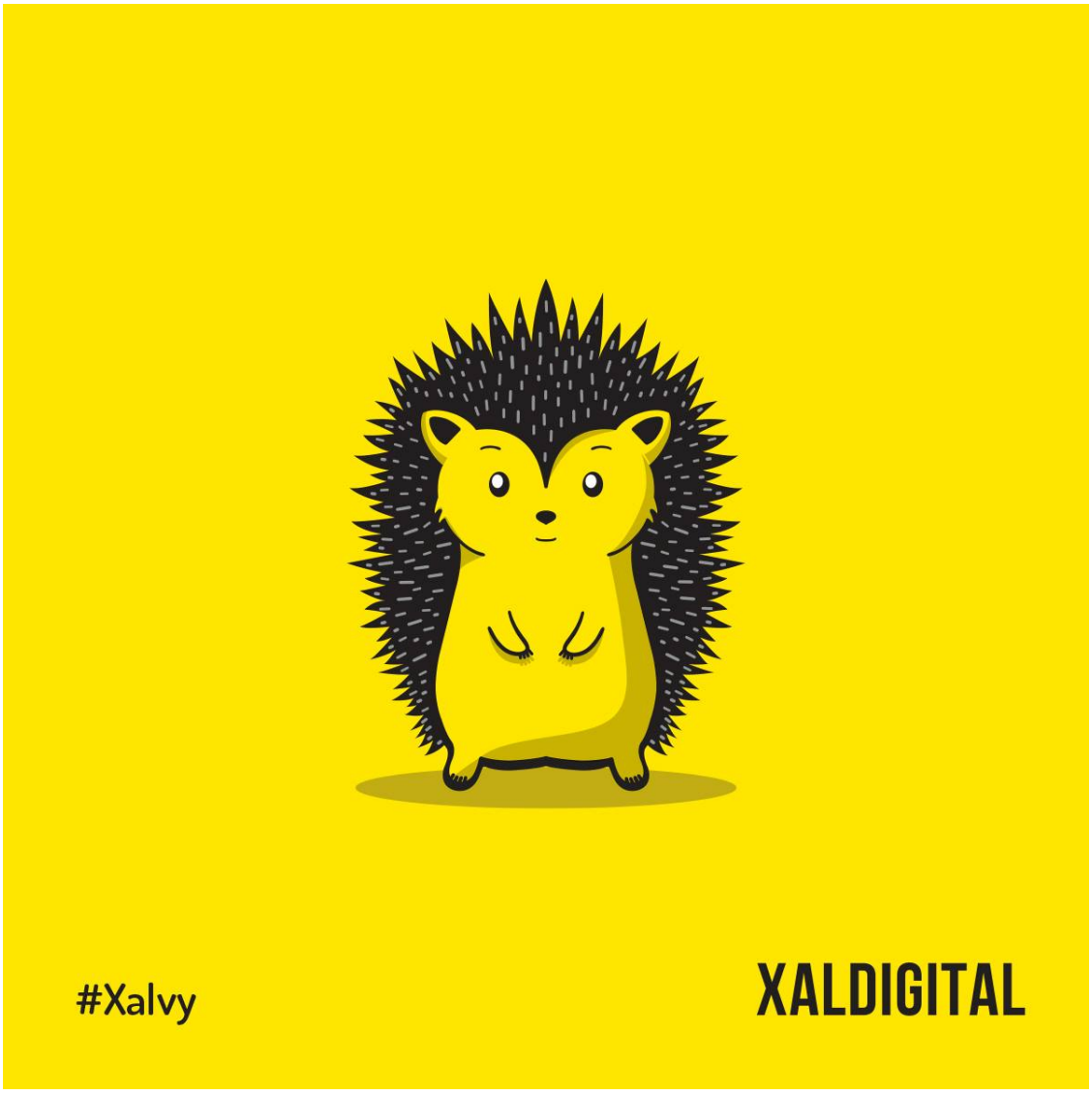

Challenge Data Engineer

Canseco García Edgar Jesús

.

Agosto de 2022

Solution

Environment setup

For environment setup, I create a VM on GCP (a budget topic), to deploy an environment from scratch.

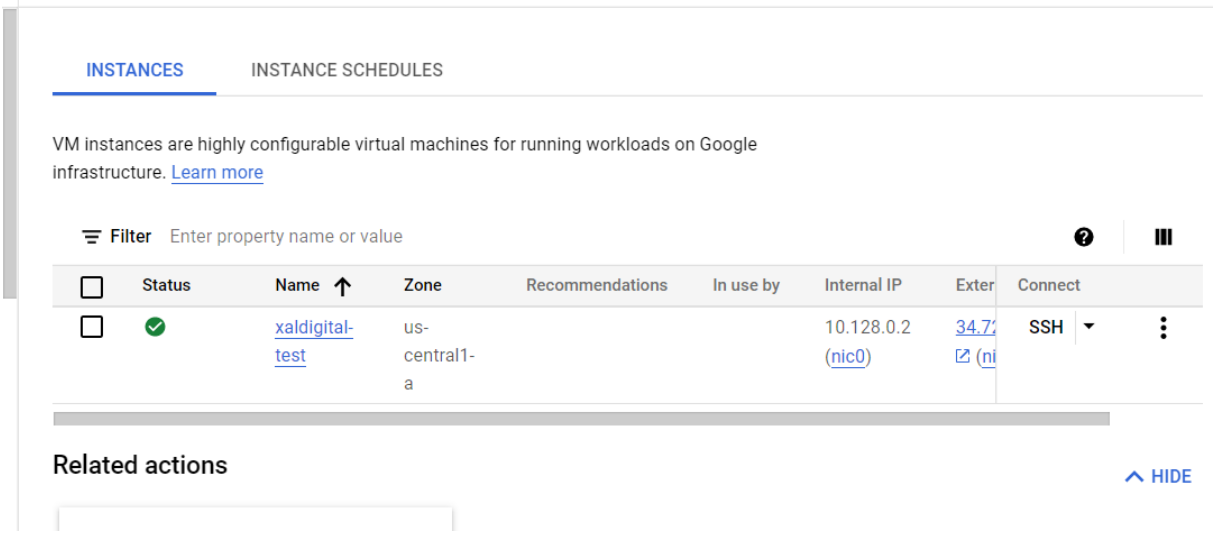

Figura 1- GCP VM

The image used is Ubuntu, I installed the necessary components to run a GUI environment, python, and docker for this challenge.

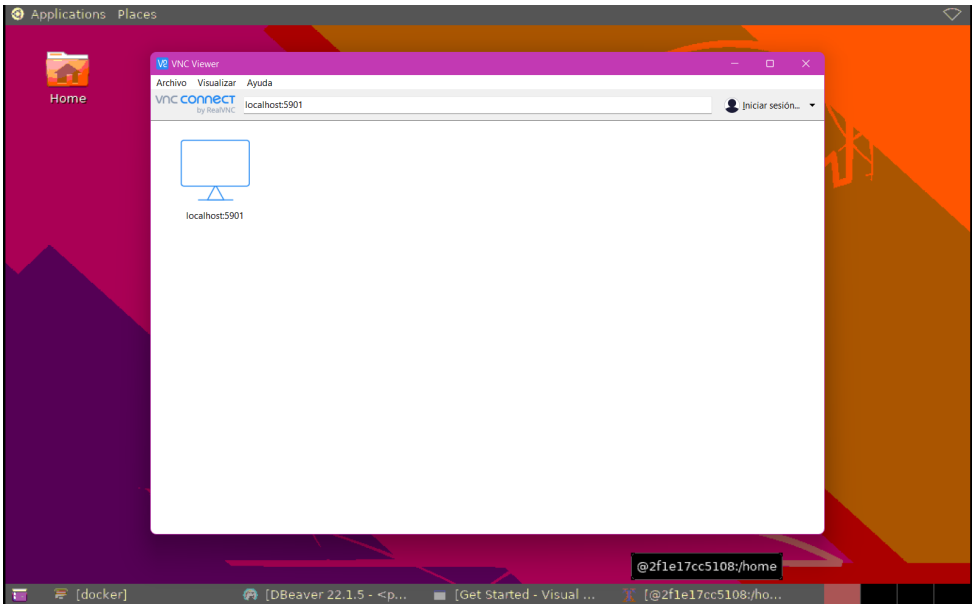

Figura 2- GUI Configuration

With the yml granted for the challenge, a Postgres database and a centos server are configured and launched with docker compose.

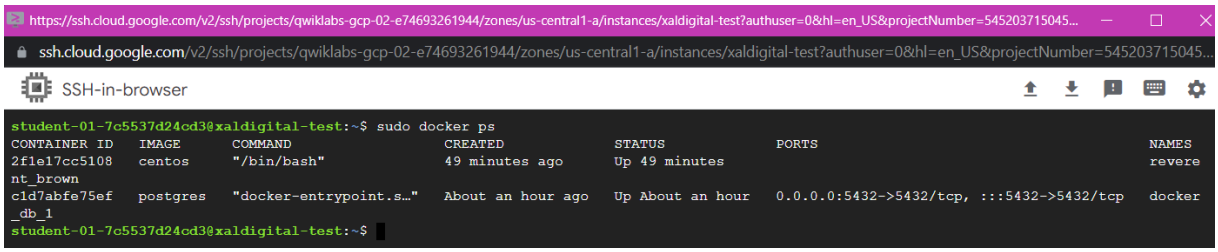

Figura 3- Containers

Design the E-R

Once the connection to PostgreSQL is made, I design the E-R from the database and create the table schema where data will be ingested.

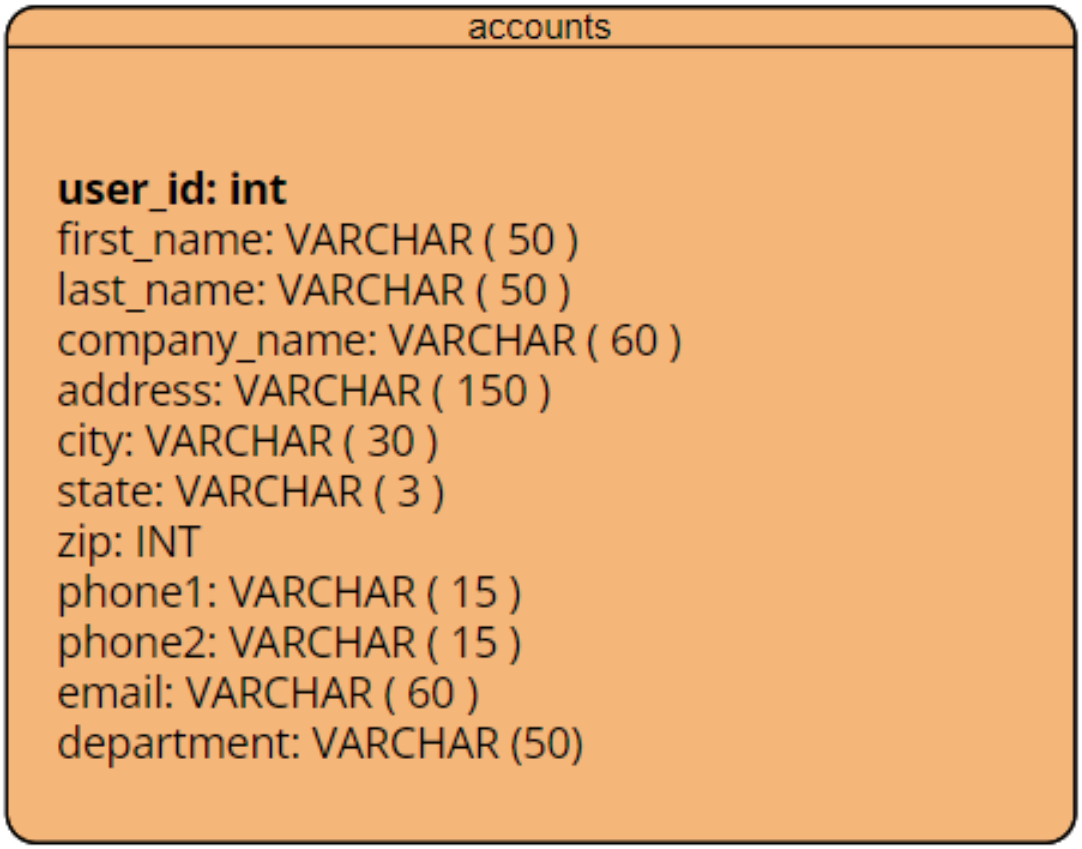

Figura 4- E-R

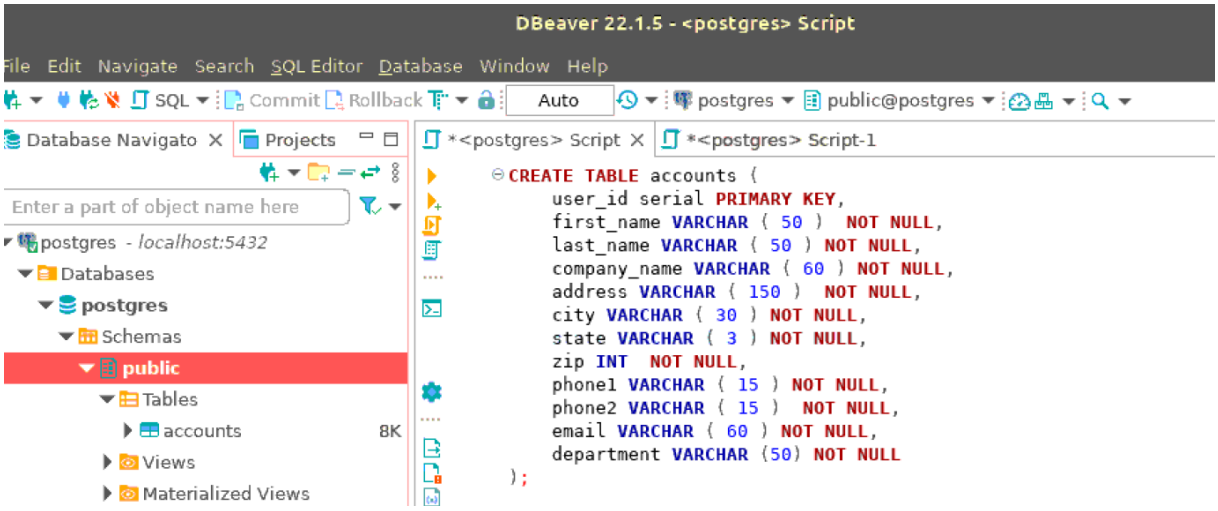

Figura 5- Table creation

Ingest the data

With the 'docker cp'command I copy the file from my ubuntu environment to centos server.

Figura 6- CSV file

I tried to make the connection from the centos server to PostgreSQL, I installed the necessary libraries but I got a bit confused with the network configurations and communication between containers. I still have the work of learning how the containers need to be configured with more time.

Figura 7- Failed connection

So to continue with the challenge, I will use my ubuntu server instead of centos. With this, I create the first version of the code that validates the state column and inserts the data into PostgreSQL

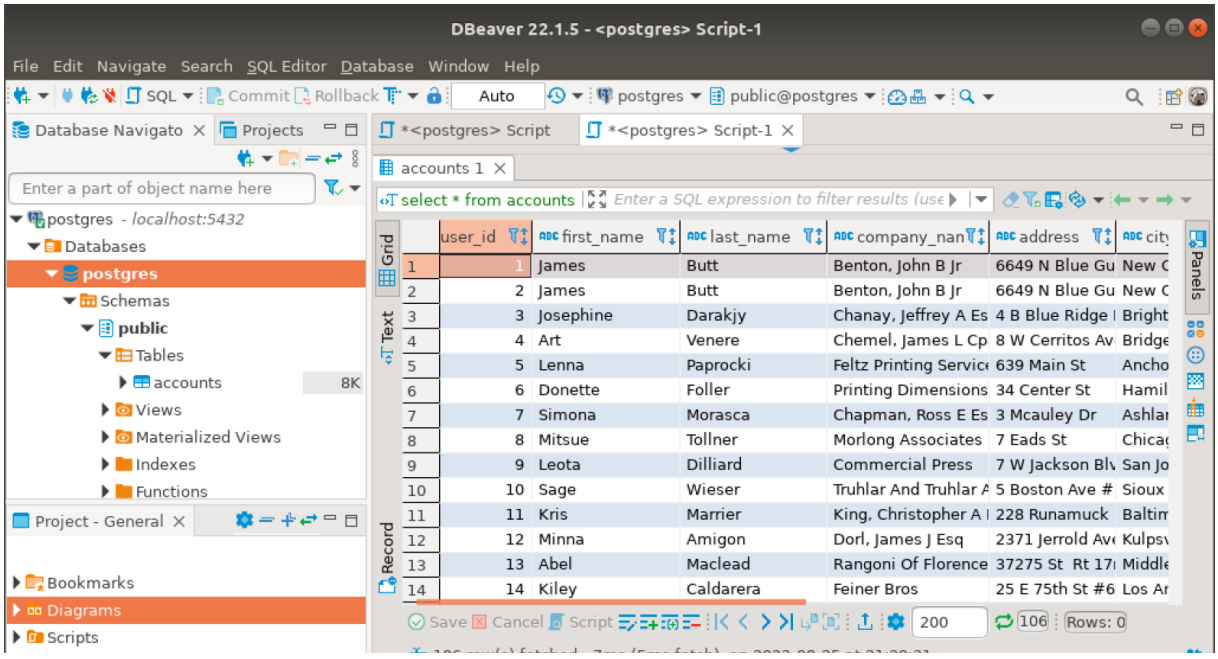

Figura 8- Table Result

Api for read

Finally, I implement the API with flask for reading the first 25 elements by the user_id of the table. The API address is: http://localhost:5000/api/accounts/

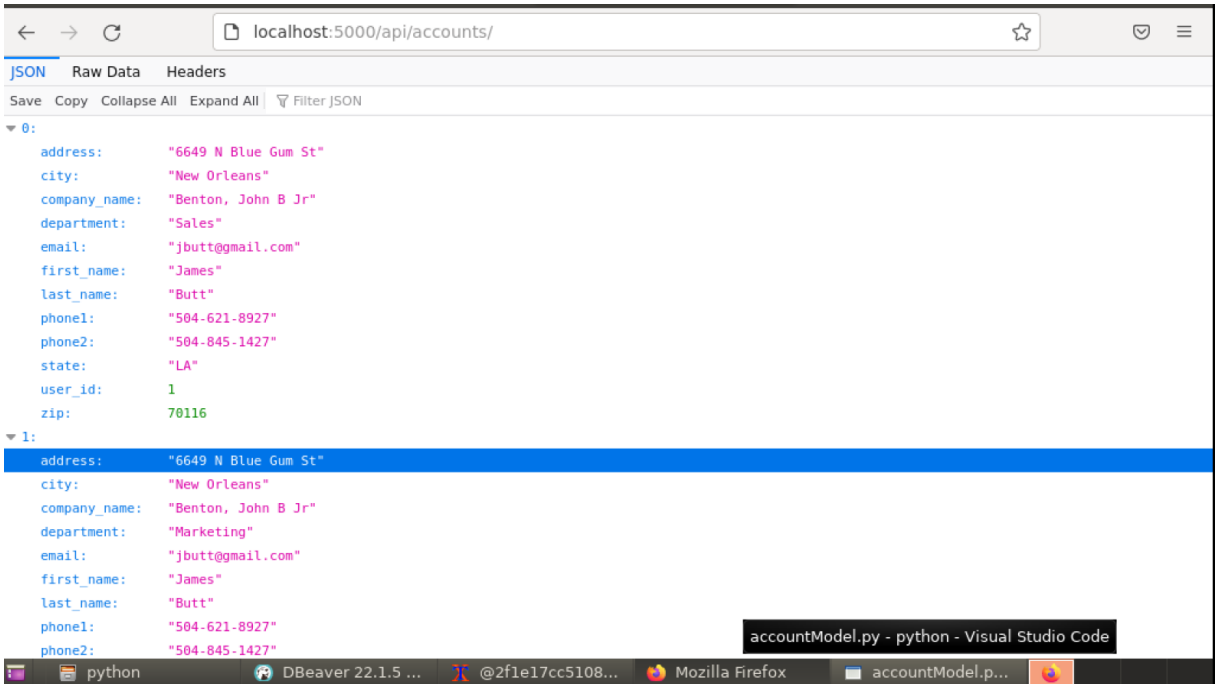

Figura 9- Table Result

Conclusion

It was fun to implement the challenge. I learned a lot implementing my first API, and also I have the work to search the Docker configurations better.