What Is Time to First Byte (TTFB) and How to Reduce It

Last updated on Feb 8th, 2022 | 8 min

Time to First Byte (TTFB) divided the SEO and web performance communities back in the day. While many considered TTFB a useless metric, other experts believed in its importance.

But more on that later.

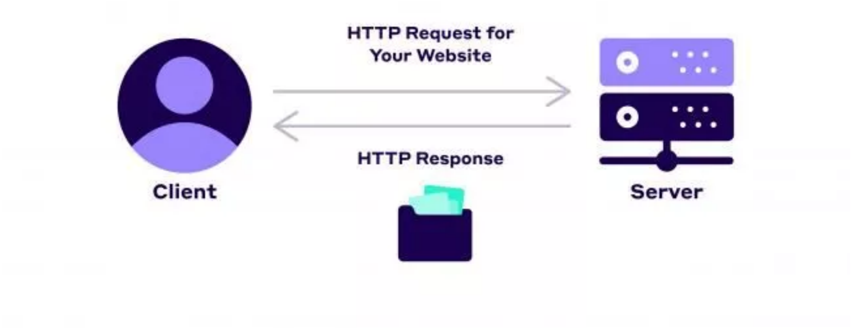
For now, you should know that our opinion on the topic is that TTFB is a valuable metric.

In this article, you'll understand why it's important to measure TTFB along with:

Let's begin!

What is Time To First Byte?

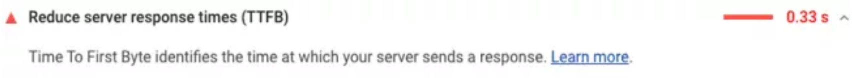
When a user tries to visit a page, their browser sends an HTTP request. The server that hosts the page has to process that request and return a response.



Time to First Byte (TTFB) measures how long it takes for a client's browser to receive the first byte of the response from the server.

The longer it takes for the server to process the request and send a response, the slower your visitors' browsers start displaying your page.

A common misconception about TTFB is that it's the same thing as Server Response Time. You will see the two terms used interchangeably, even in some popular speed testing tools:



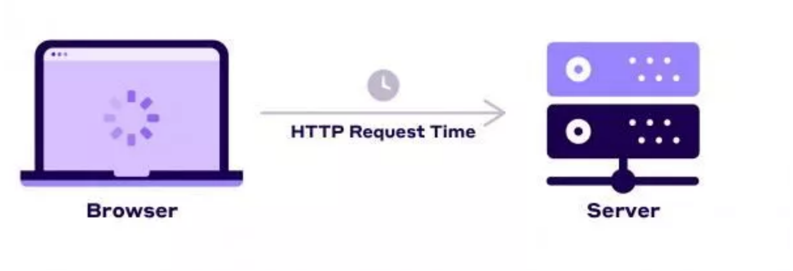
▲ Reduce server response times (TTFB) 0.33 s

Time To First Byte identifies the time at which your server sends a response. Learn more.

However, Server Response Time measures how fast the server responds, but **not how fast the response reaches the client**. It doesn't include network latency in its measurement, which is a factor that affects the real user experience.

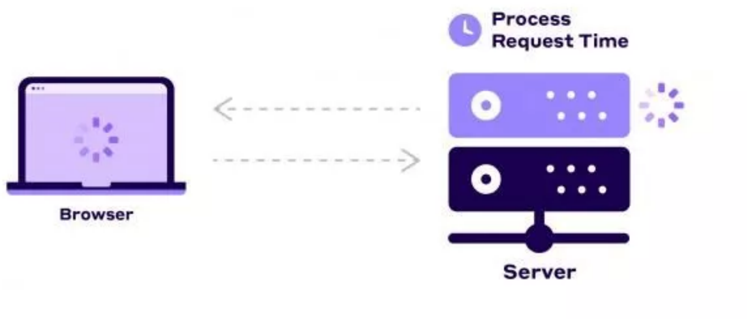
On the other hand, TTFB measures the entire process from start to finish:

1. The time it takes for the request to reach the server



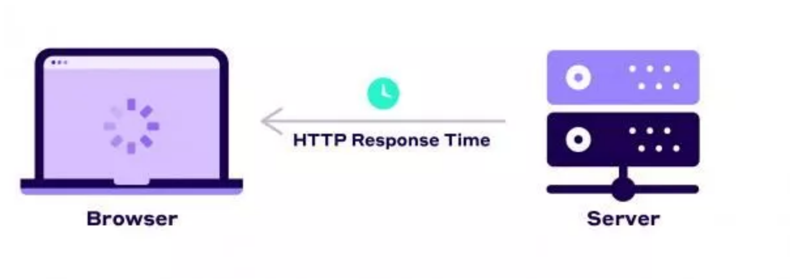
When someone visits your website, their browser sends an HTTP request to your server. In this first stage, a number of factors can cause a delay - slow DNS lookup, physical distance, and the client's internet speed.

2. The time it takes for the server to process the request



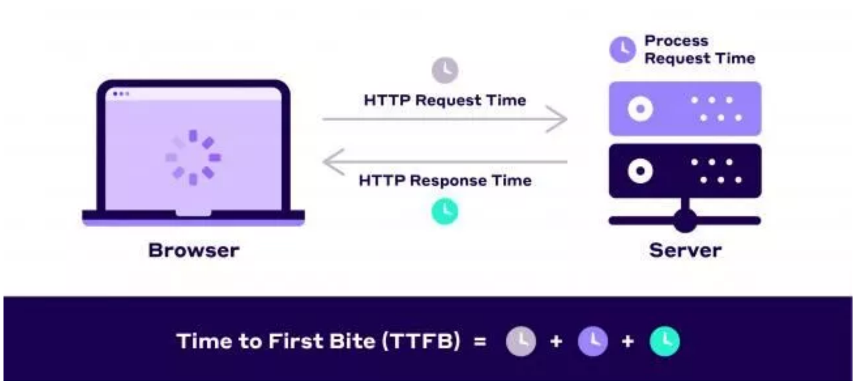
Once the request reaches its destination, the server must process it and generate a response. The process might take some time because of slow database calls, lack of caching layers, excessive scripts that must be executed on the server, a bloated theme, or insufficient server resources.

3. The time it takes for the first byte of data to reach the client



After everything is done, the server has to send the response back to the client. The leading cause of delay here is the slow network speed of both the server and the client.

Because of the entirety of its measurement, TTFB gives a better idea of the user experience in terms of request-response times.



Is TTFB a Metric Worth Focusing on?

There have always been contradictory opinions about this metric's importance.

I mentioned earlier that the SEO and web performance communities had been divided (and probably still are) when it comes to TTFB.

Back in the day, two of the most famous statements on the topic came from John Graham-Cumming from

Cloudflare and Jesse Nickles from LittleBizzy:

> *"Probably the only time TTFB is useful is as a trend. And it's best measured at the server itself so that network latency is eliminated. By examining a trend it's possible to spot whether there's a problem on the web server (such as it becoming overloaded). Measuring TTFB remotely means you're also measuring the network latency at the same time which obscures the thing TTFB is actually measuring: how fast the web server is able to respond to a request."* - **John Graham-Cumming, Cloudflare**.

> *"TTFB is meaningless. It is meaningless because it is a metric that depends completely on the unique environment of each end-user. It is meaningless because 99% of the time, it is being used to describe something that is NOT, in fact, TTFB. It is meaningless because, in reality, it is largely composed of networking elements out of the control of the end-user, web designer, and server administrator alike."* - **Jesse Nickles, LittleBizzy**.

According to these quotes, the problem with TTFB is that it depends on each user's unique environment. While true, this is a feature, not a bug.

Understanding how different people experience your site is crucial if you want to ensure a fast experience for everyone. That's why measuring the TTFB on different devices and in different regions can be so useful.

And while we can't control our users' network connections, we can make their lives much easier. Here's what our CTO has to say on this topic:

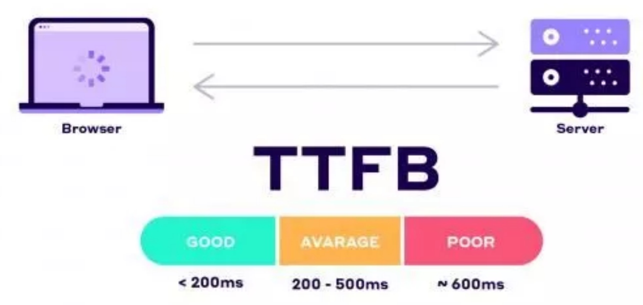
> *"You might think that you do not have direct control over the network latency of your visitors because this depends on their type and quality of the internet connection. However, you have control over how much data they need to transfer - hence you have direct control over the delays caused by network latency. Remove unused code and unnecessary requests."* - **Ivailo Hristov, CTO at NitroPack**.

In short, TTFB can give you insights into how users on slower devices or network connections experience your site. That's what makes it a valuable metric.

What is a Good Time To First Byte?

Before proceeding with the list of tools for measuring TTFB, let's set the benchmarks you should aim for.

- Anything **below 200ms is considered a good TTFB**.
- The 200 to 500ms range is pretty average, especially for dynamic content.
- Anything around 600ms or more usually points to an issue on the server-side, especially if the TTFB is high from locations close to the server.



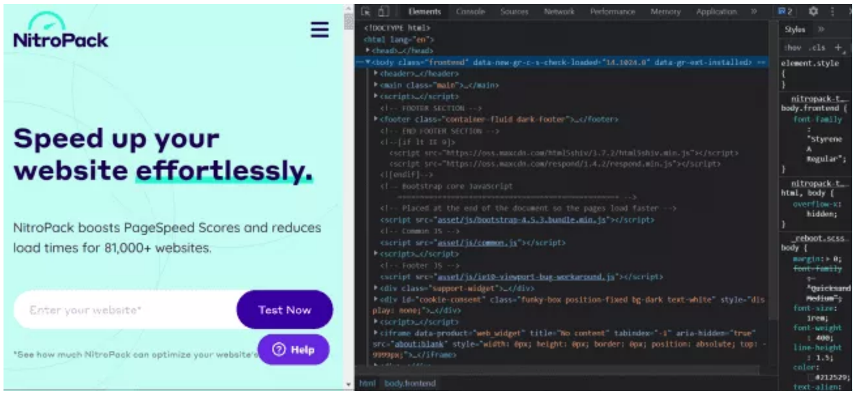
How to Measure TTFB

There are many tools that you can use to measure TTFB. However, keep in mind that each one can give you a slightly different score (due to various factors like testing methodology, testing location, etc.). Stick to one as a baseline and experiment with the others.

Chrome DevTools

Here's how you can check your TTFB via Chrome's DevTools:

Open your website and right-click on a page and select Inspect:



Speed up your website effortlessly.

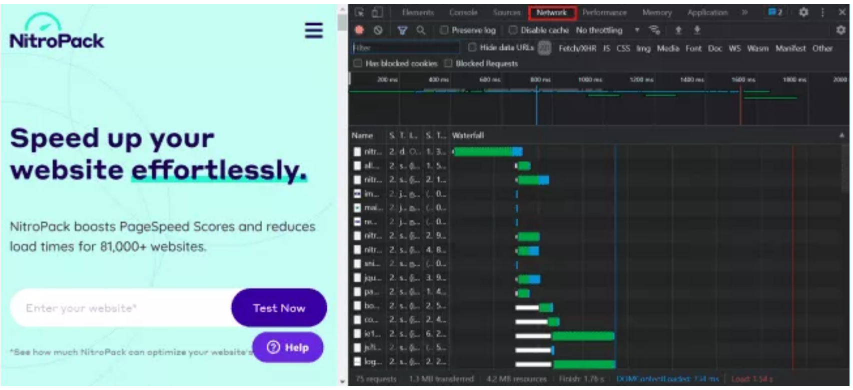
NitroPack boosts PageSpeed Scores and reduces load times for 81,000+ websites.

Enter your website*

*See how much NitroPack can optimize your website's

```
<!DOCTYPE html>
<html lang="en">
 ▸<head>…</head>
··▾<body class="frontend" data-new-gr-c-s-check-loaded="14.1024.0" data-gr-ext-installed> ==
   ▸<header>…</header>
   ▸<main class="main">…</main>
   ▸<script>…</script>
    <!-- FOOTER SECTION -->
   ▸<footer class="container-fluid dark-footer">…</footer>
    <!-- END FOOTER SECTION -->
    <!--[if lt IE 9]>
      <script src="https://oss.maxcdn.com/html5shiv/3.7.2/html5shiv.min.js"></script>
      <script src="https://oss.maxcdn.com/respond/1.4.2/respond.min.js"></script>
    <![endif]-->
    <!-- Bootstrap core JavaScript
      ================================================== -->
    <!-- Placed at the end of the document so the pages load faster -->
    <script src="asset/js/bootstrap-4.5.3.bundle.min.js"></script>
    <!-- Common JS -->
    <script src="asset/js/common.js"></script>
   ▸<script>…</script>
    <!-- Footer JS -->
    <script src="asset/js/ie10-viewport-bug-workaround.js"></script>
   ▸<div class="support-widget">…</div>
   ▸<div id="cookie-consent" class="funky-box position-fixed bg-dark text-white" style="display: none;">…</div>
   ▸<script>…</script>
   ▸<iframe data-product="web_widget" title="No content" tabindex="-1" aria-hidden="true" src="about:blank" style="width: 0px; height: 0px; border: 0px; position: absolute; top: -9999px;">…</iframe>
```

Go to the Network tab and refresh your page:



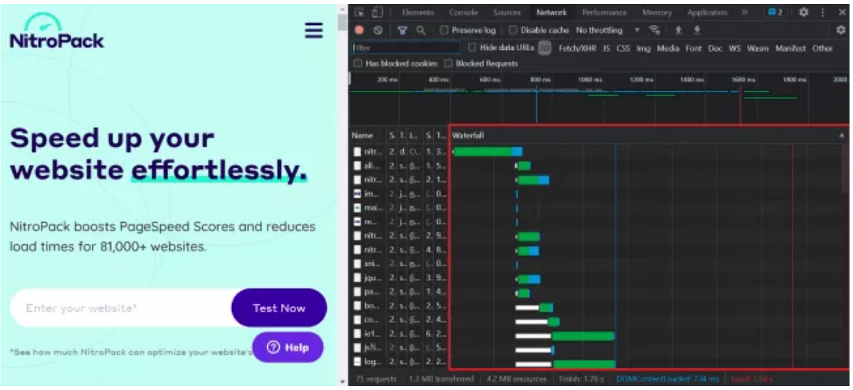
Speed up your website effortlessly.

NitroPack boosts PageSpeed Scores and reduces load times for 81,000+ websites.

Enter your website*

*See how much NitroPack can optimize your website's

Hover over the Waterfall chart:



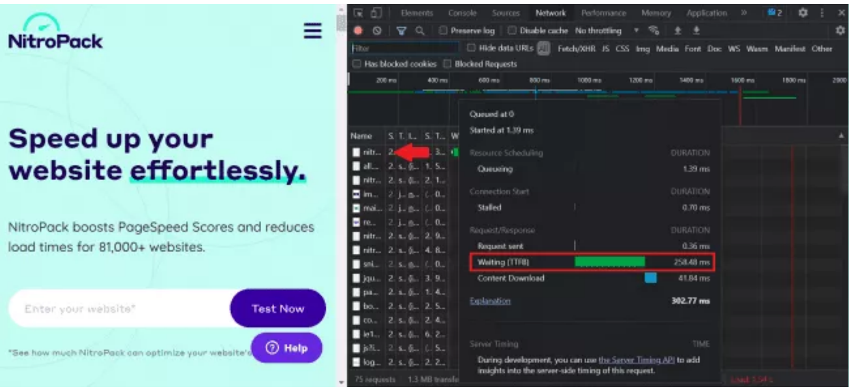
Speed up your website effortlessly.

NitroPack boosts PageSpeed Scores and reduces load times for 81,000+ websites.

Enter your website*

*See how much NitroPack can optimize your website's

And look for the TTFB of the page's main document (in this case, nitropack.io):



NitroPack

Speed up your website effortlessly.

NitroPack boosts PageSpeed Scores and reduces load times for 81,000+ websites.

Enter your website* | Test Now

*See how much NitroPack can optimize your website

Queued at 0

Started at 1.39 ms

Resource Scheduling	DURATION
Queueing	1.39 ms
Connection Start	**DURATION**
Stalled	0.70 ms
Request/Response	**DURATION**
Request sent	0.36 ms
Waiting (TTFB)	258.48 ms
Content Download	41.84 ms
Explanation	302.77 ms

Server Timing — TIME

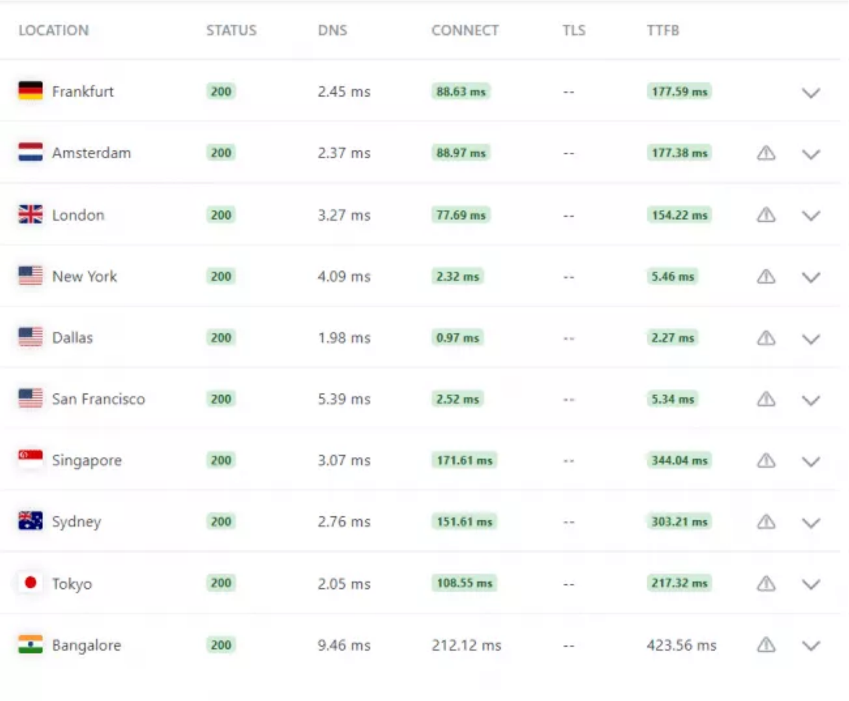
During development, you can use the Server Timing API to add insights into the server-side timing of this request.

The downside of using DevTools is that you're testing from your computer. As a result, the TTFB will be affected by your location and network connection.

If you want to understand how people over the world experience your site, you need a testing tool with more advanced features.

KeyCDN

KeyCDNs' performance test tool is an easy way to see how much distance affects our site's TTFB. If we use the example.com domain, we can see that TTFB varies quite a lot - from under 5ms in the US to over 450ms in India.



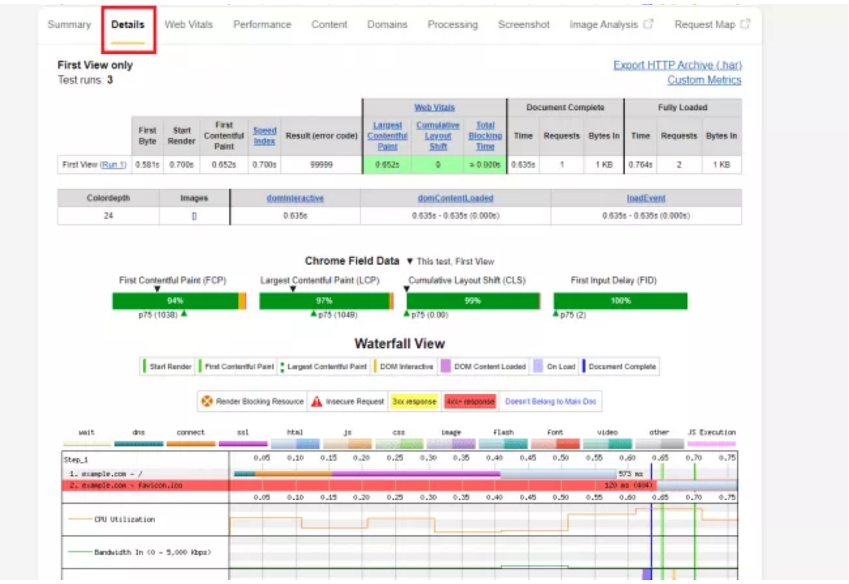
LOCATION	STATUS	DNS	CONNECT	TLS	TTFB
Frankfurt	200	2.45 ms	88.63 ms	--	177.59 ms
Amsterdam	200	2.37 ms	88.97 ms	--	177.38 ms
London	200	3.27 ms	77.69 ms	--	154.22 ms
New York	200	4.09 ms	2.32 ms	--	5.46 ms
Dallas	200	1.98 ms	0.97 ms	--	2.27 ms
San Francisco	200	5.39 ms	2.52 ms	--	5.34 ms
Singapore	200	3.07 ms	171.61 ms	--	344.04 ms
Sydney	200	2.76 ms	151.61 ms	--	303.21 ms
Tokyo	200	2.05 ms	108.55 ms	--	217.32 ms
Bangalore	200	9.46 ms	212.12 ms	--	423.56 ms

These results indicate that example.com is hosted somewhere in the US and likely doesn't provide content from servers near Bangalore or Singapore, for example.

WebPageTest

You can test your website's performance from different locations with more advanced tools like WebPageTest. It allows you to test from over twenty locations while emulating a device with a specific connection.

When you test your page, you can open the waterfall chart. Go to Details and scroll down to Waterfall view:



Summary | Details | Web Vitals | Performance | Content | Domains | Processing | Screenshot | Image Analysis | Request Map

First View only
Test runs: 3

Export HTTP Archive (.har)
Custom Metrics

						Web Vitals			Document Complete			Fully Loaded		
	First Byte	Start Render	First Contentful Paint	Speed Index	Result (error code)	Largest Contentful Paint	Cumulative Layout Shift	Total Blocking Time	Time	Requests	Bytes In	Time	Requests	Bytes In
First View (Run 1)	0.581s	0.700s	0.652s	0.700s	99999	0.652s	0	≥ 0.000s	0.635s	1	1 KB	0.764s	2	1 KB

Colordepth	Images	domInteractive	domContentLoaded	loadEvent
24	[]	0.635s	0.635s - 0.635s (0.000s)	0.635s - 0.635s (0.000s)

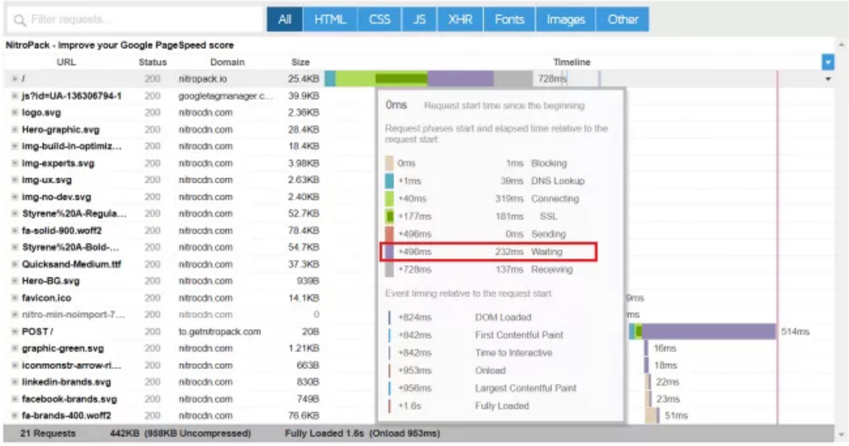
Then click on the main document and see the TTFB:

If you don’t know where your website’s visitors are located, check out the “Geo” and “Technology” reports in Google Analytics or your monitoring tool of choice:

Put simply, you want to find your visitors' locations and check the TTFB from there.

GTmetrix

In GTmetrix, the TTFB is referred to as "wait time." After the test is complete, go to the Waterfall tab and hover over it to see the statistics:



NitroPack - Improve your Google PageSpeed score

URL	Status	Domain	Size	Timeline
/	200	nitropack.io	25.4KB	728ms
js?id=UA-136306794-1	200	googletagmanager.c...	39.9KB	
logo.svg	200	nitrocdn.com	2.36KB	
Hero-graphic.svg	200	nitrocdn.com	28.4KB	
img-build-in-optimiz...	200	nitrocdn.com	18.4KB	
img-experts.svg	200	nitrocdn.com	3.98KB	
img-ux.svg	200	nitrocdn.com	2.63KB	
img-no-dev.svg	200	nitrocdn.com	2.40KB	
Styrene%20A-Regula...	200	nitrocdn.com	52.7KB	
fa-solid-900.woff2	200	nitrocdn.com	78.4KB	
Styrene%20A-Bold-...	200	nitrocdn.com	54.7KB	
Quicksand-Medium.ttf	200	nitrocdn.com	37.3KB	
Hero-BG.svg	200	nitrocdn.com	939B	
favicon.ico	200	nitrocdn.com	14.1KB	9ms
nitro-min-noimport-7...	200	nitrocdn.com	0	ms
POST /	200	to.getnitropack.com	20B	514ms
graphic-green.svg	200	nitrocdn.com	1.21KB	16ms
iconmonstr-arrow-ri...	200	nitrocdn.com	663B	18ms
linkedin-brands.svg	200	nitrocdn.com	830B	22ms
facebook-brands.svg	200	nitrocdn.com	749B	23ms
fa-brands-400.woff2	200	nitrocdn.com	76.6KB	51ms

21 Requests 442KB (958KB Uncompressed) Fully Loaded 1.6s (Onload 953ms)

0ms Request start time since the beginning

Request phases start and elapsed time relative to the request start:

0ms	1ms	Blocking
+1ms	39ms	DNS Lookup
+40ms	319ms	Connecting
+177ms	181ms	SSL
+496ms	0ms	Sending
+496ms	232ms	Waiting
+728ms	137ms	Receiving

Event timing relative to the request start:

+824ms	DOM Loaded
+842ms	First Contentful Paint
+842ms	Time to Interactive
+953ms	Onload
+956ms	Largest Contentful Paint
+1.6s	Fully Loaded

How to Reduce TTFB

You know how to measure TTFB, now let's dive into some techniques that will help you reduce it.

Rely on a Fast Host Server

The first step towards improving TTFB is relying on a high-quality Managed Service Provider (MSP) and a fast host server.

However, many owners decide to go with a shared hosting service. While cheaper, with this type of hosting, you share your server's resources (CPU, bandwidth, memory) with other websites.

This can slow down response times, so consider investing in a dedicated hosting service.

Things you should look for when upgrading/choosing your hosting plan are high availability, security, and support. Check this comparison article of the best hosting providers in 2022.

> **Friendly Reminder:** Don't forget to perform the tests we showed earlier after selecting a hosting plan.

Decrease the Number of 3rd Party Plugins and Choose a Fast Theme (for WordPress and other CMS sites)

Third-party libraries and bloated themes can seriously hurt response times.

It's tempting to add a lot of features to your site and use a theme with dynamic elements. Especially when there are thousands of options available. However, adding too many plugins and using a heavy theme increases the amount of code that has to be executed, resulting in slower page speed.

This is a common problem when working with content management systems like WordPress or Shopify.

That’s why it’s essential to choose a lightweight theme and audit each plugin or app before installing it. If page speed takes a hit after adding a new plugin, it might be better to choose another option. In a lot of cases, the added functionality isn’t worth the massive server overhead.

Implement Caching Layers

One of the easiest ways to reduce TTFB, and speed up your website in general, is setting up a caching layer.

Caching means storing a copy of a site’s resources in a different location than the origin server. Once cached, resources don’t need to be re-downloaded from the server every time. This leads to faster load times and less server overhead.

Two popular caching techniques are full page caching and object caching.

Full page cache means that the entire HTML for the page will be cached. Subsequent requests for the page will be served from the cache instead of being processed and re-built again, thus speeding up the response times.

Object caching refers to storing database queries, which results in faster data retrieval from the database.

You can set up caching by hand via your server’s HTTP headers or use a service like NitroPack to take care of the process for you.

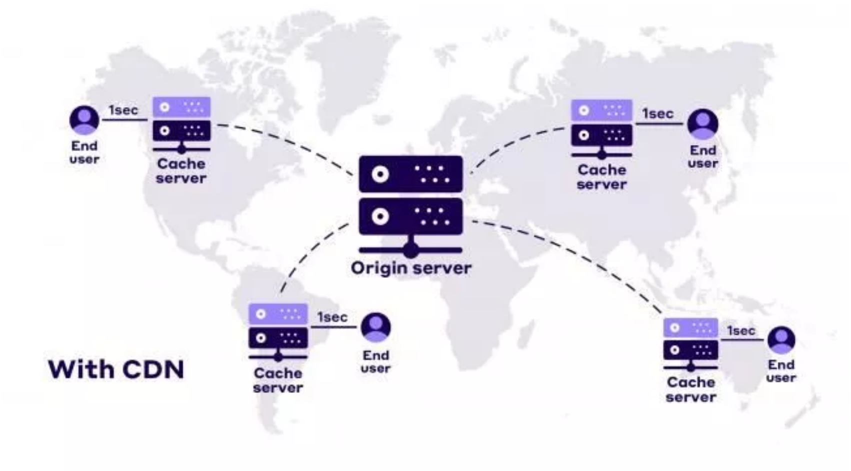
Utilize a Content Delivery Network (CDN)

A CDN consists of a number of geographically distributed servers at various points of presence (PoP) around the world.

Put simply:

A CDN helps you improve the network latency issue caused by the large physical distance between the client and server.

For example, let's say your origin server is located in the USA, but most of your clients are from India. With a CDN, your specific resources or even entire pages can be served from the server closest to them, instead of the origin in the USA.



This reduces the impact of physical distance and speeds up your website.

Use Minification and Compression

Minifying and compressing your site's resources before sending them over to the browser drastically reduces transfer time since the files are much smaller.

Minification means removing unnecessary parts like whitespace and comments from the code.

Code compression refers to applying algorithms to rewrite the files' binary code, using fewer bits than the original.

Code files usually have a lot of comments, whitespace, and repeated text, all of which are unnecessary elements that only slow down your site's load time.

When it comes to implementing these optimizations, some hosting providers apply them by default, so it's worth checking with your provider.

Additionally, you can use a service like NitroPack which can make the necessary changes for you

Optimize Your Site's Images

Did you know that about 50% of all bytes on an average page are image bytes?

Applying different image optimization techniques is another way to reduce the amount of data that needs to be transferred and, hence, speed up response times.

Image compression is one of the methods that can reduce the size of image files. There are two types of compression:

- Lossy compression removes some of the data from the original file.   This makes the files much lighter at the expense of quality. JPEG and GIF are examples of lossy image types.
- Lossless compression maintains roughly the same image quality by removing unnecessary metadata from the files. While the image quality is largely unaffected, image file size also doesn't get reduced. RAW, BMP, and PNG are all lossless image formats.

Image compression tools, such as Optimizilla and imagemin, can help you reduce image file sizes.

Another technique you can benefit from is using Next-Gen image formats like JPEG 2000, JPEG XR, AVIF, and WebP.

The difference between them and the well-known image formats is that they have better compression and quality characteristics.

Put simply:

They're smaller in size while still keeping a comparable quality level to the older formats.

**Currently, WebP is the only modern format that has enough browser support to be viable for most users.*

Conclusion

Whether people like it or not, TTFB plays its role in web performance.

Without a doubt, there are some variables that affect TTFB, but that doesn't make it meaningless as it gives us insights into how real users experience our site. And that's what you want to focus on when it comes to web performance optimization.

You might not be able to change the unique environment of each user, but you can implement some optimization techniques that will improve the network latency.

But for that, you might need some help if you aren't into speed optimization.

On that note, NitroPack is the all-in-one solution that will speed up your website in a couple of minutes. You can try it (for free) and set it up in less than 5 minutes. No coding or tech skills required.

NitroPack takes care of your site's speed optimization by applying all kinds of different techniques (some of them were mentioned in the article):

- Caching;
- Built-in CDN;
- Complete image optimization;
- HTML, CSS and JS minification & compression;
- Critical CSS, DNS prefetching, preloading, and more.

You can test NitroPack for free and see for yourself how easy it is to optimize your website's speed.

Niko Kaleev
Content Writer

Niko is one of the NitroPack storytellers. He is passionate about writing quality content and turning complex optimization concepts into engaging articles.