

Big Data and Hadoop Development

1. Introduction

In this assignment you need to perform the tasks given.

2. Objective

This assignment will help you to consolidate the concepts learnt in the session.

3. Prerequisites

None

4. Associated Data Files

N/A

5. Problem Statement

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

ApplicationAcad

Test

[Details](#)

[Settings](#)

Keys and Access Tokens

[Permissions](#)

Application Settings

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

Consumer Key (API Key) tKdfp8eLBcQ5x5XdBbL4Qqm1A

Consumer Secret (API Secret) 6gBtGapuES09TpZbu9MclrqkOcekeOVQWhd40BiERwjF6lBrTo

Access Level Read and write ([modify app permissions](#))

Owner manjudharani

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	798803756305367040-10bvmGOE3u43pqzNSN0TSZUqgOZ3bXV
--------------	--

Access Token Secret	JoJEiYHsRIgonllqq3VTxGorsMsuHkcjXUrA5wCZay2Qw
---------------------	---

Access Level	Read and write
--------------	----------------

Owner	manjudharani
-------	--------------

Owner ID	798803756305367040
----------	--------------------

```
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-yDA9xq0ljo4CV32wSnqsx2BXh4RBIRKfxZGSZrPC
TwitterAgent.sources.Twitter.accessTokenSecret=zDxePILZitS5tIWbhre0GWqps0FIj90adX8RZb6w8ZCwz
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/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
```

```
17/12/03 00:14:42 INFO twitter.TwitterSource: Processed 31,000 docs
17/12/03 00:14:42 INFO twitter.TwitterSource: Total docs indexed: 31,000, total
skipped docs: 0
17/12/03 00:14:42 INFO twitter.TwitterSource: 27 docs/second
17/12/03 00:14:42 INFO twitter.TwitterSource: Run took 1,121 seconds and process
ed:
17/12/03 00:14:42 INFO twitter.TwitterSource: 0.007 MB/sec sent to index
17/12/03 00:14:42 INFO twitter.TwitterSource: 8.238 MB text sent to index
17/12/03 00:14:42 INFO twitter.TwitterSource: There were 0 exceptions ignored:
```