

NORTHEASTERN UNIVERSITY, KHOURY COLLEGE OF COMPUTER SCIENCE

CS 6120 Natural Language Processing — Lab 6 February 13, 2025 (Week 6)

YOUR NAME + LDAP

In this lab, we're going to work in Google's Workbench, an alternative to Google Colab, when you'd like more flexibility beyond Python notebooks. Workbench is a cloud offering in Google Cloud Platform (GCP), which operates in JupyterLab and the notebook environment. In Workbench, you can install packages as you see fit in the terminals and execute Python code in Jupyter Notebooks. It is also a lighter version of procuring a virtual machine, which we will be doing later in the semester by serving Large Language Models.

Ĩ	Start your Free Trial with \$300 in credit. Don't worry-	-you won't be charged if you run out	of credits. Learn more						DISMI	ss	START	T FREE
=	Google Cloud 🛟 CS6220 👻	Select a project			NEW PROJECT]	+	>	Ċ	0	:	
		RECENT STARRED	ALL									
		Name		ID								
	🔿 Welcon	✓☆ 🍨 <u>CS6220</u> 🕑		neu-cs6220								
	You're working in CS											
	Project number: 568122555											
	Dashboard Recommend					Gemini	lvanc 1.5 Pi	ed ro				
	Create a VM					mini		÷				
	Quick access											
	RPI APIS & Services				 CANCEL	e Engine						

1 Setup - Google Cloud

If you do not have access to your \$50 Google Cloud credits, please refer to Lab 1 from Week 1. By now for those of you who have enrolled in the class, you should have received an e-mail from me. If you have not, I will provide you with a URL you will need to access in order to request a Google Cloud coupon. You will be asked to provide your school email address and name (you can use the domains: @northeastern.edu or @ccs.neu.edu). An email will be sent to you to confirm these details before a coupon is sent to you. You can redeem this coupon until May 1, 2025 and use it until January 1, 2026. Go ahead and claim your credit.

2 Creating a JupyterLab Instance

In your new project, we'll now start up Jupyterlab, install some libraries, and then do some processing. The first step is to create your own Workbench instance by clicking on **New Instance**, which will subsequently allow us to open JupyterLab. When you click **New Instance**, you will be presented with a multitude of options. We won't worry too much about that now; the defaults are OK for now. However, please do change the name of your instance so that you can find it later.

When we start coding with LLMs, you will notice that they will run painfully slow without GPU's. GCP automatically sets a quota for your GPU use to safeguard against cost overruns and resourcing issues. In our case, we would explicitly like to use GPU's, and may need to increase our quota. To do so, in your left panel, select IAM & Admin and from there, Quotas & System Limits to appropriately set GPU's.

Go ahead and select a machine with a T4 (or any available) GPU to attach to your machine. When you've selected the appropriate options for you, **go ahead and submit** to create your instance. Depending on the GPU resources (and its expense), you may have to try several times before successful obtaining a Jupyter VM. After you've submitted, this may take a little while.

Ĩ	Start your Free Trial with \$300 in credit. Don't worry—you won't be charged if you run out of credits. Learn more 🖸												DISMISS START FRE		
≡	Google Cloud	•	CS6220 🔻	Search (/) for resources, docs, products, and more Q Search +								?	*		
\gg	Vertex AI		Workben	ch	+ NEW INST/	ANCE C REFRES	H 🕨 START	STOP	也 Res	ET			:	۲	
TOOL	5			diarate v	our instances to the	new Notebooks API	which manages you	ır Workbench Not	ebooks ar	nd provi	hhe sahi	litional			
99	Dashboard		• • ·	unctional	lity with no change i	in pricing. To get starte	d, click "Enable No	tebooks API". Lea	irn more		400 444	intionial			
ф	Model Garden		E	ENABLE NOTEBOOKS API											
11	Pipelines		Instances cr functionality	Instances created by using the Compute Engine API are called legacy instances. Legacy instances don't receive the latest updates to functionality and features that are available with the Notebooks API you must enable Notebooks API											
NOTE	BOOKS	^	and <u>register</u>	☑ your le	gacy instances.					,, , o a					
со	Colab Enterprise		च Filter Er	nter prope	erty name or value								Ø		
\bigotimes	Workbench			I	nstance name 🕇		Zone	Environmen	ıt	Mach	ine type		GPUs		
VERTE	EX AI STUDIO	^		<u>0</u>	llama-server	OPEN JUPYTERL	AB us- central1·	TensorFlov	v:2.11	Efficio Insta	ent nce: 4		None	9	
* *	Overview						а			vCPU RAM	s, 16 GB	3			
+	Multimodal														

When you've started your Jupyter GCP instance, you can access it via clicking on the instance. Click on **Open JupyterLab**, and you'll have access to your instance's dashboard, where there are multiple applications, including a file navigator (and downloader) on the side panel. Note that JupyterLab will stop after a few hours of inactivity to avoid unnecessary charges, but you should generally practice good hygiene and shut down whenver you're not using the resource.

	jupyter-notebook	Kernel Git Tabs	Settings Hel	o				n1-standard-4 🔻
	+ 🗈 ±	🛛 Launcher		+				
	Filter files by name $$ $$							
	I /		Notebook					
۰.	Name •							
-	 src tutorials 		2	Spark	Spark	Ċ	Ċ	C)
0			Python	Serverless	PySpark	PyTorch 1.12	PyTorch 1.13	PyTorch 2.0
•>				Spark				
			R	1	1	1	XGBoost	
=			R	TensorFlow 2.10	TensorFlow 2.11	TensorFlow 2.12	XGBoost 1.7	
Ō								
		>	Console					
			2	Spark	6	6	0	
			Python	PySpark	PyTorch 1.12	PyTorch 1.13	PyTorch 2.0	R
			1	1		dmk XGBoost		
		Те	usorElow 2.10	TensorFlow 2 11	TensorFlow 2 12	XGBoost 17		
			1301110W 2.10					
		\$_	Other					
			Ś		Μ	2	R	
			Terminal	Text File	Markdown File	Python File	R File	Show
								Contextual Help
		\$	Dataproc S	Serverless Notel	books			
			+					
		N	ew Runtime					
			Template					
		(C)	Gooale Cla	oud Resources				

The terminal gives you access to the underlying VM, which can also be used to install packages and manipulate your hardware (e.g., NVIDIA GPU's). You can type in commands to see what files you have access to. In particular, we would like to check the GPU status. We can do so with the nvidia-smi command, which should give you the following screen:

```
(base) jupyter@jupyter-gpu:~$ nvidia-smi
Mon Dec 30 23:04:22 2024
```

+ NVIDIA-SMI 550.90.07	Driver	Version: 550.90.07	CUDA Version: 12.4
GPU Name GPU Name Fan Temp Perf 	Persistence-M Pwr:Usage/Cap	Bus-Id Disp.A Memory-Usage	Volatile Uncorr. ECC GPU-Util Compute M. MIG M.
0 Tesla T4 N/A 44C P8 	On 9W / 70W	 0000000:00:04.0 Off 1MiB / 15360MiB	0 0% Default N/A
+			

l l	GPU	GI ID	CI ID	PID	Туре	Process name		GPU Usaç	Memory Je		
==	No running processes found										
(ba	base) jupyter@jupyter-gpu:~\$										

If you see the above, congratulations! You have managed to secure a JupyterLab Launcher with a GPU. Sometimes they are short in supply.

3 Running a Jupyter Notebook with GPUs

We will now try out our notebooks in JupyterLabs. Go ahead and open a Tensorflow 2-11 Notebook and enter the below into the first cell and run it.

```
import tensorflow as tf
print("Num GPUs Available: ", len(tf.config.list_physical_devices('GPU')))
```

If the above shows that it recognizes a GPU, we know that the notebook kernel has access to the appropriate hardware. Now we must ensure that the code uses it appropriately. For this, there are several neural networks at keras.io. We're going to start with a miniature Transformer Network: the FNet Model. Let's download the FNet Notebook Training Model in a terminal with "curl" or "wget" commands. There, we will have instructions to download the data in the terminal. Go ahead and run through the instructions in the lab.

4 What to Submit

Here are the artifacts to submit:

- Screenshot of the downloaded data in your terminal. (You can type ls) or a screenshot of it downloading in the terminal.
- A picture of the results of nvidia-smi, but making sure that **GPU usage is non-zero**. This is easiest done while the F-Net is *training*