

AWS Auto Scaling Groups – Basics




Requirements

Zero EC2 instances running.

Objectives

A. Create a scaling group for web servers with the following properties:

- Amazon Linux 2 AMI
- t2.micro as the instance type
- user data:

```
 yum install -y httpd  
 systemctl start httpd  
 systemctl enable httpd
```

B. Were new instances created since you created the auto scaling group? How many? Why?

C. Change desired capacity to 2. Did it launch more instances?

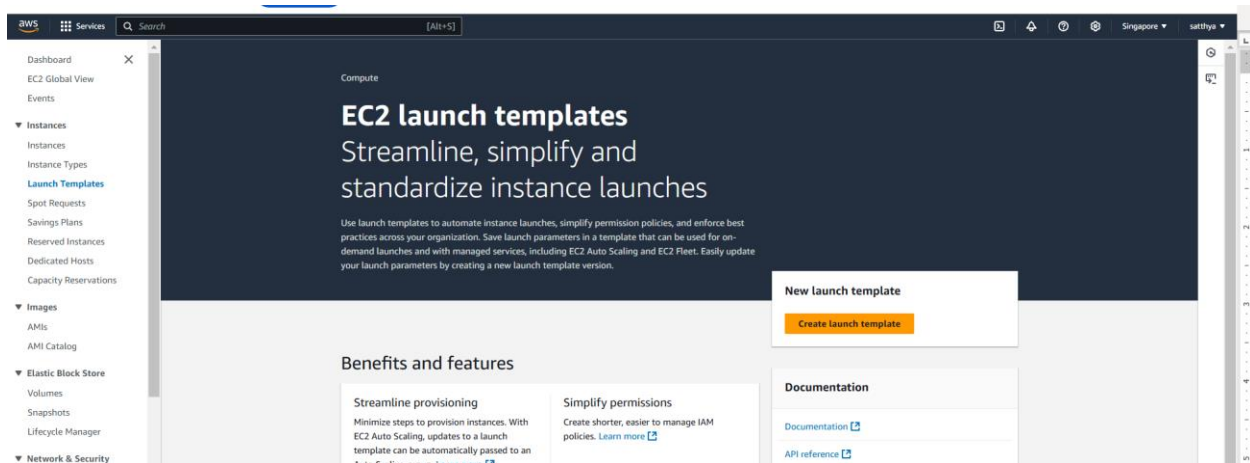
D. Change back the desired capacity to 1. What is the result of this action?

Solution

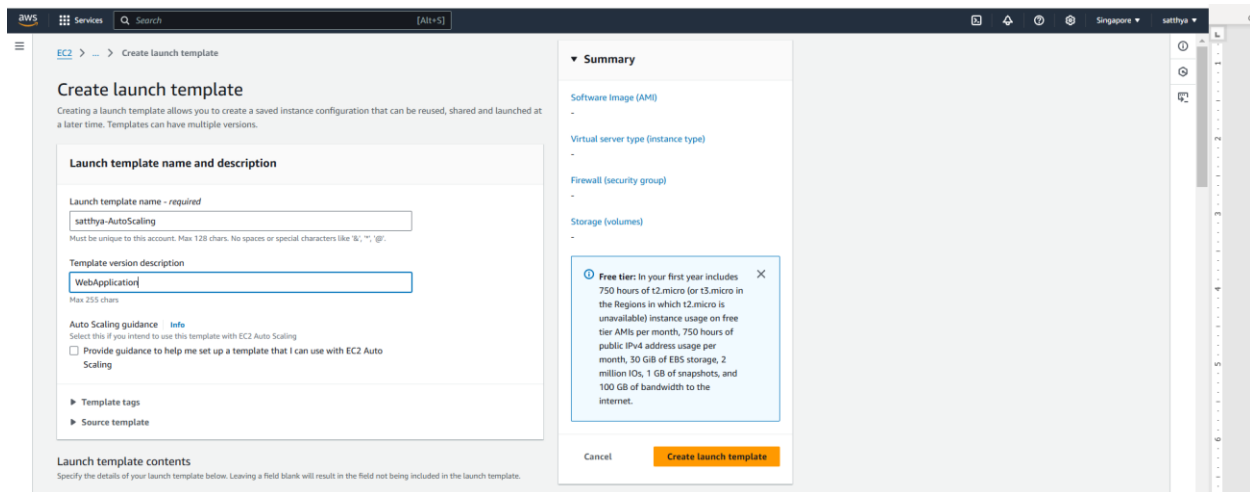
Question(A)

✓ Create template.

- Go to EC2 service.
- Click on Launch Templates



- Create Launch Templates and fill up the templates.



▼ Application and OS Images (Amazon Machine Image) [Info](#)

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. Search or Browse for AMIs if you don't see what you are looking for below

Recents Quick Start

Don't include in launch template

Amazon Linux

aws

macOS

Mac

Ubuntu

ubuntu

Windows

Microsoft

Red H

Red

Search icon

Browse more AMIs

Including AMIs from AWS, Marketplace and the Community

Amazon Machine Image (AMI)

Amazon Linux 2023 AMI

Free tier eligible

ami-07c9c7aaab42cba5a (64-bit (x86), uefi-preferred) / ami-06aa85a9a4db826fe (64-bit (Arm), uefi)

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

Description

Amazon Linux 2023 is a modern, general purpose Linux-based OS that comes with 5 years of long term support. It is optimized for AWS and designed to provide a secure, stable and high-performance execution environment to develop and run your cloud applications.

Amazon Linux 2023 AMI 2023.6.20241111.0 x86_64 HVM kernel-6.1

Architecture	Boot mode	AMI ID	Username	
64-bit (x86)	uefi-preferred	ami-07c9c7aaab42cba5a	ec2-user	<div><div>Verified provider</div></div>

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

Advanced

Instance type

t2.micro

Free tier eligible

Family: t2 1 vCPU 1 GiB Memory Current generation: true
On-Demand Ubuntu Pro base pricing: 0.0164 USD per Hour
On-Demand Linux base pricing: 0.0146 USD per Hour
On-Demand Windows base pricing: 0.0192 USD per Hour
On-Demand RHEL base pricing: 0.029 USD per Hour
On-Demand SUSE base pricing: 0.0146 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

Demo

[Create new key pair](#)

▼ **Network settings** [Info](#)

Subnet [Info](#)

subnet-0509b9524002b4fad

VPC: vpc-014ccf5a14583eefe Owner: 396913698392
Availability Zone: ap-southeast-1b Zone type: Availability Zone
IP addresses available: 4086 CIDR: 172.31.16.0/20

[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-09eccaa2f47883fe5 X
VPC: vpc-014ccf5a14583eefe

[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**

▼ **Storage (volumes)** [Info](#)

EBS Volumes

[Hide details](#)

- ▶ **Volume 1 (AMI Root) (8 GiB, EBS, General purpose SSD (gp3))**
AMI Volumes are not included in the template unless modified

 Free tier eligible customers can get up to 30 GB of EBS General Purpose (SSD) or Magnetic storage 

[Add new volume](#)

- Advance details key in user data
- Create the template.

User data - optional [Info](#)

Upload a file with your user data or enter it in the field.

 [Choose file](#)

```
yum install -y httpd
systemctl start httpd
systemctl enable httpd
```

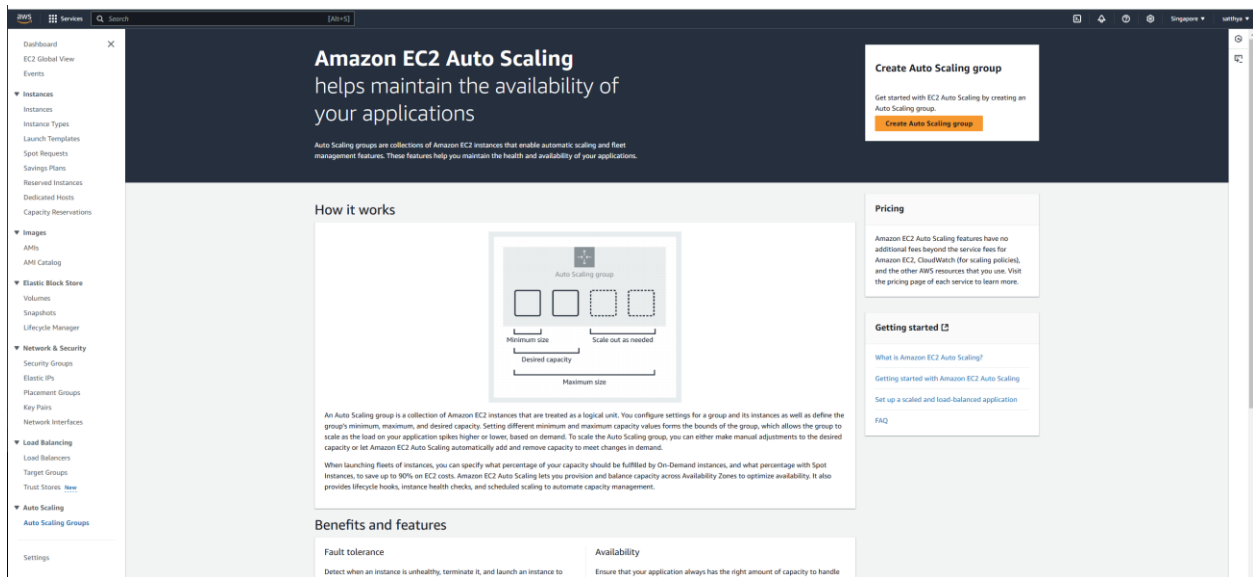


☐ User data has already been base64 encoded

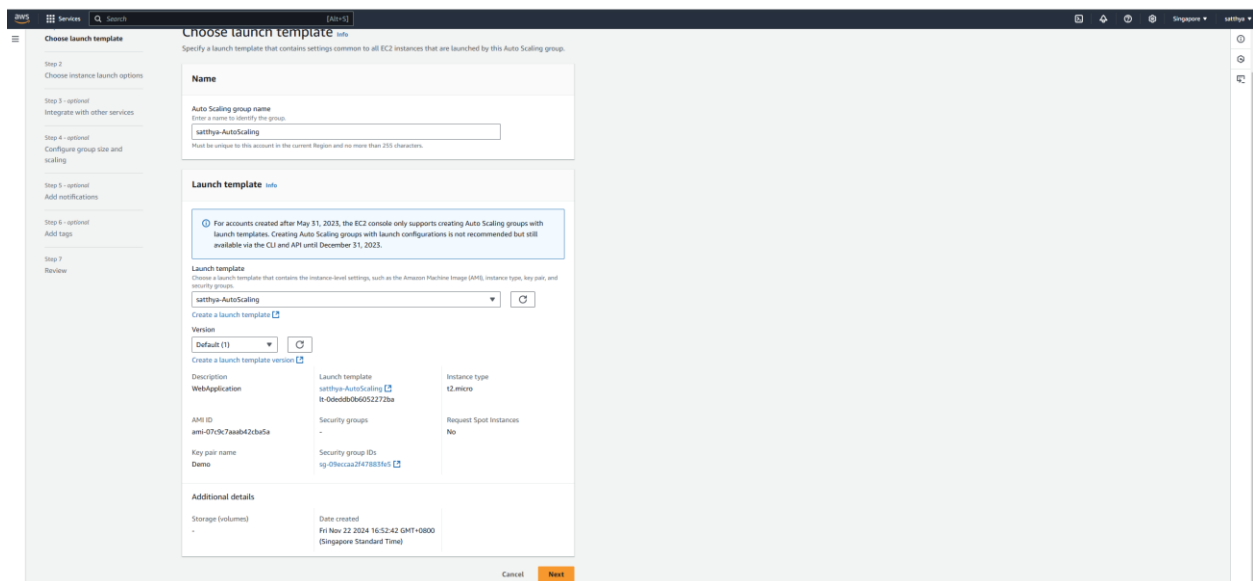


✓ Create Auto Scaling

• Create Auto Scaling



• Select the template and click next.



- Choose in which AZs to launch and click on "Next"

Choose instance launch options info

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

Instance type requirements info Override launch template

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

Launch template	Version	Description
cat-type-AutoScaling	Default	WebApplication

Instance type: t2.micro

Network info

For most applications, you can use multiple Availability Zones and let EC2 Auto Scaling balance your instances across the zones. The default VPC and default subnets are suitable for getting started quickly.

VPC

Choose the VPC that defines the virtual network for your Auto Scaling group.

vpc-014c0fa14583e0fe
172.31.0.0/20 - Default

[Create a VPC](#)

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

ap-southeast-1a | subnet-0126f7c3f08212f12 X
172.31.0.0/20 - Default

ap-southeast-1b | subnet-050b9534002b48ad X
172.31.16.0/20 - Default

[Create a subnet](#)

Availability Zone distribution - new

Auto Scaling automatically balances instances across Availability Zones. If launch failures occur in a zone, select a strategy.

☒ **Balanced best effort**
If launches fail in one Availability Zone, Auto Scaling will attempt to launch in another healthy Availability Zone.

☐ **Balanced only**
If launches fail in one Availability Zone, Auto Scaling will continue to attempt to launch in the unhealthy Availability Zone to preserve balanced distribution.

Cancel Skip to review Previous Next

- Attached it with load balance

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.

☐ 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 for your Auto Scaling group.

☒ **Choose from your load balancer target groups**
This option allows you to attach Applications, 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

Demo-Group HTTP
Application Load Balancer: Demo-ALB X

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

[Create new VPC Lattice service](#)

- Set group size and scaling

The screenshot shows the 'Configure group size and scaling' step in the AWS Management Console. The page is divided into three main sections: 'Group size', 'Desired capacity', and 'Scaling'.

- Group size:** A dropdown menu for 'Units (number of instances)' is set to '0'.
- Desired capacity:** A text input field for 'Specify your group size' is set to '0'.
- Scaling:**
 - Scaling limits:** Two input fields for 'Min desired capacity' and 'Max desired capacity' are both set to '0'.
 - Automatic scaling - optional:** Two radio buttons are present: 'No scaling policies' (selected) and 'Target tracking scaling policy'.

- Final step : review and create

The screenshot shows the 'Auto Scaling groups' page in the AWS Management Console. It features a table with the following columns: Name, Launch template/configuration, Instances, Status, Desired capacity, Min, Max, and Availability Zones.

Name	Launch template/configuration	Instances	Status	Desired capacity	Min	Max	Availability Zones
sutthya-AutoScaling	sutthya-AutoScaling Version Default	0	-	0	0	3	ap-southeast-1a, ap-southeast-1b

Question(B)

The screenshot shows the 'Instances' page in the AWS Management Console. It features a table with the following columns: Name, Instance ID, Instance state, Instance type, Status check, Alarm status, Availability Zone, Public IPv4 DNS, Public IPv4 ..., Elastic IP, IPv6 IPs, Monitoring, and Security group name.

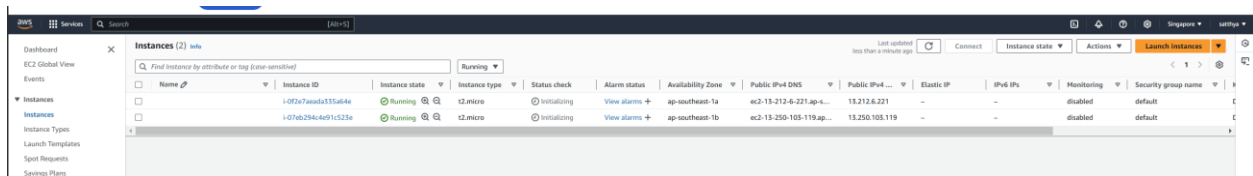
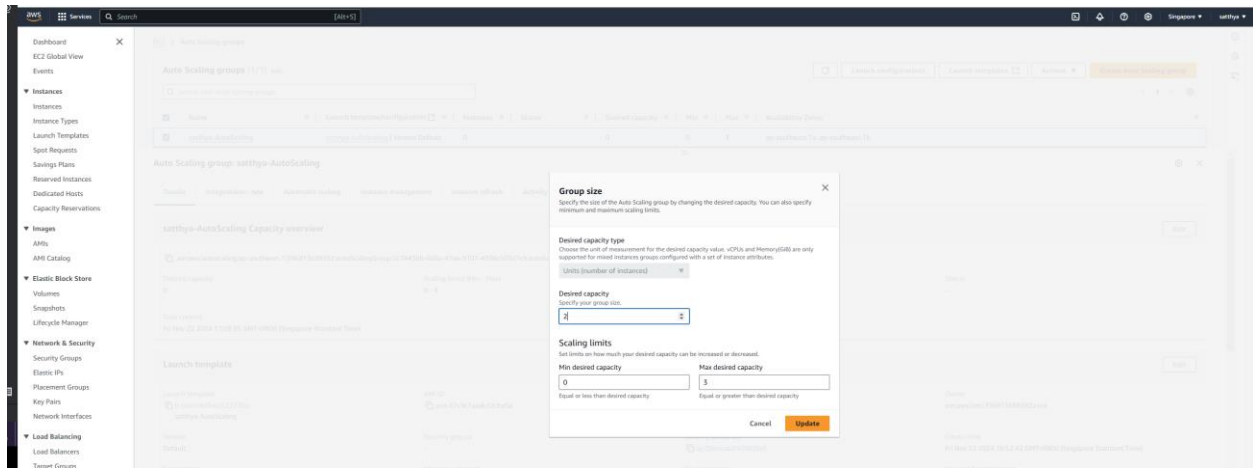
Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 ...	Elastic IP	IPv6 IPs	Monitoring	Security group name
No matching instances found												

Were new instances created since you created the auto scaling group? How many? Why?

Answer : No instances created. Because we set the desire Capacity to 0

Question(C)

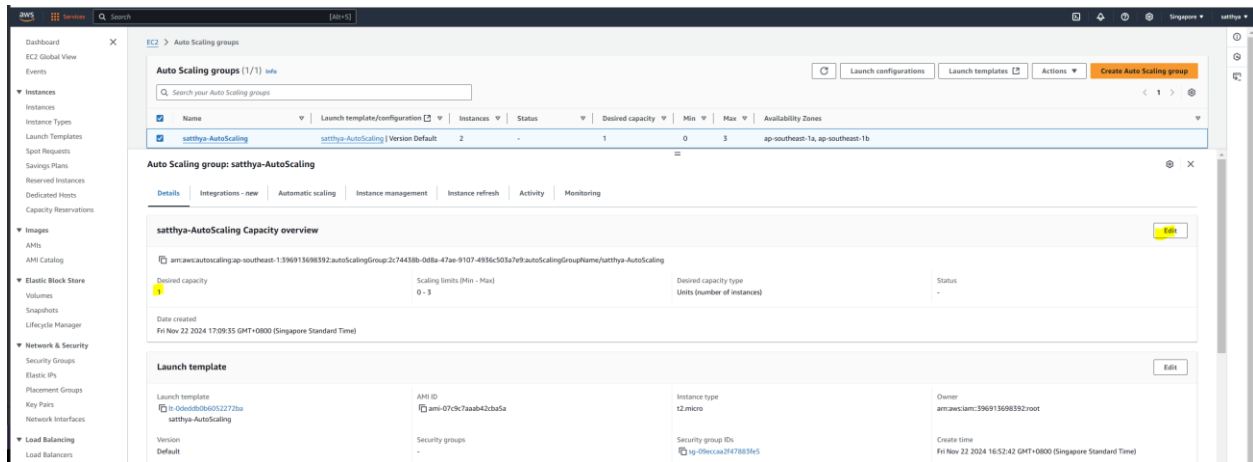
Change desired capacity to 2. Did it launch more instances?



Answer :- It launch 2 instances. Because desired capacity was increate to 2

Question(D)

Change back the desired capacity to 1. What is the result of this action?



The screenshot shows the AWS Management Console for the 'Auto Scaling groups' page. The group 'sattya-AutoScaling' is selected. The 'Desired capacity' is set to 1. The 'Launch template' section shows the configuration for the group.

Name	Launch template/configuration	Instances	Status	Desired capacity	Min	Max	Availability Zones
sattya-AutoScaling	sattya-AutoScaling	Version Default	2	1	0	3	ap-southeast-1a, ap-southeast-1b

Auto Scaling group: sattya-AutoScaling

sattya-AutoScaling Capacity overview

Desired capacity	Scaling limits (Min - Max)	Desired capacity type	Status
0 - 3	0 - 3	Units (number of instances)	-

Date created: Fri Nov 22 2024 17:09:35 GMT+0800 (Singapore Standard Time)

Launch template

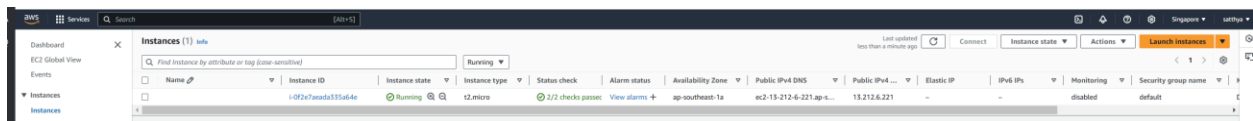
Launch template	AMI ID	Instance type	Owner
i-0d0c0b6052272ba sattya-AutoScaling	ami-07dc7aaab423ba5a	t2.micro	amazon:arn:396915698192root

Version: Default

Security groups: -

Security group IDs: sg-0f6ccaa29478839f5

Create time: Fri Nov 22 2024 16:52:42 GMT+0800 (Singapore Standard Time)



The screenshot shows the AWS Management Console for the 'Instances' page. The instance 'i-0f2a7eada135a64e' is shown in the 'Running' state.

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 ...	Elastic IP	IPv6 IPs	Monitoring	Security group name
i-0f2a7eada135a64e	i-0f2a7eada135a64e	Running	t2.micro	2/2 checks passed	View alarms	ap-southeast-1a	ec2-15-212-6-221.ap-s...	15.212.6.221	-	-	disabled	default

Answer :-It will terminate and run only one instances. Because desired capacity was change to 1.

- note:- if instance not terminate immediately, please check the cool down period