CS 5433: Bigdata Management
Programming Assignment 1
Report for PART 1

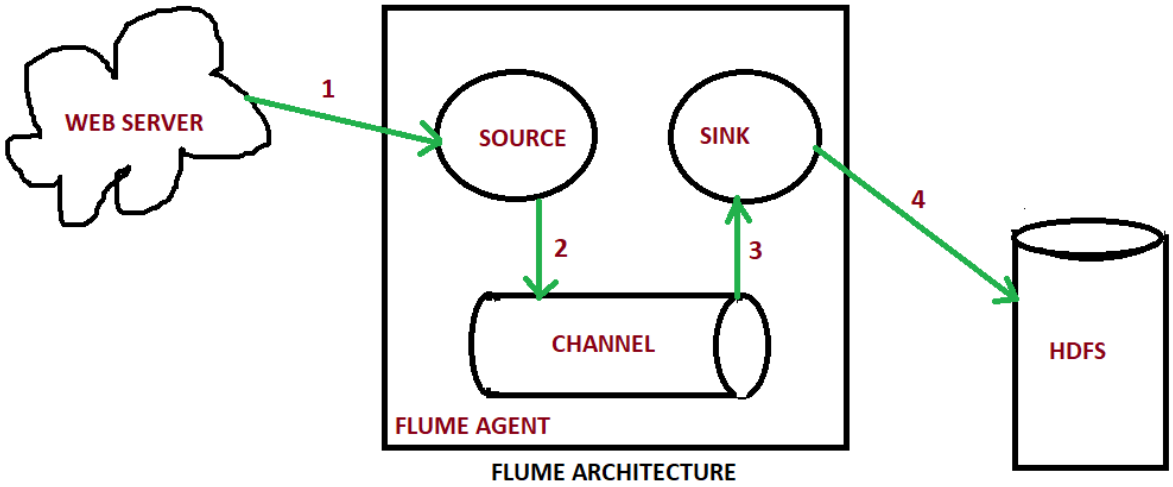
CWID: A20343337

Hadoop: It is also called as Apache Hadoop. It is an open-source framework which is used to store and process the large size of datasets (like size ranges from gigabytes to petabytes) efficiently. It mainly contains four modules,

a) Hadoop File System (HDFS)
b) Yet Another Resource Negotiator (YARN)
c) MapReduce
d) Hadoop Common

In PART1, we mainly use the HDFS which is a distributed file system that runs on low-end hardware. HDFS outperforms traditional file systems in terms of data performance, fault tolerance, and native support for huge datasets.

Flume: Flume is a distributed and reliable service for rapidly collecting, aggregating, and transporting massive amounts of log data. It is also called as Apache flume. It is an open-source software which is used to store the streamed data into HDFS. It has a simple and adaptable architecture based on the streamed data flows. Below is the architecture

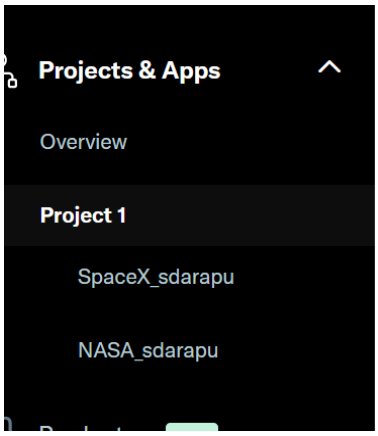


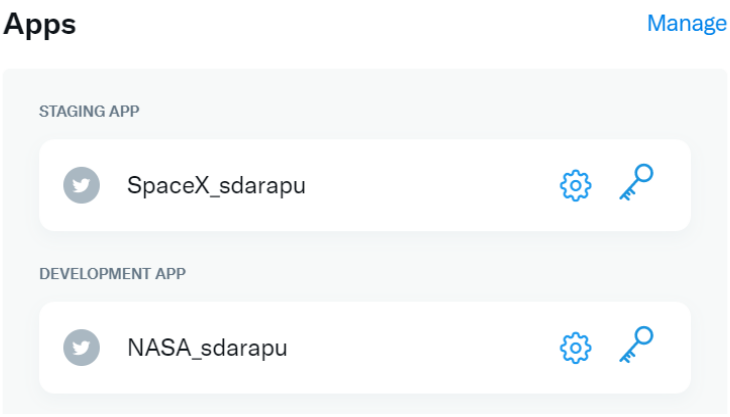
FLUME ARCHITECTURE

Approach:

Below are the steps, I have used to complete the PART 1 of the assignment

1. To collect the data from twitter using flume, I have selected two keywords “NASA” and “SpaceX”.
2. After selecting the keywords, I have filed a twitter application for developer account which is an elevated access. Within couple of days, my request for elevated access is approved.
3. Then I have created two Apps named “NASA_sdarapu” and “Spacex_sdarapu” to get my keys and tokens.





4. Now, I logged into the department’s Hadoop cluster (“hadoop-nn001.cs.okstate.edu) by using my credentials.

OpenSSH SSH client

```
Microsoft Windows [Version 10.0.19042.1526]
(c) Microsoft Corporation. All rights reserved.

C:\Users\darap>ssh sdarapu@hadoop-nn001.cs.okstate.edu
sdarapu@hadoop-nn001.cs.okstate.edu's password:
Welcome to Ubuntu 20.04.3 LTS (GNU/Linux 5.4.0-91-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

  System information as of Sat 26 Feb 2022 10:24:56 AM CST

  System load:  0.06               Processes:             335
  Usage of /:   3.6% of 1006.92GB  Users logged in:       3
  Memory usage: 48%                IPv4 address for ens3: 192.168.122.2
  Swap usage:   9%

Last login: Wed Feb 23 13:47:34 2022 from 10.200.192.236
sdarapu@hadoop-nn001:~$
```

5. After that, I have created two new directories “NASA_PA1data” and “SpaceX_PA1data” using the below command where the twitter data gets download for the keywords “NASA” and “SpaceX”.

For NASA_PA1data:

```
sdarapu@hadoop-nn001:~$ hdfs dfs -mkdir /user/sdarapu/NASA_PA1data
2022-02-26 10:26:11,640 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
sdarapu@hadoop-nn001:~$
```

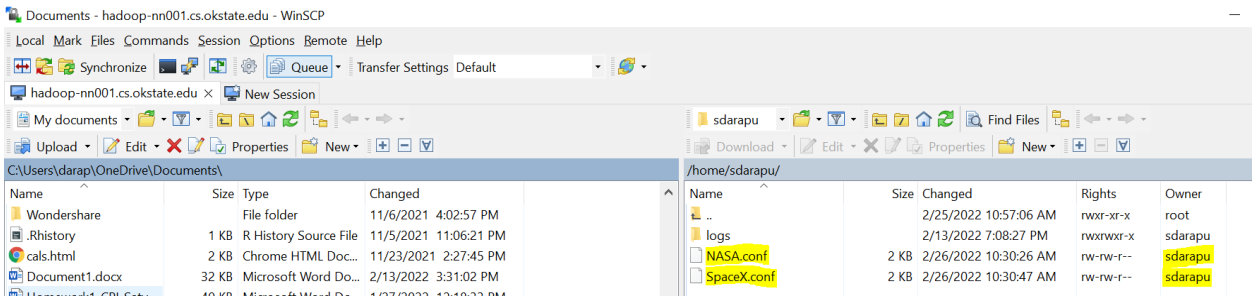
For SpaceX_PA1data:

```
sdarapu@hadoop-nn001:~$ hdfs dfs -mkdir /user/sdarapu/SpaceX_PA1data
2022-02-26 10:27:15,744 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
sdarapu@hadoop-nn001:~$
```

6. By using the “hdfs dfs -ls” command, I have checked whether the directories are created or not.

```
sdarapu@hadoop-nn001:~$ hdfs dfs -ls
2022-02-26 10:28:12,616 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
Found 2 items
drwxr-xr-x   - sdarapu sdarapu          0 2022-02-26 10:26 NASA_PA1data
drwxr-xr-x   - sdarapu sdarapu          0 2022-02-26 10:27 SpaceX_PA1data
sdarapu@hadoop-nn001:~$
```

7. Once directories are created, I have created two configuration files named “NASA.conf” and “SpaceX.conf” for connecting the twitter and flume agent directly in “WinSCP” instead of using “nano” command as shown below.



For NASA.conf:

/home/sdarapu/NASA.conf - hadoop-nn001.cs.okstate.edu - Editor - WinSCP

Encoding Color

```
# 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
TwitterAgent.sources.Twitter.consumerKey = 1ohmbRhAbTzjtZUmZlTerIhLK
TwitterAgent.sources.Twitter.consumerSecret = PgZ96EnGvf0aHK7MOuzQl5zCtwIIiN9sOxIAduHtnhKQDTKKnq
TwitterAgent.sources.Twitter.accessToken = 1491145396961304586-qZvpaB3VVkmZs7qvt1WF8NC4dd3diC
TwitterAgent.sources.Twitter.accessTokenSecret = zcsNfEqdcaPGEkI6IvyRpVcX0t5cNwRdjEjuWPKOR9vg4
TwitterAgent.sources.Twitter.keywords = NASA

# 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
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
```

For SpaceX.conf:

/home/sdarapu/SpaceX.conf - hadoop-nn001.cs.okstate.edu - Editor - WinSCP

Encoding Color

```
# 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
TwitterAgent.sources.Twitter.consumerKey = 3RJWtUQCrasVyMKZboejqB3dC
TwitterAgent.sources.Twitter.consumerSecret = 6U4hyfBrf2gH26TXv0ims8GnQBh1kPsqabNlmsVj01Dr44a5Kf
TwitterAgent.sources.Twitter.accessToken = 1491145396961304586-dGcDqkJ3lTR5x33DFLywzXgCGQOmXJ
TwitterAgent.sources.Twitter.accessTokenSecret = WZMrcmZjU7g7QJ5cNh2D8dU2S9qb7FhZH7Q5O5g7IC6jI
TwitterAgent.sources.Twitter.keywords = SpaceX

# 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
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
```

8. Once conf files are created, I have run the “nohup” command to download the twitter data into the respective directories specified in the conf files. Also, note down the process id’s to kill once the enough data is downloaded later.
   Like, data related to “NASA” keyword downloads into the directory “NASA_PA1data” and data related to “SpaceX” keyword downloads into the directory “SpaceX_PA1data”.

   For NASA data:

```
sdarapu@hadoop-nn001:~$ nohup $FLUME_HOME/bin/flume-ng agent -n TwitterAgent -f /home/sdarapu/NASA.conf --conf /usr/local/flume/conf &
[1] 367095
sdarapu@hadoop-nn001:~$ nohup: ignoring input and appending output to 'nohup.out'
```

   Process Id: 367095 → For nohup process of NASA data

   For SpaceX data:

```
sdarapu@hadoop-nn001:~$ nohup $FLUME_HOME/bin/flume-ng agent -n TwitterAgent -f /home/sdarapu/SpaceX.conf --conf /usr/local/flume/conf &
[1] 369858
sdarapu@hadoop-nn001:~$ nohup: ignoring input and appending output to 'nohup.out'
```

   Process Id: 369858 → For nohup process of SpaceX data

9. After running the nohup command, I have stopped it after waiting some time to download the enough data and then I have killed the nohup process once the enough data is downloaded from twitter. To get the process id, we can use the “ps -x” command as shown below.

   For NASA data:

```
sdarapu@hadoop-nn001:~$ ps -x
    PID TTY      STAT   TIME COMMAND
 365215 ?        Ss     0:00 /lib/systemd/systemd --user
 365218 ?        S      0:00 (sd-pam)
 365225 ?        Ssl    0:00 /usr/bin/pulseaudio --daemonize=no --log-target=journal
 365244 ?        Ss     0:00 /usr/bin/dbus-daemon --session --address=systemd: --nofork --nopidfile --systemd-activation --
 365302 ?        S      0:00 sshd: sdarapu@notty
 365303 ?        Ss     0:00 /usr/lib/openssh/sftp-server
 365928 ?        R      0:00 sshd: sdarapu@pts/5
 365930 pts/5    Ss     0:00 -bash
 366625 ?        S      0:00 sshd: sdarapu@notty
 366626 ?        Ss     0:00 /usr/lib/openssh/sftp-server
 367095 pts/5    Sl     0:38 /usr/bin/java -Xmx60m -Dtwitter4j.streamBase.URL=https://stream.twitter.com/1.1/ -cp /usr/loca
 368773 pts/5    R+     0:00 ps -x
sdarapu@hadoop-nn001:~$
```

```
sdarapu@hadoop-nn001:~$ kill -9 367095
sdarapu@hadoop-nn001:~$
```

   For SpaceX data:

```
sdarapu@hadoop-nn001:~$ ps -x
    PID TTY      STAT   TIME COMMAND
 365215 ?        Ss     0:00 /lib/systemd/systemd --user
 365218 ?        S      0:00 (sd-pam)
 365225 ?        Ssl    0:00 /usr/bin/pulseaudio --daemonize=no --log-target=journal
 365244 ?        Ss     0:00 /usr/bin/dbus-daemon --session --address=systemd: --nofork --nopidfile --systemd-activation --syslog-only
 365302 ?        S      0:00 sshd: sdarapu@notty
 365303 ?        Ss     0:00 /usr/lib/openssh/sftp-server
 365928 ?        S      0:00 sshd: sdarapu@pts/5
 365930 pts/5    Ss     0:00 -bash
 366625 ?        S      0:00 sshd: sdarapu@notty
 366626 ?        Ss     0:00 /usr/lib/openssh/sftp-server
 369858 pts/5    Sl     0:41 /usr/bin/java -Xmx60m -Dtwitter4j.streamBase.URL=https://stream.twitter.com/1.1/ -cp /usr/local/flume/conf:
 372112 pts/5    R+     0:00 ps -x
sdarapu@hadoop-nn001:~$ kill -9 369858
sdarapu@hadoop-nn001:~$
```

10. To check the data files are downloaded in the respective folders, I have executed the below command.

For NASA data:

```
sdarapu@hadoop-nn001:~$ hdfs dfs -ls /user/sdarapu/NASA_PA1data
2022-02-26 10:51:46,112 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
Found 1 items
drwxr-xr-x   - sdarapu sdarapu          0 2022-02-26 10:34 /user/sdarapu/NASA_PA1data/2022
[1]+  Killed                  nohup $FLUME_HOME/bin/flume-ng agent -n TwitterAgent -f /home/sdarapu/NASA.conf --conf /usr/local/flume/conf
sdarapu@hadoop-nn001:~$
```

```
sdarapu@hadoop-nn001:~$ hdfs dfs -ls /user/sdarapu/NASA_PA1data/2022/02/26/10
2022-02-26 10:52:34,713 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classe
Found 73 items
-rw-r--r--   3 sdarapu sdarapu      49031 2022-02-26 10:34 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266371
-rw-r--r--   3 sdarapu sdarapu      69846 2022-02-26 10:34 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266372
-rw-r--r--   3 sdarapu sdarapu      65033 2022-02-26 10:35 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266373
-rw-r--r--   3 sdarapu sdarapu      48972 2022-02-26 10:35 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266374
-rw-r--r--   3 sdarapu sdarapu      74761 2022-02-26 10:35 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266375
-rw-r--r--   3 sdarapu sdarapu      69502 2022-02-26 10:35 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266376
-rw-r--r--   3 sdarapu sdarapu      45463 2022-02-26 10:35 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266377
-rw-r--r--   3 sdarapu sdarapu      59067 2022-02-26 10:36 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266378
-rw-r--r--   3 sdarapu sdarapu      51947 2022-02-26 10:36 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266379
-rw-r--r--   3 sdarapu sdarapu      49956 2022-02-26 10:36 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266380
-rw-r--r--   3 sdarapu sdarapu      57291 2022-02-26 10:36 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266381
-rw-r--r--   3 sdarapu sdarapu      75806 2022-02-26 10:36 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266382
-rw-r--r--   3 sdarapu sdarapu      62993 2022-02-26 10:37 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266383
-rw-r--r--   3 sdarapu sdarapu      57919 2022-02-26 10:37 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266384
-rw-r--r--   3 sdarapu sdarapu      51321 2022-02-26 10:37 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266385
-rw-r--r--   3 sdarapu sdarapu      58128 2022-02-26 10:37 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266386
-rw-r--r--   3 sdarapu sdarapu      53974 2022-02-26 10:38 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266387
-rw-r--r--   3 sdarapu sdarapu      53476 2022-02-26 10:38 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266388
-rw-r--r--   3 sdarapu sdarapu      54231 2022-02-26 10:38 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266389
-rw-r--r--   3 sdarapu sdarapu      48141 2022-02-26 10:38 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266390
-rw-r--r--   3 sdarapu sdarapu      61988 2022-02-26 10:38 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266391
-rw-r--r--   3 sdarapu sdarapu      58997 2022-02-26 10:38 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266392
-rw-r--r--   3 sdarapu sdarapu      45280 2022-02-26 10:39 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266393
-rw-r--r--   3 sdarapu sdarapu      43000 2022-02-26 10:39 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266394
-rw-r--r--   3 sdarapu sdarapu      54230 2022-02-26 10:39 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266395
-rw-r--r--   3 sdarapu sdarapu      68000 2022-02-26 10:39 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266396
-rw-r--r--   3 sdarapu sdarapu      67001 2022-02-26 10:39 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266397
-rw-r--r--   3 sdarapu sdarapu      52100 2022-02-26 10:40 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266398
-rw-r--r--   3 sdarapu sdarapu      49569 2022-02-26 10:40 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266399
-rw-r--r--   3 sdarapu sdarapu      47433 2022-02-26 10:40 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266400
-rw-r--r--   3 sdarapu sdarapu      46874 2022-02-26 10:40 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266401
-rw-r--r--   3 sdarapu sdarapu      43956 2022-02-26 10:40 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266402
-rw-r--r--   3 sdarapu sdarapu      53729 2022-02-26 10:41 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266403
-rw-r--r--   3 sdarapu sdarapu      50057 2022-02-26 10:41 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266404
-rw-r--r--   3 sdarapu sdarapu      54626 2022-02-26 10:41 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266405
```

```
-rw-r--r--   3 sdarapu sdarapu      70684 2022-02-26 10:42 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266406
-rw-r--r--   3 sdarapu sdarapu      48334 2022-02-26 10:42 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266407
-rw-r--r--   3 sdarapu sdarapu      49642 2022-02-26 10:42 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266408
-rw-r--r--   3 sdarapu sdarapu      53117 2022-02-26 10:42 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266409
-rw-r--r--   3 sdarapu sdarapu      69518 2022-02-26 10:42 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266410
-rw-r--r--   3 sdarapu sdarapu      62614 2022-02-26 10:43 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266411
-rw-r--r--   3 sdarapu sdarapu      55738 2022-02-26 10:43 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266412
-rw-r--r--   3 sdarapu sdarapu      71213 2022-02-26 10:43 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266413
-rw-r--r--   3 sdarapu sdarapu      56659 2022-02-26 10:43 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266414
-rw-r--r--   3 sdarapu sdarapu      64311 2022-02-26 10:44 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266415
-rw-r--r--   3 sdarapu sdarapu      49034 2022-02-26 10:44 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266416
-rw-r--r--   3 sdarapu sdarapu      45407 2022-02-26 10:44 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266417
-rw-r--r--   3 sdarapu sdarapu      51065 2022-02-26 10:45 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266418
-rw-r--r--   3 sdarapu sdarapu      43965 2022-02-26 10:45 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266419
-rw-r--r--   3 sdarapu sdarapu      55118 2022-02-26 10:45 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266420
-rw-r--r--   3 sdarapu sdarapu      50377 2022-02-26 10:45 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266421
-rw-r--r--   3 sdarapu sdarapu      33034 2022-02-26 10:46 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266422
-rw-r--r--   3 sdarapu sdarapu      50664 2022-02-26 10:46 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266423
-rw-r--r--   3 sdarapu sdarapu      70106 2022-02-26 10:46 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266424
-rw-r--r--   3 sdarapu sdarapu      75167 2022-02-26 10:46 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266425
-rw-r--r--   3 sdarapu sdarapu      58739 2022-02-26 10:47 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266426
-rw-r--r--   3 sdarapu sdarapu      51609 2022-02-26 10:47 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266427
-rw-r--r--   3 sdarapu sdarapu      52097 2022-02-26 10:47 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266428
-rw-r--r--   3 sdarapu sdarapu      93372 2022-02-26 10:47 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266429
-rw-r--r--   3 sdarapu sdarapu      61507 2022-02-26 10:47 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266430
-rw-r--r--   3 sdarapu sdarapu      59094 2022-02-26 10:48 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266431
-rw-r--r--   3 sdarapu sdarapu      47623 2022-02-26 10:48 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266432
-rw-r--r--   3 sdarapu sdarapu      51686 2022-02-26 10:48 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266433
-rw-r--r--   3 sdarapu sdarapu      46953 2022-02-26 10:48 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266434
-rw-r--r--   3 sdarapu sdarapu      64598 2022-02-26 10:49 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266435
-rw-r--r--   3 sdarapu sdarapu      87604 2022-02-26 10:49 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266436
-rw-r--r--   3 sdarapu sdarapu      85854 2022-02-26 10:49 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266437
-rw-r--r--   3 sdarapu sdarapu      62787 2022-02-26 10:49 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266438
-rw-r--r--   3 sdarapu sdarapu      40357 2022-02-26 10:50 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266439
-rw-r--r--   3 sdarapu sdarapu      74139 2022-02-26 10:50 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266440
-rw-r--r--   3 sdarapu sdarapu      48748 2022-02-26 10:50 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266441
-rw-r--r--   3 sdarapu sdarapu      52997 2022-02-26 10:50 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266442
-rw-r--r--   3 sdarapu sdarapu      42056 2022-02-26 10:50 /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266443.tmp
sdarapu@hadoop-nn001:~$
```

For SpaceX data:

```
sdarapu@hadoop-nn001:~$ hdfs dfs -ls /user/sdarapu/SpaceX_PA1data
2022-02-26 11:22:36,076 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using
Found 1 items
drwxr-xr-x   - sdarapu sdarapu          0 2022-02-26 11:02 /user/sdarapu/SpaceX_PA1data/2022
[1]+  Killed                  nohup $FLUME_HOME/bin/flume-ng agent -n TwitterAgent -f /home/sdarapu/SpaceX.conf --co
sdarapu@hadoop-nn001:~$
```

```
sdarapu@hadoop-nn001:~$ hdfs dfs -ls /user/sdarapu/SpaceX_PA1data/2022/02/26/11
2022-02-26 11:23:25,410 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
Found 74 items
-rw-r--r--   3 sdarapu sdarapu      61262 2022-02-26 11:03 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964699
-rw-r--r--   3 sdarapu sdarapu      49669 2022-02-26 11:03 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964700
-rw-r--r--   3 sdarapu sdarapu      61912 2022-02-26 11:03 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964701
-rw-r--r--   3 sdarapu sdarapu      59084 2022-02-26 11:03 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964702
-rw-r--r--   3 sdarapu sdarapu      55412 2022-02-26 11:04 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964703
-rw-r--r--   3 sdarapu sdarapu      58106 2022-02-26 11:04 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964704
-rw-r--r--   3 sdarapu sdarapu      62556 2022-02-26 11:04 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964705
-rw-r--r--   3 sdarapu sdarapu      50443 2022-02-26 11:04 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964706
-rw-r--r--   3 sdarapu sdarapu      49147 2022-02-26 11:05 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964707
-rw-r--r--   3 sdarapu sdarapu      48598 2022-02-26 11:05 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964708
-rw-r--r--   3 sdarapu sdarapu      56231 2022-02-26 11:05 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964709
-rw-r--r--   3 sdarapu sdarapu      59453 2022-02-26 11:05 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964710
-rw-r--r--   3 sdarapu sdarapu      71578 2022-02-26 11:06 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964711
-rw-r--r--   3 sdarapu sdarapu      63325 2022-02-26 11:06 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964712
-rw-r--r--   3 sdarapu sdarapu      60639 2022-02-26 11:06 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964713
-rw-r--r--   3 sdarapu sdarapu      48888 2022-02-26 11:07 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964714
-rw-r--r--   3 sdarapu sdarapu      47519 2022-02-26 11:07 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964715
-rw-r--r--   3 sdarapu sdarapu      58856 2022-02-26 11:07 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964716
-rw-r--r--   3 sdarapu sdarapu      53773 2022-02-26 11:08 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964717
-rw-r--r--   3 sdarapu sdarapu      67972 2022-02-26 11:08 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964718
-rw-r--r--   3 sdarapu sdarapu      55499 2022-02-26 11:08 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964719
-rw-r--r--   3 sdarapu sdarapu      61846 2022-02-26 11:08 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964720
-rw-r--r--   3 sdarapu sdarapu      58379 2022-02-26 11:09 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964721
-rw-r--r--   3 sdarapu sdarapu      55931 2022-02-26 11:09 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964722
-rw-r--r--   3 sdarapu sdarapu      52617 2022-02-26 11:09 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964723
-rw-r--r--   3 sdarapu sdarapu      57969 2022-02-26 11:09 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964724
-rw-r--r--   3 sdarapu sdarapu      55370 2022-02-26 11:10 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964725
-rw-r--r--   3 sdarapu sdarapu      64909 2022-02-26 11:10 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964726
-rw-r--r--   3 sdarapu sdarapu      60614 2022-02-26 11:10 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964727
-rw-r--r--   3 sdarapu sdarapu      52513 2022-02-26 11:10 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964728
-rw-r--r--   3 sdarapu sdarapu      55351 2022-02-26 11:11 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964729
-rw-r--r--   3 sdarapu sdarapu      56288 2022-02-26 11:11 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964730
-rw-r--r--   3 sdarapu sdarapu      59462 2022-02-26 11:11 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964731
-rw-r--r--   3 sdarapu sdarapu      51914 2022-02-26 11:11 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964732
-rw-r--r--   3 sdarapu sdarapu      58900 2022-02-26 11:11 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964733
-rw-r--r--   3 sdarapu sdarapu      75405 2022-02-26 11:12 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964734
-rw-r--r--   3 sdarapu sdarapu      63331 2022-02-26 11:12 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964735
-rw-r--r--   3 sdarapu sdarapu      48641 2022-02-26 11:12 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964736
```

```
-rw-r--r--   3 sdarapu sdarapu      48641 2022-02-26 11:12 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964736
-rw-r--r--   3 sdarapu sdarapu      62715 2022-02-26 11:13 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964737
-rw-r--r--   3 sdarapu sdarapu      61020 2022-02-26 11:13 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964738
-rw-r--r--   3 sdarapu sdarapu      70260 2022-02-26 11:13 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964739
-rw-r--r--   3 sdarapu sdarapu      63834 2022-02-26 11:14 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964740
-rw-r--r--   3 sdarapu sdarapu      45462 2022-02-26 11:14 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964741
-rw-r--r--   3 sdarapu sdarapu      46502 2022-02-26 11:14 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964742
-rw-r--r--   3 sdarapu sdarapu      57795 2022-02-26 11:15 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964743
-rw-r--r--   3 sdarapu sdarapu      58025 2022-02-26 11:15 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964744
-rw-r--r--   3 sdarapu sdarapu      57354 2022-02-26 11:15 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964745
-rw-r--r--   3 sdarapu sdarapu      50778 2022-02-26 11:15 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964746
-rw-r--r--   3 sdarapu sdarapu      51560 2022-02-26 11:15 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964747
-rw-r--r--   3 sdarapu sdarapu      48214 2022-02-26 11:16 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964748
-rw-r--r--   3 sdarapu sdarapu      46828 2022-02-26 11:16 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964749
-rw-r--r--   3 sdarapu sdarapu      47833 2022-02-26 11:16 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964750
-rw-r--r--   3 sdarapu sdarapu      56978 2022-02-26 11:16 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964751
-rw-r--r--   3 sdarapu sdarapu      48626 2022-02-26 11:16 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964752
-rw-r--r--   3 sdarapu sdarapu      51662 2022-02-26 11:17 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964753
-rw-r--r--   3 sdarapu sdarapu      49302 2022-02-26 11:17 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964754
-rw-r--r--   3 sdarapu sdarapu      55203 2022-02-26 11:17 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964755
-rw-r--r--   3 sdarapu sdarapu      52304 2022-02-26 11:18 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964756
-rw-r--r--   3 sdarapu sdarapu      51732 2022-02-26 11:18 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964757
-rw-r--r--   3 sdarapu sdarapu      56892 2022-02-26 11:18 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964758
-rw-r--r--   3 sdarapu sdarapu      52978 2022-02-26 11:19 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964759
-rw-r--r--   3 sdarapu sdarapu      45904 2022-02-26 11:19 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964760
-rw-r--r--   3 sdarapu sdarapu      51829 2022-02-26 11:19 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964761
-rw-r--r--   3 sdarapu sdarapu      35892 2022-02-26 11:19 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964762
-rw-r--r--   3 sdarapu sdarapu      52258 2022-02-26 11:19 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964763
-rw-r--r--   3 sdarapu sdarapu      48903 2022-02-26 11:20 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964764
-rw-r--r--   3 sdarapu sdarapu      38768 2022-02-26 11:20 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964765
-rw-r--r--   3 sdarapu sdarapu      38720 2022-02-26 11:20 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964766
-rw-r--r--   3 sdarapu sdarapu      51050 2022-02-26 11:20 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964767
-rw-r--r--   3 sdarapu sdarapu      43081 2022-02-26 11:20 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964768
-rw-r--r--   3 sdarapu sdarapu      56772 2022-02-26 11:21 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964769
-rw-r--r--   3 sdarapu sdarapu      60545 2022-02-26 11:21 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964770
-rw-r--r--   3 sdarapu sdarapu      43672 2022-02-26 11:21 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964771
-rw-r--r--   3 sdarapu sdarapu          0 2022-02-26 11:21 /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964772.tmp
sdarapu@hadoop-nn001:~$
```

11. To view the data in the flume, I have executed the below command for one flume file.

For NASA data:

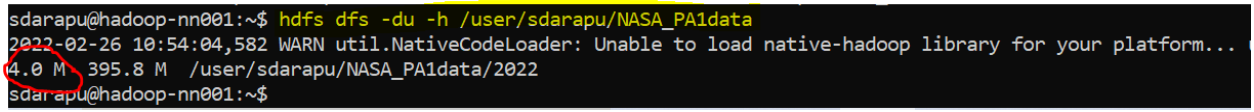
```
sdarapu@hadoop-nn001:~$ hdfs dfs -cat /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266439
2022-02-26 10:55:25,805 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
{"in_reply_to_status_id_str":null,"in_reply_to_status_id":null,"created_at":"Sat Feb 26 16:49:42 +0000 2022","in_reply_to_user_id_str":null,"source":"<a href=\"http://twitt
er.com/download/android\" rel=\"nofollow\">Twitter for Android<\/a>","retweeted_status":{"extended_tweet":{"extended_entities":{"media":[{"display_url":"pic.twitter.com/QRT
JmcVpWZ","indices":[261,284],"sizes":{"small":{"w":680,"h":529,"resize":"fit"},"large":{"w":1200,"h":933,"resize":"fit"},"thumb":{"w":150,"h":150,"resize":"crop"},"medium":
{"w":1200,"h":933,"resize":"fit"}},"id_str":"1497570165701345281","expanded_url":"https://twitter.com/haocheolkartt/status/1497570184714133504/photo/1","media_url_https":"h
ttps://pbs.twimg.com/media/FMhwIZjaAAEMvj9.jpg","id":1497570165701345281,"type":"photo","media_url":"http://pbs.twimg.com/media/FMhwIZjaAAEMvj9.jpg","url":"https://t.co/QRT
JmcVpWZ"},{"display_url":"pic.twitter.com/QRTJmcVpWZ","indices":[261,284],"sizes":{"small":{"w":680,"h":466,"resize":"fit"},"large":{"w":1200,"h":823,"resize":"fit"},"thumb
":{"w":150,"h":150,"resize":"crop"},"medium":{"w":1200,"h":823,"resize":"fit"}},"id_str":"1497570170210222082","expanded_url":"https://twitter.com/haocheolkartt/status/1497
570184714133504/photo/1","media_url_https":"https://pbs.twimg.com/media/FMhwIqWaAAI3FH_.jpg","id":1497570170210222082,"type":"photo","media_url":"http://pbs.twimg.com/media
/FMhwIqWaAAI3FH_.jpg","url":"https://t.co/QRTJmcVpWZ"},{"display_url":"pic.twitter.com/QRTJmcVpWZ","indices":[261,284],"sizes":{"small":{"w":680,"h":176,"resize":"fit"},"la
rge":{"w":864,"h":223,"resize":"fit"},"thumb":{"w":150,"h":150,"resize":"crop"},"medium":{"w":864,"h":223,"resize":"fit"}},"id_str":"1497570175264378881","expanded_url":"ht
tps://twitter.com/haocheolkartt/status/1497570184714133504/photo/1","media_url_https":"https://pbs.twimg.com/media/FMhwI9LaUAEU5KM.jpg","id":1497570175264378881,"type":"pho
to","media_url":"http://pbs.twimg.com/media/FMhwI9LaUAEU5KM.jpg","url":"https://t.co/QRTJmcVpWZ"},{"display_url":"pic.twitter.com/QRTJmcVpWZ","indices":[261,284],"sizes":{"
small":{"w":680,"h":680,"resize":"fit"},"large":{"w":1200,"h":1200,"resize":"fit"},"thumb":{"w":150,"h":150,"resize":"crop"},"medium":{"w":1200,"h":1200,"resize":"fit"}},"i
d_str":"1497570180935086084","expanded_url":"https://twitter.com/haocheolkartt/status/1497570184714133504/photo/1","media_url_https":"https://pbs.twimg.com/media/FMhwJSTacA
QOx1t.jpg","id":1497570180935086084,"type":"photo","media_url":"http://pbs.twimg.com/media/FMhwJSTacAQOx1t.jpg","url":"https://t.co/QRTJmcVpWZ"}]},"entities":{"urls":[],"ha
shtags":[],"media":[{"display_url":"pic.twitter.com/QRTJmcVpWZ","indices":[261,284],"sizes":{"small":{"w":680,"h":529,"resize":"fit"},"large":{"w":1200,"h":933,"resize":"fi
t"},"thumb":{"w":150,"h":150,"resize":"crop"},"medium":{"w":1200,"h":933,"resize":"fit"}},"id_str":"1497570165701345281","expanded_url":"https://twitter.com/haocheolkartt/s
tatus/1497570184714133504/photo/1","media_url_https":"https://pbs.twimg.com/media/FMhwIZjaAAEMvj9.jpg","id":1497570165701345281,"type":"photo","media_url":"http://pbs.twimg
.com/media/FMhwIZjaAAEMvj9.jpg","url":"https://t.co/QRTJmcVpWZ"},{"display_url":"pic.twitter.com/QRTJmcVpWZ","indices":[261,284],"sizes":{"small":{"w":680,"h":466,"resize":
"fit"},"large":{"w":1200,"h":823,"resize":"fit"},"thumb":{"w":150,"h":150,"resize":"crop"},"medium":{"w":1200,"h":823,"resize":"fit"}},"id_str":"1497570170210222082","expan
ded_url":"https://twitter.com/haocheolkartt/status/1497570184714133504/photo/1","media_url_https":"https://pbs.twimg.com/media/FMhwIqWaAAI3FH_.jpg","id":1497570170210222082
,"type":"photo","media_url":"http://pbs.twimg.com/media/FMhwIqWaAAI3FH_.jpg","url":"https://t.co/QRTJmcVpWZ"},{"display_url":"pic.twitter.com/QRTJmcVpWZ","indices":[261,284
],"sizes":{"small":{"w":680,"h":176,"resize":"fit"},"large":{"w":864,"h":223,"resize":"fit"},"thumb":{"w":150,"h":150,"resize":"crop"},"medium":{"w":864,"h":223,"resize":"f
it"}},"id_str":"1497570175264378881","expanded_url":"https://twitter.com/haocheolkartt/status/1497570184714133504/photo/1","media_url_https":"https://pbs.twimg.com/media/FM
hwI9LaUAEU5KM.jpg","id":1497570175264378881,"type":"photo","media_url":"http://pbs.twimg.com/media/FMhwI9LaUAEU5KM.jpg","url":"https://t.co/QRTJmcVpWZ"},{"display_url":"pic
.twitter.com/QRTJmcVpWZ","indices":[261,284],"sizes":{"small":{"w":680,"h":680,"resize":"fit"},"large":{"w":1200,"h":1200,"resize":"fit"},"thumb":{"w":150,"h":150,"resize":
"crop"},"medium":{"w":1200,"h":1200,"resize":"fit"}},"id_str":"1497570180935086084","expanded_url":"https://twitter.com/haocheolkartt/status/1497570184714133504/photo/1","m
edia_url_https":"https://pbs.twimg.com/media/FMhwJSTacAQOx1t.jpg","id":1497570180935086084,"type":"photo","media_url":"http://pbs.twimg.com/media/FMhwJSTacAQOx1t.jpg","url"
:"https://t.co/QRTJmcVpWZ"}],"user_mentions":[],"symbols":[]},"full_text":"UPDATE ! \n\nmakihati na kayo pls □ \n\n□  = can steal if getting the pola set\n will prio din yu
ng kukuha ng mga stbos ( 3rd pic ) and kukuha nung mga nasa last pic ^____^\n\nt. wts lfb lf kahati svt seventeen scoups cheol seungcheol assorted poca pc photocard ph only
 https://t.co/QRTJmcVpWZ","display_text_range":[0,260]},"in_reply_to_status_id_str":"1496103474140151811","in_reply_to_status_id":1496103474140151811,"created_at":"Sat Feb
26 13:52:10 +0000 2022","in_reply_to_user_id_str":"1486290326201724932","source":"<a href=\"http://twitter.com/download/android\" rel=\"nofollow\">Twitter for Android<\/a>"
,"retweet_count":13,"retweeted":false,"geo":null,"filter_level":"low","in_reply_to_screen_name":"haocheolkartt","is_quote_status":false,"id_str":"1497570184714133504","in_r
eply_to_user_id":1486290326201724932,"favorite_count":1,"id":1497570184714133504,"text":"UPDATE ! \n\nmakihati na kayo pls □ \n\n□  = can steal if getting the pola set\n wi
ll prio din yung kukuha ng mga stbos (\u2026 https://t.co/GvXRYZjSo2","place":null,"lang":"tl","quote_count":1,"favorited":false,"possibly_sensitive":false,"coordinates":nu
ll,"truncated":true,"reply_count":0,"entities":{"urls":[{"display_url":"twitter.com/i/web/status/1\u2026","indices":[116,139],"expanded_url":"https://twitter.com/i/web/stat
```

- We can also “hadoop fs -cat /path” command instead of hdfs dfs -cat /path command

For SpaceX data:

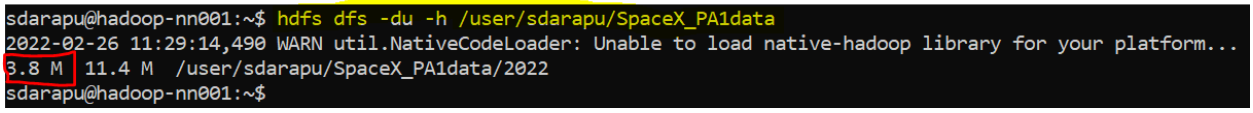
12. Now I used below command to check the size of the data that has been download after killing the respective processes.

For NASA data:



Size of the keyword NASA data downloaded is 4.0 MB

For SpaceX data:



Size of the keyword SpaceX data downloaded is 3.8 MB