

## MATERIAL DE APOIO

### Google Cloud Skill Boost

#### Laboratório 02 do Advanced: Introdução ao Vector Search e Embeddings

## **Sumário**

1.	Login no Console do Google Cloud .....	3
2.	Tarefa 1 .....	5
3.	Tarefa 2 .....	8

## 1. Login no Console do Google Cloud

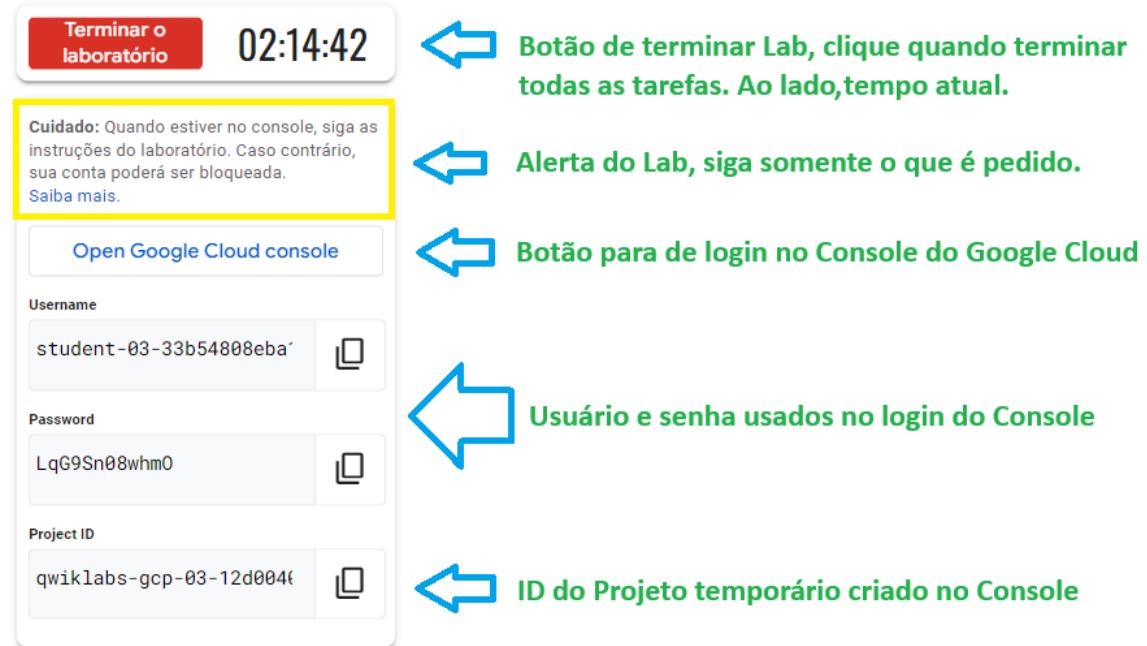
Link para o laboratório:

[https://www.cloudskillsboost.google/parts/236/course\\_templates/978/labs/488168?locale=pt\\_BR](https://www.cloudskillsboost.google/parts/236/course_templates/978/labs/488168?locale=pt_BR)

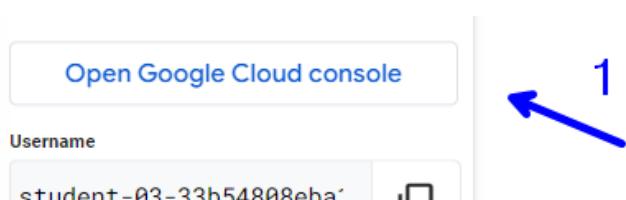
- 1) Primeiro passo é fazer login no Console do Google Cloud, clique no botão verde “Começar o laboratório”:



- 2) Após, irá abrir o Painel de login no Console:



- 3) Clique no botão “Open Google Cloud Console” (1) com o botão direito do mouse e escolha para abrir em uma aba Anônima/InPrivate, depois preencha o usuário (2) e senha (3) na página de login:





## Fazer login

Use sua Conta do Google

2

E-mail ou telefone  
student-03-33b54808eba1@qwiklabs.net

[Esqueceu seu e-mail?](#)



## Olá!

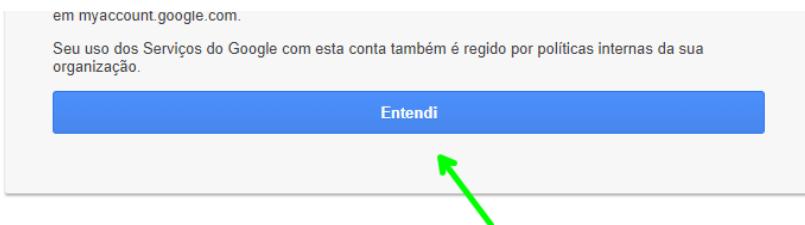
student-03-33b54808eba1@qwiklabs.net

3

Digite sua senha  
LqG9Sn08whmO

Mostrar senha

### 4) Aceite todos os termos e condições do Google Cloud (1-2):



## Google Cloud

Welcome student fe2a879d!

Create and manage your Google Cloud instances, disks, networks, and other resources in one place.

student fe2a879d  
student-03-33b54808eba1@qwiklabs.net [SWITCH ACCOUNT](#)

Country

Brazil

Terms of Service

I agree to the [Google Cloud Platform Terms of Service](#), and the terms of service of [any applicable services and APIs](#).

1

Email updates

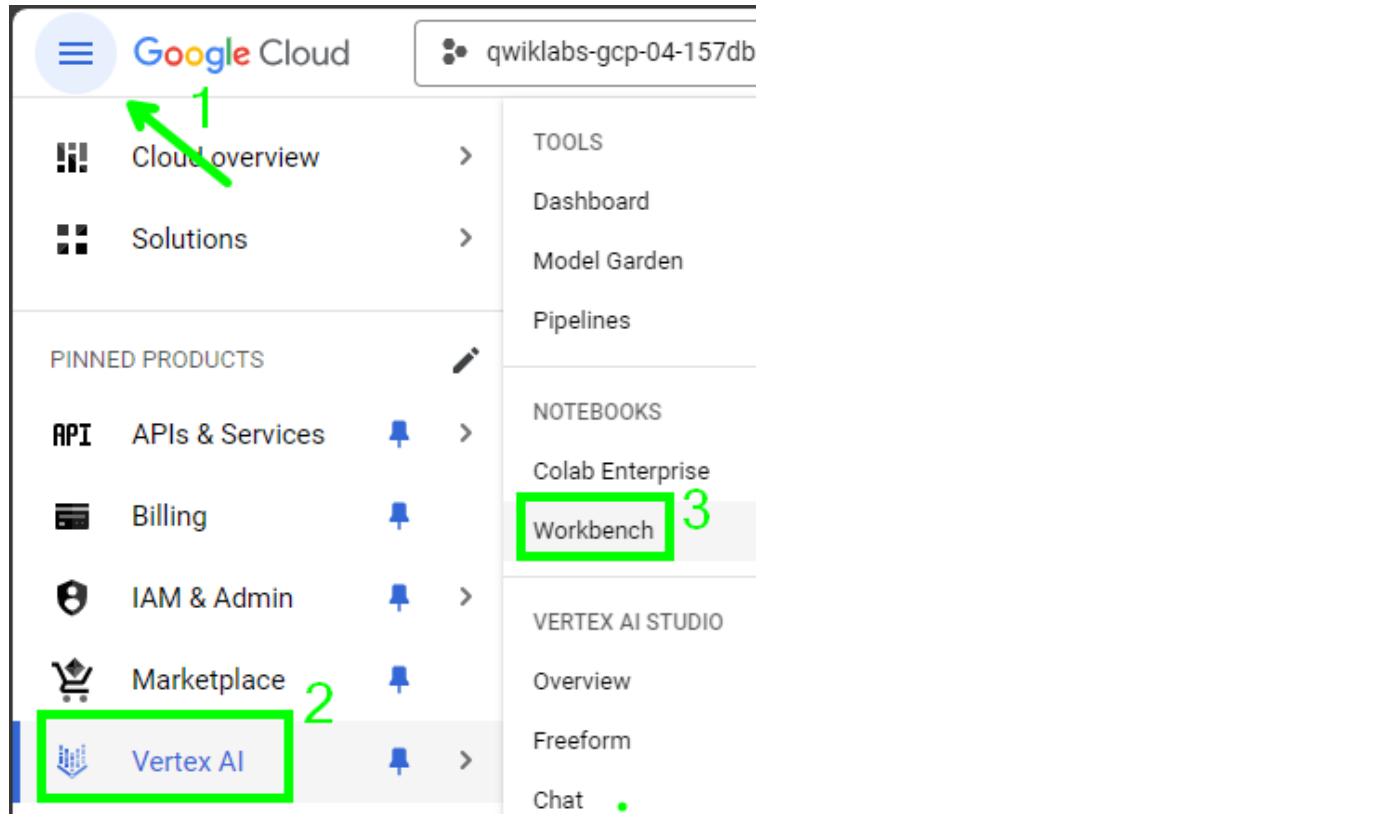
I would like to receive periodic emails on news, product updates and special offers from Google Cloud and Google Cloud Partners.

2

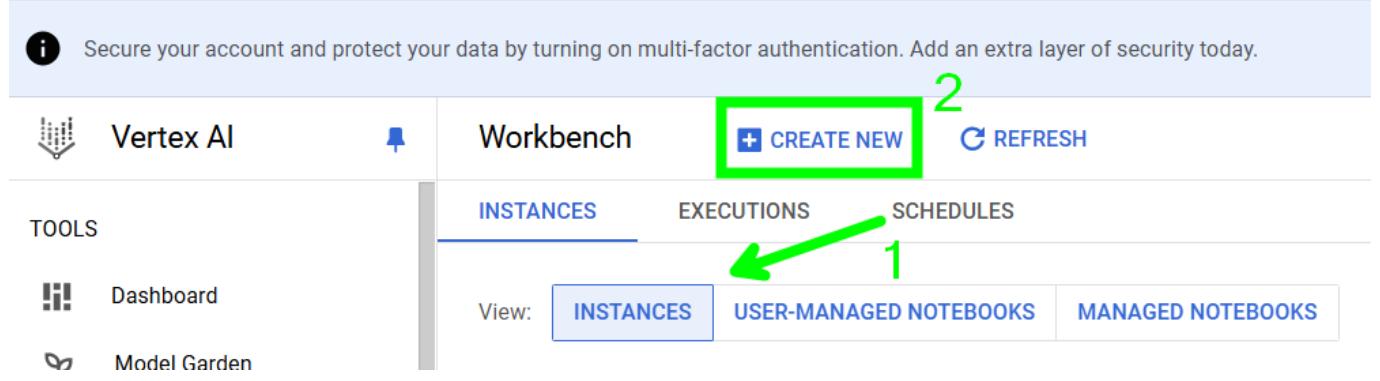
[AGREE AND CONTINUE](#)

## 2. Tarefa 1

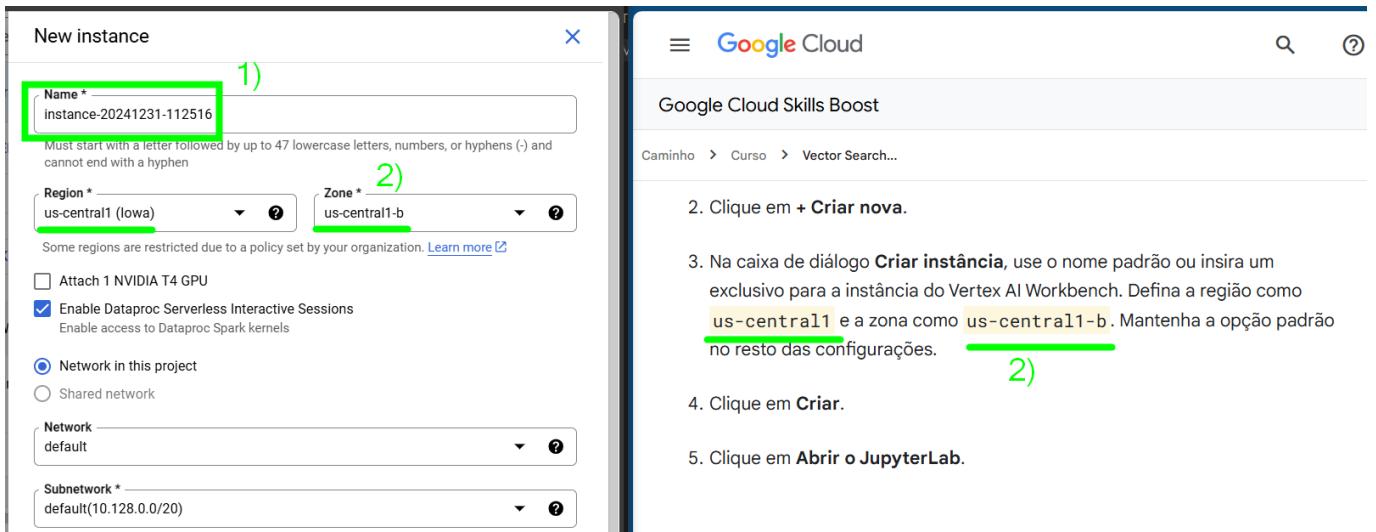
- 1) Depois que entrar no Console, clique no Menu (1) no canto superior esquerdo, procure o recurso “Vertex AI” (2), e clique em “Workbench” (3):



- 2) Em **Workbench**, clique na aba do menu **INSTANCES** (1), e depois em **CREATE NEW** (2).



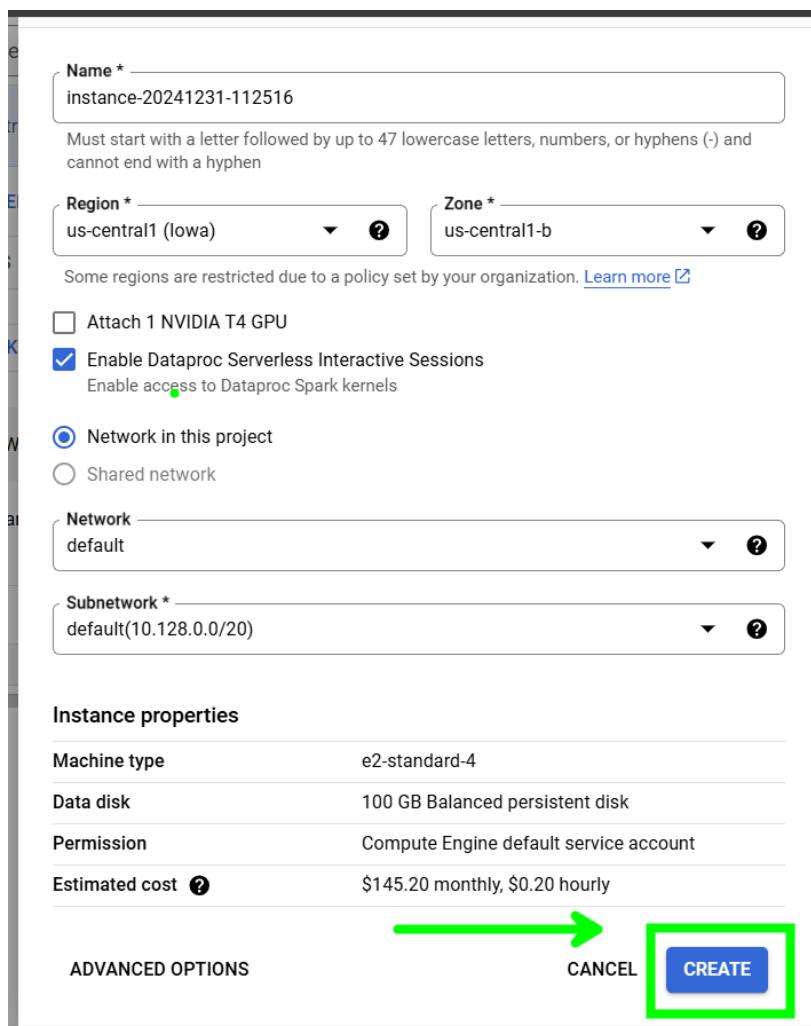
- 3) Na janela que se abre, mantenha o nome padrão como aparece (1), e garanta que a região e a zona esteja igual a qual o laboratório pede (2):



The image shows the 'New instance' dialog on the left and the 'Google Cloud Skills Boost' navigation bar on the right. The 'New instance' dialog has several fields highlighted with green boxes and numbers: 1) 'Name' field containing 'instance-20241231-112516'; 2) 'Zone' field set to 'us-central1-b'. The 'Google Cloud Skills Boost' bar shows the path: 'Caminho > Curso > Vector Search...'. Below the dialog, numbered steps 2 through 5 are listed:

2. Clique em **+ Criar nova**.
3. Na caixa de diálogo **Criar instância**, use o nome padrão ou insira um exclusivo para a instância do Vertex AI Workbench. Defina a região como **us-central1** e a zona como **us-central1-b**. Mantenha a opção padrão no resto das configurações.
4. Clique em **Criar**.
5. Clique em **Abrir o JupyterLab**.

- 4) Mantenha todo o resto como padrão e clique em **CREATE** no botão azul abaixo da configuração:



The image shows the 'New instance' configuration dialog with the following details:

- Name:** instance-20241231-112516
- Region:** us-central1 (Iowa)
- Zone:** us-central1-b
- Attachments:** Attach 1 NVIDIA T4 GPU (unchecked)
- Enable Dataproc Serverless Interactive Sessions:** checked
- Network:** Network in this project (selected)
- Subnetwork:** default(10.128.0.0/20)

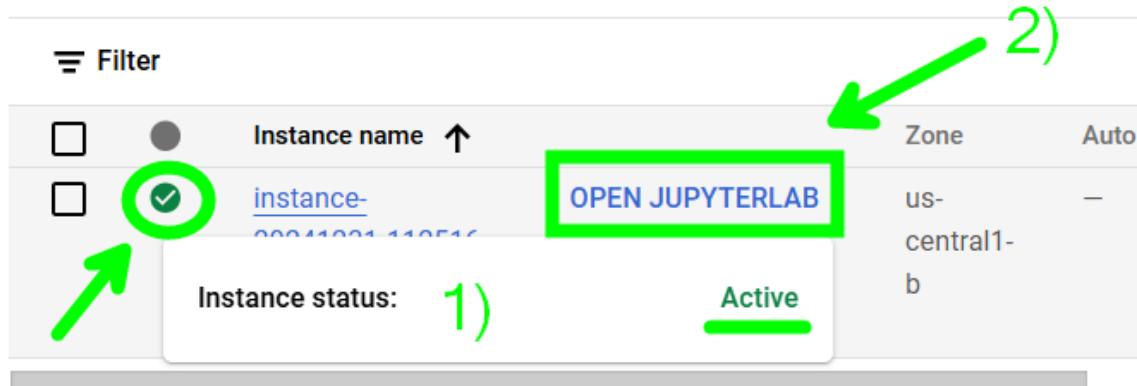
**Instance properties:**

- Machine type:** e2-standard-4
- Data disk:** 100 GB Balanced persistent disk
- Permission:** Compute Engine default service account
- Estimated cost:** \$145.20 monthly, \$0.20 hourly

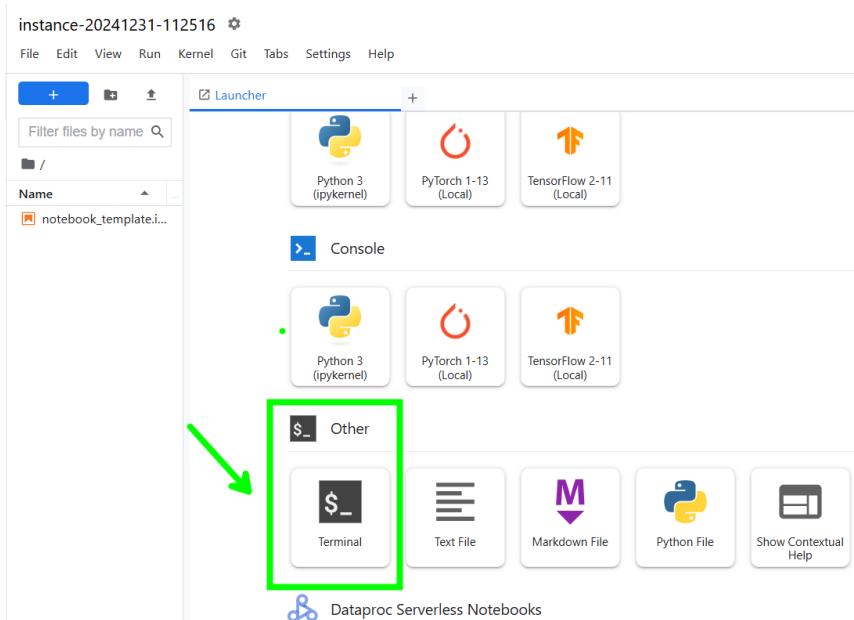
At the bottom, there are 'ADVANCED OPTIONS', 'CANCEL', and a large green arrow pointing to the 'CREATE' button, which is highlighted with a green box.

- 5) Aguarde até que a instância seja criada, você verá uma bola verde ticada no indicativo do status, ele estará como “Active” (1). Após, clique no botão **OPEN JUPYTERLAB** (2):

Workbench Instances have JupyterLab 3 pre-installed and are configured with GPU-enabled machine learning frameworks. [Learn more](#)

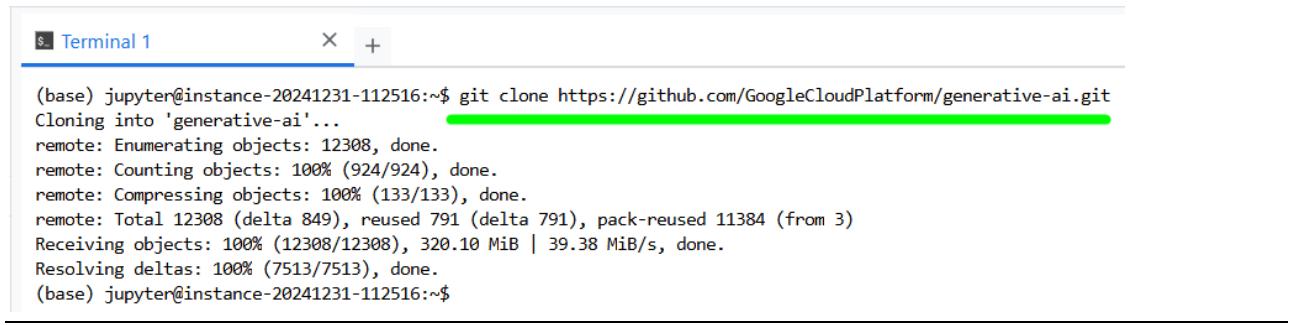


- 6) Quando estiver dentro do JupyterLab, clique no **TERMINAL**:



- 7) Execute o comando que aparece no laboratório (copie da página do laboratório se tiver erros ao copiar com a formatação do documento):

➔ git clone <https://github.com/GoogleCloudPlatform/generative-ai.git>

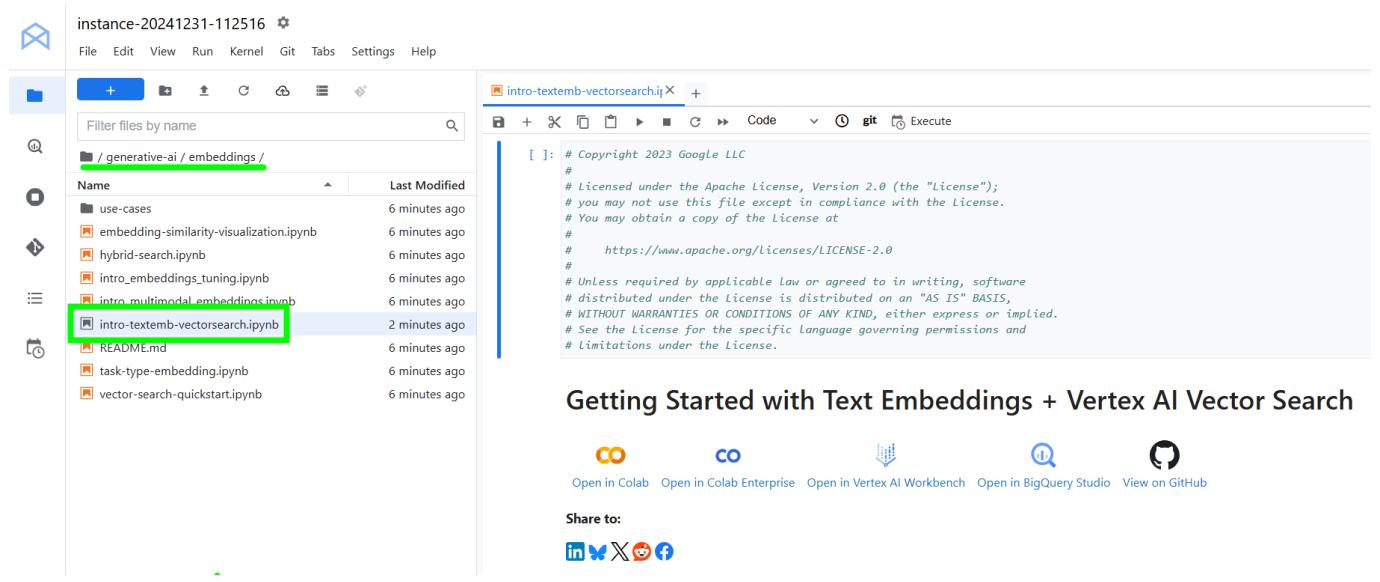


```

Terminal 1
(base) jupyter@instance-20241231-112516:~$ git clone https://github.com/GoogleCloudPlatform/generative-ai.git
Cloning into 'generative-ai'...
remote: Enumerating objects: 12308, done.
remote: Counting objects: 100% (924/924), done.
remote: Compressing objects: 100% (133/133), done.
remote: Total 12308 (delta 849), reused 791 (delta 791), pack-reused 11384 (from 3)
Receiving objects: 100% (12308/12308), 320.10 MiB | 39.38 MiB/s, done.
Resolving deltas: 100% (7513/7513), done.
(base) jupyter@instance-20241231-112516:~$
```

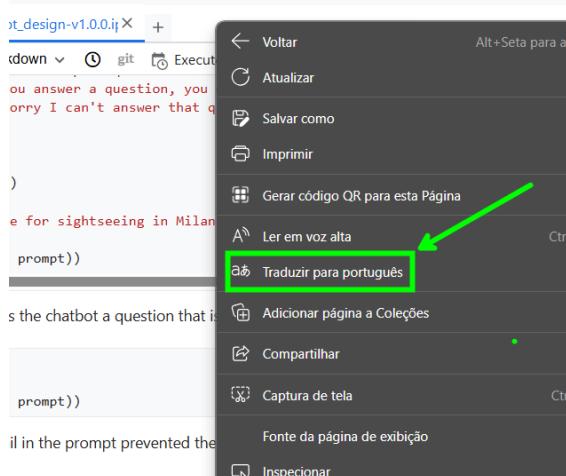
### 3. Tarefa 2

- Do lado esquerdo onde estão localizados os arquivos (File Browser), clique duas vezes no arquivo que está no caminho (Se pedir para selecionar o kernel, selecione **Python3**):
  - generative-ai / embeddings / intro-textemb-vectorsearch.ipynb



The screenshot shows a Jupyter Notebook interface. On the left, the 'File Browser' pane displays a list of files in the 'generative-ai/embeddings/' directory. The file 'intro-textemb-vectorsearch.ipynb' is selected and highlighted with a green box. On the right, the notebook cell for 'intro-textemb-vectorsearch.ipynb' is running, showing the Apache License 2.0 header. Below the notebook, a title 'Getting Started with Text Embeddings + Vertex AI Vector Search' is visible, along with sharing and collaboration links.

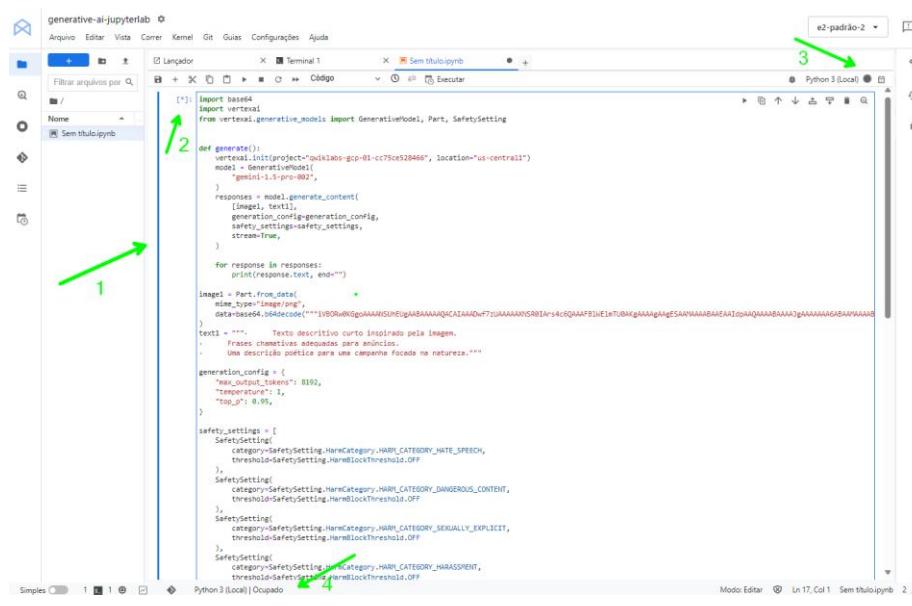
- Se desejar, clique com o botão direito do mouse dentro do notebook, e traduza o conteúdo para o português:



The screenshot shows a Jupyter Notebook cell with a context menu open. The menu includes options like 'Voltar', 'Atualizar', 'Salvar como', 'Imprimir', 'Gerar código QR para esta Página', 'Ler em voz alta', 'Traduzir para português' (which is highlighted with a green box and a red arrow), 'Adicionar página a Coleções', 'Compartilhar', 'Captura de tela', 'Fonte da página de exibição', and 'Inspecionar'. The notebook cell contains some code and text in English.

3) Em um notebook, a célula é a caixa de comando que se quer rodar, mantenha a barra azul lateral (1) sempre no comando que se quer rodar clicando na célula desejada. A célula pode ser rodada com o botão de **Play**, ou **CTRL + Botão Esquerdo do Mouse** (Roda e se mantém na célula), ou **Shift + Botão Esquerdo do Mouse** (Roda e segue para a próxima célula).

Há alguns indicativos de que a célula está rodando, é **preciso esperar** que ela termine antes de rodar o próximo comando. Se houver um **asterisco** (2) na frente da célula, se o círculo estiver **preenchido** (3), ou se estiver **ocupado** (4), deve-se esperar que o comando seja completado:



4) Na sessão **Embeddings de texto em ação**, coloque as informações de **Project ID** e **Location** que aparecem para você no laboratório, como no exemplo abaixo (Pule Set IAM Permissions):

Google Cloud Skills Boost

Caminho > Curso > Vector Search...

3. No painel de navegação à esquerda, navegue até a pasta `generative-ai/embeddings` e abra o notebook `intro-textemb-vectorsearch.ipynb`.

4. No comando **Selecionar Kernel**, deixe o **Python3** como padrão e clique em **Selecionar**.

5. Role para baixo até a seção **Embeddings de texto em ação** e execute as células de configuração.

6. Ao definir suas variáveis de ambiente, use `us-central1` para o local e `qwiklabs-gcp-02-718de790fc08` para o ID do projeto.

**Observação:** é possível pular as células do notebook que tenham a indicação *Colab only*.

6. Pule a seção **Definir permissões do IAM** porque sua conta de serviço já tem as permissões necessárias.

```

# intro-textemb-vectorsearch.ipynb
# Environment variables
# Sets environment variables. If asked, please replace the following [your-project-id] with your project ID and run it.

[1]: # get project ID
PROJECT_ID = ! gcloud config get project
PROJECT_ID = PROJECT_ID[0]
LOCATION = "us-central1"
# If PROJECT_ID == "(unset)":
#     print("Please set the project ID manually below")

# define project information
# If PROJECT_ID == "(unset)":
#     PROJECT_ID = "qwiklabs-gcp-02-718de790fc08" # @param {type:"string"}

# generate an unique id for this session
from datetime import datetime
UID = datetime.now().strftime("%Y%m%d%H%M")

```

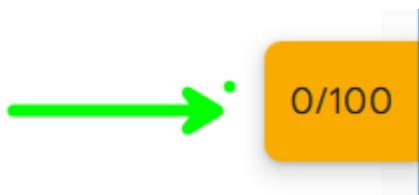
5) A seguir rode todas as células até o final, uma a uma, sempre esperando o processamento terminar como indicado anteriormente no **Item 3**), sempre leia e analise o output da resposta esperada pelo comando.

6) O que esse notebook faz é criar um index para os embeddings associados aos chunks extraídos dos documentos, que demorará longo tempo para ser criado. Você pode acompanhar a criação através do console como indicado no notebook. Depois que o index for criado, o endpoint será criado, que **também irá demorar bastante tempo** para ser criado. A criação dele também pode ser acompanhada pelo console, se toda forma é só esperar o notebook processar o comando sem fazer nada nesse meio tempo.

**OBS:** Se o endpoint apresentar erro ao ser criado, é só executar o comando novamente que dará certo, ele será criado com sucesso e não demorará tanto tempo como da primeira vez.

7) Não precisa executar o Cleaning Up ao final.

8) Assim que terminar, clique em todos os botões de progresso no laboratório, garanta que o progresso esteja em 100% na aba flutuante amarela do lado esquerdo da página:



Somente após 100/100, termine o laboratório no botão vermelho do painel.