

StreamSets Data Collector 1.2.2.0 Release Notes

March 10, 2016

New Features and Enhancements

We're happy to announce a new version of the StreamSets Data Collector. This version has a number of bug fixes and the following new features and enhancements:

- Support for MapR Streams (origin and destination)
- Support for MapR FS (destination)
- Support for InfluxDB (destination)
- Support for RabbitMQ (destination)
- Support for the Groovy language (processor)
- HTTP proxy support for the Omniture and HTTP Client origins
- MapR Streams as a pipeline error handling option
- New expression language primitives: `str:length()`, `str:matches()`, `str:concat()`

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

Upgrade Requirements

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

Fully-Qualified Protobuf Message Type

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-2529	When a Kafka Producer uses a Round-Robin partition strategy to write to Kafka 0.9.0.0, the Kafka Producer might fail to write messages to the Kafka broker.

StreamSets Data Collector 1.2.2.0 Release Notes

SDC-2508	The Solr destination throws an exception when writing to a SolrCloud instance.
SDC-2455	The Log Parser processor displays an incorrect message when the log data does not match the specified log format.
SDC-2445	<p>The Field Remover processor does not correctly handle fields that include spaces in the field path when you select the field from the UI.</p> <p>When a pipeline is valid for preview, the processor lists available fields and uses single quotation marks to escape spaces as follows:</p> <pre>/foo/'foo 2'/fieldname or /foo/foo2/'field name'</pre>
SDC-2414	<p>When an Elasticsearch destination is configured to write to a different version of Elasticsearch, the pipeline can hang indefinitely upon start.</p> <p>For example, if you configure the destination to use the Elasticsearch 2.2 stage library and you use it to write to Elasticsearch version 1.7, when you start the pipeline, it can hang with the message: "Starting Pipeline..."</p>

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-2554	MapR FS does not support Kerberos authentication at this time.
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 1.2.2.0 Release Notes

SDC-2438	<p>Trying to write a message to Kafka 0.8 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>
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: <SDCinstalldir>/data/runInfo/ <cluster pipeline name>/<revision>/pipelineState.json</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-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-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>

StreamSets Data Collector 1.2.2.0 Release Notes

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-nosuchmethoderror-kafka-consumer-simpleconsumer/.</p>
SDC-891	<p>At this time, writing to error records to file is not supported for cluster mode pipelines.</p> <p>Workaround: Write error records to Kafka or to an SDC RPC pipeline.</p>
SDC-890	<p>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.</p> <p>Workaround: To stop all pipelines, use the Stop icon in the Data Collector console.</p>

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.