ASSIGNMENT 12

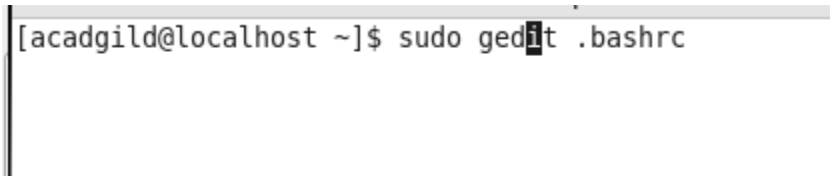
Creating a flume agent to stream data from twitter and store on HDFS

Step1: Creating a Twitter app

Step 2: Copying the consumer key and the consumer secret code

Step 3: creating the acceess token

Downlaoding flume tar file and save it under bashrc.file



```
# .bashrc

# Source global definitions
if [ -f /etc/bashrc ]; then
        . /etc/bashrc
fi

# User specific aliases and functions

export JAVA_HOME=/usr/java/jdk1.8.0_151
export PATH=$PATH:$JAVA_HOME/bin

# Below 2 lines we have to add for HADOOP Installation

export HADOOP_HOME=/home/acadgild/install/hadoop/hadoop-2.6.5
export PATH=$PATH:$JAVA_HOME/bin:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_COMMON_LIB_NATIVE_DIR=/home/acadgild/install/hadoop/hadoop-2.6.5/lib/native
export HADOOP_CONF_DIR=/home/acadgild/install/hadoop/hadoop-2.6.5/etc/hadoop/

# Below 2 lines we have to add for PIG Installation

export PIG_HOME=/home/acadgild/install/pig/pig-0.16.0
export PATH=$PATH:$PIG_HOME/bin

# Below 2 lines we have to add for HIVE Installation

export HIVE_HOME=/home/acadgild/install/hive/apache-hive-2.3.2-bin
export PATH=$PATH:$HIVE_HOME/bin

#BElow 2 lines we have to add for SPARK Installation

export SPARK_HOME=/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7
export PATH=$PATH:$SPARK_HOME/bin
```

Update bashrc file

```
[acadgild@localhost ~]$ source.bashrc
```

Creating a new file under Flume-extracted directory

Change the twitter api keys with the keys generated

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=uX0TWqkx0okYEj
jqLzxIx6mD6

TwitterAgent.sources.Twitter.consumerSecret=rzHIs3TMJnAD
bZNvdGU7LQUo0kPxPISq3RGSLfqcBip39X5END

TwitterAgent.sources.Twitter.accessToken=559516596-
yDA9xqOljo4CV32wSnqsx2BXh4RBIRKFxZGSZrPC

TwitterAgent.sources.Twitter.accessTokenSecret=zDxePILZitS
5tIWBhre0GWqps0FIj9OadX8RZb6w8ZCwz

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=hdfs://localhost:9000/use
r/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

Creating a new directory

```
[acadgild@localhost ~]$ hadoop fs -mkdir - p /user/flume/tweets/
```

fetching the data from twitter using the command

```
File  Edit  View  Search  Terminal  Help
[hadoop@acadgild ~]$ flume-ng agent -n TwitterAgent  -f /home/hadoop/HADOOP/apac
he-flume-1.6.0-bin/conf/flume.conf
```

OP is obtained by the cat command

```
[hadoop@acadgild ~]$ hadoop dfs -cat /user/flume/tweets/FlumeData.1450940925854
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
```