StreamSets Data Collector 1.1.2 Release Notes

November 2015

Fixed Issues

The following issues are fixed with this release:

JIRA	Description
SDC-1950	Performance optimization of JSON Parser and generator.
SDC-1942	Avro data generator does not support union { null, record }.
SDC-1936	initd template does not work with Ubuntu.
SDC-1934	JDBC Consumer should set offset to initialOffset if passed in offset is null.
SDC-1922	Add maxBatchSize to JDBC Consumer.
SDC-1912	JMS Consumer does not generate unique source IDs.
SDC-1905	Field model selector does not support wildcards for list entries.
SDC-1898	Missing support for List-Map data type in Expression Evaluator processor.
SDC-1896	JDBC origin should output List-Map records.
SDC-1893	Scripting processors do not handle List-Map fields correctly.
SDC-1875	Missing support for the List-Map data type in various processors and data formats.
SDC-1857	Improve File Tail origin usability.
SDC-1774	The Data Collector command line interface fails to connect when the Data Collector uses digest authentication.

StreamSets Data Collector 1.1.2 Release Notes

Known Issues

Please note the following known issues with this release:

JIRA	Description
SDC-1975	Using the JDBC Consumer to read from Netezza is not supported at this time.
SDC-1926	When using the JDBC Producer destination with the Microsoft SQL Server change log format, the destination database table cannot use composite primary keys.
SDC-1794	You cannot use a custom delimited file type when writing delimited data to destinations at this time.
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.
SDC-1626	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: `<SDCinstalldir>/data/runInfo/ <cluster pipeline name>/<revision>/pipelineState.json` In the file, change CONNECT_ERROR to STOPPED and save the file.
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/.

StreamSets Data Collector 1.1.2 Release Notes

JIRA	Description
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 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.