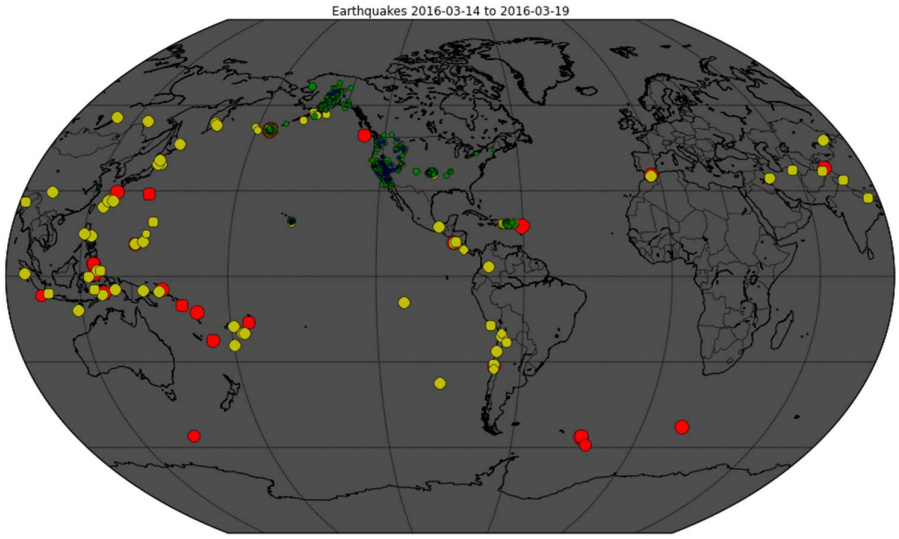
Rent-a-VM to Process Earthquake Data

GSP008

Overview

Using Google Cloud to set up a virtual machine to process earthquake data frees you from IT minutia to focus on your scientific goals. You can ingest and process data, then present the results in various formats. In this lab, you will ingest real-time earthquake data published by the United States Geological Survey (USGS) and create maps that look like the following:



In this lab you will spin up a virtual machine, access it remotely, and then manually create a pipeline to retrieve, process and publish the data.

What you will learn

In this lab, you will learn how to do the following:

- Create a Compute Engine instance with specific security permissions.
- SSH into the instance.
- Install the software package Git (for source code version control).
- Ingest data into the Compute Engine instance.
- Transform data on the Compute Engine instance.
- Store the transformed data on Cloud Storage.
- Publish Cloud Storage data to the web.

Setup

Before you click the Start Lab button

Read these instructions. Labs are timed and you cannot pause them. The timer, which starts when you click **Start Lab**, shows how long Google Cloud resources will be made available to you.

This Qwiklabs hands-on lab lets you do the lab activities yourself in a real cloud environment, not in a simulation or demo environment. It does so by giving you new, temporary credentials that you use to sign in and access Google Cloud for the duration of the lab.

What you need

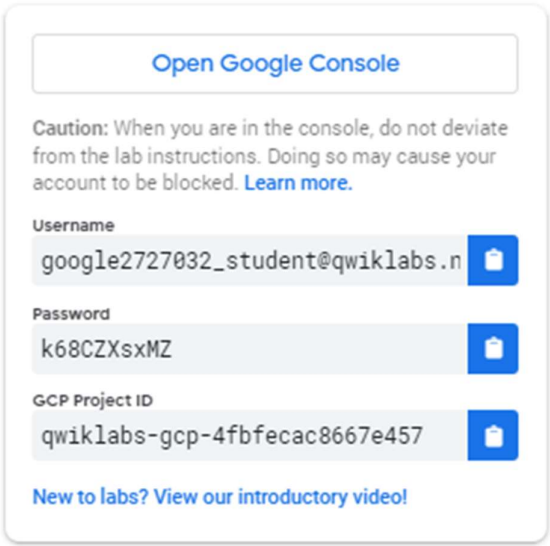
To complete this lab, you need:

- Access to a standard internet browser (Chrome browser recommended).
- Time to complete the lab.
  Note: If you already have your own personal Google Cloud account or project, do not use it for this lab.

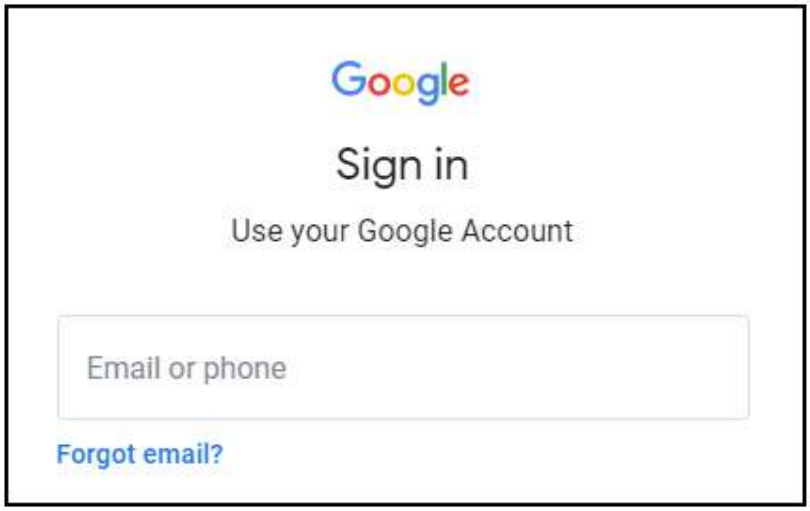
Note: If you are using a Pixelbook, open an Incognito window to run this lab.

How to start your lab and sign in to the Google Cloud Console

1. Click the **Start Lab** button. If you need to pay for the lab, a pop-up opens for you to select your payment method. On the left is a panel populated with the temporary credentials that you must use for this lab.

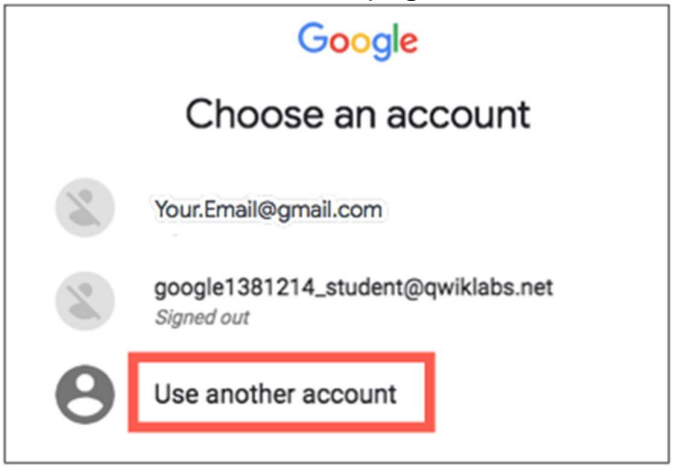


2. Copy the username, and then click **Open Google Console**. The lab spins up resources, and then opens another tab that shows the **Sign in** page.



Tip: Open the tabs in separate windows, side-by-side.

If you see the **Choose an account** page, click **Use Another Account**.



3. In the **Sign in** page, paste the username that you copied from the Connection Details panel. Then copy and paste the password.

   Important: You must use the credentials from the Connection Details panel. Do not use your Qwiklabs credentials. If you have your own Google Cloud account, do not use it for this lab (avoids incurring charges).

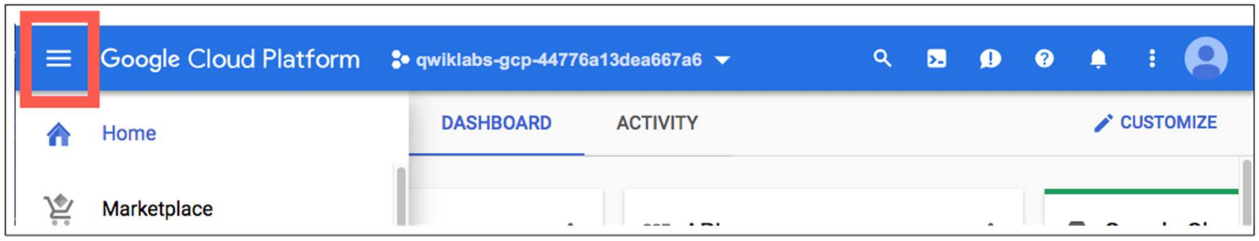
4. Click through the subsequent pages:

   - Accept the terms and conditions.
   - Do not add recovery options or two-factor authentication (because this is a temporary account).
   - Do not sign up for free trials.

After a few moments, the Cloud Console opens in this tab.

Note: You can view the menu with a list of Google Cloud Products and Services by clicking the **Navigation menu** at the top-

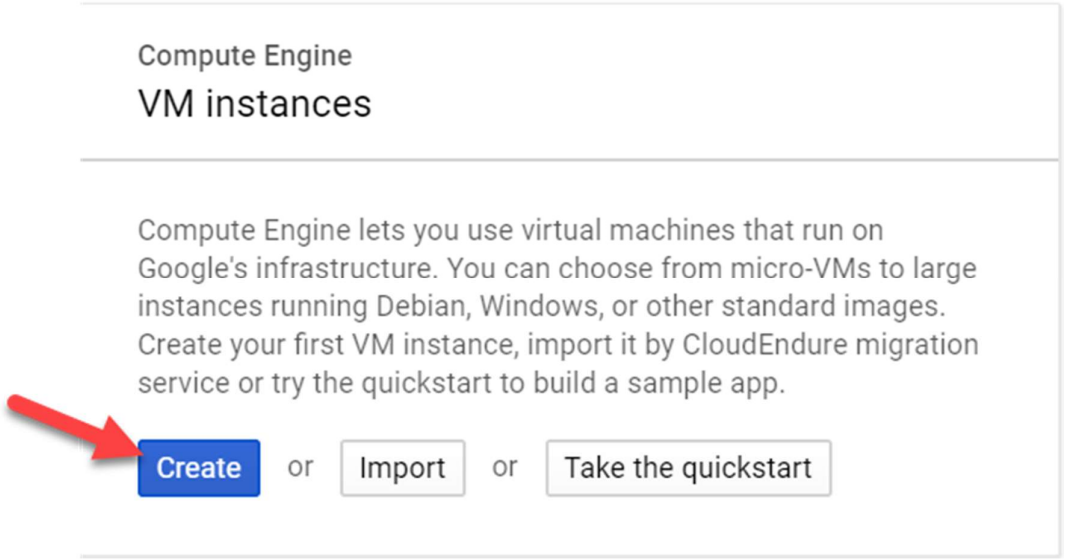
left.



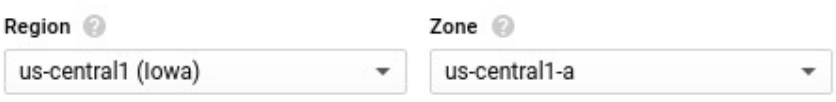
Create Compute Engine instance with the necessary API access

To create a Compute Engine instance, from the **Navigation menu** click on **Compute Engine** > **VM instances**:

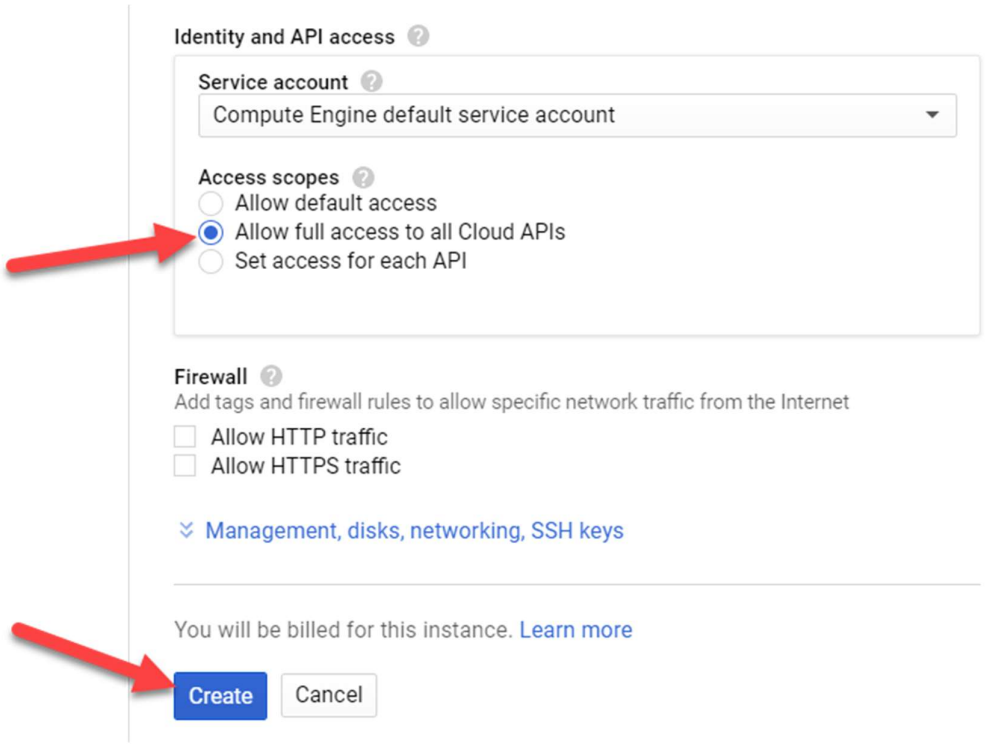
Click **Create** and wait for the "Create an instance" form to load:



Use default Region and Zone for creating the instance:



Change Identify API access for the Compute Engine default service account to **Allow full access to all Cloud APIs**, then click **Create**.



You'll see a green circle with a check when the instance is created.

Click **Check my progress** below to verify you're on track in this lab.

Create Compute Engine instance with the necessary API access

Check my progress

SSH into the instance

You can remotely access your Compute Engine instance using Secure Shell (SSH):

Click the **SSH** button next to your newly created VM:

The VM instance details displays.

Note: Make sure your browser is not blocking pop-ups.
SSH keys are automatically transferred; no extra software is needed to ssh directly from the browser.

To find some information about the Compute Engine instance, type the following into the command-line:

```
cat /proc/cpuinfo
```

You should see a similar output:

```
processor       : 0
vendor_id       : GenuineIntel
cpu family      : 6
model           : 63
model name      : Intel(R) Xeon(R) CPU @ 2.30GHz
....
```

Install software

Still in the SSH window, enter the following commands:

```
sudo apt-get update
sudo apt-get -y -qq install git
sudo apt-get install python-mpltoolkits.basemap
```

Enter **Y** when asked if it's acceptable to use additional disk space.

Verify that git is now installed:

```
git --version
```

You should see a similar output:

```
git version 2.11.0
```

Click **Check my progress** below to verify you're on track in this lab.

Install software

Check my progress

Ingest USGS data

Still in the SSH window, enter the following command to download the code from GitHub:

```
git clone https://github.com/GoogleCloudPlatform/training-data-analyst
```

Note: If you get a git authorization error, it is likely that the GitHub URL has a typo in it. Please copy and paste the above code.

Navigate to the folder corresponding to this lab:

```
cd training-data-analyst/CPB100/lab2b
```

Examine the `ingest` code using `less`:

```
less ingest.sh
```

The `less` command allows you to view the file (Press the **spacebar** to scroll down; the letter **b** to back up a page; the letter **q** to quit).

Enter **q** to exit the editor.

The program `ingest.sh` downloads a dataset of earthquakes in the past 7 days from the US Geological Survey. Notice where the file is downloaded to (disk or Cloud Storage.)

Enter the following command to run the `ingest` code:

```
bash ingest.sh
```

Click **Check my progress** below to verify you're on track in this lab.

Ingest USGS data

Check my progress

Transform the data

You will use a Python program to transform the raw data into a map of earthquake activity:

The transformation code is explained in detail in this notebook.
Feel free to read the narrative to understand what the transformation code does. The notebook itself was written in Datalab, a Google Cloud product that you will use later in this set of labs.

Still in the Compute Engine instance, enter the following command to install the necessary Python packages on the Compute Engine instance:

```
bash install_missing.sh
```

Enter the following command to run the transformation code:

```
python3 transform.py
```

You will notice a new image file `earthquakes.png` in your current directory if you enter the following command:

```
ls -l
```

```
gcpstaging6905_student@instance-1:~/training-data-analyst/CPB100/lab2b$ ls -l
total 640
-rw-r--r-- 1 gcpstaging6905_student gcpstaging6905_student    637 Oct 11 13:21 commands.sh
-rw-r--r-- 1 gcpstaging6905_student gcpstaging6905_student 314708 Oct 11 13:26 earthquakes.csv
-rw-r--r-- 1 gcpstaging6905_student gcpstaging6905_student    751 Oct 11 13:21 earthquakes.htm
-rw-r--r-- 1 gcpstaging6905_student gcpstaging6905_student 312934 Oct 11 13:29 earthquakes.png
-rwxr-xr-x 1 gcpstaging6905_student gcpstaging6905_student    759 Oct 11 13:21 ingest.sh
-rwxr-xr-x 1 gcpstaging6905_student gcpstaging6905_student    680 Oct 11 13:21 install_missing.s
drwxr-xr-x 2 gcpstaging6905_student gcpstaging6905_student   4096 Oct 11 13:21 scheduled
-rwxr-xr-x 1 gcpstaging6905_student gcpstaging6905_student   3074 Oct 11 13:21 transform.py
gcpstaging6905_student@instance-1:~/training-data-analyst/CPB100/lab2b$
```

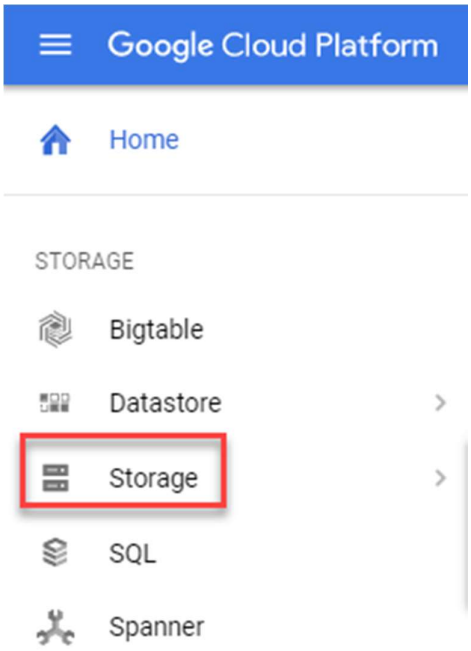
Click **Check my progress** below to verify you're on track in this lab.

Transform the data

Check my progress

Create a Cloud Storage bucket

Return to the Cloud Console for this step. From the **Navigation menu** select **Storage**:



Click on **Create Bucket**, then create your bucket with the following characteristics:

- Choose a globally unique bucket name (but not a name you'd like to use for your own projects), then click **Continue**.
- You can leave it as **Multi-Regional**, or improve speed and reduce costs by making it **Regional** (choose the same region as your Compute Engine instance).
  Then, click **Create**.

Take note of your bucket name. You will insert its name whenever the instructions ask for `<YOUR-BUCKET>`.

Store data

You will now learn how to store the original and transformed data in Cloud Storage.

In the SSH window of the Compute Engine instance, run the following, changing `<YOUR-BUCKET>` to the bucket name you created earlier:

```
gsutil cp earthquakes.* gs://<YOUR-BUCKET>/earthquakes/
```

This command copies the files to your bucket in Cloud Storage.

Return to the Cloud Console and on the Storage Browser page click on the **Refresh** button near the top of the page. Now click on the bucket name then the `/earthquakes` folder.

You should now see the following three files in the earthquakes folder:

Name	Size	Type	Storage class	Last modified	Public access
earthquakes.csv	641.82 KB	text/csv	Standard	9/6/19, 4:38:00 PM UTC+5:30	Not public
earthquakes.htm	751 B	text/html	Standard	9/6/19, 4:38:00 PM UTC+5:30	Not public
earthquakes.png	311.11 KB	image/png	Standard	9/6/19, 4:38:01 PM UTC+5:30	Not public

Click **Check my progress** below to verify you're on track in this lab.

Create bucket and Store data

Check my progress

Publish Cloud Storage files to web

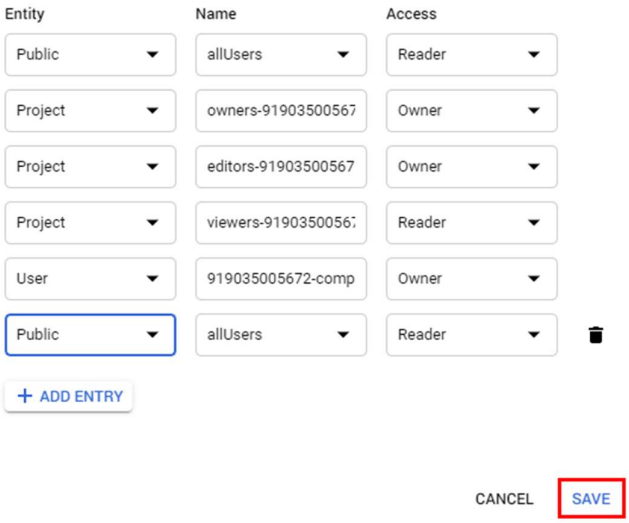
You will now publish the files in your bucket to the web.

To create a publicly accessible URL for the files, click on the `earthquakes.htm` file, then click the three dots at the end of the row and select **Edit Permissions** from the dropdown menu.

In the overlay that appears, click the **+ Add entry** button.

Add a permission for all users by entering in the following:

- Select **Public** for the Entity.
- Enter **allUsers** for the Name.
- Select **Reader** for the Access.
- Then click **Save**.

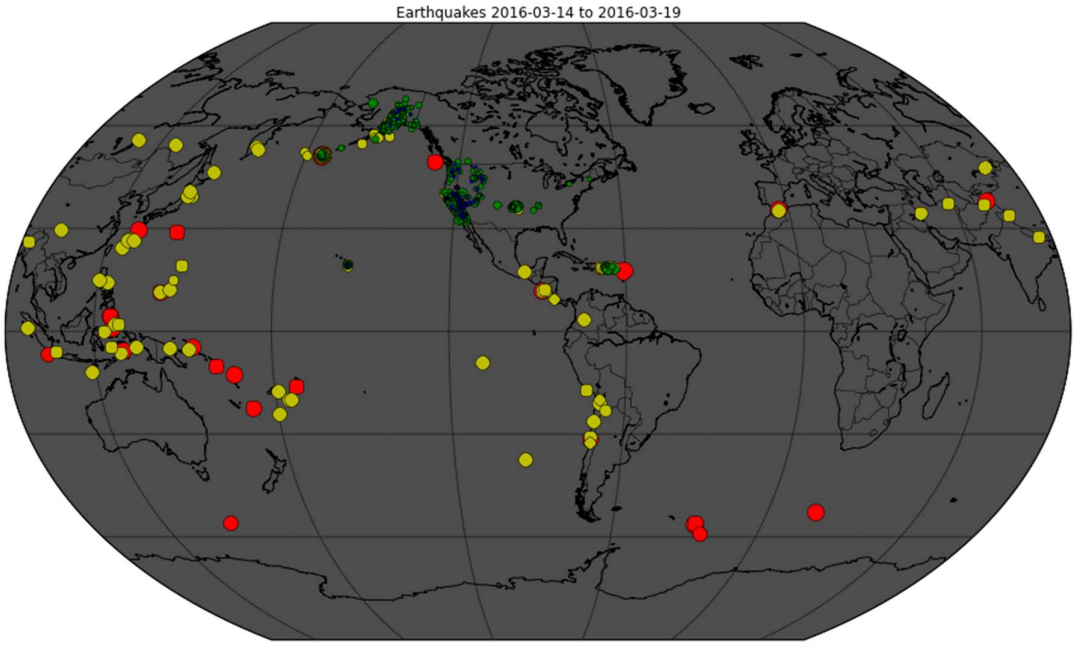


Repeat the above steps for `earthquakes.png`.

Click on the name of a file and notice the URL of the published Cloud Storage file and how it relates to your bucket name and content. It should resemble the following:

```
https://storage.cloud.google.com/YOUR-BUCKET-NAME/earthquakes/earthquakes.png
```

If you click on the `earthquakes.png` image file and then on the public URL, a new tab will be opened with the following image loaded:



Go ahead and close the SSH window.

Congratulations!

You have completed this lab and learned how to spin up a compute engine instance, access it remotely, then manually create a pipeline to retrieve, process and publish the data.

Finish Your Quest

This self-paced lab is part of the Qwiklabs Scientific Data Processing Quest. A Quest is a series of related labs that form a learning path. Completing this Quest earns you the badge above, to recognize your achievement. You can make your badge (or badges) public and link to them in your online resume or social media account. Enroll in this Quest and get immediate completion credit if you've taken this lab. See other available Qwiklabs Quests.

Take Your Next Lab

Continue your Quest with Weather Data in BigQuery, or try these suggestions:

- Distributed Image Processing in Cloud Dataproc
- Distributed Computation of NDVI from Landsat Images using Cloud Dataflow

Next Steps/Learn More

Here are some follow-up steps:

- Check out USGS.gov for complete information. For example:
  - the latest 20 large earthquakes in the world
  - geodetic data
  - hazard assessment data and models, and more.
- Sign up for automatic notifications of earthquakes in your area.

Google Cloud Training & Certification

...helps you make the most of Google Cloud technologies. Our classes include technical skills and best practices to help you get up to speed quickly and continue your learning journey. We offer fundamental to advanced level training, with on-demand, live, and virtual options to suit your busy schedule. Certifications help you validate and prove your skill and expertise in Google Cloud technologies.
Manual Last Updated November 20, 2020
Lab Last Tested November 20, 2020