

StreamSets Transformer

Cumulative 3.15.x Release Notes

September 10, 2020

We're happy to announce a new version of StreamSets Transformer. Version 3.15.x contains several new features, enhancements, and some important bug fixes in the following release:

- Version 3.15.0 - September 10, 2020

This document contains important information about the following topics:

- [New Features and Enhancements](#)
- [Upgrade Information](#)
- [Fixed Issues in 3.15.0](#)
- [Known Issues](#)
- [Additional Resources](#)

New Features and Enhancements

Transformer 3.15.x includes the following new features and enhancements:

New Cluster Type

- [Dataproc](#) - You can now run pipelines on a Google Dataproc cluster. Transformer supports Dataproc image versions 1.3.x and 1.4.x.

Cluster Enhancements

- [Hadoop YARN cluster with Kerberos authentication](#) - When configuring a pipeline to run on a Hadoop YARN cluster with Kerberos authentication, you can store the Kerberos keytab in a credential store, and then use a credential function to call the keytab from the pipeline properties.

When using a credential store, you can require Transformer group access to the keytab so that you can more securely use multiple keytabs for a Transformer instance.

- [SQL Server 2019 Big Data Cluster version support](#) - Transformer 3.15 supports Cumulative Update 4 and later for SQL Server 2019.

New Stages

You can use the following new stages only in pipelines that run on Dataproc clusters:

- **Google Cloud Storage origin** - You can use this origin to read fully written objects in Google Cloud Storage.

- **Google Cloud Storage destination** - You can use this destination to write to Google Cloud Storage.
- **Google Big Query destination** - You can use this destination to write to Google BigQuery.

Stage Enhancements

- **Amazon S3 stages** - You can specify the AWS region where the Amazon S3 bucket resides. Previously, the stages used US East (N. Virginia) us-east-1, which is the default AWS region.
- [Amazon S3 destination](#) - You can specify fields to partition the data by. You can also use the destination in a slowly changing dimension pipeline to write results to a file dimension on Amazon S3.
- [Azure SQL destination](#) - The destination is no longer considered a Technology Preview feature and is approved for use in production.
- **Hive destination** - You can enable the destination to compensate for data drift by creating columns in the destination table for new fields.
- [Slowly Changing Dimension processor](#) - The processor no longer requires that the configured tracking fields exist in the master dimension data.

Additional Enhancements

- [File-based user authentication](#) - When Transformer is configured for file-based authentication, you can use the Transformer UI to change your password.
- **Pipeline run history** - The pipeline run history displays the input, output, and error record count for each pipeline run.
- [Sample pipelines](#) - Transformer provides several sample pipelines that you can use to learn about Transformer features or as a basis for building your own pipelines.

Upgrade Information

You can upgrade previous versions of Transformer to version 3.15.x. For complete instructions on upgrading, see the [Upgrade documentation](#).

Fixed Issues in 3.15.0

Transformer 3.15.0 includes the following fixed issues:

Issue	Description
TRANSFORM-2282	When Transformer is installed through a cloud service provider marketplace or from an RPM package on CentOS 7, Oracle Linux 7, or Red Hat Enterprise Linux 7, the TRANSFORMER_RESOURCE environment variable incorrectly points to a directory that does not exist, <code>/opt/streamsets-datacollector/resources</code> .

	With this fix, the environment variable points to the following directory, as expected: <code>/var/streamsets-datacollector/resources</code> .
TRANSFORM-2153	A pipeline fails to start on a Hadoop YARN cluster that uses Kerberos authentication when the pipeline uses cluster deployment mode and Transformer is configured to impersonate the currently logged in user.
TRANSFORM-2145	The JDBC Table origins ignore the Load Data Only Once and Cache Data properties.
TRANSFORM-2072	When concurrently starting multiple pipelines on the same Amazon EMR cluster using the same staging directory, Transformer starts the pipelines sequentially with a delay between each start.
TRANSFORM-2052	A pipeline with a Kudu origin or destination that runs on a Hadoop YARN cluster using Kerberos authentication fails with the following error: <code>Cannot import authentication data from a different user</code>
TRANSFORM-1776	Due to an unresolved Microsoft SQL Server BDC issue, most Transformer pipelines run on SQL Server BDC clusters fail with errors.

Known Issues

Transformer 3.15.x includes the following known issues:

Issue	Description
TRANSFORM-2357	The Dataproc cluster Region property displays an invalid Google Cloud region, us-east2. Also, it does not display the us-east4 region. Workaround: To run a pipeline on a Dataproc cluster in the us-east4 region, set Region to Custom, then enter "us-east4".
TRANSFORM-2087	A pipeline fails to start when a Kafka origin is configured to read messages starting from a specified offset.
TRANSFORM-1943	A pipeline that joins two Snowflake origins can generate invalid SQL if the Snowflake tables are from different schemas and databases. Workaround: Configure each origin to run in Query read mode and use the fully qualified table name in the query as follows: <code><database.schema.table_name></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-495	The File origin processes files with mixed schemas.

Additional Resources

Our Documentation page provides access to all StreamSets product documentation: 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, visit our Community page: <https://streamsets.com/community/>.

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

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