# 10. What is an Elastic Load Balancer (ELB)?

Distributes inbound requests evenly across the target resource group. The group could be a fleet of EC2 instances, a range of IP addresses or Containers.

Targets can be defined across different Azs or all placed in one AZ.

ELB Example

You have created a new app and it is running on a single EC2 instance. It is being accessed by a number of users. Your architecture might initially look like this:
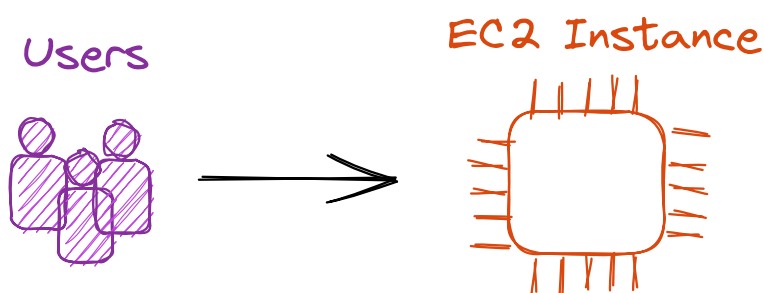

This means that if the single instance fails, your application will be down. The instance may also be unable to handlethe increased load of spikes in traffic.

Adding an ELB to distribute traffic across multiple instances avoids these problems:
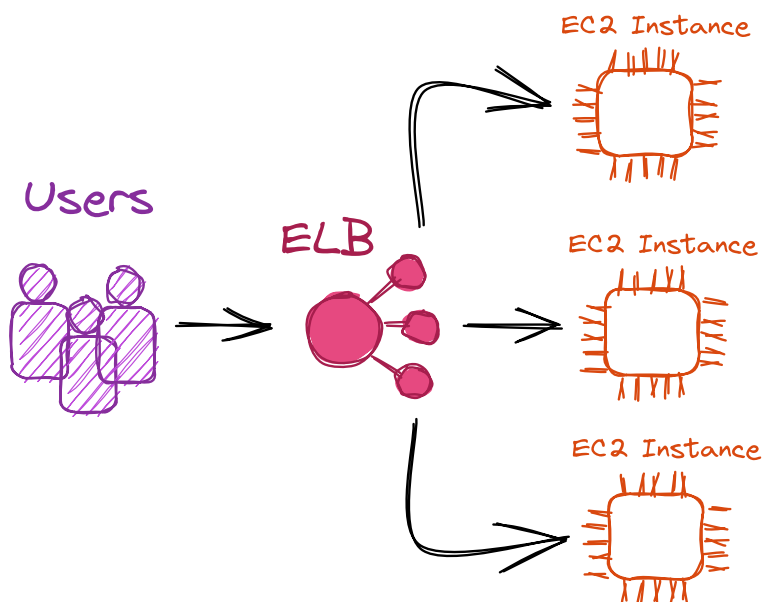
ELB is highly available because it is a managed service.

If any instances fail, the ELB will automatically detect and divert. The additional instances help with the load where there is a surge in traffic.
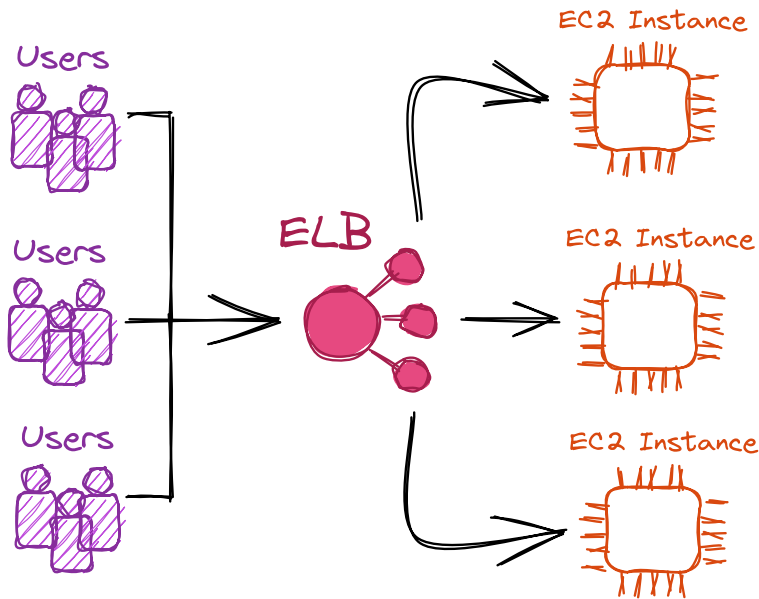
The ELB dynamically scales up and down as needed.

There are three ELBs to choose from:

Application Load Balancer	Network Load Balancer	Classic Load Balancer
Flexible features for HTTP or HTTPS web apps. Operates at the request level. Advanced routing, TLS termination and visibiility features.	Ultra high-performance, maintaining low latencies. Routing operates at the connection level. Handles millions of requests per second.	For apps built in classic EC2 environment, operating at connection and request level.

ELB Components

Listeners

Every ELB needs at least one listener to be configured. It defines how your requests are routed to target groups based on your set **conditions**
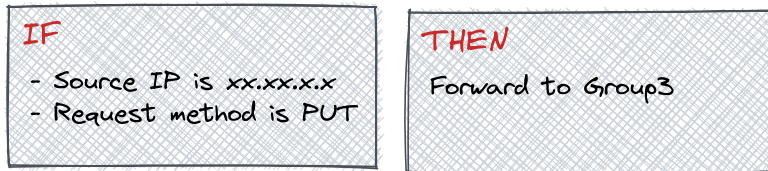
Target Groups

A group of resources to route requests to. ELB can be configured with multiple target groups, each with different listener configurations and **rules**.

Rules

Associated with each listener. Define how a request is routed to target group.

All **conditions** in a rule lead to a single action:

```
IF

- Source IP is xx.xx.x.x
- Request method is PUT
```

```
THEN

Forward to Group3
```

Health Checks

Performed against the target group resources. ELB contacts each target using a specific protocol to receive a response. If no response is received the target is marked as unhealthy and temporarily decommissioned.

Internet-Facing ELB

Nodes of ELB are accessed via the internet via a public DNS name.

Internal ELB

Only serves requests originating from within your VPC. You may have an internal ELB sitting between your web servers on the public subnet and your application servers on the private subnet.

ELB Nodes

An ELB Node is placed within each selected Availability Zone. Without this Node, traffic will not be routed to that AZ. This is because the Nodes are used by the ELB to distribute traffic to your target groups.

Cross-Zone Load Balancing

Allows traffic to be distributed evenly between all targets across multiple Availbility Zones, including AZs with their own associated ELBs. If turned off, traffic will be distributed in only the AZ associated with that load balancer.

Availability Zone A

Availability Zone B

10 10 10 10 10 10 10 10 10 10