

Using e-divisive means change detection in continuous benchmarking

June 6th, 2025
@ SPEC DevOps Performance WG

Henrik Ingo, Nyrkiö
Founder & CEO



Nyrkiö for faster software

Capture every perf regression

Using state of art statistics

algorithms and automation

Continuous Performance Engineering

Before:

- Unit testing is automated in CI, but performance testing is not
- Performance is a specialist role or team
- With custom, often in-house tooling
- Focus on Release Candidate
- -> Delay GA release
- Goal is to maintain same overall performance, not to find specific regressions.

Future:

- Continuous Benchmarking in CI
- Ultimately specialist Performance Engineers will go extinct, just like QA
- Standardized tooling:
green checkmark in your pull requests
- Left Shift:
 - Benchmarks before merging
 - ...After merging
 - ...And nightly
 - Find and fix regressions as they happen

The Need for Change Point Detection

Benchmarks don't produce binary pass/fail results

Numeric timeseries: Constant with random noise + regressions/improvements

End users might notice 10% regression

Problem: This is also the level of noise in your testing environment

No, it's not a cloud problem!

MongoDB circa 2016:

- 20-30 large benchmarks, daily
- ~100 microbenchmarks, hourly
- Graphs + **threshold based alerting**
- **99% false positive rate**
- Build Lead duty 1 week per month

E-Divisive Means (Matteson & James, 2013)

"If you understand the difference between Anomaly Detection and Change Point Detection, and the filter out all the papers dealing with seasonality or other trends... the choice was obvious."

David Daly

blog.nyrkio.com/.../welcome-apache-otava-incubating-project/

Ishita Kumar:

Outperforms DYNP & PELT

- especially with small number of data points. (1-3 vs 3-15)
- 3x higher F1 score

In:

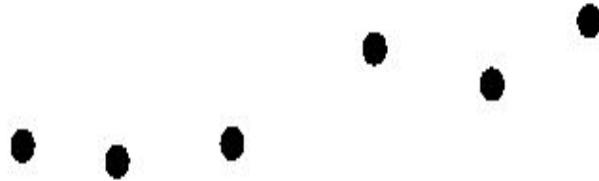
Fleming & Kolaczowski, et.al. 2023

Hunter: Using Change Point Detection to Hunt for Performance Regressions

dl.acm.org/doi/10.1145/3578244.3583719

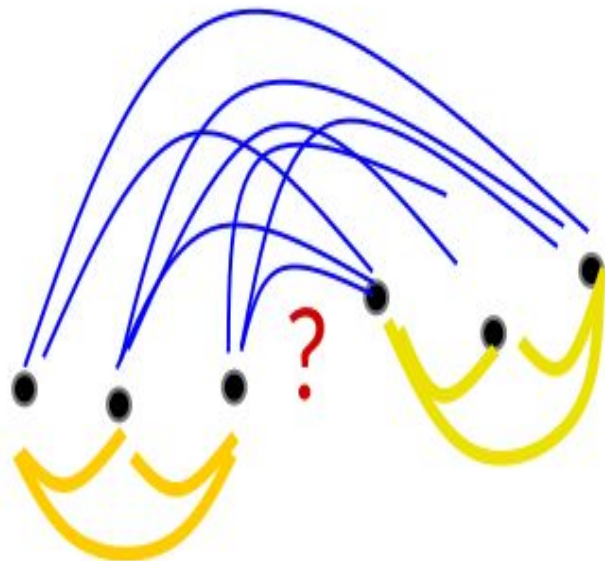
E-divisive means for dummies

(Formally: [Matteson & James, 2013](#))



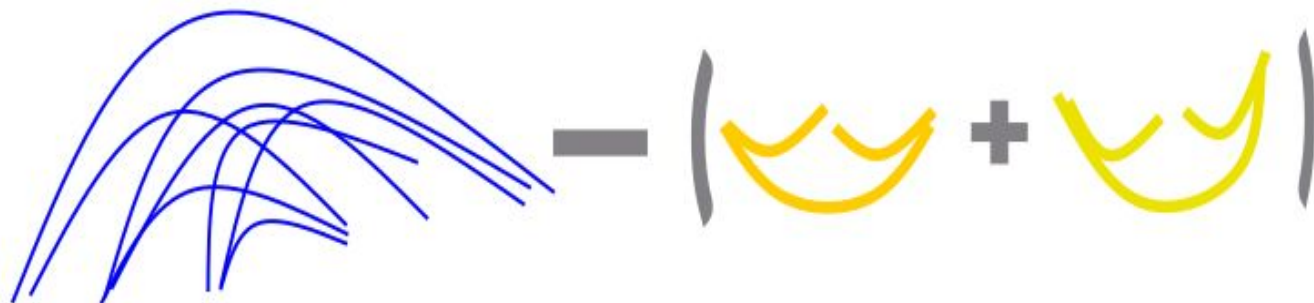
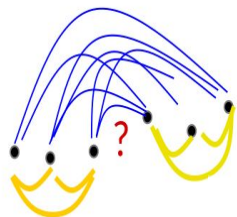
E-divisive means for dummies

(Formally: [Matteson & James, 2013](#))



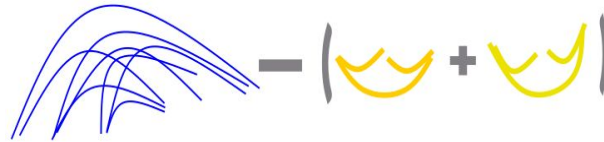
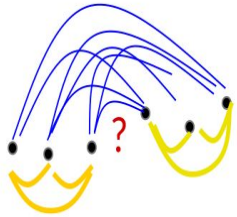
E-divisive means for dummies

(Formally: [Matteson & James, 2013](#))



E-divisive means for dummies

(Formally: [Matteson & James, 2013](#))



Results

Before

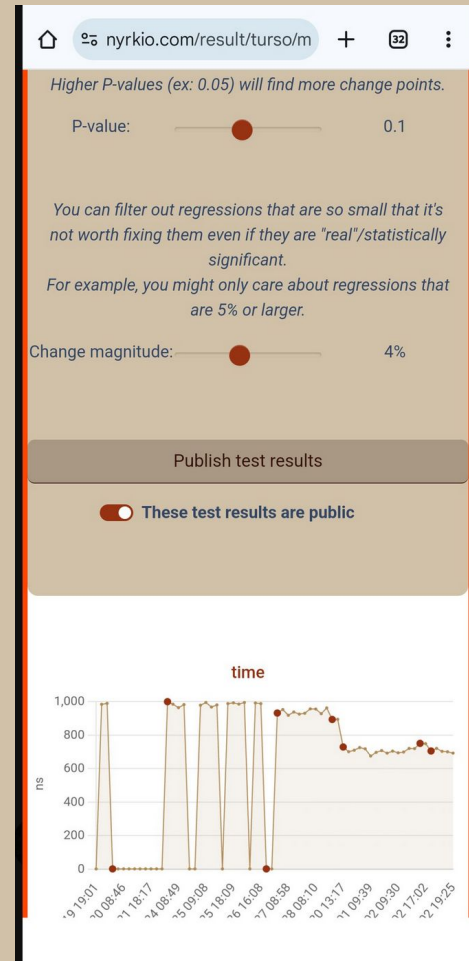
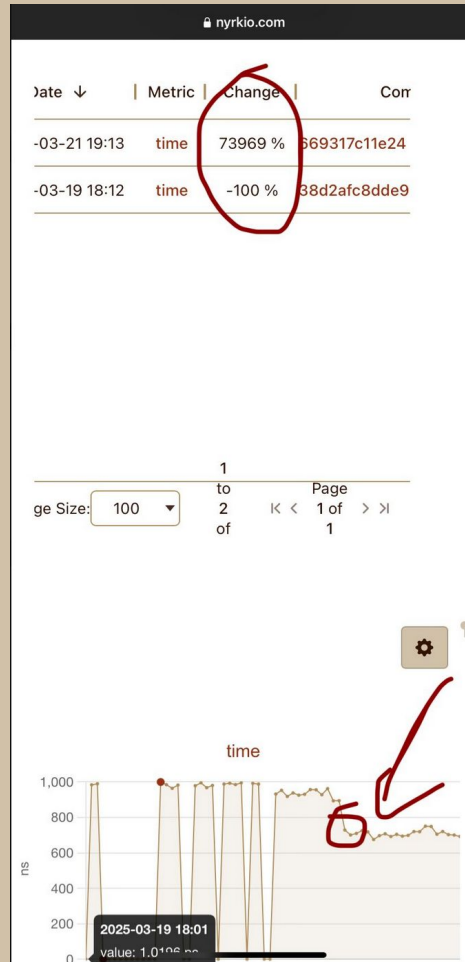
- 95-99% false positives
- Performance team responsible

After

- ~50% false positive rate, tunable
- Every team responsible for perf

Users

MongoDB, Netflix, Datastax, Confluent, Red Hat OpenShift, Nyrkiö, Dremio, TigerBeetle



Demo

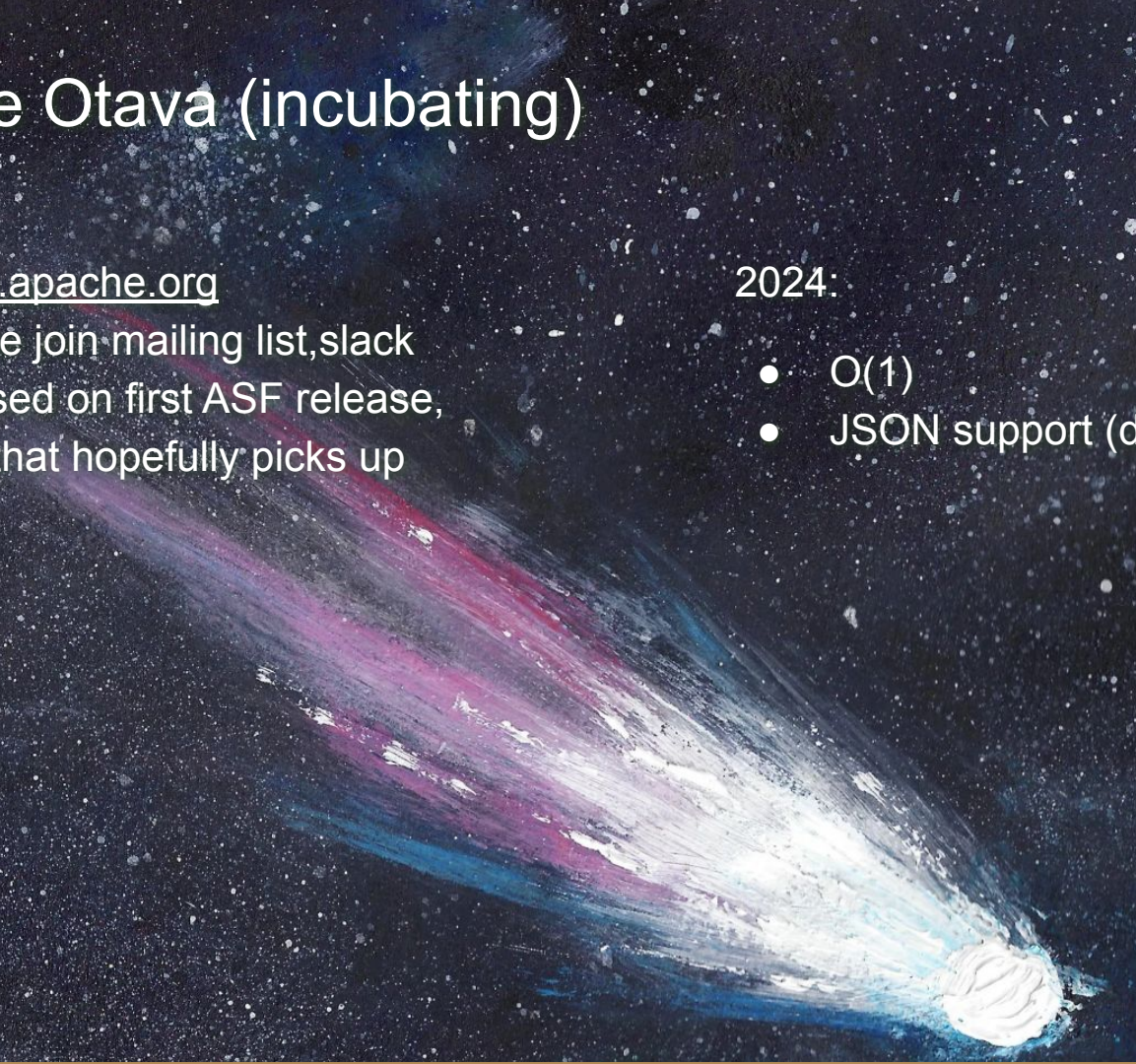
[nyrkio.com/public/https%3A%2F%2Fgithub.com%2Ftigerbeetle%2Ftigerbeetle/main/devhub](https://github.com/tigerbeetle/tigerbeetle/main/devhub)
[nyrkio.com/public/https%3A%2F%2Fgithub.com%2Ftursodatabase%2Flimbo/main/clickbench/sqlite3/main](https://github.com/tursodatabase/limbo/main/clickbench/sqlite3/main)

Apache Otava (incubating)

- otava.apache.org
- Please join mailing list, slack
- Focused on first ASF release, after that hopefully picks up

2024:

- O(1)
- JSON support (default: CSV)





Founded in 2024

Mainstream and commercialize tools we built over the years.
Ultimately goal: Eradicate Performance Engineer as a specialist role.

Self funded.

Also the web service is OSS.

**NYRKIÖ**

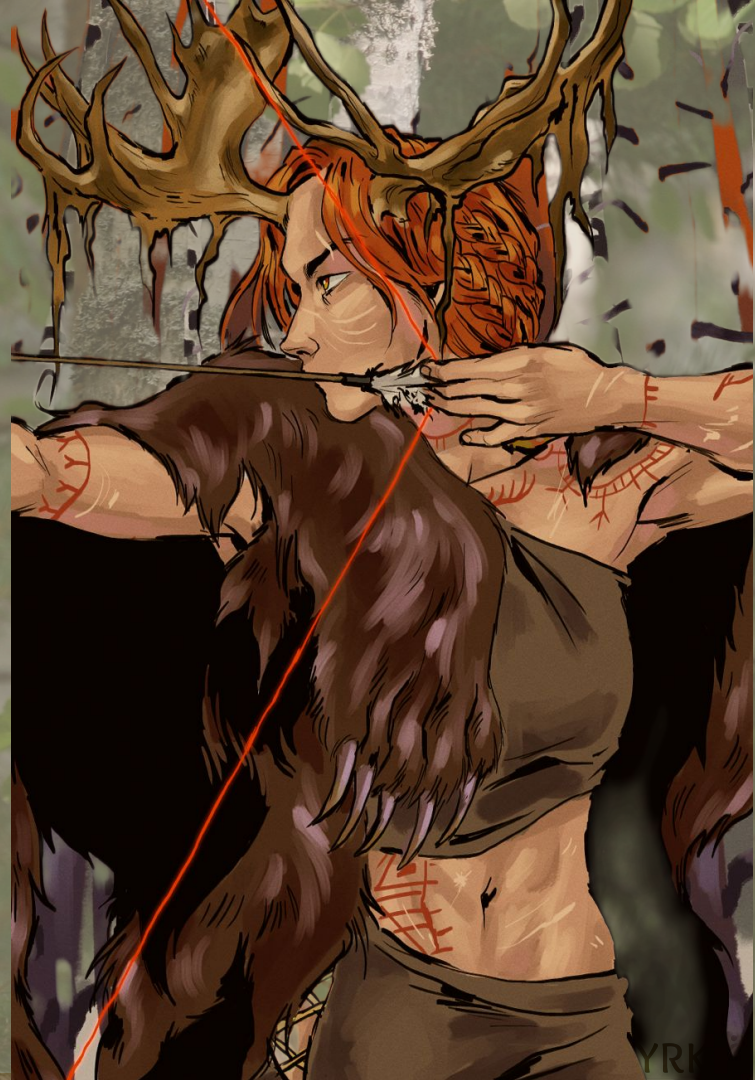
Q&A

Example: What's up with all this forest imagery?

Rene Schwietzke @ flickr

In Finnish mythology, Nyrkiö (or Nyyrikki) is the guardian spirit of hunters and wanderers in the forest. They also take care of small game like squirrels and birds.

The Legend of Nyrkiö



Rene Schwietzke @ flickr

"For Faster Software"

- assonance: repeated rhymes on the prefix of words (as opposed to end)

"Nyrkiö for faster software"

- Kalevala meter
- Trad- Finnish verse = 5 beats per line

Trivia:

- Lord of the Rings, especially Gandalf, based on Kalevala, Finnish national epic

The Legend of Nyrkiö

