Install gcloud command line tool to access quotas:

[Installing Google Cloud SDK | Cloud SDK Documentation]
Install Google Cloud on a Ubuntu machine:

```
diego@EstudioVM:~$ echo "deb [signed-by=/usr/share/keyrings/cloud.google.gpg]
https://packages.cloud.google.com/apt cloud-sdk main" | sudo tee -a
/etc/apt/sources.list.d/google-cloud-sdk.list
[sudo] contraseña para diego:
deb [signed-by=/usr/share/keyrings/cloud.google.gpg]
https://packages.cloud.google.com/apt cloud-sdk main

diego@EstudioVM:~$ sudo apt-get install apt-transport-https ca-certificates gnupg
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias
Leyendo la información de estado... Hecho
gnupg ya está en su versión más reciente (2.2.19-3ubuntu2).
fijado gnupg como instalado manualmente.
ca-certificates ya está en su versión más reciente (20210119~20.04.1).
Los paquetes indicados a continuación se instalaron de forma automática y ya no
son necesarios.
  libfprint-2-tod1 libllvm9 linux-modules-extra-5.4.0-58-generic python3-click
  python3-colorama
Utilice «sudo apt autoremove» para eliminarlos.
Se actualizarán los siguientes paquetes:
  apt-transport-https
1 actualizados, 0 nuevos se instalarán, 0 para eliminar y 106 no actualizados.
Se necesita descargar 1.704 B de archivos.
Se utilizarán 0 B de espacio de disco adicional después de esta operación.
¿Desea continuar? [S/n] s
Des:1 http://es.archive.ubuntu.com/ubuntu focal-updates/universe amd64
apt-transport-https all 2.0.4 [1.704 B]
Descargados 1.704 B en 0s (7.038 B/s)
(Leyendo la base de datos ... 275235 ficheros o directorios instalados
actualmente.)
Preparando para desempaquetar .../apt-transport-https_2.0.4_all.deb ...
Desempaquetando apt-transport-https (2.0.4) sobre (2.0.2ubuntu0.2) ...
Configurando apt-transport-https (2.0.4) …

diego@EstudioVM:~$ curl https://packages.cloud.google.com/apt/doc/apt-key.gpg |
sudo apt-key --keyring /usr/share/keyrings/cloud.google.gpg add -
  % Total  % Received % Xferd  Average Speed   Time   Time  Time  Current
                                 Dload  Upload   Total   Spent   Left  Speed
100  1974  100  1974 0     0  10175   0 --:--:-- --:--:-- --:--:-- 10123
OK

diego@EstudioVM:~$ sudo apt-get update && sudo apt-get install google-cloud-sdk
Obj:1 https://download.docker.com/linux/ubuntu bionic InRelease
Des:2 http://security.ubuntu.com/ubuntu focal-security InRelease [109 kB]
Obj:3 http://es.archive.ubuntu.com/ubuntu focal InRelease
Des:4 http://es.archive.ubuntu.com/ubuntu focal-updates InRelease [114 kB]
```

```
Des:5 http://es.archive.ubuntu.com/ubuntu focal-backports InRelease [101 kB]
Des:6 http://security.ubuntu.com/ubuntu focal-security/main amd64 DEP-11
Metadat                    a [24,3 kB]
Des:7 http://security.ubuntu.com/ubuntu focal-security/main amd64 c-n-f
Metadata                   [6.164 B]
Des:8 http://security.ubuntu.com/ubuntu focal-security/universe amd64 DEP-11
Met                   adata [56,6 kB]
Des:9 http://es.archive.ubuntu.com/ubuntu focal-updates/main i386 Packages [415
     kB]
Des:10 https://packages.cloud.google.com/apt cloud-sdk InRelease [6.739 B]
Des:11 http://es.archive.ubuntu.com/ubuntu focal-updates/main amd64 Packages
[79                    7 kB]
Des:12 https://packages.cloud.google.com/apt cloud-sdk/main i386 Packages [136
k                      B]
Des:13 https://packages.cloud.google.com/apt cloud-sdk/main amd64 Packages [156
     kB]
Des:14 http://es.archive.ubuntu.com/ubuntu focal-updates/main amd64 DEP-11
Metad                      ata [264 kB]
Des:15 http://es.archive.ubuntu.com/ubuntu focal-updates/restricted amd64
Packag                     es [144 kB]
Des:16 http://es.archive.ubuntu.com/ubuntu focal-updates/universe i386 Packages
     [547 kB]
Des:17 http://es.archive.ubuntu.com/ubuntu focal-updates/universe amd64
Packages                   [737 kB]
Des:18 http://es.archive.ubuntu.com/ubuntu focal-updates/universe amd64 DEP-11
M                     etadata [301 kB]
Des:19 http://es.archive.ubuntu.com/ubuntu focal-updates/multiverse amd64
DEP-11                     Metadata [2.468 B]
Des:20 http://es.archive.ubuntu.com/ubuntu focal-backports/universe amd64
DEP-11                     Metadata [1.768 B]
Descargados 3.919 kB en 2s (1.870 kB/s)
Leyendo lista de paquetes... Hecho
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias
Leyendo la información de estado... Hecho
Los paquetes indicados a continuación se instalaron de forma automática y ya no
son necesarios.
  libfprint-2-tod1 libllvm9 linux-modules-extra-5.4.0-58-generic python3-click
python3-colorama
Utilice «sudo apt autoremove» para eliminarlos.
Se instalarán los siguientes paquetes adicionales:
  python3-crcmod
Paquetes sugeridos:
  google-cloud-sdk-app-engine-java google-cloud-sdk-app-engine-python
  google-cloud-sdk-pubsub-emulator google-cloud-sdk-bigtable-emulator
  google-cloud-sdk-datastore-emulator kubectl
Se instalarán los siguientes paquetes NUEVOS:
  google-cloud-sdk python3-crcmod
0 actualizados, 2 nuevos se instalarán, 0 para eliminar y 106 no actualizados.
Se necesita descargar 92,6 MB de archivos.
Se utilizarán 492 MB de espacio de disco adicional después de esta operación.
¿Desea continuar? [S/n]
```

```
Des:1 http://es.archive.ubuntu.com/ubuntu focal/universe amd64 python3-crcmod
amd64 1.7+dfsg-2build2 [18,8 kB]
Des:2 https://packages.cloud.google.com/apt cloud-sdk/main amd64
google-cloud-sdk all 326.0.0-0 [92,6 MB]
Descargados 92,6 MB en 16s (5.828 kB/s)
Seleccionando el paquete python3-crcmod previamente no seleccionado.
(Leyendo la base de datos ... 275235 ficheros o directorios instalados
actualmente.)
Preparando para desempaquetar .../python3-crcmod_1.7+dfsg-2build2_amd64.deb ...
Desempaquetando python3-crcmod (1.7+dfsg-2build2) ...
Seleccionando el paquete google-cloud-sdk previamente no seleccionado.
Preparando para desempaquetar .../google-cloud-sdk_326.0.0-0_all.deb ...
Desempaquetando google-cloud-sdk (326.0.0-0) ...
Configurando python3-crcmod (1.7+dfsg-2build2) ...
Configurando google-cloud-sdk (326.0.0-0) ...
Procesando disparadores para man-db (2.9.1-1) ...
```

I select us-west1-b because it seems it is the cheapest region for a K80 Nvidia GPU.

```
diego@EstudioVM:~$ gcloud init
Welcome! This command will take you through the configuration of gcloud.

Your current configuration has been set to: [default]

You can skip diagnostics next time by using the following flag:
  gcloud init --skip-diagnostics

Network diagnostic detects and fixes local network connection issues.
Checking network connection...done.
Reachability Check passed.
Network diagnostic passed (1/1 checks passed).

You must log in to continue. Would you like to log in (Y/n)?

Go to the following link in your browser:


https://accounts.google.com/o/oauth2/auth?response_type=code&client_id=32555940
559.apps.googleuse
rcontent.com&redirect_uri=urn%3Aietf%3Awg%3Aoauth%3A2.0%3Aoob&scope=openid+http
s%3A%2F%2Fwww.googleap
is.com%2Fauth%2Fuserinfo.email+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fcloud-
platform+https%3A%2F%2
Fwww.googleapis.com%2Fauth%2Fappengine.admin+https%3A%2F%2Fwww.googleapis.com%2
Fauth%2Fcompute+https%
3A%2F%2Fwww.googleapis.com%2Fauth%2Faccounts.reauth&state=heAdLxFKVCxJxoZOZ8U6Z
TVwDQD9Lr&prompt=conse
nt&access_type=offline&code_challenge=zDcjDJyylCMcbOTHbPBup-V_dEQZzO0osMWgzTwXl
rw&code_challenge_meth   od=S256

Enter verification code:
4/1AY0e-g7Y-jz7nMo8U4yReIMnpUsX_EfDmzFpX7AgcOVA1xNUoo4syKMDwmw
```

```
You are logged in as: [diegotascon@gmail.com].

Pick cloud project to use:
 [1] aidl2021project
 [2] udemy-nodejs-279112
 [3] Create a new project
Please enter numeric choice or text value (must exactly match list
item):  1

Your current project has been set to: [aidl2021project].

Do you want to configure a default Compute Region and Zone? (Y/n)?

Which Google Compute Engine zone would you like to use as project
default?
If you do not specify a zone via a command line flag while working
with Compute Engine resources, the default is assumed.
 [1] us-east1-b
 [2] us-east1-c
 [3] us-east1-d
 [4] us-east4-c
 [5] us-east4-b
 [6] us-east4-a
 [7] us-central1-c
 [8] us-central1-a
 [9] us-central1-f
 [10] us-central1-b
 [11] us-west1-b
 [12] us-west1-c
 [13] us-west1-a
 [14] europe-west4-a
 [15] europe-west4-b
 [16] europe-west4-c
 [17] europe-west1-b
 [18] europe-west1-d
 [19] europe-west1-c
 [20] europe-west3-c
 [21] europe-west3-a
 [22] europe-west3-b
 [23] europe-west2-c
 [24] europe-west2-b
 [25] europe-west2-a
 [26] asia-east1-b
 [27] asia-east1-a
 [28] asia-east1-c
 [29] asia-southeast1-b
 [30] asia-southeast1-a
 [31] asia-southeast1-c
 [32] asia-northeast1-b
 [33] asia-northeast1-c
 [34] asia-northeast1-a
 [35] asia-south1-c
```

```
 [36] asia-south1-b
 [37] asia-south1-a
 [38] australia-southeast1-b
 [39] australia-southeast1-c
 [40] australia-southeast1-a
 [41] southamerica-east1-b
 [42] southamerica-east1-c
 [43] southamerica-east1-a
 [44] asia-east2-a
 [45] asia-east2-b
 [46] asia-east2-c
 [47] asia-northeast2-a
 [48] asia-northeast2-b
 [49] asia-northeast2-c
 [50] asia-northeast3-a
Did not print [24] options.
Too many options [74]. Enter "list" at prompt to print choices fully.
Please enter numeric choice or text value (must exactly match list
item):  us-west1-b

Your project default Compute Engine zone has been set to [us-west1-b].
You can change it by running [gcloud config set compute/zone NAME].

Your project default Compute Engine region has been set to [us-west1].
You can change it by running [gcloud config set compute/region NAME].

Created a default .boto configuration file at [/home/diego/.boto]. See this
file and
[https://cloud.google.com/storage/docs/gsutil/commands/config] for more
information about configuring Google Cloud Storage.
Your Google Cloud SDK is configured and ready to use!

* Commands that require authentication will use diegotascon@gmail.com by
default
* Commands will reference project `aidl2021project` by default
* Compute Engine commands will use region `us-west1` by default
* Compute Engine commands will use zone `us-west1-b` by default

Run `gcloud help config` to learn how to change individual settings

This gcloud configuration is called [default]. You can create additional
configurations if you work with multiple accounts and/or projects.
Run `gcloud topic configurations` to learn more.

Some things to try next:

* Run `gcloud --help` to see the Cloud Platform services you can interact with.
And run `gcloud help COMMAND` to get help on any gcloud command.
* Run `gcloud topic --help` to learn about advanced features of the SDK like
arg files and output formatting
```

Creating VMs with attached GPUs

Creating VMs with attached GPUs | Compute Engine Documentation (google.com)

Checking GPU quota

Creating VMs with attached GPUs | Compute Engine Documentation (google.com)

Do I have sufficient GPU quota in the region?

```
diego@EstudioVM:~$ gcloud compute regions describe us-west1
creationTimestamp: '1969-12-31T16:00:00.000-08:00'
description: us-west1
id: '1210'
kind: compute#region
name: us-west1
quotas:
- limit: 24.0
  metric: CPUS
  usage: 0.0
- limit: 4096.0
  metric: DISKS_TOTAL_GB
  usage: 0.0
- limit: 8.0
  metric: STATIC_ADDRESSES
  usage: 0.0
- limit: 8.0
  metric: IN_USE_ADDRESSES
  usage: 0.0
- limit: 500.0
  metric: SSD_TOTAL_GB
  usage: 0.0
- limit: 6000.0
  metric: LOCAL_SSD_TOTAL_GB
  usage: 0.0
- limit: 100.0
  metric: INSTANCE_GROUPS
  usage: 0.0
- limit: 50.0
  metric: INSTANCE_GROUP_MANAGERS
  usage: 0.0
- limit: 24.0
  metric: INSTANCES
  usage: 0.0
- limit: 50.0
  metric: AUTOSCALERS
  usage: 0.0
- limit: 20.0
  metric: REGIONAL_AUTOSCALERS
  usage: 0.0
- limit: 20.0
  metric: REGIONAL_INSTANCE_GROUP_MANAGERS
```

```
  usage: 0.0
- limit: 0.0
  metric: PREEMPTIBLE_CPUS
  usage: 0.0
- limit: 1.0
  metric: NVIDIA_K80_GPUS
  usage: 0.0
- limit: 0.0
  metric: COMMITTED_CPUS
  usage: 0.0
- limit: 0.0
  metric: COMMITTED_LOCAL_SSD_TOTAL_GB
  usage: 0.0
- limit: 0.0
  metric: COMMITMENTS
  usage: 0.0
- limit: 100.0
  metric: NETWORK_ENDPOINT_GROUPS
  usage: 0.0
- limit: 200.0
  metric: INTERNAL_ADDRESSES
  usage: 0.0
- limit: 1.0
  metric: NVIDIA_P100_GPUS
  usage: 0.0
- limit: 0.0
  metric: PREEMPTIBLE_LOCAL_SSD_GB
  usage: 0.0
- limit: 1.0
  metric: PREEMPTIBLE_NVIDIA_K80_GPUS
  usage: 0.0
- limit: 1.0
  metric: PREEMPTIBLE_NVIDIA_P100_GPUS
  usage: 0.0
- limit: 1.0
  metric: NVIDIA_P100_VWS_GPUS
  usage: 0.0
- limit: 1.0
  metric: NVIDIA_V100_GPUS
  usage: 0.0
- limit: 1.0
  metric: NVIDIA_P4_GPUS
  usage: 0.0
- limit: 1.0
  metric: NVIDIA_P4_VWS_GPUS
  usage: 0.0
- limit: 100.0
  metric: NODE_GROUPS
  usage: 0.0
- limit: 100.0
  metric: NODE_TEMPLATES
  usage: 0.0
```

```
- limit: 1.0
  metric: PREEMPTIBLE_NVIDIA_V100_GPUS
  usage: 0.0
- limit: 1.0
  metric: PREEMPTIBLE_NVIDIA_P4_GPUS
  usage: 0.0
- limit: 1.0
  metric: PREEMPTIBLE_NVIDIA_P100_VWS_GPUS
  usage: 0.0
- limit: 1.0
  metric: PREEMPTIBLE_NVIDIA_P4_VWS_GPUS
  usage: 0.0
- limit: 16.0
  metric: INTERCONNECT_ATTACHMENTS_PER_REGION
  usage: 0.0
- limit: 80000.0
  metric: INTERCONNECT_ATTACHMENTS_TOTAL_MBPS
  usage: 0.0
- limit: 5.0
  metric: RESOURCE_POLICIES
  usage: 0.0
- limit: 20.0
  metric: IN_USE_SNAPSHOT_SCHEDULES
  usage: 0.0
- limit: 1.0
  metric: NVIDIA_T4_GPUS
  usage: 0.0
- limit: 1.0
  metric: NVIDIA_T4_VWS_GPUS
  usage: 0.0
- limit: 1.0
  metric: PREEMPTIBLE_NVIDIA_T4_GPUS
  usage: 0.0
- limit: 1.0
  metric: PREEMPTIBLE_NVIDIA_T4_VWS_GPUS
  usage: 0.0
- limit: 20.0
  metric: IN_USE_BACKUP_SCHEDULES
  usage: 0.0
- limit: 10.0
  metric: PUBLIC_DELEGATED_PREFIXES
  usage: 0.0
- limit: 0.0
  metric: COMMITTED_NVIDIA_K80_GPUS
  usage: 0.0
- limit: 0.0
  metric: COMMITTED_NVIDIA_P100_GPUS
  usage: 0.0
- limit: 0.0
  metric: COMMITTED_NVIDIA_P4_GPUS
  usage: 0.0
- limit: 0.0
```

```
  metric: COMMITTED_NVIDIA_V100_GPUS
  usage: 0.0
- limit: 0.0
  metric: COMMITTED_NVIDIA_T4_GPUS
  usage: 0.0
- limit: 8.0
  metric: C2_CPUS
  usage: 0.0
- limit: 8.0
  metric: N2_CPUS
  usage: 0.0
- limit: 0.0
  metric: COMMITTED_N2_CPUS
  usage: 0.0
- limit: 0.0
  metric: COMMITTED_C2_CPUS
  usage: 0.0
- limit: 100.0
  metric: RESERVATIONS
  usage: 0.0
- limit: 0.0
  metric: COMMITTED_LICENSES
  usage: 0.0
- limit: 8.0
  metric: N2D_CPUS
  usage: 0.0
- limit: 0.0
  metric: COMMITTED_N2D_CPUS
  usage: 0.0
- limit: 128.0
  metric: STATIC_BYOIP_ADDRESSES
  usage: 0.0
- limit: 0.0
  metric: AFFINITY_GROUPS
  usage: 0.0
- limit: 0.0
  metric: NVIDIA_A100_GPUS
  usage: 0.0
- limit: 0.0
  metric: PREEMPTIBLE_NVIDIA_A100_GPUS
  usage: 0.0
- limit: 0.0
  metric: COMMITTED_NVIDIA_A100_GPUS
  usage: 0.0
- limit: 0.0
  metric: M1_CPUS
  usage: 0.0
- limit: 0.0
  metric: M2_CPUS
  usage: 0.0
- limit: 0.0
  metric: A2_CPUS
```

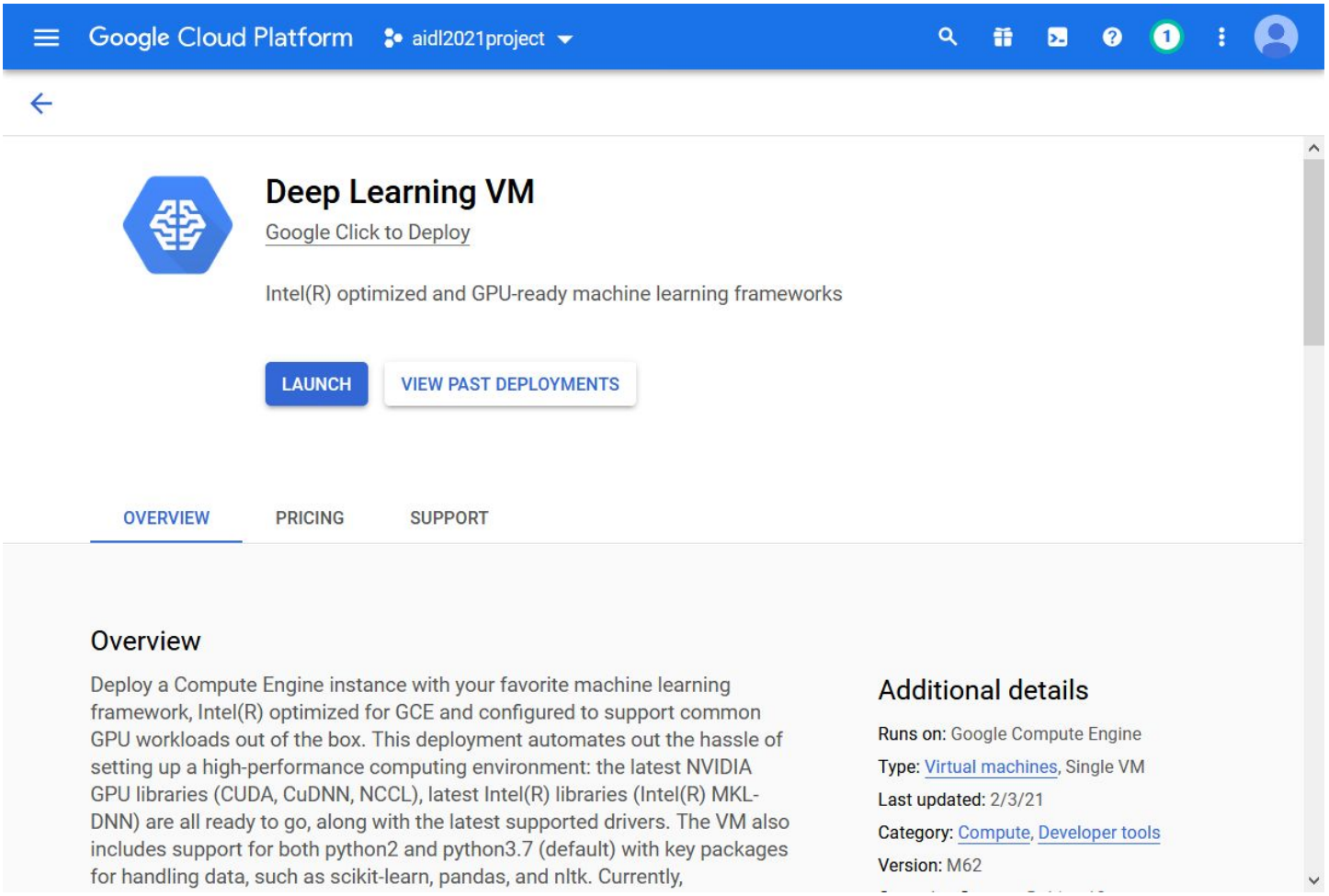
```
  usage: 0.0
- limit: 0.0
  metric: COMMITTED_A2_CPUS
  usage: 0.0
- limit: 0.0
  metric: COMMITTED_MEMORY_OPTIMIZED_CPUS
  usage: 0.0
selfLink:
https://www.googleapis.com/compute/v1/projects/aidl2021project/regions/us-west1
status: UP
supportsPzs: false
zones:
-
https://www.googleapis.com/compute/v1/projects/aidl2021project/zones/us-west1-a
-
https://www.googleapis.com/compute/v1/projects/aidl2021project/zones/us-west1-b
-
https://www.googleapis.com/compute/v1/projects/aidl2021project/zones/us-west1-c
```

It seems I have a limit of 1 Nvidia K80 GPU. That should be enough.

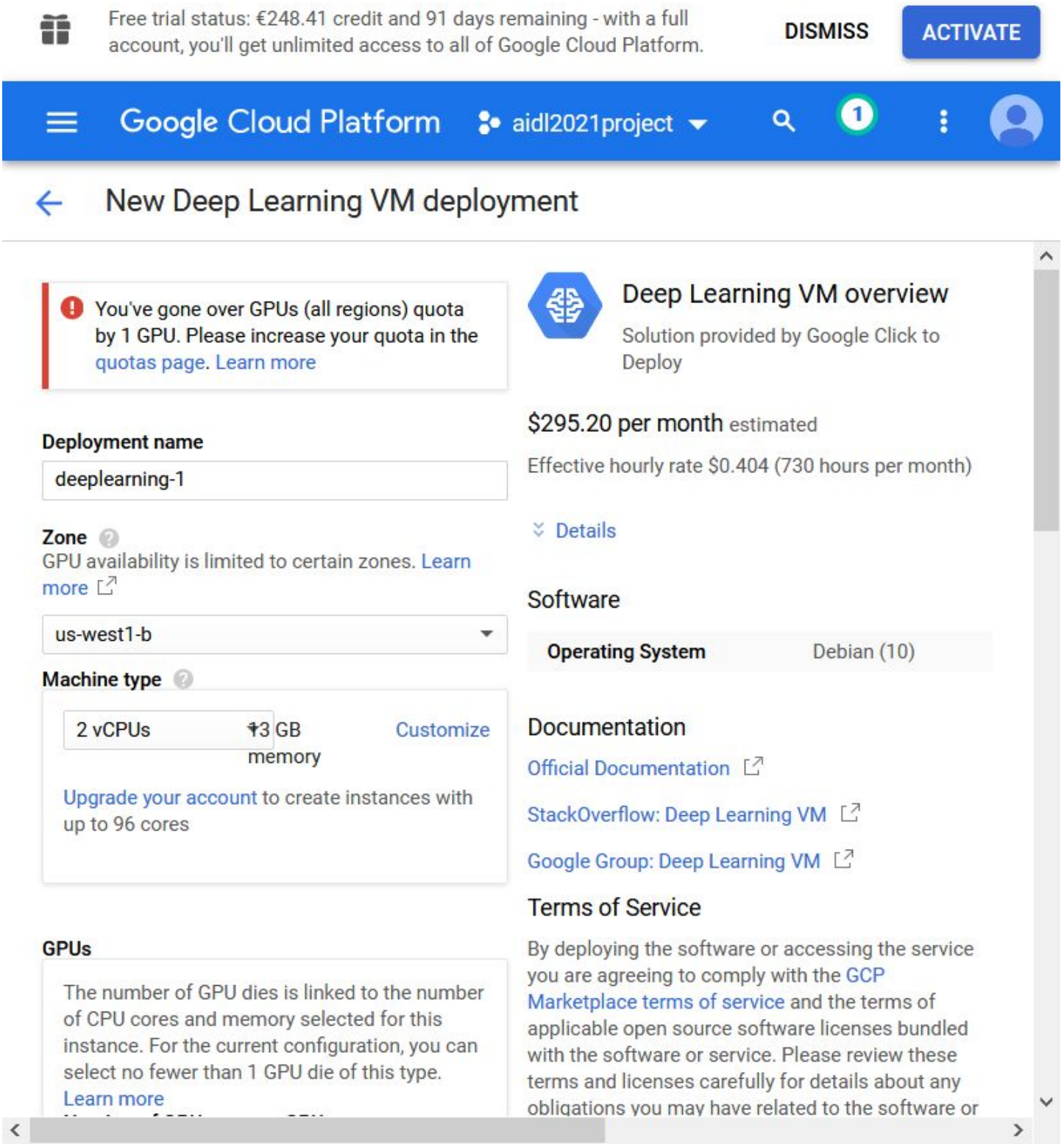
Create an instance with one GPU

[Creating a Deep Learning VM Instance From the Google Cloud Marketplace]

I open
https://console.cloud.google.com/marketplace/details/click-to-deploy-images/deeplearning?_ga=2.13134391.1007356524.1612367677-1479030961.1612367677

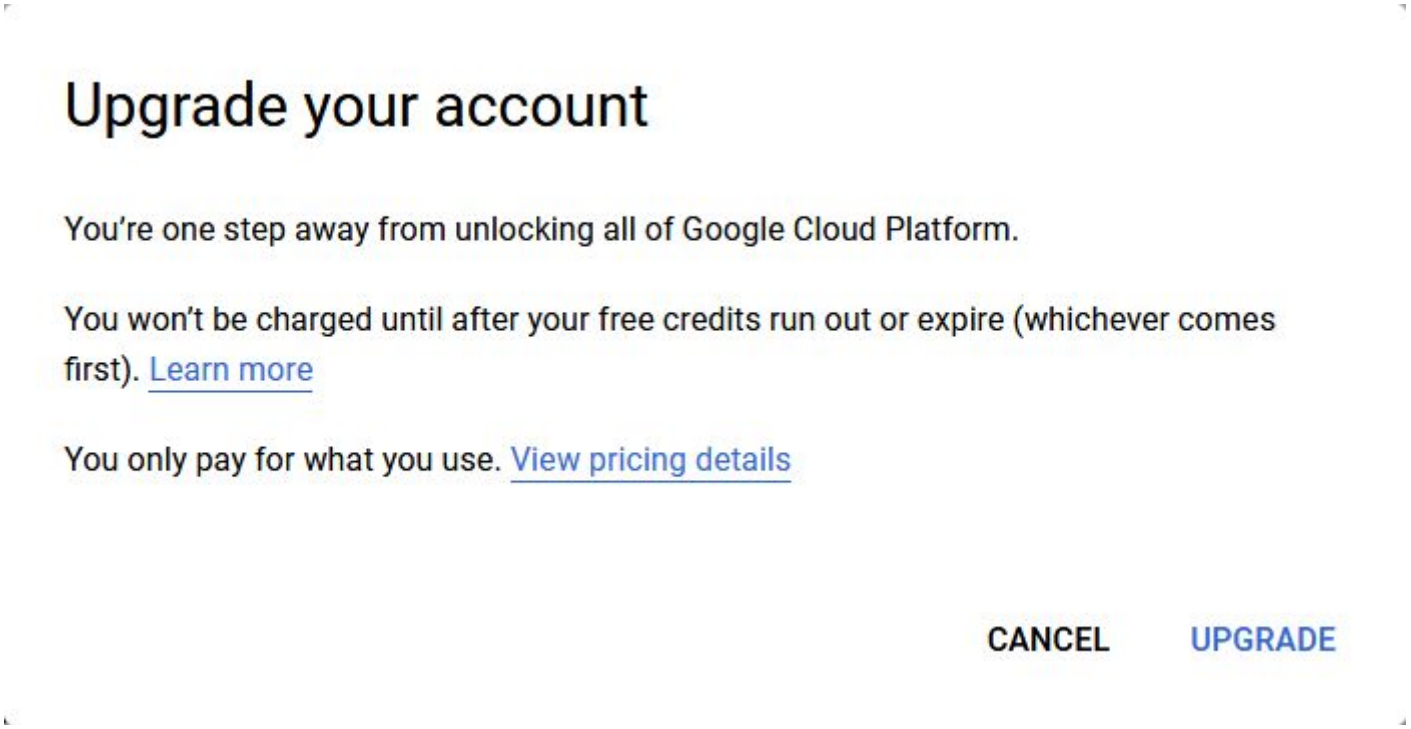


After launching the instance, an error states that I have no access to GPUs:



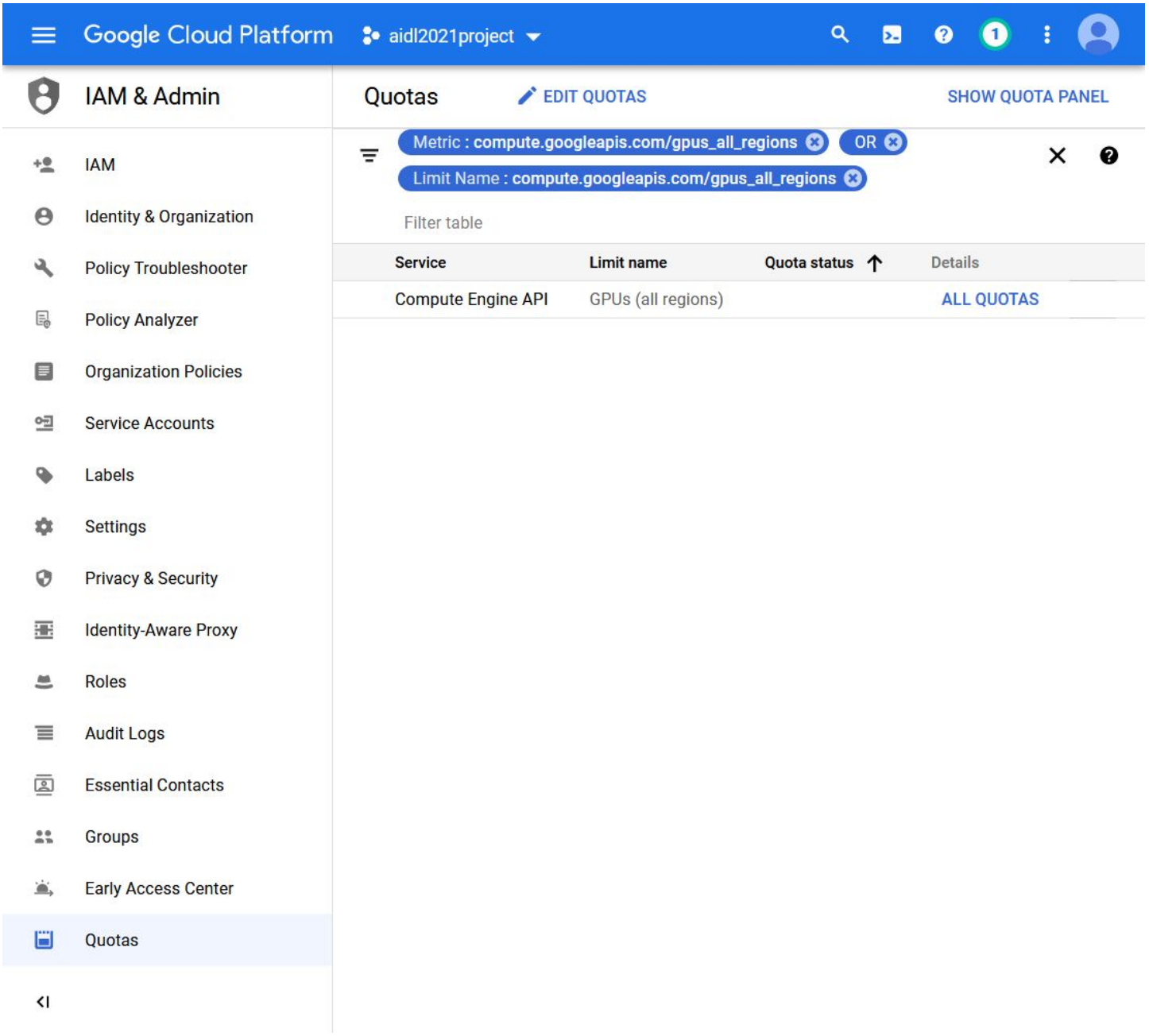
Upgrading the free account to premium

I tried to change the “GPUs (all regions)” quota, but according to https://stackoverflow.com/questions/63104914/cant-select-quota-to-edit-on-google-cloud-quota-edition-screen I must upgrade to a Premium account. So I push the gift icon (not shown in the capture) and press “Activate”.

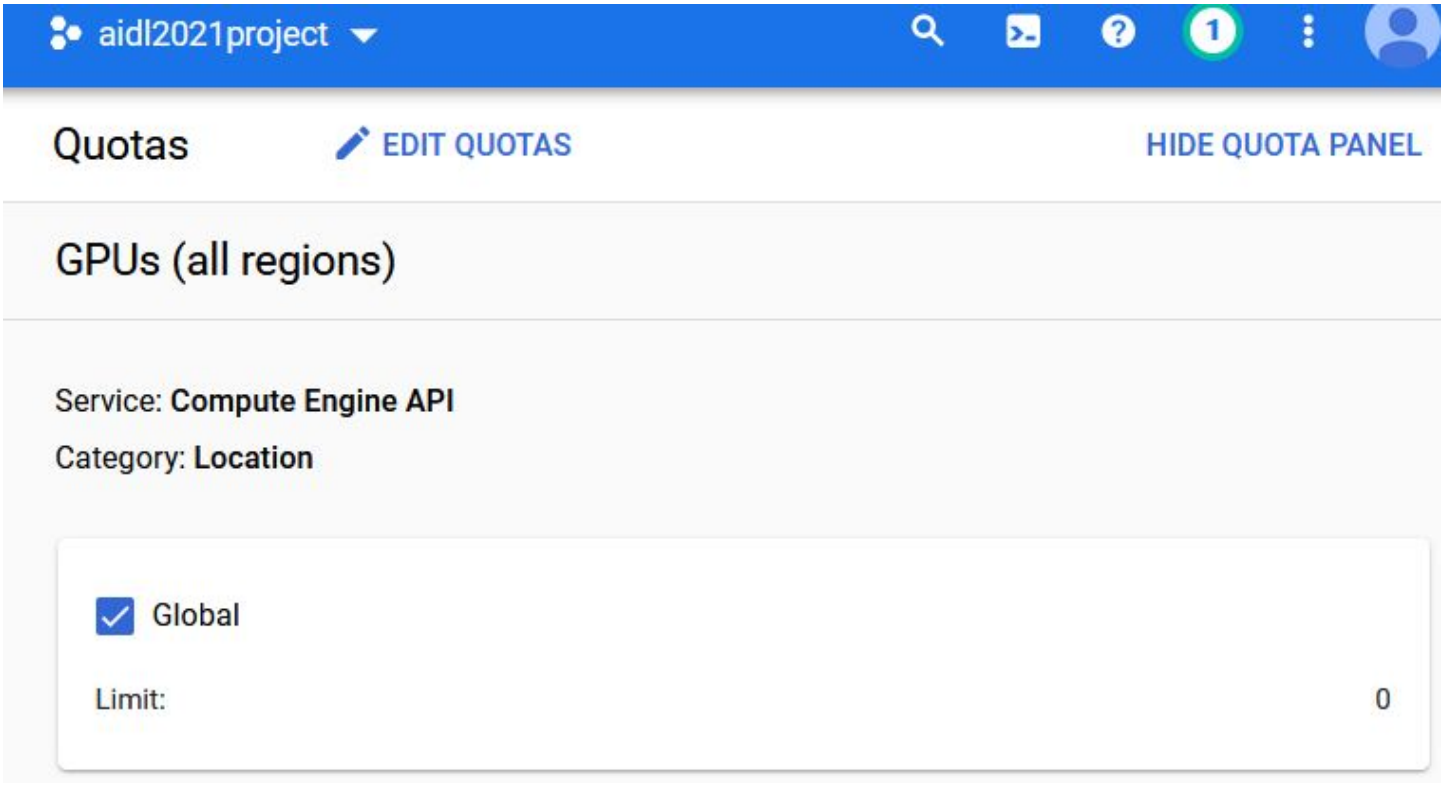


Requesting a quota upgrade for GPUs

After upgrading, I go to the quotas page link:
https://console.cloud.google.com/iam-admin/quotas?metric=compute.googleapis.com%2Fgpus_all_regions&project=aidl2021project&folder=&organizationId=



I click on the “GPUs (all regions)” text:



I select “Global” and click on “Edit quotas”:

× 1 quota selected

Quota changes

Expand each service card to change individual quotas.

Compute Engine API

Quota: GPUs (all regions)

Current limit: 0

Enter a new quota limit. Your request will be sent to your service provider for approval.

New limit *
1

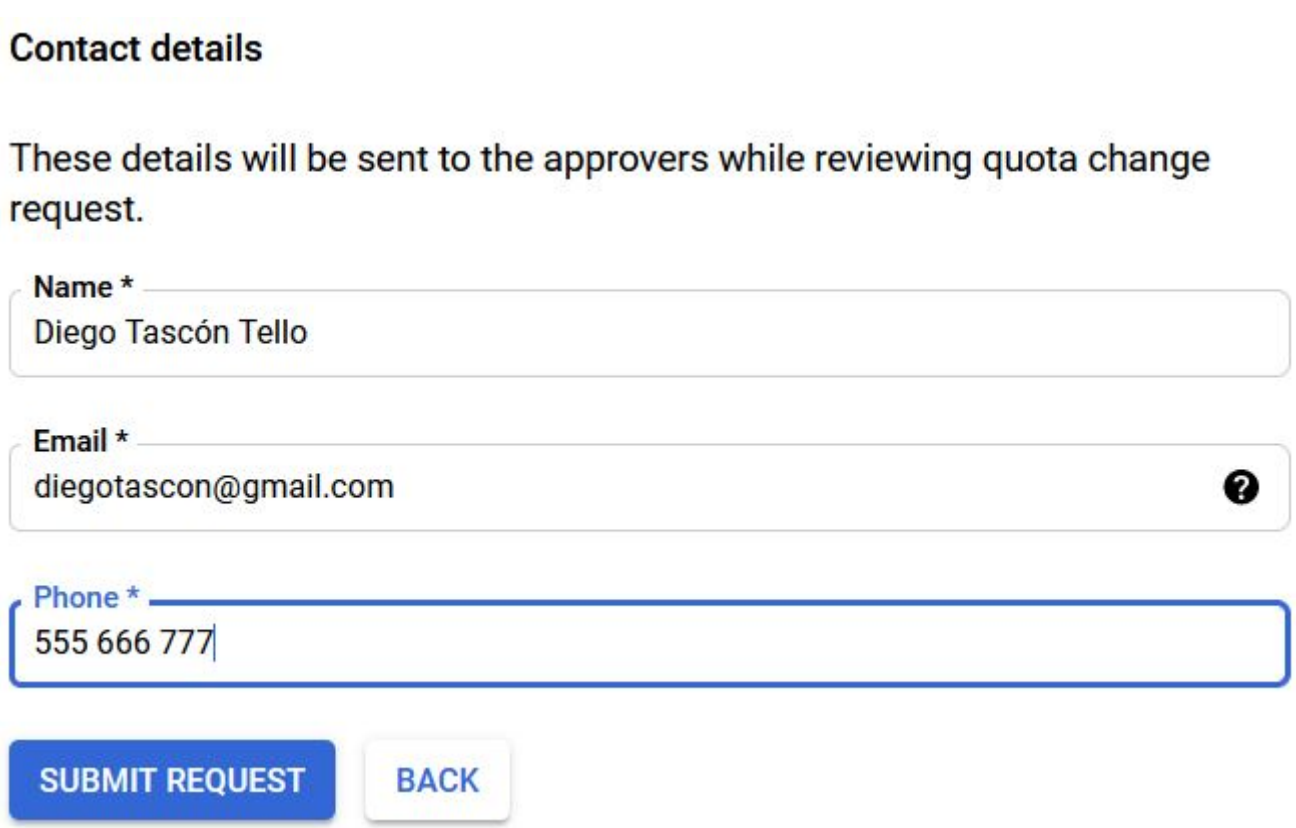
GPU increment request to be able to use an already granted 1 GPU quota on us-west1-b region. This will allow me to deploy my first Deep Learning VM instance and check Google Cloud services for DL training.
Thank you

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.

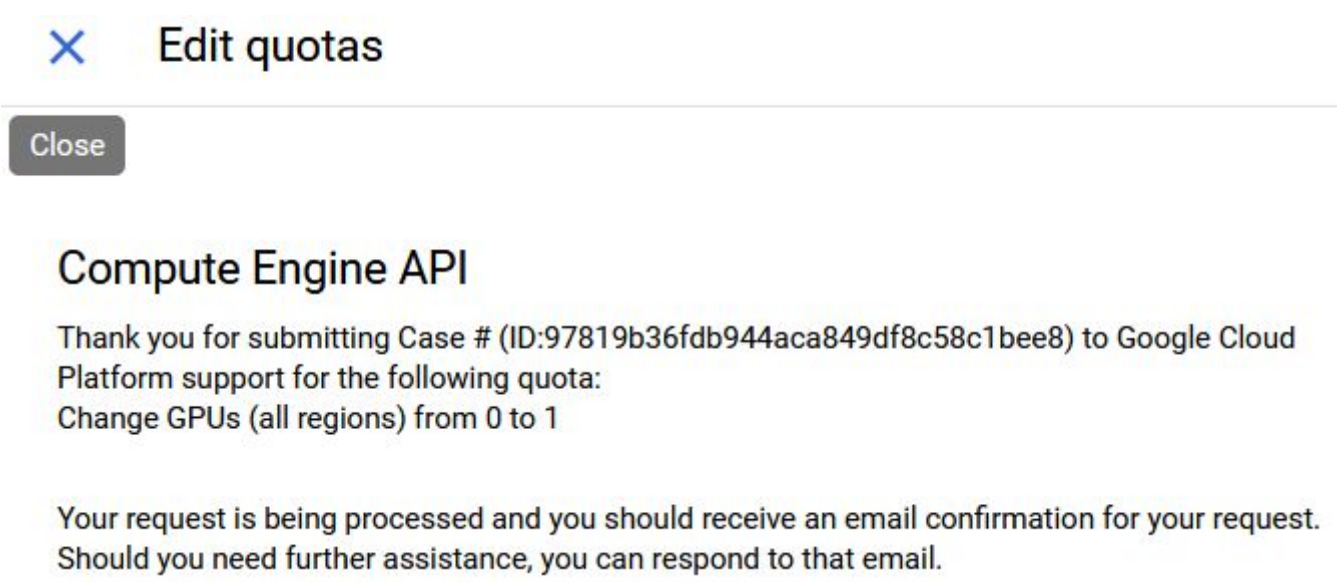
DONE

NEXT

I set the new limit to 1 and write a description of the request. After that I press “Done” and “Next”:



Contact details must be filled in and “Submit request” pressed:



Once the request is sent, an email has to be received. I press “Close” to end with the quota edition. The mail arrives in just one minute:

Google Compute Engine Quota Support <esupport@google.com>
para mí

Hello,

Your quota request for aidl2021project has been approved and your project quota
has been adjusted according to the following requested limits:

Name	Region	Requested Limit
GPUS_ALL_REGIONS	GLOBAL	1

After approved, Quotas can take up to 15 min to be fully visible in the Cloud
Console and available to you.

To verify, please navigate to https://console.cloud.google.com/iam-admin/quotas?project=aidl2021project.

If you find actual limits are greater than expected, this is normal if previous
requests were approved with higher limits. Otherwise, please let us know if approved changes are not reflected in your project.

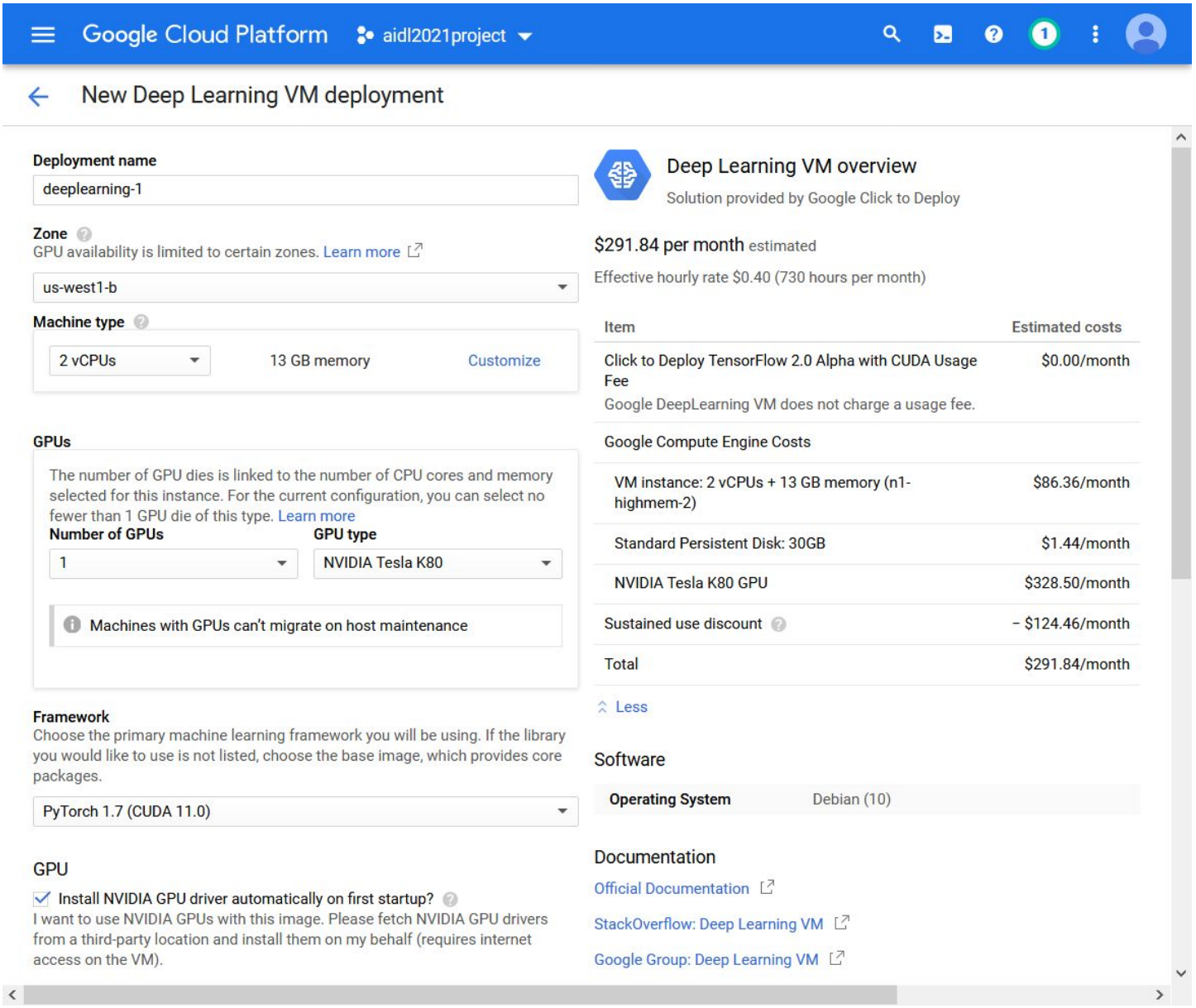
If you want to increase your quota further, please file a new request.

Best regards and happy computing!

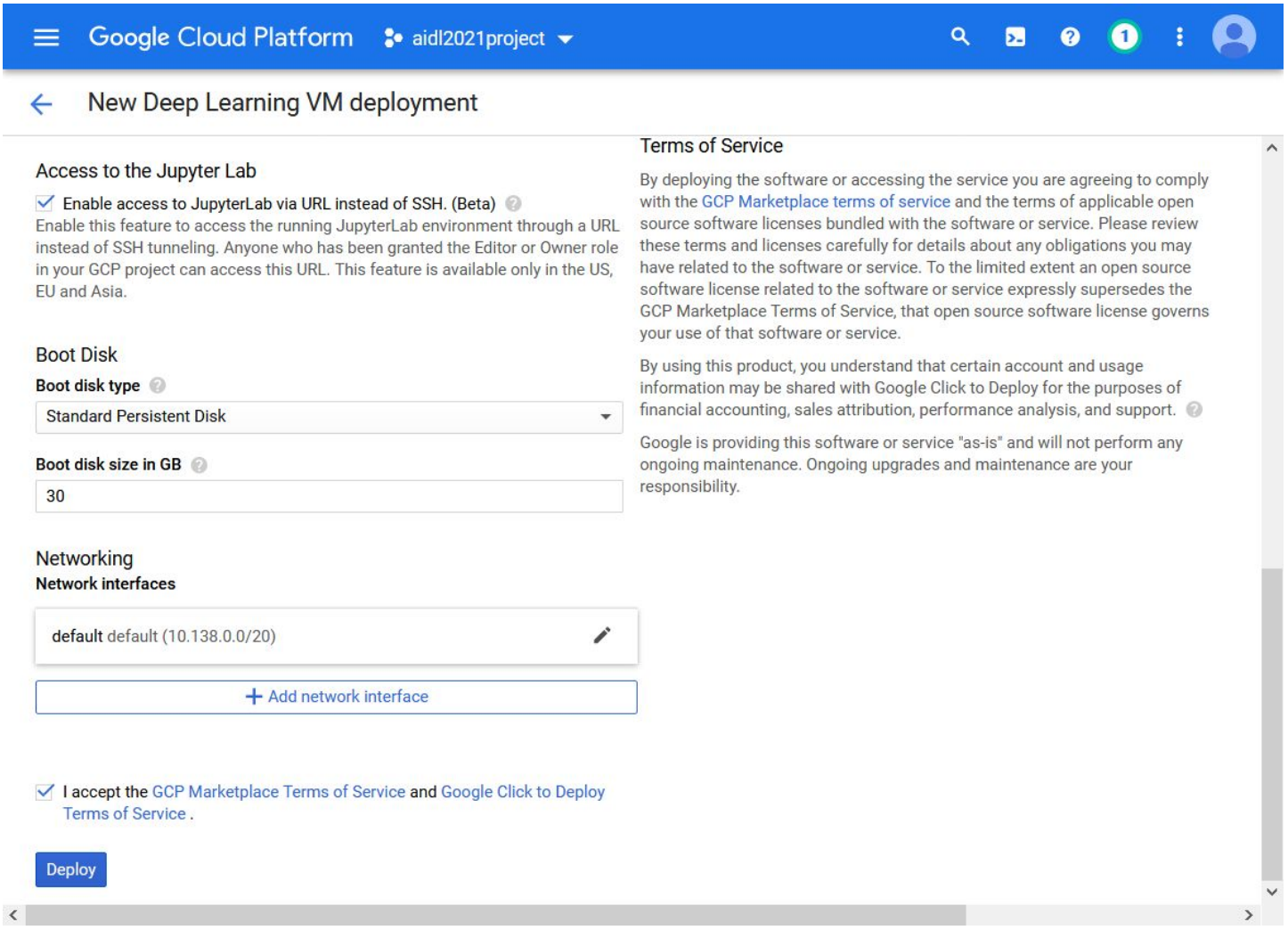
Sincerely,
Cloud Platform Support

Creating the instance

So I open again https://console.cloud.google.com/marketplace/details/click-to-deploy-images/deeplearning?_ga=2.13134391.1007356524.1612367677-1479030961.1612367677 and press “Launch”:

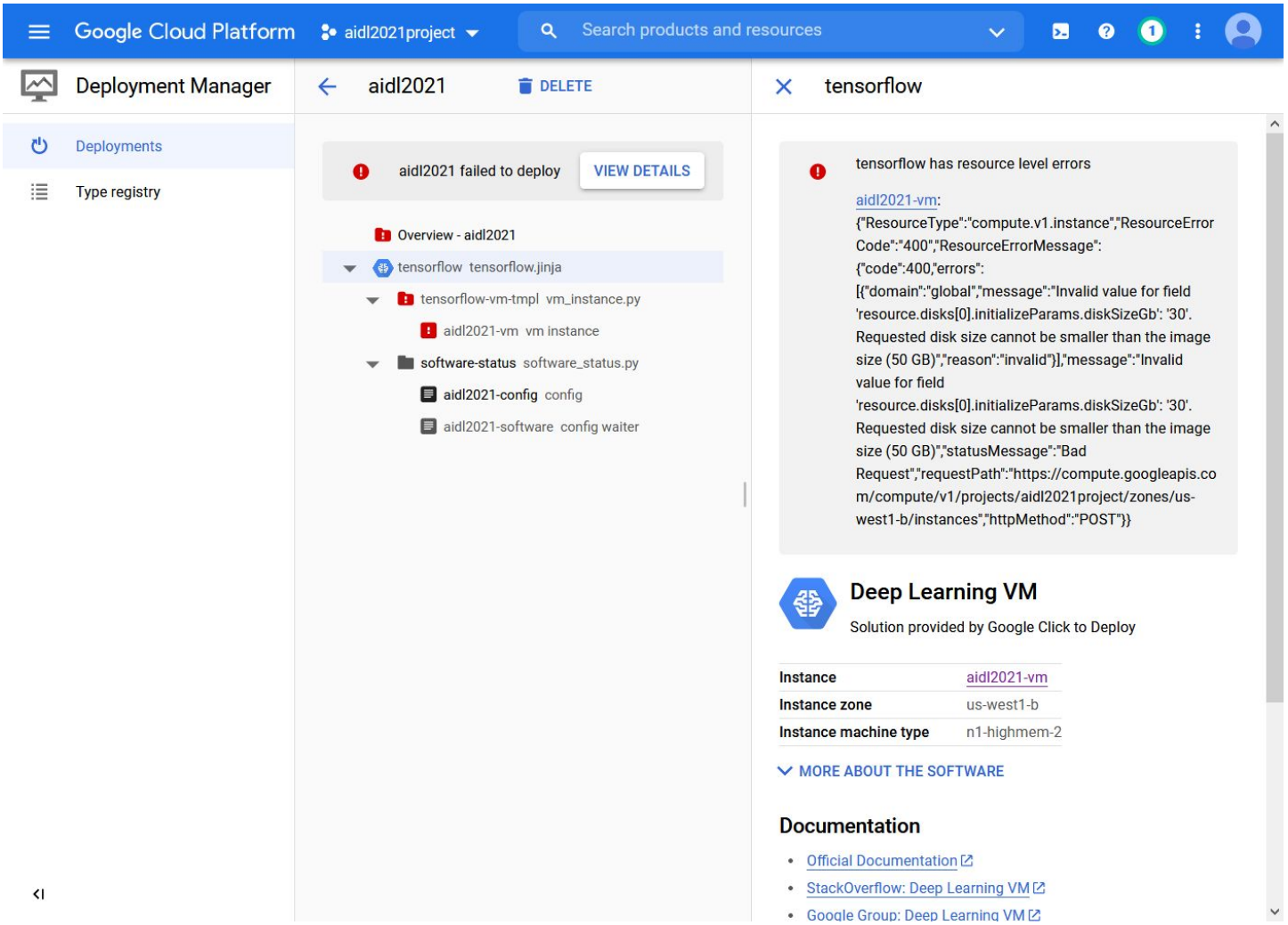


After changing the default deployment name (shown), choosing the Zone (us-west1-b seemed the cheapest), 2 CPUs, leaving the standard 13GB of RAM, setting 1 NVIDIA Tesla K80, selecting PyTorch 1.7 as framework, activating the “Install NVIDIA GPU driver automatically on first startup” ...



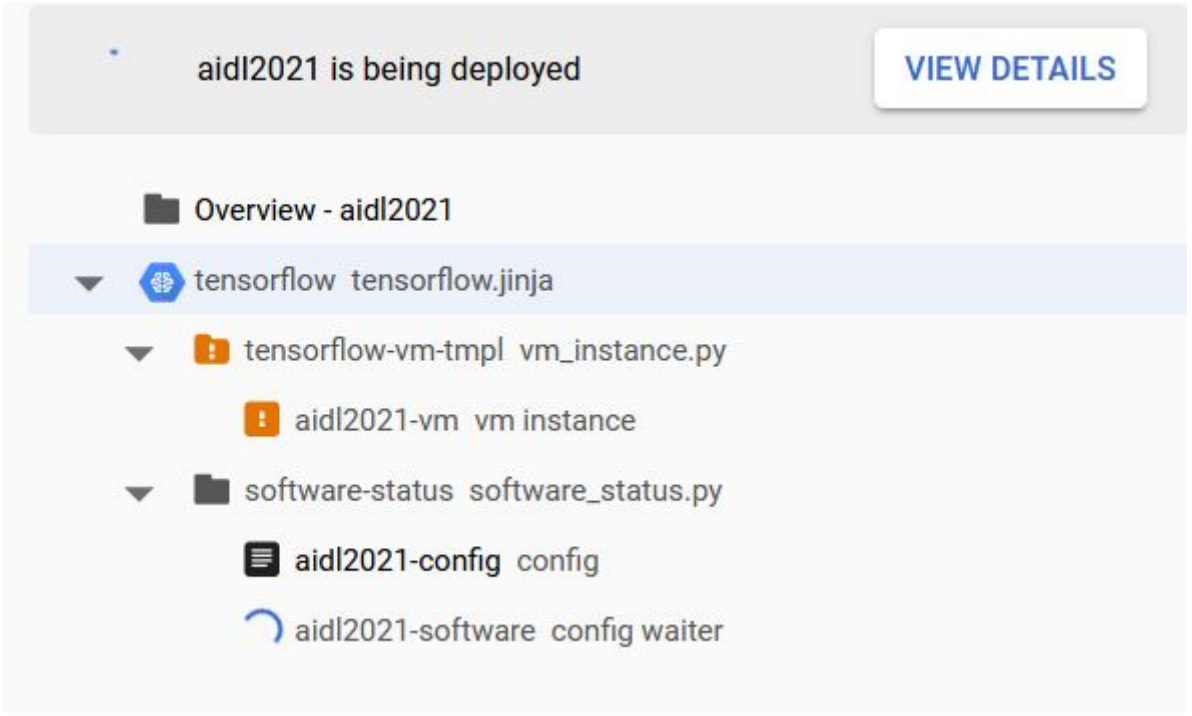
… enabling access to JupyterLab via URL, setting a disk size of 30GB (as stated in the slides), and accepting the terms of service, I press “Deploy”.

But an error is issued: *“Invalid value for field 'resource.disks[0].initializeParams.diskSizeGb': '30'. Requested disk size cannot be smaller than the image size (50 GB)"*

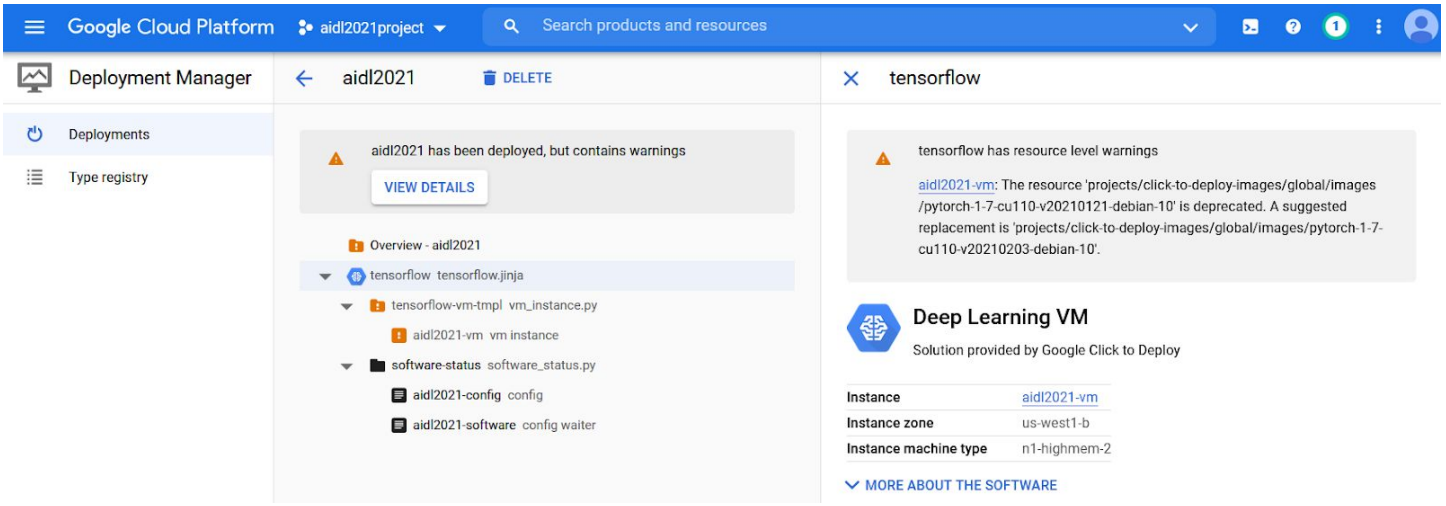


I press “Delete” to erase the resources and make a new try.

Thanks to Dani Fojo indications on Slack, I try again with 70GB (50GB for the installation and libraries, plus 20GB for the dataset).



The instance is deployed with a minor warning:



Why on earth does it let me choose a deprecated option?? Anyway, it is just 11 days old…

Get started with Deep Learning VM

SSH

Suggested next steps

- **Set up the Cloud SDK.**
  The Cloud SDK (gcloud) is the preferred command line tool for interfacing with your instance. Download it here.

- **(Optional) Copy files to your VM from your local machine.**
  You can use the gcloud tool to upload files to your machine.

```
$ gcloud compute scp --project aidl2021project --zone us-
```

- **Access the running Jupyter notebook.**
  We've already started a Jupyter notebook instance on the VM for your convenience. In order to get link that can be used to access Jupyter Lab run the following command.

```
$ gcloud compute instances describe --project aidl2021pro
```

- **Assign a static external IP address to your VM instance**
  An ephemeral external IP address has been assigned to the VM instance. If you require a static external IP address, you may promote the address to static. Learn more

Copy files to your VM from your local machine: `gcloud compute scp --project aidl2021project --zone us-west1-b --recurse <local file or directory> aidl2021-vm:~/`

Access the running Jupyter notebook: `gcloud compute instances describe --project aidl2021project --zone us-west1-b aidl2021-vm | grep googleusercontent.com | grep datalab`

Documentation

- Official Documentation
- StackOverflow: Deep Learning VM
- Google Group: Deep Learning VM

Support

If you have non-framework related issues, you can bring them up at the Deep Learning VM Stack Overflow.

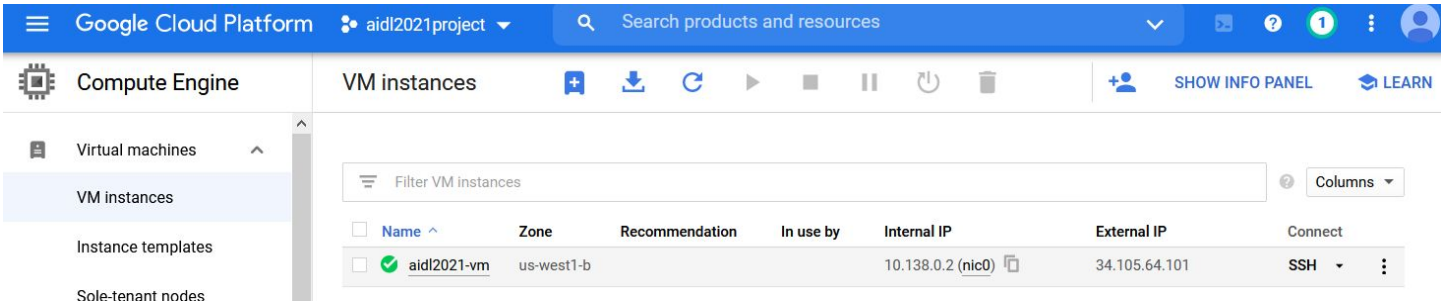
Template properties

type	tensorflow.jinja
acceleratorCount	`1.0`
acceleratorType	`nvidia-tesla-k80`
bootDiskSizeGb	`70.0`
bootDiskType	`pd-standard`
externalIP	`- EPHEMERAL`
input_enableProxy	`true`
input_framework	`PyTorch 1.7 (CUDA 11.0)`
input_installNvidiaDriver	`true`
machineType	`n1-highmem-2`
network	`- https://www.googleapis.com/compute/v1/projects/aidl2021project/global/networks/default`
subnetwork	`- https://www.googleapis.com/compute/v1/projects/aidl2021project/regions/us-west1/subnetworks/default`
zone	`us-west1-b`

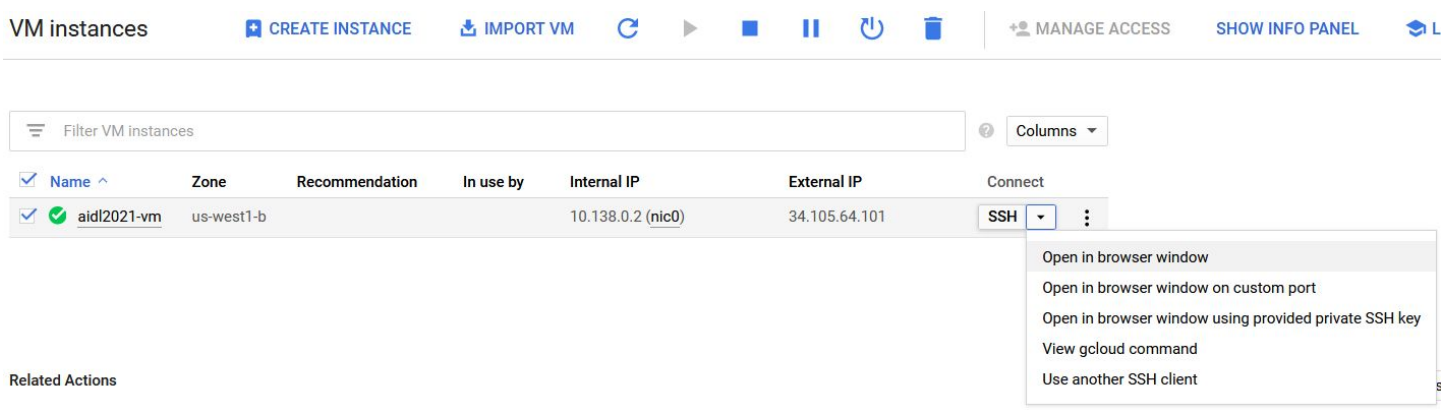
First access to the instance

https://cloud.google.com/ai-platform/deep-learning-vm/docs/quickstart-marketplace#access_your_new_instance

I go to the VM instances page in my Console: https://console.cloud.google.com/compute/instances?_ga=2.98630743.819462119.1612542875-1479030961.1612367677

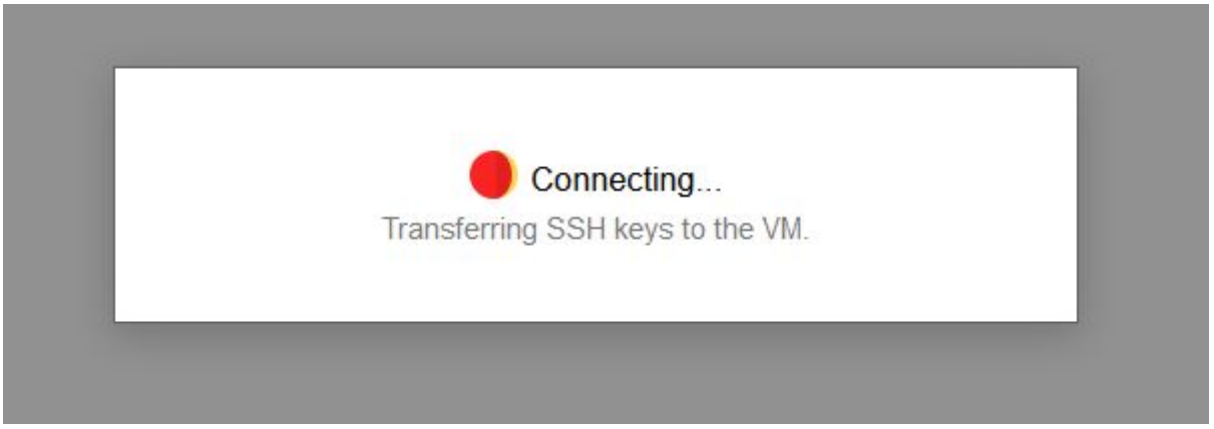


I click on the drop-down SSH list:

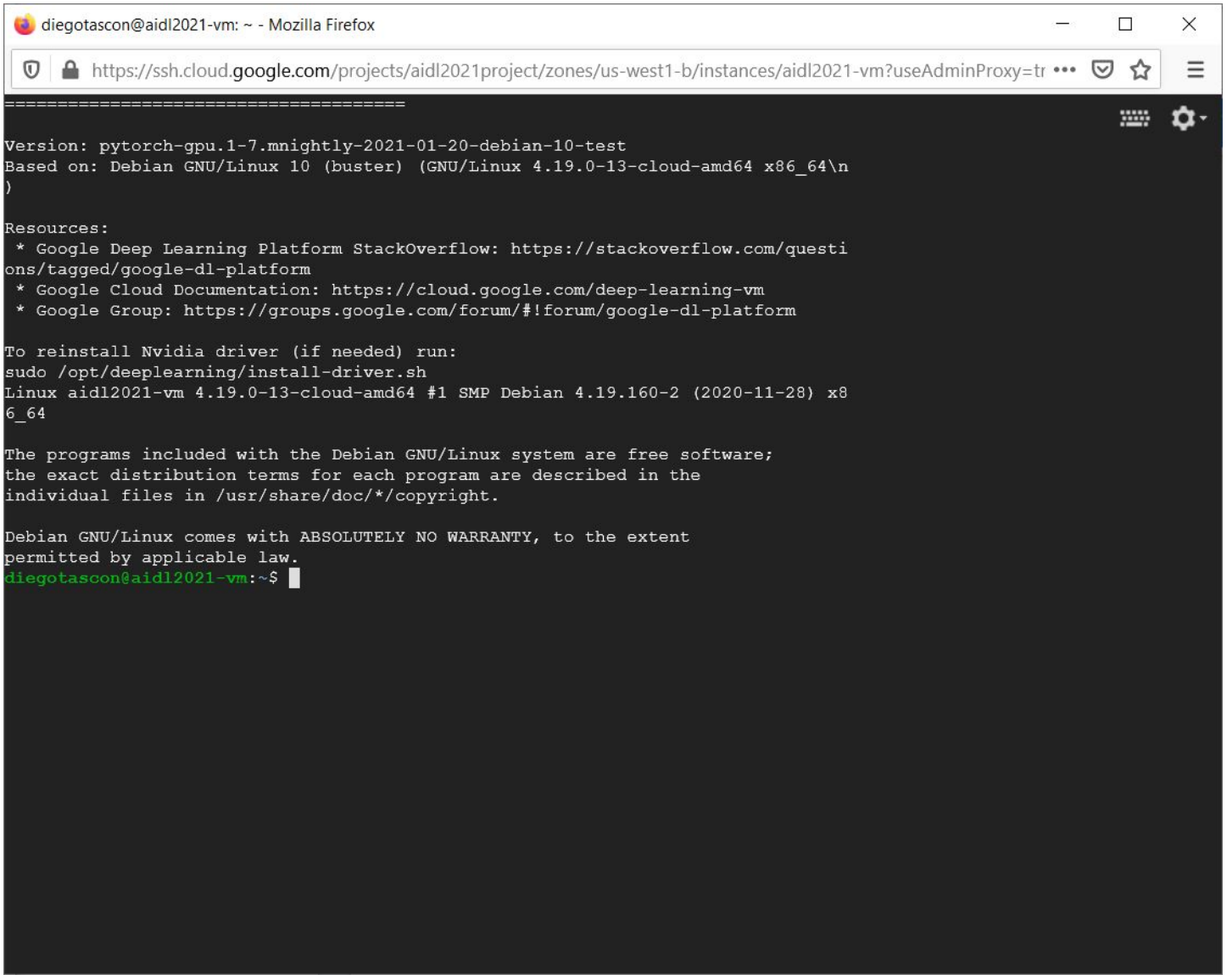


and I choose “Open in browser window”:

I setup my browser to allow Google Cloud console opening new tabs or windows:



And finally a shell window opens:



I get some info from the machine:

```
diegotascon@aidl2021-vm:~$ uname -a
Linux aidl2021-vm 4.19.0-13-cloud-amd64 #1 SMP Debian 4.19.160-2 (2020-11-28)
x86_64 GNU/Linux

diegotascon@aidl2021-vm:~$ vmstat -a
procs -----------memory---------- ---swap-- -----io---- -system--
------cpu-----
 r  b   swpd   free  inact active   si   so    bi    bo   in   cs us sy id wa
st
 0  0      0 12491360 300580 370832   0    0   322    6   76  103  1  1 96  2  0

diegotascon@aidl2021-vm:~$ nvidia-smi
Sat Feb  6 15:23:08 2021
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 450.51.06    Driver Version: 450.51.06  CUDA Version: 11.0     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
```

```
|===============================+======================+======================|
|   0  Tesla K80           Off  | 00000000:00:04.0 Off |                    0 |
| N/A   47C     P0    72W / 149W |    0MiB / 11441MiB |    100%      Default |
|                               |                      |                  N/A |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                                  |
|  GPU   GI   CI      PID   Type   Process name                  GPU Memory |
|    ID   ID                                                     Usage      |
|=============================================================================|
|  No running processes found                                                 |
+-----------------------------------------------------------------------------+

diegotascon@aidl2021-vm:~$ df -h
Filesystem      Size  Used Avail Use% Mounted on
udev            6.4G  0  6.4G   0% /dev
tmpfs           1.3G  8.5M  1.3G   1% /run
/dev/sda1       69G   33G   34G  50% /
tmpfs           6.4G  0  6.4G   0% /dev/shm
tmpfs           5.0M  0  5.0M   0% /run/lock
tmpfs           6.4G  0  6.4G   0% /sys/fs/cgroup
/dev/sda15      124M  7.9M  116M   7% /boot/efi
tmpfs           1.3G  0  1.3G   0% /run/user/1001

diegotascon@aidl2021-vm:~$ ls /home
diegotascon  jupyter

diegotascon@aidl2021-vm:~$ python --version
Python 3.7.8

diegotascon@aidl2021-vm:~$ python
Python 3.7.8 | packaged by conda-forge | (default, Nov 17 2020, 23:45:15)
[GCC 7.5.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import torch
>>> torch.cuda.is_available()
True
```

In the console, if I click on the aidl2021-vm, I get the details:

VM instance details EDIT RESET LEARN

aidl2021-vm

Details Monitoring Screenshot

Remote access

SSH

Connect to serial console

Enable connecting to serial ports

Logs

Cloud Logging

Serial port 1 (console)

More

Instance Id

4301326556626734139

Machine type

n1-highmem-2 (2 vCPUs, 13 GB memory)

Reservation

Automatically choose (default)

CPU platform

Intel Broadwell

Display device

Turn on a display device if you want to use screen capturing and recording tools.

Turn on display device

GPUs

1 x NVIDIA Tesla K80

Zone

us-west1-b

Labels

goog-dm : aidl2021

Creation time

05.02.2021, 17:31:18

Network interfaces

Name	Network	Subnetwork	Primary internal IP	Alias IP ranges	External IP	Network Tier	IP forwarding	Network details
nic0	default	default	10.138.0.2	—	34.105.64.101 (ephemeral)	Premium	Off	**View details**

Public DNS PTR Record
None

Firewalls
Allow HTTP traffic
Allow HTTPS traffic

Network tags
aidl2021-deployment, deeplearning-vm

Deletion protection
Enable deletion protection
When deletion protection is enabled, instance cannot be deleted. Learn more

Confidential VM service
Disabled

Boot disk

Name	Image	Size (GB)	Device name	Type	Encryption	Mode	When deleting instance
aidl2021-vm	pytorch-1-7-cu110-v20210121-debian-10	70	tensorflow-vm-tmpl-boot-disk	Standard persistent disk	Google managed	Boot, read/write	Delete disk

Additional disks
None

Local disks
None

Preserved state size
0 GB

Shielded VM
To edit Shielded VM features you need to stop the instance first.
Turn on all settings for the most secure configuration.
Turn on Secure Boot
Turn on vTPM
Turn on Integrity Monitoring

Availability policies

Preemptibility	Off (recommended)
On host maintenance	Terminate VM instance
Automatic restart	On (recommended)

Custom metadata

framework	PyTorch:1.7
google-logging-enable	0
google-monitoring-enable	0
install-nvidia-driver	True
proxy-mode	project_editors
proxy-url	1a45d4f37210686b-dot-us-west1.notebooks.googleusercontent.com
shutdown-script	/opt/deeplearning/bin/shutdown_script.sh
status-config-url	https://runtimeconfig.googleapis.com/v1beta1/projects/aidl2021project/configs/aidl2021-config
status-uptime-deadline	600
status-variable-path	status
title	PyTorch/CUDA11.0.GPU
version	nightly-2021-01-20-debian-10-test

SSH Keys

☐ Block project-wide SSH keys

None

Service account

602576452010-compute@developer.gserviceaccount.com

Cloud API access scopes

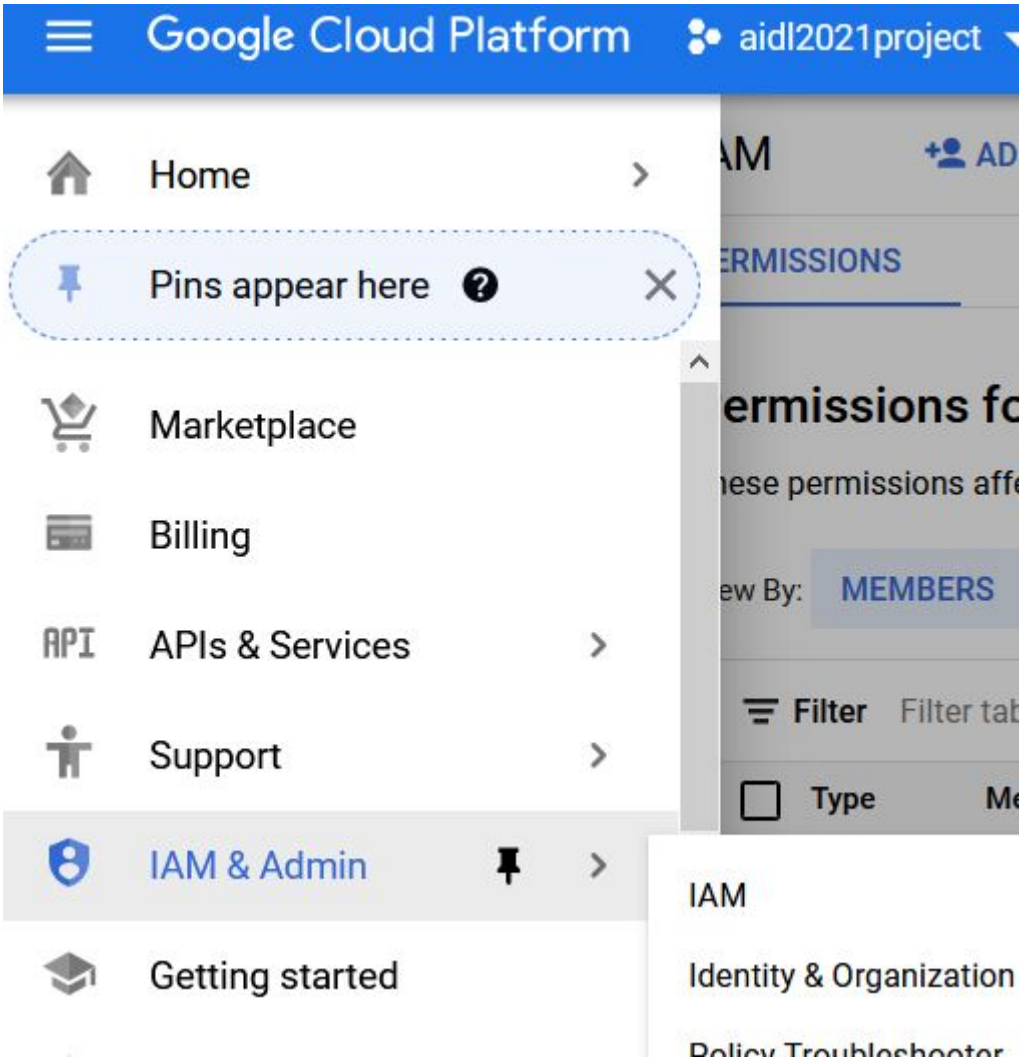
BigQuery	Disabled
Bigtable Admin	Disabled
Bigtable Data	Disabled
Cloud Datastore	Disabled
Cloud Debugger	Disabled
Cloud Pub/Sub	Disabled
Cloud Source Repositories	Disabled
Cloud SQL	Disabled
Compute Engine	Read Write
Service Control	Disabled
Service Management	Disabled
Stackdriver Logging API	Write Only
Stackdriver Monitoring API	Write Only
Stackdriver Trace	Disabled
Storage	Read Only
Task queue	Disabled
User info	Disabled

Equivalent REST

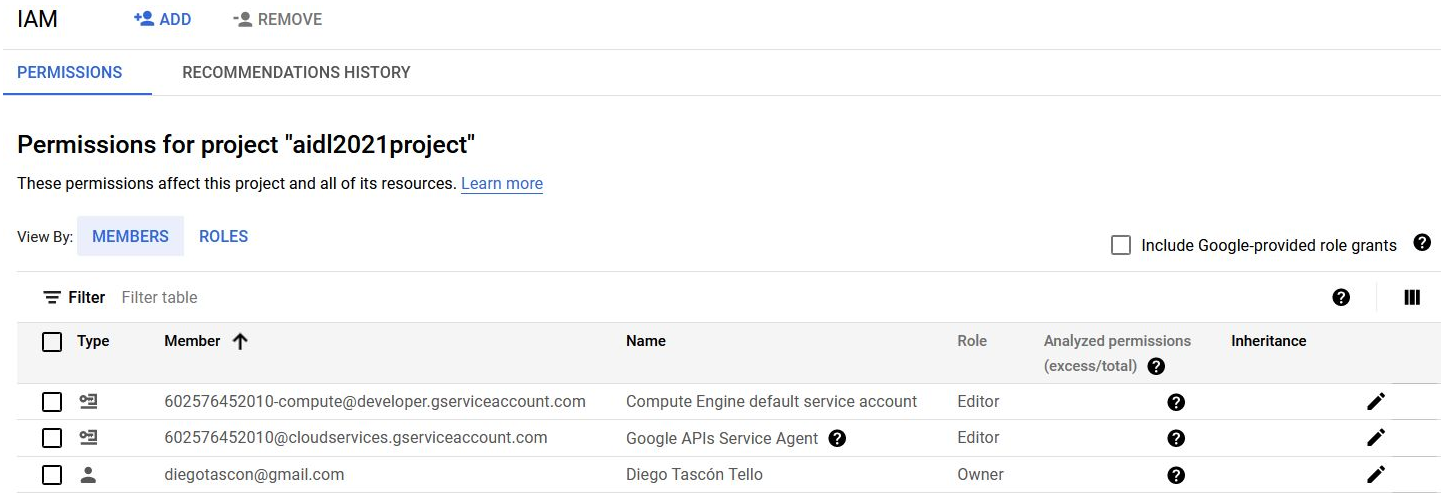
Granting permissions to colleagues over the project

Steps vaguely inspired in the following documentation:
https://cloud.google.com/compute/docs/access/managing-access-to-resources#bind-member

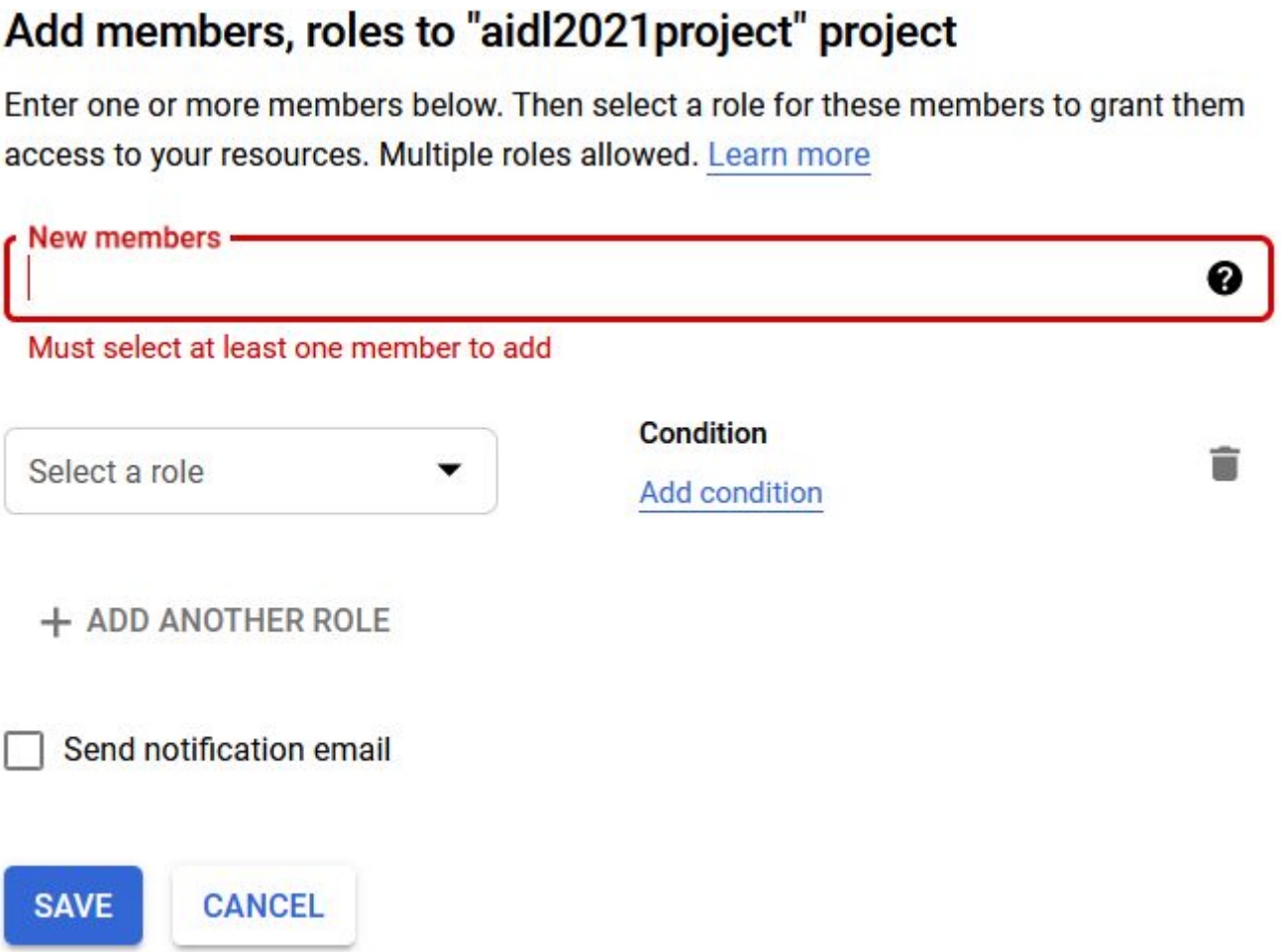
In the menu, I clicked directly the “IAM & Admin” option, not any of its suboptions:



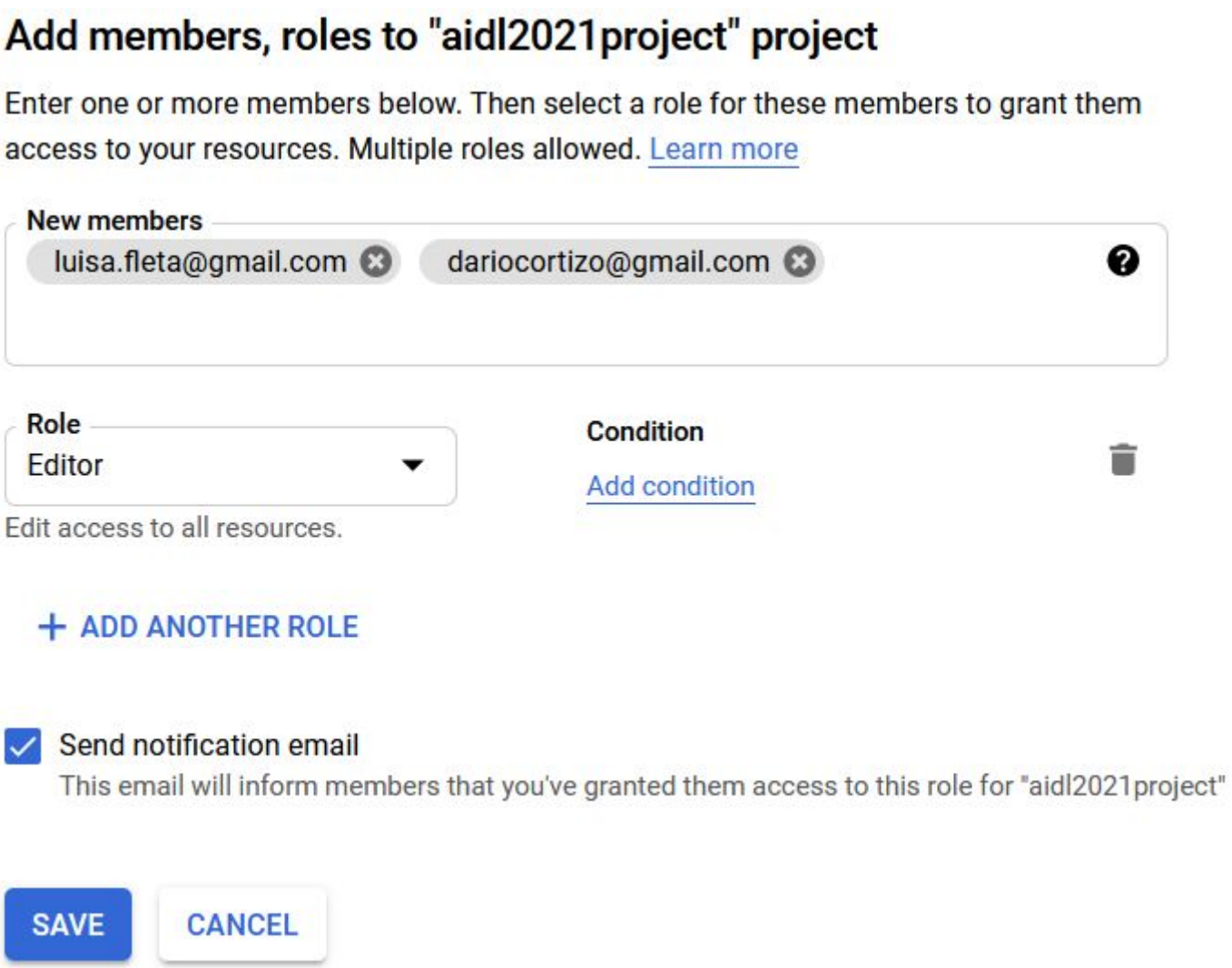
And the list of users of the project appears:



I pressed the Add button above:



I write the gmail addresses of my colleagues and grant them an editor role:



Pressing “Save”:

Permissions for project "aidl2021project"

These permissions affect this project and all of its resources. Learn more

View By: MEMBERS ROLES

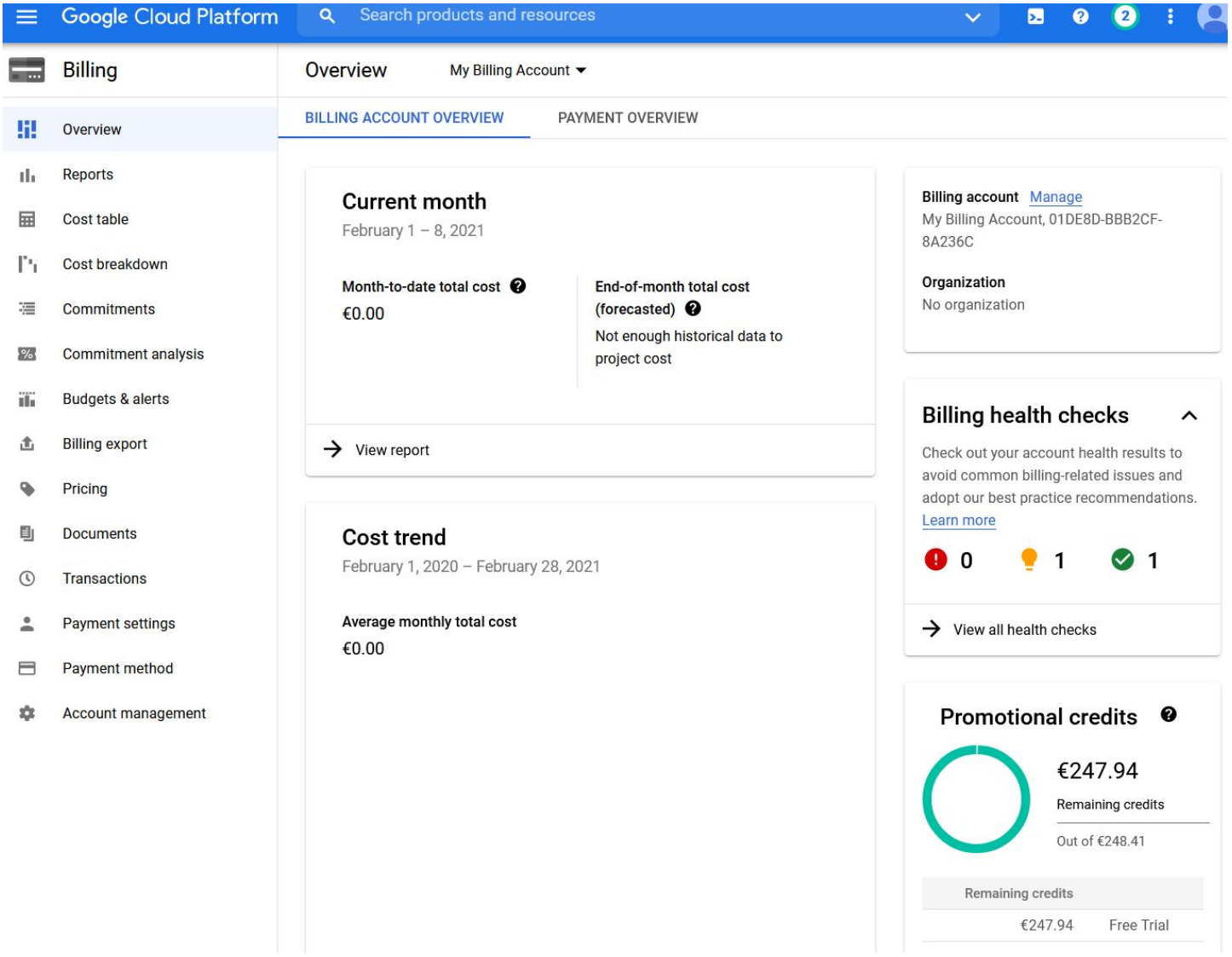
Filter Filter table

Type	Member	Name	Role
	602576452010-compute@developer.gserviceaccount.com	Compute Engine default service account	Editor
	602576452010@cloudservices.gserviceaccount.com	Google APIs Service Agent	Editor
	dariocortizo@gmail.com		Editor
	diegotascon@gmail.com	Diego Tascón Tello	Owner
	luisa.fleta@gmail.com		Editor

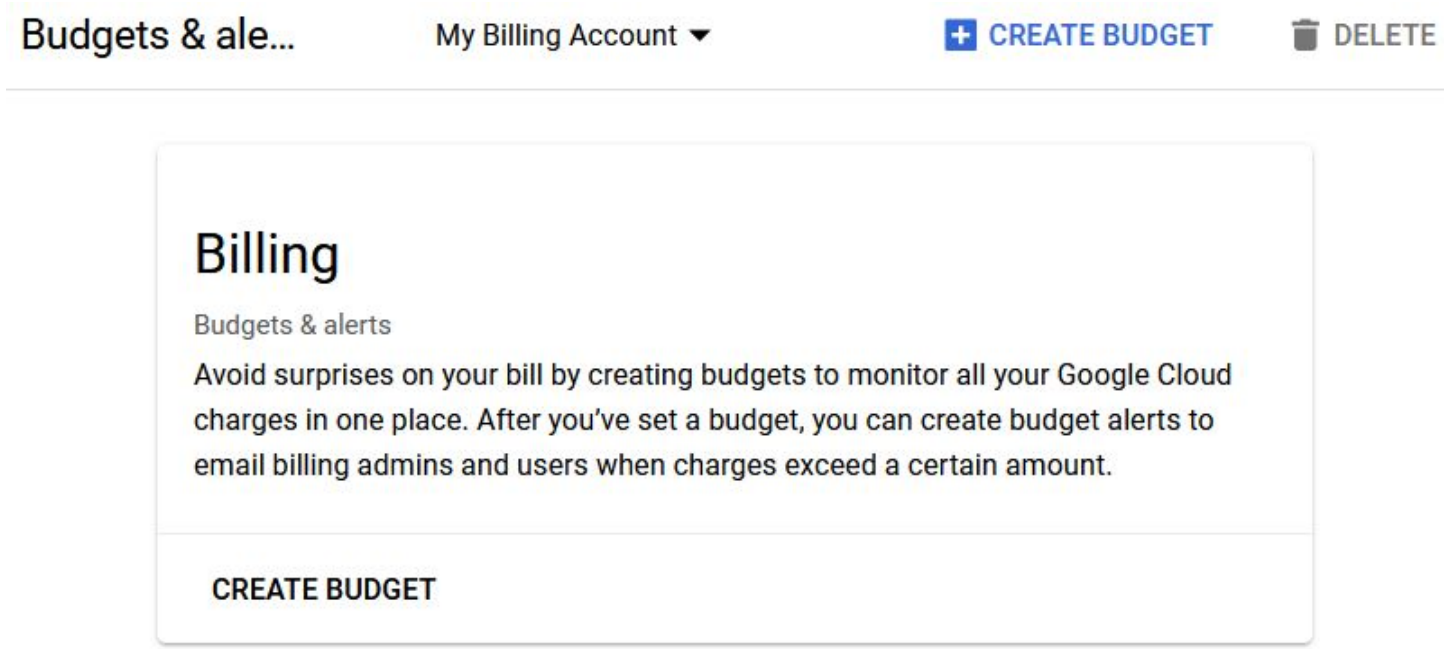
Setting a budget alarm

https://cloud.google.com/billing/docs/how-to/budgets

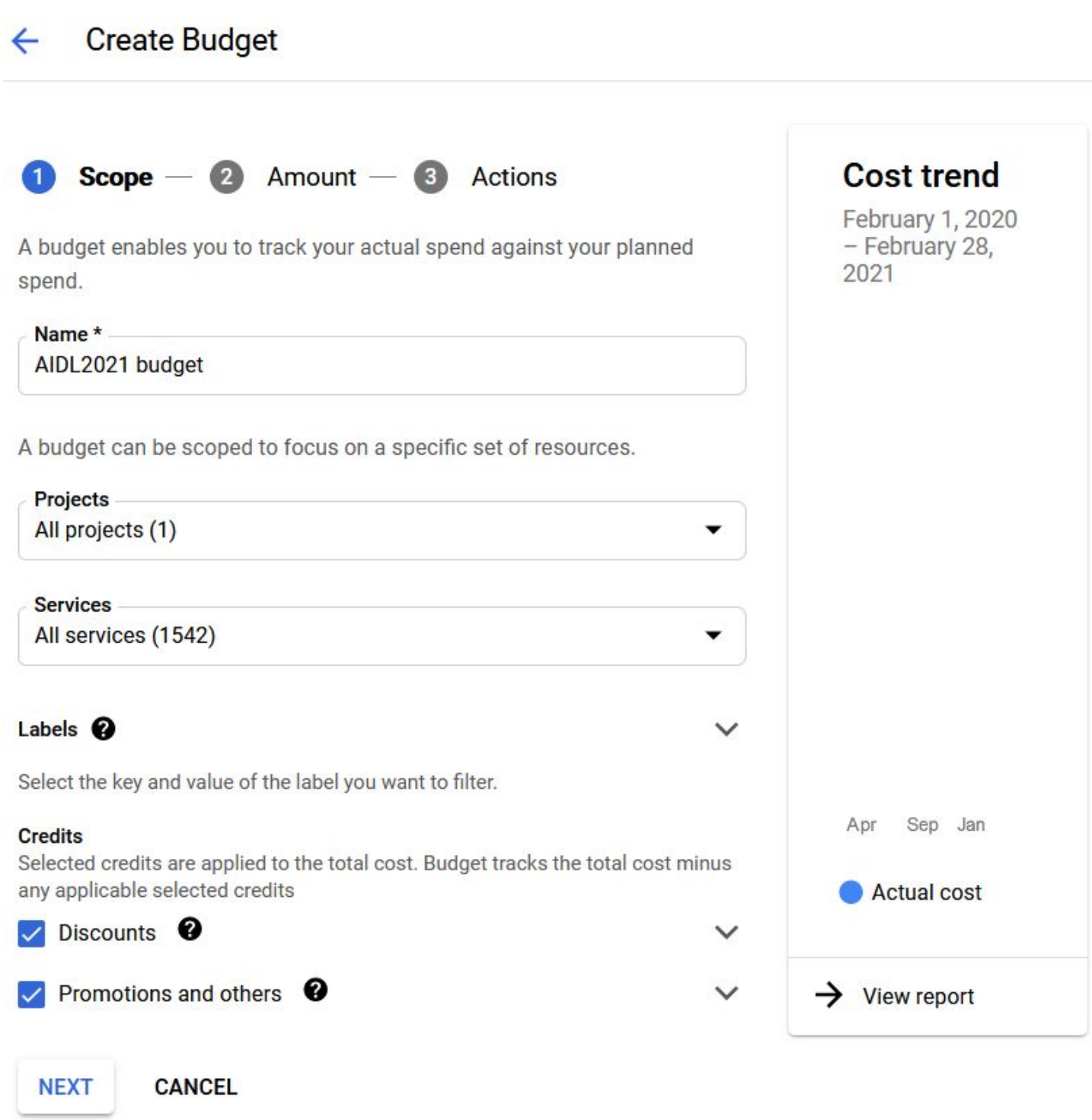
In the main menu I select Billing and the billing console appears:



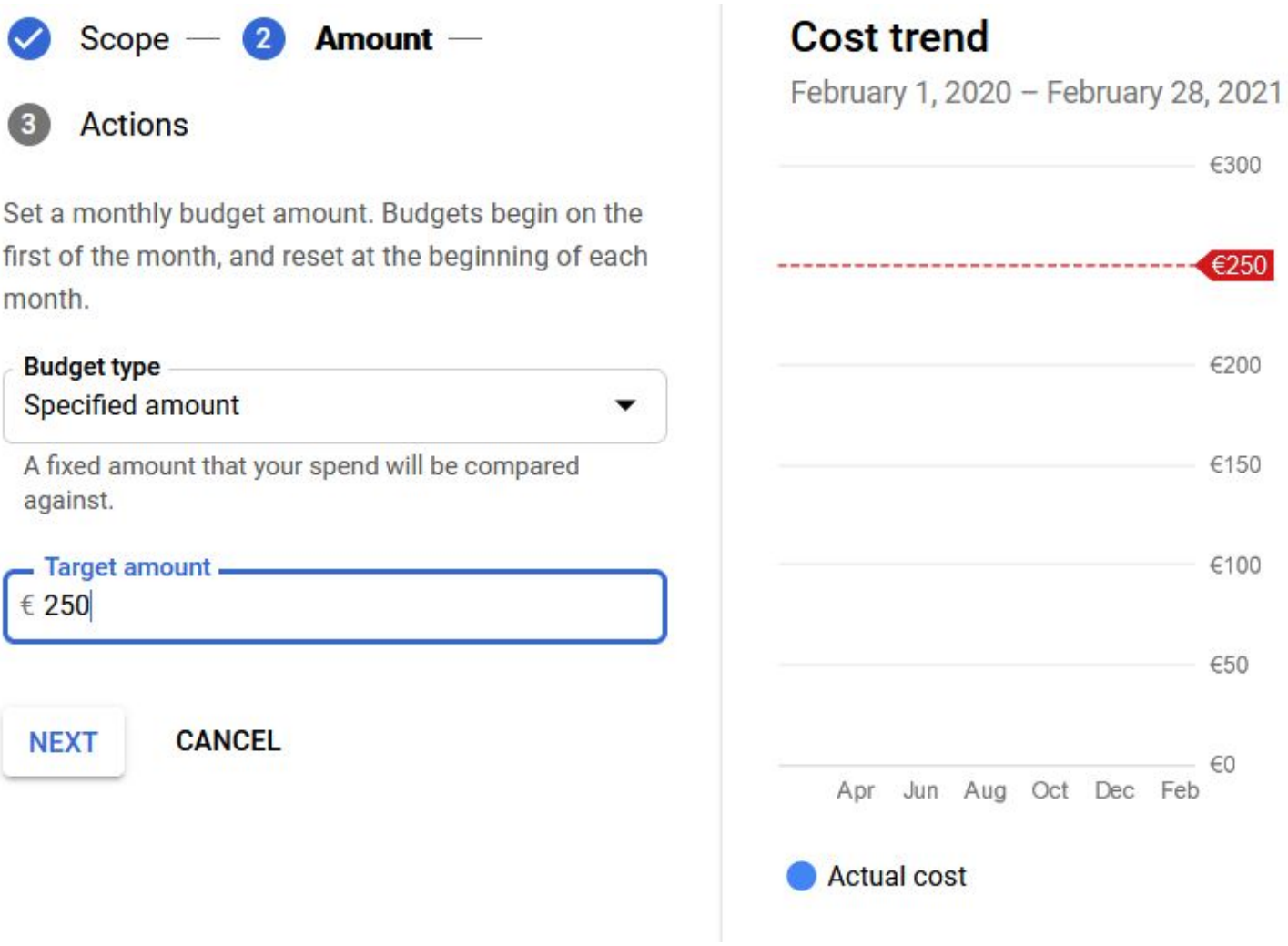
In the left menu, I select “Budgets & alerts”:



I press “Create budget”:



After writing down the name, I press “Next” (it doesn’t allow me to select just one project, so I left the scopes as they came by default):



I set that it triggers an alarm when 250€ or more are spent in a single month. I press Next:

Scope — Amount — 3 **Actions**

Set alert threshold rules

Send email alert notifications after the actual or forecasted spend exceeds a percent of the budget or a specified amount. Learn more.

Percent of budget	Amount	Trigger on
50 %	€ 125	Actual
90 %	€ 225	Actual
100 %	€ 250	Actual

+ ADD THRESHOLD

Manage notifications

Send email alert notifications to billing admins and users of this billing account.

☑ Email alerts to billing admins and users

Allow Monitoring email notification channels to receive alerts when this budget reaches thresholds.

☐ Link Monitoring email notification channels to this budget
Select a Monitoring workspace and maximum 5 Monitoring email notification channels.

Use Pub/Sub notifications to programmatically receive spend updates about this budget.

☐ Connect a Pub/Sub topic to this budget
Select a project and Pub/Sub topic. Anyone who can view this budget will also be able to view the project ID and the topic name. It may not be possible to add a Pub/Sub topic if it belongs to an organization that has domain restricted sharing enabled.

FINISH CANCEL

Cost trend

February 1, 2020 – February 28, 2021

View report

The already set thresholds are ok to me, so I just press Finish:

Budgets & alerts

My Billing Account ▾ + CREATE BUDGET 🗑 DELETE

Budgets track expenses within a Google Cloud Platform project or billing account. Your budget can be a specified amount or based on previous spend. You can set alerts to notify billing admins and users when a budget goes over a specified amount.

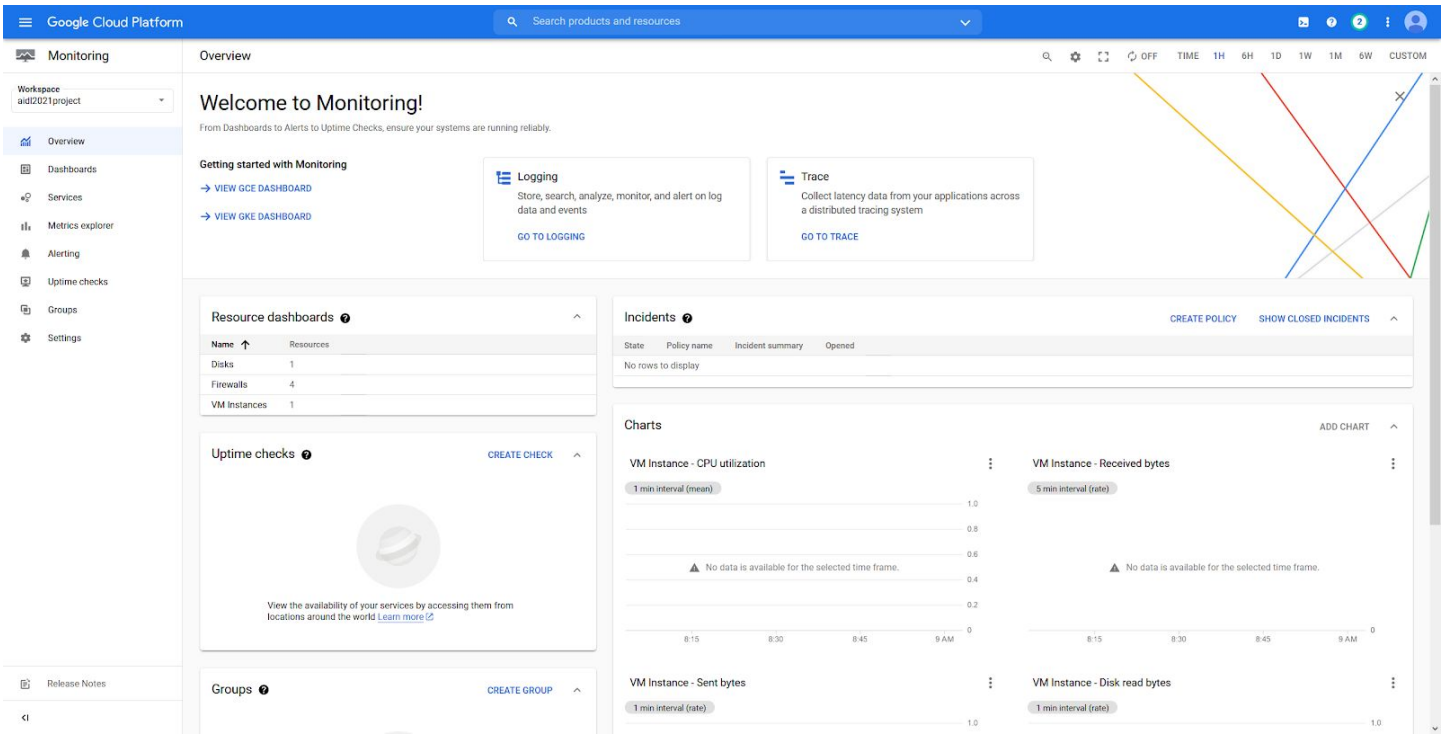
ⓘ Setting a budget does not cap resource or API consumption. Learn more.

Budget name	Budget type	Applies to	Trigger alerts at	Spend and budget amount
AIDL2021 budget	Specified amount	This billing acco ▾	50%, 90%, and 100%	€0.00 / €250.00 Includes -€0.47 credit

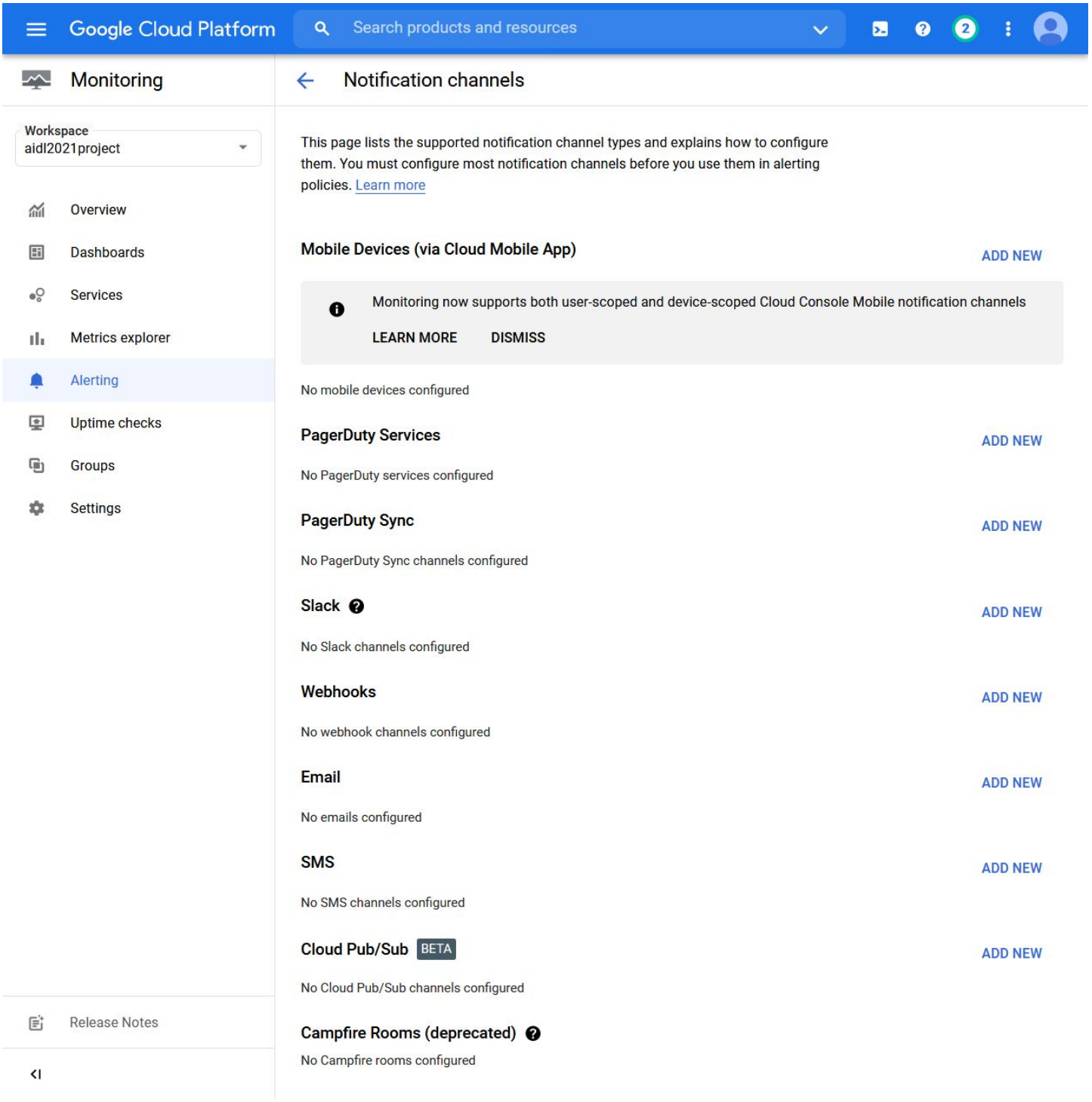
Adding other team members to the receiving alerts

To let other team members receiving the budget alerts, a Monitoring Workspace must be created: https://cloud.google.com/monitoring/workspaces/manage?_ga=2.205456150.-78066755.1601881399#create-quickly

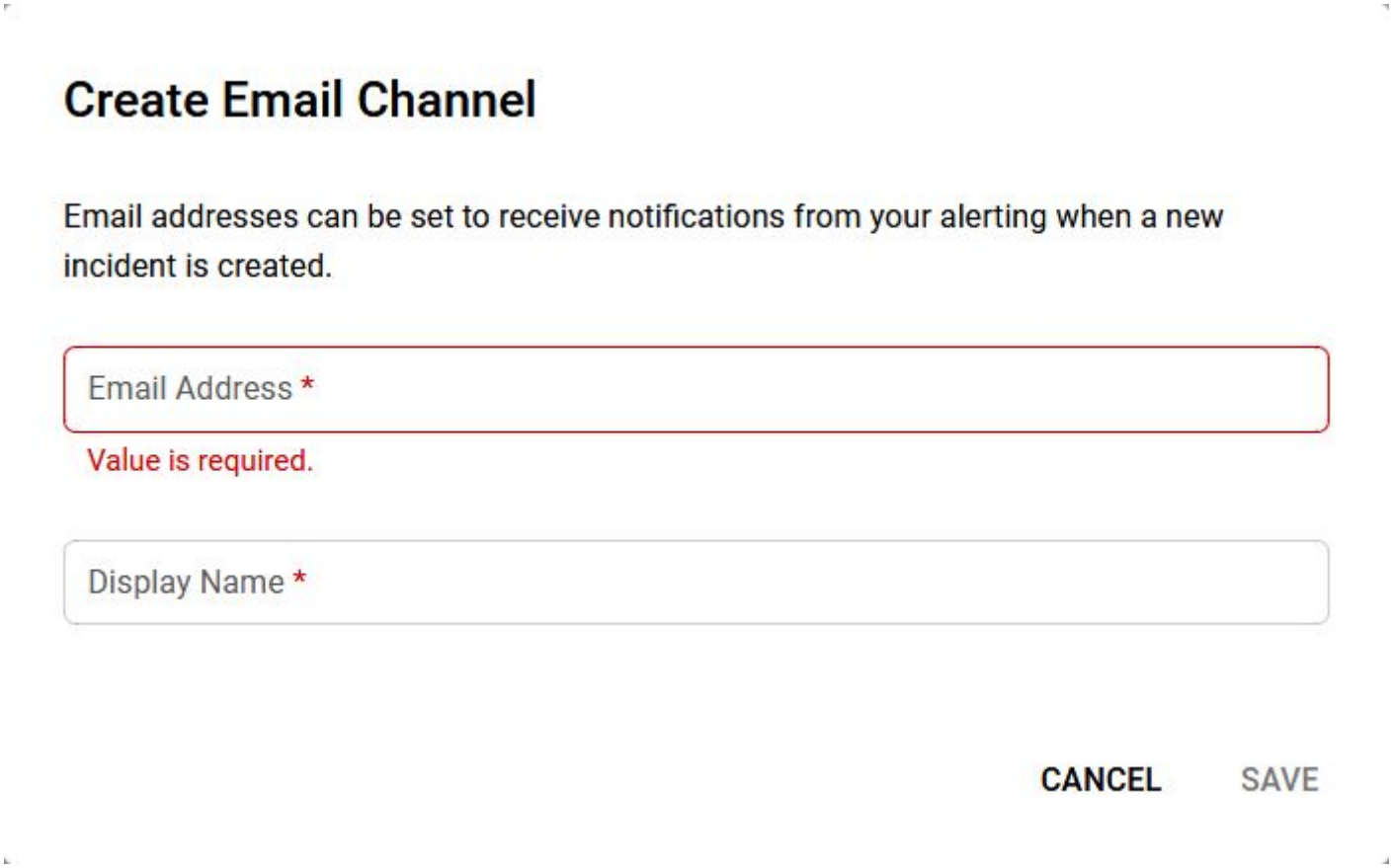
I go to “Monitoring” in the main menu:



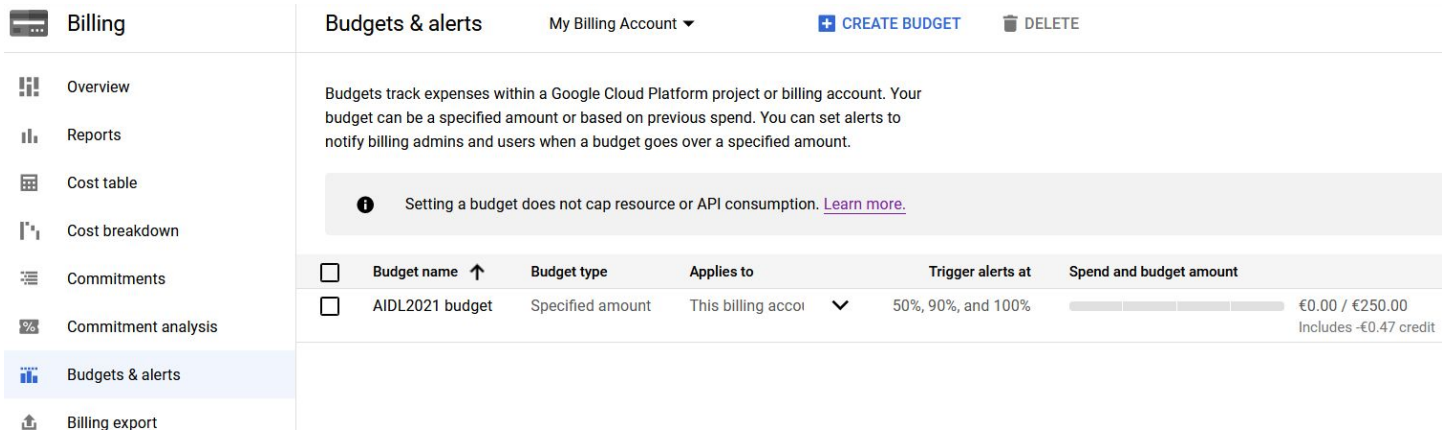
As stated in https://cloud.google.com/monitoring/workspaces/create#single-project-workspace, a Workspace for the project is automatically created (see upper left corner). I press “Alerting” on the left menu to set email notifications:



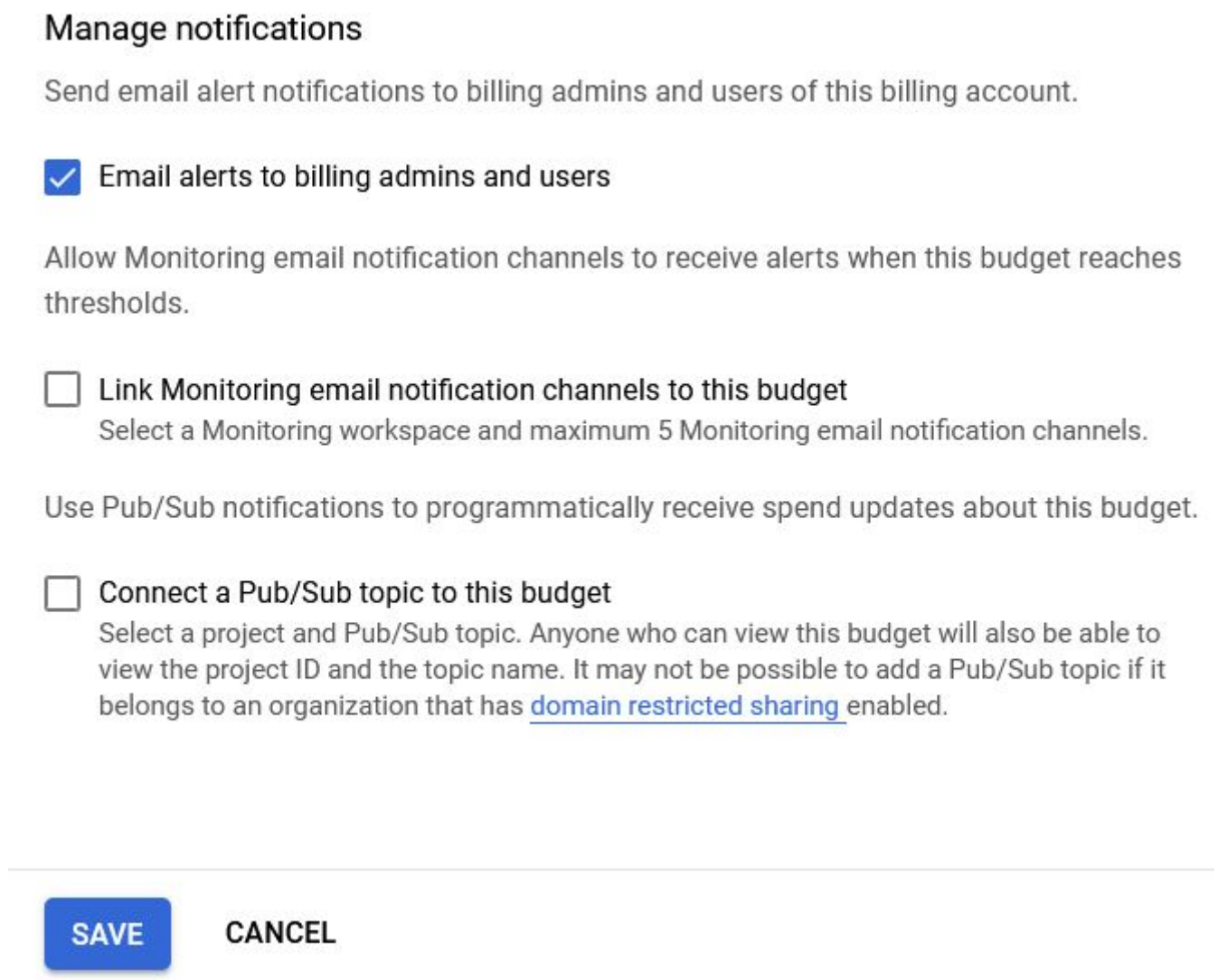
I click on the “Add new” link for email:



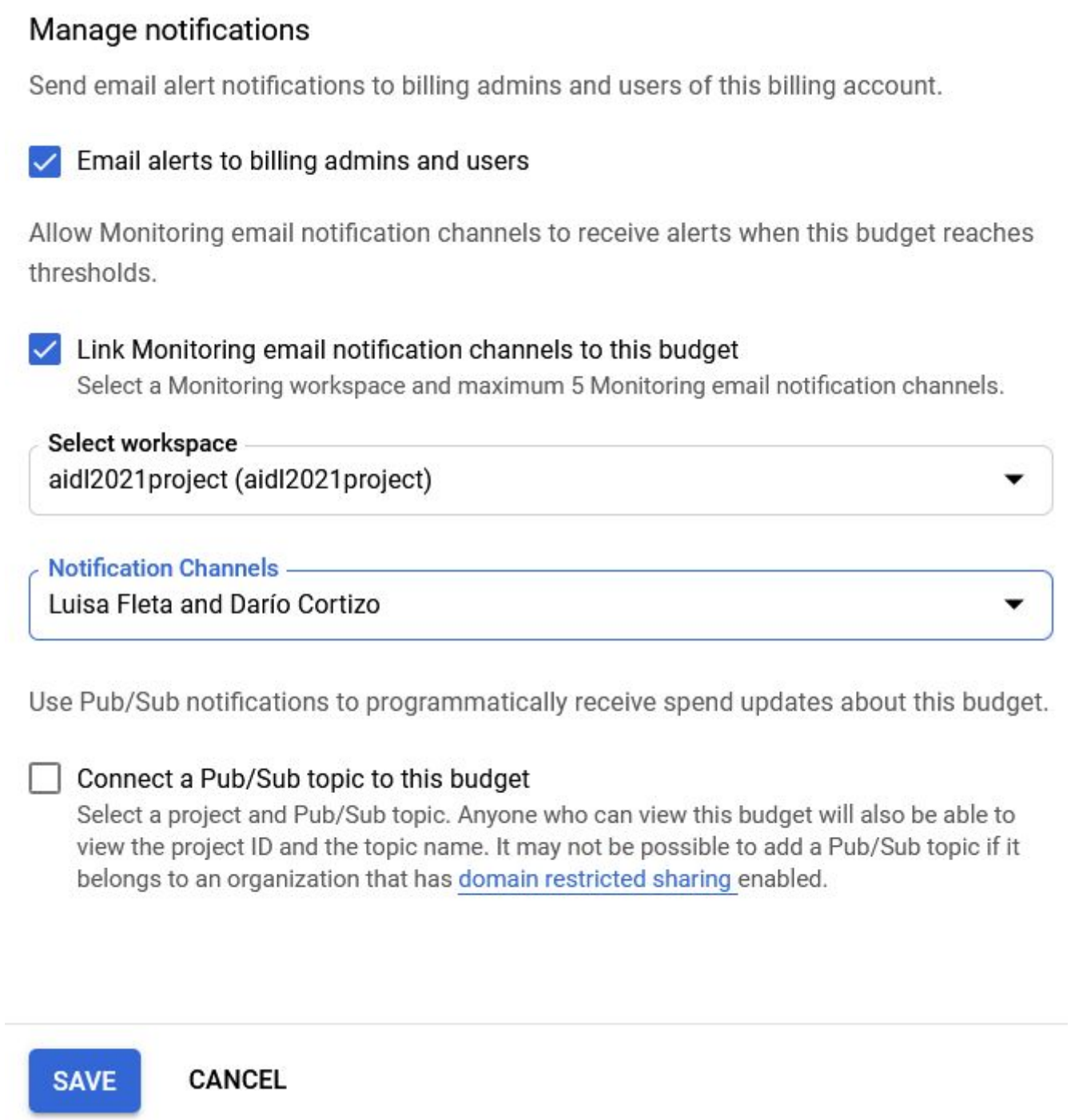
I add the mails of my team mates. Then I com back to the “Budgets & Alerts” menu option under Billing:



I click on the “AIDL2021 budget” to edit the alert:



In the bottom “Manage notifications” section I check “Link Monitoring email notification channels to this budget. Two additional fields appear. I select the aidl2021project as workspace and my team mates as notification channels:



I press “Save” to finish the setup.

Using the instance

Copying files

The gcloud command line tool can be used to copy files. My first try with gcloud compute scp failed:

```
diego@EstudioVM:~/10-Project/ldd-pix2pix$ gcloud compute scp *.py
aidl2021-vm:/home/diegotascon
WARNING: The private SSH key file for gcloud does not exist.
WARNING: The public SSH key file for gcloud does not exist.
WARNING: You do not have an SSH key for gcloud.
WARNING: SSH keygen will be executed to generate a key.
Generating public/private rsa key pair.
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/diego/.ssh/google_compute_engine
Your public key has been saved in /home/diego/.ssh/google_compute_engine.pub
The key fingerprint is:
SHA256:HmEUJ+FjWaT43Cf+UDFxKlyty+qLybeFFfAOiOAL3tc diego@EstudioVM
The key's randomart image is:
+---[RSA 3072]----+
|     .   =+= o..  |
|   . . = O + +.   |
|  . . o X + *.    |
| . o . * + +.+    |
|  . o . E o.=.    |
|     . . o *o     |
|     . +..|       |
|     . oo+        |
|     +o+o.|       |
+----[SHA256]-----+
Updating project ssh metadata...⠧Updated
[https://www.googleapis.com/compute/v1/projects/aidl2021project].
Updating project ssh metadata...done.
Waiting for SSH key to propagate.
Warning: Permanently added 'compute.4301326556626734139' (ECDSA) to the list of
known hosts.
scp: /home/diegotascon/arguments.py: Permission denied
scp: /home/diegotascon/dataset.py: Permission denied
scp: /home/diegotascon/libs.py: Permission denied
scp: /home/diegotascon/network.py: Permission denied
scp: /home/diegotascon/test.py: Permission denied
scp: /home/diegotascon/train.py: Permission denied
scp: /home/diegotascon/transform-dataset.py: Permission denied
scp: /home/diegotascon/transform-dataset-v1.0.py: Permission denied
scp: /home/diegotascon/transform.py: Permission denied
scp: /home/diegotascon/utils.py: Permission denied
ERROR: (gcloud.compute.scp) [/usr/bin/scp] exited with return code [1].
```

Thanks to https://stackoverflow.com/questions/27807018/gcloud-compute-copy-files-permission-denied-when-copying-files I find a solution:

```
diego@EstudioVM:~/10-Project/ldd-pix2pix$ gcloud compute scp *.py
diegotascon@aidl2021-vm:/home/diegotascon
Updating project ssh metadata...⠧Updated
[https://www.googleapis.com/compute/v1/projects/aidl2021project].
Updating project ssh metadata...done.
```

```
Waiting for SSH key to propagate.
arguments.py                                  100% 5819  33.9KB/s   00:00
dataset.py                                    100% 4946  28.8KB/s   00:00
libs.py                                       100%  815  4.8KB/s   00:00
network.py                                    100%   13KB  76.6KB/s   00:00
test.py                                       100% 0     0.0KB/s   00:00
train.py                                      100% 8774  51.2KB/s   00:00
transform-dataset.py                          100% 8339  49.1KB/s   00:00
transform-dataset-v1.0.py                     100% 6229  36.2KB/s   00:00
transform.py                                  100% 0     0.0KB/s   00:00
utils.py                                      100% 8880  51.8KB/s   00:00
```

Accessing remotely to the instance with gcloud

https://cloud.google.com/sdk/gcloud/reference/compute/ssh

```
diego@EstudioVM:~/10-Project/ldd-pix2pix$ gcloud compute ssh
diegotascon@aidl2021-vm
=====================================
Welcome to the Google Deep Learning VM
=====================================

Version: pytorch-gpu.1-7.mnightly-2021-01-20-debian-10-test
Based on: Debian GNU/Linux 10 (buster) (GNU/Linux 4.19.0-13-cloud-amd64
x86_64\n)

Resources:
 * Google Deep Learning Platform StackOverflow:
https://stackoverflow.com/questions/tagged/google-dl-platform
 * Google Cloud Documentation: https://cloud.google.com/deep-learning-vm
 * Google Group: https://groups.google.com/forum/#!forum/google-dl-platform

To reinstall Nvidia driver (if needed) run:
sudo /opt/deeplearning/install-driver.sh
Linux aidl2021-vm 4.19.0-13-cloud-amd64 #1 SMP Debian 4.19.160-2 (2020-11-28)
x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Sat Feb  6 15:23:01 2021 from 35.235.241.65
```

Starting and stopping remotely the instance

https://cloud.google.com/sdk/gcloud/reference/compute/instances/stop

You can just “sudo shutdown now” if you’re connected or use gcloud compute instances stop:
https://cloud.google.com/sdk/gcloud/reference/compute/instances/stop

Python installation

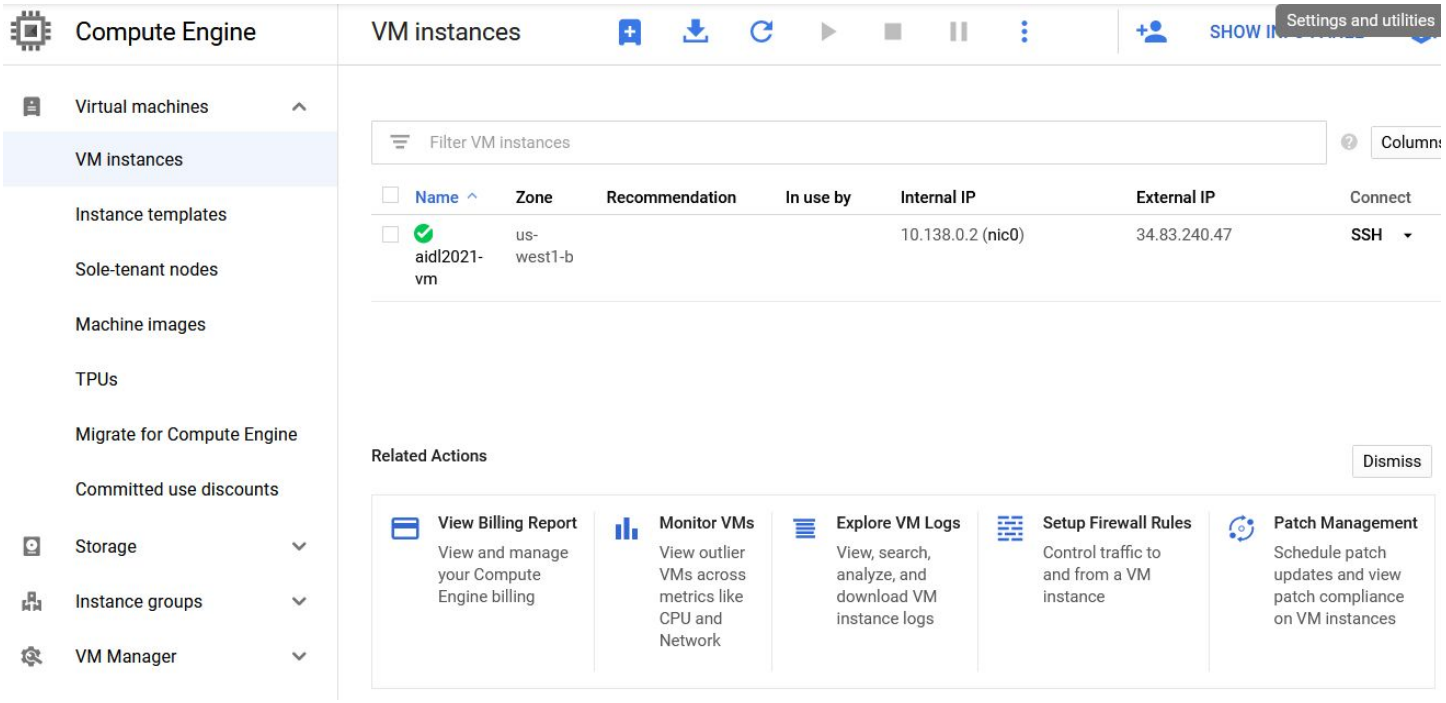
The instance comes with python and many modules already installed. For the project I had to additionally install:

- tensorboard
-

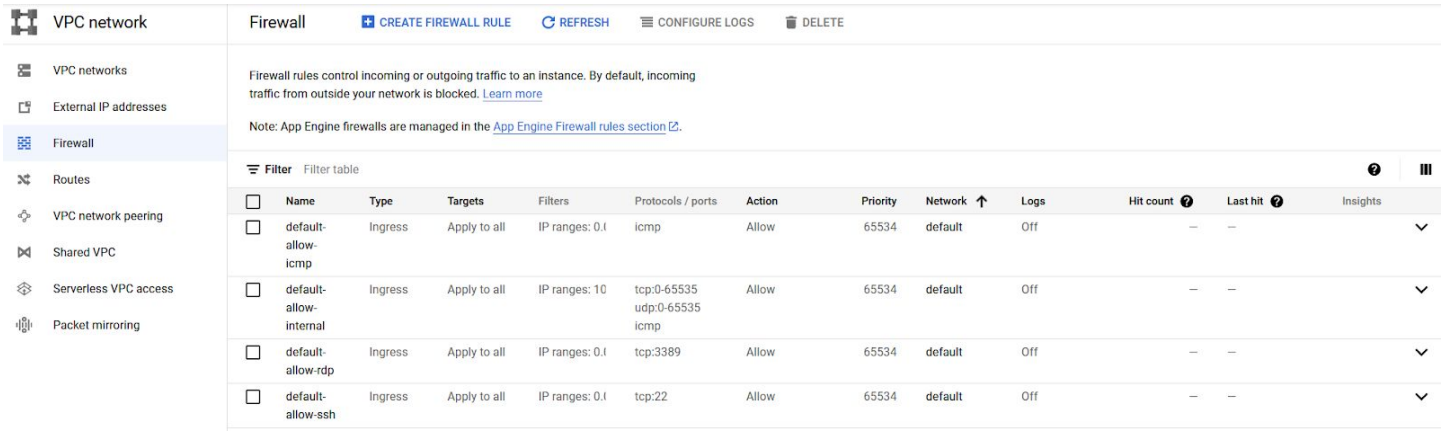
Minor settings

Changing firewall rules to allow external connections

Accessing Jupyter or Tensorboard requires allowing external connections to the instance. To modify the firewall rules, I went to the Compute Engine console:



I click on the “Setup Firewall Rules” link:



I pressed on “Create firewall rule”:

Create a firewall rule

Firewall rules control incoming or outgoing traffic to an instance. By default, incoming traffic from outside your network is blocked. Learn more

Name *

tensorboard

Lowercase letters, numbers, hyphens allowed

Description

Allows access to the tensorboard server running on aidl2021-vm

Logs

Turning on firewall logs can generate a large number of logs which can increase costs in Stackdriver. Learn more

- ○ On
- ◉ Off

Network *

default

Priority *

1000 CHECK PRIORITY OF OTHER FIREWALL RULES

Priority can be 0 - 65535

Direction of traffic

- ◉ Ingress
- ○ Egress

Action on match

- ◉ Allow
- ○ Deny

Direction of traffic

- (•) Ingress
- ( ) Egress

Action on match

- (•) Allow
- ( ) Deny

Targets

Specified target tags

Target tags *

aidl2021-vm

Source filter

IP ranges

Source IP ranges *

0.0.0.0/0 for example, 0.0.0.0/0, 192.168.2.0/24

Second source filter

None

Protocols and ports

- ( ) Allow all
- (•) Specified protocols and ports

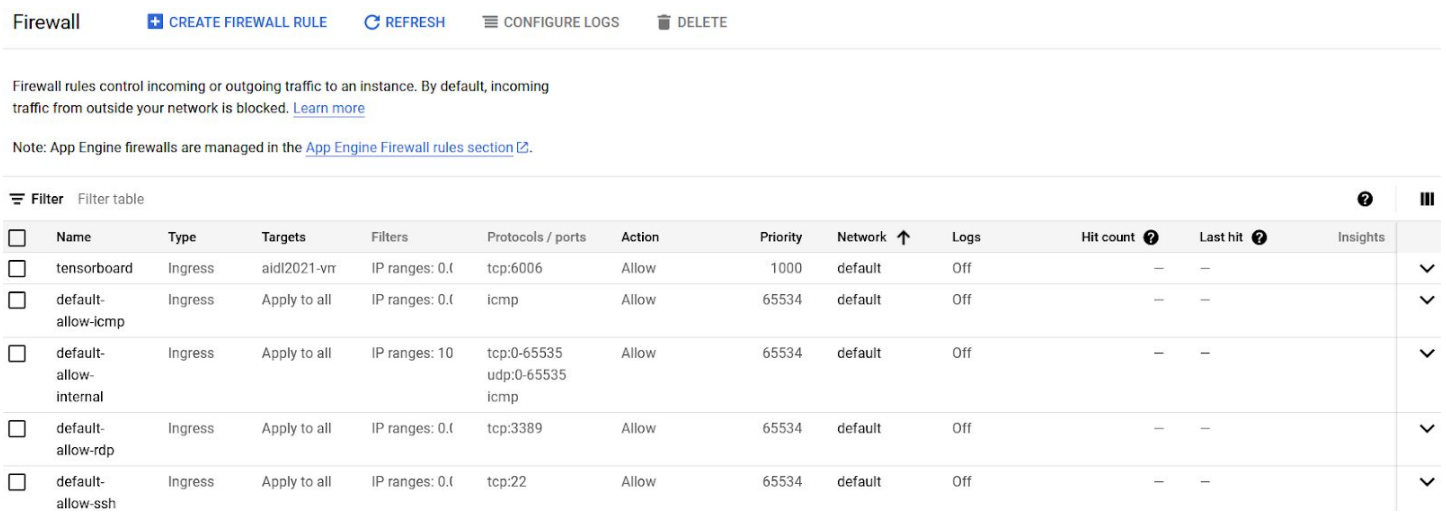
- [x] tcp : 6006
- [ ] udp : all
- [ ] Other protocols

protocols, comma separated, e.g. ah, sctp

DISABLE RULE

CREATE CANCEL

I filled in the gaps and I pressed “Create”.



Name	Type	Targets	Filters	Protocols / ports	Action	Priority	Network	Logs	Hit count	Last hit	Insights
tensorboard	Ingress	aidl2021-vm	IP ranges: 0.(	tcp:6006	Allow	1000	default	Off	–	–	
default-allow-icmp	Ingress	Apply to all	IP ranges: 0.(	icmp	Allow	65534	default	Off	–	–	
default-allow-internal	Ingress	Apply to all	IP ranges: 10	tcp:0-65535 udp:0-65535 icmp	Allow	65534	default	Off	–	–	
default-allow-rdp	Ingress	Apply to all	IP ranges: 0.(	tcp:3389	Allow	65534	default	Off	–	–	
default-allow-ssh	Ingress	Apply to all	IP ranges: 0.(	tcp:22	Allow	65534	default	Off	–	–	

The tensorboard rule appears. That didn’t work, so I edited the rule to allow any target in the network:

And that allowed me to access Tensorboard in the instance.