

Why is the Time to First Byte (TTFB) so high? *

TTFB (Time To First Byte) generally refers to the amount of time it takes for the server to respond to a client's request and send the first byte of data back to the client. In other words, it's the time between when a user enters a URL in their web browser and when the first byte of data is received from the server.

TTFB is measured in milliseconds and includes the time it takes for the server to process the request and generate a response, as well as the time it takes for the response to travel from the server to the client. Typically, TTFB is measured from the moment the client sends the request to the server until the moment the server sends the first byte of the response back to the client.

Understanding TTFB: Definitions and Variations by Tool and Environment

It is important to recognize that how TTFB is measured and reported can vary based on the tool, browser, app, or service used to measure it. Some tools measure TTFB as the time elapsed between when the request is sent and when the server begins sending the response, while others may measure it as the time elapsed between when the request is sent and when the first byte of the response is received by the client.

Additionally, TTFB can be affected by factors like network latency, server processing time, and caching. As a result, establishing a single universally applicable definition for this metric is difficult. TTFB measurements may, therefore, vary depending on the specific tool or method used to measure it and other factors related to the network, server, and client environment.

For instance, in [GTmetrix](#), TTFB is measured as the sum of "Redirect Duration" + "Connection Duration" + "Backend Duration." On the other hand, when using [WebPageTest](#) to download a page, you might encounter an unexpected surprise. Specifically, WebPageTest may report the Time To First Byte as the time the "H" was received, rather than the time the page itself was actually sent. This discrepancy becomes apparent due to the 10-second wait.

Common Causes of a High TTFB

There are several reasons why TTFB can be high. For example:

1. **Server processing time:** The server might take too long to process the request and generate a response. This may happen when the server is overloaded, the code is poorly optimized, or the database queries take too long to execute.
2. **Network latency:** TTFB may be affected by the distance between the user's device and the server and any network congestion or bottlenecks that exist between them.
3. **DNS lookup time:** Before a user's device connects to a server, it must first lookup the server's IP address via a Domain Name System (DNS) server. This can add additional time to the TTFB if the DNS lookup takes a long time to complete.
4. **Content delivery network (CDN) issues:** When a website is using a CDN, the TTFB may be impacted by the CDN's caching policies, network latency between the CDN and the origin server, or the number of hops required to route the request to the appropriate server.
5. **Third-party resources:** If a website depends on resources hosted on third-party servers, it may affect TTFB.

Diagnosing a high TTFB requires investigating these potential causes and making the necessary optimizations to improve server response time, reduce network latency, and streamline content delivery. It's important to understand that upgrading your service plan or server hardware won't reduce load times if your web application code is inefficient.

More Resources

- <https://gtmetrix.com/reduce-initial-server-response-time.html>
- <https://moz.com/blog/improving-search-rank-by-optimizing-your-time-to-first-byte>
- <https://web.dev/time-to-first-byte/>

Revision #8

Created 11 October 2022 01:44:51 by openVservers

Updated 24 July 2024 16:37:21 by openVservers