

StreamSets Data Collector 1.2.0.1 Release Notes

February 3, 2016

Fixed Issues

The following table lists some of known issues that are fixed with this release. For the full list, click [here](#).

JIRA	Description
SDC-2383	The Kafka Consumer origin duplicates the last message read when the pipeline is restarted.
SDC-2376	The Hadoop FS CDH 5.3 stage library does not include a required library.
SDC-2375	Due to refactoring, Hadoop FS origin does not work with Data Collector version 1.2.0.0.
SDC-2364	Origins cannot read concatenated gzip files.
SDC-2357	Data preview does not show the earliest available data for the Kafka Consumer origin when accessing Kafka 0.9.0.0.
SDC-2122	Due to the current Java keystore handling, you cannot use HTTPS with Cloudera Manager installations of the Data Collector.

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-2372	<p>Trying to write a message to Kafka that is longer than the Kafka maximum message size causes the pipeline to fail and restart endlessly.</p> <p>Workaround: Check the Kafka documentation for how to reconfigure Kafka brokers to allow larger messages or ensure that incoming records are within the configured limit.</p>

StreamSets Data Collector 1.2.0.1 Release Notes

JIRA	Description
SDC-2358 SDC-2359	<p>When unable to connect to a Kafka 0.9.0.0 broker, a pipeline with a Kafka Consumer or Kafka Producer can hang during validation or display initialization errors.</p> <p>The Data Collector log indicates that the broker is unavailable.</p>
SDC-2340	<p>When processing Avro files, leave the Compression Format set to None in the origin. Typically compression is handled within Avro, and the Avro file does not need to be compressed.</p>
SDC-2226	<p>When the Data Collector UI is configured for HTTPS, pipelines with the Hadoop FS origin fail to start.</p>
SDC-2190	<p>Pipelines with Hadoop FS origins that access kerberized clusters 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 PLAINTEXT.</p> <p>When enabling Kerberos, HDP 2.3 sets the <code>security.inter.broker.protocol</code> property to PLAINTEXTSASL, which is not supported.</p> <p>If the property is not set to PLAINTEXT, when the pipeline starts, validation errors indicate a problem connecting to Kafka.</p>
SDC-1626	<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><SDCinstalldir>/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-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</p>

StreamSets Data Collector 1.2.0.1 Release Notes

	http://hortonworks.com/community/forums/topic/kafka-and-spark-streaming-nosuch-methoderror-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.