

11-785 Introduction to Deep Learning

Fall 2025

GCP setup

TAs:

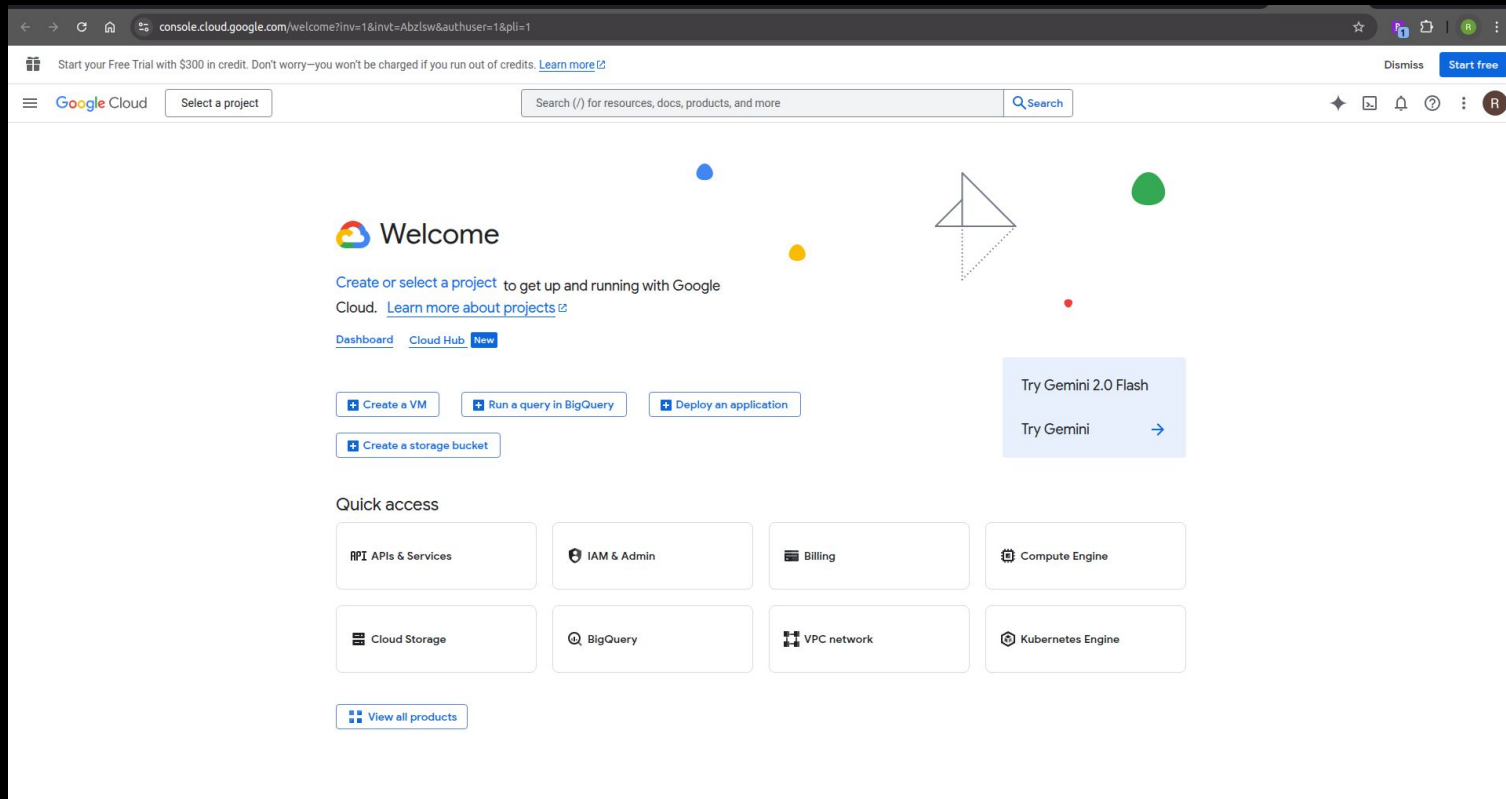
Rutvik Joshi

Alex Moker

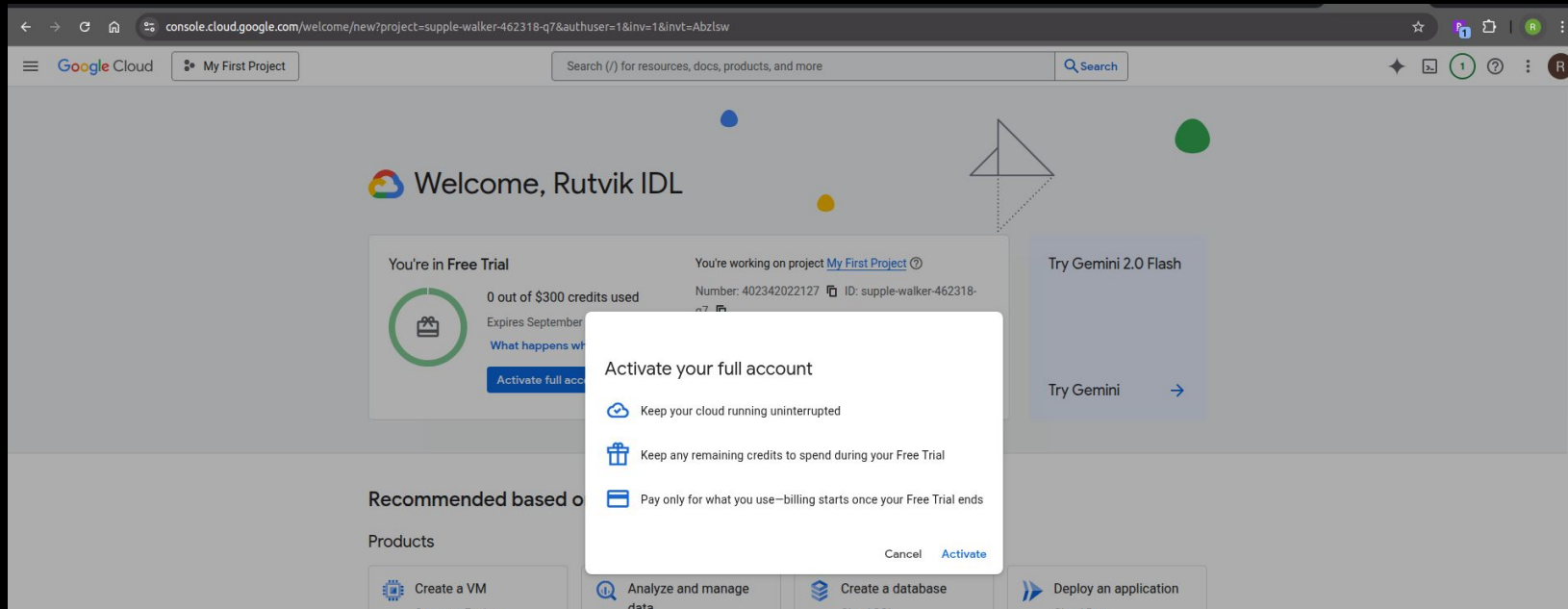


Creating GCP account:

1. Log into and avail the free 300\$ credits. If you have already utilized this before create a new account. You will be asked to enter some payment method



2. Now activate your full account



3. Check credits available

console.cloud.google.com/billing/019D43-57033A-475EE9/credits/all?authuser=1&inv=1&inv=Abzls&project=supple-walker-462318-q7

Google Cloud Search (/) for resources, docs, products, and more Search

Billing / Credits / All Credits

Billing account
My Billing Account

Overview

Cost management

- Reports
- Cost table
- Cost breakdown
- Budgets & alerts
- Billing export
- Anomalies

Cost optimization

- FinOps hub
- Committed use discou...
- CUD analysis
- Pricing
- Cost estimation
- Credits

Payments

- Documents
- Transactions

Credits

[All Credits](#)

View and download credit details here. Active committed use discounts are not included here and can be viewed on the [Commitments page](#).

Filter Filter credits

Credit name	Status ↑	Percent remaining	Remaining value	Original value	Type	Credit ID	Scope ⓘ	Start date	End date
Free Trial	Available	<div><div></div></div> 100%	\$300.00	\$300.00	One-time	FreeTrialUp...	Certain usage; see the terms for free trial.	June 8, 2025	September 7, 2025
Free Trial	Expired	—	\$300.00	\$300.00	One-time	FreeTrial:Cr...	Certain usage; see the terms for free trial.	June 8, 2025	June 8, 2025

4. Open “All Quotas” and check for value in Compute Engine API. If the value is “0”, select “Compute Engine API” and click the “Edit” button and add value of 1. You can add a description like “For my IDL project”

The screenshot shows the Google Cloud IAM & Admin console. The left sidebar contains navigation links for IAM, PAM, Principal Access Boundaries, Organizations, Identity & Organization, Policy Troubleshooter, Policy Analyzer, Organization Policies, Service Accounts, Workload Identity Federation, Workforce Identity Federation, Labels, Tags, Settings, Privacy & Security, Identity-Aware Proxy, Roles, Audit logs, Essential Contacts, Asset Inventory, **Quotas & System Limits**, Manage Resources, and Release Notes. The main content area is titled 'Quotas & System Limits for project "My First Project"'. It includes a 'QUOTAS & SYSTEM LIMITS' tab and an 'INCREASE REQUESTS' link. A message states: 'Values for Compute Engine API - CPUs per VM family are being updated. This may take 2-3 weeks to complete.' Below this, a summary shows 'Current usage > 90%' as 0 and 'All quotas & system limits' as 20,620. A table lists various quotas with columns for Service, Name, Type, Dimensions, Value, Current usage percentage, Current usage, and Adjustable. The 'Compute Engine API' service is selected, and its quotas are listed.

Service	Name	Type	Dimensions (e.g. location)	Value	Current usage percentage	Current usage	Adjustable
Compute Engine API	Networks	Quota		5	20%	1	Yes
OS Config API	OsPolicyAssignment	Quota	zone: us-central1-f	20	5%	1	Yes
Compute Engine API	Firewall rules	Quota		100	3%	3	Yes
Compute Engine API	Heavy-weight read requests per minute	Quota		2,250	0.67%	15	Yes
Compute Engine API	Subnetwork ranges per VPC Network	Quota	network_id: default	500	0.6%	3	Yes
Compute Engine API	Read requests per minute	Quota		6,000	0.51%	30.75	Yes
Compute Engine API	Static routes per VPC Network	Quota	network_id: default	200	0.5%	1	Yes
Compute Engine API	Read requests per minute per region	Quota	region: us-central1	6,000	0.45%	27	Yes

Rows per page: 50 | 1 - 50 of 20620

Recommended for you:

- Working with quotas
- Viewing your quota in the Google Cloud console
- Managing your quota using the Google Cloud console
- Service Quota Model
- Managing Service Quota
- Chart and monitor quota metrics

4. Open “All Quotas” and check for value in Compute Engine API. If the value is “0”, select “Compute Engine API and click the “Edit” button and add value of 1. You can add a description like “For my IDL project”

The screenshot shows the Google Cloud IAM & Admin console. The left sidebar contains navigation links: IAM, PAM, Principal Access Boundary, Organizations (PREVIEW), Identity & Organization, Policy Troubleshooter, Policy Analyzer, Organization Policies, Service Accounts, Workload Identity Federat..., Workforce Identity Federa..., Labels, Tags, Settings, Privacy & Security, and Identity-Aware Proxy. The main content area is titled 'Quotas & System Limits for project "My First Project"'. It shows a table with 'Current usage > 90%' and 'All quotas & system limits'. The 'Compute Engine API' quota is selected, showing a current value of 0. The 'Edit quota' modal is open, showing the 'Compute Engine API' quota with a current value of 0 and a new value of 1 entered. A description 'For my Deep Learn' is being added.

Quotas & System Limits for project "My First Project"

1 quota selected [EDIT](#)

Set up quota & system limit alerts
Get alerted if a quota is close to reaching its maximum. Click on more actions button ⓘ in a row to get started, or click [LEARN MORE](#)

Values for [Compute Engine API - Local SSD per VM family \(GB\) \(default\)](#) are being updated. This may take 2-3 weeks to complete.

Current usage > 90%	All quotas & system limits
0	17,535

[View quotas & system limits](#)

Filter [Metric : compute.googleapis.com/gpus_all_regions](#) Enter property name or value

Service	Name	Type	Dimensions (e.g. location)	Value	Current usage percentage
<input checked="" type="checkbox"/>	Compute Engine API	GPUs (all regions)	Quota	0	<div></div>

1 quota selected

Step 1/2

Quota changes
Expand each service card to change individual quotas.

Edit quota

Compute Engine API

Quota: GPUs (all regions)
Current value: 0
Enter a new quota value. A value above 0 will require approval from your service provider. ⓘ

New value *

Request description *

Your description will be sent to your service provider and is used to evaluate your request. It's useful to include the intent of the quota usage, future growth plans, region or zone spread, and any additional requirements or dependencies.

[DONE](#)

[NEXT](#)

You can also check for which GPU regions have what compute available from <https://cloud.google.com/compute/docs/gpus/gpu-regions-zones>. This would be helpful in your later assignments

Google has deprecated their Colab instance :(so when will be looking into VM instances. These are similar to EC2 instances provided by AWS.

Move to your local terminal and execute below commands. I am using Linux so you may need to modify the code for Mac or Windows. Do remove the comments when pasting.

```
$ cd ~ # Go to your home directory
```

```
$ mkdir .ssh # Create a .ssh folder if you don't have it
```

```
$ cd .ssh
```

```
$ ssh-keygen -t rsa -b 4096 # This is to create a private key
```

```
$ Enter file to save the key: IDL_GCP # This will create a file IDL_GCP.pub
```

```
$ cat IDL_GCP.pub # To check contents of the file
```

Copy the contents and open SSH keys on GCP. Paste this and remove the last part till the “@” symbol including it. Also add another space after the “==” so that now it is “== your_id”

4. Search for Compute engine and click create Instance. This would ask you to enable APIs. Please enable any APIs that are asked, this is not charged
5. Select a Machine Configuration as needed.

console.cloud.google.com/compute/instancesAdd?authuser=1&inv=1&inv=Abzls&project=supple-walker-462318-q7

Google Cloud My First Project VM Search

Create an instance Create VM from...

Machine configuration

- Machine configuration
 - n1-highmem-4 (1 NVIDIA T4), us-west2
- OS and storage
 - Debian GNU/Linux 12 (bookworm)
- Data protection
 - Snapshot schedules
- Networking
 - 1 network interface
- Observability
 - Install Ops Agent
- Security
- Advanced

Machine configuration

Name * instance-20250608-191932 ⓘ

Region * us-west2 (Los Angeles) ⓘ Zone * Any ⓘ

Region is permanent. Google will choose a zone on your behalf, maximizing VM obtainability. Zone is permanent.

General purpose Compute optimized Memory optimized Storage optimized **GPUs**

Graphics processing units (GPUs) accelerate specific workloads on your instances such as machine learning and data processing. [Learn More](#)

GPU type NVIDIA T4 ⓘ Number of GPUs 1 ⓘ

☐ Enable Virtual Workstation (NVIDIA GRID)

Series ⓘ	Description	vCPUs ⓘ	Memory ⓘ	CPU Platform
N1	Balanced price & performance	1 - 96	1.8 - 624 GB	Intel Haswell

Machine type

Choose a machine type with preset amounts of vCPUs and memory that suit most workloads. Or, you can create a custom machine for your workload's particular needs. [Learn more](#)

Preset Custom

n1-highmem-4 (4 vCPU, 2 core, 26 GB memory)

vCPU 4 (2 cores) Memory 26 GB

[Advanced configurations](#)

Monthly estimate

\$355.93

That's about \$0.49 hourly

Pay for what you use: no upfront costs and per second billing

Item	Monthly estimate
4 vCPU + 26 GB memory	\$207.46
1 NVIDIA T4	\$299.30
10 GB balanced persistent disk	\$1.20
Use discount	-\$152.03
Logging	Cost varies ⓘ
Monitoring	Cost varies ⓘ
Snapshot schedule	Cost varies ⓘ
Total	\$355.93

[Compute Engine pricing](#) ⓘ

[Cloud Operations pricing](#) ⓘ

[Less](#)

Create Cancel [Equivalent code](#)

Copy the external IP given here and open VC code.

Install Remote-ssh client extension from Microsoft.

Click on the “Connect to Host” then click “Configure SSH Hosts”, default ssh path and add below code

```

```
Host IDL_GCP_VM
```

```
HostName ExternalIP
```

```
IdentityFile ~/.ssh/
```

```
User your_name
```

```

The screenshot displays the Google Cloud Platform console interface. The left sidebar shows the navigation menu with 'Compute Engine' selected. Under 'Virtual machines', 'VM instances' is highlighted. The main content area shows a table of VM instances. One instance is listed with the name 'instance-20250608-213956', located in the 'us-west4-b' zone. The table includes columns for Status, Name, Zone, Recommendations, In use by, Internal IP, External IP, and Connect. The 'Connect' column shows an SSH button. Below the table, there are several 'Related actions' cards, including 'Explore protection summary', 'View billing report', 'Monitor VMs', 'Explore VM logs', 'Set up firewall rules', 'Patch management', and 'Load balance between VMs'. On the right side, there is a 'Get started with Compute Engine' section with various tutorials and links.

Status	Name	Zone	Recommendations	In use by	Internal IP	External IP	Connect
<input checked="" type="checkbox"/>	instance-20250608-213956	us-west4-b			10.182.0.4 (nic0)	34.16.160.147 (nic0)	SSH

Create a [setup.sh](#) file in the GCP instance and paste the code given in the notebook.

You need to update you kaggle username and key in the code.

Now you have python and torch running on your instance. If the drivers are installed correctly you will also have GPU enabled pytorch.

You are free to update the driver commands as Nvidia releases new versions of the drivers.

Good Luck!