StreamSets Data Collector
Cumulative 3.15.x Release Notes

April 16, 2020

We're happy to announce a new version of StreamSets Data Collector. Version 3.15.x contains several new features and enhancements in the following release:

- Version 3.15.0 - April 16, 2020

This document contains important information about the following topics:

New Features and Enhancements

Data Collector 3.15.x includes the following new features and enhancements:

Installation

This release includes the following installation enhancements:

- **Data Collector downloads require registration** - Data Collector installation packages downloaded from the StreamSets website now require that you register the Data Collector instance with StreamSets. Data Collector enforces the registration at different times, based on the following installation types:
    - A core version of Data Collector requires registration when you install an additional stage library.
    - A full version of Data Collector requires registration during your initial login.

  Registration involves submitting your contact information to receive a unique activation code by email and then entering the code in your Data Collector instance to activate the instance and use all functionality.

- **Download access for enterprise customers** - Customers with an enterprise license can now download all Data Collector installation packages from the StreamSets Support portal. Installation packages downloaded from the Support portal do not require that you register the Data Collector instance.

Enterprise Stage Libraries

Enterprise stage libraries are now free for use in both development and production. You can now use Enterprise stage libraries without purchasing an enterprise license.

In April 2020, StreamSets released updated Enterprise stage libraries for Protector, Oracle, and Snowflake. StreamSets also released a new Enterprise stage library for Databricks.

Data Collector Configuration

This release includes the following Data Collector configuration enhancement:

- **Java heap size for cloud service provider installations** - Data Collector installed through a cloud service provider marketplace now has a default Java heap size set to 50% of the available memory on the virtual machine.

Upgrade Information

You can upgrade previous versions of Data Collector to version 3.15.x. When you upgrade using a Data Collector installation package downloaded from the StreamSets website, you must register the Data Collector with StreamSets.

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.

Known Issues

Please note the following known issues with Data Collector 3.15.x.

For a full list of known issues, click here.

JIRA	Description
SDC-14253	Package Manager cannot install external libraries for custom stage libraries. Package Manager can only install external libraries for stage libraries that were installed using Package Manager. **Workaround:** Install external libraries manually for custom stage libraries. For example, to install a JDBC driver for a MemSQL Enterprise library that was installed as a custom stage library, copy the JDBC driver files to the following location: `<custom stage library dir>/streamsets-datacollector-memsql-lib/lib`

SDC-13918	In some cases, the PostgreSQL CDC Client origin incorrectly synchronizes with Postgres due to offset handling.
SDC-13679	Pressing the Tab key while configuring a Field Remover processor can generate a null pointer exception.
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-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.

Document revised on April 28, 2020