

CS 5433: Bigdata Management
Programming Assignment 1
ReadMe file for PART 1

CWID: A20343337

- Keywords used to download the data from twitter are NASA and SpaceX

Steps to run the code:

1. Create two configuration files NASA.conf and SpaceX.conf in the Hadoop cluster by using nano command as shown below or using the WinSCP.

(Check SatyaRajyaSaiTejaswini_Darapureddy_Program_PA1 for reference)

```
sdarapu@hadoop-nn001:~$ nano NASA.conf
```

```
sdarapu@hadoop-nn001:~$ nano SpaceX.conf
```

2. Run nohup command to download the twitter data into the specified directory which was created earlier.

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 &
```

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 &
```

3. Use the below command to check the downloaded files.

For NASA data:

```
sdarapu@hadoop-nn001:~$ hdfs dfs -ls /user/sdarapu/NASA_PA1data/2022/02/26/10
```

For SpaceX data:

```
sdarapu@hadoop-nn001:~$ hdfs dfs -ls /user/sdarapu/SpaceX_PA1data/2022/02/26/11
```

4. Use the below command to view the data of any of the flume file.

For NASA data:

```
sdarapu@hadoop-nn001:~$ hdfs dfs -cat /user/sdarapu/NASA_PA1data/2022/02/26/10/FlumeData.1645893266439
```

[For SpaceX data:](#)

```
sdarapu@hadoop-nn001:~$ hdfs dfs -cat /user/sdarapu/SpaceX_PA1data/2022/02/26/11/FlumeData.1645894964766
```

5. Use the below command to check the size of the data downloaded and after that kill nohup command once the enough amount of data is downloaded.

[For NASA data:](#)

```
sdarapu@hadoop-nn001:~$ hdfs dfs -du -h /user/sdarapu/NASA_PA1data
```

[For SpaceX data:](#)

```
sdarapu@hadoop-nn001:~$ hdfs dfs -du -h /user/sdarapu/SpaceX_PA1data
```

6. To the kill the nohup, first get its process id and then kill it by using the below commands.

```
sdarapu@hadoop-nn001:~$ ps -x
```

```
sdarapu@hadoop-nn001:~$ kill -9 369858
```