# CS422 - Project 1 - RLE and Execution Models

Liangze Jiang

March 2021

## 1 Introduction

In this project, we implemented several models and passes all 255 tests, and the entire execution time matches the baseline with a 0.9 speedup and 17.03s execution time. The models are:

- A volcano model with run length encoding data being it inputs and decodes the data in scan operator.
- A volcano model that scans single columns of run length encoding data and then merge them to only one column according to their virtual id and decode them to tuples before Join, Aggregate, Sort and so on.
- Based on the last model, we implemented some query optimization rules, i.e., we realize that some operators can be placed before others in order to make the query executed faster, namely transpose Filter and Reconstruct and transpose Decode and Aggregate, Join, Filter.
- Then we implemented operator at a time and column at a time engine, which executes on columnar inputs and use selection vectors to indicate a tuple is active or not.

## 2 Performance Analysis

After experiment, we can see that the execution time in these five model decreases one by one(volcano > volcano RLE > volcano RLE with rules > operator-at-a-time > column-at-a-time). Analysing the models one by one, we can find that the accesses of each case perfectly match the baseline and the execution time also match the baseline, in Fig 1. And Table 1 illustrates the speedup when we first pass all the test and the improvement we did after that.

Table 1: Improvements on Our First Commit

| | Speedup | Time |
|---|---|---|
| First Commit | 0.77 | 20.08s |
| First Commit + Code Cleaning | 0.85 | 18.16s |
| First Commit + Code Cleaning + Streaming the Inputs | 0.90 | 17.03s |

# 3 Implementation Details

Some techniques and implementation details in this project are:

- All the Join operators in this project are implemented by Hash Join.
- In the operator at a time and column at a time model, we eliminate inactive records before Join, Aggregate to further speed up the query.
- Use transpose-compute-transpose pattern to process column data in operator at a time and column at a time model.
- Avoid materializing the input table as much as possible in volcano model.

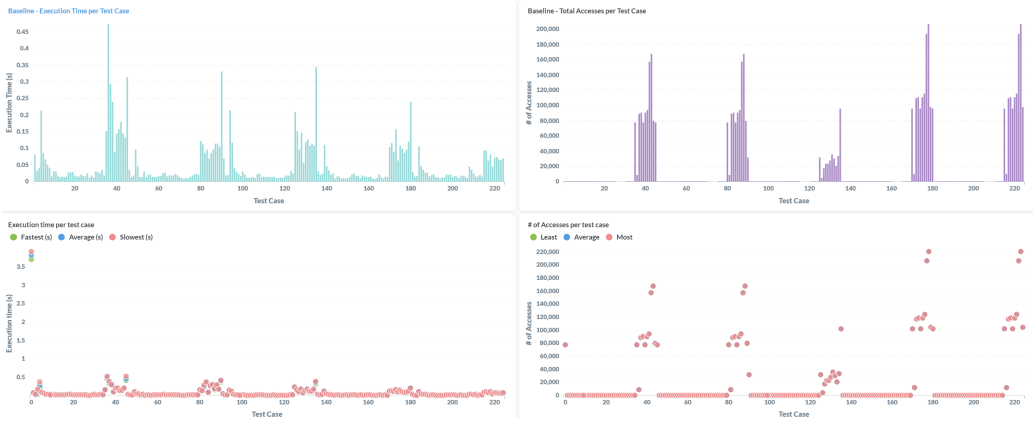


Figure 1: Comparison between baseline and our implementation