StreamSets Data Collector 1.3.0.1 Release Notes

April 20, 2016

We're happy to announce a new version of StreamSets Data Collector.

This release includes several important bug fixes. Please feel free to check out the [Documentation](#) for this release.

Upgrade Notes

Reconcile the Data Collector Configuration Files

When you upgrade Data Collector, you need to manually update the new Data Collector configuration file, `$SDC_CONF/sdc.properties`.

A new release can include new properties required for Data Collector to start or function properly. When you install a new version of Data Collector, compare the existing and new versions of the file, and update the new file as needed.

Fully-Qualified Protobuf Message Type

When upgrading Data Collector from any version earlier than 1.2.1.0, note the following requirement if you process protobuf data.

Data Collector now requires a fully-qualified message type name, as follows:

```
<package name>.<message type name>
```

If you were processing protobuf data with a previous version of Data Collector, perform the following steps:

1. Edit all pipelines that read or write protobuf data.
2. For the origin or destination, on the **Protobuf** tab, update the **Message Type** property to add the package name as follows: `<package name>.<message type name>`.

Fixed Issues

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

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

JIRA	Description
SDC-2776	The RabbitMQ Producer destination does not use the queue name as the routing key when you leave the routing key undefined and use a direct exchange binding type.

SDC-2773	In a Cloudera Manager installation, you cannot successfully restart Data Collector from the console.
SDC-2756	The Hadoop FS and Local FS destinations cannot write files larger than 2 GB at this time.
SDC-2716	Installations with Cloudera Manager cannot be configured for LDAP authentication at this time.

Known Issues

Please note the following known issues with this release.

For a full list of known issues, check out our JIRA.

JIRA	Description

SDC-2800	Installations with Cloudera Manager are missing default values for some Data Collector configuration properties. This can cause excessive polling of the REST API, resulting in a high CPU usage by the web browser.
	Workaround: Add the following Data Collector configuration properties through Cloudera Manager:
	`ui.enable.webSocket=true`
	`max.webSockets.concurrent.requests=15`
	`ui.refresh.interval.ms=2000`

SDC-2586	To process records larger than 1 MB, you must configure the DataFactoryBuilder.OverRunLimit property. However, this property is not configurable in the Data Collector configuration file, `$SDC_CONF/sdc.properties`. Workaround: Set the value of DataFactoryBuilder.OverRunLimit property in the SDC_JAVA_OPTS environment variable in the Data Collector environment file, `$SDC_DIST/libexec/sdc-env.sh` or `$SDC_DIST/libexec/sdcd-env.sh`. Set the property greater than the largest record you want to process. For example, to process records up to 2 MB, set the property to 2097152 as follows: `SDC_JAVA_OPTS="-DDataFactoryBuilder.OverRunLimit=2097152"`
SDC-2554	MapR FS does not support Kerberos authentication at this time.
SDC-2552	When an invalid topic is specified for a MapR Streams Consumer or a MapR Streams Producer, the following incorrect message displays:

StreamSets Data Collector 1.3.0.1 Release Notes

	`CONTAINER_0701 - Stage 'com_streamsets_pipeline_stage_origin_maprstreams_MapRStreamsDSource_1' initialization error: org.apache.kafka.common.config.ConfigException: No bootstrap urls given in bootstrap.servers` This message is misleading because MapR Streams does not support the `bootstrap.servers` option.
SDC-2374	A cluster mode pipeline can hang with a CONNECT_ERROR status. This can be a temporary connection problem that resolves, returning the pipeline to the RUNNING status. If the problem is not temporary, you might need to manually edit the pipeline state file to set the pipeline to STOPPED. Edit the file only after you confirm that the pipeline is no longer running on the cluster or that the cluster has been decommissioned. To manually change the pipeline state, edit the following file: `$SDC_DATA/runInfo/<cluster pipeline name>/<revision>/pipelineState.json` In the file, change CONNECT_ERROR to STOPPED and save the file.
SDC-2359	Due to a Kafka issue, a pipeline with a Kafka Consumer or Kafka Producer can hang during validation or display initialization errors when unable to connect to a Kafka 0.9.0.0 broker. The Data Collector log indicates that the broker is unavailable. For more information about the Kafka JIRA, see https://issues.apache.org/jira/browse/KAFKA-2880.
SDC-2226	When the Data Collector UI is configured for HTTPS, pipelines with the Hadoop FS origin fail to start.
SDC-2133	You cannot stop a pipeline in the middle of a long-running processor, such as a Jython Evaluator performing a complex script.
SDC-1731	When using the Kafka Consumer or Kafka Producer on HDP 2.3 with Kerberos enabled, set the Kafka broker configuration property `security.inter.broker.protocol` to PLAINTEXT. When enabling Kerberos, HDP 2.3 sets the `security.inter.broker.protocol` property to PLAINTEXTSASL, which is not supported. If the property is not set to PLAINTEXT, when the pipeline starts, validation errors indicate a problem connecting to Kafka.

StreamSets Data Collector 1.3.0.1 Release Notes

SDC-1567	You cannot use cluster mode pipelines to read from HDP 2.3 due to a HDP integration issue with Kafka and Spark Streaming.
	For more information about the HDP issue, see http://hortonworks.com/community/forums/topic/kafka-and-spark-streaming-nosuchmethoderror-kafka-consumer-simpleconsumer/.
SDC-891	At this time, writing to error records to file is not supported for cluster mode pipelines.
	Workaround: Write error records to Kafka or to an SDC RPC pipeline.
SDC-890	For cluster mode pipelines configured to stop on error or to stop upon reaching a memory limit, the Data Collector cannot stop all worker pipelines as expected.
	Workaround: To stop all pipelines, use the Stop icon in the Data Collector console.

Contact Information

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

To report an issue, ask for help, or find out about our next meetup, check out our Community page: http://streamsets.com/community/.

For general inquiries, email us at info@streamsets.com.