For the third week of the Data Talk Club dezoomcamp, we delved deeper into the world of data warehousing, specifically Google's BigQuery. As part of the Associate Cloud Engineering program, I used Pluralsight to gain exposure to BigQuery. I discovered that BigQuery combines both SQL and NoSQL features, which was new to me. Last week, I had limited knowledge of data warehouses, but thanks to Ankush Khanna's straightforward approach, even a kindergartner could grasp the concept.

This week, we learned how to use BigQuery through SQL, creating tables from data stored in Google Cloud Storage (GCS), and building machine learning (ML) models in BigQuery. I particularly appreciated the references to documentaries that helped us understand what we were writing, rather than just blindly coding.

We were also taught how to deploy an ML model using a docker image and how to test it using Postman. Ife came to my rescue when I struggled with using Postman.

In addition to data warehousing, we covered the architecture of BigQuery, best practices, and pricing. The pricing aspect was important to me after incurring a debt of $200+ last year when I left my Azure automl instance running for a month.

During the hands-on practice session, I deployed a flow to download and upload the NYC-TLC-Data for FHV 2019 to my Google Cloud Storage using Prefect. This helped me appreciate the value of orchestration tools even more. I also wrote some queries in BigQuery and created part of this content while the flow was running, killing two birds with one stone. From there, I created different types of tables in BigQuery - an external table, a non-partitioned table, and a partitioned-clustered table - and tested the pros and cons of each.

In conclusion, I learned how to deploy solutions that align with business needs, finding the optimal balance between query speed and cost. I am thrilled to be a part of the dezoomcamp and grateful to the organizers for providing this opportunity. I am eager for week four.