

StreamSets Transformer

Cumulative 3.14.x Release Notes

June 18, 2020

We're happy to announce a new version of StreamSets Transformer. Version 3.14.x contains several new features, enhancements, and some important bug fixes in the following release:

- Version 3.14.0 - June 18, 2020

This document contains important information about the following topics:

- [New Features and Enhancements](#)
- [Upgrade Information](#)
- [Fixed Issues in 3.14.0](#)
- [Known Issues](#)
- [Additional Resources](#)

New Features and Enhancements

Transformer 3.14.x includes the following new features and enhancements:

Installation

This release includes the following installation enhancements:

- **[RPM installation and service start](#)** - You can install Transformer from an RPM package and then start Transformer as a service.
- **[Transformer downloads require registration](#)** - Transformer installation packages downloaded from the StreamSets website require that you register and activate the Transformer instance with StreamSets or that you register the Transformer instance with StreamSets Control Hub.

Your StreamSets account determines the maximum number of Spark executors allowed for each pipeline.

- **[Cloud service provider marketplace offers](#)** - When you install Transformer through a cloud service provider marketplace, you select an offer based on the maximum number of Spark executors allowed for each pipeline.

New Stages

This release includes the following new stages:

- **MapR Hive origin** - You can use this origin to read from a MapR Hive table.

- **MapR Hive destination** - You can use this destination to write to a MapR Hive table.

Stage Enhancements

This release includes the following stage enhancements:

- **Amazon Redshift destination** - You can now perform both insert and merge commands for loading data into an Amazon Redshift table.
- **Field Renamer processor** - The processor can change the case of field names to uppercase, lowercase, or title case.
- **Slowly Changing Dimension processor** - The processor now supports processing multiple changes to the same record in the same batch when processing Type 1 data or Type 2 data that uses an Active Flag or Version Increment tracking field.
- **SQL Expression processor** - Use the new Cast To Type property to specify the data type of the output field instead of using the inferred type.

Credential Store Enhancements

- [AWS Secrets Manager credential store](#) - You can call secrets defined in AWS Secrets Manager from pipeline and stage properties.
- [Group secrets in credential stores](#) - You can configure Transformer to validate a user's group against a comma-separated list of groups allowed to access each secret.

Additional Enhancements

- **Cluster pipeline validation** - You can validate a pipeline against the cluster configured to run the pipeline.
- **Default users and groups for cloud service provider installations** - Transformer installed through a cloud service provider marketplace includes only a default `admin` user account and no default groups.
- [Proxy user for Hadoop YARN clusters with Kerberos authentication](#) - Before pipelines can use a proxy user to launch the Spark application on a Hadoop YARN cluster that uses Kerberos authentication, you must install the required Kerberos client packages on the Transformer machine and then configure the environment variables used by the `K5start` program.

Upgrade Information

You can upgrade previous versions of Transformer to version 3.14.x. For complete instructions on upgrading, see the [Upgrade documentation](#).

Update `NullTypes` in Spark SQL Expression Processors

With this release, the Spark `NullType` data type is no longer supported.

If upgraded pipelines include Spark SQL Expression processors that use or result in a null value, update those processors so they do not use the `NullType` data type:

- If the expression might result in a null value, you can use the new `Cast To` property to select the data type to use.
- If the expression logic uses null values, add a `cast` call to ensure that the null values are handled properly.

Fixed Issues in 3.14.0

Transformer 3.14.0 includes the following fixed issues:

| Issue | Description |
|----------------|--|
| TRANSFORM-2002 | A pipeline run from a Control Hub job fails on an Amazon EMR cluster because of spaces added to the pipeline name. |
| TRANSFORM-1874 | The Amazon Redshift destination does not support writing to Date, Timestamp, or Decimal fields with 3.13.0. |

Known Issues

Transformer 3.14.x includes the following known issues:

| Issue | Description |
|----------------|---|
| 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. |
| 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 <code>Deployment pipeline</code> property set to <code>Cluster</code> . |

| | |
|---------------|---|
| | Workaround: Set the Deployment Mode pipeline property to Client. |
| 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/>.