

# CS265 Design Document

Mali Akmanalp, Sophie Hilgard, Andrew Ross

March 22, 2017

## 1 Project Description

The goal of the project is to determine optimal allocations of a given amount of memory to cache, memtable, and bloom filters for different workloads in an LSM-tree. To explore this, we develop both mathematical models and simulations that demonstrate the cache/memtable/bloom filter size tradeoffs under a variety of workload distributions. These models and simulations can be used to predict the probability distribution of disk accesses and determine optimal memory allocations that minimize its expected number (or some other metric of the disk access distribution that better corresponds to robustness). Finally, we want to verify the correctness of our models and simulations by testing against RocksDB, and see if we can implement an adaptive optimization scheme that finds the best RocksDB configuration for a query distribution we may not even know beforehand.

Success in this project would be being able to find optimal memory allocations for specific query distributions and showing an improvement in results from an adaptive memory system as compared to standard RocksDB benchmarks.

## 2 Modeling

Initial modeling results and future directions are laid out in our midway check-in paper.

## 3 Implementation Plans

We first implemented a Python simulation to calculate expected disk accesses and component-by-component statistics for various memory configurations and workloads. We have delineated a variety of representative query distributions and hope to be able to predict what the optimal memory allocations are for each of these types using results from our modeling and simulations. We will then attempt to test our hypotheses in RocksDB and see if the results confirm our analysis.

## 4 Benchmarks

As in the MONKEY paper, we hope to be able to show that we can calculate memory allocations for which RocksDB will perform significantly faster than under its default settings and quantify these performance gains.

## 5 Expectations

We hope to be able to determine characteristics of query distributions that we can map to optimal memory allocations for those distributions.

## 6 Metrics of Success

Success in this project would be to generate RocksDB results which are significantly better than the results under default settings and to achieve results that make sense across modeling, simulation, and testing phases.