

A Digital Performance Monitoring Overview for SRE/DevOps Teams

MAIN CONCEPTS, DRIVERS AND SOLUTIONS

A few words about Boris Rogier

- Network and Application Performance Monitoring veteran
- Founded and managed Performance Vision (NAPM software vendor) from 2004 until 2020
- Focusing efforts on inventing new ways to monitor experience and client side infrastructure performance at the cloud age
- <https://www.linkedin.com/in/borisrogier/>

Objectives

- Share high level knowledge of digital performance drivers for internet facing applications
- Review possibilities for performance optimizations
- Understand ways to monitor them
- Best practices summary
- Get your feedback on concept we are working on

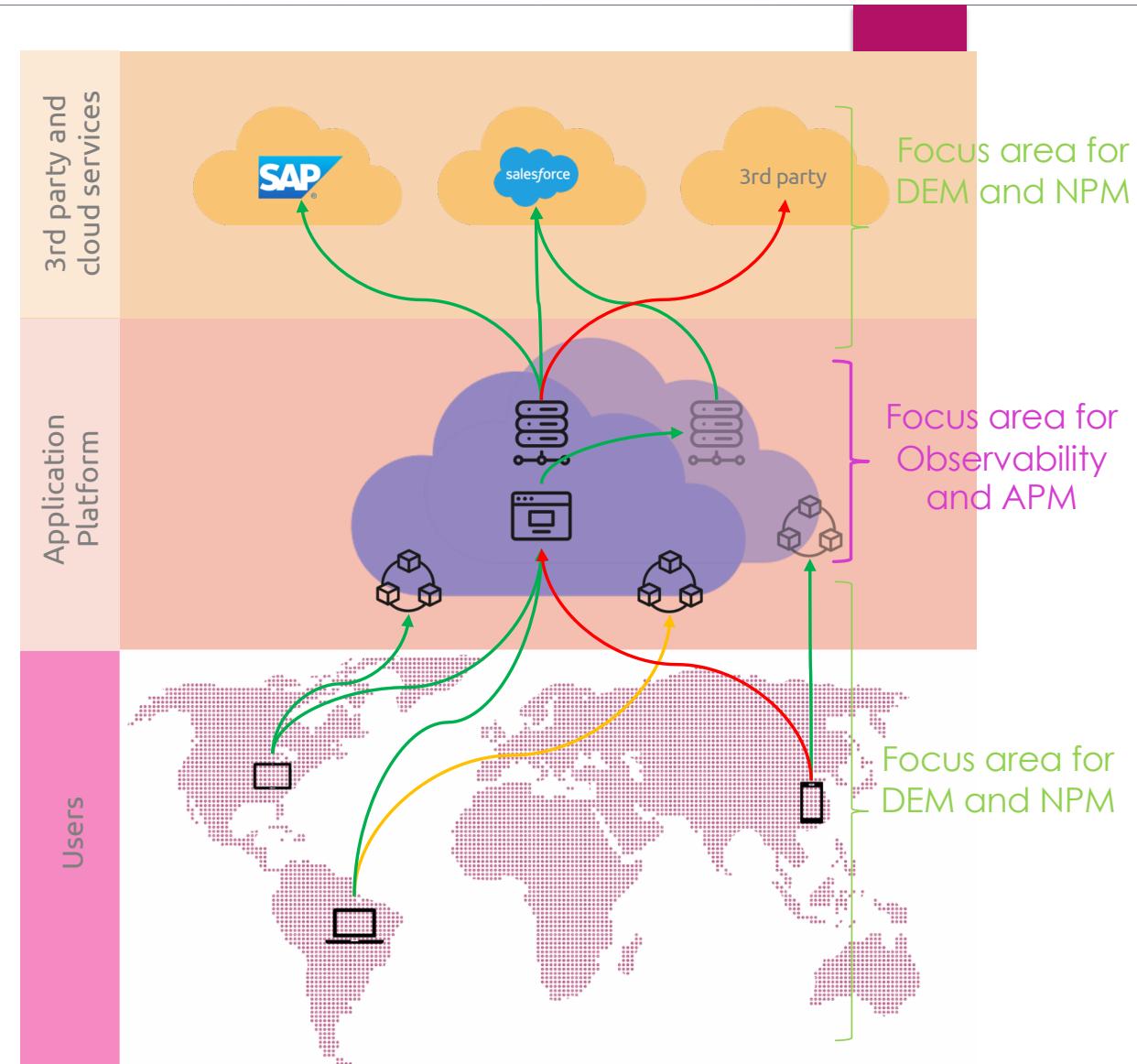
Why Digital Experience Matters?

DIGITAL EXPERIENCE IS A BUSINESS MATTER

- ▶ Drives **conversion / adoption** of digital services
- ▶ Drives **churn**
- ▶ Is a **competitive (dis)advantage**

Understand the context and the scope

- APM and Observability tend to focus on collecting metrics, events and traces inside the application platform (e.g. within and between micro services)
- Specific value of Digital Experience Monitoring is on the delivery of application services from the distributed platform to the users (and from the platform to 3rd parties)



The DRIVERS of Digital Experience

Contents
("« WHAT »")

Combined Performance
("« HOW they are rendered »")

User Profile
("Who")

Elementary Performance
("« HOW Fast »")

Users Vs platform hosting
("« WHERE »")

User Profile

« Who »

- ▶ You do not design for performance without data on:
 - ▶ Where your users are connected [LOCATION]
 - ▶ How they are connected [CONNECTIVITY]
 - ▶ Which device and browser they use [DEVICE – USER AGENT]
 - ▶ How sensitive to performance they are

Contents

« What »

- ▶ All the elements required by the application
 - ▶ Initial page load
 - ▶ Nav / HTML
 - ▶ Headers elements (CSS, Scripts)
 - ▶ Body elements (images, additional CSS, videos, Scripts)
 - ▶ Subsequent calls
 - ▶ Triggered by scripts
 - ▶ Triggered by user actions
- ▶ Overall size to be loaded
- ▶ Which elements are required for the app to be perceived as functional?

Elementary Performance

« How fast »

Page Elements



JS Elements

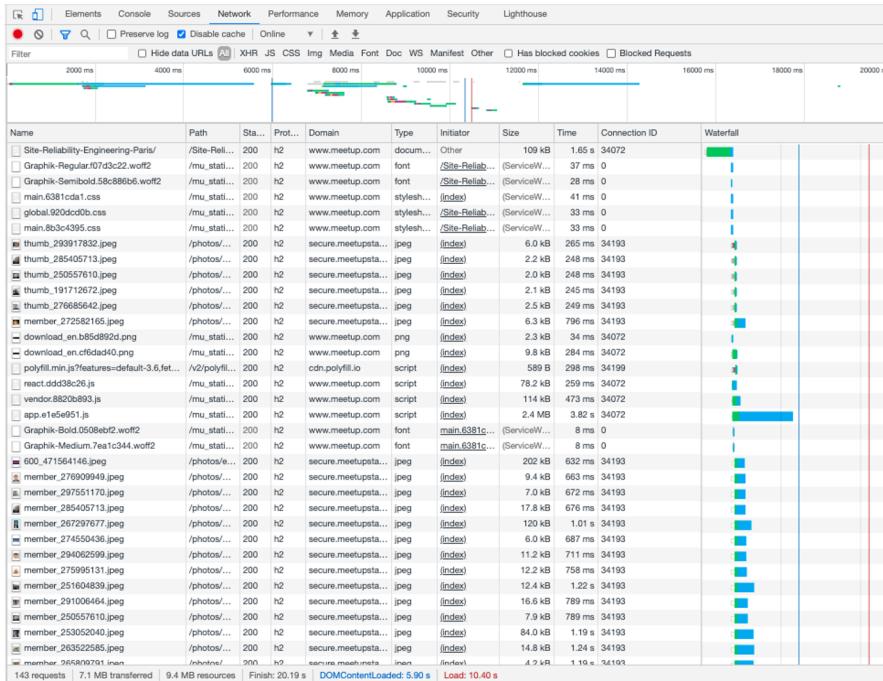


Optional

Can be skipped by caching
Mandatory

Combined Performance

« How it is rendered - Critical Path »



Combined contents

Overall size

Overall number of elements

Number of CSS files

Critical path

Async / Defer scripts

Items blocking the main thread

Appearance of user critical elements

Network sessions

of DNS resolutions required

of TCP sessions required

TLS session setup

Protocols

Parallel handling / Queueing

Users location and platform hosting

« Where »

- ▶ Drivers
 - ▶ Users network « location »
 - ▶ Where
 - ▶ Network connectivity
 - ▶ Device and browser
 - ▶ Platform hosting
 - ▶ Main platform
 - ▶ CDNs
 - ▶ 3rd parties
- ▶ Consequences
 - ▶ DNS resolution times (depending on your DNS service and the users' location)
 - ▶ Network Connection Times
 - ▶ Regional specifics (China's big firewall)

Main performance optimization paths

| Contents | Elementary performance | Combined – Critical Path performance | Users location and platform hosting |
|--|--|--|--|
| <ul style="list-style-type: none">• Be savvy! :-)• Media optimization• Minification (CSS, Scripts)• Remove unused scripts | <ul style="list-style-type: none">• Caching• Compression• Progressive images• 3rd party for libraries and fonts• Protocol• Jscript execution | <ul style="list-style-type: none">• User critical elements first• Defer / Async for scripts• HTTP multiplexing• Hunt scripts blocking the main thread | <ul style="list-style-type: none">• Device / Browser specific versions• Regional / Dynamic CDN strategy• Global accelerators• Regional front ends |

Main ways to monitor experience

| Tool Category | Tools / Vendors | Pros | Cons |
|----------------------|---|--|--|
| Chrome Webdev Tool | <ul style="list-style-type: none">Google | <ul style="list-style-type: none">Free | <ul style="list-style-type: none">Not aligned with real world (network, device, location)Expertise required |
| HTTP Synthetic | <ul style="list-style-type: none">WebpagetestSpeedcurveSolarwinds Pingdom | <ul style="list-style-type: none">Reproducible testCan partially reflect users' location | <ul style="list-style-type: none">Requires scenario configuration (hard to automate)Not aligned with real world (network, device)Does not show all transactions used |
| Real User Monitoring | <ul style="list-style-type: none">SpeedcurveAppDynamicsNewRelicDynatraceKadiska | <ul style="list-style-type: none">Align with real life (network, device, location)Instant setup | <ul style="list-style-type: none">Some data depth limitation (e.g. network and script errors) |

Monitoring Metrics and Gaps

MONITORING METRICS

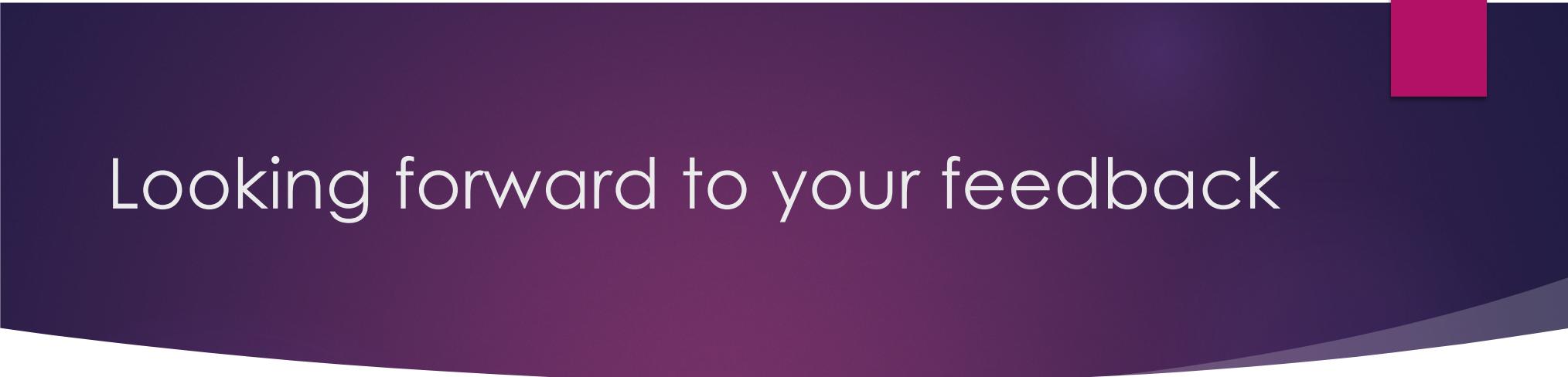
- ▶ Page Load Times
- ▶ FCP, TTI
- ▶ Web Core Vitals: LCP, FID, CLS

GAPS

- ▶ Single Page Applications / Web progressive apps
- ▶ Mobile Apps
- ▶ Infrastructure layer monitoring (DNS, Network, TLS, APIs)

Best Practices

- ▶ Understand and monitor customer profile / application usage
 - ▶ Location
 - ▶ Device
 - ▶ Network connectivity
 - ▶ Business criticality
- ▶ Define critical transactions and for each critical resource
- ▶ Define performance budgets / SLOs
- ▶ Seek simplification / minification
- ▶ Integrate performance in design and architecture decisions



Looking forward to your feedback