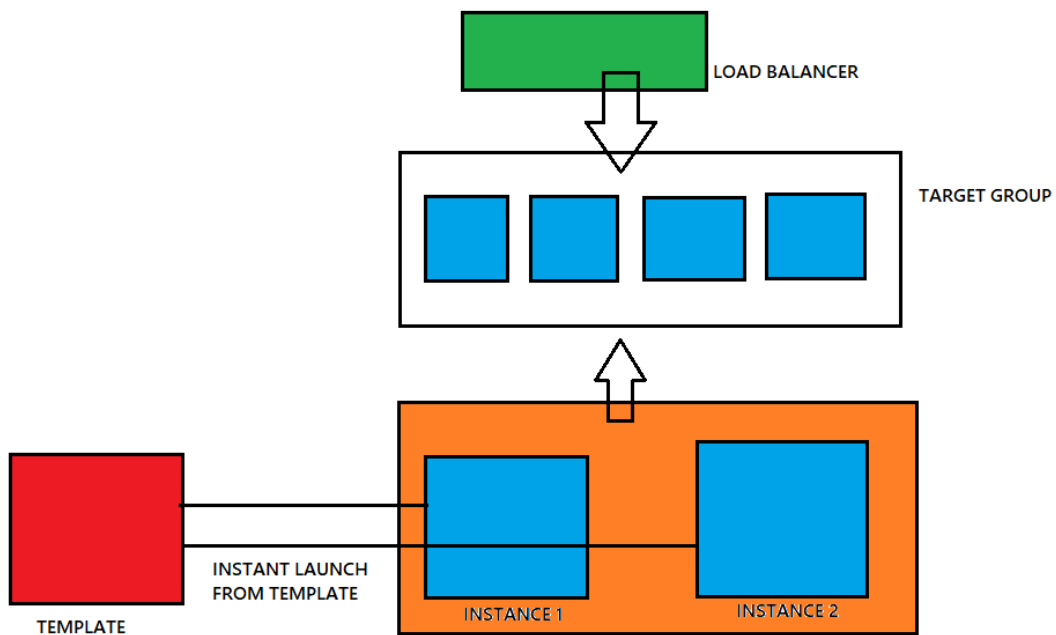
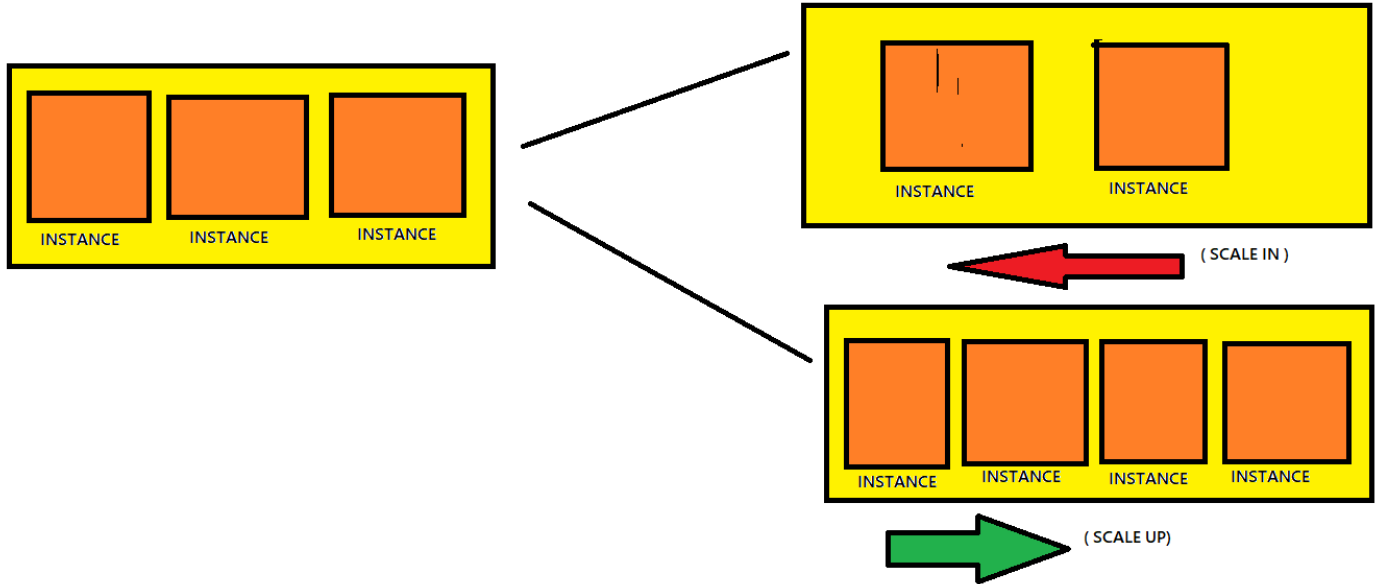


AIM- CONCEPT OF AUTO-SCALING



STEP 1- LAUNCH A TEMPLATE

Gmail

YouTube

aws

Services

Search

[Alt+S]

EC2

VPC

S3

AWS Auto Scaling

Simple Queue Service

Simple Notification Service

EC2 Dashboard

EC2 Global View

Events

Console-to-Code

Preview

Instances

Instances

Instance Types

Launch Templates

Spot Requests

Savings Plans

Reserved Instances

Use launch templates to automate instance launches, simplify permission policies, and enforce best practices across your organization. Save launch parameters in a template that can be used for on-demand launches and with managed services, including EC2 Auto Scaling and EC2 Fleet. Easily update your launch parameters by creating a new launch template version.

New launch template

Create launch template

Benefits and features

Streamline provisioning

Minimize steps to provision instances. With EC2 Auto Scaling, update to a launch template can

Simplify permissions

Create shorter, easier to manage IAM policies. [Learn more](#)

Documentation

[Documentation](#)

EC2 > Launch templates > 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

Launch template name - required

template1

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

Auto Scaling guidance

Info

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

☐ Provide evidence to help us determine templates that leverage with EC2 Auto

Summary

Software Image (AMI)

-

Virtual server type (instance type)

-

Firewall (security group)

-

Storage (volumes)

-

Free tier: In your first year includes 750 hours of t2.micro (or 47 hours in the Region us-east-1)

Cancel

Create launch template

CloudShell

Feedback

© 2021 Amazon Web Services, Inc. or its affiliates. Privacy Terms

Search our full catalog including 1000s of application and OS images

Recents

Quick Start

Don't include in launch template

Amazon Linux

aws

macOS

Mac

Ubuntu

ubuntu®

Windows

Microsoft

Red H

Red

Browse more AMIs

Including AMIs from AWS, Marketplace and the Community

Amazon Machine Image (AMI)

Ubuntu Server 22.04 LTS (HVM), SSD Volume Type

ami-07d9b9ddc6cd8dd30 (64-bit (x86)) / ami-0568072f574d822a4 (64-bit (Arm))

Virtualization: hvm ENA enabled: true Root device type: ebs

Free tier eligible

Description

Canonical, Ubuntu, 22.04 LTS, amd64 jammy image build on 2024-02-07

Architecture

AMI ID

64-bit (x86)

ami-07d9b9ddc6cd8dd30

Verified provider

CloudShell Feedback © 2024, Amazon Web Services,

▼ Summar

Software Imag

Canonical, Ubu

ami-07d9b9ddc6

Virtual server t

-

Firewall (securi

-

Storage (volum

1 volume(s) - 8

Free tie

include:

Cancel

Canonical, Ubuntu, 22.04 LTS, amd64 jammy image build on 2024-02-07

Architecture

AMI ID

64-bit (x86)

ami-07d9b9ddc6cd8dd30

Verified provider

▼ Instance type Info | Get advice

Advanced

Instance type

t2.micro

Family: t2 1 vCPU 1 GiB Memory Current generation: true

On-Demand Windows base pricing: 0.0162 USD per Hour

On-Demand SUSE base pricing: 0.0116 USD per Hour

On-Demand RHEL base pricing: 0.0716 USD per Hour

On-Demand Linux base pricing: 0.0116 USD per Hour

Free tier eligible

All generations

Compare instance types

Additional costs apply for AMIs with pre-installed software

EC2 VPC S3 AWS Auto Scaling Simple Queue Service Simple Notification Service

Subnet [Info](#)

subnet-0886539ca2b844d31
 VPC: vpc-04643f4dea97e29fa Owner: 905418179079 Availability Zone: us-east-1b
 IP addresses available: 4091 CIDR: 172.31.0.0/20

[Create new subnet](#)

When you specify a subnet, a network interface is automatically added to your template.

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.

☒ Select existing security group ☐ Create security group

Common security groups [Info](#)

Select security groups

default sg-0a59205cbaef99134 [X](#)
 VPC: vpc-04643f4dea97e29fa

[Compare security group rules](#)

Security groups that you add or remove here will be added to or removed from all your network interfaces.

[Advanced network configuration](#)

Summary

default

Storage (volumes)

1 volume(s) - 8 GiB

Free tier: In your f... includes 750 hours... t3.micro in the Reg... t2.micro is unavail... usage on free tier /... month, 30 GiB of E... million IOs, 1 GB o... and 100 GB of ban... internet.

[Cancel](#) [Create](#)

EC2 > [Launch templates](#) > Create launch template

Success
 Successfully created [template1\(lt-092c9a051f979ee3b\)](#).

[Actions log](#)

Next Steps

Launch an instance

With On-Demand Instances, you pay for compute capacity by the second (for Linux, with a minimum of 60 seconds) or by the hour (for all other operating systems) with no long-term commitments or upfront payments. Launch an On-Demand Instance from your launch template.

Step 2- launch instance from template

PC S3 AWS Auto Scaling Simple Queue Service Simple Notification Service

Launch Templates (1/1) [Info](#)

Launch Template ID	Launch Template Name	Default Version
lt-092c9a051f979ee3b	template1	1

Actions

- [Create launch template](#)
- [Launch instance from template](#)
- [Modify template \(Create new version\)](#)
- [Delete template](#)
- [Delete template version](#)
- [Set default version](#)
- [Manage tags](#)
- [Create Spot Fleet](#)

The first screenshot shows the 'Choose a launch template' page. It features a 'Source template' section with two dropdown menus: 'template1' (ID: lt-092c9a051f979ee3b) and '1 (Default)'. Below this is the 'Instance details' section, which includes a sub-section for 'Application and OS Images (Amazon Machine Image)'. A summary panel on the right lists configuration details: Number of instances (1), Software Image (AMI), Virtual server type (instance type) (t2.micro), Firewall (security group) (default), and Storage (volumes).

The second screenshot shows the 'Launch instance from template' page. It includes a search bar for AMIs and a 'Quick Start' section with tabs for 'Template or default value', 'Amazon Linux', 'macOS', 'Ubuntu', 'Windows', and 'Red H'. The 'Ubuntu' tab is selected, showing the 'Ubuntu Server 22.04 LTS (HVM), SSD Volume Type' AMI. A summary panel on the right shows the same configuration details as the first screenshot, but with the 'Launch instance' button highlighted in orange.

The third screenshot shows the 'Launch instance from template' page after the instance has been successfully launched. A green banner at the top indicates 'Success' with the message 'Successfully initiated launch of instance (i-0111dd08809eb8ba0)'. Below this is a 'Launch log' section and a 'Next steps' section with a link to 'Get notified of estimated charges'.

Step 3- now go to auto-scaling

Attach to a new load balancer

Define a new load balancer to create for attachment to this Auto Scaling group.

Load balancer type

Choose from the load balancer types offered below. Type selection cannot be changed after the load balancer is created. If you need a different type of load balancer than those offered here, visit the [Load Balancing console](#).

☒ Application Load Balancer
HTTP, HTTPS

☐ Network Load Balancer
TCP, UDP, TLS

Load balancer name

Name cannot be changed after the load balancer is created.

asg1D-1

Load balancer scheme

Scheme cannot be changed after the load balancer is created.

☐ Internal

☒ Internet-facing

Network mapping

Your new load balancer will be created using the same VPC and Availability Zone selections as your Auto Scaling group. You can select different subnets and add subnets from additional Availability Zones.

VPC

vpc-04643f4dea97e29fa

Availability Zones and subnets

You must select a single subnet for each Availability Zone enabled. Only public subnets are available for selection to support DNS

Network mapping

Your new load balancer will be created using the same VPC and Availability Zone selections as your Auto Scaling group. You can select different subnets and add subnets from additional Availability Zones.

VPC

vpc-04643f4dea97e29fa

Availability Zones and subnets

You must select a single subnet for each Availability Zone enabled. Only public subnets are available for selection to support DNS resolution.

<input checked="" type="checkbox"/> us-east-1b	subnet-0886539ca2b844d31
<input checked="" type="checkbox"/> us-east-1a	subnet-0e3cc9d94044f8c92
<input type="checkbox"/> us-east-1f	Select a subnet
<input type="checkbox"/> us-east-1d	Select a subnet
<input type="checkbox"/> us-east-1c	Select a subnet
<input type="checkbox"/> us-east-1e	Select a subnet

Listeners and routing

☐ us-east-1a

Select a subnet



Listeners and routing

If you require secure listeners, or multiple listeners, you can configure them from the [Load Balancing console](#) after your load balancer is created.

Protocol

Port

Default routing (forward to)

HTTP

80

Create a target group



New target group name

An instance target group with default settings will be created.

asg1D-1

Tags - optional

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

Add tag

50 remaining

VPC Lattice integration options [Info](#)

To improve networking capabilities and scalability, integrate your Auto Scaling group with VPC Lattice. VPC Lattice facilitates communications between AWS services and helps you connect and manage your applications across compute services in AWS.

Select VPC Lattice service to attach

☒ No VPC Lattice service

VPC Lattice will not manage your Auto Scaling group's network access and connectivity with other services.

☐ Attach to VPC Lattice service

Incoming requests associated with specified VPC Lattice target groups will be routed to your Auto Scaling group.

To improve networking capabilities and scalability, integrate your Auto Scaling group with VPC Lattice. VPC Lattice facilitates communications between AWS services and helps you connect and manage your applications across compute services in AWS.

Select VPC Lattice service to attach

☒ No VPC Lattice service

VPC Lattice will not manage your Auto Scaling group's network access and connectivity with other services.

☐ Attach to VPC Lattice service

Incoming requests associated with specified VPC Lattice target groups will be routed to your Auto Scaling group.

[Create new VPC Lattice service](#)

Health checks

Health checks increase availability by replacing unhealthy instances. When you use multiple health checks, all are evaluated, and if at least one fails, instance replacement occurs.

EC2 health checks

☒ Always enabled

Additional health check types - optional [Info](#)

☐ Turn on Elastic Load Balancing health checks **Recommended**

Elastic Load Balancing monitors whether instances are available to handle requests. When it reports an unhealthy instance, EC2 Auto Scaling can replace it on its next periodic check.

☐ Turn on VPC Lattice health checks

VPC Lattice can monitor whether instances are available to handle requests. If it considers a target as failed a health check, EC2 Auto Scaling replaces it after its next periodic check.

Health check grace period [Info](#)

This time period delays the first health check until your instances finish initializing. It doesn't prevent an instance from terminating when placed into a non-running state.

300

seconds

nd

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

2

Equal or less than desired capacity

Max desired capacity

4

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.

☒ No scaling policies

Your Auto Scaling group will remain at its initial size and

☐ Target tracking scaling policy

Choose a CloudWatch metric and target value and let the



An instance maintenance policy determines how much availability your application has when EC2 Auto Scaling replaces instances. It also establishes guardrails that limit the amount of capacity that can be added or removed when replacing instances.

Choose a replacement behavior depending on your availability requirements

Mixed behavior

☒ No policy

For rebalancing events, new instances will launch before terminating others. For all other events, instances terminate and launch at the same time.

Prioritize availability

☐ Launch before terminating

Launch new instances and wait for them to be ready before terminating others. This allows you to go above your desired capacity by a given percentage and may temporarily increase costs.

Control costs

☐ Terminate and launch

Terminate and launch instances at the same time. This allows you to go below your desired capacity by a given percentage and may temporarily reduce availability.

Flexible

☐ Custom behavior

Set custom values for the minimum and maximum amount of available capacity. This gives you greater flexibility in setting how far below and over your desired capacity EC2 Auto Scaling goes when replacing instances.

Instance scale-in protection

Scale-in protection prevents newly launched instances from being terminated by scaling activities. Make sure to remove scale-in protection for the group or individual instances when instances are ready to be terminated.

☐ Enable instance scale-in protection

Cancel

Skip to review

Previous

Next

Add tags - optional 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.

options

ons

!

i 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

Cancel

Previous

Next

search

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

24°C

No policy

Instance scale-in protection

Instance scale-in protection

☐ Enable instance protection from scale in

Step 5: Add notifications

Edit

Notifications

No notifications

Step 6: Add tags

Edit

Tags (0)

Key	Value	Tag new instances
-----	-------	-------------------

No tags

Cancel

Previous

Create Auto Scaling group

Auto Scaling groups (1) [Info](#)

Search your Auto Scaling groups

Launch configurations Launch templates Actions **Create Auto Scaling group**

Name	Launch template/configuration	Instances	Status	Desired capacity	Min	Max	Availability Zones
asg1D	template1 Version Default	2	-	2	2	4	us-east-1a, us-east-1b

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.

☐ **No scaling policies**
 Your Auto Scaling group will remain at its initial size and will not dynamically resize to meet demand.

☒ **Target tracking scaling policy**
 Choose a CloudWatch metric and target value and let the scaling policy adjust the desired capacity in proportion to the metric's value.

Scaling policy name

Target Tracking Policy

Metric type [Info](#)

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

Average CPU utilization

Target value

50

Instance warmup [Info](#)

20 seconds

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

Step 4- now go to instance and check the no of min instance

Instances (3) [Info](#)

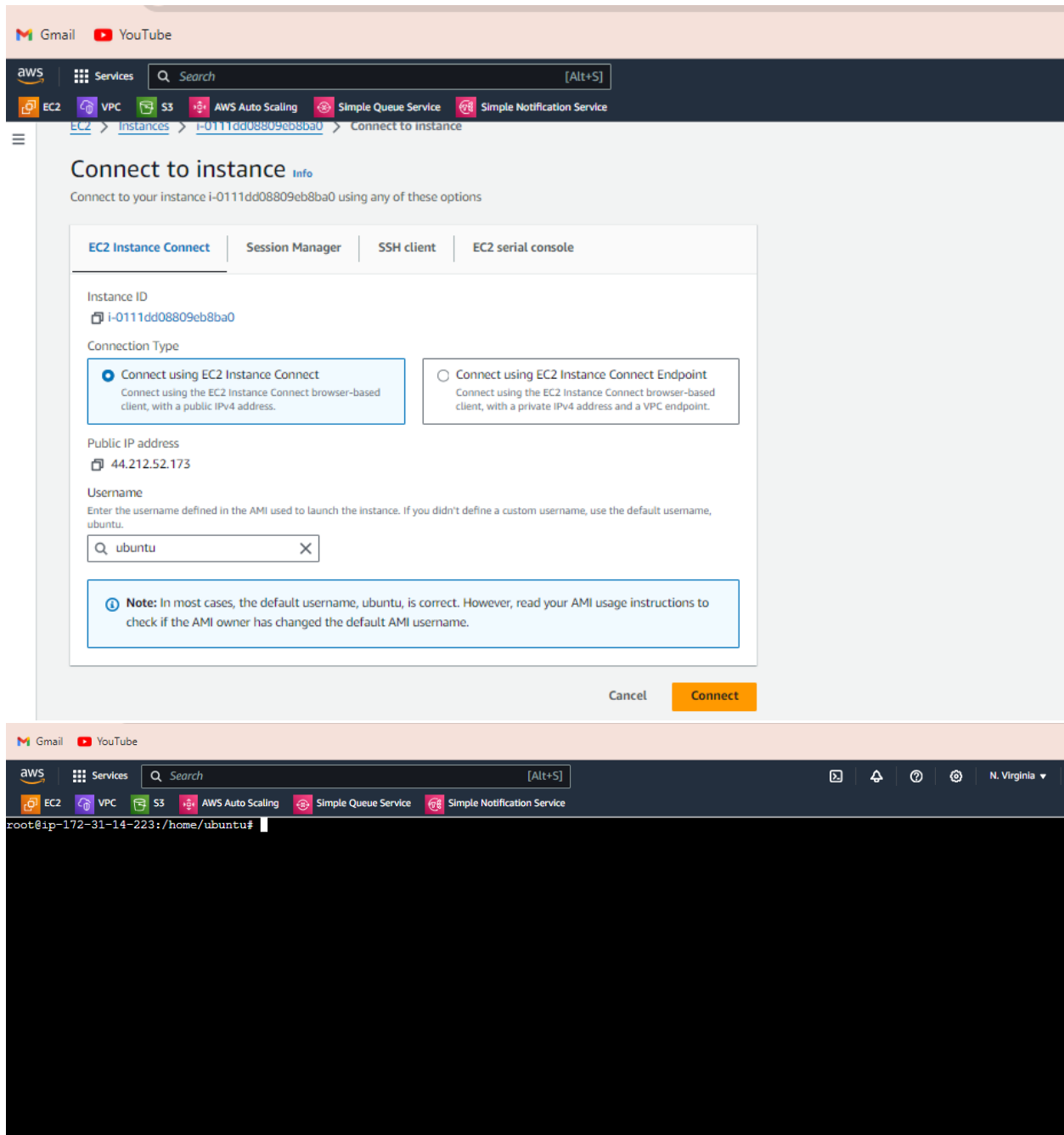
Find Instance by attribute or tag (case-sensitive) Any state

Connect Instance state Actions **Launch instances**

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 ...	Ela
	i-065925bb9992f7c49	Running	t2.micro	Initializing	View alarms	us-east-1a	ec2-54-235-33-21.com...	54.235.33.21	-
	i-0111dd08809eb8ba0	Running	t2.micro	2/2 checks passed	View alarms	us-east-1b	ec2-44-212-52-173.co...	44.212.52.173	-
	i-05353ec9ddd8651a4	Pending	t2.micro	-	View alarms	us-east-1b	ec2-44-205-15-80.com...	44.205.15.80	-

Since we set min 2 instance so we got here now auto scaling is working now its time to set manually high traffic so our auto scale will work and instance will increase

Step 5- now connect to any instance



Run some command to increase traffic

Command – top

```
>/dev/null &
```

```
EC2 VPC S3 AWS Auto Scaling Simple Queue Service Simple Notification Service
%Cpu(s):  5.0 us,  5.0 sy,  0.0 ni, 90.0 id,  0.0 wa,  0.0 hi,  0.0 si,  0.0 st
MiB Mem :  949.7 total,  180.5 free,  183.1 used,  586.1 buff/cache
MiB Swap:   0.0 total,   0.0 free,   0.0 used.  580.4 avail Mem
top - 19:59:24 up 29 min,  2 users,  load average: 0.74, 0.24, 0.09
Tasks: 106 total,  2 running, 103 sleeping,  1 stopped,  0 zombie
%Cpu(s): 64.5 us, 35.5 sy,  0.0 ni,  0.0 id,  0.0 wa,  0.0 hi,  0.0 si,  0.0 st
MiB Mem :  949.7 total,  180.5 free,  183.0 used,  586.1 buff/cache
MiB Swap:   0.0 total,   0.0 free,   0.0 used.  580.4 avail Mem

  PID USER      PR  NI  VIRT  RES  SHR S %CPU  %MEM    TIME+  COMMAND
 13872 root        20   0   6188  1792  1792 R  99.7   0.2   1:18.52 yes
 13861 root        20   0  11896  5504  4736 S   0.3   0.6   0:00.04 sudo
    1 root        20   0 167524 12000  7392 S   0.0   1.2   0:08.05 systemd
    2 root        20   0     0     0     0 S   0.0   0.0   0:00.00 kthreadd
    3 root        0 -20     0     0     0 I   0.0   0.0   0:00.00 rcu_gp
    4 root        0 -20     0     0     0 I   0.0   0.0   0:00.00 rcu_par_gp
    5 root        0 -20     0     0     0 I   0.0   0.0   0:00.00 slub_flushwq
    6 root        0 -20     0     0     0 I   0.0   0.0   0:00.00 netns
    8 root        0 -20     0     0     0 I   0.0   0.0   0:00.00 kworker/0:0H-events_highpri
   10 root        0 -20     0     0     0 I   0.0   0.0   0:00.00 mm_percpu_wq
   11 root        20   0     0     0     0 I   0.0   0.0   0:00.00 rcu_tasks_rude_kthread
   12 root        20   0     0     0     0 I   0.0   0.0   0:00.00 rcu_tasks_trace_kthread
   13 root        20   0     0     0     0 S   0.0   0.0   0:00.33 ksoftirqd/0
   14 root        20   0     0     0     0 S   0.0   0.0   0:00.54 rcu_sched
```

>/dev/null &

Instances (1/3) info								
Find Instance by attribute or tag (case-sensitive)								
	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IP
<input checked="" type="checkbox"/>		i-065925bb9992f7c49	Running	t2.micro	2/2 checks passed	View alarms +	us-east-1a	ec2-54-23
<input type="checkbox"/>		i-0111dd08809eb8ba0	Running	t2.micro	2/2 checks passed	View alarms +	us-east-1b	ec2-44-21
<input type="checkbox"/>		i-05353ec9ddd8651a4	Running	t2.micro	2/2 checks passed	View alarms +	us-east-1b	ec2-44-20

NOW WAIT FOR FEW MINT TO NOTICE YOUR INSTANCE IF IT IS INCREASE OR NOT

Instances (4) info								
Find Instance by attribute or tag (case-sensitive)								
	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IP
<input type="checkbox"/>		i-065925bb9992f7c49	Running	t2.micro	2/2 checks passed	View alarms +	us-east-1a	ec2-54-23
<input type="checkbox"/>		i-0111dd08809eb8ba0	Terminated	t2.micro	-	View alarms +	us-east-1b	-
<input type="checkbox"/>		i-05353ec9ddd8651a4	Terminated	t2.micro	-	View alarms +	us-east-1b	-
<input type="checkbox"/>		i-03eee126cc491c5e6	Running	t2.micro	Initializing	View alarms +	us-east-1b	ec2-44-21