

# How to use Google Cloud Platform with GPU

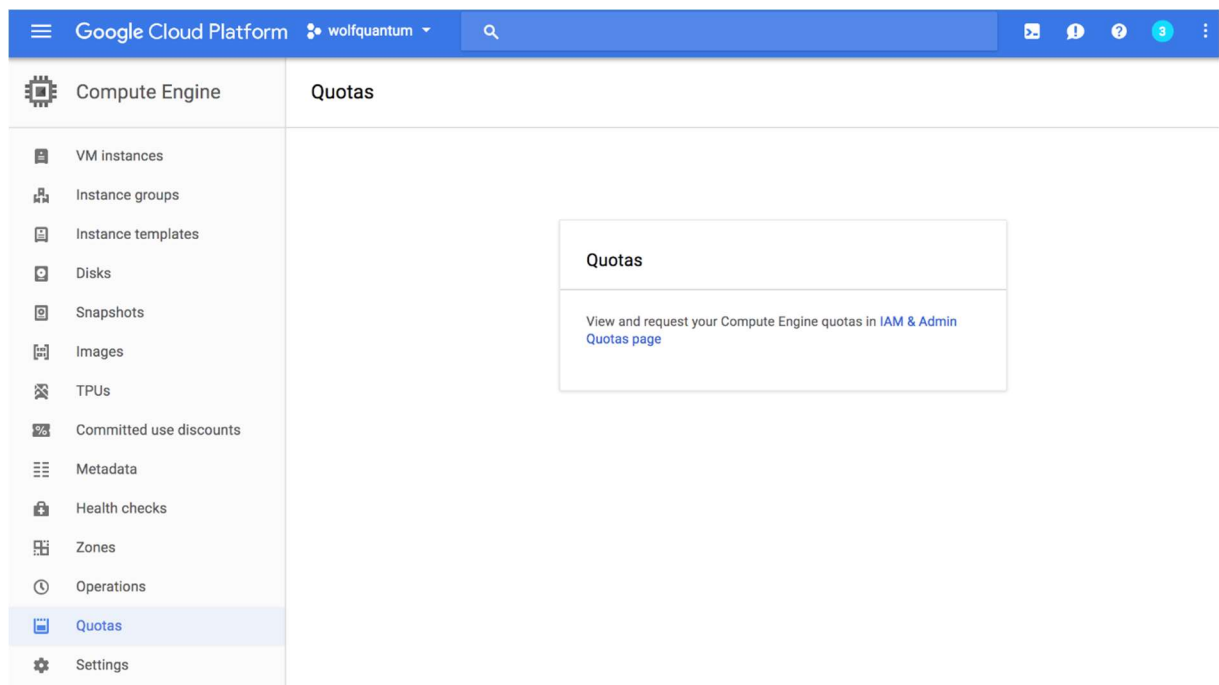
**Peiyan Gong**

6/8/2018

**Warning:** GPU instances are quite expensive. Please shut down your GPU instances when you are not using them.

To those who applied for the \$300 Google Cloud free-trial coupons: the \$300 public coupons cannot be used to apply for GPU quotas. Also, you need to upgrade your account into a paid account to issue GPU quotas. These quota is to ask for permission of using GPUs. You must issue a quota before using GPUs. These quotas take about 10 mins to a couple hours to process.

In compute engine page, select Quotas on the right, and click "IAM & Admin Quotas page."



On the IAM & Admin Quotas page, in the Metric tab choose K80 and P100 GPUs; in the Location tab, we suggest you to choose us-east1.

Metric

2 metrics

Location

us-east1

Clear

X

Select all

None

License insert requests per 100 seconds per user

License insert requests per day

List requests per 100 seconds

List requests per 100 seconds per user

Local SSD (GB)

Log read bandwidth per day

Log stored data per day

Managed instance groups

Memcache Compute Units per day

Memcache key operations per day

✓ NVIDIA K80 GPUs

✓ NVIDIA P100 GPUs

Networks

Number of indexes per day

Operation read requests per 100 seconds

Operation read requests per 100 seconds per user

Check the GPU type you want, and hit "EDIT QUOTAS." Fill in your information. P100 is faster and more expensive than K80. However, P100 is more powerful than K80. Also, GCP now offer V100 GPUs as Beta version. A more detailed information of the comparison of these GPUS can be found here: <https://cloud.google.com/compute/docs/gpus/>

IAM & admin

- IAM
- Identity
- Quotas**
- Service accounts
- Labels
- GCP Privacy & Security
- Settings
- Encryption keys
- Identity-Aware Proxy
- Roles

Quotas [EDIT QUOTAS](#)

Quota type	Service	Metric
All quotas	Google Compute Engine API	2 metrics

Service	Location
<input type="checkbox"/> Google Compute Engine API NVIDIA P100 GPUs	us-east1
<input checked="" type="checkbox"/> Google Compute Engine API NVIDIA K80 GPUs	us-east1

1 quota selected

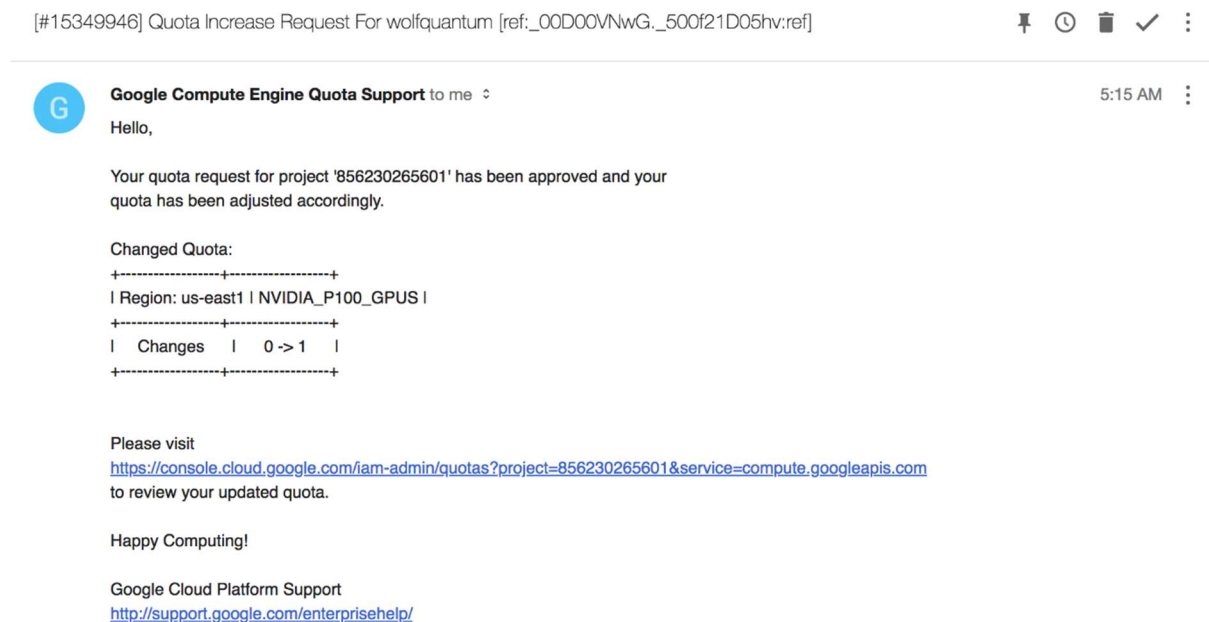
Edit quotas
Name  

Email  

Phone

Next

After you submit the form, you should receive an email like this one in two minutes.



Now create your VM machine. When you create your VM instance, please choose "us-east1-\*". Remember to select GPU as well.

Google Cloud Platform EECS442

Compute Engine

VM instances

Instance groups

Instance templates

Disks

Snapshots

Images

Committed use discounts

Metadata

Health checks

Zones

Operations

Quotas

Settings

Create an instance

Name

Zone

Machine type

Cores  vCPU 1 - 64

Memory  GB 1 - 6.5

GPUs

The number of GPU dies is linked to the number of CPU cores selected for this instance. For this machine type, you can select no fewer than 1 GPU die. [Learn more](#)

Number of GPUs  GPU type

Machines with GPUs can't be preemptible nor migrate on host maintenance

\$535.67 per month estimated

Effective hourly rate \$0.734 (730 hours per month)

[Details](#)

Finally, enter VM instances page and select instance-1. The “START” button will light up and click this button. Then click “SSH” button under “Connect”.

Google Cloud Platform CLFusion

Compute Engine

VM instances

CREATE INSTANCE

IMPORT VM

REFRESH

START

STOP

RESET

DELETE

Filter VM instances

Name	Zone	Recommendation	Internal IP	External IP	Connect
instance-1	us-east1-b		10.142.0.2 (nic0)	None	SSH

Select an instance

LABELS

MONITORING

No instances selected.

A new window will pop up. This window is the terminal of the VM instance. Have fun!