

Assignment 11.1

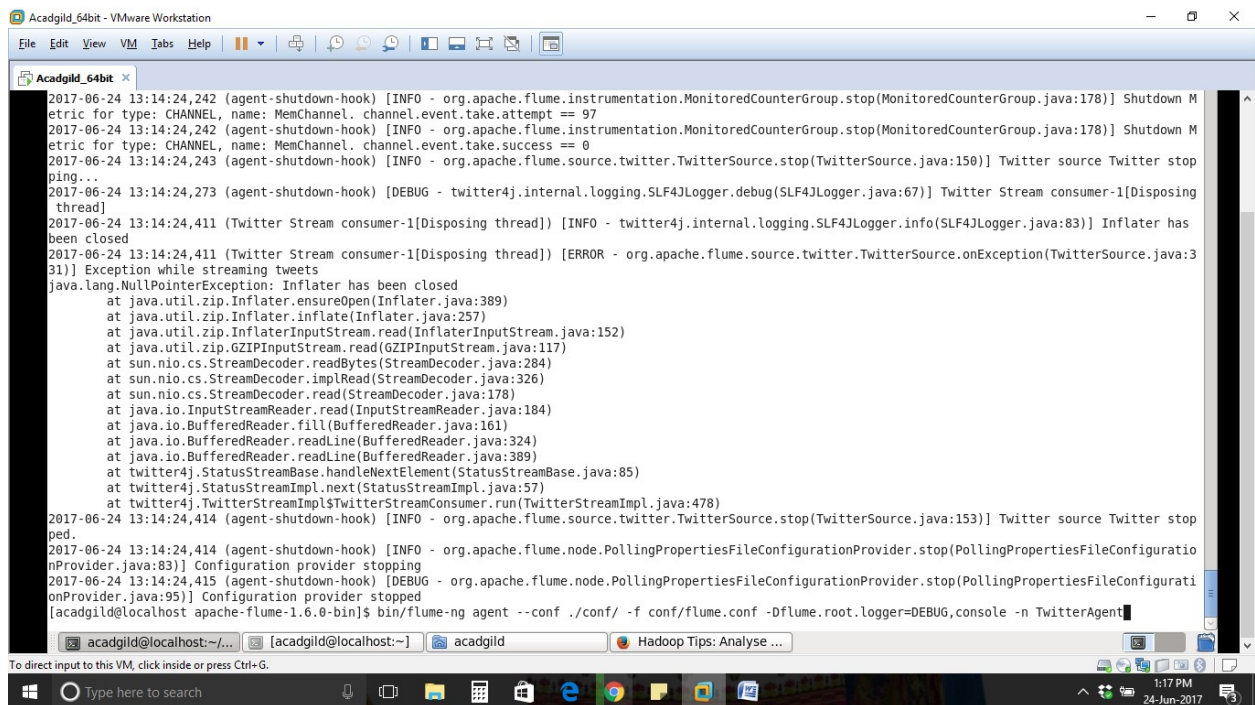
Flume

Problem Statement

Create a flume agent that streams data from Twitter and stores in the HDFS.

```
bin/flume-ng agent --conf ./conf/ -f conf/flume.conf -Dflume.root.logger=DEBUG,console -n TwitterAgent
```

Starting the Twitter Agent



The screenshot shows a VMware Workstation window titled "Acadgild_64bit - VMware Workstation". Inside the VM, a terminal window displays the following logs:

```
2017-06-24 13:14:24,242 (agent-shutdown-hook) [INFO - org.apache.flume.instrumentation.MonitoredCounterGroup.stop(MonitoredCounterGroup.java:178)] Shutdown Metric for type: CHANNEL, name: MemChannel, channel.event.take.attempt == 97
2017-06-24 13:14:24,242 (agent-shutdown-hook) [INFO - org.apache.flume.instrumentation.MonitoredCounterGroup.stop(MonitoredCounterGroup.java:178)] Shutdown Metric for type: CHANNEL, name: MemChannel, channel.event.take.success == 0
2017-06-24 13:14:24,243 (agent-shutdown-hook) [INFO - org.apache.flume.source.twitter.TwitterSource.stop(TwitterSource.java:150)] Twitter source Twitter stop ping...
2017-06-24 13:14:24,273 (agent-shutdown-hook) [DEBUG - twitter4j.internal.logging.SLF4JLogger.debug(SLF4JLogger.java:67)] Twitter Stream consumer-1[Disposing thread]
2017-06-24 13:14:24,411 (Twitter Stream consumer-1[Disposing thread]) [INFO - twitter4j.internal.logging.SLF4JLogger.info(SLF4JLogger.java:83)] Inflater has been closed
2017-06-24 13:14:24,411 (Twitter Stream consumer-1[Disposing thread]) [ERROR - org.apache.flume.source.twitter.TwitterSource.onException(TwitterSource.java:331)] Exception while streaming tweets
java.lang.NullPointerException: Inflater has been closed
    at java.util.zip.Inflater.ensureOpen(Inflater.java:389)
    at java.util.zip.Inflater.inflate(Inflater.java:257)
    at java.util.zip.InflaterInputStream.read(InflaterInputStream.java:152)
    at java.util.zip.GZIPInputStream.read(GZIPInputStream.java:117)
    at sun.nio.cs.StreamDecoder.readBytes(StreamDecoder.java:284)
    at sun.nio.cs.StreamDecoder.implRead(StreamDecoder.java:326)
    at sun.nio.cs.StreamDecoder.read(StreamDecoder.java:178)
    at java.io.InputStreamReader.read(InputStreamReader.java:184)
    at java.io.BufferedReader.fill(BufferedReader.java:161)
    at java.io.BufferedReader.readLine(BufferedReader.java:324)
    at java.io.BufferedReader.readLine(BufferedReader.java:389)
    at twitter4j.StatusStreamBase.handleNextElement(StatusStreamBase.java:85)
    at twitter4j.StatusStreamImpl.next(StatusStreamImpl.java:57)
    at twitter4j.TwitterStreamImpl$TwitterStreamConsumer.run(TwitterStreamImpl.java:478)
2017-06-24 13:14:24,414 (agent-shutdown-hook) [INFO - org.apache.flume.source.twitter.TwitterSource.stop(TwitterSource.java:153)] Twitter source Twitter stopped.
2017-06-24 13:14:24,414 (agent-shutdown-hook) [INFO - org.apache.flume.node.PollingPropertiesFileConfigurationProvider.stop(PollingPropertiesFileConfigurationProvider.java:83)] Configuration provider stopping
2017-06-24 13:14:24,415 (agent-shutdown-hook) [DEBUG - org.apache.flume.node.PollingPropertiesFileConfigurationProvider.stop(PollingPropertiesFileConfigurationProvider.java:95)] Configuration provider stopped

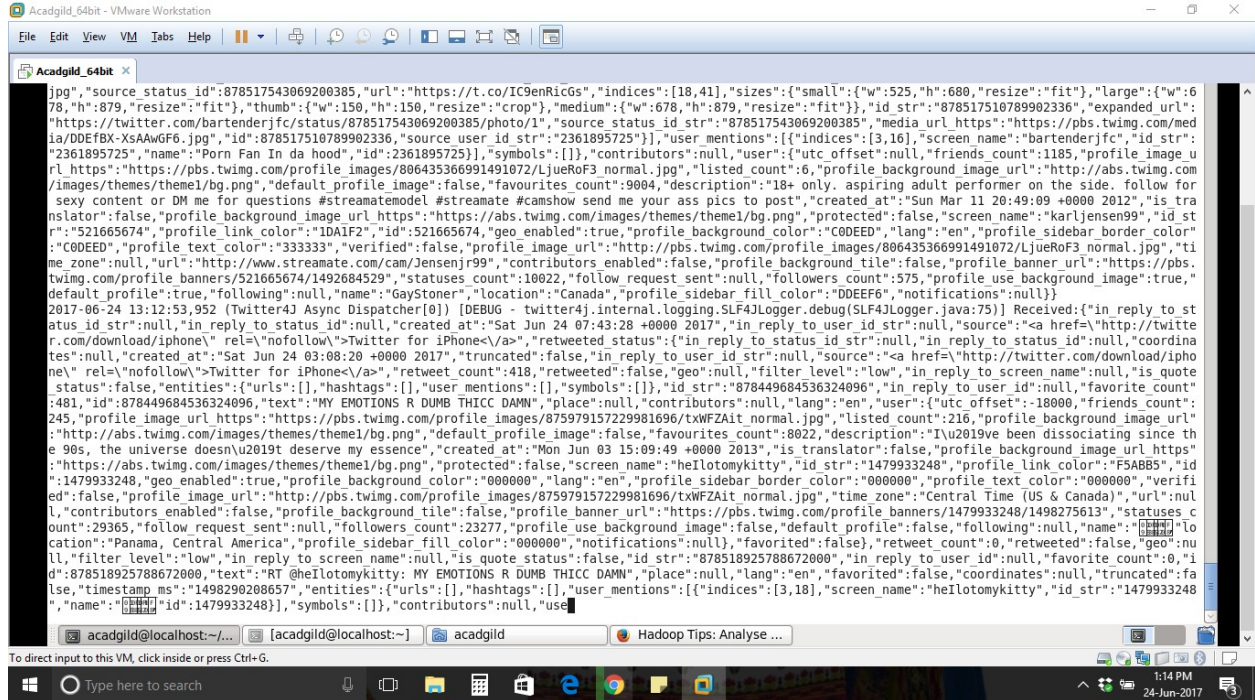
[acadgild@localhost apache-flume-1.6.0-bin]$ bin/flume-ng agent --conf ./conf/ -f conf/flume.conf -Dflume.root.logger=DEBUG,console -n TwitterAgent
```

Below the terminal, a command prompt shows the user has entered the command:

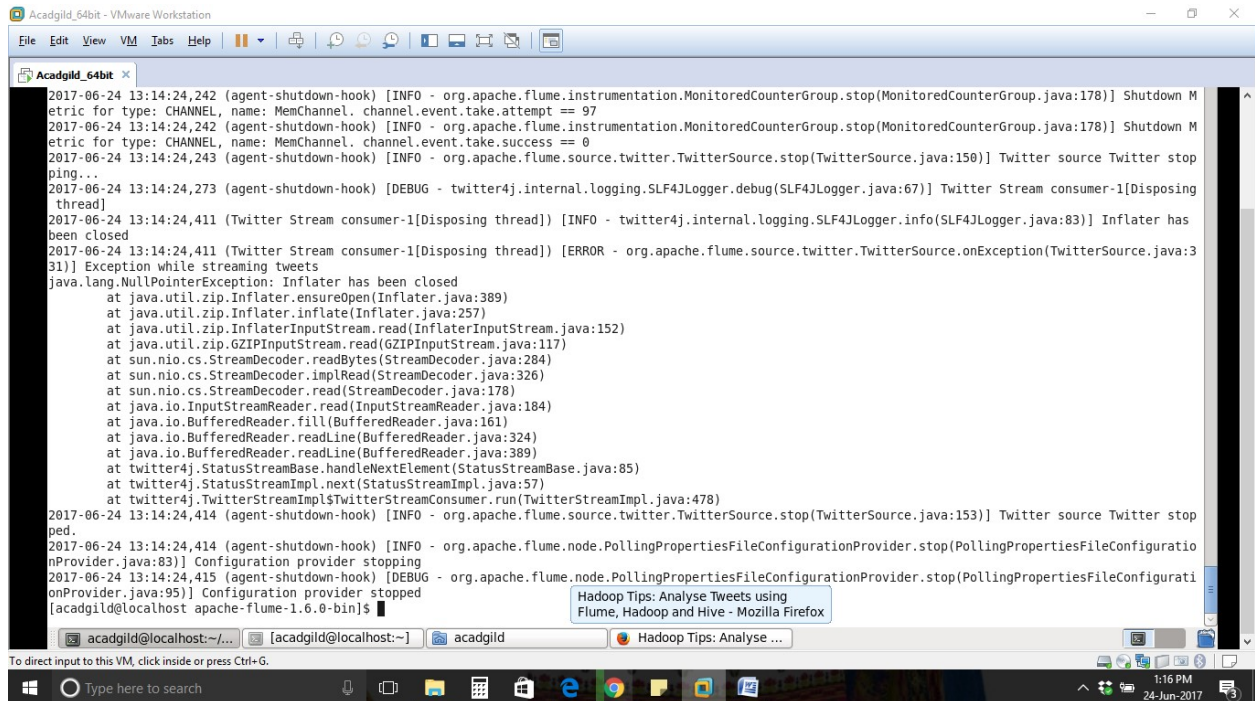
```
[acadgild@localhost:~/...]$ bin/flume-ng agent --conf ./conf/ -f conf/flume.conf -Dflume.root.logger=DEBUG,console -n TwitterAgent
```

The bottom of the image shows the Windows taskbar with the time 1:17 PM and date 24-Jun-2017.

Data being streamed into HDFS



Stop streaming by pressing Ctrl +C



Listing contents of /user/flume/tweets

```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help

Acadgild_64bit x
localhost: starting namenode, logging to /usr/local/hadoop-2.6.0/logs/hadoop-acadgild-namenode-localhost.localdomain.out
localhost: starting datanode, logging to /usr/local/hadoop-2.6.0/logs/hadoop-acadgild-datanode-localhost.localdomain.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop-2.6.0/logs/hadoop-acadgild-secondarynamenode-localhost.localdomain.out
17/06/24 12:37:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop-2.6.0/logs/yarn-acadgild-resourcemanager-localhost.localdomain.out
localhost: starting nodemanager, logging to /usr/local/hadoop-2.6.0/logs/yarn-acadgild-nodemanager-localhost.localdomain.out
[acadgild@localhost ~]$ SHADOOP_HOME/sbin/mr-jobhistory-daemon.sh --config $HADOOP_HOME/etc/hadoop start historyserver
starting historyserver, logging to /usr/local/hadoop-2.6.0/logs/mapred-acadgild-historyserver-localhost.localdomain.out
[acadgild@localhost ~]$ jps
3408 ResourceManager
3857 JobHistoryServer
3923 Jps
3511 NodeManager
2970 NameNode
3259 SecondaryNameNode
3070 DataNode
[acadgild@localhost ~]$ hdfs dfs -ls /user/flume/tweets
17/06/24 13:16:18 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ls: /user/flume/tweets: No such file or directory
[acadgild@localhost ~]$ hdfs dfs -ls /user/flume/tweets
17/06/24 13:16:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 133938 2017-06-24 13:14 /user/flume/tweets/FlumeData.1498290112889.tmp
-rw-r--r-- 1 acadgild supergroup 0 2017-06-24 13:13 /user/flume/tweets/FlumeData.1498290112890
[acadgild@localhost ~]$
```

Contents of file in which the Twitter Data is collected

The image shows a web browser window displaying a Twitter profile for 'Acadgild_64bit'. The browser's address bar shows the URL 'https://twitter.com/Acadgild_64bit'. The profile page includes a bio, location (New York, NY), and a list of tweets. The tweets contain various links, including a salary network, a shopping adviser, and a tweet about a bipolar tweet. The browser's address bar shows the URL 'https://twitter.com/Acadgild_64bit'. The page is viewed in a window titled 'Acadgild_64bit'.

