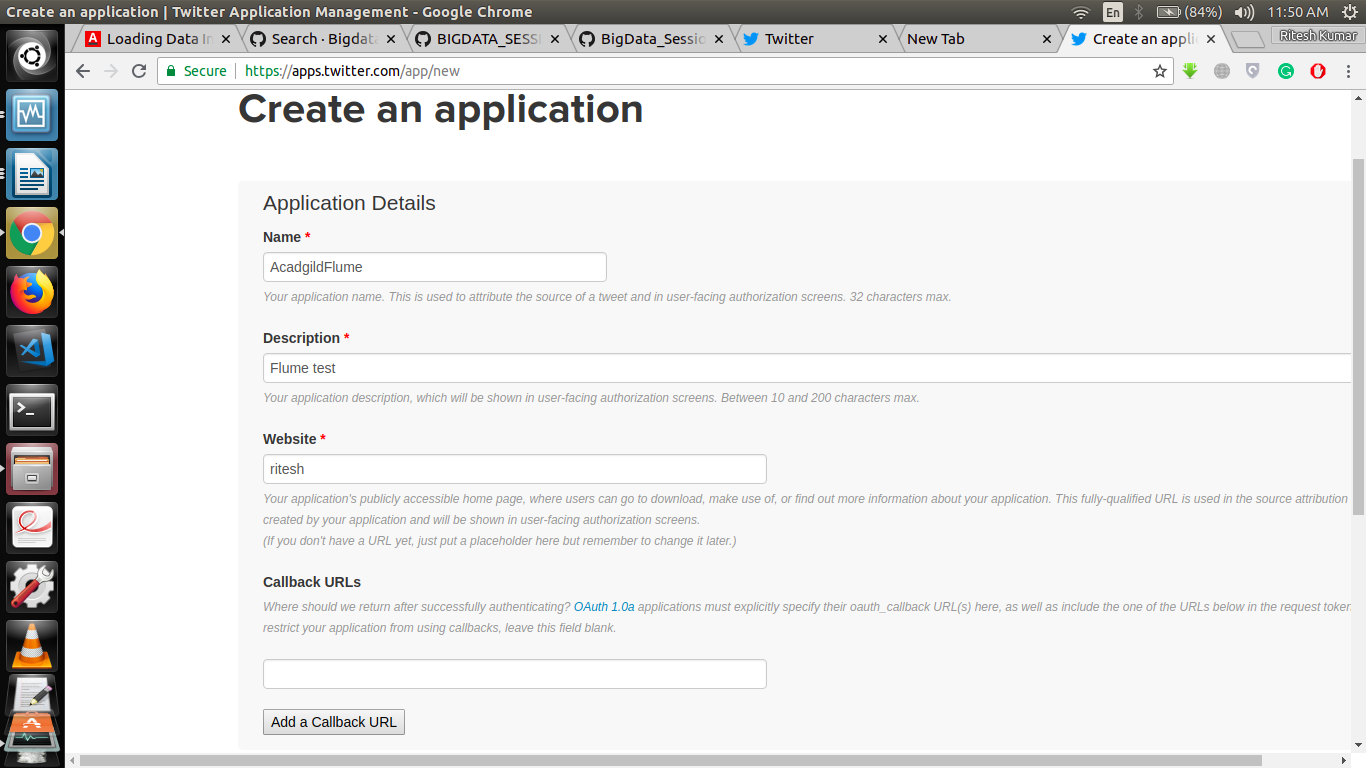
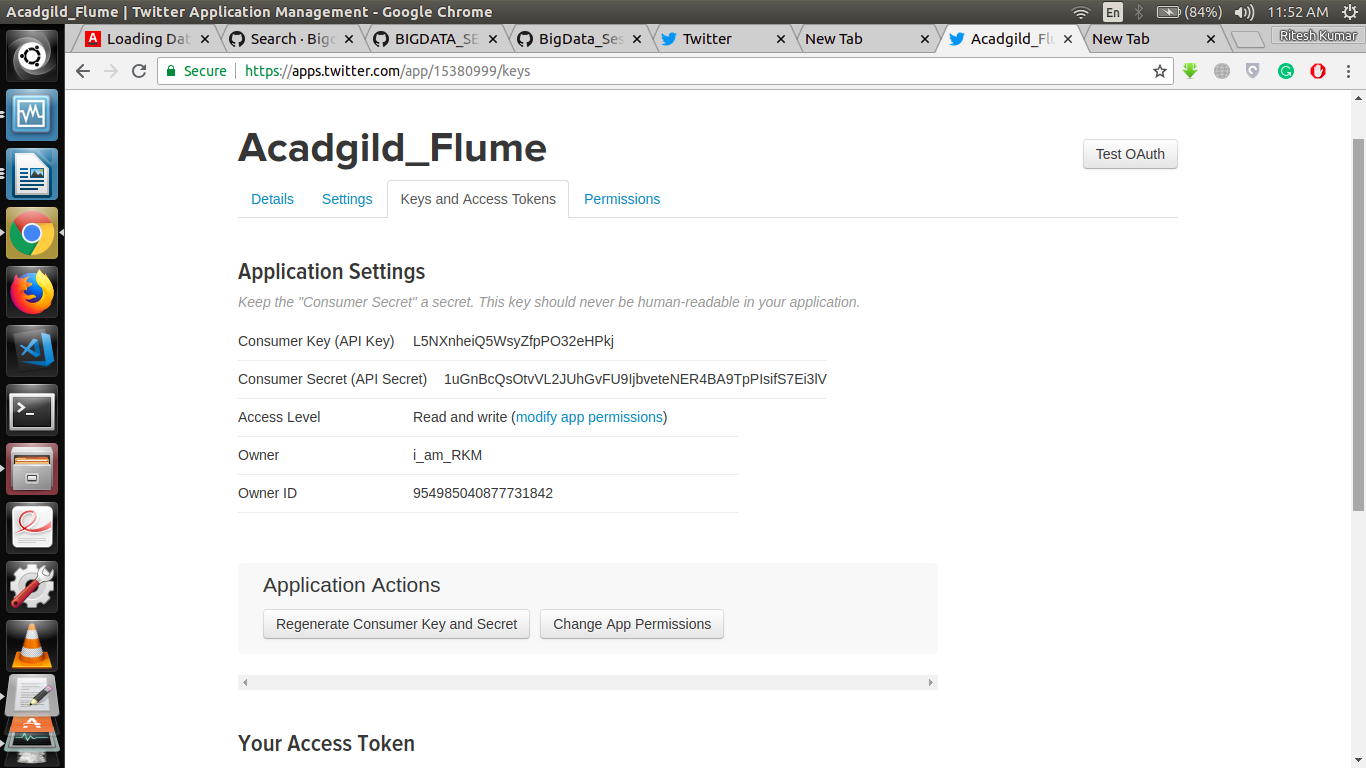
**Bigdata Assignment 4.5**

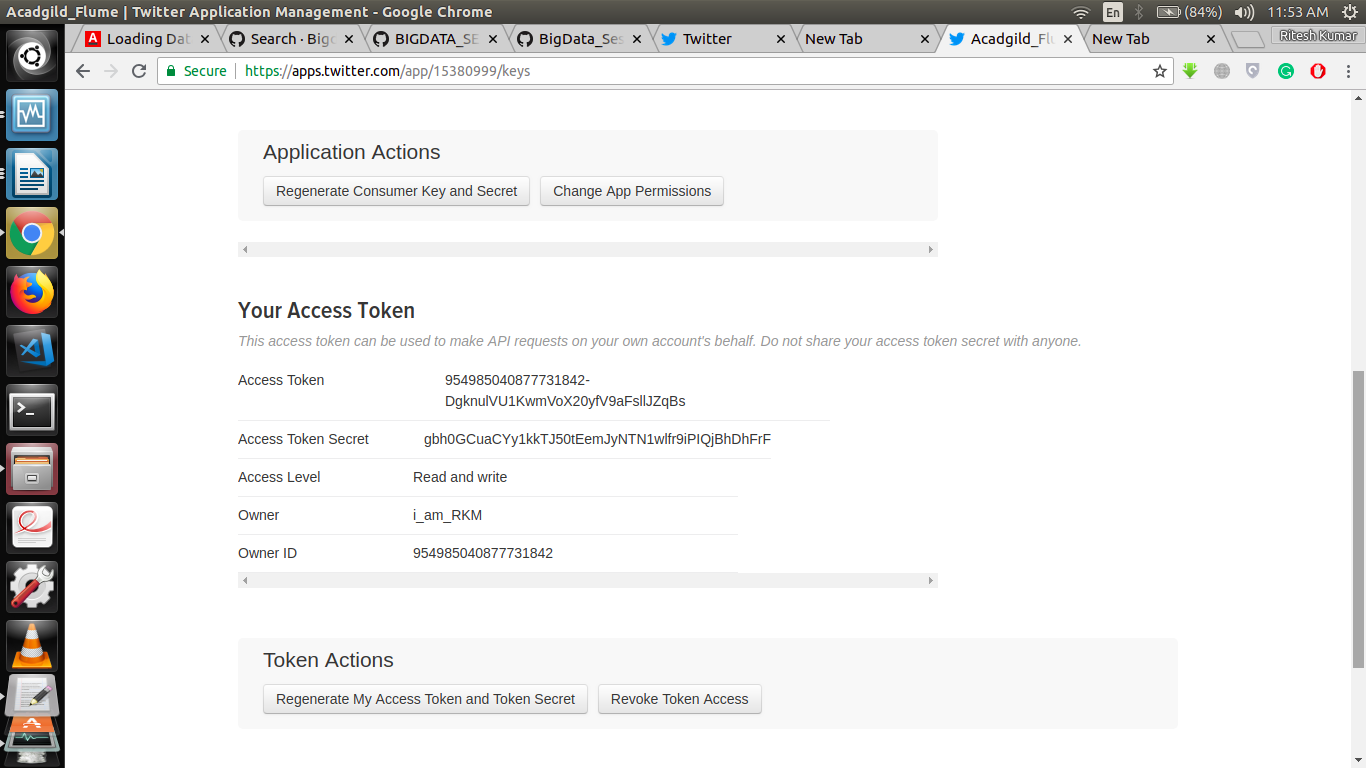
Create a flume agent that streams data from Twitter and stores in the HDFS.

Solution :

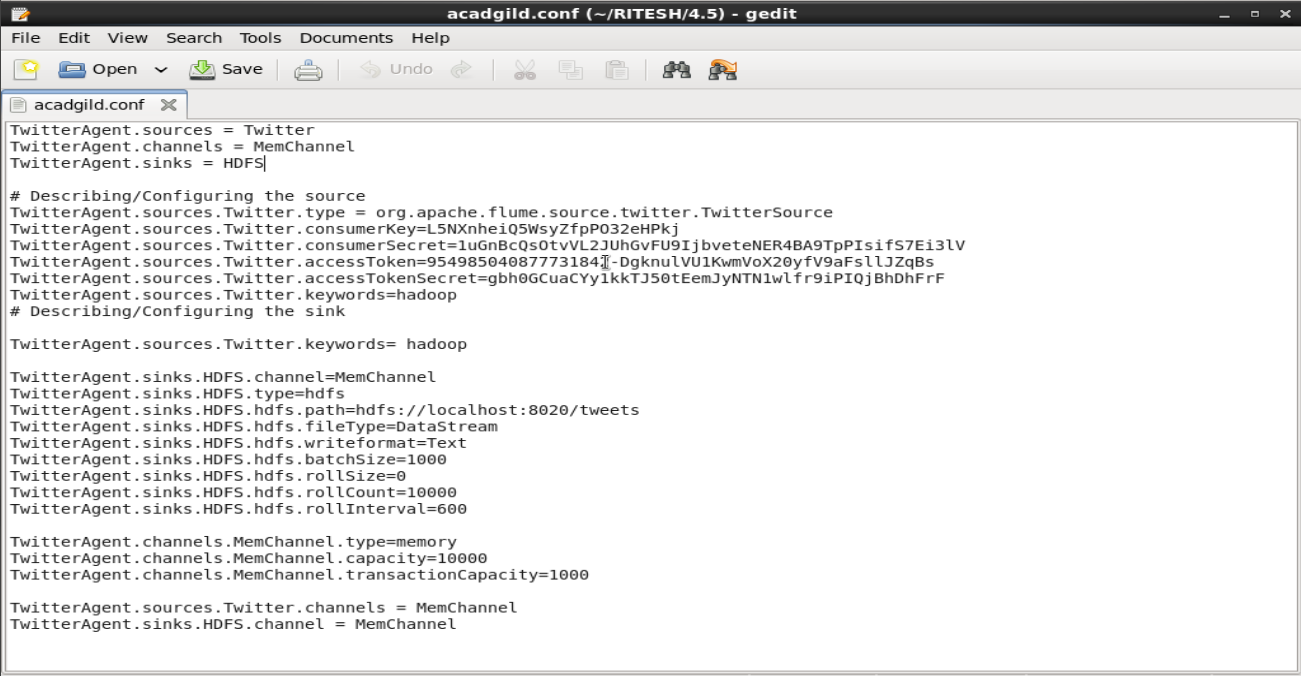
* Login to twitter
* Go to “<https://apps.twitter.come/app>” and create new app.
* Fill the details as hown in the below screenshot

****

* Go to “Keys and Access Tokens” tab and copy the “Consumer Key” and “Consumer Secrete Code”.
* Click on the “Create My Access Token” button.
* Copy “Access Token” and Access Token Secret”.

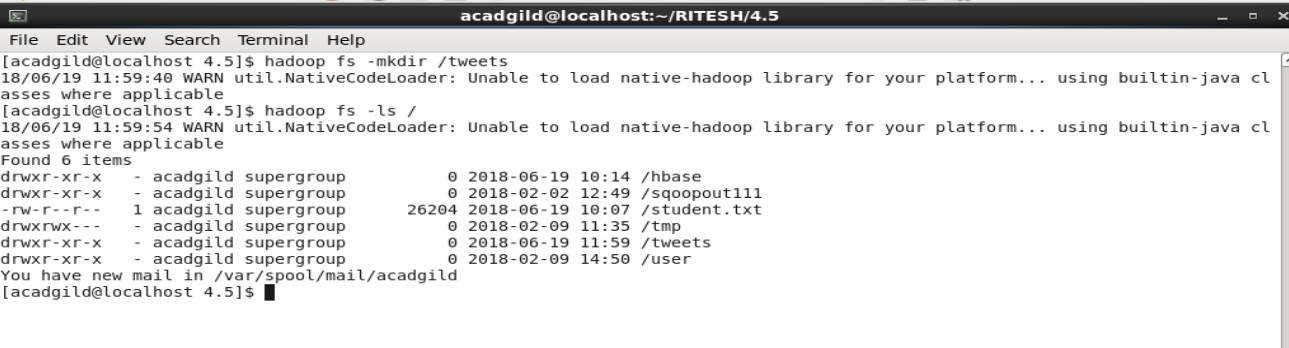


* Copy the tokens in the flume configuration file where we have set the sources , search terms , channels and sink



* Created a folder 'tweets' in hdfs.

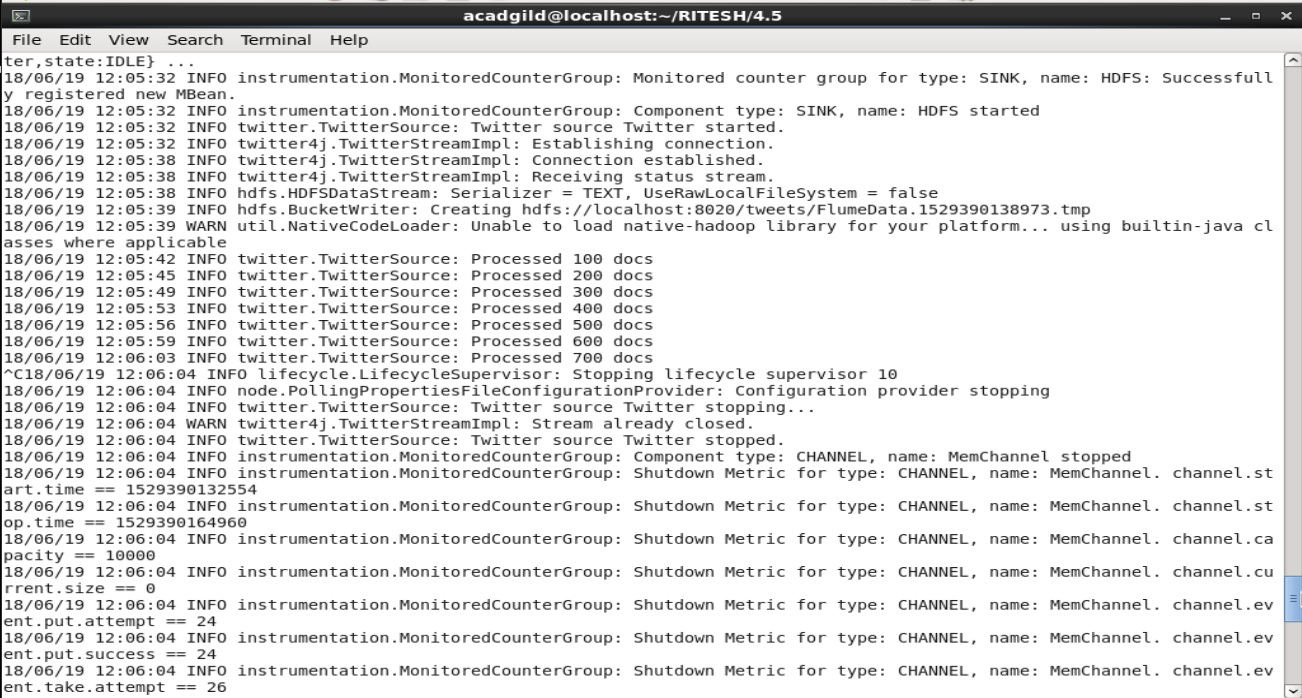
**hadoop fs -mkdir /tweets**



* Then gave flume command to get the tweets to the hdfs using flume streaming.

**flume ng-agent -n TwitterAgent -f acadgild.conf**

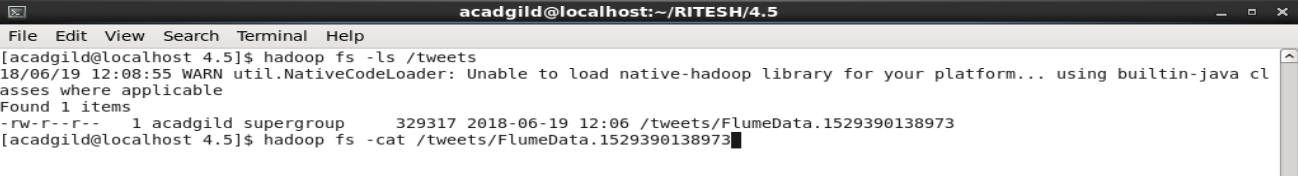


As it is visible in the screenshot , tweets are streaming into hdfs using flume from twitter source.

* Stopped the process after time using 'ctrl+d'
* Checked the tweets folder in hdfs and its contents

**hadoop fs -ls /tweets**

**hadoop fs -cat /tweets/FlumeData.1529390138973**



* As it is visible in the below screenshot , tweets are collected in the tweets folder in the hdfs.



**The configuration file for flume (Directly can't upload in github due to safety)**

**acadgild.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=L5NXnheiQ5WsyZfpPO32eHPkj**

**TwitterAgent.sources.Twitter.consumerSecret=1uGnBcQsOtvVL2JUhGvFU9IjbveteNER4BA9TpPIsifS7Ei3lV**

**TwitterAgent.sources.Twitter.accessToken=954985040877731842-DgknulVU1KwmVoX20yfV9aFsllJZqBs**

**TwitterAgent.sources.Twitter.accessTokenSecret=gbh0GCuaCYy1kkTJ50tEemJyNTN1wlfr9iPIQjBhDhFrF**

**TwitterAgent.sources.Twitter.keywords=hadoop**

**# Describing/Configuring the sink**

**TwitterAgent.sources.Twitter.keywords= hadoop**

**TwitterAgent.sinks.HDFS.channel=MemChannel**

**TwitterAgent.sinks.HDFS.type=hdfs**

**TwitterAgent.sinks.HDFS.hdfs.path=hdfs://localhost:8020/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**