

Project Progress Report #2

Local Model Heavy Hitters

Yueran Cao, Kang Zhao

April 24, 2025

What have we done so far?

Over the past week, we had limited progress due to midterm exams and assignments for other courses. However, we did receive and carefully reviewed your feedback regarding data sources. Following your suggestion, we've decided to begin by generating synthetic data ourselves. This approach allows us to control parameters such as the distribution shape and presence of heavy hitters, which should help us better understand the algorithm's behavior.

We also briefly discussed the design of the data generator component, and how we might later test different distribution scenarios (e.g. uniform vs. skewed, different domain sizes).

Deviations from original plan

According to our initial timeline, we had planned to begin implementation this week. Due to other academic responsibilities, we had to temporarily shift focus. While this caused a short delay, the overall project structure and goals remain unchanged. We are still planning to simulate both the client and server logic in a single program, as previously discussed and approved.

Next steps / Progress plan

With midterms behind us, we are now resuming work on the project. Our immediate goals for the coming week are:

- Implement the synthetic data generator, including support for inserting known heavy hitters
- Implement the local randomization mechanism using randomized response (possibly with hashing)
- Begin developing the aggregation and decoding logic
- Run initial end-to-end tests with a small number of users to verify correctness

We are aiming to complete most of the core implementation before the next report, and will also begin thinking about how to structure our evaluation and experiments.