

Lesson 03 Demo 10

Configuring the Manual and Dynamic Scaling

Objective: To understand how to configure manual and dynamic scaling for your application using Amazon Web Services (AWS) tools and services.

Tools required: AWS workspace

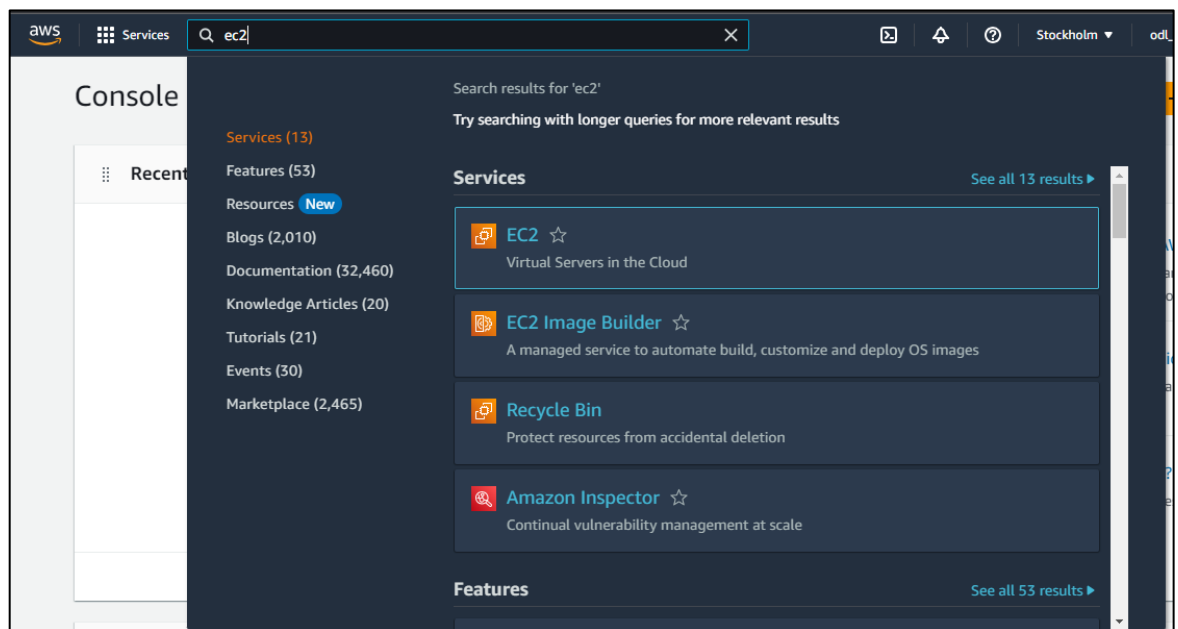
Prerequisites: Create an EC2 instance named S3

Steps to be followed:

1. Set up predefined auto scaling group
2. Set up EC2 Auto Scaling with Load Balancer

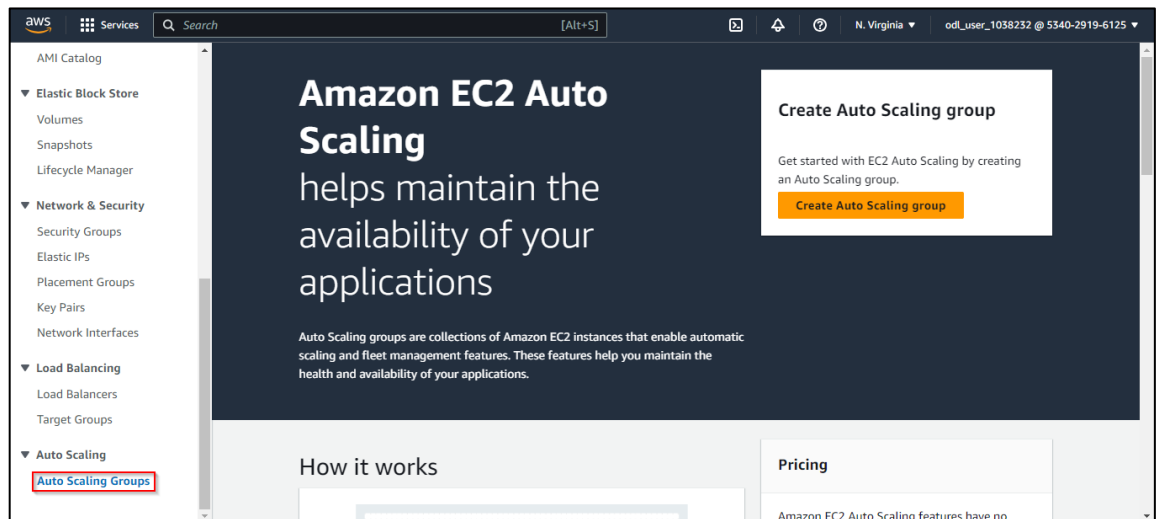
Step 1: Set up predefined auto scaling group

1.1 Open the AWS console, and search and click on **EC2**

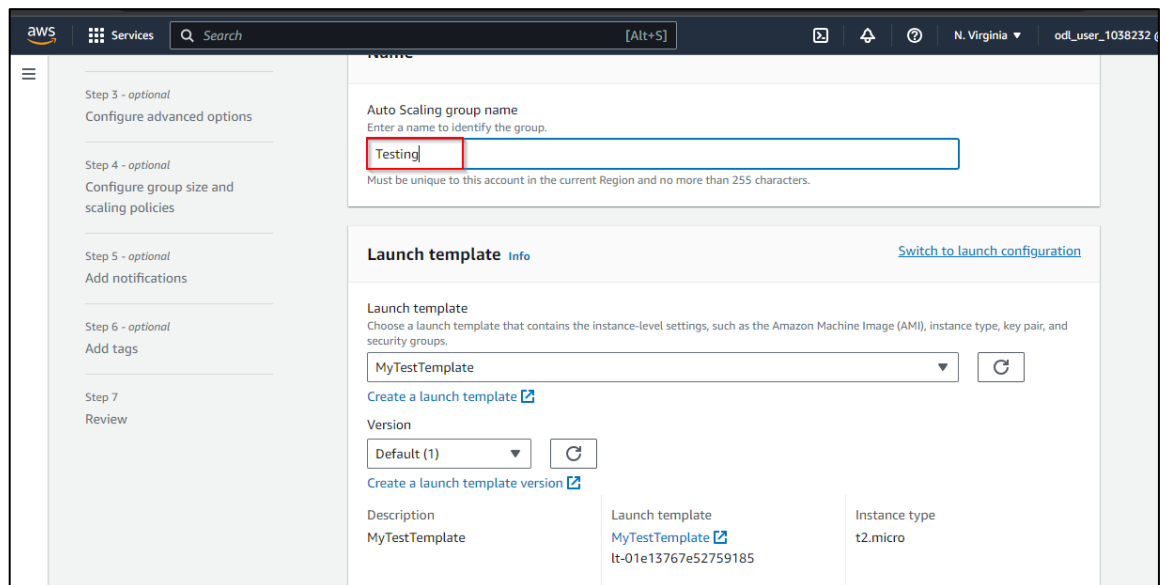


Set the region **US East (N. Virginia) us-east-1** in all the demos.

1.2 Click on **Create Auto Scaling group**



1.3 Enter the name as **Testing**, select **MyTestTemplate** in the Launch template, and click on **Next**



1.4 Click on the Availability Zones and subnets as **us-east-1a** and **us-east-1b**, click on **Next**

For most applications, you can use multiple Availability Zones and let AWS 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-0b520898d9ec3b795
172.31.0.0/16 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

us-east-1a | subnet-062cdfb96d3c2ff
172.31.16.0/20 Default

us-east-1b | subnet-0318f2926629127df
172.31.32.0/20 Default

[Create a subnet](#)

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
MyTestTemplate lt-01e13767e52759185	Default	MyTestTemplate

Instance type
t2.micro

Cancel Skip to review Previous **Next**

1.5 Enter the **Desired capacity**, **Minimum capacity**, and **Maximum capacity** as **2**, and click on **Next**

Configure group size and scaling policies - optional [Info](#)

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

Group size - optional [Info](#)

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

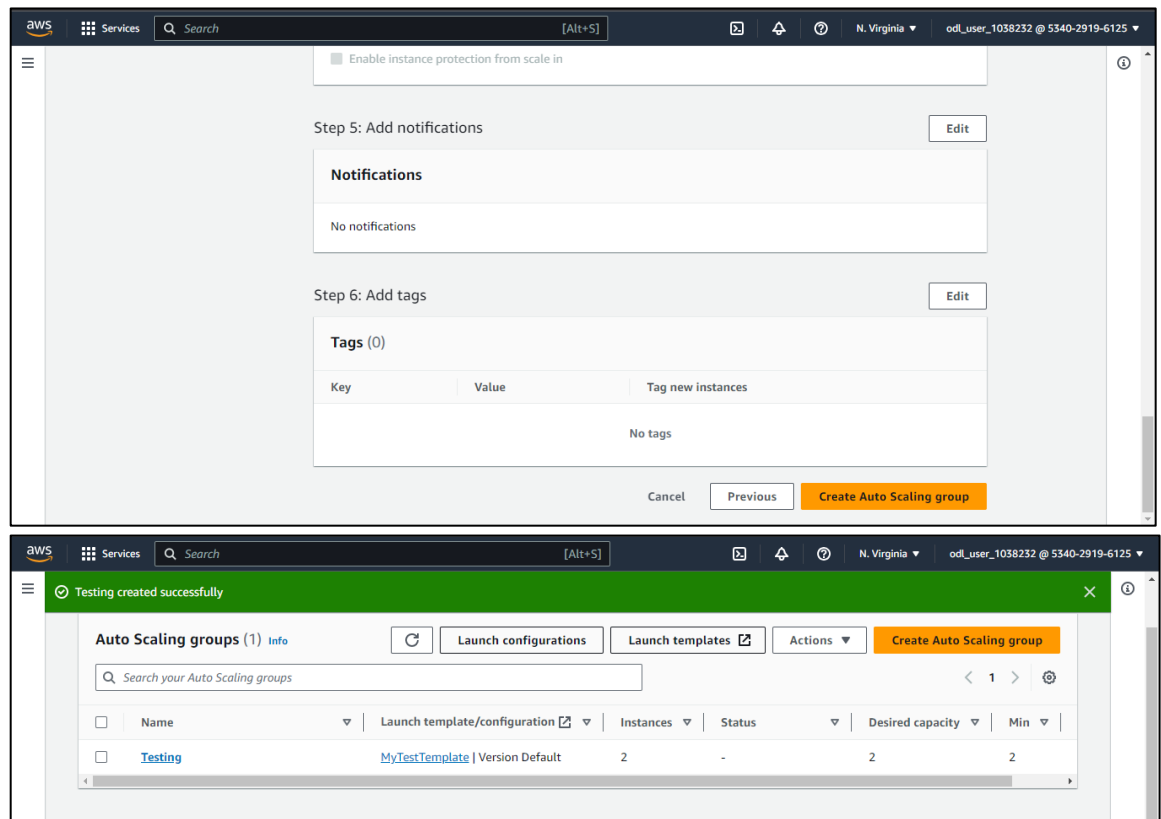
Desired capacity
2

Minimum capacity
2

Maximum capacity
2

Scaling policies - optional

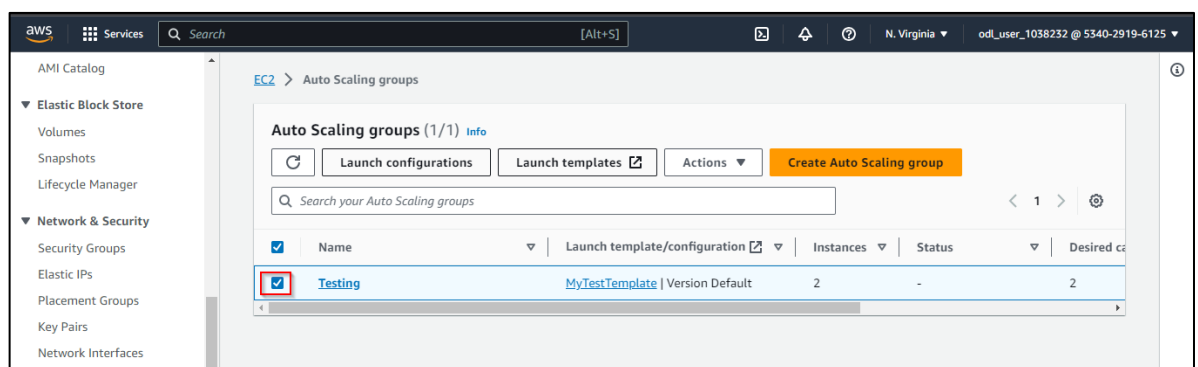
1.6 Review the steps, and click on the **Create Auto Scaling group**



Auto scaling groups are created successfully.

Step 2: Set up EC2 Auto Scaling with Load Balancer

2.1 Open the **Auto scaling groups** section created previously



2.2 Click on **Details** to verify the group details

The screenshot shows the AWS Management Console interface. On the left, the navigation pane is visible with categories like Elastic Block Store, Network & Security, Load Balancing, and Auto Scaling. The 'Auto Scaling Groups' link is highlighted. The main content area shows the 'Auto Scaling groups (1/1)' page. A search bar is present, and a table lists the groups. The 'Testing' group is selected, and its 'Details' tab is active. The 'Group details' section shows the following information:

Property	Value
Auto Scaling group name	Testing
Desired capacity	2
Status	-
Amazon Resource Name (ARN)	arn:aws:autoscaling:us-east-1:534029196125:autoScalingGroup:8e0aa92d-7d52-435b-9c25-246bba5270d3:autoScalingGroupNa
Date created	Thu Aug 10 2023 12:57:15
Minimum capacity	2

2.3 Click on **Instances**, and then create a new instance by clicking on Launch instances

The first screenshot shows the 'Instances (3)' page with three instances listed:

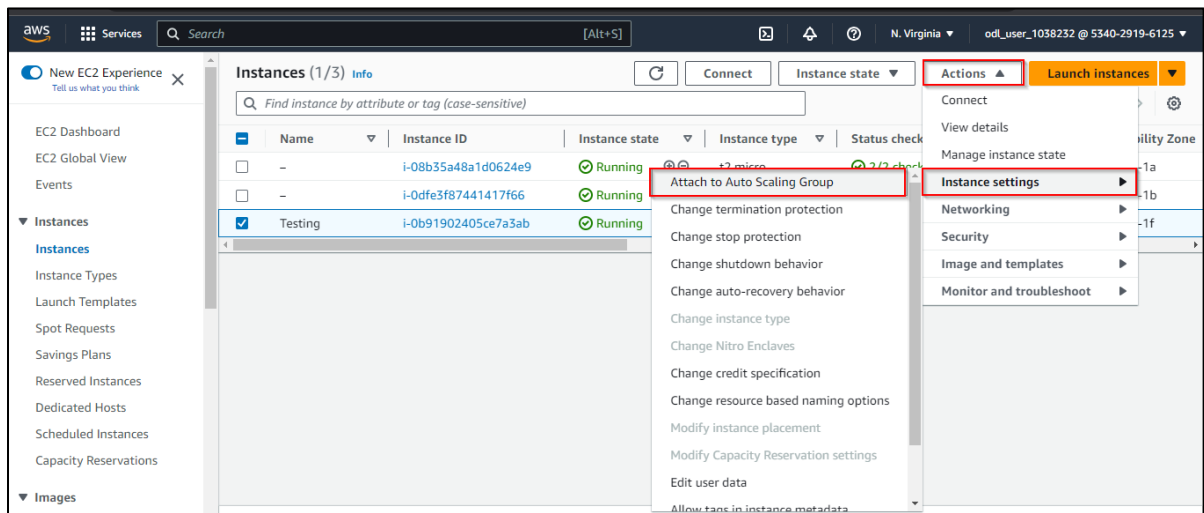
Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone
-	i-02516899c927ca8e5	Terminated	t2.micro	-	No alarms	us-east-1a
-	i-08b35a48a1d0624e9	Running	t2.micro	2/2 checks passed	No alarms	us-east-1a
-	i-0dfe3f87441417f66	Running	t2.micro	2/2 checks passed	No alarms	us-east-1b

The second screenshot shows the 'Instances (4)' page after a new instance has been launched. The 'Testing' instance is now listed with a 'Pending' state:

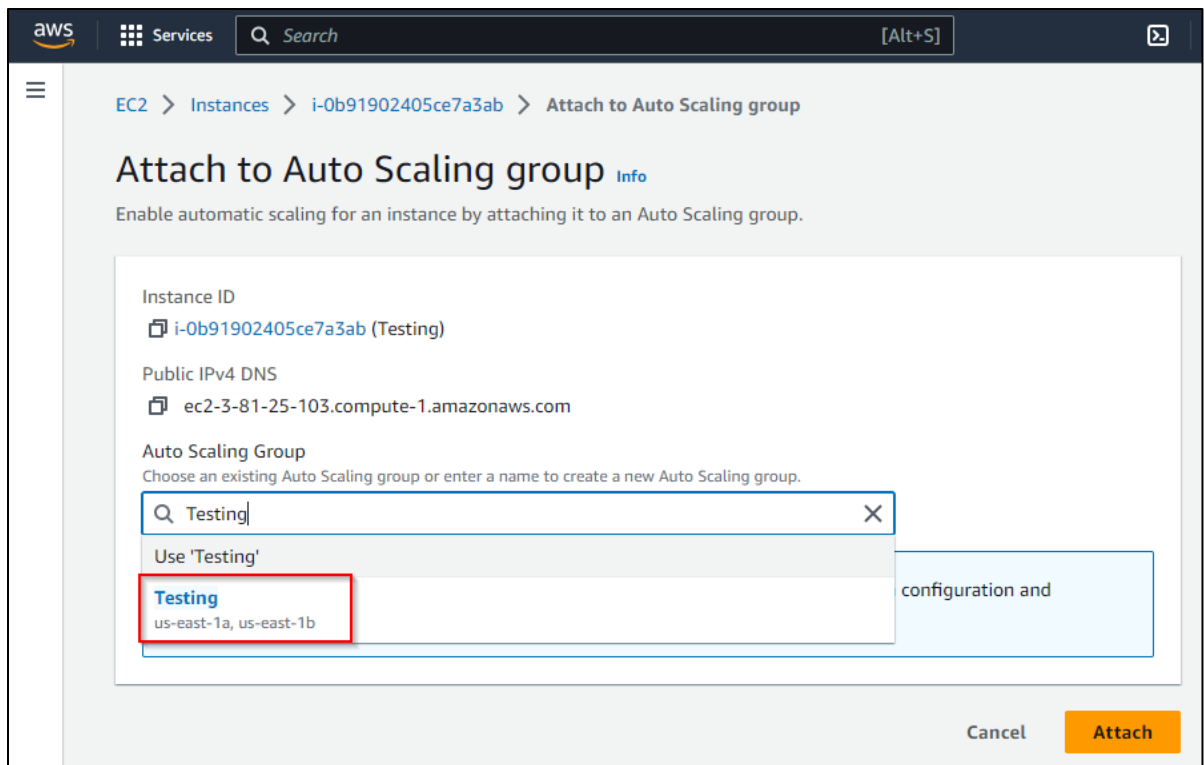
Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone
-	i-02516899c927ca8e5	Terminated	t2.micro	-	No alarms	us-east-1a
-	i-08b35a48a1d0624e9	Running	t2.micro	2/2 checks passed	No alarms	us-east-1a
-	i-0dfe3f87441417f66	Running	t2.micro	2/2 checks passed	No alarms	us-east-1b
Testing	i-0b91902405ce7a3ab	Pending	t2.micro	-	No alarms	us-east-1f

Instances are created successfully now; check out the previous demos to learn how to create instances.

2.4 Click on **Actions** > **Instance settings** > **Attach to Auto Scaling group**



2.5 Select the Auto Scaling Group name, **Testing**; click on the **Attach** button



Use different Auto Scaling group names.

We navigated through the process of configuring both manual and dynamic scaling techniques within the AWS workspace.

By following these steps, you will be able to successfully configure manual and dynamic scaling techniques. They provide the flexibility to manage resources efficiently, ensuring that your application remains responsive and available to users while adapting to changing traffic patterns and demands.