CS511 SP24 Final Project Update 1

Author: Yangchen Ye, Xixiang Liu, Zihan Shan

This week's vector

This week's vector is: could be be on a good path to reproduce the column sketch algorithm correctly. It is the core of the project around which everything else builds on that we at least have a working implementation of the algorithm and data structures in the paper that satisfies all the correctness assumptions and able to exhibit the performance benefit on a most typical workflow.

This week's plan

We came up with roadmap for achieving the goal in a modular way. Firstly, we will implement a minimal prototype of column sketch with extensive unit tests and integration tests. Secondly, we will run micro-benchmarks on the minimal prototype on the workload that we assume column sketch would perform best to get confidence that the core is implemented correctly. After that we will go on to create boundaries between the core algorithm and other parts of the project like parquet integration and extended benchmark suites.

This week's result

We reviewed the paper and crystalized correctness requirements into a suite of unit tests and integration tests and implemented a minimal prototype which passes the test suites([code could be found here](#)). This gives us confidence that the correctness property is solid.

Next week's vector

After correctness, the next week's vector is: does column sketch work well under micro-benchmark settings on dataset that most fit its strength. This will be a starting point and baseline to conduct extended benchmarks in end-to-end settings and different types of dataset.

Next week's plan

Next week we will do two things in parallel. Firstly, we will design and implement efficient query processing on top of column sketch and run experiments on the prototype to see if it works well. Secondly, we will start designing tools for generating flexible dataset in parquet format to prepare for extending the experiment.

One sentence per person

Xixiang Liu: I researched about the relevant data structures and algorithms to get ready for implementing them.

Zihan Shan: I reviewed the original paper for further understand the testbench to prepare running experiments on the prototype, and discussed plans for testing.

Yangchen Ye: I engineered the prototype of column sketch, settling down the design decisions unanswered in the original paper.