Baqend

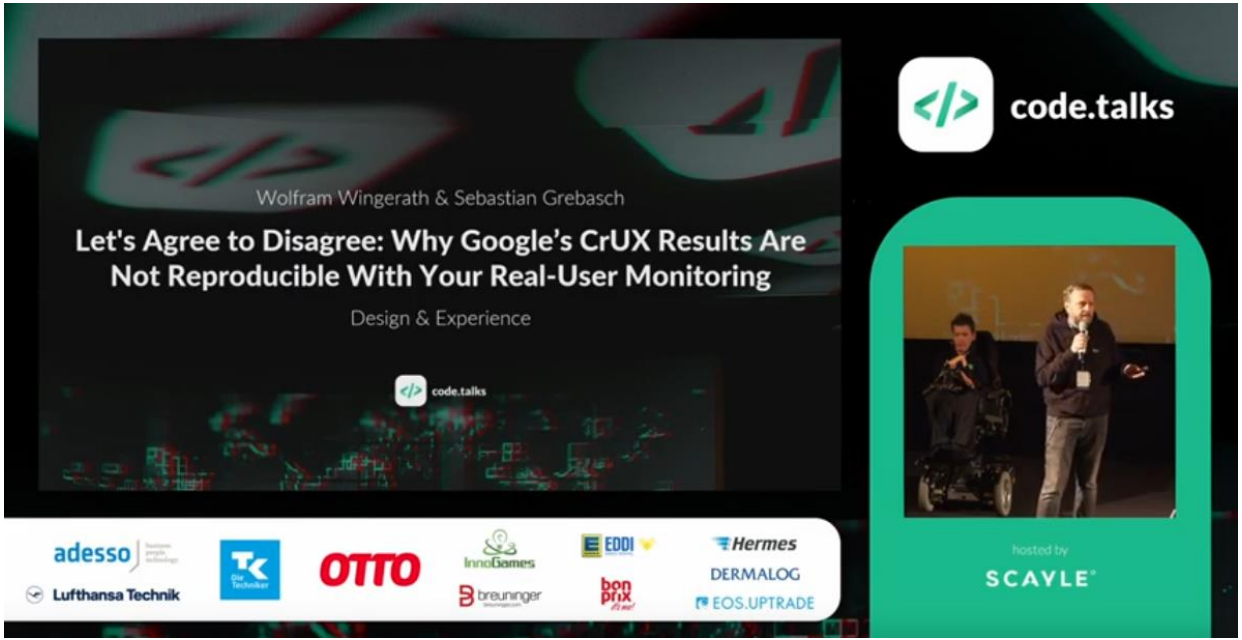
Let‘s Agree to Disagree

Why Google’s CrUX Results Are Not Reproducible With Your Real-User Monitoring

DB/DC Workshop
Dec. 07, 2022
Wolle

Carl von Ossietzky
Universität
Oldenburg

Presented at Code.Talks 2022 together with Google



Wolfram Wingerath, Sebastian Grebasch. Let's Agree to Disagree: Why Google's CrUX Results Are Not Reproducible With Your Real-User Monitoring, CodeTalks (2022)

I Am Wolle

Research:
- Stream Processing
- Real-Time Databases
- NoSQL & Cloud Systems
- ...

Practice:
- Web Caching
- Big Data Analytics
- Anger Management
- ...

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Carl von Ossietzky
Universität
Oldenburg

Baqend

WHO WE ARE

We bring performance research to practice

30+ man-years of **web performance research** at University of Hamburg

Novel technology for **caching dynamic data** went into Baqend in 2014

Baqend **launched Speed Kit** as the all-in-one page speed platform in 2018

7,000
customer websites are already using Speed Kit

$2.6 billion
in annual revenue runs with Speed Kit

160 million
users per month benefit from Speed Kit

You Heard the Stories

amazon — **100 ms** slower → **-1%** Conversion Rate

zalando — **100 ms** faster → **+0.7%** Revenue Per Session

Walmart — **100 ms** faster → **+1%** Revenue

Greg Linden. Make Data Useful. Stanford Data Mining Class CS345A, 2006

Shuhei Kagawa, Jeff Cybulski, David Martin Jones, et. al.. Loading Time Matters. Zalando Tech Blog, 2018

C. Crocker, A. Kulick, B. Ram. Real-User Monitoring at Walmart, SF & SV Web Performance Group, 2012.

You Heard the Stories

amazon

100 ms slower → -1% Conversion Rate

zalando

100 ms faster → +0.7% Revenue Per Session

Walmart

100 ms faster → +1% Revenue

Page Speed

=

Money

Greg Linden. Make Data Useful. Stanford Data Mining Class CS345A, 2006

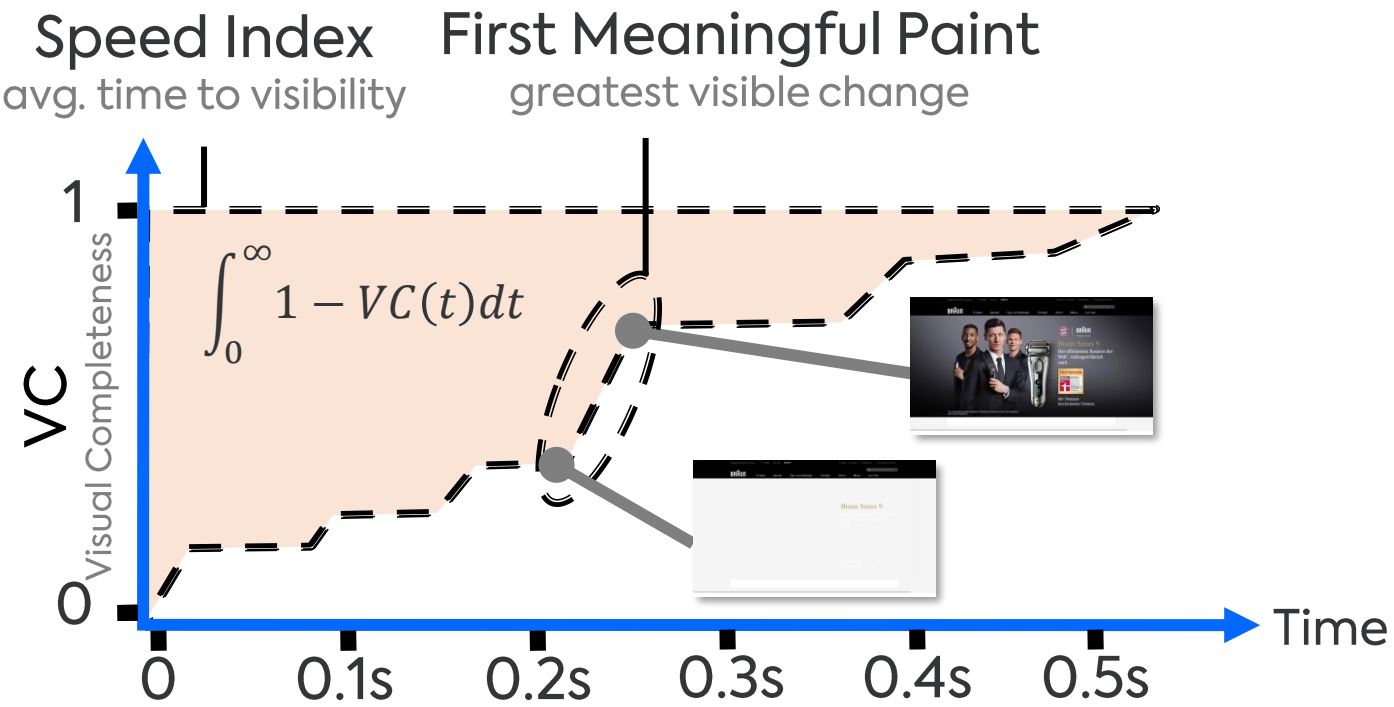
Shuhei Kagawa, Jeff Cybulski, David Martin Jones, et. al.. Loading Time Matters. Zalando Tech Blog, 2018

C. Crocker, A. Kulick, B. Ram. Real-User Monitoring at Walmart. SF & SV Web Performance Group, 2012.

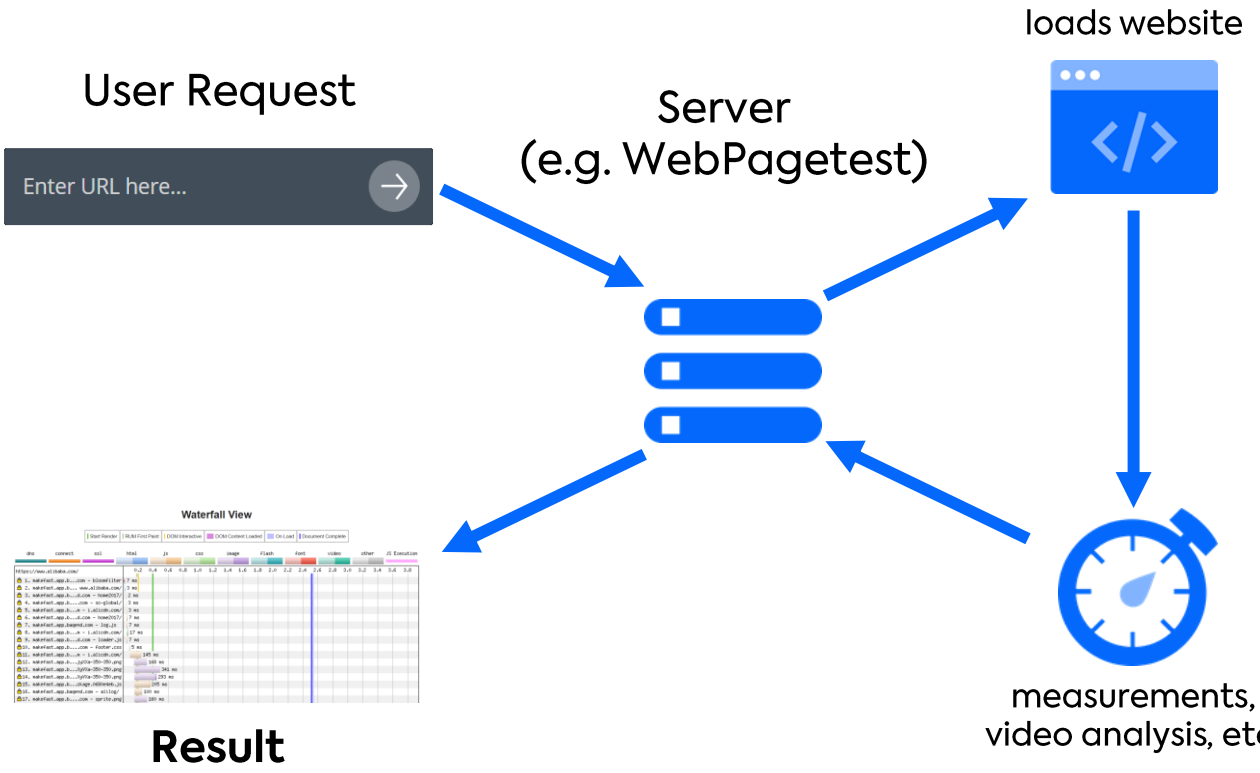
Speed Kit

How to Measure Web Performance

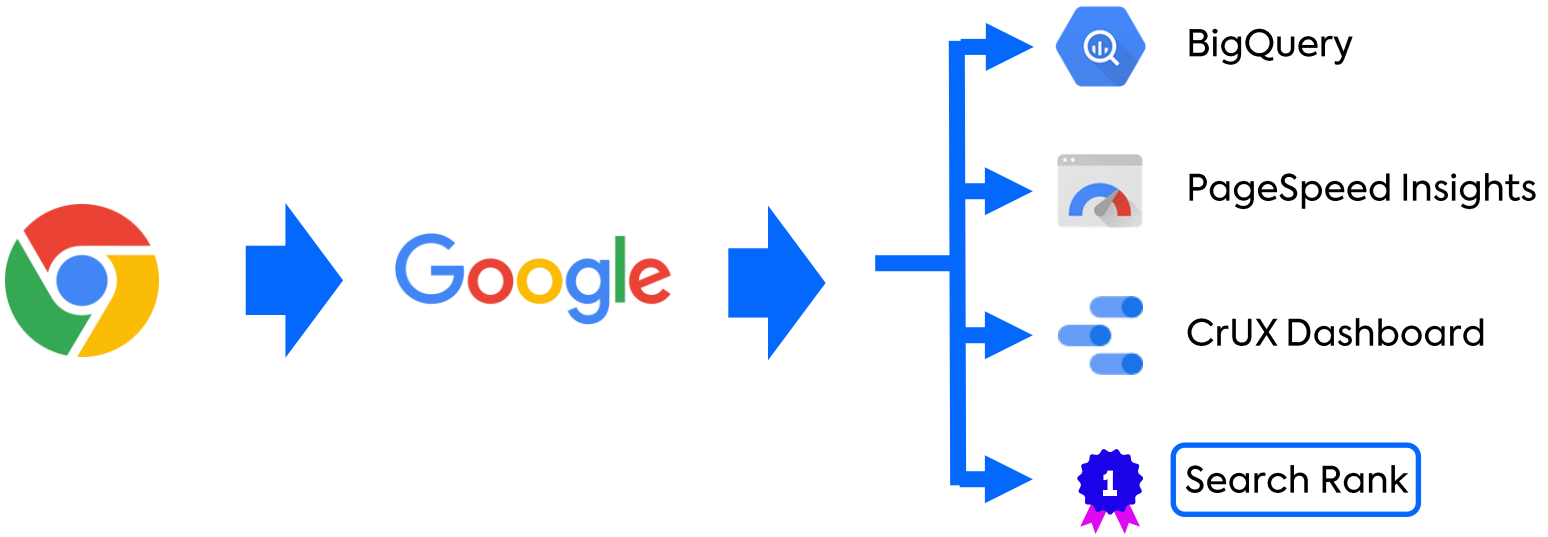
What Do Users Perceive as Fast?



Synthetic Performance Testing

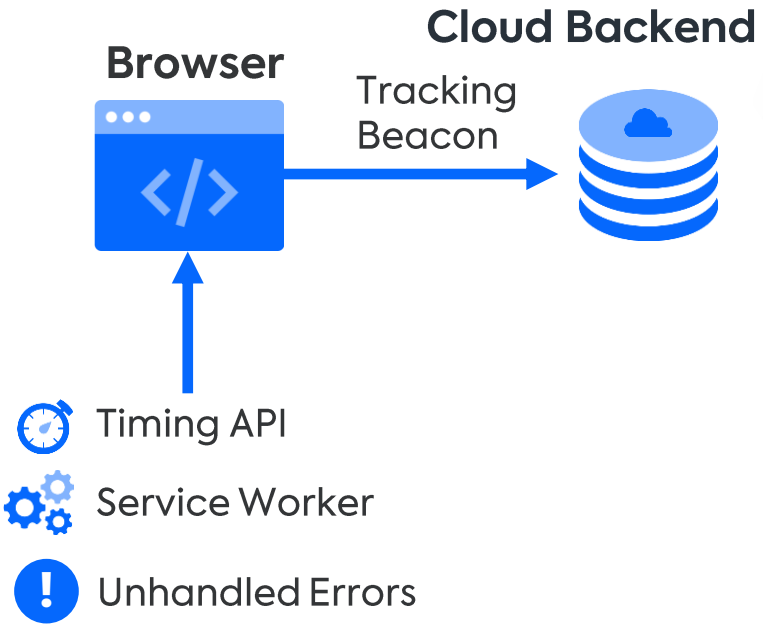


CrUX Data Analysis



- Performance data from Chrome desktop & mobile users
- Domain granularity
- Publicly available
- Performance distributions (histograms)
- SEO-critical

The Basic Idea



- Time-to-First-Byte
- First (Contentful) Paint
- DOM Timer
- First Input Delay

Performance

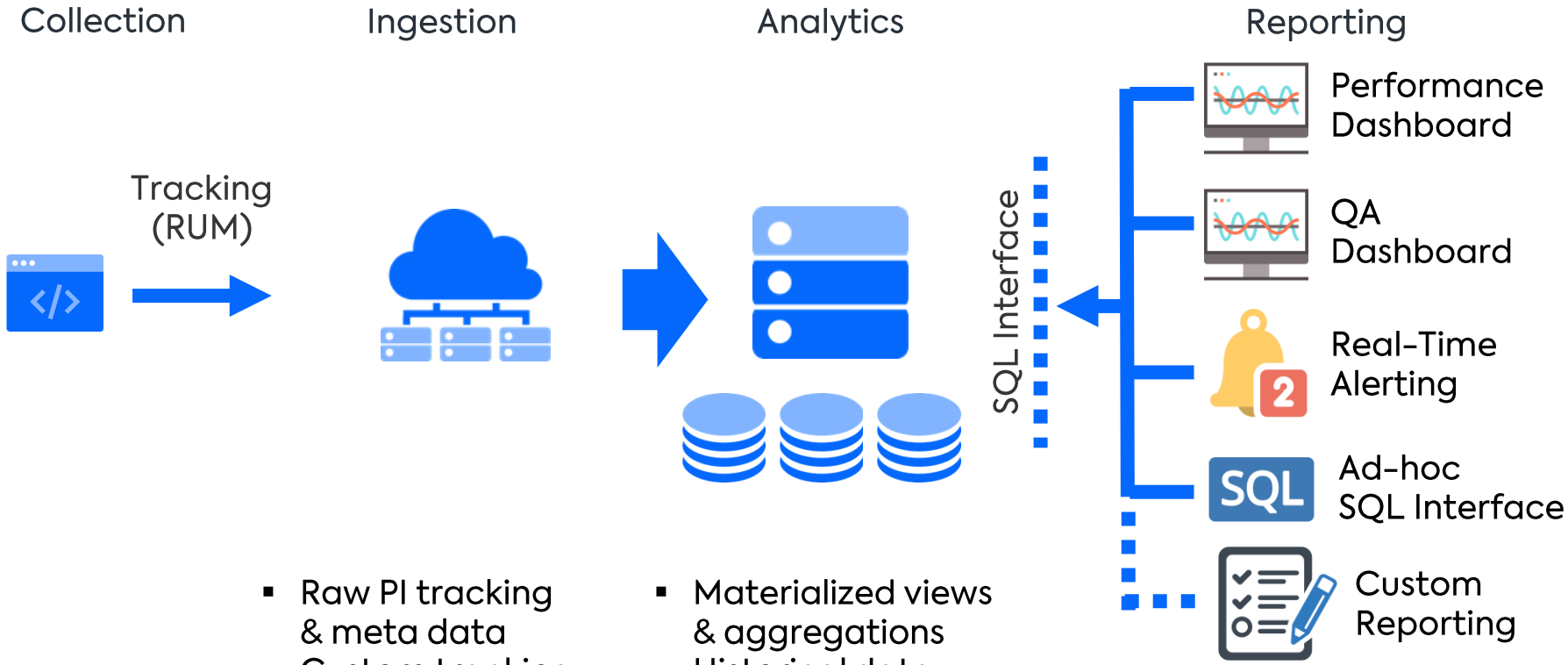
- Session Length
- Time on Site
- First User Interaction
- Bounce Rate

User Engagement

- Page Views & Sessions
- Browser Distribution
- JavaScript Errors
- Caching Insights

QA Metadata

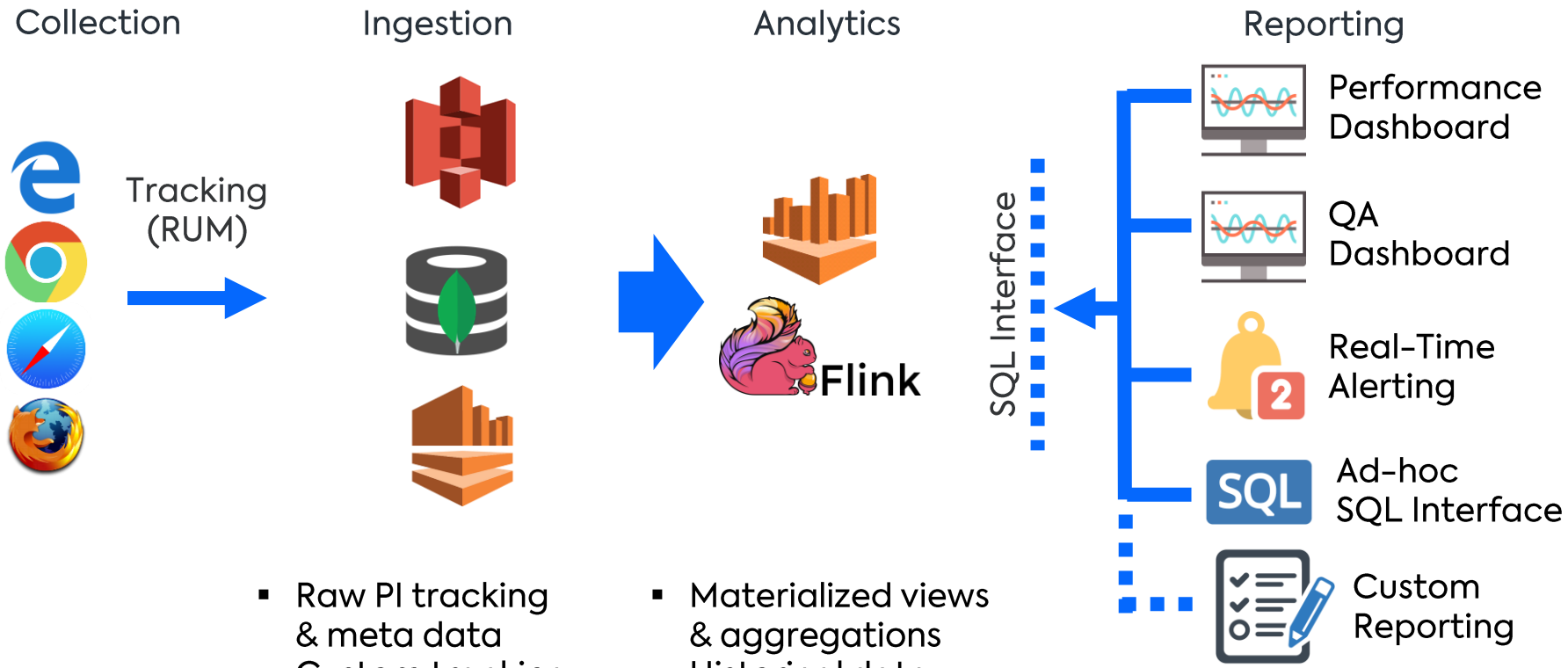
Industry Example



- Raw PI tracking & meta data
- Custom tracking

- Materialized views & aggregations
- Historical data

Industry Example



- Raw PI tracking & meta data
- Custom tracking

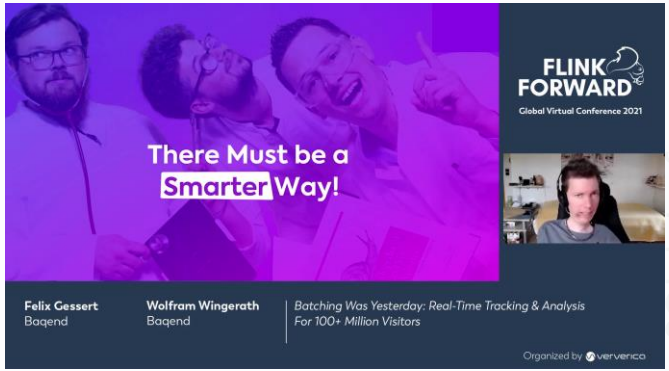
- Materialized views & aggregations
- Historical data

Industry Example

Collection

Tracking (RUM)

Ingestion



F. Gessert, W. Wingerath. Batching Was Yesterday: Real-Time Tracking & Analysis For 100+ Million Visitors, Flink Forward (2021)

- Raw PI tracking & meta data
- Custom tracking

Analytics

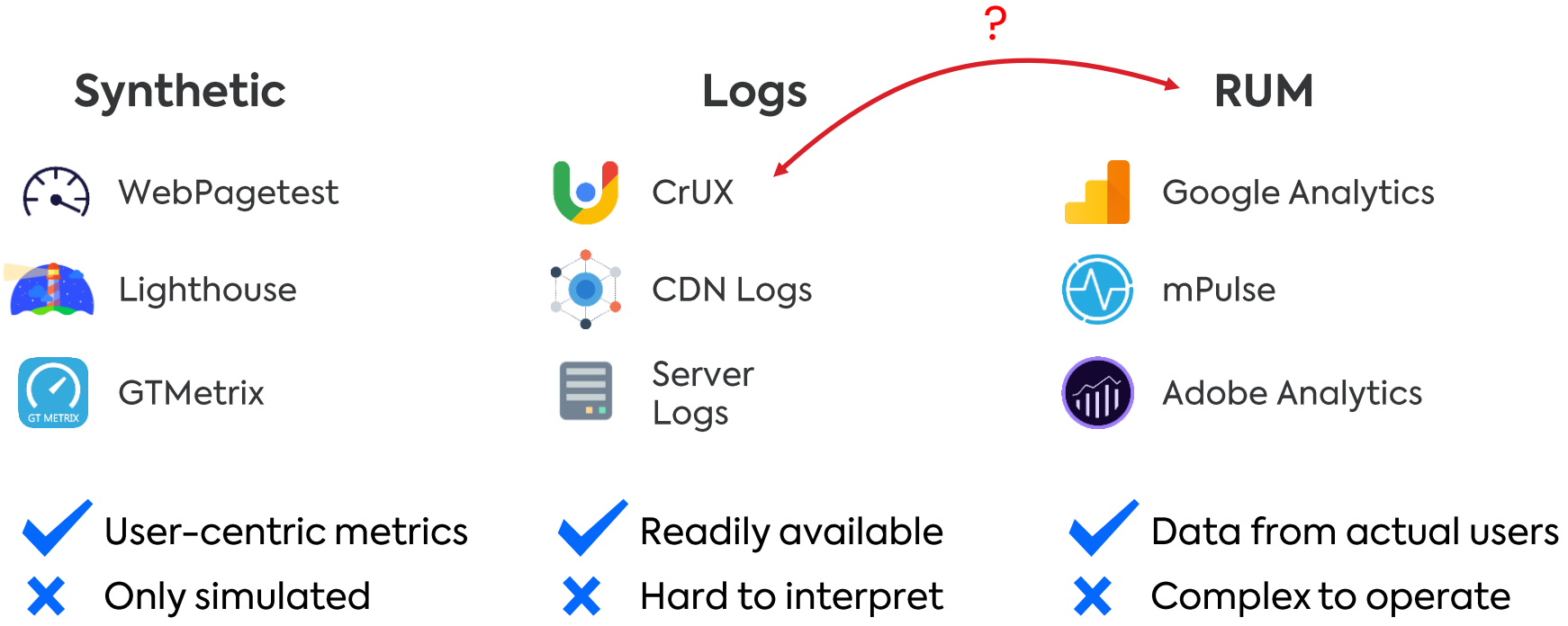
- Materialized views & aggregations
- Historical data

SQL Interface

Reporting

- Performance Dashboard
- QA Dashboard
- Real-Time Alerting
- Ad-hoc SQL Interface
- Custom Reporting

Summary: Measurement Approaches



Speed Kit

Analyzing Performance Data

Split Testing for Web Performance

Speed Kit Users vs. Normal Users

- Speed Kit enabled

- **Measurable uplift**:
  - + Performance
  - + User engagement
  - + ...

- Speed Kit disabled (no acceleration)

W. Wingerath, B. Wollmer, M. Bestehorn, S. Succo, F. Bücklers, J. Domnik, F. Panse, E. Witt, A. Sener, F. Gessert, N. Ritter. Beaconnect: Continuous Web Performance A/B-Testing at Scale. VLDB 2022

Split Testing for Web Performance

Speed Kit Users

- Speed Kit enabled

Tracking →

vs.

- **Measurable uplift**:
  - \+ Performance
  - \+ User engagement
  - \+ ...

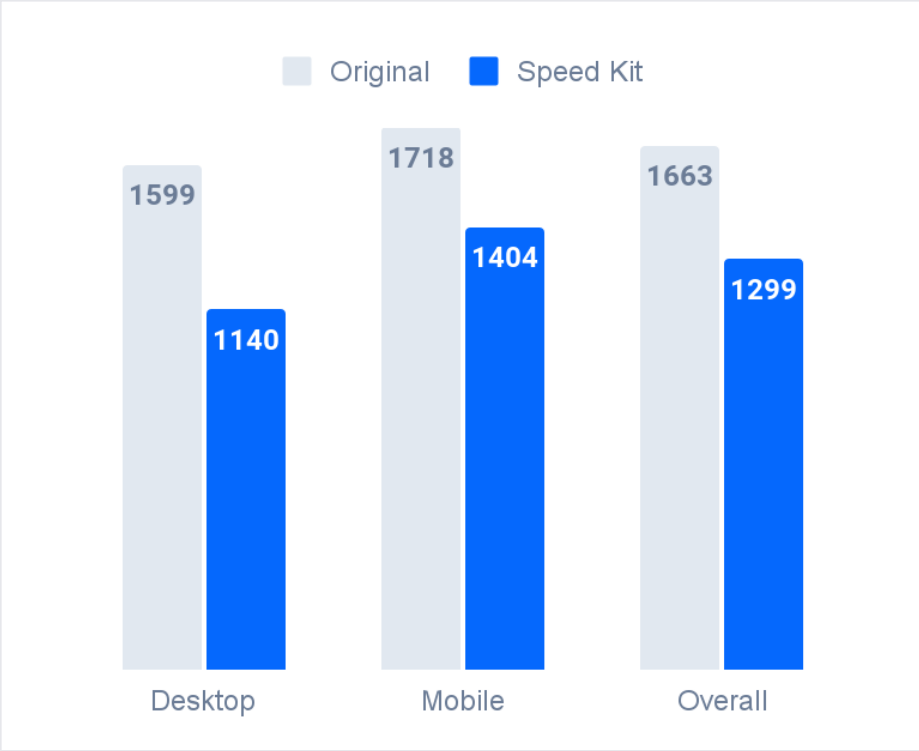
← Tracking

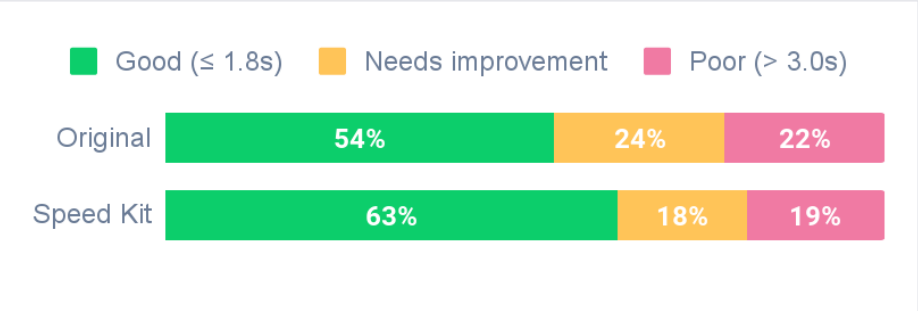
Normal Users

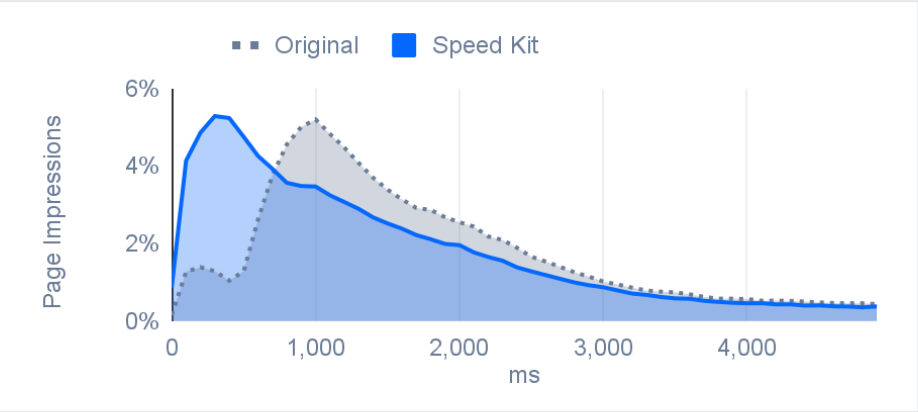
- Speed Kit disabled (no acceleration)

W. Wingerath, B. Wollmer, M. Bestehorn, S. Succo, F. Bücklers, J. Domnik, F. Panse, E. Witt, A. Sener, F. Gessert, N. Ritter. Beaconnect: Continuous Web Performance A/B-Testing at Scale. VLDB 2022

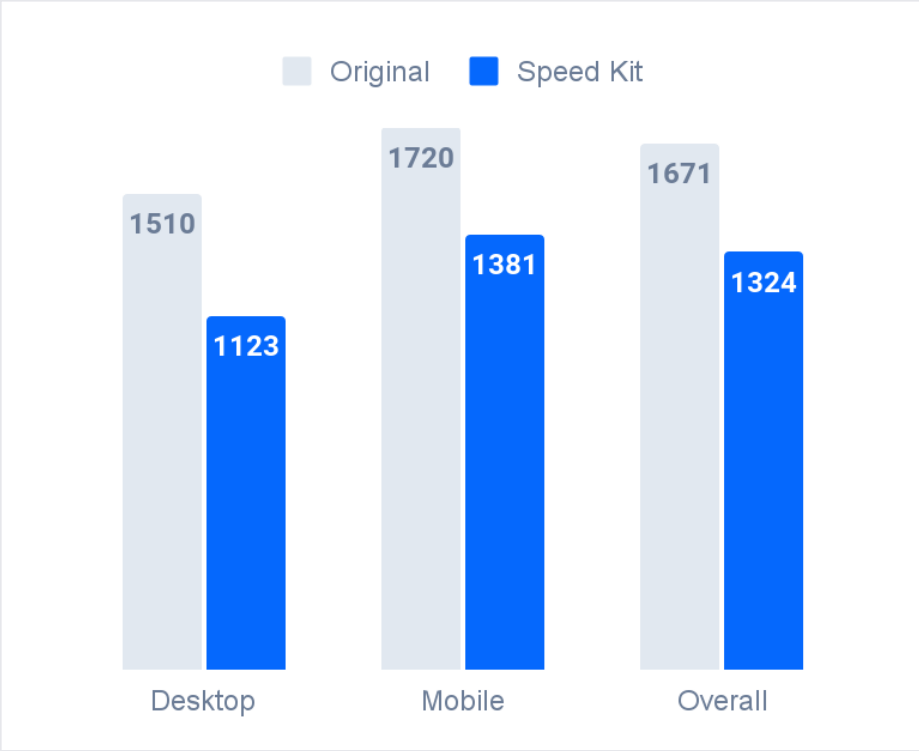
Applying Dimension Filters: All Users

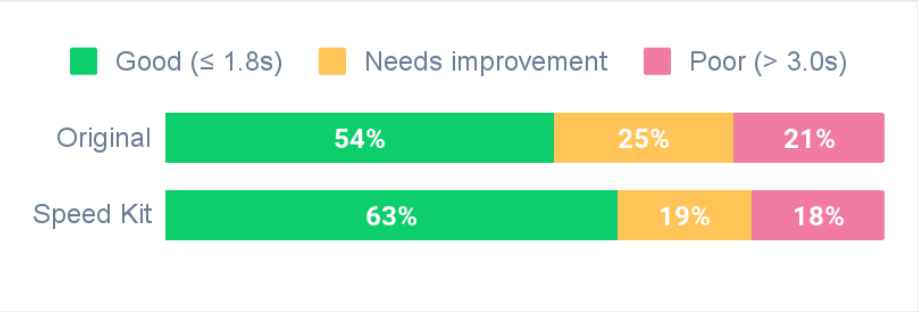


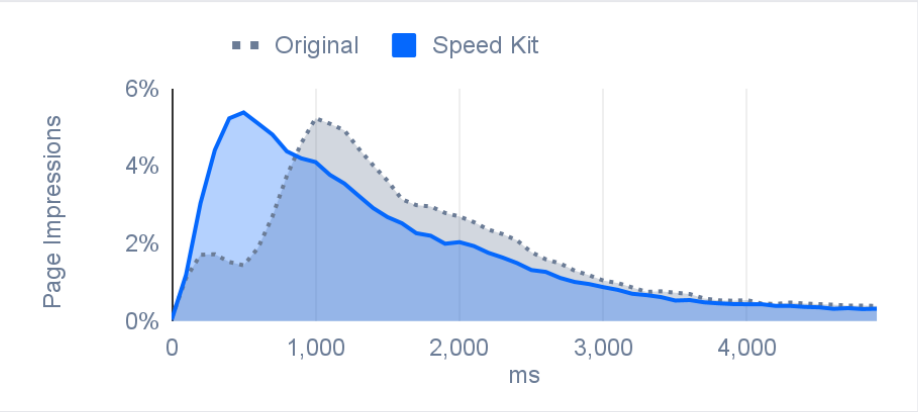




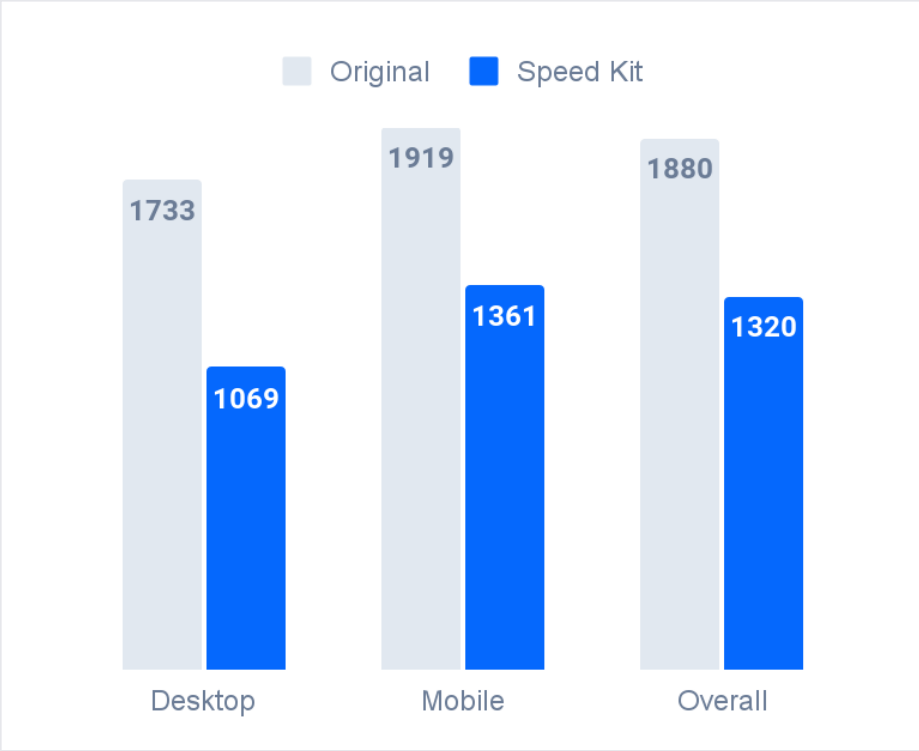
Applying Dimension Filters: Chrome

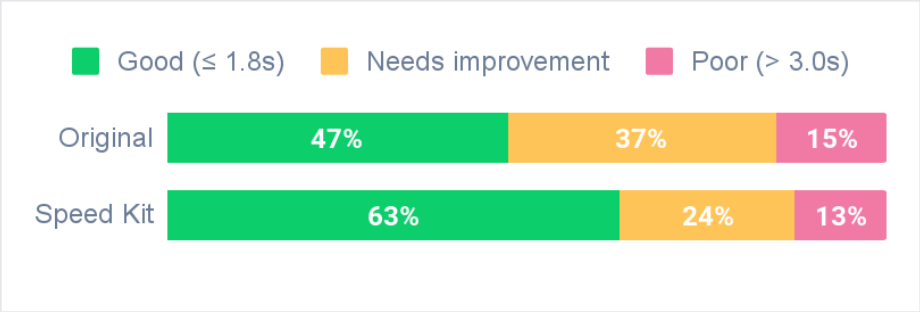


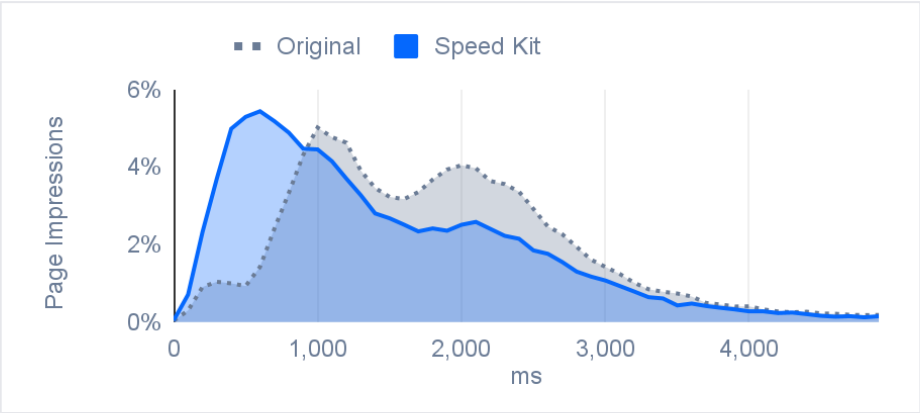




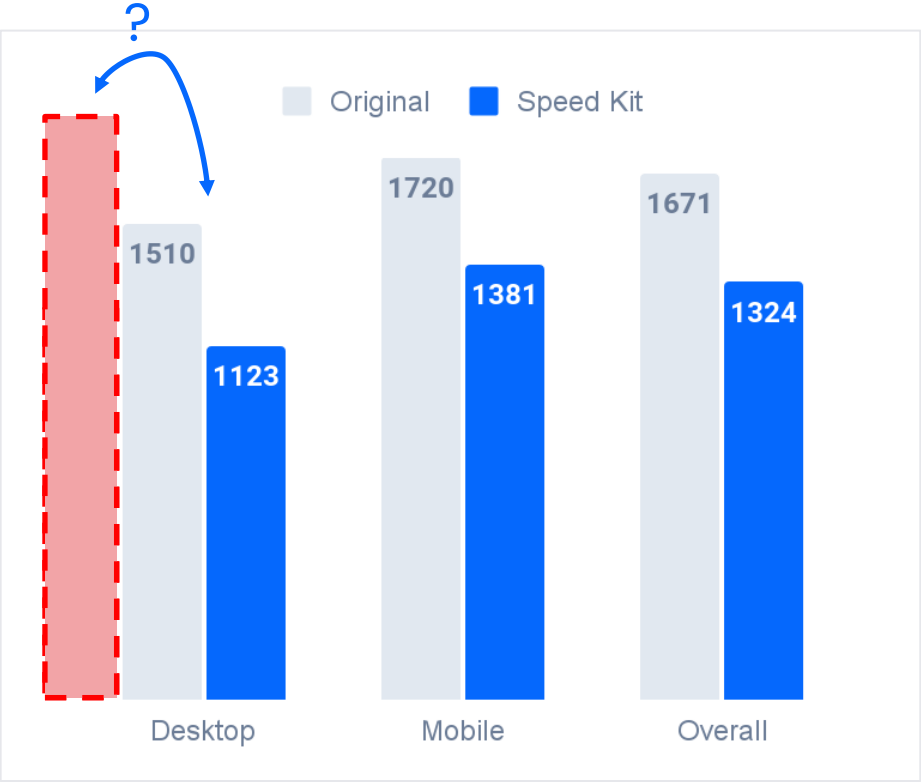
Applying Dimension Filters: Chrome, Product Pages

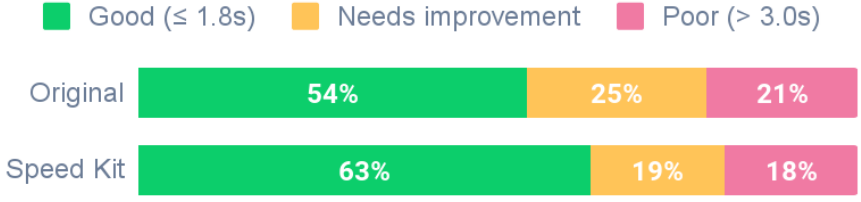


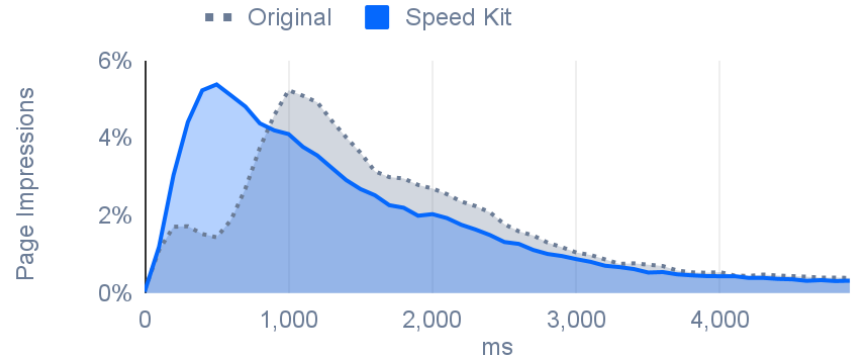




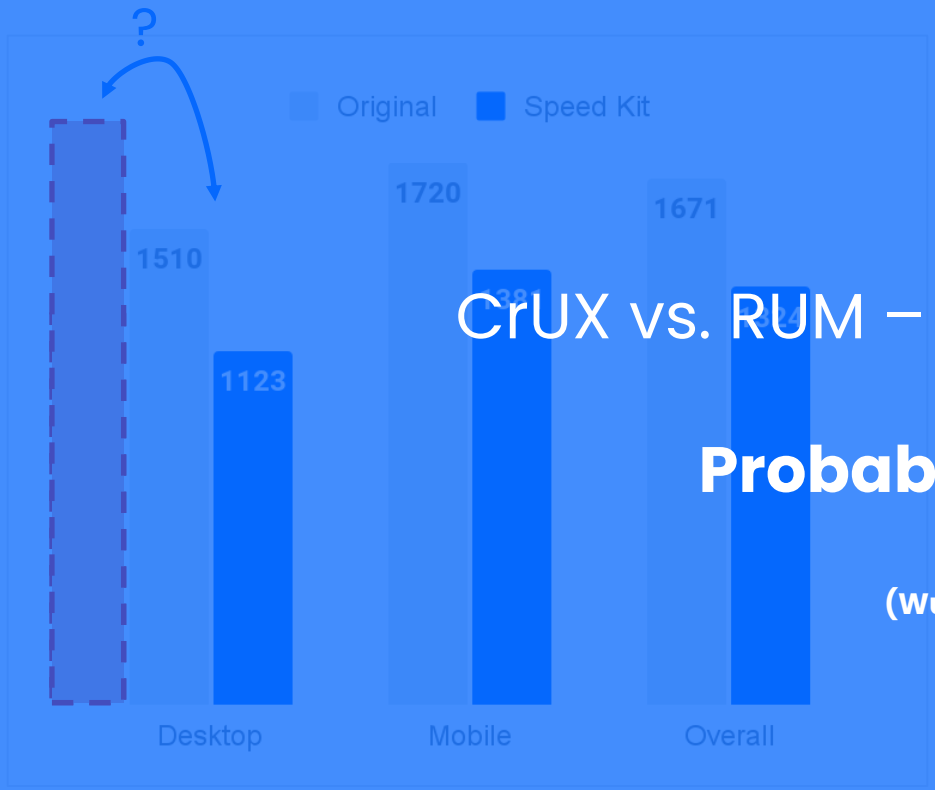
Applying Dimension Filters: Chrome

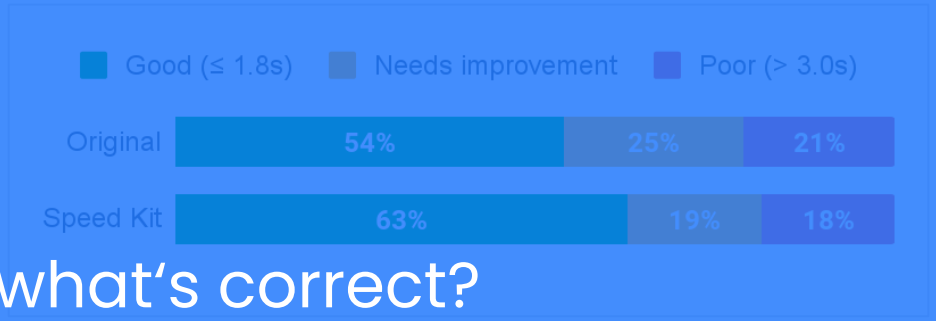






Applying Dimension Filters: Chrome





CrUX vs. RUM – what's correct?

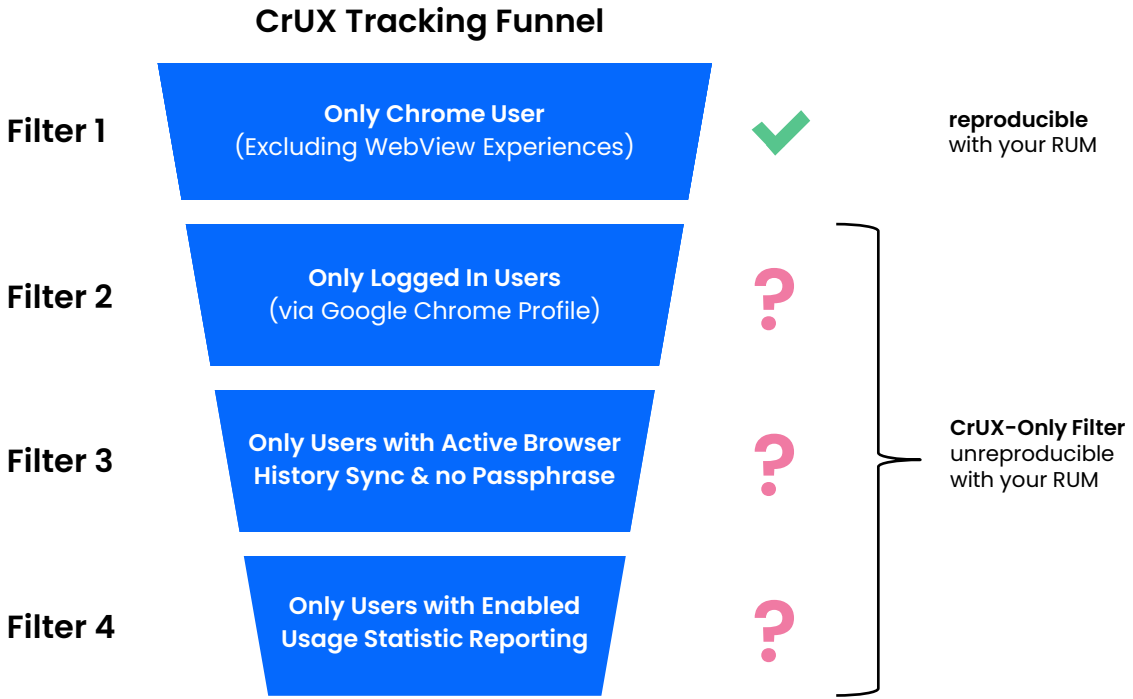
Probably both!

(Wut!?)

Speed Kit

CrUX vs. RUM

Why Google's CrUX Data is a Black Box



Sources: https://developers.google.com/web/tools/chrome-user-experience-report, https://groups.google.com/a/chromium.org/g/chrome-ux-report/c/I3ERRf7Mqio (Access May 11, 2022)

S. Ferrlein, W. Wingerath, B. Wollmer. Why Google's CrUX Results Are Not Reproducible With Your Real-User Monitoring. Baqend Tech Blog, 2022

Speed Kit

Wrapup

Wrapup

CrUX vs. RUM

CrUX

Based on field data

Public data
=> zero-effort
=> competitor data available

Fixed granularity
=> only by month / last 28 days
=> only fixed dimensions
=> only (part of) Chrome users

RUM

Based on field data

Custom deployment
=> complex to operate
=> Just your own website

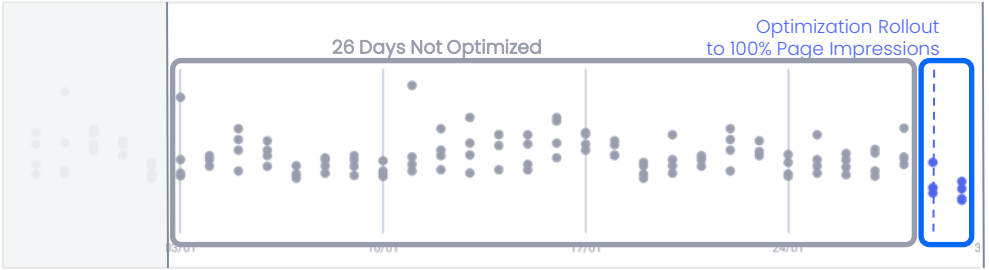
Complete freedom
=> real-time / full detail
=> custom dimensions
=> all browsers

Neither gives you the full picture!

Speed Kit

CrUX + RUM!

CrUX 28 Days Rolling Window: After 2 Days (7%)



Still looks bad!

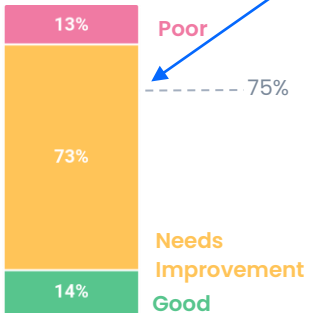
Looks good already!

After 2 Days of Optimization

Two days after the 100% Rollout of an optimization, the Google CrUX 28 Days Report still includes **26 days (93%)** of not optimized performance.

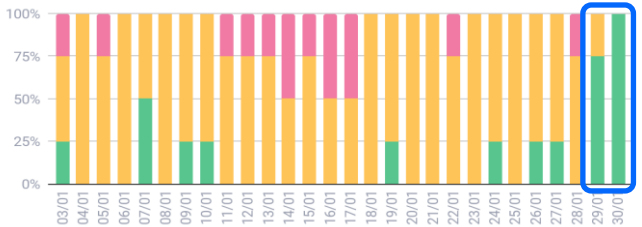
The **2 Days (7%)** of optimized performance can only impact the overall result slightly.

What Google Reports That Day



What Google Aggregates In Its Report That Day

(But doesn't report in that detail)



93% Not Optimized
(26 of 28 Days)

7% Optimized
(2 of 28 Days)

CrUX 28 Days Rolling Window: After 7 Days (25%)

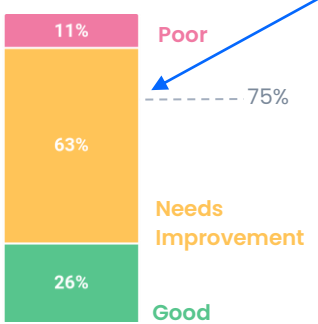
Looks better!

After 7 Days of Optimization

One week after the 100% Rollout of an optimization, the Google CrUX 28 Days Report still includes **21 days (75%)** of not optimized performance.

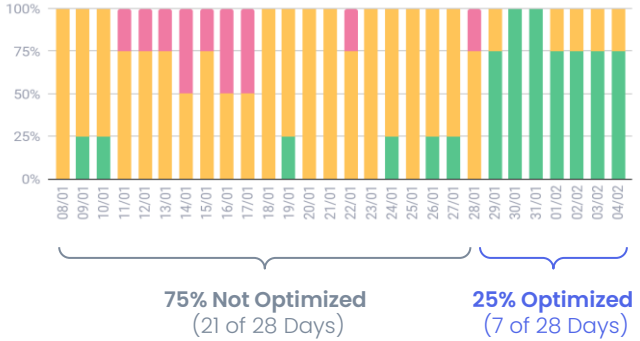
The **7 Days (25%)** of optimized performance are still not dominant in the overall result.

What Google Reports That Day

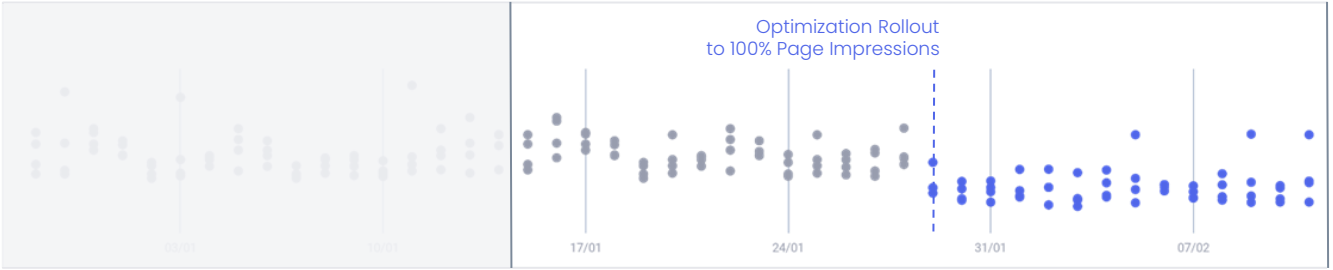


What Google Aggregates In Its Report That Day

(But doesn't report in that detail)



CrUX 28 Days Rolling Window: After 14 Days (50%)

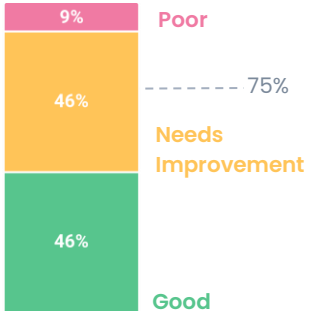


After 14 Days of Optimization

Two weeks after the 100% Rollout of an optimization, the Google CrUX 28 Days Report contains **14 days (50%)** of not optimized and **14 Days (50%)** of optimized performance.

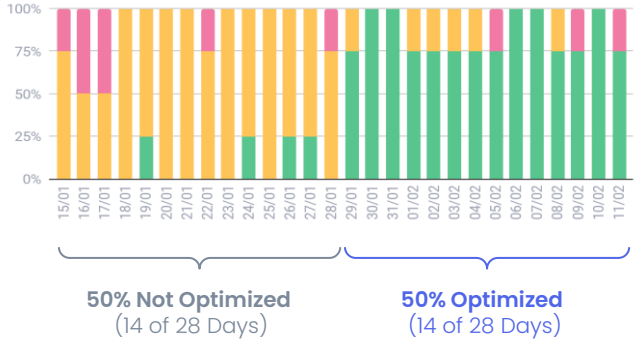
From now on the impact of the optimization will be at least dominant in the overall result.

What Google Reports That Day

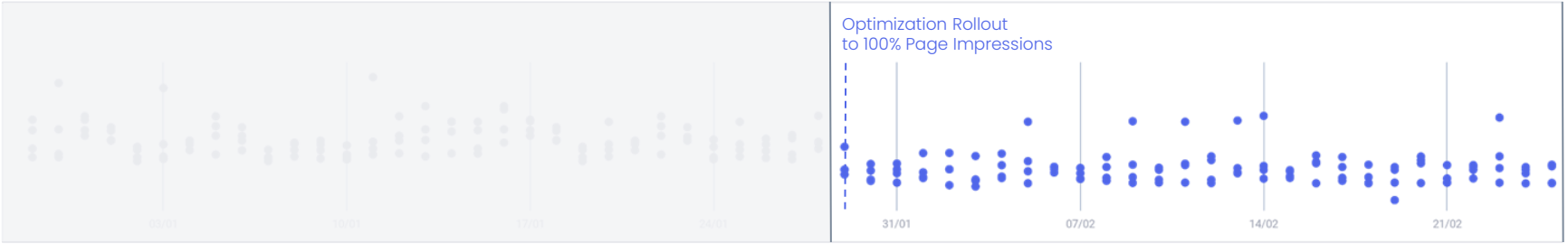


What Google Aggregates In Its Report That Day

(But doesn't report in that detail)



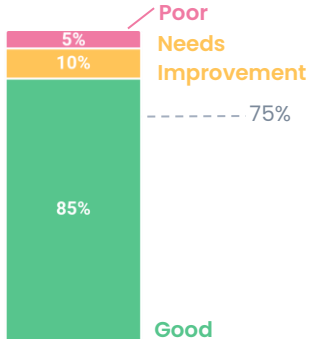
CrUX 28 Days Rolling Window: After 28 Days (100%)



After 28 Days of Optimization

Only when the page optimization is live for **28 days (100%)** the effect will be fully reflected in the Google CrUX 28 Days Report.

What Google Reports That Day



What Google Aggregates In Its Report That Day

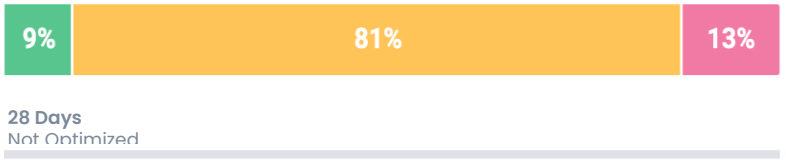
(But doesn't report in that detail)

100% Optimized
(28 of 28 Days)

CrUX 28 Days Rolling Window: FCP CrUX Uplift Example

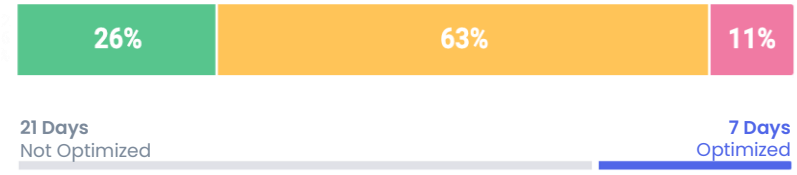
Before Optimization

Before the rollout of the optimization, the First Contentful Paint (FCP) of the last 28 Days is obtained from the CrUX API.



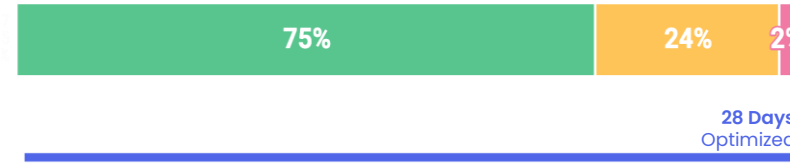
7 Days after Optimization

After 7 Days of optimizations, the 28 Days FCP reported by the CrUX API is already **202 ms faster**.



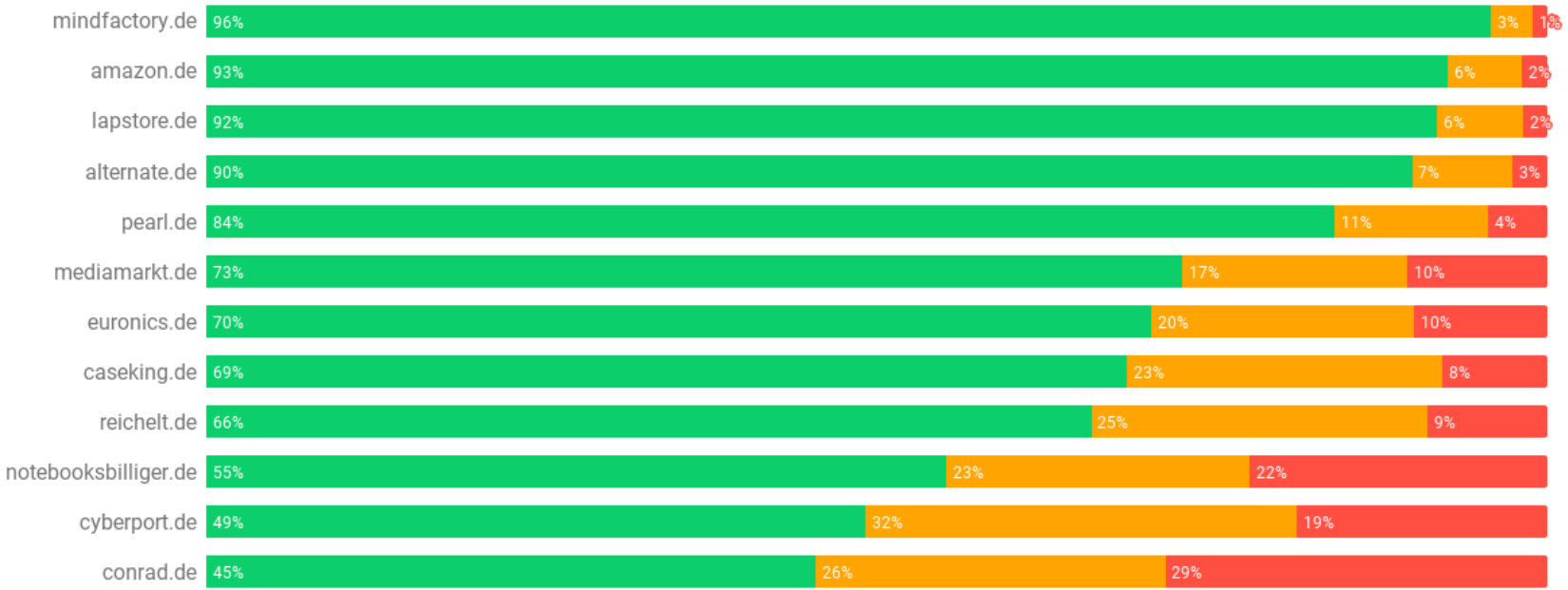
28 Days after Optimization

Extrapolated to 28 Days of Optimization (7 Days Uplift * 4) the FCP is expected to become **806 ms faster** compared to the FCP before the optimization.



CrUX + RUM!

Rank Your Score & Forecast With Competitors!



Speed Kit

Roadmap Features

Dashboard KPI Overview

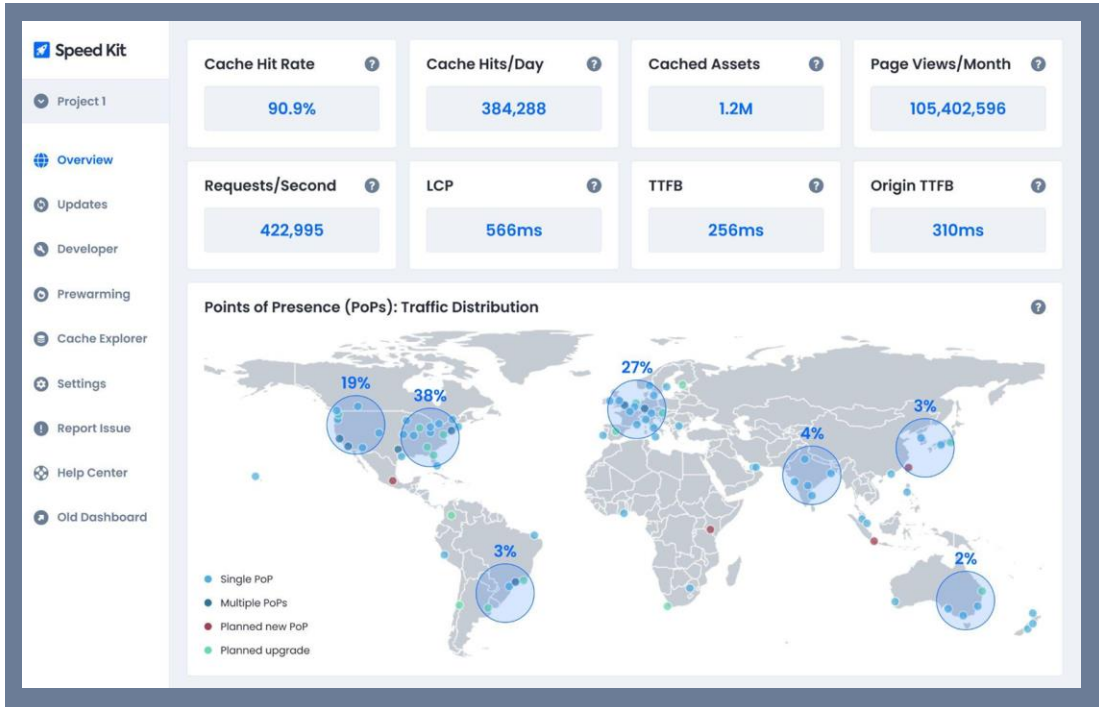
Purpose

Simple overview of the technical Speed Kit KPIs

Approach

The dashboard will provide a view to display the technical KPIs:

- See your current traffic usage and how many cache hits you have
- See how the performance of your site behaves at a single glance



Roadmap Features

Speed Analytics Dashboard

Purpose

A simple overview of the main performance metrics

Approach

An easy-to-use performance dashboard based on monthly RUM data:

- Status quo and over time development of core web vitals
- Drill-down by origin, device and page types
- Can be extended by SQL workbench for drill-downs

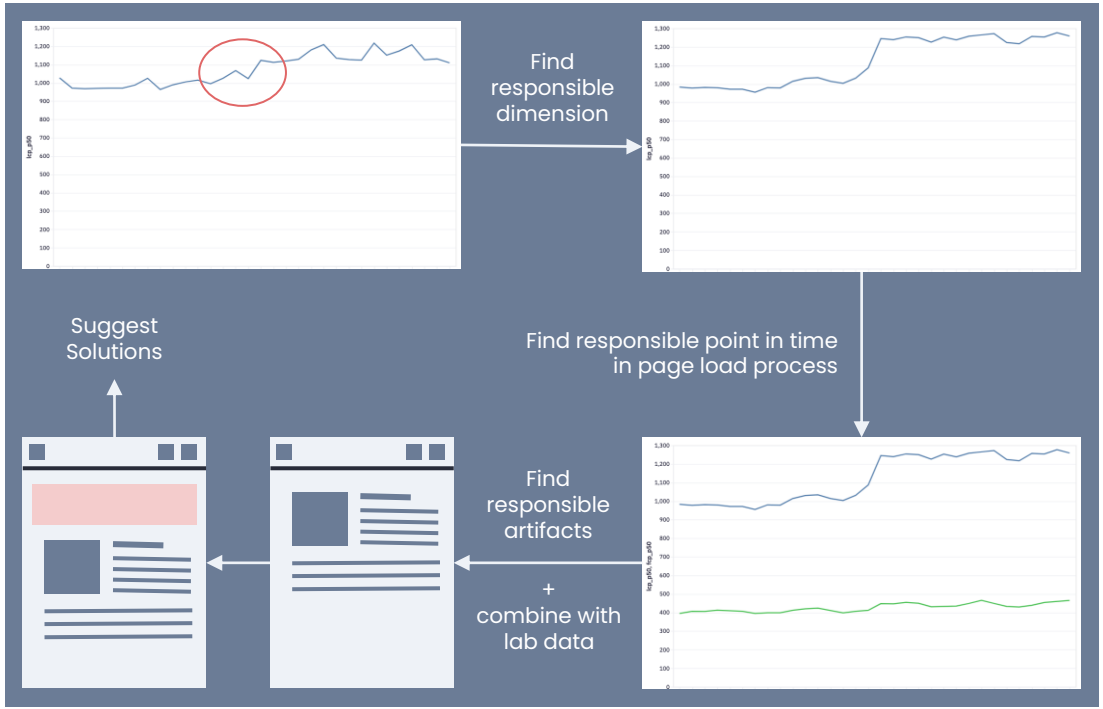
Speed Guard

Purpose

Detect performance issues with exact cause and solution

Approach

1. Monitor performance and detect anomalies
2. Find cause by automated dicing and combining RUM and lab data
3. Suggest solution based on caused and web performance best practices



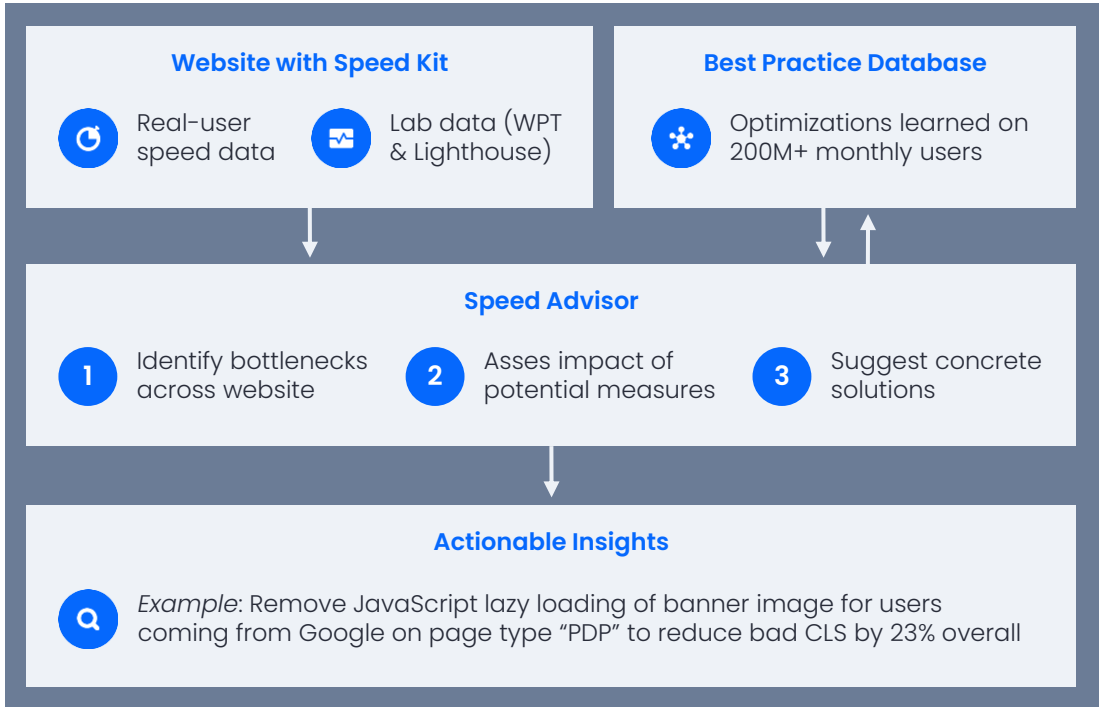
Speed Advisor

Purpose

Suggested speed improvements beyond Speed Kit

Approach

Speed Kit's real-user monitoring and synthetic testing collect detailed speed data. The advisor analyzes various dimensions (e.g. browsers) and metrics (e.g. CLS) and suggest applicable optimizations.



Wolfram Wingerath, Felix Gessert, et al.: Speed Kit: A Polyglot & GDPR-Compliant Approach For Caching Personalized Content, ICDE (2020).

Thanks!

Questions?

wolle@uol.de

uol.de/wolle