

175- [JAWS] - Lab - Uso de Auto Scaling en AWS (Linux)

Datos Generales:

Nombre: Tomás Alfredo Villaseca Constantinescu

País: Chile

Fecha: 21/10/2023

Contacto: tomas.villaseca.c@gmail.com

Después de completar este laboratorio, usted será capaz de hacer lo siguiente:

- Crear una instancia EC2 mediante un comando de la CLI de AWS.
- Crear una nueva AMI mediante la CLI de AWS.
- Crear una plantilla de lanzamiento de Amazon EC2.
- Crear una configuración de lanzamiento de Amazon EC2 Auto Scaling.
- Configurar políticas de escalado y crear un grupo de Auto Scaling para aumentar y reducir el número de servidores en función de una carga variable.

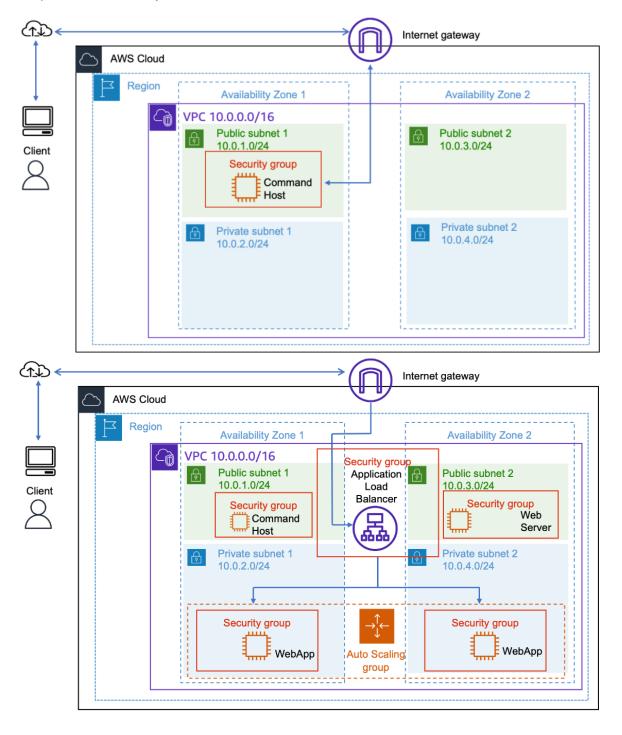
Resumen Laboratorio:

En este laboratorio, utilizará la AWS CLI para crear una instancia EC2 para alojar un servidor web y crear una AMI a partir de dicha instancia.

Utilizará esa AMI como base para lanzar un sistema que se escala automáticamente bajo una carga variable utilizando Amazon EC2 Auto Scaling.

También creará un Elastic Load Balancer para distribuir la carga entre las instancias EC2 creadas en varias zonas de disponibilidad mediante la configuración de escalado automático.

Arquitectura Inicial y final:



Tarea 1: Crear una nueva AMI para Amazon EC2 Auto Scaling

En esta tarea, lanzará una nueva instancia EC2 y luego creará una nueva AMI basada en esa instancia en ejecución.

1.1 - Conectarse a la instancia Command Host

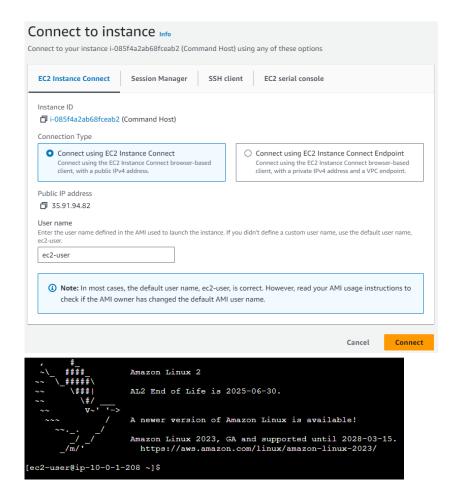
En esta tarea, se utiliza EC2 Instance Connect para conectarse a la instancia Command Host que se creó cuando se aprovisionó el laboratorio.

Esta instancia se utiliza para ejecutar comandos de AWS CLI.

Paso 1: AWS Management Console → Search → EC2 → Instances → Command Host

Instances (1/1) Info							
Q Find instance by attribute or tag (case-sensitive)							
~	Name	Instance ID	Instance state	Instance type			
✓	Command Host	i-085f4a2ab68fceab2	⊘ Running ② ○	t3.medium			

Paso 2: Command Host → Connect → EC2 Instance Connect



1.2 - Configurar AWS CLI en Command Host

La AWS CLI está preconfigurada en la instancia Command Host.

Paso 1: Para confirmar que la región en la que se está ejecutando la instancia del host de comandos es la misma que la del laboratorio (us-west-2), ejecute el siguiente comando:

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

```
[ec2-user@ip-10-0-1-208 ~]$ curl http://169.254.169.254/latest/dynamic/instance-identity/document | grep region % Total % Received % Xferd Average Speed Time Time Time Current

Dload Upload Total Spent Left Speed

100 475 100 475 0 0 277k 0 --:--:-- --:--- 463k

"region": "us-west-2",

[ec2-user@ip-10-0-1-208 ~]$
```

Paso 2: Para actualizar AWS CLI con las credenciales correctas utilizar el comando aws configure

```
aws configure
```

- AWS Access Key ID → Apretar Enter
- AWS Secret Access Key → Apretar Enter
- Default region name = us-west-2
- Default output format = json

```
[ec2-user@ip-10-0-1-208 ~]$ aws configure
AWS Access Key ID [None]:
AWS Secret Access Key [None]:
Default region name [us-west-2]: us-west-2
Default output format [None]: json
[ec2-user@ip-10-0-1-208 ~]$
```

Paso 3: Ejecutar el siguiente comando para acceder al directorio donde se encuentran el script:

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

```
[ec2-user@ip-10-0-1-208 ~]$ cd /home/ec2-user/
[ec2-user@ip-10-0-1-208 ~]$ pwd
/home/ec2-user
[ec2-user@ip-10-0-1-208 ~]$ ls
UserData.txt
[ec2-user@ip-10-0-1-208 ~]$
```

1.3 – Crear una nueva instancia EC2

En esta tarea, se utilizará AWS CLI para crear una nueva instancia que aloje un servidor web.

Paso 1: Inspeccionar el script UserData.txt

```
more UserData.txt
```

```
[ec2-user@ip-10-0-1-208 ~]$ more UserData.txt

#!/bin/bash
yum update -y --security
amazon-linux-extras install epel -y
yum -y install httpd php stress
systemctl enable httpd.service
systemctl start httpd
cd /var/www/html
wget http://aws-tc-largeobjects.s3.amazonaws.com/CUR-TF-100-TULABS-1/10-lab-autoscaling-linux/s3/ec2-stress.zip
unzip ec2-stress.zip
echo 'UserData has been successfully executed. ' >> /home/ec2-user/result
find -wholename /root/.*history -wholename /home/*/.*history -exec rm -f {} \;
find / -name 'authorized_keys' -exec rm -f {} \;
rm -rf /var/lib/cloud/data/scripts/*
[ec2-user@ip-10-0-1-208 ~]$
```

Este script realiza una serie de tareas de inicialización, incluyendo la actualización de todo el software instalado en la caja y la instalación de una pequeña aplicación web PHP que puede utilizar para simular una alta carga de CPU en la instancia.

Paso 2: Ejecutar el siguiente comando reemplazando los valores correspondientes por los entregados en los detalles del laboratorio para lanzar una nueva instancia EC2.

- AMIID = ami-0025f0db847eb6254
- HTTPACCESS = sg-07975c12874a8573a
- KEYNAME = vockey
- SUBNETID = subnet-0bd28d44407aae936

```
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'
```

```
[ec2-user@ip-10-0-1-208 ~]$ aws ec2 run-instances 73a --subnet-id subnet-0bd28d44407aae936 --associ i-0e9482d229a814b32 [ec2-user@ip-10-0-1-208 ~]$
```

La salida de este comando le proporciona un Instanceld.

Instance ID = i-0e9482d229a814b32

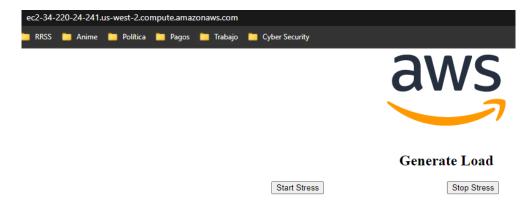
Paso 2: Para obtener el Public DNS name ejecutar el siguiente comando:

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

```
[ec2-user@ip-10-0-1-208 ~]$ aws ec2 describe-instances --in
"ec2-34-220-24-241.us-west-2.compute.amazonaws.com"
[ec2-user@ip-10-0-1-208 ~]$
```

Public DNS name = ec2-34-220-24-241.us-west-2.compute.amazonaws.com

Paso 4: Copiar Public DNS name en una nueva pestaña de navegador web.



1.4 - Crear una nueva instancia EC2

En esta tarea, creará una nueva AMI basada en la instancia que se acaba de crear.

Paso 1: Ejecutar el siguiente comando para crear una nueva AMI:

• Reemplazar NEW-INSTANCE-ID por Instance ID obtenido anteriormente.

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

Por defecto, el comando aws ec2 create-image reinicia la instancia actual antes de crear la AMI para asegurar la integridad de la imagen en el sistema de archivos.

```
[ec2-user@ip-10-0-1-208 ~]$ aws ec2 create-image --name WebServerAMI --instance-id i-0e9482d229a814b32 {
    "ImageId": "ami-021504398952a42d3"
}
[ec2-user@ip-10-0-1-208 ~]$
```

• AMI ID = ami-021504398952a42d3

Tarea 2: Crear un entorno de Auto Scaling

En esta tarea, se creará un equilibrador de carga que agrupa un grupo de instancias EC2 bajo una única dirección DNS.

Utilizará el escalado automático para crear un grupo escalable dinámicamente de instancias EC2 basado en la AMI que creó en la tarea anterior.

Por último, creará un conjunto de alarmas que reduzcan o aumenten el número de instancias de su grupo de equilibradores de carga siempre que el rendimiento de la CPU de cualquier máquina del grupo supere o quede por debajo de un conjunto de umbrales especificados.

2.1 - Crear un Application Load Balancer

En esta tarea, se creará un equilibrador de carga que puede equilibrar el tráfico entre varias instancias EC2 y zonas de disponibilidad.

Paso 1: EC2 → Load Balancing → Load Balancers → Create Load Balancer

▼ Load Balancing				
Load Balancers	G	Actions ▼	Create load balancer	•
Target Groups				

Paso 2: Create Load Balancer → Load Balancer Types

Application Load Balancer → Create

Load balancer types							
Application Load Balancer Info	Network Load Balancer Info	Gateway Load Balancer Info					

Paso 3: Create Load Balancer → Basic Configuration

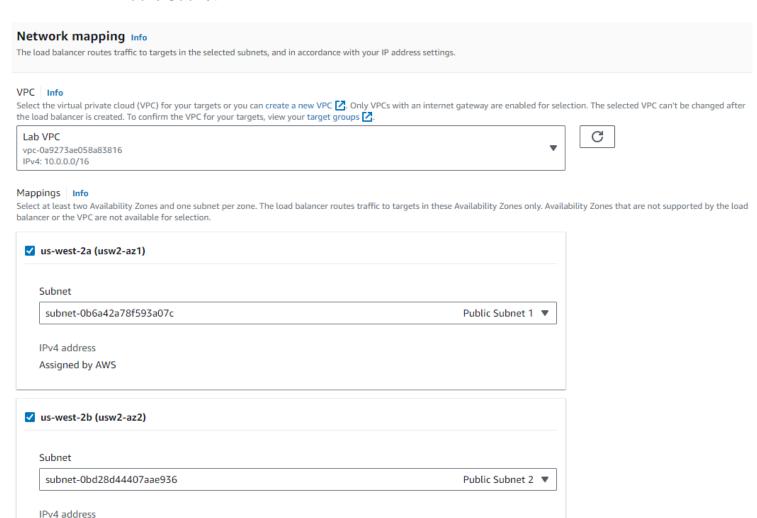
Load Balancer Name = WebServerELB

Basic configuration Load balancer name Name must be unique within your AWS account and can't be changed after the load balancer is created. WebServerELB

A maximum of 32 alphanumeric characters including hyphens are allowed, but the name must not begin or end with a hyphen.

Paso 4: Create Load Balancer → Network Mapping

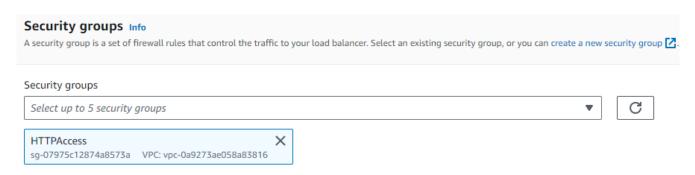
- VPC → Lab VPC
- Mappings → Selectionar ambas AZs disponibles.
- AZ 1 → Public Subnet 1
- AZ 2 → Public Subnet 2



Paso 5: Create Load Balancer → Security Groups

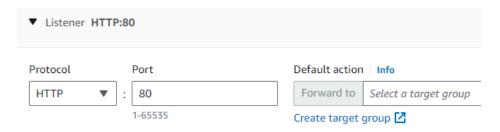
Desplegar lista → HTTPAccess

Assigned by AWS



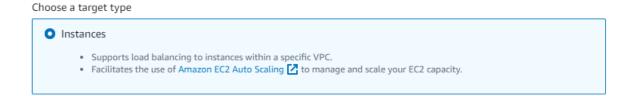
Paso 6: Create Load Balancer → Listeners and routing

Create target group → Link → Se abre una nueva pestaña del navegador web



Paso 7: Create target group

Target type → Instances



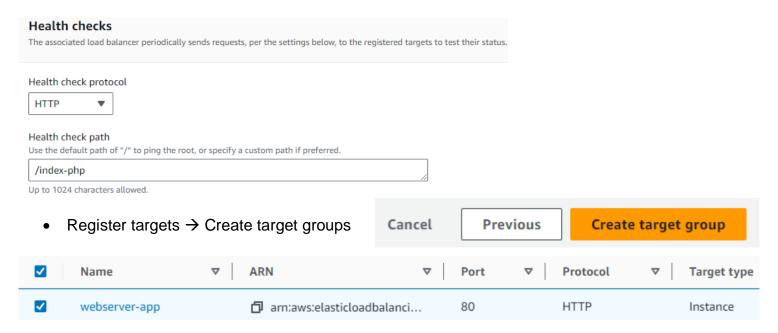
Target group name = webserver-app

Target group name

webserver-app

A maximum of 32 alphanumeric characters including hyphens are allowed, but the name must not begin or end with a hyphen.

Health checks → Health check path → /index.php



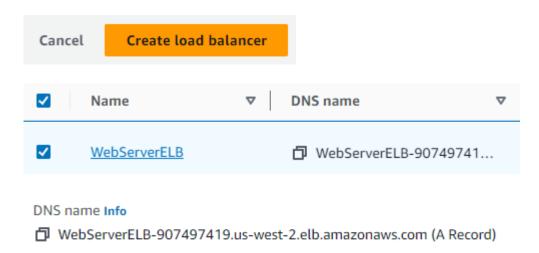
Paso 8: Volver a la pestaña de navegador de Load Balancers.

Refrescar → Desplegar lista → webserver-app



Paso 9: Create Load Balancer → Create

- View Load Balancer
- Copiar DNS name = WebServerELB-907497419.us-west-2.elb.amazonaws.com



2.2 - Crear un Launch Template

En esta tarea, se crea una plantilla de lanzamiento (launch template) para el Auto Scaling Group.

Una plantilla de lanzamiento es una plantilla que un Auto Scaling Group utiliza para lanzar instancias EC2.

Al crear una plantilla de lanzamiento, se especifica información para las instancias, como la AMI, el tipo de instancia, el par de claves, el grupo de seguridad y los discos.

Paso 1: EC2 → Instances → Launch Templates → Create Launch Template

▼ Instances Instances Instance Types Launch Templates New launch template Create launch template

Paso 2: Create Launch Template → Name and Description

- Launch Template Name = web-app-launch-template
- Template version description = a web server for the load test app.
- Auto Scaling Guidance → Seleccionar casilla asociada.

Launch template name and description

Launch template name - required

web-app-launch-template

Must be unique to this account. Max 128 chars. No spaces or special characters like '&', '*', '@'.

Template version description

a web server for the load test app

Max 255 chars

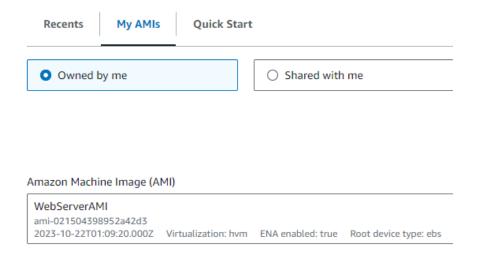
Auto Scaling guidance Info

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

Provide guidance to help me set up a template that I can use with EC2 Auto Scaling

Paso 3: Create Launch Template → Application and OS images

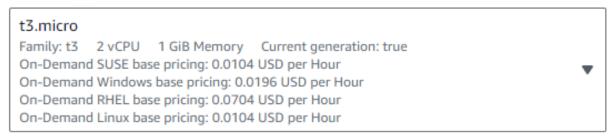
My AMIs → WebServerAMI



Paso 4: Create Launch Template → Instance type

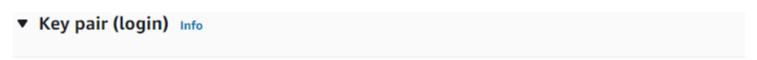
Instance type → t3.micro

Instance type



Paso 5: Create Launch Template → Key pair (login)

Desplegar lista → Don't include in launch template



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

Don't include in launch template

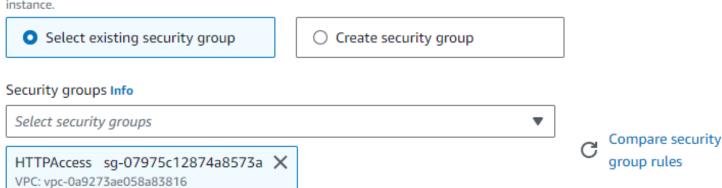
✓ Create new key pair

Paso 6: Create Launch Template → Network Settings

Security Groups → HTTPAccess

Firewall (security groups) Info

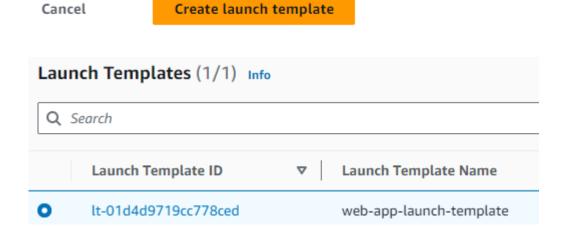
A security group is a set of firewall rules that control the traffic for your instance. Add rules to allow specific traffic to reach your instance.



Paso 7: Create Launch Template → Create

View Launch Template

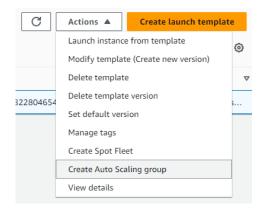
Cancel



2.3 - Crear un Auto Scaling Group

En esta tarea, utilizará la launch template para crear un Auto Scaling Group.

Paso 1: Launch Templates → web-app-launch-template → Actions → Create Auto Scaling Group



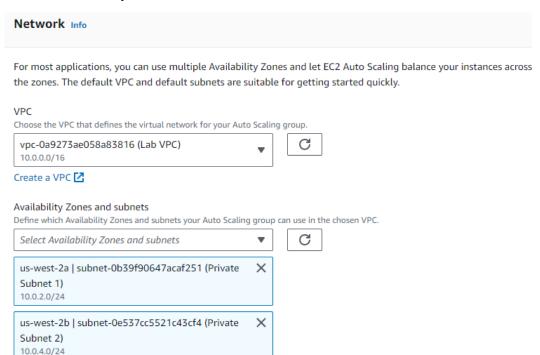
Paso 2: Create Auto Scaling Group → Name

Auto Scaling Group Name = Web App Auto Scaling Group



Paso 3: Create Auto Scaling Group → Network

- VPC → Lab VPC
- Availability Zones and subnets → Private Subner 1 / Private Subnet 2

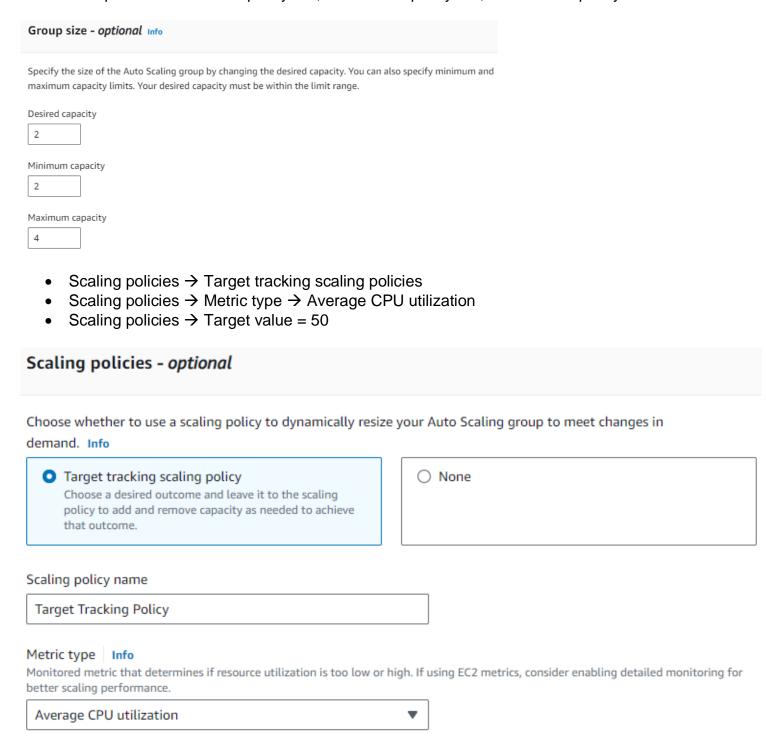


Paso 4: Create Auto Scaling Group → Configure advanced options

Load balancing → Attach to an existing load balancer 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 Attach to an existing load Attach to a new load Traffic to your Auto Scaling group balancer balancer will not be fronted by a load Quickly create a basic load Choose from your existing load balancer. balancers. balancer to attach to your Auto Scaling group. Existing load balancer target groups → webserver-app | HTTP Choose from your load balancer target groups Choose from Classic Load Balancers This option allows you to attach Application, Network, or Gateway 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. C Select target groups × webserver-app | HTTP Application Load Balancer: WebServerELB Health checks → Additional health check types → Seleccionar casilla ELB Health checks Health checks increase availability by replacing unhealthy instances. When you use multiple health checks, all are evaluated, a 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 Scaling can replace it on its next periodic check.

Paso 5: Create Auto Scaling Group → Configure group size and scaling policies

Group size → Desired capacity = 2, Minimum capacity = 2, Maximum capacity = 4

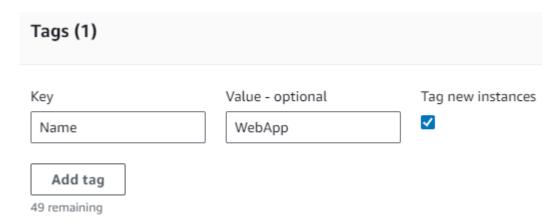


Target value

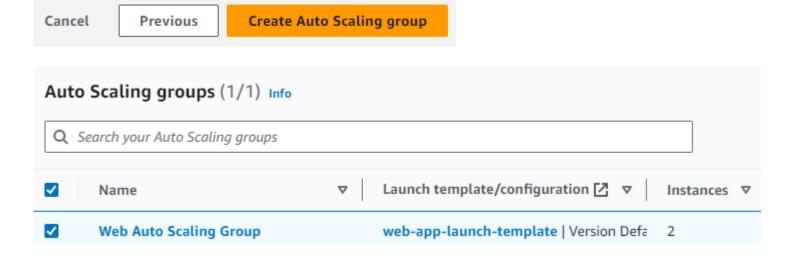
50

Paso 6: Create Auto Scaling Group → Add tags → Add tag

- Key = Name
- Value = WebApp



Paso 7: Create Auto Scaling Group → Create

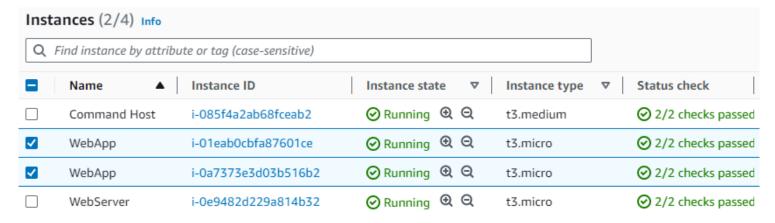


Tarea 3: Verificar la configuración de Auto Scaling

En esta tarea, verificará que tanto la configuración de Auto Scaling como el equilibrador de carga funcionan accediendo a un script preinstalado en uno de sus servidores que consumirá ciclos de CPU, lo que invoca la alarma de escalado.

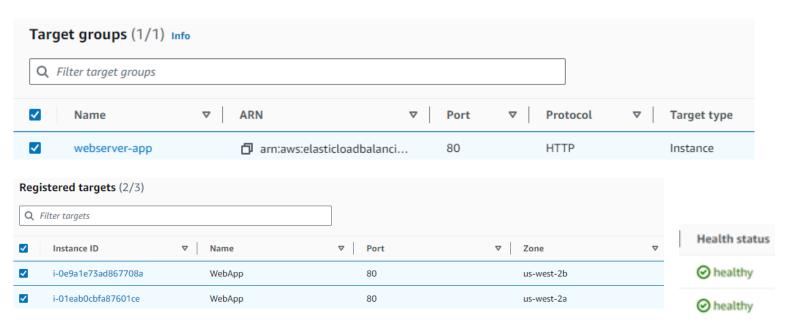
Paso 1: EC2 → Instances

• Se puede ver 2 nuevas instancias WebApp creadas por Auto Scaling Group.



Paso 2: EC2 → Load balancing → Target Groups → webserver-app

Targets → Puede ver el status de ambas instancias → Healthy Status.



Tarea 4: Probar la configuración de Auto Scaling

Paso 1: Abrir una nueva pestaña de navegador y copiar DNS name del Load Balancer.



Paso 2: En la página web seleccionar "Start Stress".

Generating load

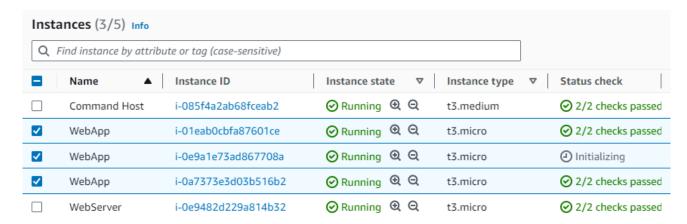
Start Stress

Paso 3: EC2 → Auto Scaling → Auto Scaling Group → Web App Auto Scaling Group → Activity

- Tras unos minutos se puede ver que su Auto Scaling Group añade una nueva instancia.
- Lo anterior ocurre porque Amazon CloudWatch detectó que la utilización media de la CPU de su grupo de Auto Scaling superó el 50 por ciento y su política de escalado se ha invocado en respuesta.



Paso 4: Comprobar que las nuevas instancias EC2 se están lanzado en el panel de instancias de EC2.



Laboratorio Completado

