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

***SOLUTION :***

**1)Modify usr/local/flume/conf/flume.conf**    
  
[acadgild@localhost conf]$ cat flume.conf   
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=uX0TWqkx0okYEjjqLzxIx6mD6  
TwitterAgent.sources.Twitter.consumerSecret=rzHIs3TMJnADbZNvdGU7LQUo0kPxPISq3RGSLfqcBip39X5END  
TwitterAgent.sources.Twitter.accessToken=559516596-yDA9xqOljo4CV32wSnqsx2BXh4RBIRKFxZGSZrPC  
TwitterAgent.sources.Twitter.accessTokenSecret=zDxePILZitS5tIWBhre0GWqps0FIj9OadX8RZb6w8ZCwz  
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/user/flume/tweetme  
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

**2)create the HDFS Directory:**

[acadgild@localhost conf]$ hadoop fs -mkdir /user/flume/tweetme

**3)Launch the flume agent:**  
  
flume-ng agent -n TwitterAgent -f /usr/local/flume/conf/flume.conf

**OUTPUT:**  
17/04/25 23:26:59 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: HDFS started  
17/04/25 23:27:05 INFO twitter4j.TwitterStreamImpl: Connection established.  
17/04/25 23:27:05 INFO twitter4j.TwitterStreamImpl: Receiving status stream.  
17/04/25 23:27:06 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false  
17/04/25 23:27:08 INFO twitter.TwitterSource: Processed 100 docs  
17/04/25 23:27:09 INFO hdfs.BucketWriter: Creating hdfs://localhost:9000/user/flume/tweetme/FlumeData.1493143026258.tmp  
17/04/25 23:27:10 INFO twitter.TwitterSource: Processed 200 docs  
17/04/25 23:27:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
17/04/25 23:27:12 INFO twitter.TwitterSource: Processed 300 docs  
17/04/25 23:27:15 INFO twitter.TwitterSource: Processed 400 docs  
17/04/25 23:27:17 INFO twitter.TwitterSource: Processed 500 docs  
  
**4)**

**Check HDFS path for the twitter stream**

[acadgild@localhost ~]$ hadoop fs -ls /user/flume/tweetme  
17/04/25 23:28:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
Found 1 items  
-rw-r--r--   1 acadgild supergroup     876637 2017-04-25 23:28 /user/flume/tweetme/FlumeData.1493143026258