



“It **Worked** on Staging!”

How to Build, Break, and Fix a Global Real-User Monitoring Data **Platform**

ICRAIS 2026, Hamburg, Germany

Wolfram „Wolle“ Wingerath

October 02, 2026

Slides Available at wolle.science and speedkit.com

I Am **Wolle**



Research:

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



Practice:

- Web Caching •
- Big Data Analytics •
- Anger Management •
- ...



Summing Up 10 Years of R&D: **Outline**

1

Why did we build our own RUM data platform?

Why should you monitor user behavior and when is a standard tool insufficient?

2

What did we learn?

Where did we encounter road blocks and how did we overcome them?

3

Where are we going from here?

How did the use cases evolve and what's in store next?

You Heard the **Stories**



100 ms slower



-1% Conversion Rate



100 ms faster



+0.7% Revenue Per Session



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.

Load Time & **SEO**



From 7s to 2s Loads



+80% Traffic



500 ms Slower Loads



-20% Traffic



40% Faster Loads



+15% SEO Traffic



Lucia Moses. [How GQ Cut Its Webpage Load Time By 80 Percent](#), Digiday, 2015.



Marissa Mayer, [Conference Keynote](#), Web 2.0, 2006



Sam Meder, Vadim Antonov, Jeff Chang. [Driving User Growth With Performance Improvements](#), Pinterest Blog, 2017

Load Time & User Engagement

FORRESTER®

-80% load time → **+60%** Session Length (Mobile)

OTTO

-42% time to FCP → **+25%** Session Length

Akamai

+2s load time → **+103%** Bounce Rate



Forrester. The Total Economic Impact
™ Of Accelerated Mobile Pages, 2017



Lars Bognar. Mobile Speed Race der Otto
Group Verbessert Mobile Ladezeiten, TWG, 2019



Akamai. Akamai Online Retail Performance
Report: Milliseconds Are Critical, Akamai Blog, 2017

Load Time & User Satisfaction



+500 ms network delay → **+26%** peak frustration

Aberdeen *Group* **+1 s** delay in response times → **-16%** customer satisfaction



+50% response time

-50% productivity



Tammy Everts. Mobile Web Stress: The Impact of Network Speed on Emotional Engagement and Brand Perception, Radware, 2013

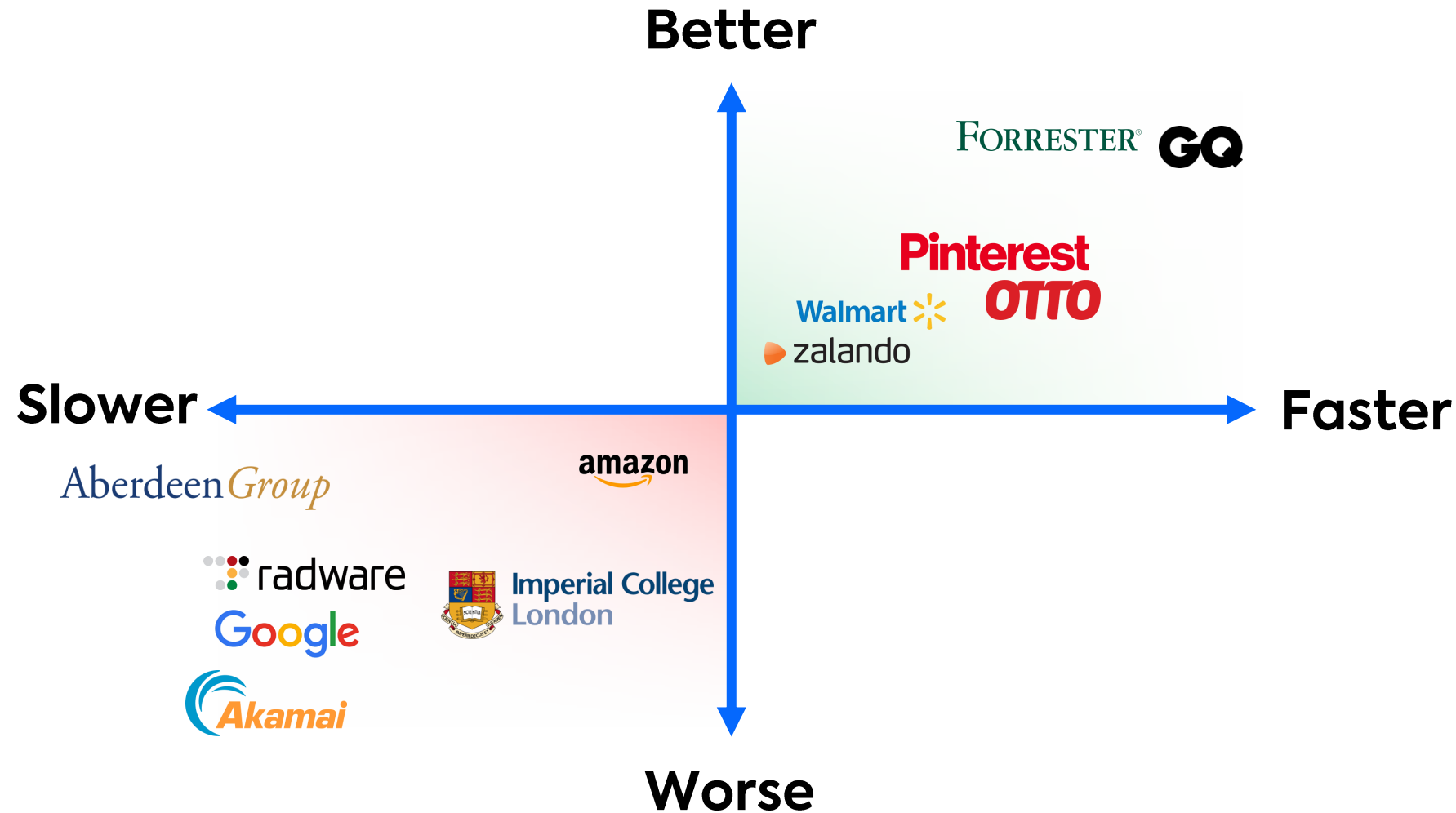


The Performance of Web Applications: Customers Are Won or Lost in One Second, Aberdeen Group, 2008.



T. Goodman, R. Spence. The Effect of System Response Time on Interactive Computer Aided Problem Solving, ICL, 1978

Business Impact : **Why** to Make a Website Fast



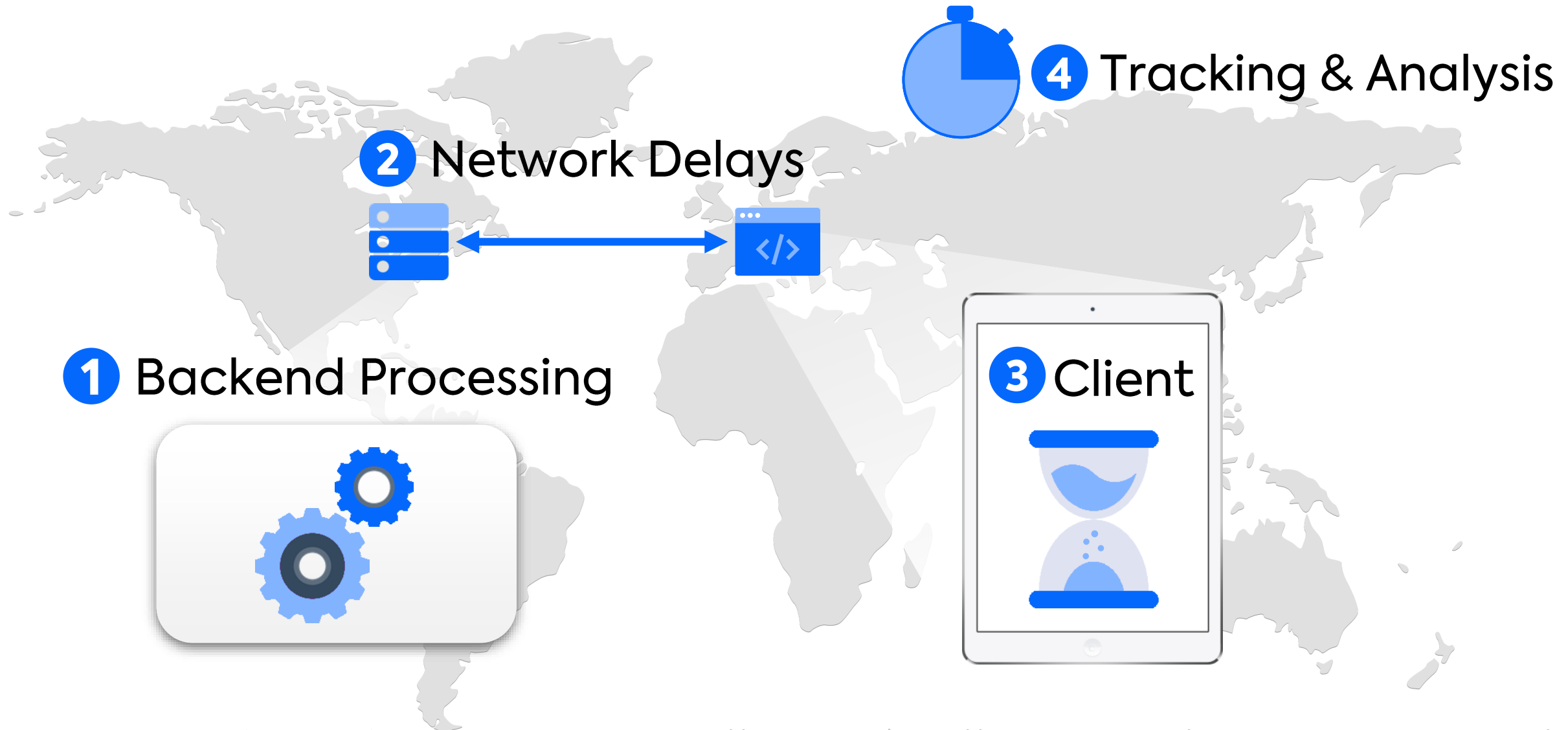
W. Wingerath, B. Wollmer, F. Gessert, S. Succo, N. Ritter. Going for Speed: Full-Stack Performance Engineering in Modern Web-Based Applications, WWW 2021

Business Impact : Why to Make a Website Fast



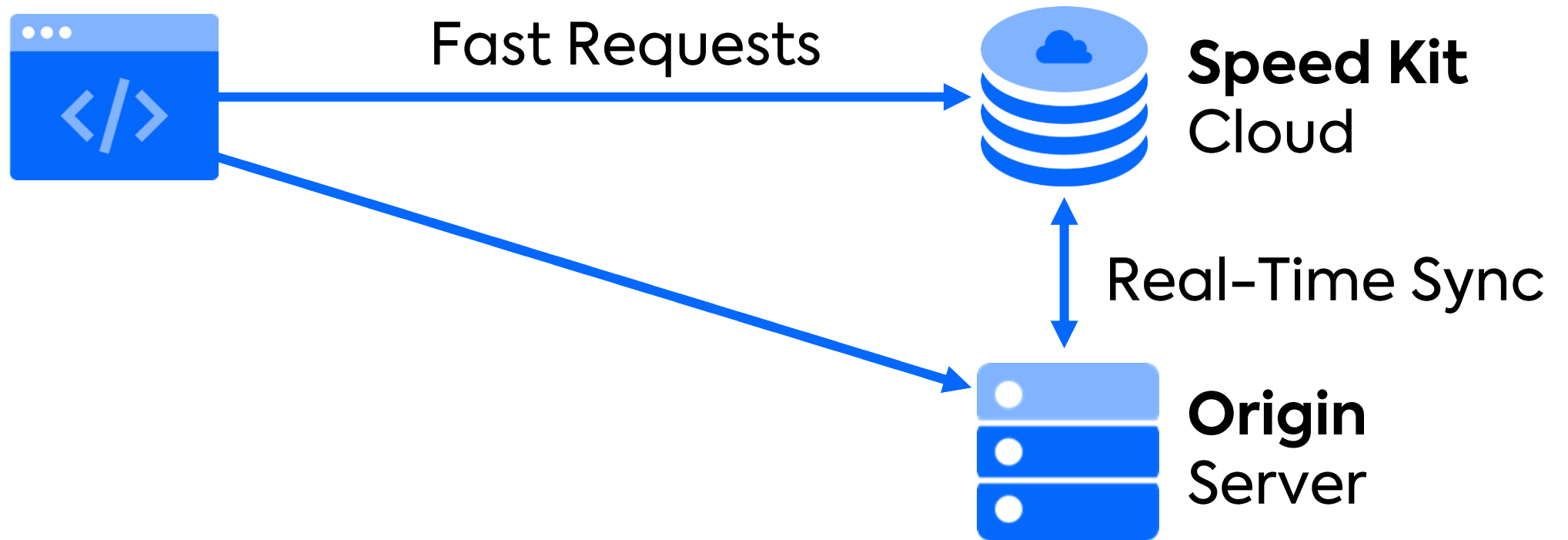
W. Wingerath, B. Wollmer, F. Gessert, S. Succo, N. Ritter. Going for Speed: Full-Stack Performance Engineering in Modern Web-Based Applications, WWW 2021

How to Make a Website Fast

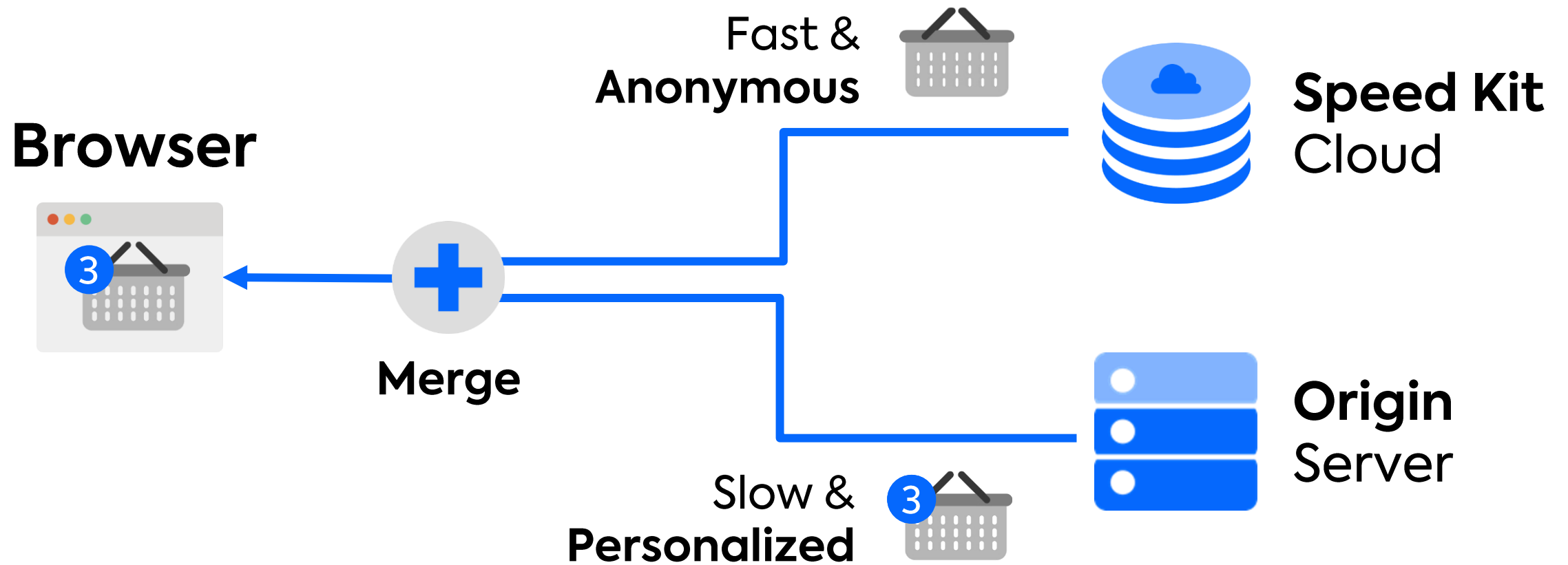


How **Speed Kit** Works

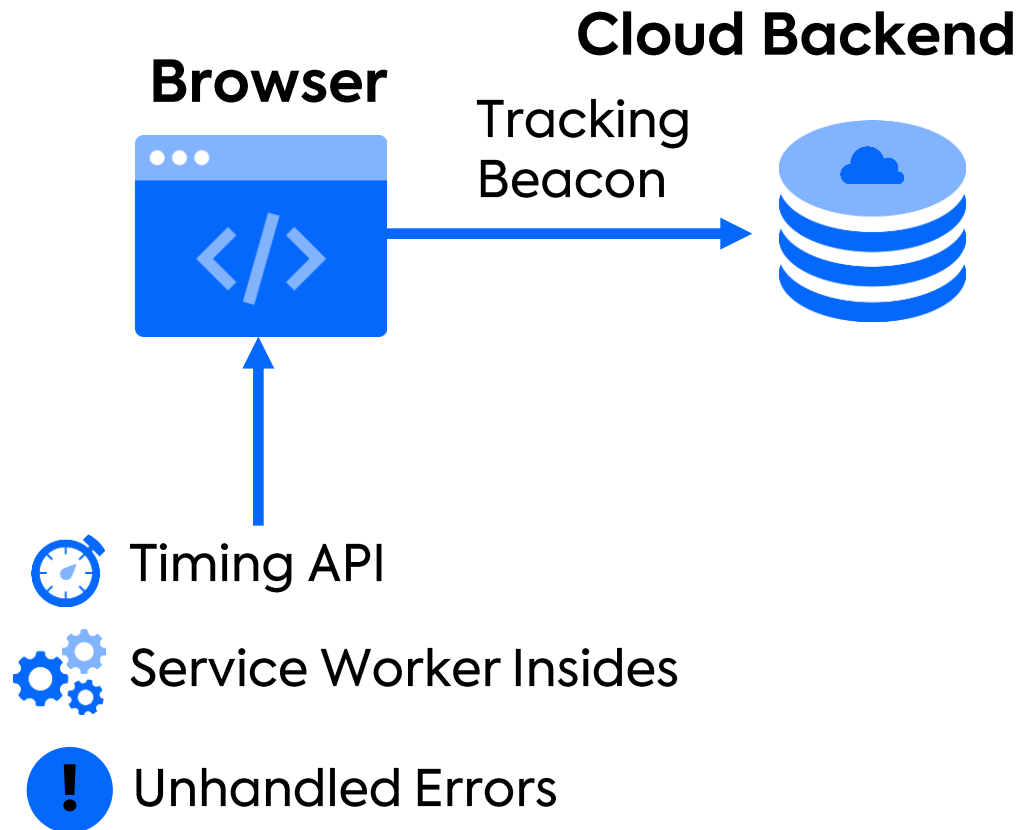
Website **Speed Kit JS**
(Service Worker)



Accelerating **Personalized** HTML



Measuring Technical & Business Performance



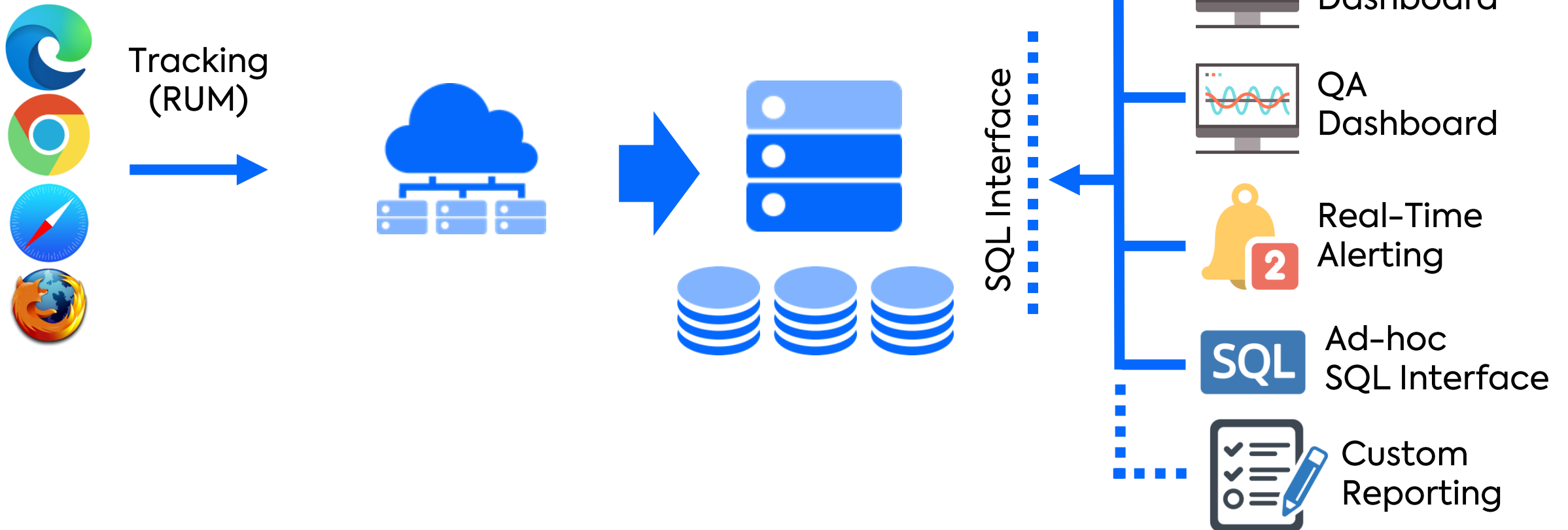
- Time-to-First-Byte - First (Contentful) Paint - DOM Timer - First Input Delay	Performance
- Session Length - Time on Site - First User Interaction - Bounce Rate	User Engagement
- Cart Size - Transactions - Conversion Rate - Revenue	Business KPIs
- Page Views & Sessions - Browser Distribution - JavaScript Errors - Caching Insights	QA Metadata

Building a RUM Data Platform: Challenges

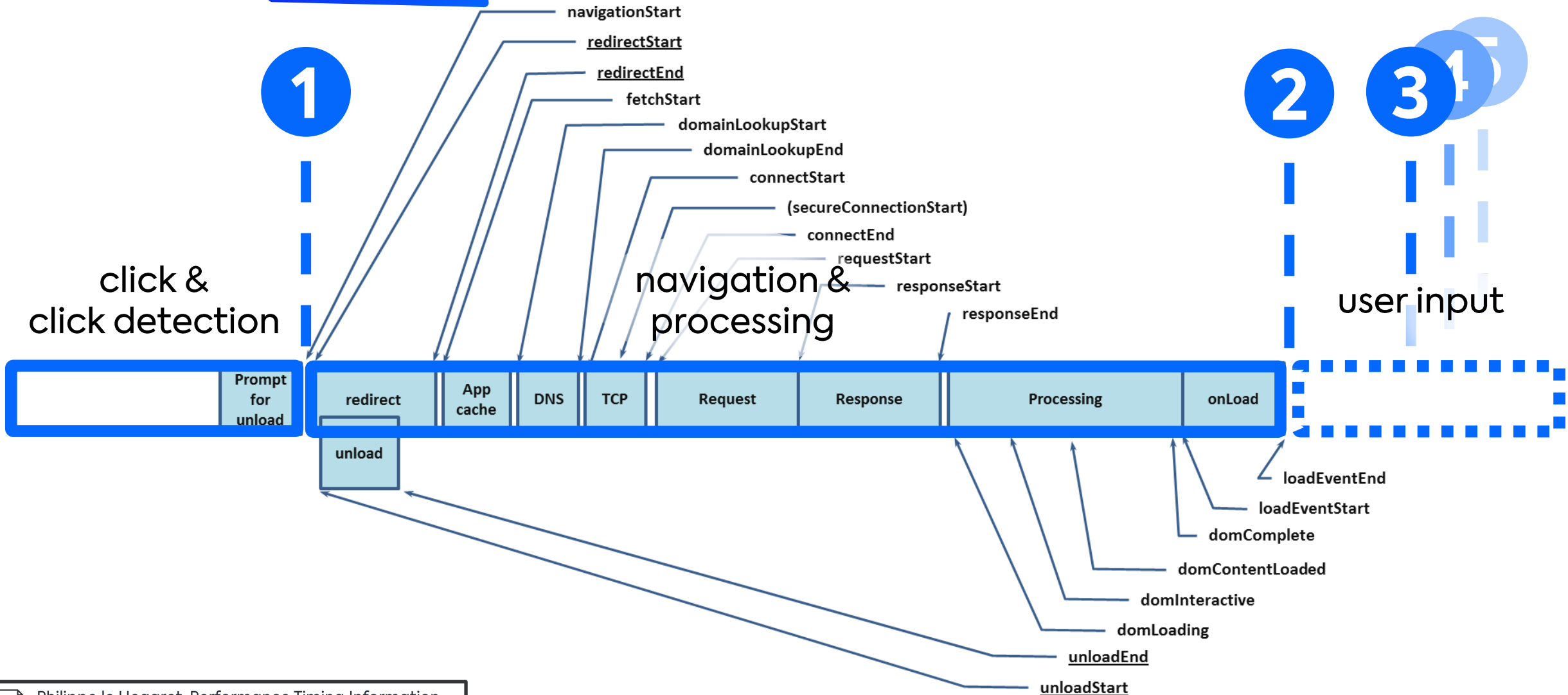
Data Collection

Data Processing & Storage

Use Cases

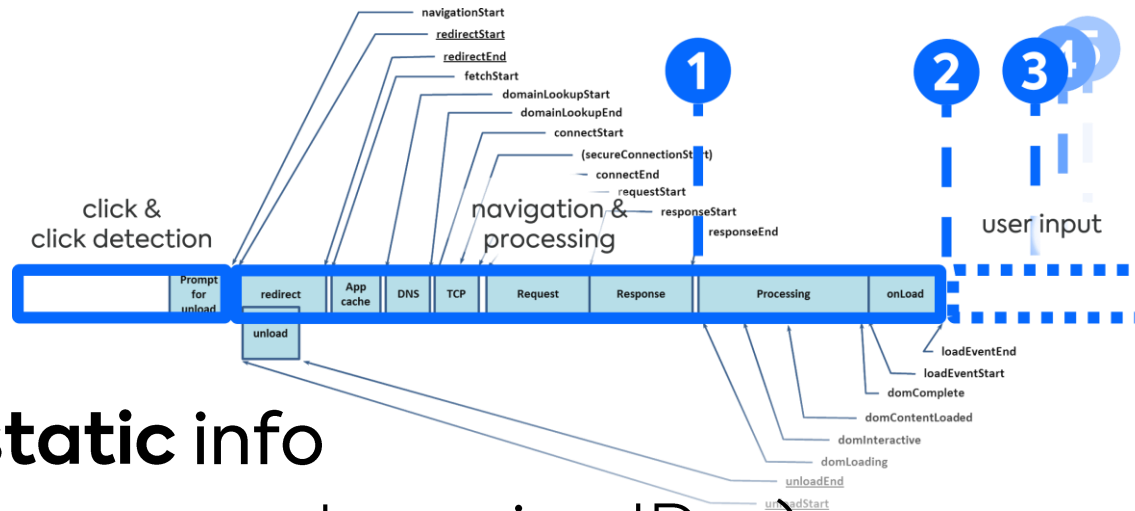


When to Send Data Beacons ?



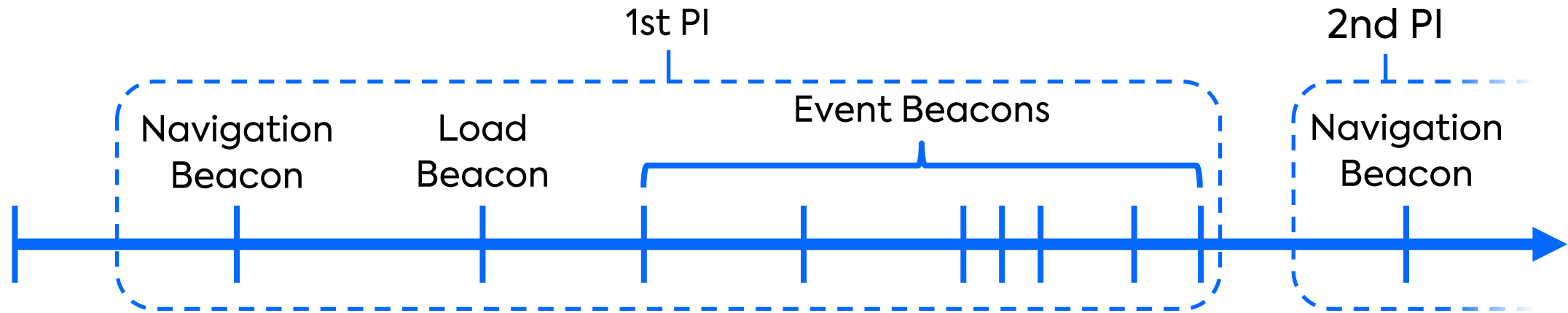
Philippe le Hegaret. Performance Timing Information: Part 1 – Navigation Timing, W3C Blog, 2012

Types of Data Beacons



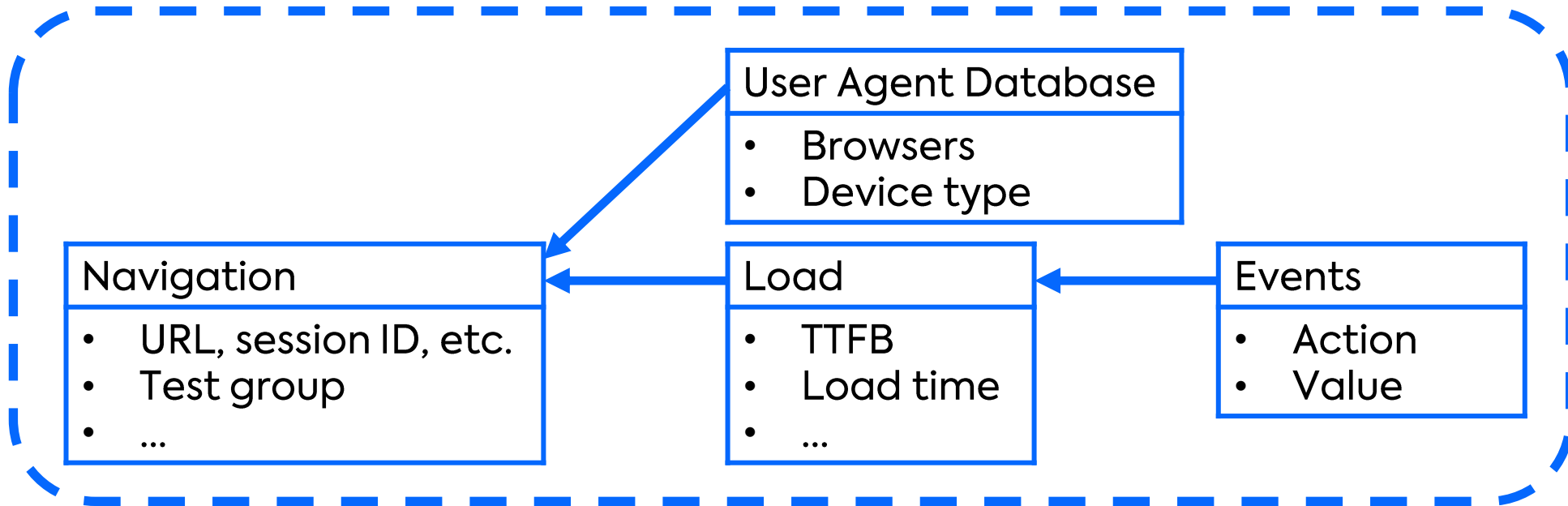
1. 1 for **static** info
(URL, user agent, session ID, ...)
2. 1 for **timings**
(TTFB, load time, FCP, ...)
3. 0-n for **events**
(first input, JavaScript errors, ...)

Schema: **PI**



- **Beacon Join → PI:** How do we handle events that come late?
 - Simply wait 5 minutes?
 - Wait for next PI or session timeout?
 - ...?
- How to resolve **user agents**?

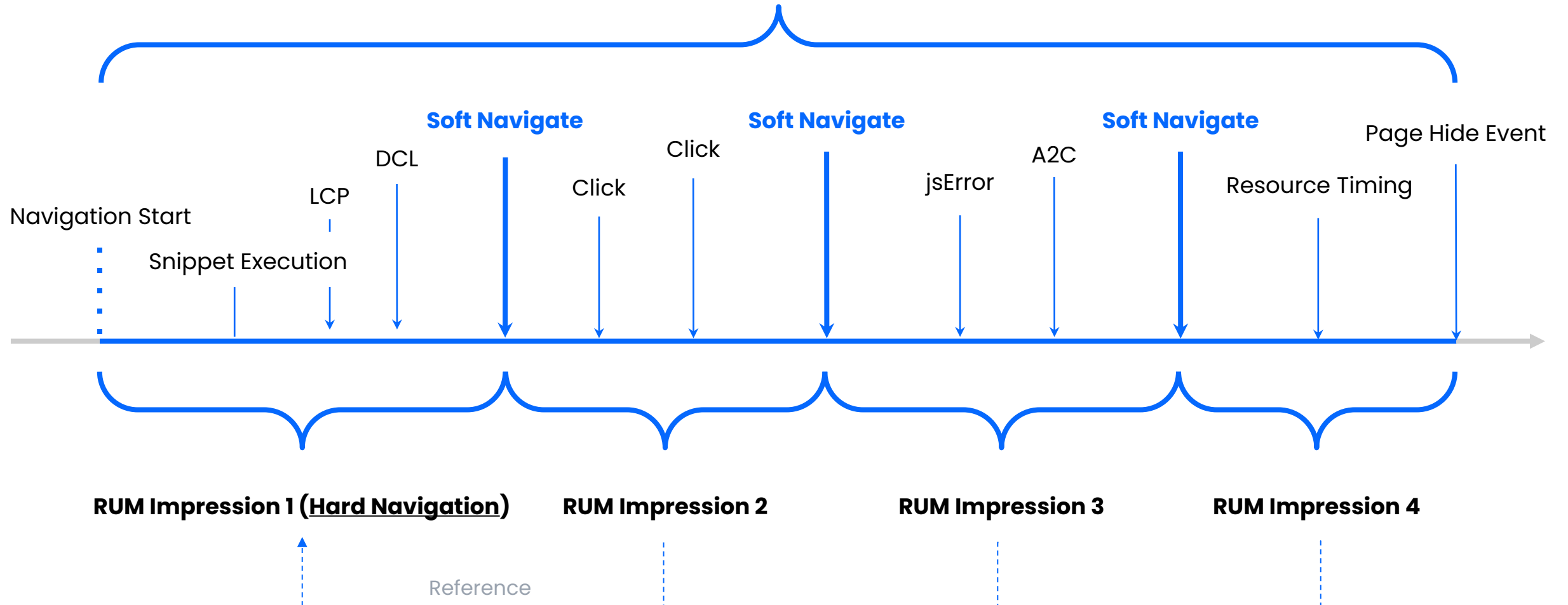
Early Schema: **PI**



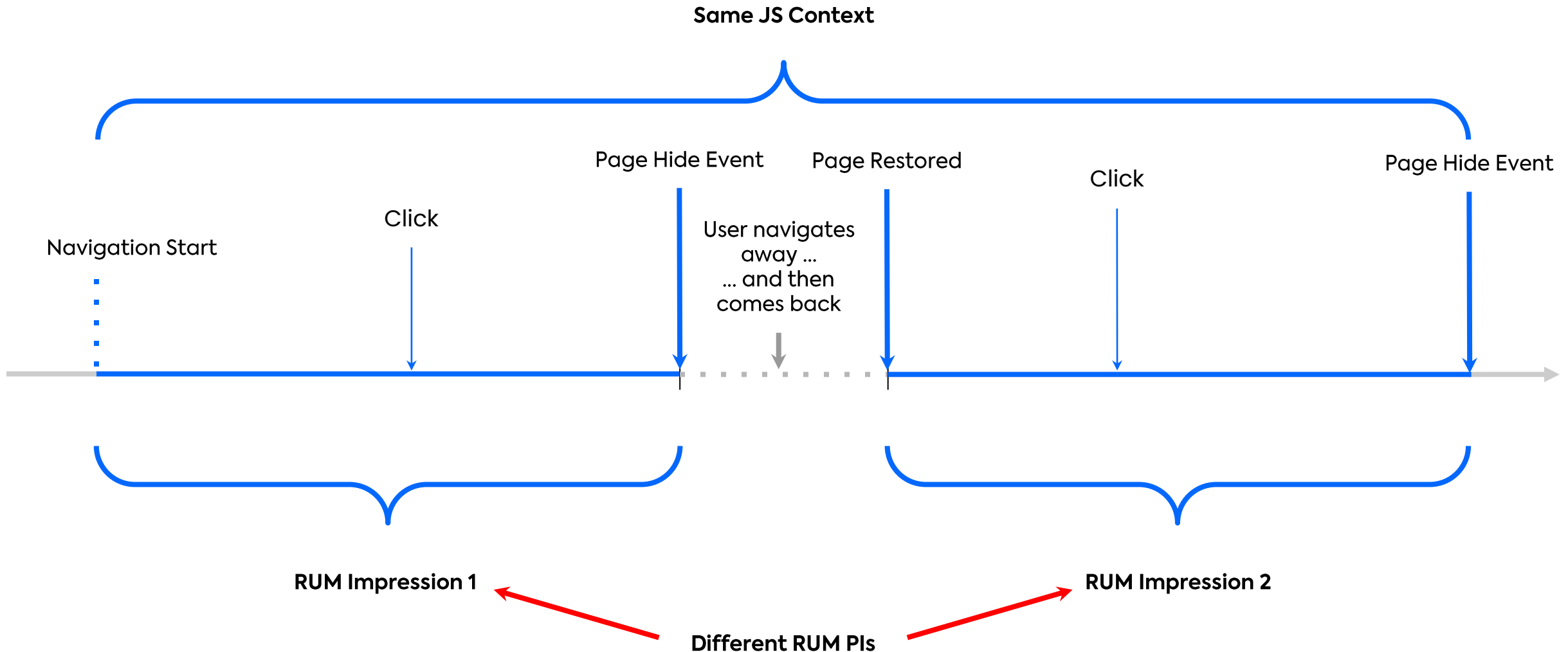
- **Aggregate events:** collect all events per PI
- **Join 3 Collections:** put together PI from navigation/load/event beacons
- **Resolve User Agents:** derive browser, device, etc. from UA string

SPA Challenge: Soft Navigations

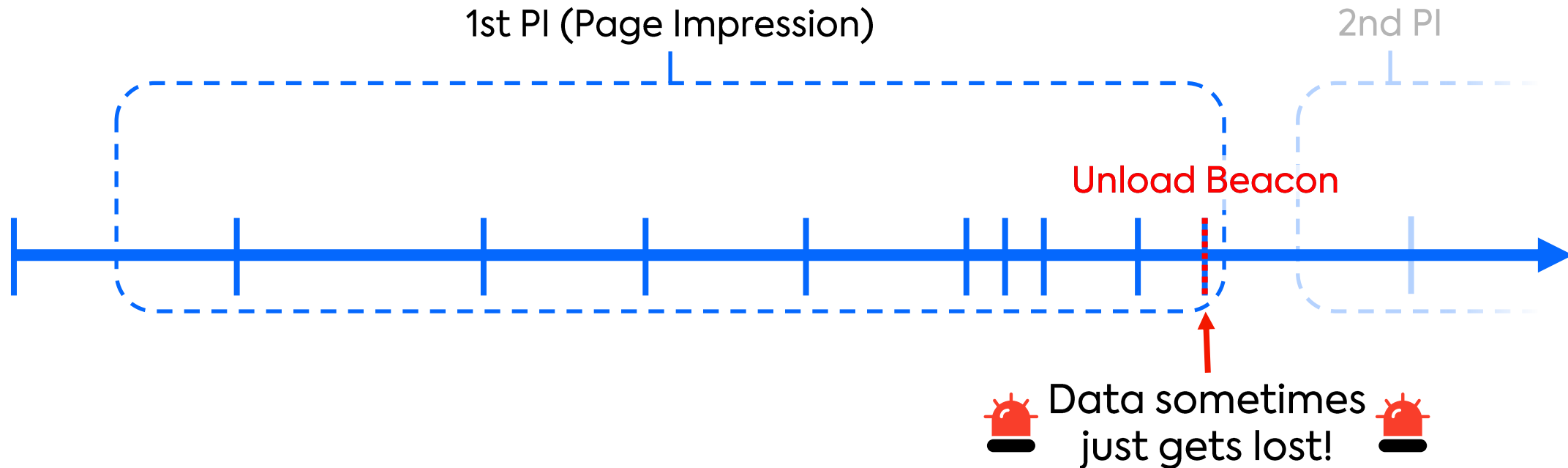
Hard Navigation for Browser



JS Challenge: Backward-Forward-Cache



Reliability Challenge: Mostly Working in Practice

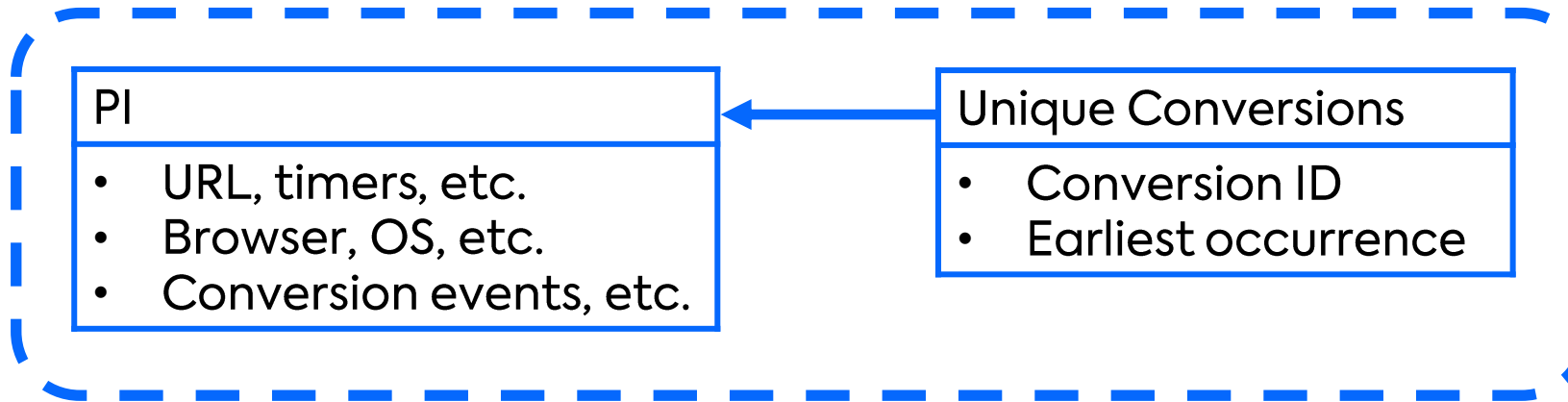


- **Dropped Beacons:** tracking data may be discarded by the browser
- **Browser „Optimizations“:** suspending inactive tabs, tracking prevention (e.g. Safari, Brave)
- **Varying Feature Support:** API availability and behavior may differ by browser vendor or version
- Also: broken APIs, non-standard customer implementations, hidden dependencies, high complexity as tracking happens on many layers ...



Wolfram „Wolle“ Wingerath. Data Validation at Scale: Managing Data Quality in Complex Data Pipelines, Code.Talks 2023

Early Schema: **Session**



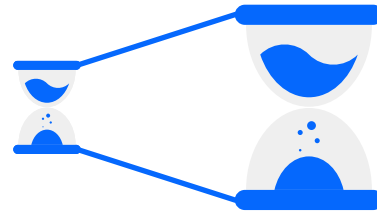
- **Aggregation by session ID:** min, max, count, sum, avg, median, ...
- **Deduplicate Conversions:** only 1st occurrence per conversion is valid

Latency Challenge: Debugging vs. Analytics

Alerting



Processing Time



Trend Analysis



- Simple metrics / little context
 - Counters
 - Extreme values
 - Specific errors

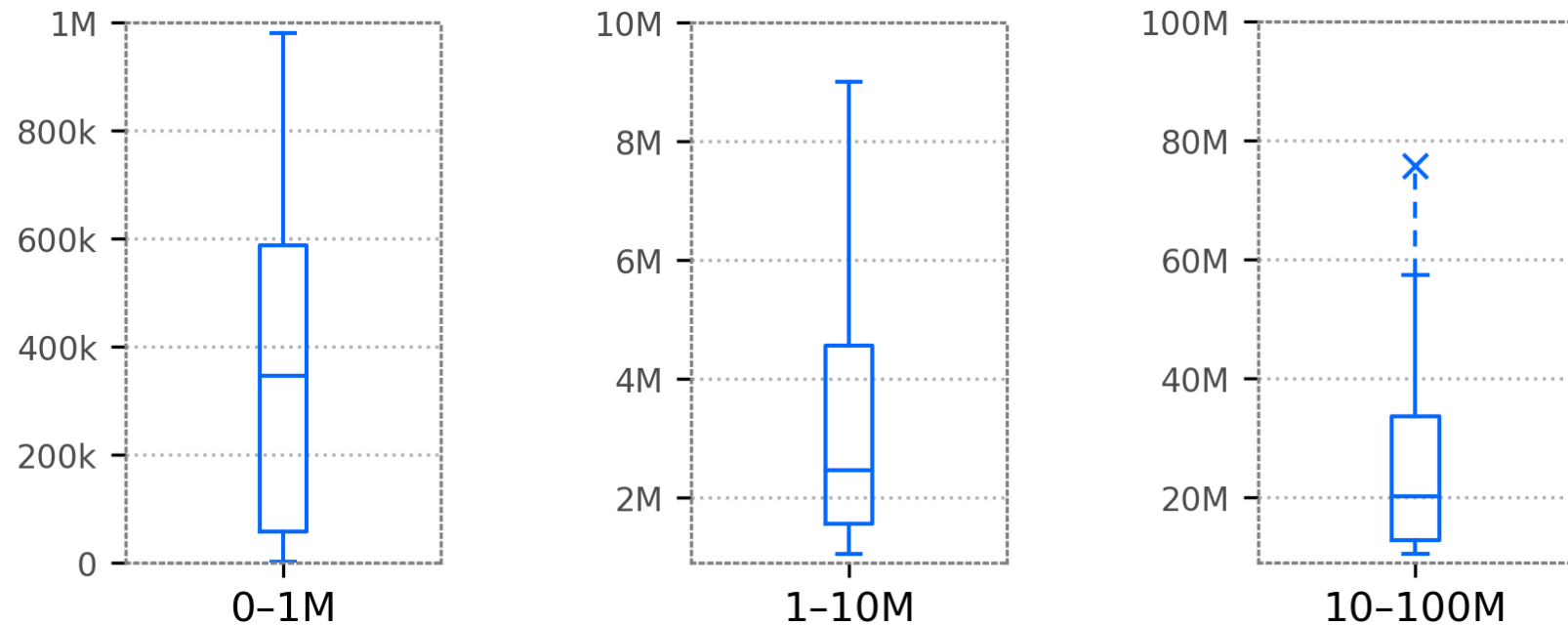
- Complex aggregations / huge time windows
 - Bounce rate
 - Performance by month
 - Seasonal effects

Multi-Tenancy Challenge: 3 Orders of Magnitude*

Overall Monthly Traffic*

- >350M users
- >500M sessions
- >2B PIs
- >10B data beacons

PI Distribution by Magnitude

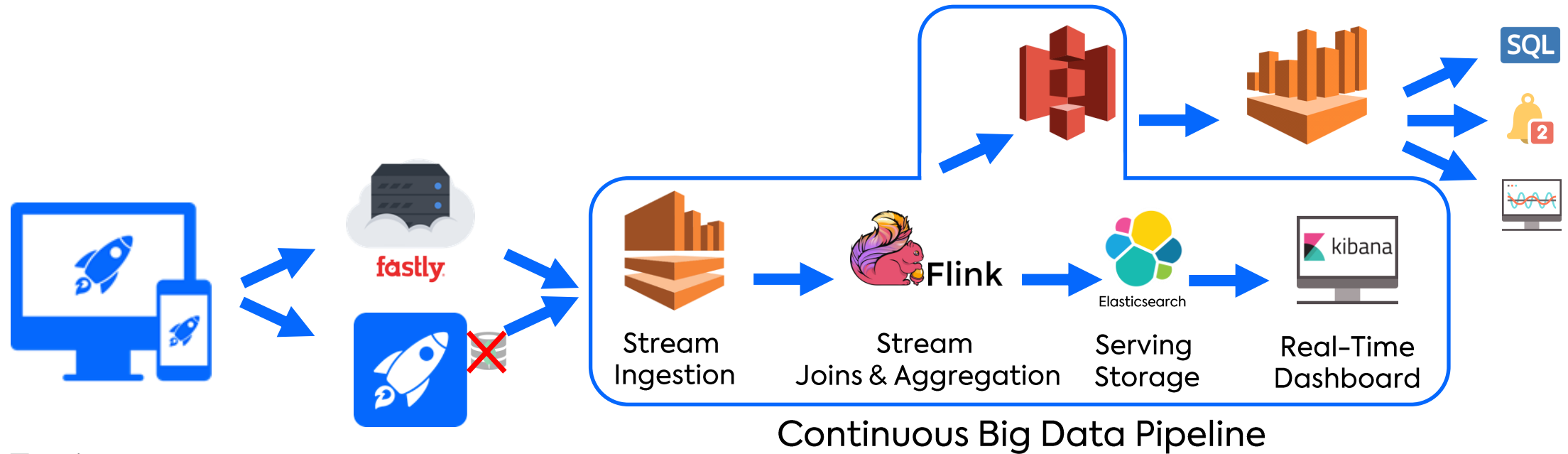


*Monthly Page Impressions (PIs), August 2026



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)

Processing Model Evolution: Batch to **Stream**



Key Topics:

- ✓ No legacy tech => stability & efficiency
- ✓ Faster ingestion => real-time reporting & analytics
- ✓ Fewer joins => faster analytics

In Collaboration With
amazon | science



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)

Incremental Processing for Efficient Monitoring

Partial Page Impressions (PPIs)
Enhanced Data Beacons

Time	Browser	Device	Test Group	First Contentful Paint (FCP)
11:05:04.578	Firefox	Mobile	Speed Kit	127ms
11:06:48.139	Chrome	Mobile	Original	958ms

1-Min. Time Windows
Immediate Aggregates (Storage)

	Browser	Device	Test Group	First Contentful Paint (FCP)
11:05	Firefox	Mobile	Speed Kit	{200ms: 1, 500ms: 2}
	Firefox	Mobile	Original	{600ms: 2, 800ms: 5}
	Safari	Desktop	Original	{1100ms: 1}
11:06	Firefox	Mobile	Speed Kit	{200ms: 3}
	Chrome	Mobile	Speed Kit	{400ms: 2}
	Opera	Tablet	Original	{700ms: 1, 1300ms: 2}
	Safari	Desktop	Original	{600ms: 4, 900ms}

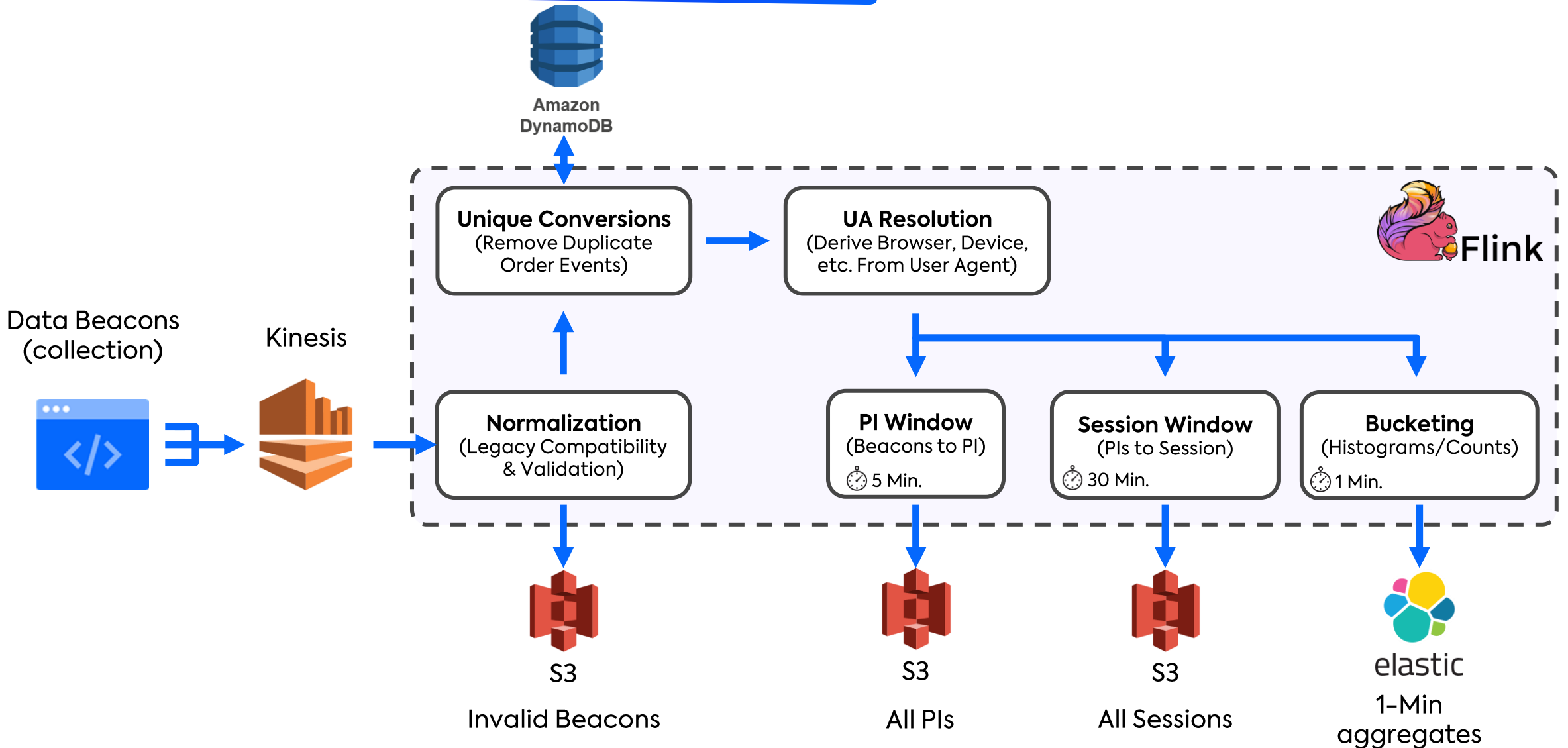
Arbitrary Time Windows
Real-Time Reporting (Dashboard Queries)

11:05	Browser	Device	Test Group	First Contentful Paint (FCP)
–				
11:06	Firefox	Mobile	Speed Kit	{200ms: 4, 500ms: 2}

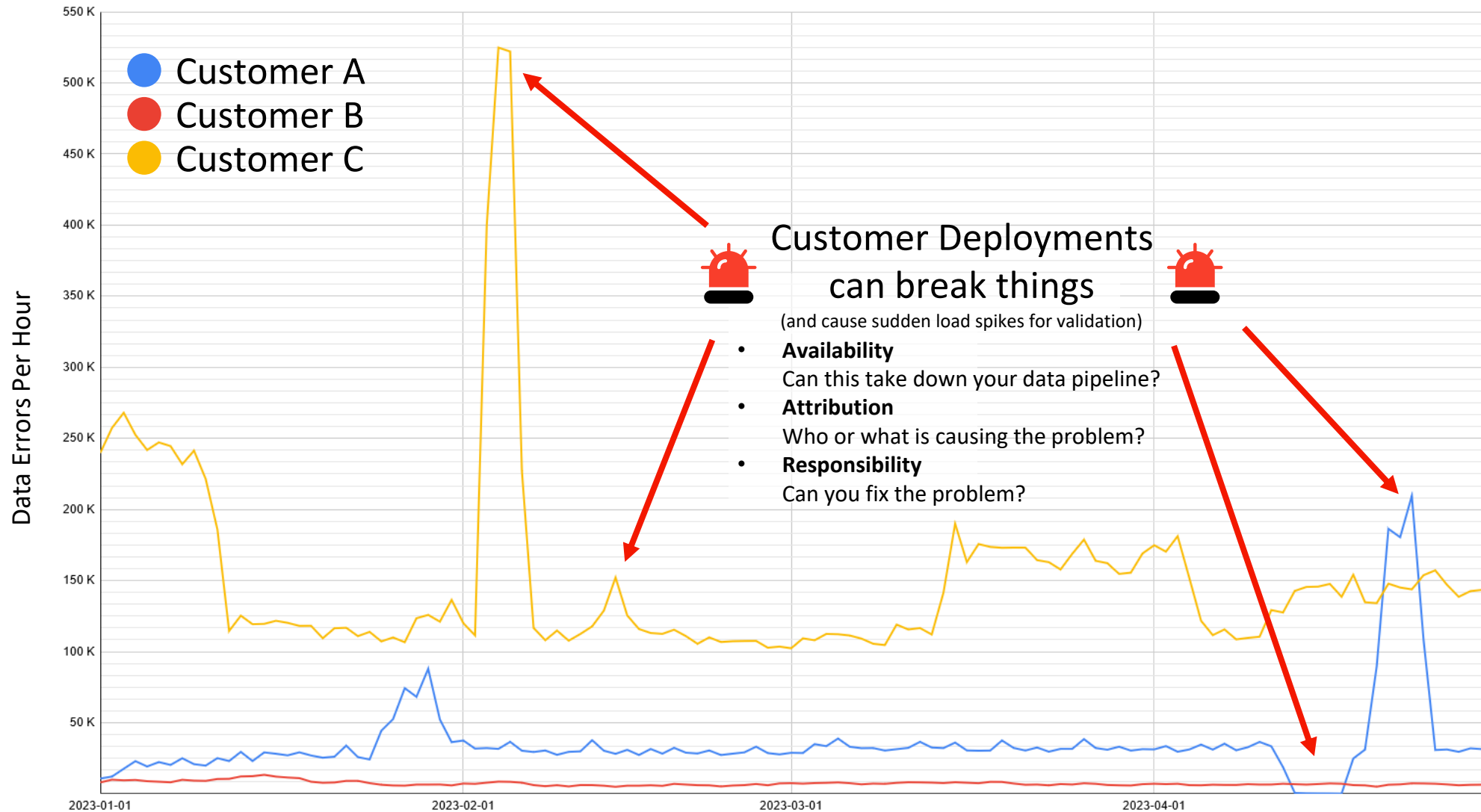


W. Wingerath, B. Wollmer, M. Bestehorn, S. Succo, F. Bücklers, J. Domnik, F. Panse, E. Witt, A. Sener, F. Gessert, N. Ritter. Beaconnect: Continuous Web Performance A/B-Testing at Scale, VLDB (2022)

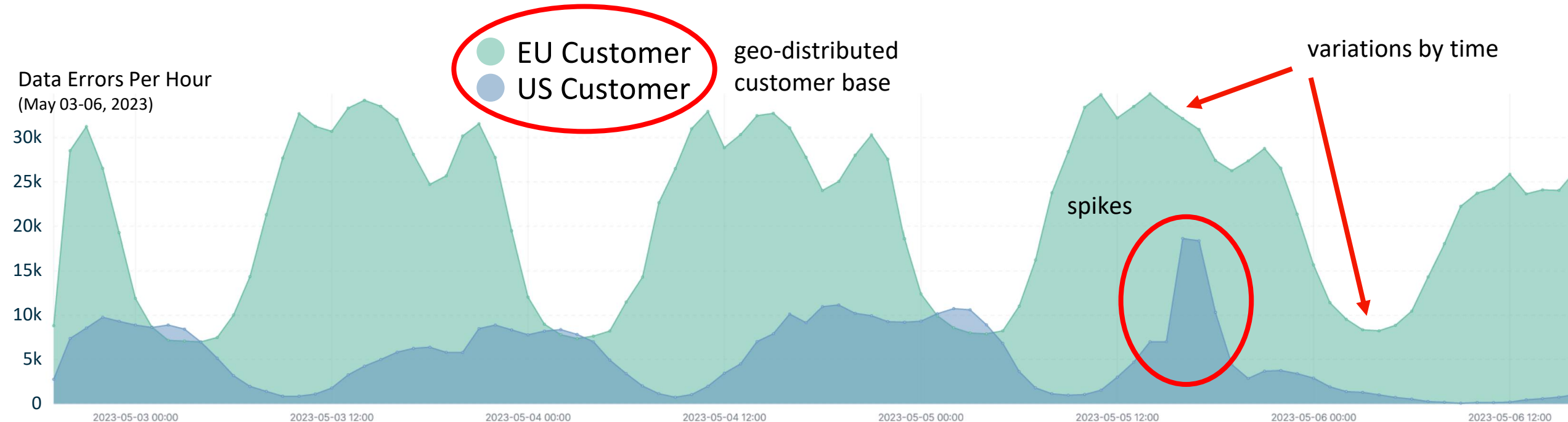
Early Processing Architecture



Robustness Challenge: Customer Deployments



Elasticity Challenge: Fluctuating Load



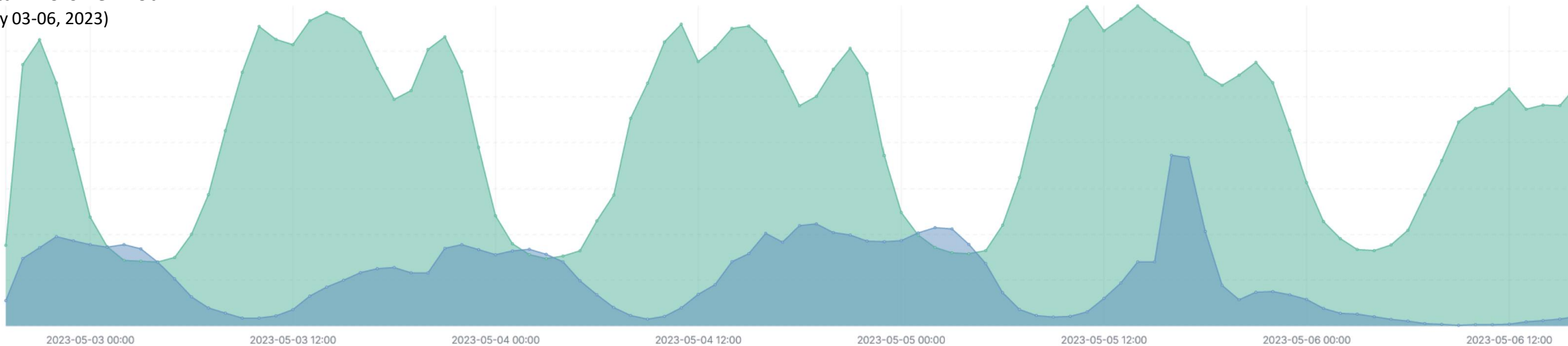
- Specifying generalized constraints can be difficult in large deployments, because of ...
 - ... temporal fluctuations (e.g. throughout the day, on black Friday, or during holidays)
 - ... multi-tenancy (e.g. different data patterns by customer timezone or domain)

→ **Elasticity & Multi-Tenancy** requirements can be challenging!

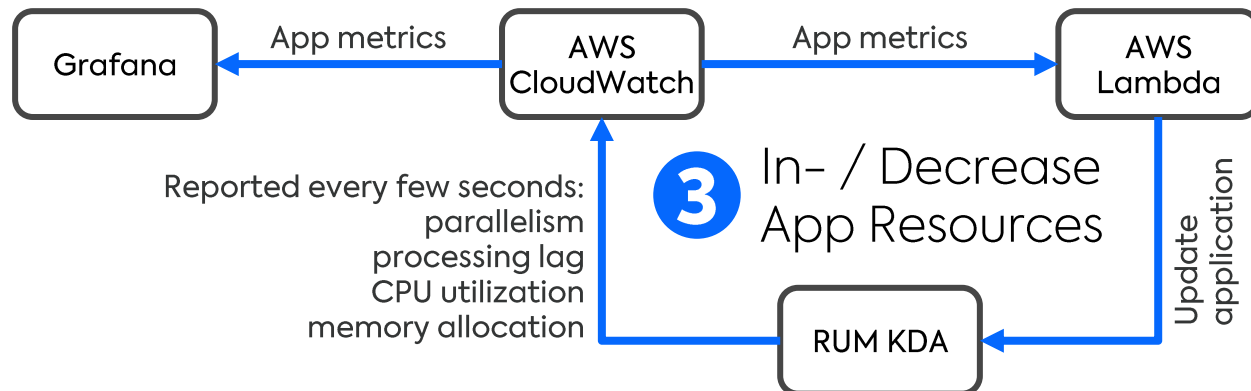
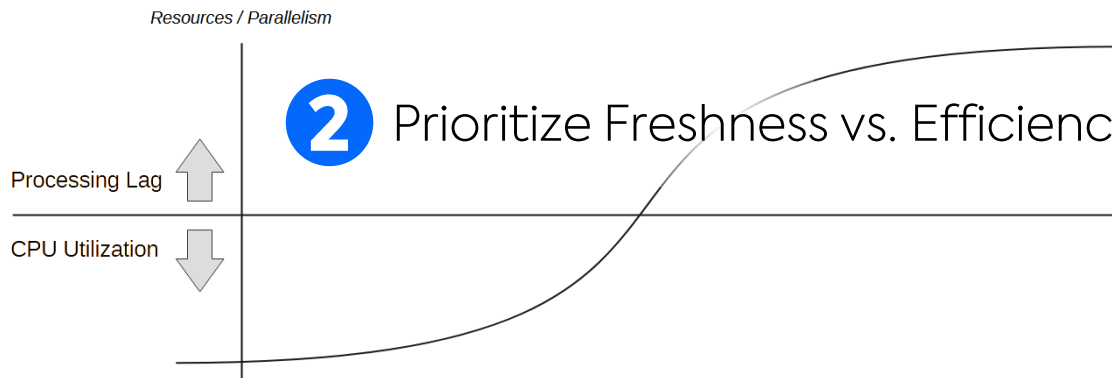
DevOps Evolution: **Autoscaling**

1 Load Monitoring

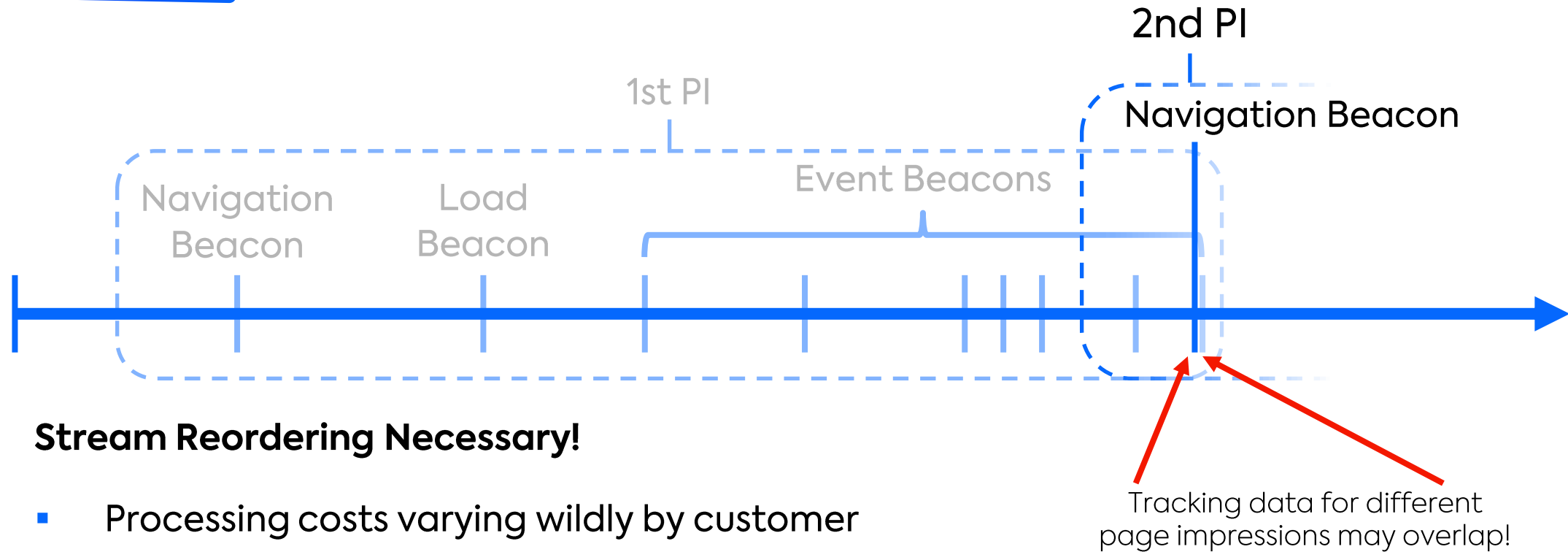
Web Traffic
~~Data Errors Per Hour~~
(May 03-06, 2023)



2 Prioritize Freshness vs. Efficiency

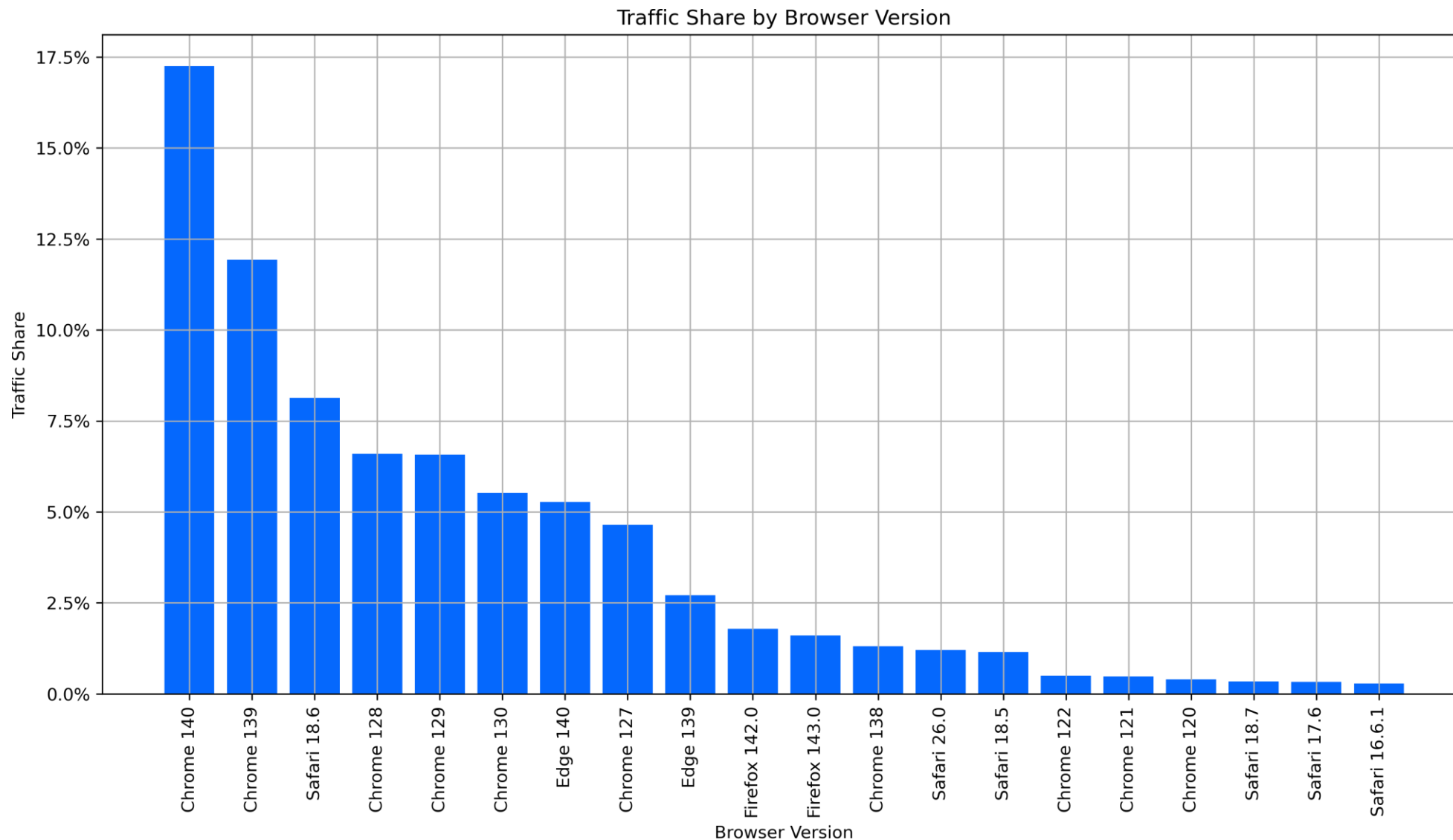


Sorting Challenge: Out-of-Order Data Arrival

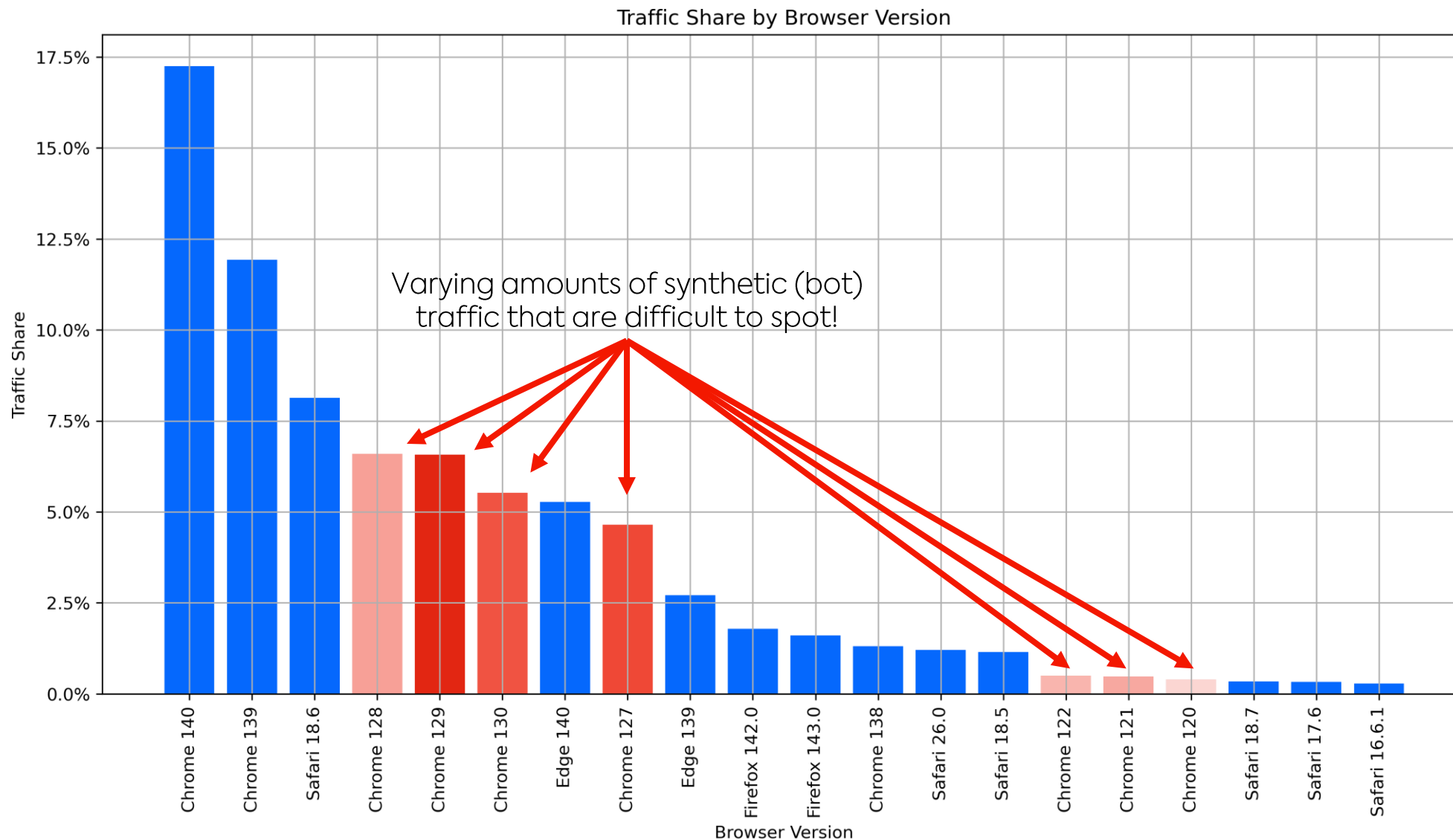


- **Stream Reordering Necessary!**
 - Processing costs varying wildly by customer
 - Increasing complexity (e.g. event time vs. processing time, client-side indexing vs. server-side indexing)
 - More complex debugging
- Even Worse for **Soft PIs!**

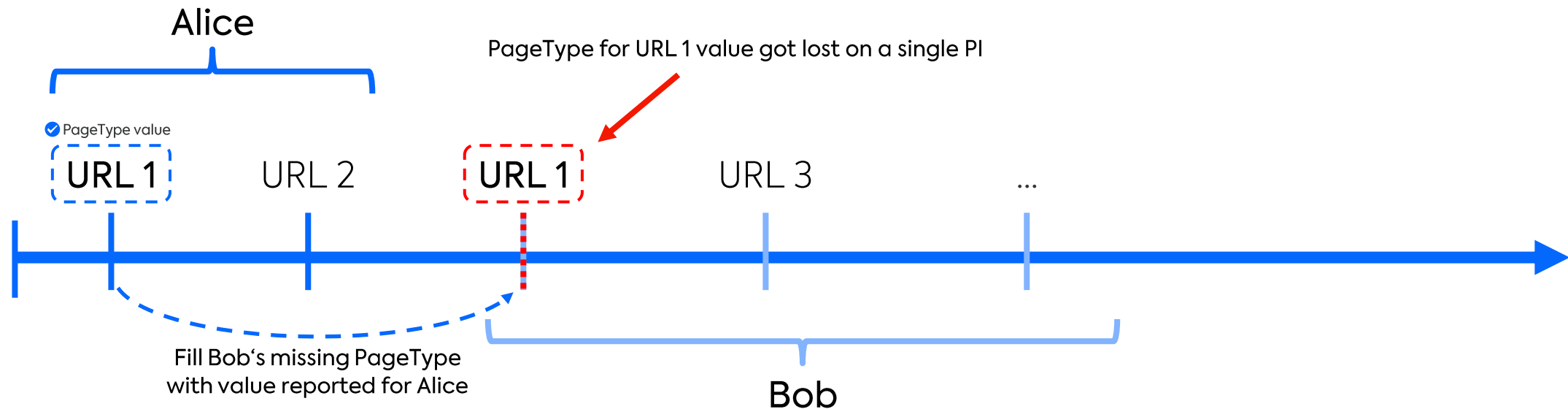
Bot Traffic Challenge: Filtering Synthetic Traffic



Bot Traffic Challenge: Filtering Synthetic Traffic

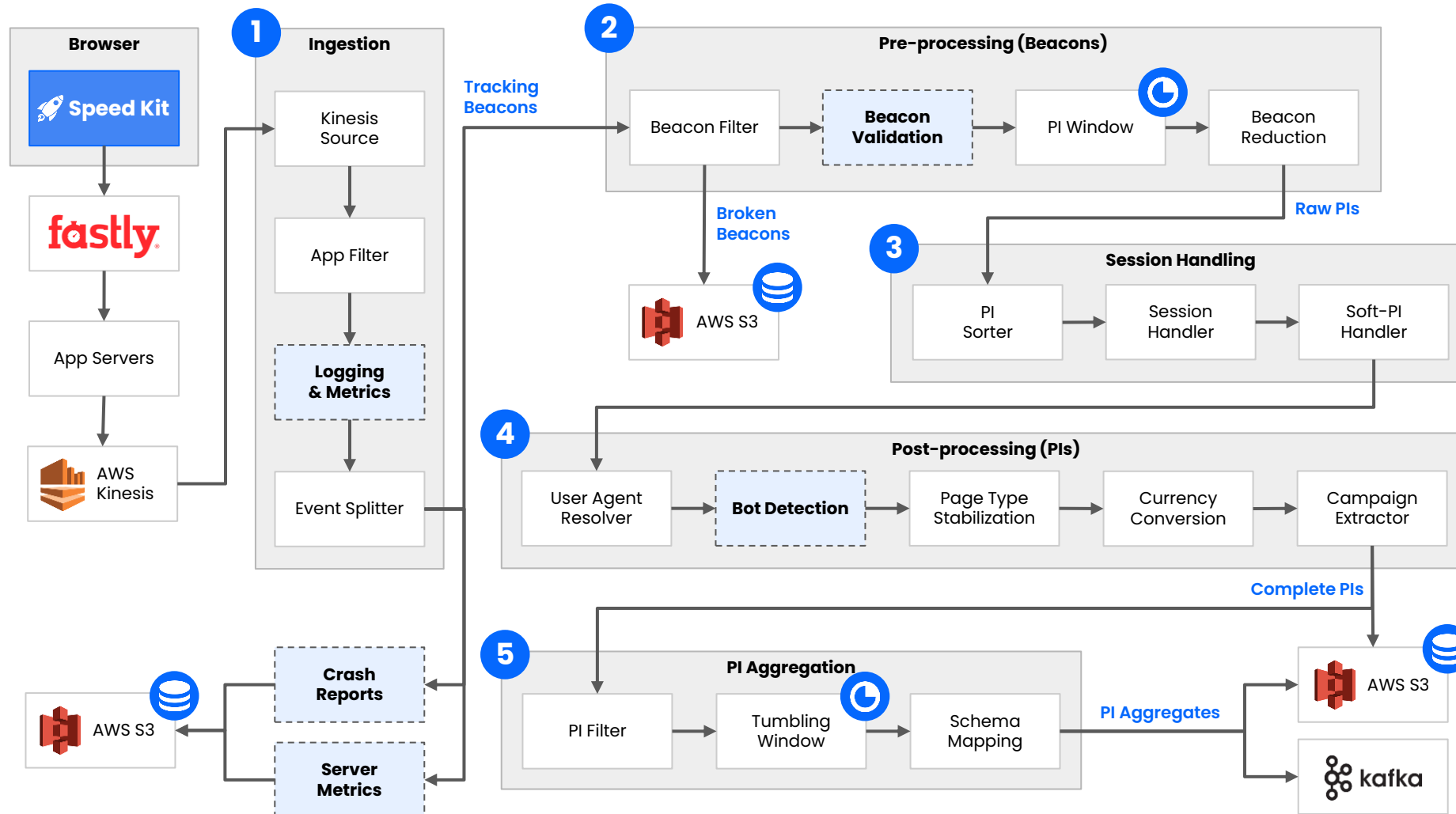


Many **Micro** Challenges: Unification, Encoding, ...



- **Stabilizing values** such as missing Page Type, inconsistent split values, etc.
- **Currency unification**, e.g. by transforming transaction revenue to USD by current exchange rate
- **Marketing channel extraction** from URL campaign parameters (source, medium, campaign)
- **Attribute (de-)encoding**, i.e. making minified tracking attributes human-readable again

Processing Architecture Evolution



Split Testing for Web Performance

Speed Kit Users

vs.

Normal Users



Tracking
→



Tracking
←



- Speed Kit enabled

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

- Speed Kit disabled
(no acceleration)



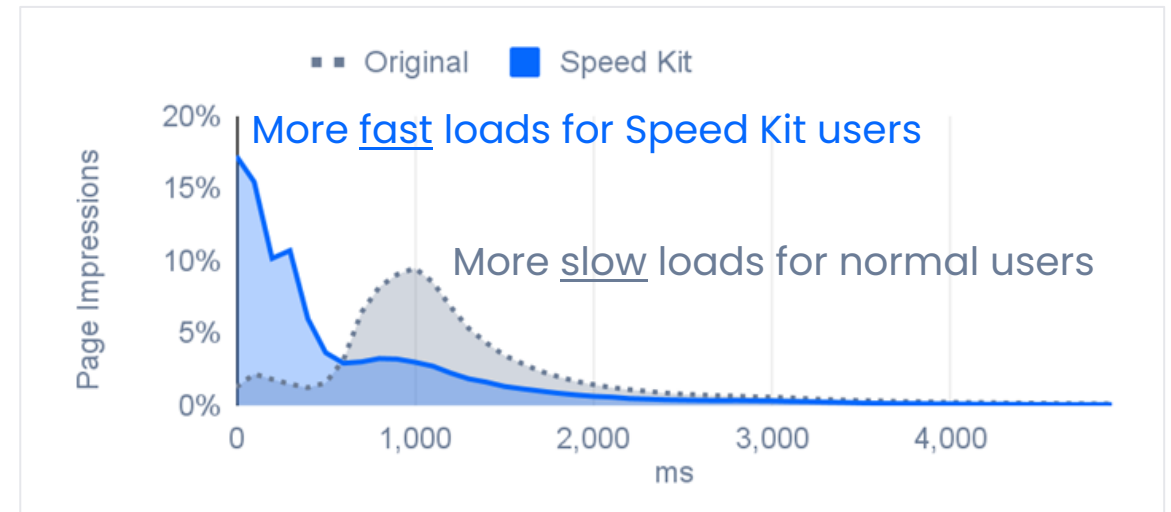
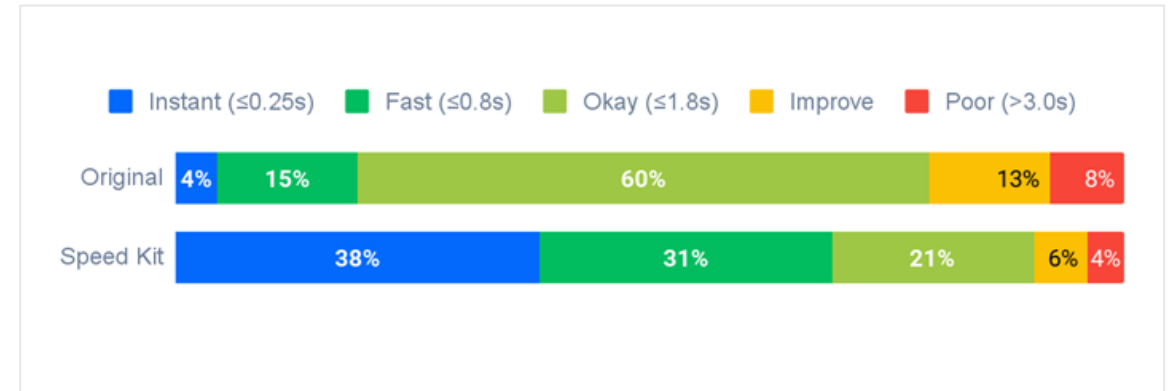
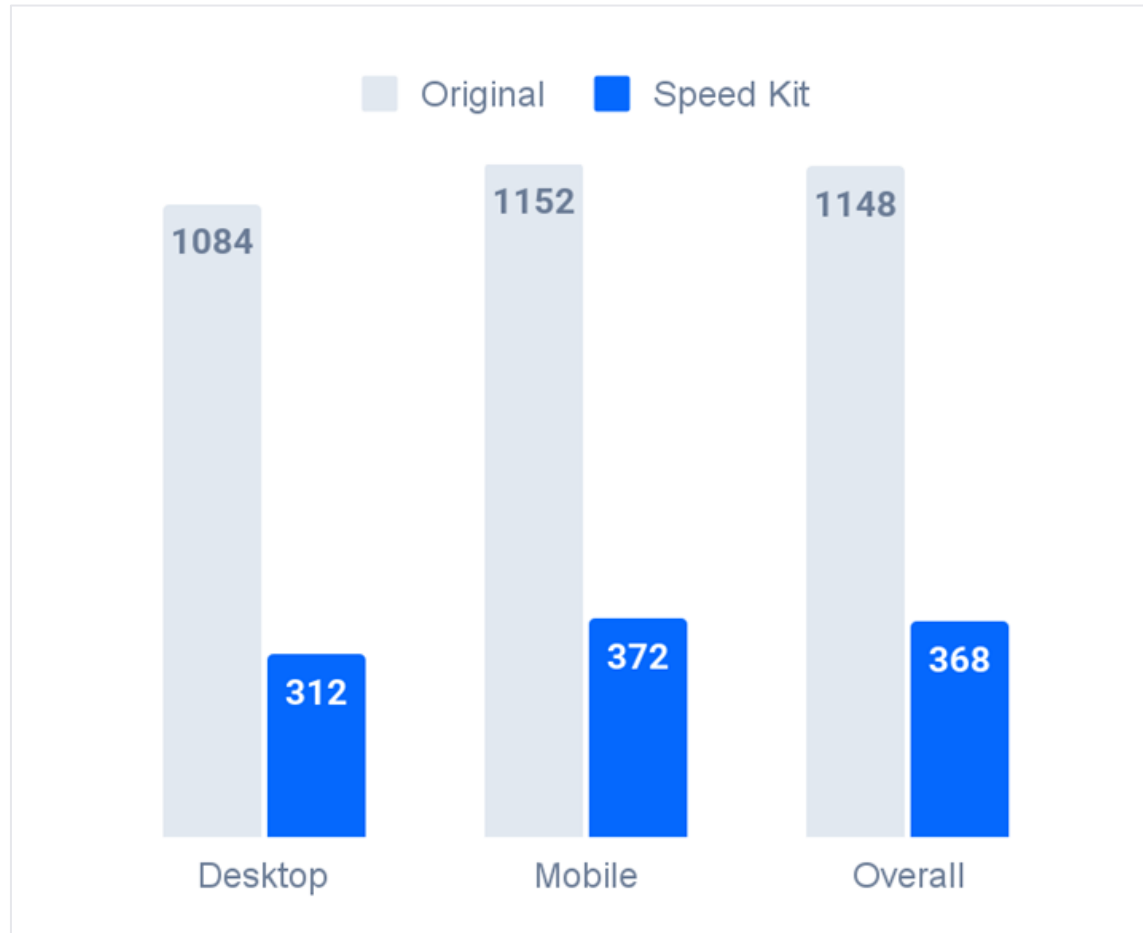
W. Wingerath, F. Gessert, E. Witt, H. Kuhlmann, F. Bücklers, B. Wollmer, N. Ritter. Speed Kit: A Polyglot & GDPR-Compliant Approach For Caching Personalized Content, ICDE 2020

Continuous Uplift Visualization

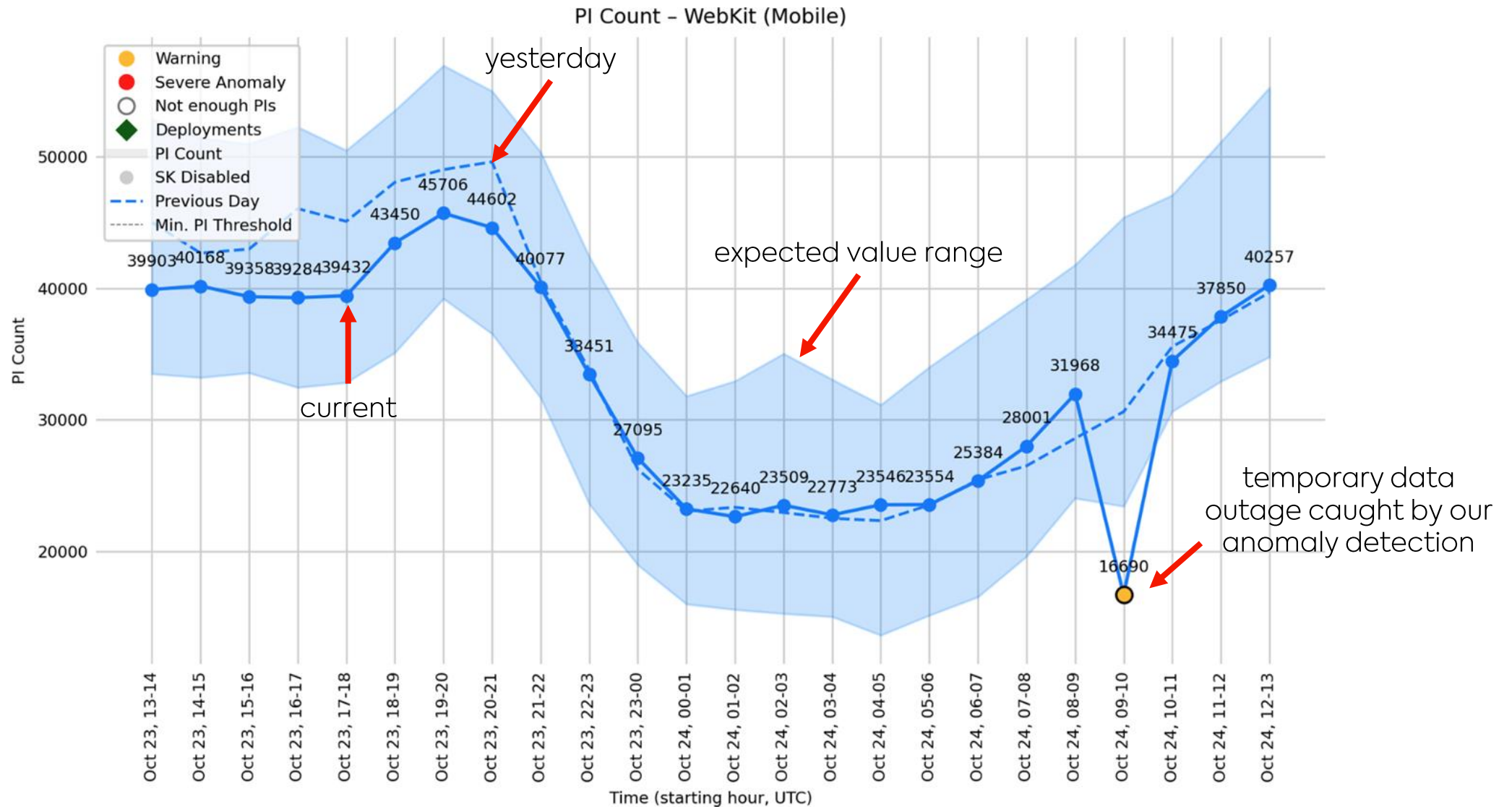
780 ms
faster (Median)

First Contentful Paint

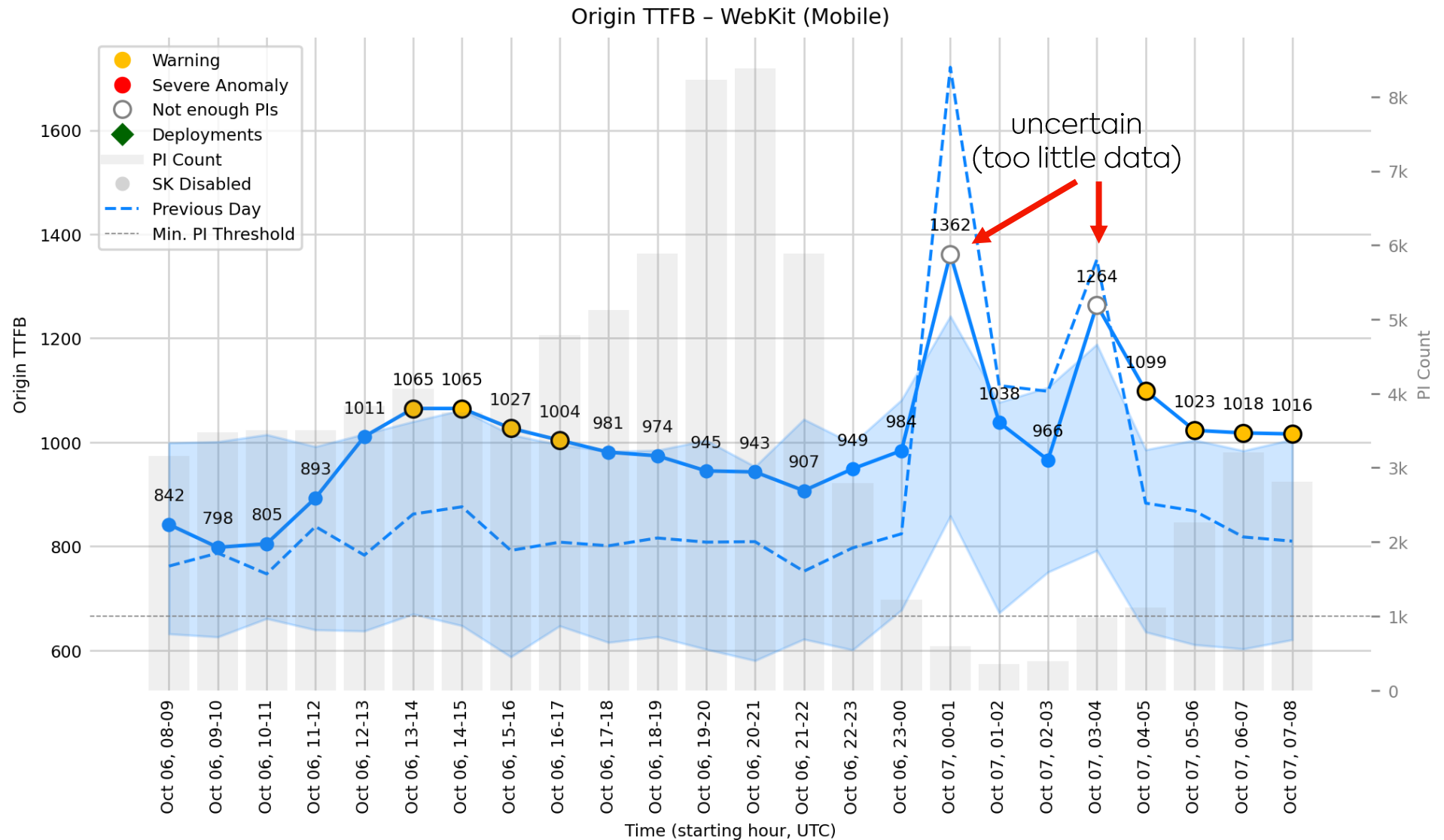
68%
faster



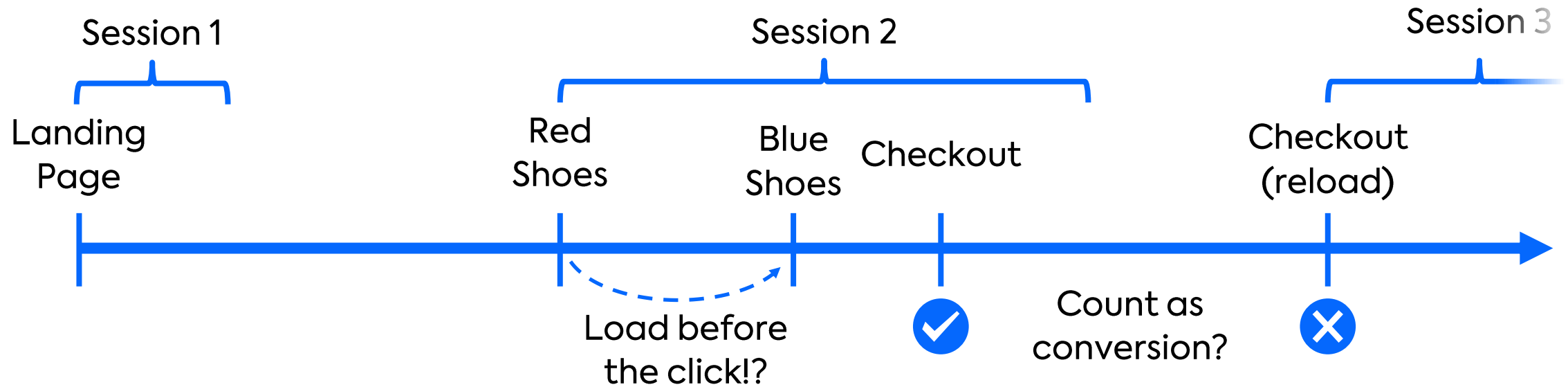
Anomaly Challenge: Finding Issues Quick



Anomaly Challenge: Finding Issues Quick

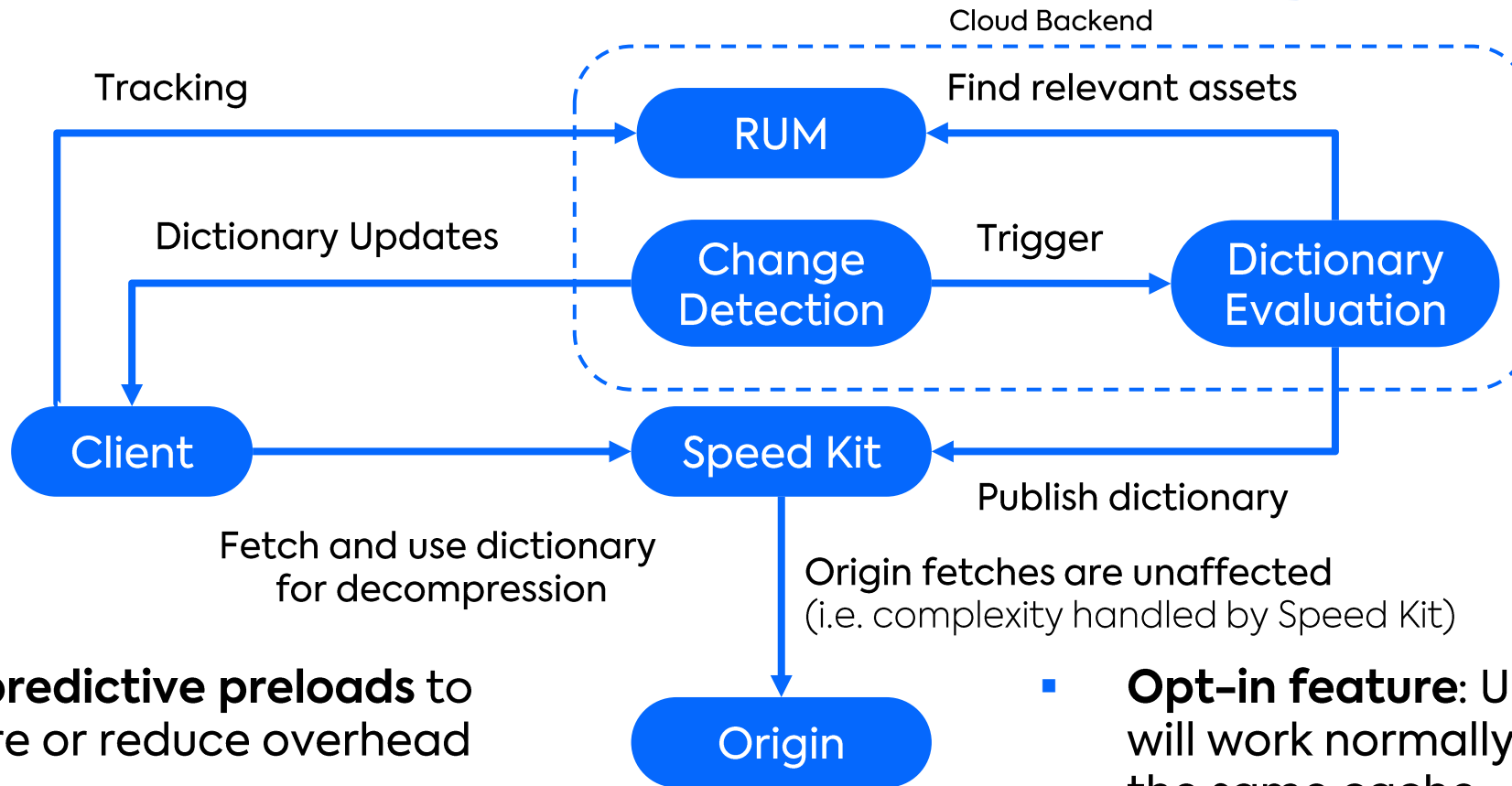


Preloading Challenge: Predicting the Next Click



- **User Behavior Analysis:** identify „successful“ user journeys, quantify satisfaction, optimize UX, ...
- **Predictive Preloading:** click path prediction (ML), end-to-end system integration, monitoring, ...
- **Delta Encoding:** delta computation, caching strategy (dictionary generation, personalized content handling, content invalidation), browser support, preloading efficiency optimization, ...

SDC Challenge: **Shared Dictionary** Compression



- Synergy to **predictive preloads** to preload more or reduce overhead
- Often **Savings of >80%** for HTML files
- Relies on Speed Kit's data tracking platform to generate optimal dictionaries from **RUM data**

- **Opt-in feature:** Unsupported clients will work normally, but benefit from the same cache
- Dictionary compression is used **throughout the stack** and recomputed at the edge if needed

SDC in Numbers: Bandwidth Savings in Practice

Date

📅 1 week ago - today

×

97,7 %

Maximum Saving

91,6 %

Median saving

62,2 %

Minimum Saving

0

PIS with no saving

SDC in Numbers: Bandwidth Savings in Practice

Date

1 week ago - today

App

Pilot-Customer

97,7 %

Maximum Saving

91,6 %

Median saving

62,2 %

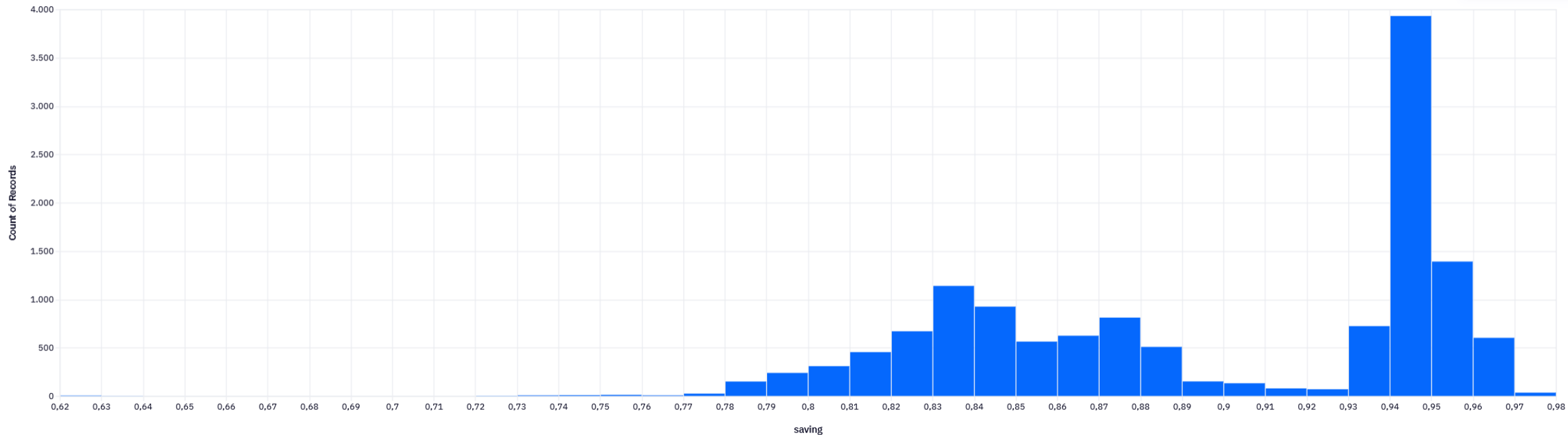
Minimum Saving

0

PIS with no saving

Saving per PI

Explore



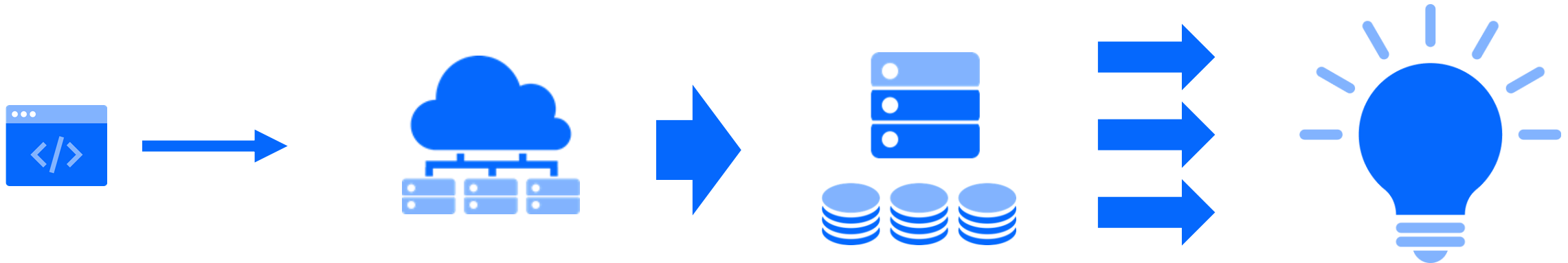
Speed Kit RUM Data Platform: **Recap**

Data Collection

- **Splitting up** PI data into beacons
- Managing multiple navigations within the same JavaScript context (esp. **Soft PIs**)
- Wrangling **different browsers**, standards, and custom implementations

Data Use Cases

- All data in **multiple granularities** to enable both detailed analytics and efficient reporting
- **Anomaly detection** for low response times and quick root cause identification
- **Predictive preloading** for next-level acceleration



Data Processing

- **Stream processing** to avoid visibility delays
- Optimal resource utilization through **autoscaling**
- **Multi-tenant architecture** with rigorous error profiling

Thanks! Questions?

