



dbt intro - optional class



dbt Introduction

What is dbt?

dbt is an open-source software that enables data analysts and engineers to transform data in their warehouse more effectively by simply writing SQL queries.

Key features:

- Version control:
- Modularization:
- Testing:
- Documentation.

Products:

- **dbt core:** a set of open source Python packages, run via the command line, that allows transform, document, and test your data
- **dbt Cloud:** a web based UI that allows you to both develop and deploy data pipelines (built on top of dbt Core)

dbt Introduction

Why is it so widely used?

- Rise of analytics engineering;
- Simplification of complex data transformations;
- Distance reduction from engineering roles and analytics roles;
- Usage of python or SQL;

dbt Introduction

Data Analyst

- Efficient collaboration;
- Data accuracy;
- Reduced development time

Data Scientist

- Streamlined data exploration;
- Collaborative research;
- Model production;

Data Engineer

- Efficient ETL workflow;
- Scalability;
- Versioned data pipelines;
- Easy to feed data to other departments;

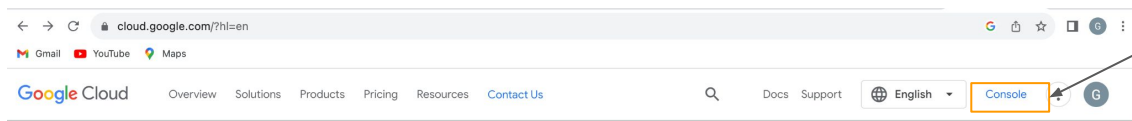
dbt Introduction

Setup BigQuery & dbt!

dbt cloud Introduction

1. Setup Google BigQuery

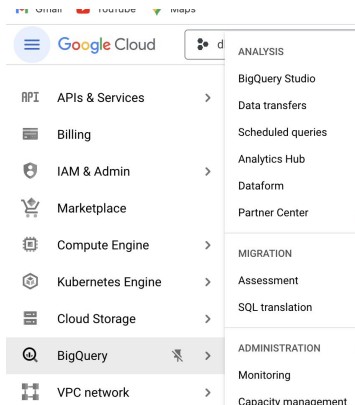
1. Setup a [google developer account](https://developers.google.com/) (<https://developers.google.com/>)
2. Access Google Cloud Console (<https://cloud.google.com/>)



The new way to cloud
starts here

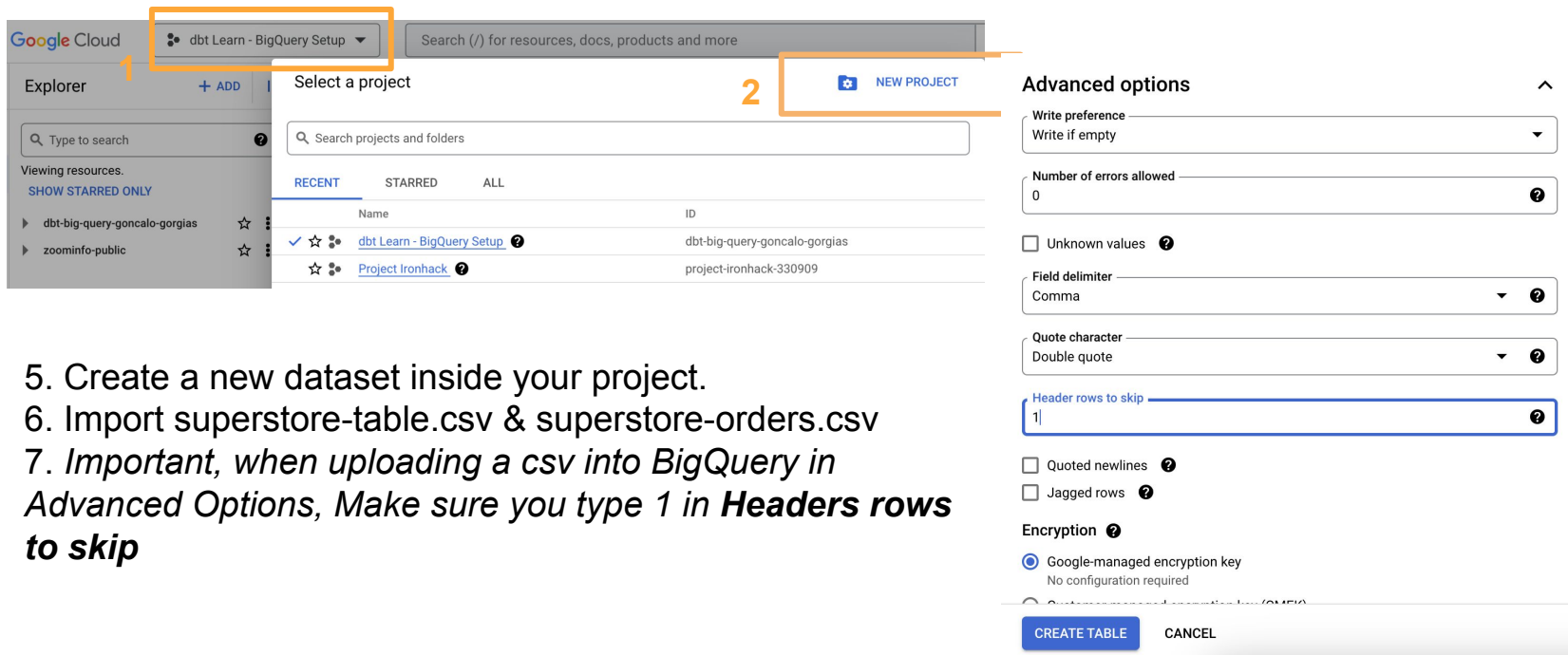


3. Go to Big Query



dbt cloud Introduction

4. Create a new project



Google Cloud

dbt Learn - BigQuery Setup

Search (/) for resources, docs, products and more

Explorer + ADD

Select a project

2 NEW PROJECT

Search projects and folders

RECENT STARRED ALL

Name	ID
✓ dbt Learn - BigQuery Setup	dbt-big-query-goncalo-gorgias
☆ Project Ironhack	project-ironhack-330909

Advanced options

Write preference
Write if empty

Number of errors allowed
0

☐ Unknown values

Field delimiter
Comma

Quote character
Double quote

Header rows to skip
1

☐ Quoted newlines

☐ Jagged rows

Encryption

☒ Google-managed encryption key
No configuration required

☐ Customer-managed encryption key (CMK)

CREATE TABLE CANCEL

5. Create a new dataset inside your project.

6. Import superstore-table.csv & superstore-orders.csv

7. *Important, when uploading a csv into BigQuery in Advanced Options, Make sure you type 1 in **Headers rows to skip***

dbt cloud Introduction

1. Setup dbt connection to BigQuery

1. Create dbt cloud account (<https://www.getdbt.com/product/dbt-cloud>)
2. Setup dbt <> Big Query connection
(<https://docs.getdbt.com/quickstarts/bigquery?step=1>)
3. **Generate BigQuery credentials**
 - a. Start the [GCP credentials wizard](#).
 - b. Create a service account for your new project from the [Service accounts page](#).
 - c. Create a service account key for your new project from the [Service accounts page](#).
4. Follow instructions on connection BigQuery in dbt
(<https://docs.getdbt.com/docs/cloud/connect-data-platform/connect-bigquery>)

If you want to learn more, dbt has multiple free courses for you
<https://courses.getdbt.com/collections>

Demo time!