

AWS Elastic Load Balancer: Overview And Types

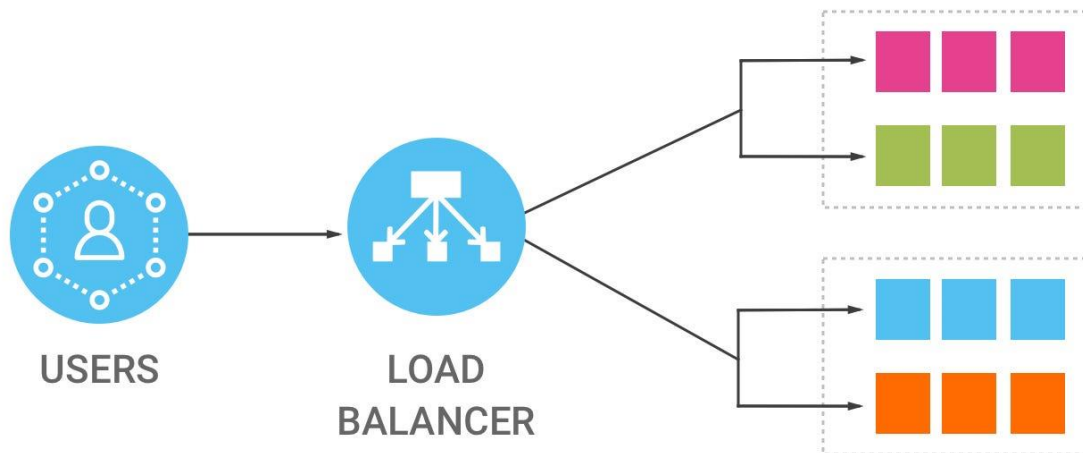
Elastic Load Balancing (ELB) is a load-balancing service for AWS deployments. It automatically distributes incoming application traffic and scales resources to fulfill high traffic demands.

What are Load Balancers and how do they work?

Load balancers detect the health of back-end resources and do not send traffic to servers that are not able to fulfill requests. **Load balancing** is the process of redistribution of workload in a distributed ensuring no **computing machine is overloaded, under-loaded, or idle**. It tries to speed up different parameters like **response time, execution time, system stability**, etc. thereby improving the performance of the cloud. ng situation. The survey papers based on load balancing could not provide a proper systematic classification of methods and techniques.

Factors responsible for load unbalancing problem and also provided some methods that can be used:

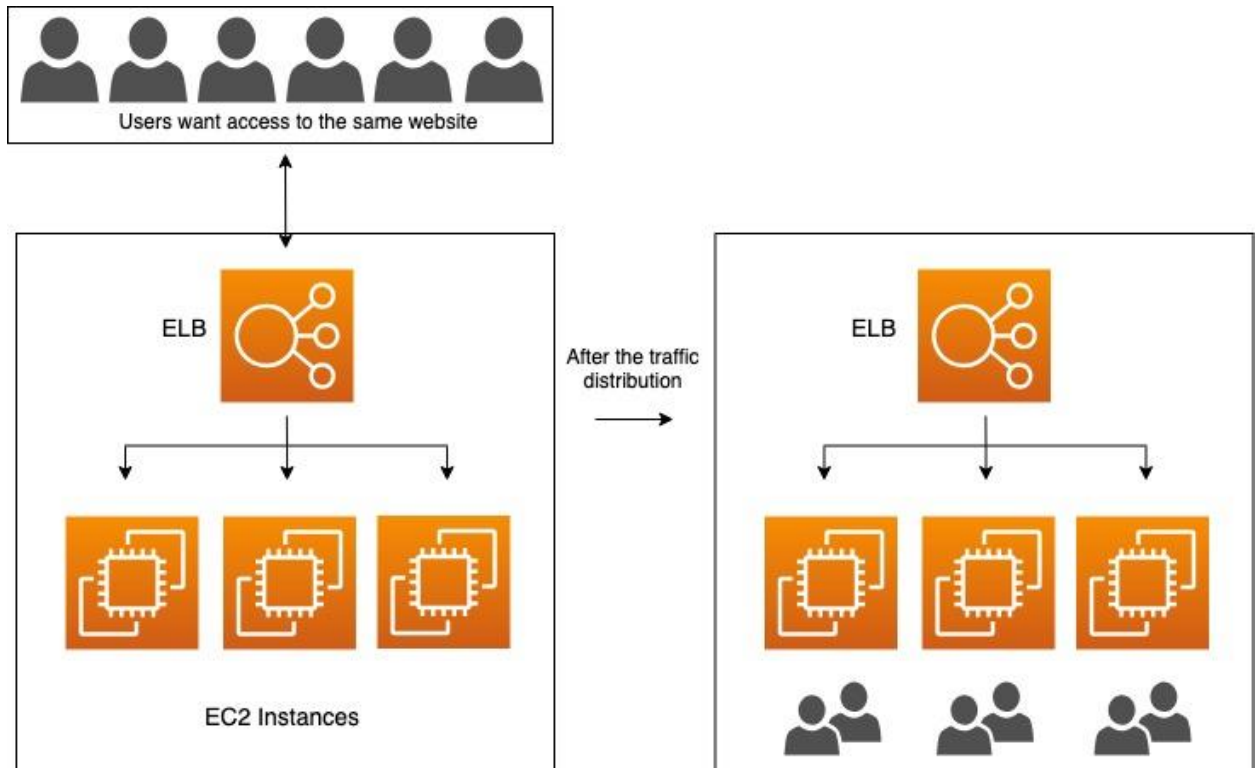
- Identify the factors that cause load unbalancing problems in cloud computing.
- Provide a systematic overview of the existing approaches that are used in the load balancing process and the way in which these approaches have been used in cloud technology.
- A physical device, a virtualized instance running on specialized hardware.
- Incorporated into application delivery controllers (ADCs) designed to more broadly improve the performance and security of web and microservices-based applications, regardless of their hosting.



Elastic Load Balancing in AWS

- Distributes incoming application traffic across multiple targets, such as EC2 instances, containers (ECS), Lambda functions, and IP addresses in multiple Availability Zones(AZs).
- Distributes Client traffic across servers.

- Improves the performance of applications.
- It Accepts incoming traffic from clients and routes requests to its registered targets.
- Monitors the condition of its registered targets and routes traffic only to healthy targets.
- Enable deletion protection to stop your load balancer from being deleted accidentally.
Disabled by default.
- Deleting ELB won't delete the instances registered to it.
- Cross Zone Load Balancing – If enabled, each load balancer node automatically distributes traffic across the registered targets in all enabled AZs.
- Supports SSL Offloading which is a feature that allows the AWS Elastic Load Balancer to bypass the SSL termination by removing the SSL-based encryption from the incoming traffic.



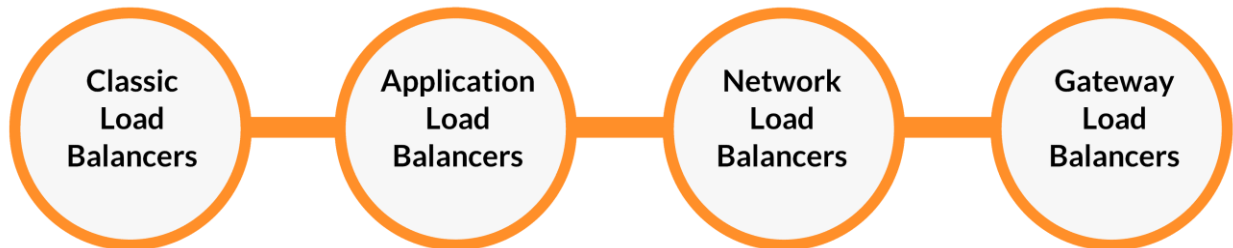
Common Features

- **High availability**
The most common service that relies on Elastic Load Balancer is Amazon's EC2 Instances, as Elastic Load Balancer performs a health checkup to confirm an instance continues to be running before sending traffic thereto. A developer can integrate Amazon Route 53 and Domain name system (DNS) failover to further boost application resiliency.
- **Automatic scaling**
A developer can use the AWS Auto Scaling feature to ensure he or she has enough EC2 instances running behind an ELB. The developer sets Auto Scaling conditions, and when a condition is met, a brand new EC2 instance can spin up to satisfy the specified minimum. A developer also can set a condition to spin up new EC2 instances to scale back latency.
- **Security**
AWS Elastic Load Balancer supports applications within an [Amazon VPC](#) for stronger network security. An IT team can specify whether it wants an internet-facing or internal load balancer. The latter option enables a developer to route traffic through an ELB using private IP addresses.

AWS Load Balancer Types

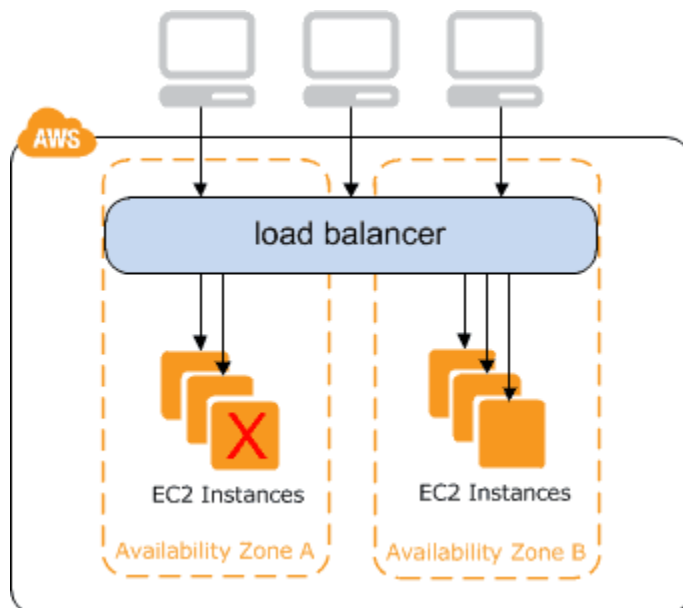
4 Types of load balancer in AWS

- Classic Load Balancer
- Application Load Balancer
- Network Load Balancer
- Gateway Load Balancer



Classic Load Balancer

- The Load Balancer which balances the traffic across multiple instances in multiple availability zones is called a Classic Load Balancer.
- It supports both EC2 Classic EC2-VPC and Increases the availability of your application by sending traffic to healthy Instance.
- Supports HTTP, HTTPS, TCP, and SSL listeners and supports sticky sessions using application-generated cookies.
- To make ensure that your registered instances are able to handle the request load In each AZ, keep approximately the same number of instances In each AZ registered with the load balancer.

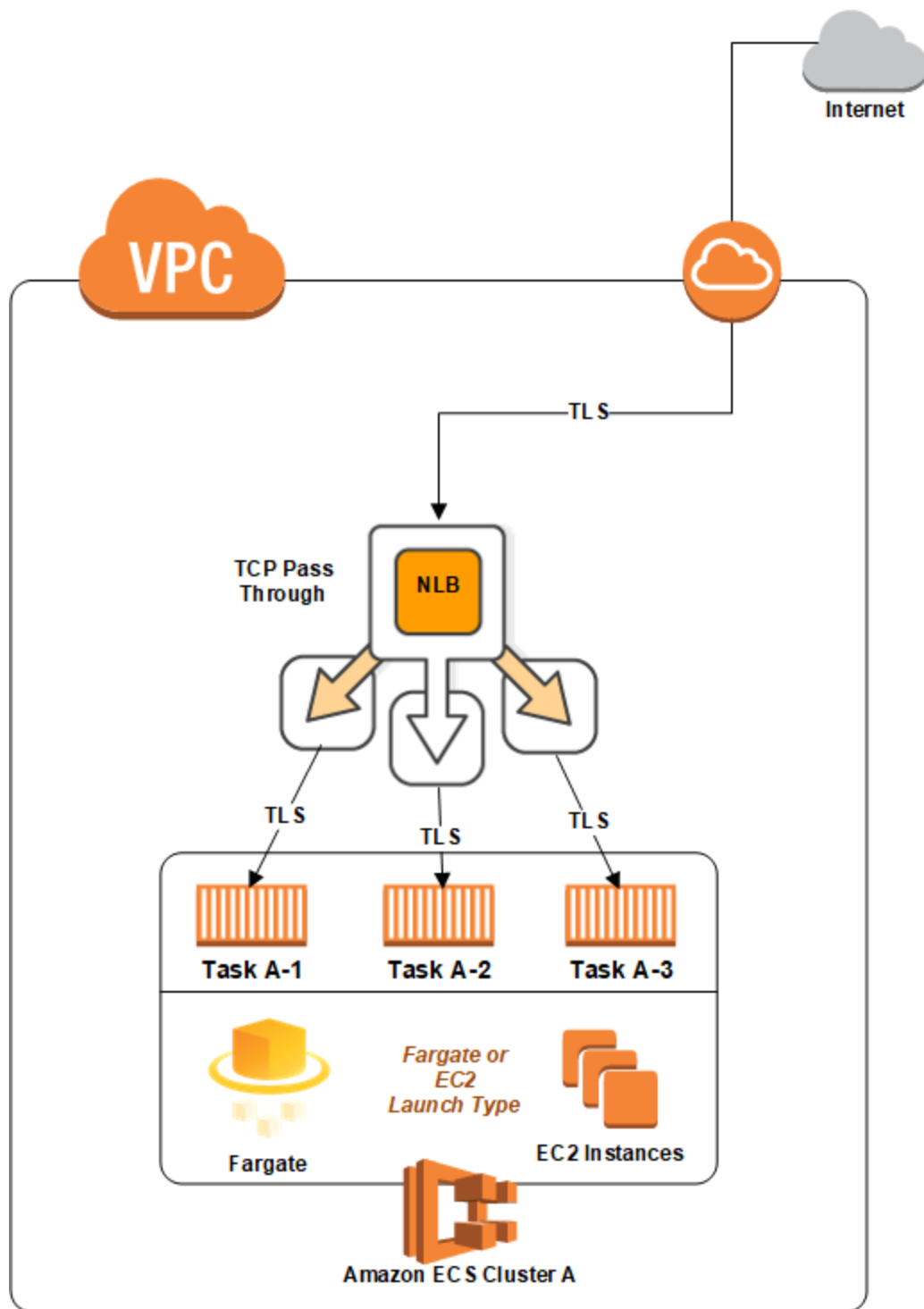


Limitations:

Regional Limit	
LB per region	20
Load Balancer Components Limit	
Listeners	50
Security groups	5
Registered instances	1000
Subnets per Availability Zone	1

Network Load balancer

- Network Load Balancer handles sudden and violates traffic across the EC2 Instances in order to avoid any latency.
- Connection baseload Balancing and it supports TCP protocol.
- Ability to handle volatile workloads and scale to millions of requests per second.
- Support for static IP addresses for the load balancer. or assign one Elastic IP address per subnet enabled for the load balancer.
- Cross-zone load balancing is disabled by default
- If you specify targets using an instance ID, the source IP addresses of the clients are preserved and provided to your applications. If you specify targets by IP address, the source IP addresses are the private IP addresses of the load balancer nodes.
- Network Load Balancers support connections from clients over inter-region VPC peering, AWS managed VPN, and third-party VPN solutions.

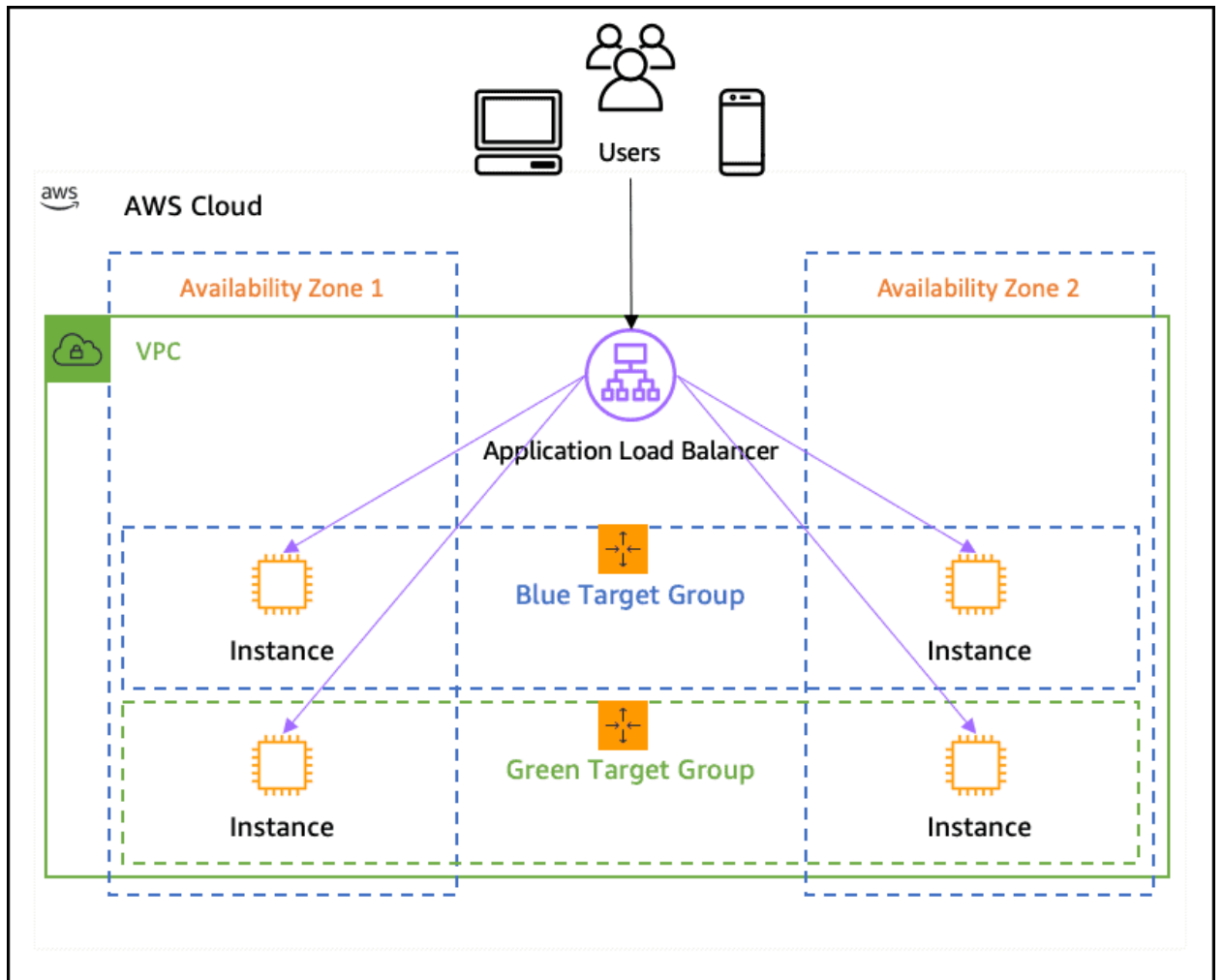


Limitations:

Regional Limit per Region	
Number of Network LB	20
Target groups	3000
Components Limit per LB	
Listeners	50
Targets per Availability Zone With Cross-zone load balancing disabled	200
Targets With Cross-zone load balancing enabled	200
Subnets per Availability Zone	1

Application Load balancer

- The Load Balancer that distributes the traffic to appropriate target groups on the basis of content is called Application Load Balancer.
- New feature-rich, layer 7 load balancing platform.
- Supports web sockets, HTTP, HTTPS, and microservices and container-based applications, including deep integration with EC2 container service.
- Support for path-based and host-based routing. Also, provide routing requests to multiple applications on a single EC2 instance.
- Cross-Zone load balancing is always enabled and you can also specify Lambda functions are targeted to serve HTTP(S) requests.
- Supports load balancer-generated cookies only for sticky sessions.

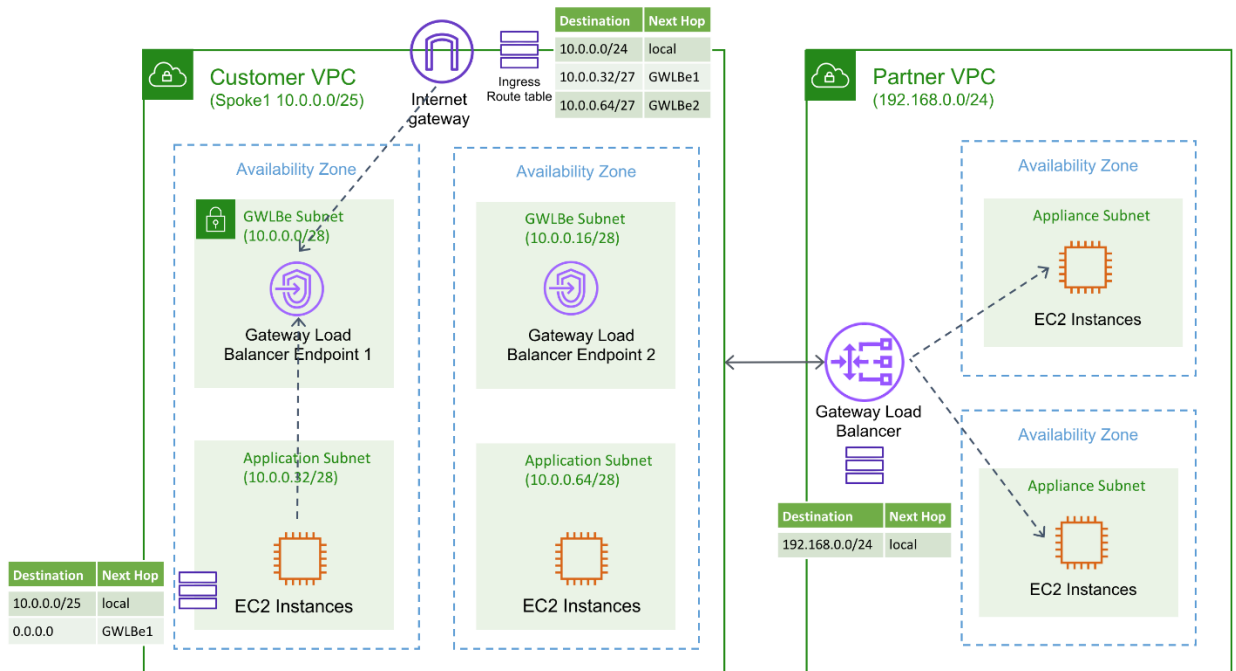


Limitations:

Regional Limit	
LB Per Region	20
Target groups/Region	3000
Load Balancer Components Limit	
Listeners/load balancer	50
Targets /load balancer	1000
Subnets /AvailabilityZone per load balancer	1
Security groups / load balancer	5
Rules(not counting default rules)	100
Certificates (not counting default certificates)	25
Number of times a target can be registered	100

Gateway Load Balancer

- It makes it easy to deploy, scale, and manage your third-party virtual appliances.
- Provide you one gateway for distributing traffic across multiple virtual appliances, while scaling them up, or down, based on demand.
- It eliminates potential points of failure in your network and increases availability.
- Users can find, test, and buy virtual appliances from third-party vendors directly in AWS Marketplace.
- This integrated experience streamlines the deployment process, so users can see value from your virtual appliances more quickly—whether you want to work with the same vendors you do today, or trying something new.



Limitations:

Regional Limit

LB Per Region	20
---------------	----

Load Balancer Components Limit

Gateway Load Balancers per VPC	10
Target groups with GENEVE protocol	100
Targets per Availability Zone per target group with GENEVE protocol	300