

³⁵Br^{eaking} ⁵⁶Ba^d Performance

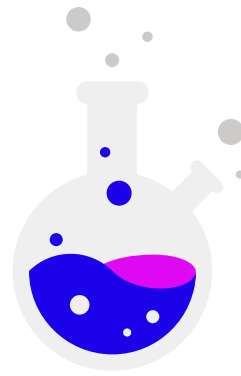
Mobile Site Speed and the Impact on E-Commerce

We Are **Baqend**

We bring performance research to practice.



40+ man-years of **web performance research**

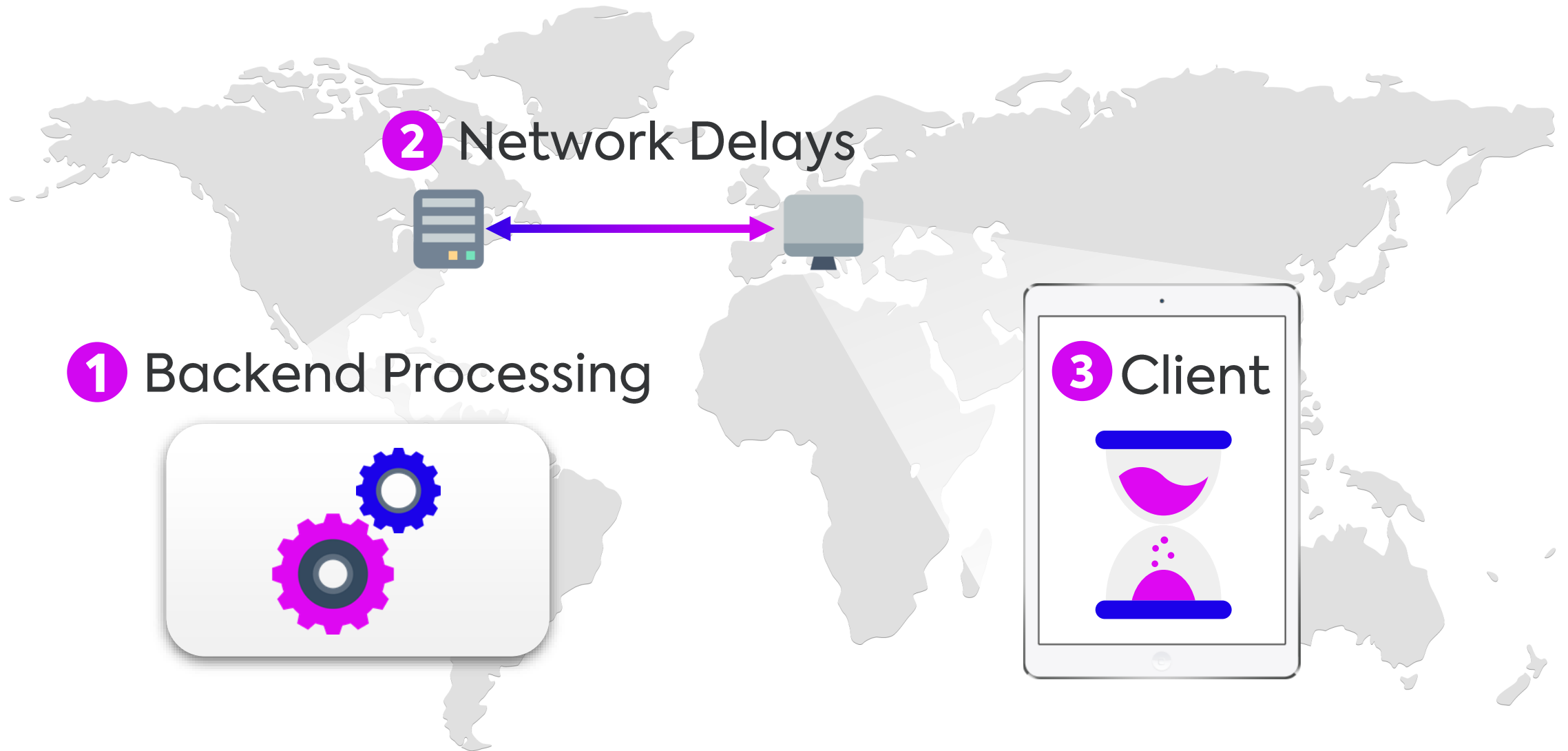


Novel technology for caching **dynamic data**



Speed Kit - SaaS for e-commerce speed

3 Things Make Your Website **Slow**



A man in a yellow hazmat suit is sitting in a chair in the center of a large, empty warehouse. The warehouse is filled with stacks of cash, some on the floor and some on pallets. The lighting is dim, and the overall atmosphere is one of mystery and wealth. The text is overlaid on the image in a white, bold font.

Why Do **Businesses** Care About Web Performance?

Delay **Psychology** : Rules of Thumb

Delay	User Perception
0 – 100 ms	Instant
100 – 300 ms	Small perceptible delay
300 – 1000 ms	Machine is working
1+ s	Mental context switch
10+ s	Task is abandoned



Stay under 1000 ms to keep users' attention



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



Load Time & **SEO**



From 7 s to 2 s 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



+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

Summary: The **Business Impact** of Site Speed



Summary: The **Business Impact** of Site Speed

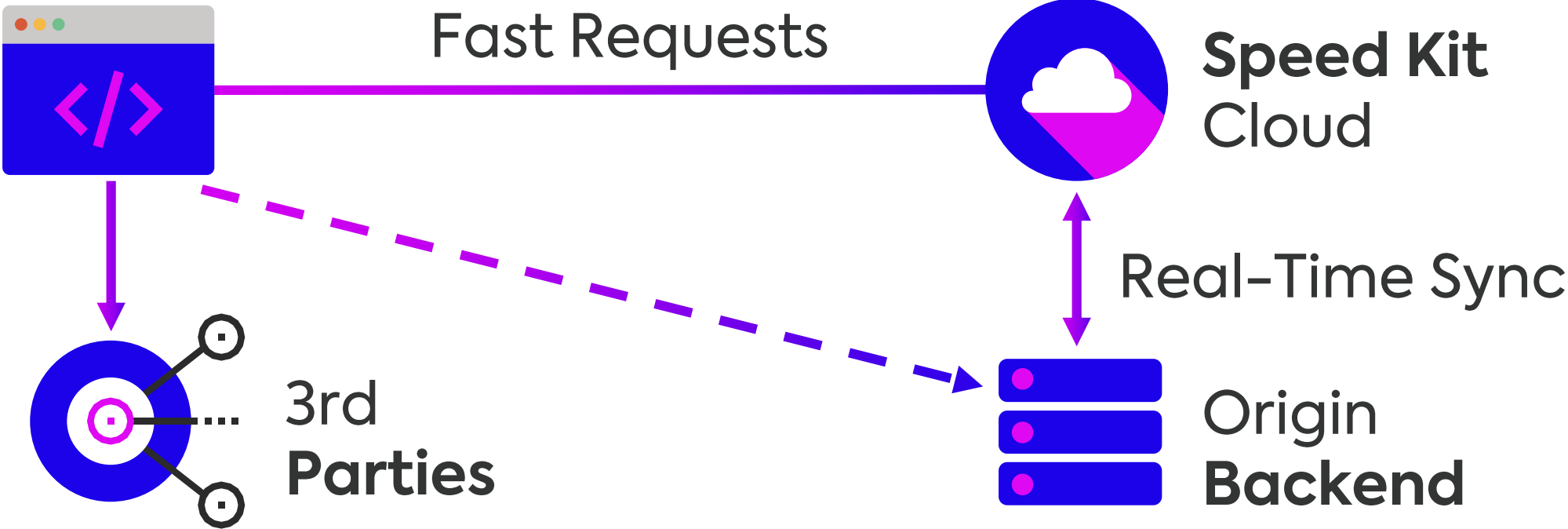


The image features two men in orange hazmat suits and night vision goggles, set against a purple, starry background. The man in the foreground is looking back over his shoulder, while the man behind him looks forward. The text 'We Kill Load Times' is overlaid on the image.

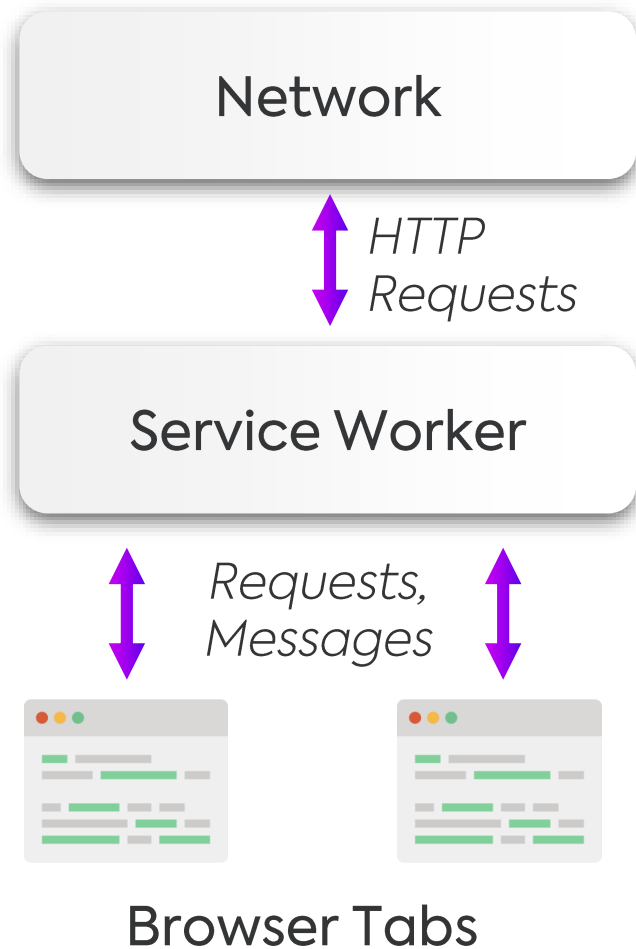
We Kill
Load Times

How **Speed Kit** Works

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

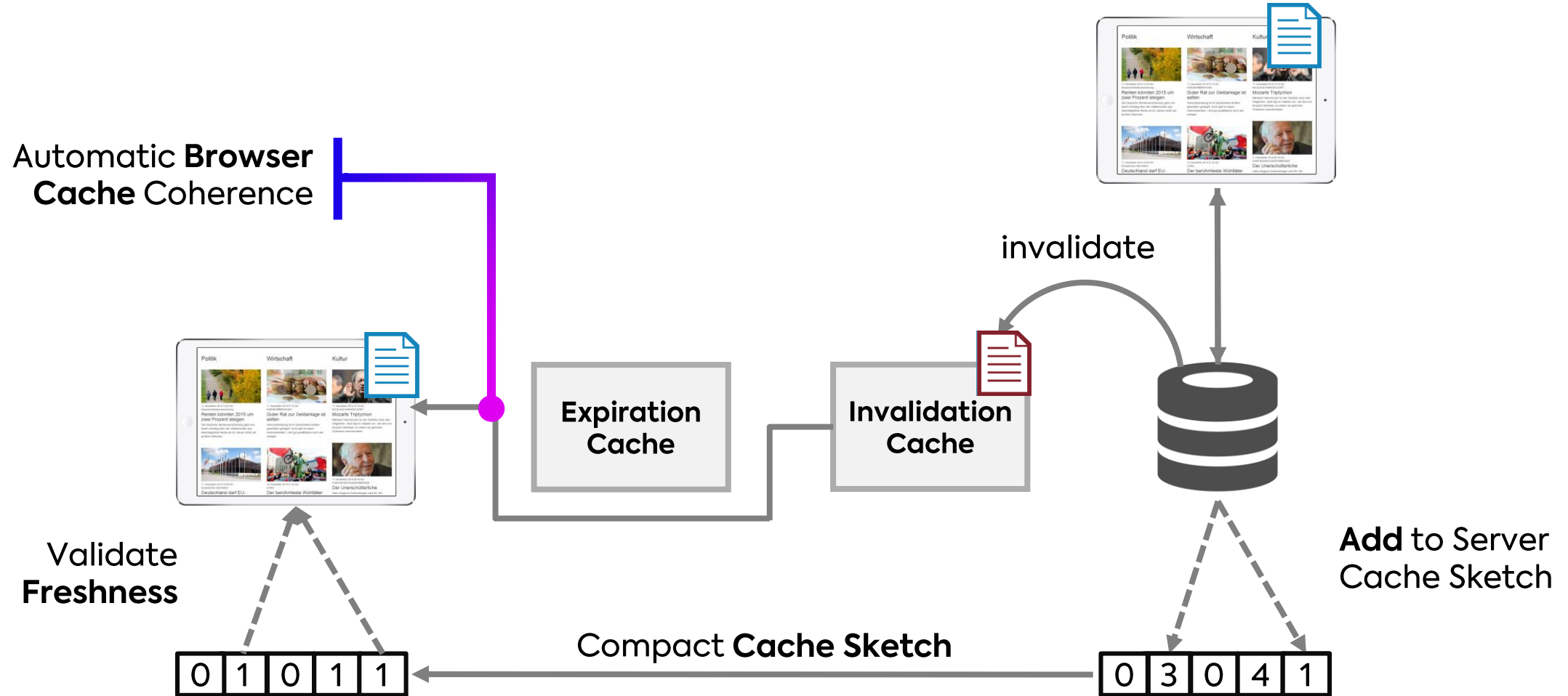


Background: **Service Workers**



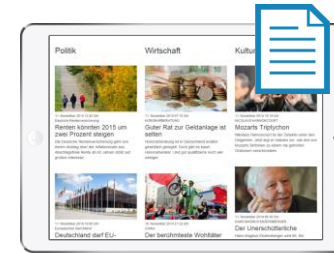
- Programmable **Network Proxy**, running as a **Background Process**, without **DOM Access**.
- Capabilities:
 - Intercept & rewrite **HTTP requests**
 - **Cache** data (CacheStorage)
 - **Store** data (IndexedDB)
 - Respond **offline** or in slow network
 - Sync data & handle push

How We Solved Cache Coherence



How We Solved Cache Coherence

Automatic **Browser**
Cache Coherence



invalidate

False-Positive Rate: $f \approx \left(1 - e^{-\frac{kn}{m}}\right)^k$ Hash-Functions: $k = \left\lceil \ln(2) \cdot \left(\frac{n}{m}\right) \right\rceil$

With 20.000 entries and a 5% false positive rate: **11 Kbyte**

Consistency: Δ -Atomicity, Read-Your-Writes, Monotonic Reads, Monotonic Writes, Causal Consistency



Add to Server
Cache Sketch

Validate
Freshness

Compact **Cache Sketch**

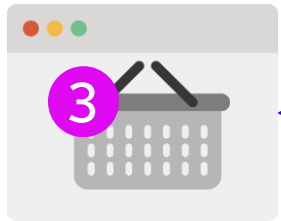
0 1 0 1 1

0 3 0 4 1

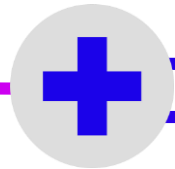


USP: HTML Caching With **Dynamic Blocks**

Browser
(blocks marked by selector)



Replace
Dynamic Blocks

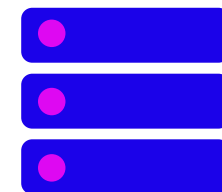


Fast &
Anonymous



**Speed Kit
Cloud**

Slow &
Personalized



**Origin
Server**

In Action: Speed Kit Waterfalls



Built-in cache coherence mechanism

Lower server time (TTFB)

On-the-fly image optimization

Automatic browser + CDN caching

Accelerated personalization

Accelerated 3rd parties

HTTP/2 multiplexing

Network stack tuning: TCP, TLS, IPv6

All resources offline-available

This Talk = Teaser of Ongoing **Study**

[Speedstudy.info](https://speedstudy.info)

Mobile Site Speed and the Impact on E-Commerce

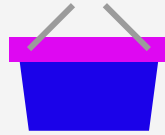


Google

 **Baqend**

 **Universität Hamburg**
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Study Setup



Implementation

A/B-Tested Speedup

- First Paint
- Time to Interaction
- ...

Evaluation

Quantified On-Site Uplift

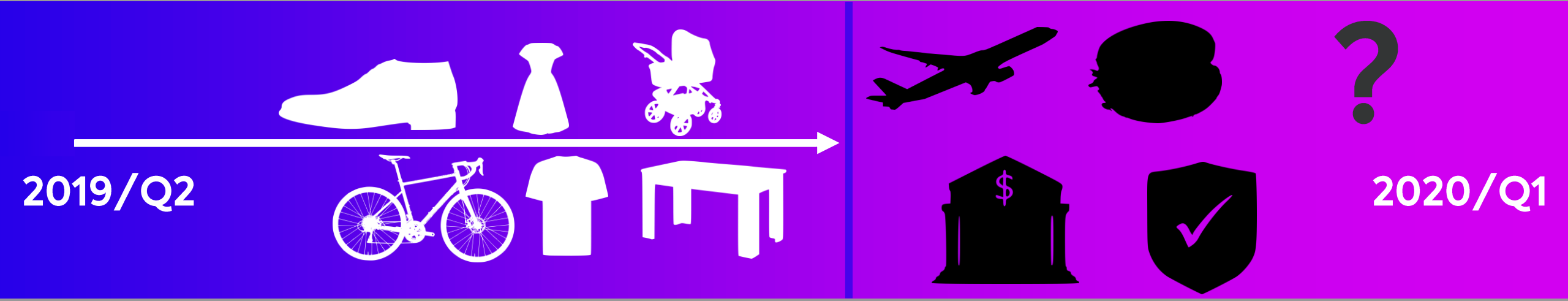
- Conversion Rate
- Time on Site
- ...

Long-Term Effects

Not Evaluated

- SEO Rank
- Returning Visitors
- ...

Study Timeline



The image features three men in white lab coats against a purple background. The man on the left has a beard and glasses, wearing a stethoscope. The man in the middle has glasses and is using a reflex hammer on his own knee. The man on the right has glasses and is pointing upwards with a wide smile. In the bottom right corner, there is a book titled 'Encyclopedia of and Very Websites' with a snail on the cover. The text 'Website is loading' is visible on a laptop screen in the foreground.

How Does Speed **Correlate** To Business Success Exactly?

Overview: Measuring Performance

Synthetic Tests



✓ User-centric metrics

✗ Only simulated

Real-User Monitoring



✓ Data from actual users

✗ Complex to operate

Log Analysis



✓ Readily available

✗ Hard to interpret

Measuring the Uplift – With **SCIENCE**

CDNs, Manual Optimizations



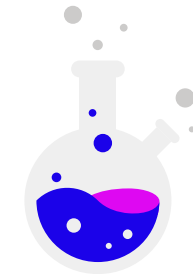
- Only before-after comparison



Speed Kit



Application Features



- Measurable business impact through A/B tests



Measuring the Uplift – With **SCIENCE**

CDNs, Manual Optimizations



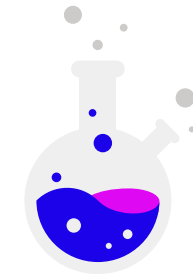
- Only before-after comparison

Speed Kit



- Statistically sound **split testing**
- Clean measurement of performance & business uplifts

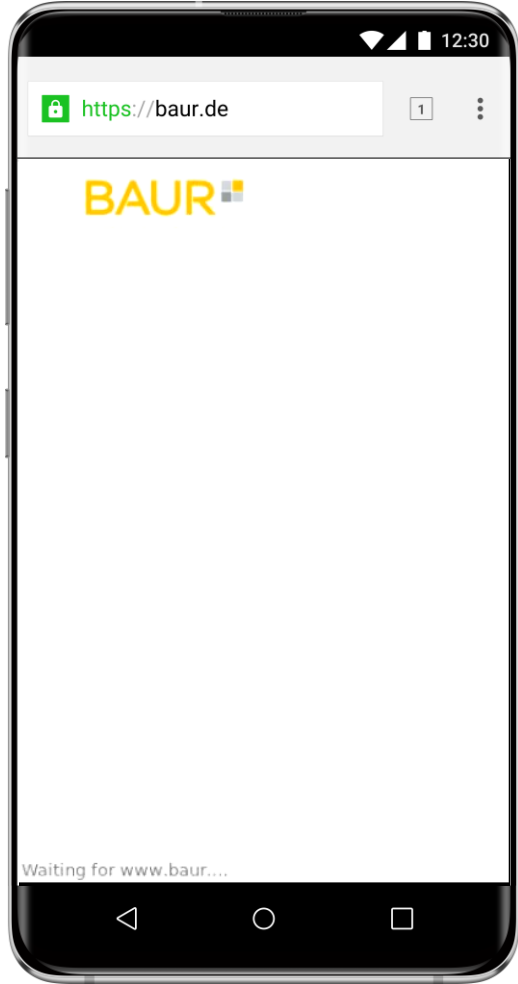
Application Features



- Measurable business impact through A/B tests

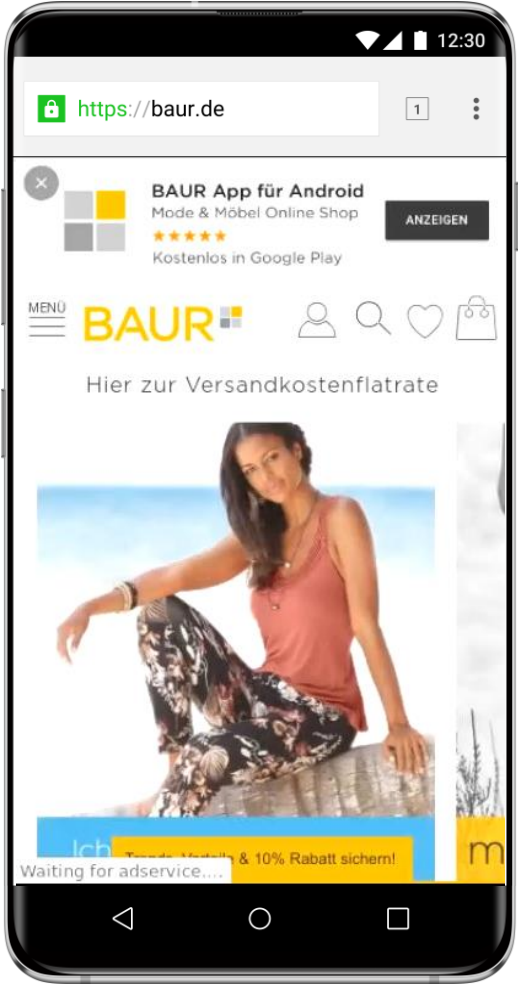
Baur.de

**Before
Speed Kit**

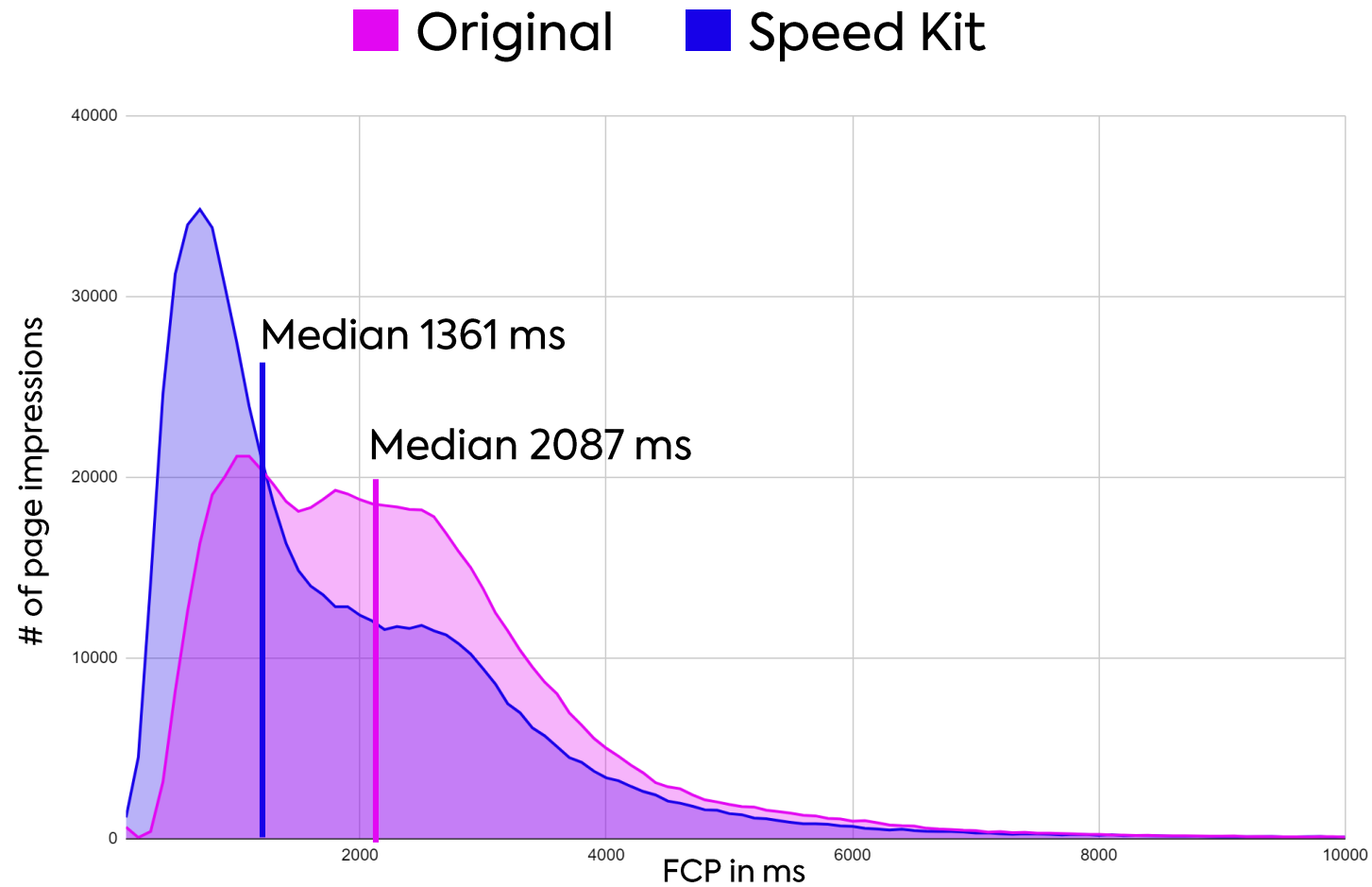


**1.5x
faster**

**After
Speed Kit**



Overall Performance **Uplift** For Baur.de

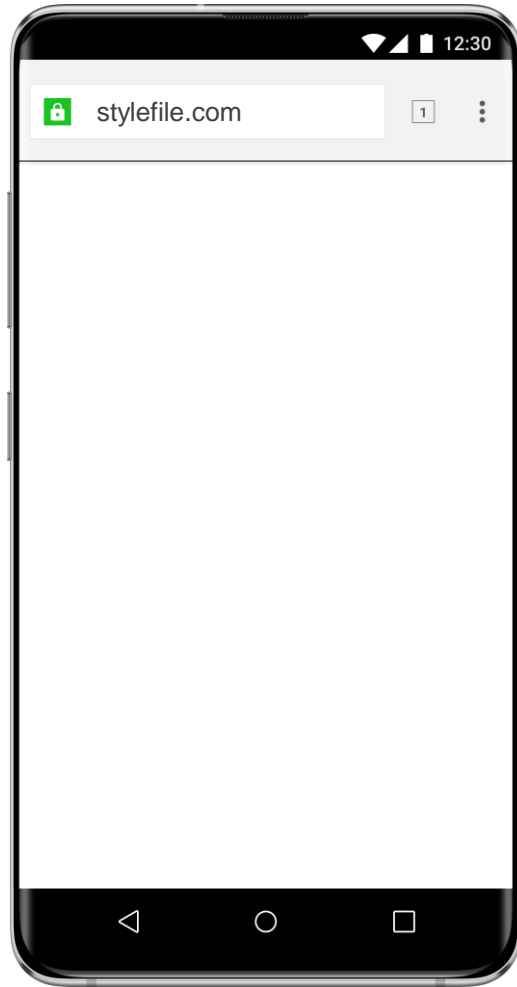


First Contentful Paint Histogram

*Histogram of first contentful paint on PDV pages compared between the two A/B test split groups

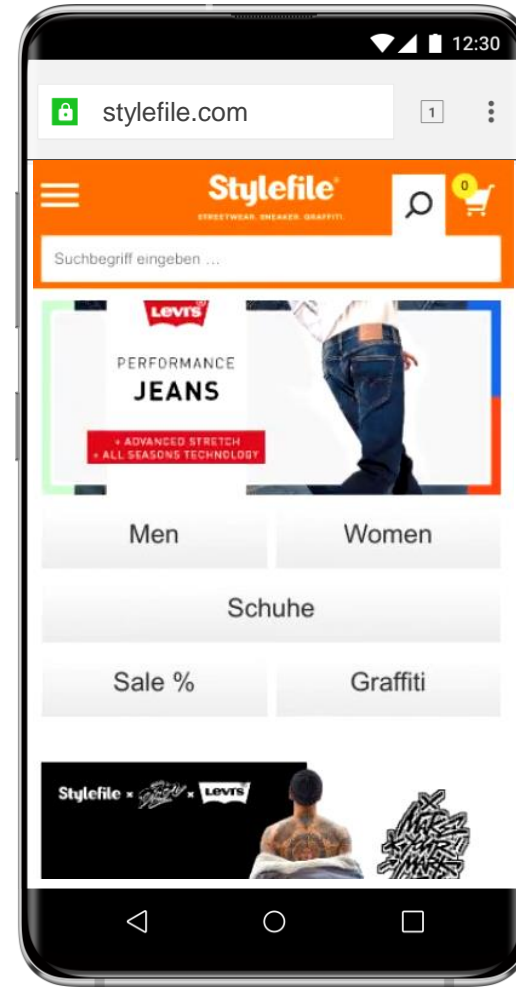
Stylefile.com

Without
Speed Kit



2.1x
faster

With
Speed Kit



Stylefile: **Business** Uplift



User-Based
Conversion Rate

+ 1.9%



Average Order
Value

+ 3.8%

Directions for **Research**

- **Prefetching & Click Prediction**
- **Anomaly detection** on Real-User Data
- Content-based **Staleness Minimization**
- **Dynamic Block Inference**
- Dual & Adaptive **Bloom Filter**
- **Vision-based**
 - Image Optimization
 - Regression Testing
 - Speed Metrics
- Workload-based **TTL Estimation**
- **Connection-Aware Compression & Push**



How Do We Measure Web Performance?

Real-User Monitoring (RUM)

Collection



Tracking (RUM)

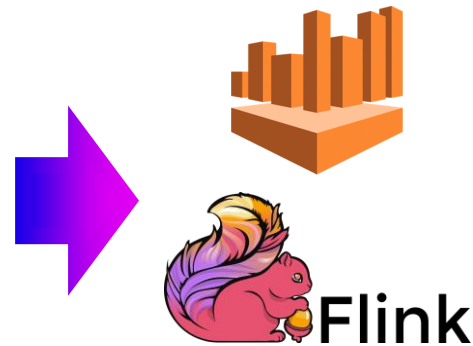


Ingestion



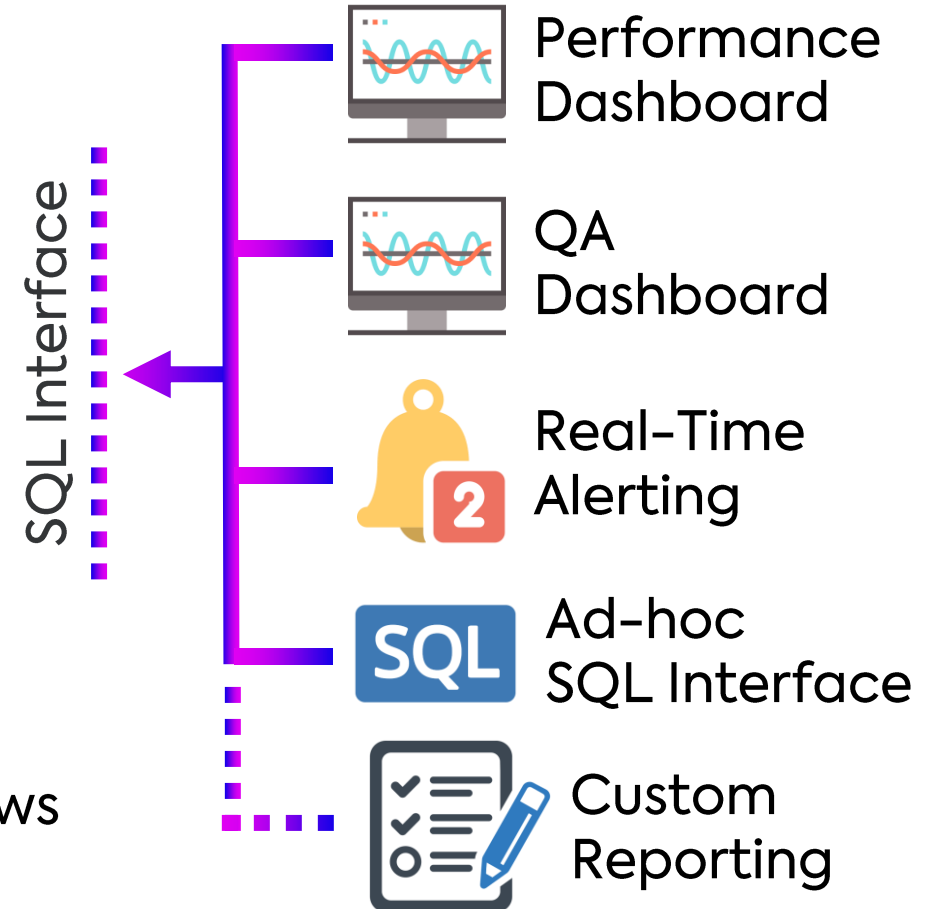
- Raw PI tracking & meta data
- Custom tracking

Analytics

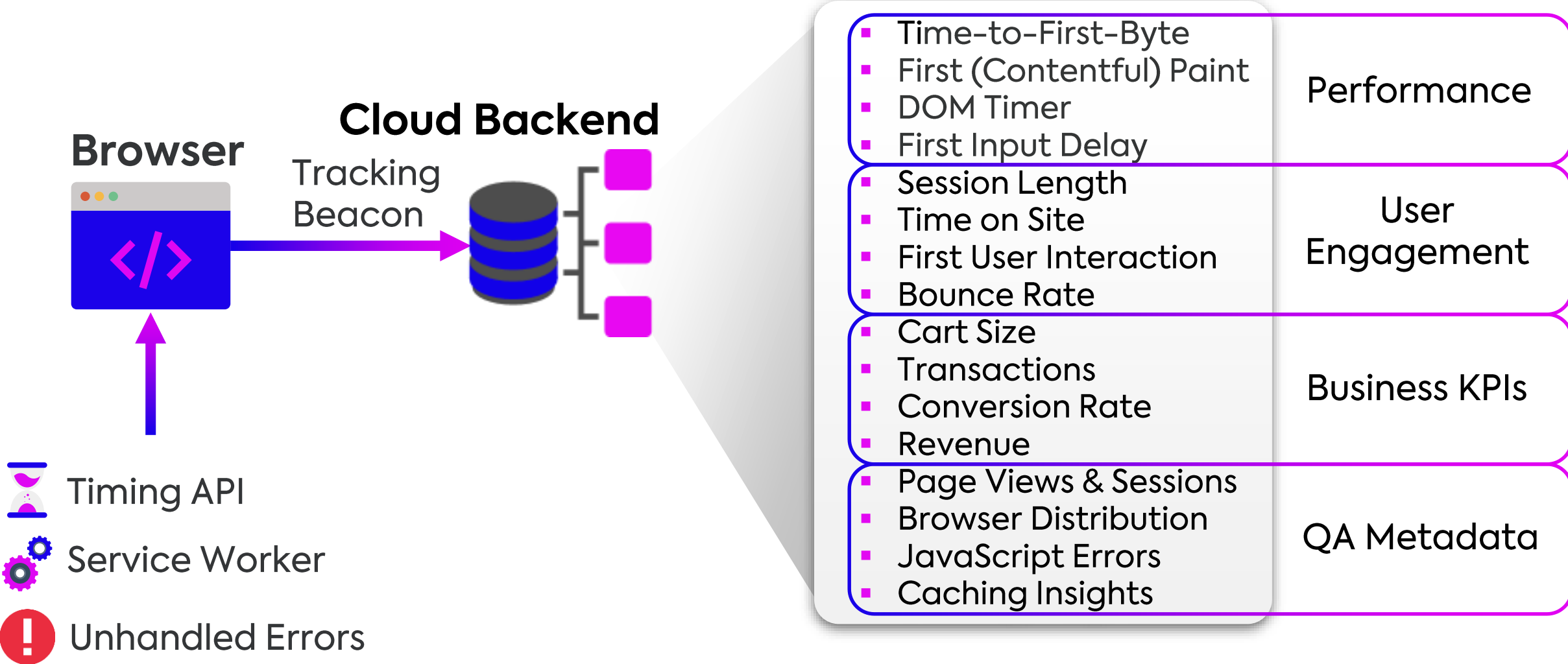


- Materialized views & aggregations
- Historical data

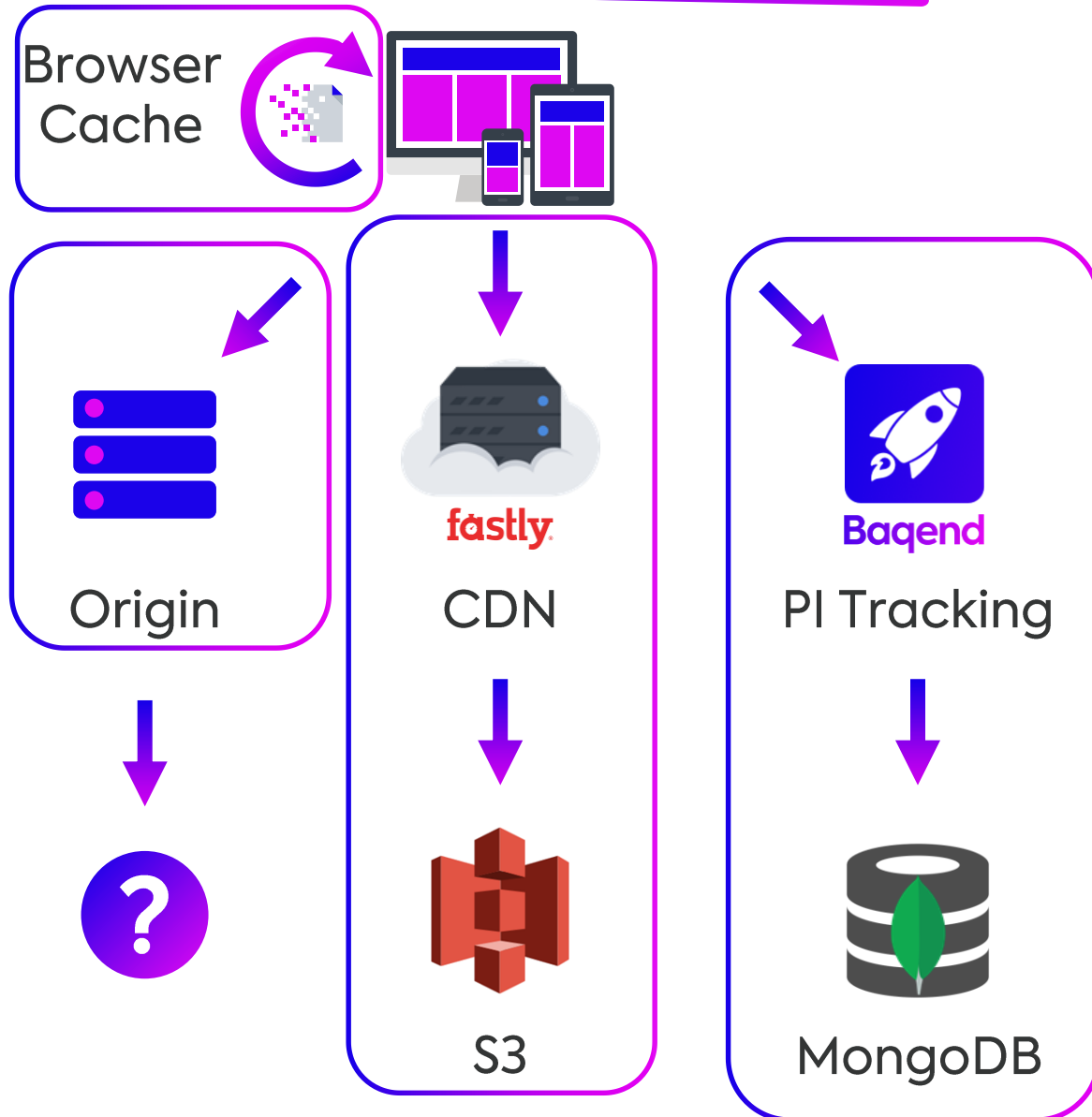
Reporting



Goal: Performance & Business Insights

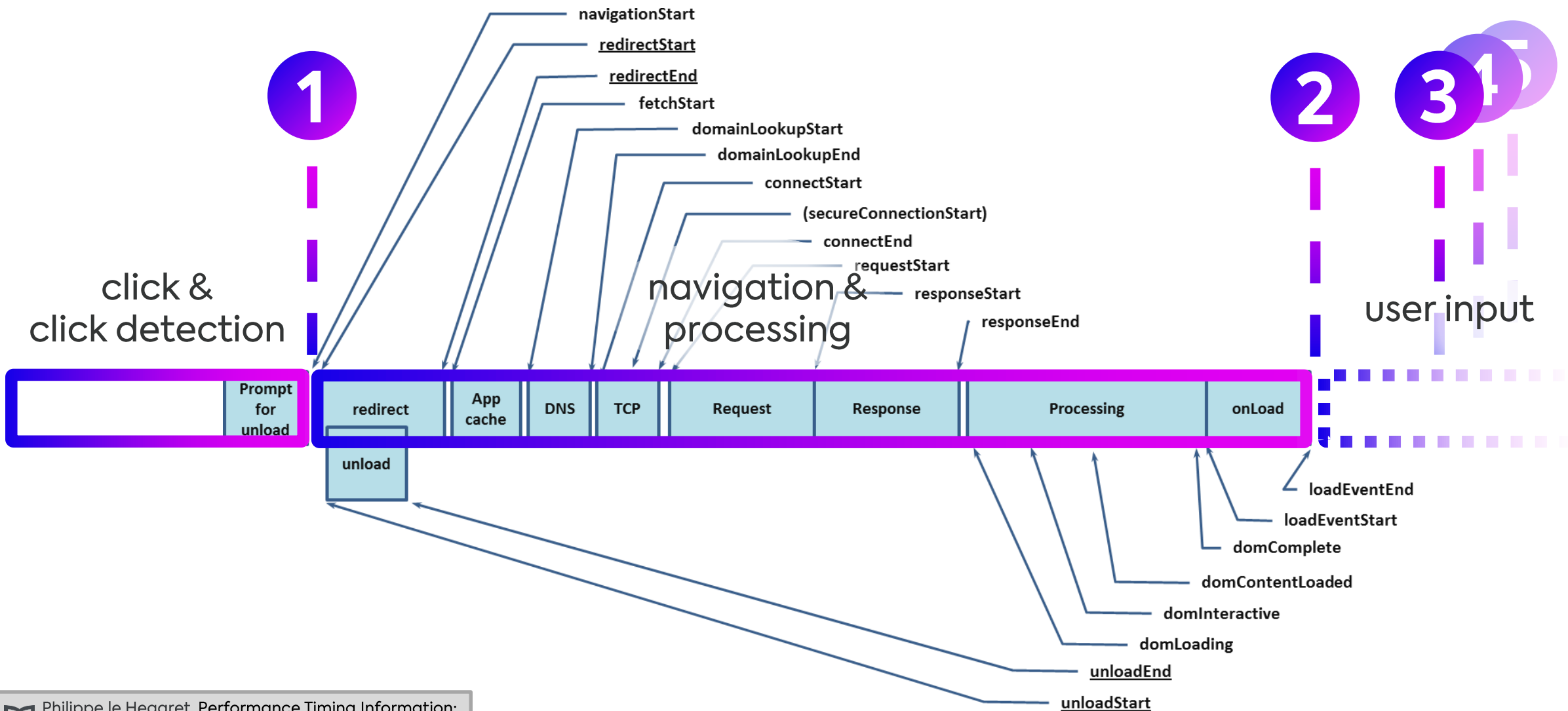


How to **Collect** the Performance Data?

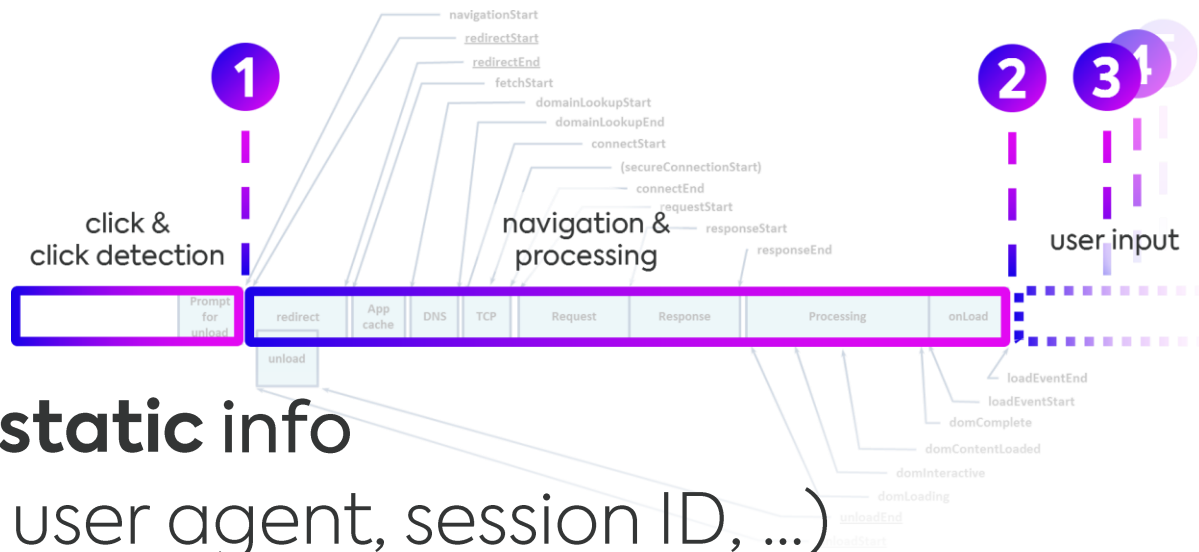


- Logging requests is not enough:
 - ✗ User? Rendering? ...
 - ✗ Browser cache (invisible)
 - ✗ Origin requests (no logs)
 - ✓ CDN requests
- Solution: **Tracking every PI** (page impression)

When to Send Data Beacons?

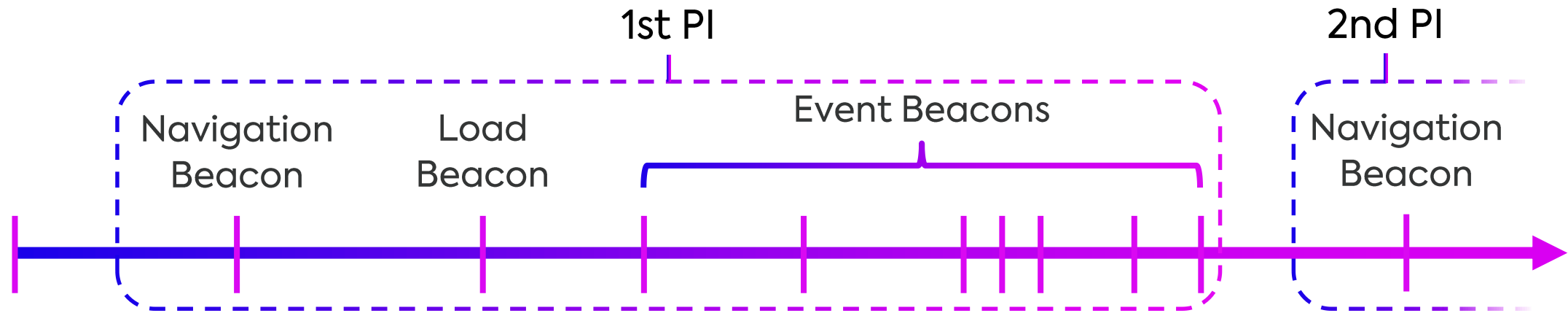


Types of Data Beacons



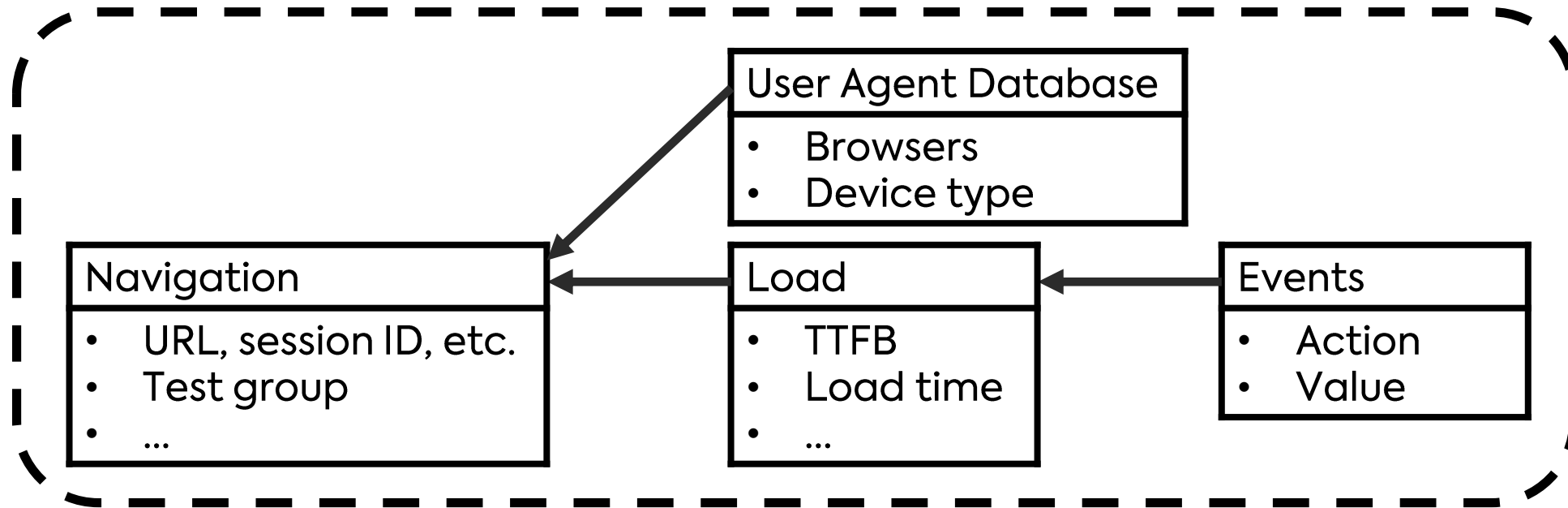
1. 1 for **static** info
(URL, user agent, session ID, ...)
2. 1 for **timings**
(TTFB, load time, FCP, ...)
3. 0-n for **events**
(first input, add-to-cart, ...)

Schema: **PI**



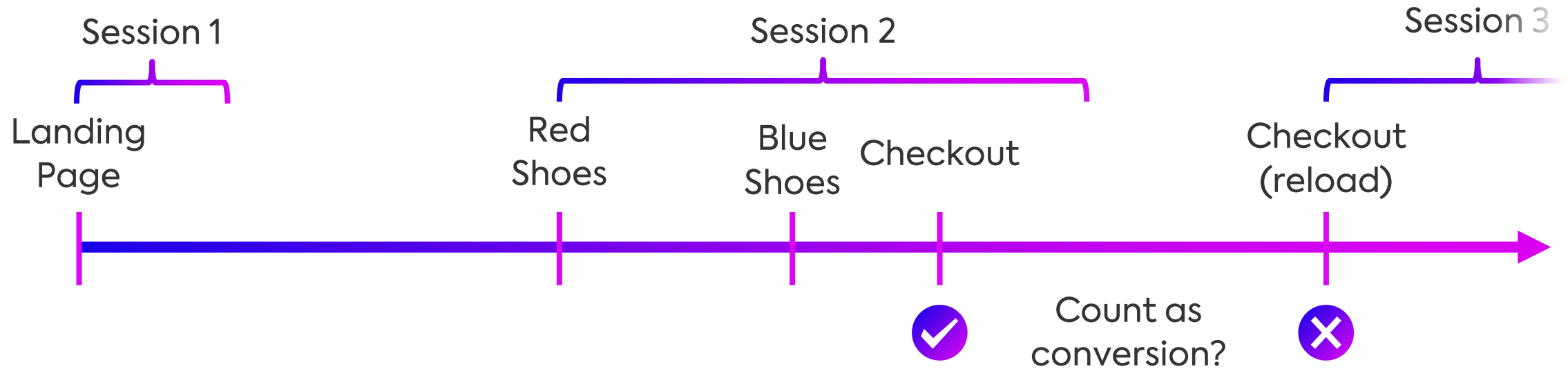
- **Beacon Join → PI:** How do we handle events that come late?
 - Simply wait 5 minutes?
 - Wait for next PI or session timeout?
 - ...?
- How to resolve **user agents**?

Schema: **PI**



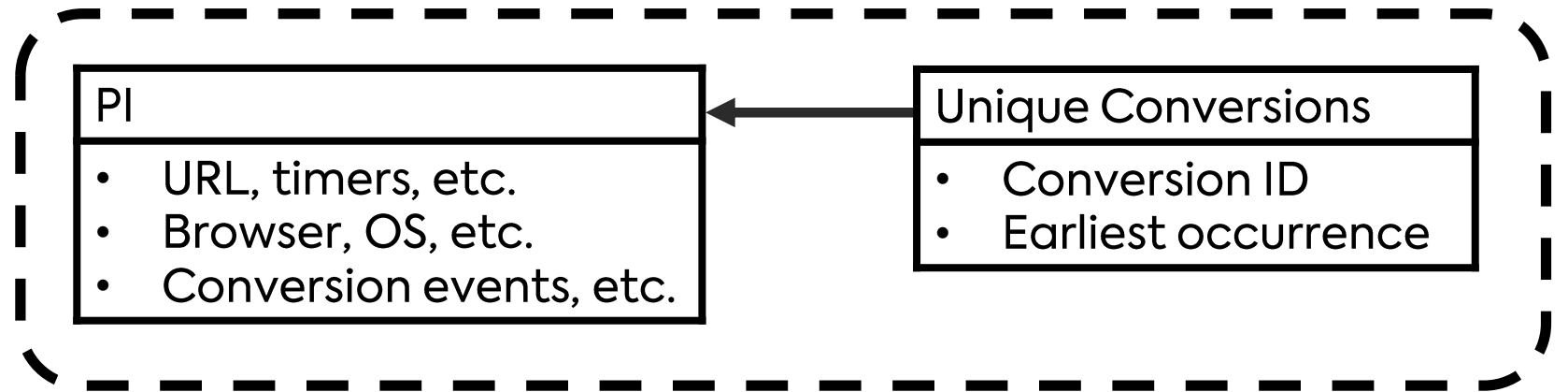
- **Aggregate events:** collect all events per PI
- **Join 3 Collections:** put together PI from navigation/load/event beacons
- **Resolve User Agents:** derive browser, device, etc. from UA string

Schema: Sessions



- **Unique conversions:** remove phantoms
- **Session timeout** after 30 minutes of *inactivity*

Schema: Sessions



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

No Way, MongoDB!



Indexing

Queries over non-indexed attributes were infeasible



Runtime

Even with indexes in place, queries could take 30+ min.



Scalability

Queries got slower with increasing amounts of data



Complexity (Joins!)

MongoDB aggregation pipelines become sophisticated quickly



Fixing My Life With
Flex Tape Athena

The „A“ Stands for „**AWS**ome“

- Desperate **attempt**:
 1. Dump MongoDB collection
 2. Upload to S3
 3. Query with Athena

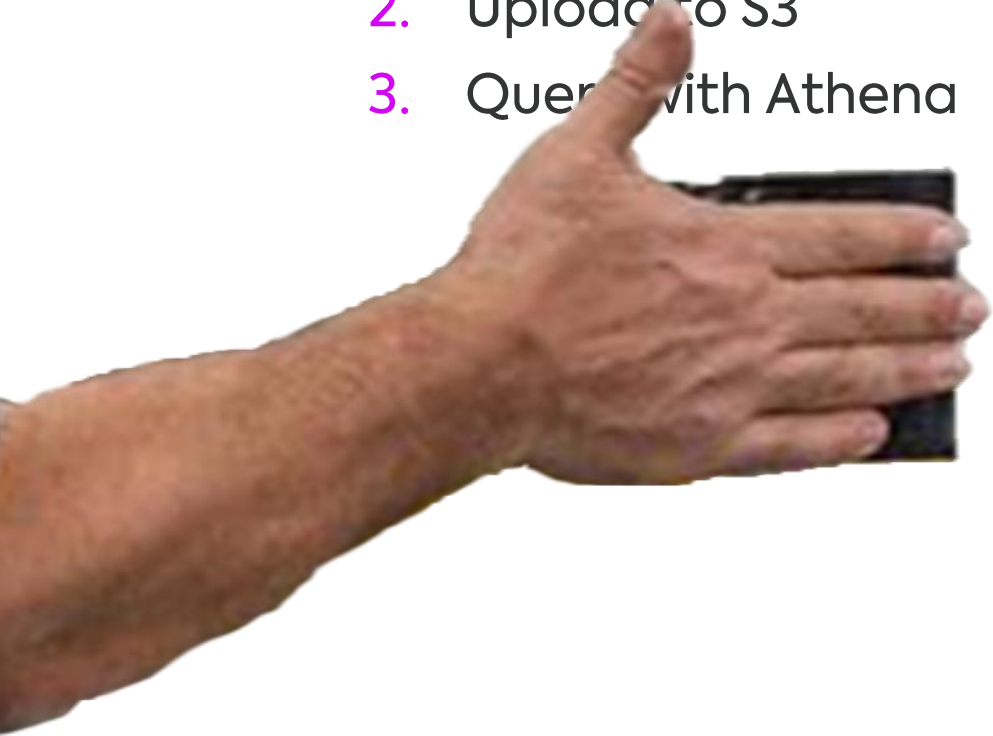


- Typical analysis:
 - 1 equi-join
 - 3 mio. Pls
 - ~15+ min.

The „A“ Stands for „**AWS**ome“

- Desperate **attempt**:

1. Dump MongoDB collection
2. Upload to S3
3. Query with Athena



- Typical analysis:

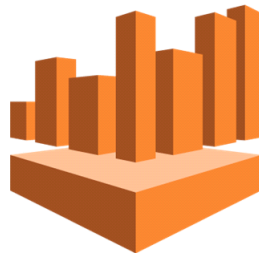
- 1 equi-join
- 3 mio. Pls
- ~15+ min.

The „A“ Stands for „**AWS**ome“



~~Desperate attempt~~: New best practice:

1. Dump MongoDB collection
2. Upload to S3
3. Query with Athena



AWS
Athena

- Typical analysis:
 - 1 equi-join
 - 3 mio. Pls
 - ~10 seconds

What's an Athena?



Athena

=

presto



- **Managed Presto:**

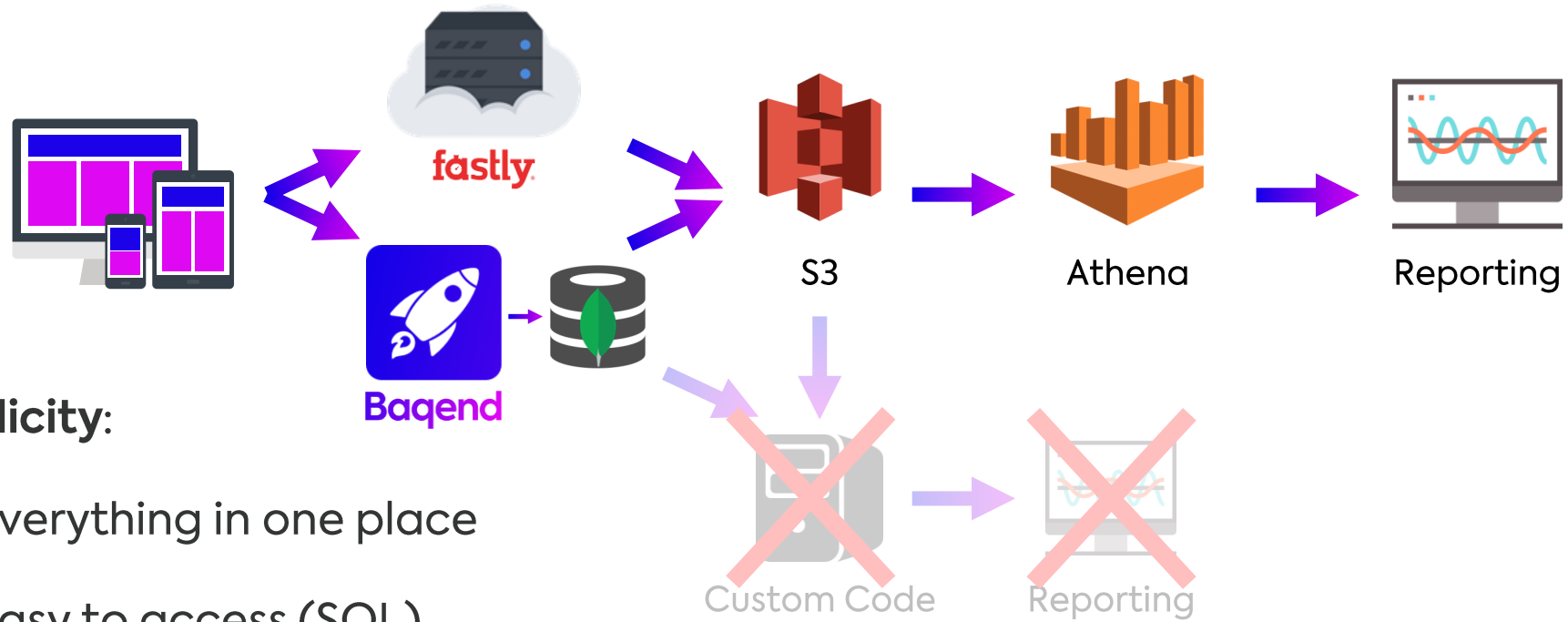
- Interactive analytics with SQL
- Heterogeneous datastores
- Petabyte-scale (Facebook)

- **Pricing by scanned data volume:**

- Efficient storage formats!
- Partitioning or clustering!
- Careful query design!



Upgrading Our **ETL** Pipeline



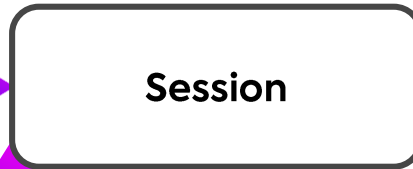
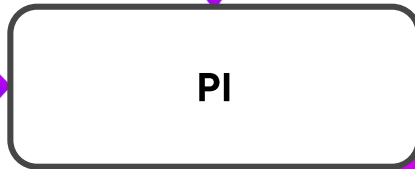
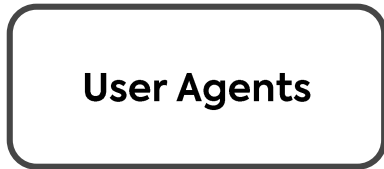
- **Simplicity:**
 - Everything in one place
 - Easy to access (SQL)
- **Scalability & efficiency:**
 - Hundreds of gigabytes scanned in a query
 - Response time on the order of seconds

Schema Overview

Tracking



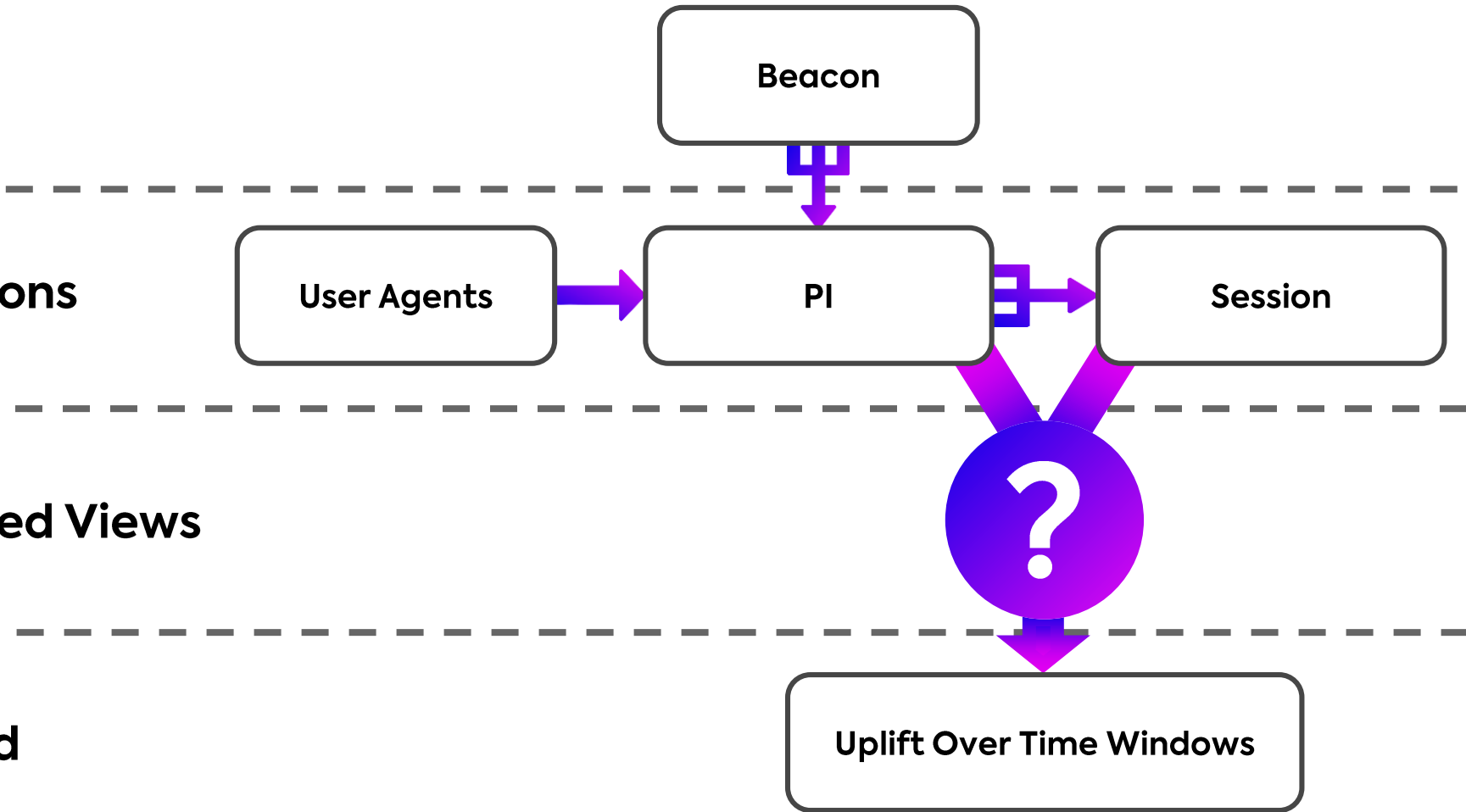
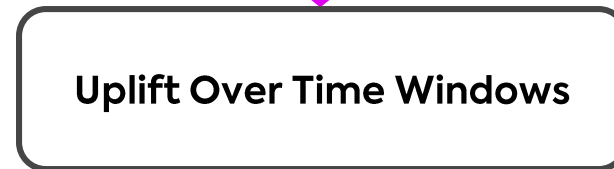
Aggregations



Materialized Views



Dashboard

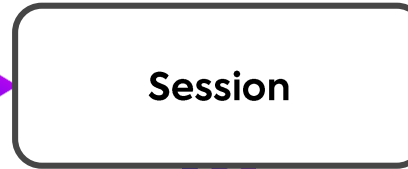
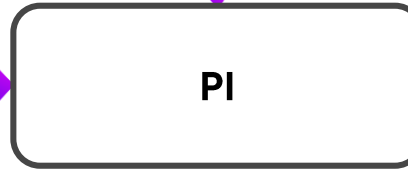
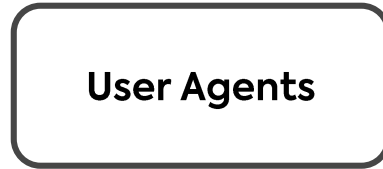


Schema Overview

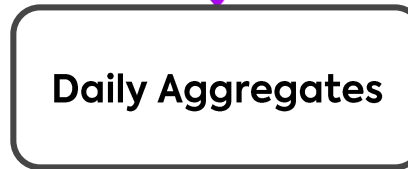
Tracking



Aggregations



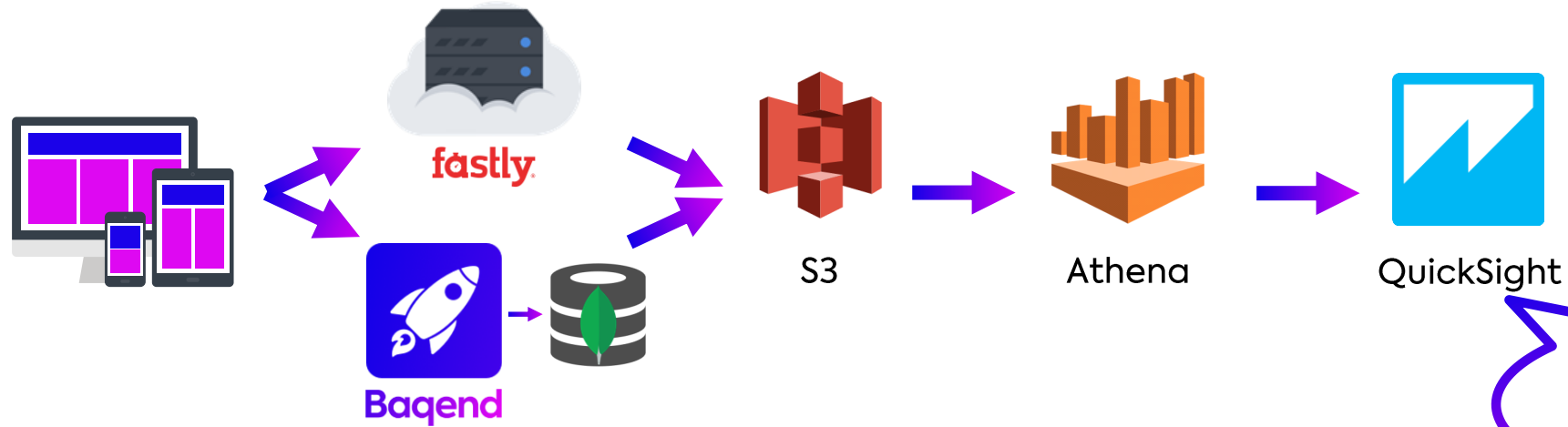
Materialized Views



Dashboard



Our **Batch Analytics** Tech Stack

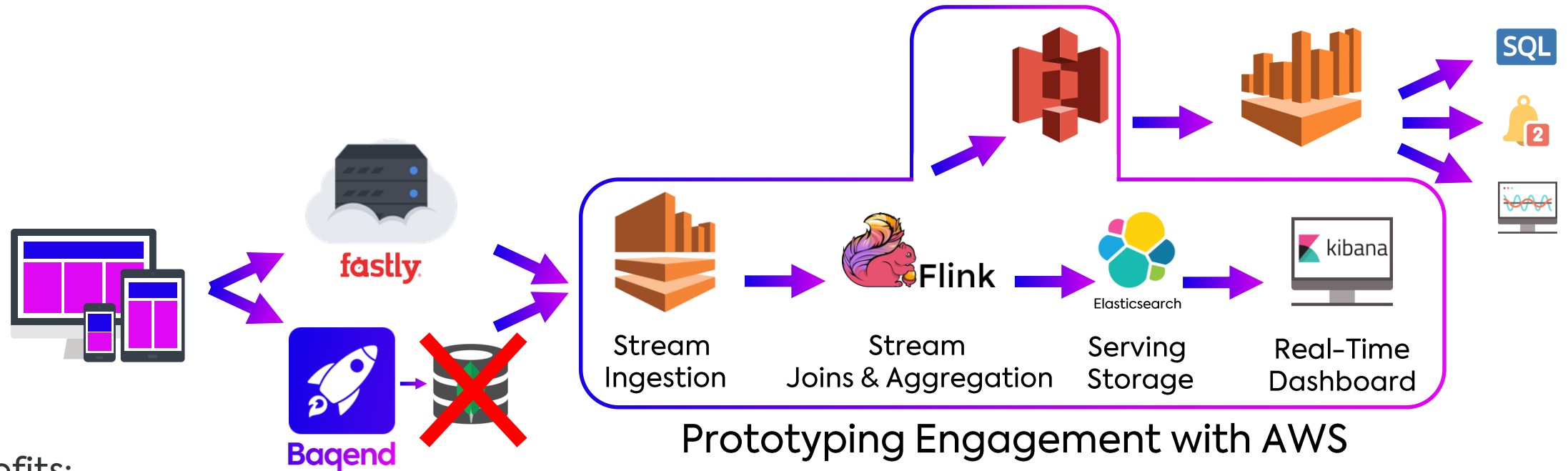


Don't get me started ..

Issues:

- ✘ Many joins → slow queries
- ✘ 90 minutes discovery time
- ✘ No continuous dashboard (daily materialization)

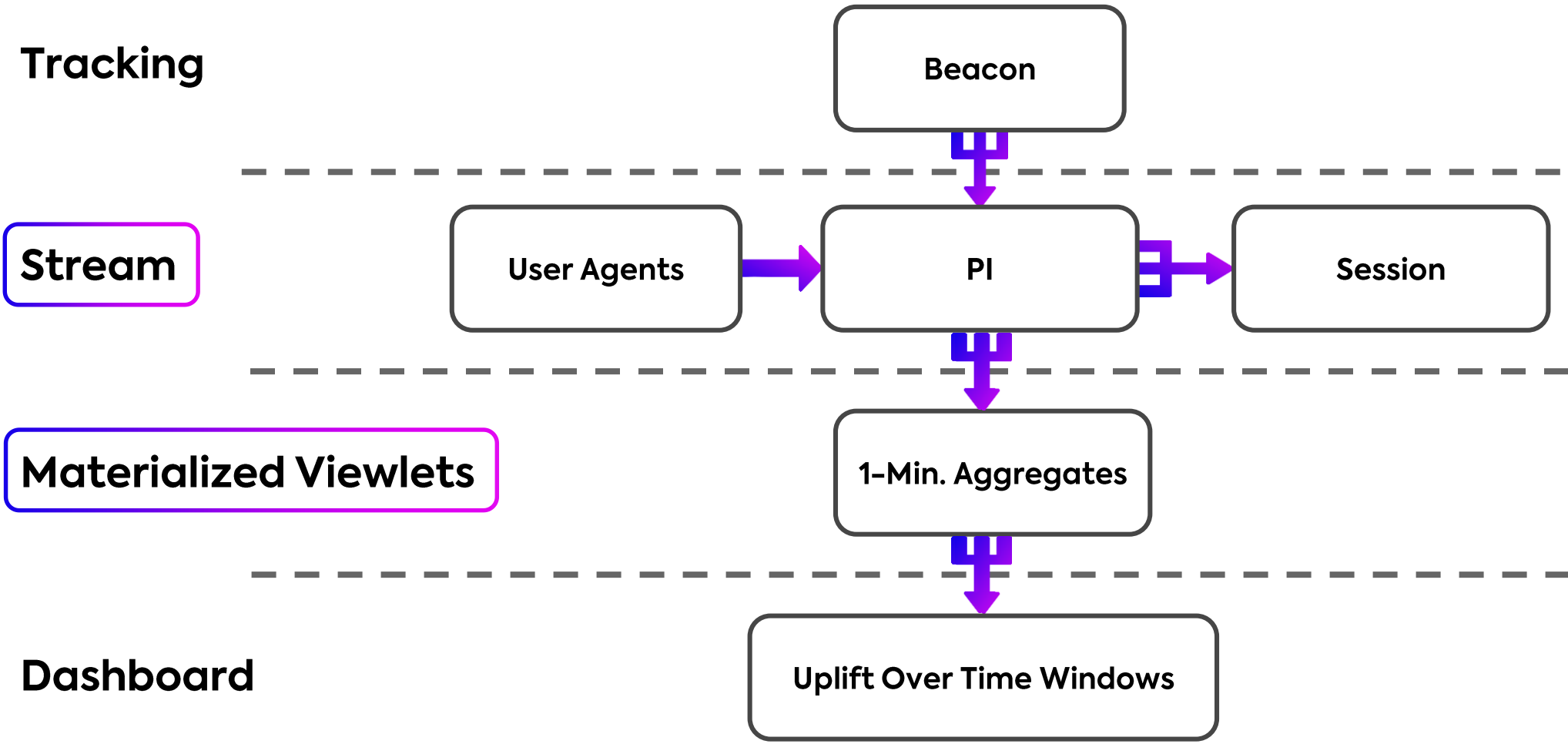
2020 Real-Time Analytics Tech Stack



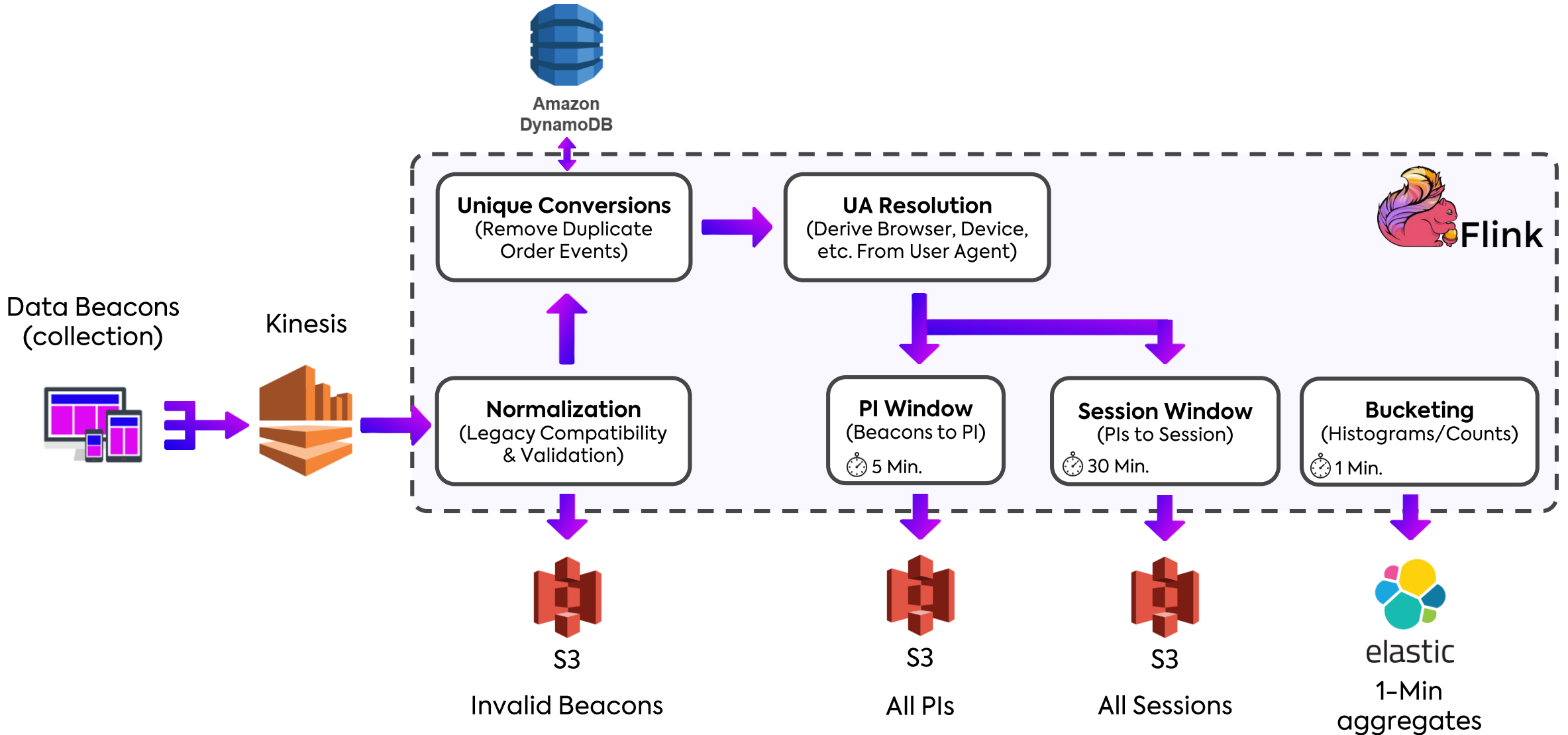
Benefits:

- ✓ No legacy tech → stability & efficiency
- ✓ Faster ingestion → Live performance charts
- ✓ Fewer joins → faster analytics

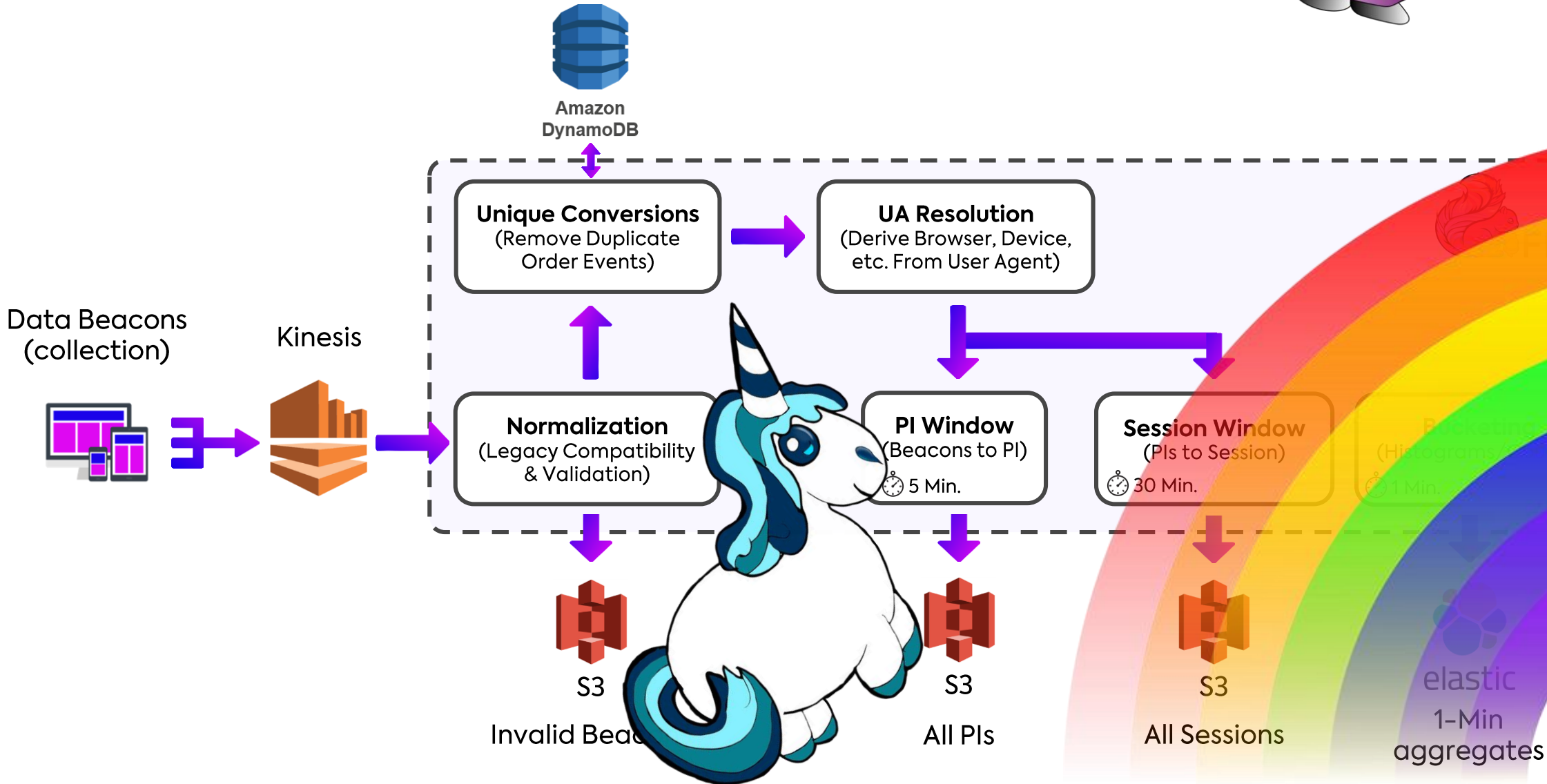
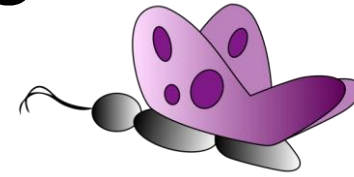
Shiny & NewOld & Lame Schema



Zero-Latency Analytics



Zero-Latency Analytics





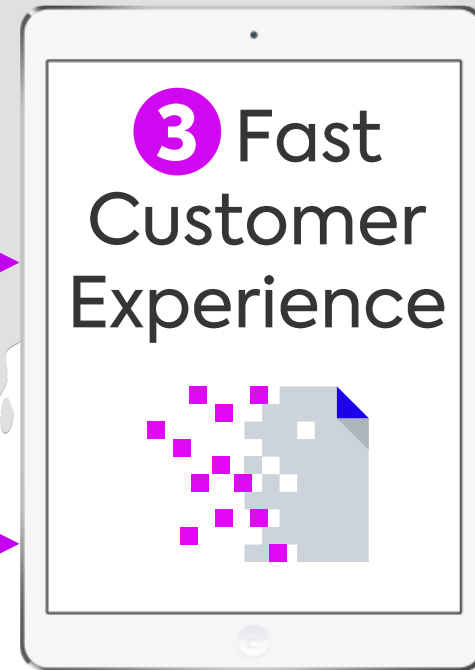
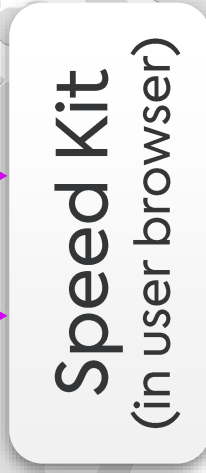
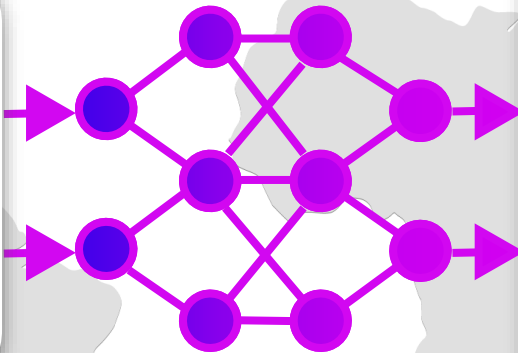
Summary & Outlook

Speed Kit Optimizes **End-To-End**

1 Offloaded Servers

2 Low Latency

3 Fast Customer Experience



Split Testing for Web Performance

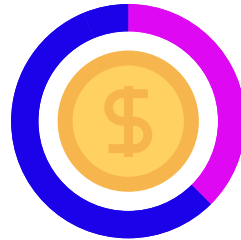
Speed Kit Users



Tracking



vs.



Tracking



Normal Users



- Speed Kit enabled

- **Measurable uplift:**
 - + Performance
 - + User engagement
 - + Business success

- Speed Kit disabled
(no acceleration)

Join & Learn More: speedstudy.info

THE LARGEST SYSTEMATIC STUDY OF

Mobile Site Speed and the Impact on E-Commerce

Your Email

SUBSCRIBE FOR UPDATES

SUBSCRIBE & PARTICIPATE



Google

 Baqend

 Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

The background image shows two men sitting in folding chairs in a desert landscape. They are wearing white protective suits and head-mounted displays (HMDs). The man on the left is looking towards the camera, while the man on the right is looking slightly away. The scene is set in a vast, open desert with a clear sky.

Ready to Load **Instantly?**

Join the study!

Details & newsletter on

speedstudy.info

Felix Gessert
Wolfram Wingerath

felix.gessert@baqend.com
wolle@baqend.com