This week, the DataTalkClub de-zoomcamp focused on batch processing with Spark, with lectures delivered by Alex. I started learning Spark last year through online resources from Udacity and Datacamp, but struggled with installing it locally. Fortunately, this week's de-zoomcamp helped me set up Spark on both Windows and Linux machines. In addition to Spark, we also used DataProc to run our Spark jobs.

Batch processing is a popular technique for processing large volumes of data, with 63% of data professionals reporting its use according to a 2018 survey by the TDWI. This involves processing data in groups (batches) rather than processing individual data records in real-time.

During this week's lectures, we learned how to set up Spark, create local Spark clusters, access the web UI, and create standalone clusters for both master and workers. We also learned how to run Spark jobs by reading data from cloud storage, running Spark jobs on DataProc, and integrating Spark with data warehouses.

One of the major challenges I faced was setting up Spark on my Windows machine, but as I solved problems, I documented them in our FAQ and helped others in the community with similar issues.

Looking forward to week 6, I'm excited about the project phase and continuing to build my skills in Spark and batch processing.