Assignment 11.1

Problem Statement

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

**Flume.conf**

*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 = O1B0dBZiazZuTtlcn3iD1zN8c*

*TwitterAgent.sources.Twitter.consumerSecret = dJ0gy05ZNCCOznICIC9RncVnIUOI6ajf5U4zC5tzXIrvoMvlss*

*TwitterAgent.sources.Twitter.accessToken = 2306605970-z7WpGnYkx7J8Y2BclnRpsqsCiI6EiUbq1QkHEET*

*TwitterAgent.sources.Twitter.accessTokenSecret = Lr31L0EwKhCAnfMP3ZMSHtkVmaFFPH0rLYYPbDik6XS7d*

*TwitterAgent.sources.Twitter.keywords = cricket*

*TwitterAgent.sinks.HDFS.channel = MemChannel*

*TwitterAgent.sinks.HDFS.type = hdfs*

*TwitterAgent.sinks.HDFS.hdfs.path = hdfs://localhost/user/ec2-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.channels.MemChannel.type = memory*

*TwitterAgent.channels.MemChannel.capacity = 10000*

*TwitterAgent.channels.MemChannel.transactionCapacity = 100*

**Flume agent start**

flume-ng agent --conf-file /usr/lib/flume-ng/conf/flume.conf Dflume.root.logger=DEBUG,console -n TwitterAgent

**data streamed to hdfs**



