

StreamSets Transformer

Cumulative 3.13.x Release Notes

March 31, 2020

We're happy to announce a new version of StreamSets Transformer. Version 3.13.x contains several new features, enhancements, and some important bug fixes in the following release:

- Version 3.13.0 - March 31, 2020

This document contains important information about the following topics:

- [New Features and Enhancements](#)
- [Upgrade Information](#)
- [Fixed Issues in 3.13.0](#)
- [Known Issues](#)
- [Documentation](#)
- [Contact Information](#)

New Features and Enhancements

Transformer 3.13.x includes the following new features and enhancements:

New Cluster Types

- [Amazon EMR support](#) - You can now run pipelines on an Amazon EMR cluster, version 5.13.0 or later.
- **Kubernetes support** - You can now run pipelines on a Kubernetes cluster.
- [MapR support](#) - You can now run pipelines on a MapR version 6.1.0 Hadoop YARN cluster.

New Stages

- [Amazon Redshift destination](#) - You can now write to an Amazon Redshift table.
- **Azure Event Hubs** [origin](#) and [destination](#) - You can now read from and write to Microsoft Azure Event Hubs.
- **MapR FS origin and destination** - You can now read from and write to MapR FS.
- **MapR Event Store origin and destination** - You can now read from and write to MapR Event Store.
- [MySQL JDBC Table origin](#) - You can now read from a MySQL database table.
- [Oracle JDBC Table origin](#) - You can now read from an Oracle database table.

- [PostgreSQL JDBC Table origin](#) - You can now read from a PostgreSQL database table.
- [SQL Server JDBC Table origin](#) - You can now read from a Microsoft SQL Server database table.

General Enhancements

- [Azure credential store](#) - You can now use an Azure credential store with Transformer.
- **Reverse proxy support** - You can now register Transformer instances deployed behind a reverse proxy with Control Hub.

Pipeline Enhancements

- **Callback URL** - You can now specify a custom callback URL that the cluster uses to communicate with Transformer. This property can be configured on the Advanced tab of the pipeline properties.
- **Cluster preview** - You can now preview pipelines using the cluster configured for the pipeline, in addition to previewing data using the local Transformer machine.

Stage Enhancements

- **Amazon stages** - Amazon stages can now connect using Amazon Web Services libraries for Hadoop 2.7.7.
- **Delta Lake destination** - When using the Upsert using Merge write mode, you can now specify a timestamp column to ensure that the latest record is upserted when a batch contains multiple versions of a record.
- **Glob patterns allowed in origin directory paths** - You can now use glob patterns to specify the read directory for the following origins: ADLS Gen1, ADLS Gen2, Amazon, File.
- **Snowflake origin** - You can now use the SELECT command as well as COPY UNLOAD to read from Snowflake.
- **Snowflake stages** - You can now configure advanced Snowflake properties to pass to Snowflake. This property can be configured on the Advanced tab of the stage properties.

Upgrade Information

To upgrade existing versions of StreamSets Transformer registered with Control Hub, complete the following steps for each registered Transformer.

We recommend upgrading one Transformer at a time and verifying that jobs continue to run as expected.

1. In Control Hub, stop all jobs running on the Transformer to be upgraded.
2. Stop the Transformer.
3. Download Transformer 3.13.x from the [StreamSets website](#).
4. [Install and configure Transformer 3.13.x](#).

5. Compare the previous and new versions of the `$TRANSFORMER_CONF/transformer.properties` file, and update the new file as needed with the same customized property values.
6. [Start the 3.13.x version of Transformer.](#)
7. [Register the 3.13.x version of Transformer with Control Hub.](#)
8. To have Transformer 3.13.x run the same jobs as the earlier version of Transformer, [assign the same label to the 3.13.x version of Transformer.](#)
9. Restart Transformer jobs.
10. If jobs run as expected, [unregister the old version of Transformer.](#)

Fixed Issues in 3.13.0

Transformer 3.13.0 includes the following fixed issues:

Issue	Description
TRANSFORM-1777	The SQL Evaluator and Field Replacer processors do not retain field types for null values. This sometimes causes errors when writing to destinations.
TRANSFORM-1401	You cannot use a PySpark processor in a pipeline that runs on a Databricks cluster at this time.

Known Issues

Transformer 3.13.x includes the following known issues:

Issue	Description
TRANSFORM-1874	The Amazon Redshift destination does not support writing to Date, Timestamp, or Decimal fields with 3.13.0. This will be fixed in a future release.
TRANSFORM-1776	Due to an unresolved Microsoft SQL Server BDC issue, most Transformer pipelines run on SQL Server BDC clusters fail with errors. Workaround: In the SQL Server BDC cluster, remove the <code>mssql-mleap-lib-assembly-1.0.jar</code> file from the following HDFS ZIP file: <code>/system/spark/spark_libs.zip</code> .
TRANSFORM-1596	Restarting a cluster pipeline shortly after starting it can cause the cluster to use the same Spark application ID for both pipeline runs, leading to errors. Workaround: Wait a few seconds before starting a cluster pipeline again.
TRANSFORM-1358	Scala compilation fails for pipelines that run in Hadoop YARN cluster deployments. This affects pipelines with a preprocessing script or Scala processor that run on Hadoop YARN with the Deployment pipeline property set to Cluster.

	Workaround: Set the Deployment Mode pipeline property to Client.
TRANSFORM-495	The File origin processes files with mixed schemas.

Documentation

You can view the latest Transformer documentation in the Control Hub online help. Transformer details appear in the following chapters: [Transformer](#) and [Pipeline Design for Transformer](#).

Contact Information

For more information about StreamSets, visit our website: <https://streamsets.com/>.

Check out our Documentation page for doc highlights and tutorials: <https://streamsets.com/docs>.

To report an issue, to get help from our Google group, Slack channel, or Ask site, or to find out about our next meetup, check out our Community page: <https://streamsets.com/community/>.

For general inquiries, email us at info@streamsets.com.