

Google Cloud Product

Who am I?

Sarah

Lead Data Scientist at zData Inc.

Contracted out to Disney for predictive maintenance

Previously:

- Sr. Statistician at Idaho National Lab
- Statistical Programmer II at The Mayo Clinic

Overview

- Cloud?
- Set up a compute unit
- Run data lab
- Query (big) data
- How to sample subset
- Exploratory Analysis

Working in the Cloud

What is the cloud?

Working in the Cloud

What is the cloud?

A remote server farm that ____ has to maintain, update, and secure.

There are several ____ :

- Azure (Microsoft)
- AWS (Amazon)
- GCP (Google)

Working in the Cloud

Why the cloud?

Working in the Cloud

Why the cloud?

- Scalability
- Ease of access
- Security of Data
- Stability

Compute Engine / Virtual Machine instances

What is CE / VM

Infrastructure as a Service (IaaS) component of Google Cloud Platform (GCP)

Same infrastructure that runs Google's search engine, Gmail, and YouTube

Compute Engine / Virtual Machine instances

2 ways to create in GCP:

1. Through the UI
2. Through the Google Cloud Shell

Compute Engine / Virtual Machine instances

2 ways to create in GCP:

1. Through the UI
2. **Through the Google Cloud Shell**

Compute Engine / Virtual Machine instances

Through the Google Cloud Shell

Click on the  at the top right of the screen

Activate Cloud Shell

Then a command line should pop up at the bottom of the screen

Type: `gcloud compute zones list`

Why did I do this?

Compute Engine / Virtual Machine instances

In cloudshell type:

```
datalab create <NAME> --zone <ZONE>
```

Replace **<NAME>** with what you want to call your VM

Replace **<ZONE>** with the zone you picked

Pick a zone for your compute engine, preferably near your location

I picked **us-west2-a** since I am in Boise, ID

```
datalab create sarahs-vm --zone us-west2-a
```

Compute Engine / Virtual Machine instances

It will take a few minutes to set up

It will ask “Do you want to continue?”

Tell it “Y”

It will provide an output similar to this:

Why did I scribble out part of the image?

```
sewing@cloudshell:~ (divine-vehicle-219619)$ datalab create sarahs-vm --zone us-west2-a
Creating the repository datalab-notebooks
Creating the instance sarahs-vm
Created [https://www.googleapis.com/compute/v1/projects/divine-vehicle-219619/zones/us-west2-a/instances/sarahs-vm]
Connecting to sarahs-vm.
This will create an SSH tunnel and may prompt you to create an rsa key pair. To manage the connection,
Waiting for Datalab to be reachable at http://localhost:8081/
This tool needs to create the directory [/home/sewing/.ssh] before
being able to generate SSH keys.

Do you want to continue (Y/n)? Y

Generating public/private rsa key pair.
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/sewing/.ssh/google_compute_engine.
Your public key has been saved in /home/sewing/.ssh/google_compute_engine.pub.
The key fingerprint is:
SHA256:W0Yh0e8FM6uvenql2thw5hfIFQgBeYoVWc3L2eKViNc sewing@cs-6000-devshell-vm-4fa55a87-3
The key's randomart image is:
+--[RSA 2048]-----+
|      .   .          |
|     o .  o          |
|    .+..+..          |
|   .+..+..          |
|  .+..+..          |
| .+..+..          |
|+.+..+..          |
|o.o+o          |
|O+o          |
|=*o          |
+----[SHA256]-----+
Updating project ssh metadata...:Updated [https://www.googleapis.com/compute/v1/projects/divine-vehicle-219619/zones/us-west2-a/instances/sarahs-vm]
Updating project ssh metadata...done.
Waiting for SSH key to propagate.

The connection to Datalab is now open and will remain until this command is killed.
Click on the *Web Preview* (square button at top-right), select *Change port > Port 8081
```

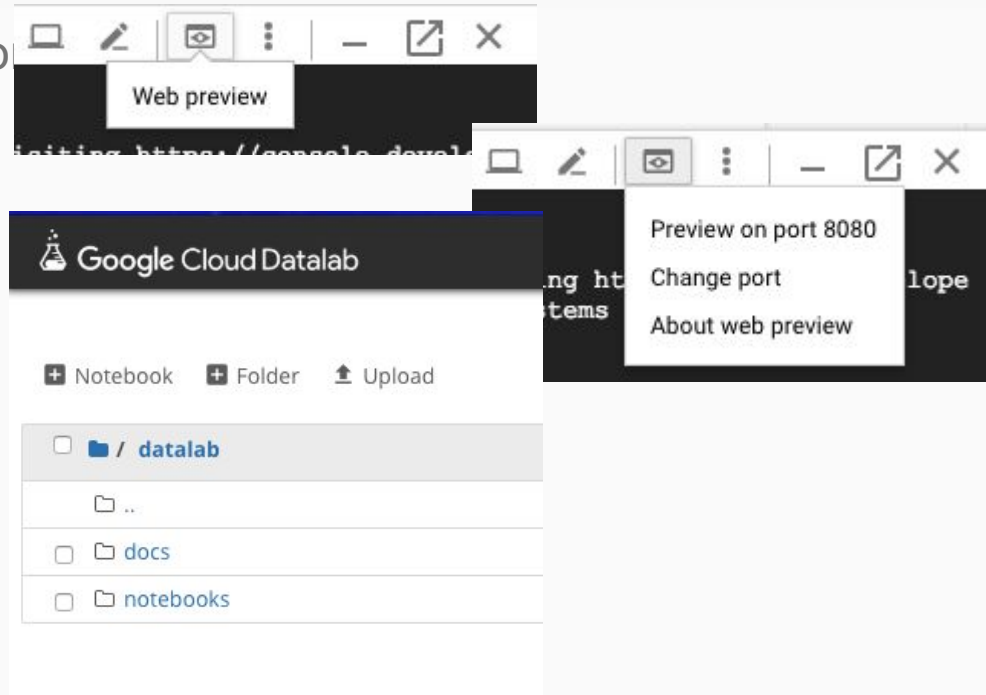
Compute Engine / Virtual Machine instances

At the right top of the cloud shell click on preview icon.

Pick **change port**

Tell it port **8081**

Then click on **Change and Preview**



Compute Engine / Virtual Machine instances

Other important command, to restart compute engine:

```
datalab connect <instance NAME>
```