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
Let’s Agree to Disagree: Why Google’s CrUX Results Are Not Reproducible With Your Real-User Monitoring
I Am Wolle

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

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

University of Hamburg
Carl von Ossietzky University of Oldenburg
Bagend
We bring performance research to practice

**WHO WE ARE**

- **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

1. Amazon: 100 ms slower → -1% Conversion Rate
2. Zalando: 100 ms faster → +0.7% Revenue Per Session
3. Walmart: 100 ms faster → +1% Revenue

References:
Page Speed

\[ \text{100 ms faster} \quad \rightarrow \quad +0.7\% \text{ Revenue Per Session} \]

\[ \text{100 ms faster} \quad \rightarrow \quad +1\% \text{ Revenue} \]

Why Do Businesses Care About Performance?

You Heard the Stories
How to Measure Web Performance
What Do Users Perceive as Fast?

**Speed Index**

*avg. time to visibility*

**First Meaningful Paint**

*greatest visible change*

\[ \int_0^\infty 1 - VC(t) \, dt \]

Let's Agree to Disagree: Why Google’s CrUX Results Are Not Reproducible With Your Real-User Monitoring
Synthetic Performance Testing

User Request

Enter URL here...

Server (e.g. WebPagetest)

loads website

Result

measurements, video analysis, etc.

Let's Agree to Disagree: Why Google’s CrUX Results Are Not Reproducible With Your Real–User Monitoring
Measuring Web Performance

CrUX Data Analysis

Google

- Performance data from Chrome desktop & mobile users
- Domain granularity

- Publicly available
- Performance distributions (histograms)
- SEO-critical

BigQuery
PageSpeed Insights
CrUX Dashboard
Search Rank

Let’s Agree to Disagree: Why Google’s CrUX Results Are Not Reproducible With Your Real-User Monitoring
The Basic Idea

Real-User Monitoring (RUM)

Let's Agree to Disagree: Why Google’s CrUX Results Are Not Reproducible With Your Real-User Monitoring

- Time-to-First-Byte
- First (Contentful) Paint
- DOM Timer
- First Input Delay
- Session Length
- Time on Site
- First User Interaction
- Bounce Rate
- Page Views & Sessions
- Browser Distribution
- JavaScript Errors
- Caching Insights

Performance
User Engagement
QA Metadata
Let's Agree to Disagree: Why Google’s CrUX Results Are Not Reproducible With Your Real-User Monitoring
Let's Agree to Disagree: Why Google's CrUX Results Are Not Reproducible With Your Real-User Monitoring
Let’s Agree to Disagree: Why Google’s CrUX Results Are Not Reproducible With Your Real-User Monitoring

Industry Example

Collection

Ingestion

Analytics

Reporting

- Raw PI tracking & meta data
- Custom tracking
- Materialized views & aggregations
- Historical data

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

Let's Agree to Disagree: Why Google’s CrUX Results Are Not Reproducible With Your Real-User Monitoring
Analyzing Performance Data
Let’s Agree to Disagree: Why Google’s CrUX Results Are Not Reproducible With Your Real-User Monitoring

Split Testing for Web Performance

**Speed Kit Users** vs. **Normal Users**

- **Speed Kit enabled**
- **Measurable uplift:**
  - Performance
  - User engagement
  - ...
- **Speed Kit disabled**
  - (no acceleration)

---

Let's Agree to Disagree: Why Google’s CrUX Results Are Not Reproducible With Your Real-User Monitoring

Analyzing Performance Data

Applying Dimension Filters: All Users

Let's Agree to Disagree: Why Google’s CrUX Results Are Not Reproducible With Your Real-User Monitoring
Applying Dimension Filters: Chrome

Analyzing Performance Data

Let's Agree to Disagree: Why Google’s CrUX Results Are Not Reproducible With Your Real-User Monitoring
Applying Dimension Filters: Chrome, Product Pages

Let's Agree to Disagree: Why Google’s CrUX Results Are Not Reproducible With Your Real-User Monitoring
Analyzing Performance Data

Applying Dimension Filters: Chrome

Let's Agree to Disagree: Why Google's CrUX Results Are Not Reproducible With Your Real-User Monitoring
Applying Dimension Filters: Chrome

Let's Agree to Disagree: Why Google's CrUX Results Are Not Reproducible With Your Real-User Monitoring

CrUX vs. RUM — what’s correct?

Probably both!

(Wut!?)
CrUX vs. RUM
Why Google’s CrUX Data is a Black Box

CrUX Tracking Funnel

Filter 1
Only Chrome User
(Excluding WebView Experiences)
reproducible with your RUM

Filter 2
Only Logged In Users
(via Google Chrome Profile)

Filter 3
Only Users with Active Browser History Sync & no Passphrase
unreproducible with your RUM

Filter 4
Only Users with Enabled Usage Statistic Reporting

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)

Let’s Agree to Disagree: Why Google’s CrUX Results Are Not Reproducible With Your Real-User Monitoring

<table>
<thead>
<tr>
<th>CrUX</th>
<th>vs.</th>
<th>RUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Based on field data</td>
<td></td>
<td>Based on field data</td>
</tr>
<tr>
<td>Public data</td>
<td></td>
<td>Custom deployment</td>
</tr>
<tr>
<td>=&gt; zero-effort</td>
<td></td>
<td>=&gt; complex to operate</td>
</tr>
<tr>
<td>=&gt; competitor data available</td>
<td></td>
<td>=&gt; Just your own website</td>
</tr>
<tr>
<td>Fixed granularity</td>
<td></td>
<td>Complete freedom</td>
</tr>
<tr>
<td>=&gt; only by month / last 28 days</td>
<td></td>
<td>=&gt; real-time / full detail</td>
</tr>
<tr>
<td>=&gt; only fixed dimensions</td>
<td></td>
<td>=&gt; custom dimensions</td>
</tr>
<tr>
<td>=&gt; only (part of) Chrome users</td>
<td></td>
<td>=&gt; all browsers</td>
</tr>
</tbody>
</table>

Neither gives you the full picture!
CrUX + RUM!
Let’s Agree to Disagree: Why Google’s CrUX Results Are Not Reproducible With Your Real-User Monitoring

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

**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**

- 13% Poor
- 73% Needs Improvement
- 14% Good

**What Google Aggregates In Its Report That Day**

- 93% Not Optimized (26 of 28 Days)
- 7% Optimized (2 of 28 Days)

Looks good already!

Still looks bad!
CrUX 28 Days Rolling Window: After 7 Days (25%)

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

- 11% Poor
- 63% Needs Improvement
- 26% Good

What Google Aggregates In Its Report That Day

- 75% Not Optimized (21 of 28 Days)
- 25% Optimized (7 of 28 Days)

Looks better!
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

- 9% Poor
- 46% Needs Improvement
- 46% Good

What Google Aggregates In Its Report That Day (But doesn’t report in that detail)

- 50% Not Optimized (14 of 28 Days)
- 50% Optimized (14 of 28 Days)
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

- **85%** Good
- **10%** Improvement
- **5%** Poor

What Google Aggregates In Its Report That Day (But doesn't report in that detail)

- **100% Optimized** (28 of 28 Days)
Let's Agree to Disagree: Why Google’s CrUX Results Are Not Reproducible With Your Real-User Monitoring

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.
Rank Your Score & Forecast With Competitors!

<table>
<thead>
<tr>
<th>Website</th>
<th>Mindfactory.de</th>
<th>Amazon.de</th>
<th>Lapstore.de</th>
<th>Alternate.de</th>
<th>Pearl.de</th>
<th>MediaMarkt.de</th>
<th>Euronics.de</th>
<th>Caseking.de</th>
<th>Reichelt.de</th>
<th>NotebooksBilliger.de</th>
<th>Cyberport.de</th>
<th>Conrad.de</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>96%</td>
<td>93%</td>
<td>92%</td>
<td>90%</td>
<td>84%</td>
<td>73%</td>
<td>70%</td>
<td>69%</td>
<td>56%</td>
<td>55%</td>
<td>49%</td>
<td>45%</td>
</tr>
<tr>
<td>Forecast</td>
<td>3%</td>
<td>5%</td>
<td>2%</td>
<td>7%</td>
<td>4%</td>
<td>10%</td>
<td>10%</td>
<td>8%</td>
<td>9%</td>
<td>22%</td>
<td>19%</td>
<td>29%</td>
</tr>
</tbody>
</table>
Roadmap Features
**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
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

---

Let’s Agree to Disagree: Why Google’s CrUX Results Are Not Reproducible With Your Real-User Monitoring
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
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.

Speed Advisor

1. Identify bottlenecks across website
2. Assess impact of potential measures
3. Suggest concrete solutions

Websites with Speed Kit

- Real-user speed data
- Lab data (WPT & Lighthouse)

Best Practice Database

- Optimizations learned on 200M+ monthly users

Actionable Insights

Example: Remove JavaScript lazy loading of banner image for users coming from Google on page type “PDP” to reduce bad CLS by 23% overall

Let’s Agree to Disagree: Why Google’s CrUX Results Are Not Reproducible With Your Real-User Monitoring
Thanks!

Questions?

wolle@uol.de

uol.de/wolle