It has been a week and three days into the DataTalk Data Engineering ZoomCamp, but it has never run short of excitement.

In this first week, I finally got to understander containerization using docker.

Docker is a platform for developers and sysadmins to build, ship, and run distributed applications on laptops, data center VMs, or the cloud using containers. The main advantage of docker is isolation, portability, scalability, version control, resource efficiency, and cost saving.

As stated earlier, docker works using containers. Containers are like VM but the virtualization of the OS, not the complete computer hardware. They don’t require the dedication of some of your hardware, like in the case of VM.

I learned how to build docker containers using the command line, docker file, and docker-compose file. In addition to this, I learned how to build a docker network and also discovered that all containers built with docker-compose are created on the same network.

The fun part about this program was learning more about what I already knew. I learned how to generate an ssh key, apply the public key on the GCP platform, and connect to it locally using the private key. I learned how to connect to GCP using SDK and refreshed my knowledge of using terraform to build resources.

I was amazed when I discovered you could get data schema to create a table in SQL using pandas. In addition to this, I learned how to connect, create tables, and work on a SQL database from a jupyter notebook.

This bootcamp was more than learning but hands-on practice. I built containers for postgres, postgres-admin, and a python script for ingesting data into the Postgres database. For this week’s assignment, I loaded some data from the NYC Taxi & Limousine Commission using the ingestion container and wrote several queries to analyze the data. In addition, I used terraform to build a GCP cloud storage bucket and BigQuery dataset from my local computer.

While it has been fun so far, it has also challenged me, maybe because I’m using windows os. At our introductory meeting, Luis Oliveira mentioned he used 6 hrs weekly during the last zoomcamp, but I found myself using a minimum of 4 hrs daily. I spend the majority of my time debugging, which the community has so far been very helpful. As I got comfortable with the program, I found myself rendering help to others that are facing similar challenges I have conquered.  
  
There is more to learn and 9 weeks to go, and you can always join the ride if you’re interested in data engineering.