

# Lab - 175

## Uso de escalado automático en AWS (Linux)



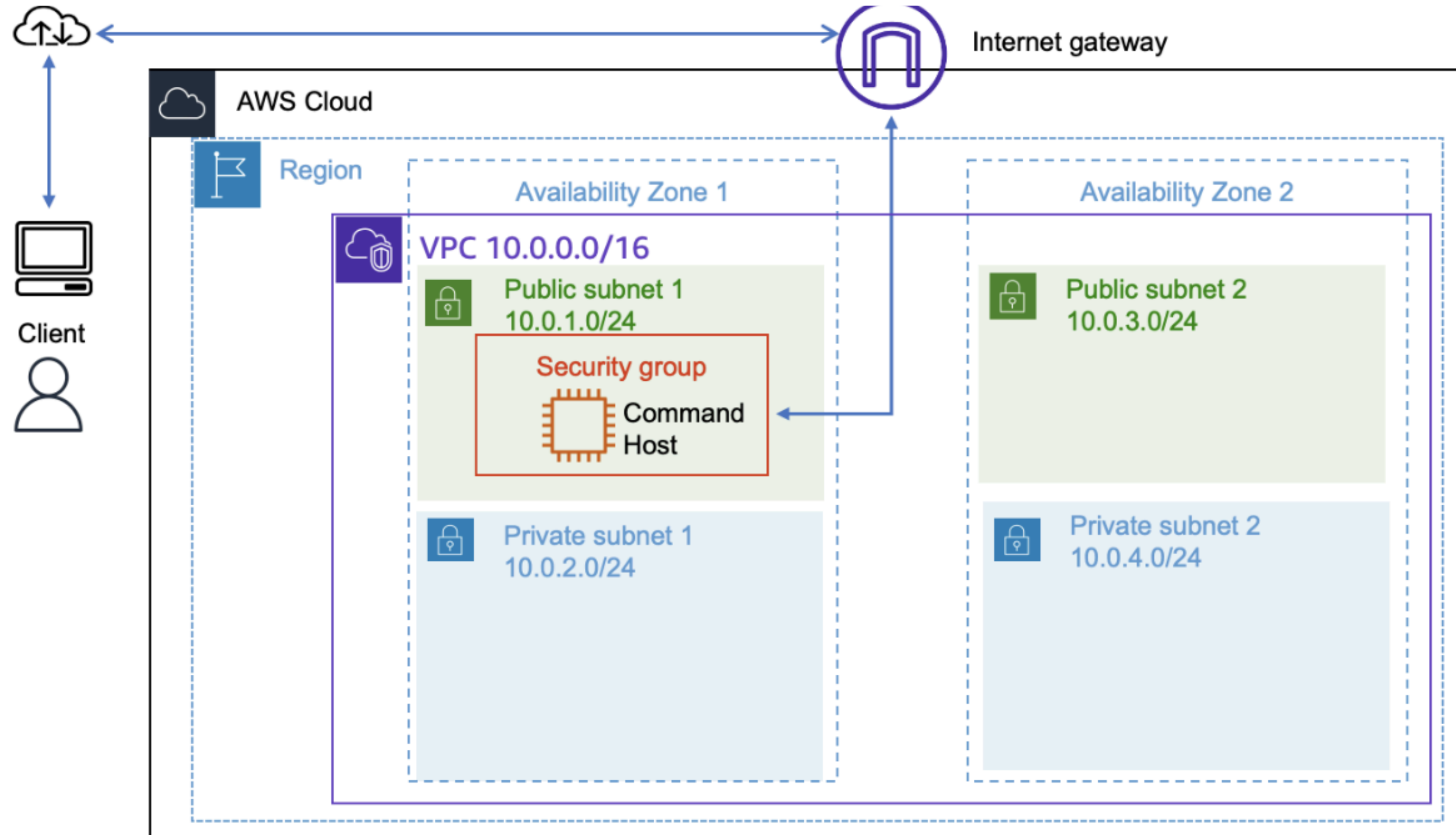
Benjamin Sabaño, Sebastian Aguilera , Joel Umpierrez,  
Juan Sansberro, Ignacio Suárez



# Objetivos

- **Crear una instancia de EC2 mediante la utilización de un comando de la AWS CLI.**
- **Crear una AMI nueva mediante la utilización de la AWS CLI.**
- **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 reducir y escalar horizontalmente el número de servidores con base en la carga variable.**

# Arquitectura inicial del Laboratorio



# Tarea 1: crear una AMI nueva para Amazon EC2 Auto Scaling

## Tarea 1.1: conectarse a la instancia Command Host

Usará **EC2 Instance Connect** para conectarse a la **instancia de EC2 Command Host** que se creó cuando se aprovisionó el laboratorio

En la **Consola de administración de AWS**, en la barra de búsqueda, ingrese y seleccione **EC2** para abrir la **Consola de administración de Elastic Compute Cloud**.

En el panel de navegación, seleccione **Instancias**.

En la lista de instancias, seleccione la instancia **Command Host**.

Elija **Conectar**.

En la pestaña **EC2 Instance Connect**, seleccione **Conectar**

Connect to instance [Info](#)

Connect to your instance i-09ef573ebf9bb957d (Command Host) using any of these options

**EC2 Instance Connect** | Session Manager | SSH client | EC2 serial console

Instance ID  
i-09ef573ebf9bb957d (Command Host)

Connection Type

☒ Connect using EC2 Instance Connect  
Connect using the EC2 Instance Connect browser-based client, with a public IPv4 address.

☐ Connect using EC2 Instance Connect Endpoint  
Connect using the EC2 Instance Connect browser-based client, with a private IPv4 address and a VPC endpoint.

Public IP address  
54.190.136.197

Username  
Enter the username defined in the AMI used to launch the instance. If you didn't define a custom username, use the default username, ec2-user.

ec2-user

**Note:** In most cases, the default username, ec2-user, is correct. However, read your AMI usage instructions to check if the AMI owner has changed the default AMI username.

Cancel **Connect**

## Tarea 1.2: configurar la AWS CLI

Ejecute el siguiente comando para confirmar que la región en la cual se ejecuta la instancia **Command Host** es la misma que la del laboratorio (la región us-west-2).

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

Ejecute el siguiente comando para actualizar el software de la **AWS CLI** con las credenciales correctas:

**aws configure**

```
[ec2-user@ip-10-0-1-133 ~]$ curl http://169.254.169.254/latest/dynamic/instance-identity/document | grep region
% Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
           %             0         0    212k         0 --:--:-- --:--:-- --:--:--  231k
"region" : "us-west-2",
[ec2-user@ip-10-0-1-133 ~]$ aws configure
AWS Access Key ID [None]: ASIAR6QQ36FBV7HWZT5L
AWS Secret Access Key [None]: hNjD41XXBV594Cd825Z3sWVgKOVjsDM32yOpVaZa
Default region name [us-west-2]: us-west-2
Default output format [None]: json
[ec2-user@ip-10-0-1-133 ~]$ cd /home/ec2-user/
```

Cuando se le solicite, ingrese la siguiente información:

- **AWS Access Key ID:** presione Intro.
- **AWS Secret Access Key:** presione Intro.
- **Default region name:** ingrese el nombre de la región según los pasos anteriores de esta tarea (por ejemplo, us-west-2). Si la región ya aparece, presione Intro.
- **Default output format:** ingrese json.



# Tarea 1.3: crear una instancia de EC2 nueva

En esta tarea, utilizaremos la AWS CLI para crear una instancia nueva que alojará un servidor web. y usaremos el comando 'more UserData.txt' para inspeccionar el script,

A continuación extraemos del apartado de Detalles de Vocareum las credenciales para elaborar el siguiente script donde cambiamos el KEYNAME, AMMID, HTTPACCESS y SUBNETID

```
ec2-user@ip-10-0-1-133 ~]$ 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/*
```

```
aws ec2 run-instances --key-name vockey --instance-type t3.micro --image-id ami-0ef45c14a8b7161ce --user-data file:///home/ec2-user/UserData.txt --security-group-ids sg-0004d06234f949648 --subnet-id subnet-07ad3792b9e7f3fe7 --associate-public-ip-address --tag-specifications 'ResourceType=instance,Tags=[{Key=Name,Value=WebServer}]' --output text --query 'Instances[*].InstanceId'
```

Este script nos proporcionará una ID de instancia nueva que utilizaremos en el transcurso del lab

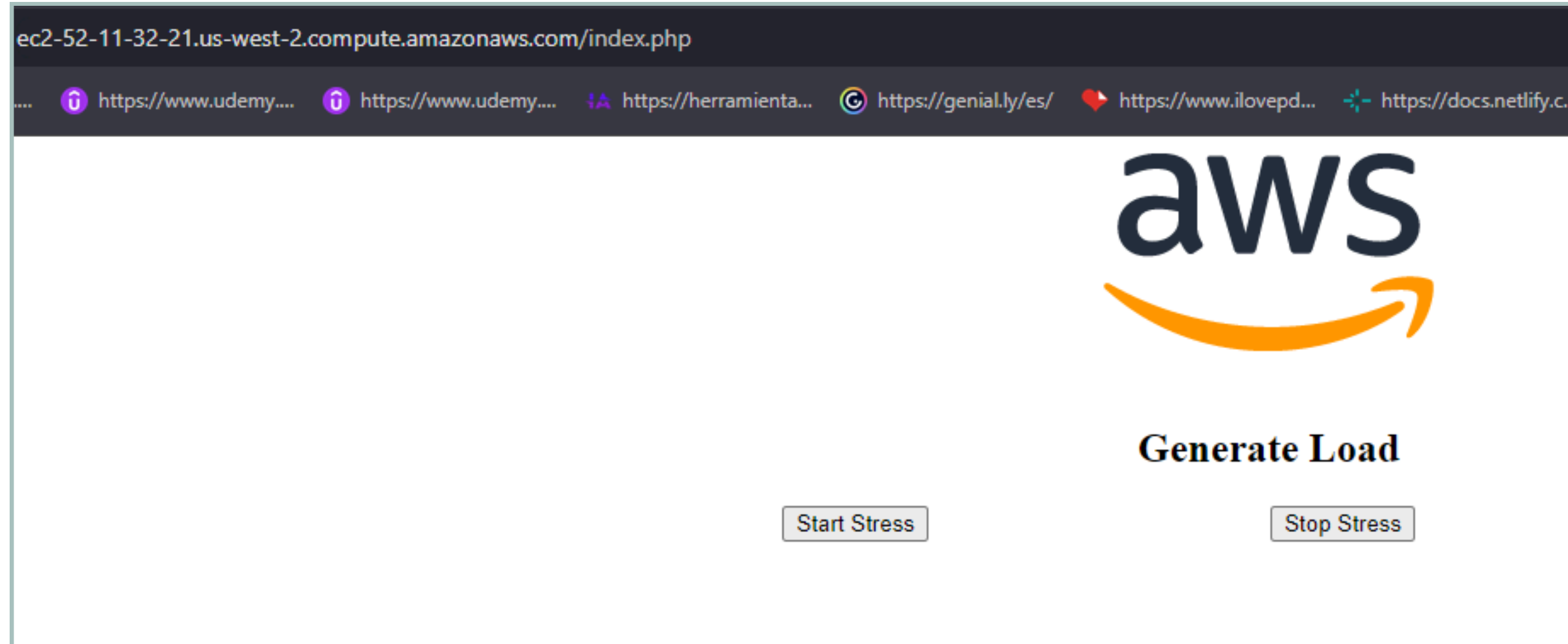
```
ec2-user@ip-10-0-1-133 ~]$ aws ec2 run-instances --key-name vockey --instance-type t3.micro --image-id ami-0ef45c14a8b7161ce --user-data file:///home/ec2-user/UserData.txt --security-group-ids sg-0004d06234f949648 --subnet-id subnet-07ad3792b9e7f3fe7 --associate-public-ip-address --tag-specifications 'ResourceType=instance,Tags=[{Key=Name,Value=WebServer}]' --output text --query 'Instances[*].InstanceId'
i-065416ebe37dde343
ec2-user@ip-10-0-1-133 ~]$
```

```
ec2-user@ip-10-0-1-133 ~]$ aws ec2 wait instance-running --instance-ids i-065416ebe37dde343
ec2-user@ip-10-0-1-133 ~]$
```

## Tarea 1.3: crear una instancia de EC2 nueva

```
[ec2-user@ip-10-0-1-133 ~]$ aws ec2 wait instance-running --instance-ids i-065416ebe37dde343  
[ec2-user@ip-10-0-1-133 ~]$ aws ec2 describe-instances --instance-id i-065416ebe37dde343 --query 'Reservations[0].Instances[0].PublicIpAddress'  
'ec2-52-11-32-21.us-west-2.compute.amazonaws.com'  
[ec2-user@ip-10-0-1-133 ~]$
```

obtuvimos el nombre del dominio público y agregándole HTTP:// al inicio y /index.php al final, ingresamos al enlace correspondiente.



## Tarea 1.4: crear una AMI personalizada

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

Para crear una AMI nueva con base en esta instancia, reemplace NEW-INSTANCE-ID en el comando `aws ec2 create-image` por el valor que copió anteriormente y ejecute el comando ajustado:

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

```
ec2-user@ip-10-0-1-133 ~]$ aws ec2 create-image --name WebServerAMI --instance-id i-065416ebe37dde343  
{"ImageId": "ami-061683703e309da69"}
```



# Tarea 2: crear un entorno de Auto Scaling

## Tarea 2.1: crear un equilibrador de carga de aplicación

En el panel de navegación izquierdo de la Consola de administración de Elastic Compute Cloud, localice la sección **Balance de carga** y elija **Balanceadores de carga**.

Elija **Crear balanceador de carga**.

En la sección **Tipos de equilibradores de carga**, en **Balanceador de carga de aplicaciones**, elija **Crear**.

En la página **Crear balanceador de carga de aplicaciones**, en la sección **Configuración básica**, configure la siguiente opción:

- En **Nombre del balanceador de carga**, ingrese WebServerELB.

Basic configuration	
Load balancer name	Name must be unique within your AWS
	WebServerELB

Default action	Info
Forward to	webserver-app Target type: Instance, IPv4

configuramos el Target Group a continuación

# Tarea 2: crear un entorno de Auto Scaling

## Tarea 2.1: configuramos el Target Group

**Basic configuration**  
Settings in this section can't be changed after the target group is created.

Choose a target type

☒ Instances

- Supports load balancing to instances within a specific VPC.
- Facilitates the use of [Amazon EC2 Auto Scaling](#) to create and manage a fleet of Amazon EC2 instances.

Target group name

webserver-app

**Health checks**  
The associated load balancer periodically sends health checks to the instances in the target group.

Health check protocol

HTTP

Health check path

Use the default path of "/" to perform health checks on the root of the site.  
/index.php

**Security groups** [Info](#)  
A security group is a set of firewall rules that control the traffic to your load balancer.

Security groups

Select up to 5 security groups

HTTPAccess  
sg-0004d06234f949648 VPC: vpc-0ab7c0411005fd812

Create target group

**Network mapping** [Info](#)  
The load balancer routes traffic to targets in the VPC.

VPC [Info](#)  
Select the virtual private cloud (VPC) for your load balancer. The load balancer is created in the same VPC as the load balancer is created. To confirm the VPC, see the VPC console.

Lab VPC  
vpc-0ab7c0411005fd812  
IPv4 VPC CIDR: 10.0.0.0/16

Mappings [Info](#)  
Select at least two Availability Zones and one subnet. If the subnets or the VPC are not available for selection, you must create them first.

☒ us-west-2a (usw2-az2)  
Subnet  
subnet-0c98ceba86407af0f  
IPv4 address  
Assigned by AWS

☒ us-west-2b (usw2-az1)  
Subnet  
subnet-0f7654d8782b10044

# Tarea 2: crear un entorno de Auto Scaling

## Tarea 2.1: Revisamos el Load Balancer

✔ Successfully created load balancer: **WebServerELB**

It might take a few minutes for your load balancer to fully set up and route traffic. Targets will also take a few minutes to complete the registration process and pass initial health checks.

### Load balancers (1/1)

Elastic Load Balancing scales your load balancer capacity automatically in response to changes in incoming traffic.

🔍 Filter load balancers

☑	Name ▾	DNS name ▾	State ▾
☑	<a href="#">WebServerELB</a>	📋 WebServerELB-48198578...	✔ Active

# Task 2.2: Creando una plantilla de lanzamiento

## ¿Qué haremos?

Ahora crearemos una plantilla de lanzamiento para el grupo de Auto Scaling.  
Es una plantilla que utiliza un grupo de Auto Scaling para iniciar instancias de EC2. Cuando crea una plantilla de lanzamiento, se especifica la información para las instancias, tales como la AMI, tipo de instancia, par de claves, grupo de seguridad y discos.



1. En la Consola de administración de Elastic Compute Cloud del panel de navegación izquierdo, ubique la sección **Instancias** y elija **Plantillas de lanzamiento**.

### Create launch template

Creating a launch template allows you to create a saved instance configuration that can be reused, shared and launched at a later time. Templates can have multiple versions.

#### 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

### Amazon Machine Image (AMI)

WebServerAMI

ami-061683703e309da69

2024-06-25T23:24:09.000Z

Virtual Machine Image

### Instance type

t3.micro

Family: t3 2 vCPU 1 GiB Memory Current generation: true

On-Demand SUSE base pricing: 0.0104 USD per Hour

On-Demand Windows base pricing: 0.0196 USD per Hour

On-Demand RHEL base pricing: 0.0704 USD per Hour

On-Demand Linux base pricing: 0.0104 USD per Hour

Additional costs apply for AMIs with pre-installed software

### Common security groups [Info](#)

Select security groups

HTTPAccess sg-0004d06234f949648

VPC: vpc-0ab7c0411005fd812

Security groups that you add or remove here will be added to the launch template

### Key pair name

Don't include in launch template

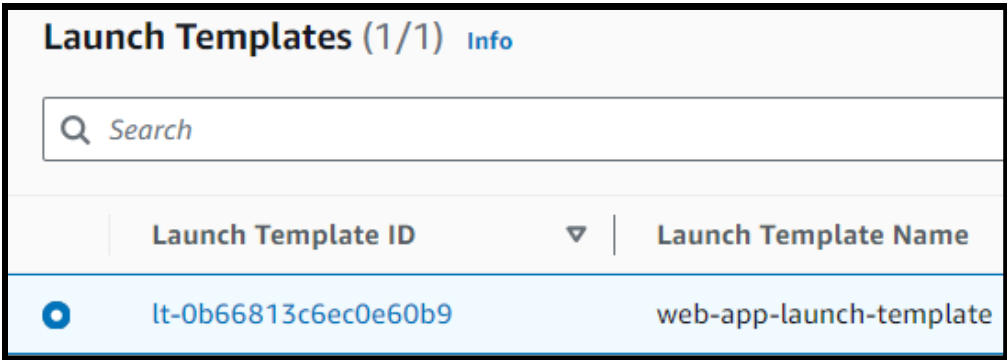
Create launch template

✓

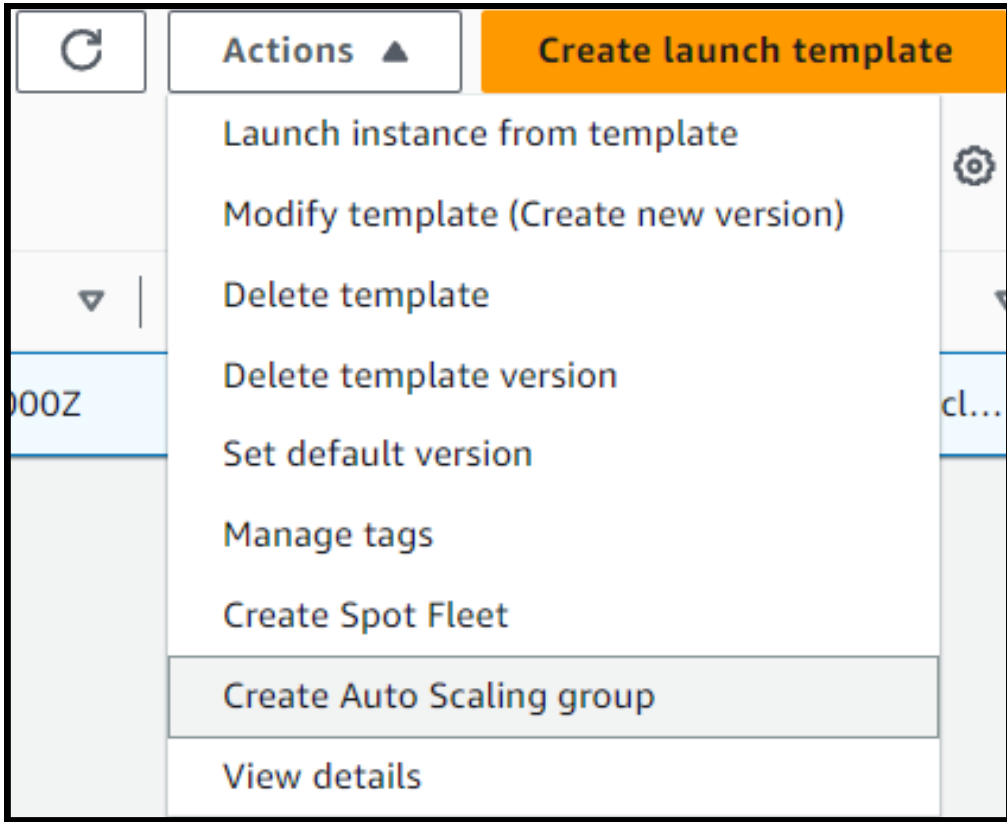
**Success**  
Successfully created web-app-launch-template(lt-0b66813c6ec0e60b9).

# Tarea 2.3: crear un grupo de Auto Scaling

Se utilizará la plantilla de lanzamiento para crear un grupo de Auto Scaling.



Crear grupo de Auto Scaling.

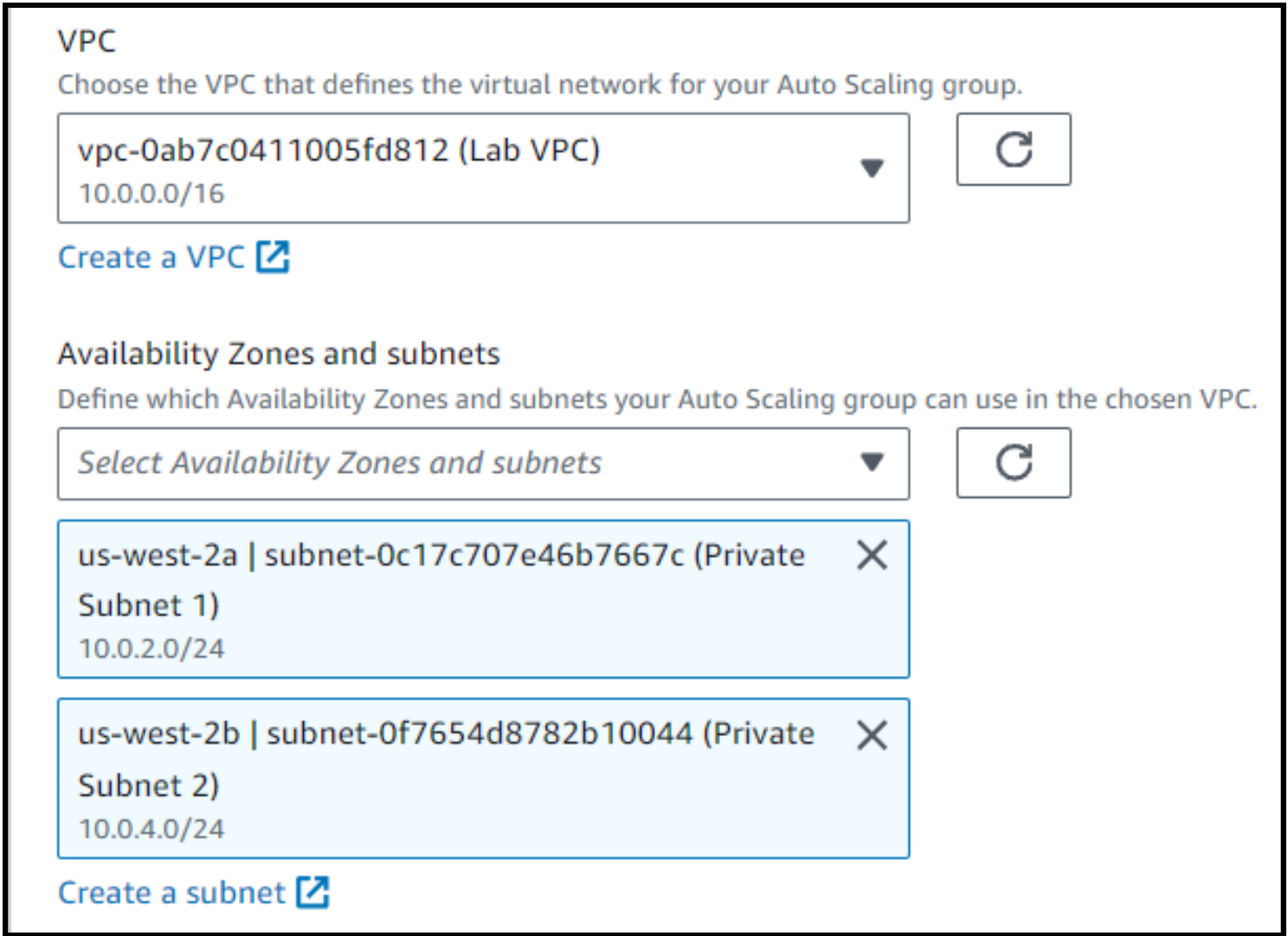


Nombre del grupo de Auto Scaling

Auto Scaling group name  
Enter a name to identify the group.

Web App Auto Scaling Group

Elegir las opciones de lanzamiento de instancias de la sección Red





# Tarea 2.3: crear un grupo de Auto Scaling

En la página Configurar las opciones avanzadas – opcional, configure las siguientes opciones:

☐ 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.

☒ Choose from your load balancer target groups  
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 shown here.

Select target groups

webservice-app | HTTP  
Application Load Balancer: WebServerELB

Configurar políticas de escalado y tamaño de grupo – opcional

Health checks

Health checks increase availability by replacing unhealthy instances. When you 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 Scaling can replace it on its next periodic check.

Desired capacity

Specify your group size.

2

Scaling limits

Set limits on how much your desired capacity can be increased or decreased

Min desired capacity

2

Equal or less than desired capacity

Max desired capacity

4

Equal or greater than desired capacity

# Tarea 2.3: crear un grupo de Auto Scaling

En la página Configurar las opciones avanzadas – opcional, configure las siguientes opciones:

☐ 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.

☒ Choose from your load balancer target groups  
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 shown here.

Select target groups

webservers-app | HTTP  
Application Load Balancer: WebServerELB

Configurar tamaño de grupo y políticas de escalado – opcional

Desired capacity  
Specify your group size.

2

☒ Target tracking scaling policy  
Choose a CloudWatch metric and target value and let the scaling policy adjust the desired capacity in proportion to the metric's value.

Scaling limits  
Set limits on how much your desired capacity can be increased or decreased

Min desired capacity  
2  
Equal or less than desired capacity

Max desired capacity  
4  
Equal or greater than desired capacity

Metric type [Info](#)  
Monitored metric that determines better scaling performance.

Average CPU utilization

Target value  
50

Create Auto Scaling group

✔ Web App Auto Scaling Group, 1 Scaling policy created successfully

Health checks  
Health checks increase availability by replacing unhealthy instances. When you 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 Scaling can replace it on its next periodic check.

# Tarea 3: verificar la configuración de Auto Scaling

En esta tarea, verificará que tanto la configuración de Auto Scaling como el balanceador de carga se encuentren en ejecución al acceder a un script preinstalado en uno de sus servidores que consumirá ciclos de CPU, lo que invocará la alarma de escalado horizontal.

En el panel de navegación izquierdo, elija **Instancias**.

Se crean dos instancias nuevas con el nombre **WebApp** como parte de su grupo de Auto Scaling. Mientras se crean estas instancias, se está Inicializando la






**Comprobación de estado** de estas dos instancias.

Observe el campo de Comprobación de estado de las instancias hasta que el estado sea 2/2 checks passed (2/2 comprobaciones aprobadas). Espere a que las dos instancias nuevas completen la inicialización antes de continuar con el siguiente paso.

Es posible que tenga que seleccionar **Actualizar** para ver el estado actualizado.

Cuando las instancias finalicen la inicialización, seleccione **Grupos de destino** en la sección **Balance de carga** del panel de navegación izquierdo y luego seleccione su grupo de destino, **webserver-app**.

En la pestaña **Destinos**, compruebe que se están creando dos instancias. Actualice esta lista hasta que el valor de **Estado** de estas instancias cambie a En buen estado.

Instances (4) <a href="#">Info</a>			
<input type="text" value="Find Instance by attribute or tag (case-sensitive)"/>			
<input type="checkbox"/>	Name 	Instance ID	Ins
<input type="checkbox"/>	Command Host	i-09ef573ebf9bb957d	
<input type="checkbox"/>	WebApp	i-0ee908bceb096ef6f	
<input type="checkbox"/>	WebApp	i-0a7bc33c95db3b714	
<input type="checkbox"/>	WebServer	i-065416ebe37dde343	

2	 2
Total targets	Healthy

# Tarea 4: Verificar la configuración del Auto Scaling

Pegue el Nombre de DNS del equilibrador de carga que copió anteriormente en la barra de direcciones.

Seleccione Start Stress.



Esto provoca que la utilización de CPU en la instancia aumente al 100 %

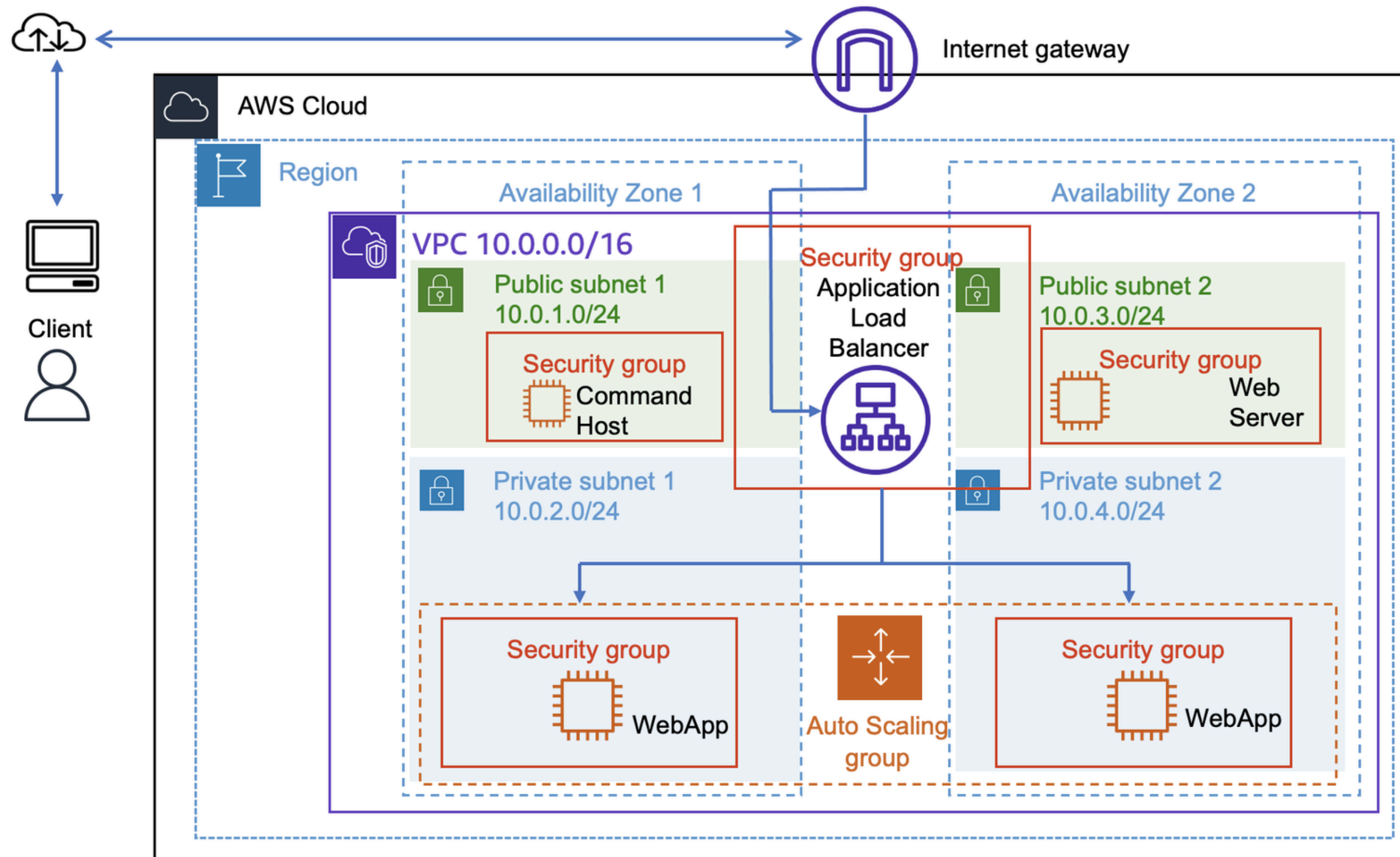
Amazon CloudWatch detectó que la utilización de CPU promedio de su grupo de Auto Scaling superó el 50% y en respuesta, se invocó su política de escalado horizontal.

Después de unos minutos, debería ver que su grupo de Auto Scaling agrega una instancia nueva.

Auto Scaling group: Web App Auto Scaling Group		
Activity history (2)		
<div>Filter activity history</div>		
Status	Description	Cause
Successful	Launching a new EC2 instance: i-0a7bc33c95db3b714	At 2024-01-01 10:00:00 from 0 to 2 desired
Successful	Launching a new EC2 instance: i-0ee908bceb096ef6f	At 2024-01-01 10:00:00 from 0 to 2 desired



# Arquitectura final del Laboratorio





**Muchas Gracias**