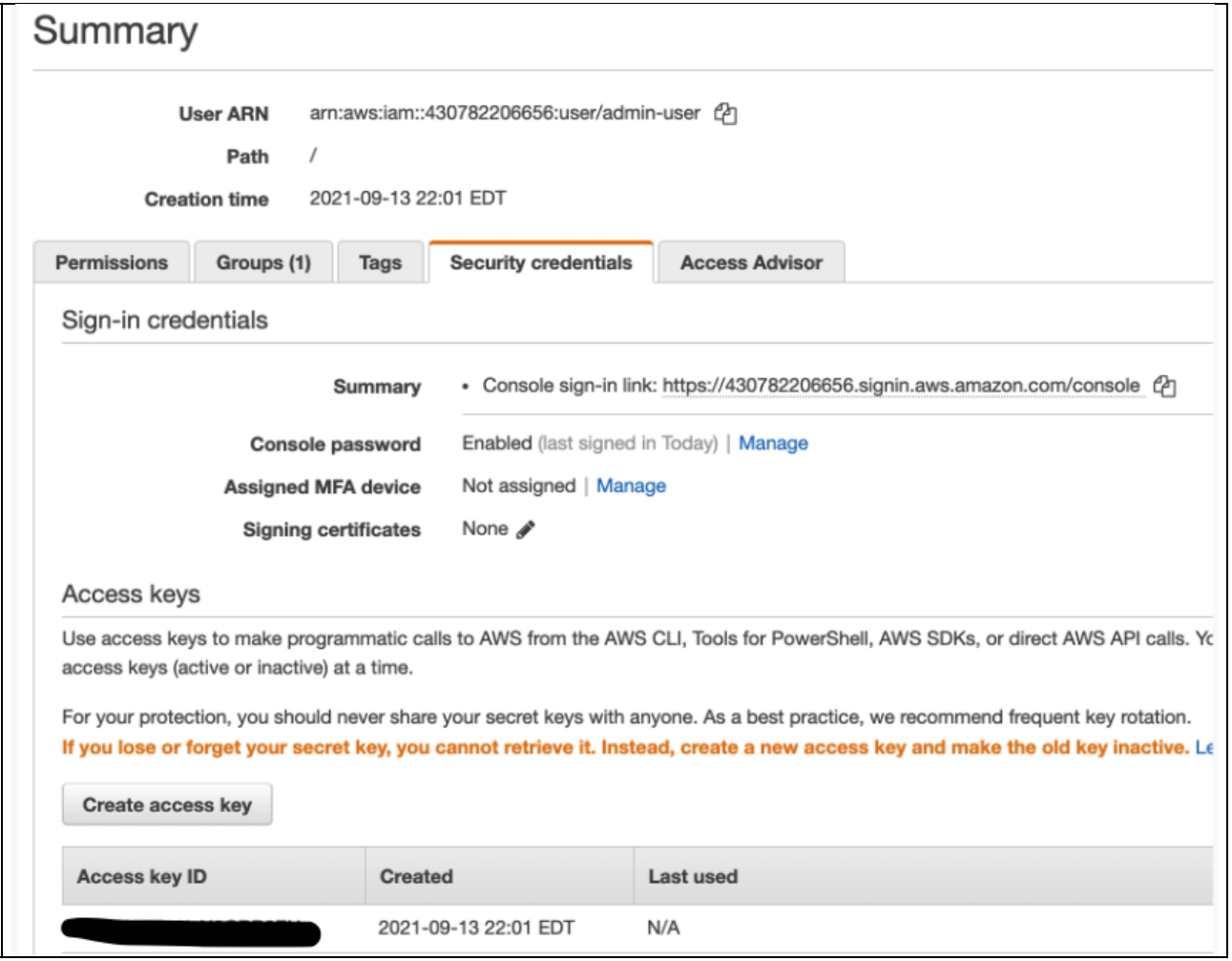
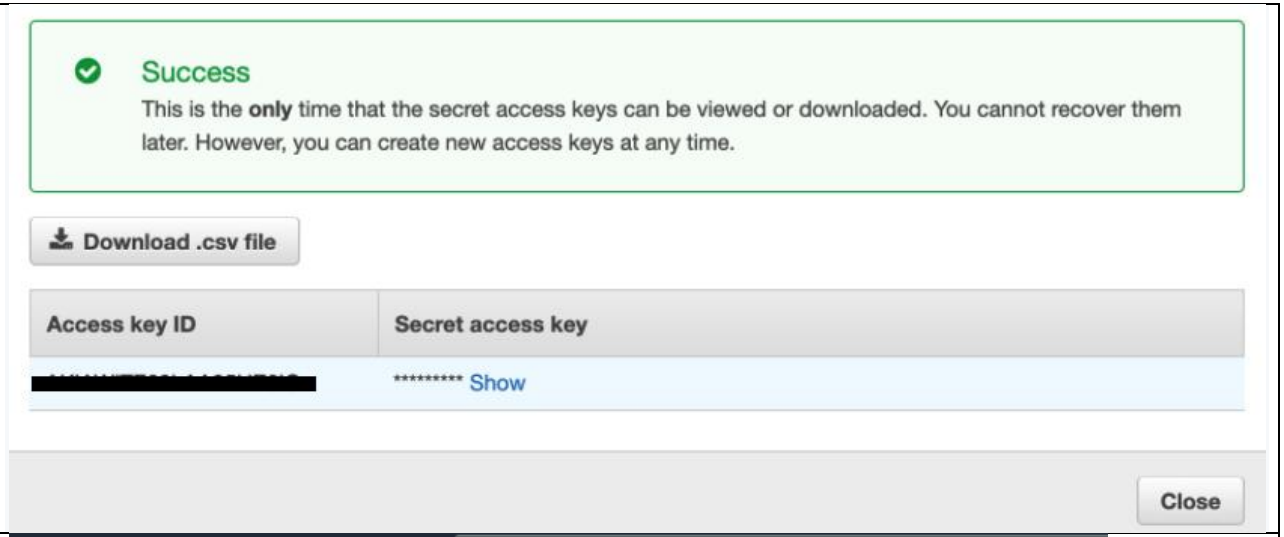
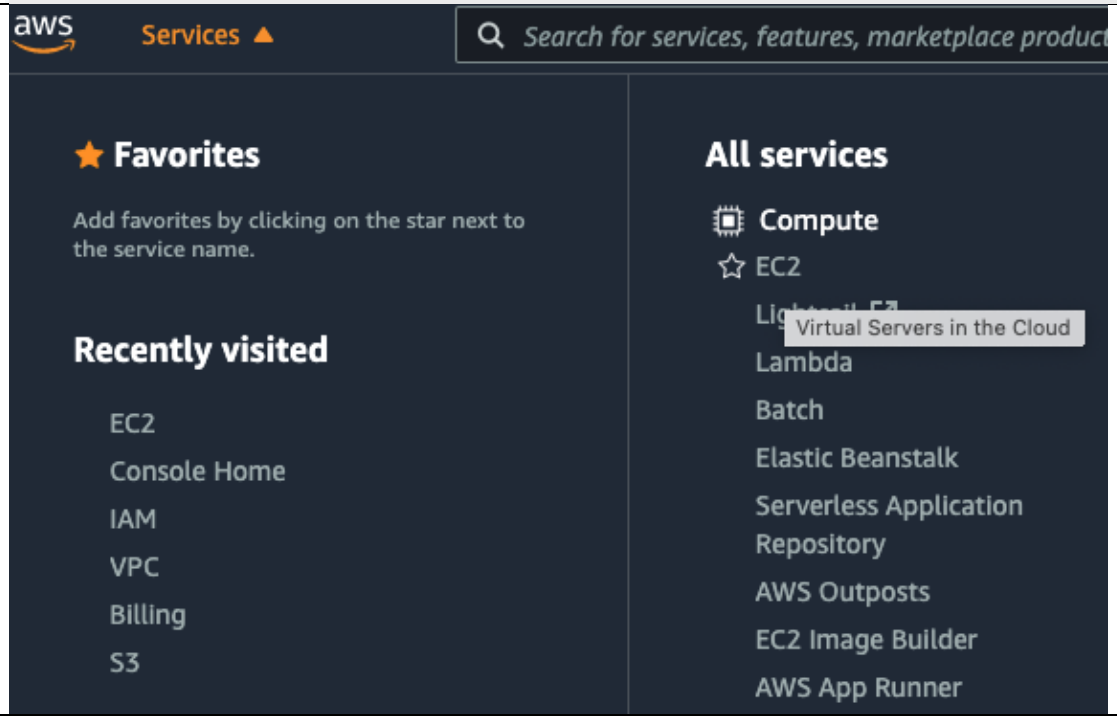
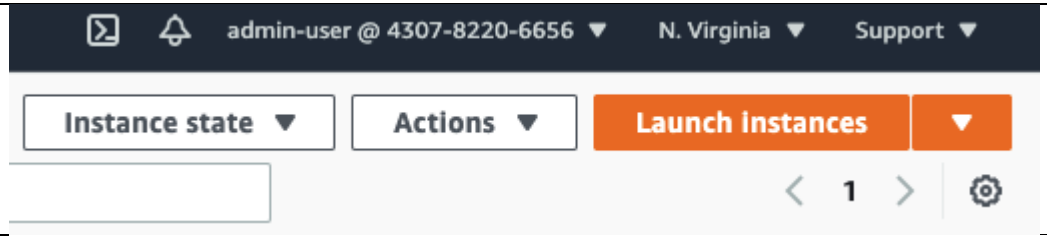
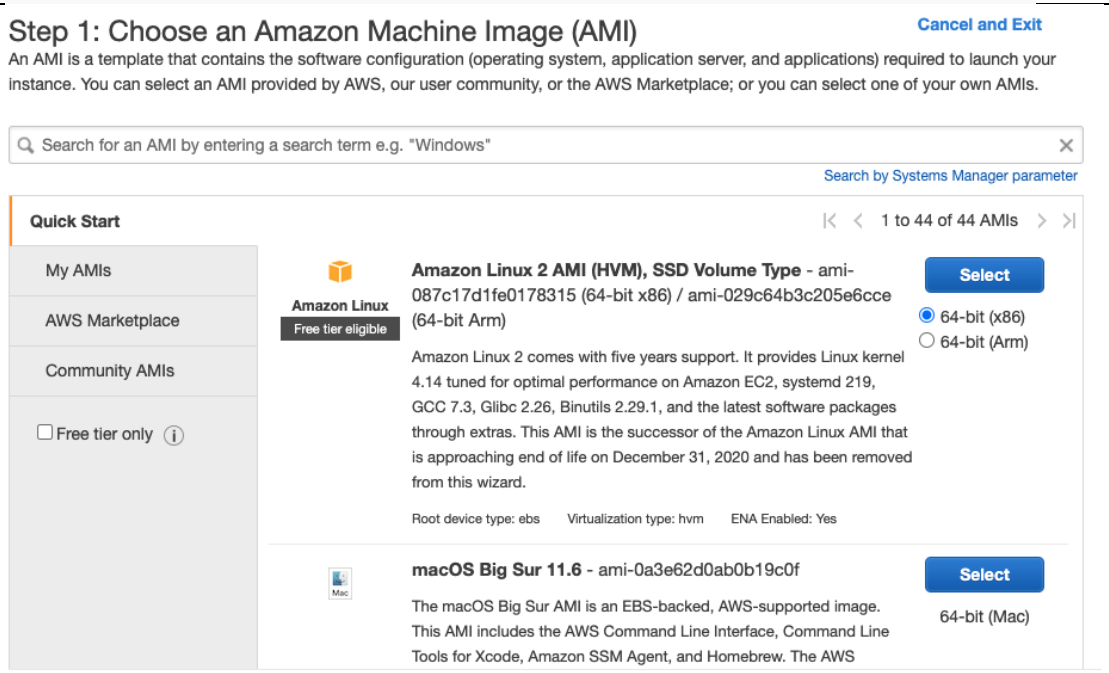
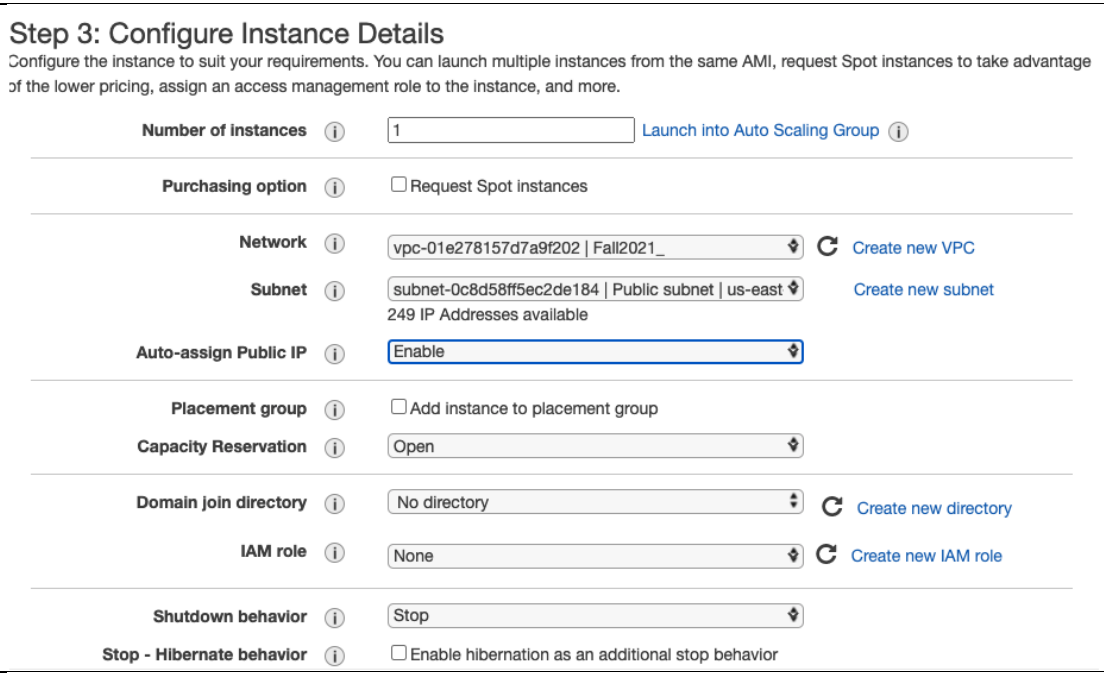
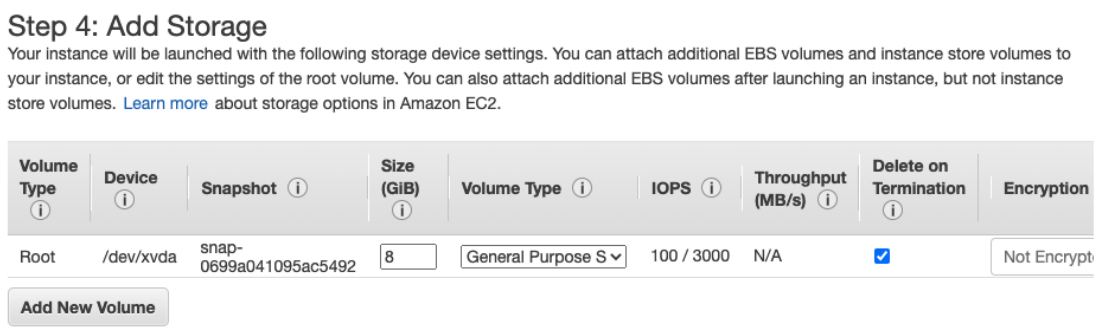
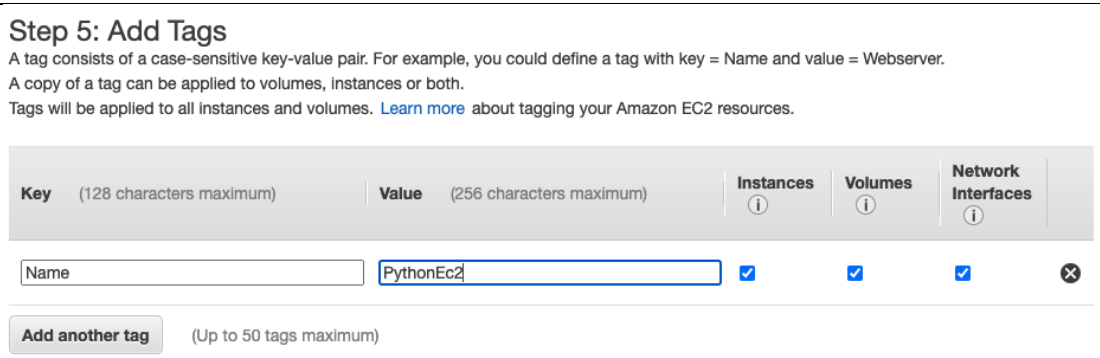
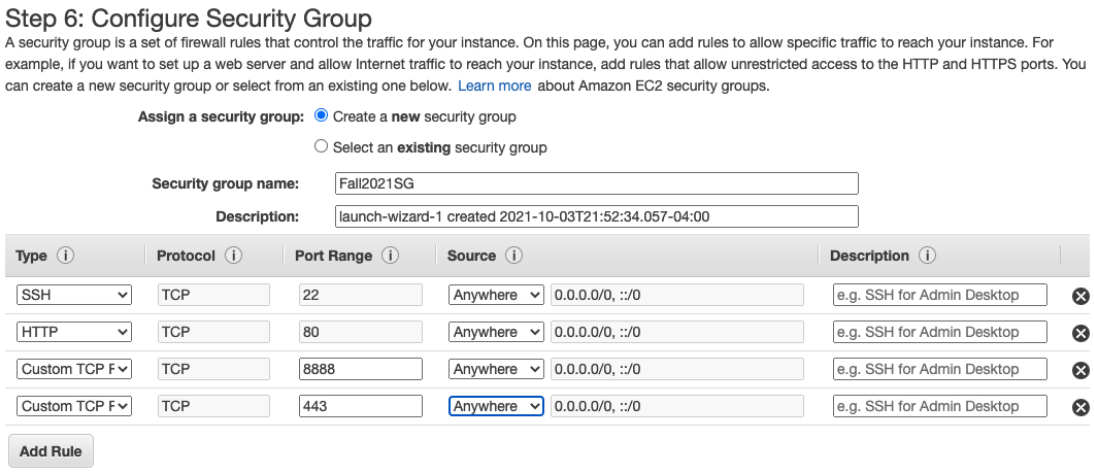
Team member	Sisi Zhang, Harish Ram
Login into you AWS account Do not use your root account!	

Go to AWS Console, service IAM, and select the user you want to allow to create S3 from an EC2	

Create an access key and keep in a safe place. You can download it as a csv file or copy and past both the access key ID and Secrete Access Key in a safe place	
Then, select Services menu option EC2	

Launch Instance. Make sure you are inside the US East (N Virginia) region	
select Amazon Lunix AMI 2018.03.0	
Select Free Tier	

Make sure you lunch it in the Public Subnet of the VPC you have created previous. In addition, enable Auto-Assign Public IP. If you do not do so, you will not be able to access it later on via the internet.	
Accept the default configuration of storage	

Name the EC2, e.g., PythonEc2	
Create a new Security groups and make sure the following ports are open from anywhere ● 22 (port used by SSH) ● 80 (Port used by HTTP) ● 8888 (Port used by Jupiter) ● 443 (Port used by Boto)	

Review and lunch EC2

Step 7: Review Instance Launch

Please review your instance launch details. You can go back to edit changes for each section. Click **Launch** to assign a key pair to your instance and complete the launch process.

Improve your instances' security. Your security group, Fall2021SG, is open to the world.

Your instances may be accessible from any IP address. We recommend that you update your security group rules to allow access from known IP addresses only.

You can also open additional ports in your security group to facilitate access to the application or service you're running, e.g., HTTP (80) for web servers. Edit security groups

AMI Details

Edit AMI

Amazon Linux 2 AMI (HVM), SSD Volume Type - ami-087c17d1fe0178315

Free tier eligible

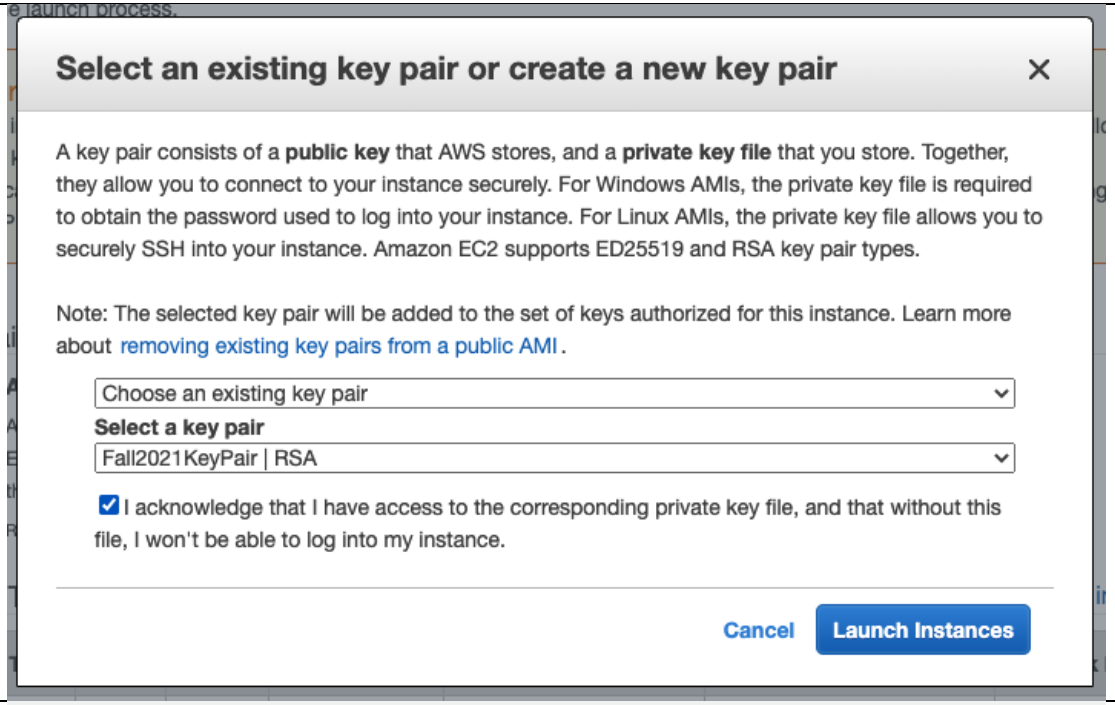
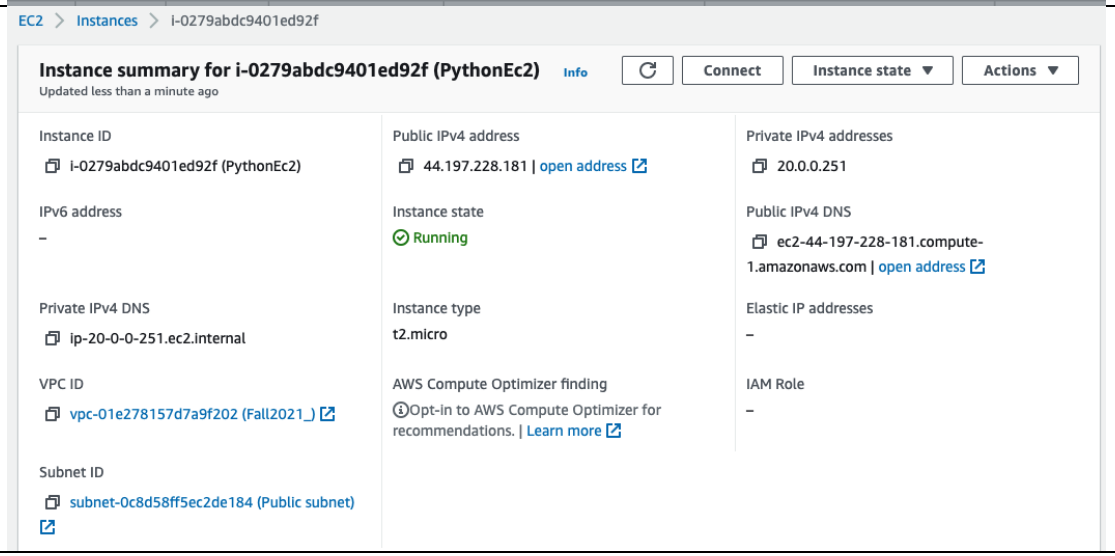
Amazon Linux 2 comes with five years support. It provides Linux kernel 4.14 tuned for optimal performance on Amazon EC2, systemd 219, GCC 7.3, Glibc 2.26, Binutils 2.29.1, and the latest software packages through extras. This AMI is the successor of the Amazon Linux AMI that is a...

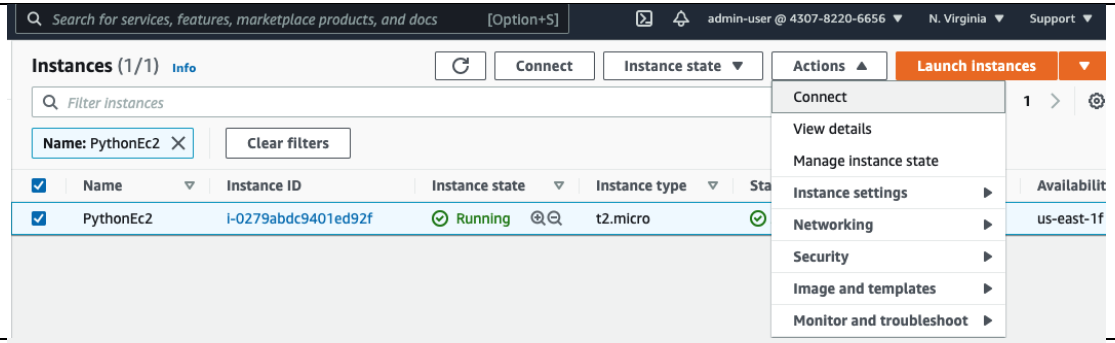
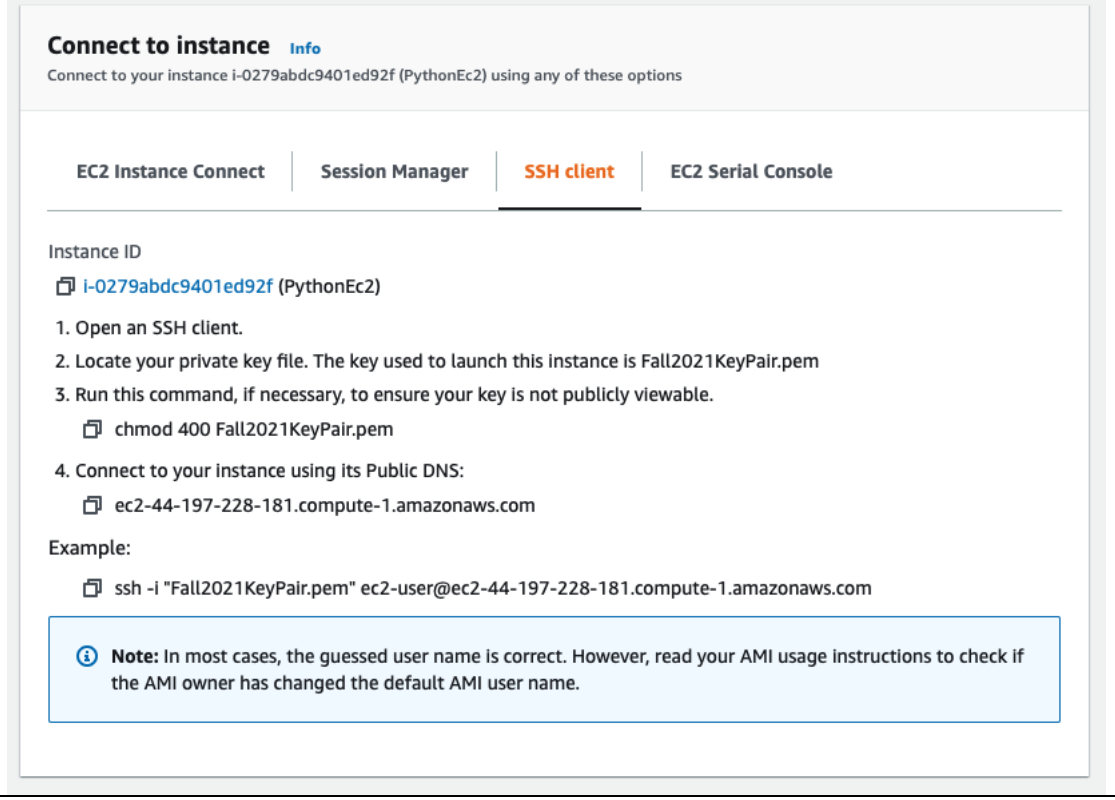
Root Device Type: ebs Virtualization type: hvm

Instance Type

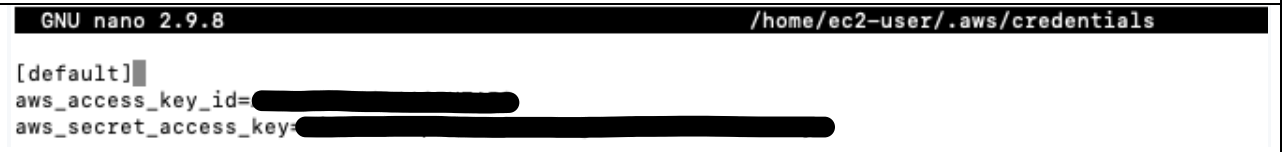
Edit instance type

Instance Type	ECUs	vCPUs	Memory (GiB)	Instance Storage (GB)	EBS-Optimized Available	Network Performance
t2.micro	-	1	1	EBS only	-	Low to Moderate

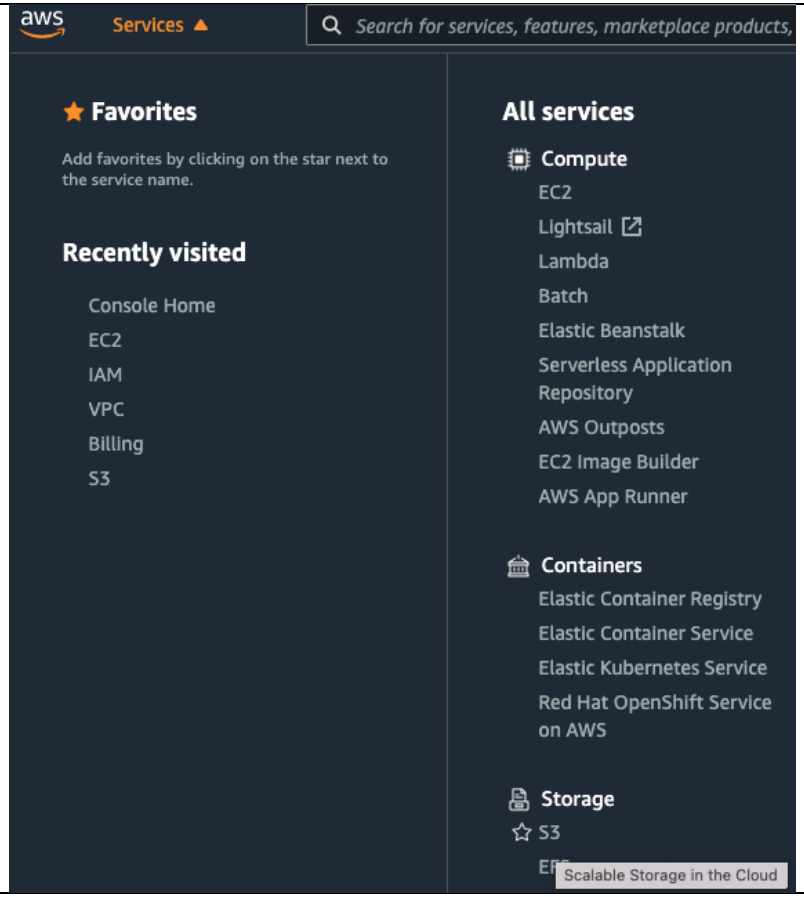
If you do not have a PEM Key, create one. If you have already one PEM key, use it. Make sure you actually have it! And Lunch the instance	
Your new instance will show up in the EC2 dashboard Copy the Public IP address. You will need it to login into your new EC2	

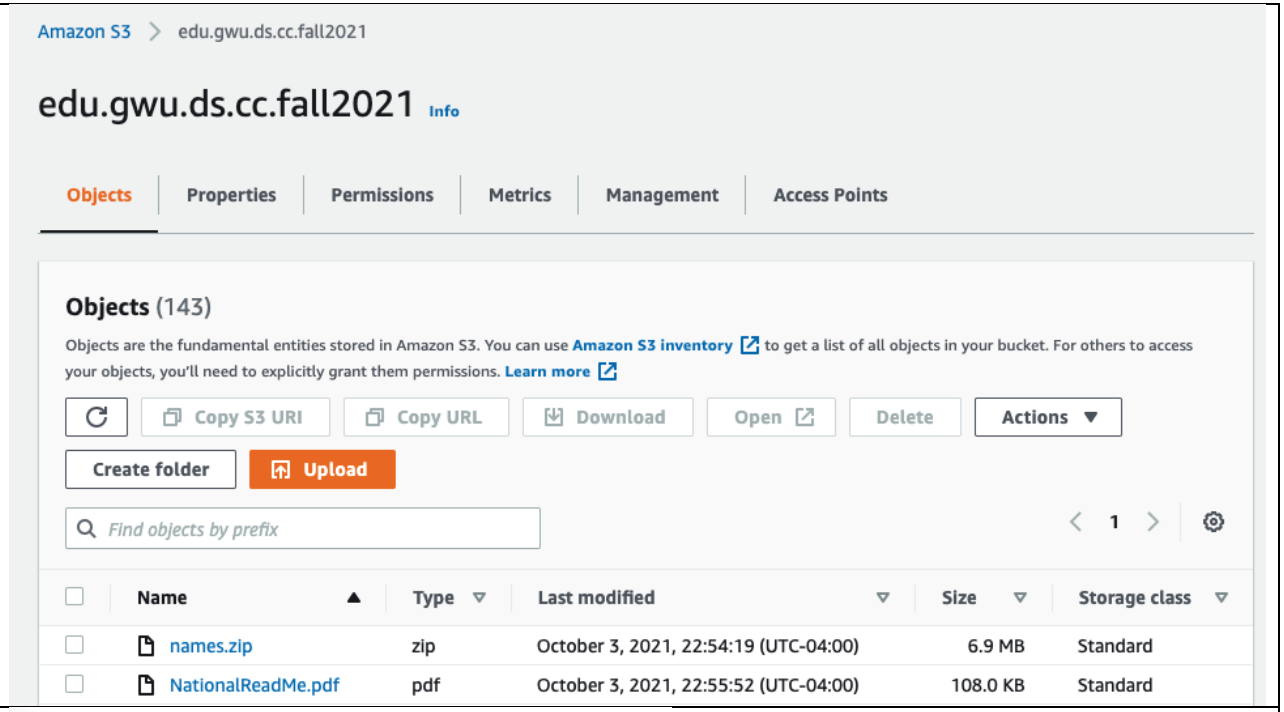
Select the instance in the EC2 dashboard and select Action, Connect	
Use the instructions provided to login into your EC2 If you have a Windows Laptop, you will have to use PuTTY: follow the link for instructions If you have a Lunix or Mac Laptop, you must open a terminal and ssh into your new EC2	

Ssh to your EC2 First time you ssh into your EC2, you will get a message asking if you want to continue, type yes	(base) sisizhang@Sisis-MacBook-Pro key % sudo ssh -i "Fall2021KeyPair.pem" ec2-user@ec2-44-197-228-181.compute-1.amazonaws.com Password: Sorry, try again. Password: The authenticity of host 'ec2-44-197-228-181.compute-1.amazonaws.com (44.197.228.181)' can't be established. ECDSA key fingerprint is SHA256:4Vcxhp3ttHSVrOBrhUER9A2cKG1wmk1nuycpnbIZtNs. Are you sure you want to continue connecting (yes/no/[fingerprint])?
As soon as you ssh into your EC2, update it by issuing the command: sudo yum update	__| __|_) _| (/ Amazon Linux 2 AMI ___|___|___| https://aws.amazon.com/amazon-linux-2/ 11 package(s) needed for security, out of 35 available Run "sudo yum update" to apply all updates. [ec2-user@ip-20-0-0-251 ~]$ sudo yum uodate Loaded plugins: extras_suggestions, langpacks, priorities, update-motd No such command: uodate. Please use /bin/yum --help [ec2-user@ip-20-0-0-251 ~]$ sudo yum update
Then execute the following commands to install s3fs-fuse: this would allow you to mount a S3 bucket as a Lunix folder (directory) If you have launched another type of EC2, e.g., Ubuntu, please check installation instructions: https://github.com/s3fs-fuse/s3fs-fuse/wiki/Installation-Notes	[ec2-user@ip-20-0-0-251 s3fs-fuse]$ sudo sed -i 's/enabled=0/enabled=1/' /etc/yum.repos.d/epel.repo [ec2-user@ip-20-0-0-251 s3fs-fuse]$ sudo yum install -y gcc libstdc++-devel gcc-c++ fuse fuse-devel curl-devel libxml make openssl-devel git make [ec2-user@ip-20-0-0-251 s3fs-fuse]$ git clone https://github.com/s3fs-fuse/s3fs-fuse [ec2-user@ip-20-0-0-251 s3fs-fuse]$ cd s3fs-fuse/ [ec2-user@ip-20-0-0-251 s3fs-fuse]$ ./autogen.sh [ec2-user@ip-20-0-0-251 s3fs-fuse]$ ./configure --prefix=/usr --with-openssl [ec2-user@ip-20-0-0-251 s3fs-fuse]$ make [ec2-user@ip-20-0-0-251 s3fs-fuse]$ sudo make install
Create a file to keep your AWS credential. You can use nano, vi, vim, or any editor available in your EC2. You can also use your laptop text editor, e.g., notepad, notepad++,	[ec2-user@ip-20-0-0-251 s3fs-fuse]$ mkdir ~/.aws [ec2-user@ip-20-0-0-251 s3fs-fuse]$ nano ~/.aws/credentials

and then sftp your credentials file into your EC2.	
Provide the credentials you created and downloaded beforehand, e.g., [default] aws_access_key_id=AKIAIOSFODNN7EXAMPLE aws_secret_access_key=wJalrXUtnFEMI/K7MDENG/bPxRfiCYEXAMPLEKEY If you use nano, to save your changes issue the following commands: control-o control-x	
Then change the access to your credential file	`[ec2-user@ip-20-0-0-251 s3fs-fuse]$ chmod 640 ~/.aws/credentials`
Create a bucket using AWS CLI. Note that you will not be able to name your bucket as edu.gwu.ds.cc.fall2010 Since I have already done so. Bucket names are unique. You will need to pick a name that does not exit. In this class, you will use the following name rule edu.gwu.ds.cc.YourGNumber	`[ec2-user@ip-20-0-0-251 s3fs-fuse]$ aws s3api create-bucket --bucket edu.gwu.ds.cc.fall2021 --region us-east-1`

YourGNumber is your GWU G#. In this way, it is unlike that somebody else names a bucket in this way	
Create a local directory that will be synchronized with your AWS S3 Bucket The option -p will create the upper level directories if they do not exist. This is just an example, you give any folder name that is consistent with Lunix directory name rules.	`[ec2-user@ip-20-0-0-251 s3fs-fuse]$ mkdir -p ~/edu/gwu/ds/cc/fall2021`
Synchronize your local folder with your aws s3 bucket	`[ec2-user@ip-20-0-0-251 s3fs-fuse]$ s3fs 'edu.gwu.ds.cc.fall2021'  ~/edu/gwu/ds/cc/fall2021 -o use_path_request_style`
Download SSA baby name file into your local folder: ~/edu/gwu/ds/cc/fall2020	`[ec2-user@ip-20-0-0-251 s3fs-fuse]$ cd ~/edu/gwu/ds/cc/fall2021` `[ec2-user@ip-20-0-0-251 fall2021]$ wget https://www.ssa.gov/oact/babynames/names.zip`
Unzip names.zip	`[ec2-user@ip-20-0-0-251 fall2021]$ unzip names.zip`
Using AWS CLI, list the files available in the AWS S3	`[ec2-user@ip-20-0-0-251 fall2021]$ aws s3 ls s3://edu.gwu.ds.cc.fall2021`

Now open AWS console and check the files that are inside edu.gwu.ds.cc.fall2010 bucket		

Navigate to the folder where you have uploaded and unzipped the names.zip file and you will see the same list of baby names per year	
Now let us write a small Python program to manipulate buckets and objects using Python Boto library Before you do so, you must install Python3. Normally, the simple "sudo yum install python3" should work, however, sometimes it does not.	[ec2-user@ip-20-0-0-251 ~]$ python3 --version Python 3.7.10
Update pip. To do so:	curl -O https://bootstrap.pypa.io/get-pip.py python3 get-pip.py -user echo 'export PATH=~/.local/bin:$PATH' >> ~/.bashrc

<table>
<tr><td>

1. Download the installation script from pypa.io.
2. Execute the script
3. Add pip3 to path

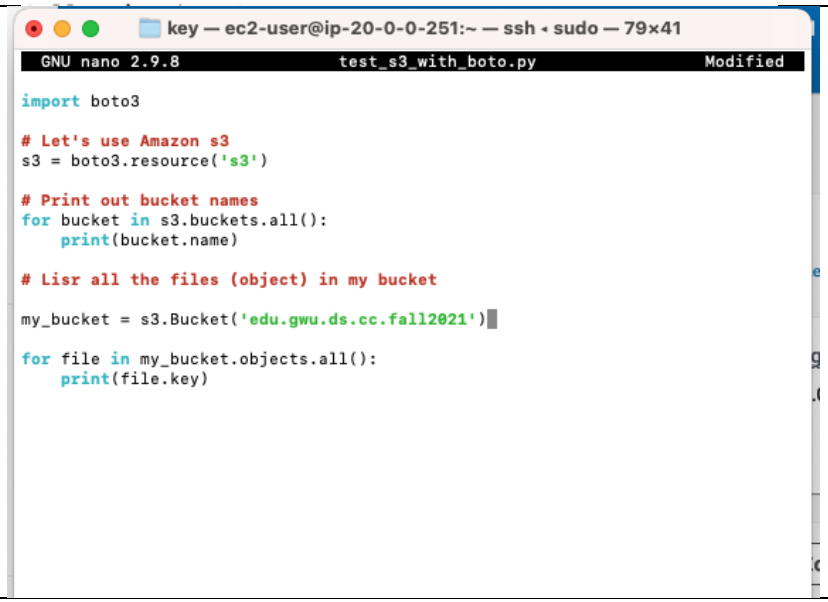
</td><td>

```
source ~/.bashrc
[ec2-user@ip-20-0-0-251 ~]$ curl -O https://bootstrap.pypa.io/get-pip.py
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100 1901k  100 1901k    0     0  34.3M      0 --:--:-- --:--:-- --:--:-- 34.3M
[ec2-user@ip-20-0-0-251 ~]$ python3 get-pip.py --user
[[ec2-user@ip-20-0-0-251 ~]$ echo 'export PATH=~/.local/bin:$PATH' >> ~/.bashrc
[[ec2-user@ip-20-0-0-251 ~]$ source ~/.bashrc
```

</td></tr>
<tr><td>Install Boto</td><td>

```
[ec2-user@ip-20-0-0-251 ~]$ pip3 install boto3 --user
```

</td></tr>
<tr><td>Using your preferred text editor create a python program to list your buckets and the files inside the bucket that you have upload previously. Name this Python program: test_s3_with_boto.py</td><td>



</td></tr>
<tr><td>Make your Python program executable</td><td>

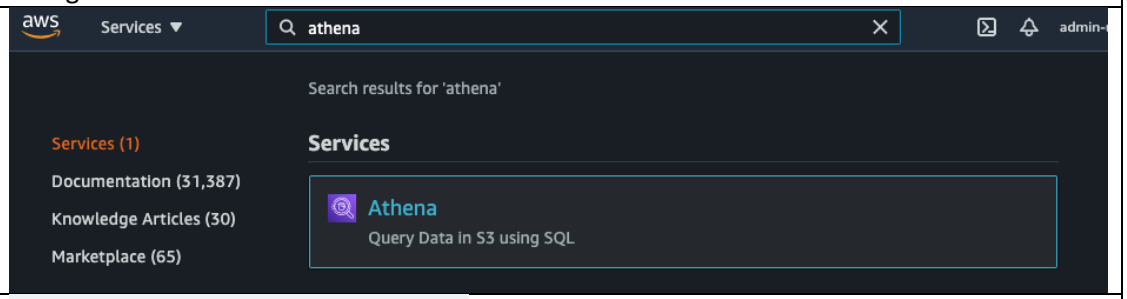
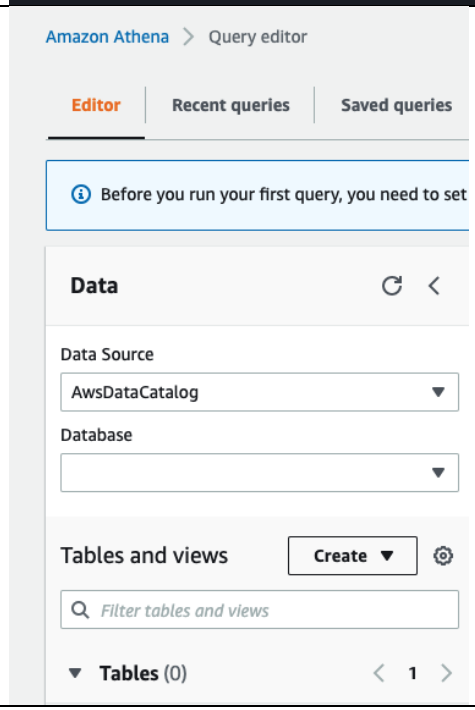
```
[ec2-user@ip-20-0-0-251 ~]$ chmod +x test_s3_with_boto.py
```

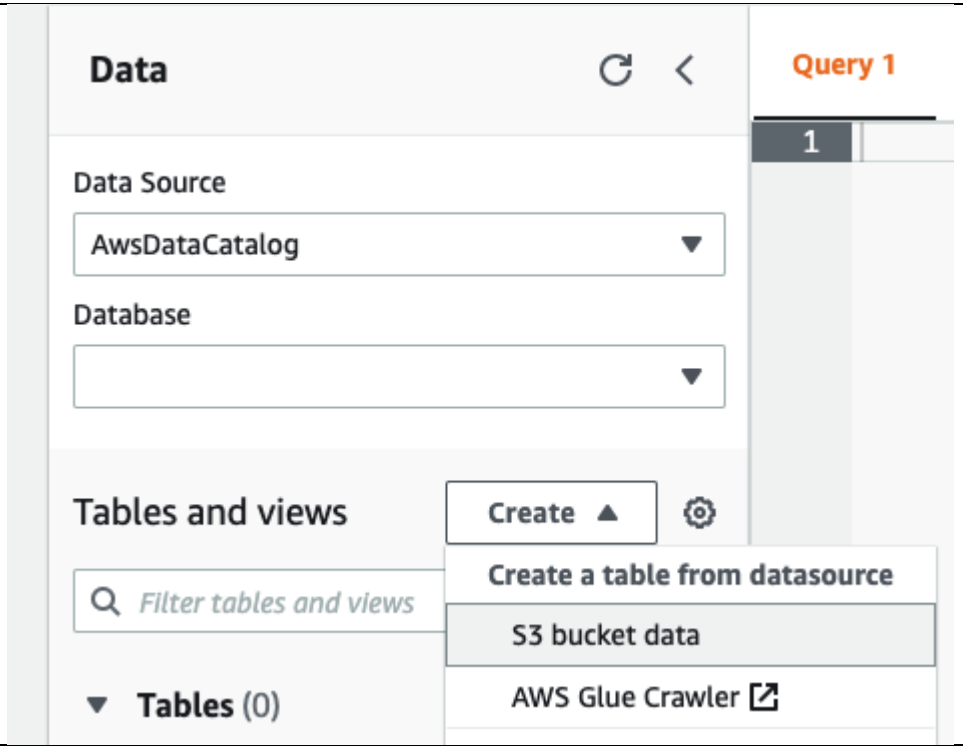
</td></tr>
<tr><td>Execute your test_s3_with_boto.py</td><td>

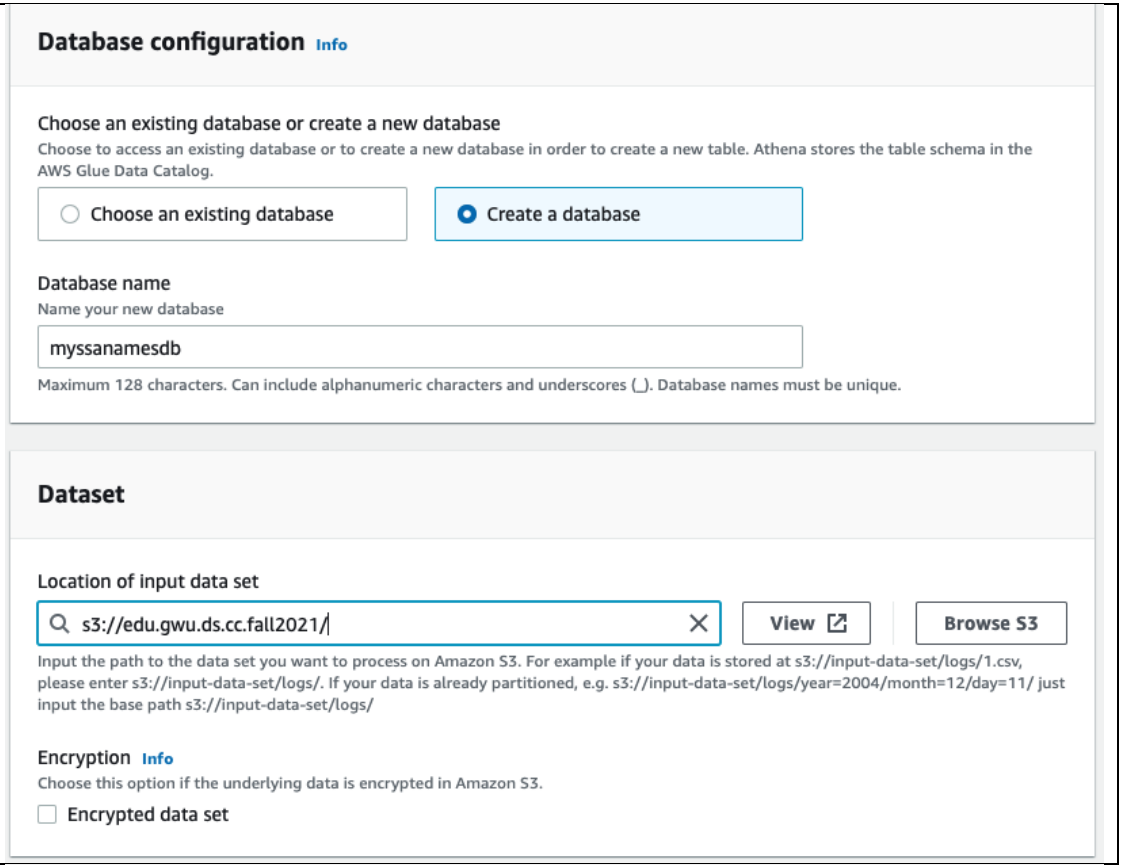
```
[[ec2-user@ip-20-0-0-251 ~]$ python3 test_s3_with_boto.py
edu.gwu.ds.cc.fall2021
NationalReadMe.pdf
names.zip
yob1880.txt
yob1881.txt
yob1882.txt
yob1883.txt
```

</td></tr>
<tr><td>Congratulations! You made it!</td><td></td></tr>
</table>

Continue to play with Boto3. Upload a file into your bucket, then download another file from the bucket. Documentation is available in the following URL: https://boto3.amazonaws.com/v1/documentation/api/latest/index.html	

Now, we will query the names uploaded as S3 files using Athena. Check documentation about Athena	
Open the console and find Athena	
Select the option Create table	

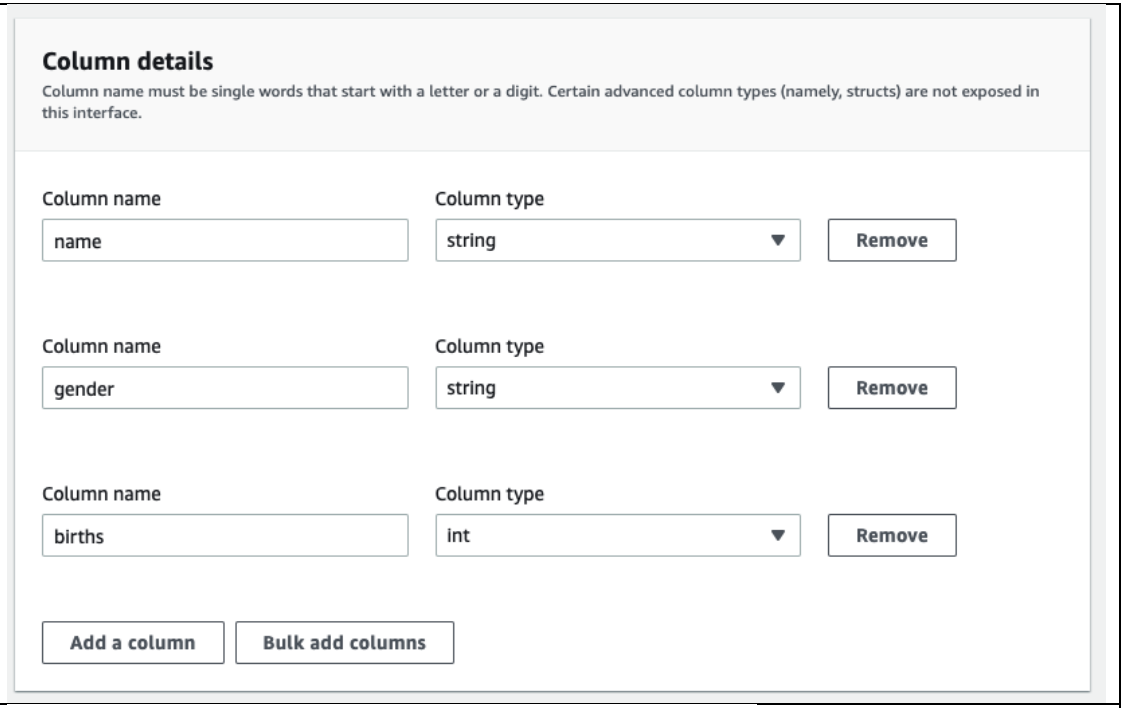
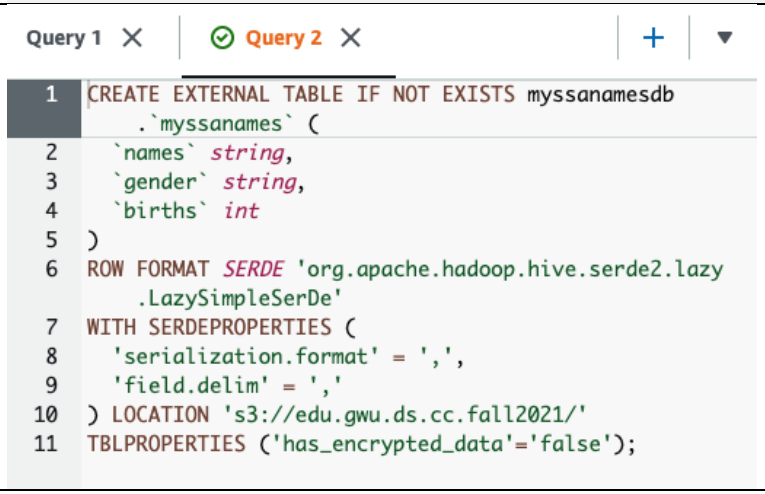
Select the option From S3 bucket data	

Select the option “Create a new database” and given a name, e.g., myssanamesdb Provide a name for the Athena table that will contain your data, e.g., myssanames Provide the name of the bucket you created previously, e.g, s3://edu.gwu.ds.cs.fall2010/ REMEMBER this is my bucket name! You must have created a bucket with your G# DO NOT FORGET the last forward slash (/) at the end of the name of the bucket Click the “next” button	
Select the CSV option	

<table>
<tr>
<td>Provide the column names and types, e.g.:
Name type String
Gender type String
Births type Int

Then type next

Do not worry about partitions

Click “Create Table”</td>
<td> 
</td>
</tr>
<tr>
<td>Athena will create the table and will show the DLL used to create it</td>
<td> 
</td>
</tr>
</table>

Create a new query that will show the total number of both female and male births

SELECT gender, SUM(births) as total
FROM myssanames
GROUP BY gender;

Execute the query and check the results