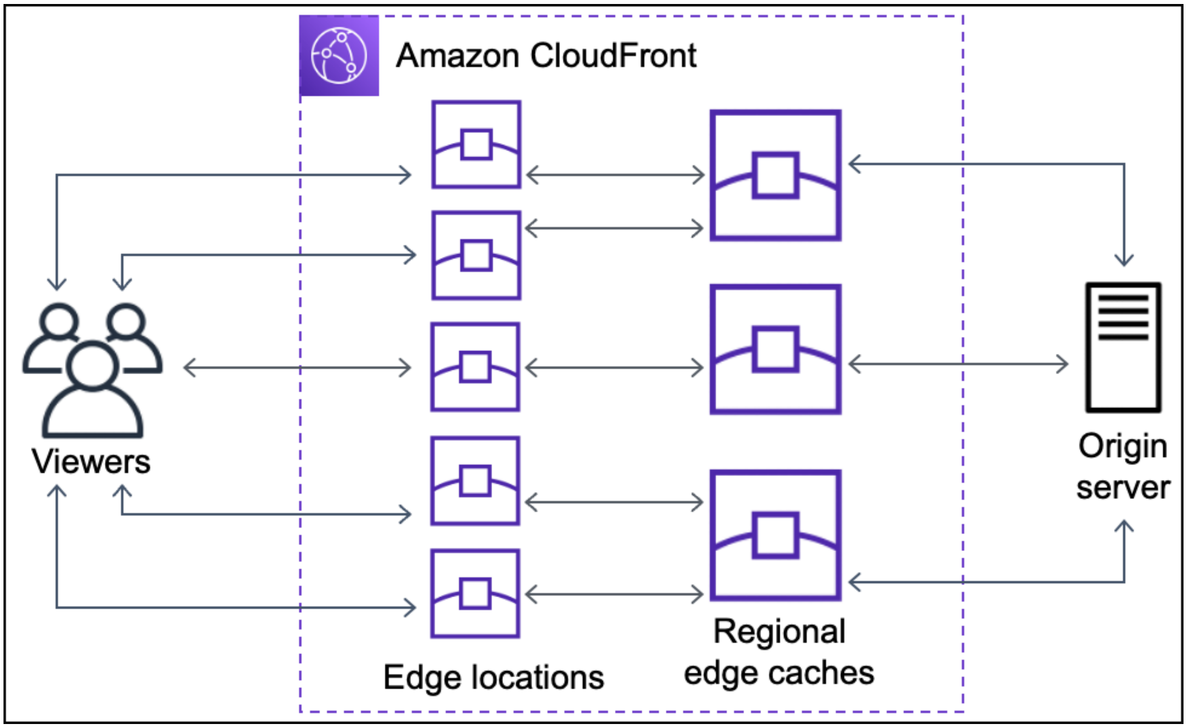
NETWORKING
IN
AWS

CloudFront

AWS CloudFront is a fast and highly secure content delivery network (CDN) service that accelerates the delivery of your websites, APIs, video content, and other web assets. It improves the performance of your applications by caching content at edge locations around the world, reducing latency and providing a better user experience.



CloudFront is often used to improve performance and reduce latency not only for static content stored in S3 but even dynamic content stored behind say, Application Load balancers and EC2 instances. (i.e. Cloudfront distributions can have multiple services as their origins.)

Note:

Cloudfront distributions also have AWS Shield Advanced installed by default and therefore offer DDoS protection.

Global Accelerator

AWS Global Accelerator is designed to reduce latency and improve data transfer speeds by using the AWS global network. It features automatic failover, ensuring high availability and reliability. This makes it an excellent fit for non-HTTP use cases such as gaming (UDP), IoT, and Voice over IP, where low latency and fast data transfer are critical.

Additionally, Global Accelerator is suitable for static HTTP use cases, providing a consistent and reliable experience.

Note: Global Accelerator has automatic failover

Cross Origin Resource Sharing

Cross-Origin Resource Sharing (CORS) is a security feature in web browsers that controls how web pages can request resources from a different domain. It ensures safe interactions between web pages and resources across different origins, protecting against cross-site scripting (XSS) attacks.

In AWS, CORS is essential for services like Amazon S3 and API Gateway. For S3, it allows web applications to directly request and interact with S3 buckets from different domains, useful for functionalities like file uploads.

With API Gateway, CORS settings enable APIs to be accessed from web applications on different domains. This integration supports client-side applications connecting to serverless backend services securely and efficiently.

Very Important:

Amazon CloudFront also leverages CORS through its `Access-Control-Allow-Origin` header. This header specifies which origins are permitted to access resources on a CloudFront distribution. By configuring `Access-Control-Allow-Origin`, you can control which domains can make cross-origin requests to your CloudFront-distributed content, ensuring secure and efficient delivery of static and dynamic content across different domains.

Direct Connect

AWS Direct Connect is a network service that establishes a dedicated, private connection between your on-premises infrastructure and AWS.

This high-speed, low-latency connection helps improve performance and provides a more consistent network experience compared to standard internet connections.

Direct Connect is ideal for applications requiring stable and high-throughput data transfer, such as large-scale data migrations, hybrid cloud architectures, and real-time data processing.

By bypassing the public internet, it enhances security and reduces the risk of interruptions, ensuring reliable and efficient connectivity to your AWS resources.

Note:

Direct Connect allows for direct connections to AWS Servers, and does not affect bandwidth at all.

Subnets

Subnets in AWS (Amazon Web Services) refer to segmented portions of a Virtual Private Cloud (VPC), which allow you to logically isolate and group resources within your cloud infrastructure. Each subnet resides in a specific Availability Zone (AZ) and can span multiple AZs for high availability and fault tolerance. Subnets are used to allocate IP addresses to resources (like EC2 instances) and enforce network access controls through Network Access Control Lists (ACLs) and Security Groups. They play a crucial role in organizing and securing your AWS resources within a VPC environment.

Note:

VPCs can span Multiple Availability Zones but each subnet maps to a single Availability Zone.

BlockSize of a VPC

A Virtual Private Cloud (VPC) is a virtual network in the AWS cloud where you can launch AWS resources. VPCs help you logically isolate and control access to your cloud resources.

Block size in the context of VPC refers to the range of IP addresses that can be used to define subnets. AWS VPC allows different block sizes, typically ranging from /16 (65,536 IP addresses) to /28 (16 IP addresses).

Note: The following IP addresses for every subnet are reserved by AWS:

.0

.1

.2

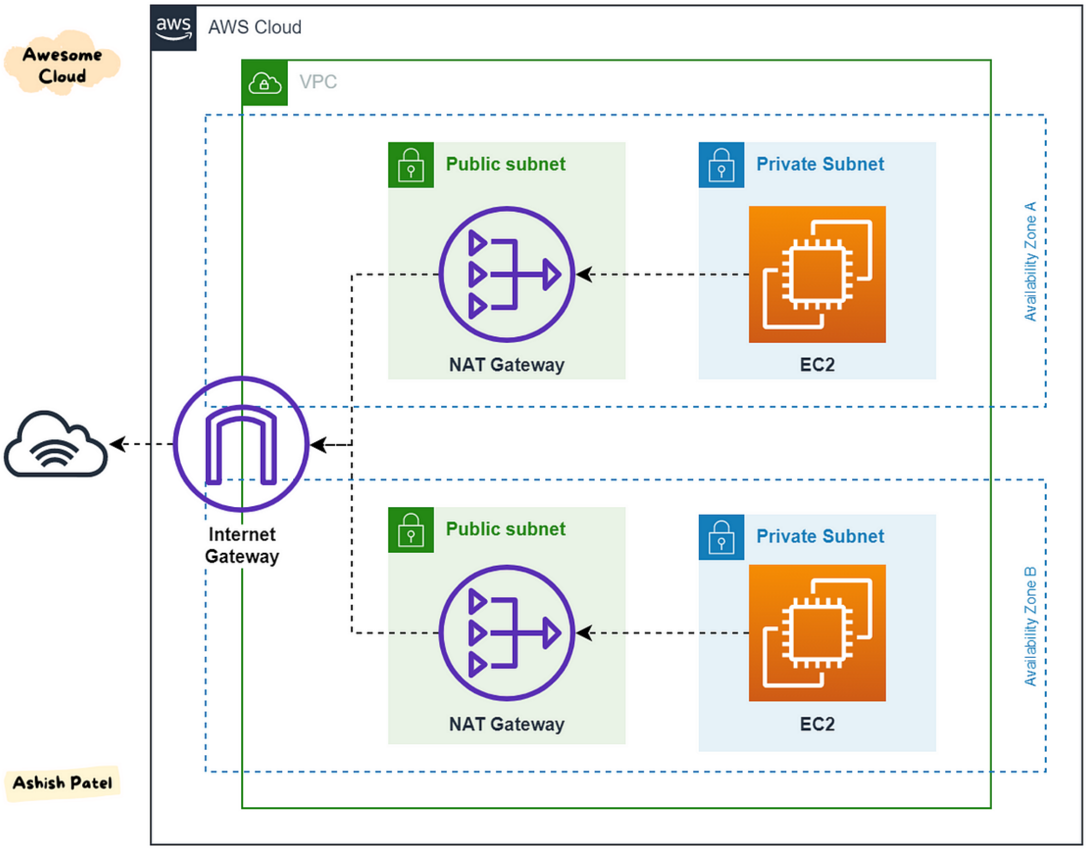
.3

and

.255

NAT Gateways

A NAT gateway is a Network Address Translation (NAT) service. You can use a NAT gateway so that instances in a private subnet can connect to services outside your VPC but external services cannot initiate a connection with those instances.



When you create a NAT gateway, you specify one of the following connectivity types:

- **Public –** (Default) Instances in private subnets can connect to the internet through a public NAT gateway, but cannot receive unsolicited inbound connections from the internet. You create a public NAT gateway in a public subnet and must associate an elastic IP address with the NAT gateway at creation. You route traffic from the NAT gateway to the internet gateway for the VPC. Alternatively, you can use a public NAT gateway to connect to other VPCs or your on-premises network. In this case, you route traffic from the NAT gateway through a transit gateway or a virtual private gateway.
- **Private –** Instances in private subnets can connect to other VPCs or your on-premises network through a private NAT gateway. You can route traffic from the NAT gateway through a transit gateway or a virtual private gateway. You cannot associate an elastic IP address with a private NAT gateway.

  You can attach an internet gateway to a VPC with a private NAT gateway, but if you route traffic from the private NAT gateway to the internet gateway, the internet gateway drops the traffic.

Routing NAT Gateway

Quick Tip:

NAT Gateways must be routed in private subnets if private EC2 instances are to talk with them.

However, NAT Gateways must be placed in a public subnet itself if internet access is required.

NAT Gateway vs NAT instance

NAT instance is you running NAT software on an EC2 and managing it. This is cheaper but the burden of management is on you.

NAT gateway is a managed service that AWS manages, you just direct outbound Internet traffic to it. This is more expensive, the AWS handles things.

NAT Gateway is newer and AWS suggests that you replace all older NAT instances with NAT gateways.

Gateway VPC Endpoint

Gateway VPC Endpoints allow you to access **S3 buckets and DynamoDB** directly from your VPC without connectivity to the internet, no NAT required.

Note: Gateway VPC Endpoints are **free** and have no additional charge associated with them.

Elastic Load Balancers

An Elastic Load Balancer (ELB) in AWS automatically distributes incoming application traffic across multiple targets, such as EC2 instances, containers, and IP addresses.

It enhances the availability and fault tolerance of your applications by balancing the load and ensuring traffic is directed to healthy instances.

ELB supports various types of load balancers, including Application Load Balancer (ALB), Network Load Balancer (NLB), and Gateway Load Balancer (GLB), each tailored for different use cases and traffic patterns.

Note: Elastic Load Balancers can only be connected to Availability Zones in a single region.

Gateway Load Balancers

Gateway Load Balancers are a new type of elastic load balancer that operates at layer 3 of the OSI model, and are marketed as being best for integration of third party network appliances.

From AWS website:

Gateway Load Balancer helps you easily deploy, scale, and manage your third-party virtual appliances. It gives you one gateway for distributing traffic across multiple virtual appliances while scaling them up or down, based on demand. This decreases potential points of failure in your network and increases availability.

Health Checks

Quick tip:

Application Load Balancers supports HTTP health checks, and is the best at it.

Network Load Balancers work with TCP/UDP protocols but is still capable of doing HTTP health checks, though ALB still better.

Also, Autoscaling Actions are a thing.

Listener Rule

Quick Tip:

Listener Rules are used in ALBs to redirect HTTP traffic to HTTPS.

API Gateway URL

To design the API Gateway URL with the company's domain name and corresponding certificate, the company needs to do the following:

1. Create a Regional API Gateway endpoint: This will allow the company to create an endpoint that is specific to a region.
2. Associate the API Gateway endpoint with the company's domain name: This will allow the company to use its own domain name for the API Gateway URL.
3. Import the public certificate associated with the company's domain name into AWS Certificate Manager (ACM) in the same Region: This will allow the company to use HTTPS for secure communication with its APIs.
4. Attach the certificate to the API Gateway endpoint: This will allow the company to use the certificate for securing the API Gateway URL.
5. Configure Route 53 to route traffic to the API Gateway endpoint: This will allow the company to use Route 53 to route traffic to the API Gateway URL using the company's

Route53 and Failover

Amazon Route 53 is a scalable Domain Name System (DNS) web service that translates domain names into IP addresses and routes users to endpoints. It offers several routing policies, including failover options like Active-Active and Active-Passive.

Active-Active failover in Route 53 allows traffic to be distributed across multiple endpoints, all of which are capable of handling requests simultaneously. This setup improves availability and responsiveness by distributing traffic among healthy endpoints.

Active-Passive failover, on the other hand, uses a standby endpoint that only becomes active when the primary endpoint fails. This configuration ensures minimal downtime by automatically redirecting traffic to the standby endpoint when the primary endpoint becomes unavailable.

Route 53's failover policies enable robust fault tolerance and high availability for applications, ensuring seamless user experience even during endpoint failures.

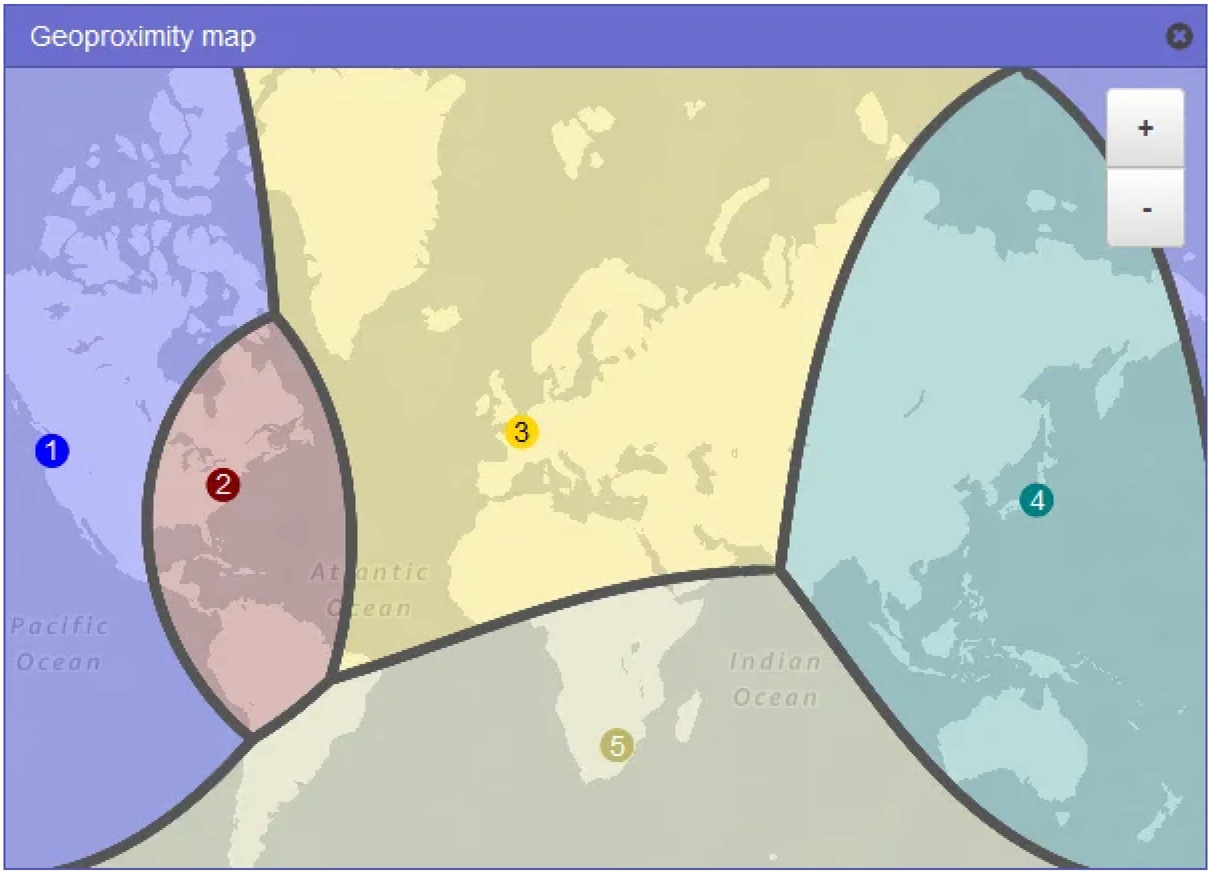
Use active-active failover configuration when you want all of your resources to be available the majority of the time. When a resource becomes unavailable, Route 53 can detect that it's unhealthy and stop including it when responding to queries.

Note: Remember that an Active-Active Failover uses all available resources all the time without a primary nor a secondary resource.

Geoproximity vs Geolocation Routing

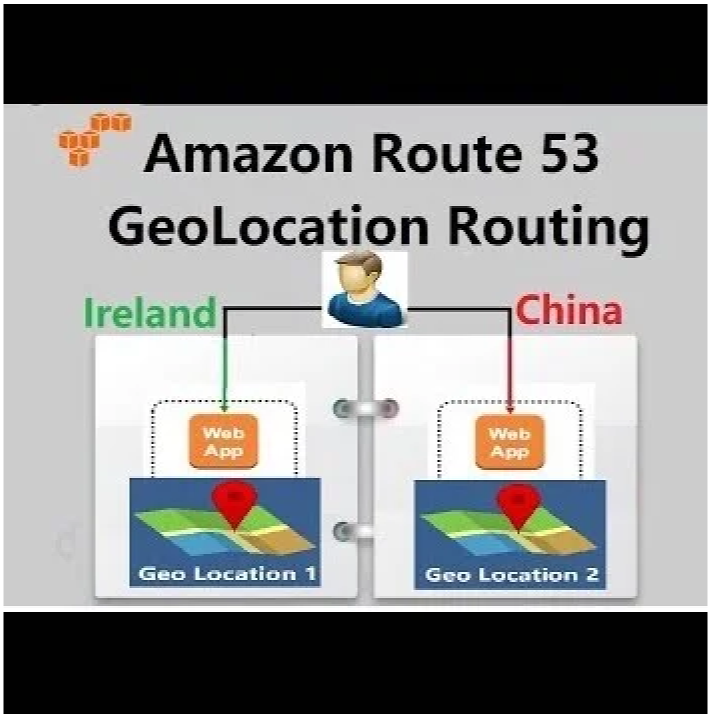
Geoproximity Routing Policy

A geoproximity routing policy is what you use when you want to route traffic based on the location of your resources, or shift traffic flow between resources. This policy allows you to direct users to different servers, even though those servers might be further away, using something called a *bias.*



Geolocation Routing Policy

A geolocation routing policy is what you use when you want to route traffic based on the location of your users. Most businesses these days have users all over the world, and they want to serve content to those users as fast as possible. A geolocation routing policy allows you to allocate the resources that serve your traffic based on the location that users' DNS queries originate from.



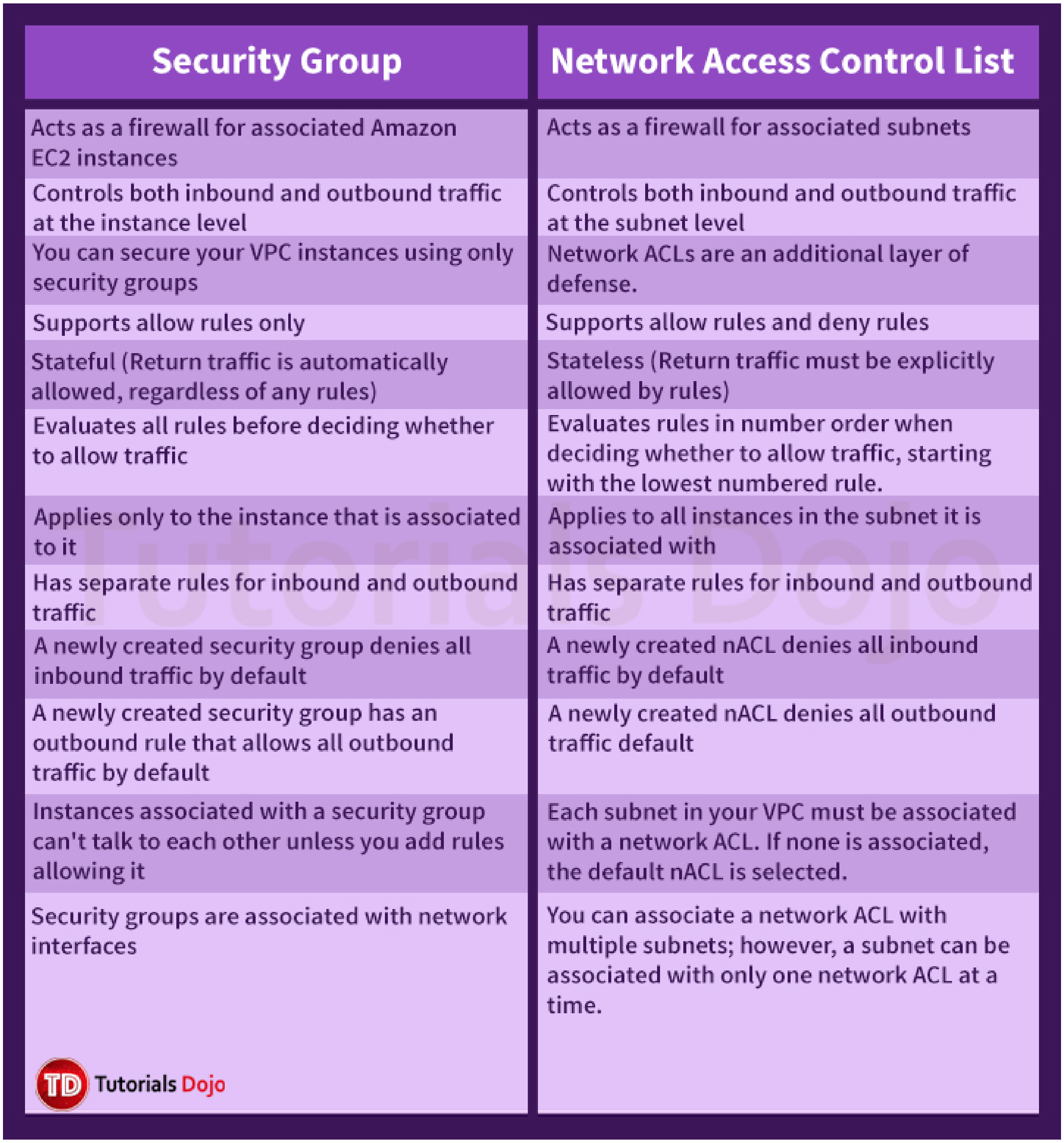
With geolocation routing, you can localize content and restrict the distribution of content to only the locations in which you are able or allowed to distribute. You can also balance the traffic load across endpoints in a predictable way.

Origin Access Identity (OAI)

Quick Tip:

OAI (Origin Access Identity) is used to give access to S3/Cloudfront info without allowing for direct navigation to the URL.

Security Groups vs NACL



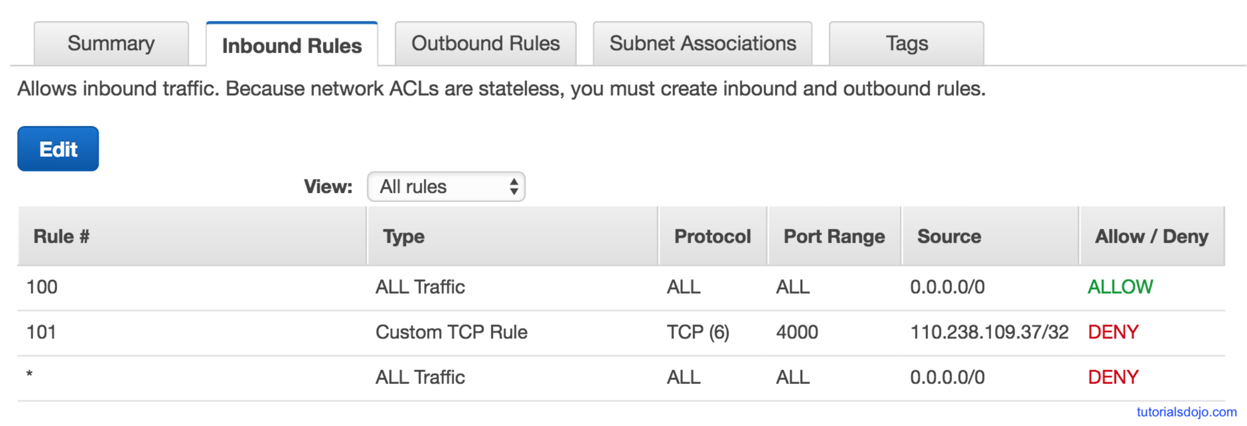
Security Group	Network Access Control List
Acts as a firewall for associated Amazon EC2 instances	Acts as a firewall for associated subnets
Controls both inbound and outbound traffic at the instance level	Controls both inbound and outbound traffic at the subnet level
You can secure your VPC instances using only security groups	Network ACLs are an additional layer of defense.
Supports allow rules only	Supports allow rules and deny rules
Stateful (Return traffic is automatically allowed, regardless of any rules)	Stateless (Return traffic must be explicitly allowed by rules)
Evaluates all rules before deciding whether to allow traffic	Evaluates rules in number order when deciding whether to allow traffic, starting with the lowest numbered rule.
Applies only to the instance that is associated to it	Applies to all instances in the subnet it is associated with
Has separate rules for inbound and outbound traffic	Has separate rules for inbound and outbound traffic
A newly created security group denies all inbound traffic by default	A newly created nACL denies all inbound traffic by default
A newly created security group has an outbound rule that allows all outbound traffic by default	A newly created nACL denies all outbound traffic default
Instances associated with a security group can't talk to each other unless you add rules allowing it	Each subnet in your VPC must be associated with a network ACL. If none is associated, the default nACL is selected.
Security groups are associated with network interfaces	You can associate a network ACL with multiple subnets; however, a subnet can be associated with only one network ACL at a time.

NACL/SG Rule priority

Network ACLs (NACLs) and Security Groups (SGs) are key components for managing security in AWS.

- **NACLs**: Stateless, apply at the subnet level, and control inbound and outbound traffic through rules evaluated in order. They allow or deny traffic based on IP addresses and protocols.
- **Security Groups**: Stateful, apply at the instance level, and control inbound and outbound traffic with rules evaluated as a collective set. They allow traffic based on port, protocol, and IP address.

Both are essential for fine-tuning access control and ensuring the security of your AWS resources.



Allows inbound traffic. Because network ACLs are stateless, you must create inbound and outbound rules.

View: All rules

Rule #	Type	Protocol	Port Range	Source	Allow / Deny
100	ALL Traffic	ALL	ALL	0.0.0.0/0	ALLOW
101	Custom TCP Rule	TCP (6)	4000	110.238.109.37/32	DENY
*	ALL Traffic	ALL	ALL	0.0.0.0/0	DENY

Lower the rule number, the higher priority it is with the "*" rule being the lowest priority and highest number rule.

Higher priority rules will overpower lower priority ones. In the above example, rule 100 will be called first, then 101, then *

AWS PrivateLink

AWS PrivateLink provides private connectivity between VPCs, AWS services, and your on-premises networks, without exposing your traffic to the public internet.

AWS PrivateLink makes it easy to connect services across different accounts and VPCs to significantly simplify your network architecture.

Interface
VPC endpoints, powered by AWS PrivateLink, connect you to services hosted by AWS Partners and supported solutions available in AWS Marketplace.