Flume - AAT notes

Arturo Alatriste Trujillo

Contents

[Examples: 2](#_Toc492033370)

[From pooldir to hdfs 2](#_Toc492033371)

[Specifiying path based on event header ( timestamp ) and compression 3](#_Toc492033372)

[setting file type to DataStream and file extension 3](#_Toc492033373)

[Typical command line invocation 3](#_Toc492033374)

[Diagrams 4](#_Toc492033375)

[Fig 1. Common Flume Sources. 4](#_Toc492033376)

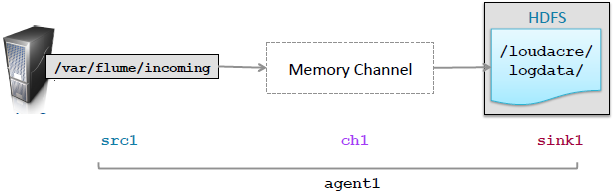
[Fig 2. Flume Agent. 5](#_Toc492033377)

[Fig 3. Flume Agent Server. 5](#_Toc492033378)

# Examples:

## From pooldir to hdfs

Example: Configure a Flume Agent to collect data from remote spool directories and save to HDFS



configuration

agent1.sources = src1

agent1.sinks = sink1

agent1.channels = ch1

agent1.channels.ch1.type = memory

Connect Source and channel

agent1.sources.src1.type = spooldir

agent1.sources.src1.spoolDir = /var/flume/incoming

agent1.sources.src1.channels = ch1

agent1.sinks.sink1.type = hdfs

agent1.sinks.sink1.hdfs.path = /loudacre/logdata

agent1.sinks.sink1.channel = ch1

Connect sink and channel

## Specifiying path based on event header ( timestamp ) and compression

(default is uncompressed)

agent1.sinks.sink1.type = hdfs

agent1.sinks.sink1.hdfs.path = /loudacre/logdata/%y-%m-%d

agent1.sinks.sink1.hdfs.codeC = snappy

agent1.sinks.sink1.channel = ch1

## setting file type to DataStream and file extension

( DataStream is raw data )

**agent1.sinks.sink1.type = hdfs**

**agent1.sinks.sink1.hdfs.path = /loudacre/logdata/%y-%m-%d**

**agent1.sinks.sink1.hdfs.fileType = DataStream**

**agent1.sinks.sink1.hdfs.fileSuffix = .txt**

**agent1.sinks.sink1.channel = ch1**

## Typical command line invocation

* The --name argument must match the agent’s name in the configuraDon file
* Setting root logger as shown will display log messages in the terminal

$ flume-ng agent \

--conf /etc/flume-ng/conf \

--conf-file /path/to/flume.conf \

--name agent1 \

-Dflume.root.logger=INFO,console

# Diagrams

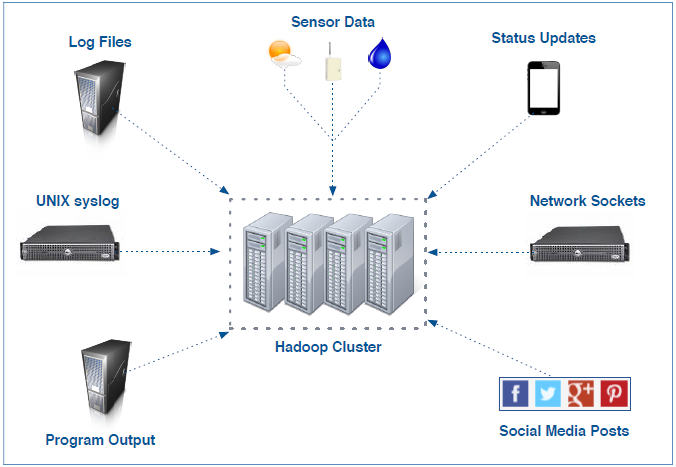
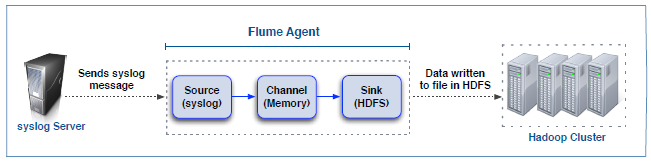
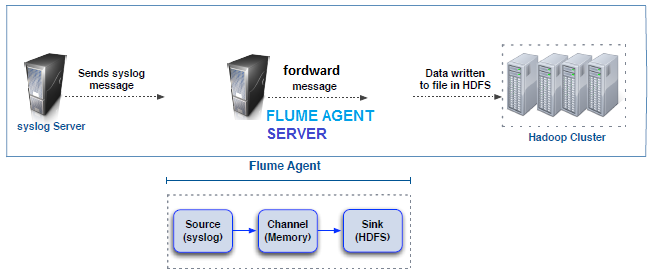


Fig 1. Common Flume Data Sources

## **Fig 1**. Common Flume Sources.



## **Fig 2**. Flume Agent.



## **Fig 3**. Flume Agent Server.

# References

## Apache Flume

<https://cwiki.apache.org/confluence/display/FLUME/Getting+Started>

## Flume documentation

<https://www.cloudera.com/documentation/enterprise/latest/topics/cdh_ig_flume_doc.html>

## Distributed Mode Diagram

<http://archive.cloudera.com/cdh/3/flume/UserGuide/#_running_in_distributed_mode>