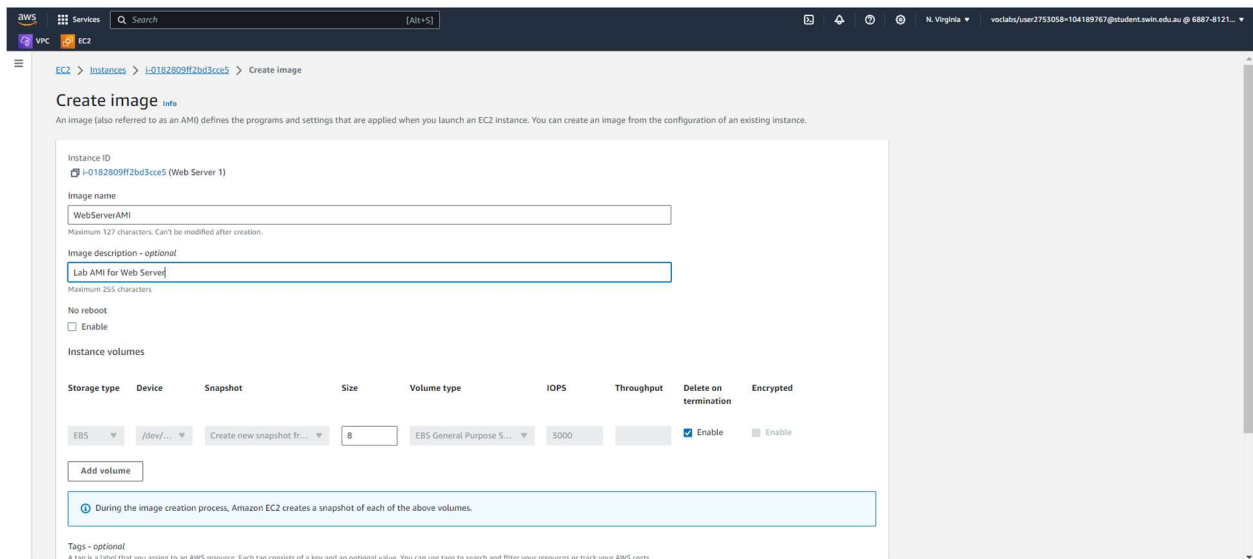


Lab 6: Scale and Load Balance Your Architecture

Name: Pham Do Tien Phong

Student ID: 104189767

Task 1: Create an AMI for Auto Scaling



The screenshot shows the AWS Management Console interface for creating a new Amazon Machine Image (AMI) from an existing EC2 instance. The breadcrumb navigation at the top indicates the path: EC2 > Instances > i-0182809f2bd3cc5 > Create image. The main heading is 'Create image' with an 'Info' link. Below this, a descriptive sentence states: 'An image (also referred to as an AMI) defines the programs and settings that are applied when you launch an EC2 instance. You can create an image from the configuration of an existing instance.'

The configuration form includes the following fields and options:

- Instance ID:** i-0182809f2bd3cc5 (Web Server 1)
- Image name:** WebServerAMI (with a note: 'Maximum 127 characters. Can't be modified after creation.')
- Image description - optional:** Lab AMI for Web Server (with a note: 'Maximum 255 characters.')
- No reboot:** ☐ Enable
- Instance volumes:** A table with columns: Storage type, Device, Snapshot, Size, Volume type, IOPS, Throughput, Delete on termination, and Encrypted. One volume is listed with Storage type 'EBS', Device '/dev/...', Snapshot 'Create new snapshot fr...', Size '8', Volume type 'EBS General Purpose S...', IOPS '3000', and 'Delete on termination' set to 'Enable'. An 'Add volume' button is located below the table.
- Tags - optional:** A note stating: 'A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS costs.'

A blue information box at the bottom of the form states: 'During the image creation process, Amazon EC2 creates a snapshot of each of the above volumes.'

I create image for the Web server instance with assigning image name and image description.

Task 2: Create a Load Balancer

Supports load balancing to VPC and on-premises resources.

- Facilitates routing to multiple IP addresses and network interfaces on the same instance.
- Offers flexibility with microservices based architectures, simplifying inter-application communication.
- Supports IPv6 targets, enabling end-to-end IPv6 communication, and IPv4-to-IPv6 NAT.

☐ Lambda function

- Facilitates routing to a single Lambda function.
- Accessible to Application Load Balancers only.

☐ Application Load Balancer

- Offers the flexibility for a Network Load Balancer to accept and route TCP requests within a specific VPC.
- Facilitates using static IP addresses and PrivateLink with an Application Load Balancer.

Target group name
LabGroup

A maximum of 32 alphanumeric characters including hyphens are allowed, but the name must not begin or end with a hyphen.

Protocol : Port
HTTP 80

IP address type
Only targets with the indicated IP address type can be registered to this target group.

☒ IPv4
Each instance has a default network interface (eth0) that is assigned the primary private IPv4 address. The instance's primary private IPv4 address is the one that will be applied to the target.

☐ IPv6
Each instance you register must have an assigned primary IPv6 address. This is configured on the instance's default network interface (eth0). [Learn more](#)

VPC
Select the VPC with the instances that you want to include in the target group. Only VPCs that support the IP address type selected above are available in this list.

Lab VPC
vpc-0d4d91f43023c5474
IPv4: 10.0.0.0/16

Protocol version
☒ HTTP1
Send requests to targets using HTTP/1.1. Supported when the request protocol is HTTP/1.1 or HTTP/2.

I create Target Group with assigning Target group name and choosing the Lab VPC.

Basic configuration

Load balancer name
Name must be unique within your AWS account and can't be changed after the load balancer is created.

LabELB

A maximum of 32 alphanumeric characters including hyphens are allowed, but the name must not begin or end with a hyphen.

Scheme [info](#)
Scheme can't be changed after the load balancer is created.

☒ Internet-facing
An internet-facing load balancer routes requests from clients over the internet to targets. Requires a public subnet. [Learn more](#)

☐ Internal
An internal load balancer routes requests from clients to targets using private IP addresses.

IP address type [info](#)
Select the type of IP addresses that your subnets use.

☒ IPv4
Recommended for internal load balancers.

☐ Dualstack
Includes IPv4 and IPv6 addresses.

Network mapping [info](#)
The load balancer routes traffic to targets in the selected subnets, and in accordance with your IP address settings.

VPC [info](#)
Select the virtual private cloud (VPC) for your targets or you can [create a new VPC](#). Only VPCs with an internet gateway are enabled for selection. The selected VPC can't be changed after the load balancer is created. To confirm the VPC for your targets, view your [target group](#).

Lab VPC
vpc-0d4d91f43023c5474
IPv4: 10.0.0.0/16

Mappings [info](#)
Select at least two Availability Zones and one subnet per zone. The load balancer routes traffic to targets in these Availability Zones only. Availability Zones that are not supported by the load balancer or the VPC are not available for selection.

☒ us-east-1a (use1-az1)

☒ us-east-1a (use1-az1)

Subnet
subnet-0f9364f99c253253 Public Subnet 1

IPv4 address
Assigned by AWS

☒ us-east-1b (use1-az2)

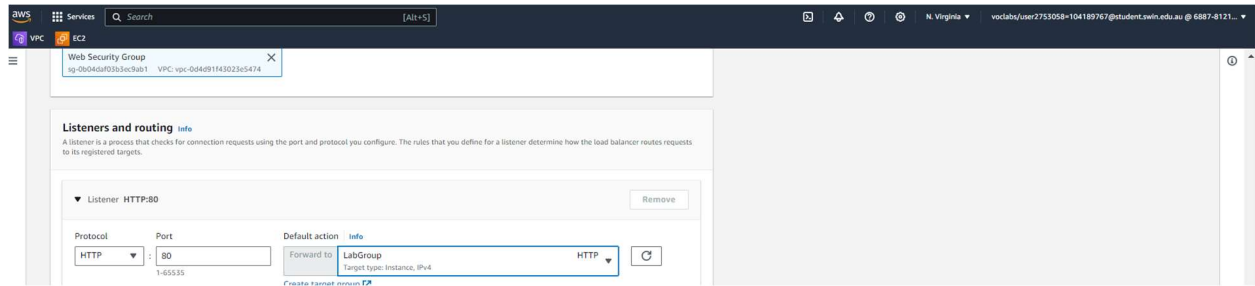
Subnet
subnet-0206fd3027d55670 Public Subnet 2

IPv4 address
Assigned by AWS

Security groups [info](#)
A security group is a set of firewall rules that control the traffic to your load balancer. Select an existing security group, or you can create a new security group.

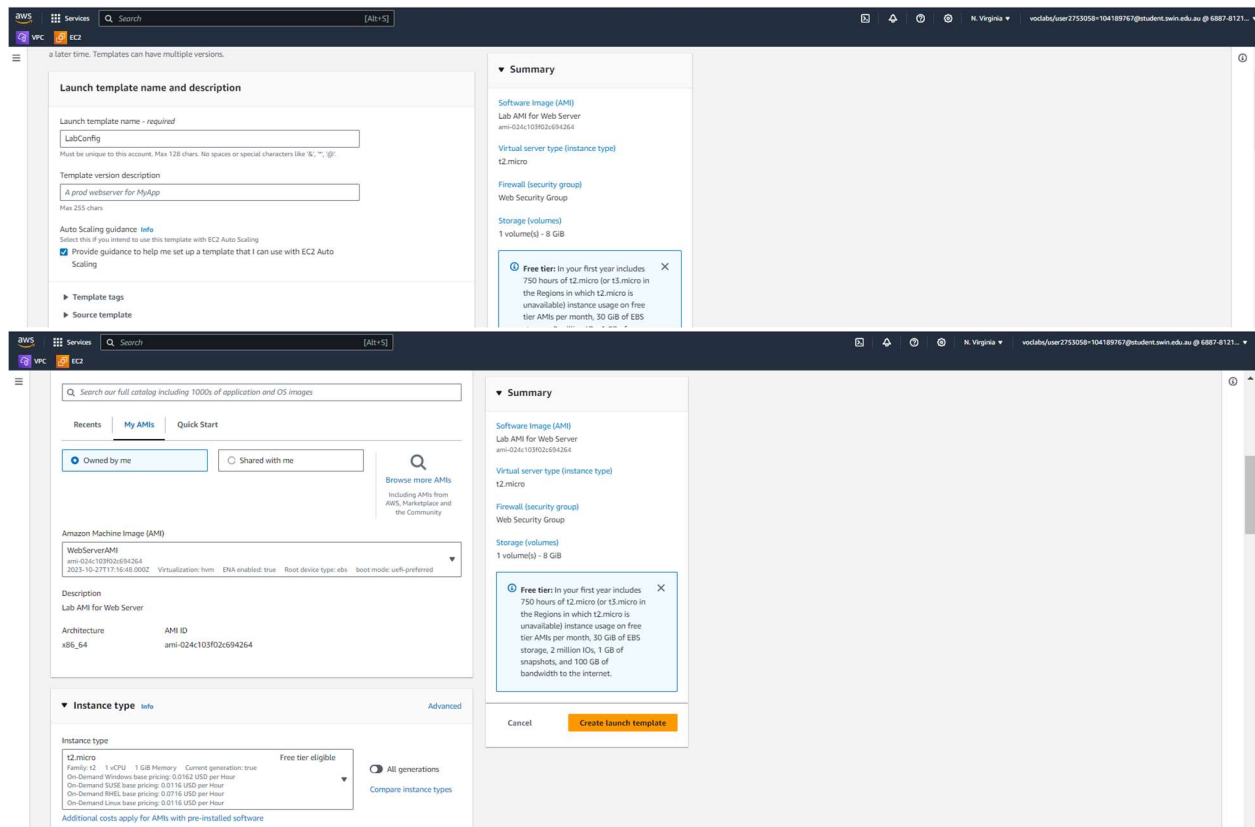
Security groups
Select up to 5 security groups

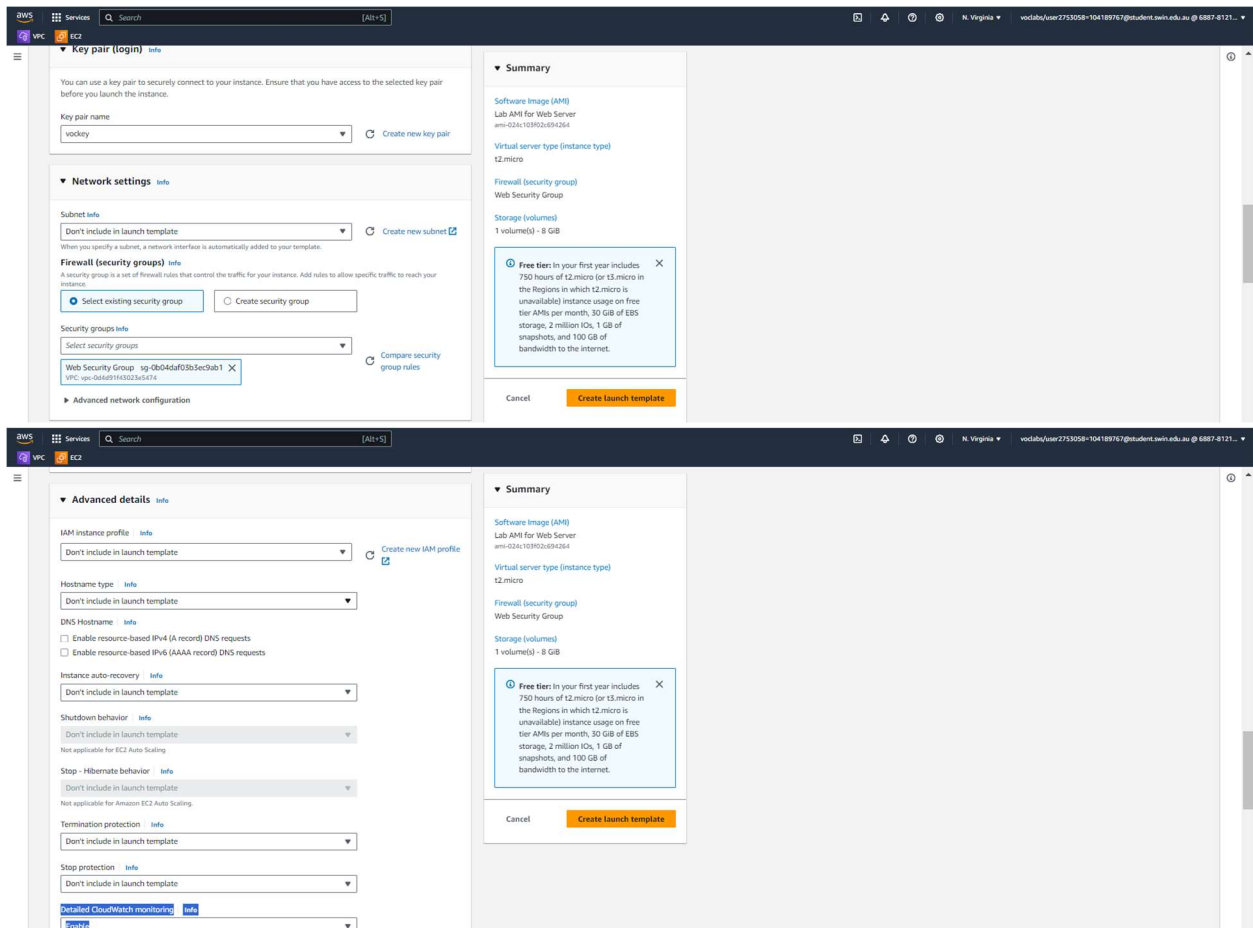
Web Security Group
sg-0b04d4f2b3ec9ab1 VPC: vpc-0d4d91f43023c5474



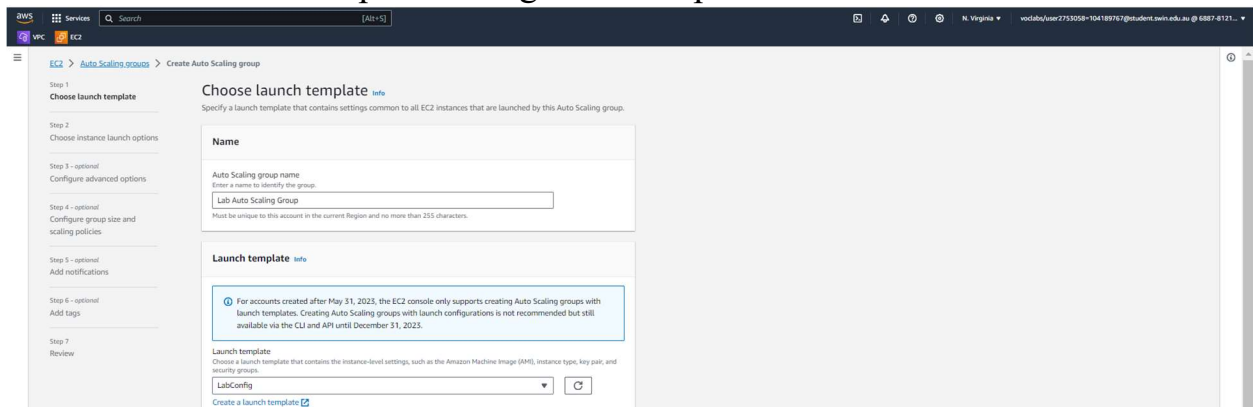
These images above are the steps I create application Load Balancer. I create Load Balancer name , choose Lab VPC, select Public subnet 1 and Public subnet 2 in 2 Availability Zone, choose Security Group (Web Security Group) and opt the default action for LabGroup.

Task 3: Create a Launch Template and an Auto Scaling Group





I create the Launch template through these steps .



I create Auto Scaling Group for the launch template I created above.

Choose instance launch options [info](#)

Choose the VPC network environment that your instances are launched into, and customize the instance types and purchase options.

Instance type requirements [info](#)

[Override launch template](#)

You can keep the same instance attributes or instance type from your launch template, or you can choose to override the launch template by specifying different instance attributes or manually adding instance types.

Launch template	Version	Description
LabConfig i	Default	-

Instance type
t2.micro

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
Choose the VPC that defines the virtual network for your Auto Scaling group:
vpc-064891f43025e474 (Lab VPC)
10.0.0.0/16 [Create a VPC](#) [i](#)

Availability Zones and subnets
Define which Availability Zones and subnets your Auto Scaling group can use in the chosen VPC.

Select Availability Zones and subnets [i](#)

us-east-1a subnet-0ab35d7e205ec909 (Private Subnet 1) 10.0.1.0/24	X
us-east-1b subnet-035d747c30c3dc94 (Private Subnet 2)	X

I choose instance launch options with choosing VPC as well as Availability Zones and subnets.

Configure advanced options - optional [info](#)

Integrate your Auto Scaling group with other services to distribute network traffic across multiple servers using a load balancer or to establish service-to-service communications using VPC Lattice. You can also set options that give you more control over health check replacements and monitoring.

Load balancing [info](#)

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 proxied 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 [i](#)

LabGroup HTTP Application Load Balancer: LabELB	X
--	---

Additional settings

Monitoring [info](#)

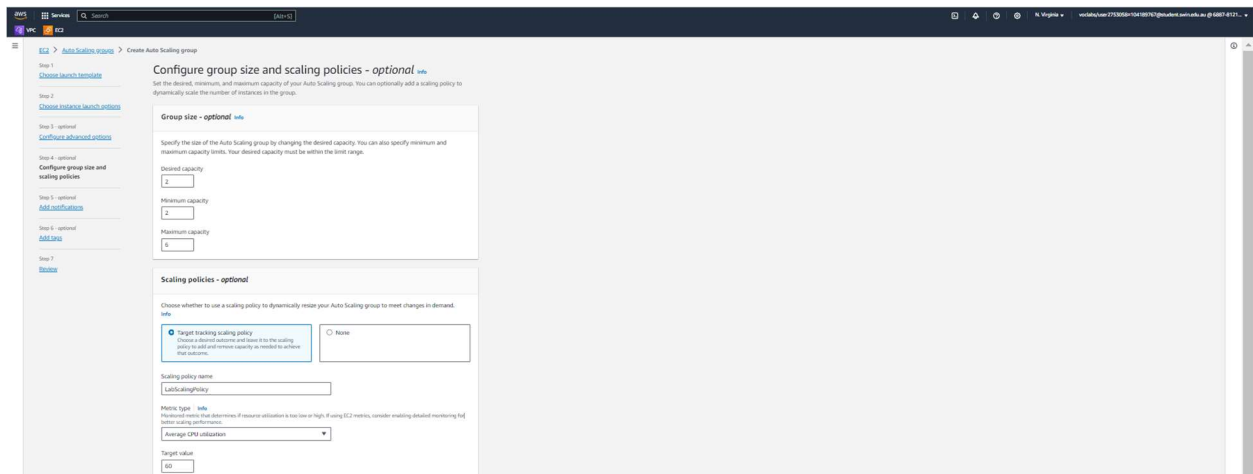
☒ Enable group metrics collection within CloudWatch

Default instance warmup [info](#)

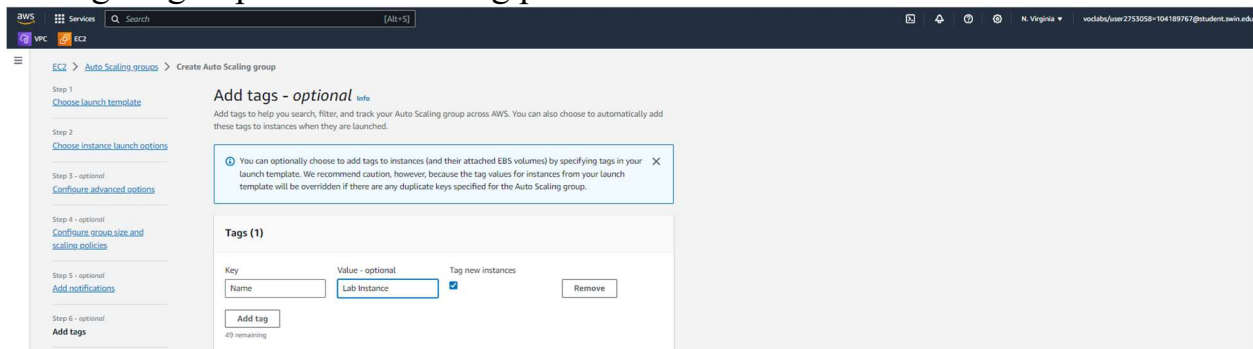
The amount of time that CloudWatch metrics for new instances do not contribute to the group's aggregated instance metrics, as their usage data is not reliable yet.

☐ Enable default instance warmup

Configure advanced options .

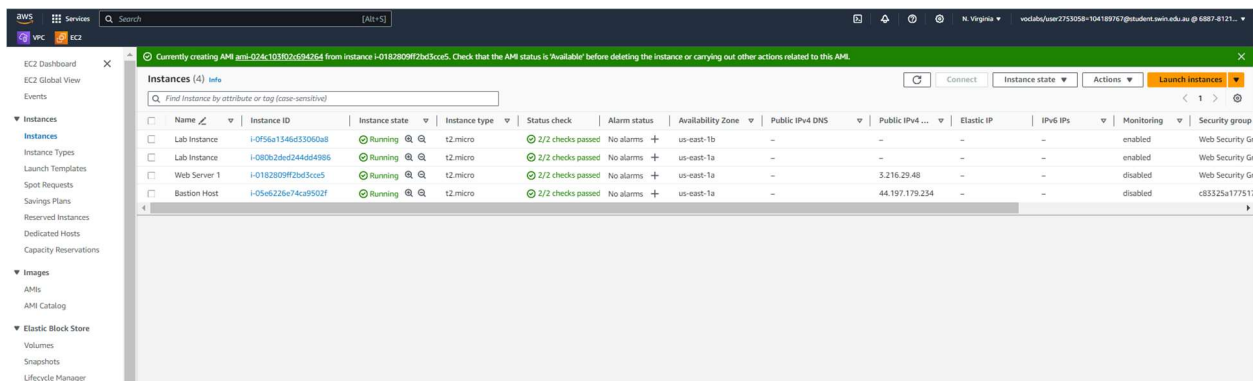


Configure group size and scaling policies.



Add tags for the auto scaling group.

Task 4: Verify that Load Balancing is Working



It has 2 Lab Instance instances.

The screenshot shows the AWS Management Console interface for the 'EC2' service, specifically the 'Target groups' page. The left sidebar contains navigation links for various AWS services. The main content area displays a table of target groups, with one group named 'LabGroup' selected. Below the table, a detailed view of the 'LabGroup' target group is shown, including tabs for 'Details', 'Targets', 'Monitoring', 'Health checks', 'Attributes', and 'Tags'. The 'Targets' tab is active, showing a list of registered targets. There are two targets listed, both with a 'Healthy' status.

Name	ARN	Port	Protocol	Target type	Load balancer	VPC ID
LabGroup	arn:aws:elasticloadbalancing:us-east-1:688781215973:loadbalancer/targetgroup/...	80	HTTP	Instance	LabELB	vpc-0d4d91f43023e5474

Instance ID	Name	Port	Zone	Health status	Health status details
i-0f56a134d633060a8	Lab Instance	80	us-east-1b	Healthy	
i-080b2aed244d45986	Lab Instance	80	us-east-1a	Healthy	

It has 2 target intances.

The screenshot shows the AWS Management Console interface for the 'EC2' service, specifically the 'Load balancers' page. The left sidebar contains navigation links for various AWS services. The main content area displays a table of load balancers, with one load balancer named 'LabELB' selected. Below the table, a detailed view of the 'LabELB' load balancer is shown, including tabs for 'Details', 'Listeners and rules', 'Network mapping', 'Security', 'Monitoring', 'Integrations', 'Attributes', and 'Tags'. The 'Details' tab is active, showing various configuration details for the load balancer, including its type, status, DNS name, and VPC ID. A tooltip indicates that the DNS name has been copied.

Name	DNS name	State	VPC ID	Availability Zones	Type	Date created
LabELB	LabELB-1750069467.us-east-1.elb.amazonaws.com	Active	vpc-0d4d91f43023e5474	2 Availability Zones	application	October 26, 2023, 01:26 (UTC+08:00)

Load balancer: LabELB

Details

- Load balancer type: Application
- Status: Active
- VPC: vpc-0d4d91f43023e5474
- IP address type: IPv4
- Scheme: Internet-facing
- Hosted zone: Z35SXDOTQ7X7K6
- Availability Zones: us-east-1a (use1-az1), us-east-1b (use1-az2)
- Date created: October 26, 2023, 01:26 (UTC+08:00)

Load balancer ARN: arn:aws:elasticloadbalancing:us-east-1:688781215973:loadbalancer/app/LabELB/46c062ccacbd9e4

DNS name copied: LabELB-1750069467.us-east-1.elb.amazonaws.com (A Record)

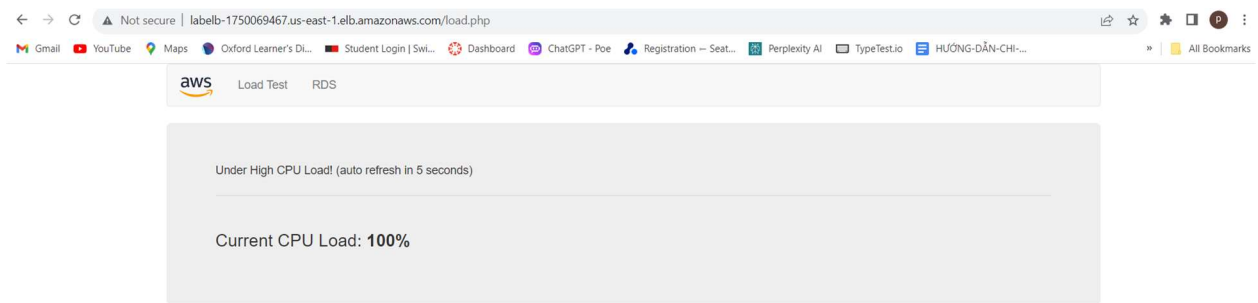
I copied the DNS name.

Task 5: Test Auto Scaling

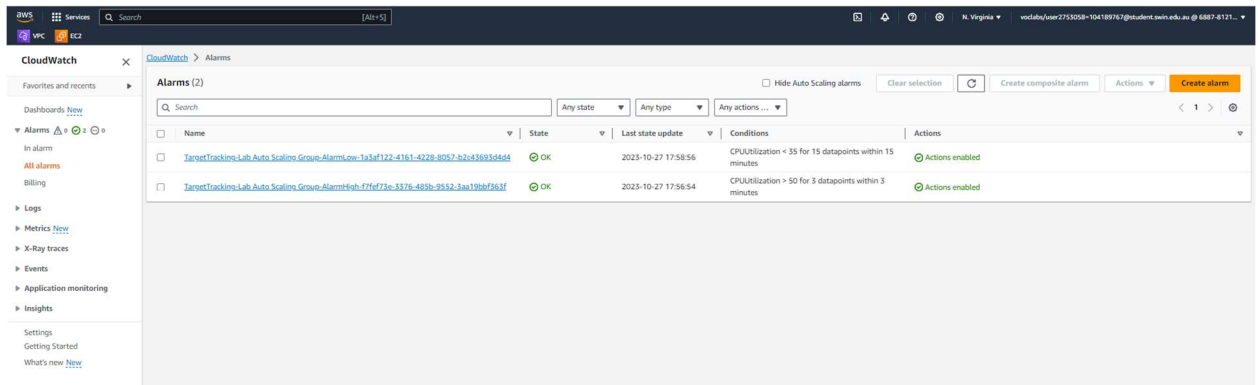
The screenshot shows the AWS Management Console interface for the 'CloudWatch' service, specifically the 'Alarms' page. The left sidebar contains navigation links for various AWS services. The main content area displays a table of alarms, with two alarms listed. Both alarms have an 'OK' state. The first alarm is named 'TargetTracking-Lab-Auto-Scaling-Group-AI-Mid-High' and the second is named 'TargetTracking-Lab-Auto-Scaling-Group-AI-Mid-Low'. Both alarms have conditions based on CPU utilization.

Name	State	Last state update	Conditions	Actions
TargetTracking-Lab-Auto-Scaling-Group-AI-Mid-High-cafab66b-7211-4800-bc55-afce82ef8c95	OK	2023-10-27 17:43:03	CPUUtilization > 60 for 3 datapoints within 5 minutes	Actions enabled
TargetTracking-Lab-Auto-Scaling-Group-AI-Mid-Low-be9f3341-3199-4682-b104-5672372948e6	OK	2023-10-27 17:40:06	CPUUtilization < 54 for 15 datapoints within 15 minutes	Actions enabled

It has OK state.



I choose Load Test



After waiting, all alarms are OK state.

Task 6: Terminate Web Server 1



I terminated the Web Server instance.