Assignment 4 – ALB and ASG

1. Run 2 web servers behind ALB.
   a. Create an SG for the ALB that allows access from the internet on port 80. Give a meaningful name like ALB. Because it will be used in the servers' SG.
   b. Create an SG for an EC2 instance. Open up port 80 for the ALB SG.
   c. Create 2 servers in 2 different AZs (us-east-1a, us-east-1b) with different HTML content. Refer to the code snippet below and put it into "User Data". Hit Edit in the "Network Settings" and select us-east-1a in instance1 and us-east1b in instance2. Select the SG for the webserver you created in the previous step.
   d. Create ALB. Select 2 AZs (us-east-1a, us-east-1b) for HA (High Availability). Create the TG and register the servers. And select the TG you created.

2. Run web servers behind NLB.

NLB operates at layer 4. There is no SG. How you are going to whitelist access only from NLB in your EC2 SG? The solution is to use elastic IPs in the NLB nodes. So we can whitelist those static IPs in the EC2 SG. We whitelist the private IPs of the elastic IP. Because all traffic inside the same network (VPC) uses private IPs.

   a. Create 2 elastic IPs.
   b. Create NLB and associate the nodes in AZ with the elastic IPs.
   c. Add the instances in us-east-1a, and us-east-1b that you created in task 1 to the target group of the NLB. Protocol must be **TCP,** not HTTP.
   d. Once NLB is being provisioned, you will find private IPs in the Elastic IP. Grab those IPs, and allow it in in the EC2 SG.
   e. Update the target group and deselect **Preserve client IP addresses.**
   f. Grab the private IP of the elastic IP in the VPC elastic IP section. Update the private instance's security group to allow **private IPs of the elastic IPs** that you created in the first step.

3. Run the web server behind the ALB in ASG.
   a. Deregister instances behind the ALB. We will register them inside ASG so it can scale automatically.
   b. Create a launch configuration. Launch template is recommended. You can use that instead.
      i. Give it a name
      ii. Select the Amazon Linux AMI. You can find the AMI ID from EC2 creation wizard.
      iii. Select instance type, t2.micro.
      iv. Expand advanced. Select IAM profile.
      v. Enter the User Data below.
      vi. Select the web server's SG. Created in the task 1.
      vii. Select any key pair. It doesn't matter. Because we use Session Manager to SSH into the instance.
   c. Create the Auto Scaling Group.

i. Select launch template/configration.
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 health checks panel.
v. Desired capacity is 2, min is 1, and max is 3. Set a target tracking scale with a target value of 50.
d. Mimic the high CPU utilization with a library to test scaling out.

Extra:

- Establish a secure connection between EC2 and ELB.

Refer:

#!/bin/bash

yum install httpd -y

cd /var/www/html

echo '<p>Hello from Cloud Computing</p>' > index.html

systemctl start httpd.service

systemctl enable httpd.service

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

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

Security group name	Security group ID	Description	VPC ID
my-lab-alb-sg	sg-03e5e025e377518eb	Lab Application Load Balancer Security Group	vpc-0b978358e22761686

Owner	Inbound rules count	Outbound rules count	
409673912482	1 Permission entry	1 Permission entry	

Inbound rules   Outbound rules   Tags

Inbound rules (1/1)

Manage tags   Edit inbound rules

Filter security group rules   < 1 >

Type	▽	Protocol	▽	Port range	▽	Source	▽	Description
HTTP		TCP		80		0.0.0.0/0		–

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

Security group name	Security group ID	Description	VPC ID
my-lab-EC2-Server-sg	sg-0a370c15c5b405b61	Web Server Security Group	vpc-0b978358e22761686

Owner	Inbound rules count	Outbound rules count	
409673912482	1 Permission entry	1 Permission entry	

Inbound rules   Outbound rules   Tags

Inbound rules (1/1)

my-lab-alb-sg Security Group

Manage tags   Edit inbound rules

Filter security group rules   < 1 >

Type	▽	Protocol	▽	Port range	▽	Source	▽	Description
HTTP		TCP		80		sg-03e5e025e377518eb		–

Create an ALB

- Create the default http:80 listener.
- Select VPC and public subnets.
- Select the ALB's SG that you created in Task 1.
- Create a target group, type as Instance. Don't register targets for now.

Go to the Load Balancers Display from the EC2 Dashboard

Snapshots

Lifecycle Manager

▼ Network & Security

Security Groups

Elastic IPs

Placement Groups

Key Pairs

Network Interfaces

▼ Load Balancing

Load Balancers

Target Groups New

▼ Auto Scaling

Launch Configurations

Auto Scaling Groups

Resources

You are using the following Amazon EC2 resources in the US East (N. Virginia) Region:

Instances (running)	0	Dedicated Hosts	0
Elastic IPs	0	Instances	0
Key pairs	4	Load balancers	0
Placement groups	0	Security groups	7
Snapshots	0	Volumes	0

1) Click on Load Balancers

e, and deploy Microsoft SQL Server Always On availability groups ✕
WS Launch Wizard for SQL Server. Learn more

...

Create Load Balancer   Actions ▾

2) Click on Create Load Balancer

🔍 Filter by tags and attributes or search by keyword

|◁ ◁ None found ▷ ▷|

☐	Name ▲	DNS name ▾	State ▾	VPC ID ▾	Availability Zones ▾

You do not have any load balancers in this region.

...

Select load balancer type

Elastic Load Balancing supports four types of load balancers: Application Load Balancers, Network Load Balancers, Gateway Load Balancers, and Classic Load Balancers. Choose the load balanc
type that meets your needs.

Learn more about which load balancer is right for you

Application Load Balancer

HTTP
HTTPS

Create

Choose an Application Load Balancer when you need a
flexible feature set for your web applications with HTTP
and HTTPS traffic. Operating at the request level,
Application Load Balancers provide advanced routing
and visibility features targeted at application
architectures, including microservices and containers.

Network Load Balancer

TCP
TLS
UDP

Create

Choose a Network Load Balancer when you need ultra-
high performance, TLS offloading at scale, centralized
certificate deployment, support for UDP, and static IP
connection level, Network Load Balancers are capable
of handling millions of requests per second securely

Gateway Load Balancer

IP

Create

Choose a Gateway Load Balancer when you need to
deploy and manage a fleet of third-party virtual
t support GENEVE. These appliances
mprove security, compliance, and policy
controls.

3) Click on Create Application Load Balancer

...

Step 1: Configure Load Balancer
Basic Configuration

To configure your load balancer, provide a name, select a scheme, specify one or more listeners, and select a network. The default configuration is an Internet-facing load balancer in the selected network with a listener that receives HTTP traffic on port 80.

Name	(i)	my-lab-alb
Scheme	(i)	● internet-facing ○ internal
IP address type	(i)	ipv4

4) Name Load Balancer

Listeners

A listener is a process that checks for connection requests, using the protocol and port that you configured.

Load Balancer Protocol	Load Balancer Port
HTTP	80

Cancel   **Next: Configure Security Settings**

Availability Zones

Specify the Availability Zones to enable for your load balancer. The load balancer routes traffic to the targets in these Availability Zones only. You can specify only one subnet per Availability Zone. You must specify subnets from at least two Availability Zones to increase the availability of your load balancer.

VPC (i)	vpc-0b978358e22761686 (10.0.0.0/16)	my-lab-vpc	
Availability Zones	☑ us-east-1a	subnet-0ef43ef1cfcb561a0 (lab-sn-public-1A)
	IPv4 address (i)	Assigned by AWS
	☑ us-east-1b	subnet-03b7f8298553c4646 (lab-sn-public-1B)
	IPv4 address (i)	Assigned by AWS
	☐ us-east-1c	subnet-028d2f8b4b0b12258 (lab-sn-public-1c)

5) Select VPC

6) Select At Least 2 AZ Zones and Subnets

Additional AWS services can be integrated with this load balancer at launch when you enable them below. You can also add these and other services after your load balancer is created by reviewing the "Integrated Services" tab for the selected load balancer.

AWS Global Accelerator	☐ Create an accelerator to get static IP addresses and improve the performance and availability of your application. Learn more
	Additional charges apply

Your Accelerator will be created with the following name that you can customize. Once your Accelerator is created you can manage it from the Global Accelerator console.

Accelerator name

Maximum 64 characters. Letters and numbers only.

▸ Tags

7) Click Next

Cancel   **Next: Configure Security Settings**

...

Step 2: Configure Security Settings

⚠ **Improve your load balancer's security. Your load balancer is not using any secure listener.**
If your traffic to the load balancer needs to be secure, use the HTTPS protocol for your front-end connection. You can go back to the first step to add/configure secure listeners under Basic Configuration section. You can also continue with current settings.

8) Click Next

Cancel   Previous   **Next: Configure Security Groups**

...

Step 3: Configure Security Groups

A security group is a set of firewall rules that control the traffic to your load balancer. On this page, you can add rules to allow specific traffic to reach your load balancer. First, decide whether to create a new security group or select an existing one.

Assign a security group   ○ Create a **new** security group
● Select an **existing** security group

Filter [VPC security groups ˅]

	Security Group ID	Name	Description	Actions
☐	sg-002d4b487ca1292d2	default	default VPC security group	Copy to new
☐	sg-0d09b0bf676ce516f	launch-wizard-2	launch-wizard-2 created 2021-07-08T19:37:07.572-05:00	Copy to new
☐	sg-0fc356187933ae278	launch-wizard-3	launch-wizard-3 created 2021-07-08T20:58:44.588-05:00	Copy to new
☑	sg-03e5e025e377518eb	my-lab-alb-sg	Lab Application Load Balancer Security Group	Copy to new
☐	sg-0a370c15c5b405b61	my-lab-EC2-Server-sg	Web Server Security Group	Copy to new

7) Select the ALB Security Group you Created

9) Click Next

Cancel   Previous   **Next: Configure Routing**

...

Step 4: Configure Routing

Your load balancer routes requests to the targets in this target group using the protocol and port that you specify here. It also performs health checks on the targets using these settings. The target group you specify in this step will apply to all of the listeners configured on this load balancer. You can edit or add listeners after the load balancer is created.

Target group

Target group ⓘ   [New target group ⬍]

Name ⓘ   [my-lab-target]   10) Name Target Group

Target type   ● Instance   11) Select Instance
○ IP
○ Lambda function

Protocol ⓘ   [HTTP ⬍]

Port ⓘ   [80]

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

Health checks

Protocol ⓘ	HTTP ⬦
Path ⓘ	/

▸ Advanced health check settings

12) Click Next

Cancel　Previous　**Next: Register Targets**

...

1. Configure Load Balancer　2. Configure Security Settings　3. Configure Security Groups　4. Configure Routing　**5. Register Targets**　6. Review

Step 5: Register Targets

Register targets with your target group. If you register a target in an enabled Availability Zone, the load balancer starts routing requests to the targets as soon as the registration process completes and the target passes the initial health checks.

Registered targets

To deregister instances, select one or more registered instances and then click Remove.

Remove

☐	Instance ▾	Name ▾	Port ▾	State ▾	Security groups ▾	Zone ▾
			No instances available.			

Instances

To register additional instances, select one or more running instances, specify a port, and then click Add. The default port is the port specified for the target group. If the instance is already registered on the specified port, you must specify a different port.

Add to registered　on port 80

13) Click Next

Cancel　Previous　**Next: Review**

...

1. Configure Load Balancer　2. Configure Security Settings　3. Configure Security Groups　4. Configure Routing　5. Register Targets　**6. Review**

Step 6: Review

Please review the load balancer details before continuing

▾ Load balancer　　　　　　　　　　　　　　　　　　　　　　　　　　　Edit

Name	my-lab-alb
Scheme	internet-facing
Listeners	Port:80 - Protocol:HTTP
IP address type	ipv4
VPC	vpc-0b978358e22761686 (my-lab-vpc)
Subnets	subnet-0ef43ef1cfcb561a0 (lab-sn-public-1A), subnet-03b7f8298553c4646 (lab-sn-public-1B)
Tags	

▾ Security groups　　　　　　　　　　　　　　　　　　　　　　　　　　Edit

Security groups	sg-03e5e025e377518eb

▾ Routing

14) Click Create　　　　　　　Edit

Target group	New target group
Target group name	my-lab-target

Cancel　Previous　**Create**

Create an NLB

Network Load Balancer Info
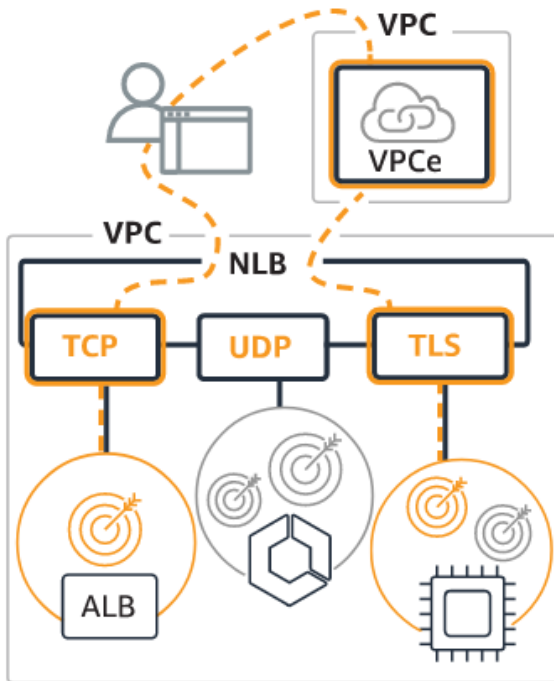
Choose a Network Load Balancer when you need ultra-high performance, TLS offloading at scale, centralized certificate deployment, support for UDP, and static IP addresses for your applications. Operating at the connection level, Network Load Balancers are capable of handling millions of requests per second securely while maintaining ultra-low latencies.

Create

a. Create 2 elastic IPs.

Elastic IP addresses (3)

	Name	Allocated IPv4 addr... ▽	Type	▽	Allocation ID	▽	Reverse DNS record	▽	As
☐	–	18.214.114.143	Public IP		eipalloc-0afb80388042162a0		–		–
☐	–	184.73.25.52	Public IP		eipalloc-06bbcf438c47f081b		–		–
☐	–	34.199.105.196	Public IP		eipalloc-07f6064d4990da144		–		–

b. Create NLB and associate it with the elastic IPs.

☑	my-first-nlb	my-first-nlb-5e570165c040b...	Active	vpc-0def861cf2ef04f24	us-east-1b, us-east-1a	network
☐	MyfirstALB	MyfirstALB-1400061722.us-...	Active	vpc-0def861cf2ef04f24	us-east-1b, us-east-1a	application

IP address type	ipv4
	Edit IP address type
VPC	vpc-0def861cf2ef04f24
Availability Zones	subnet-052ca4d8e64974a36 - us-east-1b
	IPv4 address: Elastic IP 34.199.105.196
	subnet-0976c37ff2dbe243f - us-east-1a
	IPv4 address: Elastic IP 184.73.25.52
	Edit subnets
Hosted zone	Z26RNL4JYFTOTI
Creation time	April 29, 2022 at 10:28:07 PM UTC-5

c. Spin up 2 instances with different HTML content in us-east-1a, us-east-1b AZs.

	Name	▽	Instance ID	Instance state	▽	Instance type	▽	Status check	Alarm status		Availability Zone	▽
☐	–		i-0452e560715ee832a	⊘ Running	⊕⊖	t2.micro		⊘ 2/2 checks passed	No alarms	+	us-east-1a	
☑	myNlbE1		i-03f9925915e9136a6	⊘ Running	⊕⊖	t2.micro		⏱ Initializing	No alarms	+	us-east-1a	
☑	myNLBE2		i-009a5fdf725c0d8d2	⊘ Running	⊕⊖	t2.micro		–	No alarms	+	us-east-1b	

d. Add the private instance in us-east-1a, us-east-1b to the target group of the NLB.

Target group name

mySgNlb

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

Protocol   **Port**

TCP ▼ : 80 ↕

VPC
Select the VPC with the instances that you want to include in the target group.

my-first-vpc
vpc-0def861cf2ef04f24
IPv4: 10.0.0.0/16   ▼

Register targets

This is an optional step to create a target group. However, to ensure that your load balancer routes traffic to this target group you must register your targets.

Available instances (2/3)

🔍 Filter resources by property or value   ‹ 1 › ⚙

	Instance ID	▽	Name	▽	State	▽	Security groups	Zone	▽	Subnet ID
☐	i-0452e560715ee832a				⊘ running		MyAppBehindAlb	us-east-1a		subnet-0b5aeb0697c77d6a5
☑	i-03f9925915e9136a6		myNlbE1		⊘ running		default	us-east-1a		subnet-0b5aeb0697c77d6a5
☑	i-009a5fdf725c0d8d2		myNLBE2		⊘ running		default	us-east-1b		subnet-0d2438927d9c21121

e. Update the target group and deselect Preserve client IP addresses

Edit attributes

Attributes

Restore defaults

Deregistration delay
The time to wait for in-flight requests to complete while deregistering a target. During this time, the state of the target is draining.

| 300 | ⌄ | seconds |

0-3600

☐ **Connection termination on deregistration** — *recommended*
If enabled, your Network Load Balancer will terminate active connections when deregistration delay is reached.

☐ **Stickiness**
The type of stickiness associated with this target group. If enabled, the load balancer binds a client's session to a specific instance within the target group.

☐ **Proxy protocol v2**
Before you enable proxy protocol v2, make sure that your application targets can process proxy protocol headers otherwise your application might break.

☐ **Preserve client IP addresses**
Preserve client IP addresses and ports in the packets forwarded to targets.

f. Grab private ip of the elastic ip in the AWS console, VPC, elastic ip section. Update the private instance's security group to allow private IPs of the elastic IPs that you created in the first step.

Instance: i-03f9925915e9136a6 (myNlbE1)

Instance ID	Public IPv4 address	⊘ Private IPv4 address copied
⧉ i-03f9925915e9136a6 (myNlbE1)	–	⧉ 10.0.3.215
IPv6 address	Instance state	Public IPv4 DNS
–	⊘ Running	–
Hostname type	Private IP DNS name (IPv4 only)	Answer private resource DNS name
IP name: ip-10-0-3-215.ec2.internal	⧉ ip-10-0-3-215.ec2.internal	IPv4 (A)
Instance type	Elastic IP addresses	Auto-assigned IP address
t2.micro	–	–

Inbound rules (4)

p rule... ▽	IP version ▽	Type ▽	Protocol ▽	Port range ▽	Source ▽	Descriptio
0dca9daad	IPv4	HTTP	TCP	80	10.0.4.170/32	–
637663cf	–	SSH	TCP	22	sg-06ea6b051e1d354...	–
a7002a56	–	HTTP	TCP	80	sg-06ea6b051e1d354...	–
7f318d144	IPv4	HTTP	TCP	80	10.0.3.215/32	–

Create a launch template

- Select the web server's SG created in Task 1.
- Expand advanced and enter the user data below.

Go to Launch Templates Display from EC2 Display

Create Launch Template

...

Create 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. Templates can have multiple versions.

Launch template name and description

2) Name Template

Launch template name - *required*

```
my-lab-server
```

Must be unique to this account. Max 128 chars. No spaces or special characters like '&', '*', '@'.

Template version description

```
A prod webserver for MyApp
```

Max 255 chars

Selecting Guidance will show more Details

Auto Scaling guidance   **Info**
Select this if you intend to use this template with EC2 Auto Scaling

☑ Provide guidance to help me set up a template that I can use with EC2 Auto Scaling

▶ Template tags

▶ Source template

Launch template contents

Specify the details of your launch template below. Leaving a field blank will result in the field not being included in the launch template.

▼ **Amazon machine image (AMI) - required** Info

3) Select AMI

AMI - *required*

Amazon Linux 2 AMI (HVM), SSD Volume Type
ami-0dc2d3e4c0f9ebd18
Catalog: Quick Start   virtualization: hvm   architecture: 64-bit (x86)

▼ **Instance type** Info

4) Select Instance Type

Instance type

t2.micro
Family: t2   1 vCPU   1 GiB Memory
On-Demand Linux pricing: 0.0116 USD per Hour
On-Demand Windows pricing: 0.0162 USD per Hour

Free tier eligible

Compare instance types

▼ Key pair (login)   Info

You can use a key pair to securely connect to your instance. Ensure that you have access to the selected key pair before you launch the instance.

Key pair name

| Don't include in launch template | ▼ |

Create new key pair

▼ Network settings

Networking platform   Info

| ● Virtual Private Cloud (VPC)
Launch into a virtual network in your own logically isolated area within the AWS Cloud | ○ EC2-Classic
Launch into a single flat network that you share with other customers. |

Security groups

| Select security groups | ▼ |

| my-lab-EC2-Server-sg   sg-0a370c15c5b405b61   ✕
VPC: vpc-0b978358e22761686 |

5) Select Server security Group you Created

▼ Storage (volumes)   Info

▶ Volume 1 (AMI Root) (8 GiB, EBS, General purpose SSD (gp2))
AMI Volumes are not included in the template unless modified

Add new volume

▼ Resource tags   Info

No resource tags are currently included in this template. Add a resource tag to include it in the launch template.

Add tag

50 remaining (Up to 50 tags maximum)

Network interfaces Info

No network interfaces are currently included in this template. Add a network interface to include it in the launch template.

Add network interface

Advanced details Info

#!/bin/bash --> tells the OS to invoke the specified shell to execute the script commands

Create Auto Scaling Group

- Select the launch template.
- Select the custom VPC and public subnets.
- Click on attach to an existing load balancer and select the default TG (Target Group) of the ALB.
- Desired capacity 2, min is 1, max is 3. Set target tracking scaling policy on the CPU utilization with any value from 10 to 90. AWS AutoScaling tries to maintain the CPU utilization at that level. If it exceeds, scales outs.

Go to Auto Scaling Display from EC2 Display

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

Choose launch template or configuration Info

Specify a launch template that contains settings common to all EC2 instances that are launched by this Auto Scaling group. If you currently use launch configurations, you might consider migrating to launch templates.

Name

Auto Scaling group name
Enter a name to identify the group.

my-lab-as-group

2) Name Group

Must be unique to this account in the current Region and no more than 255 characters.

Launch template Info   Switch to launch configuration

Launch template
Choose a launch template that contains the instance-level settings, such as the Amazon Machine Image (AMI), instance type, key pair, and security groups.

my-lab-server

3) Select Your Launch Template

Create a launch template ↗

Version

Default (1)

Create a launch template version ↗

Description	Launch template	Instance type
-	my-lab-server ↗ lt-0ca69b78349a3fb83	t2.micro
AMI ID ami-0dc2d3e4c0f9ebd18	**Security groups** -	**Request Spot Instances** No
Key pair name -	**Security group IDs** sg-0a370c15c5b405b61 ↗	

Additional details

Storage (volumes)	Date created
-	Sun Jul 11 2021 11:24:24 GMT-0500 (Central Daylight Time)

4) Click Next

Cancel   Next

...

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   6) Select Some Subnets

us-east-1b | subnet-03b7f8298553c4646 (lab-sn-public-1B)   ✕
10.0.2.0/24   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**

...

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 group size and scaling policies Info

Set the desired, minimum, and maximum capacity of your Auto Scaling group. You can optionally add a scaling policy to dynamically scale the number of instances in the group.

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
```
1
```

11) Set Desired, Min, and Max Capacity

Maximum capacity
```
3
```

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
```
Target Tracking Policy
```

Metric type
```
Average CPU utilization                    ▼
```

Target value
```
50
```

12) Set Target Tracking for CPU Utilization

Instances need
```
300
```
seconds warm up before including in metric

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

Instance scale-in protection - *optional*

Instance scale-in protection
If protect from scale in is enabled, newly launched instances will be protected from scale in by default.

☐ Enable instance scale-in protection

13) Click Next

Cancel   Previous   Skip to review   Next

...

Step 1
Choose launch template or configuration

Step 2
Configure settings

Step 3 *(optional)*
Configure advanced options

Add notifications Info

Send notifications to SNS topics whenever Amazon EC2 Auto Scaling launches or terminates the EC2 instances in your Auto Scaling group.

Add notification

14) Click Next

Cancel   Previous   Skip to review   **Next**

...

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

Add tags Info

Add tags to help you search, filter, and track your Auto Scaling group across AWS. You can also choose to automatically add these tags to instances when they are launched.

> ⓘ You can optionally choose to add tags to instances (and their attached EBS volumes) by specifying tags in your launch template. We recommend caution, however, because the tag values for instances from your launch template will be overridden if there are any duplicate keys specified for the Auto Scaling group.   ✕

Tags (0)

Add tag
50 remaining

15) Click Next

Cancel   Previous   **Next**

...

EC2 > Auto Scaling groups > Create Auto Scaling group

Step 1
Choose launch template or configuration

Review Info

16) Click Next

... Review What you have just Selected

Cancel   **Create Auto Scaling group**

Verify and Test the ALB
View the Health Check on your the Target Group Details. Both Instances Should be Healthy

my-lab-target

[Delete]

⊡ arn:aws:elasticloadbalancing:us-east-1:409673912482:targetgroup/my-lab-target/785ca90756d47acd

---

Details

Target type	Protocol : Port	Protocol version	VPC
Instance	HTTP: 80	HTTP1	vpc-0b978358e22761686 ↗

Load balancer
my-lab-alb ↗

Total targets	Healthy	Unhealthy	Unused	Initial	Draining
2	⊘ 2	⊗ 0	⊙ 0	⊕ 0	⊖ 0

Targets | Monitoring | Health checks | Attributes | Tags

Registered targets (2)

[↻] [Deregister] [Register targets]

🔍 Filter resources by property or value   ‹ 1 › ⚙

☐	Instance ID ▽	Name ▽	Port ▽	Zone ▽	Health status ▽	Health status details
☐	i-0179bc9cdca16967e		80	us-east-1b	⊘ healthy	
☐	i-0f05d00a3423df3ad		80	us-east-1a	⊘ healthy	

Create Load Balancer   **Actions ⌄**

Q Filter by tags and attributes or search by keyword   **DNS** |< < 1 to 1 of 1 > >|

	Name	DNS name	State	VPC ID
☐	my-lab-alb	my-lab-alb-613824474.us-east-1.elb.amazonaws.com	Active	vpc-0b978358e22761686

Test DNS with Web Browser

⊗ my-lab-alb-613824474.us-east-1. ✕ +

← → C ⚠ Not secure | my-lab-alb-613824474.us-east-1.elb.amazonaws.com

⠿ Apps G Gmail ▶ YouTube ◉ Maps ⊗ Student Portal Login ᴜ Udemy G Google
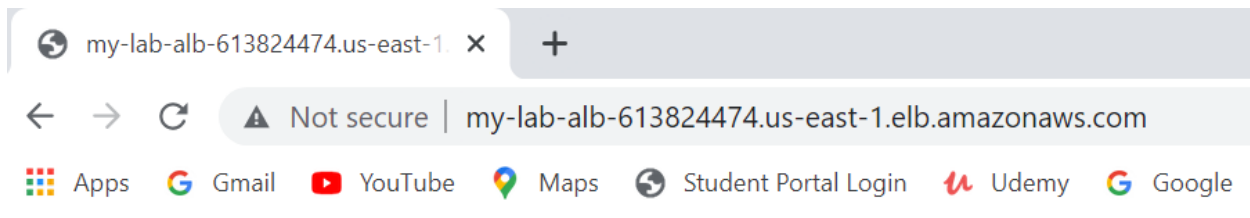
Hello from my EC2 Instance in Autoscaling Group Behind an ALB

EC2 stress tool

1-select the EC2 instance you want to install the stress tool: we can use the instance we have during the ASG class.

install stress tool using the following commands:

sudo amazon-linux-extras install epel -y

sudo yum install stress -y

```
Dependencies Resolved

Package                    Arch              Version              Repository              Size

Installing:
stress                     x86_64            1.0.4-16.el7         epel                    39 k

Transaction Summary

Install  1 Package

Total download size: 39 k
Installed size: 94 k
Downloading packages:
warning: /var/cache/yum/x86_64/2/epel/packages/stress-1.0.4-16.el7.x86_64.rpm: Header V3 RSA/SHA256 Signature, key ID 352c64e5: NOKEY
Public key for stress-1.0.4-16.el7.x86_64.rpm is not installed
stress-1.0.4-16.el7.x86_64.rpm                                                             | 39 kB  00:00:00
Retrieving key from file:///etc/pki/rpm-gpg/RPM-GPG-KEY-EPEL-7
Importing GPG key 0x352C64E5:
Userid     : "Fedora EPEL (7) <epel@fedoraproject.org>"
Fingerprint: 91e9 7d7c 4a5e 96f1 7f3m 888f 6a2f aea2 352c 64e5
Package    : epel-release-7-11.noarch (@amzn2extra-epel)
From       : /etc/pki/rpm-gpg/RPM-GPG-KEY-EPEL-7
Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
  Installing : stress-1.0.4-16.el7.x86_64                                                          1/1
  Verifying  : stress-1.0.4-16.el7.x86_64                                                          1/1

Installed:
  stress.x86_64 0:1.0.4-16.el7

Complete!
```
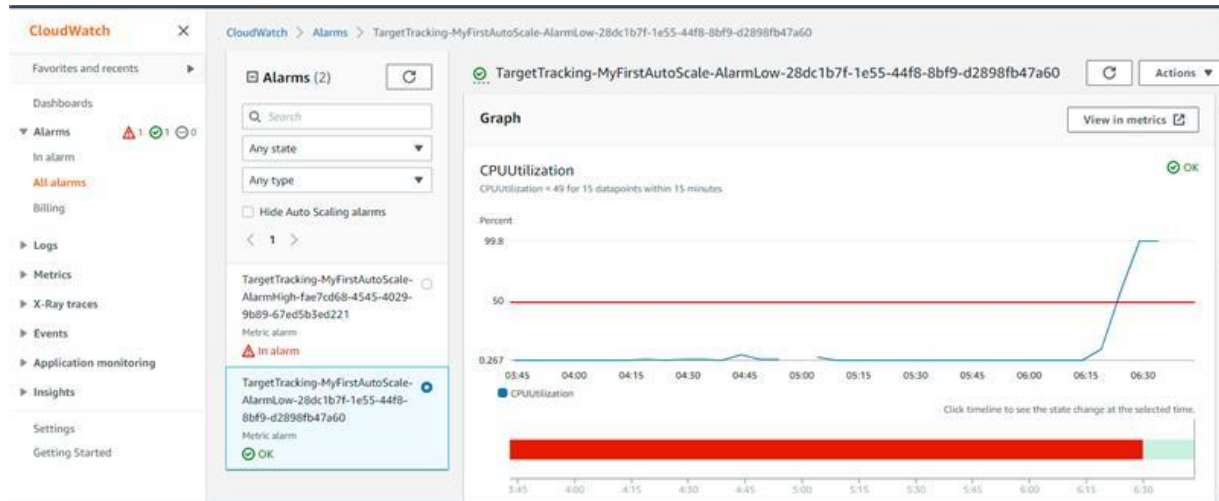
Then to visualize the CPU and memory utilization write the following commands:

sudo stress --cpu 8 --vm-bytes $(awk '/MemAvailable/{printf "%d\n", $2 * 0.9;}' < /proc/meminfo)k --vm-keep -m 1

—cpu

This will spawn 8 CPU workers spinning on a square root task (sqrt(x))

—vm-bytes

This will use 90% of the available memory from /proc/meminfo

—vm-keep

This will re-dirty memory instead of freeing and reallocating.

-m 1

This will spawn 1 worker spinning on malloc()/free()

As time goes on, it will continue to update the graph. To remove the load, press

CTRL-C to stop the stress script.

Reference: https://www.wellarchitectedlabs.com/performance-efficiency/100_labs/100_monitoring_linux_ec2