

# ELB & AS

2019年2月10日 星期日

上午1:46

## What is Elastic Load Balancing?

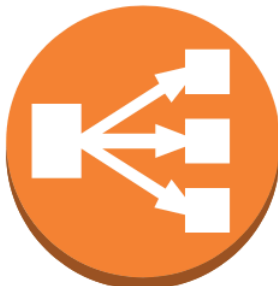
### Simplified Definition:

An ELB evenly distributes traffic between EC2 instances that are associated with it.

physically

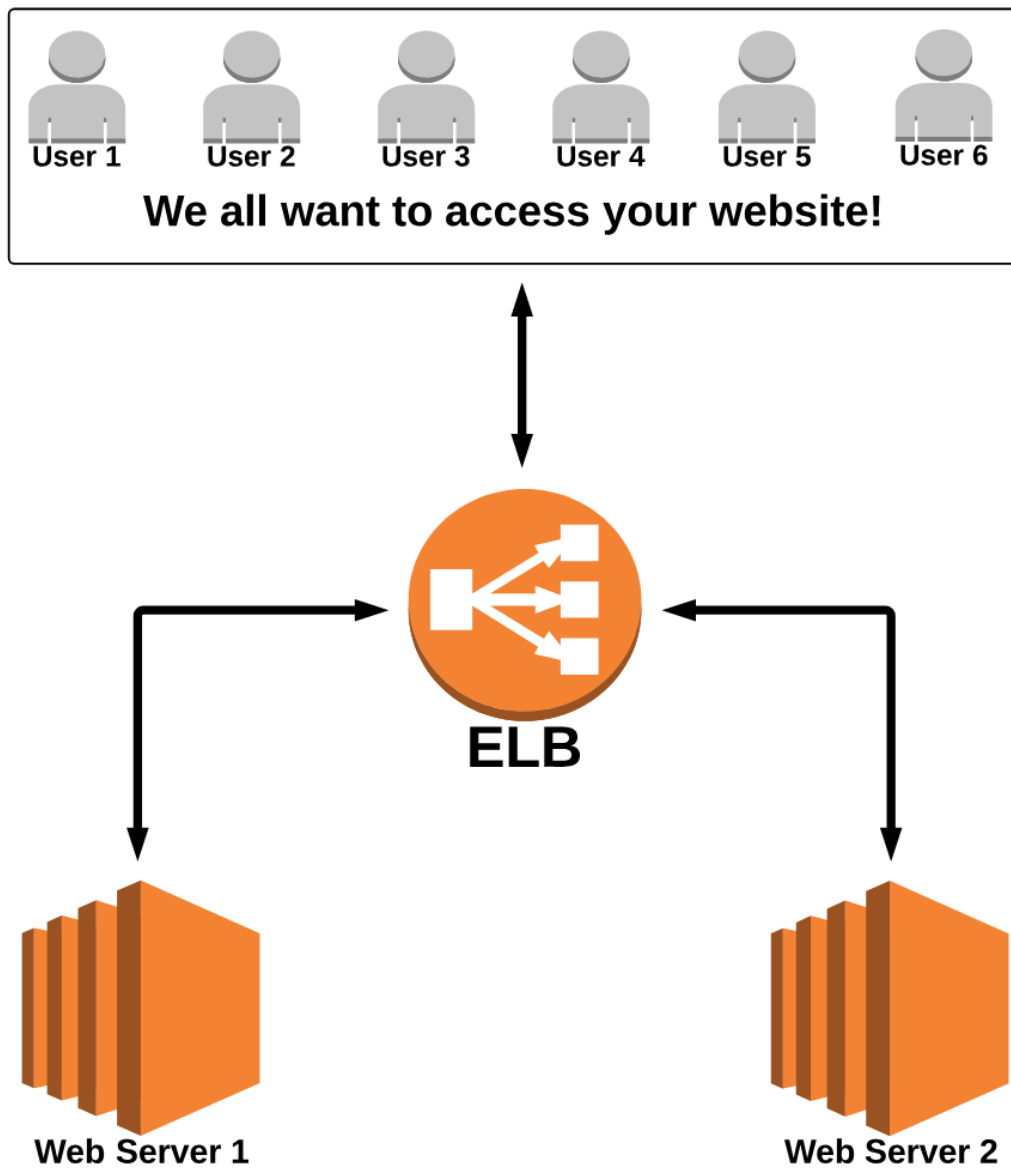
### AWS Definition:

"A load balancer **distributes incoming application traffic across multiple EC2 instances in multiple Availability Zones**. This increases the **fault tolerance** of your applications. Elastic Load Balancing **detects unhealthy instances and routes traffic only to healthy instances**."



ELB

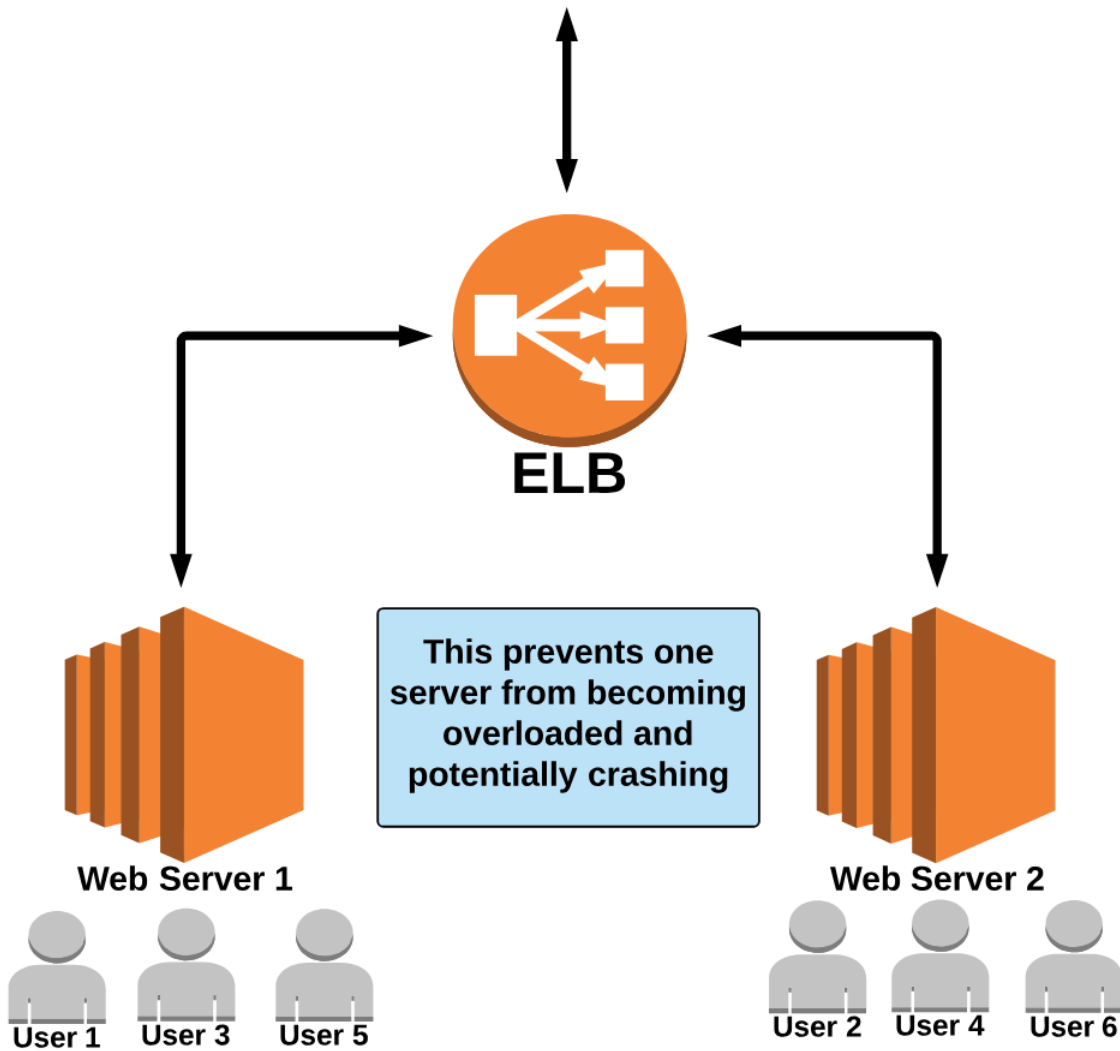
## ELB Basics.



## ELB Basics:



**We all want to access your website!**

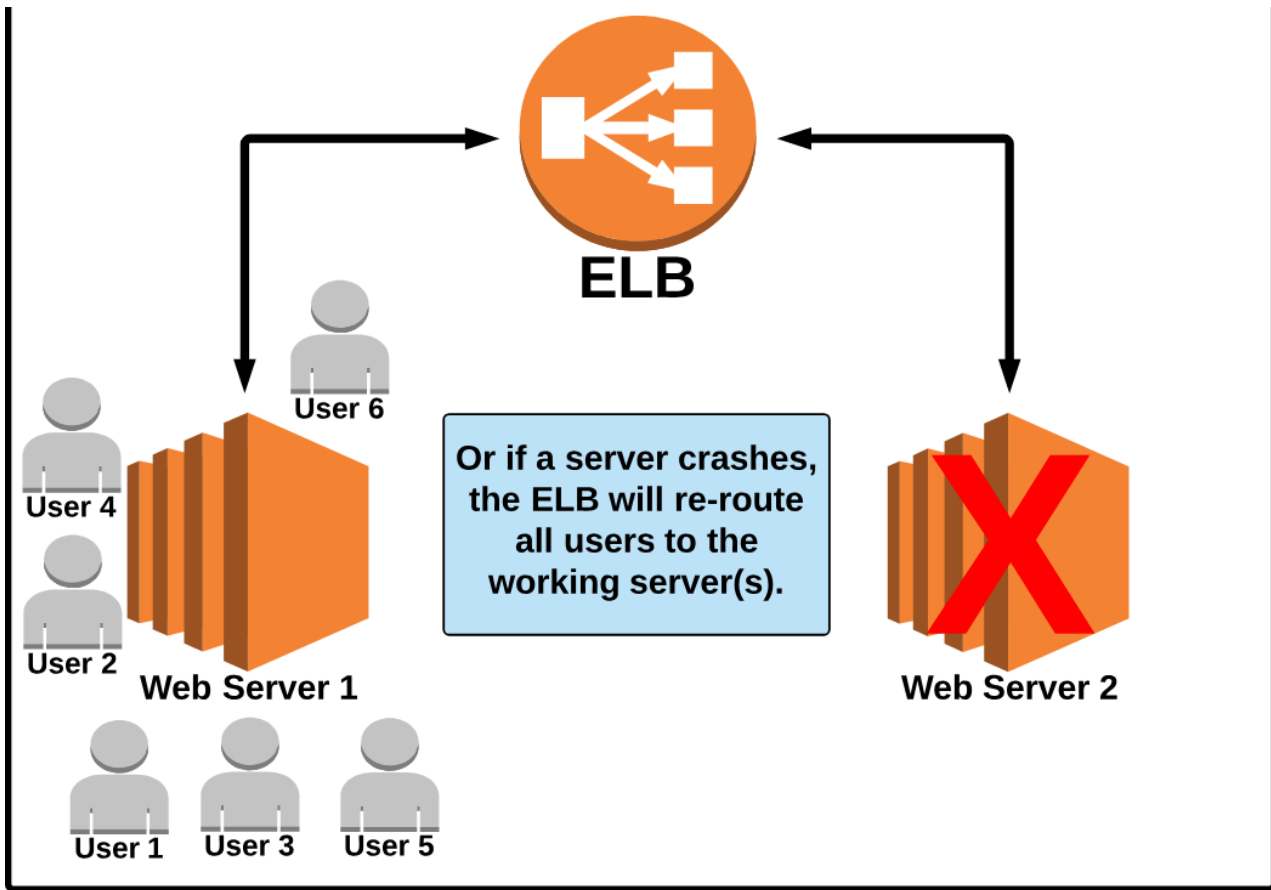


## **ELB Basics:**

**We all want to access your website!**







## ELB Basics:

### ***Elastic Load Balancing***

is a foundational component of

***High Availability***

and



# ***Fault Tolerance***

***We now know that Elastic Load Balancing can evenly distribute traffic between all active servers - but what happens if demand (traffic) is so high that the active services can't handle it?***

## **What is Auto Scaling?**

### **Simplified Definition:**

Auto Scaling automates the process of adding (***scaling up***) OR removing (***scaling down***) EC2 instances ***based on traffic demand*** for your application.

### **AWS Definition:**

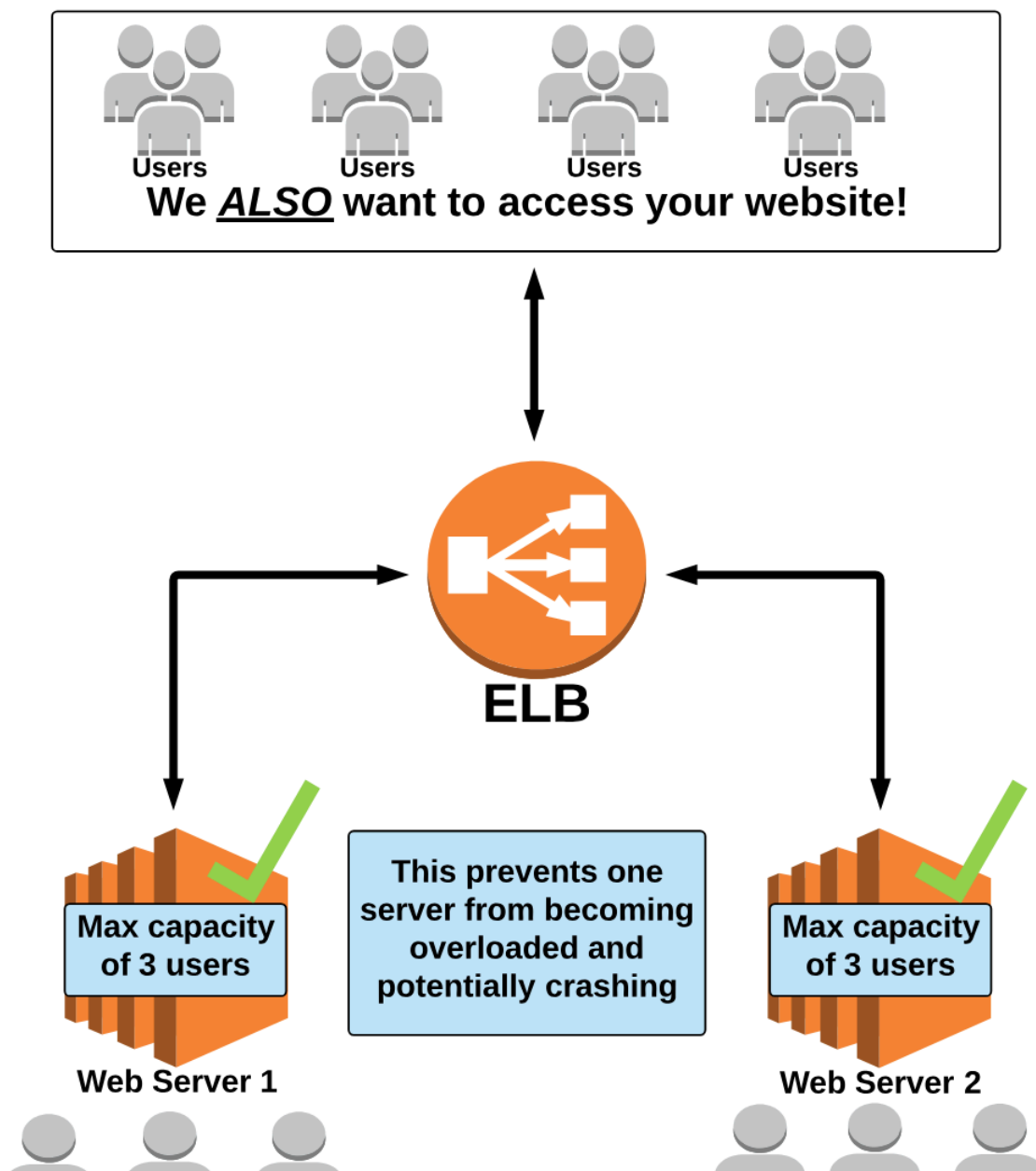
"Auto Scaling helps you ensure that you have the correct number of Amazon EC2 instances available to ***handle the load for your application***. You create collections of EC2 instances, called ***Auto Scaling groups***. You can specify the minimum number of instances in each Auto Scaling group, and Auto Scaling ensures that your group never goes below this size. You can specify the maximum number of instances in each Auto Scaling group, and Auto Scaling ensures that your group never goes above this size. If you specify the desired capacity, either when you create the group or at any time thereafter, Auto Scaling ensures that your group has this many instances. If you specify scaling policies, then Auto Scaling can launch or terminate instances as demand on your application increases or decreases."





# Auto Scaling

## Auto Scaling Basics:





User 1

User 3

User 5

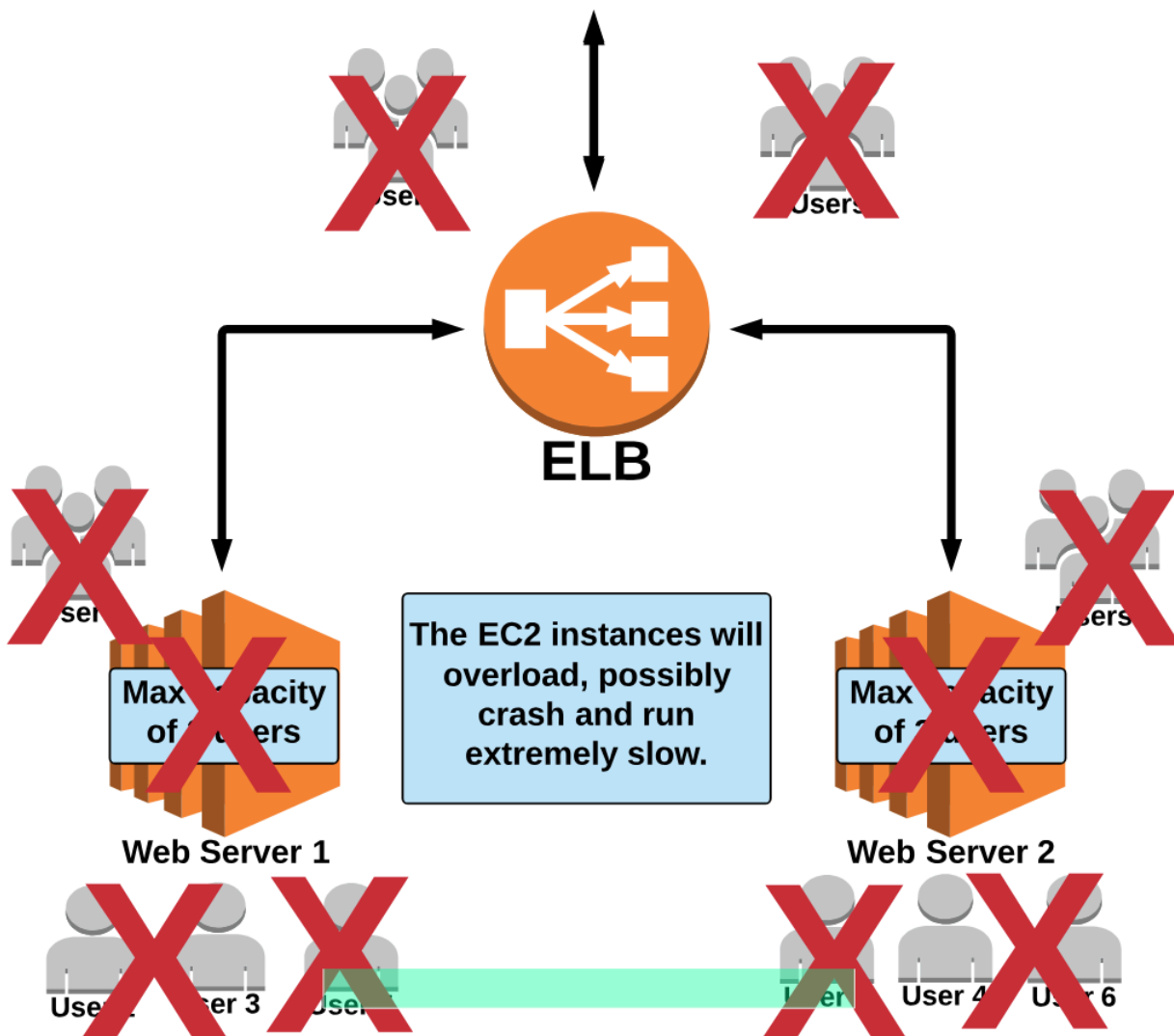
User 2

User 4

User 6

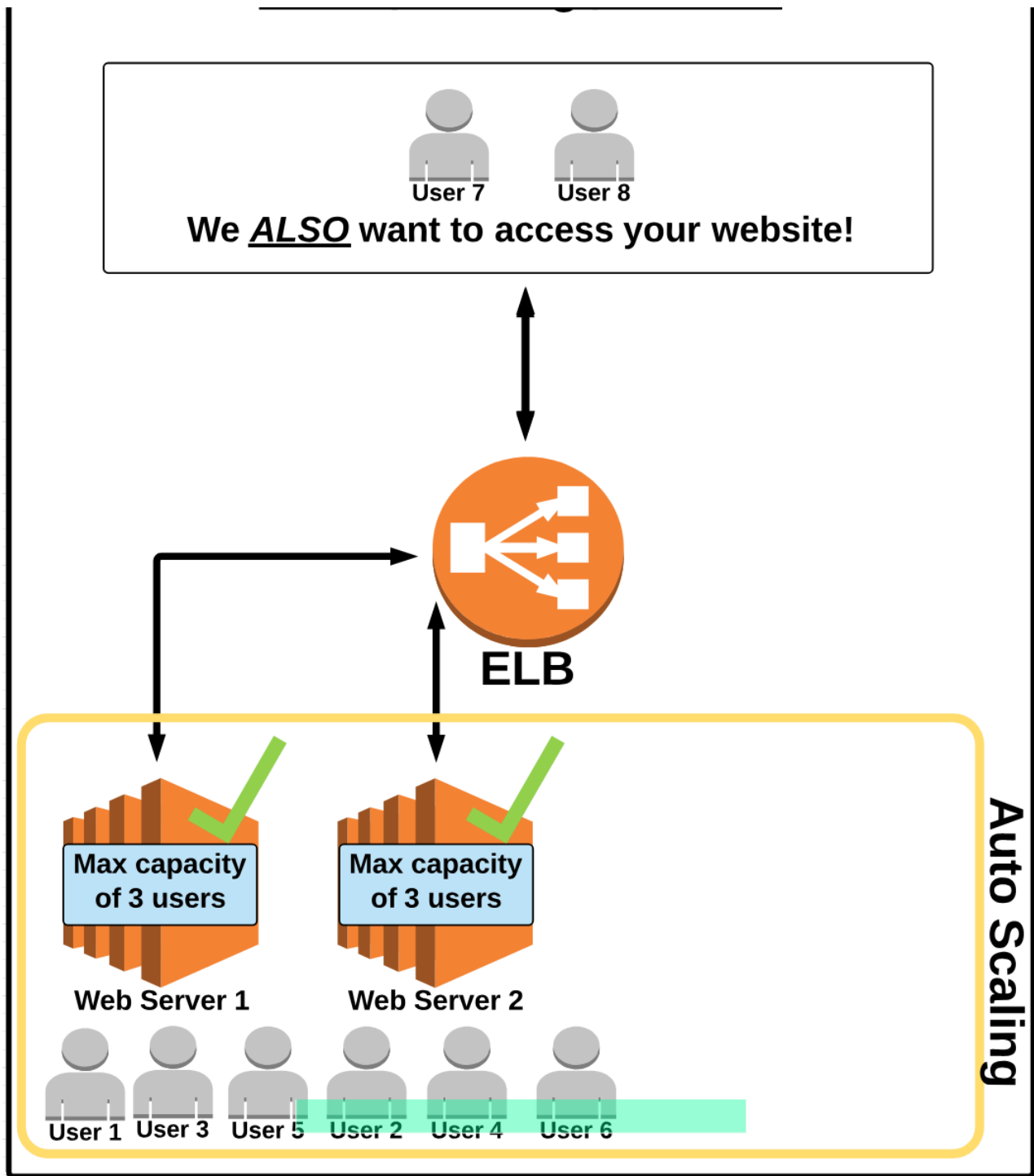
## Auto Scaling Basics:

We all want to access your website!



## Auto Scaling Basics:





## Auto Scaling Basics:



We ALSO want to access your website!

Auto Scaling will automatically add additional servers, based on demand.

ELB

Max capacity  
of 3 users

Web Server 1

Max capacity  
of 3 users

Web Server 2

Max capacity  
of 3 users

Web Server 3



User 1

User 3

User 5

User 2

User 4

User 6

User 7

User 8

Auto Scaling

## Auto Scaling Basics:



User 5

User 6

User 7

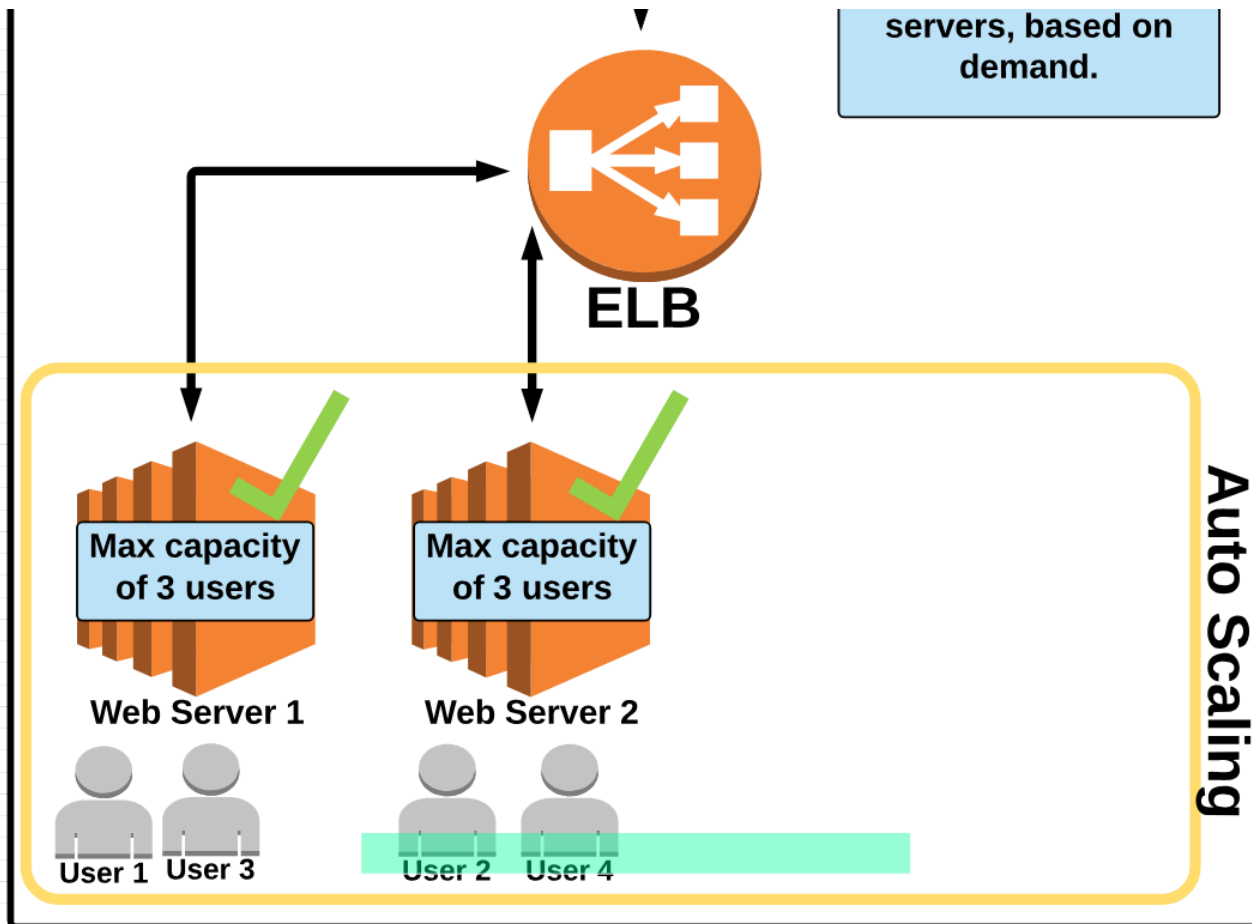
User 8

Thank you - we enjoyed your website!

Auto Scaling will automatically remove







## Auto Scaling Basics:

### *Auto Scaling*

builds on the benefits of

### *Elastic Load Balancing*

while adding the benefits of



# ***Scalability***

and

# ***Elasticity***

## Introduction to ELB

- ▶ Elastic Load Balancer allows us to distribute the incoming traffic to multiple instances similar to what a traditional load balancer does.
- ▶ ELB is capable of handling rapid change in the network traffic patterns.
- ▶ Since it's managed service, client does not have to worry about the high availability related aspects.

