

GCP GPU instance

Wed, Jan 13, 2021 at 8:02 PM

We wanted to send you these instructions include any of your students are unable to receive any GPUs

the Edu grant. To do this, you will first need to ensure you have sufficient quota to use a GPU. You will then need to ensure that the VM you configure can support GPUs.

To request GPU quota follow these three steps:

STEP 1: CHECK YOUR PROJECT

GPUs are granted to projects, not to billing accounts or user accounts, so before requesting a GPU quota increase, create the project. Then make sure of the following:

- A. Be in the project that needs GPUs.
- B. Ensure the project has an Edu account as the billing account — This should reflect the name of the course that you received the Cloud credit for. You can go to Billing from the main menu to see the billing account and change it if necessary.

Billing

You have multiple billing accounts

Billing account "2020 Fall Conferences" is linked to this project

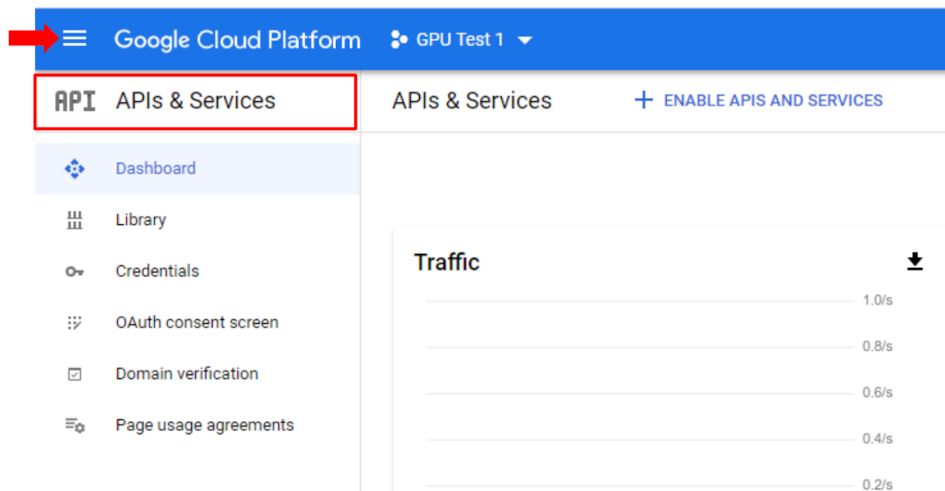
[GO TO LINKED BILLING ACCOUNT](#) [MANAGE BILLING ACCOUNTS](#)

Note: The [Free Trial program expressly does not permit GPUs to be used](#), so be sure the billing account listed is one associated to a class.

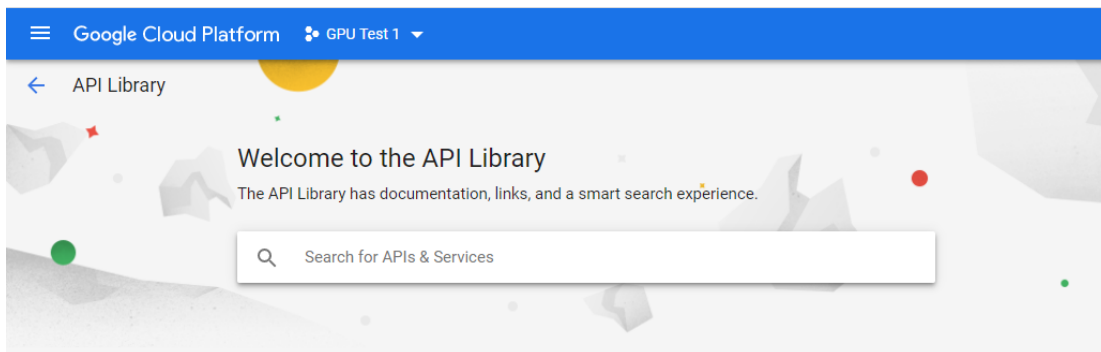
STEP 2: ENABLE COMPUTE ENGINE API FOR THE PROJECT

Ensure the **Compute Engine API** has been enabled for this project.

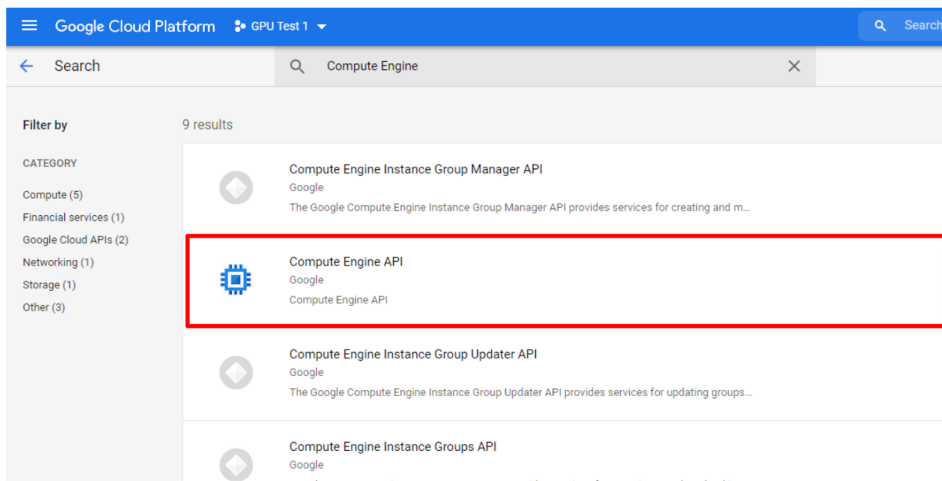
A. First go to **APIs & Services** in the main menu:



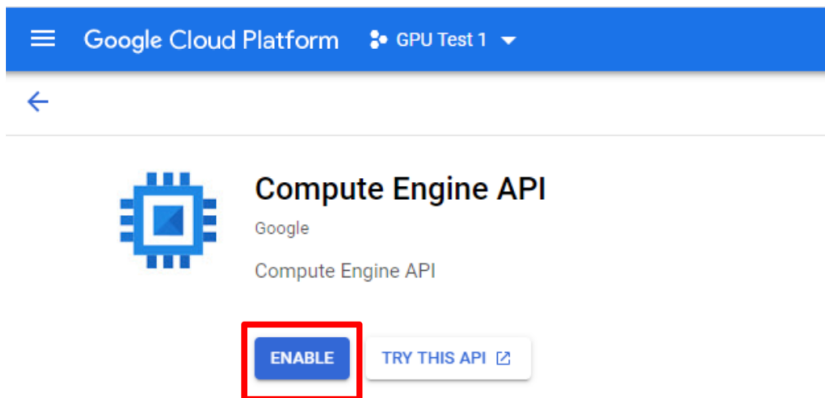
B. Then select **+ Enable APIs and Services** at the top of the console:



C. Search for Compute Engine in the search box. Select **Compute Engine API** from the list:



D. Click the button to **Enable** the API:



STEP 3: SET YOUR QUOTAS

Once you are in the correct project, you will need to ensure you have two different quotas set.

- A. First, to see your quotas, go to **Quotas**. You can do by using the search bar at the top at the top of the console or navigating through the IAM & Admin menu:
- B. Scroll to **Compute Engine API** quotas in the "Service" column, then the **GPUs** in the "Limit name" column. Depending on how many APIs you have enabled, this may be fairly far down in the list. You may find filters useful for this.

Service	Limit name	Quota status	Details
BigQuery API	Requests per day	One quota is within limit	ALL QUOTAS
BigQuery API	IamPolicy requests per minute	One quota is within limit	ALL QUOTAS
BigQuery API	Query usage per day	One quota is within limit	ALL QUOTAS
BigQuery API	Extract bytes per day	One quota is within limit	ALL QUOTAS
BigQuery API	Cloud SQL federated query cross region bytes per day	One quota is within limit	ALL QUOTAS
BigQuery Storage API	CreateWriteStream request limit per project per minute per minute	One quota is within limit	ALL QUOTAS
BigQuery Storage API	Pending stream bytes	One quota is within limit	ALL QUOTAS
Cloud Datastore API	Discovery requests per 100 seconds	One quota is within limit	ALL QUOTAS
Cloud Datastore API	Export Requests Per Minute	One quota is within limit	ALL QUOTAS
Cloud Datastore API	Import Requests Per Minute	One quota is within limit	ALL QUOTAS
Cloud Datastore API	Get Index Requests Per Minute	One quota is within limit	ALL QUOTAS
Cloud Datastore API	List Indexes Requests Per Minute	One quota is within limit	ALL QUOTAS
Cloud Datastore API	Get Operation Requests Per Minute	One quota is within limit	ALL QUOTAS
Cloud Datastore API	List Operations Requests Per Minute	One quota is within limit	ALL QUOTAS
Cloud Datastore API	Create or Delete Index Requests Per Minute	One quota is within limit	ALL QUOTAS
Cloud Debugger API	Debugger requests per day	One quota is within limit	ALL QUOTAS

C. Based on the "Service" column, select a GPU (in this case "NVIDIA P4") and see what the quota is in the panel that opens on the right side (in this case "1"). You should ensure any GPU you want to use is set to 1. Selecting more than one region will make it easier to find a region with GPUs available (if your closest region does not have enough when you need them).

Service	Limit name	Quota status	Details
Compute Engine API	NVIDIA V100 GPUs	All 24 quotas are within limit	ALL QUOTAS
Compute Engine API	NVIDIA P4 GPUs	All 24 quotas are within limit	ALL QUOTAS
Compute Engine API	NVIDIA P4 Virtual Workstation GPUs	All 24 quotas are within limit	ALL QUOTAS
Compute Engine API	Health checks	One quota is within limit	ALL QUOTAS
Compute Engine API	Node groups	All 24 quotas are within limit	ALL QUOTAS
Compute Engine API	Node templates	All 24 quotas are within limit	ALL QUOTAS

Service: Compute Engine API
Category: Location

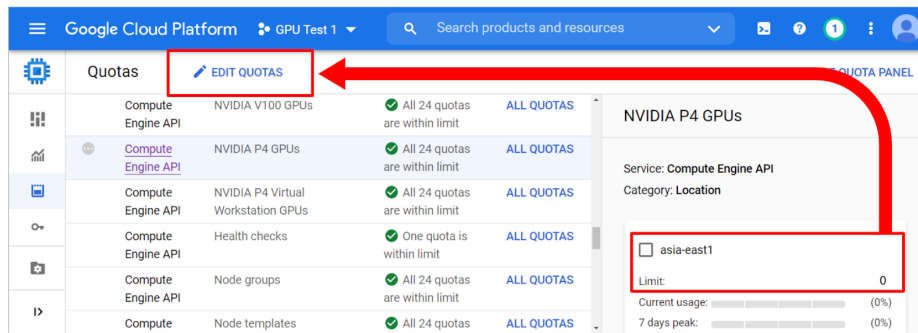
☐ asia-east1

Limit: 1

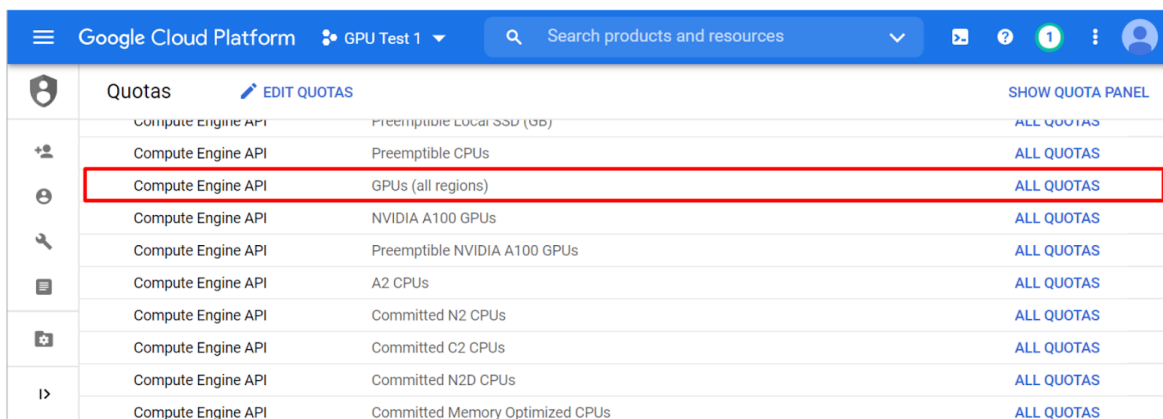
Current usage: (0%)

7 days peak: (0%)

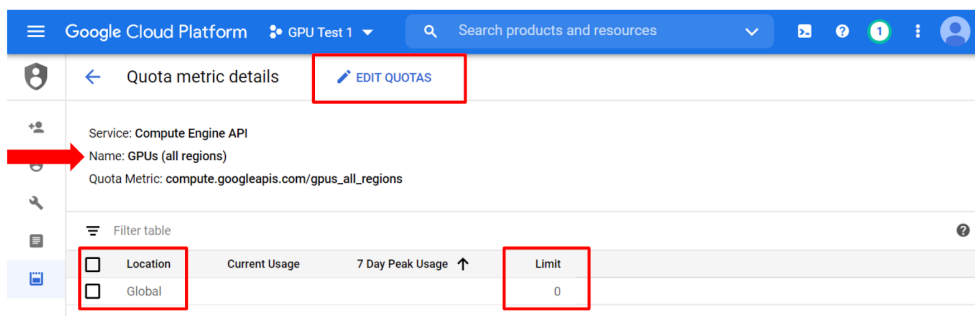
- D. If the quota for any GPU you want to use is not 1, change it to 1 by checking the desired GPU and region and selecting **Edit Quotas**. Fill in the prompts to increase quota. (See more information about this below.)
- DO NOT CHANGE THE QUOTA TO 2! Edu accounts only seem to automatically receive quota upgrades to 1 GPU.**
 - You may select multiple GPUs and regions.



E. Go to the quota **Compute Engine API > GPUs (all regions)**. This tends to be near the end of the list of quotas, not with the other GPU quotas:



- Select **GPUs (all regions)**. The quota panel will appear to the side of All Quotas or on its own page. If this quota is **0**, select the "Global" checkbox and then Edit Quotas.



- You'll then be prompted to complete the quota change request. Start by adding your name and email address, then click **Next**.

Google Cloud Platform GPU Test 1 Search products and resources

Quota metric details EDIT QUOTAS

Service: Compute Engine API
Name: GPUs (all regions)
Quota Metric: compute.googleapis.com/gpus_all_regions

Filter table

Location	Current Usage	7 Day Peak Usage	Limit
Global			0

1 quota selected

Contact details

These details will be sent to the approvers while reviewing quota change request.

Name *

Email *

Phone *

NEXT

- On the next screen, enter "1" as your new quota limit. **DO NOT CHANGE THE QUOTA TO 2! Edu accounts only seem to automatically receive quota upgrades to 1 GPU.** Your request may fail if you ask for more than one as your new limit. Include a description, then click **Done**.

Google Cloud Platform GPU Test 1 Search products and resources

Quota metric details EDIT QUOTAS

Service: Compute Engine API
Name: GPUs (all regions)
Quota Metric: compute.googleapis.com/gpus_all_regions

Filter table

Location	Current Usage	7 Day Peak Usage	Limit
Global			0

1 quota selected

Quota: GPUs (all regions)

Current limit: 0

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

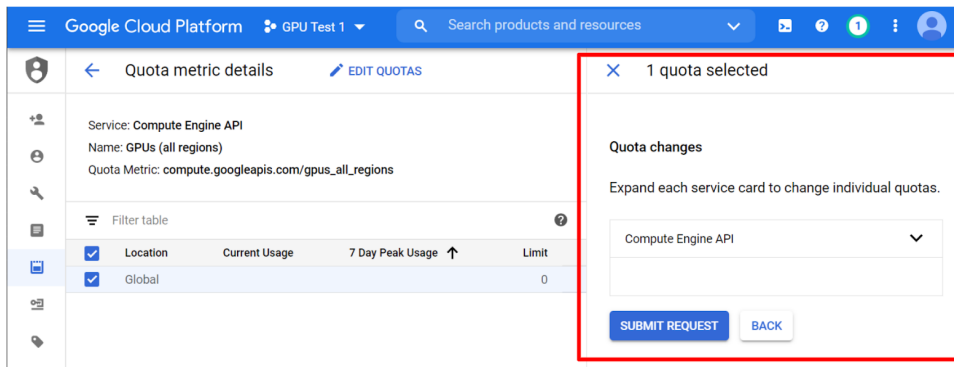
New limit *

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

- On the next screen, submit your request.



Check your email to ensure that the request was submitted and then wait. We have seen response times as quick as 1 minute, but to be safe, please have your students go through this process a few days before they will need GPUs.

To use GPUs

You cannot attach every GPU board to every VM in every region.

- For information about regional availability and restrictions, please see [this page](#).
- For information about suggested VMs, please see [this page](#).

GPUs can only be attached to general purpose N1 machine types, as per [this page](#).

[Quoted text hidden]

[Quoted text hidden]