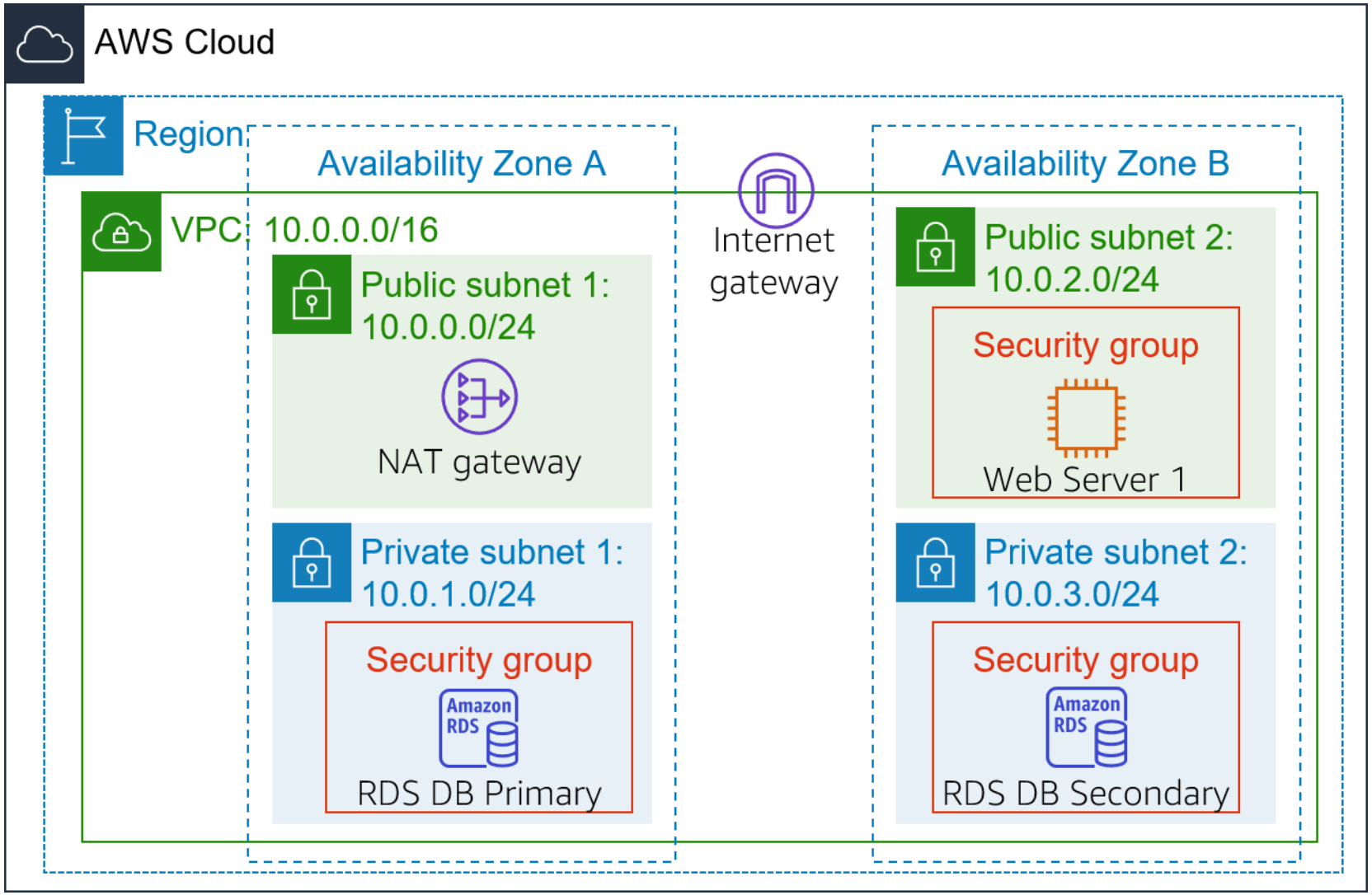
Lab 6: Scale and Load Balance Your Architecture

Objectives

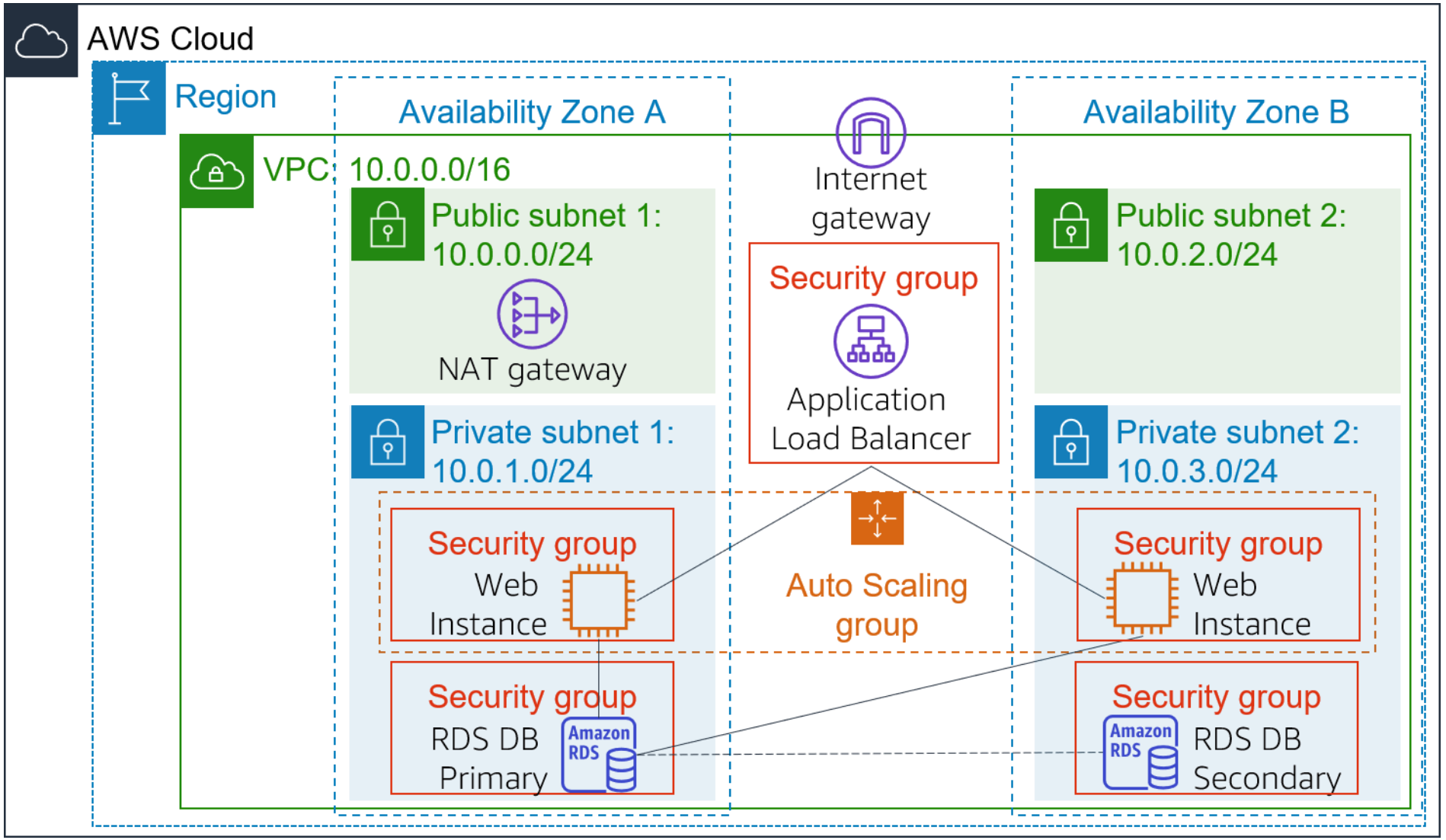
After completing this lab, you can:

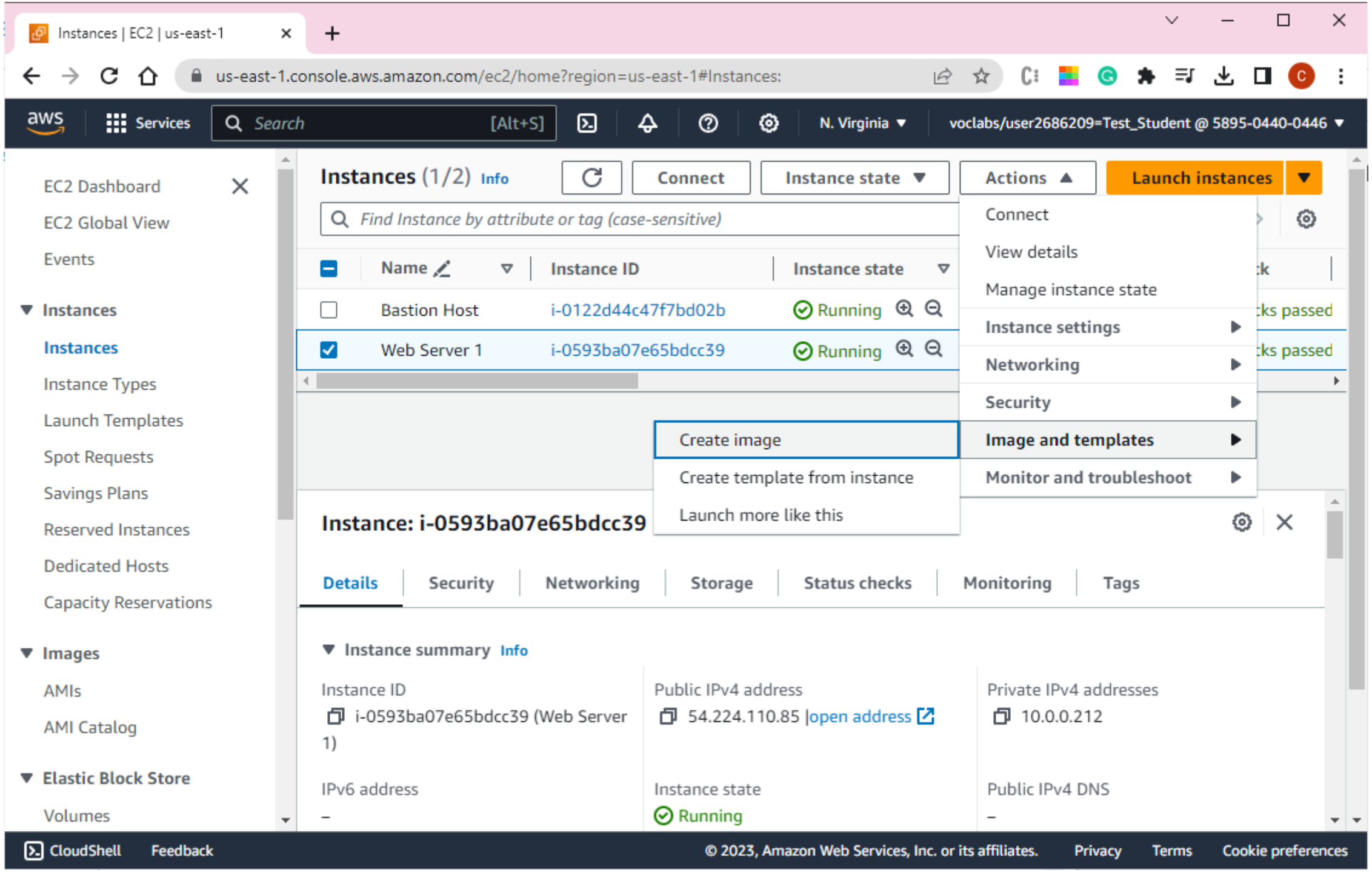
- Create an Amazon Machine Image (AMI) from a running instance.
- Create a load balancer.
- Create a launch template and an Auto Scaling group.
- Automatically scale new instances
- Create Amazon CloudWatch alarms and monitor performance of your infrastructure.

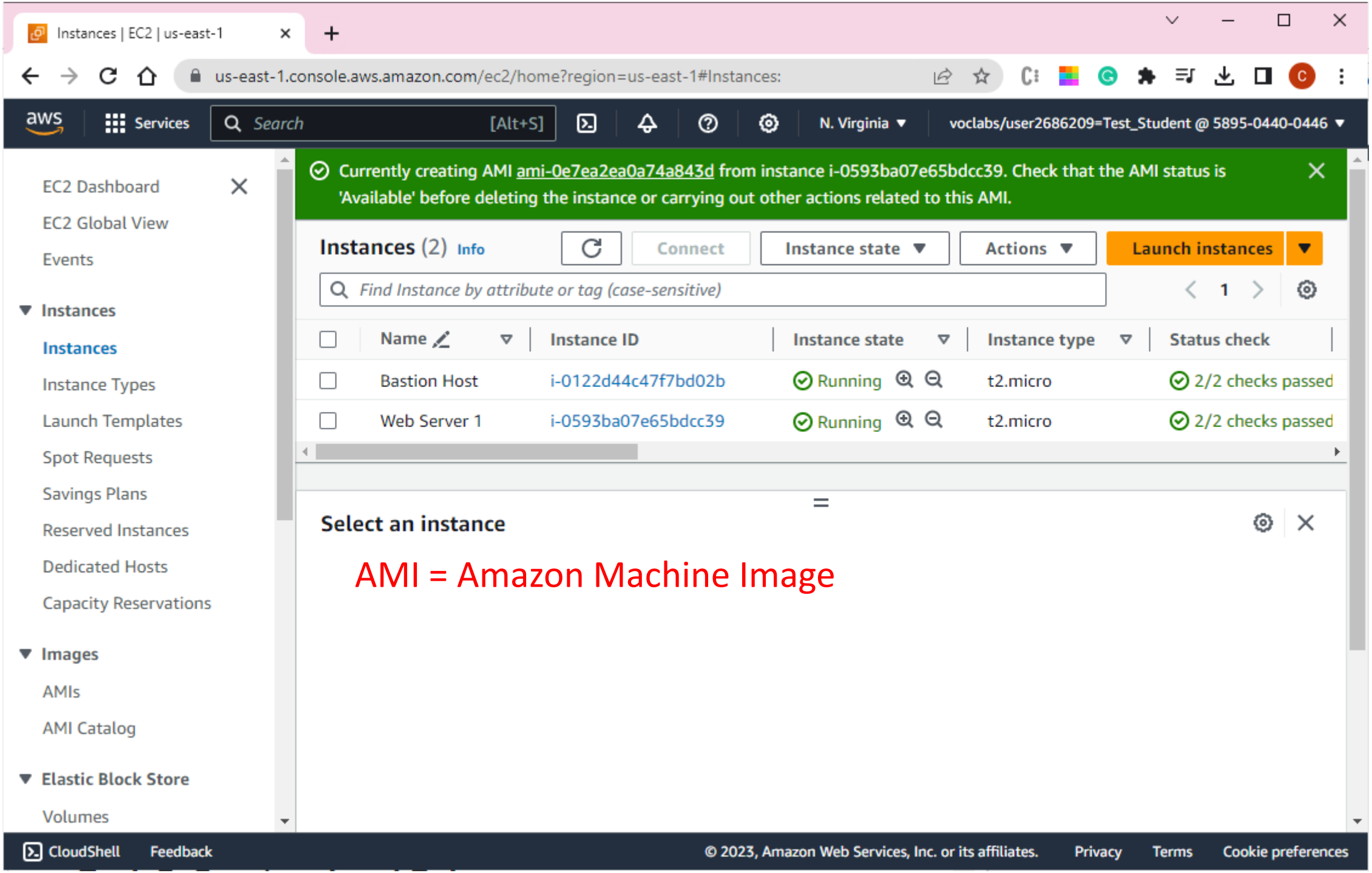
You start with the following infrastructure:



The final state of the infrastructure is:




Instances | EC2 | us-east-1
us-east-1.console.aws.amazon.com/ec2/home?region=us-east-1#Instances:
aws
Services
Search
[Alt+S]
N. Virginia
voclabs/user2686209=Test_Student @ 5895-0440-0446
EC2 Dashboard
EC2 Global View
Events
Instances
Instances
Instance Types
Launch Templates
Spot Requests
Savings Plans
Reserved Instances
Dedicated Hosts
Capacity Reservations
Images
AMIs
AMI Catalog
Elastic Block Store
Volumes
Instances (1/2) Info
Connect
Instance state
Actions
Launch instances
Find Instance by attribute or tag (case-sensitive)
Name
Instance ID
Instance state
Bastion Host
i-0122d44c47f7bd02b
Running
Web Server 1
i-0593ba07e65bdcc39
Running
Connect
View details
Manage instance state
Instance settings
Networking
Security
Image and templates
Monitor and troubleshoot
Create image
Create template from instance
Launch more like this
Instance: i-0593ba07e65bdcc39
Details
Security
Networking
Storage
Status checks
Monitoring
Tags
Instance summary Info
Instance ID
i-0593ba07e65bdcc39 (Web Server 1)
Public IPv4 address
54.224.110.85 |open address
Private IPv4 addresses
10.0.0.212
IPv6 address
–
Instance state
Running
Public IPv4 DNS
–
CloudShell
Feedback
© 2023, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

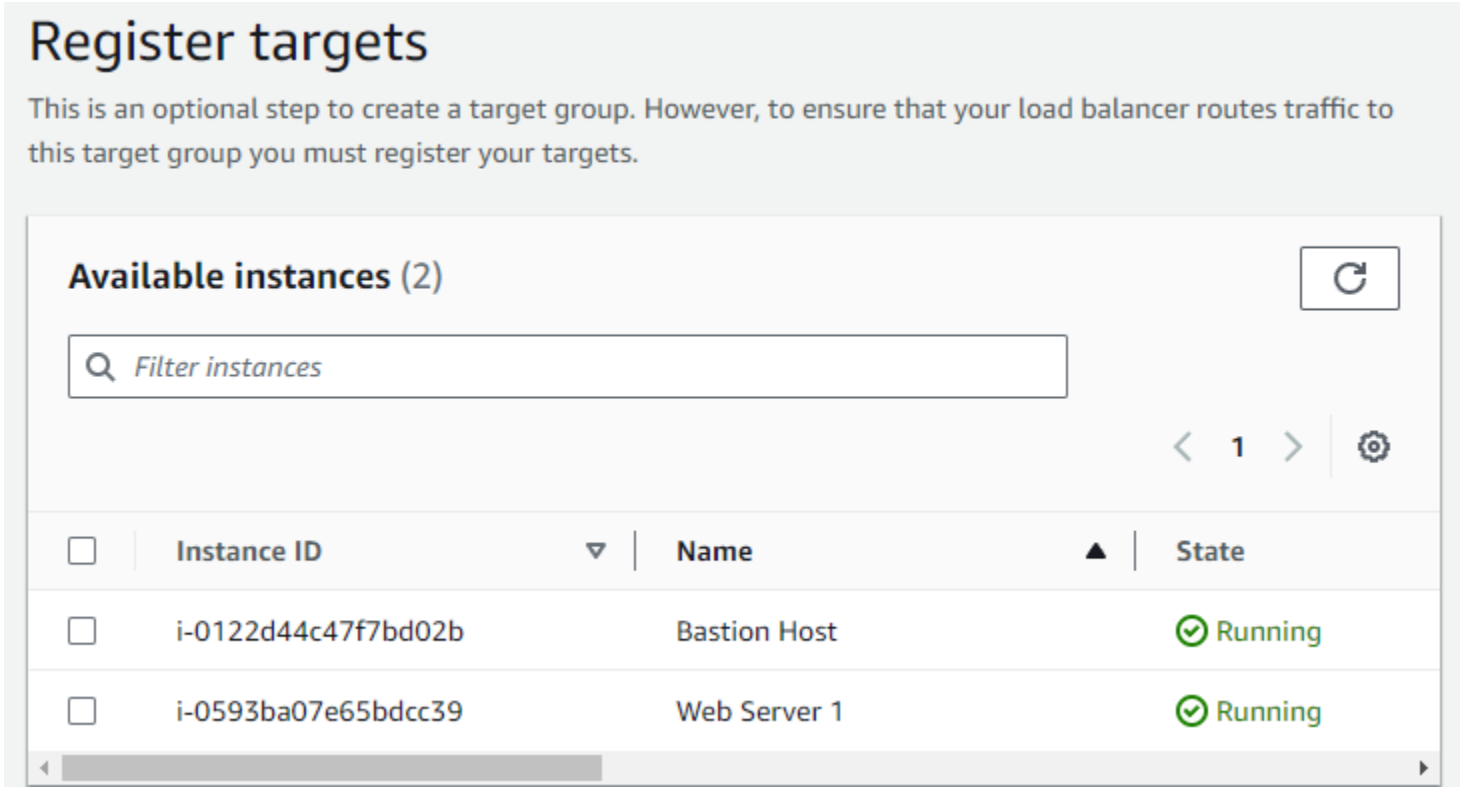

Instances | EC2 | us-east-1
us-east-1.console.aws.amazon.com/ec2/home?region=us-east-1#Instances:
aws
Services
Search
[Alt+S]
N. Virginia
voclabs/user2686209=Test_Student @ 5895-0440-0446
EC2 Dashboard
EC2 Global View
Events
Instances
Instances
Instance Types
Launch Templates
Spot Requests
Savings Plans
Reserved Instances
Dedicated Hosts
Capacity Reservations
Images
AMIs
AMI Catalog
Elastic Block Store
Volumes
Currently creating AMI ami-0e7ea2ea0a74a843d from instance i-0593ba07e65bdcc39. Check that the AMI status is 'Available' before deleting the instance or carrying out other actions related to this AMI.
Instances (2) Info
Connect
Instance state
Actions
Launch instances
Find Instance by attribute or tag (case-sensitive)
1
Name
Instance ID
Instance state
Instance type
Status check
Bastion Host
i-0122d44c47f7bd02b
Running
t2.micro
2/2 checks passed
Web Server 1
i-0593ba07e65bdcc39
Running
t2.micro
2/2 checks passed
Select an instance
AMI = Amazon Machine Image
CloudShell
Feedback
© 2023, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

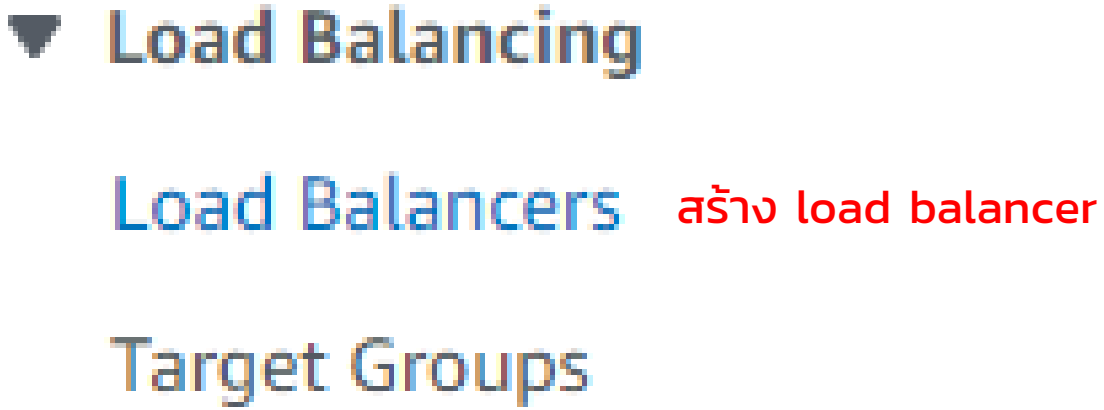
Load Balancing

Load Balancers

Target Groups สร้าง target groups ก่อน

ขั้นนี้เป็น optional ไม่ได้ทำอะไร
(สามารถเลือก target ที่เป็น instance ที่กำลังทำงานอยู่ได้ แต่เราไม่เลือก
จะใช้ target เป็น auto scaling group (default: 2, min: 2, max: 6)




Load Balancing
Load Balancers
สร้าง load balancer
Target Groups

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 groups.

Lab VPC
vpc-0da678b1637e82e2e
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-az6)

Subnet

subnet-0b2c811766cbe9394 Public Subnet 1

IPv4 address

Assigned by AWS

us-east-1b (use1-az1)

Subnet

subnet-0ac709d9e1be79255 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-0b427b214ceb899ad VPC: vpc-0da678b1637e82e2e

Listeners and routing Info

A listener is a process that checks for connection requests using the port and protocol you configure. The rules that you define for a listener determine how the load balancer routes requests to its registered targets.

▼ Listener **HTTP:80** Remove

Protocol: HTTP : Port: 80 (1-65535)

Default action Info: Forward to LabGroup target group ที่สร้างไว้ HTTP
Target type: Instance, IPv4

Create target group

Listener tags - *optional*

Consider adding tags to your listener. Tags enable you to categorize your AWS resources so you can more easily manage them.

Add listener tag

You can add up to 50 more tags.

▼ **Instances**

Instances

Instance Types

Launch Templates เลือก AMI ที่สร้างไว้

Spot Requests

Savings Plans

Reserved Instances

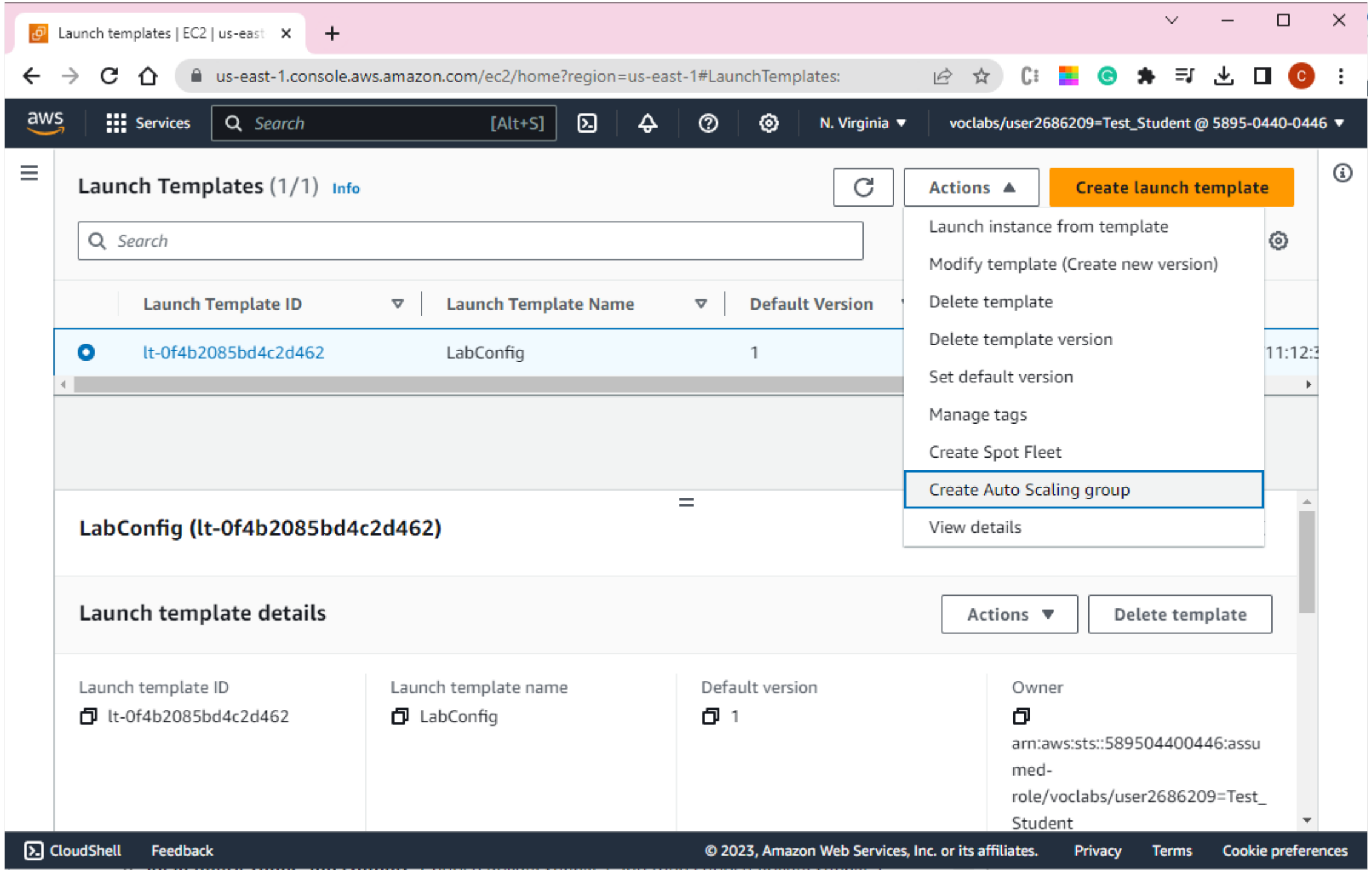
Dedicated Hosts

Capacity Reservations

Detailed CloudWatch monitoring | Info ติด cloudwatch ไว้ด้วย

Enable ▼

Additional charges apply


Launch templates | EC2 | us-east
us-east-1.console.aws.amazon.com/ec2/home?region=us-east-1#LaunchTemplates:
aws
Services
Search
[Alt+S]
N. Virginia
voclabs/user2686209=Test_Student @ 5895-0440-0446
Launch Templates (1/1) Info
Actions
Create launch template
Launch instance from template
Modify template (Create new version)
Delete template
Delete template version
Set default version
Manage tags
Create Spot Fleet
Create Auto Scaling group
View details
Search
Launch Template ID
Launch Template Name
Default Version
lt-0f4b2085bd4c2d462
LabConfig
1
LabConfig (lt-0f4b2085bd4c2d462)
Launch template details
Actions
Delete template
Launch template ID
lt-0f4b2085bd4c2d462
Launch template name
LabConfig
Default version
1
Owner
arn:aws:sts::589504400446:assumed-role/voclabs/user2686209=Test_Student
CloudShell
Feedback
© 2023, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

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 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 ▼ ⟳

LabGroup | HTTP ✕
Application Load Balancer: LabELB

Group size - *optional* Info

Specify the size of the Auto Scaling group by changing the desired capacity. You can also specify minimum and maximum capacity limits. Your desired capacity must be within the limit range.

Desired capacity

2

Minimum capacity

2

Maximum capacity

6

2 instances (limit ระหว่าง 2 ถึง 6)

Scaling policies - *optional*

Choose whether to use a scaling policy to dynamically resize your Auto Scaling group to meet changes in demand. Info

- (●) **Target tracking scaling policy**
  Choose a desired outcome and leave it to the scaling policy to add and remove capacity as needed to achieve that outcome.
- ( ) **None**

Scaling policy name

LabScalingPolicy

Metric type Info

Monitored metric that determines if resource utilization is too low or high. If using EC2 metrics, consider enabling detailed monitoring for better scaling performance.

Average CPU utilization

Target value

60

Instance warmup Info

300 seconds

- **Target value:** 60 **มีการแก้ไขเป็น 50%**

  This tells Auto Scaling to maintain an *average* CPU utilization *across all instances* at 60%. Auto Scaling will automatically add or remove capacity as required to keep the metric at, or close to, the specified target value. It adjusts to fluctuations in the metric due to a fluctuating load pattern.

☐ Disable scale in to create only a scale-out policy

Alarms 0 0 2

In alarm

All alarms

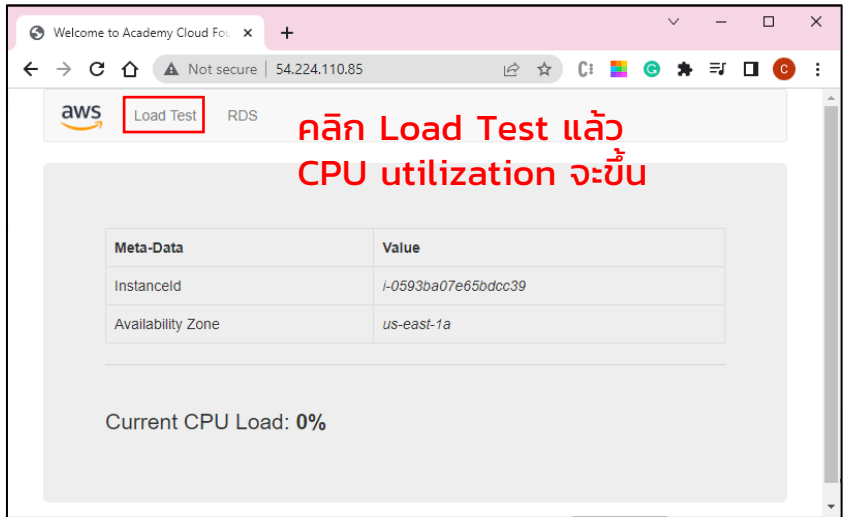
Billing

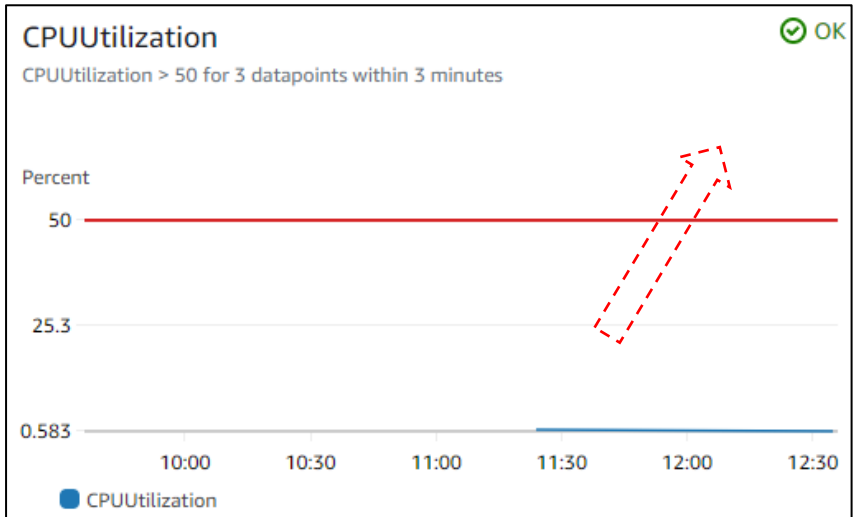
Alarms (2)

Hide Auto Scaling alarms | Clear selection | Create composite alarm

Search | Any state | Any type | Any actions ...

Name	State	Last state update	Conditions	Actions
TargetTracking-Lab Auto Scaling Group-AlarmLow-64608303-cc73-421f-a5b8-278197dd0f0c	In alarm	2023-10-28 11:40:10	CPUUtilization < 35 for 15 datapoints within 15 minutes	Actions enabled
TargetTracking-Lab Auto Scaling Group-AlarmHigh-5397d2f1-a8ff-4b8b-9606-3e2c78d95bb4	OK	2023-10-28 11:30:40	CPUUtilization > 50 for 3 datapoints within 3 minutes	Actions enabled





CPU utilization ต้องขึ้นเกิน 50% ถึงจะเพิ่ม instance

60. In the search box next to **Services**, search for and select **EC2**.
61. In the left navigation pane, choose **Instances**.

    More than two instances labeled **Lab Instance** should now be running. The new instance(s) were created by Auto Scaling in response to the CloudWatch alarm.

จะเห็น instance มากกว่า 2 (scale up)

Task 6: Terminate Web Server 1

In this task, you will terminate *Web Server 1*. This instance was used to create the AMI used by your Auto Scaling group, but it is no longer needed.

62. Select ☑ **Web Server 1** (and ensure it is the only instance selected).
63. In the **Instance state** menu, choose **Instance State** > **Terminate Instance**.
64. Choose **Terminate**

ทำลาย instance ที่เป็นต้นฉบับทิ้งไป เนื่องจากเปลืองค่าใช้จ่าย
Load Balancer ไม่ได้ส่ง traffic ไปให้ เพราะไม่ได้ register