

How to Setup Solr Sink for Flume

Summary of New Feature

This is a draft implementation of [FLUME-1687](#)

<https://issues.apache.org/jira/browse/FLUME-1687>

This sink is a great alternative to the ElasticSearchSink.

There are some flume users that have experience with Apache Solr but do not necessarily understand how to get ElasticSearch up and running.

Having a SolrSink as an alternative could be very helpful in creating a user interface for searching through event and log data collected with Flume using Apache Solr.

This sink essentially picks up events from the channel and uses the ConcurrentUpdateSolrServer client from SolrJ to send events in batches as SolrInputDocuments using multiple worker threads to Apache Solr.

It uses a thread-safe client for Apache Solr (ConcurrentUpdateSolrServer) which buffers all added documents and then transmits them using open HTTP connections.

Internally, ConcurrentUpdateSolrServer has worker threads that are used to drain the buffer

The number of worker threads and the threshold at which the documents are sent to the server and the url where the Solr index is located is configurable.

Please check it out and let me know your feedback.

How to Install

This Sink is designed to work with Flume 1.3.1

It has not yet been tested with other versions of Flume.

1. Copy the jar files below to your \$FLUME_HOME/lib folder.

<https://issues.apache.org/jira/secure/attachment/12579661/flume-new-feature-dependencies.zip>

<https://issues.apache.org/jira/secure/attachment/12579660/flume-new-features-1.3.1.jar>

2. Configure the Sink within your Agent similar to the example below and kick off flume.

```
#####
##### SINK CONFIGURATION #####
#####

# The name of the flume agent here is datastream
# The name of the sink here is solr1

# Declaring the Channels
datastream.channels = c1

# Declaring the Sinks
datastream.sinks = solr1

# FQCN (Fully-qualified class name) component type for the type
datastream.sinks.solr1.type = org.apache.flume.sink.solr.SolrSink

# Channel for this sink
datastream.sinks.solr1.channel = c1

datastream.sinks.solr1.serverUrl = http://localhost:8983/solr/flume

# Number of events to be written per transaction
datastream.sinks.solr1.batchSize = 500

# The number of background worker threads used by ConcurrentUpdateSolrServer to empty the queue
datastream.sinks.solr1.threadCount = 2

# Serializes the headers and body of an event into SolrInputDocuments that are sent to Apache Solr
datastream.sinks.solr1.serializer = org.apache.flume.sink.solr.SolrBasicEventSerializer

# A comma-delimited list of headers allowed.
# These must also be valid field names in the schema for this index
datastream.sinks.solr1.serializer.validHeaderFields = loglevel,timestamp,hostname

# The name of the field in the schema used for the event body. 'body' by default.
datastream.sinks.solr1.serializer.bodyFieldname = body
```

The jar file for the sources is available here

<https://issues.apache.org/jira/secure/attachment/12579662/flume-new-features-1.3.1-sources.jar>