MEASUREMENT · LOAD TEST

Precise, repeated, wrong

18,750 req/s: three repetitions agreed closely, and the number was three times wrong.

Muhammet Şafak • October 3, 2026



SAME RUN, SAME SERVICE

Two phases, a threefold gap

	Ladder (open loop)	Tuning phase (closed loop)
GO, 4 CORES	18,750 req/s	✓ 55,715 req/s
WITHIN ITSELF	Consistent	✓ Consistent
IN HINDSIGHT	Turned out wrong	✓ Turned out right

WHICH ONE IS WRONG?

Consistency is not accuracy.

Both measurements were consistent within themselves, and nothing said which of them was wrong.

LOOKING AT THE DATA

The repetitions ranged from 1,250 to 48,000.

The median was 18,750, because that is where the middle repetition happened to land.

FIRST DIAGNOSIS

It was correct and not enough.

The writeback of the CHECKPOINT issued before every step spilled into the measurement window: at 3,000 req/s, the write p99 was 10.6 ms, the read 3.3 ms.

SECOND ATTEMPT

A fivefold divergence, zero spoiled windows

CELL	REPETITION 1	REPETITION 2	REPETITION 3
Go, 1 core	1,250	6,500	3,500
PHP-FPM, 1 core	1,750	2,750	750

PHP-FPM, 1 CORE, 3,000 REQ/S

oha reports two latencies

LATENCY	P50	P75	P99	SLOWEST
Time to first byte	1.42 ms	47 ms	191 ms	267 ms
Latency-corrected	1.44 ms	1,539 ms	3,718 ms	3,819 ms

MECHANISM

The correction turns the limit into a cliff

1

Counts from the schedule

the moment it should have been sent

2

Responses slow down

connections fill up

3

The schedule slips

every request inherits it

4

Cliff

either passes or misses

WHAT THE SEARCH COULD NOT MEASURE

The cliff moves between repetitions.

Exponential search and bisection assume that pass/fail is monotonic in the rate. Neither was true.

THE FIX

Drop the search, set fixed rates

7

Fixed rates

Per cell, the same in every repetition

Vote

Majority rule

If most repetitions carried it, it is carried

115%

Probe threshold

Before every repetition

ONE MORE LIE

Chromium on the host machine

METRIC	VALUE
Headless Chromium processes	About thirty
Total CPU	1,019%
Left for the Docker VM	94%
Go, 4 cores	17,490 req/s

RESULT · GO, 4 CORES

Two methods, within a few percent

MEASUREMENT	REQ/S	NOTE
Tuning phase (closed loop)	55,715	Twelve hours earlier
Five repetitions (quiet machine)	55,836–57,321	Median 56,571; 45/45 repetitions passed the probe

FOR YOUR OWN SETUP

Three concrete additions

- [ ] A second phase that measures in a different loop, and a rule that voids the run when the two diverge
- [ ] A probe before every repetition that runs no application code
- [ ] A record of the environment alongside the repetition



The second road

muhammetsafak.com/research/capacity-per-core-go-php-oauth2/

The raw data of the numbers and the discarded attempts is in the research notebook.

