SESSION 11 ASSIGNMENT 3

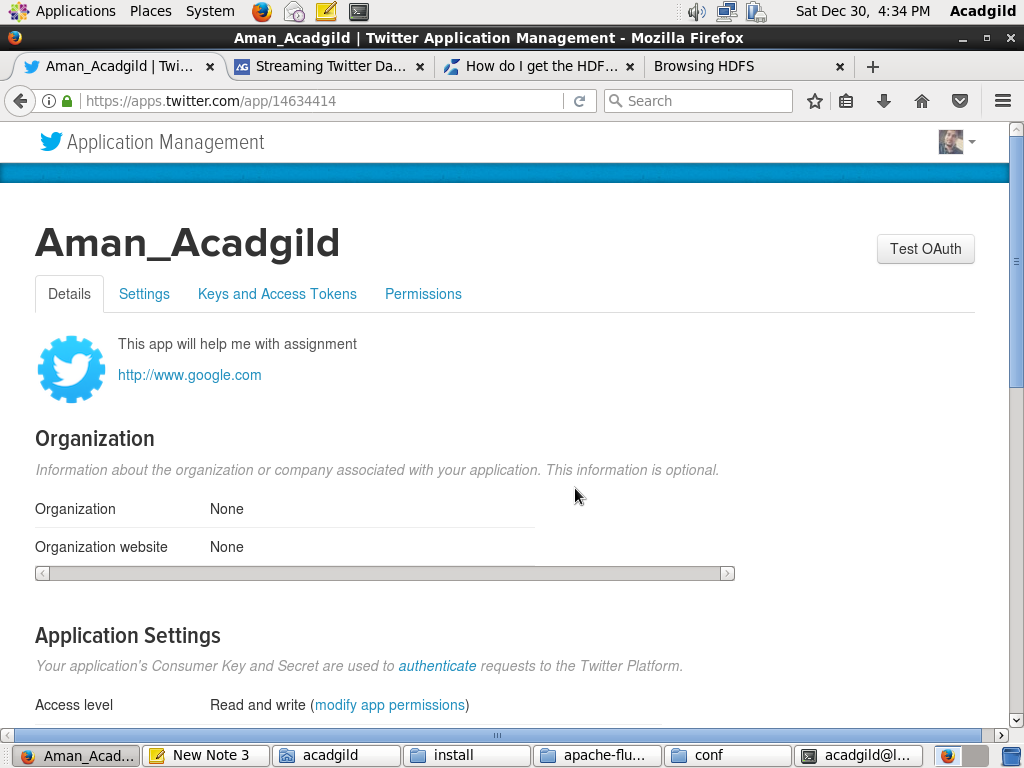
Create a flume agent that streams data from Twitter and stores in the HDFS.

Creation of App:

For streaming data from twitter we will have to create a twitter app from the following link –

<https://apps.twitter.com/app>.

After creation of the app we need to copy different user credentials which are required for flume.conf file.



Contents of flume.conf

TwitterAgent.sources = Twitter

TwitterAgent.channels = MemChannel

TwitterAgent.sinks = HDFS

# Describing/Configuring the source

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

TwitterAgent.sources.Twitter.consumerKey=<copy consumer key from app>

TwitterAgent.sources.Twitter.consumerSecret=<copy code from app>

TwitterAgent.sources.Twitter.accessToken=<copy code from app>

TwitterAgent.sources.Twitter.accessTokenSecret=<copy code from app>

TwitterAgent.sources.Twitter.keywords=India, big data

# Describing/Configuring the sink

TwitterAgent.sinks.HDFS.channel=MemChannel

TwitterAgent.sinks.HDFS.type=hdfs

TwitterAgent.sinks.HDFS.hdfs.path=hdfs://localhost:8020//user/twitter

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

TwitterAgent.channels.MemChannel.transactionCapacity=1000

TwitterAgent.sources.Twitter.channels = MemChannel

TwitterAgent.sinks.HDFS.channel = MemChannel

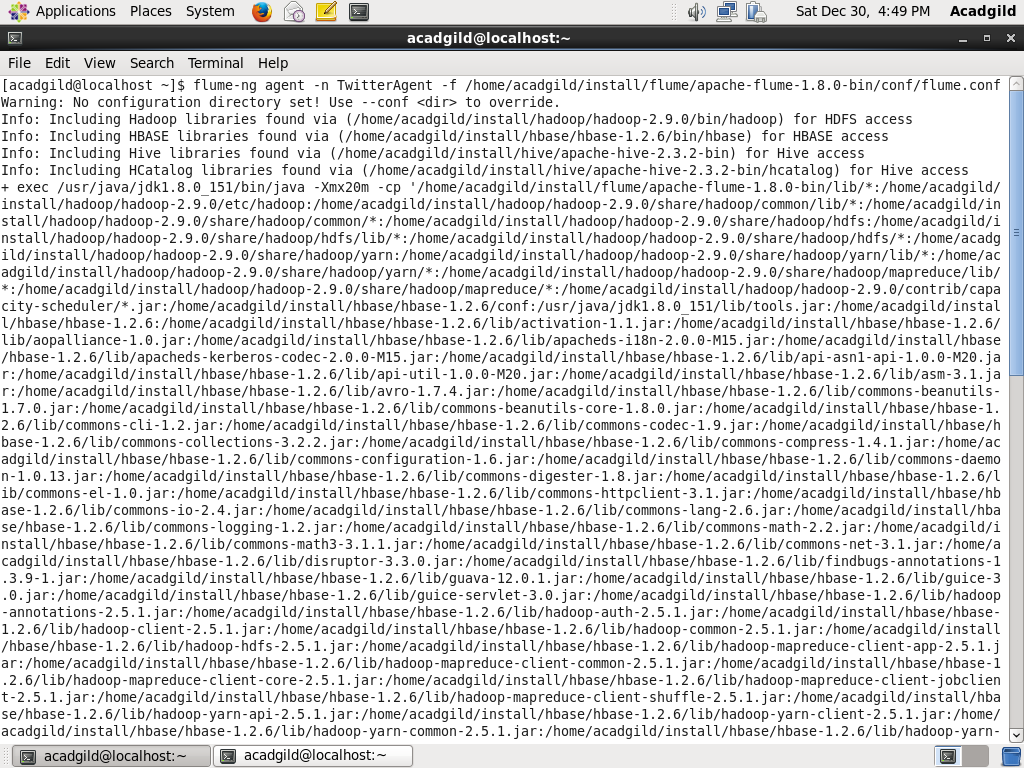
Streaming the data:

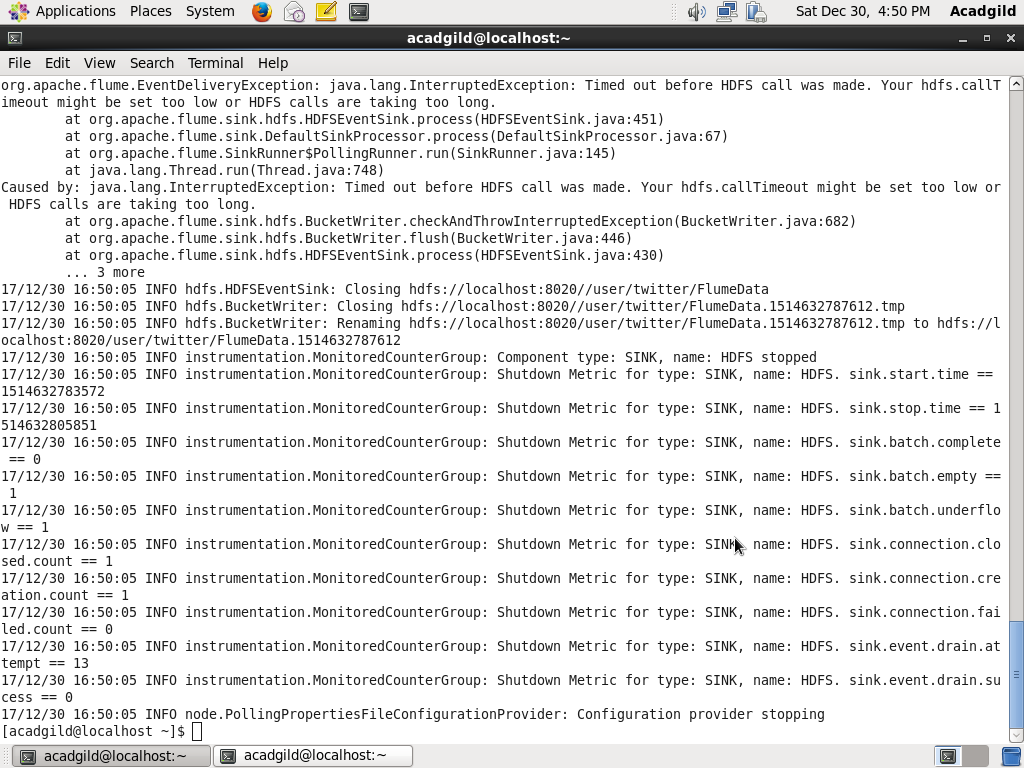
Firstly we will create HDFS directory (path same as provided in fulme.conf file).

Hadoop dfs –mkdir –p /user/twitter

Then for streaming the data from twitter we will execute the following command:

flume-ng agent -n TwitterAgent -f /home/acadgild/install/fume/apache-flume-1.8.0-bin/conf/flume.conf



Once, the tweet data started streaming it into the given HDFS path we can use ‘Ctrl+c’ command to stop the streaming process.

Viewing the contents of the file created.

hadoop dfs –cat /user/twitter/Flu\*

