**Procedure:**

1. Applied for Twitter developer Account and got the required keys to collect Stream data.
2. Installed tweepy, pymongo packages in anaconda using anaconda command prompt.
3. Installed MongoDB community edition version and configured as given in the mongoDB documentation.
4. Using the reference code provided to get the data, collected tweets data more than 10,000 and retrieved the information using mongo DB Queries. Database Name is CovidStream and Collection name is tweets.

**Source code to get the Twitter Stream data and inserting the data into mongoDB**

# -\*- coding: utf-8 -\*-

"""

Created on Wed Mar 17 19:36:34 2021

@author: srava

"""

from tweepy import Stream

from tweepy import OAuthHandler

from tweepy.streaming import StreamListener

import pymongo

from pymongo import MongoClient

import json

#consumer key, consumer secret, access token, access secret.

CONSUMER\_KEY= "gun1fifd8iaKXLAeiFJ8kRr34"

CONSUMER\_SECRET="rIpsyP6IM6DFW6O5IJGD9deyRaOLuZJjX8gALqX3rZkTOyF5LL"

OAUTH\_TOKEN="1371821070681059331-C1OauRpgy5kT7MjyMU1O7bjehdhBYq"

OAUTH\_TOKEN\_SECRET="sAEzpDbVvwLtmFes8htHC8HVoMRWM7KKHRxMZfIx9A7oO"

# The MongoDB connection info.

conn = MongoClient('localhost', 27017)

# This assumes your database name is CovidStream.

db = conn.CovidStream

# Your collection name is tweets.

collection = db.tweets

db.tweets.create\_index([("id", pymongo.ASCENDING)],unique = True,)

class getStreamData(StreamListener):

def on\_data(self, data):

# Load the Tweet into the variable "tweet"

try:

tweet = json.loads(data)

# Pull important data from the tweet to store in the database

#One at a time.

collection.insert\_one(tweet)

return True

except:

pass

if \_\_name\_\_ == "\_\_main\_\_":

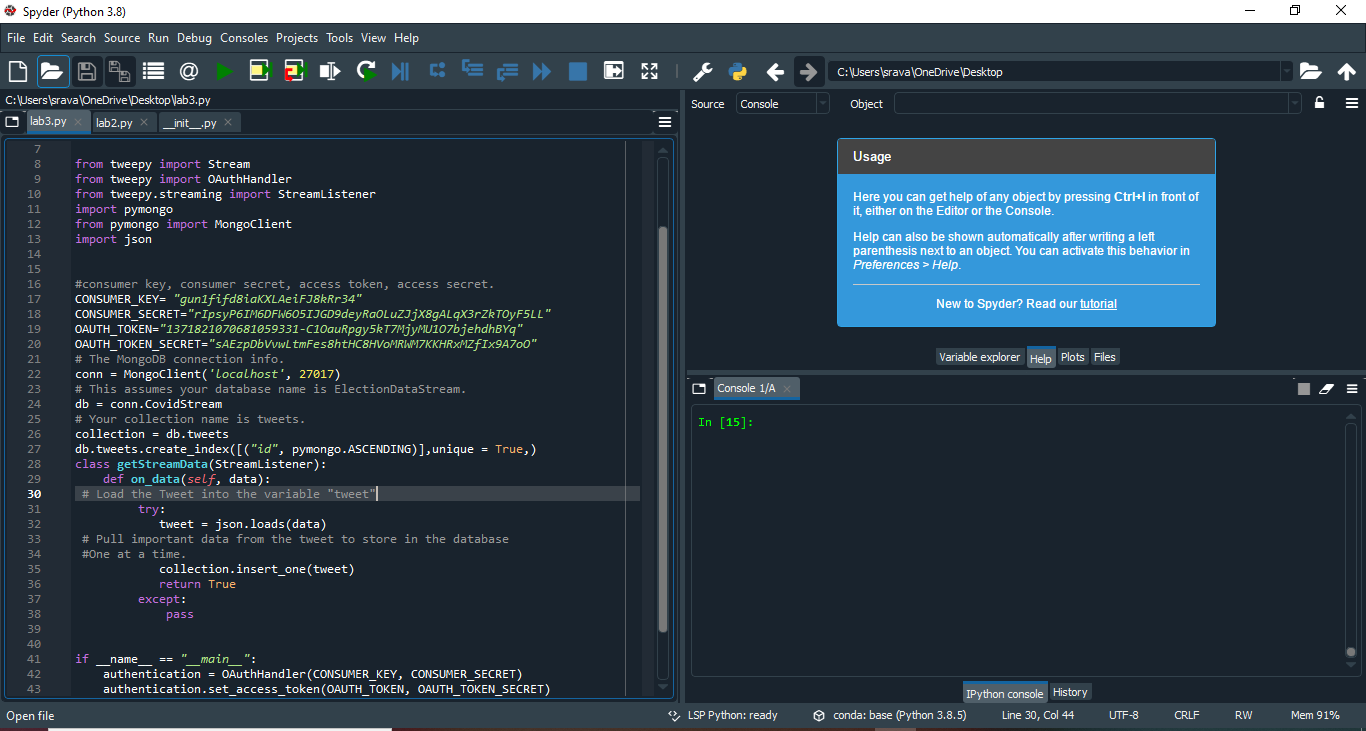
authentication = OAuthHandler(CONSUMER\_KEY, CONSUMER\_SECRET)

authentication.set\_access\_token(OAUTH\_TOKEN, OAUTH\_TOKEN\_SECRET)

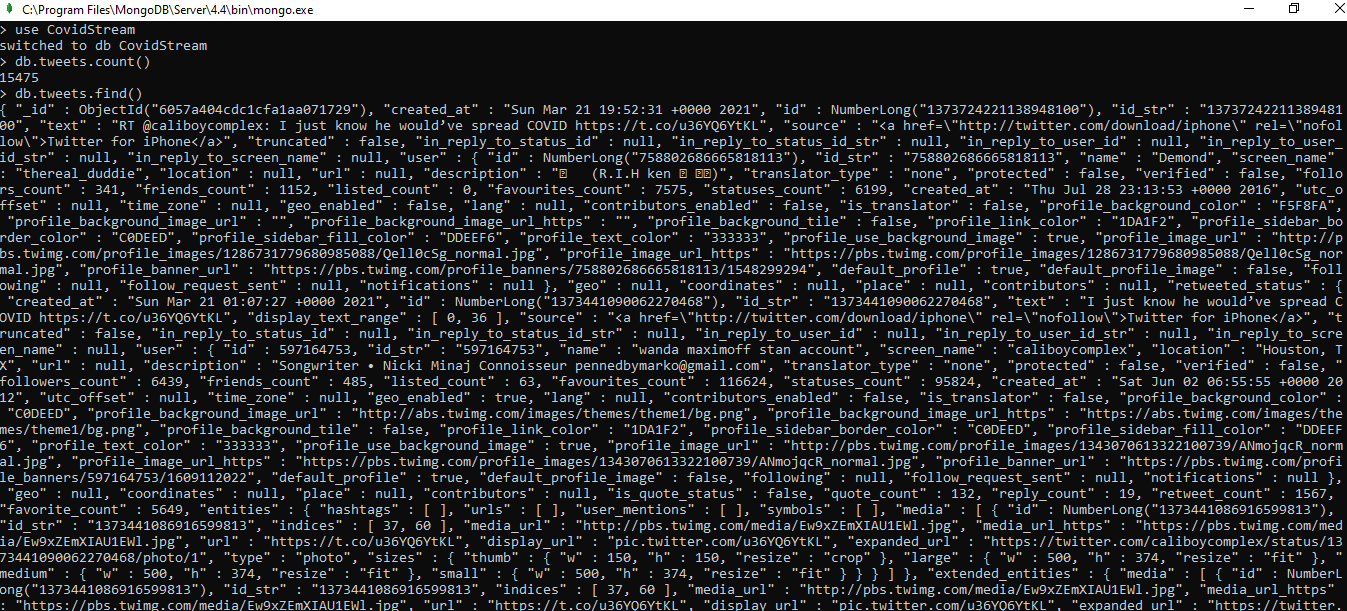
tweetStream = Stream(authentication, getStreamData())

# Here write down your keywords which you want to search for.

tweetStream.filter(track=["Coronavirus","Covid","Covid 19"])



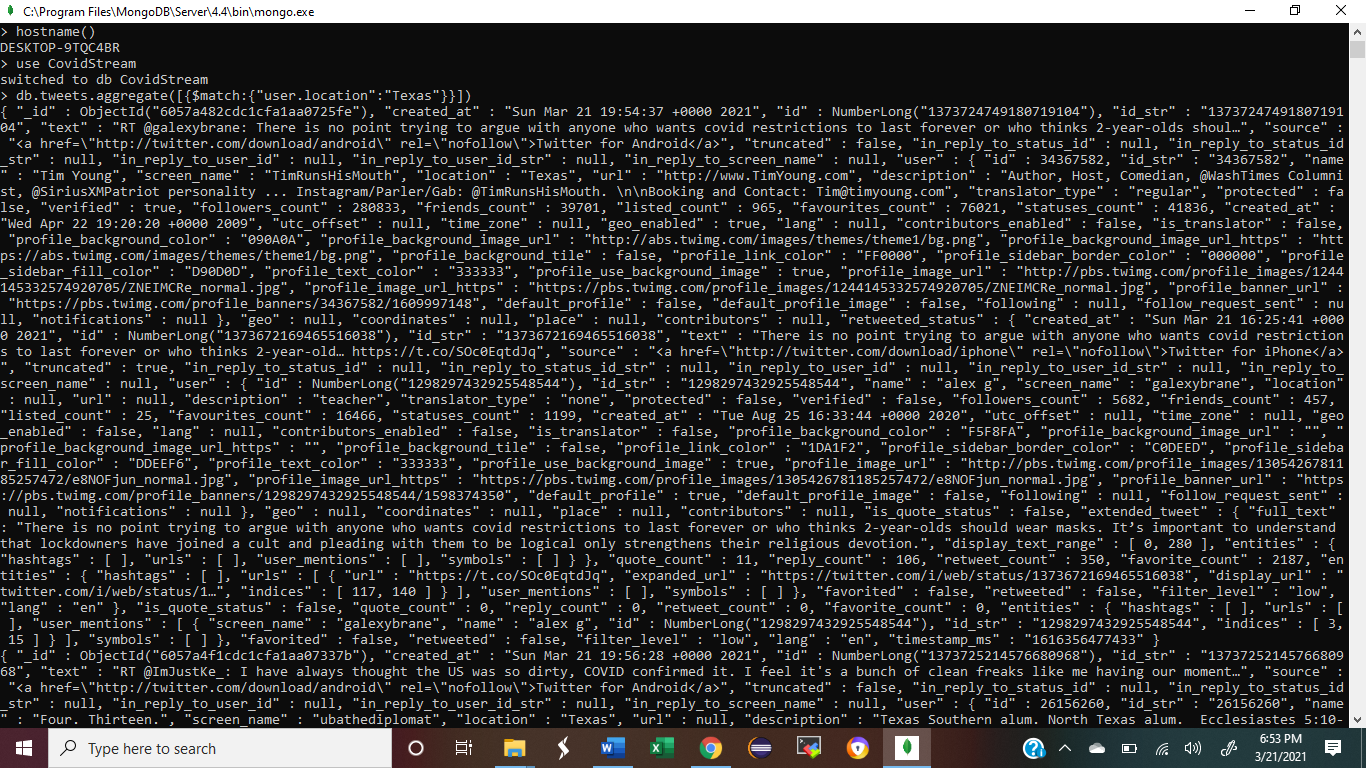
Collected more than 10,000 tweets into CovidStream Database in MongoDB and tweets as collection name.



**Query using $match to get the user location from Texas**

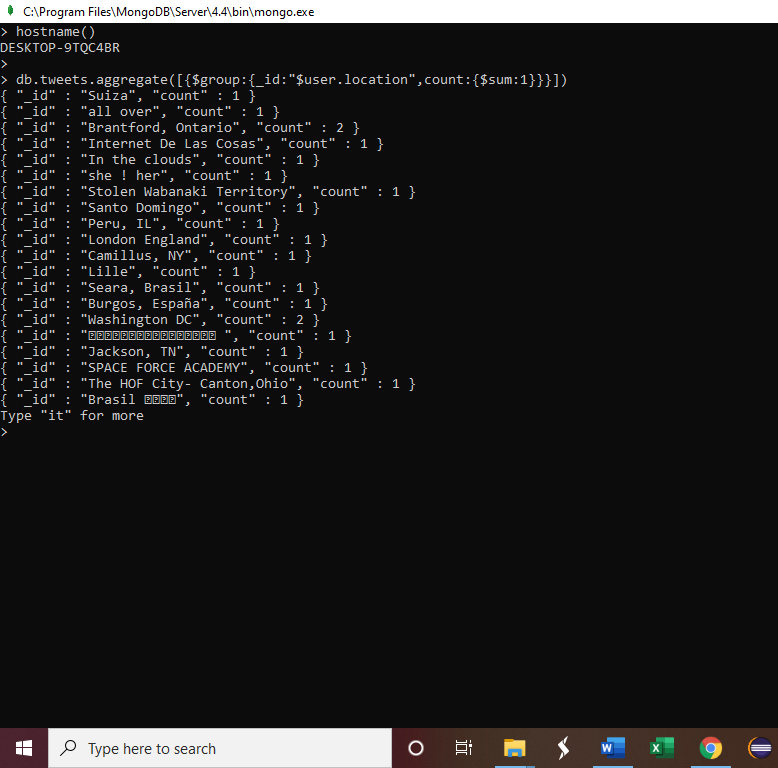
***db.tweets.aggregate([{$match:{"user.location":"Texas"}}])***

**Output:**



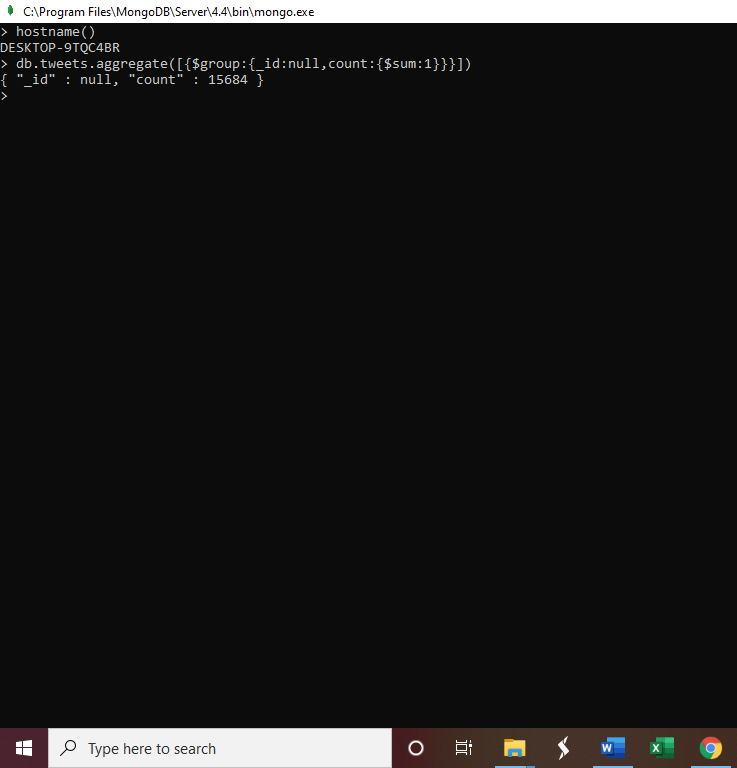
**Query to group user locations and their count using $group**

***db.tweets.aggregate([{$group:{\_id:"$user.location",count:{$sum:1}}}])***



**Query to count the number of documents in tweets collection**

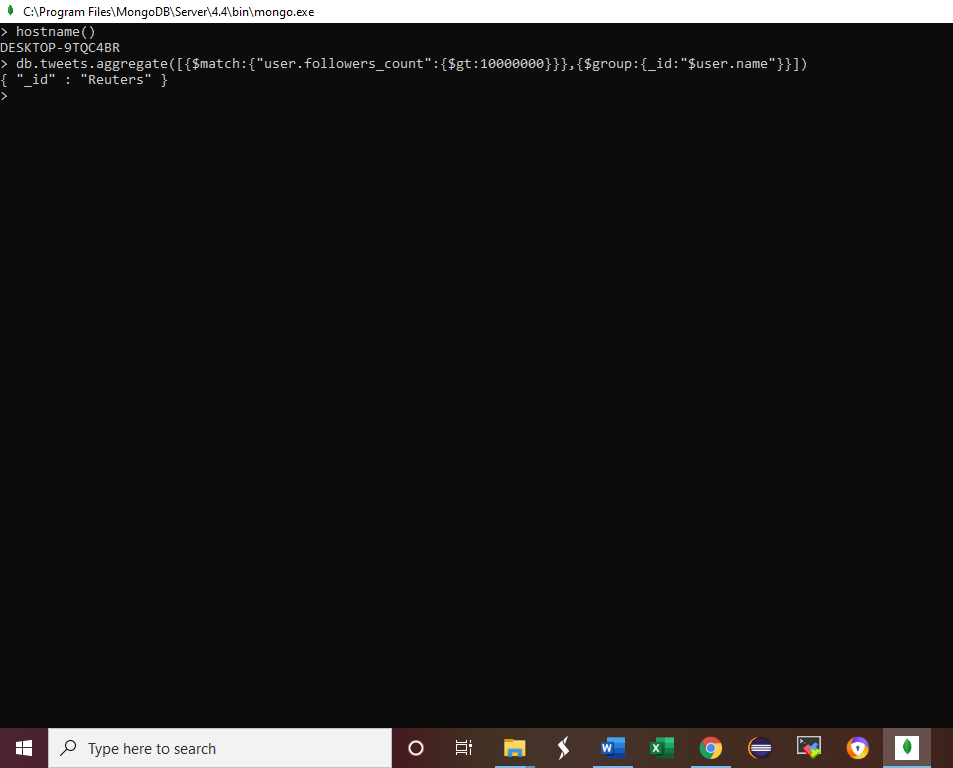
***db.tweets.aggregate([{$group:{\_id:null,count:{$sum:1}}}])***



**Aggregation pipeline Example**

**Query to retrieve the username who has followers\_count more than 10,000,000**

***db.tweets.aggregate([{$match:{"user.followers\_count":{$gt:10000000}}},{$group:{\_id:"$user.name"}}])***



**Query to retrieve username who has friends\_count more than 0.1 million**

***db.tweets.aggregate([{$match:{"user.friends\_count":{$gt:100000}}},{$group:{\_id:"$user.name"}}])***

