**Session 20: Oozie, Sqoop & Flume**

**Twitter Data Analysis Using Flume**

Flume helps hadoop to get data from live streaming.

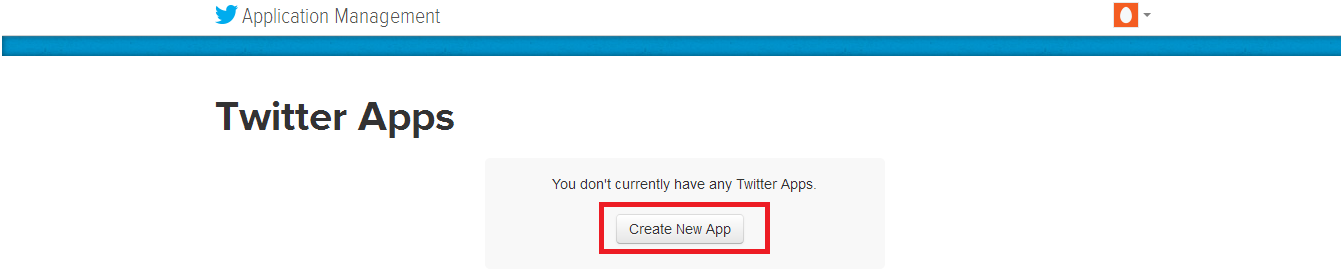
* Guarantees data delivery.
* Scales horizontally (connects commodity system in parallel) to handle additional data volume.

1. Create Twitter account

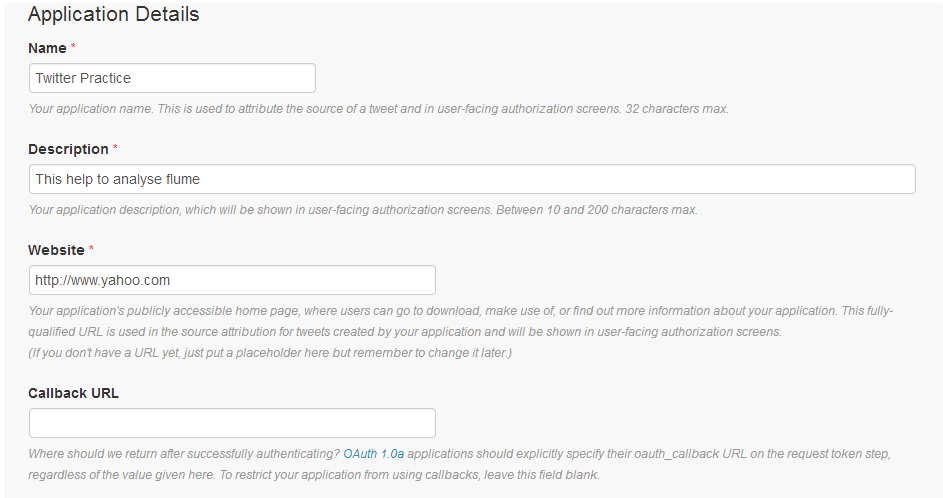
2. Login Twitter

3. Create Application

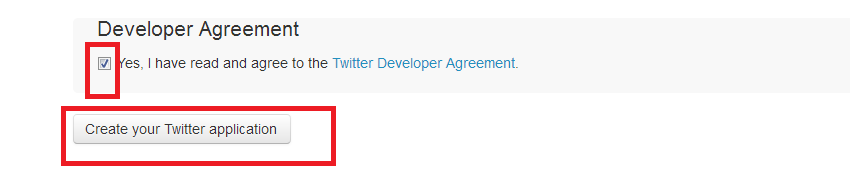
https://apps.twitter.com/

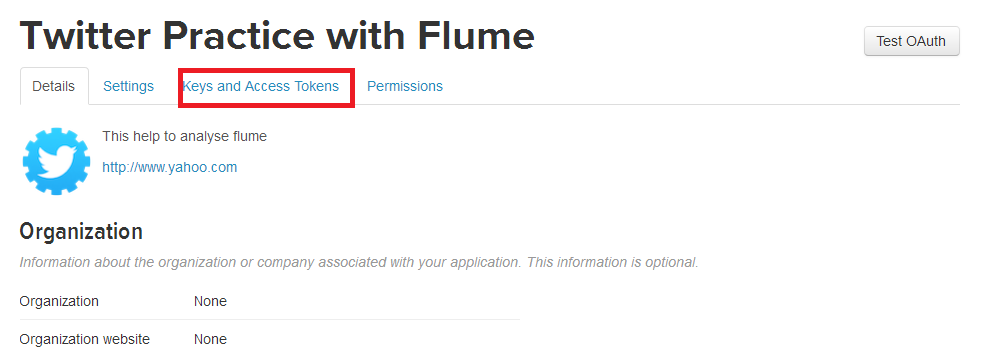


4. Create an application.

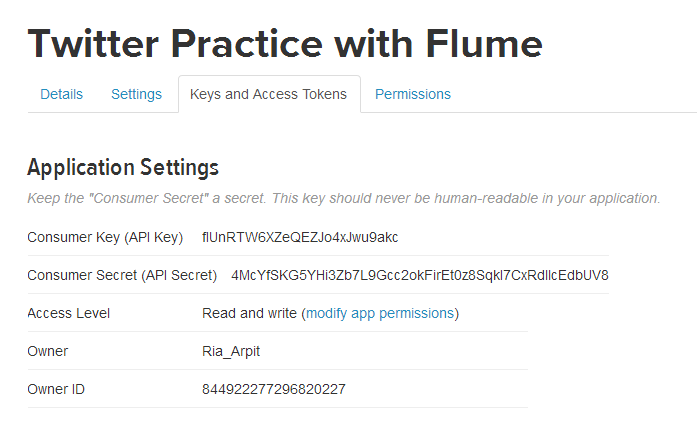


5. Accept the agreement and click on create application

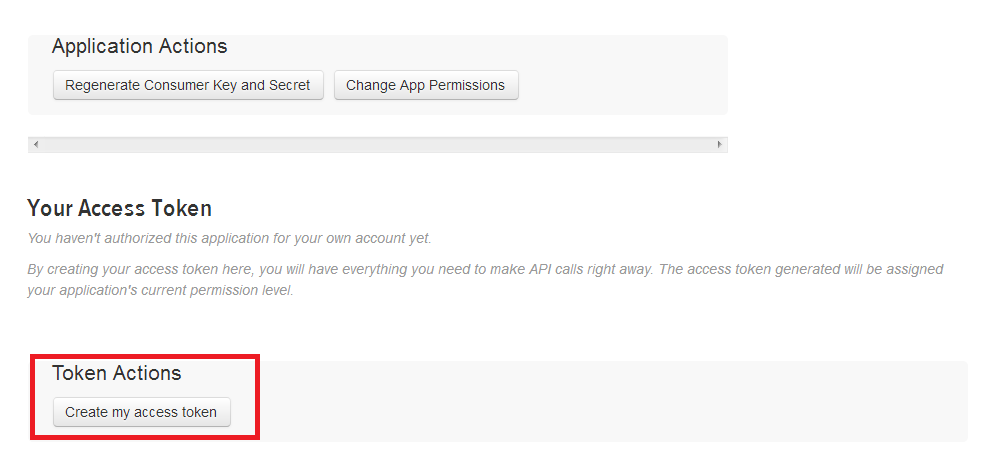
6. Select keys and access token tab

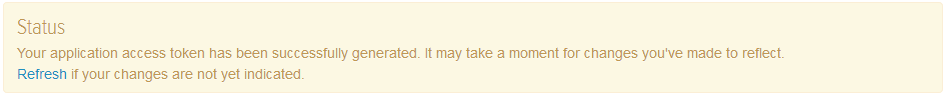


7. Copy the consumer key and the consumer secret.

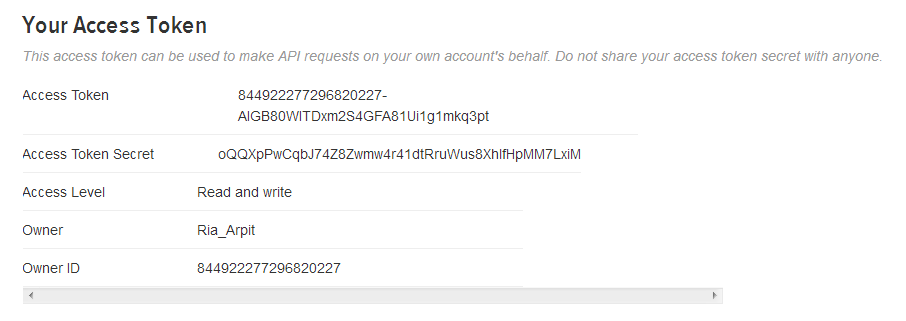


8. Scroll down further and click on ‘create my access token’.





scroll down to see access token



9. create flume-twitter.conf in /usr/lib/flume-ng/conf directory



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=consumerKey

TwitterAgent.sources.Twitter.consumerSecret=ConsumerSecret

TwitterAgent.sources.Twitter.accessToken=accessToken

TwitterAgent.sources.Twitter.accessTokenSecret=accessTokenSecret

TwitterAgent.sources.Twitter.keywords=hadoop, bigdata, mapreduce, mahout, hbase, nosql

# Describing/Configuring the sink

TwitterAgent.sources.Twitter.keywords= hadoop,election,sports, cricket,Big data

TwitterAgent.sinks.HDFS.channel=MemChannel

TwitterAgent.sinks.HDFS.type=hdfs

TwitterAgent.sinks.HDFS.hdfs.path=/user/flume/tweets

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

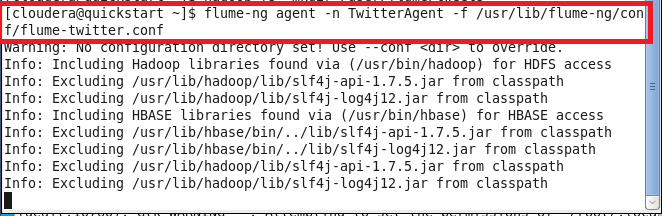
TwitterAgent.sources.Twitter.channels = MemChannel

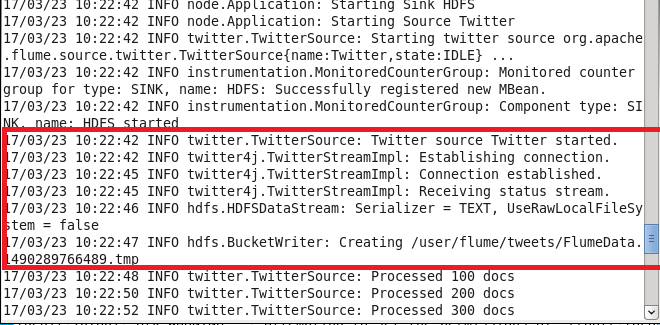
TwitterAgent.sinks.HDFS.channel = MemChannel

10. make directory in hdfs



11. Execute the app

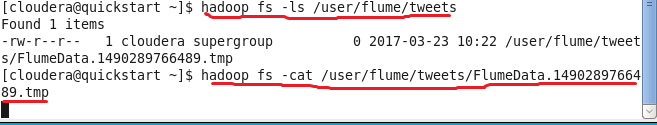




or

flume-ng agent --conf-file /usr/lib/flume-ng/conf/flume-twitter.conf --name TwitterAgent

12. Open another terminal window and check the output



We have completed the action of fetching live-streaming data from Twitter and loaded it to the HDFS, using Flume.

Press Ctrl+c to stop