

Perform Foundational Data, ML, and AI Tasks in Google Cloud: Challenge Lab

GSP323



Google Cloud Self-Paced Labs

Overview

In a challenge lab you're given a scenario and a set of tasks. Instead of following step-by-step instructions, you will use the skills learned from the labs in the quest to figure out how to complete the tasks on your own! An automated scoring system (shown on this page) will provide feedback on whether you have completed your tasks correctly.

When you take a challenge lab, you will not be taught new Google Cloud concepts. You are expected to extend your learned skills, like changing default values and reading and researching error messages to fix your own mistakes.

To score 100% you must successfully complete all tasks within the time period!

This lab is recommended for students who have enrolled in the [Perform Foundational Data, ML, and AI Tasks in Google Cloud](#) quest. Are you ready for the challenge?

Topics tested:

- Create a simple Dataproc job
- Create a simple DataFlow job
- Create a simple Dataprep job
- Perform one of the three Google machine learning backed API tasks

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:

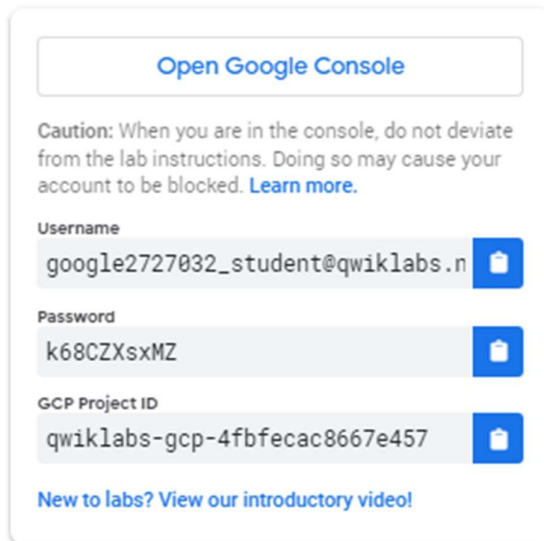
- Access to a standard internet browser (Chrome browser recommended).
- Time to complete the lab.

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.](#)

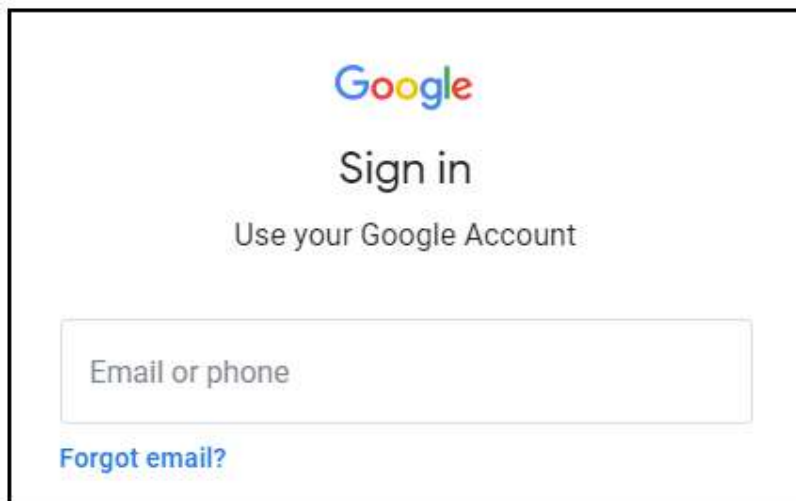
Username
google2727032_student@qwiklabs.n 

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

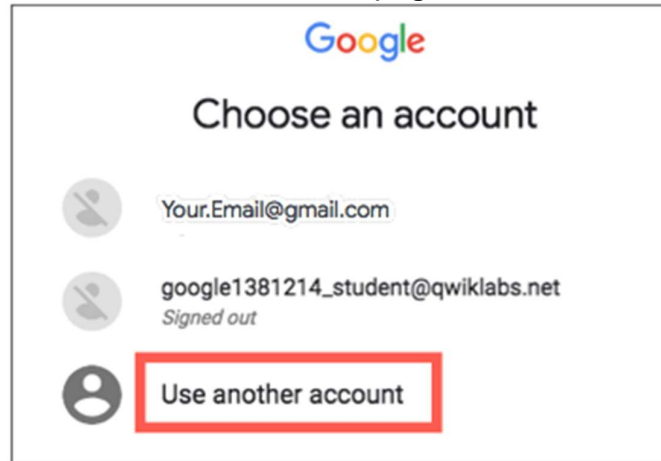
Use your Google Account

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.



Challenge scenario

As a junior data engineer in Jooli Inc. and recently trained with Google Cloud and a number of data services you have been asked to demonstrate your newly learned skills. The team has asked you to complete the following tasks.

You are expected to have the skills and knowledge for these tasks so don't expect step-by-step guides.

Task 1: Run a simple Dataflow job

You have used Dataflow in the question to load data into BigQuery from Pub/Sub, now use the Dataflow batch template **Text Files on Cloud Storage to BigQuery** under "Process Data in Bulk (batch)" to transfer data from a Cloud Storage bucket (`gs://cloud-training/gsp323/lab.csv`). The following table has the values you need to correctly configure the Dataflow job.

You will need to make sure you have:

- created a BigQuery dataset called `lab`
- replace **YOUR_PROJECT** with the lab project name
- created a Cloud Storage Bucket called `YOUR_PROJECT`

Field	Value
JavaScript UDF path in Cloud Storage	<code>gs://cloud-training/gsp323/lab.js</code>
JSON path	<code>gs://cloud-training/gsp323/lab.schema</code>
JavaScript UDF name	<code>transform</code>
BigQuery output table	<code>YOUR_PROJECT:lab.customers</code>
Cloud Storage input path	<code>gs://cloud-training/gsp323/lab.csv</code>
Temporary BigQuery directory	<code>gs://YOUR_PROJECT/bigquery_temp</code>
Temporary location	<code>gs://YOUR_PROJECT/temp</code>

Wait for the job to finish before trying to check your progress.

Click *Check my progress* to verify the objective.

If you don't get a green check mark, click on the **Score** fly-out on the top right and click **Run Step** on the relevant step. A hint pop up opens to give you advice.

Task 2: Run a simple Dataproc job

You have used Dataproc in the quest, now you must run another example Spark job using Dataproc.

Before you run the job, log into one of the cluster nodes and copy the `/data.txt` file into hdfs (use the command `hdfs dfs -cp gs://cloud-training/gsp323/data.txt /data.txt`).

Run a Dataproc job using the values below.

Field	Value
Region	us-central1
Job type	Spark
Main class or jar	org.apache.spark.examples.SparkPageRank
Jar files	file:///usr/lib/spark/examples/jars/spark-examples.jar
Arguments	/data.txt
Max restarts per hour	1

Wait for the job to finish before trying to check your progress.

Click *Check my progress* to verify the objective.

If you don't get a green check mark, click on the **Score** fly-out on the top right and click **Run Step** on the relevant step. A hint pop up opens to give you advice.

Task 3: Run a simple Dataprep job

You have used Dataprep to import data files and transformed them to gain views of the data. Use Dataprep to import one CSV file (described below) that holds data of lab executions.

`gs://cloud-training/gsp323/runs.csv` structure:

runid	userid	labid	lab_title	start	end	time	score	state
5556	545	122	Lab 122	2020-04-09 16:18:19	2020-04-09 17:10:11	3112	61.25	SUCCESS
5557	116	165	Lab 165	2020-04-09 16:44:45	2020-04-09 18:13:58	5353	60.5	SUCCESS
5558	969	31	Lab 31	2020-04-09 17:59:01	2020-04-09 18:02:09	188	0	FAILURE

Perform the following transforms to ensure the data is in the right state:

- Remove all rows with the state of "FAILURE"
- Remove all rows with 0 or 0.0 as a score (Use the regex pattern `/(^0$|^0\.0$)/`)
- Label columns with the names above

Make sure you run the job. You will need to wait until the Dataflow job completes before you can grade this task.

Click *Check my progress* to verify the objective.

If you don't get a green check mark, click on the **Score** fly-out on the top right and click **Run Step** on the relevant step. A hint pop up opens to give you advice.

Task 4: AI

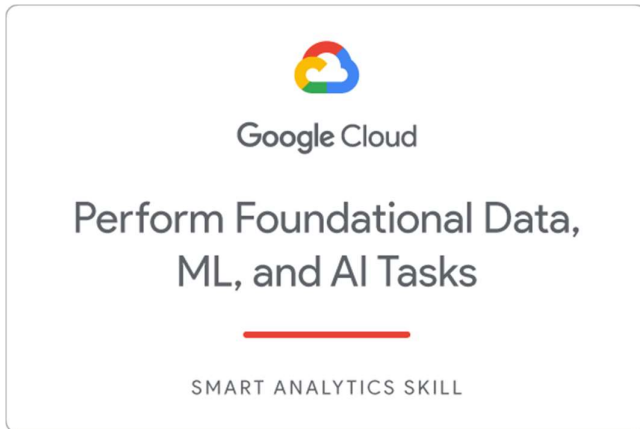
Complete one of the tasks below, **YOUR_PROJECT** must be replaced with your lab project name.

- Use Google Cloud Speech API to analyze the audio file `gs://cloud-training/gsp323/task4.flac`. Once you have analyzed the file you can upload the resulting analysis to `gs://YOUR_PROJECT-marking/task4-gcs.result`.
- Use the Cloud Natural Language API to analyze the sentence from text about Odin. The text you need to analyze is "Old Norse texts portray Odin as one-eyed and long-bearded, frequently wielding a spear named Gungnir and wearing a cloak and a broad hat." Once you have analyzed the text you can upload the resulting analysis to `gs://YOUR_PROJECT-marking/task4-cn1.result`.
- Use Google Video Intelligence and detect all text on the video `gs://spl1s/gsp154/video/train.mp4`. Once you have completed the processing of the video, pipe the output into a file and upload to `gs://YOUR_PROJECT-marking/task4-gvi.result`. Ensure the progress of the operation is complete and the service account you're uploading the output with has the Storage Object Admin role.

Click *Check my progress* to verify the objective.

If you don't get a green check mark, click on the **Score** fly-out on the top right and click **Run Step** on the relevant step. A hint pop up opens to give you advice.

Congratulations!



Earn Your Next Skill Badge

This self-paced lab is part of the [Perform Foundational Data, ML, and AI Tasks in Google Cloud](#) quest. Completing this skill badge quest earns you the badge above, to recognize your achievement. Share your badge on your resume and social platforms, and announce your accomplishment using #GoogleCloudBadge.

This skill badge quest is part of Google Cloud's [Data Analyst](#) and [Data Engineer](#) learning paths. Continue your learning journey by enrolling in the [Engineer Data with Google Cloud](#) quest.

[See other available Qwiklabs Quests](#) available in the catalog.

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 February 23, 2021

Lab Last Tested February 19, 2021

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.

solution: <https://www.youtube.com/watch?v=GQtW5c7Cu5g>

beware: there are 2 **Text Files on Cloud Storage to BigQuery** in dropdown.