

## #LatencyTipOfTheDay: MOST page loads will experience the 99%'lie server response

6-8 minutes

---

Yes. **MOST** of the page view attempts will experience the 99%'lie server response time in modern web applications. You didn't read that wrong.



This simple fact seems to surprise many people. Especially people who spend much of their time looking at pretty lines depicting averages, 50%'lie, 90%'lie or 95%'lies of server response time in feel-good monitoring charts. I am constantly amazed by how little attention is paid to the "higher end" of the percentile spectrum in most application monitoring, benchmarking, and tuning environments. Given the fact that most user interactions will experience those numbers, the adjacent picture comes to mind.

Oh, and in case the message isn't clear, I am also saying that:

- **MOST** of your users will experience the 99.9%'lie once in ten page view attempts
- **MOST** of your users will experience the 99.99%'lie once in 100

page view attempts

- **More than 5%** of your shoppers/visitors/customers will experience the 99.99%'lie once in 10 page view attempts.

**So, how does this work: Simple. It's math.**

For most (>50%) web pages to possibly avoid experiencing the 99%'lie of server response time, the number of resource requests per page would need to be smaller than 69.

Why 69?

Here is the math:

- The chances of a single resource request avoiding the 99%'lie is 99%. [Duh.]
- The chances on all N resource requests in a page avoiding the 99%'lie is  $(0.99^N) * 100\%$ .
- $(0.99^{69}) * 100\% = 49.9\%$

So with 69 resource requests or more per page, **MOST** (> 50% of) page loads are going to fail to avoid the 99%'lie. And the users waiting for those pages to fill will experience the 99%'lie for at least some portion of the web page. This is true even if you assume perfect parallelism for all resource requests within a page (non of the requests issued depend on previous requests being answered). Reality is obviously much worse than that, since requests in pages do depend on previous response, but I'll stick with what we can claim for sure.

The percentage of page view attempts that will experience your 99%'lie server response time (even assuming perfect parallelism in all requests) will be bound from below by:

% of page view attempts experiencing 99%'lie  $\geq (1 - (0.99^N)) * 100\%$

Where N is the number of [resource requests / objects / HTTP GETs] per page.

## **So, How many server requests are involved in loading a web page?**

The total number of server requests issued by a single page load obviously varies by application, but it appears to be a continually growing number in modern web applications. So to back my claims I went off to the trusty web and looked for data.

According to some older stats collected for a sample of several billions of pages processed as part of Google's crawl and indexing pipeline, the number of HTTP GET requests per page on "Top sites" hit the obvious right answer (42, Duh!) in mid 2010 (see [\[1\]](#)). For "All sites" it was 44. But those tiny numbers are so 2010...

According to other sources, the number of objects per page has been growing steadily, with ~49 around the 2009-2010 timeframe (similar to but larger than Google's estimates), and crossed 100 GETs per page in late 2012 (see [\[2\]](#)). But that was almost two years ago.

And according to a very simple and subjective measurement done with my browser just now, loading this blog's web page (before this posting) involved 119 individual resource requests. So nearly 70% of the page views of this blog are experiencing blogger's 99%'lie.

To further make sure that I'm not smoking something, I hand checked a few common web sites I happened to think of, and none of the request counts came in at 420:

<b>Site</b>	<b># of requests</b>	<b>page loads that would experience the 99%'lie</b> <b><math>[(1 - (.99^N)) ^ *</math></b>

		<b>100%]</b>
amazon.com	190	85.2%
kohls.com	204	87.1%
jcrew.com	112	67.6%
saksfifthavenue.com	109	66.5%
--	--	--
nytimes.com	173	82.4%
cnn.com	279	93.9%
--	--	--
twitter.com	87	58.3%
pinterest.com	84	57.0%
facebook.com	178	83.3%
--	--	--
google.com (yes, that simple noise-free page)	31	26.7%
google.com search for "http requests per page"	76	53.4%

So yes. There is one web page on this list for which most page loads will not experience the 99%'lie. "Only" 1/4 of visits to google.com's clean and plain home page will see that percentile. But if you actually use google search for something, you are back on my side of the boat....

**What the ^%&\*! are people spending their time looking at?**

Given these simple facts, I am constantly amazed by the number of people I meet who never look at numbers above the 95%'ile, and spend most of their attention of medians or averages. Even if we temporarily ignore the fact that the 95%'ile is irrelevant (as in too optimistic) for more than half of your page views, there is less than a 3% chance of modern web app page view avoiding the 95%'ile of server response time. This means that the 90%'ile, 75%'ile, median, and [usually] the average are completely irrelevant to 97% of your page views.

**So please, Wake Up! Start looking at the right end of the percentile spectrum...**

#### **References:**

- [1] [Sreeram Ramachandran: "Web metrics: Size and number of resources"](#), May 2010.
- [2] ["Average Number of Web Page Objects Breaks 100"](#), Nov 2012

#### **Discussion Note:**

It's been noted by a few people that these calculations assume that there is no strong time-correlation of bad or good result. This is absolutely true. The calculations I use are valid if every request has the same chance of experiencing a larger-than-percentile-X result regardless of what previous results have seen. A strong time correlation would decrease the number of pages that would see worse-than-percentile-X results (down to a theoretical X% in theoretically perfect "all responses in a given page are either above or below the X%'ile" situations). Similarly, a strong time anti-correlation (e.g. a repeated pattern going through the full range of response times values every 100 responses) will increase the number of pages that would see a worse-than-percentile-X result, up to a theoretical 100%.

So in reality, my statement of "most" (and the 99.3% computed for cnn.com above) may be slightly exaggerated. Maybe instead of >50% of your page views seeing the 99%'lie, it's "only" 20% of page views that really are that bad... ;-)

Without time-correlation (or anti-correlation) information, the best you can do is act on the basic information at hand. And the only thing we know about a given X %'ile in most systems (on its own, with no other measured information about correlative behavior) is that the chances of seeing a number above it is  $(100\%-X)$ .