StreamSets Data Collector and Data Collector Edge 3.13.0 Release Notes

January 10, 2020

We're happy to announce new versions of StreamSets Data Collector and StreamSets Data Collector Edge. Version 3.13.0 contains several new features, enhancements, and some important bug fixes.

 This document contains important information about the following topics:

- New Features and Enhancements in Version 3.13.x
- Upgrading to Version 3.13.x
- Fixed Issues in Version 3.13.0
- Known Issues in Version 3.13.x

New Features and Enhancements in Version 3.13.x

Version 3.13.x improves performance when running large numbers of pipelines and changes the API for communication between Data Collector and Control Hub.

In addition, this version updates the process for generating and uploading support bundles. From Data Collector, you now download the support bundle to your machine. You can then upload the support bundle from your machine to your StreamSets Support ticket.

Upgrading to Version 3.13.x

You can upgrade previous versions of Data Collector to version 3.13.0. 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, the latest available versions, and links to the supported versions and the stage documentation. To view the release notes for Enterprise stage libraries, see the StreamSets Documentation page.

Note: Enterprise stage libraries are free for development purposes only. For information about purchasing an Enterprise stage library for use in production, contact StreamSets.

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 Version 3.13.0

The following table lists some of the known issues that are fixed with this release.

For the full list, click here.

JIRA	Description
SDC-13084	When running large numbers of pipelines, Data Collector does not send some heartbeats to Control Hub.
SDC-12845	The Websocket Client destination generates an error when sending data if an empty header or a TLS connection is configured.
SDC-12774	The Amazon S3 origin generates an error when reading a CSV file larger than 2GB.

Known Issues in Version 3.13.x

Please note the following known issues with this release.

For a full list of known issues, click here.

JIRA	Description
SDC-13180	The SFTP/FTP/FTPS Client origin does not create error records when reading JSON data and the number of characters exceeds the value configured in the Max Object Length (chars) property, possibly leading to data loss. Workaround: Configure a higher value in Max Object Length (chars) property.
SDC-13117	The Salesforce destination cannot use the Bulk API to write data with an external ID that references more than one data type.
SDC-12704	The Salesforce destination generates an error if you specify a blank string for a foreign key and try to upsert it with the SOAP API.
SDC-9888	When record fields contain special characters, the InfluxDB destination writes invalid measurements and truncated values to the InfluxDB database.
SDC-9853	Running a cluster streaming mode pipeline using Spark 2.1 that includes the HTTP Client processor encounters a ClassCastException error. Workaround: Copy the `jersey-server-2.25.1.jar` file from the `$SDC_DIST/container-lib` directory into the

	`$SDC_DIST/streamsets-libs/streamsets-datacollector-basic-lib/lib` directory. Then, restart Data Collector and re-submit the cluster application.
SDC-9514	Runtime parameters are not supported in all configuration properties in cluster batch execution mode, such as Max Batch Size.
SDC-8855	The MySQL Binary Log origin does not start reading from the offset specified in the Initial Offset property after a pipeline restart.
SDC-8514	The Data Parser processor sends a record to the next stage for processing even when the record encounters an error. Workaround: Use a Stream Selector processor after the Data Parser. Define a condition for the Stream Selector that checks if the fields in the record were correctly parsed. If not parsed correctly, send the record to a stream that handles the error.
SDC-8474	The Data Parser processor loses the original record when the record encounters an error.
SDC-8078	The HTTP Server origin does not release the ports that it uses after the pipeline stops. Releasing the ports requires restarting Data Collector.
SDC-7761	The Java keystore credential store implementation fails to work for a Data Collector installed through Cloudera Manager. The jks-cs command creates the Java keystore file in the Data Collector configuration directory defined for the parcel. However, for Data Collector to access the Java keystore file, the file must be outside of the parcel directory. The CyberArk and Vault credential store implementations do work with a Data Collector installed through Cloudera Manager.
SDC-6554	When converting Avro to Parquet on Impala, Decimal fields seem to be unreadable. Data Collector writes the Decimal data as variable-length byte arrays. And due to Impala issue IMPALA-2494, Impala cannot read the data.
SDC-5141	Due to a limitation in the Javascript engine, the Javascript Evaluator issues a null pointer exception when unable to compile a script.
SDC-4212	If you configure a UDP Source or UDP to Kafka origin to enable multithreading after you have already run the pipeline with the option disabled, the following validation error displays: `Multithreaded UDP server is not available on your platform`. Workaround: Restart Data Collector.
SDC-3944	The Hive Streaming destination using the MapR library cannot connect to a MapR cluster that uses Kerberos or username/password login authentication.

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.