Lab: Using Flume to Export Data

|  |  |
| --- | --- |
| **Objective:** | Install, configure and run Flume. |
| **Successful Outcome:** | A running Flume agent that reads data from a network connection and writes it to a folder in HDFS. |
| **Before You Begin:** | Go to terminal on sandbox. |

### Perform the following steps:

1. Install Flume
   1. On sandbox as root, confirm install of flume-ng:

# flume-ng

* + 1. If needed install Flume using the following command:

# yum –y install flume

* 1. Verify Flume is installed by viewing the usage of the flume-ng command:

# flume-ng

1. View the Flume Agent Configuration
   1. Change directories to the ~/materials directory.
   2. A Flume agent has been written for you in a file named flume-agent-1.conf. View the file:

# less flume-agent-1.conf

1. Start a Flume agent
   1. Start flume-ng using the following command (all on a single line):

flume-ng agent --conf conf --conf-file flume-agent-1.conf --name myagent -Dflume.root.logger=INFO,console &

*NOTE that in a full deployment we would typically include one more option: --conf=<conf-dir>. The <conf-dir> directory would include a shell script* flume-env.sh *and potentially a log4j properties file. In this example, we pass a Java option to force Flume to log to the console and we go without a custom environment script.*

* 1. View the output of the command. Make sure all sinks and sources started:

INFO sink.RollingFileSink: RollingFileSink sink1 started.

INFO instrumentation.MonitoredCounterGroup: Monitoried counter group for type: SOURCE, name: source1, registered successfully.

INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: source1 started

INFO source.AvroSource: Avro source source1 started.

1. Test the Flume Agent
   1. From a separate terminal (if possible), we can then telnet port 44444 and send Flume an event:

# telnet localhost 44444

Trying 127.0.0.1...

Connected to localhost.localdomain (127.0.0.1).

Escape character is '^]'.

Hello world! <ENTER>

OK

**RESULT**: You just ran a Flume agent that reads data from a network connection and streams it into a folder in HDFS.