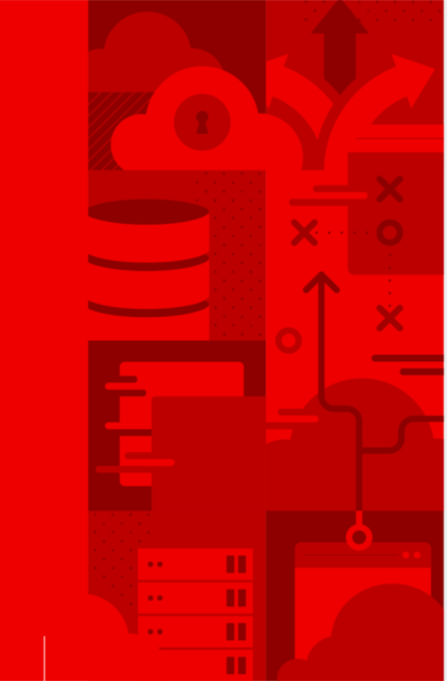


Demanding the impossible: rigorous database benchmarking

DMITRII DOLGOV

18-06-2023

Choose your fighter:

github.com/cmu-db/benchbase
github.com/akopytov/sysbench
github.com/brianfrankcooper/YCSB
github.com/TPC-Council/HammerDB
postgresql.org/docs/current/pgbench.html

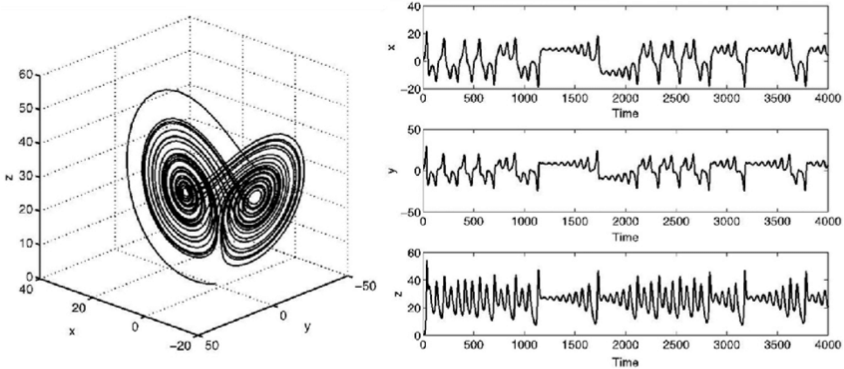
Replicated live workload

```
latency average = 0.011 ms
latency stddev = 0.002 ms
tps = 89357.630697 (without initial connection time)
```

```
latency average = 0.011 ms
latency stddev = 0.002 ms
tps = 89357.630697 (without initial connection time)

latency average = 0.014 ms
latency stddev = 0.023 ms
tps = 67107.536620 (without initial connection time)
```

Benchmarking model



The phase space plot of the Lorenz attractor,

"Nonlinear time series methods for analyzing behavioral sequences"

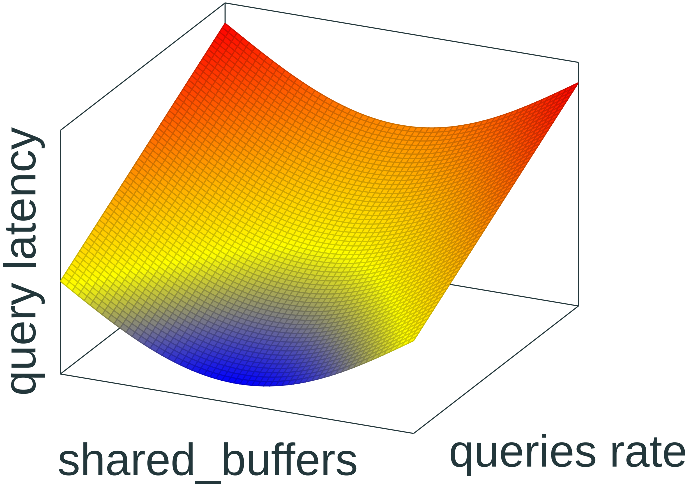
Dimentions?

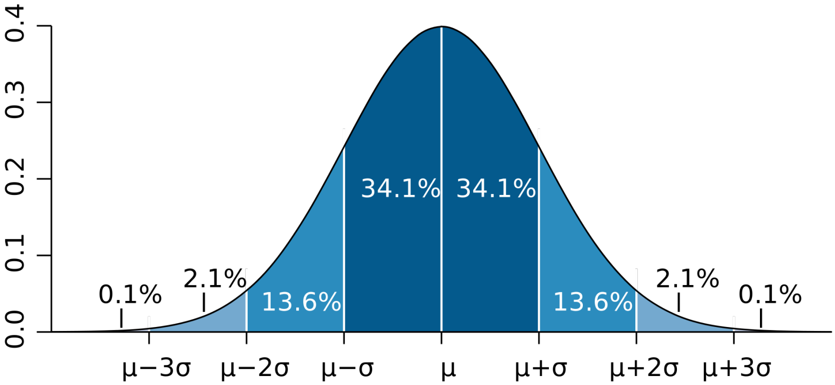
DB parameters

Hardware resources

Workload parameters

Performance results


query latency
shared_buffers
queries rate



Probability distribution

Benchmarking is exploring the system's **known** properties in presence of **unknown** factors.

Red Hat

PostgreSQL specifics

Too low or too high?

```
shared_buffers
max_wal_size
work_mem
checkpoint_timeout
checkpoint_completion_target
wal_writer_flush_after
checkpoint_flush_after
[ … ]
```

Too low or too high?

```
vm.nr_hugepages
vm.dirty_background_bytes
vm.dirty_bytes
block/<dev>/queue/read_ahead_kb
block/<dev>/queue/scheduler
[ … ]
```

Noise

CPU/NUMA pinning, p-state, frequency scaling

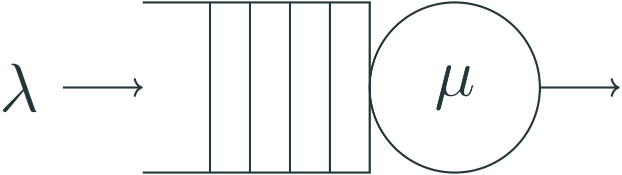
Files creation, NVMe trim

Noisy neighbors, virtualized infrastructure

How long?

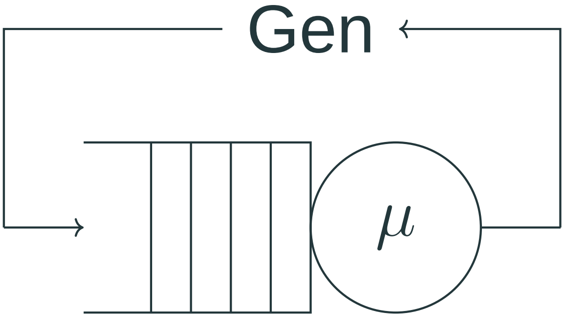
```
autovacuum_naptime = 1min
autovacuum_vacuum_threshold = 50
autovacuum_vacuum_insert_threshold = 1000
autovacuum_vacuum_scale_factor = 0.2
autovacuum_vacuum_insert_scale_factor = 0.2
autovacuum_vacuum_cost_delay = 2ms
autovacuum_vacuum_cost_limit = -1
```

The load generator?



”Open versus closed: A cautionary tale”. Schroeder, B., Wierman, A. and Harchol-Balter, M., USENIX. 2006.

The load generator?



”Open versus closed: A cautionary tale”. Schroeder, B., Wierman, A. and Harchol-Balter, M., USENIX. 2006.

Statistics

Now any series of experiments is only of value in so far as it enables us to form a judgement as to the statistical constants of the population to which the experiment belong.

Student, 1908. The probable error of a mean. Biometrika, 6(1), pp.1-25.

Population, metrics

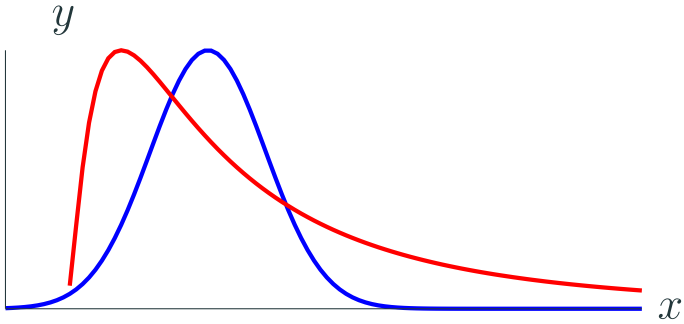
$$\mu = E(x), \sigma = \sqrt{E[(X-\mu)^2]}$$

Samples, statistics

$$\overline{X}, s_N = \sqrt{\tfrac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

t-test

$$t = \tfrac{\overline{X}-\mu}{s/\sqrt{n}}, [\overline{x} \pm \tfrac{cs}{\sqrt{n}}]$$



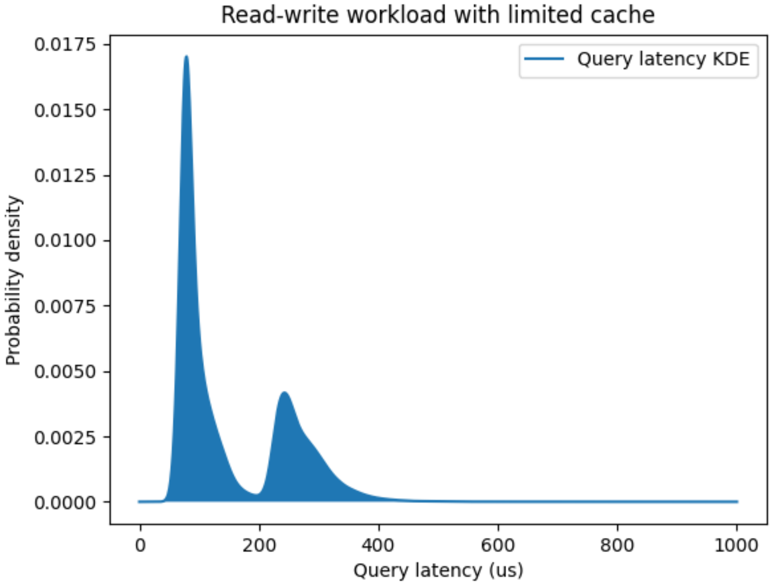
Hoefler, T. and Belli, R., 2015, November. Scientific benchmarking of parallel computing systems: twelve ways to tell the masses when reporting performance results. In Proceedings of the international conference for high performance computing, networking, storage and analysis (pp. 1-12).

Tests for comparison of means are not affected very much by the distributional nonnormality, but this does not extend to the comparison of variances.

Box, G.E., Hunter, J.S. and Hunter, W.G., 2005. Statistics for experimenters. In Wiley series in probability and statistics. Hoboken, NJ: Wiley.

Fleming, M., Kolaczkowski, P., Kumar, I., Das, S., McCarthy, S., Pattabhiraman, P. and Ingo, H., 2023, April. Hunter: Using Change Point Detection to Hunt for Performance Regressions. In Proceedings of the 2023 ACM/SPEC International Conference on Performance Engineering (pp. 199-206).

clickhouse.com/docs/en/operations/utilities/clickhouse-benchmark


Read-write workload with limited cache
Query latency KDE
Probability density
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
0
200
400
600
800
1000
Query latency (us)

```
@usecs:
[16, 32)        32 |                                    |
[32, 64)       202 |                                    |
[64, 128)   169897 |@@@@@@                              |
[128, 256)  679545 |@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@|
[256, 512)   20950 |                                    |
[512, 1K)      378 |                                    |
[1K, 2K)       118 |                                    |
[2K, 4K)       133 |                                    |
[4K, 8K)       306 |                                    |
```

Median, quantiles, IQR

```
scipy.stats.mannwhitneyu
```

How many runs, $E(1\%, 95\%, X)$?

$$CoV \approx 0.3\% \rightarrow E(1\%, 95\%, X) \approx 10$$
$$CoV \approx 9.0\% \rightarrow E(1\%, 95\%, X) \approx 240$$

Maricq, A., Duplyakin, D., Jimenez, I., Maltzahn, C., Stutsman, R. and Ricci, R., 2018. Taming performance variability. In 13th USENIX Symposium on Operating Systems Design and Implementation (OSDI 18) (pp. 409-425).

Time average vs ensemble average?

For an ergodic system: $\overline{N}^{TimeAvg} = \overline{N}^{Ensemble}$

Harchol-Balter, M., 2013. Performance modeling and design of computer systems: queueing theory in action. Cambridge University Press.

Paired difference test.

Randomized testing.

Final thoughts

Benchmarking is exploring

Known vs Unknown

Common vs Particular

Statistical approach for clear communication

Questions?

@erthalion@fosstodon.org

✉ ddolgov at redhat dot com