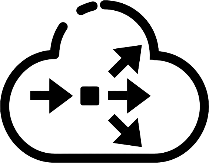
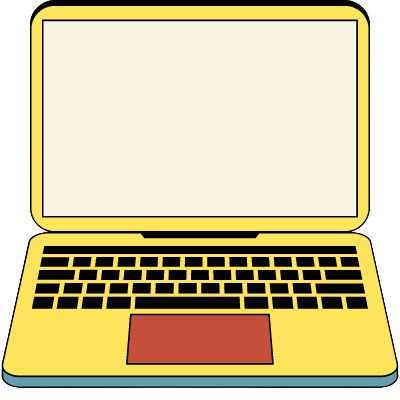
Habilitar Balanceo de Carga

**Christian Millán Soria**





- Curso: 2023-2024

- Clase: 2º DAW Tarde

- Profesor: David Hormigo Ramírez

- Módulo: Despliegue de Aplicaciones Web

- Fecha: 21/11/2023

**1. Segunda instancia** 3

**1.1. Imagen de la instancia** 3

**1.2. Creación de la instancia basada en imagen** 4

**2. Grupo de destino** 7

**3. Balanceador de carga** 9

**4. Comprobación de cada servidor** 13

**5. Atributos del grupo de destinos** 14

El hecho de contar con un único servidor de cara al cliente es una mala práctica, debido a que cualquier imprevisto puede ocasionar una paralización del servicio y unos futuros costes económicos.

Para evitar esto, se suelen utilizarlos llamados [**balanceadores de carga**🔗](https://docs.aws.amazon.com/elasticloadbalancing/latest/userguide/what-is-load-balancing.html).

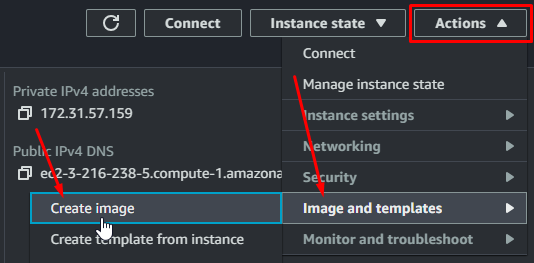
Estos hacen que la carga de las peticiones realizadas contra nuestro servidor no se realicen siempre en una única máquina, sino de distribuyen el servicio en varias máquinas que actúan de la misma manera.

**1. Segunda instancia**

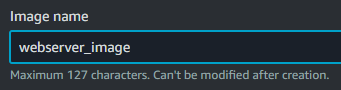
Para poder conseguir la configuración mencionada, primero debemos crear una segunda instancia que actúe como una copia de nuestro primer servidor.

**1.1. Imagen de la instancia**

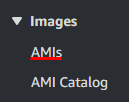
Para llevar a cabo esto, nos dirigimos a AWS, en el apartado del resumen de nuestra instancia principal, hacemos clic en el botón **Actions** y creamos una imagen a partir del estado actual de nuestra máquina.



Le damos un nombre a la imagen y la creamos.



En el panel lateral izquierdo podemos acceder a la pestaña de **AMIs** y ver, en el resumen de esta nueva imagen, el estado de la misma. Esperamos a que se termine de crear por completo.

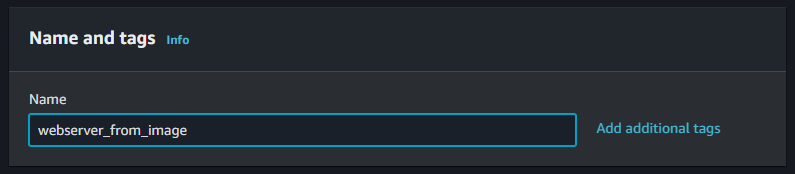


**1.2. Creación de la instancia basada en imagen**

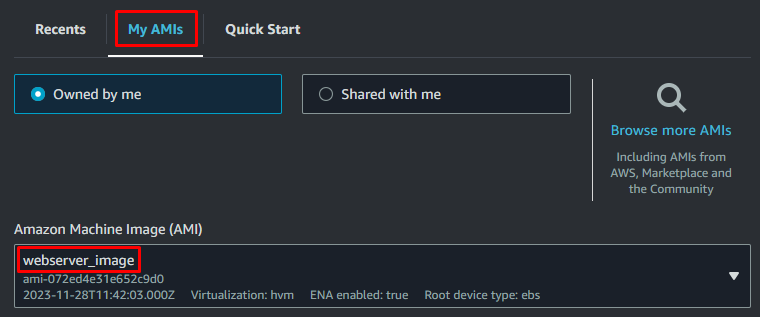
Una vez la imagen está disponible, podemos pasar a crear la nueva instancia.

Nos vamos al apartado para lanzar una nueva instancia en EC2.

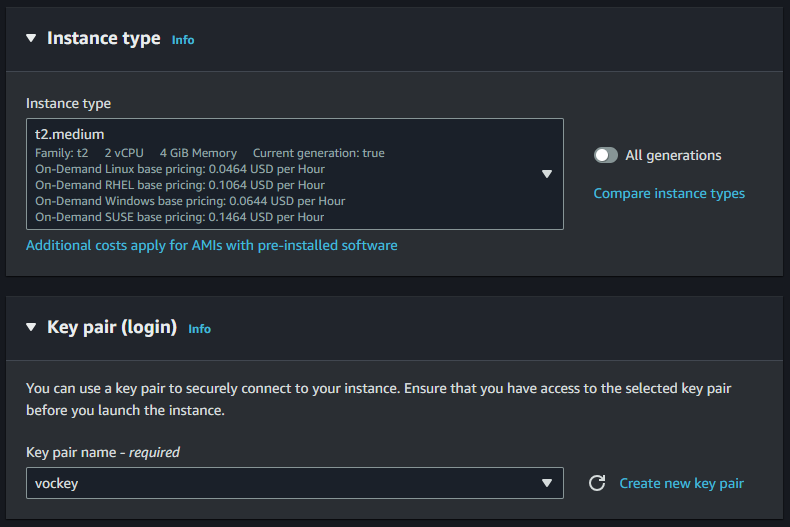
Le ponemos un nombre distinto al de la anterior.



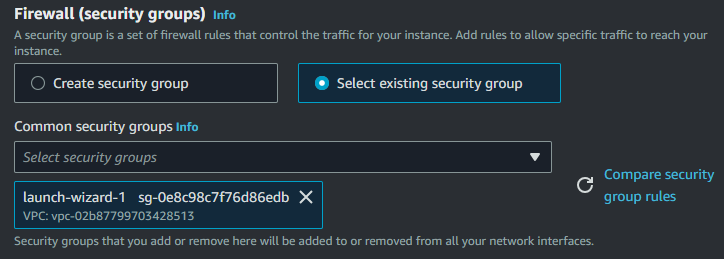
En el siguiente apartado es donde **seleccionamos la imagen que acabamos de crear**, en lugar de una imagen de un sistema operativo predeterminado de los que nos ofrece AWS.



Establecemos el tipo de instancia y su par de claves de igual forma que con la primera.

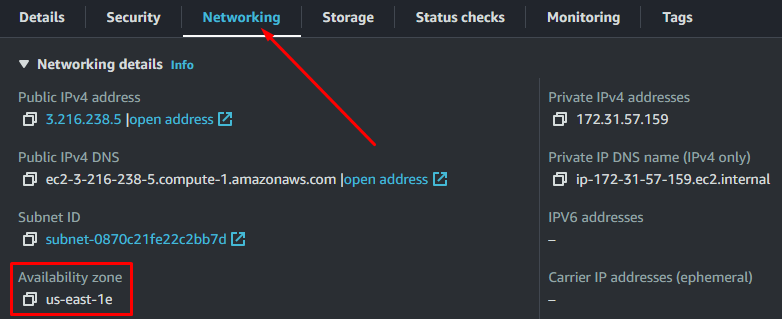


En el apartado de red, primero elegimos que el grupo de seguridad de la nueva instancia será el mismo que el de la anterior.



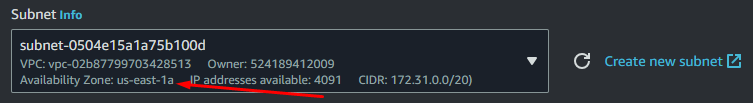
Ahora pasamos a entrar a la configuración de este apartado de una forma un poco más avanzada, haciendo clic en el botón de edición en su esquina superior derecha.

Si abrimos, en otra pestaña diferente a esta, el resumen de la primera instancia de nuestra cuenta de AWS, podemos acceder a su subapartado de redes, donde podemos comprobar la zona de disponibilidad que tiene configurada.



En mi caso, la zona de disponibilidad es la “**us-east-1e**”.

Volvemos a la pestaña donde tenemos la configuración de la nueva instancia y seleccionamos, en el subapartado de **Subnet** una zona de disponibilidad diferente a la anterior.



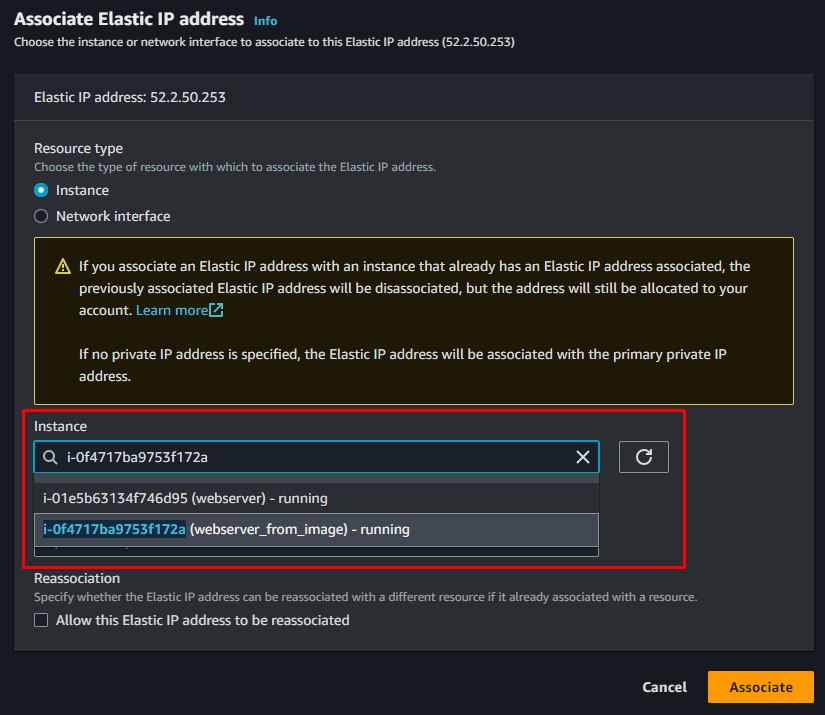
De esta forma, simulamos que un servidor se encuentra en un [**CPD🔗**](https://ucloudglobal.com/en/blog/data-processing-center/#:~:text=A%20data%20processing%20center%20or,and%20applications%20Of%20the%20same.)distinto al de la primera máquina.

Lanzamos la instancia.



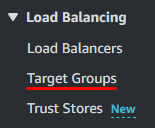
Adicionalmente, vamos a asociar una nueva IP Elástica a esta nueva instancia para que no haya problemas a la hora de conectarnos.

Nos dirigimos al apartado de las mismas en el panel izquierdo en AWS, creamos la IP y la asociamos a la nueva instancia.



**2. Grupo de destino**

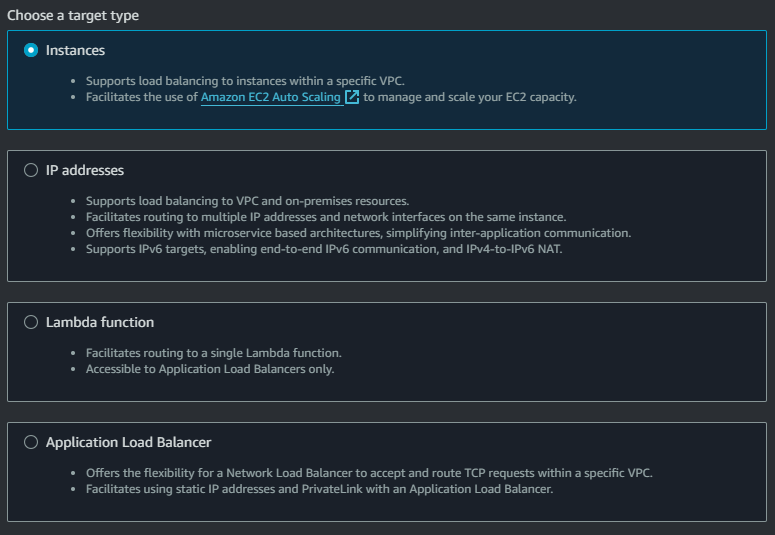
Dentro del apartado de EC2, en el panel de la izquierda de la página, accedemos al apartado de [**Grupos de destino**🔗](https://docs.aws.amazon.com/elasticloadbalancing/latest/application/load-balancer-target-groups.html).



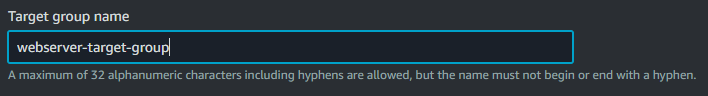
Una vez dentro, creamos un nuevo Grupo de destino.



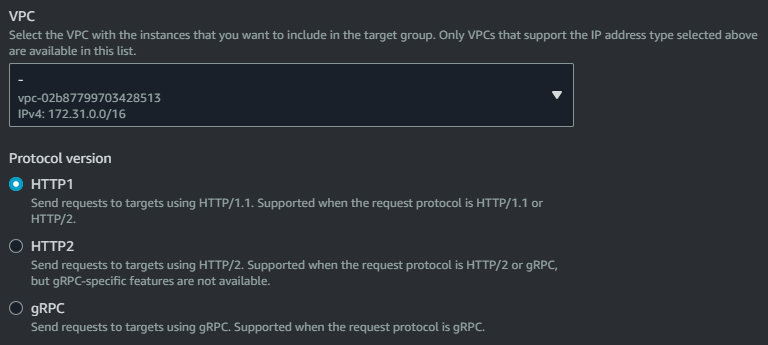
En la configuración del nuevo grupo, seleccionamos primero que el target serán unas instancias.



Establecemos un nombre para el nuevo grupo de destino.

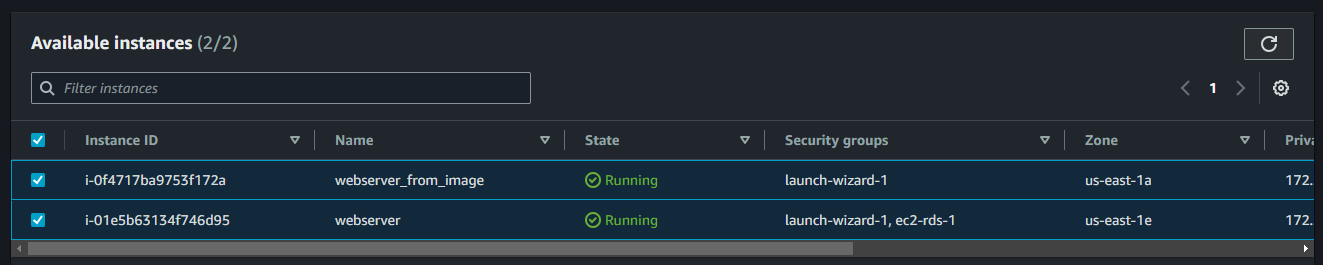


Dejamos la VPC de manera predeterminada y seleccionamos el protocolo [**HTTP1**🔗](https://www.digitalocean.com/community/tutorials/http-1-1-vs-http-2-what-s-the-difference).

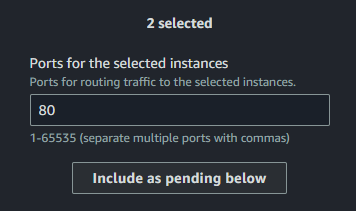


Le damos al botón de Siguiente.

Se nos muestra un menú donde seleccionaremos las dos instancias a las que irá dirigido este grupo de destino.



Dejamos el puerto predeterminado en el 80 y le damos al botón que aparece debajo del mismo.

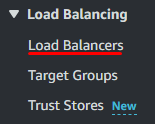


Creamos el grupo de destino.



**3. Balanceador de carga**

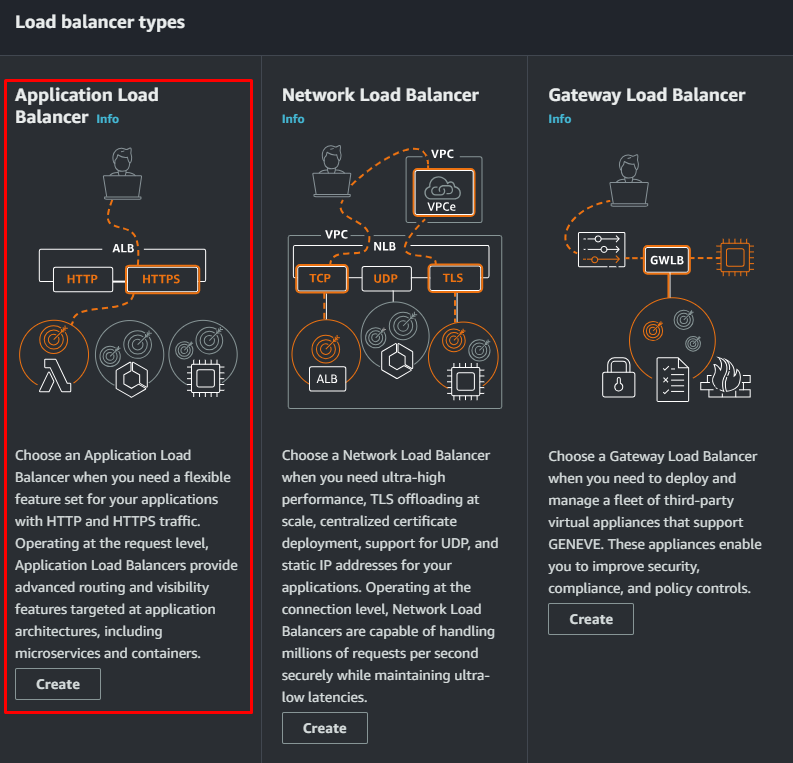
Abrimos la sección del **Balanceador de carga** en el panel de la izquierda.



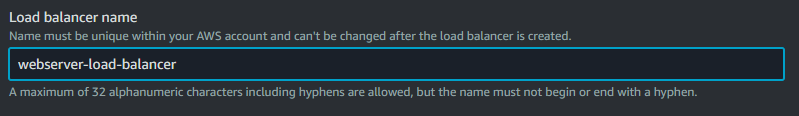
Creamos un nuevo balanceador de carga.



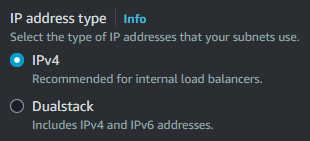
Escogemos el tipo de balanceador para **aplicación**.



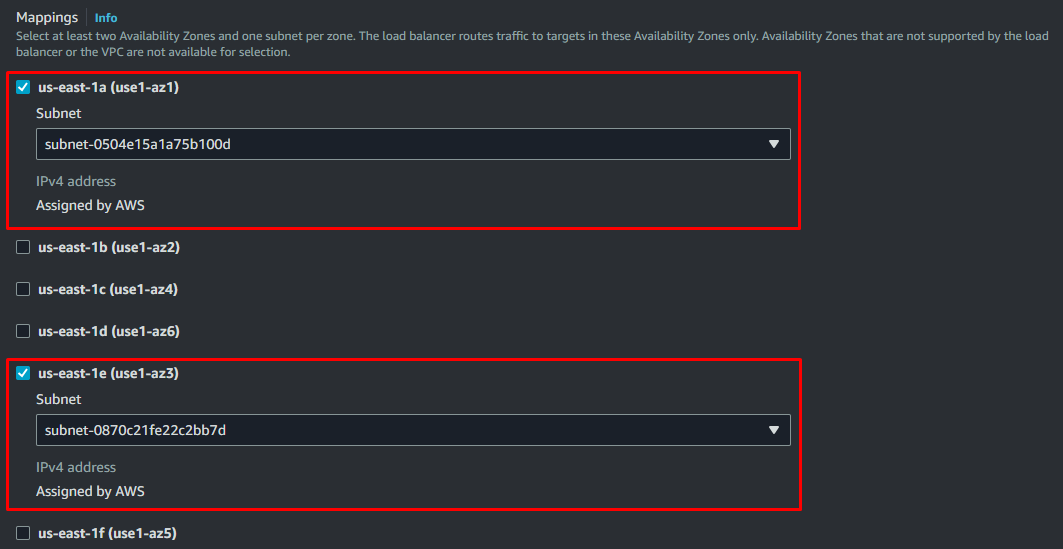
Establecemos un nuevo nombre para el mismo.



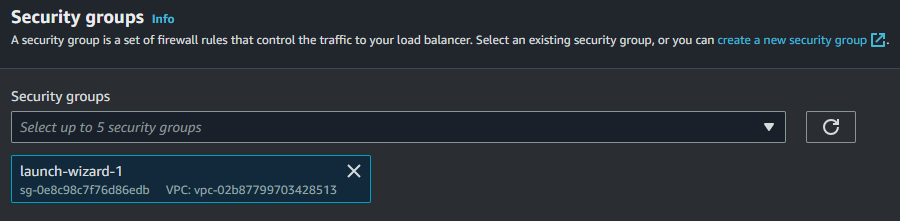
Dejamos el tipo de dirección IP en **IPv4**.



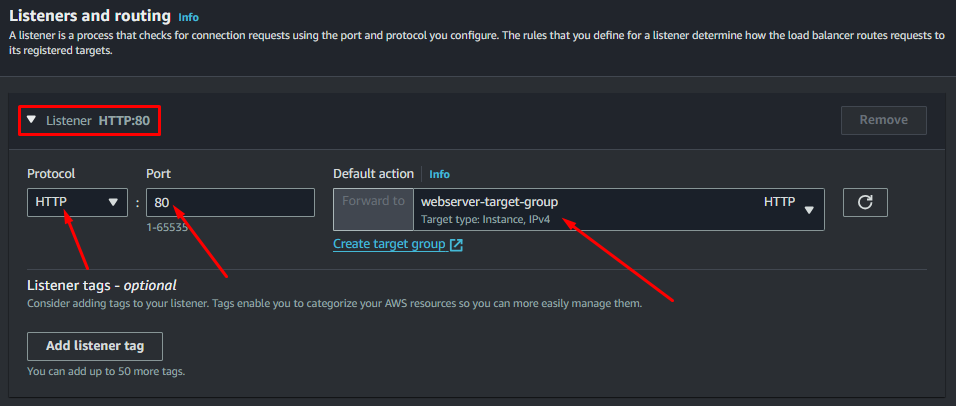
En el siguiente subapartado seleccionamos las dos subredes que establecimos para cada instancia.



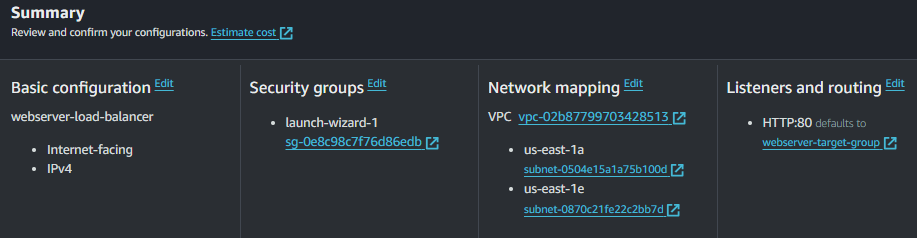
En el subapartado dedicado a los grupos de seguridad, eliminamos el grupo **default** y seleccionamos el grupo que utilizamos a la hora de crear la nueva instancia, el cual es el mismo que tiene la instancia original.



Establecemos los **listeners** para este nuevo balanceador, dejando el predeterminado configurado como HTTP.



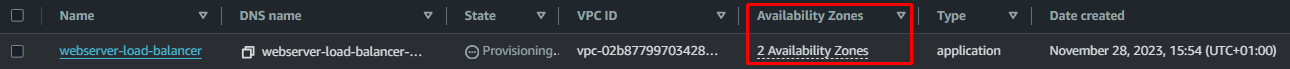
Aquí podemos ver un resumen de la configuración elegida:



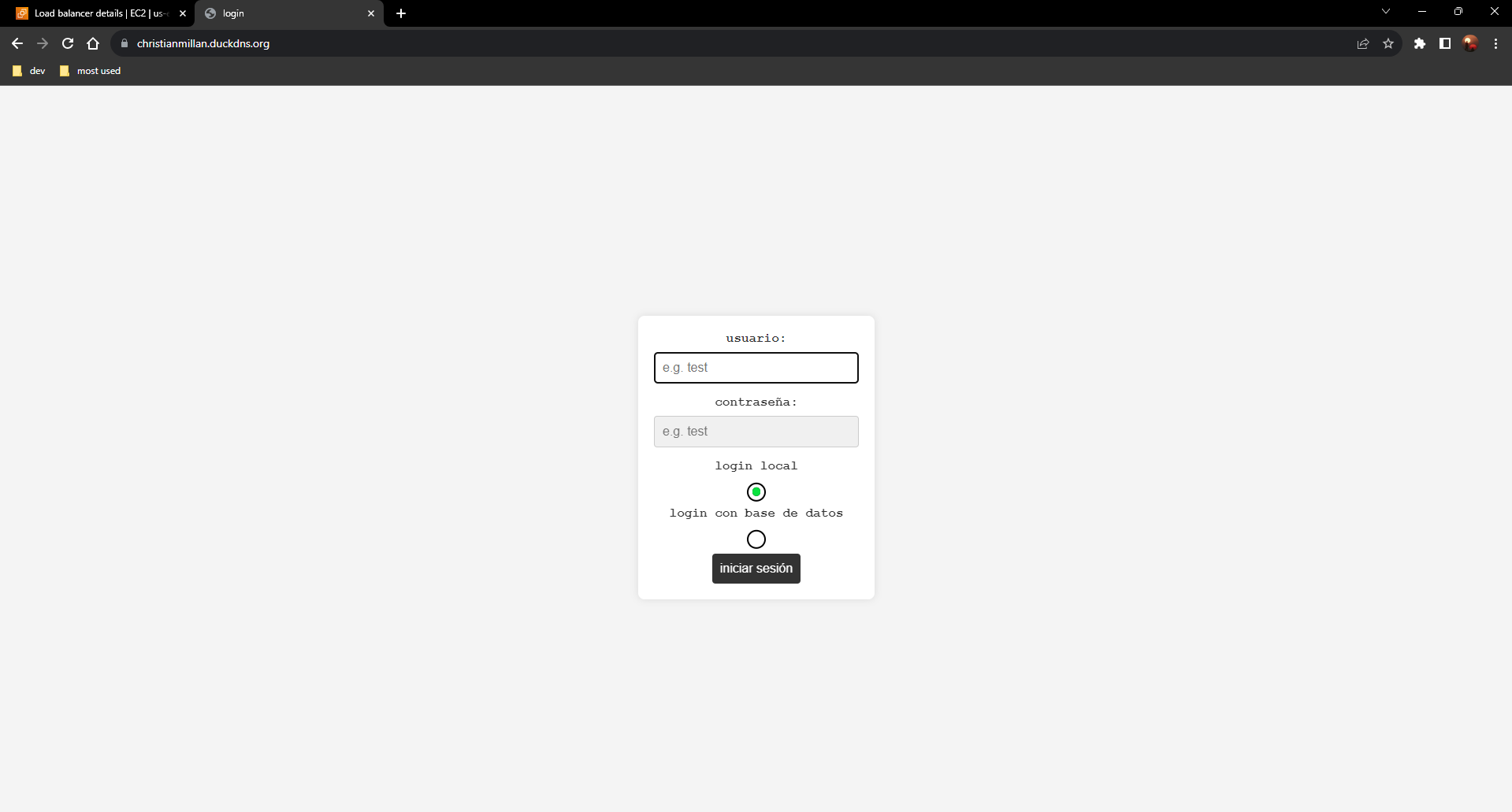
Creamos el balanceador.



Podemos ver como el nuevo balanceador cuenta con dos zonas de disponibilidad.



Esperamos a que el estado del balanceador cambia a estar disponible y cogemos el DNS público que se nos muestra en su resumen. Lo introducimos en un navegador web y podemos ver como visitamos la página de nuestro servidor.

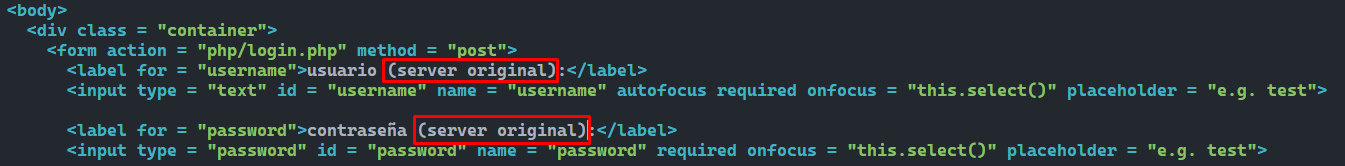


**4. Comprobación de cada servidor**

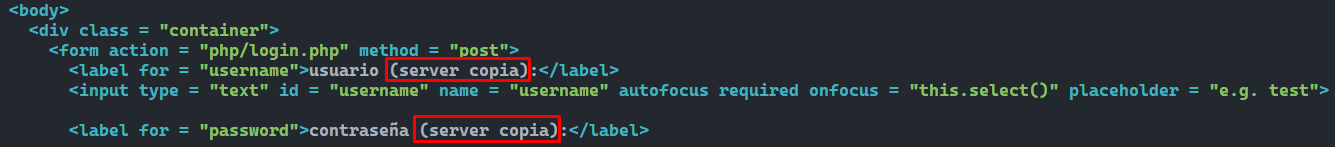
Como se puede ver en la captura anterior, no se distingue a qué servidor se ha redirigido, por lo que vamos a cambiar el código de cada servidor para diferenciarlos.

Nos conectamos por SSH a la primera instancia.

Nos dirigimos al código de la página principal del login y cambiamos el código para que se vea de qué servidor se trata.



Hacemos lo mismo en la nueva instancia.



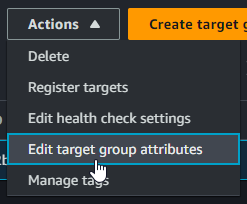
Ahora al cargar la página web podemos diferenciar los servidores.



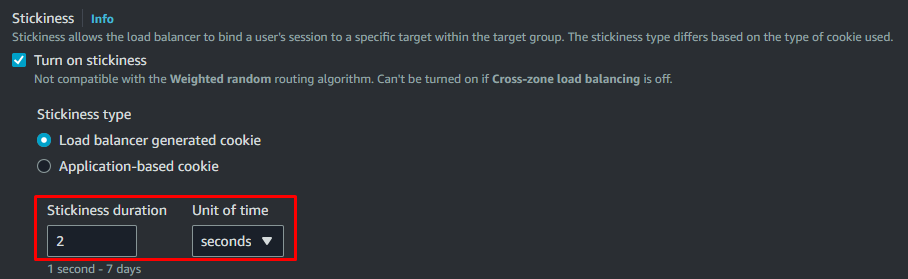
**5. Atributos del grupo de destinos**

El siguiente paso es configurar el grupo de destino para que las peticiones al servicio se lleven a cabo de forma intercalada y se divida la carga entre las dos instancias.

Para ello primero seleccionamos el grupo de destinos en el listado de los mismos y hacemos clic en el botón de **Editar atributos del grupo de destinos**.



Una vez dentro, habilitamos el [**Stickiness🔗**](https://docs.aws.amazon.com/elasticloadbalancing/latest/application/sticky-sessions.html) y establecemos un tiempo de intercambio de 2 segundos.



Guardamos los cambios.

Ahora al recargar la página, cada x tiempo la petición se cambia de un servidor a otro.

