Section 12 : Leveraging the AWS Global Infrastructure

* Why make a global application?

- A global appl^{n} is an application deployed in multiple geographies

- On AWS : this could be Regions and/or Edge locations

- Decreased Latency :

1] Latency is the time it takes for network packet to reach a server.

2] It takes time for a packet from Asia to reach the US.

3] Deploy your applications closer to your users to decrease latency, better experience

- Disaster Recovery :-

1] If AWS region goes down (earthquake, storms, power shutdown, politics)....

2] You can fail-over to another region & have your application still working

3] A DR plan is important to increase the availability to your application.

- Attack Protection :- ~~Dis Disaster~~

Distributed global infrastructure is harder to attack.

Global Applications in AWS

① Global DNS : Route 53

- Great to route users to the closest deployment with least latency.
- Great for disaster recovery strategies.

② Global Content Delivery Network (CDN): CloudFront

- Replicate part of your application to AWS Edge locations - decrease latency
- Cache common requests - improved user experience & decreased latency.

③ S3 Transfer Acceleration:

- Accelerate global uplods & downlods into Amazon S3.

④ AWS Global Accelerator:

- Improve global appl^n availability & performan using the AWS global network.

Amazon Route 53 Overview

- Route 53 is a Managed DNS (Domain Name System)
- DNS is a collection of rules & records which helps clients understand how to reach a server through URLs.

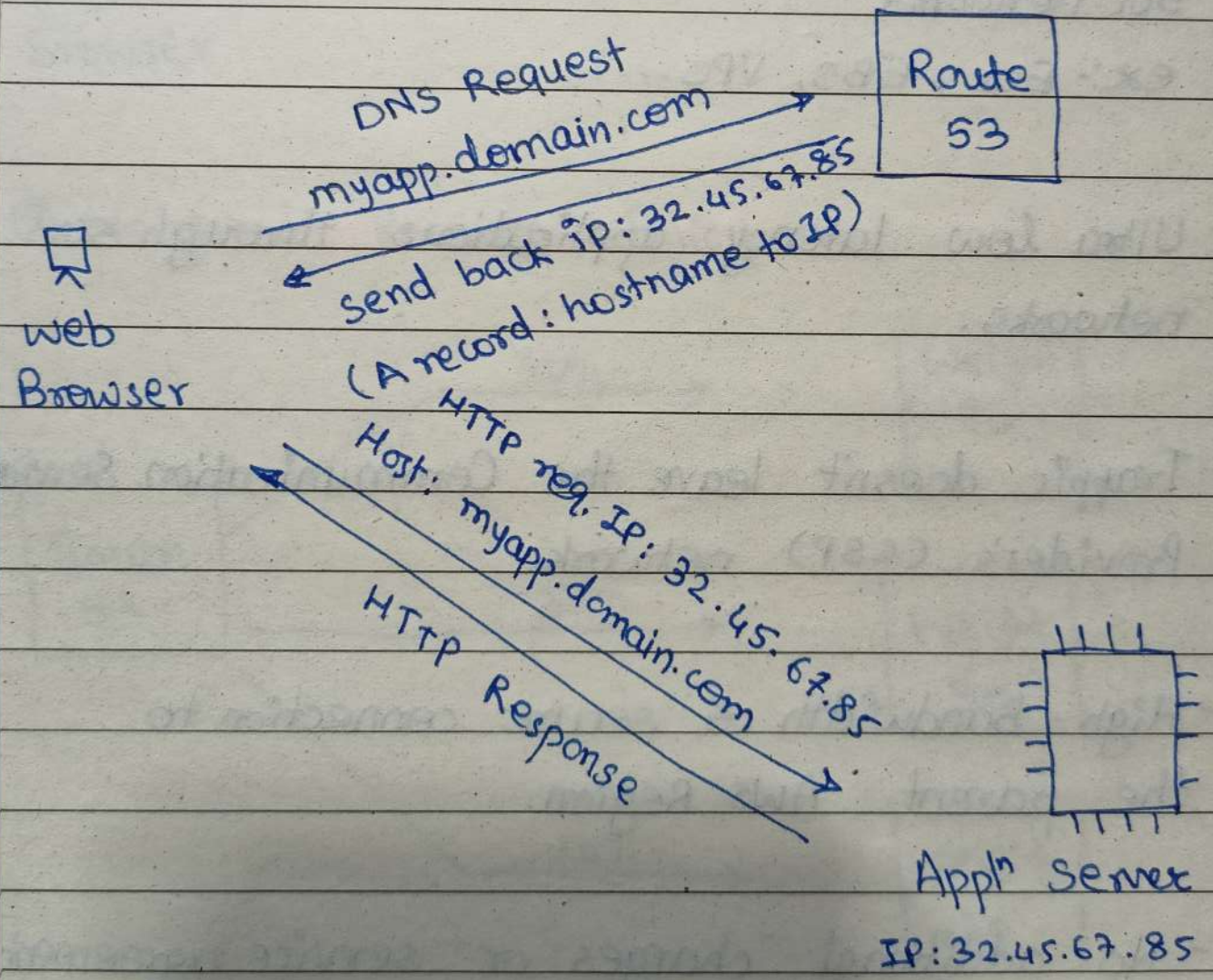
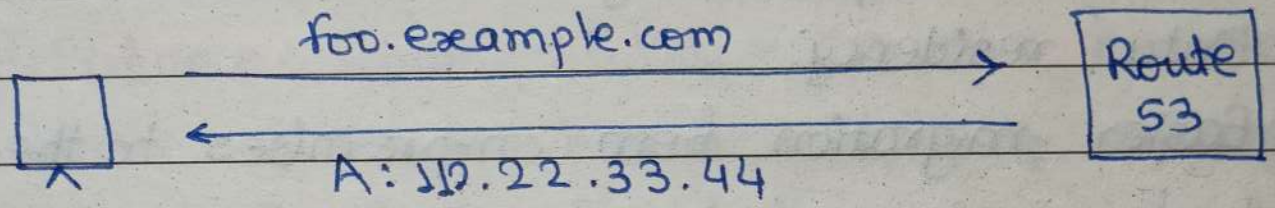


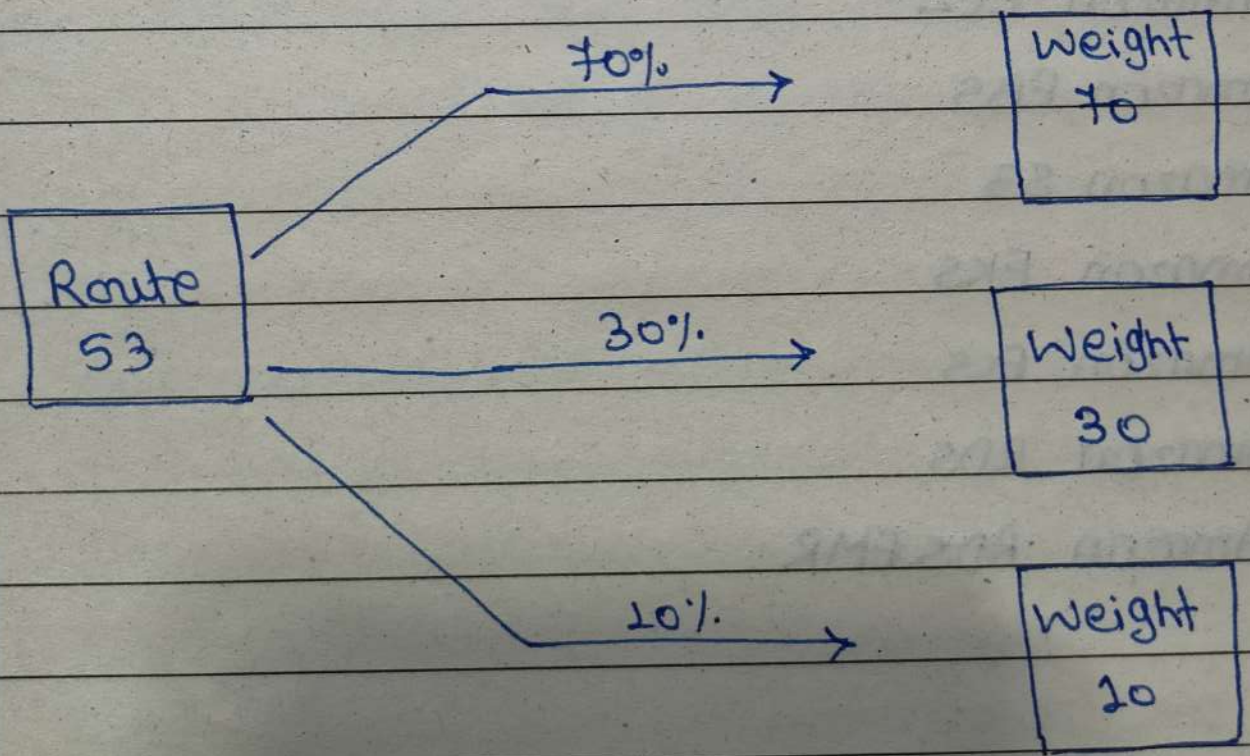
fig. Route 53 : Diagram for a Record

Route 53 Routing Policies

① Simple Routing Policy: No health checks



② Weighted Routing Policy



③ Latency Routing Policies :-
Based on location of user, it will route user request to closest server.

④ Failover Routing Policy :- Disaster Recovery
Router will do health check of primary server if it is failed then route user request to next server.

AWS CloudFront

- Content Delivery Network (CDN)
- Improves read performance, content is cached at the edge.
- Improves user experience.
- 216 point of presence globally (edge loc.)
- DDOS Protection (because worldwide), integration with Shield, AWS web application Firewall (WAF)

* CloudFront - Origins

① S3 Bucket :-

For distributing files & caching them at the edge.

- Enhanced security with cloudfront Origin Access Control (OAC)
- OAC is replacing Origin Access Identity (OAI)
- CloudFront can be used as an ingress (to upload files to S3)

AWS Global Accelerator vs Cloudfront

* Custom Origin (HTTP):-
- Application load balancer
- EC2 instance
- S3 website (must first enable the bucket as a static S3 website)
- Any HTTP backend you want.

CloudFront Vs S3 Cross Region Replication

① CloudFront :-

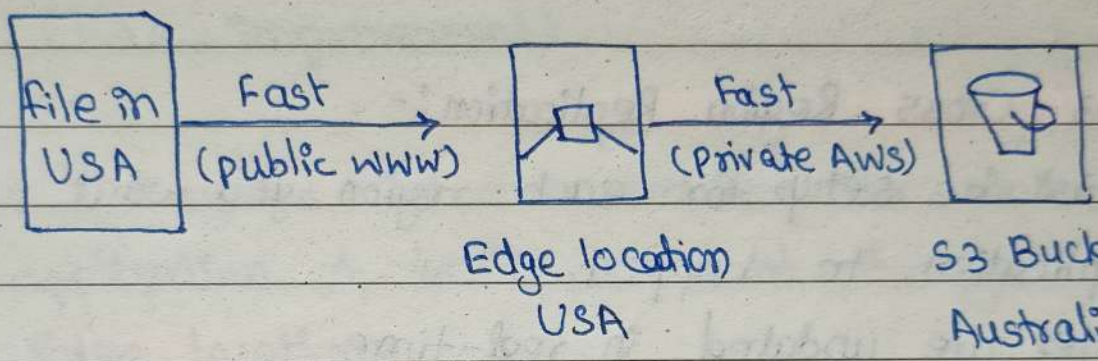
1. Global Edge network.
2. Files are cached for TTL (Maybe a day)
3. Great for static content that must be available everywhere.

② S3 Cross Region Replication :-

1. Must be setup for each region you want replication to happen.
2. Files are updated in real-time
3. Read only
4. Great for dynamic content that needs to be available at ~~you~~ low latency in few regions.

S3 Transfer Acceleration

- Increase transfer speed by transferring file to an AWS edge location which will forward the data to the S3 bucket in the target region



AWS Global Accelerator

- Improves global application availability & performance using the AWS global network.

- leverage the AWS internal network to optimize the route to your application (60% improvement)

- 2 Anycast IP are created for your application & traffic is sent through Edge locations.

- The Edge locations send the traffic to your application.

AWS Global Accelerator Vs CloudFront

- They both use AWS global network & its edge locations around the world.

- Both services integrate with AWS shield for DDoS protection.

- **CloudFront - Content Delivery Networks**
- Improves performance for your cacheable content (such as images & videos)
- Content is **served at the edge.**

- **Global Accelerator**
- No caching, **proxying packets** at the edge to appli^ns running in one or more AWS Regions.
- Improves performance for a wide range of applications over TCP or UDP.
- Good for HTTP use cases that requires static IP addresses.
- Good for HTTP use cases that required deterministic, fast regional failover.

AWS Outposts

- **Hybrid Cloud**: business that keep an on-premi-ses infrastructure alongside a cloud infrastructure

- Therefore, two ways of dealing with IT systems:
  ① One for AWS cloud (using the AWS console, CLI, & AWS APIs).
  ② One for their on-premises infrastructure.

- AWS Outposts are "server racks" that offers the same AWS infrastructure services, APIs & tools to build your own applications on-premises just as in the cloud.

- AWS will setup & manage "Outposts Racks" within your on-premises infrastructure and you can start leveraging AWS services on-premises.

- You are responsible for the Outposts Rack physical security.

#Benefits of Aws Outposts :→

① Low-latency access to on-premises systems.
② Local data processing
③ Data residency
④ Easier migration from on-premises to the cloud.
⑤ Fully managed service.

• Some services that work on Outposts :-

① Amazon EC2
② Amazon EBS
③ Amazon S3
④ Amazon EKS
⑤ Amazon ECS
⑥ Amazon RDS
⑦ Amazon ~~RDS~~ EMR

AWS Wavelength

- Wavelength Zones are infrastructure deployments embedded within the telecommu-nications providers datacenters at the edge of 5G networks.

- Brings AWS services to the edge of the 5G networks.
  ex:- EC2, EBS, VPC.....

- Ultra low latency applications through 5G networks.

- Traffic doesn't leave the Communication Service Provider's (CSP) network.

- High-bandwidth & secure connection to the parent AWS Region.

- No additional charges or service agreement

- Use Cases: Smart Cities, ML-assisted diagnostics, Connected Vehicles, Interactive Live video streams, AR|VR, Real-time Gaming..

AWS Local Zones

- Places AWS Compute, storage, database & other selected AWS services closer to end users to run latency-sensitive applications.

- Extend your VPC to more locations - "Extension of an AWS Region"

- Compatible with EC2, RDS, ECS, EBS, Elasticache, Direct connect....

- Example :
  - AWS Region : N.Virginia (us-east-1)
  - AWS Local Zones : Boston, Chicago, Dallas, Houston, miami....

Global Applications Architecture

① Single Region, Single AZ :-

High Availability - X

Global Latency - X

Difficultly - low

② Single Region, Multiple AZ :-

High Availability - ✓

Global Latency - X

Difficulty - medium

③ Multi Region, Active-Passive

- write in only one Active Availability zone.
- Read can be performed from any active / passive availability zone.

So,

Global reads Latency - ✓

Global writes Latency - X

Difficulty - medium

④ Multi Region, Active - Active

- Here, all availability zones are active so, Read & write can be performed on any availability zone.

So,

Global Read's Latency - ✓

Global Write's Latency - ✓

Difficulty - High

Global Application in AWS - Summary

① Global DNS - Route 53

- Great for route users to the closest deployment with least latency.
- Great for disaster recovery strategies.

② Global Content Delivery Network (CDN): Cloudfront

- Replicate part of your application to AWS Edge locations - decrease latency.
- Cache common requests - improved user experience & decreased latency

③ S3 Transfer Acceleration:-

- Accelerate global uploads & downloads in S3.

④ AWS Global Accelerator:-

- Improves global applications availability & performance using the AWS Global

⑤ AWS Outposts :-

Deploy Outposts Racks in your own Data Centers to extend AWS services

⑥ AWS Wavelength :-

- Brings AWS services to the edge of the 5G networks.
- Ultra-low latency applications

⑦ AWS Local Zones :-

- Bring AWS resources (compute, database, storage, ...) closer to your users.
- Good for latency sensitive applications.