



Google Cloud Platform

GCP Crash Course for p1rcc Hackathon

May 19, 2018
Clayton Mellina

Topics

- Basic Concepts and Tools
- Console Tour
- Google Cloud Storage
- Google Compute Engine

Basic Concepts and Tools

Interacting with GCP

- Sign up at <https://cloud.google.com/>
- There are 4 primary ways to interact with GCP:
 - Google Cloud Console: <https://console.cloud.google.com/>
 - The **gcloud** command line tool, Google Cloud SDK: <https://cloud.google.com/sdk/downloads>
 - Language-specific client libraries: <https://cloud.google.com/apis/docs/cloud-client-libraries>
 - HTTP APIs

Authenticating with gcloud

Show your authenticated accounts

```
gcloud auth list
```

Authenticate

```
gcloud auth login
```

Authenticate in a browserless environment

```
gcloud auth login --no-launch-browser
```

Switch accounts (if you have multiple)

```
gcloud config set account name@example.com
```

GCP Projects

- Projects are the basic organizational unit of GCP
 - Resources are associated with the projects that create them.
 - Default resource access controls provide convenient settings for projects and their members.
 - GCP quota is managed on a per-project basis.
- Many resource creation/access tasks are scoped to a project

Setting your project

- In Console



- In gcloud

```
# Check default project setting
```

```
gcloud config get-value project
```

```
# Set default project
```

```
gcloud config set project plrcc-example
```

Regions

- Resources are organized into **regions** with names like us-west1, us-central2, etc.
- Here is an overview of regions:
<https://cloud.google.com/compute/docs/regions-zones/>
- **Zones** are a grouping of resources in a region and have an additional letter in their name, e.g. us-west1-b
- Some resources are regional (e.g. static IP addresses) while some are zonal (e.g. VMs, disks).

Enabling APIs

- Almost all services and APIs need to be enabled for your project before you can use them.
- You can do this directly in Console.

Console Tour

Stops along the Console Tour

- Project selector
- Side panel + pins
- Search
- Notifications
- Cloud Shell
- Enabling an API
- IAM
- Quota

What has already been done

Before you begin

1. Select or create a GCP project.

[GO TO THE MANAGE RESOURCES PAGE](#)

2. Make sure that billing is enabled for your project.

[LEARN HOW TO ENABLE BILLING](#)

Hackathon setup

- Projects were created and linked to SVAI's billing account
- You were given project Editor permission
- Only members on your project can create or use your project's resources
- If you still are curious: <https://github.com/pumpikano/google-cloud-hackathon>

Google Cloud Storage

GCS

- Basic organizational unit is a **bucket** which must have a globally unique name.
- No storage limits.
- Example path:
 - `gs://p1rcc/somatic/results/variants/somatic.snvs.snpeff.vcf.gz`
- Paths look like there is a directory hierarchy, but storage is actually flat.
 - This means you cannot “move” a folder—this is really a copy, so will take time proportional to the amount of data.

GCS in Console

- Canonical URL is <https://console.cloud.google.com/storage/browser/<bucket>>
 - E.g. <https://console.cloud.google.com/storage/browser/p1rcc/>

<input type="checkbox"/> Name	Default storage class [?]	Location	Lifecycle [?]	Labels [?]	Requester pays [?]	
<input type="checkbox"/> p1rcc	Multi-Regional	US	<u>None</u>		<input checked="" type="radio"/> Off	⋮
<input type="checkbox"/> p1rcc-staging	Multi-Regional	US	None		<input checked="" type="radio"/> Off	⋮

Command line

- GCS has its own command line tool: **gsutil**
- Reference docs here: <https://cloud.google.com/storage/docs/gsutil>
 - Or: **gsutil help <subcommand>**
- Subcommands mirror the Unix file commands, e.g. **cp**, **mv**, **mkdir**, etc.

```
# Recursively copy objects
```

```
gsutil -m cp -r gs://p1rcc all_the_data/
```

Command line

- GCS has its own command line tool: **gsutil**
- Reference docs here: <https://cloud.google.com/storage/docs/gsutil>
 - Or: **gsutil help <subcommand>**
- Subcommands mirror the Unix file commands, e.g. **cp**, **mv**, **mkdir**, etc.

```
# Recursively copy objects  
gsutil -m cp -r gs://p1rcc all_the_data/
```

Means "go fast"

Google Compute Engine

GCE

- Configurable VMs in the cloud. This is the “EC2” of Google Cloud.
- Here is a simple guide:

<https://cloud.google.com/compute/docs/quickstart-linux>

Creating an instance

Connecting to an instance

<input type="checkbox"/> Name ^	Zone	Recommendation	Internal IP	External IP	Connect
<input type="checkbox"/> claytonvm	us-west1-b		10.138.0.2	35.230.84.81	SSH
<input type="checkbox"/> claytonvm2	us-west1-b		10.138.0.20	104.198.11.135	SSH
<input type="checkbox"/> ggp-13062498541936268504	us-west1-b		10.138.0.3	35.230.53.92	SSH
<input type="checkbox"/> ggp-15082817988080681366	us-west1-b		10.138.0.6	35.199.174.194	SSH
<input type="checkbox"/> nnn-15816786394455391538	us-west1-b		10.138.0.19	35.203.175.254	SSH

- Open in browser window
- Open in browser window on custom port
- View gcloud command
- Use another SSH client

Connect in
Cloud Shell

Connect in
terminal

Networking

- By default, VMs are not exposed to external HTTP traffic.
- Here is a StackOverflow walking through opening a port by creating a firewall rule:
<https://stackoverflow.com/questions/21065922/how-to-open-a-specific-port-such-as-9090-in-google-compute-engine>
- When creating an instance, allow HTTP traffic!

Firewall

Add tags and firewall rules to allow specific network traffic from the Internet

- Allow HTTP traffic
- Allow HTTPS traffic

“One-liner” to create a permission firewall rule

```
# Create a firewall rule to open up the 8000 port range
gcloud compute firewall-rules create open-range-8000 \
  --action allow \
  --direction ingress \
  --rules tcp:8000-9000 \
  --source-ranges 0.0.0.0/0 \
  --priority 1000 \
  --target-tags open-range-8000
```


Acronym Cheatsheet

- GCP = Google Cloud Platform
- GCS = Google Cloud Storage
- GCE = Google Cloud Engine
- VPC = Virtual Private Cloud
- IAM = Identity & Access Management

Misc Resources

- Public genomic datasets:
http://googlegenomics.readthedocs.io/en/latest/use_cases/discover_public_data/reference_genomes.html
- [Example](#) of querying TCGA variant with BigQuery