Azure Guide for CS224n

This guide will help you setup and use Azure Virtual Machines for your final project. Before we start, it cannot be stressed enough: **do not leave your machine running when you are not using it**. The expected time to complete the setup guide is **15 min** to **1 hour**, depending on which configuration you opt to take.

Contents

Your Azure subscription for this class

Microsoft has generously agreed to sponsor CS224n, and has provided us with Azure credit to distribute to CS224n students. We expect that there will be enough credit for teams to run as many experiments as they need for their projects. **However, it's very important for students to manage their credit carefully, so that they can get the most out of it (see next section).**

You need Azure credits for assignment 4, assignment 5, and final project. When assignment 4 is released, you will receive an email containing an invitation to claim your Azure credits.

For the specifics of how much credits you will be provided for each of the assignments, refer to the Azure posts we will make on Ed.

For the final project, a credit of $150 will be assigned per team (according to the teams you gave us in your project proposal), with the same amount allocated regardless of team size. The $150 corresponds to about **150 hours**, or slightly over **6 days** on a NC6 machine.

The $150 is an initial allocation. If you use it up running *genuine* experiments, that's **perfectly OK and completely expected** – we expect that most teams will need more credit, and we have plenty more to give you. However, **please don't use up your credit by leaving your machine running when you're not using it**! Nor should you use up many hours of credit using your VM to write your code (see next section).

When you run out of credit (or before you run out), you can ask us for more on Ed using the "azure" tag.

Best practices for managing your Azure credit

Azure virtual machines are charged at a flat rate, for each minute that they are turned on. This is irrespective of:

- whether you are ssh'd to the machine at that time
- whether you are running any processes on the machine at that time
- the computational intensity of the the processes you're running
- whether you're using GPUs

Therefore, the most important thing you need to do to, to manage your Azure credit, is to **carefully turn your VM on and off just when you need it**. If you are using a NC6 VM, it is charged at $0.9/hour while it is turned on.

We advise you to **develop your code on your local machine** (for example your laptop with the CPU version of PyTorch installed) for debugging (i.e., work on your new code until you are able to complete several training iterations without errors), then run your code on your Azure VM when it's time to train on a GPU.

Note: we have provided you with a Practical Tips for Final Projects document which gives tips on how to sync your code between your laptop and your VM, how to use tmux to manage your sessions in your VM, and how to monitor your memory/CPU/GPU usage.

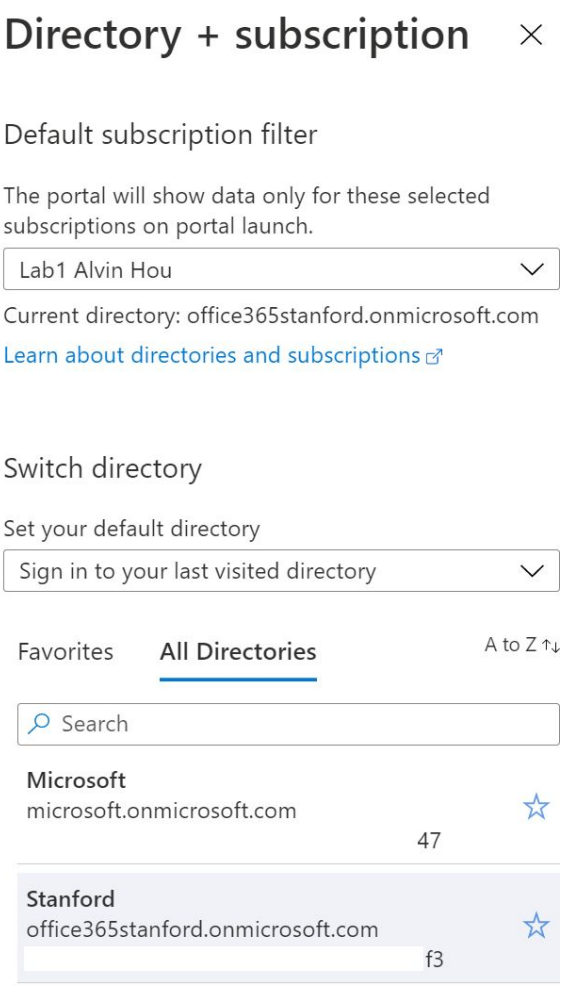
Azure also has an auto-shutdown feature that allows you to specify a time when you want your VM to turn off - this allows you to turn off the machine at a time when you are unable to do it manually. For example, if you start an experiment at 9 p.m., and you want to stop it after 5 hours, you can set auto-shutdown to turn your VM off at 2 a.m. This will prevent you spending credit that you would have otherwise spent until you woke up many hours later to turn off the VM.

See FAQs of this document to learn how to check your balance.

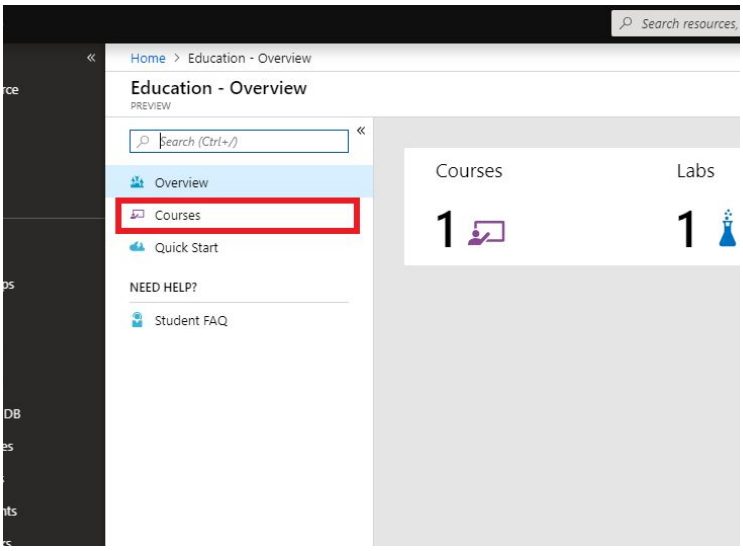
Configuring your Azure VM

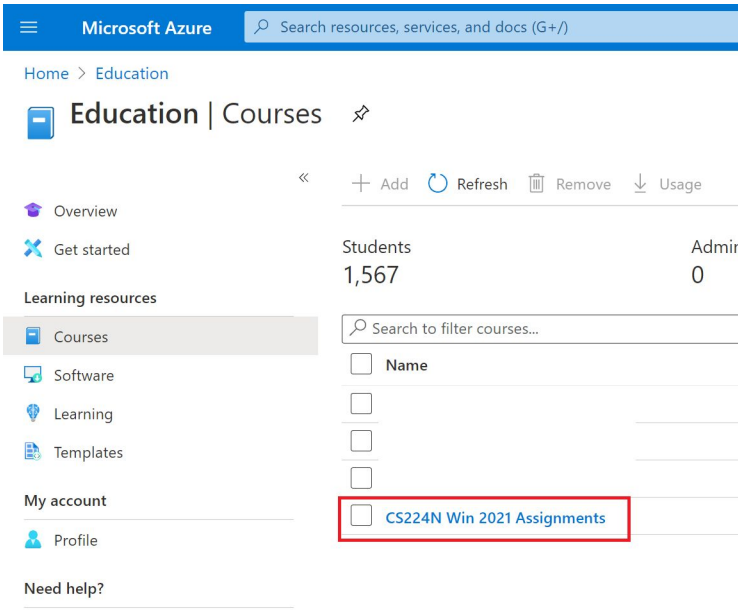
Creating an Azure account (5 min)

Login to your account at portal.azure.com using your stanford.edu email address and make sure your **Active Directory** (shown under your email address in the top right corner) is **Leland Stanford Junior University.** If you have multiple subscriptions (e.g. you're sharing the same email account for CS 224N with another course using Azure like CS 234 or CS 273B), click on the **Account Menu** in the top-right corner, select **Switch directory**, and choose **`Stanford - office365stnaford.onmicrosoft.com`.**



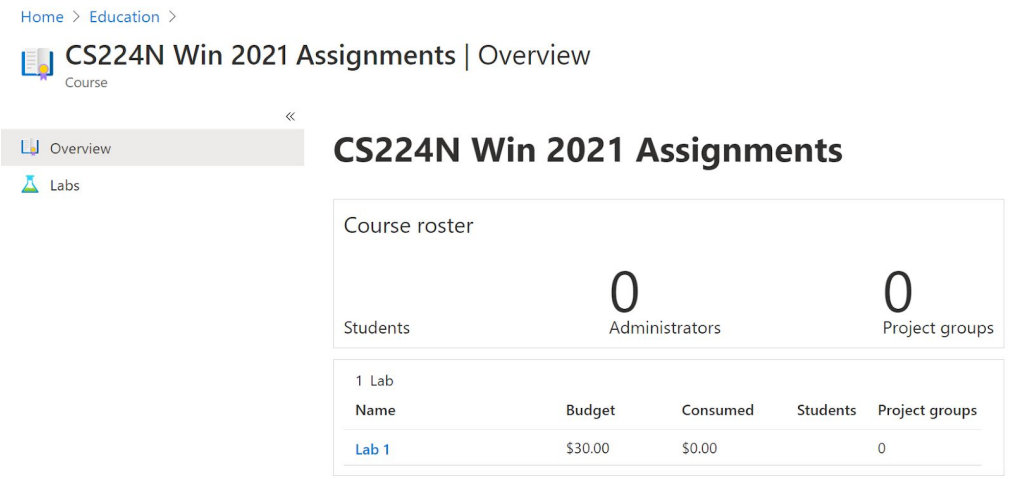
Go to portal.azure.com/#blade/Microsoft_Azure_Education/EducationMenuBlade/overview. Click on **Courses**. You should see **CS224N Win 2021 Assignments**(if you are working on assignments) or **CS224N Win 2021 Project** (if you are working on the project) in your list of courses. If you don't see the course(s) for CS224N, see Ed for detailed instructions.

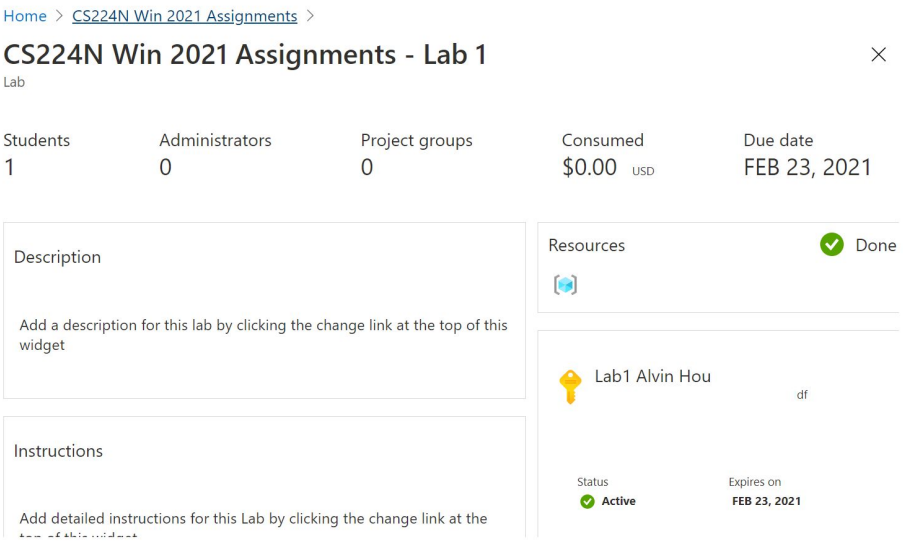




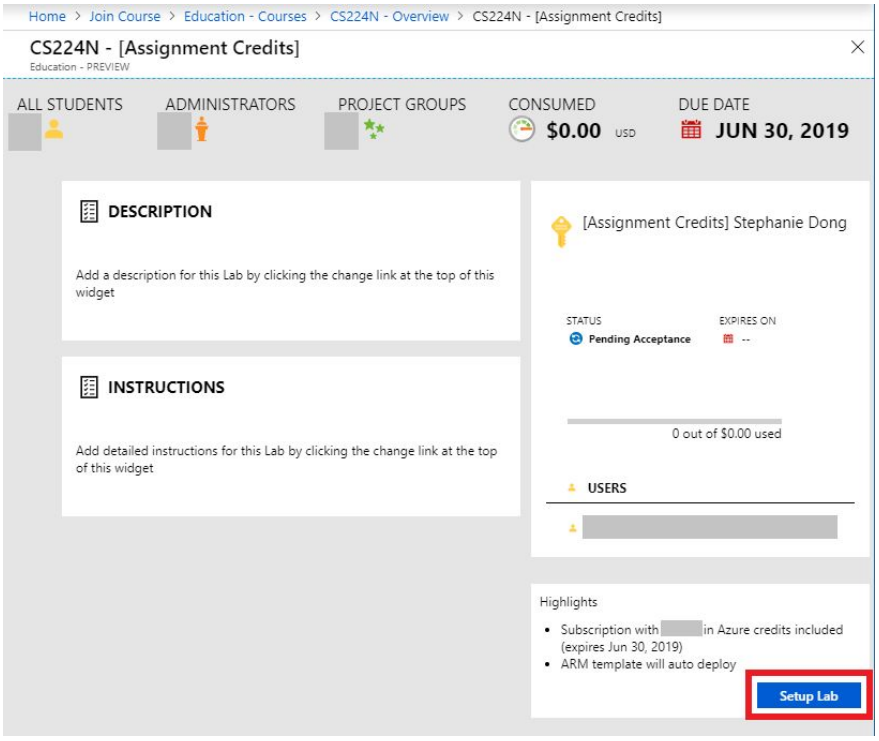
Activating your subscription (5 min)

Click on **CS224N Win 2021 Assignments** or **C224N Win 2021 Project** as applicable.
Under **Labs**, is where you will see your Azure credit subscriptions. You will be receiving credits for assignments and project separately.
If you are following this guide for assignments that require Azure, you should see **[Lab 1]**. Click on **[Lab 1]**.
If you are following this guide for the final project, you should see **[Project Credits]**. Click on **[Project Credits]**.

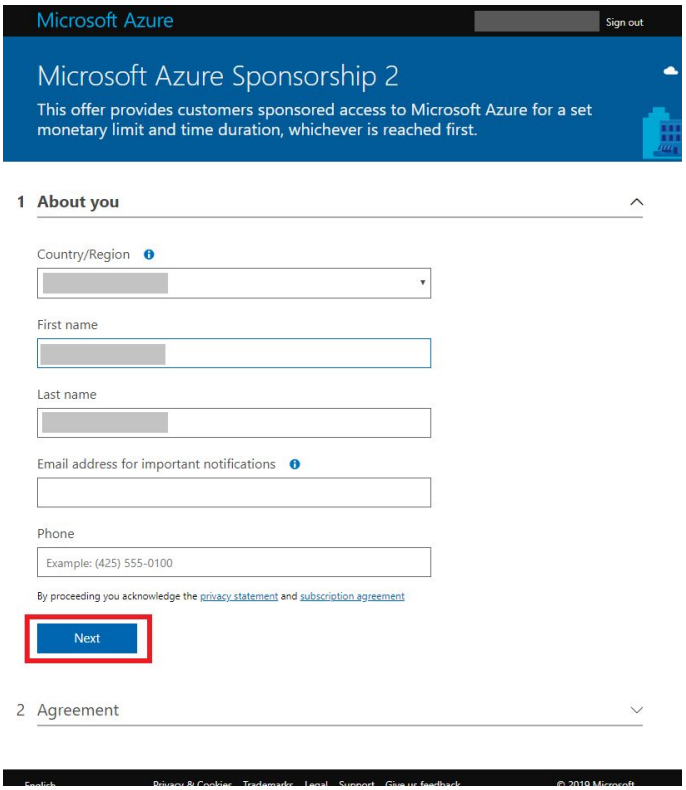


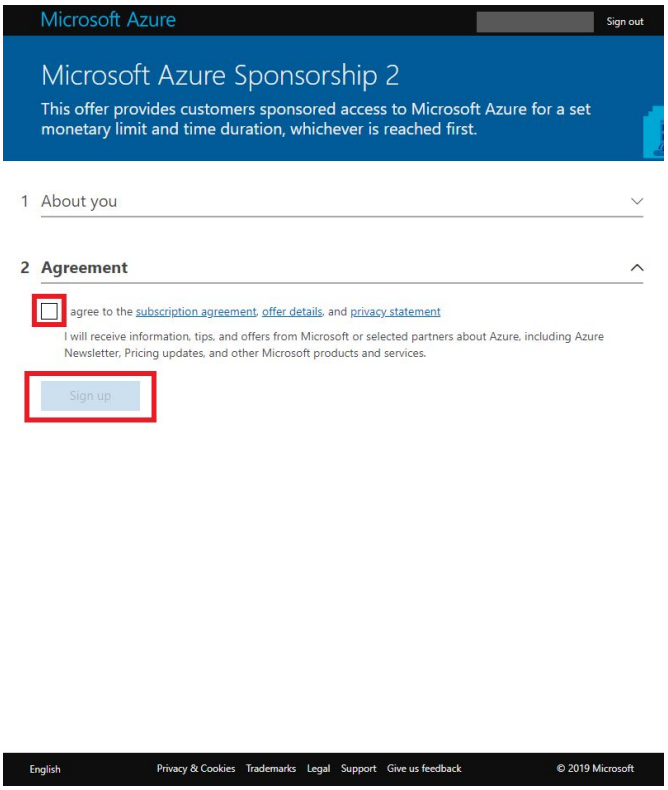


You should be brought to an overview page for your Azure subscription for either the assignments or the project. Click on **Setup Lab** to activate your subscription. If you don't see the option for **Setup Lab**, and your **STATUS** says **Accepted**, then you have already done this step.



If this is your first time activating a subscription under Azure, you may be brought to the agreement page. **Fill in your information** and click **Next** and **Sign up**. It may take a few minutes for the next page to load after you click **Sign up**.



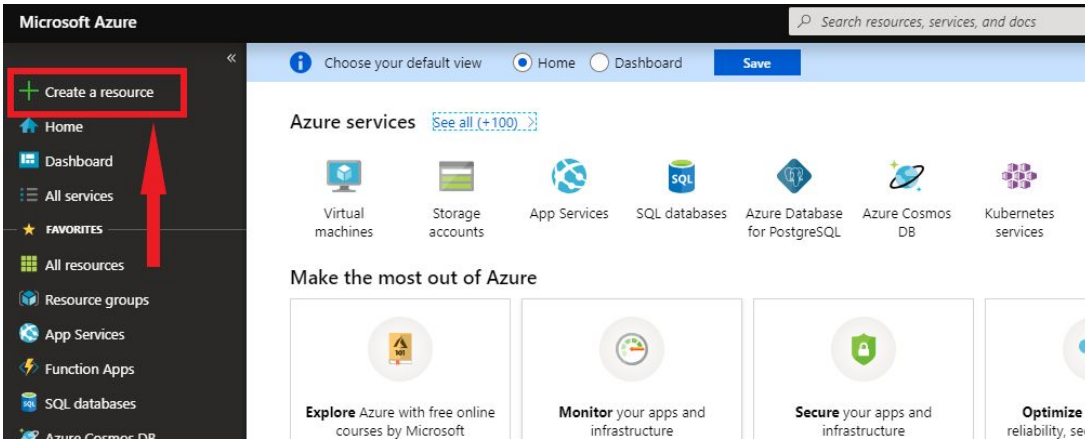


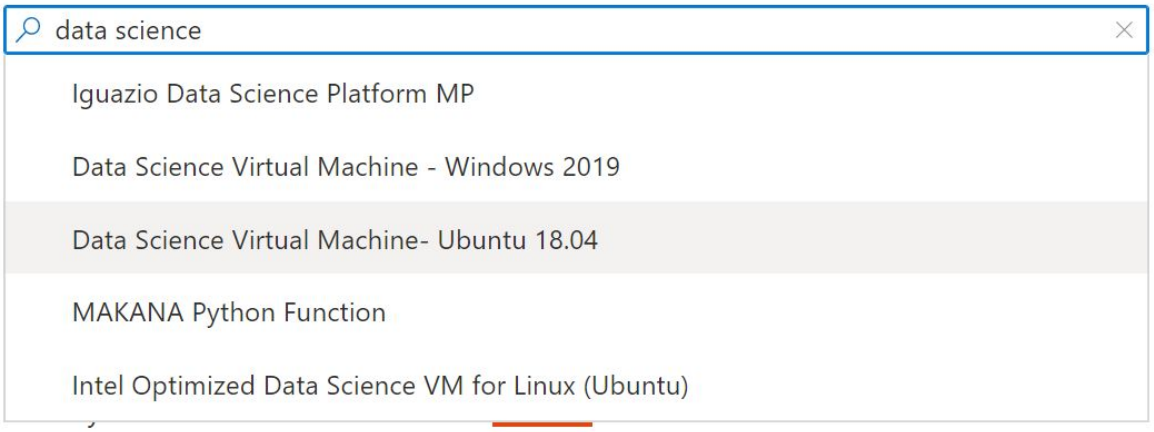
Creating a VM (15-45 min)

Using a predefined image (15 min)

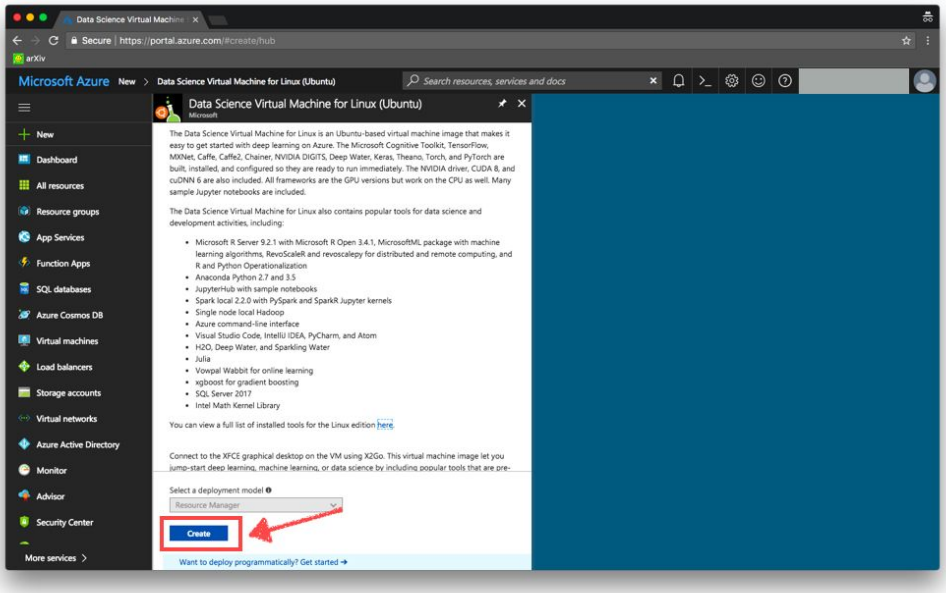
If you use a predefined image, we recommend using the `Data Science Virtual Machine- Ubuntu 18.04` image, which comes installed with Python 3.7, -gpu, tensorflow-gpu, CUDA, and cuDNN.

1. Click the `+ Create a Resource` in the left sidebar menu and type in `Data Science Virtual Machine- Ubuntu 18.04`. It's essential that you select the Ubuntu and **not** CentOS distribution.





2. Click `Create`.

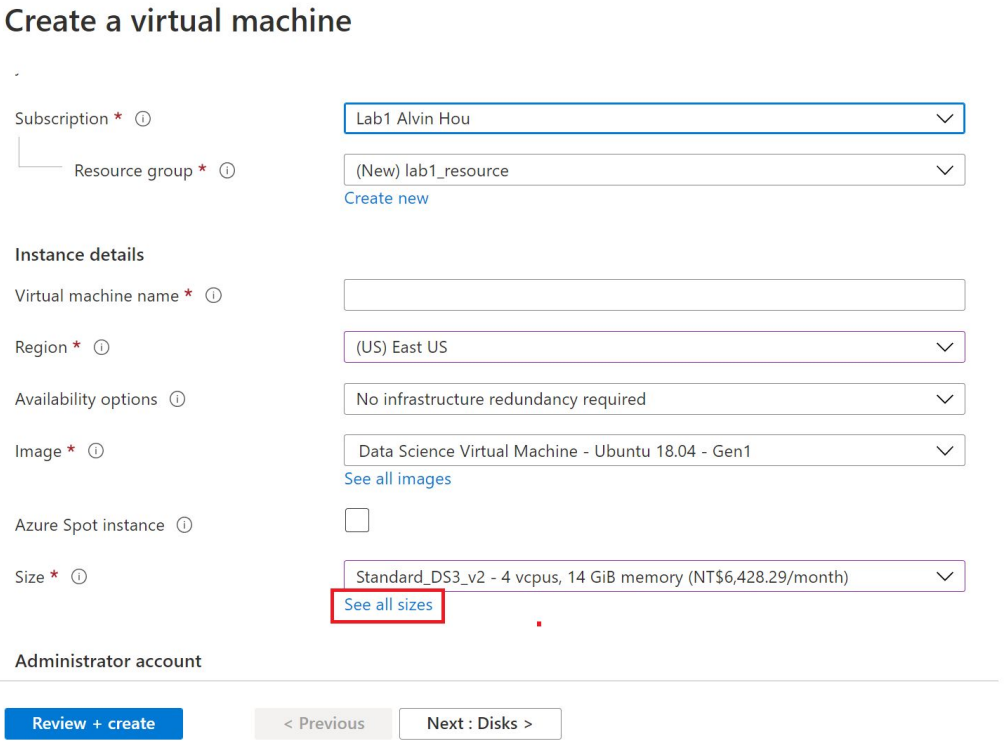


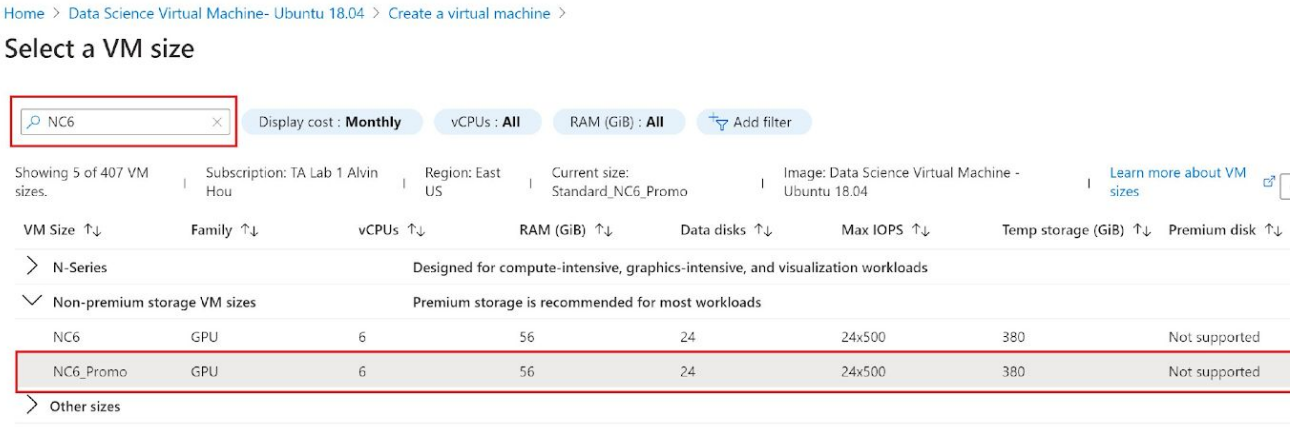
3. Fill in the following fields:

   - **Subscription.**
     i. If this is your first time using Azure or Azure for CS224N on this account, you might see `Microsoft Azure Sponsorship 2`. Choose this option
     i. If you are working on assignments and you see the subscription starting with `[Lab 1]`, choose this one.
     ii. If you are working on projects, you should see an option starting with `[Project Credits]`. Choose this one.
     iii. The VM that you create will use Azure credits from the subscription chosen, and sometimes may not be transferable to a different subscription. If you don't see the subscription that you are looking for, make sure you follow the section above on *Activating your subscription* carefully. If that still does not resolve your issue, post on Ed for assistance.

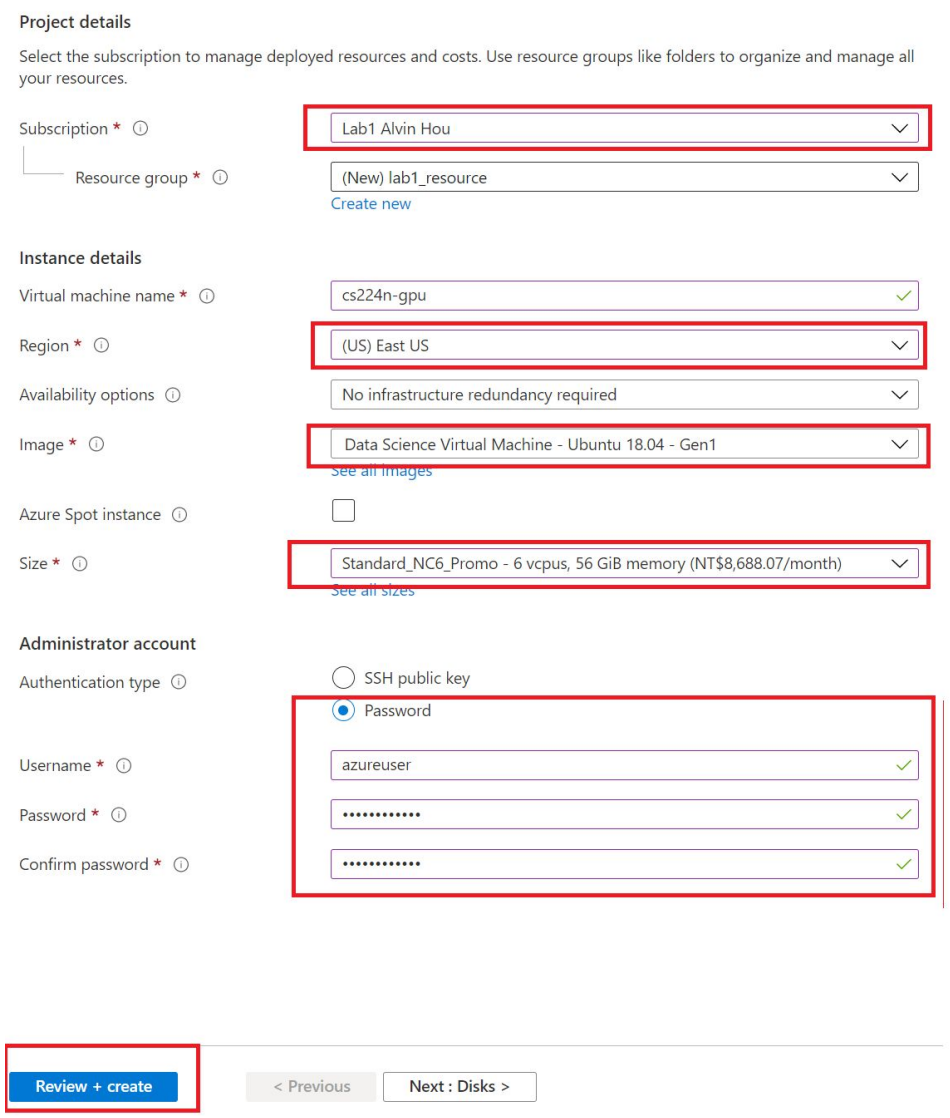
- **Resource group.** If you create multiple VMs, those within the same resource group will share resources. Unless you create multiple VMs, this configuration does not matter, so click `Create New` and type `cs224n-gpu`.
    i. **IMPORTANT**. If you are switching to a new subscription (for example from [Lab 1] to [Project Credits], you need to create a new resource group.
- **Virtual Machine Name.** This will be the name of your VM. You can name it whatever you want.
- **Region**. Choose `East US`
- **Image. IMPORTANT** Choose `Data Science Virtual Machine- Ubuntu 18.04`

- **Size**. Click on `Change size` and search for `NC6` . Select `NC6_promo` if applicable. If not, pick `NC6`. You may need to `clear all filters`. See screenshots below.
- **User name.** This will be the username used on the VM. You can name yourself whatever you want. I named myself `steph`. Since it's most convenient for all of the people in your group to share one user account, it might make more sense to use the name `group` or `team` or `<team-name>` like `purple-elephants`. (I bet your favorite language model didn't expect to see purple elephants in an Azure walkthrough…)
- **Authentication type.** If you are not familiar with SSH keys, authenticate using password; otherwise, choose whichever you prefer. I chose a secret password.

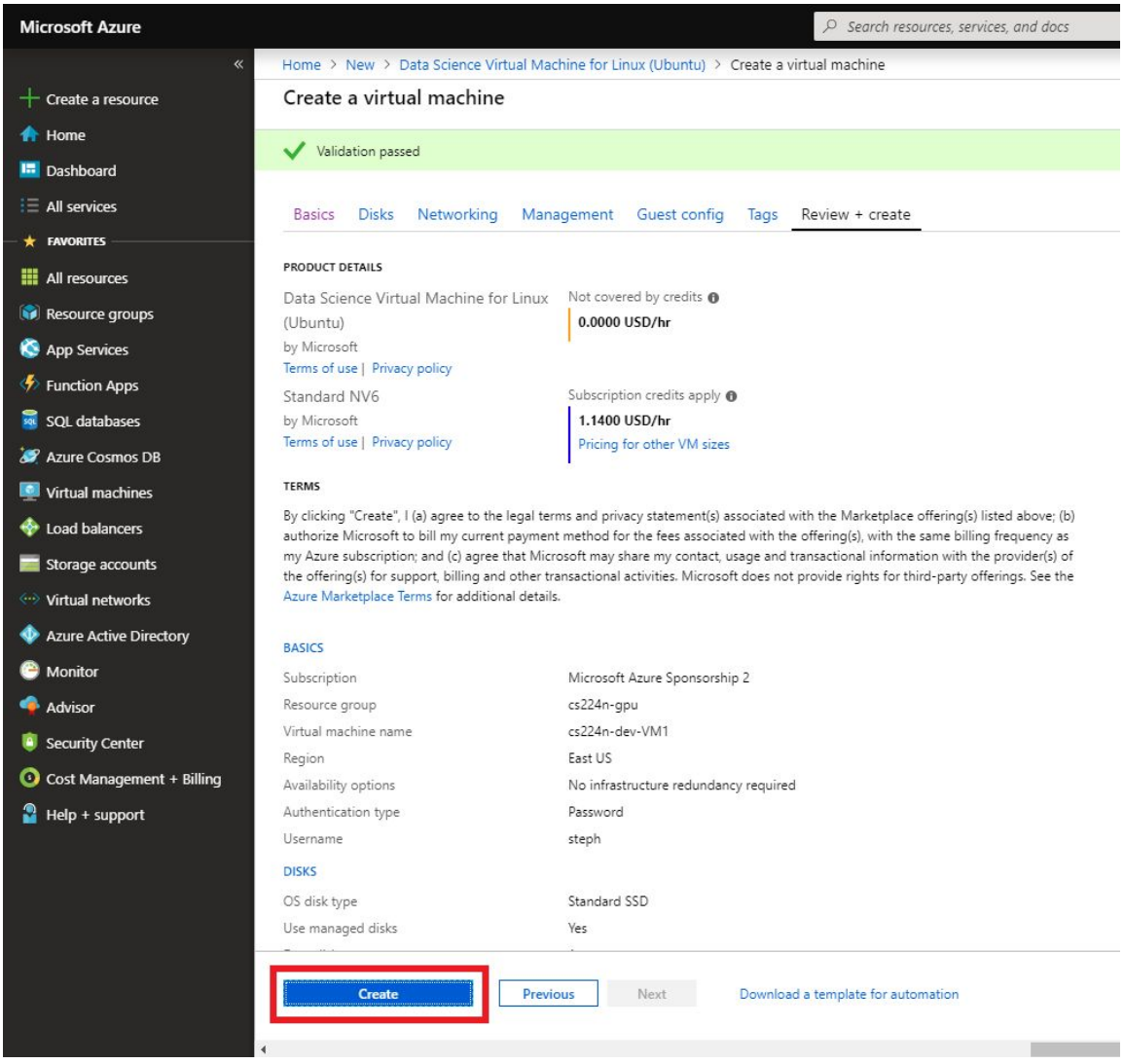




4. Double check the fields outlined in red below are filled in according to the spec above. Click Review + create.



5. Wait for the configuration to validate. Click `Create`. Sometimes, the validation errors. If you don't see `Validation passed`, click on `Basics`, confirm the fields you filled in from the previous step are still there and click Click Review + create to try again.

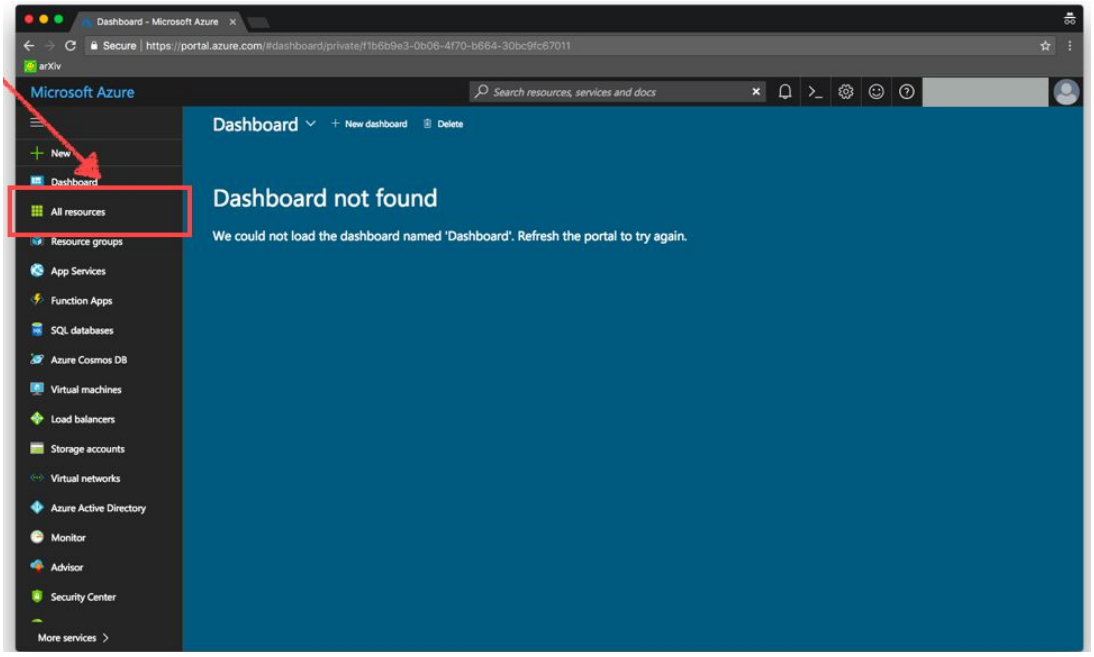


6. You've created a VM! Continue to Using Azure.
   NOTE: If you do not plan on using your VM right now, stop the instance **right now.** The VM is automatically started up when it is created. Follow the instructions below to stop your VM.

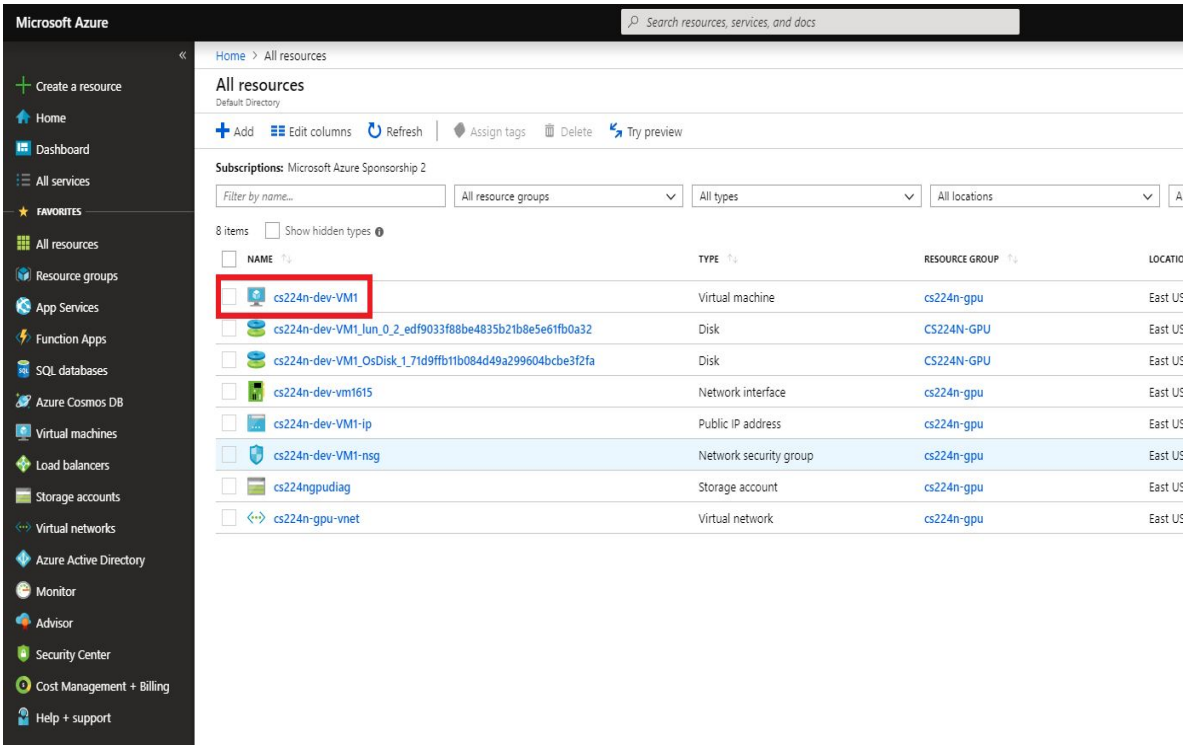
Using Azure

Managing a VM

1. Click the `All resources` in the left sidebar menu. If it is not on the left sidebar, click on All services in the sidebar, and All resources from there.

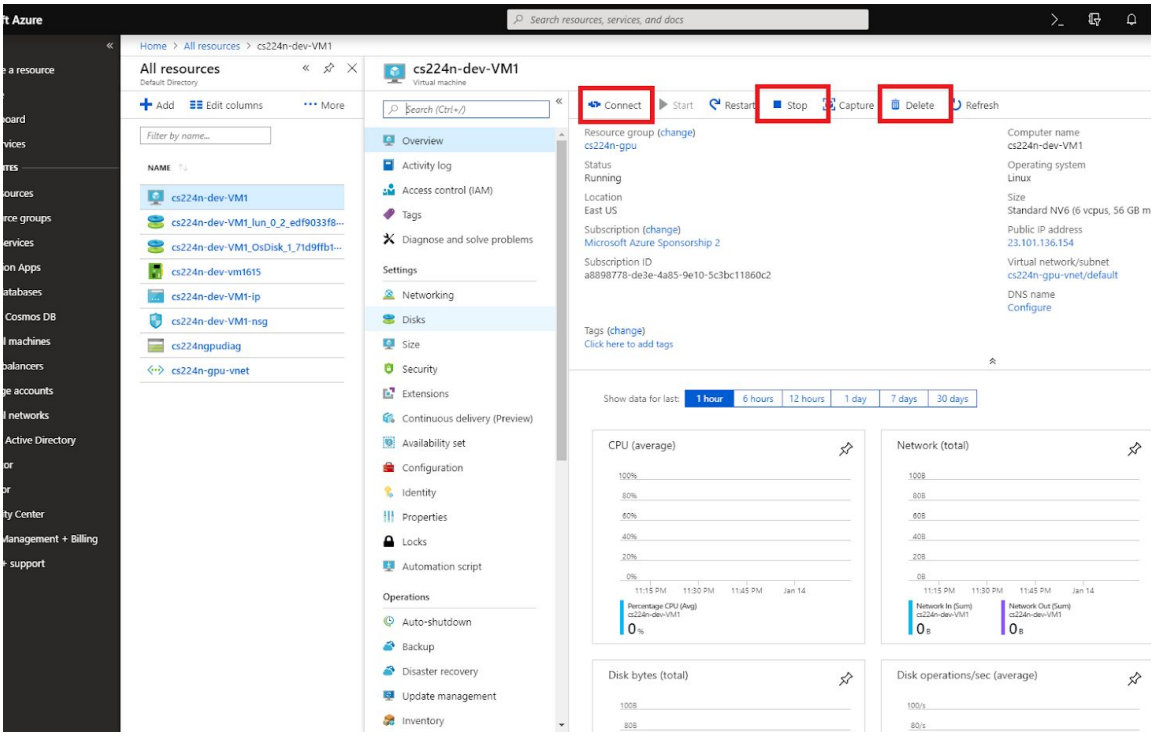


2. Click the name of your VM. You might need to **wait up to 10 minutes** after creating the VM for it to appear on this menu.



3. There are a few important options. Click `Connect` for an ssh command to connect to your instance. Click `Start`/`Stop` to start or stop the instance. If you want to delete the instance, click `Delete`.

   Note that if your instance is stopped but not deleted, it will still accrue charge for storage. (This cost is minimal). Again, ***do not leave your machine running when you are not using it****.*

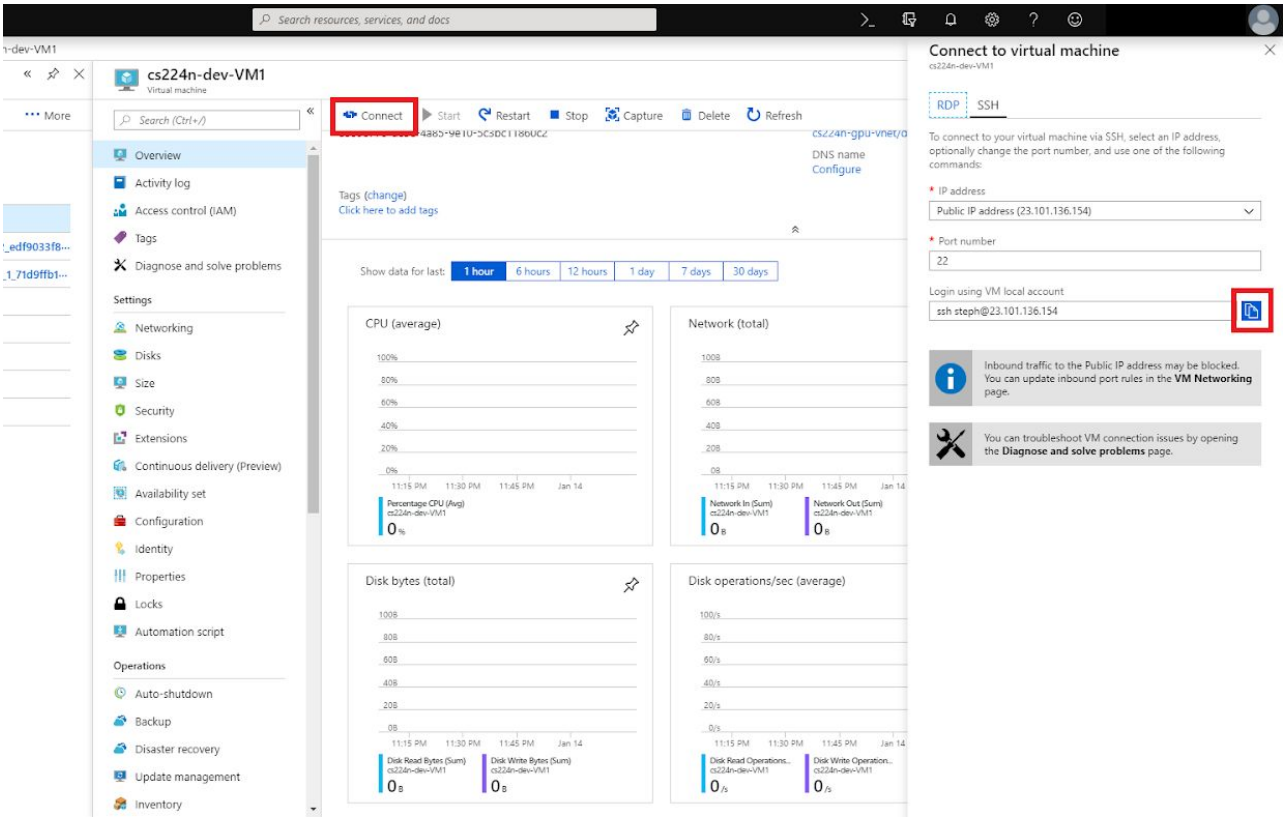


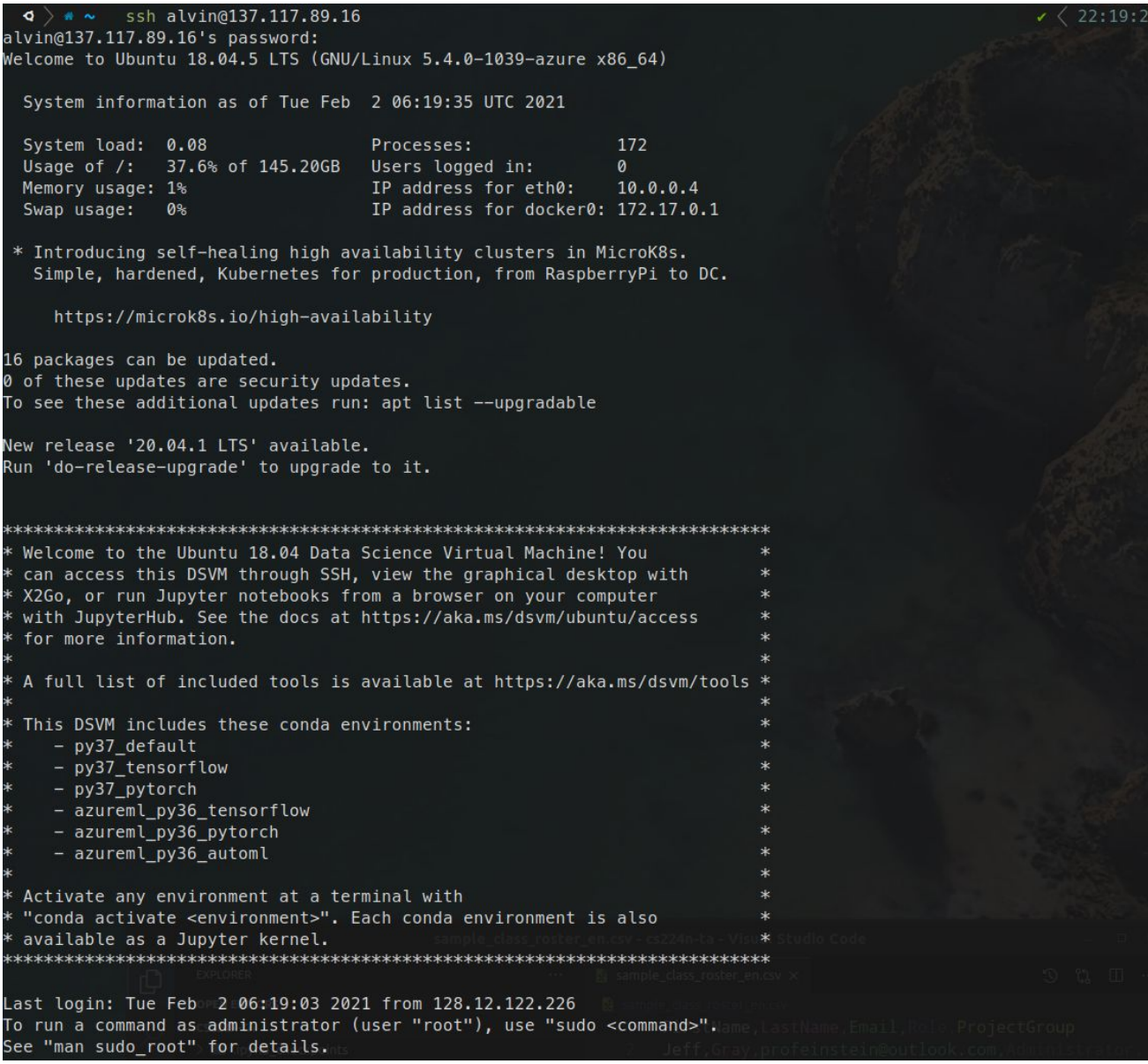
us

Connecting to a VM

Check out [CS224N: Practical Tips for Using Virtual Machines 2021] for more tips on using Azure.

1. Click `Connect` from the previous menu. In the right side panel that pops up, click on the copy icon, and paste the ssh command into your terminal.

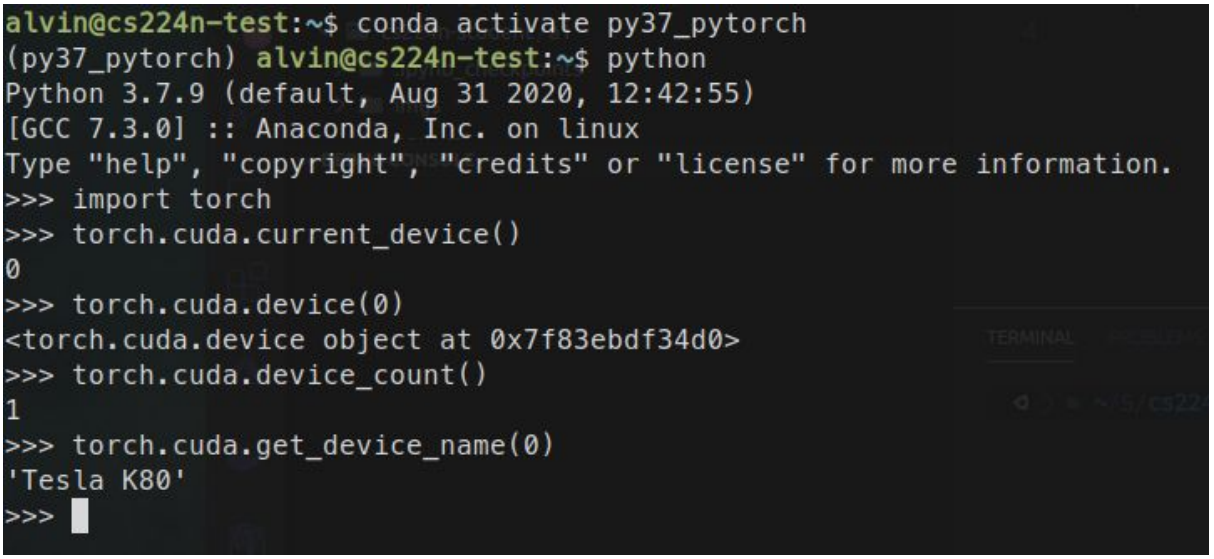




2. Check that Pytorch can access the GPUs by activating the conda environment and opening Python. See the following example

```
$ conda activate py37_pytorch
$ python
>>> import torch
>>> torch.cuda.current_device()
>>> torch.cuda.device(0)
>>> torch.cuda.device_count()
```

You should see something like this:



If you see an error message about CUDA, post to Ed for assistance.

FAQs

How do I check my remaining balance?

Go to the Labs under the CS224N Azure page from https://portal.azure.com/#blade/Microsoft_Azure_Education/EducationMenuBlade/overview
Note that Azure bills at midnight every business day, so this figure usually reflects your credit as of the last billing time. Also, note that you will only see your subscription after it is activated. Instructions for activating your subscription(s) is in section *Activate your subscription* above.

How do I share my instances with other students in my group?

For shared subscriptions only, once an instance and user account on that instance has been created using a subscription, all accounts linked to that subscription can see that instance on their dashboard and follow the directions in Using Azure to manage and connect to their VM. Only the subscription created for the final project is shared.

How do I create new user accounts?

If your group feels strongly about using separate user accounts instead of a shared one on your instance, please post privately on Ed.

What happens when I exceed my credit?

Your subscription will be disabled. Please shut down your VM(s) and follow the instructions on Ed.

Can I add a personal credit card to the account?

Sure, though we do not recommend it. If you exhaust the funds from your CS 224N subscription, your personal credit card will be charged without warning.

Can I select more powerful instances?

Though we recommend the NC6, you are free to use any of the instances. Just keep in mind that you have a budget!

Appendix

How do I create an SSH key for VM connection?

On your local machine, create SSH key pairs:

- Run `ssh-keygen -m PEM -t rsa -b 4096` (Linux / MacOS)
- Or use the PuTTYgen tool (Windows)

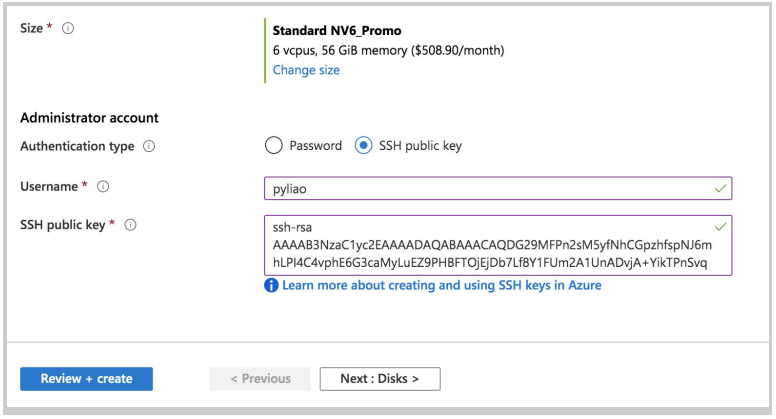
When prompted for a passphrase, either enter a passphrase to secure your private key, or leave it empty.

The public key will be saved to `~/.ssh/id_rsa.pub` by default. The public key looks like:

```
ssh-rsa
AAAAB3NzaC1yc2EAABADAQABAAACAQC1/KanayNr+Q7ogR5mKnGpKWRBQU7F3Jjhn7utdf7Z2i
UFykaYx+MInSnT3XdnBRS8KhC0IP8ptbngIaNOWd6zM8hB6UrcRTlTpwk/SuGMw1Vb40xlEFph
BkVEUgBolOoANIEXriAMvlDMZsgvnMFiQ12tD/u14cxy1WNEMAftey/vX3Fgp2vEq4zHXEliY/
sFZLJUJzcRUI0MOfHXAuCjg/qyqqbIuTDFyfg8k0JTtyGFEMQhbXKcuP2yGx1uw0ice62LRzr8
w0mszftXyMik1PnshRXbmE2xgINYg5xo/ra3mq2imwtOKJpfdtFoMiKhJmSNHBSkK7vFTeYgg0
v2cQ2+vL38lcIFX4Oh+QCzvNF/AXoDVlQtVtSqfQxRVG79Zqio5p12gHFktlfV7reCBvVIhyxc
2LlYUkrq4DHzkxNY5c9OGSHXSle9YsO3F1J5ip18f6gPq4xFmo6dVoJodZm9N0YMKCkZ4k1qJD
ESsJBk2ujDPmQQeMjJX3FnDXYYB182ZCGQzXfzlPDC29cWVgDZEXNHuYrOLmJTmYtLZ4WkdUhL
Llt5XsdoKWqlWpbegyYtGZgeZNRtOOdN6ybOPJqmYFd2qRtb4sYPniGJDOGhx4VodXAjT09omh
QJpE6wlZbRWDvKC55R2d/CSPHJscEiuudb+1SG2uA/oik/WQ== username@domainname
```

Copy this public key, or run: `cat ~/.ssh/id_rsa.pub | pbcopy`

Now in VM creation, choose SSH public key instead of Password in **Authentication type**. Enter your preferred Username. In the SSH public key field, paste the public key you just generated and copied.



Proceed with the remaining process. Now when you login to the VM, you won't be prompted for a password!