Amazon Auto Scaling Groups

A Message Queuing Service

Case Study



**Introduction**

Amazon EC2 Auto Scaling helps you ensure that you have the correct number of Amazon EC2 instances available to handle the load for your application. You create collections of EC2 instances, called Auto Scaling groups. You can specify the minimum number of instances in each Auto Scaling group, and Amazon EC2 Auto Scaling ensures that your group never goes below this size. You can specify the maximum number of instances in each Auto Scaling group, and Amazon EC2 Auto Scaling ensures that your group never goes above this size. If you specify the desired capacity, either when you create the group or at any time thereafter, Amazon EC2 Auto Scaling ensures that your group has this many instances. If you specify scaling policies, then Amazon EC2 Auto Scaling can launch or terminate instances as demand on your application increases or decreases.

For example, the following Auto Scaling group has a minimum size of one instance, a desired capacity of two instances, and a maximum size of four instances. The scaling policies that you define adjust the number of instances, within your minimum and maximum number of instances, based on the criteria that you specify.


   An illustration of a basic Auto Scaling group.
  

When you use Amazon EC2 Auto Scaling, you must use certain building blocks to get started. This tutorial walks you through the process for setting up building blocks to create a basic infrastructure for Amazon EC2 Auto Scaling.

Before you create an Auto Scaling group for use with your application, review your application thoroughly as it runs in the AWS Cloud. Consider the following:

* How many Availability Zones the Auto Scaling group should span.
* What existing resources can be used, such as security groups or Amazon Machine Images (AMIs).
* Whether you want to scale to increase or decrease capacity, or if you just want to ensure that a specific number of servers are always running. Keep in mind that Amazon EC2 Auto Scaling can do both simultaneously.
* What metrics have the most relevance to your application's performance.
* How long it takes to launch and configure a server.

The better you understand your application, the more effective you can make your Auto Scaling architecture.

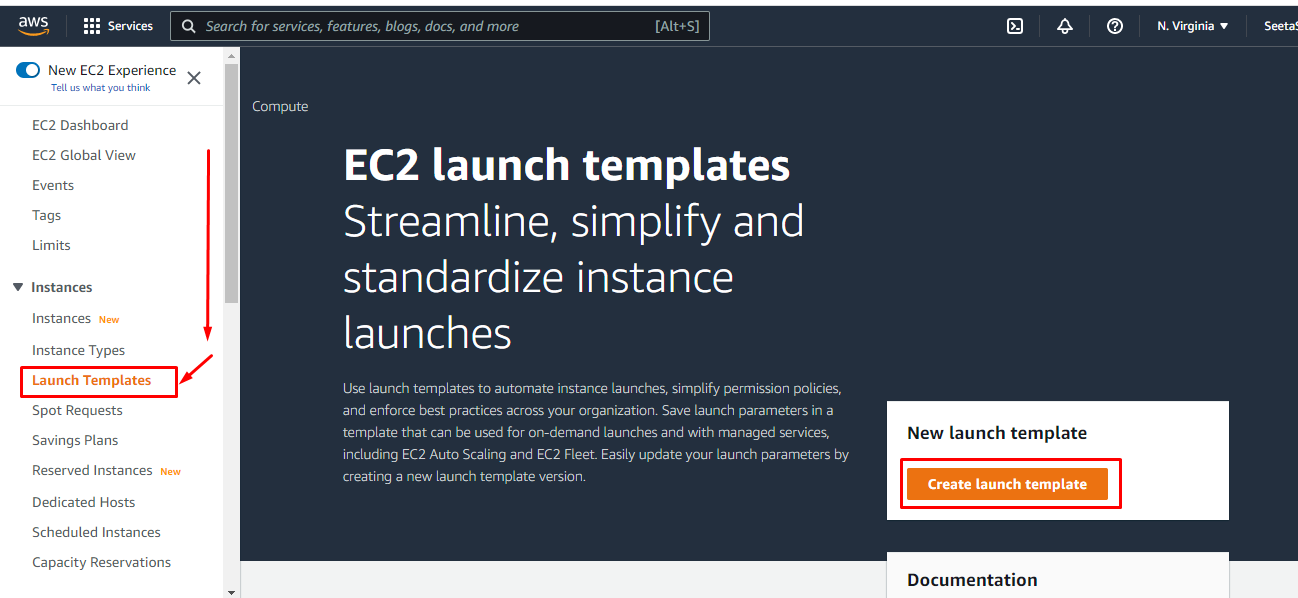
**Tasks**

* Walkthrough summary
* Prepare for the walkthrough
* Step 1: Create a launch template
* Step 2: Create a single-instance Auto Scaling group
* Step 3: Verify your Auto Scaling group
* Step 4: Terminate an instance in your Auto Scaling group
* Step 5: Next steps
* Step 6: Clean up

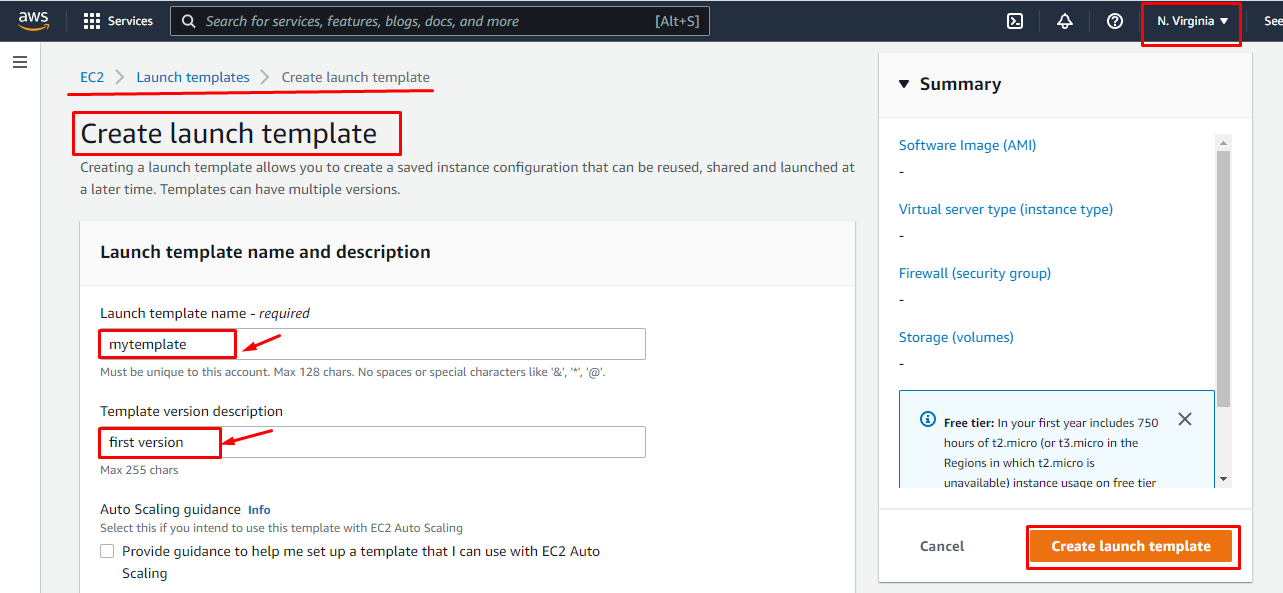
**Step 1: Create a launch template**

**To create a launch template**

1. Open the Launch templates page of the Amazon EC2 console.



1. On the navigation bar at the top of the screen, select an AWS Region. The launch template and Auto Scaling group that you create are tied to the Region that you specify.
2. Choose **Create launch template**.
3. For **Launch template name**, enter **my-template-for-auto-scaling**.



1. Under **Auto Scaling guidance**, select the check box.

Graphical user interface, text, application, email

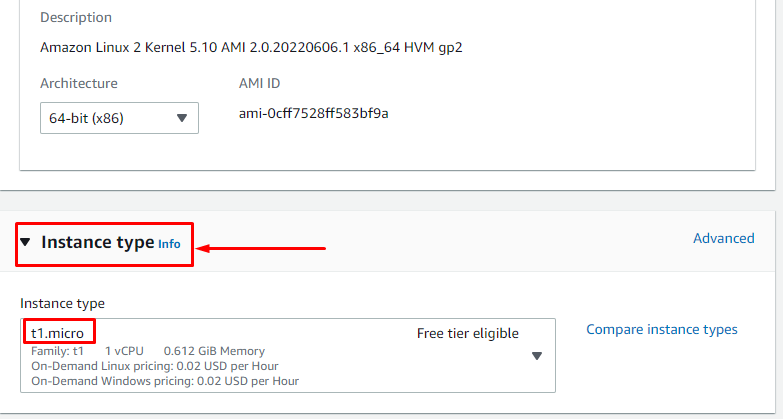
Description automatically generated

1. For **Application and OS Images (Amazon Machine Image)**, choose a version of Amazon Linux 2 (HVM) from the **Quick Start** list. The AMI serves as a basic configuration template for your instances.

Graphical user interface, application, email

Description automatically generated

1. For **Instance type**, choose a hardware configuration that is compatible with the AMI that you specified.



1. (Optional) For **Key pair (login)**, choose an existing key pair. You use key pairs to connect to an Amazon EC2 instance with SSH. Connecting to an instance is not included as part of this tutorial. Therefore, you don't need to specify a key pair unless you intend to connect to your instance using SSH.

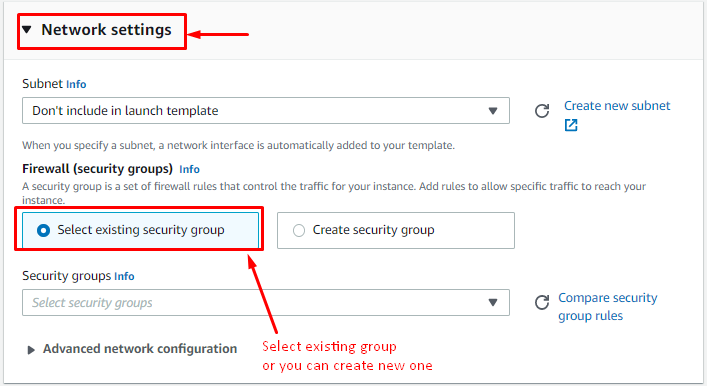
Application

Description automatically generated with medium confidence

Graphical user interface, text, application, email

Description automatically generated

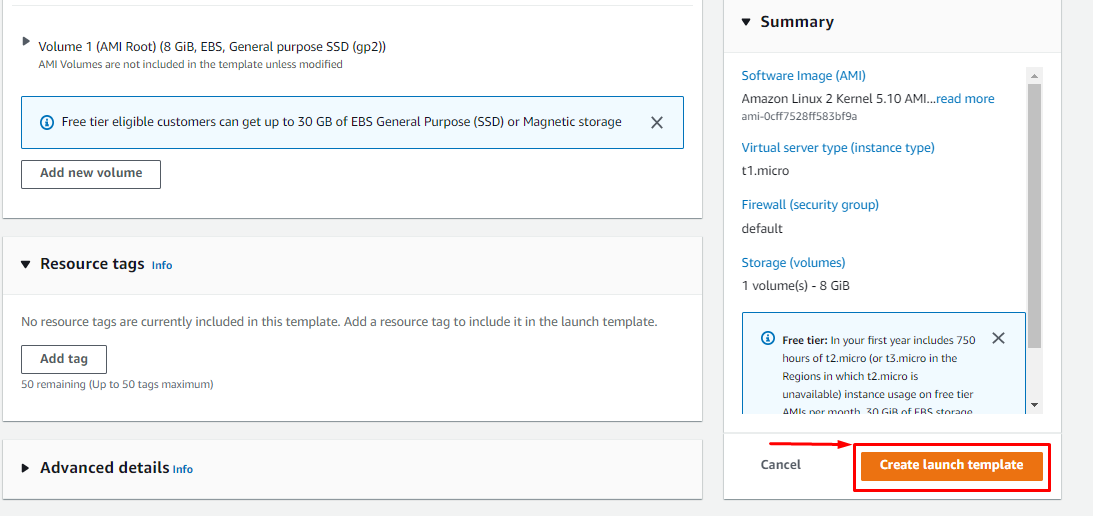
1. For **Network settings**, **Security groups**, choose a security group in the same VPC that you plan to use as the VPC for your Auto Scaling group. If you don't specify a security group, your instance is automatically associated with the default security group for the VPC.



Graphical user interface, text, application

Description automatically generated

1. You can leave **Advanced network configuration** empty. Leaving the setting empty creates a primary network interface with IP addresses that we select for your instance based on the subnet to which the network interface is established. If instead you choose to configure a network interface, the security group must be a part of it.



1. Choose **Create launch template**.

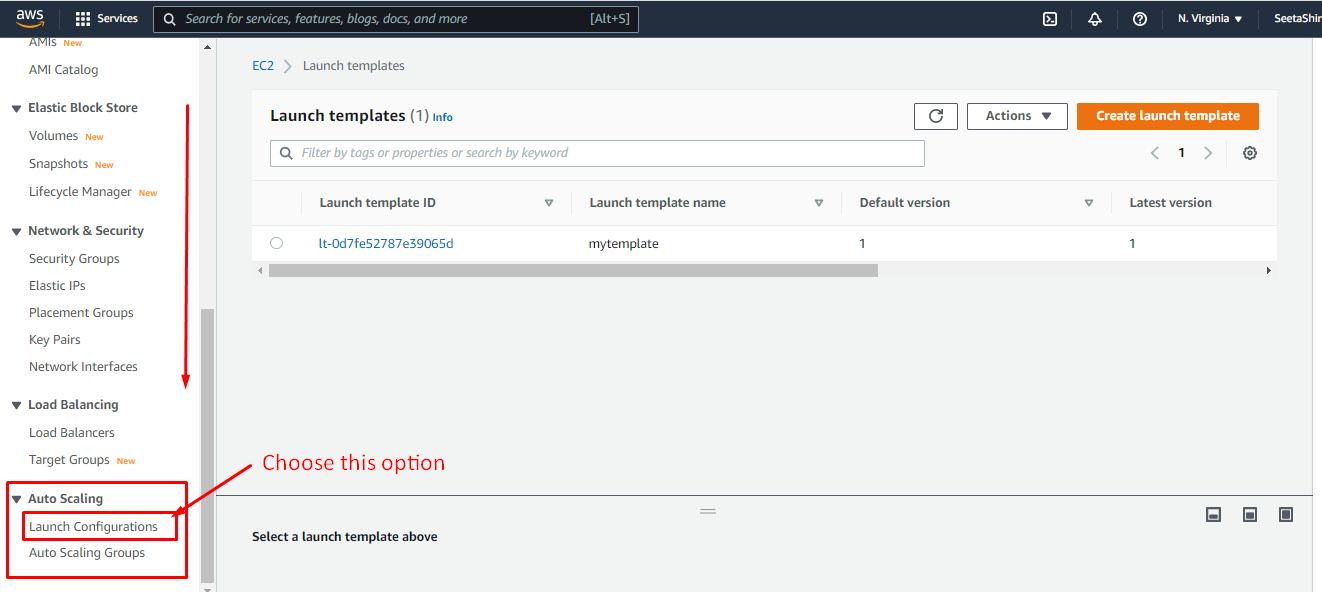
Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

1. On the confirmation page, choose **Create Auto Scaling group**.



**To create a launch configuration**

1. Open the [Launch configurations page](https://console.aws.amazon.com/ec2/v2/home?#LaunchConfigurations) of the Amazon EC2 console.

Graphical user interface, text, application, email

Description automatically generated

1. On the navigation bar, select an AWS Region. The launch configuration and Auto Scaling group that you create are tied to the Region that you specify.

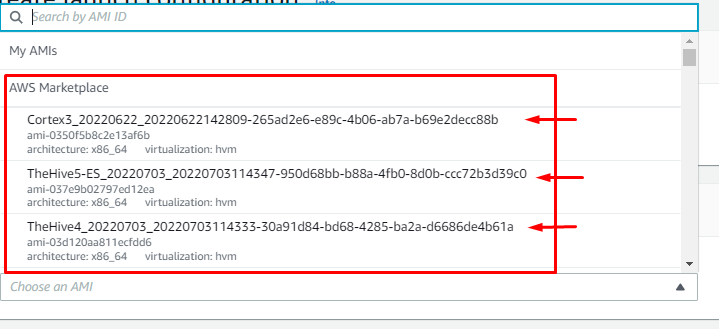
Graphical user interface, text, application

Description automatically generated

1. Choose **Create launch configuration**, and then enter **my-first-launch-configuration** in the **Name** field.

Graphical user interface, text, application, email

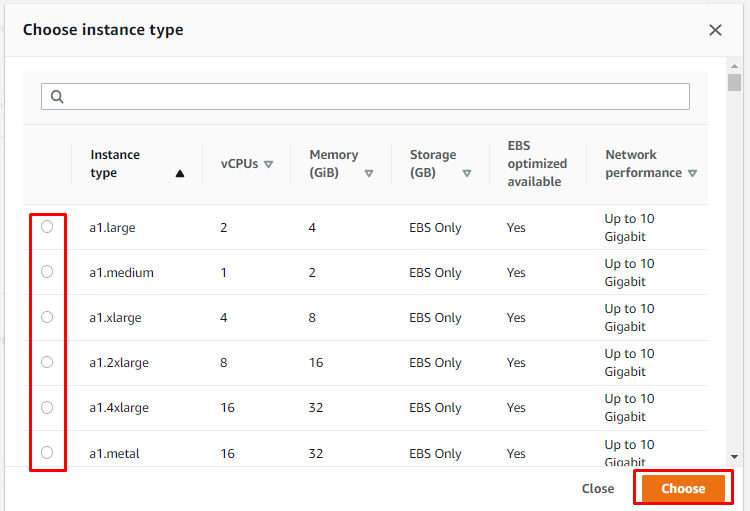
Description automatically generated



1. For **Amazon machine image (AMI)**, choose an AMI. To find a specific AMI, you can[find a suitable AMI](https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/finding-an-ami.html), make note of its ID, and enter the ID as search criteria.

To get the ID of the Amazon Linux 2 AMI:

* 1. Open the [Amazon EC2 console](https://console.aws.amazon.com/ec2/).



* 1. In the navigation pane, under **Instances**, choose **Instances**, and then choose **Launch instances**.
  2. On the **Quick Start** tab of the **Choose an Amazon Machine Image** page, note the ID of the AMI next to **Amazon Linux 2 AMI (HVM)**. Notice that this AMI is marked "Free tier eligible."

1. For **Instance type**, select a hardware configuration for your instance.

Graphical user interface

Description automatically generated with medium confidence

1. Under **Additional configuration**, for **Advanced details**, **IP address type**, make a selection. To provide internet connectivity to instances in a VPC, choose an option that assigns a public IP address. If an instance is launched into a default VPC, the default is to assign a public IP address. If you want to provide internet connectivity to your instance but aren't sure whether you have a default VPC, choose **Assign a public IP address to every instance**.

Graphical user interface, text, application, email

Description automatically generated

1. For **Security groups**, choose an existing security group. If you leave the **Create a new security group** option selected, a default SSH rule is configured for Amazon EC2 instances running Linux. A default RDP rule is configured for Amazon EC2 instances running Windows.

Graphical user interface, text, application, email

Description automatically generated

1. For **Key pair (login)**, choose an option under **Key pair options** as instructed. Connecting to an instance is not included as part of this tutorial. Therefore, you can select **Proceed without a key pair** unless you intend to connect to your instance using SSH.

Graphical user interface, application

Description automatically generated

1. Choose **Create launch configuration**.

Graphical user interface, text, application, email

Description automatically generated

1. Select the check box next to the name of your new launch configuration and choose **Actions**, **Create Auto Scaling group**

Graphical user interface, text, application, email

Description automatically generated