AWS Solution Architect Associate Certification Training – Module 18

18. Cloud Front

Introduction to Cloud Front distributions and Cloud Front Edge locations

Amazon Cloud Front 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. Cloud Front 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 user is routed to the edge location that provides the lowest latency (time delay), 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 it immediately.
- If the content is not in that edge location, CloudFront retrieves it from an origin that you've defined—such as an Amazon S3 bucket, a MediaPackage channel, or an HTTP server (for example, a web server) that you have identified as the source for the definitive version of your content.

As an example, suppose that you're serving an image from a traditional web server, not from CloudFront. For example, you might serve an image, sunsetphoto.png, using the URL http://example.com/sunsetphoto.png.

Your users can easily navigate to this URL and see the image. But they probably don't know that their request was routed from one network to another—through the complex collection of interconnected networks that comprise the internet—until the image was found.

CloudFront speeds up the distribution of your content by routing each user request through the AWS backbone network to the edge location that can best serve your content. Typically, this is a CloudFront edge server that provides the fastest delivery to the viewer. Using the AWS network dramatically reduces the number of networks that your users' requests must pass through, which improves performance. Users get lower latency—the time it takes to load the first byte of the file—and higher data transfer rates.

You also get increased reliability and availability because copies of your files (also known as *objects)* are now held (or cached) in multiple edge locations around the world.

An edge location is where end users access services located at AWS. They are located in most of the major cities around the world and are specifically used by CloudFront (CDN) to distribute content to end user to reduce latency. It is like frontend for the service we access which are located in AWS cloud.

Access Control and security

To perform any operation on CloudFront resources, such as creating a web distribution or invalidating an object, AWS Identity and Access Management (IAM) requires you to authenticate that you're an approved AWS user. If you're using the CloudFront console, you authenticate your identity by providing your AWS user name and a password. To create, update, delete, or list CloudFront resources, you need permissions to perform the operation, and you need permissions to access the corresponding resources.

Content Caching

One of the purposes of using CloudFront is to reduce the number of requests that your origin server must respond to directly. This reduces the load on your origin server and also reduces latency because more objects are served from CloudFront edge locations, which are closer to your users.The more requests that CloudFront is able to serve from edge caches as a proportion of all requests (that is, the greater the cache hit ratio), the fewer viewer requests that CloudFront needs to forward to your origin to get the latest version or a unique version of an object.