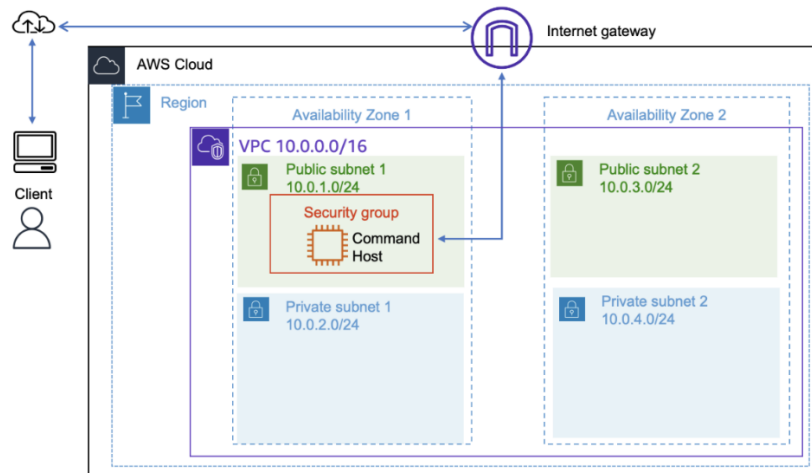


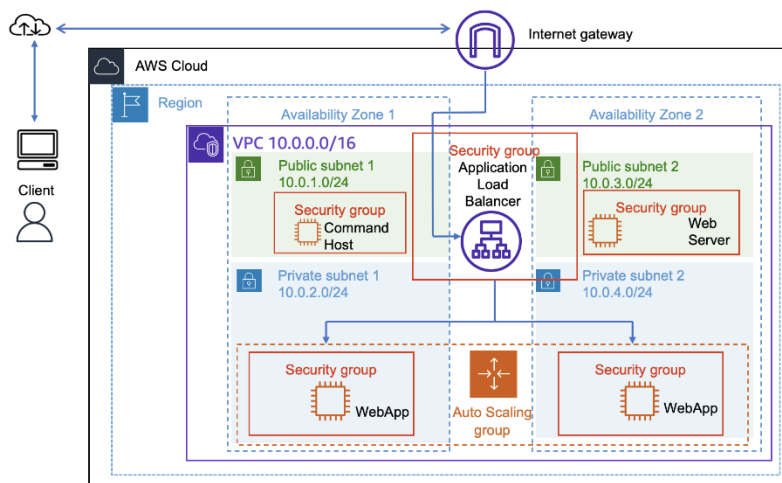
Using Auto Scaling in AWS (Linux)

Lab overview

In this lab, you use the AWS Command Line Interface (AWS CLI) to create an Amazon Elastic Compute Cloud (EC2) instance to host a web server and create an Amazon Machine Image (AMI) from that instance. You then use that AMI as the basis for launching a system that scales automatically under a variable load by using Amazon EC2 Auto Scaling. You also create an Elastic Load Balancer to distribute the load across EC2 instances created in multiple Availability Zones by the auto scaling configuration.



Final architecture:



Objectives

After completing this lab, you will be able to do the following:

- Create an EC2 instance by using an AWS CLI command.

- Create a new AMI by using the AWS CLI.
- Create an Amazon EC2 launch template.
- Create an Amazon EC2 Auto Scaling launch configuration.
- Configure scaling policies and create an Auto Scaling group to scale in and scale out the number of servers based on a variable load.

Task 1: Creating a new AMI for Amazon EC2 Auto Scaling

In this task, you launch a new EC2 instance and then create a new AMI based on that running instance. You use the AWS CLI on the Command Host EC2 instance to perform all of these operations.

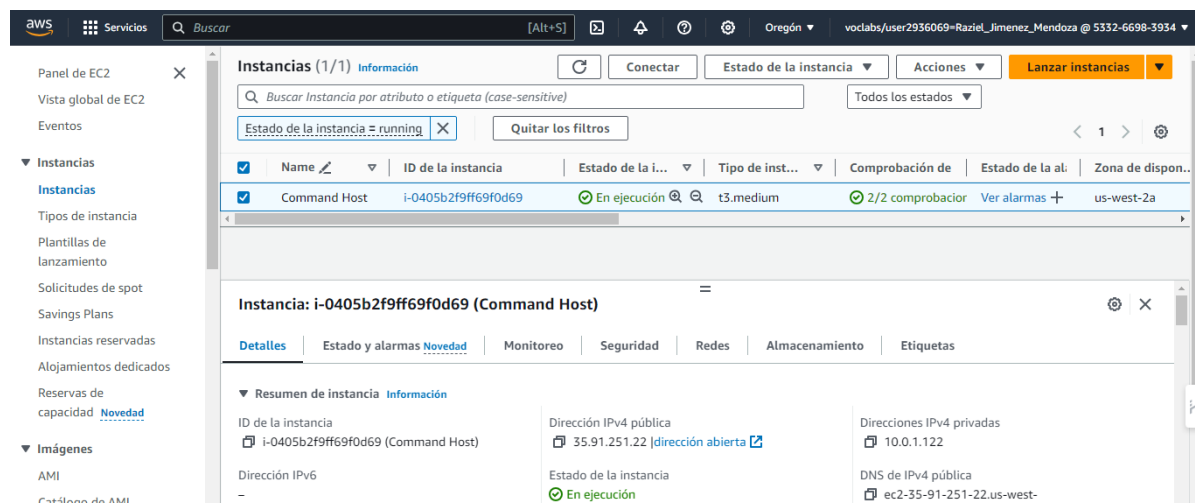
Task 1.1: Connecting to the Command Host instance

In this task, you use EC2 Instance Connect to connect to the Command Host EC2 instance that was created when the lab was provisioned. You use this instance to run AWS CLI commands.

5. On the **AWS Management Console**, in the **Search** bar, enter and choose **EC2** to open the **EC2 Management Console**.
6. In the navigation pane, choose **Instances**.
7. From the list of instances, select the **Command Host** instance.
8. Choose **Connect**.
9. On the **EC2 Instance Connect** tab, choose **Connect**.

Note: If you prefer to use an SSH client to connect to the EC2 instance, see the guidance to [Connect to Your Linux Instance](#).

Now that you are connected to the Command Host instance, you can configure and use the AWS CLI to call AWS services.



The screenshot shows the AWS Management Console interface for connecting to an EC2 instance. The breadcrumb navigation at the top reads: [EC2](#) > [Instancias](#) > [i-0405b2f9ff69f0d69](#) > Conectarse a la instancia. The main heading is 'Conectarse a la instancia' with a sub-link for 'Información'. Below this, a message states: 'Conéctese a la instancia i-0405b2f9ff69f0d69 (Command Host) mediante cualquiera de estas opciones'. There are four tabs: 'Conexión de la instancia EC2' (selected), 'Administrador de sesiones', 'Cliente SSH', and 'Consola de serie de EC2'. The 'Conexión de la instancia EC2' tab contains the following fields: 'ID de la instancia' (i-0405b2f9ff69f0d69 (Command Host)), 'Tipo de conexión' (with two radio button options: 'Conectarse mediante la Conexión de la instancia EC2' which is selected, and 'Conectarse mediante punto de conexión de EC2 Instance Connect'), 'Dirección IP pública' (35.91.251.22), and 'Nombre de usuario' (a search box containing 'ec2-user'). A note at the bottom states: 'Nota: En la mayoría de los casos, el nombre de usuario predeterminado, ec2-user, es correcto. Sin embargo, lea las instrucciones de uso de la AMI para comprobar si el propietario de la AMI ha cambiado el nombre de usuario predeterminado.' At the bottom right are 'Cancelar' and 'Conectar' buttons.

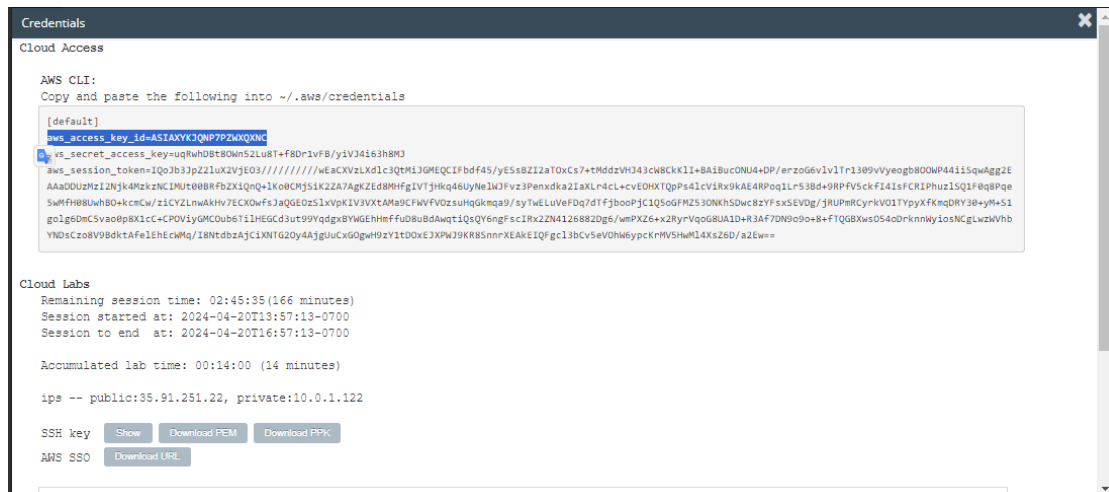
Task 1.2: Configuring the AWS CLI

The AWS CLI is preconfigured on the Command Host instance.

10. To confirm that the Region in which the Command Host instance is running is the same as the lab (the us-west-2 Region), run the following command:

```
curl http://169.254.169.254/latest/dynamic/instance-identity/document | grep region
```

You use this Region information in the next steps.

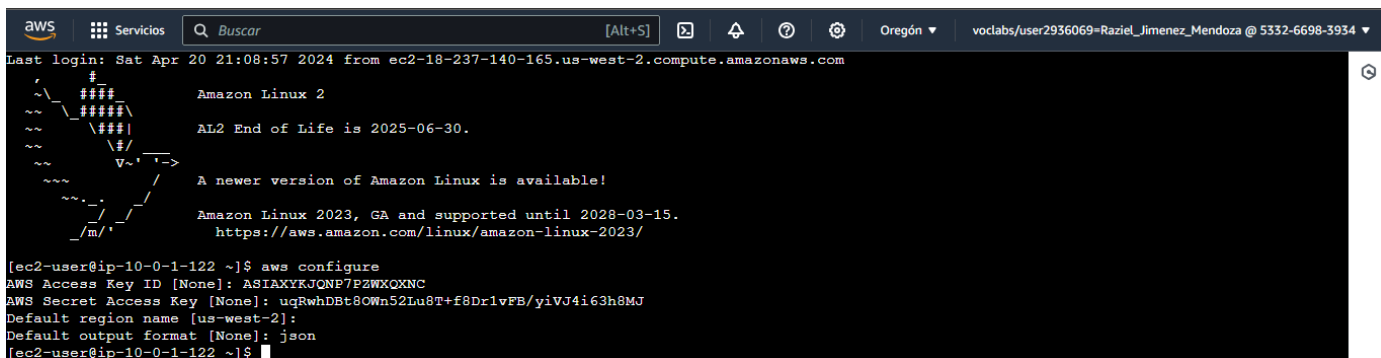


11. To update the AWS CLI software with the correct credentials, run the following command:

```
aws configure
```

12. At the prompts, enter the following information:

- **AWS Access Key ID:** Press Enter.
- **AWS Secret Access Key:** Press Enter.
- **Default region name:** Enter the name of the Region from the previous steps in this task (for example, `us-west-2`). If the Region is already displayed, press Enter.
- **Default output format:** Enter `json`



Now you are ready to access and run the scripts detailed in the following steps.

13. To access these scripts, enter the following command to navigate to their directory:

```
cd /home/ec2-user/
```

Task 1.3: Creating a new EC2 Instance

In this task, you use the AWS CLI to create a new instance that hosts a web server.

14. To inspect the UserData.txt script that was installed for you as part of the Command Host creation, run the following command:

```
more UserData.txt
```

```
[ec2-user@ip-10-0-1-122 ~]$ cd /home/ec2-user/  
[ec2-user@ip-10-0-1-122 ~]$ ls  
UserData.txt  
[ec2-user@ip-10-0-1-122 ~]$
```

This script performs a number of initialization tasks, including updating all installed software on the box and installing a small PHP web application that you can use to simulate a high CPU load on the instance. The following lines appear near the end of the script:

```
find -wholename /root/.history -wholename /home/.history -exec rm -f {} \;  
find / -name 'authorized_keys' -exec rm -f {} \;  
rm -rf /var/lib/cloud/data/scripts/*
```

These lines erase any history or security information that might have accidentally been left on the instance when the image was taken.

15. At the top of this page, choose **Details**, and choose **Show**.
16. Copy the **KEYNAME**, **AMIID**, **HTTPACCESS**, and **SUBNETID** values into a text editor document, and then choose **X** to close the **Credentials** panel.
17. In the following script, replace the corresponding text with the values from the previous step.

```
aws ec2 run-instances --key-name KEYNAME --instance-type t3.micro --image-id AMIID --  
user-data file:///home/ec2-user/UserData.txt --security-group-ids HTTPACCESS --subnet-id  
SUBNETID --associate-public-ip-address --tag-specifications  
'ResourceType=instance,Tags=[{Key=Name,Value=WebServer}]' --output text --query  
'Instances[*].InstanceId'
```

```
aws ec2 run-instances --key-name vockey --instance-type t3.micro --image-id ami-06883a492f195064e  
--user-data file:///home/ec2-user/UserData.txt --security-group-ids sg-0604fcdbe3175fb2 --subnet-  
id subnet-0eb902ec12dcb6881 --associate-public-ip-address --tag-specifications  
'ResourceType=instance,Tags=[{Key=Name,Value=WebServer}]' --output text --query  
'Instances[*].InstanceId'
```

Cloud Access

AWS CLI:

Show

Cloud Labs

Remaining session time: 02:28:42(149 minutes)

Session started at: 2024-04-20T13:57:13-0700

Session to end at: 2024-04-20T16:57:13-0700

Accumulated lab time: 00:31:00 (31 minutes)

ips -- public:35.91.251.22, private:10.0.1.122

SSH key

Show

Download PEM

Download PPK

AWS SSO

Download URL

AMIID	ami-06883a492f195064e
HTTPACCESS	sg-0604fcdbed3175fb2
COMMANDHOSTIP	35.91.251.22
KEYNAME	vockey
SUBNETID	subnet-0eb902ec12dcb6881

Note: An AWS token was copied and added to the **credentials** file.

Copy and paste the following into ~/.aws/credentials

```
[default]
aws_access_key_id=ASIAXYKJQNP7PZWQXNC
aws_secret_access_key=uqrwhDBT80Wn5ZLu8T+f8Dr1vFB/y1VJ4163h8MJ
aws_session_token=IQoJb3JpZ2l1uXZVJEO3////////wEaCXvzLXdlc3QTMlJGMEQCIbdf45/yESsBZ12aToXcs7+tMddzVHJ43cW8CkKl1+BA18ucONU4+DP/erzoG6v1v1Tr1309vVyeogb800WP441i5qWAgg2E
AAADDUzMr12NjK4MzkZNCIMUt00BRfbZXIqNq+1Ko8CMj51K2ZATAgKZEd8MHfgIVTjHk46UyNe1WJfVz3Penxdka2IaXLR4cL+cvEOHXTQpPs41cVIRx9KAE4RPQ1Lr53Bd+9RPfV5ckFI4IsFCRIPhuz15Q1F0q8Pqe
5wMfH88Uwh80+kcmCw/z1CYZLnaAKhv7ECKOwFs3aQGE0zS1xVpK1V3VXtAMa9CFWfV0ZsuHQgkmaq9/syTwELuVeFdQ7d1fjboopJC1Q5oGFmZ53ONKhSDwc8ZyFxsxSEVDg/jRUPmRCyrkVO1TpyXfKmqDRY38+yM+S1
go1g6DmC5vaop8X1cc+CPOViyGMCoub6T1IHGcd3ut99YqdgxBYNGEhHffuD8u8dAwqt1QsQY6ngfsc1Rx2Z4126882Dg6/wmPXZ6+x2RyrVqog8UA1D+R3AF7DN9o9o+8+FTQGBXws054oDrknkMylosNCg1wZkVhb
YNDsCzo8V9BdktAFelEhEcMwq/18NtdbZAJC1XNTG20y4AJgUuCXvG0gW9zY1tD0xEJXPNJ9KR8SnnrXEAKEIQFgc13bCy5eVohW6ypcKrmV5HwM14XsZ6D/a2Ew==
```

```
[ec2-user@ip-10-0-1-122 ~]$ ls -l
total 4
-rw-r--r-- 1 root root 548 Apr  8 07:20 UserData.txt
[ec2-user@ip-10-0-1-122 ~]$ ls -a
. . . .aws .bash_logout .bash_profile .bashrc .ssh UserData.txt
[ec2-user@ip-10-0-1-122 ~]$ cd .aws
[ec2-user@ip-10-0-1-122 .aws]$ ls
config credentials
[ec2-user@ip-10-0-1-122 .aws]$ vi credentials
```

```
[default]
aws_access_key_id = ASIAXYKJQNP7PZWQXNC
aws_secret_access_key = uqrwhDBT80Wn5ZLu8T+f8Dr1vFB/y1VJ4163h8MJ
aws_session_token=IQoJb3JpZ2l1uXZVJEO3////////wEaCXvzLXdlc3QTMlJGMEQCIbdf45/yESsBZ12aToXcs7+tMddzVHJ43cW8CkKl1+BA18ucONU4+DP/erzoG6v1v1Tr1309vVyeogb800WP441i5qWAgg2EAAADDUzMr12NjK4MzkZNCIMUt00BRfbZXIqNq+1Ko8CMj51K2ZATAgKZEd8MHfgIVTjHk46UyNe1WJfVz3Penxdka2IaXLR4cL+cvEOHXTQpPs41cVIRx9KAE4RPQ1Lr53Bd+9RPfV5ckFI4IsFCRIPhuz15Q1F0q8Pqe5wMfH88Uwh80+kcmCw/z1CYZLnaAKhv7ECKOwFs3aQGE0zS1xVpK1V3VXtAMa9CFWfV0ZsuHQgkmaq9/syTwELuVeFdQ7d1fjboopJC1Q5oGFmZ53ONKhSDwc8ZyFxsxSEVDg/jRUPmRCyrkVO1TpyXfKmqDRY38+yM+S1go1g6DmC5vaop8X1cc+CPOViyGMCoub6T1IHGcd3ut99YqdgxBYNGEhHffuD8u8dAwqt1QsQY6ngfsc1Rx2Z4126882Dg6/wmPXZ6+x2RyrVqog8UA1D+R3AF7DN9o9o+8+FTQGBXws054oDrknkMylosNCg1wZkVhbYNDsCzo8V9BdktAFelEhEcMwq/18NtdbZAJC1XNTG20y4AJgUuCXvG0gW9zY1tD0xEJXPNJ9KR8SnnrXEAKEIQFgc13bCy5eVohW6ypcKrmV5HwM14XsZ6D/a2Ew==
```

18. Enter your modified script into the terminal window, and run the script.

The output of this command provides you with an **InstanceId**.
Subsequent steps in this lab refer to this value as **NEW-INSTANCE-ID**.
Replace this value as needed throughout this lab.

19. Copy and paste the **InstanceId** value into a text editor to use later.
20. To use the `aws ec2 wait instance-running` command to monitor this instance's status, replace *NEW-INSTANCE-ID* in the following command with the **InstanceId** value that you copied in the previous step. Run your modified command.

```
aws ec2 wait instance-running --instance-ids NEW-INSTANCE-ID
```

Wait for the command to return to a prompt before proceeding.

Your instance starts a new web server. To test that the web server was installed properly, you must obtain the public DNS name.

21. To obtain the public DNS name, in the following command, replace *NEW-INSTANCE-ID* with the value that you copied previously, and run your modified command:

```
aws ec2 describe-instances --instance-id NEW-INSTANCE-ID --query  
'Reservations[0].Instances[0].NetworkInterfaces[0].Association.PublicDnsName'
```

22. Copy the output of this command without the quotation marks.

The value of this output is referred to as **PUBLIC-DNS-ADDRESS** in the next steps.

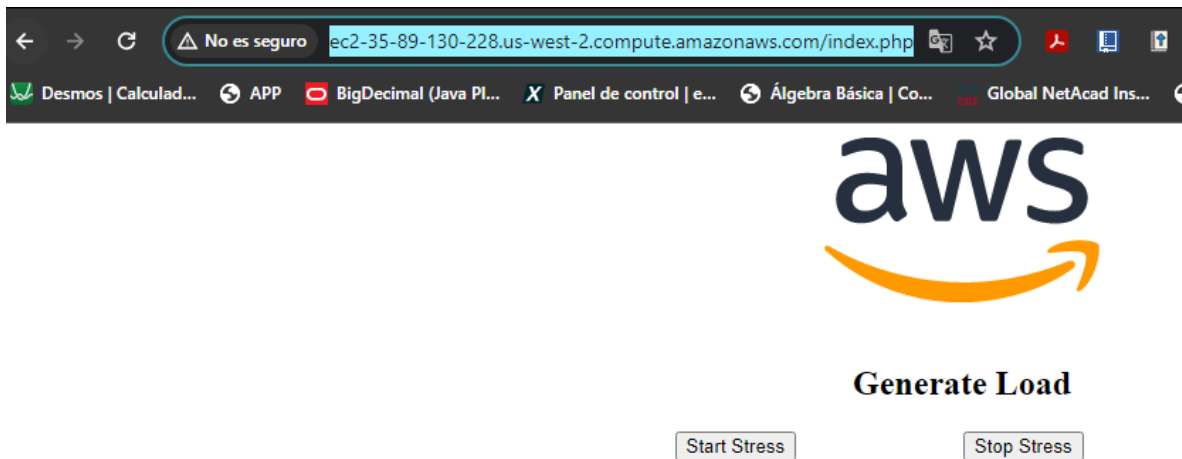
23. In a new browser tab, enter the output that you copied from the previous step.

It could take a few minutes for the web server to be installed. Wait 5 minutes before continuing to the next steps.

Do not choose **Start Stress** at this stage.

24. In the following command, replace *PUBLIC-DNS-ADDRESS* with the value that you copied in the previous steps, and then run your modified command.

```
http://PUBLIC-DNS-ADDRESS/index.php
```



Task 1.4: Creating a Custom AMI

In this task, you create a new AMI based on that instance that you just created.

25. To create a new AMI based on this instance, in the following aws ec2 create-image command, replace *NEW-INSTANCE-ID* with the value that you copied previously, and run your adjusted command:

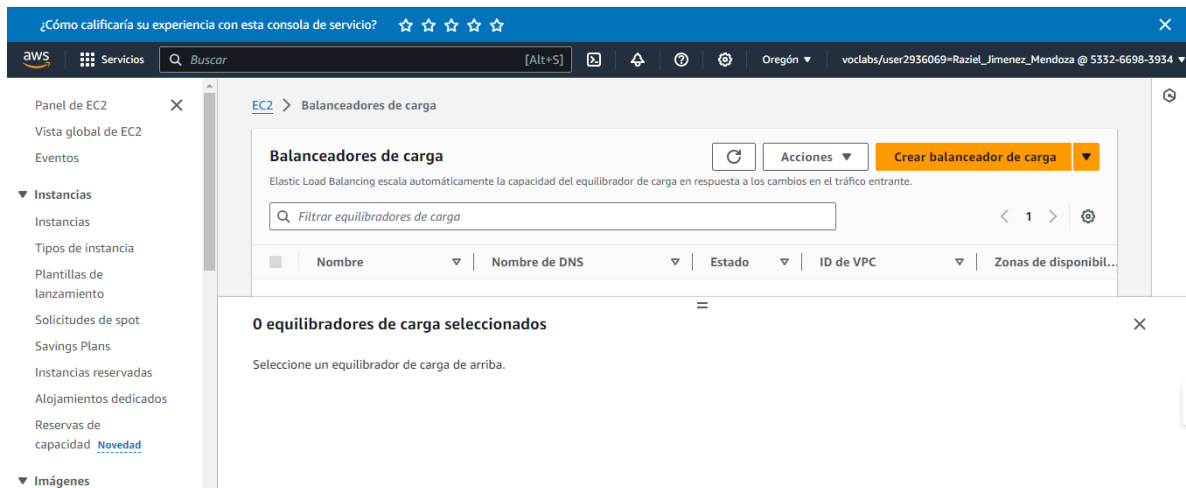
```
aws ec2 create-image --name WebServerAMI --instance-id NEW-INSTANCE-ID
```

```
[ec2-user@ip-10-0-1-122 ~]$ aws ec2 create-image --name WebServerAMI --instance-id i-077265d5840673db4
{
  "ImageId": "ami-05ba4e2a0836057fa"
}
[ec2-user@ip-10-0-1-122 ~]$
```

Task 2.1: Creating an Application Load Balancer

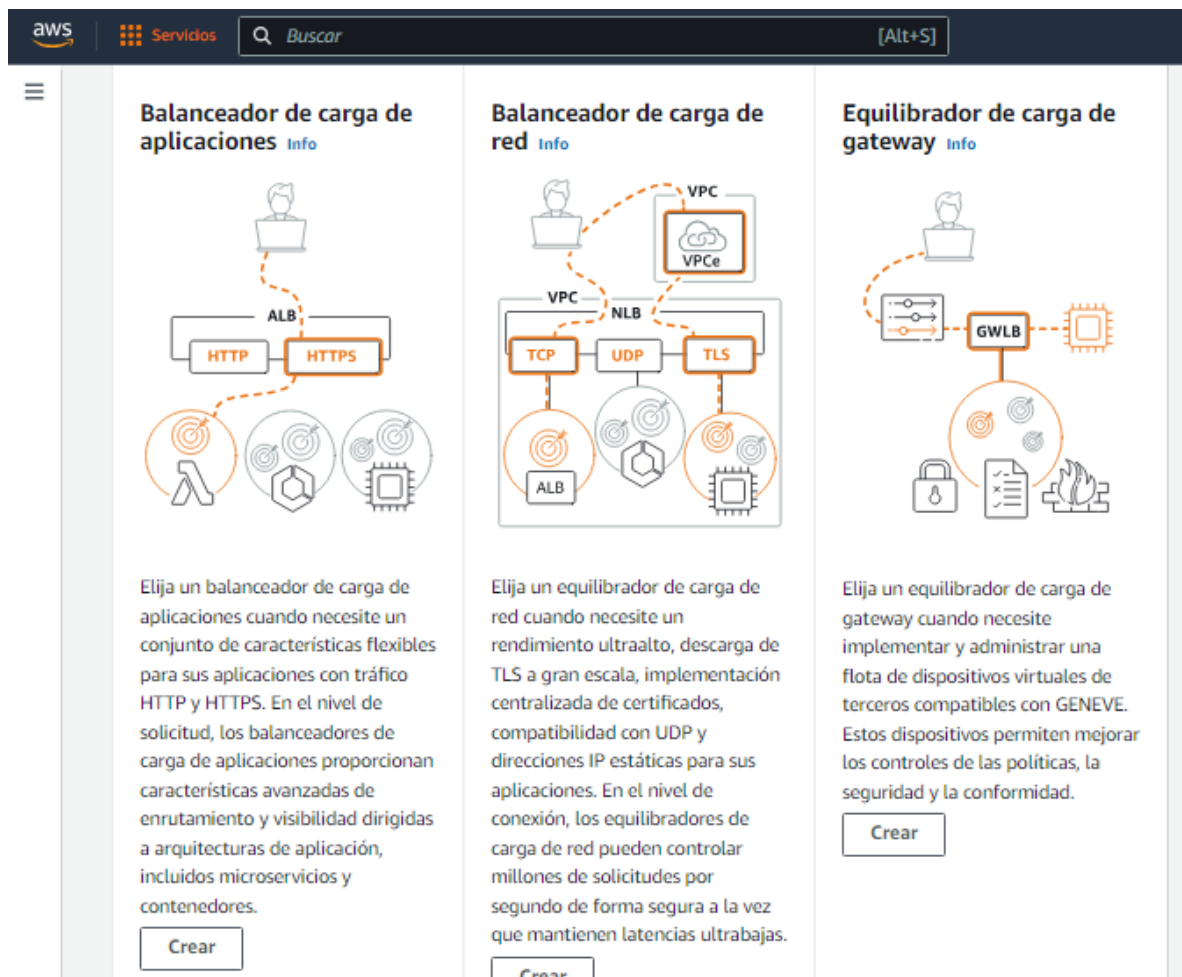
In this task, you create a load balancer that can balance traffic across multiple EC2 instances and Availability Zones.

26. On the EC2 Management Console, in the left navigation pane, locate the **Load Balancing** section, and choose **Load Balancers**.



27. Choose **Create load balancer**.

28. In the **Load balancer types** section, for **Application Load Balancer**, choose **Create**.



29. On the **Create Application Load Balancer** page, in the **Basic configuration** section, configure the following option:

- For **Load balancer name**, enter **WebServerELB**

The screenshot shows the 'Configuración básica' (Basic configuration) section of the AWS Management Console for creating an Application Load Balancer. The 'Nombre del balanceador de carga' (Load balancer name) field is set to 'WebServerELB'. The 'Esquema' (Scheme) is set to 'Expuesto a Internet' (Exposed to Internet). The 'Tipo de dirección IP' (IP address type) is set to 'IPv4'.

Configuración básica

Nombre del balanceador de carga
Debe ser nombre único dentro de su cuenta de AWS y no puede cambiarse después de crear el equilibrador de carga.

Se permite un máximo de 32 caracteres alfanuméricos, incluidos guiones, pero el nombre no puede comenzar ni terminar por un guion.

Esquema [Info](#)
El esquema no se puede cambiar después de crear el equilibrador de carga.
☒ **Expuesto a Internet**
Un balanceador de carga expuesto a Internet dirige las solicitudes de los clientes a través de Internet a los destinos. Requiere una subred pública. [Más información](#)
☐ **Interno**
Un balanceador de carga interno dirige las solicitudes de los clientes a los destinos mediante direcciones IP privadas.

Tipo de dirección IP [Info](#)
Seleccione el tipo de direcciones IP que utilizan las subredes.
☒ **IPv4**
Recomendado para balanceadores de carga internos.
☐ **Dualstack**
Incluye direcciones IPv4 e IPv6.

30. In the **Network mapping** section, configure the following options:

- For **VPC**, choose **Lab VPC**.

The screenshot shows the 'Mapeo de red' (Network mapping) section of the AWS Management Console for creating an Application Load Balancer. The 'VPC' is set to 'Lab VPC'. The 'Mapeos' (Mappings) section shows two availability zones: 'us-west-2a (usw2-az2)' and 'us-west-2b (usw2-az1)', both of which are currently unchecked.

Mapeo de red [Info](#)
El balanceador de carga dirige el tráfico a los destinos de las subredes seleccionadas y en función de la configuración de las direcciones IP.

VPC [Info](#)
Seleccione la nube privada virtual (VPC) para los destinos o [cree una nueva VPC](#). Solo las VPC con una puerta de enlace de Internet están habilitadas para la selección. La VPC seleccionada no se podrá cambiar después de crear el equilibrador de carga. Para confirmar la VPC para los destinos, consulte los [grupos de destinos](#).

vpc-06915e49a72034f56
IPv4: 10.0.0.0/16

Mapeos [Info](#)
Seleccione al menos dos zonas de disponibilidad y una subred por zona. El equilibrador de carga solo dirige el tráfico a los destinos de estas zonas de disponibilidad. Las zonas de disponibilidad que no son compatibles con el equilibrador de carga o la VPC no están disponibles para seleccionarse.
☐ **us-west-2a (usw2-az2)**
☐ **us-west-2b (usw2-az1)**

- For **Mappings**, choose both Availability Zones listed.
 - For the first Availability Zone, choose **Public Subnet 1**.
 - For the second Availability Zone, choose **Public Subnet 2**.

These options configure the load balancer to operate across multiple Availability Zones.

Mapeos [Info](#)

Seleccione al menos dos zonas de disponibilidad y una subred por zona. El equilibrador de carga solo dirige el tráfico a los destinos de estas zonas de disponibilidad. Las zonas de disponibilidad que no son compatibles con el equilibrador de carga o la VPC no están disponibles para seleccionárselas.

☒ **us-west-2a (usw2-az2)**

Subred

subnet-00ec30b5f79820208

Public Subnet 1 ▼

Dirección IPv4


Asignado por AWS

☒ **us-west-2b (usw2-az1)**

Subred

subnet-055159c7873d1c359

Private Subnet 2 ▼

 La subred seleccionada no tiene ninguna ruta a una puerta de enlace de Internet. Esto significa que el equilibrador de carga no recibirá tráfico de Internet. Puede continuar con esta selección; sin embargo, para que el tráfico de Internet llegue al enlace de carga, debe actualizar la tabla de enrutamiento de la subred en el [Consola de la VPC](#).

Dirección IPv4

Asignado por AWS

31. In the **Security groups** section, choose the **X** for the **default** security group to remove it.


32. From the **Security groups** dropdown list, choose **HTTPAccess**.

The **HTTPAccess** security group has already been created for you, which permits HTTP access.

Grupos de seguridad [Info](#)

Un grupo de seguridad consiste en un conjunto de reglas de firewall que controlan el tráfico hacia el equilibrador de carga. Seleccione un grupo de seguridad existente o [cree un nuevo grupo de seguridad](#).

Grupos de seguridad

Seleccione hasta 5 grupos de seguridad ▼ 

HTTPAccess

sg-0604fcdbe3175fb2 VPC: vpc-06915e49a72034f56

✕

33. In the **Listeners and routing** section, choose the **Create target group** link.

Note: This link opens a new browser tab with the **Create target group** configuration options.

▼ Agente de escucha HTTP:80 Eliminar

Protocolo Puerto Acción predeterminada [Info](#)

HTTP : 80 Reenviar a Seleccione un grupo de destino ↻

1-65535 [Crear un grupo de destino](#)

Etiquetas del agente de escucha - opcional

Considere la posibilidad de agregar etiquetas al agente de escucha. Las etiquetas permiten clasificar los recursos de AWS para que pueda administrarlos con mayor facilidad.

Agregar etiqueta de agente de escucha

Puede agregar hasta 50 etiquetas más.

[Agregar agente de escucha](#)

34. On the **Specify group details** page, in the **Basic configuration** section, configure the following options:

- For **Choose a target type**, choose **Instances**.
- For **Target group name**, enter `webserver-app`

Nombre del grupo de destino

webserver-app

Se permite un máximo de 32 caracteres alfanuméricos, incluidos guiones, pero el nombre no puede comenzar ni terminar por un guion.

Protocolo : Puerto

Elija un protocolo para su grupo de destinos que corresponda al tipo de equilibrador de carga que enrutará el tráfico hacia él. Algunos protocolos ahora incluyen la detección de anomalías para los destinos y se pueden establecer opciones de mitigación una vez creado el grupo de destinos. Esta elección no se puede cambiar después de la creación

HTTP

80

1-65535

Tipo de dirección IP

Solo los destinos con el tipo de dirección IP indicado pueden registrarse en este grupo de destino.

☒ IPv4

Cada instancia tiene una interfaz de red predeterminada (eth0) a la que se le asigna la dirección IPv4 privada principal. La dirección IPv4 privada principal de la instancia es la que se aplicará al destino.

☐ IPv6

Cada instancia que registre debe tener asignada una dirección IPv6 principal. Ésta se configura en la interfaz de red predeterminada de la instancia (eth0). [Obtenga más información](#)

35. In the **Health checks** section, for **Health check path**, enter `/index.php`

Comprobaciones de estado

El balanceador de carga asociado envía periódicamente solicitudes, según las configuraciones que aparecen a continuación, a los destinos registrados para comprobar su estado.

Protocolo de comprobación de estado

HTTP ▼

Ruta de comprobación de estado

Utilice la ruta predeterminada "/" para realizar comprobaciones de estado en la raíz, o especifique una ruta personalizada si lo prefiere.

/index.php

Hasta 1024 caracteres permitidos.

► Configuración avanzada de comprobación de estado

36. At the bottom of the page, choose **Next**.

37. On the **Register targets** page, choose **Create target group**.

Once the target group has been created successfully, close the **Target groups** browser tab.

Registrar destinos

Se trata de un paso opcional para crear un grupo de destino. Sin embargo, para asegurarse de que el balanceador de carga direcciona el tráfico a este grupo de destino, debe registrar los destinos.

Instancias disponibles (2)

Q Filtrar instancias

<input type="checkbox"/>	ID de instancia ▼	Nombre ▼	Estado ▼	Grupos de seguridad ▼	Zona ▼
<input type="checkbox"/>	i-077265d5840673db4	WebServer	✓ Ejecutando	HTTPAccess	us-west-2b
<input type="checkbox"/>	i-0405b2f9ff69f0d69	Command Host	✓ Ejecutando	CommandHostSecurityGroup	us-west-2a

Revisar destinos

Destinos (0)

Eliminar todos los pendientes

Q Filtrar destinos

☐ Mostrar solo pendientes

ID de instancia ▼	Nombre ▼	Puerto ▼	Estado ▼	Grupos de seguridad ▼	Zona ▼	Dirección IPv4 privada	ID de subred ▼	Hora de lanzamiento
Aún no se han agregado instancias.								
Especifique las instancias anteriores o deje el grupo vacío si prefiere agregar destinos más adelante.								

0 pendientes

Cancelar Anterior **Crear un grupo de destino**

38. Return to the **Load balancers** browser tab, and locate the **Listeners and routing** section. For **Default action**, choose **Refresh** to the right of the **Forward to** dropdown list.
39. From the **Forward to** dropdown list, choose **webserver-app**.

Agentes de escucha y direccionamiento [Info](#)

Un agente de escucha es un proceso que comprueba las solicitudes de conexión mediante el puerto y el protocolo que configure. Las reglas que defina para un agente de escucha determinan cómo el equilibrador de carga dirige las solicitudes a sus destinos registrados.

▼ Agente de escucha **HTTP:80** Eliminar

Protocolo: **HTTP** : Puerto: **80**
1-65535

Acción predeterminada [Info](#)

Reenviar a: **webserver-app** Tipo de destino: Instancia, IPv4 **HTTP** ↺

[Crear un grupo de destino](#)

Etiquetas del agente de escucha - opcional

Considere la posibilidad de agregar etiquetas al agente de escucha. Las etiquetas permiten clasificar los recursos de AWS para que pueda administrarlos con mayor facilidad.

Agregar etiqueta de agente de escucha

Puede agregar hasta 50 etiquetas más.

Agregar agente de escucha

40. At the bottom of the page, choose **Create load balancer**.

You should receive a message similar to the following:

Successfully created load balancer: WebServerELB

Integraciones de servicios [Editar](#)

AWS WAF: Ninguno
AWS Global Accelerator: Ninguno

Etiquetas [Editar](#)

Ninguno

Atributos

ℹ Algunos atributos predeterminados se aplicarán al balanceador de carga. Puede verlos y editarlos después de crear el balanceador de carga.

Estado y flujo de trabajo de creación

► **Estado y tareas del lado del servidor**

Tras completar y enviar los pasos anteriores, todas las tareas del servidor y sus estados estarán disponibles para su supervisión.

Cancelar Crear balanceador de carga

41. To view the **WebServerELB** load balancer that you created, choose **View load balancer**.

42. To copy the **DNS name** of the load balancer, use the copy option , and paste the DNS name into a text editor.

Task 2.2: Creating a launch template

In this task, you create a launch template for your Auto Scaling group. A launch template is a template that an Auto Scaling group uses to launch EC2 instances. When you create a launch template, you specify information for the instances, such as the AMI, instance type, key pair, security group, and disks.

43. On the EC2 Management Console, in the left navigation pane, locate the **Instances** section, and choose **Launch Templates**.



44. Choose **Create launch template**.

45. On the **Create launch template** page, in the **Launch template name and description** section, configure the following options:

- For **Launch template name** - *required*, enter `web-app-launch-template`
- For **Template version description**, enter `A web server for the load test app`
- For **Auto Scaling guidance**, select **Provide guidance to help me set up a template that I can use with EC2 Auto Scaling**.

Nombre y descripción de la plantilla de lanzamiento

Nombre de la plantilla de lanzamiento - *obligatorio*

web-app-launch-template

Debe ser única para esta cuenta. Máximo de 128 caracteres. Sin espacios ni caracteres especiales, como "&", "*", "@".

Descripción de la versión de la plantilla

A web server for the load test app

Máximo de 255 caracteres

Orientación sobre Auto Scaling | [Información](#)

Selecciónelo si va a utilizar esta plantilla con EC2 Auto Scaling

- ☒ Proporcionar orientación que me ayude a configurar una plantilla que pueda utilizar con EC2 Auto Scaling

46. In the Application and OS Images (Amazon Machine Image) - required section, choose the My AMIs tab.

Notice that **WebServerAMI** is already chosen.

▼ Imágenes de aplicaciones y sistemas operativos (Imagen de máquina de Amazon) - obligatorio [Información](#)

Una AMI es una plantilla que contiene la configuración de software (sistema operativo, servidor de aplicaciones y aplicaciones) necesaria para lanzar la instancia. Busque o examine las AMI si no ve lo que busca a continuación.

🔍 Busque en nuestro catálogo completo que incluye miles de imágenes de sistemas operativos y aplicaciones

Recientes

Mis AMI

Inicio rápido

☒ De mi propiedad

☐ Compartido conmigo



Buscar más AMI

Inclusión de AMI de
AWS, Marketplace y la
comunidad

Imágenes de máquina de Amazon (AMI)

WebServerAMI
ami-05ba4e2a0836057fa
2024-04-20T22:01:25.000Z Virtualización: hvm Activado para ENA: true Tipo de dispositivo raíz: ebs

Descripción

-

Arquitectura

x86_64

ID de AMI

ami-05ba4e2a0836057fa

47. In the **Instance type** section, choose the **Instance type** dropdown list, and choose **t3.micro**.
48. In the **Key pair (login)** section, confirm that the **Key pair name** dropdown list is set to **Don't include in launch template**.

Amazon EC2 uses public key cryptography to encrypt and decrypt login information. To log in to your instance, you must create a key pair, specify the name of the key pair when you launch the instance, and provide the private key when you connect to the instance.

Note: In this lab, you do not need to connect to the instance.

The screenshot shows the Amazon EC2 console interface. The top section is titled '▼ Tipo de instancia' with links for 'Información' and 'Obtener asesoramiento', and a 'Avanzado' button. Below this, the 'Tipo de instancia' dropdown is set to 't3.micro'. A tooltip is visible showing details for the t3 family: 'Familia: t3', '2 vCPU', '1 GiB Memoria', 'Generación actual: true', and pricing for SUSE, Windows, RHEL, and Linux. To the right, the 'Todas las generaciones' radio button is selected, and there is a 'Comparar tipos de instancias' link. Below the instance type section, the '▼ Par de claves (inicio de sesión)' section is shown with an 'Información' link. It contains a paragraph explaining the use of key pairs. The 'Nombre del par de claves' dropdown is set to 'No incluir en la plantilla de lanzamiento'. To the right of this dropdown is a 'Crear un nuevo par de claves' button with a refresh icon.

49. In the **Network settings** section, choose the **Security groups** dropdown list, and choose **HTTPAccess**.

When you launch an instance, you can pass user data to the instance. The data can be used to run configuration tasks and scripts.

▼ Configuraciones de red

Información

Subred

Información

No incluir en la plantilla de lanzamiento

↺

↻ Crear nueva subred

Al especificar una subred, se agrega automáticamente una interfaz de red a la plantilla.

Firewall (grupos de seguridad)

Información

Un grupo de seguridad es un conjunto de reglas de firewall que controlan el tráfico de la instancia. Agregue reglas para permitir que un tráfico específico llegue a la instancia.

● Seleccionar un grupo de seguridad existente

○ Crear grupo de seguridad

Grupos de seguridad

Información

Seleccionar grupos de seguridad

HTTPAccess sg-0604fcdbe3175fb2

×

VPC: vpc-06915e49a72034f56

↺

Compare reglas de grupo de seguridad

▼ Configuración de red avanzada

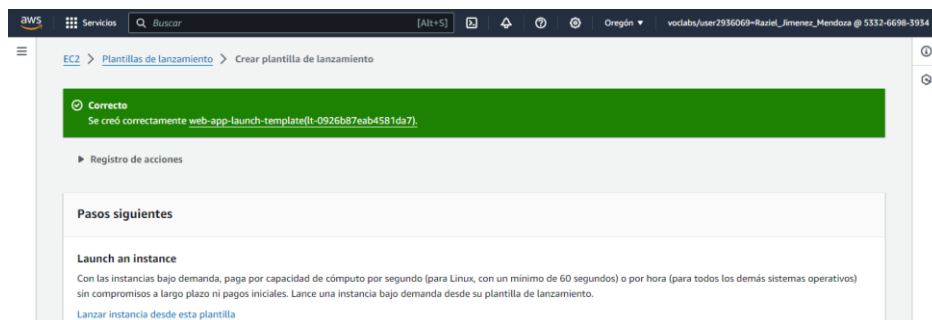
Actualmente, no hay ninguna interfaz de red incluida en esta plantilla. Agregue una interfaz de red para incluirla en la plantilla de lanzamiento.

Agregue interfaz de red

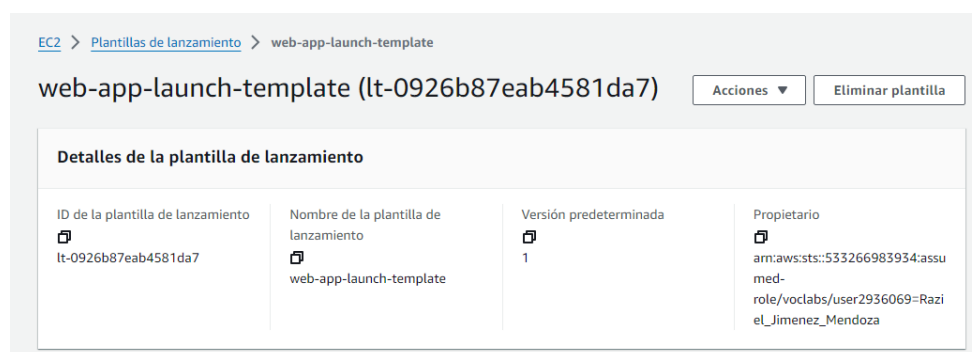
50. Choose **Create launch template**.

You should receive a message similar to the following:

Successfully created web-app-launch-template.



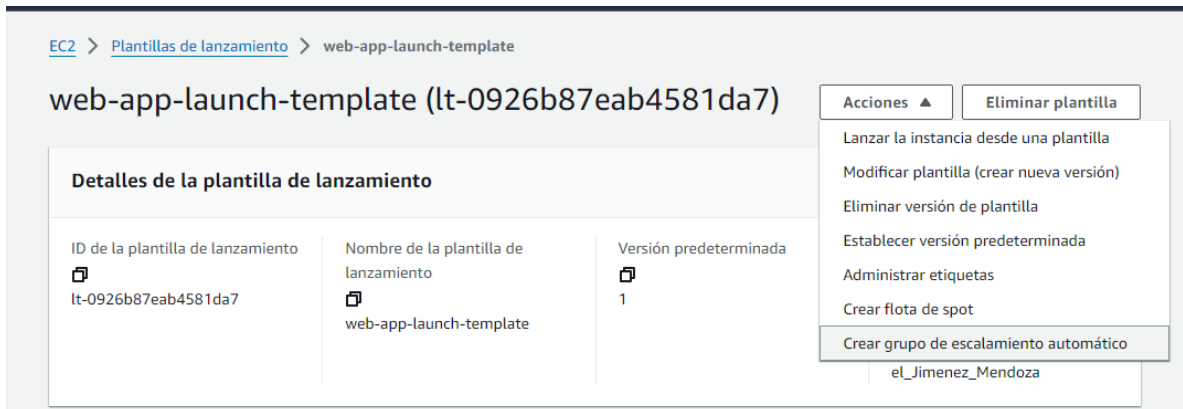
51. Choose **View launch templates**.



Task 2.3: Creating an Auto Scaling group

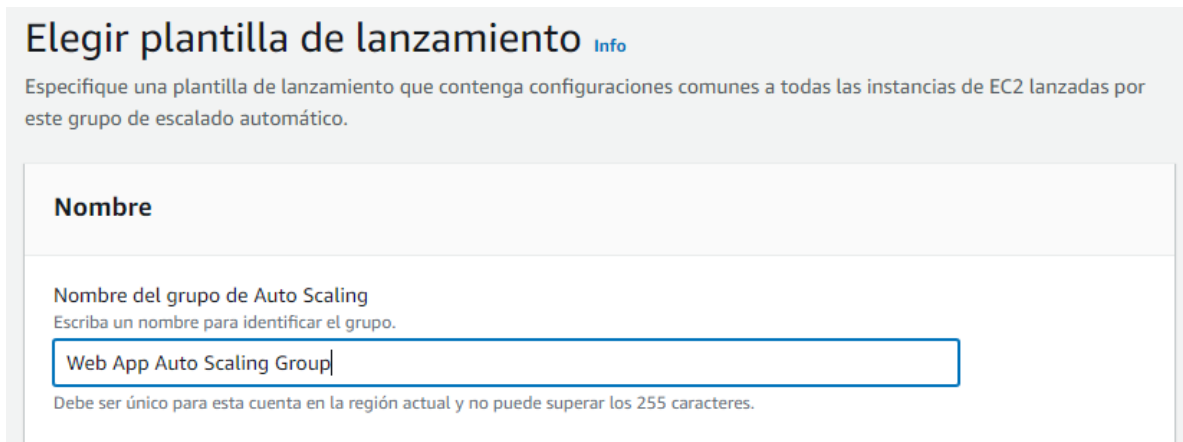
In this task, you use your launch template to create an Auto Scaling group.

52. Choose **web-app-launch-template**, and then from the **Actions** dropdown list, choose **Create Auto Scaling group**.



The screenshot shows the AWS Management Console page for the 'web-app-launch-template'. The breadcrumb navigation is 'EC2 > Plantillas de lanzamiento > web-app-launch-template'. The title is 'web-app-launch-template (lt-0926b87eab4581da7)'. Below the title, there's a section 'Detalles de la plantilla de lanzamiento' with three columns: 'ID de la plantilla de lanzamiento' (lt-0926b87eab4581da7), 'Nombre de la plantilla de lanzamiento' (web-app-launch-template), and 'Versión predeterminada' (1). To the right, there's an 'Acciones' dropdown menu with options: 'Lanzar la instancia desde una plantilla', 'Modificar plantilla (crear nueva versión)', 'Eliminar versión de plantilla', 'Establecer versión predeterminada', 'Administrar etiquetas', 'Crear flota de spot', and 'Crear grupo de escalamiento automático'. The 'Crear grupo de escalamiento automático' option is highlighted, and the user 'el_Jimenez_Mendoza' is visible at the bottom.

53. On the **Choose launch template or configuration** page, in the **Name** section, for **Auto Scaling group name**, enter **Web App Auto Scaling Group**



The screenshot shows the 'Elegir plantilla de lanzamiento' page. The title is 'Elegir plantilla de lanzamiento' with an 'Info' link. Below the title, there's a description: 'Especifique una plantilla de lanzamiento que contenga configuraciones comunes a todas las instancias de EC2 lanzadas por este grupo de escalado automático.' Below this, there's a section 'Nombre' with a sub-section 'Nombre del grupo de Auto Scaling'. The text says 'Escriba un nombre para identificar el grupo.' and there's a text input field containing 'Web App Auto Scaling Group'. Below the input field, there's a note: 'Debe ser único para esta cuenta en la región actual y no puede superar los 255 caracteres.'

54. Choose **Next**.
55. On the **Choose instance launch options** page, in the **Network** section, configure the following options:
- From the **VPC** dropdown list, choose **Lab VPC**.
 - From the **Availability Zones and subnets** dropdown list, choose **Private Subnet 1 (10.0.2.0/24)** and **Private Subnet 2 (10.0.4.0/24)**.

Red [Info](#)

Para la mayoría de las aplicaciones, puede utilizar varias zonas de disponibilidad y dejar que EC2 Auto Scaling equilibre sus instancias entre las zonas. La VPC predeterminada y las subredes predeterminadas son adecuadas para comenzar rápidamente.

VPC
Elija la VPC que define la red virtual para el grupo de Auto Scaling.

vpc-06915e49a72034f56 (Lab VPC)
10.0.0.0/16

[Crear una VPC](#)

Zonas de disponibilidad y subredes
Defina qué zonas de disponibilidad y subredes puede utilizar el grupo de Auto Scaling en la VPC elegida.

Seleccionar zonas de disponibilidad y subredes

us-west-2a | subnet-0d713db24d83abfa0 (Private Subnet 1)
10.0.2.0/24

us-west-2b | subnet-055159c7873d1c359 (Private Subnet 2)
10.0.4.0/24

[Crear una subred](#)

56. Choose **Next**.

57. On the **Configure advanced options – optional** page, configure the following options:

- In the **Load balancing – optional** section, choose **Attach to an existing load balancer**.

Balance de carga [Info](#)

Utilice las siguientes opciones para asociar su grupo de Auto Scaling a un balanceador de carga existente o a uno nuevo que defina.

☐ No se encontró ningún balanceador de carga
El tráfico a su grupo de Auto Scaling no se llevará a cabo por un balanceador de carga.

☒ Asociar a un balanceador de carga existente
Elija entre los balanceadores de carga existentes.

☐ Asociar a un nuevo balanceador de carga
Cree rápidamente un balanceador de carga básico para asociarlo al grupo de Auto Scaling.

- In the **Attach to an existing load balancer** section, configure the following options:
 - Choose **Choose from your load balancer target groups**.
 - From the **Existing load balancer target groups** dropdown list, choose **webserver-app | HTTP**.

Asociar a un balanceador de carga existente

Seleccione los balanceadores de carga que desea asociar al grupo de Auto Scaling.

- ☒ Elegir entre los grupos de destino del balanceador de carga
- Esta opción le permite asociar balanceadores de carga de puerta de enlace, red o aplicaciones.

- ☐ Elegir entre balanceadores de carga clásicos

Grupos de destino del balanceador de carga existentes

Solo están disponibles para su selección los grupos de destino de instancias que pertenecen a la misma VPC que el grupo de Auto Scaling.

Seleccionar grupos de destino

webserver-app | HTTP X
Application Load Balancer: WebServerELB

- In the **Health checks** section, under **Additional health check types**, select **Turn on Elastic Load Balancing health checks**.

Comprobaciones de estado

Las comprobaciones de estado aumentan la disponibilidad reemplazando instancias en mal estado. Cuando se utilizan varias comprobaciones de estado, se evalúan todas y, si se produce un error en al menos una, se lleva a cabo la sustitución de instancias.

Comprobaciones de estado de EC2

[Siempre habilitadas](#)

Tipos de comprobaciones de estado adicionales - *opcional* | [Info](#)

- ☒ Activar las comprobaciones de estado de Elastic Load Balancing **Recomendado**
- Elastic Load Balancing puede monitorear si las instancias están disponibles para gestionar solicitudes. Cuando informa de una instancia en mal estado, EC2 Auto Scaling puede sustituirla en la siguiente comprobación periódica.

[i](#) EC2 Auto Scaling comenzará a detectar y actuar en función de las comprobaciones de estado realizadas por Elastic Load Balancing. Para evitar terminaciones inesperadas, primero verifique la configuración de estas comprobaciones de estado en la [consola del equilibrador de carga](#) [↗](#)

58. Choose **Next**.

59. On the **Configure group size and scaling policies – *optional*** page, configure the following options:

- In the **Group size – *optional*** section, enter the following values:
 - **Desired capacity:** 2

Tamaño del grupo [Info](#)

Defina el tamaño inicial del grupo de escalamiento automático. Después de crear el grupo, puede cambiar su tamaño para satisfacer la demanda, ya sea en forma manual o mediante el escalamiento automático.

Tipo de capacidad deseado

Elija la unidad de medida para el valor de capacidad deseado. Las vCPU y la memoria (GiB) solo son compatibles con grupos de instancias mixtos configurados con un conjunto de atributos de instancia.

Unidades (número de instancias) ▼

Capacidad deseada

Especifique el tamaño de su grupo.

4

- Minimum capacity: 2
- Maximum capacity: 4

Escalado [Info](#)

Puede cambiar el tamaño de su grupo de escalamiento automático de forma manual o automática para cumplir con los cambios en la demanda.

Límites de escalamiento

Establezca límites sobre cuánto puede aumentarse o disminuirse la capacidad deseada.

Capacidad deseada mínima

2

Capacidad igual o inferior a la deseada

Capacidad deseada máxima

4

Capacidad igual o superior a la deseada

- In the **Scaling policies – optional** section, configure the following options:
 - Choose **Target tracking scaling policy**.
 - For **Metric type**, choose **Average CPU utilization**.
 - For **Target value**, enter 50

This change tells auto scaling to maintain an average CPU utilization across all instances of 50 percent. Auto scaling automatically adds or removes capacity as required to keep the metric at or close to the specified target value. It adjusts to fluctuations in the metric due to a fluctuating load pattern.

Escalamiento automático - *opcional*

Elija si desea utilizar una política de seguimiento de destino [Info](#)

Puede configurar otras políticas de escalado basadas en métricas y un escalado programado después de crear su grupo de escalamiento automático.

☐ Sin políticas de escalamiento

Su grupo de escalamiento automático mantendrá su tamaño inicial y no se redimensionará de forma dinámica para satisfacer la demanda.

☒ Política de escalado de seguimiento de destino

Elija una métrica y un valor objetivo de CloudWatch y deje que la política de escalamiento ajuste la capacidad deseada en proporción al valor de la métrica.

Nombre de la política de escalado

Target Tracking Policy

Tipo de métrica [Info](#)

Métrica supervisada que determina si la utilización de recursos es demasiado baja o alta. Si utiliza métricas de EC2, considere la posibilidad de habilitar la supervisión detallada para obtener un mejor rendimiento de escalado.

Utilización promedio de la CPU

Valor de destino

50

Preparación de la instancia [Info](#)

300

segundos

60. Choose **Next**.

61. On the **Add notifications - *optional*** page, choose **Next**.

Añadir notificación - *opcional* [Info](#)

Envíe notificaciones a temas de SNS siempre que Amazon EC2 Auto Scaling lance o termine las instancias EC2 de su grupo de Auto Scaling.

Añadir notificación

Cancelar

Omitir para revisar

Anterior

Siguiente

62. On the **Add tags - *optional*** page, choose **Add tag** and configure the following options:

- For **Key**, enter **Name**
- For **Value - optional**, enter **WebApp**

Etiquetas (1)

Clave

Valor: opcional

Etiquetar instancias nuevas

☒

Eliminar

Agregar etiqueta

49 restante

Cancelar

Anterior

Siguiente

63. Choose **Next**.

64. On the **Review** page, choose **Create Auto Scaling group**.

These options launch EC2 instances in private subnets across both Availability Zones.

Your Auto Scaling group initially shows an **Instances** count of zero, but new instances will be launched to reach the desired count of two instances.

Note: If you experience an error related to the t3.micro instance type not being available, then rerun this task by choosing the t2.micro instance type instead.

Paso 5: Añadir notificaciones

Editar

Notificaciones

Sin notificaciones

Paso 6: Añadir etiquetas

Editar

Etiquetas (1)

Clave	Valor	Etiquetar instancias nuevas
Name	WebApp	Sí

Cancelar

Anterior

Crear grupo de Auto Scaling

Task 3: Verifying the auto scaling configuration

In this task, you verify that both the auto scaling configuration and the load balancer are working by accessing a pre-installed script on one of your servers that will consume CPU cycles, which invokes the scale out alarm.

65. In the left navigation pane, choose **Instances**.

Two new instances named **WebApp** are being created as part of your Auto Scaling group. While these instances are being created, the **Status check** for these two instances is *Initializing*.

Observe the **Status check** field for the instances until the status is *2/2 checks passed*. Wait for the two new instances to complete initialization before you proceed to the next step.

You might need to choose **Refresh** to see the updated status.

Instancias (4) Información								
<input type="text" value="Buscar instancia por atributo o etiqueta (case-sensitive)"/>					Todos los e...		< 1 >	
<input type="checkbox"/>	Name	ID de la instancia	Estado de la i...	Tipo de inst...	Comprobación de	Estado de la al	Zona de dispon...	DNS de
<input type="checkbox"/>	WebApp	i-08809efe9225bbcdc	En ejecución	t3.micro	Inicializando	Ver alarmas +	us-west-2b	-
<input type="checkbox"/>	WebServer	i-077265d5840673db4	En ejecución	t3.micro	2/2 comprobacion	Ver alarmas +	us-west-2b	ec2-35-
<input type="checkbox"/>	WebApp	i-0a3752a7b31bd262b	En ejecución	t3.micro	Inicializando	Ver alarmas +	us-west-2a	-
<input type="checkbox"/>	Command Host	i-0405b2f9ff69f0d69	En ejecución	t3.medium	2/2 comprobacion	Ver alarmas +	us-west-2a	ec2-35-

66. Once the instances have completed initialization, in the left navigation pane in the **Load Balancing** section, choose **Target Groups**, and then select your target group, **webserver-app**.

[EC2](#) > [Grupos de destino](#) > webserver-app

webserver-app

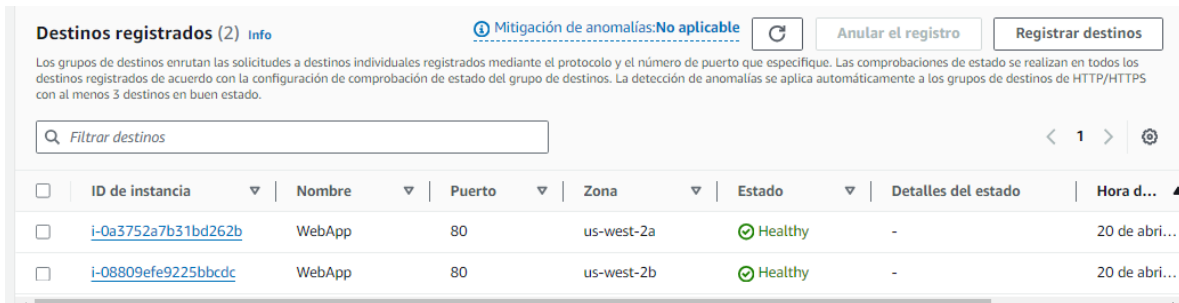
Acciones

Detalles
 `arn:aws:elasticloadbalancing:us-west-2:533266983934:targetgroup/webserver-app/d61f0339f30a153f`

Tipo de destino Instancia	Protocolo : Puerto HTTP: 80	Versión del protocolo HTTP1	VPC vpc-06915e49a72034f56
Tipo de dirección IP IPv4	Balanceador de carga Ninguno asociado		
2 Destinos totales	2 En buen estado 0 Anómalo	0 En mal estado	0 Sin utilizar
		0 Inicial	0 Vaciado

67. On the **Targets** tab, verify that two instances are being created. Refresh this list until the **Health status** of these instances changes to *healthy*.

You can now test the web application by accessing it through the load balancer.



Destinos registrados (2) [Info](#) [Mitigación de anomalías: No aplicable](#) [Anular el registro](#) [Registrar destinos](#)

Los grupos de destinos enrutan las solicitudes a destinos individuales registrados mediante el protocolo y el número de puerto que especifique. Las comprobaciones de estado se realizan en todos los destinos registrados de acuerdo con la configuración de comprobación de estado del grupo de destinos. La detección de anomalías se aplica automáticamente a los grupos de destinos de HTTP/HTTPS con al menos 3 destinos en buen estado.

<input type="checkbox"/>	ID de instancia	Nombre	Puerto	Zona	Estado	Detalles del estado	Hora d...
<input type="checkbox"/>	i-Oa3752a7b31bd262b	WebApp	80	us-west-2a	Healthy	-	20 de abri...
<input type="checkbox"/>	i-08809efe9225bbcdc	WebApp	80	us-west-2b	Healthy	-	20 de abri...

Task 4: Testing auto scaling configuration

68. Open a new web browser tab, and paste the **DNS name** of the load balancer that you copied earlier into the address bar, and press Enter.

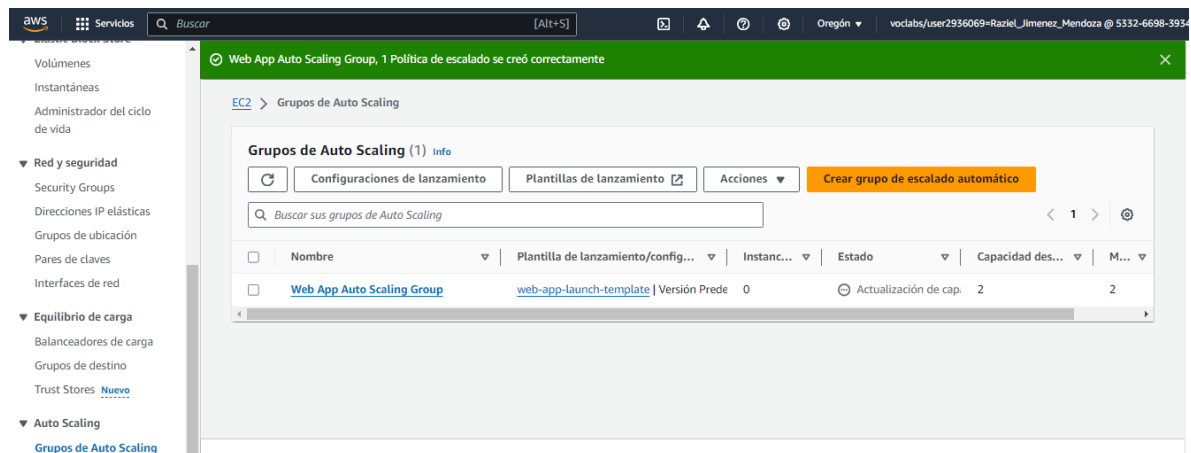


69. On the web page, choose **Start Stress**.

This step calls the application **stress** in the background, which causes the CPU utilization on the instance that serviced this request to spike to 100 percent.



70. On the EC2 Management console, in the left navigation pane in the **Auto Scaling** section, choose **Auto Scaling Groups**.

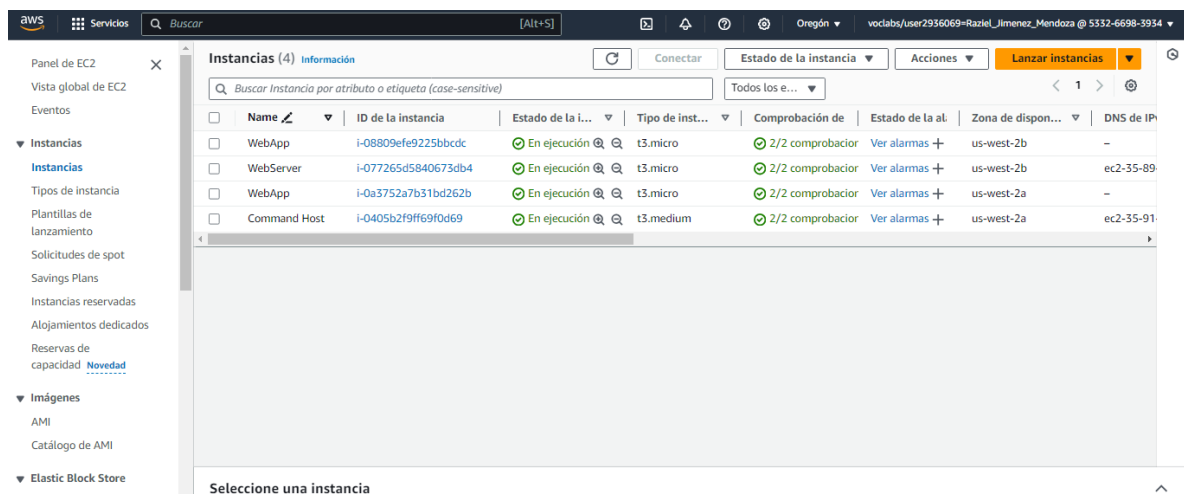


71. Select **Web App Auto Scaling Group**.

72. Choose the **Activity** tab.

After a few minutes, you should see your Auto Scaling group add a new instance. This occurs because Amazon CloudWatch detected that the average CPU utilization of your Auto Scaling group exceeded 50 percent, and your scale-up policy has been invoked in response.

You can also check the new instances being launched on the EC2 Dashboard.



Conclusion

Congratulations! You now have successfully done the following:

- Created an EC2 instance by using an AWS CLI command
- Created a new AMI by using the AWS CLI
- Created an Amazon EC2 launch template
- Created an Amazon EC2 Auto Scaling launch configuration
- Configured scaling policies and created an Auto Scaling group to scale in and scale out the number of servers based on variable load

Lab complete

Congratulations! You have completed the activity.

73. At the top of this page, choose **End Lab** and then choose **Yes** to confirm that you want to end the activity.

A panel appears indicating that "You may close this message box now. Lab resources are terminating."

74. To close the **End Lab** panel, choose the X in the upper-right corner.