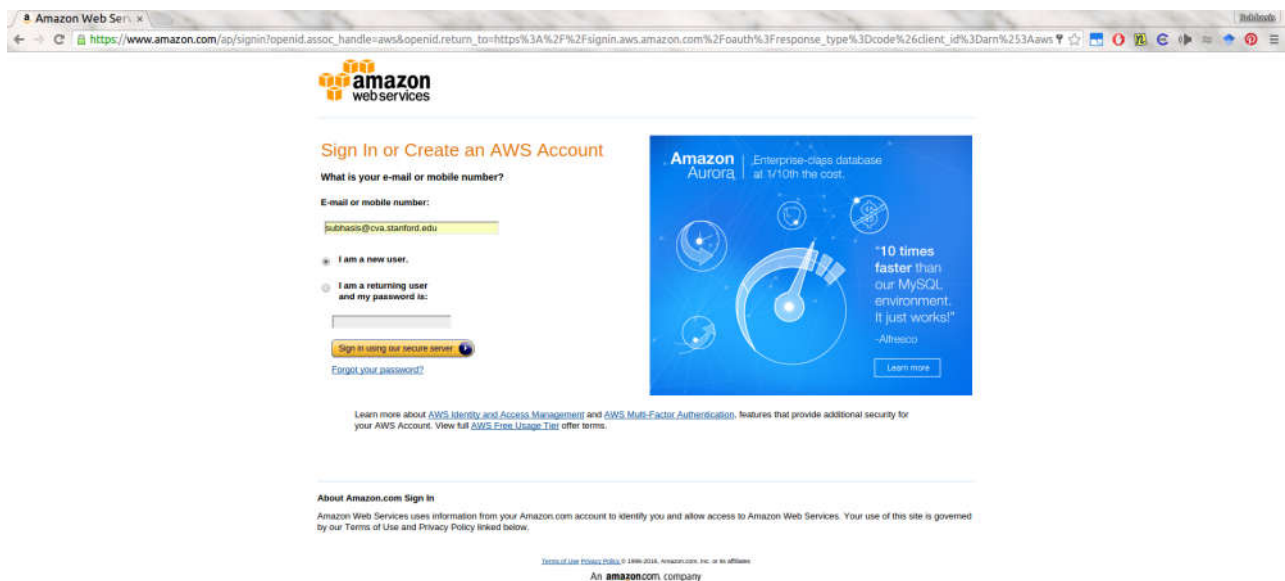


AWS Tutorial

For GPU instances, we also have an Amazon Machine Image (AMI) that you can use to launch GPU instances on Amazon EC2. This tutorial goes through how to set up your own EC2 instance with the provided AMI. **We do not currently distribute AWS credits to CS231N students but you are welcome to use this snapshot on your own budget.**

TL;DR for the AWS-savvy: Our image is `cs231n_caffe_torch7_keras_lasagne_v2`, AMI ID: `ami-125b2c72` in the us-west-1 region. Use a `g2.2xlarge` instance. Caffe, Torch7, Theano, Keras and Lasagne are pre-installed. Python bindings of caffe are available. It has CUDA 7.5 and CuDNN v3.

First, if you don't have an AWS account already, create one by going to the [AWS homepage](#), and clicking on the yellow "Sign In to the Console" button. It will direct you to a signup page which looks like the following.



Select the "I am a new user" checkbox, click the "Sign in using our secure server" button, and follow the subsequent pages to provide the required details. They will ask for a credit card information, and also a phone verification, so have your phone and credit card ready.

Once you have signed up, go back to the [AWS homepage](#), click on “Sign In to the Console”, and this time sign in using your username and password.

The screenshot shows the AWS sign-in page. At the top, it says "Sign In or Create an AWS Account". Below this, it asks "What is your e-mail or mobile number?". There is a text input field containing "subhasis@cva.stanford.edu". Below the input field, there are two radio buttons: "I am a new user." and "I am a returning user and my password is:". The "I am a returning user" option is selected. Below the radio buttons, there is a password input field. To the right of the input fields, there is a blue box with the "Amazon Aurora" logo and text: "Enterprise-class database at 1/10th the cost." and "10 times faster than our MySQL environment. It just works!" -Alfred. Below the input fields, there is a "Sign in using our secure server" button and a "Forgot your password?" link. At the bottom, there is a section titled "About Amazon.com Sign In" with text: "Amazon Web Services uses information from your Amazon.com account to identify you and allow access to Amazon Web Services. Your use of this site is governed by our Terms of Use and Privacy Policy linked below." and a link to "Terms of Use Privacy Policy".

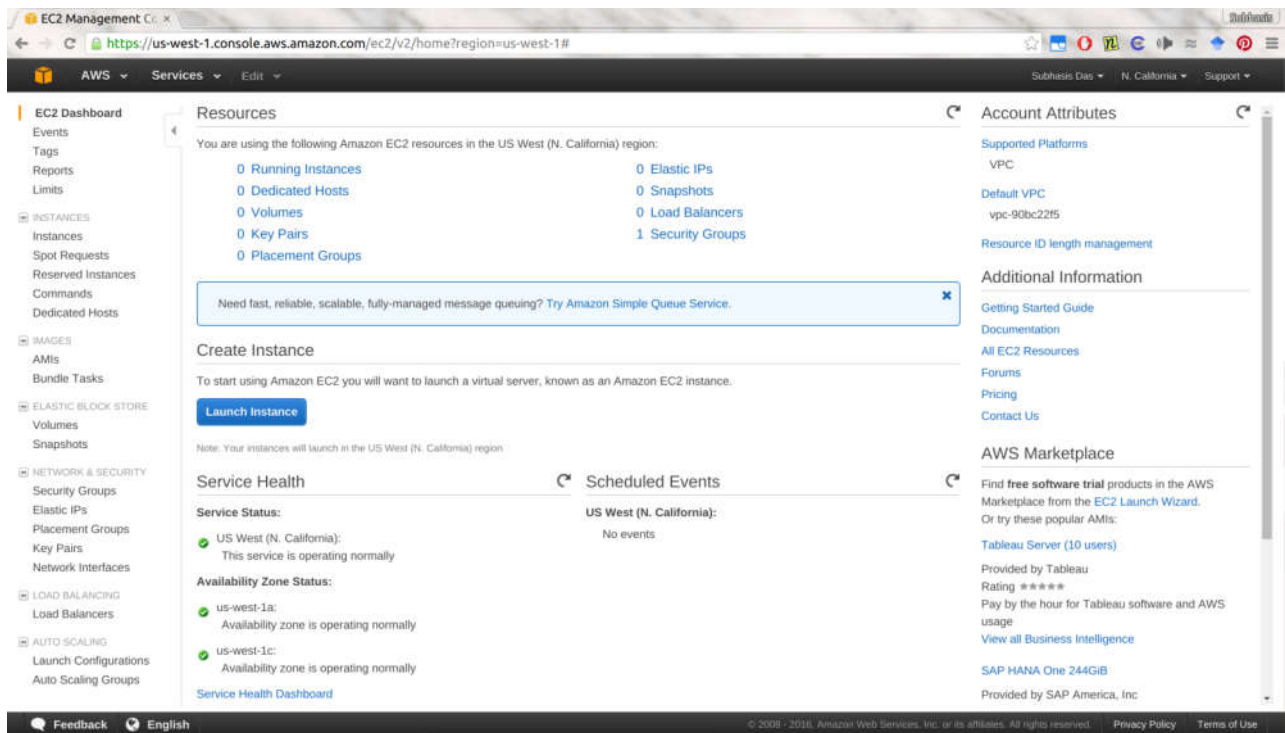
Once you have signed in, you will be greeted by a page like this:

The screenshot shows the AWS Management Console home page. The top navigation bar includes the "AWS" logo, "Services", "Edit", and a dropdown menu for "Subhasis Das" with "N. California" and "Support" options. The main content area is titled "Amazon Web Services" and displays a grid of service categories: Compute (EC2, EC2 Container Service, Elastic Beanstalk, Lambda), Storage & Content Delivery (S3, CloudFront, Elastic File System, Glacier, Import/Export Snowball, Storage Gateway), Database (RDS, DynamoDB, ElastiCache, Redshift, DMS), Networking (VPC), Developer Tools (CodeCommit, CodeDeploy, CodePipeline, CloudWatch, CloudFormation, CloudTrail, Config, OpsWorks, Service Catalog, Trusted Advisor), Security & Identity (Identity & Access Management, Directory Service, Inspector, WAF, Certificate Manager), Analytics (EMR), Internet of Things (AWS IoT), Mobile Services (Mobile Hub, Cognito, Device Farm, Mobile Analytics, SNS), Application Services (API Gateway, AppStream, CloudSearch, Elastic Transcoder, SES, SQS, SWF), and Enterprise Applications (WorkSpaces, WorkDocs, WorkMail). On the right side, there is a "Resource Groups" section with a "Create a Group" button and a "Tag Editor" button. Below this, there is an "Additional Resources" section with links to "Getting Started", "AWS Console Mobile App", "AWS Marketplace", "AWS re-Invent Announcements", and "Service Health". The "Service Health" section shows a green checkmark and the text "AWS Storage Gateway (Tokyo)" with an update timestamp of "Jan 27 2016 07:39:00 GMT-0800".

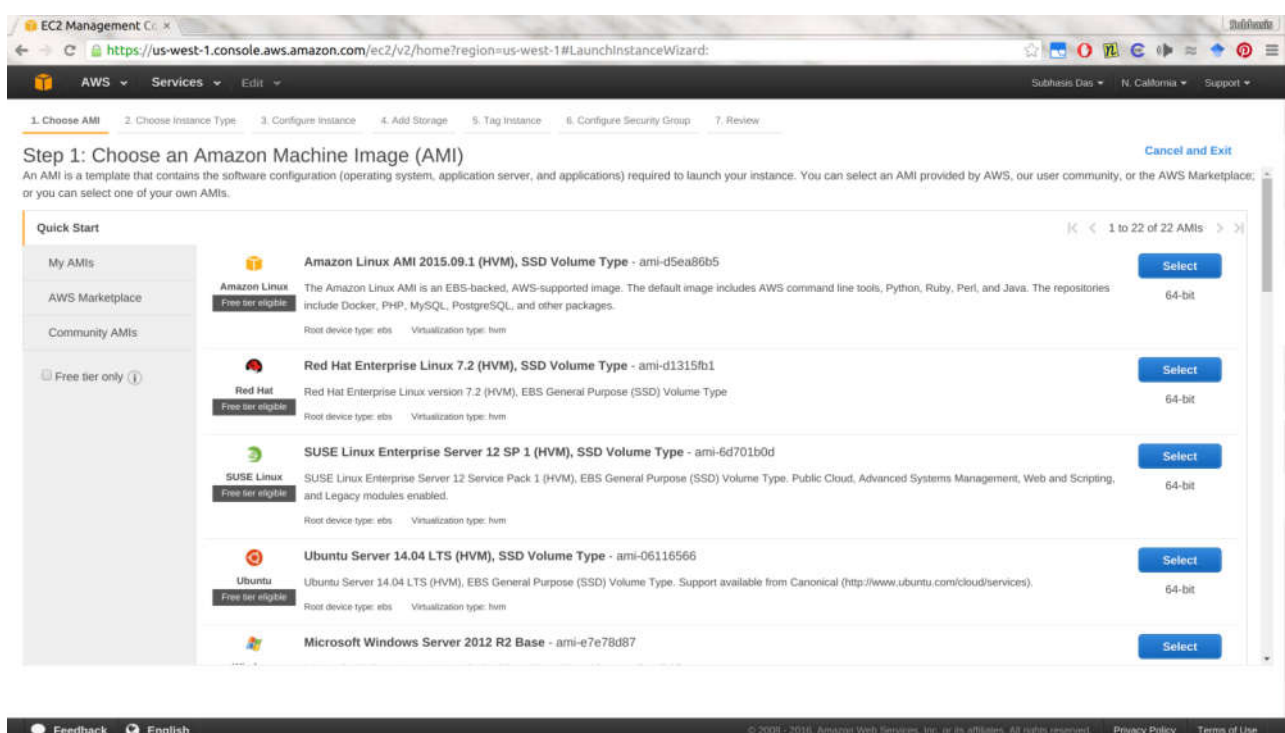
Make sure that the region information on the top right is set to N. California. If it is not, change it to N. California by selecting from the dropdown menu there.

(Note that the subsequent steps requires your account to be “Verified” by Amazon. This may take up to 2 hrs, and you may not be able to launch instances until your account verification is complete.)

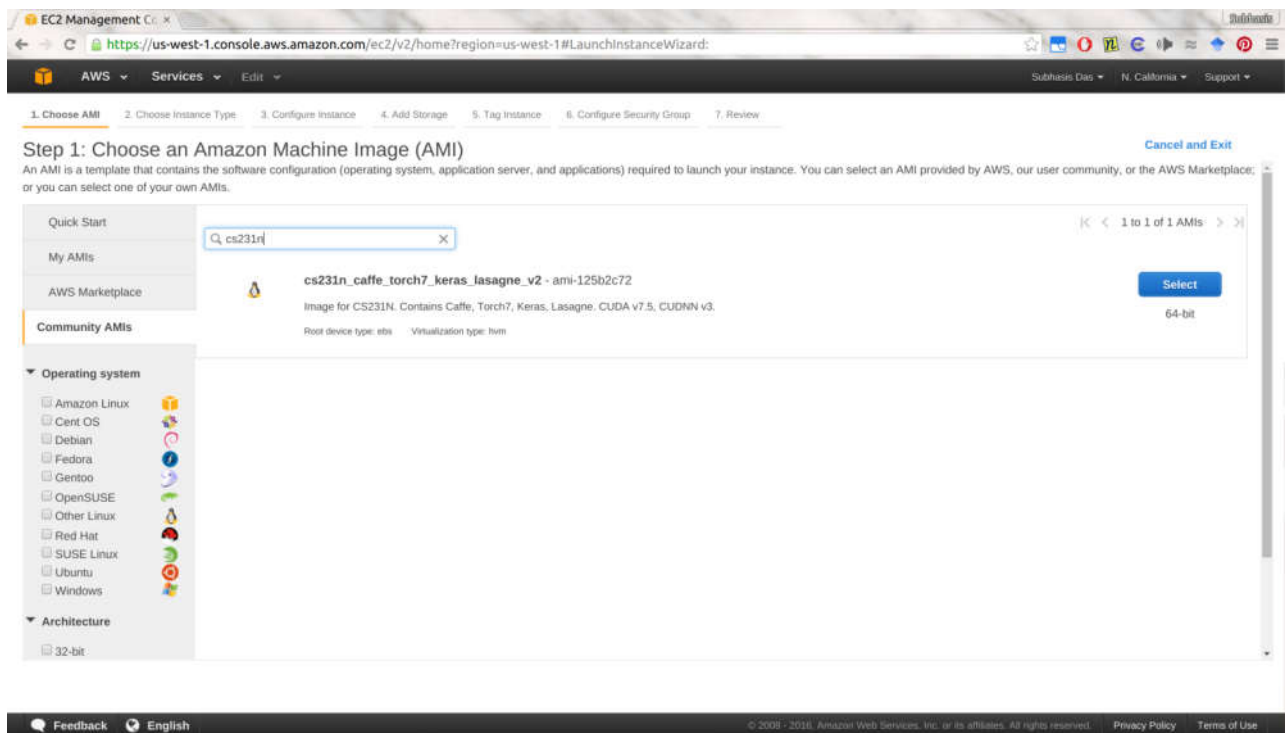
Next, click on the EC2 link (first link under the Compute category). You will go to a dashboard page like this:



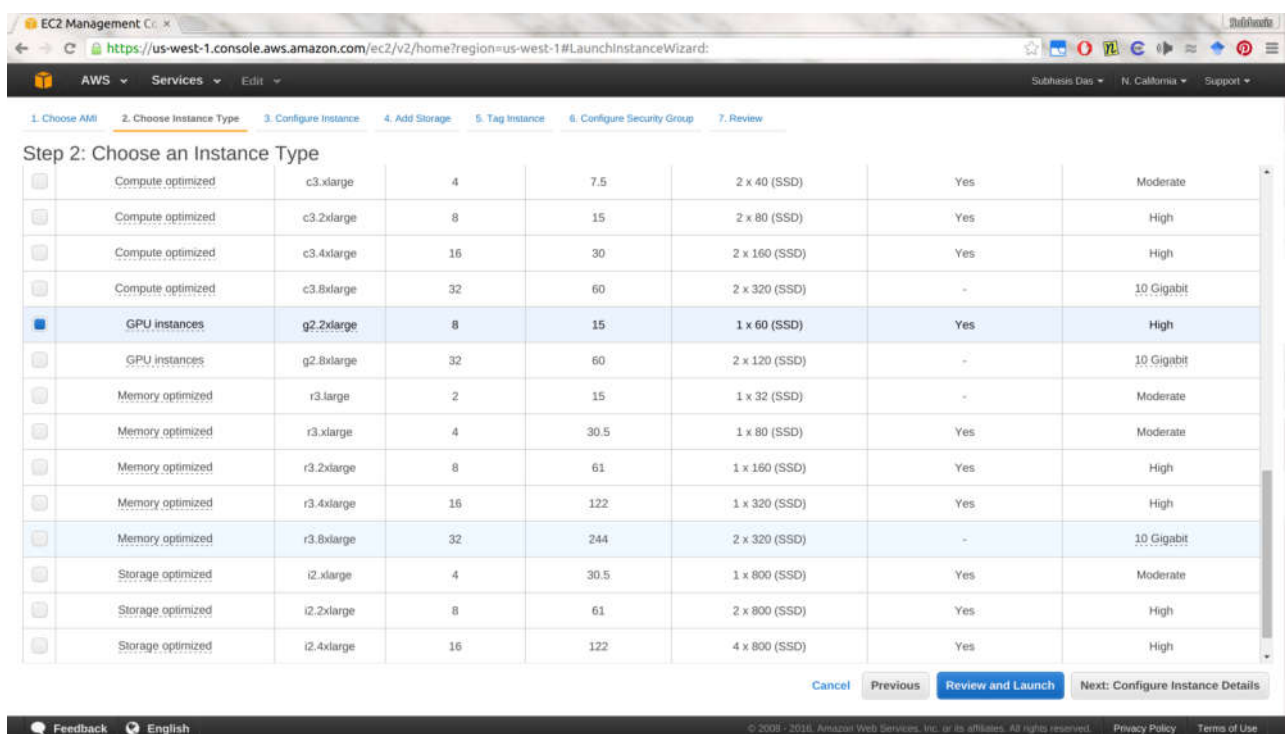
Click the blue “Launch Instance” button, and you will be redirected to a page like the following:



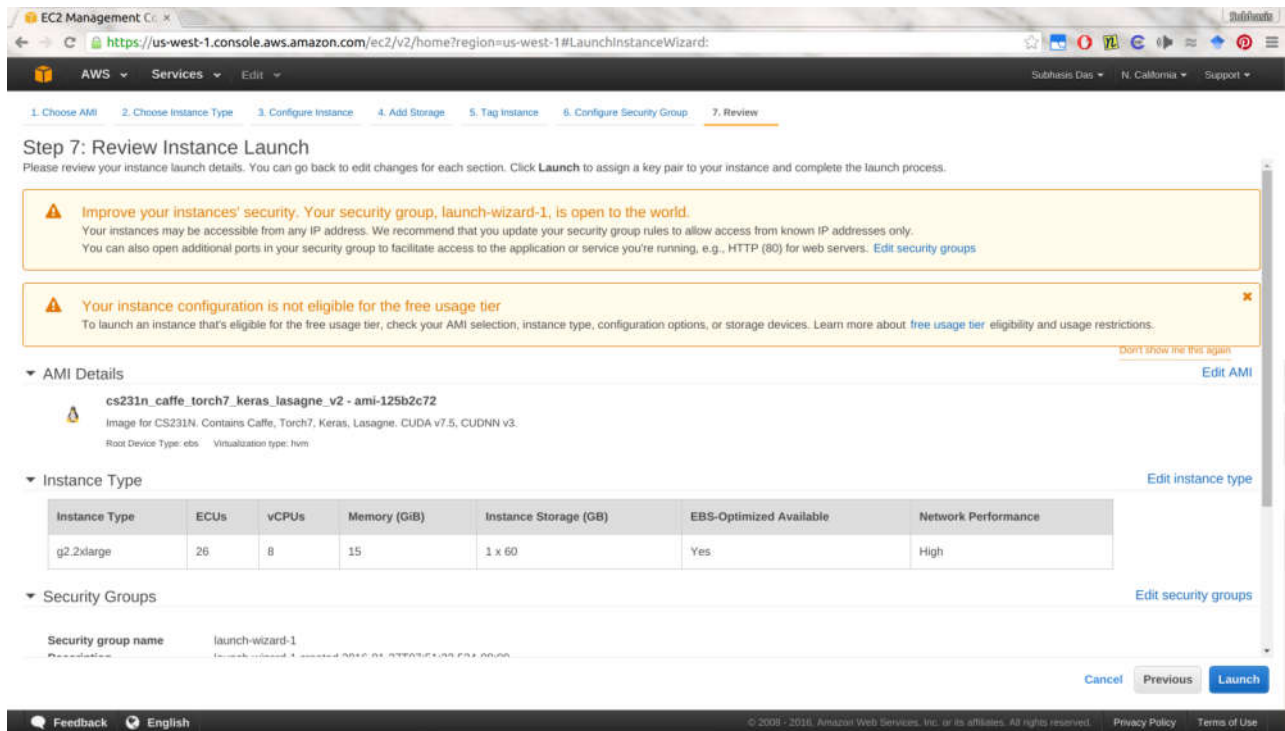
Click on the “Community AMIs” link on the left sidebar, and search for “cs231n” in the search box. You should be able to see the AMI `cs231n_caffe_torch7_keras_lasagne_v2` (AMI ID: `ami-125b2c72`). Select that AMI, and continue to the next step to choose your instance type.



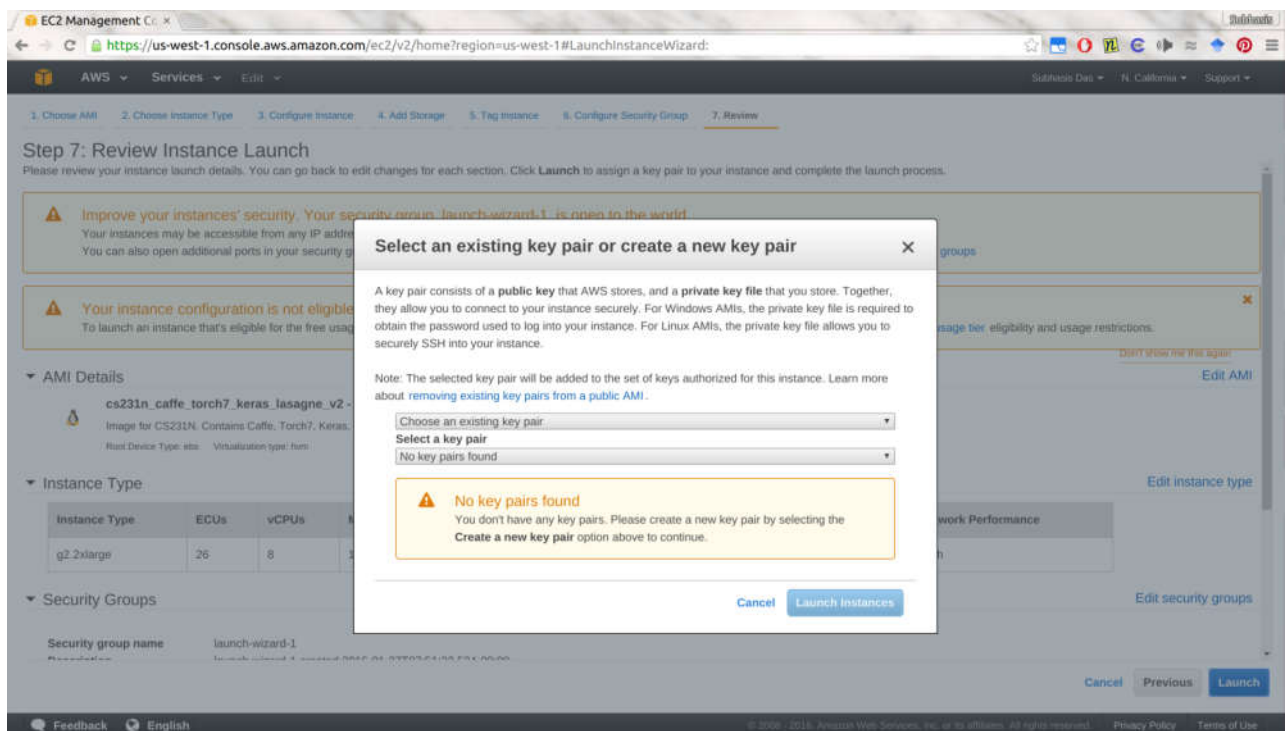
Choose the instance type `g2.2xlarge` , and click on “Review and Launch”.

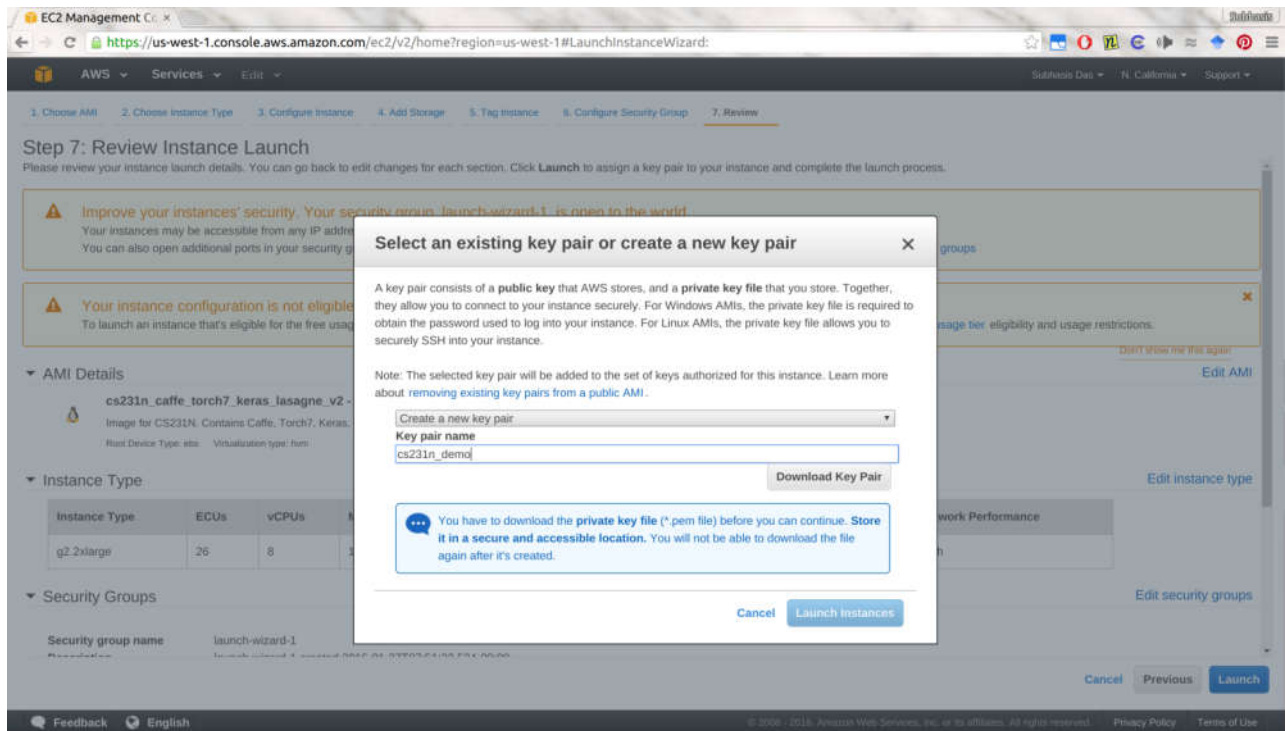


In the next page, click on Launch.



You will be then prompted to create or use an existing key-pair. If you already use AWS and have a key-pair, you can use that, or alternately you can create a new one by choosing “Create a new key pair” from the drop-down menu and giving it some name of your choice. You should then download the key pair, and keep it somewhere that you won’t accidentally delete. Remember that there is **NO WAY** to get to your instance if you lose your key.





Once you download your key, you should change the permissions of the key to user-only RW, In Linux/OSX you can do it by:

```
$ chmod 600 PEM_FILENAME
```

Here `PEM_FILENAME` is the full file name of the .pem file you just downloaded.

After this is done, click on “Launch Instances”, and you should see a screen showing that your instances are launching:

EC2 Management

https://us-west-1.console.aws.amazon.com/ec2/v2/home?region=us-west-1#LaunchInstanceWizard

AWS Services Edit

Subhash Das N. California Support

Launch Status

Your instances are now launching
The following instance launches have been initiated: i-a41511d [View launch log](#)

Get notified of estimated charges
Create billing alerts to get an email notification when estimated charges on your AWS bill exceed an amount you define (for example, if you exceed the free usage tier).

How to connect to your instances

Your instances are launching, and it may take a few minutes until they are in the **running** state, when they will be ready for you to use. Usage hours on your new instances will start immediately and continue to accrue until you stop or terminate your instances. Click **View instances** to monitor your instances' status. Once your instances are in the **running** state, you can **connect** to them from the Instances screen. [Find out](#) how to connect to your instances.

Here are some helpful resources to get you started

- How to connect to your Linux instance
- Amazon EC2: User Guide
- Learn about AWS Free Usage Tier
- Amazon EC2: Discussion Forum

While your instances are launching you can also

- [Create status check alarms](#) to be notified when these instances fail status checks. (Additional charges may apply)
- [Create and attach additional EBS volumes](#) (Additional charges may apply)
- [Manage security groups](#)

[View Instances](#)

Feedback English

© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Click on “View Instances” to see your instance state. It should change to “Running” and “2/2 status checks passed” as shown below within some time. You are now ready to ssh into the instance.

EC2 Management

https://us-west-1.console.aws.amazon.com/ec2/v2/home?region=us-west-1#Instances

AWS Services Edit

Subhash Das N. California Support

EC2 Dashboard

Launch Instance Connect Actions

Filter by tags and attributes or search by keyword

Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS	Public IP
	i-a41511d	g2.xlarge	us-west-1c	running	2/2 checks ...	None	ec2-54-183-60-80.us-w...	54.183.60.80

Instance: i-a41511d Public DNS: ec2-54-183-60-80.us-west-1.compute.amazonaws.com

Description	Status Checks	Monitoring	Tags
Instance ID	i-a41511d	Public DNS	ec2-54-183-60-80.us-west-1.compute.amazonaws.com
Instance state	running	Public IP	54.183.60.80
Instance type	g2.xlarge	Elastic IP	-
Private DNS	ip-172-31-0-241.us-west-1.compute.internal	Availability zone	us-west-1c
Private IPs	172.31.0.241	Security groups	launch-wizard-3, view rules
Secondary private IPs		Scheduled events	No scheduled events
VPC ID	vpc-90bc22f5	AMI ID	cs231n_caffe_torch7_keras_lasagne_v2 (ami-125b2c72)
Subnet ID	subnet-7f199526	Platform	-

Feedback English

© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

First, note down the Public IP of the instance from the instance listing. Then, do:

```
ssh -i PEM_FILENAME ubuntu@PUBLIC_IP
```

Now you should be logged in to the instance. You can check that Caffe is working by doing:

```
$ cd caffe
$ ./build/tools/caffe time --gpu 0 --model examples/mnist/lenet.prototxt
```

We have Caffe, Theano, Torch7, Keras and Lasagne pre-installed. Caffe python bindings are also available by default. We have CUDA 7.5 and CuDNN v3 installed.

If you encounter any error such as

```
Check failed: error == cudaSuccess (77 vs. 0) an illegal memory access was encountered
```

you might want to terminate your instance and start over again. I have observed this rarely, and I am not sure what causes this.

About how to use these instances:

- The root directory is only 12GB, and only ~ 3GB of that is free.
- There should be a 60GB `/mnt` directory that you can use to put your data, model checkpoints, models etc.
- Remember that the `/mnt` directory won't be persistent across reboots/terminations.
- Stop your instances when are done for the day to avoid incurring charges. GPU instances are costly. Use your funds wisely. Terminate them when you are sure you are done with your instance (disk storage also costs something, and can be significant if you have a large disk footprint).
- Look into creating custom alarms to automatically stop your instances when they are not doing anything.
- If you need access to a large dataset and don't want to download it every time you spin up an instance, the best way to go would be to create an AMI for that and attach that AMI to your machine when configuring your instance (before launching but after you have selected the AMI).

