CloudFront

12. CloudFront

Amazon Cloudfront is a Web service that gives businesses and Web application developers an easy and Cost effective way to distribute content with low latency and high data transfer Speed.

- Cloudfront is a global service.
- Amazon cloudfront is a web service that speeds up distribution of your Static and dynamic web content such as HTML, CSS, JS and image files to your users.
- Amazon cloud front is a global content delivery network service that accelerates delivery of your websites, API, video content or other web assets.
- The major issues now a days with file or site access over the internet is latency, hence there is another layer of service added over your normal web app infra that is called CDN (connect delivery network)
- CDN provider has multiple storage or you can say cache locations distributed all over the world, which help data/web-app access easy and fast with reduced latency.
- Such CDN service is provided by Amazon with the help of Cloud Front.
- Cloud front dealers you are content through a worldwide network of data centers called edge locations.
- When a user request content that you are serving with cloudfront the user is routed (via DNS resolution) to edge location that provides the lowest latency so that content is delivered with the best possible performance.
- If the content is already in the edge location with the lowest latency cloudfront delivers to immediately.
- This dramatically reduces the numbers of networks that you are users request must pass through which improves performance.
- If not, cloud font retrievers it from an Amazon S3 bucket or an HTTP or web server that you have identified as the source of the definitive version of content (original server).
- Cloudfront also keeps persistent connection with origin server so files are fetched from the Origins as quickly as possible.

You can access Amazon cloud front in the following ways-

1. AWS management console

2. AWS SDKs

3. Cloudfront API

4. AWS Command Line Interface

Cloudfront Edge Locations-

- Edge locations are not tide to availability zones or regions.
- CloudFront delivers your content through a worldwide network of data centers called edge locations. When a user requests content that you're serving with CloudFront, the request is routed to the edge location that provides the lowest latency (time delay), so that content is delivered with the best possible performance.
- Amazon CloudFront uses a global network of 310+ Points of Presence (300+ Edge locations and 13 regional mid-tier caches) in 90+ cities across 47 countries.
- https://aws.amazon.com/cloudfront/features/?whats-new-cloudfront.sort-by=item.additionalFields.postDateTime&whats-new-cloudfront.sort-order=desc

Cloudfront – Regional Edge Cache

TechData-Infinity-Devops with MultiCloud

- Amazon cloud front has added several regional edge cache locations globally at close proximity to your viewers.
- They are located between your origin web server and the Global edge locations that serve content directly to you are viewer.
- As objects become less popular individual edge locations may remove those objects to make room for more popular content.
- Regional edge cache working as a alternative of origin to reduce the burden of origin.
- Regional edge cache have a large catchy with than any individual edge location so object remain in the catch a longer at the nearest regional edge caches.
- Regional Edge Caches have larger cache-width than any individual edge location, so your objects remain in cache longer at these locations.
- This helps keep more of your content closer to your viewers, reducing the need for CloudFront to go back to your origin webserver, and improving overall performance for viewers

CloudFront Regional Edge Cache- Working

- When a viewer makes a request on your website or through your application, DNS routes the request to the cloud front edge location that can based serve the users request.
- This location is typically the nearest cloudfront edge location in terms of latency.
- In the edge location, cloudfront checks its cache for the requested files.
- If the files are in the catch a cloudfront returns them to the user.
- If the files are not in the cache, the edge servers go to the nearest regional edge cache to fetch the object.
- Regional edge catches have feature parity with edge locations for example _ a catch a invalidation request remove an object from both edge catch catches and regional edge catches before it expires.
- The next time a viewer request the object cloudfront returns to the origin to fetch the latest version of the Object.
- Proxy method PUT/ POST/ PATCH/OPTIONS/DELETE go directly to the origin from the edge locations and do not proxy through the original edge catches.
- Dynamic content as determined at request time, does not flow through regional edge cache, but goes directly to the Origin.

Features of CF

- Can support both static and dynamic content
- Supports PUT/POST and other http methods.
- Can add custom SSL certificate.
- Invalidation enabled.
- We can add custom error Reponses.
- Low TTL (time to live)
- Geo targeting
- Wildcard/zone apex certificate support

New Concepts:

- TTL (Time to Live) :-

1.Time to live (TTL) is the time that an object is stored in a caching system before it's deleted or refreshed.

2. Minimum TTL Specify the minimum amount of time, in seconds, that you want objects to stay in CloudFront caches before CloudFront forwards another request to your origin to determine whether the object has been updated.

3. The default value for Minimum TTL is 0 seconds. Maximum is 1 year and default value is 1 day. These values are specified in seconds.

- Invalidation:-
    1. Amazon Cloudfront's invalidation feature, which allows you to remove an object from the CloudFront cache before it expires.
    2. To invalidate files, you can specify either the path for individual files or a path that ends with the * wildcard, which might apply to one file or to many, as shown in the following example:
        - /images/image1.jpg
        - /images/image*
        - /images/*
        - Pricing

You incur CloudFront charges when CloudFront responds to requests for your objects.

https://aws.amazon.com/cloudfront/pricing/

How much load can CF cache server handle?

Amazon Cloud front can handle 250,000 requests per second per distribution. The data transfer rate per distribution is 150 Gbps, the maximum length of the request including headers, query strings is 20,480 bytes

The maximum size of a single file that can be delivered through Amazon CloudFront is 20 GB. This limit applies to all Amazon CloudFront distributions.

https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/cloudfront-limits.html

- Use Cases:-

    1. Accelerate static website content delivery. (S3 combined)

2. CloudFront can speed up the delivery of your static content to viewers across the globe.

https://aws.amazon.com/blogs/networking-and-content-delivery/amazon-s3-amazon-cloudfront-a-match-made-in-the-cloud/

3. Cloud front can be attached over the load balancer to reduce some load and latency.