StreamSets Transformer 3.11.0 Release Notes

October 10, 2019

We're happy to announce a new version of StreamSets Transformer. This version contains several new features and some bug fixes. We recommend that you try this latest release.

This document contains the following information:

- Upgrade Information
- New Features and Enhancements
- Known Issues
- Contact Information

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 3.11.0 from the StreamSets website.
4. Install and configure Transformer 3.11.0.
5. Start the 3.11.0 version of Transformer.
6. Register the 3.11.0 version of Transformer with Control Hub.
7. To have Transformer 3.11.0 run the same jobs as the earlier version of Transformer, assign the same label to the 3.11.0 version of Transformer.
8. Restart Transformer jobs.
9. If jobs run as expected, unregister the old version of Transformer.

New Features and Enhancements

This release includes the following new features and enhancements:

- **Field Existence Selector processor -** You can use this new processor to route records to a downstream stage when only the specified fields exist in those records. Records that do not meet the criteria pass to a separate output for error handling.

- **JDBC Lookup processor enhancement** - You can configure the processor to return the first matching row or multiple matching rows. You can determine if any matching rows exist or perform a count of matching rows. You can also specify how to order multiple matches.

- **Delta Lake destination enhancement -** You can now use the Delta Lake destination to perform updates, upserts, and deletes in addition to inserts.

- **Scala preprocessing script** - You can now define a pipeline preprocessing script that registers a user-defined function (UDF) before the pipeline starts, enabling the UDF to be called by a pipeline stage. For more information, see the sample script on the Advanced tab of the pipeline configuration properties.

- **Standalone Spark cluster** - You can now configure a pipeline to run on a standalone Spark cluster.

- **Avro advanced read properties** - When reading Avro data using Spark 2.4 or later, you can now specify the Avro schema to use in JSON format. You can also configure the origin to process all files, including those that do not use the `.avro` extension.

Known Issues

Please note the following known issue with this release:

Issue	Description
TRANSFORM-495	The File origin processes files with mixed schemas.

Contact Information

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

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.

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.