AWS Question (AWSQ) Documentation.

run time on ec2 t2-large machine (8 gb RAM) →
`5.576925992965698`

run time on local machine (mac book pro intel i7 32 gb RAM)
→ `2.7405030727386475`

comment:

laboratory machine configuration

- first the **ALLv1-59257** lab was started, following that aws cli was configured and installed on my pc by entering the SDKs provided by the lab into the ~/.aws/credentials file
- below was downloaded the ssh key also made available by the lab, labuser.pem
- below was launched the lab, once ready you can access aws console

ec2 machine configuration

- for ec2 machine, go to ec2, "start instance", from here you can enter a name, choose operating system(ubuntu), instance type (t2-large), select key pair.pem to access ssh, provided already by the lab; enter network specifications, the default vpc with subnet-a was chosen, from the way the network of the default vpc is designed all subnets have free access to the internet (internet gateway and routing tables properly connected), for access to the instance is 3 security group firewall rules were opened :
  - port 22 - TCP protocol - ssh type - Origin my-ip-address

- port 80 - TCP protocol - http type - Origin 0.0.0.0/0
- port 8080 - TCP protocol - TCP custom type - Origin 0.0.0.0/0

In addition, an EBS root disk was mounted, of 32 Gb gp2 (general purpose) IOPS 100/3000

in the end the "LabRole" IAM role was assigned to the machine

configuration bucket s3

- To allow the dataset to be read to the ec2 machine, the dataset on my local machine and immediately uploaded to a s3 bucket created specifically to host the python code and dataset (list.json), command used is

```
aws s3 cp list.json s3://dataset-diego --region us-east-1
aws s3 cp most_popular_tags.ipynb s3://dataset-diego/code --region us-east-1
```

alternatively to load dataset and code into the machine you could use the SCP command

Machine application configuration machine

access with ssh

```
chmod 400 labuser.pem
ssh -i "labuser.pem" ubuntu@ec2-3-222-189-25.compute-1.amazonaws.com
```

machine upgrades have been made

```
sudo apt-get update
```

has been installed pip

```
sudo apt-get install python3-pip
```

aws cli was installed

```
pip install awscli
```

is being created the dataset hammer

```
mkdir -p dataset
```

datasets were downloaded from the s3 bucket

```
aws s3 cp s3://dataset-diego/code .
aws s3 cp s3://dataset-diego/list.json ./dataset
```

jupyter notebook installed

```
pip install notebook
```

command was run to launch jupyter notebook and leave it listening

```
jupyter notebook --no-browser
```

in another terminal was given the command

```
zx
```

if the security groups have been configured correctly we will be able to connect to the jupyter notebook hosted on our ec2 machine, at port http://localhost:8088/

we are now able to launch our script

opening the data set required the use of chunksize, otherwise the ec2 machine would suffer by going to shut down the process and restarting the machine.

Running this script on this virtual machine turns out to have worse times, compared to my local pc

most_popular_tags_ec2.ipynb

most_popular_tags.ipynb