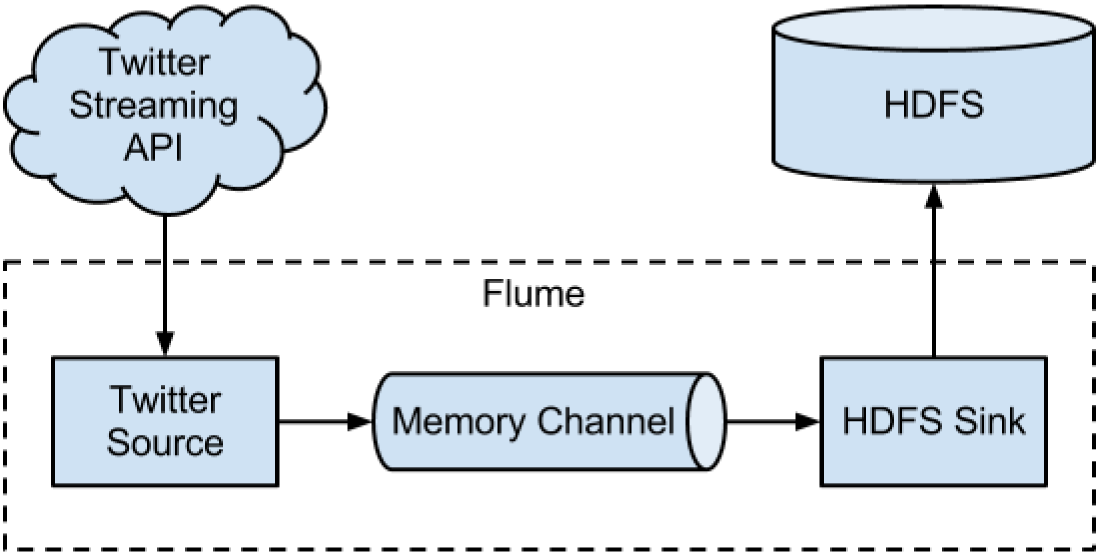
Documentation of Streaming of Twitter data using Apache Flume

-Prepared by Vignesh.R (15CSE107)

Abstract:

Apache Flume is a distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of log data. It has a simple and flexible architecture based on streaming data flows. It is robust and fault tolerant with tunable reliability mechanisms and many failover and recovery mechanisms. It uses a simple extensible data model that allows for online analytic application. This document depicts the streaming of Twitter data using Flume engine.



There are 3 major components, namely: Source, Channel, and Sink, which are involved in ingesting data, moving data and storing data, respectively.

Below is the breakdown of the parts applicable in this scenario:

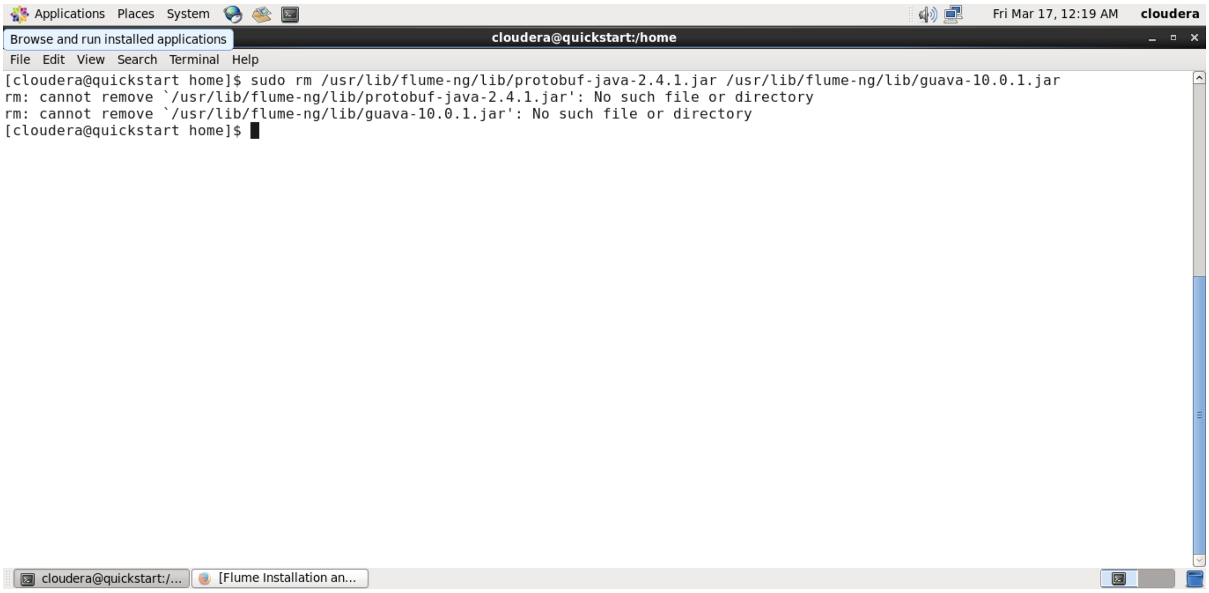
- **Event** – A singular unit of data that is transported by Flume (typically a single log entry).
- **Source** – The entity through which data enters into the Flume. Sources either actively samples the data or passively waits for data to be delivered to them. A variety of sources such as log4j logs and syslogs, allows data to be collected.
- **Sink** – The unit that delivers the data to the destination. A variety of sinks allow data to be streamed to a range of destinations. Example: HDFS sink writes events to the HDFS.
- **Channel** – It is the connection between the Source and the Sink. The Source ingests Event into the Channel and the Sink drains the Channel.
- **Agent** – Any physical Java virtual machine running Flume. It is a collection of Sources, Sinks and Channels.
- **Client** – It produces and transmits the Event to the Source operating within the Agent

Streaming Twitter data from Flume:-

Step 1:

Remove protobuf-java-2.4.1.jar and guava-10.1.1.jar from library directory of flume-ng. Super user permission is given for this modification. If no such directories are available, then ignore it.

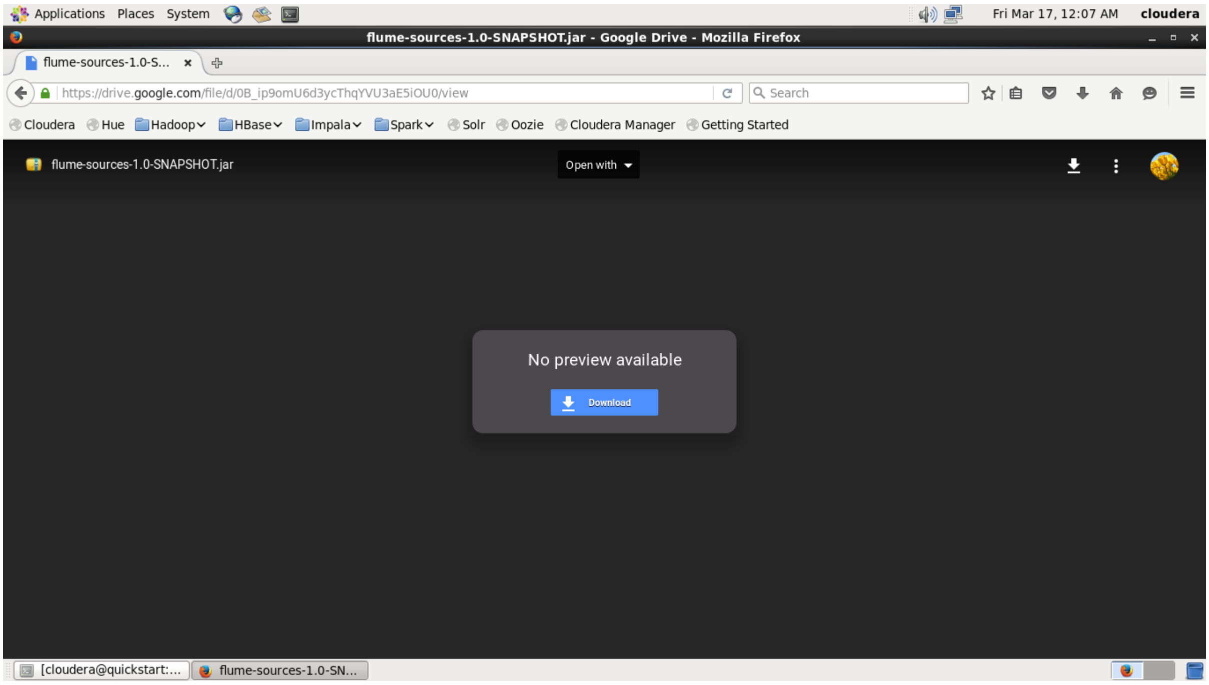
Command: sudo rm /usr/lib/flume-ng/lib/protobuf-java-2.4.1.jar /usr/lib/flume-ng/lib/guava-10.0.1.jar


Applications Places System
Fri Mar 17, 12:19 AM
cloudera
Browse and run installed applications
cloudera@quickstart:/home
File Edit View Search Terminal Help
cloudera@quickstart:/...
[Flume Installation an...

```
[cloudera@quickstart home]$ sudo rm /usr/lib/flume-ng/lib/protobuf-java-2.4.1.jar /usr/lib/flume-ng/lib/guava-10.0.1.jar
rm: cannot remove `/usr/lib/flume-ng/lib/protobuf-java-2.4.1.jar': No such file or directory
rm: cannot remove `/usr/lib/flume-ng/lib/guava-10.0.1.jar': No such file or directory
[cloudera@quickstart home]$ 
```

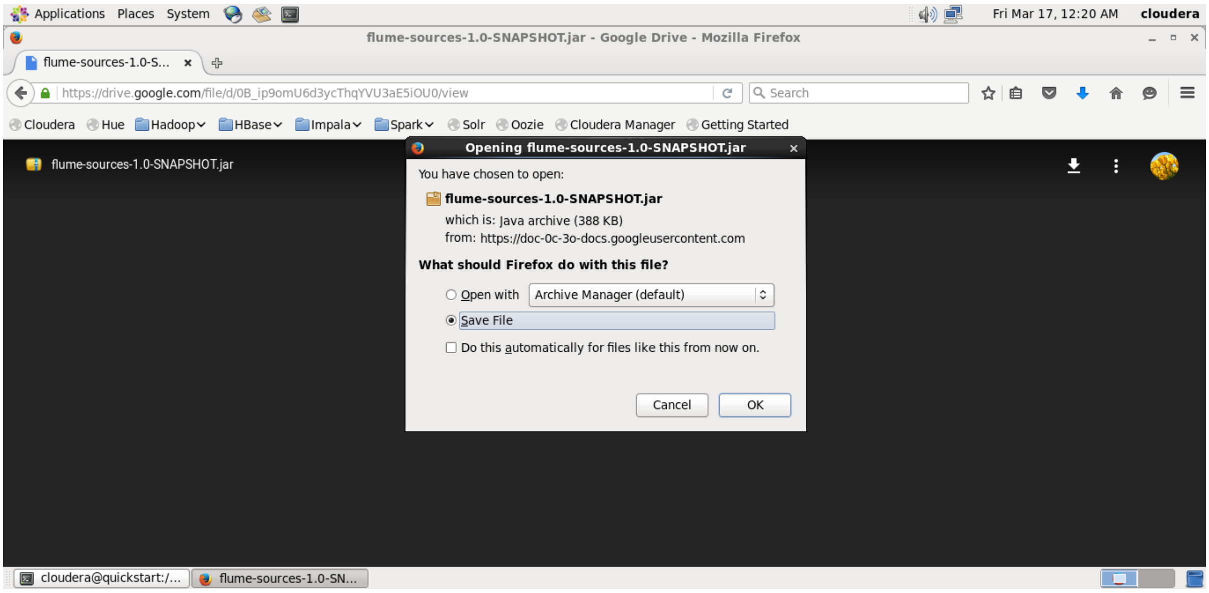
Step 2 :

The flume-sources-1.0-SNAPSHOT.jar file is downloaded from https://drive.google.com/file/d/0B_ip9omU6d3ycThqYVU3aE5iOU0/view?usp=sharing link.



Step 3:

Click on Download option and then save the jar file.



Opening flume-sources-1.0-SNAPSHOT.jar

You have chosen to open:

flume-sources-1.0-SNAPSHOT.jar

which is: Java archive (388 KB)

from: https://doc-0c-3o-docs.googleusercontent.com

What should Firefox do with this file?

○ Open with Archive Manager (default)

● Save File

□ Do this automatically for files like this from now on.

Cancel OK

Step 4:

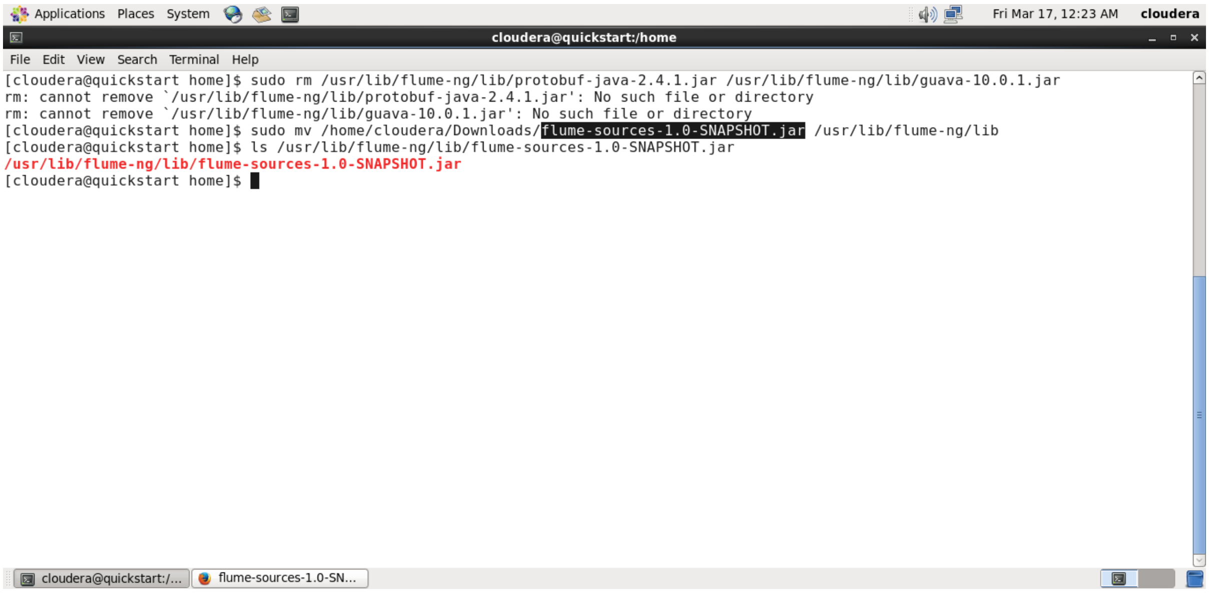
Move the flume SNAPSHOT.jar file to the lib folder of flume from downloads.

Command: sudo mv /home/cloudera/Downloads/flume-sources-1.0-SNAPSHOT.jar/usr/lib/flume-ng/lib

Step 5:

Check whether SNAPSHOT.jar file is moved to the library folder of flume-ng.

Command: ls /usr/lib/flume-ng/lib/flume-sources-1.0-SNAPSHOT.jar

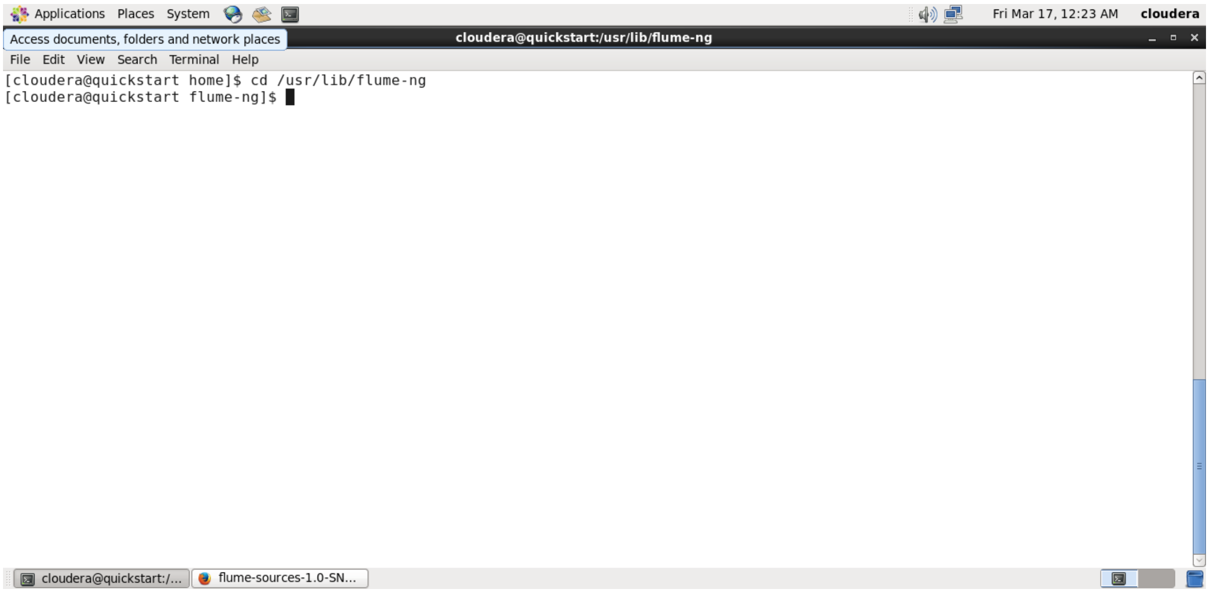


```
[cloudera@quickstart home]$ sudo rm /usr/lib/flume-ng/lib/protobuf-java-2.4.1.jar /usr/lib/flume-ng/lib/guava-10.0.1.jar
rm: cannot remove `/usr/lib/flume-ng/lib/protobuf-java-2.4.1.jar': No such file or directory
rm: cannot remove `/usr/lib/flume-ng/lib/guava-10.0.1.jar': No such file or directory
[cloudera@quickstart home]$ sudo mv /home/cloudera/Downloads/flume-sources-1.0-SNAPSHOT.jar /usr/lib/flume-ng/lib
[cloudera@quickstart home]$ ls /usr/lib/flume-ng/lib/flume-sources-1.0-SNAPSHOT.jar
/usr/lib/flume-ng/lib/flume-sources-1.0-SNAPSHOT.jar
[cloudera@quickstart home]$ 
```

Step 6:

Then the present working directory is moved to the flume-ng directory.

Command: cd /usr/lib/flume-ng/



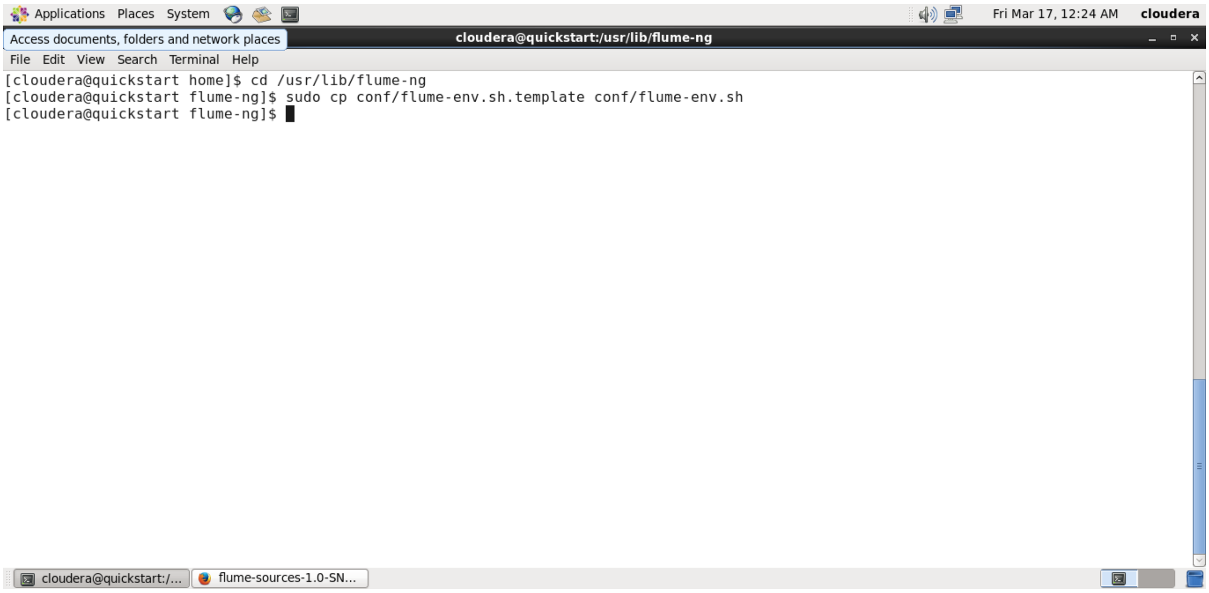
```
[cloudera@quickstart home]$ cd /usr/lib/flume-ng
[cloudera@quickstart flume-ng]$ 
```

Applications Places System
Fri Mar 17, 12:23 AM cloudera
Access documents, folders and network places
cloudera@quickstart:/usr/lib/flume-ng
File Edit View Search Terminal Help
cloudera@quickstart:/...
flume-sources-1.0-SN...

Step 7:

The flume-env.sh.template content is copied to flume-env.sh. Super user permission is given for this modification.

Command: sudo cp conf/flume-env.sh.template conf/flume-env.sh



```
[cloudera@quickstart home]$ cd /usr/lib/flume-ng
[cloudera@quickstart flume-ng]$ sudo cp conf/flume-env.sh.template conf/flume-env.sh
[cloudera@quickstart flume-ng]$ 
```

Applications Places System
Fri Mar 17, 12:24 AM cloudera
Access documents, folders and network places
cloudera@quickstart:/usr/lib/flume-ng
File Edit View Search Terminal Help
cloudera@quickstart:/...
flume-sources-1.0-SN...

Step 8:

Then the JAVA_HOME and the FLUME_CLASSPATH of the copied flume-env.sh file is edited using gedit command. Super user permission is given for this modification.

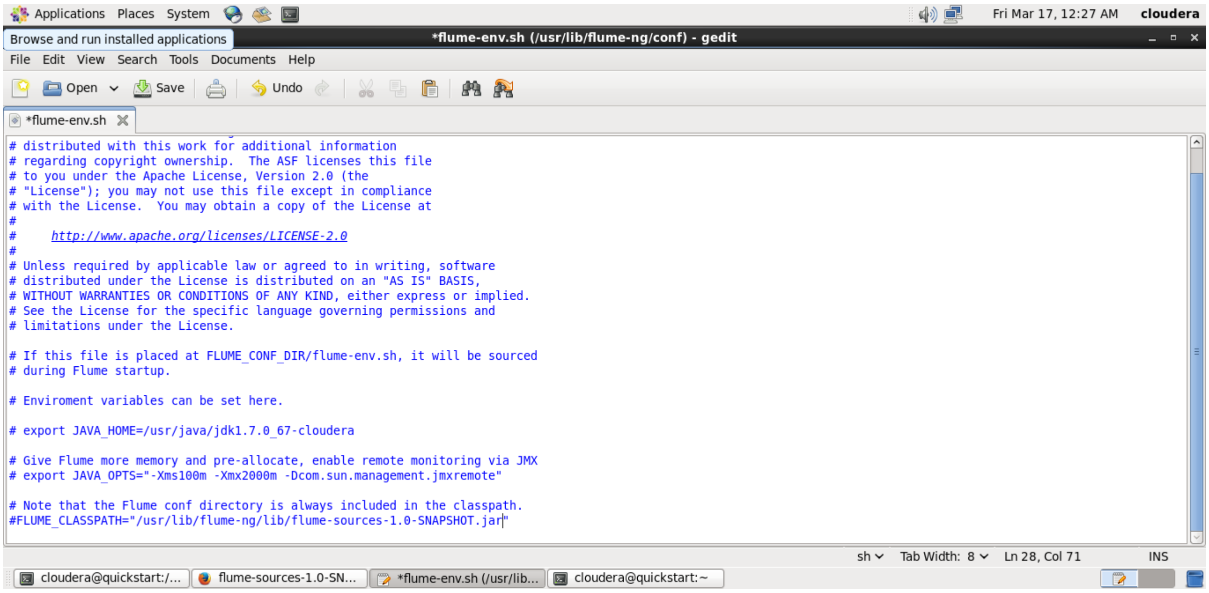
Command: sudo gedit conf/flume-env.sh

Step 9:

Here, the java path has to be identified from the local machine and java path is set. The class path of flume is set to the downloaded SNAPSHOT.jar file location. Then save the file and close it.

Command: JAVA_HOME= /usr/java/jdk1.7.0_67-cloudera

Command: FLUME_CLASSPATH=”/usr/lib/flume-ng/lib/flume-sources-1.0-SNAPSHOT.jar”

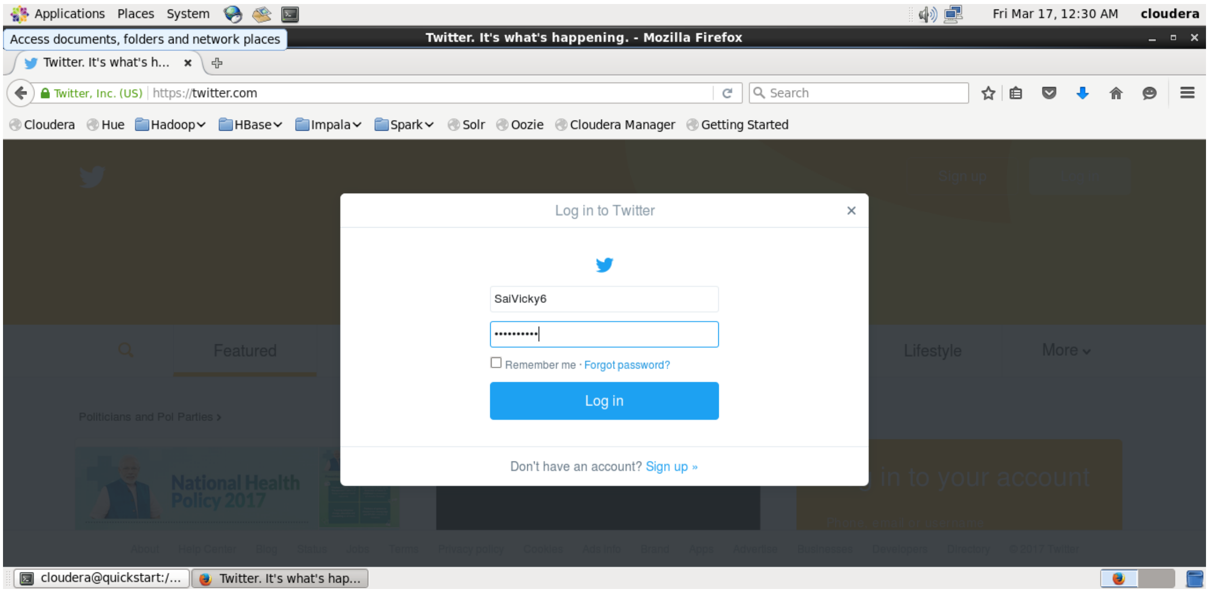

Applications Places System
Fri Mar 17, 12:27 AM cloudera
Browse and run installed applications
*flume-env.sh (/usr/lib/flume-ng/conf) - gedit
File Edit View Search Tools Documents Help
Open Save Undo
*flume-env.sh
sh Tab Width: 8 Ln 28, Col 71 INS
cloudera@quickstart:/...
flume-sources-1.0-SN...
*flume-env.sh (/usr/lib...
cloudera@quickstart:~

```
# distributed with this work for additional information
# regarding copyright ownership.  The ASF licenses this file
# to you under the Apache License, Version 2.0 (the
# "License"); you may not use this file except in compliance
# with the License.  You may obtain a copy of the License at
#
#     http://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.

# If this file is placed at FLUME_CONF_DIR/flume-env.sh, it will be sourced
# during Flume startup.

# Enviroment variables can be set here.

# export JAVA_HOME=/usr/java/jdk1.7.0_67-cloudera

# Give Flume more memory and pre-allocate, enable remote monitoring via JMX
# export JAVA_OPTS="-Xms100m -Xmx2000m -Dcom.sun.management.jmxremote"

# Note that the Flume conf directory is always included in the classpath.
#FLUME_CLASSPATH="/usr/lib/flume-ng/lib/flume-sources-1.0-SNAPSHOT.jar"
```

Step 10:

Open the web browser and sign in to your account in twitter. Then enter your login credentials and open your account.

Website link: https://twitter.com/login



Applications Places System
Fri Mar 17, 12:30 AM cloudera
Access documents, folders and network places
Twitter. It's what's happening. - Mozilla Firefox
Twitter. It's what's h...
Twitter, Inc. (US) https://twitter.com
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Log in to Twitter
SaiVicky6
••••••••••
Remember me · Forgot password?
Log in
Don't have an account? Sign up »
cloudera@quickstart:/...
Twitter. It's what's hap...

Step 11:

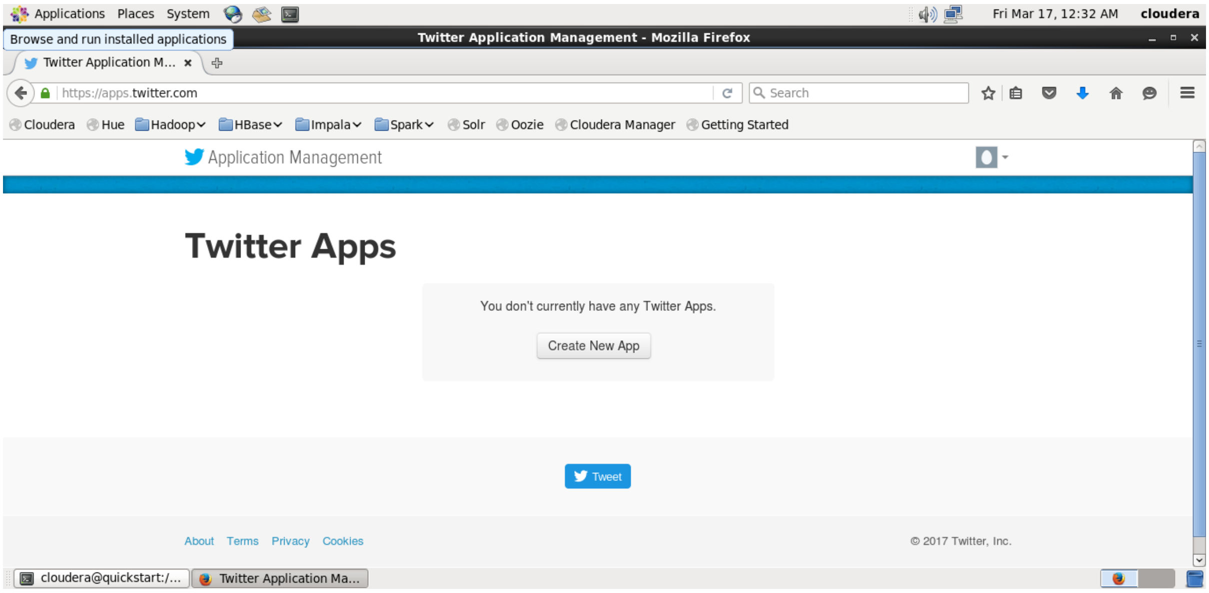
Then switch to the apps site from your twitter account.

Website link: https://apps.twitter.com

Step 12:

Application window appears. Click on the Create Application option.

Note: To create an application, your phone number has to be added in your twitter account.



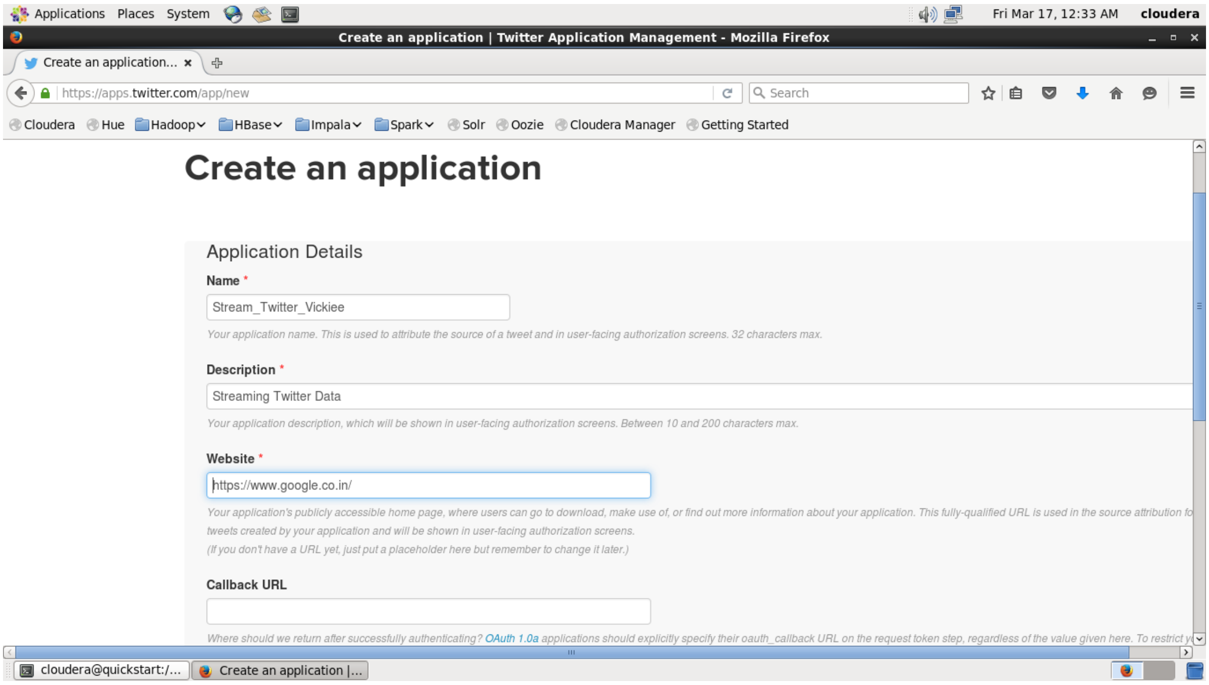
Applications Places System
Fri Mar 17, 12:32 AM cloudera
Browse and run installed applications
Twitter Application Management - Mozilla Firefox
Twitter Application M...
https://apps.twitter.com
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Application Management
Twitter Apps
Create New App
Tweet
About Terms Privacy Cookies
© 2017 Twitter, Inc.
cloudera@quickstart:/...
Twitter Application Ma...

You don't currently have any Twitter Apps.

Step 13:

Enter the Application name, its Description and the website.

Note: Application name should be unique. Search engine site is given in the website field. Callback URL field is left blank.


Applications Places System
Fri Mar 17, 12:33 AM cloudera
Create an application | Twitter Application Management - Mozilla Firefox
Create an application...
https://apps.twitter.com/app/new
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
cloudera@quickstart:/...
Create an application |...

Create an application

Application Details

Name *

Stream_Twitter_Vickiee

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description *

Streaming Twitter Data

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website *

https://www.google.co.in/

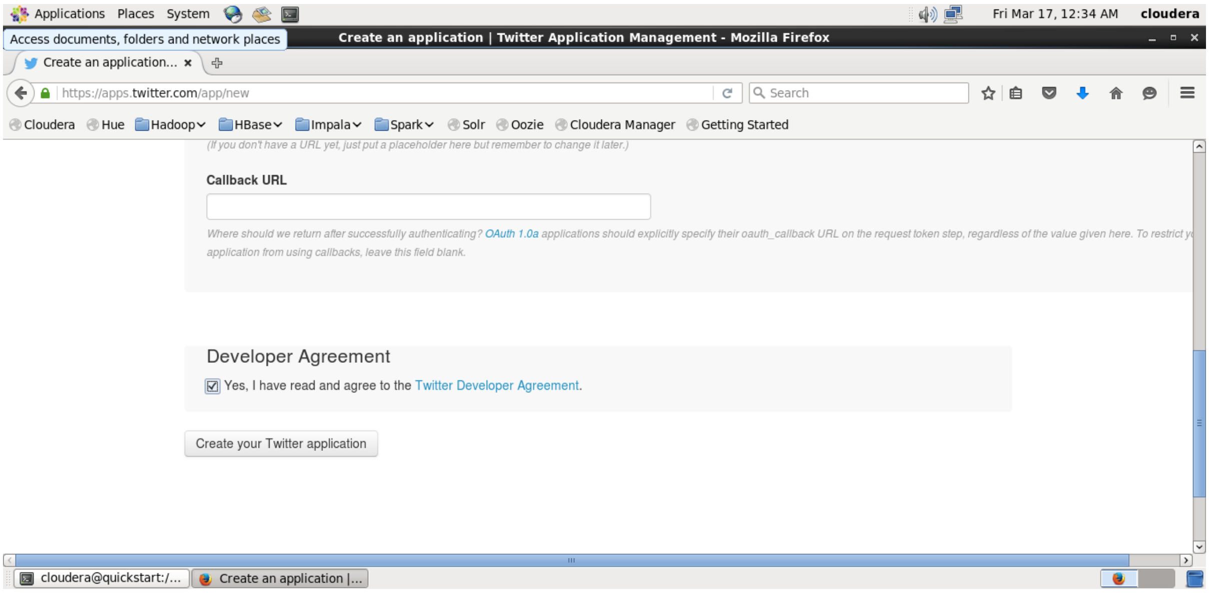
Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution fo tweets created by your application and will be shown in user-facing authorization screens.
(If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Callback URL

Where should we return after successfully authenticating? OAuth 1.0a applications should explicitly specify their oauth_callback URL on the request token step, regardless of the value given here. To restrict y

Step 14:

Check the ‘Yes, I agree’ and click on ‘Create your Twitter application’.


Applications Places System
Fri Mar 17, 12:34 AM cloudera
Access documents, folders and network places
Create an application | Twitter Application Management - Mozilla Firefox
Create an application...
https://apps.twitter.com/app/new
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
cloudera@quickstart:/...
Create an application |...

(If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Callback URL

Where should we return after successfully authenticating? OAuth 1.0a applications should explicitly specify their oauth_callback URL on the request token step, regardless of the value given here. To restrict y application from using callbacks, leave this field blank.

Developer Agreement

☑ Yes, I have read and agree to the Twitter Developer Agreement.

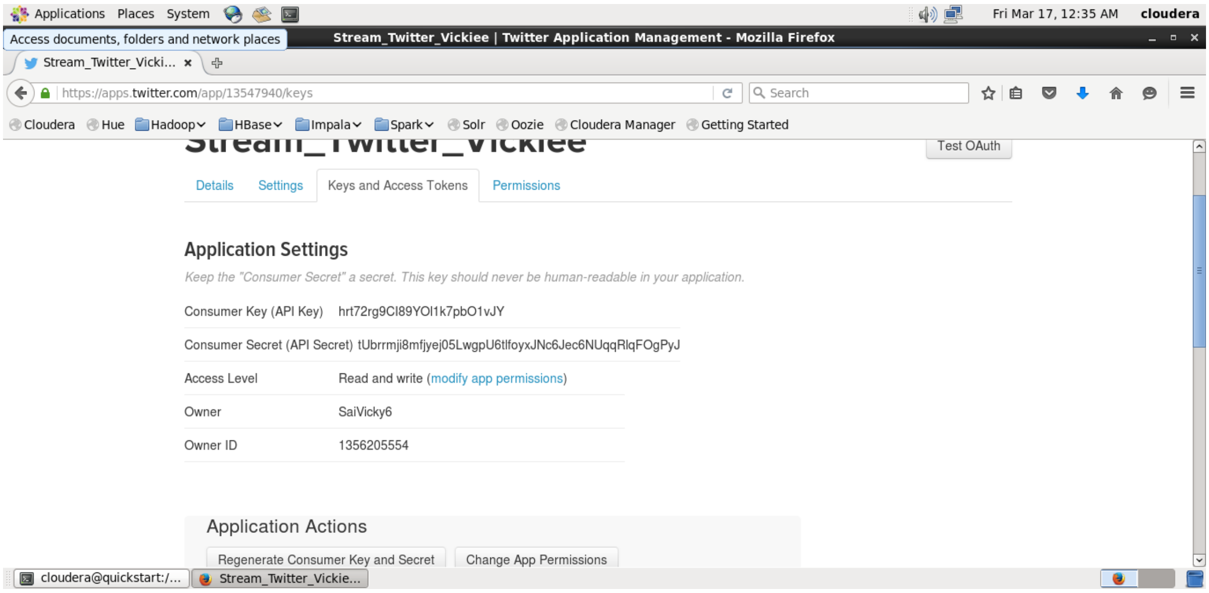
Create your Twitter application

Step 15:

Now your application will be created.

Step 16:

Then click on Keys and Access tokens tab. The Consumer key and Consumer Secret keys will be generated.



Applications Places System
Fri Mar 17, 12:35 AM cloudera
Access documents, folders and network places
Stream_Twitter_Vickiee | Twitter Application Management - Mozilla Firefox
Stream_Twitter_Vicki...
https://apps.twitter.com/app/13547940/keys
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Stream_Twitter_Vickiee
Test OAuth
Details
Settings
Keys and Access Tokens
Permissions
Regenerate Consumer Key and Secret
Change App Permissions
cloudera@quickstart:/...
Stream_Twitter_Vickie...

Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key)	hrt72rg9Cl89YOl1k7pbO1vJY
Consumer Secret (API Secret)	tUbrrmji8mfjyej05LwgpU6tlfoyxJNc6Jec6NUqqRlqFOgPyJ
Access Level	Read and write (modify app permissions)
Owner	SaiVicky6
Owner ID	1356205554

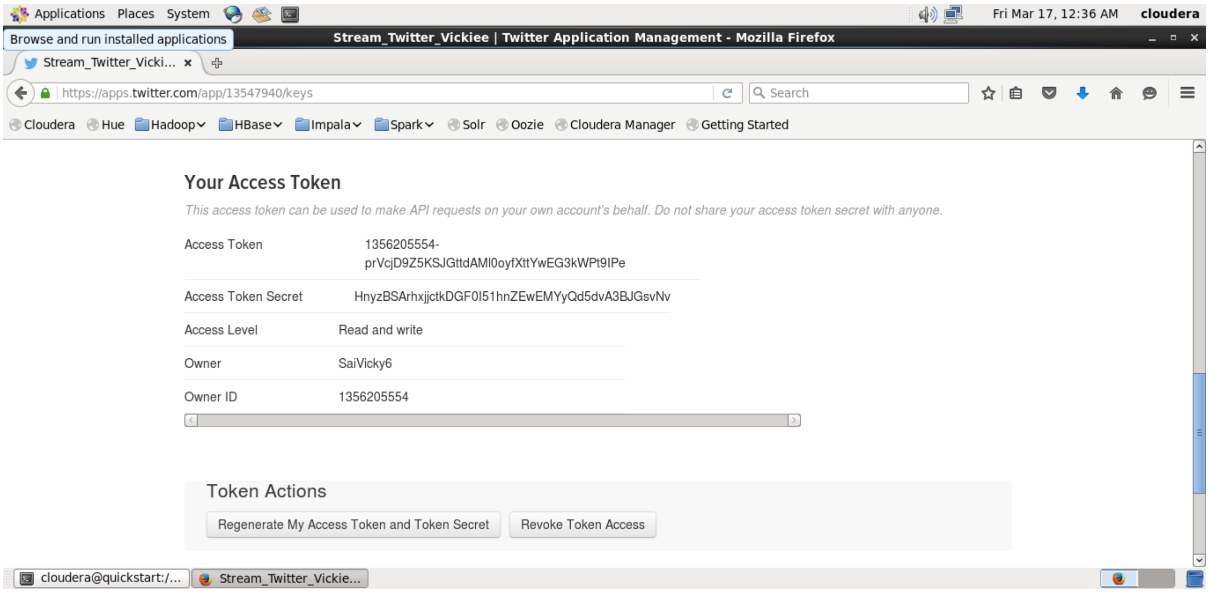
Application Actions

Step 17:

Scroll down. Click on ‘Create my Access token’.

Step 18:

The Access token and Access token Secret will be generated.



Applications Places System
Fri Mar 17, 12:36 AM cloudera
Browse and run installed applications
Stream_Twitter_Vickiee | Twitter Application Management - Mozilla Firefox
Stream_Twitter_Vicki...
https://apps.twitter.com/app/13547940/keys
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Regenerate My Access Token and Token Secret
Revoke Token Access
cloudera@quickstart:/...
Stream_Twitter_Vickie...

Your Access Token

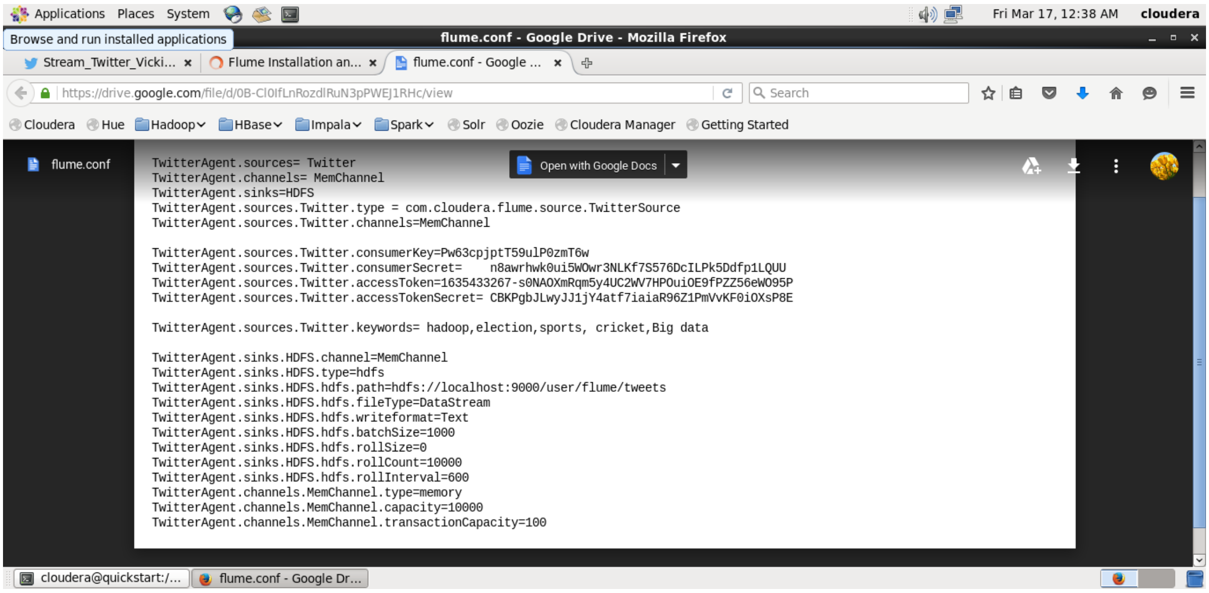
This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

Access Token	1356205554- prVcjD9Z5KSJGttdAMl0oyfXttYwEG3kWPt9IPe
Access Token Secret	HnyzBSArhxjjctkDGF0I51hnZEwEMYyQd5dvA3BJGsvNv
Access Level	Read and write
Owner	SaiVicky6
Owner ID	1356205554

Token Actions

Step 19:

Then download the flume.conf file from https://drive.google.com/file/d/0B-Cl0IfLnRozdlRuN3pPWEJ1RHc/view?usp=sharing



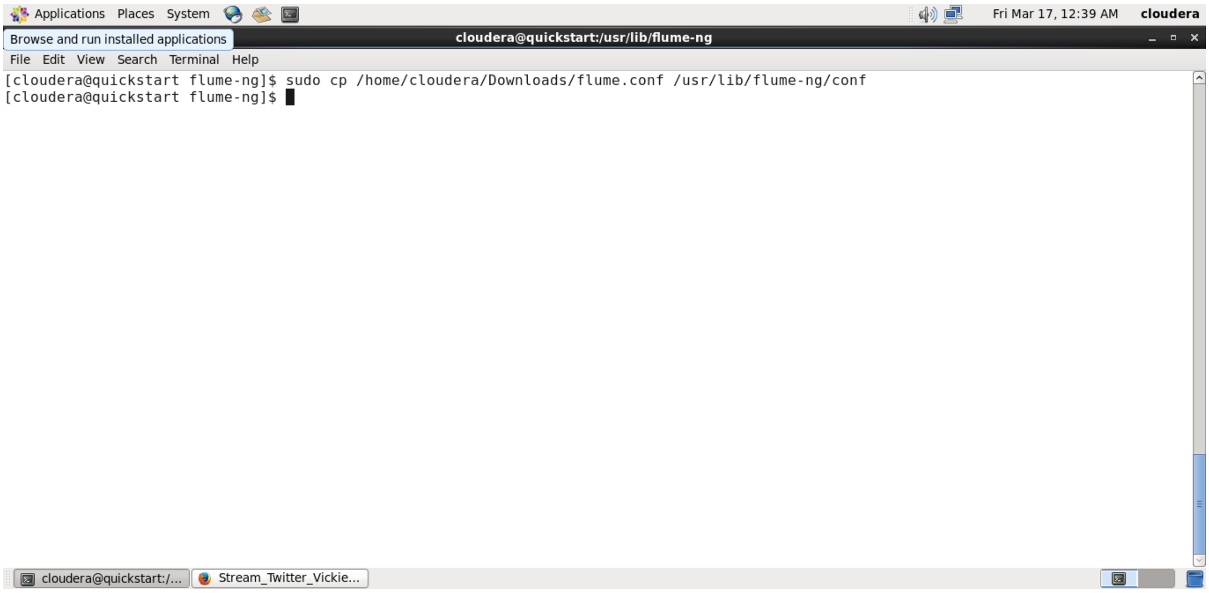
Applications Places System
Fri Mar 17, 12:38 AM cloudera
Browse and run installed applications
flume.conf - Google Drive - Mozilla Firefox
Stream_Twitter_Vicki...
Flume Installation an...
flume.conf - Google ...
https://drive.google.com/file/d/0B-Cl0IfLnRozdlRuN3pPWEJ1RHc/view
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
flume.conf
Open with Google Docs
cloudera@quickstart:/...
flume.conf - Google Dr...

```
TwitterAgent.sources= Twitter
TwitterAgent.channels= MemChannel
TwitterAgent.sinks=HDFS
TwitterAgent.sources.Twitter.type = com.cloudera.flume.source.TwitterSource
TwitterAgent.sources.Twitter.channels=MemChannel

TwitterAgent.sources.Twitter.consumerKey=Pw63cpjptT59ulP0zmT6w
TwitterAgent.sources.Twitter.consumerSecret=    n8awrhwk0ui5WOwr3NLKf7S576DcILPk5Ddfp1LQUU
TwitterAgent.sources.Twitter.accessToken=1635433267-s0NAOXmRqm5y4UC2WV7HPOuiOE9fPZZ56eWO95P
TwitterAgent.sources.Twitter.accessTokenSecret= CBKPgbJLwyJJ1jY4atf7iaiaR96Z1PmVvKF0iOXsP8E

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/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=100
```

Step 20:

Move the downloaded flume.conf file to configuration folder of flume-ng. Super user permission is given for this modification.

Command: sudo cp /home/cloudera/Downloads/flume.conf /usr/lib/flume-ng/conf/



Applications Places System
Fri Mar 17, 12:39 AM cloudera
Browse and run installed applications
cloudera@quickstart:/usr/lib/flume-ng
File Edit View Search Terminal Help
cloudera@quickstart:/...
Stream_Twitter_Vickie...

```
[cloudera@quickstart flume-ng]$ sudo cp /home/cloudera/Downloads/flume.conf /usr/lib/flume-ng/conf
[cloudera@quickstart flume-ng]$ 
```

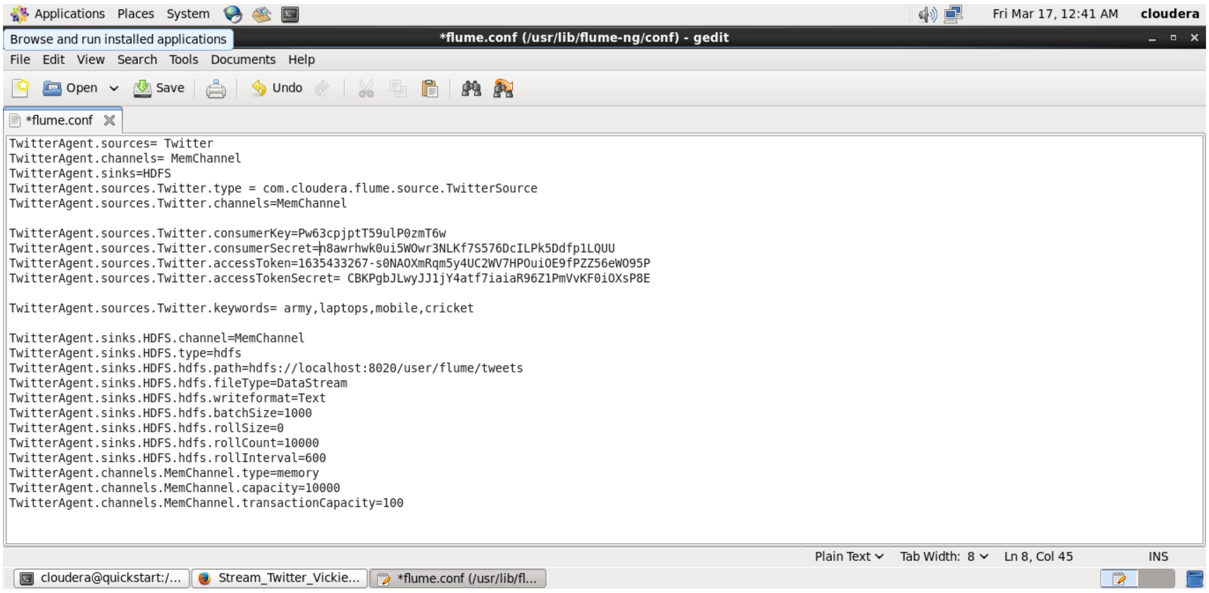
Step 21:

Open flume.conf file using gedit to make changes in the access credentials.

Command: sudo gedit conf/flume.conf

Step 22:

Replace the TwitterAgent.sources.Twitter keys and tokens with the one provided by the twitter in the application Keys and Access tokens steps. Replace the keywords of TwitterAgent.sources.Twitter.keywords to the one which we are analyzing. Also change the localhost number of TwitterAgent.sinks.HDFS.hdfs.path to the localhost number of your host. Save the file and close it.

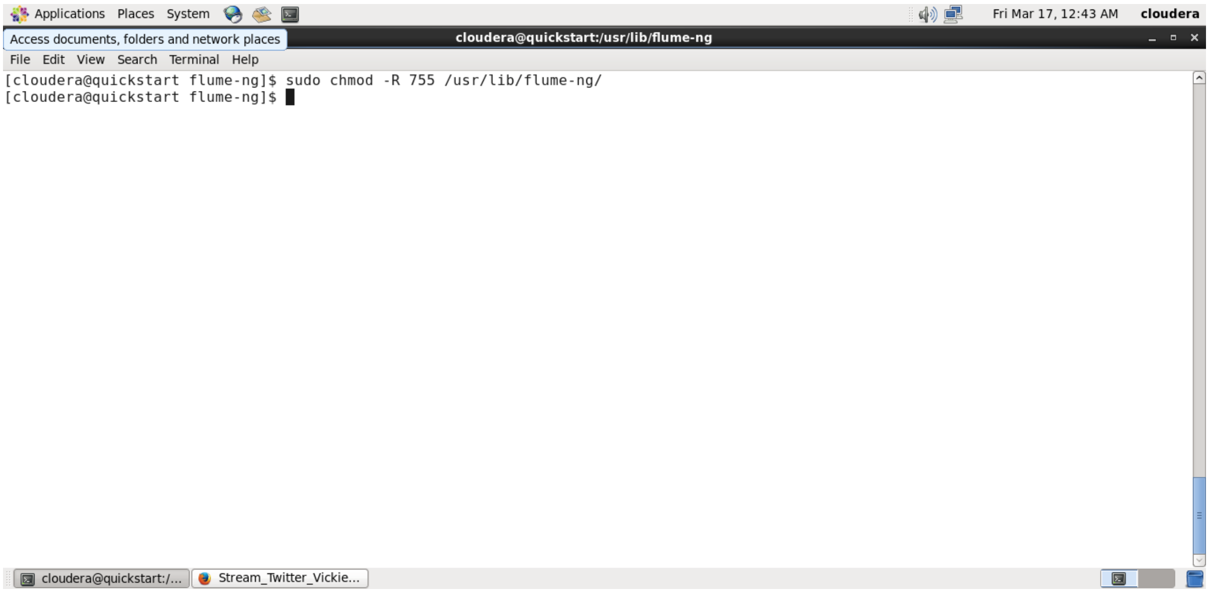

Applications Places System
Fri Mar 17, 12:41 AM cloudera
Browse and run installed applications
*flume.conf (/usr/lib/flume-ng/conf) - gedit
File Edit View Search Tools Documents Help
Open Save Undo
*flume.conf
Plain Text Tab Width: 8 Ln 8, Col 45 INS
cloudera@quickstart:/... Stream_Twitter_Vickie... *flume.conf (/usr/lib/fl...

```
TwitterAgent.sources= Twitter
TwitterAgent.channels= MemChannel
TwitterAgent.sinks=HDFS
TwitterAgent.sources.Twitter.type = com.cloudera.flume.source.TwitterSource
TwitterAgent.sources.Twitter.channels=MemChannel

TwitterAgent.sources.Twitter.consumerKey=Pw63cpjptT59ulP0zmT6w
TwitterAgent.sources.Twitter.consumerSecret=h8awrhwk0ui5WOwr3NLKf7S576DcILPk5Ddfp1LQUU
TwitterAgent.sources.Twitter.accessToken=1635433267-s0NAOXmRqm5y4UC2WV7HPOuiOE9fPZZ56eWO95P
TwitterAgent.sources.Twitter.accessTokenSecret= CBKPgbJLwyJJ1jY4atf7iaiaR96Z1PmVvKF0iOXsP8E

TwitterAgent.sources.Twitter.keywords= army,laptops,mobile,cricket

TwitterAgent.sinks.HDFS.channel=MemChannel
TwitterAgent.sinks.HDFS.type=hdfs
TwitterAgent.sinks.HDFS.hdfs.path=hdfs://localhost:8020/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=100
```

Step 23:

Change the permissions for flume directory. Super user permission is given for this modification.

Command: sudo chmod -R 755 /usr/lib/flume-ng/

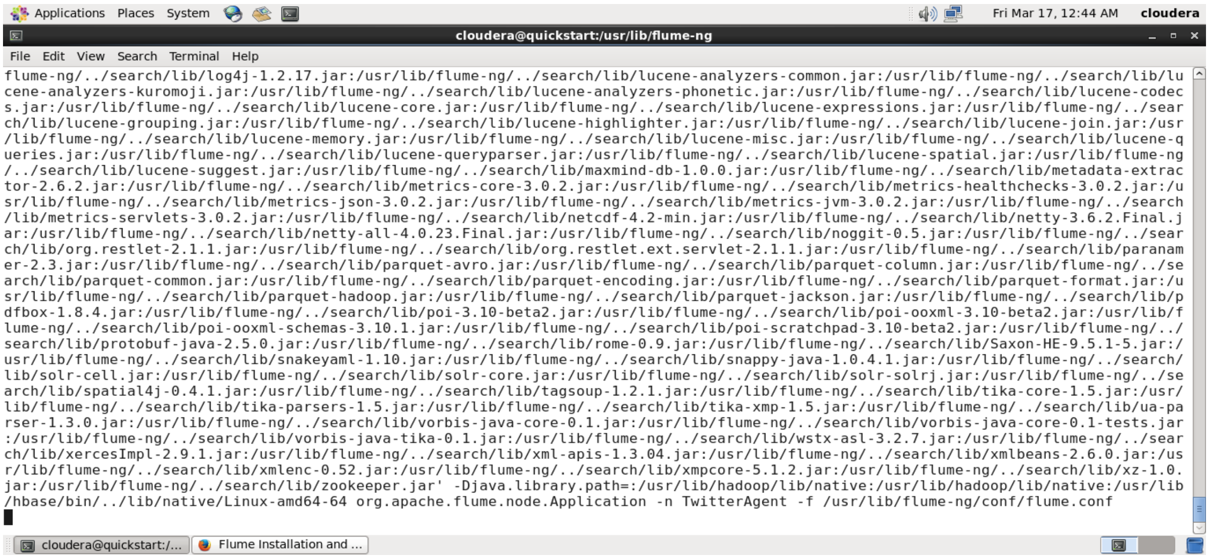

Applications Places System
Fri Mar 17, 12:43 AM cloudera
Access documents, folders and network places
cloudera@quickstart:/usr/lib/flume-ng
File Edit View Search Terminal Help
cloudera@quickstart:/... Stream_Twitter_Vickie...

```
[cloudera@quickstart flume-ng]$ sudo chmod -R 755 /usr/lib/flume-ng/
[cloudera@quickstart flume-ng]$ 
```

Step 24:

The final step is to start fetching the data from twitter.

Command: ./bin/flume-ng agent -n TwitterAgent -c conf -f /usr/lib/ flume-ng/conf/flume.conf



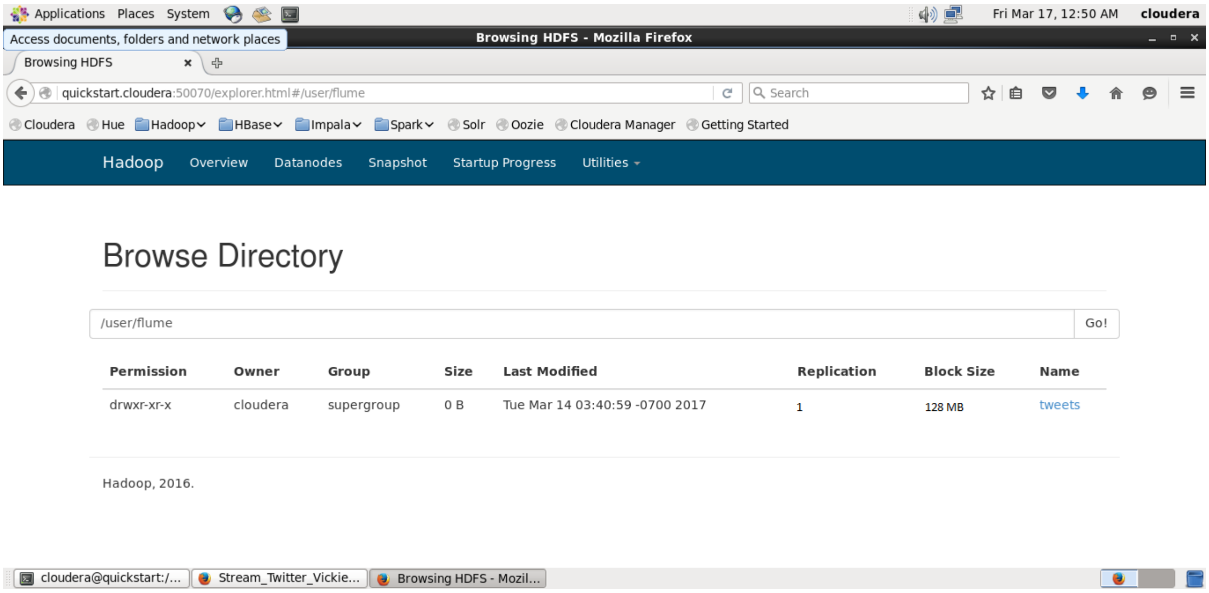
```
flume-ng/../search/lib/log4j-1.2.17.jar:/usr/lib/flume-ng/../search/lib/lucene-analyzers-common.jar:/usr/lib/flume-ng/../search/lib/lu
cene-analyzers-kuromoji.jar:/usr/lib/flume-ng/../search/lib/lucene-analyzers-phonetic.jar:/usr/lib/flume-ng/../search/lib/lucene-codec
s.jar:/usr/lib/flume-ng/../search/lib/lucene-core.jar:/usr/lib/flume-ng/../search/lib/lucene-expressions.jar:/usr/lib/flume-ng/../sear
ch/lib/lucene-grouping.jar:/usr/lib/flume-ng/../search/lib/lucene-highlighter.jar:/usr/lib/flume-ng/../search/lib/lucene-join.jar:/usr
/lib/flume-ng/../search/lib/lucene-memory.jar:/usr/lib/flume-ng/../search/lib/lucene-misc.jar:/usr/lib/flume-ng/../search/lib/lucene-q
ueries.jar:/usr/lib/flume-ng/../search/lib/lucene-queryparser.jar:/usr/lib/flume-ng/../search/lib/lucene-spatial.jar:/usr/lib/flume-ng
/../search/lib/lucene-suggest.jar:/usr/lib/flume-ng/../search/lib/maxmind-db-1.0.0.jar:/usr/lib/flume-ng/../search/lib/metadata-extrac
tor-2.6.2.jar:/usr/lib/flume-ng/../search/lib/metrics-core-3.0.2.jar:/usr/lib/flume-ng/../search/lib/metrics-healthchecks-3.0.2.jar:/u
sr/lib/flume-ng/../search/lib/metrics-json-3.0.2.jar:/usr/lib/flume-ng/../search/lib/metrics-jvm-3.0.2.jar:/usr/lib/flume-ng/../search
/lib/metrics-servlets-3.0.2.jar:/usr/lib/flume-ng/../search/lib/netcdf-4.2-min.jar:/usr/lib/flume-ng/../search/lib/netty-3.6.2.Final.j
ar:/usr/lib/flume-ng/../search/lib/netty-all-4.0.23.Final.jar:/usr/lib/flume-ng/../search/lib/noggit-0.5.jar:/usr/lib/flume-ng/../sear
ch/lib/org.restlet-2.1.1.jar:/usr/lib/flume-ng/../search/lib/org.restlet.ext.servlet-2.1.1.jar:/usr/lib/flume-ng/../search/lib/paranam
er-2.3.jar:/usr/lib/flume-ng/../search/lib/parquet-avro.jar:/usr/lib/flume-ng/../search/lib/parquet-column.jar:/usr/lib/flume-ng/../se
arch/lib/parquet-common.jar:/usr/lib/flume-ng/../search/lib/parquet-encoding.jar:/usr/lib/flume-ng/../search/lib/parquet-format.jar:/u
sr/lib/flume-ng/../search/lib/parquet-hadoop.jar:/usr/lib/flume-ng/../search/lib/parquet-jackson.jar:/usr/lib/flume-ng/../search/lib/p
dfbox-1.8.4.jar:/usr/lib/flume-ng/../search/lib/poi-3.10-beta2.jar:/usr/lib/flume-ng/../search/lib/poi-ooxml-3.10-beta2.jar:/usr/lib/f
lume-ng/../search/lib/poi-ooxml-schemas-3.10.1.jar:/usr/lib/flume-ng/../search/lib/poi-scratchpad-3.10-beta2.jar:/usr/lib/flume-ng/../
search/lib/protobuf-java-2.5.0.jar:/usr/lib/flume-ng/../search/lib/rome-0.9.jar:/usr/lib/flume-ng/../search/lib/Saxon-HE-9.5.1-5.jar:/
usr/lib/flume-ng/../search/lib/snakeyaml-1.10.jar:/usr/lib/flume-ng/../search/lib/snappy-java-1.0.4.1.jar:/usr/lib/flume-ng/../search/
lib/solr-cell.jar:/usr/lib/flume-ng/../search/lib/solr-core.jar:/usr/lib/flume-ng/../search/lib/solr-solrj.jar:/usr/lib/flume-ng/../se
arch/lib/spatial4j-0.4.1.jar:/usr/lib/flume-ng/../search/lib/tagsoup-1.2.1.jar:/usr/lib/flume-ng/../search/lib/tika-core-1.5.jar:/usr/
lib/flume-ng/../search/lib/tika-parsers-1.5.jar:/usr/lib/flume-ng/../search/lib/tika-xmp-1.5.jar:/usr/lib/flume-ng/../search/lib/ua-pa
rser-1.3.0.jar:/usr/lib/flume-ng/../search/lib/vorbis-java-core-0.1.jar:/usr/lib/flume-ng/../search/lib/vorbis-java-core-0.1-tests.jar
:/usr/lib/flume-ng/../search/lib/vorbis-java-tika-0.1.jar:/usr/lib/flume-ng/../search/lib/wstx-asl-3.2.7.jar:/usr/lib/flume-ng/../sear
ch/lib/xercesImpl-2.9.1.jar:/usr/lib/flume-ng/../search/lib/xml-apis-1.3.04.jar:/usr/lib/flume-ng/../search/lib/xmlbeans-2.6.0.jar:/us
r/lib/flume-ng/../search/lib/xmlenc-0.52.jar:/usr/lib/flume-ng/../search/lib/xmpcore-5.1.2.jar:/usr/lib/flume-ng/../search/lib/xz-1.0.
jar:/usr/lib/flume-ng/../search/lib/zookeeper.jar' -Djava.library.path=:/usr/lib/hadoop/lib/native:/usr/lib/hadoop/lib/native:/usr/lib
/hbase/bin/../lib/native/Linux-amd64-64 org.apache.flume.node.Application -n TwitterAgent -f /usr/lib/flume-ng/conf/flume.conf
█
```

Applications Places System
Fri Mar 17, 12:44 AM
cloudera
cloudera@quickstart:/usr/lib/flume-ng
File Edit View Search Terminal Help
cloudera@quickstart:/...
Flume Installation and ...

Now the live data is streamed from Twitter. Wait for about 50 to 60 seconds to fetch the streaming data from twitter. Ctrl+C is pressed to terminate the streaming. The system may throw few exceptions as the process is terminated in the middle but ignore it.

Step 25:

Now the browser is opened in the the VM machine. Locate the tweets data path in HDFS at /user/flume/tweets.



Applications Places System
Fri Mar 17, 12:50 AM
cloudera
Access documents, folders and network places
Browsing HDFS - Mozilla Firefox
Browsing HDFS
quickstart.cloudera:50070/explorer.html#/user/flume
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Hadoop Overview Datanodes Snapshot Startup Progress Utilities
Go!
cloudera@quickstart:/...
Stream_Twitter_Vickie...
Browsing HDFS - Mozil...

Browse Directory

/user/flume

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	cloudera	supergroup	0 B	Tue Mar 14 03:40:59 -0700 2017	1	128 MB	tweets

Hadoop, 2016.

Step 26:

The data is yielded as shown in the snapshot and modified according to demands of the user.

Conclusion:

Thus the Twitter data data is successfully streamed from the Twitter using Apache Flume engine.