**CS 5433: Bigdata Management**

**Programming Assignment 1**

**PART 1 - Flume Configuration files for data download**

**CWID: A20343337**

## 1. NASA data configuration file

```
# Naming the components on the current agent

TwitterAgent.sources = Twitter
TwitterAgent.channels = MemChannel
TwitterAgent.sinks = HDFS

# Describing/Configuring the source

TwitterAgent.sources.Twitter.type = com.cloudera.flume.source.TwitterSource  #This downloads the data in Json format
TwitterAgent.sources.Twitter.consumerKey = 1ohmbRhAbTzjtZUmZlTerIhLK #Consumer API key obtained from twitter developer account
TwitterAgent.sources.Twitter.consumerSecret = PgZ96EnGvf0aHK7MOuzQl5zCtwIIiN9sOxIAduHtnhKQDTKKnq #Consumer API token obtained from twitter developer account
TwitterAgent.sources.Twitter.accessToken = 1491145396961304586-qZvpaB3VVkmZs7qvt1WF8NC4dd3diC #Access key obtained from twitter developer account
TwitterAgent.sources.Twitter.accessTokenSecret = zcsNfEqdcaPGEkI6IvyRpVcX0t5cNwRdjEjuWPKOR9vg4 #Access token obtained from twitter developer account
TwitterAgent.sources.Twitter.keywords = NASA #keyword used for downloading the data from twitter

# Describing/Configuring the sink

TwitterAgent.sinks.HDFS.type = hdfs
TwitterAgent.sinks.HDFS.hdfs.path = hdfs://hadoop-nn001.cs.okstate.edu:9000/user/sdarapu/NASA_PA1data/%Y/%m/%d/%H #Specified path in which the streamed data gets downloaded.
TwitterAgent.sinks.HDFS.hdfs.useLocalTimeStamp = true
TwitterAgent.sinks.HDFS.hdfs.fileType = DataStream
TwitterAgent.sinks.HDFS.hdfs.writeFormat = Text
TwitterAgent.sinks.HDFS.hdfs.batchSize = 100
TwitterAgent.sinks.HDFS.hdfs.rollSize = 0
TwtterAgent.sinks.HDFS.hdfs.rollCount = 0
```

```
# Describing/Configuring the channel

TwitterAgent.channels.MemChannel.type = memory
TwitterAgent.channels.MemChannel.capacity = 10000
TwitterAgent.channels.MemChannel.transactionCapacity = 10000

# Binding the source and sink to the channel

TwitterAgent.sources.Twitter.channels = MemChannel
TwitterAgent.sinks.HDFS.channel = MemChannel
```

## 2. SpaceX data configuration file

```
# Naming the components on the current agent

TwitterAgent.sources = Twitter
TwitterAgent.channels = MemChannel
TwitterAgent.sinks = HDFS

# Describing/Configuring the source

TwitterAgent.sources.Twitter.type = com.cloudera.flume.source.TwitterSource  #This downloads the data in Json format
TwitterAgent.sources.Twitter.consumerKey = 3RJWtUQCrasVyMKZboejqB3dC #Consumer API key obtained from twitter developer account
TwitterAgent.sources.Twitter.consumerSecret = 6U4hyfBrf2gH26TXv0ims8GnQBh1kPsqabNlmsVj01Dr44a5Kf #Consumer API token obtained from twitter developer account
TwitterAgent.sources.Twitter.accessToken = 1491145396961304586-dGcDqkJ3lTR5x33DFLywzXgCGQOmXJ #Access key obtained from twitter developer account
TwitterAgent.sources.Twitter.accessTokenSecret = WZMrcmZjU7g7QJ5cNh2D8dU2S9qb7FhZH7Q5O5g7IC6jI #Access token obtained from twitter developer account
TwitterAgent.sources.Twitter.keywords = SpaceX  #keyword used for downloading the data from twitter

# Describing/Configuring the sink

TwitterAgent.sinks.HDFS.type = hdfs
TwitterAgent.sinks.HDFS.hdfs.path = hdfs://hadoop-nn001.cs.okstate.edu:9000/user/sdarapu/SpaceX_PA1data/%Y/%m/%d/%H  #Specified path in which the streamed data gets downloaded.
TwitterAgent.sinks.HDFS.hdfs.useLocalTimeStamp = true
TwitterAgent.sinks.HDFS.hdfs.fileType = DataStream
TwitterAgent.sinks.HDFS.hdfs.writeFormat = Text
TwitterAgent.sinks.HDFS.hdfs.batchSize = 100
TwitterAgent.sinks.HDFS.hdfs.rollSize = 0
TwtterAgent.sinks.HDFS.hdfs.rollCount = 0


# Describing/Configuring the channel

TwitterAgent.channels.MemChannel.type = memory
TwitterAgent.channels.MemChannel.capacity = 10000
TwitterAgent.channels.MemChannel.transactionCapacity = 10000

# Binding the source and sink to the channel

TwitterAgent.sources.Twitter.channels = MemChannel
TwitterAgent.sinks.HDFS.channel = MemChannel
```