Google Cloud Datalab: Qwik Start

GSP085

Overview

Cloud Datalab is an interactive data analysis and machine learning environment designed for Google Cloud Platform (GCP) for data exploration, analysis, visualization, and machine learning.

Cloud Datalab uses notebooks instead of text files containing code. Notebooks bring together code, documentation written in markdown, and the results of code execution — whether as text, image, or HTML/JavaScript. Like a code editor or IDE, notebooks allow you to execute code in an interactive and iterative manner, rendering the results alongside the code. When you share a notebook with team members, you can include code, markdown-formatted documentation, and

results that include interactive charts, to provide them with context that goes beyond what Python or SQL code files alone can provide.

With Cloud Datalab you'll be able to analyze your data on Google BigQuery, Cloud Machine Learning Engine, Google Compute Engine, and Google Cloud Storage using Python, SQL, and JavaScript (for BigQuery user-defined functions).

What you'll do

In this lab you'll:

- Create a Datalab instance
- Create a new notebook
- Use git and ungit for source control

Qwiklabs setup

What you'll need

To complete this lab, you'll need:

- Access to a standard internet browser (Chrome browser recommended).
- Time. Note the lab's **Completion** time in Qwiklabs. This is an estimate of the time it should take to complete all steps. Plan your schedule so you have time to complete the lab. Once you start the lab, you will not be able to pause and return later (you begin at step 1 every time you start a lab).

- The lab's **Access** time is how long your lab resources will be available. If you finish your lab with access time still available, you will be able to explore the Google Cloud Platform or work on any section of the lab that was marked "if you have time". Once the Access time runs out, your lab will end and all resources will terminate.

- You **DO NOT** need a Google Cloud Platform account or project. An account, project and associated resources are provided to you as part of this lab.

- If you already have your own GCP account, make sure you do not use it for this lab.

- If your lab prompts you to log into the console, **use only the student account provided to you by the lab**. This prevents you from incurring charges for lab activities in your personal GCP account.

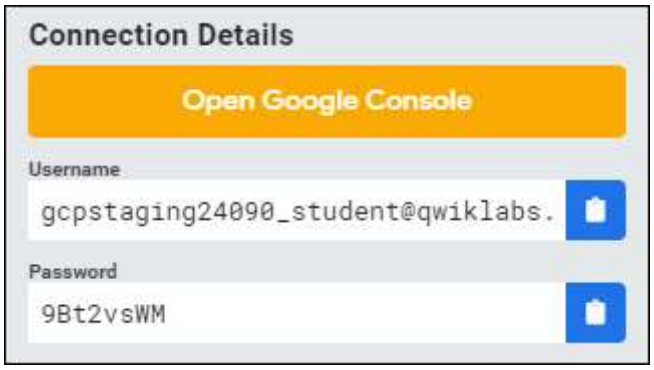
Start your lab

When you are ready, click **Start Lab**. You can track your lab's progress with the status bar at the top of your screen.

> **Important:** What is happening during this time? Your lab is spinning up GCP resources for you behind the scenes, including an account, a project, resources within the project, and permission for you to control the resources needed to run the lab. This means that instead of spending time manually setting up a project and building resources from scratch as part of your lab, you can begin learning more quickly.

Find Your Lab's GCP Username and Password

To access the resources and console for this lab, locate the Connection Details panel in Qwiklabs. Here you will find the account ID and password for the account you will use to log in to the Google Cloud Platform:

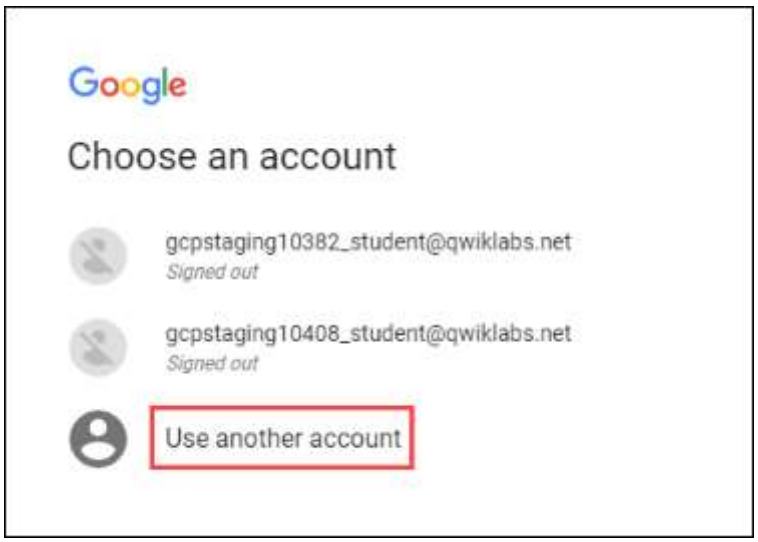


If your lab provides other resource identifiers or connection-related information, it will appear on this panel as well.

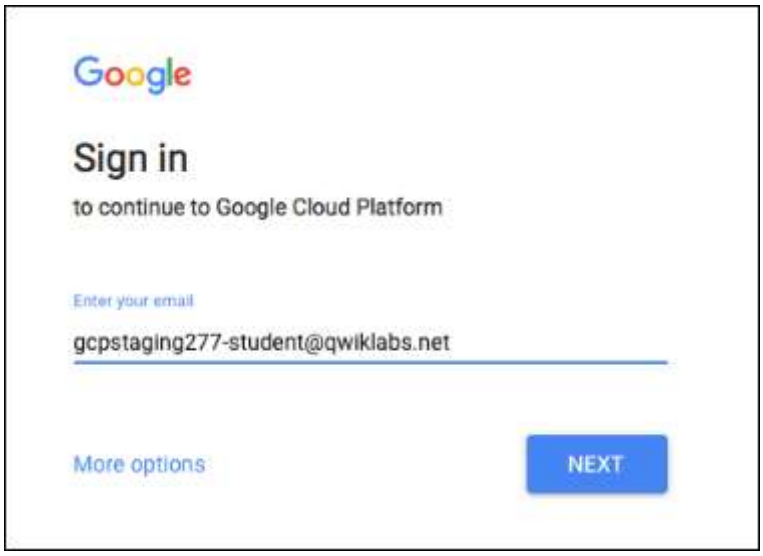
Log in to Google Cloud Console

Using the Qwiklabs browser tab/window or the separate browser you are using for the Qwiklabs session, copy the Username from the Connection Details panel and click the “Open Google Console” button.

You'll be asked to choose an account. Click **Use another account**.



Paste in the Username, and then the Password as prompted:

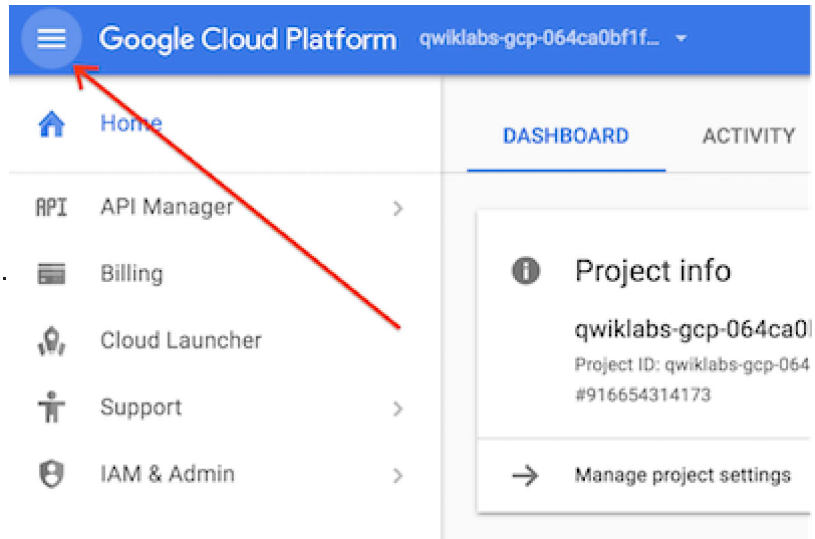


Accept the terms and conditions.

Since this is a temporary account, which you will only have access to for this one lab:

- Do not add recovery options
- Do not sign up for free trials

Note: You can view the menu with a list of GCP Products and Services by clicking the **Navigation menu** at the top-left next to “Google Cloud Platform”.

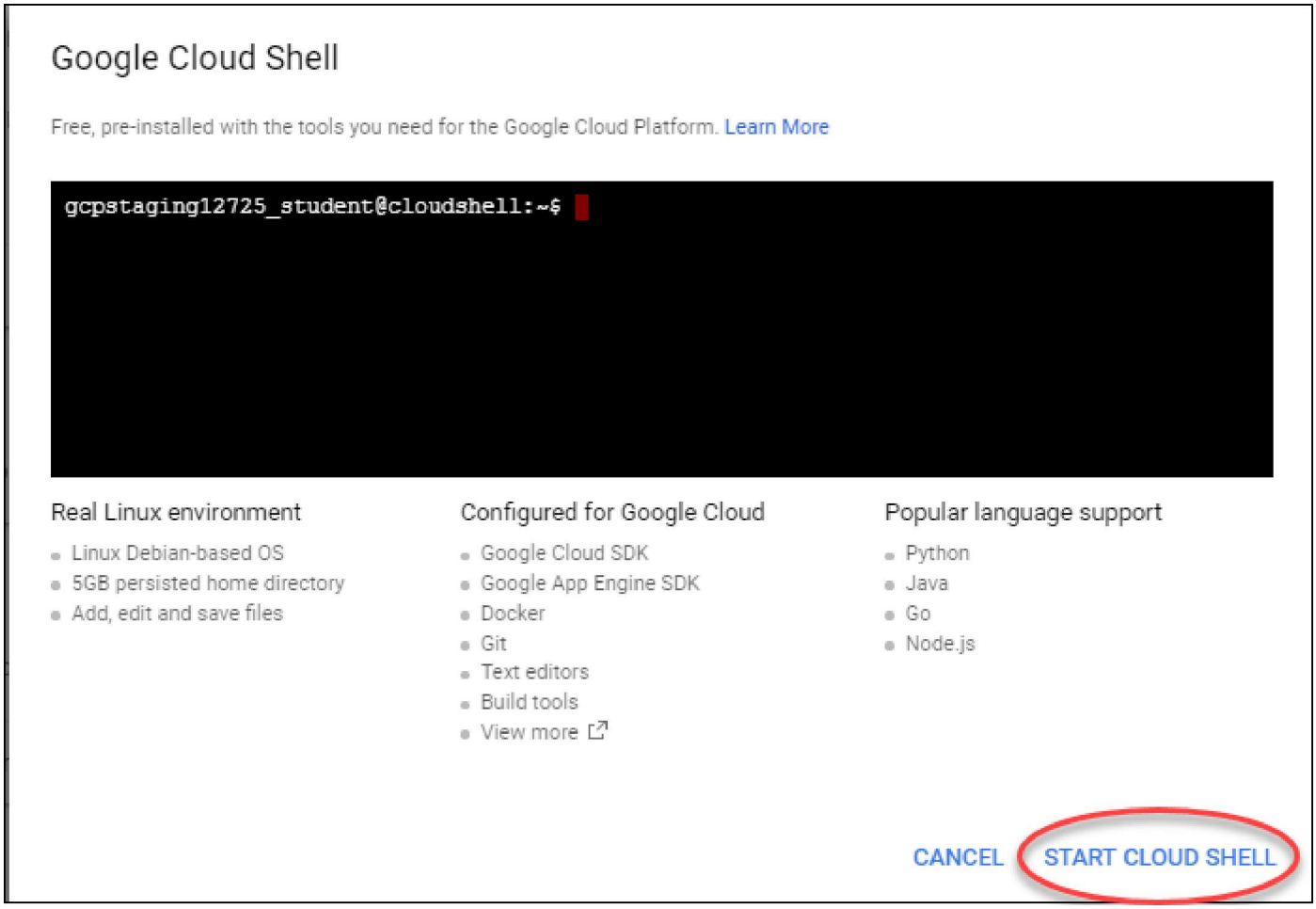


Activate Google Cloud Shell

Google Cloud Shell provides command-line access to your GCP resources.

From the GCP Console click the Cloud Shell icon on the top right toolbar:

Then click **START CLOUD SHELL**:



You can click "START CLOUD SHELL" immediately when the dialog comes up.

It takes a few moments to provision and connect to the environment:

The Cloud Shell is a virtual machine loaded with all the development tools you'll need. It offers a persistent 5GB home directory, and runs on the Google Cloud, greatly enhancing network performance and authentication.

Once connected to the cloud shell, you'll see that you are already authenticated and the project is set to your *PROJECT_ID*:

```
gcloud auth list
```

Output:

```
Credentialed accounts:
 - <myaccount>@<mydomain>.com (active)
```

Note: gcloud is the powerful and unified command-line tool for Google Cloud Platform. Full documentation is available on Google Cloud gcloud Overview. It comes pre-installed on Cloud Shell and supports tab-completion.

```
gcloud config list project
```

Output:

```
[core]
project = <PROJECT_ID>
```

Creating a new Datalab instance

When you create a new Datalab instance, GCP adds a `datalab-notebooks` Cloud Source Repository to the project. In this lab, the repository is called the "Cloud Remote Repo". It's a remote repository for the `/content/datalab/notebooks` git repository created in the Docker container that runs in your Cloud Datalab VM instance. The VM instance is called the "Cloud Datalab VM repo".

- In Cloud Shell, enter the following command to create a database instance named `my-datalab`:

```
datalab create my-datalab
```

Database instance name rules: The name of the instance must start with a lowercase letter, followed by up to 63 lowercase letters, numbers, or hyphens, and cannot end with a hyphen.

- When prompted, type in the zone number for `us-central1-c`.
- After you specify the zone, the system creates and connects the new Datalab VM Instance.

You will see something like this:

```
Creating the network datalab-network
Creating the firewall rule datalab-network-allow-ssh
Creating the disk my-datalab-pd
Creating the repository datalab-notebooks
Creating the instance my-datalab
Creating the instance my-datalab
Created
[https://www.googleapis.com/compute/v1/projects/qwiklabs-gcp-
a71230ef4a2cf32e/zones/us-east1-d/instances/my-datalab].
Connecting to my-datalab.
This will create an SSH tunnel and may prompt you to create an rsa
key pair. To manage these keys, see
https://cloud.google.com/compute/docs/instances/adding-removing-
ssh-keys
Waiting for Datalab to be reachable at http://localhost:8081/
```

```
This tool needs to create the directory
[/home/gcpstaging10172_student/.ssh] before being able to generate
SSH
 keys.
```

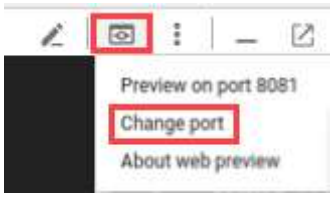
- When prompted, enter **Y** to continue.
- When prompted for a passphrase, press **Enter** for no passphrase, and then again to confirm.
- Datalab will take 5 - 10 minutes to set up. Go watch the *Datalab: Notebooks on the Cloud* video while it's building.

When you see the following output, you have successfully created and connected to the Datalab VM instance:

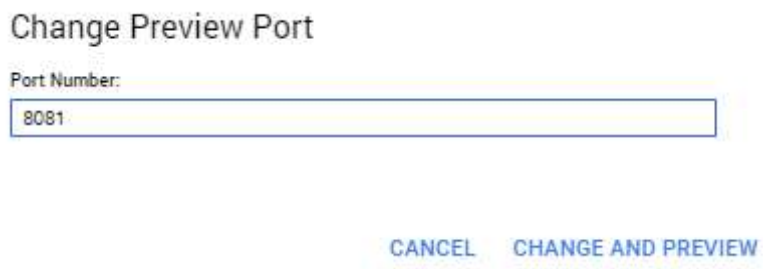
```
The connection to datalab is now open and will remain until this
command is killed.
Click on the *Web Preview* (square button at top-right), select
*Change port > Port 8081*, and start using Datalab.
```

Datalab is ready when you see a message prompting you to use "Web Preview".

At the top of Cloud Shell, click the Web Preview icon, then "Change port". If you do not see the Web preview icon, click the hamburger icon at the top of the left menu to close it.



Now type **8081** and click "Change and Preview".



Click **Check my progress** below to check your lab progress.

Creating a new Datalab instance

Check my progress

You can browse the Cloud Remote Repo from the Cloud Platform Console Repositories page from the left menu. After clicking the **Navigation menu** (hamburger), navigate to **Source Repositories** > **Repositories**.

Working with notebooks

You can store Cloud Datalab notebooks in Google Cloud Source Repo, which is a git repository. The system clones this git repository onto a persistent disk attached to the VM. The clone is the workspace where you can add, remove, and modify files. To share your work with other users,

commit your changes in your VM workspace and then use the git client to push your changes to the Cloud Source Repo. Notebooks are automatically saved to the persistent disk periodically, and you can save them yourself whenever you want.

> If you delete the persistent disk, any notebooks that are not explicitly pushed to the git repository may be lost. Deleting the persistent disk is not recommended.

Now that you've created and connected your Cloud Datalab, time to explore different tasks associated with Cloud Datalab notebooks.

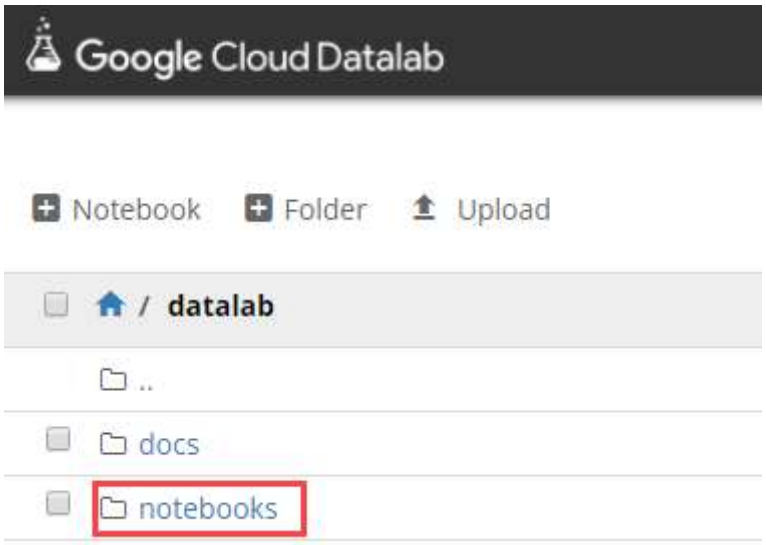
Using ungit in your browser

The Cloud Datalab container includes ungit, a web-based git client which allows you to make commits to the Cloud Datalab VM Repo and push notebooks to the Cloud Remote Repo from the Cloud Datalab browser UI.

In this section you'll create a new notebook, then commit and push the notebook to the Cloud Datalab VM repo.

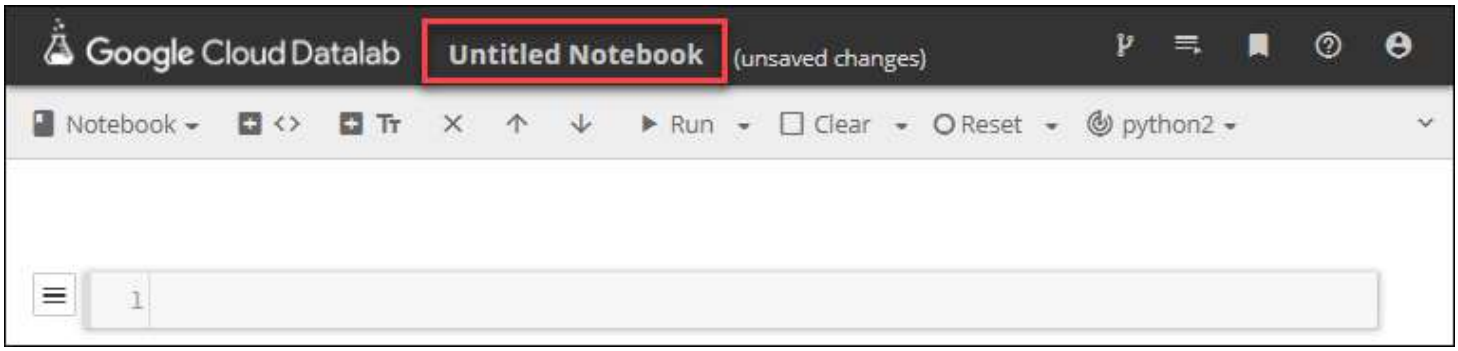
First, add a notebook to the Cloud Remote Repo.

. Navigate to the `/datalab/notebooks` folder by clicking "Notebooks".



. Add a new notebook from the /datalab/notebooks folder by clicking the **+ Notebook** icon.

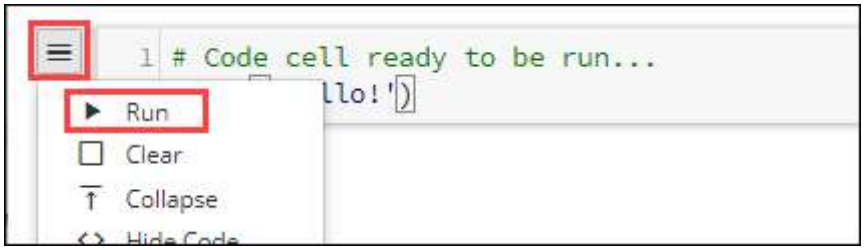
. Rename the notebook by clicking **Untitled Notebook**in the menu bar, then change the name to "New Notebook", and click **OK**.



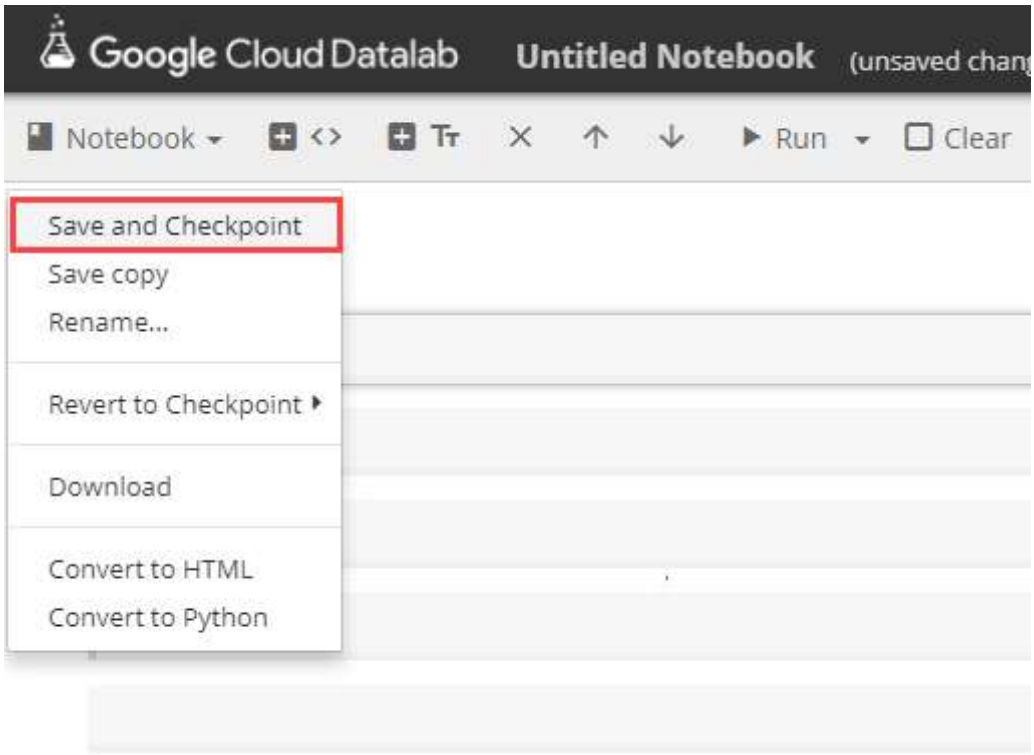
. Type the following into the cell on line 1:

```
# Code cell ready to be run...
print('Hello!')
```

. Run the code by clicking the menu button to the left of the cell then **Run**.



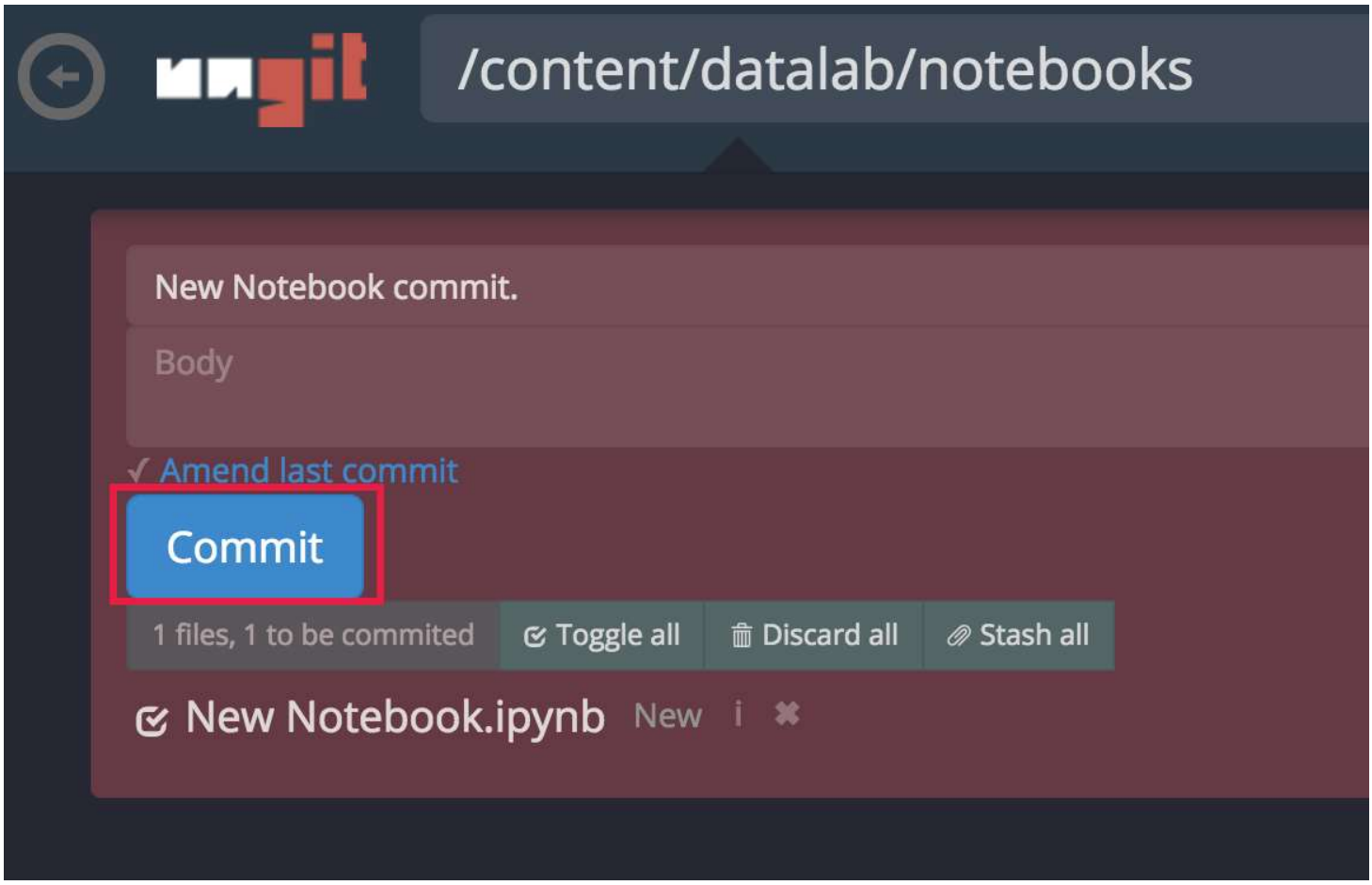
. Select **Notebook** > **Save and Checkpoint** or **Ctrl-s** to save the notebook. Notebooks are also autosaved.



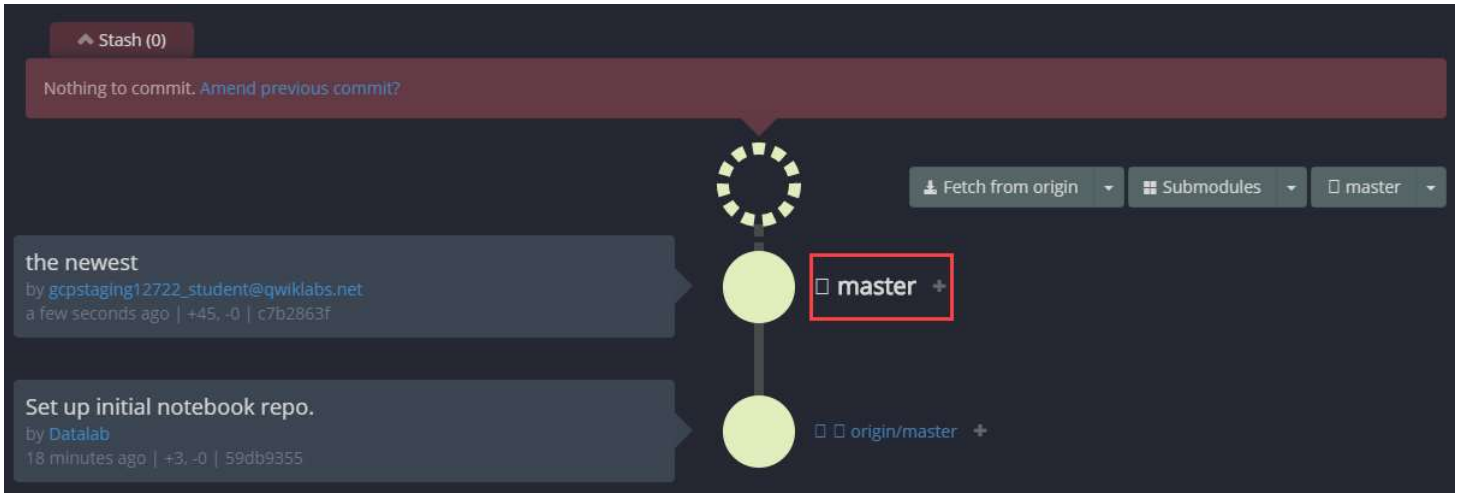
. Go back to the **Google Cloud Datalab** window and click the **ungit** icon in the notebooks repository to open ungit on the /content/datalab/notebooks repo.

A browser window opens on the Cloud Datalab VM.

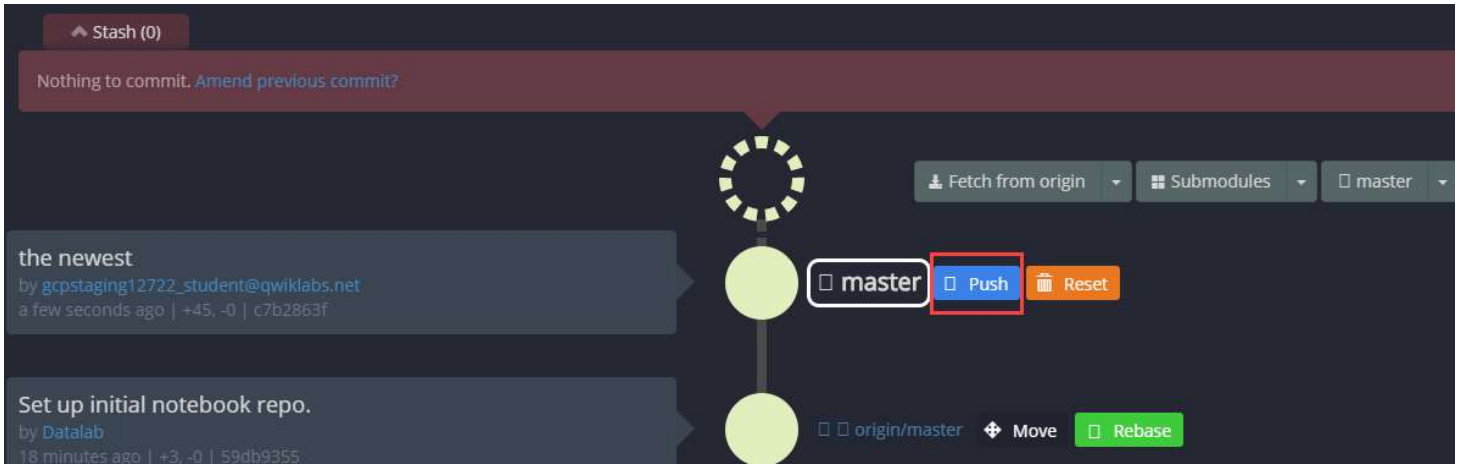
. Add a title for your commit, then click **Commit** to commit the new notebook to the Cloud Datalab VM Repo.



. After committing the notebook, click **master** to see your options for what you can do with the commit.



. Click **Push** to add the changes to the `datalab-notebooks` Cloud Remote Repo.



. Return to the Console and click the name of the repository name (`datalab-noteboooks`) and you'll see the notebook that you created and the commit you just pushed.

Click **Check my progress** below to check your lab progress.

Use ungit for committing the code

Check my progress

Using git from the command line

Instead of using ungit in the Cloud Datalab for source control, you can SSH into the Cloud Datalab VM and run git from a Cloud Shell terminal. You'll use this tecnique next.

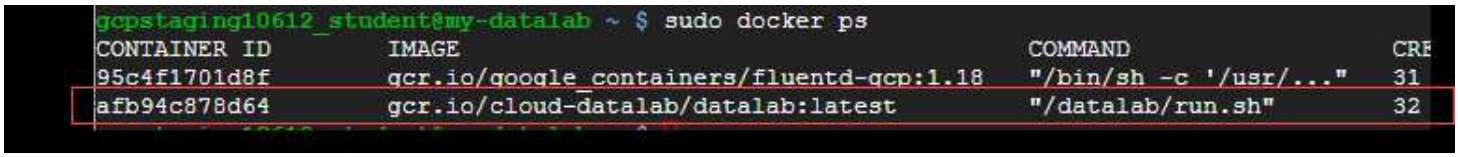
. In the Console, click **Navigation menu** > **Compute Engine** > **VM instances**. Your datalab will be listed in the VM instances window.

. Click the **SSH** button to SSH to the Cloud Datalab VM.

. In the SSH window, list the Container ID of the Cloud Datalab docker image running in the VM.

```
sudo docker ps
```

. Copy the Container ID that is associated with the `/datalab/run.sh` command.



. Open an interactive shell session inside the container, replacing `<container-id>` with the ID you copied by running:

```
docker exec -it <container-id> bash
```

. Change to the `/content/datalab/notebooks` directory in the container:

```
cd /content/datalab/notebooks
```

. This is the root directory for Cloud Datalab VM Repo from which you can issue git commands. As an example, run this command to show the commit logs from your earlier work:

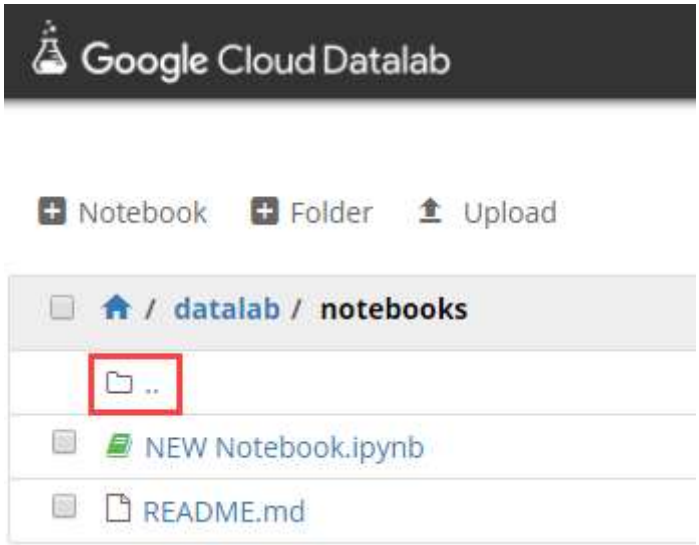
```
git log
```

Sample output:

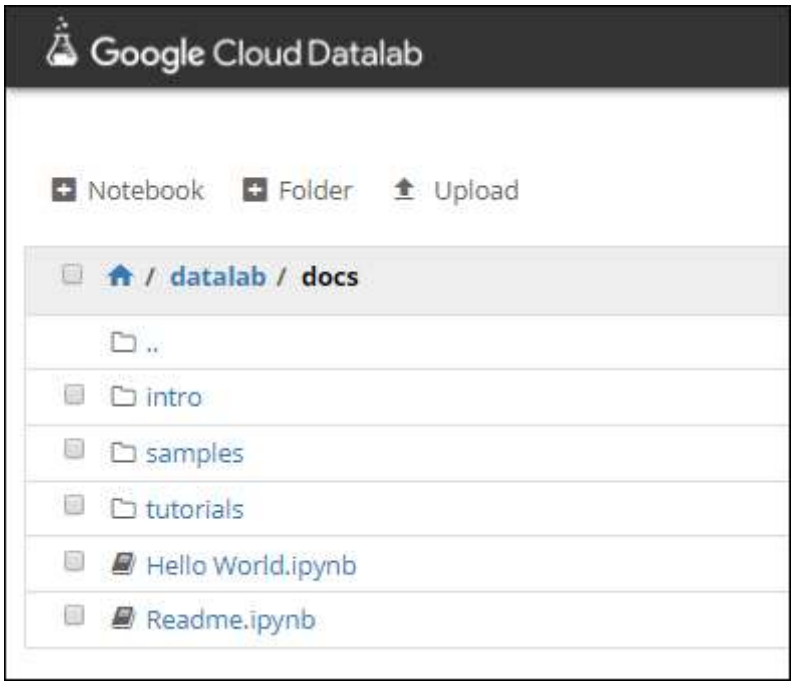
```
commit b768195849fe9824fbd784262e90ddbdea5318c2
Author: gcpstaging10612_student@qwiklabs.net
<gcpstaging10612_student@qwiklabs.net>
Date:   Fri Jan 26 22:55:10 2018 +0000
    New Notebook commit.
commit ac2ed6d01cb42c1cb8b680b8255eea3277899823
Author: Datalab <nobody>
Date:   Fri Jan 26 22:35:56 2018 +0000
    Set up initial notebook repo.
```

For a list of other git commands, enter `git` in the command line.

If you have time, go back to the **Cloud Datalab**window and click on the folder icon, then navigate to `datalab/docs` .



The docs folder has example labs, samples, and other resources to for you to explore.



Test your knowledge

Test your knowledge about Google cloud Platform by taking our quiz.

Notebooks help you write code: they allow you to execute code in an interactive and iterative manner, rendering the results alongside the code. Further, when you share a notebook with team members, you can include code, markdown-formatted documentation, and results that include interactive charts, to provide them with context that goes beyond what Python or SQL code files alone can provide.

True False

Congratulations!

You created a Datalab and explored how to use ungit and git to manage the notebook.

Finish Your Quest

Continue your Quest with Baseline: Data, ML, AI. 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

This lab is also part of a series of labs called Qwik Starts. These labs are designed to give you a little taste of the many features available with Google Cloud. Search for "Qwik Starts" in the lab catalog to find the next lab you'd like to take!

Next Steps /Learn More

- Check out our Quests for a series of labs to concentrate on one area. For example, Networking in the Google Cloud.
- Read the Google Cloud Datalab Documentation for more information.
- To learn more about ungit, watch this ungit video.
- Watch Datalab: Notebook in the cloud (11 minutes).