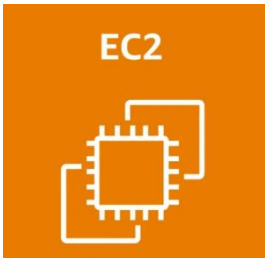
EC2



Los servicios de máquinas virtuales fueron los primeros servicios tanto de AWS como de Azure, los cuales proporcionan infraestructura como servicio (*IaaS*). Posteriormente se añadieron otros servicios como tecnología sin servidor (*serverless*), tecnología basada en contenedores y plataforma como servicio (*PaaS*).

Ya hemos comentado el coste de ejecutar servidores *in-house* (compra, mantenimiento del centro de datos, personal, etc..) además de la posibilidad de que la capacidad del servidor podría permanecer sin uso e inactiva durante gran parte del tiempo de ejecución de los servidores, lo que implica un desperdicio.

Amazon EC2

Amazon Elastic Compute Cloud (**Amazon EC2** - https://docs.aws.amazon.com/ec2/) proporciona máquinas virtuales en las que podemos alojar el mismo tipo de aplicaciones que podríamos ejecutar en un servidor en nuestras oficinas. Además, ofrece capacidad de cómputo segura y de tamaño ajustable en la nube. Las instancias EC2 admiten distintas cargas de trabajo (servidores de aplicaciones, web, de base de datos, de correo, multimedia, de archivos, etc..)

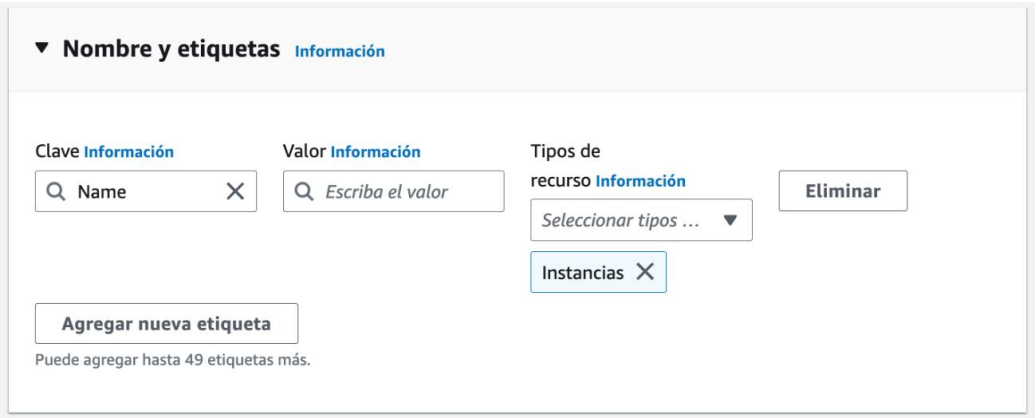
La computación elástica (*Elastic Compute*) se refiere a la capacidad para aumentar o reducir fácilmente la cantidad de servidores que ejecutan una aplicación de manera automática, así como para aumentar o reducir la capacidad de procesamiento (CPU), memoria RAM o almacenamiento de los servidores existentes.

La primera vez que lancemos una instancia de Amazon EC2, utilizaremos el asistente de lanzamiento de instancias de la consola de administración de AWS, el cual nos facilita paso a paso la configuración y creación de nuestra máquina virtual.

Antes de nada, le pondremos un nombre a la instancia de EC2. Si queremos añadir más información, podemos emplear las etiquetas a modo de metadatos.

Paso 0: Etiquetas

Las etiquetas son marcas que se asignan a los recursos de AWS. Cada etiqueta está formada por una clave y un valor opcional, siendo ambos campos *case sensitive*.



Paso 0 - Etiquetas

El etiquetado es la forma en que asocia metadatos a una instancia EC2. De esta manera podemos clasificar los recursos de AWS, como las instancias EC2, de diferentes maneras, por ejemplo, en función de la finalidad, el propietario o el entorno.

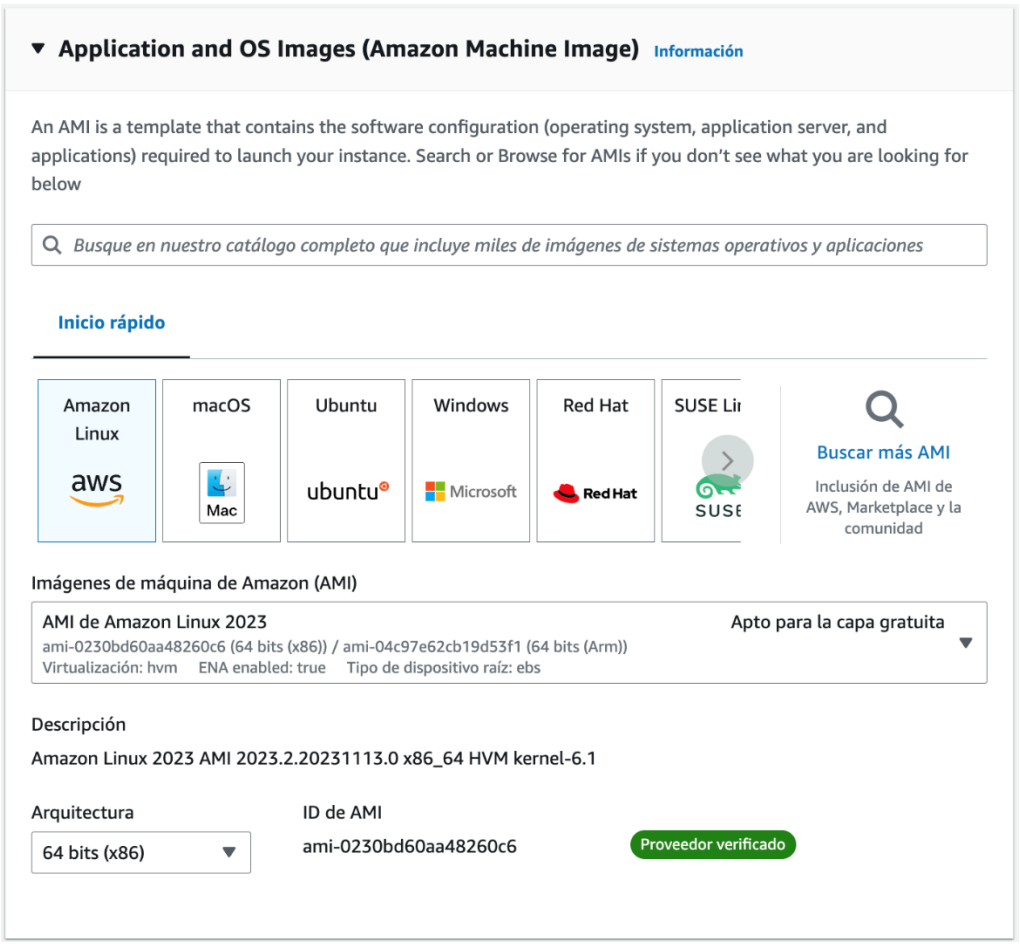
Los beneficios potenciales del etiquetado son la capacidad de filtrado, la automatización, la asignación de costes y el control de acceso.

Paso 1: AMI

Una **imagen de Amazon Machine** (AMI) proporciona la información necesaria para lanzar una instancia EC2. Podríamos decir que es una plantilla que contiene la configuración de software (sistema operativo, servidor de aplicaciones y aplicaciones instaladas) necesaria para lanzar la instancia.

Así pues, el primer paso consiste en elegir cual será la AMI de nuestra instancia. Por ejemplo, una AMI que contenga un servidor de aplicaciones, o una que contenga un servidor de base de datos.

Si vamos a montar un clúster, también podemos lanzar varias instancias a partir de una sola AMI.



Paso 1 - Seleccionar la AMI

Las AMI incluyen los siguientes componentes:

- Una plantilla para el volumen raíz de la instancia, el cual contiene un sistema operativo y todo lo que se instaló en él (aplicaciones, librerías, etc.). Amazon EC2 copia la plantilla en el volumen raíz de una instancia EC2 nueva y, a continuación, la inicia.
- Permisos de lanzamiento que controlan qué cuentas de AWS pueden usar la AMI.
- La asignación de dispositivos de bloques que especifica los volúmenes que deben asociarse a la instancia en su lanzamiento, si corresponde.

Tipos de AMI

Puede elegir entre los siguientes tipos de AMI:

- *Quick Start*: AWS ofrece una serie de AMI prediseñadas, tanto Linux como Windows, para lanzar las instancias.
- *Mis AMI*: estas son las AMI que hemos creado nosotros, ya sea a partir de máquinas locales mediante VmWare, VirtualBox, o una previa que hayamos creado en una instancia EC2, configurado y luego exportado.

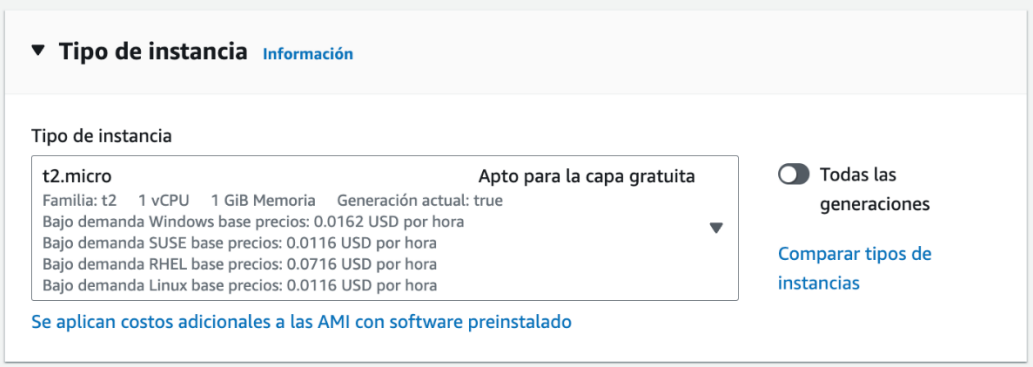
- *AWS Marketplace*: catálogo que incluye miles de soluciones de software creadas por empresas terceras (las cuales pueden cobrar por su uso). Estas AMI pueden ofrecer casos de uso específicos para que pueda ponerse en marcha rápidamente.
- *AMI de la comunidad*: estas son AMI creadas por personas de todo el mundo. AWS no controla estas AMI, así que deben utilizarse bajo la propia responsabilidad, evitando su uso en entornos corporativos o de producción.

Las AMI dependen de la región

Las AMI que creamos se hacen en la región en la que estamos conectados. Si la necesitamos en otra región, debemos realizar un proceso de copia.

Paso 2: Tipo de instancias

El segundo paso es seleccionar un tipo de instancia, según nuestro caso de uso. Los tipos de instancia incluyen diversas combinaciones de capacidad de CPU, memoria, almacenamiento y red.



Paso 2 - Eligiendo el tipo de instancia

Cada tipo de instancia se ofrece en uno o más tamaños, lo cual permite escalar los recursos en función de los requisitos de la carga de trabajo de destino.

Categorías

Las categorías de tipos de instancia incluyen instancias de uso general, optimizadas para informática, memoria, almacenamiento y de informática acelerada.

Categoría	Tipo de instancia	Caso de uso
Uso general	a1, m4, m5, m6, t2, **t3**, **t4g**	Amplio
Computación	c5, c6, **c7**	Alto rendimiento
Memoria	r5, r6, **r7**, x1, x2, z1	*Big Data*
Informática acelerada	f1, g3, g4, g5, p3, p4, p5	*Machine Learning*
Almacenamiento	d2, d3, h1, i3, i4	Sistemas de archivos distribuidos

Por ejemplo, instancias del tipo **r**, las cuales optimizan la memoria, tendrán más memoria que CPU comparada, por ejemplo, con la familia de propósito general del tipo m.

Instancia graviton

AWS ofrece máquinas basadas en ARM que ofrecen un mayor rendimiento con un menor consumo energético, y por tanto, un menor coste por instancia. Si elegimos una instancia de tipo **g**, como pueda ser una *t4g* o *m6g*, nuestras aplicaciones y las librerías dependientes deben soportar la arquitectura ARM64.

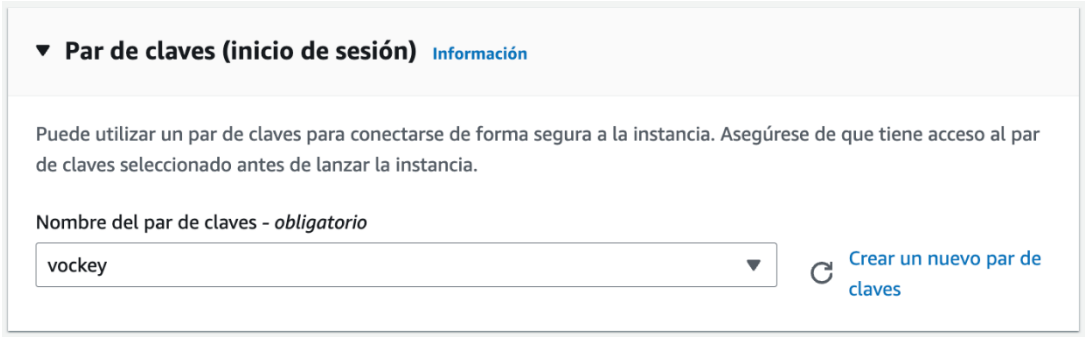
Tipos de instancias

Los tipos de instancias (https://aws.amazon.com/es/ec2/instance-types/) ofrecen familias, generaciones y tamaños. Así pues, el tipo de instancia t3.large referencia a la familia T, de la tercera generación y con un tamaño **large.**

En general, los tipos de instancia que son de una generación superior son más potentes y ofrecen una mejor relación calidad/precio.

Paso 3: Par de claves

Para poner conectarnos a la instancia vía SSH y poder configurarla, necesitaremos hacerlo a través de un par de claves SSH. Así pues, el siguiente paso es elegir un **par de claves** existente (formato X.509), continuar sin un par de claves o crear un par de claves nuevo antes de crear y lanzar la instancia EC2.



Paso 3 - Eligiendo el par de claves

Claves en AWS Academy

Nuestro usuario tiene creado por defecto un par de claves que se conocen como **vockey**. Esta claves se pueden descargar desde la opción *AWS Details* del laboratorio de *Learner Lab*. Más adelante veremos cómo utilizarlas.

Amazon EC2 utiliza la criptografía de clave pública para cifrar y descifrar la información de inicio de sesión. La clave pública la almacena AWS, mientras que la clave privada la almacenamos nosotros.

Guarda tus claves

Si creamos una par de claves nuevas, hemos de descargarlas y guardarlas en un lugar seguro. Esta es la única oportunidad de guardar el archivo de clave privada. Si perdemos las claves, tendremos que destruir la instancia y volver a crearla.

Para conectarnos a la instancia desde nuestra máquina local, necesitamos hacerlo via un cliente SSH / Putty adjuntando el par de claves descargado. Si la AMI es de Windows, utilizaremos la clave privada para obtener la contraseña de administrador que necesita para iniciar sesión en la instancia. En cambio, si la AMI es de Linux, lo haremos mediante ssh:

```
ssh -i /path/miParClaves.pem miNombreUsuarioInstancia@miPublicDNSInstancia
```

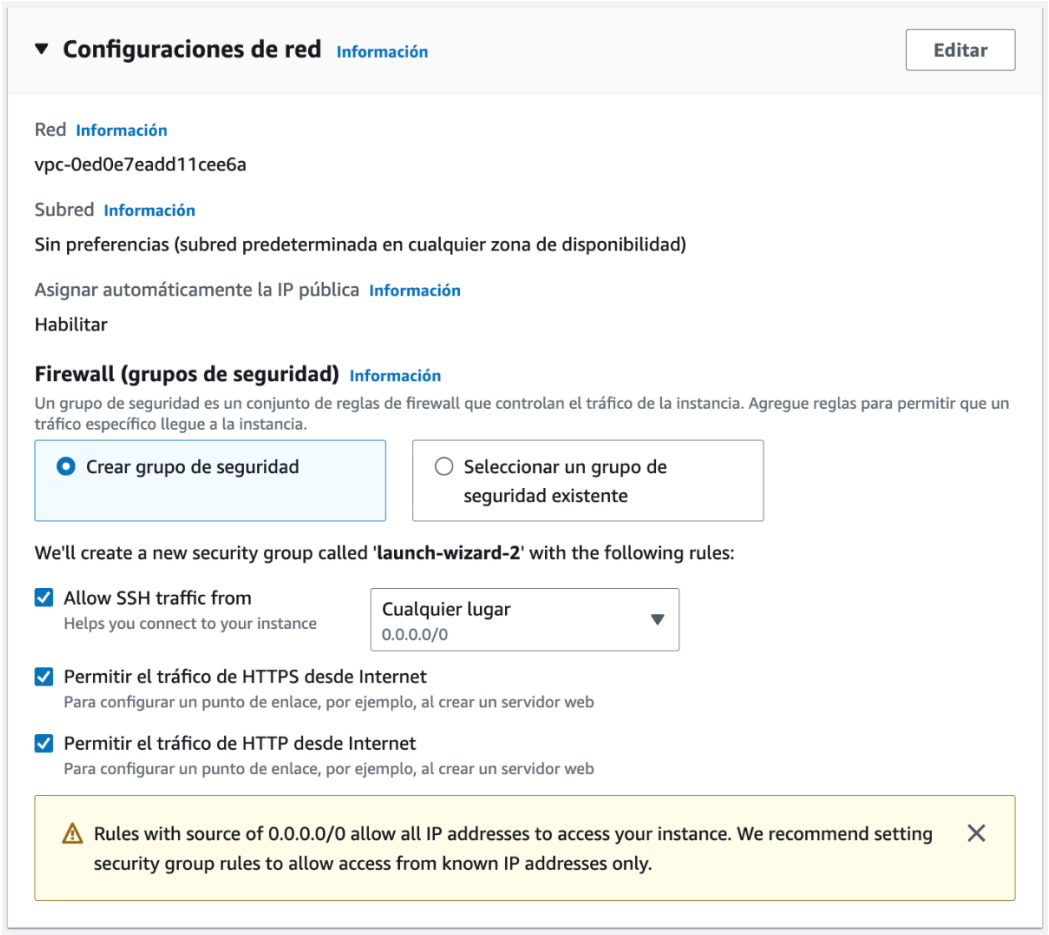
Por ejemplo, si utilizamos la *Amazon Linux AMI* y descargamos las claves de *AWS Academy* (suponiendo que la IP pública de la máquina que hemos creado es 3.83.80.52) nos conectaríamos mediante:

```
ssh -i labsuser.pem ec2-user@3.83.80.52
```

Más información
en: https://docs.aws.amazon.com/es_es/AWSEC2/latest/UserGuide/AccessingInstances.html

Paso 4: Configuración de la red

El siguiente paso es especificar la ubicación de red en la que se implementará la instancia EC2, teniendo en cuenta la región donde nos encontramos antes de lanzar la instancia. En este paso, elegiremos la **VPC** y la **subred** dentro de la misma, ya sea de las que tenemos creadas o pudiendo crear los recursos en este paso.



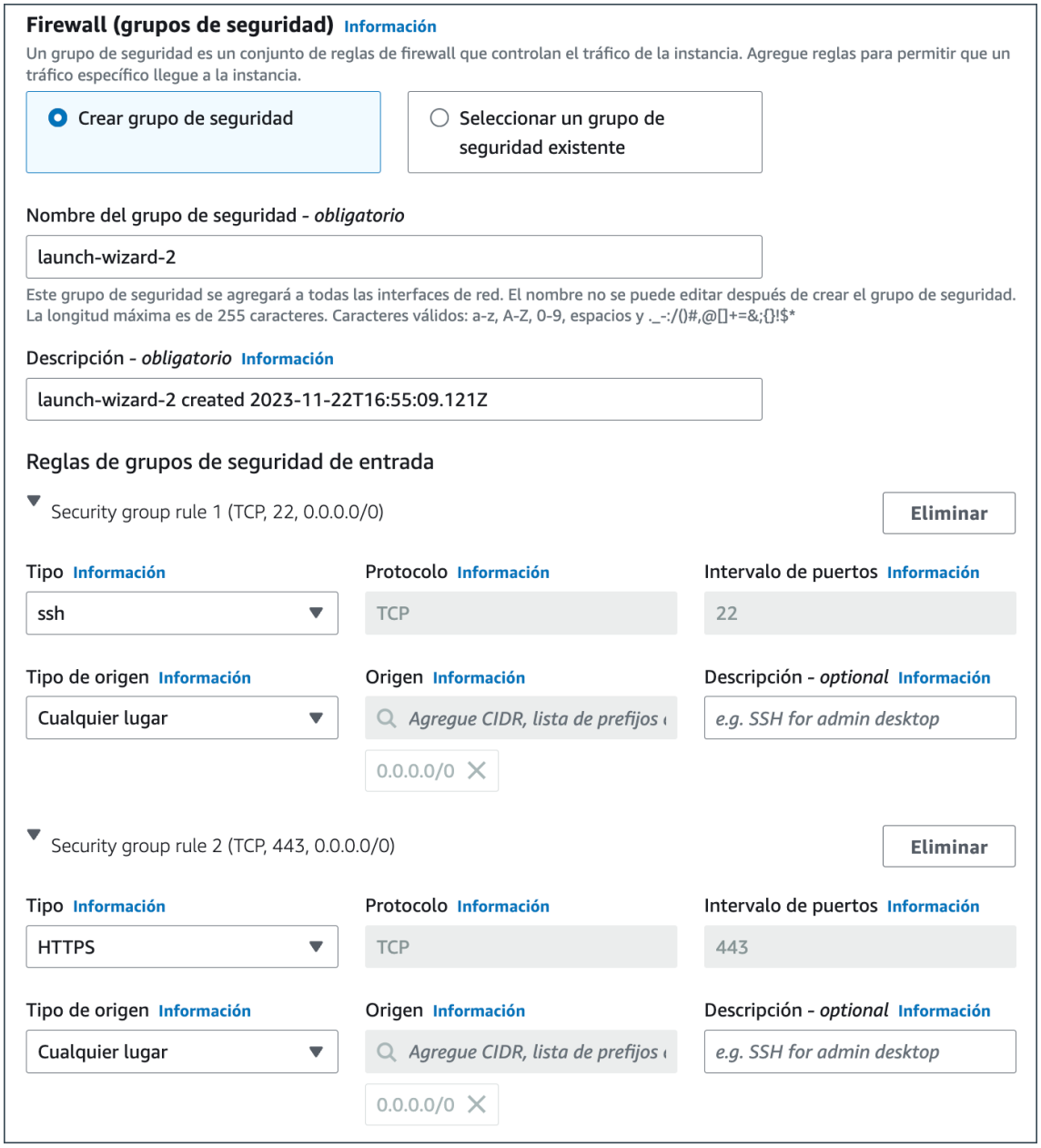
Paso 4 - Configurando la red

Respecto a la asignación pública de IP sobre esta instancia, cuando se lanza una instancia en una VPC predeterminada, AWS le asigna una dirección IP pública de forma predeterminada. En caso contrario, si la VPC no es la predeterminada, AWS no asignará una dirección IP pública, a no ser que lo indiquemos de forma explícita.

Grupo de seguridad

Un grupo de seguridad es un conjunto de reglas de firewall que controlan el tráfico de red de una o más instancias, por lo que se encuentra fuera del sistema operativo de la instancia, formando parte de la VPC.

Si editamos la red (botón cuadrado en la esquina superior derecha), podemos seleccionar unos de los grupos de seguridad existente, o crear uno nuevo:



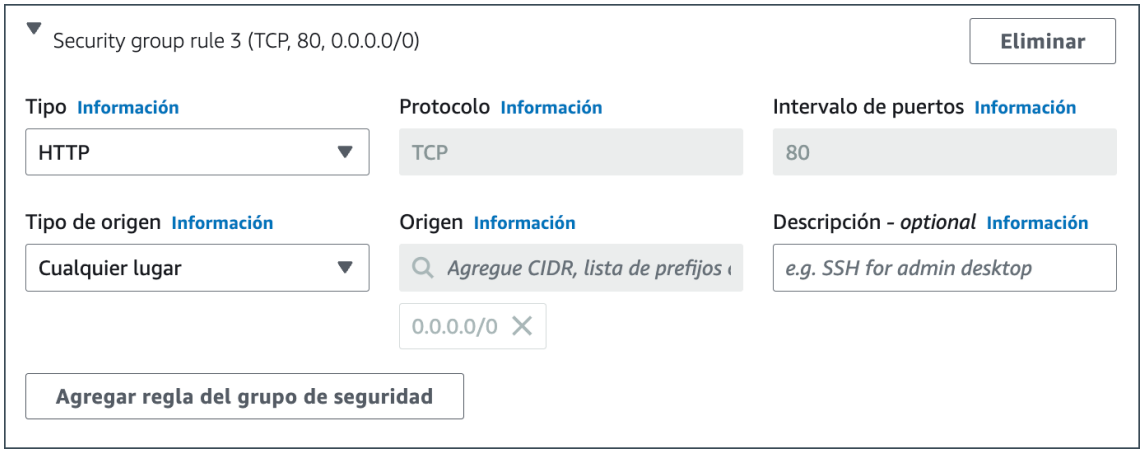
Grupo de seguridad

Dentro del grupo, agregaremos reglas para habilitar el tráfico hacia o desde nuestras instancias asociadas. Para cada una de estas reglas especificaremos el puerto, el protocolo (TCP, UDP, ICMP), así como el origen (por ejemplo, una dirección IP u otro grupo de seguridad) que tiene permiso para utilizar la regla.

De forma predeterminada, se incluye una regla de salida que permite todo el tráfico saliente. Es posible quitar esta regla y agregar reglas de salida que solo permitan tráfico saliente específico.

Servidor Web

Para un ejemplo que vamos a realizar en siguientes pasos, debemos habilitar las peticiones entrantes en el puerto 80. Para ello crearemos una regla que permita el tráfico HTTP.



Regla HTTP

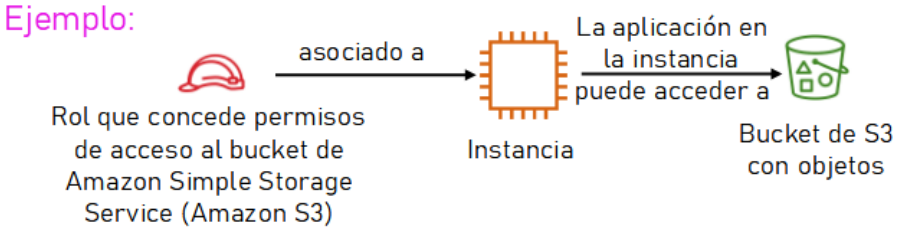
AWS evalúa las reglas de todos los grupos de seguridad asociados a una instancia para decidir si permite que el tráfico llegue a ella. Si desea lanzar una instancia en una nube virtual privada (VPC), debe crear un grupo de seguridad nuevo o utilizar uno que ya exista en esa VPC.

Las reglas de un grupo de seguridad se pueden modificar en cualquier momento, y las reglas nuevas se aplicarán automáticamente a todas las instancias que estén asociadas al grupo de seguridad.

Asociar un rol de IAM

Si necesitamos que nuestras instancias EC2 ejecuten una aplicación que debe realizar llamadas seguras de la API a otros servicios de AWS, en vez de dejar anotadas las credenciales en el código de la aplicación (esto es una muy mala práctica que puede acarrear problemas de seguridad), debemos asociar un rol de IAM a una instancia EC2.

El rol de IAM asociado a una instancia EC2 se almacena en un **perfil de instancia**. Si creamos el rol desde esta misma pantalla, AWS creará un perfil de instancia automáticamente y le otorgará el mismo nombre que al rol. En el desplegable la lista que se muestra es, en realidad, una lista de nombres de perfiles de instancia.



Paso 4 - Rol IAM

Cuando definimos un rol que una instancia EC2 puede utilizar, estamos configurando qué cuentas o servicios de AWS pueden asumir dicho rol, así como qué acciones y recursos de la API puede utilizar la aplicación después de asumir el rol. Si cambia un rol, el cambio se extiende a todas las instancias que tengan el rol asociado.

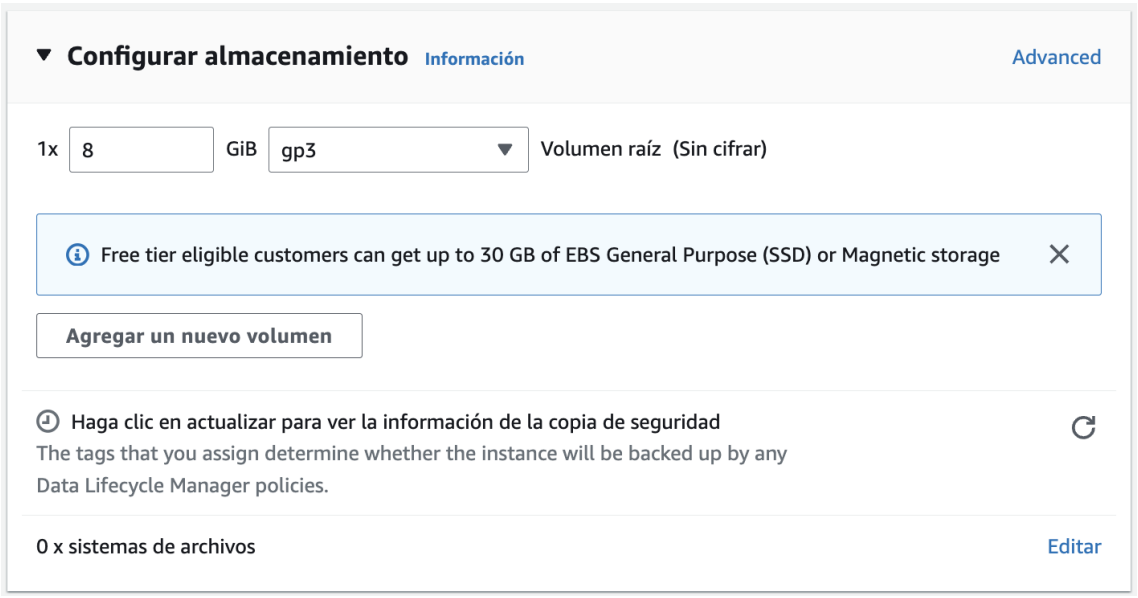
La asociación del rol no está limitada al momento del lanzamiento de la instancia, también se puede asociar un rol a una instancia que ya exista.

Paso 5: Almacenamiento

Al lanzar la instancia EC2 configuraremos las opciones de almacenamiento. Por ejemplo el tamaño del volumen raíz en el que está instalado el sistema operativo invitado o volúmenes de almacenamiento adicionales cuando lance la instancia.

Algunas AMI están configuradas para lanzar más de un volumen de almacenamiento de forma predeterminada y, de esa manera, proporcionar almacenamiento independiente del volumen raíz. Para

cada volumen que tenga la instancia, podemos indicar el tamaño de los discos, los tipos de volumen, si el almacenamiento se conservará en el caso de terminación de la instancia y si se debe utilizar el cifrado.



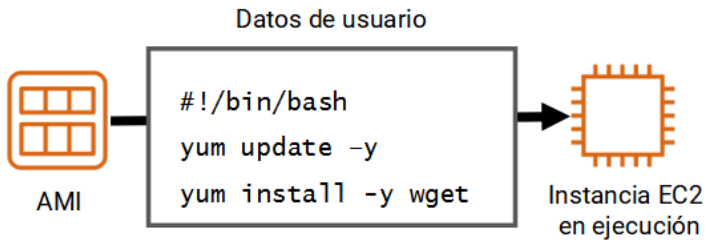
Paso 5 - Almacenamiento

Paso 6: Detalles avanzados / Script de datos de usuario

Al momento de crear las instancias EC2, de forma opcional, podemos especificar un script de datos de usuario durante el lanzamiento de la instancia. Los datos de usuario pueden automatizar la finalización de las instalaciones y las configuraciones durante el lanzamiento de la instancia. Por ejemplo, un script de datos de usuario podría colocar parches en el sistema operativo de la instancia y actualizarlo, recuperar e instalar claves de licencia de software, o instalar sistemas de software adicionales.

Por ejemplo, si queremos instalar un servidor de Apache, de manera que arranque automáticamente y que muestre un *Hola Mundo* podríamos poner

```
#!/bin/bash
yum update -y
yum -y install httpd
systemctl enable httpd
systemctl start httpd
echo '<html><h1>Hola Mundo desde el Severo!</h1></html>' > /var/www/html/index.html
```



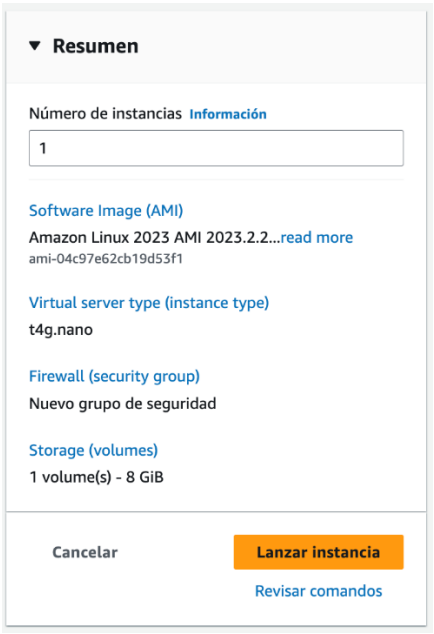
Script en Windows

Si nuestra instancia es de Windows, el script de datos de usuario debe escribirse en un formato que sea compatible con una ventana del símbolo del sistema (comandos por lotes) o con *Windows PowerShell*.

De forma predeterminada, los datos de usuario sólo se ejecutan la primera vez que se inicia la instancia.

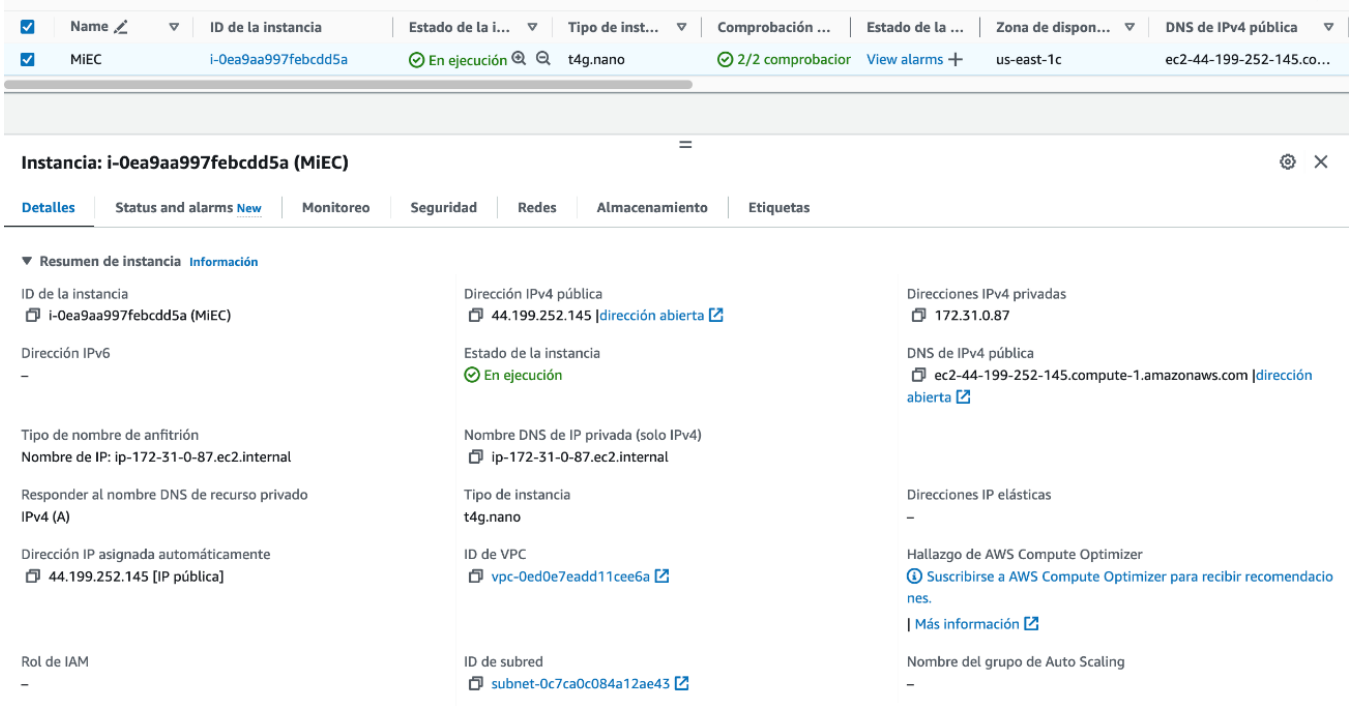
Paso 7: Resumen

El paso final es una página resumen con todos los datos introducidos. A la derecha se nos muestra todas las características seleccionadas, pudiendo indicar cuantas instancias del mismo tipo queremos crear.



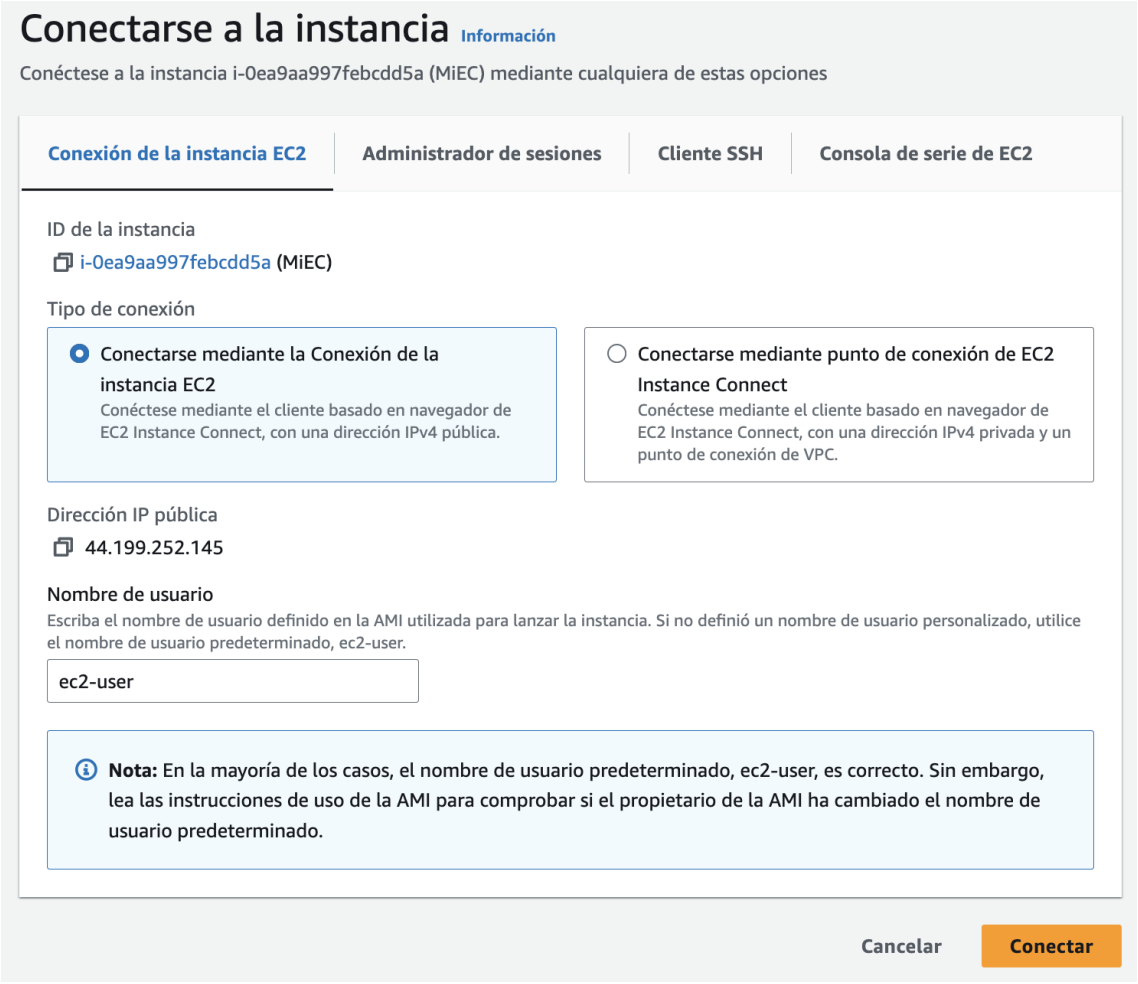
Paso 7 - Resumen

Por último, una vez lanzada la instancia, podemos observar la información disponible sobre la misma: dirección IP y la dirección DNS, el tipo de instancia, el ID de instancia único asignado a la instancia, el ID de la AMI que utilizó para lanzar la instancia, el ID de la VPC, el ID de la subred, etc...



Características de la instancia creada

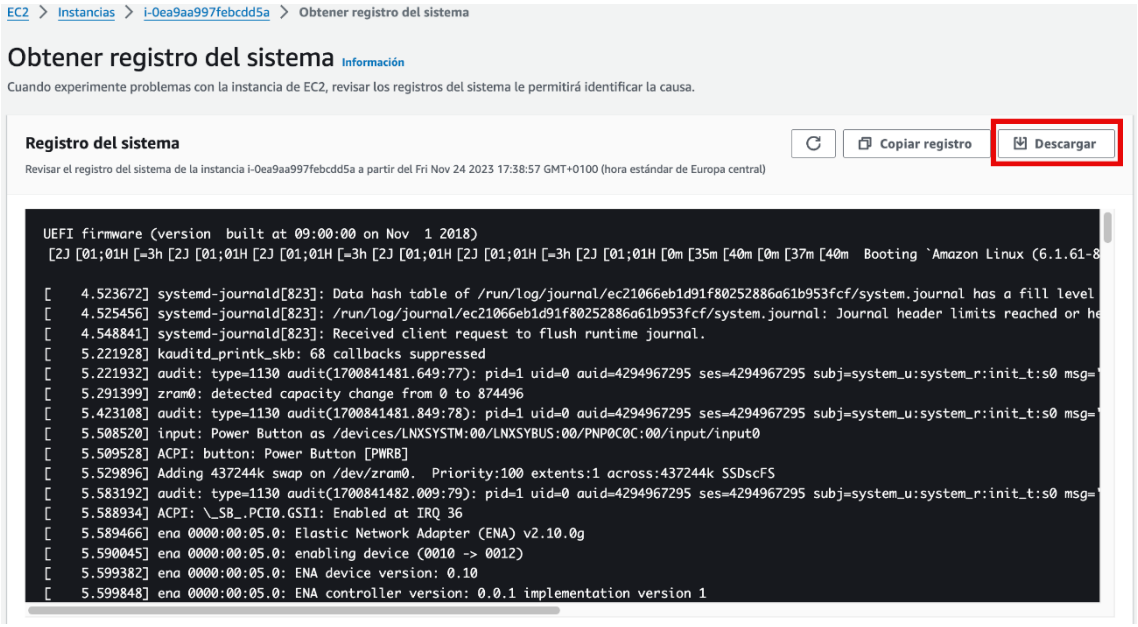
Aunque más adelante veremos cómo conectarnos mediante SSH, desde la propia consola de AWS podemos conectarnos a las instancias. Para ello, seleccionamos la instancia, y en la parte superior pulsamos sobre el botón *Conectar*, y utilizaremos una conexión a la instancia EC2:



Conexión a la instancia EC2

¿Y si algo ha fallado en el arranque?

Desde la opción *Acciones → Monitoreo y solución de problemas → Obtener registro del sistema* podemos visualizar qué aparecería si tuviéramos un monitor conectado a la instancia y comprobar y descargar el fichero de *log* de la máquina arrancada, para ver qué ha sucedido.



Log de la instancia EC2

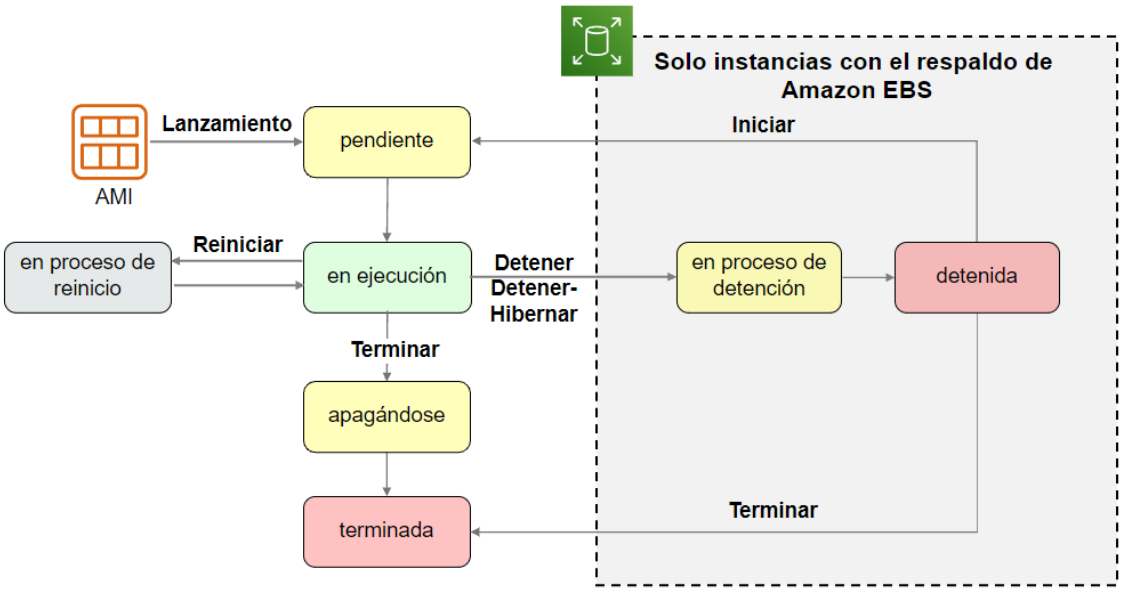
En resumen, las instancias EC2 se lanzan desde una plantilla de AMI en una VPC de nuestra cuenta. Podemos elegir entre muchos tipos de instancias, con diferentes combinaciones de CPU, RAM,

almacenamiento y redes. Además, podemos configurar grupos de seguridad para controlar el acceso a las instancias (especificar el origen y los puertos permitidos). Al crear una instancia, mediante los datos de usuario, podemos especificar un script que se ejecutará la primera vez que se lance una instancia.

Ciclo de vida de las instancias

Las instancias en todo momento tienen un estado que se puede consultar:

- *Pending (pendiente)*: nada más lanzarse o al arrancar una instancia detenida.
- *Running (en ejecución)*: cuando arrancó la instancia por completo y está lista para su uso. En este momento se empieza a facturar.
- *Rebooting (reiniciada)*: AWS recomienda reiniciar las instancias con la consola de Amazon EC2, la CLI de AWS o los SDK de AWS, en lugar de utilizar el reinicio desde el sistema operativo invitado. Una instancia reiniciada permanece en el mismo host físico, mantiene el mismo DNS público y la misma IP pública y, si tiene volúmenes del almacén de instancias, conserva los datos en ellos.
- *Shutting down (en proceso de terminación / apagándose)*
- ***Terminated*** *(terminada)*: las instancias terminadas permanecen visibles en la consola de Amazon EC2 durante un tiempo antes de que se destruya la máquina virtual. Sin embargo, no es posible conectarse a una instancia terminada ni recuperarla.
- *Stopping (deteniéndose)*: las instancias que cuentan con volúmenes EBS se pueden detener.
- *Stopped (detenida)*: no generará los mismos costes que una instancia en el estado *running*. Sólo se paga por el almacenamiento de datos. Solo se pueden detener las instancias que utilizan como almacenamiento EBS.



Ciclo de vida de una instancia

IPs estáticas

A cada instancia que recibe una IP pública se le asigna también un DNS externo. Por ejemplo, si la dirección IP pública asignada a la instancia es 203.0.113.25, el nombre de host DNS externo podría ser *ec2-203-0-113-25.compute-1.amazonaws.com*.
AWS libera la dirección IP pública de la instancia cuando la instancia se detiene o se termina. *La instancia detenida recibe una dirección IP pública nueva cuando se reinicia*.
Si necesitamos una IP pública fija, se recomienda utilizar una IP elástica, asociándola primero a la región donde vaya a residir la instancia EC2. Recuerda que las IP elásticas se pagan por cada hora que las tenemos reservadas y se deja de pagar por ellas si están asociadas a una instancia en ejecución.