

# Cloud Computing - AWS



Presentado por **Alejandro Chacón**

[www.consultec-ti.com](http://www.consultec-ti.com)

# Agenda

- Intro
- Almacenamiento S3
- Networking
- Cómputo y Almacenamiento 
- Integración y Monitoreo
- Bases de Datos
- Seguridad
- Despliegues y Gestión de Infraestructura

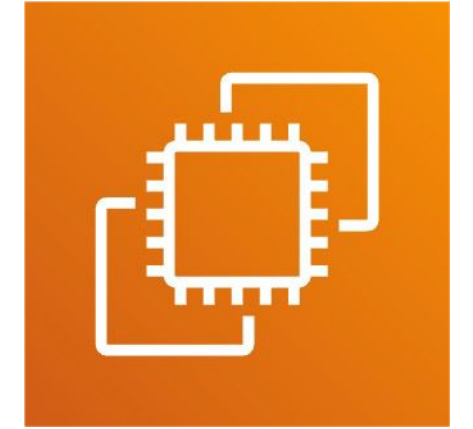
# AWS Compute

## Cómputo en la Nube

- Intro
- EC2 – User Data / nomenclatura
- Tipos de Instancias
- Grupos de Seguridad
- Opciones de compra
- Almacenamiento (EBS, EFS, Instance Store, AMI)
- Escalabilidad y Alta disponibilidad
  - Grupos de Escalamiento
  - Balanceadores de Carga

# Cómputo en AWS

## Introducción



- EC2 es uno de los más populares de la oferta de AWS
- EC2 = Elastic Compute Cloud = Infraestructura como servicio
- Consiste principalmente en la capacidad de:
  - Alquiler de máquinas virtuales (EC2)
  - Almacenamiento de datos en unidades virtuales (EBS)
  - Distribución de carga entre máquinas (ELB)
  - Escalar los servicios usando un grupo de escalado automático (ASG)
- Conocer EC2 es fundamental para entender cómo funciona la Nube

# Cómputo en AWS

## Opciones Generales

- Sistema operativo (SO): Linux, Windows o Mac OS
- Cuánta potencia de Cómputo y núcleos (CPU)
- Cuánta memoria de acceso aleatorio (RAM)
- Cuánto espacio de almacenamiento:
  - Conectado a la red (EBS y EFS)
  - hardware (EC2 Instance Store)
- Tarjeta de red: velocidad de la tarjeta, dirección IP pública
- Reglas de firewall: grupo de seguridad
- Script Bootstrap (configurar en el primer lanzamiento):  
datos de usuario de EC2 (user-data)



# Cómputo en AWS

## EC2 – User Data

- Es posible arrancar nuestras instancias utilizando un script de datos de usuario de EC2.
- bootstrapping significa ejecutar comandos cuando se inicia una máquina
- Ese script solo se ejecuta una vez en el primer inicio de la instancia
- Los datos de usuario de EC2 se utilizan para automatizar tareas de arranque como:
  - Instalar actualizaciones
  - Instalación de software
  - Descarga de archivos comunes de Internet
  - Cualquier cosa que puedas pensar ...
- El script de datos de usuario de EC2 se ejecuta con el usuario raíz

# Cómputo en AWS

## EC2 User Data – Lab

- Lanzaremos nuestro primer servidor virtual utilizando la consola de AWS
- Obtendremos una primera aproximación de alto nivel a los diversos parámetros.
- Veremos que nuestro servidor web se inicia utilizando los datos de usuario de EC2
- Aprenderemos cómo iniciar/detener/terminar nuestra instancia.

# Cómputo en AWS

## Convención en tipos de instancias

- Podemos usar diferentes tipos de instancias EC2 que están optimizadas para diferentes casos de uso:  
:(<https://aws.amazon.com/ec2/instance-types/>)
- AWS tiene la siguiente convención de nomenclatura:

m5.2xlarge

- m: clase de instancia
- 5: generación (AWS los mejora con el tiempo)
- 2xlarge: tamaño dentro de la clase de instancia

General Purpose

Compute Optimized

Memory Optimized

Accelerated Computing

Storage Optimized

Instance Features

Measuring Instance  
Performance



# Cómputo en AWS

## Instancias – Uso general

- Excelente para una diversidad de cargas de trabajo, como servidores web o repositorios de código
- Balance entre:
  - Cómputo
  - Memoria
  - Redes
- En el curso, utilizaremos t2.micro, que es una instancia EC2 de uso general.

### General Purpose

General purpose instances provide a balance of compute, memory and networking resources, and can be used for a variety of diverse workloads. These instances are ideal for applications that use these resources in equal proportions such as web servers and code repositories.

Mac	T4g	T3	T3a	T2	M6g	M5	M5a	M5n	M5zn	M4	A1
-----	-----	----	-----	----	-----	----	-----	-----	------	----	----

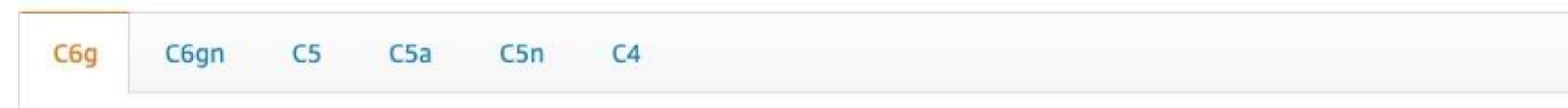
# Cómputo en AWS

## Instancias – Optimizada en Cómputo

- Ideal para tareas de computación intensiva que requieren procesadores de alto rendimiento:
  - Cargas de trabajo de procesamiento por lotes
  - Servidores web de alto rendimiento
  - Informática de alto rendimiento (HPC)
  - Modelado científico y aprendizaje automático
  - Servidores de juegos dedicados

### Compute Optimized

Compute Optimized instances are ideal for compute bound applications that benefit from high performance processors. Instances belonging to this family are well suited for batch processing workloads, media transcoding, high performance web servers, high performance computing (HPC), scientific modeling, dedicated gaming servers and ad server engines, machine learning inference and other compute intensive applications.



# Cómputo en AWS

## Instancias – Optimizada en Memoria

- Rendimiento rápido para cargas de trabajo que procesan grandes conjuntos de datos en la memoria
- Casos de uso:
  - Bases de datos relacionales/no relacionales de alto rendimiento
  - Almacenes de caché distribuidos a escala web
  - Bases de datos en memoria optimizadas para BI (inteligencia empresarial)
  - Aplicaciones que realizan procesamiento en tiempo real de grandes datos no estructurados

### Memory Optimized

Memory optimized instances are designed to deliver fast performance for workloads that process large data sets in memory.



# Cómputo en AWS

## Instancias – Optimizada en Almacenamiento

- Ideal para tareas de uso intensivo de almacenamiento que requieren un alto acceso secuencial de lectura y escritura a grandes conjuntos de datos en el almacenamiento local
- Casos de uso:
  - Sistemas de procesamiento de transacciones en línea de alta frecuencia (OLTP)
  - Bases de datos relacionales y NoSQL
  - Caché para bases de datos en memoria (por ejemplo, Redis)
  - Aplicaciones de almacenamiento de datos
  - Sistemas de archivos distribuidos

### Storage Optimized

Storage optimized instances are designed for workloads that require high, sequential read and write access to very large data sets on local storage. They are optimized to deliver tens of thousands of low-latency, random I/O operations per second (IOPS) to applications.

I3	I3en	D2	D3	D3en	H1
----	------	----	----	------	----



# Cómputo en AWS

## Instancias - Ejemplos

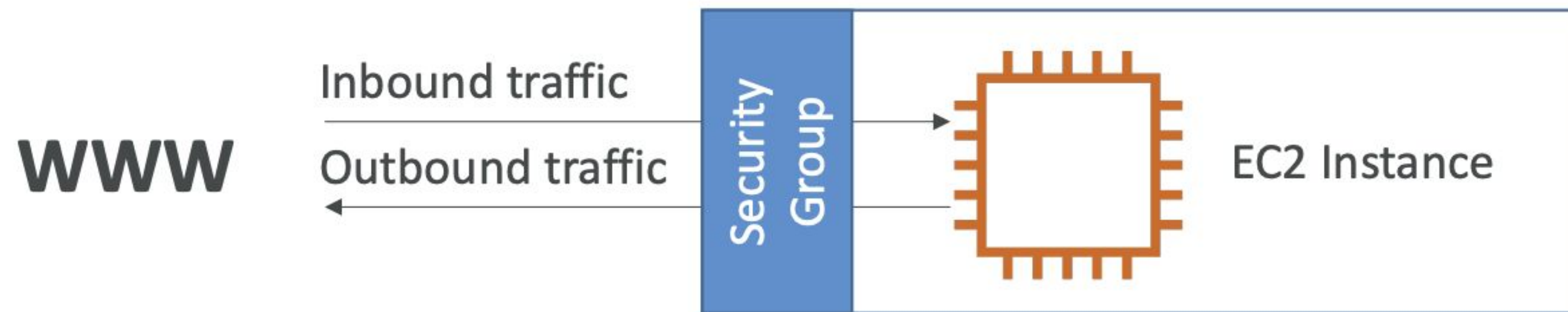
Instance	vCPU	Mem (GiB)	Storage	Network Performance	EBS Bandwidth (Mbps)
t2.micro	1	1	EBS-Only	Low to Moderate	
t2.xlarge	4	16	EBS-Only	Moderate	
c5d.4xlarge	16	32	1 x 400 NVMe SSD	Up to 10 Gbps	4,750
r5.16xlarge	64	512	EBS Only	20 Gbps	13,600
m5.8xlarge	32	128	EBS Only	10 Gbps	6,800



# Cómputo en AWS

## Grupos de Seguridad – Intro

- Los grupos de seguridad son fundamentales para la seguridad de la red en AWS
- Controlan cómo se permite el tráfico dentro o fuera de nuestras instancias EC2.



- Los grupos de seguridad solo contienen reglas para permitir.
- Las reglas de los grupos de seguridad pueden hacer referencia por IP o por grupo de seguridad

# Cómputo en AWS

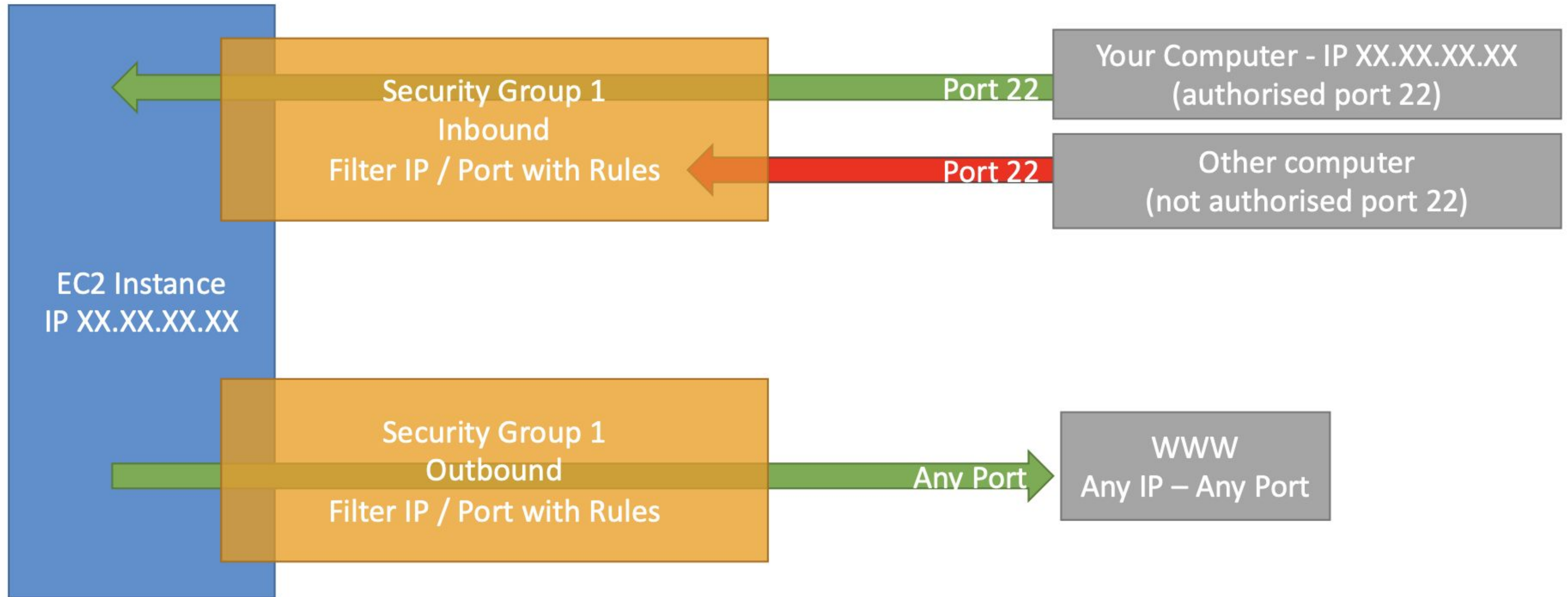
## Grupos de Seguridad – Deep

- Los grupos de seguridad actúan como un "cortafuegos" en las instancias EC2
- Ellos regulan:
  - Acceso a Puertos
  - Rangos de IP autorizados: IPv4 e IPv6
  - Control de la red de entrada (de otros a la instancia)
  - Control de red de salida (de una instancia a otra)

Type ⓘ	Protocol ⓘ	Port Range ⓘ	Source ⓘ	Description ⓘ
HTTP	TCP	80	0.0.0.0/0	test http page
SSH	TCP	22	122.149.196.85/32	
Custom TCP Rule	TCP	4567	0.0.0.0/0	java app

# Cómputo en AWS

## Grupos de Seguridad – Diagrama Simple



# Cómputo en AWS

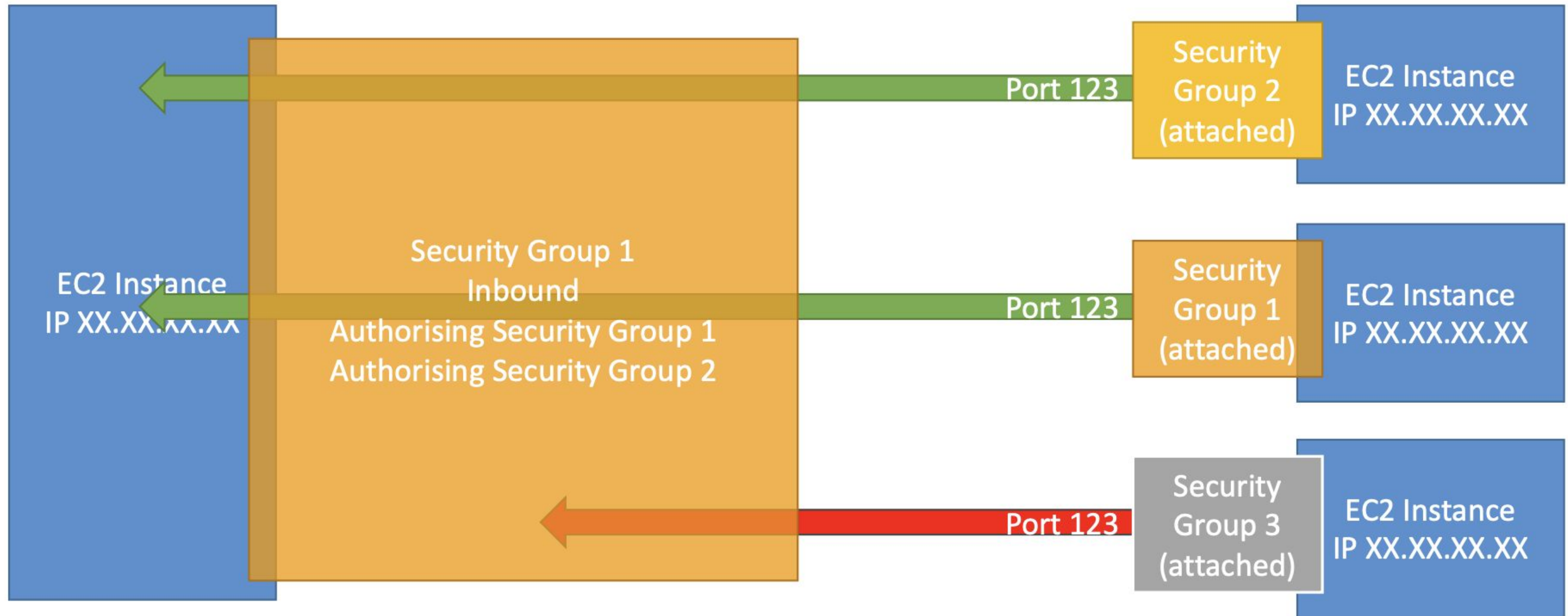
## Grupos de Seguridad – Generalidades

- Se puede adjuntar a varias instancias
- Solo puede ser usado en región/VPC
- Vive "fuera" de EC2: si el tráfico está bloqueado, la instancia de EC2 no lo verá
- Es bueno mantener un grupo de seguridad separado para el acceso SSH
- Si no se puede acceder a su aplicación (tiempo de espera), puede ser un problema del grupo de seguridad
- Si su aplicación da un error de "conexión rechazada", entonces es un error de la aplicación o no se inicia
- Todo el tráfico entrante está bloqueado de forma predeterminada
- Todo el tráfico saliente está autorizado por defecto



# Cómputo en AWS

## Grupos de Seguridad – Referenciando otros SG





# Cómputo en AWS

## Grupos de Seguridad – Puertos Básicos

- 22 = SSH (Secure Shell): inicie sesión en una instancia de Linux
- 21 = FTP (Protocolo de transferencia de archivos): cargue archivos en un recurso compartido de archivos
- 22 = SFTP (Protocolo seguro de transferencia de archivos): cargue archivos usando SSH
- 80 = HTTP: accede a sitios web no seguros
- 443 = HTTPS: acceda a sitios web seguros
- 3389 = RDP (Protocolo de escritorio remoto): inicie sesión en una instancia de Windows

# Cómputo en AWS

## Grupos de Seguridad –Tools para conectar ssh

	SSH	Putty	EC2 Instance Connect
Mac	✓		✓
Linux	✓		✓
Windows < 10		✓	✓
Windows >= 10	✓	✓	✓

# Cómputo en AWS

## EC2 – Opciones de Compra

- Instancias bajo demanda: carga de trabajo corta, precios predecibles, pago por segundo
- Reservado (1 y 3 años)
  - Instancias reservadas: cargas de trabajo largas
  - Instancias reservadas convertibles: cargas de trabajo largas con instancias flexibles
- Planes de ahorro (1 y 3 años): compromiso con una cantidad de uso, carga de trabajo prolongada
- Instancias puntuales (spot): cargas de trabajo cortas, económicas, pueden perder instancias (menos confiable)
- Hosts dedicados: reserve un servidor físico completo, controle la ubicación de la instancia
- Instancias dedicadas: ningún otro cliente compartirá su hardware
- Reservas de capacidad: reserva de capacidad en una zona de disponibilidad específica para cualquier duración

# Cómputo en AWS

## EC2 – Opciones de Compra – On-demand

- Paga por lo que usas:
  - Linux o Windows: facturación por segundo, después del primer minuto
  - Todos los demás sistemas operativos: facturación por hora
- Tiene el costo más alto pero no hay pago por adelantado
- Sin compromiso a largo plazo
- Recomendado para cargas de trabajo ininterrumpidas y a corto plazo, donde no se puede predecir cómo se comportará la aplicación



# Cómputo en AWS

## EC2 – Opciones de Compra – Reservadas

- Varía el % de descuento en comparación con On-demand (Ahorro)
- Reserva atributos de instancia específicos (tipo de instancia, región, tenencia, sistema operativo)
- Periodo de Reserva – 1 año (+descuento) o 3 años (+++descuento)
- Opciones de pago: sin pago por adelantado (+), pago por adelantado parcial (++) , pago por adelantado (+++)
- Ámbito de instancia reservada: regional o zonal (capacidad de reserva en una zona de disponibilidad)
- Recomendado para aplicaciones de uso de estado estable (piense en la base de datos)
- Puede comprar y vender en el Marketplace de instancias reservadas
- Instancia reservada convertible
  - Puede cambiar el tipo de instancia EC2, la familia de instancias, el sistema operativo, el alcance y la tenencia
  - Hasta 66% de descuento



# Cómputo en AWS

## EC2 – Opciones de Compra – Planes de Ahorro

- Obtenga un descuento basado en el uso a largo plazo (hasta un 70 %\*, igual que las instancias reservadas)
- Comprometerse a cierto tipo de uso (\$10/hora por 1 o 3 años)
- El uso más allá de EC2 Savings Plans se factura al precio On-Demand
- no permitido en familias de instancias y una región de AWS específicas (por ejemplo, M5 en us-east-1)
- Flexible en:
  - Tamaño de instancia (por ejemplo, m5.xlarge, m5.2xlarge)
  - SO (por ejemplo, Linux, Windows)
  - Tenencia (Host, Dedicado, Predeterminado)

# Cómputo en AWS

## EC2 – Opciones de Compra – Spot



- Puede obtener un descuento de hasta el 90% en comparación con On-demand
- Instancias que puede "perder" en cualquier momento si su precio máximo es menor que el precio al contado actual
- Las instancias MÁS rentables en AWS
- Útil para cargas de trabajo resistentes a fallas
  - Trabajos por lotes
  - Análisis de los datos
  - Procesamiento de imágenes
  - Cualquier carga de trabajo distribuida
  - Cargas de trabajo con una hora de inicio y finalización flexible
- No apto para trabajos críticos o bases de datos.

# Cómputo en AWS

## EC2 – Opciones de Compra – Dedicados

- Un servidor físico con capacidad de instancia EC2 completamente dedicado a su uso
- Le permite abordar los requisitos de cumplimiento y utilizar sus licencias de software vinculadas al servidor existentes (licencias de software por socket, por núcleo, por VM)
- Opciones de compra:
  - Bajo demanda: pago por segundo por host dedicado activo
  - Reservado: 1 o 3 años (sin pago inicial, pago inicial parcial, pago inicial total)
- La opción más cara
- Útil para software que tiene un modelo de licencia complicado (BYOL – Traiga su propia licencia)
- O para empresas que tienen fuertes necesidades regulatorias o de cumplimiento.

# Cómputo en AWS

## EC2 – Opciones de Compra – Dedicadas

- Las instancias se ejecutan en hardware dedicado a usted
- Puede compartir hardware con otras instancias en la misma cuenta
- Sin control sobre la ubicación de la instancia (puede mover el hardware después de detener/iniciar)

Characteristic	Dedicated Instances	Dedicated Hosts
Enables the use of dedicated physical servers	X	X
Per instance billing (subject to a \$2 per region fee)	X	
Per host billing		X
Visibility of sockets, cores, host ID		X
Affinity between a host and instance		X
Targeted instance placement		X
Automatic instance placement	X	X
Add capacity using an allocation request		X



# Cómputo en AWS

## EC2 – Reserva de capacidad

- Reserve la capacidad de las instancias bajo demanda en una zona de disponibilidad específica para cualquier duración
- Siempre tiene acceso a la capacidad de EC2 cuando la necesita
- Sin compromiso de tiempo (crear/cancelar en cualquier momento), sin descuentos de facturación
- Combínalo con Instancias Reservadas Regionales y Planes de Ahorro para beneficiarse de descuentos en la facturación
- Se le cobra según la tarifa bajo demanda, ya sea que ejecute instancias o no.
- Adecuado para cargas de trabajo ininterrumpidas a corto plazo que deben estar en una zona de disponibilidad específica



# Cómputo en AWS

## EC2 – Opciones de Compra Resumen

- Bajo demanda: venir y alojarse en el resort cuando queramos, pagamos el precio completo
- Reservado: como planificar con anticipación y si planeamos quedarnos por mucho tiempo, podemos obtener un buen descuento.
- Planes de Ahorro: paga una cierta cantidad por hora por cierto período y permanecer en cualquier tipo de habitación (por ejemplo, King, Suite, Sea View, ...)
- Instancias spot: el hotel permite a las personas ofertar por las habitaciones vacías y el mejor postor se queda con las habitaciones. Puede ser expulsado en cualquier momento
- Hosts dedicados: Reservamos un edificio completo del resort
- Reservas de Capacidad: reservas una habitación por un periodo a precio completo aunque no te alojes en ella



# Cómputo en AWS

## EC2 – Opciones de Compra – Comparaciones

Tipo de precio	Precio (por hora)
Bajo demanda	\$0.10
Instancia de spot (precio de spot)	\$0.038 - \$0.039 (hasta 61% de descuento)
Instancia reservada (1 año)	\$0.062 (sin pago inicial) - \$0.058 (todo por adelantado)
Instancia reservada (3 años)	\$0.043 (sin pago inicial) - \$0.037 (todo por adelantado)
Plan Ahorro EC2 (1 año)	\$0.062 (sin pago inicial) - \$0.058 (todo por adelantado)
Instancia convertible reservada (1 año)	\$0.071 (sin pago inicial) - \$0.066 (todo por adelantado)
Anfitrión dedicado	Precio bajo demanda
Reserva de host dedicado	Hasta un 70% de descuento
Reservas de Capacidad	Precio bajo demanda



# Cómputo en AWS

## EC2 – Responsabilidad Compartida



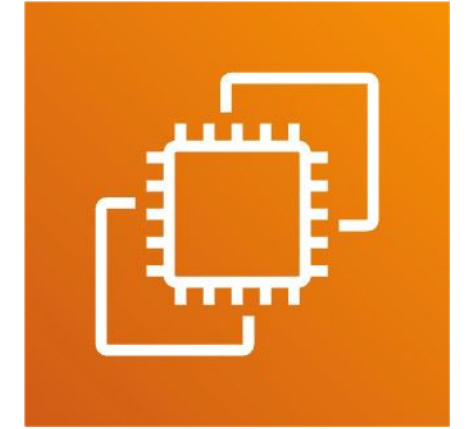
- Infraestructura (seguridad de red global)
- Aislamiento en hosts físicos
- Sustitución de hardware defectuoso
- Validación de cumplimiento



- Reglas de grupos de seguridad
- Parches y actualizaciones del sistema operativo
- Software y utilidades instalados en la instancia EC2
- Funciones de IAM asignadas a la administración de acceso de usuarios de EC2 e IAM
- Seguridad de datos en su instancia

# Cómputo en AWS

## EC2 – Resumen

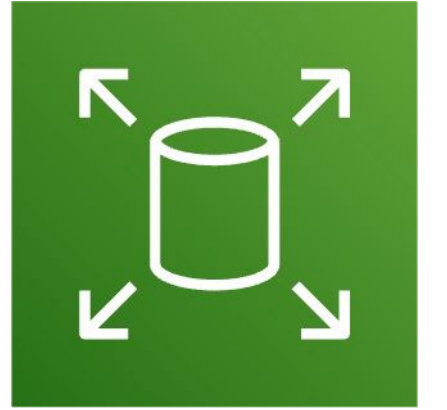


- Instancia EC2: AMI (SO) + Tamaño de instancia (CPU + RAM) + Almacenamiento + grupos de seguridad + Datos de usuario de EC2
- Grupos de seguridad: cortafuegos adjunto a la instancia EC2
- Datos de usuario de EC2: secuencia de comandos iniciada en el primer inicio de una instancia
- SSH: inicie una terminal en nuestras instancias EC2 (puerto 22)
- Rol de instancia EC2: enlace a roles de IAM
- Opciones de compra: On-Demand, Spot, Reserved  
(Standard + Convertible + Scheduled), Dedicated Host, Dedicated Instance



# Cómputo en AWS

## Almacenamiento – EBS Intro



- Un volumen EBS (Elastic Block Store) es una unidad de red que puede conectar a sus instancias mientras se ejecutan
- Permite que sus instancias conserven los datos, incluso después de su terminación
- Solo se pueden montar en una instancia a la vez
- Están vinculados a una zona de disponibilidad específica
- Analogía: Piense en ellos como una "memoria USB de red"
- Nivel gratuito: 30 GB de almacenamiento EBS gratuito de tipo Propósito general (SSD) o Magnético por mes

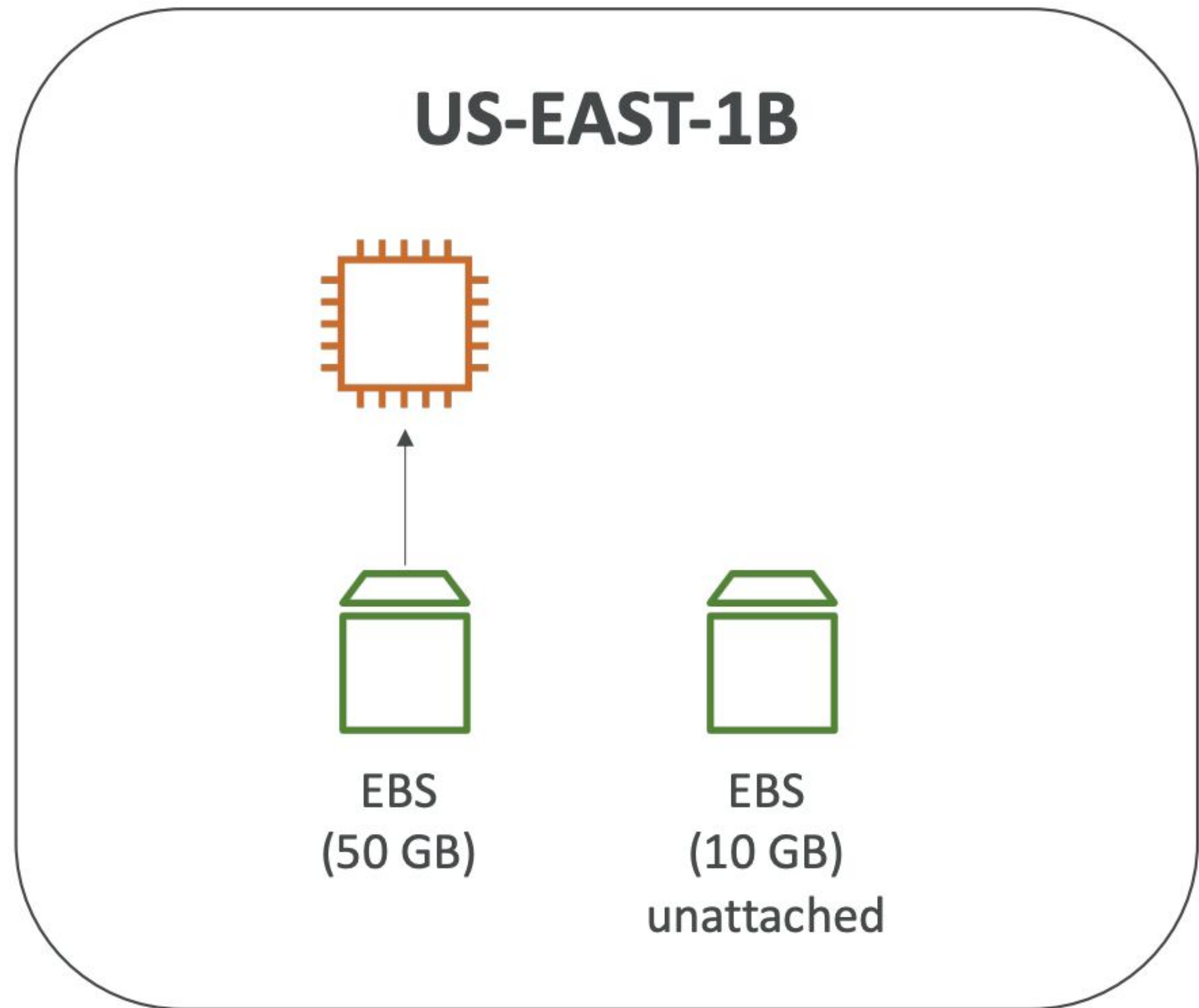
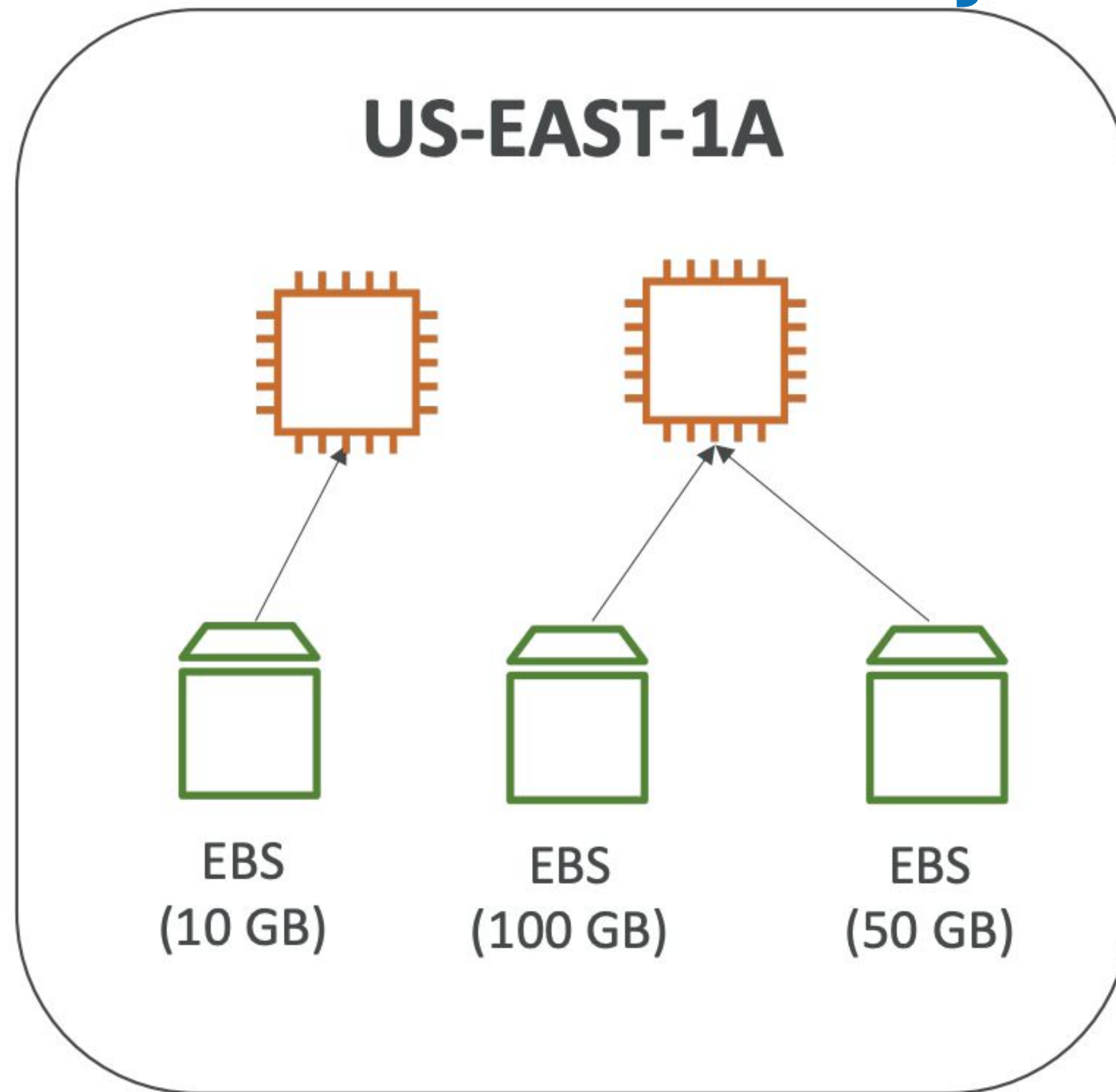
# Cómputo en AWS

## Almacenamiento – EBS Intro

- Es una unidad de red (es decir, no es una unidad física)
  - Utiliza la red para comunicar la instancia, lo que significa que puede haber un poco de latencia.
  - Se puede desconectar de una instancia EC2 y adjuntar a otra rápidamente
- Está bloqueado en una zona de disponibilidad (AZ)
  - Un volumen de EBS en us-east-1a no se puede adjuntar a us-east-1b
  - Para mover un volumen, primero debe tomar un snapshot
- Tener una capacidad aprovisionada (tamaño en GB e IOPS)
  - Se le factura por toda la capacidad aprovisionada
  - Puede aumentar la capacidad de la unidad con el tiempo

# Cómputo en AWS

## Almacenamiento – EBS Ejemplo



# Cómputo en AWS

## Almacenamiento – EBS Eliminar

Volume Type ⓘ	Device ⓘ	Snapshot ⓘ	Size (GiB) ⓘ	Volume Type ⓘ	IOPS ⓘ	Throughput (MB/s) ⓘ	Delete on Termination ⓘ	Encryption ⓘ
Root	/dev/xvda	snap-09f18f682fd23a1b1	8	General Purpose SSD (gp2) ▼	100 / 3000	N/A	<input checked="" type="checkbox"/>	Not Encrypted ▼
EBS ▼	/dev/sdb ▼	Search (case-insensit	8	General Purpose SSD (gp2) ▼	100 / 3000	N/A	<input type="checkbox"/>	Not Encrypted ▼

Add New Volume

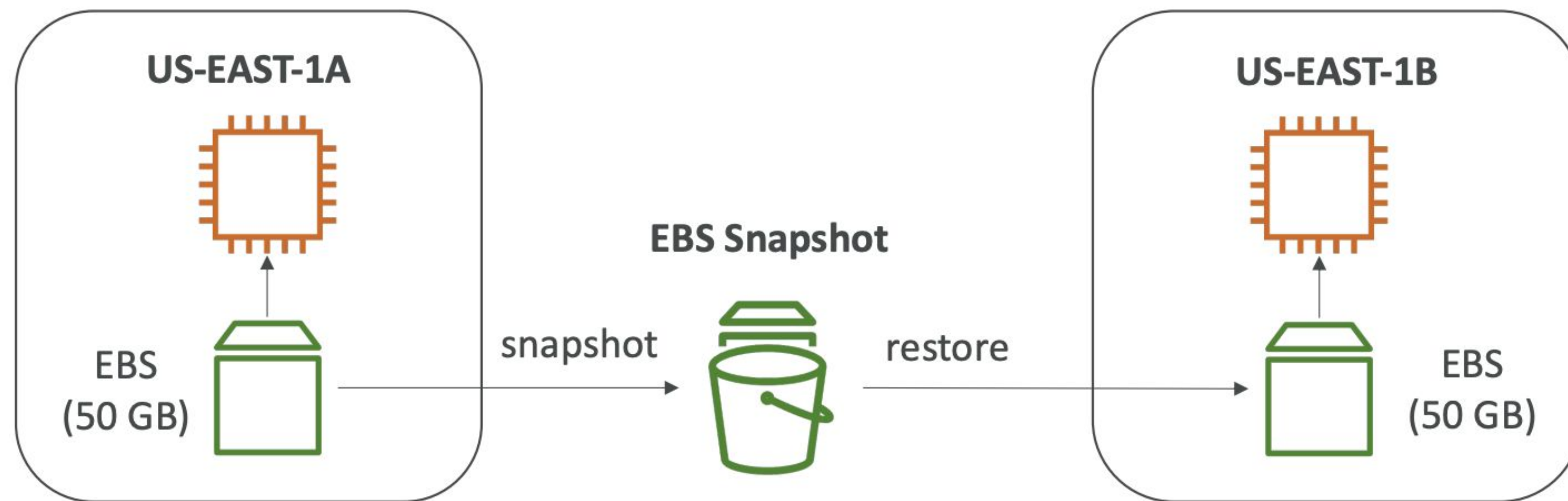
- Controla el comportamiento de EBS cuando finaliza una instancia EC2
- De forma predeterminada, el volumen de EBS raíz se elimina (atributo habilitado)
- De forma predeterminada, no se elimina ningún otro volumen de EBS adjunto (atributo deshabilitado)
- Esto puede ser controlado por la consola de AWS / AWS CLI
- Caso de uso: preservar el volumen raíz cuando finaliza la instancia



# Cómputo en AWS

## Almacenamiento – EBS Snapshots

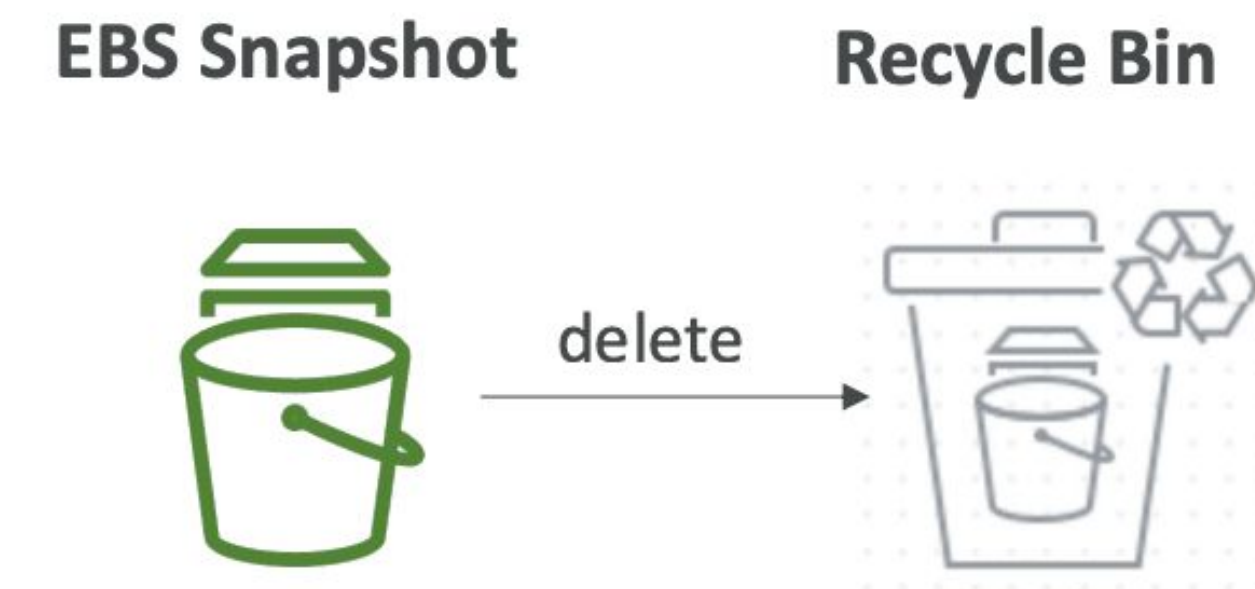
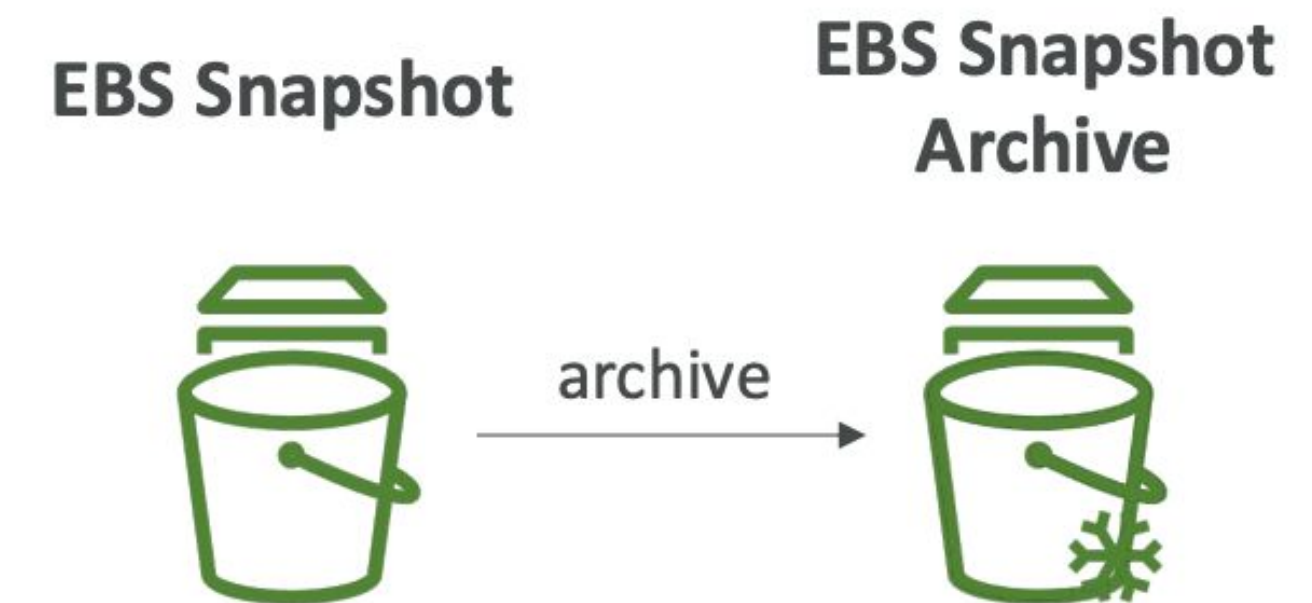
- Realiza una copia de seguridad (instantánea) del volumen de EBS en un momento dado
- No es necesario desconectar el volumen para hacer una snapshot, pero se recomienda
- Puede copiar Snapshots en AZ o región



# Cómputo en AWS

## Almacenamiento – EBS Snapshots

- Archivo de Snapshots de EBS
  - Mueva una instantánea a un "nivel de archivo" que es un 75 % más económico
  - La restauración del archivo demora entre 24 y 72 horas
- Papelera de reciclaje para Snapshots de EBS
  - Configure reglas para conservar las Snapshots eliminadas para que pueda recuperarlas después de una eliminación accidental
  - Especificar retención (de 1 día a 1 año)



# Cómputo en AWS

## Almacenamiento – AMI Intro

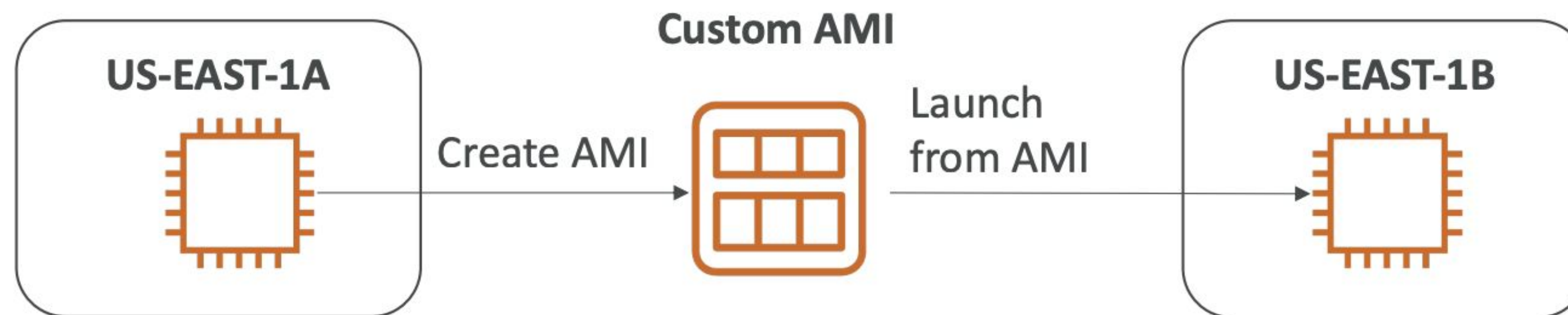


- AMI = Imagen de máquina de Amazon
- Las AMI son una personalización de una instancia EC2
  - Añades tu propio software, configuración, sistema operativo, monitorización...
  - Tiempo de configuración/arranque más rápido porque todo su software está preempaquetado
- Las AMI se construyen para una región específica (y se pueden copiar entre regiones)
- Puede lanzar instancias EC2 desde:
  - Una AMI pública: es proporcionada por AWS
  - Tus propias AMI: las creas y las mantienes tú mismo
  - Una AMI de AWS Marketplace: una AMI que alguien más hizo (y potencialmente vende)

# Cómputo en AWS

## Almacenamiento – AMI Intro

- Inicie una instancia EC2 y personalízala
- Detener la instancia (para la integridad de los datos)
- Cree una AMI: esto también creará Snapshots de EBS
- Lanzar instancias desde otras AMI



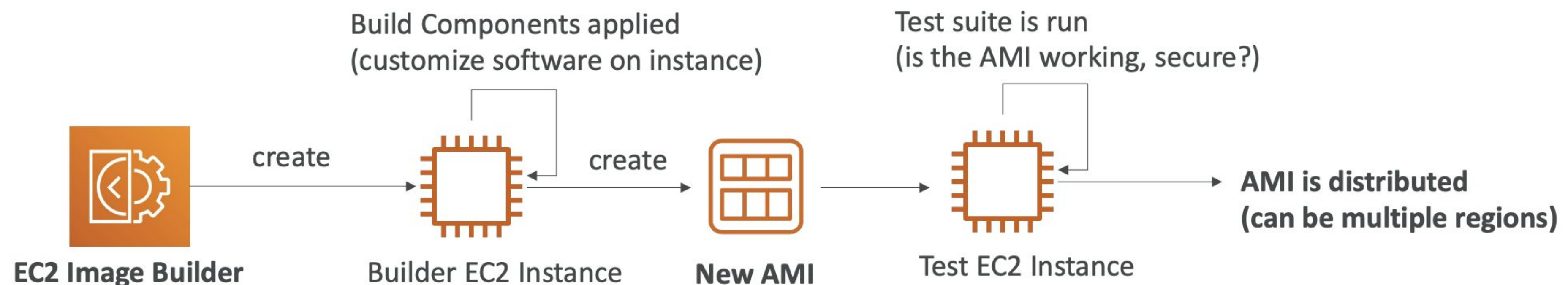


# Cómputo en AWS

## Almacenamiento – AMI Builder

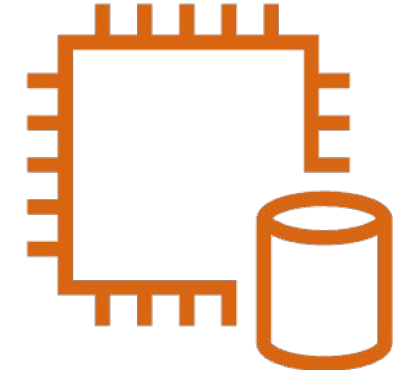


- Se utiliza para automatizar la creación de máquinas virtuales o imágenes de contenedores
- Automatiza la creación, mantenga, valide y pruebe las AMI de EC2
- Se puede ejecutar según un cronograma (semanalmente, siempre que se actualicen los paquetes, etc.)
- Servicio gratuito (solo paga por los recursos subyacentes)



# Cómputo en AWS

## Almacenamiento – Instance Store



- Los volúmenes de EBS son unidades de red con un rendimiento bueno pero "limitado"
- Si necesita un disco de hardware de alto rendimiento, use EC2 Instance Store
- Mejor rendimiento de E/S
- EC2 Instance Store pierde su almacenamiento si se detiene (efímero)
- Bueno para búfer/caché/datos temporales/contenido temporal
- Riesgo de pérdida de datos si falla el hardware
- Las copias de seguridad y la replicación son su responsabilidad

# Cómputo en AWS

## Almacenamiento – Instance Store

IOPS muy alto

Instance Size	100% Random Read IOPS	Write IOPS
i3.large *	100,125	35,000
i3.xlarge *	206,250	70,000
i3.2xlarge	412,500	180,000
i3.4xlarge	825,000	360,000
i3.8xlarge	1.65 million	720,000
i3.16xlarge	3.3 million	1.4 million
i3.metal	3.3 million	1.4 million
i3en.large *	42,500	32,500
i3en.xlarge *	85,000	65,000
i3en.2xlarge *	170,000	130,000
i3en.3xlarge	250,000	200,000
i3en.6xlarge	500,000	400,000
i3en.12xlarge	1 million	800,000
i3en.24xlarge	2 million	1.6 million
i3en.metal	2 million	1.6 million

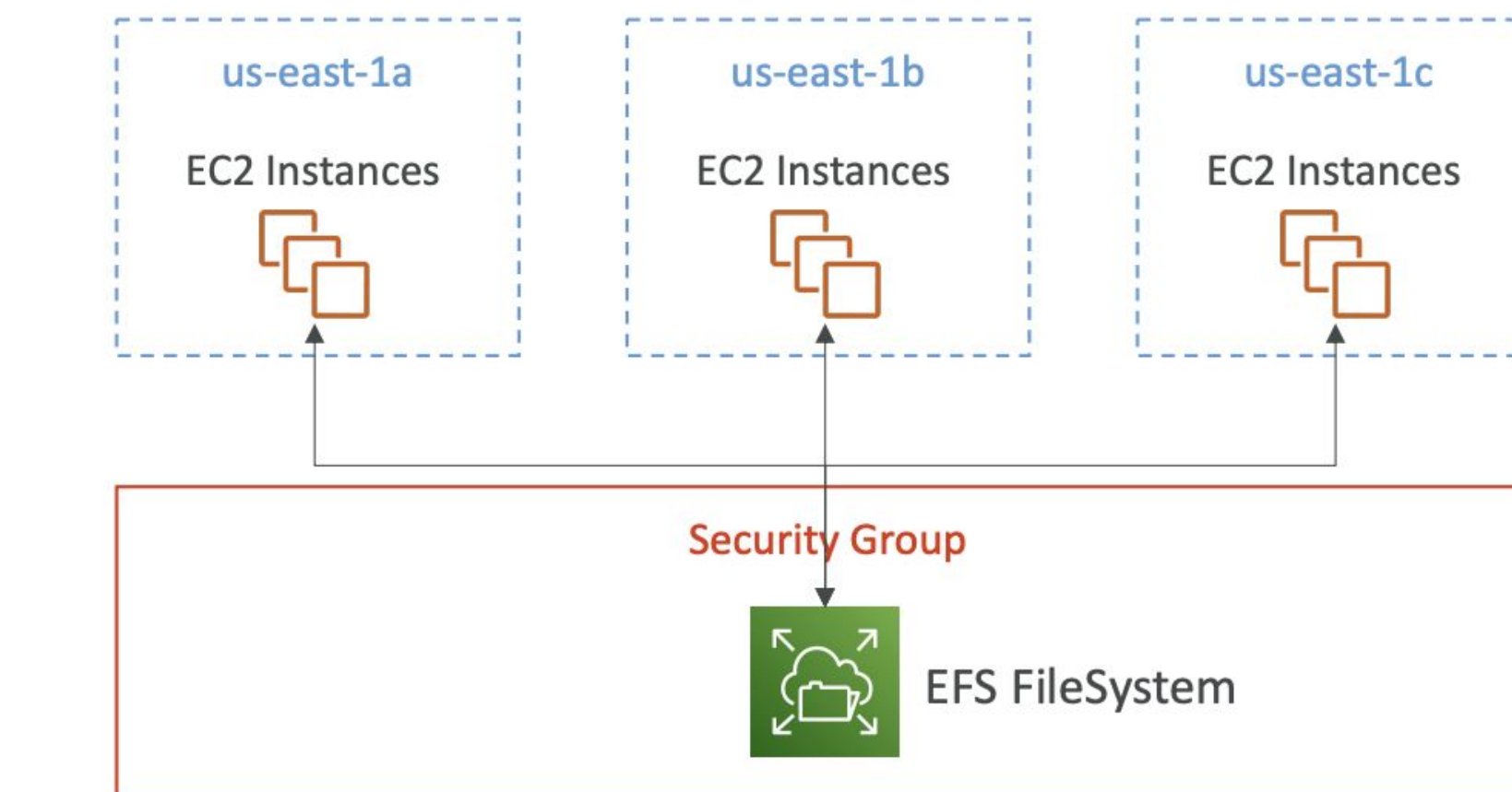


# Cómputo en AWS

## Almacenamiento – EFS Intro



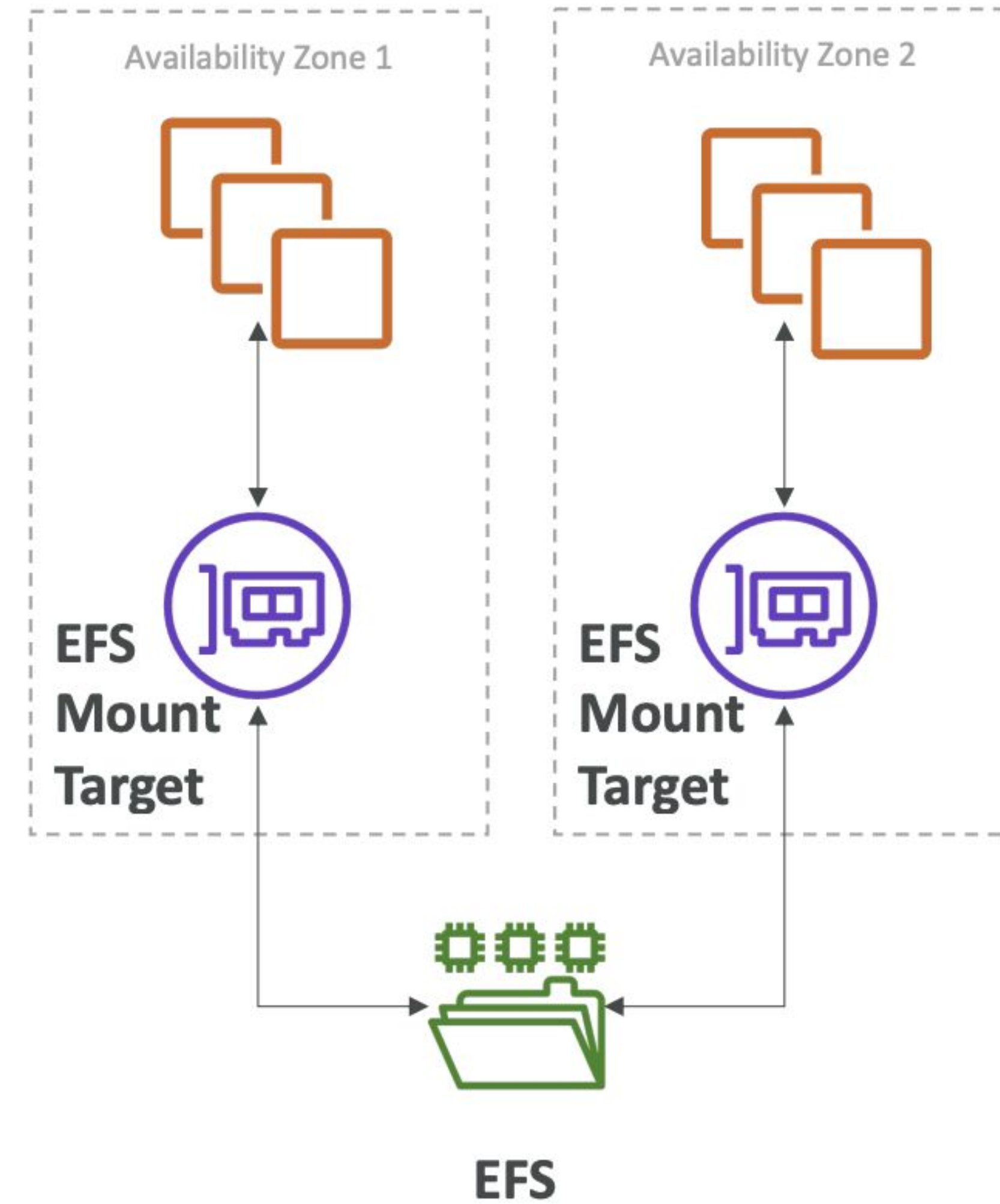
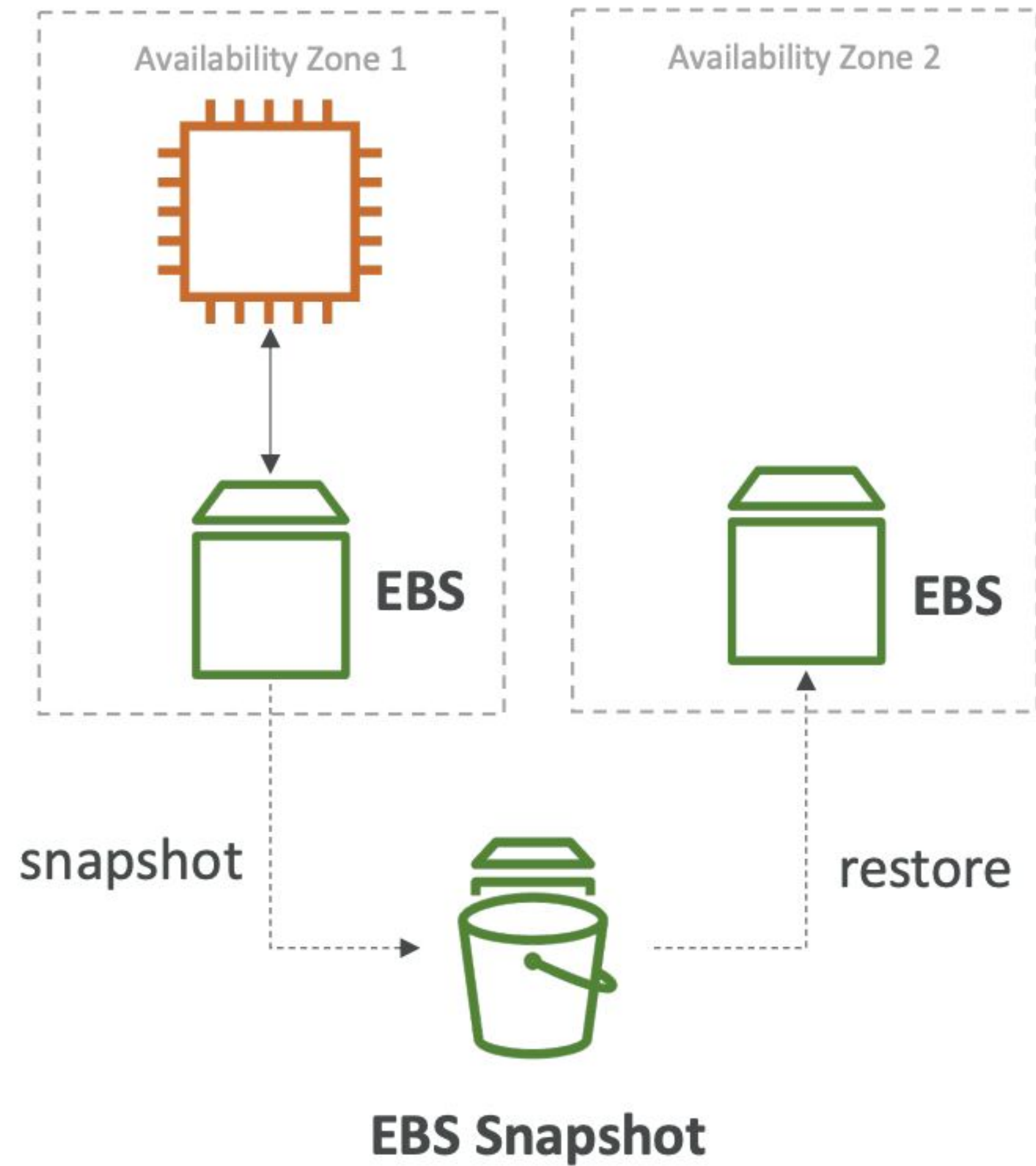
- NFS administrado (sistema de archivos de red) que se puede montar en cientos de EC2
- EFS funciona con instancias Linux EC2 en multi-AZ
- Altamente disponible, escalable, costoso (3x gp2), pago por uso, sin planificación de capacidad





# Cómputo en AWS

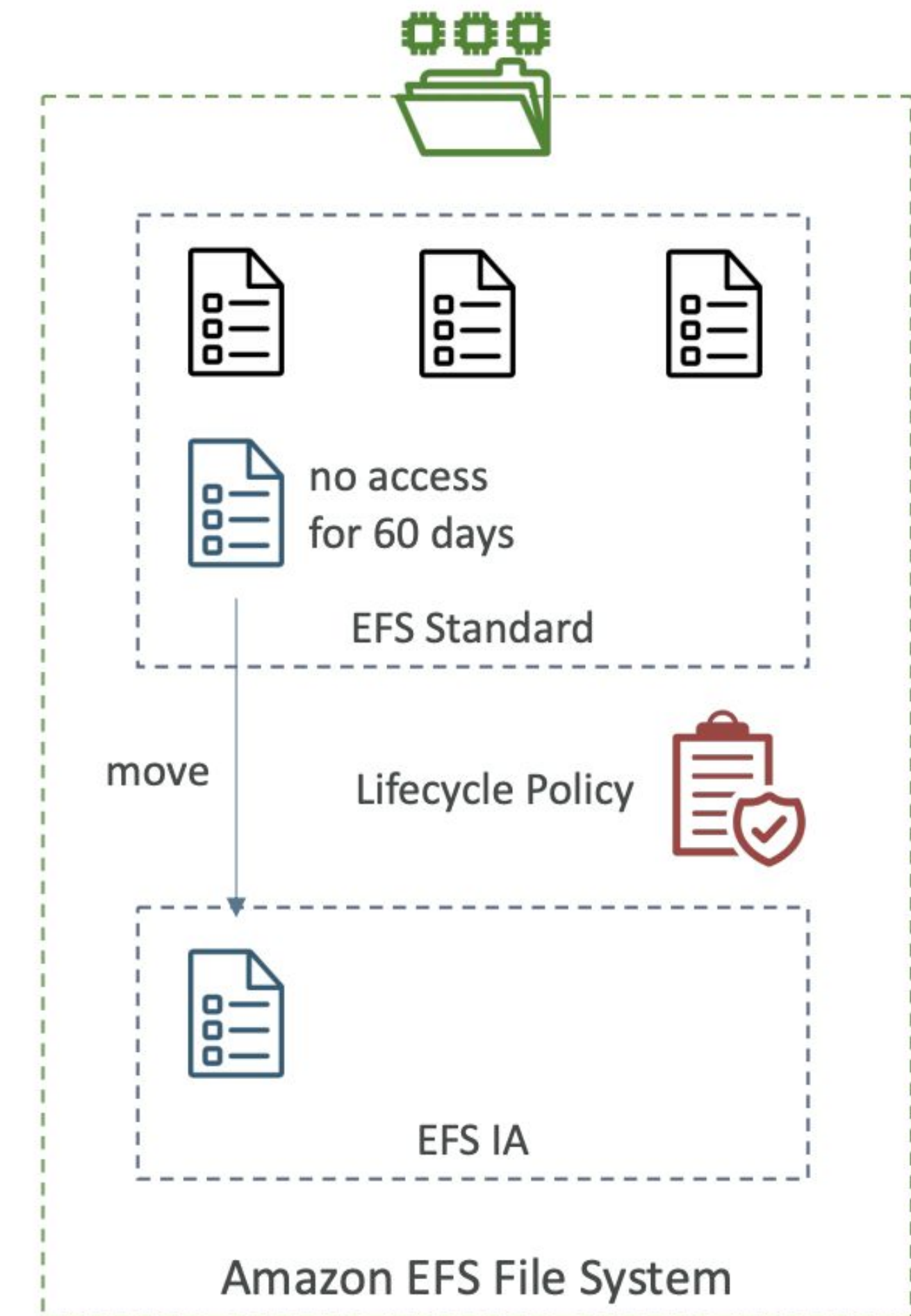
## Almacenamiento – EFS vs EBS



# Cómputo en AWS

## Almacenamiento – EFS-IA

- Clase de almacenamiento con costo optimizado para archivos a los que no se accede todos los días
- Hasta un 92 % menos de costo en comparación con el estándar EFS
- EFS moverá automáticamente sus archivos a EFS-IA en función de la última vez que se accedió a ellos
- Habilite EFS-IA con una política de ciclo de vida
- Ejemplo: mover archivos a los que no se accede durante 60 días a EFS-IA
- Transparente para las aplicaciones que acceden a EFS



# Cómputo en AWS

## Almacenamiento – Responsabilidad Compartida



- Infraestructura
  - Replicación de datos para volúmenes EBS y unidades EFS
  - Sustitución de hardware defectuoso
  - Asegurarse de que sus empleados no puedan acceder a sus datos
- Configuración de procedimientos de copia de seguridad/instantánea
  - Configurar el cifrado de datos
  - Responsabilidad de cualquier dato en las unidades
  - Comprender el riesgo de usar EC2 Instance Store



# Cómputo en AWS

## Almacenamiento – FSX Intro



- Lanzamiento de sistemas de archivos de alto rendimiento de terceros en AWS
- Servicio totalmente gestionado



**FSx for Lustre**



**FSx for  
Windows File  
Server**



**FSx for  
NetApp ONTAP**

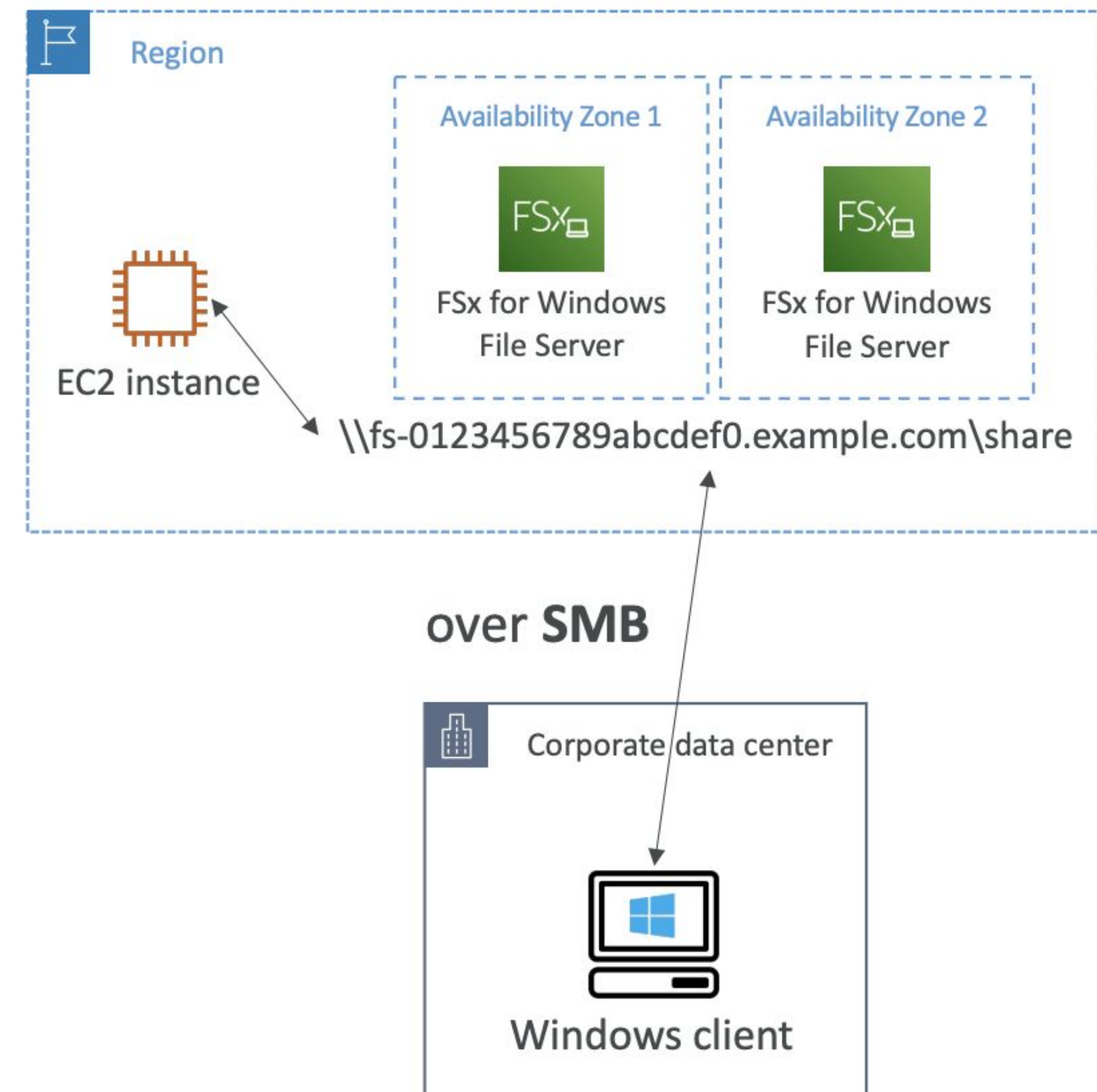


# Cómputo en AWS

## Almacenamiento – FSX for Windows



- Un sistema de archivos compartido nativo de Windows completamente administrado, altamente confiable y escalable
- Basado en el servidor de archivos de Windows
- Soporta protocolo SMB y Windows NTFS
- Integrado con Microsoft Active Directory
- Se puede acceder desde AWS o desde su infraestructura local

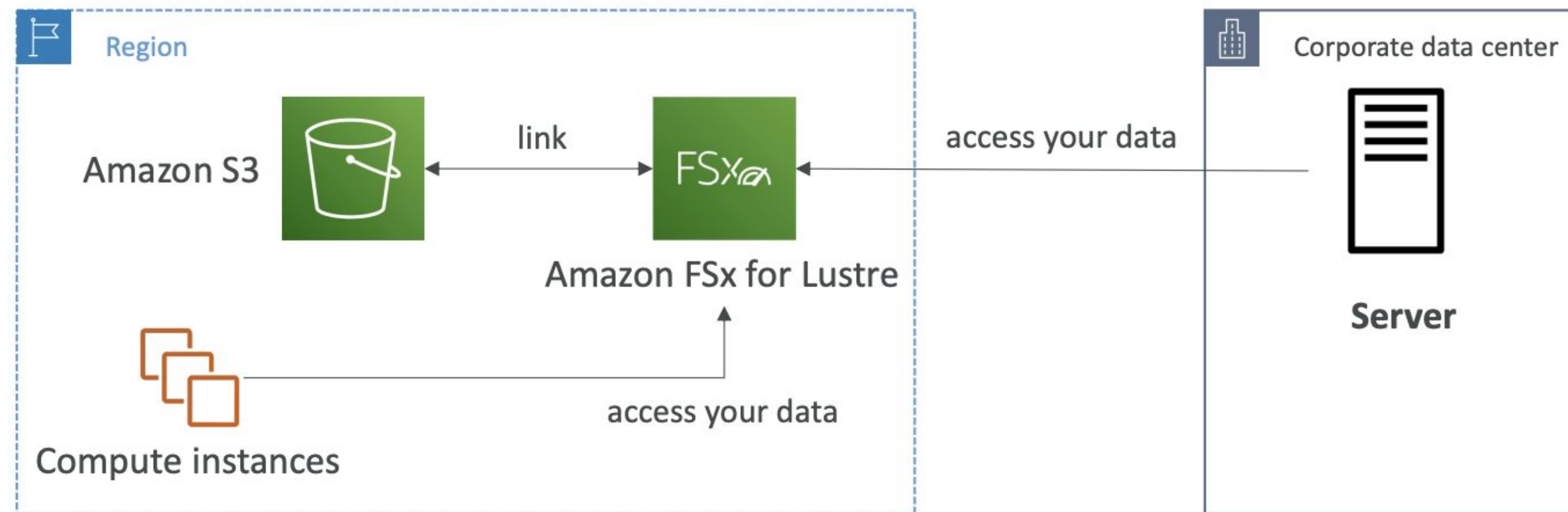


# Cómputo en AWS

## Almacenamiento – FSX for Lustre (Linux)



- Un almacenamiento de archivos escalable, de alto rendimiento y totalmente administrado para informática de alto rendimiento (HPC)
- El nombre Lustre se deriva de “Linux” y “cluster”
- Aprendizaje automático, análisis, procesamiento de video, modelado financiero, ...
- Escala hasta 100 GB/s, millones de IOPS, latencias de ms



# Cómputo en AWS

## Almacenamiento – Resumen

- Volúmenes de EBS:
- unidades de red conectadas a una instancia EC2 a la vez
- Asignado a una zona de disponibilidad
- Puede usar Snapshots de EBS para copias de seguridad/transferencia de volúmenes de EBS a través de AZ
- AMI: cree instancias EC2 listas para usar con nuestras personalizaciones
- EC2 Image Builder: cree, pruebe y distribuya automáticamente AMI
- Instance Stores EC2:
- Disco de hardware de alto rendimiento conectado a nuestra instancia EC2
- Perdido si nuestra instancia se detiene/termina
- EFS: sistema de archivos de red, se puede adjuntar a cientos de instancias en una región
- EFS-IA: clase de almacenamiento rentable para archivos a los que se accede con poca frecuencia
- FSx para Windows: Sistema de archivos de red para servidores Windows
- FSx for Lustre: sistema de archivos Linux para computación de alto rendimiento

# Cómputo en AWS

## Escalabilidad y Alta disponibilidad

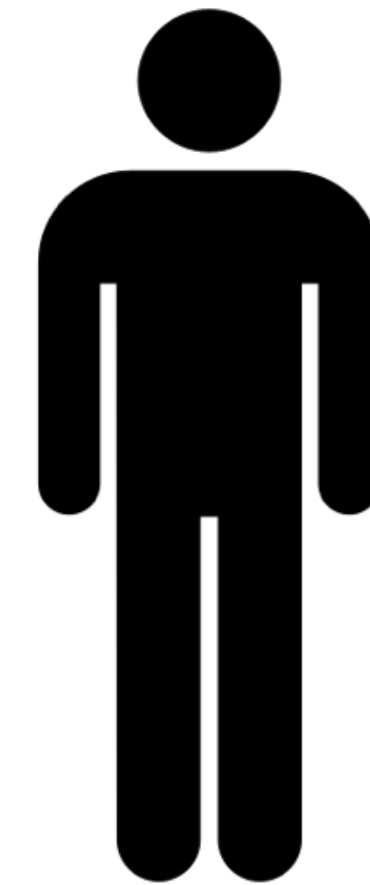
- Escalabilidad significa que una aplicación/sistema puede manejar mayores cargas adaptándose.
- Hay dos tipos de escalabilidad:
  - Escalabilidad Vertical
  - Escalabilidad horizontal (= elasticidad)
- La escalabilidad está vinculada pero es diferente a la alta disponibilidad
- Profundicemos en la distinción, usando un centro de llamadas como ejemplo



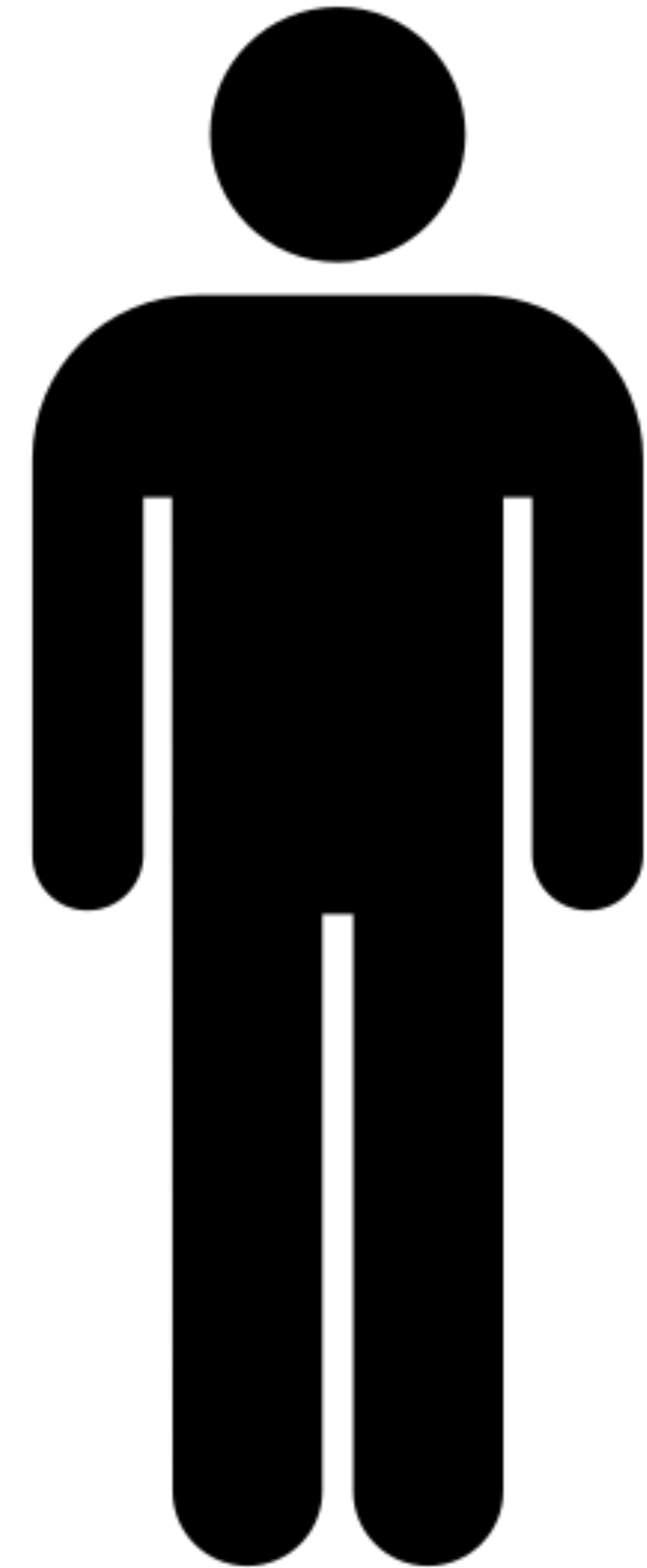
# Cómputo en AWS

## Escalabilidad Vertical

- Escalabilidad vertical significa aumentar el tamaño de la instancia
- Por ejemplo, su aplicación se ejecuta en un t2.micro
- Escalar esa aplicación verticalmente significa ejecutarla en un t2.large
- La escalabilidad vertical es muy común para los sistemas no distribuidos, como una base de datos.
- Por lo general, hay un límite de cuánto puede escalar verticalmente (límite de hardware)



operador  
junior

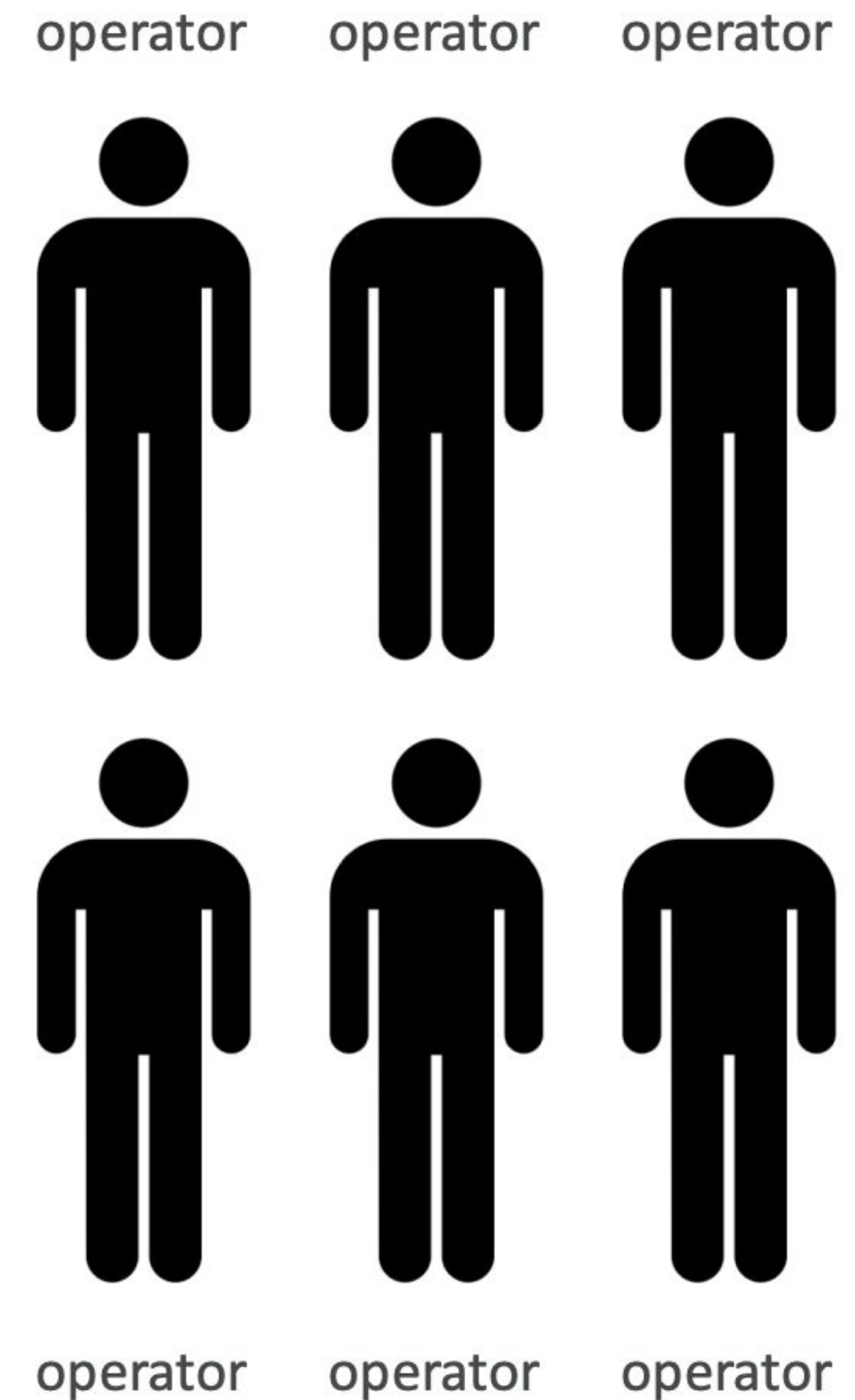


operador  
senior

# Cómputo en AWS

## Escalabilidad Horizontal

- Escalabilidad horizontal significa aumentar el número de instancias/sistemas para su aplicación
- El escalado horizontal implica sistemas distribuidos.
- Esto es muy común para aplicaciones web/aplicaciones modernas
- Es fácil de escalar horizontalmente gracias a las ofertas en la nube como Amazon EC2

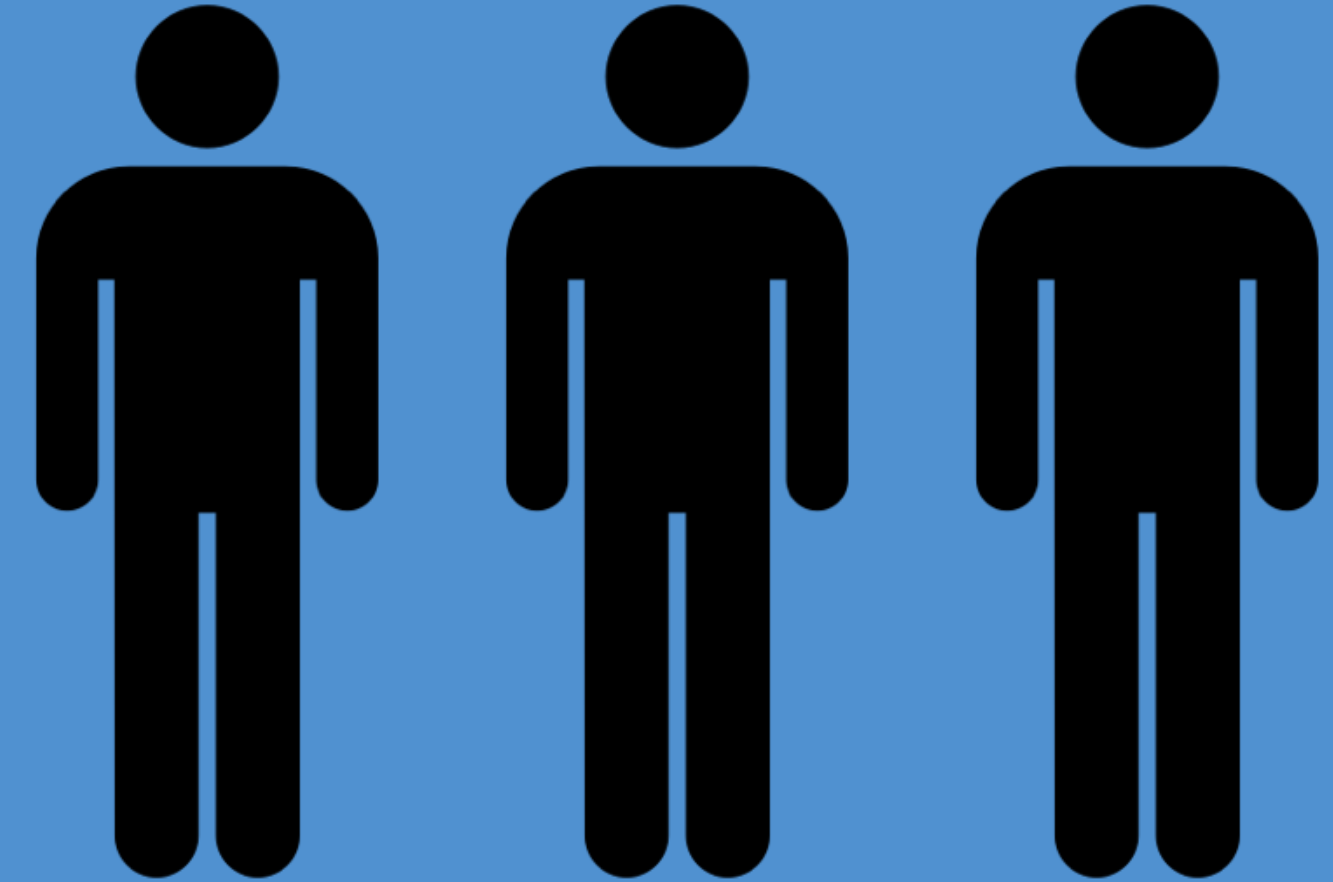


# Cómputo en AWS

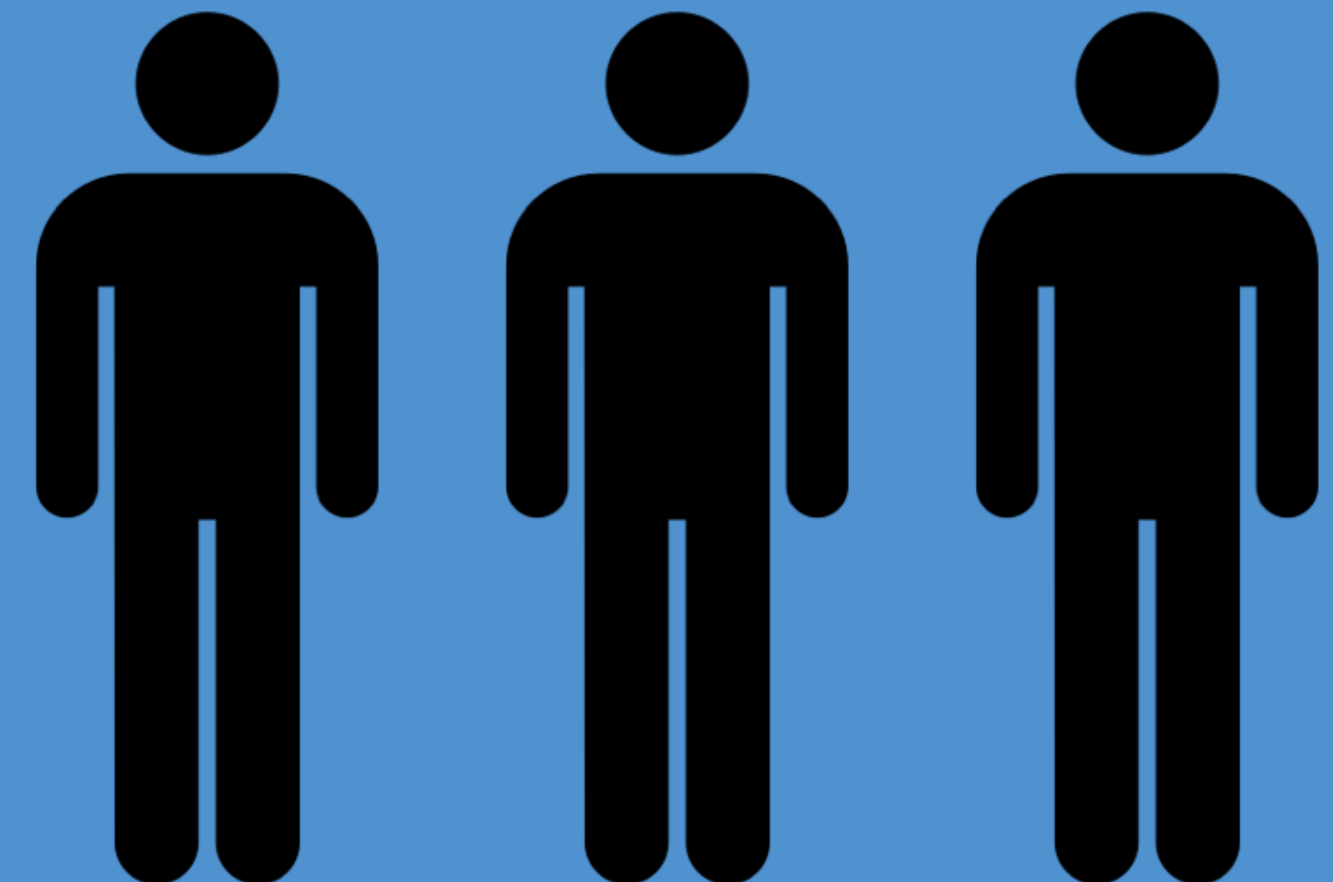
## Alta disponibilidad

- La alta disponibilidad suele ir de la mano con el escalado horizontal
- Alta disponibilidad significa ejecutar su aplicación/sistema en al menos 2 zonas de disponibilidad
- El objetivo de la alta disponibilidad es sobrevivir a una pérdida del centro de datos (desastre)

primer edificio en nueva york



segundo edificio en san francisco



# Cómputo en AWS

## Escalabilidad y Alta disponibilidad para EC2

- Escalado vertical: aumentar el tamaño de la instancia (= escalar hacia arriba o hacia abajo)
  - Desde: t2.nano – 0.5G de RAM, 1 vCPU
  - Hasta: u-12tb1.metal: 12,3 TB de RAM, 448 vCPU
- Escalado horizontal: aumentar el número de instancias (= escalado hacia afuera/adentro)
  - Grupo de Auto Scaling
  - Balanceador de carga
- Alta disponibilidad: ejecuta instancias para la misma aplicación en múltiples AZ
  - Auto Scaling Grupo multi AZ
  - Balanceador de carga multi AZ



# Cómputo en AWS

## Escalabilidad frente Alta disponibilidad

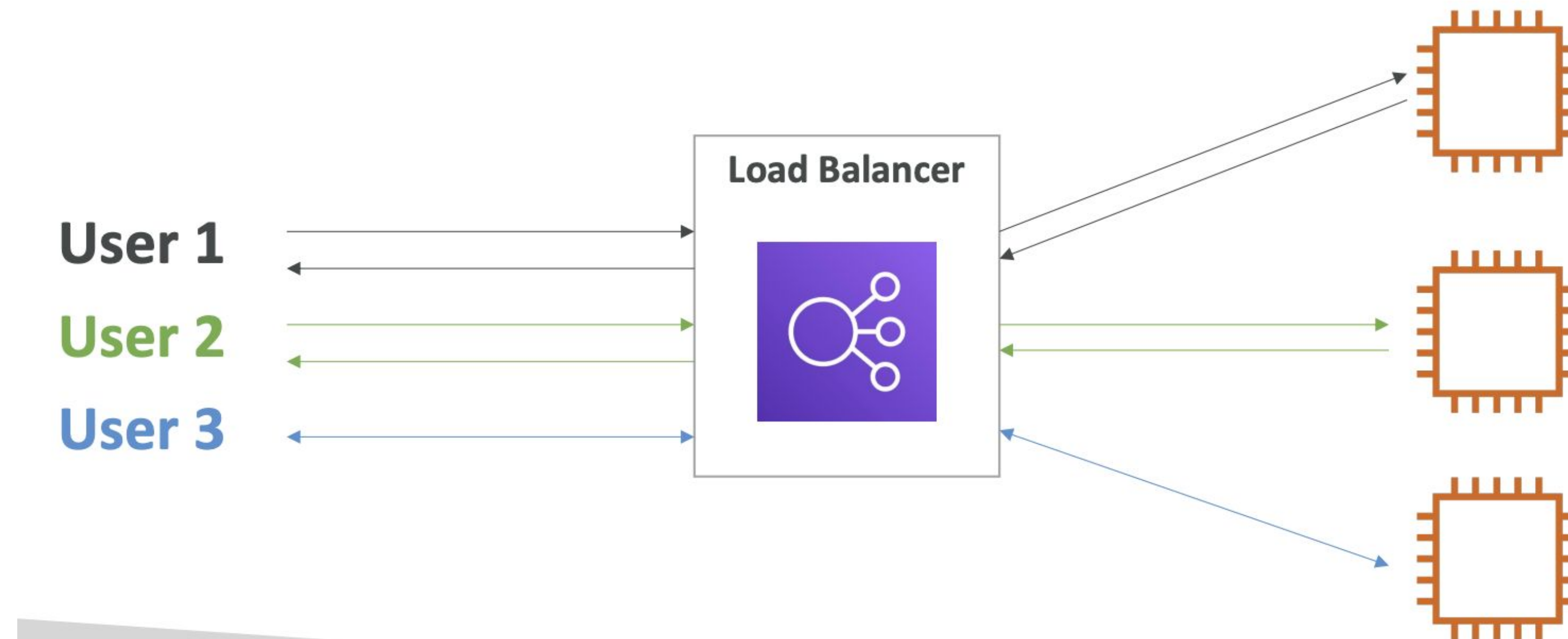
- Escalabilidad: capacidad de acomodar una carga más grande fortaleciendo el hardware (escalamiento vertical) o agregando nodos (escalamiento horizontal)
- Elasticidad: una vez que un sistema es escalable, la elasticidad significa que habrá algo de "escalado automático" para que el sistema pueda escalar en función de la carga. Esto es "amigable con la nube": pago por uso, satisfacer la demanda, optimizar costos
- Agilidad: (no relacionado con la escalabilidad – distractor) los nuevos recursos de TI están a solo un clic de distancia, lo que significa que reduce el tiempo para que esos recursos estén disponibles para sus desarrolladores de semanas a solo minutos.

# Cómputo en AWS

## Balanceo de Cargas



- Los balanceadores de carga son servidores que reenvían el tráfico de Internet a varios servidores (instancias EC2) en sentido descendente.



# Cómputo en AWS

## ¿Porque balancear?

- Distribuya la carga entre varias instancias posteriores
- Exponga un único punto de acceso (DNS) a su aplicación
- Manejar sin problemas las fallas de las instancias posteriores
- Realice comprobaciones periódicas del estado de sus instancias
- Proporcione terminación SSL (HTTPS) para sus sitios web
- Alta disponibilidad en todas las zonas

# Cómputo en AWS

## Elastic Load Balancers

- Un ELB (Elastic Load Balancer) es un balanceador de carga administrado
  - AWS garantiza que funcionará
  - AWS se encarga de las actualizaciones, el mantenimiento y la alta disponibilidad
  - AWS proporciona sólo unas pocas perillas de configuración
- Cuesta menos configurar su propio balanceador de carga, pero requerirá mucho más esfuerzo de su parte (mantenimiento, integraciones)
- 4 tipos de balanceadores de carga ofrecidos por AWS:
  - Balanceador de carga de aplicaciones (solo HTTP/HTTPS): capa 7
  - Balanceador de carga de red (alto rendimiento, permite TCP) – Capa 4
  - Balanceador de carga de puerta de enlace: capa 3
  - Classic Load Balancer (retirado en 2023) – Capa 4 y 7



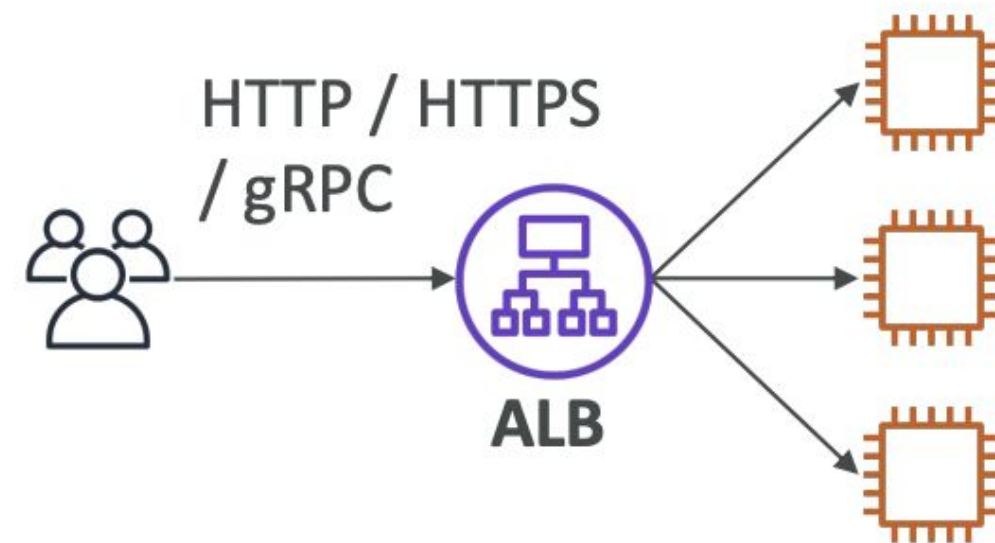
# Cómputo en AWS

## ELB – Comparativa

### Application Load Balancer



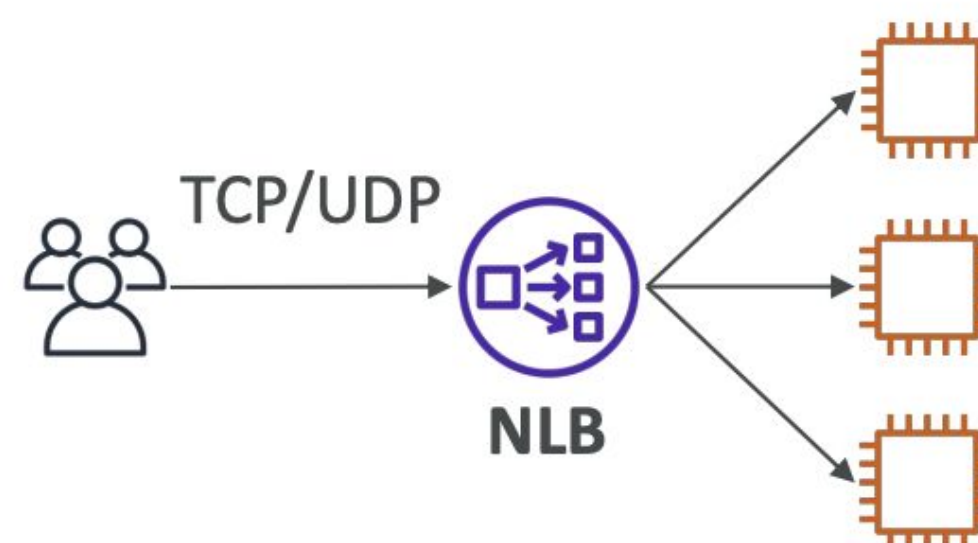
- HTTP / HTTPS / gRPC protocols (Layer 7)
- HTTP Routing features
- Static DNS (URL)



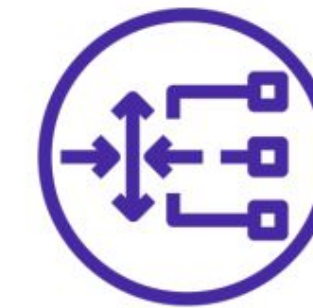
### Network Load Balancer



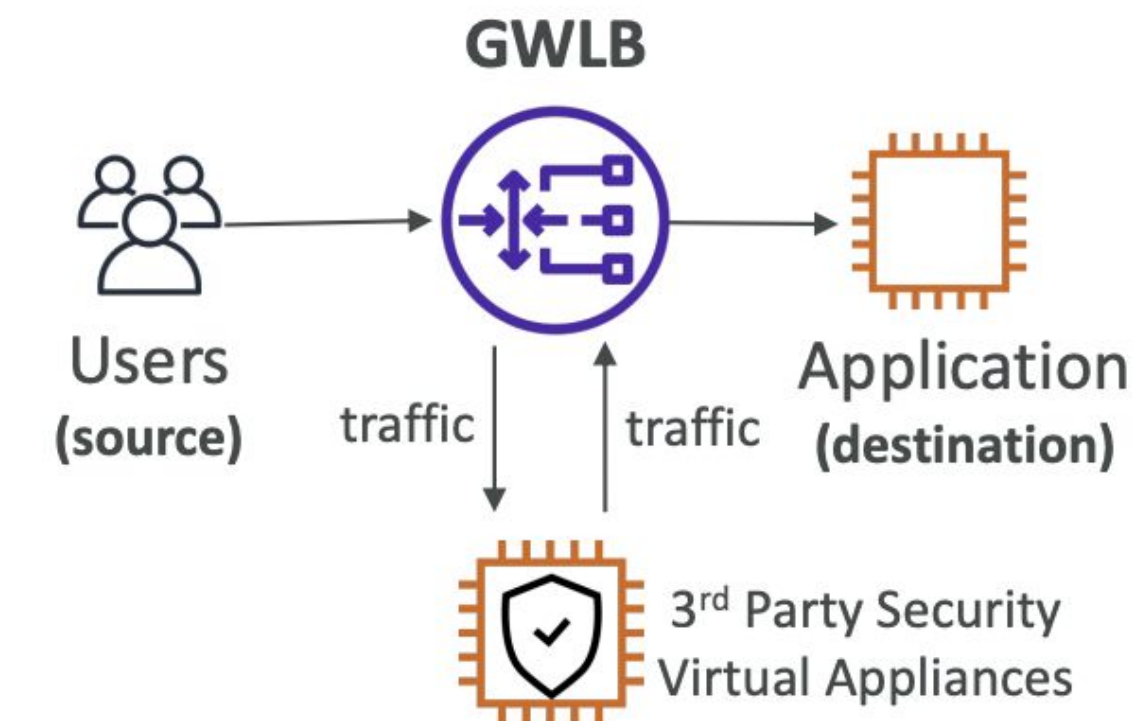
- TCP / UDP protocols (Layer 4)
- High Performance: millions of request per seconds
- Static IP through Elastic IP



### Gateway Load Balancer

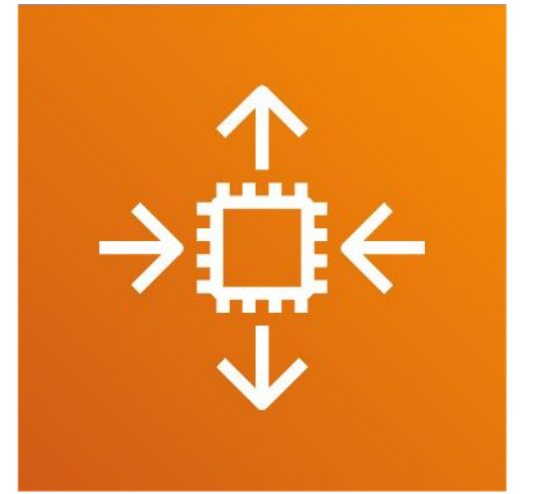


- GENEVE Protocol on IP Packets (Layer 3)
- Route Traffic to Firewalls that you manage on EC2 Instances
- Intrusion detection



# Cómputo en AWS

## Autoscaling Intro

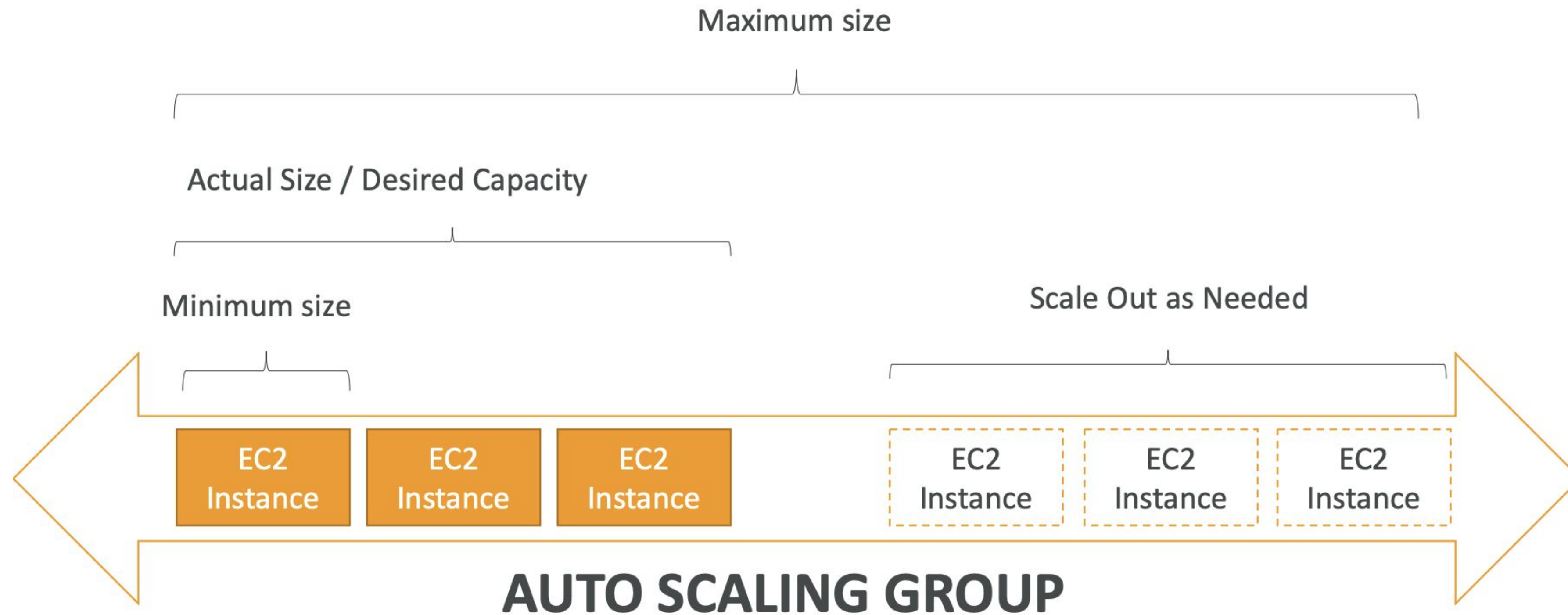


- En la vida real, la carga en sus sitios web y aplicaciones puede cambiar
- En la nube, puedes crear y deshacerte de servidores muy rápidamente
- El objetivo de un Grupo de Auto Scaling (ASG) es:
  - Escale horizontalmente (agregue instancias EC2) para adaptarse a una mayor carga
  - Escale (elimine las instancias EC2) para que coincida con una carga reducida
  - Asegurarnos de que tenemos un número mínimo y máximo de máquinas en funcionamiento
  - Registre automáticamente nuevas instancias en un balanceador de carga
  - Reemplazar instancias en mal estado
- Ahorro de costes: solo se ejecuta a una capacidad óptima (principio de la nube)



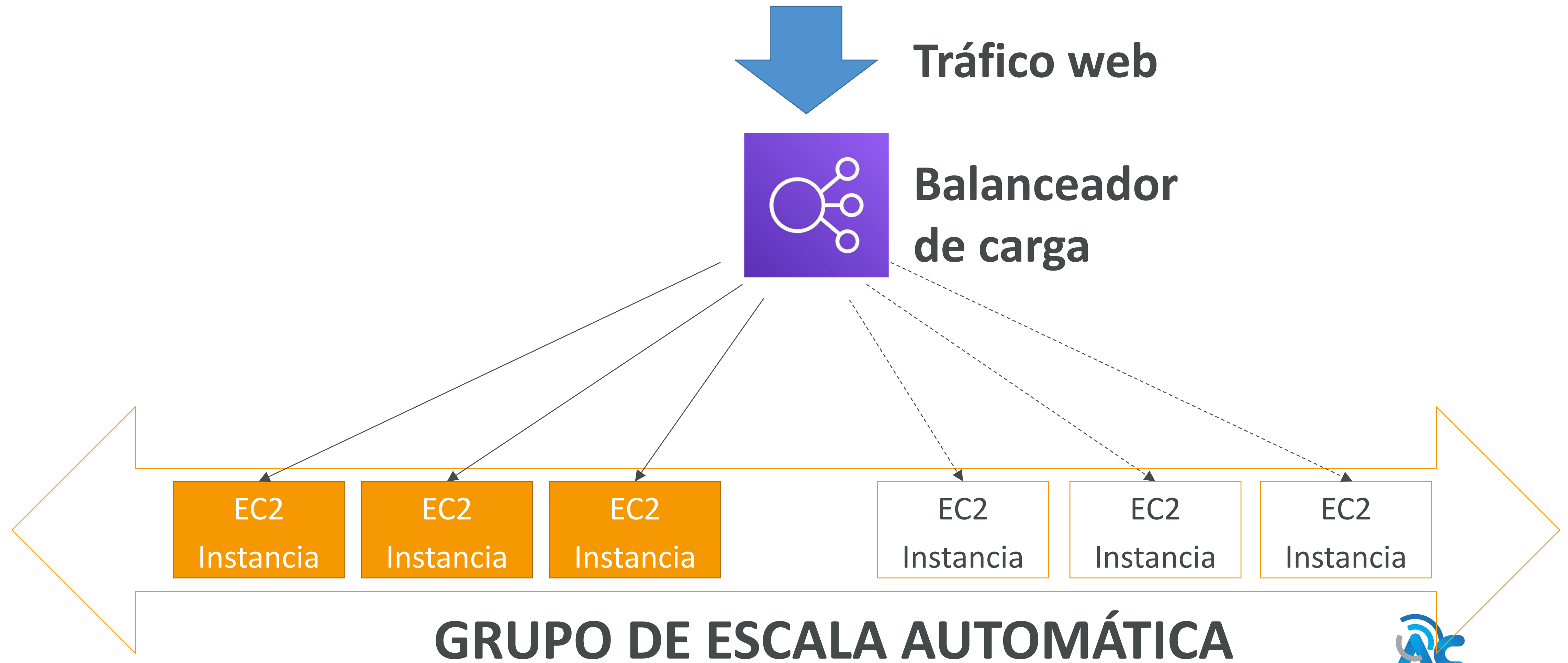
# Cómputo en AWS

## Grupos de Autoescalado



# Cómputo en AWS

## Autoescalado con Balanceadores





# Cómputo en AWS

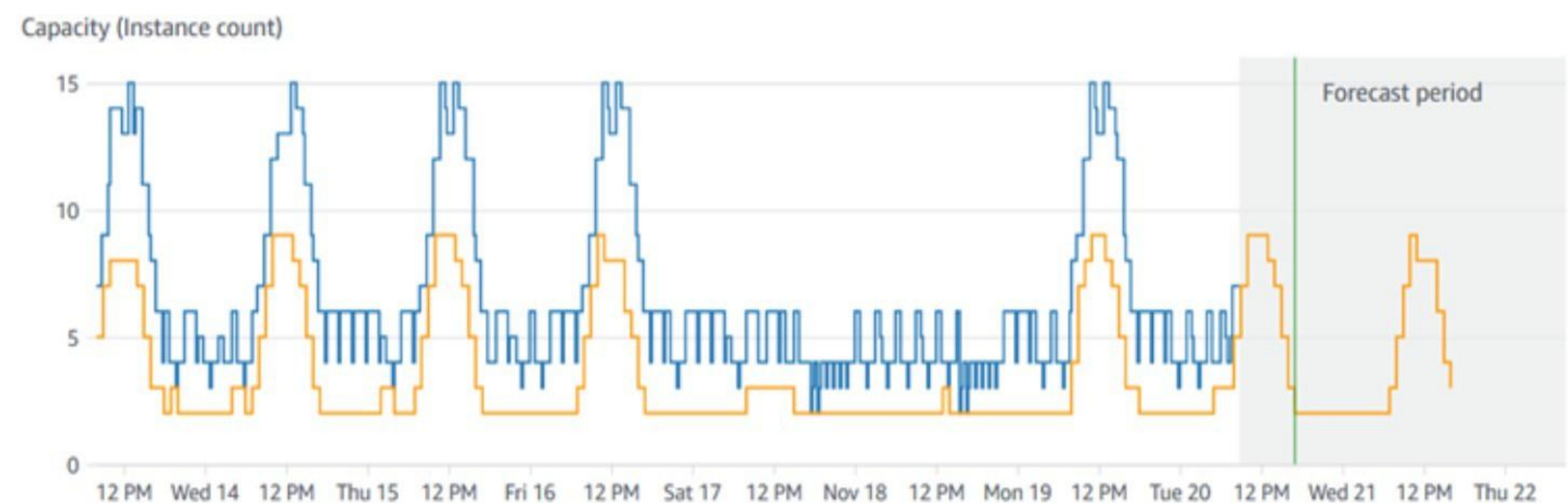
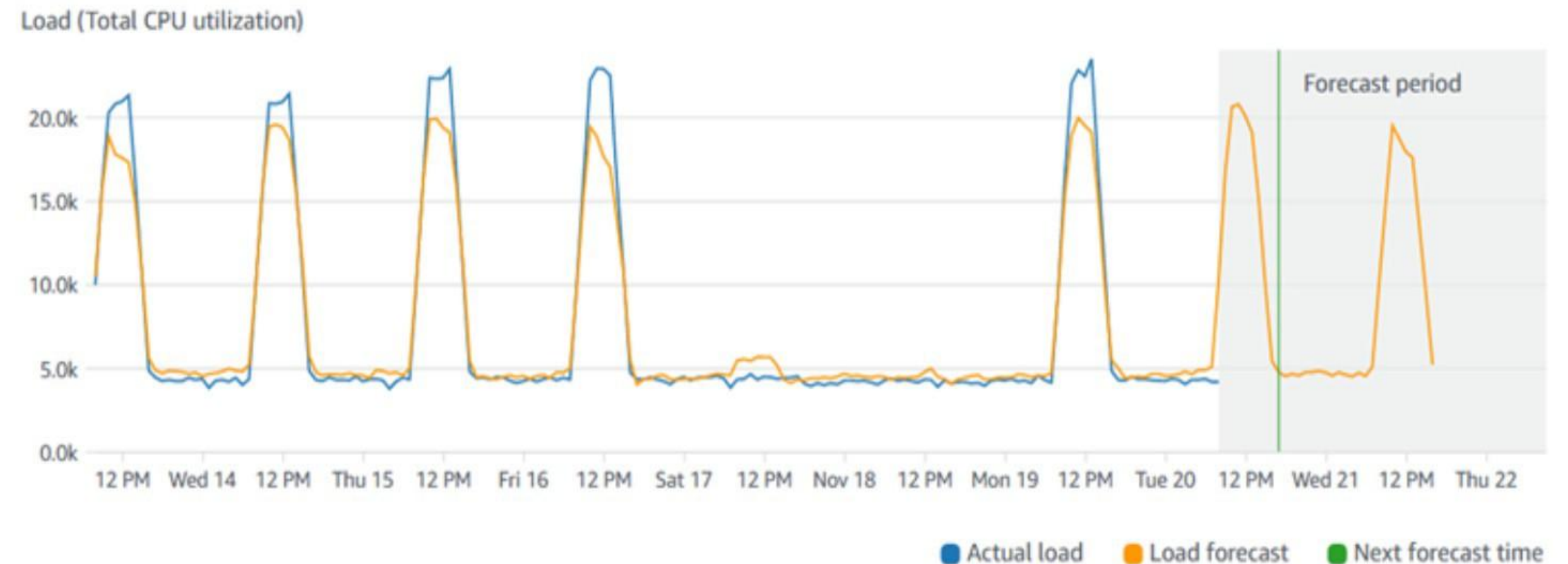
## Estrategias de escalabilidad

- Escalado manual: actualice el tamaño de un ASG manualmente
- Escalado dinámico: responda a la demanda cambiante
- Escalado simple / escalonado
  - Cuando se activa una alarma de CloudWatch (por ejemplo, CPU > 70 %), agregue 2 unidades
  - Cuando se activa una alarma de CloudWatch (por ejemplo, CPU < 30 %), elimine 1
- Escalado de seguimiento de objetivos
  - Ejemplo: quiero que la CPU ASG promedio se mantenga en alrededor del 40 %
- Escalado programado
  - Anticipar una escala basada en patrones de uso conocidos
  - Ejemplo: aumentar el min. de instancias a 10 a las 17 h los viernes

# Cómputo en AWS

## Estrategias de escalabilidad

- Escalado predictivo
  - Utiliza el aprendizaje automático para predecir el tráfico futuro con anticipación
  - Ajusta automáticamente el número correcto de instancias EC2 por adelantado
- Útil cuando su carga tiene patrones predecibles basados en el tiempo



# Cómputo en AWS

## Autoescalado y Balanceo – Resumen

- Alta Disponibilidad y Escalabilidad (vertical y horizontal) vs Elasticidad y Agilidad en la Nube
- Balanceadores de carga elásticos (ELB)
  - Distribuya el tráfico a través de instancias EC2 de back-end, puede ser Multi-AZ
  - Admite controles de salud (Health Checks)
  - 4 tipos: Clásico (antiguo), Aplicación (HTTP – L7), Red (TCP – L4), Puerta de enlace (L3)
- Grupos de Auto Scaling (ASG)
  - Implemente Elasticity para su aplicación, a través de múltiples AZ
  - Escale las instancias EC2 en función de la demanda de su sistema, reemplace las que no estén en buen estado
  - Integrado con el ELB



# Contacto

---

[achacon@consultec-ti.com](mailto:achacon@consultec-ti.com)

 [info@consultec-ti.com](mailto:info@consultec-ti.com)

 [@consulteclatam](https://www.instagram.com/consulteclatam)

 [@consultec-ti](https://www.linkedin.com/company/consultec-ti)

 [consultec-ti.com](https://www.consultec-ti.com)





# Gracias

¡Nos vemos pronto!