MutPanning Software User Guide

Version: July 2018
Boston, USA

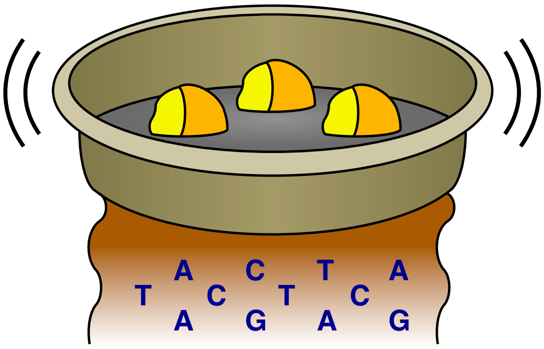


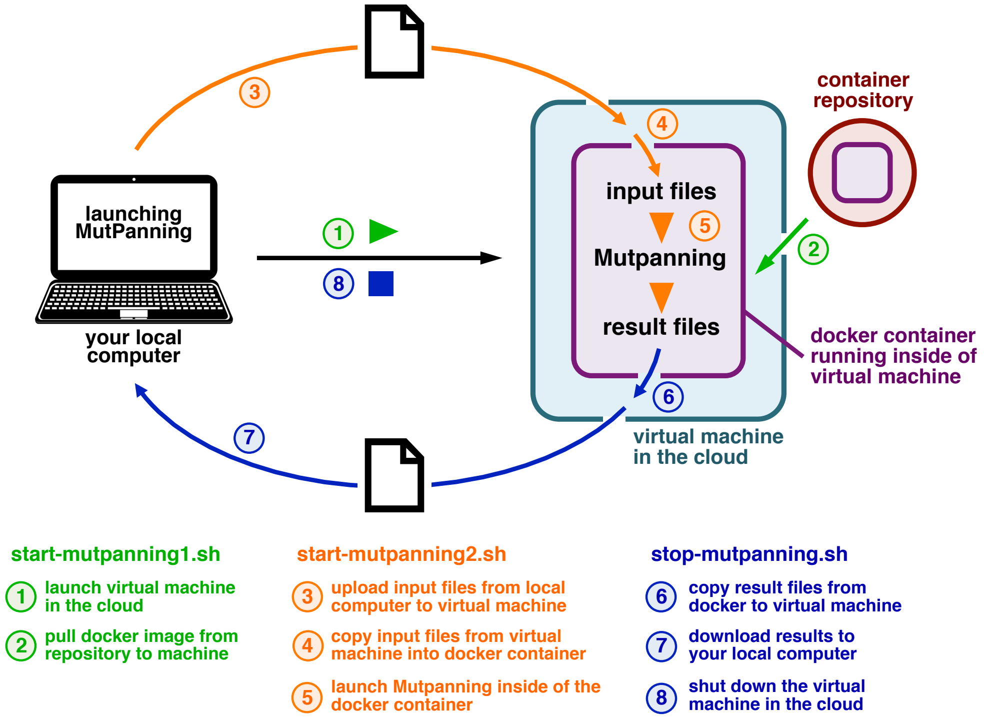
Table of Contents

1. Software Architecture of MutPanning

The MutPanning Software is designed to run as a Docker Container in a Computing Cloud. Docker containers allow distributing a software together with a standardized operating system. When running a docker, it spins up its own operating system, in which all variables and software dependencies are defined. Hence, docker files allow platform-independent execution of software, as the environment (e.g. necessary compilers, software libraries, file structure) is encoded in the docker.

As MutPanning launches multiple processes in parallel, a multi-CPU system is needed, so that MutPanning will not be run on your local computer (i.e. no particular hardware resources required), but on a server, for instance in a computing cloud. To facilitate the usability of our software, we have summarized these functions in 3 bash scripts that fully handle the communication with the cloud. In particular, they launch a virtual machine in the cloud, spin up the docker , launch the Mutpanning script in the cloud, as well as upload and download your data to and from the cloud.

This software manual is primarily intended for less experienced users, who have never worked with dockers or the cloud. Here, we will provide you with a step-by-step instruction, on how to launch MutPanning in the Google Cloud. No prior knowledge on how to work in the cloud is required. We will start from creating a Google Cloud Account and configuring it for MutPanning. More experienced users, who are already familiar with the Docker concept, will probably be able to skip many of these steps. Further, we note that Docker is a general conepct and tha most larger computing clouds allow the deployment of Docker Containers. Hence, MutPanning is not a Google Cloud product, but can be run in most clouds or local servers. We would just like to give you an exemplary step-by-step instruction, on how to run it in the Google Cloud, so that less experienced users do not have to familiarize with these concepts, if they wish to use our tool.


container repository
launching MutPanning
your local computer
input files
Mutpanning
result files
docker container running inside of virtual machine
virtual machine in the cloud
1
2
3
4
5
6
7
8
start-mutpanning1.sh
1 launch virtual machine in the cloud
2 pull docker image from repository to machine
start-mutpanning2.sh
3 upload input files from local computer to virtual machine
4 copy input files from virtual machine into docker container
5 launch Mutpanning inside of the docker container
stop-mutpanning.sh
6 copy result files from docker to virtual machine
7 download results to your local computer
8 shut down the virtual machine in the cloud

2. Prerequisites before your first MutPanning run

The following prerequisites are needed in order to run the MutPanning Software.

- Local computer with MacOS or other UNIX-based operating system. Note that if you use Windows, you are still able to run MutPanning in the cloud. Only the bash scripts that automatically spin up a Virtual Machine require a UNIX-based system, so that you would need to do this manually.
- Google Chrome (recommended)
- gcloud command line tool with beta extension
- Python 2.7.x or later
- Google account configured for Cloud Computing (see instructions below). Please note that MutPanning comes as a dockerized container (docker.com) and most compute clouds support container deployment. We will show in this manual how to run this docker in the Google Cloud. However, you can use any cloud which supports container deployment to run MutPanning.

This chapter will provide you wiht step-by-step instructions on how to meet these prerequisites. Note that MutPanning will be entirely run in the cloud so that no particular hardware requirements are needed for your local computer. Furhter, please note that this chapter is mainly intended for users that have never worked in the Google Cloud before. If you have worked with Google Cloud previously or wish to use a different cloud, you can skip this chapter.

2.1 Install Google Chrome

We highly recommend using this software with the Google Chrome browser. In principle any browser can be used to work with the Google Cloud. However, Chrome has been optimized to interact with the cloud and the gcloud command line tool. The browser can be downloaded here:

https://www.google.com/chrome/

2.2 Install the gcloud command line tool

gcloud is a command line tool, which interacts with the Google Cloud. Our bash scripts, use this command line tool. Please note that gcloud requires Python 2.7.x or later. If you do not have python installed on your local computer, you should first install Python, which can be downloaded here:

https://www.python.org/download/releases/2.7/

Further you will need a Google Account (gmail) to be able to download and install gcloud. If you do not have a gmail account, Sign Up for a Google Account first.

There are multiple ways to install the gcloud tool. If you never worked with gcloud before, we highly recommend using the interactive installation, as this will not only install gcloud with standard settings, but also take care of all post processing steps, such as adding gcloud to your PATH, which will allow you to use it in your command line.

For this purpose, type into the command line (Terminal)

```
curl https://sdk.cloud.google.com | bash
```

which will automatically download and install the files. Follow the instructions on the screen and select the standard settings to avoid any unusual behavior. Note that if you installed the Python in an unusual location, you might be prompted to define the path to your python installation, so that gcloud can find it. This step will take 3-5 minutes. Next type into the command line

```
exec -l $SHELL
```

to restart your command line and make changes effective. Finally, you need to configure your gcloud command line tool. For this purpose, type into the command line

```
gcloud init
```

and follow the instructions on the screen. Select the standard settings to avoid any unusual behavior. At this stage, you will also be asked for a Google Account. If you already know under which Google Account you would like to perform Cloud Computing, use this account here. If you have not decided at this stage, proceed with any Google Account, which you have access to. This can easily be changed at a later stage. For more detailed instructions, go to

https://cloud.google.com/sdk/docs/downloads-interactive

and make sure that the right OS is selected.

Finally, the beta extension of gcloud needs to be installed. After gcloud is installed and initiated, type

```
gcloud components install beta
```

into the command line to install the beta extension of gcloud. Now the gcloud tool should be completely configured for all subsequent steps.

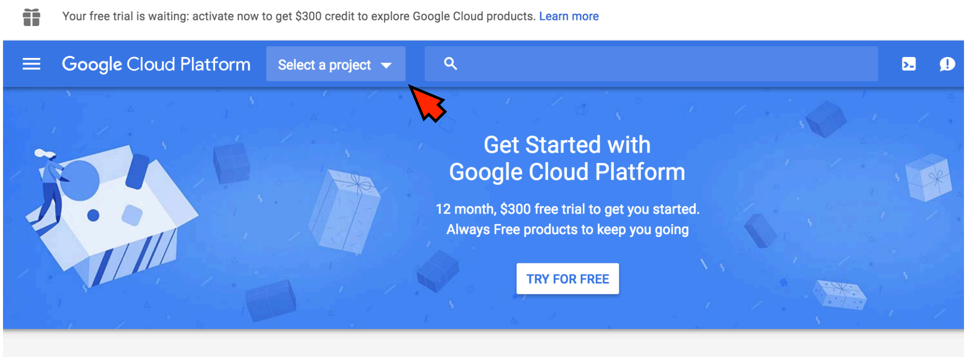
2.3 Setup a Google Account for Cloud Computing

We recommend carrying out all the following steps in the Google Chrome browser. Using your Google Account, log in to the google cloud console under

http://console.cloud.google.com/

Preferably you would use the same Google Account here, which you also used to install the gcloud tool in the previous section. However, this not mandatory. If this you log into the Google Cloud Console for the first time, you need to agree to the Terms of Service to be able to use the Google Cloud (select Yes in the second question).

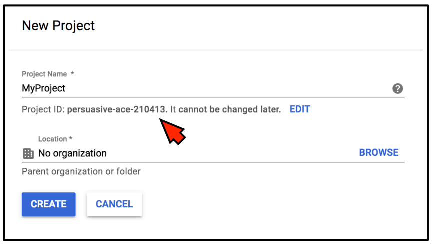
In the upper left corner click on *Select a Project*.



Your free trial is waiting: activate now to get $300 credit to explore Google Cloud products. Learn more

A window pops up, which asks you to select a project. Currently, you do not have any project, so click in the upper right corner on *New Project* to create a new project.

Select a *Project Name*. Although supported by the Google Cloud, we strongly recommend not using any special characters, blanks, or underscores here, as they might interfere with the bash scripts later. Ideally, use letters and numbers only.



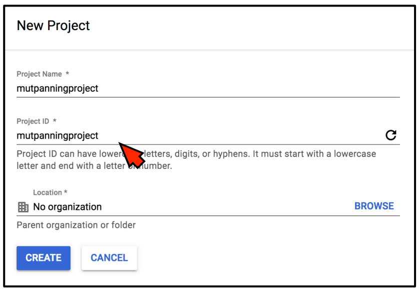
New Project

Project Name *
MyProject

Project ID: persuasive-ace-210413. It cannot be changed later. EDIT

Location *
No organization
Parent organization or folder

Please also note that Google Cloud automatically generates a Project ID for you (displayed below the project name). This ID will be needed in all subsequent scripts (such as *MyProject* in this example), so make sure that it is as simple as possible. If you choose common names as the project name, Google Cloud will assign an arbitrary ID to your project, as the name has already been taken by somebody else and the Project ID needs to be unique in the entire cloud. Hence, we recommend using less common or longer names, which are still free, as in this case the Project ID will be the same as the Project name, which will make the handling of the subsequent steps a lot easier.



New Project

Project Name *
mutpanningproject

Project ID *
mutpanningproject

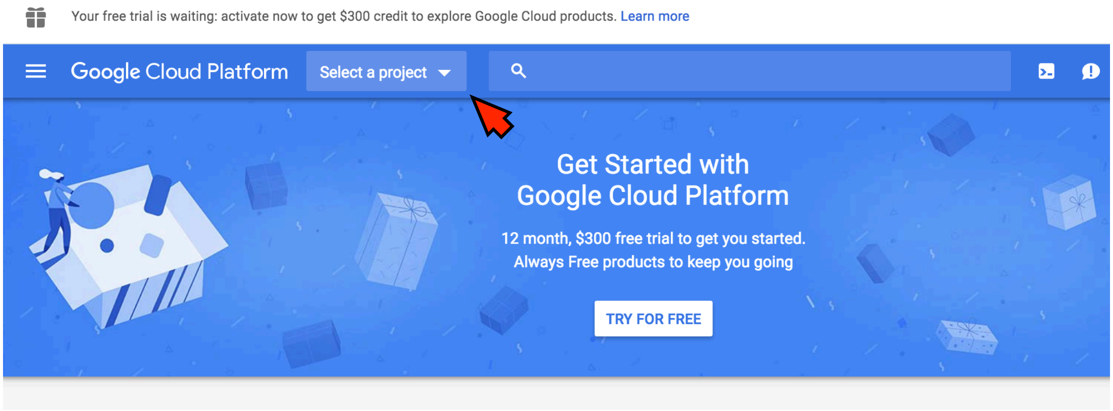
Project ID can have lowercase letters, digits, or hyphens. It must start with a lowercase letter and end with a letter or number.

Location *
No organization
Parent organization or folder

Alternatively, you also have the possibility to change the Project ID manually by clicking on *EDIT* next to the automatically generated Project ID.

Finally, click on *CREATE* to create a new project.

You will return to the initial google console window. Click again on *Select a project*.

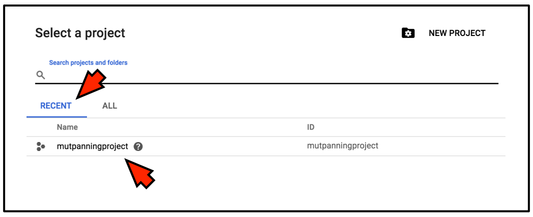


Your free trial is waiting: activate now to get $300 credit to explore Google Cloud products. Learn more

Get Started with Google Cloud Platform

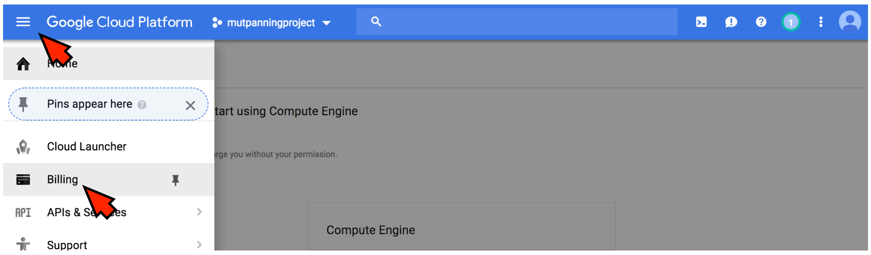
12 month, $300 free trial to get you started.
Always Free products to keep you going

Under *RECENT* you will now find the project, you just created. Click on its name and *OPEN* to open the project.

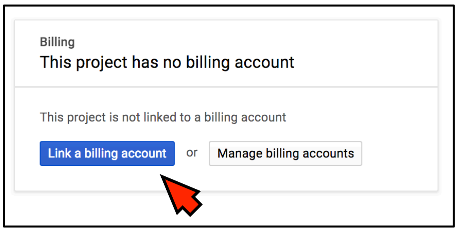


Next, you need add a billing account to your project. In the Navigation Menu in the top left (icon left next to Google Clout Platform)

select *Billing*.



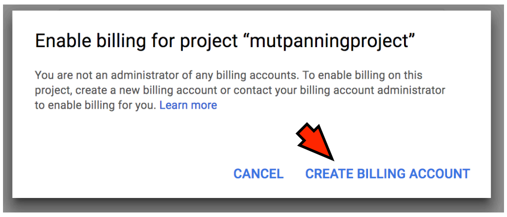
Click on *Link a billing account* to add a billing account to this project.


Billing

This project has no billing account

This project is not linked to a billing account

As you have not configured any billing accounts, click on *CREATE BILLING ACCOUNT*.

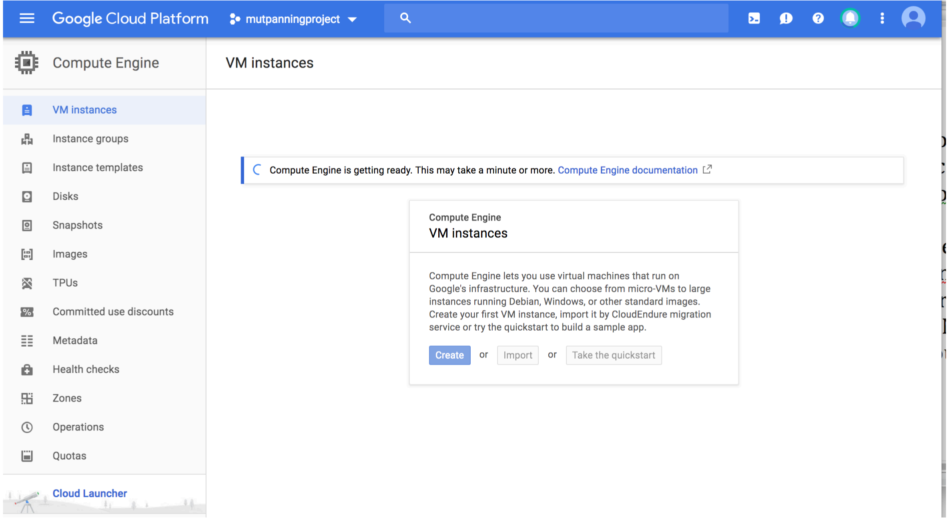

Enable billing for project “mutpanningproject”

You are not an administrator of any billing accounts. To enable billing on this project, create a new billing account or contact your billing account administrator to enable billing for you. Learn more

Agree to the Terms of conditions and follow the instructions to enter your billing information. With the current rates, the expected Google Cloud Computing costs for a complete MutPanning run on 500 samples are approx. $5 (compute time: 3-4 hrs), the cost of a run on 12,000 samples are approx. $30 (compute time: 24 hrs). Please note that if you create a new Google Cloud Compute account, you will receive a credit of $300.

It might now take a couple of minutes until your account is configured to launch virtual machines in the cloud. You can monitor the status of your account by clicking on the Navigation Menu

in the top left corner and selecting *Compute Engine* -> *VM Instances*.
Your screen should look like this, while your account is being configured. However, you can already proceed with the next step to extend your quotas, while your account is getting ready.



Compute Engine is getting ready. This may take a minute or more. Compute Engine documentation

Compute Engine
VM instances

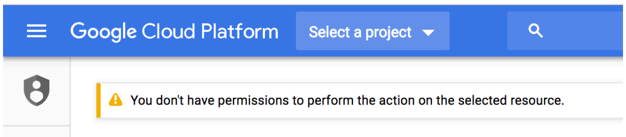
Compute Engine lets you use virtual machines that run on Google's infrastructure. You can choose from micro-VMs to large instances running Debian, Windows, or other standard images. Create your first VM instance, import it by CloudEndure migration service or try the quickstart to build a sample app.

2.4 Extend Resource Quotas

MutPanning needs to launch a Virtual Machine in the cloud with 24 CPUs, 60GB memory, as well as 2000 GB space on a persistent SSD volume. However, there are certain standard resource limits, which you can allocate in the cloud and you will first need to ask for an extension of these limits, before you can launch larger instances. For the SSD volume space, the standard limit is currently set to 500 GB, so that you will need to ask for an extension of this limit. To ask for a quota extension log into your google account. Then launch

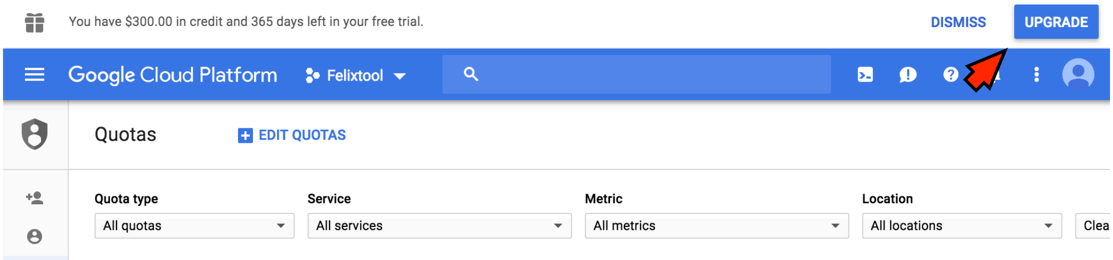
console.cloud.google.com/iam-admin/quotas

If you receive a message that you do not have the permissions to ask for an extension

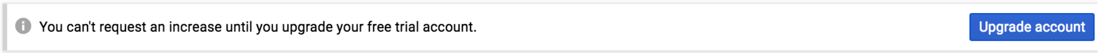


You don't have permissions to perform the action on the selected resource.

this usually indicates that you have not selected a project. Quota extensions always apply to a certain project and not the account in general. In this case, click on *Select a project*, and select the project you would like to ask for extensions.
If you enrolled in the free trial, you will first need to upgrade your account by clicking on *UPGRADE* in the upper right corner.

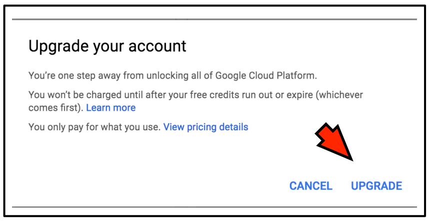

You have $300.00 in credit and 365 days left in your free trial.

Sometimes this button appears somewhere else, like in this example, which was displayed below the selection menus. In this case, click on *Upgrade account* to proceed.



You can't request an increase until you upgrade your free trial account.

The following dialog that pops up.



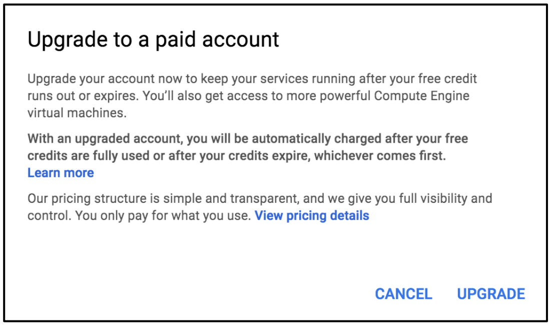
Upgrade your account

You're one step away from unlocking all of Google Cloud Platform.

You won't be charged until after your free credits run out or expire (whichever comes first). Learn more

You only pay for what you use. View pricing details

In some cases, the message looked like this



Upgrade to a paid account

Upgrade your account now to keep your services running after your free credit runs out or expires. You'll also get access to more powerful Compute Engine virtual machines.

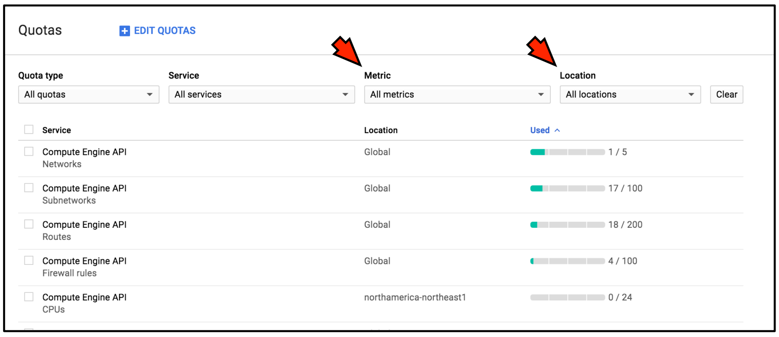
With an upgraded account, you will be automatically charged after your free credits are fully used or after your credits expire, whichever comes first. Learn more

Our pricing structure is simple and transparent, and we give you full visibility and control. You only pay for what you use. View pricing details

In either case, click on *UPGRADE* to proceed.

This upgrade allows Google Cloud to bill you for computational costs and you will not be able to ask for a quota extension otherwise. Note that this step might not be necessary if you have not enrolled into the free trial or your free trial has already expired. If you do not find the upgrade button, try to proceed with the following steps for quota extension. If you are able to proceed, your account already has been upgraded to a billable account.

Now you should be able to see an overview over all resources available. To ask for a quota extension of the persistent SSD hard drive space, click on the menu *Metric* (3rd column from the left). In the pop down menu first click on *NONE* and then type into the search field *SSD*. Select *Persistent SSD (GB)* from the menu.



Quotas

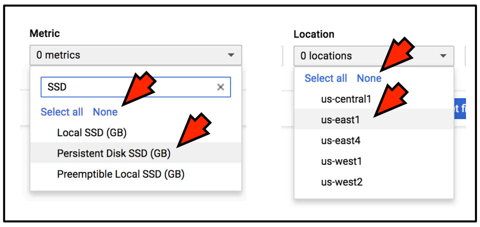
Quota type: All quotas

Service: All services

Metric: All metrics

Location: All locations

Service	Location	Used
Compute Engine API Networks	Global	1 / 5
Compute Engine API Subnetworks	Global	17 / 100
Compute Engine API Routes	Global	18 / 200
Compute Engine API Firewall rules	Global	4 / 100
Compute Engine API CPUs	northamerica-northeast1	0 / 24



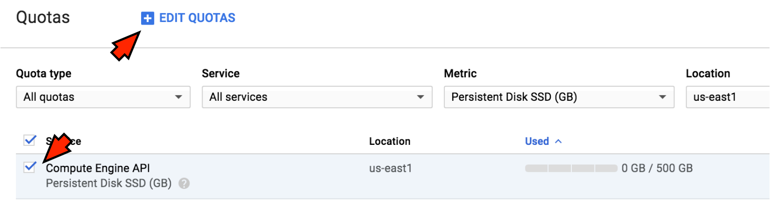
Metric: 0 metrics

SSD

Location: 0 locations

Similarly, you need to select a location in which you would like to request for the quota extension. For this purpose, Click on the menu *Location*, select *NONE* in the pop down menu and then select the location in which you would like to launch your Virtual Machine. Please note that we have successfully tested MutPanning in zones us-east1, us-east4 and us-central1. However, we generally recommend selecting the zone that is closest to you, in order to facilitate the data transfer.

With these filters you should now only see exactly one item left. Select this item by clicking on the box on the left next to this item. Then click on *EDIT QUOTAS*.

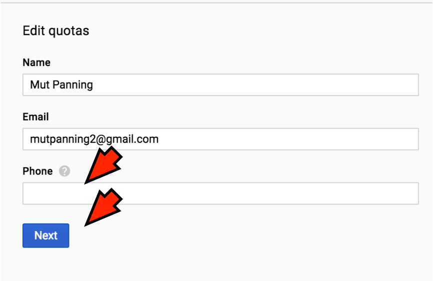


Quotas

Quota type	Service	Metric	Location
All quotas	All services	Persistent Disk SSD (GB)	us-east1

☑ Service	Location	Used
☑ Compute Engine API Persistent Disk SSD (GB)	us-east1	0 GB / 500 GB

In the dialog, provide your phone number and click on *Next*. In the next dialog window upper line, type *2000* for the *New quota limit*. Please note that this is the space needed for one run. If you would like to perform multiple runs in parallel or you already have other Virtual Machines running, you might need to request a higher threshold.

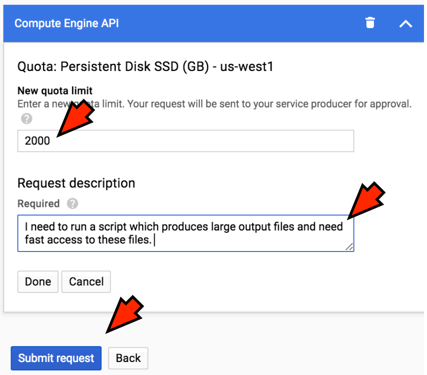


Edit quotas

Name: Mut Panning

Email: mutpanning2@gmail.com

Phone



Compute Engine API

Quota: Persistent Disk SSD (GB) - us-west1

New quota limit

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

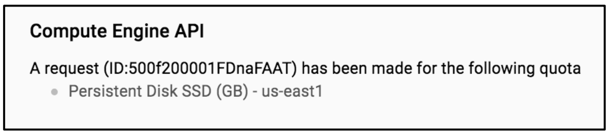
2000

Request description

Required

I need to run a script which produces large output files and need fast access to these files.

Finally, provide a quick *Request description*, for instance “I need to run a script which produces large output files and needs fast access to these files.”. Finally, click on *Submit request*. You might need to wait a couple of seconds, until your request has been successfully submitted, which should look like this.

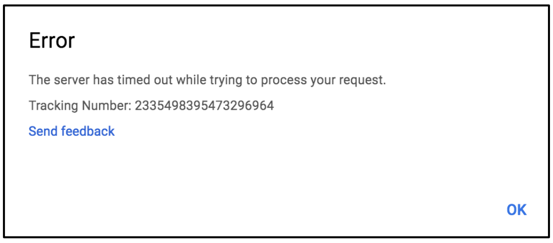


Compute Engine API

A request (ID:500f200001FDnaFAAT) has been made for the following quota

- Persistent Disk SSD (GB) - us-east1

In very rare cases, we observed that the request was not successfully submitted, most likely as the cloud was overloaded. In this case, we recommend selecting a different zone. If the error should persist, trying to submit the quota extension later, solved this problem. The error message we received typically looked like this.

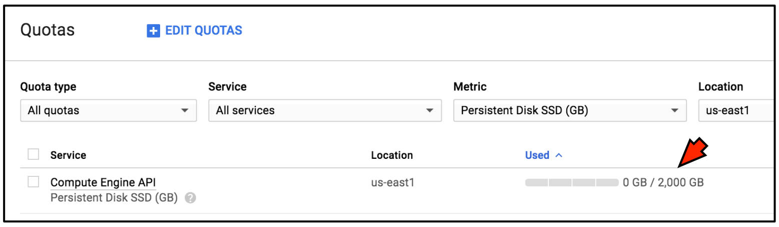


Error

The server has timed out while trying to process your request.

Tracking Number: 2335498395473296964

In principle, this request can take up 48 hours to be granted. However, as this is a relatively small request, in our experience these requests were fulfilled within less than 5 minutes. When you go back to your quotas page, you should be able to see that your space limit was extended to 2000 GB.



Quota type: All quotas
Service: All services
Metric: Persistent Disk SSD (GB)
Location: us-east1

Service	Location	Used
Compute Engine API Persistent Disk SSD (GB)	us-east1	0 GB / 2,000 GB

Please note that additional quota extensions for the CPUs might be needed if you would like to run multiple runs of MutPanning in parallel. These can be performed in a very similar manner. Please note that CPU quota extensions can be avoided if you distribute your runs across different zones (current standard limit is 24 CPUs per zone and MutPanning needs 24 CPUs).

Congratulations - With this your Google Cloud Account should be all set to run MutPanning.

2.5 Download the MutPanning Software

Finally, you need to download the MutPanning software from the same website, where you downloaded this manual. The following zip folders are available:

1. mutpanning_scripts (mandatory) This folder contains four small bash scripts, which will perform all the communication with the Google Cloud for you.

2. mutpanning_testfiles (optional). This folder contains files on which you can test the MutPanning software. Either you can use the PanCancer dataset, which was used in our study. If you rather prefer a quick test of MutPanning (runtime approx. 3 hrs),

In the folder PanCancer you find the entire dataset used in this study. If you wish to have a quick test of MutPanning (approx. 3 hrs) on a smaller dataset, we also prepared a smaller test dataset of the 598 melanoma samples included in our study cohort. If you wish to use your own data, you do not necessarily need to download these files. However, we encourage you to download these files, in order to prepare the structure of your data in a similar way. A detailed instruction how to prepare your own data will be given in the next chapter.

3. mutpanning_docker (optional). This folder contains all the complete source code of MutPanning together with all the files it is dependent on and the docker file. This folder is relatively large (approx. 1.5GB) and you will typically not need this folder, unless you would like to recompile MutPanning, for instance if you would like to execute MutPanning with non-standard settings. Otherwise, you can fully execute MutPanning with our precompiled docker images in the Google Container Registry instead. As recompiling takes a relatively large time, we recommend running MutPanning with the standard settings, if you are new to this software.

2.6 Build your own MutPanning docker file (optional)

If you wish to recompile MutPanning starting from the docker file (mutpanning_docker), we will provide you with a step-by-step instruction in this paragraph. However, as this step is time and resource-intensive, we do not recommend this, if you are new to this software. Instead, you can fully run MutPanning form our precompiled images in the Google Container Registry with standard settings. Rebuilding the docker file allows you to run MutPanning with non-standard settings.

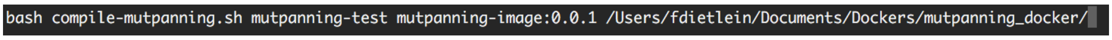
The first step to build your own docker image, is to download the zip folder mutpanning_docker.zip and unzip the folder, unless you have not already done so in the previous step. Please do not unzip any of the subfolders, as the scripts will automatically do this step later. Now run the bash script **compile-mutpanning.sh** with the following arguments

1. **Project ID** The name / ID of the project under which you would like to build and deposit your local docker image. This has to be exactly the project ID that you chose when creating the project. Note that all Virtual Machines within the same project will have automatically access to this local image. However, if you launch VMs in a different project, you will have to adjust the access rights manually.

2. Image name The name under which you would like to deposit the local image. This name can be fully arbitrary, however, please avoid blanks, underscores and special characters as they might interfere with the execution of the bash script. Further you can add tags to your image, by using ":" after the image name (e.g. mutpanning:0.0.1). These tags can be used for version control and if you compile under the same name with a different tag, both of these images will be saved together under the same path.

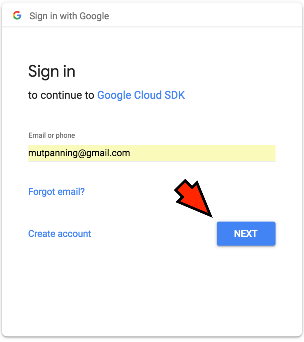
3. Docker folder The full path to the unzipped docker folder mutpanning_docker.

An exemplary execution of this script is shown below.



```
bash compile-mutpanning.sh mutpanning-test mutpanning-image:0.0.1 /Users/fdietlein/Documents/Dockers/mutpanning_docker/
```

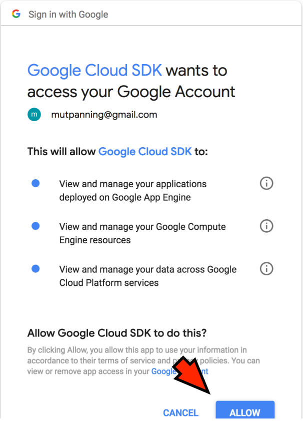
The script will first ask you to login in a browser window. The script will then ask you for permission to access your google account. Click on ALLOW to continue.



Sign in

to continue to Google Cloud SDK

Email or phone: mutpanning@gmail.com



Google Cloud SDK wants to access your Google Account

mutpanning@gmail.com

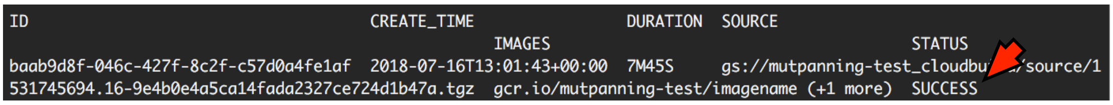
This will allow Google Cloud SDK to:

- View and manage your applications deployed on Google App Engine
- View and manage your Google Compute Engine resources
- View and manage your data across Google Cloud Platform services

Allow Google Cloud SDK to do this?

By clicking Allow, you allow this app to use your information in accordance to their terms of service and privacy policies. You can view or remove app access in your Google Account

Please note that the first step of the compilation is the upload for the docker file to the Google Cloud (approx. 2 GB), so that a fast and stable internet connection is required in this step. Please avoid interruption of the internet connection during this time. This is the time-limiting step of this script.
After the upload is completed the compilation takes approx.. 5-7 minute depending on the work load in the cloud. Finally you should see a screen like this which indicates that your docker image was built successfully and stored in the Google Container Repository.



```
ID                                    CREATE_TIME                DURATION   SOURCE
                                                   IMAGES                                          STATUS
baab9d8f-046c-427f-8c2f-c57d0a4fe1af  2018-07-16T13:01:43+00:00  7M45S      gs://mutpanning-test_cloudbu[illegible]/source/1
531745694.16-9e4b0e4a5ca14fada2327ce724d1b47a.tgz  gcr.io/mutpanning-test/imagename (+1 more)  SUCCESS
```

You can manually administrate your Docker images on the Google Cloud under
console.cloud.google.com/gcr/images/<project-name>
By clicking on an image you can see the different versions and you can manually change the tags or delete images that are not needed anymore.

3. Prepare your data

MutPanning is designed to detect cancer driver mutations from aggregated whole-exome sequencing data. MutPanning needs two input files
1. A file listing the positions of all somatic mutations in a standard Mutation Annotation Format (MAF).
2. A file which contains the identifiers of all samples and organizes them into subcohorts (typically the cancer types).
Both of these files should be tab-delimited. Make sure that the files adhere exactly to the column headers listed below (case-sensitive), as MutPanning uses these column names to extract the data from these tables. The file can contain more columns or columns in a different order.

3.1 Sample annotation file

This file organizes samples into subcohorts. Samples in each subcohort are analyzed together for mutational significance. Further, MutPanning automatically analyzes all samples in the cohort together ("PanCancer" run) for mutational significance. This file should contain the following column headers.

ID The sample index used for internal reference in MutPanning. Typically, samples are numbered from 0 to max in this file. However, if you would like to compare intermediate files produced in different runs of MutPanning, you can choose arbitrary nonnegative integer numbers. Note, however, that these sample indices should be unique.

Sample The same sample identifier used in the mutation annotation file. Note that these sample identifiers should be unique. Avoid special characters.

Cohort The cohort name, in which the samples should be analyzed together for significance. This is typically the cancer type. However, you can also choose e.g. subtypes of cancer types or combine different cancer types together in order to enhance the significance.

3.2 Mutation Annotation File

This file lists the positions of all somatic mutations in your cohort. This file should follow the standard mutation annotation format and have the following colums.

Hugo_Symbol The nomenclature of the symbol in Hugo nomenclature (genenames.org).

Chromosome The chromosome, on which the mutation was found. Please use X and Y for the sex chromosomes and not 23 and 24.

Start_Position The start position of the mutation (Hg19).

End_Position The end position of the mutation (Hg19).

Strand The strand on which the mutation was detected (both 1/-1 or +/- nomenclature are fine).

Variant_Classification The functional class of the mutation (e.g. Silent, Missense_Mutations etc.). Please use the standard nomenclature used in MAF files. If in doubt, please see our exemplary data sets.

Variant_Type The type of the mutation, such as single base substitution (SNP), insersions (INS) or deletions (DEL). .). Please use the standard nomenclature used in MAF files. If in doubt, please see our exemplary data sets.

Reference_Allele The nucleotide expected in the Hg19 reference genome. Please note that if this mutation does not with the in the Hg19 genome, this mutation is ignored.

Tumor_Seq_Allele1 The alternative nucleotide1 found in the reads

Tumor_Seq_Allele2 The alternative nucleotide2 found in the reads . Typically, either of these nucleotides is the reference allele, whereas the other column contains the alternative read. Different callers handle differently how they assign these columns. MutPanning will first look in Tumor_Seq_Allele1 whether this equals the reference allele. If so, will take the change in the second Tumor_SEq_Allele column as alternative read.

Tumor_Sample_Barcode
The name of the tumor sample. These identifies have to be unique and be exactly the same as in the sample annotation file (case sensitive).

3.3 Test datasets

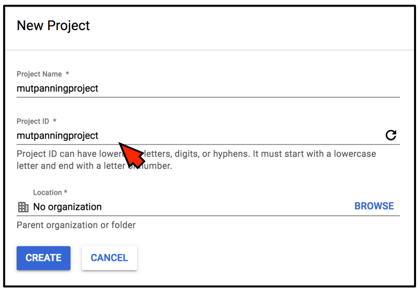
We prepared two datasets, which you can use to test MutPanning (mutpanning_testfiles.zip). Even if you do not intend to use these data for testing purposes, we encourage you to download these data, to familiarize yourself with the format and prepare your own data accordinling. The PanCancer dataset contains 12,004 samples. This is the dataset that was used in our manuscript. MutPanning needs approx. 24 hours to run on these data. The Melanoma dataset contains all 598 melanoma samples from our study cohort. MutPanning needs approx. 3 hrs to run on these data.

4. Run MutPanning on your data

If you made it until here, you already managed the largest challenges to configure your account and prepare your data. You are just two steps away from running MutPanning on your data.
You will need to run two bash scripts locally on your computer (start-mutpanning1.sh and start-mutpanning2.sh) to launch MutPanning. These scripts will take care of all communications with the Google Cloud. In particular, they will launch a Virtual Machine in the cloud, upload your data and launch the execution of MutPanning. Once your results are ready, you will need to execute a third bash script (stop-mutpanning.sh) to download your results and shut down the Virtual Machine.
In this chapter, we will give detailed step-by-step instructions on how to run these scripts. Note that you will not be required to write any bash scripts. However, you should be able to run a bash script. To launch a bash script you should first change into the directory in which the mutpanning scripts are stored (typically this would look like cd <path_to_scripts>/mutpanning_scripts/ in your terminal). Then type bash <script_name> <parameters separated by blank> into the command line to launch a bash script. If you are unfamiliar with running bash scripts in general, we found this quick introduction on how to run bash scripts very helpful to get a basic idea on what bash scripts are.
https://ryanstutorials.net/bash-scripting-tutorial/bash-script.php

4.1 Run MutPanning on your data

There are two main steps to launch MutPanning in the cloud.
The first script **start-mutpanning1.sh** is responsible to launch a Virtual Machine in the cloud and launch the Docker container. This script expects the following three parameters:
1. Project ID: The name / ID of the project under which you would like to launch MutPanning. This has to be exactly the project ID that you chose when creating the project.



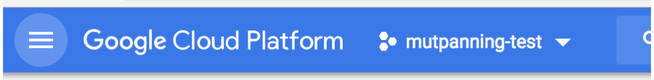
New Project

Project Name *
mutpanningproject

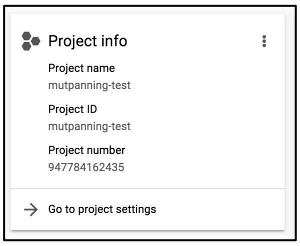
Project ID *
mutpanningproject
Project ID can have lowerc letters, digits, or hyphens. It must start with a lowercase letter and end with a letter o umber.

Location *
No organization
Parent organization or folder

Unless you chose a very common name, that had already been given in the cloud, or manually changed the project ID, the project ID is the same as the project name, displayed in the upper left corner.



If you are unsure about the project ID, you can always go to the home page of your google cloud account (click "Home" in the Navigation menu) and find this information in the upper left corner under "Project Info".



Project info

Project name
mutpanning-test

Project ID
mutpanning-test

Project number
947784162435

Go to project settings

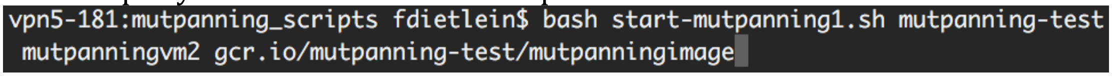
Make sure that you configured this project according to our descriptions in the prerequisites chapter. In particular, make sure you requested enough quotas (24 CPUs, 60 GB memory, 2000 GB of SSD size).

2. Instance name: This can be an arbitrary name, but make sure that you remember it because you will need this name in the subsequent steps. Also make sure that this name contains letters and numbers only, in particular no special characters, no blanks, no underscores. Although Google Cloud supports some of these names, they might interfere with the bash scripts. Further, make sure that you have not launched an instance under the same cloud previously, as the script identifies the Virtual Machine based on this name only.

3. Path to the Docker image: Typically, you would a path to one of the docker images that we compiled for you and deposited into the Google Container Repository. However, if you rather choose to run an image, which you compiled yourself in the prerequisites chapter, enter the path here. Also make sure that you include the tag into the path (if there is any), in order to launch the right version. Most paths look like

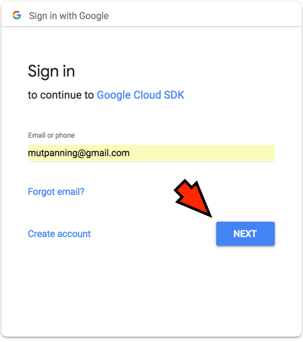
gcr.io/<project-name>/<image-name>:<tag>

An exemplary execution of this this script is shown below:



```
vpn5-181:mutpanning_scripts fdietlein$ bash start-mutpanning1.sh mutpanning-test
 mutpanningvm2 gcr.io/mutpanning-test/mutpanningimage
```

The script will first open a Google Chrome browser window



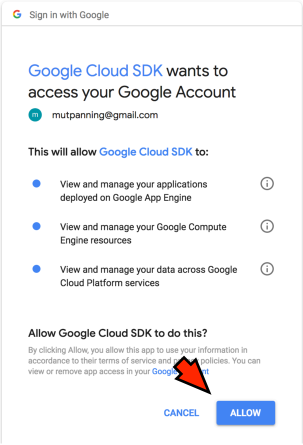
Sign in with Google

Sign in
to continue to Google Cloud SDK

Email or phone
mutpanning@gmail.com

Forgot email?

Create account



Sign in with Google

Google Cloud SDK wants to access your Google Account
mutpanning@gmail.com

This will allow Google Cloud SDK to:

- View and manage your applications deployed on Google App Engine
- View and manage your Google Compute Engine resources
- View and manage your data across Google Cloud Platform services

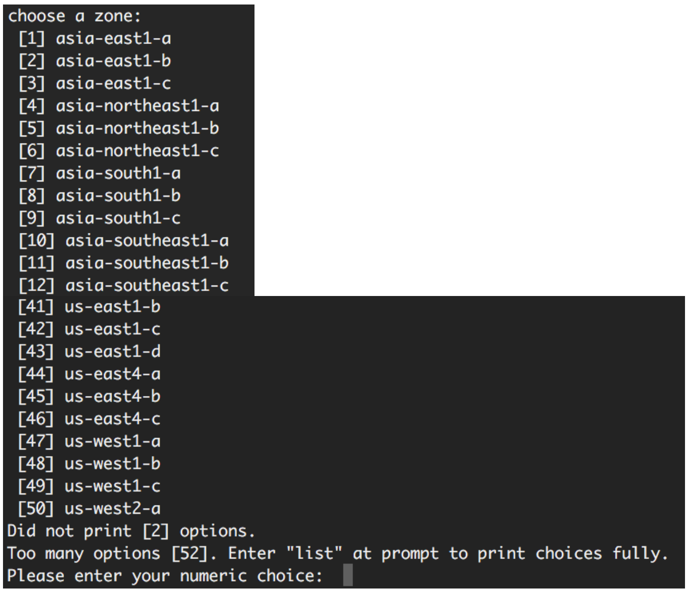
Allow Google Cloud SDK to do this?

By clicking Allow, you allow this app to use your information in accordance to their terms of service and [illegible] policies. You can view or remove app access in your Googl[illegible]nt

Which asks you to login into your account. If you are using multiple accounts, please make sure that you choose the account under which you created the project that you specified in the command line. The script will ask you for your permission for Google Cloud SDK to access your account. Click on Allow to continue.

Next, the script will ask you in which zone you would like to launch the Virtual Machine. Make sure that you still have a quota of 24 CPUs, 60GB memory and

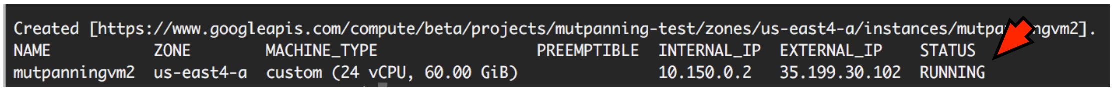
2000GB SSD space disk in this zone available. Typically this is the zone in which you requested the quota extension in the prerequisites chapter. Please note that it usually does not make any difference which subzone (labeled by a-c) you launch the instance.



```
choose a zone:
 [1] asia-east1-a
 [2] asia-east1-b
 [3] asia-east1-c
 [4] asia-northeast1-a
 [5] asia-northeast1-b
 [6] asia-northeast1-c
 [7] asia-south1-a
 [8] asia-south1-b
 [9] asia-south1-c
 [10] asia-southeast1-a
 [11] asia-southeast1-b
 [12] asia-southeast1-c
 [41] us-east1-b
 [42] us-east1-c
 [43] us-east1-d
 [44] us-east4-a
 [45] us-east4-b
 [46] us-east4-c
 [47] us-west1-a
 [48] us-west1-b
 [49] us-west1-c
 [50] us-west2-a
Did not print [2] options.
Too many options [52]. Enter "list" at prompt to print choices fully.
Please enter your numeric choice:
```

Enter your selection as a number and press enter to continue.

The script will next spin up a Virtual Machine and launch the Docker. Please note that this can take a couple of minutes and do not terminate the script while you are waiting. As the final result you should see the following screen, which indicates that the Virtual Machine was launched successfully and is RUNNING



```
Created [https://www.googleapis.com/compute/beta/projects/mutpanning-test/zones/us-east4-a/instances/mutpanningvm2].
NAME           ZONE        MACHINE_TYPE                 PREEMPTIBLE  INTERNAL_IP  EXTERNAL_IP    STATUS
mutpanningvm2  us-east4-a  custom (24 vCPU, 60.00 GiB)               10.150.0.2   35.199.30.102  RUNNING
```

If this is not the case (e.g. the script terminates prematurely or the status is FAILURE), the most typical error is that you did not have enough quotas in the zone you requested or used you used a bad name for the instance. In this case, we suggest that you check your quotas under

console.cloud.google.com/iam-admin/quotas

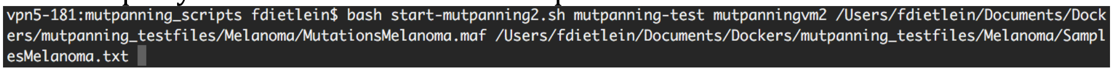
and extend them so that you have 24 CPUs available, 60 GB memeory and 2000 GB SSD available in the zone that you requested. If you already have multiple VMs running in the cloud, please note that there are also global quotas on how many resources you are able to use. Even though you might be under the limit in each zone individually, you might have exceeded these global limits. To increase your global quotas, select zone "global" instead of a local zone.

The second script **start-mutpanning2.sh** uploads your data into the cloud and launches MutPanning on the Virtual Machine, after they are copied. This script can only be successfully executed after both the Virtual Machine and the Docker image have been successfully launched. Hence, we suggest to wait approximately 5 minutes after the completion of the first script before you launch start-mutpanning2.sh

This script expects the following parameters:

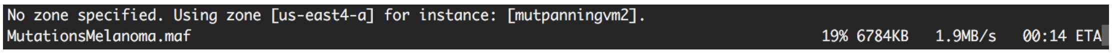
1. Project ID: Please use exactly (case-sensitive) the same name that you used in the first step.
2. Instance name: Please use exactly (case-sensitive) the same name that you used in the first step.
3. Full path to the mutation annotation file: This is the *.maf file that contains all your mutations. You may either use your own data or our test datasets (<full_path>/mutpanning_testfiles/Melanoma/MutationsMelanoma.maf) or (<full_path>/mutpanning_testfiles/PanCancer/MutationsComplete.maf).
4. **Full path to the sample annotation file.** This *.txt file that contains the identifiers of all your samples and associates them into subcohorts. You may either use your own data or our test datasets
(<full_path>/mutpanning_testfiles/Melanoma/SamplesMelanoma.txt) or (<full_path>/mutpanning_testfiles/PanCancer/SamplesComplete.txt).

An exemplary execution of this bash script can be seen below



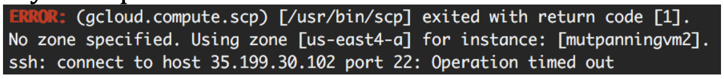
```
vpn5-181:mutpanning_scripts fdietlein$ bash start-mutpanning2.sh mutpanning-test mutpanningvm2 /Users/fdietlein/Documents/Dock
ers/mutpanning_testfiles/Melanoma/MutationsMelanoma.maf /Users/fdietlein/Documents/Dockers/mutpanning_testfiles/Melanoma/Sampl
esMelanoma.txt
```

The script will first ask you to login and use your google account (browser window). Shortly after this you should see the progress of the upload of your data to the cloud.



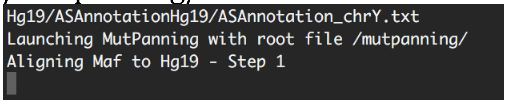
```
No zone specified. Using zone [us-east4-a] for instance: [mutpanningvm2].
MutationsMelanoma.maf                                                          19% 6784KB   1.9MB/s   00:14 ETA
```

Make sure that your internet connection is fast and stable during the upload. Note that it might take a couple of minutes to upload your files. During this time you should not close the connection.
If you experience an error like this



```
ERROR: (gcloud.compute.scp) [/usr/bin/scp] exited with return code [1].
No zone specified. Using zone [us-east4-a] for instance: [mutpanningvm2].
ssh: connect to host 35.199.30.102 port 22: Operation timed out
```

during the execution of the this script, this indicates that the Virtual Machine or the Docker have not been fully loaded yet. Note that the error message does not always look like this. Instead of a time out, you might also receive an ssh key error (“It is possible that your SSH key has not propagated to the instance yet. Try running this command again.”). All these messages do not indicate that there is a problem with your Virtual Machine, you just tried to log into your Virtual Machine too early, so that it could not handle your log in request for various reasons. In all this cases, just give it a little bit more time to load and try again the execution of the same script with exactly the same parameters after 2 minutes. This should usually solve the problem. Another typical error is that your script will repetitively ask you for a zone (it should not ask for a zone at all). This indicates that it cannot find the instance name that you specified in the parameters. In this case, cancel the script and make sure that you use exactly the same name (context-sensitive) that you used to launch a Virtual Machine in the first script.
As soon as you see the message “Launching MutPanning with root file /mutpanning/” it is safe to close the connection. This should look like below.



```
Hg19/ASAnnotationHg19/ASAnnotation_chrY.txt
Launching MutPanning with root file /mutpanning/
Aligning Maf to Hg19 - Step 1
```

The script is now running in the cloud and even if you close the connection, the script will continue to run. Depending on the data size the script typically runs between 3 hours (~500 samples) to 24hrs (12,000 samples).

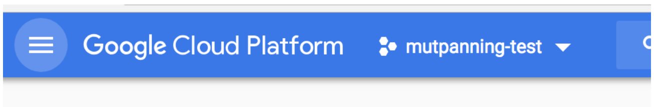
4.2 Terminate MutPanning and download the results

You can use the script **stop-mutpanning.sh** to regularly check whether your MutPanning results are ready. If so, this script will automatically download the results and shut down the Virtual Machine. If not, you will be notified that the results are not ready, but the Virtual Machine will continue to work on your data. Hence, it is safe to execute this script regularly in order check on the status of your results.

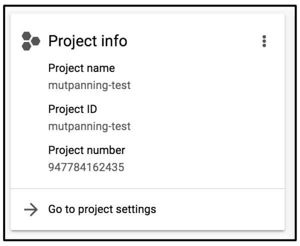
Please note that you will be billed by the total time that the Virtual Machine runs, independent of whether MutPanning has already terminated or not. So in order to safe costs, we recommend to download and shut down the Virtual Machine shortly after the results are ready. If you still need intermediate files you will still be able to download them later by restarting the Virtual Machine.

The script **stop-mutpanning.sh** expects the following parameters

1. Project ID: Please use exactly (case-sensitive) the same ID that you used when launching the Virtual Machine. If you are unsure, you can usually find the name in the top left corner in the console



in case you used the same project ID as the project name. If not, you will be able to find the project ID in the Project Info box, provided on the home page of your google console.


Project info

Project name
mutpanning-test

Project ID
mutpanning-test

Project number
947784162435

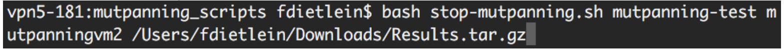
→ Go to project settings

2. Instance name: Please use exactly (case-sensitive) the same name that you used in the first step to launch MutPanning. If you are unsure, you can find the names of all instances in your project under Navigation Menu (icon in top left next to "Google Cloud Platform") -> Compute Engine -> VM instances or under

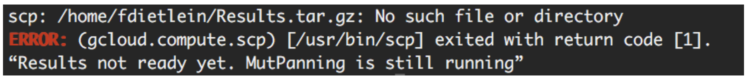
console.cloud.google.com/compute/instances

3. Full path to the output file. The full path to the output file, to which you would like to download your results on your local computer (e.g.: /Users/<user_name>/Downloads/Results.tar.gz). Please include the full path as well as the full file name (*.tar.gz archive). Please make sure that the file, to which you would like to download your results, does not already exist, as the script will terminate immaturely otherwise.

An exemplary execution of this script is shown below.

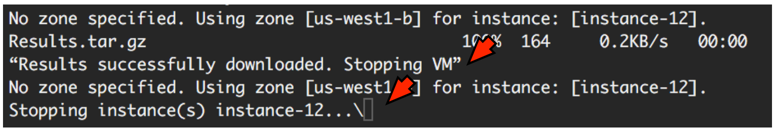

```
vpn5-181:mutpanning_scripts fdietlein$ bash stop-mutpanning.sh mutpanning-test m
utpanningvm2 /Users/fdietlein/Downloads/Results.tar.gz
```

The script will first ask you to login and ask or your permission to access your account (browser window). Please make sure that you log in under the same account under which you launched the MutPanning run. If your results are not ready yet you will see a message like this.



```
scp: /home/fdietlein/Results.tar.gz: No such file or directory
ERROR: (gcloud.compute.scp) [/usr/bin/scp] exited with return code [1].
"Results not ready yet. MutPanning is still running"
```

In this case, just execute the same script with the same parameters a few hours later. If your results are ready for download, you will see an output like that



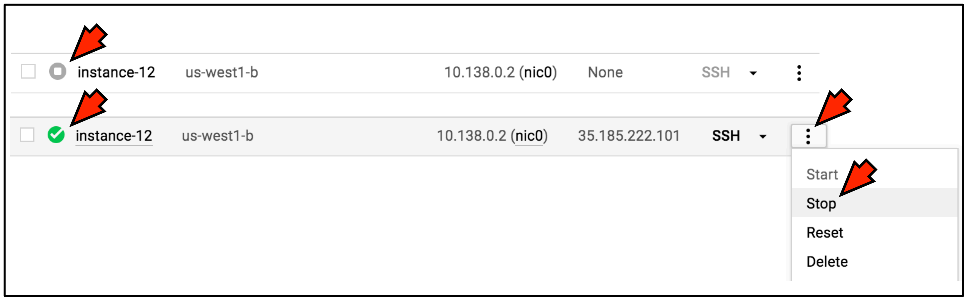
```
No zone specified. Using zone [us-west1-b] for instance: [instance-12].
Results.tar.gz                                 100%  164     0.2KB/s   00:00
"Results successfully downloaded. Stopping VM"
No zone specified. Using zone [us-west1-b] for instance: [instance-12].
Stopping instance(s) instance-12...\
```

which shows the status of the download of your results, as well as that the Virtual Machine is being shut down. Please note that it might take 1-2 minutes to successfully shut down your machine. Please do not terminate the script while the machine is being shut down. Unzip the result folder and you should have a separate result file for each cancer type /cohort that you specified in the sample annotation file. Further you will find a file ("PanCancer") which analyzes the mutational significance of all samples together.

After successful download of the manuscript, the Virtual Machine should be stopped. However, we recommend a quick manual check in the console under

console.cloud.google.com/compute/instances

to avoid unnecessary cost. If for some reason the machine is still running (e.g. ssh connection was interrupted while stopping the machine), you can stop it manually.



□ ○	instance-12	us-west1-b	10.138.0.2 (nic0)	None	SSH ▾
□ ✓	instance-12	us-west1-b	10.138.0.2 (nic0)	35.185.222.101	SSH ▾

In case you would like to additionally download any intermediate files from the MutPanning run, you can restart the machine and manually download all intermediate files. If you need to release the resources for new MutPanning runs, you can permanently delete the machine from the cloud.