A Fast Path to Benchmarking

Mark Brown
LPC 2025, Tokyo

Agenda

- What do we mean by benchmarking?
- Tooling
  - What sort of structure is useful to put around benchmarks?
  - Through the medium of Fastpath
- Interesting benchmarks

What is a benchmark?

- Most testing focuses on tests that either pass or fail
  - Clear and easy to think about
  - Doesn’t really work for performance
  - Not ideal for SLAs
- Benchmarks track how numbers change
  - Did the number change?
  - How did it change?
  - How confident are we in that number anyway?

Complexities

- Generality
- Sensitivity
- Reproducibility
- Resource requirements
- Complexity

Fastpath

- Command line tool
  - Run benchmarks
    - Interactively
    - Automatically in CI
  - Manage results
  - UI for analysis
- Released earlier this year
  - https://fastpath.docs.arm.com/en/latest/
- Written by
  - Ryan Roberts
  - Aishwarya Rambhadran
  - Aishwarya TCV

Execution environment

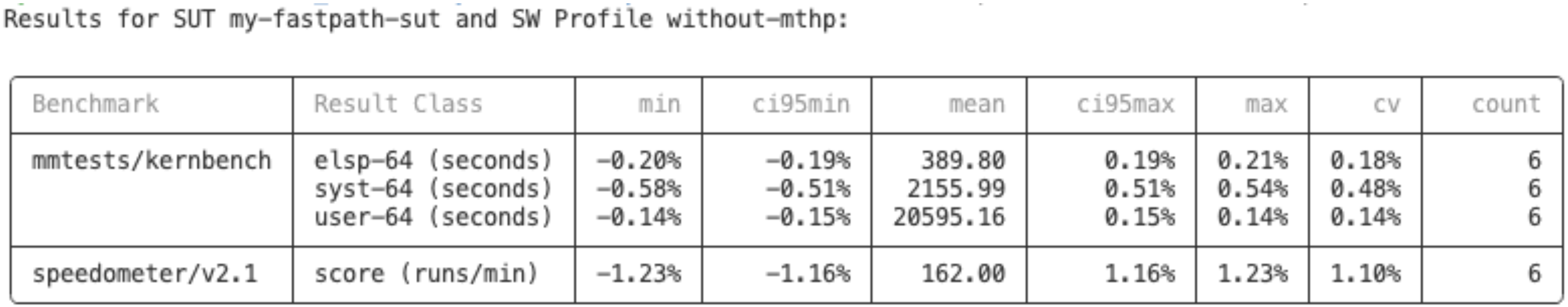
- Standard problem for testing
  - Hardware and software stack impacts results
- Full system testing brings in a much bigger software stack
  - Not great if you deal with a diverse range of systems

- Containerize the tests and their dependencies
  - Move a minimal set of software into the system description

Trusting results

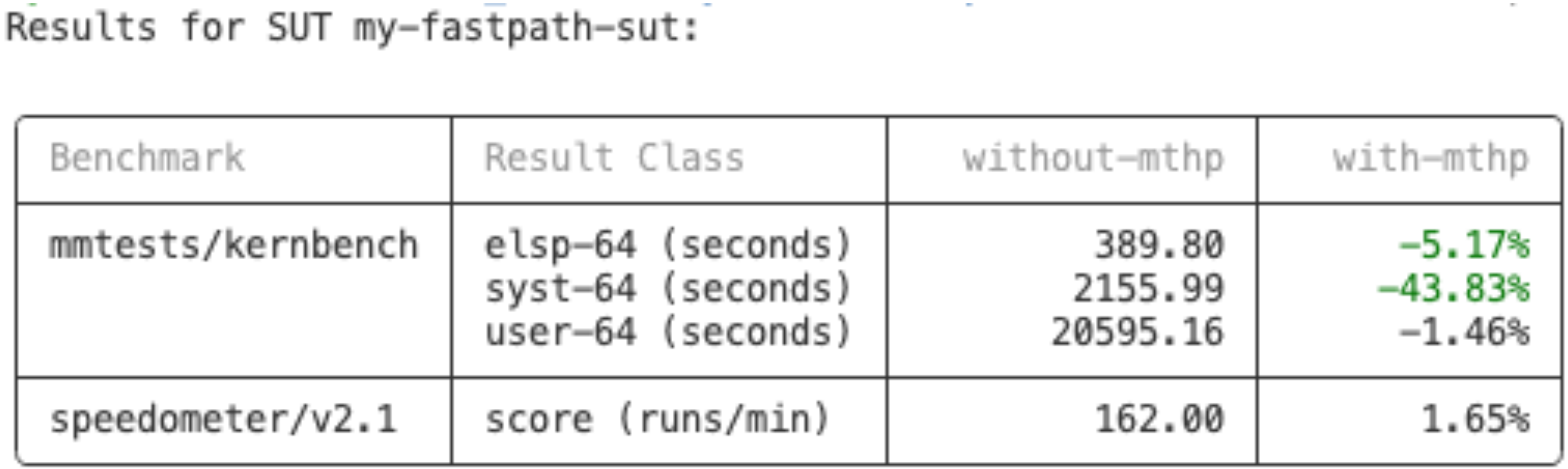
- Exact reproducibility is typically unachievable – time for statistics!
- Run things repeatedly
  - Possibly with warmup runs to get caches warm
- Combine into summary numbers
  - Averages
  - Minimum and maximum
  - Standard deviation
  - Confidence intervals
  - Coefficient of variation
  - Quartile coefficient of dispersion

Compare results



Results for SUT my-fastpath-sut and SW Profile without-mthp:

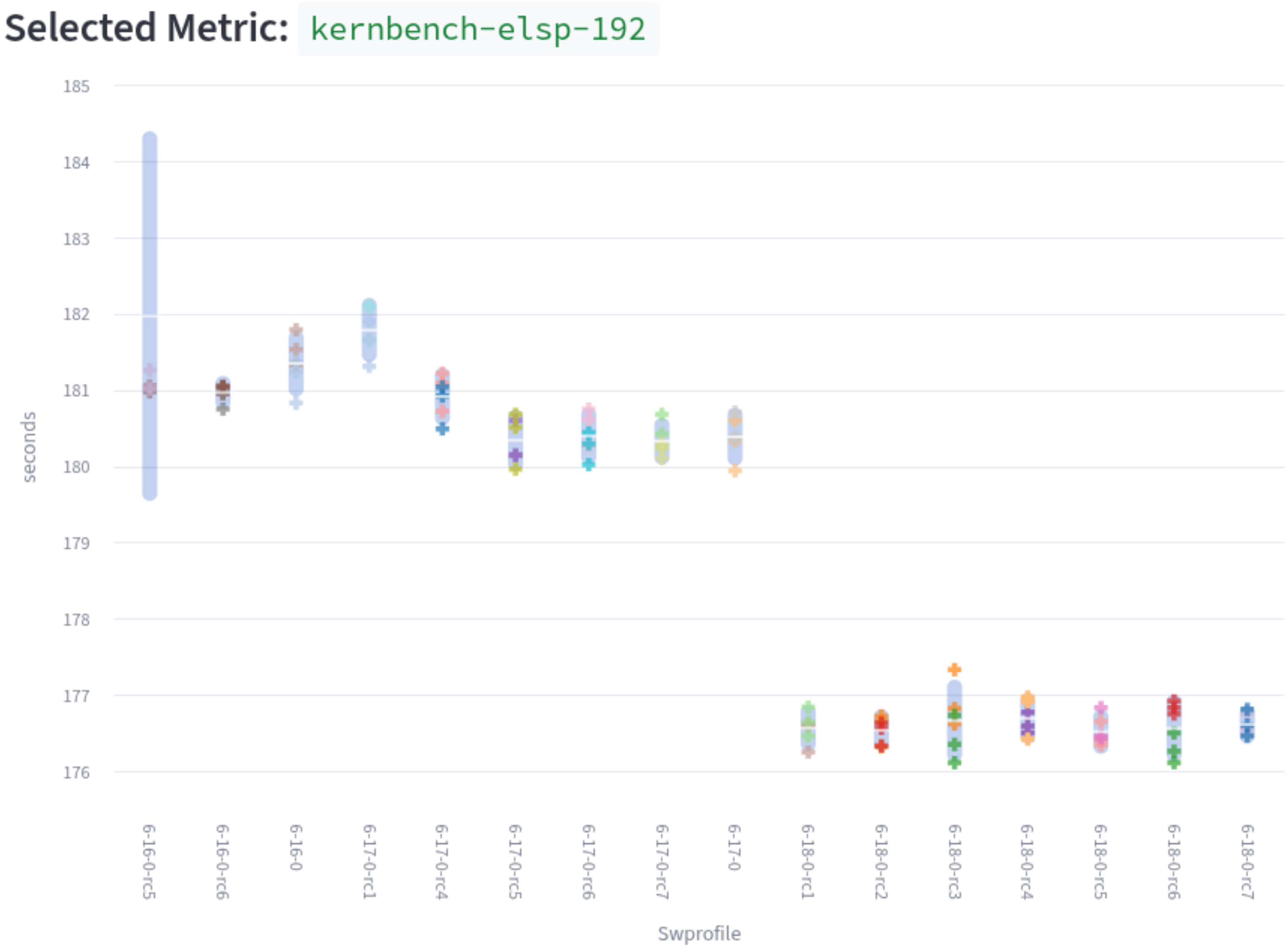
Benchmark	Result Class	min	ci95min	mean	ci95max	max	cv	count
mmtests/kernbench	elsp-64 (seconds)	-0.20%	-0.19%	389.80	0.19%	0.21%	0.18%	6
	syst-64 (seconds)	-0.58%	-0.51%	2155.99	0.51%	0.54%	0.48%	6
	user-64 (seconds)	-0.14%	-0.15%	20595.16	0.15%	0.14%	0.14%	6
speedometer/v2.1	score (runs/min)	-1.23%	-1.16%	162.00	1.16%	1.23%	1.10%	6



Results for SUT my-fastpath-sut:

Benchmark	Result Class	without-mthp	with-mthp
mmtests/kernbench	elsp-64 (seconds)	389.80	-5.17%
	syst-64 (seconds)	2155.99	-43.83%
	user-64 (seconds)	20595.16	-1.46%
speedometer/v2.1	score (runs/min)	162.00	1.65%

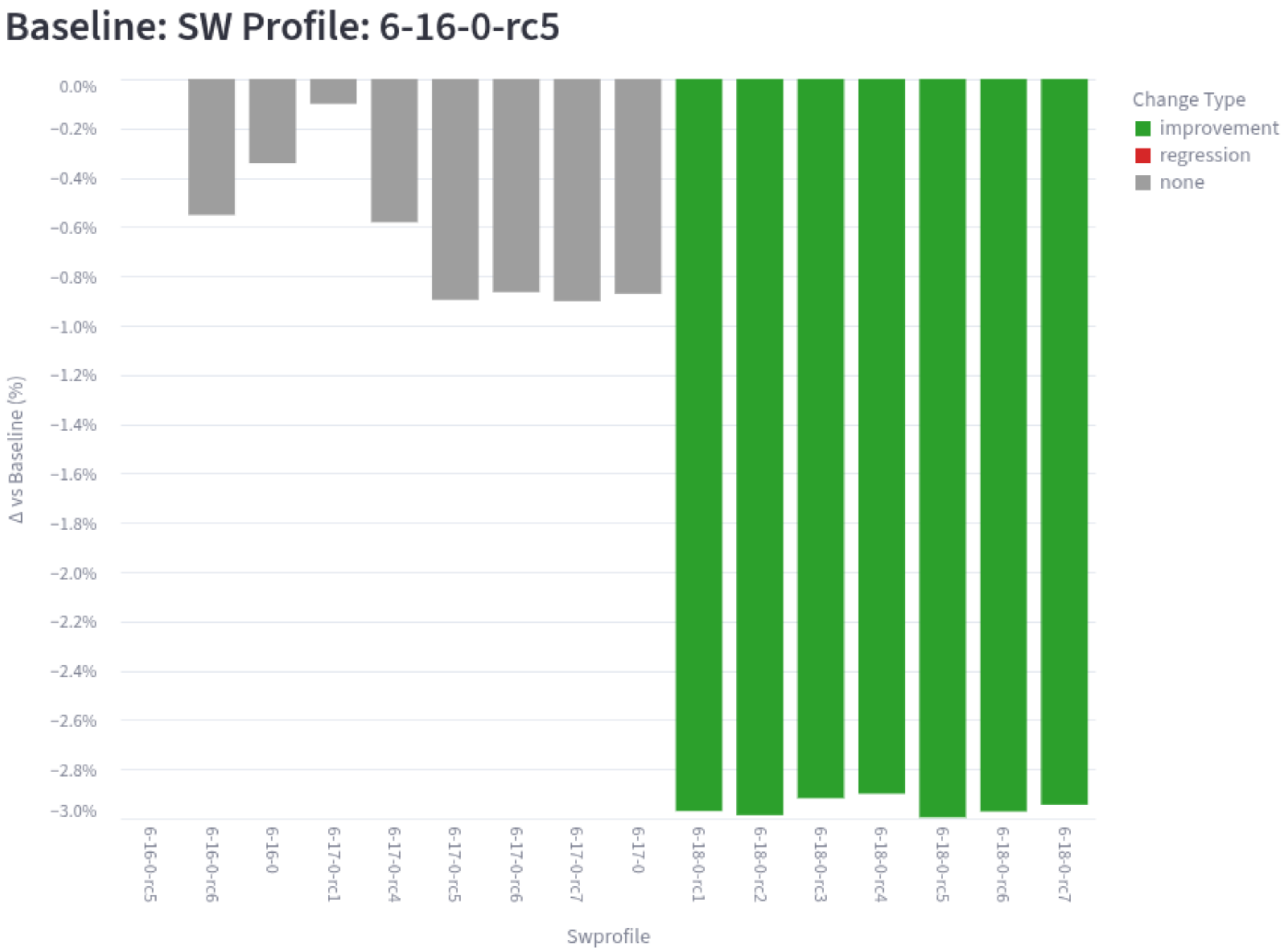
Compare results – with pictures!



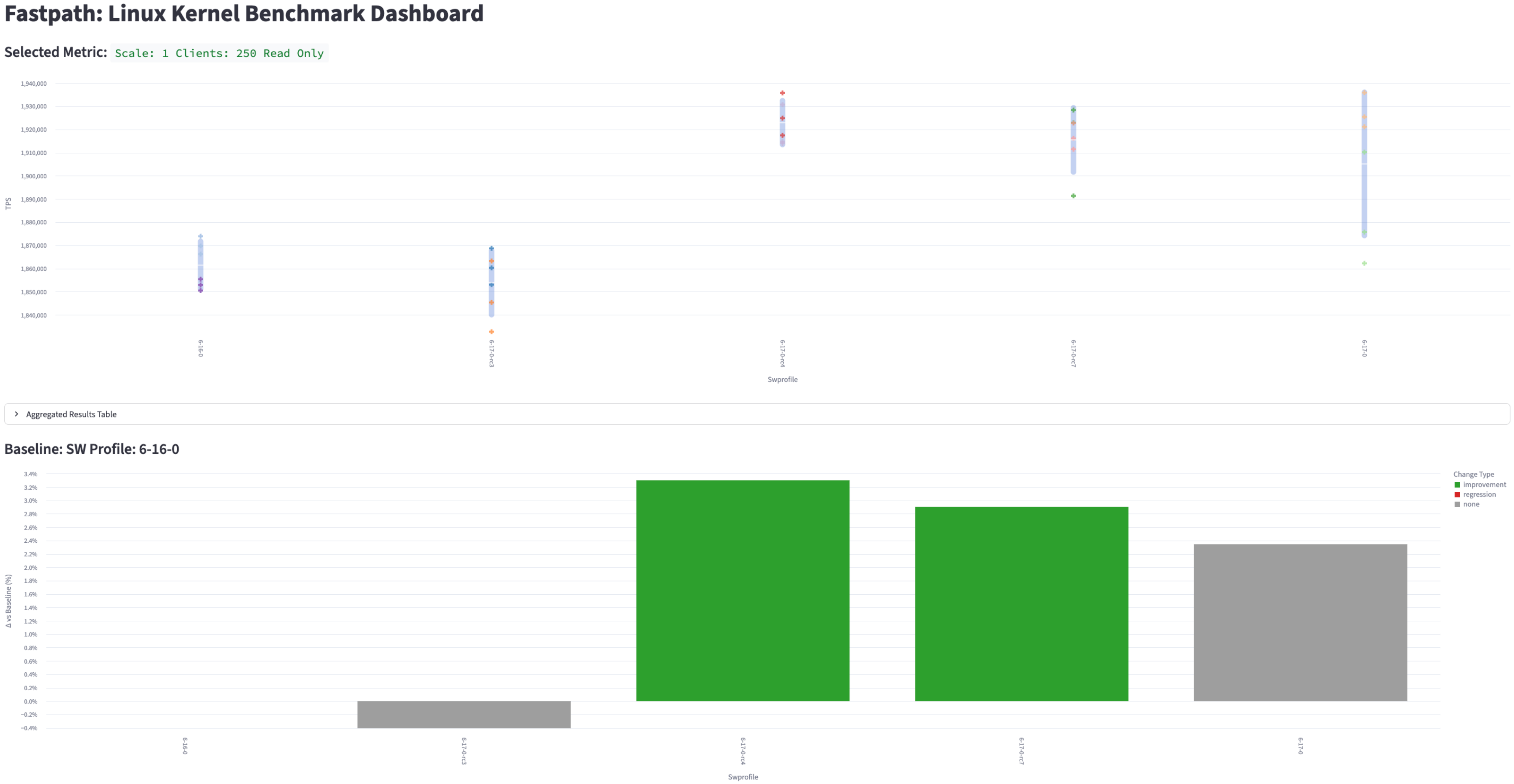
Selected Metric: `kernbench-elsp-192`

arm

Compare results – with pictures!



Compare results – with pictures!



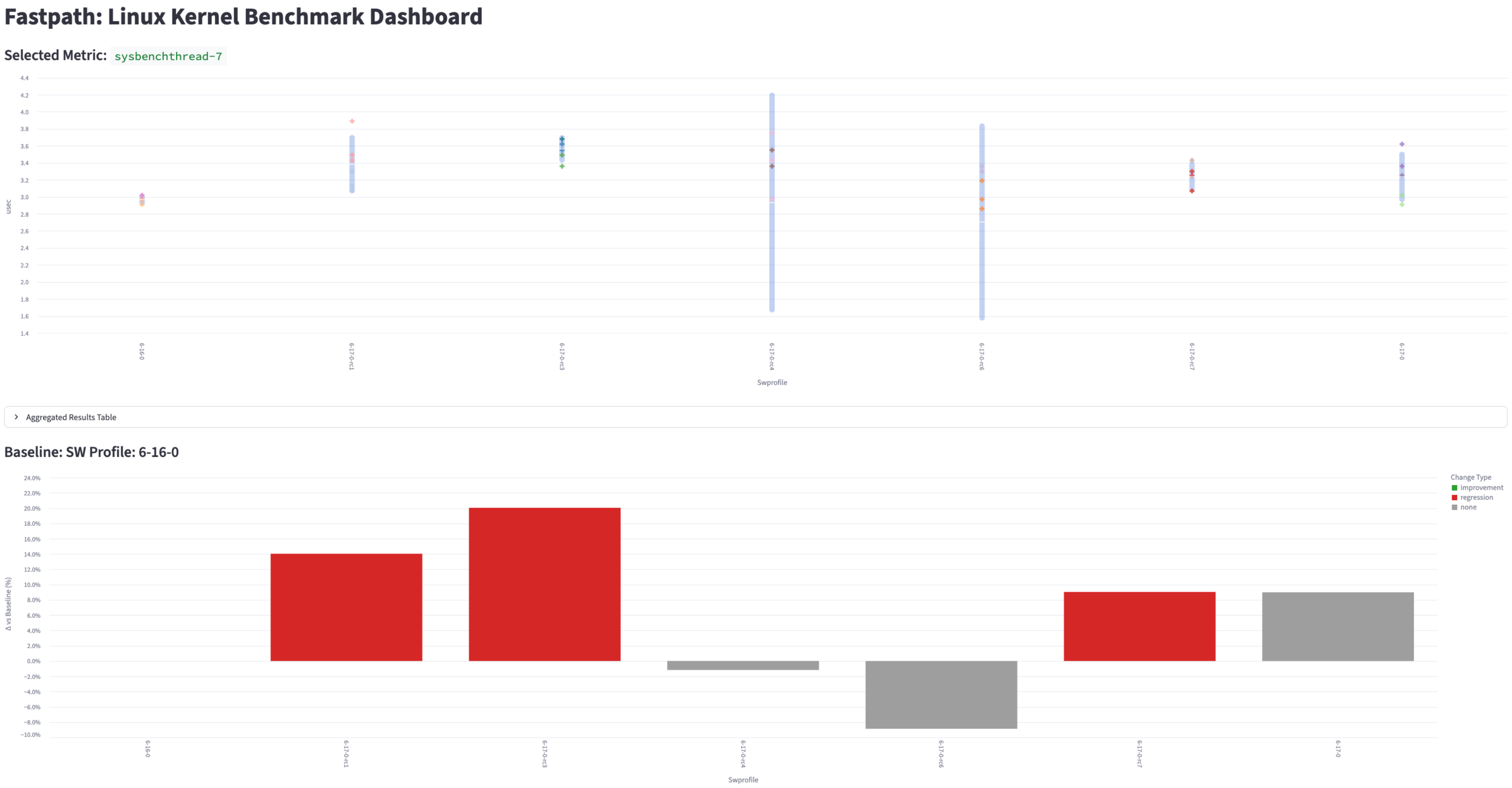
Fastpath: Linux Kernel Benchmark Dashboard

Selected Metric: `Scale: 1 Clients: 250 Read Only`

› Aggregated Results Table

Baseline: SW Profile: 6-16-0

Compare results – with pictures!



Fastpath: Linux Kernel Benchmark Dashboard

Selected Metric: `sysbenchthread-7`

› Aggregated Results Table

Baseline: SW Profile: 6-16-0

Compare results – with automation!

- Turn the statistics into pass/fail results
  - Adequate stability
  - Non-overlapping 95% confidence intervals
- Turn pass/fail results into
  - Bisections
  - Email reports
- Optimisations
  - Scale number of iterations based on statistical separation

Benchmarks

Microbenchmarks

- bonnie
- cyclictest
- fio
- lmbench
- netperf
- pidbench
- timerlat

Omnibus benchmarks

- mmtests
  - hackbench
  - kernbench
  - Sysbench
  - …
- Phoronix Test Suite
  - memtier
  - nginx
  - pgbench
  - pybench
  - …
- Geekbench
- sysbench
- Will-it-scale

Application benchmarks

- compose-samples
- glmark2
- schbench
- perf bench
- speedometer
- uibench

arm

Merci
Danke
Gracias
Grazie
谢谢
ありがとう
Asante
Thank You
감사합니다
धन्यवाद
Kiitos
شكرًا
ধন্যবাদ
תודה
ధన్యవాదములు
Köszönöm