

Analyzing Natality Data Using AI Platform and BigQuery

GSP012



Overview

In this lab, you analyze a large (137 million rows) natality dataset using BigQuery and AI Platform.

What you learn

In this lab, you:

- Launch AI Platform notebook
- Invoke a BigQuery query
- Create charts in Jupyter
- Export data for machine learning

This lab illustrates how you can carry out data exploration of large datasets, but continue to use familiar tools like Pandas and Jupyter. The trick is to do the first part of your aggregation in BigQuery, get back a Pandas DataFrame, then work with the smaller Pandas DataFrame locally. AI Platform provides a managed Jupyter experience, so you don't need to run notebook servers yourself.

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:

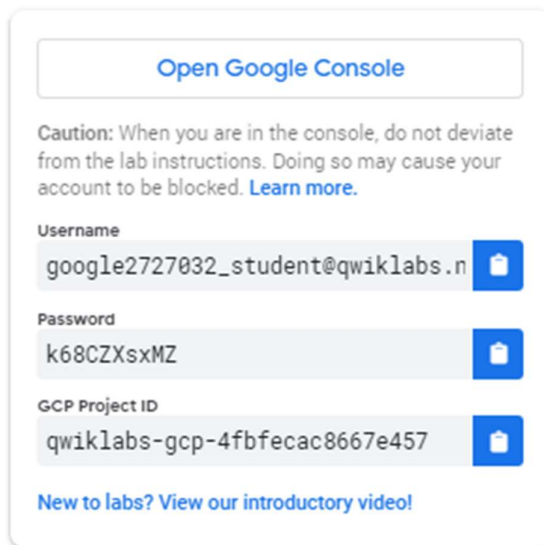
- Access to a standard internet browser (Chrome browser recommended).
- Time to complete the lab.

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.



Open Google Console

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. [Learn more.](#)

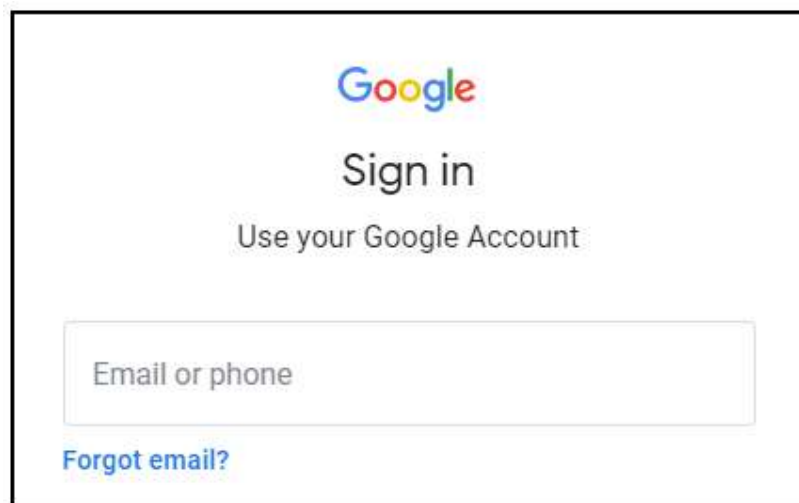
Username
google2727032_student@qwiklabs.n

Password
k68CZxsxMZ

GCP Project ID
qwiklabs-gcp-4fbfecac8667e457

[New to labs? View our introductory video!](#)

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.



Google

Sign in

Use your Google Account

Email or phone

[Forgot email?](#)

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.

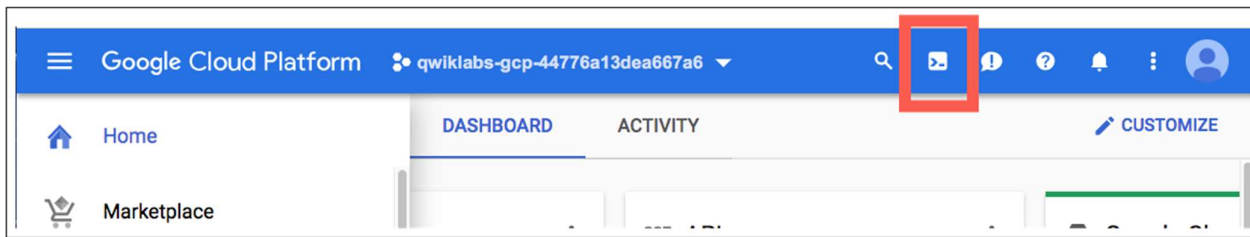
Note: You can view the menu with a list of Google Cloud Products and Services by clicking the **Navigation menu** at the top-left.



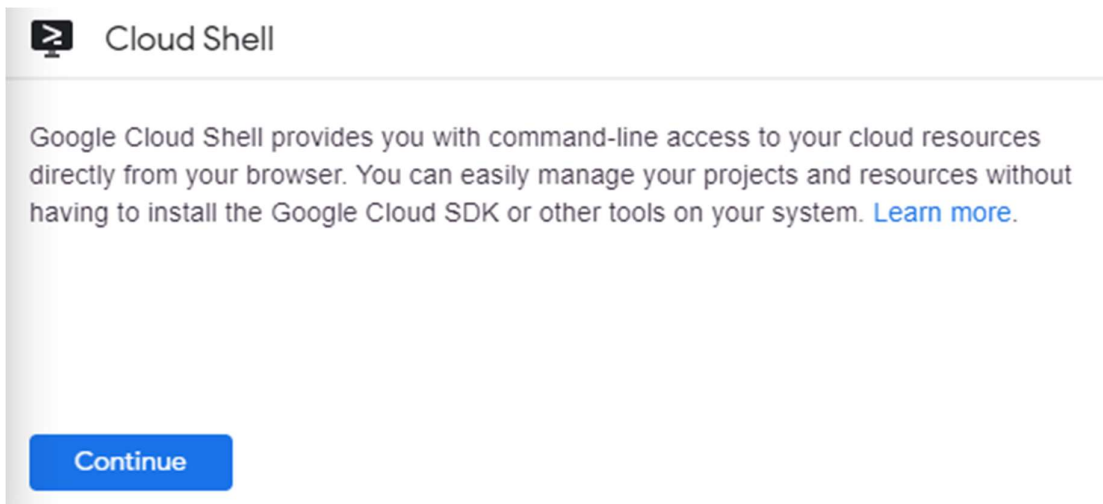
Activate Cloud Shell

Cloud Shell is a virtual machine that is loaded with development tools. It offers a persistent 5GB home directory and runs on the Google Cloud. Cloud Shell provides command-line access to your Google Cloud resources.

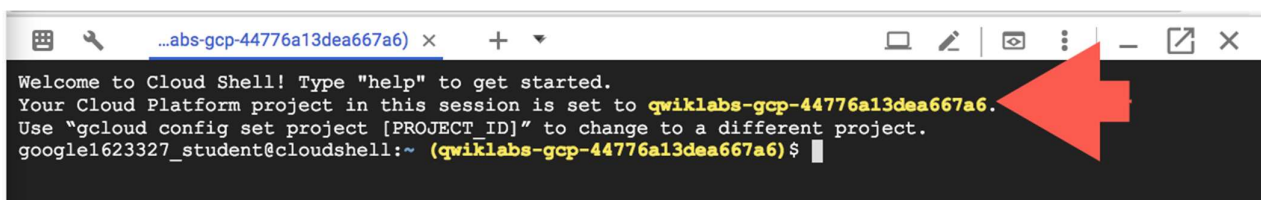
In the Cloud Console, in the top right toolbar, click the **Activate Cloud Shell** button.



Click **Continue**.



It takes a few moments to provision and connect to the environment. When you are connected, you are already authenticated, and the project is set to your *PROJECT_ID*. For example:



`gcloud` is the command-line tool for Google Cloud. It comes pre-installed on Cloud Shell and supports tab-completion.

You can list the active account name with this command:

```
gcloud auth list
```

(Output)

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

(Example output)

```
Credentialed accounts:  
- google1623327 student@gwiklabs.net
```

You can list the project ID with this command:

```
gcloud config list project
```

(Output)

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

(Example output)

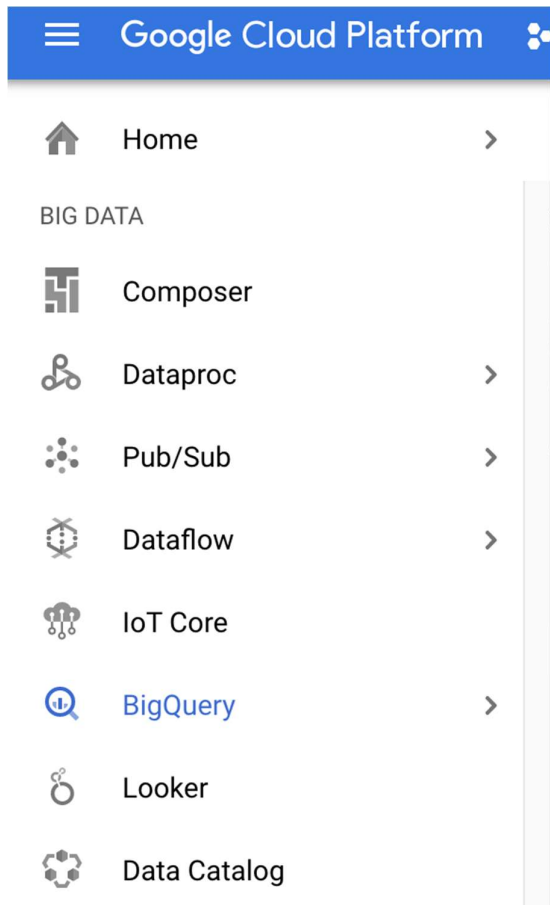
```
[core]  
project = qwiklabs-gcp-44776a13dea667a6
```

For full documentation of `gcloud` see the [gcloud command-line tool overview](#).

Invoke BigQuery

Open BigQuery Console

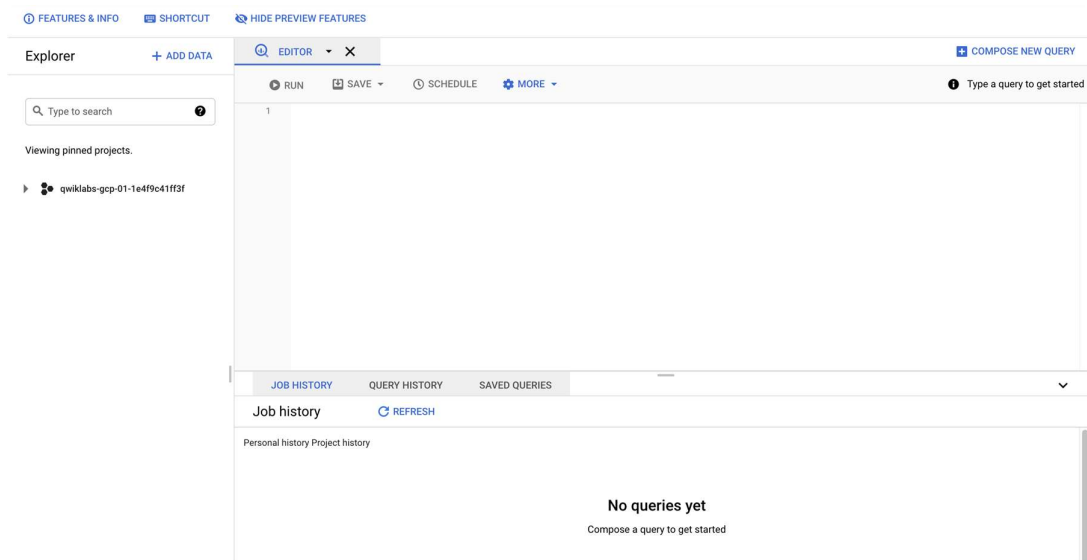
In the Google Cloud Console, select **Navigation menu** > **BigQuery**:



The **Welcome to BigQuery in the Cloud Console** message box opens. This message box provides a link to the quickstart guide and the release notes.

Click **Done**.

The BigQuery console opens.



In the Query Editor enter the following query:

```
SELECT
  plurality,
  COUNT(1) AS num_babies,
  AVG(weight_pounds) AS ave_weight
FROM
  `bigquery-public-data.samples.natality`
WHERE
  year > 2000 AND year < 2005
GROUP BY
  plurality
```

Now click **Run**.

Review the result. How many triplets were born in the US between 2000 and 2005?

Click **Check my progress** to verify the objective.

Invoke Bigquery and run a query.

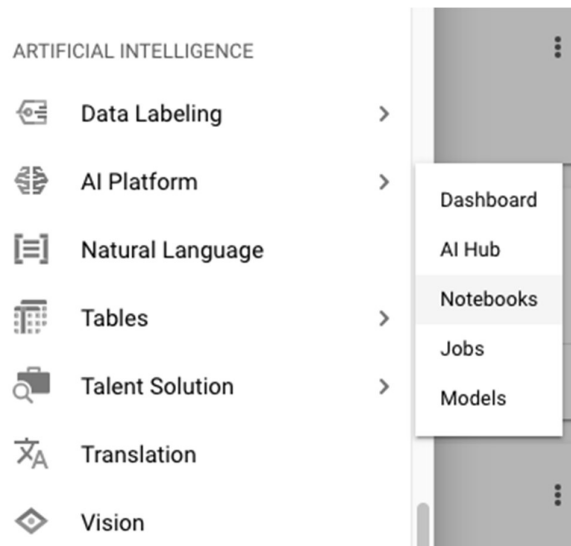
Check my progress

Launch AI Platform Notebooks

To launch AI Platform Notebooks:

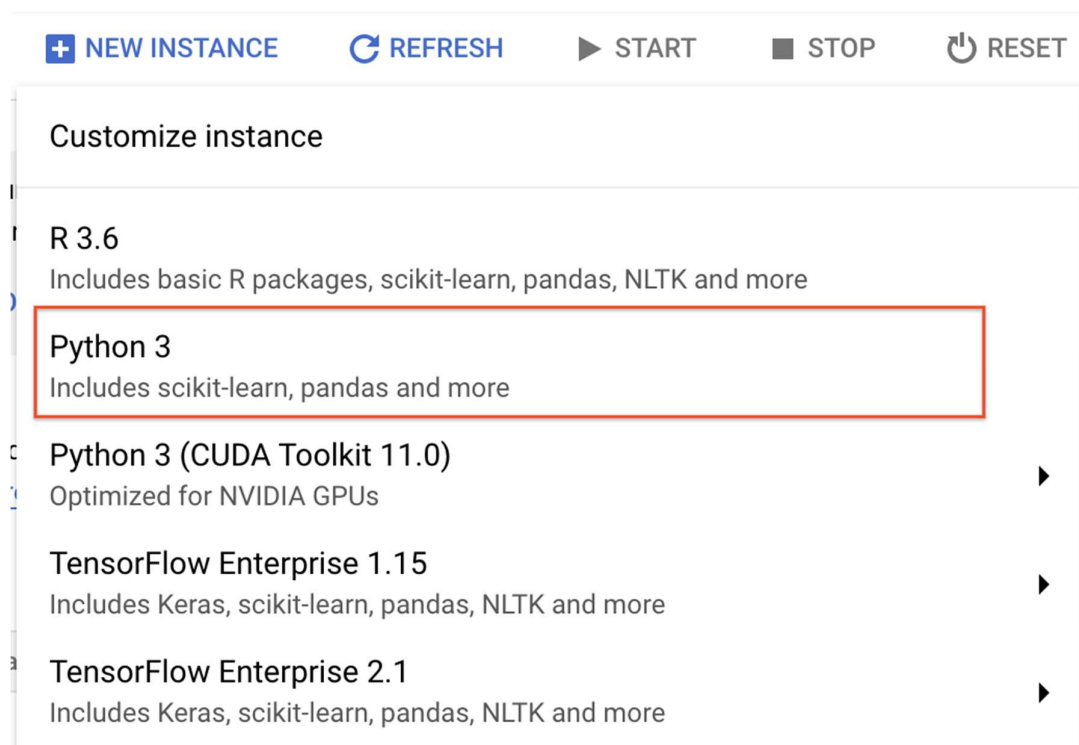
Step 1

Click on the **Navigation Menu**. Navigate to **AI Platform**, then to **Notebooks**.



Step 2

On the Notebook instances page, click [+ NEW INSTANCE](#). Select Python 3.

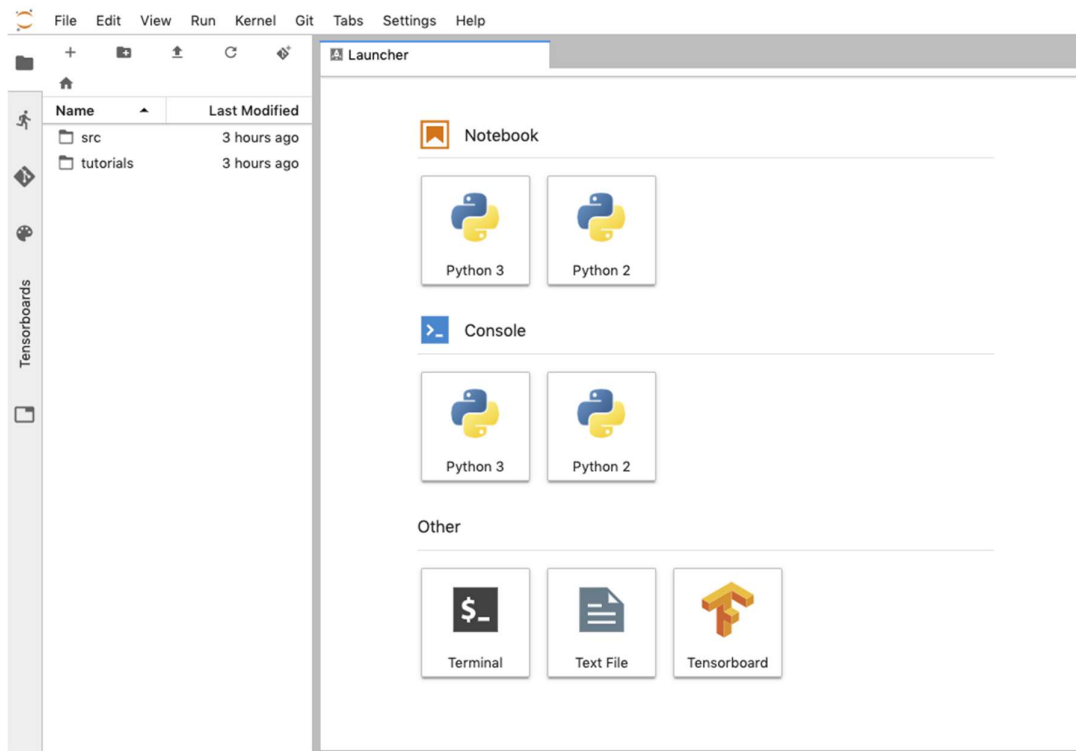


In the pop-up, confirm the name of the deep learning VM, for **Region**, select `us-central1` and for **Zone**, select a zone within that region. Leave the remaining fields with their default and click **Create**.

The new VM will take 2-3 minutes to start.

Step 3

Click **Open JupyterLab**. A JupyterLab window will open in a new tab.



Click **Check my progress** to verify the objective.

Create AI Platform Notebooks instance.

Check my progress

Visualize data in AI Platform

In JupyterLab, start a new notebook by clicking on **Notebook > Python 3**.

Start by updating to the latest version of the BigQuery Python Client Library.

Enter the following into the first cell of the notebook, then click the **Run** button at the top of the screen:

```
!pip install google-cloud-bigquery==1.25.0
!pip install --upgrade google-cloud-bigquery-storage
```

Kindly ignore the deprecation warning: "After October 2020 you may experience errors when installing or updating...."

Restart the kernel by clicking on **Restart the kernel**.

Next, insert the following code to import the BigQuery Python Client Library and initialize a client. The BigQuery client will be used to send and receive messages from the BigQuery API.

```
from google.cloud import bigquery
client = bigquery.Client()
```

Run the cell with **Shift + Enter**.

Add the following into the next cell of the notebook to run a query on the BigQuery natality public dataset:

```
sql = """
SELECT
    plurality,
    COUNT(1) AS count,
    year
FROM
    `bigquery-public-data.samples.natality`
WHERE
    NOT IS_NAN(plurality) AND plurality > 1
GROUP BY
    plurality, year
ORDER BY
    count DESC
"""
df = client.query(sql).to_dataframe()
df.head()
```

This dataset describes all United States births registered from 1969 to 2008. This query returns the annual count of plural births by plurality (2 for twins, 3 for triplets, etc.).

Run the cell with **Shift + Enter**.

You just ran a query in the cloud! The head of the DataFrame (the first 5 rows) is displayed below the code cell. Full results are available for further analysis in a Pandas DataFrame.

Run a query to get annual count of plural births by plurality

Check my progress

Insert the following code into the next cell to pivot the data and create a stacked bar chart of the count of plural births over time:

```
pivot_table = df.pivot(index='year', columns='plurality', values='count')
pivot_table.plot(kind='bar', stacked=True, figsize=(15,7));
```

Next, take a look at baby weight by gender.

In the next cell, enter the following, then run the cell:

```
sql = """
SELECT
    is_male,
    AVG(weight_pounds) AS ave_weight
FROM
    `bigquery-public-data.samples.natality`
GROUP BY
    is_male
"""
df = client.query(sql).to_dataframe()
df.plot(x='is_male', y='ave_weight', kind='bar');
```

Are male babies heavier or lighter than female babies? Does this align with your expectations?

Run a query to get baby weight by gender

Check my progress

For your last visualization, see how the baby's weight fluctuates according to the number of gestation weeks.

Enter the following into the next cell and run it:

```
sql = """
SELECT
    gestation_weeks,
    AVG(weight_pounds) AS ave_weight
FROM
    `bigquery-public-data.samples.natality`
WHERE
    NOT IS NAN(gestation_weeks) AND gestation_weeks <> 99
GROUP BY
    gestation_weeks
ORDER BY
    gestation_weeks
"""
df = client.query(sql).to_dataframe()
df.plot(x='gestation_weeks', y='ave_weight', kind='bar');
```

Note: Because the `gestation_weeks` field allows null values and stores unknown values as 99, this query excludes records where `gestation_weeks` is null or 99.

Now you have a chart that shows how the weight of the baby relates to the number of weeks of gestation.

Run a query to get the weight of the baby which relates to the number of weeks of gestation.

Check my progress

Congratulations!

You learned how to launch AI Platform notebook, run queries with BigQuery, and create a chart to show your results.



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 [Predicting Baby Weight with TensorFlow on Cloud ML Engine](#), or try below one:

- [Weather Data in BigQuery](#)

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 December 25, 2020

Lab Last Tested December 25, 2020

Copyright 2021 Google LLC All rights reserved. Google and the Google logo are trademarks of Google LLC. All other company and product names may be trademarks of the respective companies with which they are associated.