**AUTOSCALING**

**Auto-scaling** is a way to automatically scale up or down the number of compute resources that are being allocated to your application based on its needs at any time...

**Snapshot** is a point in time backup of a specific volume. (It’s not a bootable copy)

**AMI** is the backup of entire EC2 instance that might have multiple attached volumes. contains a software configuration



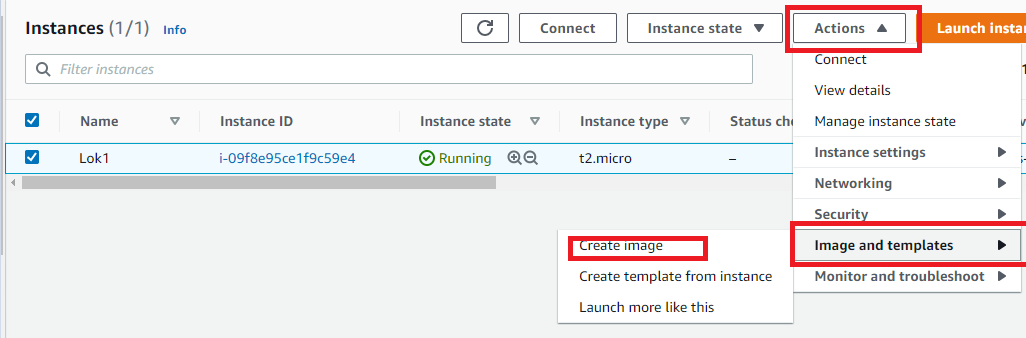
**Why do we need autoscaling:**

Amazon Ec2 auto scaling helps you to ensure correct number of Amazon Ec2 instances are available to handle the load of user traffic.It scales up/down the instances based on the load.

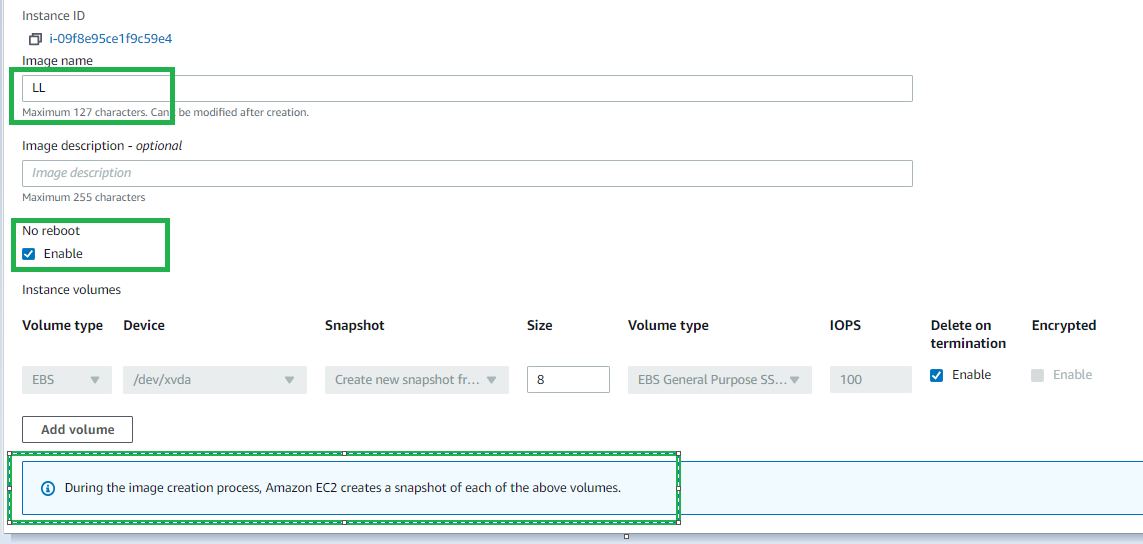
**Note :**: To distribute the traffic equally we need to use load balancer.

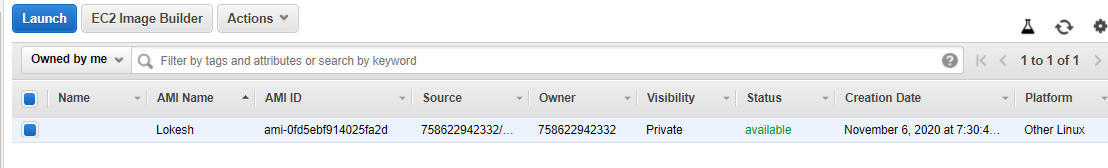
**To create an AMI:**

1. Select an Ec2 🡪actions🡪 Image and template 🡪 create image

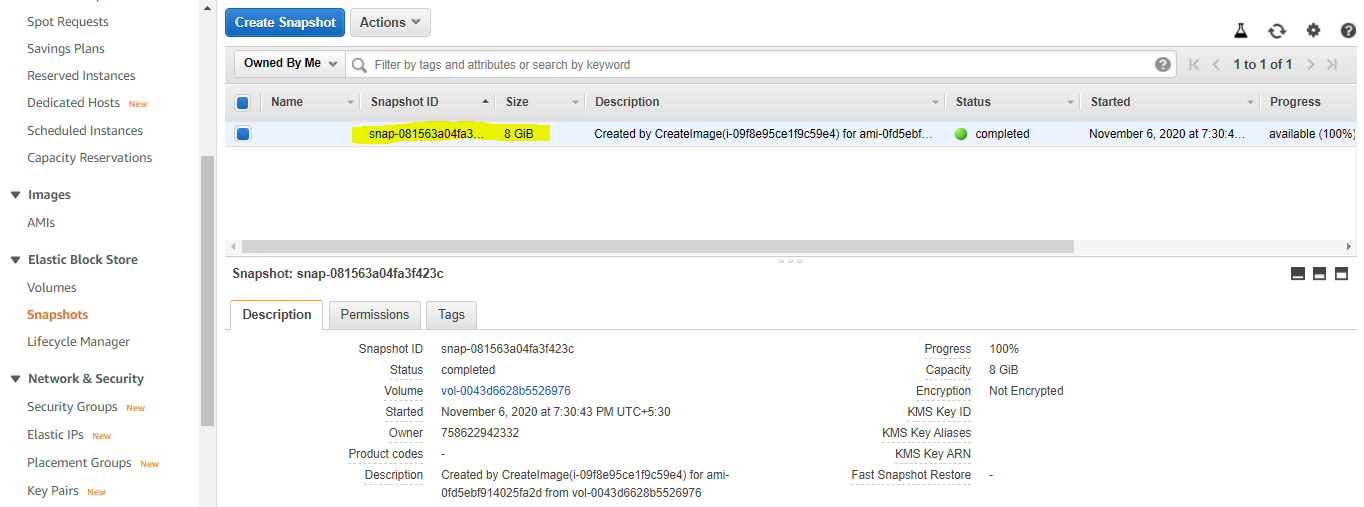


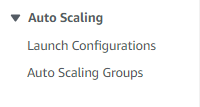
1. Put AMI name and click create image.





**Note** :: When we create an AMI ,it will intern create a snapshot. And it will be of the same size as the instance volume from which we created an AMI





## Launch configuration and Launch template:

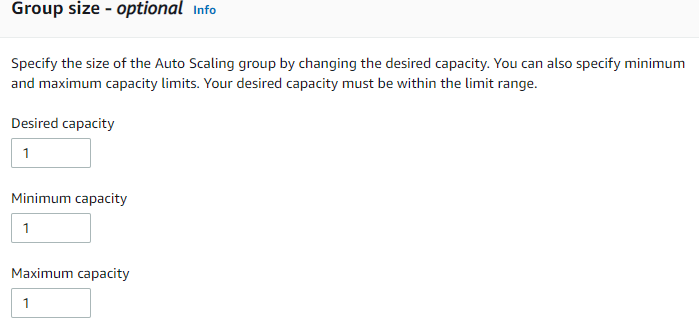
**Launch Template :** Creating a launch template allows you to create a saved instance configuration that can be reused, shared and launched at a later time.( save your settings and continue working on them later,)

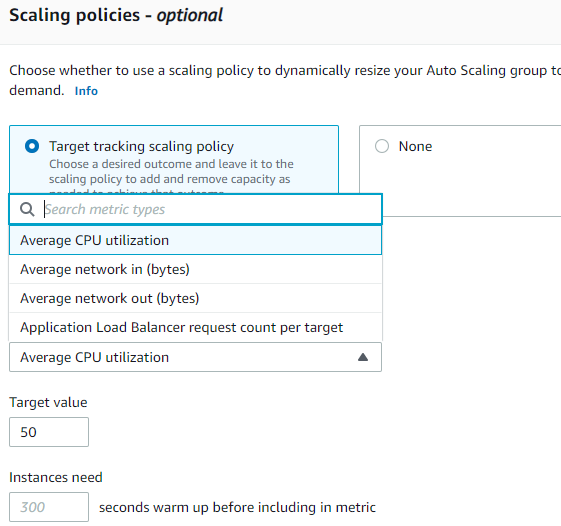
**Launch configuration:**

It contains AMI details,instance type,Purchase option,u can attach IAM role,enabling cloud watch monitoring,we can add additional volumes,,selecting security group,,Key pair…

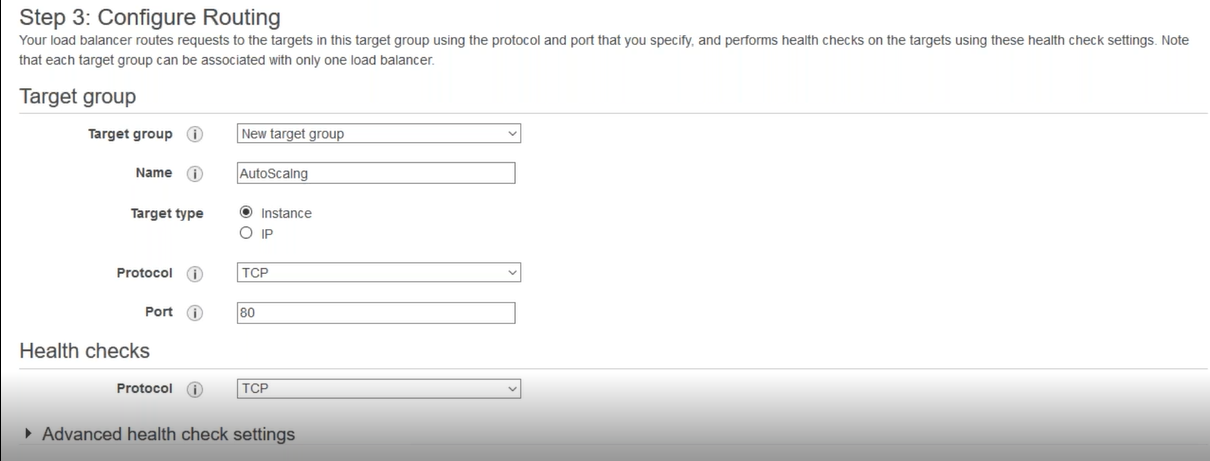
**Note ::** Later, if you want to use a different launch configuration, you can create a new one and apply it to any Auto Scaling group. Existing launch configurations cannot be edited.

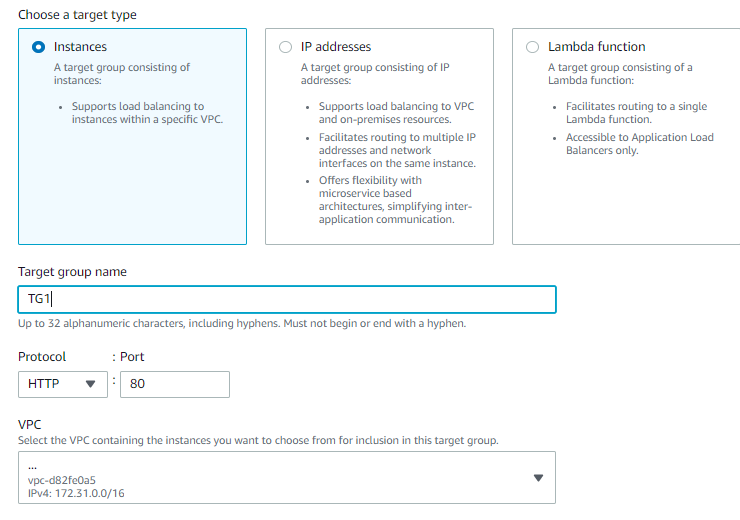
**Autoscaling groups:** select the launch configuration,,,choose VPC and subnet ,,Choose type of load balacer(App /Net ) or Classic and create a target group.Group size(Desired,maximum and minimum),,Scaling policies( Cpu ,network in/out bytes),SNS notification.

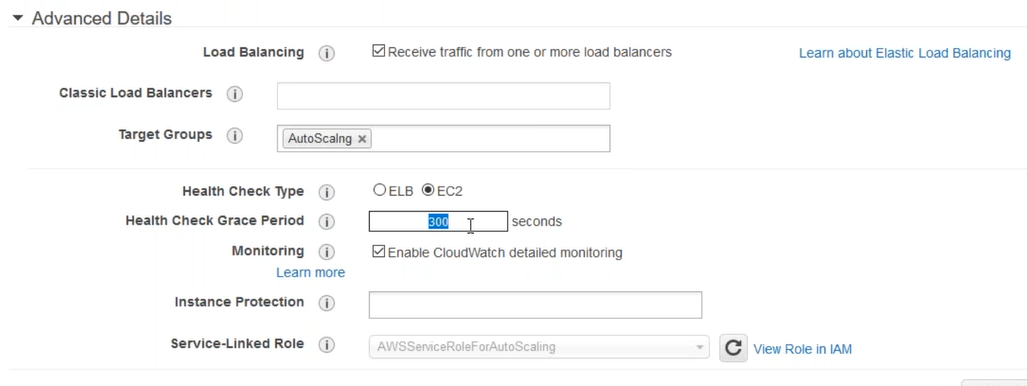


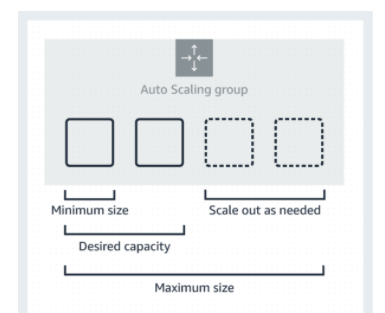


**Target Group**:: Your load balancer routes requests to the targets in a target group and performs health checks on the targets.’









**Horizontal scaling** means that you scale by adding more ec2 machines into your pool of resources

whereas **Vertical scaling** means that you scale by **adding more power (CPU, RAM)** to an existing ec2 machine.

<https://digitalcloud.training/certification-training/aws-solutions-architect-associate/compute/elastic-load-balancing/>

**what is health check grace period in aws?**

Your Load Balancer periodically sends requests to its registered instances to test their status. These tests are called health checks. ... The load balancer routes requests only to the healthy instances. When the load balancer determines that an instance is unhealthy, it stops routing requests to that instance.