Menu

Category: Network Services

These articles provide information about Fastly products that focus on performance (speed), availability, and media and that accelerate content delivery with control from an edge cloud platform.

Cache Reservation

Last updated: 2025-05-06

/products/cache-reservation

Cache Reservation provides access to the caching layer at Fastly's edge where you can reserve cache space specifically for your content in Fastly shielding locations. By prioritizing your content's cache storage, Cache Reservation allows that content to stay in cache longer by minimizing content eviction in these multi-tenant environments. Cache Reservation helps optimize your origin's offload from any CDN, including Fastly, reducing your cloud egress costs.

To learn more about Cache Reservation, contact your account manager or email sales@fastly.com for additional details.

Prerequisites

To purchase Cache Reservation you must have a paid account for full-site delivery or streaming delivery and you must enable shielding. In addition, all cache reservations are subject to pre-qualification, specific to the size of your content and the shielding locations requested. Specifically:

- the ID of the specific service you want to have associated with your cache reservation.
- the shielding location for the traffic managed by the specific service.
- the approximate size of the objects included in this reservation.

- an estimate of how much space (in GB) the actively consumed content for each service will require at the chosen shielding location. We can help you with this estimate. Fastly will use this information in the qualification process to determine if it is possible for your company to use Cache Reservation. To see if your company meets the qualification criteria, contact sales@fastly.com.

Limitations and considerations

Keep in mind the following limitations and considerations:

- Cache reservation is not compatible with Fastly's Compute platform.
- Reservations are restricted to shield POPs only.
- Each reservation covers a single shielding location.
- Reservations apply to your cache as a whole, not to individual objects, and you cannot see what is in the cache at any given time.
- Reservations minimize cache storage eviction via prioritization up to the specified reservation size, but do not ultimately prevent eviction.

Billing

Cache Reservation charges are billed based on reservation size (in GB) in specific shielding locations.

---

* * *

Capacity Reservation

Last updated: 2023-04-05

/products/capacity-reservation

Capacity Reservation allows you to reserve Fastly traffic capacity for events based on data bandwidth (in gigabits per second), duration, and type of delivery (Media Shield for VOD, Media Shield for Live, Fastly Streaming Delivery, or Fastly Full Site Delivery). You are

required to purchase capacity reservations any time you're expecting a utilization spike from your planned events. When purchasing capacity reservations for your events, you are also required to purchase Fastly's Live Event Monitoring service for the duration of the event.

Billing

> NOTE
>
> Billing limits for this product may be different if you've purchased a packaged offering or are using a product or feature trial.

Fees for Capacity Reservation are based on the duration, in hours, of the event and the reserved bandwidth, in gigabits per second (Gbps), for Fastly Full Site Delivery and Streaming Delivery traffic (Capacity Reservation - Edge) or Media Shield for VOD and Media Shield for Live traffic (Capacity Reservation - Media Shield). Fees do not include delivery fees for Fastly Full Site Delivery, Fastly Streaming Delivery, Media Shield for VOD, or Media Shield for Live.

> Some Fastly products can be purchased directly in the web interface on the Products page. For more details about a product, including pricing information or for help purchasing it, contact your account manager or email sales@fastly.com.

---

* * *

Cloud Optimizer

Last updated: 2020-12-18

/products/cloud-optimizer

> IMPORTANT
>
> This information is part of a limited availability release. For additional details, read our product and feature lifecycle descriptions.

Fastly's Cloud Optimizer product allows customers using one or more non-Fastly content delivery networks (CDNs) to take advantage of Fastly's Full-Site Delivery features without migrating edge delivery traffic to Fastly. Cloud Optimizer works with your existing content delivery infrastructure by designating Fastly as the origin for all of your end-user-serving CDNs. Using Cloud Optimizer provides you with real-time visibility of origin traffic, granular load balancing for your origin infrastructure, and request collapsing to decrease traffic to origin.

To learn more about Fastly's Cloud Optimizer, contact your account manager or email sales@fastly.com for more details.

> NOTE
>
> Cloud Optimizer is not available for video streaming activities. Check out Media Shield for Live and Media Shield for VOD instead.

Some Fastly products can be purchased directly in the web interface on the Products page. For more details about a product, including pricing information or for help purchasing it, contact your account manager or email sales@fastly.com.

---

* * *

Dedicated IP addresses

Last updated: 2023-09-28

/products/dedicated-ip-addresses

Fastly's Dedicated Internet Protocol (IP) addresses provide you with a pool of IPv4 and IPv6 addresses, maintained and managed by us, across Fastly's global Edge Cloud. They can be used to support TLS certificate management for non-SNI clients, to support custom cipher suites or IP-to-service pinning, or to help manage zero-rated billing endpoints or security allowlisting.

Purchase of Fastly's Platform TLS product requires you to also have purchased Dedicated IP addresses.

TLS non-SNI client support

Fastly-managed certificates require clients to support TLS v1.2 and Server Name Indication (SNI) by default. When there is not an SNI match, a fallback certificate is used. Dedicated IP addresses can be used to host fallback certificates for non-SNI client support. Using the Fastly API, you can place your self-managed or Fastly-managed certificates at a dedicated set of IP addresses in the event there is no SNI match.

Additionally, for self-managed certificates only, you can also indicate a default ECDSA and RSA certificate. When no SNI match is found, Fastly will first check if the client supports ECDSA. If it does, we will send the fallback ECDSA certificate. If there is no SNI match and the client does not support ECDSA, we send the RSA fallback certificate.

Use your DNS records to associate the set of dedicated IP addresses to use to direct incoming traffic to the Fastly edge network. There are three possible network routing options (sometimes referred to as network maps or domain maps) that allow you to choose which sub-regions of the Fastly network to use.

Read Fastly's TLS offerings for a more detailed description of the supported TLS options at Fastly.

Custom cipher suites

Fastly supports a number of standard cipher suites. Should you require more personalized control, Fastly supports the creation of custom cipher suites by providing you with dedicated IP addresses that support these custom sets.

IP-to-service pinning

IP-to-service pinning uses dedicated IP addresses to map customer services to specific endpoint IP addresses and direct an end user's request to a specific service based on the requested endpoint IP address.

Zero-rated IP addresses

Zero-rated IP addresses (ZRIPs) allow you to use dedicated IP addresses within Fastly's global Edge Cloud to identify traffic for special treatment. For example, if you need to waive billing charges going to or from specific web pages, ZRIPs can help you to identify traffic for zero billing.

Security allowlisting

Security allowlisting uses dedicated IP addresses to control the set of Fastly global IP addresses seen by third parties. You can incorporate dedicated IPs into access control lists (ACLs) to tighten security between a customer and a third party.

> Some Fastly products can be purchased directly in the web interface on the Products page. For more details about a product, including pricing information or for help purchasing it, contact your account manager or email sales@fastly.com.

---

* * *

Domain Research API

Last updated: 2025-11-04

/products/domain-research-api

Fastly's Domain Research API allows you to programmatically retrieve algorithmic domain search results and check domain availability. The Domain Research API might be useful if you want to:

- check if domain names are available for registration or for sale in aftermarkets.
- add a domain discovery experience to your product.
- monitor domain name state changes over time.

How it works

The Domain Research API is part of the Fastly API, which requires a Fastly API token. You can use the Fastly API to enable the Domain Research API for your account.

The Domain Research API has two endpoints. The Suggest endpoint allows you to retrieve domain search results, and the Status endpoint allows you to check domain availability.

The Status endpoint has two variants. The Precise variant allows you to check the registry-level availability of a domain (e.g., for new registrations), and the Estimated variant allows you to check DNS and aftermarket-level availability of a domain (e.g., not registry-level).

Limitations and considerations

The Domain Research API has the following limitations and considerations:

- The most common use case of the Status endpoint is to determine if a domain is available for registration from a registrar. The `inactive` status means a domain is available for registration.
- By design, the Status endpoint accepts only a single domain per API request. If your use case calls for checking multiple domains, you'll need to check them one at a time.
- Our platform limits Domain Research API requests to a total runtime of 30 seconds. If an upstream provider fails to respond within that timeframe, our API will return a best-effort value, typically `unknown` or `undelegated`.
- The Suggest and Status endpoints are intentionally decoupled. Domain availability information is not present in Suggest endpoint responses, and Status endpoint responses do not include search result suggestions.

Billing

The Domain Research API is disabled by default. Users assigned the superuser role can enable it on the Products page.

Billing for the Domain Research API is based on Domain Research API requests. There are three types of billable Domain Research API requests:

- **Suggest API requests:** The number of HTTP requests sent to the Suggest endpoint.
- **Status-Precise API requests:** The number of HTTP requests sent to the Status endpoint.
- **Status-Estimated API requests:** The number of HTTP requests sent to the Status endpoint, with the `&scope=estimate` parameter explicitly set. These responses include a subset of the data returned by the Status-Precise API consisting of (i) DNS-level domain delegation (or undelegated) data and (ii) Aftermarket listing

metadata (if present for a given domain). These responses do not include domain registry-level availability data.

---

* * *

Fastly's Full-Site Delivery

Last updated: 2024-10-22

/products/fastlys-full-site-delivery

Fastly's Full-Site Delivery allows you to speed up websites and mobile applications by pushing content closer to users, providing improved and secure experiences across the world. Full-Site Delivery includes the following features.

Content serving, caching, and control

Full-Site Delivery uses Fastly's global content delivery capabilities to cache and accelerate the delivery of your HTTP-based file content such as video, images, CSS, JavaScript files, as well as HTML and API responses. Specifically:

- **HTTP header controls.** Full-Site Delivery obeys standard HTTP caching headers and support forwarding, adding, removing, and modifying the HTTP headers we receive from your origin servers and send to end users, allowing you to send one set of instructions to your Fastly services and another set of instructions to downstream caches, proxies or browsers.
- **Time to Live controls.** Content expiration is controlled via Time to Live (TTL) settings you configure that work as timers on your cached content. You have the option of configuring a global default TTL to control cached content which, when set, will cache objects in a consistent manner even if you have multiple origins or server applications with inconsistent TTL settings.
- **Request collapsing.** When your content expires, the fetch and refresh process from your origin may take one second or more. During that time, your Full-Site Delivery may receive dozens or hundreds of end-user requests for that content. Fastly's request collapsing feature groups those requests and fulfills them together when it receives the refreshed content from your origin. Request collapsing decreases load

receives the refreshed content from your origin. Request collapsing decreases load on your origin servers by keeping your Fastly services from sending duplicate requests for the same expired content to them. Request collapsing is enabled by default.

- **Grace mode (Serving stale content).** If your origin servers become unavailable for any reason, grace mode can instruct your Fastly services to continue to serve stale or expired (but likely still valid) content to end users for a set amount of time. This allows you some extra time to return your unavailable servers to normal operations while still serving content instead of error messages to end users. Grace mode is not configured by default. To enable it, you must specifically configure your services to serve stale content.
- **Compression.** To help you speed up information transmission, we allow you to compress static content during transmission thereby making it available to your customers more quickly. To enable static or dynamic content compression, you must either enable automatic compression or set up an advanced compression policy.
- **Purging.** For dynamic or event-based content that doesn't lend itself to predetermined TTL-based content expiration, you can proactively remove or invalidate your content within milliseconds with Fastly's purging features. We limit purging to an average of 100K purges per hour per customer account, inclusive of all services within that account and according to any packaged offering you've purchased.

Edge logic and advanced content delivery control

Fastly's content delivery capabilities are based on a heavily extended version of the Varnish caching software. Varnish software gives you direct access to content delivery, control and edge logic capabilities, via the expressive HTTP inspection and modification scripting language, Varnish Configuration Language (VCL).

Streaming content delivery

Fastly's Streaming Delivery allows you to stream live and video-on-demand streaming content by leveraging Fastly's native support of common streaming formats. Fastly streaming format support includes HTTP Live Streaming (HLS), HTTP Dynamic Streaming (HDS), Dynamic Adaptive Streaming over HTTP (MPEG-DASH) and HTTP Smooth Streaming.

Precision Path

Precision Path traffic routing proactively identifies network congestion and poorly performing paths and automatically switches your traffic over to better performing alternatives. This improves service availability and resilience. Provisioned at strategic locations across our global fleet, this feature is available to all Fastly customers as part of our platform.

Origin shielding

You can designate a Fastly point of presence (POP) to serve as a shield for your origin servers, thus enabling increased cache hit rates for your Fastly services and potentially protecting your origin servers from unexpected spikes in requests for content. You can optimize this shielding geographically by configuring different shield POPs for different origin server locations. Origin shielding is not enabled by default. To use it, you must specifically enable it.

Load balancing

Services configured with multiple origin servers will automatically distribute requests to those servers evenly. You can modify this default load balancing behavior with a variety of conditions and load balancing rules.

Health checks

The health of your origin servers can be monitored with configurable health checks to help ensure only responsive origin servers are being sent requests.

Fastly control panel

All Fastly accounts have access to Fastly's control panel, allowing it to be managed by multiple users within your organization. You can control each user's role, as well as control the scope of their service access and their specific permission levels. Fastly services can be created, monitored, and managed through the Fastly control panel via any standard, modern web browser.

Application programming interface (API)

Fastly provides an application programming interface (API), accessible via HTTPS,

through which Fastly services can be created and configured, and customers can access account information and analytics.

Real-time log streaming

To help you tune the performance of your Fastly services, we support real-time log streaming to a variety of locations, including third-party services, for storage and analysis. You can find our supported logging endpoints in our list of streaming log guides. We limit real-time log usage to a monthly average of two log statements per request, per service. If you require a higher volume of logs, Fastly offers High Volume Logging.

Transport Layer Security

Fastly supports a variety of Transport Layer Security (TLS) services that allow websites and applications to serve traffic over HTTP Secure (HTTPS), providing added privacy and data security for your services and end users. All Fastly services have access to our free shared domain option, plus a variety of additional paid TLS services to meet your TLS business and technical needs.

Always-on DDoS mitigation

Fastly's globally distributed network was built to absorb Distributed Denial of Service (DDoS) attacks. As part of Fastly's standard, Full Site Delivery, all customers receive access to a combination of features inherent in Fastly Edge Cloud network capabilities that help protect the availability of your content from DDoS threats.

- **Access to origin shielding.** Fastly allows you to designate a specific point of presence (POP) to host cached content from your origin servers. This POP acts as a *shield* that protects those servers from every cache miss or pass through the Fastly network, reducing the load that directly reaches them.
- **Automatic resistance to availability attacks.** Before they're even processed by our caching infrastructure, we filter out Layer 3 and 4 attacks (e.g., Ping floods, ICMP floods, UDP abuse) as well as distributed reflection and amplification (DRDoS) attacks that rely on anonymity to abuse internet protocols (e.g., DNS and NTP).
- **Access to Fastly cache IP space.** Fastly provides an API endpoint to any customer who would like to know which IP addresses our caches will use to send traffic from our CDN to your origin servers. We make this data available so you can update firewalls at your origin to ensure only our cache traffic can access your resources.

- **Custom DDoS filter creation abilities.** Using custom VCL, we allow you to craft your own DDoS protection rules to protect your network from complex Layer 7 attacks. Once you identify signs of a potential DDoS attack, you can mix and match Fastly VCL with custom VCL to construct filter configurations based on a variety of client and request criteria (e.g., headers, cookies, request path, client IP, geographic location) that block malicious requests before they hit your origin servers.

In addition to these included mitigation capabilities, Fastly offers Fastly DDoS Protection. For more information about this or any of our advanced services, including their subscription costs, contact sales@fastly.com.

Pricing and billing

> NOTE
>
> Billing limits for this product may be different if you've purchased a packaged offering or are using a product or feature trial.

Full-Site Delivery charges are based on the volume of content delivered to your end users and the location of the POPs from which that content was served. Fastly billing is done in arrears based on actual usage with month-to-date usage being available via both our control panel and APIs.

> NOTE
>
> Fastly maintains partnerships with Google and Microsoft that may provide discounts on outbound data transfer traffic to customers who qualify and configure their Fastly services correctly. See our integrations guides for additional details.

---

* * *

Fastly's On-the-Fly Packager service

Last updated: 2024-01-16

/products/fastlys-onthefly-packager-service

Fastly offers an "on-the-fly," dynamic, video-on-demand content packager service. Rather than requiring you to pre-package all protocols of a viewer-requested video, Fastly allows you to dynamically package video content in different HTTP streaming formats in real time, using source files. That video content then becomes immediately available to viewers.

> IMPORTANT
>
> Fastly's On-the-Fly Packager (OTFP) for On Demand Streaming service is an add-on service that requires the purchase of or use of additional professional service hours for first-time implementations. Our Professional Services team will assist with configuration and testing of this product and its features. To enable OTFP and begin this process, contact your account manager or email sales@fastly.com for more details.

Supported on-the-fly packager features

Fastly's OTFP service supports the following specific features.

Supported HTTP streaming formats and codecs

- **HDS, HLS, and MPEG-DASH packaging.** Fastly provides support for version 1 of the Adobe HTTP Dynamic Streaming (HDS) specification and support for the ISO/IEC 23009-1:2014 specification defining Dynamic Adaptive Streaming over HTTP (MPEG-DASH). We support all features included in up to version 3 (draft 6) of the HTTP Live Streaming (HLS) specification and popular features from later versions such as subtitle, trick play and media segmentation in fragmented MPEG-4 (fMP4) format (per ISO/IEC 14996-12:2015 specification).
- **Standard codecs.** Fastly supports Advanced Video Coding (H.264/AVC/MPEG-4 Part 10) and High Efficiency Video Coding (H.265/HEVC) video codecs. Fastly also supports Advanced Audio Coding (AAC, AAC-LC, HE-AAC), Dolby Digital (AC-3) and MPEG-1 Audio Layer III (MP3) audio codecs.
- **Source video container format.** Fastly supports the Progressive MP4 specification (specifically the .mp4, unencrypted .mov, and audio-only .m4a extensions) as source container format for packaging into all supported HTTP streaming formats.

Accessibility and user experience

- **HLS multi-language subtitles and closed captions.** Fastly provides support for both in-band (EIA-608 and CEA-708) and out-of-band (Web Video Text Tracks or WebVTT) subtitle and closed caption delivery.
- **HLS trick play.** Fastly supports trick play (also called trick mode), a feature that displays video scenes during fast-forwarding and rewinding. The HLS Authoring Specification requires this feature for distributing video on the Apple TV.

Content protection

- **Media encryption.** Fastly can encrypt videos packaged into HLS (supports both Envelope/AES-128 and SAMPLE-AES methods) and MPEG-DASH (ISO/IEC 23001-7, a common encryption in ISO base media file format file) streaming formats by generating a unique content encryption key for each video, enabling secure video delivery to viewers.
- **Multi-DRM.** Fastly can support multiple Digital Rights Management (DRM) technologies including Apple FairPlay for HLS and Microsoft PlayReady, Google Widevine and Marlin DRM for MPEG-DASH streaming formats. OTFP is integrated with Multi-DRM service providers that are responsible for content rights management and DRM license delivery.

Dynamic Ad Insertion (DAI) readiness

- **HLS timed metadata injection.** Fastly supports HLS time-based metadata, which allows you embed custom metadata or ad markers about a stream into video segments at specified time instances in ID3v2 format.
- **Content preconditioning.** Fastly can segment video at the intended break points, such as for ad markers via HLS and MPEG-DASH protocols. Fastly can also add any third-party service-specific cues or metadata into video manifests at those break points to implement server or client-side ad stitching.

Clip creation

- **Clip creation (also known as "timeline trimming").** Fastly supports clip creation features for all supported packaging formats, allowing you to deliver sections of video without segmenting a longer, archived video. Time query parameters ("start" and "end") allow you to break up videos into discrete sections so users don't have to find the relevant section using the timeline.

Standard content delivery network features

Fastly also provides the following features as part of standard content delivery network services:

- Token-based validation for decreasing response time by placing validation at the edge
- Geolocation and device detection for content targeting
- Dictionaries for real-time business rules and decision making at the edge
- Remote log streaming for data aggregation and viewer diagnostics
- Transport Layer Security (TLS) for secure communications delivery

How the on-the-fly packager service works

Fastly's OTFP service gets configured between our caching network and your origin storage (e.g., Amazon S3, Google Cloud Storage, or Rackspace Cloud Files).

When users request manifests or video segments, those requests initially come to Fastly caches instead of going to your origin storage. Fastly's edge caches deliver those objects if they are available and valid. If the objects don't already exist in the edge caches, the requests will be passed on to a designated shield cache to be delivered instead as long as the objects are available and valid. If neither the edge caches nor the shield cache can deliver the objects, the requests for those objects will go directly to and be fulfilled by the OTFP service which acts as an origin for Fastly's cache nodes.

The OTFP service will make the necessary request to your origin storage to fulfill the original request from the user. The OTFP service also maintains a small, local, in-memory cache for video metadata indexes. These indexes are created using mp4 moov atom (or movie atom) that provide information about the video file such as its timescale, duration, audio and video codec information, and video resolution (among other characteristics).

For adaptive bitrate playback, the OTFP service will cache indexes of each quality level

requested. If a user requests a manifest, OTFP will look for the corresponding indexes and, if it is available and valid, OTFP will generate the manifest and deliver it to the user. Otherwise, OTFP will fetch the moov atom from origin storage to generate the corresponding index. If a user requests video segments, OTFP will look for the corresponding audio and video sample entries in the cached index, download those samples from origin storage, and package them in the format requested.

---

* * *

Fastly's Streaming Delivery

Last updated: 2025-12-19

/products/fastlys-streaming-delivery

Fastly's Streaming Delivery allows you to scale the delivery of your streaming content independently of any other HTTP content delivery supported by Fastly's Full Site Delivery. You can also configure and control live and video on demand (VOD) caching using full site delivery, but there are advantages to using streaming delivery, such as more favorable streaming-specific pricing and a lower traffic load on your full site delivery service. Additionally, all of the features available to full site delivery services are available to streaming delivery services.

> IMPORTANT
>
> Streaming media delivery refers to the delivery of streaming media content using supported streaming protocols for real-time consumption as the content is received, rather than after it has been downloaded. Streaming media content consists of audio and video segments, and associated content intended to enable or to be consumed with them, such as manifests and files used for captions or subtitles.

If you have your own video packaging infrastructure, Fastly can act as a globally distributed HTTP streaming network to improve quality of service and increase viewer capacity for both your live and VOD content. When a manifest or video segment is requested by an end user's player, your Fastly Streaming Delivery will pull the requested content from your origin media servers and subsequent requests for that stream will be

served from Fastly's points of presence (POPs) instead of your origin servers.

Request collapsing

If many users request the same content at the same time and that content is not cached in Fastly's POPs, your origin will have to serve that content. It doesn't, however, need to know about every individual user request made to Fastly's POPs and it would be inefficient to send the same content out many times. So, Fastly will only request the in-demand content from your origin once, essentially *collapsing* all of the user requests into a single request to your origin. Then we'll respond to each user individually.

Read more about request collapsing.

Streaming miss

When Fastly needs to fetch content from your origin to serve a user request, we minimize the time until that user receives the first response (also called first-byte latency), by sending pieces of your origin's response to the user as soon as Fastly receives them, instead of first buffering the response from your origin, caching the data, and then streaming the data back to the user.

Read more about Streaming Miss.

Origin shielding

You can designate a Fastly POP to serve as a shield for your origin servers, intercepting user requests on behalf of your origins to protect them from spikes in request traffic and also potentially increasing your overall cache hit rates. You can optimize this shielding geographically by configuring different shield POPs for different origin servers.

Origin shielding is not enabled by default. To use it, you must specifically enable it.

Read more about origin shielding.

Real-time log streaming

To help you tune the performance of your Fastly services, we support real-time log streaming to a variety of locations, including third-party services, for storage and analysis. You can find our supported logging endpoints in our list of streaming log guides. We limit real-time log usage to a monthly average of two log statements per request, per service. If you require a higher volume of logs, Fastly offers High Volume Logging

service. If you require a higher volume of logs, Fastly offers High Volume Logging.

Supported streaming protocols

Fastly's Streaming Delivery supports the following HTTP-based media streaming protocols:

- Dynamic Adaptive Streaming over HTTP (MPEG-DASH)
- High Efficiency Streaming Protocol (HESP)
- HTTP Dynamic Streaming (HDS)
- HTTP Live Streaming (HLS)
- HTTP Smooth Streaming (HSS)
- Low-Latency HTTP Live Streaming (LL-HLS)

Limitations and billing

Fastly's Streaming Delivery is a subset of Fastly's Full Site Delivery for the delivery of streaming media content. Fastly's Streaming Delivery must be configured in a separate Fastly account, for exclusive use with streaming media content, in order to use separate billing plans and invoices as part of calculating your bill.

> Some Fastly products can be purchased directly in the web interface on the Products page. For more details about a product, including pricing information or for help purchasing it, contact your account manager or email sales@fastly.com.

> NOTE
>
> Fastly maintains partnerships with Google and Microsoft that may provide discounts on outbound data transfer traffic to customers who qualify and configure their Fastly services correctly. See our integrations guides for additional details.

* * *

HIPAA-Compliant Caching and

Delivery

Last updated: 2018-08-01

/products/hipaa-compliant-caching-and-delivery

You can configure the Fastly CDN service to cache and transmit protected health information (PHI) in keeping with Health Information Portability and Accountability Act (HIPAA) security requirements. Use the following features to ensure secure handling of cache data that contains PHI:

- Configure frontend and backend TLS to encrypt transmitted data from your origin to your end users.
- Add the `beresp.hipaa` variable to objects containing PHI to keep that data out of non-volatile disk storage at the edge.

Contact sales@fastly.com for more information on how to enable the `beresp.hipaa` feature for your account. For accounts that have this feature enabled, Fastly will enter into a HIPAA business associate agreement (BAA) as an addendum to our terms of service.

> IMPORTANT
>
> If you have purchased Fastly's PCI-compliant caching or HIPAA-compliant caching products Fastly will enforce a minimum version of TLS 1.2 or higher for all connections to meet the compliance requirements mandated by the PCI Security Standards Council.

> NOTE
>
> Fastly's security and technology compliance program includes safeguards for the entire Fastly CDN service, independent of using the `beresp.hipaa` variable. The Fastly security program and technology compliance content provide more information about these safeguards.

---

* * *

Image Optimizer

Last updated: 2025-11-18

/products/image-optimizer

The Fastly Image Optimizer (Fastly IO) is a real-time image transformation and optimization service that caches and serves pixel-optimized, bandwidth-efficient images requested from your origin server. Fastly IO specifically supports a variety of input and output image formats.

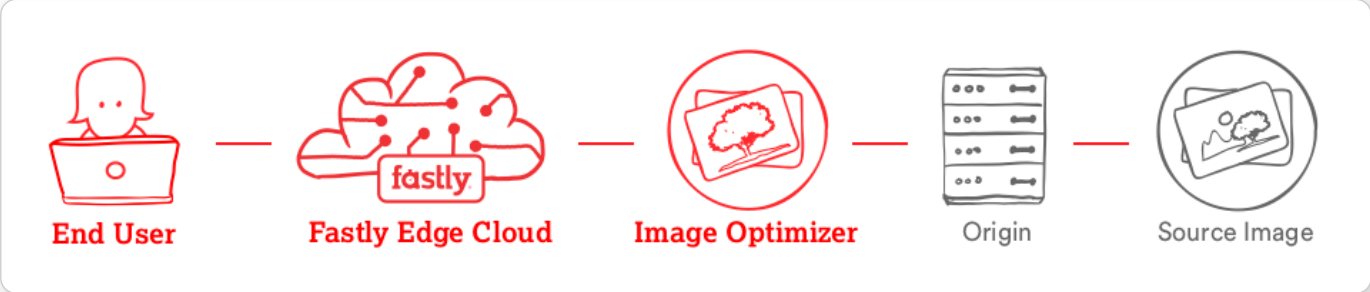

Image transformation and optimization

When an image is requested from your origin server, Fastly IO can perform transformation tasks before serving and caching the optimized version. Image transformations can be applied programmatically and through dynamic URLs in real-time. You can make images responsive so they automatically adjust to fit the size of the screen viewing the content. As a result, image pre-processing can be offloaded to the edge. Multiple copies of the images, each appropriately sized for different devices, are served from cache instead, which allows you to reduce the number of requests to your origin.

Debugging and troubleshooting

To aid in debugging when serving images, special HTTP headers will be present in a response when an image is requested. The specific header included depends on the response's result. For successful transformations and optimizations, the HTTP header returned provides general information that allows you to compare image dimensions, file sizes, and formats. Additional HTTP headers are included for source image issues that aren't fatal enough to cause an error but could still be problematic, as well as

transformations and optimizations that fail outright.

Billing

> NOTE
>
> Billing limits for this product may be different if you've purchased a packaged offering or are using a product or feature trial.

Billing for Fastly IO is based on the number of image requests that Image Optimizer processes and delivers for you. A request becomes a billable IO request when your service configuration invokes Image Optimizer.

What triggers a billable IO request

Billing triggers differ depending on whether you are using a VCL service or a Compute service:

- **VCL services:** Any request that sets the `req.http.X-Fastly-Imageopto-Api` header is counted as a billable image request. Services using shielding may execute their edge logic twice—once at the edge POP and once at the shield POP. Both executions count as billable image requests.
- **Compute services:** Image Optimizer can be accessed through Compute SDKs as part of Fastly's Beta program. Any request sent to Image Optimizer, including requests made through the SDKs such as `Request::with_image_optimizer`, counts as a billable image request.

How image requests are counted

Once Image Optimizer is invoked, we count image requests as follows:

- Image Optimizer requests are billed whether there is a cache HIT or MISS.
- When using animated GIF-to-video functionality, each frame delivered as video counts as a separate image request. GIF-to-GIF transformations count as one request.
- Non-image content cannot be optimized, but will still be counted as an image request if it is sent through Image Optimizer. Only send image content through IO.

Purchasing

On most accounts, anyone assigned the role of superuser can purchase this product from the Fastly control panel. If you have not been assigned that role, you can use the control panel to request that a superuser purchase it for you.

Some features, such as AVIF and JPEGXL encoding format, may require an account upgrade. Existing customers can contact sales@fastly.com for details about pricing, features, and upgrade options.

> This article describes a product that may use third-party cloud infrastructure to process or store content or requests for content. For more information, check out our cloud infrastructure security and compliance program.

* * *

Media Shield for Live

Last updated: 2020-12-18

/products/media-shield-for-live

Fastly Media Shield for Live offers customers the ability to decrease origin traffic by reducing multiple CDN requests of live video events or live linear channels into a single request back to your origin. Media Shield for Live works with your existing architecture by making Fastly the origin to all of your end-user-serving CDNs. This also allows you to take advantage of Fastly's Observability features in a multi-CDN environment.

> Some Fastly products can be purchased directly in the web interface on the Products page. For more details about a product, including pricing information or for help purchasing it, contact your account manager or email sales@fastly.com.

* * *

Media Shield for VOD

	Last updated: 2020-12-18
	/products/media-shield-for-vod

Fastly Media Shield for VOD offers video-on-demand customers the ability to decrease origin traffic by reducing multiple CDN requests into a single request back to your origin. Media Shield for VOD works with your existing architecture by making Fastly the origin to all of your end-user-serving CDNs. Fastly Media Shield for VOD is compatible with Fastly's On-the-Fly-Packaging (OTFP) service.

Media Shield for VOD allows you to take advantage of Fastly's Observability features in a multi-CDN environment.

> Some Fastly products can be purchased directly in the web interface on the Products page. For more details about a product, including pricing information or for help purchasing it, contact your account manager or email sales@fastly.com.

---

* * *

Object Storage

	Last updated: 2024-12-19
	/products/object-storage

Fastly Object Storage is an Amazon S3-compatible large object storage solution that works seamlessly with both Deliver and Compute services. Using Fastly Object Storage, you can store larger file sizes with Fastly, improving latency, increasing cache hit ratios, and reducing egress charges. Objects stored in Fastly Object Storage are accessible via an S3 compatible interface.

Fastly Object Storage might be useful if you want to:

- replace an existing storage solution with one that's part of Fastly's network.

- store data in a neutral location to be used across different vendors.
- access data at the edge to reduce overall costs.

Limitations and considerations

Use of Fastly Object Storage is subject to the following limitations:

- Data is stored in containers called buckets, which are limited to 100 per region. There is no limit to the amount of data storage per bucket.
- Object keys are limited to 1024 bytes.
- Objects are limited to 5 TB per object. Object metadata is limited to 1,000 bytes per object.
- No more than 5 GB of data may be included in a single upload, but multipart upload is supported.
- The minimum size of any part of a multipart upload, except the last one, is 5 MB. The maximum number of parts in a multipart upload is 10,000.
- Uploads must complete within 120 seconds or you'll receive a `408 Request Timeout` error. One way to decrease the likelihood of timeouts is to upload objects in smaller parts, making sure to meet the requirements for multipart upload above.
- Metadata set on objects using `x-amz-meta-` headers is limited to 1000 bytes, including the length of the header name, and must not begin with `x-amz-meta-fst` or `x-amz-meta-fastly`.
- Data accessed from public, Amazon S3-compatible endpoints are subject to a maximum speed of ~150Mbps and limited to a total of 100 requests per second per bucket. Data accessed from within Fastly's network is not subject to these limits. If your needs exceed these limits, reach out to sales@fastly.com.

Additional limitations set by the S3-compatible API can be found in the documentation.

Billing

Fastly Object Storage is an add-on and is priced in addition to Fastly services. On most accounts, anyone assigned the role of superuser can purchase this product from the Fastly control panel. If you have not been assigned that role, you can use the control panel to request that a superuser purchase it for you.

panel to request that a superuser purchase it for you.

Billing for Fastly Object Storage is based on a combination of total storage charges and data processing operations for the month. Storage is calculated using GB-month, rounded to the nearest hour, and processing operations are categorized into two groups, each with differing prices:

- **Class A Operations** are write operations and include `CreateMultipartUpload`, `CompleteMultipartUpload`, `PutObject`, and `UploadPart`.
- **Class B Operations** are read operations and include `GetObject` and `HeadObject`.

A full list of supported operations are available in our documentation.

You are responsible for removing your data and you will be charged for storage as long as your data is present in Fastly Object Storage.

For more details about this product, contact your account manager or email sales@fastly.com.

---

* * *

Oblivious HTTP Relay

Last updated: 2023-05-31

/products/oblivious-http-relay

The Fastly Oblivious HTTP Relay (OHTTP Relay) implements the relay portion of the Oblivious HTTP specification, which allows you to create an OHTTP-compliant service using Fastly. It can be used to build double-blind privacy-enabled Fastly services that transmit requests and responses without direct knowledge of personally identifiable information linked to customers.

How it works

Oblivious HTTP is a protocol for forwarding encrypted messages via HTTP. Specifically, OHTTP facilitates the transmission of an encrypted, encapsulated message to an HTTP endpoint from a client to a gateway through a trusted relay service, without delivering

identifying information about the end user who made the request or other information that is unnecessary for request processing. Fastly's Oblivious HTTP Relay acts as that relay service.

Using Oblivious HTTP, encrypted messages are created by a client and forwarded via HTTPS to a trusted relay, in this case, Fastly's OHTTP Relay. That relay then forwards it via HTTPS to a gateway. The gateway then removes any request encryption and generates an encrypted response to the original request, forwarding it to a target without ever exposing the client originally making the request.

Fastly's OHTTP Relay product serves as the relay portion of the Oblivious HTTP transmission process. Specifically, the Fastly OHTTP Relay does the following:

- **Routes requests and responses.** The OHTTP relay routes encrypted, encapsulated messages and corresponding responses between clients and configured backends (OHTTP Gateways).
- **Performs simple request and response validation.** The OHTTP relay performs simple request and response validation, which you can specify. For example, OHTTP can confirm the message's content type, that the request was received via HTTPS, and that the request was received with a known host and path that maps to a known backend endpoint.
- **Removes non-essential request information.** The OHTTP relay strips all request headers except those that are required for the correct operation of the Fastly service or that must be passed to the OHTTP Gateway. At your request, Fastly can configure specific headers as long as they don't contain personally identifiable information.

Limitations and considerations

To maintain the privacy hygiene of messages and their corresponding responses, OHTTP Relay will not permit the following:

- **You will not be able to use the control panel or API to control your OHTTP-enabled service configuration.** After the OHTTP-enabled service is created, you must contact Fastly to make modifications to the service configuration.
- **You cannot decrypt encapsulated messages.** No visibility or introspection into the nature of the end user request is possible within Fastly's OHTTP Relay. Fastly does not have the keys to decrypt messages.

- **You will not be able to log any personally identifiable information.** No personally identifying data is available for log delivery.

Implementing OHTTP Relay

To implement Fastly's OHTTP Relay, you must contact Fastly at sales@fastly.com to begin the onboarding process. As part of that process, you will be expected to provide Fastly's Professional Services team with a frontend hostname for the relay service and a backend hostname for the gateway service through which headers will pass.

In addition, you can also request the inclusion of additional HTTP headers beyond `Content-Type`, `Content-Length`, and `Host` that should not be stripped from requests and responses during validation. If you specify additional headers, you must confirm that they will not contain personally identifiable information that can be linked to customers.

Our Professional Services staff will use this information to guide you through the onboarding process as part of the initial setup and configuration process for your Fastly service.

Once your service configuration settings are confirmed, they will be enabled for you by Fastly. You will have a Fastly account created for you and will be assigned the role of User so that you can view real-time and historical stats about your service. As a standard User, you will not be able to directly control and make changes to your OHTTP-enabled service. Requests for service configuration changes can be submitted directly to Fastly via support@fastly.com.

Billing

We bill you for OHTTP Relay based on a combination of bandwidth (per GB) and requests (per 10,000) for content delivered to clients from Fastly and then for bandwidth for traffic sent from Fastly to your customers' origin.

> Some Fastly products can be purchased directly in the web interface on the Products page. For more details about a product, including pricing information or for help purchasing it, contact your account manager or email sales@fastly.com.

---

* * *

Origin Connect

Origin Connect

Last updated: 2024-12-10

/products/origin-connect

Origin Connect provides you with a direct fiber connection between your origin servers and a Fastly shield POP thus reducing the number of organizations (and by association, the number of servers) handling your data.

Prerequisites

To be considered for Origin Connect, you need to:

- have at least one Fastly shield POP configured
- have servers in the same data center as the selected Fastly shield POP (e.g., IAD, AMS, SJC)
- be interviewed by Fastly so we can identify your customer-specific business needs
- have Enterprise-level support
- have a publicly routed Autonomous System Number (ASN)

If you are approved for Origin Connect, we'll issue you with a Letter of Authorization and Connecting Facility Assignment (LOA-CFA) that the data center provider will need when you order your cross-network connection (or cross connect). You will need to pay for the cross connect with your facility provider.

For each cross connect, you, as subscriber, will need to provide Fastly with:

- a minimum of a globally unique (non RFC-1918) /31 IPv4 network prefix
- a minimum of a /127 IPv6 network prefix
- a 100G or 10G port (we recommend two and will accept up to 4× 100/10G ports for redundancy)

Both you, as the subscriber, and Fastly will each need to:

- provide the ASN intended for Border Gateway Protocol (BGP) peering use

- provision BGP peering on each interconnect
- provide a BGP prefix filter list
- comply with any other reasonable request to technically provision the Origin Connect product

If the cross connect is not completed within 90 days, the authority granted by the LOA-CFA expires.

In the event of Origin Connect service degradation, congestion, or a failure of one of these interconnects, public internet transit will be used for origin connectivity, and the subscriber will prefer the carrier of Fastly's reasonable request. There is no Service Level Agreement (SLA) available for Origin Connect.

If your origin server is located within a cloud storage provider or your traffic doesn't meet our minimum threshold for Origin Connect, contact us at sales@fastly.com to discuss other options.

---

* * *

PCI-Compliant Caching and Delivery

Last updated: 2018-08-01

/products/pci-compliant-caching-and-delivery

We have designed Fastly's core CDN service with Payment Card Industry Data Security Standard (PCI DSS) compliance in mind. With proper authorization on your account, you can use Fastly's `beresp.pci` VCL variable to automatically cache content in a manner that satisfies PCI DSS requirements.

Adding the `beresp.pci` variable to an object prevents writing of that object to non-volatile disk storage on the edge. Combined with frontend and backend TLS, this feature allows you to cache and transmit flagged content through the Fastly network in compliance with our PCI certification.

Contact sales-ecommerce@fastly.com for more information on how to enable this product for your account.