**Hadoop Applications:**

Initially we used Apache Hadoop and it's Eco-systems to build the following applications, all the ouputs were stored into Mysql.

Developed the following hadoop applications using various Eco-systems mentioned below.

Channel D/type Data Using HDFS Processing output RDBMS

======= ====== ========= ==== ======== ==== ======

LOG LOG Local Files files Hive Table Mysql

XML XML Local Files xml Pig & Spark File Mysql

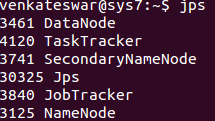
Twitter JSON Flume Json Hive Table Mysql

**HADOOP TWITTER APPLICATION:**

**Tools Used: (Flume, HDFS, Hive, Sqoop, MySQL)**

1. Hadoop Availability
2. Twitter Application
3. Flume Configuration
4. Hadoop Distributed File System (HDFS)
5. Hive Table
6. Mysql Data
7. Hadoop Availability:

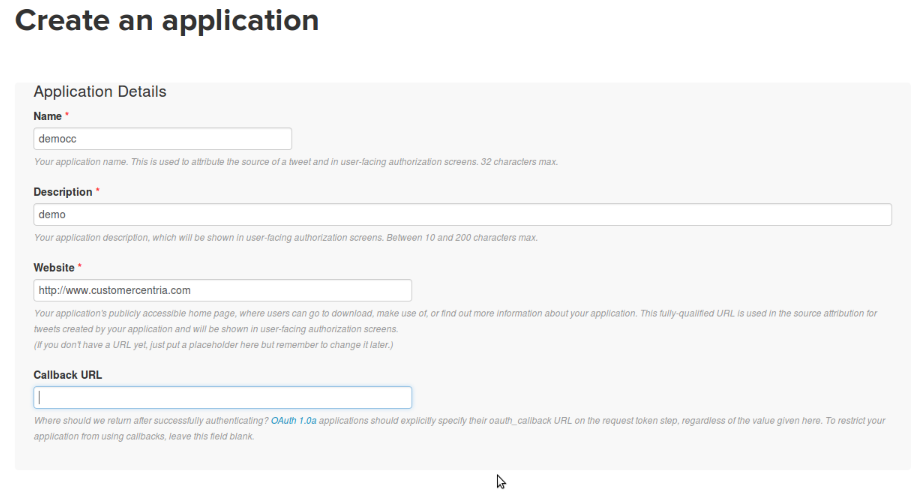
Hadoop must be up and running. Verify it by typing jps command in the terminal.and check if all 5 Hadoop daemons are running.



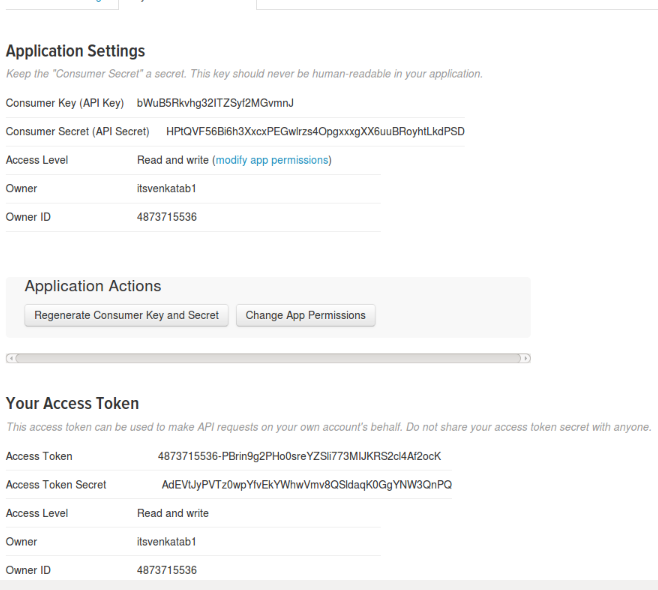
2. Twitter Application:

1. Go to <https://twitter.com/> and create a new Twitter account and log in to the account.

2. Go to <https://apps.twitter.com/> and click on create New App, fill in the below details and accept the developer agreement.



3. Go to Keys and Acces tokens tab and click on create my access token. Get the following keys Consumer Key (API Key), Consumer Secret (API Secret), Access Token, Access Token Secret

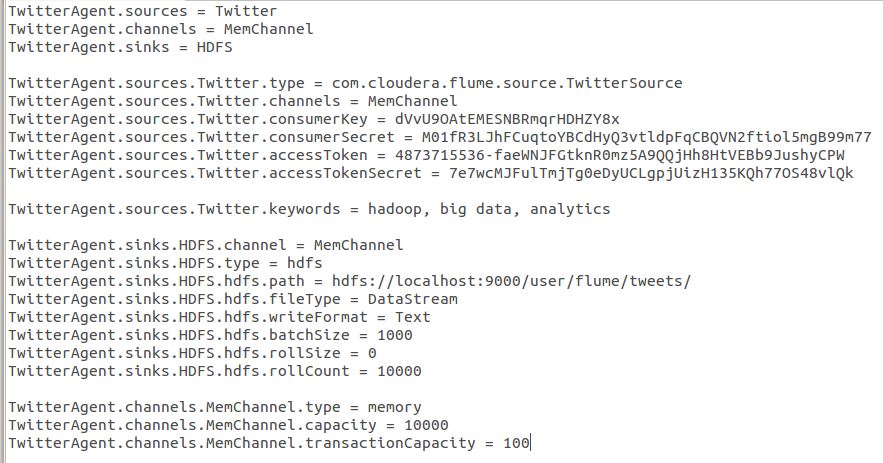


3. Flume Configuration:

1. Flume must be installed and configured . Verify it by typing echo $FLUME\_HOME, it ensures flume availability



2. Edit the following flume-twitter.conf file with the latest keys and tokens from twitter.

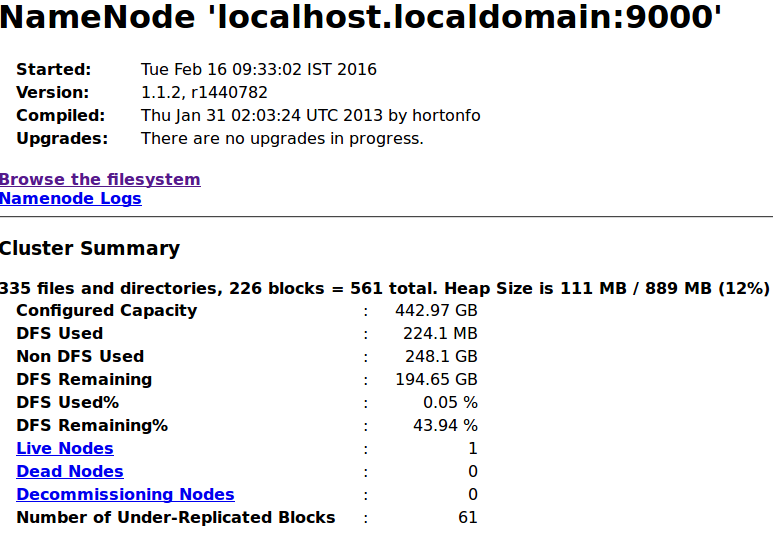


* 1. Enter the following commands in the terminal to run the flume agents.

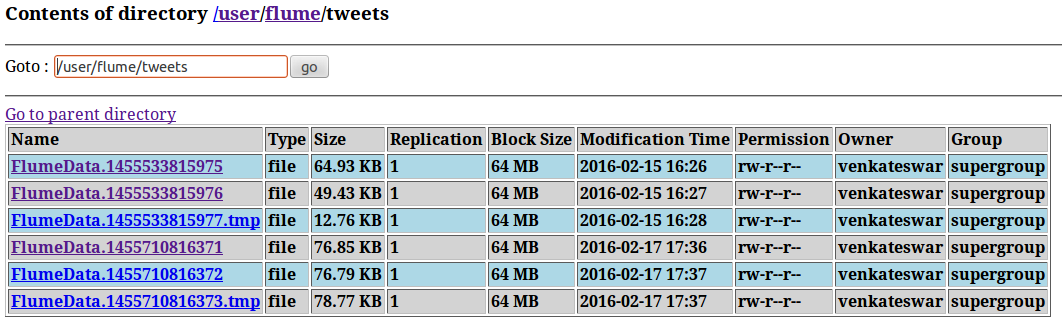


4. Flume Agents starts writing tweets into HDFS.

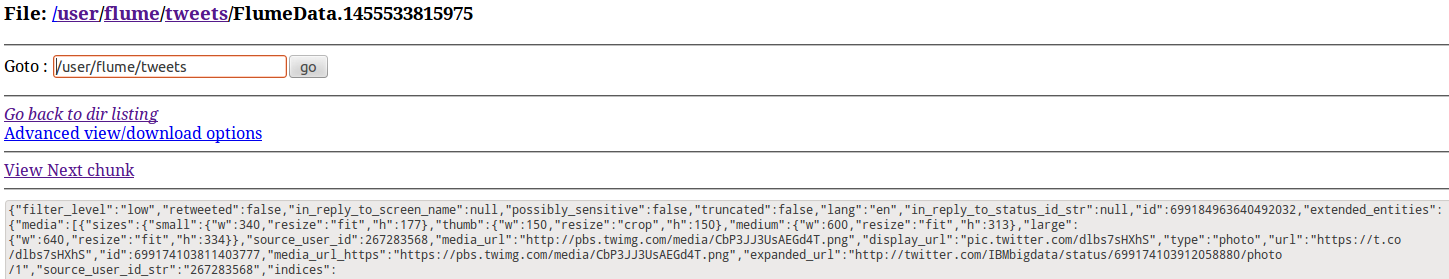
1. HDFS:
   * 1. Twitter feed will be stored in the following hdfs location.



* + 1. Tweets Location

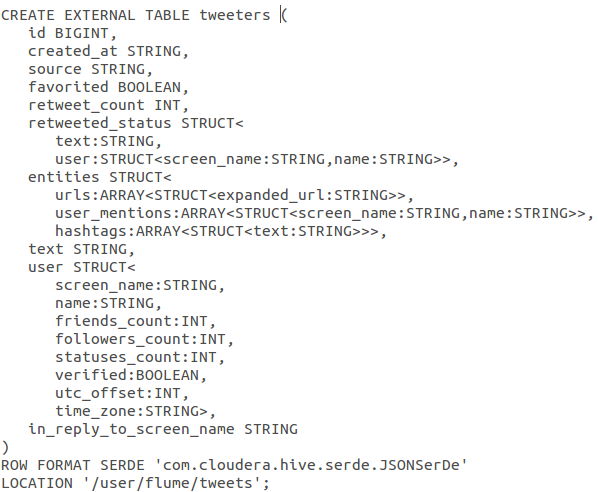


* + 1. Tweets in the form of JSON



1. Hive Table:

1. Create aHive table with the following structure to store the tweets informations.



* + 1. Select \* from tweeters2 limit 100;
    2. Store these results into HDFS by using typing the following commands in the terminal.
    3. Mysql Data:

1.

