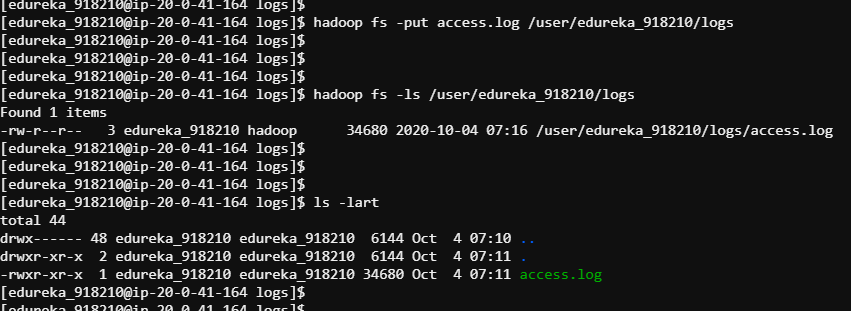
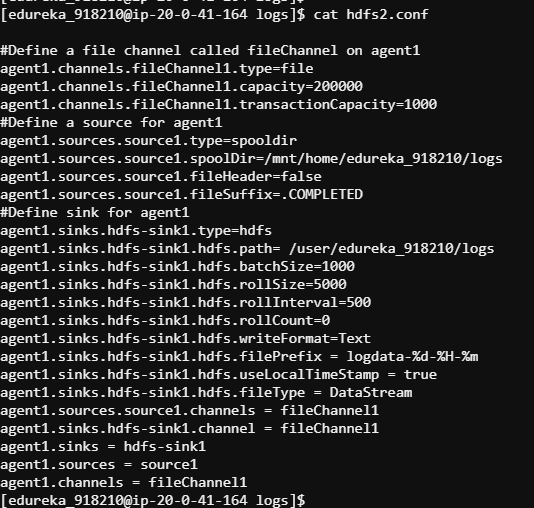
**WORKING WITH FLUME :**

**Schenario 1 : Extract Streaming Data with Apache Flume from Spool Directory to HDFS**

Putting logs in HDFS System



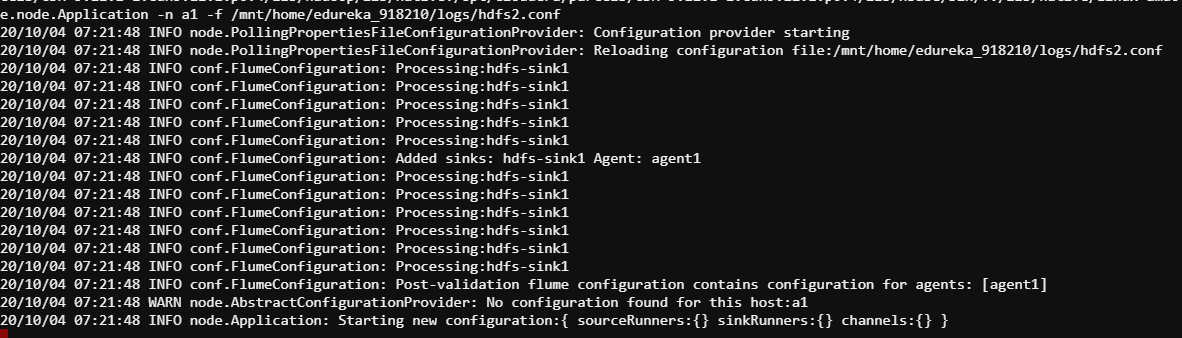
hdfs.conf



Run the Flume agent using below command to get the logs in HDFS sink.

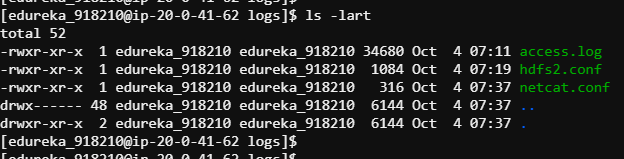
COMMANDS:

flume-ng agent -n a1 -f /mnt/home/edureka\_918210/logs/hdfs2.conf -Dflume.root.logger=INFO,console

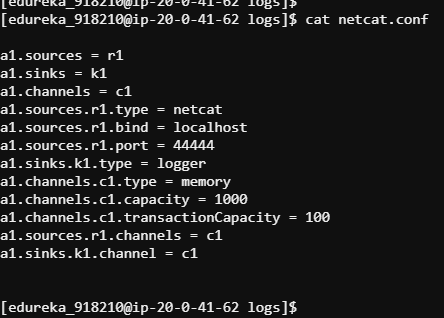


**Schenario 2 : Extract Streaming Data with Apache Flume from NetCat to Logger**

In this demo, we will learn how to extract the streaming data with Apache Flume by using **NetCat** as the data source and **Logger** as the sink.



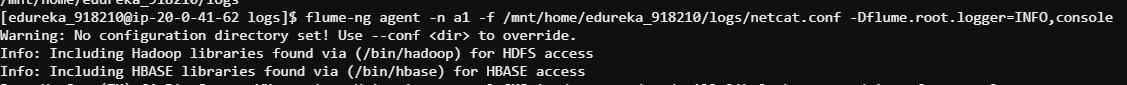
netcat.conf

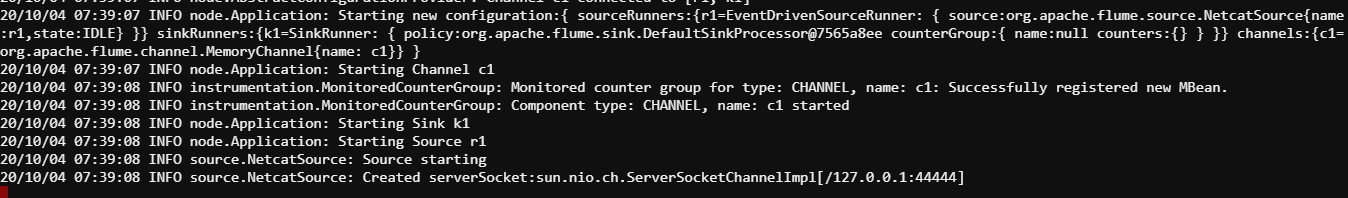


Run the Flume agent using the following command:

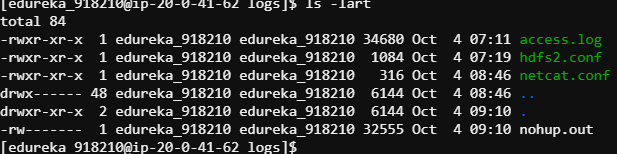
**Command:**

flume-ng agent -n a1 -f /mnt/home/edureka\_918210/logs/netcat.conf -Dflume.root.logger=INFO,console



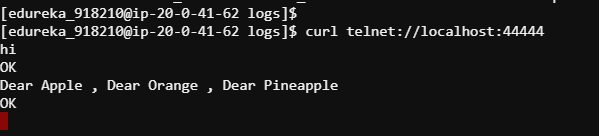


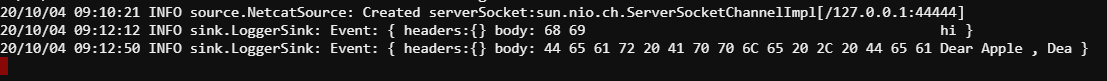
Running the process in the background



**Command:**

curl telnet://localhost:44444



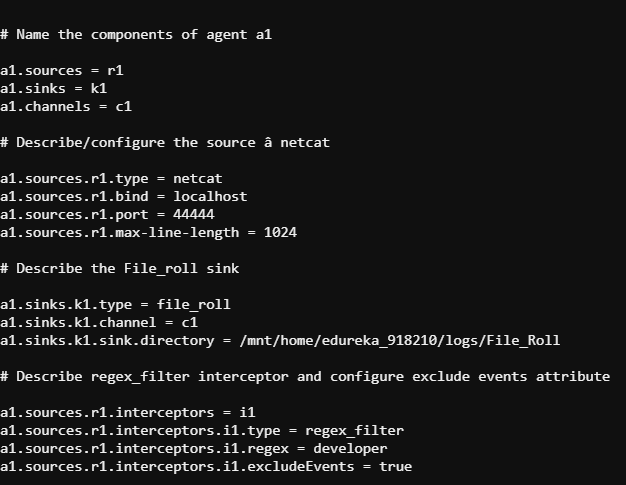


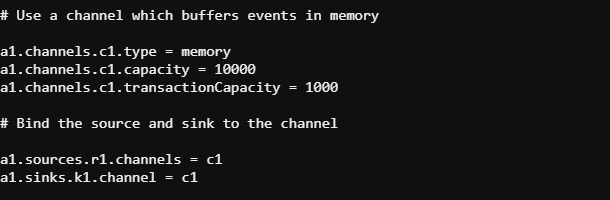
edureka+ 11402 26809 0 09:10 pts/1 00:00:01 /usr/java/jdk1.8.0\_144-cloudera/bin/java -Xmx20m -Dflume.root.logger=INFO,console -cp /opt/cloudera/parcels/CDH-5.1

**Schenario 3: Extract Streaming Data with Apache Flume from NetCat to File Roll Data**

mkdir File\_Roll

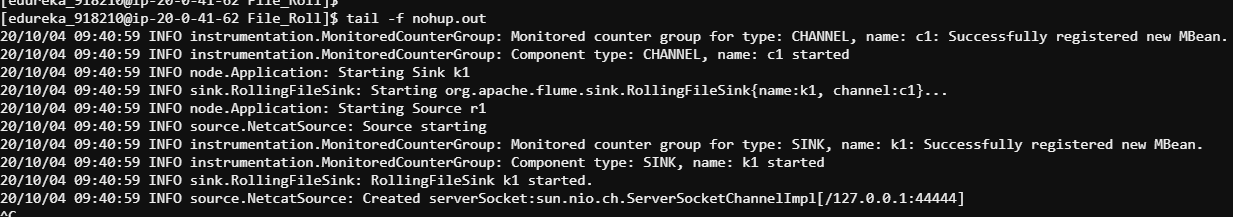
vi file\_roll.conf



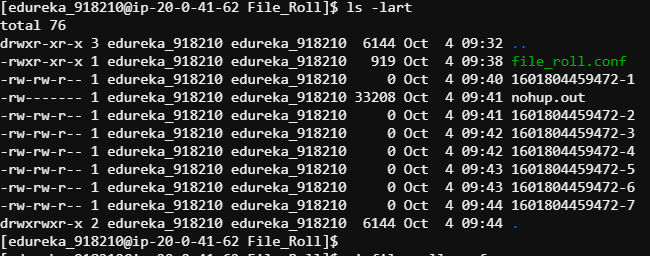


**FLUME AGENT**

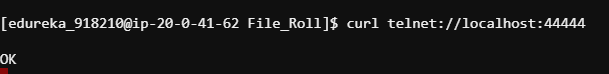
flume-ng agent -n a1 -f /mnt/home/edureka\_918210/logs/File\_Roll/file\_roll.conf -Dflume.root.logger=INFO,console

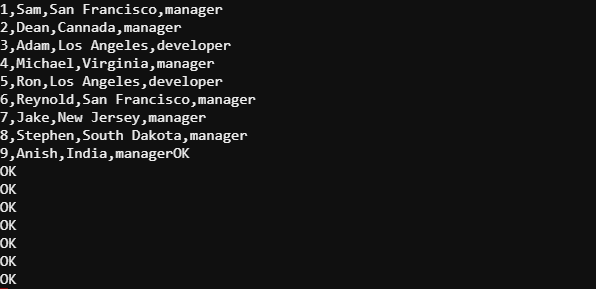


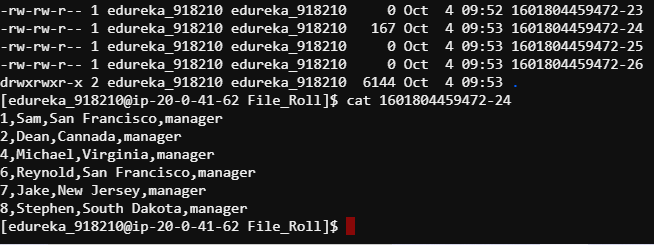
**File System :**



curl <telnet://localhost:44444>



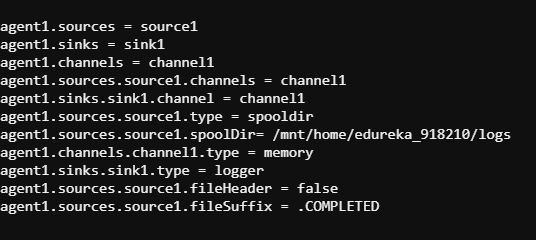




**Schenario 4: Spool Directory to Logger Data Streaming Using Apache Flume**

Create a Flume configuration file using the command vi flume.conf

Define source, channel and sink.

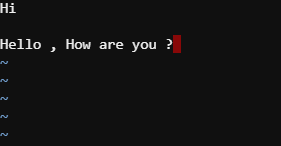


flume-ng agent -n agent1 -f /mnt/home/edureka\_918210/logs/flume.conf -Dflume.root.logger=INFO,console



You will be able to see the texts that have been written in the text file present in the defined directory.

Try placing some other text file through a different console. You will be able to see the text.

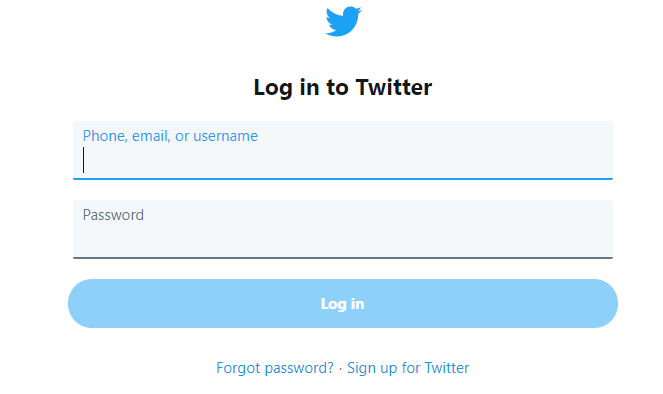


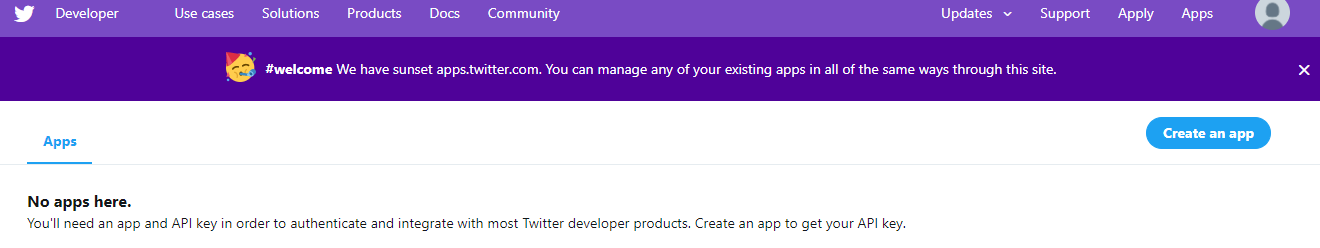
**Schenario 5: Extract Streaming Data with Apache Flume from Twitter to HDFS**

Step 1: Go to “apps.twitter.com.”

Step 2: Create a developer account.

Step 3: Log in to your Twitter account and click on “Create an app” button.





TwitterAgent.sources = Twitter

TwitterAgent.channels = MemChannel

TwitterAgent.sinks = HDFS

TwitterAgent.sources.Twitter.type = org.apache.flume.source.twitter.TwitterSource

TwitterAgent.sources.Twitter.channels = MemChannel

TwitterAgent.sources.Twitter.consumerKey = zz3c3kMSCM0wFzpItMmeDHOom

TwitterAgent.sources.Twitter.consumerSecret =

g31gbwNambj3B2VEApWLoYJCsb5fWTpOz0FEwBAD6l3gAavlQH

SybN56o9OlPfl23OpJL6C8ahMSSdGuOmLOlUL8Lh

TwitterAgent.sources.Twitter.accessToken = 3027110343-X73G5HhGPhZHCtRWeqHQhEzVqQp0nTwAf4iybLb

TwitterAgent.sources.Twitter.accessTokenSecret = FKLRDtnsYbAhx1KXyjmZkwdosAdzlr1sBwpLcCpyAOOlh

TwitterAgent.sources.Twitter.keywords = spark, scientist, hadoop, big data,analytics, bigdata, cloudera, data science, data scientist, business intelligence,mapreduce, data warehouse, data warehousing, mahout, hbase, nosql, newsql,businessintelligence, cloudcomputing

TwitterAgent.sinks.HDFS.channel = MemChannel

TwitterAgent.sinks.HDFS.type = hdfs

TwitterAgent.sinks.HDFS.hdfs.path = /user/edureka\_918210/logs

TwitterAgent.sinks.HDFS.hdfs.fileType = DataStream

TwitterAgent.sinks.HDFS.hdfs.writeFormat = Text

TwitterAgent.sinks.HDFS.hdfs.batchSize = 1000

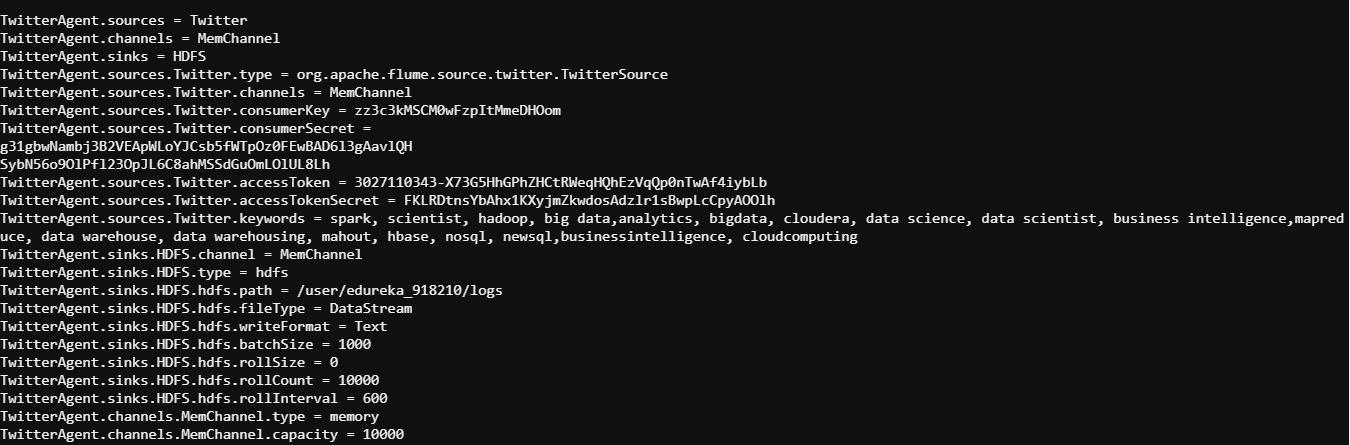
TwitterAgent.sinks.HDFS.hdfs.rollSize = 0

TwitterAgent.sinks.HDFS.hdfs.rollCount = 10000

TwitterAgent.sinks.HDFS.hdfs.rollInterval = 600

TwitterAgent.channels.MemChannel.type = memory

TwitterAgent.channels.MemChannel.capacity = 10000





flume-ng agent -n TwitterAgent -c conf -f /mnt/home/edureka\_918210/logs/flume\_twitter.conf -Dflume.root.logger=INFO,console

