Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Systems@ETH Zürich

Cloud computing architecture

Semester project report

Group XXX
Author 1 - Legi Nr
Author 2 - Legi Nr
Author 3 - Legi Nr

Systems Group
Department of Computer Science
ETH Zurich
February 26, 2021

Instructions

Please do not modify the template, except for putting your solutions, names and legi-NR. Parts 1 and 2 should be answered in maximum six pages (including the questions). **If you exceed the space, points might be subtracted**.

Part 1 [20 points]

(a) **[10 points]** Plot a single line graph with 95th percentile latency on the y-axis (the y-axis should range from 0 to 5 ms) and QPS on the x-axis (the x-axis should range from 0 to 70K). Label your axes. State how many runs you averaged across (we recommend three) and include error bars. There should be 7 lines in total in your plot, showing the performance of memcached running with no interference and six different sources of interference: cpu, l1d, l1i, l2, l3, membw. The readability of your plot will be part of your grade.

(b) **[6 points]** Describe how the tail latency and saturation point (the "knee in the curve") of memcached is affected by each type of interference. Also describe your hypothesis for why memcached performance is affected in this way.

(c) **[2 points]** Explain the use of the taskset command in the container commands for memcached and iBench in the provided scripts. Why do we run some of the iBench benchmarks on the same core as memcached and others on a different core?

(d) **[2 point]** Assuming a service level objective (SLO) for memcached of up to 2 ms 95th percentile latency at 40K QPS, which iBench source of interference can safely be collocated with memcached without violating this SLO? Explain your reasoning.

Part 2 [25 points]

1. **[12 points]** Fill in the following table with the normalized execution time of each batch job with each source of interference. The execution time should be normalized to the job's execution time with no interference. Color-code each field in the table as follows: green if the normalized execution time is less than or equal to 1.3, orange if the normalized execution time is over 1.3 and up to 2, and red if the normalized execution time is greater than 2. Summarize in a paragraph the resource interference sensitivity of each batch job.

Workload	`none`	`cpu`	`l1d`	`l1i`	`l2`	`llc`	`memBW`
dedup	1.0						
blackscholes	1.0						
ferret	1.0						
freqmine	1.0						
canneal	1.0						
fft	1.0						

2. **[3 points]** Explain in a few sentences what the interference profile table tells you about the resource requirements for each application. Which jobs (if any) seem like good candidates to collocate with memcached from Part 1, without violating the SLO of 2 ms P95 latency at 40K QPS?

3. **[10 points]** Plot a single line graph with speedup as the y-axis (normalized time to the single thread config, Time_1 / Time_n) vs. number of threads on the x-axis. Briefly discuss the scalability of each application: e.g., linear/sub-linear/super-linear. Do any of the applications gain a significant speedup with more threads? Explain what you consider to be "significant".