

# CT 420 Real-Time Systems

# Web Performance Benchmarking

Dr. Jawad Manzoor Assistant Professor School of Computer Science

#### Contents



- User QoE
- Case Studies
- Core Web Vitals
- Performance Measurements Tools

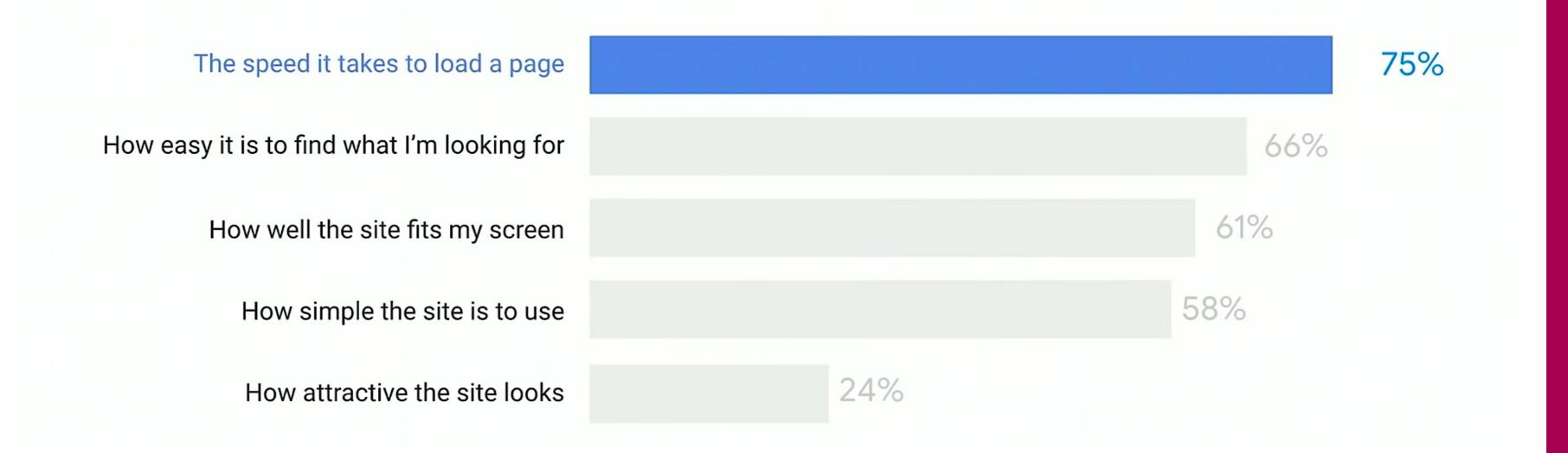
#### Speed and Performance



- Speed and performance plays a vital role in the success of any online venture.
- Websites that load quickly and respond to user input in a timely fashion engage and retain users better than websites that are slow to load, and feel sluggish.
- Modern websites ship lots of code and mobile devices in particular have limited CPU power and memory. Coupled with adverse network conditions, this can create poor performance.
- Monitoring performance is of utmost importance today.

#### User Experience Factors





Source: SPEED MATTERS: Designing for Mobile Performance, Awwwards

#### Speed Matters

■ When page load time increases from..

1s to 3s the probability of bounce increases 32%

1s to 5s the probability of bounce increases 90%

Source: Google/SOASTA Research, 2017.



### How to measure performance?

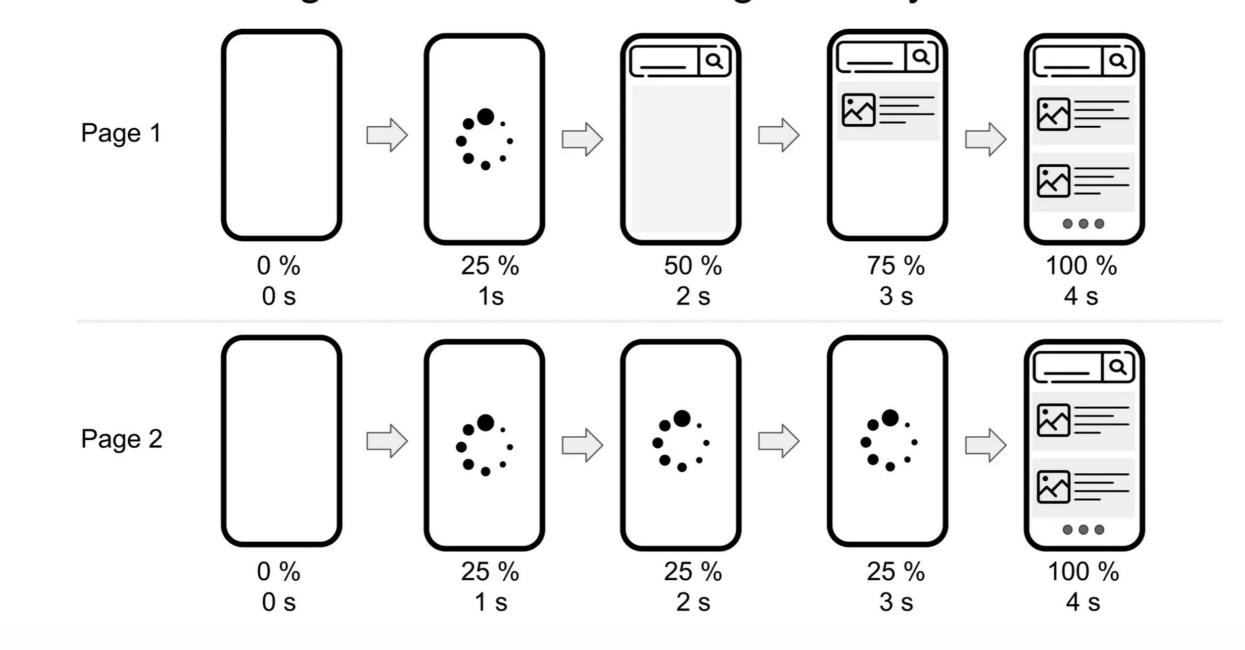


- Speed and performance are relative terms.
- Each application dictates its own set of requirements based on:
  - Business criteria
  - Context
  - User expectations
  - Complexity of the task
- ☐ In the context of web, we need to plan and design for specific, user-centric perceptual processing time constants.

### User's Perspective



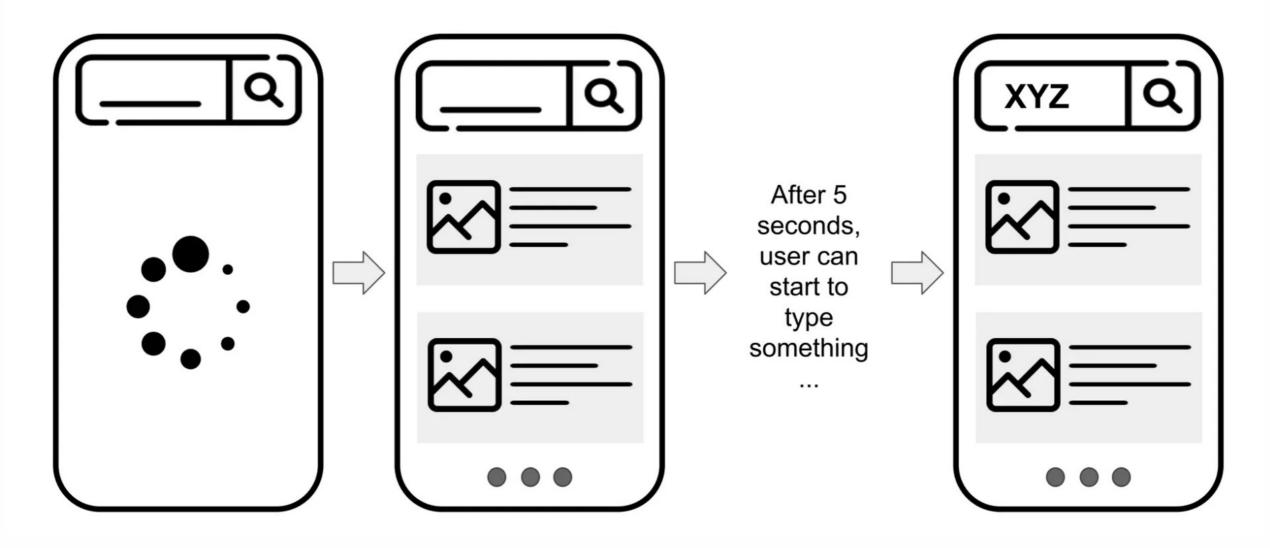
#### Scenario 1: Pages Load Content Progressively Seems Faster



## User's Perspective



Scenario 2: Pages Appear Quickly But Respond to User Interaction Slowly



#### Older Metrics of Measurement



#### ■ Document: DOMContentLoaded event

- The DOMContentLoaded event fires when the HTML document has been completely parsed, and all deferred scripts (<script defer src="..."> and <script type="module">) have downloaded and executed.
- It doesn't wait for other things like images, subframes, and async scripts to finish loading.

#### ■ Window: load event

- The load event is fired when the whole page has loaded, including all dependent resources such as stylesheets, scripts, iframes, and images.
- □ Older metrics like load or DOMContentLoaded are not good because they don't necessarily correspond to what the user sees on their screen.



# Core Web Vitals

#### Core Web Vitals



- ☐ Web Vitals is an initiative by Google to provide unified guidance for quality signals that are essential to delivering a great user experience on the web.
- ☐ Core Web Vitals are the subset of Web Vitals that apply to all web pages and should be measured by all site owners.
- Each of the Core Web Vitals represents a distinct facet of the user experience, is measurable in the field, and reflects the real-world experience of a critical user-centric outcome.
- ☐ The current set for Core Web Vitals focuses on three aspects of the user experience—loading, interactivity, and visual stability.

#### Core Web Vitals





#### **Perceived Load Speed**

First Contentful Paint (FCP)

Largest Contentful Paint (LCP)

Speed Index



#### **Load Responsiveness**

First Input Delay (FID)

Time To Interactive(TTI)

Total Blocking Time (TBT)



#### **Visual Stability**

Cumulative Layout Shift (CLS)



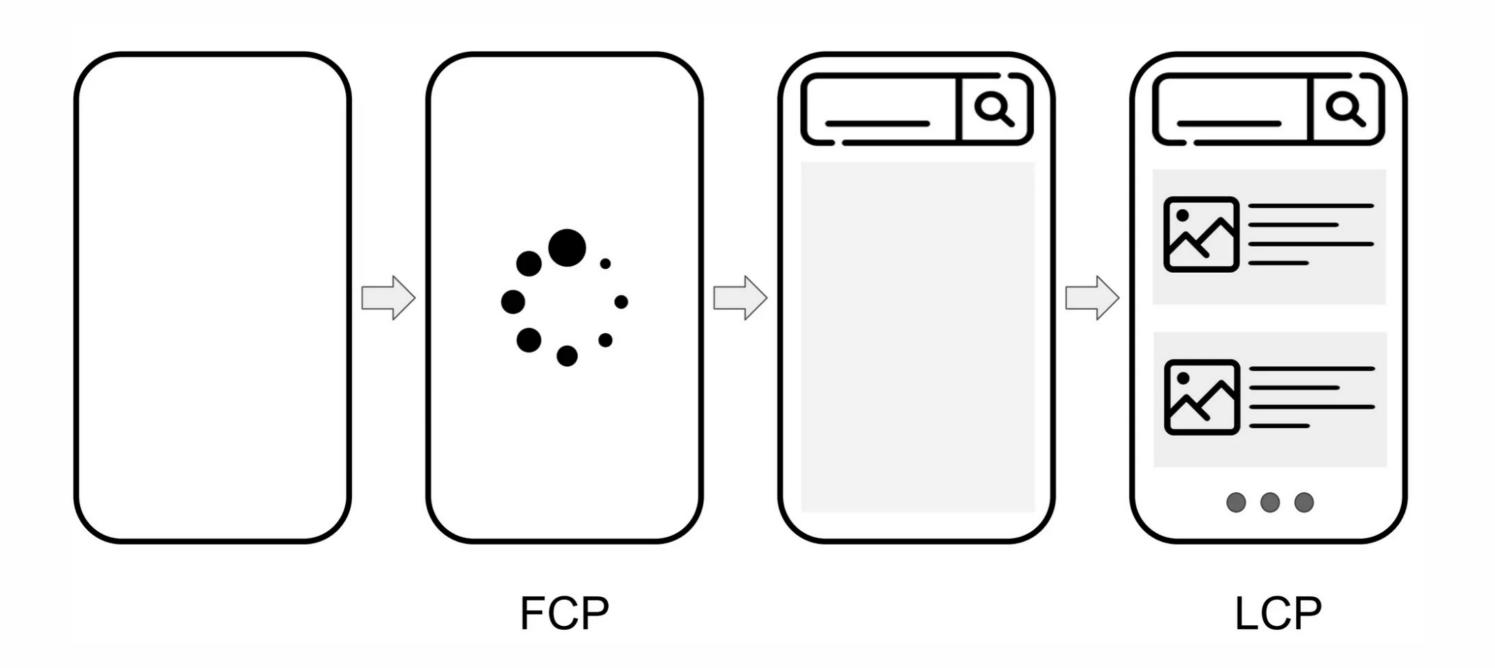
- ☐ First Contentful Paint (FCP)
  - First Contentful Paint marks the time at which the first text or image is painted
  - FCP measures how long it takes the browser to render the first piece of DOM content after a user navigates to your page.
  - FCP score is a comparison of your page's FCP time and FCP times for real websites, based on data from the HTTP Archive.
  - For example:
    - Sites performing in the ninety-ninth percentile render FCP in about 1.2 seconds.
    - If your website's FCP is 1.2 seconds, your FCP score is 99.



- Largest Contentful Paint (LCP)
  - LCP measures when the largest content element in the viewport is rendered to the screen.
  - This approximates when the main content of the page is visible to users.
  - Lighthouse extracts LCP data from Chrome's tracing tool.
  - The table below shows how to interpret your LCP score:

LCP time (in seconds)	Color-coding
0-2.5	Green (fast)
2.5-4	Orange (moderate)
Over 4	Red (slow)







#### Speed Index

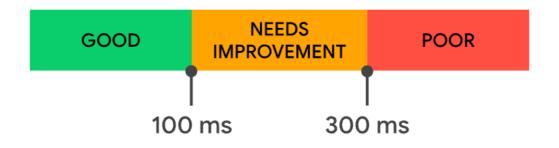
- Speed Index measures how quickly content is visually displayed during page load.
- Lighthouse first captures a video of the page loading in the browser and computes the visual progression between frames.
- Lighthouse then generates the Speed Index score, which is a comparison of your page's speed index and the speed indices of real websites, based on data from the HTTP Archive.
- This table shows how to interpret your mobile Speed Index score:

Speed Index (in seconds)	Color-coding
0-3.4	Green (fast)
3.4-5.8	Orange (moderate)
Over 5.8	Red (slow)



- First Input Delay
  - It is important it is to make a good first impression
  - The First Input Delay (FID) metric helps measure your user's first impression of your site's interactivity and responsiveness.
  - FID measures the time from when a user first interacts with a page (that is, when they click a link, tap on a button, or use a custom, JavaScript-powered control) to the time when the browser is actually able to begin processing event handlers in response to that interaction.







- ☐ Time to Interactive (TTI)
  - TTI measures how long it takes a page to become fully interactive. A page is considered fully interactive when:
    - The page displays useful content, which is measured by the First Contentful Paint,
    - Event handlers are registered for most visible page elements, and
    - The page responds to user interactions within 50 milliseconds.
  - Measuring TTI is important because some sites optimize content visibility at the expense of interactivity. This can create a frustrating user experience:

TTI metric (in seconds)	Color-coding
0-3.8	Green (fast)
3.9-7.3	Orange (moderate)
Over 7.3	Red (slow)

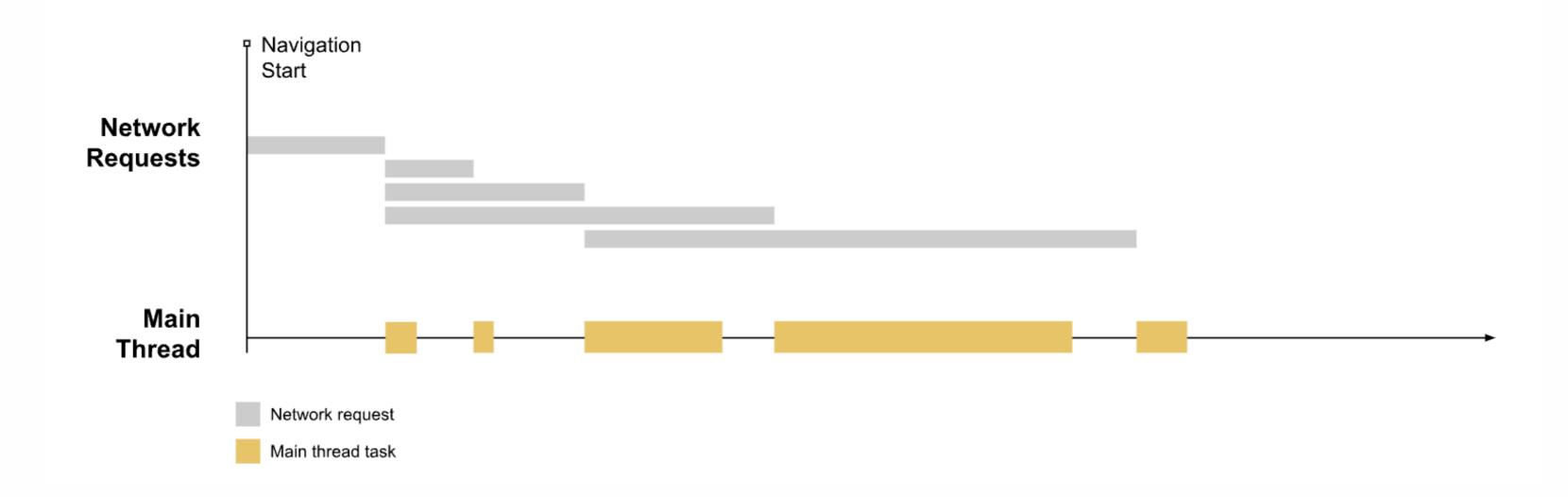


- Total Blocking Time (TBT)
  - TBT measures the total amount of time that a page is blocked from responding to user input, such as mouse clicks, screen taps, or keyboard presses.
  - The sum is calculated by adding the blocking portion of all long tasks between First Contentful Paint and Time to Interactive.
  - Any task that executes for more than 50 ms is a long task. The amount of time after 50 ms is the blocking portion.
  - For example: if Lighthouse detects a 70 ms long task, the blocking portion would be 20 ms.

TBT time (in milliseconds)	Color-coding
0-200	Green (fast)
200-600	Orange (moderate)
Over 600	Red (slow)

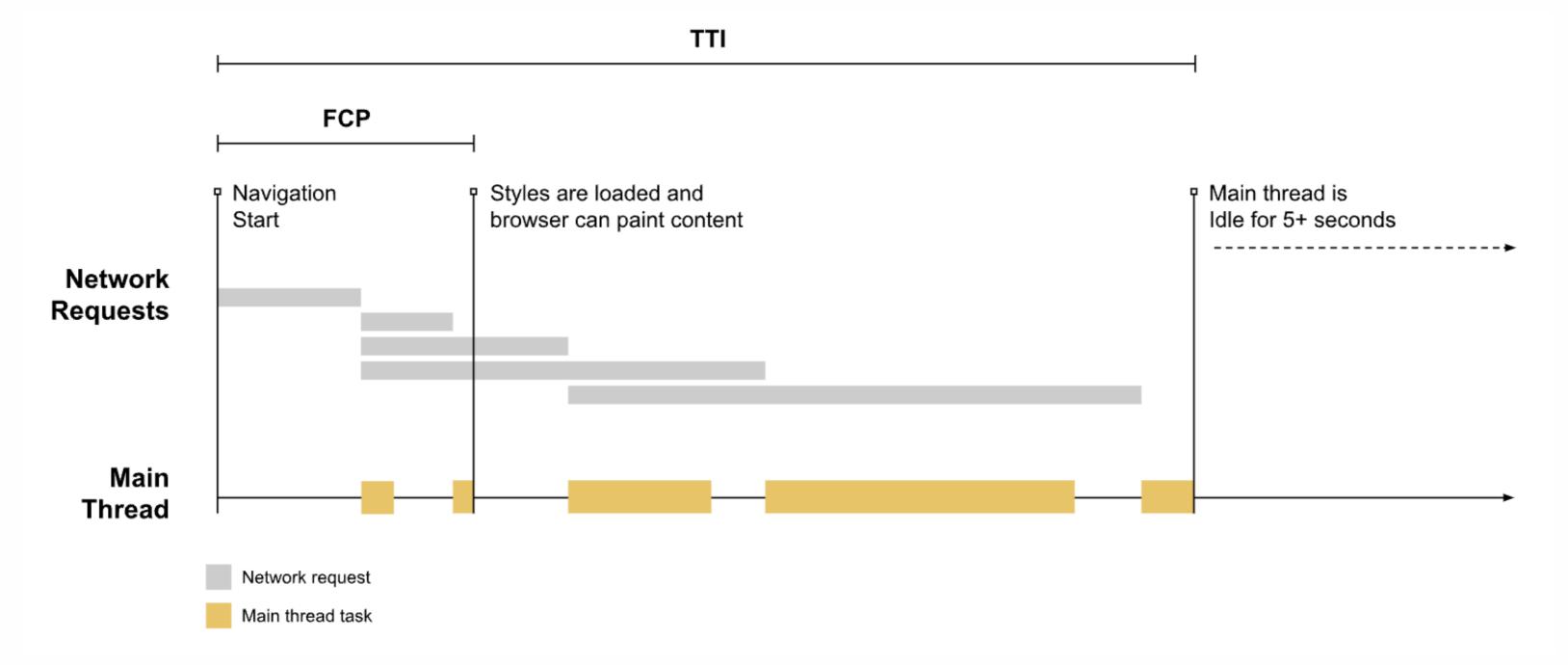


- ☐ Timeline of a typical web page load
  - A page makes some network requests for resources (CSS and JS files)
  - After those resources are finished downloading, they're processed on the main thread.



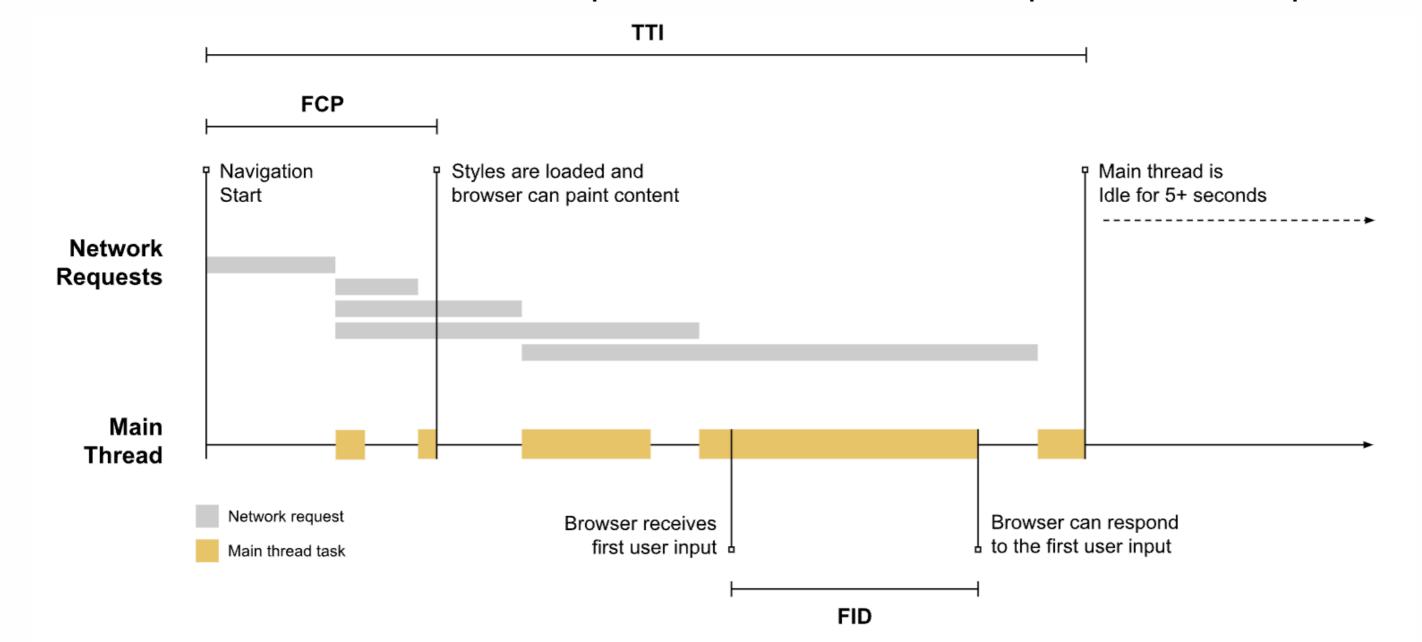


 Long first input delays typically occur between FCP and TTI because the page has rendered some of its content but isn't yet reliably interactive.





- Suppose a user tried to interact with the page near the beginning of the longest task
- The input occurs while the browser is in the middle of running a task, therefore, it
  has to wait until the task completes before it can respond to the input.





- ☐ We have often been in a situation when we are:
- □ Reading an article online and something suddenly changes on the page without warning, the text moves, and you've lost your place.
- Or you're about to tap a link or a button, but in the instant before your finger lands—BOOM—the link moves, and you end up clicking something else!
- ☐ This is annoying and can sometimes cause real damage.







- Unexpected movement of page content usually happens because resources are loaded asynchronously or DOM elements get dynamically added to the page above existing content.
- ☐ The culprit might be:
  - An image or video with unknown dimensions
  - A font that renders larger or smaller than its fallback
  - A third-party ad or widget that dynamically resizes itself.



- □ Cumulative Layout Shift (CLS):
  - A layout shift occurs any time a visible element changes its position from one rendered frame to the next.
  - A burst of layout shifts, known as a session window, is when one or more individual layout shifts occur in rapid succession with less than 1-second in between each shift and a maximum of 5 seconds for the total window duration.
  - The largest burst is the session window with the maximum cumulative score of all layout shifts within that window.
  - CLS is a measure of the largest burst of layout shift scores for every unexpected layout shift that occurs during the entire lifespan of a page.



☐ Cumulative Layout Shift (CLS)



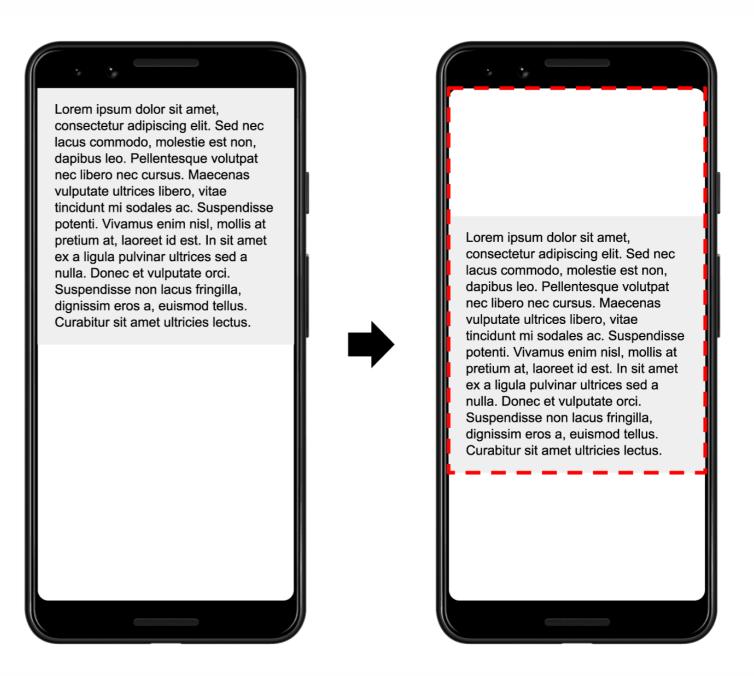


- **□** Layout Shift Score:
  - The browser looks at the viewport size and the movement of unstable elements in the viewport between two rendered frames and calculates impact fraction and the distance fraction.

```
layout shift score = impact fraction * distance fraction
```

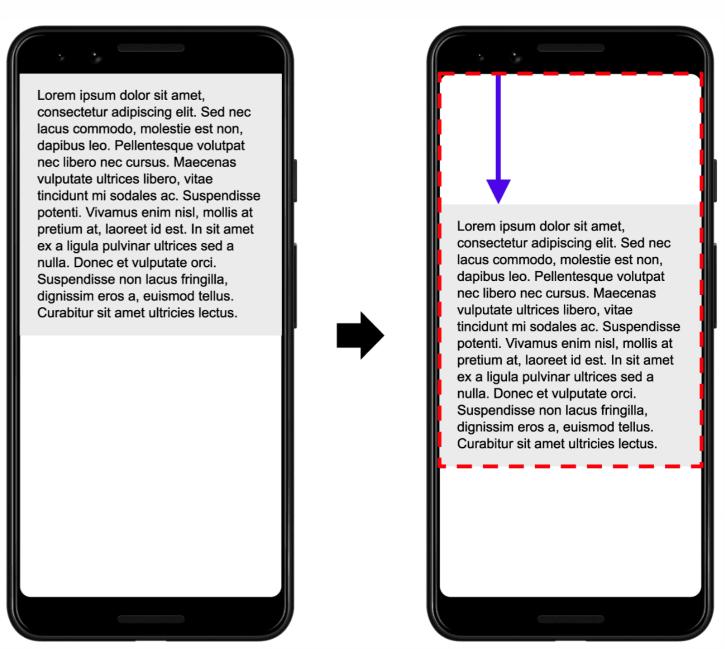


- ☐ The impact fraction is the union of the visible areas of all unstable elements for the previous frame and the current frame—as a fraction of the total area of the viewport.
- The element in the image takes up half of the viewport in one frame.
- Then, in the next frame, the element shifts down by 25% of the viewport height.
- The red, dotted rectangle indicates the union of the element's visible area in both frames, which is 75% of the total viewport, so its impact fraction is 0.75.





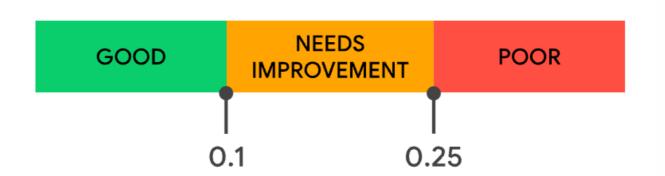
- ☐ The distance fraction is the distance that unstable elements have moved, relative to the viewport. The distance fraction is the greatest distance any unstable element has moved in the frame divided by the viewport's largest dimension.
- In the image, the largest viewport dimension is the height, and the unstable element has moved by 25% of the viewport height, which makes the distance fraction 0.25.
- In this example the impact fraction is 0.75 and the distance fraction is 0.25, so the layout shift score is 0.75 \* 0.25 = 0.1875.





- ☐ Cumulative Layout Shift (CLS):
  - To provide a good user experience, sites should strive to have a CLS score of 0.1 or less.

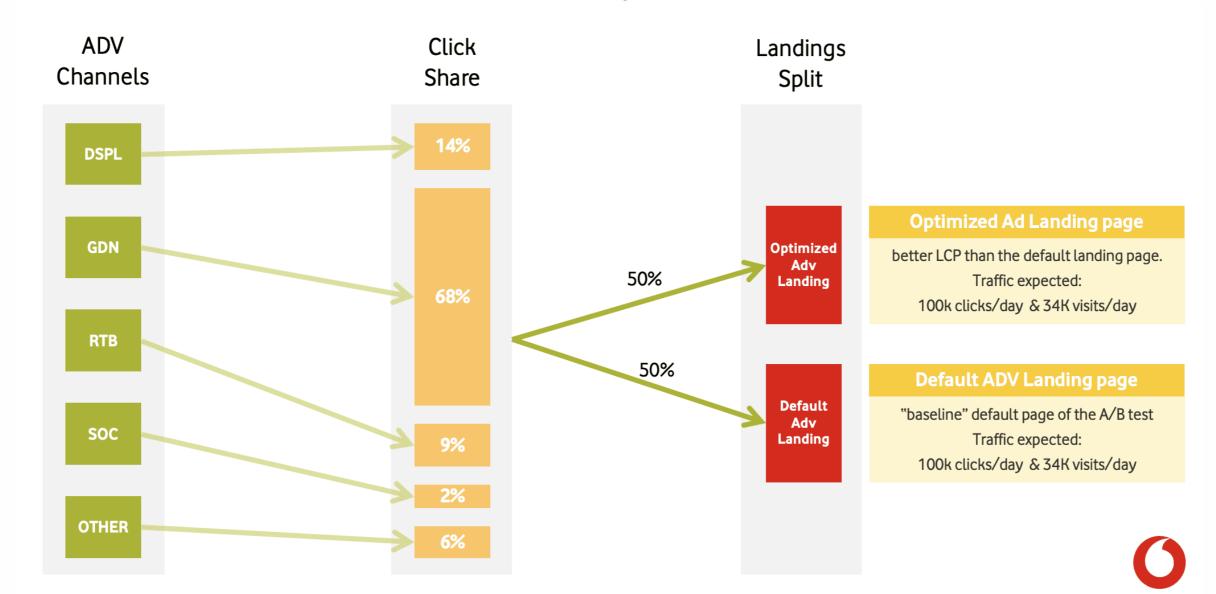




## Case Study: Vodafone



- Vodafone is a leading telecommunications company operating fixed and mobile networks in 21 countries.
- □ An A/B test was run on a landing page (where version A was optimized for Web Vitals and had a 31% better LCP score)



## Case Study: Vodafone



■ Vodafone determined that the optimization resulted in a very positive Return on Investment (ROI)

#### **IMPACT**



More sales



Improvement in the cart to visit rate



15%
Improvement in the lead to visit rate

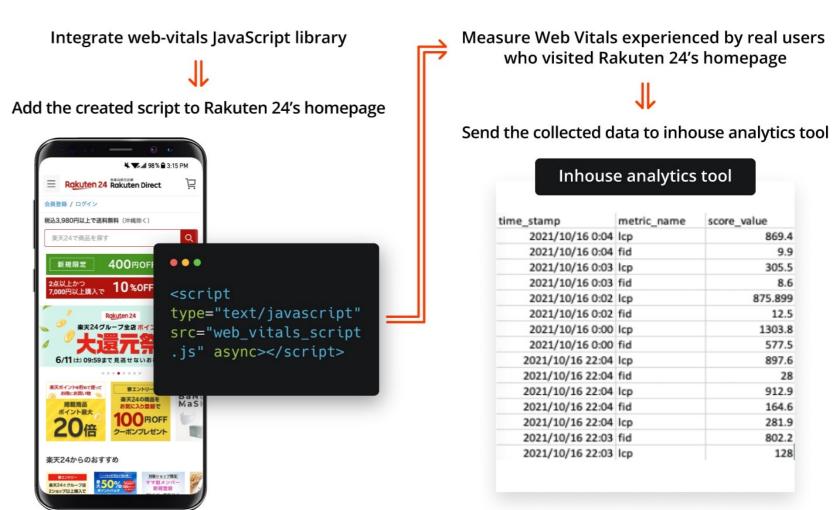


31% Improvement in LCP

## Case Study: Rakuten 24



- □ Rakuten 24 is an online store that collaborates with both major multinational and domestic consumer goods manufacturers to offer a great variety of daily necessities.
- ☐ This store is provided by Rakuten Group, Inc.—a global leader in internet services, and is among the top performers on their digital marketplace platform in Japan.
- Rakuten 24 decided to use the web-vitals JavaScript library to measure Core Web Vitals and other metrics in the field and send the data to inhouse analytics tool.
- Rakuten 24 also ran an A/B test focused on optimizing Core Web Vitals and related metrics.



## Case Study: Rakuten 24









1.2s 1.7s 1.3s 1.4s 1.5s 1.6s 1.8s 1.9s 2.0s Layout shift 54% 43% 43% 43% 43% 43% 43% 100% 54%

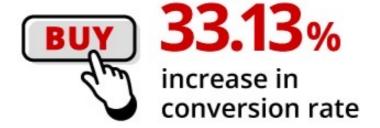
### Case Study: Rakuten 24

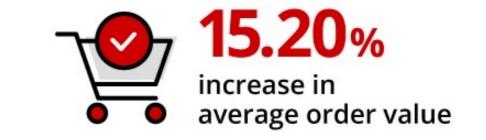


Web performance optimization is challenging but rewarding.

#### **IMPACT**













# Performance Measurement Tools

## Performance Pillars

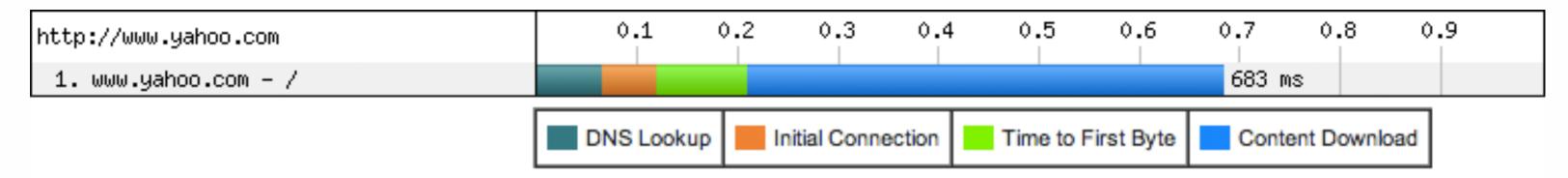


- Networking fetching resources
- Rendering page layout and rendering
- Computing JavaScript execution
- ☐ The rendering and computing steps follow a single-threaded approach
  - it is not possible to perform concurrent modifications of the resulting Document Object Model (DOM).
  - optimizing how the rendering and script execution runtimes work together is of critical importance.
- ☐ Fast and efficient delivery of network resources is also a performance keystone of web applications.
  - Optimizations of rendering won't do much good if the browser is blocked on the network, waiting for the resources to arrive





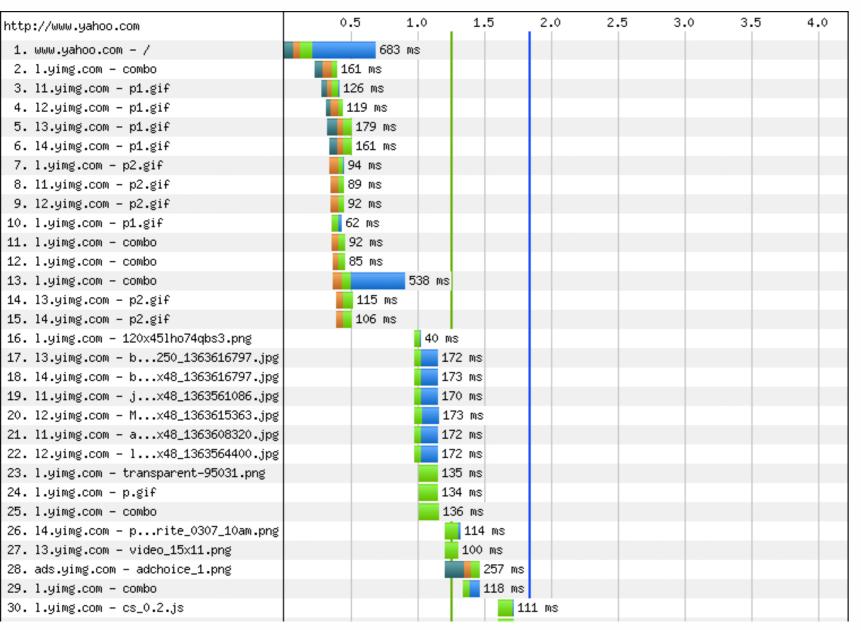
- ☐ The resource waterfall is likely the single most insightful network performance and diagnostics tool at our disposal.
- Every HTTP request is composed of several separate stages



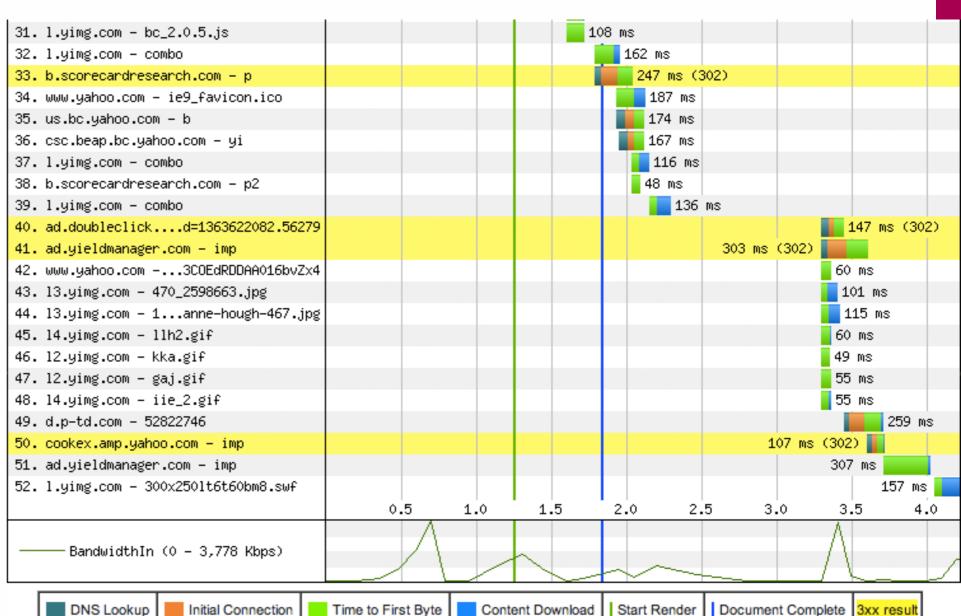
- ☐ In this example, the Yahoo! homepage took 683 ms to download.
- Over 200 ms of that time was spent waiting on the network, which amounts to 30% of total
- ☐ The rest of the time is spent downloading resources.



- A modern web application needs a wide variety of resources to produce the final output.
- ☐ In the given example, to load the Yahoo! homepage, the browser requires 52 resources, fetched from 30 different hosts, all adding up to 486 KB in total.
- ☐ The resource waterfall reveals several important insights about the structure of the page and the browser processing pipeline.
- □ It is a powerful tool that can help reveal the chosen optimizations, or lack thereof, for any page or application.





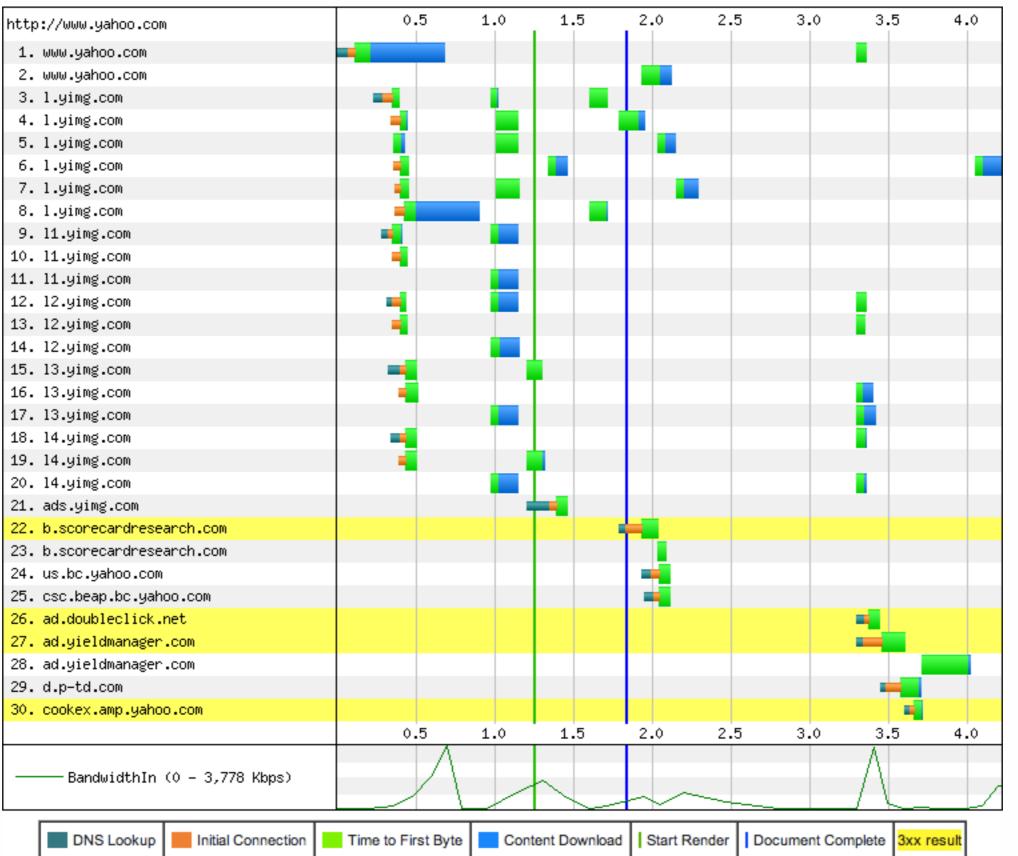




- Resource waterfall analysis
  - Notice that HTML parsing is performed incrementally, allowing the browser to discover required resources early and dispatch the HTTP requests in parallel.
  - The incremental discovery of each resource in the document is what creates the distinct resource "waterfall effect."
  - The "Start Render" (green vertical line) occurs well before all the resources are fully loaded, allowing the user to begin interacting with the page while the page is being built.
  - the "Document Complete" event (blue vertical line), also fires early and well before the remaining assets are loaded.
- □ Different browsers implement different logic for when, and in which order, the individual resource requests are dispatched.

### Connection View





#### Connection View



- ☐ The connection view shows the life of each TCP connection
- ☐ In this example:
  - 14 DNS lookups
  - 22 TCP handshakes
  - A lot of network latency (indicated in green) while waiting to receive the first byte of each response
  - The download time (indicated in blue) is a small fraction of the total time.
- Why some requests are showing the green bar (time to first byte) only?
- □ In the bandwidth chart, why is the utilization of the available connection very low?



# Browser Developer Tools

## DevTools



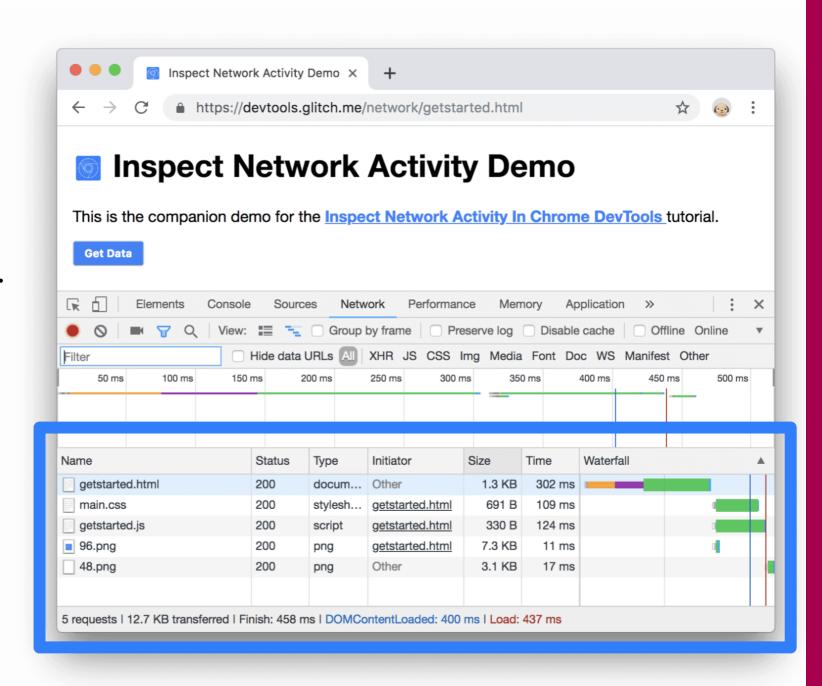
- ☐ Chrome DevTools is a set of web developer tools built directly into the Google Chrome browser.
- It provides in-depth analysis on everything that happens while your page loads or runs.
- Features:
  - Evaluate site performance
  - Log network requests
  - ☐ Simulate mobile devices with Device Mode
  - Many more ...
- Other browsers (Firefox, Safari etc.) also have developer tools that offer similar functionality.

## DevTools



#### Features

- View the network activity
- Inspect a resource's details
- Filter resources
- View/Search headers, responses, cookies, timing etc.
- Customize columns
- Simulate a slower network connection
- Recorder
- Emulate a first-time visitor
- Emulate slow network connections
- Export data





# Lighthouse

# Lighthouse



- ☐ Lighthouse is an open-source, automated tool for improving the performance, quality, and correctness of your web apps.
- ☐ Give Lighthouse an URL and it runs a series of audits against the web page.
- ☐ It generates a report that offers opportunities for optimization.
- ☐ It has audits for:
  - Performance
  - Accessibility
  - Progressive web apps
  - SEO

## Lighthouse



- How to run Lighthouse?
  - Chrome DevTools Easily audit pages that require authentication, and read your reports in a user-friendly format.
  - Command line Automate your Lighthouse runs via shell scripts.
  - Node module Integrate Lighthouse into your continuous integration systems.
  - Web UI Run Lighthouse and link to reports without installing a thing

#### Demo



- 1. Navigate to <a href="PageSpeed Insights">PageSpeed Insights</a>.
- 2. Enter a web page URL.
- 3. Click Analyze.

# Audit Report



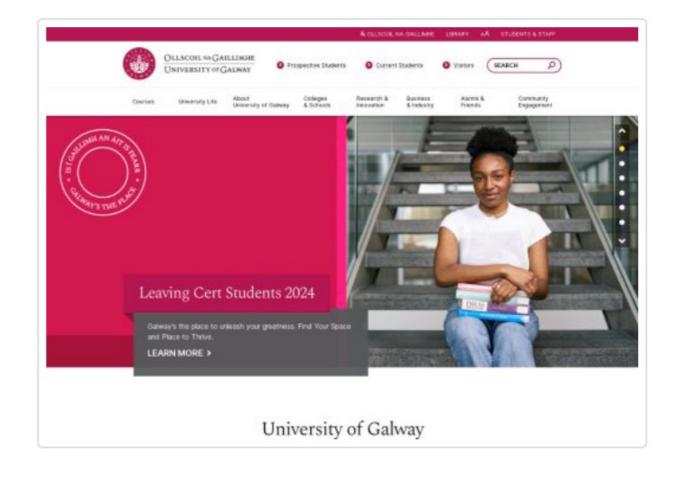




Values are estimated and may vary. The performance score is calculated directly from these metrics. See calculator.

Performance





# Audit Report



#### **METRICS**

First Contentful Paint

 $0.9 \, s$ 

Total Blocking Time

30 ms

Speed Index

 $0.9 \, s$ 

Largest Contentful Paint

1.3 s

Cumulative Layout Shift

0.003

#### **OPPORTUNITIES**

#### Opportunity

▲ Serve images in next-gen formats

Efficiently encode images

Properly size images

**Estimated Savings** 

2.56s ×

1.16s ×

■ 0.36s ∨

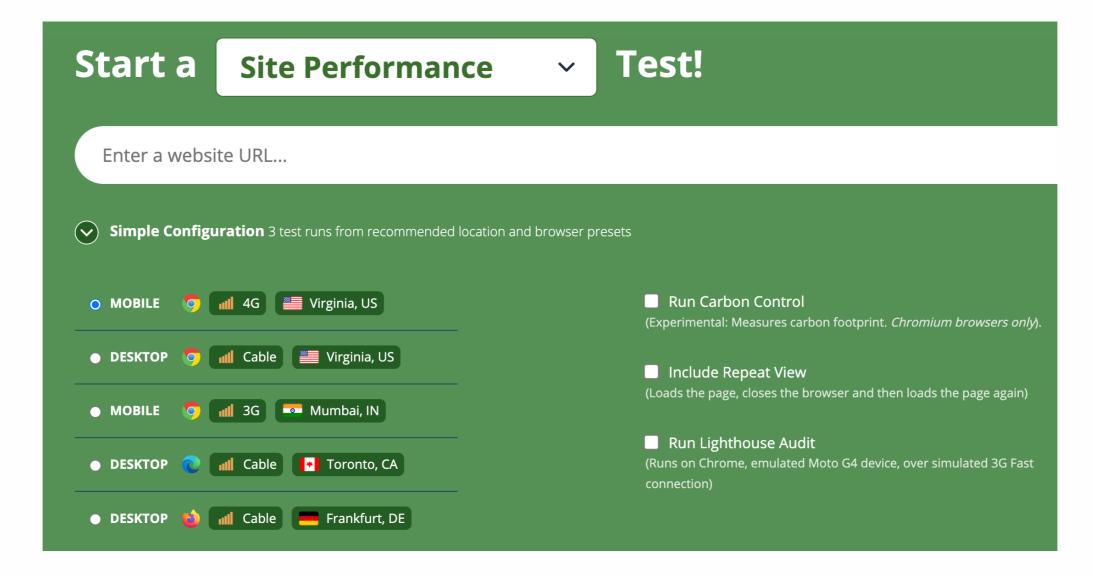


# WebPage Test

## WebPage Test

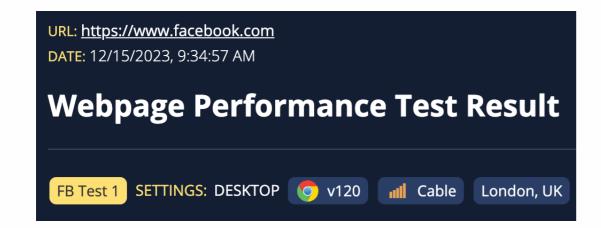


- It is an open-source project and a free web service that provides a system for testing the performance of web pages from multiple locations around the world.
- It also allows you to add packet-level network emulation on top, including aspects such as packet loss



# WebPage Test







(Based on Median Run by: ▼ Speed Index)

Note: Metric availability will vary

First View (Run 2)

**Time to First Byte** 

When did the content start downloading?

**Start Render** 

When did pixels first start to appear?

**First Contentful Paint** 

How soon did text and images start to appear?

**Speed Index** 

How soon did the page look usable?

**Largest Contentful Paint** 

When did the largest visible content finish loading?

Cumulative Layout Shift Total Blocking Time Page Weight

How much did the design shift while loading?

Was the main thread

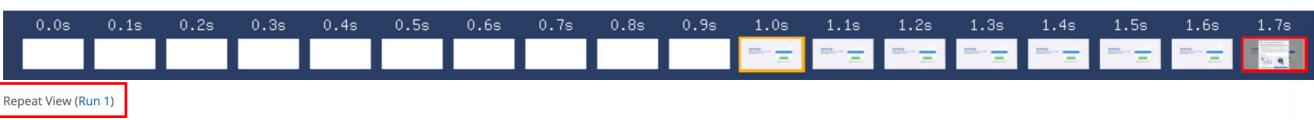
blocked?

blocked?

How many bytes

downloaded?

Visual Page Loading Process (Explore)



**Time to First Byte** 

**.317**s

When did the content start downloading?

**Start Render** 

.700s

When did pixels first start to appear?

**First Contentful Paint** 

**.665**s

How soon did text and images start to appear?

**Speed Index** 

**1.056**s

How soon did the page look

**Largest Contentful Paint** 

**1.019**s

When did the largest visible content finish loading?

Cumulative Layout Shift Total Blocking Time Page Weight

How much did the design shift while loading?

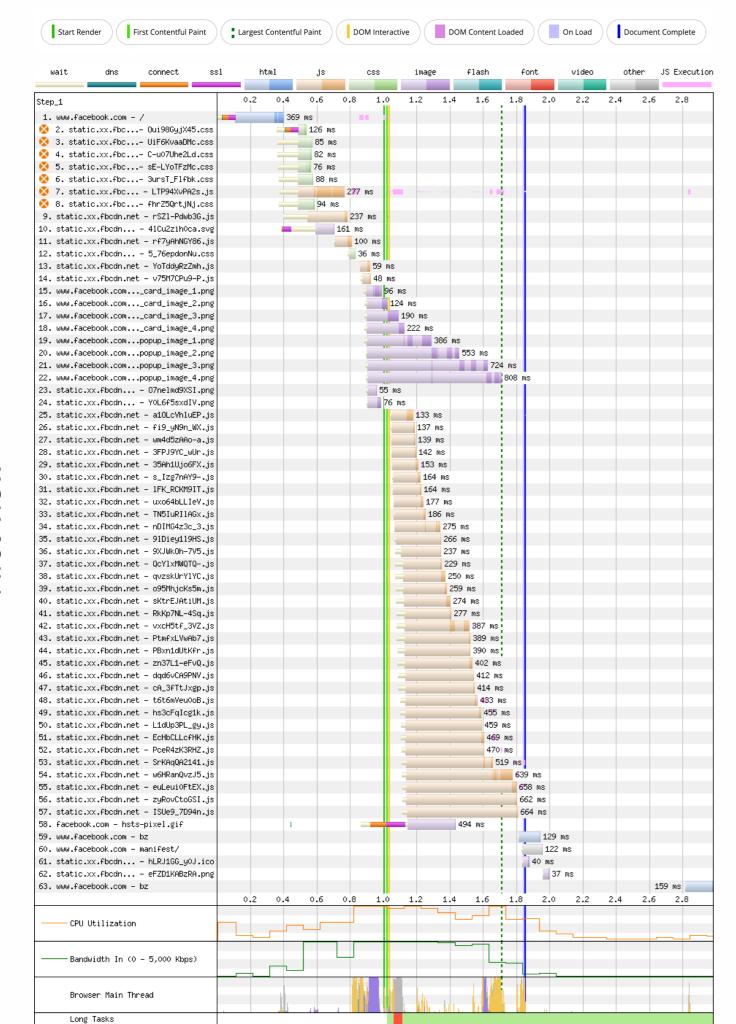
**26**<sub>KB</sub>

Was the main thread How many bytes downloaded?

**Visual Page Loading Process** (Explore)

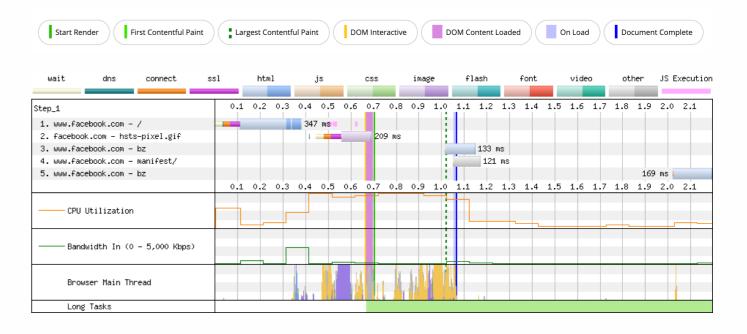








#### Resource Waterfall of the Test



Repeat view

#### Connection View of the Test

First view

Repeat view





# Acknowledgements

OLLSCOIL NA GAILLIMИE
UNIVERSITY OF GALWAY

- □ https://developer.chrome.com/docs/lighthouse
- □ <a href="https://web.dev/">https://web.dev/</a>
- □ <a href="https://www.webpagetest.org/">https://www.webpagetest.org/</a>

