

## Resolução do desafio – Laboratório 4 (<https://www.cloudskillsboost.google/quests/117>)

### Tarefa 1: execute um job simples com o Dataflow

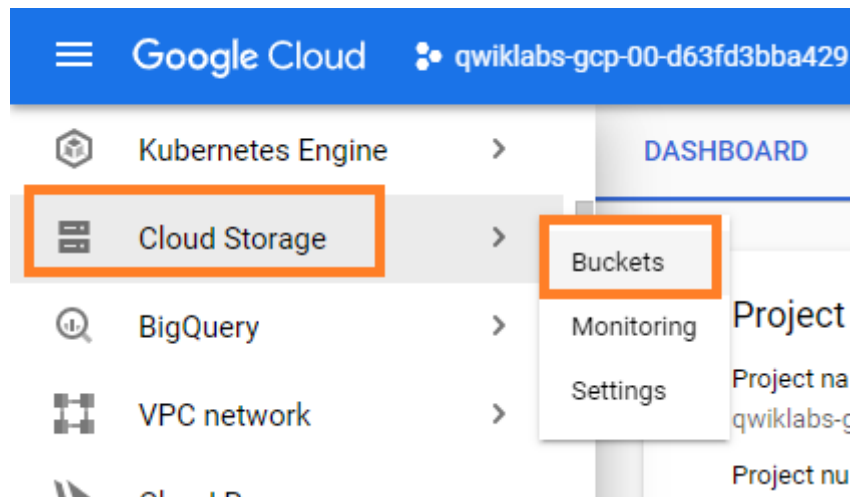
Você usou o Dataflow na Quest para carregar dados do Pub/Sub no BigQuery. Agora use o modelo de lote do Dataflow **Arquivos de texto do Cloud Storage para o BigQuery**, que está em "Processar dados em massa (lote)", para transferir dados de um bucket do Cloud Storage (`gs://cloud-training/gsp323/lab.csv`). Veja na tabela a seguir os valores necessários para configurar corretamente o job do Dataflow.

Você precisará certificar-se que:

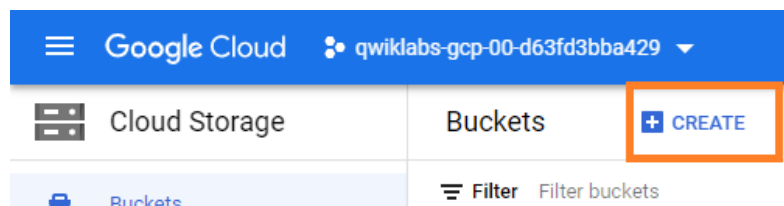
- criou um conjunto de dados do BigQuery com o nome `lab_274` ;
- criou um bucket do Cloud Storage com o nome `3c553672-3cff-4485-a755-5cbc6ed6c59e` .

Acessar o console e ir ao painel de serviços do lado esquerdo da tela:

Procurar por “Cloud Storage” e clicar em “Buckets”.

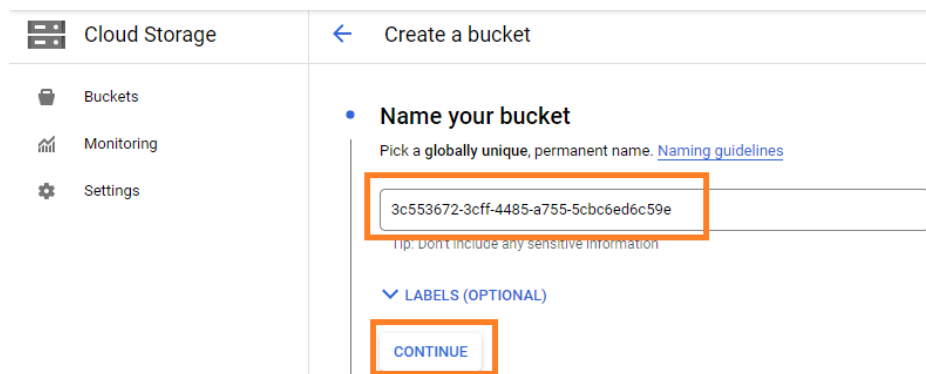


Clique em “Create”



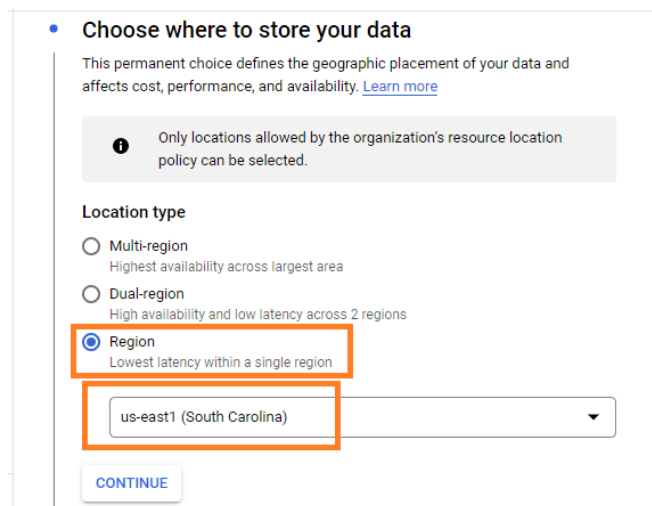
Informe o nome conforme especificado no seu laboratório.

Clique em “Continue”



Escolha a opção “Region” para indicar uma única região.

Informe a região que foi indicada no seu laboratório. A região está informada mais abaixo no painel da esquerda, no mesmo local aonde você encontra o usuário e senha para acesso ao Console.



• **Choose where to store your data**

This permanent choice defines the geographic placement of your data and affects cost, performance, and availability. [Learn more](#)

**Location type**

- ☐ Multi-region  
Highest availability across largest area
- ☐ Dual-region  
High availability and low latency across 2 regions
- ☒ **Region**  
Lowest latency within a single region

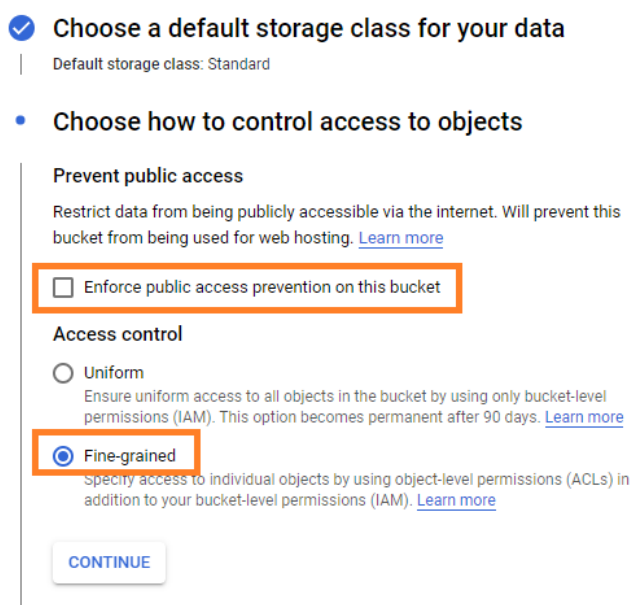
us-east1 (South Carolina)

[CONTINUE](#)

Clique em “Continue” duas vezes até chegar na seção de configuração de acesso.

Desmarque a caixa indicada com **“Enforce public access prevention on this bucket”**

Selecione a opção do **“Fine-grained”**



✓ **Choose a default storage class for your data**

Default storage class: Standard

• **Choose how to control access to objects**

**Prevent public access**

Restrict data from being publicly accessible via the internet. Will prevent this bucket from being used for web hosting. [Learn more](#)

☐ **Enforce public access prevention on this bucket**

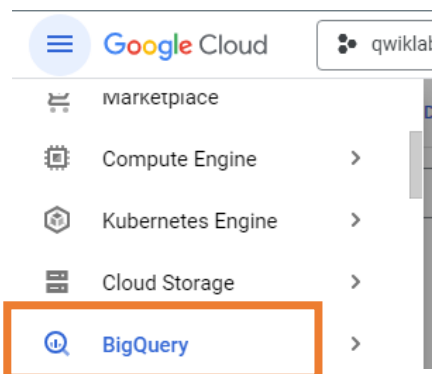
**Access control**

- ☐ Uniform  
Ensure uniform access to all objects in the bucket by using only bucket-level permissions (IAM). This option becomes permanent after 90 days. [Learn more](#)
- ☒ **Fine-grained**  
Specify access to individual objects by using object-level permissions (ACLs) in addition to your bucket-level permissions (IAM). [Learn more](#)

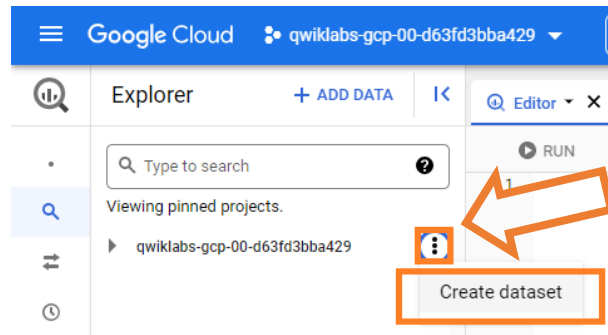
[CONTINUE](#)

Clique em “Create” ao final da página para criar o Bucket.

Depois de criar o Bucket, no menu da esquerda procure por BigQuery. (Será preciso descer a lista de serviços)



Na próxima tela, clique nos três pontos na frente do nome do ID do projeto e depois selecione “Create Dataset”



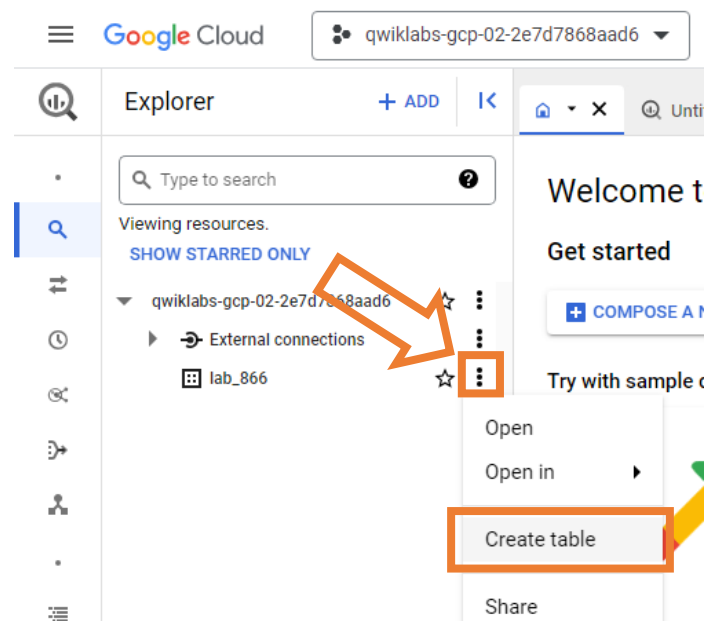
Informe o “DatasetID” conforme o seu laboratório.

Indique a região conforme o seu laboratório.

Clique em “Create dataset”

A screenshot of the 'Create dataset' form in Google Cloud. The form has several sections: 'Project ID' (qwiklabs-gcp-02-2e7d7868aad6), 'Dataset ID' (lab\_866), 'Location type' (Region), 'Region' (us-central1 (Iowa)), 'Default table expiration' (with a checkbox for 'Enable table expiration'), and 'Advanced options' (expanded). The 'CREATE DATASET' button is highlighted with an orange box, and an orange arrow points to it from the left.

Depois clique sobre os 3 pontos na frente do Dataset que foi montado e escolha a opção “Create table”.



O nome da tabela você encontra **no final da seguinte linha do seu laboratório (copie somente o nome que está depois do ponto final)**:

Tabela de saída do BigQuery	qwiklabs-gcp-02-2e7d7868aad6:lab_866.customers_519
-----------------------------	--

Preencha o nome da tabela no campo indicado abaixo:

Create table

Source

Create table from  
Empty table

Destination

Project \*  
qwiklabs-gcp-02-2e7d7868aad6

Dataset \*  
lab\_866

Table \*  
customers\_519

Unicode letters, marks, numbers, connectors, dashes or spaces allowed.

Table type  
Native table

Schema

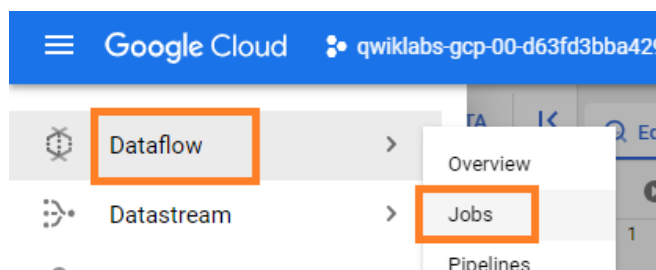
☐ Edit as text

CREATE TABLE CANCEL

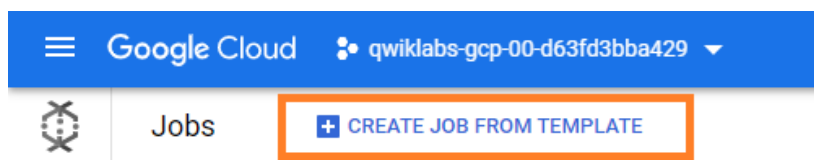
Clique em “Create table” para finalizar.

No menu da esquerda procure por “Dataflow” (será necessário descer a lista de serviços).

Clique em “Jobs”



Clique na opção de “Create a job from template”.



Dê um nome à sua escolha para o campo “Job name”.

**Indique a região conforme o seu laboratório.**

**Escolha o modelo do “Dataflow template” como “Text Files on Cloud Storage to BigQuery” da categoria “Process Data in Bulk (batch)”**

Atenção! Existem dois modelos com o mesmo nome, porém o correto é o indicado na categoria “Process Data in Bulk (batch)”

Google Cloud | qwiklabs-gcp-01-4ea01dbcdbae | Search | Create job from template

Job name \*  
jobtarefa1  
Must be unique among running jobs

Regional endpoint \*  
us-west1 (Oregon)  
Choose a Dataflow regional endpoint to deploy worker instances and store job metadata. You can optionally deploy worker instances to any available Google Cloud region or zone by using the worker region or worker zone parameters. Job metadata is always stored in the Dataflow regional endpoint. [Learn more](#)

Dataflow template \*  
Filter: text Files on Cloud Storage to BigQuery  
Process Data Continuously (stream)  
Text Files on Cloud Storage to BigQuery **X**  
Text Files on Cloud Storage to BigQuery  
CANCEL OK

Na mesma tela, um pouco mais abaixo, preencha os demais campos conforme o seu laboratório

Required Parameters

Cloud Storage Input File(s) \*  
☒ gs:// cloud-training/gsp323/lab.csv BROWSE  
Path of the file pattern glob to read from. (Example: gs://your-bucket/path/\*.csv)

Cloud Storage location of your BigQuery schema file described as a JSON \*  
☒ gs:// cloud-training/gsp323/lab.schema BROWSE  
JSON file with BigQuery Schema description. JSON Example: { "BigQuery Schema": [ { "name": "location", "type": "STRING" }, { "name": "name", "type": "STRING" }, { "name": "age", "type": "STRING" }, { "name": "color", "type": "STRING" }, { "name": "coffee", "type": "STRING" } ] }

BigQuery output table \*  
☒ qwiklabs-gcp-03-be7121296192:lab\_105.customers\_226 BROWSE

Temporary directory for BigQuery loading process \*  
☒ gs:// qwiklabs-gcp-03-be7121296192-marking/bigquery\_temp BROWSE  
Temporary directory for BigQuery loading process (Example: gs://your-bucket/your-files/temp\_dir)

Temporary location \*  
☒ gs:// qwiklabs-gcp-03-be7121296192-marking/temp BROWSE  
Path and filename prefix for writing temporary files. Ex: gs://your-bucket/temp

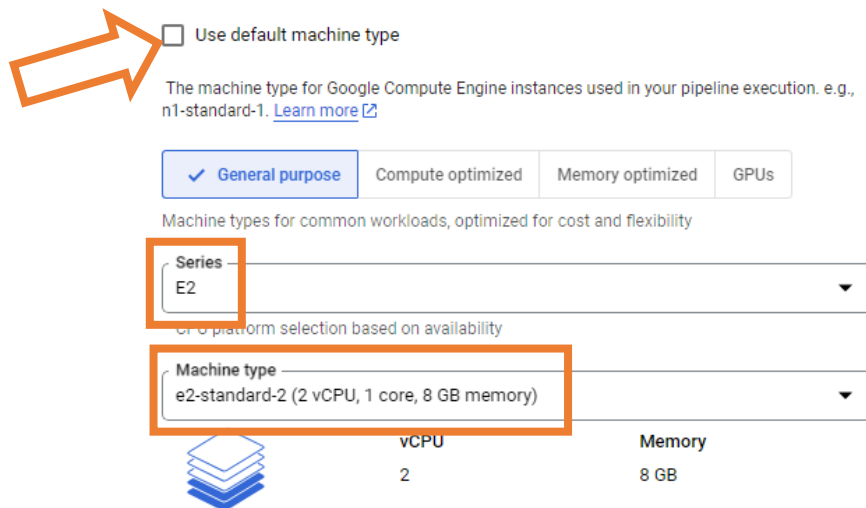
Clique em “Optional Parameters” e preencha os campos abaixo conforme especificado em seu laboratório.

Optional Parameters

JavaScript UDF path in Cloud Storage  
☒ gs:// cloud-training/gsp323/lab.js BROWSE  
The Cloud Storage path pattern for the JavaScript code containing your user-defined functions.

JavaScript UDF name  
transform  
The name of the function to call from your JavaScript file. Use only letters, digits, and underscores. (Example: transform\_udf1)

Desça um pouco a tela e procure pela opção “Use default machine type”. Desmarque essa opção e preencha os campos abaixo **conforme seu laboratório indicar**.



☐ Use default machine type

The machine type for Google Compute Engine instances used in your pipeline execution. e.g., n1-standard-1. [Learn more](#)

**General purpose** Compute optimized Memory optimized GPUs

Machine types for common workloads, optimized for cost and flexibility

Series: E2


Machine type: e2-standard-2 (2 vCPU, 1 core, 8 GB memory)

	vCPU	Memory
	2	8 GB

Clique em “Run Job” e aguarde. (o processo leva um tempo, demora alguns minutos)

Após isso, aguarde o processo e valide a primeira tarefa.

Clique em *Verificar meu progresso* para ver o objetivo.



Execute um job simples com o Dataflow

Verificar meu progresso

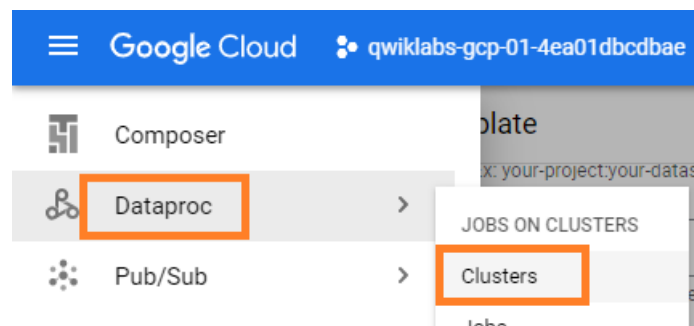
*Assessment Completed!*

## Tarefa 2: execute um job simples com o Dataproc

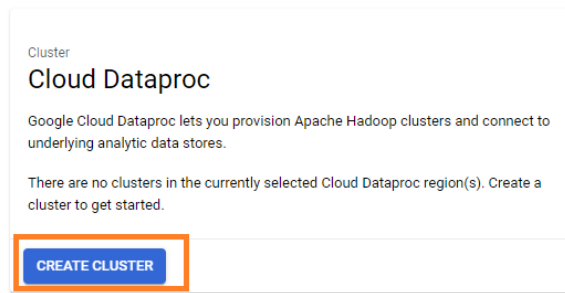
Você já usou o Dataproc na Quest e agora executará um job do Spark com ele.

Antes disso, faça login em um dos nós do cluster e copie o arquivo /data.txt no hdfs. Use o comando `hdfs dfs -cp gs://cloud-training/gsp323/data.txt /data.txt`.

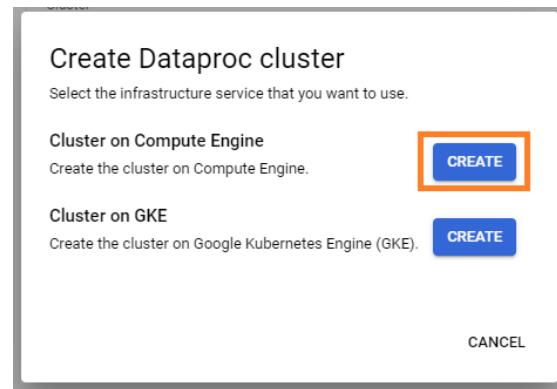
No menu da esquerda procure por “Dataproc” e depois clique em “Clusters”. (Será necessário descer na lista de serviços).



Clique no botão “Create a cluster”

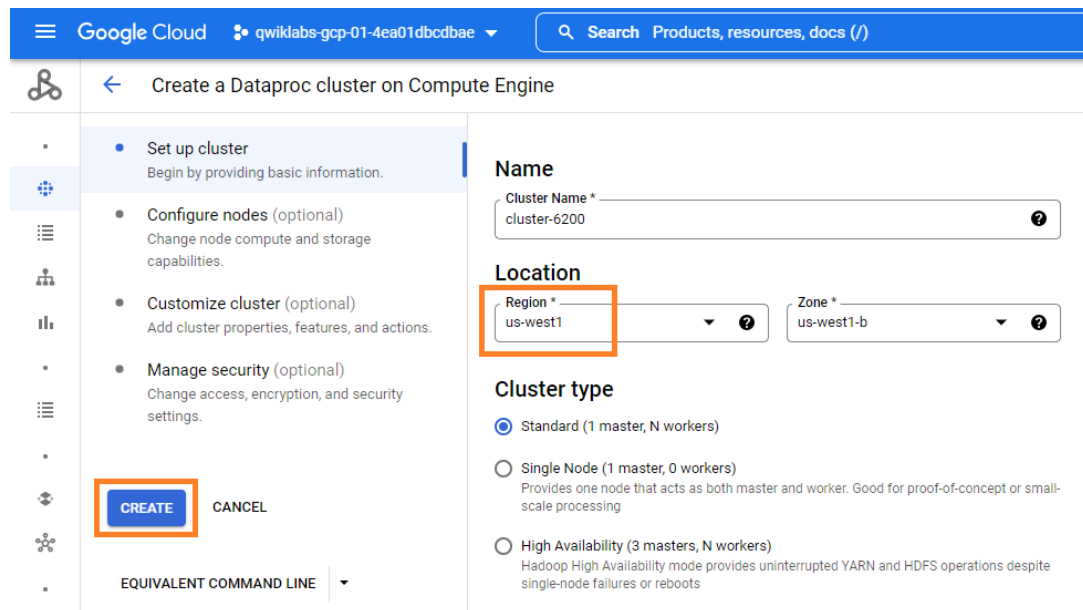


Selecione a primeira opção.



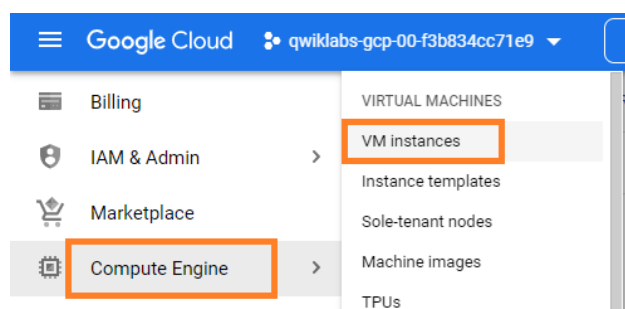
Altere somente a região para a qual o seu laboratório foi indicado.

Depois clique em “Create”

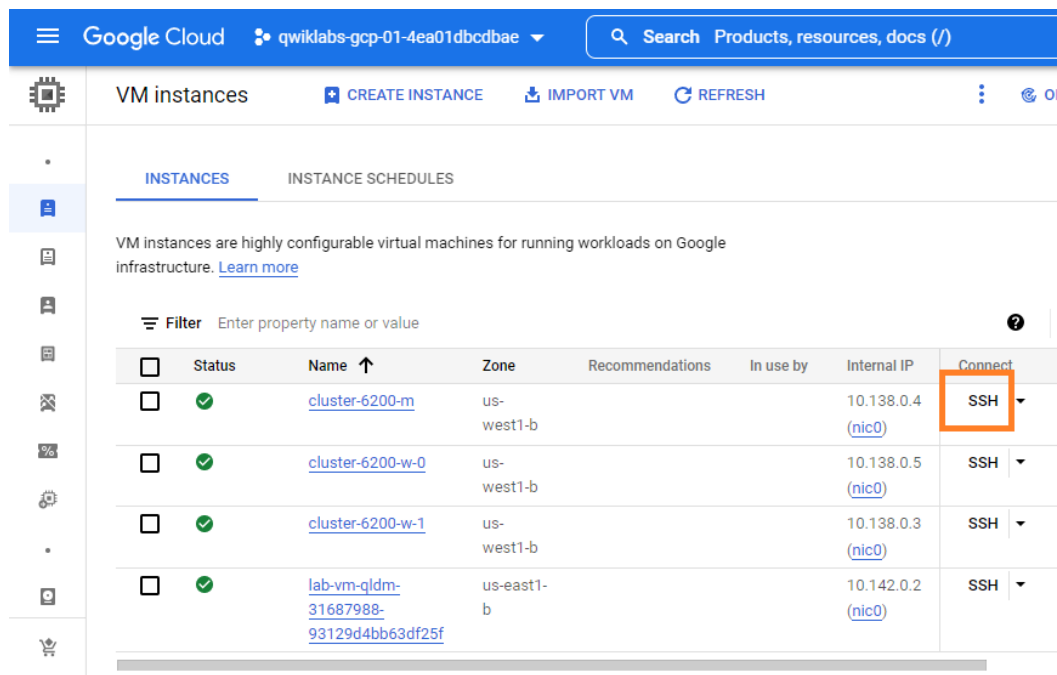


Aguarde a criação do cluster finalizar.

Volte no menu da esquerda em “Compute engine” e depois clique em “VM instances”

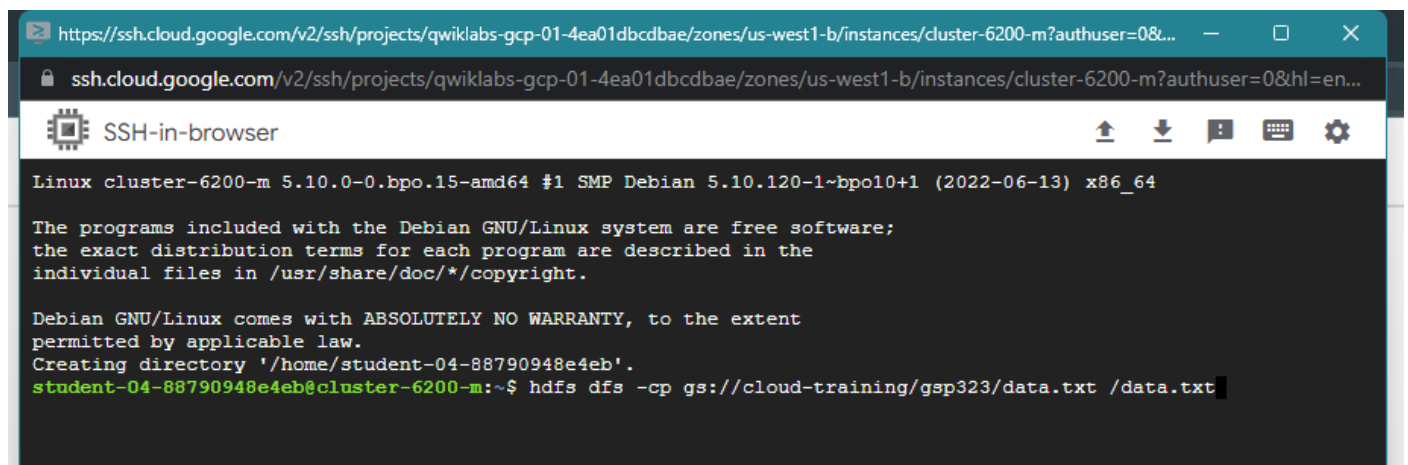


Escolha uma instância associada ao cluster e clique em SSH para iniciar a conexão.

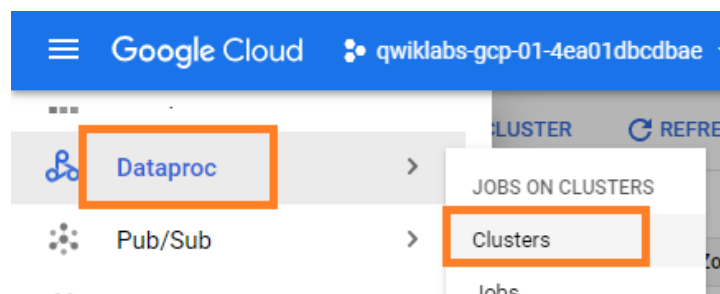


Faça uma cópia do arquivo data.txt digitando o comando abaixo na tela do terminal SSH que foi aberta:

```
hdfs dfs -cp gs://cloud-training/gsp323/data.txt /data.txt
```

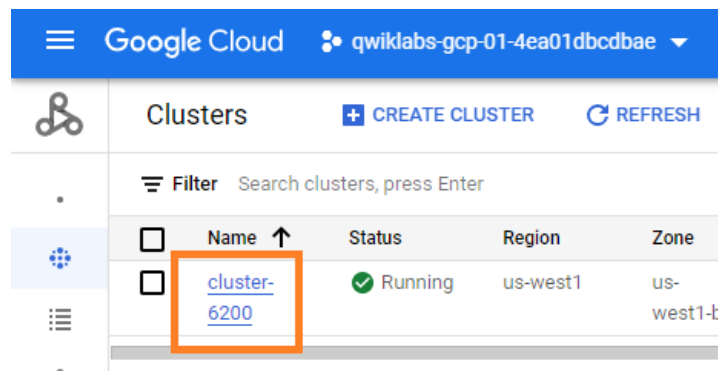


No menu da esquerda, volte em “Dataproc” e clique em “Cluster”.

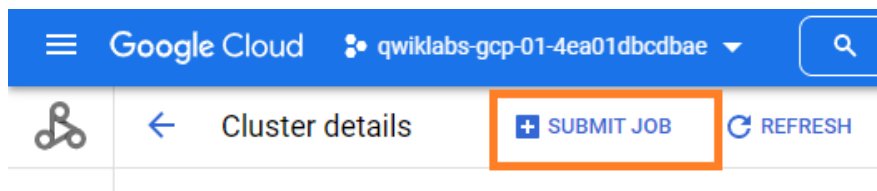


Clique sobre o nome do Cluster que foi criado





Clique sobre o botão “Submit job” na parte de cima da tela.



Preencha os campos conforme indicado no seu laboratório.

Depois de preencher, clique no botão “Submit”

### Submit a job

Job ID \*  
job-9f9ed21b

Job type \*  
Spark

Main class or jar \*  
org.apache.spark.examples.SparkPageRank  
The fully qualified name of a class in a provided or standard jar file, for example, com.example.wordcount, or a provided jar file to use the main class of that jar file

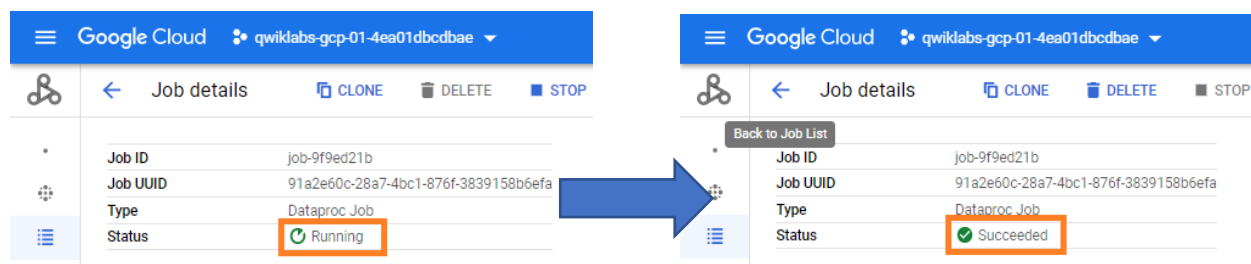
Jar files  
file:///usr/lib/spark/examples/jars/spark-examples.jar ✕  
Enter file path, for example, hdfs://example/example.jar  
Jar files are included in the CLASSPATH. Can be a GCS file with the gs:// prefix, an HDFS file on the cluster with the hdfs:// prefix, or a local file on the cluster with the file:// prefix.

Archive files  
Archive files are extracted in the Spark working directory. Can be a GCS file with the gs:// prefix, an HDFS file on the cluster with the hdfs:// prefix, or a local file on the cluster with the file:// prefix. Supported file types: .jar, .tar, .tar.gz, .tgz, .zip.

Arguments  
/data.txt ✕ Press <Return> to add more arguments  
Additional arguments to pass to the main class. Press Return after each argument.

Max restarts per hour  
1  
Leave blank if you don't want to allow automatic restarts on job failure. [Learn more](#)

Aguarde o Status mudar de “Running” para “Succeeded”



Valide a segunda tarefa.

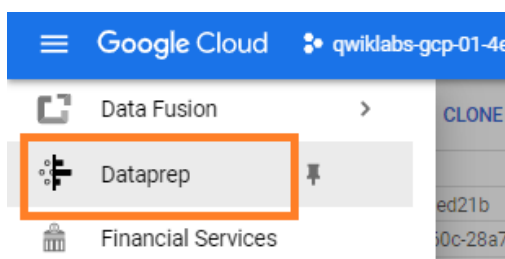


### Tarefa 3: execute um job simples com o Dataprep

Você já importou arquivos de dados com o Dataprep e os transformou para gerar visualizações. Agora use o Dataprep para importar um arquivo CSV (descrito abaixo) com os dados de execução do laboratório.

Estrutura do arquivo `gs://cloud-training/gsp323/runs.csv`:

No menu da esquerda procure por “Dataprep” (será necessário descer a lista)



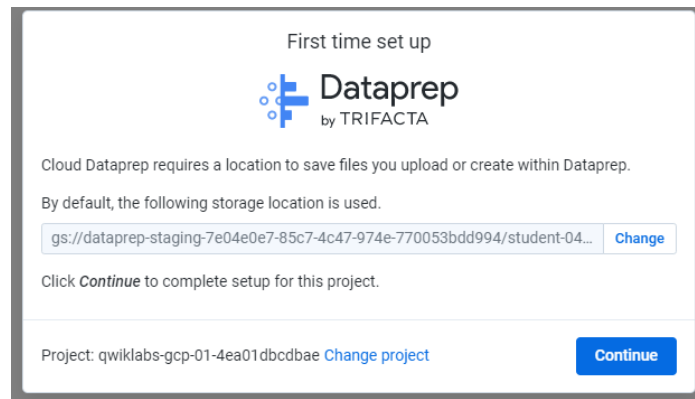
Será aberta uma nova janela.

Aceite os termos para continuar.

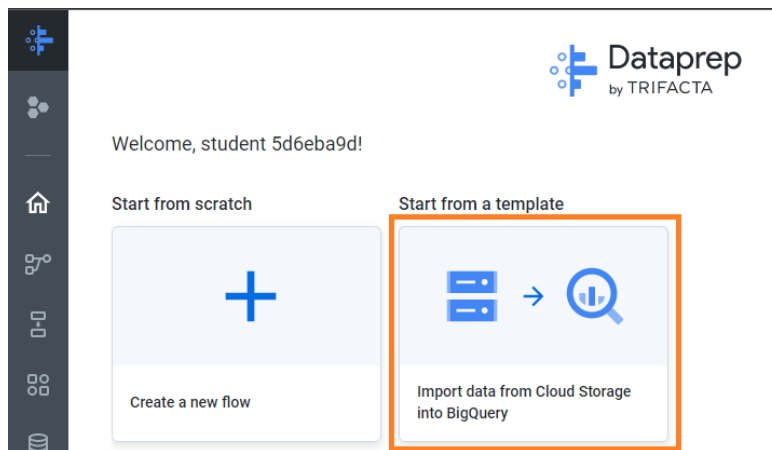
Dê as permissões necessárias que aparecerem nas mensagens seguintes.

Escolha a sua conta de email que foi indicada pelo laboratório e dê as permissões para ela também.

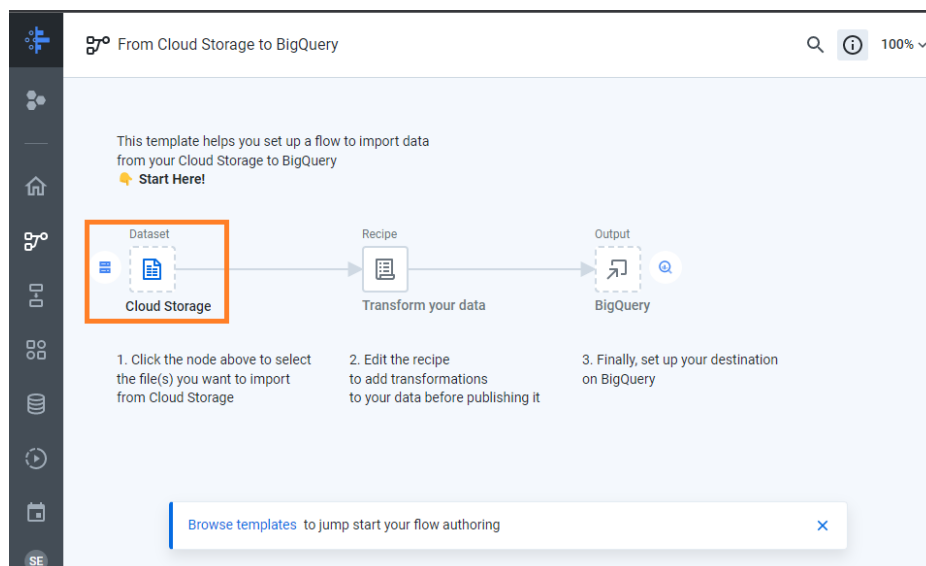
Chegando na tela abaixo, clique em “Continue”



Na próxima tela clique em “Import data from Cloud Storage into Bigquery”

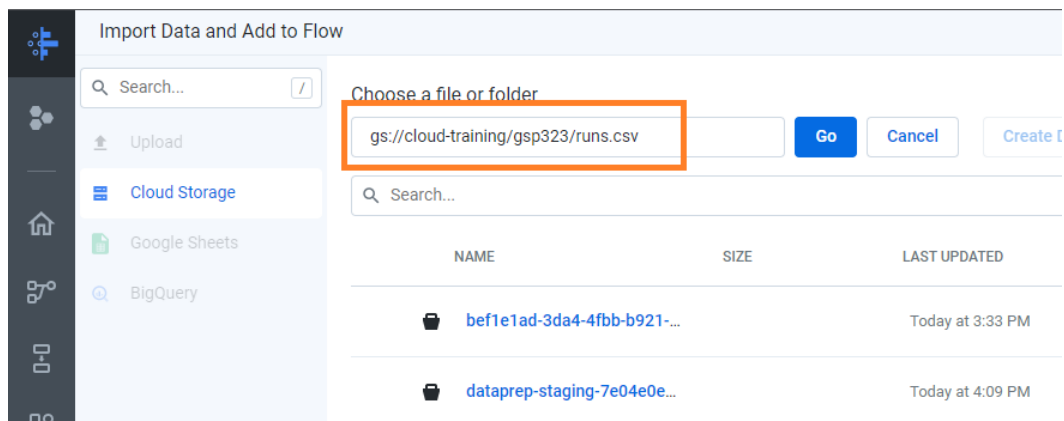


Clique no primeiro botão para selecionar o Bucket.

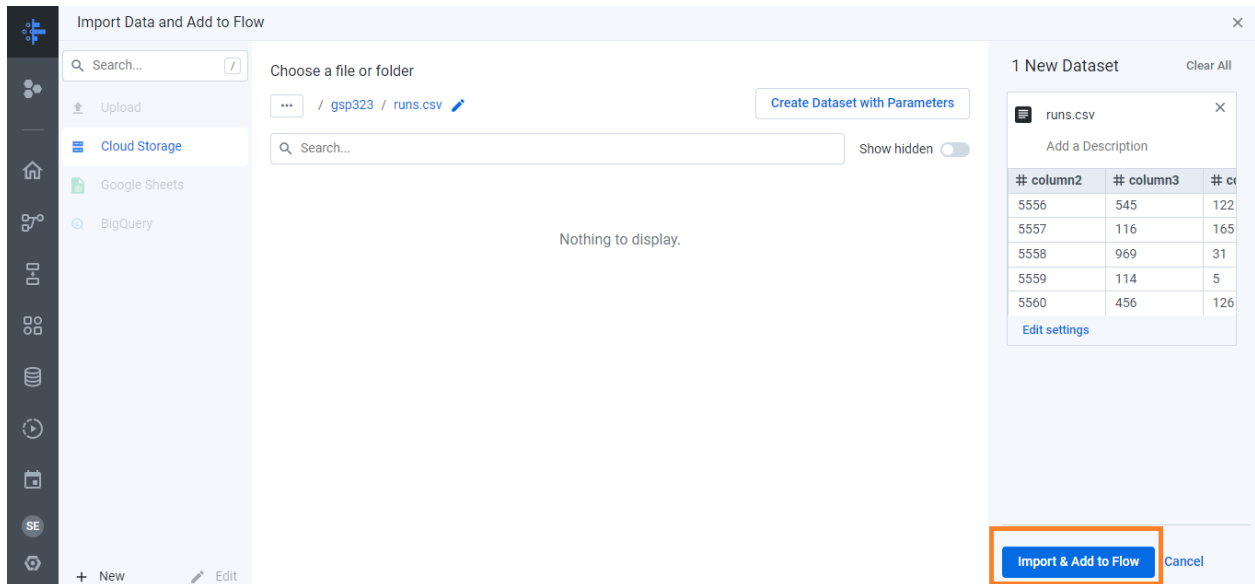


Informe o caminho do arquivo conforme especificado no seu laboratório.

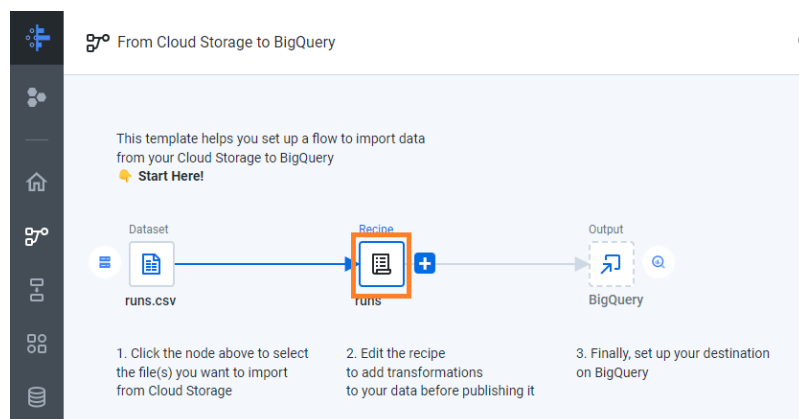
Depois clique sobre o botão “Go”.



Aguarde os dados carregarem e depois clique em “Import & Add to flow”

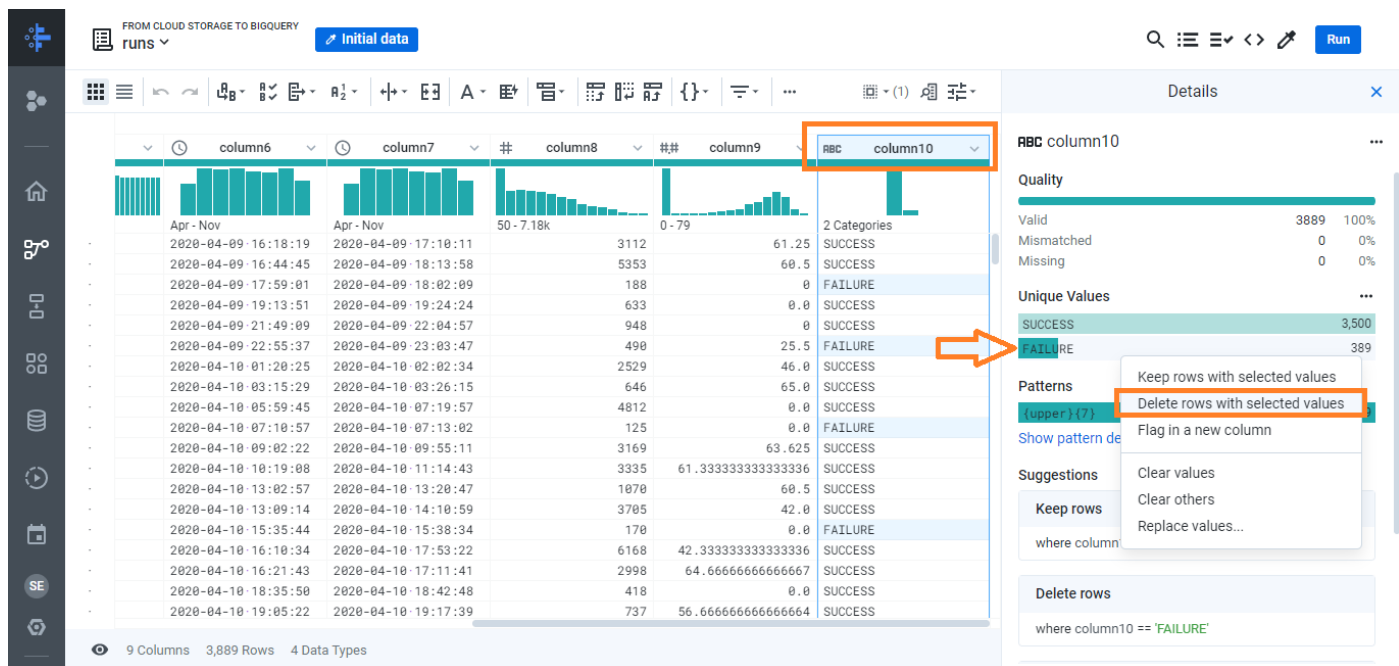


Clique duas vezes sobre o item “Recipe”



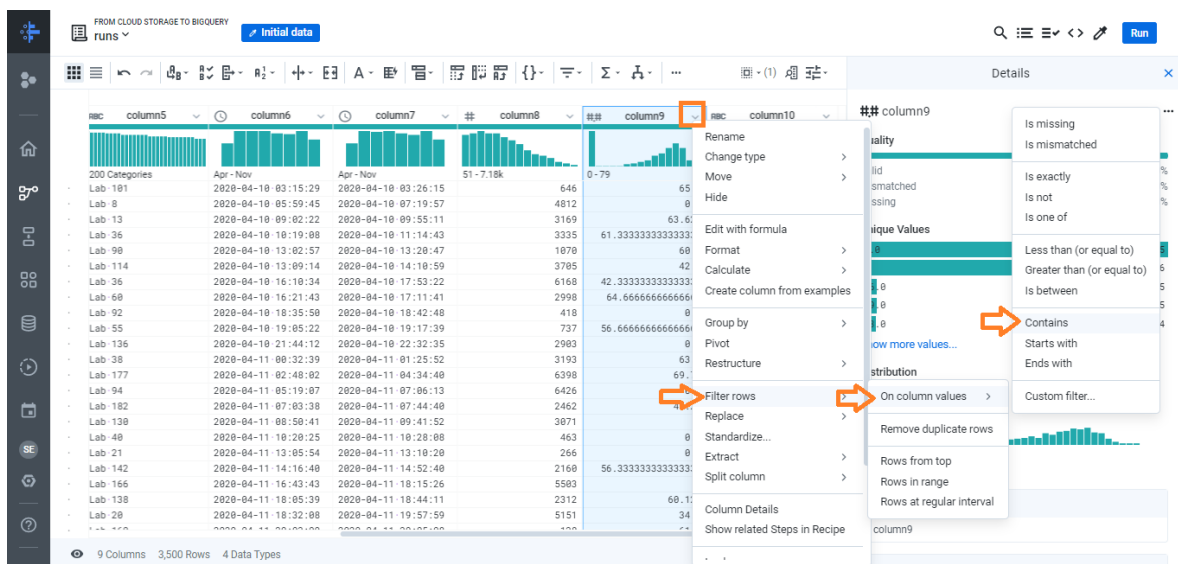
Ir até a coluna 10, clicar sobre ela.

No painel direito, clique com o botão direito do mouse sobre o item FAILURE e depois clique sobre a opção “Delete rows with selected values”.



Clique no ícone de filtro da coluna 9.

Vá até a opção “Filter rows”, “On column values”, “Contains”.



No painel que abrir à direita, informe o padrão conforme o seu laboratório e depois clique em “Add”

Filter rows

Condition required

Contains

Filter rows that contain a specified value or pattern

Column required

## column9

Pattern to match required

/(\*0\$|\*0\.\$0\$)/

Action required

Keep matching rows

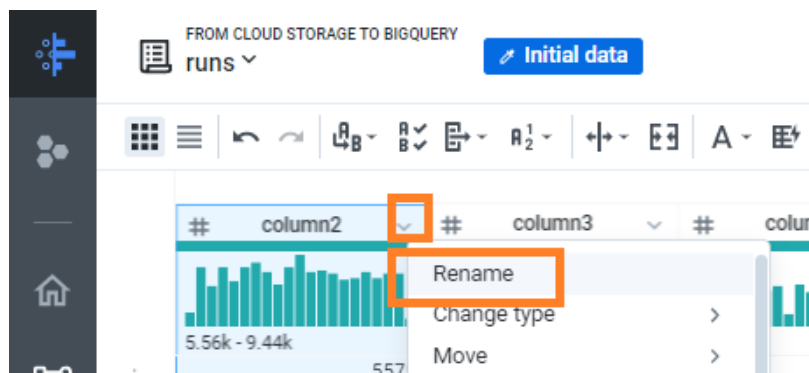
Delete matching rows

Cancel Add

Renomeie as colunas conforme indicado no seu laboratório.

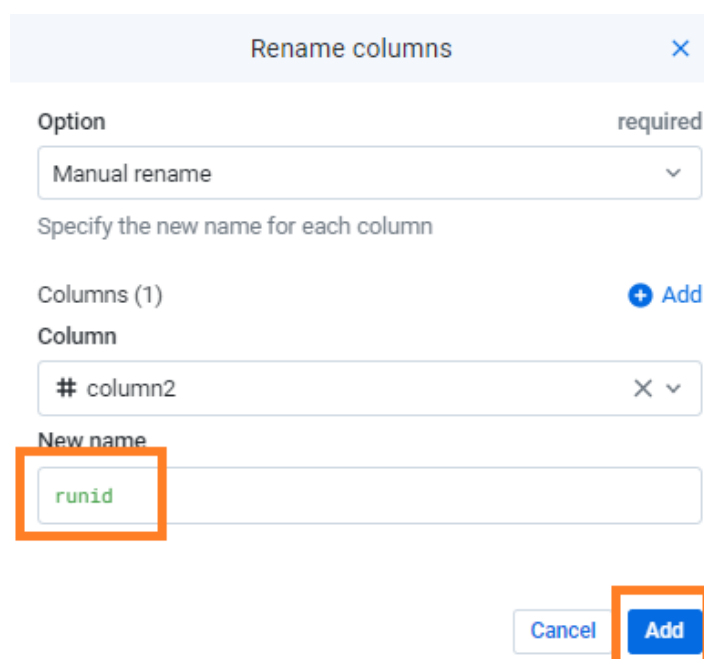
Clique sobre o ícone de filtro ao lado do nome da coluna.

Clique sobre a opção “Rename”.



No lado direito da tela informe o novo nome conforme o seu laboratório.

Clique em Add.



Repita o processo para as demais colunas conforme o seu laboratório.

Clique em “Run” para executar.

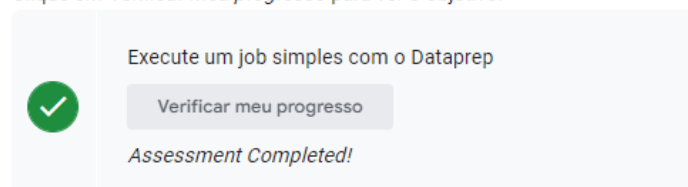


Clique em “Run” novamente, localizado no canto inferior direito da tela.

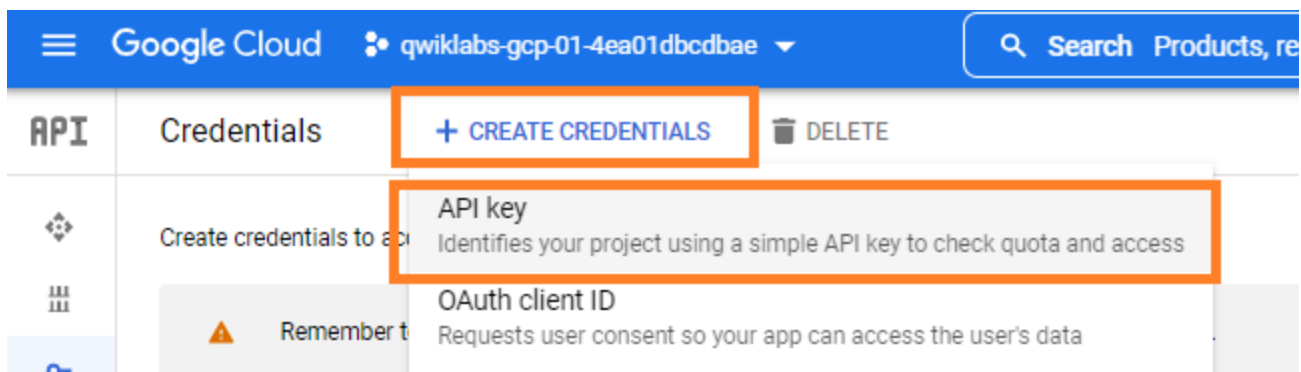
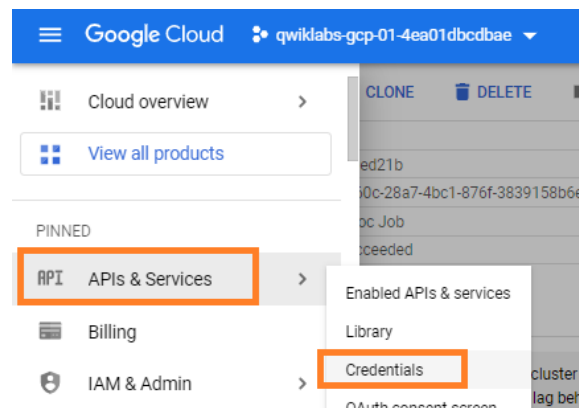
Aguarde o processo finalizar.

Valide a tarefa.

Clique em *Verificar meu progresso* para ver o objetivo.



Volte no menu da esquerda



Copie o valor da chave e armazene em algum lugar.

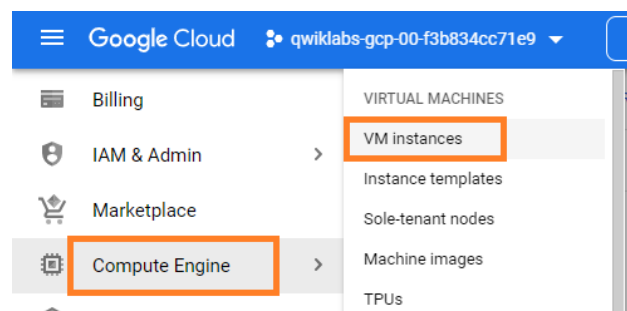
## API key created

Use this key in your application by passing it with the `key=API_KEY` parameter.

Your API key  
AIzaSyCwSJXezGs\_90fZUHGRtdyU9TAFIg0w0pQ

⚠ This key is unrestricted. To prevent unauthorized use, we recommend restricting where and for which APIs it can be used. [Edit API key](#) to add restrictions. [Learn more](#)

Volte no menu da esquerda em “Compute engine” e clique em “VM instances”



Procure pela VM que já veio criada junto com o laboratório (observação: não utilize a instância do cluster)

Clique para conectar nela via SSH.

Google Cloud qwiklabs-gcp-01-4ea01dbcdbae Search Products, resources, docs (/)

VM instances CREATE INSTANCE IMPORT VM REFRESH START / RESUME STOP OPERATIONS

VM instances are highly configurable virtual machines for running workloads on Google infrastructure. [Learn more](#)

**Filter** Enter property name or value

Status	Name	Zone	Recommendations	In use by	Internal IP	External IP	Connect
✓	<a href="#">cluster-6200-m</a>	us-west1-b			10.138.0.4 ( <a href="#">nic0</a> )	34.168.1.81 ( <a href="#">nic0</a> )	SSH
✓	<a href="#">cluster-6200-w-0</a>	us-west1-b			10.138.0.5 ( <a href="#">nic0</a> )	35.230.66.235 ( <a href="#">nic0</a> )	SSH
✓	<a href="#">cluster-6200-w-1</a>	us-west1-b			10.138.0.3 ( <a href="#">nic0</a> )	34.83.195.63 ( <a href="#">nic0</a> )	SSH
✓	<a href="#">lab-vm-qldm-31687988-93129d4bb63df25f</a>	us-east1-b			10.142.0.2 ( <a href="#">nic0</a> )	34.138.225.125 ( <a href="#">nic0</a> )	SSH

Na janela que aparecer, digite o seguinte comando, informando a sua chave de API gerada a pouco:

```
export API_KEY=APAGUE ESSA PARTE E COLOQUE SUA CHAVE
```

ssh.cloud.google.com/v2/ssh/projects/qwiklabs-gcp-01-4ea01dbcdbae/zones/us-east1-b/instances/lab-vm-qldm-31687988-93129d4bb...

SSH-in-browser

```
Linux lab-vm-qldm-31687988-93129d4bb63df25f 4.19.0-21-cloud-amd64 #1 SMP Debian 4.19.249-2 (2022-06-30) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Creating directory '/home/student-04-88790948e4eb'.
student-04-88790948e4eb@lab-vm-qldm-31687988-93129d4bb63df25f:~$ export API_KEY=AIzaSyCwSJXEzGs_90fZUHGRtdyU9TAFIgOwOpQ
student-04-88790948e4eb@lab-vm-qldm-31687988-93129d4bb63df25f:~$
```

Dê ENTER para executar.

Depois digite o seguinte comando e dê ENTER. (esse comando cria o arquivo request.json)

```
touch request.json
```

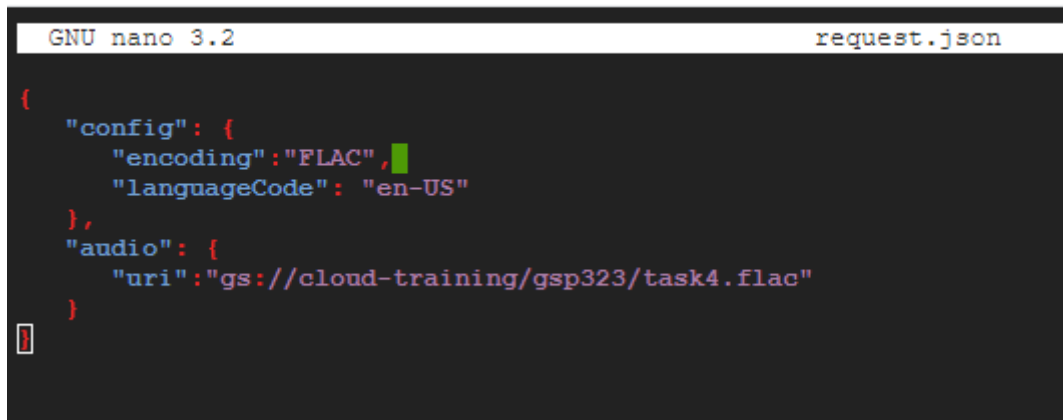
Depois digite o seguinte comando e dê ENTER. (esse comando serve para editar o arquivo)

```
sudo nano request.json
```

Dentro do editor que foi aberto, cole o seguinte trecho:

```
{
  "config": {
    "encoding": "FLAC",
    "languageCode": "en-US"
  },
  "audio": {
    "uri": "gs://cloud-training/gsp323/task4.flac"
  }
}
```





```
GNU nano 3.2 request.json
{
  "config": {
    "encoding": "FLAC",
    "languageCode": "en-US"
  },
  "audio": {
    "uri": "gs://cloud-training/gsp323/task4.flac"
  }
}
```

Para salvar e fechar o editor aperte: CTRL O e dê ENTER. Depois aperte CTRL X.

Na sequência digite o seguinte comando e dê ENTER.

```
curl -s -X POST -H "Content-Type: application/json" --data-binary @request.json
"https://speech.googleapis.com/v1/speech:recognize?key=${API_KEY}"
```

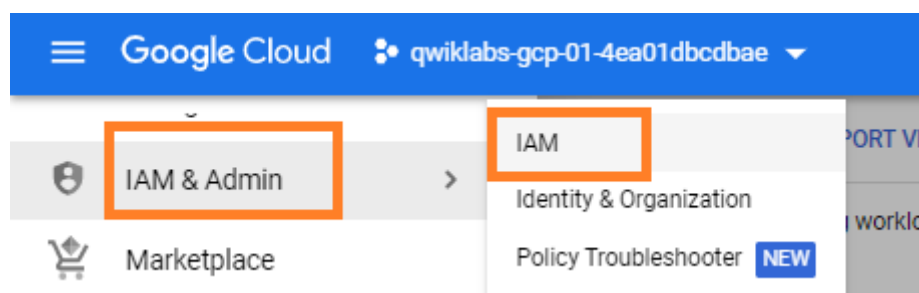
A seguinte mensagem deverá aparecer.

```
student-04-88790948e4eb@lab-vm-qldm-31687988-93129d4bb63df25f:~$ curl -s -X POST -H "Content-Type: application/
json" --data-binary @request.json "https://speech.googleapis.com/v1/speech:recognize?key=${API_KEY}"
{
  "results": [
    {
      "alternatives": [
        {
          "transcript": "welcome the Baseline data machine learning and artificial intelligence challenge lab",
          "confidence": 0.900355
        }
      ],
      "resultEndTime": "5.220s",
      "languageCode": "en-us"
    }
  ],
  "totalBilledTime": "15s",
  "requestId": "4088197278215526688"
}
student-04-88790948e4eb@lab-vm-qldm-31687988-93129d4bb63df25f:~$
```

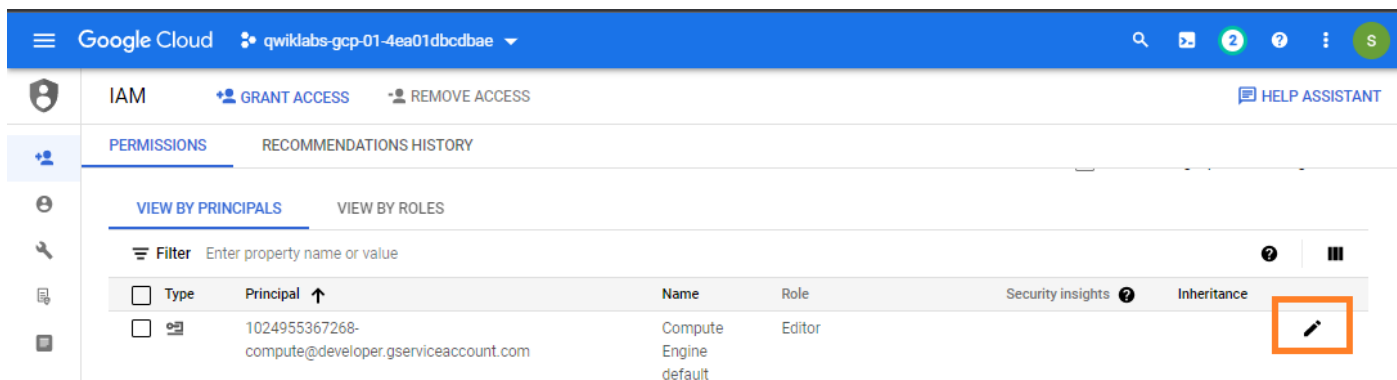
Agora, copie a mensagem de resposta para o arquivo "result.json" com o comando abaixo.

```
curl -s -X POST -H "Content-Type: application/json" --data-binary @request.json
"https://speech.googleapis.com/v1/speech:recognize?key=${API_KEY}" > result.json
```

Antes da próxima etapa, volte na página do Google Console e acesse o menu da esquerda em "IAM & Admin" e clique em "IAM".

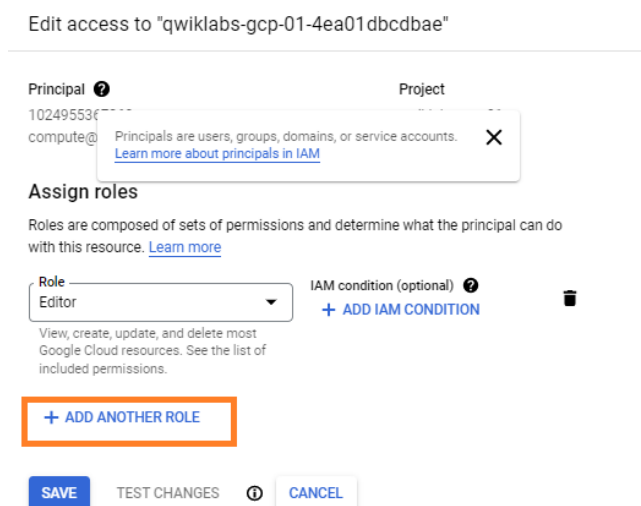


Procure pelo usuário que tem o "compute" antes do arroba e clique no lápis para editar.

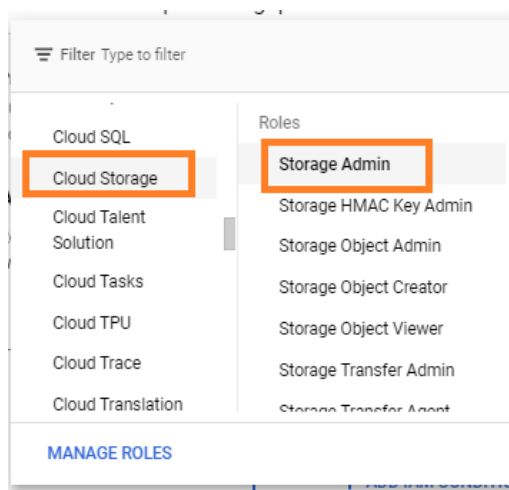


Esse usuário já tem uma regra, mas será necessário adicionar outra.

Clique em “Add Another role”



Escolha “Cloud Storage” e depois clique em “Storage admin” na lista que irá abrir na direita.



Clique em “Save” para salvar.

Por fim, utilize o comando abaixo para copiar o arquivo de resultado para o seu Bucket.

Altere o final do comando com o caminho especificado em seu laboratório.

`gsutil cp result.json COLOQUE O CAMINHO INFORMADO NO SEU LABORATÓRIO`

Valide a tarefa

Clique em *Verificar meu progresso* para ver o objetivo.



Conclua uma das tarefas de IA

[Verificar meu progresso](#)

*Assessment Completed!*