

Lab 7 – Flume Assignment

2. Flume Config file settings

```
agent1.sources = mySource
agent1.channels = channel1a channel1b
agent1.sinks = sink1a sink1b
agent1.sources.mySource.type = exec
agent1.sources.mySource.channels = channel1a channel1b
agent1.sinks.sink1a.channel = channel1a
agent1.sinks.sink1b.channel = channel1b

agent1.sources.mySource.type = spooldir
agent1.sources.mySource.spoolDir = /home/cloudera/flume/spooldir

agent1.sinks.sink1a.type = hdfs
agent1.sinks.sink1b.type = logger

agent1.channels.channel1a.type = file
agent1.channels.channel1b.type = memory

agent1.sinks.sink1a.hdfs.path = hdfs://localhost/user/cloudera/flumeOut/
agent1.sinks.sink1a.hdfs.filePrefix = apache
agent1.sinks.sink1a.hdfs.fileSuffix = .log
agent1.sinks.sink1a.hdfs.fileType = DataStream
```

Spool directory path:

/home/cloudera/flume/spooldir

HDFS Path:

/user/cloudera/flumeOut/

I have also attached the conf file into the flume folder.

Command to start the flume agent:

```
flume-ng agent -n agent1 -c /home/cloudera/flume/conf/ -f
/home/cloudera/flume/conf/flumeConf.conf
```

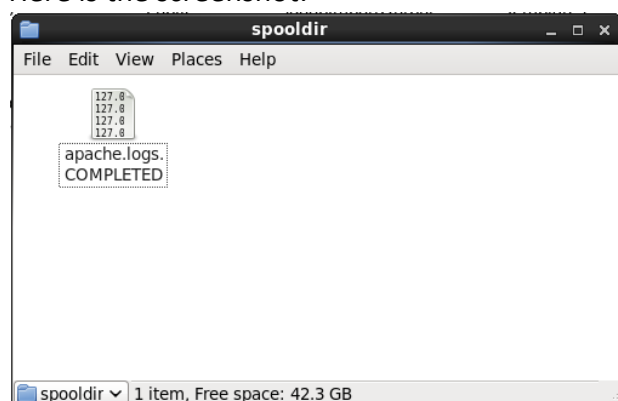
```

cloudera@quickstart:~$ flume-ng agent -n agent1 -c /home/cloudera/flume/conf/ -f /home/cloudera/flume/conf/flumeConf.conf
Info: Including Hadoop libraries found via (/usr/bin/hadoop) for HDFS access
Info: Including HBASE libraries found via (/usr/bin/hbase) for HBASE access
Info: Including Hive libraries found via () for Hive access
+ exec /usr/java/jdk1.7.0_67-cloudera/bin/java -Xmx20m -cp '/home/cloudera/flume/conf:/usr/lib/flume-ng/lib/*:/etc/hadoop/conf:/usr/lib/hadoop/lib/*:/usr/lib/hadoop/./*/:/usr/lib/hadoop-hdfs/lib/*:/usr/lib/hadoop-hdfs/./*/:/usr/lib/hadoop-yarn/lib/*:/usr/lib/hadoop-yarn/./*/:/usr/lib/hadoop-mapreduce/lib/*:/usr/lib/hadoop-mapreduce/./*/:/usr/lib/hbase/bin/./conf:/usr/java/jdk1.7.0_67-cloudera/lib/tools.jar:/usr/lib/hbase/bin/./usr/lib/hbase/bin/./lib/activation-1.1.jar:/usr/lib/hbase/bin/./lib/apacheds-i18n-2.0.0-M15.jar:/usr/lib/hbase/bin/./lib/apacheds-kerberos-codec-2.0.0-M15.jar:/usr/lib/hbase/bin/./lib/api-asn1-api-1.0.0-M20.jar:/usr/lib/hbase/bin/./lib/api-util-1.0.0-M20.jar:/usr/lib/hbase/bin/./lib/asm-3.2.jar:/usr/lib/hbase/bin/./lib/avro.jar:/usr/lib/hbase/bin/./lib/aws-java-sdk-bundle-1.11.134.jar:/usr/lib/hbase/bin/./lib/commons-beanutils-1.9.2.jar:/usr/lib/hbase/bin/./lib/commons-beanutils-core-1.8.0.jar:/usr/lib/hbase/bin/./lib/commons-cli-1.2.jar:/usr/lib/hbase/bin/./lib/commons-codec-1.9.jar:/usr/lib/hbase/bin/./lib/commons-collections-3.2.2.jar:/usr/lib/hbase/bin/./lib/commons-compress-1.4.1.jar:/usr/lib/hbase/bin/./lib/commons-configuration-1.6.jar:/usr/lib/hbase/bin/./lib/commons-daemon-1.0.13.jar:/usr/lib/hbase/bin/./lib/commons-digester-1.8.jar:/usr/lib/hbase/bin/./lib/commons-el-1.0.jar:/usr/lib/hbase/bin/./lib/commons-httpclient-3.1.jar:/usr/lib/hbase/bin/./lib/commons-io-2.4.jar:/usr/lib/hbase/bin/./lib/commons-lang-2.6.jar:/usr/lib/hbase/bin/./lib/commons-logging-1.2.jar:/usr/lib/hbase/bin/./lib/commons-math-2.1.jar:/usr/lib/hbase/bin/./lib/commons-math3-3.1.1.jar:/usr/lib/hbase/bin/./lib/commons-net-3.1.jar:/usr/lib/hbase/bin/./lib/commons-3.1.1.jar:/usr/lib/hbase/bin/./lib/curator-client-2.7.1.jar:/usr/lib/hbase/bin/./lib/curator-framework-2.7.1.jar:/usr/lib/hbase/bin/./lib/curator-recipes-2.7.1.jar:/usr/lib/hbase/bin/./lib/disruptor-3.3.0.jar:/usr/lib/hbase/bin/./lib/findbugs-annotation-1.3.9-1.jar:/usr/lib/hbase/bin/./lib/gson-2.2.4.jar:/usr/lib/hbase/bin/./lib/guava-12.0.1.jar:/usr/lib/hbase/bin/./lib/hamcrest-core-1.3.jar:/usr/lib/hbase/bin/./lib/hbase-annotations-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/./lib/hbase-annotations-1.2.0-cdh5.13.0-tests.jar:/usr/lib/hbase/bin/./lib/hbase-client-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/./lib/hbase-common-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/./lib/hbase-examples-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/./lib/hbase-external-blockcache-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/./lib/hbase-hadoop2-compat-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/./lib/hbase-hadoop2-compat-1.2.0-cdh5.13.0-tests.jar:/usr/lib/hbase/bin/./lib/hbase-hadoop-compat-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/./lib/hbase-hadoop-compat-1.2.0-cdh5.13.0-tests.jar:/usr/lib/hbase/bin/./lib/hbase-it-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/./lib/hbase-it-1.2.0-cdh5.13.0-tests.jar:/usr/lib/hbase/bin/./lib/hbase-prefix-tree-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/./lib/hbase-procedure-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/./lib/hbase-protocol-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/./lib/hbase-resource-bundle-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/./lib/hbase-rest-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/./lib/hbase-rsgroup-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/./lib/hbase-rsgroup-1.2.0-cdh5.13.0-tests.jar:/usr/lib/hbase/bin/./lib/hbase-server-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/./lib/hbase-server-1.2.0-cdh5.13.0-tests.jar:/usr/lib/hbase/bin/./lib/hbase-shell-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/./lib/hbase-spark-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/./lib/hbase-thrift-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/./lib/high-scale-lib-1.1.1.jar:/usr/lib/hbase/bin/./lib/hsqldb-1.8.0.10.jar:/usr/lib/hbase/bin/./lib/htrace-core-3.2.0-incubating.jar:/usr/lib/hbase/bin/./lib/htrace-core4-4.0.1-incubating.jar:/usr/lib/hbase/bin/./lib/htrace-core.jar:/usr/lib/hbase/

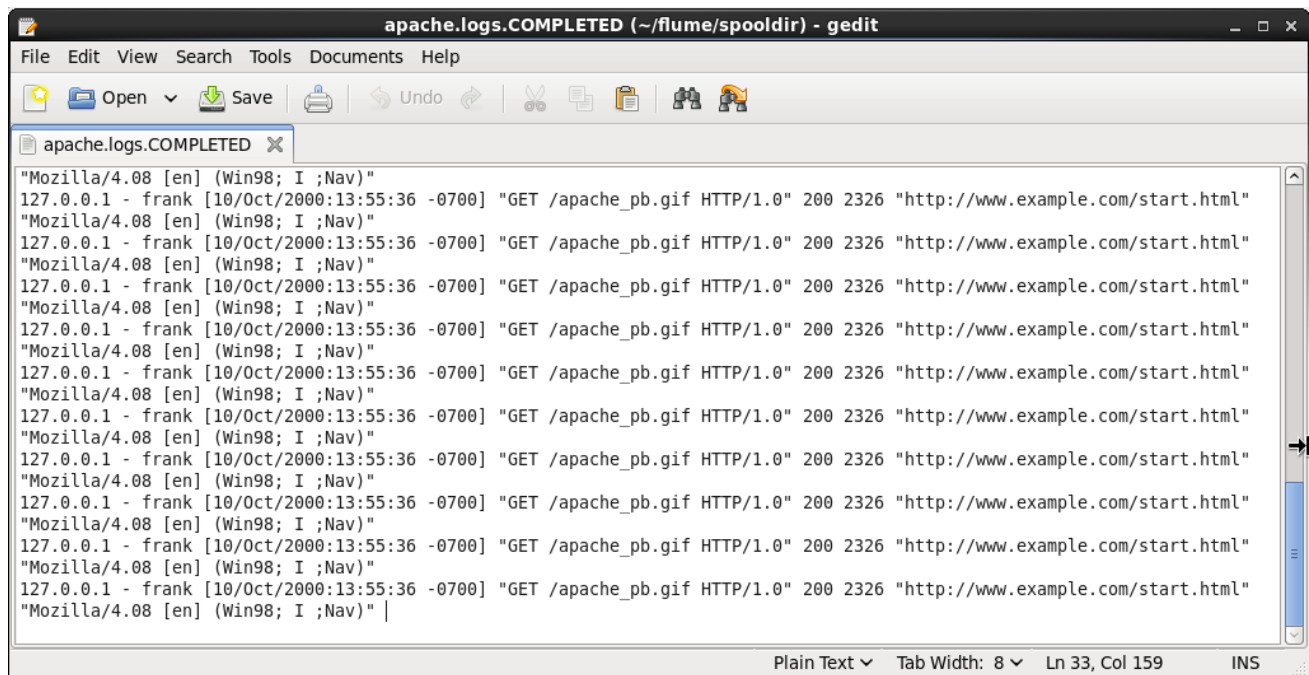
```

Now the agent will start and start observing on the **spooldir**. The file I put there is a log file and I have started adding new lines to the log file and agent starts saving the file in HDFS according to the setting.

Here is the screenshot:



Start adding in the apache.logs (it was renamed while reading) file.

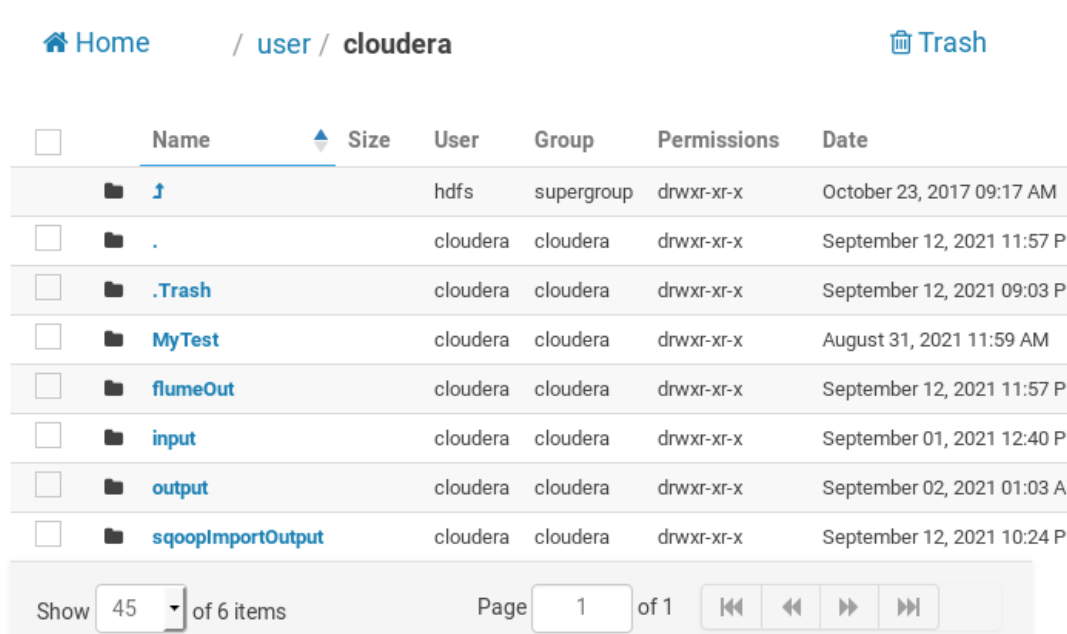


The screenshot shows a gedit editor window titled "apache.logs.COMPLETED (~/.flume/spooldir) - gedit". The window contains a list of log entries, each representing an HTTP GET request for a file named "apache_pb.gif". The log entries are formatted as follows:

```
"Mozilla/4.08 [en] (Win98; I ;Nav)"
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html"
"Mozilla/4.08 [en] (Win98; I ;Nav)"
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html"
"Mozilla/4.08 [en] (Win98; I ;Nav)"
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html"
"Mozilla/4.08 [en] (Win98; I ;Nav)"
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html"
"Mozilla/4.08 [en] (Win98; I ;Nav)"
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html"
"Mozilla/4.08 [en] (Win98; I ;Nav)"
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html"
"Mozilla/4.08 [en] (Win98; I ;Nav)"
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html"
"Mozilla/4.08 [en] (Win98; I ;Nav)"
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html"
"Mozilla/4.08 [en] (Win98; I ;Nav)"
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html"
"Mozilla/4.08 [en] (Win98; I ;Nav)"
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html"
|Mozilla/4.08 [en] (Win98; I ;Nav)" |
```

The status bar at the bottom of the window indicates "Plain Text", "Tab Width: 8", "Ln 33, Col 159", and "INS".

Flume Out Directory in HDFS – Screenshot:




The screenshot shows the Cloudera web interface for the HDFS file system. The breadcrumb navigation shows the path: Home / user / cloudera. The interface displays a table of files and directories in the HDFS file system.

<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	↑		hdfs	supergroup	drwxr-xr-x	October 23, 2017 09:17 AM
<input type="checkbox"/>	.		cloudera	cloudera	drwxr-xr-x	September 12, 2021 11:57 PM
<input type="checkbox"/>	.Trash		cloudera	cloudera	drwxr-xr-x	September 12, 2021 09:03 PM
<input type="checkbox"/>	MyTest		cloudera	cloudera	drwxr-xr-x	August 31, 2021 11:59 AM
<input type="checkbox"/>	flumeOut		cloudera	cloudera	drwxr-xr-x	September 12, 2021 11:57 PM
<input type="checkbox"/>	input		cloudera	cloudera	drwxr-xr-x	September 01, 2021 12:40 PM
<input type="checkbox"/>	output		cloudera	cloudera	drwxr-xr-x	September 02, 2021 01:03 AM
<input type="checkbox"/>	sqoopImportOutput		cloudera	cloudera	drwxr-xr-x	September 12, 2021 10:24 PM








At the bottom of the table, there is a pagination control showing "Show 45 of 6 items" and "Page 1 of 1".

File in HDFS – Screenshot

 **File Browser**

⚙️ Actions ▾ ✖ Move to trash ▾ 📁 Upload ▾

[🏠 Home](#) / [user](#) / [cloudera](#) / **flumeOut** 🗑️ Trash

<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	 ↑		cloudera	cloudera	drwxr-xr-x	September 13,
<input type="checkbox"/>	 .		cloudera	cloudera	drwxr-xr-x	September 13,
<input type="checkbox"/>	 apache.1631516445692.log	1.1 KB	cloudera	cloudera	-rw-r--r--	September 13,
<input type="checkbox"/>	 apache.1631516445693.log	1.1 KB	cloudera	cloudera	-rw-r--r--	September 13,
<input type="checkbox"/>	 apache.1631516445694.log	1.1 KB	cloudera	cloudera	-rw-r--r--	September 13,
<input type="checkbox"/>	 apache.1631516445695.log	1.1 KB	cloudera	cloudera	-rw-r--r--	September 13,
<input type="checkbox"/>	 apache.1631516445696.log.tmp	319 bytes	cloudera	cloudera	-rw-r--r--	September 13,