Start Lab 01:30:00

Loading Data into Cloud SQL

Lab 1 hour No cost Intermediate

Rate Lab

GSP196

Google Cloud Self-Paced Labs

SQL

Lab 1 hour No cost Intermediate

Rate Lab

GSP196

Google Cloud Self-Paced Labs

SQL

Lab 1 hour No cost Intermediate

Rate Lab

GSP196

Google Cloud Self-Paced Labs

Google Cloud Self-Paced Labs

SQL

Lab 1 hour No cost Intermediate

Rate Lab

GSP196

SQL

Lab 1 hour No cost Intermediate

Rate Lab

GSP196

SQL

Lab 1 hour No cost Intermediate

Rate Lab

GSP196

SQL

Lab 1 hour No cost Intermediate

Rate Lab

GSP196

SQL

Lab 1 hour No cost Intermediate

Rate Lab

GSP196

SQL

Lab 1 hour No cost Intermediate

Rate Lab

GSP196

Google Cloud Self-Paced Labs

SQL

Lab 1 hour No cost Intermediate

Rate Lab

GSP196

Google Cloud Self-Paced Labs

```
export PROJECT_ID=$(gcloud info --
format='value(config.project)')
export BUCKET=${PROJECT_ID}-ml
```

4. Enter following command to stage the file into Cloud Storage bucket:

```
gsutil cp create_table.sql \
    gs://$BUCKET/create_table.sql
```

Task 2. Create a Cloud SQL instance

1. Enter the following commands to create a Cloud SQL instance:

```
gcloud sql instances create flights \
    --database-version=POSTGRES_13 --cpu=2 --memory=8GiB \
    --region="REGION" --root-password=Passw0rd
```

This takes a few minutes to complete.

Test completed task

Click **Check my progress** to verify your performed task. If you have successfully created a Cloud SQL instance, you will see an assessment score.

Create a Cloud SQL instance

Check my progress

2. Create an environment variable with the Cloud Shell IP address:

```
export ADDRESS=$(curl -s http://ipecho.net/plain)/32
```

3. Allowlist the Cloud Shell instance for management access to your SQL instance:

```
gcloud sql instances patch flights --authorized-networks
$ADDRESS
```

4. When prompted press, Y to accept the change.

Test completed task

Click **Check my progress** to verify your performed task. If you have successfully allowlisted Cloud Shell to access the SQL instance, you will see an assessment score.

Allowlist the Cloud Shell instance to access your SQL instance.

Check my progress

Create database and table

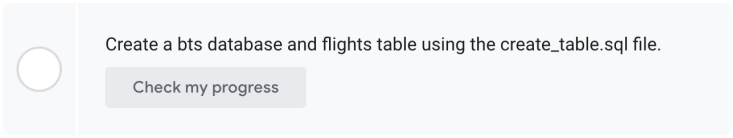
To import data into a Postgres table, you first create an empty database and a table with the correct schema.

1. In the Cloud Console, on the **Navigation menu** (☰), click **SQL**.
2. To open the Overview page of an instance, click the instance name `flights`.
3. Select **Databases** from the SQL navigation menu on the left.
4. Click **Create database**.
5. In the New database dialog, name the database `bts`.
6. Click **Create**.
7. To open the Overview page of an instance, select **Overview** from the SQL navigation menu.
8. Click **IMPORT** on the top.
9. In the Cloud Storage file field, click **Browse**.
10. In the Buckets section, click the arrow opposite your bucket name.
11. Select the file `create_table.sql`.
12. Click **Select**.
13. In the **File format** section, select SQL.
14. Specify the Database `bts` in your Cloud SQL instance.
15. Click **Import** start the import.

A few seconds later, the empty table will be created.

Test completed task

Click **Check my progress** to verify your performed task. If you have successfully created a bts database and flights table using the create_table.sql file, you will see an assessment score.



Task 3. Add data to a Cloud SQL instance

You created the empty database and table, now load the CSV files into this table. Loading the January data by browsing to 201501.csv in your bucket and specifying CSV as the format, bts as the database, and flights as the table.

1. In your Cloud SQL instance page, click **IMPORT**.
2. In the Cloud Storage file field, click **Browse**, and then click the arrow opposite your bucket name, and then click `201501.csv`.
3. Click **Select**.
4. Select **CSV** as File format.
5. Select the `bts` database and type in `flights` as your table.
6. Click **IMPORT**.

Task 4. Interact with the database

1. Connect to the Cloud SQL instance from Cloud Shell using:

```
gcloud sql connect flights --user=postgres
```

2. When prompted for a password enter `Passw0rd`. You may not see the letters as you type.
3. In the prompt that comes up, connect to the bts database:

```
\c bts;
```

4. When prompted for a password enter `Passw0rd`.
5. Then, run a query to obtain the 5 busiest airports:

```
SELECT "Origin", COUNT(*) AS num_flights
FROM flights GROUP BY "Origin"
ORDER BY num_flights DESC
LIMIT 5;
```

While this query is performant because the dataset is relatively small (only January!),

the database will slow as you add more months.

Relational databases are suited to smallish datasets on which you perform ad hoc queries that return a small a subset of the data. For larger datasets, you tune the performance of a relational database by indexing the columns of interest. Further, because relational databases typically support transactions and guarantee strong consistency, they are an excellent choice for data that will be updated often.

However, a relational database is a poor choice if:

- Your data is primarily read-only
- If your dataset sizes go into the terabyte range
- You have a need to scan the full table (such as to compute the maximum value of a column) or if your data streams in at high rates.

This describes the flight delay use case. For this case you would switch from a relational database to an analytics data warehouse – BigQuery. The analytics data warehouse will allow us to use SQL and is much more capable of dealing with large datasets and ad hoc queries (i.e. doesn't need the columns to be indexed).

Congratulations!

Now you know how to create tables and import text data that has been stored on Cloud Storage into Cloud SQL.

Next steps / learn more

Here are some follow-up steps:

- Data Science on the Google Cloud Platform, 2nd Edition: O'Reilly Media, Inc.

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 Updated November 28, 2023

Lab Last Tested November 28, 2023