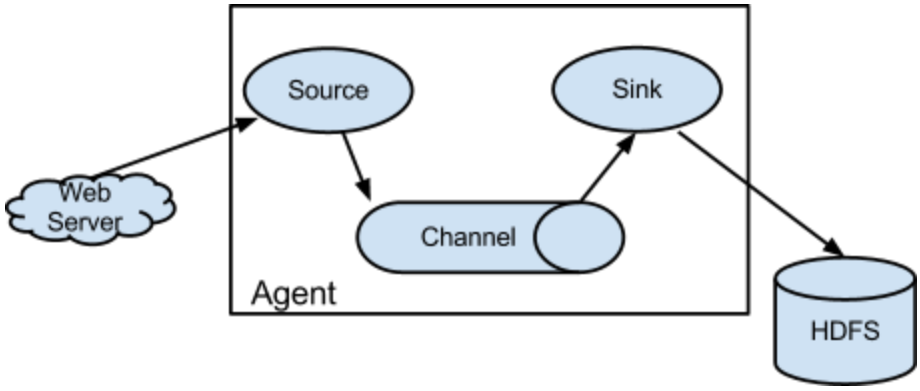
CloudxLab Flume Introduction

Flume is a distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of log data. It has a simple and flexible architecture based on streaming data flows. It is robust and fault tolerant with tunable reliability mechanisms and many failover and recovery mechanisms. It uses a simple extensible data model that allows for online analytic application.



Flume supports a large variety of sources Including:

• tail (like unix tail -f),
• syslog,
• log4j - allowing java applications to write logs to HDFS via flume

Demo Steps -

1) Login to Web console
2) Download sample flume.properties (in /data/flume/conf in HDFS) to your local directory in CloudxLab

   hadoop fs -copyToLocal /data/flume/conf .

3) Modify source in the flume.properties so that it can take it will take data from localhost and port 44444

   vi ~/conf/flume.properties
   # Modify these lines

   a1.sources.r1.type = netcat

a1.sources.r1.bind = localhost
a1.sources.r1.port = 44444

a1.sinks.hdfs-Cluster1-sink.hdfs.path =hdfs:///user/abhinav8248/flume/webdata

(Change abhinav8248 to your username)

4) Run the flume agent

flume-ng agent --conf conf --conf-file conf/flume.properties --name a1 Dflume.root.logger=INFO,console

5) If flume.properties file is correct, flume agent will start running in your terminal

6) Now we will generate some data so that flume agent can consume it. Login to web console in other tab and generate some data. Run this command

nc localhost 44444

Now type in some text and it will saved in HDFS.

Please find screencast for the same here https://youtu.be/SxpfSGGp_Ws