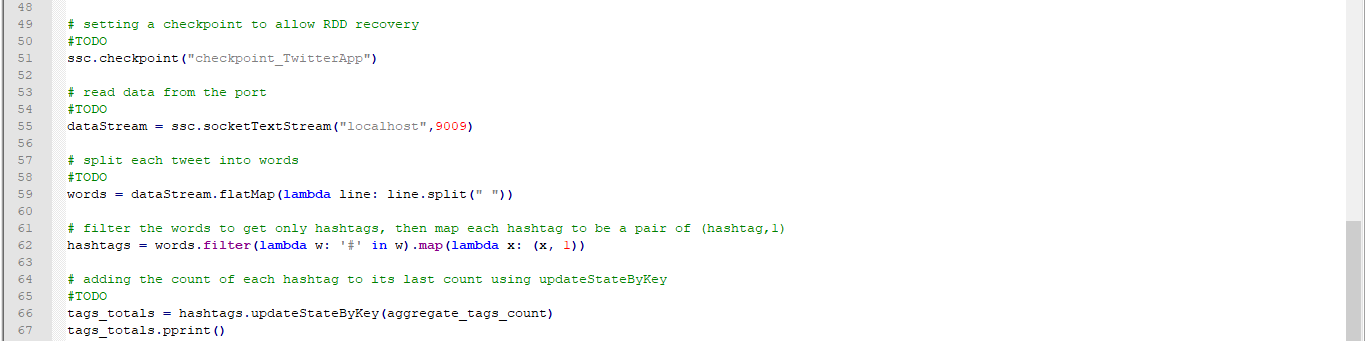
**REQUESTING STEPS:**

1. Fill the missing parts in the Spark Streaming application. (the TODO parts)

Solution of missing parts are available in Spark\_streaming\_application\_1.py



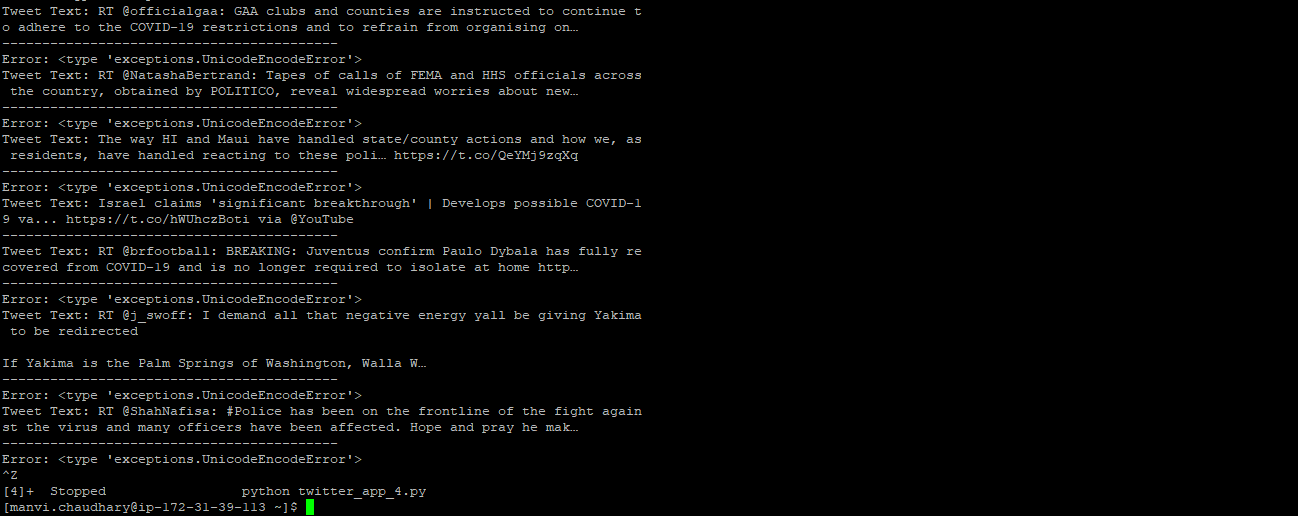
1. Connect to the edge machine and upload the files or create them with a text editor. Please note that you need to change the application port from 9009 to a different one to avoid conflicts with running programs from other colleagues

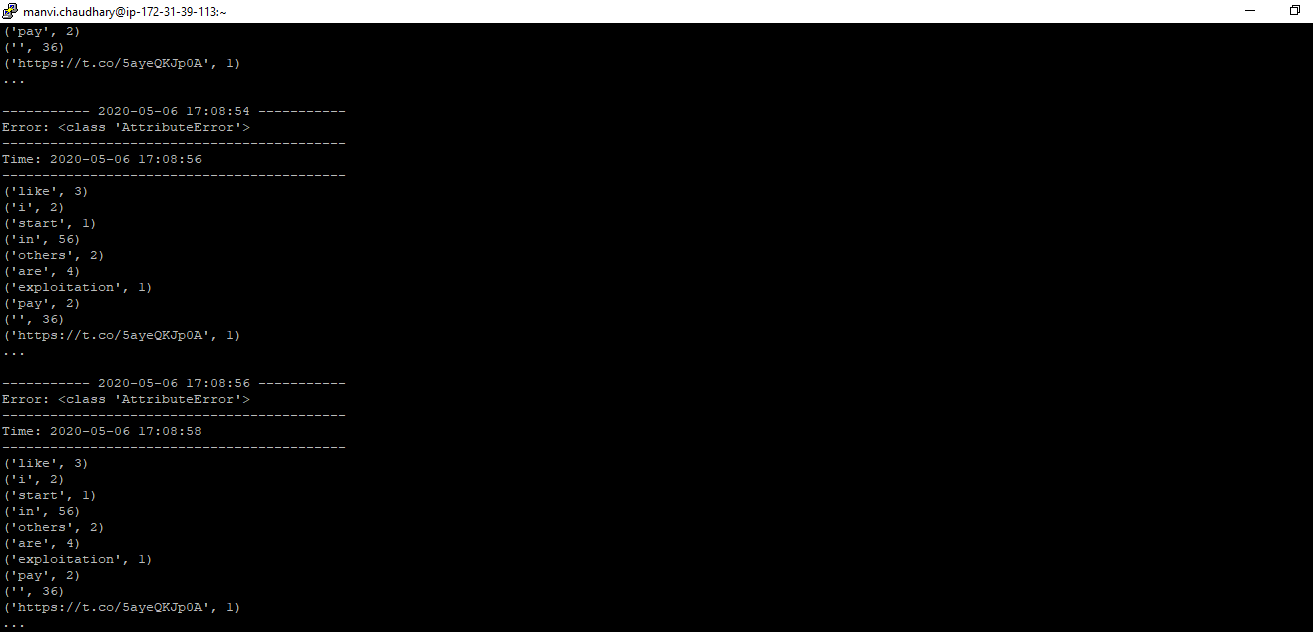
Connected to the edge machine and uploaded files twitter\_app\_2.py and Spark\_streaming\_application\_2.py and changed the application port to 2001.

1. In the query from twitter\_app.py instead of getting the tweets of the location we have, use another location and word instead of #.

I used Madrid bounding box -3.7834,40.3735,-3.6233,40.4702 and tracked tweets that has **Covid-19** word, the code for this problem is present in twitter\_app\_2.py and Spark\_streamng\_application\_2.py file

Output :





1. When the top10 elements are computed, copy the output and paste in your submission.

Top 10 elements computed:

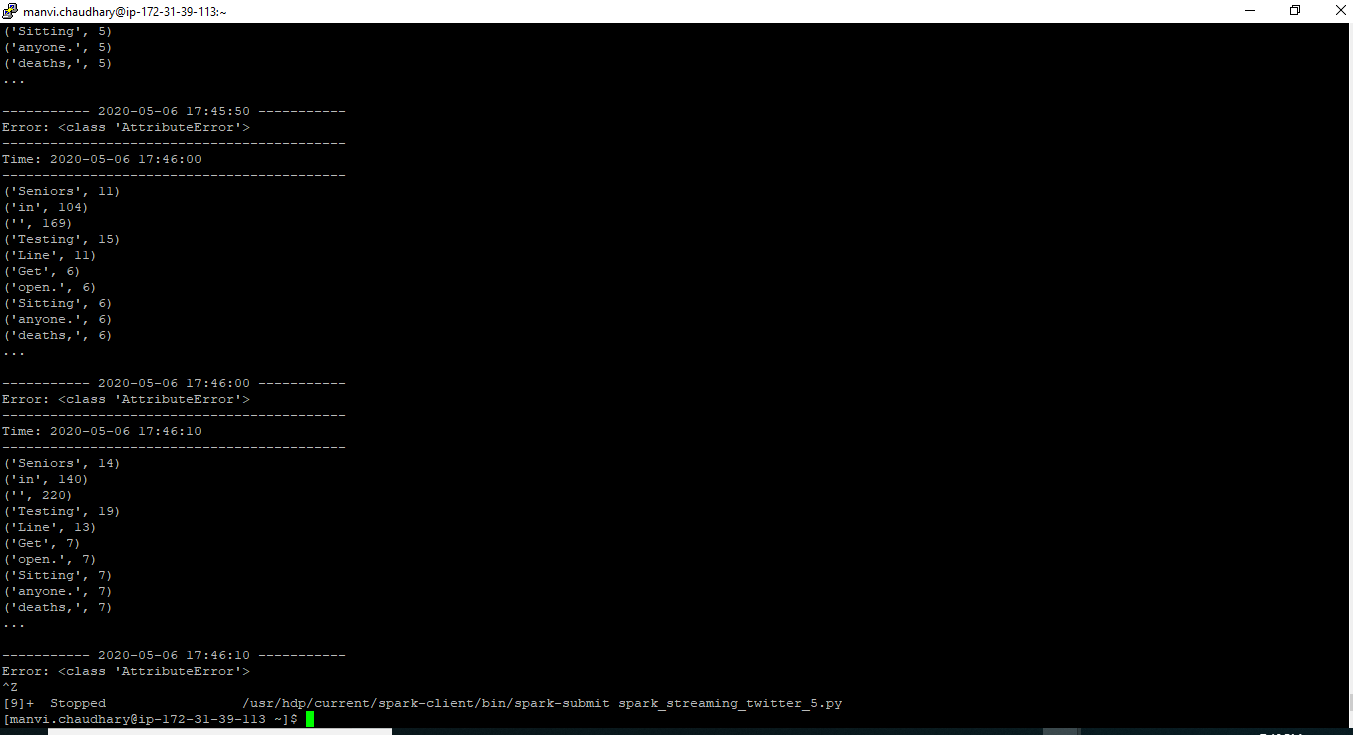


1. Instead of computing the top10 elements with Spark SQL, change the code to obtain the Top10 words (not only hashtags) using a moving window of 10 minutes every 30 seconds. Copy & paste the result.

Code to obtain top 10 words using a moving window of 10 minutes every 30 seconds is present in spark\_streaming\_twitter\_3.py

Output:





1. Use your own Access credentials. Tip: You need to create an application from your twitter account.

ACCESS\_TOKEN = '2650901850-NSjHW1FPRoUCxY3TokT4jFpfv4JS4CQpusyrW6L'

ACCESS\_SECRET = 'vlw4ILjE8w9JGbBFBtqMzDOdXWHxEZraTHGdqF4ydEevp'

CONSUMER\_KEY = 'jj4jwZNT5CknR2nsc6cOYkkdX'

CONSUMER\_SECRET = 'U3M6vvpQ530Il1SbZXqoM8P2cxUmu7dhqLxPyOxrlRhScWY8rF'