Assignment 4 – ALB and ASG

ACCEPTANCE CRITERIA – Include the followings in the PDF:

- 2 links for ALB and NLB.
- Screenshot of healthy instances in both TGs of ALB and NLB.
- Screenshot of the Activity tab in the ASG.

Tasks:

Task 1. Run 2 web servers behind ALB

a. Create an SG for the ALB that allows access from the internet on port 80 (HTTP). Give a meaningful name like "alb-sg". The meaningful name will help when whitelisting this SG in the web servers' SG.
b. Create an SG for an EC2 instance (web servers). Open up port 80 from the ALB SG. That means the web servers only allow access from the load balancer.
c. Create 2 web servers in us-east-1a and us-east-1b AZs with different HTML content. To do that, hit "Edit" in the "Network Settings" and select subnets with "**us-east-1a**" for the first instance and "**us-east1b**" for the second instance. It is important because these AZs are where your load balancer nodes will be created. Lets say you created 2 instances in AZ 1a and 1b. But your load balancer nodes are created in AZ 1d and 1f, the load balancer cannot route the requests to the servers and you will see "**unused**" state in the target group.
d. Select the SG for the webserver you created in the previous step.
e. Put the following script in "User Data". So, your web server starts automatically when the server starts.

```
#!/bin/bash
yum install -y httpd
systemctl start httpd
systemctl enable httpd
echo "<h1>Hello YourName from $(hostname -f)</h1>" > /var/www/html/index.html
```

#!/bin/bash – is equivalent to "sudo -s" in bash.

f. Create ALB. Select **us-east-1a** and **us-east-1b** AZs for HA (High Availability). Create the TG and register the servers. And select the TG you created.

Task 2. Run web servers behind NLB

Similar to launching an ALB. The only difference is, to change TG protocol to **TCP** (Layer 4)**,** not HTTP (Layer 7).

Task 3. Run the web server in ASG

a. Deregister instances behind the ALB. We will register them through ASG. So they can scale automatically.

b. Create a launch template. Not launch configuration because the launch template is recommended. Launch template allows you to select AMI like EC2 where launch configuration requires you to enter AMI ID.
    i. Give it a name
    ii. Select the Amazon Linux AMI.
    iii. Select instance type, t2.micro.
    iv. Expand advanced. Select the IAM profile. Just in case you want to debug your web app, for example, to see if the web server is up with a custom HTML. But the web app is already configured automatically with user data.
    v. Enter the previous User Data above.
    vi. Select the web server's SG. Created in task 1.
    vii. Select any key pair. It doesn't matter. Because we use Session Manager to SSH into the instance if needed.
c. Create the Auto Scaling Group.
    i. Select launch template/configuration.
    ii. Select AZs (Subnets). That is where your instances launched.
    iii. Click on attach to an existing load balancer and select the default TG of the ALB.
    iv. Select ELB in the health checks panel.
    v. Set desired, min, and max capacity. Set a target tracking scale policy.
d. Mimic the high CPU utilization with the "stress" library to test scaling out behavior. See the last page for further reference.

Task 4. Clean up ALB, NLB, EC2 instances. They cost huge amounts.

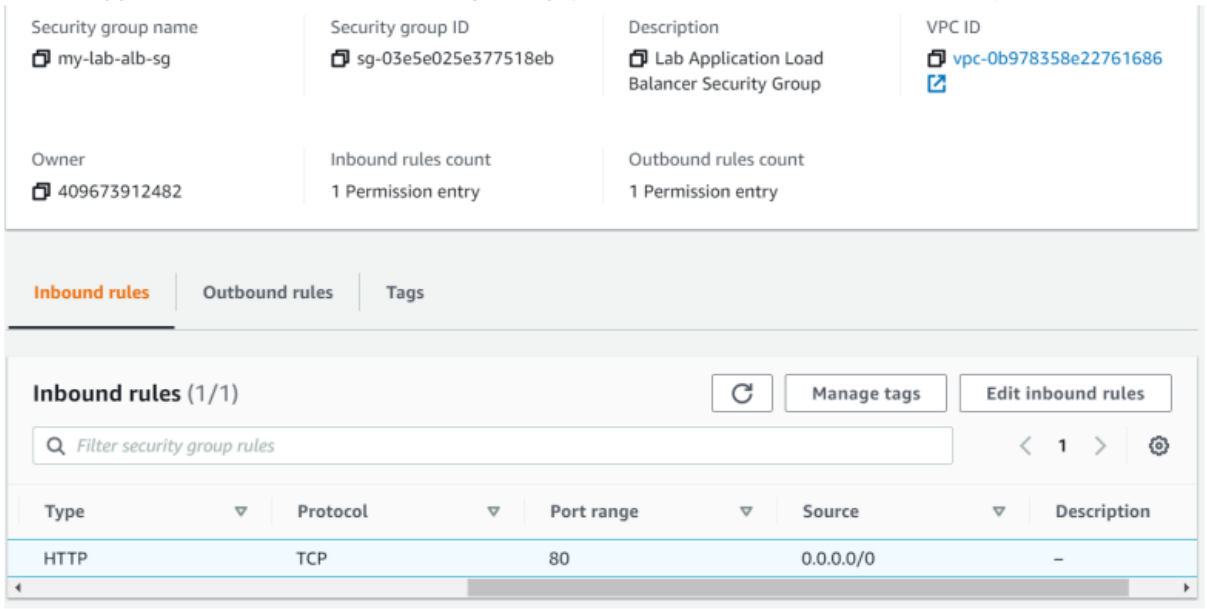
Step by step

Task 1 - Run 2 web servers behind ALB

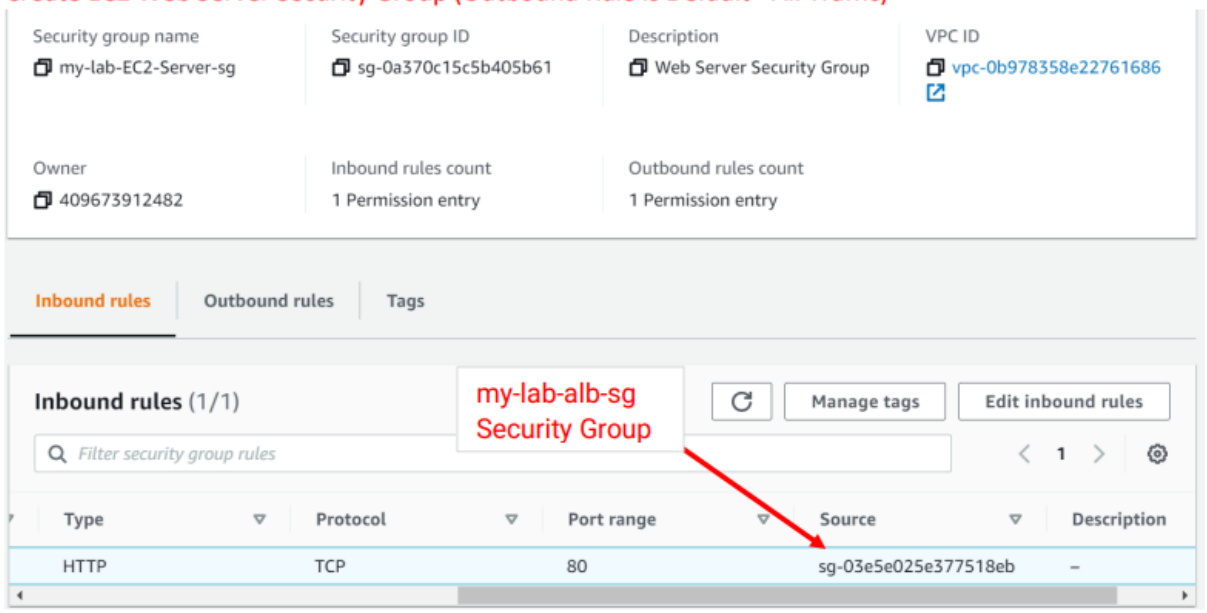
Create Security Groups for ALB

- Create an SG for the ALB which is open to the world.
- Create an SG for web servers that allows ALB's SG

Create Application Load Balancer Security Group (Outbound Rule is Default - All Traffic)

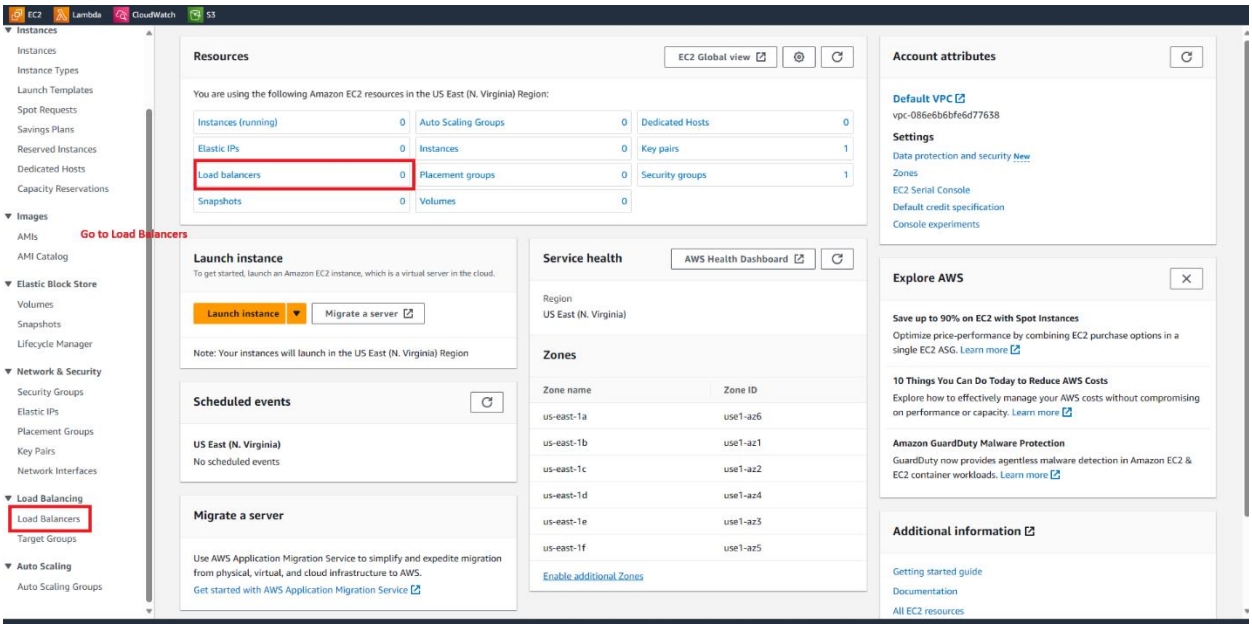


Create EC2 Web Server Security Group (Outbound Rule is Default - All Traffic)

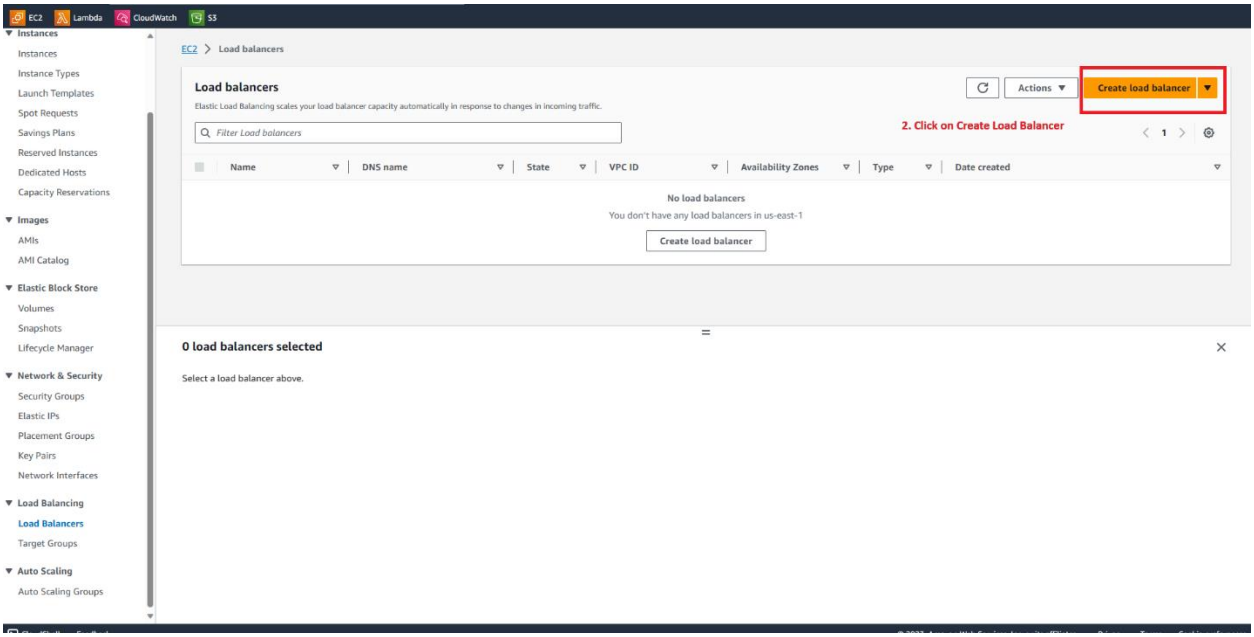


Create an ALB

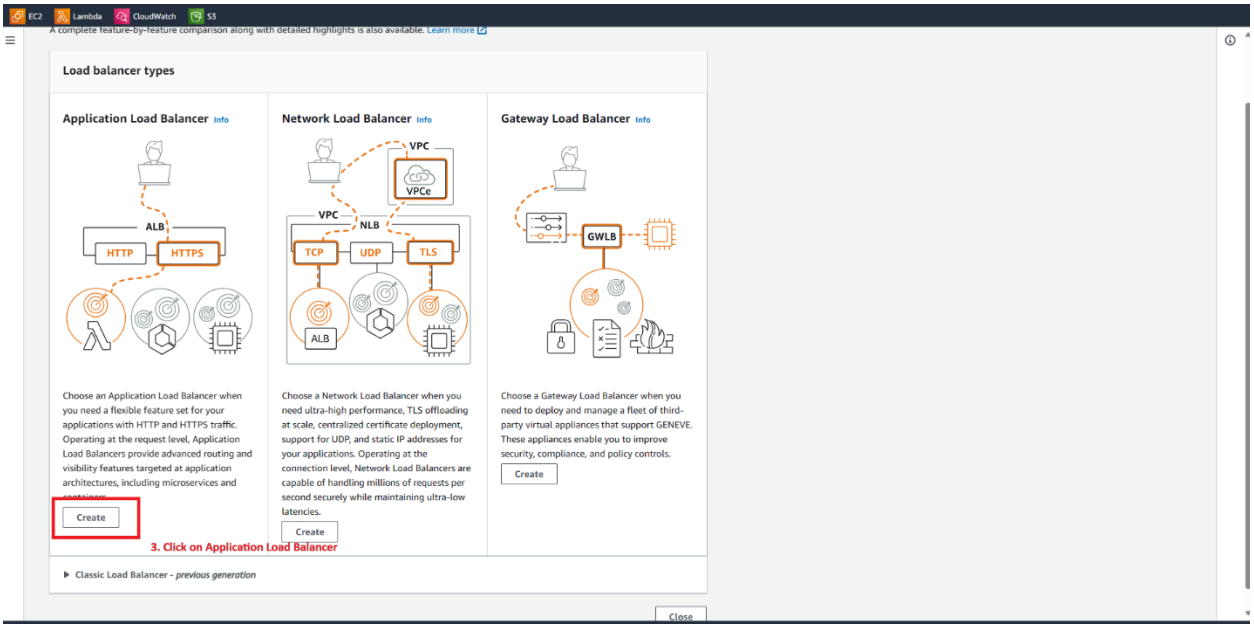
1: Go to Load Balancers Display in EC2 Dashboard.



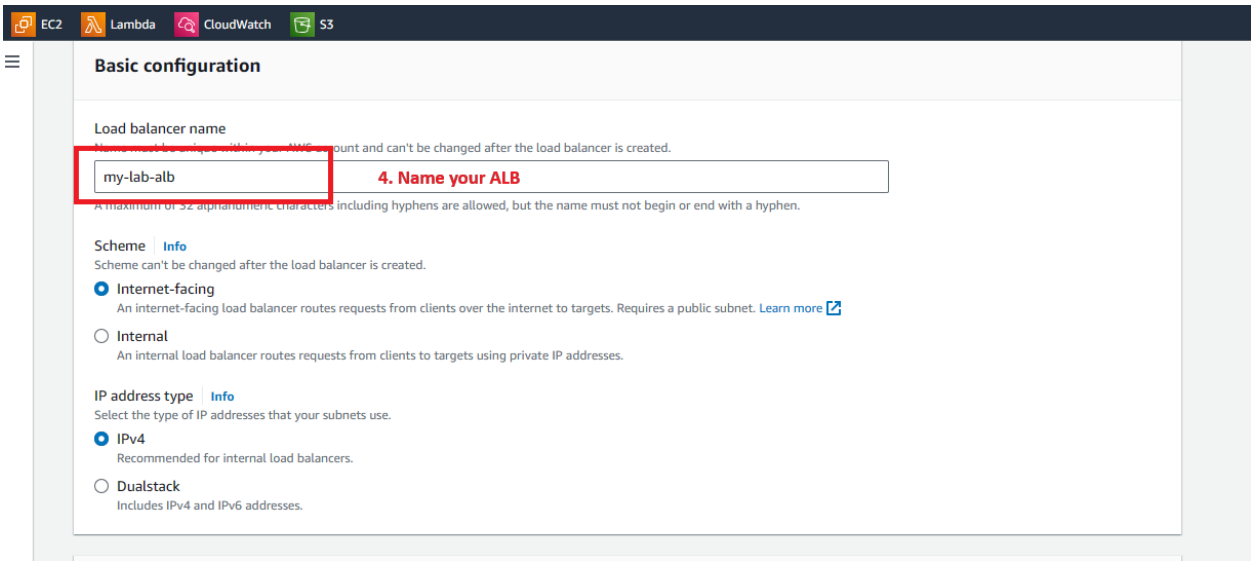
2: Once in Load Balancers Display, click on Create Load Balancer.



3: Click on Create Application Load Balancer.

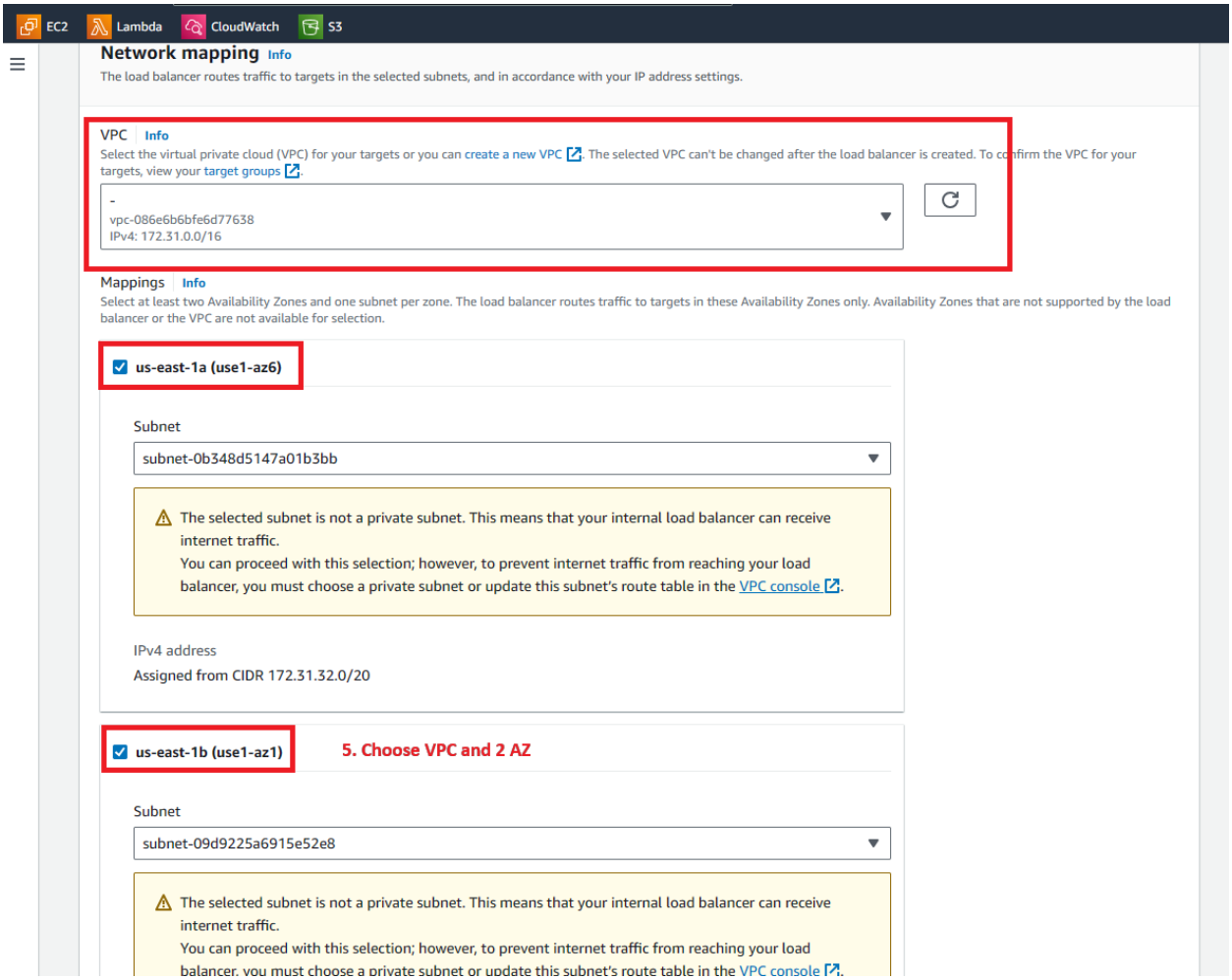


4: Name your ALB.

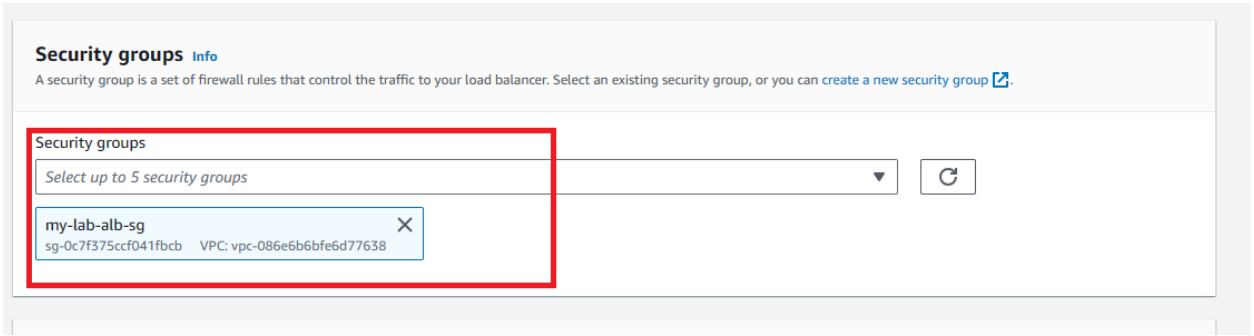


5: Select VPC

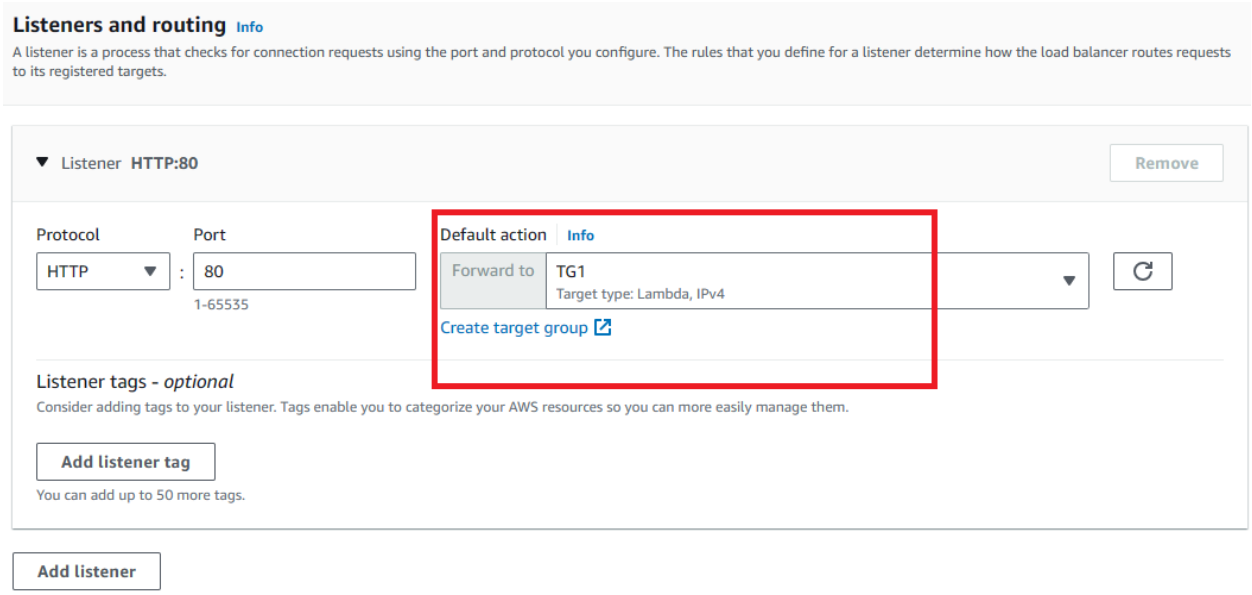
6: Select at least 2 AZ zones/subnets



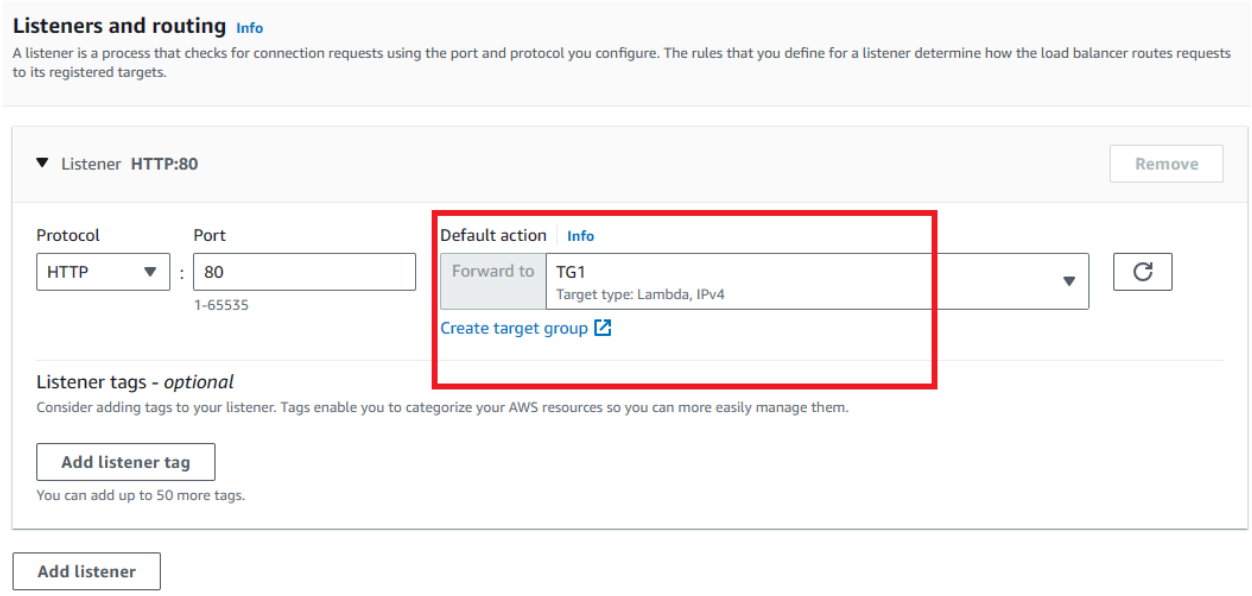
7: Select ALB SG you created



8: Select TG you created

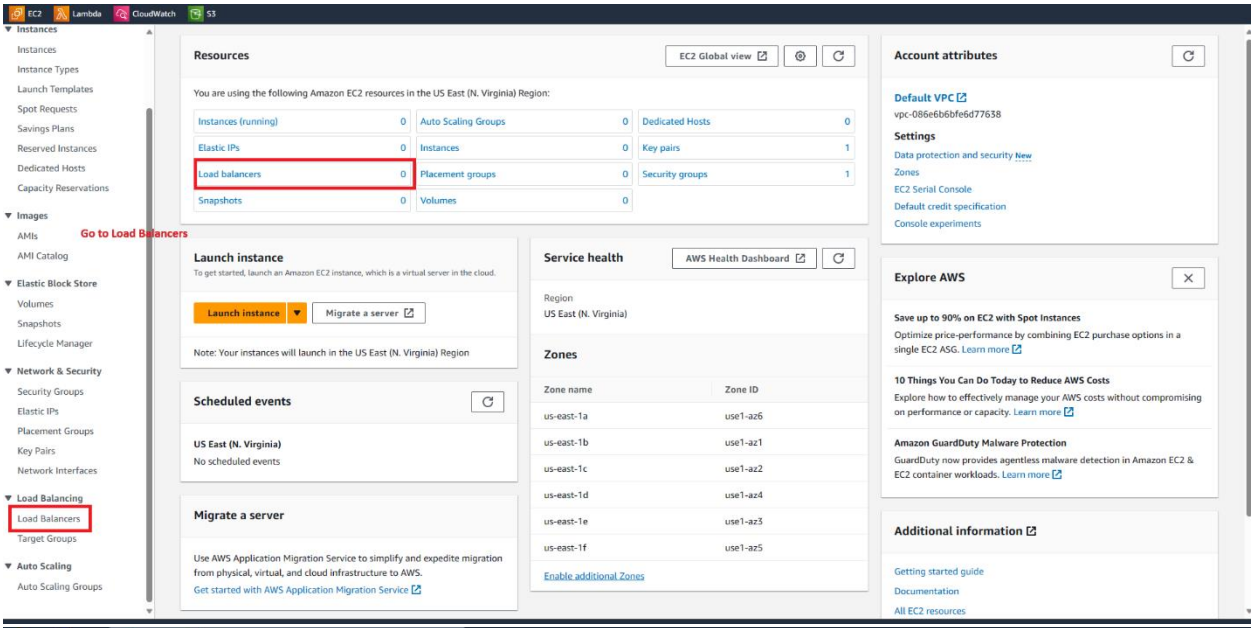


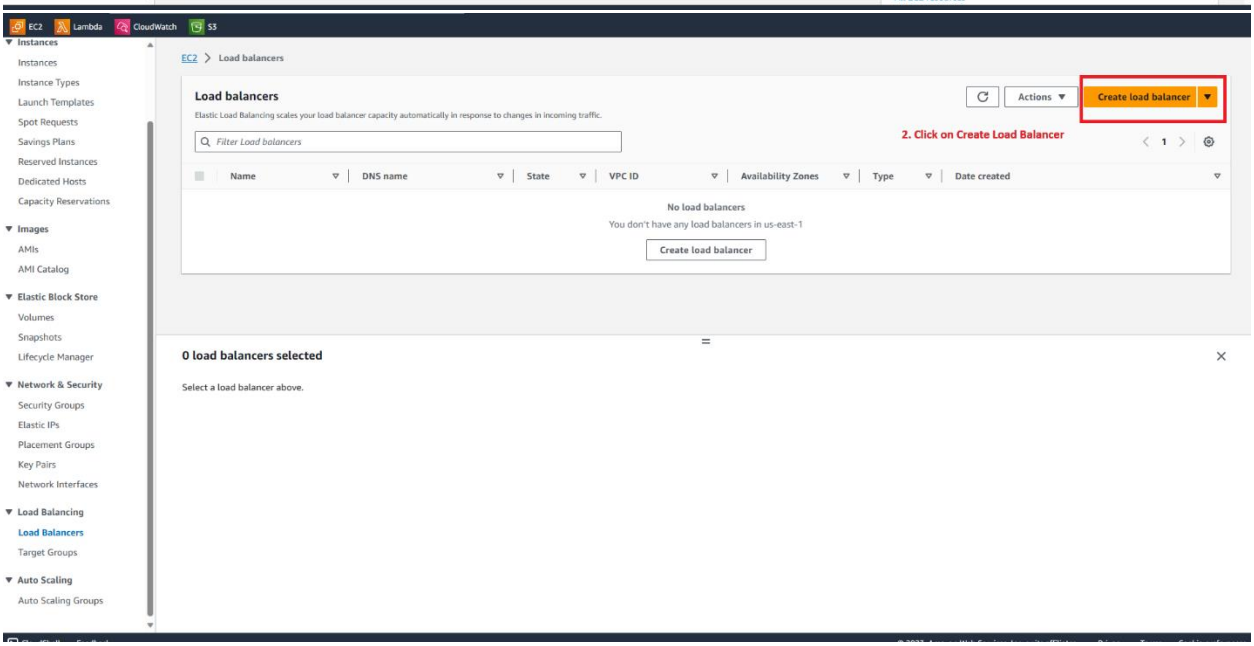
9: Create your load balancer.



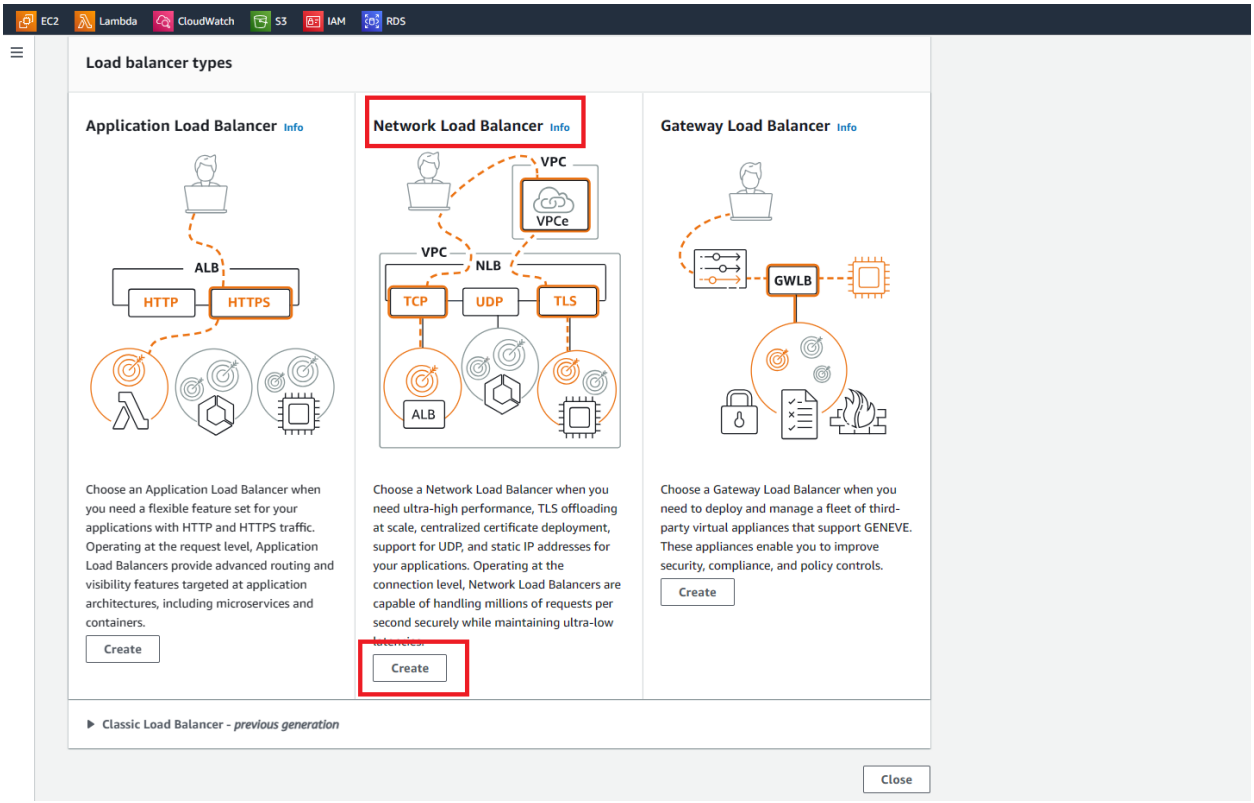
Task 2 – Create an NLB

1: Go to Load Balancers from EC2 dashboard and hit create Load Balancer.

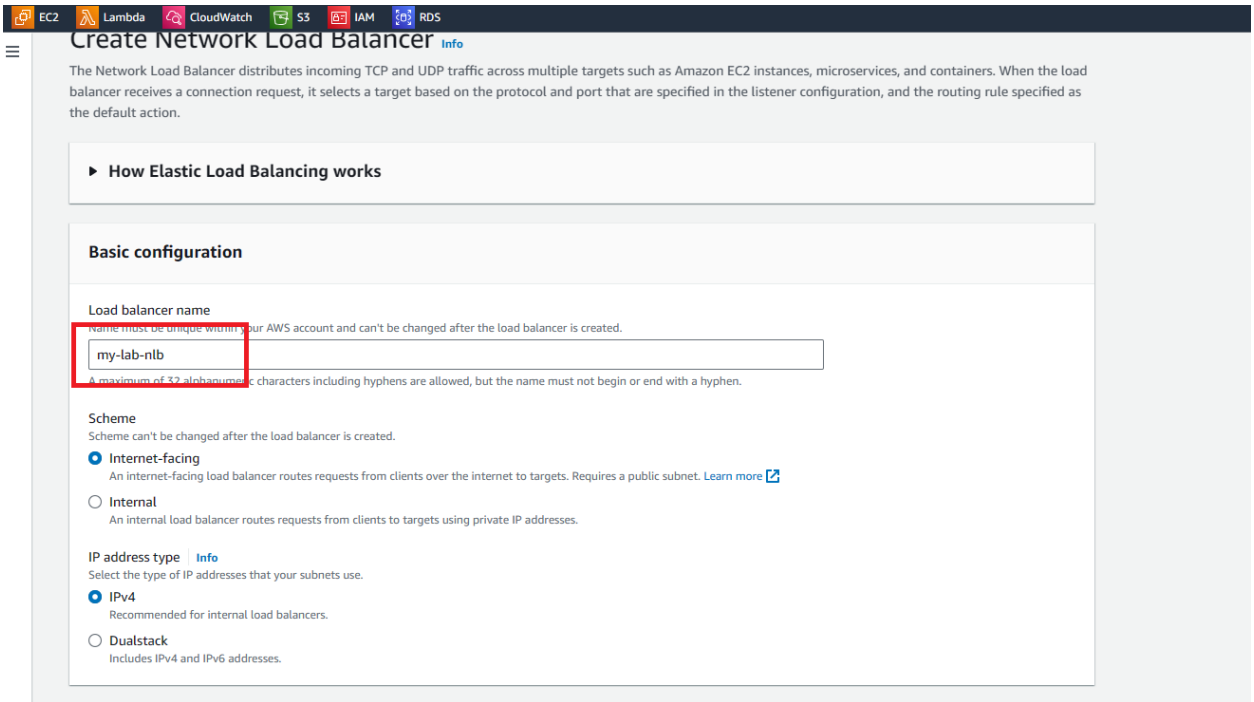




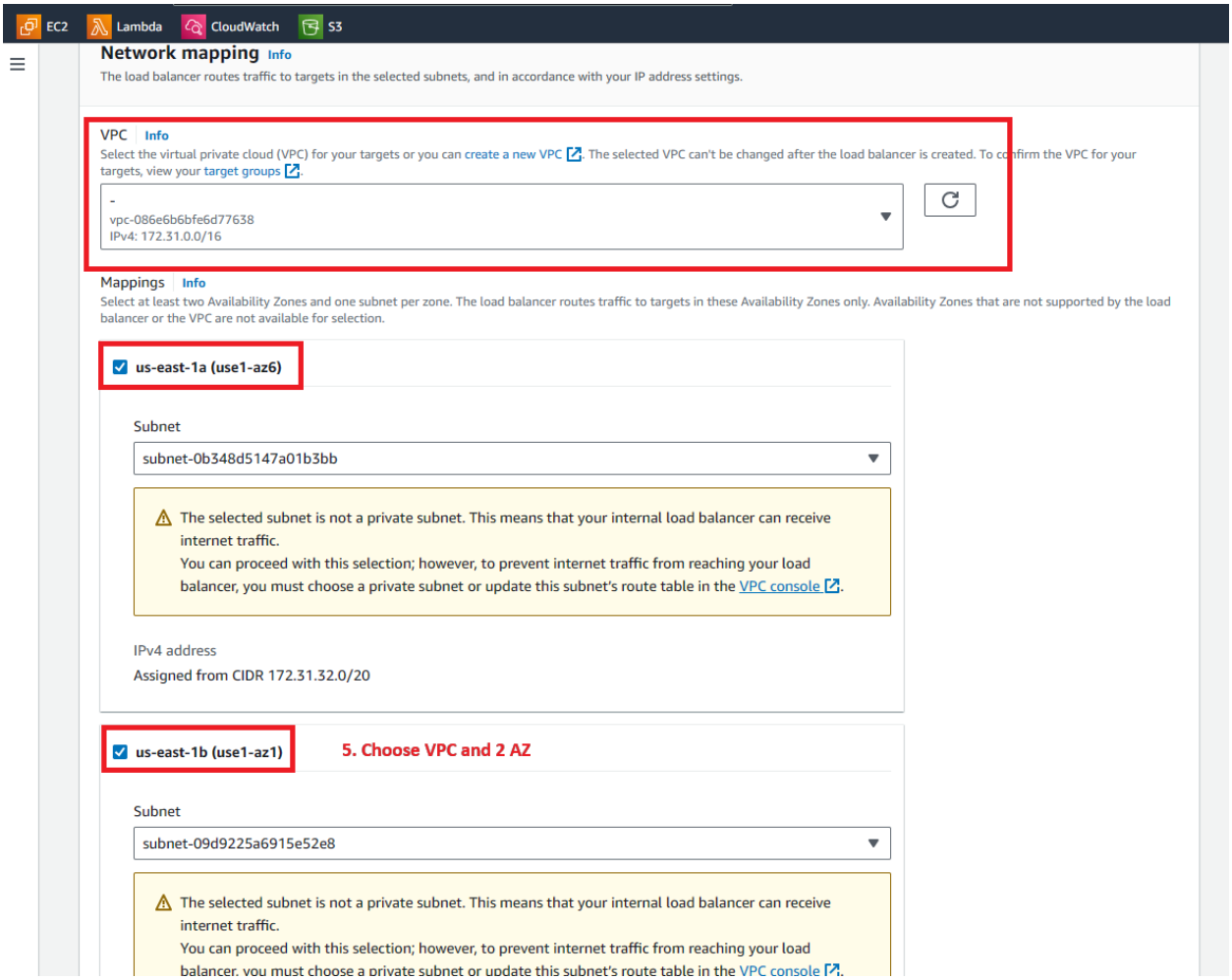
2: Select Network Load Balancer and hit create.



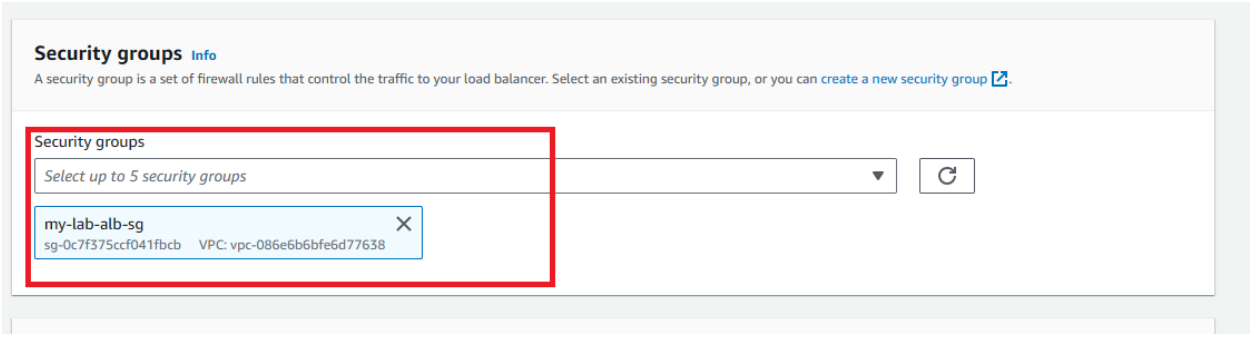
3: Name your NLB



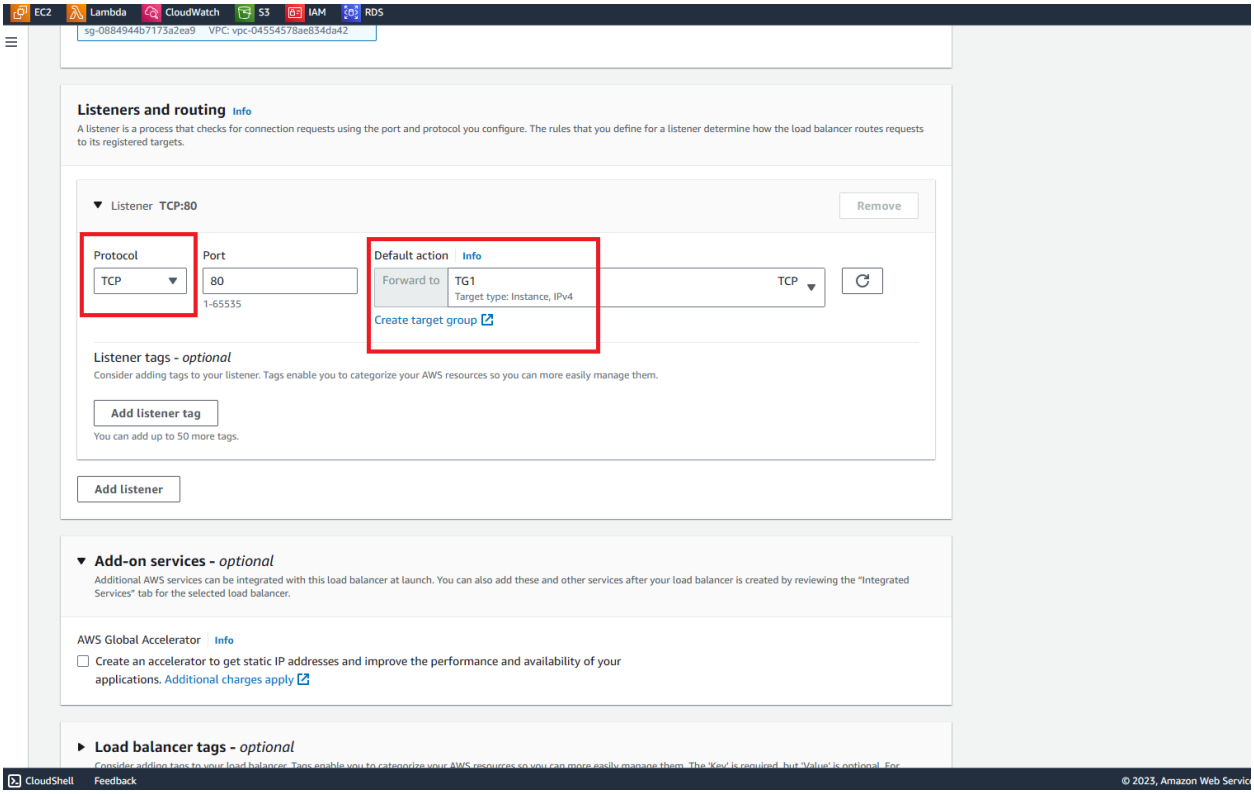
4: Select VPC and 2 AZs



5: Select SG



6: Select TG (remember, the protocol for NLB is TCP)

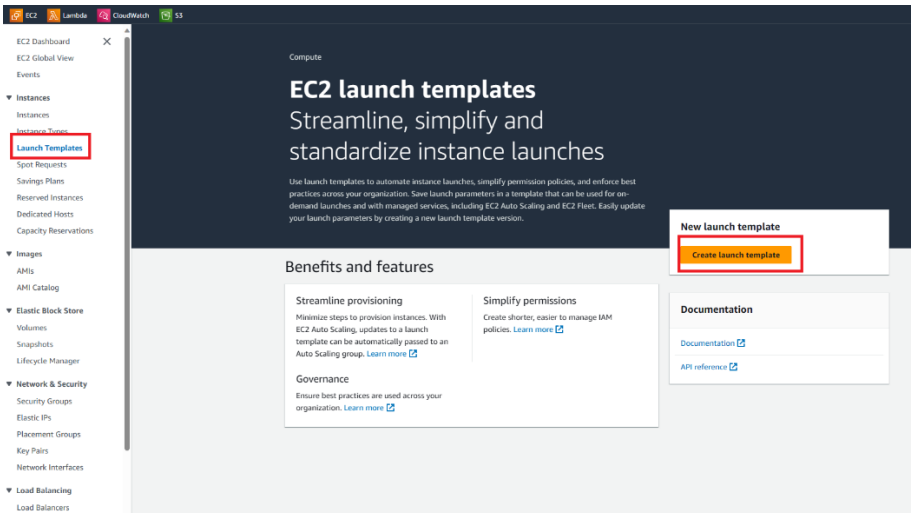


7: Create a Load Balancer.

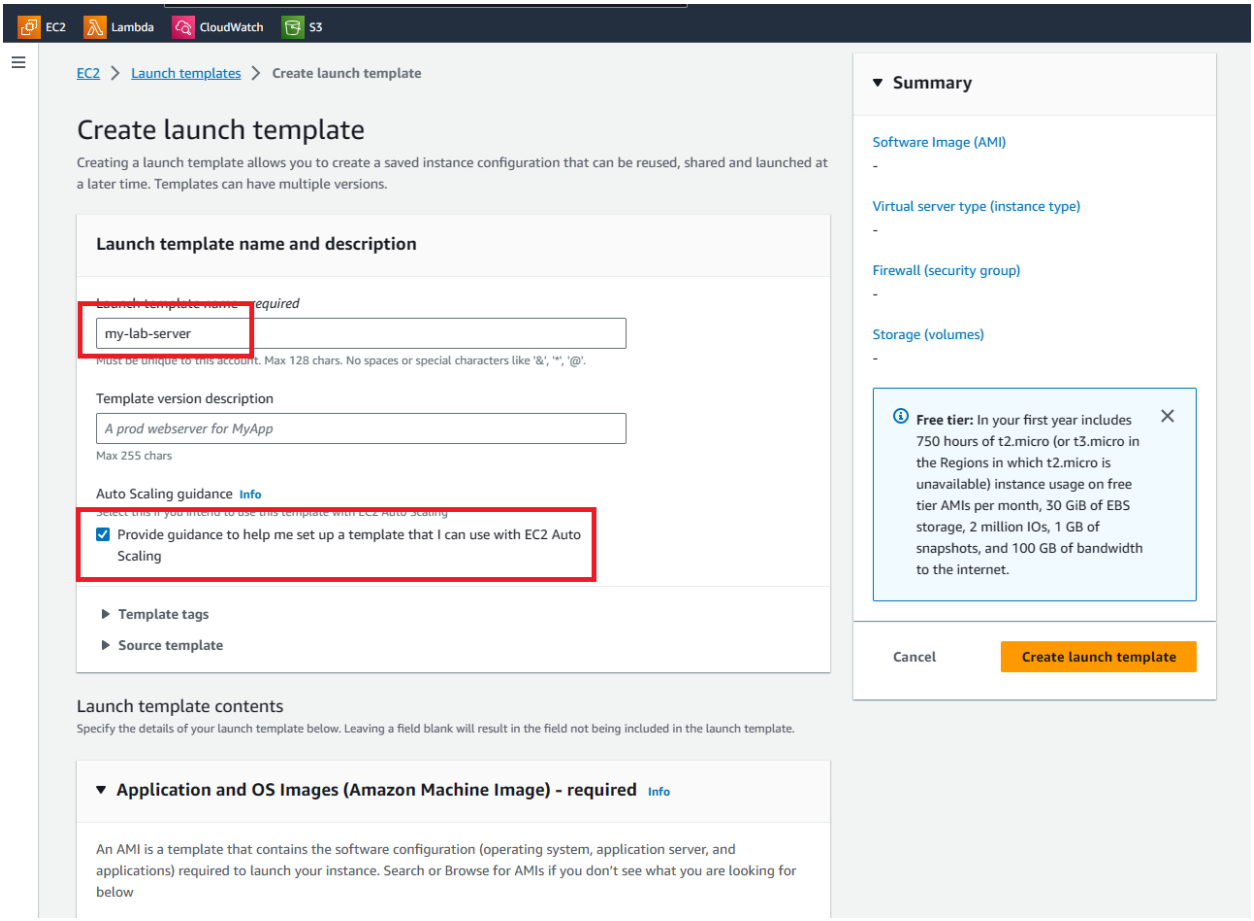
8: Follow the steps in Part D of this assignment to create 2 links for app1 and app2. Using rules and condition

Task 3 – Run the Web Server behind the ALB in ASG

1: Go to Launch Template in EC2 dashboard and hit create a launch template

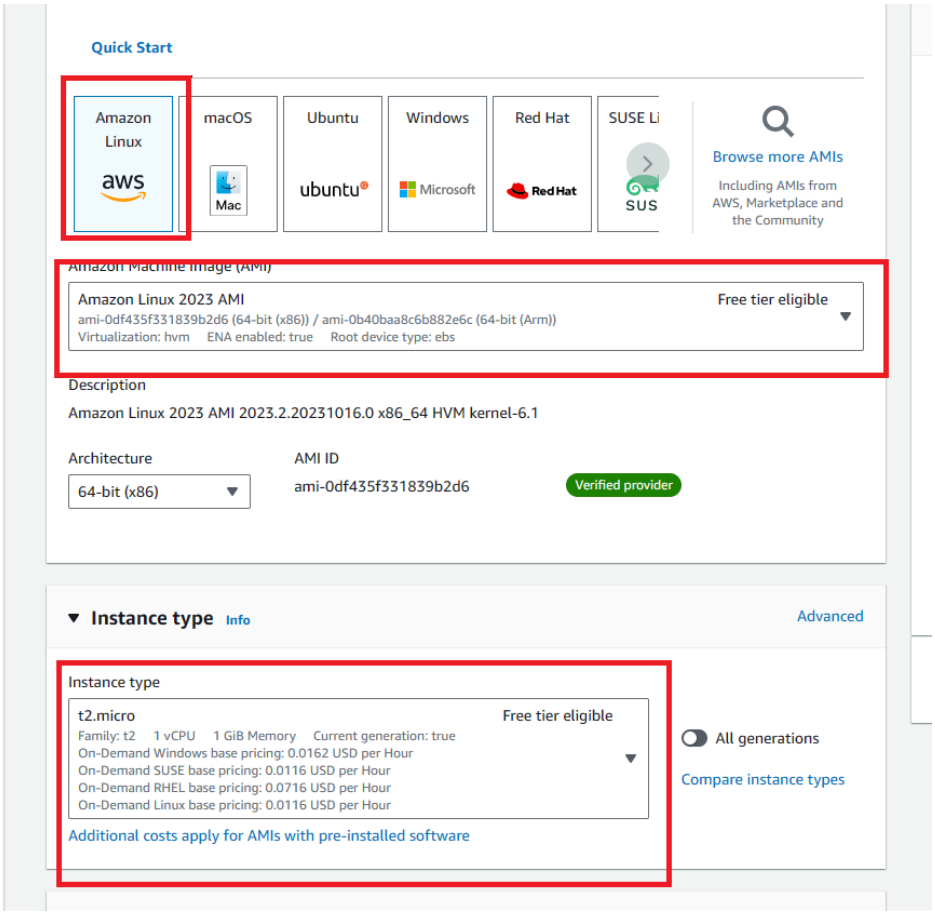


2: Provide a name and select guidance for a detailed assistance

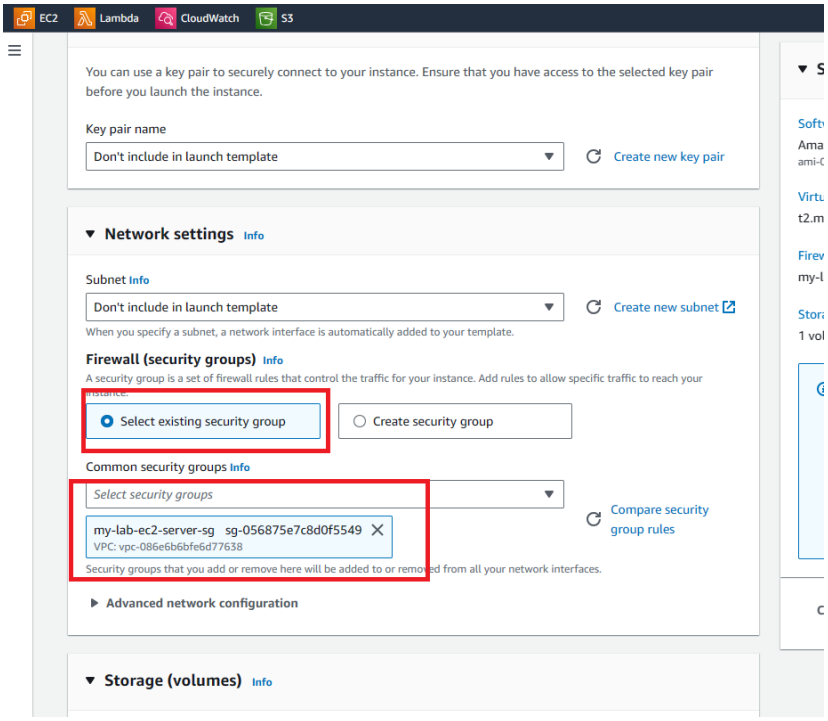


3: Select AMI

4: Select Instance Type

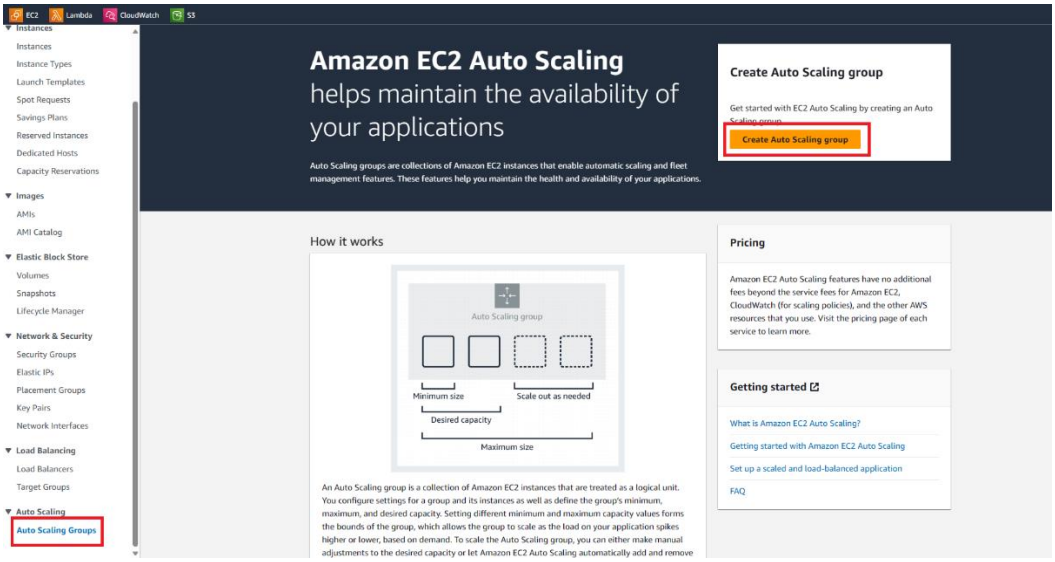


5: Select the security group you created. Then hit create.

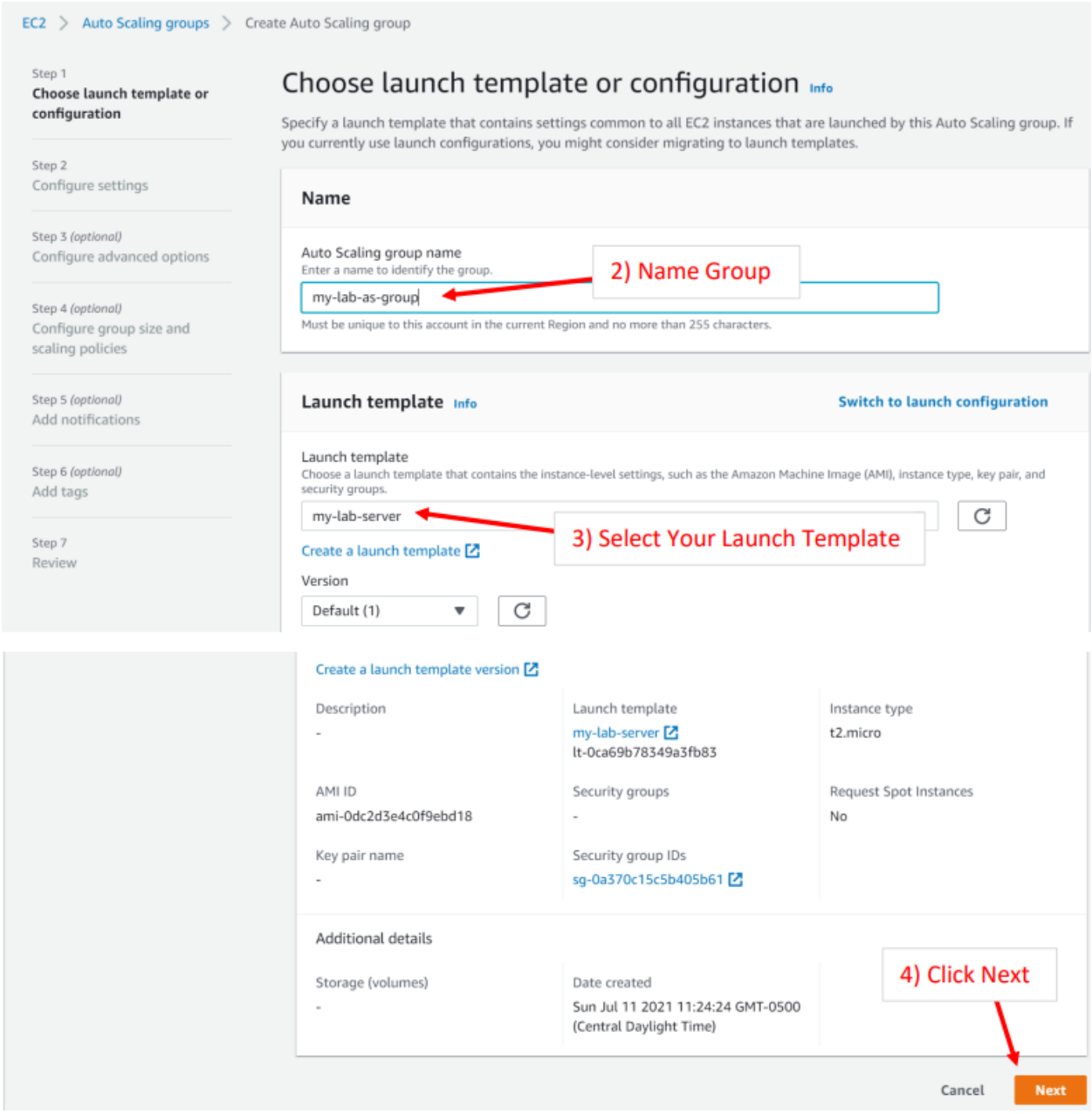


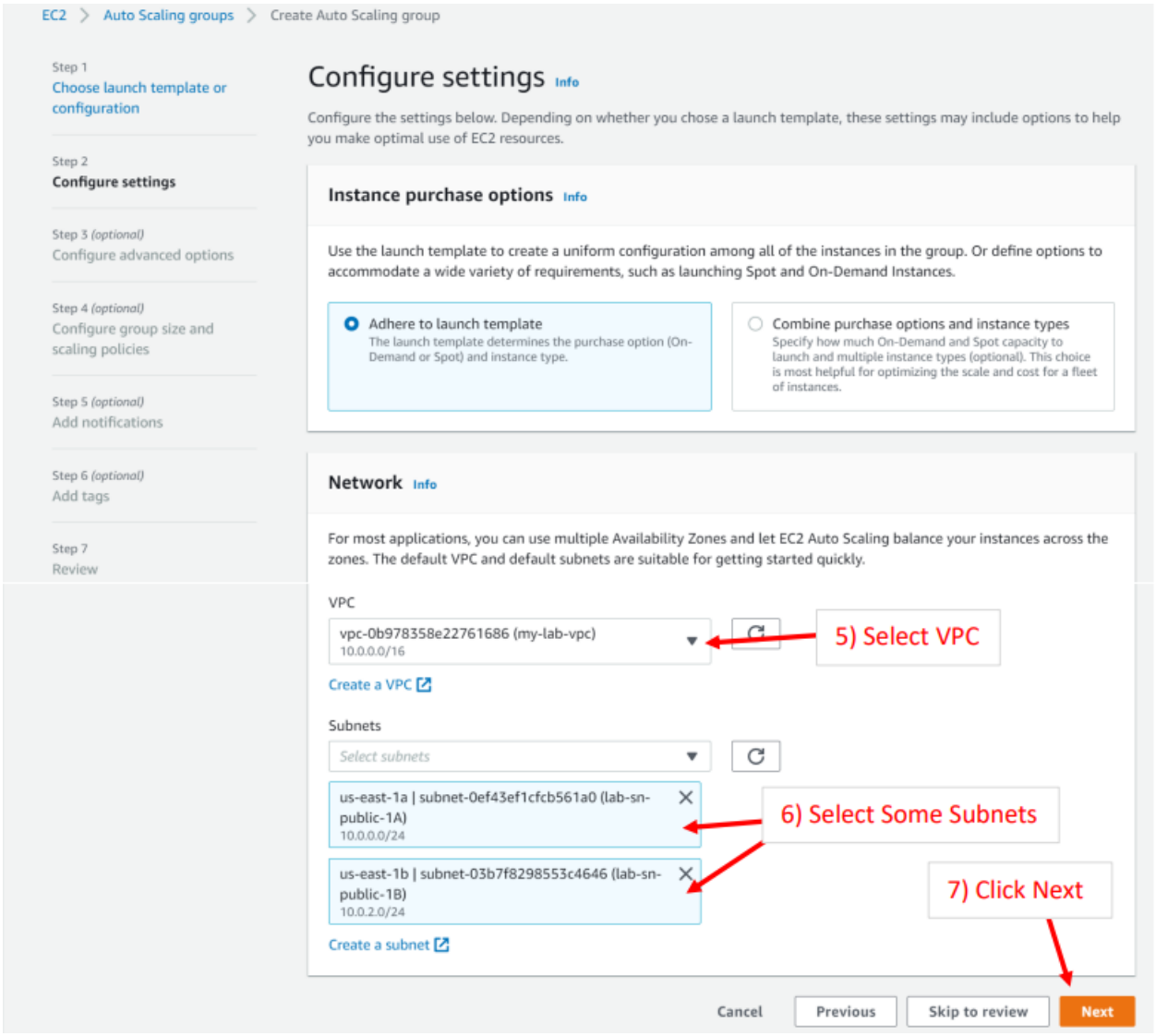
Create an Auto-Scaling Group

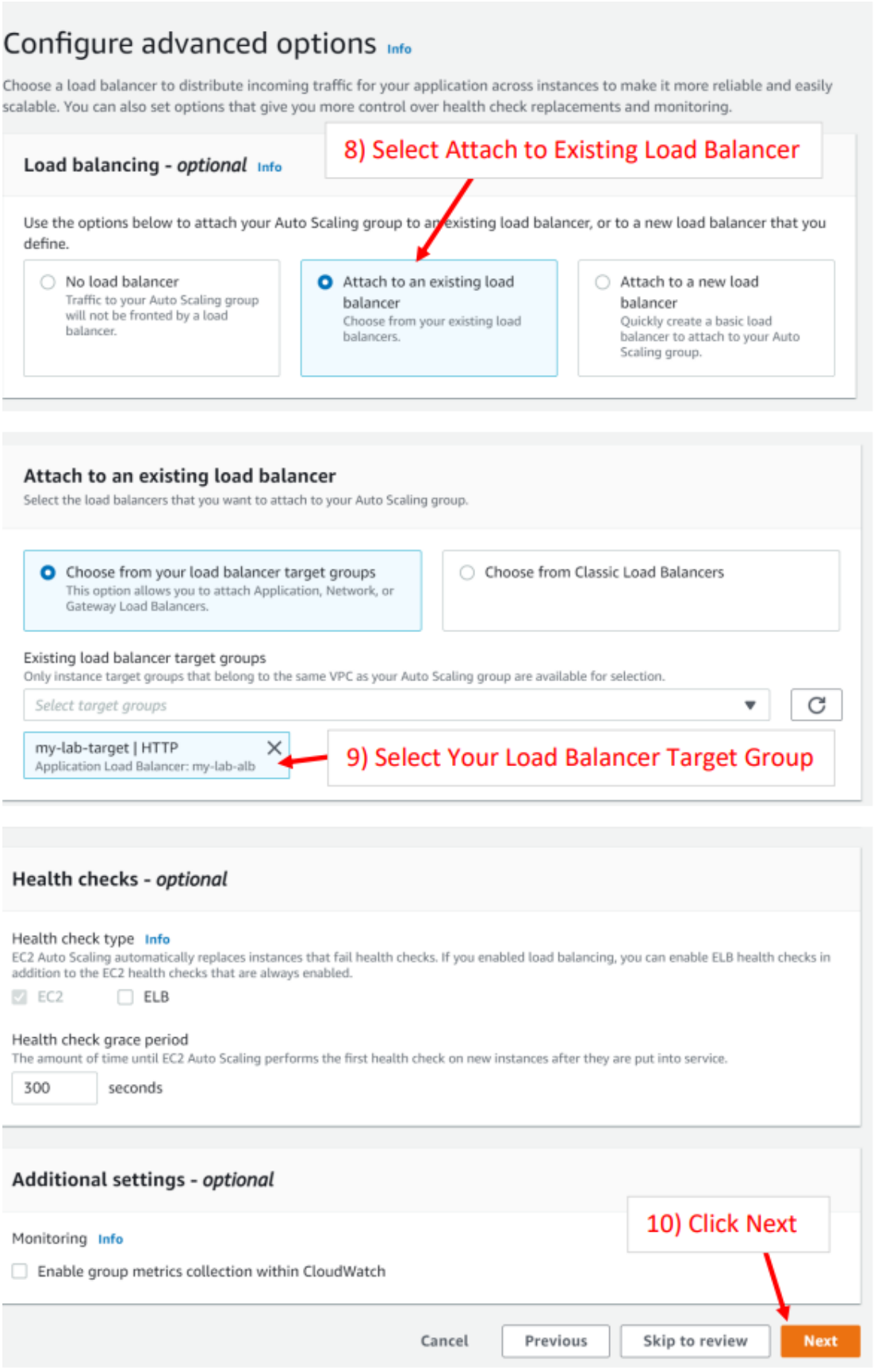
1: From EC2 dashboard, go to Auto Scaling Groups and click on Create ASG

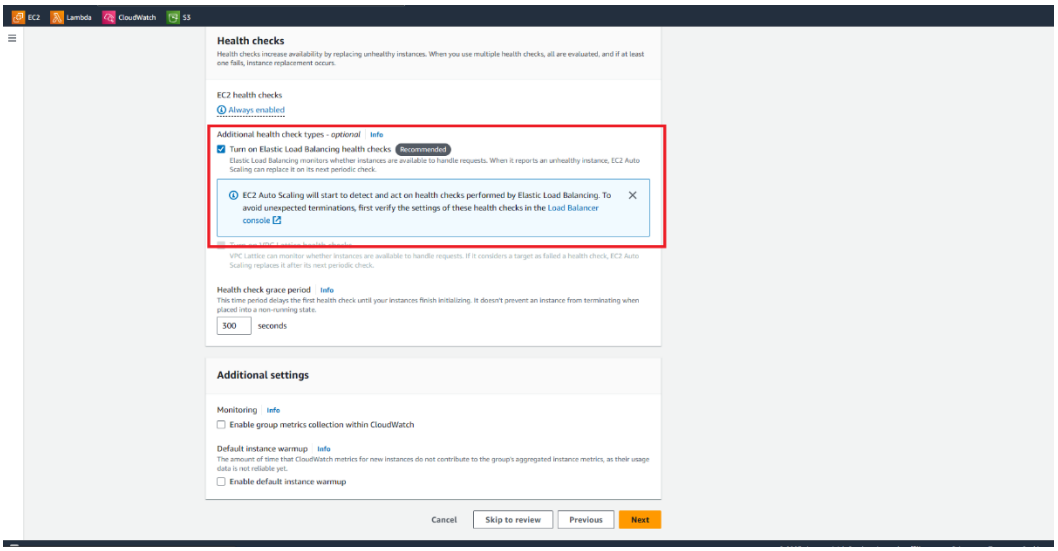


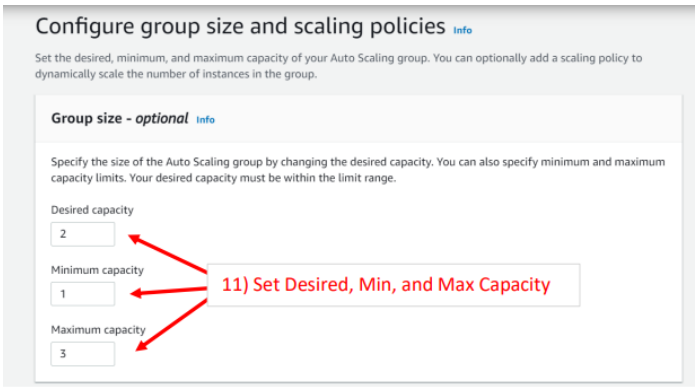
2: Name your group and select your Launch template you previously created.

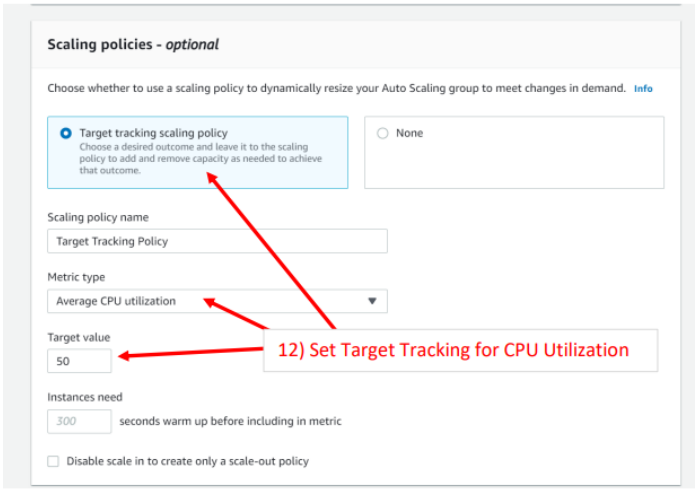


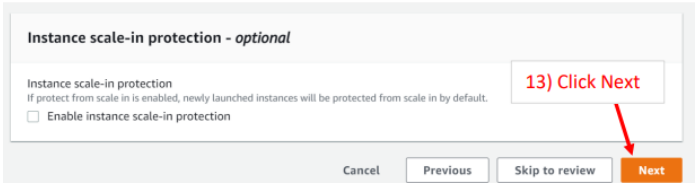

EC2 > Auto Scaling groups > Create Auto Scaling group
Step 1
Choose launch template or configuration
Step 2
Configure settings
Step 3 (optional)
Configure advanced options
Step 4 (optional)
Configure group size and scaling policies
Step 5 (optional)
Add notifications
Step 6 (optional)
Add tags
Step 7
Review
Configure settings Info
Configure the settings below. Depending on whether you chose a launch template, these settings may include options to help you make optimal use of EC2 resources.
Instance purchase options Info
Use the launch template to create a uniform configuration among all of the instances in the group. Or define options to accommodate a wide variety of requirements, such as launching Spot and On-Demand Instances.
Adhere to launch template
The launch template determines the purchase option (On-Demand or Spot) and instance type.
Combine purchase options and instance types
Specify how much On-Demand and Spot capacity to launch and multiple instance types (optional). This choice is most helpful for optimizing the scale and cost for a fleet of instances.
Network Info
For most applications, you can use multiple Availability Zones and let EC2 Auto Scaling balance your instances across the zones. The default VPC and default subnets are suitable for getting started quickly.
VPC
vpc-0b978358e22761686 (my-lab-vpc)
10.0.0.0/16
5) Select VPC
Create a VPC
Subnets
Select subnets
us-east-1a | subnet-0ef43ef1cfcb561a0 (lab-sn-public-1A)
10.0.0.0/24
us-east-1b | subnet-03b7f8298553c4646 (lab-sn-public-1B)
10.0.2.0/24
6) Select Some Subnets
7) Click Next
Create a subnet
Cancel
Previous
Skip to review
Next


Configure advanced options Info
Choose a load balancer to distribute incoming traffic for your application across instances to make it more reliable and easily scalable. You can also set options that give you more control over health check replacements and monitoring.
Load balancing - optional Info
8) Select Attach to Existing Load Balancer
Use the options below to attach your Auto Scaling group to an existing load balancer, or to a new load balancer that you define.
No load balancer
Traffic to your Auto Scaling group will not be fronted by a load balancer.
Attach to an existing load balancer
Choose from your existing load balancers.
Attach to a new load balancer
Quickly create a basic load balancer to attach to your Auto Scaling group.
Attach to an existing load balancer
Select the load balancers that you want to attach to your Auto Scaling group.
Choose from your load balancer target groups
This option allows you to attach Application, Network, or Gateway Load Balancers.
Choose from Classic Load Balancers
Existing load balancer target groups
Only instance target groups that belong to the same VPC as your Auto Scaling group are available for selection.
Select target groups
my-lab-target | HTTP
Application Load Balancer: my-lab-alb
9) Select Your Load Balancer Target Group
Health checks - optional
Health check type Info
EC2 Auto Scaling automatically replaces instances that fail health checks. If you enabled load balancing, you can enable ELB health checks in addition to the EC2 health checks that are always enabled.
EC2
ELB
Health check grace period
The amount of time until EC2 Auto Scaling performs the first health check on new instances after they are put into service.
300
seconds
Additional settings - optional
Monitoring Info
Enable group metrics collection within CloudWatch
10) Click Next
Cancel
Previous
Skip to review
Next





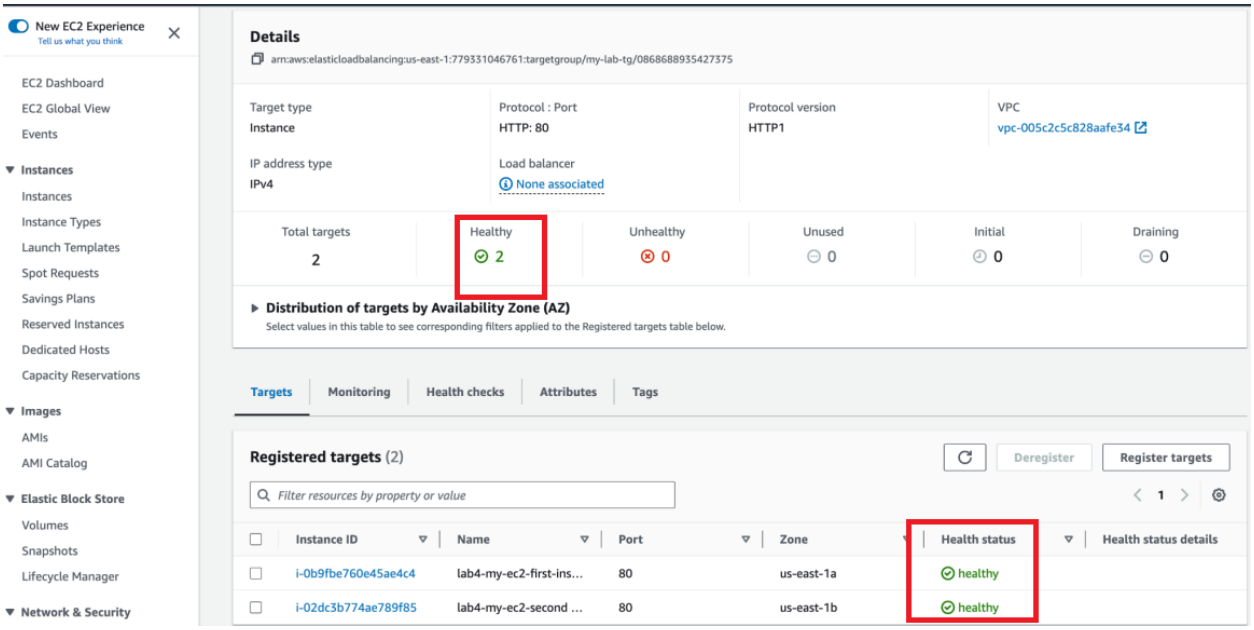




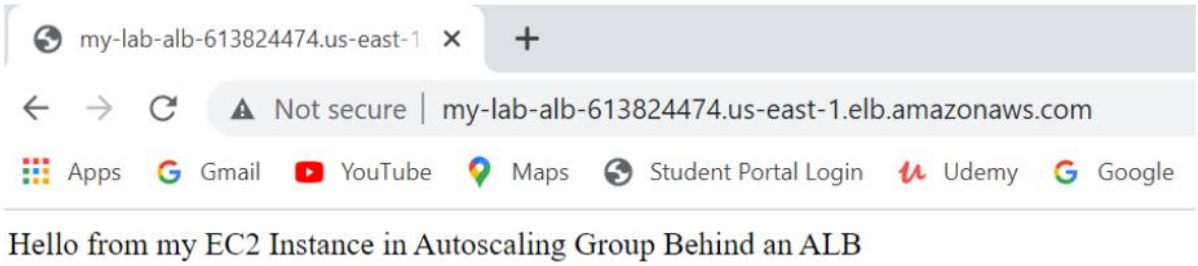
4: Finally create the ASG.

Verify and Test the ALB

View the Health check in your Target Group Details. Both instances should be healthy.



Test DNS with Web Browser



2: You can use EC2 stress tool to test out the scaling out.