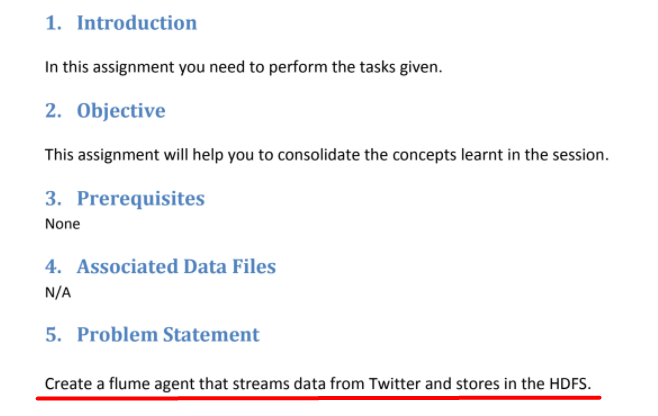
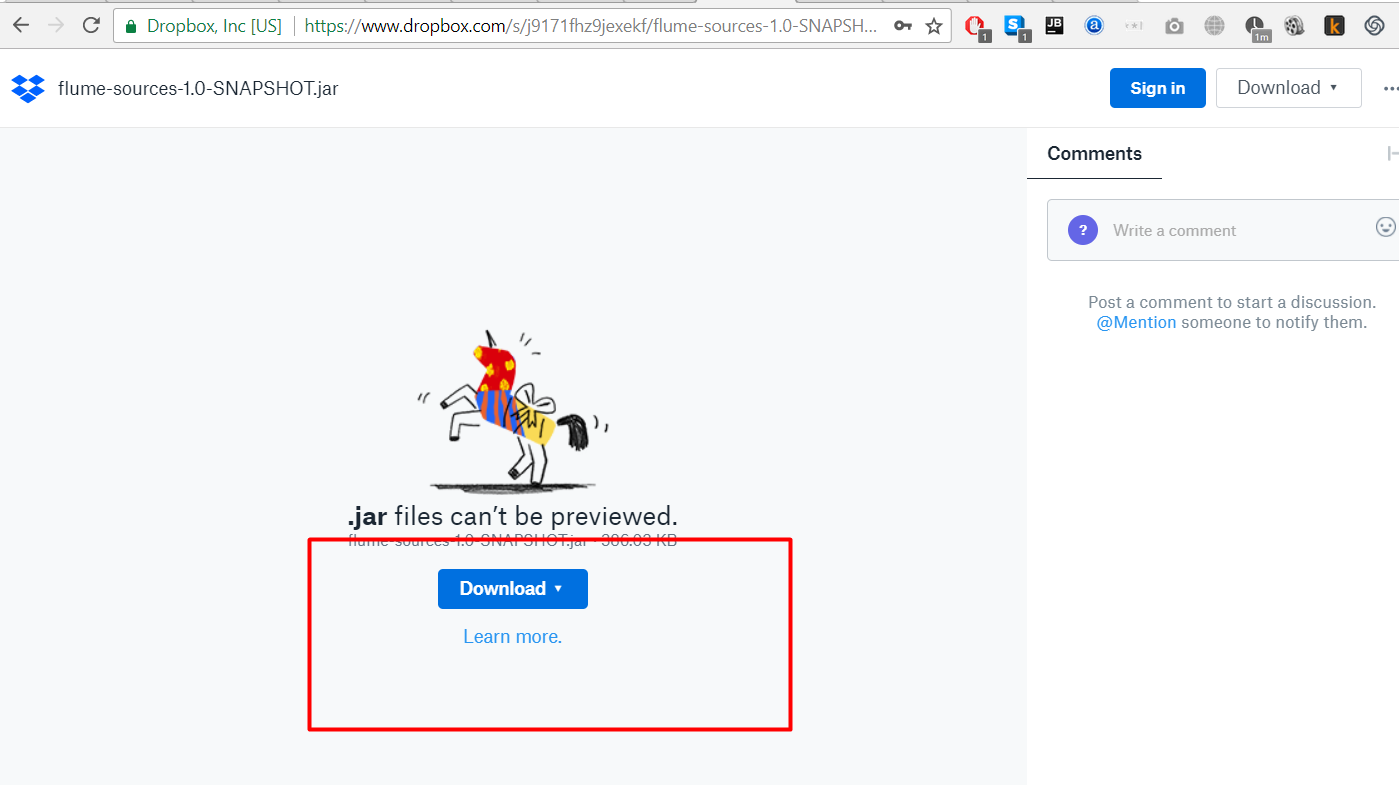


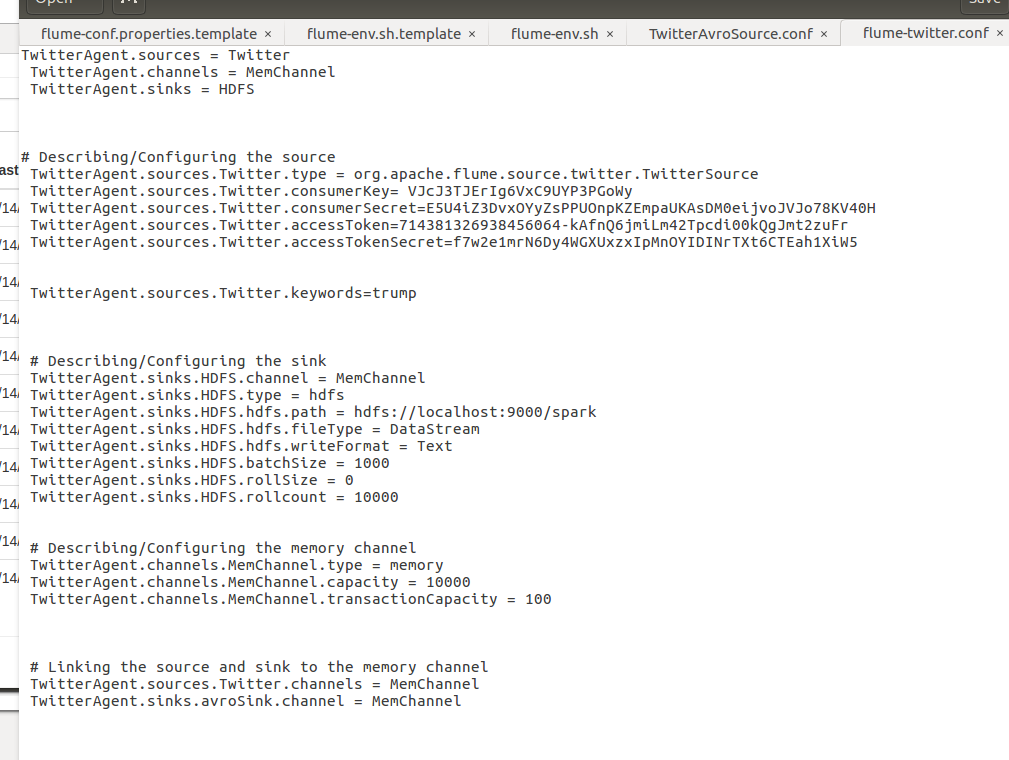
**Assignment 2**



First of all we need to download cloudera jar file for streaming data from the twitter which we can find here :

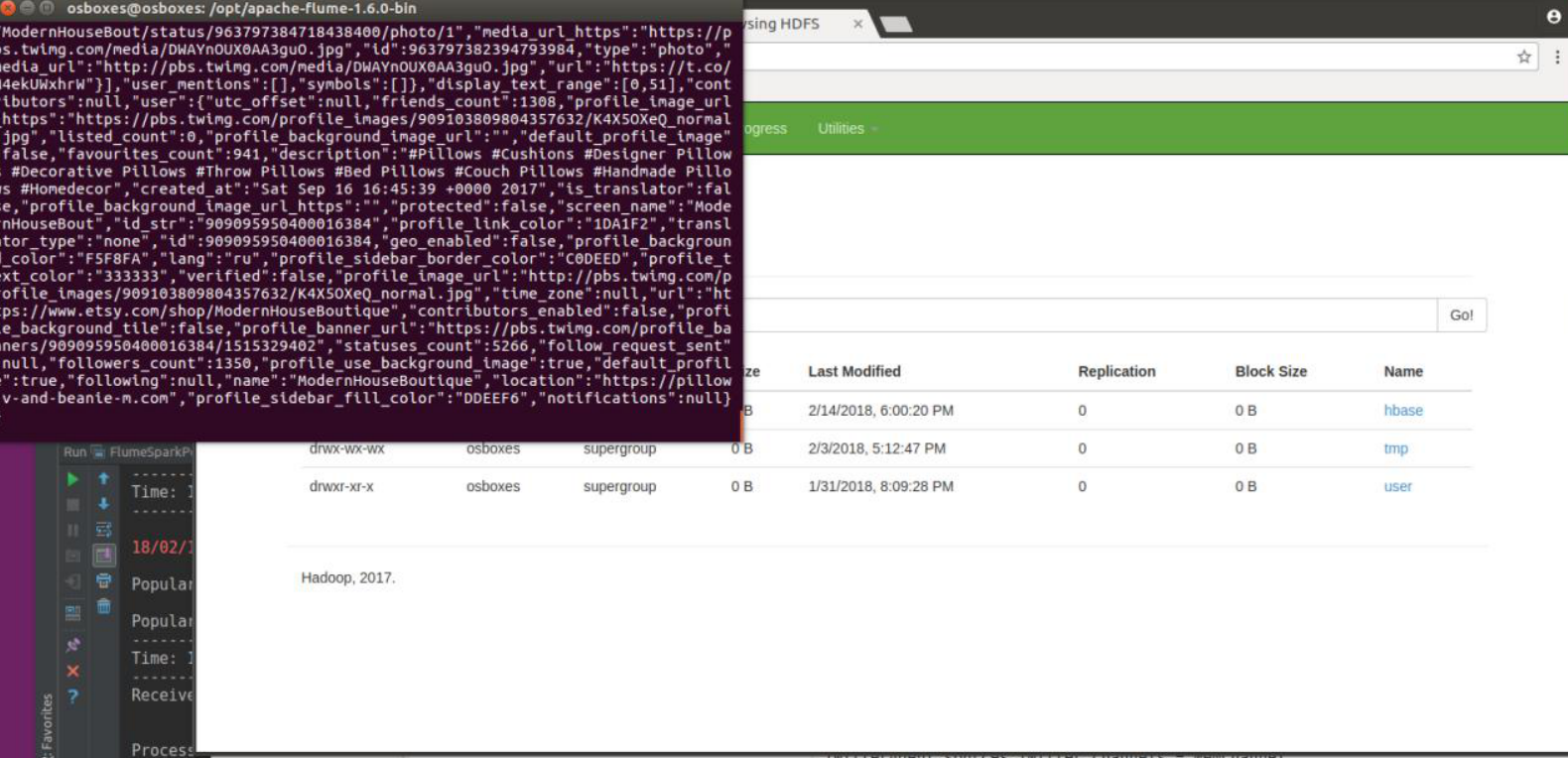


I drop the point where we install and configure environmental vars for the flume. Now we need to create config for twitter stream with sink point to hdfs folder with name spark

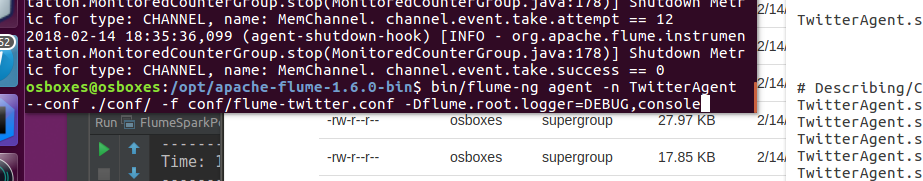


And put it under ./conf dir

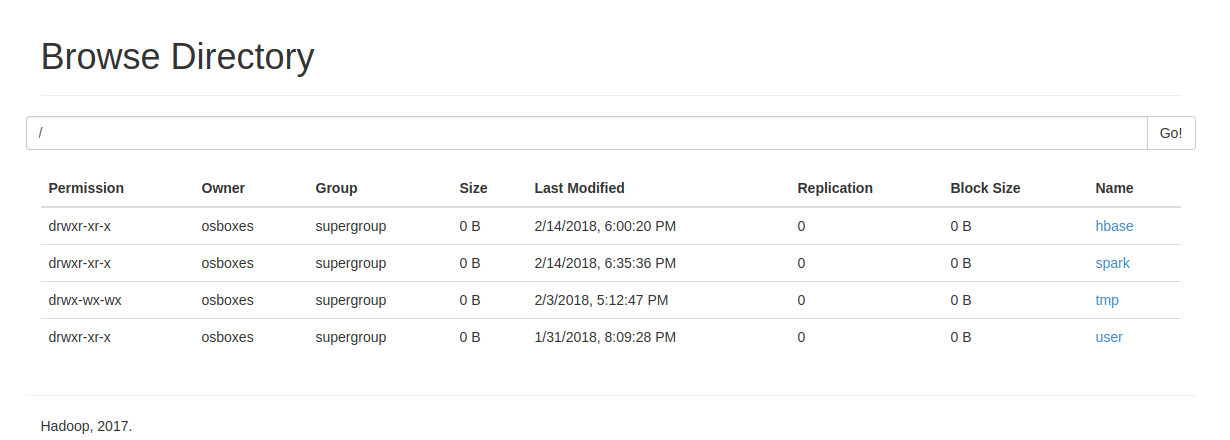
Let make sure that no spark folder exist before we run the flume



Now we can run the flume by using the command line below



As u can see the park dir was created



And the dataset which contain keyword “trump” also present in it

