

SOLUTIONS

Why your APIs need an edge cloud

Caching ecommerce APIs to optimize web and mobile performance

Overview

Consumers have come to expect more from their online shopping experiences. They have little tolerance for slow sites that feature generic content. To engage these consumers and drive conversion rates, you need fast response times, personalized interactions, and device-specific user experiences.

Application Program Interfaces (APIs) are a critical component to any successful ecommerce platform designed to achieve these goals. APIs are the building blocks used to send, receive, and modify business-critical data. But, while APIs enable a more dynamic shopping experience, performance is often a challenge.

Content delivery networks (CDNs) are typically used to enhance web and mobile performance. But, there's a problem — legacy CDNs cannot cache frequently and unpredictably changing content like that delivered via API.

This is where Fastly comes in. As an edge cloud platform, Fastly allows you to process, serve, and secure data at the edge of your network. That enables us to cache content previously considered “uncacheable,” including API content like catalog, pricing, and, inventory data. In this brief you will learn how Fastly can cache APIs to optimize performance, enhance personalized interactions, and dramatically reduce infrastructure costs without replacing legacy solutions.

The heat is on

Online retailers face several technical challenges in today's marketplace. To survive and thrive, you must implement the elements and best practices that will help you deliver an optimized shopping experience — all the while controlling both capital expenditure (CapEx) and operating expenditure (OpEx).

Speed matters

Consumers expect a snappy online shopping experience. Otherwise they quickly migrate to the competition. One survey concludes that 20% of end-users will abandon a mobile site that takes longer than three seconds to load.¹ Several other studies have been published that show a strong correlation between slow page load times and cart abandonment. Given that speed is such an important factor, the tradeoff of fast delivery for the ability to serve constantly changing content is becoming increasingly problematic.

The New York Times

When The New York Times sends news alerts to its mobile users, 20 to 30 million push alerts are sent out in a single minute. The Times uses Fastly to mirror those alerts and take load off their central systems. The company saves around \$25,000 per month by putting Fastly in front of its alerts API.

¹ “The Very Real Costs of Bad Website Performance,” Aberdeen Research, August 2016

WHY YOUR APIS NEED AN EDGE CLOUD (CONTINUED)

Omni-channel strategy required

Consumers today expect a unified web and mobile experience. Inventory, products, and prices should be consistent across all platforms. This is critical because shoppers often shop on one channel and buy on another. According to 2018 shopping data for Cyber Monday, mobile accounted for 51% of site visits but only 34% of revenue.² This data suggests consumers have a tendency to browse on mobile platforms before buying on desktops.

Recent research from PwC concludes that “customers expect the same brand experience, quality, price and service regardless of channel, and it is that connectivity and seamless experience across channels that is the real differentiator.”³ By delivering data via API, you can stop serving shoppers stale data and thereby offer a true omni-channel experience.

Know thy consumer

“The right content for the right consumer” is the name of the game in online retailing today. To stay competitive and relevant, your website and applications need to be able to target specific content to specific users based on location, language, and browsing preferences. You must also tailor content delivery to the consumer’s device, and address shoppers based on actions they’ve taken within a site or app.

The hits add up

Delivering personalized ecommerce applications can be expensive — very expensive. Since legacy CDNs cannot cache APIs, every page view or app interaction results in a hit to the origin infrastructure and generates both infrastructure CapEx and OpEx costs. Furthermore, you’re double charged if you use data hosting providers — legacy CDN vendors charge to facilitate the API call and hosting providers charge to serve it.

The mighty API

In today’s distributed application platforms, many key services are provided by APIs. APIs are at the core of both web and mobile ecommerce applications. They enable faster development, real-time data exchange, and more secure transactions. If your online store has a mobile application, chances are it was developed on top of APIs. A study by industry analyst firm Freeform Dynamics found that most organizations are using APIs to build mobile and Web apps (76 percent and 78 percent, respectively)⁴. Examples of common ecommerce API content include store details and locations, product inventory and pricing, shopping carts, payment options, and purchase transactions.

Mobile accounted for more than half of shopping visits but only 34% of revenue on Cyber Monday 2018

Adobe Press Release, November 2018

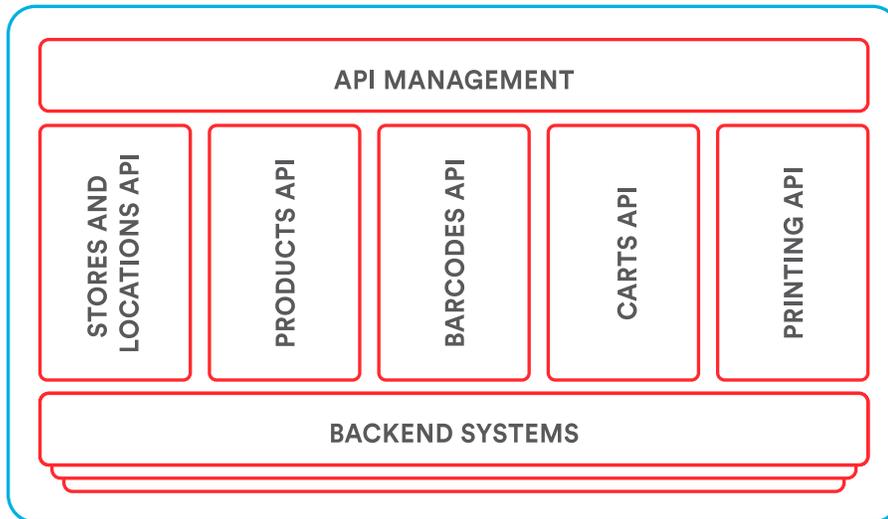
² “Adobe Analytics Data Shows Cyber Monday Broke Online Sales Record with \$7.9 Billion,” Adobe Press Release, November 26, 2018

³ “Customers Are Calling the Shots,” PwC, January 2017

⁴ “APIs abound, but challenges remain,” Thor Olavsrud, CIO, April 19, 2016

WHY YOUR APIS NEED AN EDGE CLOUD (CONTINUED)

Example of API stack



If only you could cache your APIs, you would improve response times, ensure data consistency across channels, and future-proof your applications by making them secure and scalable. But there's a problem...

The challenge of caching APIs

Static content like images and text can be reliably cached at the edge, speeding delivery and reducing infrastructure costs. Content like price and inventory that change frequently and unpredictably, however, are considered too "dynamic" to cache by legacy CDNs since they are not designed to instantly invalidate and remove cached content.

This dynamic content typically remains at the origin server — slowing delivery and racking up infrastructure costs. Many ecommerce companies employ a workaround strategy by caching content using a low "Time-To-Live" (TTL) setting in the CDN. TTL determines how long an object should remain in the CDN cache before it expires and therefore has to be refreshed from origin.

Unfortunately this workaround strategy has a dangerous side-effect: stale, outdated content may be served if it changes at the origin before the TTL expires. A 60-second TTL is a large enough window to cause an inconsistent or confusing user experience. For example, if inventory levels change frequently and unpredictably, the site could display a product as available for purchase while it is actually out of stock. On the flip side, a very low TTL leads to premature expiration of content, which results in the origin infrastructure bearing more load than is necessary.

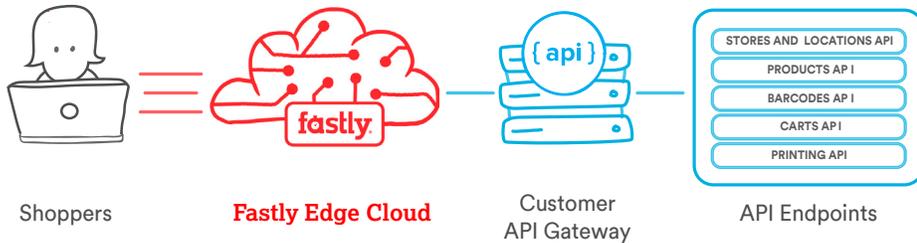
Introducing Fastly

Fastly is an edge cloud platform designed to meet the needs of today's ecommerce companies by providing an unprecedented level of performance and control. Fastly is trusted by leading ecommerce companies including Airbnb, Boots UK, Etsy, HotelTonight, KAYAK, Nordstrom Rack, Shopify, Stripe, and Ticketmaster.

WHY YOUR APIS NEED AN EDGE CLOUD (CONTINUED)

Our architecture allows you to serve content faster than legacy CDNs due in part to our ability to cache frequently and unpredictably changing content. We call this type of content “event-driven.” APIs can be classified as event-driven content since they contain frequently and unpredictably changing information such as catalog, pricing, and inventory data.

Example of traffic flow with Fastly



Fastly is based on a highly customized and distributed version of Varnish, an open source web accelerator designed for high-performance caching and content delivery. A primary benefit of a Varnish-based platform is the ability to implement more logic at the edge, allowing customers to implement functionality such as device detection, geo-fencing, request routing, and token authentication. If requests come via API call, Fastly can serve cached responses based on predefined custom logic for faster response times.

Legacy CDNs are unable to cache APIs, but Fastly makes it easy to deliver tailored customer experiences in real time, regardless of device type or network. In short, we give you an unprecedented level of performance and control on a single, secure platform.

Clear outdated content in a flash

Fastly’s global content invalidation has a mean purge time of 150 milliseconds. This near real-time purge capability is fundamental to caching event-driven content such as API responses. Legacy CDNs take hours or even days to fully clear content from cache, making them unable to cache APIs.

For customers who shop on one channel (say mobile) and buy on another (say desktop), Instant Purge of outdated API content ensures consistency across all channels.

Our near instant purge capability also enables action-triggered changes. For example, when an inventory level drops to zero, this event displays immediately to all applications. When a consumer writes a review and hits submit, it is immediately visible to other shoppers. If you run a 24-hour flash sale, the start and finish times are exactly as advertised. In each case, immediacy is essential.



There was a big unknown for us with the geo-fencing approach. We’d toyed with the idea of geo-fencing requests for some time. As with all new features on the platform we A/B tested it and the performance increase was astonishing. Historically the hotel list has been the most expensive to serve; now (once the cache is warm) we can return results in under 50ms. It’s awesome!

Harlow Ward, Rails Engineering Lead, HotelTonight

WHY YOUR APIS NEED AN EDGE CLOUD (CONTINUED)

Easily update content site-wide

To simplify the potentially complicated task of triggering multiple purges of related content, Fastly makes it possible to combine several API purge requests into a single action with Surrogate Keys. Administrators can tag APIs with keywords to create relational dependencies such as by product type, list price, and manufacturer.

For example, if the API call for a computer keyboard called “ABC” is tagged with categories “peripherals” and “keyboards”, then a purge request can clear the specific item, the product category, or all peripherals. Furthermore, if the marketing department initiates a 10% sale on peripherals, all cached peripherals pricing data can be immediately invalidated and refreshed with new pricing. So customers always see the most up-to-date information.

Deliver customized content based on location

With Fastly’s GeoIP module, you can pinpoint the location of an end-user down to their longitude, latitude, continent, country, city, postal code, telephone area code, and metro code. Using this information, you can create virtual boundaries (geo-fences) to direct users to localized content and logic.

HotelTonight, the popular mobile app for last minute hotel reservations, relies on Fastly’s GeoIP module to geo-fence URLs that group requests from users within the same proximity. This allows the company to return the same inventory and pricing information to users within a predefined geography from cache, thus reducing unique requests to the hotel inventory API by several orders of magnitude. HotelTonight can thereby reduce the load on origin while also increasing responsiveness for users.

With GeoIP, you can deliver localized content like coupons to a city or zip code, serve versions of your site in different languages by location, and show product availability by region — all the while keeping your API content cached at the edge.

Serve optimized content for every type of device

The ability to detect mobile devices and customize content in real-time allows retailers to create the best possible user experience. By enabling Device Detection on Fastly, you can serve content optimized for a specific platform via API.

For instance, if a user visits your site from an iPhone, you’ll want to serve images with smaller dimensions so as to fit the smaller screen area, but with higher resolution to take advantage of the iPhone’s retina display. When serving content to a user viewing your site with a different smartphone, you can automatically send them lower resolution images that utilize less bandwidth.

Gain real-time performance insights

To monitor API usage and health, Fastly makes it very easy for you to get real-time, secure access to your log data. You gain visibility into how your customers are engaging with your content, allowing you to identify trends and resolve any API delivery problems. Moreover, you can monitor the impact of new code deployments and, with Fastly’s versioned platform, roll back to previous stable code in case of issues.



We have confidence in Fastly that we don't have for other vendors we looked at. Because we move money around, we need the best available option for any number of things related to performance, security, and stability. We only considered what we thought was best in class, and Fastly was the obvious choice.

Marc Hedlund, VP of Engineering, Stripe

WHY YOUR APIS NEED AN EDGE CLOUD (CONTINUED)

Secure your customers' information

Protecting your customers' identities and transactions, as well as the integrity and reputation of your website, requires a comprehensive approach to security. Since Fastly is a certified Level 1 Service provider, compliant with the Payment Card Industry Data Security Standard (PCI DSS), it meets the requirements for secure API delivery. While other CDNs have different levels of PCI certification for different physical networks, Fastly provides a single platform that is fully PCI compliant.

Fastly also provides Transport Layer Security (TLS), the next generation encryption protocol. As part of the standard service, you can terminate secure TLS connections at our network edge, closer to users, offloading encrypted API traffic from your web server for better performance.

To protect against API targeted distributed denial of service attacks (DDoS), the Fastly network of globally distributed POPs has capacity large enough to absorb attack traffic.

Full transparency and first class support

Unlike legacy CDN vendors who provide black box systems that require their customers to always open a support ticket with slow response times, Fastly delivers unmatched transparency and support to speed deployments and address uptime issues. A Fastly service is fully self-provisioning and quick to set up. You can have a full deployment delivering production traffic to your APIs in a matter of minutes.

We allow customers to view configurations and make changes themselves, or reach out to Fastly's 24/7 dedicated customer support for assistance with some or all changes. Through Fastly's developer-friendly control panel, you can easily make changes to your configurations – such as modifying the default TTL or enabling Gzip of assets for improved performance.

Fastly customers love our support. Our knowledgeable support engineers are available across a variety of channels to provide assistance in near real-time. It's no wonder our customer support satisfaction rating stands over 95% as of December 31, 2018.

The Fastly way

Optimized web and mobile content delivery benefits both you and your customers. Your customers get a better shopping experience. Your faster sites yield higher conversion rates, and your tailored content increases average cart value. Caching APIs at the edge of our network enables you to attain these goals across all of your content — not just static assets. In the highly competitive world of ecommerce, delivering API content at speed can be a major source of competitive advantage. Isn't it about time you tried Fastly?

To learn more about how we can help support your mobile ecommerce and API strategy, contact Fastly at sales@fastly.com.



A Walgreens subsidiary

Fastly's GeoIP edge module allows us to determine where customers are coming from and make a logical guess based on their location as to which Boots site they'd like to access.

Peter Dawson, Solution Architect,
Boots UK