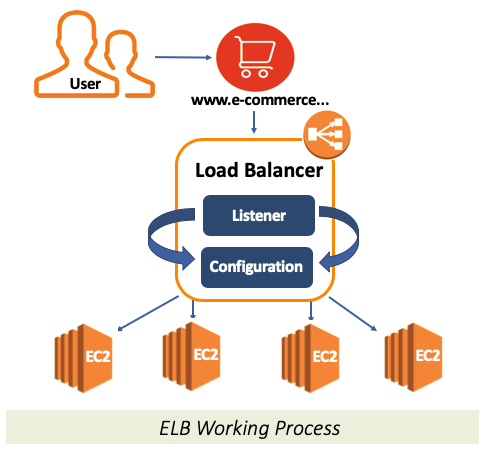
Elastic Load Balancing (ELB)

ELB ( Elastic Load Balancing), as the name suggests, is a service allowing to make the load distribution for your existing resources (instances) in AWS. It automatically distributes incoming application traffic across multiple targets, such as Amazon EC2 instances, containers.

We have set up 1 server and started our web site. Let's assume that this is an e-commerce website and it should be running 24/7.

But, if our virtual machine fails for any reason, it will probably take 30 minutes to make it work again because of installation, adjustment, etc.. Therefore, you will lose customers and money.

So, there are many shortcomings in setting up a single machine so, we need to construct the infrastructure over multiple servers. And this time you need to direct the traffic between these servers. This is what we call the **Load Balancer** that tackle this job for you.



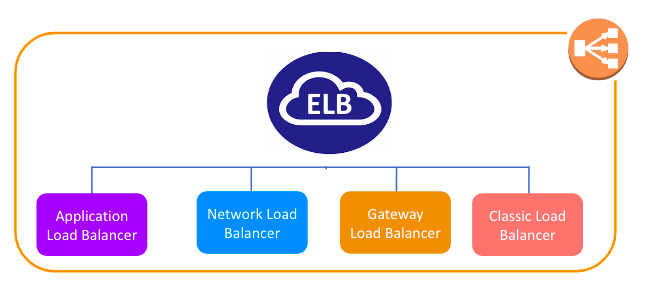
Load Balancer basically consists of 2 components.

* **Listener**
* **Configuration**

**Listener**, as the name suggests, listens to the incoming traffic through a specific port according to the component adjustments.

**Configuration** is a set of rules that takes into account the Listener's analysis and provides to direct traffic.

### **Types of ELB**



### **Application Load Balancer**

It is operated at Layer 7 of the OSI Model

It identifies the incoming traffic and forwards it to the right resources. You can route the request according to the information here by reading the **contents of the package**. In other words, it directs traffic according to the contents of the package. For example, if a URL has /API extensions, then it is routed to the appropriate application resources. t is best suited for the load balancing of **HTTP and HTTPS** traffic.

Like Network Load Balancer, you can use the Application Load Balancer compatible with **Auto Scaling** function of AWS.

### **Network Load Balancer**

It is operated at Layer 4 of the OSI model.

It makes routing at the transport layer (TCP/SSL), and it can handle millions of requests per second. So It is best suited for load balancing the **TCP traffic** when **high performance is required**.

If we need a **simple and fast** load balancer over a basic TCP-based port, this is the Network Load Balancer.

### **Gateway Load Balancer**

Gateway Load Balancers enable you to **deploy, scale, and manage virtual appliances,** such as **firewalls, intrusion detection and prevention systems**, and deep packet inspection systems.

A Gateway Load Balancer operates at the **third layer** of the Open Systems Interconnection (OSI) model, **the network layer**.

It listens for all IP packets across all ports and forwards traffic to the target group that's specified in the listener rule.

It maintains stickiness of flows to a specific target appliance using 5-tuple (for TCP/UDP flows) or 3-tuple (for non-TCP/UDP flows). The Gateway Load Balancer and its registered virtual appliance instances exchange application traffic using the GENEVE protocol on port 6081.

### **Classic Load Balancer**

It has been serving at Layer 4 of the **OSI** model since 2009 and the oldest type of ELB.

Classic Load Balancer provides basic load balancing across multiple Amazon EC2 instances and operates at both the request level and connection level.

It can also be used for load balancing the HTTP or HTTPs traffic and use layer 7-specific features.

AWS doesn't recommend you to use Classic Load Balancer anymore.