

CloudFront & Global Accelerator

AWS CloudFront

Content delivery network, they cache content at different locations world-wide that are closer to the user to improve read performance.

CloudFront has DDos protection since it the network is world wide and have integration with shield.

Origins

There are many resources that CloudFront can get the original content from, remember it is still a server that is caching content so it has to get the content from somewhere.

1. S3 bucket: You can get files from a S3 bucket and cache them in the CDN. OAC (origin access control) is used for security. OAC is only for S3 bucket, and it will permit request to be only from CloudFront.
2. HTTP backend: You can also get content from a HTTP endpoint, either from ALB, EC2 instances, for S3 websites. They can all be cached in the CDN because they are still serving contents

CloudFront vs S3 cross region replication

CloudFront is used for CDN, it has over 200 servers forming as a network to speed up the delivering content to the user. The files will be cached for up to a TTL, so it will be expired. Good for static content.

S3 replication on the other hand has to be configured to do replication on other regions. So it is only good if you only need to speed up the content deliver for a couple region and not world wide. There will be no caching happening because the files are stored. Good for dynamic content.

Using CloudFront

After you have setup the origin for CloudFront you will be getting a new DNS name, that DNS name will be used to access the content that you would like to be cached in the network, to be fast deliver to the customer.

ALB or EC2 as origin

Since CloudFront's edge location (servers) aren't part of VPC, in order to make ALB or EC2 as the origin you have to make them public for the CloudFront to access it. The user will be visiting the CloudFront DNS name, then CloudFront will make the request on behalf of the user and then cache the content, to serve it quicker.

CloudFront geo restriction

Restrict who can access your CDN based on the country the user is visiting from.

You can set up allowlist (whitelist) or blacklist.

This is good for control access to contents. Very similar to geo location with Route 53, to restrict content.

Cache invalidation

CloudFront won't know if you updated your content that it needs to refresh the cache since the TTL is long. It will only refresh only after TTL is expired.

But you can force an entire or partial cache refresh by using CloudFront Invalidation. you can invalidate all files or a specific path of the origin. Invalidated cache will refetch the content from the origin.

CloudFront pricing

Because CloudFront is a world-wide CDN it is priced differently based on which edge location (server) is used to retrieve the content.

Some countries are going to be more expensive than other of course.

There are three pricing models for CloudFront

1. Price class all: You pick all regions, you will use ALL CDN servers, this will get you best performance but at cost of those higher pricing for some countries
2. Price class 200: Get you CDN in most regions, but exclude the most expensive region
3. Price class 100: Only pick the least expensive regions.

Global accelerator

So to understand why this service is needed here is the problem it solves: Say you deployed an application in India but you have users all over the world. The customer in say US would have to do many packet hops in order to reach the application in India. We want to minimize that latency as

fast as possible.

Unicast IP: One server hold one IP address, this is the IP address that we are familiar with

Anycast IP: All server will hold the same IP and client is routed to the nearest one

Global accelerator solves the problem of needing user's packet to reach far away via public internet by minimizing the amount of public internet traffic that they need to go through. When you use global accelerator you will get 2 **Anycast IP** that you can use for your application (again they will reroute the traffic to the closest server to them).

The Anycast IP route traffic to the edge locations (the CDN servers) which then sends traffic to your actual application in AWS through the private AWS traffic (which is much faster and optimize and have less traffic from the public internet).

Benefits

You can use global accelerator for elastic IP, EC2 instances, ALB, NLB, public or private resources

You will get consistent performance because it routes with lower latency. It has built-in health check for fast failover.

It also have the same DDos protection because it is using the edge location like CloudFront.

CloudFront vs Accelerator

They both use same edge location network all around the world. Both use AWS Shield for DDos protection

CloudFront is used for caching purposes and the content is delivered from the edge location not from your actual application. It can also used for serving dynamic content.

Accelerator on the other hand is meant to minimize public internet hops, and the traffic is sent to the private AWS network to speed up the traffic. There is no caching going on because the content is served from directly from your application. And it will have fast failover due to built-in health checks.

Revision #1

Created 20 February 2023 00:37:16 by Tamarine

Updated 20 February 2023 01:53:12 by Tamarine