Epi4 Performance Analysis

Ruohao Wu

Table of Contents

Assumptions

1. The datafile should have exactly the same format as example code

2. The property of variable should make sense (eg. time should not be negative)

3. The query index should match among files

Introduction of Variables in Data

Real: Wall Clock Time, all the time spent by the program waiting for resources

User: The amount of time the CPU spends running the program's code

Sys: Time spent handling system calls, interrupts, and other low-level operations

Data Cleaning

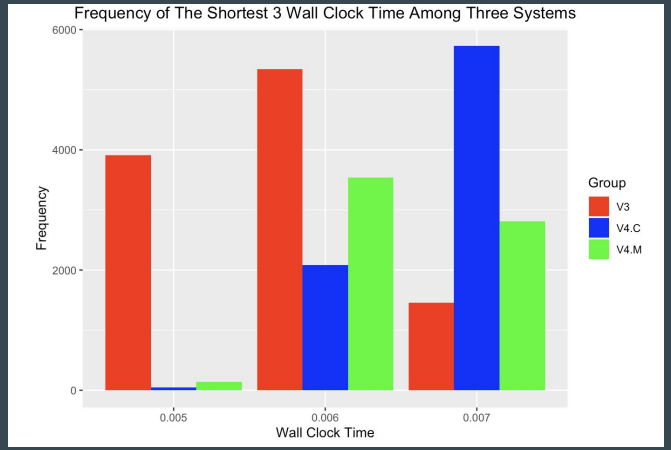
Data		
history	19365 obs. of 2 variables	
V3	19359 obs. of 5 variables	
V4.C	19359 obs. of 5 variables	
V4.M	19359 obs. of 5 variables	

The Observations should be exactly the same

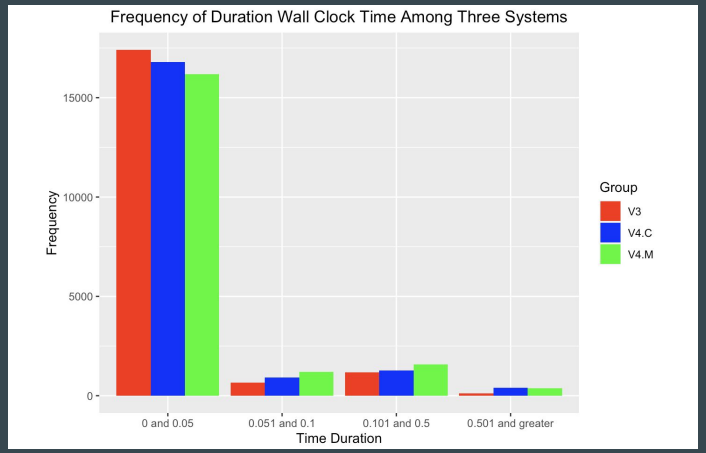
```
[1] 3985
[1] 3987 3988 3989
[1] 3992
[1] 4398
```

Index that which queries are missing

EDA Plot of Real Run Time

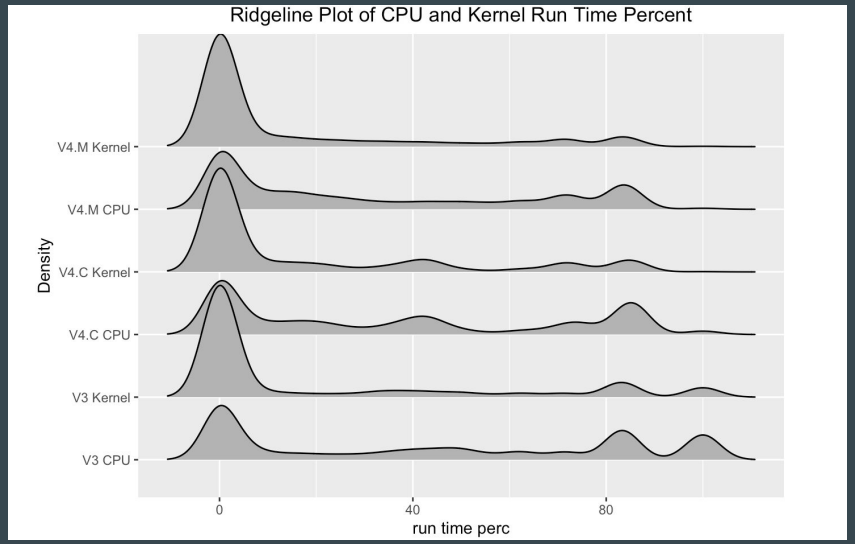


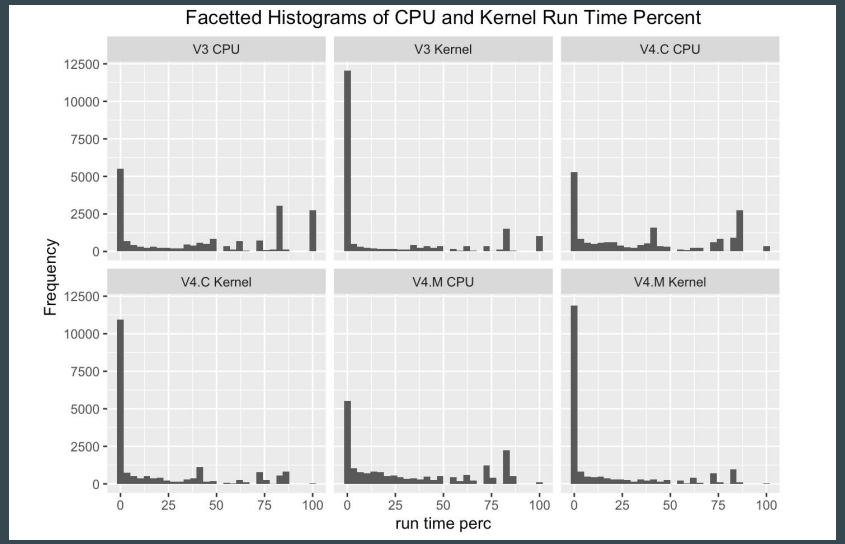
Frequency of top 3 shortest real time



Frequency of time duration of real time

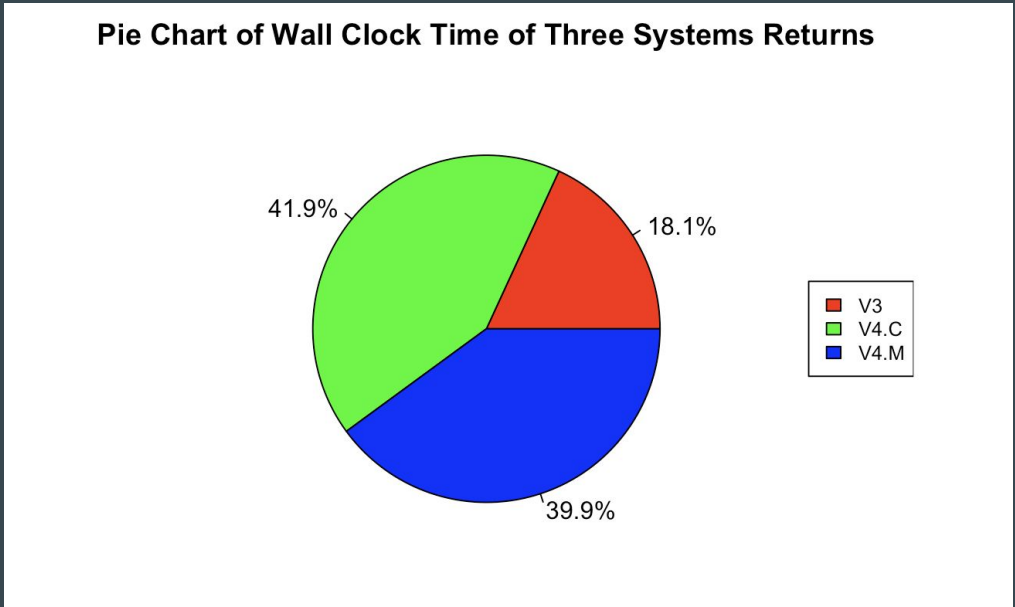
EDA Plot of CPU and Kernel Run Time





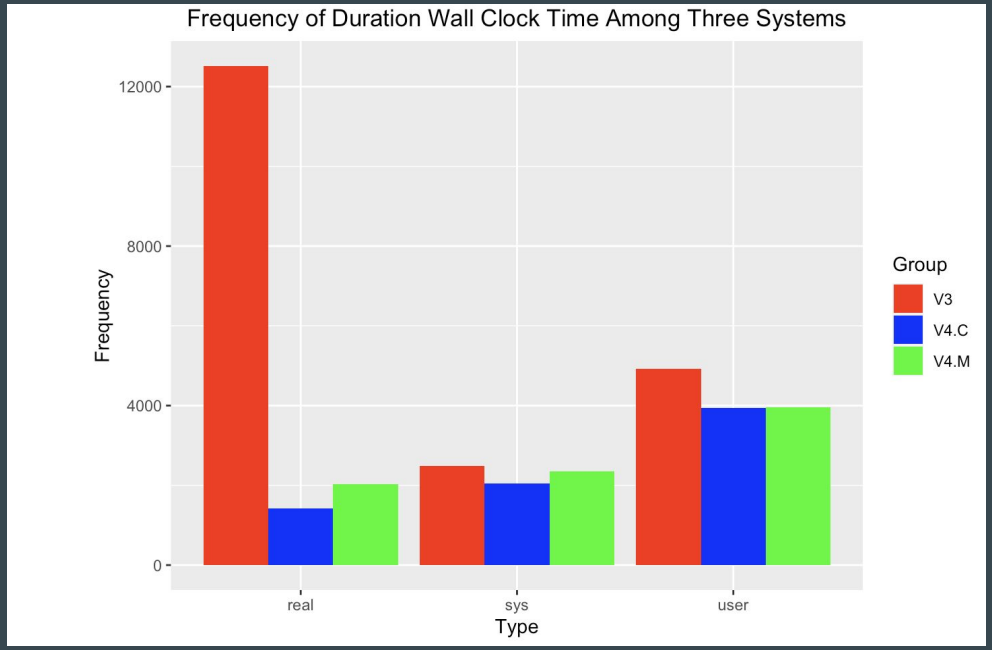
V4.M Kernel and V4.M CPU seems to have advantage

EDA Plot of Query That Returns Something on Real Time



This is particularly interesting to see for query that returns something V3 outperforms

EDA of Pairwise Comparison



V3 outperforms in wall-clock time, but not for CPU and Kernel

Conclusion & Future Work

- In this particular dataset, V3 outperformed compare on wall-clock time
- CPU and Kernel Run time are roughly the same
- Have done some work relates to query history
- Assumption about V3, V4.C, and V4.M

Some Work About Query History

	type1 <chr>	type2 <chr>	type3 <chr>	type4 <chr>
3987	asof-0x1	geo-0x2	signal-0x1	time-0xRANGE
4485	geo-0x200	lag-0x1	signal-0x1	time-0x18
44851	geo-0x200	lag-0x1	signal-0x1	time-0x18
39871	asof-0x1	geo-0x2	signal-0x1	time-0xRANGE
39872	asof-0x1	geo-0x2	signal-0x1	time-0xRANGE
39873	asof-0x1	geo-0x2	signal-0x1	time-0xRANGE
5241	geo-0xSTAR	issue-0xRANGE	signal-0x1	time-0x1
1	asof-0x1	geo-0x1	signal-0x1	time-0x120
6071	geo-0xSTAR	issue-0xRANGE	signal-0x1	time-0x1

```
## $type1
##
##   asof-0x1    geo-0x1 geo-0xSTAR
##         267         65        3578
##
## $type2
##
##       geo-0x1    geo-0xSTAR issue-0xRANGE
##           261             6           3643
##
## $type3
##
## signal-0x1
##       3910
##
## $type4
##
## time-0x1
##     3910
```

Query with max run time for each system

The frequency of each query type in smallest 5% run time in each system

Q&A