

Name :Roshan Saral Kumar(CLOUD-COMPUTING-INTERN)
DATE:30-10-2025

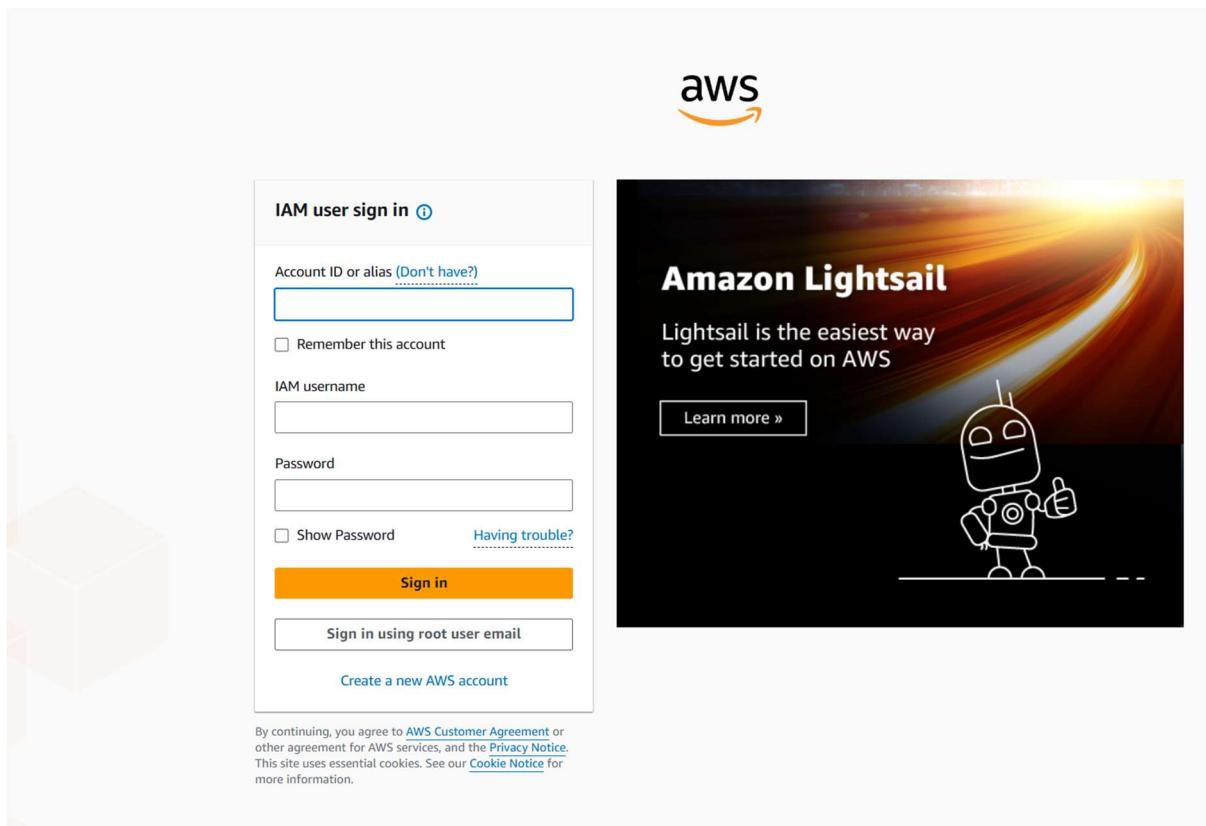
TASK 7

- **Task 7: Host and Deploy a Web Application on the Cloud**
-

- **Objective:- To understand scalability and fault tolerance in cloud computing by setting up a Load Balancer and Auto Scaling group for a simple web application running on multiple virtual machines.**
 - **This helps interns learn how cloud systems handle *high availability, traffic management, and cost-efficient scaling automatically.**
-

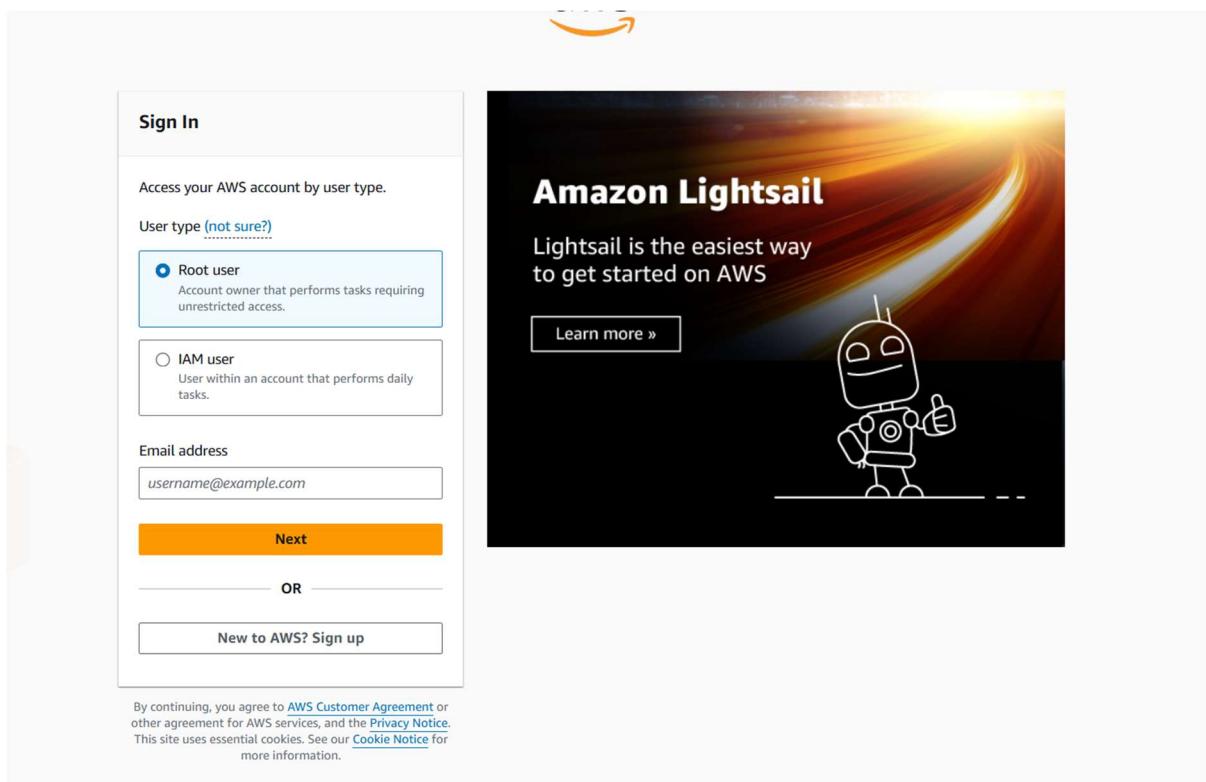
- **TOOLS USED FOR AWS (Free Tier):**
 - **EC2 + Elastic Load Balancer + Auto Scaling**
-

- **THIS IS THE SIGN IN CONSOLE PAGE AND IAM USING SIGN USING ROOT USER EMAIL (ROOT USER MEANS I HAVE ACCESS TO ALL THE SERVICES OF AWS IF IT IS AN IAM (IDENTITY AND ACCESS MANAGEMENT USER THEN THE USER WILL HAVE ACCESS TO ONLY A LIMITED NUMBER OF SERVICES THAT AWS OFFERS AND THE IAM USERS WILL HAVE IAM POLICIES ATTACHED TO THEM SO ONLY THOSE SERVICES THEY WILL BE ABLE TO USE) CURRENTLY LOGGING IN AS ROOT USER EMAIL.**



The image shows the AWS IAM User Sign In page. At the top right is the AWS logo. Below it is a sidebar with the title "IAM user sign in ⓘ". The main form contains fields for "Account ID or alias (Don't have?)", "Remember this account", "IAM username", "Password", and "Show Password". There is also a link "Having trouble?". A large orange "Sign in" button is centered. Below the button are links for "Sign in using root user email" and "Create a new AWS account". At the bottom of the page is a small legal notice about cookie usage.

- AFTER CLICKING ON SIGN IN USING ROOT USER EMAIL I WILL GET THIS PAGE WHERE I HAVE TO ENTER MY EMAIL-ID. AND AS WE CAN SEE ROOT USER IS SELECTED BY DEFAULT. WE CAN ALSO SELECT IAM USER IN THIS PAGE IN CASE IF THERE IS A CHANGE.



The image shows the AWS Sign In page. At the top right is the AWS logo. Below it is a sidebar with the title "Sign In". The main form asks "Access your AWS account by user type." It has two options: "Root user" (selected) and "IAM user". Below these are fields for "Email address" (containing "username@example.com") and a large orange "Next" button. To the right of the form is a sidebar for "Amazon Lightsail" with the text "Lightsail is the easiest way to get started on AWS" and a "Learn more »" button. At the bottom of the page is a small legal notice about cookie usage.

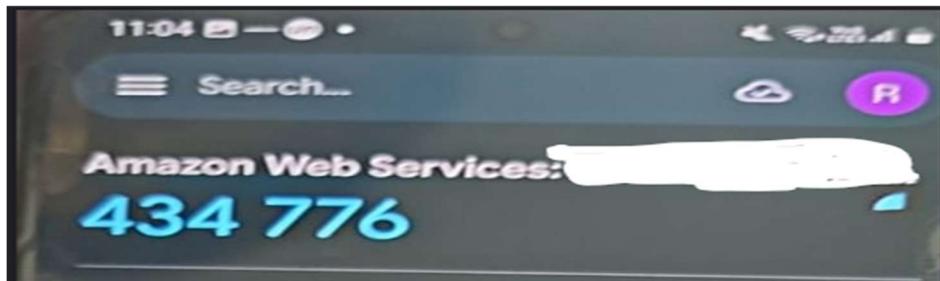
- AFTER ENTERING OUR EMAIL WE NEED TO GIVE OUR PASSWORD

The image shows two side-by-side screenshots. The left screenshot is a 'Root user sign in' interface from AWS. It has a light blue header with the text 'Root user sign in'. Below it is a form with the placeholder 'Enter the password for' followed by '(not you?)'. There is a 'Password' input field containing a single character '|'. Below the input field are two buttons: 'Show password' (unchecked) and 'Forgot password?'. At the bottom are two buttons: a large orange 'Sign in' button and a smaller white 'Sign in to a different account' button. The right screenshot is an 'Amazon Lightsail' landing page. It features a dark background with a bright, glowing orange and yellow swoosh graphic. The text 'Amazon Lightsail' is prominently displayed in white. Below it, the text 'Lightsail is the easiest way to get started on AWS' is shown. There is a white rectangular button with the text 'Learn more »'. To the right of the text is a simple line-art illustration of a robot with a smiling face, one arm raised, and a thumbs-up gesture.

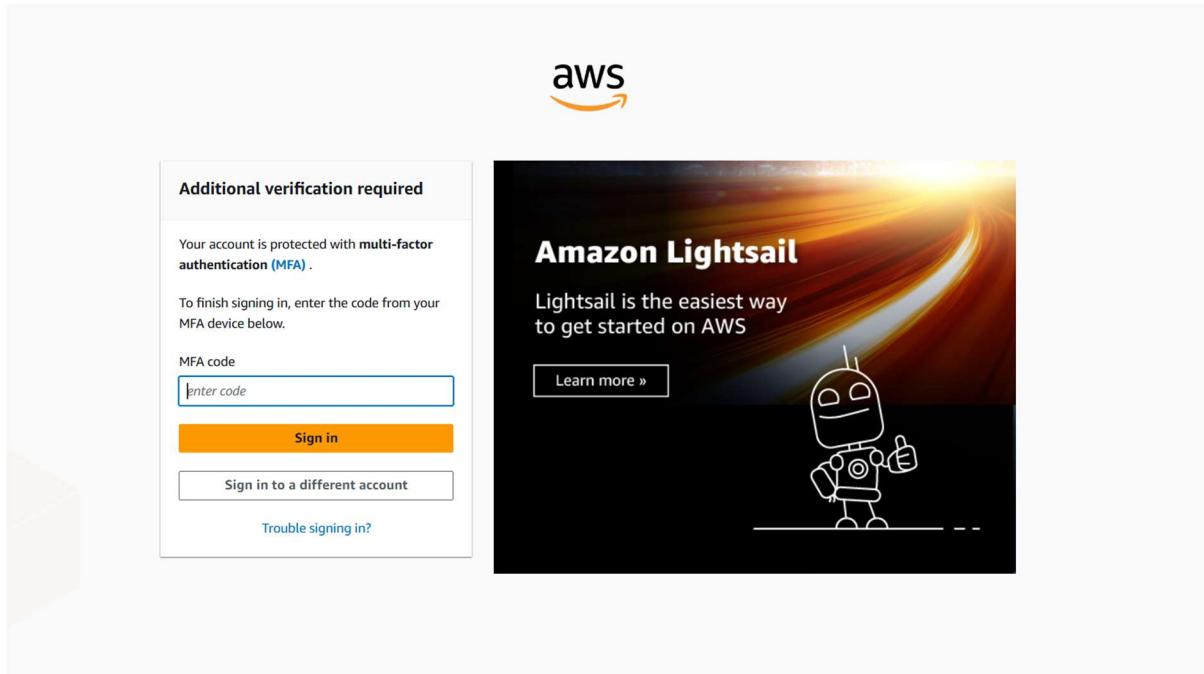
- SINCE I HAVE AN AWS ACCOUNT ALREADY CREATED I ACTUALLY MADE AN MFA FOR MY ROOT USER . MFA STANDS FOR (MULTI-FACTOR AUTHENTICATION CODE) IT ACTS AS A DOUBLE LAYER PROTECTION FOR OUR ROOT USER OTHERWISE WE CAN GET HACKED SO ALREADY PROVIDE AN EXTRA LAYER OF PROTECTION TO BE ON THE SAFER SIDE.
- MFA CODE IS GIVEN BY TOTEM TOKENS OR U CAN ALSO USE THE “GOOGLE AUTHENTICATOR APP” IN YOUR MOBILE WHICH IS MOSTLY PREFERRED.

The image shows a screenshot of the Google Authenticator app on a mobile device. The screen displays a QR code at the top, followed by a list of six-digit verification codes. The first code is '234567'. Below the list is a button labeled 'Generate new key'.

- THE SCREEN IN MY MOBLIE LOOKS LIKE THIS:-THE CONTENT WHICH IAM HIDING USING A WHITE COLOR IS MY PRIVATE INFORMATION OF MY ACCOUNT WHICH CANNOT BE SHARED BUT HIS IS HOW A SIMPLE MFA CODE LOOKS LIKE AND THE MFA CODE KEEPS ON CHANGING ON A DAILY BASIS.ALWAYS USE MFA TO ENSURE ENHANCED SECURITY.



- THIS IS THE CODE THAT I HAVE TO ENTER IN MY AWS WEB BROWSER IN THE BELOW IMAGE IT IS SHOWN.



- AFTER SIGNING IN WE ENTER THE AWS MANAGEMENT CONSOLE AND THE DEFAULT REGION IS NORTH VIRGINIA THAT IS “us-east-1” WE CAN SELECT DIFFERENT REGIONS. IN THE IMAGE IT IS CLEARLY SHOWN IN THE RIGHT HAND SIDE WHICH REGION IAM CURRENTLY WORKING IN. THE BELOW IMAGE SHOWS THE SERVICES WHICH I VISITED AND THE CURRENT REGION WHICH IAM WORKING IN.
- TERMINOLOGIES:-
- Region:- It is a geothermal location which consists of a group of AZs (AVAILABILITY ZONES) and the network of a region is called a VPC(VIRTUAL PRIVATE CLOUD).
- AZ(Availability Zone):- they are a group of data centers.the network of a Az is a subnet.
- Edge Location:- it is a location in between various Azs or it is found at the boundary of an az In order to cache the incoming data acts like a storage for easier retrieval.

PLS NOTE CHANGE STARTS FROM HERE BELOW:-

- NOTE:-WE ARE WORKING IN THE PUBLIC SUBNET THAT IS “us-east-1” if we are working in any azs it will show us—east-1a,us-east-1b and so on up to us-east-1f. We can check the azs in our aws management console itself that is given in below figure.

SO NOW FOR THE deployment of web application SETTINGS IAM CURRENTLY WORKING IN THE NORTH VIRGINIA REGION SO IAM USING “us-east-1”.

IF WE CLICK ON VPC:-

HERE WE WILL BE ABLE TO SEE THE SUBNETS IN THE VPC DASHBOARD IF WE CLICK ON SUBNETS THEN TOTALLY THERE ARE 6 SUBNETS THAT IS PRIVATE SUBNETS FROM “us-east-1a” to “us-east-1f” totally we have 6 private subnets or we can also say we have 6 azs under NORTH VIRGINIA REGION THAT IS “US-EAST-1” REGION. This is provided by aws itself by default we cannot create any Availability Zones.

Logging into aws amangement console and searching for VPC(VIRTUAL PRIVATE CLOUD)

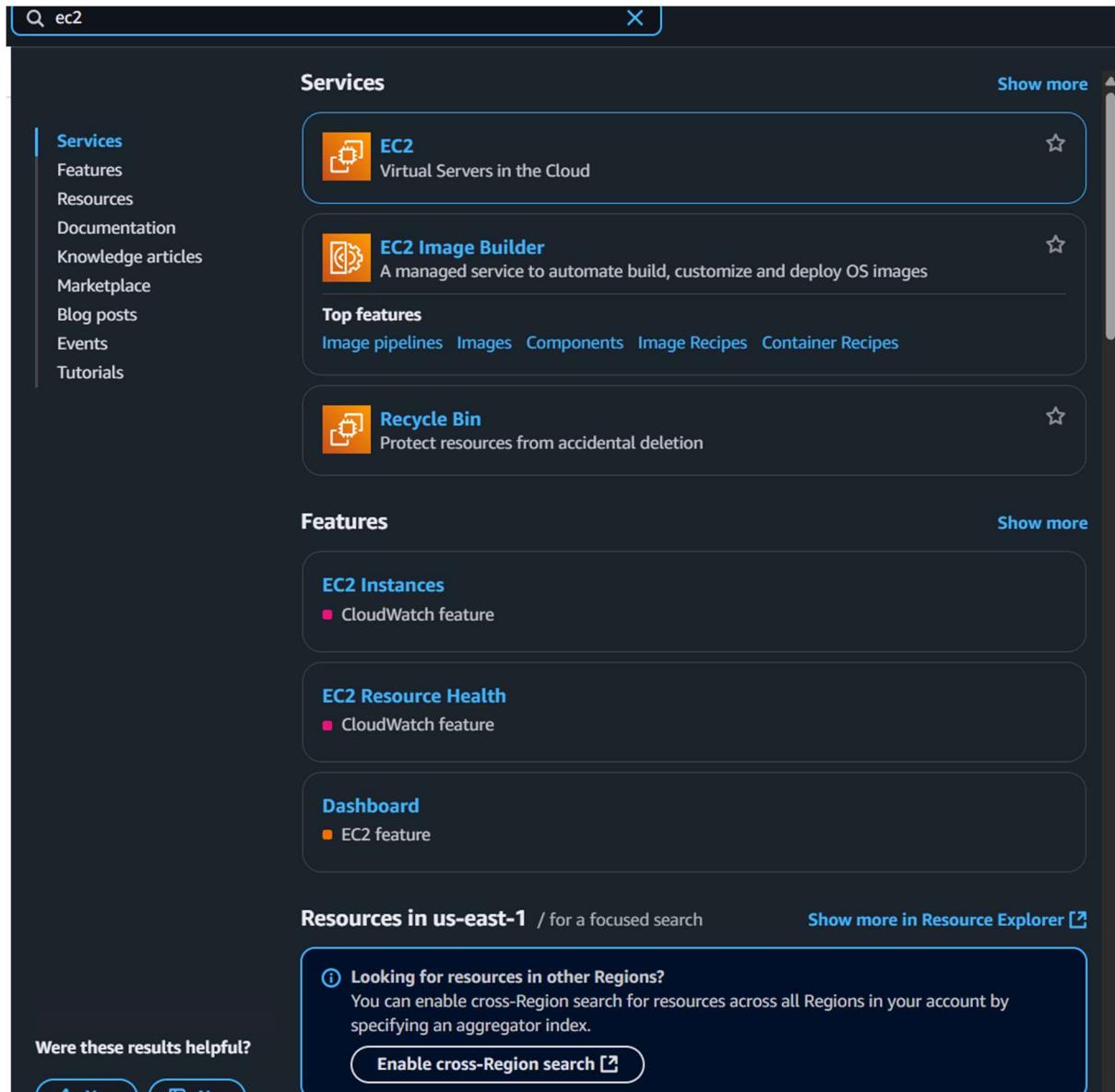
THIS IS THE AWS MANAGEMENT CONSOLE AFTER LOGGING IN CURRENTLY WORKING IN NORTH VIRGINIA REGION(US-EAST-1) the below image shows that

The screenshot shows the AWS Management Console Home page. At the top, there's a navigation bar with the AWS logo, a search bar, and a 'United States (N. Virginia)' dropdown. Below the navigation is the 'Console Home' section with a 'Recently visited' sidebar containing links to S3, IAM, VPC, Aurora and RDS, EC2, CloudWatch, Lambda, and API Gateway. To the right of this is the 'Applications' section, which is currently empty. Further down are sections for 'Welcome to AWS' (with links to Getting started with AWS, Training and certification, and AWS Builder Center), 'AWS Health' (showing 0 open issues and 0 scheduled changes), and 'Cost and usage' (showing credits remaining of \$158.62 USD, current month cost of \$1.34, and a 3.25% increase). On the far right, a large sidebar lists all AWS regions with their respective IDs:

Region	ID
N. Virginia	us-east-1
Ohio	us-east-2
N. California	us-west-1
Oregon	us-west-2
Asia Pacific	
Hyderabad	ap-south-2
Mumbai	ap-south-1
Osaka	ap-northeast-3
Seoul	ap-northeast-2
Singapore	ap-southeast-1
Sydney	ap-southeast-2
Tokyo	ap-northeast-1
Canada	ca-central-1
Central	eu-central-1
Europe	
Frankfurt	eu-central-1
Ireland	eu-west-1
London	eu-west-2
Paris	eu-west-3
Stockholm	eu-north-1
South America	
São Paulo	sa-east-1
There are 16 Regions that are not enabled for this account	

STEP 1 :- FIRST LAUNCH THE EC2 INSTANCES BY GOING TO THE EC2 DASHBOARD.

SEARCHING FOR EC2



THE BELOW FIGURE SHOWS MY EC2 DASHBOARD.

EC2

- [Dashboard](#)
- [EC2 Global View](#)
- [Events](#)
- Instances**
 - [Instances](#)
 - [Instance Types](#)
 - [Launch Templates](#)
 - [Spot Requests](#)
 - [Savings Plans](#)
 - [Reserved Instances](#)
 - [Dedicated Hosts](#)
 - [Capacity Reservations](#)
 - [Capacity Manager New](#)
- Images**
 - [AMIs](#)
 - [AMI Catalog](#)
- Elastic Block Store**
 - [Volumes](#)
 - [Snapshots](#)
 - [Lifecycle Manager](#)
- Network & Security**
 - [Security Groups](#)
 - [Elastic IPs](#)
 - [Placement Groups](#)
 - [Key Pairs](#)
 - [Network Interfaces](#)

Resources

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

Instances (running)	0	Auto Scaling Groups	0	Capacity Reservations	0
Dedicated Hosts	0	Elastic IPs	0	Instances	0
Key pairs	2	Load balancers	0	Placement groups	0
Security groups	3	Snapshots	0	Volumes	0

Launch instance

To get started, launch an Amazon EC2 instance, which is a virtual server in the cloud.

[Launch instance](#) [Migrate a server](#)

Note: Your instances will launch in the United States (N. Virginia) Region

Instance alarms

[View in CloudWatch](#)

⚠ 0 in alarm	✔ 0 OK	⌚ 0 insufficient data
--	---	---

[Instances in alarm](#)

Scheduled events

[Edit](#) [United States \(N. Virginia\)](#)

No scheduled events

Service health

[AWS Health Dashboard](#)

Region
United States (N. Virginia)
Status
✔ This service is operating normally.

Zones

Zone name	Zone ID
us-east-1a	use1-az2
us-east-1b	use1-az4
us-east-1c	use1-az6
us-east-1d	use1-az1
us-east-1e	use1-az3
us-east-1f	use1-az5

[Enable additional Zones](#)

SINCE I DON'T HAVE t2.micro as a free tier eligible in aws that is why iam taking t3.micro and even if we use the cost is less for t3.micro compared to t3.small in USD.

▼ Instance type [Info](#) | [Get advice](#)

Instance type

t3.small

Family: t3 2 vCPU 2 GiB Memory Current generation: true On-Demand SUSE base pricing: 0.0518 USD per Hour
 On-Demand Linux base pricing: 0.0208 USD per Hour On-Demand RHEL base pricing: 0.0496 USD per Hour
 On-Demand Ubuntu Pro base pricing: 0.0243 USD per Hour On-Demand Windows base pricing: 0.0392 USD per Hour

Free tier eligible



Get advice on instance type selection...

t3.micro

Family: t3 2 vCPU 1 GiB Memory Current generation: true On-Demand Ubuntu Pro base pricing: 0.0139 USD per Hour
 On-Demand SUSE base pricing: 0.0104 USD per Hour On-Demand Linux base pricing: 0.0104 USD per Hour
 On-Demand RHEL base pricing: 0.0392 USD per Hour On-Demand Windows base pricing: 0.0196 USD per Hour

Free tier eligible

t3.small

Family: t3 2 vCPU 2 GiB Memory Current generation: true On-Demand SUSE base pricing: 0.0518 USD per Hour
 On-Demand Linux base pricing: 0.0208 USD per Hour On-Demand RHEL base pricing: 0.0496 USD per Hour
 On-Demand Ubuntu Pro base pricing: 0.0243 USD per Hour On-Demand Windows base pricing: 0.0392 USD per Hour

Free tier eligible



c7i-flex.large

Family: c7i-flex 2 vCPU 4 GiB Memory Current generation: true On-Demand RHEL base pricing: 0.11359 USD per Hour
 On-Demand Ubuntu Pro base pricing: 0.08829 USD per Hour On-Demand Windows base pricing: 0.17219 USD per Hour
 On-Demand Linux base pricing: 0.08479 USD per Hour On-Demand SUSE base pricing: 0.14109 USD per Hour

Free tier eligible

m7i-flex.large

Family: m7i-flex 2 vCPU 8 GiB Memory Current generation: true

Free tier eligible

It seems like you may be new to launching instances in EC2. Take a walkthrough to learn about EC2, how to launch instances and about best practices

[Take a walkthrough](#) [Do not show me this message again.](#)

Launch an instance Info

Amazon EC2 allows you to create virtual machines, or instances, that run on the AWS Cloud. Quickly get started by following the simple steps below.

Name and tags Info

Name: ELEVATE-LABS-SERVER [Add additional tags](#)

Application and OS Images (Amazon Machine Image) Info

An AMI contains the operating system, application server, and applications for your instance. If you don't see a suitable AMI below, use the search field or choose [Browse more AMIs](#).

Recents	Quick Start					
Amazon Linux	macOS	Ubuntu	Windows	Red Hat	SUSE Linux	Debian
 Amazon Linux	 macOS	 Ubuntu	 Windows	 Red Hat	 SUSE Linux	 Debian

[Browse more AMIs](#) Including AMIs from AWS, Marketplace and the Community

Software Image (AMI) Canonical, Ubuntu, 24.04, amd6... [read more](#) ami-0360c520857e3138f

Virtual server type (instance type) t3.micro

Firewall (security group) New security group

Storage (volumes) 1 volume(s) - 8 GiB

[Cancel](#) [Launch instance](#) [Preview code](#)

IN THE ABOVE FIGURE IAM CREATING TWO EC2 INSTANCES AND CHHOSING UBUNTU 22.04 LTS AS MY OS. IN ONE GO AS U CAN SEE IAM CREATING 2 EC2 INSTANCES IN THE AWS MANAGEMENT CONSOLE AND AFTER THAT WE CAN RENAME IT IF WE WANT.

Description
Ubuntu Server 24.04 LTS (HVM), EBS General Purpose (SSD) Volume Type. Support available from Canonical (<http://www.ubuntu.com/cloud/services>).

Canonical, Ubuntu, 24.04, amd64 noble image

Architecture	AMI ID	Publish Date	Username
64-bit (x86) ▼	ami-0360c520857e3138f	2025-08-21	ubuntu Verified provider

Instance type Info | Get advice

Instance type

t3.micro Free tier eligible [▼](#)

Family: t3 2 vCPU 1 GiB Memory Current generation: true
 On-Demand Ubuntu Pro base pricing: 0.0139 USD per Hour On-Demand SUSE base pricing: 0.0104 USD per Hour
 On-Demand Linux base pricing: 0.0104 USD per Hour On-Demand RHEL base pricing: 0.0392 USD per Hour
 On-Demand Windows base pricing: 0.0196 USD per Hour

[All generations](#) [Compare instance types](#)

Additional costs apply for AMIs with pre-installed software

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 - required

[Default value](#) [▼](#) [Create new key pair](#)

THE ABOVE FIGURE SHOWS THAT IAM NOT TAKING ANY KEY – PAIR VALUE SINCE WE ARE NOT DOING AN SSH FOR THE EC2 INSTANCE INSTEAD WE ARE TRYING TO CREATE A LOAD BALANCER AND FROM THERE WE WIL BE ABLE TO ACCESS OUR WEBSITE.

▼ Network settings [Info](#)

VPC - required | [Info](#)

vpc-08652b60c7ce5df5b 172.31.0.0/16	(default) ▾	
--	-------------	--

Subnet | [Info](#)

No preference	▼	Create new subnet ?
---------------	---	-------------------------------------

Availability Zone | [Info](#)

No preference	▼	Enable additional zones ?
---------------	---	---

Auto-assign public IP | [Info](#)

Enable	▼
--------	---

Firewall (security groups) | [Info](#)

A security group is a set of firewall rules that control the traffic for your instance. Add rules to allow specific traffic to reach your instance.

<input checked="" type="radio"/> Create security group	<input type="radio"/> Select existing security group
--	--

Security group name - required

This security group will be added to all network interfaces. The name can't be edited after the security group is created. Max length is 255 characters. Valid characters: a-z, A-Z, 0-9, spaces, and _-:/()#@[]+=;&{}\$^

Description - required | [Info](#)

Inbound Security Group Rules

▼ Security group rule 1 (TCP, 22, 0.0.0.0/0)

Type Info	Protocol Info	Port range Info
ssh	TCP	22

Source type | [Info](#)

Anywhere	Source Info	Description - optional Info
▼	<input type="text" value="0.0.0.0"/>	e.g. SSH for admin desktop

▼ Security group rule 2 (TCP, 80, 0.0.0.0/0)

Type Info	Protocol Info	Port range Info
HTTP	TCP	80

Source type | [Info](#)

Anywhere	Source Info	Description - optional Info
▼	<input type="text" value="0.0.0.0"/>	e.g. SSH for admin desktop

Rules with source of 0.0.0.0/0 allow all IP addresses to access your instance. We recommend setting security group rules to allow access from known IP addresses only.

[Add security group rule](#)

IN THE ABOVE FIGURE THE DEFAULT VPC IS USED AND I HAVE CONFIGURED THE NETWORK SETTINGS TO ALLOW HTTP THAT IS PORT 80 AND SSH TRAFFIC THAT IS PORT 22 INTO MY EC2 INSTANCES.THAT IS SHOWN IN THE BELOW FIGURE.

Description - required | [Info](#)

Inbound Security Group Rules

▼ Security group rule 1 (TCP, 22, 0.0.0.0/0)

Type Info	Protocol Info	Port range Info
ssh	TCP	22

Source type | [Info](#)

Anywhere	Source Info	Description - optional Info
▼	<input type="text" value="0.0.0.0"/>	e.g. SSH for admin desktop

▼ Security group rule 2 (TCP, 80, 0.0.0.0/0)

Type Info	Protocol Info	Port range Info
HTTP	TCP	80

Source type | [Info](#)

Anywhere	Source Info	Description - optional Info
▼	<input type="text" value="0.0.0.0"/>	e.g. SSH for admin desktop

Rules with source of 0.0.0.0/0 allow all IP addresses to access your instance. We recommend setting security group rules to allow access from known IP addresses only.

[Add security group rule](#)

▼ Configure storage [Info](#)

1x GiB [▼](#) Root volume, 3000 IOPS, Not encrypted

[Add new volume](#)

The selected AMI contains instance store volumes, however the instance does not allow any instance store volumes. None of the instance store volumes from the AMI will be accessible from the instance

[① Click refresh to view backup information](#) [⟳](#)
The tags that you assign determine whether the instance will be backed up by any Data Lifecycle Manager policies.

0 x File systems [Edit](#)

THE ABOVE FIGURE REPRESENTS MY STORAGE OF MY EC2 INSTANCES.

Metadata version [Info](#)
V2 only (token required) [▼](#)

⚠ For V2 requests, you must include a session token in all instance metadata requests.
Applications or agents that use V1 for instance metadata access will break.

Metadata response hop limit [Info](#)
2

Allow tags in metadata [Info](#)
Select [▼](#)

User data - optional [Info](#)
Upload a file with your user data or enter it in the field.
[Choose file](#)

```
#!/bin/bash
sudo yum update -y
sudo yum install -y httpd
sudo systemctl start httpd
echo '<!DOCTYPE html><html><head><title>Cloud Auto Scaling Demo</title></head><body>
<h2>Hello from my Cloud Instance!</h2><p>This page is served from a cloud VM behind a load
balancer.</p></body></html>' > /var/www/html/index.html
```

User data has already been base64 encoded

HERE IAM USIJNG A BASH CSCRIPT TO INSTALL MY APACHE SERVER FOR THESE 2 EC2 INSTANCES.

Instances (1/2) Info

Last updated less than a minute ago

Find Instance by attribute or tag (case-sensitive)

All states ▾

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 ...	Elastic IP	IPv6 IPs
<input checked="" type="checkbox"/> ELEVATE-LAB...	i-0fb5012231445d3c	Running	t3.micro	Initializing	View alarms +	us-east-1b	ec2-54-209-251-72.co...	54.209.251.72	-	-
<input type="checkbox"/> ELEVATE-LAB...	i-06b9d2b0a177b5369	Running	t3.micro	Initializing	View alarms +	us-east-1b	ec2-54-227-121-41.co...	54.227.121.41	-	-

i-0fb5012231445d3c (ELEVATE-LABS-SERVER-1)

Details Status and alarms Monitoring Security Networking Storage Tags

Instance summary [Info](#)

Instance ID: i-0fb5012231445d3c

Public IPv4 address: 54.209.251.72 | [open address](#)

Instance state: Running

Private IPv4 addresses: 172.31.23.201

Public DNS: ec2-54-209-251-72.compute-1.amazonaws.com | [open address](#)

IPv6 address: -

Hostname type: IP name: ip-172-31-23-201.ec2.internal

Private IP DNS name (IPv4 only): ip-172-31-23-201.ec2.internal

THE ABOVE FIGURE REPRESENTS SERVER 1 IN EC2 DASHBOARD.

Instances (1/2) Info

Last updated 1 minute ago

Find Instance by attribute or tag (case-sensitive)

All states ▾

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 ...	Elastic IP	IPv6 IPs
<input type="checkbox"/> ELEVATE-LAB...	i-0fb5012231445d3c	Running	t3.micro	Initializing	View alarms +	us-east-1b	ec2-54-209-251-72.co...	54.209.251.72	-	-
<input checked="" type="checkbox"/> ELEVATE-LAB...	i-06b9d2b0a177b5369	Running	t3.micro	Initializing	View alarms +	us-east-1b	ec2-54-227-121-41.co...	54.227.121.41	-	-

i-06b9d2b0a177b5369 (ELEVATE-LABS-SERVER-2)

Details Status and alarms Monitoring Security Networking Storage Tags

Instance summary [Info](#)

Instance ID: i-06b9d2b0a177b5369

Public IPv4 address: 54.227.121.41 | [open address](#)

Instance state: Running

Private IPv4 addresses: 172.31.21.203

Public DNS: ec2-54-227-121-41.compute-1.amazonaws.com | [open address](#)

IPv6 address: -

Hostname type: Unknown

Private IP DNS name (IPv4 only): ip-172-31-21-203.ec2.internal

THE ABOVE FIGURE REPRESENTS SERVER 2 IN EC2 DASHBOARD AND BOTH ARE IN RUNNING STATE. NOW CURRENTLY I WILL STOP THE INSTANCE TO SAVE SOME CHARGES.

Instances (1/2) Info											
Last updated 2 minutes ago Connect Instance state Actions Launch instances											
<input type="text" value="Q. Find Instance by attribute or tag (case-sensitive)"/> All states											
Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 ...	Elastic IP	IPv6 IPs	
ELEVATE-LAB...	i-0fb5012231445d3c	Stopping	t3.micro	Initializing	View alarms	us-east-1b	ec2-54-209-251-72.co...	54.209.251.72	-	-	
ELEVATE-LAB...	i-06b9d2b0a177b5369	Stopping	t3.micro	Initializing	View alarms	us-east-1b	ec2-54-227-121-41.co...	54.227.121.41	-	-	

I HAVE STOPPED THE EC2 INSTANCES WHICH WERE RUNNING TO SAVE SOME COSTS.

STEP 2:- CREATE THE LOAD BALANCER . IT IS AVAILABE IN EC2 ITSELF JUST TOWARDS THE LEFT PANE.

The screenshot shows the AWS EC2 Load Balancers page. On the left, there is a navigation sidebar with various services like Spot Requests, Savings Plans, Reserved Instances, and Auto Scaling. The main content area has a heading "Introducing URL rewrite for Application Load Balancer" with a sub-instruction: "Modify host headers and URL paths of incoming requests before they reach your targets. To get started, add a rule to your listener and configure a transform." Below this is a section titled "Load balancers" with a sub-instruction: "Elastic Load Balancing scales your load balancer capacity automatically in response to changes in incoming traffic." A search bar labeled "Filter load balancers" is present. A central illustration features a robot holding a wrench and working on a cloud, with the text "No load balancers" below it. At the bottom, a message says "0 load balancers selected" and "Select a load balancer above." Action buttons include "Actions" and "Create load balancer".

CURRENTLY IN THE ABOVE FIGURE THERE ARE NO LOAD BALANCERS SO IAM GOING TO CREATE A LOAD BALANCER.

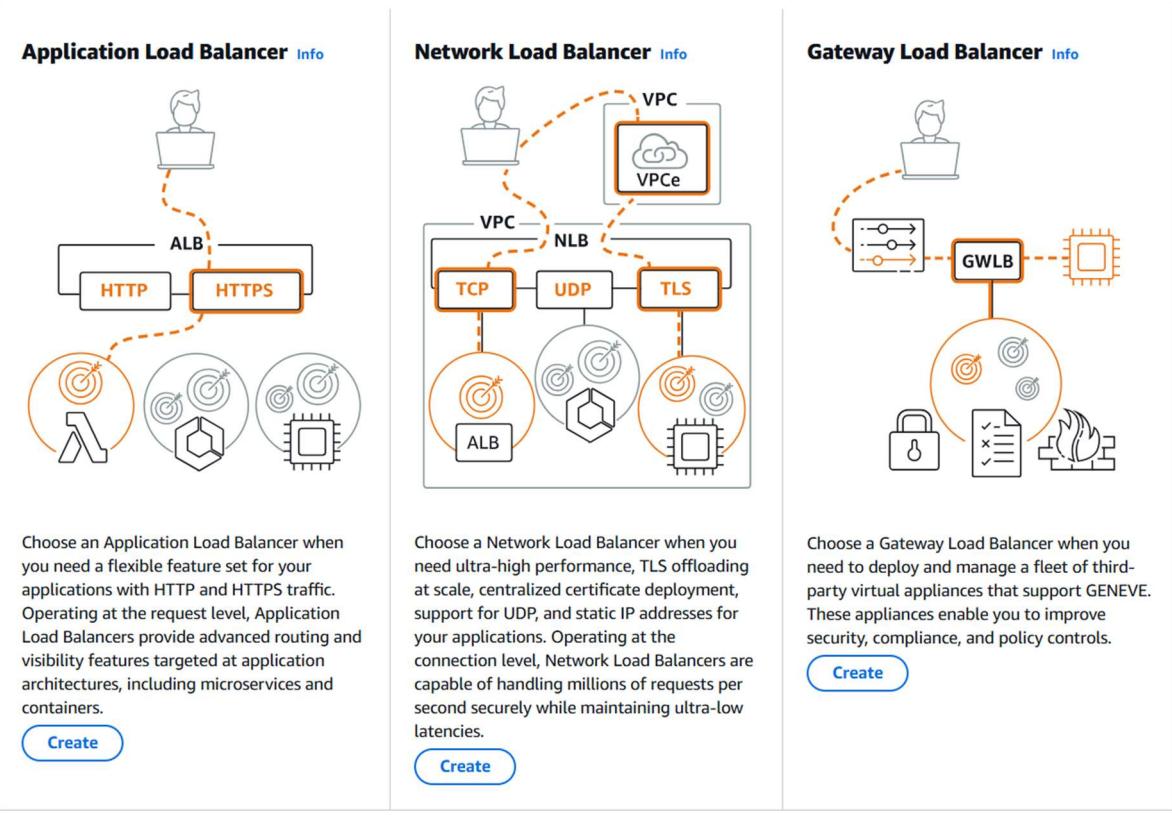
The screenshot shows the "Compare and select load balancer type" page. It compares three types of load balancers: Application Load Balancer, Network Load Balancer, and Gateway Load Balancer. Each type has a diagram and a brief description. The "Create" button is visible at the bottom of each section.

- Application Load Balancer**: Handles HTTP and HTTPS traffic from a browser. It routes requests to Lambda functions, API Gateways, and microservices. Choose this for 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**: Handles TCP, UDP, and TLS traffic. It routes requests to VPC endpoints, Lambda functions, and API Gateways. Choose this when you need ultra-low latency, TLS offloading at a centralized certificate management point, support for UDP and static IP addresses for your applications, and handling millions of requests per second securely while maintaining ultra-low latencies.
- Gateway Load Balancer**: Handles traffic for third-party virtual appliances supporting GENEVE. It routes requests to Lambda functions, API Gateways, and VPC endpoints. Choose this when you need to deploy and manage a fleet of third-party virtual appliances that support GENEVE. These appliances enable you to improve security, compliance, and policy controls.

WE HAVE THREE LOAD BALNCERS THAT WE CAN USE WHEN TO USE WHICH LOAD BALNCER IS GIVEN BELOW IT ALL DEPENDS UPON THE USE CASE:-

Load Balancer	Best For	Benefits	Disadvantages	Real-Time Use Cases
Application Load Balancer (ALB)	HTTP/HTTPS traffic with advanced routing (Layer 7)	- Path/host-based routing - WebSocket & HTTP/2 support - Native integration with ECS/EKS - Detailed CloudWatch metrics	- Slightly higher latency - Only supports HTTP/HTTPS	- Routing /api/* to microservices - Hosting multiple domains on one ALB - Blue/Green deployments
Network Load Balancer (NLB)	High-performance TCP/UDP traffic (Layer 4)	- Ultra-low latency - Handles millions of requests/sec - Static IP support - TLS termination	- No content-based routing - Limited protocol-level features	- Load balancing for financial trading apps - Gaming/VoIP apps - IoT backends
Classic Load Balancer (CLB)	Legacy apps needing basic Layer 4/7 support	- Simple setup - Supports EC2-Classic - HTTP & TCP support	- Deprecated for new apps - No path-based routing - Limited observability	- Legacy monoliths on EC2-Classic - Basic web apps without microservices
Gateway Load Balancer (GWLB)	Deploying, scaling, and managing third-party virtual appliances (Layer 3)	- Transparent traffic inspection - Integrates with firewalls, IDS/IPS, deep packet inspection tools - Preserves source IP - Scales elastically	- Not for direct client traffic - Requires VPC endpoint service setup - Higher complexity	- Inserting Palo Alto/Checkpoint firewalls - Centralized traffic inspection across VPCs - Inline threat detection

Use Case	Recommended Load Balancer
Microservices with path-based routing	ALB
Ultra-low latency TCP/UDP apps	NLB
Legacy EC2-Classic workloads	CLB
Inline security appliances (firewalls, IDS)	GWLB



OUT OF THESE 3 LOAD BALANCERS FOR THIS TASK 7 I WILL BE CHOOSING AN APPLICATION LOAD BALNCER.

Create Application Load Balancer [Info](#)

The Application Load Balancer distributes incoming HTTP and HTTPS traffic across multiple targets such as Amazon EC2 instances, microservices, and containers, based on request attributes. When the load balancer receives a connection request, it evaluates the listener rules in priority order to determine which rule to apply, and if applicable, it selects a target from the target group for the rule action.

▶ How Application Load Balancers work

Basic configuration

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

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

- Serves internet-facing traffic.
- Has public IP addresses.
- DNS name resolves to public IPs.
- Requires a public subnet.

Internal

- Serves internal traffic.
- Has private IP addresses.
- DNS name resolves to private IPs.
- Compatible with the IPv4 and Dualstack IP address types.

Load balancer IP address type [Info](#)
Select the front-end IP address type to assign to the load balancer. The VPC and subnets mapped to this load balancer must include the selected IP address types. Public IPv4 addresses have an additional cost.

IPv4

- Includes only IPv4 addresses.

Dualstack

- Includes IPv4 and IPv6 addresses.

Dualstack without public IPv4

- Includes a public IPv6 address, and private IPv4 and IPv6 addresses. Compatible with internet-facing load balancers only.

HERE WIL BE CREATING THE LOAD BALNCER AND IT IS INTERNET FACING WHICH MENAS WE HAVE DIRECT ACCESS TO THE INTERNET.

The figure shows the configuration of an Application Load Balancer (ALB) in the AWS Cloud. It includes sections for IP pools, Availability Zones and subnets, and Security groups.

IP pools: You can optionally choose to configure an IPAM pool as the preferred source for your load balancers IP addresses. Create or view Pools in the Amazon VPC IP Address Manager console.

Availability Zones and subnets: Select at least two Availability Zones and a subnet for each zone. A load balancer node will be placed in each selected zone and will automatically scale in response to traffic. The load balancer routes traffic to targets in the selected Availability Zones only.

Security groups: 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.

FOR THE LOAD BALNCER IN THE ABOVE FIGURE IAM USING DEFAULT VPC AND SECURITY GROUP.

Listeners and routing: 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.

Default action: The default action is used if no other rules apply. Choose the default action for traffic on this listener.

Routing action: Forward to target groups, Redirect to URL, or Return fixed response.

Forward to target group: Choose a target group and specify routing weight or create target group.

Target group	Weight	Percent
Select a target group	1	100%

Target group stickiness: Enables the load balancer to bind a user's session to a specific target group. To use stickiness the client must support cookies. If you want to bind a user's session to a specific target, turn on the Target Group attribute Stickiness.

THE ABOVE FIGURE SHOWS THAT FOR ALB WE NEED A LISTNER PORT TO UNDERSTAND THE REQUESTS WHICH ARE BEING MADE VIA THE HTTP PROTOCOL.

THE BELOW FIGURE REPRESNTS THE TARGET GROUP FOR MY LOAD BALNCER WIHICH IAM CREATING FOR THE EC2 INSTANCES.

Settings - immutable

Choose a target type and the load balancer and listener will route traffic to your target. These settings can't be modified after target group creation.

Target type

Indicate what resource type you want to target. Only the selected resource type can be registered to this target group.

Instances

Supports load balancing to instances in a VPC. Integrate with Auto Scaling Groups or ECS services for automatic management.

Suitable for: ALB NLB GWLB

IP addresses

Supports load balancing to VPC and on-premises resources. Facilitates routing to IP addresses and network interfaces on the same instance. Supports IPv6 targets.

Suitable for: ALB NLB GWLB

Lambda function

Supports load balancing to a single Lambda function. ALB required as traffic source.

Suitable for: ALB

Application Load Balancer

Allows use of static IP addresses and PrivateLink with an Application Load Balancer. NLB required as traffic source.

Suitable for: NLB

Target group name

Name must be unique per Region per AWS account.

EC2-INSTANCES

Accepts: a-z, A-Z, 0-9, and hyphen (-). Can't begin or end with hyphen. 1-32 total characters; Count: 13/32

Protocol

Protocol for communication between the load balancer and targets.

HTTP

Port

Port number where targets receive traffic. Can be overridden for individual targets during registration.

80

1-65535

THE BELOW FIGURE SHOWS THE HEALTH CHECKS FOR THE “/” PATH WHICH MEANS ANY PATH. AND HEALTH CHECK PROTOCOL IS HTTP AT PORT 80

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.

vpc-08652b60c7ce5df5b (default) 172.31.0.0/16 Create VPC

Protocol version

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

HTTP2
Send requests to targets using HTTP/2. Supported when the request protocol is HTTP/2 or gRPC, but gRPC-specific features are not available.

gRPC
Send requests to targets using gRPC. Supported when the request protocol is gRPC.

Health checks
The associated load balancer periodically sends requests, per the settings below, to the registered targets to test their status.

Health check protocol

HTTP

Health check path
Use the default path of “/” to perform health checks on the root, or specify a custom path if preferred.

/

Up to 1024 characters allowed.

► Advanced health check settings

Advanced health check settings

Health check port
The port the load balancer uses when performing health checks on targets. By default, the health check port is the same as the target group's traffic port. However, you can specify a different port as an override.

Traffic port
 Override

Healthy threshold
The number of consecutive health checks successes required before considering an unhealthy target healthy.

5
2-10

Unhealthy threshold
The number of consecutive health check failures required before considering a target unhealthy.

2
2-10

Timeout
The amount of time, in seconds, during which no response means a failed health check.

5 seconds
2-120

Interval
The approximate amount of time between health checks of an individual target.

30 seconds
5-300

Success codes
The HTTP codes to use when checking for a successful response from a target. You can specify multiple values (for example, "200,202") or a range of values (for example, "200-299").

200

Restore defaults

THIS EVERYTHING REMAINS THE SAME THERE IS NO CHANGE FOR HEALTH CHECK PORTS FOR THE TARGET GROUPS NOW I HAVE TO REGISTER MY TARGET GROUPS FOR THE LOAD BALANCER.

Available instances (2)

<input type="checkbox"/>	Instance ID	Name	State	Security groups	Zone	Private IPv4 address	Subnet ID
<input type="checkbox"/>	i-0fb5012231445d3c	ELEVATE-LABS-SERVER-1	Running	launch-wizard-3	us-east-1b	172.31.23.201	subnet-0200b59ddc95923
<input type="checkbox"/>	i-06b9d2b0a177b5369	ELEVATE-LABS-SERVER-2	Running	launch-wizard-3	us-east-1b	172.31.21.203	subnet-0200b59ddc95923

0 selected

Ports for the selected instances
Ports for routing traffic to the selected instances.

80
1-65535 (separate multiple ports with commas)

Include as pending below

2 selections are now pending below. Include more or register targets when ready.

Review targets

Targets (2)

<input type="checkbox"/>	Name	Port	State	Security groups	Zone	Private IPv4 address	Subnet ID	Launch time
<input type="checkbox"/>	ELEVATE-LABS-SERVER-2	80	Running	launch-wizard-3	us-east-1b	172.31.21.203	subnet-0200b59ddc9592303	October 31, 2025, 09:09 (UTC+05:30)
<input type="checkbox"/>	ELEVATE-LABS-SERVER-1	80	Running	launch-wizard-3	us-east-1b	172.31.23.201	subnet-0200b59ddc9592303	October 31, 2025, 09:09 (UTC+05:30)

Remove all pending

THE ABOVE FIGURE INDICATES I HAVE REGISTERED MY EC2 INSTANCES WHICH ARE THE TARGETS TO THE TARGET GROUP.

SUMMARY OF TARGET GROUP WHICH IS GOING TO BE CREATED.

Review and create

Review your target group configuration before creating

Step 1: Target group details

[Edit](#)

Target group details

Name AUTOSCALING-LB-GROUP	Target type Instance	Protocol : Port HTTP: 80	Protocol version HTTP1
VPC vpc-08652b60c7ce5df5b	IP address type IPv4		

Health check details

Health check protocol HTTP	Health check path /	Health check port traffic-port	Interval 30 seconds
Timeout 5 seconds	Healthy threshold 5	Unhealthy threshold 2	Success codes 200

Step 2: Register targets

[Edit](#)

Targets (2)

Instance ID	Name	Port	Zone
i-06b9d2b0a177b5369	ELEVATE-LABS-SERVER-2	80	us-east-1b
i-0fb5012231445d3c	ELEVATE-LABS-SERVER-1	80	us-east-1b

THE BELOW FIGURE REPRESENTS WE HAVE SUCCESSFULLY CREATED THE TARGET GROUP.

Successfully created the target group: AUTOSCALING-LB-GROUP. Anomaly detection is automatically applied to all registered targets. Results can be viewed in the Targets tab.

AUTOSCALING-LB-GROUP

Details

Target type Instance	Protocol : Port HTTP: 80	Protocol version HTTP1
IP address type IPv4	Load balancer None associated	VPC vpc-08652b60c7ce5df5b
2 Total targets	0 Healthy 0 Anomalous	0 Unhealthy 0 Unused 0 Initial 0 Draining

Distribution of targets by Availability Zone (AZ)
Select values in this table to see corresponding filters applied to the Registered targets table below.

Registered targets (2) [Info](#)

Target groups route requests to individual registered targets using the protocol and port number specified. Health checks are performed on all registered targets according to the target group's health check settings. Anomaly detection is automatically applied to HTTP/HTTPS target groups with at least 3 healthy targets.

Instance ID	Name	Port	Zone	Health status	Health status details	Administrative o...	Override details	Launch...	Anomaly c...
i-0fb5012231445d3c	ELEVATE-LABS...	80	us-east-1b (us...)	Unused	Target group is not co...	-	-	October 3...	Normal
i-06b9d2b0a177b5369	ELEVATE-LABS...	80	us-east-1b (us...)	Unused	Target group is not co...	-	-	October 3...	Normal

NEXT I HAVE ADDED MY TARGET GROUP TO THE ALB.

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 **Port**

HTTP 80 1-65535

Default action [Info](#)

The default action is used if no other rules apply. Choose the default action for traffic on this listener.

Routing action

Forward to target groups Redirect to URL Return fixed response

Forward to target group [Info](#)

Choose a target group and specify routing weight or [create target group](#).

Target group

Target group	Weight	Percent
AUTOSCALING-LB-GROUP Target type: Instance, IPv4 Target stickiness: Off	HTTP C	1 0-999

+ Add target group
You can add up to 4 more target groups.

Target group stickiness [Info](#)

Enables the load balancer to bind a user's session to a specific target group. To use stickiness the client must support cookies. If you want to bind a user's session to a specific target, turn on the Target Group attribute Stickiness.

Turn on target group stickiness

THE BELOW FIGURE SHOWS WHERE MY ALB LIVES.

Availability Zones and subnets [Info](#)

Select at least two Availability Zones and a subnet for each zone. A load balancer node will be placed in each selected zone and will automatically scale in response to traffic. The load balancer routes traffic to targets in the selected Availability Zones only.

us-east-1a (use1-az2)

Subnet

Only CIDR blocks corresponding to the load balancer IP address type are used. At least 8 available IP addresses are required for your load balancer to scale efficiently.

subnet-057b40416a98890f3
IPv4 subnet CIDR: 172.31.80.0/20

us-east-1b (use1-az4)

us-east-1c (use1-az5)

Subnet

Only CIDR blocks corresponding to the load balancer IP address type are used. At least 8 available IP addresses are required for your load balancer to scale efficiently.

subnet-0d5cabf48603fed34
IPv4 subnet CIDR: 172.31.32.0/20

us-east-1d (use1-az1)

us-east-1e (use1-az3)

us-east-1f (use1-az5)

THE BELOW FIGURE SHOWS THAT ALB IS CREATED SUCCESSFULLY

Successfully created load balancer: ELEVATE-ALB
It might take a few minutes for your load balancer to fully set up and route traffic. Targets will also take a few minutes to complete the registration process and pass initial health checks.

Introducing URL rewrite for Application Load Balancer
Modify host headers and URL paths of incoming requests before they reach your targets. To get started, add a rule to your listener and configure a transform. [Learn more](#)

ELEVATE-ALB

[Actions](#)

Details

Load balancer type Application	Status Provisioning	VPC vpc-08652b60c7ce5df5b	Load balancer IP address type IPv4
Scheme Internet-facing	Hosted zone Z355XDOTRQ7X7K	Availability Zones subnet-0d5cabf48603fed34 us-east-1c (use1-az6) subnet-057b40416a98890f3 us-east-1a (use1-az2)	Date created October 31, 2025, 09:24 (UTC+05:30)
Load balancer ARN arn:aws:elasticloadbalancing:us-east-1:518286664533:loadbalancer/app/ELEVATE-ALB/6bc791a831a3f9a8		DNS name info ELEVATE-ALB-378360910.us-east-1.elb.amazonaws.com (A Record)	

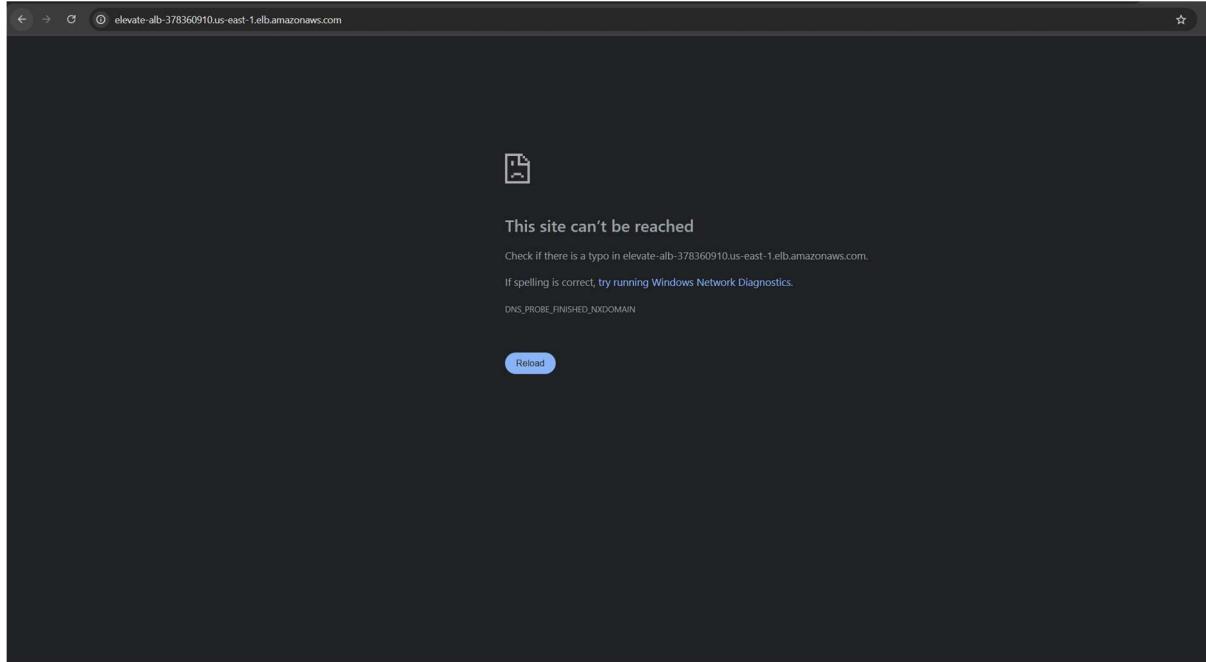
[Listeners and rules](#) [Network mapping](#) [Resource map](#) [Security](#) [Monitoring](#) [Integrations](#) [Attributes](#) [Capacity](#) [Tags](#)

Listeners and rules (1) [Info](#)

A listener checks for connection requests on its configured protocol and port. Traffic received by the listener is routed according to the default action and any additional rules.

Protocol:Port	Default action	Rules	ARN	Security policy	Default SSL/TLS certificate	mTLS	Trust store
HTTP:80	Forward to target group AUTOSCALING-LB-GROUP 1 (100%) Target group stickiness: Off	1 rule	ARN	Not applicable	Not applicable	Not applicable	Not applicable

WE GOT AN ISSUE THAT OUR WEBSITE IS NOT WORKING LIKE THIS BELOW FIGURE BCZ INITIALLY THE ALB IS UNDER PROVISIONING STATE SO WE CANT SEE ANYTHING.



THE ELEVATE-ALB IS PROVISIONING IN THE ABOVE FIGURE.

**ONCE IT IS SUCCESSFULLY DONE THE ALB NOW IS IN THE ACTIVE STATE
NOW IF WE OPEN VIA THE DNS URL OF THE ALB WE WILL GET THE WEBSITE AS
SHOWN IN THE BELOW FIGURE.**

Iam creating a launch template for an auto scaling group so that manually we do not have to go to very ec2 instance and ssh and install the apache server so now iam going to launch a template

Instance type requirements [Info](#)

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
for-auto-scaling-groups [edit] lt-035c61d5d5fab8fd9	Default	
Instance type		
t3.micro		

[Override launch template](#)

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-08652b60c7ce5df5b 172.31.0.0/16 Default	[edit]
--	------------------------

[Create a VPC](#) [\[edit\]](#)

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	[edit]
use1-az2 (us-east-1a) subnet-057b40416a98890f3 172.31.80.0/20 Default	[edit]
use1-az4 (us-east-1b) subnet-0200b59ddc9592303 172.31.16.0/20 Default	[edit]
use1-az6 (us-east-1c) subnet-0d5cabf48603fed34 172.31.32.0/20 Default	[edit]

[Create a subnet](#) [\[edit\]](#)

For these auto scaling groups azs of alb should be same and should be matching then only it will be able to transfer the traffic.

Choose launch template [Info](#)

Specify a launch template that contains settings common to all EC2 instances that are launched by this Auto Scaling group.

Name

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

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

Launch template [Info](#)

For accounts created after May 31, 2023, the EC2 console only supports creating Auto Scaling groups with launch templates. Creating Auto Scaling groups with launch configurations is not recommended but still available via the CLI and API until December 31, 2023.

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.

for-auto-scaling-groups [edit]	[create]
--	--------------------------

[Create a launch template](#) [\[edit\]](#)

Version

Default (1) [edit]	[create]
--	--------------------------

[Create a launch template version](#) [\[edit\]](#)

Description -	Launch template for-auto-scaling-groups [edit] lt-035c61d5d5fab8fd9	Instance type t3.micro
AMI ID ami-0360c520857e3138f	Security groups -	Request Spot Instances No
Key pair name -	Security group IDs sg-062fad1a6bc29909 [edit]	

[Additional details](#)

The above figure shows that we are attaching the template for the auto scaling groups.

Integrate with other services - optional Info

Use a load balancer to distribute network traffic across multiple servers. Enable service-to-service communications with VPC Lattice. Shift resources away from impaired Availability Zones with zonal shift. You can also customize 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.

Select Load balancing options

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 to attach

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 ▼ 

AUTOSCALING-LB-GROUP | HTTP X
Application Load Balancer: ELEVATE-ALB

I am attaching my auto scaling groups to an existing load balancer and enabling the auto health checks while load balancing the traffic.

THIS IS FOR THE THRESHOLD WHICH IAM GIVING AND WE CAN HANDLE THE LIMITS OF AUTOSCALING

Group size Info

Set the initial size of the Auto Scaling group. After creating the group, you can change its size to meet demand, either manually or by using automatic scaling.

Desired capacity type
Choose the unit of measurement for the desired capacity value. vCPUs and Memory(GiB) are only supported for mixed instances groups configured with a set of instance attributes.

Units (number of instances) ▼

Desired capacity
Specify your group size.

2

Scaling Info

You can resize your Auto Scaling group manually or automatically to meet changes in demand.

Scaling limits
Set limits on how much your desired capacity can be increased or decreased.

Min desired capacity	Max desired capacity
1	4
Equal or less than desired capacity	Equal or greater than desired capacity

Automatic scaling - optional
Choose whether to use a target tracking policy | Info

You can set up other metric-based scaling policies and scheduled scaling after creating your Auto Scaling group.

Review [Info](#)

Step 1: Choose launch template [Edit](#)

Group details		
Auto Scaling group name	tg-for-ec2-group	
Launch template		
Launch template	Version	Description
for-auto-scaling-groups [?]	Default	

Step 2: Choose instance launch options [Edit](#)

Network		
VPC	vpc-08652b60c7ce5df5b [?]	
Availability Zones and subnets		
Availability Zone	Subnet	Subnet CIDR range
use1-az2 (us-east-1a)	subnet-057b40416a98890f3 [?]	172.31.80.0/20
use1-az4 (us-east-1b)	subnet-0200b59ddc9592303 [?]	172.31.16.0/20
use1-az6 (us-east-1c)	subnet-0d5cabf48603fed34 [?]	172.31.32.0/20
Availability Zone distribution		
Balanced best effort		

THIS IS THE SUMMARY OF THE AUTO SCALING GROUP IN THE ABOVE FIGURE AND THE BELOW FIGURE.

Step 2: Choose instance launch options [Edit](#)

Network		
VPC	vpc-08652b60c7ce5df5b [?]	
Availability Zones and subnets		
Availability Zone	Subnet	Subnet CIDR range
use1-az2 (us-east-1a)	subnet-057b40416a98890f3 [?]	172.31.80.0/20
use1-az4 (us-east-1b)	subnet-0200b59ddc9592303 [?]	172.31.16.0/20
use1-az6 (us-east-1c)	subnet-0d5cabf48603fed34 [?]	172.31.32.0/20
Availability Zone distribution		
Balanced best effort		

Instance type requirements

This Auto Scaling group will adhere to the launch template.

Step 3: Integrate with other services [Edit](#)

Load balancing		
Load balancer 1		
Name	Type	Target group
ELEVATE-ALB [?]	Application/HTTP	AUTOSCALING-LB-GROUP [?]

The screenshot shows the AWS Auto Scaling Groups page. At the top, there's a search bar and several navigation links: 'Launch configurations', 'Launch templates', 'Actions', and 'Create Auto Scaling group'. Below the search bar, a table lists the auto scaling group with the following details:

Name	Launch template/configuration	Instances	Status	Desired capacity	Min	Max	Availability Zones	Creation time
tg-for-ec2-group	for-auto-scaling-groups Version Default	0	Updating capacity...	2	1	4	3 Availability Zones	Fri Oct 31 2025 10:20:16 GMT+0530 (India Standard Time)

Below this, there's a detailed view for the 'tg-for-ec2-group' capacity overview, showing desired capacity as 2, scaling limits from 1 to 4, and a status message indicating 'Updating capacity'.

NOW IN THE ABOVE FIGURE I HAVE CREATED MY AUTO SCALING GROUP AND IMA GOING TO ATTACH THIS TO MY LOAD BALANCER.

IAM CREATING THE THRESHOLD BASED ON THE INFORMATION PROVIDED BY ELEVATE LABS.60 % CPU UTILIZATION

Create dynamic scaling policy

Policy type: Target tracking scaling

Scaling policy name: 60%_UTILISATION

Metric type: Average CPU utilization

Target value: 60

Instance warmup: 300 seconds

Disable scale in to create only a scale-out policy

```
ubuntu@ip-172-31-23-201:~$ ab -n 500 -c 50 http://ELEVATE-ALB-378360910.us-east-1.elb.amazonaws.com/
This is ApacheBench, Version 2.3 <$Revision: 1903618 $>
Copyright 1996 Adam Twiss, Zeus Technology Ltd, http://www.zeustech.net/
Licensed to The Apache Software Foundation, http://www.apache.org/

Benchmarking ELEVATE-ALB-378360910.us-east-1.elb.amazonaws.com (be patient)
Completed 100 requests
Completed 200 requests
Completed 300 requests
Completed 400 requests
Completed 500 requests
Finished 500 requests

Server Software:      Apache/2.4.58
Server Hostname:     ELEVATE-ALB-378360910.us-east-1.elb.amazonaws.com
Server Port:          80

Document Path:        /
Document Length:     10671 bytes

Concurrency Level:   50
Time taken for tests: 0.083 seconds
Complete requests:   500
Failed requests:     250
                     (Connect: 0, Receive: 250, Exceptions: 0)
Total transferred:   2807000 bytes
HTML transferred:    2677000 bytes
Requests per second: 6035.29 [#/sec] (mean)
Time per request:    8.285 [ms] (mean)
Time per request:    0.166 [ms] (mean, across all concurrent requests)
Transfer rate:       33088.03 [Kbytes/sec] received
```

We can see in the above figure that these many requests are processed.the two figures indicate the output of how many requests were processed and how many instances were created.

```

Finished 500 requests

Server Software: Apache/2.4.58
Server Hostname: ELEVATE-ALB-378360910.us-east-1.elb.amazonaws.com
Server Port: 80

Document Path: /
Document Length: 10671 bytes

Concurrency Level: 50
Time taken for tests: 0.083 seconds
Complete requests: 500
Failed requests: 250
    (Connect: 0, Receive: 0, Length: 250, Exceptions: 0)
Total transferred: 2807000 bytes
HTML transferred: 2677000 bytes
Requests per second: 6035.29 [#/sec] (mean)
Time per request: 8.285 [ms] (mean)
Time per request: 0.166 [ms] (mean, across all concurrent requests)
Transfer rate: 33088.03 [Kbytes/sec] received

Connection Times (ms)
              min     mean[+/-sd] median     max
Connect:        1       1     0.1      1       2
Processing:     1       6     3.6      5      16
Waiting:        1       6     3.6      5      16
Total:          2       7     3.6      6      17

Percentage of the requests served within a certain time (ms)
 50%      6
 66%      8
 75%     10
 80%     11
 90%     13
 95%     14
 98%     16
 99%     16
100%    17 (longest request)

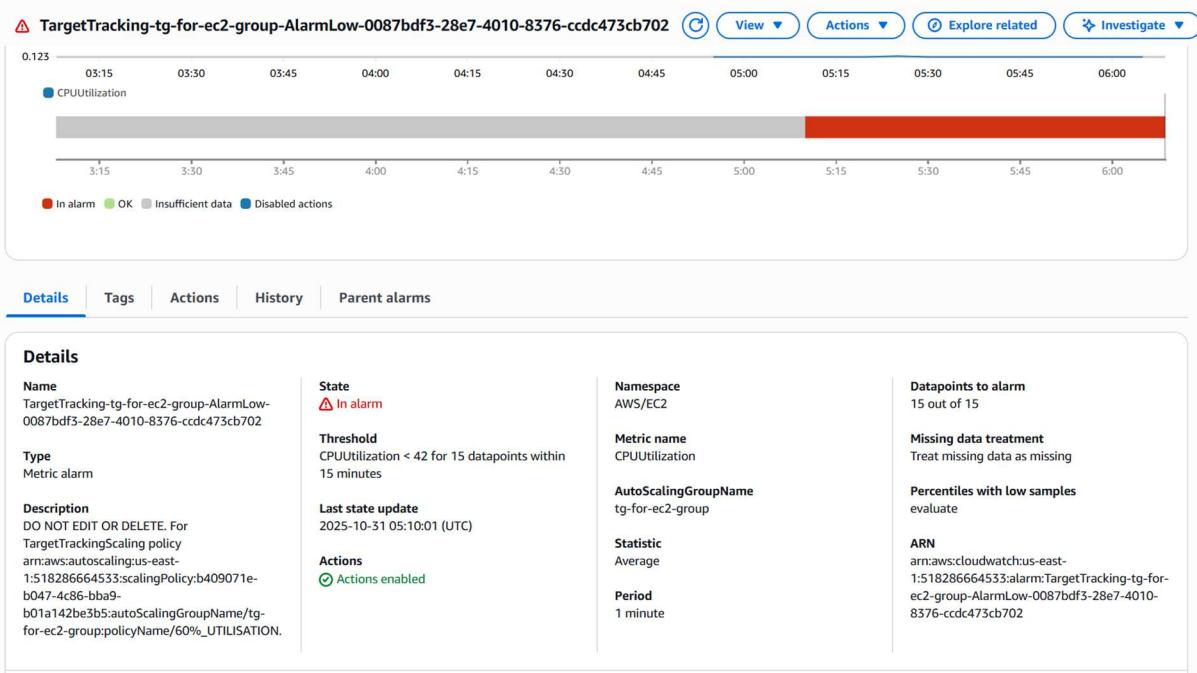
```

The below figure indicates the activity that has happened while executing the command which was mentioned and created an new ec2 instance.

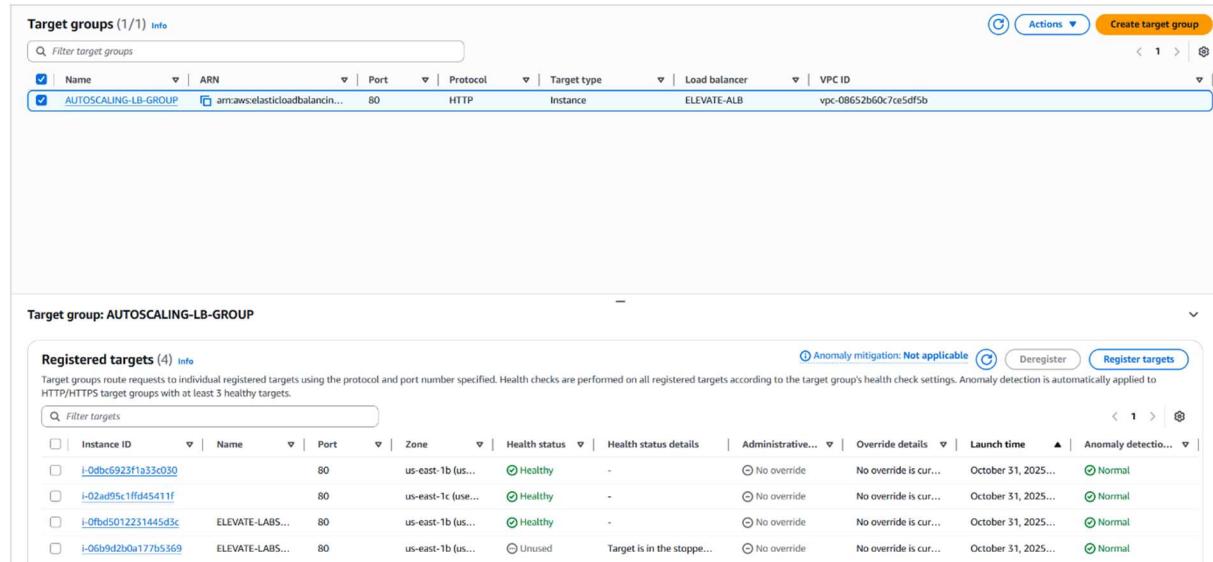
The screenshot shows the AWS Auto Scaling Groups console. At the top, there's a search bar and a table for managing Auto Scaling groups. One group, 'tg-for-ec2-group', is selected. Below the table, a detailed view for 'tg-for-ec2-group' is shown, including a 'Create notification' button. The 'Activity history' section at the bottom lists three events:

Status	Description	Cause	Start time	End time
Successful	Terminating EC2 instance: i-0dbc6923f1a33c030	At 2025-10-31T05:10:01Z a monitor alarm TargetTracking-tg-for-ec2-group-AlarmLow-0087bd3-28e7-4010-8376-ccdf473cb702 in state ALARM triggered policy 60%, UTILISATION changing the desired capacity from 2 to 1. At 2025-10-31T05:10:11Z an instance was taken out of service in response to a difference between desired and actual capacity, shrinking the capacity from 2 to 1. At 2025-10-31T05:10:12 instance i-0dbc6923f1a33c030 was selected for termination.	2025 October 31, 10:40:11 AM +05:30	2025 October 31, 10:46:14 AM +05:30
Successful	Launching a new EC2 instance: i-02ad95c1fffd45411f	At 2025-10-31T04:50:16Z a user request created an AutoScalingGroup changing the desired capacity from 0 to 2. At 2025-10-31T04:50:16Z an instance was started in response to a difference between desired and actual capacity, increasing the capacity from 0 to 2.	2025 October 31, 10:20:19 AM +05:30	2025 October 31, 10:25:23 AM +05:30
Successful	Launching a new EC2 instance: i-0dbc6923f1a33c030	At 2025-10-31T04:50:16Z a user request created an AutoScalingGroup changing the desired capacity from 0 to 2. At 2025-10-31T04:50:16Z an instance was started in response to a difference between desired and actual capacity, increasing the capacity from 0 to 2.	2025 October 31, 10:20:18 AM +05:30	2025 October 31, 10:25:23 AM +05:30

So now we can see that another ec2 instance was created and it is in running state after sometime it was terminated because the cpu utilization was around 60% and the actual value was 21.1%



This we can see in cloud watch when the alarm was triggered because of a low cpu utilisation.



Initially the new ec2 instance was running and after some time it was automatically terminated. For a low cpu utilization because iam running a static website with no high processes.

The top 2 were the newly

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 IP	Elastic IP	IPv6 IPs
i-0dbx6923f1a33c030	i-0dbx6923f1a33c030	Terminated	t3.micro	-	View alarms +	us-east-1b	-	-	-	-
i-02ad95c1ff45411f	i-02ad95c1ff45411f	Running	t3.micro	3/3 checks passed	View alarms +	us-east-1c	ec2-3-94-6-109.compute-1.amazonaws.com	3.94.6.109	-	-
ELEVATE-LAB...	i-0fb5012231445d3c	Running	t3.micro	3/3 checks passed	View alarms +	us-east-1b	ec2-34-239-245-130.compute-1.amazonaws.com	34.239.245.130	-	-
ELEVATE-LAB...	i-06b9d2b0a177b5369	Stopped	t3.micro	-	View alarms +	us-east-1b	-	-	-	-

The top two where the newly created instances and the topmost one is now terminated after sometime so we can see that based on the 60 % utilization that is around 60 % the ec2 instace was automatically scaled in and whenever we have anything greater than 60 % cpu utilization it is scaled out.

Now if we check the dns public url of load balncer then we first get an apache server website and hten when we refresh the page we get hello world form auto scaling group.

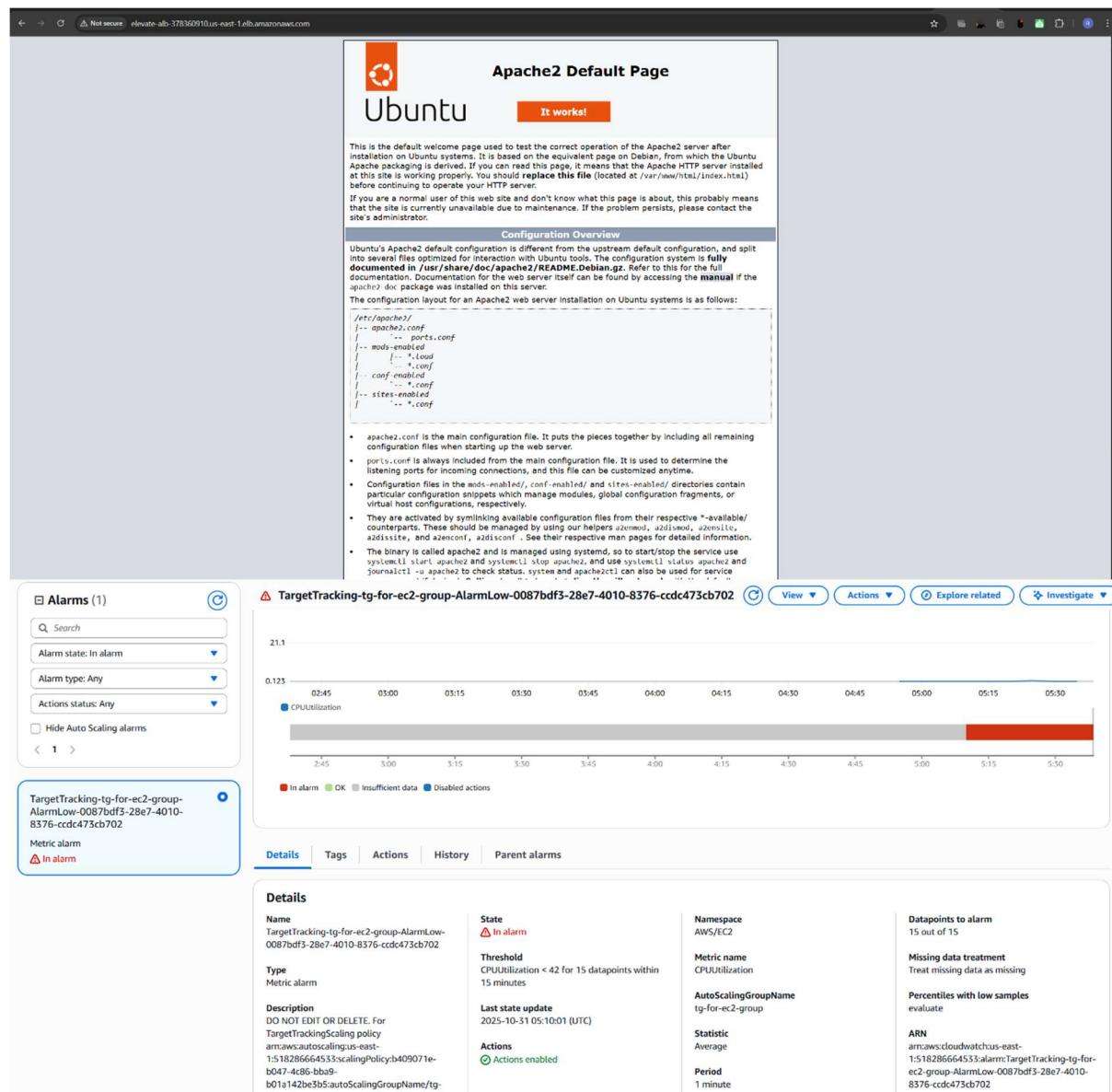
AFTER REFRESHING THE PAGE WE GET

Hello from Auto Scaling EC2

And again we refresh we get the apache server in a round robin fashion.

So this indicates that auto scaling groups can control our website and we can have this so that we can reduce ur costs and utilization can be made in a cheaper manner but more effectively.

The cloud watch metrics looks like this and an alarm will be indicated the below figures represent the cloud watch metrics and shows the cpu utilizations and why the alarm was triggered.



CloudWatch > Alarms > TargetTracking-tg-for-ec2-group-AlarmLow-0087bdf3-28e7-4010-8376-ccdc473cb702

CloudWatch

- Favorites and recents
- Dashboards
- AI Operations New
- Alarms **1** In alarm OK Disabled actions
- All alarms New
- Billing
- Logs
- Metrics All metrics New
- Explorer
- Streams
- Application Signals (APM) New
- GenAI Observability New
- Network Monitoring
- Insights

Alarms (1)

At this time: 2025-10-31 05:10:01 (UTC) Actions View Explore related Investigate

TargetTracking-tg-for-ec2-group-AlarmLow-0087bdf3-28e7-4010-8376-ccdc473cb702

Percent

0.123 21.1 42

02:45 03:00 03:15 03:30 03:45 04:00 04:15 04:30 04:45 05:00 05:15 05:30

CPUUtilization

In alarm OK Insufficient data Disabled actions

TargetTracking-tg-for-ec2-group-AlarmLow-0087bdf3-28e7-4010-8376-ccdc473cb702

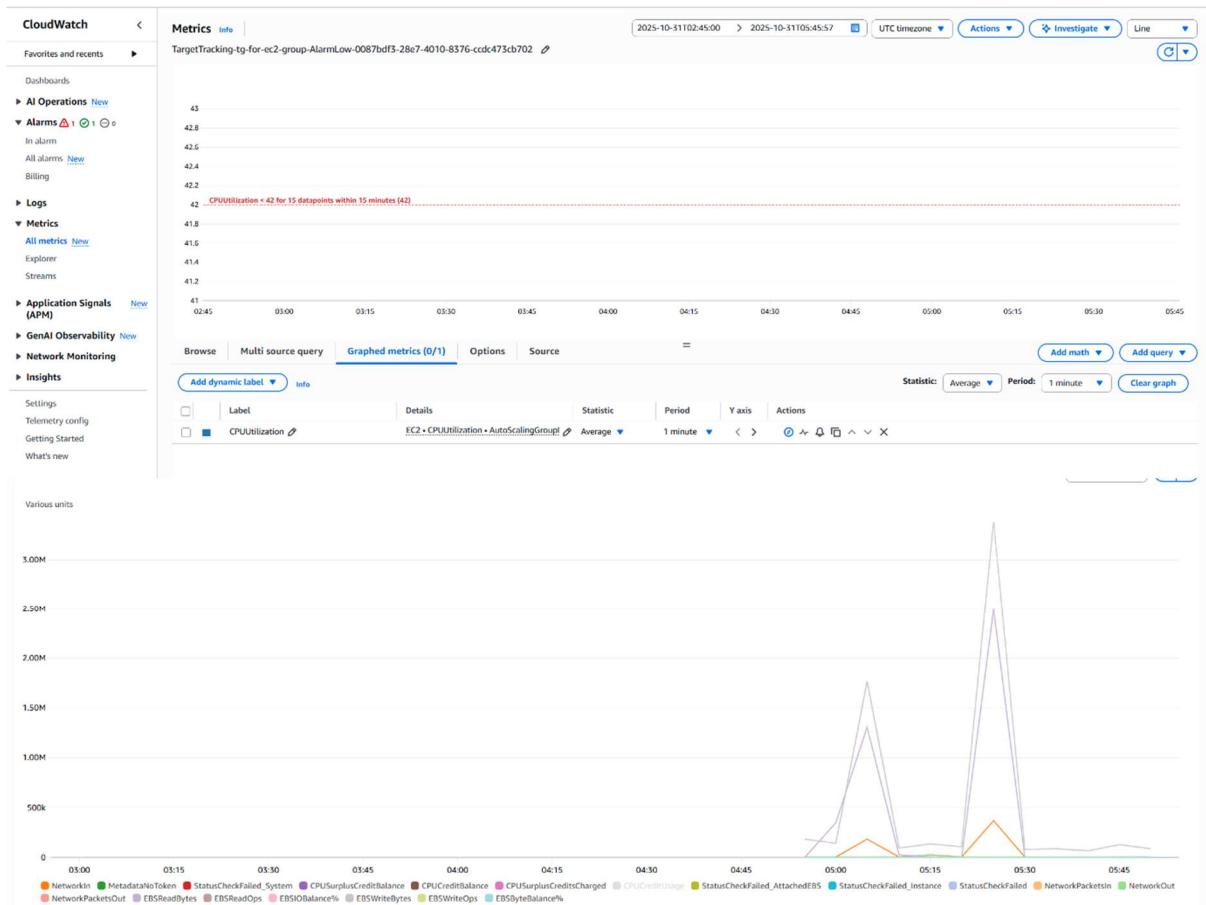
Metric alarm In alarm

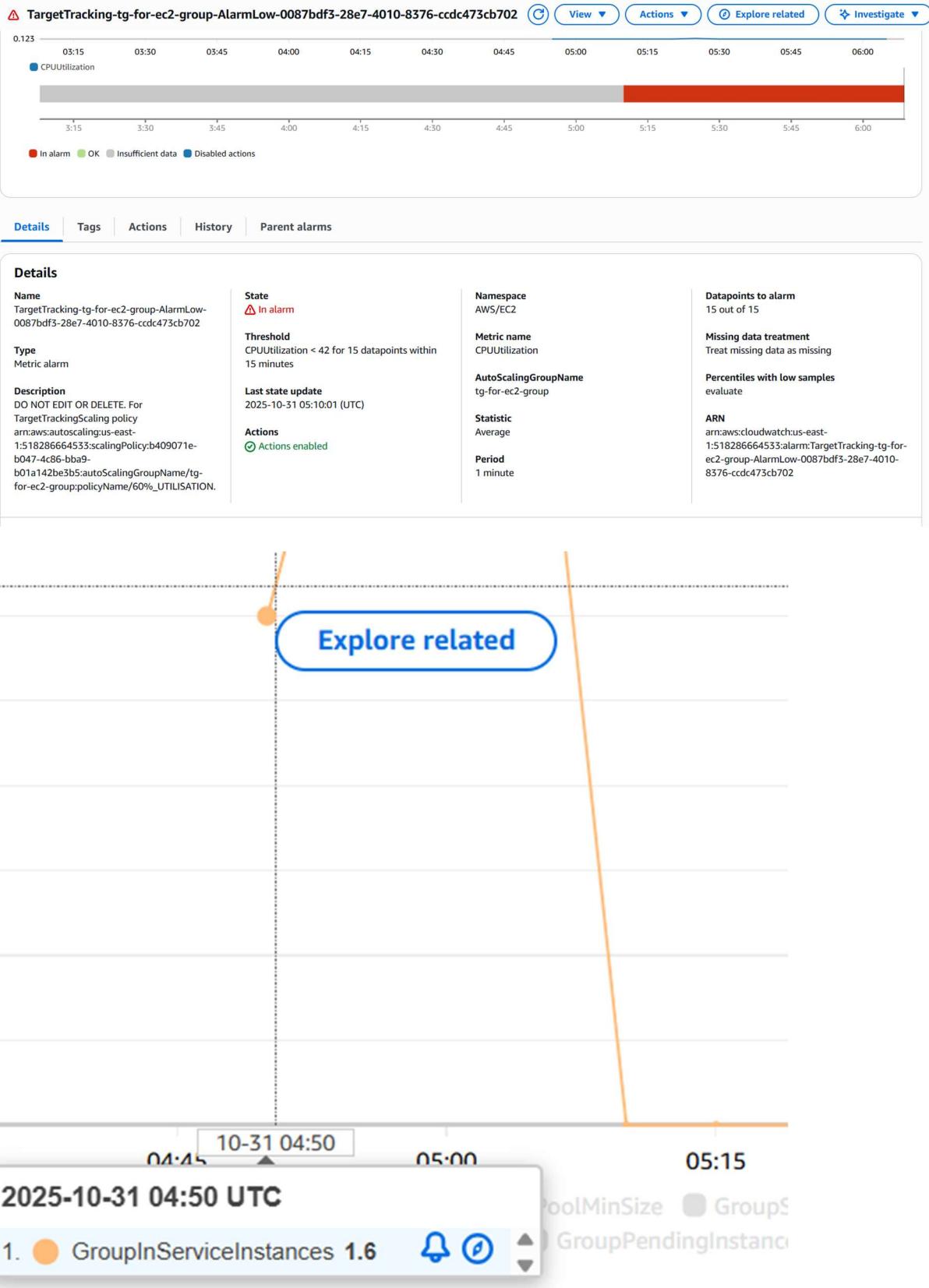
History (4)

At this time: 2025-10-31 05:10:01 (UTC)

Date (UTC)	Type	Description
2025-10-31 05:10:01	Action	Successfully executed action <code>awslogsautoscaling:us-east-1:518286664533:scalingPolicy:b409071e-b047-4c86-bba9-b01a142be3b5:autoScalingGroupName/tg-for-ec2-group;policyName/60%,_UTILISATION</code>
2025-10-31 05:10:01	Configuration update	Alarm "TargetTracking-tg-for-ec2-group-AlarmLow-0087bdf3-28e7-4010-8376-ccdc473cb702" updated
2025-10-31 05:10:01	State update	Alarm updated from Insufficient data to In alarm.
2025-10-31 04:55:56	Configuration update	Alarm "TargetTracking-tg-for-ec2-group-AlarmLow-0087bdf3-28e7-4010-8376-ccdc473cb702" created

Details Tags Actions History Parent alarms







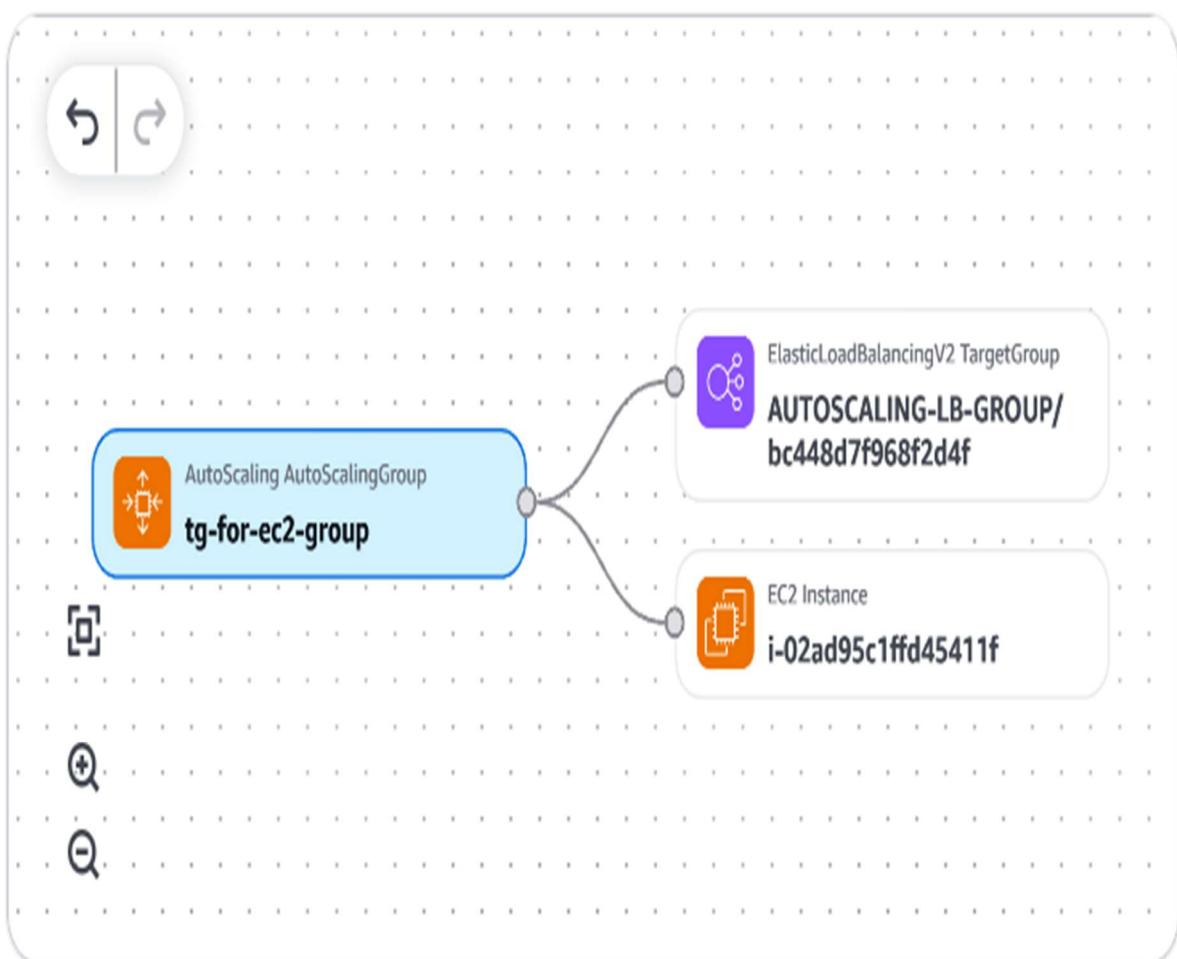
AutoScaling AutoScalingGroup

tg-for-ec2-group



▼ Topology map

Find other resources



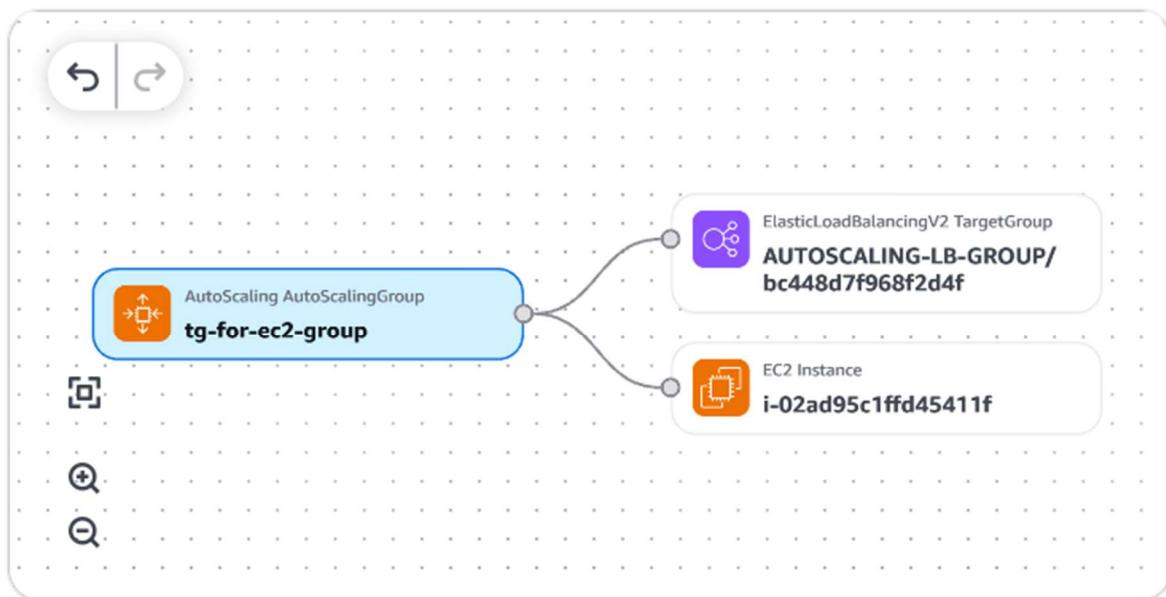
(2025-1... > 2025-1... () Local timezone ▾) (

AutoScaling AutoScalingGroup

tg-for-ec2-group

▼ Topology map

Find other resources



Metrics Logs

Related metrics for Oct 31, 2025 10:19 - Oct 31, 2025 10:29
(LOCAL) at 1 minute intervals

▼ AWS/AutoScaling

GroupInServiceInstances: Average

No unit



GroupDesiredCapacity: Average

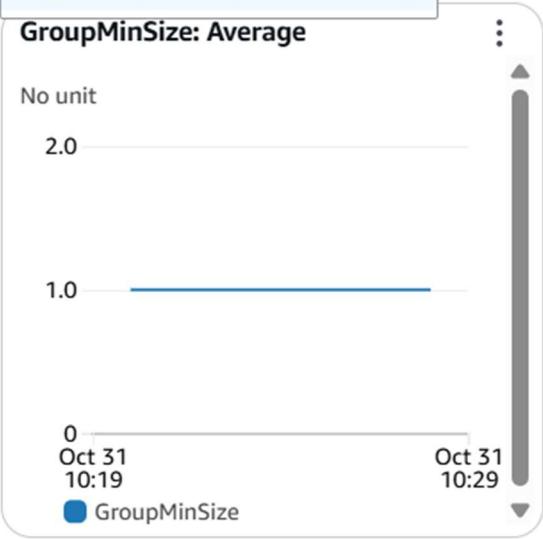
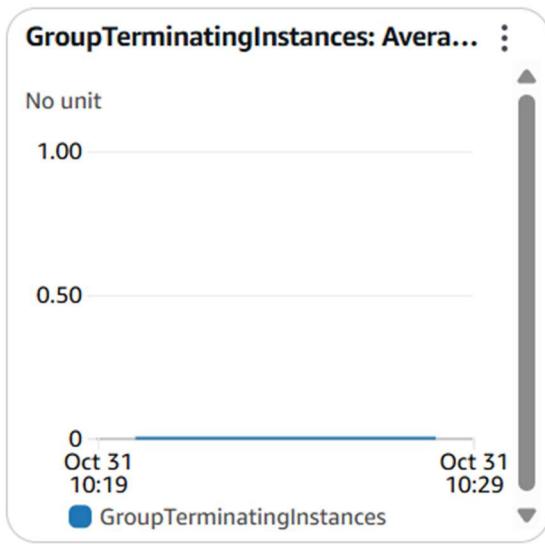
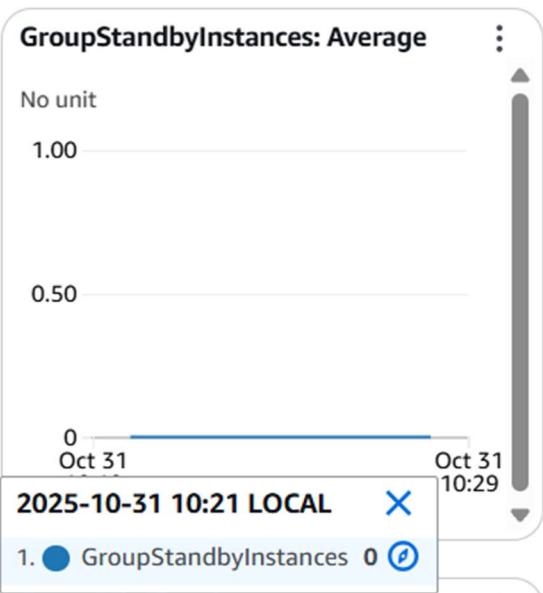
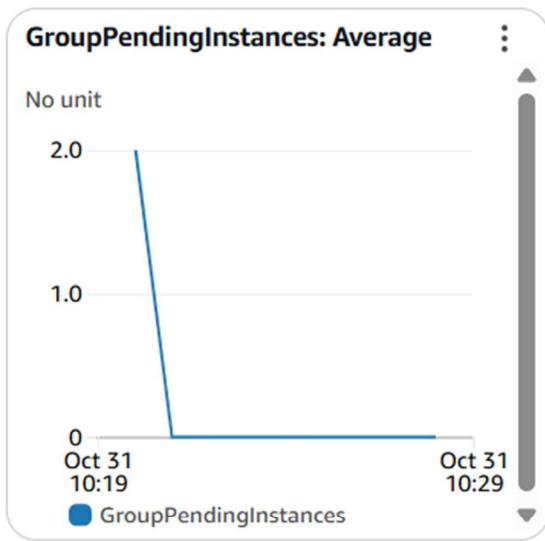
No unit



◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏹ ⏸ ⏹ ⏵

► Topology map ⓘ

🔍 Find other resources



GroupMaxSize: Average

No unit

5.0

4.0

3.0

Oct 31
10:19

Oct 31
10:29

GroupMaxSize

GroupTotalInstances: Average

No unit

3.0

2.0

1.0

Oct 31
10:19

Oct 31
10:29

GroupTotalInstances

GroupAndWarmPoolDesiredCapacity: Average

No unit

3.0

2.0

1.0

Oct 31
10:19

Oct 31
10:29

GroupAndWarmPoolDesiredCapacity

GroupAndWarmPoolTotalCapacity: Average

No unit

3.0

2.0

1.0

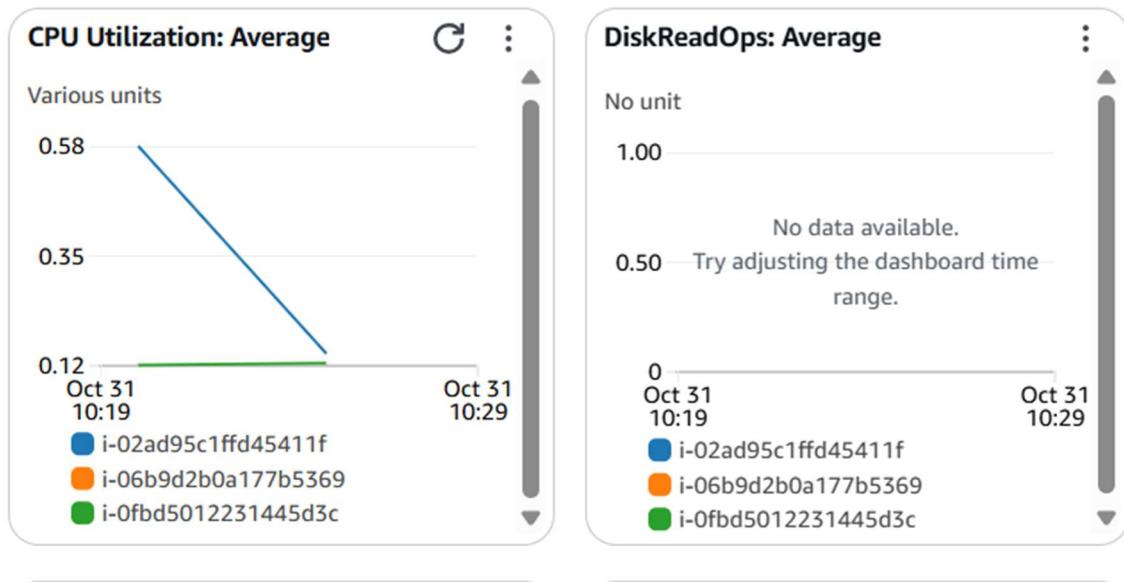
Oct 31
10:19

Oct 31
10:29

GroupAndWarmPoolTotalCapacity

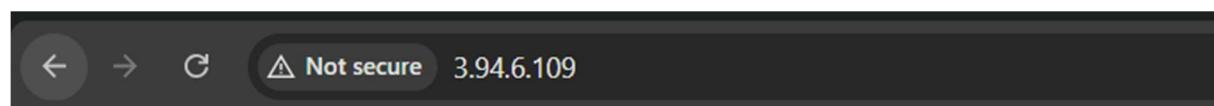
Related metrics for Oct 31, 2025 10:19 - Oct 31, 2025 10:29 (LOCAL) at 1 minute intervals ⓘ

▼ AWS/EC2



So based on this I got to know how to use ec2 instances and how to attach auto scaling groups to my load balancer when I directly created a load balancer I selected existing load balancer and attached The auto scaling group to my alb and I could see that a new ec2 instance was launched and was terminated after some time when it was below 60% cpu utilization. That is why I used the command previously to handle 500 requests and it had 50.5 cpu utilization and was custom made to test the auto scaling groups.

New ec2 instance created from an auto scaling group showing this I did not even do anything in user data script



Hello from Auto Scaling EC2