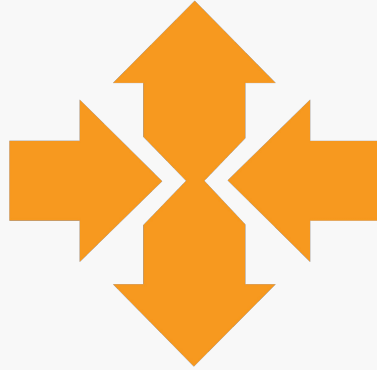


Auto Scaling



Auto Scaling

- Auto Scaling helps you ensure that you have the correct number of Amazon EC2 instances available to handle the load for your application.
- Automatically increase the number of Amazon EC2 instances during demand spikes to maintain performance and decrease capacity during lulls
- Auto Scaling is well suited both to applications that have stable demand patterns or that experience hourly, daily, or weekly variability in usage
- Reduce costs
- Monitors health of the instances and terminates the and replaces with new one
- Auto Scaling automatically balances EC2 instances across zones when multiple zones are configured

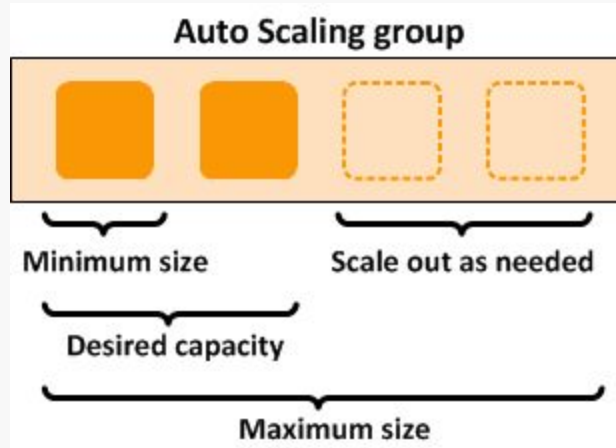


Auto Scaling Features and Benefits

- Reduce costs
- Always ensures we have desired capacity to handle traffic demands
- Monitors health of the instances and replaces unhealthy instances and replaces with healthy instances.
- Auto Scaling automatically launches EC2 instances evenly across zones when multiple zones are configured



Auto Scaling Group



Auto Scaling Components

Groups	<p>Your EC2 instances are organized into <i>groups</i> so that they can be treated as a logical unit for the purposes of scaling and management.</p> <p>When you create a group, you can specify its minimum, maximum, and, desired number of EC2 instances</p>
Launch configurations	<p>Your group uses a launch configuration as a template for its EC2 instances.</p> <p>When you create a launch configuration, you can specify information such as the AMI ID, instance type, key pair, security groups, and block device mapping for your instances.</p>
Scaling plans	<p>A scaling plan tells Auto Scaling when and how to scale</p>



Launch Configurations

- A launch configuration is a template that an Auto Scaling group uses to launch EC2 instances
- When you create a launch configuration, you specify information for the instances such as
 - Amazon Machine Image (AMI)
 - Instance type
 - Key pair
 - One or more security groups



Auto Scaling Demo

