Reflection report

In this reflection report, I will reflect on my contributions to the Final Project for the CloudDB course.

For this project, we got divided in teams of three. Me, together with my colleagues Luca and Tim, needed to implement a Flow-Join, following the paper from Wolf Roediger, *Flow-Join: Adaptive Skew Handling for Distributed Joins over High-Speed Networks* [Roediger W. et al., 2016].

During the development of the project, I was in charge of several tasks:
First, I created and managed the project repository, ensuring an organised structure and proper version control.
Within the Python development, I implemented the `selective_broadcast` function in the `adaptive_redistribution.py` file, following the specifications of Roediger's article [p.3 A, Roediger W. et al., 2016]. This function is essential for redistributing data when it is identified as heavy hitters. I also worked on the `FJ.cpp` file, where I tried to implement the flow join. Although I have to admit that I didn't quite make sure that it worked perfectly. This was because within the team we decided to switch from Python to C++ and it was necessary to start with the C++ implementation right away.
At the C++ level, I implemented the `inner_join` function in the `helper_functions.cpp` file, guided by the Python code provided by Luca. I used the structures defined by Tim in `helper_functions.h` to ensure proper integration of the code. I also added blocks to be able to measure the execution time, necessary to generate the plots.
Finally, I actively collaborated in debugging and troubleshooting together with the team, which contributed to a smoother and more efficient development of the project.

Given the estimated time frame, I knew it would not be easy. However, I approached the task with optimism and a positive attitude.
Throughout the development, there was a very good atmosphere in the team. We motivated and encourage each other. We collaborated effectively, checking in regularly, helping each other whenever necessary. The communication was spot on: we update each other regularly on WhatsApp and, if needed, met quickly on Discord. These aspects contributed significantly to the success of the project. At the start, we made sure to have a clear task distribution, which speed up the starting phase.
Of course, there were areas that could have been improved. It was difficult to finalize some aspects of our initial established goals. For example, the networking, which is one of the key points of the flow join implementation in the paper, couldn’t be finished. We started working with one library to implement this, and ad the end phase of the project we discovered it was not working out as expected. So we quickly switched to another library.

In future projects, I intend to prioritize early testing and decision-making to avoid last-minute rushes and ensure smoother project execution, as well as ensuring that all project goals be completed.
I will also take with me the communication and collaboration practices done here. Since they work out so well, I think all my feature project will benefit from it.

Sources

- Roediger W. et al. Flow-Join: Adaptive Skew Handling for Distributed Joins over High-Speed Networks. 2016