So, I am finally wrapping up week two of the DataTalk Data Engineering ZoomCamp, and the excitement keeps growing.

At this time last year, I attended my first Microsoft virtual training on using Azure synaptic for analytics, and the majority of the terms were mind-blowing. Initially, I felt dull, but I had to do some research after the training.

With week two of this ZoomCamp, I understood most of this content better, to the extent that I can easily explain them to someone if I am just waking up from a very deep sleep. These contents are;

* Data Lake and Data warehouse
* ETL and ELT
* Data orchestration

Which I made a short summary of in my attached slide. Ankush Khanna did just to the first two, while Jeff Hale and Kalise Richmond did just to the last using Prefect, which was what accounted for the majority of the week.

As I highlighted in week one, at the Zoomcamp, we don’t just learn but practice and solve problems as we learn. I learned and built several prefect workflows. Currently, I can build a prefect deployment to run on my local computer just using a python script. I can run this same deployment from a docker container, and I can also run it by storing my codes on GitHub. In addition to building and running deployment, I learned how to schedule flow runs and use prefect block to connect to GCP, GitHub and even save secrets.   
The fun part about this week was the assignment; I got stuck in question three for three days, thinking I had a poor network, not knowing the files we were to upload to GCP were large that we needed to edit the timeout variable of the prefect GCS bucket function. I felt like a dummy after discovering this.

While I wasn’t able to give much community support last week due to my health condition, I was still able to make some input to the FQA.

There is more to learn and 8 weeks to go, and you can always join the ride if you’re interested in data engineering.

You can still join!!!

Week two of the DataTalk Data Engineering ZoomCamp has come to an end, and my excitement for the course continues to grow. Last year, I attended my first Microsoft virtual training on using Azure Synapse for analytics, which left me feeling overwhelmed by the new terms I encountered. However, after doing some research, I was able to grasp the concepts better.

This week, I have a much better understanding of key topics such as Data Lake and Data Warehouse, ETL and ELT, and Data Orchestration. I have even created a summary of these topics in my slide deck. The instructors, Ankush Khanna, Jeff Hale, and Kalise Richmond, taught us how to apply these concepts using Prefect.

At the ZoomCamp, we don't just learn, but also practice what we've learned. I was able to build multiple Prefect workflows and learned how to run them locally, using a Docker container, and even from codes stored on GitHub. I also discovered how to schedule flow runs and, connect to GCP, GitHub, and securely store secrets using Prefect blocks.

This week's assignment was a bit challenging for me, as I spent three days trying to solve question three. I later found out that I needed to adjust the timeout variable of the Prefect GCS bucket function because the files we needed to upload to GCP were too large. Despite feeling like a dummy at first, I was proud of myself for finding a solution.

Although my health prevented me from offering significant support to the community last week, I was still able to contribute to the FAQ.

For this week, I have one troubling question: Are there opportunities for Prefect, like airflow, in the data engineering labor market?

There is still so much to learn and eight weeks left in the course. If you're interested in data engineering, there's still time to join the ride.