Quick tip: Review the prerequisites before you run the lab

End Lab 00:34:32

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.

Open Google Cloud console

Username
student-02-fc646f56522c

Password
Y96PLKTE005P

Project ID
qwiklabs-gcp-00-4d5bb68

Utilize the Streamlit Framework with Cloud Run and the Gemini API in Vertex AI

Lab 1 hour No cost Intermediate

★★★★★

This lab may incorporate AI tools to support your learning.

GSP1229

Previous Next

Google Cloud Self-Paced Labs

Overview

In this lab, you will learn how to build a generative AI application using the Gemini API in Vertex AI and deploying it on Cloud Run. You'll use the Streamlit framework to create an interactive interface for generating stories.

The lab involves running the application locally in Cloud Shell to test its functionality and then deploying it to Cloud Run for scalable serving. You'll gain practical experience integrating Gemini with a user interface and leveraging Cloud Run for efficient deployment.

Gemini is a family of powerful generative AI models developed by Google DeepMind, capable of understanding and generating various forms of content, including text, code, images, audio, and video.

Gemini API in Vertex AI

The Gemini API in Vertex AI provides a unified interface for interacting with Gemini models. This allows developers to easily integrate these powerful AI capabilities into their applications. For the most up-to-date details and specific features of the latest versions, please refer to the official Gemini documentation.

Gemini Models

- Gemini Pro: Designed for complex reasoning, including:
  - Analyzing and summarizing large amounts of information.
  - Sophisticated cross-modal reasoning (across text, code, images, etc.).
  - Effective problem-solving with complex codebases.
- Gemini Flash: Optimized for speed and efficiency, offering:
  - Enhanced multimodal capabilities, including improved spatial understanding, new output modalities (text, audio, images), and native tool use (Google Search, code execution, and third-party functions).

Prerequisites

Before starting this lab, you should be familiar with:

- Basic Python programming.
- General API concepts.
- Running Python code in a Jupyter notebook on Vertex AI Workbench.

Objectives

- Integrate Gemini API in Vertex AI with applications
- Build and deploy the developed sample application on Google Cloud Run
- Use the Streamlit framework to build a Cloud Run application

Setup and requirements

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 are made available to you.

This hands-on lab lets you do the lab activities in a real cloud environment, not in a

To complete this lab, you need:

- Access to a standard internet browser (Chrome browser recommended).

Note: Use an Incognito (recommended) or private browser window to run this lab. This prevents conflicts between your personal account and the student account, which may cause extra charges incurred to your personal account.

- Time to complete the lab—remember, once you start, you cannot pause a lab.

Note: Use only the student account for this lab. If you use a different Google Cloud account, you may incur charges to that account.

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

following:

- The Open Google Cloud console button
- Time remaining
- The temporary credentials that you must use for this lab
- Other information, if needed, to step through this lab

2. Click **Open Google Cloud console** (or right-click and select **Open Link in Incognito Window** if you are running the Chrome browser).

   The lab spins up resources, and then opens another tab that shows the Sign in page.

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

   Note: If you see the **Choose an account** dialog, click **Use Another Account**.

```
student-02-fc646f56522c@qwiklabs.net
```

You can also find the Username in the Lab Details pane.

4. Click **Next**.

5. Copy the **Password** below and paste it into the **Welcome** dialog.

```
Y96PLKTE005P
```

You can also find the Password in the Lab Details pane.

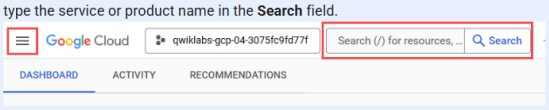
6. Click **Next**.

> **Important:** You must use the credentials the lab provides you. Do not use your Google Cloud account credentials.

charges.

7. 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 Google Cloud console opens in this tab.

> **Note:** To access Google Cloud products and services, click the **Navigation menu** or type the service or product name in the **Search** field.
>
> 
>

Task 1. Build and Deploy the Application to Cloud Run

In this section, you will deploy the Streamlit Application in Cloud Run.

Clone the Repository

1. Open a new Cloud Shell terminal by clicking on the Cloud Shell icon in the top right corner of the Cloud console.
2. Run the following commands to clone the repo and navigate to `gemini-streamlit-cloudrun` directory in Cloud Shell using the following commands.

```
git clone https://github.com/GoogleCloudPlatform/generative-
```

```
cd generative-ai/gemini/sample-apps/gemini-streamlit-cloudrun
```

To run the Streamlit Application on Cloud Run, you will need to perform some additional steps.

Configuration

1. Setup the Python virtual environment and install the dependencies:

```
python3 -m venv gemini-streamlit
source gemini-streamlit/bin/activate
pip install -r requirements.txt
```

2. Your application requires access to two environment variables:

- `GOOGLE_CLOUD_PROJECT` : This the Google Cloud project ID.

These variables are needed since the Vertex AI initialization needs the Google Cloud project ID and the region. The specific code line from the `app.py` function is shown here: `vertexai.init(project=PROJECT_ID, location=LOCATION)`

In Cloud Shell, execute the following commands:

```
GOOGLE_CLOUD_PROJECT='qwiklabs-gcp-00-4d5bb68b8b6c'
GOOGLE_CLOUD_REGION='us-central1'
```

3. You will now build the Docker image for the application and push it to Artifact Registry. To do this, you will need one environment variable set that will point to the Artifact Registry name. The commands below will create this Artifact Registry repository for you.

Note: This step will take several minutes to complete.

4. In Cloud Shell, execute the following command:

```
SERVICE_NAME='gemini-streamlit-app'
gcloud artifacts repositories create "$AR_REPO" --
location="$GOOGLE_CLOUD_REGION" --repository-format=Docker
gcloud builds submit --tag "$GOOGLE_CLOUD_REGION-
docker.pkg.dev/$GOOGLE_CLOUD_PROJECT/$AR_REPO/$SERVICE_NAME"
```

Output:

```
DONE
--------------------------------------------------------------------------------
ID                                    CREATE_TIME                DURATION
a601ffd1-c282-43d2-942c-53cc13f43bf2  2023-12-18T11:37:30+00:00  2M29S
```

6. The final step is to deploy the service in Cloud Run with the image that we had built and had pushed to the Artifact Registry in the previous step.

```
gcloud run deploy "$SERVICE_NAME" \
  --port=8080 \
  --allow-unauthenticated \
  --region=$GOOGLE_CLOUD_REGION \
  --platform=managed  \
  --project=$GOOGLE_CLOUD_PROJECT \
  --set-env-
vars=GOOGLE_CLOUD_PROJECT=$GOOGLE_CLOUD_PROJECT,GOOGLE_CLOUD_REGION
```

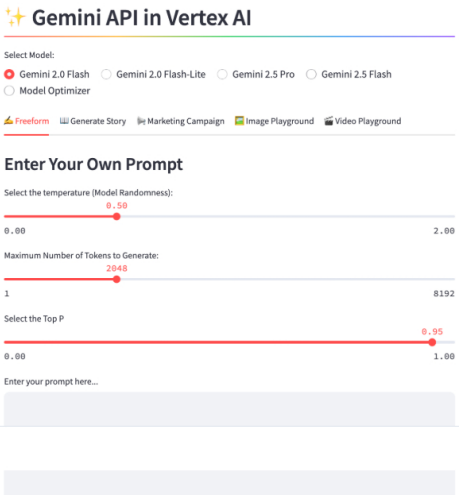
On successful deployment, you will be provided a URL to the Cloud Run service. You can visit that in the browser to view the Cloud Run application that you just deployed.

Output:

```
⠹ Deploying new service... Done.

✓ Deploying new service... Done.

Done.
Service [gemini-streamlit-app] revision [gemini-streamlit-app-00001-srg] h
```



Generate

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

Build and Deploy the Application to Cloud Run

Check my progress

Assessment Completed!

Congratulations!

with applications and build and deploy the developed sample application on Google Cloud Run.

Next steps / learn more

Check out the following resources to learn more about Gemini:

- Gemini Overview
- Generative AI on Vertex AI Documentation
- Generative AI on YouTube
- Explore the Vertex AI Cookbook for a curated, searchable gallery of notebooks for Generative AI.
- Explore other notebooks and samples in the Google Cloud Generative AI repository.

Google Cloud training and 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 May 14, 2025

Lab Last Tested May 14, 2025

with Gemini and