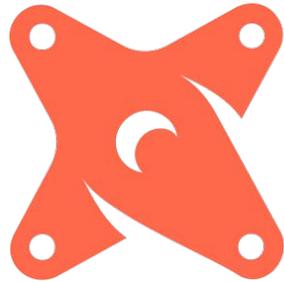


# TP DBT & BigQuery



**dbt**



Google  
BigQuery

Chafik NOUIRA, LesFurets.com  
tp-bigdata@lesfurets.com

# 1 - Create a GCP account using BigQuery Sandbox Web UI <https://console.cloud.google.com/bigquery>

No credit card or phone number are needed



## Créer votre compte Google

Accéder à Google Cloud Platform

Prénom  Nom

Nom d'utilisateur

Vous pouvez utiliser des lettres, des chiffres et des points

Noms d'utilisateurs disponibles : [tpbigdata7](#)  
[tpbigdata906](#) [bigdatatp37](#)

[Utiliser mon adresse e-mail actuelle à la place](#)

Mot de passe  Confirmer

Utilisez au moins huit caractères avec des lettres, des chiffres et des symboles

Afficher le mot de passe

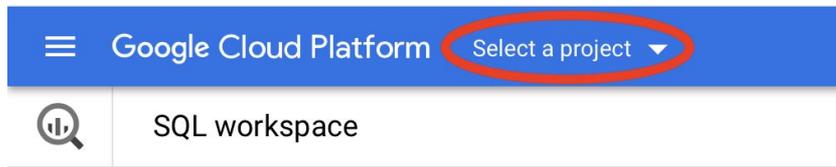
[Se connecter à un compte existant](#)



Tout Google avec un seul compte.

## 2 - Create a new project and test the query in *customers\_original.sql*

1



2



3

### New Project



You have 11 projects remaining in your quota. Request an increase or delete projects. [Learn more](#)

[MANAGE QUOTAS](#)

Project name \*

Project ID: tp-bigdata-337017. It cannot be changed later. [EDIT](#)

Location \*

Parent organization or folder

### 3 - Create a Service Account for dbt-user

BigQuery Credentials page : <https://console.cloud.google.com/apis/credentials/wizard>

1

The screenshot shows the Google Cloud Platform console interface. The top navigation bar includes the Google Cloud Platform logo and the user's profile 'tp-bigdata'. The left sidebar shows the 'APIs & Services' menu with 'Credentials' selected. The main content area is titled 'Create credentials' and contains the following sections:

- 1 Credential Type**
- Which API are you using?**

Different APIs use different auth platforms and some credentials can be restricted to only call certain APIs.

Select an API \*  
BigQuery API
- What data will you be accessing? \***

Different credentials are required to authorize access depending on the type of data that you request. [Learn more](#)

**i** This Google Cloud Platform API is usually accessed from a server using a service account. To create a service account, select "Application data".

User data  
Data belonging to a Google user, like their email address or age. User consent required. This will create an OAuth client.

Application data  
Data belonging to your own application, such as your app's Cloud Firestore backend. This will create a service account.
- Are you planning to use this API with Compute Engine, Kubernetes Engine, App Engine, or Cloud Functions?**

Applications running on GCE, GKE, GAE, and GCF can use Application Default Credentials and don't require that you create a credential.

Yes, I'm using one or more

No, I'm not using them
- NEXT** (highlighted with a red circle)
- 2 Your Credentials**

At the bottom of the page, there are 'DONE' and 'CANCEL' buttons.

2

The screenshot shows the 'Service account details' step of the wizard, labeled '1'. It contains the following fields and options:

- 1 Service account details**
- Service account name \***  
dbt-user  
Display name for this service account
- Service account ID...**  
dbt-user @tp-bigdata-337017.iam.gserviceaccount.com X ↻
- Service account description**  
Describe what this service account will do
- CREATE AND CONTINUE** (highlighted with a red oval)
- 2 Grant this service account access to project (optional)**
- 3 Grant users access to this service account (optional)**
- DONE** and **CANCEL** buttons at the bottom.

### 3 - Create a Service Account for dbt-user

BigQuery Credentials page : <https://console.cloud.google.com/apis/credentials/wizard>

3

#### ✓ Service account details

2

#### Grant this service account access to project (optional)

Grant this service account access to tp-bigdata so that it has permission to complete specific actions on the resources in your project. [Learn more](#)

Role

BigQuery Admin

Condition

[Add condition](#)

Administer all BigQuery resources and data

+ ADD ANOTHER ROLE

CONTINUE

3

#### Grant users access to this service account (optional)

DONE

CANCEL

## 4 - Generate and download BigQuery API Key (JSON file). We'll use it later to connect *dbt* to *BigQuery*

1 **API** APIs & Services

Credentials [+ CREATE CREDENTIALS](#) DELETE

Dashboard

Library

**Credentials**

OAuth consent screen

Domain verification

Page usage agreements

Create credentials to access your enabled APIs. [Learn more](#)

Remember to configure the OAuth consent screen with informa

### API Keys

<input type="checkbox"/>	Name	Creation date ↓
No API keys to display		

### OAuth 2.0 Client IDs

<input type="checkbox"/>	Name	Creation date ↓
No OAuth clients to display		

### Service Accounts

<input type="checkbox"/>	Email
<input type="checkbox"/>	dbt-user@tp-bigdata-337017.iam.gserviceaccount.com

## 4 - Generate and download BigQuery API Key (JSON file). We'll use it later to connect *dbt* to *BigQuery*

2

← dbt-user

DETAILS PERMISSIONS **KEYS** METRICS LOGS

### Service account details

Name  
dbt-user SAVE

Description SAVE

3

← dbt-user

DETAILS PERMISSIONS **KEYS** METRICS LOGS

### Keys

Service account keys could pose a security risk if compromised. We recommend more about the best way to authenticate service accounts on Google Cloud.

Add a new key pair or upload a public key certificate from an existing key pair.

Block service account key creation using [organization policies](#).  
[Learn more about setting organization policies for service accounts](#)

**ADD KEY** ▾

- Create new key**
- Upload existing key

	Key creation date	Key expiration date
--	-------------------	---------------------

4 - the json file is downloaded automatically

### Create private key for "dbt-user"

Downloads a file that contains the private key. Store the file securely because this key can't be recovered if lost.

**Key type**

- JSON  
Recommended
- P12  
For backward compatibility with code using the P12 format

CANCEL **CREATE**

5 - Create a dbt account : <https://cloud.getdbt.com/signup/>

6 - Configure the new project using the JSON file generated in step 4

1

dbt | lesfurets / Analytics

### Set Up "Analytics"

Continue >

This guide will help you set up a new dbt project.  
Got a question during setup? Click the speech-bubble in the top-right corner of the app to get in touch with support.  
Ready to get started? Click "Continue" to setup your dbt Cloud project.

- ✓ Create Project
- ↻ Database Connection
- 📁 Add Repository
- 👤 Invite Users

2

### Set Up a Database Connection

Which type of data warehouse should dbt Cloud connect to?

- PostgreSQL
- Redshift
- Snowflake
- BigQuery
- Apache Spark
- Databricks

skip >

## 6 - Configure the new project using the JSON file generated in step 4

3

### Set Up a Database Connection

Test Continue >

TYPE	bigquery
NAME	Bigquery

#### Bigquery Settings

CREATE FROM FILE **Upload a Service Account JSON file**

RETRIES (OPTIONAL) 1  
*Added in dbt 0.15.1. The number of times to retry queries that fail with BigQuery internal errors.*

LOCATION (OPTIONAL)  
*Added in dbt 0.12.2. Location to create new Datasets in. For more information, see [the Bigquery docs](#).*

4

Uploaded file successfully!

### Set Up a Database Connection

1 Test 2 Continue >

Connection Test Succeeded! Click "Continue" to save your connection and move onto the next step.

TYPE	bigquery
NAME	Bigquery

5

### Set Up a Repository

Continue >

Connect your repository below to finish setting up your dbt project. If you use GitHub, you can log in with GitHub to authorize access to your existing dbt project. If you use a different git provider, you can import your project using a Deploy Key. Check out the [detailed docs](#) for more information on authorizing your repository with a Deploy Key.

**Add repository from:**

Managed Git Clone Github GitLab

#### dbt Cloud managed repository

Initialize your project with a dbt Cloud-managed repository. You can use this repository to build and deploy your dbt project, then export it from dbt Cloud at any time.

tp-dbt Create

skip >

## 6 - Configure the new project using the JSON file generated in step 4

6

### Set Up a Repository

Connect your repository below to finish setting up your dbt project. If you use GitHub, you can log in with GitHub to authorize access to your existing dbt project. If you use a different git provider, you can import your project using a Deploy Key. Check out the [detailed docs](#) for more information on authorizing your repository with a Deploy Key.

**Add repository from:**

Managed  Git Clone  Github  GitLab

Successfully imported repository.

#### dbt Cloud managed repository

Initialize your project with a dbt Cloud-managed repository. You can use this repository to build and deploy your dbt project, then export it from dbt Cloud at any time.

tp-dbt

7

 Your account is on a free trial ending 16/01/2022. To continue using dbt Cloud after that date, please select a plan. [Go to Billing.](#)

## Welcome to lesfurets on dbt Cloud!

### Getting Started

Welcome to dbt Cloud! Now that you're set up with a repository and a connection, you can start building your dbt projects, be sure to check out [the docs](#) and [join the discussion on Discourse](#). If you have any questions or comments as a speech bubble in the top-right corner of the application.

7 - Initialize your project

**Note : in the next steps, you should always hit Save (CTRL + S) to save your files before executing dbt commands**

**Note 2 : for every change in the model, files must be saved and dbt run executed to rebuild models before testing or generating docs**

8 - In the **dbt\_project.yml** file, replace **my\_new\_project** by **tp\_dbt** (Lines 5 and 35)

8.1 - Go to line 38 : **materialized : view**, what does it mean ?

8.2 - Go to the file **models/example/my\_first\_dbt\_model.sql**, line 10, what does it mean ?

8.3 - Explain the relationship between the two models.

8.4 - Explain **models/example/schema.yml** file.

8.5 - At the bottom of the screen, write **dbt run** then click Enter. The project should compile with no errors.

8.6 - Execute the command **dbt test** to run tests. As you can see, there's a test failure for the first model. Fix it.

9 - Make your first commit.

10 - The master branch is now “read-only”. Go ahead and create a new branch named **add-customers-model**.

11 - Create a file named **customers\_original.sql** in **models** folder and copy the query from the provided file with the same name (in the .zip file).

12 - In **dbt\_project.yml** file, make sure that models in “**models**” folder are materialized as tables (do not apply this change to “**example**” folder)

13 - Execute **dbt run** and check the new table created in BigQuery (you’ll need to refresh the page)

14 - Duplicate **models/customers\_original.sql** in a file named **customers.sql** (same folder)

15 - Separate **customer.sql** model in 3 different models : **stg\_customers**, **stg\_orders** and **customers**. <

Make sure to add **stg\_customers** and **stg\_orders** references in the **customers** model.

(see **models/example/my\_second\_dbt\_model.sql**)

16 - Run “**dbt run**” and then check BigQuery tables.

17 - Add tests and documentation to **stg\_customers**, **stg\_orders** and **customers** models.

18 - Run “ **dbt test** ”

# 19 - Commit your changes and then merge to master

1

The screenshot shows the dbt web interface. At the top, there is a navigation bar with the dbt logo, the username 'lesfurets', and the project name 'Analytics'. Below this is a 'Project' header with a 'view docs' link and a menu icon. A green button labeled 'commit...' is circled in red. The left sidebar shows a file tree with folders like 'analysis', 'logs', 'macros', 'models', and 'seeds'. The 'models' folder is expanded, showing files like 'my\_first\_dbt\_model.sql', 'my\_second\_dbt\_model.sql', 'schema.yml', 'customers\_original.sql', 'customers.sql', 'schema.yml', 'stg\_customers.sql', and 'stg\_orders.sql'. The 'schema.yml' file is selected. The main area displays the content of 'schema.yml', which defines three models: 'customers', 'stg\_customers', and 'stg\_orders'. At the bottom, there are tabs for 'Query Results', 'Compiled SQL', and 'Lineage'. A keyboard shortcut icon (Command + Enter) is visible at the bottom left.

2

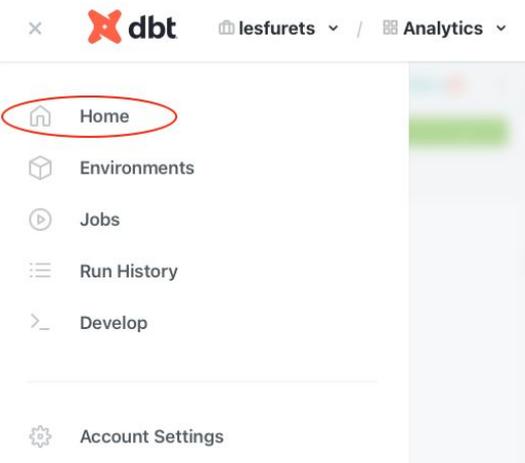
The screenshot shows a dialog box titled 'Write a Commit Message'. It has a text input field containing 'Add customers model' and a 'COMMIT MESSAGE' label. Below the input field are two buttons: a green 'Commit' button and a grey 'Cancel' button.

3

The screenshot shows the dbt web interface again. The navigation bar is the same. Below the 'Project' header, a green button labeled 'merge to master' is circled in red. The text 'branch: add-customers-model' is visible below the button.

# 20 - Deploy your project in production

1



2

## Deployment Environments >

**There's nothing here!**

Your account doesn't have any environments yet. You can [create a new one now](#).

3

The screenshot shows the 'Deployment Credentials' form in dbt Cloud. The form is titled 'Deployment Credentials' and includes a 'Save' button in the top right corner. The form fields are as follows:

NAME	Production 1
TYPE	Deployment
DBT VERSION	1.0 (latest)
CUSTOM BRANCH	<input type="checkbox"/>

**Deployment Credentials**  
Enter your deployment credentials here. dbt will use these credentials to connect to your database and run scheduled jobs in this environment.

CONNECTION	Bigquery >
PROJECT ID	tp-bigdata-337017
CLIENT EMAIL	dbt-user@tp-bigdata-337017.iam.gserviceaccount.com
CLIENT ID	109449542634079853730
DATASET	analytics 2

## 21 - Create a job to run you models in production

### < Environments / Production

Production

NAME

1.0.0

DBT VERSION

[Jobs >](#)

#### There's nothing here!

This environment doesn't have any jobs yet. You can [create a new one now.](#)

## 21 - Create a job to run you models in production

6 Save

NAME  1

ENVIRONMENT  2

DBT VERSION

THREADS

TARGET NAME

GENERATE DOCS?  3

RUN SOURCE FRESHNESS

RUN TIMEOUT   
Number of seconds a run will execute before it is cancelled by dbt Cloud. Set to 0 to never time out runs for this job.

DEFER TO PREVIOUS RUN STATE?   
Use the selected job's most recent successful run manifest for state comparison.

1.  4

2.

You can configure notifications for this job from your [profile page](#).

Schedule OFF

Webhooks

API

RUN ON SCHEDULE?  5

TIMING  every day

## 22 - Run your job and check BigQuery tables

< [Jobs](#) / **Production**

 Run now

 Settings

**No recent runs.**

LAST RESULT

 [View Documentation](#) >

ARTIFACTS

[Production](#) >

ENVIRONMENT