

StreamSets Data Collector 1.3.0.2 Release Notes

May 3, 2016

We're happy to announce a patch release for StreamSets Data Collector.

This patch release includes an important bug fix for the Kinesis Producer destination. If your pipelines use Kinesis Producer, please upgrade to this release and note the possible upgrade impact below. If you don't use Kinesis Producer, feel free to skip this version.

Feel free to check out the [Documentation](#) for this release.

Upgrade Notes

Kinesis Producer Round Robin Partition Strategy Change

With this release, the Kinesis Producer destination no longer includes the Round-Robin partition strategy. Destinations configured to use Round-Robin are updated to use the Random partition strategy.

If you don't want to use the Random strategy, configure the destination to use the Partition Expression strategy and define the expression to use.

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.

Fixed Issues

This release includes the following fixed issue:

JIRA	Description
SDC-2804	Kinesis Producer Round Robin and Random partition strategies do not work as expected.

StreamSets Data Collector 1.3.0.2 Release Notes

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	<p>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.</p> <p>Workaround: Add the following Data Collector configuration properties through Cloudera Manager:</p> <pre>ui.enable.webSocket=true max.webSockets.concurrent.requests=15 ui.refresh.interval.ms=2000</pre>
SDC-2586	<p>To process records larger than 1 MB, you must configure the <code>DataFactoryBuilder.OverRunLimit</code> property. However, this property is not configurable in the Data Collector configuration file, <code>\$SDC_CONF/sdc.properties</code>.</p> <p>Workaround: Set the value of <code>DataFactoryBuilder.OverRunLimit</code> property in the <code>SDC_JAVA_OPTS</code> environment variable in the Data Collector environment file, <code>\$SDC_DIST/libexec/sdc-env.sh</code> or <code>\$SDC_DIST/libexec/sdcd-env.sh</code>.</p> <p>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:</p> <pre>SDC_JAVA_OPTS="-DDataFactoryBuilder.OverRunLimit=2097152"</pre>
SDC-2554	<p>MapR FS does not support Kerberos authentication at this time.</p>
SDC-2552	<p>When an invalid topic is specified for a MapR Streams Consumer or a MapR Streams Producer, the following incorrect message displays:</p> <pre>CONTAINER_0701 - Stage 'com_streamsets_pipeline_stage_origin_maprstreams_MapRStreamsD Source_1' initialization error: org.apache.kafka.common.config.ConfigException: No bootstrap urls given in bootstrap.servers</pre> <p>This message is misleading because MapR Streams does not support the <code>bootstrap.servers</code> option.</p>

StreamSets Data Collector 1.3.0.2 Release Notes

SDC-2374	<p>A cluster mode pipeline can hang with a <code>CONNECT_ERROR</code> status. This can be a temporary connection problem that resolves, returning the pipeline to the <code>RUNNING</code> status.</p> <p>If the problem is not temporary, you might need to manually edit the pipeline state file to set the pipeline to <code>STOPPED</code>. Edit the file only after you confirm that the pipeline is no longer running on the cluster or that the cluster has been decommissioned.</p> <p>To manually change the pipeline state, edit the following file: <code>\$SDC_DATA/runInfo/<cluster pipeline name>/<revision>/pipelineState.json</code></p> <p>In the file, change <code>CONNECT_ERROR</code> to <code>STOPPED</code> and save the file.</p>
SDC-2359	<p>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.</p> <p>For more information about the Kafka JIRA, see https://issues.apache.org/jira/browse/KAFKA-2880.</p>
SDC-2226	<p>When the Data Collector UI is configured for HTTPS, pipelines with the Hadoop FS origin fail to start.</p>
SDC-2133	<p>You cannot stop a pipeline in the middle of a long-running processor, such as a Jython Evaluator performing a complex script.</p>
SDC-1731	<p>When using the Kafka Consumer or Kafka Producer on HDP 2.3 with Kerberos enabled, set the Kafka broker configuration property <code>security.inter.broker.protocol</code> to <code>PLAINTEXT</code>.</p> <p>When enabling Kerberos, HDP 2.3 sets the <code>security.inter.broker.protocol</code> property to <code>PLAINTEXTSASL</code>, which is not supported.</p> <p>If the property is not set to <code>PLAINTEXT</code>, when the pipeline starts, validation errors indicate a problem connecting to Kafka.</p>
SDC-1567	<p>You cannot use cluster mode pipelines to read from HDP 2.3 due to a HDP integration issue with Kafka and Spark Streaming.</p> <p>For more information about the HDP issue, see http://hortonworks.com/community/forums/topic/kafka-and-spark-streaming-nosuch-methoderror-kafka-consumer-simpleconsumer/.</p>

StreamSets Data Collector 1.3.0.2 Release Notes

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.