalization operator and all the Modeling operators that use the Vector Assembler take longer to complete. This is an issue with Apache Spark version 3.3. Workaround: None.
TDS-1178	Summary: When the server and user browsers are in different timezones, the logs are displayed without indicating the actual timezone.
	Workaround: None.
TDS-1155	Summary: The operators may display truncated results while you are using the CDP7 cluster.
	Workaround: In the TIBCO Data Virtualization configuration of CDP7, set the Default String Length parameter to 32678 and Concurrent Request Limit to 0 .
TDS-1069	Summary: The Null Value Replacement operator fails to run on wide data with 3,000 columns.
	Workaround: Use the Null Value Replacement operator in multiple phases, where the operator in each step replaces a different subset of columns.
TDS-1040	Summary: Jupyter Notebook may get timed out for Job Scheduler when it runs for more than 60 minutes.
	Workaround: Increase the value of the timeout parameter in the jupyterhub_config.py file, and then restart the Notebook container.

Key Summary and Workaround

For example, the value of str(60*60*1000) is increased to str (90*60*1000) in the following snippet.

```
c.ConfigurableHTTPProxy.command =
[
'configurable-http-proxy',
'--log-level', proxy_log_level,
'--timeout', str(90*60*1000)
]
```

TDS-1016

Summary: When a workflow is exported from 7.0.0 and imported in 7.1.0, the proxy operators do not appear in the workflow.

Workaround: After the workflow is imported, the Linux administrator must run the following command:

tds alpine amend

TDS-972

Summary: When an S3 data source is used as the shared file system, the Jupyter Notebook may not work.

Workaround: Use different data sources as the Shared File System.

Note: It is recommended not to use an S3 data source as the Shared File System.

TDS-958

Summary: While using the Spark Standalone cluster, an error may appear when you try to write data on the shared volume.

Workaround: Add the following line to the PYSPARK_SUBMIT_ ARGS environment variable:

--conf spark.hadoop.fs.permissions.umask-mode=000

TDS-941

Summary: Duplicate entries of data sources result in an error

Key	Summary and Workaround
	while running a workflow.
	Workaround: Delete the data source and reenter the data source.
TDS-928	Summary: The workflow variables provided with the REST API to run a workflow override the user-defined workflow variable in Alpine (managed from Actions > Workflow Variables option on the UI).
	Workaround: Validate the workflow variables before running the workflow through the UI.
TDS-923	Summary: The dollar (\$) symbol is not accepted as a workflow variable for the regular expression in the Dynamic Column Filter operator.
	Workaround: Use \Z instead of \$ character as a workflow variable.
TDS-855	Summary: The operator can fail due to OOTB CDATA configuration in the Amazon Redshift adapter.
	Workaround: Contact the TIBCO Data Virtualization support for the required configuration.
TDS-849	Summary: When the Disjoint parameter is set to True, the Random Sampling operator takes a substantial amount of time to run with the S3 data source.
	Workaround: None.
TDS-831	Summary: When the Null Value Replacement operator is not configured, an error might occur while running the workflow.
	Workaround: Configure the Null Value Replacement operator.
TDS-826	Summary: The Join operator fails when you join by columns where the column name contains multibyte characters such as Chinese, Japanese, or Korean. This is an issue with Spark.

Key	Summary and Workaround
	Workaround: Do not use multibyte characters on the joiner columns.
TDS-824	Summary: When a user creates a clause in script mode to filter the data in the Row Filter operator, an error message appears if the column name contains Chinese, Japanese, Korean, or special characters such as a dot, or only a number.
	Workaround: Enclose the column names with backticks (` `) in the clause. For example, if a column name is ab.c, then use `ab.c`.
	Note: The use of dots in column names is not recommended. Some operators may not work when the column name contains a dot.
TDS-802	Summary: The Docker container cannot resolve the DNS name mapped with the TIBCO Data Virtualization services.
	Workaround: Install TIBCO Data Science - Team Studio and TIBCO Data Virtualization on a different machine.
TDS-762	Summary: The Stop button in the workflow canvas does not stop the actual Spark job in a New Workflow engine.
	Workaround: To stop the application using the Spark web user interface, perform these steps:
	 Access the Spark master web user interface of the configured data source.
	Navigate to a driver or application that you want to stop, and click Kill.
	To stop the application using the Hadoop YARN user interface, perform these steps:
	 Access the Hadoop YARN web user interface of the configured data source.

Key **Summary and Workaround** 2. Navigate to a driver or application that you want to stop, and click Kill Application. TDS-746 Summary: A temporary schema (Compute_temp) is available in TIBCO Data Science - Team Studio for the model trainer operators that do not provide UI for the user to specify the output schema. **Note:** Do not use **Compute_temp** schema for your workflow. Workaround: None. TDS-745 **Summary:** The model imported in the **New Workflow** engine using the Load Model operator does not display the summary of the loaded model. Workaround: None TDS-742 Summary: The Export to SBDF and Export Model to **Workspace** operators fail for larger data sets. The following error message appears due to the file size limit of the Chorus workspace: java.lang.Exception: could not create file because of: status=422 Workaround: Increase the file size of work files in file_sizes_ mb.workfiles. Note: The maximum limit applicable for the file_sizes_ mb.workfiles configuration is 2,048 MB. TDS-737 **Summary:** The **PCA** operator fails with a data set that has more than 1,000 columns because of an issue in Apache Spark. Workaround: None. TDS-731 **Summary:** The following errors may appear when you edit a

Key	Summary and Workaround
	workflow:
	actor name is not unique!
	Workaround: Close the workflow and then reopen it.
TDS-723	Summary: When using a Notebook as a Python Execute operator in legacy workflows, the operator remains invalid when passing three substitute inputs.
	Workaround: The cc.write_output_file in the notebook python code (use_output_substitution=True argument) should either have an explicit header=False argument or should not have any header argument.
	Note: The CSV file data used as inputs should not have a header line.
	If the Python Execute operator is a terminal operator in a workflow or if the cc.write_output_file does not have use_output_substitution as True, then the header argument can be either True or False.
TDS-676	Summary: When all the columns are removed while running the Column Cleanser operator, the results are not visible.
	Workaround: Refresh the workflow page.
TDS-659	Summary: If your computer goes into sleep mode while running a workflow, then Click on individual operators to see their output logs may appear on the result tab.
	Workaround: Refresh your web browser or reopen the workflow.
TDS-571	Summary: An error might occur while using a Output Table name that exists in the underlying data source.
	Workaround: Use a new name in the Output Table parameter

Key	Summary and Workaround
	or contact the system administrator to remove the existing table from the underlying data source.
TDS-568	Summary: The Set Operations operator accepts a maximum of two data sets as inputs.
	Workaround: None.
TDS-567	Summary: The Join operator accepts a maximum of two data sets as inputs.
	Workaround: None.
TDS-537	Summary: When the Run Now button is clicked in the workflow summary page, the Activity tab does not provide information about the workflow execution. However, the results can be viewed from the Your Workflow Results option.
	Workaround: None.
TDS-534	Summary: Some operators do not respond to the Store All Results option in Legacy Workflow Engine.
	Workaround: None.
TDS-319	Summary: The Copy Between Databases option is not supported for TDV object storage.
	Workaround: None.
TSDEV- 25979	Summary: When you insert data sets into a TIBCO® Data Virtualization data source from a Python notebook, ensure that the tdv_scripts_location in the ChorusCommander field matches the one set in your alpine.config file. By default, cc.tdv_scripts_location is set to Util.
	Workaround: For example, if your startBatch, addBatch, and loadBatch scripts are in a TIBCO® DV schema called Utilities, then use cc.tdv_scripts_location='Utilities' before calling cc.write_output_table.

Key	Summary and Workaround
TSDEV- 25838	Summary: In the case of parquet library conflict, operators cannot write or infer parquet files successfully for EMR data sources.
	Workaround: In this case, use Avro or CSV as the output format from operators.
TSDEV- 25835	Summary: R Execute operator does not work with EMR S3 and CDP7.
	Workaround: None.
TSDEV- 25810	Summary: The column names for TIBCO Data Virtualization metadata tables are stored in the order in which they are modified, so when TIBCO® Data Science - Team Studio queries the metadata table, the column names can be displayed in the wrong order in its user interface.
	Workaround: You can correct this problem by republishing the view/table in TIBCO Data Virtualization with the correct column order.
TSDEV- 25806	Summary: TIBCO Data Virtualization object storage does not support the Python Execute operator with Notebooks. Workaround: None.
TSDEV- 25776	Summary: For the TIBCO Data Virtualization data source, the Copy to Database operator (Copy mode > Parallel option) does not work. The operator displays an explicit error message, prompting you to use the Simple mode for copying.
	Workaround: To use the Copy to Database operator with a TIBCO Data Virtualization data source, use Simple mode.
TSDEV- 25774	Summary: The Copy to Hadoop operator does not support appending to an existing Parquet or Avro file on HDFS. This process is not supported by the Hadoop API used by TIBCO® Data Science - Team Studio.

Key	Summary and Workaround
	Workaround: None.
TSDEV- 25747	Summary: The Use Array parameter of Pivot operator does not work with a TIBCO Data Virtualization data source because TIBCO Data Virtualization does not support array types.
	Workaround: None.
TSDEV- 25657	Summary: An exported workflow lost many operators from the original workflow.
	Workaround: By using SSH access, the administrator can edit the workfileid.inf file to change the version number in the file to the latest version of the work file.
	You can find this file under /usr/chorus/shared/ALPINE_DATA_ REPOSITORY/flow/Public/.
	The latest version of the file can be found at /usr/chorus/shared/ALPINE_DATA_REPOSITORY/flow_version directory.
TSDEV-	Summary: A PySpark notebook that uses Avro can fail.
25472	Workaround: Change the PySpark initialization to os.environ ['PYSPARK_SUBMIT_ARGS'] = "master yarn-clientnum-executors 1executor-memory 1gpackages org.apache.spark:sparkavro_ 2.11:2.4.3 pyspark-shell". Older PySpark notebooks accessing avrò files must be migrated to use this change for the notebooks to work with the new notebook server.
	This is a known issue with Spark. See SPARK-27623 for more information.
TSDEV- 23858	Summary: After updating the email configuration through the user interface, you must restart the server for the configuration changes to take effect.
	Workaround: None.

Key	Summary and Workaround
TSDEV- 22930	Summary: The compression format, Snappy, is not supported by the CSV file format.
	Workaround: For this file format, use another compression format.
TSDEV- 22859	Summary: If you try to run Spark normalization on a wide data set, it fails due to an issue in Spark versions before 2.3.
	Workaround: Use the MR version on a wide dataset, or set the option not to use Spark.
TSDEV- 22778	Summary: In the Hadoop Join operator, clicking a checkbox to add or remove a column does not work.
	Workaround: Double-click the checkbox to select the checkbox for adding or removing columns in a Hadoop Join operator. This issue occurs due to a limitation in the user interface framework.

TIBCO Documentation and Support Services

For information about this product, you can read the documentation, contact TIBCO Support, and join TIBCO Community.

How to Access TIBCO Documentation

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

The Product Documentation website is updated frequently and is more current than any other documentation included with the product.

Product-Specific Documentation

The documentation for this product is available on the TIBCO® Data Science - Team Studio Product Documentation page.

How to Contact Support for TIBCO Products

You can contact the Support team in the following ways:

- To access the Support Knowledge Base and getting personalized content about products you are interested in, visit our product Support website.
- To create a Support case, you must have a valid maintenance or support contract
 with a Cloud Software Group entity. You also need a username and password to log
 in to the our product Support website. If you do not have a username, you can
 request one by clicking Register on the website.

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

23 TIBCO Documentation and Support Services	
requests from within the TIBCO Ideas Portal. For a free registration, go to TIBCO Community.	

SOME CLOUD SOFTWARE GROUP, INC. ("CLOUD SG") SOFTWARE AND CLOUD SERVICES EMBED, BUNDLE, OR OTHERWISE INCLUDE OTHER SOFTWARE, INCLUDING OTHER CLOUD SG SOFTWARE (COLLECTIVELY, "INCLUDED SOFTWARE"). USE OF INCLUDED SOFTWARE IS SOLELY TO ENABLE THE FUNCTIONALITY (OR PROVIDE LIMITED ADD-ON FUNCTIONALITY) OF THE LICENSED CLOUD SG SOFTWARE AND/OR CLOUD SERVICES. THE INCLUDED SOFTWARE IS NOT LICENSED TO BE USED OR ACCESSED BY ANY OTHER CLOUD SG SOFTWARE AND/OR CLOUD SERVICES OR FOR ANY OTHER PURPOSE.

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

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 Cloud Software Group, Inc.

TIBCO, the TIBCO logo, the TIBCO O logo, Spotfire, and Alpine Data Labs are either registered trademarks or trademarks of Cloud Software Group, 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. You acknowledge that all rights to these third party marks are the exclusive property of their respective owners. Please refer to Cloud SG's Third Party Trademark Notices (https://www.cloud.com/legal) for more information.

This document includes fonts that are licensed under the SIL Open Font License, Version 1.1, which is available at: https://scripts.sil.org/OFL

Copyright (c) Paul D. Hunt, with Reserved Font Name Source Sans Pro and Source Code Pro.

Cloud SG 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. See the "readme" file for the availability of a specific version of Cloud SG software 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. CLOUD SG MAY MAKE IMPROVEMENTS AND/OR CHANGES IN THE PRODUCT(S), THE PROGRAM(S), AND/OR THE SERVICES DESCRIBED IN THIS DOCUMENT AT ANY TIME WITHOUT NOTICE.

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 "README" FILES.

This and other products of Cloud SG may be covered by registered patents. For details, please refer to the Virtual Patent Marking document located at https://www.tibco.com/patents.

Copyright © 2017-2023. Cloud Software Group, Inc. All Rights Reserved.