



SB 8016 - Big Data Analytics

Mini Project Report

Name of the Student : MOHAMMED ARMASH S.M.

Register Number : 211720118019

Year : 3rd

Semester: 5

Section: A

Branch: COMPUTER AND COMMUNICATION ENGINEERING

Course Coordinator : Mr.M.Ashok/Ms.R.Saranya

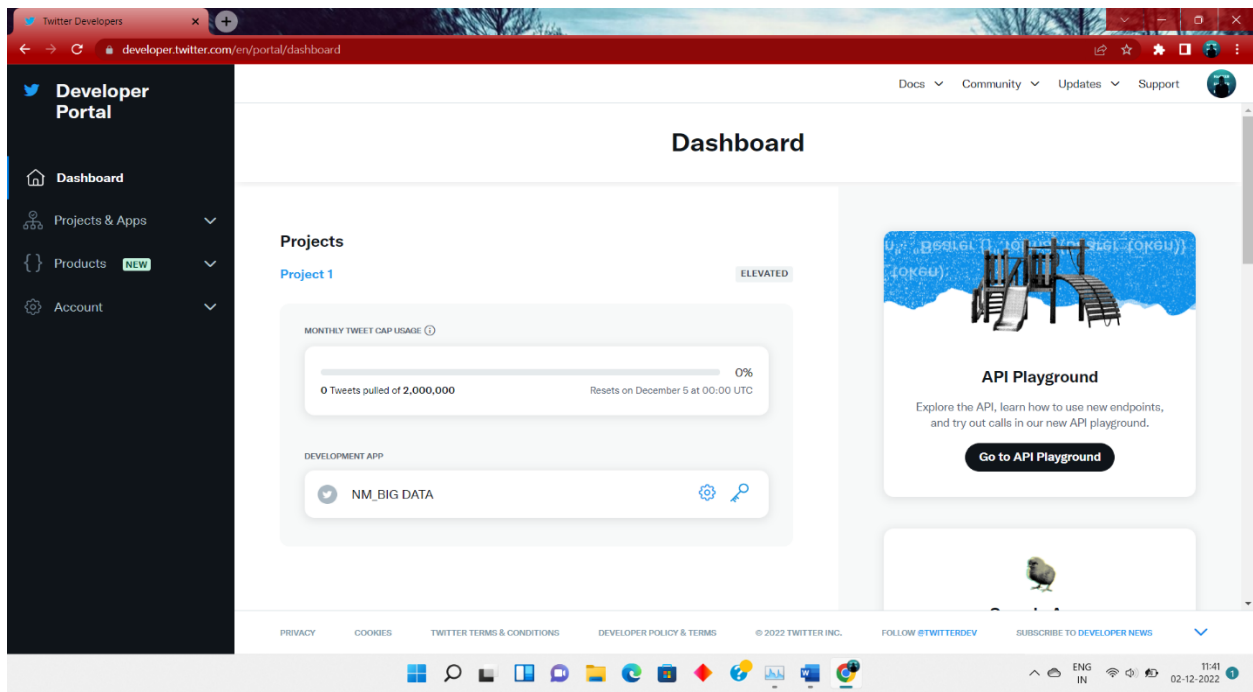
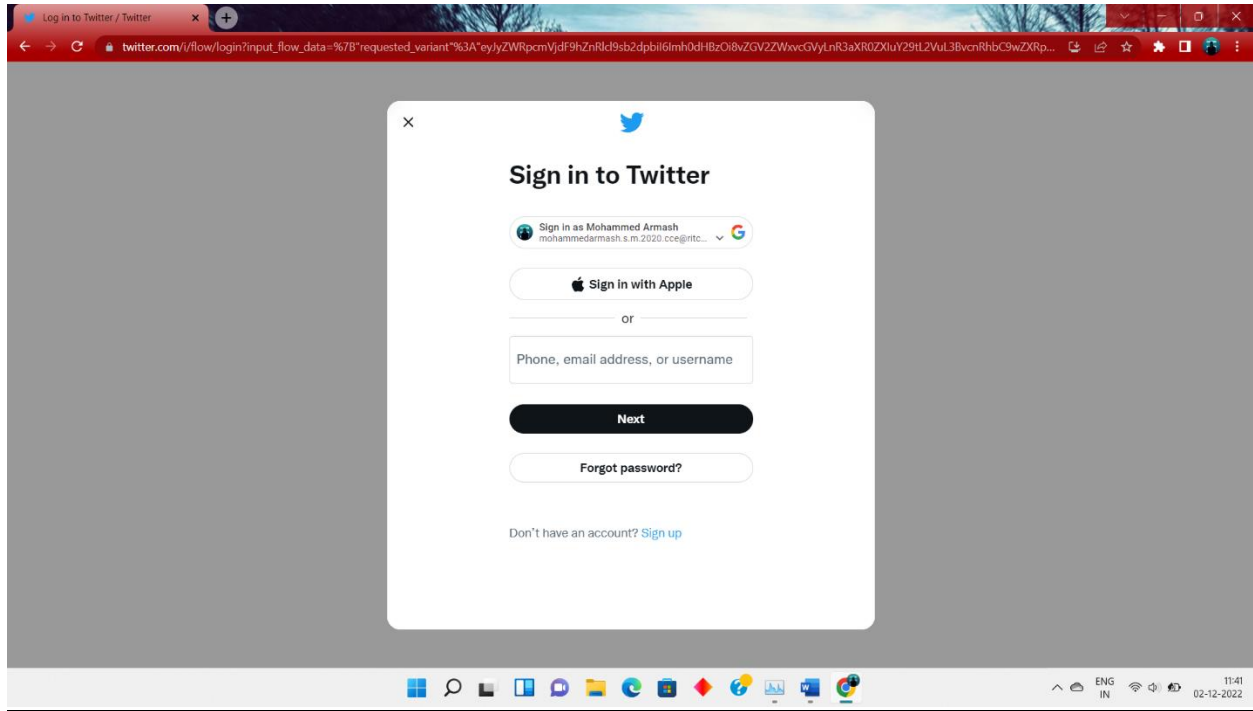
Mini Project Mark Evaluation

S.No	Rubrics	Marks
1	Twitter Developer Account Creation	
2	Generating “Consumer API keys”, “Consumer secret”, “Access token,” and “Access token secret”	
3	Creating Flume Configuration file	
4	Writing Flume Command	
5	Twitter Data in HDFS	

**Course Coordinator
Signature**

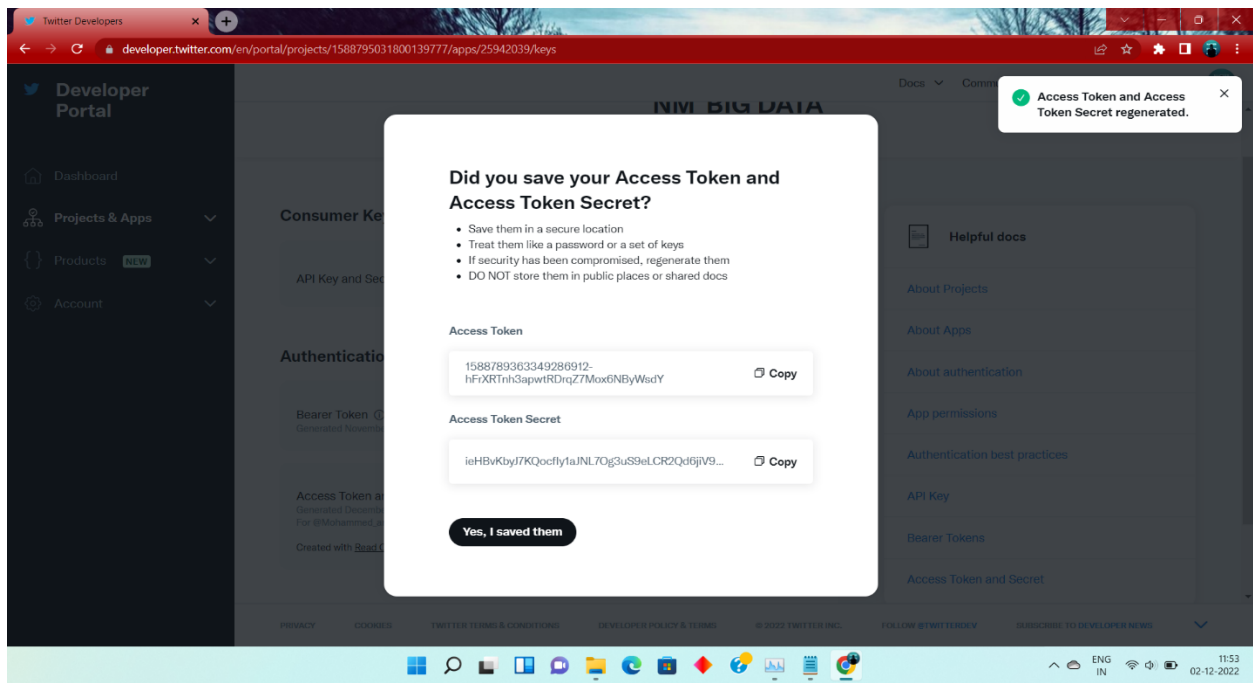
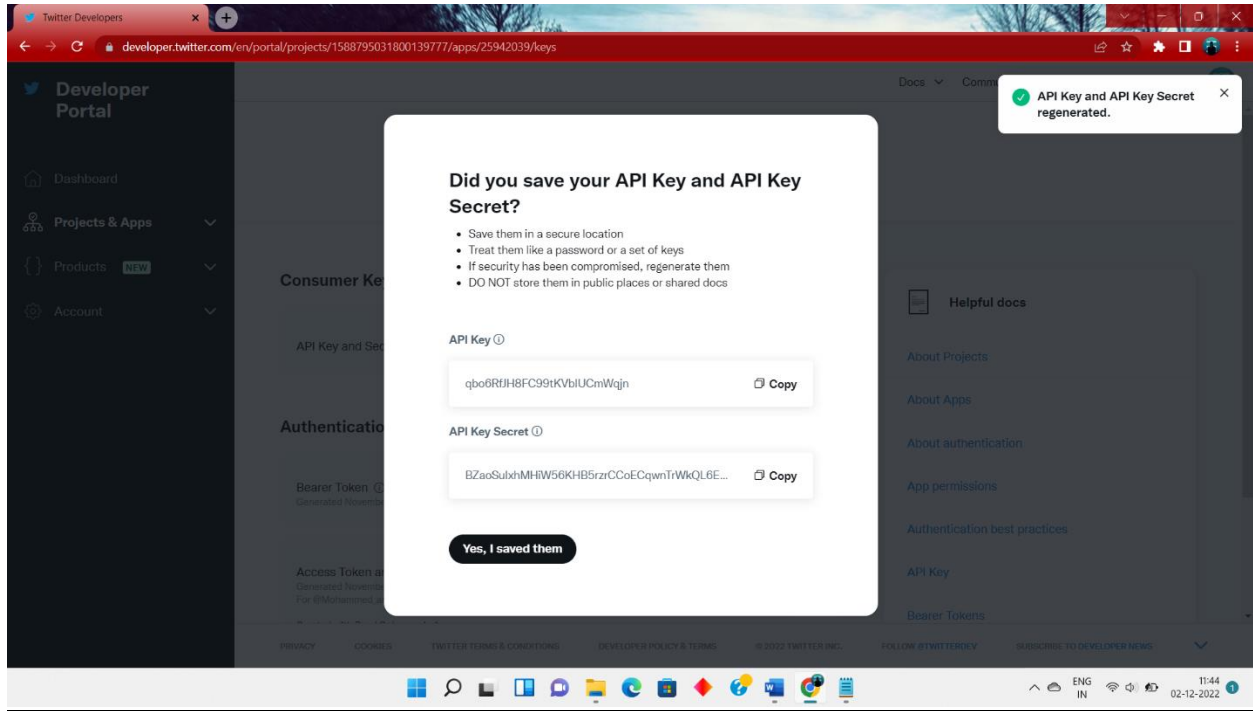
Twitter Developer Account Creation

211720118019



Generating “Consumer API keys”, “Consumer secret”, “Access token,” and “Access token secret”

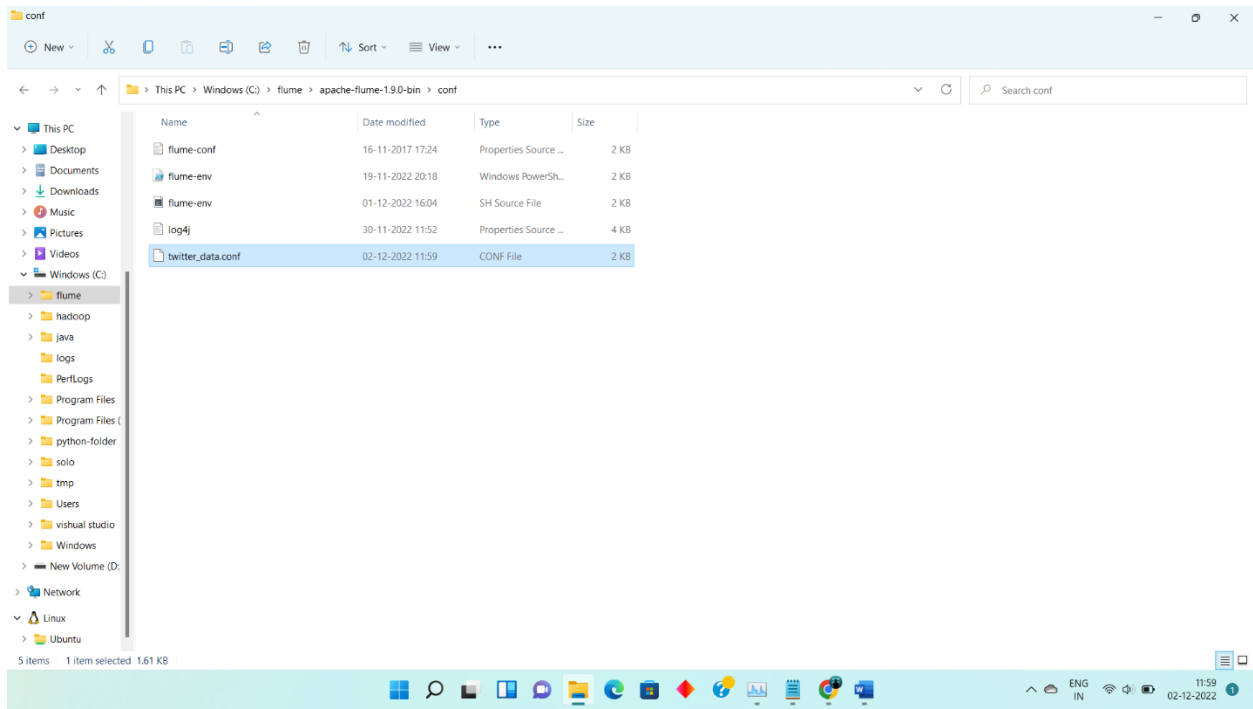
211720118019



Creating Flume Configuration file

211720118019

```
C:\flume\apache-flume-1.9.0-bin\conf\twitter_data.conf - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
twitter_data.conf
1 # Naming the components on the current agent.
2 TwitterAgent.sources = Twitter
3 TwitterAgent.channels = MemChannel
4 TwitterAgent.sinks = HDFS
5
6 # DescribingConfiguring the source
7 TwitterAgent.sources.Twitter.type = org.apache.flume.source.twitter.TwitterSource
8 TwitterAgent.sources.Twitter.consumerKey = q648E2H8FC94s1PbIDCmWqin
9 TwitterAgent.sources.Twitter.consumerSecret = 82as5u1ab0U8460B3vrrCCoEOpwTrWqLAKTHQ0ks0KTSS
10 TwitterAgent.sources.Twitter.accessToken = 158878936349284912-hfrXRTnh3apvtRDrqL7Moxe0NByWsdY
11 TwitterAgent.sources.Twitter.accessTokenSecret = i6B5Vf0pJ7RQocfTylaJNL70q3u59eLCR2Qd6j1VpW0gv
12 TwitterAgent.sources.Twitter.keywords = spark, scientist, hadoop, big data,
13 analytics, bigdata, cloudera, data science, data scientist, business intelligence,
14 mapreduce, data warehouse, data warehousing, mahout, hbase, nosql, newsql,
15 businessintelligence, cloudcomputing
16
17 # DescribingConfiguring the sink
18
19 TwitterAgent.sinks.HDFS.type = hdfs
20 TwitterAgent.sinks.HDFS.hdfs.path = hdfs://localhost:9000/user/hadoop/twitter_data/
21 TwitterAgent.sinks.HDFS.hdfs.fileType = DataStream
22 TwitterAgent.sinks.HDFS.hdfs.writeFormat = Text
23 TwitterAgent.sinks.HDFS.hdfs.batchSize = 1000
24 TwitterAgent.sinks.HDFS.hdfs.rollSize = 0
25 TwitterAgent.sinks.HDFS.hdfs.rollCount = 10000
26
27 # DescribingConfiguring the channel
28 TwitterAgent.channels.MemChannel.type = memory
29 TwitterAgent.channels.MemChannel.capacity = 10000
30 TwitterAgent.channels.MemChannel.transactionCapacity = 1000
31
32 # Binding the source and sink to the channel
33 TwitterAgent.sources.Twitter.channels = MemChannel
34 TwitterAgent.sinks.HDFS.channel = MemChannel
35
```



Writing Flume Command

211720118019

```
Administrator: Command Prompt
Microsoft Windows [Version 10.0.22000.978]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>cd ..

C:\Windows>cd ..

C:\>cd flume

C:\flume>dir
Volume in drive C is Windows
Volume Serial Number is 2658-C19A

Directory of C:\flume

19-11-2022  19:48    <DIR>          .
19-11-2022  19:48    <DIR>          apache-flume-1.9.0-bin
               0 File(s)                0 bytes
               2 Dir(s)  175,259,402,240 bytes free

C:\flume>cd apache-flume-1.9.0-bin

C:\flume\apache-flume-1.9.0-bin>cd bin

C:\flume\apache-flume-1.9.0-bin\bin>flume-ng agent -n TwitterAgent -c C:\flume\apache-flume-1.9.0-bin\conf -f C:\flume\apache-flume-1.9.0-bin\conf\twitter_data.conf -property flume.root.logger=DEBUG,console
```

```
Administrator: Command Prompt - flume-ng agent -n TwitterAgent -c C:\flume\apache-flume-1.9.0-bin\conf -f C:\flume\apache-flume-1.9.0-bin\conf\twitter_data.conf -property flume.root.logger=DEBUG,console
Microsoft Windows [Version 10.0.22000.978]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>cd ..

C:\Windows>cd ..

C:\>cd flume

C:\flume>dir
Volume in drive C is Windows
Volume Serial Number is 2658-C19A

Directory of C:\flume

19-11-2022  19:48    <DIR>          .
19-11-2022  19:48    <DIR>          apache-flume-1.9.0-bin
               0 File(s)                0 bytes
               2 Dir(s)  175,259,402,240 bytes free

C:\flume>cd apache-flume-1.9.0-bin

C:\flume\apache-flume-1.9.0-bin>cd bin

C:\flume\apache-flume-1.9.0-bin\bin>flume-ng agent -n TwitterAgent -c C:\flume\apache-flume-1.9.0-bin\conf -f C:\flume\apache-flume-1.9.0-bin\conf\twitter_data.conf -property flume.root.logger=DEBUG,console

C:\flume\apache-flume-1.9.0-bin\bin>powershell.exe -NoProfile -InputFormat none -ExecutionPolicy unrestricted -File C:\flume\apache-flume-1.9.0-bin\bin\flume-ng.ps1 agent -n TwitterAgent -c C:\flume\apache-flume-1.9.0-bin\conf -f C:\flume\apache-flume-1.9.0-bin\conf\twitter_data.conf -property flume.root.logger=DEBUG,console

WARN: Config directory not set. Defaulting to C:\flume\apache-flume-1.9.0-bin\conf
Sourcing environment configuration script C:\flume\apache-flume-1.9.0-bin\conf\flume-env.ps1
Including Hadoop libraries found in (C:\hadoop\hadoop-3.3.4) for DFS access
WARN: HBASE_HOME not found
WARN: HIVE_HOME not found

Running FLUME agent :
class: org.apache.flume.node.Application
arguments: -n TwitterAgent -f "C:\flume\apache-flume-1.9.0-bin\conf\twitter_data.conf"
```

Twitter Data in HDFS

211720118019

Namenode Information

localhost:9870/dfshealth.html#tab-overview

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

Overview 'localhost:9000' (✓active)

Started:	Fri Dec 02 12:01:59 +0530 2022
Version:	3.3.4, ra585a73c3e02ac62350c136643a5e7f6095a3dbb
Compiled:	Fri Jul 29 18:02:00 +0530 2022 by stevel from branch-3.3.4
Cluster ID:	CID-0c0bda37-30f9-4723-b053-d185475b1c28
Block Pool ID:	BP-924729748-192.168.56.1-1668353888897

Summary

Security is off.
Safemode is off.

1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).
Heap Memory used 98.81 MB of 202.5 MB Heap Memory. Max Heap Memory is 889 MB.
Non Heap Memory used 51.13 MB of 52.69 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	238.08 GB
Configured Remote Capacity:	0 B
DFS Used:	321 B (0%)

Hadoop

Overview

Datanodes

Snapshot

Startup Progress

Utilities

Browse Directory

/user/Hadoop/twitter_data/

Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	Armash	Armash	76.39 KB	02/11/2022 12:26:19	3	128 MB	FlumeData.1496834774363
-rw-r--r--	Armash	Armash	75.07 KB	02/11/2022 12:26:24	3	128 MB	FlumeData.1496834774364
-rw-r--r--	Armash	Armash	79.7 KB	02/11/2022 12:26:27	3	128 MB	FlumeData.1496834774365
-rw-r--r--	Armash	Armash	74.67 KB	02/11/2022 12:26:31	3	128 MB	FlumeData.1496834774366
-rw-r--r--	Armash	Armash	68.58 KB	02/11/2022 12:26:35	3	128 MB	FlumeData.1496834774367
-rw-r--r--	Armash	Armash	69.9 KB	02/11/2022 12:26:39	3	128 MB	FlumeData.1496834774368
-rw-r--r--	Armash	Armash	74.64 KB	02/11/2022 12:26:43	3	128 MB	FlumeData.1496834774369
-rw-r--r--	Armash	Armash	79.11 KB	02/11/2022 12:26:46	3	128 MB	FlumeData.1496834774370
-rw-r--r--	Armash	Armash	69.21 KB	02/11/2022 12:26:51	3	128 MB	FlumeData.1496834774371
-rw-r--r--	Armash	Armash	20.29 KB	02/11/2022 12:26:57	3	128 MB	FlumeData.1496834774372