CS553 Programming Assignment#3

CloudKon clone with Amazon EC2, S3, SQS and DynamoDB

1)Introduction

A distributed task execution framework similar to CloudKon was implemented on Amazon EC2 using SQS and DynamoDB. The framework has two components. One client to push tasks into the SQS and one worker to retrieve works from the SQS and execute them. The Assignment was done using Python

2) Design

2.1) Client

The client was implemented in python and the tool can be run using the command

Python client.py <Instruction Queue> <Work_Load_File>

The Work_Load_File is generated by a command line tool

Python gnrtr.py <number of instructions> <instruction>

So if the tool is run as python gnrtr.py 10 sleep 10000 will generate a file wrkr with 10 sleep 10000 instructions

A unique id is generated in the client program and this id is send to the SQS by the client

The client generates unique id for a session using a counter. While putting the instruction into the SQS the id is appended to the message

2.2) Local Back End Worker

The Local Back End Worker is run using the command

Python local.py < number of threads>

The tool creates a pool of specified number of threads. The tool reads the instructions from the work load file and is put into a queue. The threads take these instructions from this queue, executes them and puts the status in the Response queue. In the end the status of the Response queue is checked

2.3) Remote Back End Worker

The worker tool is run using the command

Python wrkr.py <Instruction Queue> <DynamoDB> <Response Queue>

The worker takes the tasks out of the Instruction Queue, executes them, updates them in DynamoDB and Response Queue. The message is cut into two pieces. The first piece has the id and this is checked with the ids available in dynamodb. If it is not there then the instruction is executed deleted from instruction queue and then updated to dynamo db and response queue

Then we have a tool to check the response Queue which is run using the command

Python chk.py < Response Queue>

This will check the content of the Response Queue and check the status of the completed instructions

2.4) Duplicate Tasks:

Dynamo DB was used to avoid duplicate tasks. The worker after fetching values from the Instruction Queue checks whether the task is there in Dynamo DB. If it is there that means the instruction has already been executed. If it is a new task the task is executed and updated in response queue and DynamoDB

2.5) Animoto Clone

The animoto generator generates links into a file wrkr. The animoto client takes these values and is populated into the SQS. The worker takes these links from the SQS and then downloads these images. After downloading ffmpeg is used to make a video with this images

2)Manual

Step 1: Starting Instances, SQS and DynamoDB

This can be done from the local

Python strt_inst.py <number of Instances> <name of Instruction SQS> <name of Response SQS> <name of DynamoDB>

This tool will start the specified number of instances along with 2 SQS and 1 DynamoDB with the name specified

There is another tool to delete the instances and Queues

It is run as follows

Python terminate.py

This will terminate all the running instances and queues in the account. This is run just in case something goes wrong and u need to start making instances and queues again

17 instances, 1DynamoDB and 2 SQS are made using this tool

All the public DNS of instances are updated in the file worker

Step 2: Transferring Files

The following files has to be transferred to the client using scp

client.py

wrkr.py

gnrtr.py

local.py

sender

worker

chk.py

animoto.py

animoto_client.py

animoto_wrkr.py

Step 3: Transferring files to the workers

Password less ssh is set up on every instance

Python-boto is installed on every instance

Then the script sender is used to push wrkr.py and animoto_wrkr.py to all the workers

Step 4: Generating instructions

The Work_Load_File is generated by a command line tool

Python gnrtr.py < number of instructions > < instruction >

So if the tool is run as python gnrtr.py 10 sleep 10000 will generate a file wrkr with 10 sleep 10000 instructions

Step 5: Running the client

The client is run using the command

Python client.py <Instruction Queue> <Work_Load_File>

The client will take values from the work load file, generates the ids, combines the id and the message and pushes them to the Instruction Queue

Step 6: Running the Remote Workers

All the workers in the file worker is run using pssh at the same time

pssh -h wrkr -t 100000000 -l ubuntu -A -i "python wrkr.py <Instruction Queue> <DynamoDB> <Response Queue>

The worker is run using the command

Python wrkr.py <Instruction Queue> <DynamoDB> <Response Queue>

The workers will run till the instruction Queue is empty

The Response Queue is checked using the tool check.py

It is run using the following command

Python chk.py < Response Queue >

This will check the number of success messages in the given response queue

Step 7: Running the Local Worker

The worker is run using the following command

Python local.py <no of threads>

This will execute the number of instructions in the Work_Load_File and execute them using the number of threads specified

Step 8: Checking the Response Queue

The response queue is checked using the command

Python chk.py < Response Queue>

This will check the number of success messages in the response queue and this number is displayed

Step 9: Generating Links

Animoto.py is run as follows

Python animoto.py <n>

Where n is the number of workers

Step 10: Running the Animoto client

Animoto client is run as follows

Python animoto_client.py <instruction queue><Work load file>

This will make the client load the instructions form the work load file to the instruction queue

Step 11: Running the Animoto Worker

All the workers in the file worker is run using pssh at the same time

pssh -h wrkr -t 100000000 -l ubuntu -A -i "python animoto_wrkr.py <Instruction Queue> <DynamoDB> <Response Queue>

The worker is run using the command

Python animoto_wrkr.py <Instruction Queue> <DynamoDB> <Response Queue>

The workers will run till the instruction Queue is empty

3) Performance Evaluation

3.1) Efficiency

Remote:

Table showing the time taken

workers/Instruction	10s x 10	1s x 100	10ms x 1000
1	100.7198	103.5925	40.60159
2	100.7125	105.2598	40.31233
4	100.6499	105.1859	44.01281
8	100.8699	117.2482	64.98935
16	100.6964	104.3814	41.19066

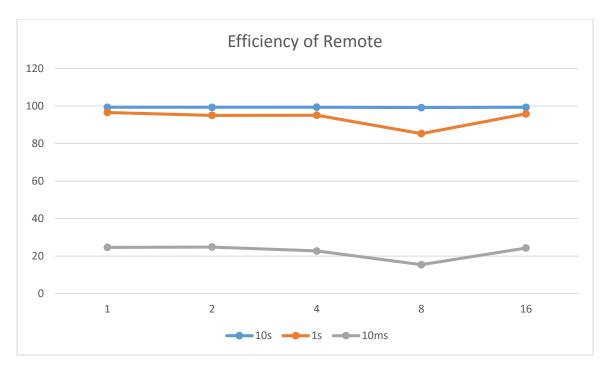
Table showing idle time taken

workers/Instruction	10s x 10	1s x 100	10ms x 1000
1	100	100	10
2	100	100	10
4	100	100	10
8	100	100	10
16	100	100	10

The efficiency is calculated as Idle time * 100/Time taken

Table showing idle time taken

workers/Instruction	10s x 10	1s x 100	10ms x 1000
1	99.28531	96.53208	24.62958
2	99.29259	95.00303	24.8063
4	99.35434	95.06974	22.72066
8	99.13758	85.28919	15.38714
16	99.30842	95.80252	24.27735



It can be observed that the Efficiency remains almost constant and we can conclude that the Cloudkon is scalable

It can also be observed that the efficiency is very less for 10ms instructions. This could be because of the high number of instructions being fetched and written into the db and SQS. So it can be concluded that as the number of instructions increases the efficiency decreases but is highly scalable

Local: Table showing the time taken

Threads/Instructions	10s x 10	1s x 100	10ms x 1000
1	100.104	100.1499	10.357
2	200.2079	200.298	20.722
4	400.4077	400.6006	41.469
8	800.829	801.1957	82.8923
16	1601.652	1602.389	165.774

Table showing idle time taken

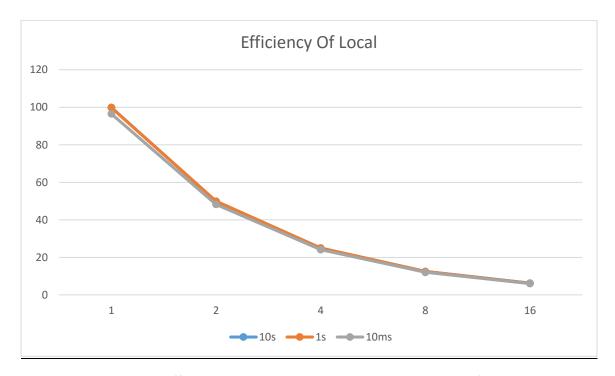
Threads/Instructions	10s x 10	1s x 100	10ms x 1000
1	100	100	10
2	100	100	10
4	100	100	10
8	100	100	10

16	100	100	10
10	100	100	

The efficiency is calculated as Idle time * 100/Time taken

Table showing idle time taken

Threads/Instructions	10s x 10	1s x 100	10ms x 1000
1	99.89611	99.85032	96.55306
2	49.94809	49.92562	48.25789
4	24.97454	24.96252	24.1144
8	12.48706	12.48135	12.06385
16	6.243553	6.240682	6.032309



It can be noted that the efficiency in local is not dependent on the number of instructions submitted

And it can also be noted that the efficiency decreases as the number of threads increases. This is because of the reason that this experiment was done a t2.micro system that has only one cpu and one core. So using threads will slow down the system instead if increasing the efficiency

3.2) Throughput

Remote:

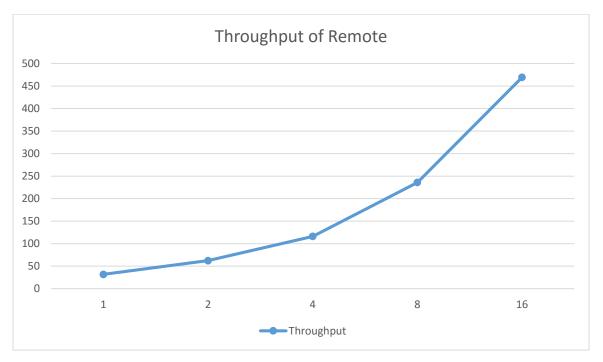
Local:

Time Taken for Executing 10k Sleep 0 Tasks is as follows

Worker/Instructions	10s x 10
1	315.326
2	160.7
4	86.131
8	42.4
16	21.303

Throughput is calculated as no of instructions/ Time taken

Workers/	10s x 10
Instructions	
1	31.71321
2	62.22775
4	116.1022
8	235.8491
16	469.4175



It can be noted that the throughput increases as the number of workers are increased. This will show that the system is truly distributional and the total work is being efficiently divided and thus adding more workers is going to increase the processing. Thus we can conclude that the system is scalable

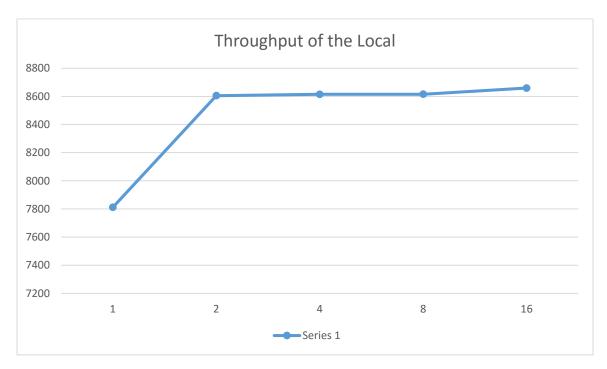
Local:

Time Taken for Executing 10k Sleep 0 Tasks is as follows

Threads/Instructions	10s x 10
1	12.801
2	11.6211
4	11.6087
8	11.6079
16	11.549

Throughput is calculated as no of instructions/ Time taken

Threads/Instructions	10s x 10
1	7811.89
2	8605.037
4	8614.23
8	8614.823
16	8658.758

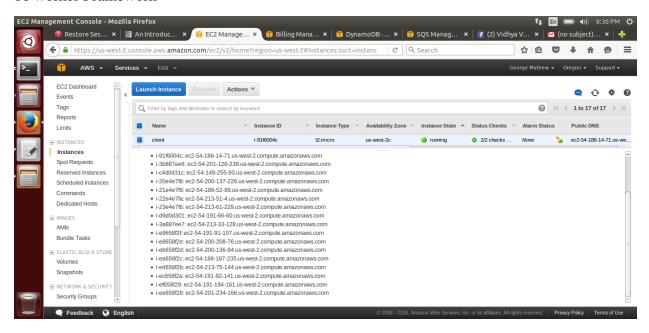


There is an increase in throughput when there are two threads but after that the throughput remains more or less the same. So it can be concluded that in a t2.micro system two threads can

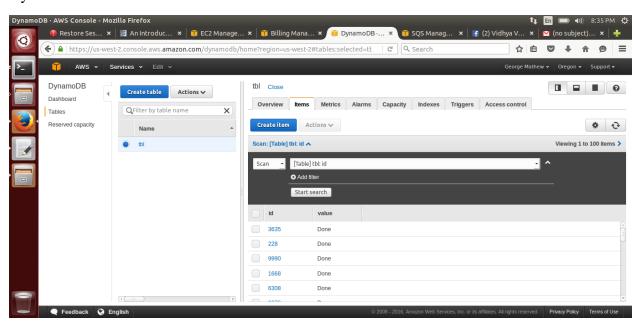
handle instruction read and write better than one thread. In case of one thread the resources are not being fully utilized

3)Screenshots

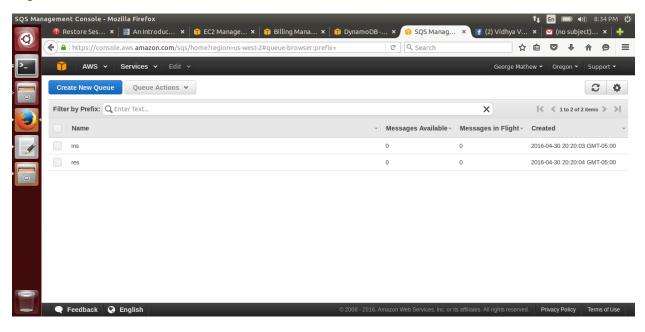
16 worker Framework



DynamoDB

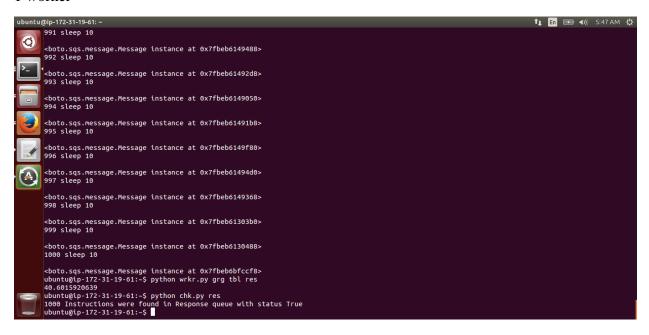


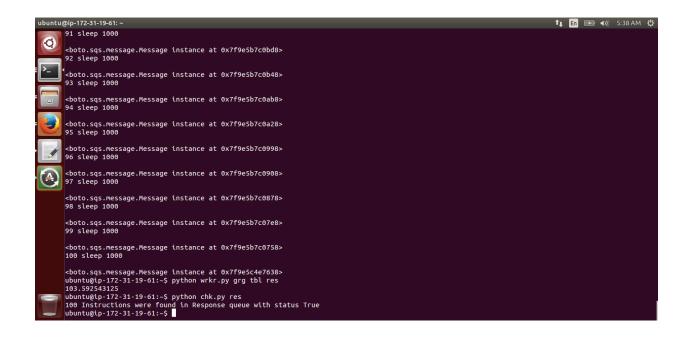
SQS

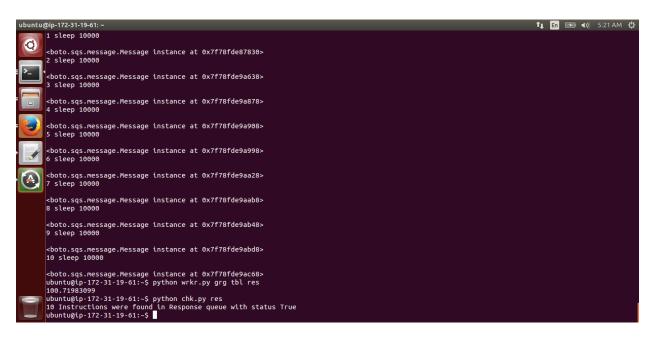


Experiments:

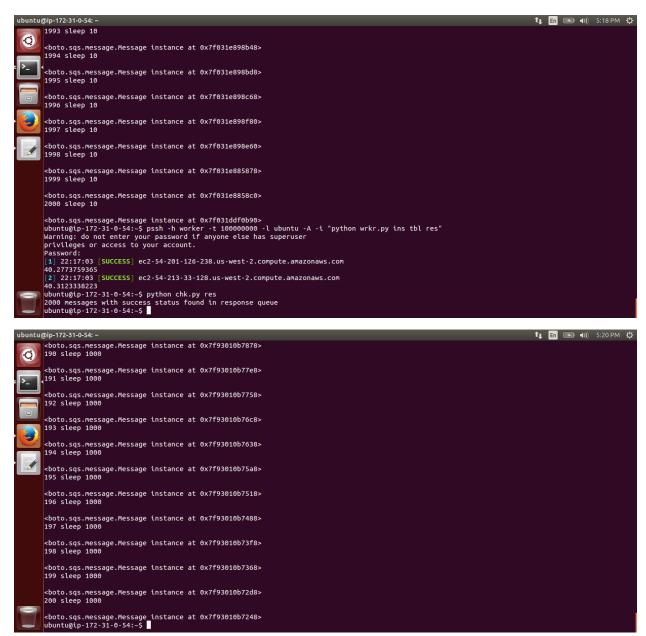
1 worker

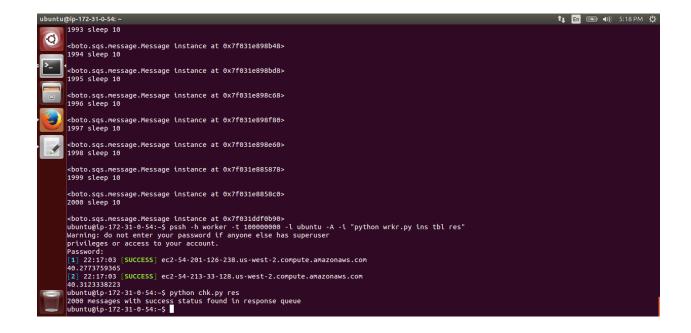




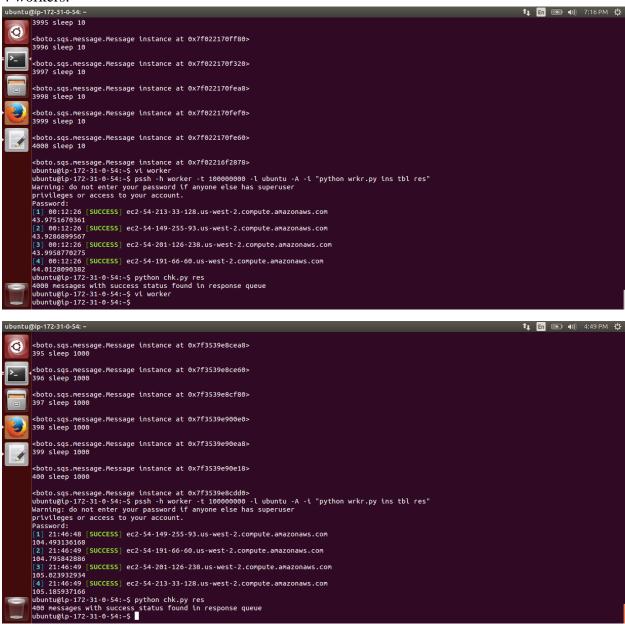


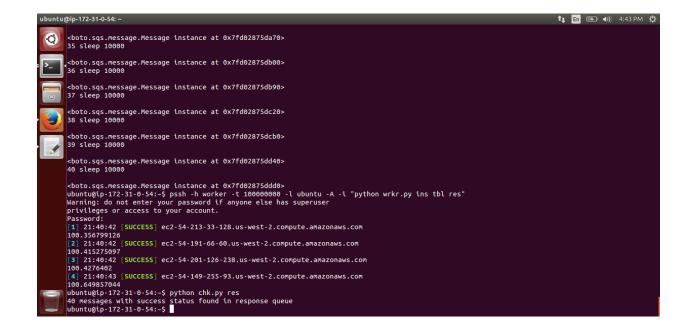
2 Workers



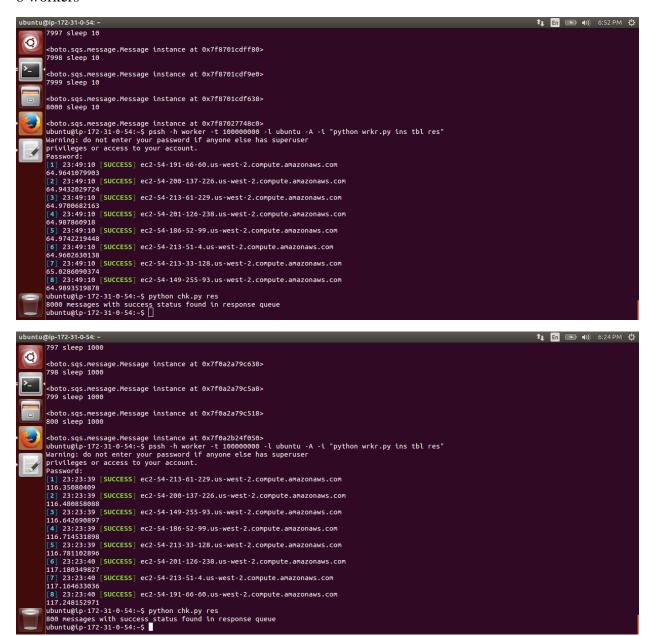


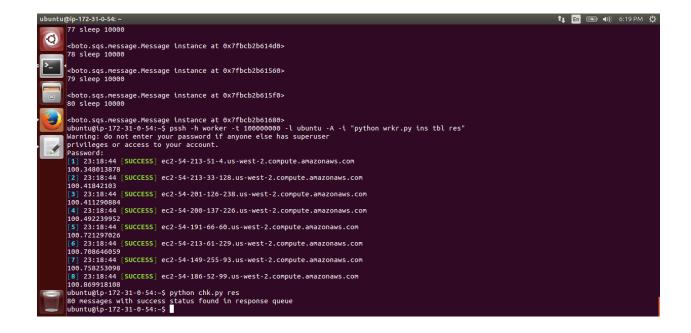
4 workers:



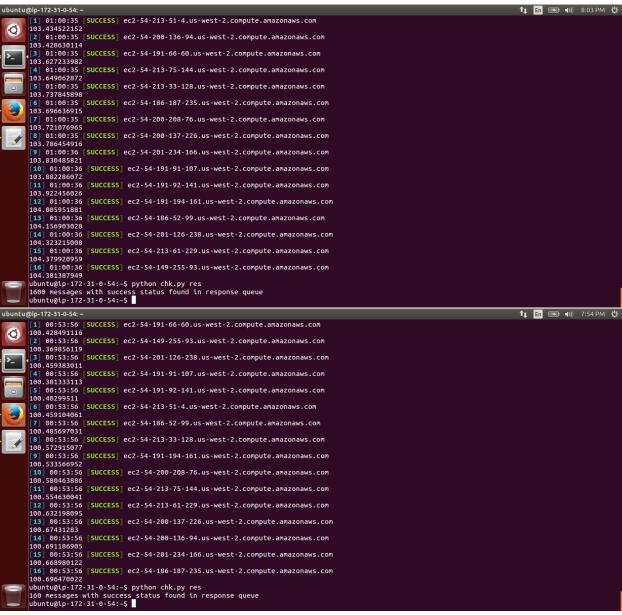


8 workers





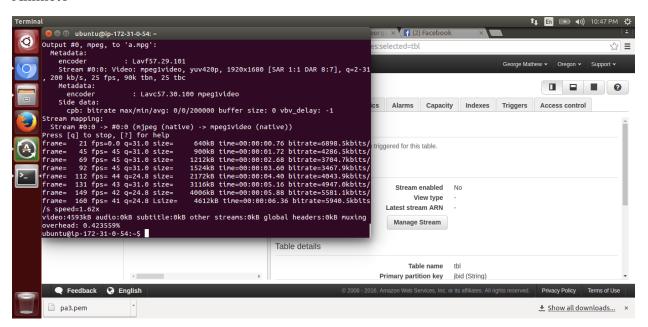
16 workers

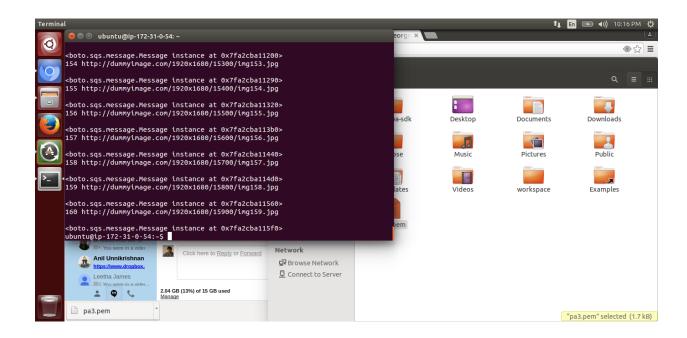


```
ubuntu@ip-172-31-0-54:
                                                                                                                                                                                                                                                                     t₁ En 🗩 •0)) 8:32 PM 😃
            [1] 01:24:04 [SUCCESS] ec2-54-186-187-235.us-west-2.compute.amazonaws.com
41.1072471142
[2] 01:24:04 [SUCCESS] ec2-54-201-234-166.us-west-2.compute.amazonaws.com
 (0)
             41.0683569908
[3] 01:24:04 [SUCCESS] ec2-54-186-52-99.us-west-2.compute.amazonaws.com
              41.1201379299
            41.12013/9299
[4] 01:24:04 [SUCCESS] ec2-54-200-136-94.us-west-2.compute.amazonaws.com
41.1075668335
[5] 01:24:04 [SUCCESS] ec2-54-213-61-229.us-west-2.compute.amazonaws.com
41.1525688171
[6] 01:24:04 [SUCCESS] ec2-54-201-126-238.us-west-2.compute.amazonaws.com
             41.1407089233
              11.140/109233
[7] 01:24:04 [SUCCESS] ec2-54-191-92-141.us-west-2.compute.amazonaws.com
41.1025049686
[8] 01:24:04 [SUCCESS] ec2-54-200-208-76.us-west-2.compute.amazonaws.com
             [9] 01:24:04 [SUCCESS] ec2-54-191-91-107.us-west-2.compute.amazonaws.com
             41.1328999996
             [10] 01:24:04 [SUCCESS] ec2-54-191-66-60.us-west-2.compute.amazonaws.com
41.2061648369
            [11] 01:24:04 [SUCCESS] ec2-54-200-137-226.us-west-2.compute.amazonaws.com
41.1446881294
[12] 01:24:04 [SUCCESS] ec2-54-191-194-161.us-west-2.compute.amazonaws.com
             41.1259419918
            41.1259419918
[13] 01:24:04 [SUCCESS] ec2-54-213-75-144.us-west-2.compute.amazonaws.com
41.0879609585
[14] 01:24:04 [SUCCESS] ec2-54-213-51-4.us-west-2.compute.amazonaws.com
41.174424867
[15] 01:24:04 [SUCCESS] ec2-54-213-33-128.us-west-2.compute.amazonaws.com
             41.2241380215
           41.2241380213 [SUCCESS] ec2-54-149-255-93.us-west-2.compute.amazonaws.com
41.1906019072

ubuntu@tp-172-31-0-54:~$ python chk.py res
16000 messages with success status found in response queue
ubuntu@tp-172-31-0-54:~$
```

Animoto





Local



