EC2 Auto Scaling

CLARUSWAY
WAY TO REINVENT YOURSELF

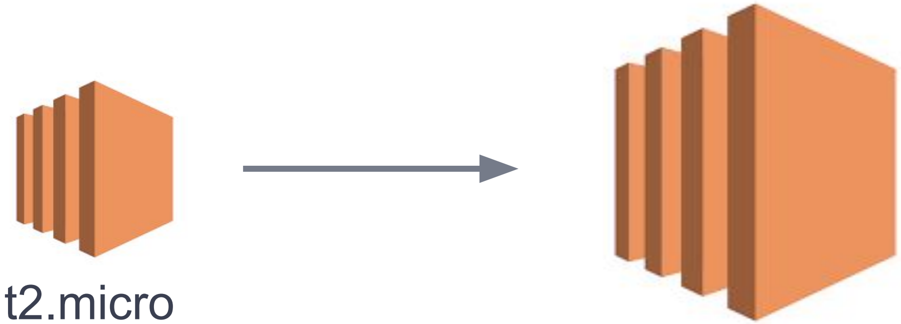
Today’s Takeaways

- Part 1: Anatomy of Auto Scaling
- Part 2: Hands on

Auto Scaling

What is Scaling?

Vertical Scaling

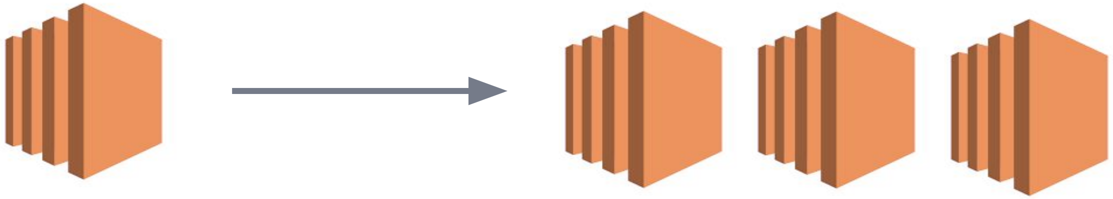


t2.micro

m4.large

Scale Up / Down

Horizontal Scaling



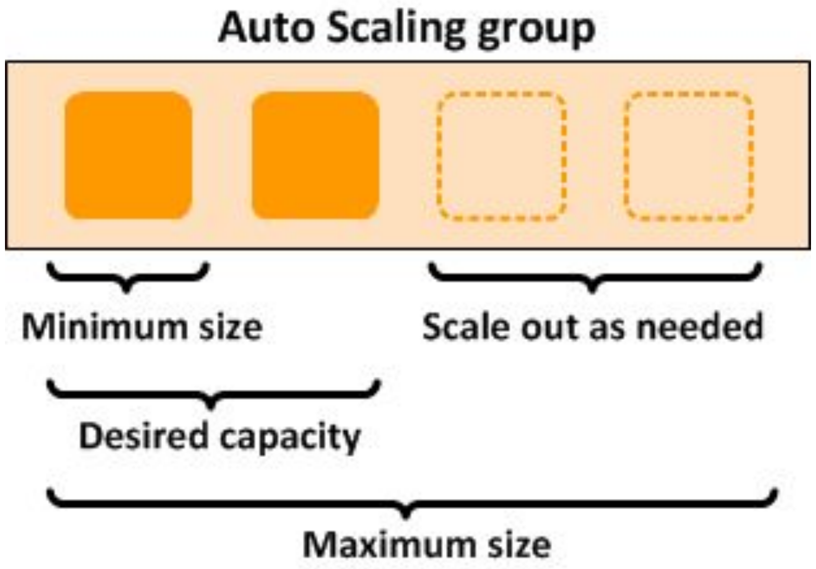
t2.micro

3 x t2.micro

Scale Out / In

Auto Scaling

What is Auto Scaling?



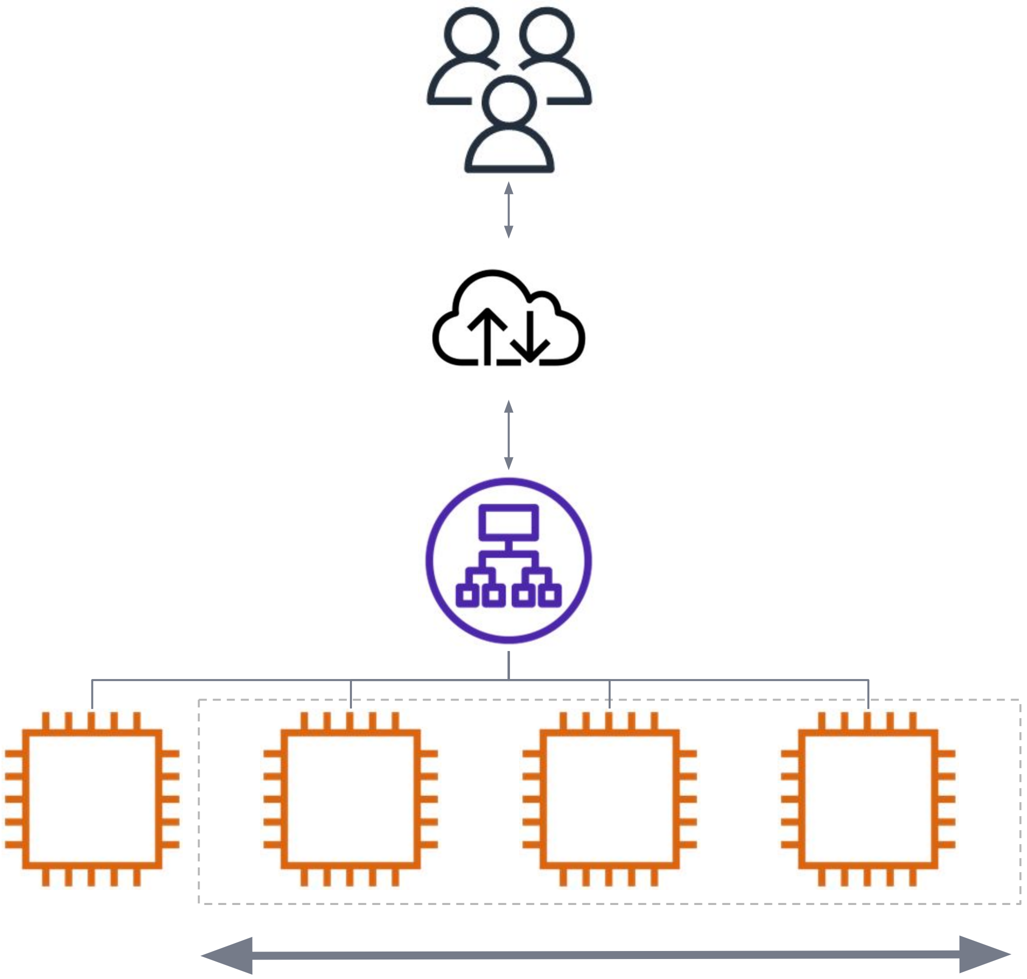
- Amazon EC2 Auto Scaling is a component that helps you ensure that you have the correct number of Amazon EC2 instances available to handle the load for your application.
- Auto Scaling adds or removes instances to keep your system steady state.
- You can automate the increasing or decreasing of virtual machines depending on your policy.

Auto Scaling

- A **load balancer** distributes traffic between instances

AWS Auto Scaling:

- allows you to **automatically increase or decrease** the number of instances behind the load balancer
- **monitors the health** of instances and **replaces impaired instances**
- provides an **efficient & cost-effective** way to meet **capacity requirements**



Scale Out (add) or **Scale In** (remove) to meet capacity needs

CLARUSWAY
WAY TO REINVENT YOURSELF

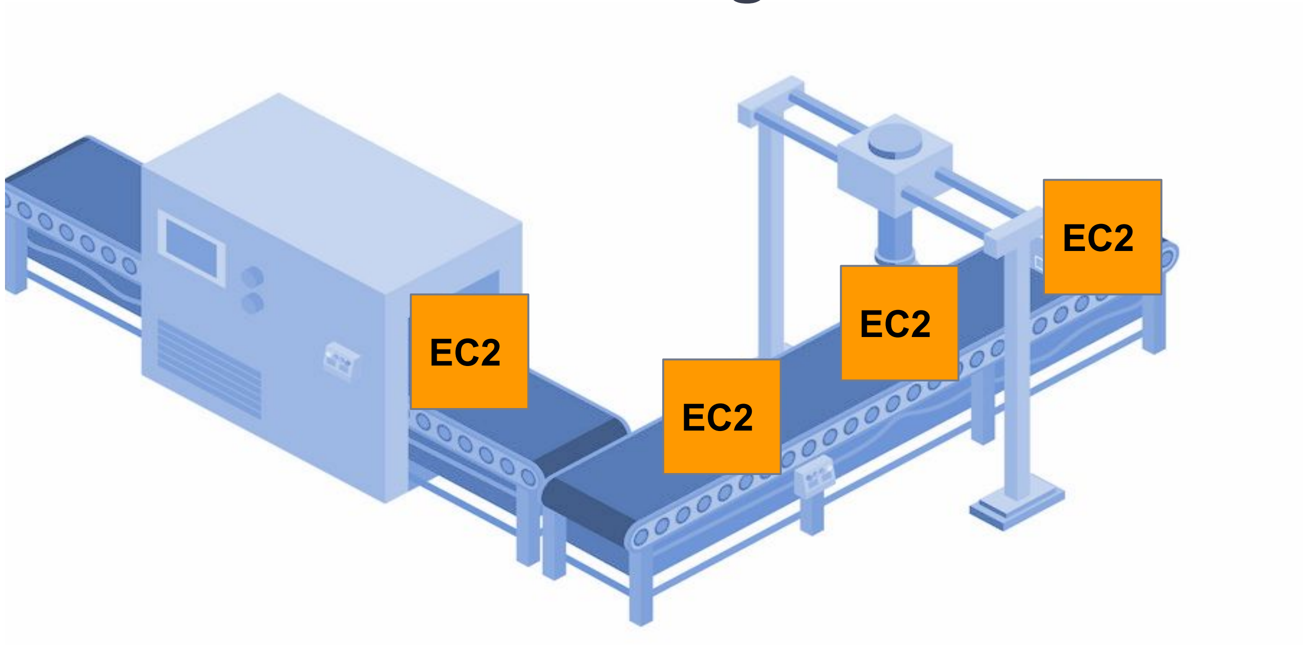
Auto Scaling

Auto Scaling vs Load Balancer

Load Balancer

Distributing the traffic across the **existing** instances

Auto Scaling



- **Launch** new instances
- **Terminate** the existing instances

Auto Scaling

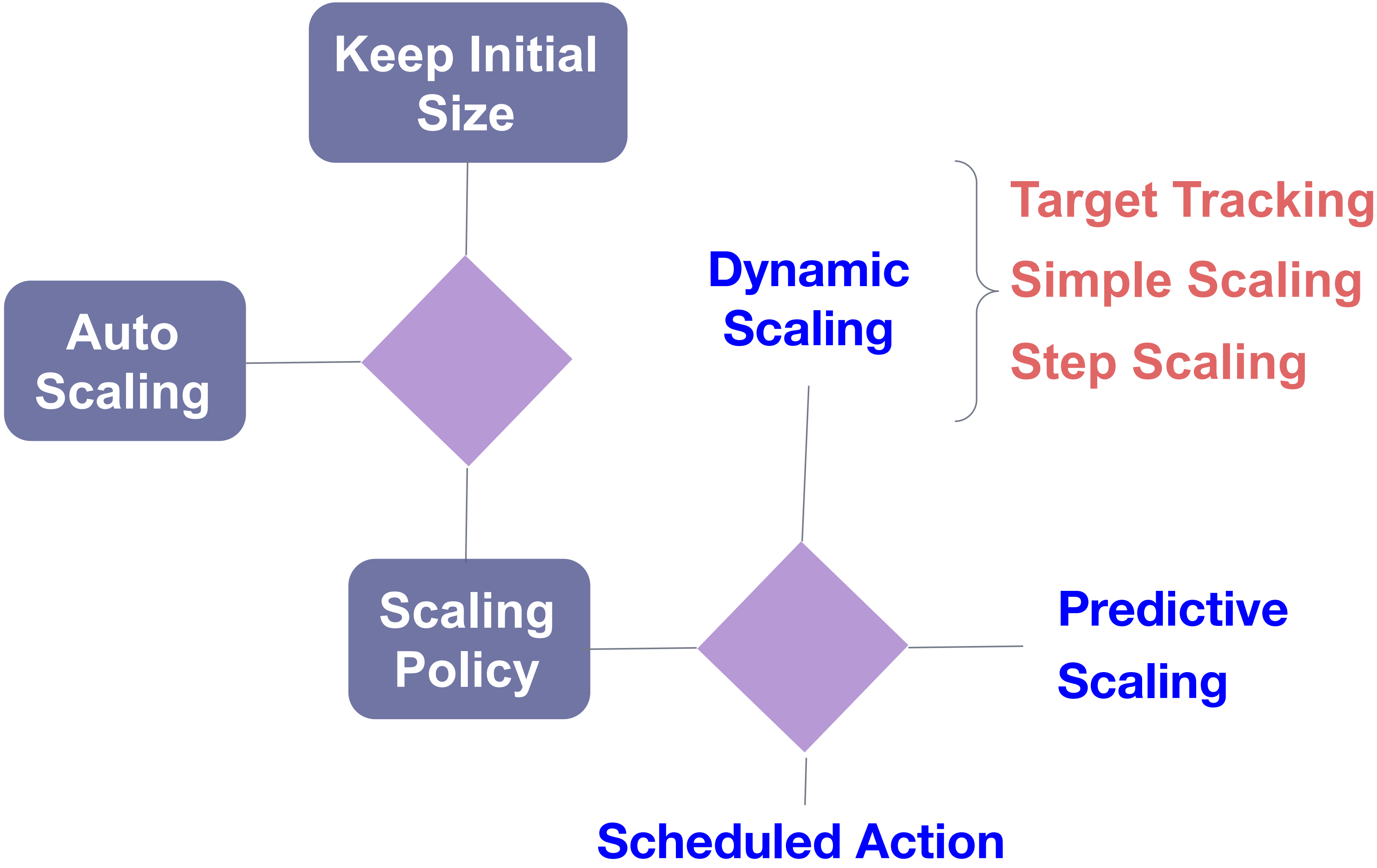
Features of Auto Scaling

SAMPLE

- Auto Scaling Policy
- Launch Templates (or previously Launch Configuration)
- Fault Tolerance
- High Availability
- Compatible with Elastic Load Balancer
- Better Cost Management.

Auto Scaling

Auto Scaling Options



Keep Initial Size

Auto Scaling

Scaling Policy

Dynamic Scaling

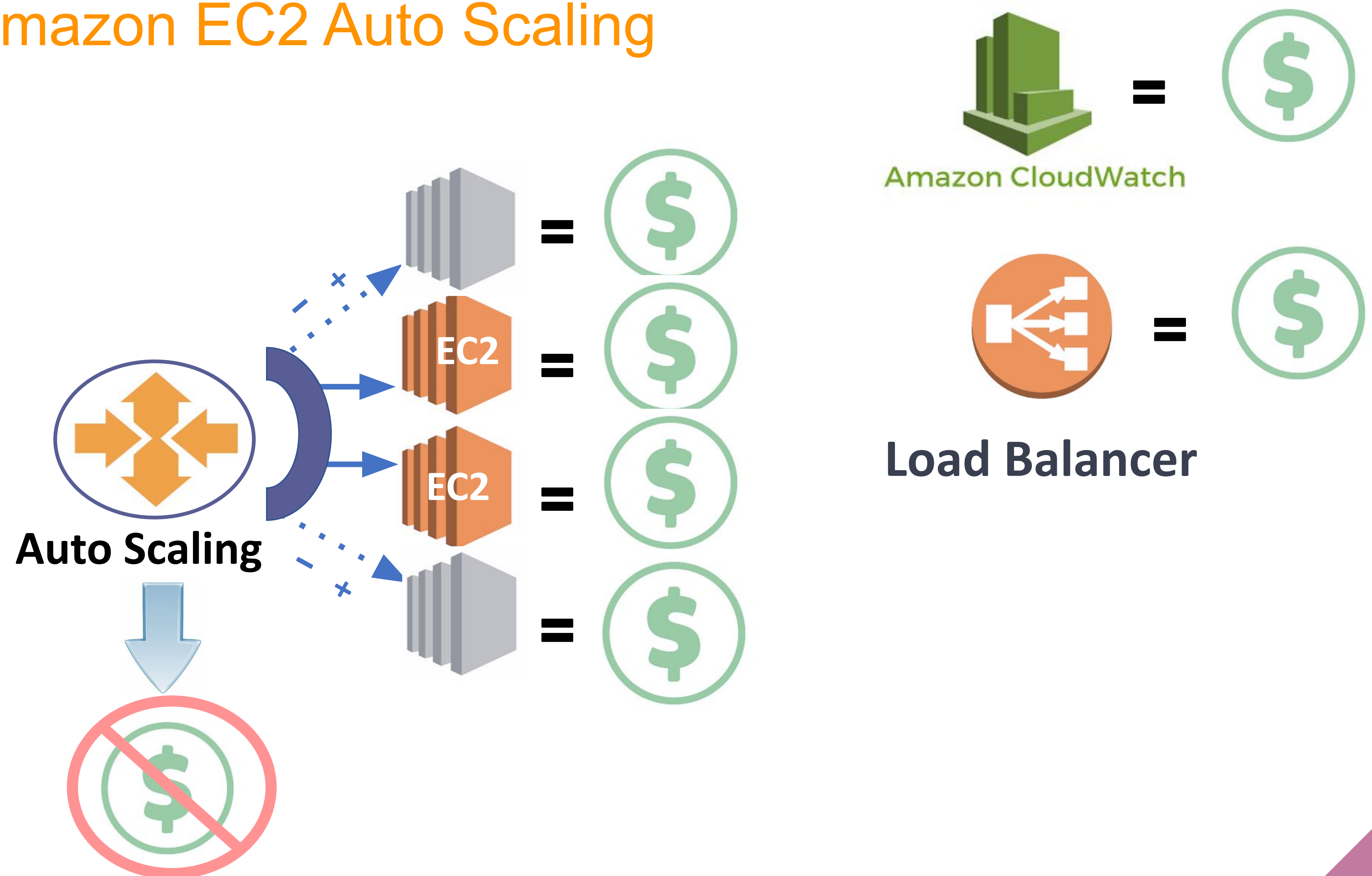
- Target Tracking
- Simple Scaling
- Step Scaling

Predictive Scaling

Scheduled Action

Auto Scaling

Pricing for Amazon EC2 Auto Scaling



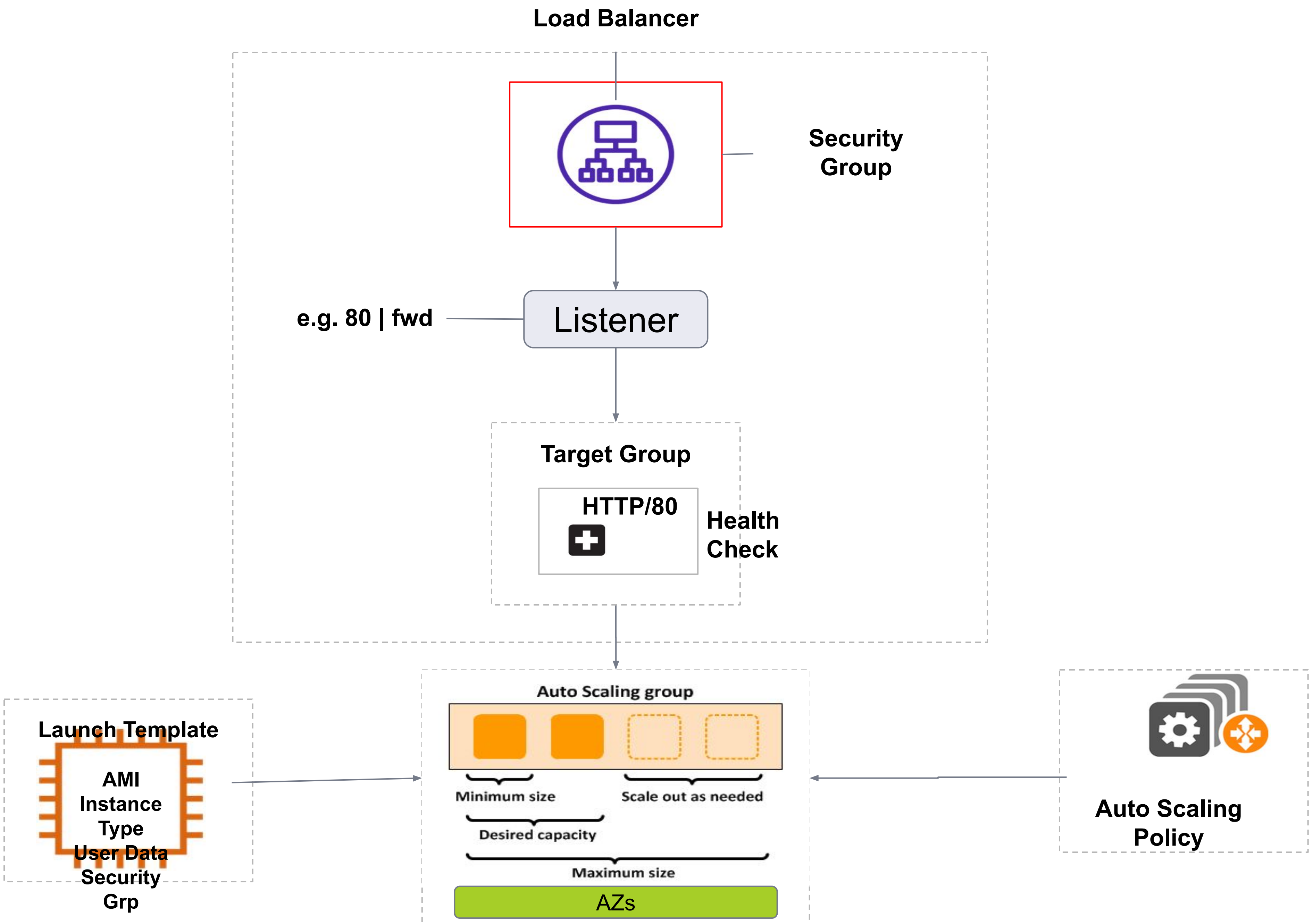
Lifecycle Auto Scaling

Let's get our hands dirty!

- Creating an Auto Scaling Group

THANKS!

Any questions?


Load Balancer
Security Group
e.g. 80 | fwd
Listener
Target Group
HTTP/80
Health Check
Launch Template
AMI
Instance Type
User Data
Security Grp
Auto Scaling group
Minimum size
Scale out as needed
Desired capacity
Maximum size
AZs
Auto Scaling Policy