Batching Was Yesterday

Real-Time Tracking & Analysis
For 100+ Million Visitors

Techcamp 2022
Felix Gessert, Wolfram Wingerath
September 28, 2022
Who We Are

First half of the talk

Felix
CEO

Wolle
Data Engineering & Research

Second half of the talk

Research:
• Web Caching
• Stream Processing
• NoSQL & Cloud Systems
• ...

Practice:
Website Acceleration
Real-User Monitoring
Continuous Analytics
...

Baqend

Speed Kit
Table of Contents

**Why** should you care about website performance?

**What** does tracking data tell you about it?

**How** do you build a scalable analytics stack?

**When** can you see the results?
Why Do Businesses Care About Performance?
Why Do Businesses Care About Performance?

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
Page Speed

100 ms faster

Money

100 ms faster

+0.7% Revenue Per Session

+1% Revenue

You Heard the Stories

Why Do Businesses Care About Performance?

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
Why Do Businesses Care About Performance?

3 things make your website slow

- **Browser rendering**: Page size and weight, image scaling, critical rendering path, ...
- **Network delays**: Network latency, access operations, overhead, protocols, ...
- **Backend processing**: Shop system, CMS, database, image servers, microservices, ...

Batching Was Yesterday: Real-Time Tracking & Analysis For 100+ Million Visitors
Why Do Businesses Care About Performance?

What makes your website slow

1. Request page
2. Process and respond
3. Load dependencies
4. Execute and render

Browser

Network

Backend
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
We Kill Load Times
How Speed Kit Works

Speed Kit Makes Websites Fast!

Browser

Speed Kit service worker

Network

Static + dynamic content

Speed Kit service

Automatic sync

Backend

No changes to infrastructure

Personalized + blacklisted content

Batching Was Yesterday: Real-Time Tracking & Analysis For 100+ Million Visitors
How Speed Kit Works

Personalization with dynamic blocks

Browser

Network

Backend

Anonymous page with dynamic blocks

Speed Kit service

Fill dynamic blocks with personalized content
Measuring the Uplift – With SCIENCE!

CDNs, Manual Optimizations

- Only before-after comparison

Speed Kit

Application Features

- Measurable business impact through A/B tests

How Speed Kit Works
Measuring the Uplift – With SCIENCE!

How Speed Kit Works

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
How Do You Measure Web Performance?
The Basic Idea

- **Real-User Monitoring (RUM)**
- **Batching Was Yesterday: Real-Time Tracking & Analysis For 100+ Million Visitors**

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

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

- **QA Metadata**
  - Page Views & Sessions
  - Browser Distribution
  - JavaScript Errors
  - Caching Insights

- **Browser**
  - Tracking Beacon
  - Timing API
  - Service Worker
  - Unhandled Errors

- **Cloud Backend**
Baqend’s RUM Pipeline

Collection

Ingestion

Analytics

Reporting

- Raw PI tracking & meta data
- Custom tracking

- Materialized views & aggregations
- Historical data

SQL Interface

Performance Dashboard

QA Dashboard

Real-Time Alerting

Ad-hoc SQL Interface

Custom Reporting
Real-User Monitoring (RUM)

Baqend’s RUM Pipeline

Collection

Ingestion

Analytics

Reporting

- Raw PI tracking & meta data
- Custom tracking

- Materialized views & aggregations
- Historical data

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

Batching Was Yesterday: Real-Time Tracking & Analysis For 100+ Million Visitors
When to Send Data Beacons?

1. Click & click detection
2. Navigation & processing
3. User input


Batching Was Yesterday: Real-Time Tracking & Analysis For 100+ Million Visitors
Real-User Monitoring (RUM)

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: Page Impression (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: Sessions

- **Bounces & Session End**: find out when and where people leave
- **Session timeout** after 30 minutes of *inactivity*
So What’s the Problem?
Speaker Shuffle

Felix
Speaker Shuffle

Wolle
The Challenge

Schema Overview

Tracking

Aggregations

Materialized Views

Dashboard

Beacon

User Agents

PI

Session

Analyses Over Time Windows
Schema Overview

The Challenge

Batching Was Yesterday: Real-Time Tracking & Analysis For 100+ Million Visitors
The Challenge

Our Batch Analytics Stack

Issues:

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

Processing Stages & Latency

Alerting

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

Processing Time

Trend Analysis

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

Batching Was Yesterday: Real-Time Tracking & Analysis For 100+ Million Visitors
There Must Be a Smarter Way!
Continuous Data Analytics

Early 2020: AWS Prototyping

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

Batching Was Yesterday: Real-Time Tracking & Analysis For 100+ Million Visitors

Continuous Data Analytics

Shiny & New Schema

Tracking

Aggregations

Materialized View

Dashboard

Batching Was Yesterday: Real-Time Tracking & Analysis For 100+ Million Visitors
Shiny & New Schema

Tracking

Stream

1-Min. Summaries

Beacon

User Agents

PI

Session

1-Min. Aggregates

Analyses Over Time Windows

Continuous Data Analytics

Batching Was Yesterday: Real-Time Tracking & Analysis For 100+ Million Visitors
## 3 Levels of Aggregation

### Partial Page Impressions (PPIs)
Enhanced Data Beacons

<table>
<thead>
<tr>
<th>Time</th>
<th>Browser</th>
<th>Device</th>
<th>Test Group</th>
<th>First Contentful Paint (FCP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:05:04.578</td>
<td>Firefox</td>
<td>Mobile</td>
<td>Speed Kit</td>
<td>127ms</td>
</tr>
<tr>
<td>11:06:48.139</td>
<td>Chrome</td>
<td>Mobile</td>
<td>Original</td>
<td>958ms</td>
</tr>
</tbody>
</table>

### 1-Min. Time Windows
Immediate Aggregates (Storage)

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

### Arbitrary Time Windows
Real-Time Reporting (Dashboard Queries)

<table>
<thead>
<tr>
<th>Time</th>
<th>Browser</th>
<th>Device</th>
<th>Test Group</th>
<th>First Contentful Paint (FCP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:05</td>
<td>Firefox</td>
<td>Mobile</td>
<td>Speed Kit</td>
<td>{200ms: 4, 500ms: 2}</td>
</tr>
</tbody>
</table>

**Speed Kit**

Batching Was Yesterday: Real-Time Tracking & Analysis For 100+ Million Visitors
Continuous Data Analytics

Visualizing Uplift in a Dashboard

Continuous

Batching Was Yesterday: Real-Time Tracking & Analysis For 100+ Million Visitors

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Speed Kit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desktop</td>
<td>1599</td>
<td>1140</td>
</tr>
<tr>
<td>Mobile</td>
<td>1718</td>
<td>1404</td>
</tr>
<tr>
<td>Overall</td>
<td>1663</td>
<td>1299</td>
</tr>
</tbody>
</table>

- Good (≤ 1.8s)
- Needs improvement
- Poor (> 3.0s)

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Speed Kit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desktop</td>
<td>54%</td>
<td>63%</td>
</tr>
<tr>
<td>Mobile</td>
<td>24%</td>
<td>18%</td>
</tr>
<tr>
<td>Overall</td>
<td>22%</td>
<td>19%</td>
</tr>
</tbody>
</table>

- More fast loads
- More slow loads

Speed Kit

Batching Was Yesterday: Real-Time Tracking & Analysis For 100+ Million Visitors

35
Zero-Latency Analytics

Core Components:
- Data Beacons (collection)
- Kinesis
- Amazon DynamoDB
- Flink

Processes:
- Bot Detection (Identify Suspicious User Agents)
- UA Resolution (Derive Browser, Device, etc. From User Agent)
- Normalization (Legacy Compatibility & Validation)
- PI Window (Beacons to PI)
- Session Window (PIs to Session)
- Bucketing (Histograms/Counts)

Outputs:
- Invalid Beacons
- All PIs
- All Sessions
- 1-Min aggregates

Batching Was Yesterday: Real-Time Tracking & Analysis For 100+ Million Visitors
Zero-Latency Analytics

Continuous Data Analytics

Batching Was Yesterday: Real-Time Tracking & Analysis For 100+ Million Visitors
Summary & Outlook
Speed Kit: Web Performance Plugin

1. Offloaded Servers
2. Low Latency
3. Fast Customer Experience

Shop Backend (unmodified infrastructure)

Speed Kit (in user browser)


Batching Was Yesterday: Real-Time Tracking & Analysis For 100+ Million Visitors
Split Testing for Web Performance

Summary & Outlook

Speed Kit Users vs. Normal Users

- Speed Kit enabled
- **Measurable uplift:**
  - Performance
  - User engagement
  - ...

- Speed Kit disabled (no acceleration)

Summary & Outlook

Learn More: speedhub.org

Faster Page Speed
Bounce Rate

Learn how faster page speed can impact the key metrics of your business.

---

Impact

Studies

100 ms faster page speed decreased bounce rate by 5%
www.akamai.com

---

Powerful Apps Fueled by the Web
Learn how developers are engaging an expanding Chrome OS audience through progressive web apps.

Vodafone Increased Sales by 8%
An A/B test focused on optimizing Web Vitals showed that a 37% improvement in LCP led to 8% more sales.
Join Baqend’s Mission of Building a Faster Web

Baqend is Hiring

Product Integration
Team Lead, Seniors & Juniors

Performance Engineering
Seniors & Juniors

Fullstack Development
Seniors & Juniors

Data Engineering
Seniors & Juniors

www.speedkit.com/careers
Summary & Outlook

Thanks!

Questions?

research@baqend.com
baqend.com/publications