

OPS 813: Cloud Computing

-Today's plan:

0) Textbook Update

1) Homework #2: Time to get another badge

2) Case #1: Stock Market data repository & initial analysis

3) Ingesting Data & Data Analytics on your instance

Google Cloud Study Jam – Qwik Labs (HW2)

- 1) We are going to get you started on your next Quest:
Baseline: Data, ML, AI.
 - 2) If you have already signed into your Qwiklabs accounts, please log out of it.
 - 3) Open this URL in a new *incognito* window:
<https://google.qwiklabs.com/quests/34?qlcampaign=1s-moraga-5569>
 - 4) Click on the "**Sign in**" button and enter your credentials.
 - 5) Check the "**Credits and Subscriptions**" section of account.
You should have received 30-day pass after completing the process above.
- *** Remember to please follow these instructions:
https://support.google.com/qwiklabs/answer/9222527?hl=en&ref_topic=9139328 and email me (ns27) the link to your public profile.

Quest Outline

HANDS-ON LAB

[Introduction to SQL for BigQuery and Cloud SQL](#)

In this lab you will learn fundamental SQL clauses and will get hands on practice running structured queries on BigQuery and Cloud SQL.



1h 15m

Introductory

1 Credit



HANDS-ON LAB

[BigQuery: Qwik Start - Console](#)

This lab shows you how to query public tables and load sample data into BigQuery using the Web UI. Watch the short videos [Get Meaningful Insights with Google BigQuery](#) and [BigQuery: Qwik Start - Qwiklabs Preview](#).



30m

Introductory

1 Credit



OR

HANDS-ON LAB

[BigQuery: Qwik Start - Command Line](#)

This hands-on lab shows you how to query public tables and load sample data into BigQuery using the

HW #2

For your homework #2:

- Please obtain a badge for completing the **Baseline: Data, ML, AI** Quest (all the labs in this Quest).
- Read Chapters 3, 4, 5 of textbook

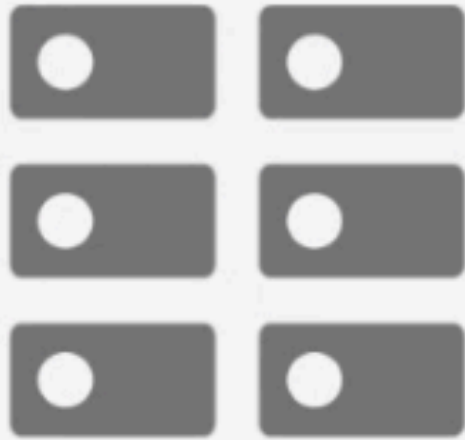


Case #1

For your case #1:

- Please download the stock data from [this link](#) to your instance.
- Then (1) find all stocks that are currently below the 5th percentile of their historical prices, (2) find all stocks that are currently above the 95th percentile of their historical prices, and (3) as a group choose 6 or 7 stocks that seem interesting to buy and hold for a month.

Data



Unstructured

Collections of files



Structured

Databases ...

Multi-Regional



Geo-redundant
99.95% SLA
2.6¢/GB



Millisecond access
Consistent API



Content storage
and delivery,
business continuity



For highest availability
of frequently
accessed data



Video
Multimedia
Business continuity

Regional



Single-region
99.9% SLA
2.0¢/GB



Millisecond access
Consistent API



Store data and run data
analytics within a region



For data accessed
frequently
within a region



Transcoding
Data analytics
General compute

Nearline



Infrequent Access
99% SLA
1.0¢/GB



Millisecond access
Consistent API



Store infrequently
accessed content



For data accessed
less than once a month



Backup
Long-tail content
Rarely accessed docs

Coldline



Cold Archival
99% SLA
0.7¢/GB



Millisecond access
Consistent API



Archive storage, rare
disaster recovery



Data accessed
less than once a year



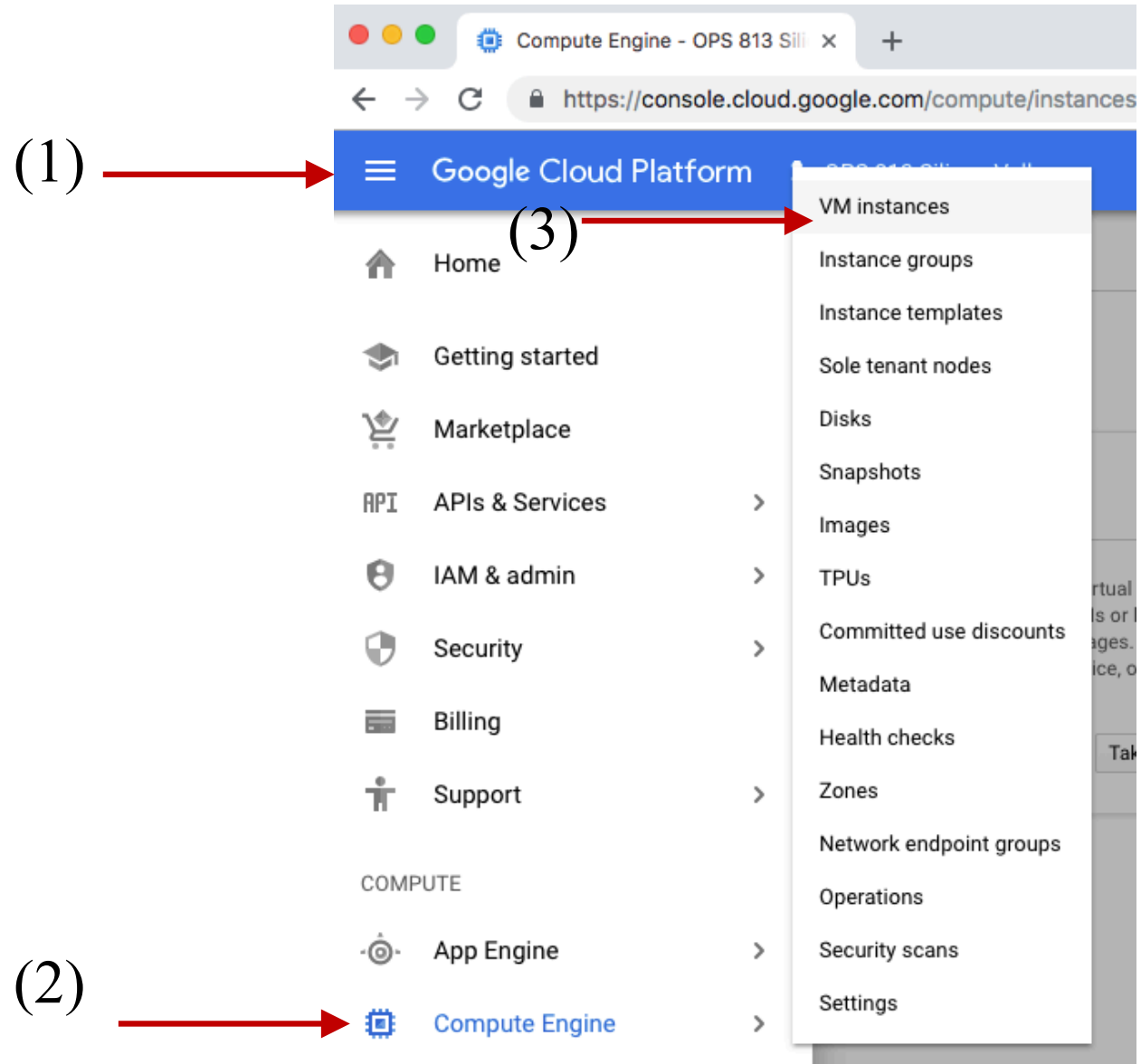
Archive
Source file backup
Disaster recovery

	<i>Bandwidth (assuming 100% utilization)</i>					
<i>Data Size</i>	<i>1 Mbps</i>	<i>10 Mbps</i>	<i>100 Mbps</i>	<i>1 Gbps</i>	<i>10 Gbps</i>	<i>100 Gbps</i>
1 GB	3 hrs	18 mins	2 mins	11 secs	1 sec	0.1 secs
10 GB	30 hrs	3 hrs	18 mins	2 mins	11 secs	1 sec
100 GB	12 days	30 hrs	3 hrs	18 mins	2 mins	11 secs
1 TB	124 days	12 days	30 hrs	3 hrs	18 mins	2 mins
10 TB	3 years	124 days	12 days	30 hrs	3 hrs	18 mins
100 TB	34 years	3 years	124 days	12 days	30 hrs	3 hrs
1 PB	340 years	34 years	3 years	124 days	12 days	30 hrs
10 PB	3404 years	340 years	34 years	3 years	124 days	12 days
100 PB	34048 years	3404 years	340 years	34 years	3 years	124 days



Last time: Setup your own instance

Setup your own instance



(4) Then click on Create in the popup window.

Setup your own instance

- Now click on ‘Create new instance’.
- Name your instance, select zone as ‘us-west1 (Oregon)’.
- Choose your ‘machine type’. (I chose 1vCPU for now).
- Select your boot disk as ‘Ubuntu 16.04 LTS’.
- Change the Size of the Boot disk to 30 GB.
- Allow full access to all Cloud APIs.
- Under the firewall options tick both ‘http’ and ‘https’ (very important).
- Click **Create**.


Name ?
cloudexample

Region ?
us-west1 (Oregon)

Zone ?
us-west1-b

Machine type
Customize to select cores, memory and GPUs.
1 vCPU 3.75 GB memory [Customize](#)

Container ?
☐ Deploy a container image to this VM instance. [Learn more](#)

Boot disk ?
 New 30 GB standard persistent disk
Image
Ubuntu 16.04 LTS [Change](#)

Identity and API access ?
Service account ?
Compute Engine default service account

Access scopes ?
☐ Allow default access
☒ Allow full access to all Cloud APIs
☐ Set access for each API

Firewall ?
Add tags and firewall rules to allow specific network traffic from the Internet
☒ Allow HTTP traffic
☒ Allow HTTPS traffic

Setup your own instance

<input type="checkbox"/>	Name ^	Zone	Recommendation	Internal IP	External IP	Connect
<input type="checkbox"/>	<input checked="" type="checkbox"/> clouDEXample	us-west1-b		10.138.0.3 (nic0)	None	SSH ▾ ⋮

Make external IP address static
(a)

Looking for
OFF button?

(b)

(c)

The image shows a screenshot of the Google Cloud Platform (GCP) interface. At the top, there is a blue header bar with the text "Google Cloud Platform" and "OPS 813 Silicon Valley". Below this is a navigation menu with various icons and labels: Home, Kubernetes Engine, Cloud Functions, STORAGE, Bigtable, Datastore, Firestore, Storage, SQL, Spanner, Memorystore, Filestore, NETWORKING, VPC network, and Network services. A red arrow points from the text "(a)" to the "VPC network" label. Another red arrow points from the text "(b)" to the "Network services" label. A third red arrow points from the text "(c)" to the "VPC networks" sub-menu item, which is expanded to show "VPC networks", "External IP addresses", and "Firewall rules".

Setup your own instance

1) Click on type for your VM instance and choose "Static". Then choose a static IP address, e.g., "myaddress".

<input type="checkbox"/>	Name	External Address	Region	Type ▾	Version	In use by	Network Tier ?	Labels
<input type="checkbox"/>	anything	35.203.171.61	us-west1	Static	IPv4	⚠ None	Premium	
	—	35.185.193.0	us-west1	<div>Static Ephemeral</div>	IPv4	VM instance <u>cloudexample</u> (Zone b)		

2) Let's add firewall rules to control who gets into the instance.

(a) →

(b) →

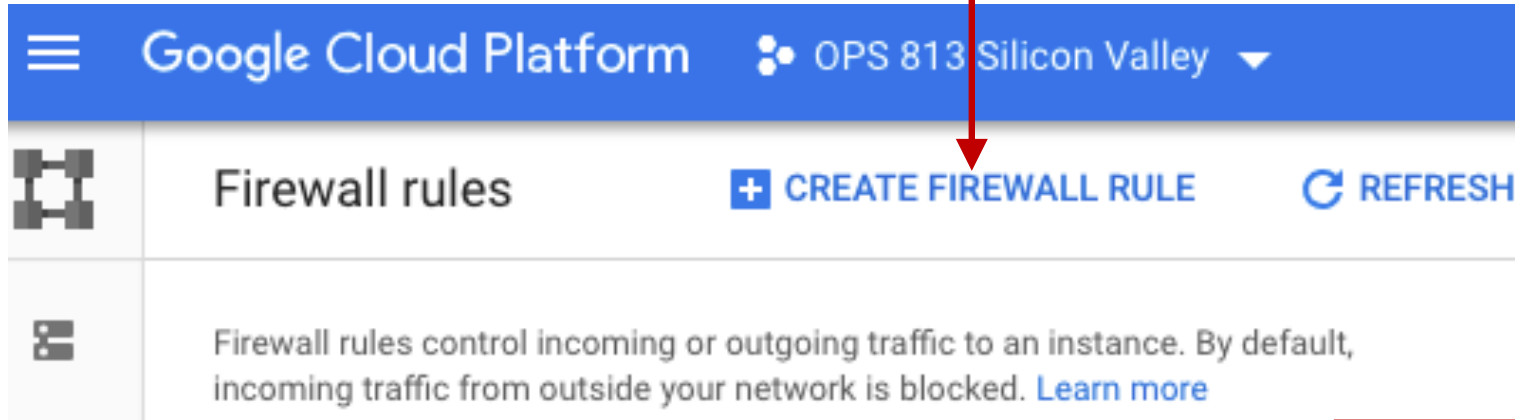
(c) →

The screenshot shows the Google Cloud Platform navigation menu. The menu is open, displaying various services. A red arrow labeled (a) points to the menu icon. A red arrow labeled (b) points to the 'VPC network' option. A red arrow labeled (c) points to the 'Firewall rules' option in the expanded 'VPC network' submenu.

- Google Cloud Platform OPS 813 Silicon Valley
- Home
- Memorystore
- Filestore
- NETWORKING
 - VPC network
 - VPC networks
 - External IP addresses
 - Firewall rules
 - Routes
 - VPC network peering
 - Shared VPC
 - Network services
 - Hybrid Connectivity
 - Network Service Tiers
 - Network Security

Setup your own instance

1) Click on Create Firewall Rule



2) Set up the firewall rule according to this screenshot:

And then click Create.

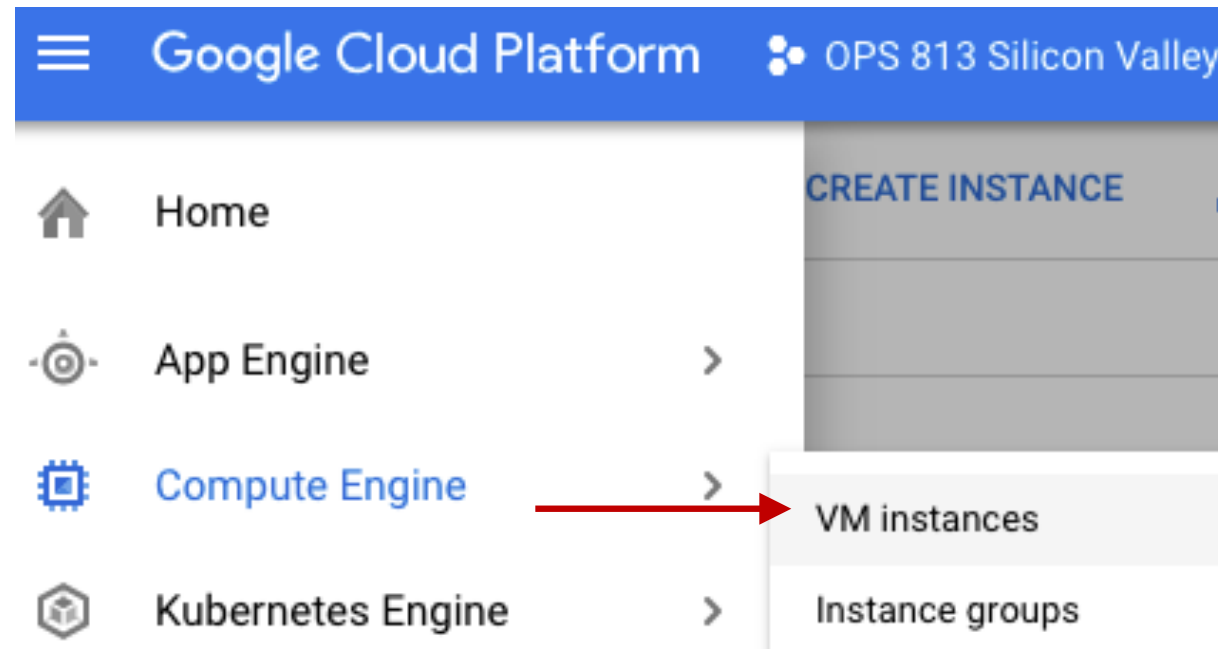
The screenshot shows the configuration form for a Firewall rule. The form includes the following fields and options:

- Network**: default
- Priority**: 1000 (Priority can be 0 - 65535 [Check pri](#))
- Direction of traffic**: ☒ Ingress, ☐ Egress
- Action on match**: ☒ Allow, ☐ Deny
- Targets**: All instances in the network
- Source filter**: IP ranges
- Source IP ranges**: 0.0.0.0/0 (with a close button)
- Second source filter**: None
- Protocols and ports**: ☐ Allow all, ☒ Specified protocols and por
- tcp**: ☒ tcp : 5000

Red arrows point to the 'All instances in the network' target, the '0.0.0.0/0' source IP range, and the 'Specified protocols and ports' option.

Setup your own instance

1) Now lets **use** the instance




2) Click on Open in browser window:



Setup your own instance

1) Remember to turn the instance **OFF** after you are done or you will lose up all your cloud credits.

<input type="checkbox"/>	Name ^	Zone	Recommendation	Internal IP	External IP	Connect	
<input type="checkbox"/>	<input checked="" type="checkbox"/> <u>cloudexample</u>	us-west1-b		10.138.0.3 (nic0)	<u>35.185.193.0</u> ↗	SSH ▾	<div><div>⋮</div><div>Start</div><div>Stop</div><div>Reset</div><div>Delete</div><div>View network c</div><div>View logs</div></div>



Software on the machine:

Using sudo, apt-get, wget, bash

```
ns27@cloudexample:~$ sudo apt-get update
Hit:1 http://us-west1.gce.archive.ubuntu.com/ubuntu xenial InRelease
Get:2 http://us-west1.gce.archive.ubuntu.com/ubuntu xenial-updates InRelease [109 kB]
Get:3 http://us-west1.gce.archive.ubuntu.com/ubuntu xenial-backports InRelease [107 kB]
Hit:4 http://archive.canonical.com/ubuntu xenial InRelease
Hit:5 http://security.ubuntu.com/ubuntu xenial-security InRelease
Fetched 216 kB in 0s (384 kB/s)
Reading package lists... Done
```

If you visit <https://repo.continuum.io/archive/> you will see there are many versions of Anaconda for various platforms.

Magic commands:

```
wget https://repo.continuum.io/archive/Anaconda3-5.3.1-Linux-x86_64.sh
```

```
bash Anaconda3-5.3.1-Linux-x86_64.sh
```

```
installation finished.
Do you wish the installer to initialize Anaconda3
in your /home/ns27/.bashrc ? [yes|no]
[no] >>> yes
```

Configure Jupyter Notebook Server

Make use of your configuration file:

```
source ~/.bashrc
```

Generate a Jupyter configuration file:

```
jupyter notebook --generate-config
```

Choose a password:

```
jupyter notebook password
```

Now edit the file:

```
nano ~/.jupyter/jupyter_notebook_config.py
```

```
# The string should be of the form type:salt:hashed-password.
c.NotebookApp.password = 'sha1:dff6d7b36dd4:71512fb17fa6b6c72e'
## The port the notebook server will listen on.
c.NotebookApp.port = 5000
## The IP address the notebook server will listen on.
c.NotebookApp.ip = '*'
```

Workflow on Jupyter Notebook

Let's run Jupyter

```
jupyter-notebook --no-browser --port=5000
```

Try it out.

Let's upload our mtg02.zip file to our instance.

Let's uncompress it...how? Hint: `sudo apt-get install zip`
(this installs zip and unzip on your instance)

Want Jupyter notebook to run in background?
Use `&` at the end.

Want to run Jupyter even if you logout? `nohup`

Job management on your instance

jobs

ps -ef

grep

And pipes |

bg, fg, %number

Book

Code in book:

```
git clone https://github.com/GoogleCloudPlatform/data-science-on-gcp
```

Case #1

Now let's get together in groups corresponding to your case groups.

- Create a new instance for case 1.
- Enable your teammates to log into the instance too.
- Also enable me (ns27@stmarys-ca.edu) to log in.
- install Jupyter and set it up so that you can start the case.
- To download the data directly to your cloud instance use <https://github.com/Kaggle/kaggle-api>

Case #1

As a reminder, please remember to turn off your instance when you are not using it in order to conserve your Google Cloud credits.

Please hand in three parts for your Case #1:

Part 1:

- a) Describe the challenges you faced in working in the Cloud and using the dataset.
- b) Describe the things you learned about working in the Cloud for this case and in working with the dataset.
- c) Provide some analytical justification for the 6 or 7 stocks you picked.

Part 2:

- Provide the Python source code for your analytics as an appendix.

Part 3:

- Please grant me access to the instance you used for the case.

****** In addition, for the group that is presenting each week, please keep your presentation less than 15 minutes using a slide deck.