

StreamSets Data Collector

Cumulative 3.18.x Release Notes

August 26, 2020

We're happy to announce a new version of StreamSets Data Collector. Version 3.18.x contains several new features, enhancements, and some important bug fixes in the following release:

- Version 3.18.1 - August 26, 2020
- Version 3.18.0 - August 21, 2020

This document contains important information about the following topics:

- [New Features and Enhancements](#)
- [Upgrade Information](#)
- [Fixed Issues in 3.18.1](#)
- [Fixed Issues in 3.18.0](#)
- [Known Issues](#)
- [Contact Information](#)

New Features and Enhancements

Data Collector 3.18.x includes the following new features and enhancements:

Pipeline Advanced Options

Pipelines and most pipeline stages include [advanced options](#) with default values that should work in most cases. By default, new and upgraded pipelines hide the advanced options. Advanced options can include individual properties or complete tabs.

As you start designing pipelines, StreamSets recommends configuring the basic properties and using the default values for the advanced options. Then as you continue working with pipelines, explore the advanced options to find ways to fine tune processing.

Sample Pipelines

Data Collector provides several [sample pipelines](#) that you can use to learn about Data Collector features or as a basis for building your own pipelines.

Enterprise Stage Libraries

In September 2020, StreamSets released updated Enterprise stage libraries for Databricks and Snowflake.

For a list of available Enterprise libraries, see [Enterprise Stage Libraries](#). For more information about the new features, fixed issues, and known issues in an Enterprise stage library, see the release notes

for the Enterprise stage library, available under [Enterprise Libraries](#) on the StreamSets Documentation page.

Enterprise stage libraries are free for use in both development and production.

Stage Enhancements

- **HTTP Client processor** - The Missing Values Behavior property determines the action that occurs when a request generates no response.
- **Amazon S3 stages** - “IAM roles” are now known as “instance profiles” in all Amazon S3 stages.

Additional Enhancements

- **Error records** - Error records include Control Hub job information in `jobName` and `jobId` record header attributes.

Upgrade Information

You can upgrade previous versions of Data Collector to version 3.18.x. For complete instructions on upgrading, see the [Upgrade documentation](#).

Upgrade Enterprise Stage Libraries

When you upgrade Data Collector, you must determine whether to upgrade your Enterprise stage libraries. See [Enterprise Stage Libraries](#) in the Data Collector documentation for a list of available Enterprise stage libraries and links to the supported versions and the stage documentation. To view the release notes for Enterprise stage libraries, see the [StreamSets Documentation page](#).

1. Uninstall the previous version of the Enterprise stage library.
 - a. In Package Manager, select the installed version.
 - b. Click the **Uninstall** icon.
 - c. Restart Data Collector.
2. Follow the stage documentation to install the new version of the Enterprise stage library and restart Data Collector.

Fixed Issues in 3.18.1

The following table lists some of the known issues that are fixed in Data Collector 3.18.1.

For the full list, click [here](#).

JIRA	Description
SDC-15503 SDC-15505	You cannot configure webhooks in Data Collector and Control Hub Pipeline Designer.

Fixed Issues in 3.18.0

The following table lists some of the known issues that are fixed in Data Collector 3.18.0.

For the full list, click [here](#).

JIRA	Description
SDC-15193	<p>Pipeline failures can occur when Data Collector 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, because the installation incorrectly points the SDC_RESOURCE environment variable to a directory that does not exist, <code>/opt/streamsets-datacollector/resources</code>.</p> <p>With this fix, the environment variable points to the following directory, as expected: <code>/var/streamsets-datacollector/resources</code>.</p>
SDC-15356	<p>When removing processed records from the Amazon SQS message queue, the Amazon SQS Consumer origin attempts to delete all messages twice, when only some messages have duplicates.</p>
SDC-15108	<p>The SFTP/FTP/FTPS Client origin generates an exception when it cannot reach the specified server.</p>

Known Issues

Please note the following known issues with Data Collector 3.18.x.

For a full list of known issues, click [here](#).

JIRA	Description
SDC-15408	<p>The Oracle CDC Client origin does not honor the Discard Old Uncommitted Transactions property. When selected, the origin still processes uncommitted transactions outside of the specified transaction window.</p> <p>Workaround: Clear the Buffer Changes Locally property. Note that clearing this option can increase the PGA memory usage on the database.</p>
SDC-14338	<p>When the Field Mapper processor encounters an error creating a new field, the error stops the pipeline.</p>

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/datacollector/latest/help>.

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.

Updated on 9/30/2020.