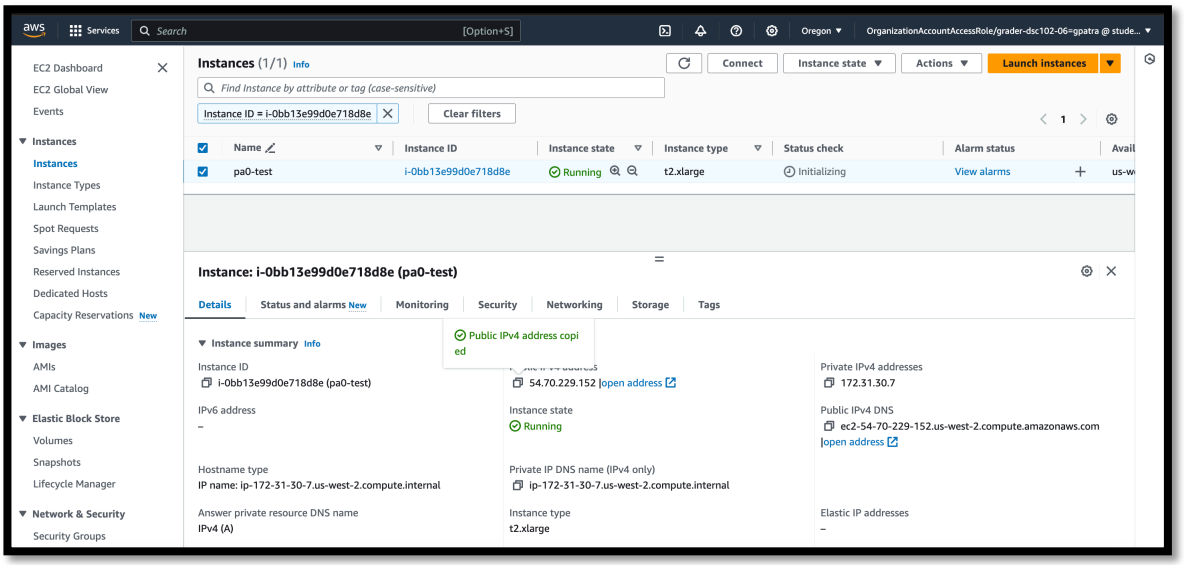
DSC 102 PA0 System Setup Tutorial

Step 1 : Setup Client i.e Jupyter Notebook and Port Forwarding for Jupyter Notebook onto localhost.

ETS LINK : https://ets-apps.ucsd.edu/individual/DSC102_WI24_A00

a) After creating your EC2 instance note down its IP address as shown below



b) Open a Terminal Window and do the following:

i) **Change permission of key file :** chmod 400 dask-key.pem

ii) **SSH Into the Scheduler EC2 Instance:** ssh -i dask-key.pem ubuntu@54.70.229.152

iii) **Activate the Dask Environment:** source dask_env/bin/activate

```
(base) golokeshpatra@Golokeshs-MacBook-Air Downloads % ssh -i gp-pa1.pem ubuntu@
54.70.229.152
The authenticity of host '54.70.229.152 (54.70.229.152)' can't be established.
ED25519 key fingerprint is SHA256:7n+v/xnnMgZGTFhnsVY7ckN7X4qjSDcTy9en70znips.
This key is not known by any other names.
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added '54.70.229.152' (ED25519) to the list of known hosts.
Welcome to Ubuntu 22.04.2 LTS (GNU/Linux 5.19.0-1025-aws x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

  System information as of Mon Jan  8 05:53:36 UTC 2024

  System load:  0.080078125       Processes:             120
  Usage of /:   6.9% of 48.27GB   Users logged in:       0
  Memory usage: 1%                IPv4 address for eth0: 172.31.30.7
  Swap usage:   0%


Expanded Security Maintenance for Applications is not enabled.

124 updates can be applied immediately.
63 of these updates are standard security updates.
To see these additional updates run: apt list --upgradable

7 additional security updates can be applied with ESM Apps.
Learn more about enabling ESM Apps service at https://ubuntu.com/esm


The list of available updates is more than a week old.
To check for new updates run: sudo apt update

Last login: Wed Oct  4 03:10:23 2023 from 76.53.230.125
ubuntu@ip-172-31-30-7:~$ source dask_env/bin/activate
```

iv) Launch Jupyter Notebook on the EC2:

jupyter notebook --port=8888

```
(dask_env) ubuntu@ip-172-31-13-215:~$ jupyter notebook --port=8888
[I 12:49:50.842 NotebookApp] Writing notebook server cookie secret to /home/ubuntu/.local/
share/jupyter/runtime/notebook_cookie_secret

  _   _                       _       _
 | | | |_ __   __| |__ _| |_ ___
 | |_| | '_ \/ _` / _` |  _/ -_)
  \___/| .__/\__,_\__,_|\__\___|
       |_|

Read the migration plan to Notebook 7 to learn about the new features and the actions to t
ake if you are using extensions.

https://jupyter-notebook.readthedocs.io/en/latest/migrate_to_notebook7.html

Please note that updating to Notebook 7 might break some of your extensions.

[I 12:49:54.030 NotebookApp] Serving notebooks from local directory: /home/ubuntu
[I 12:49:54.030 NotebookApp] Jupyter Notebook 6.5.4 is running at:
[I 12:49:54.030 NotebookApp] http://localhost:8888/?token=ace5c6b61bf24461ee067412cea642df
40a5809e38cbe756
[I 12:49:54.030 NotebookApp]  or http://127.0.0.1:8888/?token=ace5c6b61bf24461ee067412cea6
42df40a5809e38cbe756
[I 12:49:54.030 NotebookApp] Use Control-C to stop this server and shut down all kernels (
twice to skip confirmation).
[W 12:49:54.035 NotebookApp] No web browser found: could not locate runnable browser.
[C 12:49:54.035 NotebookApp]

    To access the notebook, open this file in a browser:
        file:///home/ubuntu/.local/share/jupyter/runtime/nbserver-7696-open.html
    Or copy and paste one of these URLs:
        http://localhost:8888/?token=ace5c6b61bf24461ee067412cea642df40a5809e38cbe756
     or http://127.0.0.1:8888/?token=ace5c6b61bf24461ee067412cea642df40a5809e38cbe756
```

v) Copy the link to the Jupyter Server (shown in the last line of the above screenshot) You can paste it in the browser AFTER performing step (c) below

c) Open **New Terminal Window** and run the following command:

DSC 102 PA0 System Setup Tutorial

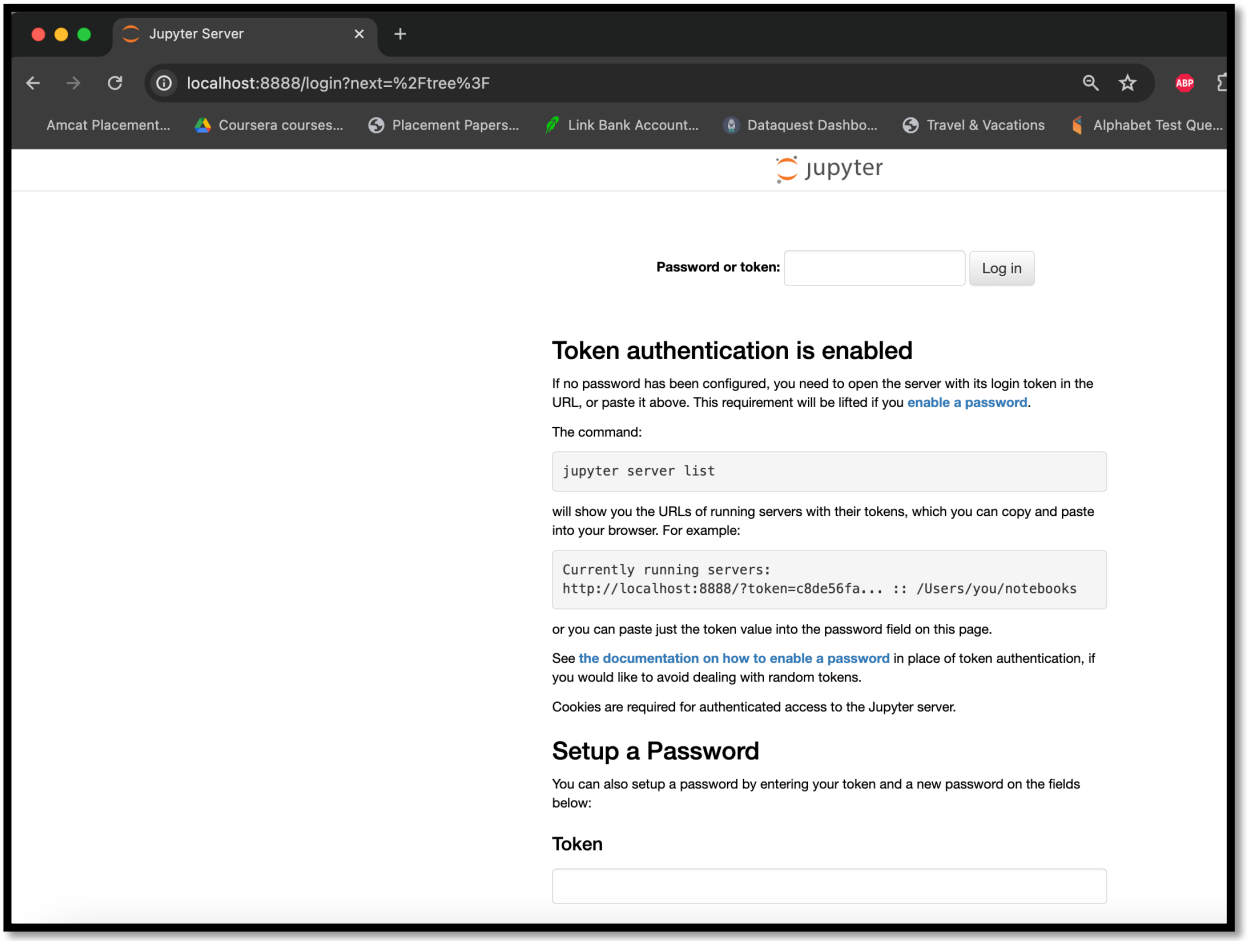
- **Port Forwarding Jupyter Notebook running on port 8888 on the EC2 to port 8888 on local system:**

  ssh -i dask-key.pem ubuntu@54.70.229.152 -L 8888:localhost:8888

```
(base) golokeshpatra@Golokeshs-MacBook-Air Downloads % ssh -i gp-pa
1.pem ubuntu@54.70.229.152 -L 8888:localhost:8888
Welcome to Ubuntu 22.04.2 LTS (GNU/Linux 5.19.0-1025-aws x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

  System information as of Mon Jan  8 05:59:06 UTC 2024

  System load:  0.013671875       Processes:             128
  Usage of /:   6.9% of 48.27GB   Users logged in:       1
  Memory usage: 2%                IPv4 address for eth0: 172.31.30.
7
  Swap usage:   0%


Expanded Security Maintenance for Applications is not enabled.

124 updates can be applied immediately.
63 of these updates are standard security updates.
To see these additional updates run: apt list --upgradable

7 additional security updates can be applied with ESM Apps.
Learn more about enabling ESM Apps service at https://ubuntu.com/es
m


The list of available updates is more than a week old.
To check for new updates run: sudo apt update

Last login: Mon Jan  8 05:57:06 2024 from 49.37.113.40
ubuntu@ip-172-31-30-7:~$
```

II) **Verify the port forwarding by -**

Go to your web-browser -> localhost:8888

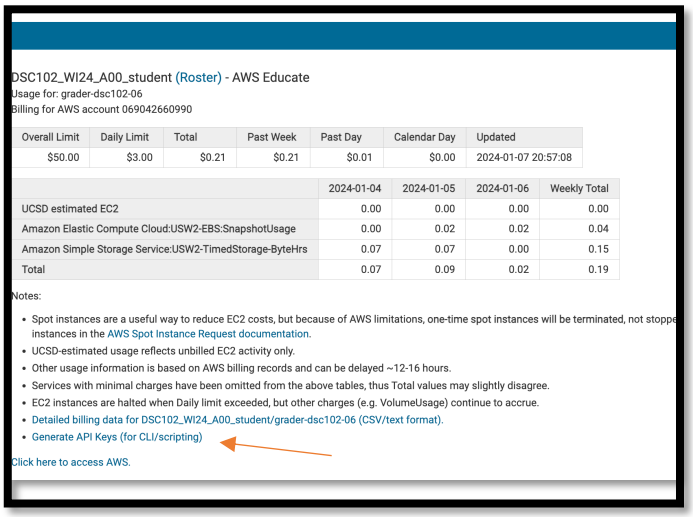


DSC 102 PA0 System Setup Tutorial

Step 2 : Download data from S3

a) In the most recently opened Terminal Window:

- Copy and paste the AWS ACCESS KEY ID, AWS SECRET ACCESS KEY, and AW SSESSION TOKEN



DSC102_WI24_A00_student (Roster) - AWS Educate
Usage for: grader-dsc102-06
Billing for AWS account 069042660990

Overall Limit	Daily Limit	Total	Past Week	Past Day	Calendar Day	Updated
$50.00	$3.00	$0.21	$0.21	$0.01	$0.00	2024-01-07 20:57:08

	2024-01-04	2024-01-05	2024-01-06	Weekly Total
UCSD estimated EC2	0.00	0.00	0.00	0.00
Amazon Elastic Compute Cloud:USW2-EBS:SnapshotUsage	0.00	0.02	0.02	0.04
Amazon Simple Storage Service:USW2-TimedStorage-ByteHrs	0.07	0.07	0.00	0.15
Total	0.07	0.09	0.02	0.19

Notes:

- Spot instances are a useful way to reduce EC2 costs, but because of AWS limitations, one-time spot instances will be terminated, not stoppe
  instances in the AWS Spot Instance Request documentation.
- UCSD-estimated usage reflects unbilled EC2 activity only.
- Other usage information is based on AWS billing records and can be delayed ~12-16 hours.
- Services with minimal charges have been omitted from the above tables, thus Total values may slightly disagree.
- EC2 instances are halted when Daily limit exceeded, but other charges (e.g. VolumeUsage) continue to accrue.
- Detailed billing data for DSC102_WI24_A00_student/grader-dsc102-06 (CSV/text format).
- Generate API Keys (for CLI/scripting)

Click here to access AWS.

```
export AWS_ACCESS_KEY_ID=ASIARAE2DMZ7KAVAULIC
export AWS_SECRET_ACCESS_KEY=mF+NccIXCEk2eBdV6ArHZsRVriJt+PU7Tn/X+dqB
export
AWS_SESSION_TOKEN=FwoGZXIvYXdzEEcaDNFWCSWK3sD941eL+CK8AdFpUBaMuj1RGLbCi9S4lDSs6sljcSl2CmFi23S7/45glOuImybuiTWsnRn88H4w7qX+SfrQbNxDnw5aAe11v/aZyAS/nYkU10kFhgcXEiBkV/328QLRdbjoqwPQaYeJ/3G0HkB7BHRu1NTnPmOX3riGQAWzXoKGtpgjd92b+8Eei
H1Pp+pnuDurkNZnf9Itk9hgKxyiqPQe+cMtcE1yJVoohpEZ2sI36VxnmnW2O/fgQ6SC1ZEcOkOpqalxKL+L7qwGMi1gdYYQPyI0Ufngyhh0IH/NYDapOktFJysWtPgsCoe/jKA4c/ZrZLpPpRMB2Ws=
```

i) **Download all the files from the S3 :**

aws s3 sync s3://dsc102-public /home/ubuntu/

```
ubuntu@ip-172-31-13-215:~$ aws s3 sync s3://dsc102-public /home/ubuntu/
download: s3://dsc102-public/PA0.py to ./PA0.py
download: s3://dsc102-public/OutputSchema_PA0.json to ./OutputSchema_PA0.json
download: s3://dsc102-public/results_PA0.json to ./results_PA0.json
download: s3://dsc102-public/user_reviews.csv to ./user_reviews.csv
ubuntu@ip-172-31-13-215:~$
```

jupyter

File View Settings Help

Files Running

Select items to perform actions on them. New Upload

/

Name	Last Modified	File Size
dask_env	3 months ago	
dask_demo_notebook.ipynb	7 minutes ago	55.8 KB
Untitled.ipynb	29 seconds ago	2.7 KB
expected_results_PA0.json	3 months ago	678 B
PA0.py	7 minutes ago	2.8 KB
user_reviews.csv	3 months ago	26.6 GB

Now, on navigating to the link copied at the end of Step 1 b), you should see the following. You can now create a new notebook and are ready to code up.

Step 3 : Dask UI Port forwarding

a) Open a **New Terminal Window** and run the following command:

i) Port Forwarding the Dask dashboard UI running on port 8787 on the EC2 to port 8787 on local system: ssh -i dask-key.pem ubuntu@54.70.229.152 -L 8787:localhost:8787