StreamSets Data Protector 1.4.0 Release Notes

October 27, 2018

We're happy to announce a new version of StreamSets Data Protector. This version contains several new features and enhancements.

This document contains important information about the following topics for this release:

- New Features
- Upgrade Information
- Known Issues

New Features

This version of Data Protector includes the following new features and enhancements:

Procedure enhancements - You can now specify the Data Collector to use when configuring procedures. You can use any registered Data Protector enabled Data Collector or the system Data Collector.

Classification rule enhancements - When viewing a classification rule, you can now also see all of the classifiers associated with the rule. This makes it easier to see how the classifiers work with the rule and makes it easier to create and edit classifiers.

Classifier enhancements - Classifiers can now classify based on field names in addition to field path and field values. Also, the default classifiers for new classification rules are now based on field names and field values, instead of field paths and field values. Existing classification rules are not affected by this change.

New protection methods - You can use the following protection methods in procedures to alter and protect sensitive data:

- **Standard Mask** - Masks email addresses, phone numbers, social security numbers, and zip codes classified by the StreamSets classification rules.

- **Custom Mask** - Masks any data with a user-defined mask.

- **Hash Data** - Enables hashing classified data with one of several algorithms, hashing Base64 encoded data, performing Base64 encoding on hashed data, and appending a salt value before hashing.

- **Expression Evaluator** - Enables protecting any classified data using a user-defined expression. You can use most of the StreamSets expression language in the expression, including the new category and data generation functions.

New protection functions - You can use the following functions to protect data in the new Expression Evaluator protection method:

- **Category functions** - Return parts of classified data, such as the area code of phone numbers or the last four digits of social security numbers. You can also use these functions to standardize the format of categories of data.

- **Data generation functions** - Generate fake data that can be used to replace classified sensitive data. There are two types of data generation functions:

  - **Faker functions** - Generate specific types of data, such as street addresses or first and last names.
  - **Xeger functions** - Generates fake data based on user-defined regular expressions.

  Data generation functions can be used to generate random data for every input value or to use the same generated data for the same input values.

StreamSets classification enhancements - The StreamSets classification rules now include classifiers for the telecom industry, such as ESN, IMEI, IMPI, IMSI, IMPU, and MSISDN. They also include a new relaxed URI, a new credit score classifier, and updates to the AMEX credit card and phone classifiers.

Also, the ETHNICITY classification rule has been renamed US_ETHNICITY. Update any procedures that use the ETHNICITY classification category to use US_ETHNICITY.

Upgrade Information

To upgrade to Data Protector 1.4.0, perform the following steps:

1. Stop all active jobs.
2. Uninstall the previous Data Protector stage library from registered Data Collectors.
3. Install the Data Protector 1.4.0 stage library on registered Data Collectors.
4. Start all jobs.

Fixed Issues

This release fixes the following known issue:

- Users with the Policy Manager role cannot share protection policies that they do not own. *(4794)*

Known Issues

Please note the following known issue with this release:

- Draft classification rules are incorrectly sent to newly registered Data Collectors.

  Workaround: Commit all classification rules before registering additional Data Collectors. *(4859)*

Contact Information

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

Check out our Documentation page for doc highlights, what's new, and tutorials: streamsets.com/docs

Or you can go straight to our latest documentation here:
https://streamsets.com/documentation/controlhub/latest/help/index.html

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.