CLOUDFLARE®

Cloudflare CDN

A global content delivery network with unique
performance optimization capabilities

Overview

Cloudflare provides a global content delivery network (CDN) with unique performance optimization capabilities: we cache static content, accelerate dynamic content, and make it easy to optimize outbound content. Cloudflare goes beyond a traditional CDN because, in addition to bringing your content closer to your site's visitors, Cloudflare also optimizes content for device, browser, and bandwidth needs. We've built a best-in-class, global infrastructure from the ground up using only next-generation, high-performance equipment—no legacy software or hardware. The result is a CDN that's easier to set up, more affordable, and built to outperform any legacy CDN on the market.

Cloudflare makes websites run faster, and having a fast site is an important part of visitor satisfaction. Site speed is crucial to providing a positive experience for website visitors because slow load times result in higher bounce rates, shorter time spent on the site, and lower conversion rates. To put it simply, latency hurts your bottom line.

In February of 2012, the multinational retailer Walmart conducted an analysis to understand how website page performance affects e-commerce conversion rates[1]. The study focused on Walmart's own e-commerce store: walmart.com. The webstore did $7.7 billion in sales, hit billions of page requests, and hosted millions of active product SKUs in 2012.

This study revealed a shocking relationship between page load times and conversion rates. As the graph below shows, there is a sharp decline in conversion rates as load times increase from 1 to 4 seconds.

Conversion rate vs. load time Load time in seconds against conversion rate percentage

Load time (seconds)
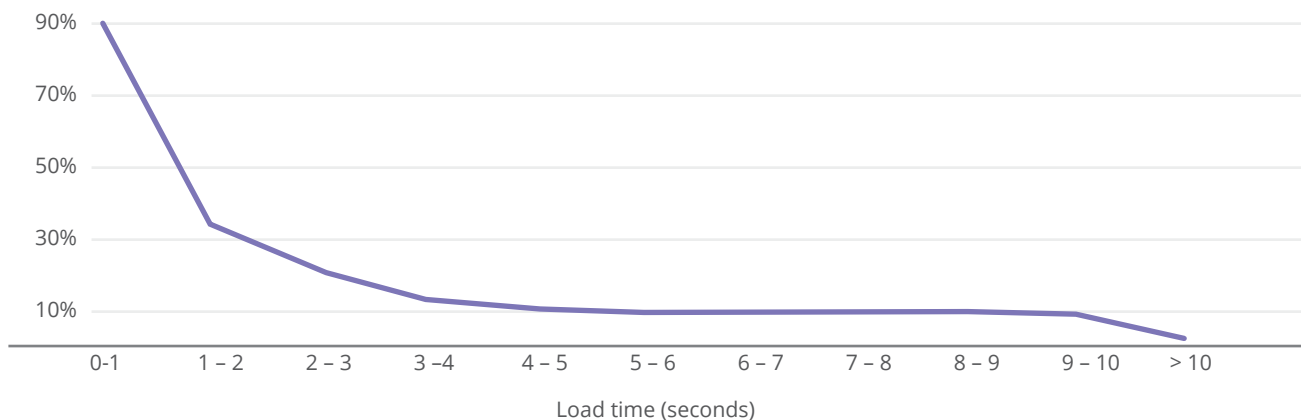
Cloudflare's performance features dramatically improve load times for users accessing your static and dynamic web content. This increases customer satisfaction, and potentially drives conversion rates up. Because our CDN reduces hops and lowers latency, a request on Cloudflare's network takes less than 23 milliseconds to serve on average.

This white paper will cover:

- Global distributed network
- Static content caching
- Custom caching management
- Edge side code

- Dynamic content acceleration
- Front-end optimization and client intelligence
- Reporting

1. Walmart.com study (Feb, 2012) http://minus.com/msM8y8nyh/2e

Global distributed network

Cloudflare has created a purpose-built network for today's web. Our global network stands out from legacy CDNs in both performance and security capabilities. Cloudflare inspects traffic by operating as a reverse proxy, and because we fully own and operate our network's equipment, we can ensure that security is built into every layer and every protocol.

Cloudflare's network has 102 PoPs globally and continues to grow our network
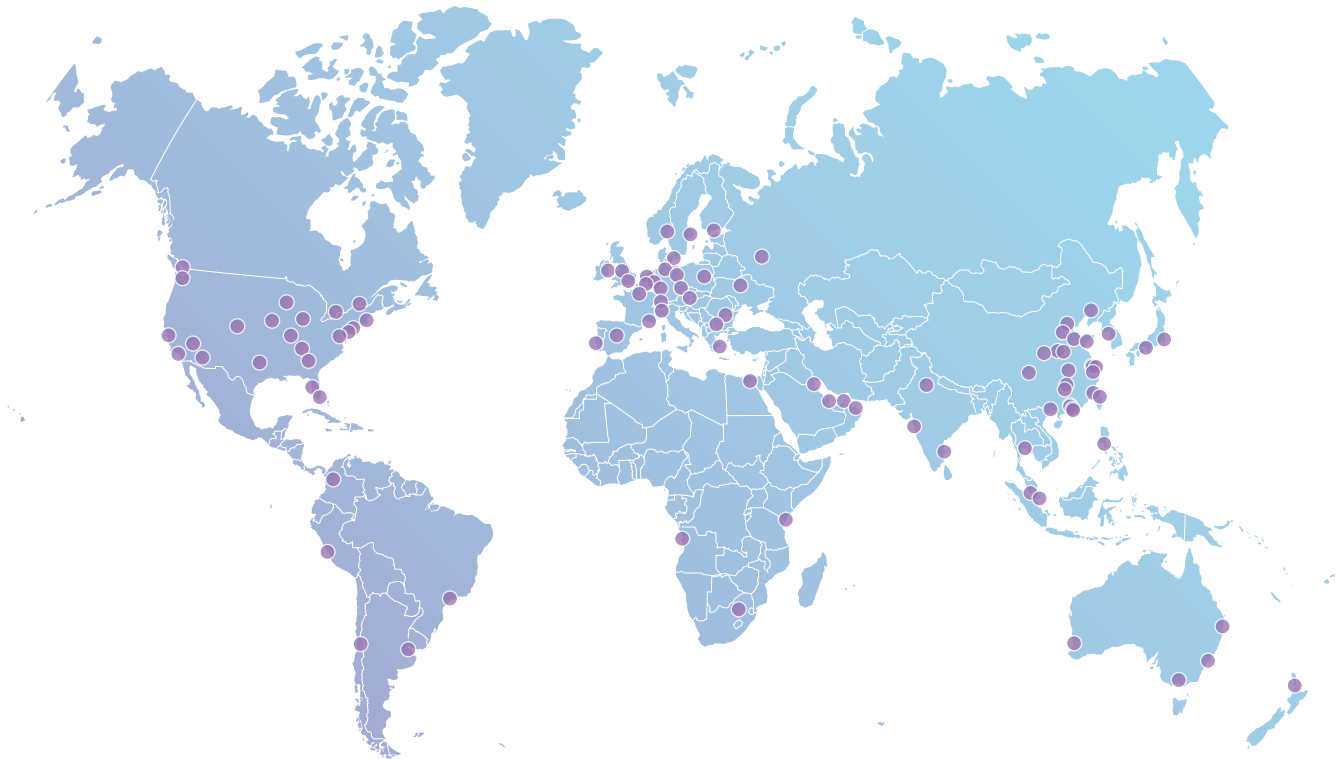
Our CDN is unique because it is a massive horizontally scaled architecture in which every node can perform DNS requests, security checks, and performance transformations. Additionally, Cloudflare employs Anycast routing to ensure web users are automatically routed to their nearest data center and around any failures. The combination of this architecture and network produces a reliable, high-performance service.

Cloudflare makes any website, web app and API faster and safer around the world. Third party performance testing shows that Cloudflare outperforms competitors on average in major regions across the world. According to Cedexis, Cloudflare is the top performing CDN, delivering content faster than the competition by 18% in North America, 23% in Europe, and 69% in Asia.

Cloudflare outperforms competitors' performance Cedexis, 2014
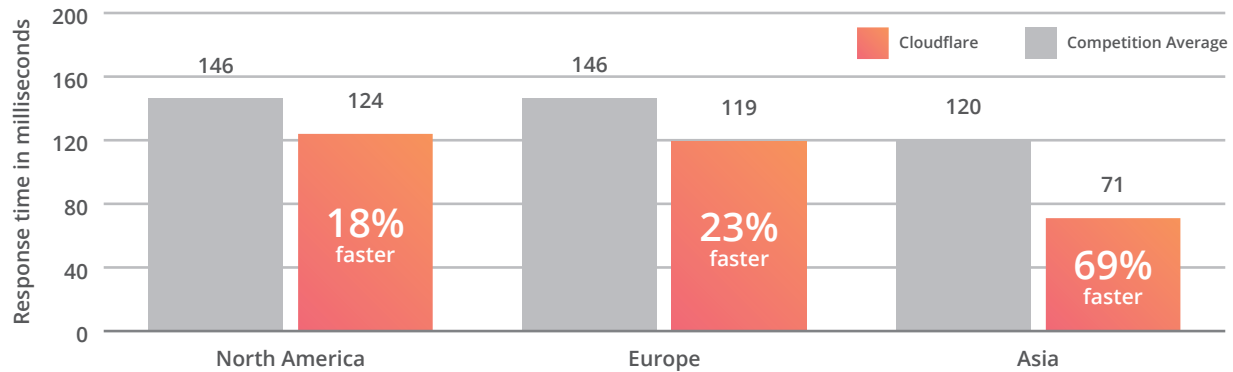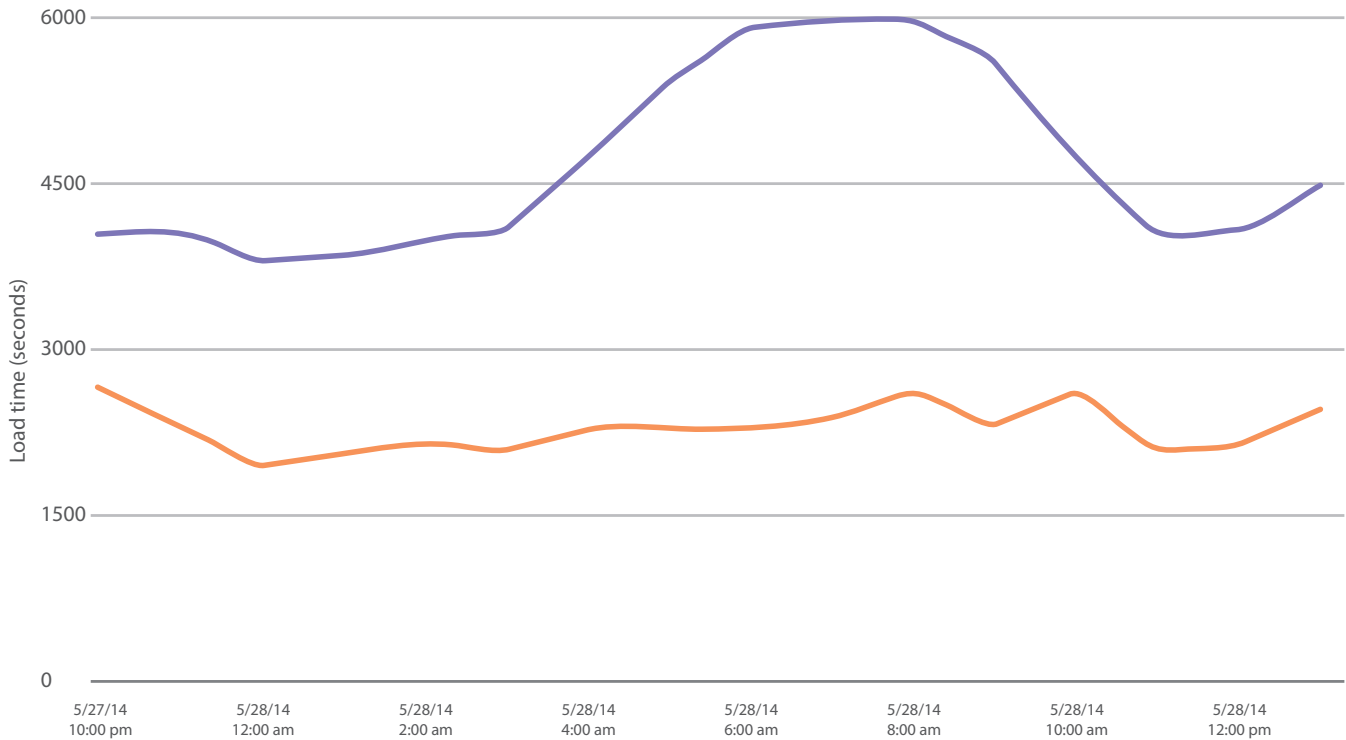

Static content caching

Cloudflare is a zero configuration CDN. As soon as you activate your site on Cloudflare, we cache your site's static content by default. Our CDN caches more than 35 file extensions automatically, and allows users to write page rules to extend the list of file extensions we cache.

Cloudflare serves site requests directly from cached content in whichever data center is physically closest to a site visitor, and because Cloudflare has data centers around the world, this means that whether you are in Chicago, Prague, Frankfurt or Singapore, web pages are delivered quickly, even when the origin web server is thousands of miles away. For example, a website might be hosted in the US, but accessed mainly by web visitors in the UK. With Cloudflare, the site will be served from a UK data center eliminating costly delays. Cloudflare's ability to make a web site appear to be hosted close to web surfers is key to accelerating web surfing.

In addition to faster load times, a website on Cloudflare sees 65% fewer requests to the origin yielding a 60% reduction in bandwidth consumption on your origin web servers.

By using Cloudflare's network, Big 5 Sporting Goods reduced their site's load time by approximately 100%. The graph below shows a comparison of the load time for content from http://www.big5sportinggoods.com cached by Cloudflare (orange line) versus the load time when the request is sent directly to Big 5 Sporting Goods' origin (blue line).

Big 5 Sporting Goods

Custom caching management

In addition to caching static content, Cloudflare allows you to write page rules to cache HTML content for specific URLs directly in our administrative control panel. Use cases for page rules could be creating a directory for static content, appending a unique file extension to static pages, or adding a query parameter to mark HTML content as static. The following examples are rules that can be created for each of those options:

- *example.com/static/* [/static/ subdirectory for static HTML pages]
- *example.com/*.shtml [.shtml file extension to signify HTML that is static]
- *example.com/*?*static=true* [adding static=true query parameter]

You can also create custom cache keys allowing you to cache a different version of your page based on browser language, cookie, or any arbitrary HTTP request header. For example, a cache key rule can be implemented that would require Cloudflare to serve different content for logged in, versus not logged in users.

Caching settings such as expiration time, or time to live (TTL), can be customized for each page rule. Cloudflare's enterprise service plan offers a TTL as low as 30 seconds. The purge feature allows you to instantly purge content from Cloudflare's cache to ensure we are serving the latest version of your site's content. Custom page rules and caching settings are easily managed in both the Cloudflare administrative interface, and the API. All configuration changes take effect immediately around the globe.

Edge side code

Cloudflare's system is built from the ground-up for performance and scalability. With Edge Side Code, Cloudflare deploys powerful logic for customers to alter HTTP requests and responses on the fly — without added latency. Any part of a request or response can be altered according to a customer's rules, enabling a range of options to tightly integrate Cloudflare's edge into applications, which improves serving performance for users.

This has been used to expand what can be cached for visitors according to user agent or geographic region. The result is improved performance for users. Examples of what customers have implemented include:

- Serving a different version of a cached page based on request headers (e.g. logged-in cookie, or browser user agent)
- Including or excluding page widgets based on the requestor's country
- Triggering "cache warming" — loading a page's related images/assets into Cloudflare's cache
- Redirecting a visitor to a particular URL based on HTTP referrer (e.g. go to a different part of a website if referred from Twitter)
- Adding, removing, or altering cookies
- Directing different URL paths to different origin servers
- Load-balancing logic — based on geographic region or automatic failover
- Sharding subdomains over a number of origin server to allow for more effective origin caching
- Sanitizing HTTP headers for requests (e.g. removing the X-Powered-By header)

Dynamic content acceleration (Railgun™)

Cloudflare WAN protocol (dynamic content compression)

HTTP/SPDY

Origin server

Cloudflare data center

Visitors
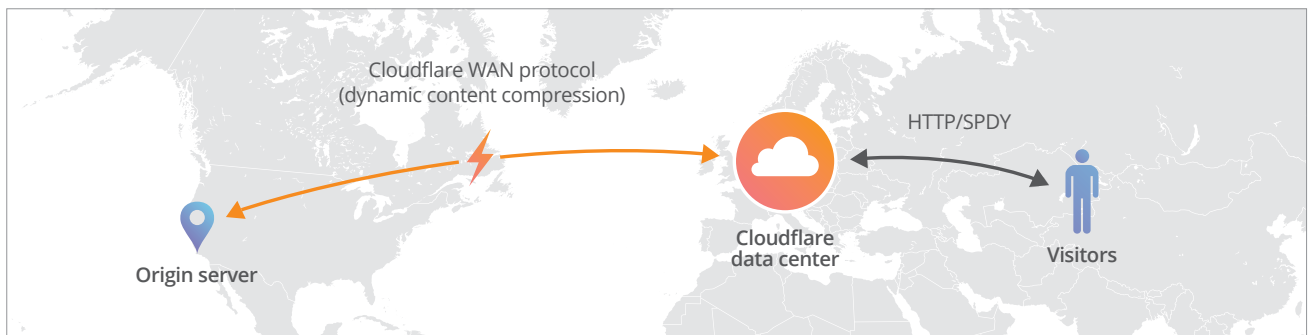
What Railgun does

Approximately 1/3 of requests made to Cloudflare are for dynamic content, and have to be sent to the origin server for processing. This happens because many web pages are not cacheable, whether due to misconfiguration, or, more commonly, because the web page changes frequently or is personalized.

Railgun accelerates and caches previously uncacheable web pages so that even if the origin server must be consulted, web pages are delivered quickly. Railgun even works for rapidly changing pages like news sites and personalized content.

For websites that present dynamically generated content it is impossible to cache the entire page; however, Cloudflare research has found the differences between page versions is often minimal. For example, the New York Times home page changes throughout the day as news stories are written, but the boilerplate HTML of the page mostly stays the same, and many stories stay on the front page all day.

For dynamically changing sites, the boilerplate HTML is the same with only small pieces of content changing (such as a person's Twitter timeline or Facebook news feed). This means there's a huge opportunity to compress web pages for transmission if the unchanging parts of a page can be detected, and only the differences transmitted.

How Railgun works

Railgun accelerates connections between Cloudflare data centers and an origin server speeding up requests that can't be served directly from Cloudflare cache.

Railgun tracks changes to pages down to the byte, and only sends the bytes that have actually changed across the network. When a request is made to a Cloudflare server for a web page that is not in cache, Cloudflare makes an HTTP connection to the origin server to request the page. It's that HTTP connection that Railgun accelerates and secures. Railgun consists of two software components: the "Listener" and the "Sender". The Railgun Listener is installed at your web host on an origin server. It's a small piece of software that runs on a standard server, and services requests from Cloudflare using the encrypted, binary Railgun protocol. The Railgun Sender is installed in all Cloudflare data centers around the world, and maintains connections with Railgun Listeners.

When an HTTP request comes in that must be handled by an origin server, Cloudflare determines whether it is destined for a Railgun-enabled website. If not, standard HTTP is used, but, if so, the HTTP request is routed to the Railgun Sender for handling.

The Railgun Sender turns the request into a compressed, binary chunk that's transmitted to the corresponding Railgun Listener. The Railgun Listener handles the request, and performs an HTTP request to the origin server. From the origin server's perspective, it's as if the HTTP connection came directly from Cloudflare, but because it comes from inside the hosting partner's infrastructure, the request suffers no latency related delay.

Railgun connections are secured by TLS to prevent eavesdropping, and secured by certificates preventing man-in-the-middle attacks. In order to eliminate the slow start up of a TCP connection, the TCP connection between Cloudflare and the origin server is kept alive so it can be reused for subsequent requests.

Railgun requests are multiplexed onto the same connection, and can be handled asynchronously. This means that Railgun is able to handle many simultaneous requests without blocking, which maximizes the use of the TCP connection.

Railgun achieves up to a 99.6% compression ratio for previously uncacheable web objects, (taking, for example, a 100k web page down to 400 bytes) and a speedup of over 700%.

Front end optimization and client intelligence

Cloudflare customers can use our one-click features to perform a wide range of content optimizations from removing unnecessary characters from HTML, CSS, and JavaScript, to reducing the number of connections needed to serve 3rd party widgets (e.g. Facebook, Twitter, ad servers, etc). Cloudflare automatically detects

the type of browser and connection a visitor is using, and delivers your content the fastest way possible. Your pages look the same as they did before, including any mobile version, but they are optimized for the web visitors' desktop or mobile environment.

Reporting

Cloudflare provides insights to website traffic that cannot be obtained from other analytics programs. In addition to visitor analytics, customers can monitor threats and search engine crawlers. For most websites, these crawlers make up 20% to 50% of traffic. It's traffic every website should understand, but most analytics services ignore. We display the analytics report in a user friendly, easy-to-read interface. Traffic to a site can be sorted by time, visitor type, and traffic types. Cloudflare includes the geographic location of a site's visitors allowing businesses to see which countries their visitors come from. Our personalized analytics reports also include exact speed benefits and savings from when the site transitioned to Cloudflare.

All enterprise customers receive raw log access for every user request, and this is available via a secure SSH FTP connection.

Enterprise engineering support

As part of the enterprise service plan, all Cloudflare enterprise customers receive an onboarding session with a Cloudflare solutions engineer to optimize your performance settings. Cloudflare's engineers can work with you to craft a configuration plan that best meets your needs.

For more information on pricing and plans, contact our sales team at:

+1 888 99 FLARE (US office), +44 20 3713 4479 (UK office) or email enterprise@cloudflare.com.
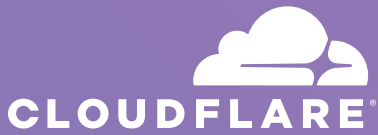
CLOUDFLARE®

1 888 99 FLARE | enterprise@cloudflare.com | www.cloudflare.com