Carregando e exportando dados

**visão global**

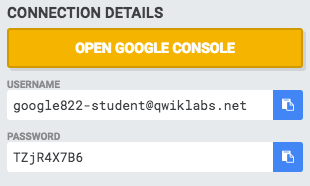
Neste laboratório, você carregará dados em diferentes formatos nas tabelas do BigQuery. Você carregará dados no BigQuery de várias maneiras, transformará os dados carregados e consultará os dados.

* Carregar um arquivo CSV em uma tabela do BigQuery usando a interface do usuário da web
* Carregar um arquivo JSON em uma tabela do BigQuery usando a CLI
* Exportar uma tabela usando a interface do usuário da web

Para cada laboratório, você obtém um novo projeto de GCP e um conjunto de recursos por um tempo fixo, sem nenhum custo.

1. Certifique-se de ter entrado no Qwiklabs usando uma **janela anônima** .
2. Observe o tempo de acesso do laboratório (por exemplo, img / time.pnge certifique-se de que você pode terminar nesse bloco de tempo.

Não há recurso de pausa. Você pode reiniciar, se necessário, mas você deve começar no começo.

1. Quando estiver pronto, clique em img / start_lab.png.
2. Anote suas credenciais de laboratório. Você os usará para fazer login no Cloud Platform Console. 
3. Clique em **Abrir o Google Console** .
4. Clique em **Usar outra conta** e copie / cole credenciais para **este** laboratório nos prompts.

Se você usar outras credenciais, receberá erros ou **incorrerá em cobranças** .

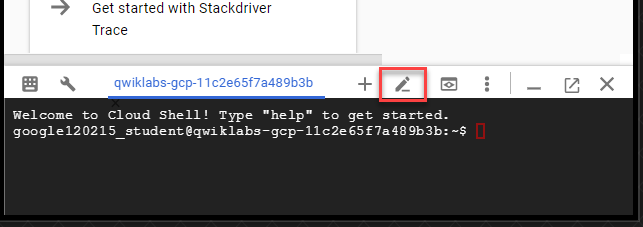
1. Aceite os termos e pule a página de recursos de recuperação.

Não clique em **Terminar, a** menos que tenha terminado o laboratório ou queira reiniciá-lo. Isso limpa o seu trabalho e remove o projeto.

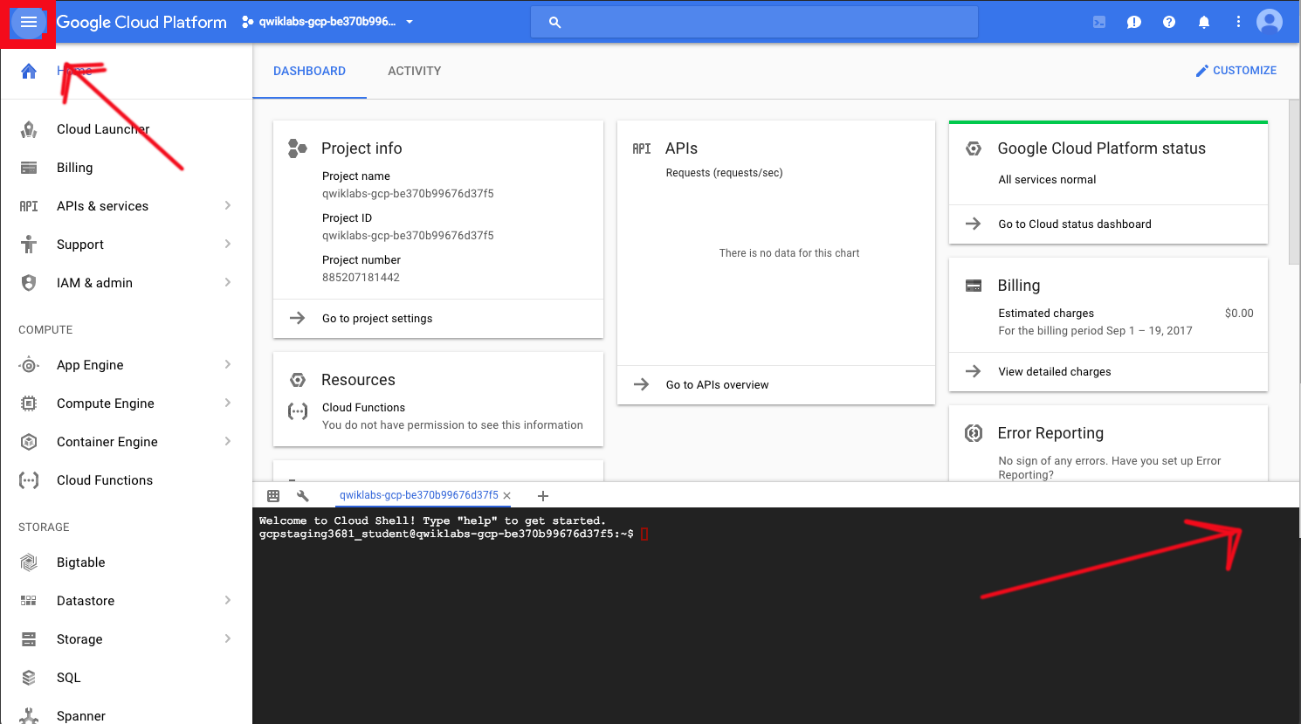
Inicie o Editor de códigos do Google Cloud Shell

Use o Editor de códigos do Google Cloud Shell para criar e editar facilmente diretórios e arquivos na instância do Cloud Shell.

Depois de ativar o Google Cloud Shell, clique no botão **Iniciar editor de códigos**(parece um lápis) para abrir o Cloud Shell Code Editor.

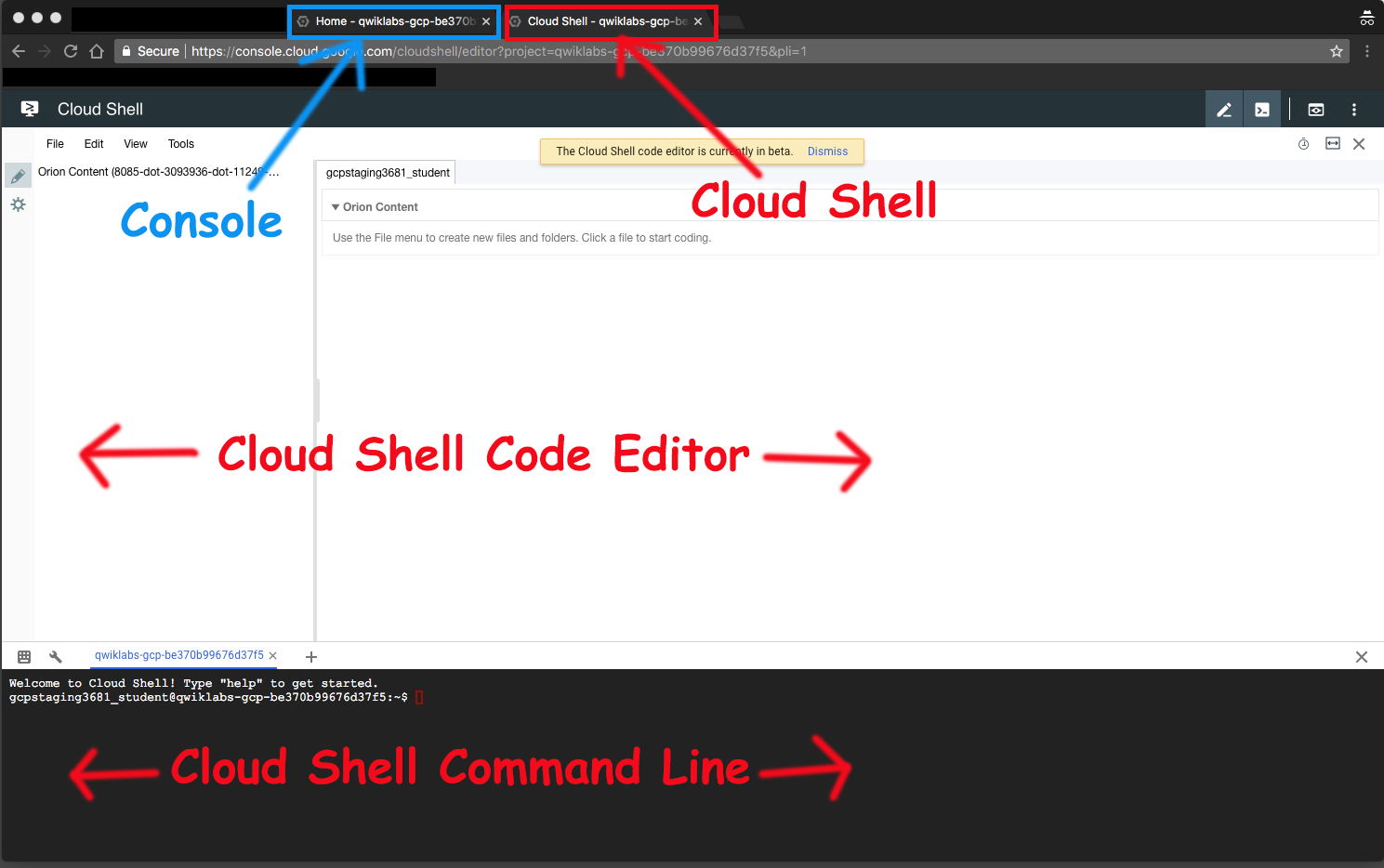


**Nota** : O botão **Launch code editor** pode estar fora da tela à direita. Pode ser necessário clicar no botão do **menu Navegação** para fechar o menu para ver os botões.



Agora você tem três interfaces disponíveis:

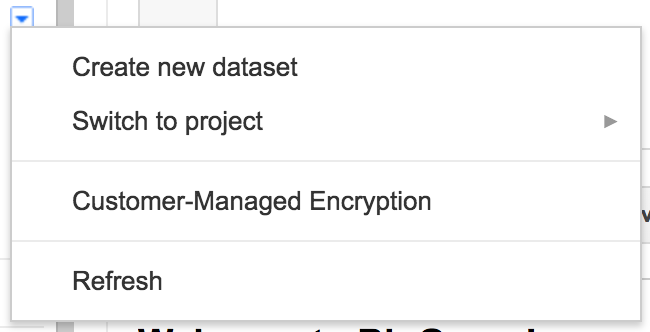
* O editor de código do Cloud Shell
* A linha de comando do Cloud Shell
* Console (clicando na guia). Você pode alternar entre o console e o Cloud Shell clicando na guia.



**Tarefa 1. Carregar os dados usando a interface do usuário da web**

Nesta tarefa, você faz o upload de um arquivo CSV para o BigQuery usando a interface da Web do BigQuery. O BigQuery suporta os seguintes formatos de dados ao carregar dados em tabelas: CSV, JSON, AVRO ou backups do Cloud Datastore. Este exemplo se concentra no carregamento de um arquivo CSV no BigQuery.

1. Volte para a guia do navegador que contém o Console.
2. No console, no **menu Navegação** ( 7a91d354499ac9f1.png), clique em **BigQuery** .
3. Quando o console abrir, você vai alterar a interface clássica. Clique em **Ir para a interface clássica** .
4. O console do BigQuery será aberto em uma nova guia do navegador. Agora, selecione seu projeto Qwiklabs clicando na seta para baixo ao lado de **Recursos Qwiklabs** , selecionando **Alternar para projeto> Seu projeto Qwiklab** .
5. À direita do nome do projeto, clique na **seta azul** . Escolha **Criar novo conjunto de dados** .



1. Na caixa de diálogo **Criar conjunto de** dados, para **ID do conjunto** de **dados** , digite cpb101\_flight\_datae clique em **OK** .
2. Baixe o seguinte arquivo para sua máquina local. Este arquivo contém os dados que preencherão a primeira tabela.

[Baixar aeroportos.csv](https://storage.googleapis.com/cloud-training/CPB200/BQ/lab4/airports.csv)

1. Crie uma nova tabela no conjunto de dados **cpb101\_flight\_data** para armazenar os dados do arquivo CSV. Passe o mouse sobre a linha com o nome do conjunto de dados. Isto irá revelar o ícone de criar tabela, que se parece com um sinal de mais. Clique no ícone de criação de tabela (o sinal de mais) à direita do conjunto de dados cpb101\_flight\_data.



1. Na página **Criar Tabela** , na seção **Dados de Origem** :

* Para **Local** , deixe **o upload do arquivo** selecionado.
* À direita do upload do arquivo, clique em **Escolher arquivo** , navegue até e selecione **aeroportos.csv** .
* Verifique se **o formato do arquivo** está definido como **CSV** .

**Nota:** Quando você criou uma tabela anteriormente, a opção Criar a partir da tarefa anterior permite que você use rapidamente suas configurações para criar tabelas semelhantes.

1. Na seção **Tabela de destino** :

* Para o **nome da tabela** , deixe cpb101\_flight\_dataselecionado.
* Para o **nome da tabela de destino** , digite **AEROPORTOS** .
* Para o **tipo Tabela** , a **tabela nativa** deve ser selecionada e imutável.

1. Na seção **Esquema** :
2. Adicione campos um de cada vez. O airports.csv possui os seguintes campos: IATA, AIRPORT, CITY, STATE, COUNTRYque são do tipo **STRING** e LATITUDE, LONGITUDEque são do tipo **FLOAT** . Torne todos esses campos NECESSÁRIOS.
3. Na seção **Opções** :

* Para **delimitador de campo** , verifique se **Vírgula** está selecionada.
* Como airports.csv contém uma única linha de cabeçalho, para que as **linhas de cabeçalho ignorem** , digite **1** .
* Aceite os valores padrão restantes e clique em **Criar Tabela** . O BigQuery cria um trabalho de carga para criar a tabela e carregar dados na tabela (isso pode levar alguns segundos). Você pode acompanhar o andamento do **trabalho**clicando em **Histórico de** tarefas .

1. **Quando** o trabalho de carregamento estiver concluído, clique em **cpb101\_flight\_data> AEROPORTOS** .
2. Na página **Detalhes** da **tabela** , clique em **Detalhes** para exibir as propriedades da tabela e, em seguida, clique em **Visualizar** para exibir os dados da tabela.

**Tarefa 2. Carregar os dados usando o CLI**

Nesta tarefa, você fará o upload de vários arquivos JSON e um arquivo de esquema associado para o BigQuery usando a CLI.

1. Volte para a guia do navegador que contém o Cloud Shell.
2. No Cloud Shell, digite o seguinte comando para fazer o download do arquivo de esquema da tabela em seu diretório de trabalho. (O arquivo é[schema\_flight\_performance.json](https://storage.googleapis.com/cloud-training/CPB200/BQ/lab4/schema_flight_performance.json) )

curl https://storage.googleapis.com/cloud-training/CPB200/BQ/lab4/schema\_flight\_performance.json -o schema\_flight\_performance.json

Em seguida, você criará uma tabela no conjunto de dados usando o arquivo de esquema que você baixou para o Cloud Shell e os dados dos arquivos JSON que estão no Cloud Storage. Os arquivos JSON têm URIs como os seguintes:

gs://cloud-training/CPB200/BQ/lab4/domestic\_2014\_flights\_\*.json

Observe que sua ID do projeto é armazenada como uma variável no Cloud Shell `$ DEVSHELL\_PROJECT\_ID`, portanto não há necessidade de você se lembrar dela. Se você precisar, você pode ver o seu ID do projeto na linha de comando à direita do seu nome de usuário (após o símbolo @).

1. No Cloud Shell, crie uma tabela chamada **flights\_2014** no conjunto de dados **cpb101\_flight\_data** com este comando:

bq load --source\_format=NEWLINE\_DELIMITED\_JSON $DEVSHELL\_PROJECT\_ID:cpb101\_flight\_data.flights\_2014 gs://cloud-training/CPB200/BQ/lab4/domestic\_2014\_flights\_\*.json ./schema\_flight\_performance.json

Existem vários arquivos JSON no intervalo do Cloud Storage. Eles são nomeados de acordo com a convenção: *domestic\_2014\_flights \* .json* \_. O caractere curinga (\*) no comando é usado para incluir todos os arquivos .json no intervalo.

Se você for solicitado a selecionar um projeto a ser definido como padrão, escolha a ID do projeto que foi configurada ao iniciar este qwiklab (Procure na guia **Conectar** de sua janela do Qwiklabs, o id do projeto normalmente se parece com **qwiklabs-gcp-123xyz** ) .

1. Depois que a tabela for criada, digite o seguinte comando para verificar se a tabela **flights\_2014** existe no conjunto de dados **cpb101\_flight\_data** .

bq ls $DEVSHELL\_PROJECT\_ID:cpb101\_flight\_data

A saída deve ficar assim:

tableId Type

-------------- -------

AIRPORTS TABLE

flights\_2014 TABLE

**Tarefa 3. Tabela de Exportação**

Nesta tarefa, você exporta uma tabela do BigQuery usando a interface do usuário da web.

Crie um bloco para receber os dados exportados

O BigQuery pode usar um intervalo do Cloud Storage para armazenar arquivos ao exportar dados. Você pode usar um bloco que você já criou em um laboratório anterior ou seguir estas instruções para criar um bloco.

1. No console, no **menu Navegação** ( 7a91d354499ac9f1.png), clique em **Página inicial** .
2. **Selecione e copie** o ID do projeto. Para simplificar, você usará o ID do projeto Qwiklabs, que já é globalmente exclusivo, como o nome do intervalo.
3. No console, no **menu Navegação** ( 7a91d354499ac9f1.png), clique em **Armazenamento** > **Navegador** .
4. Clique em **Criar Balde** .
5. Especifique o seguinte e deixe as configurações restantes como seus padrões:

|  |  |
| --- | --- |
| **Propriedade** | **Valor**  (digite o valor ou selecione a opção conforme especificado) |
| **Nome** | <your unique bucket name (Project ID)> |
| **Classe de armazenamento padrão** | [x] Regional |
| **Localização** | <Your location> |

1. Clique em **Criar** .
2. Registre o nome do seu balde. Você precisará disso nas tarefas subseqüentes.
3. No Cloud Shell, digite o seguinte para criar uma variável de ambiente denominada BUCKETe verifique se ela existe com o comando echo.

BUCKET="<your unique bucket name (Project ID)>"

echo $BUCKET

Você pode usar o **$ BUCKET** nos comandos do Cloud Shell. E se você precisar inserir o nome do intervalo <your-bucket>em um campo de texto no Console, poderá recuperar rapidamente o nome com **echo $ BUCKET** .

1. Retornar à interface da web do BigQuery. Se ainda não estiver aberto, abra o[Console](http://console.cloud.google.com/) . No **menu Navegação** ( 7a91d354499ac9f1.png), clique em **BigQuery.**
2. Selecione a tabela AIRPORTS que você criou recentemente e, usando o botão no canto superior direito, selecione a opção **Export Table** .
3. Na caixa de diálogo, especifique gs://<YOUR-BUCKET>/bq/airports.csve clique em **OK** .
4. Use o CLI para exportar a tabela:

bq extract cpb101\_flight\_data.AIRPORTS gs://$BUCKET/bq/airports2.csv

1. No console, no **menu Navegação** ( 7a91d354499ac9f1.png), clique em **Armazenamento** > **Navegador** . Navegue até o seu bloco e verifique se ambos os arquivos .csv foram criados.

**Agradecimento** : Este laboratório é baseado em um artigo de Felipe Hoffa: [As principais linguagens de programação de final de semana de acordo com o código do GitHub](https://medium.com/@hoffa/the-top-weekend-languages-according-to-githubs-code-6022ea2e33e8#.8oj2rp804)

**Termine seu laboratório**

Quando você tiver concluído seu laboratório, clique em **Finalizar Laboratório** . O Qwiklabs remove os recursos que você usou e limpa a conta para você.

Você terá a oportunidade de avaliar a experiência do laboratório. Selecione o número de estrelas aplicável, digite um comentário e clique em **Enviar** .

O número de estrelas indica o seguinte:

* 1 estrela = muito insatisfeito
* 2 estrelas = insatisfeito
* 3 estrelas = neutra
* 4 estrelas = Satisfeito
* 5 estrelas = muito satisfeito

Você pode fechar a caixa de diálogo se não quiser fornecer feedback.

Para comentários, sugestões ou correções, use a guia **Suporte** .

Última atualização: 2018-09-14

© 2018 Google LLC Todos os direitos reservados. Google e o logotipo do Google são marcas registradas do Google LLC. Todos os outros nomes de empresas e produtos podem ser marcas registradas das respectivas empresas com as quais estão associados.