

AWS Certified Solutions Architect Associate

Por Stéphane Maarek y Joan Amengual



CURSO →



EXÁMENES PRÁCTICOS →

Descargo de responsabilidad: **Estas diapositivas están protegidas por derechos de autor y son estrictamente para uso personal**

- Este documento está reservado a las personas inscritas en el curso [AWS Solutions Architect Associate](#)
- Por favor, no compartas este documento, está destinado únicamente a uso personal y a la preparación de exámenes, gracias.
- Si has obtenido estas diapositivas de forma gratuita en un sitio web que no es el del curso, por favor, ponte en contacto con joan@blockstellart.com. ¡Gracias!
- ¡Mucho suerte para el examen y feliz aprendizaje!

Curso AWS Certified Solutions Architect Associate

¡Bienvenidos! Empezamos en 5 minutos



- Vamos a preparar el examen de Arquitecto de Soluciones de AWS
- Es una certificación exigente, por lo que este curso será largo e interesante
- Es necesario tener conocimientos básicos de IT
- Este curso contiene vídeos...
 - Del curso Cloud Practitioner, Developer y SysOps - conocimientos compartidos
 - Específicos para el examen de Arquitecto de Soluciones - ¡excelentes sobre arquitectura!
- Cubriremos más de 30 servicios de AWS
- ¡Los principiantes de AWS / IT son bienvenidos! (pero tómate tu tiempo, no es una carrera)

Sobre nosotros

- **¡Stephane Maarek!**
- Ha trabajado como consultor de IT y arquitecto de soluciones de AWS, desarrollador y SysOps
- Ha trabajado con AWS durante muchos años: ha construido sitios web, aplicaciones, plataformas de streaming
- Instructor veterano en AWS (Certificaciones, CloudFormation, Lambda, EC2...)
- Puedes encontrar a Stephane en:
 - GitHub: <https://github.com/simplesteph>
 - LinkedIn: <https://www.linkedin.com/in/stephanemaarek>
 - Medium: <https://medium.com/@stephane.maarek>
 - Twitter: <https://twitter.com/stephanemaarek>



Sobre nosotros

- **¡Joan Amengual!**
- Ingeniero Full Stack en una empresa tecnológica en Silicon Valley, USA
- He trabajado con AWS varios años en diversas empresas para la migración y el escalado de servicios en el Cloud
- Instructor en Blockchain y Amazon Web Services (AWS)
- Puedes encontrarme en:
 - LinkedIn: <https://www.linkedin.com/in/joanamengual7>
 - Twitter: <https://twitter.com/joanamengual7>
 - Instagram: <https://www.instagram.com/joanamengual7>
 - Youtube: <https://www.youtube.com/@joanamengual7>



¿Qué es AWS?



- AWS (Amazon Web Services) es un proveedor de Cloud
- Te proporcionan servidores y servicios que puedes utilizar bajo demanda y escalar fácilmente
- AWS ha revolucionado la IT a lo largo del tiempo
- AWS impulsa algunos de los mayores sitios web del mundo
 - Amazon.com
 - Netflix

Lo que aprenderemos en este curso (¡y más!)



Amazon EC2



Amazon ECR



Amazon ECS



AWS Elastic Beanstalk



AWS Lambda



Auto Scaling



IAM



AWS KMS



Amazon S3



Amazon SES



Amazon RDS



Amazon Aurora



Amazon DynamoDB



Amazon ElastiCache



Amazon SQS



Amazon SNS



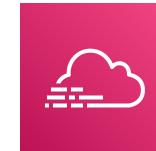
AWS Step Functions



Amazon CloudWatch



AWS CloudFormation



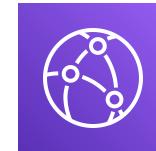
AWS CloudTrail



Amazon API Gateway



Elastic Load Balancing



Amazon CloudFront



Amazon Kinesis



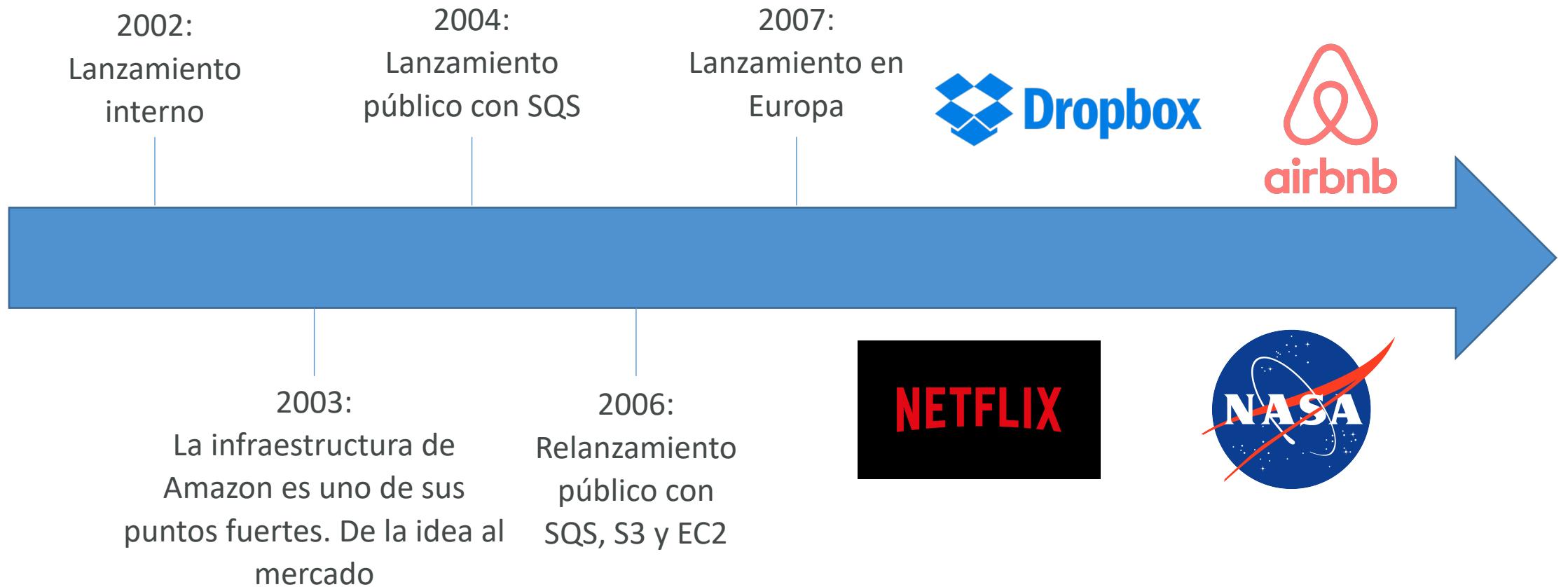
Amazon Route 53

Navegando por el plato de espaguetis de AWS



Cómo empezar con AWS

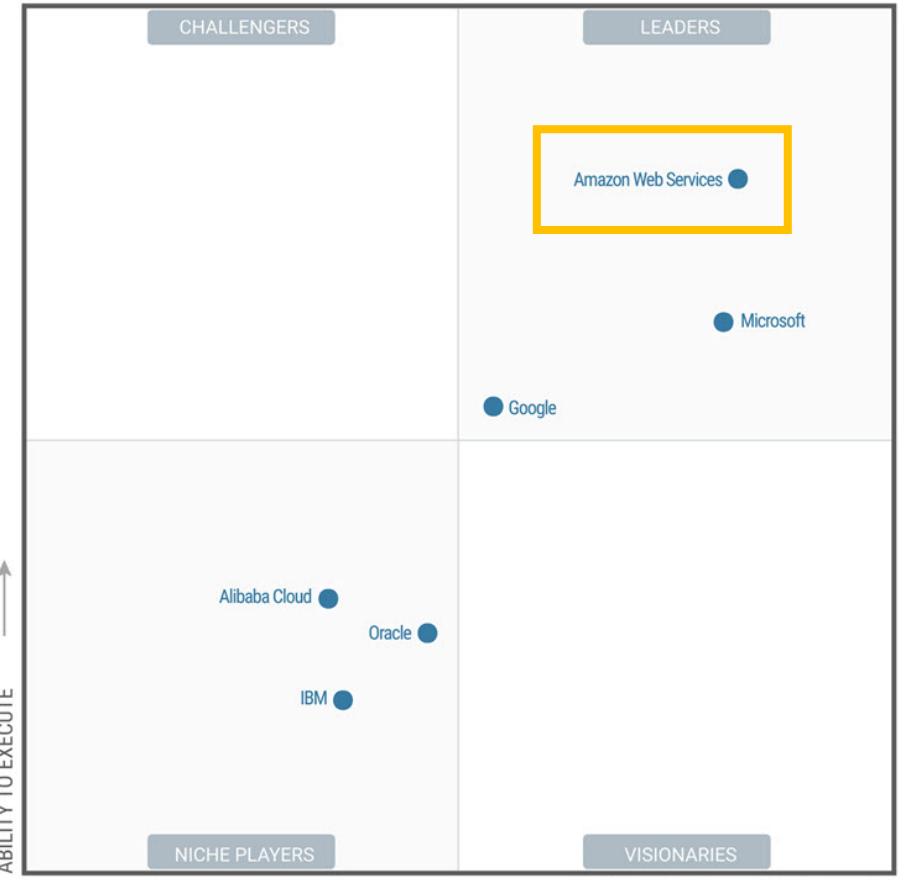
Historia del Cloud de AWS



Números de AWS Cloud

- En 2019, AWS tuvo 35.020 millones de dólares de ingresos anuales
- AWS representa el 47% del mercado en 2019 (Microsoft es el segundo con el 22%)
- Pionero y líder del mercado de Cloud de AWS por noveno año consecutivo
- Más de 1.000.000 de usuarios activos

Figure 1. Magic Quadrant for Cloud Infrastructure as a Service, Worldwide



Cuadrante mágico de Gartner

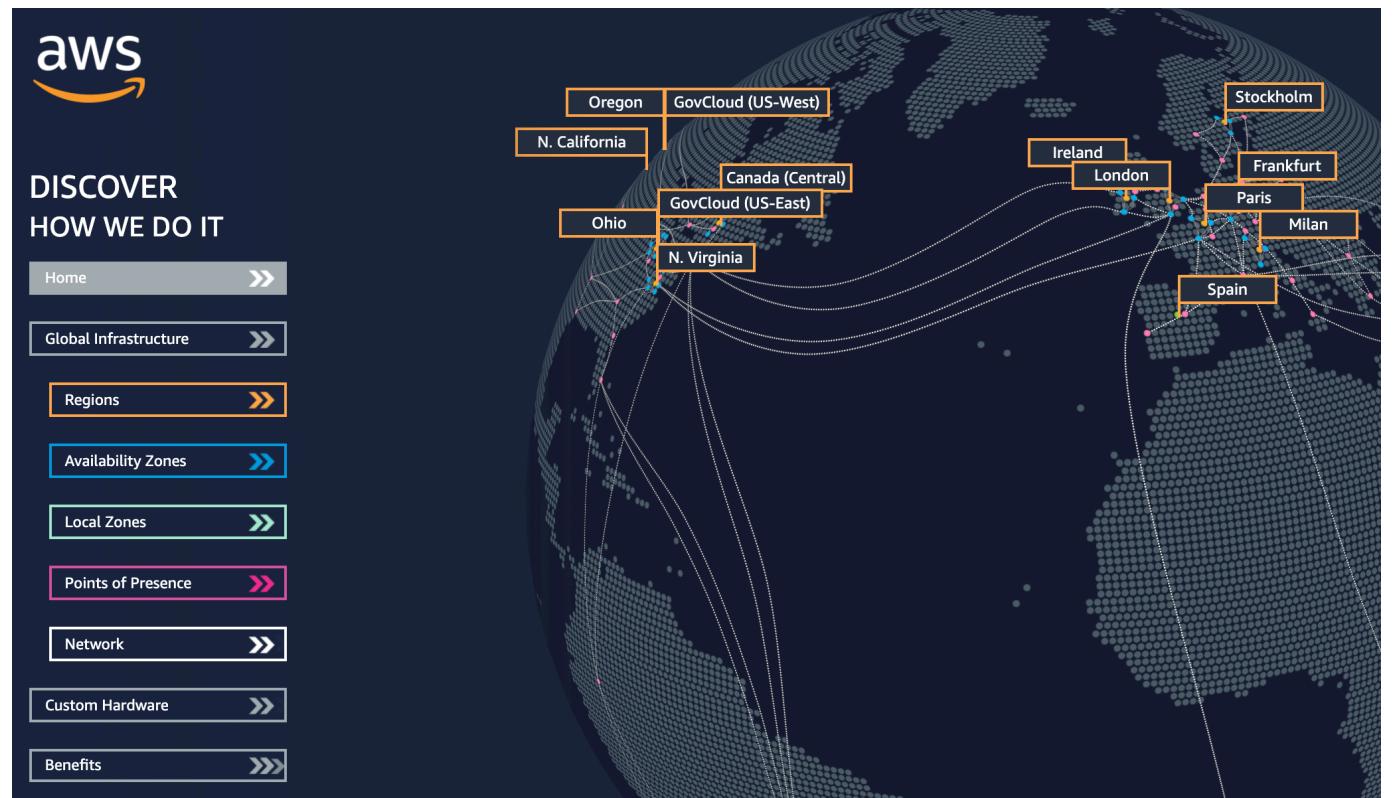
Casos de uso del Cloud de AWS

- AWS permite crear aplicaciones sofisticadas y escalables
- Aplicable a un conjunto diverso de industrias
- Los casos de uso incluyen
 - Enterprise IT, copias de seguridad y almacenamiento, análisis de Big Data
 - Alojamiento de sitios web, aplicaciones móviles y sociales
 - Juegos



Infraestructura global de AWS

- AWS Regions
- Regiones de AWS
- AWS Availability Zones
- Zonas de disponibilidad de AWS
- AWS Data Centers
- Centros de datos de AWS
- AWS Edge Locations / Points of Presence
- Puntos de presencia de AWS
- <https://infrastructure.aws/>



Regiones de AWS

- AWS tiene **Regiones** en todo el mundo
- Los nombres pueden ser us-east-1, eu-west-3...
- Una región es un **grupo de centros de datos**
- **La mayoría de los servicios de AWS son de ámbito regional**



<https://aws.amazon.com/about-aws/global-infrastructure/>

US East (N. Virginia) us-east-1

US East (Ohio) us-east-2

US West (N. California) us-west-1

US West (Oregon) us-west-2

Africa (Cape Town) af-south-1

Asia Pacific (Hong Kong) ap-east-1

Asia Pacific (Mumbai) ap-south-1

Asia Pacific (Seoul) ap-northeast-2

Asia Pacific (Singapore) ap-southeast-1

Asia Pacific (Sydney) ap-southeast-2

Asia Pacific (Tokyo) ap-northeast-1

Canada (Central) ca-central-1

Europe (Frankfurt) eu-central-1

Europe (Ireland) eu-west-1

Europe (London) eu-west-2

Europe (Paris) eu-west-3

Europe (Stockholm) eu-north-1

Middle East (Bahrain) me-south-1

South America (São Paulo) sa-east-1

¿Cómo elegir una región de AWS?

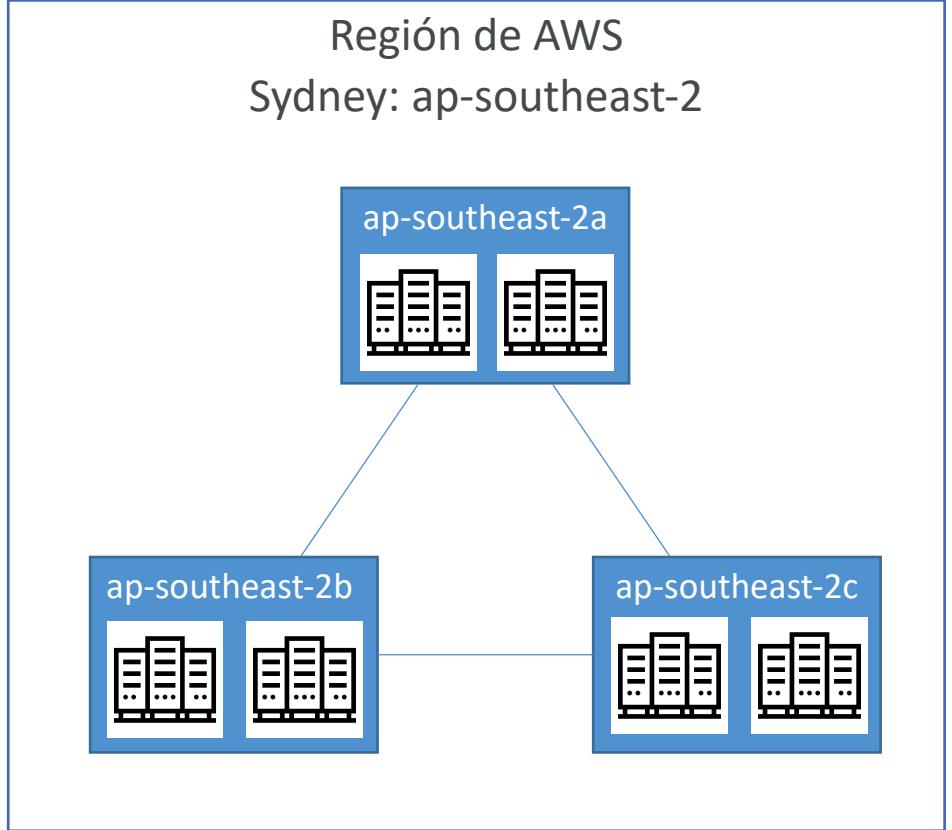
Si necesitas lanzar una nueva aplicación,
¿dónde debes hacerlo?



- **Cumplimiento de los requisitos legales y de gobernanza de datos:** los datos nunca salen de una región sin tu permiso explícito
- **Proximidad a los clientes:** latencia reducida
- **Servicios disponibles en una región:** los nuevos servicios y las nuevas funciones no están disponibles en todas las regiones
- **Precios:** los precios varían de una región a otra y son transparentes en la página de precios del servicio

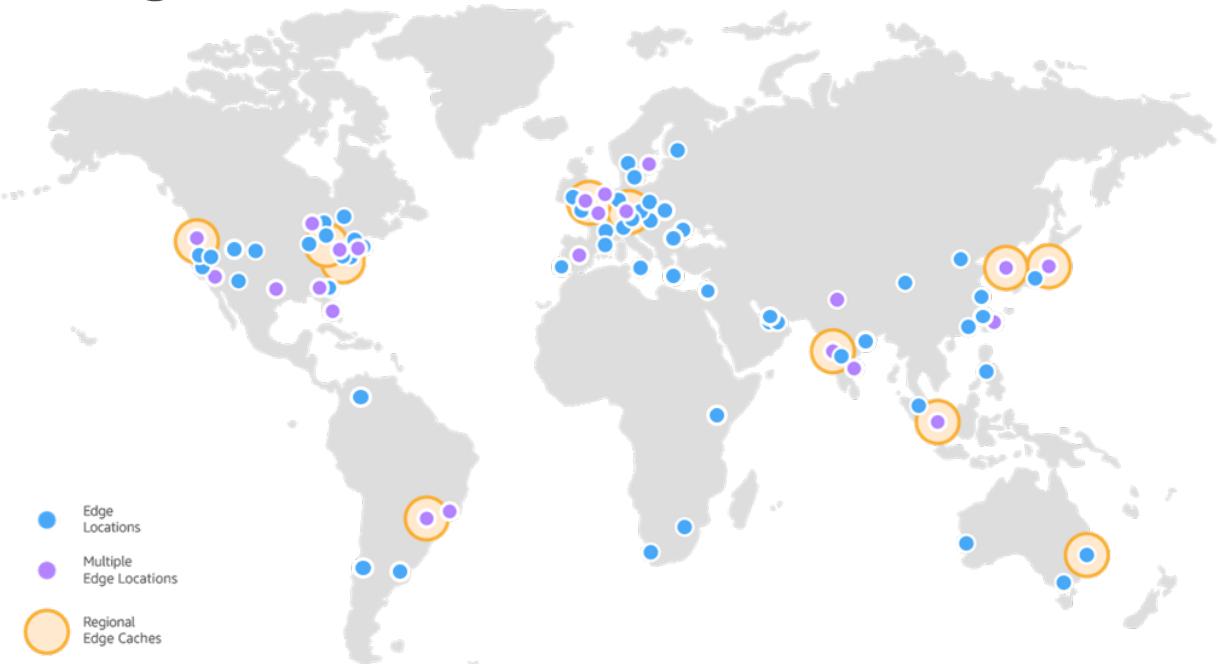
Zonas de disponibilidad de AWS

- Cada región tiene muchas zonas de disponibilidad (normalmente 3, el mínimo es 3, el máximo es 6).
Ejemplo:
 - ap-sudeste-2a
 - ap-sudeste-2b
 - ap-sudeste-2c
- Cada zona de disponibilidad (AZ) es uno o varios centros de datos discretos con alimentación, red y conectividad redundantes
- Están separadas unas de otras, de modo que están aisladas de las catástrofes
- Están conectadas con redes de alto ancho de banda y latencia ultrabaja



Puntos de presencia de AWS

- Amazon tiene 216 puntos de presencia (205 puntos de presencia y 11 cachés regionales) en 84 ciudades de 42 países
- El contenido se entrega a los usuarios finales con menor latencia



<https://aws.amazon.com/cloudfront/features/>

Tour por la consola de AWS



- **AWS cuenta con servicios globales:**

- Identity and Access Management (IAM)
- Route 53 (servicio DNS)
- CloudFront (Red de entrega de contenido)
- WAF (Firewall de aplicaciones web)



- **La mayoría de los servicios de AWS son de ámbito regional:**

- Amazon EC2 (Infraestructura como servicio)
- Elastic Beanstalk (Plataforma como servicio)
- Lambda (Función como servicio)
- Rekognition (Software como servicio)



- **Tabla de regiones:** <https://aws.amazon.com/about-aws/global-infrastructure/regional-product-services>

IAM

IAM: Usuarios y Grupos



- IAM = Identity and Access Management, servicio **global**
- **Cuenta root / raíz** creada por defecto, no debe ser utilizada ni compartida
- Los **usuarios** son personas dentro de tu organización, y pueden ser agrupados
- Los **grupos** sólo contienen usuarios, no otros grupos
- Los usuarios no tienen que pertenecer a un grupo, y el usuario puede pertenecer a varios grupos



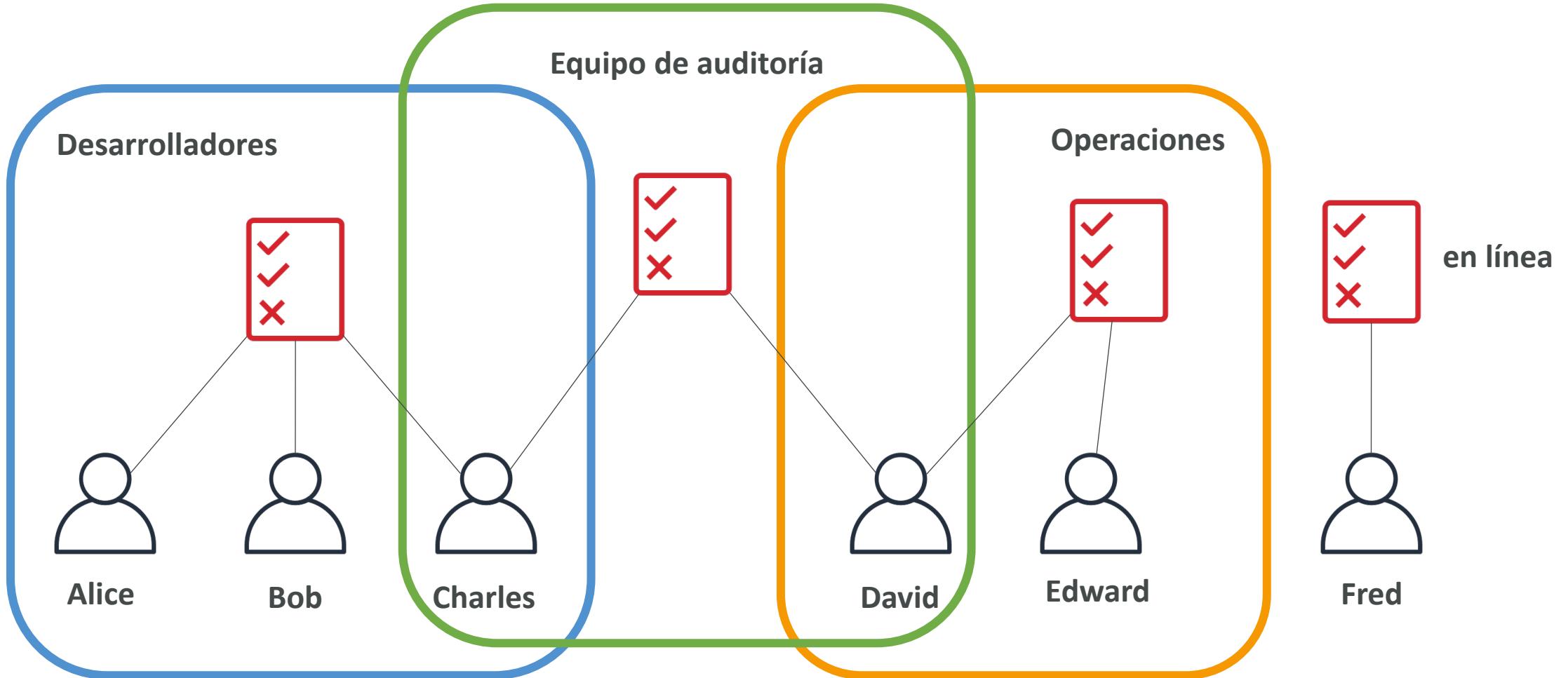
IAM: Permisos

- A los **usuarios o grupos** se les pueden asignar documentos JSON llamados políticas
- Estas políticas definen los **permisos** de los usuarios
- En AWS se aplica el **principio de mínimo privilegio**: no dar más permisos de los que un usuario necesita

```
{  
    "Version": "2012-10-17",  
    "Statement": [  
        {  
            "Effect": "Allow",  
            "Action": "ec2:Describe*",  
            "Resource": "*"  
        },  
        {  
            "Effect": "Allow",  
            "Action": "elasticloadbalancing:Describe*",  
            "Resource": "*"  
        },  
        {  
            "Effect": "Allow",  
            "Action": [  
                "cloudwatch>ListMetrics",  
                "cloudwatch:GetMetricStatistics",  
                "cloudwatch:Describe*"  
            ],  
            "Resource": "*"  
        }  
    ]  
}
```



Herencia de políticas IAM



Estructura de las políticas IAM

- Consta de:
 - **Version:** versión del lenguaje de la política, siempre incluye "2012-10-17"
 - **Id:** un identificador para la política (opcional)
 - **Statement:** una o más declaraciones individuales (obligatorio)
- Las declaraciones constan de:
 - **Sid:** un identificador para la declaración (opcional)
 - **Effect:** si la sentencia permite o deniega el acceso (Permitir, Denegar)
 - **Principal:** cuenta/usuario/rol al que se aplica esta política
 - **Action:** lista de acciones que esta política permite o deniega
 - **Resource:** lista de recursos a los que se aplican las acciones
 - **Condition:** condiciones para cuando esta política está en efecto (opcional)

```
{  
  "Version": "2012-10-17",  
  "Id": "S3-Account-Permissions",  
  "Statement": [  
    {  
      "Sid": "1",  
      "Effect": "Allow",  
      "Principal": {  
        "AWS": ["arn:aws:iam::123456789012:root"]  
      },  
      "Action": [  
        "s3:GetObject",  
        "s3:PutObject"  
      ],  
      "Resource": ["arn:aws:s3:::mybucket/*"]  
    }  
  ]  
}
```

IAM - Política de contraseñas

- Contraseñas fuertes = mayor seguridad para tu cuenta
- En AWS, puedes configurar una política de contraseñas:
 - Establecer una longitud mínima de contraseña
 - Requerir tipos de caracteres específicos:
 - incluyendo letras mayúsculas
 - letras minúsculas
 - números
 - caracteres no alfanuméricos
 - Permitir a todos los usuarios de IAM cambiar sus propias contraseñas
 - Requerir a los usuarios que cambien su contraseña después de un tiempo (caducidad de la contraseña)
 - Impedir la reutilización de la contraseña

Multi Factor Authentication - MFA



- Los usuarios tienen acceso a tu cuenta y posiblemente pueden cambiar configuraciones o eliminar recursos en tu cuenta de AWS
- **Quieres proteger tus cuentas root y los usuarios de IAM**
- MFA = contraseña que conoces + dispositivo de seguridad que posees



Contraseña +



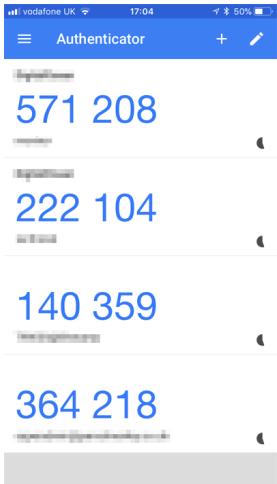
=>

Login exitoso

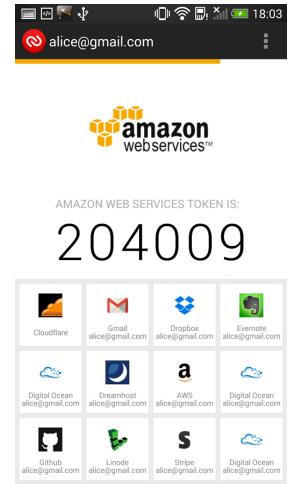
- **Principal beneficio de MFA:** si una contraseña es robada o hackeada, la cuenta no se ve comprometida

Opciones de dispositivos MFA en AWS

Dispositivo virtual MFA



Autenticador de Google
(sólo en el teléfono)



Authy
(multi-dispositivo)

Soporte para múltiples tokens en un solo dispositivo.

Clave de seguridad del segundo factor universal (U2F)

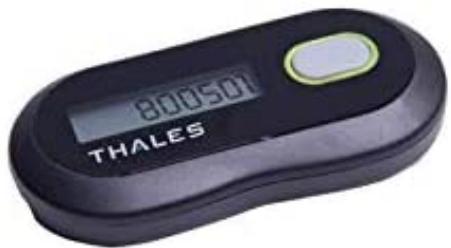


YubiKey de Yubico (3rd party)

Soporte para múltiples usuarios root e IAM utilizando una única clave de seguridad

Opciones de dispositivos MFA en AWS

Dispositivo MFA de llavero por hardware



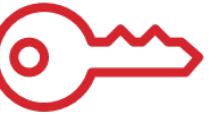
Proporcionado por Gemalto (3rd party)

Dispositivo MFA de llavero por hardware para AWS GovCloud (US)



Proporcionado por SurePassID (3rd party)

¿Cómo pueden los usuarios acceder a AWS?



- Para acceder a AWS, tienes tres opciones:
 - **Consola de administración de AWS** (protegida por contraseña + MFA)
 - **Interfaz de línea de comandos de AWS (CLI)**: protegida por claves de acceso
 - **AWS Software Developer Kit (SDK)** - para el código: protegido por claves de acceso
 - Las claves de acceso se generan a través de la consola de AWS
- Los usuarios gestionan sus propias claves de acceso
- **Las claves de acceso son secretas, como una contraseña. No las compartas**
- ID de la clave de acceso ~ = nombre de usuario
- Clave de acceso secreta ~ = contraseña

Ejemplo de claves de acceso (falsas)

Access keys

Use access keys to make secure REST or HTTP Query protocol requests to AWS service APIs. For your protection, you should never share your secret keys with anyone. As a best practice, we recommend frequent key rotation. [Learn more](#)

[Create access key](#)

Access key ID	Created	Last used	Status	
AKIASK4E37PV4TU3RD6C	2020-05-25 15:13 UTC+0100	N/A	Active	Make inactive X

- ID de la llave de acceso: AKIASK4E37PV4983d6C
- Clave de acceso secreta: AZPN3z0jWozWCndljhB0Unh8239a1bzBzO5fqkZq
- **Recuerda: no compartas tus claves de acceso**

¿Qué es la CLI de AWS?

- Una herramienta que permite interactuar con los servicios de AWS mediante comandos en tu shell de línea de comandos
- Acceso directo a las API públicas de los servicios de AWS
- Puedes desarrollar scripts para gestionar tus recursos
- Es de código abierto <https://github.com/aws/aws-cli>
- Alternativa al uso de la consola de administración de AWS

```
→ ~ aws s3 cp myfile.txt s3://ccp-mybucket/myfile.txt
upload: ./myfile.txt to s3://ccp-mybucket/myfile.txt
→ ~ aws s3 ls s3://ccp-mybucket
2021-05-14 03:22:52          0 myfile.txt
→ ~ █
```

¿Qué es el SDK de AWS?

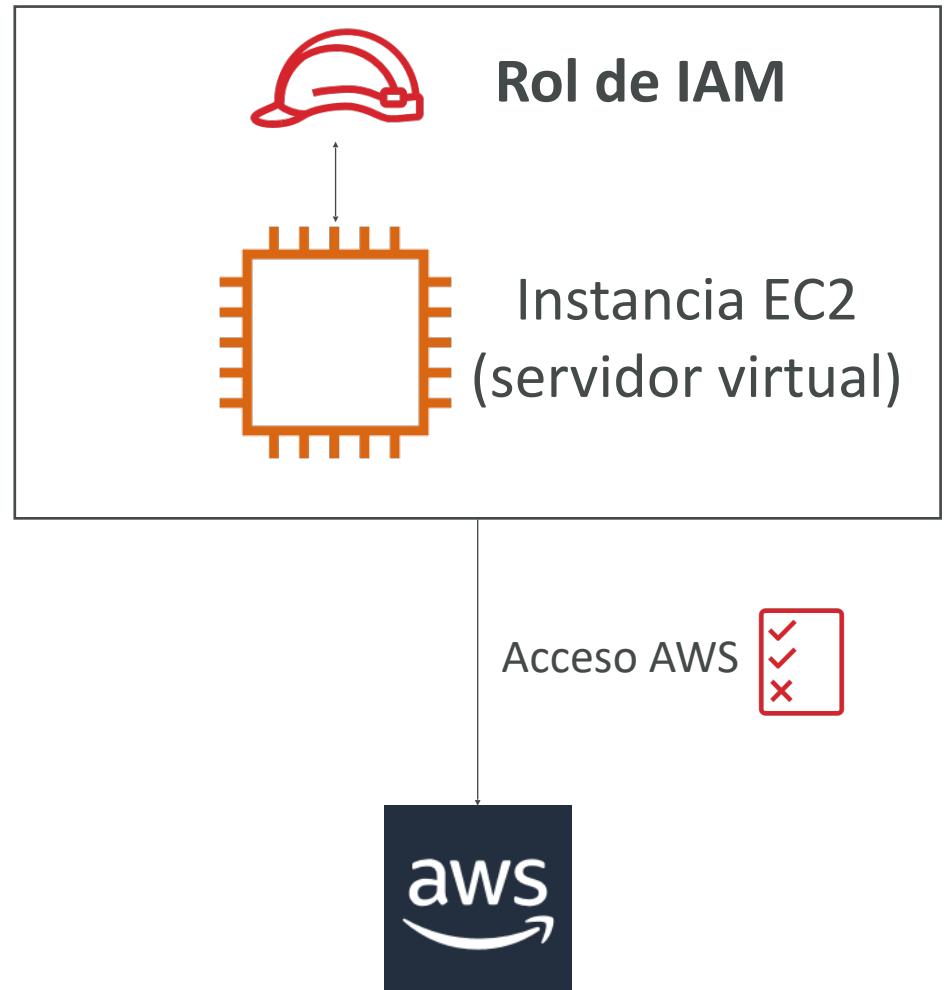


- Kit de desarrollo de software de AWS (AWS SDK)
- APIs específicas para cada lenguaje (conjunto de bibliotecas)
- Permite acceder y administrar los servicios de AWS mediante programación
- Integrado en la aplicación
- Admite:
 - SDKs (JavaScript, Python, PHP, .NET, Ruby, Java, Go, Node.js, C++)
 - SDKs para móviles (Android, iOS, ...)
 - SDKs para dispositivos IoT (Embedded C, Arduino, ...)
- Ejemplo: AWS CLI está construido sobre AWS SDK para Python



Roles IAM para los servicios

- Algún servicio de AWS tendrá que realizar acciones en tu nombre
- Para ello, asignaremos **permisos** a los servicios de AWS con **Roles IAM**
- Roles comunes:
 - Roles de Instancia EC2
 - Roles de la función Lambda
 - Roles para CloudFormation



Herramientas de seguridad IAM

- **IAM Credentials Report / Informe de credenciales de IAM (a nivel de cuenta)**
 - Un informe que enumera todos los usuarios de tu cuenta y el estado de tus diversas credenciales
- **IAM Access Advisor / Asesor de acceso de IAM (a nivel de usuario)**
 - Muestra los permisos de servicio concedidos a un usuario y cuando se accedió a esos servicios por última vez
 - Puedes utilizar esta información para revisar tus políticas

Directrices y buenas prácticas de IAM



- No utilices la cuenta root excepto para la configuración de la cuenta AWS
- Un usuario físico = Un usuario AWS
- **Asignar usuarios a grupos** y asignar permisos a grupos
- Crear una **política de contraseñas fuerte**
- Utilizar y reforzar el uso de la **autenticación multifactor (MFA)**
- Crear y utilizar **Roles** para dar permisos a los servicios de AWS
- Utilizar claves de acceso para el acceso programático (CLI / SDK)
- Revisar los permisos de tu cuenta con el informe de credenciales de IAM
- **No compartir nunca los usuarios de IAM ni las claves de acceso**

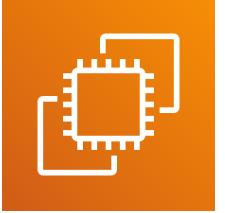
Resumen - IAM



- **Usuarios:** mapeado a un usuario físico, tiene una contraseña para la consola de AWS
- **Grupos:** contiene sólo usuarios
- **Políticas:** Documento JSON que describe los permisos para los usuarios o grupos
- **Roles:** para instancias EC2 o servicios AWS
- **Seguridad:** MFA + Política de contraseñas
- **AWS CLI:** gestiona tus servicios de AWS mediante la línea de comandos
- **AWS SDK:** gestiona tus servicios de AWS utilizando un lenguaje de programación
- **Claves de acceso:** accede a AWS mediante la CLI o el SDK
- **Auditoría:** Informes de credenciales de IAM y Asesor de acceso de IAM

EC2

Amazon EC2



- EC2 es una de las ofertas más populares de AWS
- EC2 = Elastic Compute Cloud = Infraestructura como servicio (IaaS)
- Consiste principalmente en la capacidad de :
 - Alquilar máquinas virtuales (EC2)
 - Almacenar datos en unidades virtuales (EBS)
 - Distribuir la carga entre las máquinas (ELB)
 - Escalar los servicios mediante un Auto Scaling Group (ASG) o también conocido en español como Grupo de Autoescalado
- Conocer EC2 es fundamental para entender el funcionamiento del Cloud

Opciones de tamaño y configuración de EC2

- Sistema operativo (**OS**): Linux, Windows o Mac OS
- Cuánta potencia de cálculo 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 de arranque (configurar en el primer lanzamiento): Datos de usuario de EC2

Datos del usuario de EC2

- Es posible arrancar nuestras instancias utilizando un script de datos de usuario de EC2.
- bootstrapping significa lanzar comandos cuando una máquina se inicia
- Ese script sólo se ejecuta una vez en el primer arranque 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 se te ocurra
- El script de datos de usuario de EC2 se ejecuta con el usuario root

Lanzamiento de una instancia EC2 con Linux

- Vamos a lanzar nuestro primer servidor virtual utilizando la consola de AWS
- Tendremos una primera aproximación de alto nivel a los distintos parámetros
- Veremos que nuestro servidor web se lanza utilizando los datos de usuario de EC2
- Aprenderemos a iniciar / parar / terminar nuestra instancia.

Tipos de instancias de EC2 - Visión general

- Puedes utilizar diferentes tipos de instancias EC2 optimizadas para diferentes casos de uso (<https://aws.amazon.com/ec2/instance-types/>)
- AWS tiene la siguiente convención de nombres:

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

Tipos de instancias de EC2 - Propósito general

- Excelente para una diversidad de cargas de trabajo, como servidores web o repositorios de código
- Equilibrio entre:
 - Computación
 - Memoria
 - Red
- En el curso, utilizaremos la instancia t2.micro que es una instancia EC2 de propósito 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
-----	-----	----	-----	----	-----	----	-----	-----	------	----	----

* Esta lista evolucionará con el tiempo, por favor, consulta el sitio web de AWS para obtener la información más reciente

Tipos de instancias EC2 - Computación optimizada

- Ideal para tareas de cálculo intensivo que requieren procesadores de alto rendimiento:
 - Cargas de trabajo de procesamiento por lotes
 - Transcodificación de medios
 - Servidores web de alto rendimiento
 - Computación de alto rendimiento (HPC)
 - Modelado científico y aprendizaje automático
 - Servidores dedicados a juegos

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.

C6g C6gn C5 C5a C5n C4

* Esta lista evolucionará con el tiempo, por favor, consulta el sitio web de AWS para obtener la información más reciente

Tipos de instancias EC2 - Memoria optimizada

- Rápido rendimiento para cargas de trabajo que procesan grandes conjuntos de datos en memoria
- Casos de uso:
 - Alto rendimiento, bases de datos relacionales/no relacionales
 - Almacenes de caché distribuidos a escala web
 - Bases de datos en memoria optimizadas para BI (business intelligence)
 - Aplicaciones que realizan el 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.

R6g

R5

R5a

R5b

R5n

R4

X1e

X1

High Memory

z1d

* Esta lista evolucionará con el tiempo, por favor, consulta el sitio web de AWS para obtener la información más reciente

Tipos de instancias EC2 - Almacenamiento optimizado

- Ideal para tareas de almacenamiento intensivo que requieran un acceso alto y secuencial de lectura y escritura a grandes conjuntos de datos en el almacenamiento local
- Casos de uso:
 - Sistemas de procesamiento de transacciones en línea (OLTP) de alta frecuencia
 - 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

* Esta lista evolucionará con el tiempo, por favor, consulta el sitio web de AWS para obtener la información más reciente

Tipos de instancias de EC2: ejemplo

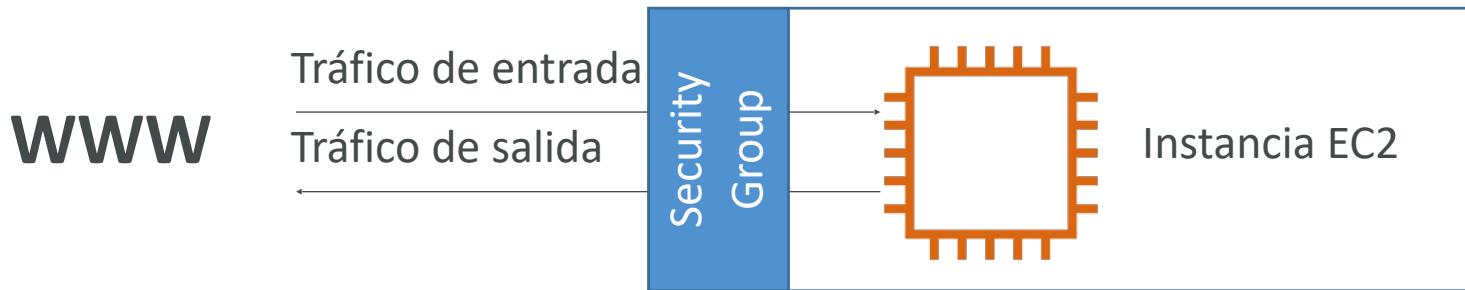
Instancia	vCPU	Mem (GiB)	Almacenamiento	Rendimiento de la red	Ancho de banda de EBS (Mbps)
t2.micro	1	1	Sólo EBS	Bajo a moderado	
t2.xlarge	4	16	Sólo EBS	Moderado	
c5d.4xlarge	16	32	1 x 400 NVMe SSD	Hasta 10 Gbps	4,750
r5.16xlarge	64	512	Sólo EBS	20 Gbps	13,600
m5.8xlarge	32	128	Sólo EBS	10 Gbps	6,800

t2.micro forma parte de la capa gratuita de AWS (hasta 750 horas al mes)

<https://instances.vantage.sh>

Introducción a los grupos de seguridad

- Los grupos de seguridad son la base de la seguridad de la red en AWS
- Controlan cómo se permite el tráfico dentro o fuera de nuestras Instancias EC2



- Los grupos de seguridad sólo contienen reglas de **permiso**
- Las reglas de los grupos de seguridad pueden hacer referencia por IP o por grupo de seguridad

Grupos de seguridad

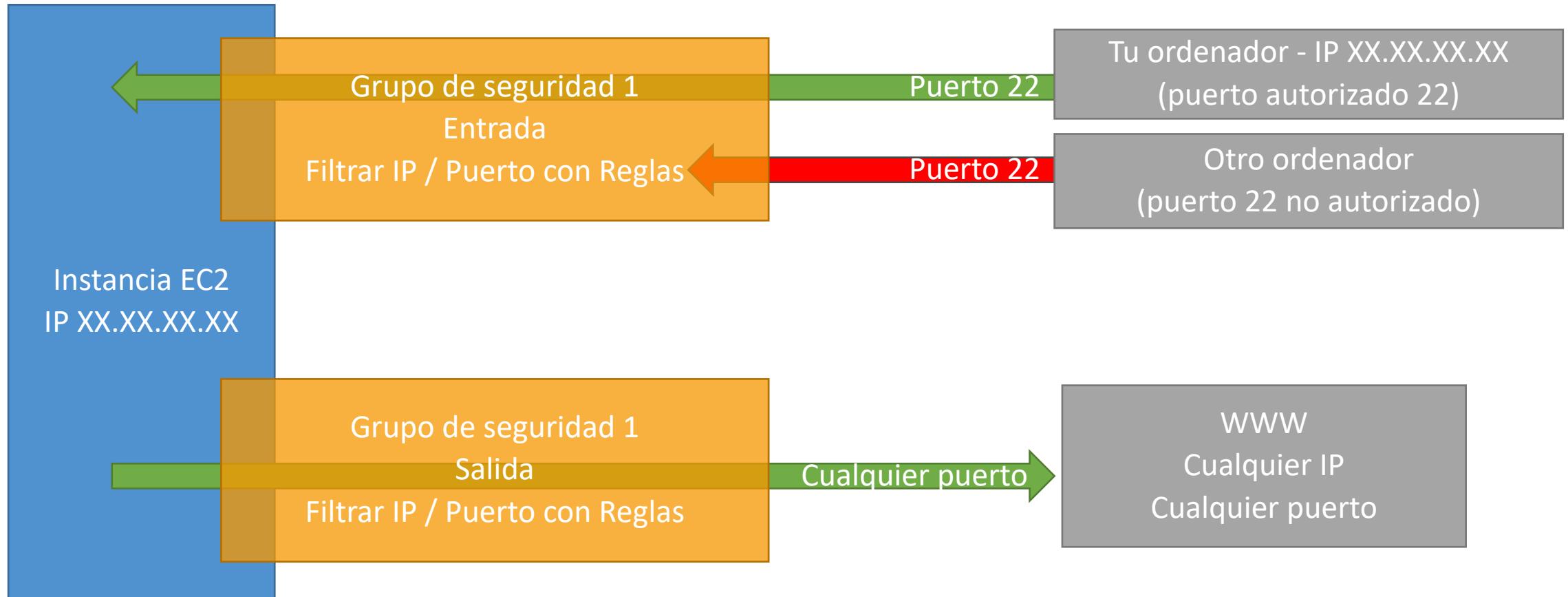
Inmersión más profunda

- Los grupos de seguridad actúan como un “firewall” en las instancias de EC2
- Regulan:
 - El acceso a los puertos
 - Rangos de IP autorizados - IPv4 e IPv6
 - Control de la red de entrada (de otros a la instancia)
 - Control de la red saliente (desde la instancia hacia 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

Grupos de seguridad

Diagrama



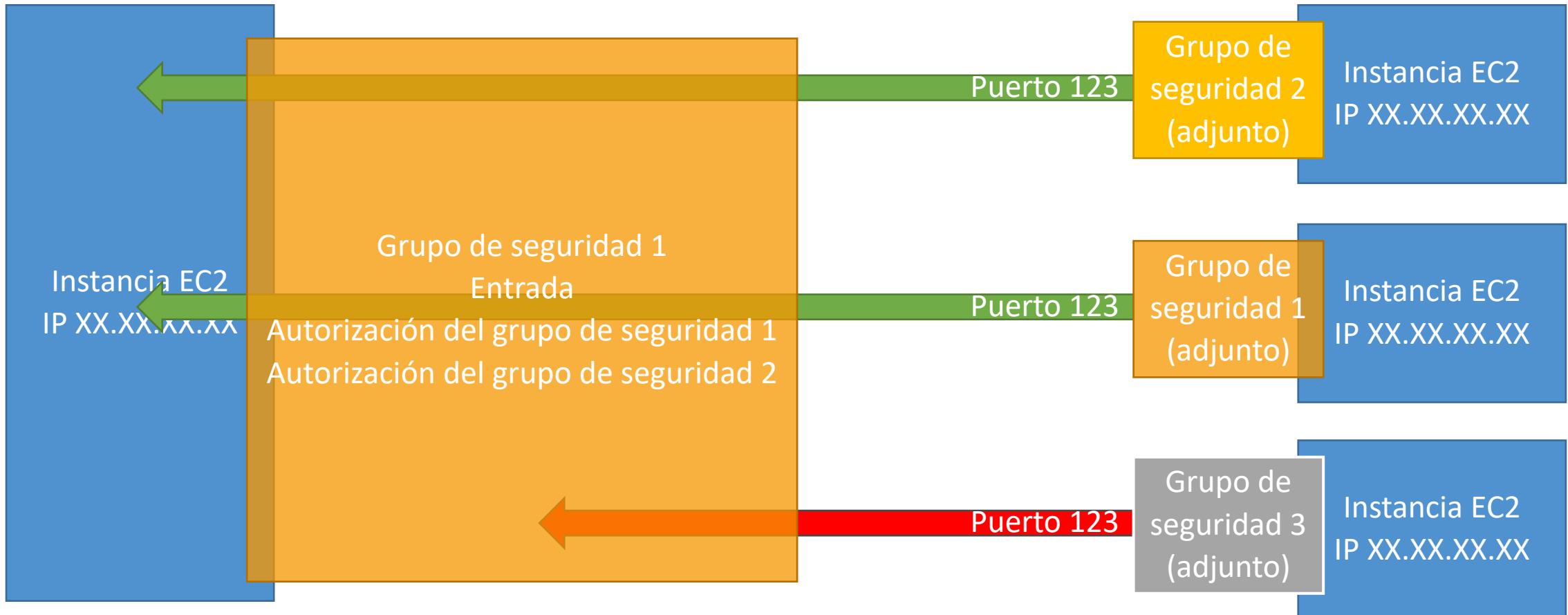
Grupos de seguridad

Es bueno saber

- Puede adjuntarse a múltiples instancias
- Bloqueado a una combinación de región / VPC
- Vive "fuera" del EC2 - si el tráfico está bloqueado, la instancia EC2 no lo verá
- Es bueno mantener un grupo de seguridad separado para el acceso SSH
- Si tu aplicación no es accesible (tiempo de espera), entonces es un problema de grupo de seguridad
- Si tu aplicación da un error de "conexión rechazada", entonces es un error de la aplicación o no se ha lanzado
- Todo el tráfico de entrada está **bloqueado** por defecto
- Todo el tráfico de salida está **autorizado** por defecto

Referencia a otros grupos de seguridad

Diagrama



Puertos clásicos que hay que conocer

- 22 = SSH (Secure Shell) - iniciar sesión en una instancia de Linux
- 21 = FTP (File Transfer Protocol) - subir archivos a un archivo compartido
- 22 = SFTP (Secure File Transfer Protocol) - subir archivos usando SSH
- 80 = HTTP - acceso a sitios web no seguros
- 443 = HTTPS - acceso a sitios web seguros
- 3389 = RDP (Remote Desktop Protocol) - iniciar sesión en una instancia de Windows

Tabla resumen SSH

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

Qué clases hay que ver

- **Mac / Linux:**

- Clase de SSH en Mac/Linux

- **Windows:**

- Clase sobre Putty
 - Si Windows 10: Clase sobre SSH en Windows 10

- **Todos los estudiantes:**

- Clase de Instance Connect EC2

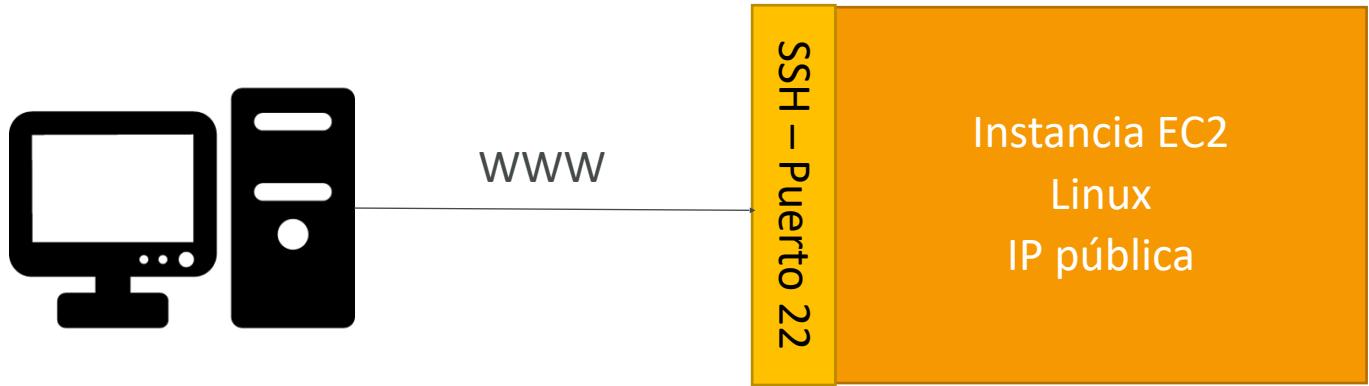
Solución de problemas de SSH

- **Los estudiantes son los que más problemas tienen con SSH**
- Si las cosas no funcionan...
 - Vuelve a ver la clase. Puede que te hayas perdido algo
 - Lee la guía de solución de problemas
 - Prueba con EC2 Instance Connect
- **Si uno de los métodos funciona (SSH, Putty o EC2 Instance Connect) estás bien**
- Si ningún método funciona, no pasa nada, el curso no utilizará mucho SSH

Cómo usar SSH en tu instancia EC2

Linux / Mac OS X

- Vamos a aprender cómo usar SSH en tu instancia EC2 usando Linux / Mac
- SSH es una de las funciones más importantes. Permite controlar una máquina remota, todo ello utilizando la línea de comandos.

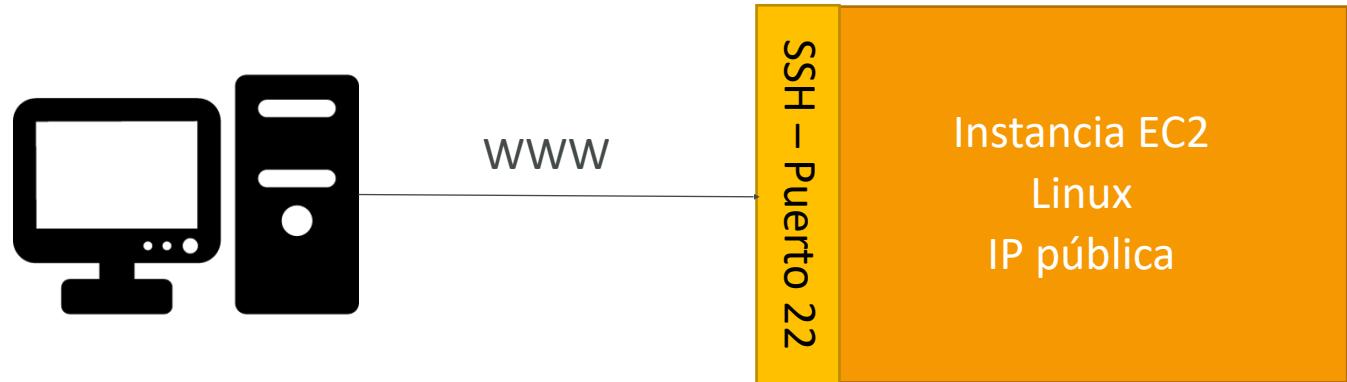


- Vamos a ver cómo podemos configurar OpenSSH `~/.ssh/config` para facilitar el SSH en nuestras instancias EC2

Cómo usar SSH en tu instancia EC2

Windows

- Vamos a aprender cómo usar SSH en tu instancia EC2 usando [Windows](#)
- SSH es una de las funciones más importantes. Permite controlar una máquina remota, todo ello utilizando la línea de comandos.



- Configuraremos todos los parámetros necesarios para hacer SSH en Windows utilizando la herramienta gratuita [Putty](#).

Instance Connect EC2

Conexión de instancias EC2

- Conéctate a tu instancia EC2 desde el navegador
- No es necesario utilizar el archivo de claves que se ha descargado
- La "magia" es que una clave temporal es cargada en EC2 por AWS
- **Funciona sólo out-of-the-box con Amazon Linux 2**
- Necesitas asegurarte de que el puerto 22 sigue abierto

Opciones de compra de instancias EC2

- **Instancias bajo demanda**: carga de trabajo corta, precio predecible, pago por segundos
- **Reservadas** (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 larga
- **Instancias Spot** - cargas de trabajo cortas, baratas, pueden perder instancias (menos fiables)
- **Hosts dedicados**: reserva un servidor físico completo, controla la ubicación de las instancias
- **Instancias dedicadas** - ningún otro cliente compartirá tu hardware
- **Reservas de capacidad** - reserva de capacidad en una AZ específica para cualquier duración

EC2 bajo demanda

- 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 coste más elevado, pero no hay que pagar por adelantado
 - Sin compromiso a largo plazo
-
- Recomendado para **cargas de trabajo a corto plazo y sin interrupciones**, cuando no se puede predecir el comportamiento de la aplicación

Instancias reservadas de EC2

- Hasta un **72%** de descuento en comparación con el servicio bajo demanda
- Reserva de atributos de instancia específicos (**tipo de instancia, región, ocupación, sistema operativo**)
- **Periodo de reserva - 1 año** (+descuento) o **3 años** (+++descuento)
- **Opciones de pago - Sin pago inicial** (+), **Pago inicial parcial** (++) , **Pago inicial total** (+++)
- **Alcance de la instancia reservada** - Por **región** o por **zona** (capacidad de reserva en una AZ)
- Recomendado para aplicaciones de uso constante (piensa en una base de datos)
- Puedes comprar y vender en el Marketplace de instancias reservadas
- **Instancia reservada convertible:**
 - Puedes cambiar el tipo de instancia EC2, la familia de instancias, el SO, etc.
 - Hasta un **66%** de descuento

Nota: los % de descuento pueden ser diferentes a los del video ya que AWS los cambia con el tiempo - los números exactos no son necesarios para el examen. Esto es solo para fines ilustrativos ☺.

Planes de ahorro EC2

- Obtén un descuento basado en el uso a largo plazo (hasta el 72%)
- Comprométete a un determinado tipo de uso (10 \$/hora durante 1 o 3 años)
- El uso más allá de los planes de ahorro de EC2 se factura al precio bajo demanda
- Bloqueado a una familia de instancias específica y a una región de AWS (por ejemplo, M5 en us-east-1)
- Flexible a través de:
 - Tamaño de instancia (por ejemplo, m5.xlarge, m5.2xlarge)
 - Sistema operativo (por ejemplo, Linux, Windows)
 - Tenencia (Anfitrión, Dedicado, Por defecto)



Instancias EC2 Spot

- Puedes obtener un **descuento de hasta el 90%** en comparación con la demanda
- Instancias que puedes "perder" en cualquier momento si su precio máximo es inferior al precio spot actual
- Las instancias **MÁS rentables** de AWS
- **Útil para las cargas de trabajo que son resistentes a los fallos**
 - Trabajos por lotes (Batch Jobs)
 - Análisis de datos
 - Procesamiento de imágenes
 - Cualquier carga de trabajo distribuida
 - Cargas de trabajo con una hora de inicio y finalización flexible
- **No es adecuado para trabajos críticos o bases de datos**

Hosts dedicados EC2

- Un servidor físico con capacidad de instancia EC2 totalmente dedicado a su uso
- Permite abordar los requisitos de **normativas y utilizar licencias de software vinculadas al servidor existentes** (licencias de software por socket, por núcleo, por VM)
- Opciones de compra:
 - **Bajo demanda** - pago por segundo para el host dedicado activo
 - **Reservado** - 1 o 3 años (sin pago inicial, pago inicial parcial, pago inicial total)
- La opción más cara
- Útil para el software que tiene un modelo de licencia complicado (BYOL - Bring Your Own License)
- O para empresas que tienen fuertes necesidades de regulación o cumplimiento

Instancias dedicadas de EC2

- Las instancias se ejecutan en un hardware dedicado para ti
- Puedes compartir el hardware con otras instancias de la misma cuenta
- No hay control sobre la ubicación de las instancias (se puede mover el hardware después de la parada/arranque)

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

Reservas de capacidad de EC2

- Reserva la capacidad de las instancias **bajo demanda** en una AZ específica para cualquier duración
- Siempre tendrás acceso a la capacidad de EC2 cuando la necesites
- **Sin compromiso de tiempo** (crear/cancelar en cualquier momento), **sin descuentos de facturación**
- Combina con las instancias regionales reservadas y los planes de ahorro para beneficiarte de descuentos en la facturación
- Se te cobra la tarifa bajo demanda tanto si ejecuta instancias como si no
- Adecuado para cargas de trabajo ininterrumpidas a corto plazo que necesitan estar en una AZ específica

¿Qué opción de compra me conviene?



- **Bajo demanda (On demand)**: venir y quedarse en el complejo cuando queramos, pagamos el precio completo
- **Reservada (Reserved)**: cómo planificar con antelación y si planeamos quedarnos durante mucho tiempo, podemos obtener un buen descuento
- **Planes de ahorro (Savings Plans)**: pagamos una cantidad por hora durante un periodo determinado y nos alojamos en cualquier tipo de habitación (por ejemplo, King, Suite, Vista al mar, ...)
- **Instancias de spot (Spot instances)**: el hotel permite que la gente puje por las habitaciones vacías y el mejor postor se queda con ellas. Puede ser expulsado en cualquier momento
- **Hosts dedicados (Dedicated Hosts)**: Se reserva un edificio entero del complejo turístico
- **Reservas de capacidad (Capacity Reservations)**: reservas una habitación por un periodo con el precio completo aunque no te alojes en ella

Comparación de precios

Ejemplo - m4.large - us-east-1

Tipo de precio	Precio (por hora)
Precio bajo demanda (On-demand)	0.10\$
Instancias de spot (Spot instances)	0.038\$ - 0.039\$ (hasta 61% de descuento)
Instancia reservada (1 año) (Reserved)	0,062\$ (sin anticipo) - 0,058\$ (todo por adelantado)
Instancia reservada (3 años) (Reserved)	0,043\$ (sin anticipo) - 0,037\$ (todo por adelantado)
Plan de ahorro EC2 (1 año) (Saving plan)	0,062\$ (sin anticipo) - 0,058\$ (todo por adelantado)
Instancia reservada convertible (1 año)	0,071\$ (sin anticipo) - 0,066\$ (todo por adelantado)
Host dedicado (Dedicated host)	Precio bajo demanda (On-demand)
Reserva de host dedicado (Dedicated host reservation)	Hasta el 70% de descuento
Reservas de capacidad (Capacity reservation)	Precio bajo demanda (On-demand)

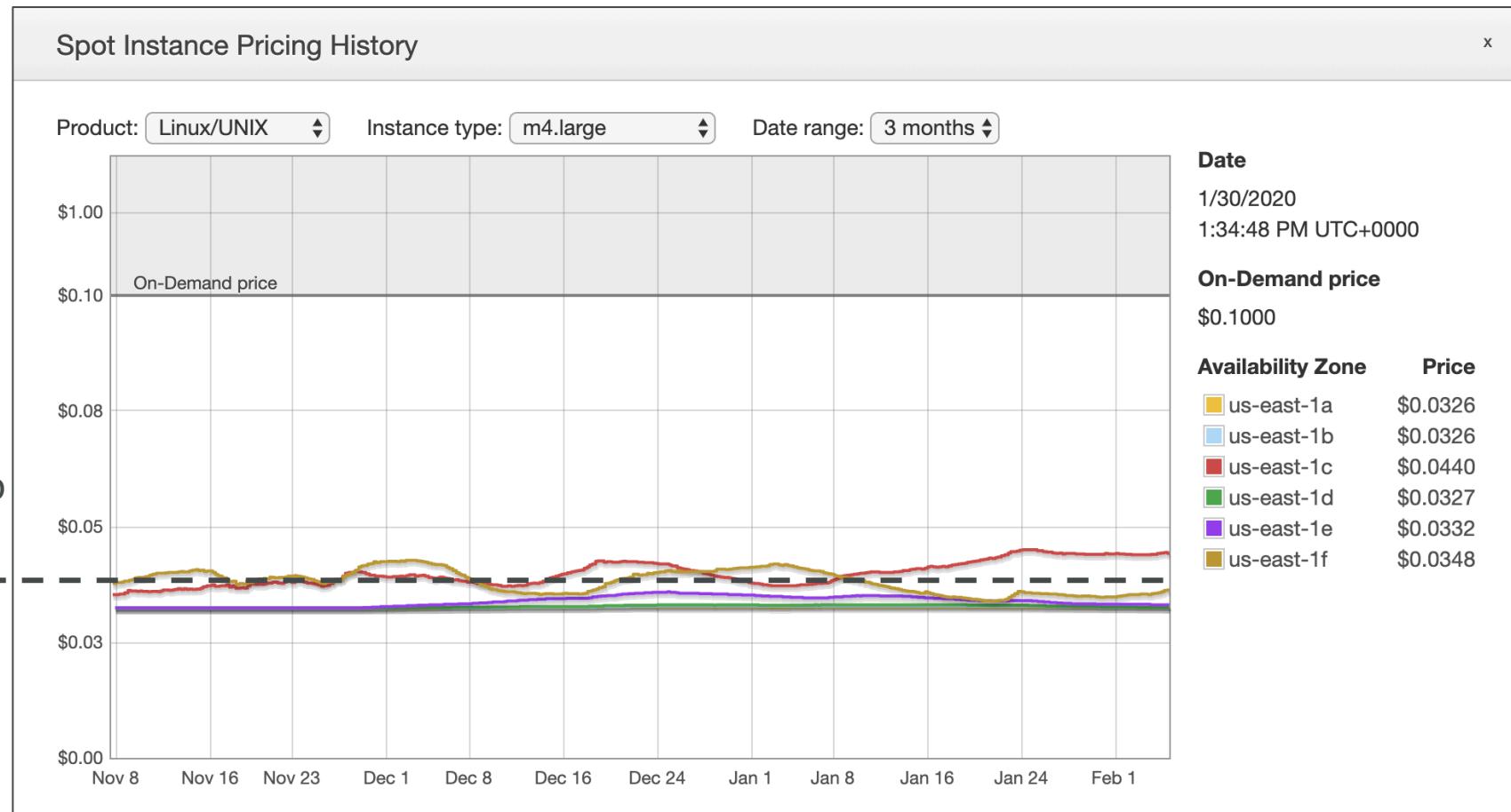


Peticiones de Instancias Spot de EC2

- Puedes obtener un descuento de hasta el 90% en comparación con la demanda
- Define el precio **spot máximo** y obtén la instancia mientras el precio **spot actual sea < máximo**
 - El precio spot por hora varía en función de la oferta y la capacidad
 - Si el precio spot actual > tu precio máximo, puedes elegir **parar** o **terminar** tu instancia con un periodo de gracia de 2 minutos.
- Otra estrategia: **Bloqueo de Spot**
 - "Bloquea" la instancia Spot durante un periodo de tiempo determinado (de 1 a 6 horas) sin interrupciones
 - En raras situaciones, la instancia puede ser reclamada
- **Se utiliza para trabajos por lotes, análisis de datos o cargas de trabajo resistentes a los fallos**
- **No es ideal para trabajos críticos o bases de datos**

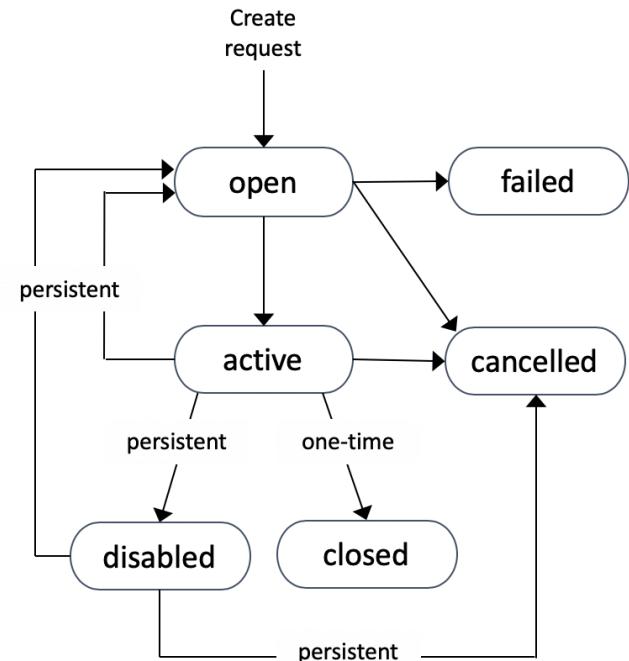
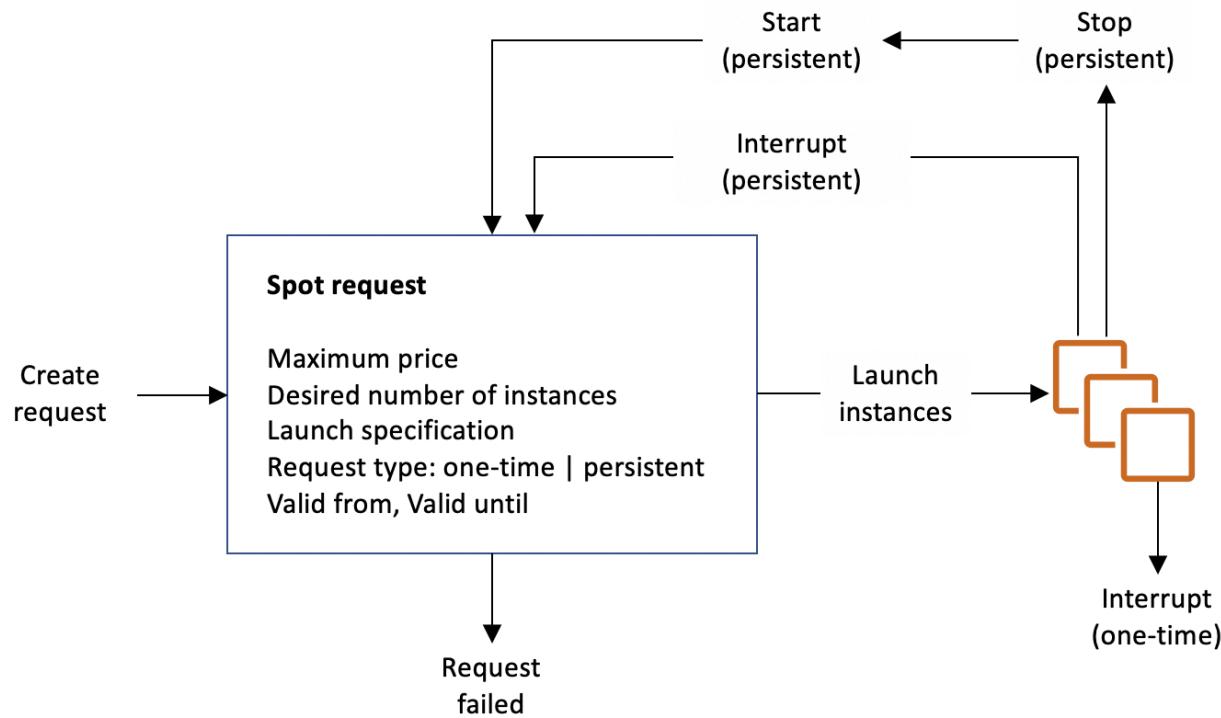
Precios de las Instancias Spot de EC2

Precio máximo definido por el usuario



<https://console.aws.amazon.com/ec2sp/v1/spot/home?region=us-east-1#>

¿Cómo terminar las Instancias Spot?



Sólo puedes cancelar las solicitudes de Instancias Spot que estén abiertas, activas o desactivadas.
La cancelación de una Solicitud Spot no termina las instancias
 Primero debes cancelar una Solicitud de Spot, y luego terminar las Instancias de Spot asociadas

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/spot-requests.html>

Flotas Spot (Spot Fleets)

- Flotas Spot = conjunto de Instancias de Spot + (opcional) Instancias bajo demanda
- La Flota Spot tratará de alcanzar la capacidad objetivo con restricciones de precio
 - Define los posibles pools de lanzamiento: tipo de instancia (m5.large), SO, Zona de Disponibilidad
 - Puede tener varios pools de lanzamiento, para que la flota pueda elegir
 - La Flota Spot deja de lanzar instancias cuando alcanza la capacidad o el coste máximo
- Estrategias para asignar Instancias de Spot:
 - **bajo precio:** desde el pool con el precio más bajo (optimización de costes, carga de trabajo corta)
 - **diversificado:** distribución en todos los pools (gran disponibilidad, cargas de trabajo largas)
 - **capacidad optimizada:** pool con la capacidad óptima para el número de instancias
- Las flotas de Spot nos permiten solicitar automáticamente las Instancias de Spot con el precio más bajo

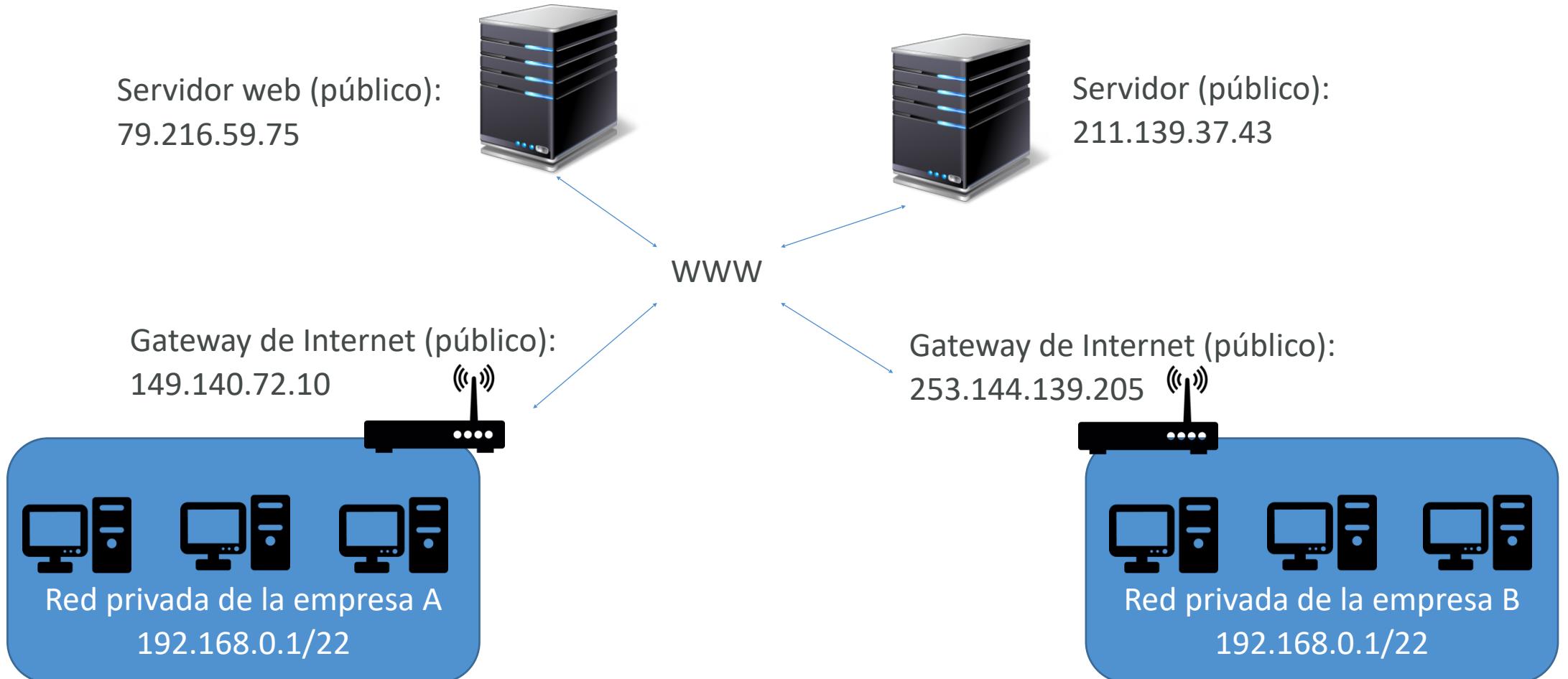
EC2 – Associate

IP privada vs. IP pública (IPv4)

- La red tiene dos tipos de IPs. IPv4 e IPv6:
 - IPv4: **1.160.10.240**
 - IPv6: **3ffe:1900:4545:3:200:f8ff:fe21:67cf**
- **En este curso, sólo utilizaremos IPv4.**
- IPv4 sigue siendo el formato más utilizado en Internet.
- IPv6 es más reciente y resuelve los problemas del Internet de las Cosas (IoT).
- El IPv4 permite **3.7 mil millones** diferentes direcciones en el espacio público
- IPv4: [0-255].[0-255].[0-255].[0-255].

IP privada vs. IP pública (IPv4)

Ejemplo



Diferencias fundamentales entre IP privada y pública (IPv4)

- **IP pública:**

- La IP pública significa que la máquina puede ser identificada en Internet (WWW)
- Debe ser única en toda la red (no puede haber dos máquinas con la misma IP pública)
- Se puede geolocalizar fácilmente

- **IP privada:**

- La IP privada significa que la máquina sólo puede ser identificada en una red privada
- La IP debe ser única en toda la red privada
- PERO dos redes privadas diferentes (dos empresas) pueden tener las mismas IP.
- Las máquinas se conectan a la WWW mediante un NAT + Gateway de Internet (un proxy)
- Sólo se puede utilizar un rango específico de IPs como IP privada

IPs elásticas

- Cuando paras y luego arrancas una instancia EC2, puede cambiar su IP pública.
- Si necesitas tener una IP pública fija para tu instancia, necesitas una IP elástica
- Una IP elástica es una IPv4 pública que te pertenece mientras no la elimines
- Puedes asignarla a una instancia a la vez

IP elástica

- Con una dirección IP elástica, puedes enmascarar el fallo de una instancia o software reasignando rápidamente la dirección a otra instancia de tu cuenta.
- Sólo puedes tener 5 Elastic IP en tu cuenta (puedes pedir a AWS que lo aumente).
- En general, [intenta evitar el uso de IP elásticas](#):
 - Suelen reflejar malas decisiones de arquitectura.
 - En su lugar, utiliza una IP pública aleatoria y registra un nombre DNS en ella
 - O, como veremos más adelante, utiliza un Load Balancer y no uses una IP pública

IP privada frente a IPv4 en AWS EC2 - Práctica

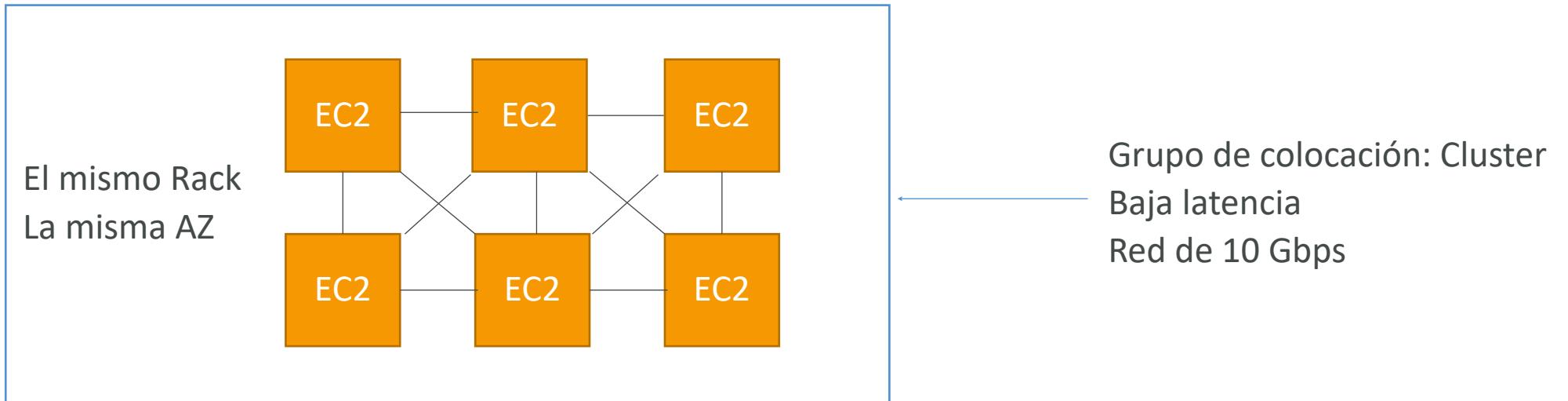
- Por defecto, tu máquina EC2 viene con
 - Una IP privada para la red interna de AWS
 - Una IP pública, para la WWW.
- Cuando estamos haciendo SSH en nuestras máquinas EC2:
 - No podemos usar una IP privada, porque no estamos en la misma red
 - Sólo podemos utilizar la IP pública.
- Si tu máquina se detiene y luego se inicia, **la IP pública puede cambiar**

Grupos de ubicación o colocación

- A veces quieres controlar la estrategia de colocación de la Instancia EC2
- Esa estrategia puede definirse mediante grupos de colocación
- Cuando creas un grupo de colocación, especificas una de las siguientes estrategias para el grupo:
 - **Cluster:** agrupa las instancias en un grupo de baja latencia en una única Zona de Disponibilidad
 - **Distribuida:** coloca estrictamente un pequeño grupo de instancias en distintos equipos de hardware subyacentes para reducir los fallos correlacionados (máximo 7 instancias por grupo por AZ)
 - **Partición:** reparte las instancias en muchas particiones diferentes (que dependen de diferentes conjuntos de racks) dentro de una AZ. Escala a cientos de instancias EC2 por grupo (Hadoop, Cassandra, Kafka)

Grupos de ubicación o colocación

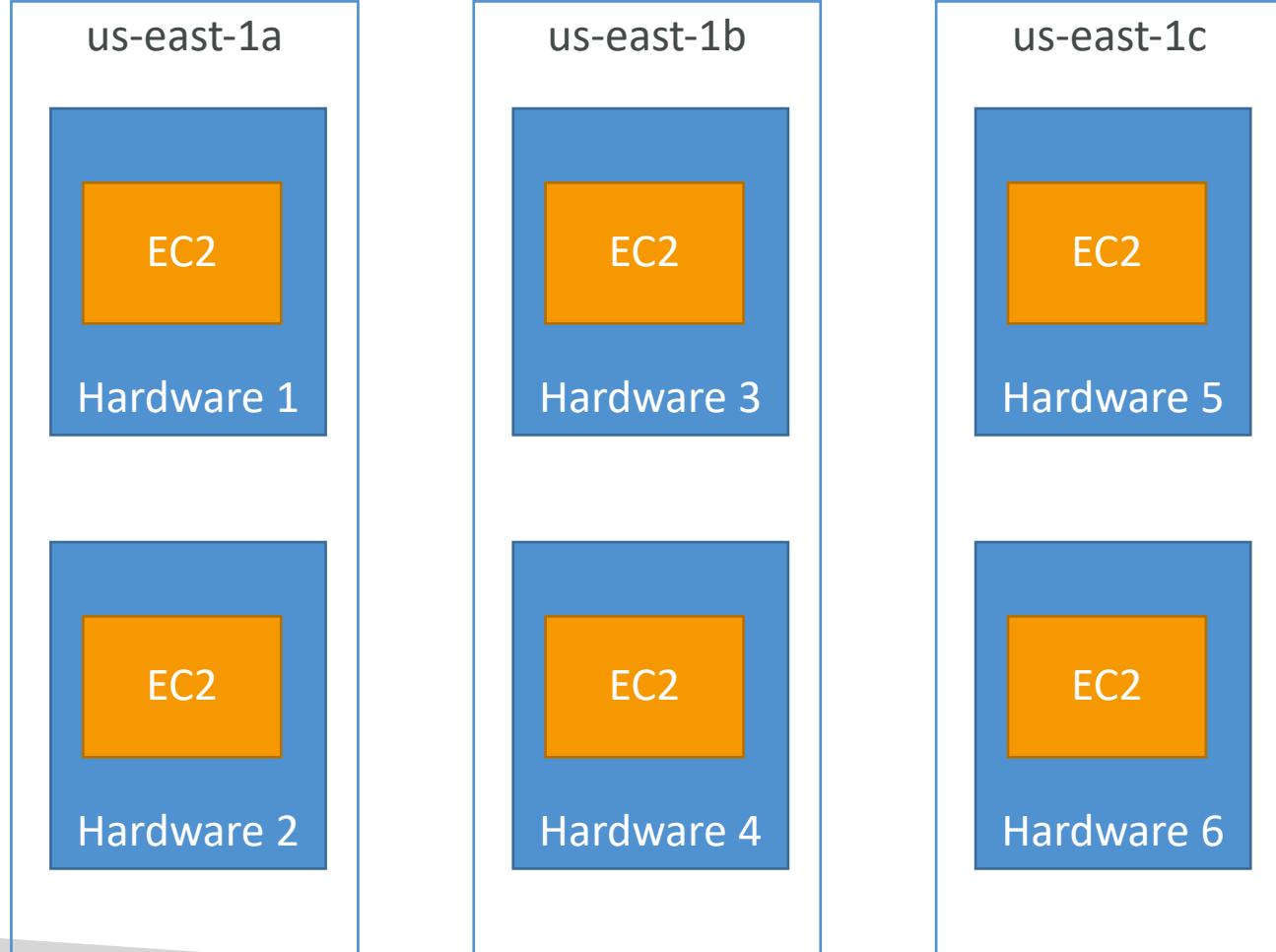
Cluster



- Ventajas: Gran red (10 Gbps de ancho de banda entre instancias con la red mejorada activada - recomendada)
- Contras: Si el rack falla, todas las instancias fallan al mismo tiempo
- Caso de uso:
 - Trabajo de Big Data que necesita completarse rápidamente
 - Aplicación que necesita una latencia extremadamente baja y un alto rendimiento de la red

Grupos de ubicación o colocación

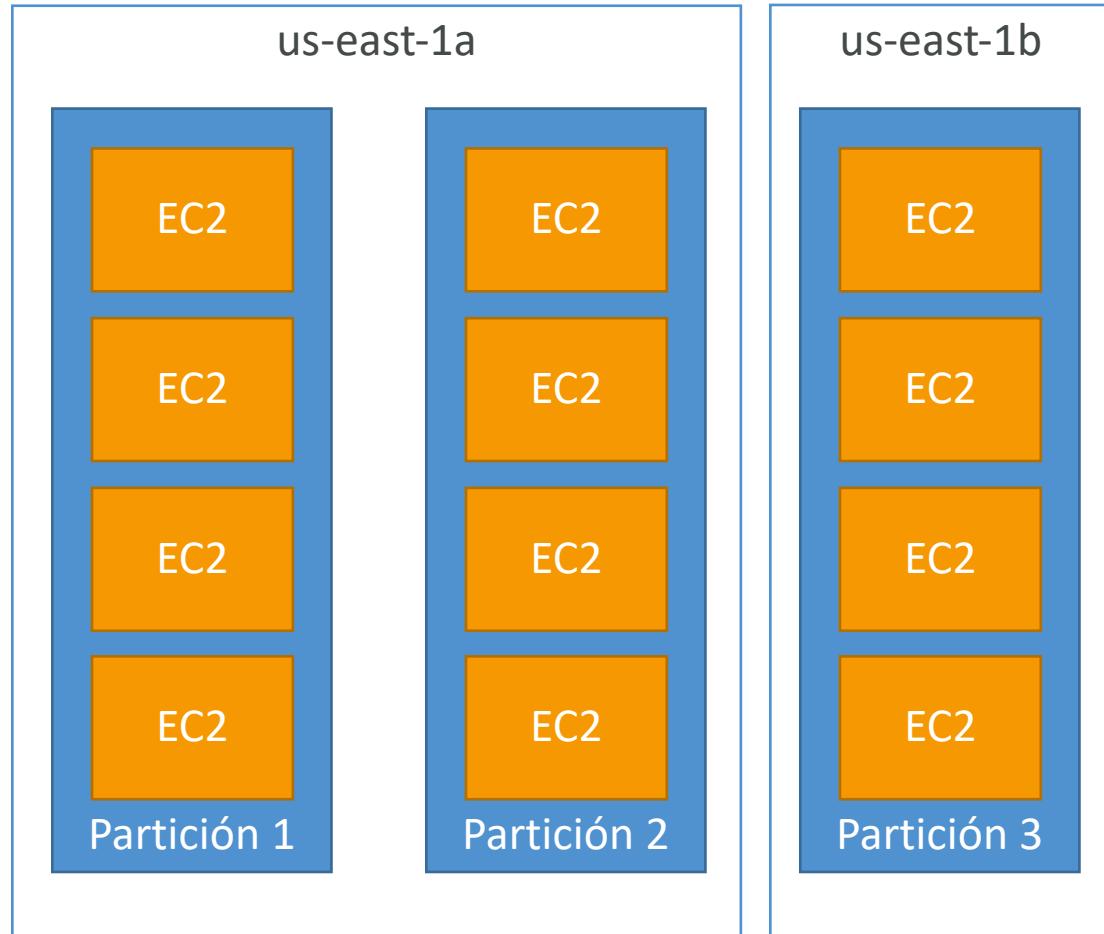
Distribuida



- Ventajas:
 - Puede abarcar varias zonas de disponibilidad (AZ)
 - Se reduce el riesgo de fallos simultáneos
 - Las instancias EC2 están en hardware físico diferente
- Contras:
 - Limitado a 7 instancias por AZ por grupo de colocación
- Caso de uso:
 - Aplicación que necesita maximizar la alta disponibilidad
 - Aplicaciones críticas en las que cada instancia debe estar aislada de los fallos de las demás

Grupos de ubicación o colocación

Partición



- Hasta 7 particiones por AZ
- Puede abarcar varias AZ en la misma región
- Hasta 100 instancias EC2
- Las instancias de una partición no comparten Racks con las instancias de las otras particiones
- Un fallo en la partición puede afectar a muchos EC2 pero no afectará a otras particiones
- Las instancias EC2 tienen acceso a la información de la partición como metadatos
- Casos de uso: HDFS, HBase, Cassandra, Kafka

Elastic Network Interfaces (ENI)

- Componente lógico de una VPC que representa una **tarjeta de red virtual**
- La ENI puede tener los siguientes atributos
 - IPv4 privada primaria, una o más IPv4 secundarias
 - Una IP elástica (IPv4) por IPv4 privada
 - Una IPv4 pública
 - Uno o más grupos de seguridad
 - Una dirección MAC
- Puedes crear ENI independientes y adjuntarlas sobre la marcha (moverlas) en instancias EC2 para la conmutación por error
- Vinculadas a una zona de disponibilidad (AZ) específica

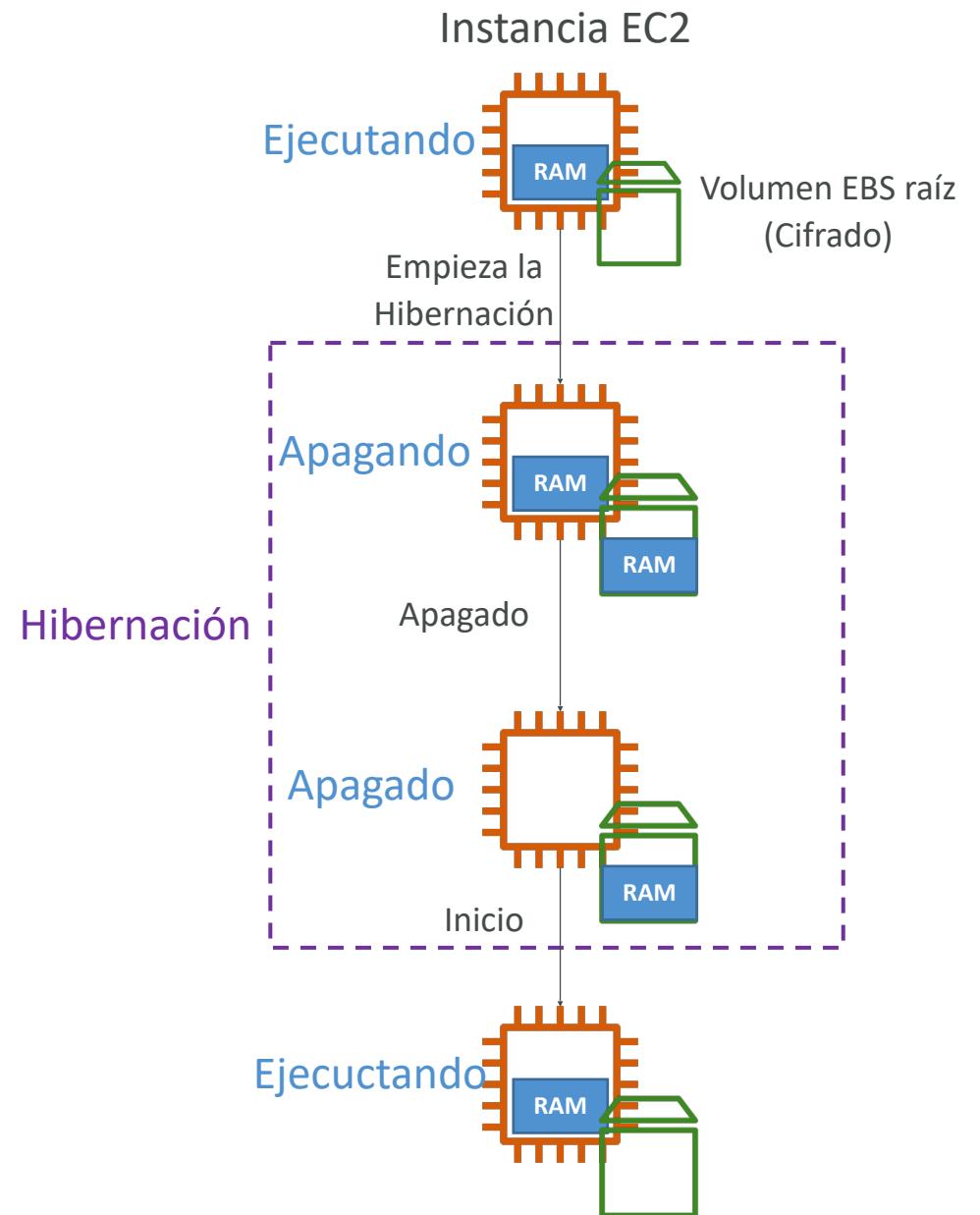


Hibernación de EC2

- Sabemos que podemos parar, terminar las instancias
 - Parar - los datos del disco (EBS) se mantienen intactos en el siguiente arranque
 - Terminar - se pierden los volúmenes EBS (root) que también están preparados para ser destruidos
- En el arranque, ocurre lo siguiente
 - Primer arranque: el SO arranca y se ejecuta el script EC2 User Data
 - Siguientes arranques: el SO arranca
 - Después se inicia tu aplicación, se calientan las cachés, ¡y eso puede llevar tiempo!

Hibernación de EC2

- Presentación de **EC2 Hibernate**:
 - Se conserva el estado en memoria (RAM)
 - El arranque de la instancia es mucho más rápido (el sistema operativo no se detiene/reinicia)
 - Bajo el capó: el estado de la RAM se escribe en un archivo en el volumen EBS raíz
 - El volumen EBS raíz debe estar encriptado
- Casos de uso:
 - Procesamiento de larga duración
 - Guardar el estado de la RAM
 - Servicios que tardan en inicializarse



Hibernación de EC2 – Es bueno saber que

- Familias de instancias soportadas - C3, C4, C5, I3, M3, M4, R3, R4, T2, T3, ...
- Tamaño de la RAM de la instancia - debe ser inferior a 150 GB.
- Tamaño de la Instancia - no se soporta para instancias bare metal.
- AMI - Amazon Linux 2, Linux AMI, Ubuntu, RHEL, CentOS y Windows...
- Volumen root - debe ser EBS, encriptado
- Disponible para instancias bajo demanda, reservadas y Spot
- Una instancia NO puede estar hibernada más de 60 días

Almacenamiento de la instancia EC2

¿Qué es un volumen EBS?



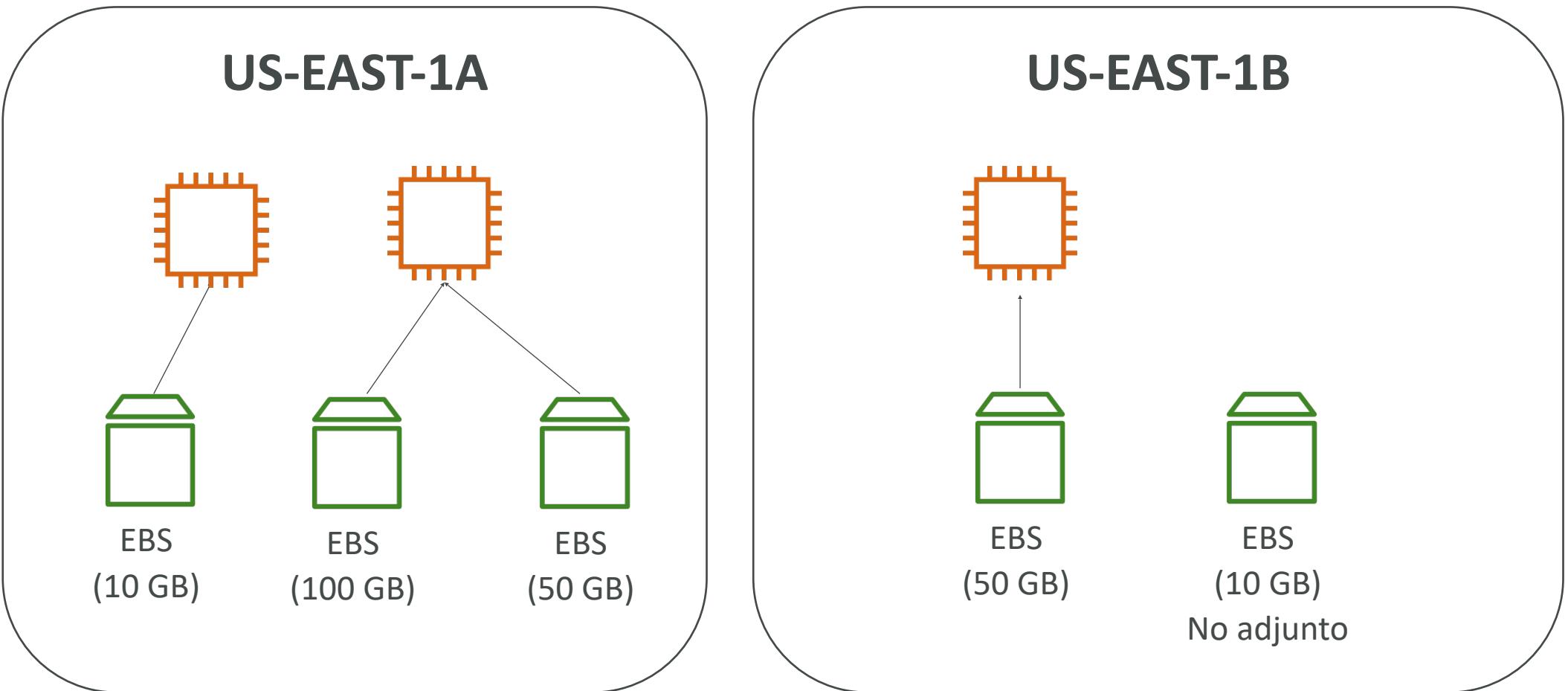
- Un **volumen EBS (Elastic Block Store)** es una **unidad de red** que puede adjuntar a las instancias mientras se ejecutan
- Permite que las instancias persistan los datos, incluso después de su finalización
- **Sólo pueden montarse en una instancia a la vez** (a nivel de CCP)
- Están vinculados **a una zona de disponibilidad específica**

- Analogía: Piensa en ellos como una "memoria USB de red"
- Nivel gratuito: 30 GB de almacenamiento EBS gratuito de tipo Propósito General (SSD) o Magnético al mes

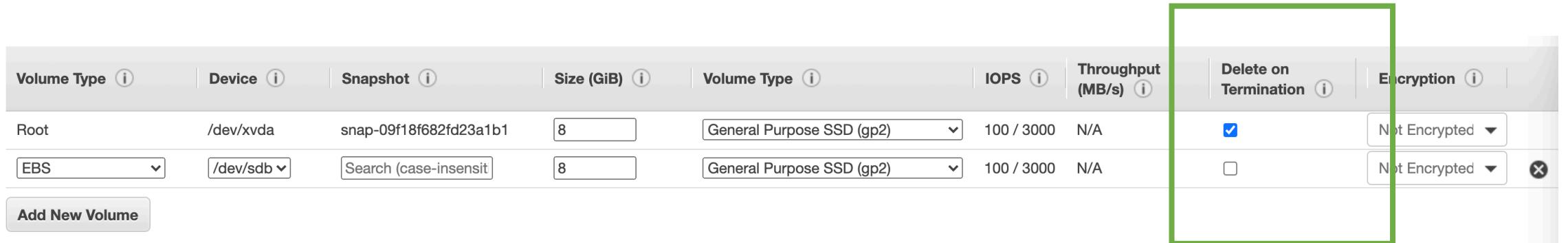
Volumen EBS

- 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 separar de una instancia EC2 y conectarla a otra rápidamente
- Está bloqueado en una Zona de Disponibilidad (AZ)
 - Un volumen EBS en us-east-1a no puede adjuntarse a us-east-1b
 - Para trasladar un volumen, primero hay que hacer un snapshot del mismo
- Tener una capacidad provisionada (tamaño en GBs, e IOPS)
 - Se facturará toda la capacidad aprovisionada
 - Puede aumentar la capacidad de la unidad con el tiempo

Volumen EBS - Ejemplo



EBS - Atributo "Borrar al terminar"



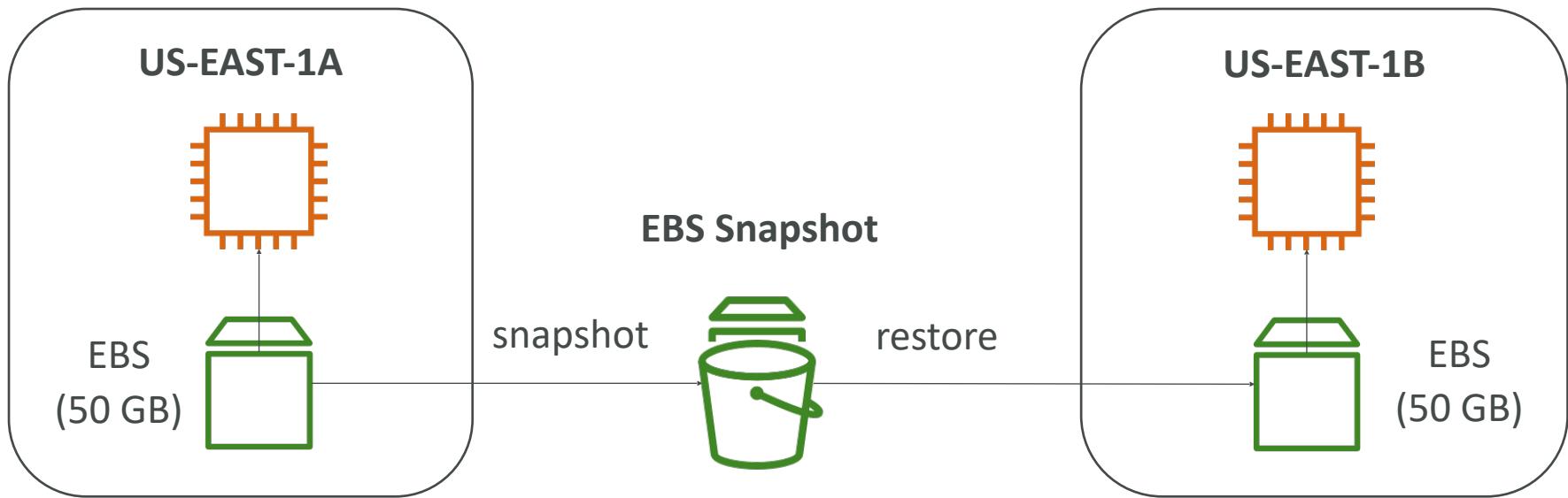
The screenshot shows the AWS EBS volume configuration interface. It lists two volumes: a 'Root' volume and another EBS volume. The 'Delete on Termination' checkbox is checked for the Root volume, indicating it will be deleted when the instance terminates. This checkbox is highlighted with a green border.

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

- Controla el comportamiento de EBS cuando una instancia EC2 termina
 - Por defecto, se elimina el volumen EBS root / raíz (atributo habilitado)
 - Por defecto, cualquier otro volumen EBS adjunto no se elimina (atributo deshabilitado)
- Esto puede ser controlado por la consola de AWS / AWS CLI
- **Caso de uso: preservar el volumen root / raíz cuando se termina la instancia**

Snapshot / Instantáneas de EBS

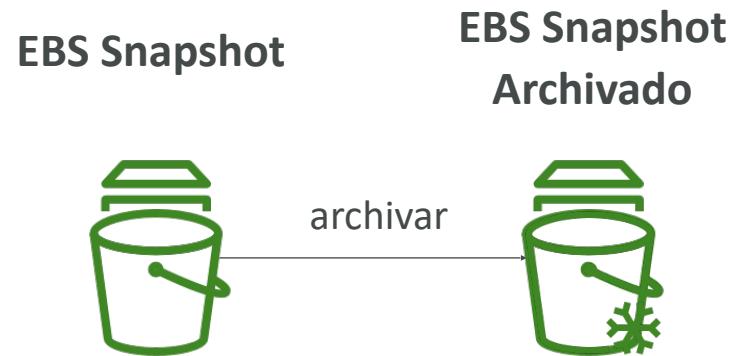
Haz una copia de seguridad (snapshot) de tu volumen EBS en un momento dado
No es necesario separar el volumen para hacer la instantánea, pero se recomienda
Puedes copiar las instantáneas a través de AZ o Región



Características de los Snapshots de EBS

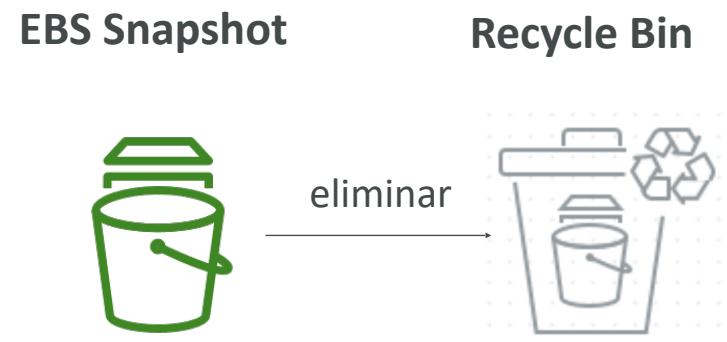
• Archivo de Snapshots de EBS

- Mover un snapshot a un "nivel de archivo" que es un 75% más barato
- La restauración del archivo tarda entre 24 y 72 horas



• Papelera de reciclaje para Snapshots EBS

- Configura reglas para retener los snapshots eliminados para poder recuperarlos después de un borrado accidental
- Especifica la retención (de 1 día a 1 año)



Visión general de AMI



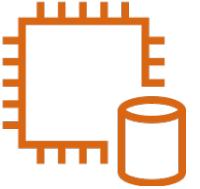
- AMI = Amazon Machine Image
- Las AMI son una **personalización** de una instancia EC2
 - Añades tu propio software, configuración, sistema operativo, monitorización...
 - Tiempo de arranque/configuración más rápido porque todo el software está preempaquetado
- Las AMI se construyen para una **región específica** (y pueden copiarse entre regiones)
- Puedes lanzar instancias EC2 desde:
 - **Una AMI pública:** proporcionada por AWS
 - **Tu propia AMI:** la creas y la mantienes tú mismo
 - **Una AMI de AWS Marketplace:** una AMI hecha por otra persona (y potencialmente vendida)

Proceso AMI (desde una instancia EC2)

- Iniciar una instancia EC2 y personalizarla
- Detener la instancia (para la integridad de los datos)
- Construir una AMI - esto también creará instantáneas de EBS
- Lanzar instancias desde otras AMIs



Almacén de instancias EC2



- Los volúmenes EBS son **unidades de red** con un rendimiento bueno pero “limitado”
- **Si necesitas un disco de hardware de alto rendimiento, utilizas EC2 Instance Store**
- Mejor rendimiento de E/S
- Los almacenes de instancias EC2 pierden su almacenamiento si se detienen (son efímeros)
- Bueno para el buffer / cache / datos de memoria virtual / contenido temporal
- Riesgo de pérdida de datos si el hardware falla
- Las copias de seguridad y la replicación son responsabilidad tuya

Almacén local de instancias EC2

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

IOPS muy altas

Tipos de volúmenes EBS

- Los volúmenes EBS vienen en 6 tipos
 - **gp2 / gp3 (SSD)**: Volumen SSD de uso general que equilibra el precio y el rendimiento para una amplia variedad de cargas de trabajo
 - **io1 / io2 (SSD)**: El volumen SSD de mayor rendimiento para cargas de trabajo de misión crítica de baja latencia o alto rendimiento
 - **st1 (HDD)**: Volumen de disco duro de bajo coste diseñado para cargas de trabajo de acceso frecuente y alto rendimiento
 - **sc1 (HDD)**: El volumen de disco duro más barato, diseñado para cargas de trabajo de acceso menos frecuente
- Los volúmenes EBS se caracterizan en Tamaño | Rendimiento | IOPS (I/O Ops Per Sec)
- En caso de duda, consulta siempre la documentación de AWS: ¡es buena!
- **Sólo se pueden utilizar gp2/gp3 y io1/io2 como volúmenes de arranque**

Tipos de volúmenes EBS - Casos de uso

SSD de uso general

- Almacenamiento rentable, baja latencia
- Volúmenes de arranque del sistema, escritorios virtuales, entornos de desarrollo y prueba
- 1 GiB - 16 TiB
- gp3:
 - Línea de base de 3.000 IOPS y rendimiento de 125 MiB/s
 - Puede aumentar las IOPS hasta 16.000 y el rendimiento hasta 1000 MiB/s de forma independiente
- gp2:
 - Los volúmenes gp2 pequeños pueden reventar las IOPS hasta 3.000
 - El tamaño del volumen y las IOPS están vinculados, las IOPS máximas son 16.000
 - 3 IOPS por GB, lo que significa que con 5.334 GB estamos en el máximo de IOPS

Tipos de volúmenes EBS - Casos de uso

IOPS provisionadas (PIOPS) SSD

- Aplicaciones empresariales críticas con un rendimiento sostenido de IOPS
- O aplicaciones que necesitan más de 16.000 IOPS
- Excelente para las **cargas de trabajo de las bases de datos** (sensibles al rendimiento y la consistencia del almacenamiento)
- io1/io2 (4 GiB - 16 TiB):
 - PIOPS máximos: 64.000 para instancias Nitro EC2 y 32.000 para otras
 - Puede aumentar los PIOPS independientemente del tamaño del almacenamiento
 - io2 tiene más durabilidad y más IOPS por GiB (al mismo precio que io1)
- io2 Block Express (4 GiB - 64 TiB):
 - Latencia de menos de un milisegundo
 - PIOPS máximas: 256.000 con una relación IOPS:GiB de 1.000:1
- Soporta EBS Multi-attach

Tipos de volúmenes EBS - Casos de uso

Unidades de disco duro (HDD)

- No puede ser un volumen de arranque
- De 125 GiB a 16 TiB
- Disco duro de rendimiento optimizado (stl)
 - Big Data, almacenes de datos, procesamiento de logs
 - **Rendimiento máximo** de 500 MiB/s - IOPS máximo de 500
- Disco duro frío (sc1):
 - Para datos a los que se accede con poca frecuencia
 - Escenarios en los que el menor coste es importante
 - **Rendimiento máximo** de 250 MiB/s - IOPS máximas de 250

EBS - Resumen de los tipos de volúmenes

	General Purpose SSD	
Tipo de volumen	gp3	gp2
Durabilidad	99,8 % - 99,9 % de durabilidad (0,1 % - 0,2 % tasa anual de errores)	
Casos de uso	<ul style="list-style-type: none">Aplicaciones interactivas de baja latenciaEntornos de desarrollo y pruebasEscritorios virtualesBases de datos de tamaño mediano y una sola instanciaVolúmenes de arranque	
Tamaño del volumen	1 GiB - 16 TiB	
Máximo de IOPS por volumen (E/S de 16 KiB)**	16,000	
Rendimiento máximo por volumen**	1000 MiB/s	250 MiB/s *
Amazon EBS Multi-attach	No admitido	
Volumen de arranque	Soportado	

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-volume-types.html#solid-state-drives>

EBS - Resumen de los tipos de volúmenes

Provisioned IOPS SSD			
Tipo de volumen	io2 Block Express*	io2	io1
Durabilidad	99,999 % de durabilidad (0,001 % tasa anual de errores)	99,999 % de durabilidad (0,001 % tasa anual de errores)	99,8 % - 99,9 % de durabilidad (0,1 % - 0,2 % tasa anual de errores)
Casos de uso	<p>Cargas de trabajo que requieren lo siguiente:</p> <ul style="list-style-type: none"> • Latencia inferior a milisegundos • Rendimiento de IOPS sostenido • Más de 64 000 IOPS o 1000 MiB/s de rendimiento 	<ul style="list-style-type: none"> • Cargas de trabajo que requieren un rendimiento sostenido de IOPS o más de 16,000 IOPS • Cargas de trabajo de bases de datos con uso intensivo de operaciones de E/S 	
Tamaño del volumen	4 GiB - 64 TiB	4 GiB - 16 TiB	
IOPS máximo por volumen (E/S de 16 KiB)	256 000	64 000†	
Rendimiento máximo por volumen	4000 MiB/s	1000 MiB/s†	
Amazon EBS Multi-attach	Compatible		
Volumen de arranque	Compatible		

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-volume-types.html#solid-state-drives>

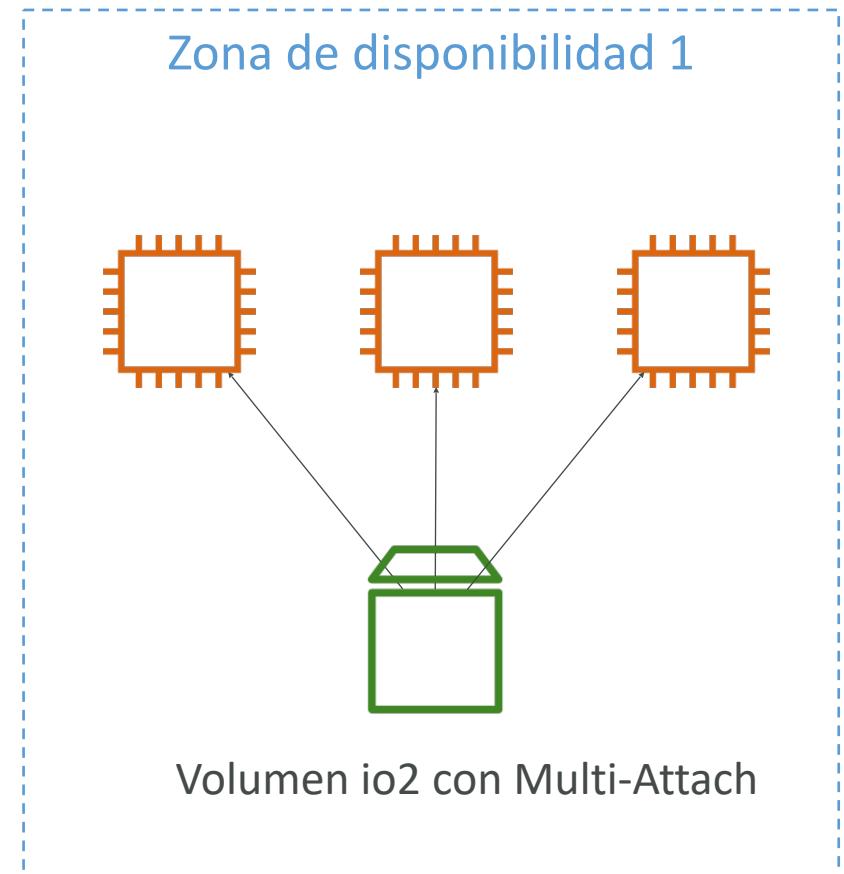
EBS - Resumen de los tipos de volúmenes

	HDD con rendimiento optimizado	HDD en frío
Tipo de volumen	st1	sc1
Durabilidad	99,8 % - 99,9 % de durabilidad (0,1 % - 0,2 % tasa anual de errores)	99,8 % - 99,9 % de durabilidad (0,1 % - 0,2 % tasa anual de errores)
Casos de uso	<ul style="list-style-type: none"> • Big data • Data warehouses • Procesamiento de registros 	<ul style="list-style-type: none"> • Almacenamiento orientado al rendimiento para datos a los que se accede con poca frecuencia • Escenarios en los que es importante el costo de almacenamiento más bajo
Tamaño del volumen	125 GiB - 16 TiB	125 GiB - 16 TiB
IOPS máximo por volumen (E/S de 1 MiB)	500	250
Rendimiento máximo por volumen	500 MiB/s	250 MiB/s
Amazon EBS Multi-attach	No admitido	No admitido
Volumen de arranque	No admitido	No admitido

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-volume-types.html#solid-state-drives>

EBS Multi-Attach - familia io1/io2

- Adjunta el mismo volumen EBS a varias instancias EC2 en la misma AZ
- Cada instancia tiene permisos completos de lectura y escritura en el volumen de alto rendimiento
- Caso de uso:
 - Conseguir una mayor disponibilidad de las aplicaciones en clusters de Linux (por ejemplo, Teradata)
 - Las aplicaciones deben gestionar operaciones de escritura concurrentes
- Hasta 16 instancias EC2 a la vez
- Debe utilizar un sistema de archivos que sea compatible con el clúster (no XFS, EX4, etc...)



Encriptación de EBS

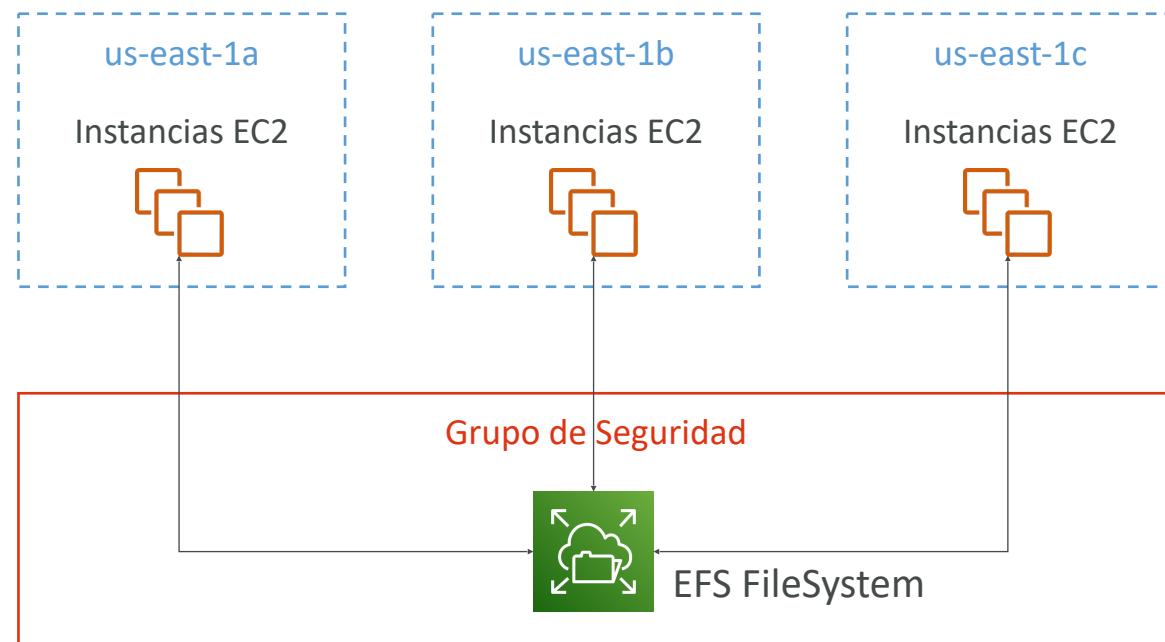
- Cuando creas un volumen EBS encriptado, obtienes lo siguiente:
 - Los datos en reposo están encriptados dentro del volumen
 - Todos los datos en movimiento entre la instancia y el volumen están encriptados
 - Todas las instantáneas están encriptadas
 - Todos los volúmenes creados a partir de la instantánea
- El cifrado y el descifrado se gestionan de forma transparente (no tienes que hacer nada)
- El cifrado tiene un impacto mínimo en la latencia
- El cifrado de EBS aprovecha las claves de KMS (AES-256)
- La copia de una Snapshot no cifrada permite el cifrado
- Las instantáneas de los volúmenes encriptados están encriptadas

Cifrado: cifrar un volumen EBS

- Crea una Snapshot de EBS del volumen
- Encripta la instantánea EBS (utilizando la copia)
- Crea un nuevo volumen EBS a partir de la instantánea (el volumen también estará encriptado)
- Ahora puedes adjuntar el volumen encriptado a la instancia original

Amazon EFS – Elastic File System

- NFS gestionado (sistema de archivos de red) que puede montarse en muchas EC2
- EFS funciona con instancias EC2 en multi-AZ
- Alta disponibilidad, escalable, caro (3x gp2), pago por uso



Amazon EFS – Elastic File System

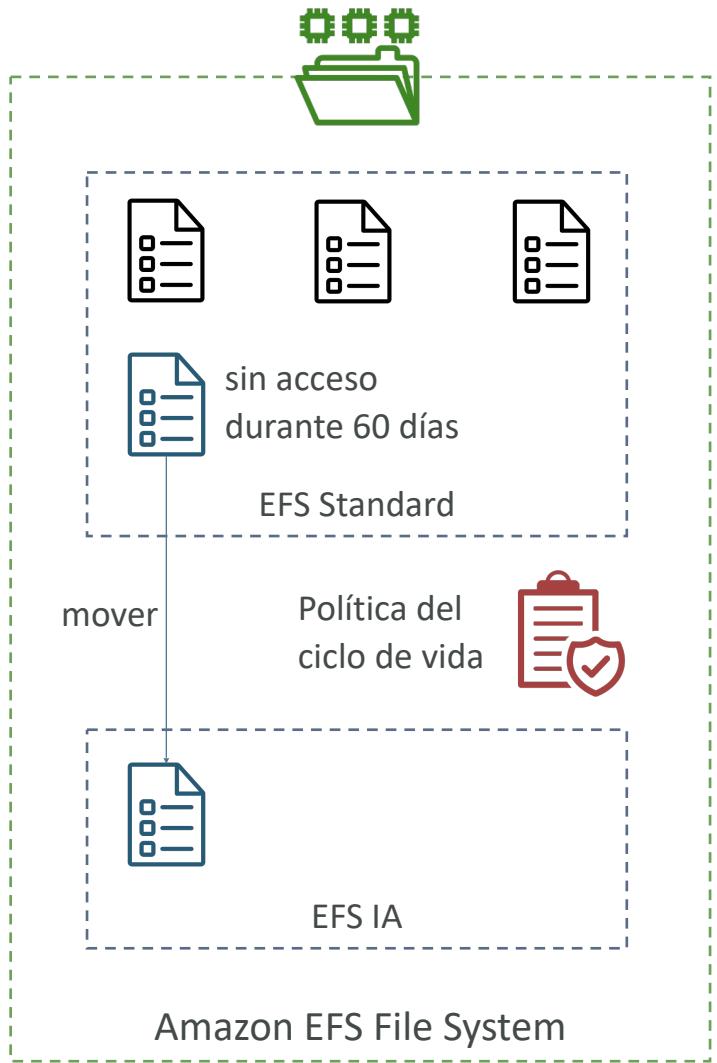
- Casos de uso: gestión de contenidos, servicio web, intercambio de datos, Wordpress
- Utiliza el protocolo NFSv4.1
- Utiliza el grupo de seguridad para controlar el acceso al EFS
- **Compatible con AMI basadas en Linux (no en Windows)**
- Cifrado en reposo mediante KMS
- Sistema de archivos POSIX (~Linux) que tiene una API de archivos estándar
- El sistema de archivos se escala automáticamente, paga por uso, ¡no hay que planificar la capacidad!

EFS - Clases de rendimiento y almacenamiento

- **Escala EFS**
 - 1000s de clientes NFS concurrentes, 10 GB+ /s de rendimiento
 - Crece hasta convertirse en un sistema de archivos en red a escala de petabytes, de forma automática
- **Modo de rendimiento (establecido en el momento de la creación del EFS)**
 - Propósito general (por defecto): casos de uso sensibles a la latencia (servidor web, CMS, etc.)
 - E/S máxima: mayor latencia, rendimiento, altamente paralelo (big data, procesamiento de medios)
- **Modo de rendimiento (Throughput)**
 - Ráfaga ($1\text{ TB} = 50\text{MiB/s} + \text{ráfaga de hasta } 100\text{MiB/s}$)
 - Aprovisionado: fija tu rendimiento independientemente del tamaño del almacenamiento, por ejemplo: 1 GiB/s para un almacenamiento de 1 TB

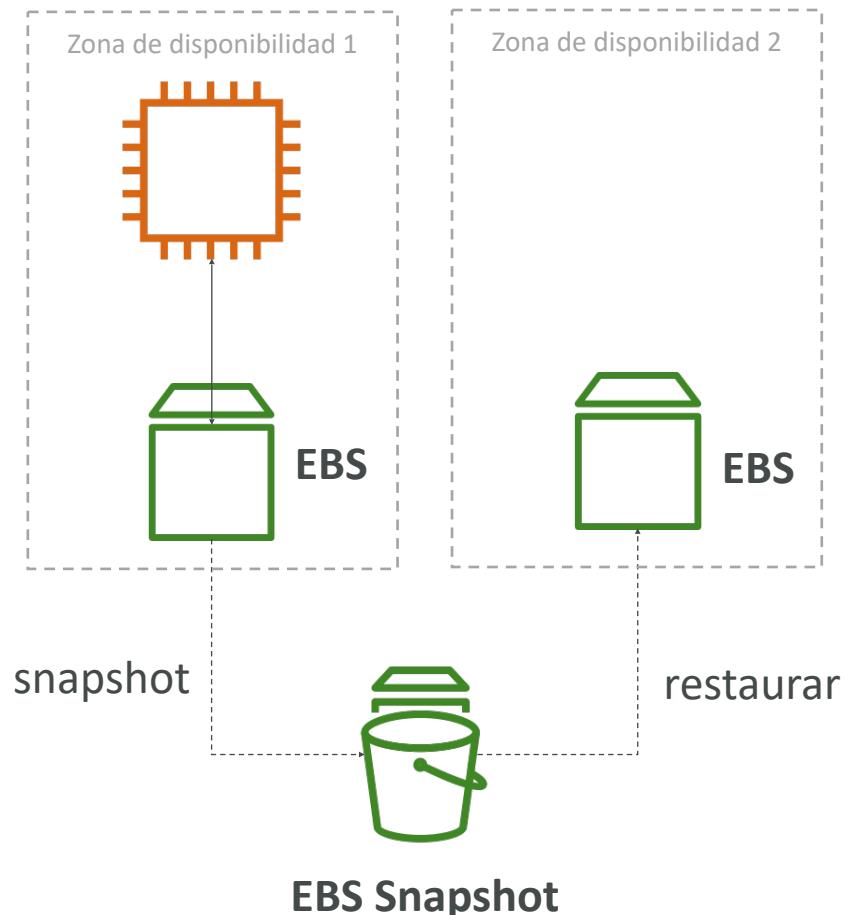
EFS - Clases de almacenamiento

- **Niveles de almacenamiento (función de gestión del ciclo de vida: mover el archivo después de N días)**
 - Estándar: para archivos de acceso frecuente
 - Acceso infrecuente (EFS-IA): coste de recuperación de los archivos, menor precio de almacenamiento. Habilita EFS-IA con una política de ciclo de vida
- **Disponibilidad y durabilidad**
 - Estándar: Multi-AZ, ideal para prod
 - Una zona: Una AZ, genial para dev, copia de seguridad activada por defecto, compatible con IA (EFS One Zone-IA)
- Más del 90% de ahorro de costes



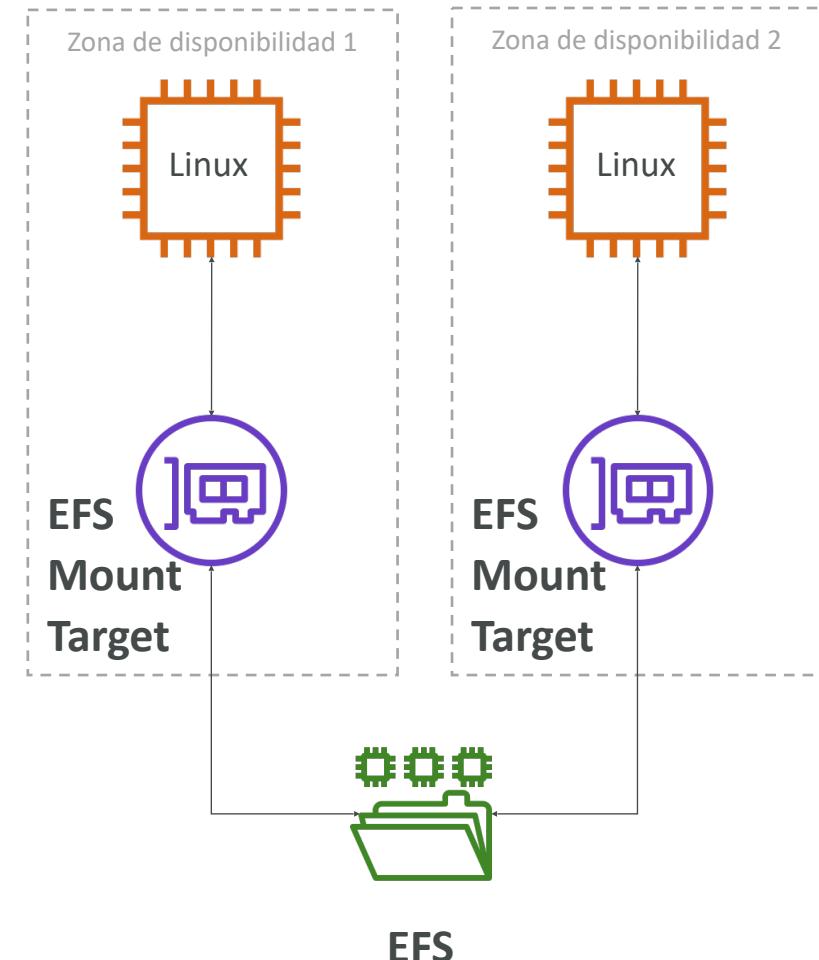
EBS vs EFS – Elastic Block Storage

- Los volúmenes EBS...
 - sólo pueden adjuntarse a una instancia a la vez
 - están bloqueados a nivel de Zona de Disponibilidad (AZ)
 - gp2: la IO aumenta si aumenta el tamaño del disco
 - io1: puede aumentar la IO de forma independiente
- Para migrar un volumen EBS a través de la AZ
 - Haz una Snapshot
 - Restaura la instantánea en otra AZ
 - Las copias de seguridad de EBS utilizan IO y no deberías ejecutarlas mientras tu aplicación esté manejando mucho tráfico
- Los volúmenes EBS root de las instancias se terminan por defecto si la instancia EC2 se termina. (puedes desactivarlo)



EBS vs EFS – Elastic File System

- Montar 100s de instancias a través de AZ
 - Compartir archivos del sitio web EFS (WordPress)
 - Sólo para instancias Linux (POSIX)
-
- EFS tiene un precio más elevado que EBS
 - Puede aprovechar EFS-IA para ahorrar costes
-
- Recuerda: EFS vs EBS vs Instance Store



Fundamentos de AWS – Parte II

Escalabilidad y alta disponibilidad

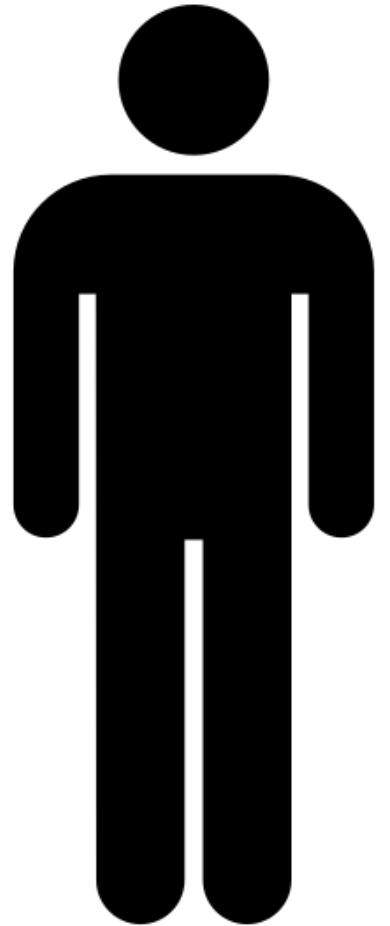
- La escalabilidad significa que una aplicación/sistema puede manejar mayores cargas adaptándose.
- Hay dos tipos de escalabilidad:
 - Escalabilidad vertical
 - Escalabilidad horizontal (= elasticidad)
- **La escalabilidad está vinculada pero es diferente a la Alta Disponibilidad**
- Vamos a profundizar en la distinción, utilizando un centro de llamadas como ejemplo

Escalabilidad vertical

- La escalabilidad vertical significa aumentar el tamaño de la instancia
- Por ejemplo, tu 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.
- RDS, ElastiCache son servicios que pueden escalar verticalmente.
- Suele haber un límite en cuanto a la escalabilidad vertical (límite de hardware)



operador junior

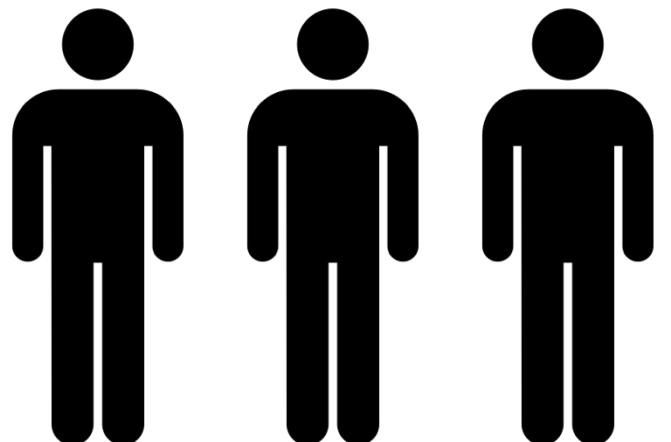
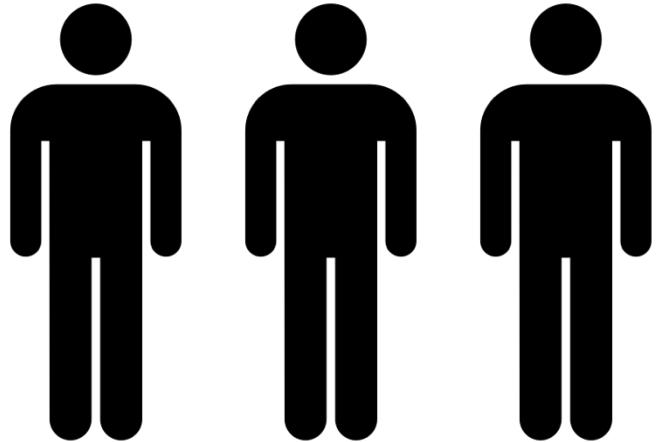


senior operator

Escalabilidad horizontal

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

operador operador operador

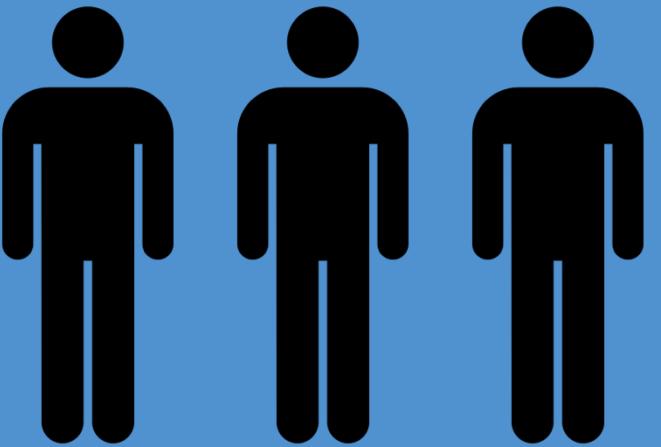


operador operador operador

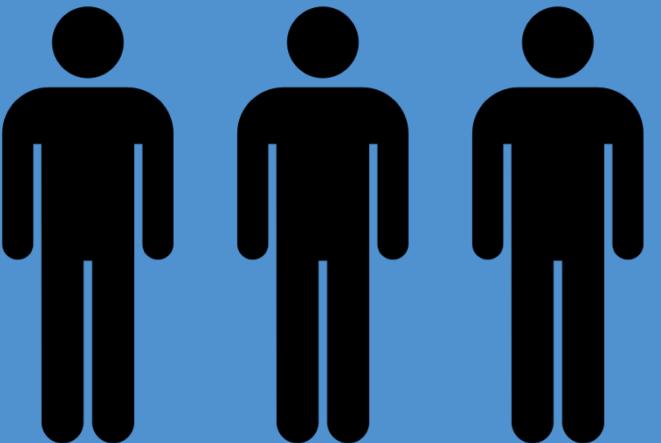
Alta disponibilidad

- La alta disponibilidad suele ir de la mano del escalado horizontal
- La alta disponibilidad significa ejecutar tu aplicación/sistema en al menos 2 centros de datos (== Zonas de Disponibilidad)
- El objetivo de la alta disponibilidad es sobrevivir a la pérdida de un centro de datos
- La alta disponibilidad puede ser pasiva (para RDS Multi AZ, por ejemplo)
- La alta disponibilidad puede ser activa (para el escalado horizontal)

primer edificio en Nueva York



segundo edificio en San Francisco

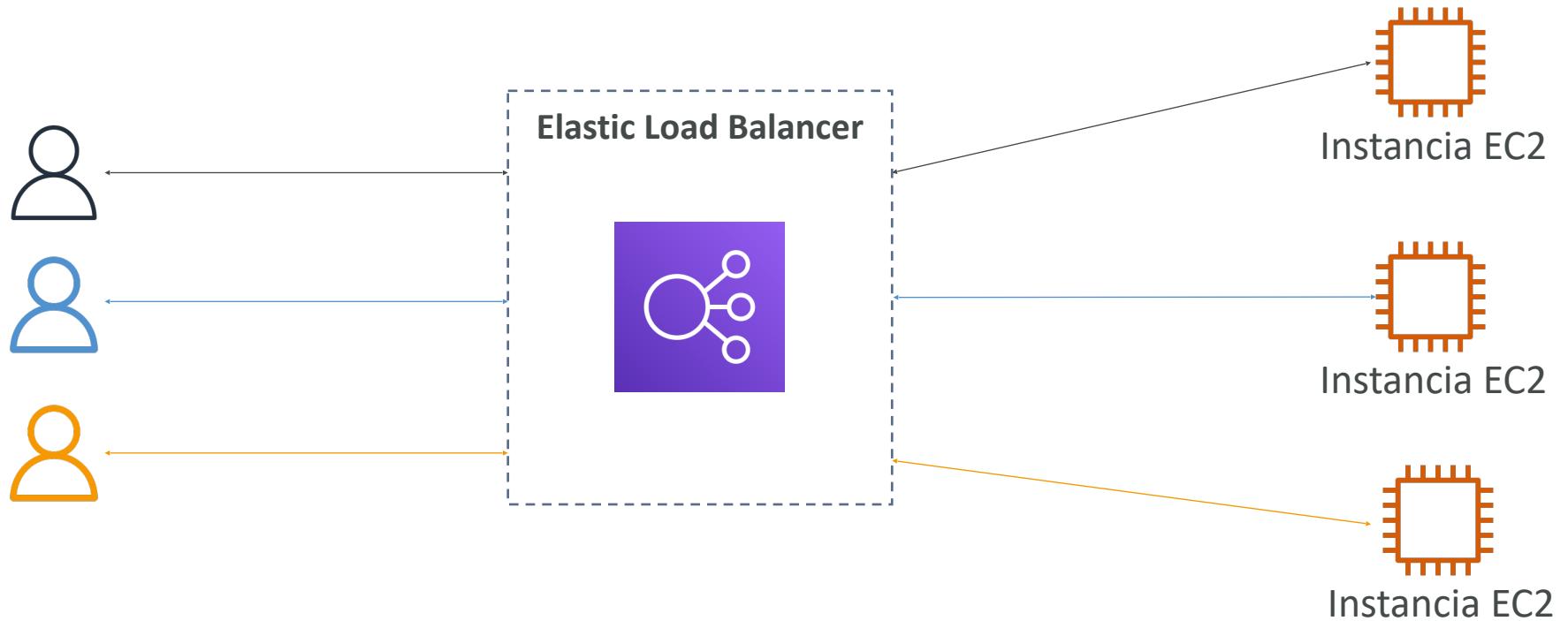


Alta disponibilidad y escalabilidad para EC2

- Escalado vertical: Aumentar el tamaño de la instancia (= escalar hacia arriba/hacia abajo)
 - De: t2.nano - 0,5G de RAM, 1 vCPU
 - A: u-12tb1.metal - 12,3 TB de RAM, 448 vCPUs
- Escalado horizontal: Aumenta el número de instancias (= escalar hacia fuera / hacia dentro)
 - Grupo de Auto Scaling
 - Load Balancer
- Alta disponibilidad: Ejecuta instancias para la misma aplicación a través de múltiples AZ
 - Auto Scaling Groups multi AZ
 - Load Balancer multi AZ

¿Qué es el balanceo de carga (load balancing)?

- Los Load Balancers son servidores que reenvían el tráfico a varios servidores (por ejemplo, instancias EC2) en sentido descendente



¿Por qué utilizar un Load Balancer?

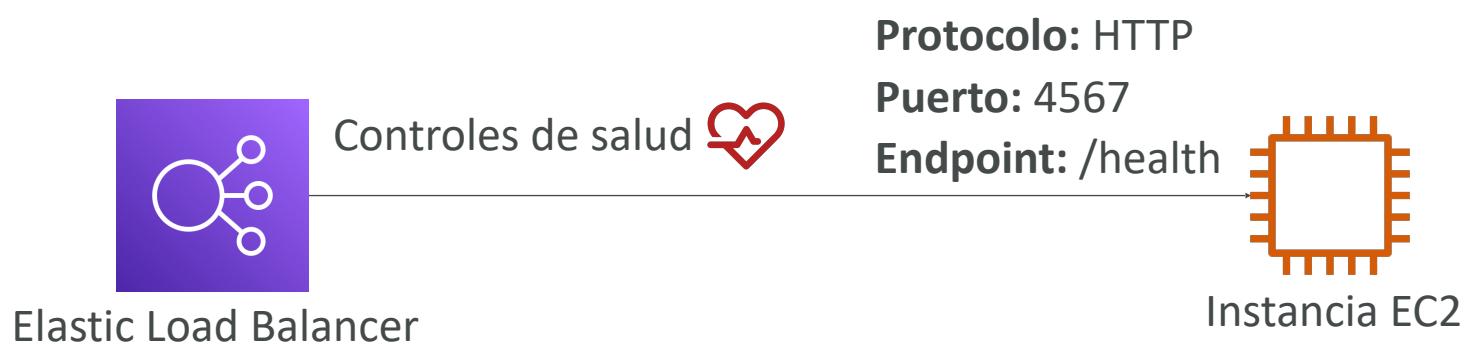
- Repartir la carga entre varias instancias descendentes
- Exponer un único punto de acceso (DNS) a tu aplicación
- Manejar sin problemas los fallos de las instancias descendentes
- Realiza comprobaciones periódicas de la salud de tus instancias
- Proporcionar terminación SSL (HTTPS) para tus sitios web
- Imponer la adherencia con las cookies
- Alta disponibilidad entre zonas
- Separar el tráfico público del privado

¿Por qué utilizar un Elastic Load Balancer?

- Un Elastic Load Balancer es un [equilibrador de carga gestionado](#)
 - AWS garantiza su funcionamiento
 - AWS se encarga de las actualizaciones, el mantenimiento y la alta disponibilidad
 - AWS sólo proporciona unos pocos mandos de configuración
- Cuesta poco configurar tu propio balanceador de carga, y te supondrá una mejora en la disponibilidad y escalabilidad
- Está integrado con muchas ofertas/servicios de AWS
 - EC2, EC2 Auto Scaling Groups, Amazon ECS
 - AWS Certificate Manager (ACM), CloudWatch
 - Route 53, AWS WAF, AWS Global Accelerator

Controles de salud

- Las comprobaciones de salud son cruciales para los Load Balancer
- Permiten al Load Balancer saber si las instancias a las que reenvía el tráfico están disponibles para responder a las peticiones
- La comprobación de salud se realiza en un puerto y una ruta (/health es común)
- Si la respuesta no es 200 (OK), la instancia no está sana

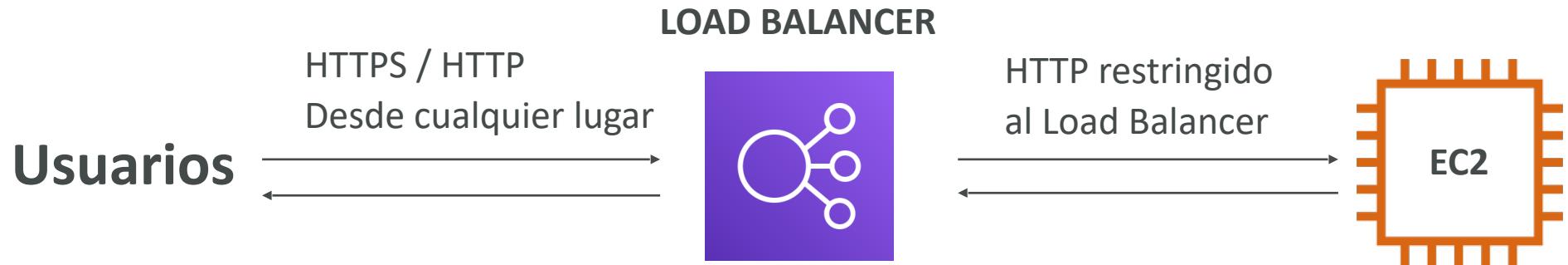




Tipos de Load Balancer en AWS

- AWS tiene **4 tipos de Load Balancer gestionados**
- **Classic Load Balancer** (v1 - antigua generación) - 2009 - CLB
 - HTTP, HTTPS, TCP, SSL (TCP seguro)
- **Application Load Balancer** (v2 - nueva generación) - 2016 - ALB
 - HTTP, HTTPS, WebSocket
- **Network Load Balancer** (v2 - nueva generación) - 2017 - NLB
 - TCP, TLS (TCP seguro), UDP
- **Gateway Load Balancer** - 2020 - GWLB
 - Funciona en la capa 3 (capa de red) - Protocolo IP
- En general, se recomienda utilizar los load balancer de nueva generación, ya que ofrecen más funciones
- Algunos Load Balancer pueden configurarse como ELB internos (privados) o externos (públicos)

Grupos de seguridad del Load Balancer



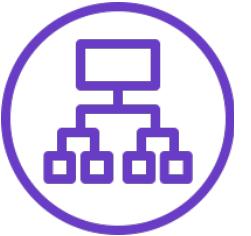
Grupo de seguridad del Load Balancer:

Type <small>i</small>	Protocol <small>i</small>	Port Range <small>i</small>	Source <small>i</small>	Description <small>i</small>
HTTP	TCP	80	0.0.0.0/0	Allow HTTP from an...
HTTPS	TCP	443	0.0.0.0/0	Allow HTTPS from a...

Grupo de seguridad de aplicaciones: Permitir el tráfico sólo desde el Load Balancer

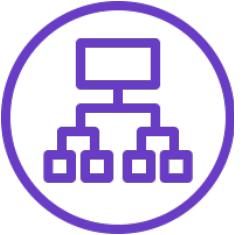
Type <small>i</small>	Protocol <small>i</small>	Port Range <small>i</small>	Source <small>i</small>	Description <small>i</small>
HTTP	TCP	80	sg-054b5ff5ea02f2b6e (load-b	Allow Traffic only...

Application Load Balancer (v2)



- El Application Load Balancer es de capa 7 (HTTP)
- Equilibrio de carga para múltiples aplicaciones HTTP en distintas máquinas (grupos de destino)
- Equilibrio de carga para múltiples aplicaciones en la misma máquina (por ejemplo, contenedores)
- Soporte para HTTP/2 y WebSocket
- Soporta redireccionamientos (de HTTP a HTTPS, por ejemplo)

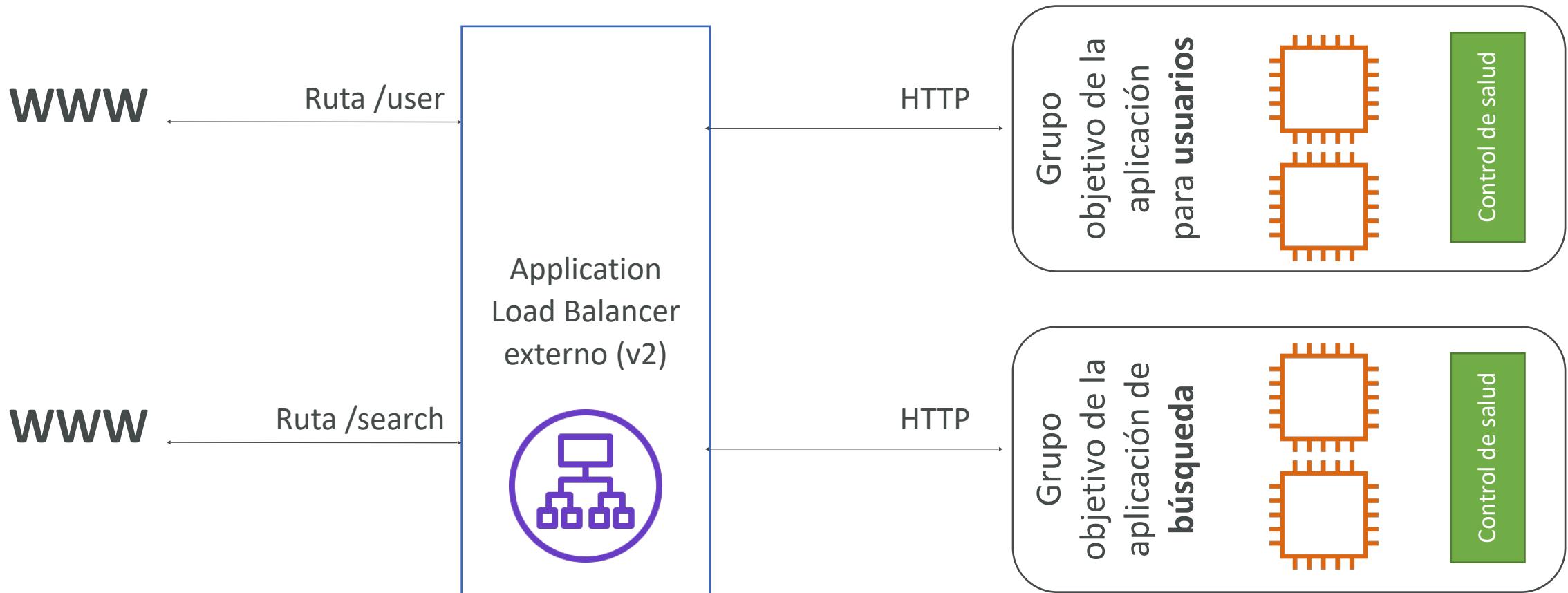
Application Load Balancer (v2)



- Tablas de enrutamiento a diferentes grupos de destino:
 - Enrutamiento basado en la ruta en la URL (example.com/users & example.com/posts)
 - Enrutamiento basado en el nombre de host en la URL (one.example.com & other.example.com)
 - Enrutamiento basado en la cadena de consulta, las cabeceras (example.com/users?id=123&order=false)
- Los ALB son muy adecuados para los microservicios y las aplicaciones basadas en contenedores (ejemplo: Docker y Amazon ECS)
- Tiene una función de mapeo de puertos para redirigir a un puerto dinámico en ECS
- En comparación, necesitaríamos varios Classic Load Balancer por aplicación

Application Load Balancer (v2)

Tráfico basado en HTTP



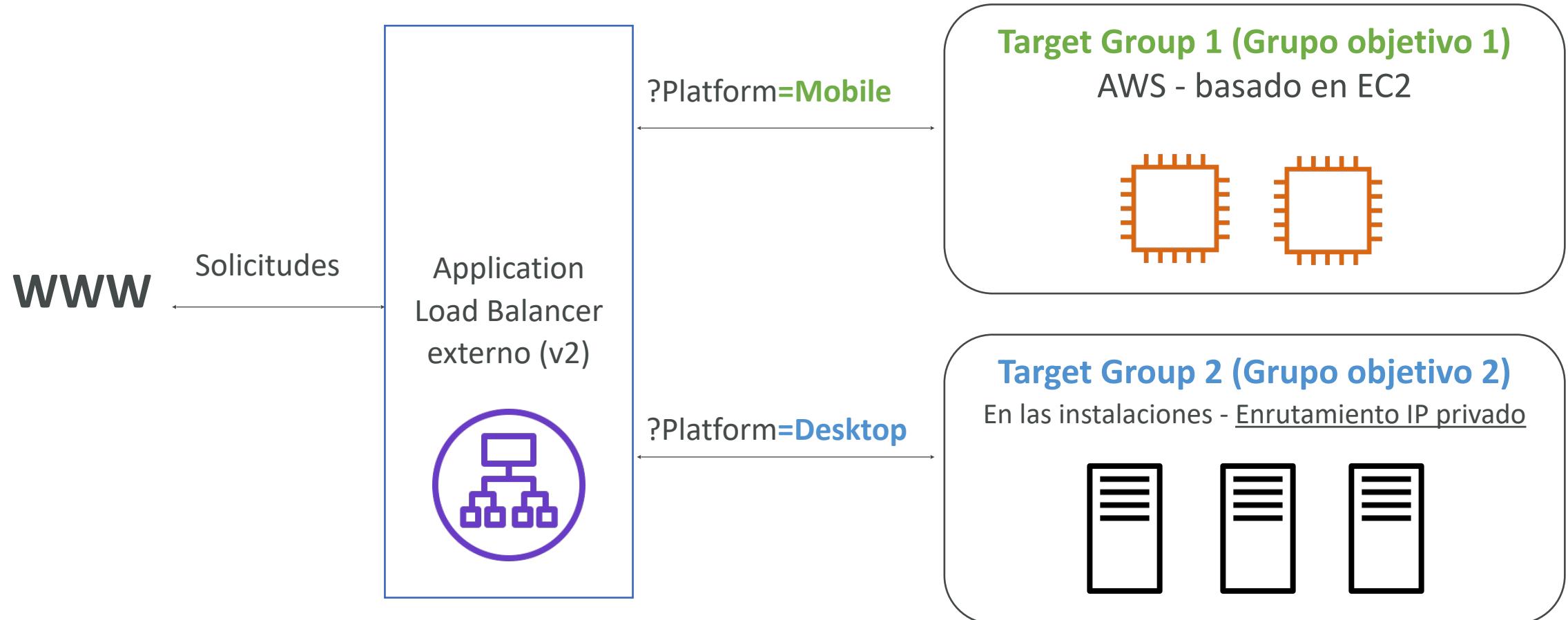
Application Load Balancer (v2)

Target Groups (Grupos objetivo)

- Instancias EC2 (pueden ser gestionadas por un Auto Scaling Groups) - HTTP
 - Tareas de ECS (gestionadas por el propio ECS) - HTTP
 - Funciones Lambda - La petición HTTP se traduce en un evento JSON
 - Direcciones IP - deben ser IPs privadas
-
- El ALB puede enrutar a múltiples grupos de destino
 - Las comprobaciones de salud son a nivel de grupo de destino

Application Load Balancer (v2)

Strings de consulta/enrutamiento de parámetros

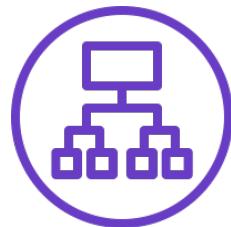


Application Load Balancer (v2)

Es bueno saber que

- Nombre de host fijo (XXX.region.elb.amazonaws.com)
- Los servidores de aplicaciones no ven directamente la IP del cliente
 - La verdadera IP del cliente se inserta en la cabecera X-Forwarded-For
 - También podemos obtener el puerto (X-Forwarded-Port) y el protocolo (X-Forwarded-Proto)

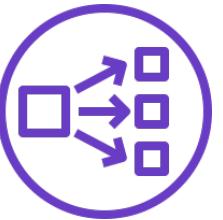
IP del cliente
12.34.56.78



IP del Load Balancer
(IP privada)

Terminación de la conexión



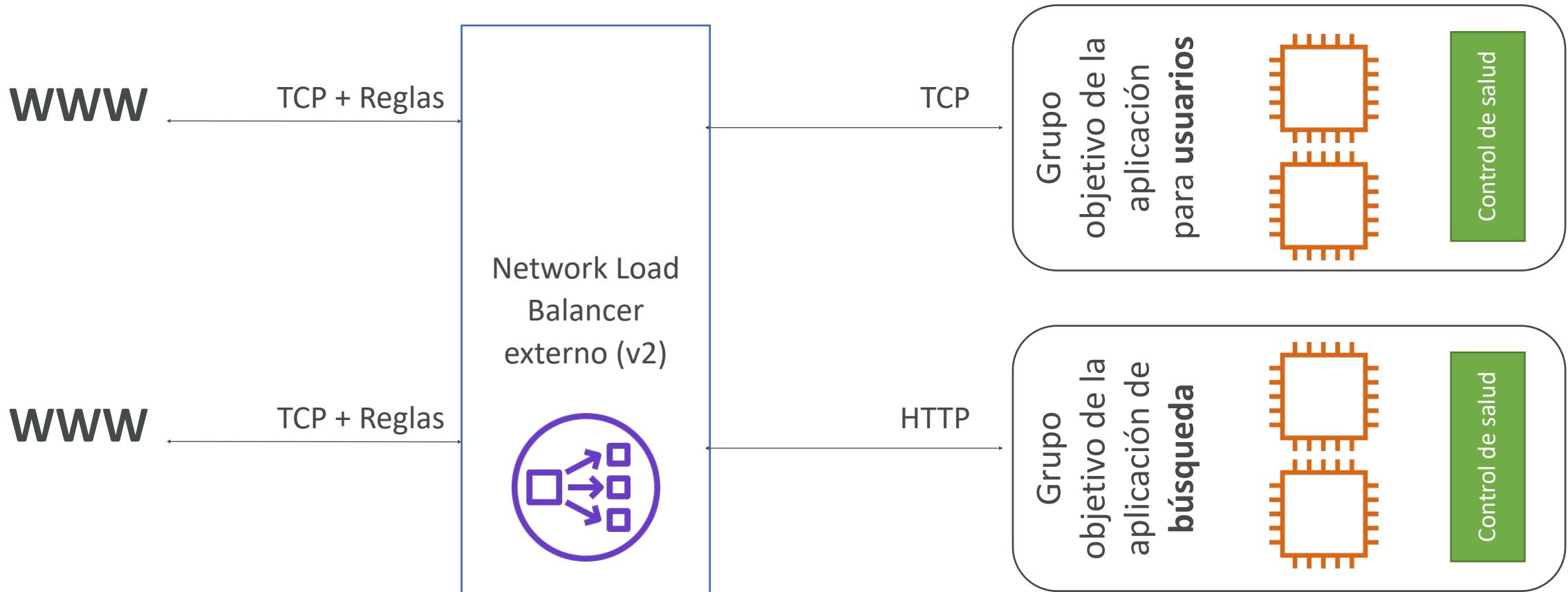


Network Load Balancer (v2)

- Los Network Load Balancer (Capa 4) permiten:
 - **Reenviar el tráfico TCP y UDP a tus instancias**
 - Manejar millones de peticiones por segundo
 - Menor latencia ~100 ms (frente a los 400 ms del ALB)
- **El NLB tiene una IP estática por AZ, y soporta la asignación de IP elástica**
(útil para poner en lista blanca una IP específica)
- Los NLB se utilizan para un rendimiento extremo, tráfico TCP o UDP
- **No está incluido en la capa gratuita de AWS**

Network Load Balancer (v2)

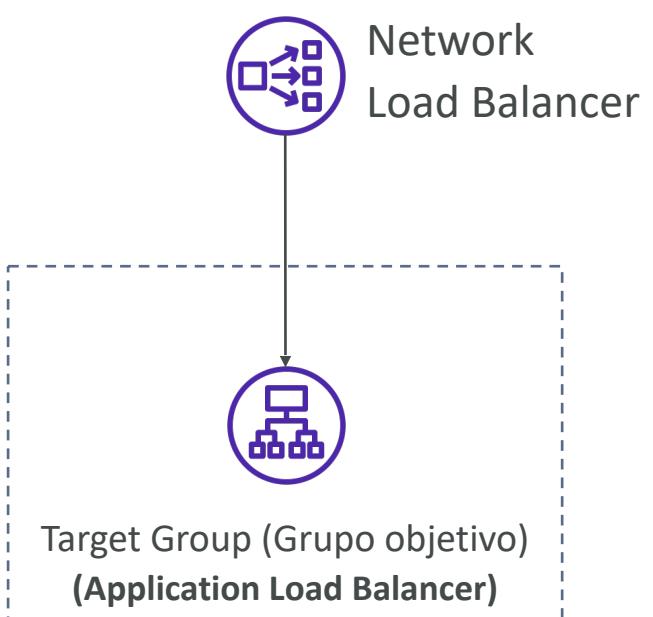
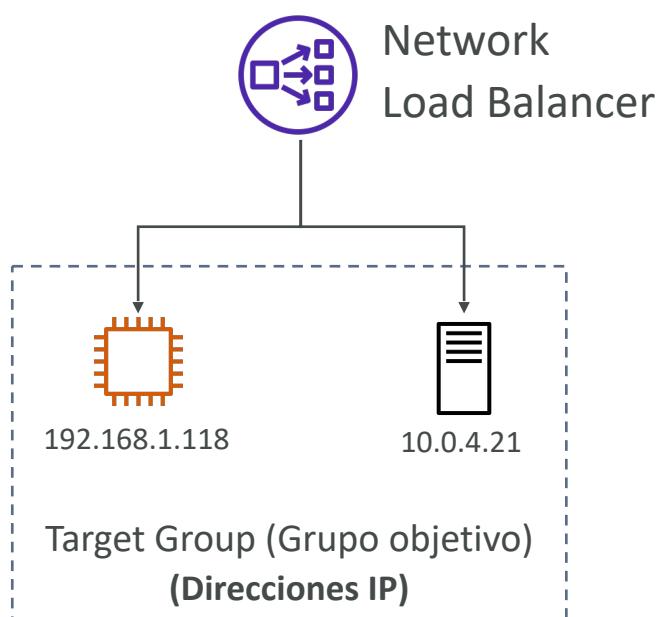
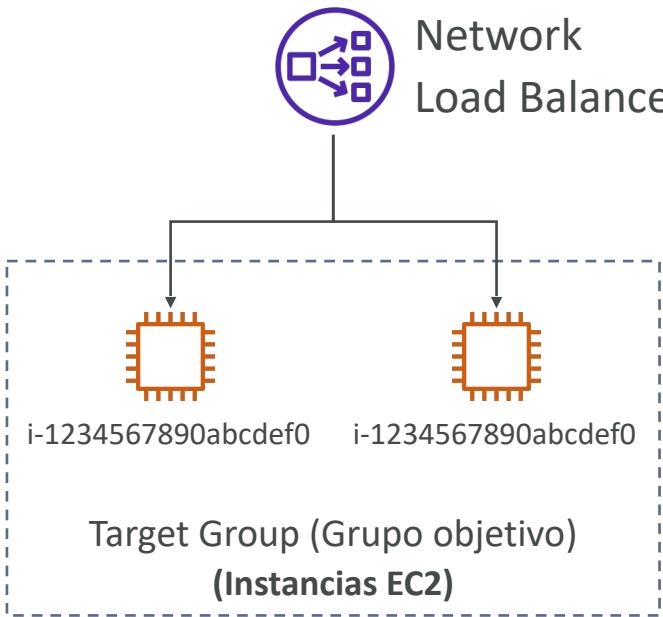
Tráfico basado en TCP (capa 4)



Network Load Balancer

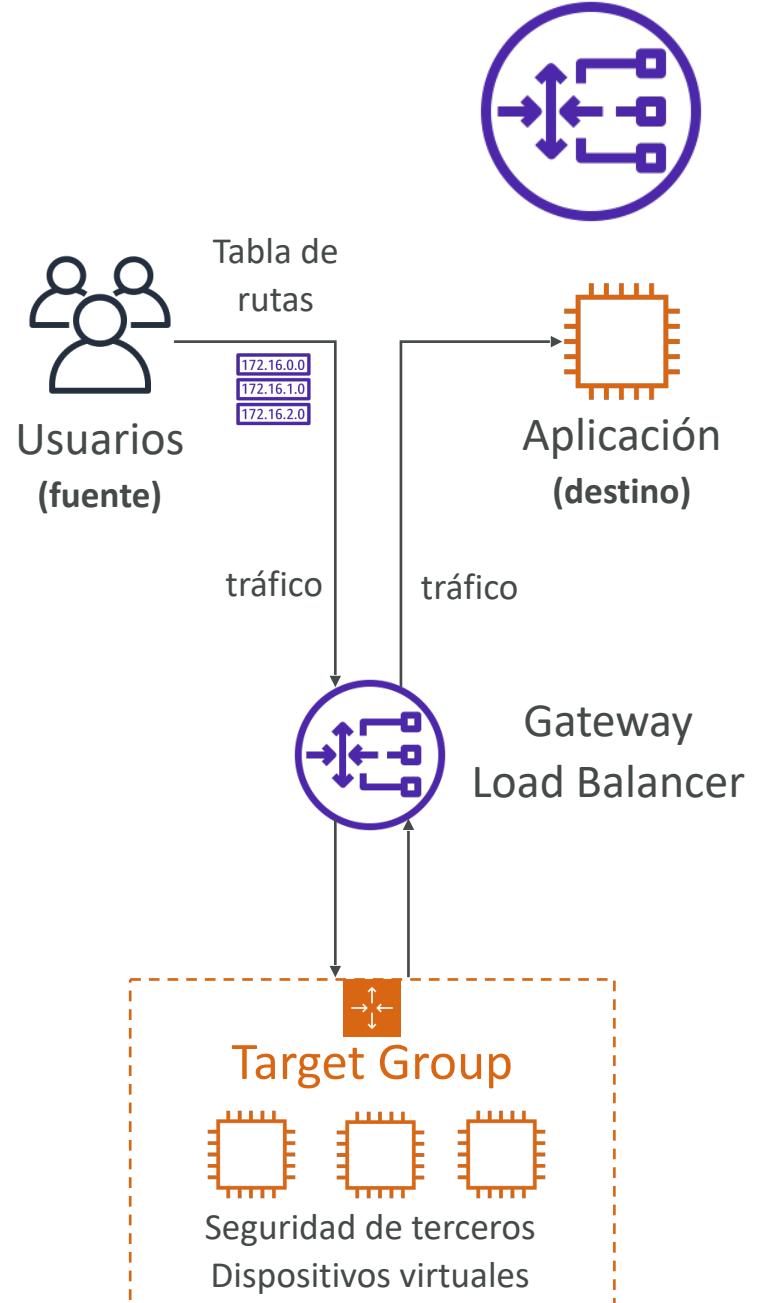
Target Groups (Grupos objetivo)

- **Instancias EC2**
- **Direcciones IP** - deben ser IPs privadas
- **Application Load Balancer**
- Los controles de salud soportan los **protocolos TCP, HTTP y HTTPS**



Gateway Load Balancer

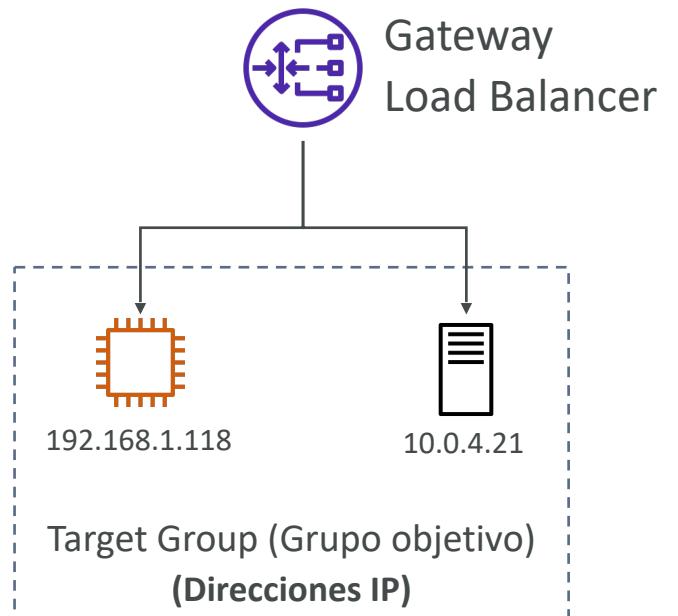
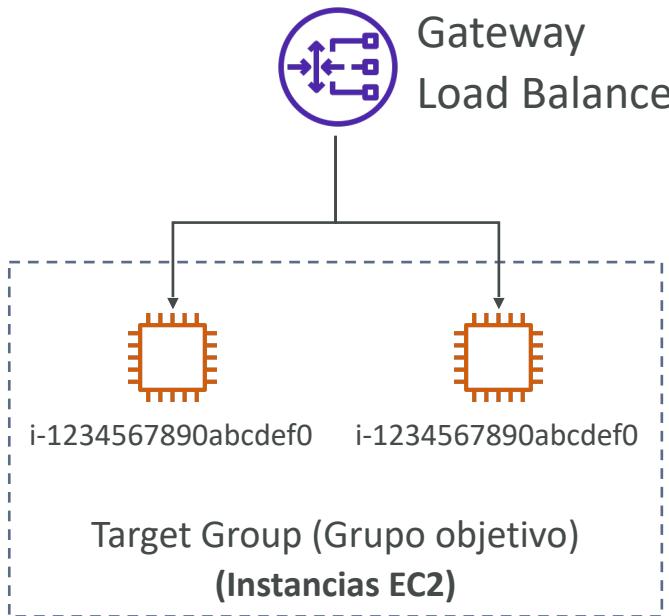
- Implementa, escala y administra una flota de dispositivos virtuales de red de terceros en AWS
- Ejemplo: Firewalls, Sistemas de Detección y Prevención de Intrusiones, Sistemas de Inspección Profunda de Paquetes, manipulación de cargas útiles, ...
- Opera en la Capa 3 (Capa de Red) - Paquetes IP
- Combina las siguientes funciones
 - **Gateway de Red Transparente** - entrada/salida única para todo el tráfico
 - **Load Balancer** - distribuye el tráfico a tus dispositivos virtuales
- Utiliza el protocolo **GENEVE** en el puerto 6081



Gateway Load Balancer

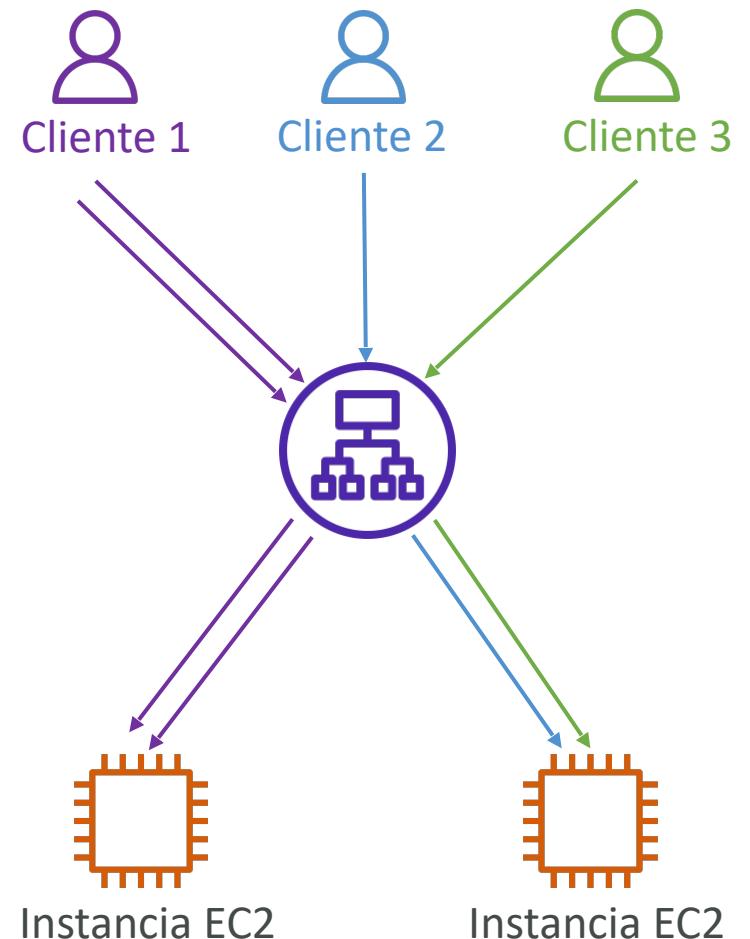
Target Groups (Grupos objetivo)

- **Instancias EC2**
- **Direcciones IP** - deben ser IPs privadas



Sticky Sessions (sesiones persistentes)

- Es posible implementar la “pegajosidad / adherencia” para que el mismo cliente sea siempre redirigido a la misma instancia detrás de un balanceador de carga
- Esto funciona para los Classic Load Balancer y los Application Load Balancer
- La "cookie" utilizada para la adherencia tiene una fecha de caducidad que tú controlas
- Caso de uso: asegurarse de que el usuario no pierde sus datos de sesión
- Activar la adherencia puede provocar un desequilibrio de la carga en las instancias EC2 del backend



Sticky Sessions – Nombres de las cookies

- **Cookies basadas en la aplicación**

- Cookie personalizada
 - Generada por el objetivo
 - Puede incluir cualquier atributo personalizado requerido por la aplicación
 - El nombre de la cookie debe especificarse individualmente para cada grupo de destino
 - No utilices **AWSALB**, **AWSALBAPP** o **AWSALBTG** (reservadas para el uso del ELB)
- Cookie de la aplicación
 - Generada por el Load Balancer
 - El nombre de la cookie es **AWSALBAPP**

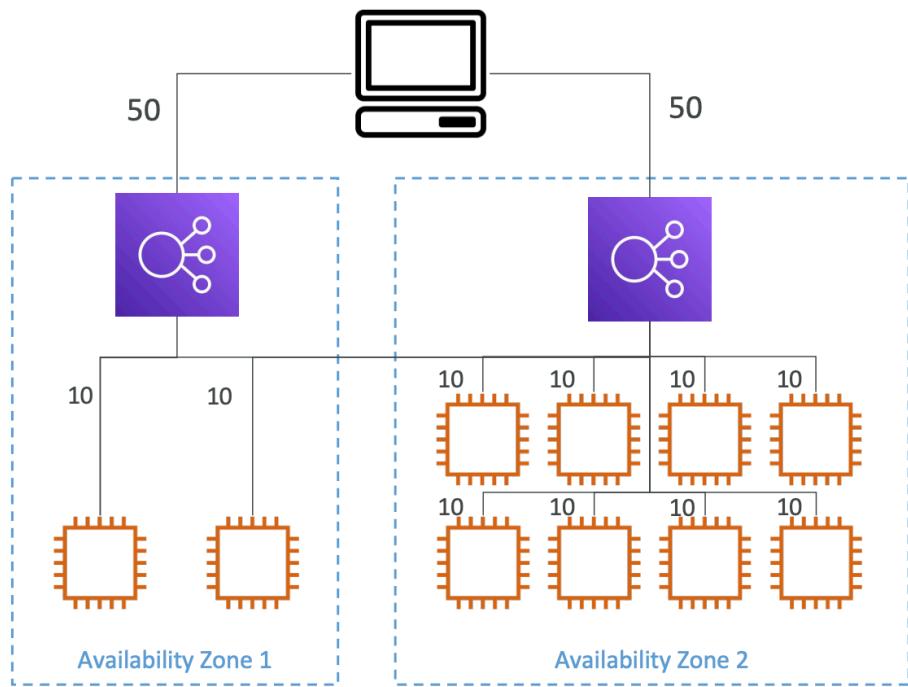
- **Cookies basadas en la duración**

- Cookie generada por el equilibrador de carga
- El nombre de la cookie es **AWSALB** para ALB, **AWSELB** para CLB

Load Balancer entre zonas

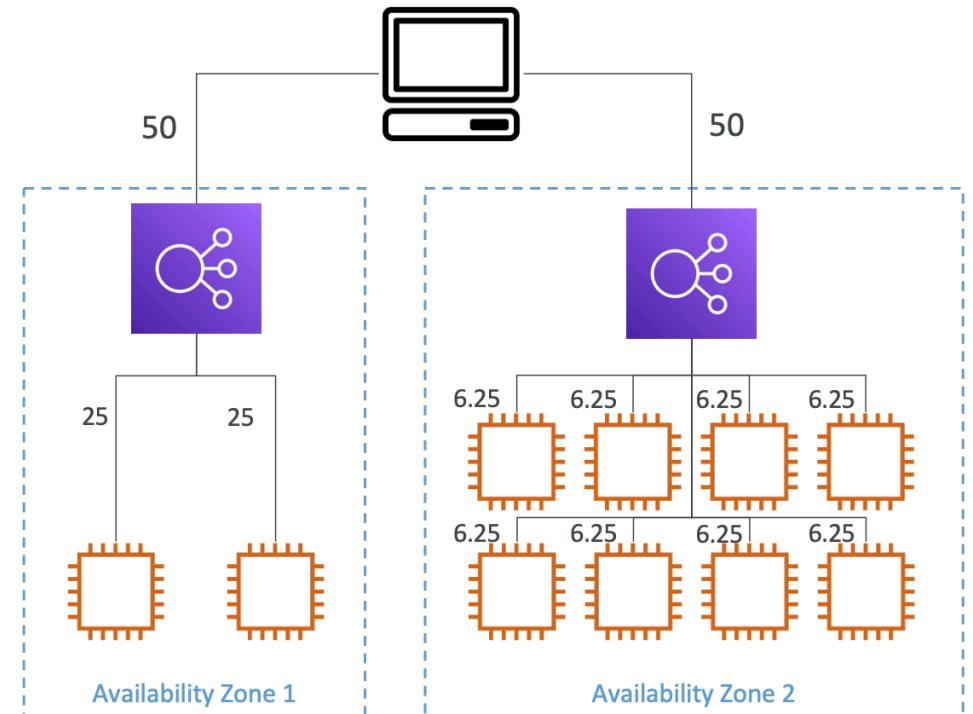
Con el Load Balancer de zona cruzada:

Cada instancia del Load Balancer distribuye uniformemente entre todas las instancias registradas en todas las AZ



Sin Load Balancer de zona cruzada:

Las solicitudes se distribuyen en las instancias del nodo del Elastic Load Balancer



Load Balancer entre zonas

- **Application Load Balancer**

- Siempre activado (no se puede desactivar)
- No se cobra por los datos inter AZ

- **Network Load Balancer & Gateway Load Balancer**

- Desactivado por defecto
- Si está activado, pagas una tarifa (\$) por los datos entre zonas geográficas

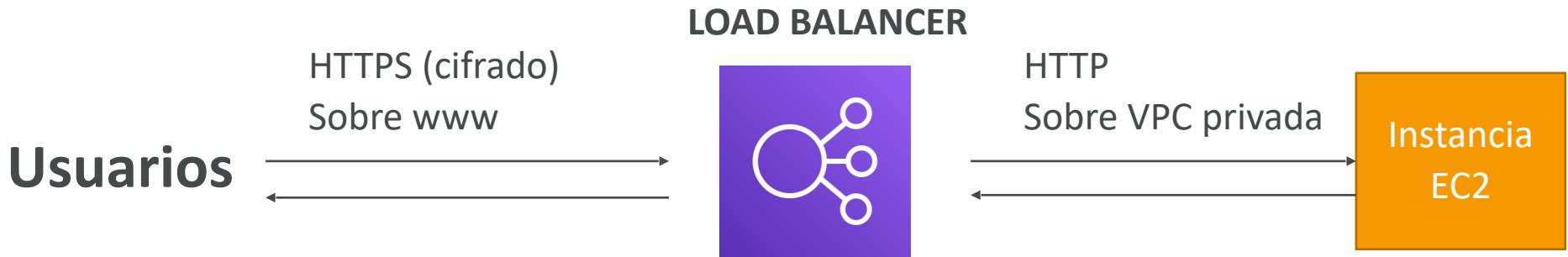
- **Classic Load Balancer**

- Desactivado por defecto
- No se cobra por los datos inter AZ si está activado

SSL/TLS - Conceptos básicos

- Un certificado SSL permite que el tráfico entre tus clientes y tu Load Balancer esté cifrado en tránsito (cifrado en vuelo)
- **SSL** hace referencia a Secure Sockets Layer, que se utiliza para cifrar las conexiones
- **TLS** se refiere a Transport Layer Security, que es una versión más reciente
- Hoy en día, **se utilizan principalmente los certificados TLS**, pero la gente sigue refiriéndose a ellos como SSL
- Los certificados SSL públicos son emitidos por las Autoridades de Certificación (CA)
- Comodo, Symantec, GoDaddy, GlobalSign, DigiCert, LetsEncrypt, etc.
- Los certificados SSL tienen una fecha de caducidad (que tú estableces) y deben ser renovados

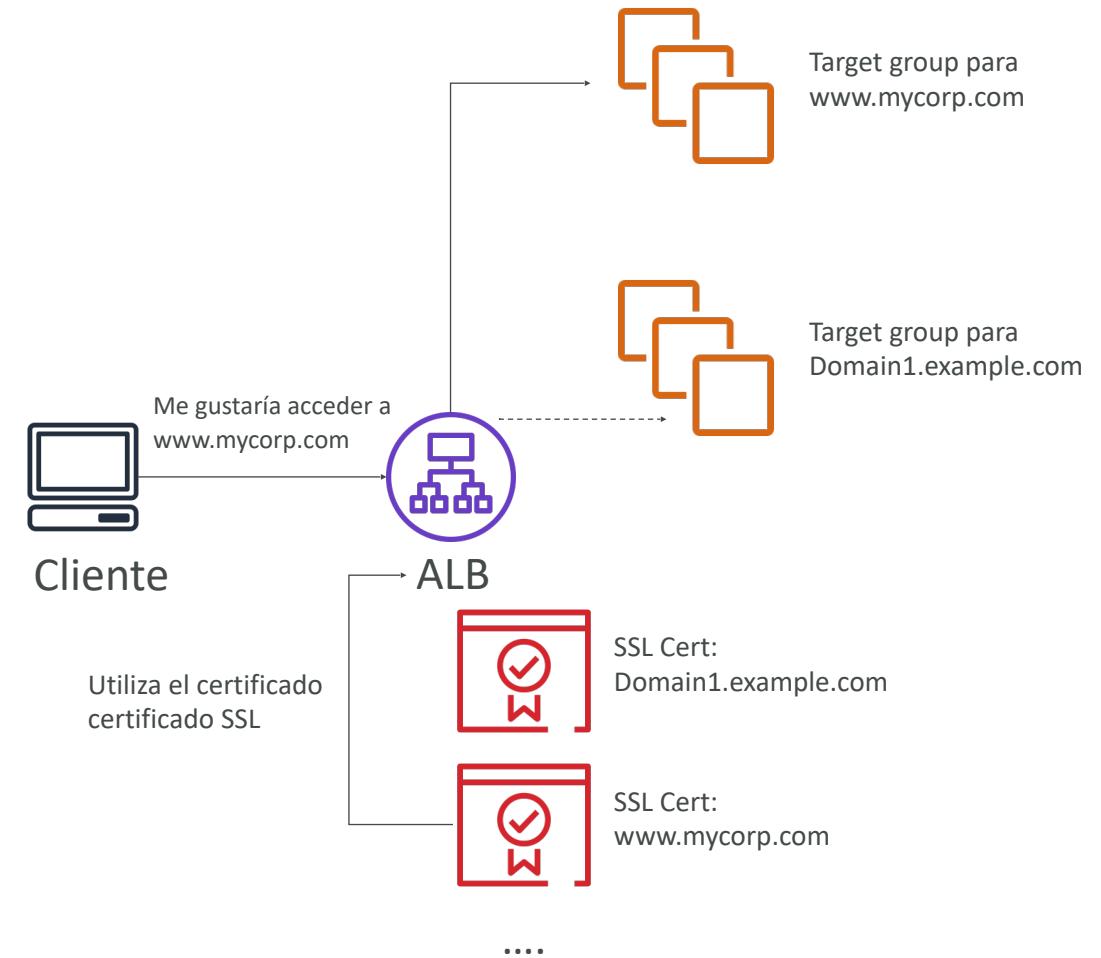
Load Balancer - Certificados SSL



- El Load Balancer utiliza un certificado X.509 (certificado de servidor SSL/TLS)
- Puedes gestionar los certificados mediante ACM (AWS Certificate Manager)
- También puedes crear y subir tus propios certificados
- Listener HTTPS:
 - Debes especificar un certificado por defecto
 - Puedes añadir una lista opcional de certificados para dar soporte a múltiples dominios
 - **Los clientes pueden utilizar SNI (Server Name Indication) para especificar el nombre de host al que llegan**
 - Posibilidad de especificar una política de seguridad para soportar versiones antiguas de SSL / TLS (clientes heredados)

SSL – Server Name Indication (SNI)

- SNI resuelve el problema de **cargar varios certificados SSL en un servidor web** (para servir a varios sitios web)
- Es un protocolo "más nuevo", y requiere que el cliente **indique** el nombre del servidor de destino en el apretón de manos SSL inicial (*handshake*)
- El servidor encontrará entonces el certificado correcto, o devolverá el predeterminado
- Nota:
 - Sólo funciona para ALB y NLB (generación más reciente), CloudFront
 - No funciona con CLB (generación anterior)



Elastic Load Balancers - Certificados SSL

- **Classic Load Balancer (v1)**

- Soporta sólo un certificado SSL
- Debe utilizar varios CLB para varios nombres de host con varios certificados SSL

- **Application Load Balancer (v2)**

- Soporta múltiples oyentes con múltiples certificados SSL
- Utiliza la indicación del nombre del servidor (SNI) para que funcione

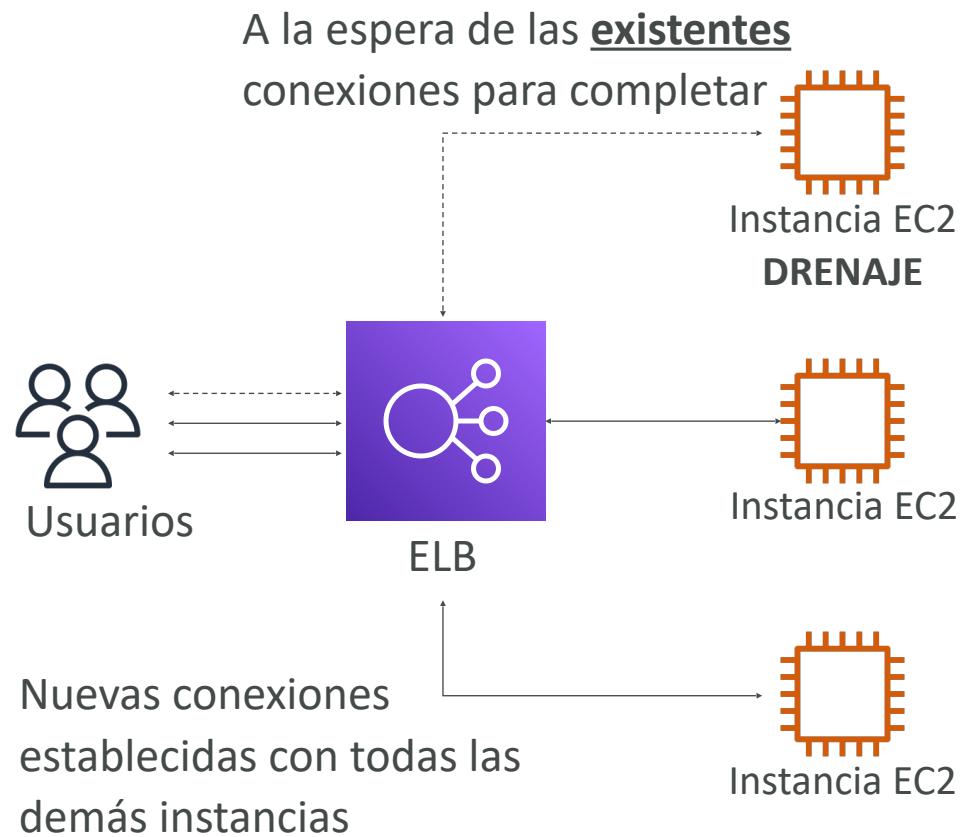
- **Network Load Balancer (v2)**

- Soporta múltiples oyentes con múltiples certificados SSL
- Utiliza la Indicación del Nombre del Servidor (SNI) para hacerlo funcionar

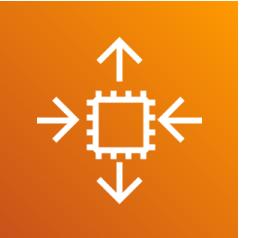
Connection Draining (Drenaje de la conexión)

- **Nombre de la característica**

- Drenaje de la conexión - para el CLB
- Retraso en el desregistro - para ALB y NLB
- Tiempo para completar las "peticiones en vuelo" mientras la instancia se está desregistrando o no está sana
- Deja de enviar nuevas peticiones a la instancia EC2 que se está desregistrando
- Entre 1 y 3600 segundos (por defecto: 300 segundos)
- Se puede desactivar (fijar el valor en 0)
- Establece un valor bajo si tus peticiones son cortas

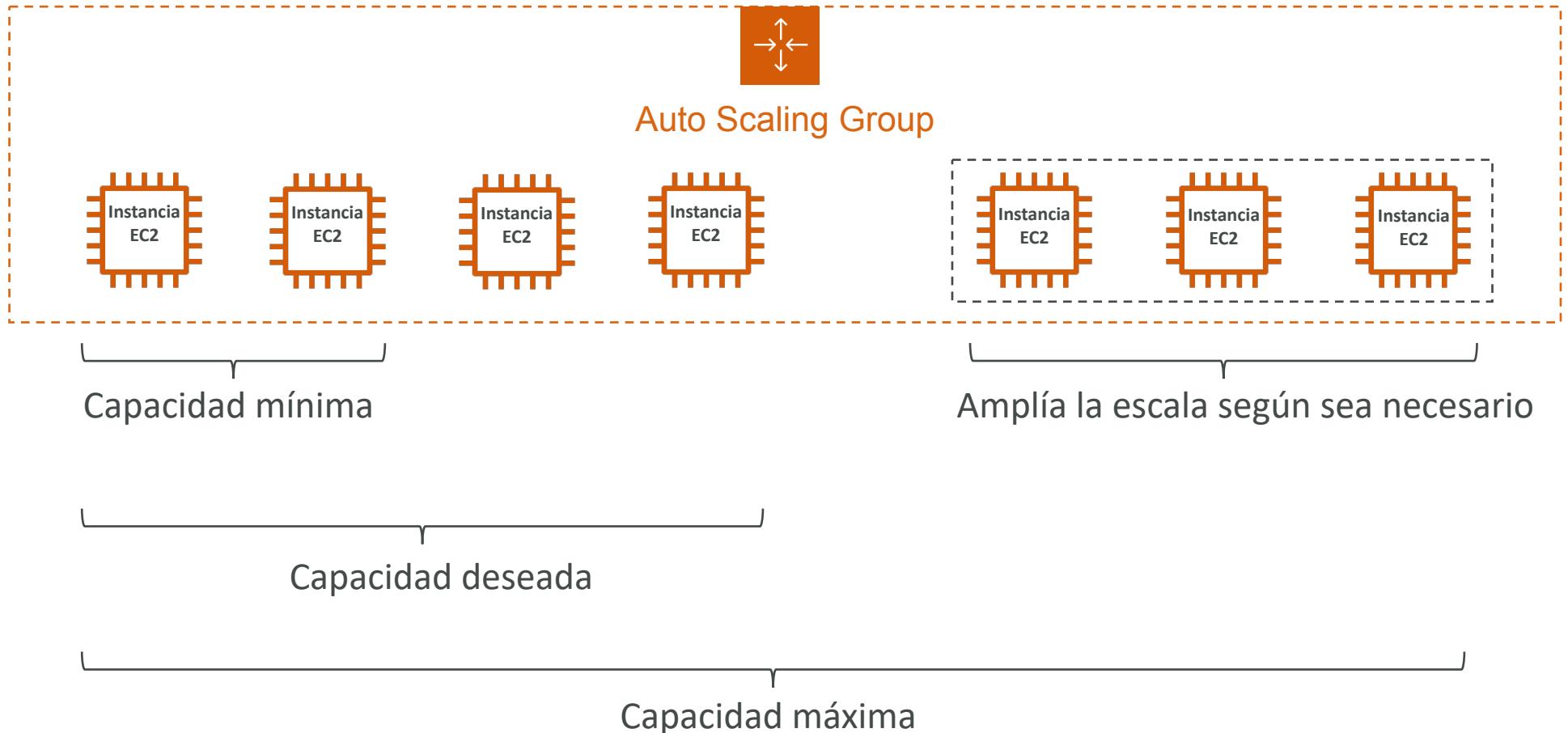


¿Qué es un Auto Scaling Group?

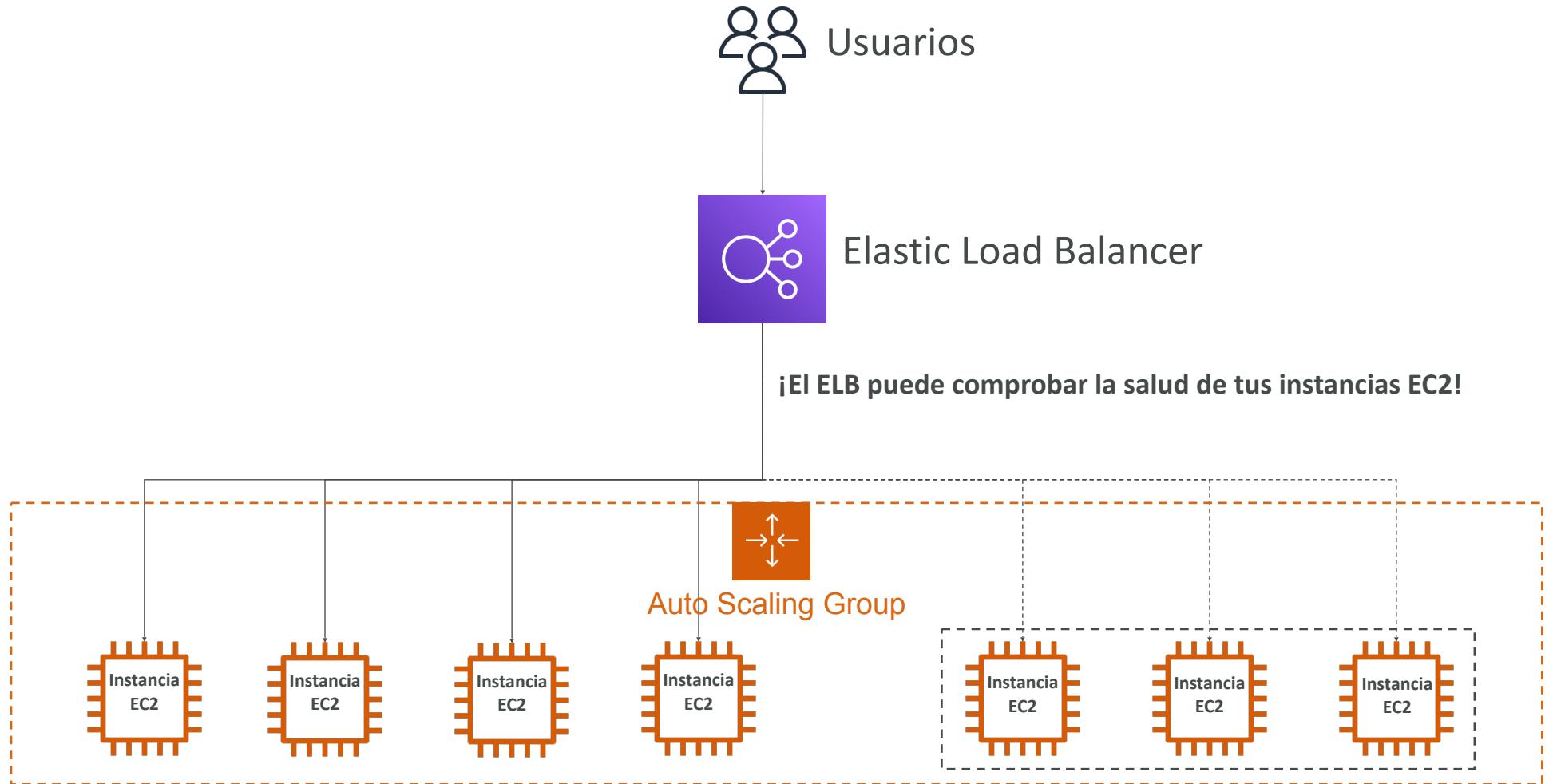


- En la vida real, la carga de tus sitios web y aplicaciones puede cambiar
- En el Cloud, puedes crear y deshacerte de servidores muy rápidamente
- El objetivo de un Auto Scaling Group (ASG) es
 - Reducir (añadir instancias EC2) para adaptarse a un aumento de la carga
 - Aumentar (eliminar instancias EC2) para que coincida con una disminución de la carga
 - Asegurar que tenemos un número mínimo y máximo de instancias EC2 en funcionamiento
 - Registrar automáticamente nuevas instancias en un Load Balancer
 - Volver a crear una instancia EC2 en caso de que se elimine una anterior (por ejemplo, si no está sana)
- Los ASG son gratuitos (sólo pagas por las instancias EC2 subyacentes)

Auto Scaling Group en AWS



Auto Scaling Group en AWS con Load Balancer



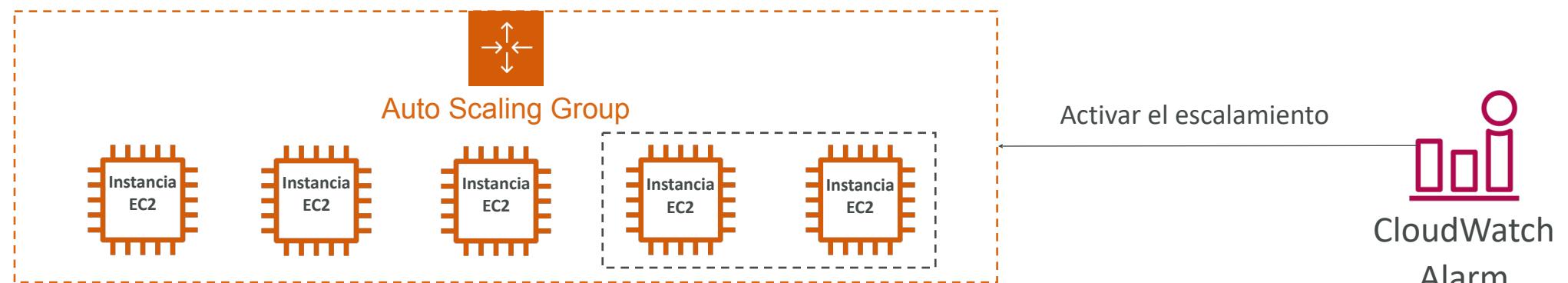
Atributos del Auto Scaling Group

- Una **Plantilla de Lanzamiento** (las antiguas "Configuraciones de Lanzamiento" están obsoletas)
 - AMI + Tipo de Instancia
 - Datos de usuario de EC2
 - Volúmenes EBS
 - Grupos de seguridad
 - Par de claves SSH
 - Roles IAM para tus instancias EC2
 - Información sobre la red y las subredes
 - Información del Load Balancer
- Tamaño mínimo / Tamaño máximo / Capacidad inicial
- Políticas de escalado



Escalado automático Alarmas y escalado de CloudWatch

- Es posible escalar un ASG basándose en las alarmas de CloudWatch
- Una alarma monitoriza una métrica (como la CPU media, o una métrica personalizada)
- Las métricas, como la CPU media, se calculan para todas las instancias del ASG
- Basándonos en la alarma
 - Podemos crear políticas de escalado (aumentar el número de instancias)
 - Podemos crear políticas de ampliación (reducir el número de instancias)



Auto Scaling Group

Políticas de escalado dinámico

- **Escala de seguimiento de objetivos**

- Lo más sencillo y fácil de configurar
- Ejemplo: Quiero que la media de la CPU de ASG se mantenga en torno al 40%

- **Escalado simple / escalonado**

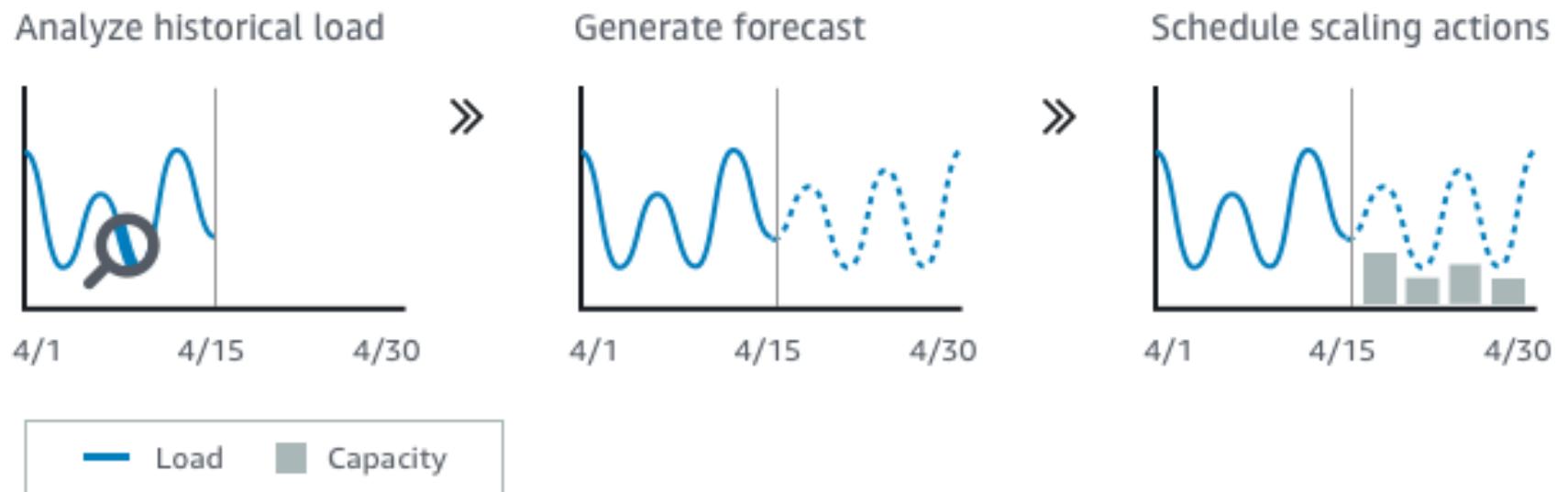
- Cuando se active una alarma de CloudWatch (por ejemplo, CPU > 70%), añade 2 unidades
- Cuando se active una alarma de CloudWatch (ejemplo CPU < 30%), entonces elimina 1

- **Acciones programadas**

- Anticipa un escalado basado en patrones de uso conocidos
- Ejemplo: aumentar la capacidad mínima a 10 a las 17 horas de los viernes

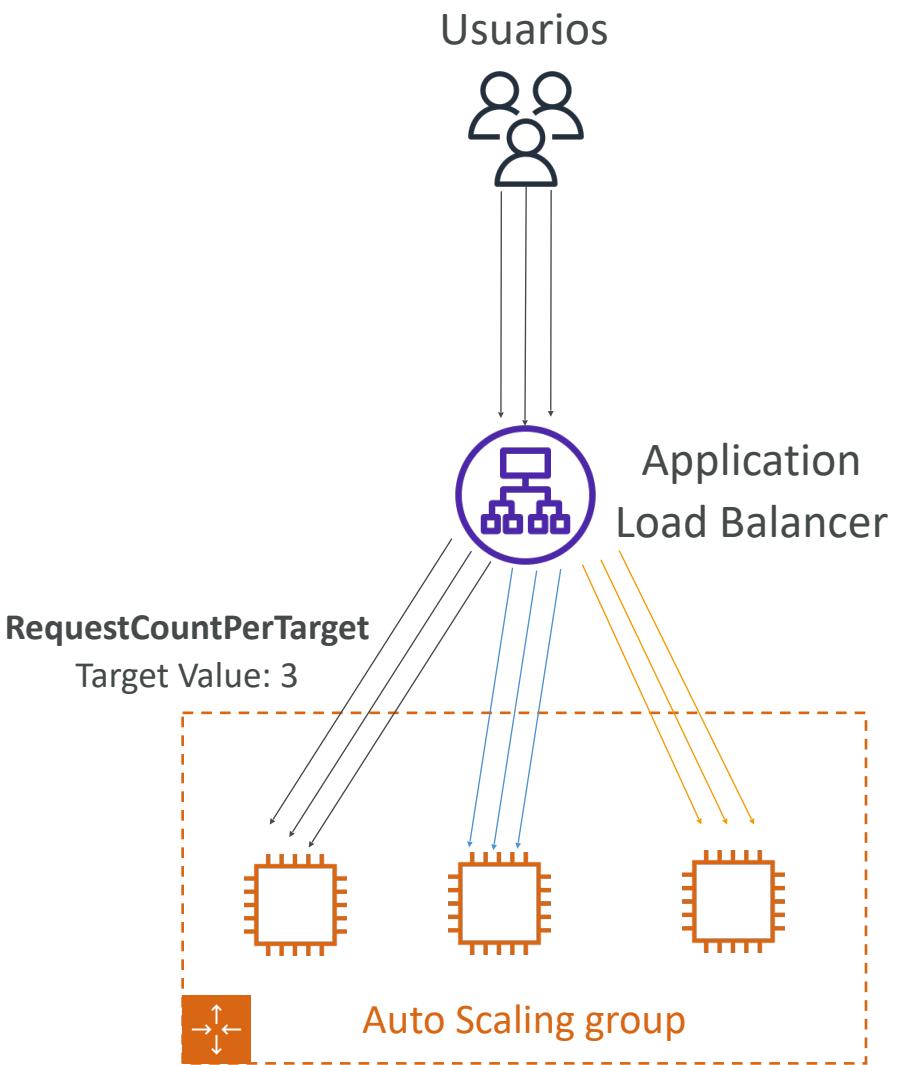
Auto Scaling Group - Escalado predictivo

- **Escalado predictivo:** previsión continua de la carga y programación del escalado por adelantado



Buenas métricas para escalar

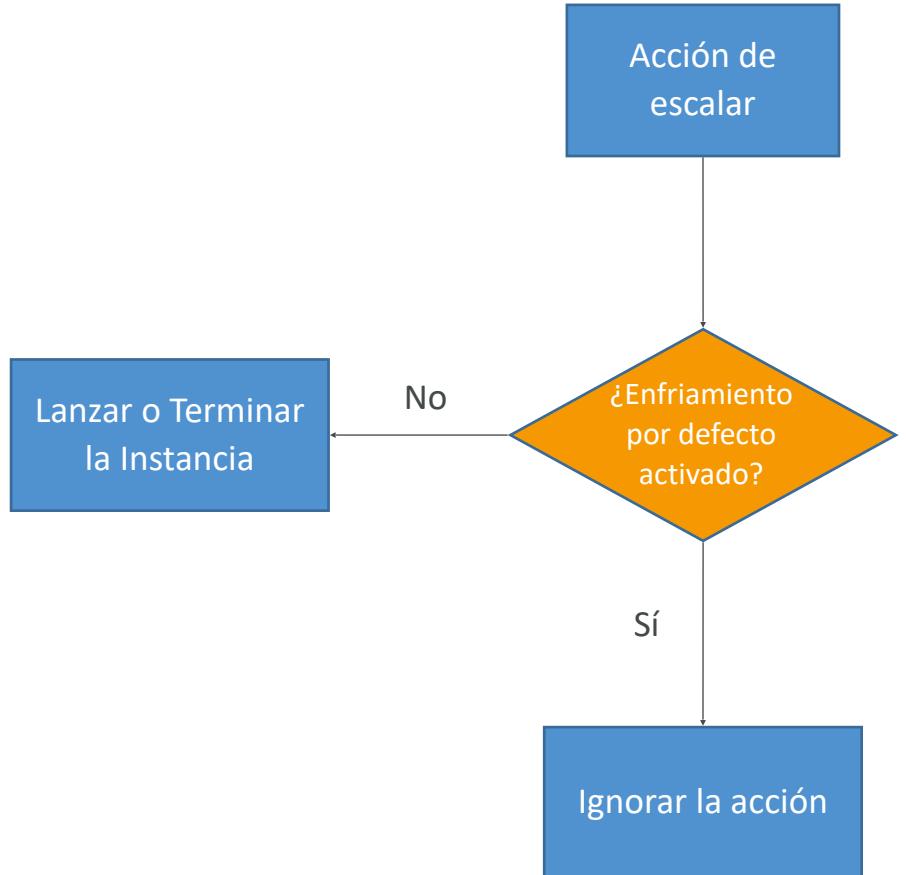
- **CPUUtilization:** Utilización media de la CPU en tus instancias
- **RequestCountPerTarget:** para asegurarse de que el número de peticiones por instancias EC2 es estable
- **Promedio de entrada/salida** de red (si tu aplicación está vinculada a la red)
- **Cualquier métrica personalizada** (que impulse con CloudWatch)



Auto Scaling Groups

Enfriamiento de la escala

- Después de que se produzca una actividad de escalado, **estarás en el periodo de enfriamiento (por defecto, 300 segundos)**
- Durante el periodo de enfriamiento, el ASG no lanzará ni terminará instancias adicionales (para permitir que las métricas se estabilicen)
- Consejo: Utiliza una AMI lista para usar para reducir el tiempo de configuración y así poder servir las peticiones más rápidamente y reducir el periodo de enfriamiento



RDS, Aurora y ElastiCache

Visión general de Amazon RDS



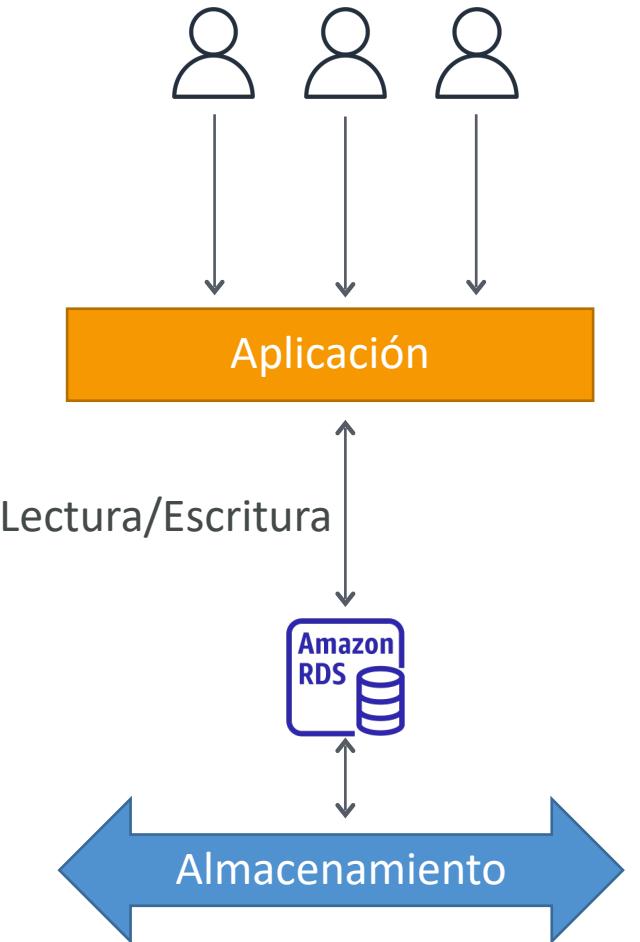
- RDS significa Servicio de Base de Datos Relacional
- Es un servicio de bases de datos gestionado para que las bases de datos utilicen SQL como lenguaje de consulta.
- Permite crear bases de datos en el Cloud que son gestionadas por AWS
 - Postgres
 - MySQL
 - MariaDB
 - Oracle
 - Microsoft SQL Server
 - Aurora (base de datos propia de AWS)

Ventaja sobre el uso de RDS frente al despliegue de la BD en EC2

- El RDS es un servicio gestionado:
 - Aprovisionamiento automatizado, parcheo del SO
 - Copias de seguridad continuas y restauración a una fecha determinada (Point in Time Restore)
 - Dashboards de monitorización
 - Rélicas de lectura para mejorar el rendimiento de lectura
 - Configuración multi AZ para DR (Disaster Recovery)
 - Ventanas de mantenimiento para actualizaciones
 - Capacidad de escalado (vertical y horizontal)
 - Almacenamiento respaldado por EBS (gp2 o io1)
- **PERO no puedes acceder por SSH a tus instancias**

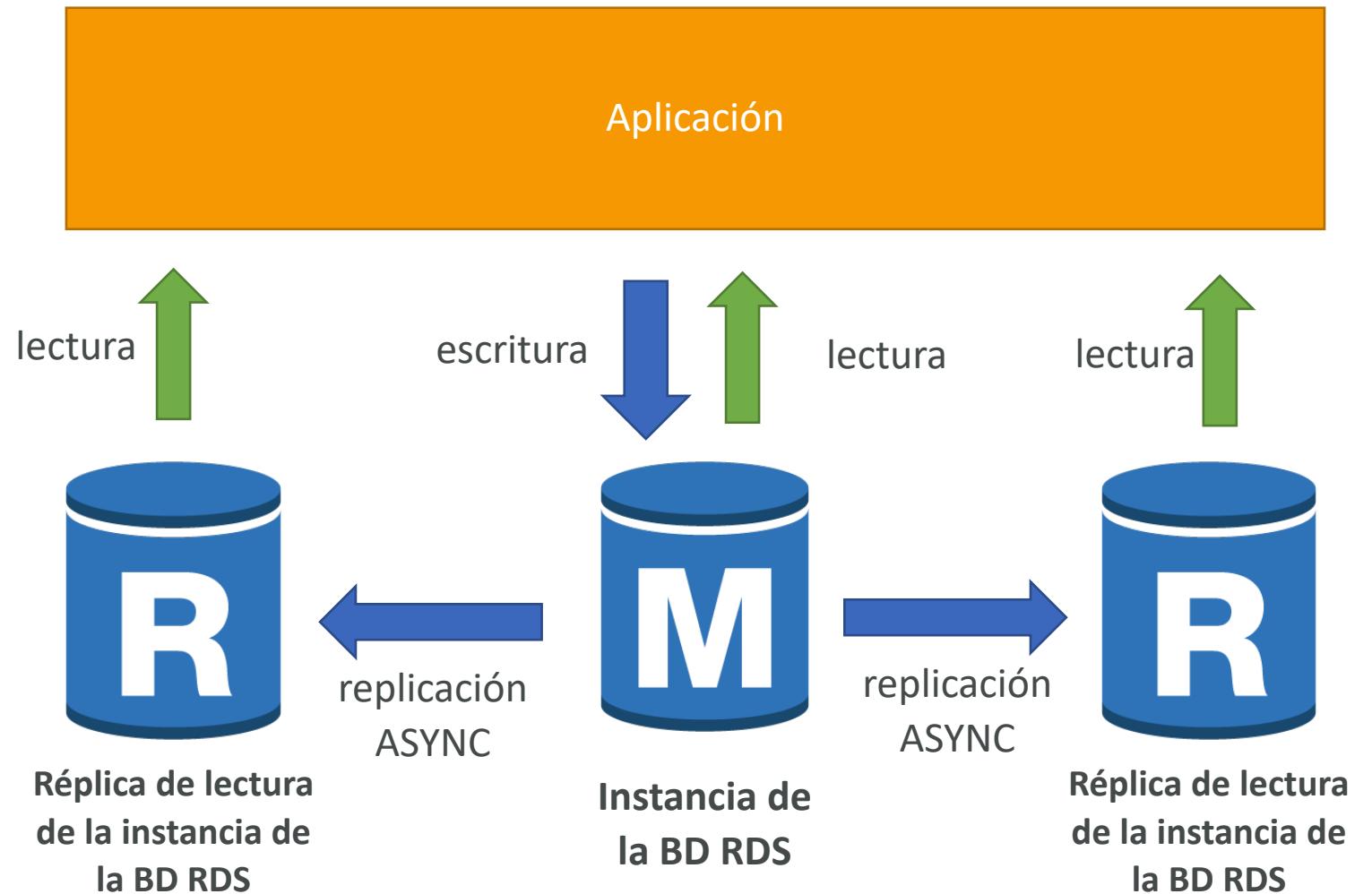
RDS – Autoescalamiento de almacenamiento

- Te ayuda a aumentar el almacenamiento de tu instancia de base de datos RDS de forma dinámica
- Cuando RDS detecta que te estás quedando sin almacenamiento gratuito en la base de datos, escala automáticamente
- Evita escalar manualmente el almacenamiento de tu base de datos
- Tienes que establecer el **Umbral Máximo de Almacenamiento** (límite máximo de almacenamiento de la BD)
- Modifica automáticamente el almacenamiento si
 - El almacenamiento gratuito es inferior al 10% del almacenamiento asignado
 - El almacenamiento bajo dura al menos 5 minutos
 - Han pasado 6 horas desde la última modificación
- Útil para aplicaciones con **cargas de trabajo imprevisibles**
- Soporta todos los motores de bases de datos RDS (MariaDB, MySQL, PostgreSQL, SQL Server, Oracle)



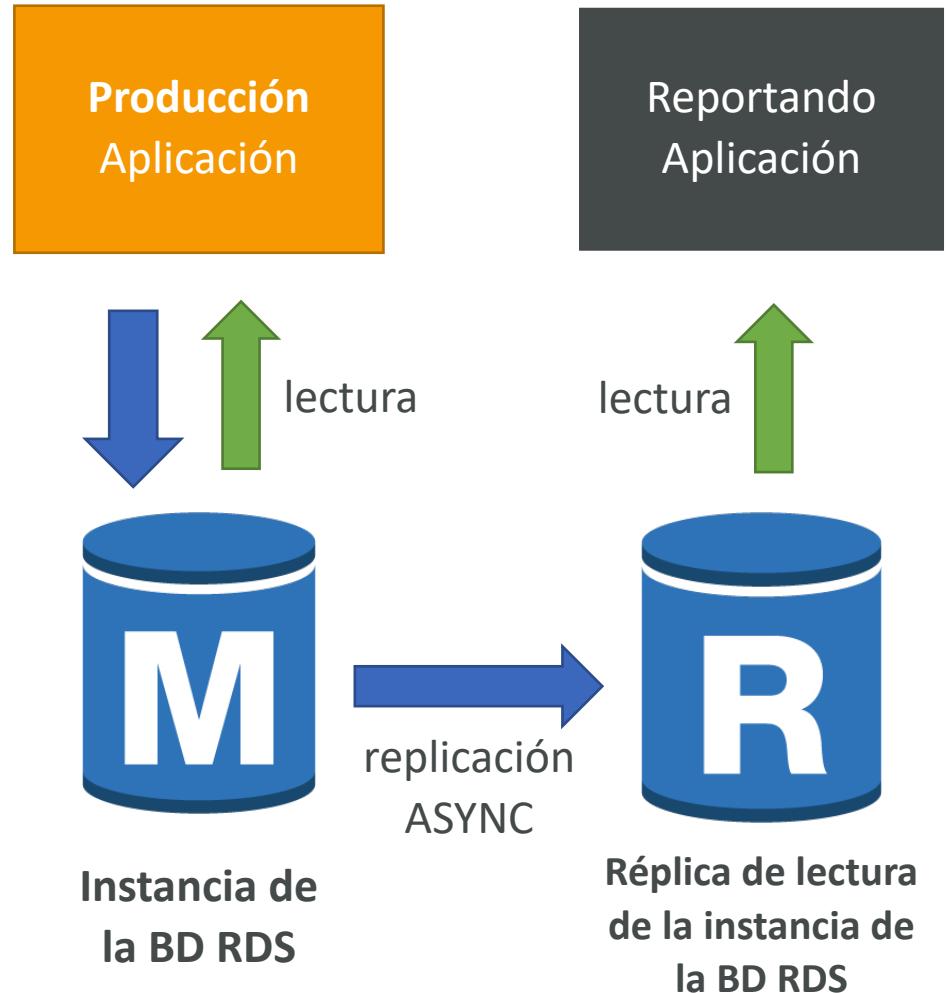
Réplicas de lectura RDS para la escalabilidad de lectura

- Hasta 5 réplicas de lectura
- Dentro de AZ, a través de AZ o a través de la región
- La replicación es **ASYNC**, por lo que las lecturas son finalmente consistentes
- Las réplicas pueden ser promovidas a su propia BD
- Las aplicaciones deben actualizar la cadena de conexión para aprovechar las réplicas de lectura



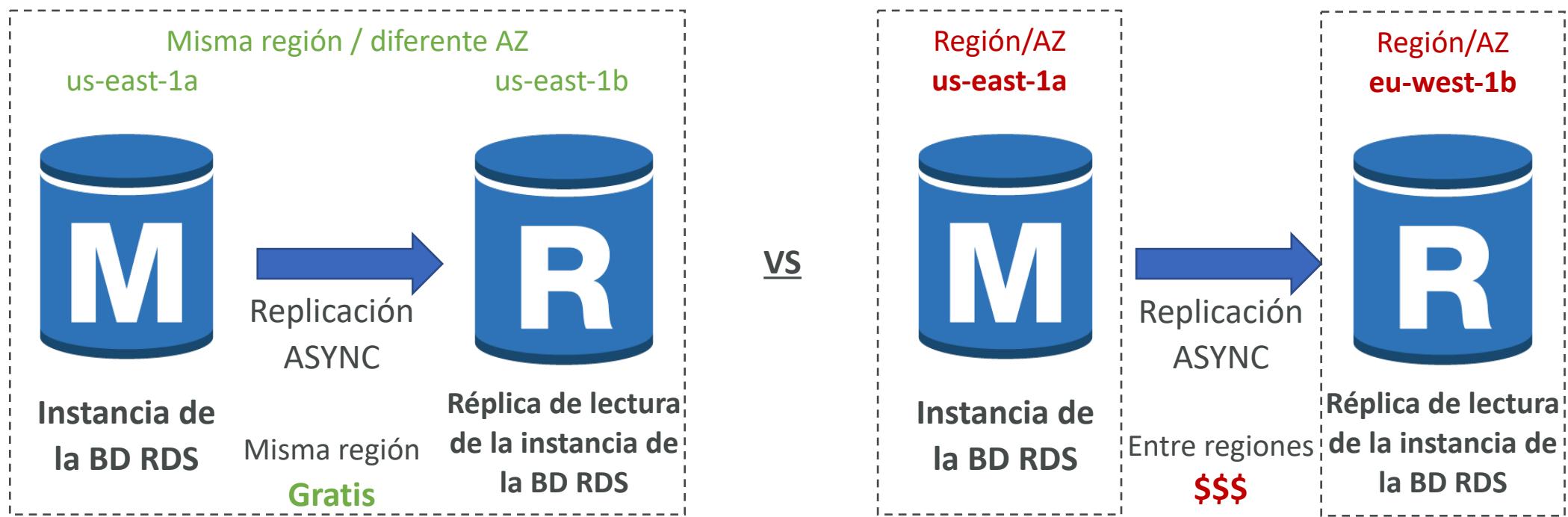
Réplicas de lectura RDS - Casos de uso

- Tienes una base de datos de producción que está recibiendo una carga normal
- Quieres ejecutar una aplicación de informes para realizar algunos análisis
- Creas una réplica de lectura para ejecutar allí la nueva carga de trabajo
- La aplicación de producción no se ve afectada
- Las réplicas de lectura se utilizan sólo para sentencias del tipo SELECT (=lectura) (no INSERT, UPDATE, DELETE)



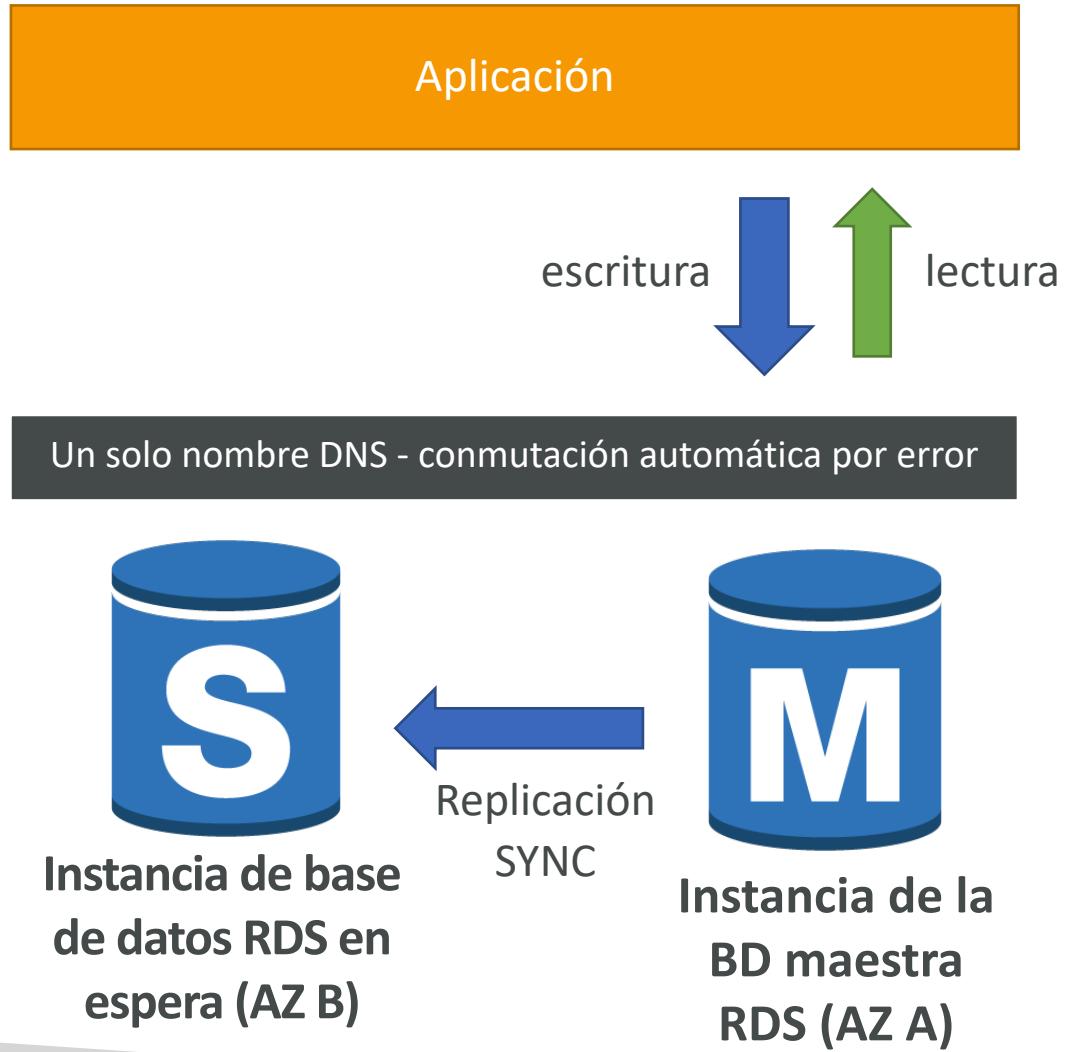
Réplicas de lectura RDS - Coste de la red

- En AWS hay un coste de red cuando los datos van de una AZ a otra
- Para las Réplicas de Lectura RDS dentro de la misma región, no pagas esa tarifa



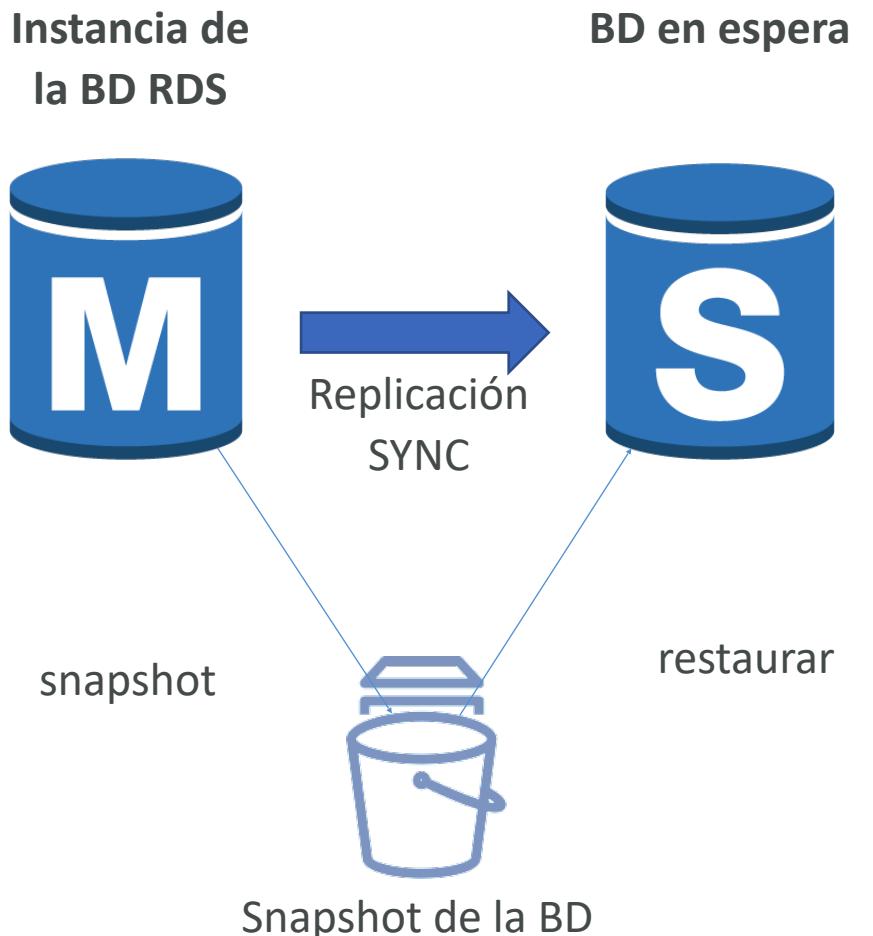
RDS Multi AZ (recuperación de desastres)

- Replicación SYNC
- Un nombre DNS - conmutación automática de la aplicación a la espera
- Aumenta la disponibilidad
- Conmutación por error en caso de pérdida de AZ, pérdida de red, fallo de instancia o de almacenamiento
- Sin intervención manual en las apps
- No se utiliza para escalar
- La replicación multi-AZ es gratis
- Nota: Las réplicas de lectura deben configurarse como Multi AZ para la recuperación de desastres (DR)



RDS – De una AZ a múltiples AZ

- Operación sin tiempo de inactividad (no es necesario parar la BD)
- Sólo tienes que hacer clic en "modificar" la base de datos
- Internamente ocurre lo siguiente
 - Se toma un Snapshot
 - Se restaura una nueva BD a partir del Snapshot en una nueva AZ
 - Se establece la sincronización entre las dos bases de datos



RDS Personalizada

- **Base de datos gestionada de Oracle y Microsoft SQL Server con personalización del sistema operativo y de la base de datos**
- RDS: automatiza la configuración, el funcionamiento y el escalado de la base de datos en AWS
- Personalizada: acceso a la base de datos subyacente y al SO para que puedas
 - Configurar los ajustes
 - Instalar parches
 - Habilitar las funciones nativas
 - Acceder a la instancia EC2 subyacente mediante **SSH** o **SSM Session Manager**
- **Desactivar el Modo de Automatización** para realizar tu personalización, mejor tomar una Snapshot de la BD antes
- RDS vs. RDS Personalizada
 - RDS: toda la base de datos y el SO serán gestionados por AWS
 - RDS Personalizada: acceso administrativo completo al SO subyacente y a la base de datos



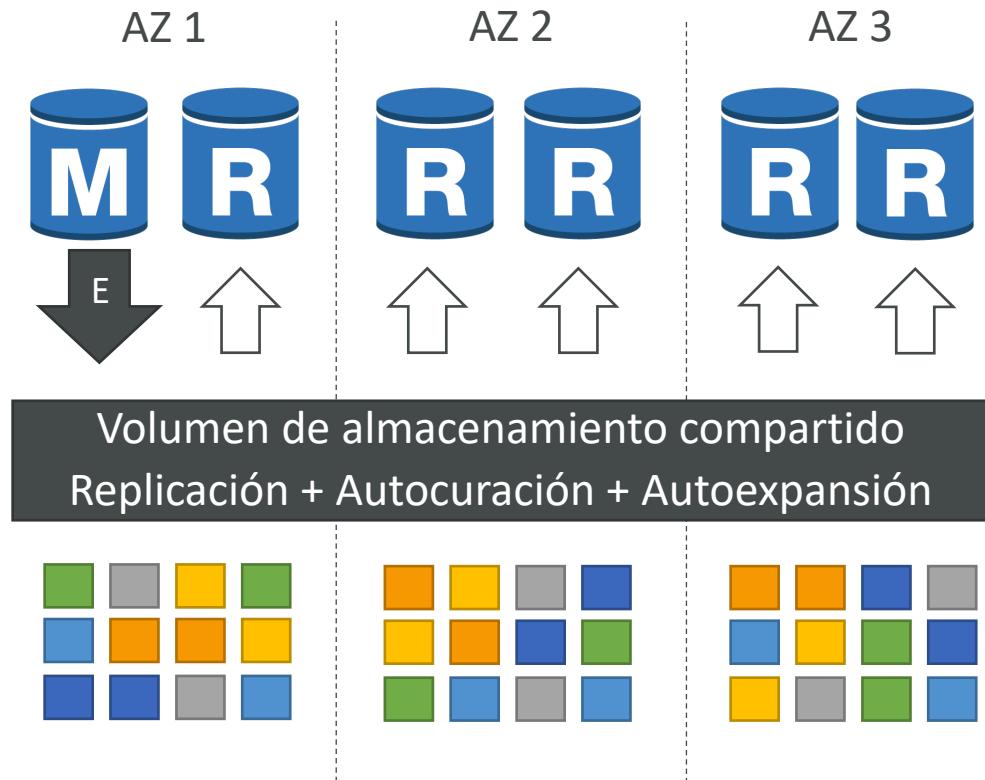


Amazon Aurora

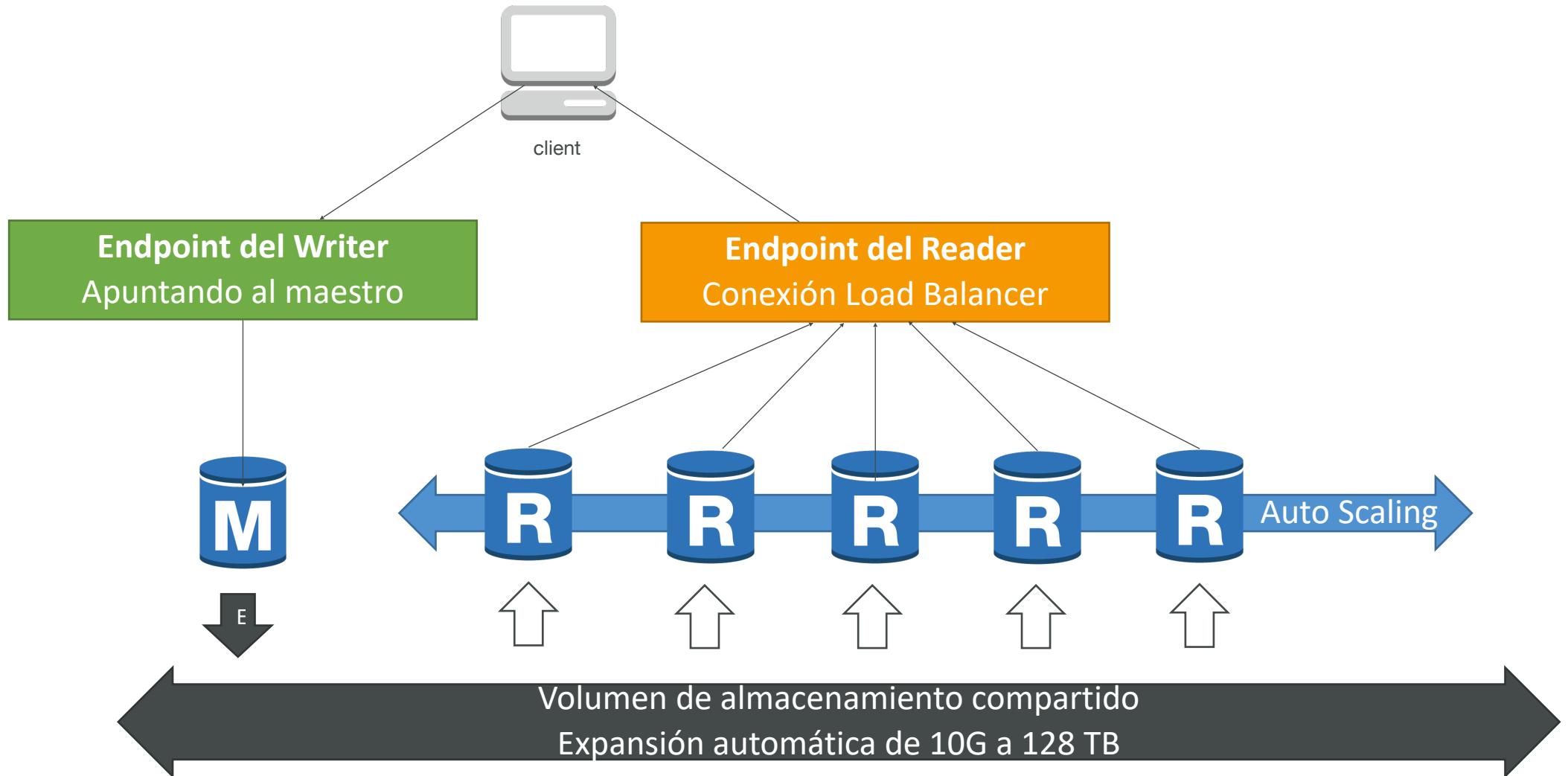
- Aurora es una tecnología propietaria de AWS (no es de código abierto)
- Tanto Postgres como MySQL se soportan como base de datos de Aurora (eso significa que tus controladores funcionarán como si Aurora fuera una base de datos Postgres o MySQL)
- Aurora está "optimizada para la nube de AWS" y afirma que su rendimiento es 5 veces superior al de MySQL en RDS, y más de 3 veces superior al de Postgres en RDS
- El almacenamiento de Aurora crece automáticamente en incrementos de 10 GB, hasta 128 TB.
- Aurora puede tener 15 réplicas mientras que MySQL tiene 5, y el proceso de replicación es más rápido (retraso de réplica inferior a 10 ms)
- La conmutación por error en Aurora es instantánea. Es nativo de Alta Disponibilidad.
- Aurora cuesta más que RDS (20% más), pero es más eficiente

Alta disponibilidad y escalado de lectura de Aurora

- 6 copias de tus datos en 3 AZ:
 - 4 copias de las 6 necesarias para las escrituras
 - 3 copias de las 6 necesarias para las lecturas
 - Autoreparación con replicación entre pares
 - El almacenamiento está dividido en 100 volúmenes
- Una instancia de Aurora se encarga de las escrituras (maestra)
- Recuperación automática del maestro en menos de 30 segundos
- El maestro + hasta 15 réplicas de lectura de Aurora realizan lecturas
- Soporte para la replicación entre regiones



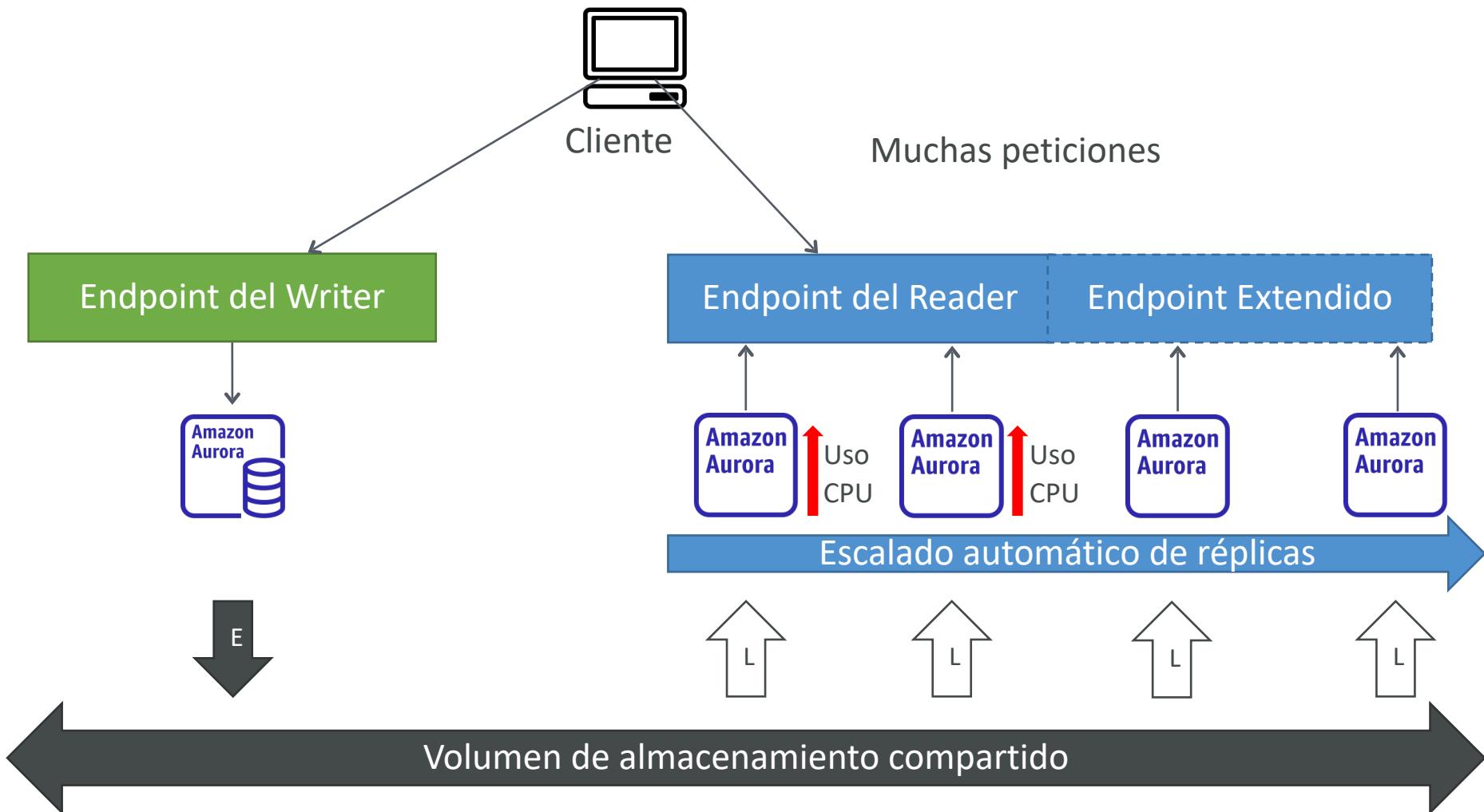
Cluster de BD Aurora



Características de Aurora

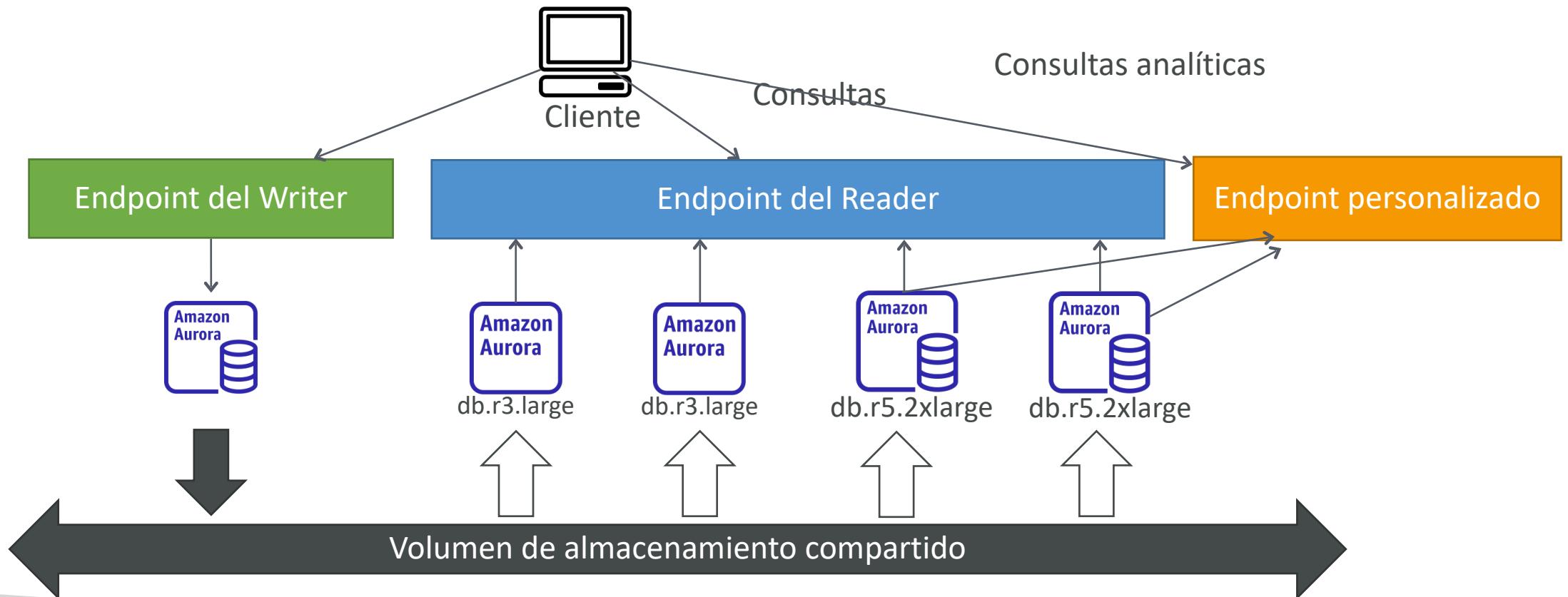
- Comutación automática por error
- Copia de seguridad y recuperación
- Aislamiento y seguridad
- Cumplimiento de la normativa del sector
- Escalado con un botón
- Parches automáticos con cero tiempo de inactividad
- Supervisión avanzada
- Mantenimiento rutinario
- Backtrack: restaura los datos en cualquier momento sin usar copias de seguridad

Réplicas de Aurora - Escalado automático



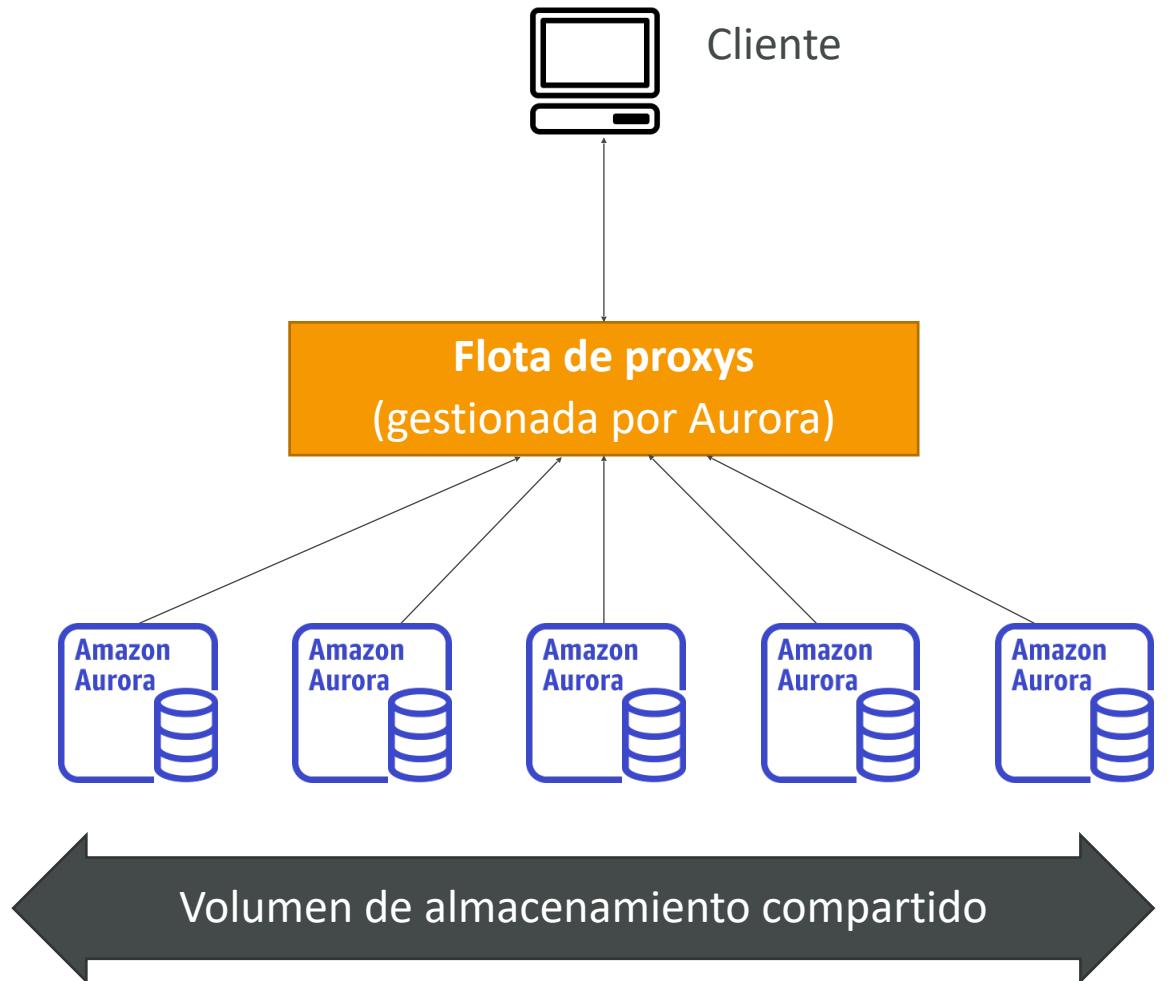
Aurora - Endpoints personalizados

- Definir un subconjunto de instancias de Aurora como endpoint personalizado
- Ejemplo: Ejecutar consultas analíticas en réplicas específicas
- El Endpoint Reader (Lector) generalmente no se utiliza después de definir Endpoints Personalizados



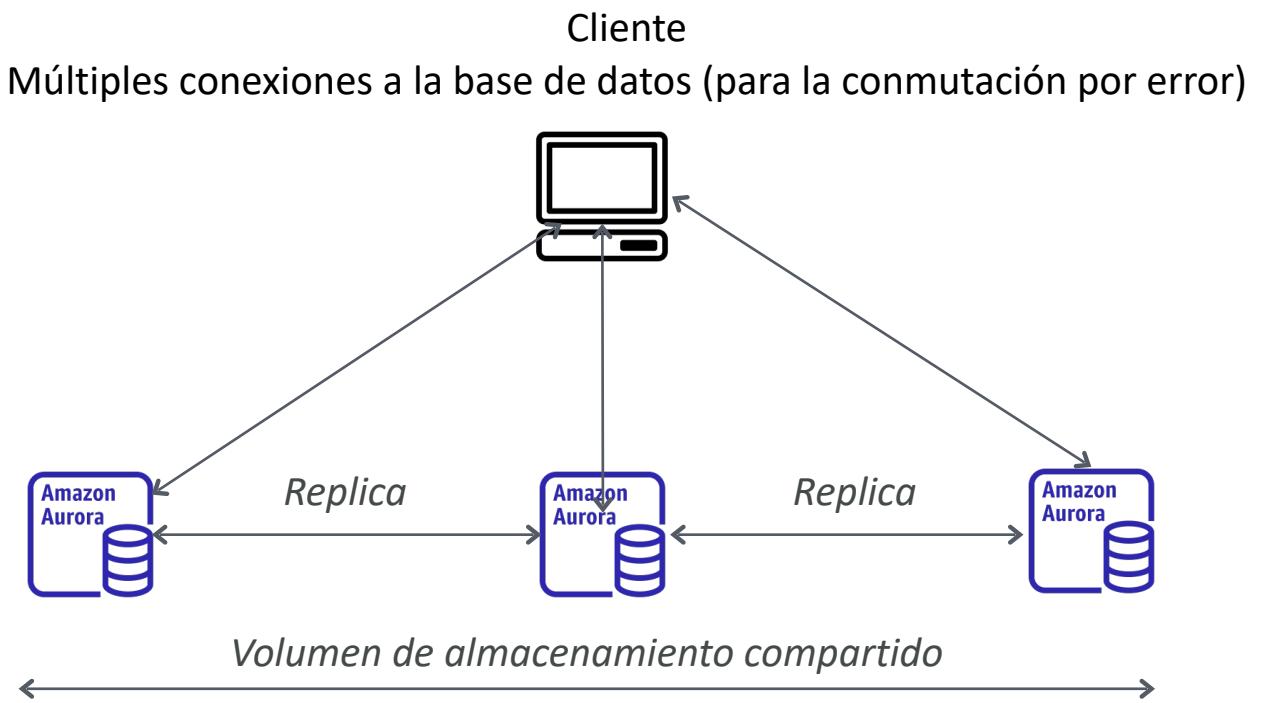
Aurora Serverless (sin servidor)

- Instanciación automática de la base de datos y autoescalado en función del uso real
- Bueno para cargas de trabajo poco frecuentes, intermitentes o imprevisibles
- No es necesario planificar la capacidad
- Pagas por segundo, puede ser más rentable



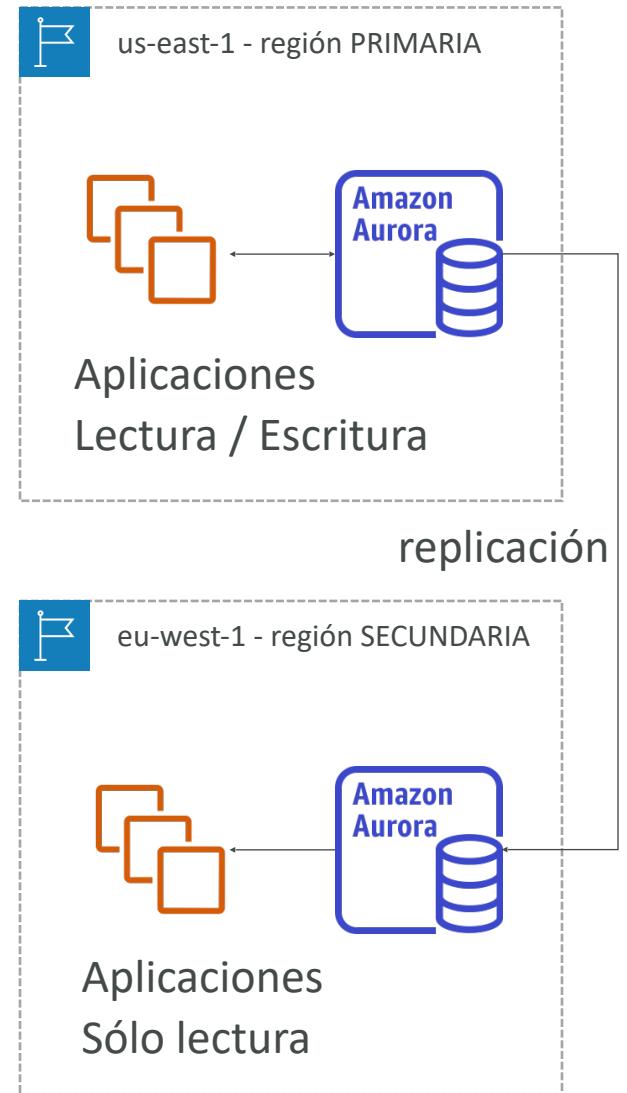
Aurora Multi-Master

- En caso de que quieras una comutación por error inmediata para el nodo de escritura (HA) -
- Cada nodo hace R/W - frente a la promoción de un RR como nuevo maestro



Aurora Global

- Réplicas de lectura entre regiones de Aurora:
 - Útiles para la recuperación de desastres
 - Fácil de poner en marcha
- Base de datos global Aurora (recomendada):
 - 1 región primaria (lectura/escritura)
 - Hasta 5 regiones secundarias (sólo de lectura), el retraso de la replicación es inferior a 1 segundo
 - Hasta 16 réplicas de lectura por región secundaria
 - Ayuda a disminuir la latencia
 - La promoción de otra región (para la recuperación de desastres) tiene un RTO de < 1 minuto
 - La replicación típica entre regiones tarda menos de 1 segundo

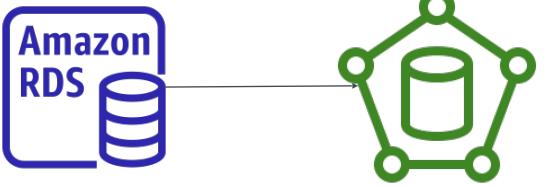


Aurora Machine Learning

- Te permite añadir predicciones basadas en ML a tus aplicaciones a través de SQL
- Integración sencilla, optimizada y segura entre Aurora y los servicios de ML de AWS
- Servicios soportados
 - Amazon SageMaker (se utiliza con cualquier modelo ML)
 - Amazon Comprehend (para el análisis de sentimientos)
- No necesitas tener experiencia en ML
- Casos de uso: detección de fraudes, orientación de anuncios, análisis de sentimientos, recomendaciones de productos

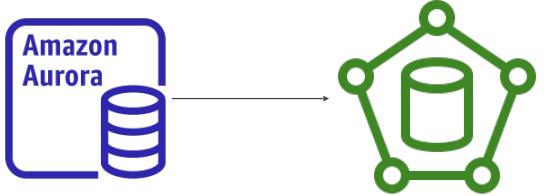


Copias de seguridad RDS



- Copias de seguridad automatizadas:
 - Copia de seguridad completa diaria de la base de datos (durante la ventana de mantenimiento)
 - Los logs de transacciones son respaldados por el RDS cada 5 minutos
 - => posibilidad de restaurar a cualquier punto en el tiempo (desde la copia de seguridad más antigua hasta hace 5 minutos)
 - De 1 a 35 días de retención, establece 0 para desactivar las copias de seguridad automáticas
- Snapshots manuales de la BD
 - Activadas manualmente por el usuario
 - Retención de la copia de seguridad durante el tiempo que quieras
- Truco: en una base de datos RDS parada, seguirás pagando por el almacenamiento. Si planeas detenerla durante mucho tiempo, deberías hacer un Snapshot y restaurar en su lugar

Copias de seguridad de Aurora



- Copias de seguridad automatizadas
 - De 1 a 35 días (no se pueden desactivar)
 - Recuperación puntual en ese intervalo de tiempo
- Snapshots manuales de la BD
 - Activadas manualmente por el usuario
 - Retención de la copia de seguridad durante el tiempo que quieras

Opciones de restauración de RDS y Aurora

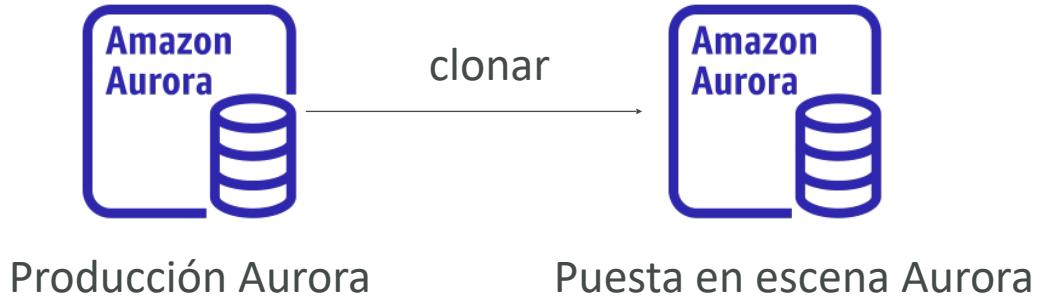


- La restauración de una copia de seguridad de RDS / Aurora o un Snapshot crea una nueva base de datos
- Restaurar la base de datos RDS de MySQL desde S3
 - Crea una copia de seguridad de tu base de datos local
 - Almacénala en Amazon S3 (almacenamiento de objetos)
 - Restaura el archivo de copia de seguridad en una nueva instancia RDS que ejecute MySQL
- Restaurar el Cluster Aurora de MySQL desde S3
 - Crea una copia de seguridad de tu base de datos local utilizando Percona XtraBackup
 - Almacena el archivo de copia de seguridad en Amazon S3
 - Restaura el archivo de copia de seguridad en un nuevo Cluster Aurora que ejecute MySQL



Clonación de la base de datos Aurora

- Crea un nuevo Cluster de BD de Aurora a partir de uno existente
- Más rápido que el Snapshot y la restauración
- El nuevo Cluster de BD utiliza el mismo volumen de cluster y los mismos datos que el original, pero cambiará cuando se actualicen los datos
- Muy rápido y rentable
- Útil para crear una base de datos "de ensayo" a partir de una base de datos "de producción" sin afectar a la base de datos de producción

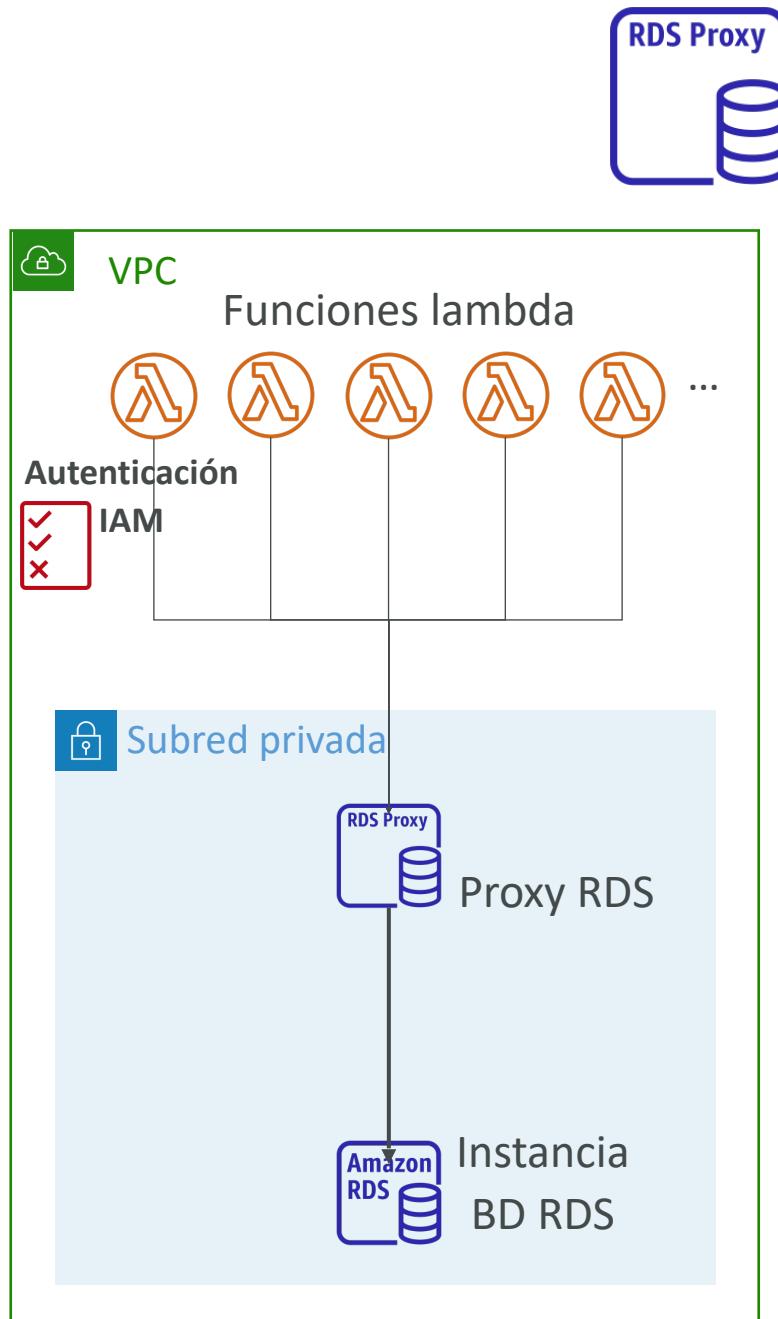


Seguridad RDS y Aurora

- **Cifrado en reposo:**
 - Cifrado de la base de datos maestra y de las réplicas mediante AWS KMS - debe definirse en el momento del lanzamiento.
 - Si la base maestra no está cifrada, las réplicas de lectura no pueden ser cifradas
 - Para cifrar una base de datos no cifrada, pasa por un Snapshot de la base de datos y restaura como cifrada
- **Cifrado en vuelo:** Preparado para TLS por defecto, utiliza certificados root del lado del cliente de AWS TLS
- **Autenticación IAM:** Roles de IAM para conectarse a tu base de datos (en lugar de nombre de usuario/pw)
- **Grupos de seguridad:** Controla el acceso de red a tu RDS / Aurora DB
- **No hay SSH disponible** excepto en RDS Custom
- **Los logs de auditoría pueden ser activados** y enviados a CloudWatch Logs para una mayor retención

Proxy de Amazon RDS

- Proxy de base de datos totalmente gestionado para RDS
- Permite a las apps agrupar y compartir las conexiones a la base de datos establecidas
- **Mejora la eficiencia de la base de datos reduciendo el estrés de los recursos de la base de datos (por ejemplo, CPU, RAM) y minimizando las conexiones abiertas (y los tiempos de espera)**
- Sin servidor, con autoescalado y alta disponibilidad (multi-AZ)
- Reduce el tiempo de comutación por error de RDS y Aurora hasta en un 66%
- **Soporta RDS (MySQL, PostgreSQL, MariaDB) y Aurora (MySQL, PostgreSQL)**
- No se requieren cambios de código para la mayoría de las aplicaciones
- **Aplica la autenticación IAM para la base de datos y almacena de forma segura las credenciales en AWS Secrets Manager**
- **El proxy RDS nunca es accesible al público (debe accederse desde la VPC)**



Visión general de Amazon ElastiCache

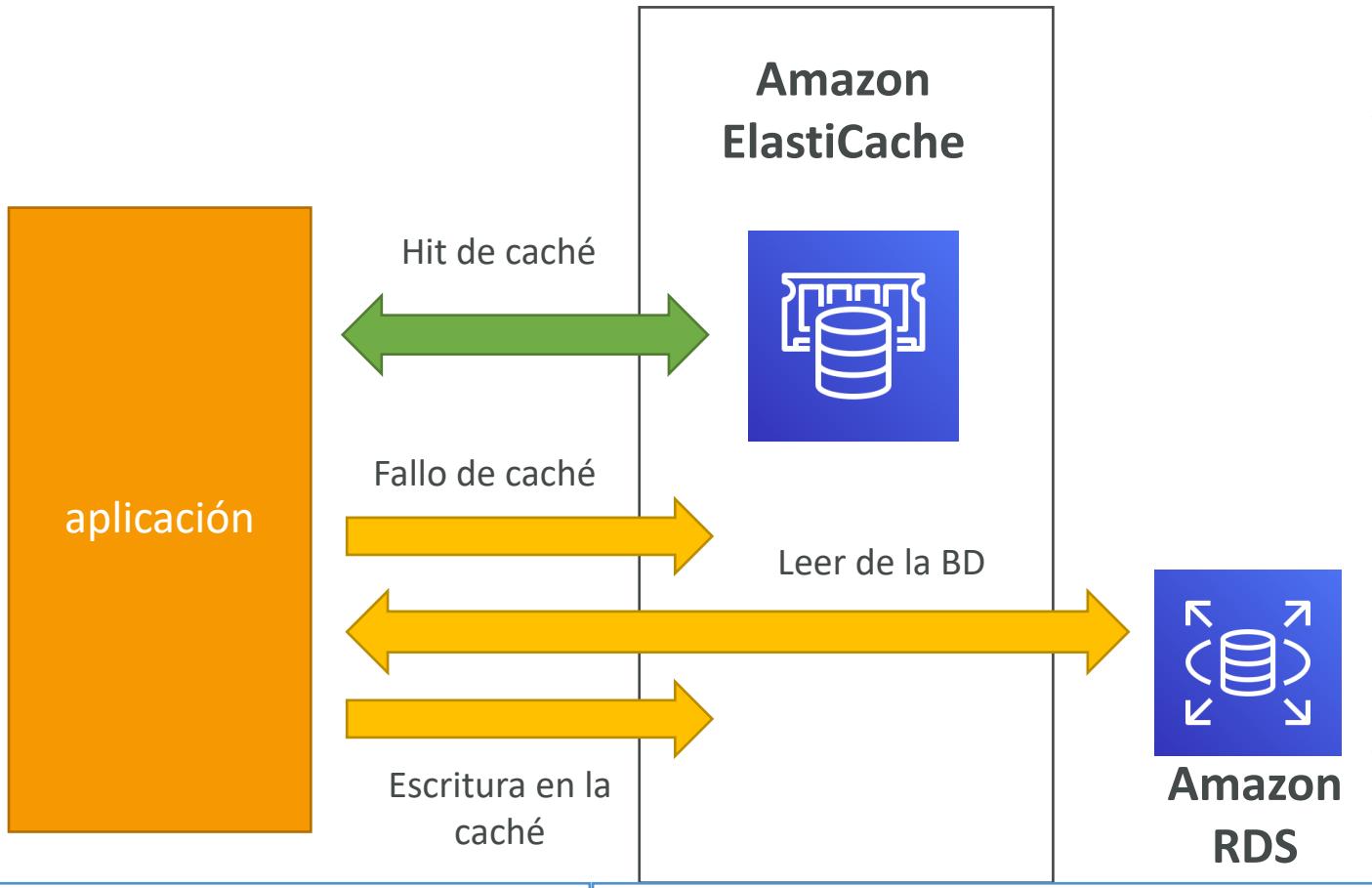


- De la misma manera que RDS es para conseguir bases de datos relacionales gestionadas...
- ElastiCache es para obtener Redis o Memcached gestionados
- Las cachés son bases de datos en memoria con un rendimiento realmente alto y baja latencia
- Ayuda a reducir la carga de las bases de datos para cargas de trabajo de lectura intensiva
- Ayuda a que tu aplicación no tenga estado
- AWS se encarga del mantenimiento/parche del sistema operativo, las optimizaciones, la instalación, la configuración, la supervisión, la recuperación de fallos y las copias de seguridad
- **El uso de ElastiCache implica grandes cambios en el código de la aplicación**

ElastiCache

Solución de arquitectura - DB Cache

- Las aplicaciones consultan ElastiCache, si no está disponible, lo obtienen del RDS y lo almacenan en ElastiCache.
- Ayuda a aliviar la carga en RDS
- La caché debe tener una estrategia de invalidación para asegurarse de que sólo se utilizan allí los datos más actuales.



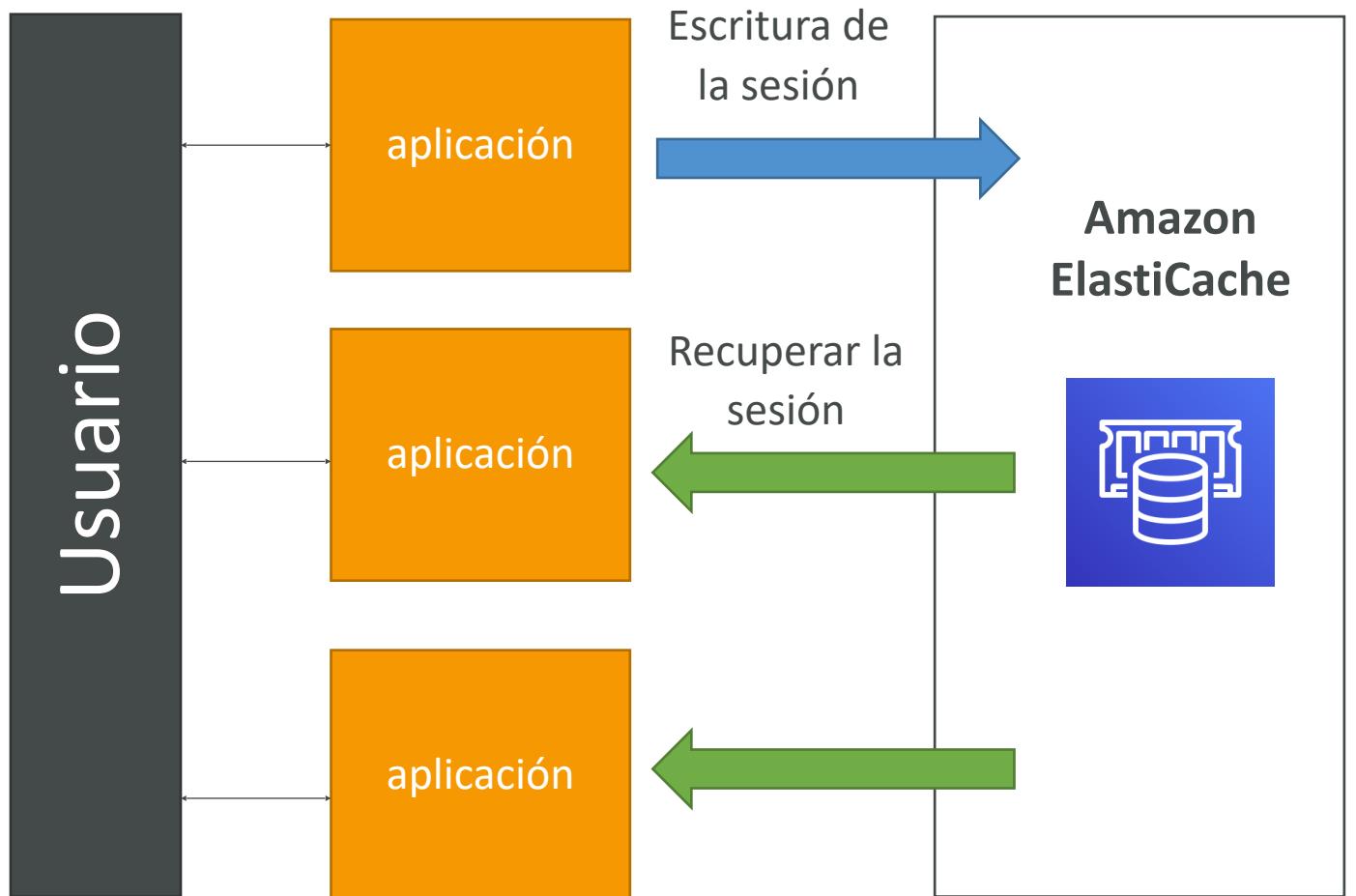
Un **hit de caché** es un estado en el que los datos solicitados para el procesamiento por un componente o aplicación se encuentran en la memoria caché.

Un **fallo de caché** es cuando los datos que solicita un sistema o una aplicación no se encuentran en la memoria caché.

ElastiCache

Solución de arquitectura - Almacén de sesiones de usuario

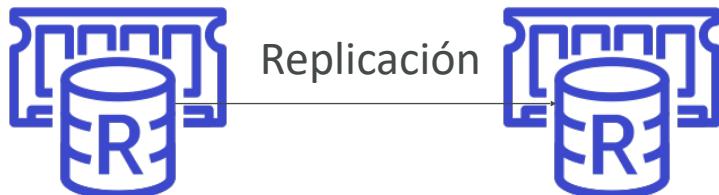
- El usuario se loguea en cualquiera de las aplicaciones
- La aplicación escribe los datos de la sesión en ElastiCache
- El usuario accede a otra instancia de nuestra aplicación
- La instancia recupera los datos y el usuario ya ha iniciado la sesión



ElastiCache – Redis vs Memcached

REDIS

- **Multi AZ** con Auto-Failover
- **Rélicas de lectura** para escalar las lecturas y tener **alta disponibilidad**
- Durabilidad de los datos mediante la persistencia AOF
- **Funciones de copia de seguridad y restauración**



MEMCACHED

- Múltiples nodos para la partición de datos (sharding)
- **Sin alta disponibilidad (replicación)**
- **No es persistente**
- **No hay copia de seguridad ni restauración**
- Arquitectura multihilo



ElastiCache - Seguridad de la caché

- Todas las cachés de ElastiCache:
 - **No soportan la autenticación IAM**
 - Las políticas de IAM en ElastiCache sólo se utilizan para la seguridad a nivel de API de AWS
- **Redis AUTH**
 - Puedes establecer una "contraseña/token" cuando crees un Cluster de Redis
 - Se trata de un nivel adicional de seguridad para tu caché (además de los grupos de seguridad)
 - Soporta el cifrado SSL en vuelo
- Memcached
 - Soporta la autenticación basada en SASL (avanzada)



Patrones para ElastiCache

- **Carga lenta (lazy)**: todos los datos leídos se almacenan en la caché, los datos pueden quedar obsoletos en la caché
- **Escribir a través**: añade o actualiza los datos en la caché cuando se escriben en una BD (no hay datos obsoletos)
- **Almacenamiento de sesión**: almacena los datos temporales de la sesión en una caché (utilizando las características del TTL)
- *Cita: Sólo hay dos cosas difíciles en Informática: la invalidación de la caché y el nombramiento de las cosas*

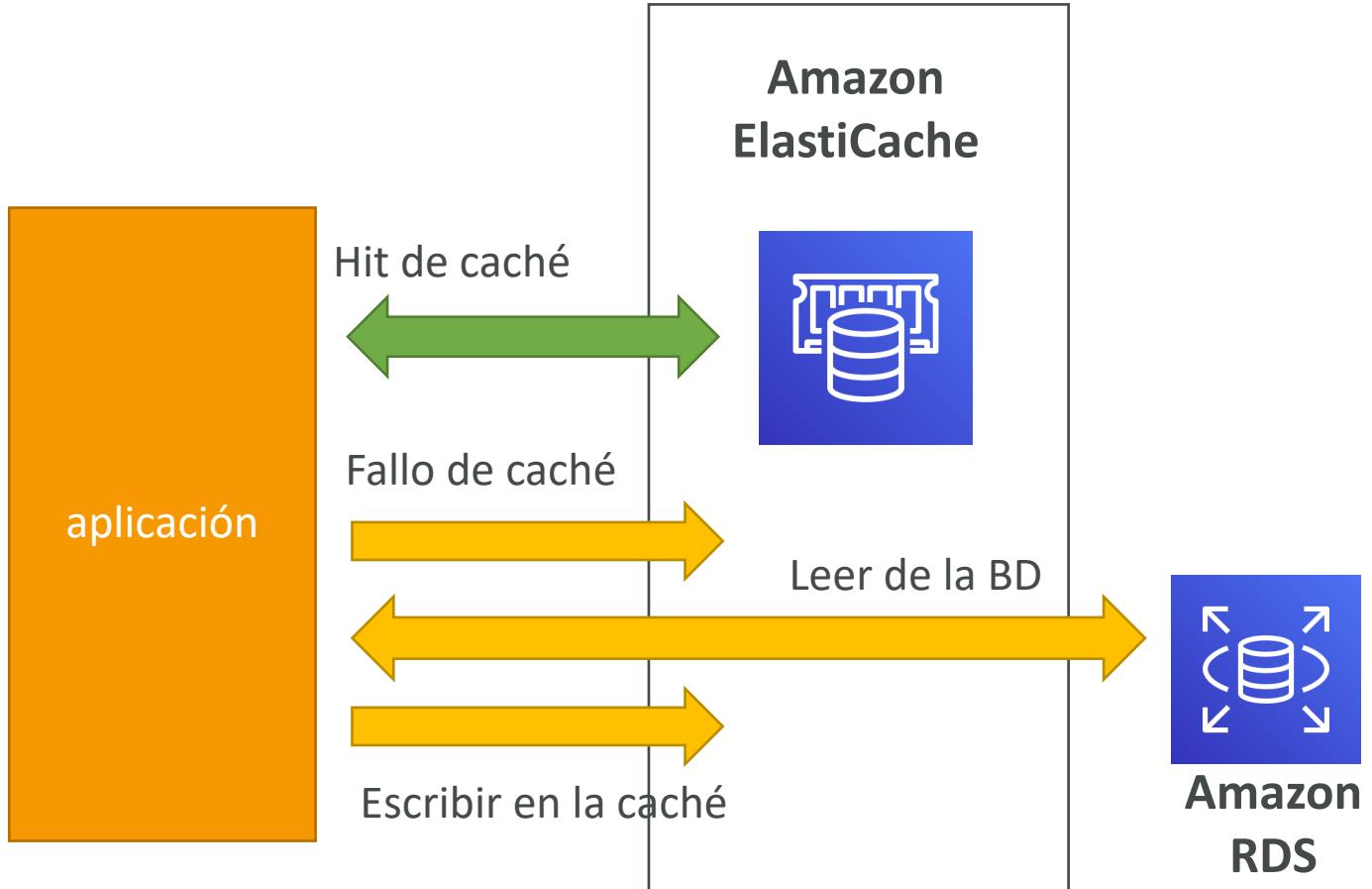
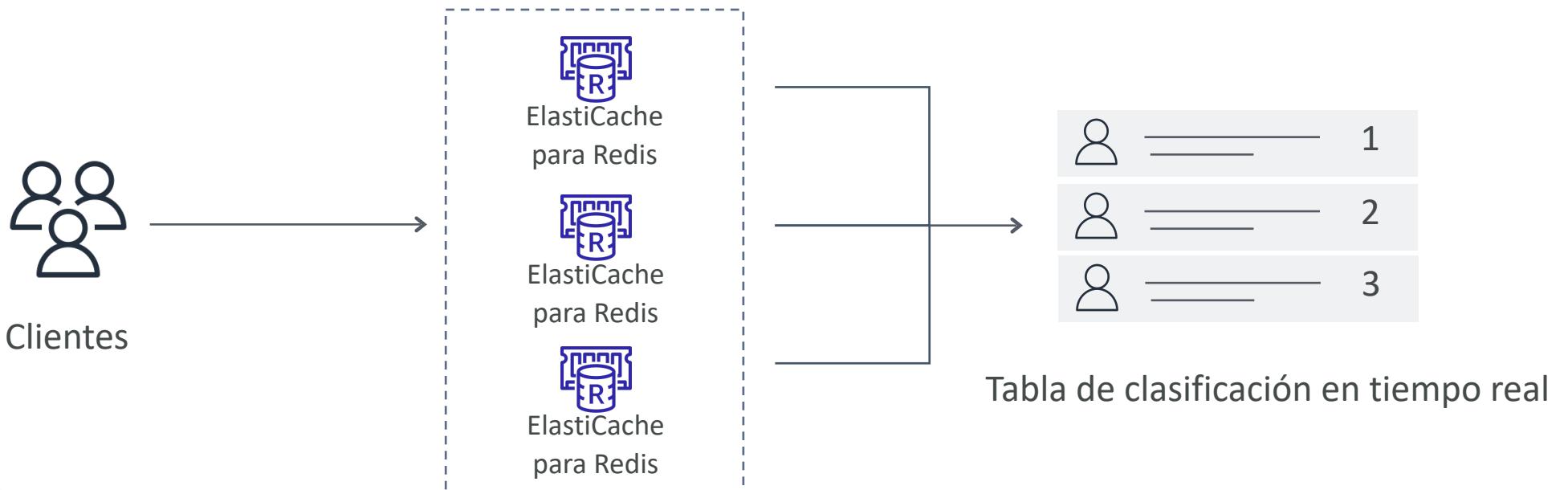


Ilustración de la carga lenta

Caso de uso de ElastiCache - Redis

- Las tablas de clasificación de juegos son computacionalmente complejas
- **Los conjuntos ordenados de Redis** garantizan tanto la unicidad como el orden de los elementos
- Cada vez que se añade un nuevo elemento, se clasifica en tiempo real y se añade en el orden correcto



Route 53

¿Qué es DNS?

- Sistema de nombres de dominio que traduce los nombres de host amigables con el ser humano en las direcciones IP de las máquinas
- www.google.com => 172.217.18.36
- El DNS es la columna vertebral de Internet
- El DNS utiliza una estructura jerárquica de nombres

.com

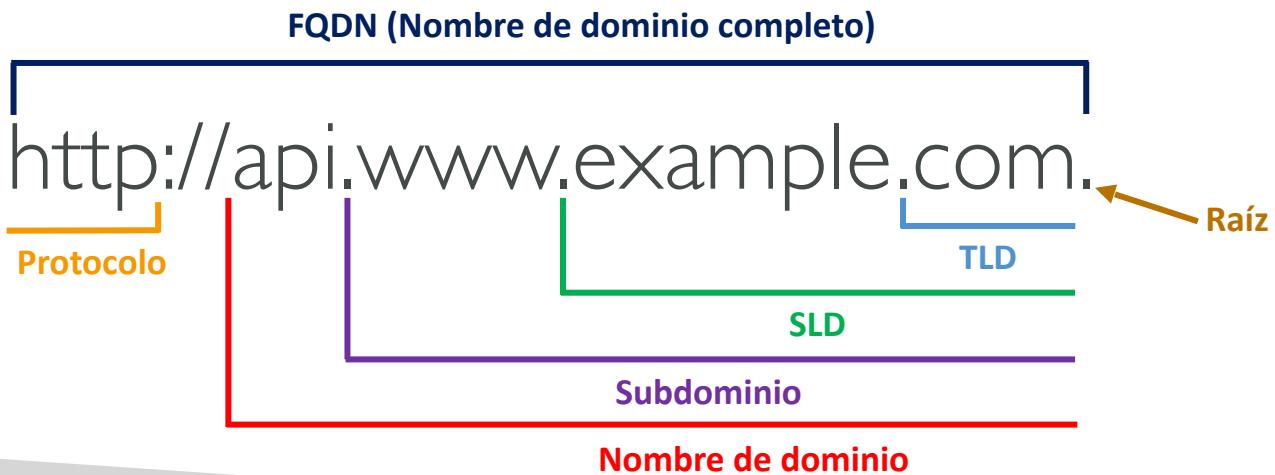
example.com

www.example.com

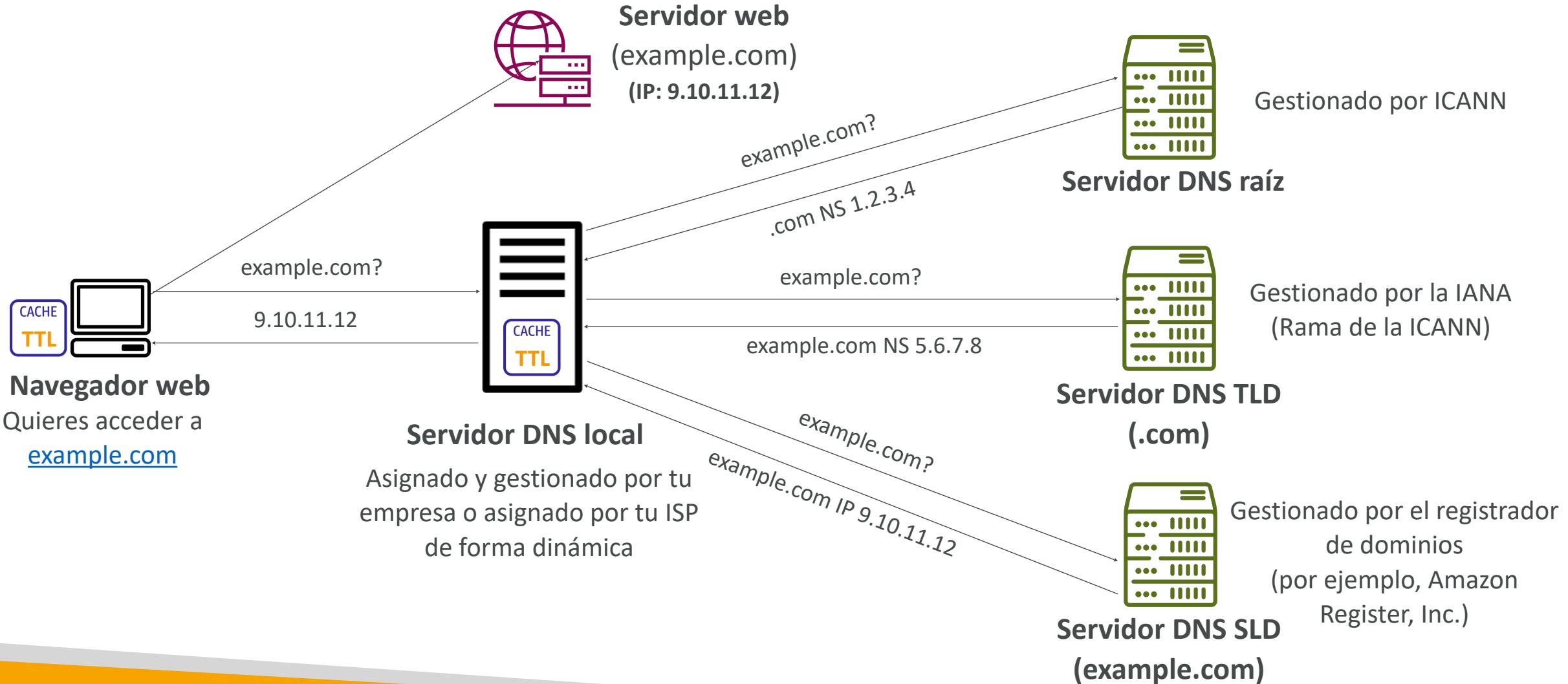
api.example.com

Terminología de DNS

- **Registrador de dominios:** Amazon Route 53, GoDaddy, ...
- **Registros DNS:** A, AAAA, CNAME, NS, ...
- **Archivo de zona:** contiene registros DNS
- **Servidor de nombres:** resuelve las consultas DNS (autorizadas o no autorizadas)
- **Dominio de primer nivel (TLD):** .com, .us, .in, .gov, .org, ...
- **Dominio de segundo nivel (SLD):** amazon.com, google.com, ...

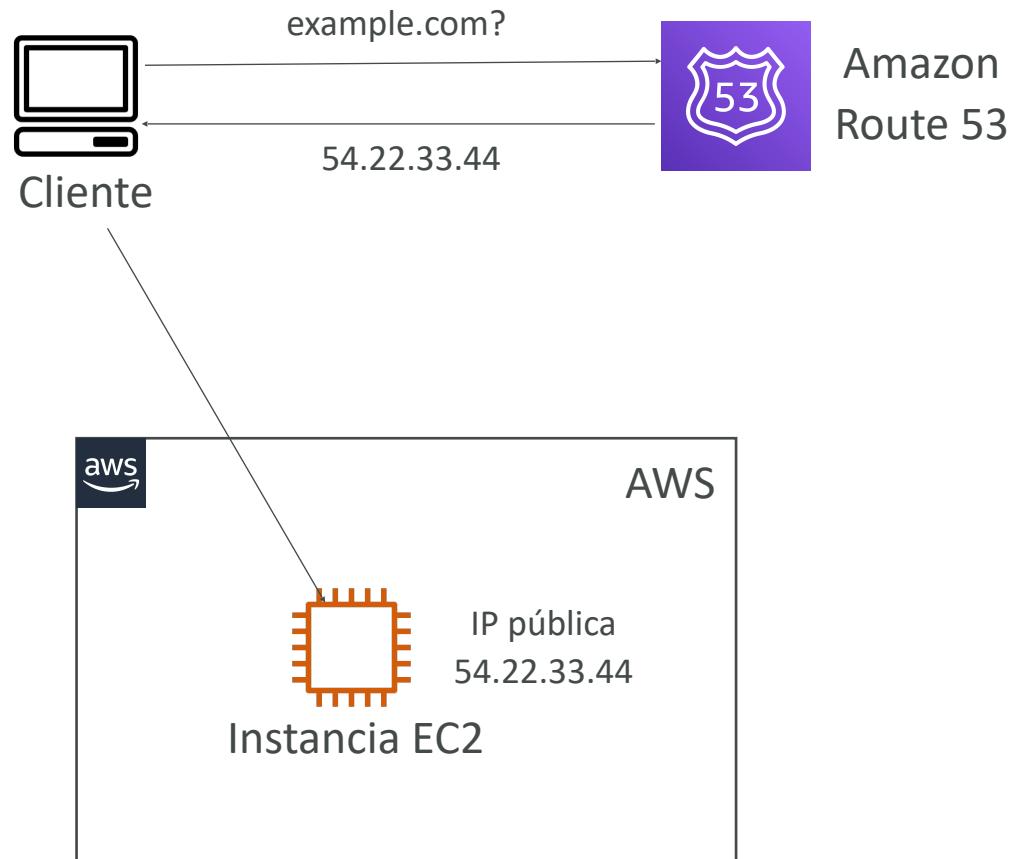


Cómo funciona el DNS



Amazon Route 53

- Un DNS altamente disponible, escalable, totalmente gestionado y autoritativo
 - Autoritario = el cliente (tú) puede actualizar los registros DNS
- Route 53 también es un registrador de dominios
- Posibilidad de comprobar la salud de tus recursos
- El único servicio de AWS que ofrece un SLA de disponibilidad del 100%
- ¿Por qué Route 53? 53 es una referencia al puerto DNS tradicional



Route 53 - Registros

- Cómo quieras dirigir el tráfico de un dominio
- Cada registro contiene
 - **Nombre del dominio/subdominio** - por ejemplo, ejemplo.com
 - **Tipo de registro** - por ejemplo, A o AAAA
 - **Valor** - por ejemplo, 12.34.56.78
 - **Política de enrutamiento** - cómo responde Route 53 a las consultas
 - **TTL** - cantidad de tiempo que el registro se almacena en caché en los Resolvers DNS
- Route 53 soporta los siguientes tipos de registros DNS:
 - (obligatorio) A / AAAA / CNAME / NS
 - (avanzado) CAA / DS / MX / NAPTR / PTR / SOA / TXT / SPF / SRV

Route 53 - Tipos de registro

- **A** - asigna un nombre de host a IPv4
- **AAAA** - asigna un nombre de host a IPv6
- **CNAME** - asigna un nombre de host a otro nombre de host
 - El objetivo es un nombre de dominio que debe tener un registro A o AAAA
 - No puedes crear un registro CNAME para el nodo superior de un espacio de nombres DNS (Zona Apex)
 - Ejemplo: no puedes crear para example.com, pero sí para www.example.com
- **NS** - Servidores de nombres para la Zona Alojada
 - Controla cómo se enruta el tráfico de un dominio

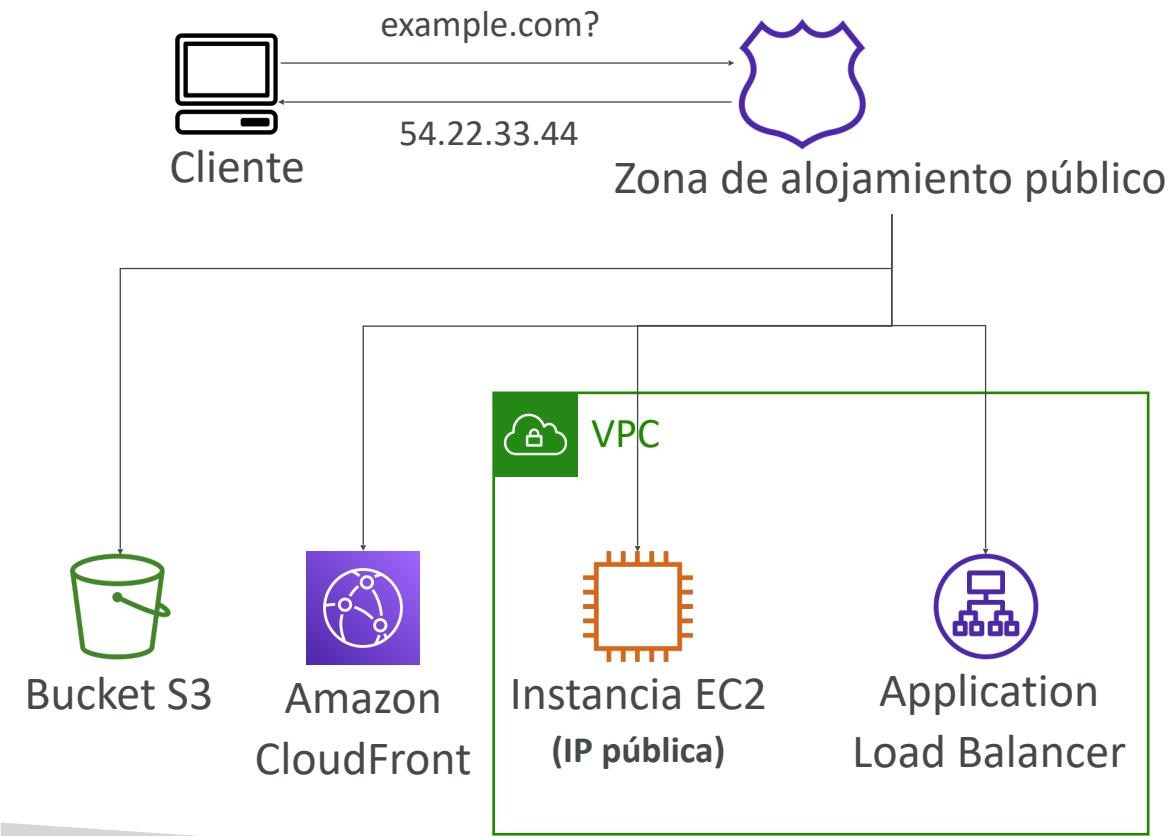


Route 53 - Zonas de alojamiento

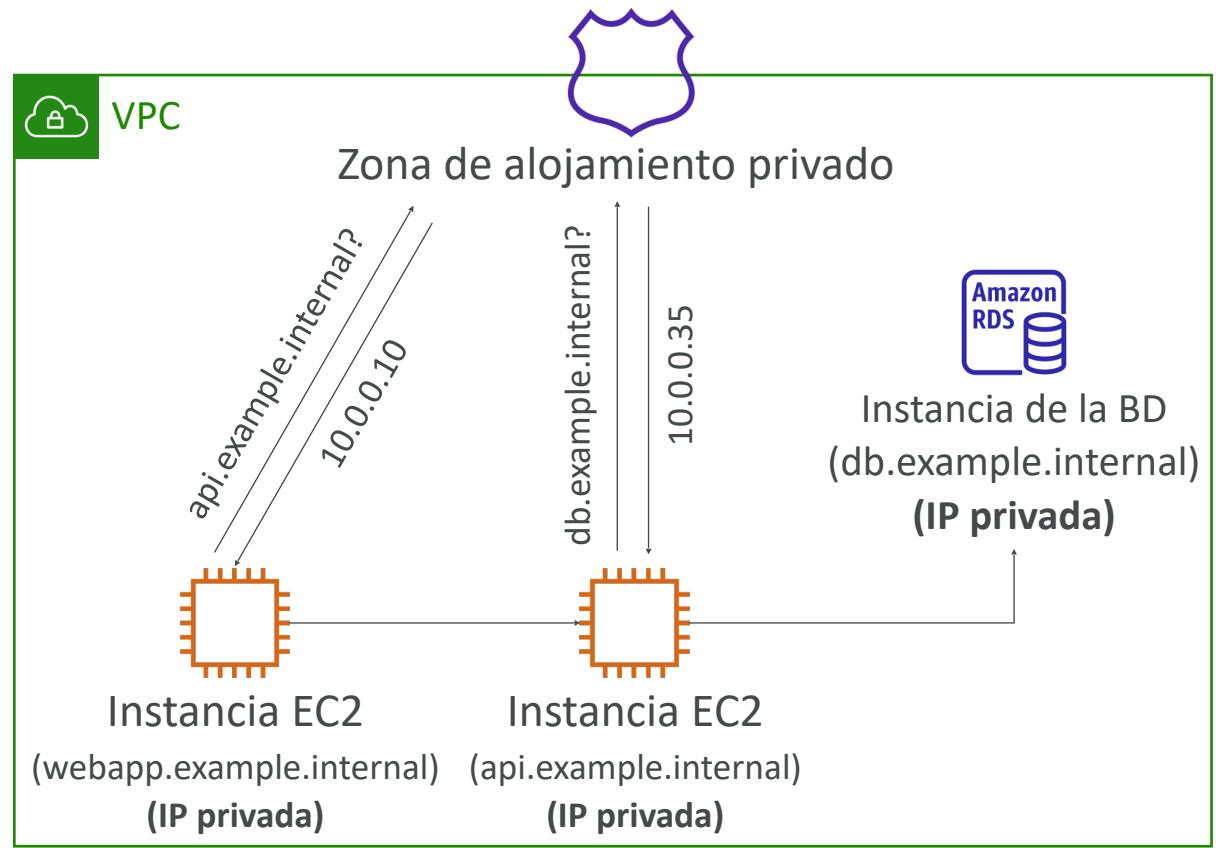
- Un contenedor para los registros que definen cómo dirigir el tráfico a un dominio y sus subdominios
- **Zonas de alojamiento público:** contiene registros que especifican cómo enrutar el tráfico en Internet (nombres de dominio público)
application1.mypublicdomain.com
- **Zonas de alojamiento privadas:** contienen registros que especifican cómo enrutar el tráfico dentro de una o más VPC (nombres de dominio privados)
application1.company.internal
- Pagas 0,50\$ al mes por zona alojada

Route 53 - Zonas de alojamiento públicas frente a privadas

Zona de alojamiento público

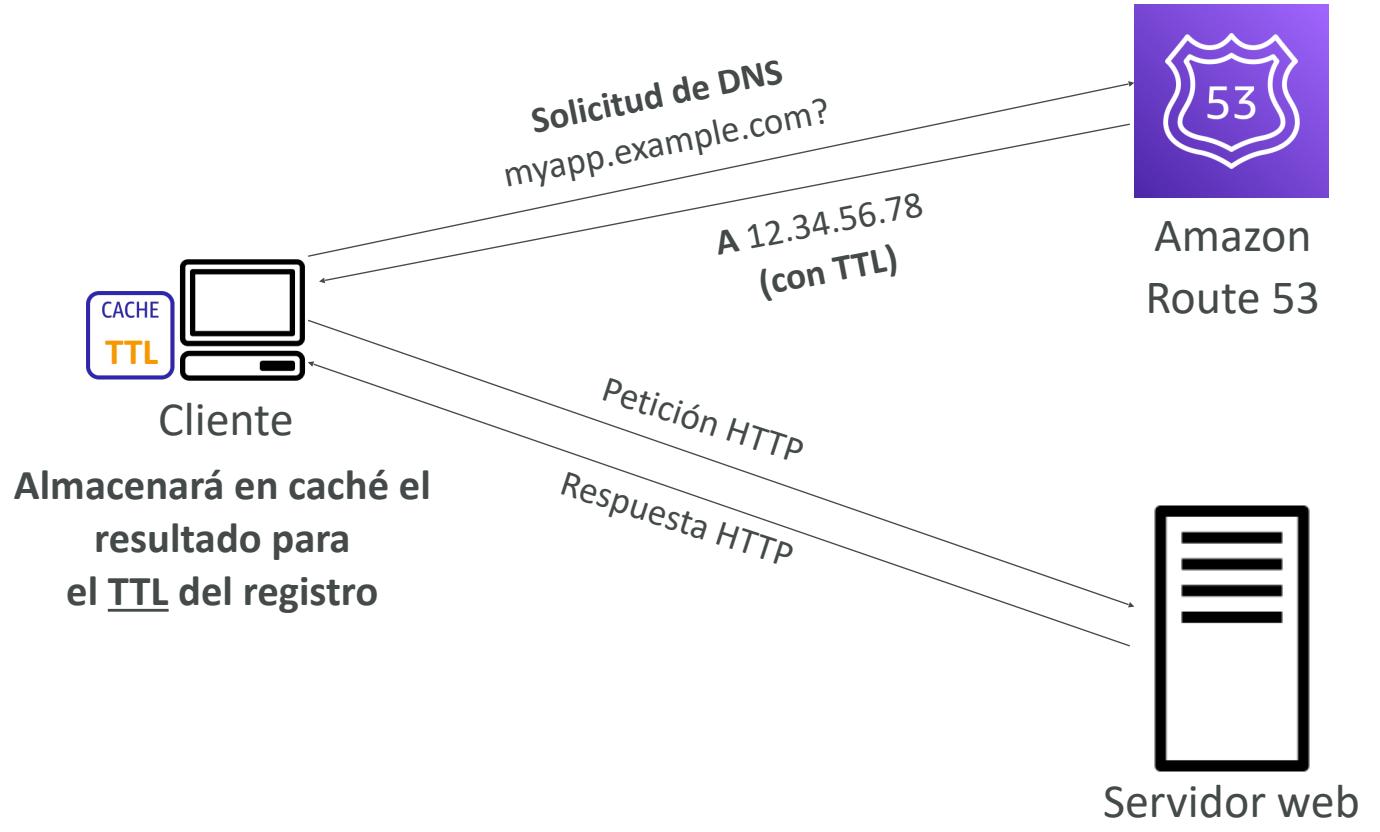


Zona de alojamiento privado



Route 53 - Registros TTL (Tiempo de vida)

- **TTL alto - por ejemplo, 24 horas**
 - Menos tráfico en Route 53
 - Registros posiblemente obsoletos
- **TTL bajo - por ejemplo, 60 seg.**
 - Más tráfico en la Route 53 (\$\$)
 - Los registros están desfasados durante menos tiempo
 - Facilidad para cambiar los registros
- **Excepto los registros de Alias, el TTL es obligatorio para cada registro DNS**



CNAME vs Alias

- Los recursos de AWS (Load Balancer, CloudFront...) exponen un nombre de host de AWS:
 - **IbI-I234.us-east-2.elb.amazonaws.com** y quieres **myapp.midominio.com**
- CNAME:
 - Apunta un nombre de host a cualquier otro nombre de host. (app.midominio.com => blabla.algo.com)
 - **SÓLO PARA DOMINIOS NO ROOT (algo.midominio.com)**
- Alias:
 - Apunta un nombre de host a un recurso de AWS (app.mydomain.com => blabla.amazonaws.com)
 - **Funciona para DOMINIO RAÍZ y DOMINIO NO RAÍZ (mydomain.com)**
 - Gratis
 - Comprobación de salud nativa

Route 53 - Registros con Alias

- Asigna un nombre de host a un recurso de AWS
- Una extensión de la funcionalidad del DNS
- Reconoce automáticamente los cambios en las direcciones IP del recurso
- A diferencia de CNAME, puede utilizarse para el nodo superior de un espacio de nombres DNS (Zona Apex), por ejemplo: example.com
- El Registro Alias es siempre del tipo A/AAAA para los recursos AWS (IPv4 / IPv6)
- No puedes establecer el TTL



Route 53 – Objetivos Registros con Alias

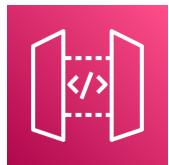
- Elastic Load Balancers
- Distribuciones CloudFront
- API Gateway
- Entornos Elastic Beanstalk
- Sitios web S3
- Endpoints de interfaz VPC
- Global Accelerator
- Registro Route 53 en la misma Zona alojada
- **No puedes establecer un registro ALIAS para un nombre DNS de EC2**



Elastic
Load Balancer



Amazon
CloudFront



Amazon
API Gateway



Elastic Beanstalk



Sitios web S3



Interfaz VPC
Endpoints



Global Accelerator



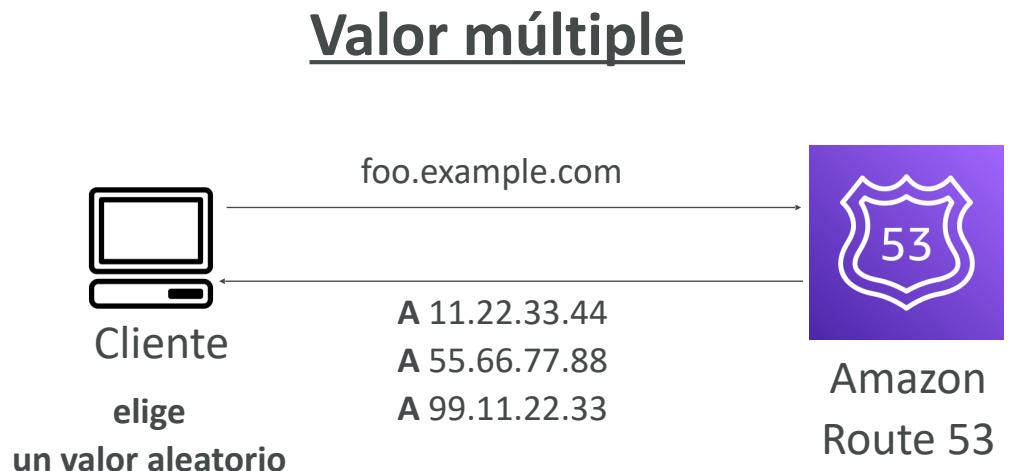
Registro de la Route 53
(misma Zona de alojamiento)

Route 53 - Políticas de enrutamiento

- Definir cómo responde Route 53 a las consultas DNS
- No te confundas con la palabra "*Enrutamiento*"
 - No es lo mismo que el enrutamiento del Load Balancer, que enruta el tráfico
 - El DNS no enruta ningún tráfico, sólo responde a las consultas del DNS
- Route 53 soporta las siguientes políticas de enrutamiento
 - Simple
 - Ponderada
 - Comutación por error
 - Basada en la latencia
 - Geolocalización
 - Respuesta multivalente
 - Geoproximidad (utilizando la función de flujo de tráfico de Route 53)

Políticas de enrutamiento - Simple

- Normalmente, dirige el tráfico a un solo recurso
- Puede especificar varios valores en el mismo registro
- **Si se devuelven varios valores, el cliente elige uno al azar**
- Cuando se habilita el Alias, sólo se puede especificar un recurso de AWS
- No se puede asociar a los controles de salud

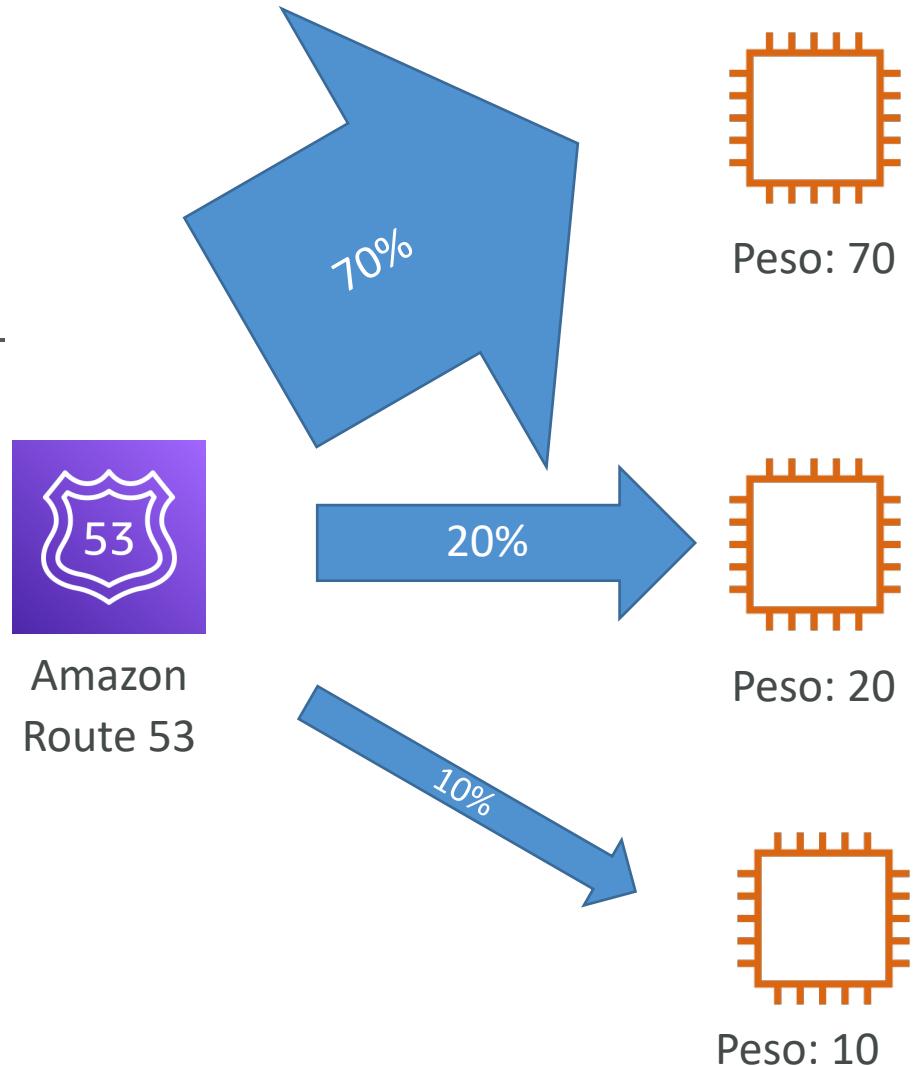


Políticas de enrutamiento - Ponderadas

- Controla el % de las solicitudes que van a cada recurso específico
- Asigna a cada registro un peso relativo:

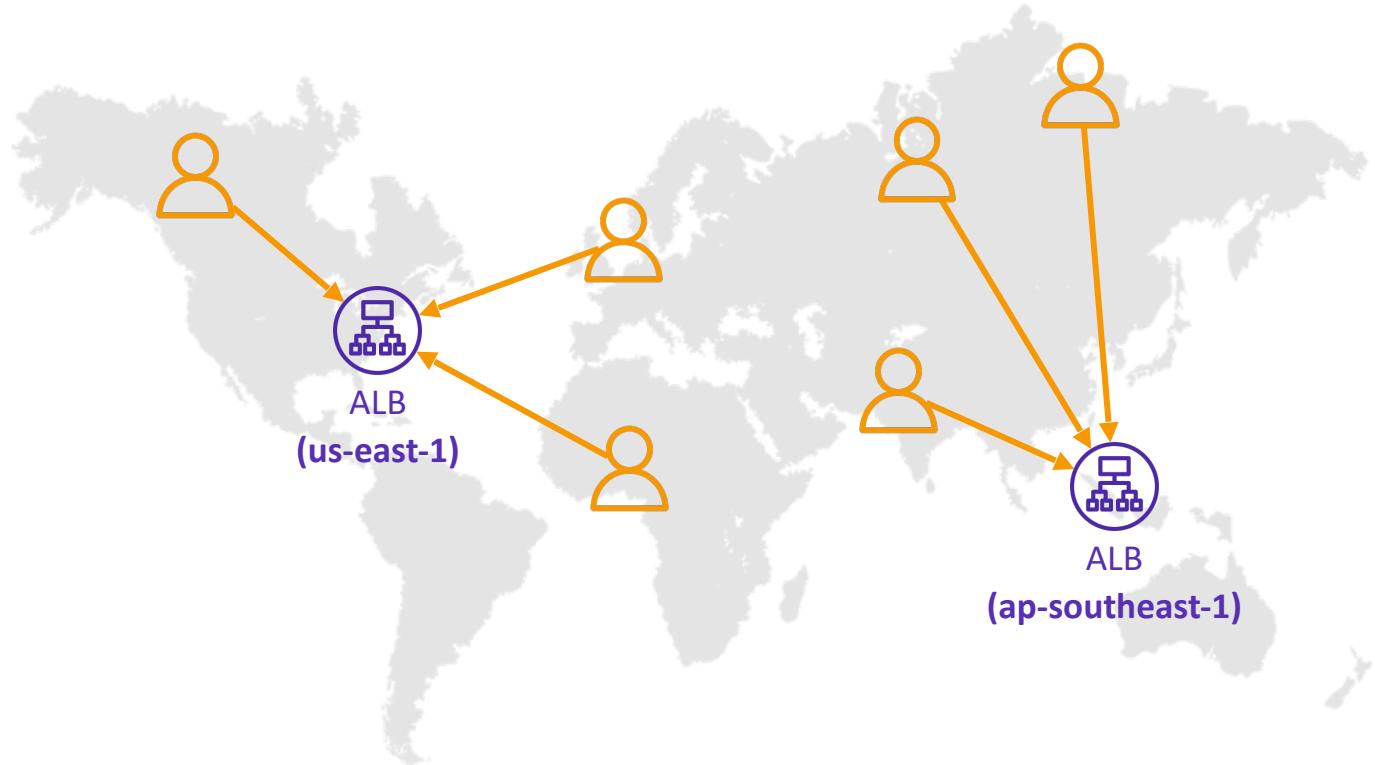
$$\bullet \text{ traffic } (\%) = \frac{\text{Weight for a specific record}}{\text{Sum of all the weights for all records}}$$

- No es necesario que los pesos sumen 100
- Los registros DNS deben tener el mismo nombre y tipo
- Pueden asociarse a las comprobaciones de salud
- Casos de uso: equilibrar la carga entre regiones, probar nuevas versiones de aplicaciones...
- **Asigna un peso de 0 a un registro para dejar de enviar tráfico a un recurso**
- **Si todos los registros tienen un peso de 0, se devolverán todos los registros por igual**



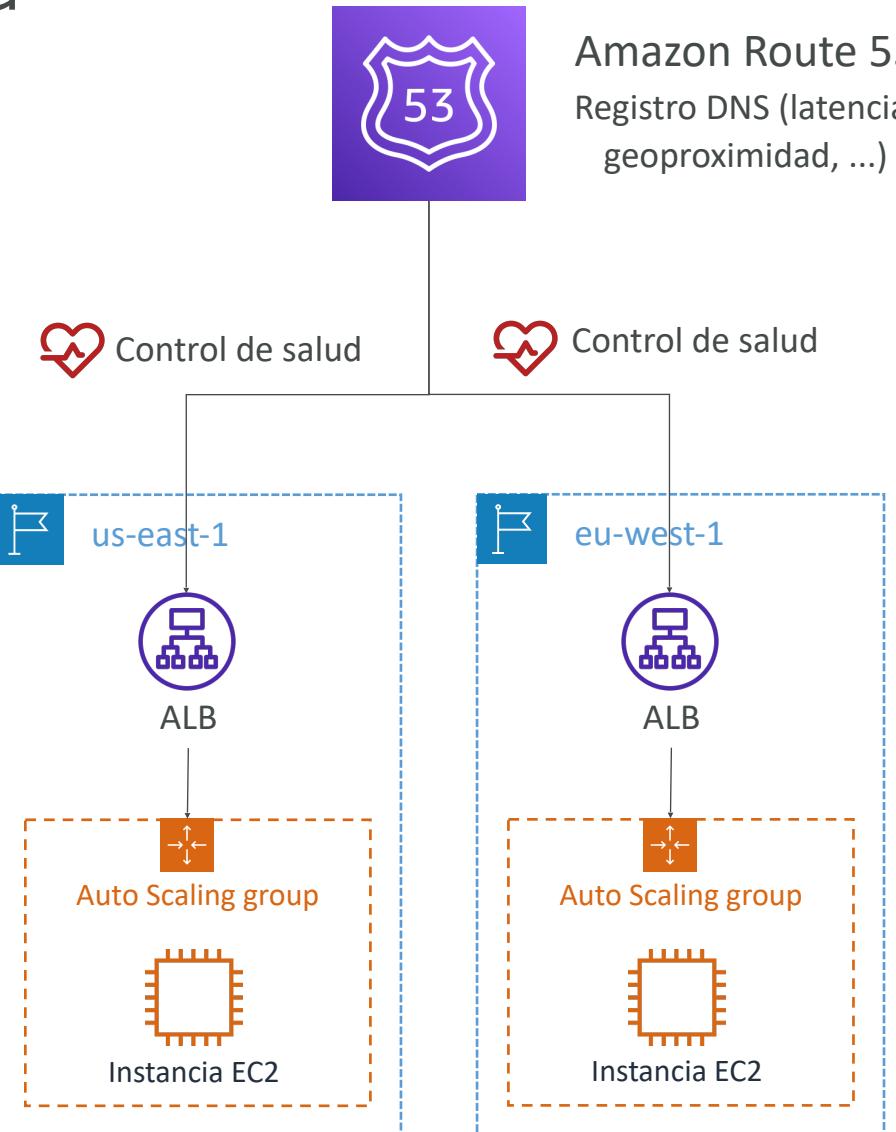
Políticas de enrutamiento - basadas en la latencia

- Redirigir al recurso que tenga la menor latencia cerca de nosotros
- Muy útil cuando la latencia para los usuarios es una prioridad
- **La latencia se basa en el tráfico entre los usuarios y las regiones de AWS**
- Los usuarios de Alemania pueden ser dirigidos a EEUU (si esa es la latencia más baja)
- Se puede asociar a los controles de salud (tiene capacidad de commutación por error)



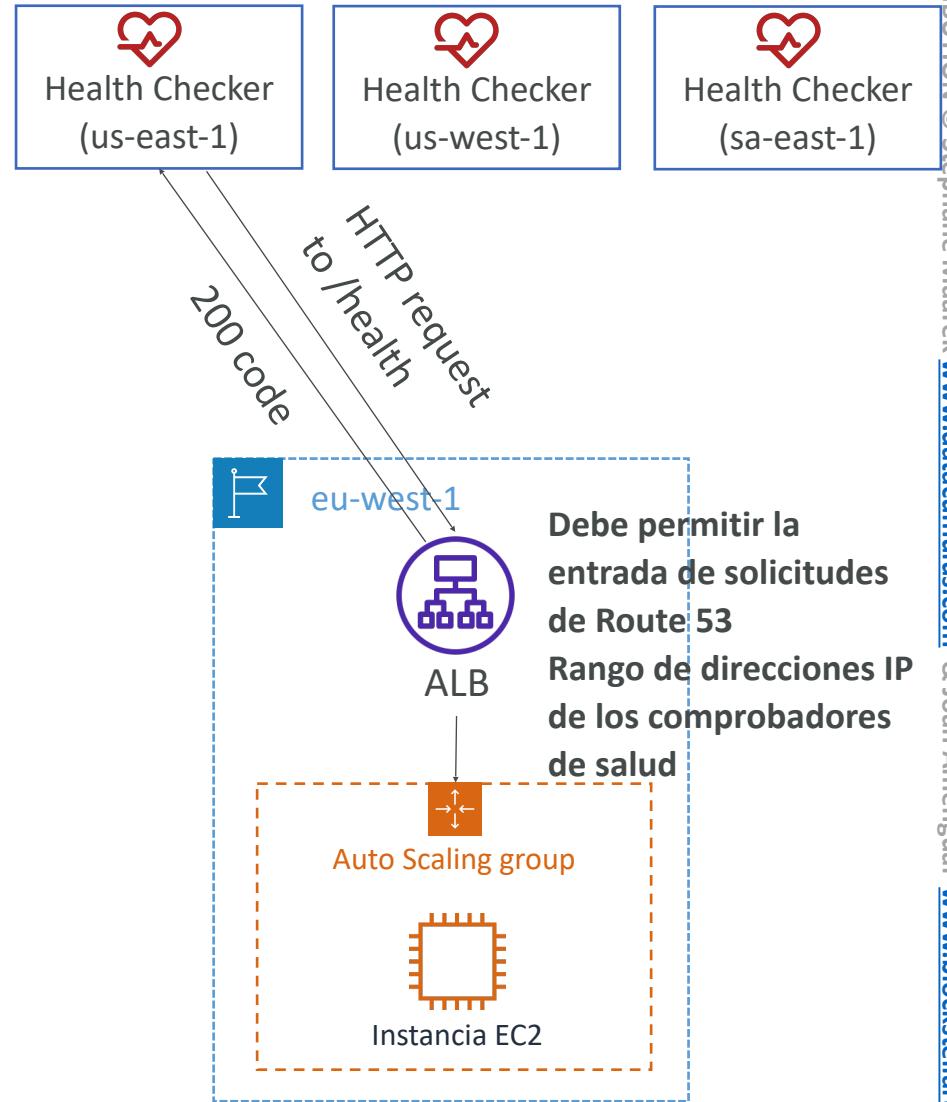
Route 53 - Controles de salud

- Las comprobaciones de salud HTTP son sólo para los **recursos públicos**
- Comprobación de la salud => Conmutación por error de DNS automatizada:
 1. Comprobaciones de salud que supervisan un endpoint (aplicación, servidor, otro recurso de AWS)
 2. Controles de salud que controlan otros controles de salud (Controles de salud calculados)
 3. Controles de salud que supervisan las alarmas de CloudWatch (control total!) - por ejemplo, alarmas en RDS, métricas personalizadas, ... (útil para recursos privados)
- Los Controles de Salud se integran con las métricas del CloudWatch



Comprobaciones de salud - Monitorizar un endpoint

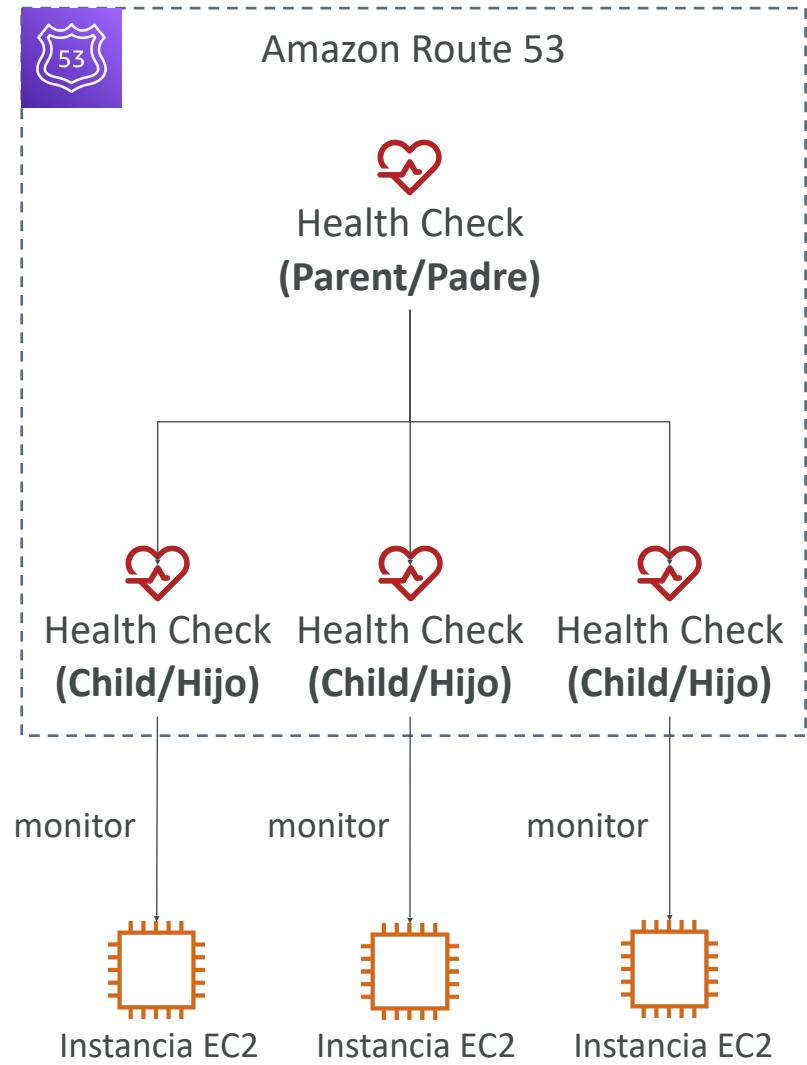
- **Unos 15 verificadores de salud globales comprobarán la salud del endpoint**
 - Umbral de salud/no salud - 3 (por defecto)
 - Intervalo - 30 segundos (se puede ajustar a 10 segundos - mayor coste)
 - Protocolo soportado: HTTP, HTTPS y TCP
 - Si $> 18\%$ de los comprobadores de salud informan de que el endpoint está sano, Route 53 lo considera **sano**. En caso contrario, se considera **no saludable**.
 - Posibilidad de elegir qué ubicaciones quieras que utilice Route 53
- Las comprobaciones de salud sólo pasan cuando el endpoint responde con los códigos de estado 2xx y 3xx
- Las comprobaciones de salud pueden configurarse para que pasen/no pasen en función del texto de los primeros **5120** bytes de la respuesta
- Configura tu router/firewall para permitir las solicitudes entrantes de los Health Checkers de Route 53



<https://ip-ranges.amazonaws.com/ip-ranges.json>

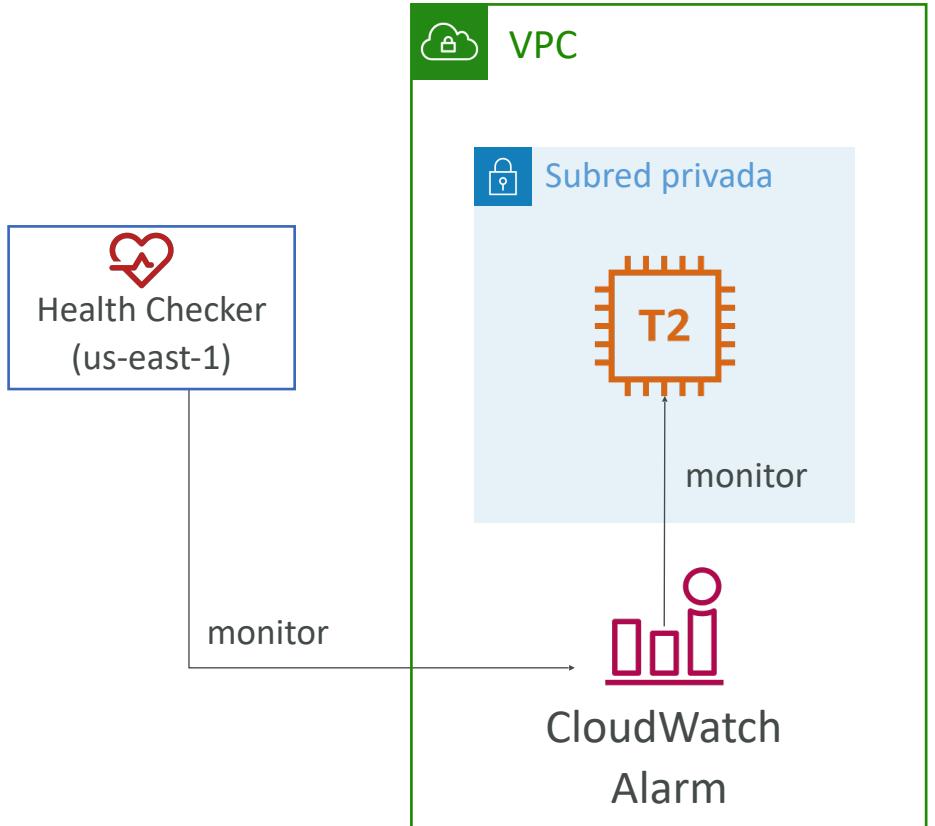
Route 53 - Controles de salud calculados

- Combinar los resultados de varios chequeos de salud en un solo chequeo de salud
- Puedes utilizar **OR, AND o NOT**
- Puedes controlar hasta 256 chequeos médicos de los “hijos” (child)
- Especifica cuántas de las comprobaciones de salud deben pasar para que “el padre” (parent) pase
- Uso: realiza el mantenimiento de tu sitio web sin que fallen todas las comprobaciones de salud

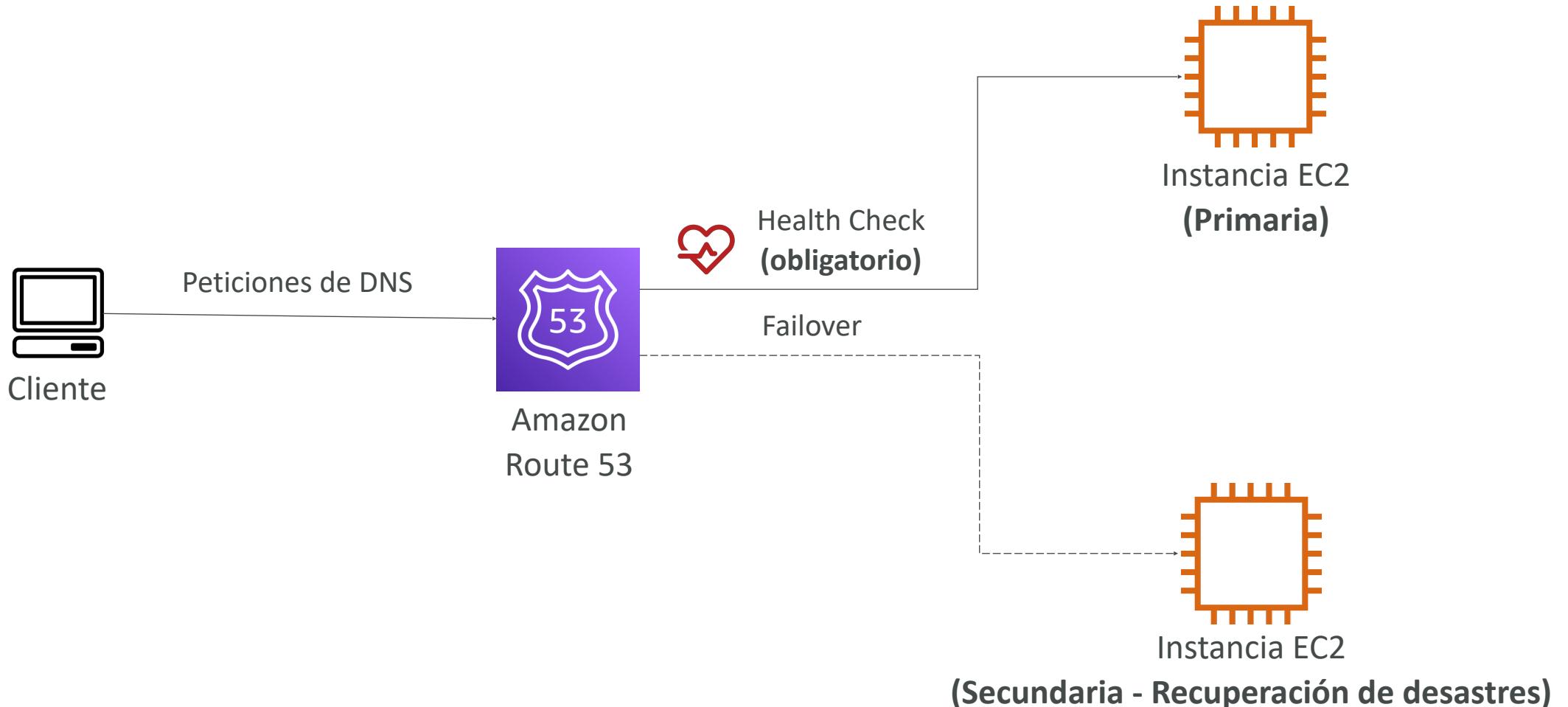


Controles de salud - Zonas de alojamiento privadas

- Los comprobadores de salud (Health Checker) de Route 53 están fuera de la VPC
- No pueden acceder a endpoints **privados** (VPC privada o recurso local)
- Puedes crear una **métrica de CloudWatch** y asociar una **alarma de CloudWatch**, y luego crear un chequeo de salud que compruebe la propia alarma

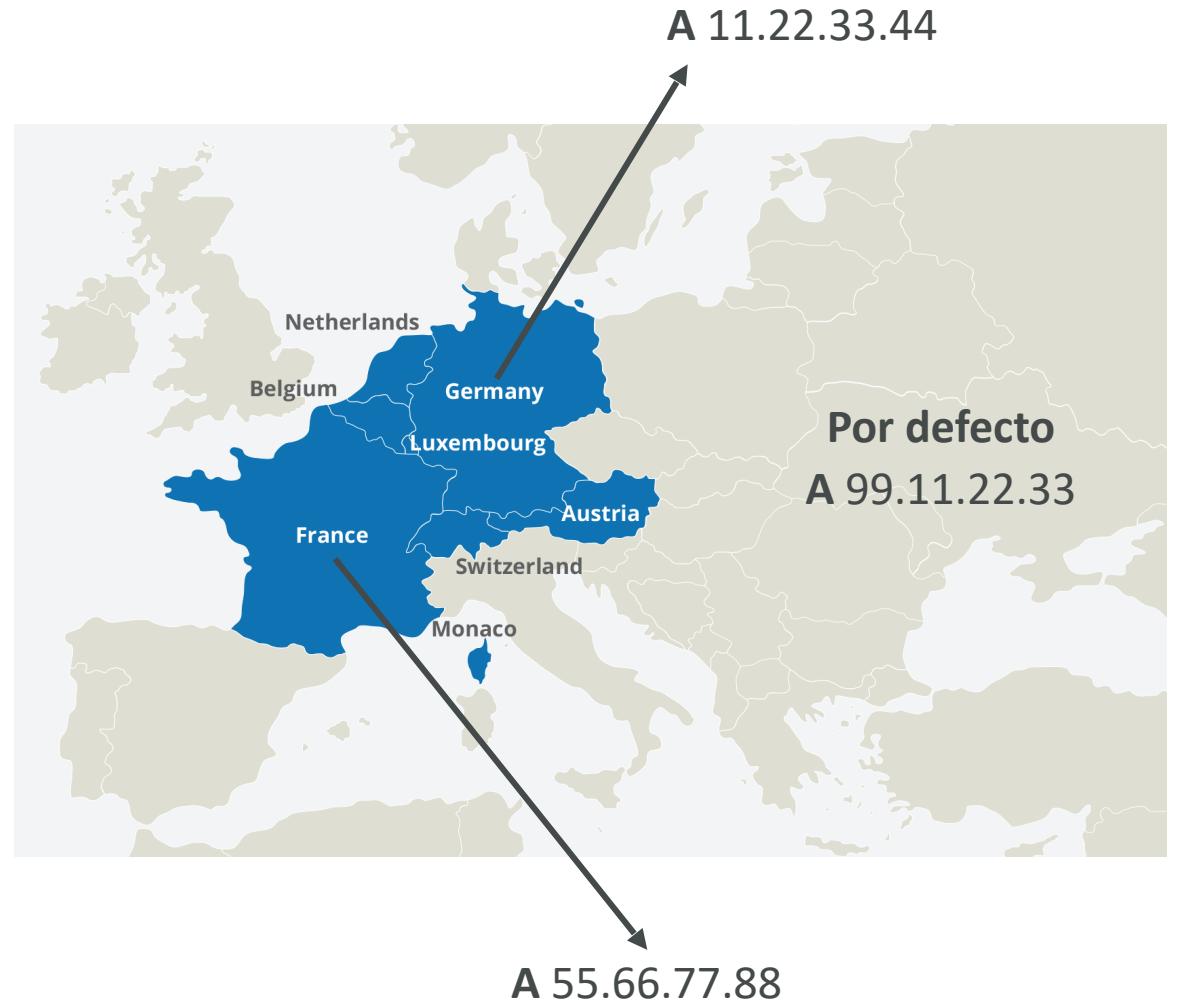


Políticas de enrutamiento - Comutación por error (activo-pasivo)



Políticas de enrutamiento - Geolocalización

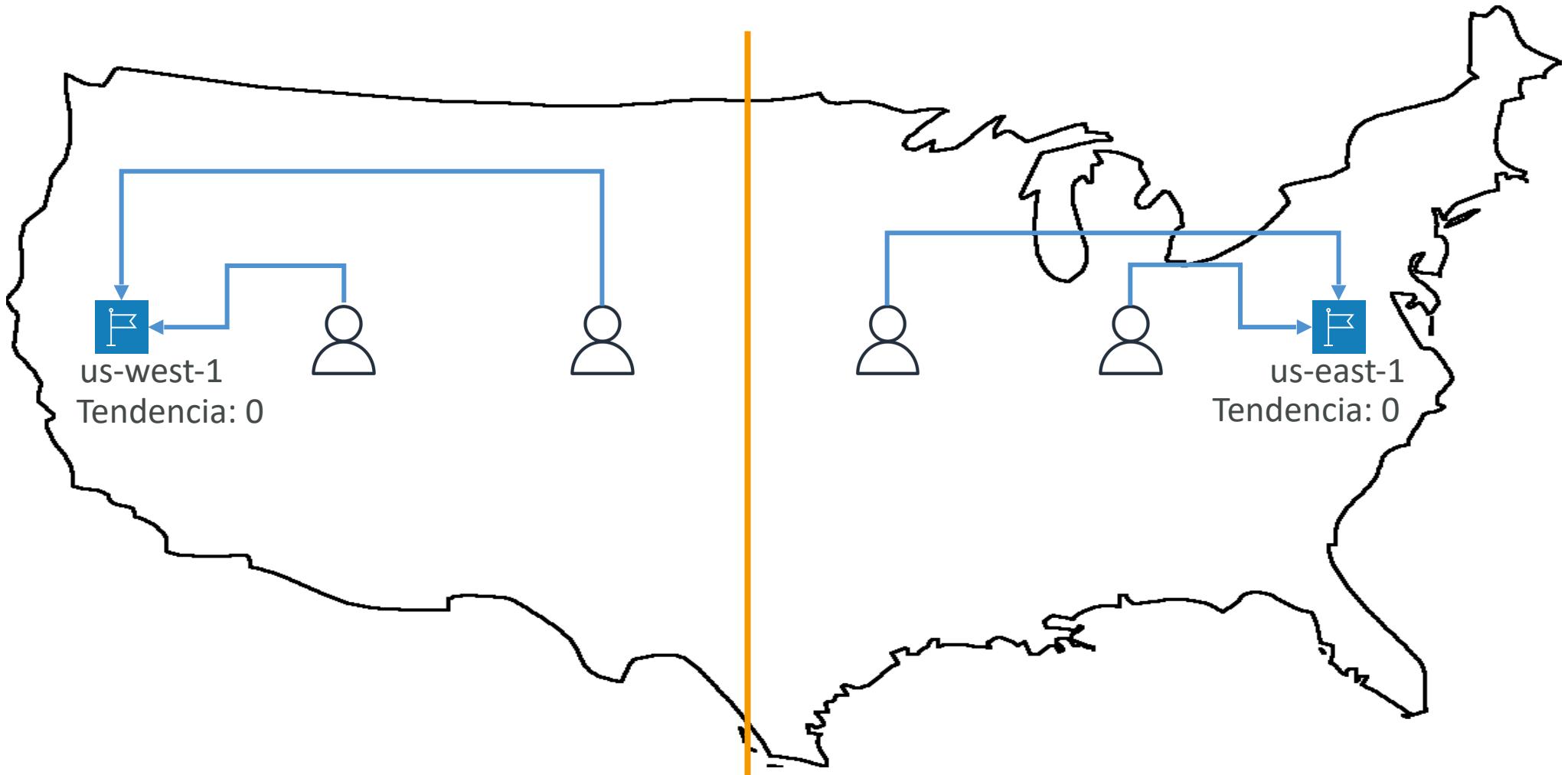
- ¡Diferente al basado en la latencia!
- **Este enrutamiento se basa en la ubicación del usuario**
- Especifica la ubicación por continente, país o estado de EE.UU. (si hay solapamiento, se selecciona la ubicación más precisa)
- Debe crear un registro "**por defecto**" (en caso de que no haya coincidencia en la ubicación)
- Casos de uso: localización de sitios web, restringir la distribución de contenidos, equilibrar la carga, ...
- Puede asociarse a los controles de salud



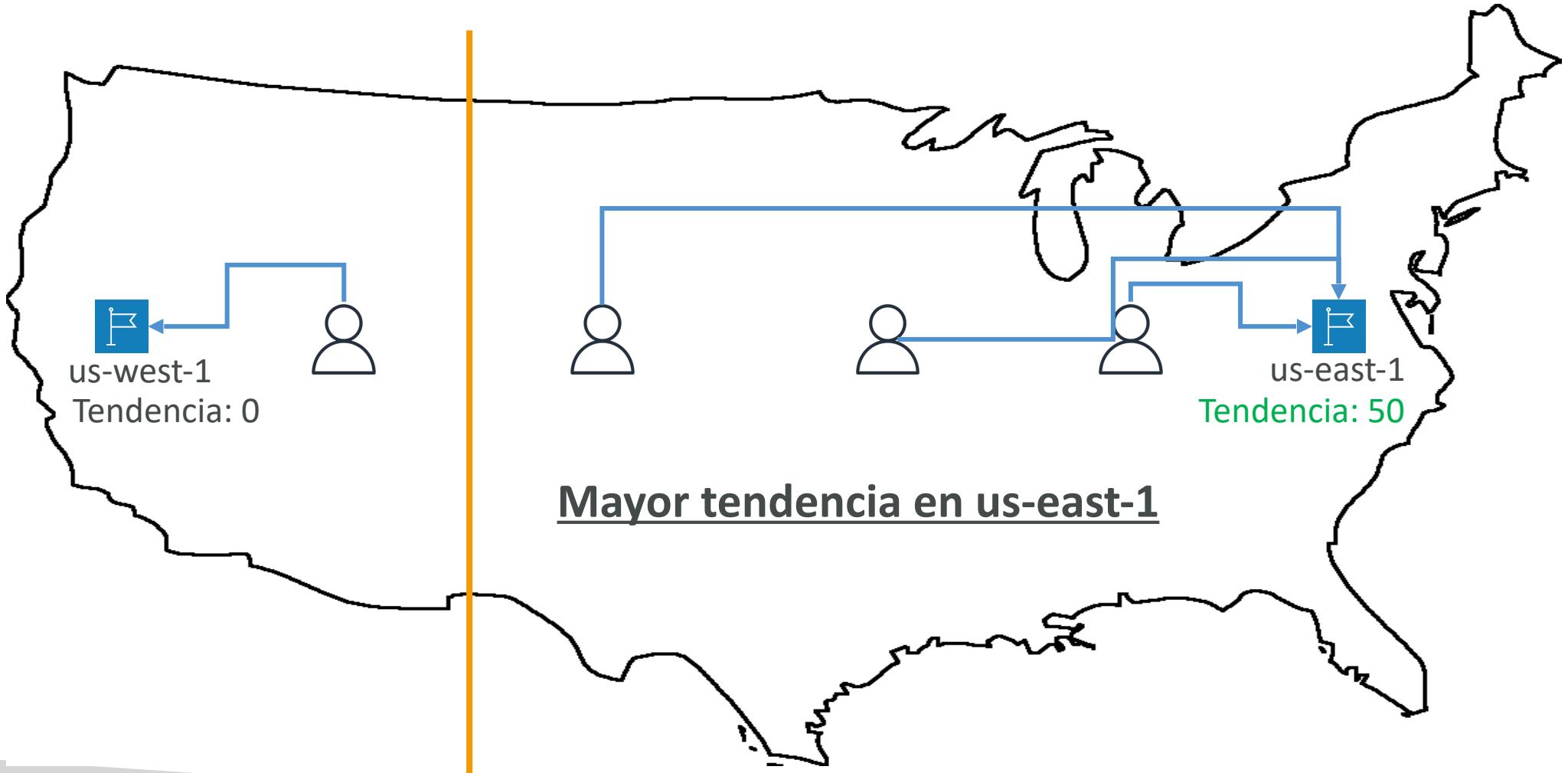
Políticas de enrutamiento - Geoproximidad

- Dirige el tráfico a tus recursos en función de la ubicación geográfica de los usuarios y los recursos
- Posibilidad de desplazar más tráfico a los recursos en función del sesgo definido
- Para cambiar el tamaño de la región geográfica, especifica los valores del sesgo:
 - Para ampliar (1 a 99) - más tráfico hacia el recurso
 - Para reducir (-1 a -99) - menos tráfico hacia el recurso
- Los recursos pueden ser
 - Recursos AWS (especifica la región AWS)
 - Recursos no AWS (especifica latitud y longitud)
- Debes utilizar el flujo de tráfico de Route 53 para utilizar esta función

Políticas de enrutamiento - Geoproximidad



Políticas de enrutamiento - Geoproximidad



Políticas de enrutamiento - Multivalores

- Utilízalo cuando dirijas el tráfico a múltiples recursos
- Route 53 devuelve múltiples valores/recursos
- Puede asociarse a comprobaciones de salud (devuelve sólo los valores de los recursos sanos)
- Se devuelven hasta 8 registros sanos por cada consulta Multi-Value
- **El Multi-Valor no sustituye a tener un ELB**

Name	Type	Value	TTL	Set ID	Health Check
www.example.com	A Record	192.0.2.2	60	Web1	A
www.example.com	A Record	198.51.100.2	60	Web2	B
www.example.com	A Record	203.0.113.2	60	Web3	C

Registrar el dominio vs. Servicio DNS

- Compras o registras tu nombre de dominio con un Registrador de Dominios, normalmente pagando una cuota anual (por ejemplo, GoDaddy, Amazon Registrar Inc., ...)
- El Registrador de dominios suele proporcionarte un servicio de DNS para gestionar tus registros de DNS
- Pero puedes utilizar otro servicio DNS para gestionar tus registros DNS
- Ejemplo: compra el dominio a GoDaddy y utiliza Route 53 para gestionar tus registros DNS



GoDaddy como registrador y Route 53 como servicio DNS



Records

We can't display your DNS information because your nameservers aren't managed by us.

Nameservers

Using custom nameservers [Change](#)

Nameserver
ns-1083.awsdns-07.org
ns-932.awsdns-52.net
ns-1911.awsdns-46.co.uk
ns-481.awsdns-60.com



Amazon
Route 53

Zona de alojamiento público
jamengual.com

▼ Hosted zone details [Edit hosted zone](#)

Hosted zone ID	Type
Z30IJCCWPKZUV	Public hosted zone
Description	Record count
HostedZone created by Route53 Registrar	22
Query log	

Name servers

- ns-252.awsdns-31.com
- ns-1468.awsdns-55.org
- ns-633.awsdns-15.net
- ns-1800.awsdns-33.co.uk

Registrador de terceros con Amazon Route 53

- **Si compras tu dominio en un registrador de terceros, puedes seguir utilizando Route 53 como proveedor de servicios DNS**
 1. Crear una Zona de alojamiento en Route 53
 2. Actualizar los registros NS en el sitio web de terceros para utilizar los servidores de nombres de Route 53
- **Registrador de dominios != Servicio DNS**
- Pero todos los Registradores de Dominio suelen tener algunas funciones DNS

Soluciones clásicas de arquitectura

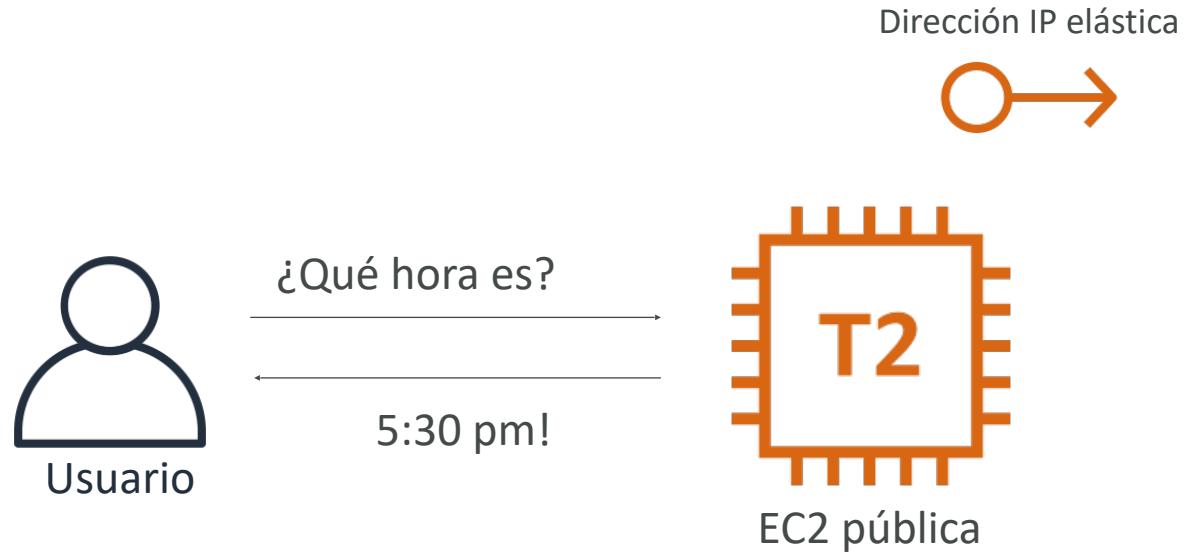
Introducción de la sección

- Estas arquitecturas de soluciones son la mejor parte de este curso
- Vamos a entender cómo funcionan juntas todas las tecnologías que hemos visto
- Esta es una sección con la que debes sentirte 100% cómodo
- Veremos la progresión de la mentalidad de un arquitecto de soluciones a través de muchos ejemplos de casos prácticos:
 - WhatIsTheTime.Com
 - MyClothes.Com
 - MyWordPress.Com
 - Instanciar aplicaciones rápidamente
 - Beanstalk

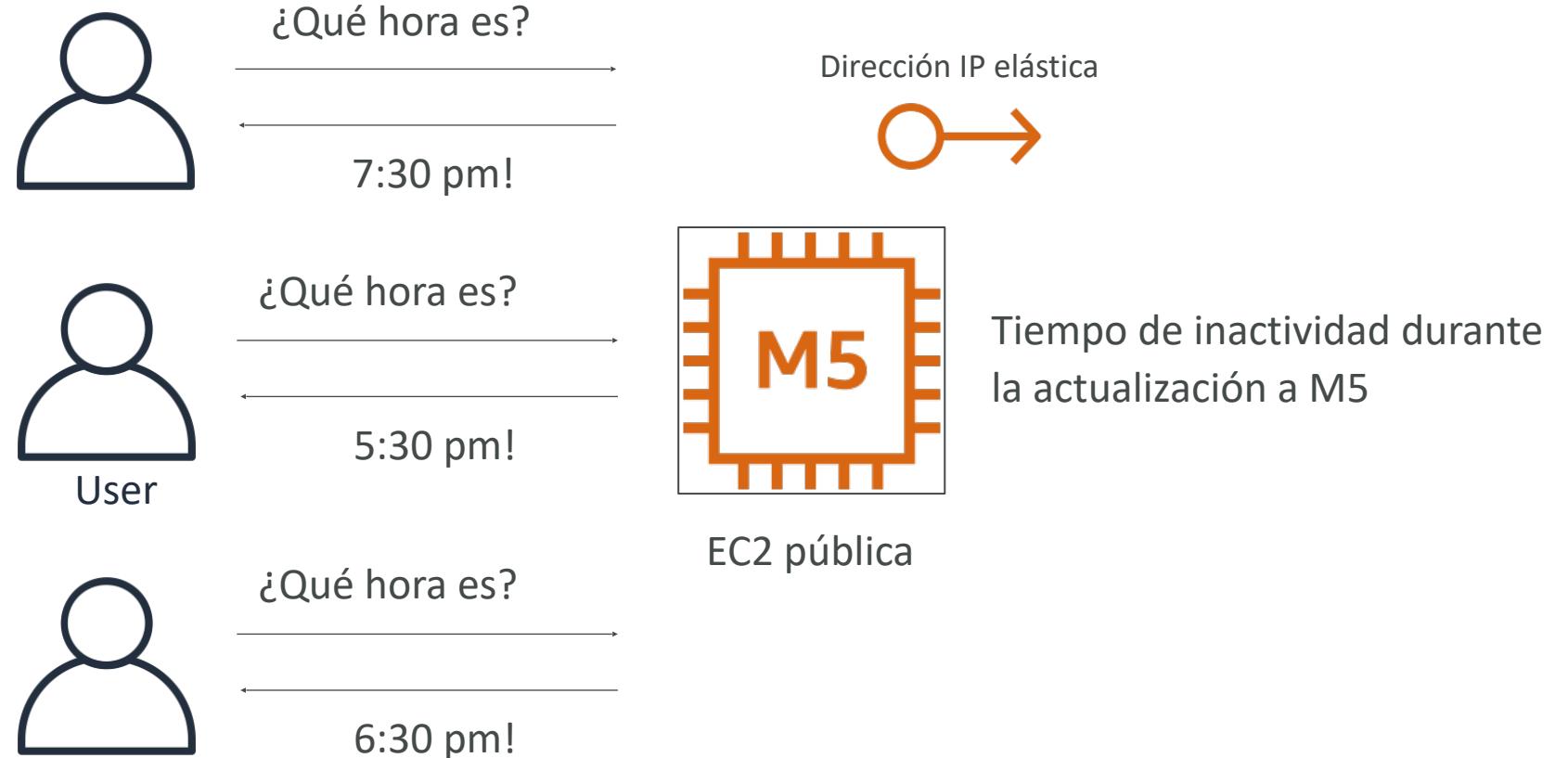
Aplicación web sin estado: WhatIsTheTime.com

- WhatIsTheTime.com permite a la gente saber qué hora es
- No necesitamos una base de datos
- Queremos empezar poco a poco y podemos aceptar el tiempo de inactividad
- Queremos escalar completamente de forma vertical y horizontal, sin tiempo de inactividad
- Vamos a recorrer el camino del Arquitecto de Soluciones para esta aplicación
- ¡Veamos cómo podemos proceder!

Aplicación web sin estado: ¿Qué hora es? Inicio simple



Aplicación web sin estado: ¿Qué hora es? Escalando verticalmente



Aplicación web sin estado: ¿Qué hora es? Escalando horizontalmente



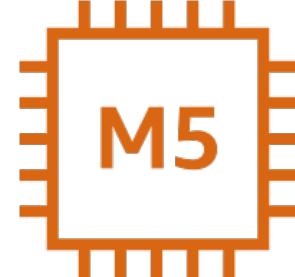
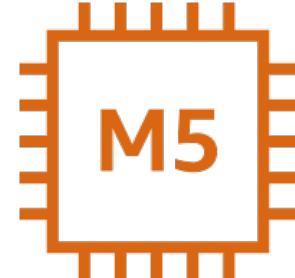
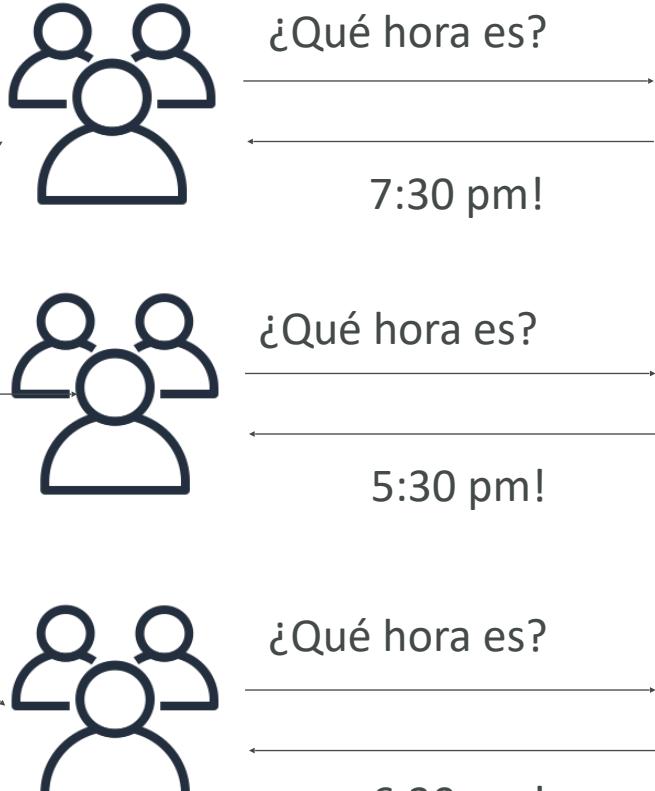
Aplicación web sin estado: ¿Qué hora es? Escalando horizontalmente

Consulta DNS

Para api.whatisthetime.com

Registro A

TTL 1 hora

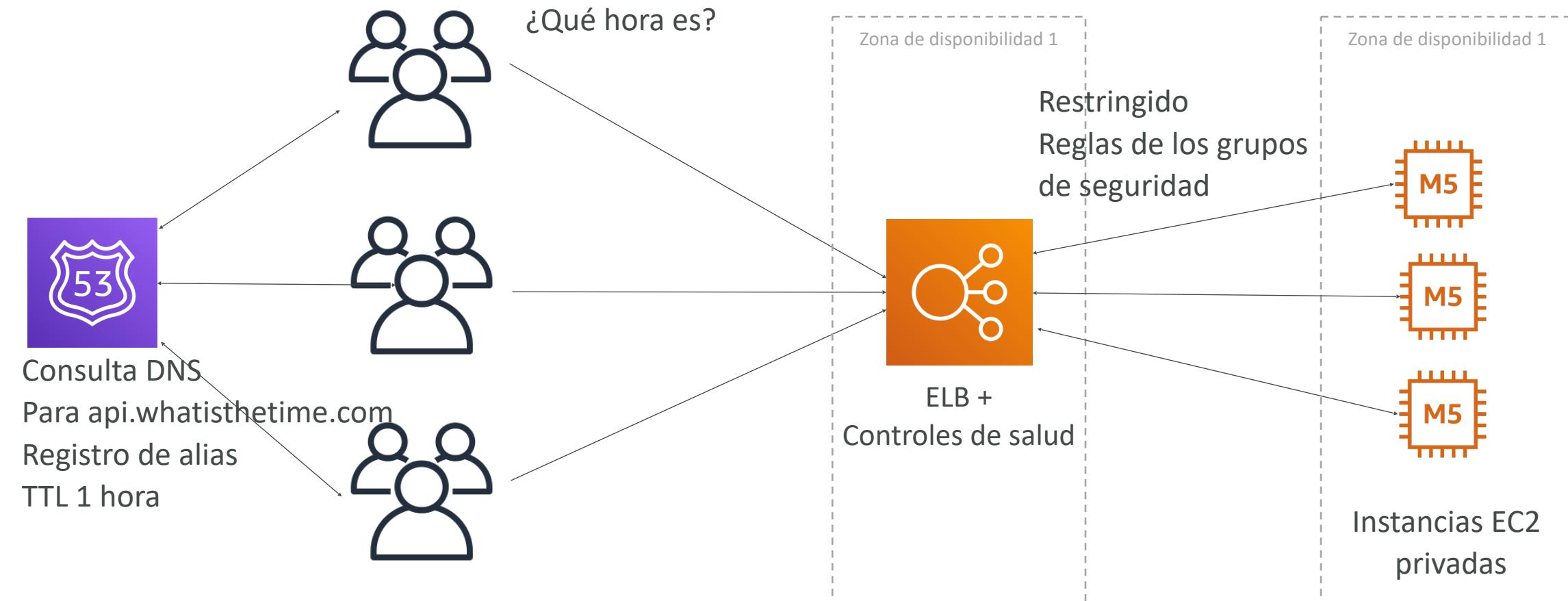


Instancia EC2 pública,
sin IP elástica

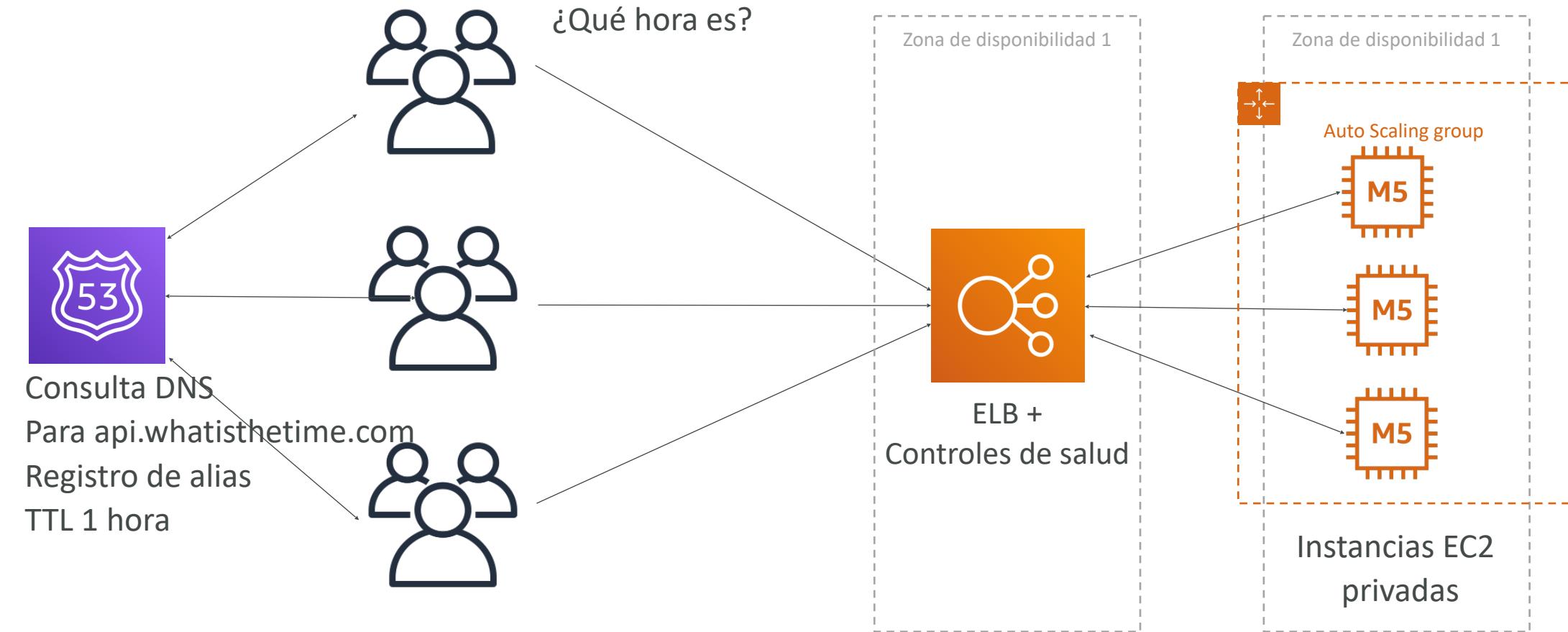
Aplicación web sin estado: ¿Qué hora es? Escalando horizontalmente: añadir y eliminar instancias



Aplicación web sin estado: ¿Qué hora es? Escalando horizontalmente, con un Load Balancer



Aplicación web sin estado: ¿Qué hora es? Escalando horizontalmente, con un grupo de autoescalamiento



Aplicación web sin estado: ¿Qué hora es? Hacer que nuestra aplicación sea multi-AZ



Mínimo 2 AZ => Reservemos capacidad



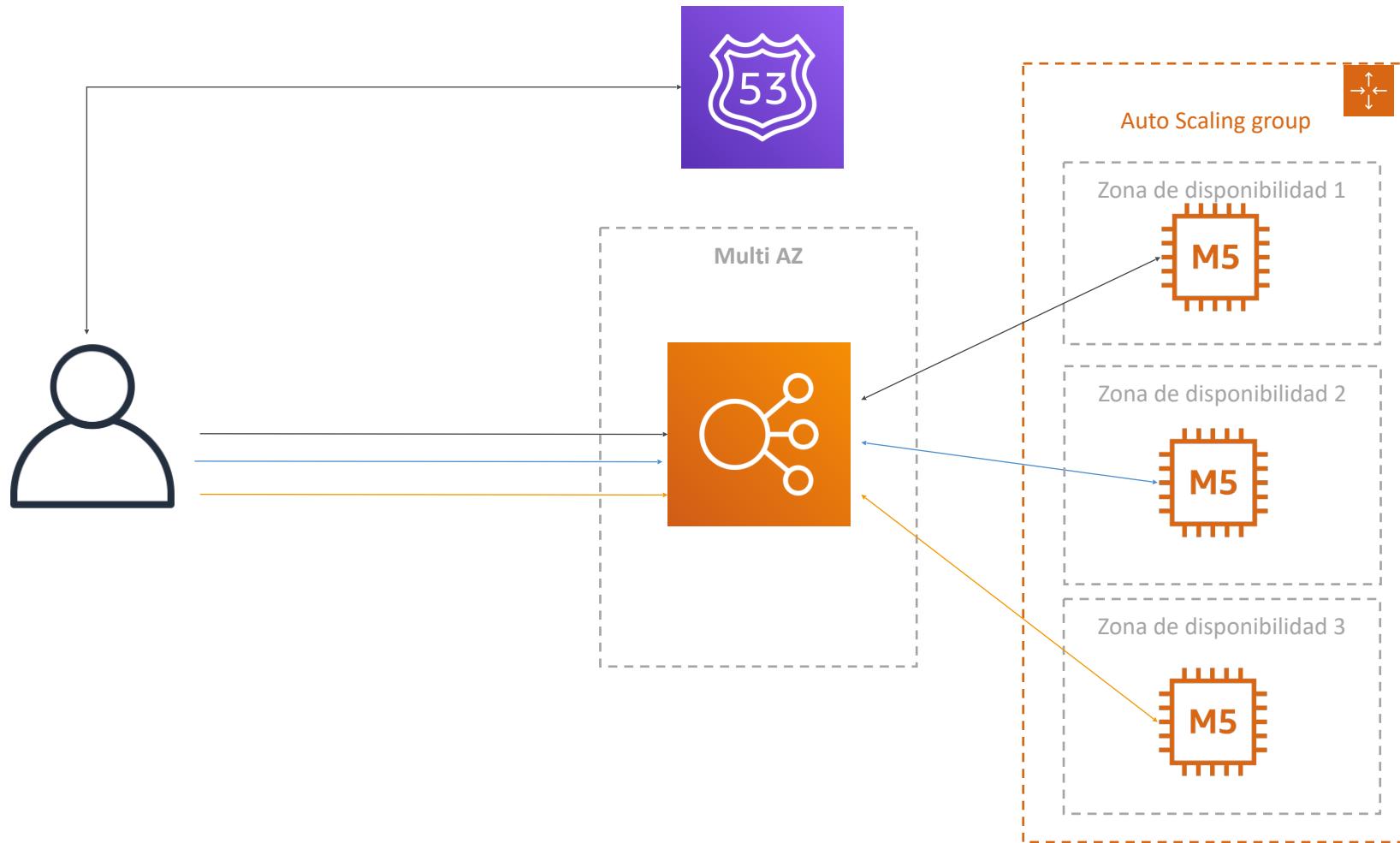
En esta clase hemos hablado de...

- IP pública vs privada e instancias EC2
- Elastic IP vs Route 53 vs Load Balancers
- Route 53 TTL, registros A y registros Alias
- Mantener las instancias EC2 manualmente vs Auto Scaling Group
- Multi AZ para sobrevivir a los desastres
- Comprobaciones de salud del ELB
- Reglas de grupos de seguridad
- Reserva de capacidad para ahorrar costes cuando sea posible
- Estamos considerando 5 pilares para una aplicación bien arquitecturada: costes, rendimiento, fiabilidad, seguridad, excelencia operativa

Aplicación web con estado: MyClothes.com

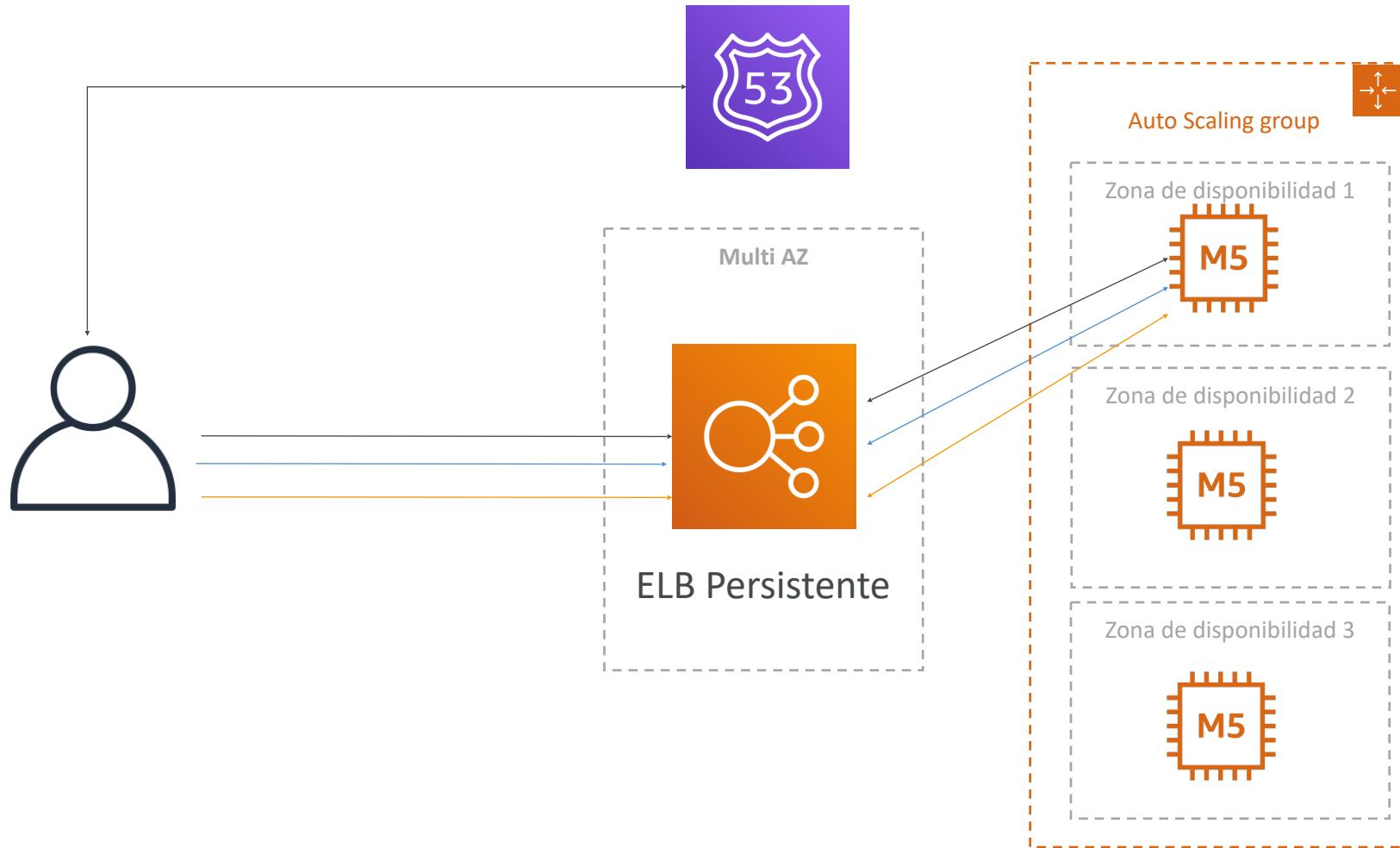
- MyClothes.com permite comprar ropa en línea.
- Hay un carrito de la compra
- Nuestro sitio web tiene cientos de usuarios al mismo tiempo
- Necesitamos escalar, mantener la escalabilidad horizontal y mantener nuestra aplicación web lo más sin estado posible
- Los usuarios no deben perder su carrito de la compra
- Los usuarios deben tener sus datos (dirección, etc.) en una base de datos
- ¡Vemos cómo podemos proceder!

Aplicación web con estado: MyClothes.com



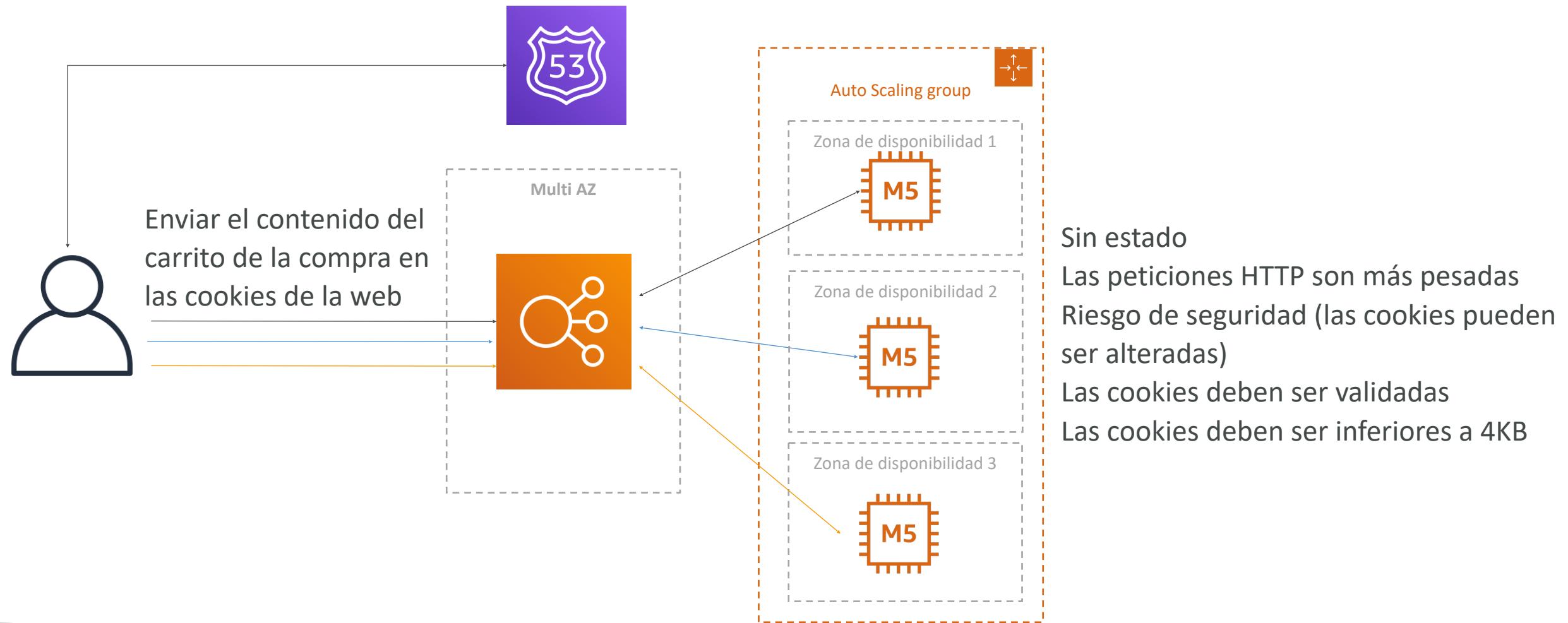
Aplicación web con estado: MyClothes.com

Introduce a Stickiness (sesiones persistentes)



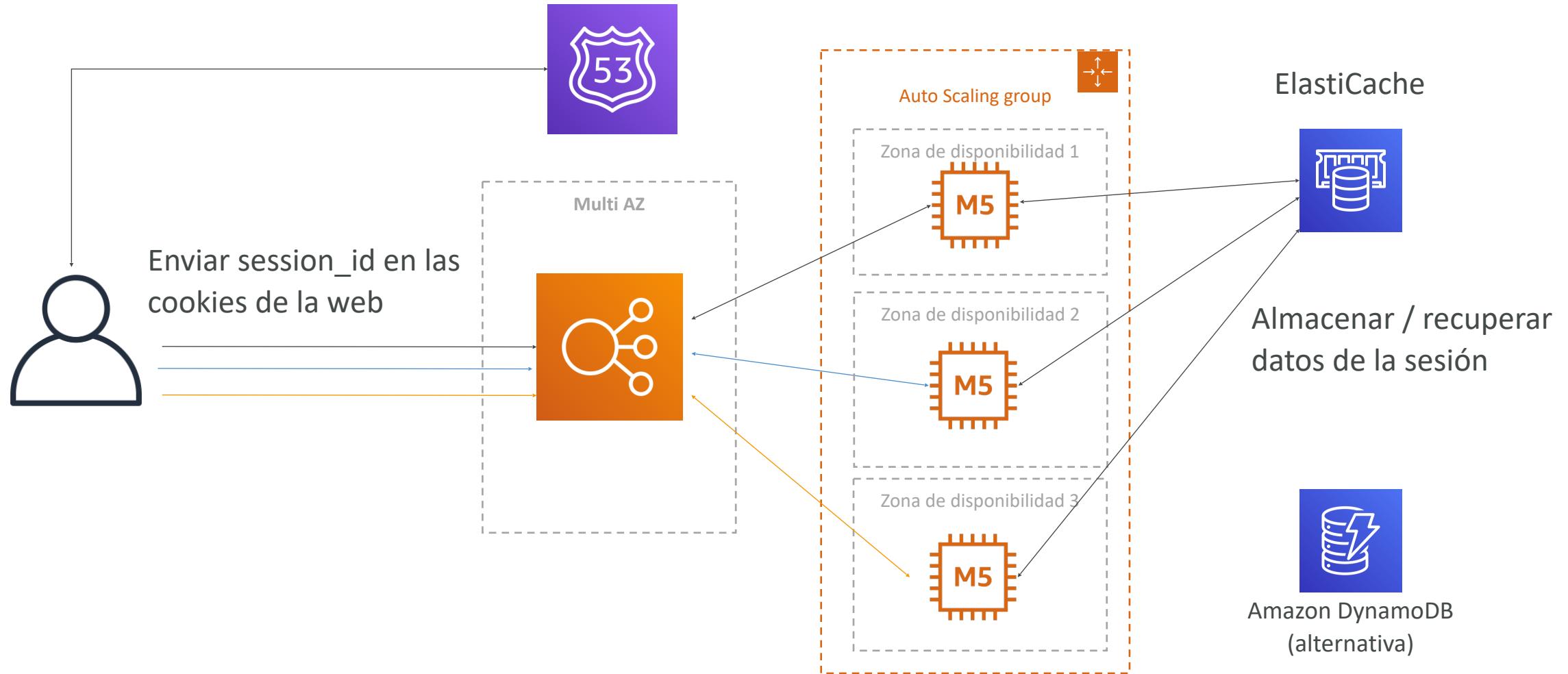
Aplicación web con estado: MyClothes.com

Introduce las cookies de usuario



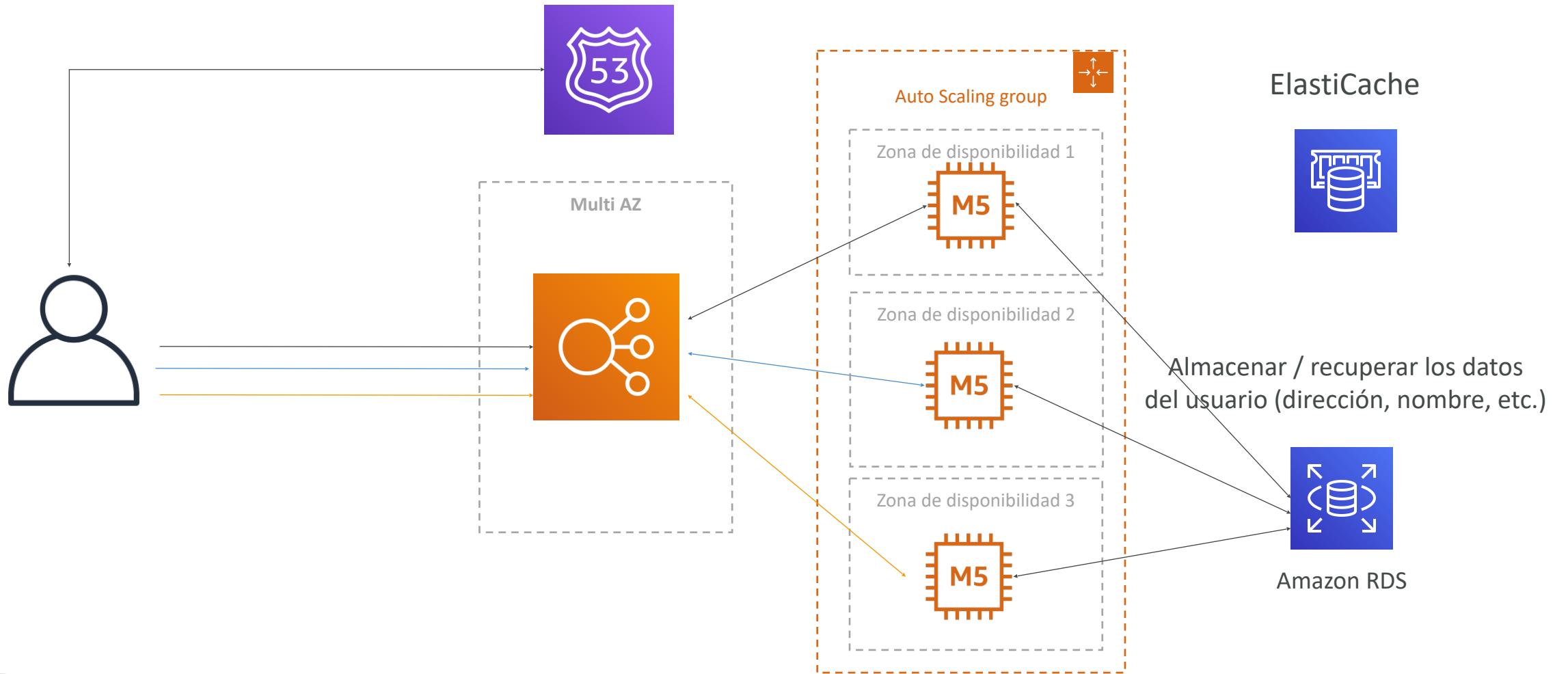
Aplicación web con estado: MyClothes.com

Introduce la sesión del servidor



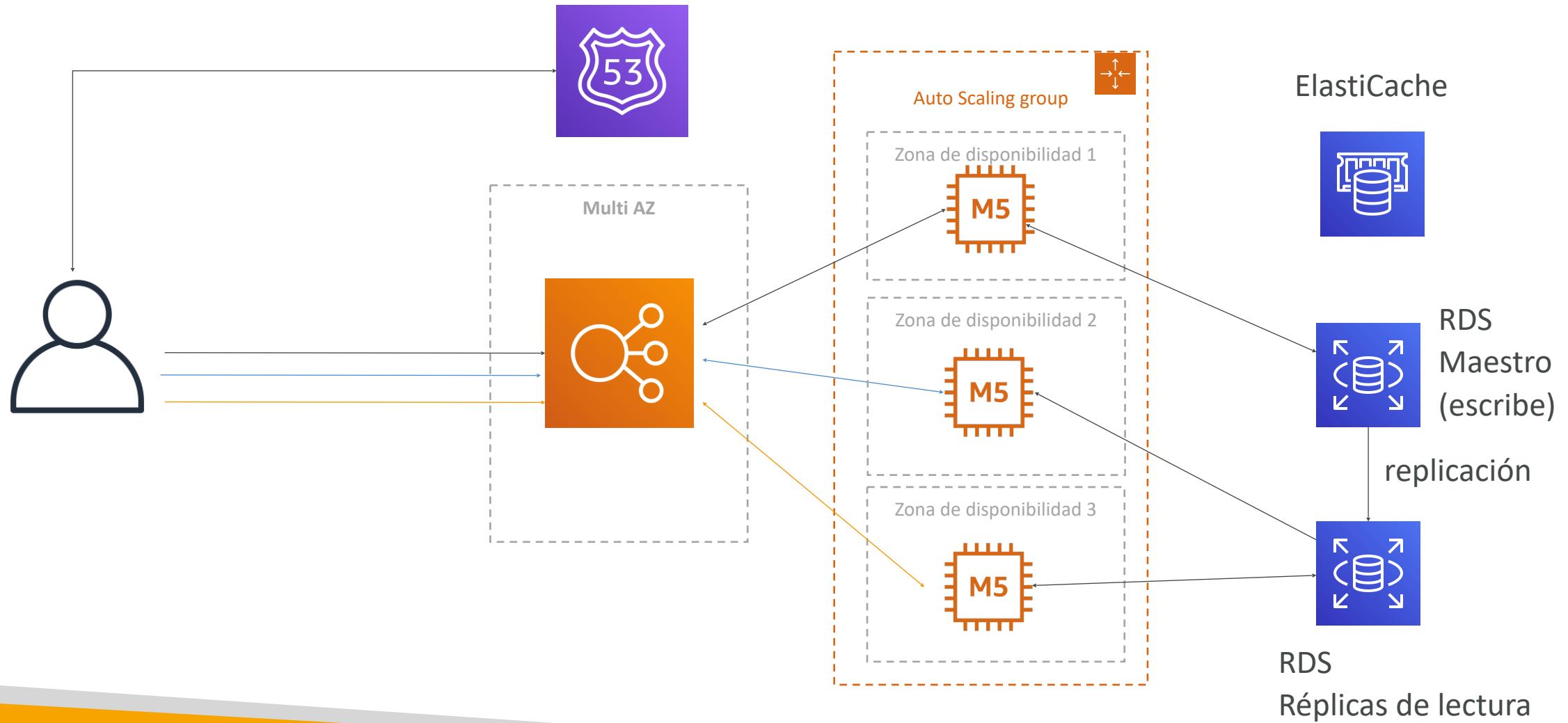
Aplicación web con estado: MyClothes.com

Almacenamiento de los datos del usuario en una base de datos



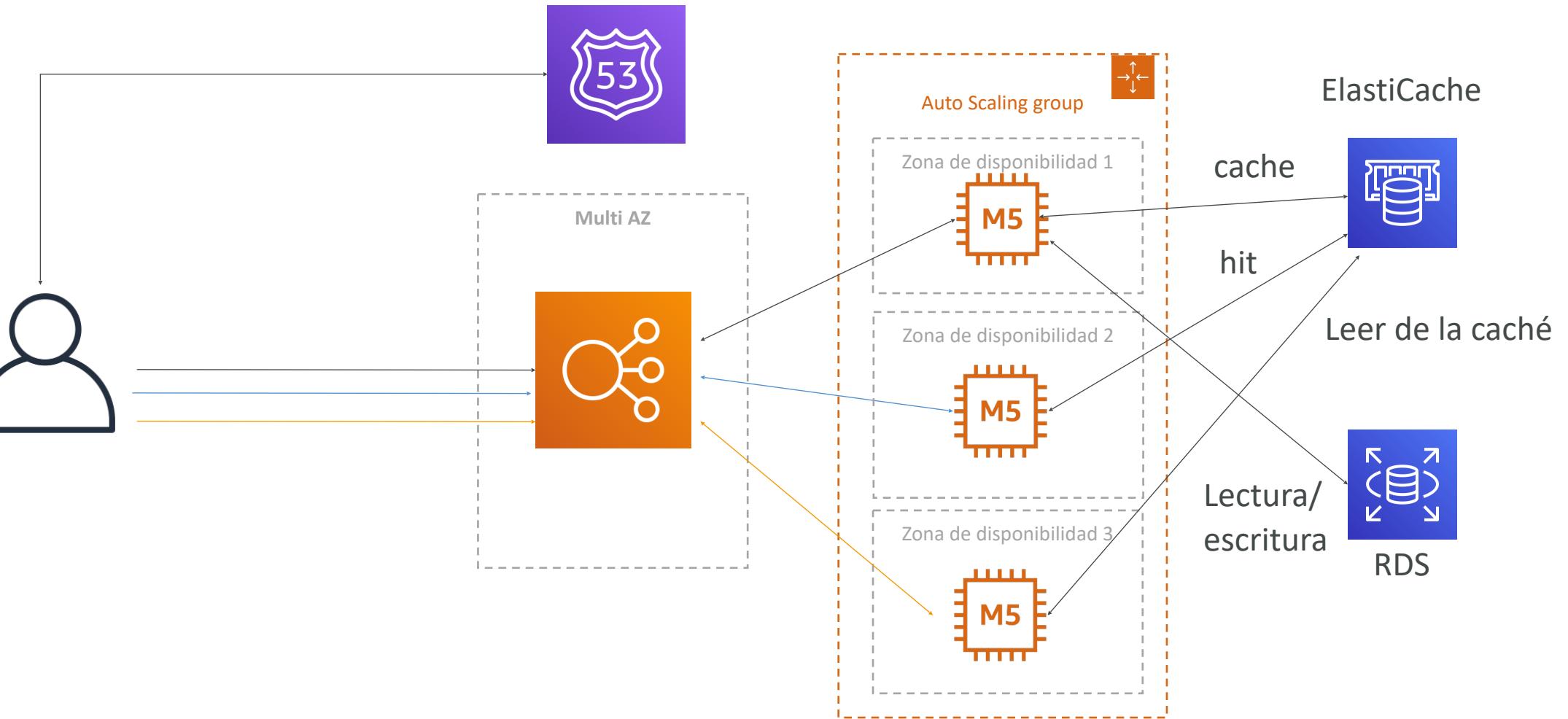
Aplicación web con estado: MyClothes.com

Lecturas de escalado



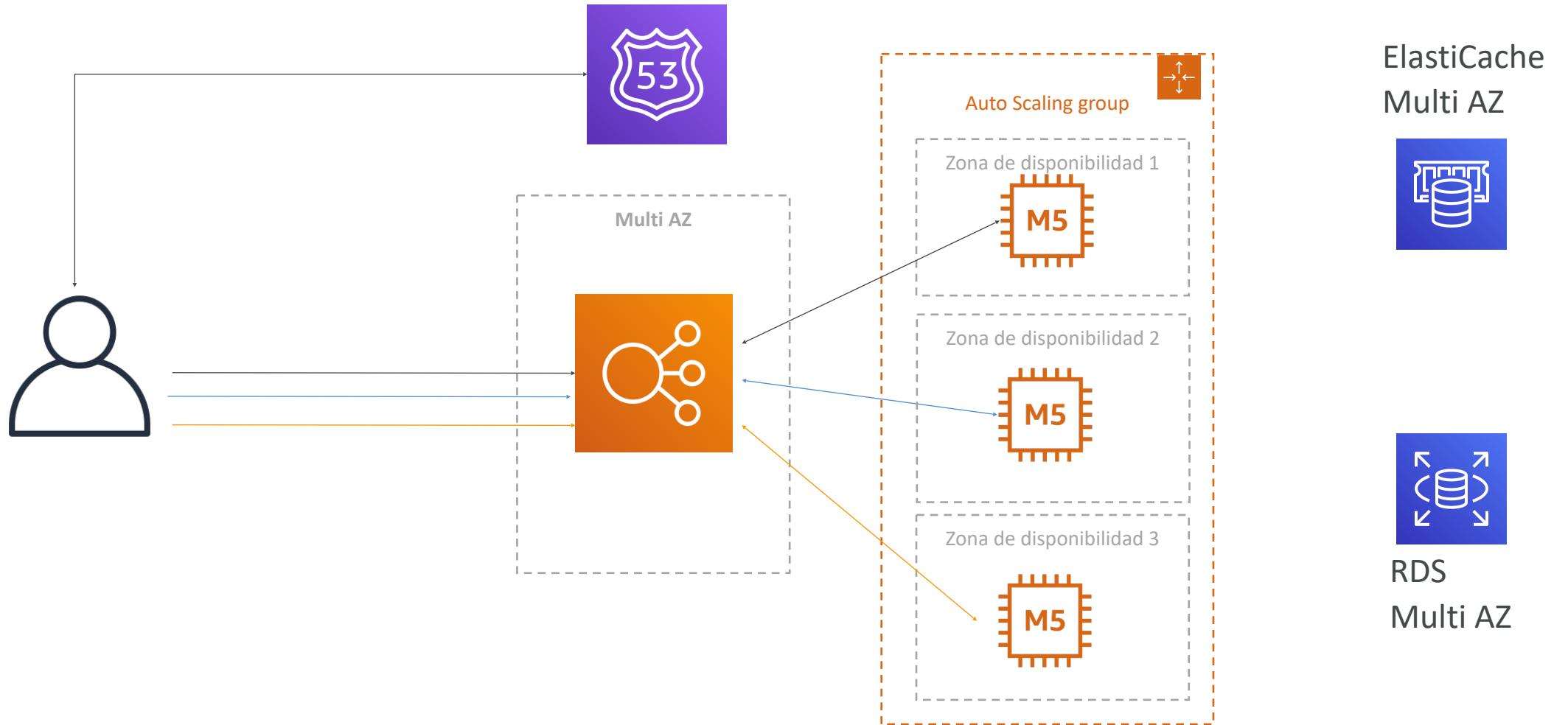
Aplicación web con estado: MyClothes.com

Escalando lecturas (Alternativa) - Escritura



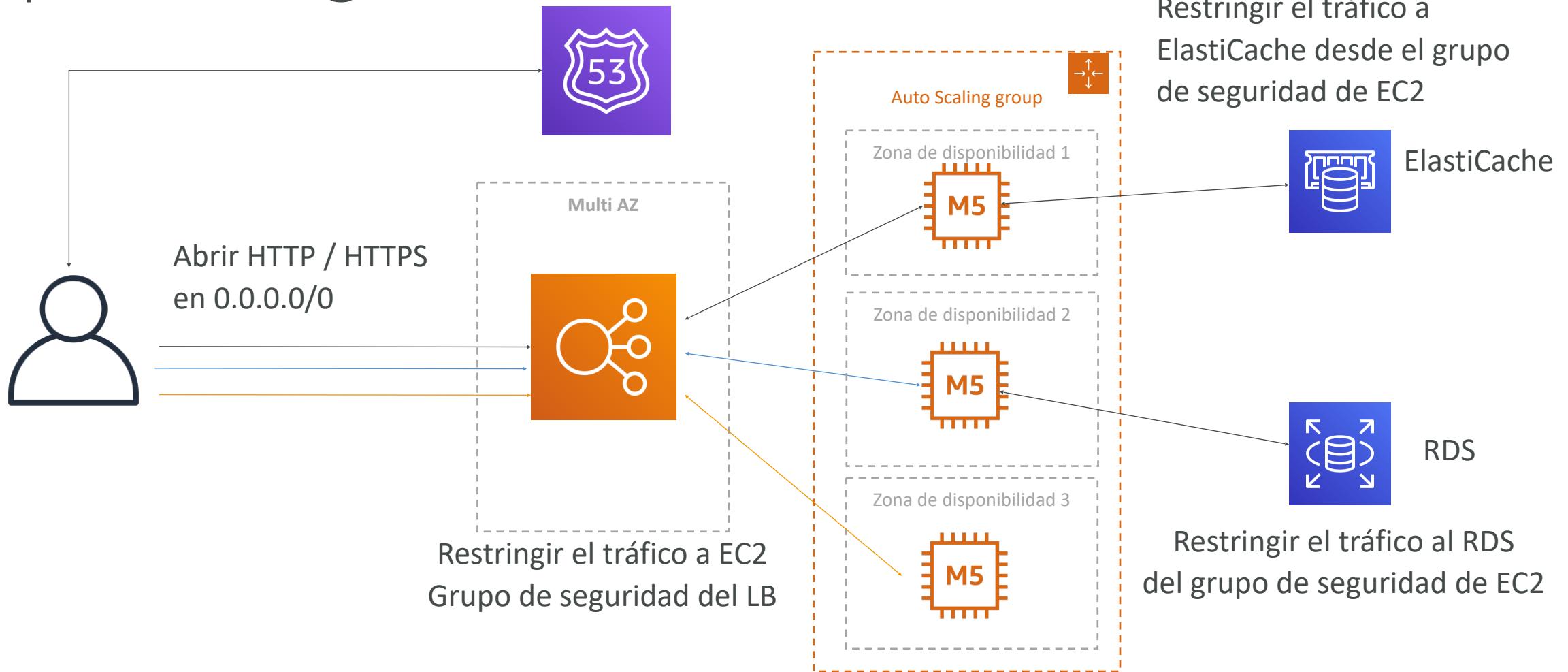
Aplicación web con estado: MyClothes.com

Multi AZ - Sobrevivir a los desastres



Aplicación web con estado: MyClothes.com

Grupos de seguridad



En esta clase hemos hablado de...

Arquitecturas de 3 niveles para aplicaciones web

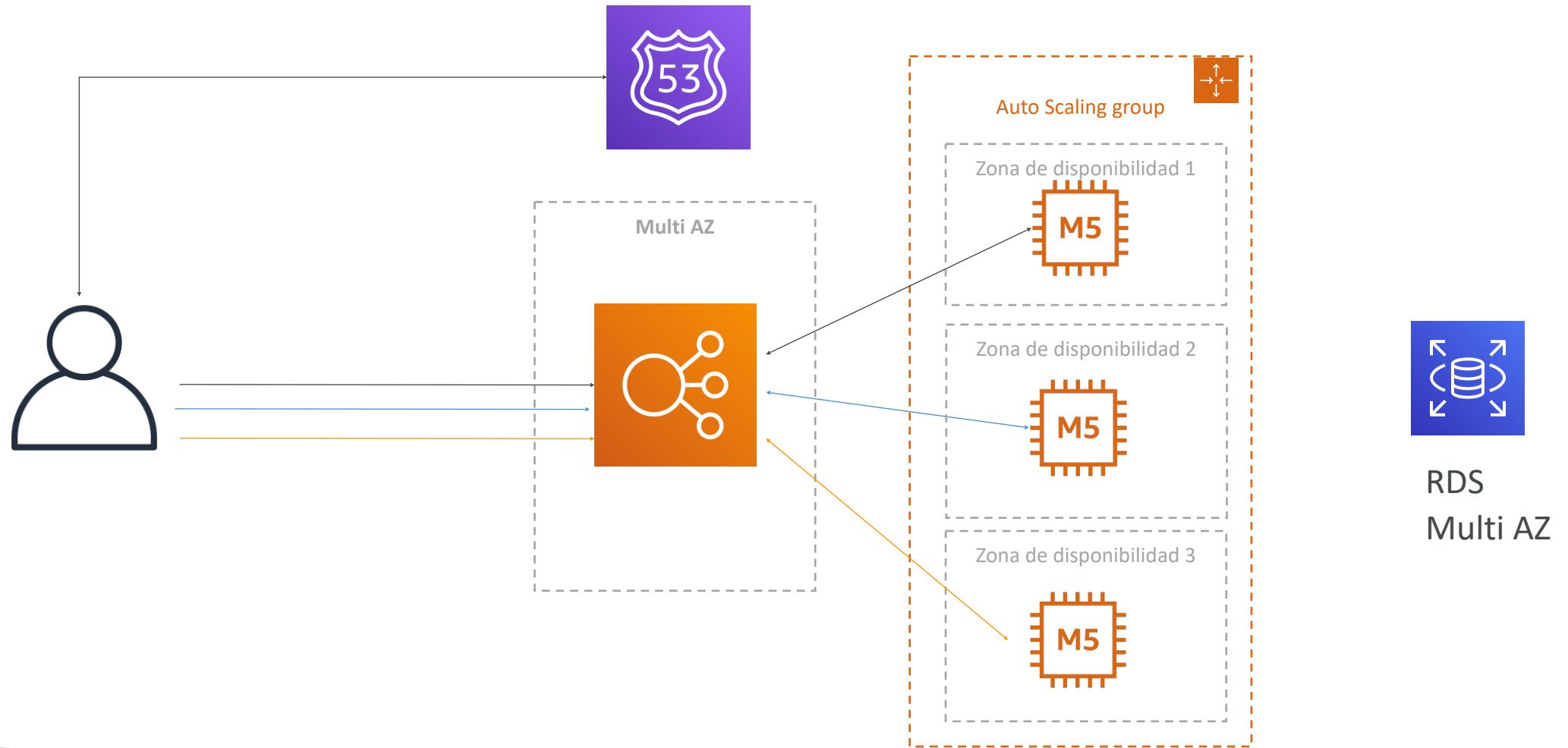
- Sesiones persistentes del ELB
- Clientes web para almacenar cookies y hacer que nuestra aplicación web sea sin estado
- ElastiCache
 - Para almacenar sesiones (alternativa: DynamoDB)
 - Para almacenar en caché los datos de RDS
 - Multi AZ
- RDS
 - Para almacenar los datos de los usuarios
 - Rélicas de lectura para escalar las lecturas
 - Multi AZ para la recuperación de desastres
- Seguridad estricta con grupos de seguridad que se refieren entre sí

Aplicación web con estado: MyWordPress.com

- Estamos intentando crear un sitio web de WordPress totalmente escalable
- Queremos que ese sitio web acceda y muestre correctamente las subidas de imágenes
- Los datos de nuestros usuarios y el contenido del blog deben almacenarse en una base de datos MySQL.
- ¡Veamos cómo podemos conseguirlo!

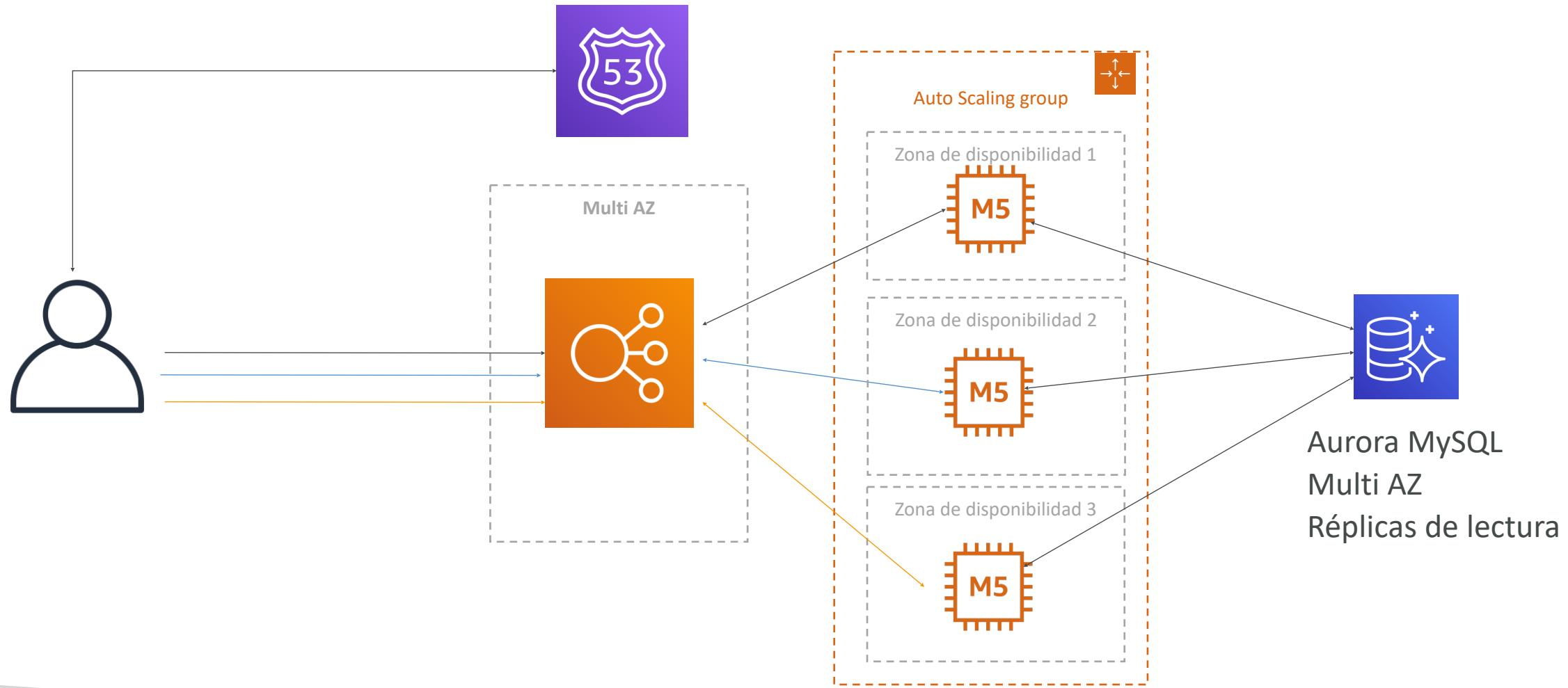
Aplicación web con estado: MyWordPress.com

Capa RDS



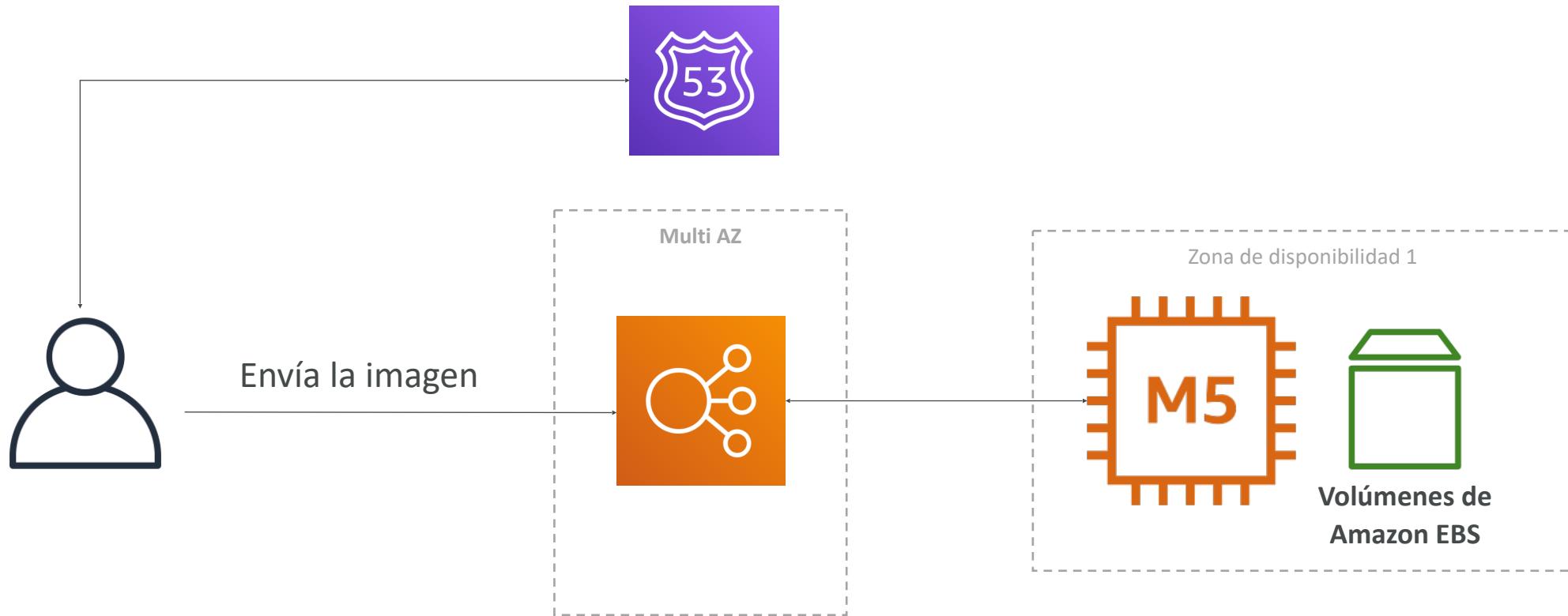
Aplicación web con estado: MyWordPress.com

Escalando con Aurora: Réplicas multi AZ y de lectura



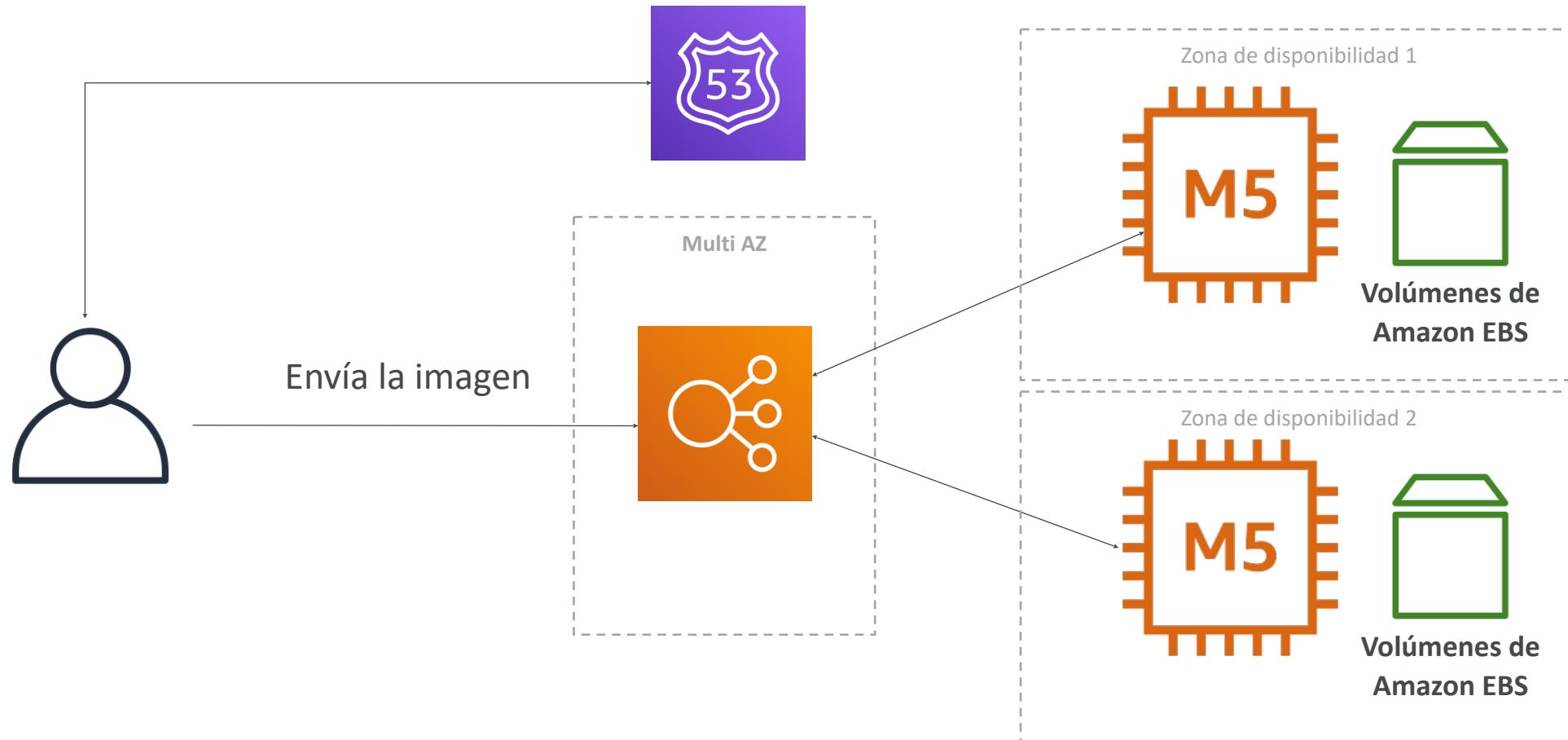
Aplicación web con estado: MyWordPress.com

Almacenamiento de imágenes con EBS



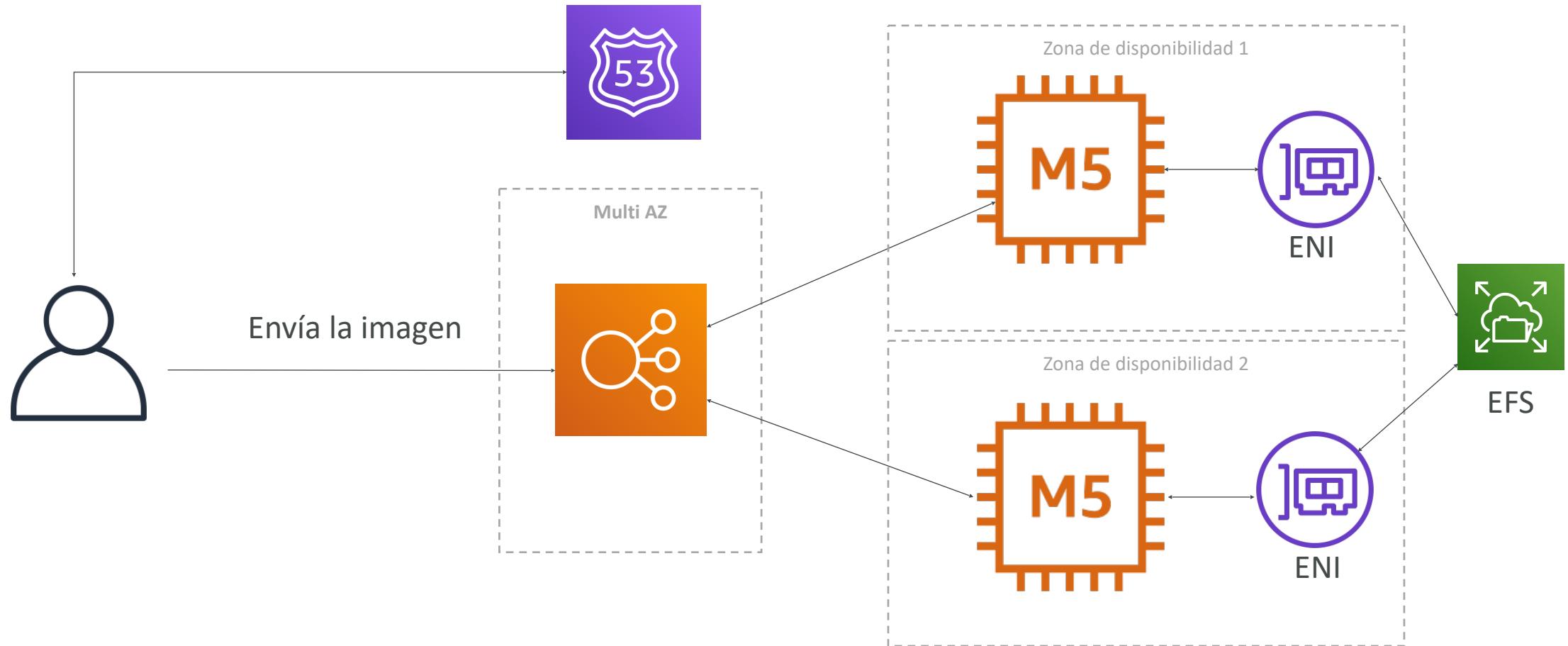
Aplicación web con estado: MyWordPress.com

Almacenamiento de imágenes con EBS



Aplicación web con estado: MyWordPress.com

Almacenamiento de imágenes con EFS



En esta clase hemos hablado de...

- Base de datos Aurora para tener fácilmente Multi-AZ y Rélicas de Lectura
- Almacenamiento de datos en EBS (aplicación de instancia única)
- Vs Almacenamiento de datos en EFS (aplicación distribuida)

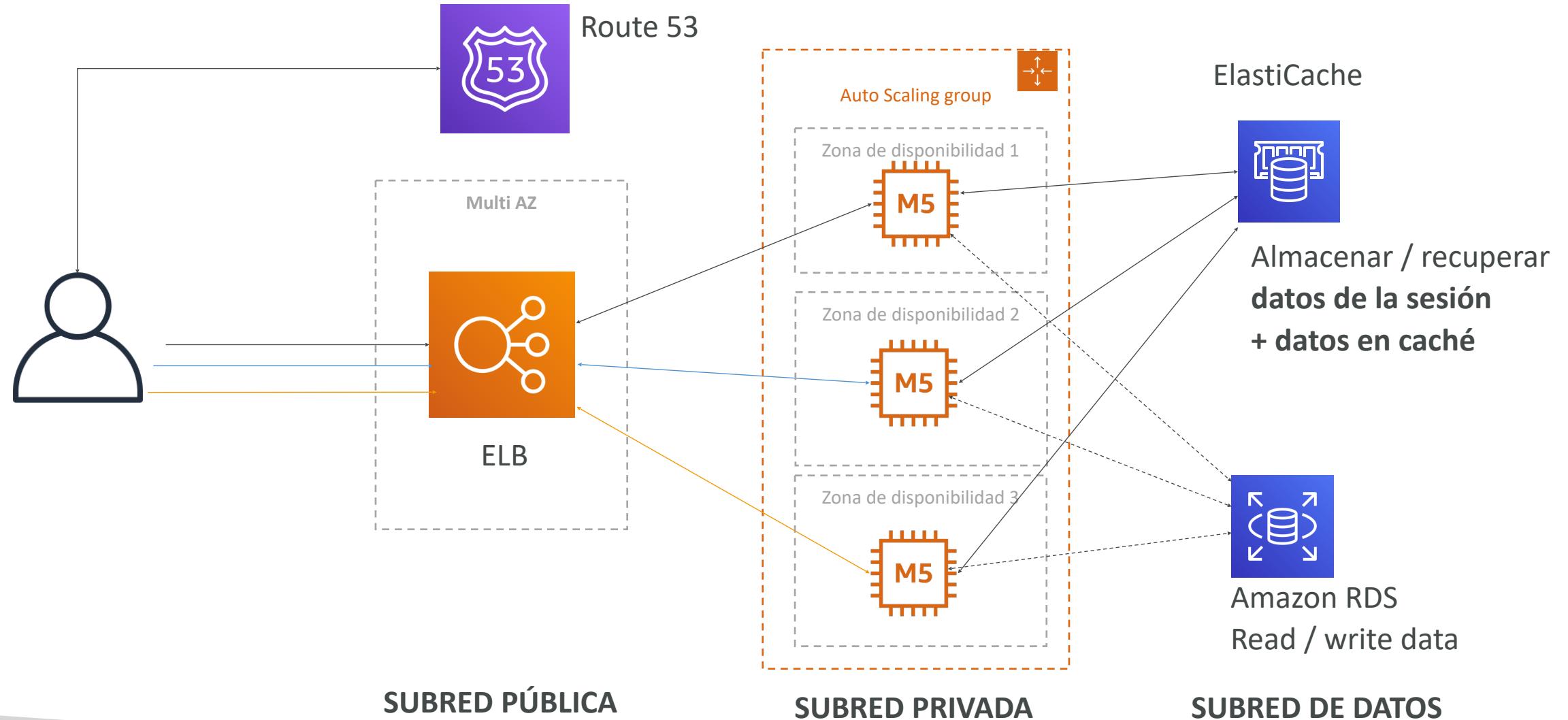
Instanciar rápidamente las aplicaciones

- Cuando se lanza un stack completo (EC2, EBS, RDS), puede llevar tiempo
 - Instalar las aplicaciones
 - Insertar los datos iniciales (o de recuperación)
 - Configurar todo
 - Lanzar la aplicación
- ¡Podemos aprovechar el Cloud para acelerar eso!

Instanciar rápidamente las aplicaciones

- Instancias EC2:
 - **Utiliza una Golden AMI:** Instala tus aplicaciones, dependencias del SO, etc. de antemano y lanza tu instancia EC2 desde la AMI dorada
 - **Arranca usando Datos de Usuario:** Para una configuración dinámica, utiliza scripts de Datos de Usuario
 - **Híbrido:** mezcla Golden AMI y Datos de usuario (Elastic Beanstalk)
- Bases de datos RDS:
 - Restaura desde un Snapshot: ¡la base de datos tendrá esquemas y datos listos!
- Volúmenes EBS:
 - Restaura a partir de una Snapshot: ¡el disco ya estará formateado y tendrá datos!

Arquitectura típica: Web App de 3 niveles



Problemas de los desarrolladores en AWS

- Gestión de la infraestructura
 - Desplegar el código
 - Configurar todas las bases de datos, balanceadores de carga, etc.
 - Problemas de escalado
-
- La mayoría de las aplicaciones web tienen la misma arquitectura (ALB + ASG)
 - Lo único que quieren los desarrolladores es que su código se ejecute
 - Posiblemente, de forma consistente en diferentes aplicaciones y entornos

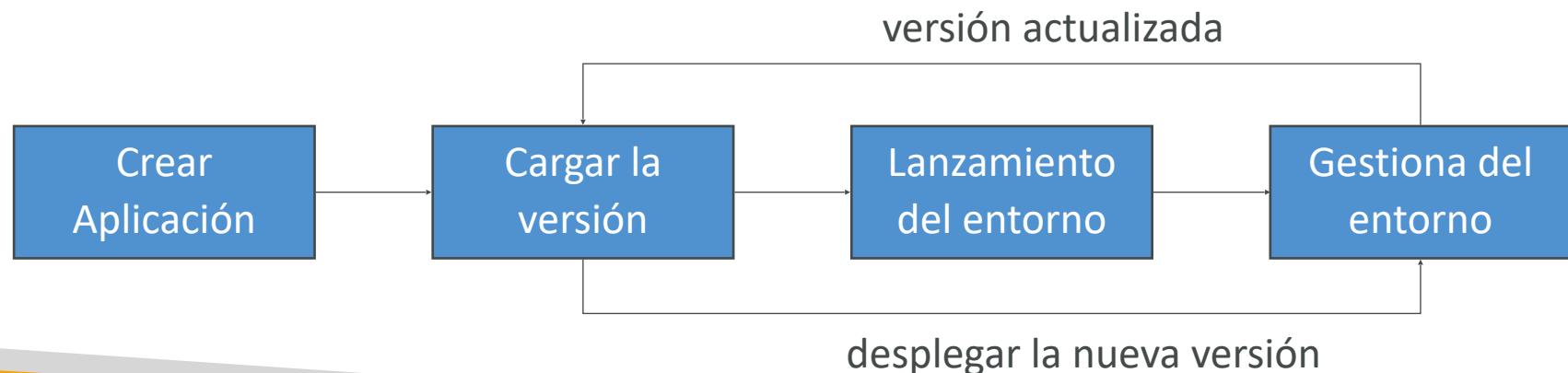
Elastic Beanstalk - Visión general



- Elastic Beanstalk es una visión centrada en el desarrollador de la implementación de una aplicación en AWS
- Utiliza todos los componentes que hemos visto antes: EC2, ASG, ELB, RDS, ...
- Servicio gestionado
 - Gestiona automáticamente el aprovisionamiento de capacidad, el equilibrio de carga, el escalado, la supervisión del estado de la aplicación, la configuración de las instancias, ...
 - Sólo el código de la aplicación es responsabilidad del desarrollador
 - Seguimos teniendo el control total de la configuración
 - Beanstalk es gratis pero pagas por las instancias subyacentes

Elastic Beanstalk - Componentes

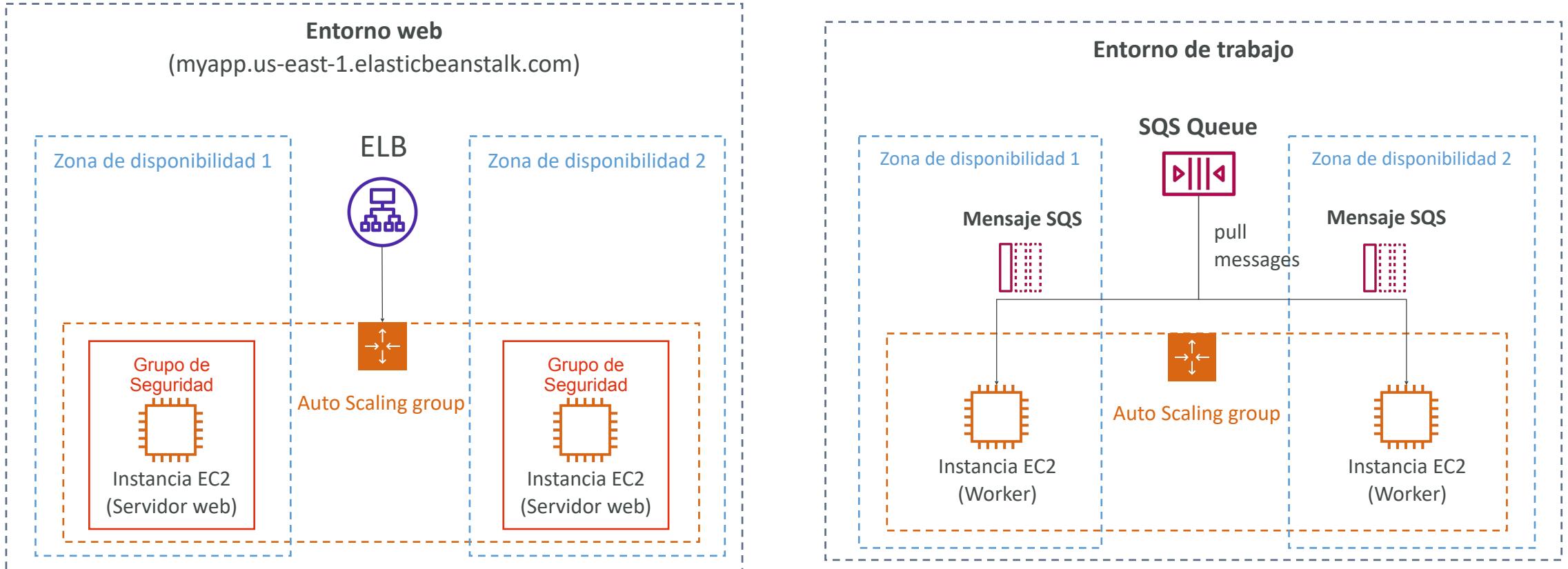
- **Aplicación:** colección de componentes de Elastic Beanstalk (entornos, versiones, configuraciones, ...)
- **Versión de la aplicación:** una iteración del código de tu aplicación
- **Entorno**
 - Colección de recursos de AWS que ejecutan una versión de la aplicación (sólo una versión de la aplicación a la vez)
 - **Niveles:** Nivel de entorno del servidor web y nivel de entorno del trabajador
 - Puedes crear varios entornos (dev, test, prod, ...)



Elastic Beanstalk - Plataformas soportadas

- Go
- Java SE
- Java con Tomcat
- .NET Core en Linux
- .NET en Windows Server
- Node.js
- PHP
- Python
- Ruby
- Packer Builder
- Contenedor único Docker
- Docker multicontenedor
- Docker Preconfigurado
- Si no tiene soporte, puedes escribir tu plataforma personalizada (avanzada)

Nivel de entorno web vs. entorno de trabajo



- Escala basada en el número de mensajes SQS
- Puede enviar mensajes a la cola SQS desde otro nivel de servidor web

Amazon S3

Introducción de la sección



- Amazon S3 es uno de los principales bloques de construcción de AWS
- Se anuncia como almacenamiento de "escala infinita".
- Muchos sitios web utilizan Amazon S3 como columna vertebral
- Muchos servicios de AWS utilizan Amazon S3 como una integración también
- Tendremos una aproximación paso a paso a S3

S3 Casos de uso

- Copia de seguridad y almacenamiento
- Recuperación de desastres
- Almacenamiento en el Cloud híbrido
- Alojamiento de aplicaciones
- Alojamiento de medios
- Data Lakes y análisis de big data
- Entrega de software
- Sitio web estático



Nasdaq almacena 7 años de datos en S3 Glacier



Sysco analiza sus datos y obtiene información comercial

Visión general de Amazon S3 - Buckets

- Amazon S3 permite almacenar objetos (archivos) en "buckets" (directorios)
- Los buckets deben tener un **nombre único a nivel global (en todas las regiones, todas las cuentas)**
- Los buckets se definen a nivel de región
- S3 parece un servicio global pero los buckets se crean en una región
 - Convención de nombres
 - Sin mayúsculas
 - Sin guiones bajos
 - 3-63 caracteres de longitud
 - No es una IP
 - Debe comenzar con una letra minúscula o un número



Amazon S3 - Objetos

- Los objetos (archivos) tienen una clave
- La **clave** es la ruta **COMPLETA**:
 - s3://mi-bucket/mi-archivo.txt
 - s3://mi-bucket/mi_carpetal/otra_carpetal/mi_archivo.txt
- La clave se compone de **prefijo** + **nombre del objeto**
 - s3://mi-bucket/mi_carpetal/otra_carpetal/mi_fichero.txt
- No existe el concepto de "directorios" dentro de los buckets (aunque la interfaz de usuario te hará pensar lo contrario)
- Sólo claves con nombres muy largos que contienen barras inclinadas ("/')

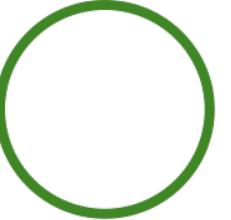


Object



Bucket S3
con objetos

Amazon S3 – Objetos (cont.)

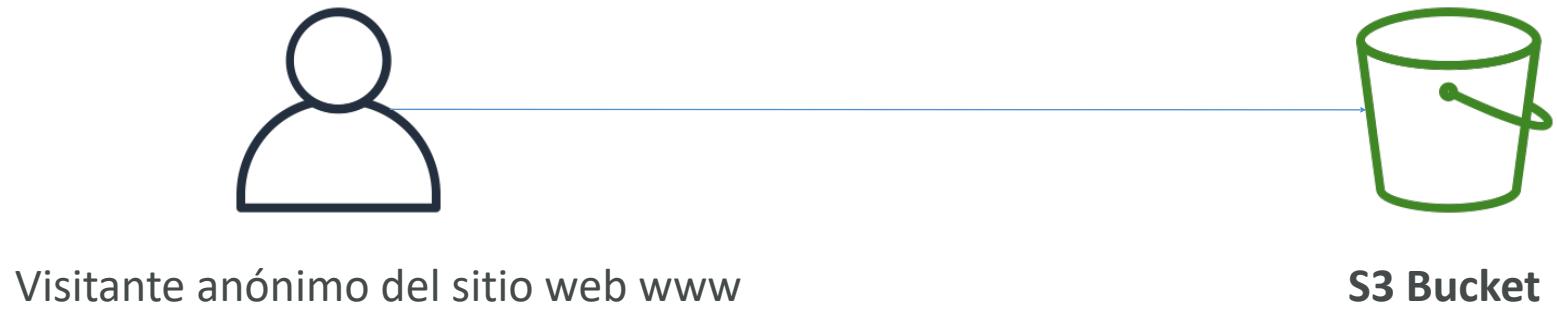


- Los valores de los objetos son el contenido del cuerpo:
 - El tamaño máximo del objeto es de 5TB (5000GB)
 - Si se suben más de 5GB, se debe usar "carga de varias partes"
- Metadatos (lista de pares clave/valor de texto - metadatos del sistema o del usuario)
- Etiquetas (par clave/valor Unicode - hasta 10) - útil para la seguridad/ciclo de vida
- ID de la versión (si el versionado está activado)

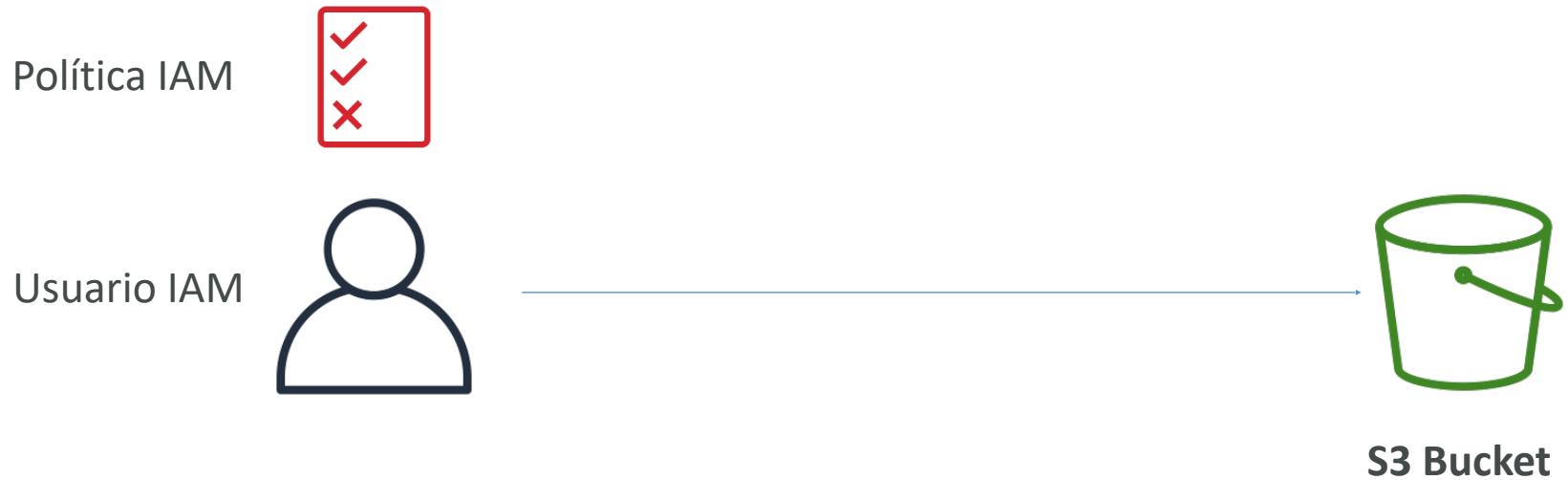
Amazon S3 – Seguridad

- **Basada en el usuario**
 - Políticas de IAM: qué llamadas a la API deben permitirse para un usuario específico desde la consola de IAM
- **Basadas en recursos**
 - Políticas de bucket - reglas a nivel de bucket desde la consola de S3 - permite cuentas cruzadas
 - Lista de control de acceso a objetos (ACL)
 - Lista de control de acceso a buckets (ACL)
- **Nota:** un usuario de IAM puede acceder a un objeto de S3 si
 - los permisos IAM del usuario lo permiten Y la política de recursos lo PERMITE
 - Y no hay una DENEGACIÓN explícita
- **Encriptación:** cifra los objetos en Amazon S3 utilizando claves de encriptación

Ejemplo: Acceso público Política de uso de bucket

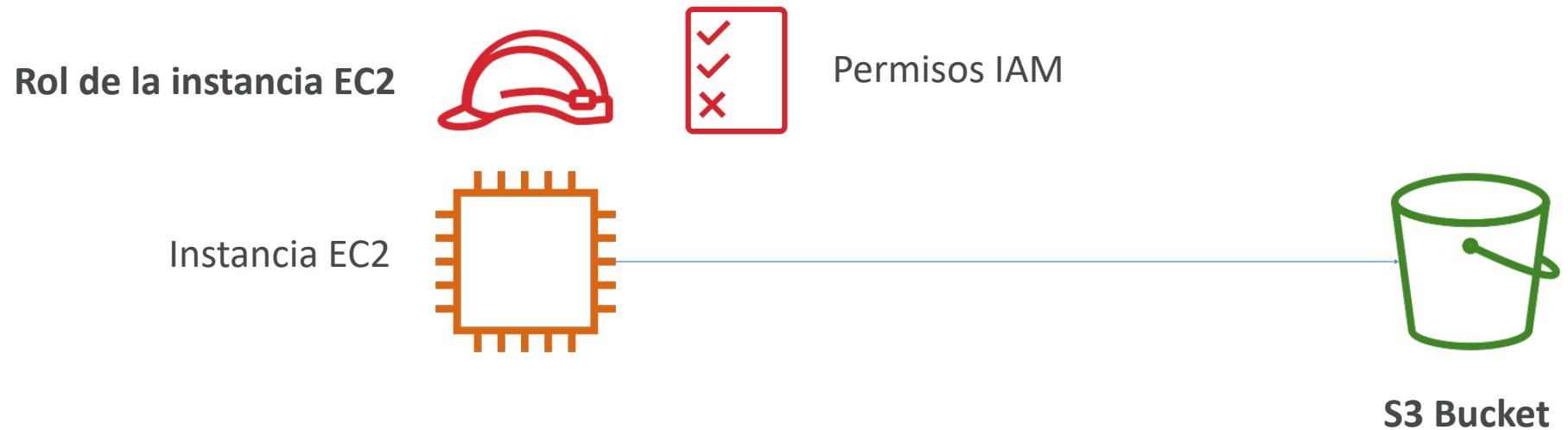


Ejemplo: Acceso del usuario al S3 Permisos IAM



Ejemplo: Acceso a la instancia EC2

Utilizar roles IAM



Avanzado: Acceso entre cuentas

Usar política de bucket

Usuario IAM
Otra cuenta de AWS



Política de bucket S3
Permite las cuentas cruzadas



S3 Bucket

Políticas de bucket S3

- **Políticas basadas en JSON:**

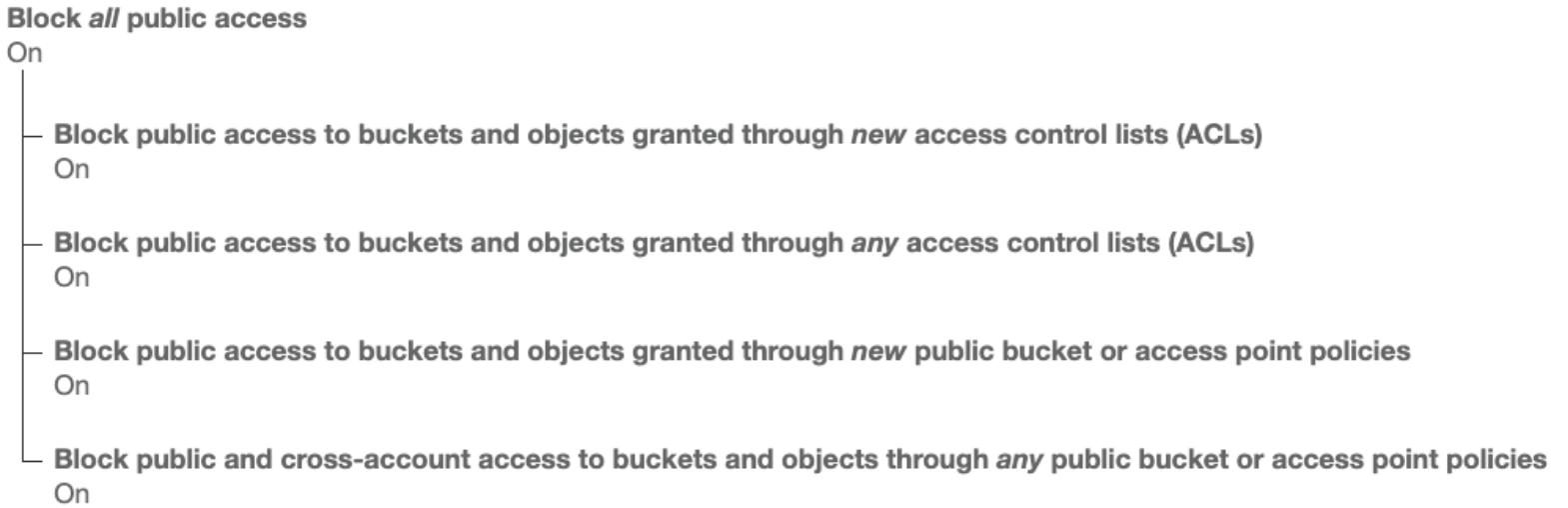
- Resources: buckets y objetos
- Effect: Permitir / Denegar
- Actions: Conjunto de API para permitir o denegar
- Principal: La cuenta o usuario al que aplicar la política

- **Usar el bucket S3 para aplicar la política:**

- Conceder acceso público al bucket
- Forzar que los objetos se cifren al subirlos
- Conceder acceso a otra cuenta (cuenta cruzada)

```
{  
  "Version": "2012-10-17",  
  "Statement": [  
    {  
      "Sid": "PublicRead",  
      "Effect": "Allow",  
      "Principal": "*",  
      "Action": [  
        "s3:GetObject"  
      ],  
      "Resource": [  
        "arn:aws:s3:::examplebucket/*"  
      ]  
    }  
  ]  
}
```

Configuración del bucket para bloquear el acceso público



- Estos ajustes se crearon para evitar la filtración de datos de la empresa
- Si sabes que tu bucket no debe ser nunca público, déjalo activado
- Pueden establecerse a nivel de cuenta

Amazon S3

Alojamiento de sitios web estáticos

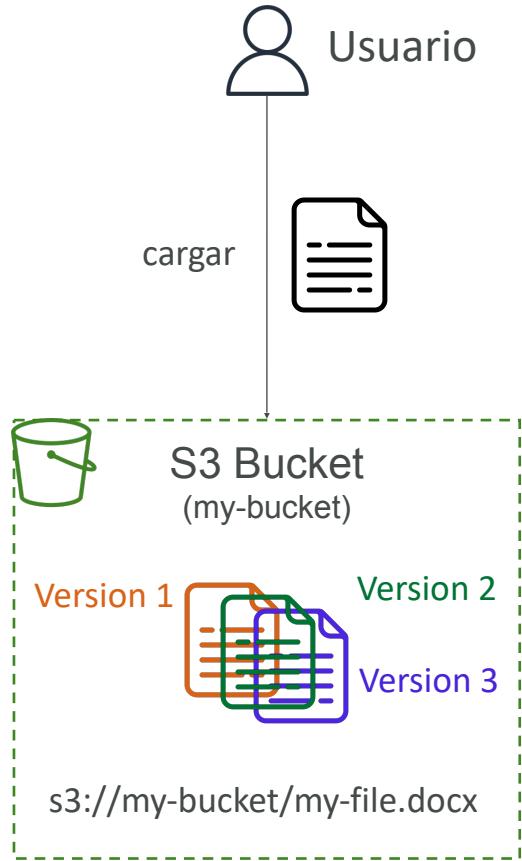
- S3 puede alojar sitios web estáticos y tenerlos accesibles en Internet
- La URL del sitio web será (según la región)
 - `http://bucket-name.s3-website-aws-region.amazonaws.com`
 -
 - `http://bucket-name.s3-website.aws-region.amazonaws.com`
- Si obtienes un error 403 Prohibido, ¡asegúrate de que la política del bucket permite las lecturas públicas!

`http://demo-bucket.s3-website-us-west-2.amazonaws.com`
`http://demo-bucket.s3-website.us-west-2.amazonaws.com`



Amazon S3 - Versionado

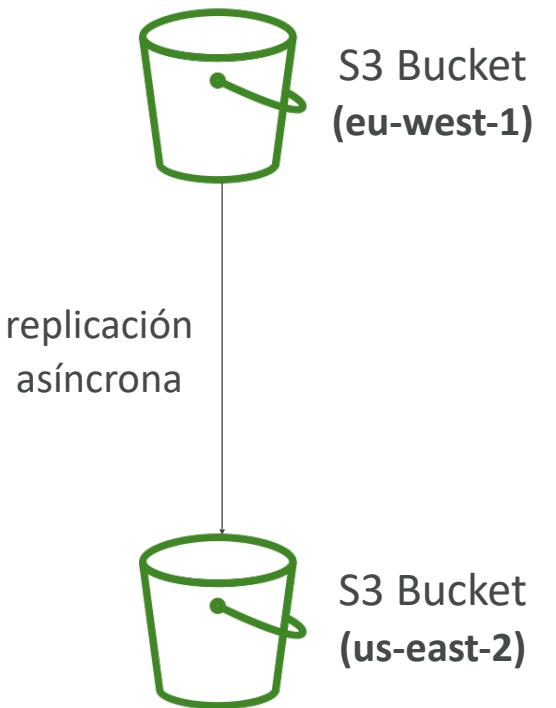
- Puedes versionar tus archivos en Amazon S3
- Se activa a nivel de bucket
- La misma clave de sobrescritura cambiará la "versión": 1, 2, 3...
- Es una buena práctica versionar tus buckets
 - Protege contra los borrados involuntarios (posibilidad de restaurar una versión)
 - Facilidad para volver a la versión anterior
- Notas:
 - Cualquier archivo que no esté versionado antes de activar el versionado tendrá la versión "nula"
 - Suspender el versionado no elimina las versiones anteriores



Amazon S3 - Replicación (CRR y SRR)



- Debe activar el control de versiones en los buckets de origen y destino
- Replicación entre regiones (CRR)
- Replicación en la misma región (SRR)
- Los buckets pueden estar en diferentes cuentas de AWS
- La copia es asíncrona
- Debes dar los permisos IAM adecuados a S3
- Casos de uso:
 - CRR - normativa, acceso de menor latencia, replicación entre cuentas
 - SRR - agregación de logs, replicación en vivo entre cuentas de producción y de prueba



Amazon S3 - Replicación (Notas)

- Después de activar la Replicación, sólo se replican los objetos nuevos
- Opcionalmente, puedes replicar los objetos existentes utilizando la **Replicación por lotes de S3**
 - Replica los objetos existentes y los objetos que fallaron en la replicación
- Para las operaciones de borrado
 - **Puede replicar los marcadores de borrado** del origen al destino (configuración opcional)
 - Los borrados con un ID de versión no se replican (para evitar borrados maliciosos)
- **No hay "encadenamiento" de la replicación**
 - Si el bucket 1 tiene replicación en el bucket 2, que tiene replicación en el bucket 3
 - Entonces los objetos creados en el bucket 1 no se replican en el bucket 3

Clases de almacenamiento S3

- Amazon S3 Standard - General Purpose
- Amazon S3 Standard-Infrequent Access (IA)
- Amazon S3 One Zone-Infrequent Access (IA)
- Amazon S3 Glacier Instant Retrieval
- Amazon S3 Glacier Flexible Retrieval
- Amazon S3 Glacier Deep Archive
- Amazon S3 Intelligent Tiering
- Se puede pasar de una clase a otra manualmente o utilizando las configuraciones del ciclo de vida de S3

S3 Durabilidad y disponibilidad

- **Durabilidad:**

- Alta durabilidad (99,99999999%, 11 9's) de los objetos a través de múltiples AZ
- Si almacenas 10.000.000 de objetos con Amazon S3, puedes esperar una media de pérdida de un solo objeto una vez cada 10.000 años
- Lo mismo para todas las clases de almacenamiento

- **Disponibilidad:**

- Mide la disponibilidad de un servicio
- Varía en función de la clase de almacenamiento
- Ejemplo: El Estándar S3 tiene una disponibilidad del 99,99% = no está disponible 53 minutos al año

Estándar S3 - Propósito general



- Disponibilidad del 99,99%.
 - Se utiliza para datos de acceso frecuente
 - Baja latencia y alto rendimiento
 - Soporta 2 fallos concurrentes de la instalación
-
- Casos de uso: Análisis de Big Data, aplicaciones móviles y de juegos, distribución de contenidos...

Clases de almacenamiento S3 - Acceso infrecuente

- Para datos a los que se accede con menos frecuencia, pero que requieren un acceso rápido cuando se necesitan
- Coste inferior al de S3 Estándar
- **Amazon S3 Standard-Infrequent Access (S3 Standard-IA)**
 - Disponibilidad del 99,9%.
 - Casos de uso: Recuperación de desastres, copias de seguridad
- **Amazon S3 One Zone-Infrequent Access (S3 One Zone-IA)**
 - Alta durabilidad (99,999999999%) en una sola AZ; los datos se pierden cuando se destruye la AZ
 - Disponibilidad del 99,5%.
 - Casos de uso: Almacenamiento de copias de seguridad secundarias de datos locales, o de datos que puedes recrear



Clases de almacenamiento de Amazon S3 Glacier

- Almacenamiento de objetos de bajo coste pensado para archivar / hacer copias de seguridad
- Precio: precio del almacenamiento + coste de recuperación del objeto
- **Amazon S3 Glacier Instant Retrieval**
 - Recuperación en milisegundos, ideal para datos a los que se accede una vez al trimestre
 - Duración mínima de almacenamiento de 90 días
- **Amazon S3 Glacier Flexible Retrieval** (antes Amazon S3 Glacier)
 - Acelerada (de 1 a 5 minutos), Estándar (de 3 a 5 horas), Masiva (de 5 a 12 horas) - gratis
 - Duración mínima de almacenamiento de 90 días
- **Amazon S3 Glacier Deep Archive** - para almacenamiento a largo plazo:
 - Estándar (12 horas), Masiva (48 horas)
 - Duración mínima de almacenamiento de 180 días





S3 Intelligent-Tiering

- Pequeña cuota mensual de monitorización y jerarquización automática
 - Mueve los objetos automáticamente entre los niveles de acceso en función del uso
 - No hay cargos por recuperación en S3 Intelligent-Tiering
-
- *Frequent Access tier (automático)*: nivel por defecto
 - *Infrequent Access tier (automático)*: objetos no accedidos durante 30 días
 - *Archive Instant Access tier (automático)*: objetos no accedidos durante 90 días
 - *Archive Access tier (opcional)*: configurable de 90 a más de 700 días
 - *Deep Archive Access tier (opcional)*: configurable de 180 días a 700+ días

Comparación de las clases de almacenamiento de S3

	Standard	Intelligent-Tiering	Standard-IA	One Zone-IA	Glacier Instant Retrieval	Glacier Flexible Retrieval	Glacier Deep Archive
Durabilidad	99.999999999% == (11 9's)						
Disponibilidad	99.99%	99.9%	99.9%	99.5%	99.9%	99.99%	99.99%
Acuerdo de nivel de servicio de disponibilidad	99.9%	99%	99%	99%	99%	99.9%	99.9%
Zonas de disponibilidad	>= 3	>= 3	>= 3	1	>= 3	>= 3	>= 3
Min. Duración del almacenamiento	Ninguno	Ninguno	30 Días	30 Días	90 Días	90 Días	180 Días
Min. Tamaño del objeto facturable	Ninguno	Ninguno	128 KB	128 KB	128 KB	40 KB	40 KB
Tasa de recuperación	Ninguno	Ninguno	Por GB recuperado	Por GB recuperado	Por GB recuperado	Por GB recuperado	Por GB recuperado

<https://aws.amazon.com/s3/storage-classes/>

Clases de almacenamiento S3 - Comparación de precios

Ejemplo: us-east-1

	Standard	Intelligent-Tiering	Standard-IA	One Zone-IA	Glacier Instant Retrieval	Glacier Flexible Retrieval	Glacier Deep Archive
Coste de almacenamiento (por GB al mes)	0.023\$	0.0025\$ - 0.023\$	%0.0125	0.01\$	0.004\$	0.0036\$	0.00099\$
Coste de recuperación (por cada 1000 solicitudes)	GET: 0.0004\$ POST: 0.005\$	GET: 0.0004\$ POST: 0.005\$	GET: 0.001\$ POST: 0.01\$	GET: \$0.001\$ POST: 0.01\$	GET: 0.01\$ POST: 0.02\$	GET: 0.0004\$ POST: 0.03\$ Expedited: 10\$ Standard: 0.05\$ Bulk: gratis	GET: 0.0004\$ POST: 0.05\$ Standard: 0.10\$ Bulk: 0.025\$
Tiempo de recuperación	Instantáneo					Expedited (1 – 5 mins) Standard (3 – 5 hours) Bulk (5 – 12 hours)	Standard (12 horas) Bulk (48 horas)
Coste de la monitorización (1000 objetos)		0.0025\$					

<https://aws.amazon.com/s3/pricing/>

Desarrollo en AWS

Metadatos de la instancia de AWS EC2

- Los metadatos de las instancias de AWS EC2 son potentes pero una de las características menos conocidas por los desarrolladores
- Permite que las instancias de AWS EC2 "aprendan sobre sí mismas" **sin necesidad de utilizar un rol de IAM para ello.**
- La URL es <http://169.254.169.254/latest/meta-data>
- Puedes recuperar el nombre del rol IAM de los metadatos, pero NO puedes recuperar la política IAM.
- Metadatos = Información sobre la instancia EC2

Visión general del SDK de AWS

- ¿Y si quieres realizar acciones en AWS directamente desde el código de tus aplicaciones? (sin usar la CLI).
- ¡Puedes utilizar un SDK (kit de desarrollo de software) !
- Los SDK oficiales son...
 - Java
 - .NET
 - Node.js
 - PHP
 - Python (llamado boto3 / botocore)
 - Go
 - Ruby
 - C++

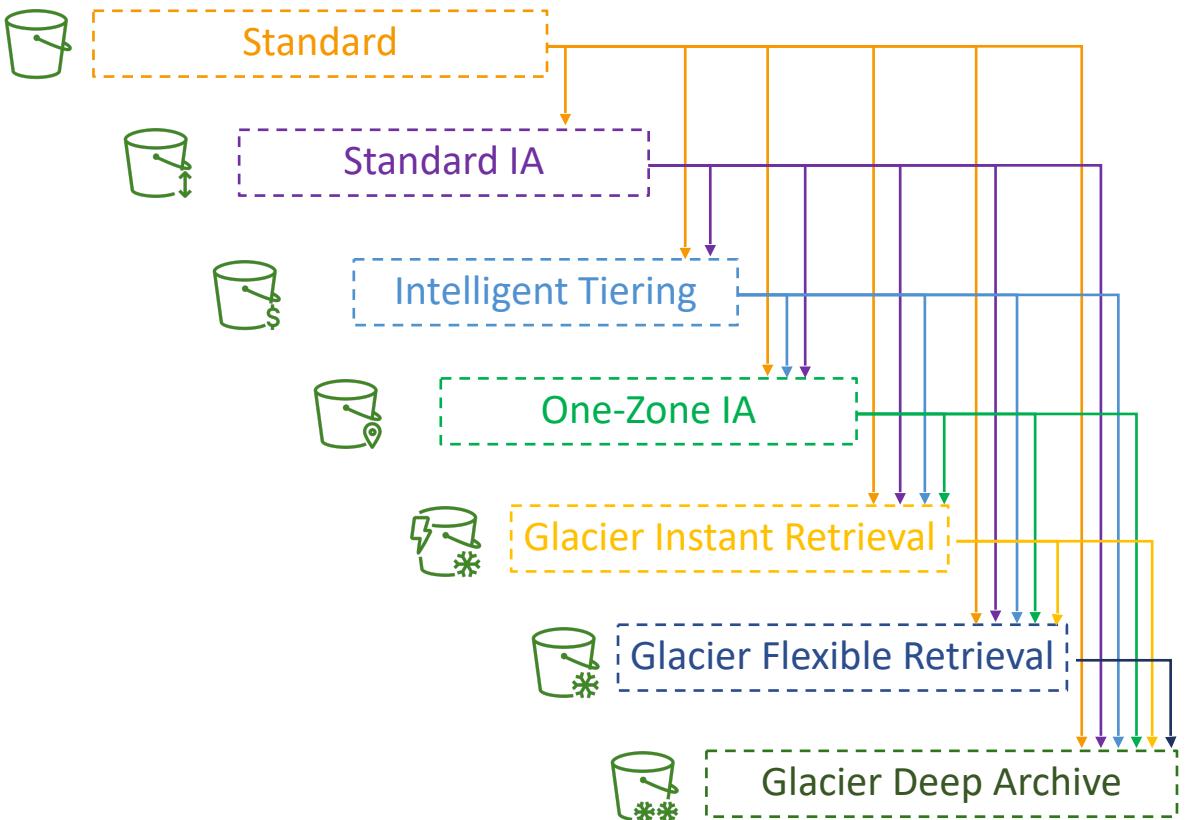
Visión general del SDK de AWS

- Tenemos que utilizar el SDK de AWS cuando programamos contra servicios de AWS como DynamoDB
- Dato curioso... la CLI de AWS utiliza el SDK de Python (boto3)
- El examen espera que sepas cuándo debes utilizar un SDK
- Practicaremos el SDK de AWS cuando lleguemos a las funciones Lambda
- Es bueno saberlo: si no especificas o configuras una región por defecto, se elegirá por defecto us-east-1

S3 avanzado

Amazon S3 - Movimiento entre clases de almacenamiento

- Puedes pasar los objetos entre las clases de almacenamiento
- Para los objetos a los que se accede con poca frecuencia, muévelos a IA Estándar
- Para los objetos de archivo a los que no necesitas acceder rápidamente, muévelos a Glacier o Glacier Deep Archive
- El movimiento de los objetos puede automatizarse mediante las Reglas del Ciclo de Vida



Amazon S3 - Reglas del ciclo de vida



- **Acciones de transición:** configura los objetos para que pasen a otra clase de almacenamiento
 - Mover los objetos a la clase IA Estándar 60 días después de su creación
 - Mover a Glacier para archivar después de 6 meses
- **Acciones de expiración** - configura los objetos para que caduquen (se eliminen) después de un tiempo
 - Los archivos de logs de acceso pueden configurarse para que se eliminen después de 365 días
 - **Se puede utilizar para eliminar versiones antiguas de archivos (si el versionado está activado)**
 - Se puede utilizar para eliminar subidas incompletas de Multipartes
- Se pueden crear reglas para un determinado prefijo (ejemplo: s3://mybucket/mp3/*)
- Se pueden crear reglas para determinados objetos Etiquetas (ejemplo: Departamento: Finanzas)

Amazon S3 - Reglas del ciclo de vida (escenario I)

- Tu aplicación en EC2 crea imágenes en miniatura después de subir las fotos del perfil a Amazon S3. Estas miniaturas pueden recrearse fácilmente, y sólo deben conservarse durante 60 días. Las imágenes de origen deben poder recuperarse inmediatamente durante estos 60 días, y después, el usuario puede esperar hasta 6 horas. ¿Cómo diseñarías esto?
- Las imágenes de origen de S3 pueden estar en **Estándar**, con una configuración del ciclo de vida para que pasen a **Glacier** después de 60 días
- Las miniaturas de S3 pueden estar en **IA de Zona Única**, con una configuración del ciclo de vida para que caduquen (se eliminen) después de 60 días

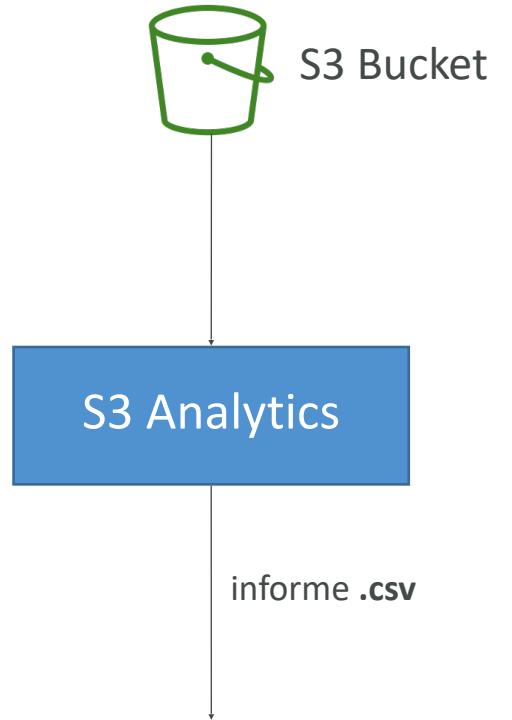
Amazon S3 - Reglas del ciclo de vida (escenario 2)

- Una norma de tu empresa establece que deberías poder recuperar tus objetos S3 eliminados inmediatamente durante 30 días, aunque esto puede ocurrir en raras ocasiones. Después de este tiempo, y durante un máximo de 365 días, los objetos eliminados deberían poder recuperarse en 48 horas.
- **Activa el Versionado de S3** para tener versiones de los objetos, de modo que los "objetos eliminados" queden de hecho ocultos por un "marcador de eliminación" y puedan ser recuperados
- Transitar las "versiones no actuales" del objeto a la **IA estándar**
- Transita después las "versiones no actuales" a **Glacier Deep Archive**

Amazon S3 Analytics

Análisis de la clase de almacenamiento

- Te ayuda a decidir cuándo pasar los objetos a la clase de almacenamiento adecuada
- Recomendaciones para **IA estándar** y **estándar**
 - NO sirve para IA de una Zona o Glacier
- El informe se actualiza diariamente
- De 24 a 48 horas para empezar a ver el análisis de los datos
- Buen primer paso para elaborar las Reglas del Ciclo de Vida (o mejorarlas)

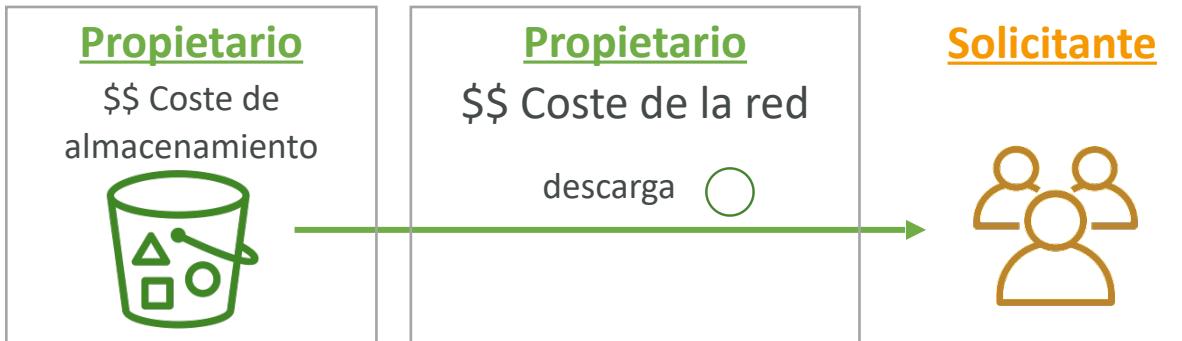


Fecha	StorageClass	ObjectAge
8/22/2022	STANDARD	000-014
8/25/2022	STANDARD	030-044
9/6/2022	STANDARD	120-149

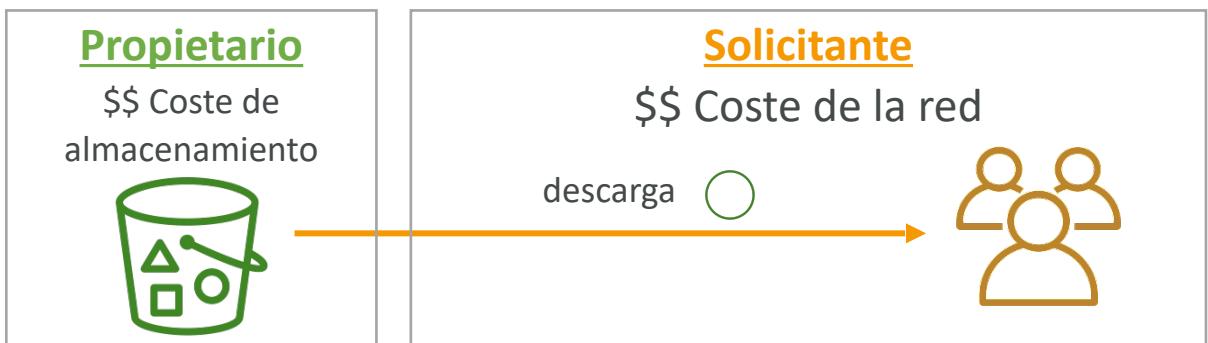
S3 - El solicitante paga

- En general, los propietarios de los buckets pagan todos los costes de almacenamiento y transferencia de datos de Amazon S3 asociados a su bucket
- **Con los buckets donde el solicitante paga**, el solicitante en lugar del propietario del bucket, paga el coste de la petición y la descarga de datos del bucket
- Es útil cuando quieras compartir grandes conjuntos de datos con otras cuentas
- El solicitante debe estar autenticado en AWS (no puede ser anónimo)

Bucket estándar

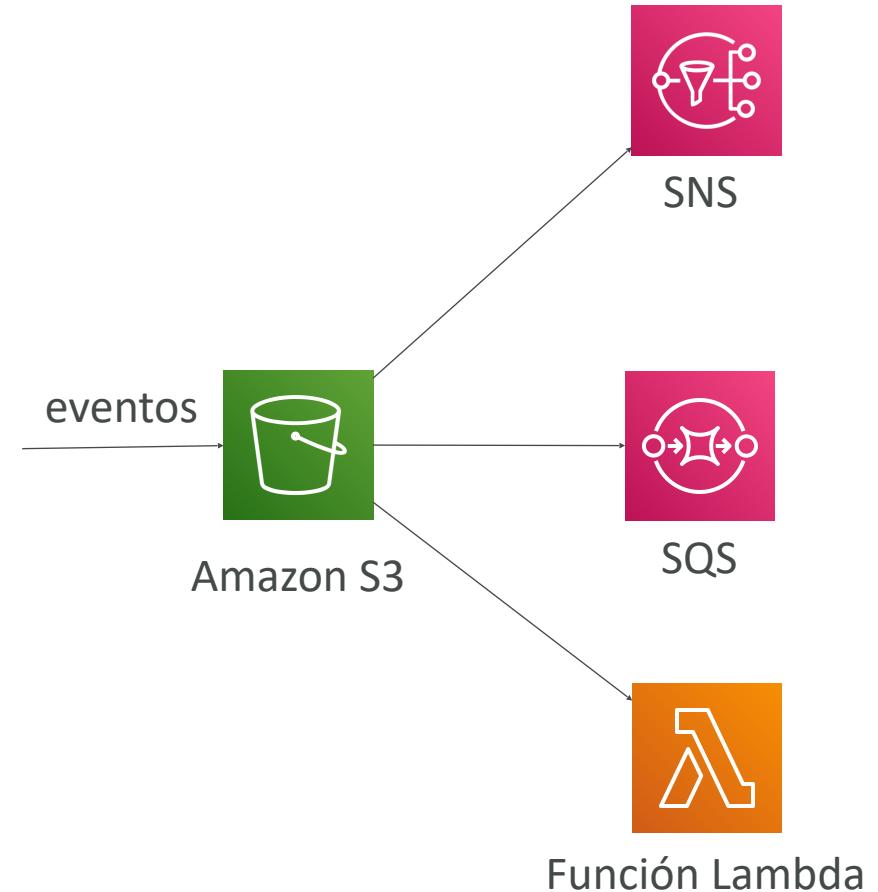


El solicitante paga el bucket



Notificaciones de eventos S3

- S3:ObjectCreated, S3:ObjectRemoved, S3:ObjectRestore, S3:Replication...
- Posibilidad de filtrar el nombre del objeto (*.jpg)
- Caso de uso: generar miniaturas de imágenes subidas a S3
- **Se pueden crear tantos "eventos S3" como se desee**
- Las notificaciones de eventos S3 suelen entregar los eventos en segundos, pero a veces pueden tardar un minuto o más



Notificaciones de eventos S3 con Amazon EventBridge



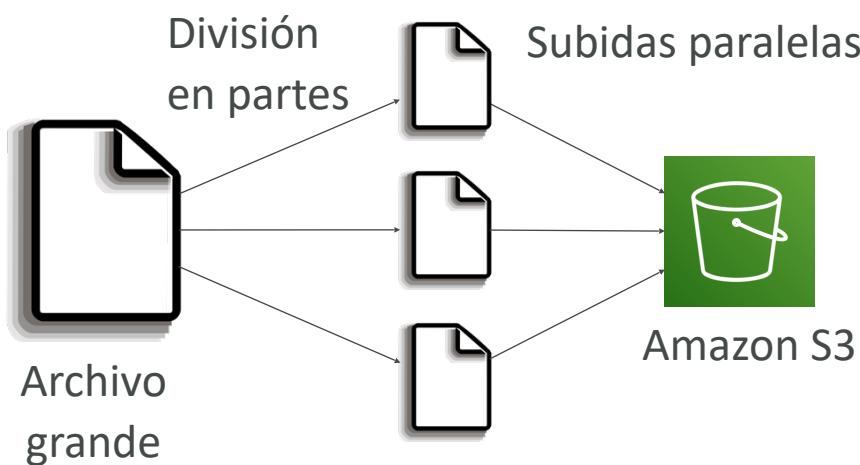
- **Opciones avanzadas** de filtrado con reglas JSON (metadatos, tamaño del objeto, nombre...)
- **Múltiples destinos** - Step Functions, Kinesis Streams / Firehose...
- **Capacidades de EventBridge** - Repetición de eventos, entrega fiable

S3 - Rendimiento básico

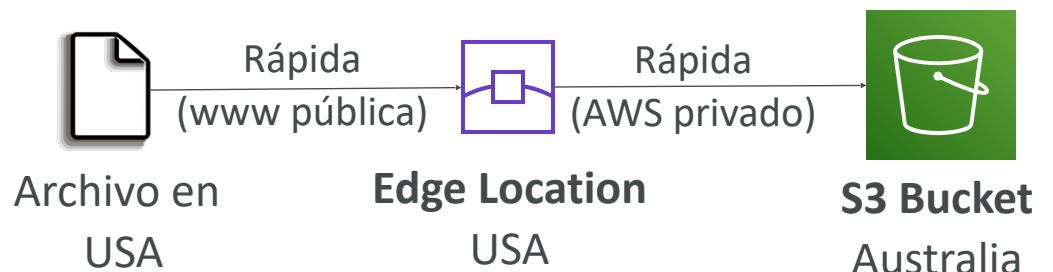
- Amazon S3 escala automáticamente a altas tasas de petición, latencia 100-200 ms
- Tu aplicación puede alcanzar al menos **3.500 peticiones PUT/COPY/POST/DELETE y 5.500 GET/HEAD por segundo por prefijo en un bucket.**
- No hay límites en el número de prefijos de un bucket.
- Ejemplo (ruta del objeto => prefijo):
 - bucket/carpeta1/sub1/fichero => /carpeta1/sub1/
 - bucket/carpeta1/sub2/fichero => /carpeta1/sub2/
 - bucket/1/fichero => /1/
 - bucket/2/fichero => /2/
- Si distribuyes las lecturas entre los cuatro prefijos de manera uniforme, puedes conseguir 22.000 peticiones por segundo para GET y HEAD

Rendimiento S3 (Performance)

- **Carga de varias partes (Multipartes):**
 - Recomendado para archivos > 100MB, obligatorio para archivos > 5GB
 - Puede ayudar a paralelizar las subidas (acelerar las transferencias)



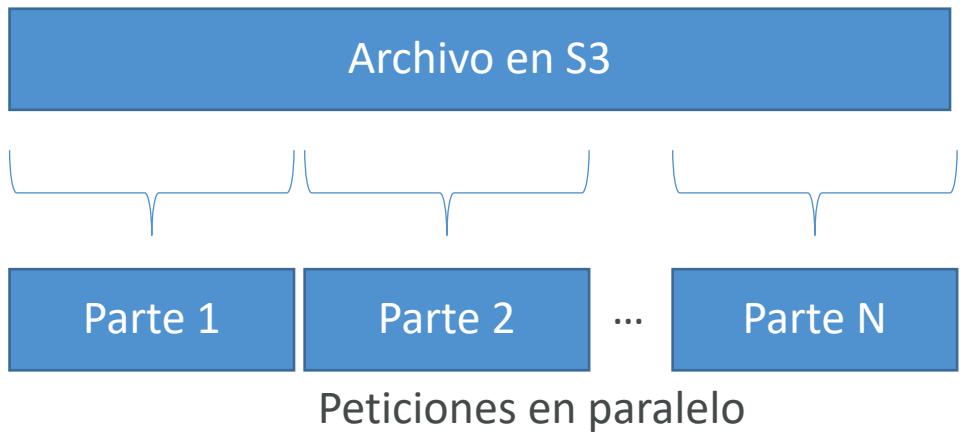
- **Aceleración de la transferencia en S3:**
 - Aumenta la velocidad de transferencia transfiriendo el archivo a un Edge Location de AWS que reenviará los datos al bucket de S3 en la región de destino
 - Compatible con la carga de varias partes



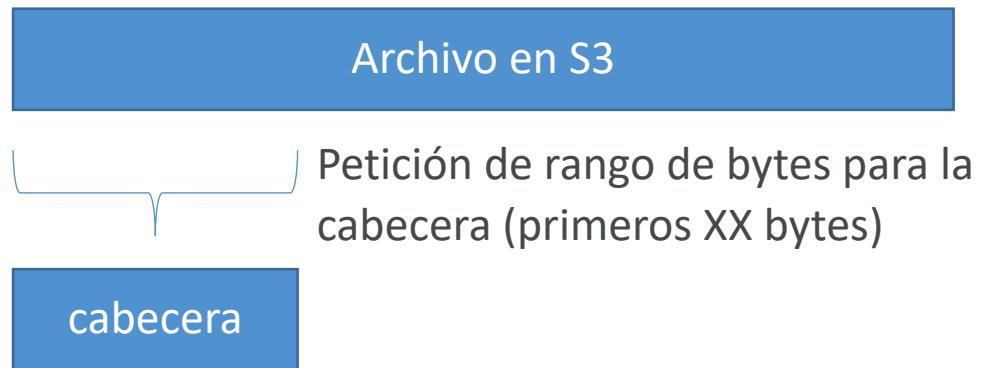
Rendimiento del S3

Recuperación del rango de bytes del S3

- Paraleliza los GETs solicitando rangos de bytes específicos
- Mejor resiliencia en caso de fallos
- Puede utilizarse para acelerar las descargas

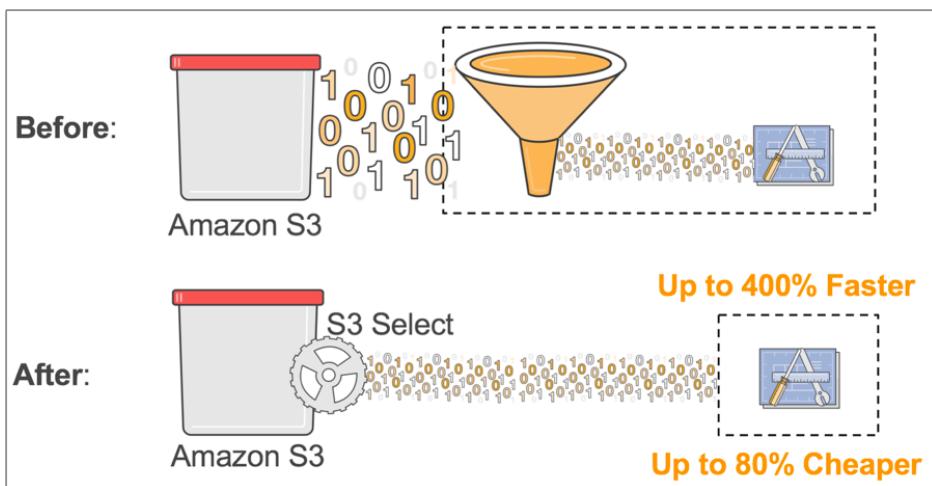


Puede utilizarse para recuperar sólo datos parciales

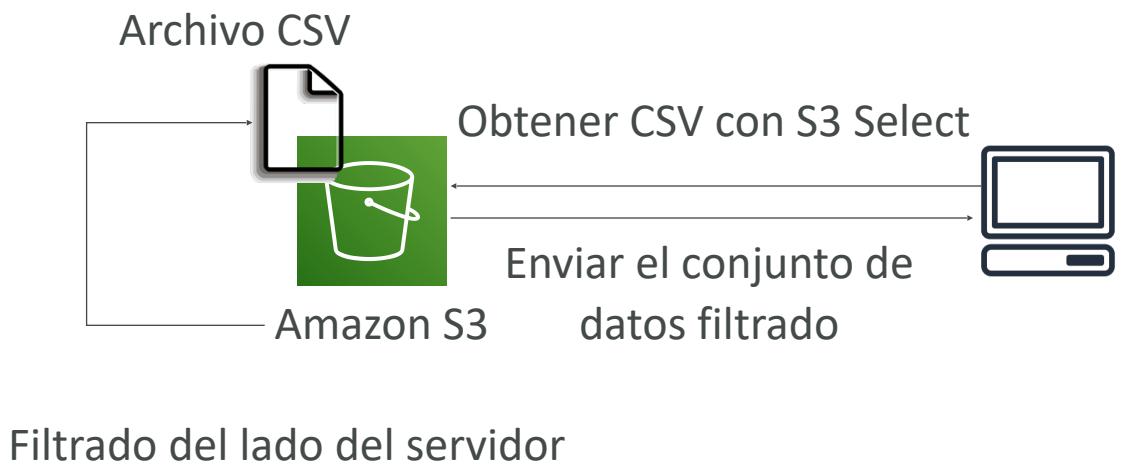


S3 Select y Glacier Select

- Recupera menos datos mediante SQL realizando un filtrado del lado del servidor
- Puedes filtrar por filas y columnas (simples sentencias SQL)
- Menos transferencia de red, menos coste de CPU en el lado del cliente

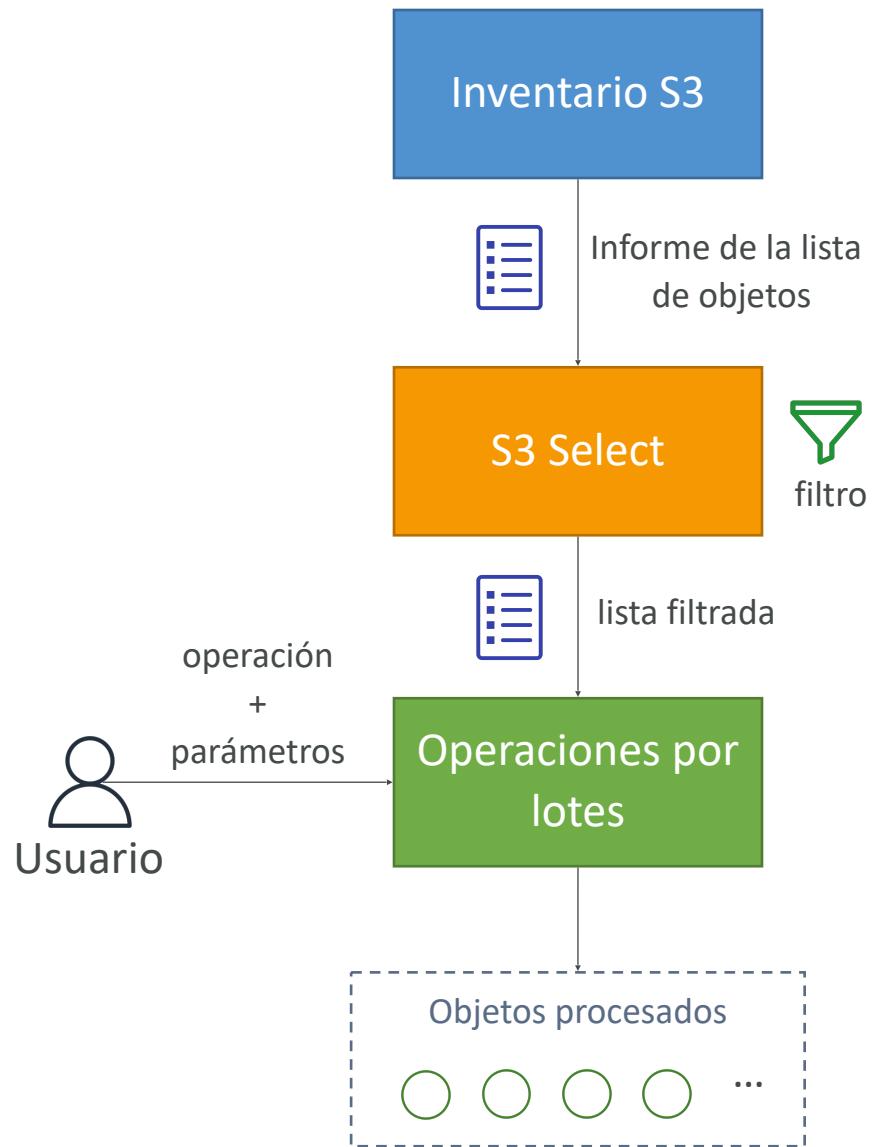


<https://aws.amazon.com/blogs/aws/s3-glacier-select/>



Operaciones por lotes de S3 (Batch Operations)

- Realiza operaciones masivas en objetos S3 existentes con una sola petición, por ejemplo:
 - Modificar los metadatos y las propiedades de los objetos
 - Copiar objetos entre buckets de S3
 - **Cifrar objetos no cifrados**
 - Modificar ACLs, etiquetas
 - Restaurar objetos desde S3 Glacier
 - Invocar una función Lambda para realizar una acción personalizada en cada objeto
- Un trabajo consiste en una lista de objetos, la acción a realizar y parámetros opcionales
- S3 Batch Operations gestiona los reintentos, sigue el progreso, envía notificaciones de finalización, genera informes ...
- **Puedes utilizar el Inventario de S3 para obtener la lista de objetos y utilizar S3 Select para filtrarlos**



Seguridad de Amazon S3



Amazon S3 - Cifrado de objetos

- Puedes cifrar objetos en buckets de S3 utilizando uno de los 4 métodos siguientes
- **Cifrado del lado del servidor (SSE)**
 - **Cifrado del lado del servidor con claves gestionadas por Amazon S3 (SSE-S3) -**
 - Activado por defecto
 - Cifra los objetos de S3 utilizando claves manejadas, gestionadas y propiedad de AWS
 - **Cifrado del lado del servidor con claves KMS almacenadas en AWS KMS (SSE-KMS)**
 - Aprovecha el servicio de administración de claves de AWS (AWS KMS) para gestionar las claves de cifrado
 - **Cifrado del lado del servidor con claves proporcionadas por el cliente (SSE-C)**
 - Cuando quieras gestionar tus propias claves de cifrado
- **Cifrado del lado del cliente**
- Es importante entender cuáles son para cada situación para el examen

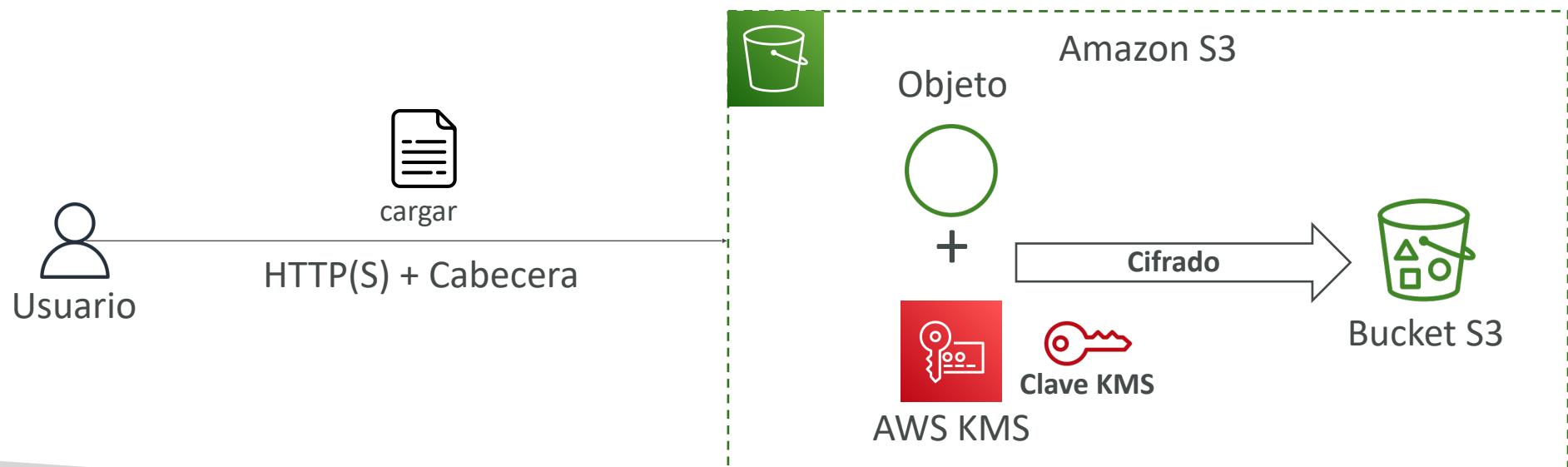
Cifrado de Amazon S3 - SSE-S3

- Cifrado mediante claves manejadas, gestionadas y propiedad de AWS
- El objeto está cifrado en el lado del servidor
- El tipo de cifrado es **AES-256**
- Debe establecerse la cabecera "x-amz-server-side-encryption": "AES256"
- **Activado por defecto para nuevos buckets y nuevos objetos**



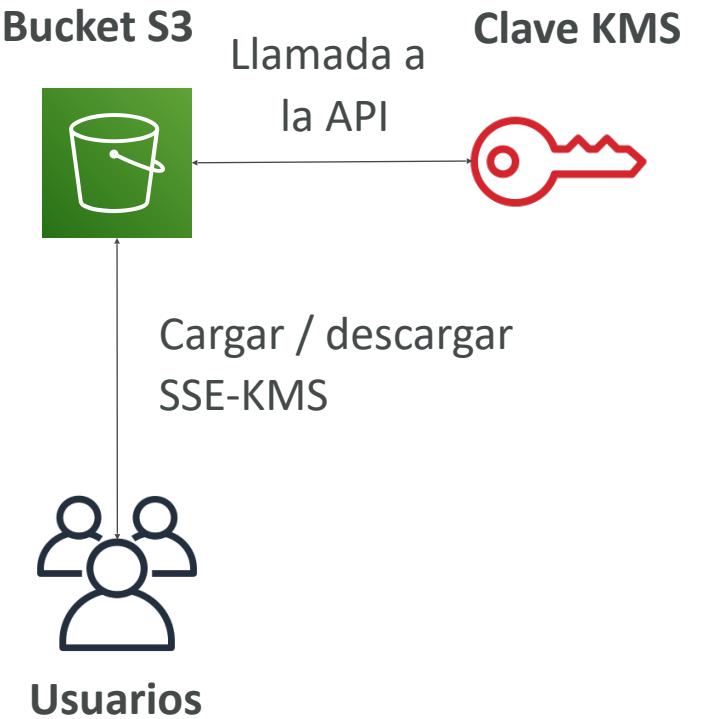
Cifrado de Amazon S3 - SSE-KMS

- Cifrado mediante claves manejadas y gestionadas por AWS KMS (Key Management Service)
- Ventajas del KMS: control del usuario + auditoría del uso de las claves mediante CloudTrail
- El objeto está cifrado en el lado del servidor
- Debes establecer la cabecera "**x-amz-server-side-encryption": "aws:kms"**



Limitación de SSE-KMS

- Si utilizas SSE-KMS, puede que te afecten los límites del KMS
- Cuando subes, llama a la API KMS **GenerateDataKey**
- Cuando descargas, llama a la API KMS **Decrypt**
- Cuenta para la cuota KMS por segundo (5500, 10000, 30000 req/s según la región)
- Puedes solicitar un aumento de cuota mediante la Consola de Cuotas de Servicio



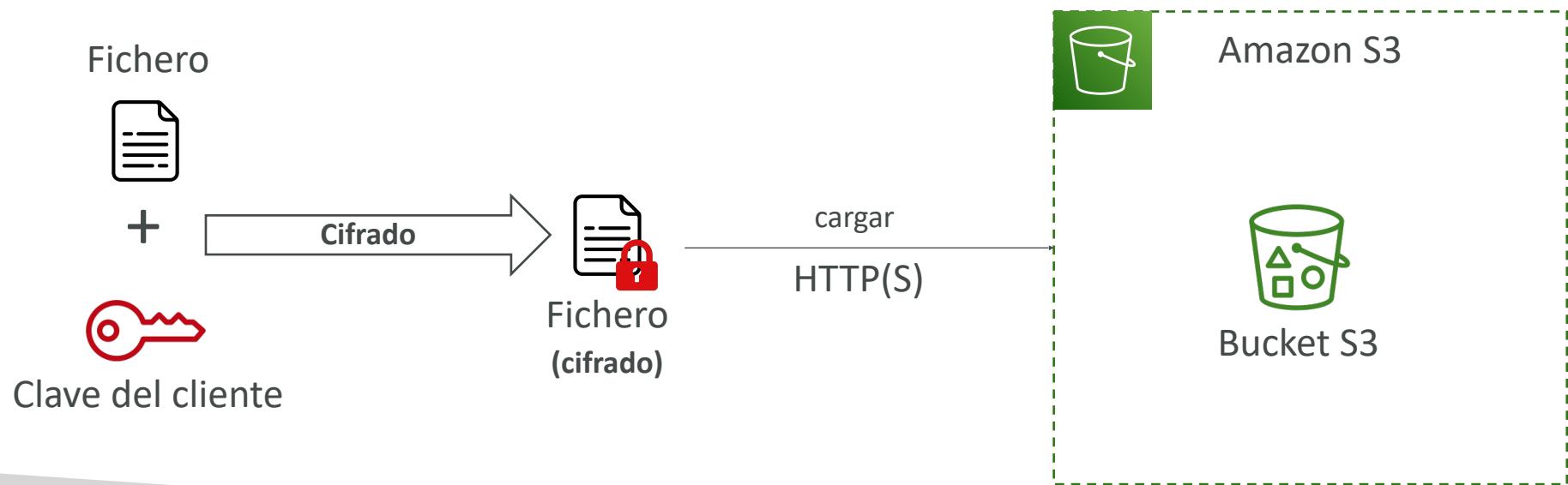
Cifrado de Amazon S3 - SSE-C

- Cifrado del lado del servidor mediante claves totalmente gestionadas por el cliente fuera de AWS
- Amazon S3 **NO** almacena la clave de cifrado que proporcionas
- **Se tiene que utilizar HTTPS**
- La clave de cifrado debe proporcionarse en las cabeceras HTTP, para cada petición HTTP realizada



Cifrado de Amazon S3 - Cifrado del lado del cliente

- Utiliza bibliotecas de clientes como la **biblioteca de cifrado del lado del cliente de Amazon S3**
- Los clientes deben cifrar los datos ellos mismos antes de enviarlos a Amazon S3
- Los clientes deben descifrar los datos ellos mismos al recuperarlos de Amazon S3
- El cliente gestiona completamente las claves y el ciclo de cifrado



Amazon S3 - Cifrado en tránsito (SSL/TLS)

- El cifrado en vuelo también se llama SSL/TLS
- Amazon S3 expone dos endpoints:
 - **HTTP Endpoint** - no cifrado
 - **Endpoint HTTPS** - cifrado en vuelo
- **Se recomienda HTTPS**
- **HTTPS es obligatorio para SSE-C**
- La mayoría de los clientes usarán el endpoint HTTPS por defecto



Amazon S3 - Políticas de cifrado por defecto vs. bucket

- Una forma de "forzar el cifrado" es utilizar una política de bucket y rechazar cualquier llamada a la API para PUT un objeto S3 sin cabeceras de cifrado

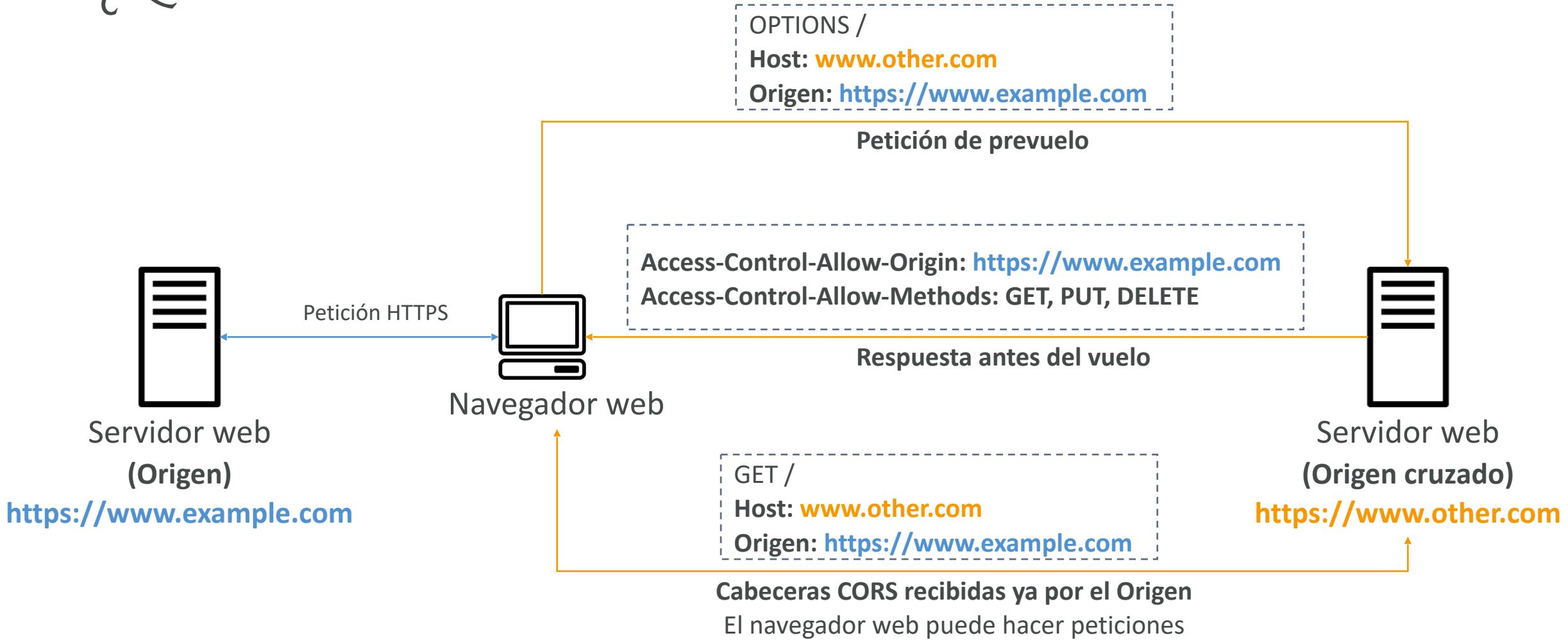
```
{  
    "Version": "2012-10-17",  
    "Statement": [  
        {  
            "Sid": "DenyIncorrectDecryptionHeader",  
            "Effect": "Deny",  
            "Principal": "*",  
            "Action": [ "s3:PutObject" ],  
            "Resource": [ "arn:aws:s3:::examplebucket/*" ],  
            "Condition": {  
                "StringNotEquals": {  
                    "s3:x-amz-server-side-encryption": "AES256"  
                }  
            }  
        }  
    ]  
}  
  
{  
    "Version": "2012-10-17",  
    "Statement": [  
        {  
            "Sid": "DenyUnencryptedObjectUploads",  
            "Effect": "Deny",  
            "Principal": "*",  
            "Action": [ "s3:PutObject" ],  
            "Resource": [ "arn:aws:s3:::examplebucket/*" ],  
            "Condition": {  
                "Null": {  
                    "s3:x-amz-server-side-encryption": true  
                }  
            }  
        }  
    ]  
}
```

- Otra forma es utilizar la opción de "cifrado por defecto" en S3
- Nota: Las políticas de los buckets se evalúan antes del "cifrado por defecto"

¿Qué es el CORS?

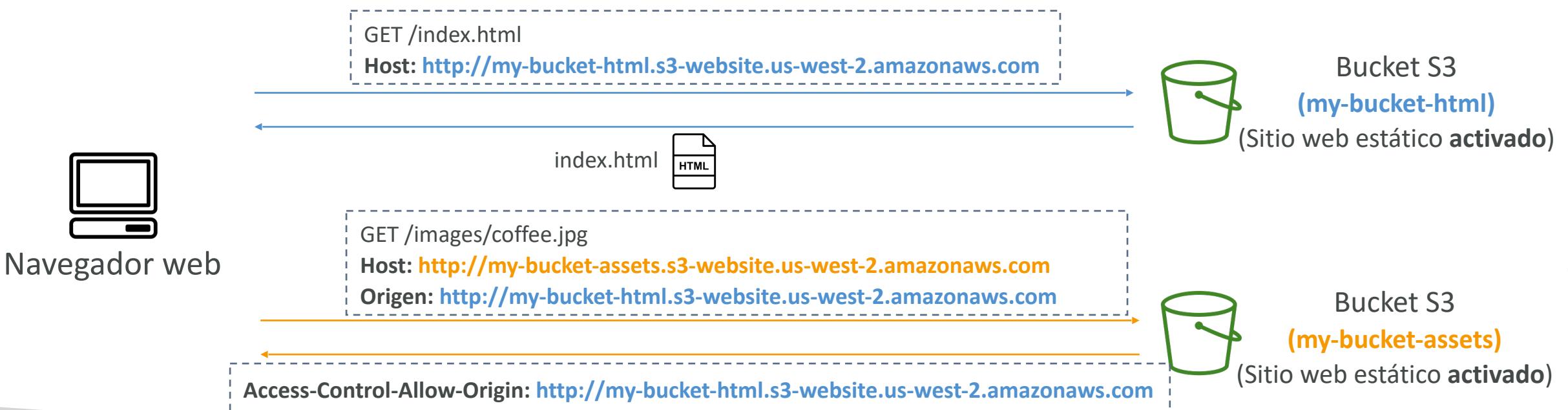
- **Compartir recursos entre orígenes (CORS)**
- **Origen = esquema (protocolo) + host (dominio) + puerto**
 - Ejemplo: <https://www.example.com> (el puerto implícito es 443 para HTTPS, 80 para HTTP)
- Mecanismo **basado en el navegador web** para permitir peticiones a otros orígenes mientras se visita el origen principal
- El mismo origen: <http://example.com/app1> y <http://example.com/app2>
- Diferentes orígenes: <http://www.example.com> y <http://other.example.com>
- Las peticiones no se cumplirán a menos que el otro origen permita las peticiones, utilizando **cabeceras CORS** (ejemplo: **Access-Control-Allow-Origin**)

¿Qué es el CORS?



Amazon S3 – CORS

- Si un cliente hace una petición de origen cruzado en nuestro bucket de S3, tenemos que habilitar las cabeceras CORS correctas
- Es una pregunta de examen muy popular
- Puedes permitir un origen específico o * (todos los orígenes)



Amazon S3 - Eliminación de MFA

- **MFA (Autenticación de Factores Múltiples)**: obliga a los usuarios a generar un código en un dispositivo (normalmente un teléfono móvil o un hardware) antes de realizar operaciones importantes en el S3
- MFA será necesario para:
 - Eliminar permanentemente una versión de un objeto
 - Suspender el control de versiones en el bucket
- MFA no será necesario para:
 - Habilitar el control de versiones
 - Listar las versiones eliminadas
- Para utilizar MFA Delete, **el control de versiones debe estar activado** en el bucket
- **Sólo el propietario del bucket (cuenta root) puede activar/desactivar MFA Delete**



Autenticador de Google



Dispositivo de hardware MFA

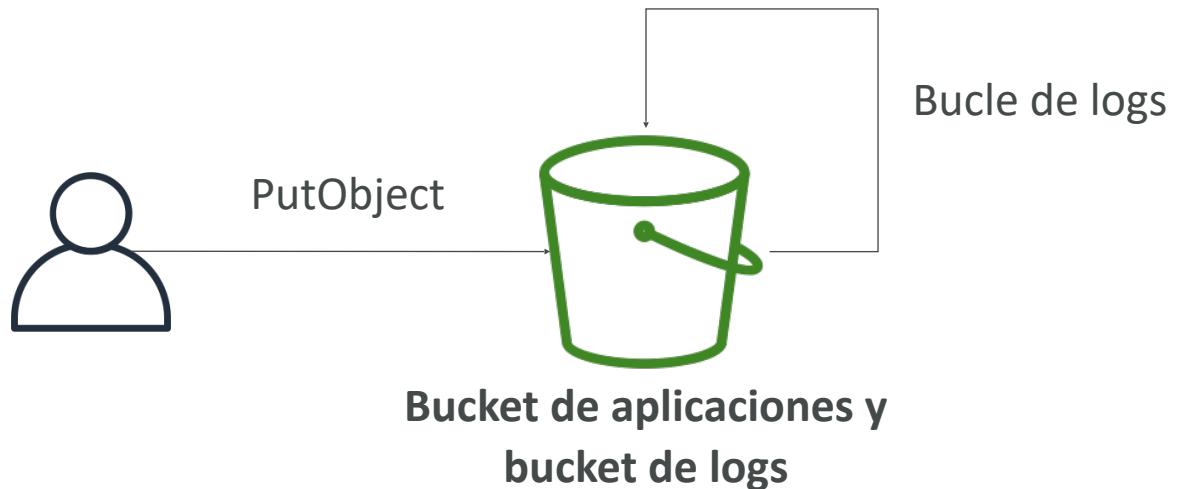
S3 Access Logs

- Para fines de auditoría, es posible que quieras registrar todos los accesos a los buckets de S3
- Cualquier petición realizada a S3, desde cualquier cuenta, autorizada o denegada, se registrará en otro bucket de S3
- Esos datos pueden ser analizados con herramientas de análisis de datos...
- El bucket de logs de destino debe estar en la misma región de AWS
- El formato de los logs está en: <https://docs.aws.amazon.com/AmazonS3/latest/dev/LogFormat.html>



S3 Access Logs: Warning

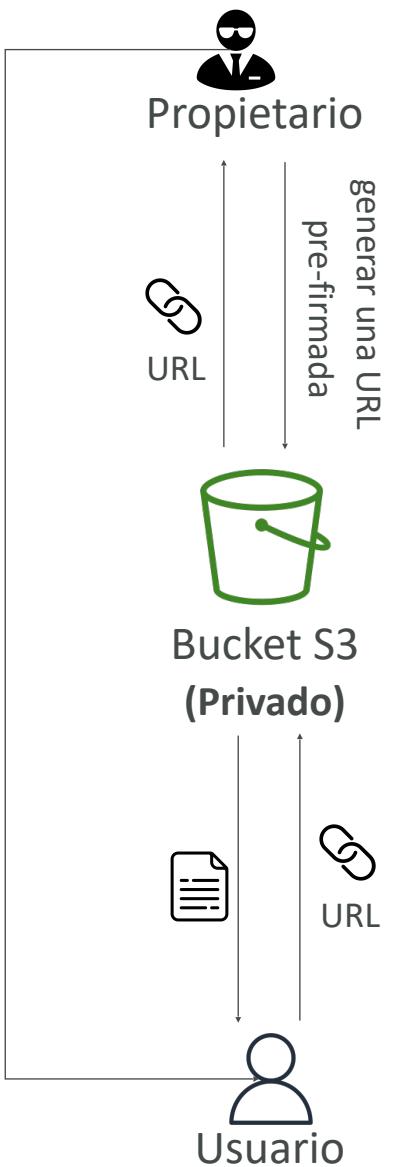
- No configures tu bucket de logs para que sea el bucket monitorizado
- Se creará un bucle de logs, y **tu bucket crecerá exponencialmente**



No lo intentes en casa ☺

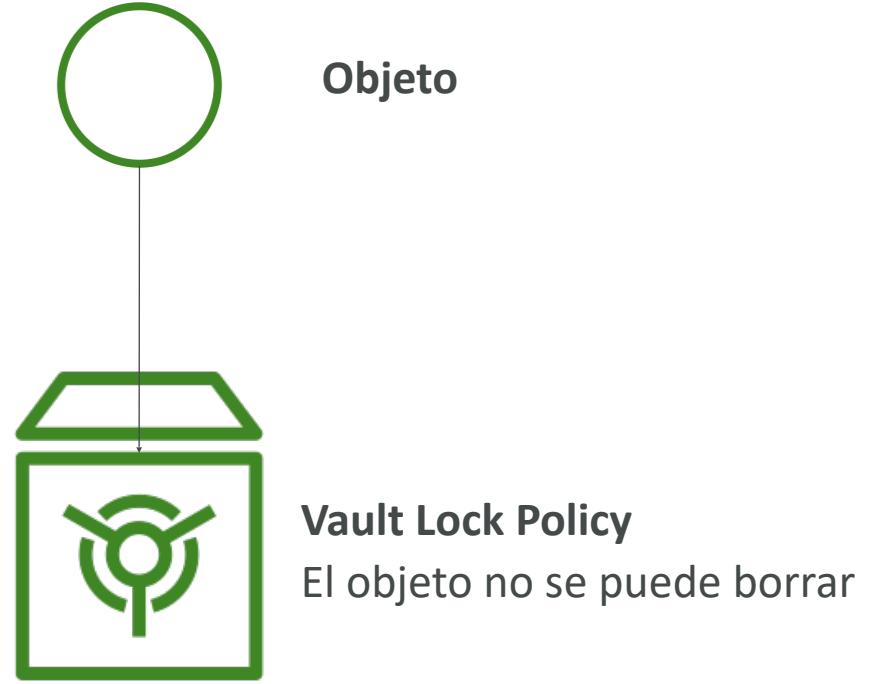
Amazon S3 - URLs pre-firmadas

- Generar URLs pre-firmadas usando la consola de **S3, la CLI de AWS o el SDK**
- **Expiración de la URL**
 - **Consola S3** - de 1 minuto a 720 minutos (12 horas)
 - **CLI de AWS** - configurar la caducidad con el parámetro --expires-in en segundos (por defecto 3600 segs, máx. 604800 segs ~ 168 horas)
- Los usuarios a los que se les da una URL pre-firmada heredan los permisos del usuario que generó la URL para GET / PUT
- Ejemplos:
 - Permite que sólo los usuarios que han iniciado sesión descarguen un vídeo premium de tu bucket de S3
 - Permitir que una lista cambiante de usuarios descargue archivos generando URLs dinámicamente
 - Permitir temporalmente que un usuario suba un archivo a una ubicación precisa en tu bucket de S3



S3 Glacier Vault Lock

- Adoptar un modelo WORM (Write Once Read Many)
- Crea una política de bloqueo de bóveda
- Bloquea la política para futuras ediciones (ya no se puede modificar ni borrar)
- Útil para el cumplimiento de la normativa y la retención de datos



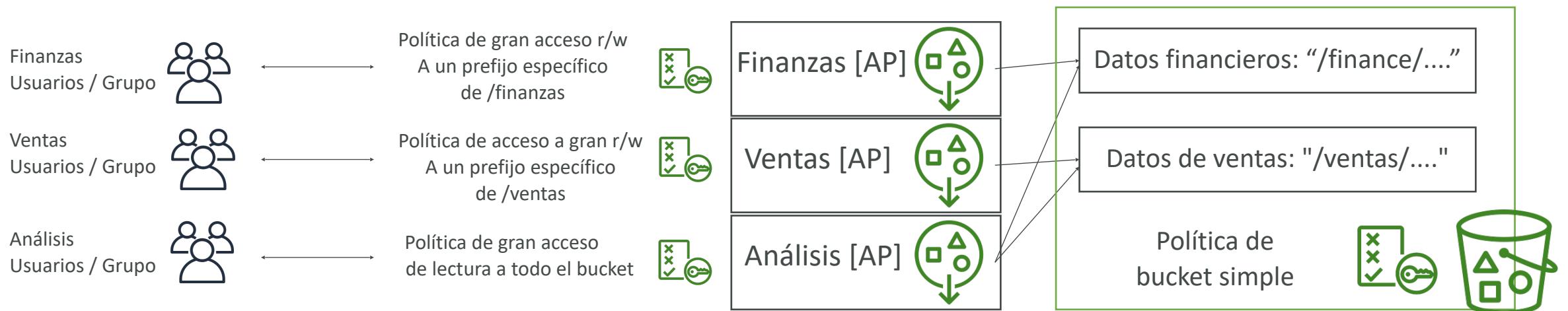
S3 Object Lock (el versionado debe estar activado)

- Adoptar un modelo WORM (Write Once Read Many)
- Bloquear el borrado de una versión del objeto durante un tiempo determinado
- **Modo de retención - Normativa:**
 - Las versiones de los objetos no pueden ser sobrescritas ni borradas por ningún usuario, incluido el usuario root
 - Los modos de retención de los objetos no pueden cambiarse, y los periodos de retención no pueden acortarse
- **Modo de retención - Gobernanza:**
 - La mayoría de los usuarios no pueden sobrescribir o eliminar una versión de un objeto ni alterar su configuración de bloqueo
 - Algunos usuarios tienen permisos especiales para cambiar la retención o eliminar el objeto
- **Periodo de retención:** protege el objeto durante un periodo fijo, que puede ser ampliado
- **Retención legal:**
 - protege el objeto indefinidamente, independientemente del periodo de retención
 - puede colocarse y eliminarse libremente mediante el permiso IAM s3:PutObjectLegalHold



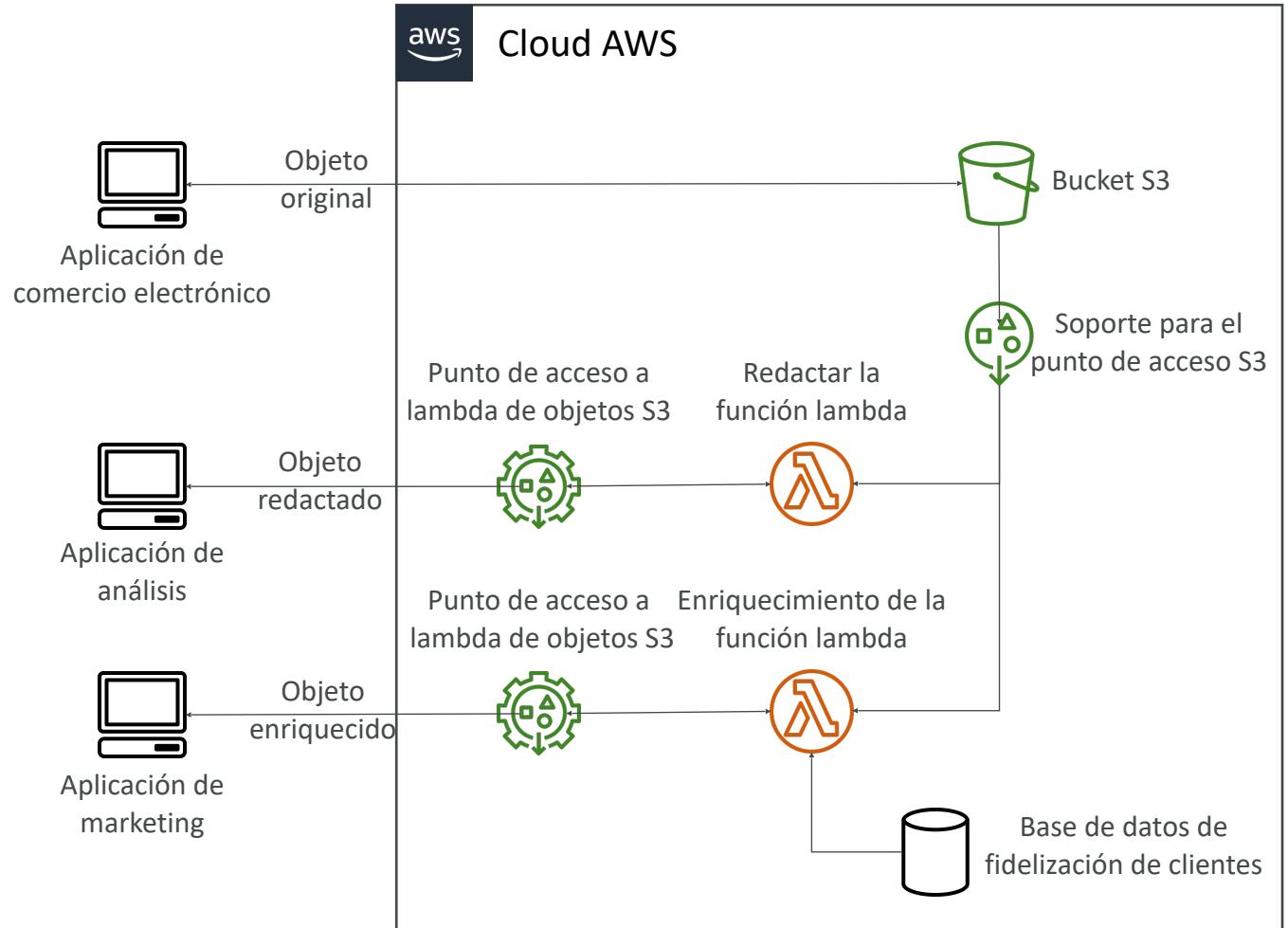
S3 - Puntos de acceso

- Cada punto de acceso tiene su propio DNS y política para limitar quién puede acceder a él
 - Un usuario / grupo IAM específico
 - Una política por Punto de Acceso => **Más fácil de gestionar que las complejas políticas de bucket**



Objeto S3 Lambda

- Utiliza las Funciones Lambda de AWS para modificar el objeto antes de que lo recupere la aplicación que lo llama
- Sólo se necesita un bucket de S3, sobre el que creamos **puntos de acceso de S3** y **puntos de acceso de S3 Object Lambda**.
- Casos de uso:
 - Redactar información de identificación personal para entornos de análisis o de no producción.
 - Convertir entre formatos de datos, como convertir XML a JSON.
 - Redimensionar y poner marcas de agua a las imágenes sobre la marcha utilizando detalles específicos de la persona que llama, como el usuario que solicitó el objeto.

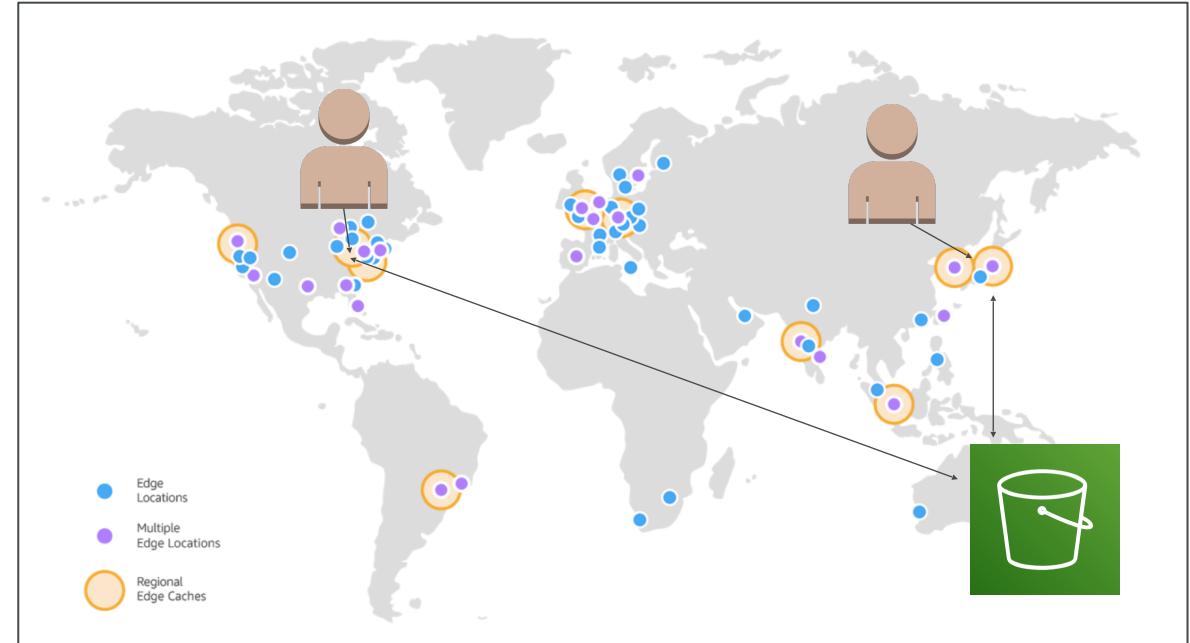


Infraestructura global

AWS CloudFront



- Red de entrega de contenidos (CDN)
- **Mejora el rendimiento de lectura, el contenido se almacena en caché en edge location**
- Mejora la experiencia de los usuarios
- 216 puntos de presencia a nivel mundial (ubicaciones edge)
- **Protección DDoS, integración con Shield, AWS Web Application Firewall**



Fuente: <https://aws.amazon.com/cloudfront/features/?nc=sn&loc=2>

CloudFront - Orígenes

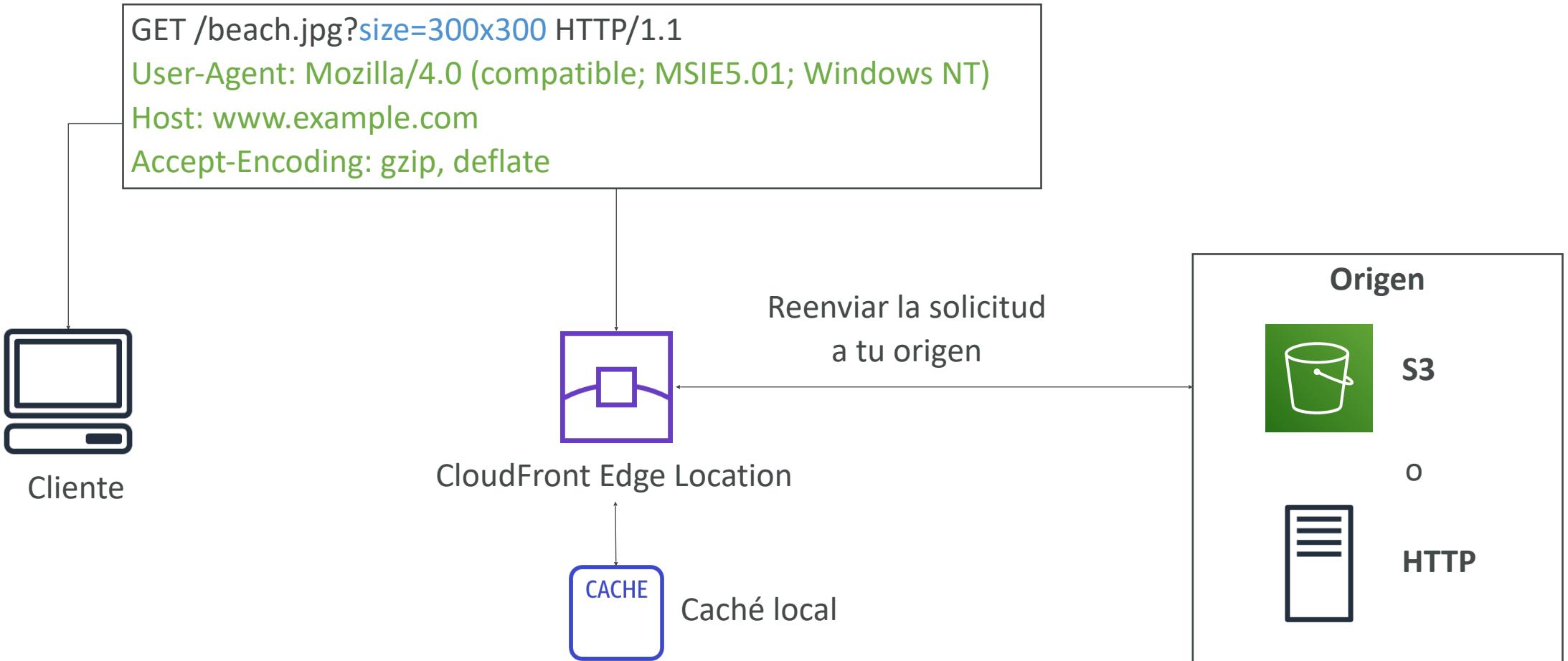
- **Bucket S3**

- Para distribuir archivos y almacenarlos en caché en el borde
- Seguridad mejorada con CloudFront Origin Access Identity (OAI)
- OAC sustituye a Origin Access Identity (OAI)
- CloudFront puede utilizarse como entrada (para subir archivos a S3)

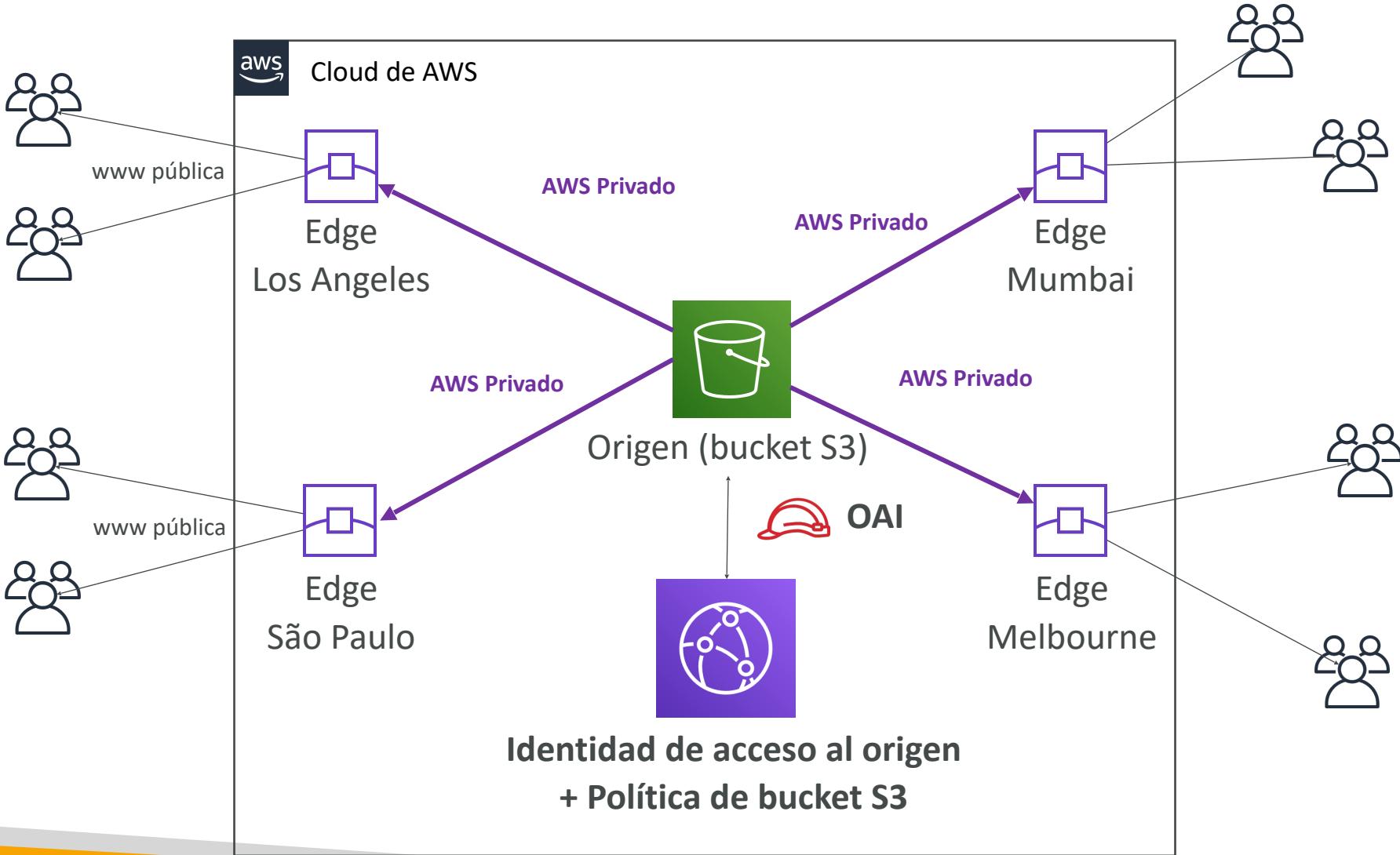
- **Origen personalizado (HTTP)**

- Application Load Balancer
- Instancia EC2
- Sitio web de S3 (primero debes habilitar el bucket como sitio web estático de S3)
- Cualquier backend HTTP que quieras

CloudFront a alto nivel



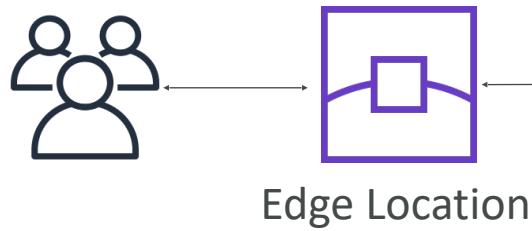
CloudFront - S3 como origen



CloudFront vs S3 Cross Region Replication (CRR)

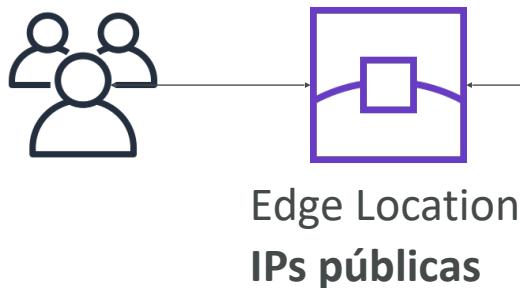
- CloudFront:
 - Red Global Edge
 - Los archivos se almacenan en caché durante un TTL (quizás un día)
 - **Es ideal para contenidos estáticos que deben estar disponibles en todas partes**
- S3 Cross Region Replication (CRR):
 - Debe configurarse para cada región en la que quieras que se produzca la replicación
 - Los archivos se actualizan casi en tiempo real
 - Sólo lectura
 - **Ideal para contenidos dinámicos que deben estar disponibles con baja latencia en pocas regiones**

CloudFront - ALB o EC2 como origen



Permitir IP pública de Edge Locations

<http://d7uri8nf7uskq.cloudfront.net/tools/list-cloudfront-ips>



Permitir la IP pública de Edge Locations



Permitir Grupo de Seguridad del Load Balancer



Restricción geográfica de CloudFront

- Puedes restringir quién puede acceder a tu distribución
 - **Lista de permitidos:** Permite que tus usuarios accedan a tu contenido sólo si están en uno de los países de una lista de países aprobados.
 - **Lista de bloqueo:** Evita que tus usuarios accedan a tu contenido si se encuentran en uno de los países de la lista de países prohibidos.
- El "país" se determina utilizando una base de datos Geo-IP de terceros
- Caso de uso: Leyes de derechos de autor para controlar el acceso a los contenidos

CloudFront - Precios

- Las Edge Locations de CloudFront están por todo el mundo
- El coste de los datos por Edge Locations varía

Per Month	United States, Mexico, & Canada	Europe & Israel	South Africa, Kenya, & Middle East	South America	Japan	Australia & New Zealand	Hong Kong, Philippines, Singapore, South Korea, Taiwan, & Thailand	India
First 10TB	\$0.085	\$0.085	\$0.110	\$0.110	\$0.114	\$0.114	\$0.140	\$0.170
Next 40TB	\$0.080	\$0.080	\$0.105	\$0.105	\$0.089	\$0.098	\$0.135	\$0.130
Next 100TB	\$0.060	\$0.060	\$0.090	\$0.090	\$0.086	\$0.094	\$0.120	\$0.110
Next 350TB	\$0.040	\$0.040	\$0.080	\$0.080	\$0.084	\$0.092	\$0.100	\$0.100
Next 524TB	\$0.030	\$0.030	\$0.060	\$0.060	\$0.080	\$0.090	\$0.080	\$0.100
Next 4PB	\$0.025	\$0.025	\$0.050	\$0.050	\$0.070	\$0.085	\$0.070	\$0.100
Over 5PB	\$0.020	\$0.020	\$0.040	\$0.040	\$0.060	\$0.080	\$0.060	\$0.100

bajo → más alto

CloudFront - Clases de precios

- Puedes reducir el número de Edge Locations para **reducir los costes**
- Tres clases de precios:
 1. Clase de precio Todos: todas las regiones - mejor rendimiento
 2. Clase de precio 200: la mayoría de las regiones, pero excluye las regiones más caras
 3. Clase de precio 100: sólo las regiones menos caras

Edge Locations Included Within	United States, Mexico, & Canada	Europe & Israel	South Africa, Kenya, & Middle East	South America	Japan	Australia & New Zealand	Hong Kong, Philippines, Singapore, South Korea, Taiwan, & Thailand	India
Price Class All	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Price Class 200	Yes	Yes	Yes	x	Yes	x	Yes	Yes
Price Class 100	Yes	Yes	x	x	x	x	x	x

CloudFront - Clase de precio

Precios Clase: 100



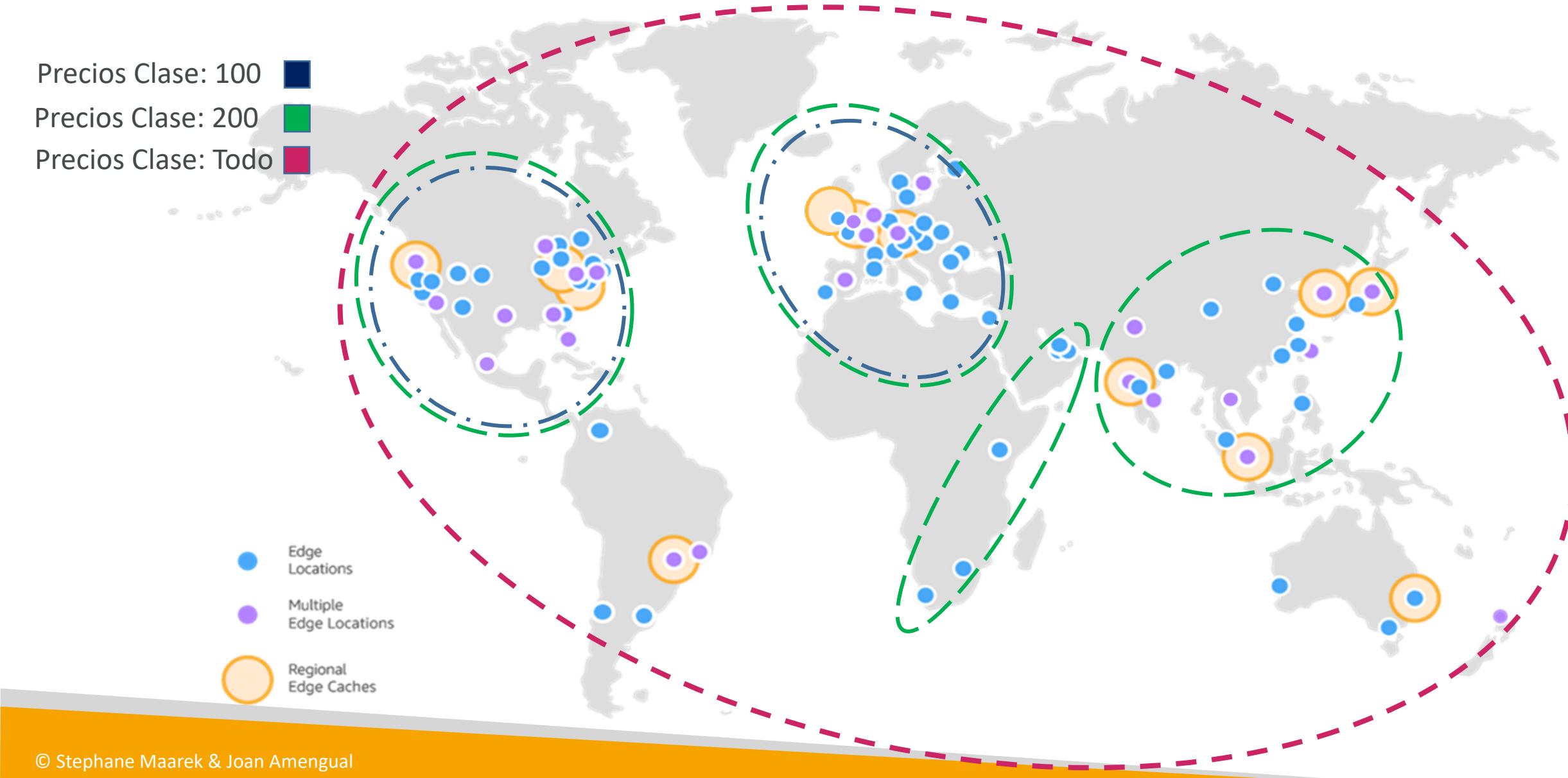
Precios Clase: 200



Precios Clase: Todo

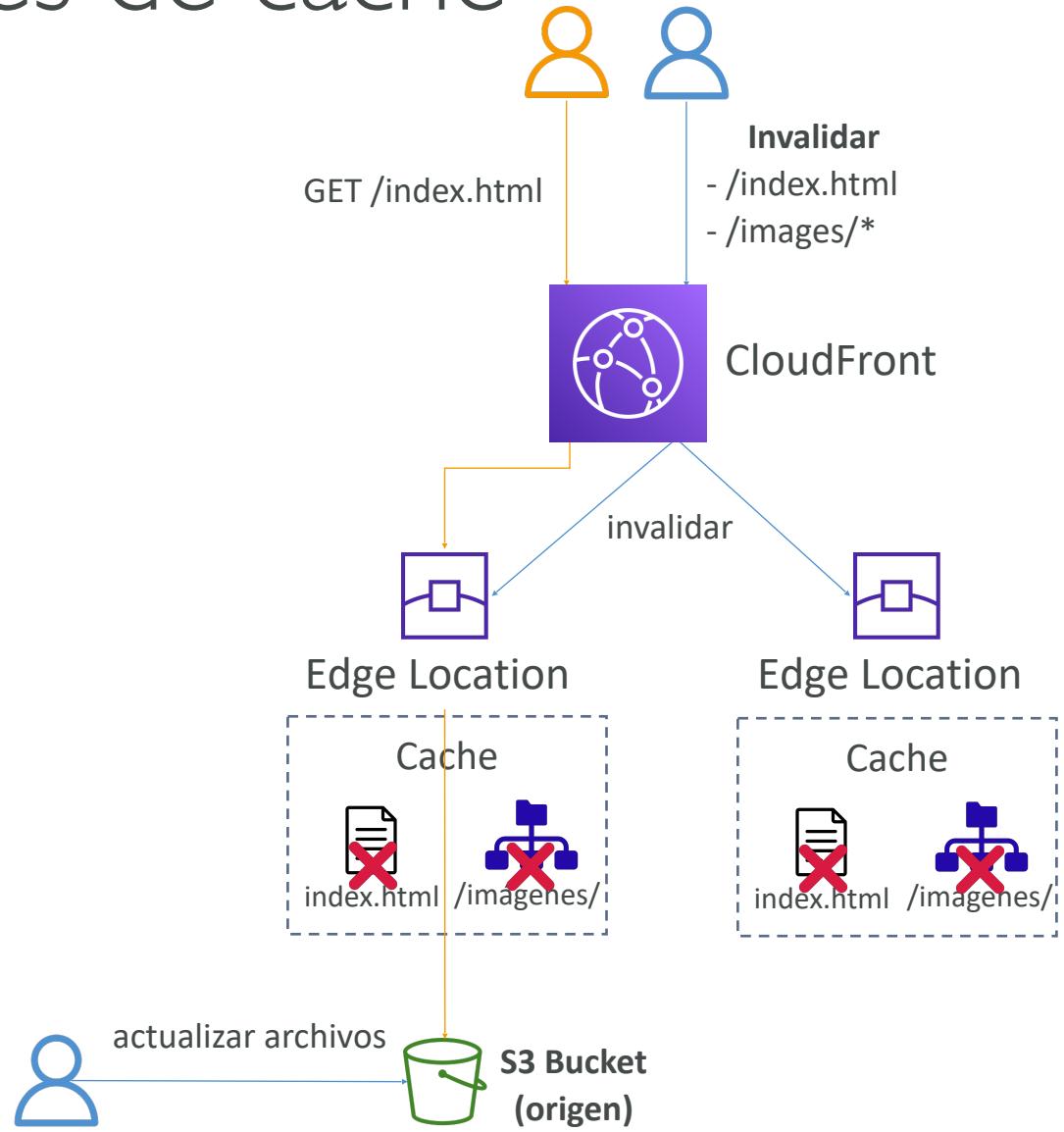


- Edge Locations
- Multiple Edge Locations
- Regional Edge Caches



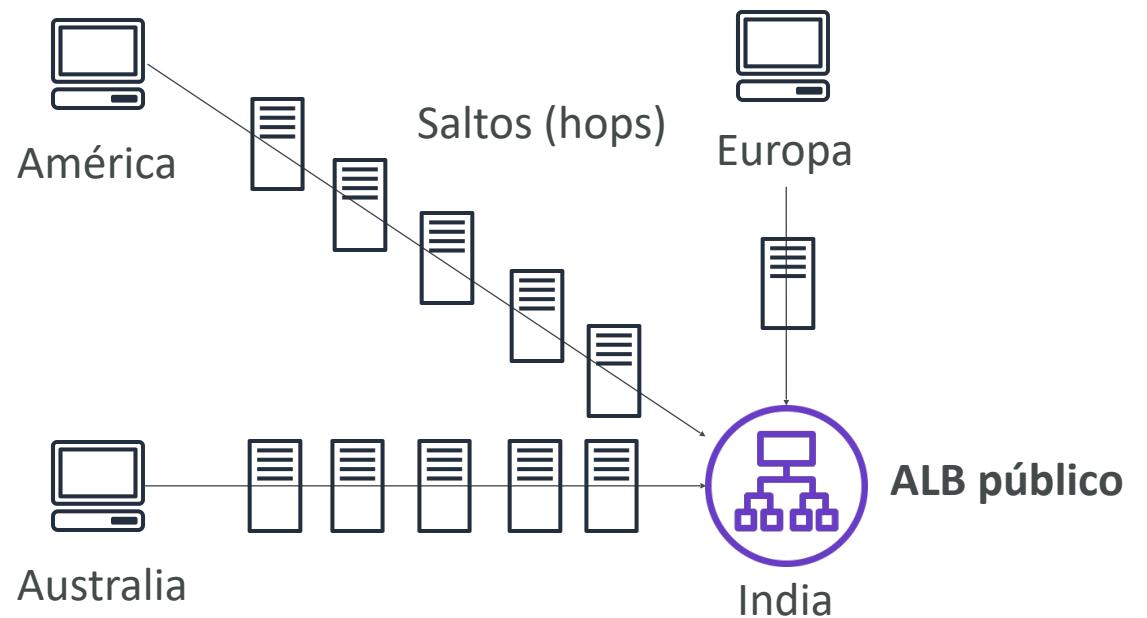
CloudFront - Invalidaciones de caché

- En caso de que actualices el origen del back-end, CloudFront no lo sabe y sólo obtendrá el contenido refrescado cuando el TTL haya expirado
- Sin embargo, puedes forzar una actualización total o parcial de la caché (obviando así el TTL) realizando una **Invalidación de CloudFront**
- Puedes invalidar todos los archivos (*) o una ruta especial (/imágenes/*)



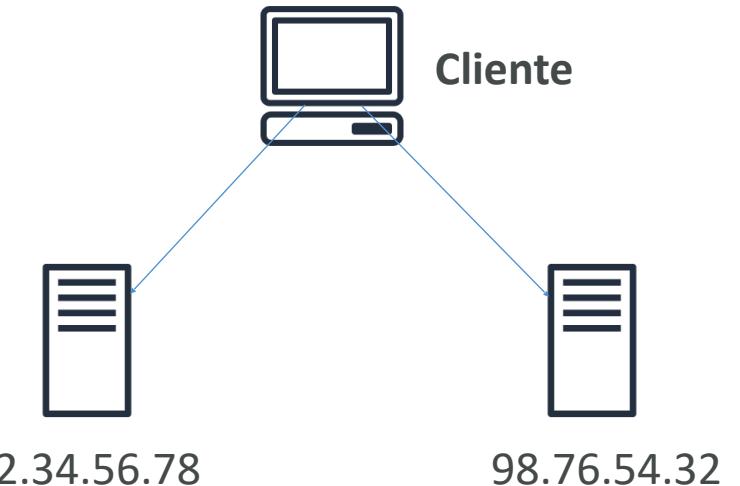
Usuarios globales para nuestra aplicación

- Has desplegado una aplicación y tienes usuarios globales que quieren acceder a ella directamente.
- Van a través de la Internet pública, lo que puede añadir mucha latencia debido a los numerosos saltos
- Queremos ir lo más rápido posible a través de la red de AWS para minimizar la latencia

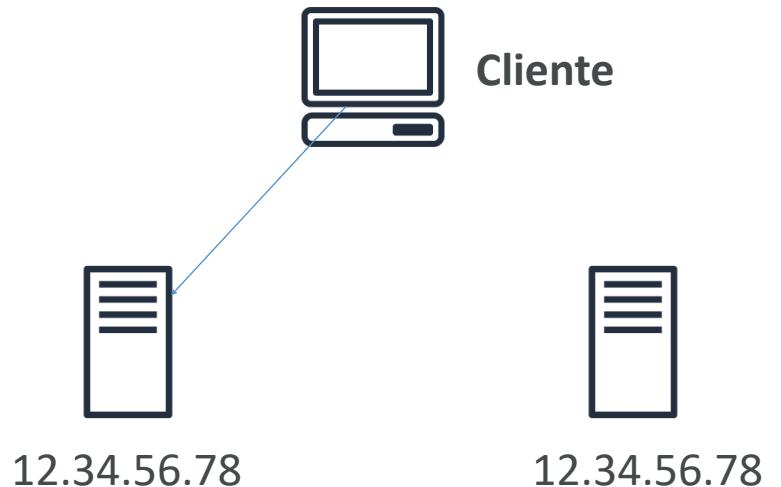


IP unicast vs IP anycast

- **IP unicast:** un servidor tiene una dirección IP



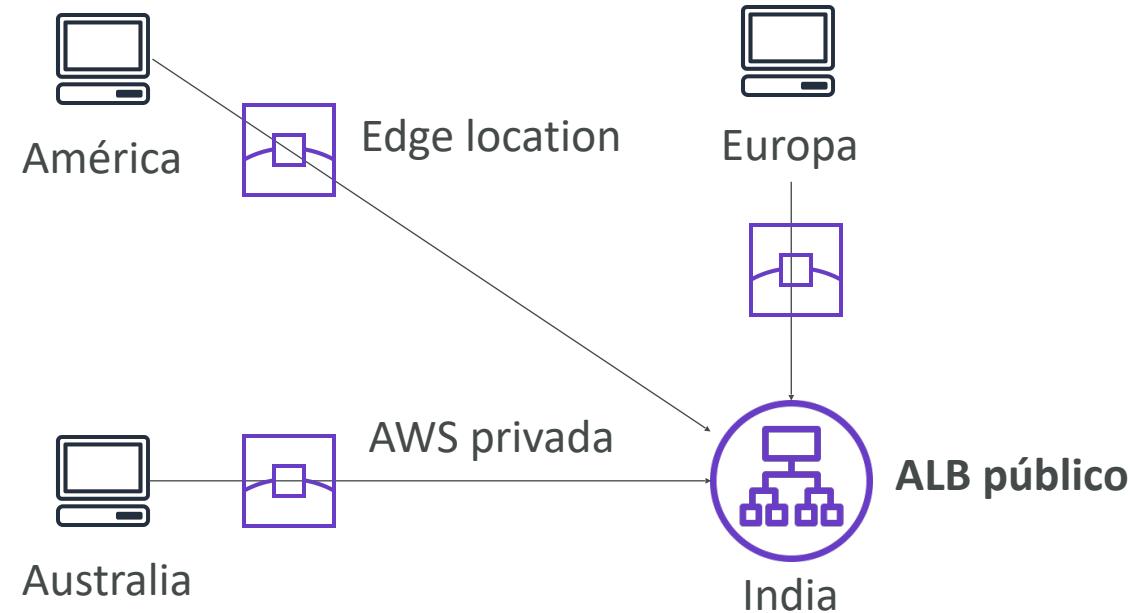
- **IP Anycast:** todos los servidores tienen la misma dirección IP y el cliente es dirigido al más cercano



AWS Global Accelerator



- Aprovecha la red interna de AWS para dirigirte a tu aplicación
- Se **crean 2 IP Anycast** para tu aplicación
- Las IP Anycast envían el tráfico directamente a las Edge Locations
- Las Edge Locations envían el tráfico a tu aplicación



AWS Global Accelerator

- Funciona con **Elastic IP, instancias EC2, ALB, NLB**
- Rendimiento consistente
 - Enrutamiento inteligente para lograr la menor latencia y una rápida conmutación por error regional
 - No hay problemas con la caché del cliente (porque la IP no cambia)
 - Red interna de AWS
- Comprobaciones de salud
 - Global Accelerator realiza una comprobación de la salud de tus aplicaciones
 - Ayuda a que tu aplicación sea global (comutación por error en menos de 1 minuto en caso de no ser saludable)
 - Genial para la recuperación de desastres (gracias a las comprobaciones de salud)
- Seguridad
 - Sólo hay que poner en la lista blanca 2 IP externas
 - Protección DDoS gracias a AWS Shield

AWS Global Accelerator vs CloudFront

- Ambos utilizan la red global de AWS y sus Edge Locations en todo el mundo
- Ambos servicios se integran con AWS Shield para la protección DDoS.
- **CloudFront**
 - Mejora el rendimiento tanto del contenido almacenable en caché (como imágenes y videos)
 - Contenido dinámico (como la aceleración de la API y la entrega de sitios dinámicos)
 - El contenido se sirve en el borde
- **Global Accelerator**
 - Mejora el rendimiento de una amplia gama de aplicaciones sobre TCP o UDP
 - Proxy de paquetes en el borde a las aplicaciones que se ejecutan en una o más regiones de AWS.
 - Es adecuado para casos de uso no HTTP, como juegos (UDP), IoT (MQTT) o voz sobre IP
 - Bueno para casos de uso HTTP que requieren direcciones IP estáticas
 - Bueno para casos de uso de HTTP que requieran una comutación por error regional determinista y rápida

Almacenamiento avanzado en AWS

Familia AWS Snow

- Dispositivos portátiles de alta seguridad para **recopilar, procesar datos, y migrar datos hacia y desde AWS**

- **Migración de datos:**



Snowcone



Snowball Edge



Snowmobile

- **Edge computing:**



Snowcone



Snowball Edge

Migraciones de datos con AWS Snow Family

	Tiempo de transferencia		
	100 Mbps	1Gbps	10Gbps
10 TB	12 días	30 horas	3 horas
100 TB	124 días	12 días	30 horas
1 PB	3 años	124 días	12 días

Desafíos:

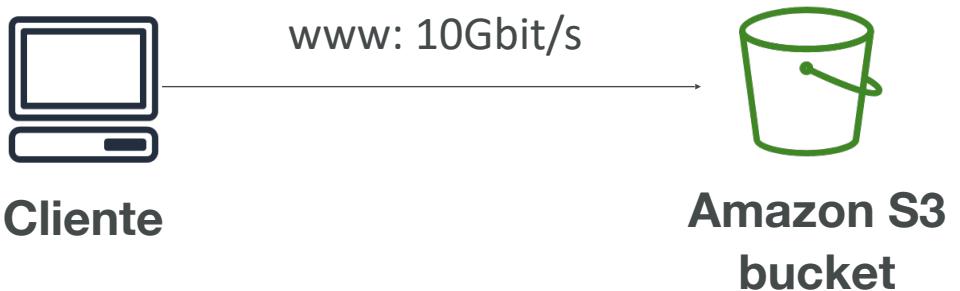
- Conectividad limitada
- Ancho de banda limitado
- Alto coste de la red
- Ancho de banda compartido
(no se puede maximizar la línea)
- Estabilidad de la conexión

Familia AWS Snow: dispositivos sin conexión para realizar migraciones de datos

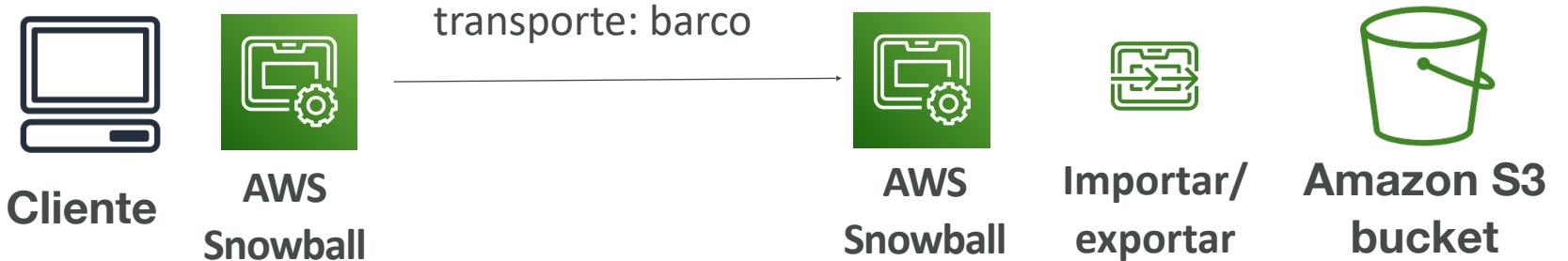
Si la transferencia a través de la red tarda más de una semana, ¡utiliza los dispositivos Snowball!

Diagramas

- Subida directa a S3:



- Con la Familia Snow:



Snowball Edge (para las transferencias de datos)



- Solución de transporte físico de datos: mover TBs o PBs de datos dentro o fuera de AWS
- Alternativa al traslado de datos por la red (y al pago de tarifas de red)
- Paga por trabajo de transferencia de datos
- Proporciona almacenamiento en bloque y almacenamiento de objetos compatible con Amazon S3
- **Snowball Edge – Storage Optimized**
 - 80 TB de capacidad de disco duro para volumen de bloques y almacenamiento de objetos compatible con S3
- **Snowball Edge – Compute Optimized**
 - 42 TB de capacidad de disco duro para volumen de bloques y almacenamiento de objetos compatible con S3
- Casos de uso:
 - Migraciones a el Cloud de grandes datos
 - Recuperación de desastres



AWS Snowcone



- **Computación pequeña y portátil, en cualquier lugar, robusta y segura, soporta entornos difíciles**
- Ligero (4,5 libras, 2,1 kg)
- Dispositivo utilizado para la computación, el almacenamiento y la transferencia de datos
- **8 TB de almacenamiento utilizable**
- Utiliza Snowcone donde no quepa Snowball (entorno con limitaciones de espacio)
- Debes proporcionar tu propia batería / cables
- Puede enviarse a AWS sin conexión, o conectarlo a Internet y utilizar **AWS DataSync** para enviar los datos



AWS Snowmobile



- Transfiere exabytes de datos ($1 \text{ EB} = 1.000 \text{ PB} = 1.000.000 \text{ TBs}$)
- Cada Snowmobile tiene 100 PB de capacidad (utiliza varias en paralelo)
- Alta seguridad: temperatura controlada, GPS, videovigilancia 24/7
- **Mejor que la Snowball si transfieres más de 10 PB**

Familia AWS Snow para las migraciones de datos



Snowcone



Snowball Edge



Snowmobile



	Snowcone	Snowball Edge Storage Optimized	Snowmobile
Capacidad de almacenamiento	8 TB usable	80 TB usable	< 100 PB
Tamaño de la migración	Hasta 24 TB, online y offline	Hasta petabytes, offline	Hasta exabytes, offline
Agente DataSync	Preinstalado		
Clusters de almacenamiento		Hasta 15 nodes	

Familia AWS Snow - Proceso de uso

1. Solicita la entrega de dispositivos Snowball desde la consola de AWS
2. Instala el cliente Snowball / AWS OpsHub en tus servidores
3. Conecta el Snowball a tus servidores y copia los archivos utilizando el cliente
4. Devuelve el dispositivo cuando hayas terminado (va a la instalación de AWS adecuada)
5. Los datos se cargarán en un bucket de S3
6. La Snowball se borrará por completo

Arquitectura de soluciones: De Snowball a Glacier

- **Snowball no puede importar a Glacier directamente**
- Debes utilizar primero Amazon S3, en combinación con una política de ciclo de vida de S3



Amazon FSx - Visión general



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



FSx para Lustre



FSx para
Windows
File Server



FSx para
ONTAP de NetApp



FSx para
OpenZFS

Amazon FSx para Windows (Servidor de archivos)



- **FSx para Windows** es una unidad compartida del sistema de archivos de **Windows** totalmente gestionada
- Soporta el protocolo SMB y el NTFS de Windows
- Integración con Microsoft Active Directory, ACLs, cuotas de usuario
- **Se puede montar en instancias EC2 de Linux**
- Soporta **los espacios de nombres del Sistema de Archivos Distribuido** (DFS) de Microsoft (agrupa archivos en varios FS)
- Escala hasta 10s de GB/s, millones de IOPS, 100s PB de datos
- Opciones de almacenamiento:
 - **SSD** - cargas de trabajo sensibles a la latencia (bases de datos, procesamiento de medios, análisis de datos, ...)
 - **HDD** - amplio espectro de cargas de trabajo (directorio personal, CMS, ...)
- Se puede acceder desde tu infraestructura local (VPN o Direct Connect)
- Puede configurarse para ser Multi-AZ (alta disponibilidad)
- Los datos se respaldan diariamente en S3

Amazon FSx para Lustre



- Lustre es un tipo de sistema de archivos distribuido en paralelo, para la informática a gran escala
- El nombre Lustre deriva de "Linux" y "Cluster".
- Machine Learning, **computación de alto rendimiento (HPC)**
- Procesamiento de vídeo, modelado financiero, automatización del diseño electrónico
- Escala hasta 100s GB/s, millones de IOPS, latencias sub-ms
- Opciones de almacenamiento:
 - **SSD** - baja latencia, cargas de trabajo intensivas en IOPS, operaciones de archivos pequeños y aleatorios
 - **HDD** - cargas de trabajo intensivas en rendimiento, operaciones de archivos grandes y secuenciales
- **Perfecta integración con S3**
 - Puede "leer S3" como un sistema de archivos (a través de FSx)
 - Puede escribir la salida de los cálculos de vuelta a S3 (a través de FSx)
- Puede utilizarse desde servidores locales (VPN o Direct Connect)

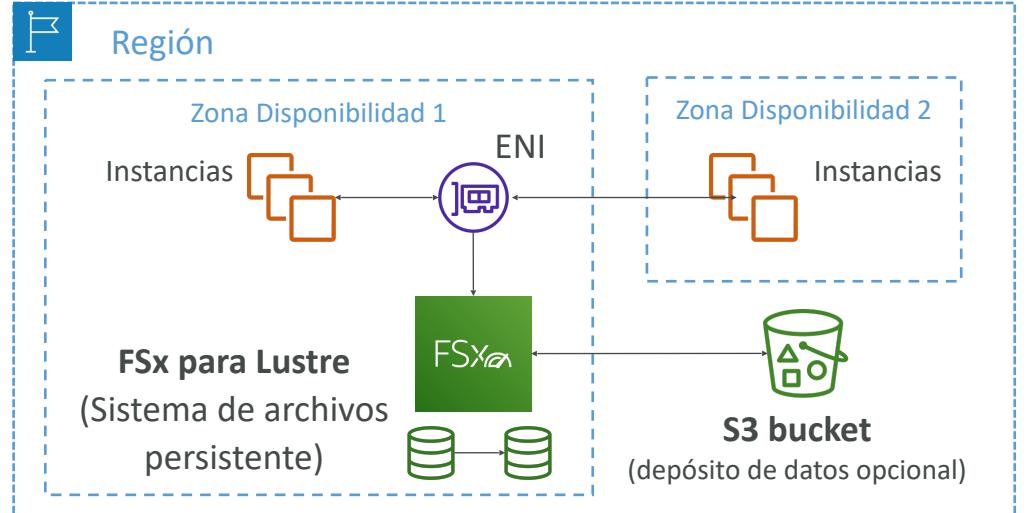
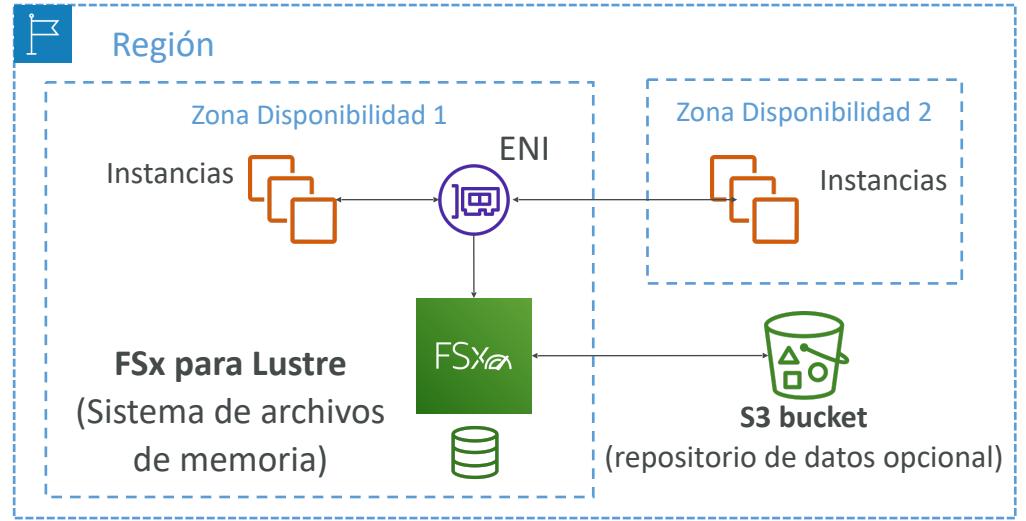
FSx Lustre - Opciones de despliegue del sistema de archivos

• Sistema de archivos de memoria

- Almacenamiento temporal
- Los datos no se replican (no persisten si falla el servidor de archivos)
- Alta velocidad (6 veces más rápido, 200MBps por TiB)
- Uso: procesamiento a corto plazo, optimizar los costes

• Sistema de archivos persistente

- Almacenamiento a largo plazo
- Los datos se replican dentro del mismo AZ
- Reemplaza los archivos fallidos en cuestión de minutos
- Uso: procesamiento a largo plazo, datos sensibles



Amazon FSx para NetApp ONTAP



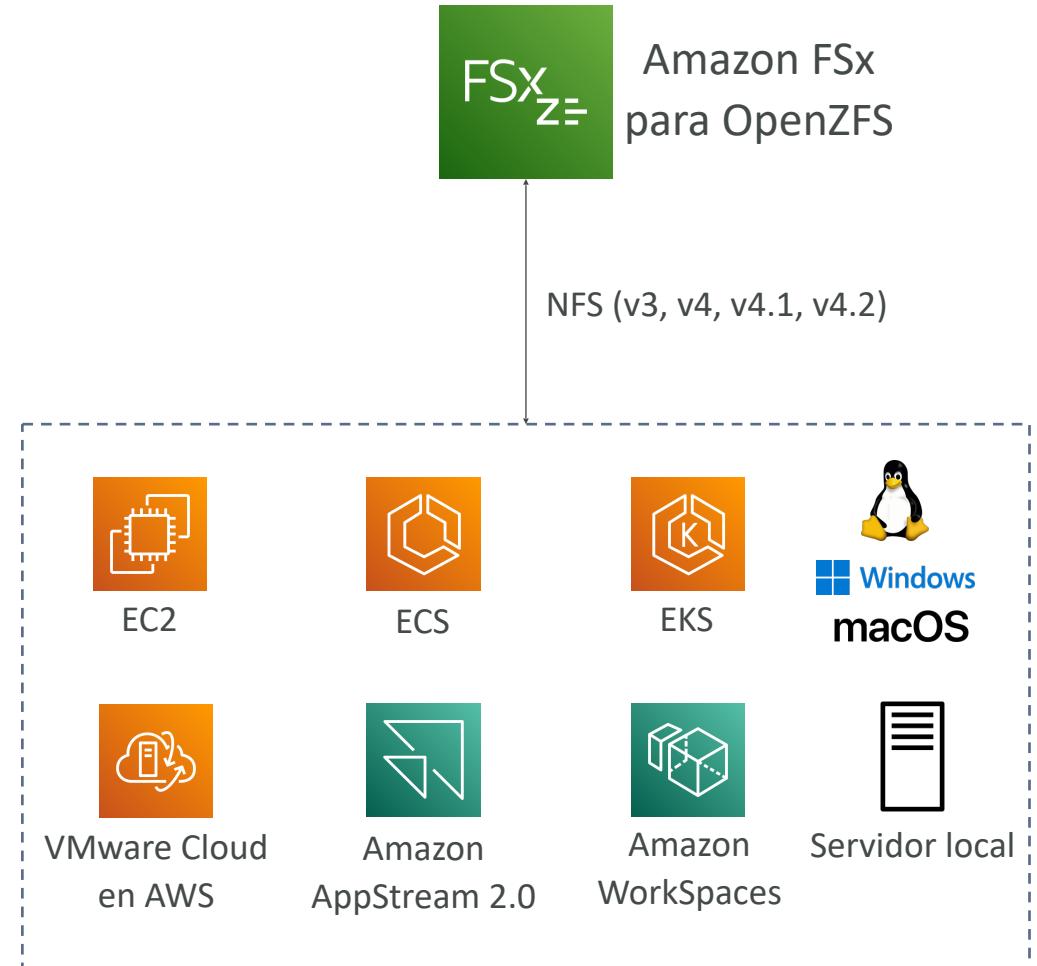
- NetApp ONTAP gestionado en AWS
- **Sistema de archivos compatible con el protocolo NFS, SMB, iSCSI**
- Mueve las cargas de trabajo que se ejecutan en ONTAP o NAS a AWS
- Funciona con:
 - Linux
 - Windows
 - MacOS
 - VMware Cloud en AWS
 - Amazon Workspaces y AppStream 2.0
 - Amazon EC2, ECS y EKS
- El almacenamiento se reduce o crece automáticamente
- Snapshots, replicación, bajo coste, compresión y desduplicación de datos
- **Clonación instantánea puntual (útil para probar nuevas cargas de trabajo)**



Amazon FSx para OpenZFS



- Sistema de archivos OpenZFS gestionado en AWS
- Sistema de archivos compatible con NFS (v3, v4, v4.1, v4.2)
- Mueve las cargas de trabajo que se ejecutan en ZFS a AWS
- Funciona con:
 - Linux
 - Windows
 - MacOS
 - VMware Cloud en AWS
 - Amazon Workspaces y AppStream 2.0
 - Amazon EC2, ECS y EKS
- Hasta 1.000.000 de IOPS con una latencia de < 0,5 ms
- Snapshots, compresión y bajo coste
- **Clonación instantánea puntual (útil para probar nuevas cargas de trabajo)**



Cloud híbrido para el almacenamiento

- AWS está impulsando la "nube híbrida"
 - Parte de tu infraestructura está en el Cloud
 - Parte de tu infraestructura está en las instalaciones
- Esto puede deberse a:
 - Largas migraciones a el Cloud
 - Requisitos de seguridad
 - Requisitos de normativa
 - Estrategia de IT
- S3 es una tecnología de almacenamiento propia (a diferencia de EFS / NFS), así que ¿cómo expones los datos de S3 en las instalaciones?
- ¡AWS Storage Gateway!

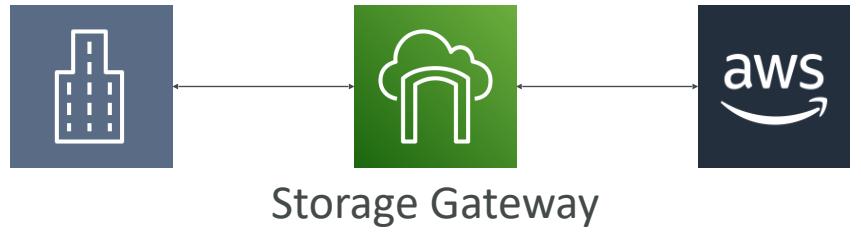
Opciones nativas de la nube de almacenamiento de AWS



AWS Storage Gateway

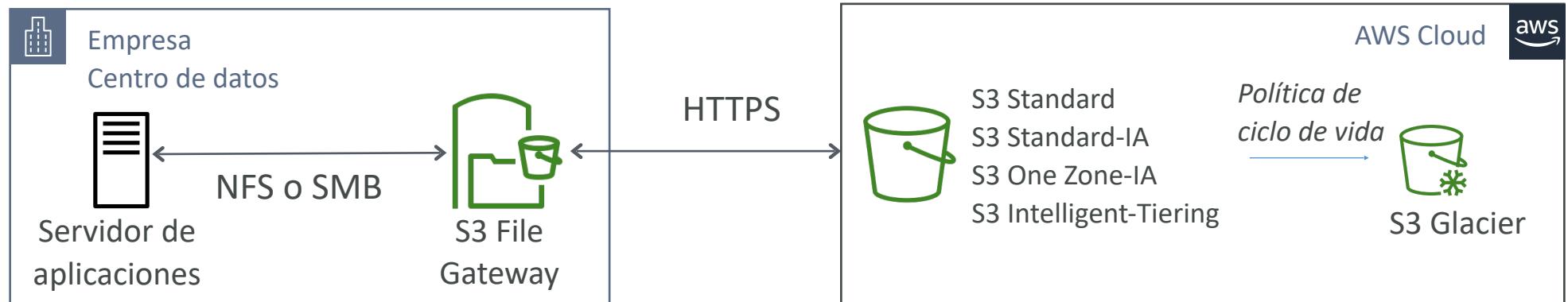


- Puente entre los datos locales y los del Cloud
- **Casos de uso:**
 - recuperación de desastres
 - copia de seguridad y restauración
 - almacenamiento por niveles
 - caché local y acceso a archivos de baja latencia
- Tipos de Gateway de almacenamiento:
 - **S3 File Gateway**
 - **Gateway de archivos FSx**
 - **Gateway de volumen**
 - **Tape Gateway**



Amazon S3 File Gateway (Puerta de enlace de archivo de Amazon S3)

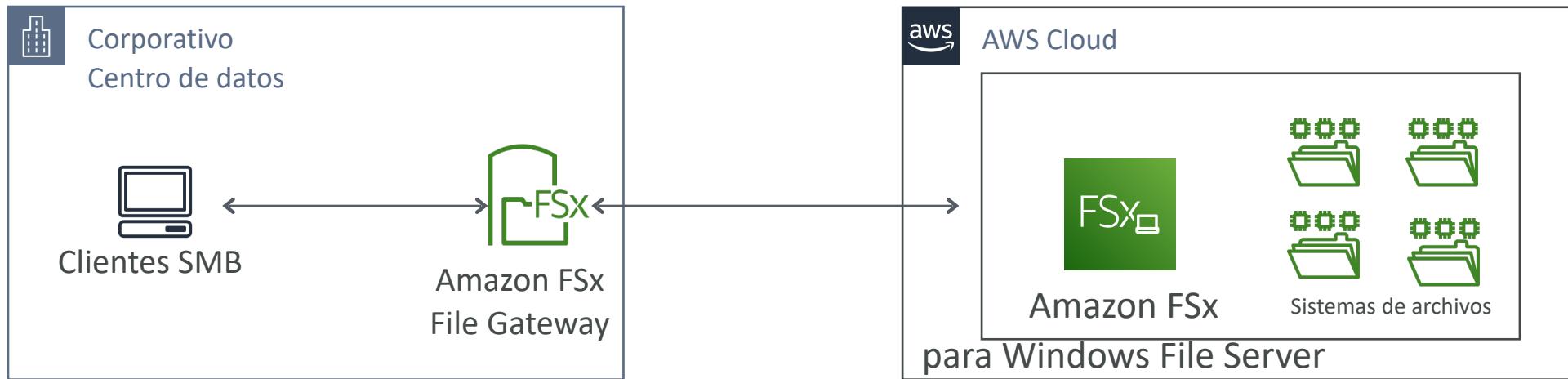
- Los buckets S3 configurados son accesibles mediante el protocolo NFS y SMB
- **Los datos utilizados más recientemente se almacenan en caché en el File Gateway**
- Soporta Estándar S3, Estándar S3 IA, S3 One Zone A, S3 Intelligent Tiering
- **Transición a S3 Glacier mediante una política de ciclo de vida**
- Acceso a buckets mediante roles IAM para cada Gateway de archivos
- El protocolo SMB tiene integración con Active Directory (AD) para la autenticación de usuarios



Amazon FSx File Gateway

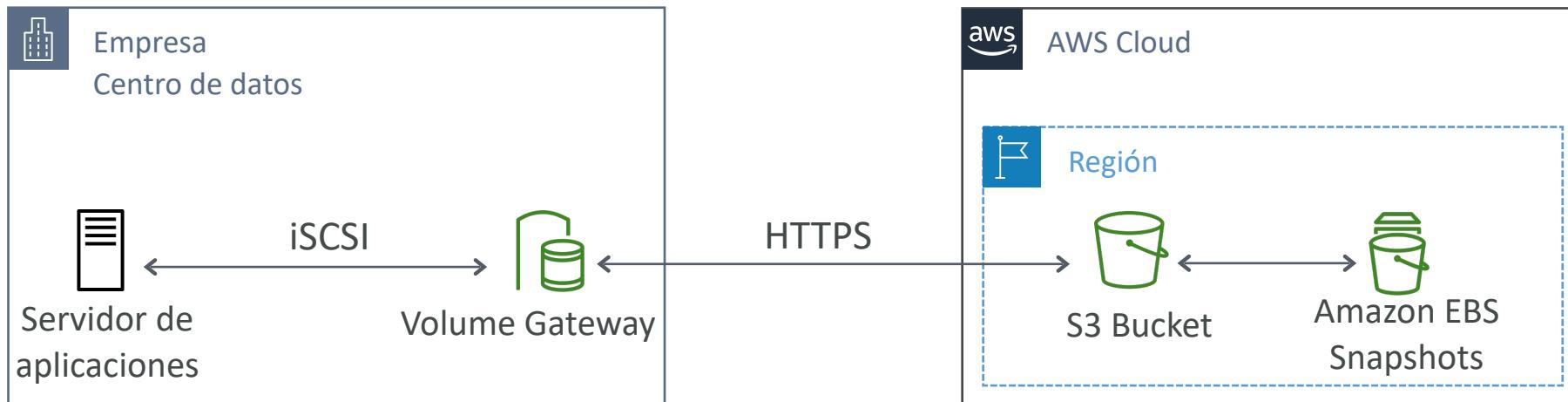
(Puerta de enlace de archivo de Amazon FSx)

- Acceso nativo a Amazon FSx para Windows File Server
- **Caché local para los datos a los que se accede con frecuencia**
- Compatibilidad nativa con Windows (SMB, NTFS, Active Directory...)
- Útil para grupos de archivos compartidos y directorios personales



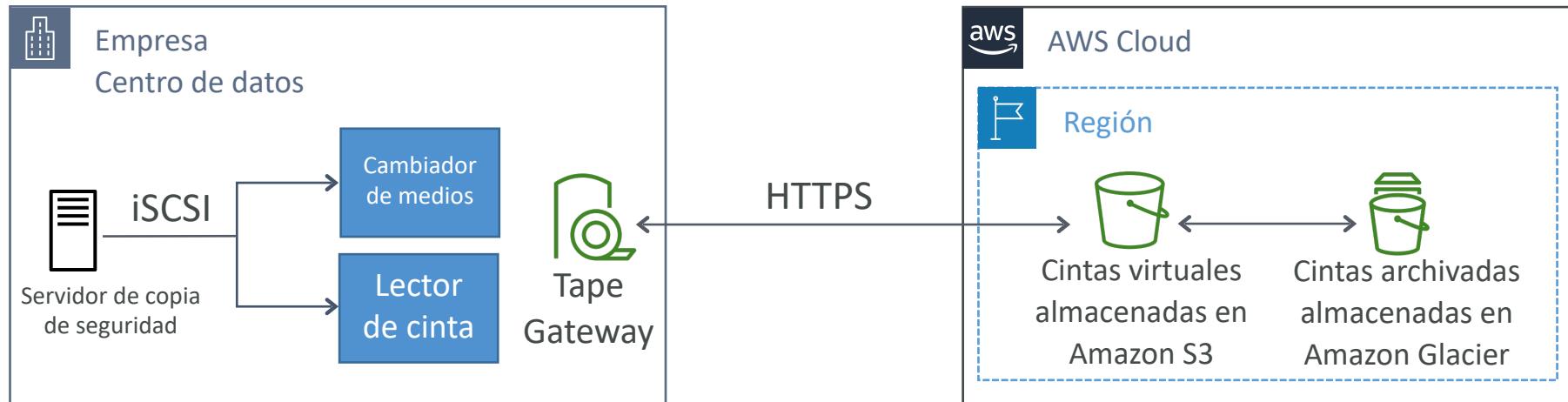
Volume Gateway (Puerta de enlace de volumen)

- Almacenamiento en bloque con protocolo iSCSI respaldado por S3
- Respaldado por Snapshots de EBS que pueden ayudar a restaurar los volúmenes locales
- **Volúmenes en caché**: acceso de baja latencia a los datos más recientes
- **Volúmenes almacenados**: todo el conjunto de datos está en las instalaciones, copias de seguridad programadas en S3



Tape Gateway (Puerta de enlace de cinta)

- Algunas empresas tienen procesos de copia de seguridad que utilizan cintas físicas (!)
- Con Tape Gateway, las empresas utilizan los mismos procesos pero, en el Cloud
- Biblioteca Virtual de Cintas (VTL) respaldada por Amazon S3 y Glacier
- Realiza copias de seguridad de los datos utilizando los procesos existentes basados en cintas (y la interfaz iSCSI)
- Funciona con los principales proveedores de software de copia de seguridad



Storage Gateway - Dispositivo de hardware

- Utilizar Storage Gateway significa que necesitas virtualización in situ.
- Si no, puedes utilizar un **dispositivo de hardware Storage Gateway**
- Puedes comprarlo en amazon.com
- Funciona con File Gateway, Volume Gateway, Tape Gateway
- Tiene los recursos necesarios de CPU, memoria, red y caché SSD
- Útil para copias de seguridad NFS diarias en centros de datos pequeños

Select host platform

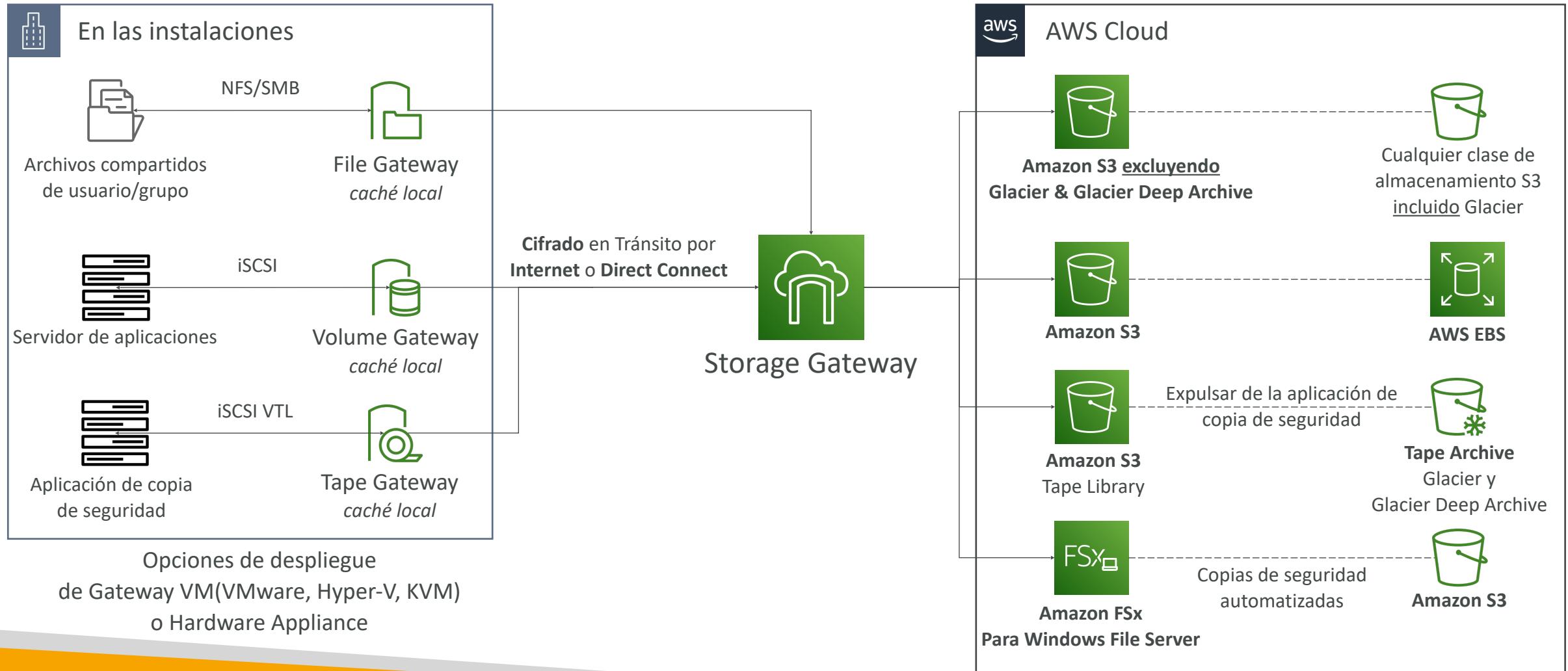
- VMware ESXi
- Microsoft Hyper-V 2012R2/2016
- Linux KVM
- Amazon EC2
- Hardware Appliance

[Buy on Amazon](#)

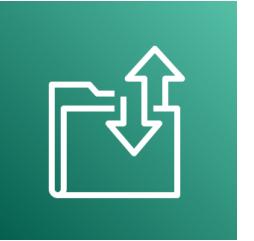
[Activate Appliance](#)



AWS Storage Gateway

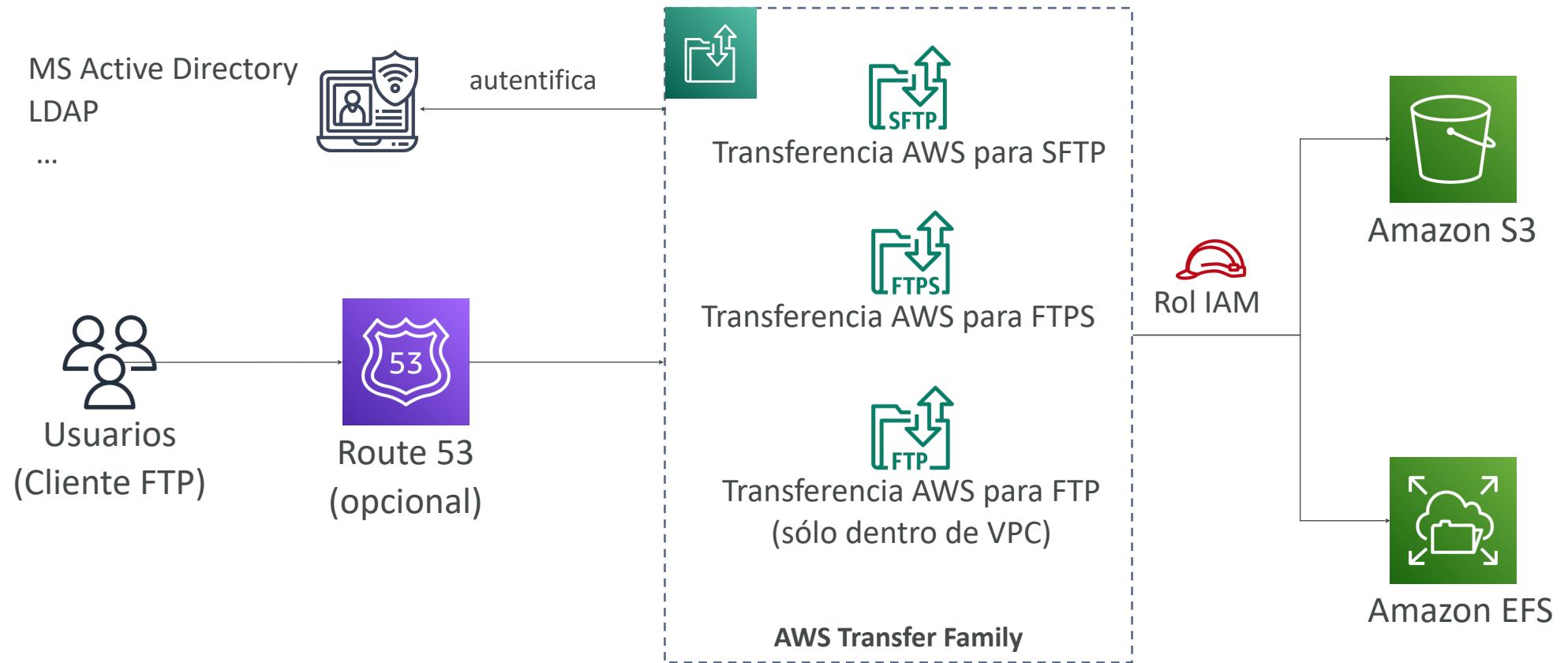


Familia de transferencia AWS



- Un servicio totalmente gestionado para la transferencia de archivos hacia y desde Amazon S3 o Amazon EFS mediante el protocolo FTP
- Protocolos soportados
 - **AWS Transfer para FTP** (Protocolo de Transferencia de Archivos (FTP))
 - **AWS Transfer para FTPS** (Protocolo de Transferencia de Archivos sobre SSL (FTPS))
 - **AWS Transfer para SFTP** (Protocolo de Transferencia de Archivos Seguro (SFTP))
- Infraestructura administrada, escalable, fiable, altamente disponible (multi-AZ)
- Paga por endpoint aprovisionado por hora + transferencias de datos en GB
- Almacena y gestiona las credenciales de los usuarios dentro del servicio
- Se integra con los sistemas de autenticación existentes (Microsoft Active Directory, LDAP, Okta, Amazon Cognito, personalizado)
- Uso: compartir archivos, conjuntos de datos públicos, CRM, ERP, ...

Familia de transferencia AWS (AWS Transfer)



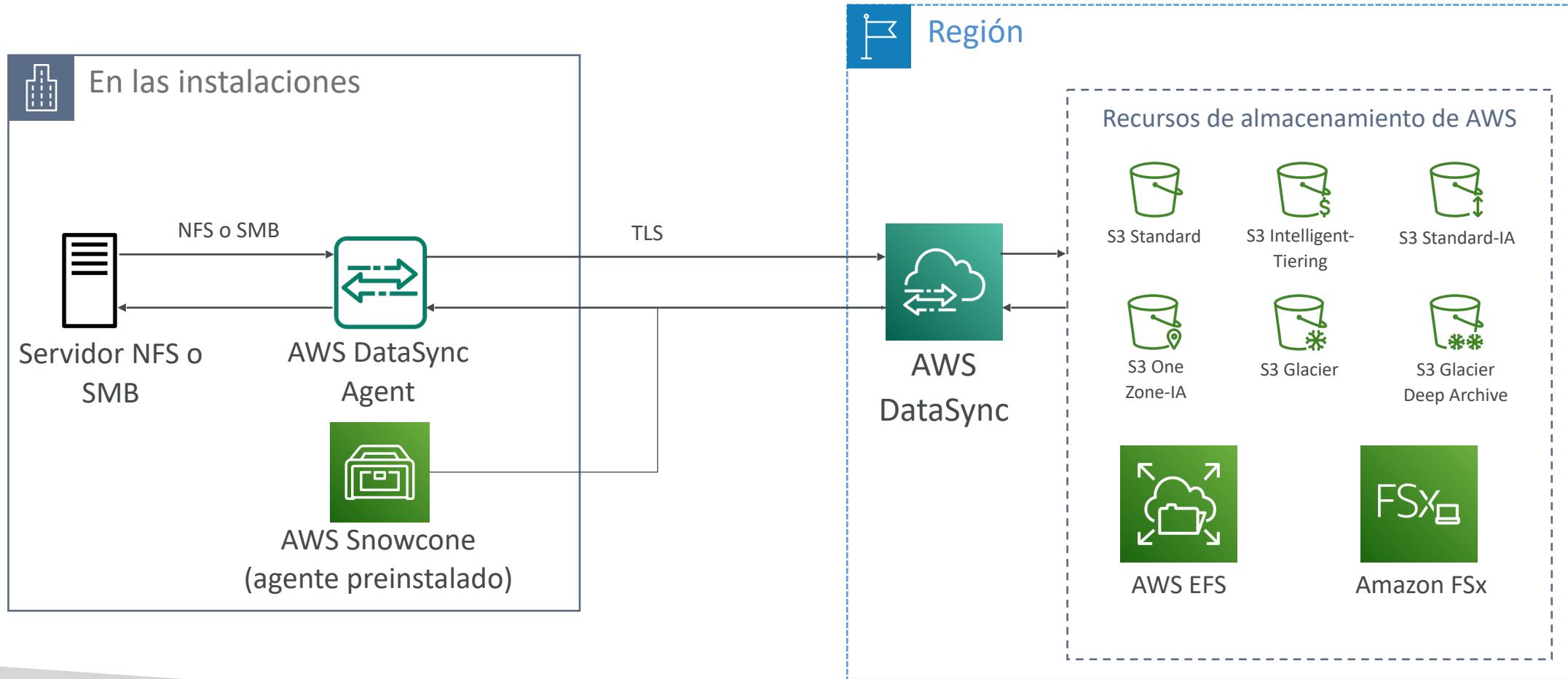
AWS DataSync



- Mover grandes cantidades de datos hacia y desde
 - En las instalaciones / otra nube a AWS (NFS, SMB, HDFS, API S3...) - **necesita agente**
 - De AWS a AWS (diferentes servicios de almacenamiento) - **no necesita agente**
- Puedes sincronizar a:
 - Amazon S3 (cualquier clase de almacenamiento - incluido Glacier)
 - Amazon EFS
 - Amazon FSx (Windows, Lustre, NetApp, OpenZFS...)
- Las tareas de replicación se pueden programar cada hora, cada día, cada semana
- **Se conservan los permisos y metadatos de los archivos** (NFS POSIX, SMB...)
- Una tarea de agente puede utilizar 10 Gbps, se puede configurar un límite de ancho de banda

AWS DataSync

NFS / SMB a AWS (S3, EFS, FSx...)



AWS DataSync

Transferencia entre servicios de almacenamiento de AWS



Comparación de almacenamiento

- **S3:** Almacenamiento de objetos
- **S3 Glacier:** Archivo de objetos
- **Volúmenes EBS:** Almacenamiento en red para una instancia EC2 a la vez
- **Almacenamiento de instancia:** Almacenamiento físico para tu instancia EC2 (altas IOPS)
- **EFS:** Sistema de archivos de red para instancias Linux, sistema de archivos POSIX
- **FSx para Windows:** Sistema de archivos de red para servidores Windows
- **FSx para Lustre:** Sistema de archivos Linux de computación de alto rendimiento
- **FSx para NetApp ONTAP:** Alta compatibilidad con sistemas operativos
- **FSx para OpenZFS:** Sistema de ficheros ZFS gestionado
- **Storage Gateway:** S3 & FSx File Gateway, Volume Gateway (caché y almacenado), Tape Gateway
- **Familia de transferencia:** Interfaz FTP, FTPS, SFTP sobre Amazon S3 o Amazon EFS
- **Sincronización de datos:** Programa la sincronización de datos desde las instalaciones a AWS, o de AWS a AWS
- **Snowcone / Snowball / Snowmobile:** para mover grandes cantidades de datos a el Cloud, físicamente
- **Base de datos:** para cargas de trabajo específicas, normalmente con indexación y consulta

Integración y mensajería de AWS

SQS, SNS & Kinesis

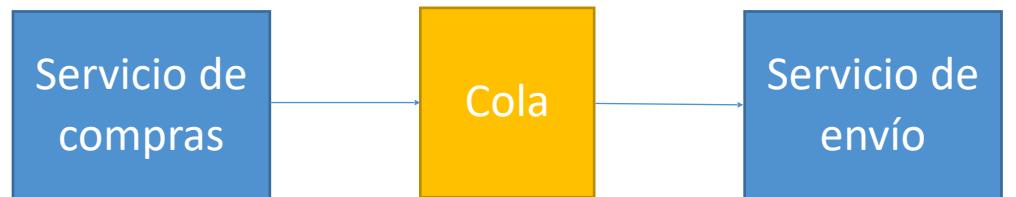
Introducción a la sección

- Cuando empezamos a desplegar varias aplicaciones, es inevitable que tengan que comunicarse entre sí.
- Existen dos patrones de comunicación entre aplicaciones

**1) Comunicaciones sincrónicas
(de aplicación a aplicación)**



**2) Asíncrono / basado en eventos
(aplicación a cola a aplicación)**

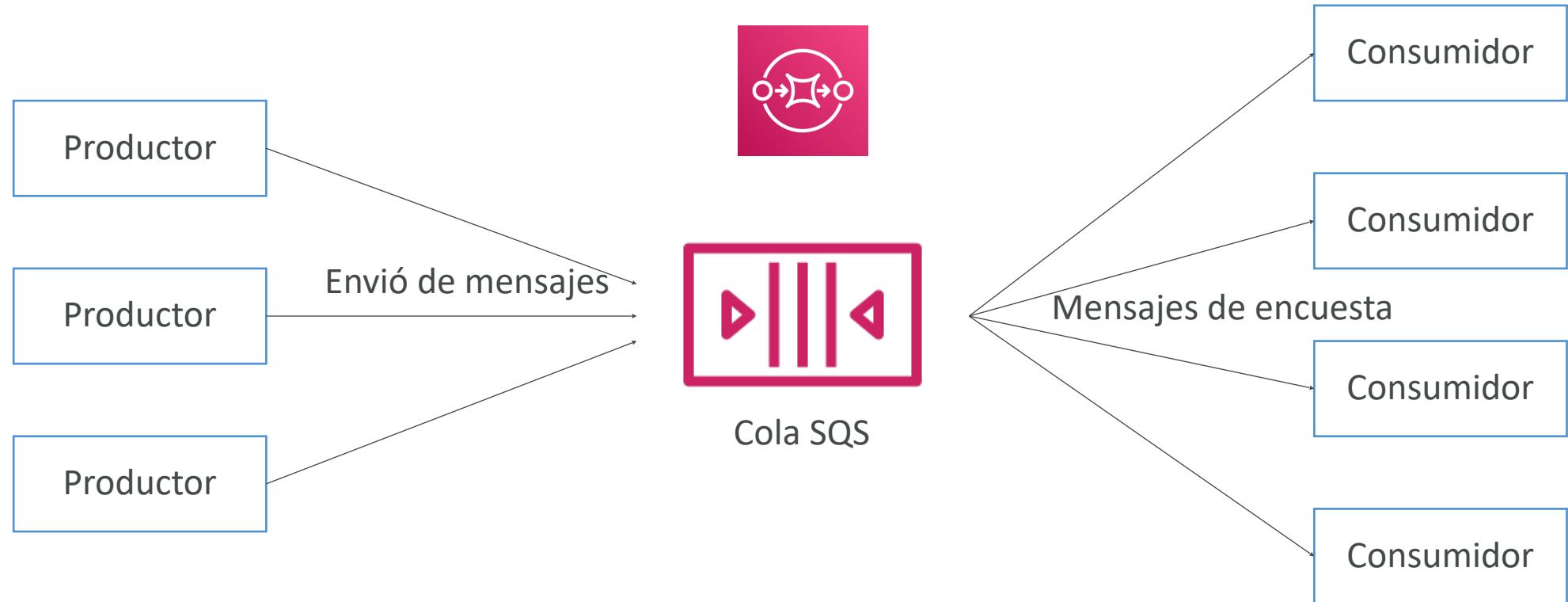


Introducción a la sección

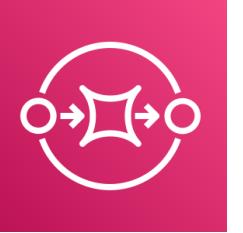
- La sincronización entre aplicaciones puede ser problemática si hay picos repentinos de tráfico
- ¿Qué pasa si de repente necesitas codificar 1000 vídeos pero normalmente son 10?
- En ese caso, es mejor **desacoplar** tus aplicaciones,
 - utilizando SQS: modelo de cola
 - utilizando SNS: modelo pub/sub
 - usando Kinesis: modelo de streaming en tiempo real
- ¡Estos servicios pueden escalar independientemente de nuestra aplicación!

Amazon SQS

¿Qué es una cola?



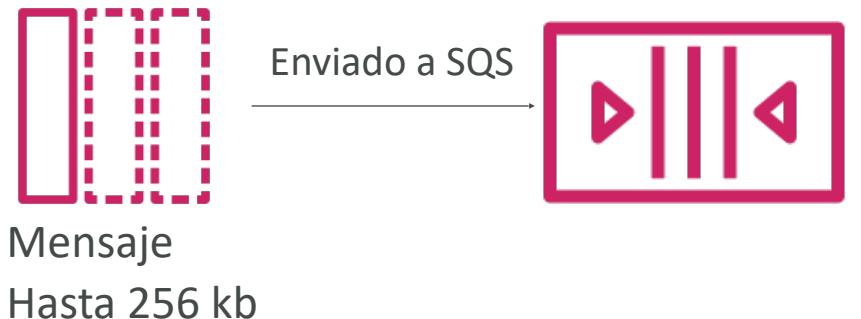
Amazon SQS – Cola estándar



- La oferta más antigua (más de 10 años)
- Servicio totalmente gestionado, utilizado para **desacoplar aplicaciones**
- Atributos:
 - Rendimiento ilimitado, número ilimitado de mensajes en cola
 - Retención de mensajes por defecto 4 días, máximo de 14 días
 - Baja latencia (<10 ms en publicación y recepción)
 - Limitación de 256 KB por mensaje enviado
- Puede haber mensajes duplicados (al menos una entrega, ocasionalmente)
- Puede haber mensajes fuera de orden (orden de mejor esfuerzo)

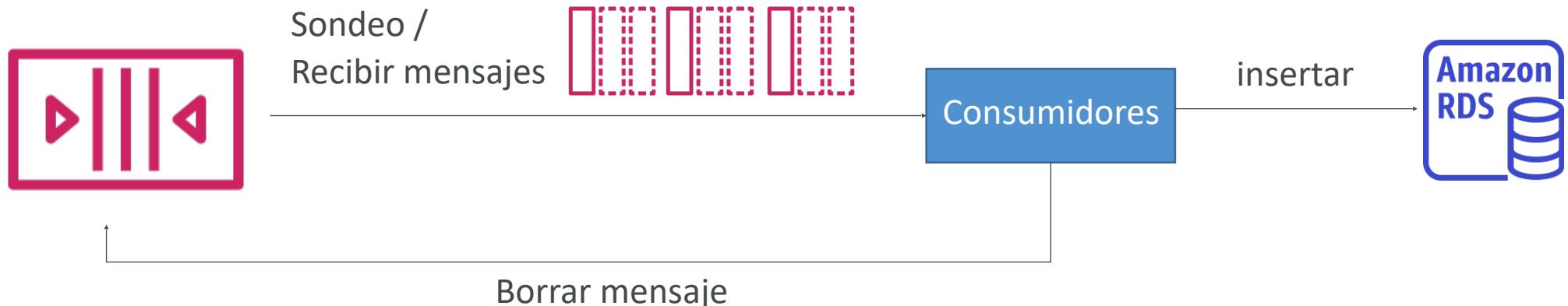
SQS - Producción de mensajes

- Producido a SQS utilizando el SDK (API SendMessage)
- El mensaje se **conserva** en SQS hasta que un consumidor lo elimina
- Retención del mensaje: por defecto 4 días, hasta 14 días
- Ejemplo: enviar un pedido para ser procesado
 - Id de pedido
 - Id de cliente
 - Los atributos que quieras
- SQS estándar: rendimiento ilimitado



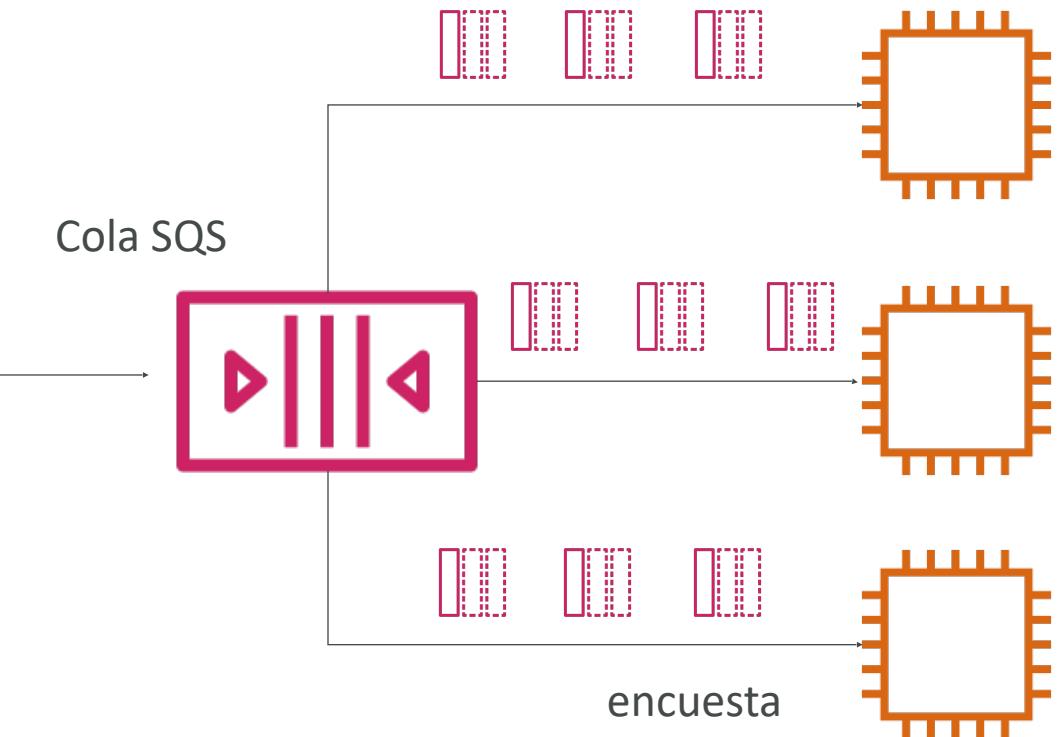
SQS – Consumir mensajes

- Consumidores (ejecutándose en instancias EC2, servidores o AWS Lambda)...
- Sondeo (encuestas) SQS en busca de mensajes (recibir hasta 10 mensajes a la vez)
- Procesar los mensajes (ejemplo: insertar el mensaje en una base de datos RDS)
- Eliminar los mensajes utilizando la API DeleteMessage



SQS

Consumidores de múltiples instancias EC2

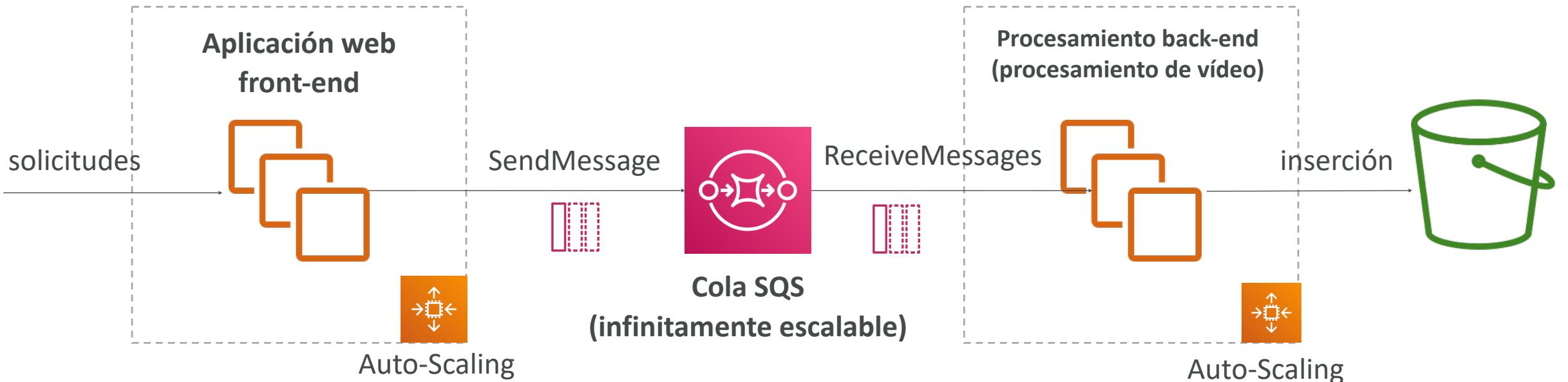


- Los consumidores reciben y procesan los mensajes en paralelo
- Al menos una entrega
- Ordenación de mensajes al mejor esfuerzo
- Los consumidores borran los mensajes después de procesarlos
- Podemos escalar los consumidores horizontalmente para mejorar el rendimiento del procesamiento

SQS con Auto Scaling Group (ASG)



SQS para **desacoplar** los niveles de aplicación

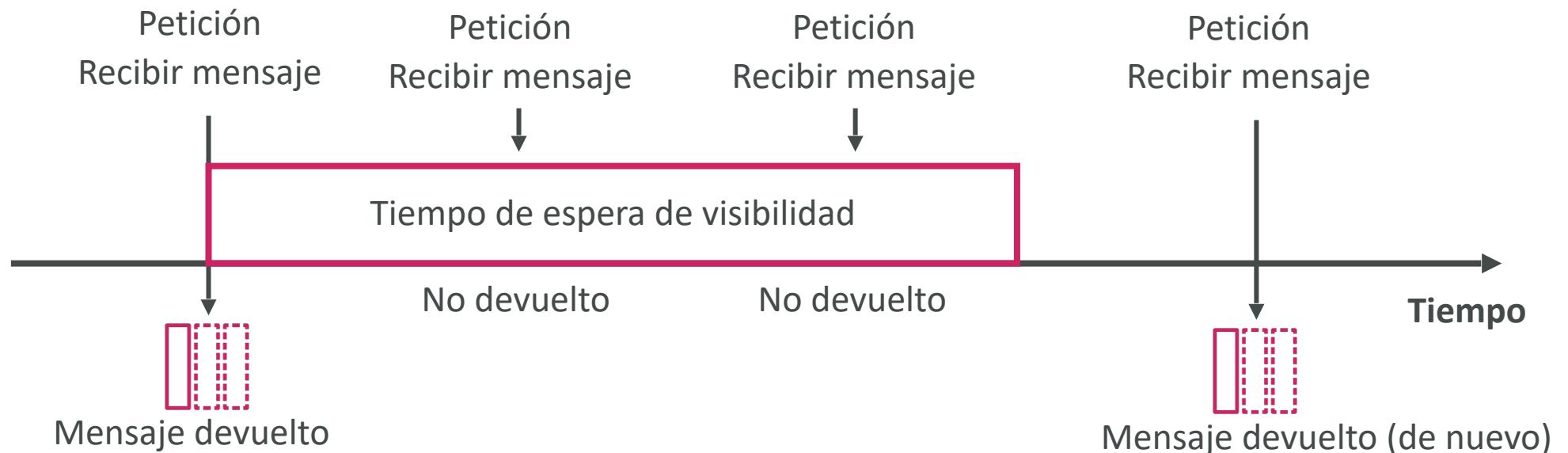


Amazon SQS - Seguridad

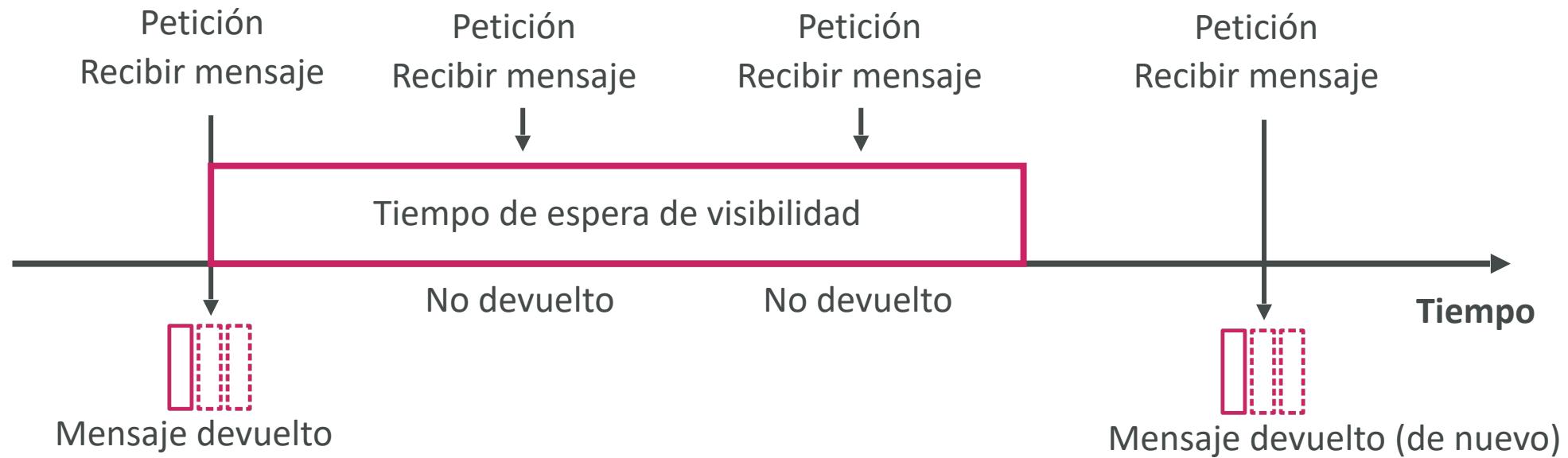
- **Cifrado:**
 - Cifrado en vuelo mediante API HTTPS
 - Cifrado en reposo mediante claves KMS
 - Cifrado del lado del cliente si el cliente desea realizar el cifrado/descifrado por sí mismo
- **Controles de acceso:** Políticas IAM para regular el acceso a la API SQS
- **Políticas de acceso a SQS** (similares a las políticas de bucket de S3)
 - Útil para el acceso entre cuentas a las colas SQS
 - Útil para permitir a otros servicios (SNS, S3...) escribir en una cola SQS

SQS –Tiempo de espera de visibilidad de mensajes

- Después de que un consumidor sondee un mensaje, éste se vuelve **invisible** para los demás consumidores
- Por defecto, el "tiempo de visibilidad del mensaje" **es de 30 segundos**
- Esto significa que el mensaje tiene 30 segundos para ser procesado
- Una vez transcurrido el tiempo de espera, el mensaje es "visible" en SQS.



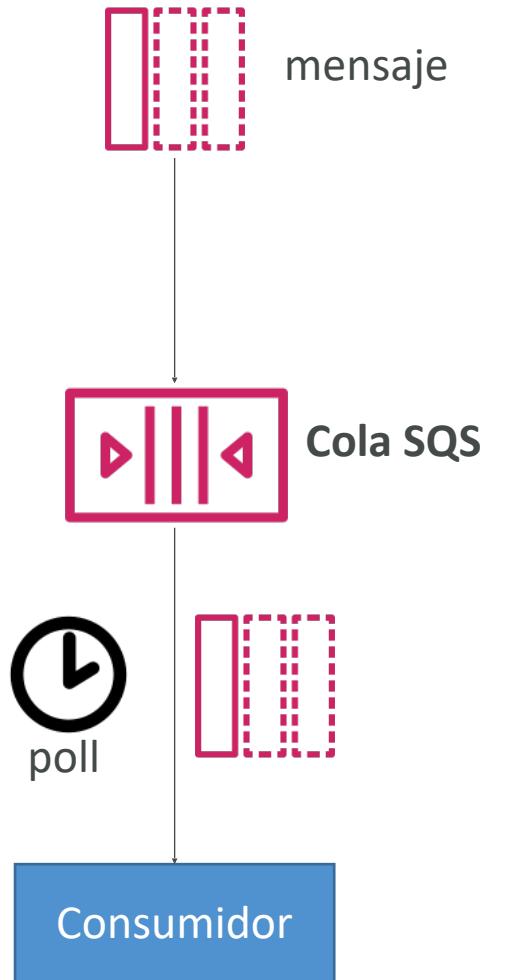
SQS – Tiempo de espera de visibilidad de mensajes



- Si un mensaje no se procesa dentro del tiempo de visibilidad, se procesará **dos veces**.
- El consumidor puede llamar a la API **ChangeMessageVisibility** para obtener más tiempo
- Si el tiempo de espera de visibilidad es alto (horas) y el consumidor se bloquea, el reprocesamiento llevará tiempo.
- Si el tiempo de visibilidad es demasiado bajo (segundos), puede haber duplicados.

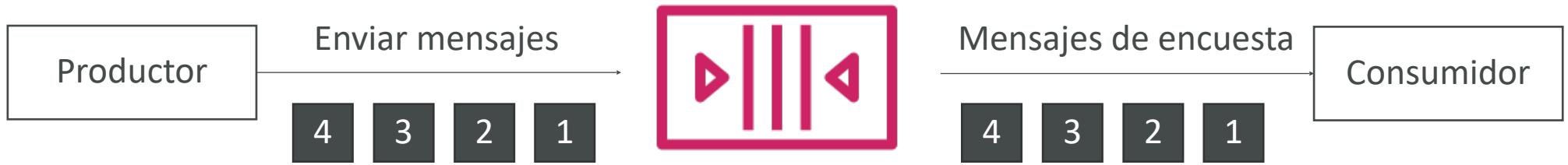
Amazon SQS - Sondeo largo (Long Polling)

- Cuando un consumidor solicita mensajes de la cola, puede opcionalmente "esperar" a que lleguen los mensajes si no hay ninguno en la cola
- Esto se llama Sondeo Largo
- **LongPolling disminuye el número de llamadas API realizadas a SQS, al tiempo que aumenta la eficiencia y reduce la latencia de tu aplicación**
- El tiempo de espera puede oscilar entre 1 y 20 segundos (preferiblemente 20 segundos)
- El sondeo largo es preferible al sondeo corto
- El sondeo largo puede activarse a nivel de cola o a nivel de API utilizando **WaitTimeSeconds**



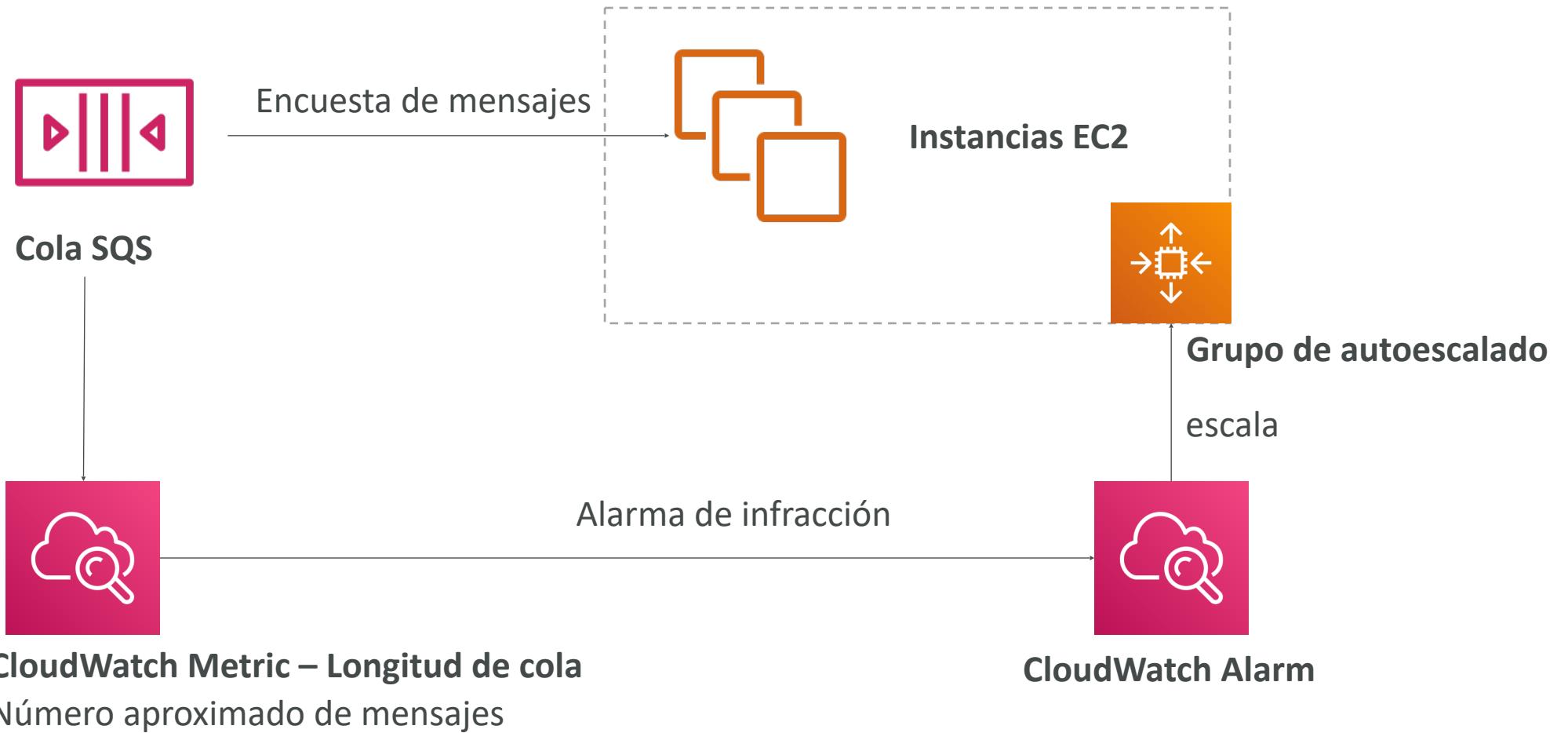
Amazon SQS - Cola FIFO

- FIFO = First In First Out (ordenación de los mensajes en la cola)

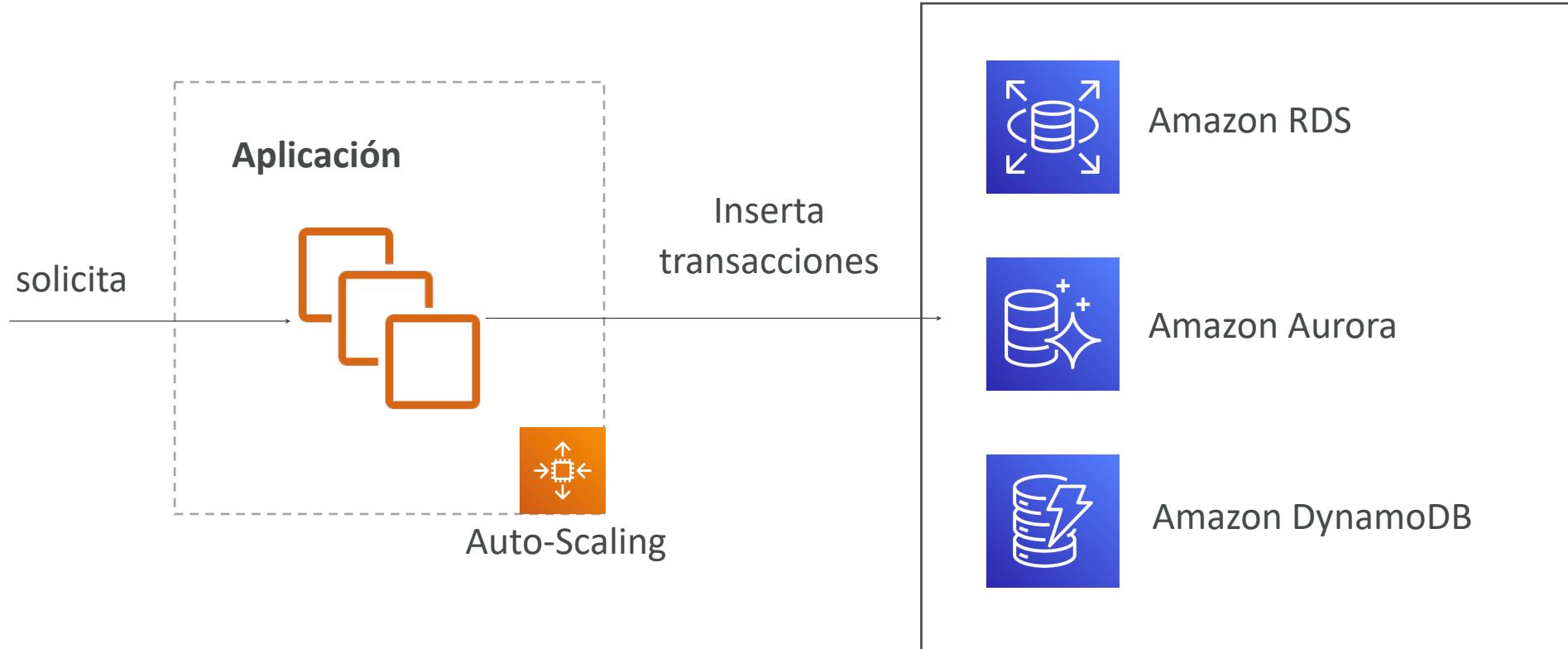


- Rendimiento limitado: 300 msg/s sin procesamiento por lotes, 3000 msg/s con procesamiento por lotes
- Capacidad de envío exactamente una vez (eliminando los duplicados)
- El consumidor procesa los mensajes en orden

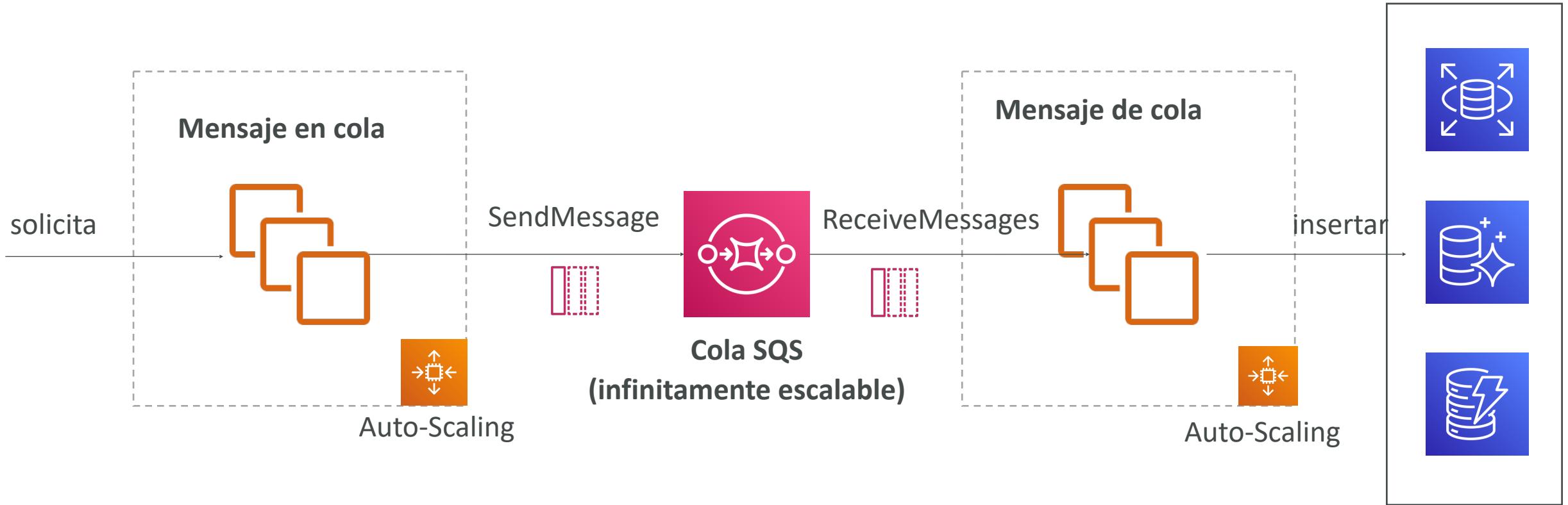
SQS con Auto Scaling Group (ASG)



Si la carga es demasiado grande, pueden perderse algunas transacciones



SQS como buffer de escrituras en la base de datos



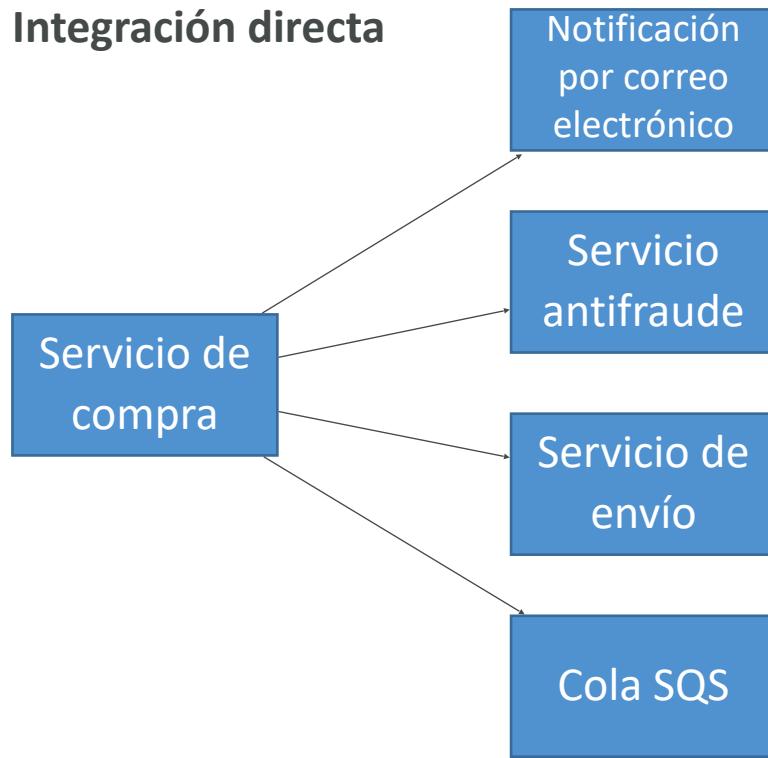
SQS para **desacoplar** los niveles de aplicación



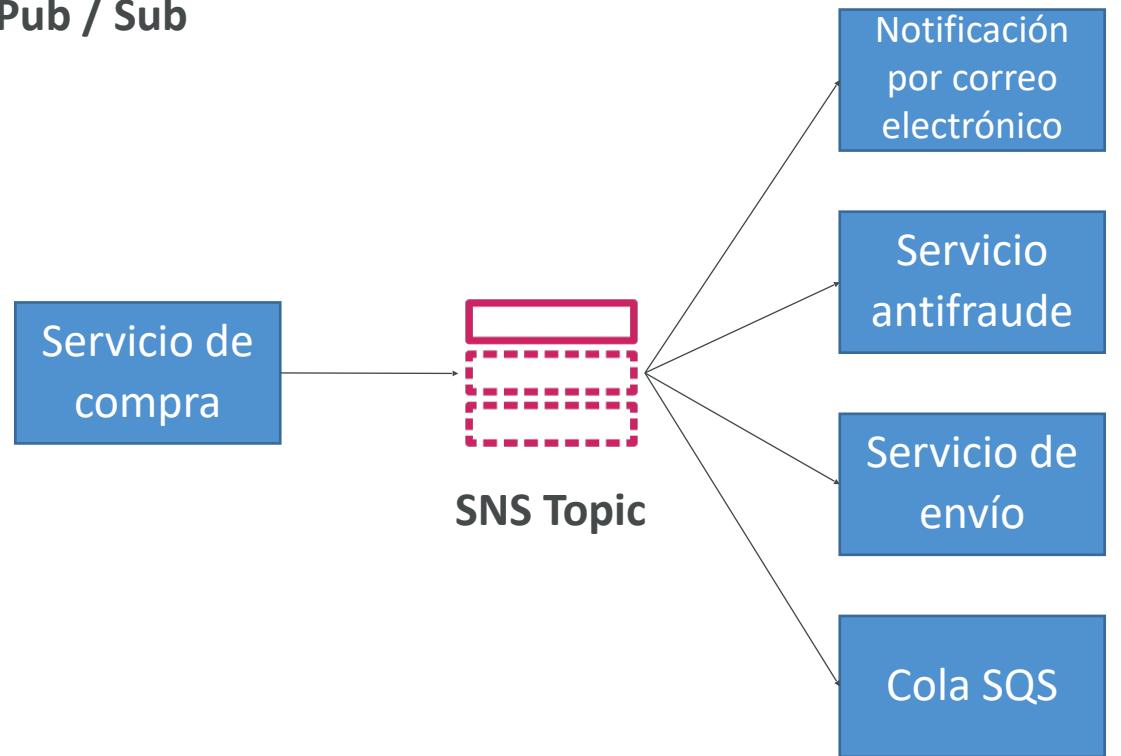
Amazon SNS

- ¿Y si quieres enviar un mensaje a muchos destinatarios?

Integración directa



Pub / Sub



Amazon SNS



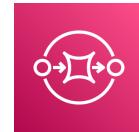
- El "productor de eventos" sólo envía mensajes a un tema SNS
- Tantos "receptores de eventos" (suscriptores) como queramos para escuchar las notificaciones del tema SNS
- Cada suscriptor al tema recibirá todos los mensajes (nota: nueva función para filtrar mensajes)
- Hasta 12.500.000 suscripciones por tema
- Límite de 100.000 temas



SNS

publicar

Suscriptores



SQS



Lambda



Kinesis Data
Firehose



Emails



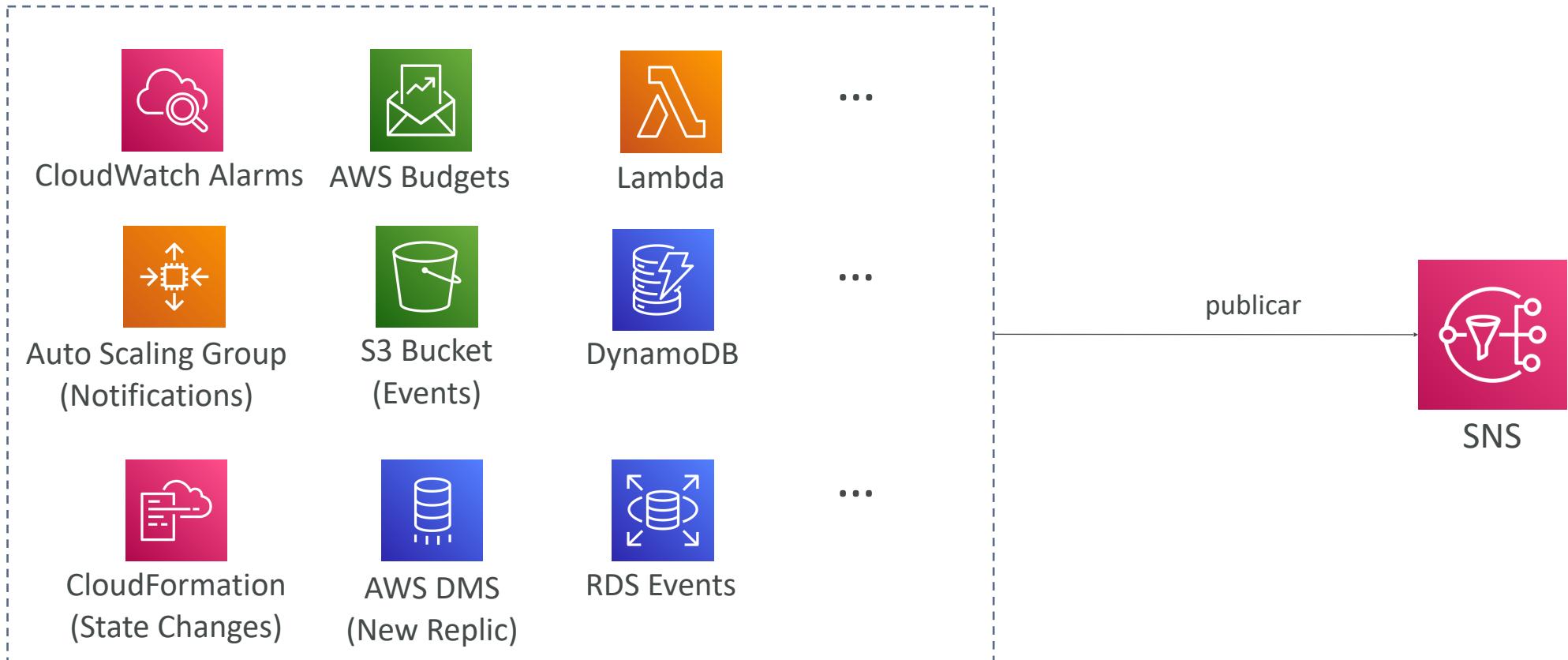
SMS y
notificaciones por móvil



Endpoints

SNS se integra con muchos servicios de AWS

- Muchos servicios de AWS pueden enviar datos directamente a SNS para notificaciones



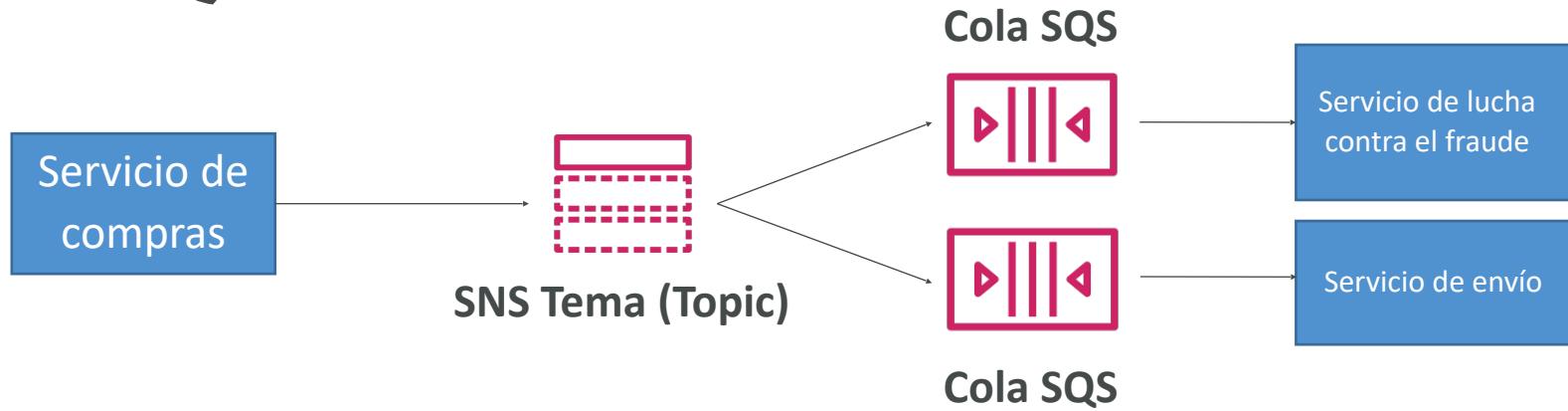
Amazon SNS - Cómo publicar

- Publicación de temas (mediante el SDK)
 - Crear un tema
 - Crea una suscripción (o varias)
 - Publicar en el tema
- Publicación directa (para aplicaciones móviles SDK)
 - Crear una aplicación de plataforma
 - Crear un punto final de plataforma
 - Publicar en el punto final de la plataforma
 - Funciona con Google GCM, Apple APNS, Amazon ADM...

Amazon SNS – Seguridad

- **Cifrado:**
 - Cifrado en vuelo mediante API HTTPS
 - Cifrado en reposo mediante claves KMS
 - Cifrado del lado del cliente si el cliente desea realizar el cifrado/descifrado por sí mismo
- **Controles de acceso:** Políticas IAM para regular el acceso a la API SNS
- **Políticas de acceso SNS** (similares a las políticas de bucket S3)
 - Útil para el acceso entre cuentas a temas SNS
 - Útil para permitir que otros servicios (S3...) escriban en un tema SNS

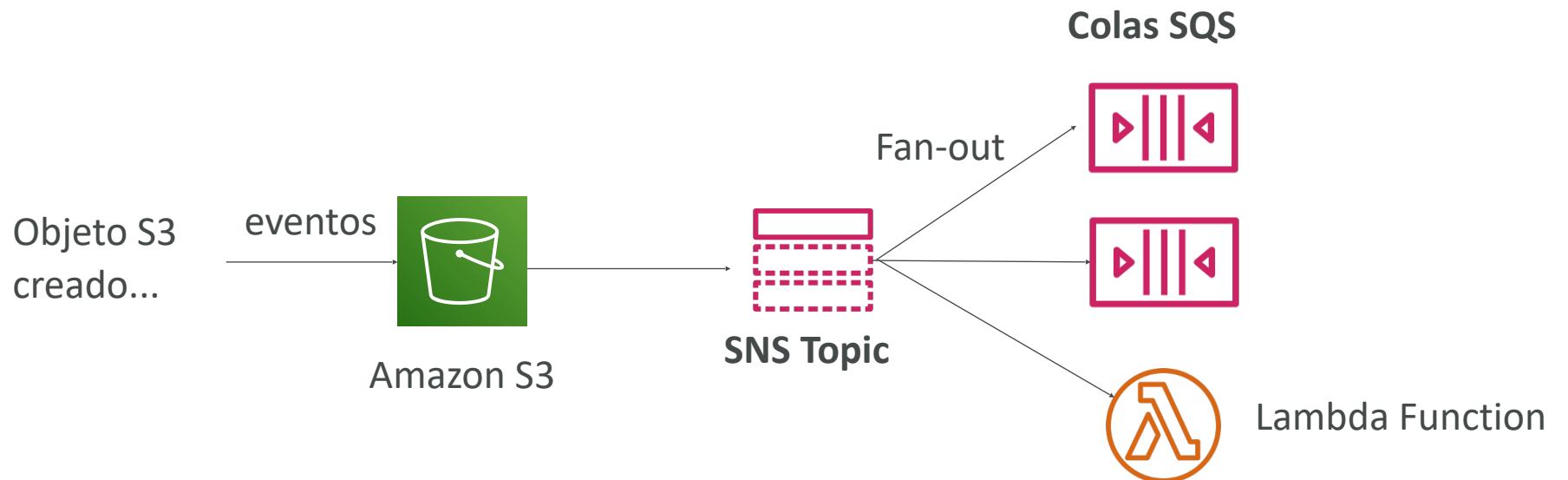
SNS + SQS: Fan Out



- Haz el push una vez en SNS, recibe en todas las colas SQS que son suscriptores
- Totalmente desacoplado, sin pérdida de datos
- SQS permite: persistencia de datos, procesamiento diferido y reintentos de trabajo
- Posibilidad de añadir más suscriptores SQS con el tiempo
- Asegúrate de que la **política de acceso** a la cola SQS permite que SNS pueda escribir

Aplicación: Eventos S3 a múltiples colas

- Para la misma combinación de: **tipo de evento** (p.e. creación de objeto) y **prefijo** (p.e. imágenes/) sólo puedes tener una regla de Evento S3
- Si quieres enviar el mismo evento S3 a muchas colas SQS, utiliza fan-out



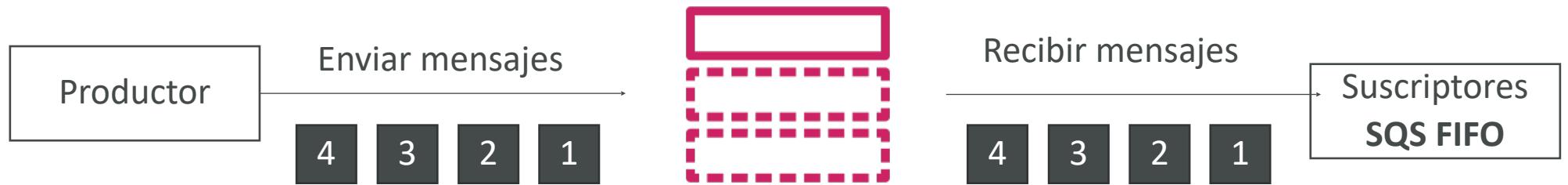
Aplicación: SNS a Amazon S3 a través de Kinesis Data Firehose

- SNS puede enviar a Kinesis y por lo tanto podemos tener la siguiente arquitectura de soluciones:



Amazon SNS - Tema (topic) FIFO

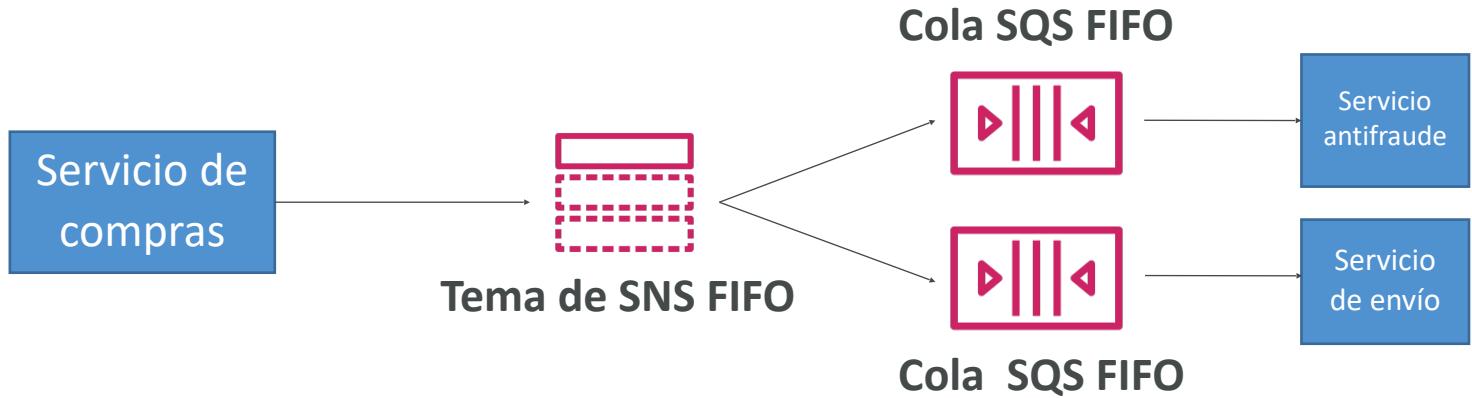
- FIFO = First In First Out (orden de los mensajes en el tema)



- Características similares a SQS FIFO:
 - **Ordenación** por ID de grupo de mensajes (se ordenan todos los mensajes del mismo grupo)
 - **Deduplicación** mediante ID de deduplicación o deduplicación basada en contenido
- Sólo puede tener colas SQS FIFO como suscriptores
- Rendimiento limitado (el mismo que SQS FIFO)

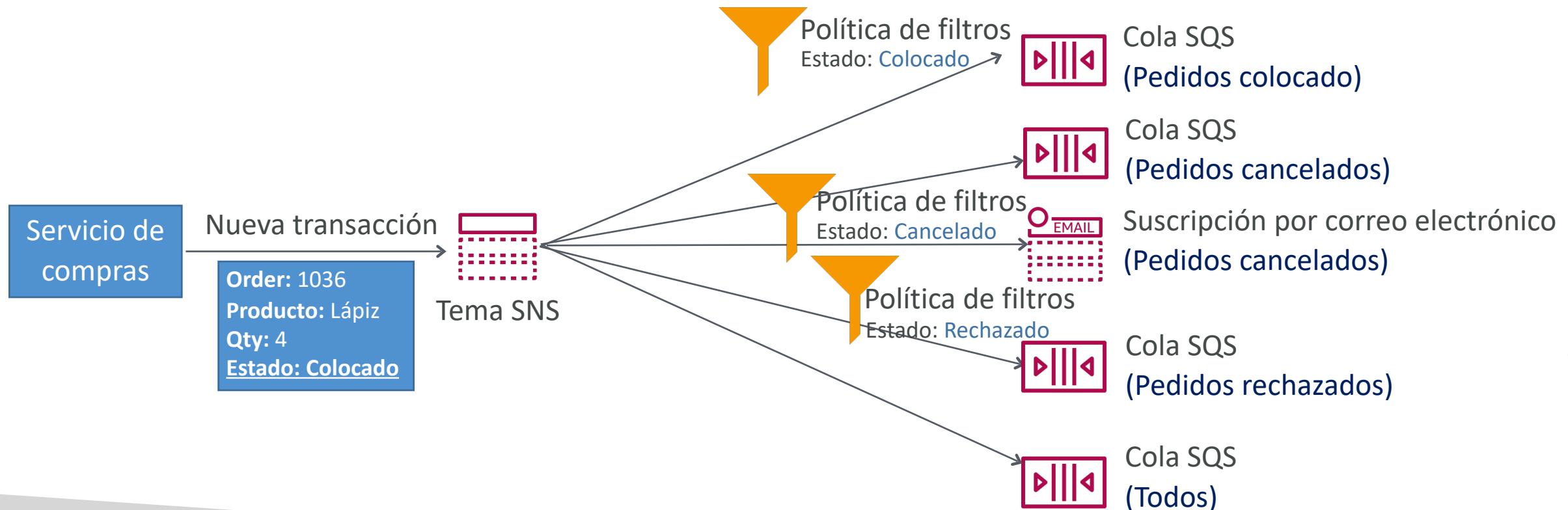
SNS FIFO + SQS FIFO: Fan Out

- En caso de que necesites fan-out + ordenación + deduplicación



SNS - Filtrado de mensajes

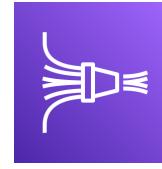
- Política JSON utilizada para filtrar los mensajes enviados a las suscripciones del tema SNS
- Si una suscripción no tiene una política de filtrado, recibe todos los mensajes



Visión general de Kinesis

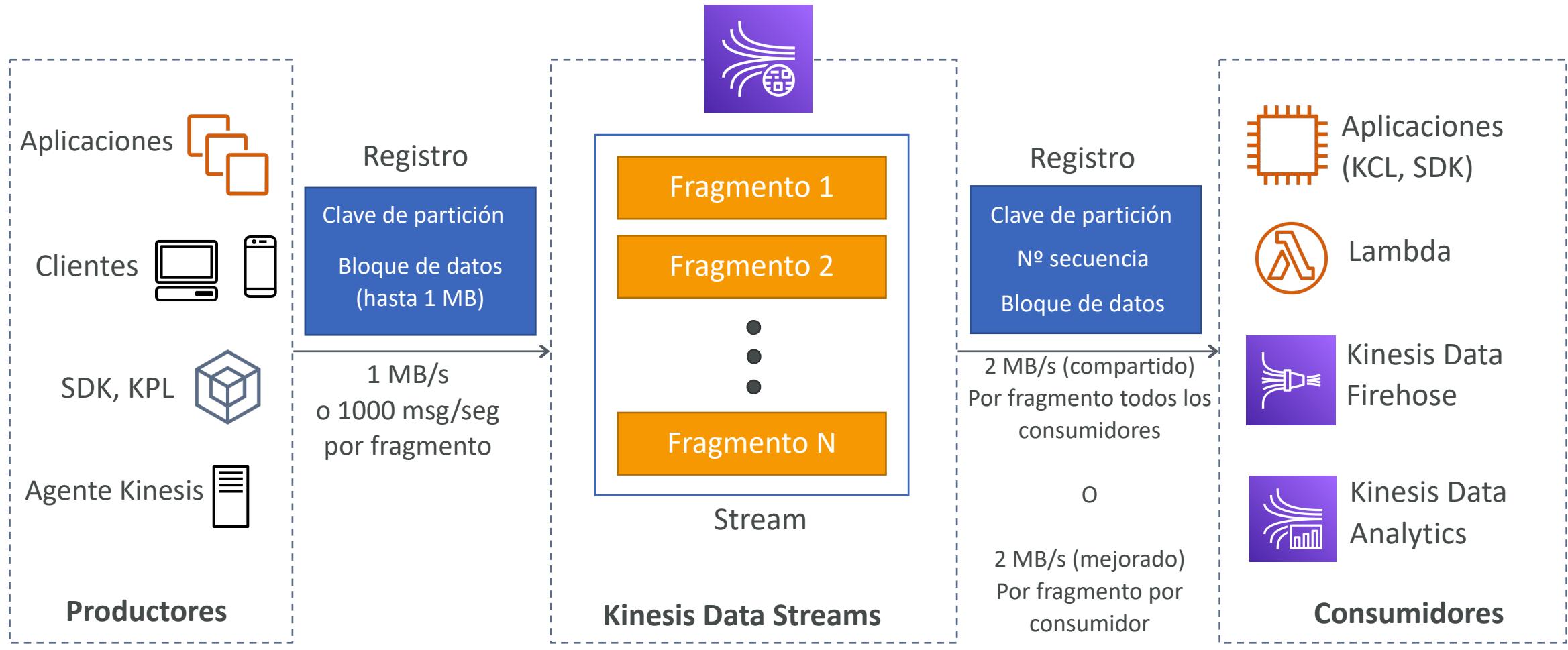


- Facilita la **recopilación**, el **procesamiento** y el **análisis** de datos de flujo continuo en tiempo real
- Ingesta de datos en tiempo real como: Registros de aplicaciones, métricas, secuencias de clics de sitios web, datos telemétricos de IoT...



- **Kinesis Data Streams**: captura, procesa y almacena flujos de datos
- **Kinesis Data Firehose**: carga flujos de datos en almacenes de datos de AWS
- **Kinesis Data Analytics**: analiza flujos de datos con SQL o Apache Flink
- **Kinesis Video Streams**: captura, procesa y almacena transmisiones de vídeo

Kinesis Data Streams





Kinesis Data Streams

- Retención entre 1 día y 365 días
- Posibilidad de volver a procesar (reproducir) los datos
- Una vez que los datos se insertan en Kinesis, no pueden borrarse (inmutabilidad)
- Los datos que comparten la misma partición van al mismo fragmento (ordenación)
- Productores: SDK de AWS, biblioteca de productores de Kinesis (KPL), agente de Kinesis
- Consumidores:
 - Escriba el suyo propio: Kinesis Client Library (KCL), AWS SDK
 - Administrados: AWS Lambda, Kinesis Data Firehose, Kinesis Data Analytics,

Kinesis Data Streams - Modos de capacidad

- **Modo aprovisionado:**

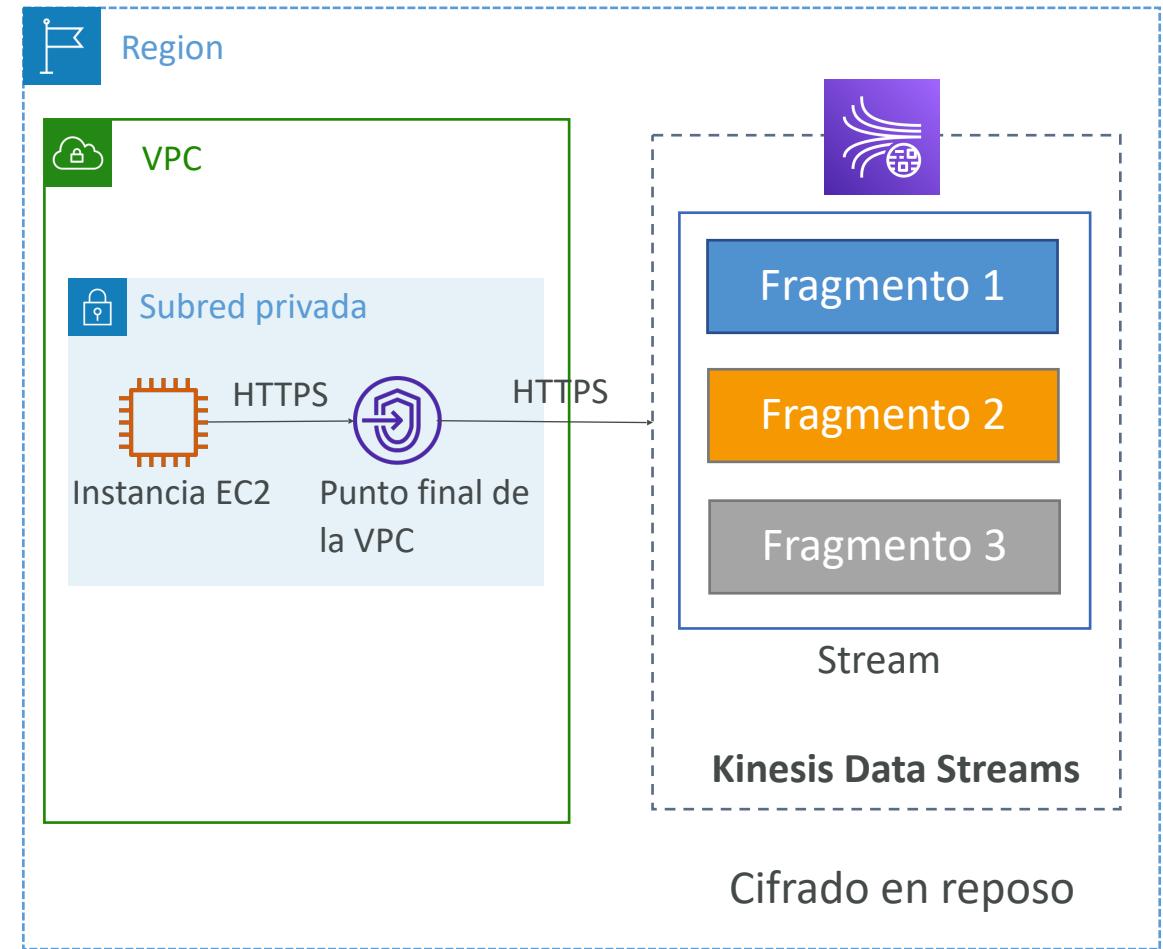
- Tú eliges el número de shards aprovisionados, escala manualmente o usando API
- Cada fragmento recibe 1 MB/s (o 1000 registros por segundo)
- Cada fragmento recibe 2 MB/s de salida (consumo en fan-out clásico o mejorado)
- Se paga por cada fragmento aprovisionado por hora

- **Modo bajo demanda:**

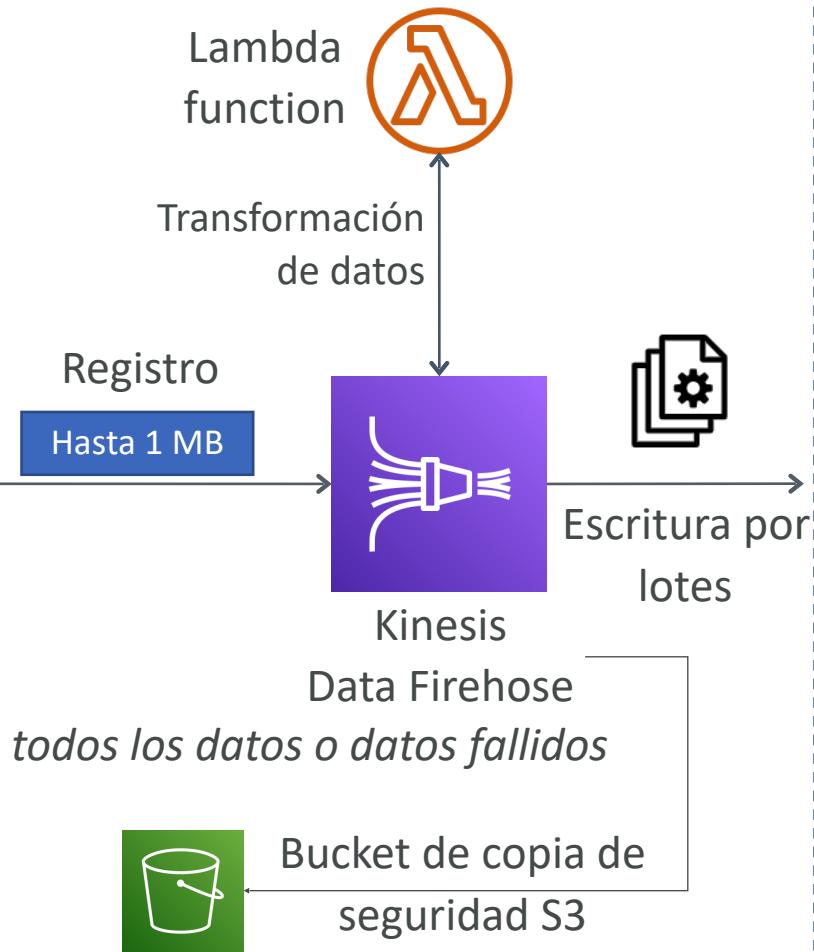
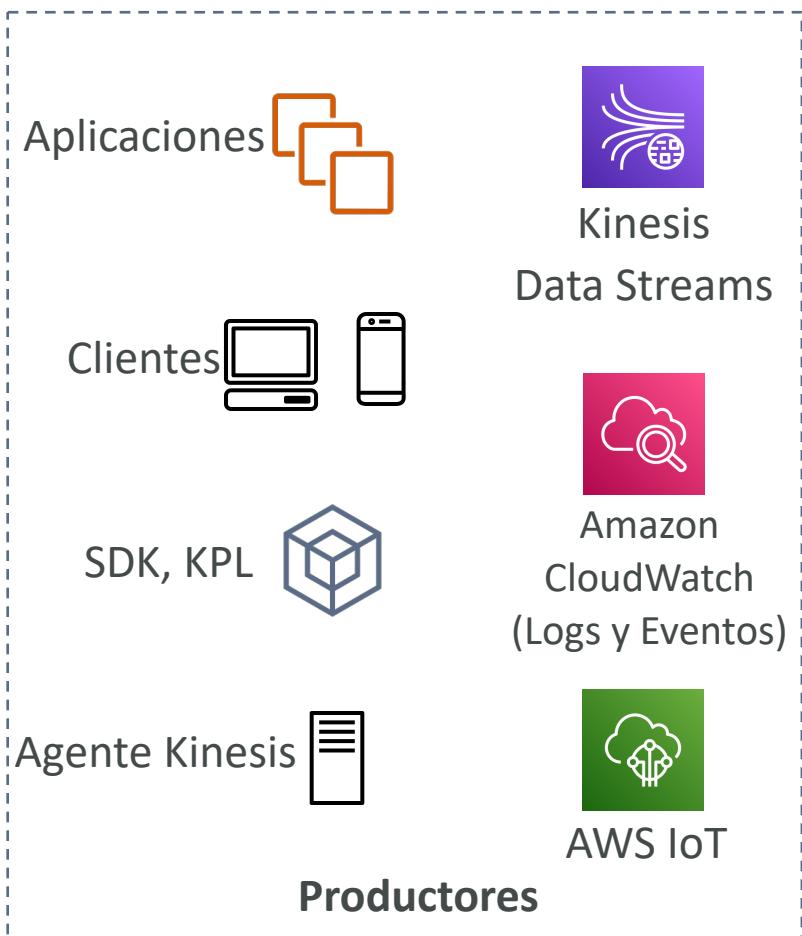
- No es necesario aprovisionar ni gestionar la capacidad
- Capacidad provisionada por defecto (4 MB/s de entrada o 4000 registros por segundo)
- Escala automáticamente en función del pico de rendimiento observado durante los últimos 30 días
- Pago por flujo por hora y entrada/salida de datos por GB

Seguridad de Kinesis Data Streams

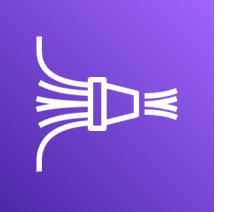
- Control de acceso / autorización mediante políticas IAM
- Cifrado en vuelo mediante puntos finales HTTPS
- Cifrado en reposo mediante KMS
- Puede implementar el cifrado/ descifrado de datos en el lado del cliente (más difícil)
- Puntos finales VPC disponibles para que Kinesis acceda dentro de VPC
- Supervisión de las llamadas a la API mediante CloudTrail



Kinesis Data Firehose



Kinesis Data Firehose



- Servicio totalmente administrado, sin administración, escalado automático, sin servidor
 - AWS: Redshift / Amazon S3 / ElasticSearch
 - Socio de terceros: Splunk / MongoDB / DataDog / NewRelic / ...
 - Personalizado: enviar a cualquier punto final HTTP
- Pague por los datos que pasan por Firehose
- **Casi en tiempo real**
 - Latencia mínima de 60 segundos para lotes no completos
 - Un mínimo de 1 MB de datos a la vez
- Admite muchos formatos de datos, conversiones, transformaciones y compresión
- Admite transformaciones de datos personalizadas mediante AWS Lambda
- Puede enviar datos fallidos o todos los datos a un bucket de S3 de backup

Kinesis Data Streams vs Kinesis Data Firehose



Kinesis Data Streams

- Servicio de streaming para la ingesta a escala
- Escribir código personalizado (productor / consumidor)
- En tiempo real (~200 ms)
- Gestión del escalado (división/fusión de fragmentos)
- Almacenamiento de datos de 1 a 365 días
- Capacidad de reproducción

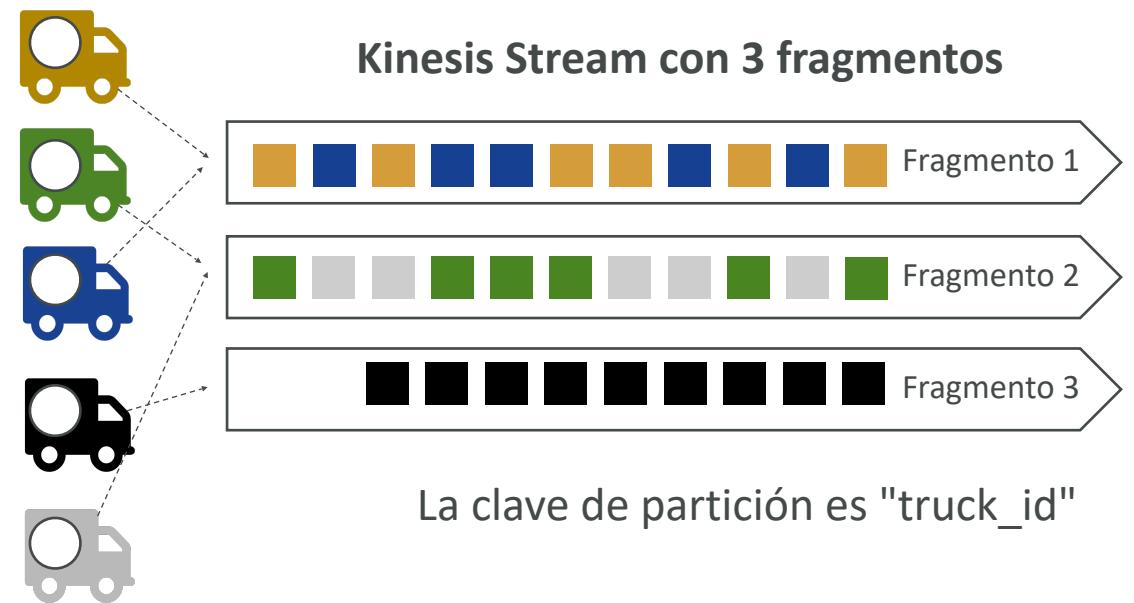


Kinesis Data Firehose

- Carga de datos de streaming en S3 / Redshift / ES / terceros / HTTP personalizado
- Totalmente gestionado
- Casi en tiempo real (tiempo de buffer min. 60 seg)
- Escalado automático
- Sin almacenamiento de datos
- No soporta capacidad de repetición

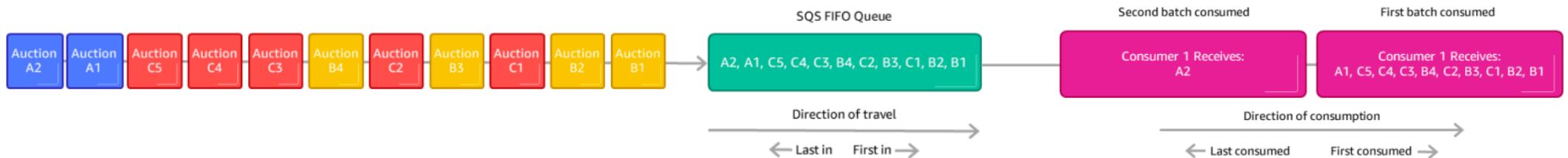
Ordenar datos en Kinesis

- Imagina que tienes 100 camiones (camión_1, camión_2, ... camión_100) en la carretera enviando sus posiciones GPS regularmente a AWS.
 - Se desea consumir los datos en orden para cada camión, de modo que pueda seguir su movimiento con precisión.
 - ¿Cómo debe enviar esos datos a Kinesis?
-
- **Respuesta:** enviar utilizando un valor de "Partition Key" del "truck_id" ("camion_id").
 - **La misma clave irá siempre al mismo fragmento**

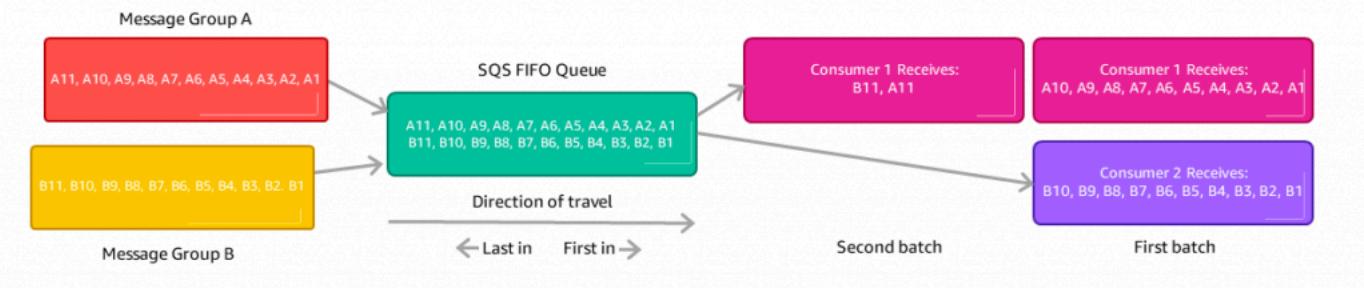


Ordenar datos en SQS

- Para SQS estándar, no hay ordenación.
- Para SQS FIFO, si no se utiliza un ID de grupo, los mensajes se consumen en el orden en que se envían, **con un solo consumidor**



- Deseas escalar el número de consumidores, pero quieres que los mensajes estén "agrupados" cuando se relacionan entre sí
- En ese caso, puedes utilizar un ID de grupo (similar a la clave de partición en Kinesis).



Ordenación Kinesis vs SQS

- **Supongamos 100 camiones, 5 fragmentos kinesis, 1 SQS FIFO**
- Kinesis Data Streams:
 - Por término medio, tendrás 20 camiones por fragmento.
 - Los camiones tendrán sus datos ordenados dentro de cada fragmento
 - La cantidad máxima de consumidores en paralelo que podemos tener es 5
 - Puede recibir hasta 5 MB/s de datos
- SQS FIFO
 - Sólo tienes una cola SQS FIFO
 - Tendrás 100 ID de grupo
 - Podrás tener hasta 100 Consumidores (debido al 100 Group ID)
 - Tendrás hasta 300 mensajes por segundo (o 3000 si utilizamos batching)

SQS vs SNS vs Kinesis

SQS:

- Los consumidores "tiran de los datos"
- Los datos se borran después de ser consumidos
- Podemos tener tantos trabajadores (consumidores) como queramos
- No es necesario aprovisionar rendimiento
- Garantías de ordenación sólo en colas FIFO
- Capacidad de retardo de mensajes individuales



SNS:

- Envío de datos a muchos suscriptores
- Hasta 12.500.000 suscriptores
- Los datos no se conservan (se pierden si no se entregan)
- Pub/Sub
- Hasta 100.000 temas (topics)
- No es necesario aprovisionar caudal
- Se integra con SQS para un patrón de arquitectura en fan-out
- Capacidad FIFO para SQS FIFO

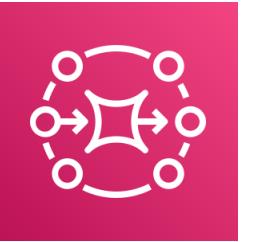


Kinesis:

- Estándar: extraer datos
 - 2 MB por fragmento
- Fan-out reforzado: datos push
 - 2 MB por fragmento y consumidor
- Posibilidad de reproducir los datos
- Pensado para big data en tiempo real, análisis y ETL
- Ordenación a nivel de fragmento
- Los datos caducan a los X días
- Modo aprovisionado o modo de capacidad bajo demanda



Amazon MQ

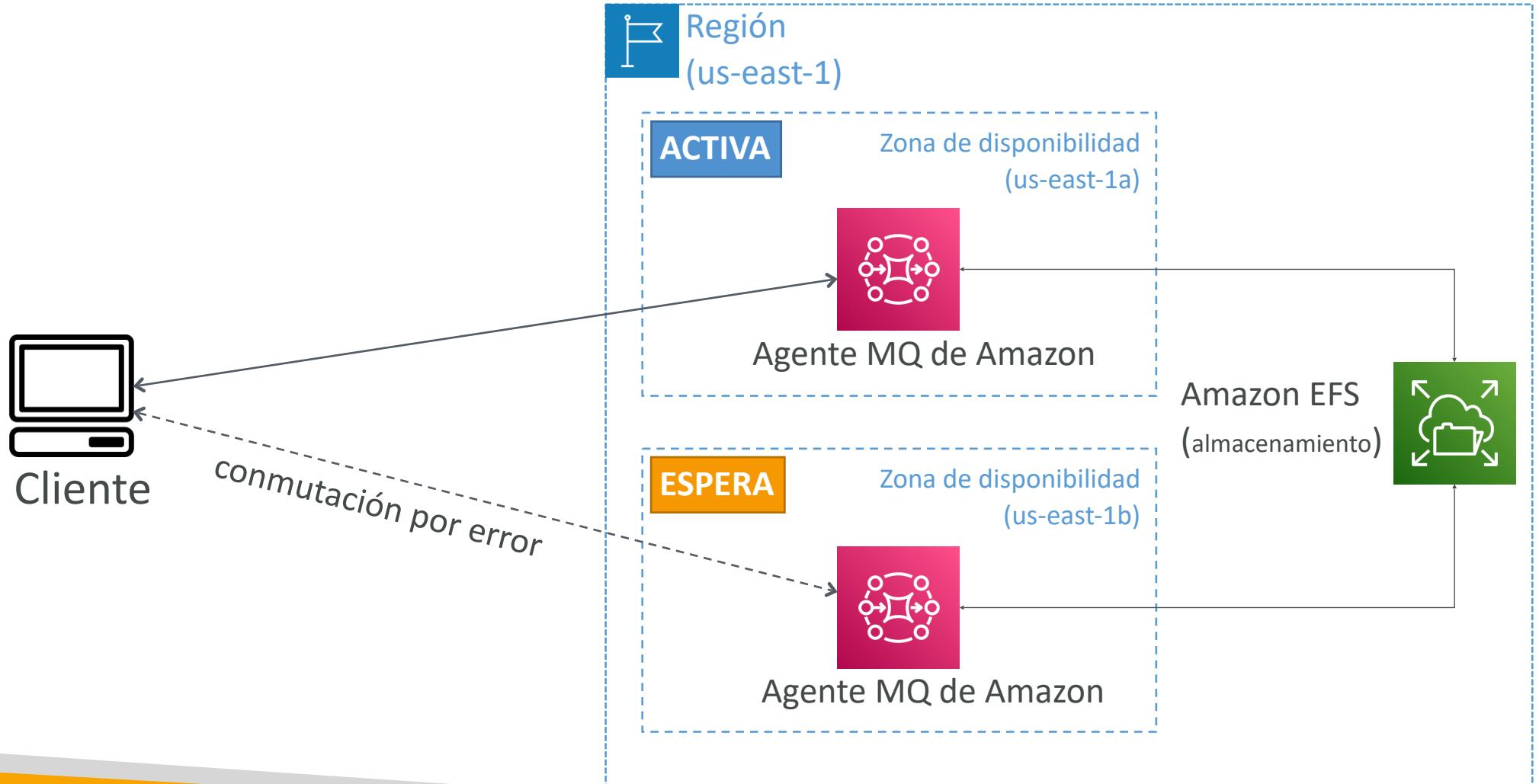


- SQS, SNS son servicios "nativos de la nube": protocolos propietarios de AWS
- Las aplicaciones tradicionales que se ejecutan desde las instalaciones pueden utilizar protocolos abiertos como: MQTT, AMQP, STOMP, Openwire, WSS
- **Al migrar a la nube**, en lugar de rediseñar la aplicación para utilizar SQS y SNS, podemos utilizar Amazon MQ
- **Amazon MQ es un servicio de intermediario de mensajes administrado para**



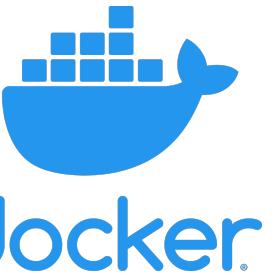
- Amazon MQ no "escala" tanto como SQS / SNS
- Amazon MQ se ejecuta en servidores, puede ejecutarse en Multi-AZ con failover
- Amazon MQ tiene características de cola (~SQS) y características de tema (~SNS)

Amazon MQ – Alta disponibilidad



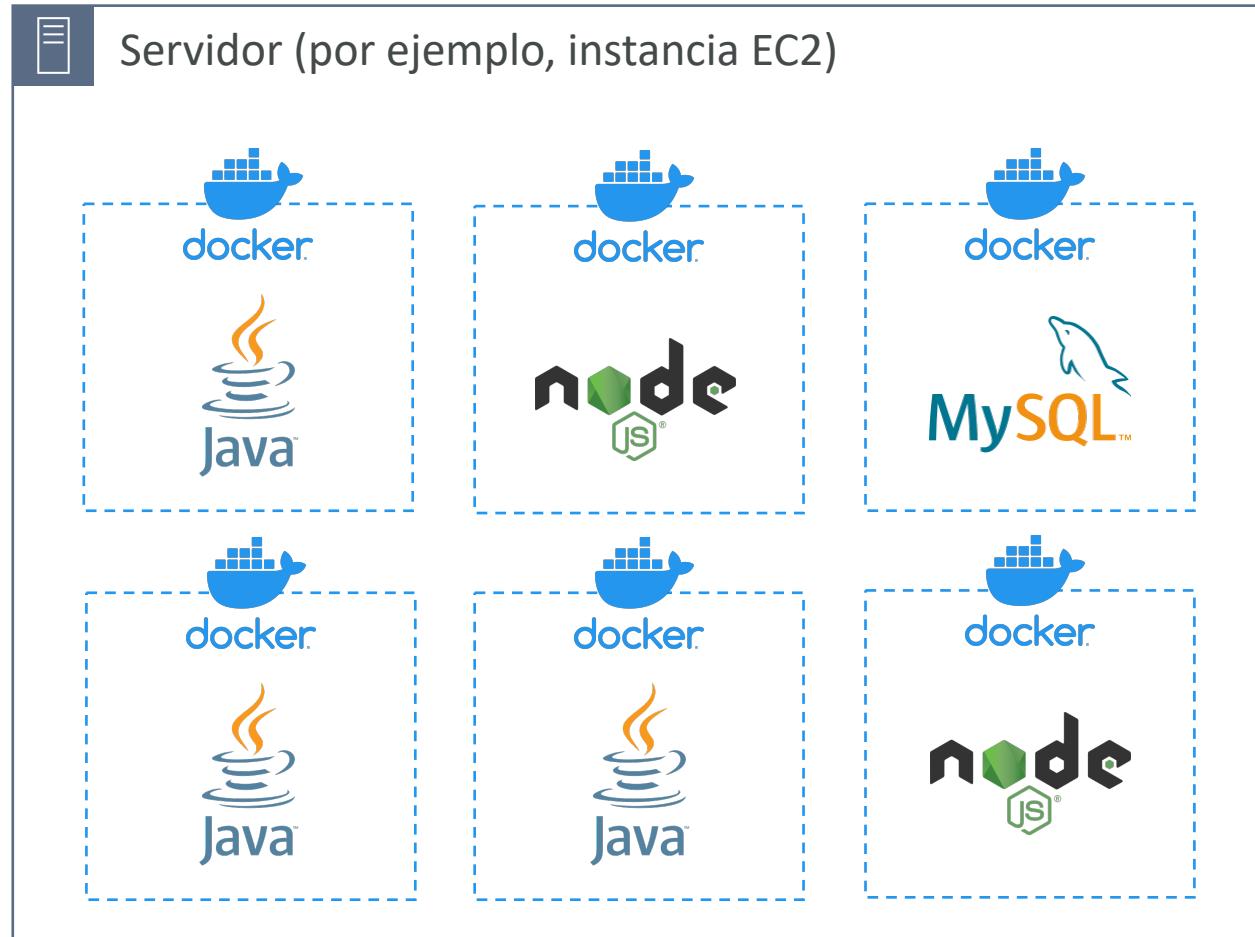
Sección de contenedores

¿Qué es Docker?



- Docker es una plataforma de desarrollo de software para desplegar aplicaciones.
- Las aplicaciones se empaquetan en **contenedores** que pueden ejecutarse en cualquier sistema operativo.
- Las aplicaciones se ejecutan igual, independientemente de dónde se ejecuten.
 - Cualquier máquina
 - Sin problemas de compatibilidad
 - Comportamiento predecible
 - Menos trabajo
 - Más fácil de mantener e implantar
 - Funciona con cualquier lenguaje, sistema operativo y tecnología
- Casos de uso: arquitectura de microservicios, aplicaciones lift-and-shift de on-premises a la nube de AWS, ...

Docker en un Sistema Operativo

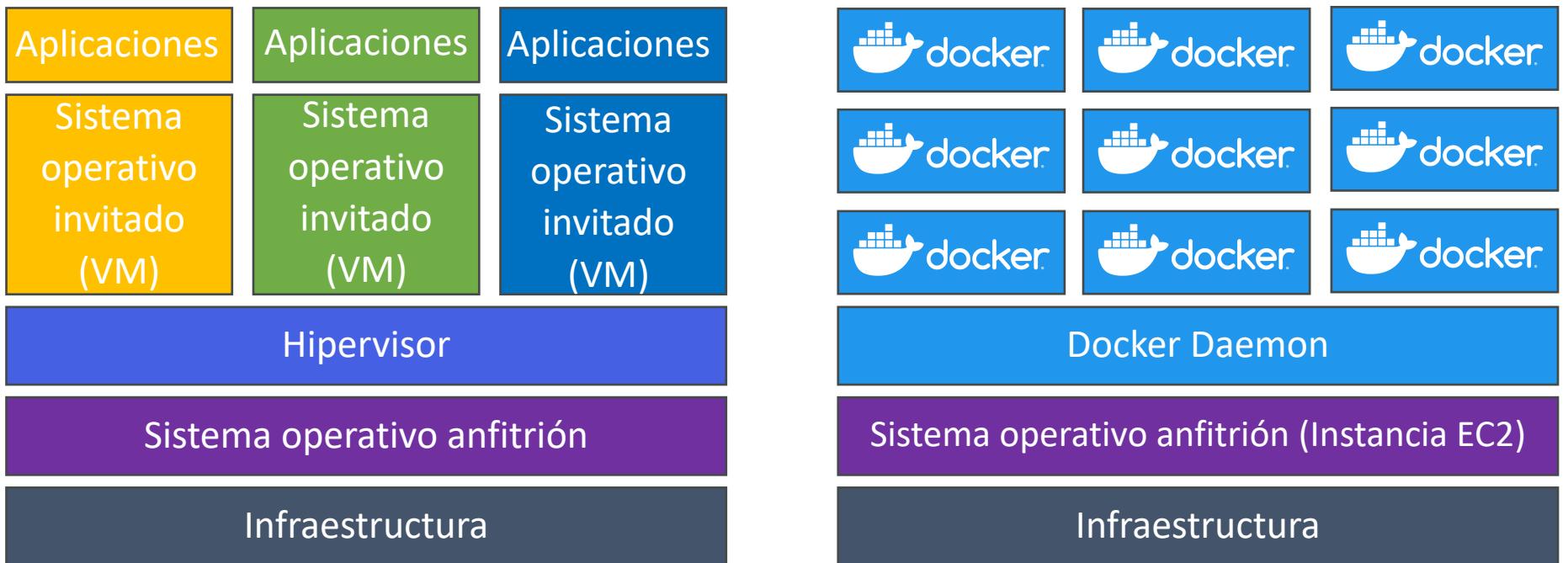


¿Dónde se almacenan las imágenes Docker?

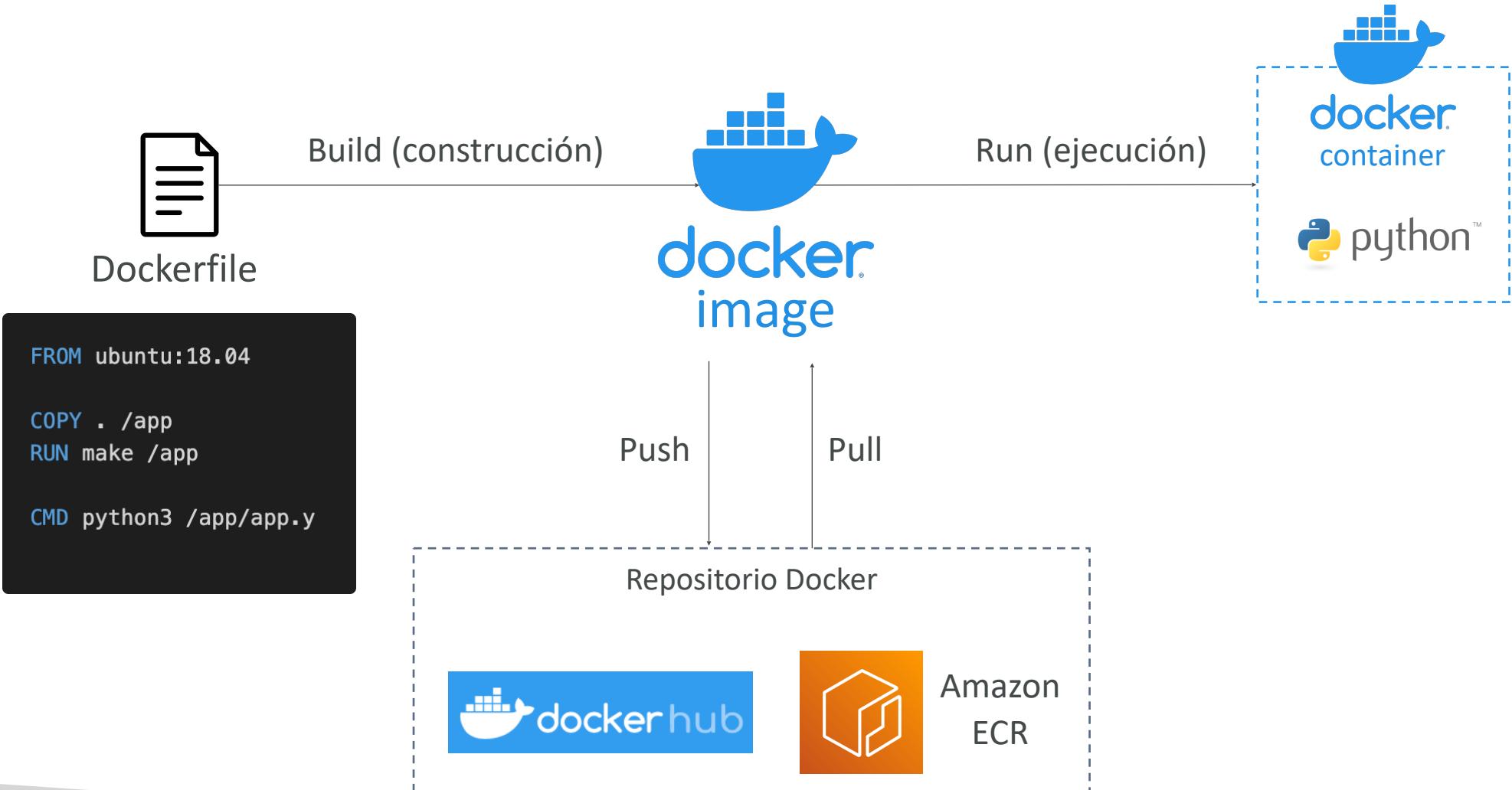
- Las imágenes Docker se almacenan en repositorios Docker
- **Docker Hub (<https://hub.docker.com>)**
 - Repositorio **público**
 - Encuentre imágenes base para muchas tecnologías o sistemas operativos (por ejemplo, Ubuntu, MySQL, ...)
- **Amazon ECR (Registro elástico de contenedores de Amazon)**
 - Repositorio **privado**
 - Repositorio **público** (**Galería pública de Amazon ECR** <https://gallery.ecr.aws>)

Docker vs máquinas virtuales

- Docker es "algo así" como una tecnología de virtualización, pero no exactamente
- Los recursos se comparten con el host => muchos contenedores en un servidor



Primeros pasos con Docker



Gestión de contenedores Docker en AWS

- **Amazon Elastic Container Service (Amazon ECS)**

- Plataforma de contenedores propia de Amazon



Amazon ECS

- **Servicio Amazon Elastic Kubernetes (Amazon EKS)**

- Kubernetes administrado por Amazon (código abierto)



Amazon EKS

- **AWS Fargate**

- Plataforma de contenedores sin servidor propia de Amazon
- Funciona con ECS y con EKS



AWS Fargate

- **Amazon ECR**

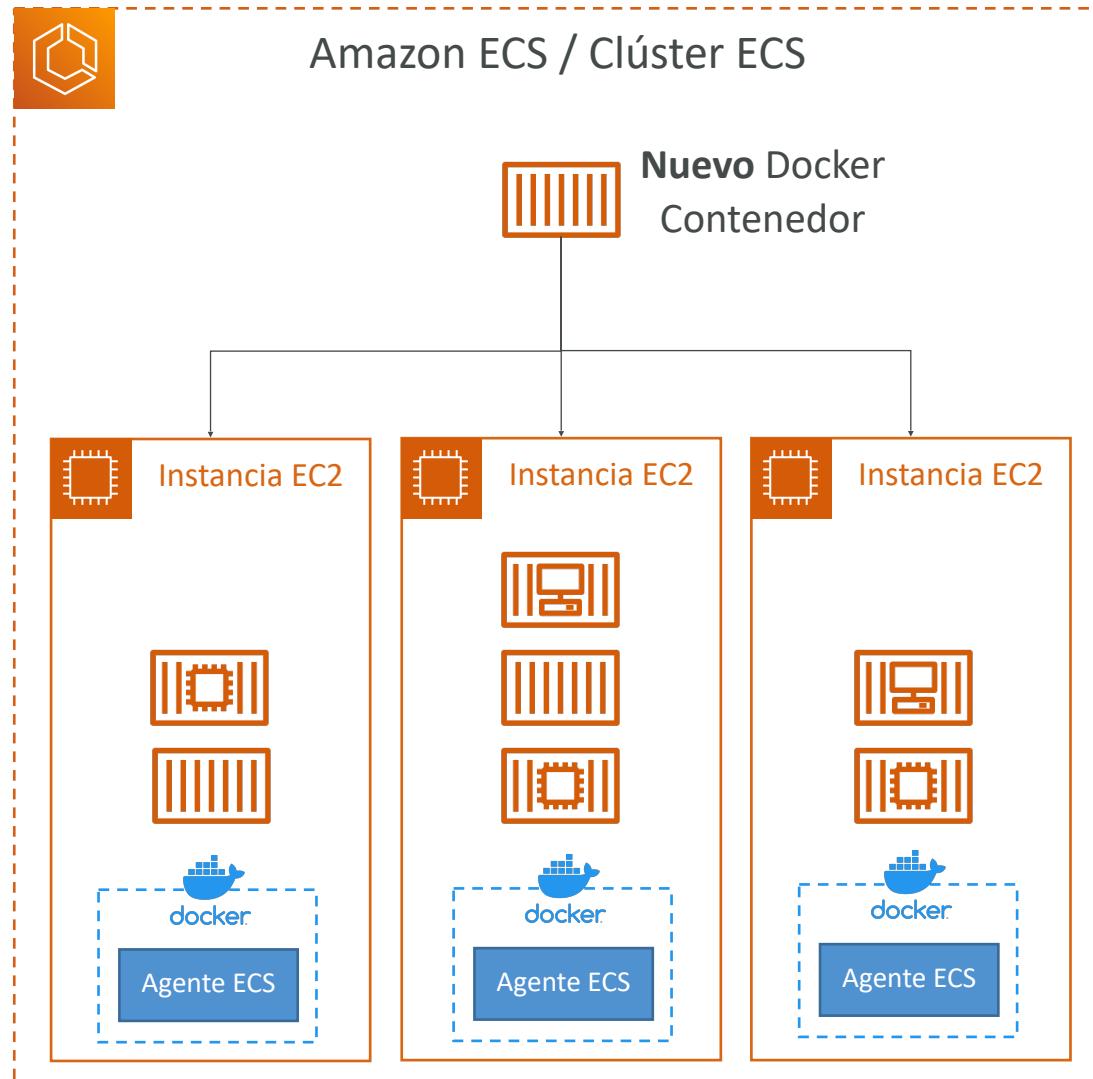
- Almacena imágenes de contenedores



Amazon ECR

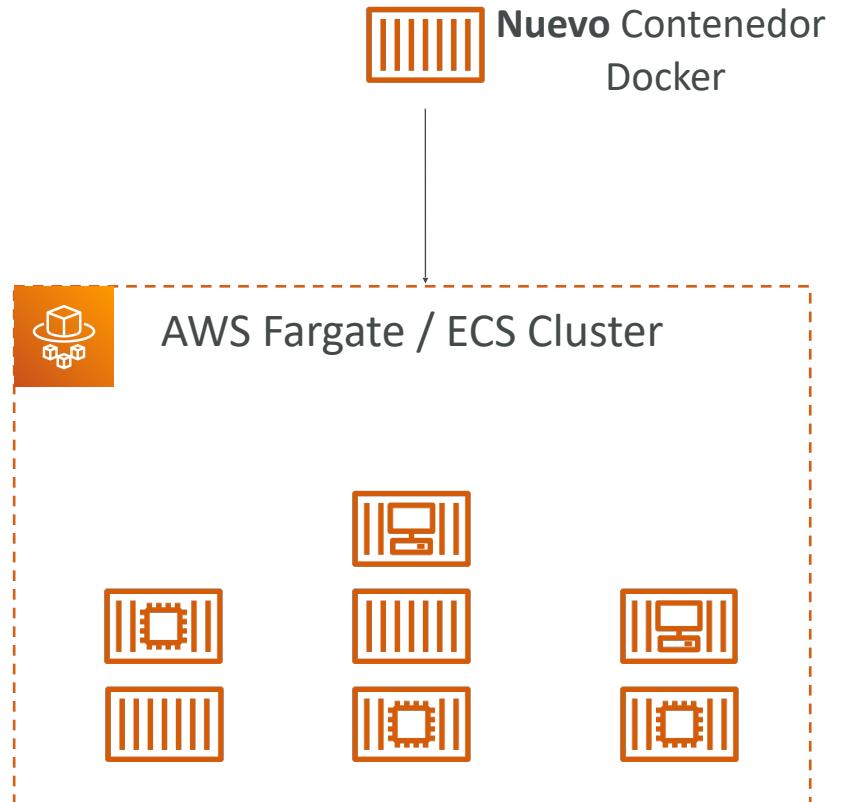
Amazon ECS - Tipo de lanzamiento EC2

- ECS = Elastic Container Service
- Lanzar contenedores Docker en AWS = Lanzar **tareas ECS** en clústeres ECS
- **Tipo de lanzamiento EC2: debe aprovisionar y mantener la infraestructura (las instancias EC2)**
- Cada Instancia EC2 debe ejecutar el Agente ECS para registrarse en el Cluster ECS
- AWS se encarga de iniciar / detener los contenedores



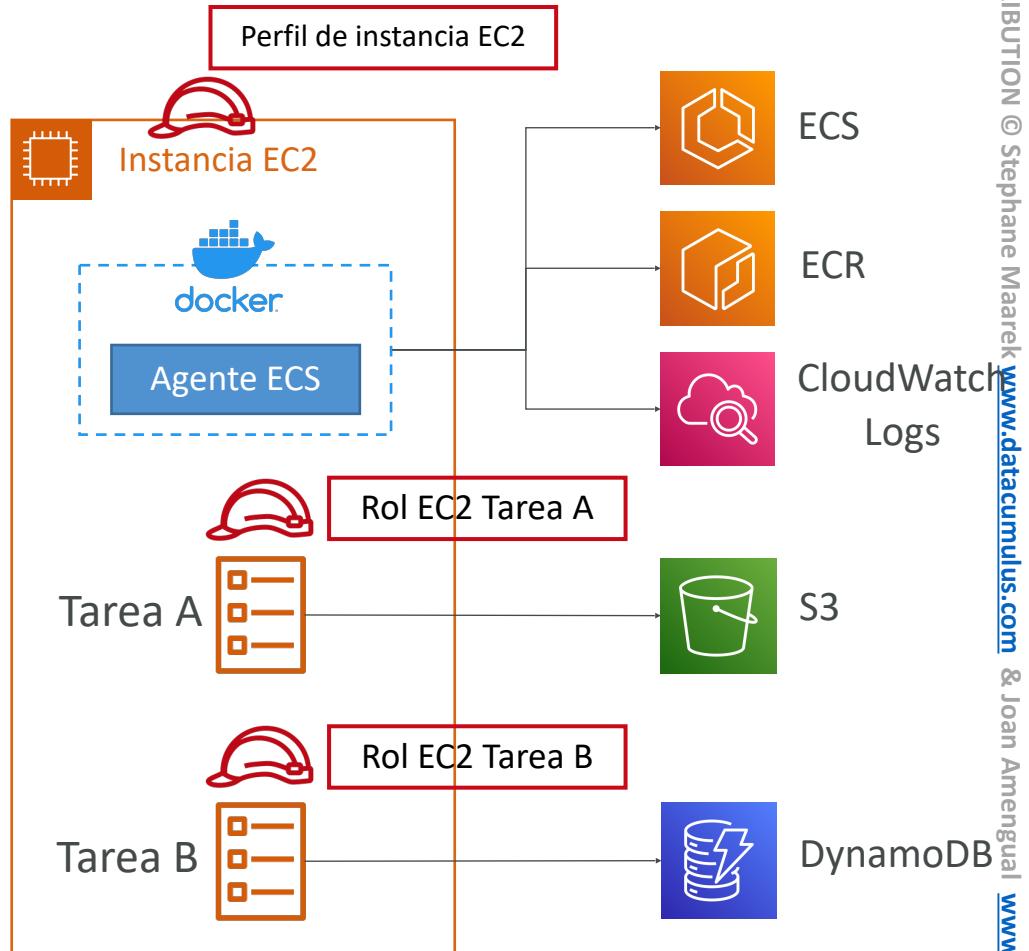
Amazon ECS - Tipo de lanzamiento Fargate

- Lanzar contenedores Docker en AWS
- **No aprovisionas la infraestructura (no hay instancias EC2 que administrar)**
- **¡Todo es Serverless!**
- Sólo tienes que crear definiciones de tareas
- AWS ejecuta las tareas ECS por ti en función de la CPU / RAM que necesites.
- Para escalar, basta con aumentar el número de tareas. Simple - no más instancias EC2



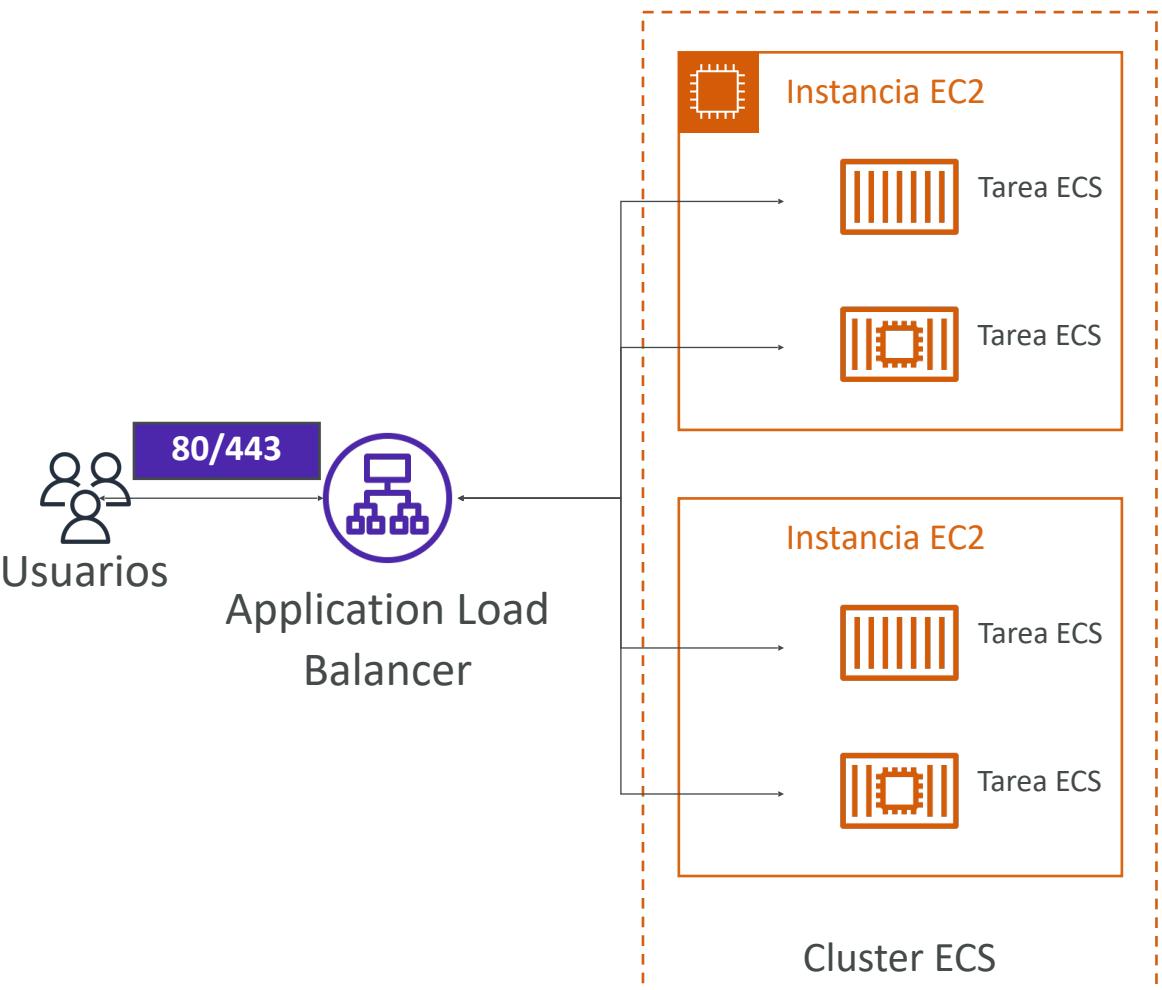
Amazon ECS - Roles IAM para ECS

- **Perfil de instancia EC2 (sólo tipo de lanzamiento EC2):**
 - Utilizado por el agente ECS
 - Realiza llamadas API al servicio ECS
 - Envía logs de contenedores a CloudWatch Logs
 - Extrae imagen de Docker de ECR
 - Hace referencia a datos sensibles en Secrets Manager o SSM Parameter Store
- **Rol de tarea ECS:**
 - Permite que cada tarea tenga un rol específico
 - Utiliza diferentes roles para los diferentes Servicios ECS que ejecute
 - El rol de la tarea se define en la definición de la tarea



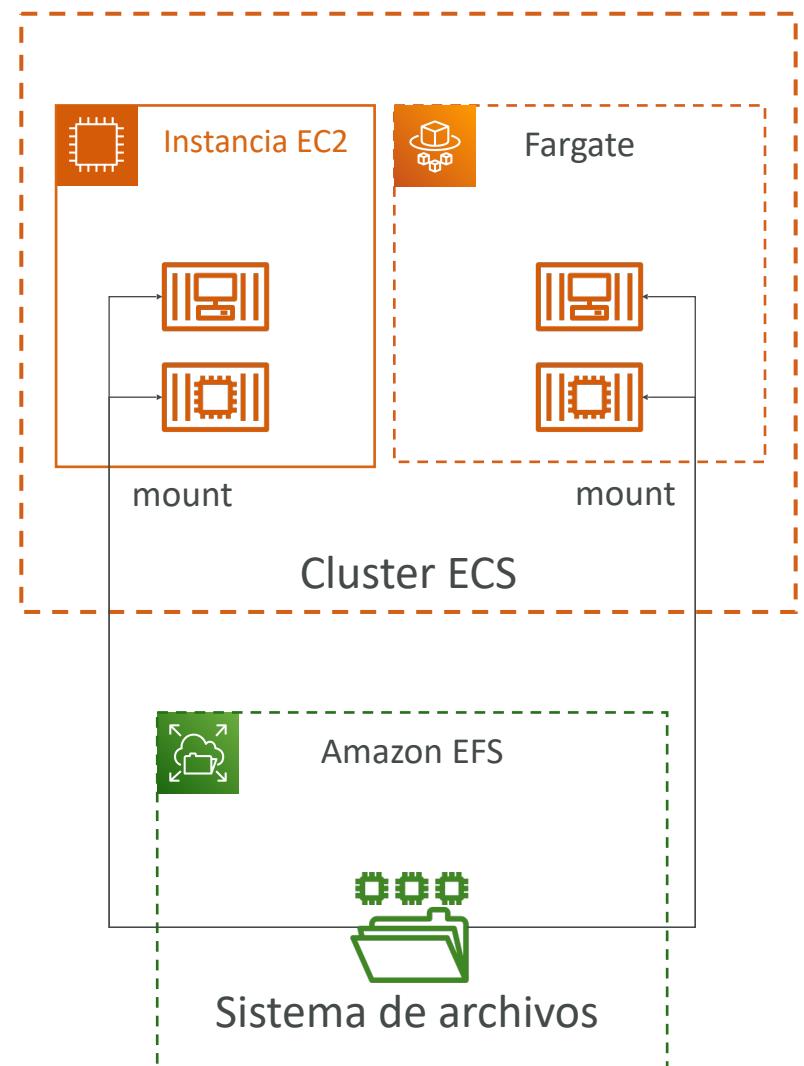
Amazon ECS - Integraciones de balanceadores de carga

- **Application Load Balancer** es compatible y funciona para la mayoría de los casos de uso.
- **Network Load Balancer** recomendado solo para casos de uso de alto rendimiento o para combinarlo con AWS Private Link
- **Elastic Load Balancer** es compatible pero no se recomienda (sin características avanzadas - sin Fargate)

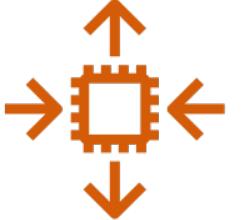


Amazon ECS - Volúmenes de datos (EFS)

- Montar sistemas de archivos EFS en tareas ECS
- Funciona tanto para los tipos de lanzamiento **EC2** como **Fargate**
- Las tareas que se ejecuten en cualquier AZ compartirán los mismos datos en el sistema de archivos EFS
- **Fargate + EFS = Sin servidor**
- Casos de uso: almacenamiento compartido multi-AZ persistente para sus contenedores
- Nota:
 - Amazon S3 no se puede montar como sistema de archivos



Escalado automático del servicio ECS



- Aumentar/disminuir automáticamente el número deseado de tareas ECS
- Amazon ECS Auto Scaling utiliza **AWS Application Auto Scaling**
 - Utilización media de la CPU del servicio ECS
 - Utilización media de memoria del servicio ECS - Escalado en RAM
 - Recuento de solicitudes de ALB por objetivo - métrica procedente del ALB
- **Seguimiento de objetivo** - escala basada en el valor objetivo para una métrica específica de CloudWatch
- **Escalado por pasos** - escalado basado en una alarma CloudWatch específica
- **Escalado programado** - escalado basado en una fecha/hora especificada (cambios predecibles)
- Autoescalado del servicio ECS (nivel de tarea) \neq Autoescalado de EC2 (nivel de instancia de EC2)
- Fargate Auto Scaling es mucho más fácil de configurar (porque es **Serverless**)

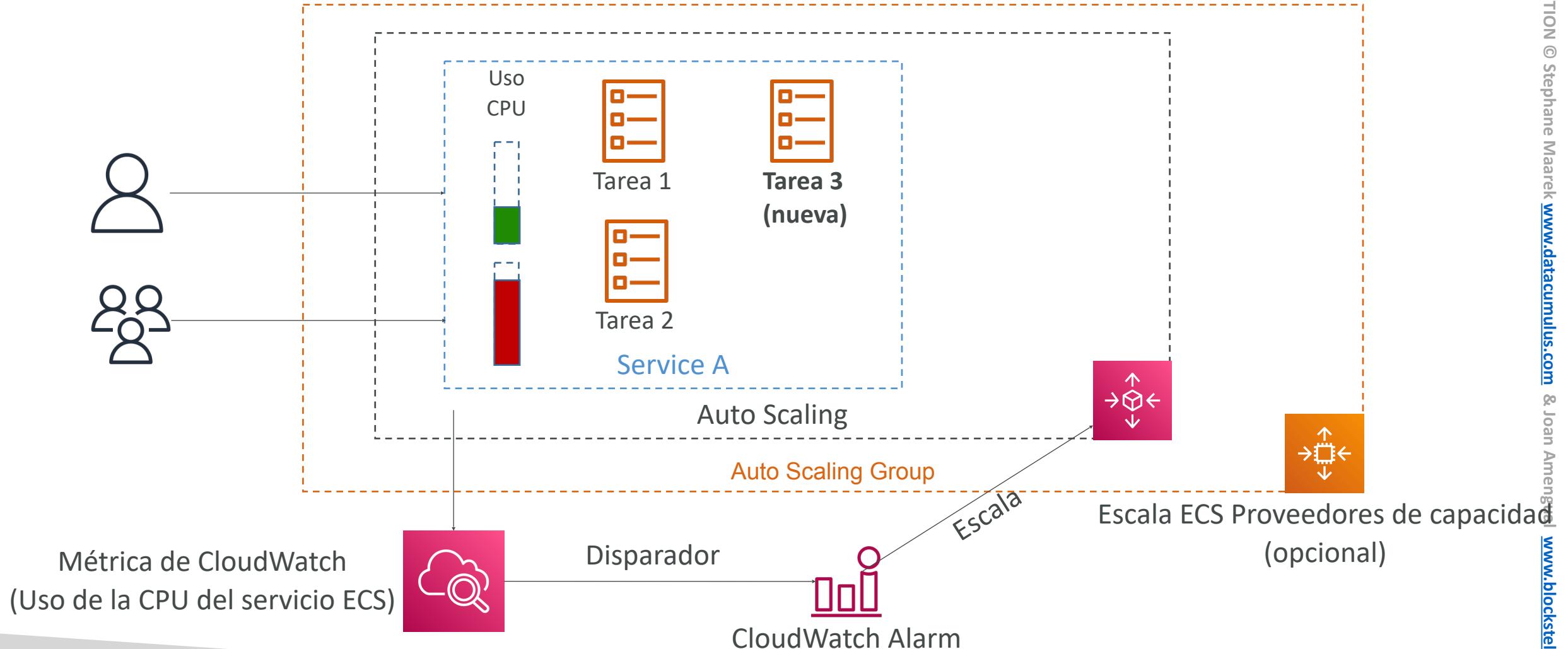
Tipo de lanzamiento EC2

Escalado automático de instancias EC2

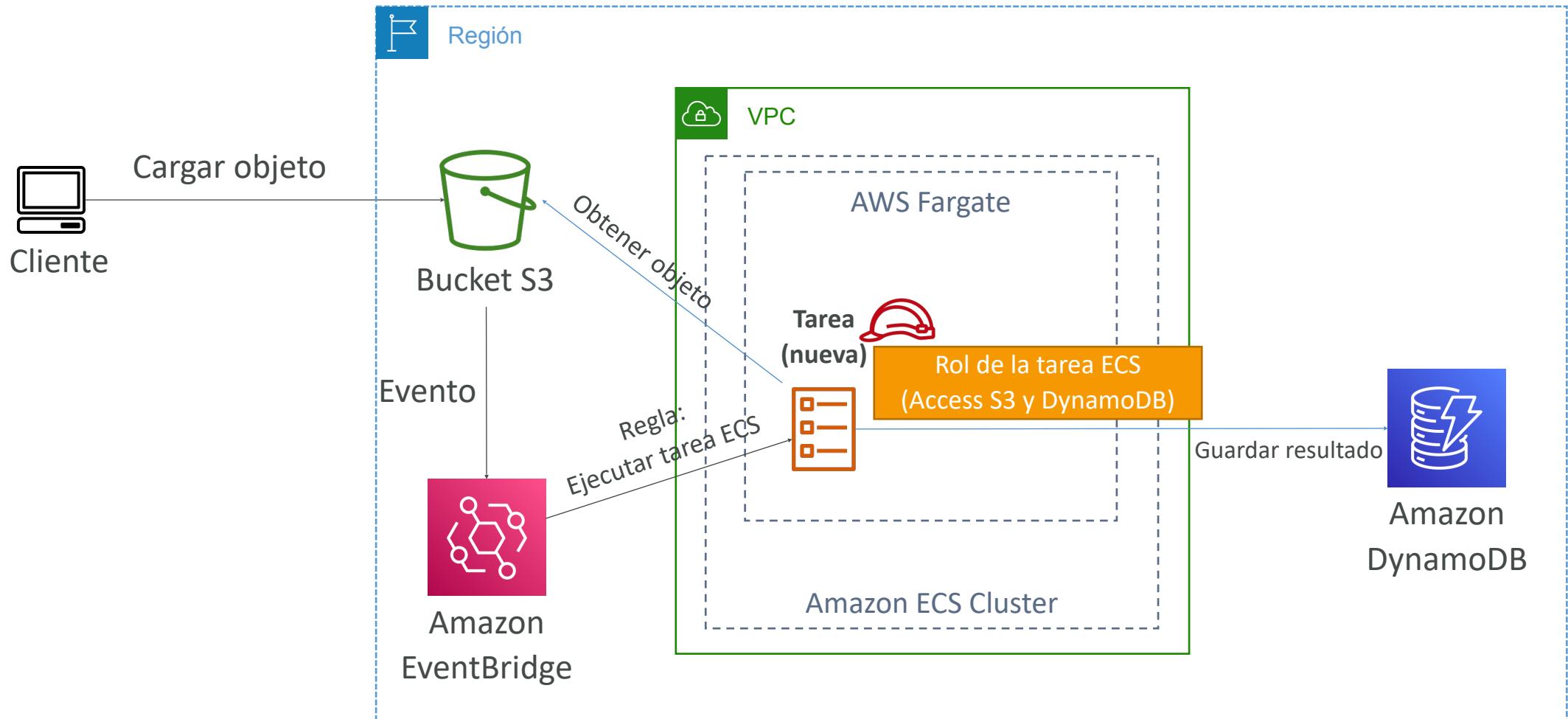
- Acomodar el escalado de servicios ECS añadiendo instancias EC2 subyacentes
- **Escalado automático de grupos (Auto Scaling Group Scaling)**
 - Escala el ASG en función de la utilización de la CPU
 - Añadir instancias EC2 con el tiempo
- **Proveedor de capacidad de clúster ECS**
 - Se utiliza para aprovisionar y escalar automáticamente la infraestructura para tareas ECS
 - Proveedor de capacidad emparejado con un Auto Scaling Group
 - Añade instancias EC2 cuando falte capacidad (CPU, RAM...)

Escalado ECS

Ejemplo de uso de CPU de servicio



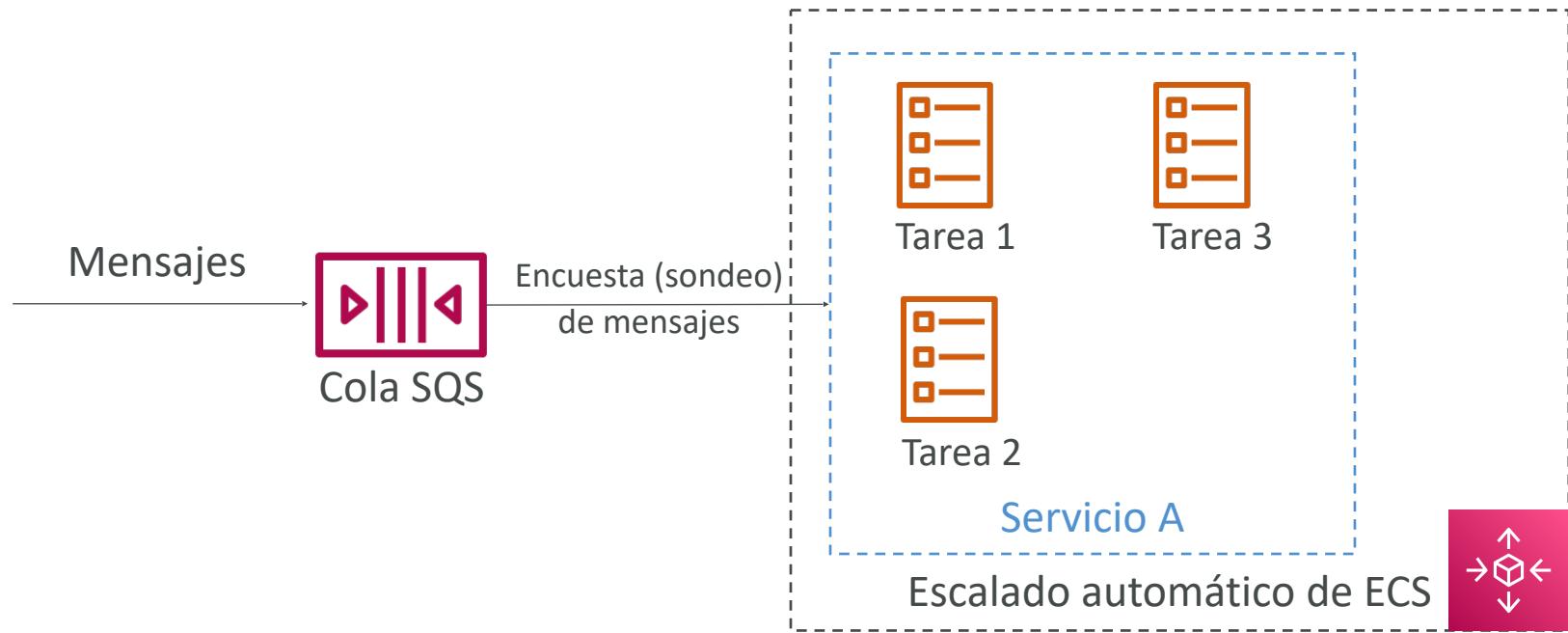
Tareas ECS invocadas por Event Bridge



Tareas ECS invocadas por el EventBridge



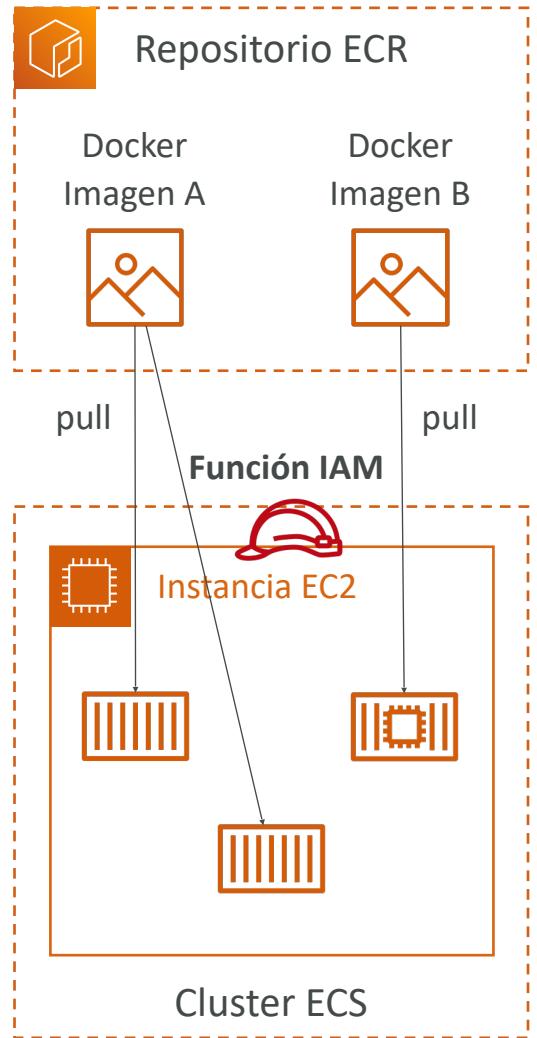
ECS - Ejemplo de cola SQS



Amazon ECR



- ECR = Registro elástico de contenedores
- Almacenar y administrar imágenes Docker en AWS
- Repositorio **privado** y **público** (**Amazon ECR Public Gallery** <https://gallery.ecr.aws>)
- Totalmente integrado con ECS, respaldado por Amazon S3
- El acceso se controla a través de IAM (errores de permiso => política)
- Soporta escaneo de vulnerabilidades de imágenes, versionado, etiquetas de imágenes, ciclo de vida de imágenes, ...

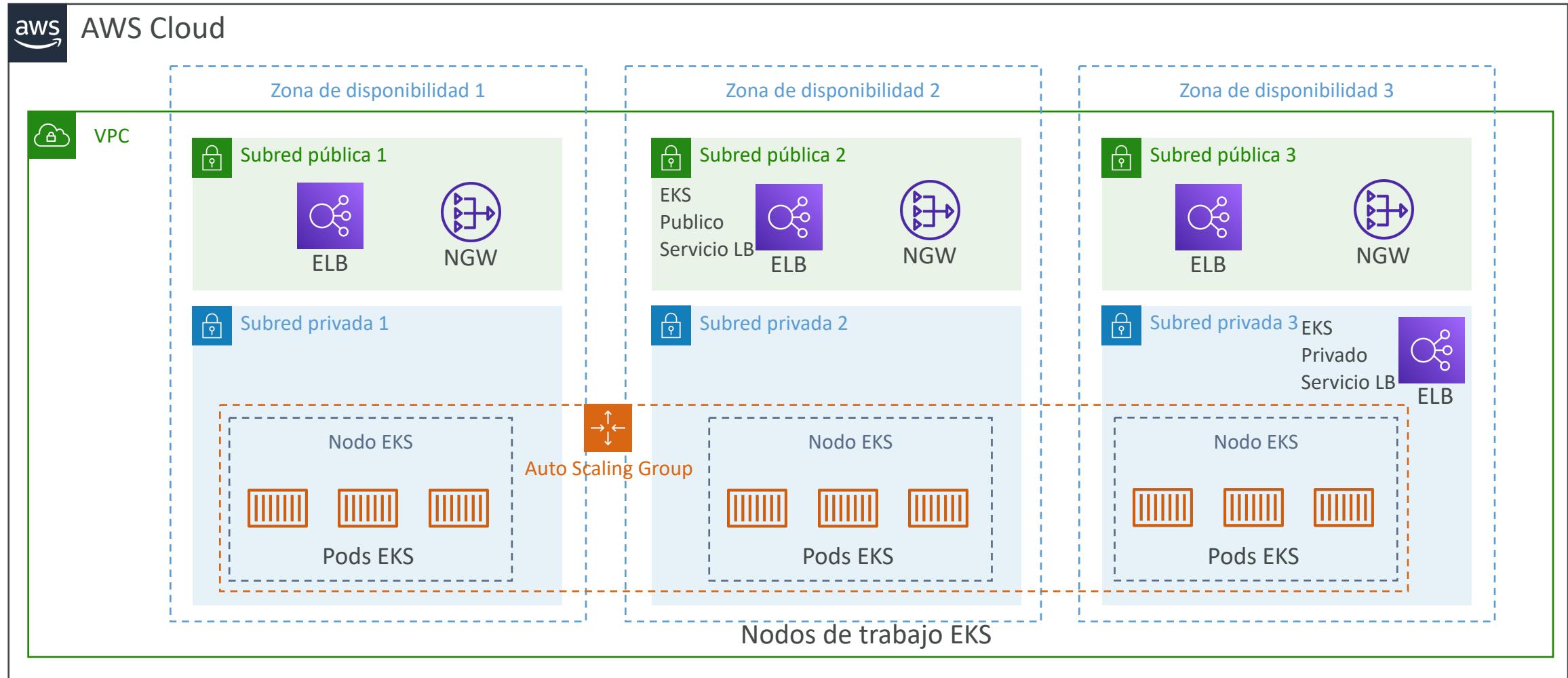


Visión general de Amazon EKS



- Amazon EKS = Servicio Amazon Elastic **Kubernetes**
- Es una forma de lanzar **clústeres Kubernetes administrados en AWS**
- Kubernetes es un sistema de código abierto para el despliegue, escalado y gestión automáticos de aplicaciones en contenedores (normalmente Docker)
- Es una alternativa a ECS, objetivo similar pero API diferente
- EKS soporta **EC2** si quieres desplegar nodos trabajadores o **Fargate** para desplegar contenedores sin servidor
- **Caso de uso:** si tu empresa ya utiliza Kubernetes on-premises o en otra nube, y quiere migrar a AWS utilizando Kubernetes
- **Kubernetes es agnóstico a la nube (puede utilizarse en cualquier nube - Azure, GCP...)**

Amazon EKS - Diagrama



Amazon EKS - Tipos de nodos

- **Grupos de nodos gestionados**

- Crea y gestiona Nodos (instancias EC2) para ti
- Los nodos forman parte de un ASG gestionado por EKS
- Admite instancias bajo demanda o puntuales

- **Nodos autogestionados**

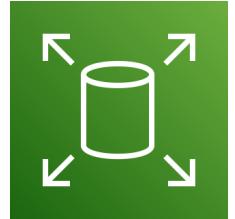
- Nodos creados por ti y registrados en el clúster EKS y gestionados por un ASG
- Puede utilizar AMI preconstruidas - Amazon EKS Optimized AMI
- Admite instancias bajo demanda o puntuales

- **AWS Fargate**

- No requiere mantenimiento; no se administran nodos

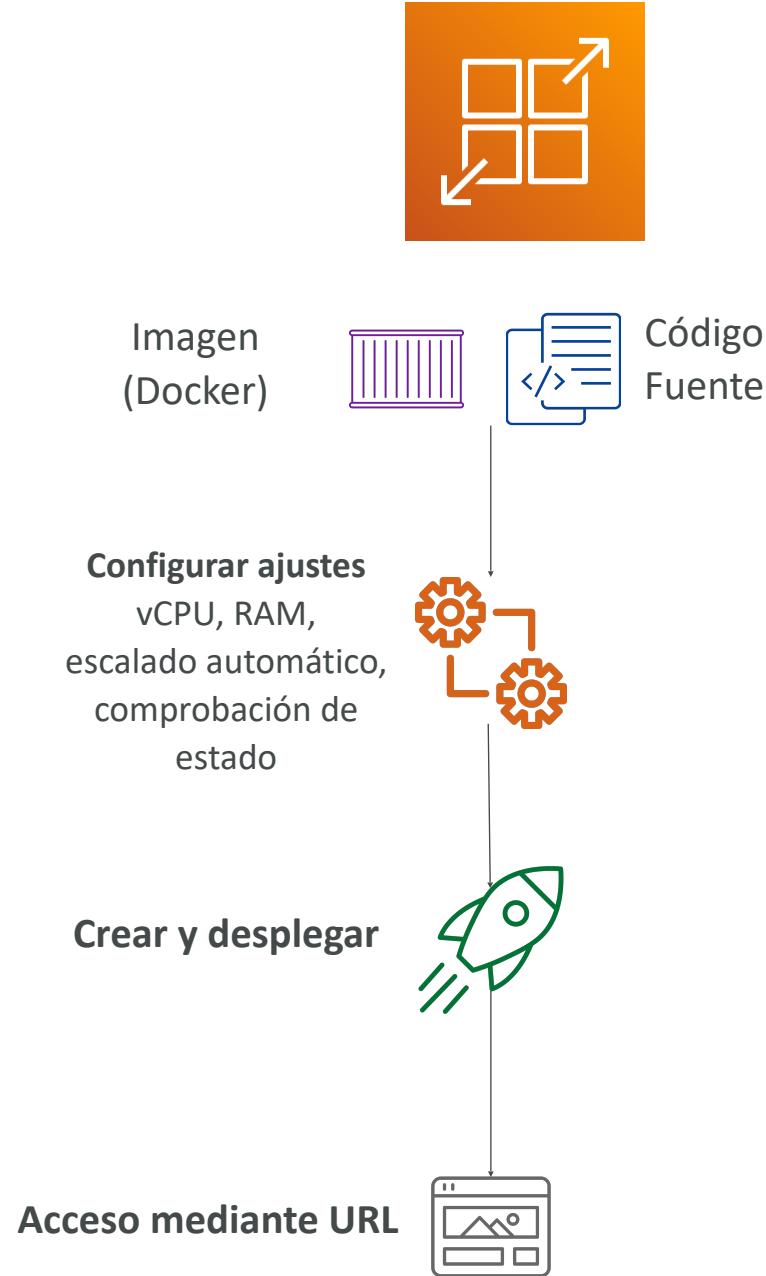
Amazon EKS - Volúmenes de datos

- Necesidad de especificar el **StorageClass** en el clúster EKS
- Aprovecha un controlador compatible con **Container Storage Interface (CSI)**
- Compatible con...
 - Amazon EBS
 - Amazon EFS (funciona con Fargate)
 - Amazon FSx para Lustre
 - Amazon FSx para NetApp ONTAP



AWS App Runner (Corredor de aplicaciones AWS)

- Servicio totalmente gestionado que facilita el despliegue de aplicaciones web y API a escala
- No se requiere experiencia en infraestructura
- Empieza con tu código fuente o una imagen de docker
- Crea y despliega automáticamente la aplicación web
- Escalado automático, alta disponibilidad, equilibrador de carga, cifrado
- Soporte de acceso VPC
- Conexión a servicios de base de datos, caché y cola de mensajes
- Casos de uso: aplicaciones web, API, microservicios, despliegues rápidos en producción



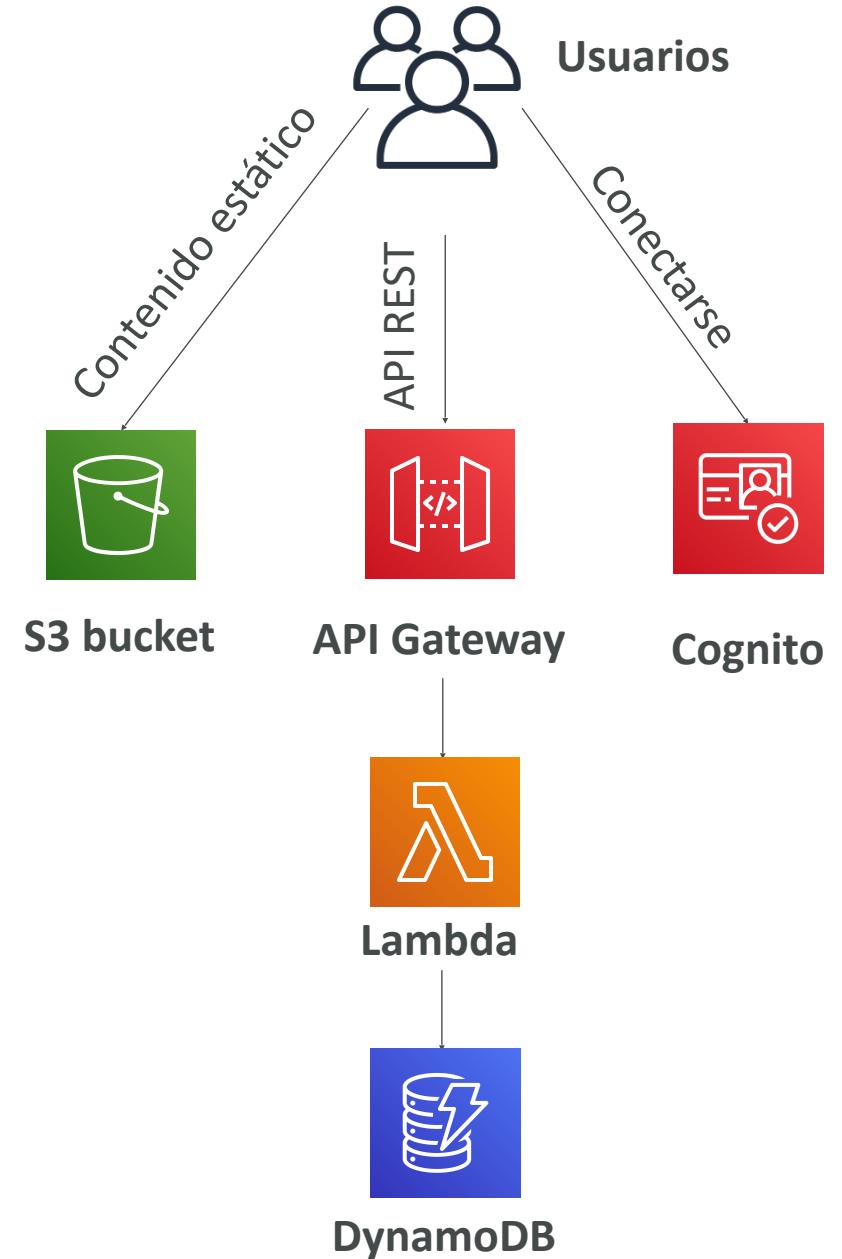
Visión general de Serverless

¿Qué es serverless?

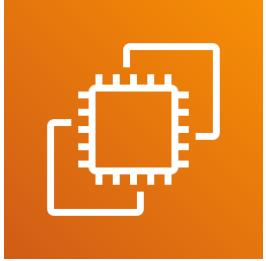
- Serverless es un nuevo paradigma en el que los desarrolladores ya no tienen que gestionar servidores...
- Sólo despliegan código
- Sólo despliegan... ¡funciones!
- Inicialmente... Serverless == FaaS (Función como Servicio)
- Serverless fue pionero en AWS Lambda pero ahora también incluye cualquier cosa que se gestione: "bases de datos, mensajería, almacenamiento, etc".
- **Serverless no significa que no haya servidores...** significa que simplemente no los gestionas / aprovisionas / ves

Sin servidor en AWS

- AWS Lambda
- DynamoDB
- AWS Cognito
- AWS API Gateway
- Amazon S3
- AWS SNS y SQS
- Kinesis Data Firehose
- Aurora Serverless
- Funciones por pasos (Step Functions)
- Fargate



Por qué AWS Lambda



Amazon EC2

- Servidores virtuales en la nube
- Limitado por RAM y CPU
- Funcionamiento continuo
- Escalado significa intervención para añadir/eliminar servidores



Amazon Lambda

- **Funciones** virtuales: sin servidores que gestionar
- Limitado por el tiempo - **ejecuciones cortas**
- Ejecución **bajo demanda**
- **Escalado automatizado**

Beneficios de AWS Lambda

- Precios sencillos:
 - Pago por solicitud y tiempo de cómputo
 - Capa gratuita de 1.000.000 de solicitudes de AWS Lambda y 400.000 GB de tiempo de cómputo
- Integrado con todo el conjunto de servicios de AWS
- **Dirigido por eventos:** las funciones son invocadas por AWS cuando se necesitan
- Integrado con muchos lenguajes de programación
- Fácil monitorización a través de AWS CloudWatch
- Fácil de obtener más recursos por funciones (¡hasta 10 GB de RAM!)
- ¡El aumento de la RAM también mejorará la CPU y la red!

Soporte del lenguaje AWS Lambda

- Node.js (JavaScript)
- Python
- Java (compatible con Java 8)
- C# (.NET Core)
- Golang
- C# / Powershell
- Ruby
- API de tiempo de ejecución personalizado (compatible con la comunidad, ejemplo Rust)
- Imagen de contenedor Lambda
 - La imagen del contenedor debe implementar la API de tiempo de ejecución Lambda
 - Se prefiere ECS / Fargate para ejecutar imágenes Docker arbitrarias

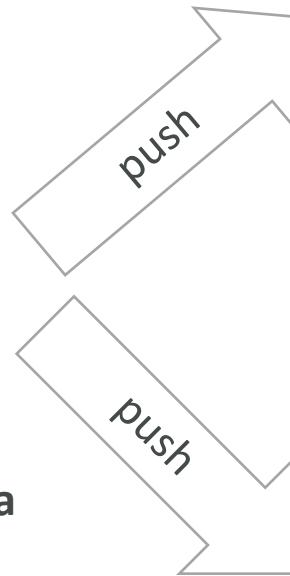
Ejemplo: Creación de miniaturas Serverless



Nueva imagen en S3



La función AWS Lambda
crea una miniatura



Nueva miniatura en S3



Nombre de la imagen
Tamaño de la imagen
Fecha de creación
etc.



Metadata en DynamoDB

Ejemplo: Trabajo CRON Serverless



CloudWatch Events
EventBridge



Función AWS Lambda
realiza una tarea

Precios de AWS Lambda: ejemplo

- Encontrará información general sobre precios en el enlace:
 - <https://aws.amazon.com/lambda/pricing/>
- Pago por **llamadas**:
 - Las primeras 1.000.000 de solicitudes son gratuitas
 - 0,20 \$ por cada millón de solicitudes (0,0000002 \$ por solicitud)
- Pago por **duración**: (en incrementos de 1 ms)
 - 400.000 GB-segundos de tiempo de cálculo al mes GRATIS
 - == 400.000 segundos si la función es de 1GB RAM
 - == 3.200.000 segundos si la función es de 128 MB RAM
 - Después, 1 dólar por 600.000 GB-segundos
- Suele ser muy barato ejecutar AWS Lambda, por lo que es muy popular

Límites de AWS Lambda que debe conocer - **por región**

- **Ejecución:**

- Asignación de memoria: 128 MB - 10GB (incrementos de 1 MB)
- Tiempo máximo de ejecución: 900 segundos (15 minutos)
- Variables de entorno (4 KB)
- Capacidad de disco en el "contenedor de funciones" (en /tmp): 512 MB a 10GB
- Conurrencia de ejecuciones: 1000 (puede aumentarse)

- **Despliegue:**

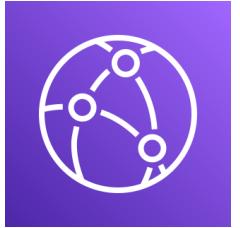
- Tamaño del despliegue de la función lambda (.zip comprimido): 50 MB
- Tamaño del despliegue sin comprimir (código + dependencias): 250 MB
- Puede utilizar el directorio /tmp para cargar otros archivos al inicio
- Tamaño de las variables de entorno: 4 KB



Personalización en el borde (Edge)

- Muchas aplicaciones modernas ejecutan alguna forma de la lógica en el borde
- **Función de borde (Edge Function):**
 - Un código que escribes y adjuntas a las distribuciones de CloudFront
 - Se ejecuta cerca de los usuarios para minimizar la latencia
- CloudFront proporciona dos tipos:
 - **Funciones CloudFront y Lambda@Edge**
- No tienes que administrar ningún servidor, implementado globalmente
- Caso de uso: personalizar el contenido de la CDN
- Paga solo por lo que utilizas
- Totalmente sin servidor

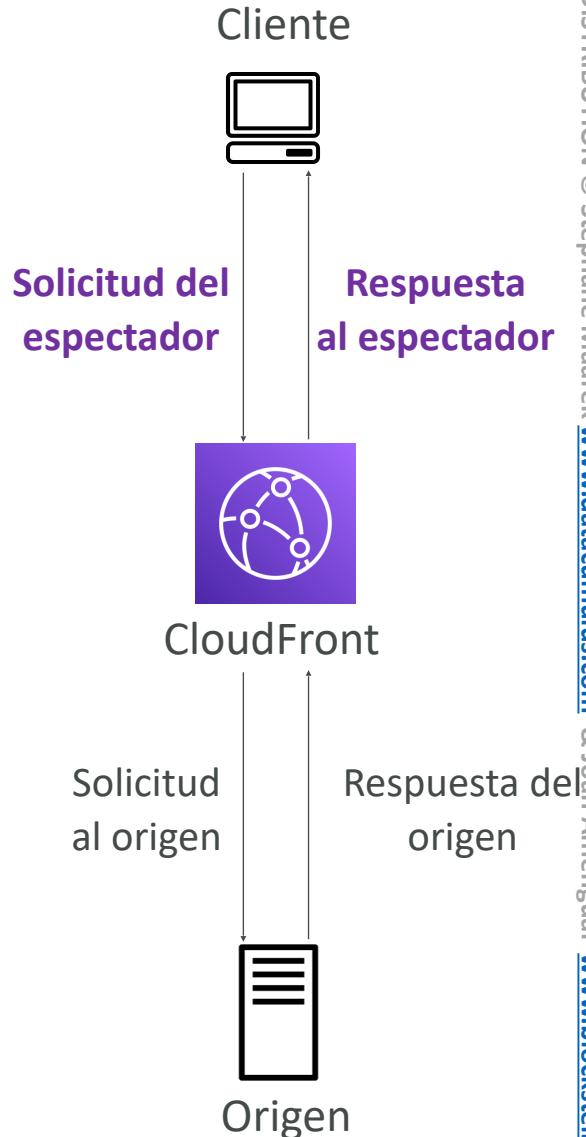
Funciones de CloudFront y casos de uso de Lambda@Edge



- Seguridad y privacidad de sitios web
- Aplicaciones web dinámicas en el Edge
- Optimización para motores de búsqueda (SEO)
- Enrutamiento inteligente entre orígenes y centros de datos
- Mitigación de bots en el Edge
- Transformación de imágenes en tiempo real
- Pruebas A/B
- Autenticación y autorización de usuarios
- Priorización de usuarios
- Seguimiento y análisis de usuarios

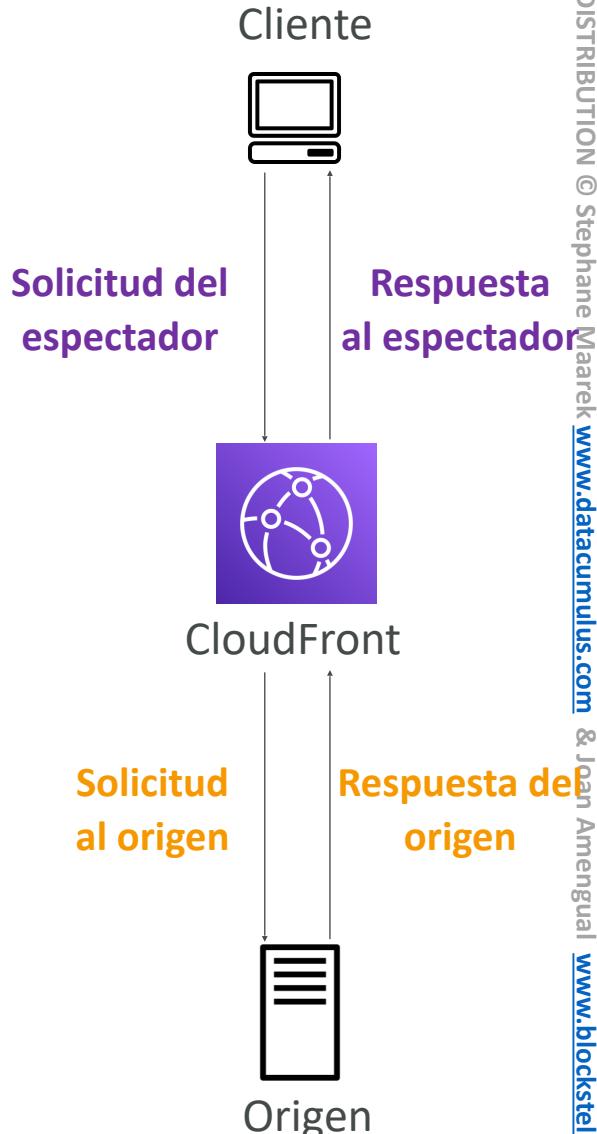
Funciones de CloudFront

- Funciones ligeras escritas en JavaScript
- Para personalizaciones de CDN a gran escala y sensibles a la latencia
- Tiempos de arranque inferiores a milisegundos, **millones de solicitudes/segundo**
- Se utiliza para modificar las solicitudes y respuestas de los espectadores:
 - **Solicitud del espectador:** después de que CloudFront reciba una solicitud de un espectador
 - **Respuesta al espectador:** antes de que CloudFront envíe la respuesta al espectador
- Característica nativa de CloudFront (gestiona el código completamente dentro de CloudFront)



Lambda@Edge

- Funciones lambda escritas en NodeJS o Python
- Escala a **1000s de peticiones/segundo**
- Se utiliza para modificar las solicitudes y respuestas de CloudFront:
 - **Solicitud del espectador**: después de que CloudFront reciba una solicitud de un espectador.
 - **Solicitud al origen**: antes de que CloudFront reenvíe la solicitud al origen.
 - **Respuesta del origen**: después de que CloudFront reciba la respuesta del origen
 - **Respuesta al espectador**: antes de que CloudFront reenvíe la respuesta al espectador.
- Crea tus funciones en una región de AWS (us-east-1), luego CloudFront replica a tus ubicaciones



Funciones de CloudFront vs Lambda@Edge

	Funciones CloudFront	Lambda@Edge
Soporte en tiempo de ejecución	JavaScript	Node.js, Python
Número de solicitudes	Millones de solicitudes por segundo	Miles de solicitudes por segundo
Triggers de CloudFront	- Solicitud/Respuesta del espectador	- Solicitud/Respuesta del espectador - Origen Solicitud/Respuesta
Máx. Tiempo de ejecución	< 1 ms	5 – 10 segundos
Max. Memoria	2 MB	128 MB hasta 10 GB
Tamaño total del paquete	10 KB	1 MB – 50 MB
Acceso a la red, acceso al sistema de archivos	No	Si
Acceso al cuerpo de la solicitud	No	Si
Precios	Nivel gratuito disponible, 1/6 del precio de @Edge	No hay nivel gratuito, se cobra por solicitud y duración

Funciones de CloudFront vs Lambda@Edge

Casos de uso

Funciones de CloudFront

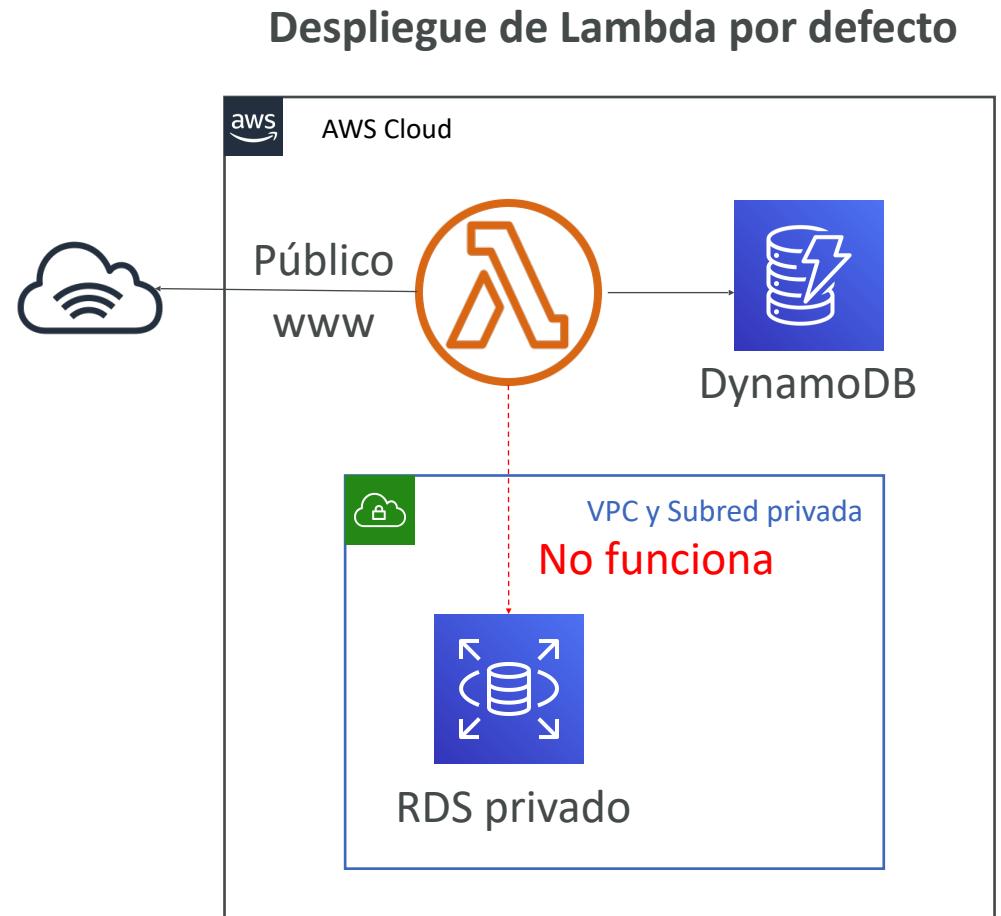
- Normalización de la clave de caché
 - Transformación de los atributos de la solicitud (cabeceras, cookies, cadenas de consulta, URL) para crear una clave de caché óptima
- Manipulación de cabeceras
 - Inserción/modificación/eliminación de cabeceras HTTP en la solicitud o la respuesta
- Reescritura o redireccionamiento de URL
- Autenticación y autorización de solicitudes
 - Creación y validación de tokens generados por el usuario (por ejemplo, JWT) para permitir/denegar solicitudes

Lambda@Edge

- Mayor tiempo de ejecución (varios ms)
- CPU o memoria ajustables
- El código depende de una tercera biblioteca (p. ej., AWS SDK para acceder a otros servicios de AWS)
- Acceso a la red para utilizar servicios externos para el procesamiento
- Acceso al sistema de archivos o acceso al cuerpo de las solicitudes HTTP

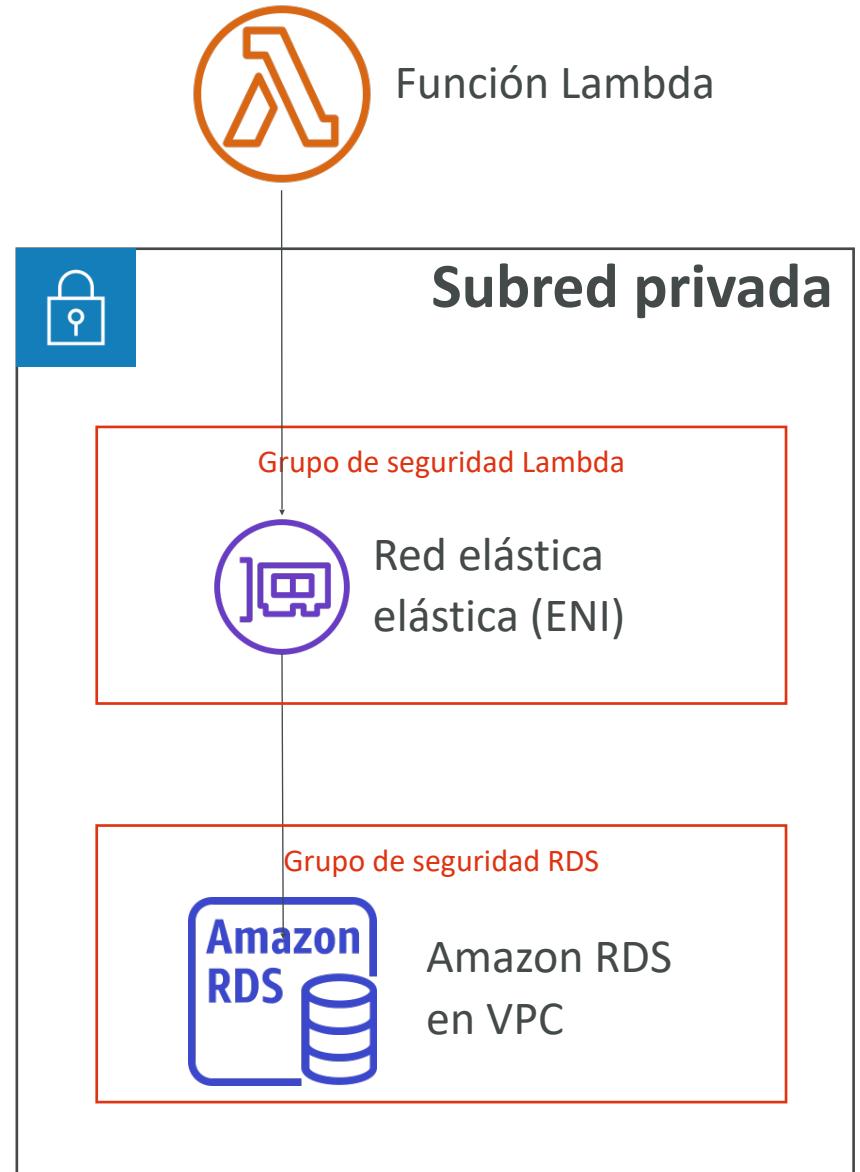
Lambda por defecto

- Por defecto, la función Lambda se lanza fuera de la propia VPC (en una VPC propiedad de AWS).
- Por lo tanto, no puedes acceder a los recursos de la VPC (RDS, ElastiCache, ELB interno...)



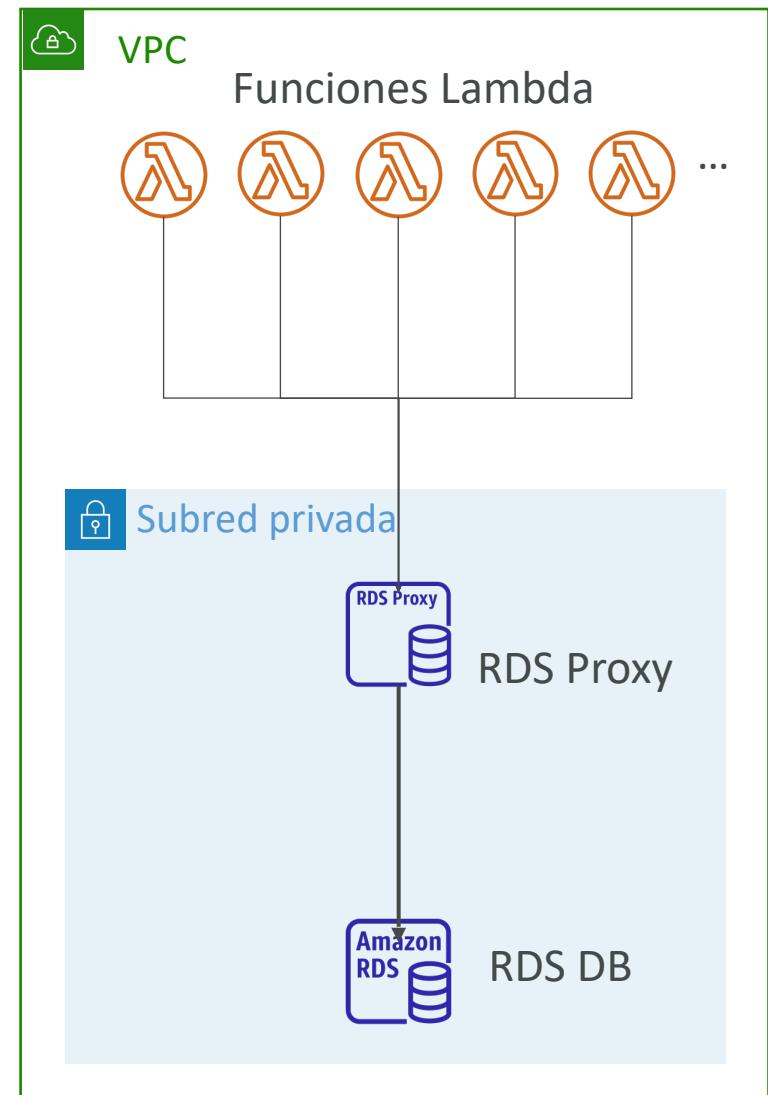
Lambda en VPC

- Debes definir el ID de la VPC, las subredes y los grupos de seguridad
- Lambda creará una ENI (Elastic Network Interface) en tus subredes



Lambda con proxy RDS

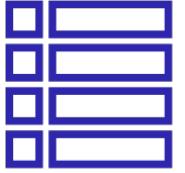
- Si las funciones Lambda acceden directamente a tu base de datos, pueden abrir demasiadas conexiones bajo carga elevada
- Proxy RDS
 - Mejora la escalabilidad agrupando y compartiendo las conexiones a la base de datos
 - Mejora la disponibilidad reduciendo en un 66% el tiempo de conmutación por error y conservando las conexiones
 - Mejora la seguridad aplicando la autenticación IAM y almacenando las credenciales en Secrets Manager
- **La función Lambda debe desplegarse en la VPC, ya que RDS Proxy nunca es accesible públicamente**





Amazon DynamoDB

- Totalmente gestionado, de alta disponibilidad con replicación a través de múltiples AZs
- Base de datos NoSQL - no relacional - con soporte de transacciones
- Escala a cargas de trabajo masivas, base de datos distribuida
- Millones de peticiones por segundo, billones de filas, cientos de TB de almacenamiento
- Rendimiento rápido y constante (milisegundos de un solo dígito)
- Integración con IAM para seguridad, autorización y administración
- Bajo coste y capacidad de autoescalado
- Sin mantenimiento ni parches, siempre disponible
- Clase de tabla de acceso estándar e infrecuente (IA)



DynamoDB - Conceptos básicos

- DynamoDB se compone de **Tablas**
- Cada tabla tiene una **clave primaria** (debe decidirse en el momento de la creación)
- Cada tabla puede tener un número infinito de elementos (= filas)
- Cada elemento tiene **atributos** (pueden añadirse con el tiempo - pueden ser nulos)
- El tamaño máximo de un elemento es de **400 KB**
- Los tipos de datos soportados son:
 - **Tipos escalares** - Cadena, Número, Binario, Booleano, Nulo
 - **Tipos de documento** - Lista, Mapa
 - **Tipos de conjuntos** - Conjunto de cadenas, Conjunto de números, Conjunto binario
- **Por lo tanto, en DynamoDB puedes evolucionar rápidamente los esquemas**

DynamoDB - Ejemplo de tabla

Clave primaria		Atributos	
Clave de partición	Clave de clasificación		
Usuario_ID	Juego_ID	Puntuación	Resultado
7791a3d6...	4421	92	Gana
873e0634...	1894	14	Pierde
873e0634...	4521	77	Gana

DynamoDB

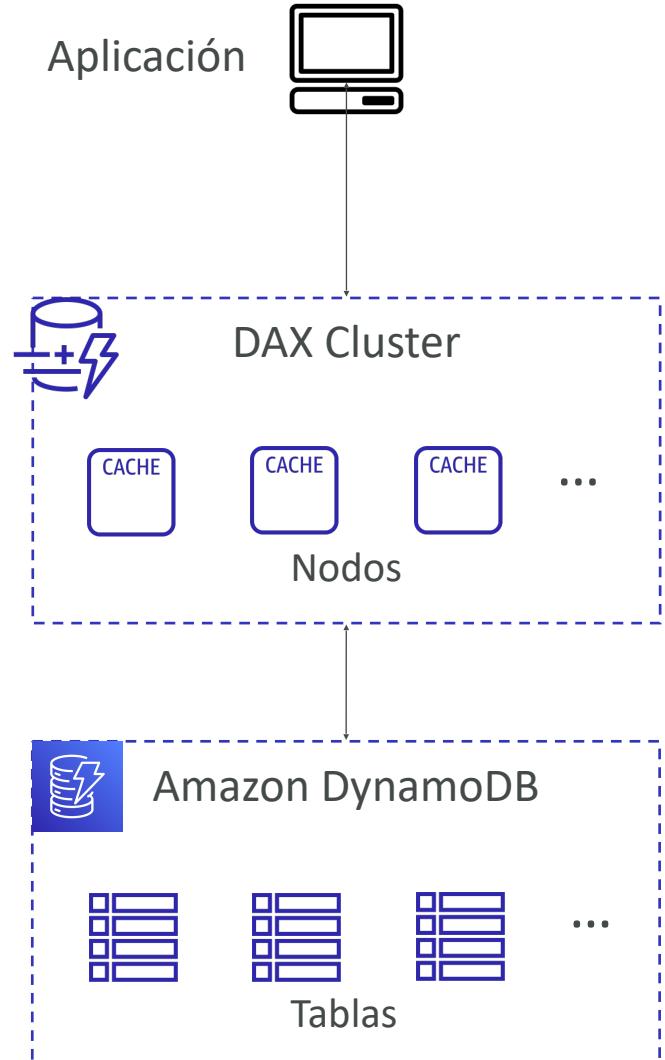
Modos de capacidad de lectura/escritura

- Controla cómo gestionar la capacidad de tu tabla (rendimiento de lectura/escritura)
- **Modo aprovisionado (por defecto)**
 - Especificas el número de lecturas/escrituras por segundo
 - Es necesario planificar la capacidad de antemano
 - Pagas por unidades de capacidad de lectura (RCU) y unidades de capacidad de escritura (WCU) provisionadas
 - Posibilidad de añadir el modo de autoescalado para RCU y WCU
- **Modo bajo demanda**
 - Las lecturas/escrituras aumentan/dismiñuyen automáticamente con tus cargas de trabajo
 - No es necesario planificar la capacidad
 - Pagas por lo que utilizas, más caro (\$\$\$)
 - Ideal para cargas de trabajo impredecibles, picos repentinos pronunciados

Acelerador de DynamoDB (DAX)



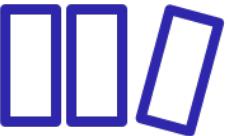
- Caché en memoria totalmente gestionada, de alta disponibilidad y sin interrupciones para DynamoDB
- **Ayuda a resolver la congestión de lectura mediante el almacenamiento en caché**
- **Latencia de microsegundos para los datos almacenados en caché**
- No requiere modificación de la lógica de la aplicación (compatible con las API de DynamoDB existentes)
- TTL de 5 minutos para la caché (predeterminado)



Acelerador DynamoDB (DAX) vs ElastiCache



DynamoDB - Procesamiento de flujos



- Flujo ordenado de modificaciones a nivel de artículo (crear/actualizar/borrar) en una tabla
- Casos prácticos:
 - Reaccionar a los cambios en tiempo real (correo electrónico de bienvenida a los usuarios)
 - Análisis de uso en tiempo real
 - Implementar la replicación entre regiones
 - Invocar AWS Lambda en los cambios de la tabla de DynamoDB

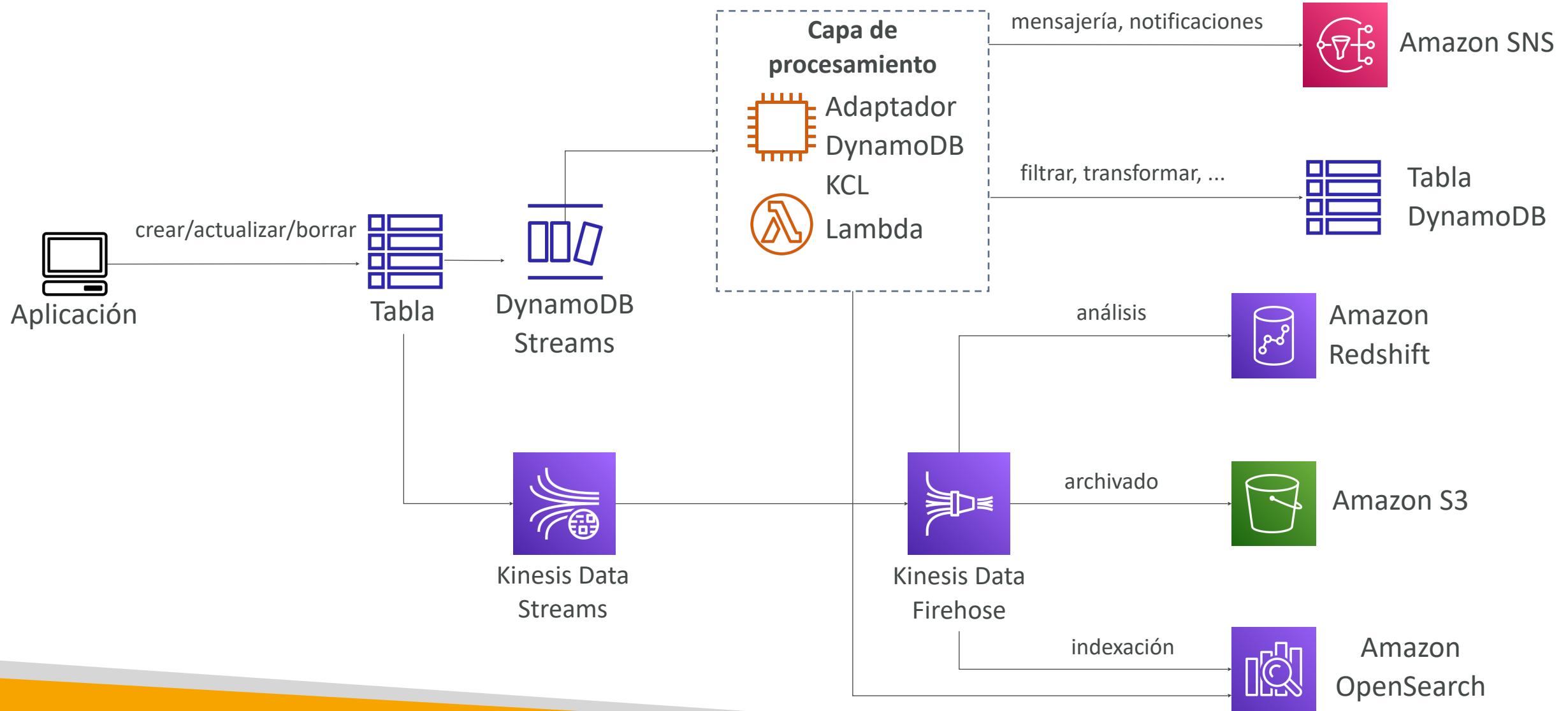
DynamoDB Streams

- Retención de 24 horas
- Número limitado de consumidores
- Procesamiento mediante activadores de AWS Lambda o el adaptador de Kinesis de DynamoDB Stream

Kinesis Data Streams (más reciente)

- Retención de 1 año
- Alto número de consumidores
- Proceso con AWS Lambda, Kinesis Data Analytics, Kinesis Data Firehose, AWS Glue Streaming ETL...

DynamoDB Streams



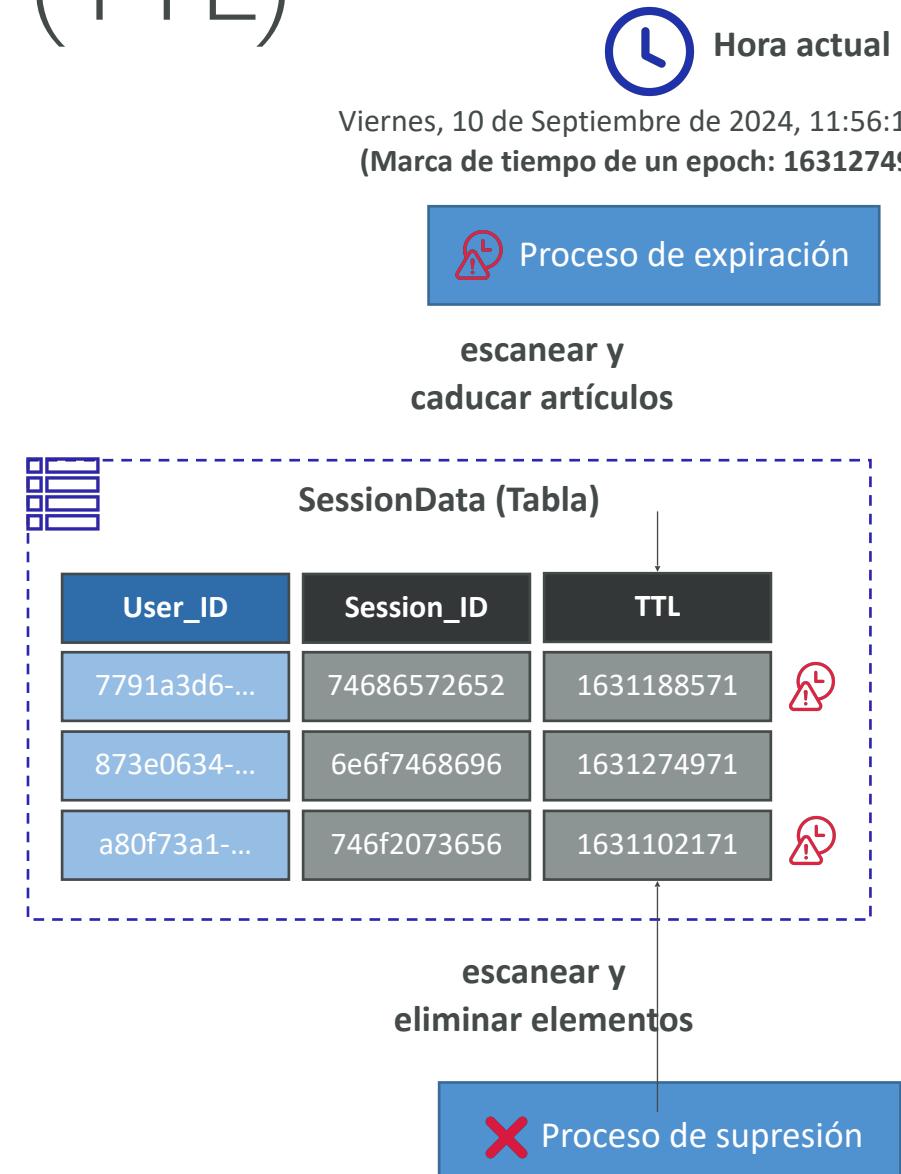
Tablas globales de DynamoDB



- Hacer accesible una tabla de DynamoDB con **baja latencia** en varias regiones
- Replicación activa-activa
- Las aplicaciones pueden **LEER** y **ESCRIBIR** en la tabla en cualquier región
- Debe habilitar DynamoDB Streams como requisito previo

DynamoDB - Tiempo de vida (TTL)

- Borrar automáticamente los elementos después de una fecha de caducidad
- Casos prácticos: reducir los datos almacenados conservando sólo los elementos actuales, cumplir las obligaciones normativas, gestión de sesiones web...



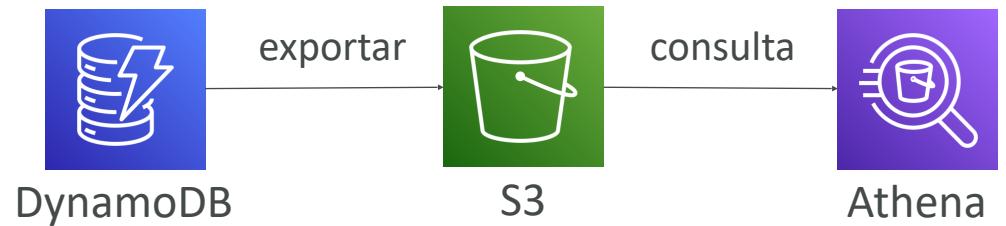
DynamoDB - Copias de seguridad para recuperación en caso de desastre

- Copias de seguridad continuas mediante recuperación puntual (PITR)
 - Activación opcional para los últimos 35 días
 - Recuperación puntual en cualquier momento dentro de la ventana de copia de seguridad
 - El proceso de recuperación crea una nueva tabla
- Copias de seguridad bajo demanda
 - Copias de seguridad completas para su conservación a largo plazo, hasta su eliminación explícita
 - No afecta al rendimiento ni a la latencia
 - Se puede configurar y administrar en AWS Backup (permite la copia entre regiones)
 - El proceso de recuperación crea una tabla nueva

DynamoDB - Integración con Amazon S3

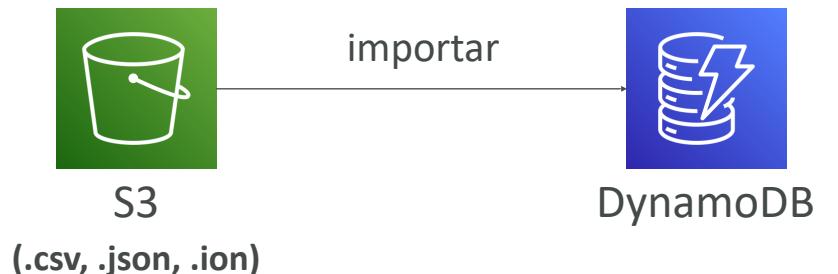
- **Exportación a S3 (debes habilitar PITR)**

- Funciona para cualquier momento de los últimos 35 días
- No afecta a la capacidad de lectura de tu tabla
- Conserva Snapshots para auditorías
- ETL sobre los datos de S3 antes de volver a importarlos a DynamoDB
- Exportación en formato DynamoDB JSON o ION



- **Importación a S3**

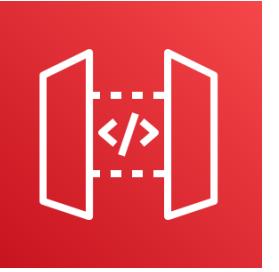
- Importación en formato CSV, DynamoDB JSON o ION
- No consume capacidad de escritura
- Crea una nueva tabla
- Los errores de importación se registran en CloudWatch Logs



Ejemplo: Creación de una API sin servidor



AWS API Gateway



- AWS Lambda + API Gateway: Sin infraestructura que administrar
- Compatibilidad con el protocolo WebSocket
- Gestión de versiones de API (v1, v2...)
- Gestión de diferentes entornos (desarrollo, pruebas, producción...)
- Gestión de la seguridad (autenticación y autorización)
- Creación de claves de API, gestión de la limitación de solicitudes
- Importación de Swagger / Open API para definir API rápidamente
- Transformación y validación de solicitudes y respuestas
- Generación de SDK y especificaciones de API
- Almacenamiento en caché de respuestas de API

API Gateway - Integraciones de alto nivel

- **Función Lambda**

- Invocar función Lambda
- Manera sencilla de exponer la API REST respaldada por AWS Lambda

- **HTTP**

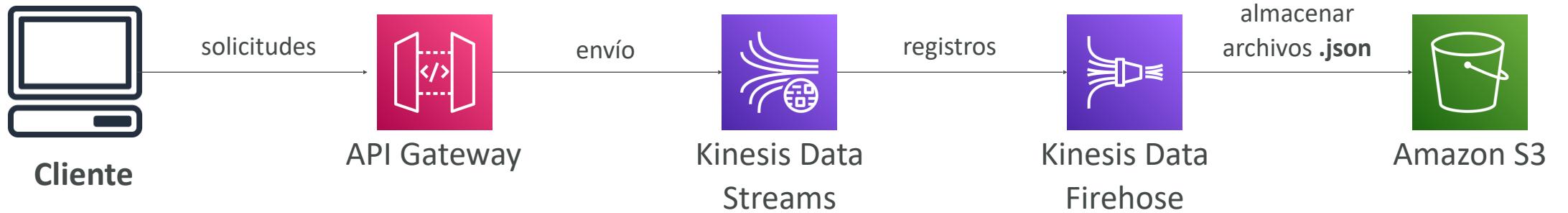
- Exponer puntos de enlace HTTP en el backend
- ¿Por qué? Añadir limitación de velocidad, almacenamiento en caché, autenticaciones de usuario, claves API, etc...

- **Servicio AWS**

- Exponer cualquier API de AWS a través de API Gateway
- Ejemplo: iniciar un flujo de trabajo AWS Step Function, enviar un mensaje a SQS
- ¿Por qué? Añadir autenticación, desplegar públicamente, control de tasa (rate) ...

Ejemplo de API Gateway

Integración de servicios de AWS Kinesis Data Streams



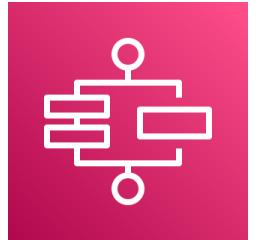
API Gateway - Tipos de endpoints

- **Edge-Optimized (por defecto):** Para clientes globales
 - Las solicitudes se enrutan a través de las ubicaciones de CloudFront Edge (mejora la latencia).
 - API Gateway sigue viviendo en una sola región
- **Regional:**
 - Para clientes dentro de la misma región
 - Podría combinarse manualmente con CloudFront (más control sobre las estrategias de almacenamiento en caché y la distribución)
- **Privada:**
 - Solo se puede acceder desde tu VPC utilizando un punto final de VPC de interfaz (ENI)
 - Utiliza una política de recursos para definir el acceso

API Gateway – Seguridad

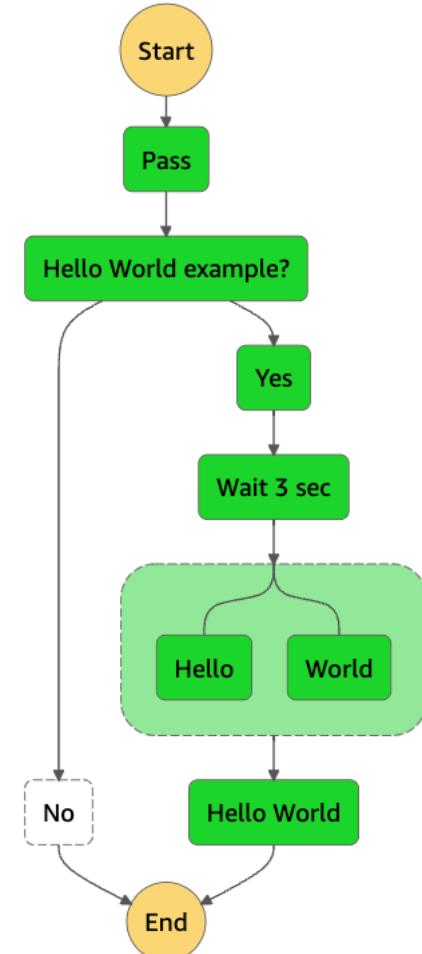
- **Autenticación de usuarios mediante**
 - Roles IAM (útil para aplicaciones internas)
 - Cognito (identidad para usuarios externos - ejemplo usuarios móviles)
 - Autorizador personalizado (tu propia lógica)
- **Seguridad HTTPS de nombre de dominio** personalizado a través de la integración con AWS Certificate Manager (ACM)
 - Si utilizas el punto de enlace Edge-Optimized, el certificado debe estar en **us-east-1**
 - Si utilizas el punto de enlace regional, el certificado debe estar en la región de API Gateway
 - Debes configurar el registro CNAME o A-alias en Route 53

AWS Step Functions



- Construye un flujo de trabajo visual sin servidor para orquestar tus funciones Lambda
- **Características:** secuencia, paralelo, condiciones, tiempos de espera, manejo de errores, ...
- Puede integrarse con EC2, ECS, servidores locales, API Gateway, colas SQS, etc...
- Posibilidad de implementar la función de aprobación humana
- **Casos de uso:** cumplimiento de pedidos, procesamiento de datos, aplicaciones web, cualquier flujo de trabajo

■ In Progress ■ Succeeded ■ Failed ■ Cancelled ■ Caught Error



Arquitecturas sin servidor

Aplicación móvil: MyTodoList

- Queremos crear una aplicación móvil con los siguientes requisitos
- Exponer como REST API con HTTPS
- Arquitectura sin servidor
- Los usuarios deberían poder interactuar directamente con su propia carpeta en S3
- Los usuarios deben autenticarse a través de un servicio gestionado sin servidor
- Los usuarios pueden escribir y leer *to-dos*, pero sobre todo leerlos
- La base de datos debe escalar y tener un alto rendimiento de lectura

Aplicación móvil: API REST



Aplicación móvil: acceso de los usuarios al S3



Aplicación móvil: alto rendimiento de lectura, datos estáticos



Aplicación móvil: almacenamiento en caché en la API Gateway



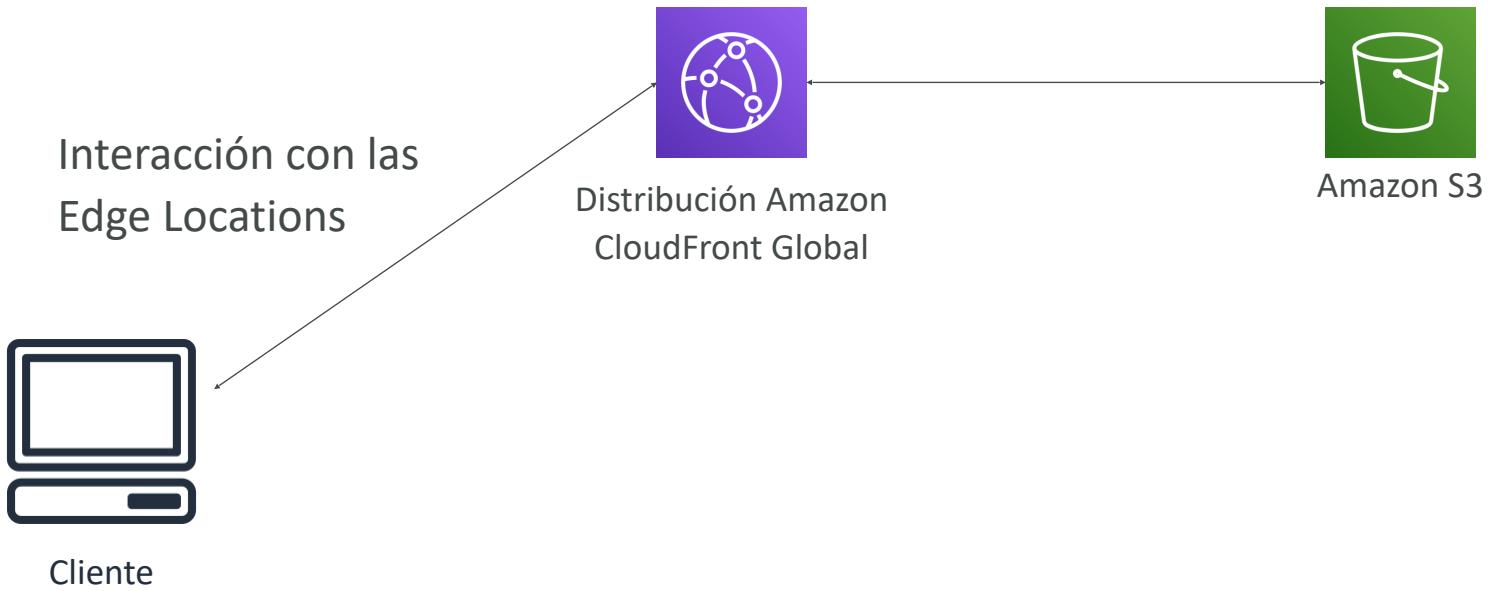
En esta clase...

- API REST sin servidor: HTTPS, API Gateway, Lambda, DynamoDB.
- Uso de Cognito para generar credenciales temporales con STS para acceder al bucket de S3 con política restringida. Los usuarios de la aplicación pueden acceder directamente a los recursos de AWS de esta forma
- Almacenamiento en caché de las lecturas en DynamoDB mediante DAX
- Almacenamiento en caché de las solicitudes REST en el nivel de API Gateway
- Seguridad para la autenticación y la autorización con Cognito, STS

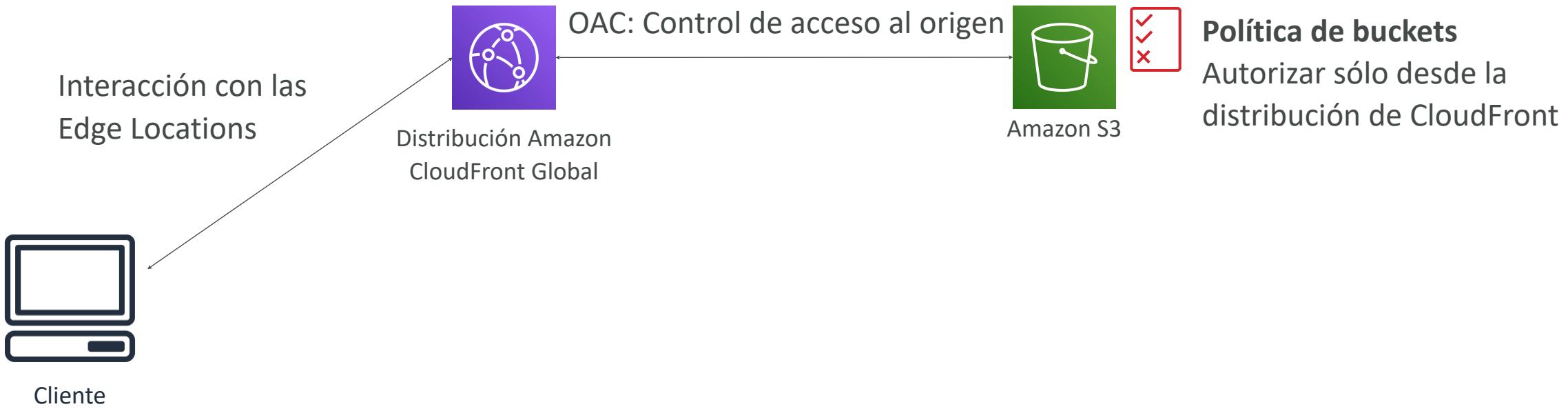
Sitio web alojado sin servidor: MyBlog.com

- Este sitio web debe escalar globalmente
- Los blogs rara vez se escriben, pero a menudo se leen
- Parte del sitio web son archivos estáticos, el resto es una API REST dinámica
- El almacenamiento en caché debe implementarse siempre que sea posible
- Cualquier nuevo usuario que se suscriba debe recibir un correo electrónico de bienvenida
- Cualquier foto subida al blog debe generar una miniatura

Servir contenido estático, globalmente



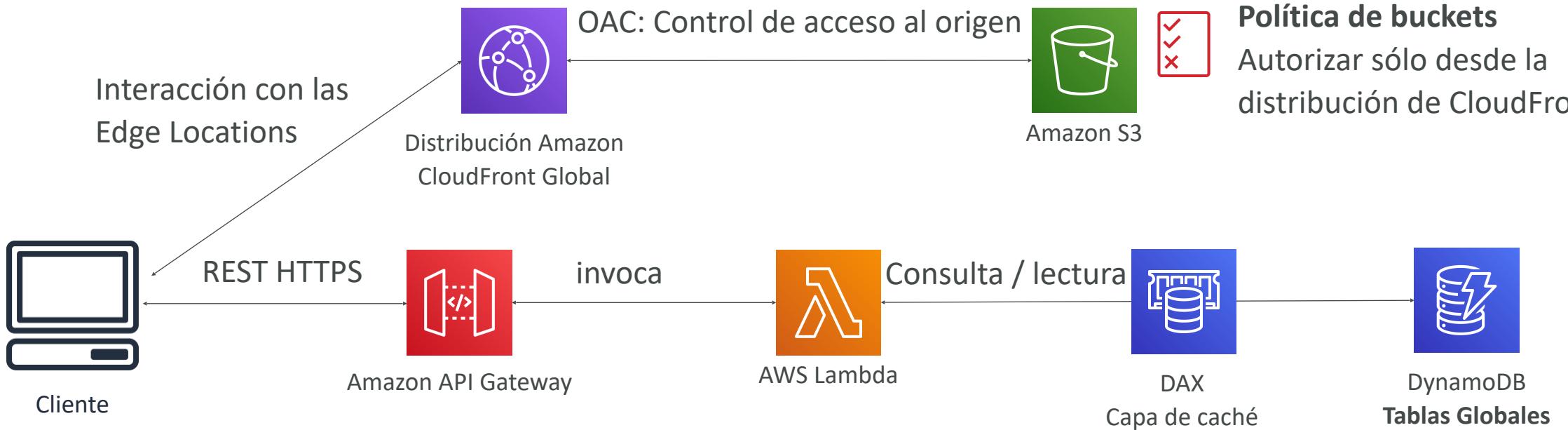
Servir contenidos estáticos de forma global y segura



Añadir una API REST pública sin servidor



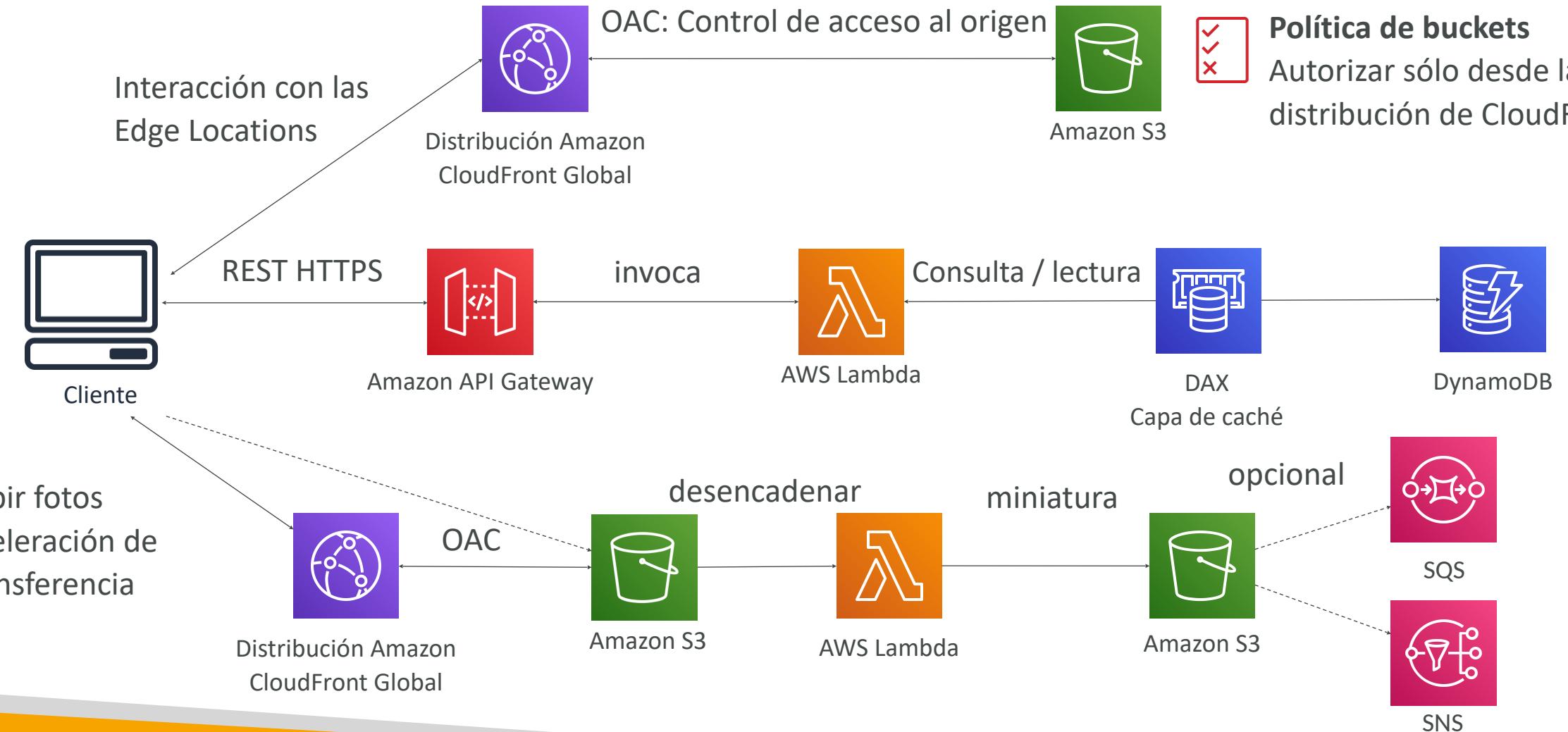
Aprovechamiento de las tablas globales de DynamoDB



Flujo de correo electrónico de bienvenida al usuario



Flujo de generación de miniaturas



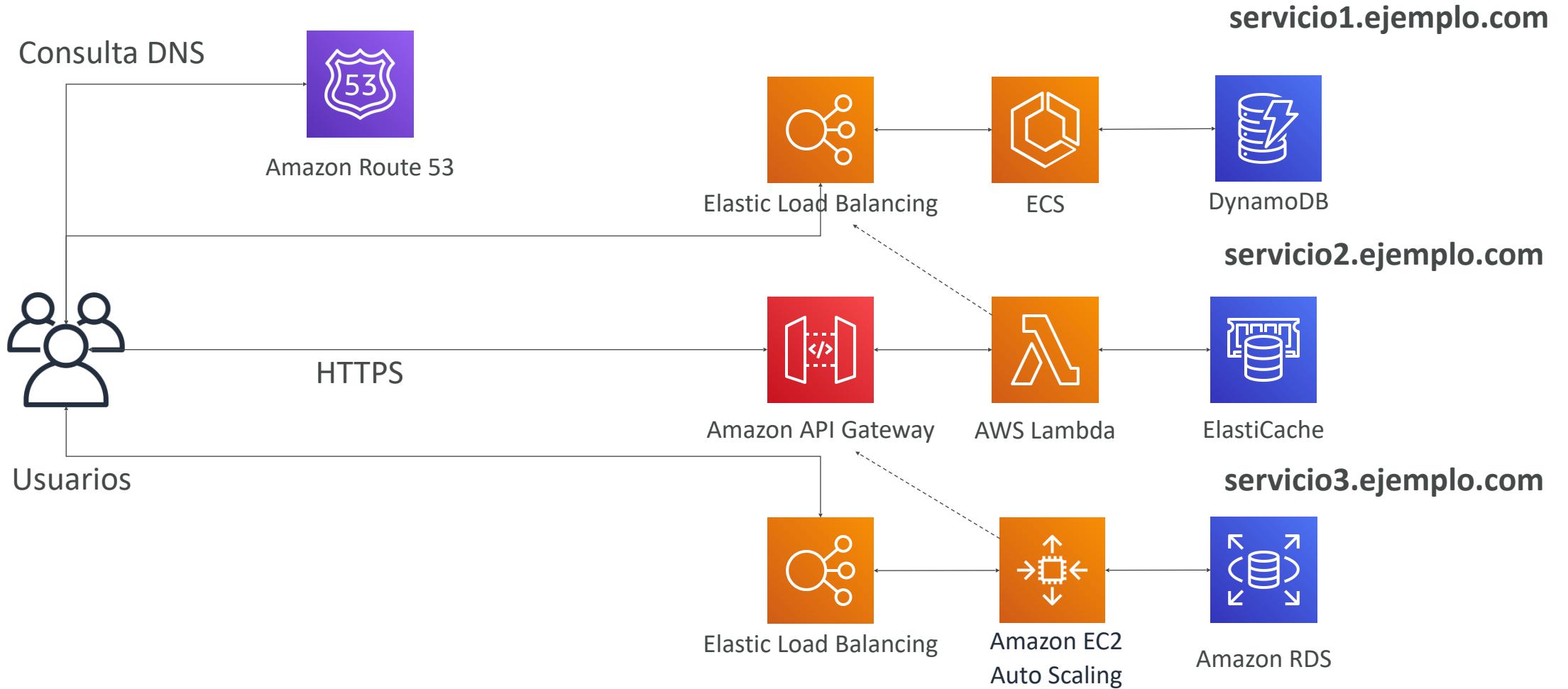
Resumen del sitio web hospedado en AWS

- Hemos visto contenido estático siendo distribuido usando CloudFront con S3
- La API REST era sin servidor, no necesitaba Cognito porque era pública.
- Aprovechamos una tabla Global DynamoDB para servir los datos globalmente
(podríamos haber utilizado Aurora Global Database)
- Habilitamos los flujos de DynamoDB para activar una función Lambda.
- La función Lambda tenía un rol IAM que podía utilizar SES.
- SES (Simple Email Service) se utilizó para enviar correos electrónicos de una manera sin servidor
- S3 puede activar SQS / SNS / Lambda para notificar eventos

Arquitectura de microservicios

- Queremos cambiar a una arquitectura de microservicios
- Muchos servicios interactúan entre sí directamente mediante una API REST
- Cada arquitectura para cada microservicio puede variar en forma y figura
- Queremos una arquitectura de microservicios para poder tener un ciclo de vida de desarrollo más ágil para cada servicio

Entorno de microservicios



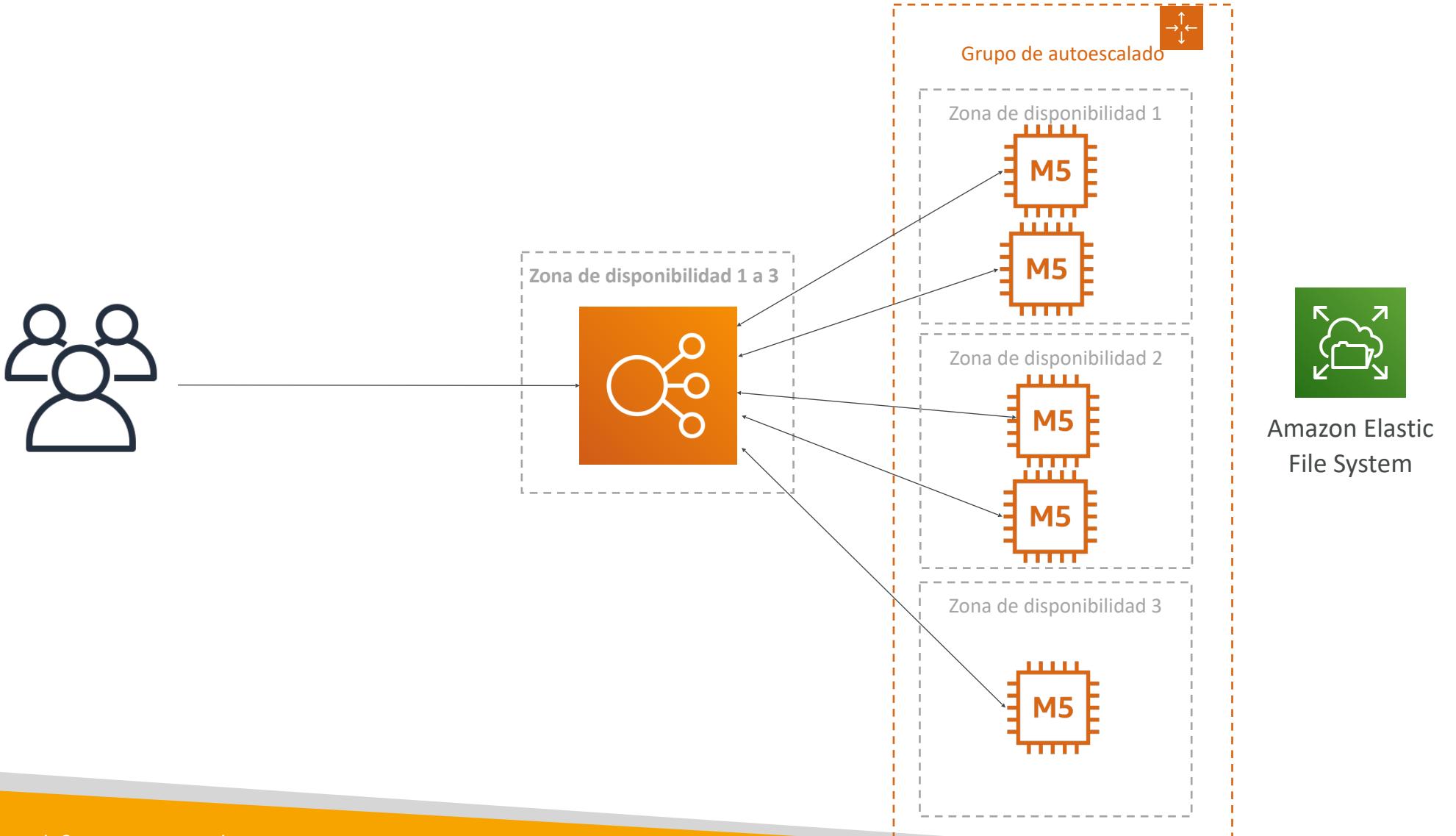
Debates sobre microservicios

- **Eres libre de diseñar cada microservicio como quieras**
- Patrones síncronos: API Gateway, Load Balancers
- Patrones asíncronos: SQS, Kinesis, SNS, Lambda triggers (S3)
- Desafíos con los microservicios:
 - sobrecarga repetida para crear cada nuevo microservicio,
 - problemas con la optimización de la densidad/utilización del servidor
 - complejidad de ejecutar varias versiones de varios microservicios simultáneamente
 - proliferación de requisitos de código del lado del cliente para integrarse con muchos servicios distintos.
- Algunos de los retos se resuelven con los patrones Serverless:
 - API Gateway, Lambda escalan automáticamente y pagas por uso
 - Puedes clonar API fácilmente, reproducir entornos
 - SDK de cliente generado a través de la integración de Swagger para API Gateway

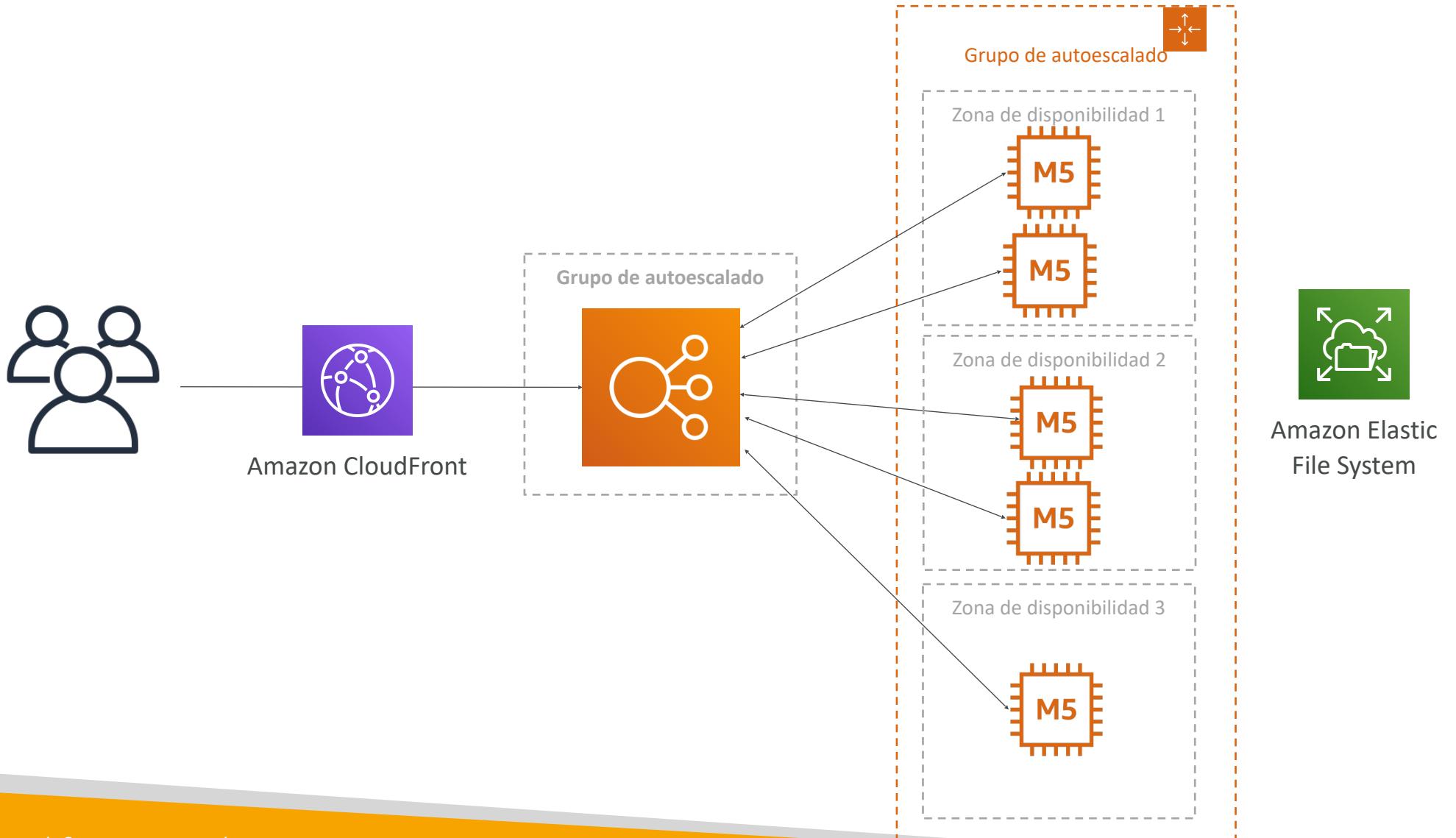
Descarga de actualizaciones de software

- Tenemos una aplicación que se ejecuta en EC2, que distribuye actualizaciones de software de vez en cuando
- Cuando sale una nueva actualización de software, recibimos muchas peticiones y el contenido se distribuye en masa por la red. Es muy costoso
- No queremos cambiar nuestra aplicación, pero queremos optimizar nuestro coste y CPU, ¿cómo podemos hacerlo?

Estado actual de nuestra aplicación



Una forma fácil de solucionar las cosas



¿Por qué CloudFront?

- Sin cambios en la arquitectura
- Almacenará en caché los archivos de actualización de software en el borde
- Los archivos de actualización de software no son dinámicos, son estáticos (nunca cambian)
- Nuestras instancias EC2 no son sin servidor
- Pero CloudFront sí, y escalará para nosotros
- Nuestro ASG no escalará tanto, y ahorraremos enormemente en EC2
- También ahorraremos en disponibilidad, coste de ancho de banda de red, etc.
- ¡Manera fácil de hacer una aplicación existente más escalable y más barata!

Bases de datos

Elegir la base de datos adecuada

- Tenemos muchas bases de datos administradas en AWS para elegir
- Preguntas para elegir la base de datos adecuada en función de tu arquitectura:
 - ¿Mucha lectura, mucha escritura o carga de trabajo equilibrada? ¿Necesidades de rendimiento? ¿Va a cambiar, necesita escalar o fluctuar durante el día?
 - ¿Cuántos datos almacenar y durante cuánto tiempo? ¿Va a crecer? ¿Tamaño medio de los objetos? ¿Cómo se accede a ellos?
 - ¿Durabilidad de los datos?
 - ¿Requerimientos de latencia? ¿Usuarios simultáneos?
 - ¿Modelo de datos? ¿Cómo se consultarán los datos? ¿Juntas? ¿Estructurados? ¿Semi-estructurados?
 - ¿Esquema sólido? ¿Más flexibilidad? ¿Información? ¿Búsqueda? ¿RDBMS / NoSQL?
 - ¿Costes de licencia? ¿Cambiar a una base de datos nativa de la nube como Aurora?



Tipos de bases de datos

- **RDBMS (= SQL / OLTP):** RDS, Aurora - ideal para uniones
- **Base de datos NoSQL - sin uniones, sin SQL:** DynamoDB (~JSON), ElastiCache (pares clave / valor), Neptune (gráficos), DocumentDB (para MongoDB), Keyspaces (para Apache Cassandra)
- **Almacén de objetos:** S3 (para objetos grandes) / Glacier (para copias de seguridad / archivos)
- **Almacén de datos - Data Warehouse** (= SQL Analytics / BI): Redshift (OLAP), Athena, EMR
- **Búsquedas:** OpenSearch (JSON) - texto libre, búsquedas no estructuradas
- **Gráficos:** Amazon Neptune - muestra las relaciones entre los datos
- **Ledger:** Base de datos de Amazon Quantum Ledger
- **Series temporales:** Amazon Timestream
- Nota: algunas bases de datos se tratan en la sección Datos y análisis

Amazon RDS – Resumen



- PostgreSQL Gestionado / MySQL / Oracle / SQL Server / MariaDB / Personalizado
- Tamaño de instancia RDS aprovisionada y tipo y tamaño de volumen EBS
- Capacidad de autoescalamiento para almacenamiento
- Soporte para réplicas de lectura y Multi AZ
- Seguridad a través de IAM, Grupos de Seguridad, KMS , SSL en tránsito
- Copia de seguridad automatizada con función de restauración puntual (hasta 35 días)
- Instantánea manual de la base de datos para una recuperación a largo plazo
- Mantenimiento gestionado y programado (con tiempo de inactividad)
- Soporte para autenticación IAM, integración con Secrets Manager
- RDS Custom para acceder y personalizar la instancia subyacente (Oracle y SQL Server)
- Caso de uso: Almacenar conjuntos de datos relacionales (RDBMS / OLTP), realizar consultas SQL, transacciones

Amazon Aurora – Resumen



- API compatible para PostgreSQL / MySQL, separación de almacenamiento y computación
- Almacenamiento: los datos se almacenan en 6 réplicas, a través de 3 AZ - alta disponibilidad, auto-reparación, auto-escalado
- Computación: Clúster de Instancia DB a través de múltiples AZ, auto-escalado de réplicas de lectura
- Clúster: Puntos finales personalizados para instancias de base de datos escritoras y lectoras
- Mismas características de seguridad / monitorización / mantenimiento que RDS
- Conozca las opciones de copia de seguridad y restauración de Aurora
- Aurora Serverless - para cargas de trabajo impredecibles / intermitentes, sin planificación de capacidad
- Aurora Multi-Master: para conmutación por error de escritura continua (alta disponibilidad de escritura)
- Aurora Global: hasta 16 instancias de lectura de BD en cada región, replicación de almacenamiento < 1 segundo
- Aurora Machine Learning: realiza ML usando SageMaker & Comprehend en Aurora
- Aurora Database Cloning: nuevo cluster a partir de uno existente, más rápido que restaurar una instantánea
- Caso de uso: igual que RDS, pero con menos mantenimiento / más flexibilidad / más rendimiento / más funciones

Amazon ElastiCache - Resumen



- Managed Redis / Memcached (oferta similar a RDS, pero para cachés)
- Almacén de datos en memoria, latencia de submilisegundos
- Debe aprovisionar un tipo de instancia EC2
- Soporte para Clustering (Redis) y Multi AZ, Read Replicas (sharding)
- Seguridad a través de IAM, Grupos de Seguridad, KMS, Redis Auth
- Función de copia de seguridad / instantánea / restauración puntual
- Mantenimiento gestionado y programado
- Requiere algunos cambios en el código de la aplicación para ser aprovechado
- Caso de uso: Almacén de claves/valores, lecturas frecuentes, menos escrituras, caché de resultados para consultas a la base de datos, almacenamiento de datos de sesión para sitios web, no puede utilizar SQL.



Amazon DynamoDB – Resumen

- Tecnología propietaria de AWS, base de datos NoSQL sin servidor administrada, latencia de milisegundos
- Modos de capacidad: capacidad aprovisionada con autoescalado opcional o capacidad bajo demanda
- Puede sustituir a ElastiCache como almacén de claves/valores (almacenamiento de datos de sesión, por ejemplo, utilizando la función TTL)
- Alta disponibilidad, Multi AZ por defecto, lectura y escritura desacopladas, capacidad de transacción
- Clúster DAX para caché de lectura, latencia de lectura de microsegundos
- Seguridad, autenticación y autorización a través de IAM
- Procesamiento de eventos: DynamoDB Streams para integrarse con AWS Lambda, o Kinesis Data Streams
- Función de tabla global: configuración activa-activa
- Copias de seguridad automatizadas de hasta 35 días con PITR (restauración a una tabla nueva), o copias de seguridad bajo demanda
- Exportación a S3 sin usar RCU dentro de la ventana PITR, importación desde S3 sin usar WCU
- Excelente para evolucionar rápidamente los esquemas
- Caso de uso: desarrollo de aplicaciones sin servidor (documentos pequeños de 100 KB), caché distribuida sin servidor, no dispone de lenguaje de consulta SQL

Amazon S3 – Resumen



- S3 es un... almacén de claves / valores para objetos
- Genial para objetos grandes, no tan genial para muchos objetos pequeños
- Sin servidor, escala infinitamente, tamaño máximo de objeto 5 TB, capacidad de versionado
- Niveles: S3 Standard, S3 Infrequent Access, S3 Intelligent, S3 Glacier + política de ciclo de vida
- Funciones: Versionado, Cifrado, Replicación, Eliminación de MFA, Registros de Acceso...
- Seguridad: IAM, políticas de bucket, ACL, puntos de acceso, Object Lambda, CORS, Object/Vault Lock
- Cifrado: SSE-S3, SSE-KMS, SSE-C, del lado del cliente, TLS en tránsito, cifrado por defecto
- Operaciones por lotes en objetos mediante S3 Batch, listado de archivos mediante S3 Inventory
- Rendimiento: Carga multiparte, aceleración de transferencias de S3, S3 Select
- Automatización: Notificaciones de eventos S3 (SNS, SQS, Lambda, EventBridge)
- Casos de uso: archivos estáticos, almacén de valores clave para archivos grandes, alojamiento de sitios web

DocumentDB



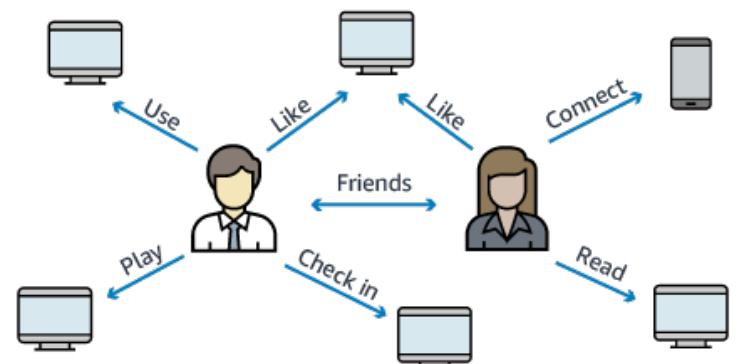
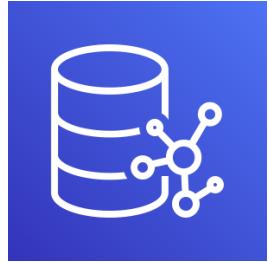
- Aurora es una “implementación de AWS” de PostgreSQL / MySQL ...
- **DocumentDB es lo mismo para MongoDB (que es una base de datos NoSQL)**

- MongoDB se utiliza para almacenar, consultar e indexar datos JSON
- Conceptos de despliegue similares a los de Aurora
- Totalmente gestionado, de alta disponibilidad con replicación a través de 3 AZ
- El almacenamiento de DocumentDB crece automáticamente en incrementos de 10 GB, hasta 64 TB.

- Escala automáticamente a cargas de trabajo con millones de peticiones por segundo

Amazon Neptune

- Base de datos **gráfica** totalmente gestionada
- Un **conjunto de datos de grafos** popular sería una **red social**
 - Los usuarios tienen amigos
 - Las publicaciones tienen comentarios
 - Los comentarios tienen likes de los usuarios
 - Los usuarios comparten y les gustan las publicaciones...
- Alta disponibilidad a través de 3 AZ, con hasta 15 réplicas de lectura
- Construye y ejecuta aplicaciones que trabajan con conjuntos de datos altamente conectados - optimizados para estas consultas complejas y difíciles
- Puede almacenar hasta miles de millones de relaciones y consultar el grafo con una latencia de milisegundos.
- Alta disponibilidad con réplicas a través de múltiples AZs
- Excelente para grafos de conocimiento (Wikipedia), detección de fraudes, motores de recomendación y redes sociales.



Amazon Keyspaces (para Apache Cassandra)

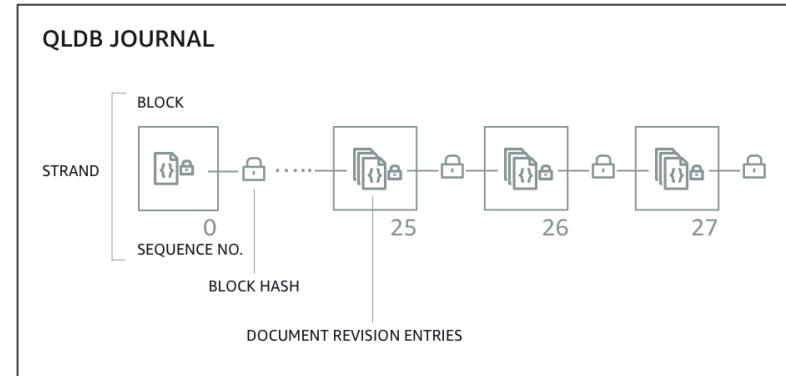


- Apache Cassandra es una base de datos distribuida NoSQL de código abierto
- Un servicio de base de datos administrado compatible con Apache Cassandra
- Sin servidor, escalable, de alta disponibilidad, totalmente administrado por AWS
- Escala automáticamente las tablas hacia arriba/abajo en función del tráfico de la aplicación
- Las tablas se replican 3 veces en múltiples AZ
- Uso del lenguaje de consulta Cassandra (CQL)
- Latencia de un milisegundo a cualquier escala, miles de solicitudes por segundo
- Capacidad: Modo bajo demanda o modo provisionado con autoescalado
- Cifrado, copia de seguridad, recuperación puntual (PITR) de hasta 35 días
- Casos de uso: almacenar información de dispositivos IoT, datos de series temporales, ...

Amazon QLDB



- QLDB son las siglas de "Quantum Ledger Database".
- Un libro mayor es un libro que **registra las transacciones financieras**
- Totalmente gestionado, sin servidor, de alta disponibilidad, replicación a través de 3 AZ
- Se utiliza para **revisar el historial de todos los cambios realizados en los datos de su aplicación** a lo largo del tiempo
- Sistema **inmutable**: ninguna entrada puede ser eliminada o modificada, verificable criptográficamente



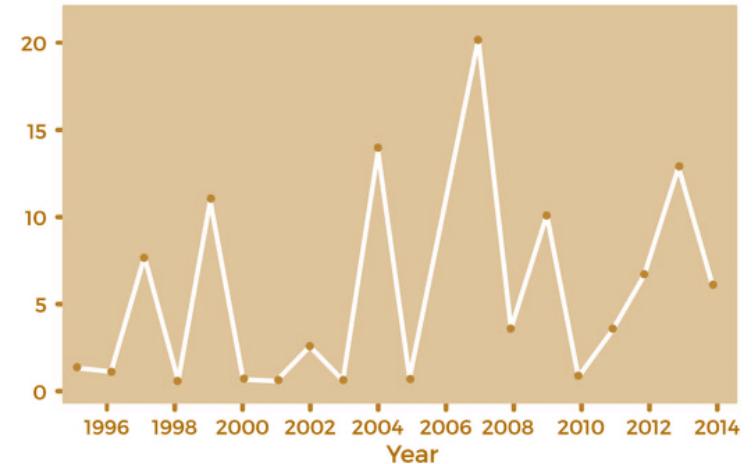
- Rendimiento 2-3 veces mejor que los marcos de blockchain de libro mayor común
- Diferencia con Amazon Managed Blockchain: **no hay componente de descentralización**, de acuerdo con las normas de regulación financiera

<https://docs.aws.amazon.com/qldb/latest/developerguide/ledger-structure.html>

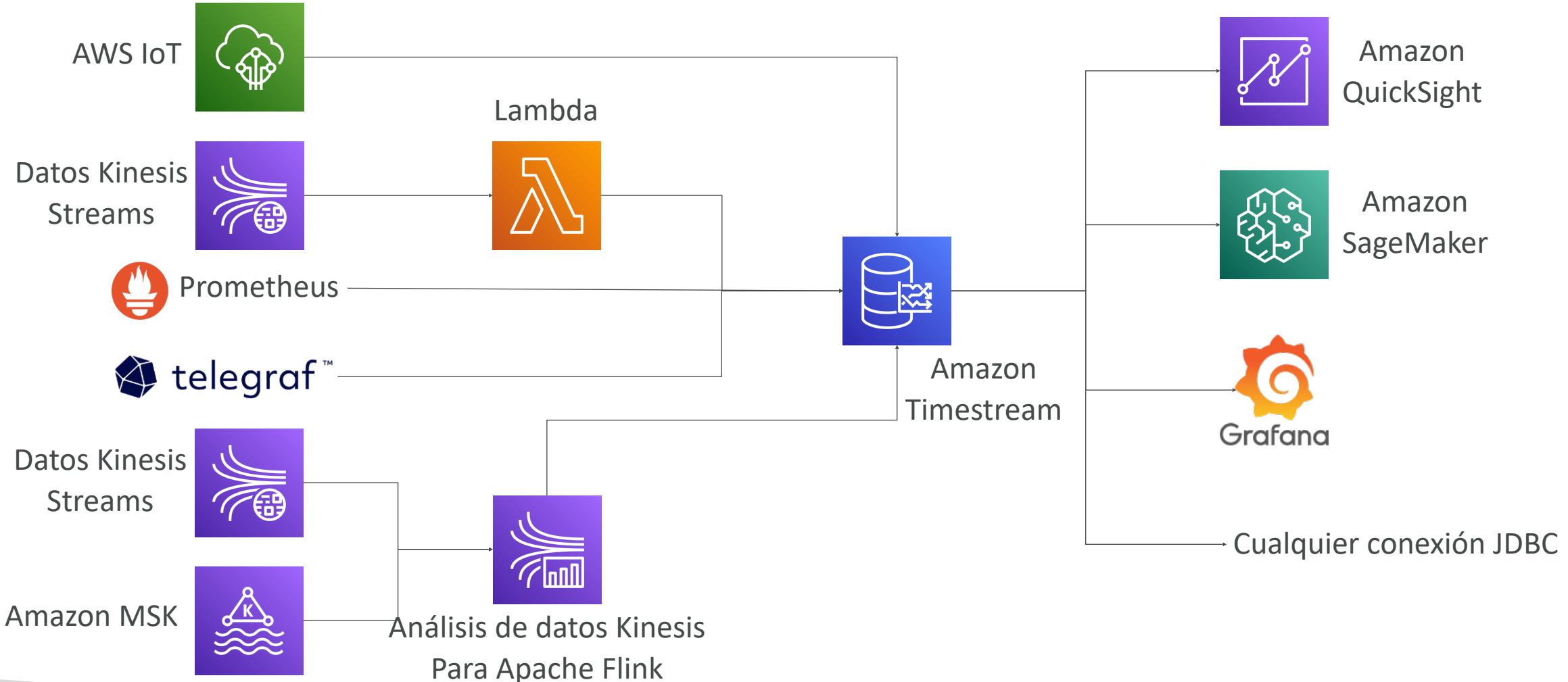
Amazon Timestream



- Base de datos de series temporales totalmente gestionada, rápida, escalable y sin servidor
- Se amplía y reduce automáticamente para ajustar la capacidad
- Almacena y analiza billones de eventos al día
- 1000 veces más rápida y 1/10 del coste de las bases de datos relacionales
- Consultas programadas, registros multimedida, compatibilidad con SQL
- Almacenamiento de datos por niveles: los datos recientes se guardan en la memoria y los históricos en un almacenamiento de coste optimizado
- Funciones integradas de análisis de series temporales (ayuda a identificar patrones en los datos casi en tiempo real)
- Cifrado en tránsito y en reposo
- Casos de uso: Aplicaciones IoT, aplicaciones operativas, análisis en tiempo real, ...



Amazon Timestream – Arquitectura



Datos y análisis

Amazon Athena



- Servicio de consulta **sin servidor** para analizar datos almacenados en Amazon S3.
- Utiliza el lenguaje SQL estándar para consultar los archivos
- Admite CSV, JSON, ORC, Avro y Parquet
- Precio: 5 dólares por TB de datos analizados
- Se utiliza habitualmente con Amazon Quicksight para la elaboración de informes y Dashboards
- **Casos de uso:** Inteligencia empresarial / análisis / informes, analizar y consultar registros de flujo de VPC, registros de ELB, **CloudTrail trails**, etc....
- **Sugerencia de examen:** analizar datos en S3 utilizando SQL sin servidor, utilizar Athena

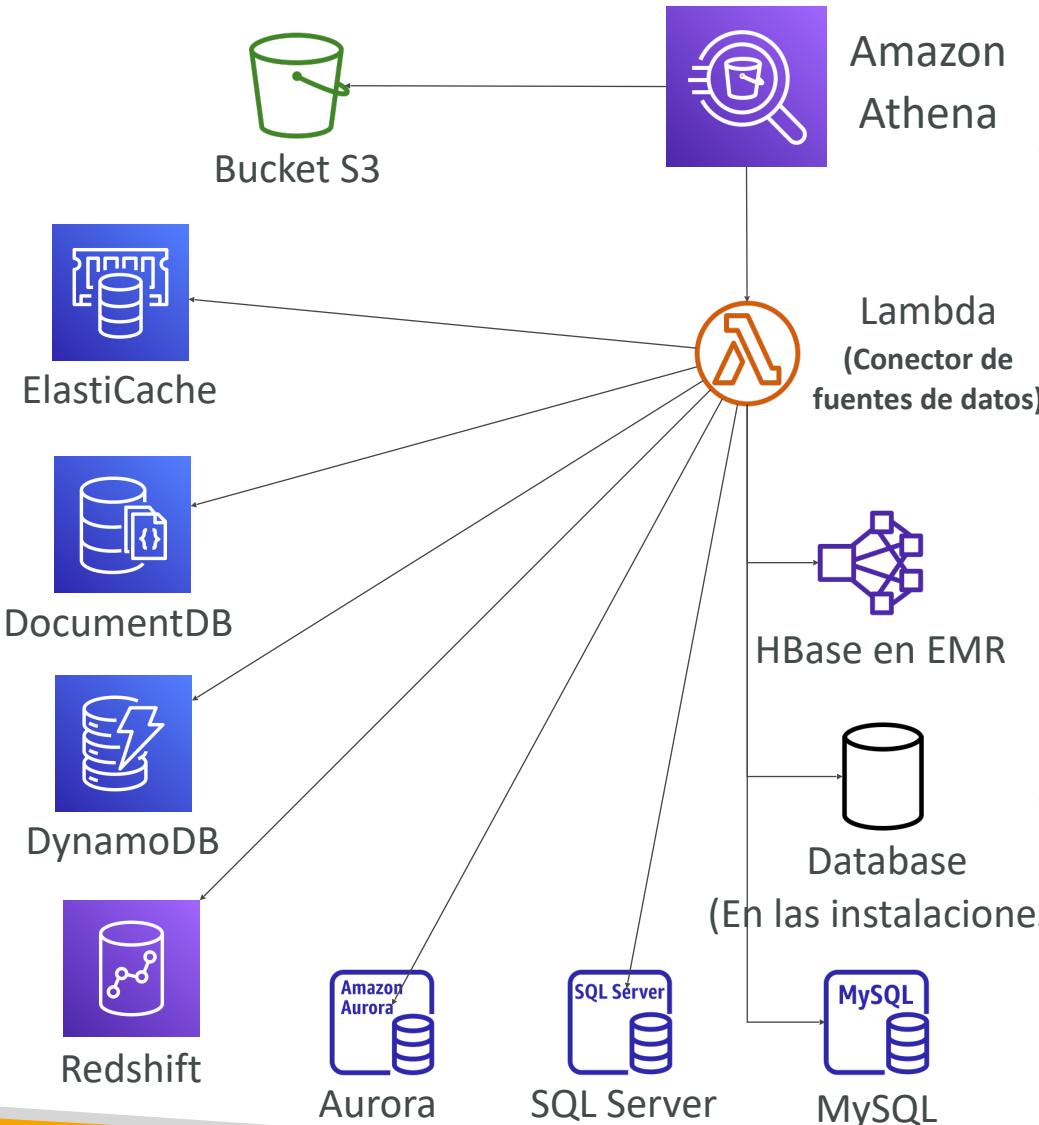


Amazon Athena – Mejora del rendimiento

- Utiliza **datos en columnas** para ahorrar costes
 - Se recomienda Apache Parquet u ORC
 - Enorme mejora del rendimiento
 - Utiliza Glue para convertir tus datos en Parquet u ORC
- **Comprime los datos** para recuperaciones más pequeñas (bzip2, gzip, lz4, snappy, zlip, zstd...)
- **Particionar** conjuntos de datos en S3 para facilitar la consulta en columnas virtuales
 - s3://yourBucket/pathToTable
 /<PARTITION_COLUMN_NAME>=<VALUE>
 /<PARTITION_COLUMN_NAME>=<VALUE>
 /<PARTITION_COLUMN_NAME>=<VALUE>
 /etc...
 - Ejemplo: s3://athena-examples/flight/parquet/year=1991/month=1/day=1/
- **Utiliza archivos de mayor tamaño (> 128 MB)** para minimizar la sobrecarga

Amazon Athena – Consulta federada

- Permite ejecutar consultas SQL en datos almacenados en fuentes de datos relacionales, no relacionales, de objetos y personalizadas (AWS o en las instalaciones).
- Utiliza conectores de fuentes de datos que se ejecutan en AWS Lambda para ejecutar consultas federadas (por ejemplo, CloudWatch Logs, DynamoDB, RDS, ...)
- Almacena los resultados de nuevo en Amazon S3

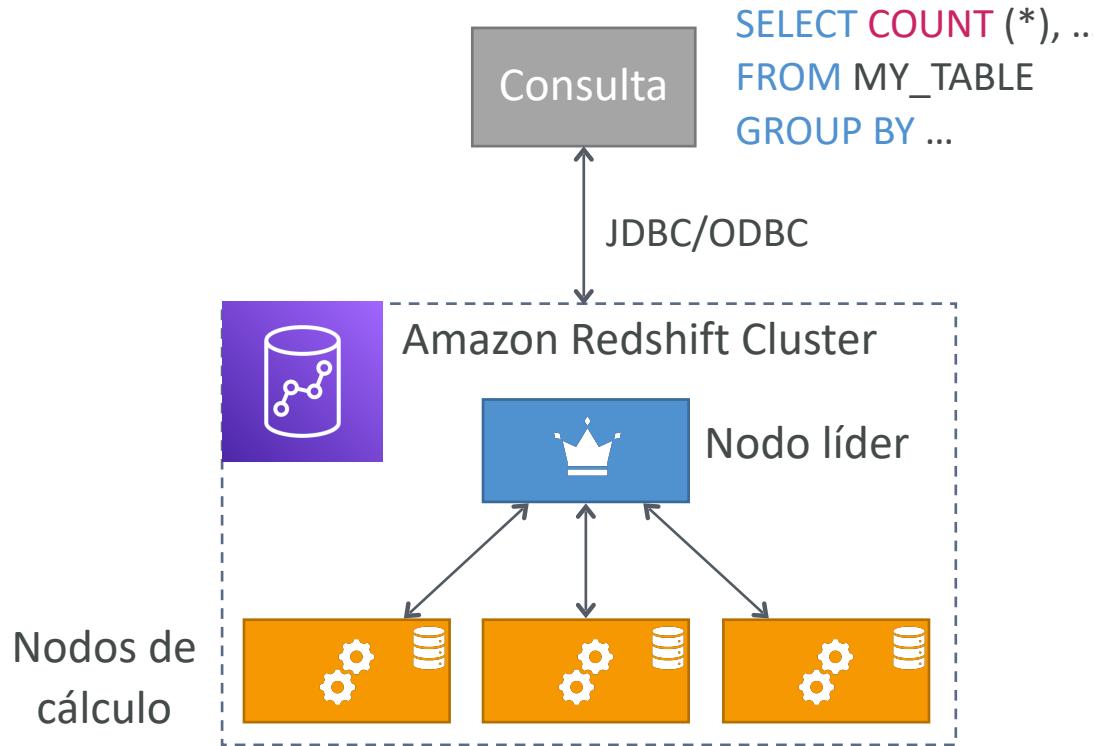


Visión general de Redshift



- Redshift se basa en PostgreSQL, pero **no se utiliza para OLTP**
- **Es OLAP - procesamiento analítico en línea (análisis y almacenamiento de datos)**
- 10 veces mejor rendimiento que otros almacenes de datos, escala a PBs de datos
- Almacenamiento de datos **en columnas** (en lugar de filas) y motor de consulta paralelo
- Pago por uso en función de las instancias aprovisionadas
- Dispone de una interfaz SQL para realizar consultas
- Se integra con herramientas de BI como Amazon Quicksight o Tableau
- **vs Athena:** consultas / uniones / agregaciones más rápidas gracias a los índices

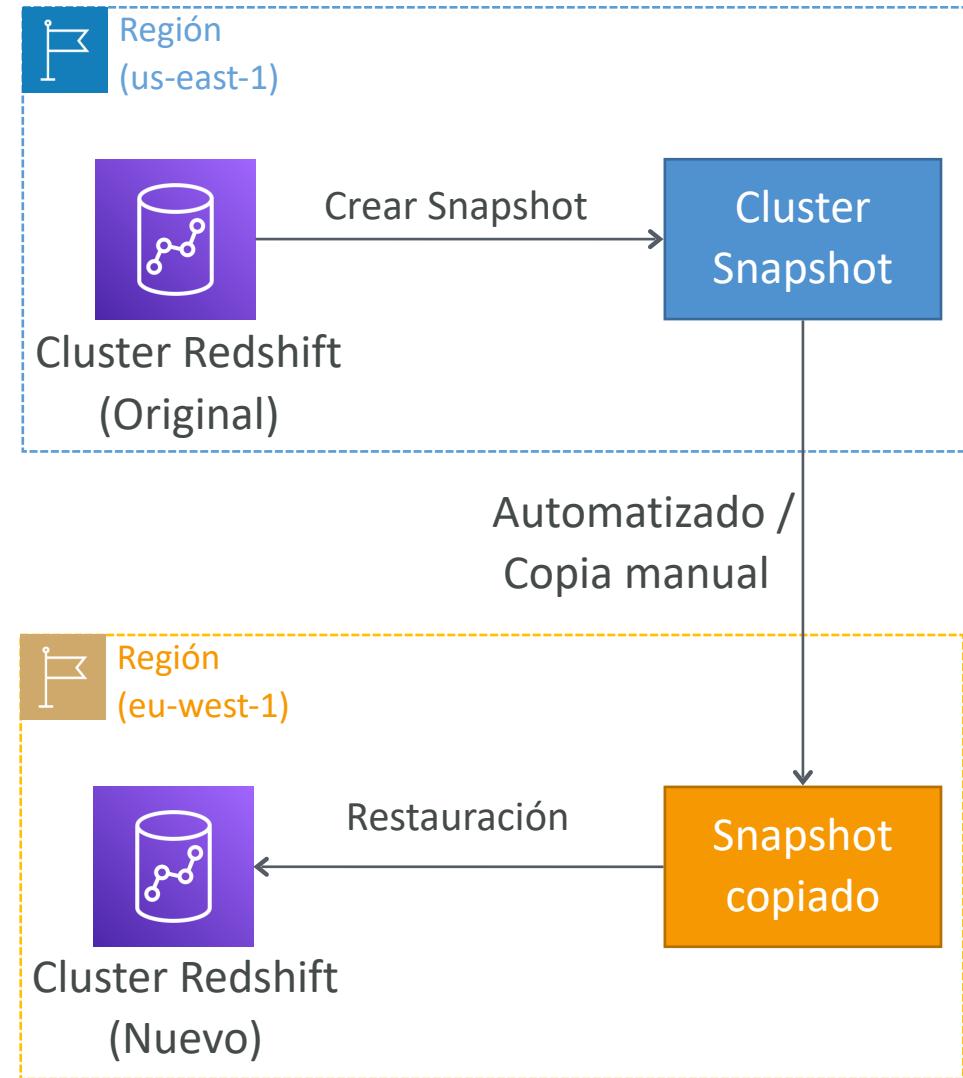
Clúster Redshift



- Nodo líder: planifica las consultas y agrega los resultados.
- Nodo de cálculo: realiza las consultas y envía los resultados al líder.
- Se aprovisiona el tamaño del nodo por adelantado
- Puedes utilizar instancias reservadas para ahorrar costes

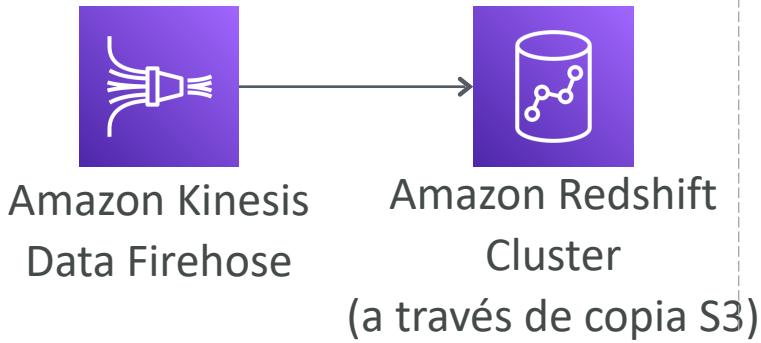
Redshift – Snapshots & Recuperación de desastres

- **Redshift no tiene modo Multi-AZ**
- Las instantáneas son copias de seguridad puntuales de un clúster, almacenadas internamente en S3
- Las instantáneas son incrementales (sólo se guarda lo que ha cambiado)
- Se puede restaurar una instantánea en un **nuevo clúster**
- Automatizado: cada 8 horas, cada 5 GB, o en un horario. Retención de 1 a 35 días
- Manual: la snapshot se conserva hasta que se elimina
- **Puedes configurar Amazon Redshift para que copie automáticamente las snapshots (automatizadas o manuales) de un clúster a otra región de AWS**

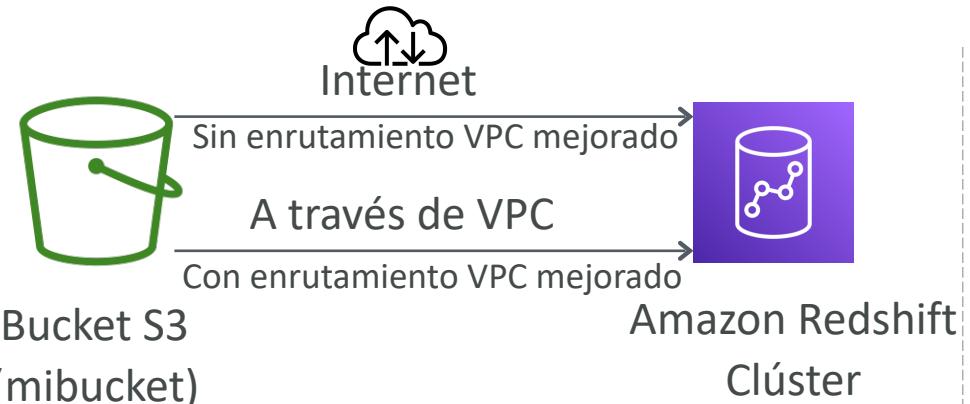


Carga de datos en Redshift: Las inserciones grandes son MUCHO mejores

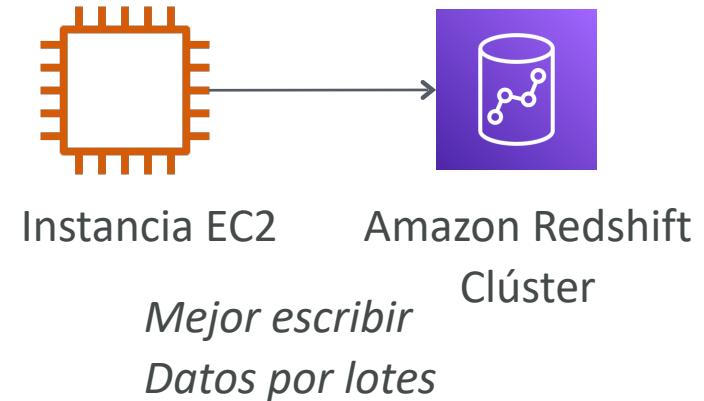
Amazon Kinesis Data Firehose



S3 utilizando el comando COPY



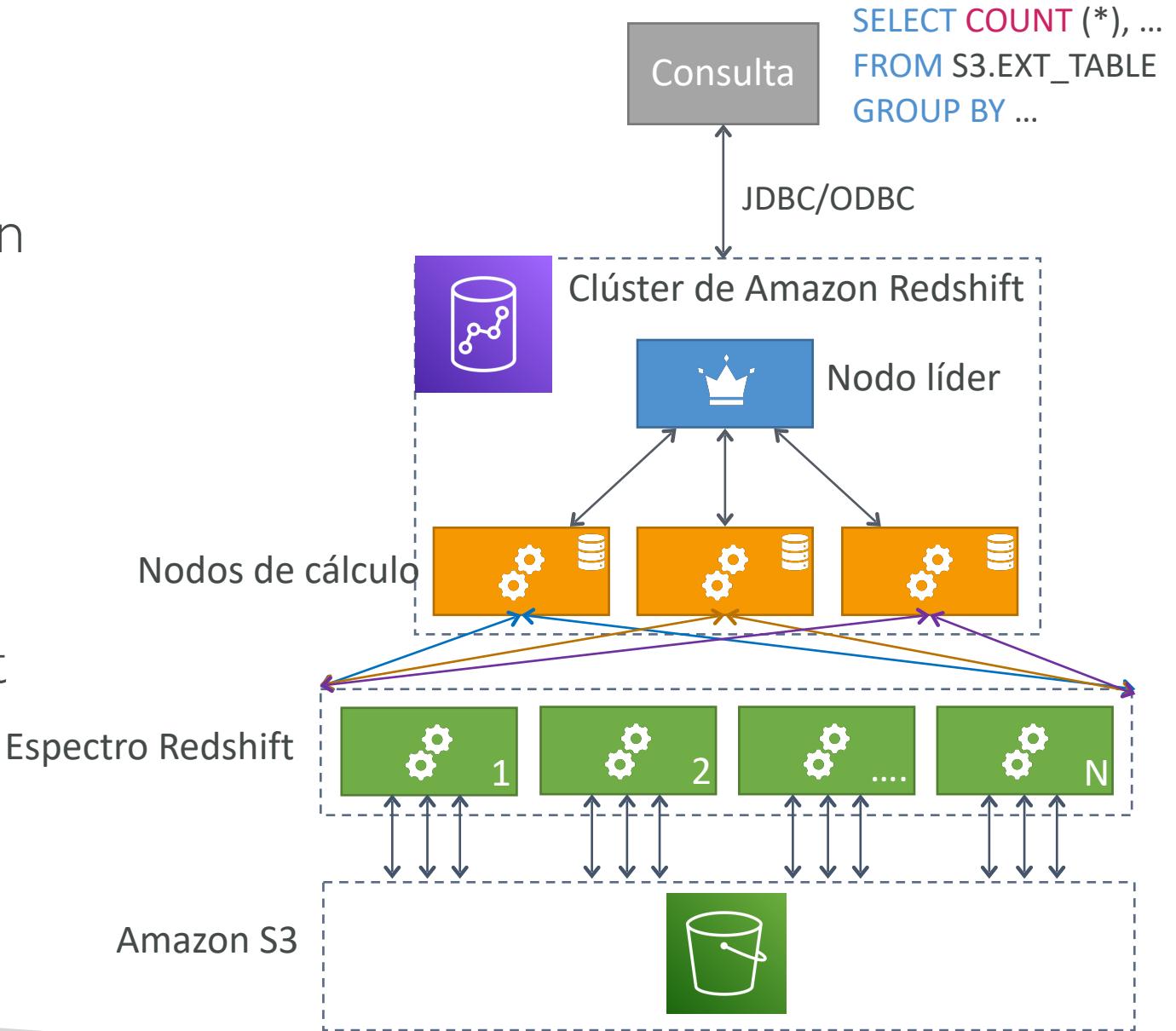
Instancia EC2 Controlador JDBC



```
copy customer
from 's3://mybucket/mydata'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

Espectro Redshift

- Consultar datos que ya están en S3 sin cargarlos
- **Debe haber un clúster Redshift disponible para iniciar la consulta**
- A continuación, la consulta se envía a miles de nodos Redshift Spectrum

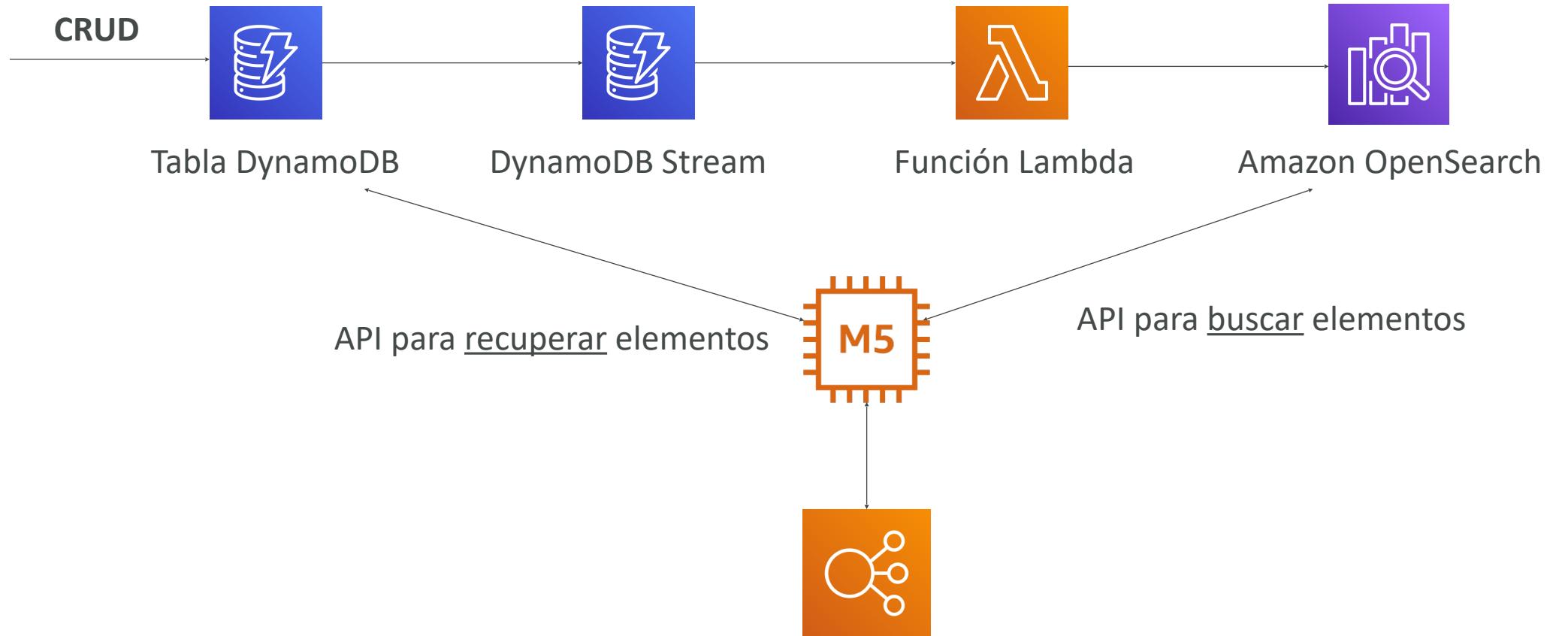


Amazon OpenSearch

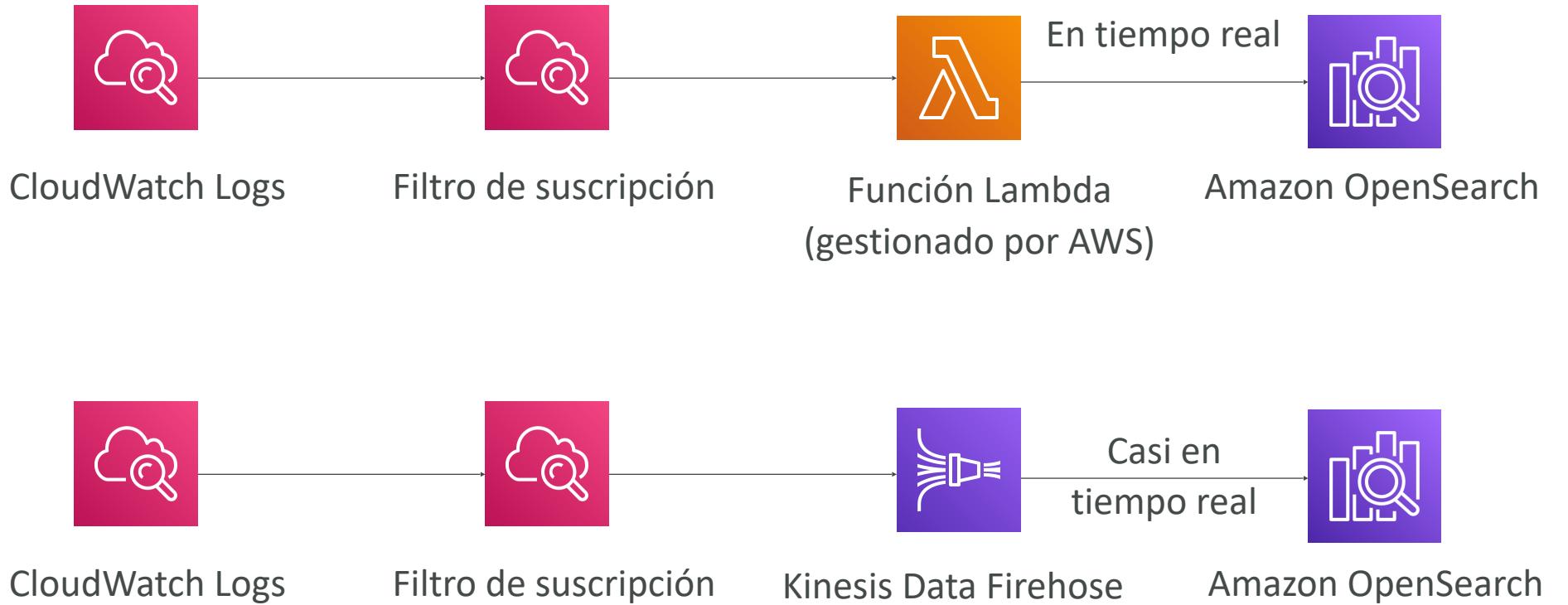


- **Amazon OpenSearch es el sucesor de Amazon Elasticsearch**
- En DynamoDB, las consultas sólo existen por clave primaria o índices...
- **Con OpenSearch, puedes buscar en cualquier campo, incluso coincidencias parciales**
- Es habitual utilizar OpenSearch como complemento de otra base de datos
- OpenSearch requiere un Cluster de instancias (no sin servidor)
- No soporta SQL (tiene su propio lenguaje de consulta)
- Ingestión desde Kinesis Data Firehose, AWS IoT y CloudWatch Logs
- Seguridad mediante Cognito & IAM, cifrado KMS, TLS
- Viene con OpenSearch Dashboards (visualización)

Patrones de OpenSearch DynamoDB

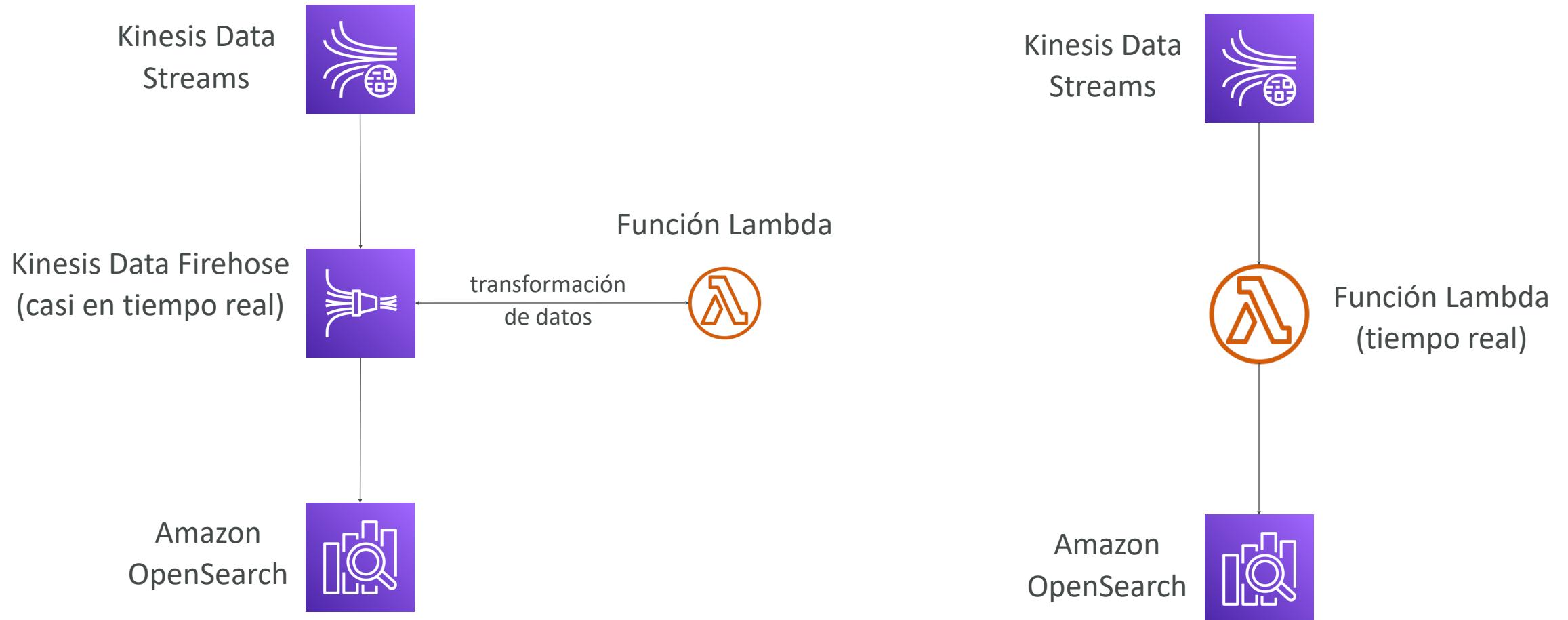


Patrones de OpenSearch CloudWatch Logs



Patrones de OpenSearch

Kinesis Data Streams y Kinesis Data Firehose



Amazon EMR



- EMR son las siglas de "Elastic MapReduce".
- EMR ayuda a **crear clusters Hadoop (Big Data)** para analizar y procesar grandes cantidades de datos
- Los clusters pueden estar formados por **cientos de instancias EC2**
- EMR viene con Apache Spark, HBase, Presto, Flink...
- EMR se encarga de todo el aprovisionamiento y configuración
- Auto-escalado e integrado con instancias Spot
- **Casos de uso: procesamiento de datos, aprendizaje automático, indexación web, big data...**

Amazon EMR - Tipos de nodos y opciones de compra

- **Nodo maestro:** Gestiona el clúster, coordina, gestiona la salud - larga duración
- **Nodo central:** Ejecuta tareas y almacena datos - larga duración
- **Nodo de tareas (opcional):** Sólo para ejecutar tareas - normalmente Spot
- **Opciones de compra:**
 - Bajo demanda: fiable, predecible, no se dará por terminado
 - Reservado (mínimo 1 año): ahorro de costes (EMR lo utilizará automáticamente si está disponible)
 - Instancias Spot: más baratas, pueden cancelarse, menos fiables
- Puedes tener un clúster de larga duración o un clúster transitorio (temporal).

Amazon QuickSight



- **Servicio de inteligencia empresarial basado en aprendizaje automático sin servidor para crear Dashboards interactivos**
- Rápido, escalable automáticamente, integrable y con precios por sesión
- Casos de uso:
 - Análisis empresarial
 - Creación de visualizaciones
 - Realizar análisis ad hoc
 - Obtener perspectivas de negocio utilizando datos
- Integrado con RDS, Aurora, Athena, Redshift, S3...
- **Computación en memoria utilizando el motor SPICE** si los datos se importan en QuickSight
- Edición Enterprise: Posibilidad de configurar la **seguridad a nivel de columna (CLS)**



<https://aws.amazon.com/quicksight/>

Integraciones de QuickSight

Fuentes de datos (servicios de AWS)



RDS



Aurora



Redshift



Athena



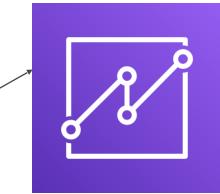
S3



OpenSearch



Timestream



QuickSight

teradata.



En las Bases de datos (JDBC)
de las instalaciones



XLSX



CSV



JSON

Fuentes de datos (importaciones)



.TSV



ELF & CLF
(Log Format)

QuickSight - Dashboards y análisis

- Definir usuarios (versiones estándar) y grupos (versión empresarial)
 - Estos usuarios y grupos sólo existen en QuickSight, no en IAM!!
- Un *dashboard*...
 - es una instantánea de sólo lectura de un análisis que se puede compartir
 - conserva la configuración del análisis (filtrado, parámetros, controles, ordenación)
- **Puedes compartir el análisis o el Dashboard con Usuarios o Grupos**
- Para compartir un Dashboard, primero debes publicarlo
- Los usuarios que ven el Dashboard también pueden ver los datos subyacentes

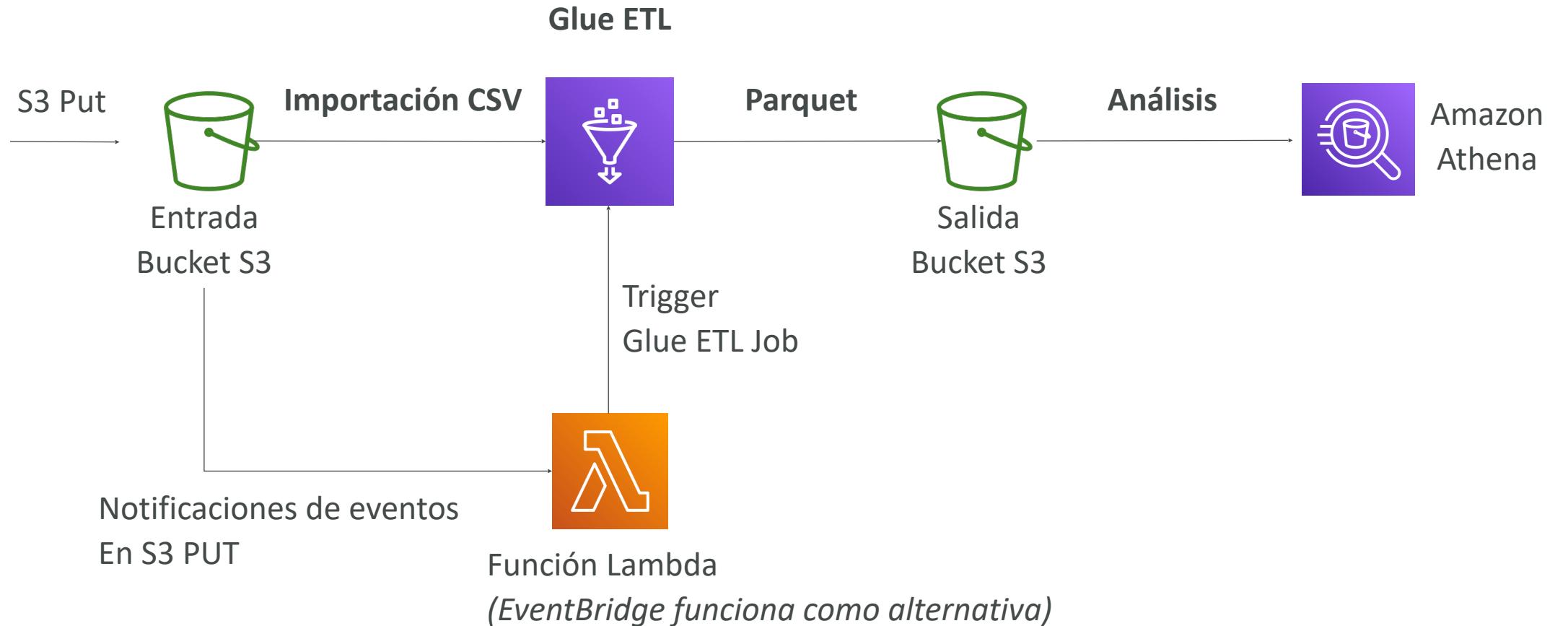
AWS Glue



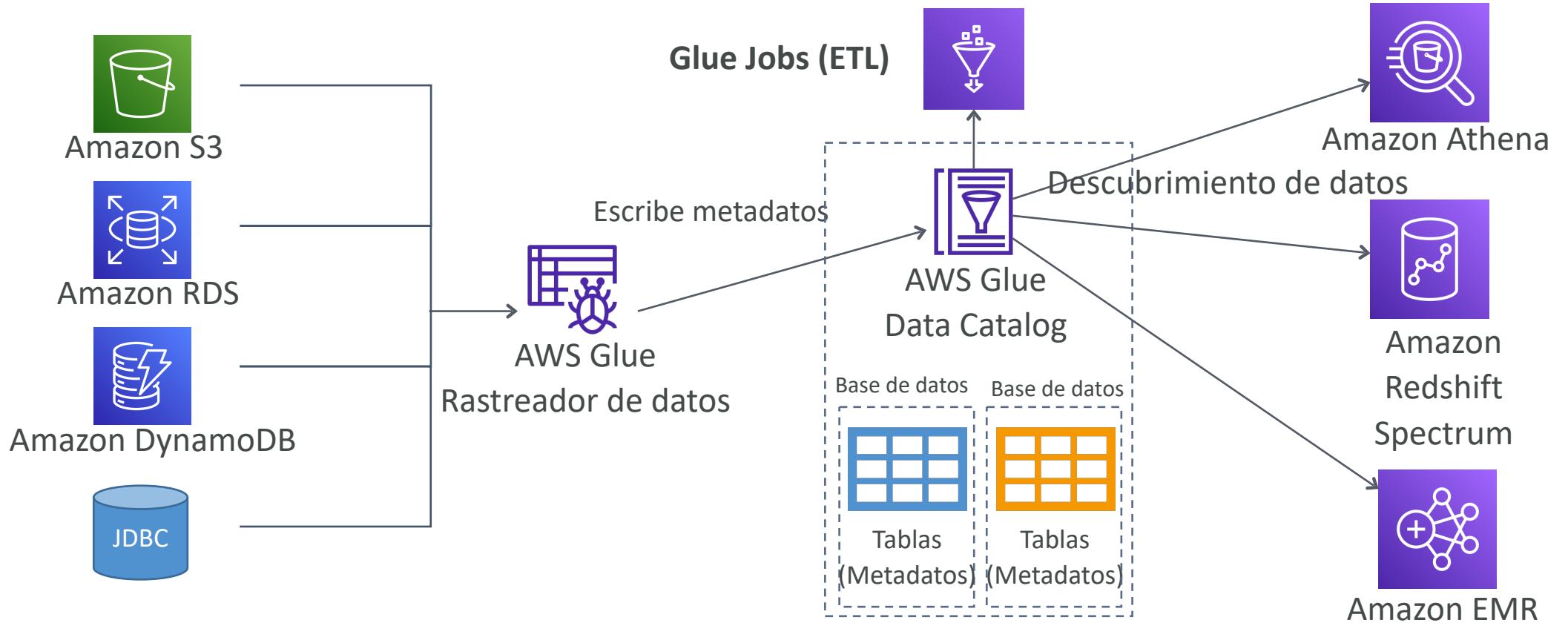
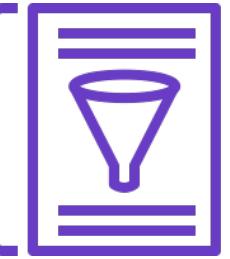
- Servicio gestionado de **extracción, transformación y carga (ETL)**
- Útil para preparar y transformar datos para análisis
- Servicio totalmente **sin servidor**



AWS Glue - Convertir datos en formato Parquet



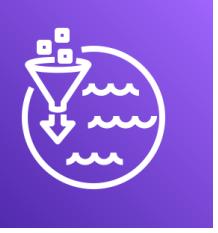
Catálogo de datos Glue



AWS Glue - lo que hay que saber a alto nivel

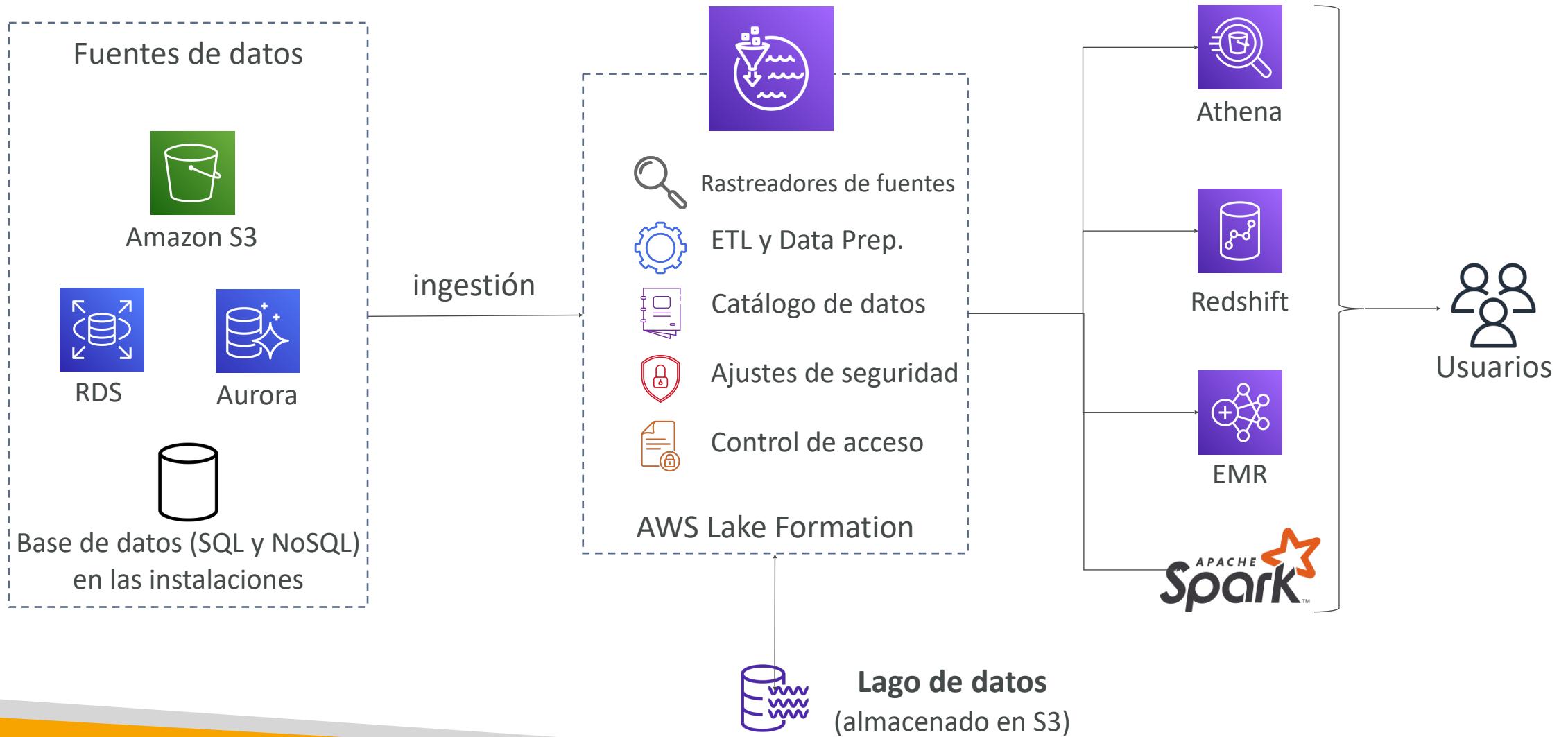
- **Glue Job Bookmarks:** evitar el reprocesamiento de datos antiguos
- **Vistas elásticas Glue (Glue Elastic Views):**
 - Combina y replica datos a través de múltiples almacenes de datos utilizando SQL
 - Sin código personalizado, Glue supervisa los cambios en los datos de origen, sin servidor
 - Aprovecha una "tabla virtual" (vista materializada)
- **Glue DataBrew:** limpia y normaliza los datos mediante transformaciones predefinidas
- **Glue Studio:** nueva interfaz gráfica de usuario para crear, ejecutar y supervisar trabajos ETL en Glue
- **Glue Streaming ETL** (basado en Apache Spark Structured Streaming): compatible con Kinesis Data Streaming, Kafka, MSK (managed Kafka)

AWS Lake Formation



- **Lake Formation / Lago de datos = lugar central para tener todos los datos con fines analíticos**
- Servicio totalmente gestionado que facilita la configuración de un **lago de datos** en cuestión de días
- Descubre, limpia, transforma e ingiere datos en el lago de datos
- Automatiza muchos pasos manuales complejos (recopilar, limpiar, mover, catalogar datos, ...) y de-duplicar (utilizando ML Transforms)
- Combina datos estructurados y no estructurados en el lago de datos
- **Planos de origen listos para usar:** S3, RDS, BD relacionales y NoSQL...
- **Control de acceso detallado para las aplicaciones (a nivel de fila y columna)**
- Construido sobre AWS Glue

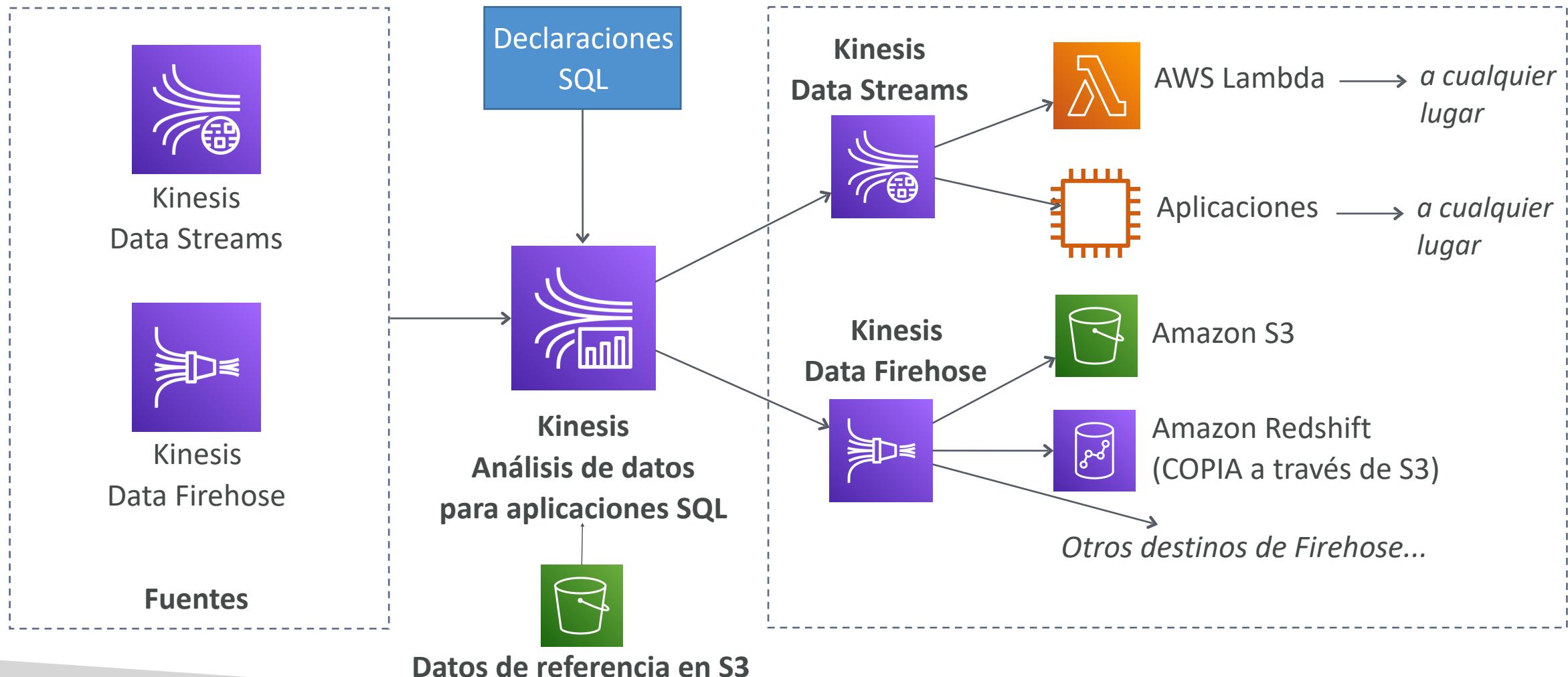
AWS Lake Formation



Ejemplo de permisos centralizados de AWS Lake Formation



Kinesis Data Analytics para aplicaciones SQL



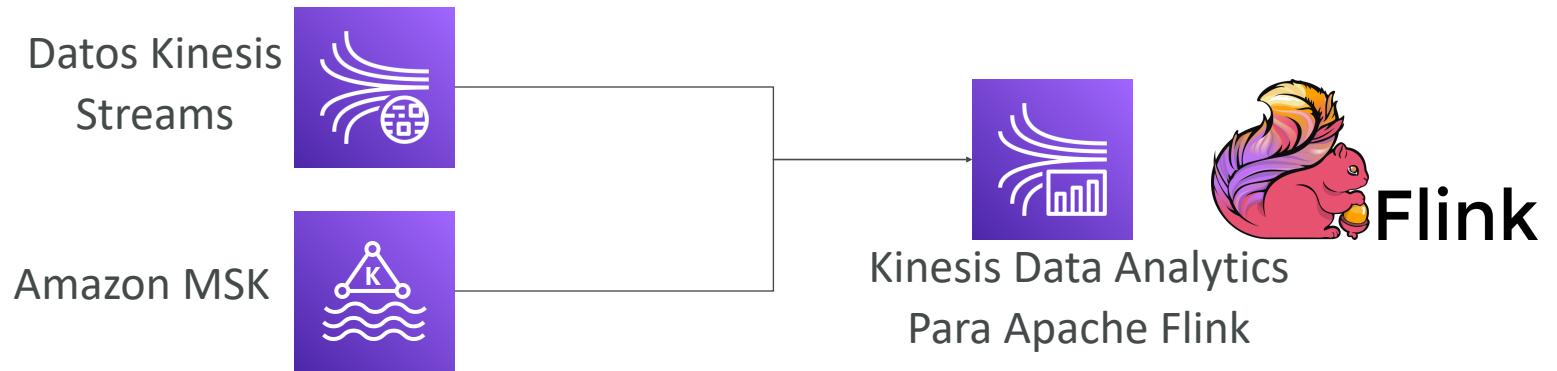


Kinesis Data Analytics (aplicación SQL)

- Análisis en tiempo real en **Kinesis Data Streams & Firehose** mediante SQL
- Añadir datos de referencia de Amazon S3 para enriquecer los datos de streaming
- Totalmente administrado, sin servidores que aprovisionar
- Escalado automático
- Paga por la tasa de consumo real
- Salida:
 - Kinesis Data Streams: crea flujos a partir de las consultas analíticas en tiempo real
 - Kinesis Data Firehose: envío de resultados de consultas analíticas a destinos
- Casos de uso:
 - Análisis de series temporales
 - Dashboards en tiempo real
 - Métricas en tiempo real

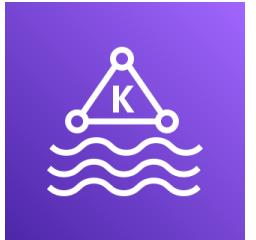
Kinesis Data Analytics para Apache Flink

- Utilización de Flink (Java, Scala o SQL) para procesar y analizar datos en streaming



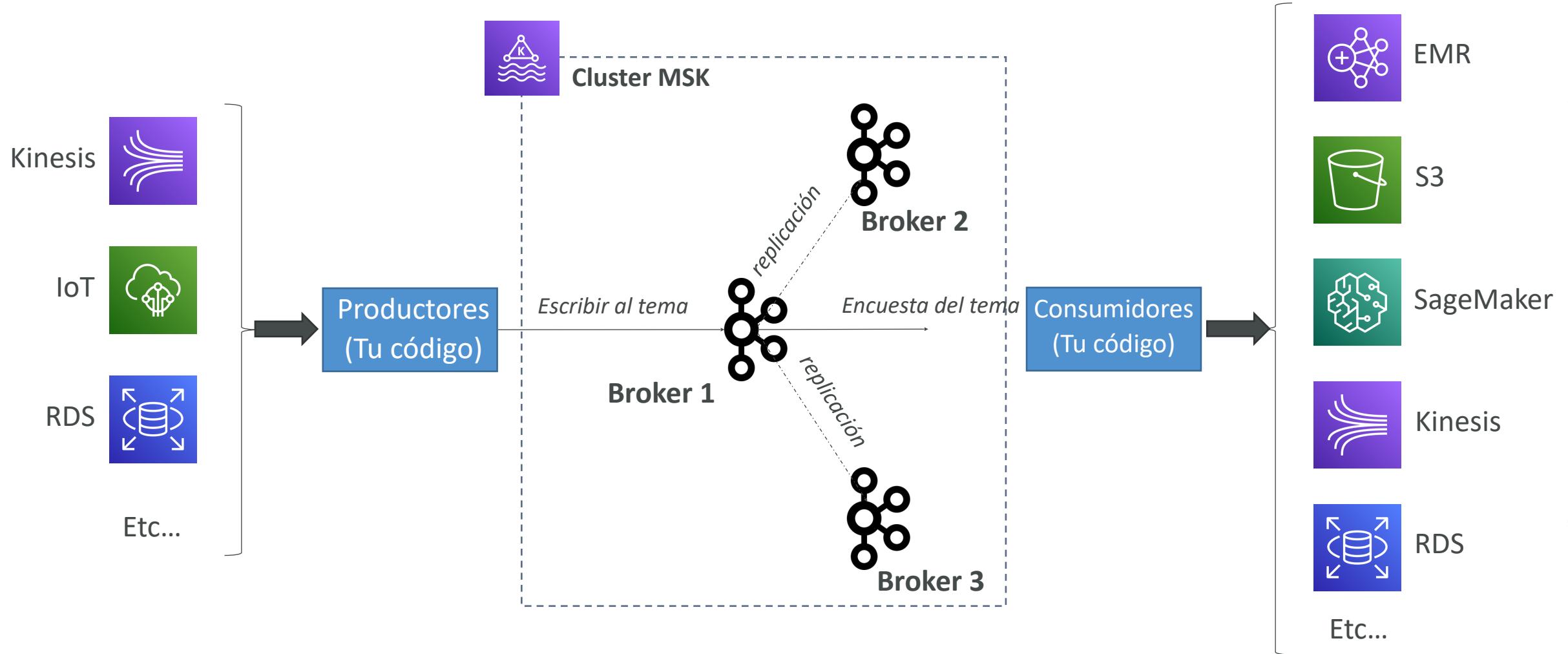
- Ejecuta cualquier aplicación Apache Flink en un clúster administrado en AWS
 - aprovisionamiento de recursos informáticos, computación paralela, escalado automático
 - backups de aplicaciones (implementados como puntos de control e instantáneas)
 - Utiliza cualquier característica de programación de Apache Flink
 - Flink no lee de Firehose (utiliza Kinesis Analytics para SQL en su lugar)

Amazon Managed Streaming para Apache Kafka (Amazon MSK)



- Alternativa a Amazon Kinesis
- Apache Kafka totalmente administrado en AWS
 - Permite crear, actualizar y eliminar clústeres
 - MSK crea y administra nodos brokers de Kafka y nodos Zookeeper por ti
 - Implementa el clúster MSK en tu VPC, multi-AZ (hasta 3 para Alta Disponibilidad)
 - Recuperación automática de fallos comunes de Apache Kafka
 - Los datos se almacenan en volúmenes EBS **durante todo el tiempo que deseas**
- **MSK sin servidor**
 - Ejecuta Apache Kafka en MSK sin gestionar la capacidad
 - MSK aprovisiona automáticamente los recursos y escala la computación y el almacenamiento

Apache Kafka a alto nivel



Kinesis Data Streams frente a Amazon MSK



Kinesis Data Streams

- Límite de tamaño de los mensajes: 1 MB
- Flujos de datos con shards
- División y fusión de fragmentos
- Cifrado TLS en vuelo
- Cifrado KMS en reposo



Amazon MSK

- 1MB por defecto, configurar para mayor (ej: 10MB)
- Temas Kafka con particiones
- Sólo se pueden añadir particiones a un tema
- Cifrado en vuelo PLAINTEXT o TLS
- Cifrado KMS en reposo

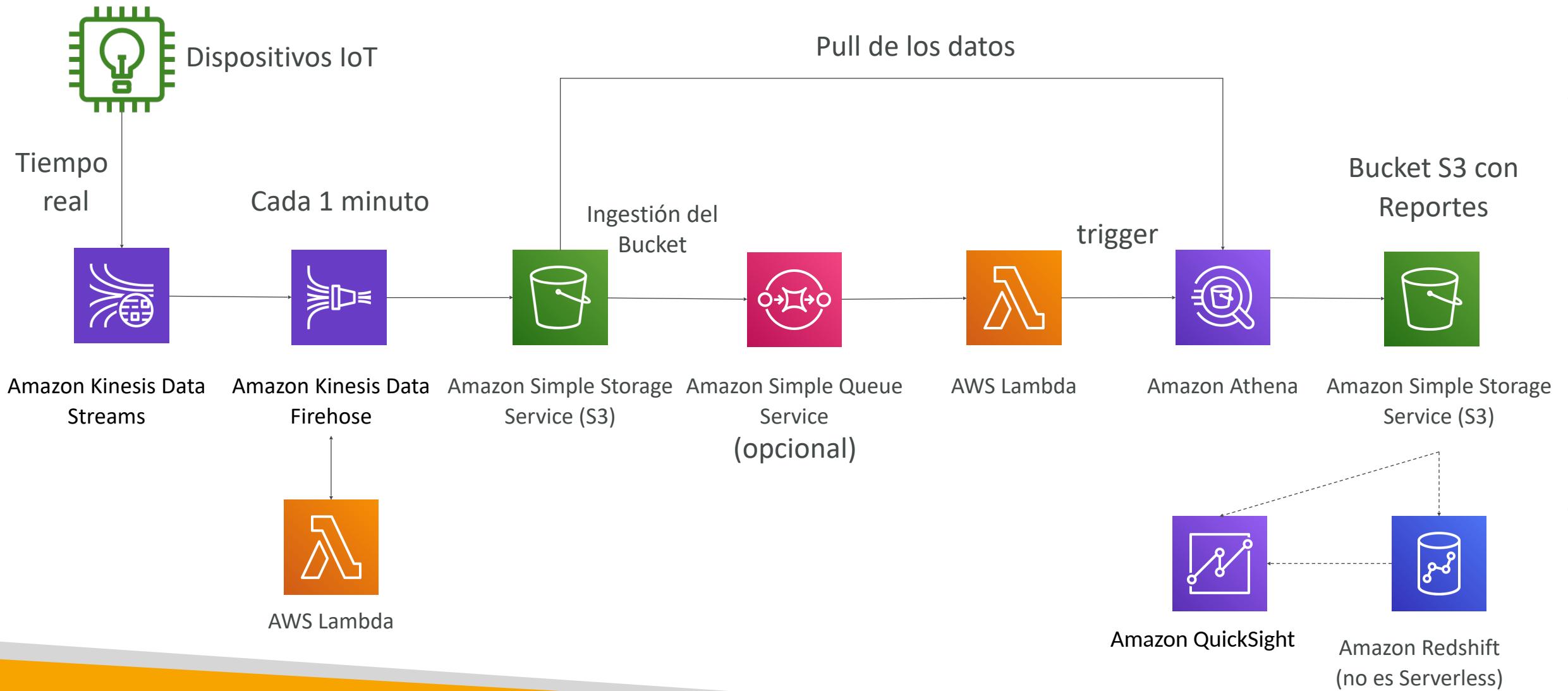
Consumidores Amazon MSK



Pipeline de ingestión de Big Data

- Queremos que el proceso de ingesta sea totalmente sin servidor.
- Queremos recopilar datos en tiempo real
- Queremos transformar los datos
- Queremos consultar los datos transformados utilizando SQL
- Los informes creados utilizando las consultas deben estar en S3
- Queremos cargar esos datos en un almacén y crear Dashboards

Pipeline de ingestión de Big Data



Debate sobre el Pipeline de ingestión de Big Data

- IoT Core permite recopilar datos de dispositivos IoT
- Kinesis es ideal para la recopilación de datos en tiempo real
- Firehose ayuda con la entrega de datos a S3 en tiempo casi real (1 minuto)
- Lambda puede ayudar a Firehose con las transformaciones de datos
- Amazon S3 puede activar notificaciones a SQS
- Lambda puede suscribirse a SQS (podríamos tener un conector S3 a Lambda)
- Athena es un servicio SQL sin servidor y los resultados se almacenan en S3
- El bucket de reportes contiene datos analizados y puede ser utilizado por herramientas de informes como AWS QuickSight, Redshift, etc...

Machine Learning

Amazon Rekognition



- Encuentra **objetos, personas, texto, escenas en imágenes y videos** utilizando ML
- **Análisis facial y búsqueda facial** para hacer verificación de usuarios, recuento de personas
- Crear una base de datos de "caras conocidas" o comparar con famosos
- Casos de uso:
 - Etiquetado
 - Moderación de contenidos
 - Detección de texto
 - Detección y Análisis de Caras (sexo, rango de edad, emociones...)
 - Búsqueda y verificación de caras
 - Reconocimiento de famosos
 - Trazado de trayectorias (por ejemplo, para análisis de partidos deportivos)

Amazon Rekognition - Moderación de contenidos

- Detectar contenido inapropiado, no deseado u ofensivo (imágenes y vídeos)
- Se utiliza en redes sociales, medios de difusión, publicidad y situaciones de comercio electrónico para crear una experiencia de usuario más segura
- Establece un Umbral Mínimo de Confianza para los elementos que se marcarán
- **Marcar contenido sensible para su revisión manual en la IA aumentada de Amazon (A2I)**
- Ayuda a cumplir la normativa



Amazon Transcribe



- **Convierte** automáticamente el **habla en texto**
- Utiliza un **proceso de deep learning** llamado **reconocimiento automático del habla** (ASR) para convertir el habla en texto de forma rápida y precisa
- Casos de uso:
 - transcribir llamadas de atención al cliente
 - automatizar el subtitulado y los subtítulos
 - generar metadatos para los activos de los medios de comunicación para crear un archivo con todas las posibilidades de búsqueda



*"Hello my name is Stéphane.
I hope you're enjoying the course!*

Amazon Polly



- Convierte el texto en voz real utilizando el aprendizaje profundo
- Te permite crear aplicaciones que hablan

*Hi! My name is Stéphane
and this is a demo of Amazon Polly*



Amazon Polly - Léxico y SSML

- Personaliza la pronunciación de las palabras con los **léxicos de pronunciación**
 - Palabras estilizadas: Jo7n => "Joan"
 - Acrónimos: AWS => "Amazon Web Services"
- Sube los léxicos en un fichero y la conversión se realizará directamente
- Genera voz a partir de texto plano o de documentos marcados con el **Lenguaje de Marcado de Síntesis de Voz (SSML)**: permite una mayor personalización
 - enfatizar palabras o frases concretas
 - utilizando pronunciación fonética
 - incluyendo sonidos respiratorios, susurros

Amazon Translate



- **Traducción** natural y precisa de **idiomas**
- Amazon Translate te permite **localizar contenidos** -como sitios web y aplicaciones- para **usuarios internacionales**, y traducir fácilmente grandes volúmenes de texto de forma eficiente.

Source language

Auto (auto) ▾

Hi my name is Stéphane

Target language

French (fr) ▾

Bonjour, je m'appelle Stéphane.

Portuguese (pt) ▾

Oi, meu nome é Stéphane.

Hindi (hi) ▾

हाय मेरा नाम स्टीफन है

Amazon Lex & Connect

- **Amazon Lex:** (la misma tecnología que impulsa a Alexa)
 - Reconocimiento automático del habla (ASR) para convertir el habla en texto
 - Comprensión del Lenguaje Natural para reconocer la intención del texto, de las personas que llaman
 - Ayuda a crear chatbots, bots de centros de llamadas
- **Amazon Connect:**
 - Recibe llamadas, crea flujos de contacto, **centro de contacto virtual** basado en la nube
 - Puede integrarse con otros sistemas CRM o AWS
 - Sin pagos iniciales, un 80% más barato que las soluciones tradicionales de centro de contacto



Amazon Comprehend



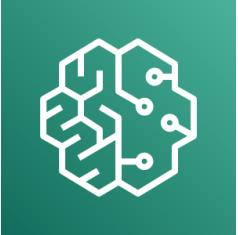
- Para el **Natural Language Processing – NLP (Procesamiento del Lenguaje Natural - PNL)**
- Servicio totalmente gestionado y sin servidor
- Utiliza el Machine Learning para encontrar ideas y relaciones en el texto
 - Lenguaje del texto
 - Extrae frases clave, lugares, personas, marcas o eventos
 - Comprende lo positivo o negativo del texto
 - Analiza el texto utilizando la tokenización y las partes del discurso
 - Organiza automáticamente una colección de archivos de texto por temas
- Ejemplos de casos de uso:
 - Analiza las interacciones con los clientes (correos electrónicos) para encontrar lo que conduce a una experiencia positiva o negativa
 - Crea y agrupa artículos por temas que Comprehend descubrirá

Amazon Comprehend Medical

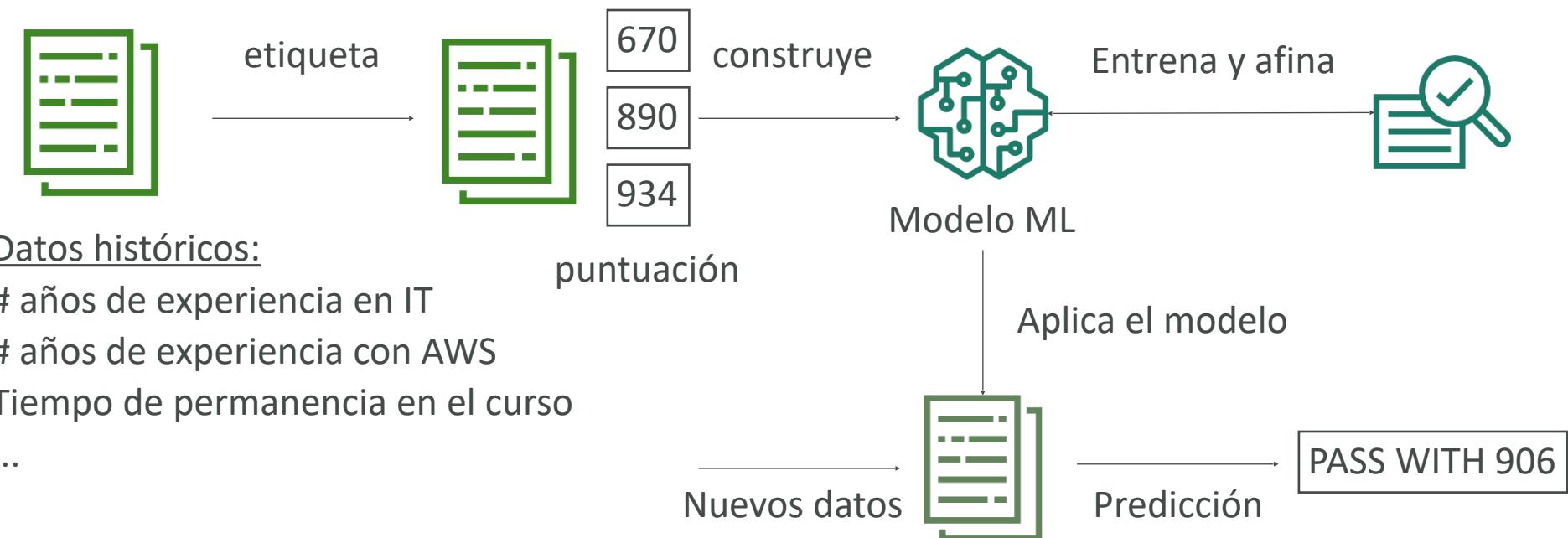


- Amazon Comprehend Medical detecta y devuelve información útil en texto clínico no estructurado:
 - Notas del médico
 - Resúmenes de alta
 - Resultados de pruebas
 - Notas de casos
- Utiliza NLP para detectar Información Sanitaria Protegida (PHI) - API DetectPHI
- Almacena tus documentos en Amazon S3, analiza los datos en tiempo real con Kinesis Data Firehose, o utiliza Amazon Transcribe para transcribir las narraciones de los pacientes en texto que pueda ser analizado por Amazon Comprehend Medical.

Amazon SageMaker



- Servicio totalmente gestionado para que los desarrolladores/científicos de datos construyan modelos ML
- Normalmente, es difícil hacer todos los procesos en un solo lugar + aprovisionar servidores
- Proceso de Machine Learning (simplificado): predecir la nota de tu examen



Amazon Forecast



- Servicio totalmente gestionado que utiliza el ML para ofrecer previsiones muy precisas
- Ejemplo: predecir las futuras ventas de un chubasquero
- Un 50% más de precisión que mirando los datos por sí mismos
- Reduce el tiempo de previsión de meses a horas
- Casos de uso: Planificación de la demanda de productos, planificación financiera, planificación de recursos, ...

Datos históricos de series temporales:

Características del producto

Precios

Descuentos

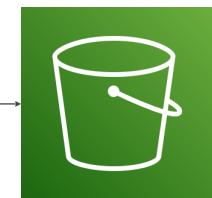
Tráfico del sitio web

Ubicación de las tiendas

...



carga



Amazon S3



Amazon Forecast

produce



Forecasting Model

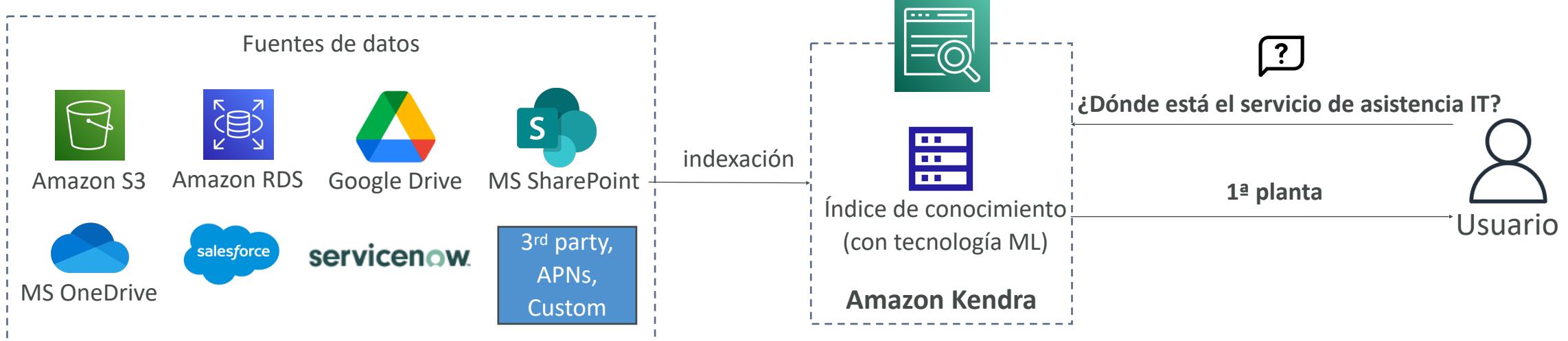


Ventas futuras
del impermeable:
500.000 dólares

Amazon Kendra



- **Servicio de búsqueda de documentos** totalmente gestionado y potenciado por Machine Learning
- Extrae respuestas de un documento (texto, pdf, HTML, PowerPoint, MS Word, preguntas frecuentes...)
- Capacidades de búsqueda en lenguaje natural
- Aprende de las interacciones/retroalimentación de los usuarios para promover los resultados preferidos (aprendizaje incremental)
- Capacidad de afinar manualmente los resultados de la búsqueda (importancia de los datos, frescura, personalización, ...)



Amazon Personalize



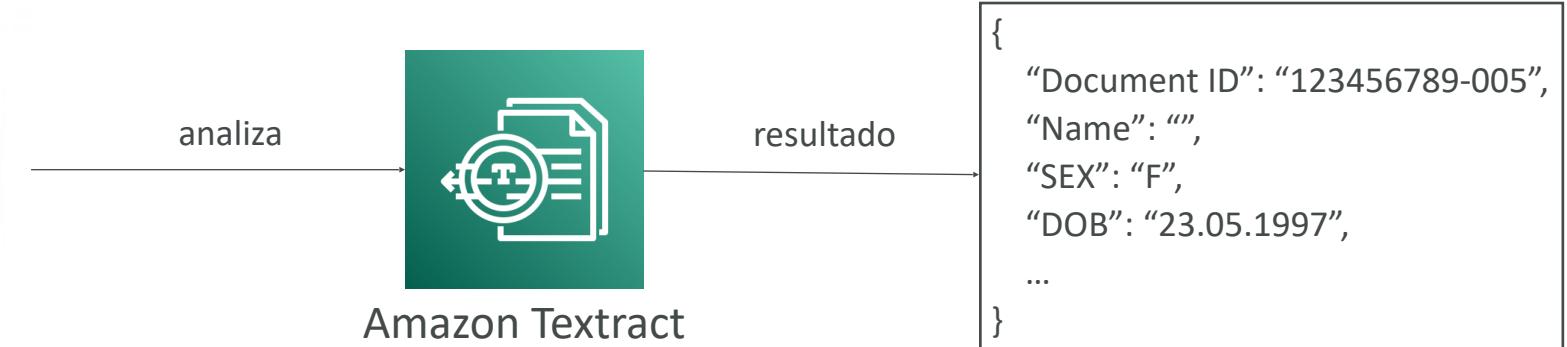
- Servicio de ML totalmente gestionado para crear aplicaciones con recomendaciones personalizadas en tiempo real
- Ejemplo: recomendaciones/reclasificación de productos personalizados, marketing directo personalizado
 - Ejemplo: El usuario compró herramientas de jardinería, proporciona recomendaciones sobre la próxima que debe comprar
- La misma tecnología utilizada por Amazon.com
- Se integra en sitios web existentes, aplicaciones, SMS, sistemas de marketing por correo electrónico, ...
- Se implementa en días, no en meses (no es necesario construir, formar y desplegar soluciones de ML)
- Casos de uso: tiendas minoristas, medios de comunicación y entretenimiento...



Amazon Textract



- Extrae automáticamente el texto, la escritura y los datos de cualquier documento escaneado utilizando IA y ML



- Extrae datos de formularios y tablas
- Leer y procesar cualquier tipo de documento (PDFs, imágenes, ...)
- Casos de uso:
 - Servicios financieros (por ejemplo, facturas, informes financieros)
 - Sanidad (por ejemplo, historiales médicos, reclamaciones de seguros)
 - Sector público (por ejemplo, formularios fiscales, documentos de identidad, pasaportes)

Resumen - Machine Learning

- **Rekognition:** detección de caras, etiquetado, reconocimiento de famosos
- **Transcribe:** de audio a texto (por ejemplo, subtítulos)
- **Polly:** de texto a audio
- **Translate:** traducciones
- **Lex:** construir bots conversacionales - chatbots
- **Connect:** centro de contacto en el Cloud
- **Comprehend:** procesamiento del lenguaje natural
- **SageMaker:** Machine Learning para todos los desarrolladores y científicos de datos
- **Forecast:** construye previsiones muy precisas
- **Kendra:** motor de búsqueda con ML
- **Personalize:** recomendaciones personalizadas en tiempo real
- **Textract:** detecta texto y datos en los documentos

Monitorización, auditoría y rendimiento de AWS

CloudWatch, CloudTrail & AWS Config

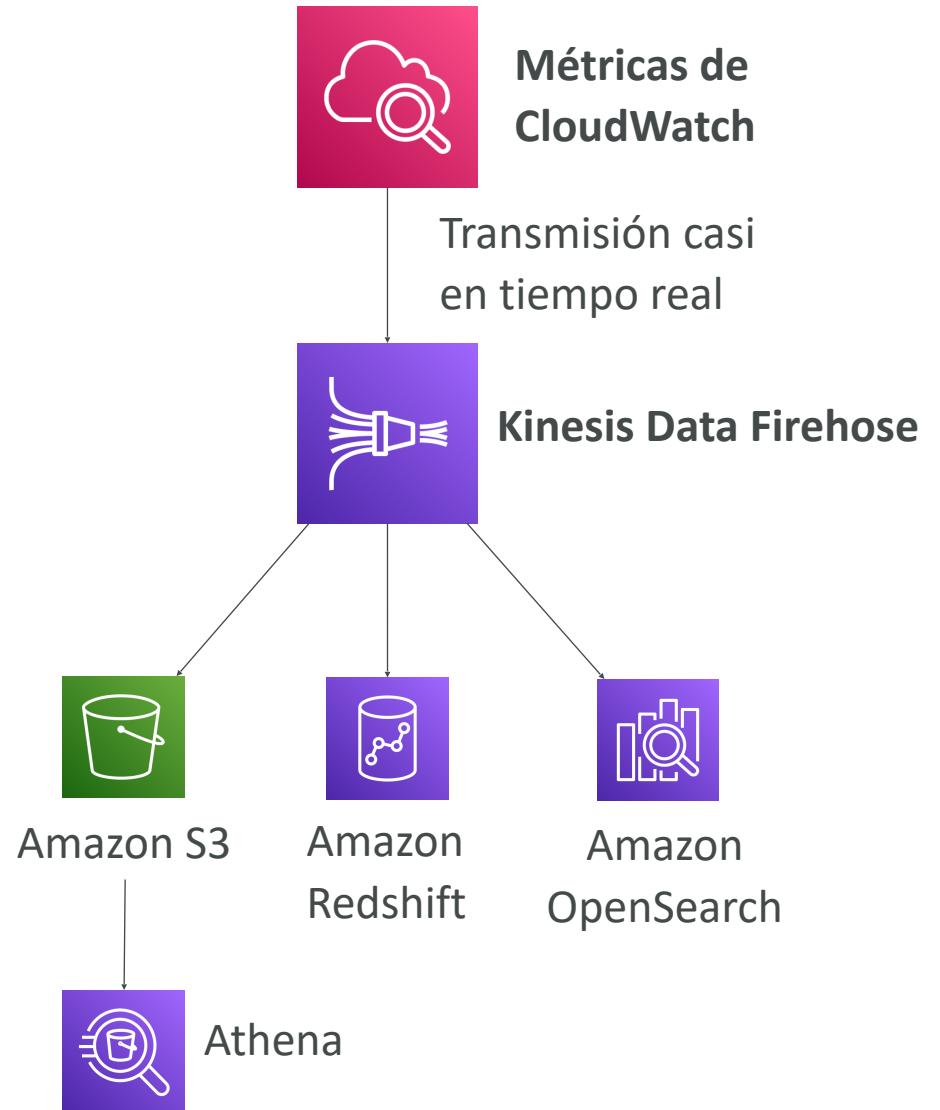
Métricas de Amazon CloudWatch



- CloudWatch proporciona métricas para cada servicio en AWS
- La **métrica** es una variable a monitorizar (CPUUtilization, NetworkIn...)
- Las métricas pertenecen a **espacios de nombres**
- Las métricas tienen **marcas de tiempo**
- Se pueden crear Dashboards de métricas de CloudWatch
- Se pueden crear **métricas personalizadas de CloudWatch** (para la RAM, por ejemplo)

Flujos de métricas de CloudWatch

- Transmite continuamente métricas de CloudWatch a un destino de tu elección, **con entrega casi en tiempo real** y baja latencia.
 - Amazon Kinesis Data Firehose (y luego sus destinos)
 - Proveedor de servicios de terceros: Datadog, Dynatrace, New Relic, Splunk, Sumo Logic...
- Opción de **filtrar métricas** para transmitir sólo un subconjunto de ellas





Registros de CloudWatch (Logs)

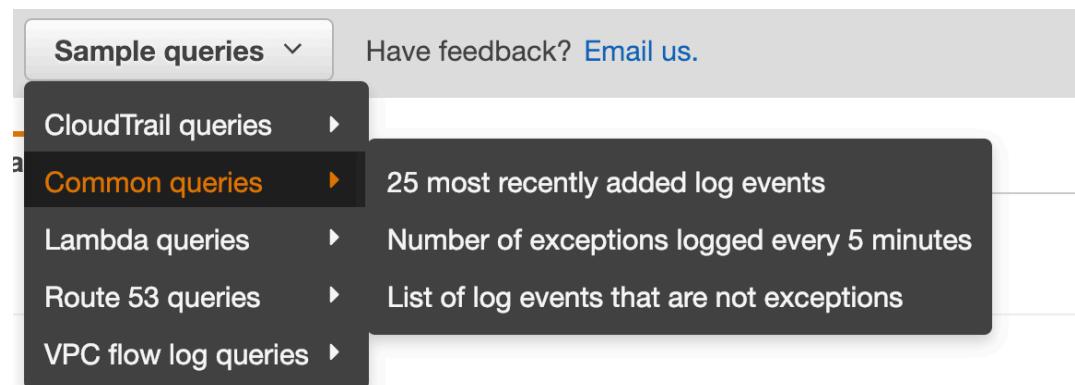
- **Grupos de registro:** nombre arbitrario, normalmente representando una aplicación
- **Flujo de registro:** instancias dentro de la aplicación / archivos de registro / contenedores
- Puede definir políticas de expiración de logs (nunca expiran, 30 días, etc..)
- **CloudWatch Logs puede enviar logs a:**
 - Amazon S3 (exportaciones)
 - Flujos de datos de Kinesis
 - Kinesis Data Firehose
 - AWS Lambda
 - ElasticSearch

CloudWatch Logs - Fuentes

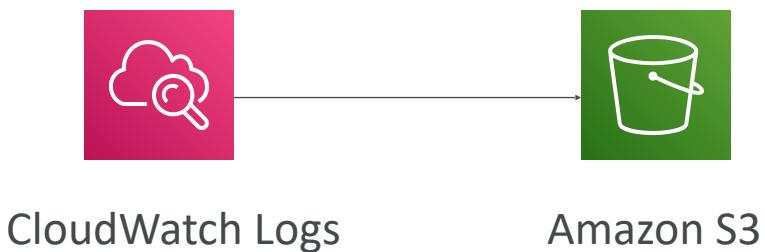
- SDK, agente de CloudWatch Logs, agente unificado de CloudWatch
- Elastic Beanstalk: recogida de logs desde la aplicación
- ECS: recopilación desde contenedores
- AWS Lambda: recopilación de registros de funciones
- Registros de flujo de VPC: Registros específicos de VPC
- API Gateway
- CloudTrail basado en filtro
- Route53: registro de consultas DNS

Filtro de métricas e información de CloudWatch Logs

- CloudWatch Logs puede utilizar expresiones de filtro
 - Por ejemplo, encontrar una IP específica dentro de un log
 - O contar ocurrencias de "ERROR" en sus registros
- Los filtros de métricas pueden utilizarse para activar alarmas de CloudWatch Metrics.
- CloudWatch Logs Insights puede utilizarse para consultar registros y añadir consultas a CloudWatch Dashboards.

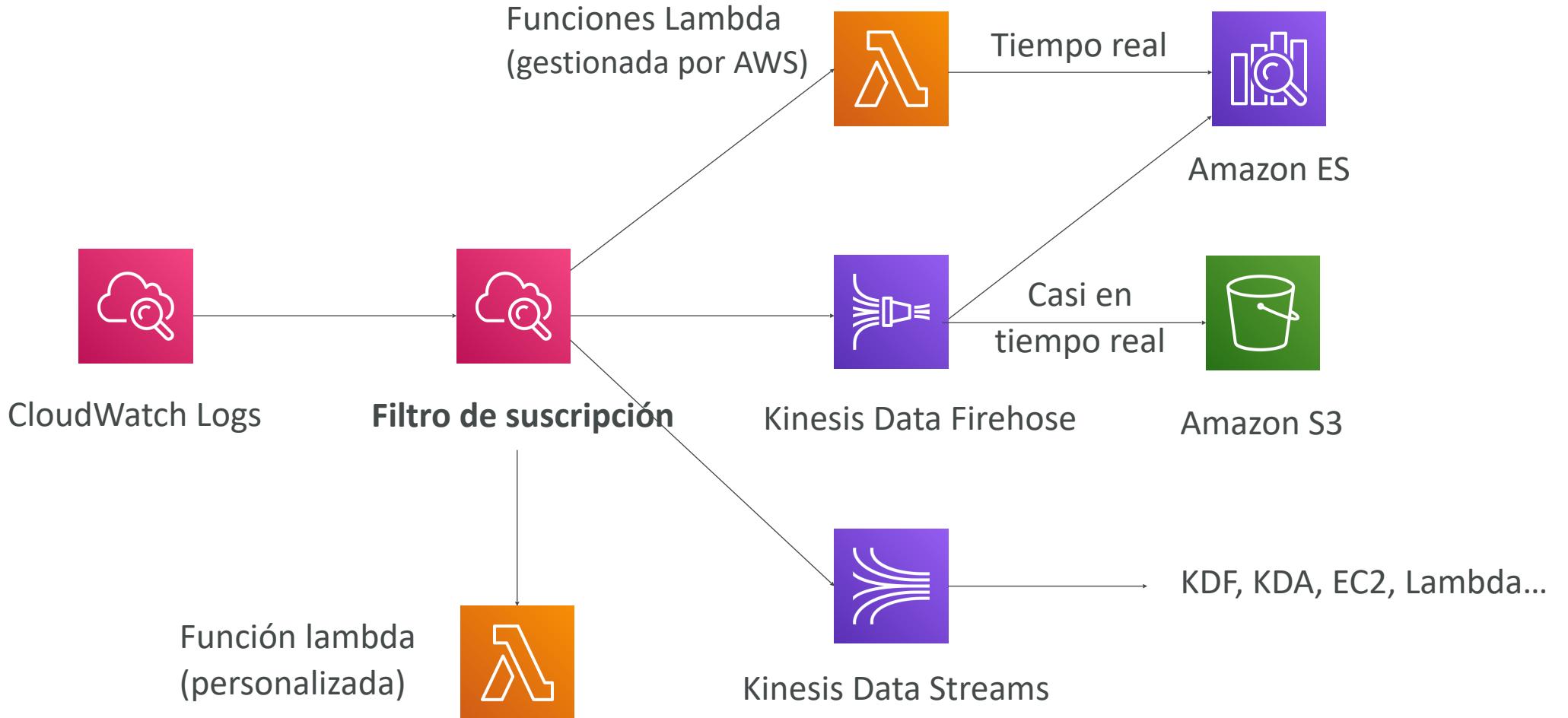


CloudWatch Logs - Exportación a S3

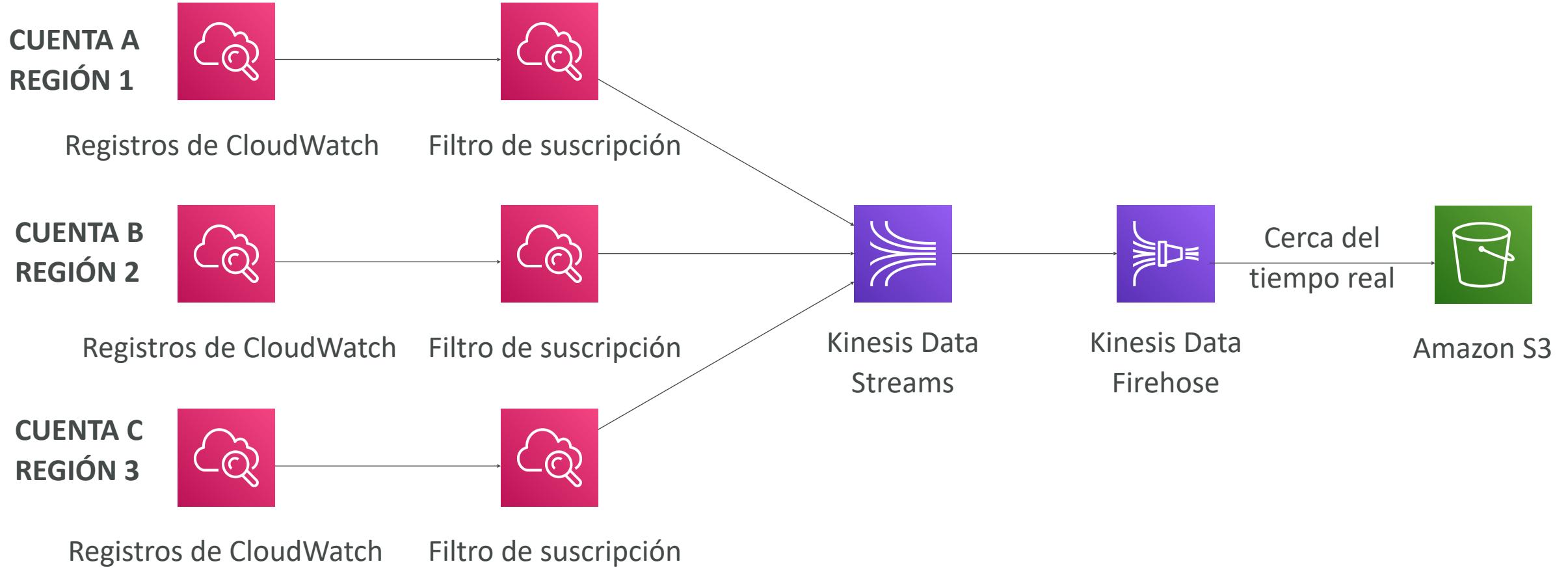


- Los datos de registro pueden tardar **hasta 12 horas** en estar disponibles para su exportación.
- La llamada a la API es **CreateExportTask**
- No en tiempo casi real ni en tiempo real... utilice suscripciones a registros en su lugar

Suscripciones a CloudWatch Logs

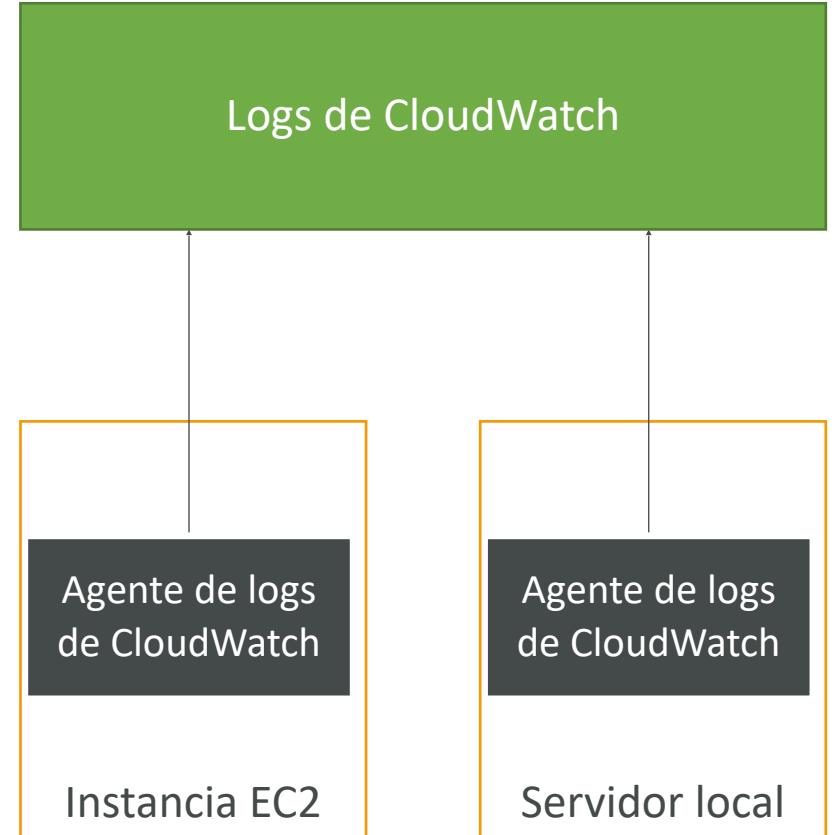


Agregación de CloudWatch Logs Multi-Cuenta y Multi-Región



CloudWatch Logs para EC2

- Por defecto, los logs de tu máquina EC2 no irán a CloudWatch
- Necesitas ejecutar un agente de CloudWatch en EC2 para enviar los logs que deseas.
- Asegúrate de que los permisos IAM son correctos
- El agente de logs de CloudWatch también puede configurarse en local



Agente CloudWatch Logs y Agente Unificado

- Para servidores virtuales (instancias EC2, servidores locales...)
- **Agente de logs de CloudWatch**
 - Versión antigua del agente
 - Sólo puede enviar a CloudWatch Logs
- **Agente Unificado CloudWatch**
 - Recoge métricas adicionales a nivel de sistema, como RAM, procesos, etc...
 - Recoge logs para enviarlos a CloudWatch Logs
 - Configuración centralizada mediante el Almacén de Parámetros SSM

Agente Unificado CloudWatch - Métricas

- Recopilados directamente en tu servidor Linux / instancia EC2
- **CPU** (activa, huésped, inactiva, sistema, usuario)
- **Métricas de disco** (libre, usado, total), IO de disco (escrituras, lecturas, bytes, iops)
- **RAM** (gratis, inactiva, usada, total, en caché)
- **Netstat** (número de conexiones TCP y UDP, paquetes netos, bytes)
- **Procesos** (totales, muertos, bloqueados, inactivos, en ejecución, en reposo)
- **Espacio de intercambio** (gratis, usado, % usado)

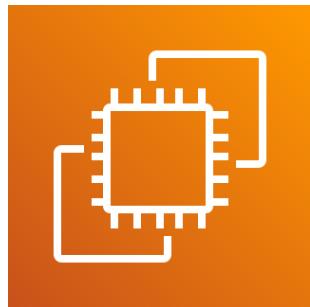
Alarmas de CloudWatch



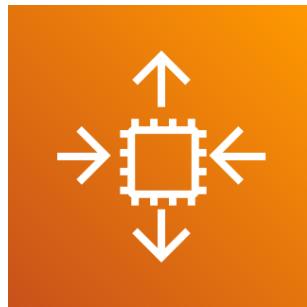
- Las alarmas se utilizan para activar notificaciones para cualquier métrica
- Varias opciones (muestreo, %, máx, mín, etc...)
- Estados de alarma:
 - OK
 - DATOS_INSUFICIENTES
 - ALARMA
- Periodo:
 - Tiempo en segundos para evaluar la métrica
 - Métricas personalizadas de alta resolución: 10 seg, 30 seg o múltiplos de 60 seg

Objetivos de alarma de CloudWatch

- Detener, Terminar, Reiniciar o Recuperar una Instancia EC2
- Activar la acción de Autoescalado
- Enviar notificación a SNS (desde donde puedes hacer prácticamente cualquier cosa)



Amazon EC2



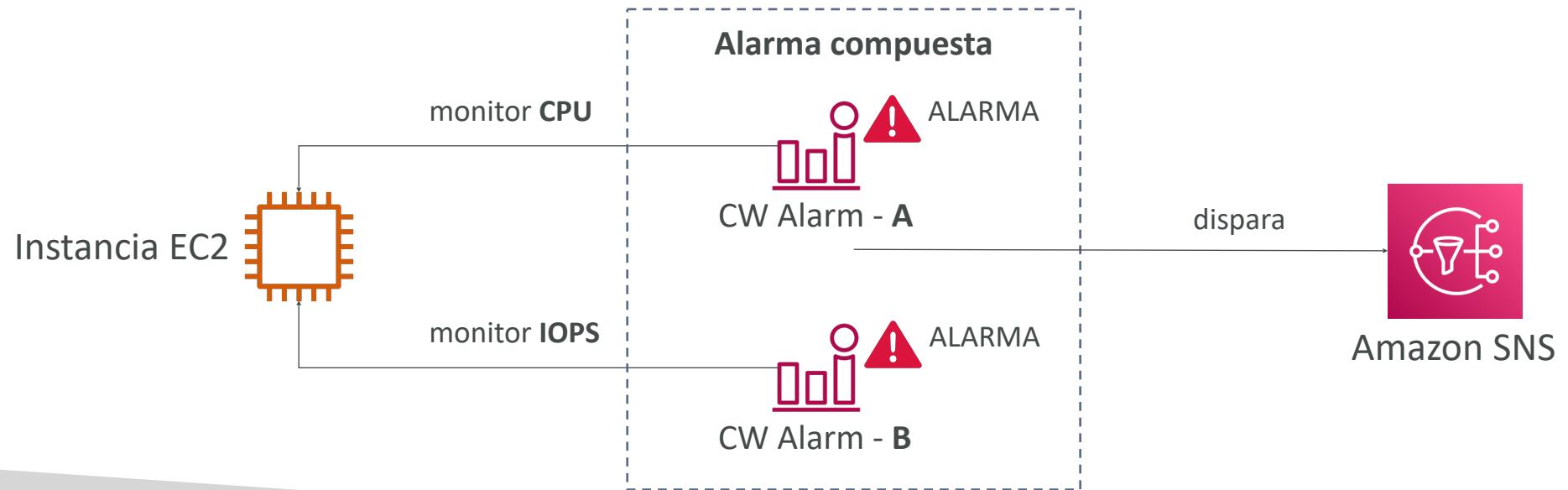
Auto Scaling Group
de EC2



Amazon SNS

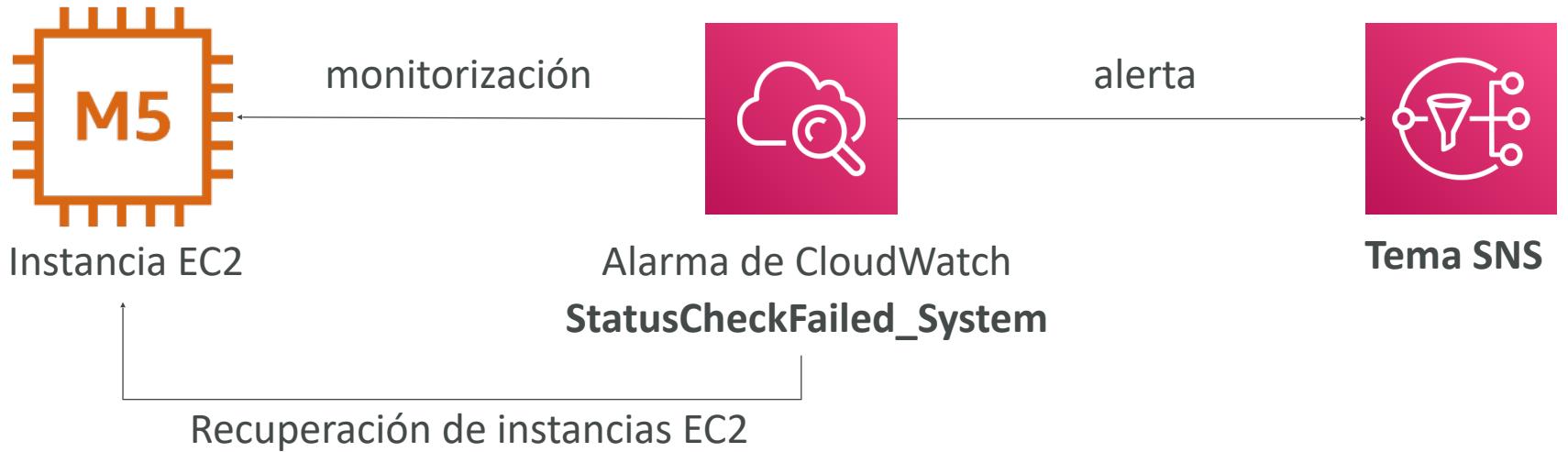
Alarmas de CloudWatch - Alarmas compuestas

- Las Alarmas CloudWatch son sobre una única métrica
- **Las alarmas compuestas supervisan los estados de otras alarmas múltiples**
- Condiciones AND y OR
- Útiles para reducir el "ruido de alarma" creando alarmas compuestas complejas



Recuperación de instancias EC2

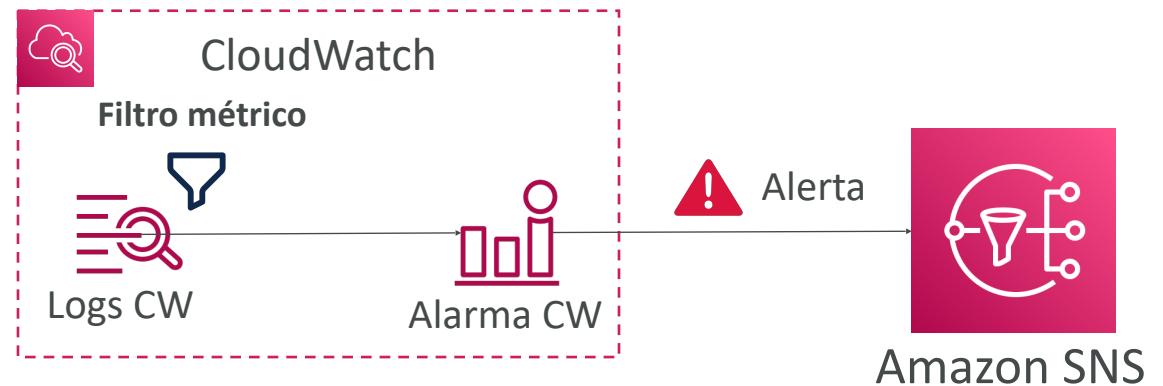
- Comprobación de estado:
 - Estado de la instancia = comprueba la máquina virtual EC2
 - Estado del sistema = comprueba el hardware subyacente



- **Recuperación:** La misma IP privada, pública, elástica, metadatos, grupo de colocación

Alarma de CloudWatch: es bueno saber que...

- Se pueden crear alarmas basadas en los filtros de métricas de CloudWatch Logs

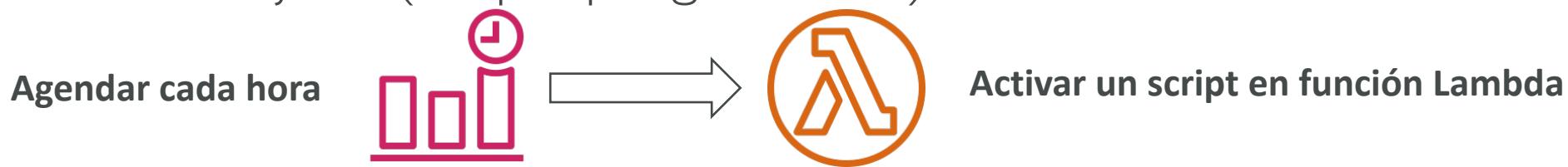


- Para probar las alarmas y notificaciones, establece el estado de alarma mediante la CLI aws cloudwatch **set-alarm-state --alarm-name "myalarm" --state-value ALARM --state-reason "fines de prueba"**

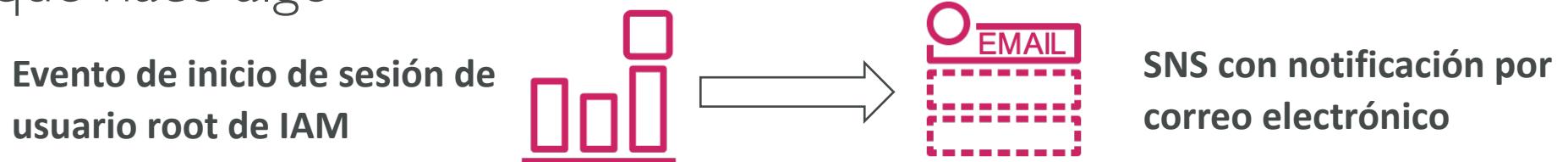
Amazon EventBridge (Antes CloudWatch Events)



- Programar: Cron jobs (scripts programados)

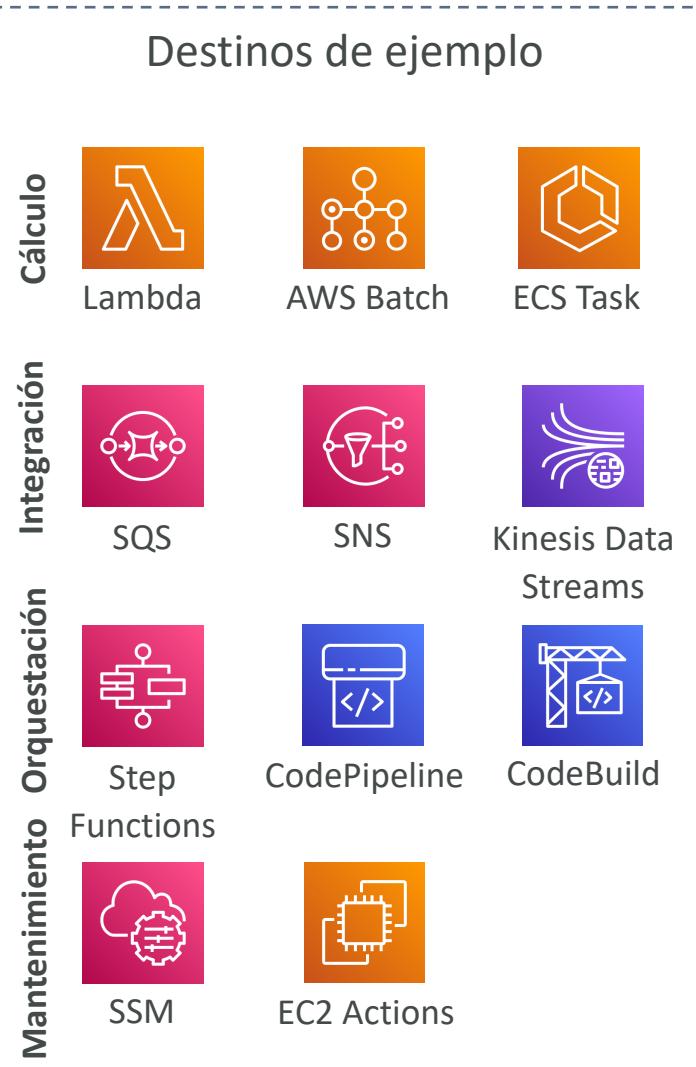
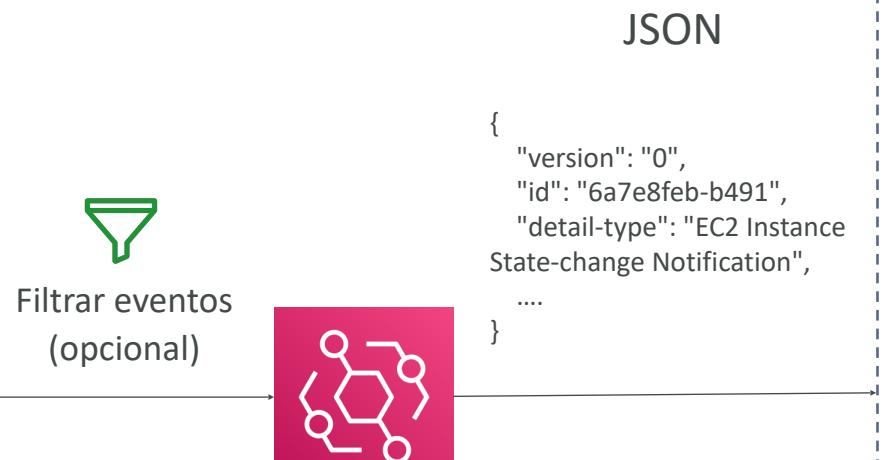


- Patrón de eventos: Reglas de eventos para reaccionar ante un servicio que hace algo



- Activa funciones Lambda, envía mensajes SQS/SNS...

Reglas de Amazon EventBridge



Amazon EventBridge



- Otras cuentas de AWS pueden acceder a los buses de eventos mediante políticas basadas en recursos
- Puedes **archivar eventos** (todos/filtro) enviados a un bus de eventos (indefinidamente o por un periodo determinado)
- Posibilidad de **reproducir eventos archivados**

Amazon EventBridge - Registro de esquemas

- EventBridge puede analizar los eventos de tu bus e inferir el **esquema**
- El **Registro de esquemas** te permite generar código para tu aplicación, que sabrá de antemano cómo se estructuran los datos en el bus de eventos
- El esquema puede versionarse

The screenshot shows the AWS Schema Registry interface. At the top, there is a header with the URL 'aws.codepipeline@CodePipelineActionExecutionStateChange'. Below it, a section titled 'Schema details' provides the following information:

Schema name	Last modified	Schema ARN
aws.codepipeline@CodePipelineActionExecutionStateChange	Dec 1, 2019, 12:11 AM GMT	-
Description	Schema for event type CodePipelineActionExecutionStateChange, published by AWS service aws.codepipeline	Schema registry aws.events Number of versions 1 Schema type OpenAPI 3.0

Below this, another section titled 'Version 1' shows the creation date 'Created on Dec 1, 2019, 12:11 AM GMT'. It includes a 'Action' dropdown and a 'Download code bindings' button. The schema definition is displayed as a JSON-like code block:

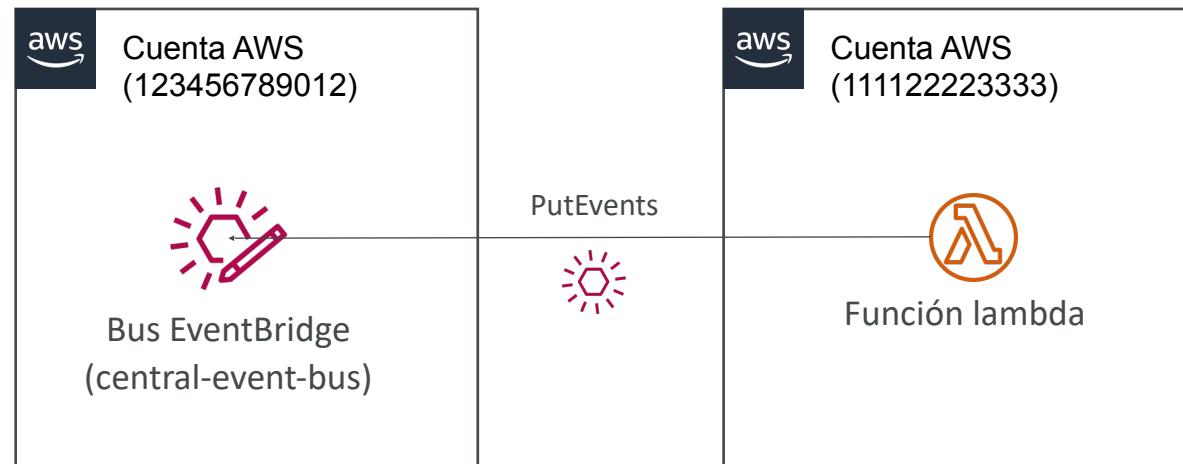
```
1 {
2   "openapi": "3.0.0",
3   "info": {
4     "version": "1.0.0",
5     "title": "CodePipelineActionExecutionStateChange"
6   },
7   "paths": {},
8   "components": {
9     "schemas": {
10       "AWSEvent": {
```

Amazon EventBridge - Política basada en recursos

- Gestionar permisos para un bus de eventos específico
- Ejemplo: permitir/denegar eventos de otra cuenta AWS o región AWS
- Caso práctico: agregar todos los eventos de tu Organización AWS en una única cuenta AWS o región AWS

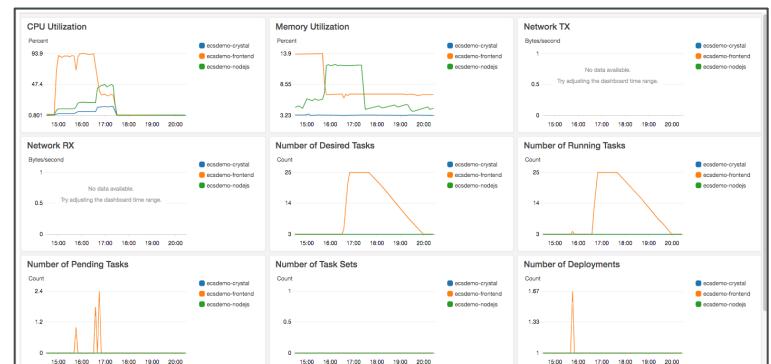
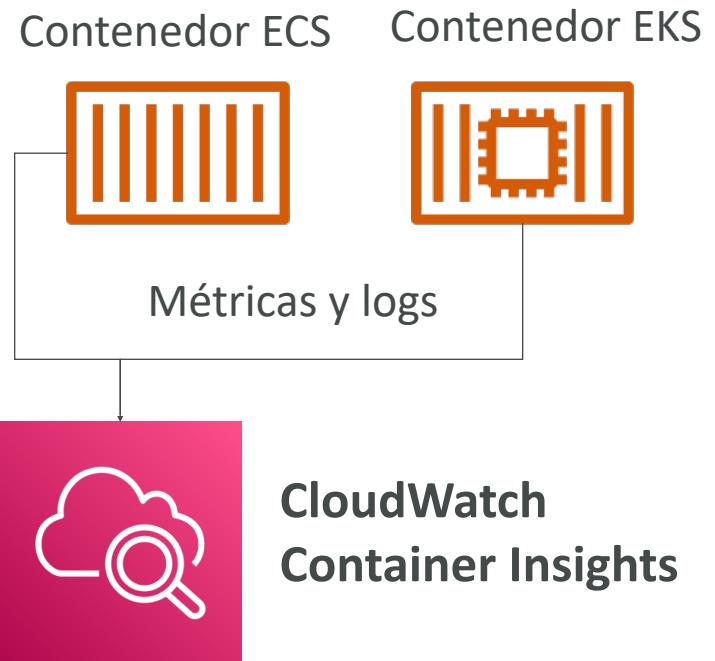
```
{  
    "Version": "2012-10-17",  
    "Statement": [  
        {  
            "Effect": "Allow",  
            "Action": "events:PutEvents",  
            "Principal": { "AWS": "111122223333" },  
            "Resource": "arn:aws:events:us-east-1:123456789012:  
event-bus/central-event-bus"  
        }  
    ]  
}
```

Permitir **eventos** desde otra cuenta AWS



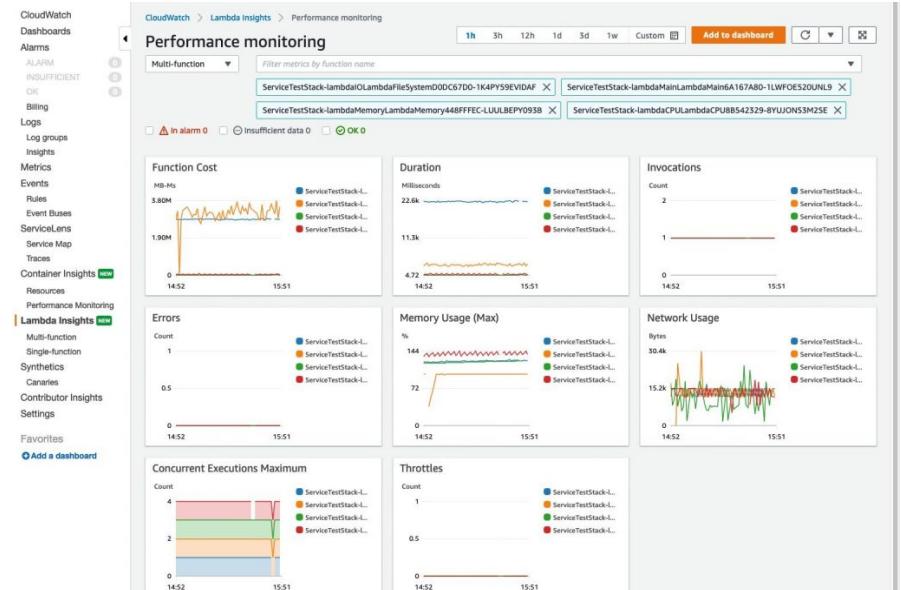
CloudWatch Container Insights

- Recoge, agrega y resume las **métricas y logs** de los contenedores
- Disponible para contenedores en...
 - Servicio Elástico de Contenedores de Amazon (Amazon ECS)
 - Servicios Elásticos de Kubernetes de Amazon (Amazon EKS)
 - Plataformas Kubernetes en EC2
 - Fargate (tanto para ECS como para EKS)
- **En Amazon EKS y Kubernetes, CloudWatch Insights utiliza una versión en contenedores del Agente CloudWatch para descubrir contenedores**



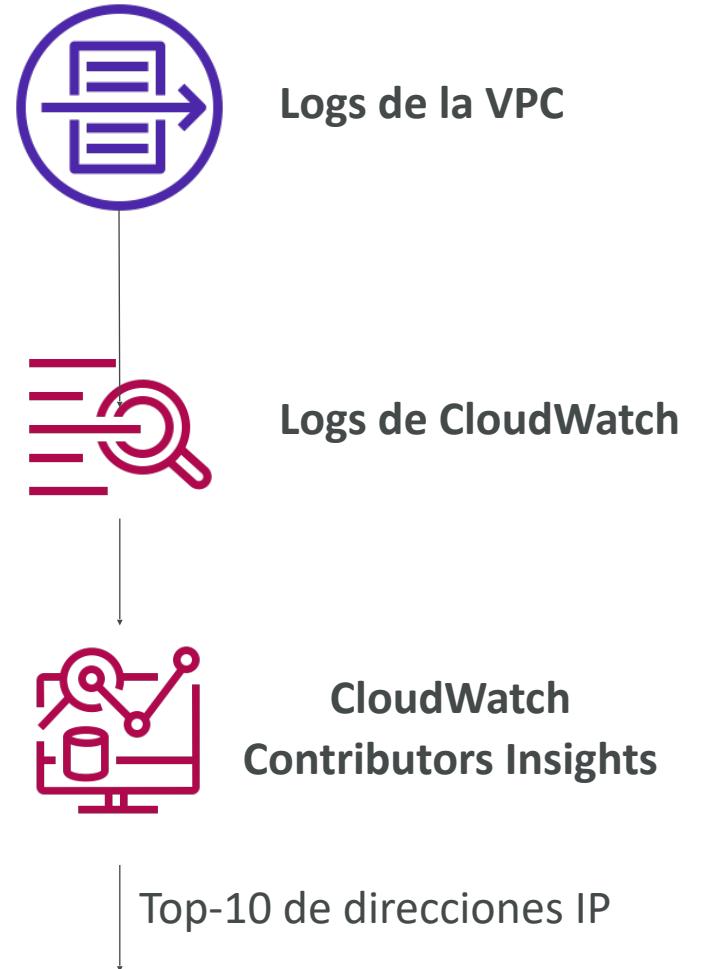
CloudWatch Lambda Insights

- Solución de monitorización y resolución de problemas para aplicaciones sin servidor que se ejecutan en AWS Lambda
- Recopila, agrega y resume métricas a nivel de sistema, incluyendo tiempo de CPU, memoria, disco y red
- Recopila, agrega y resume información de diagnóstico, como arranques en frío y cierres de trabajadores Lambda.
- Lambda Insights se proporciona como una capa de Lambda



CloudWatch Contributors Insights

- Analiza los logs y crea series temporales que muestren los datos de los colaboradores.
 - **Ver métricas sobre los N colaboradores principales**
 - El número total de colaboradores únicos, y su uso.
- Esto te ayuda a encontrar a los que más hablan y a comprender quién o qué está afectando al rendimiento del sistema.
- Funciona para cualquier logs generado por AWS (VPC, DNS, etc.)
- Por ejemplo, puedes encontrar hosts defectuosos, **identificar a los usuarios de red más pesados** o encontrar las URL que generan más errores.
- Puedes crear tus reglas desde cero, o también puedes utilizar reglas de muestra que AWS ha creado: **aprovecha tus logs de CloudWatch**
- CloudWatch también proporciona reglas integradas que puedes utilizar para analizar métricas de otros servicios de AWS.



Amazon CloudWatch Application Insights

- **Proporciona dashboards automatizados que muestran problemas potenciales con las aplicaciones monitorizadas, para ayudar a aislar los problemas en curso**
- Tus aplicaciones se ejecutan en instancias de Amazon EC2 sólo con determinadas tecnologías (Java, .NET, Microsoft IIS Web Server, bases de datos...)
- Y puedes utilizar otros recursos de AWS como Amazon EBS, RDS, ELB, ASG, Lambda, SQS, DynamoDB, S3 bucket, ECS, EKS, SNS, API Gateway...
- Desarrollado por SageMaker
- Mayor visibilidad del estado de tus aplicaciones para reducir el tiempo que tardarás en solucionar y reparar tus aplicaciones
- Las conclusiones y alertas se envían a Amazon EventBridge y SSM OpsCenter

CloudWatch Insights y visibilidad operativa

- **CloudWatch Container Insights**

- ECS, EKS, Kubernetes en EC2, Fargate, necesita agente para Kubernetes
- Métricas y logs

- **CloudWatch Lambda Insights**

- Métricas detalladas para solucionar problemas de aplicaciones sin servidor

- **CloudWatch Contributors Insights**

- Encuentra los contribuidores "Top-N" a través de los logs de CloudWatch

- **CloudWatch Application Insights**

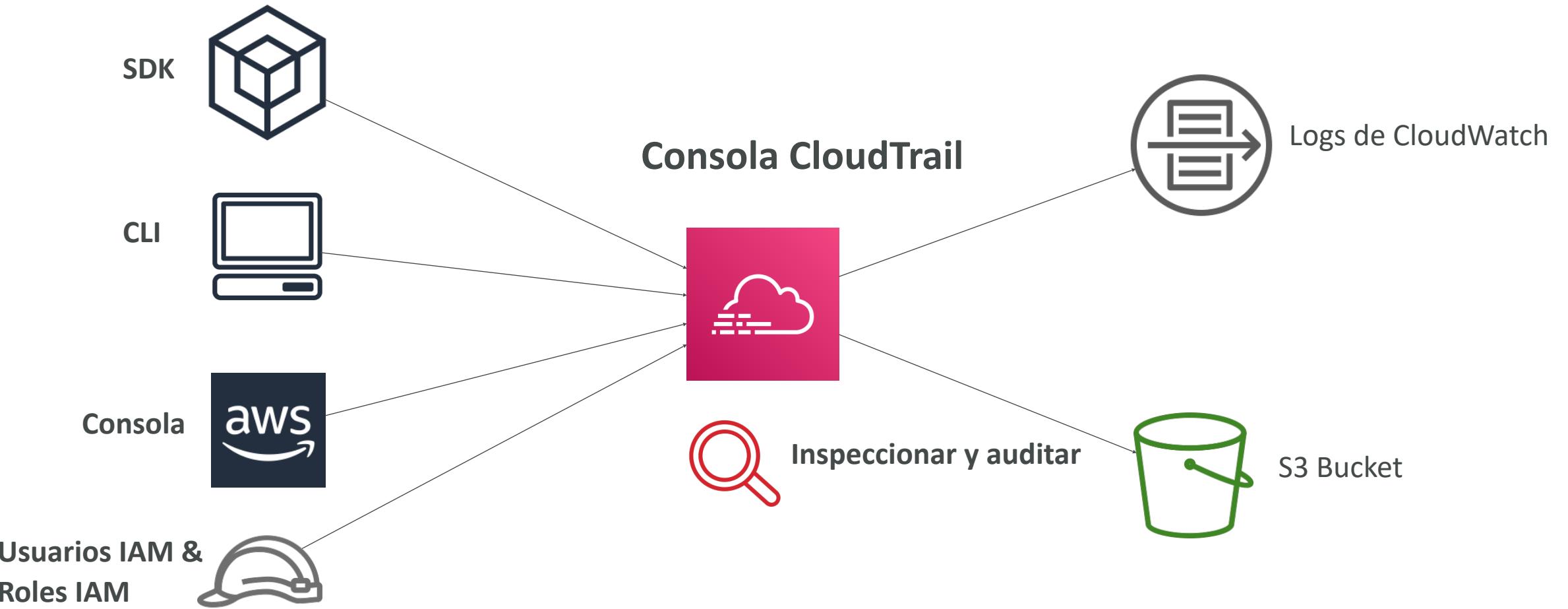
- Dashboards automáticos para solucionar problemas de tu aplicación y los servicios de AWS relacionados

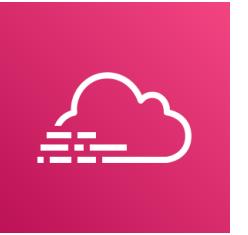
AWS CloudTrail



- **Proporciona gobernanza, normativa y auditoría para tu cuenta de AWS**
- CloudTrail está activado por defecto
- Obtén **un historial de eventos / llamadas a la API realizadas dentro de tu Cuenta de AWS** por:
 - Consola
 - SDK
 - CLI
 - Servicios de AWS
- Puedes poner logs de CloudTrail en CloudWatch Logs o S3
- **Un rastro (trail) puede aplicarse a todas las regiones (por defecto) o a una sola región**
- Si se elimina un recurso en AWS, ¡investiga primero en CloudTrail!

Diagrama de CloudTrail





Eventos CloudTrail

- **Eventos de gestión:**

- Operaciones que se realizan en los recursos de tu cuenta de AWS
- Ejemplos:
 - Configurar la seguridad (IAM **AttachRolePolicy**)
 - Configurar reglas para enrutar datos (Amazon EC2 **CreateSubnet**)
 - Configurar logs (AWS CloudTrail **CreateTrail**)
- **Por defecto, los trails están configurados para logs de eventos de gestión.**
- Pueden separar los **eventos de lectura** (que no modifican los recursos) de los Eventos de Escritura (que pueden modificar los recursos)

- **Eventos de datos:**

- **Por defecto, los eventos de datos no se registran (porque son operaciones de gran volumen)**
- Actividad a nivel de objeto de Amazon S3 (ej: **GetObject**, **DeleteObject**, **PutObject**): puede separar los Eventos de Lectura de los de Escritura
- Actividad de ejecución de funciones de AWS Lambda (la API **Invoke**)

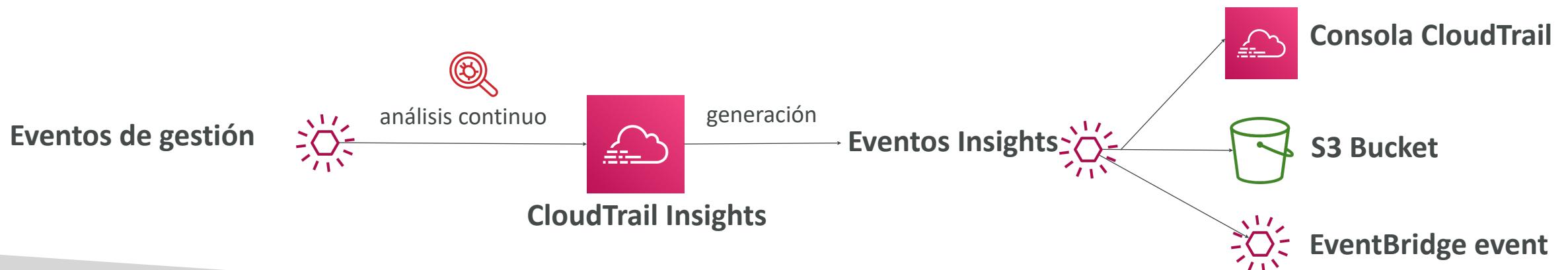
- **CloudTrail Insights Events:**

- Ver siguiente diapositiva ☺

CloudTrail Insights

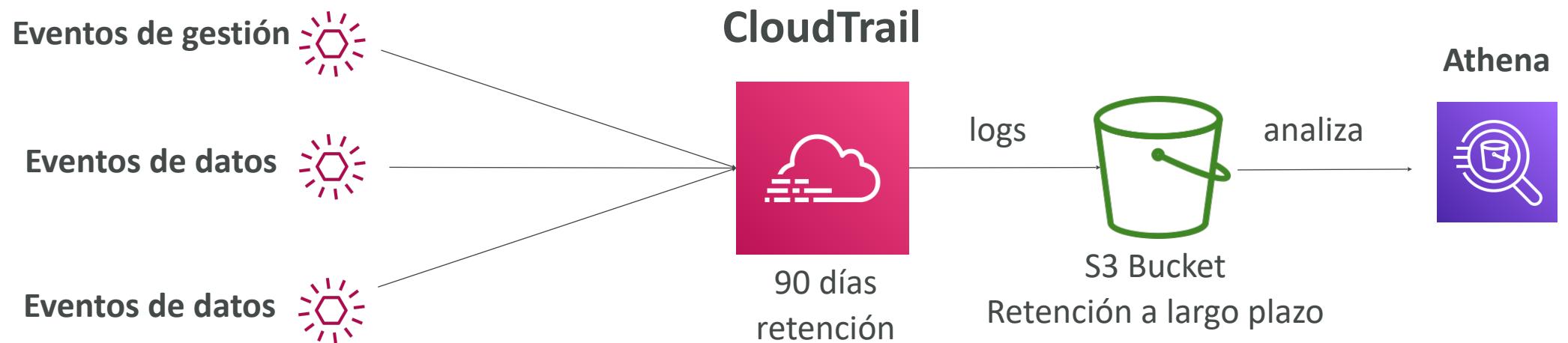


- Activa **CloudTrail Insights para detectar actividad inusual** en tu cuenta:
 - aprovisionamiento inexacto de recursos
 - superación de los límites de servicio
 - ráfagas de acciones de AWS IAM
 - lagunas en la actividad de mantenimiento periódico
- CloudTrail Insights analiza los eventos de gestión normales para crear una línea de base
- Y después **analiza continuamente los eventos de escritura para detectar patrones inusuales**
 - Las anomalías aparecen en la consola de CloudTrail
 - El evento se envía a Amazon S3
 - Se genera un evento EventBridge (para necesidades de automatización)

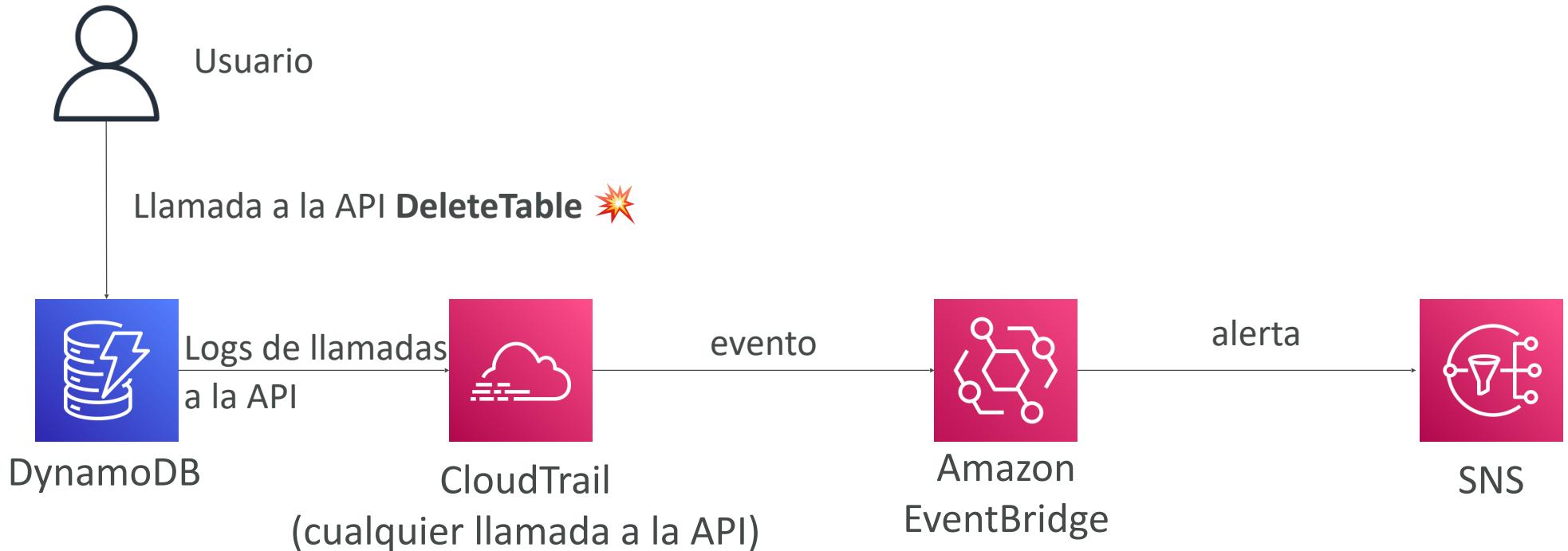


Retención de Eventos CloudTrail

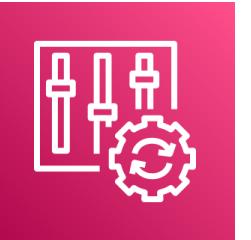
- Los eventos se almacenan durante 90 días en CloudTrail
- Para conservar los eventos más allá de este periodo, regístralos en S3 y utiliza Athena



Amazon EventBridge - Intercepción de llamadas a la API



AWS Config



- Ayuda a auditar y registrar la **normativa** de tus recursos de AWS
- Ayuda a registrar configuraciones y cambios a lo largo del tiempo
- Preguntas que se pueden resolver con AWS Config:
 - ¿Hay acceso SSH sin restricciones a mis grupos de seguridad?
 - ¿Mis buckets tienen acceso público?
 - ¿Cómo ha cambiado la configuración de mi ALB a lo largo del tiempo?
- Puedes recibir alertas (notificaciones SNS) de cualquier cambio
- AWS Config es un servicio por región
- Puede agregarse entre regiones y cuentas
- Posibilidad de almacenar los datos de configuración en S3 (analizados por Athena)

Reglas de configuración

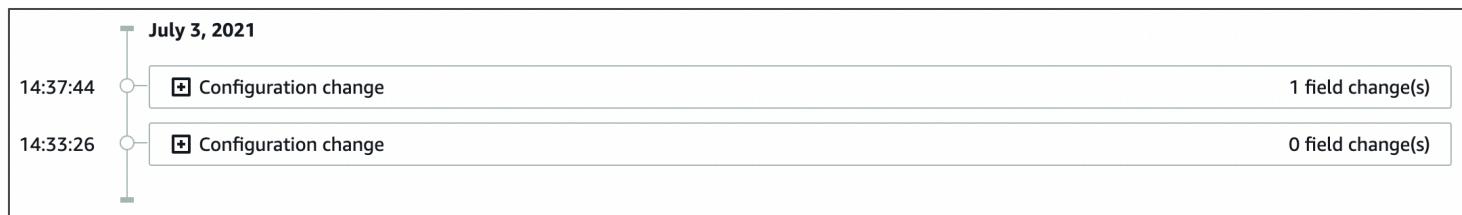
- Puedes utilizar reglas de configuración administradas por AWS Config (más de 75)
- Puedes crear reglas de configuración personalizadas (deben definirse en AWS Lambda)
 - Ej: evaluar si cada disco EBS es de tipo gp2
 - Ej: evaluar si cada instancia EC2 es t2.micro
- Las reglas se pueden evaluar / activar
 - Por cada cambio de configuración
 - Y / o: a intervalos de tiempo regulares
- **AWS Config Rules no impide que se produzcan acciones (no deniega)**
- Precios: sin capa gratuita, 0,003 \$ por elemento de configuración registrado por región, 0,001 \$ por evaluación de regla de configuración por región

Recurso AWS Config

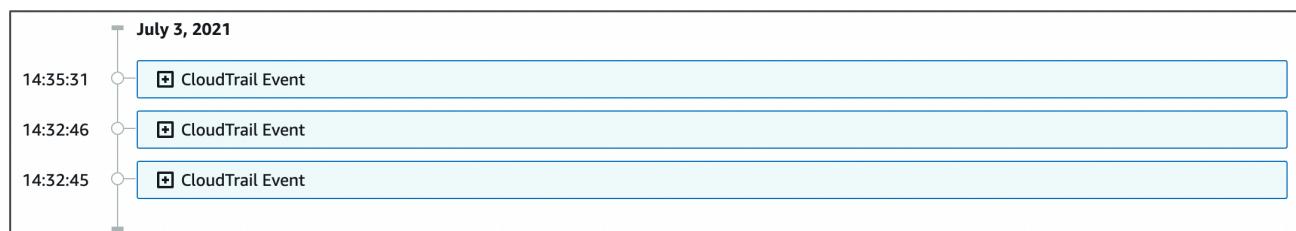
- Ver la normativa de un recurso a lo largo del tiempo

○ sg-077b425b1649da83e	EC2 SecurityGroup	✓ Compliant
○ sg-0831434f1876c0c74	EC2 SecurityGroup	⚠ Noncompliant
○ sg-09f10ed254d464f30	EC2 SecurityGroup	✓ Compliant

- Ver la configuración de un recurso a lo largo del tiempo

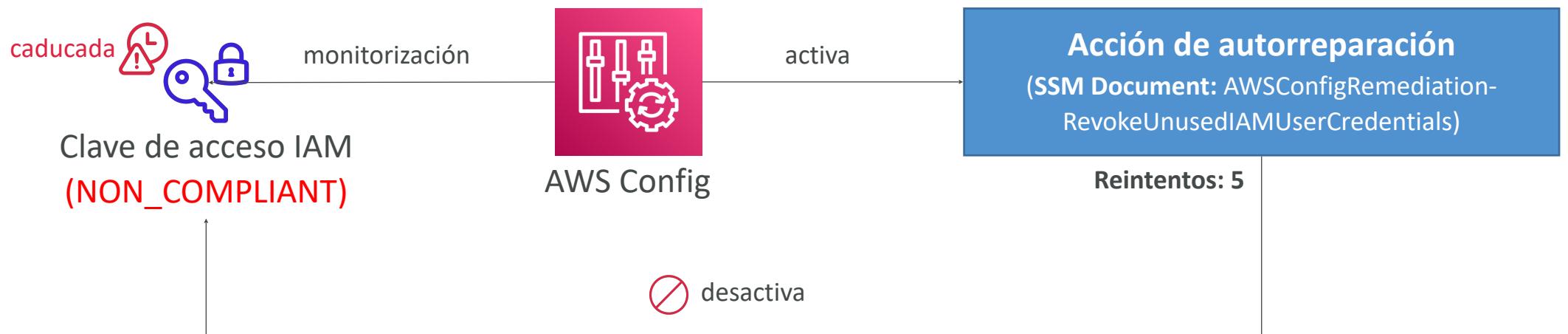


- Ver las llamadas a la API CloudTrail de un recurso a lo largo del tiempo



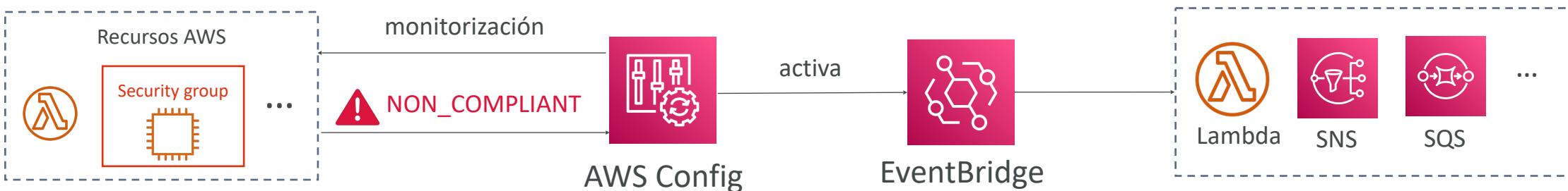
Reglas de configuración - Remediaciones

- Automatiza la corrección de recursos no conformes utilizando Documentos de Automatización SSM
- Utiliza Documentos de Automatización Administrados por AWS o crea Documentos de Automatización personalizados
 - Consejo: puedes crear Documentos de Automatización personalizados que invoquen la función Lambda
- Puedes establecer **reintentos de remediación** si el recurso sigue siendo no conforme después de la auto-remediation



Reglas de configuración - Notificaciones

- Utiliza EventBridge para activar notificaciones cuando los recursos de AWS no cumplan las normas



- Posibilidad de enviar notificaciones de cambios de configuración y de estado de la normativa al SNS (todos los eventos - utiliza el Filtrado SNS o filtra en el lado del cliente)



CloudWatch vs CloudTrail vs Config

- **CloudWatch**

- Monitorización del rendimiento (métricas, CPU, red, etc...) y dashboards
- Eventos y alertas
- Agregación y análisis de logs

- **CloudTrail**

- Registra las llamadas a la API realizadas dentro de tu Cuenta por cualquier persona
- Puedes definir trails para recursos específicos
- Servicio Global

- **Config**

- Registra los cambios de configuración
- Evalúa los recursos según las normas de cumplimiento
- Obtén una cronología de los cambios y de la normativa

Para un Elastic Load Balancer

- **CloudWatch:**

- Monitorización de la métrica de conexiones entrantes
- Visualiza los códigos de error en % a lo largo del tiempo
- Crea un dashboards para hacerte una idea del rendimiento de tu Load Balancer

- **Config:**

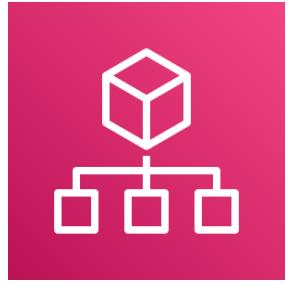
- Seguimiento de las reglas de grupos de seguridad para el Load Balancer
- Seguimiento de los cambios de configuración del Load Balancer
- Asegúrate de que siempre se asigna un certificado SSL al Load Balancer (normativa)

- **CloudTrail:**

- Rastrea quién ha realizado cambios en el Load Balancer con llamadas a la API

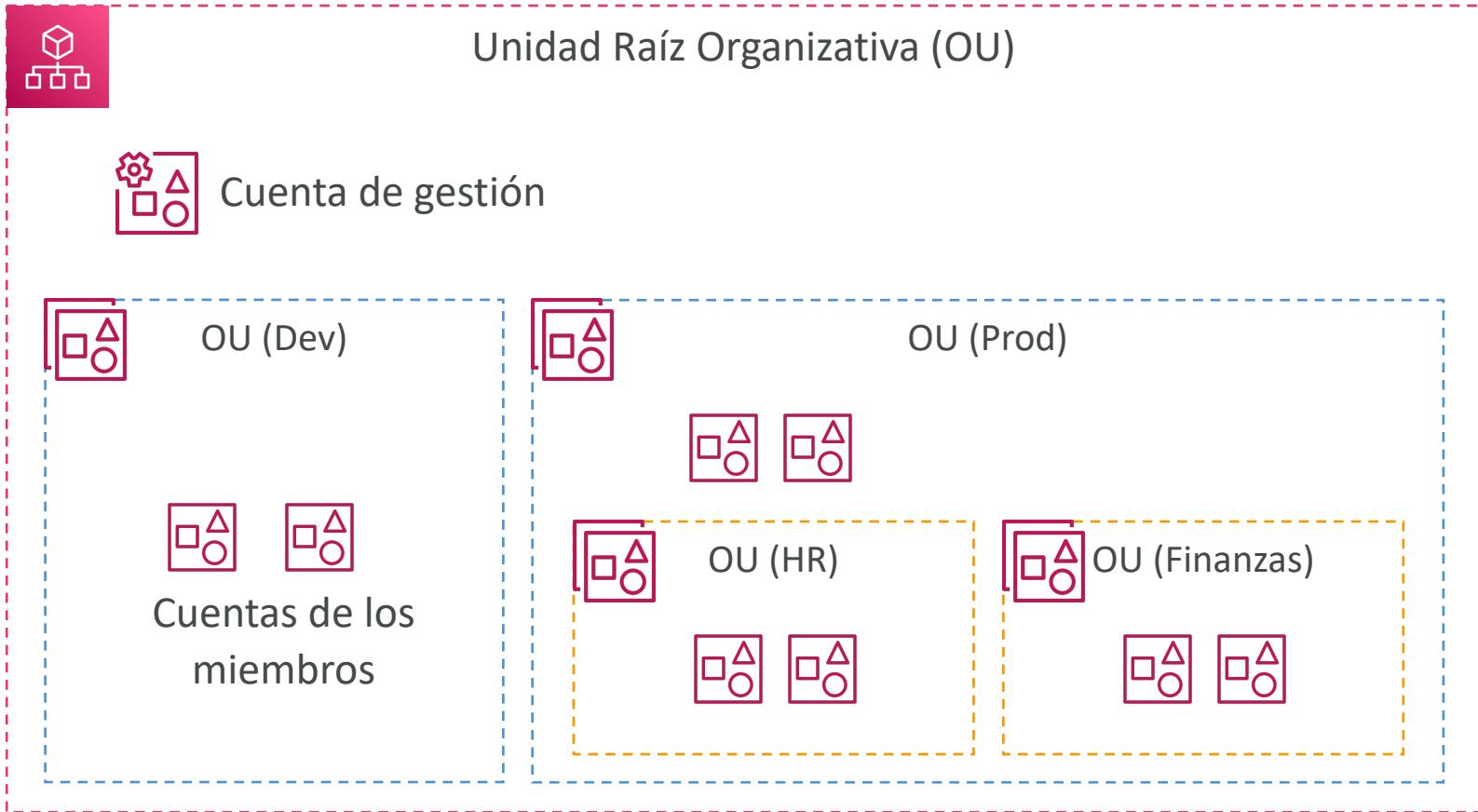
Identidad avanzada en AWS

AWS Organizations



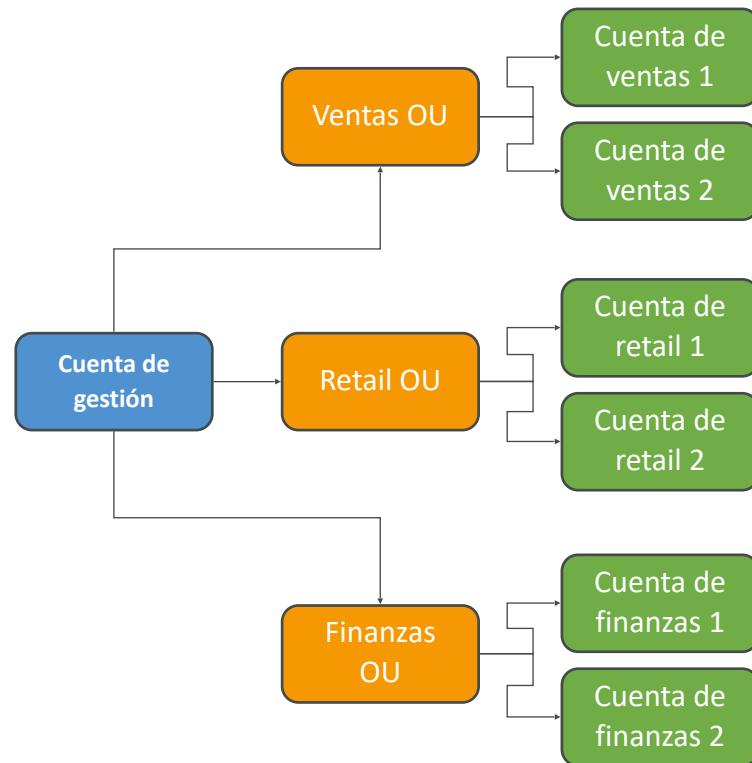
- Servicio global
- Permite administrar varias cuentas de AWS
- La cuenta principal es la cuenta de administración
- Las demás cuentas son cuentas miembro
- Las cuentas miembro sólo pueden formar parte de una organización
- Facturación consolidada en todas las cuentas: un único método de pago
- Los precios se benefician del uso agregado (descuento por volumen para EC2, S3...)
- **Instancias reservadas compartidas y descuentos de Planes de Ahorro en todas las cuentas**
- API disponible para automatizar la creación de cuentas AWS

AWS Organizations

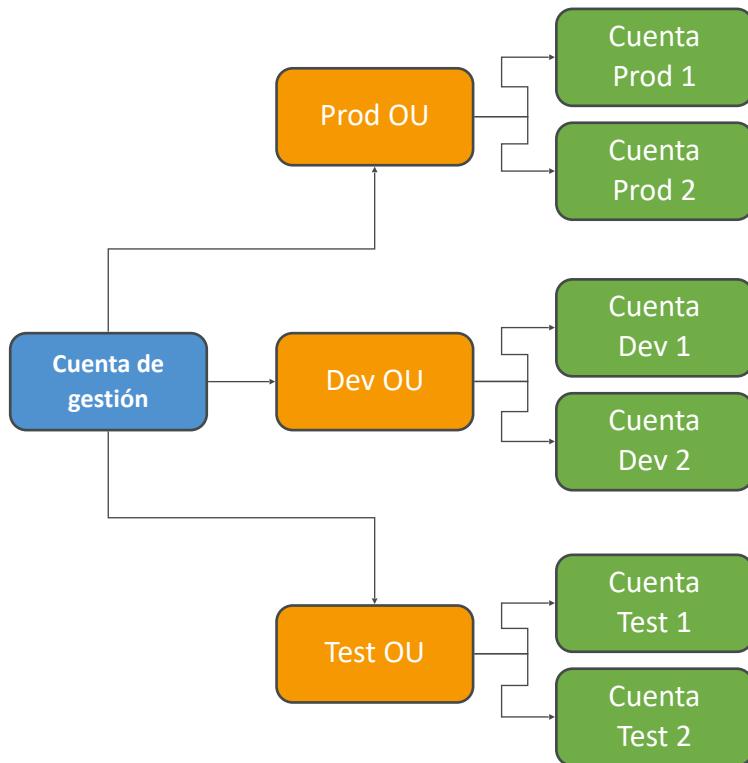


Unidades organizativas (UO) - Ejemplos

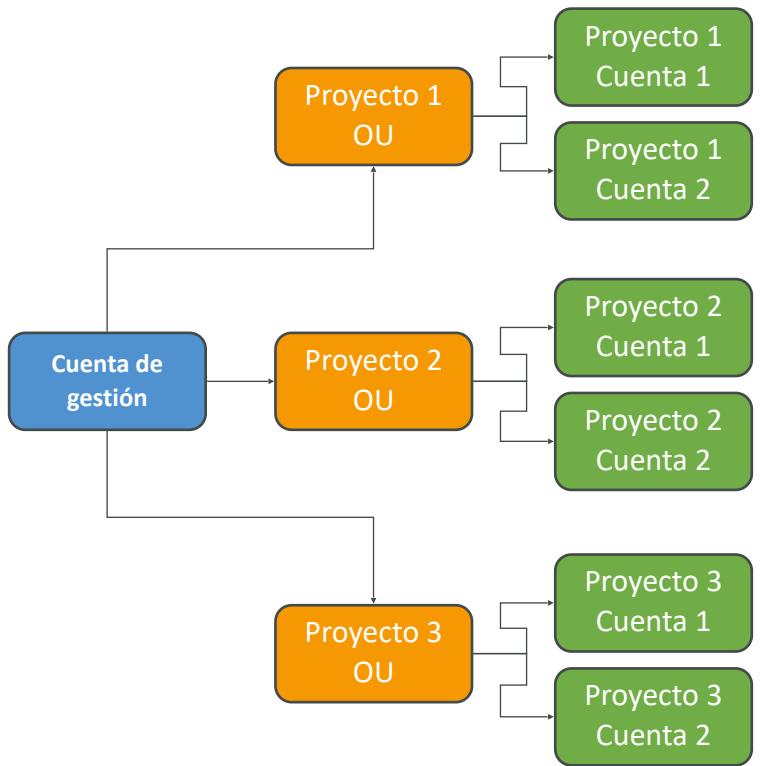
Unidad de negocio



Ciclo de vida medioambiental



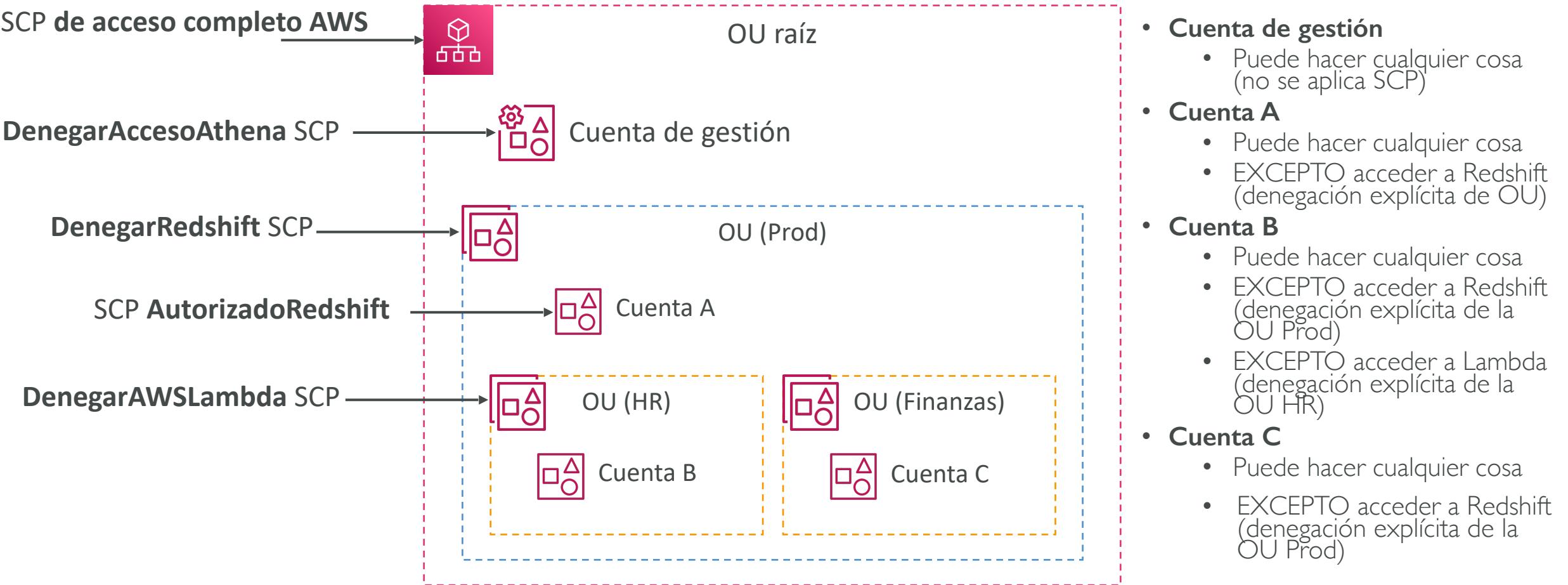
Basado en proyectos



AWS Organizations

- **Ventajas**
 - Multi Cuenta vs Una Cuenta Multi VPC
 - Utiliza normas de etiquetado con fines de facturación
 - Activar CloudTrail en todas las cuentas, enviar logs a cuenta S3 central
 - Enviar logs de CloudWatch a la cuenta central de logs
 - Establece roles entre cuentas para fines administrativos
- **Seguridad: Políticas de Control de Servicios (SCP)**
 - Políticas IAM aplicadas a OU o Cuentas para restringir usuarios y roles
 - No se aplican a la cuenta de gestión (plenos poderes de administrador)
 - Deben tener un permiso explícito (no permiten nada por defecto - como IAM)

Jerarquía SCP



Ejemplos SCP Estrategias Blocklist y Allowlist

```
"Version": "2012-10-17",
"Statement": [
    {
        "Sid": "AllowsAllActions",
        "Effect": "Allow",
        "Action": "*",
        "Resource": "*"
    },
    {
        "Sid": "DenyDynamoDB",
        "Effect": "Deny",
        "Action": "dynamodb:*",
        "Resource": "*"
    }
]
```

```
"Version": "2012-10-17",
"Statement": [
    {
        "Effect": "Allow",
        "Action": [
            "ec2:*",
            "cloudwatch:*
```

Más ejemplos: https://docs.aws.amazon.com/organizations/latest/userguide/orgs_manage_policies_example-scps.html

Condiciones IAM

aws:SourceIP

restringe la IP del cliente desde el que se realizan las llamadas a la API

```
{  
  "Version": "2012-10-17",  
  "Statement": [  
    {  
      "Effect": "Deny",  
      "Action": "*",  
      "Resource": "*",  
      "Condition": {  
        "NotIpAddress": {  
          "aws:SourceIp": ["192.0.2.0/24", "203.0.113.0/24"]  
        }  
      }  
    }  
  ]  
}
```

aws:RequestedRegion

restringir la región a la que se hacen las llamadas a la API

```
{  
  "Version": "2012-10-17",  
  "Statement": [  
    {  
      "Effect": "Deny",  
      "Action": ["ec2:*", "rds:*", "dynamodb:*"],  
      "Resource": "*",  
      "Condition": {  
        "StringEquals": {  
          "aws:RequestedRegion": ["eu-central-1", "eu-west-1"]  
        }  
      }  
    }  
  ]  
}
```

Condiciones IAM

ec2:ResourceTag

restringir en función de las etiquetas

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": ["ec2:startInstances", "ec2:StopInstances"],
      "Resource": "arn:aws:ec2:us-east-1:123456789012:instance/*",
      "Condition": {
        "StringEquals": {
          "ec2:ResourceTag/Project": "DataAnalytics",
          "aws:PrincipalTag/Department": "Data"
        }
      }
    }
  ]
}
```

aws:MultiFactorAuthPresent

para forzar el uso de MFA

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "ec2:*",
      "Resource": "*"
    },
    {
      "Effect": "Deny",
      "Action": ["ec2:StopInstances", "ec2:TerminateInstances"],
      "Resource": "*",
      "Condition": {
        "BoolIfExists": {
          "aws:MultiFactorAuthPresent": false
        }
      }
    }
  ]
}
```

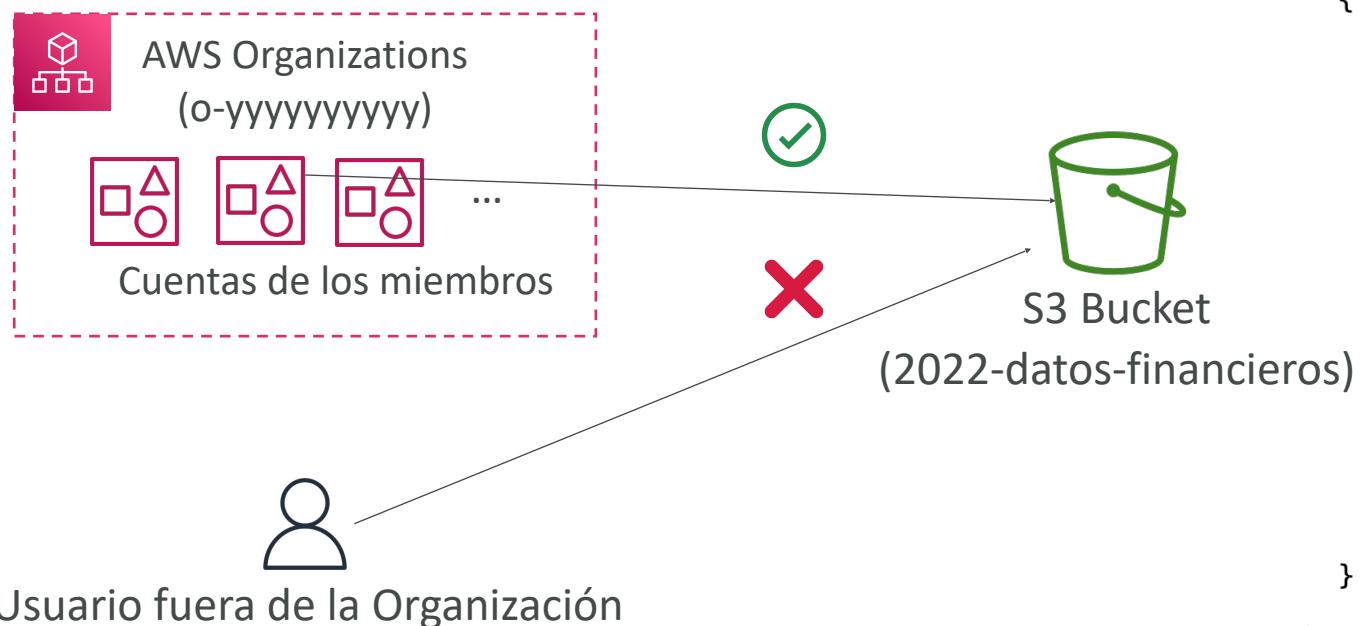
IAM para S3

- El permiso **s3>ListBucket** se aplica a **arn:aws:s3:::test**
- => permiso a nivel de bucket
- **s3GetObject, s3PutObject, s3DeleteObject** se aplica a **arn:aws:s3:::test/***
- => permiso de nivel de objeto

```
{  
    "Version": "2012-10-17",  
    "Statement": [  
        {  
            "Effect": "Allow",  
            "Action": ["s3>ListBucket"],  
            "Resource": "arn:aws:s3:::test"  
        },  
        {  
            "Effect": "Allow",  
            "Action": [  
                "s3:PutObject",  
                "s3:GetObject",  
                "s3>DeleteObject"  
            ],  
            "Resource": "arn:aws:s3:::test/*"  
        }  
    ]  
}
```

Políticas de recursos & aws:PrincipalOrgID

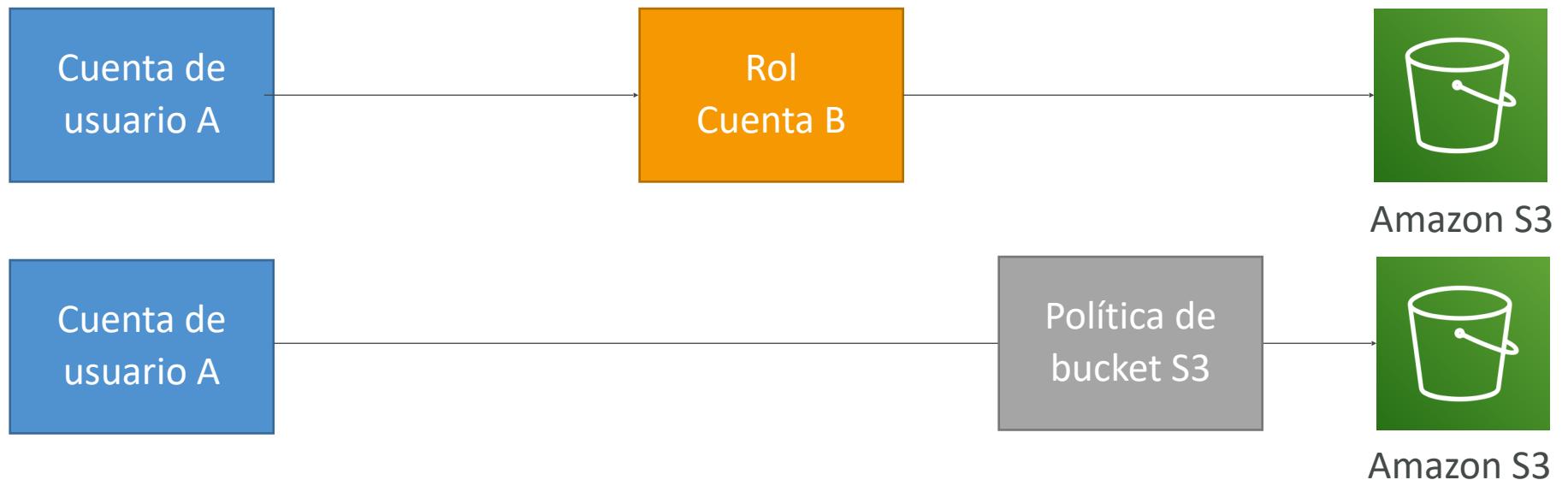
- **aws:PrincipalOrgID** puede utilizarse en cualquier política de recursos para restringir el acceso a cuentas que sean miembros de una Organización de AWS



```
{  
  "Version": "2012-10-17",  
  "Statement": [  
    {  
      "Effect": "Allow",  
      "Action": ["s3:PutObject", "s3:GetObject"],  
      "Resource": "arn:aws:s3:::2022-financial-data/*",  
      "Condition": {  
        "StringEquals": {  
          "aws:PrincipalOrgID": ["o-yyyyyyyyyy"]  
        }  
      }  
    }  
  ]  
}
```

Roles IAM vs Políticas basadas en recursos

- Cuenta cruzada:
 - adjuntando una política basada en recursos a un recurso (ejemplo: política de bucket S3)
 - O utilizando un rol como proxy

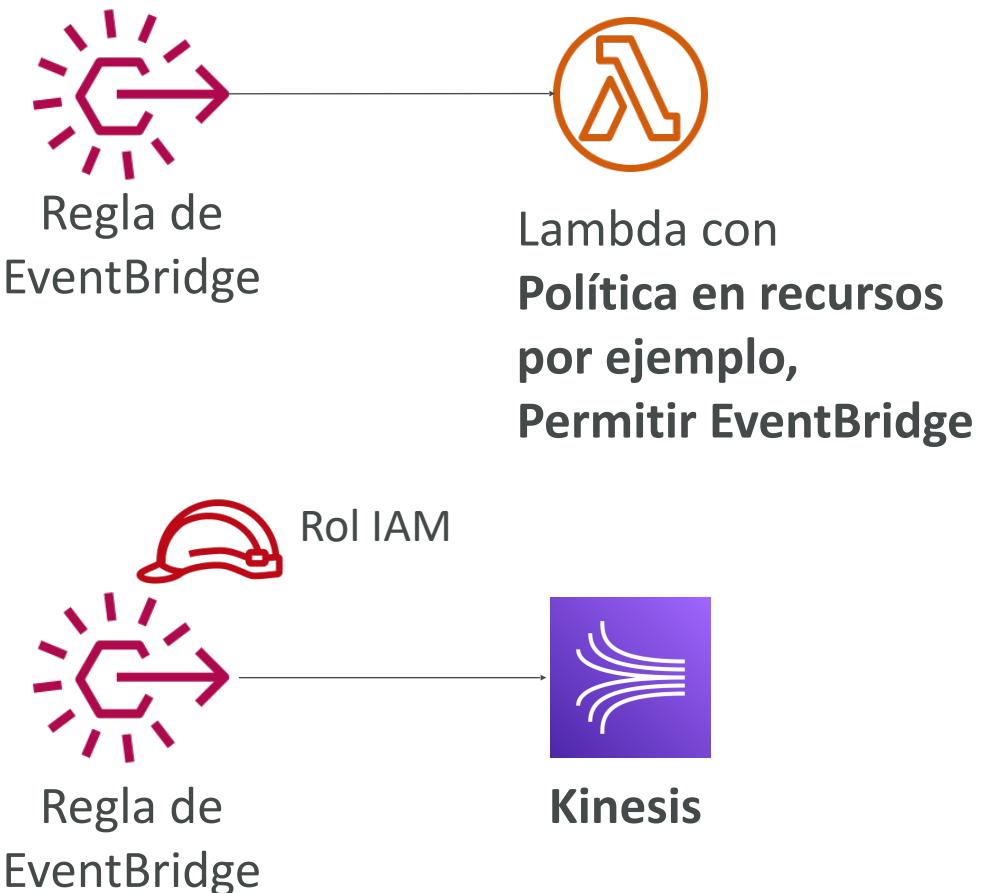


Roles IAM vs Políticas basadas en recursos

- **Cuando asumes un rol (usuario, aplicación o servicio), renuncias a tus permisos originales y tomas los permisos asignados al rol**
- Cuando se utiliza una política basada en recursos, el principal no tiene que renunciar a sus permisos
- Ejemplo: El usuario de la cuenta A necesita escanear una tabla DynamoDB de la cuenta A y volcarla en un bucket S3 de la cuenta B.
- Soportado por: Amazon S3 buckets, SNS Topic, SQS queues, etc...

Amazon EventBridge - Seguridad

- Cuando se ejecuta una regla, necesita permisos en el objetivo
- Política basada en recursos:
Lambda, SNS, SQS, CloudWatch Logs, API Gateway...
- Rol IAM: Kinesis stream, Systems Manager Run Command, ECS task...



Límites de permisos de IAM

- Los Límites de Permisos de IAM se soportan para usuarios y roles (no para grupos)
- Función avanzada para utilizar una política gestionada para establecer los permisos máximos que puede obtener una entidad IAM.

Ejemplo:

```
{  
    "Version": "2012-10-17",  
    "Statement": [  
        {  
            "Effect": "Allow",  
            "Action": [  
                "s3:*",  
                "cloudwatch:*",  
                "ec2:*"  
            ],  
            "Resource": "*"  
        }  
    ]  
}
```



```
{  
    "Version": "2012-10-17",  
    "Statement": [  
        {  
            "Effect": "Allow",  
            "Action": "iam>CreateUser",  
            "Resource": "*"  
        }  
    ]  
}
```



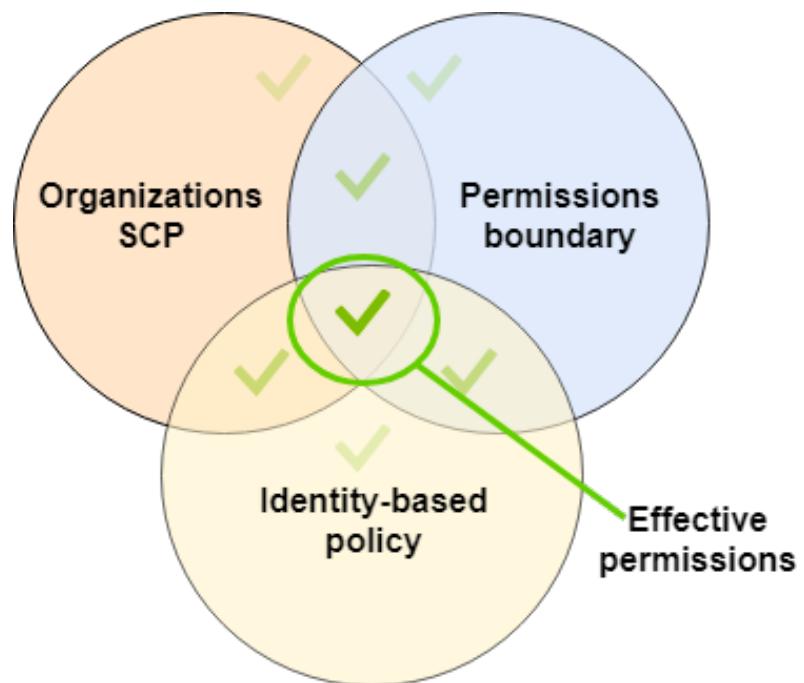
Sin permisos

Límite del permiso IAM

Permisos IAM
A través de la Política IAM

Límites de permisos de IAM

- Puede utilizarse en combinaciones de AWS Organizations SCP

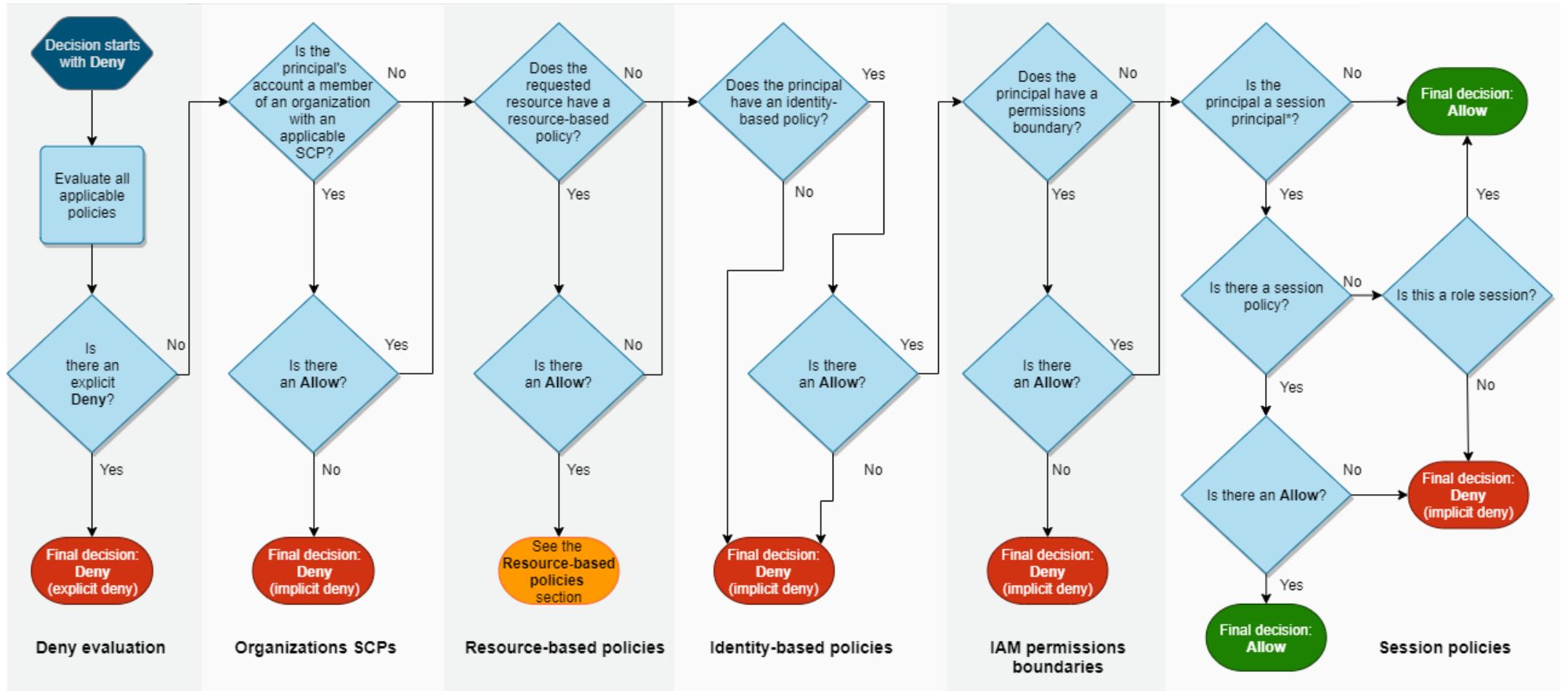


https://docs.aws.amazon.com/IAM/latest/UserGuide/access_policies_boundaries.html

Casos prácticos

- Delegar responsabilidades a no administradores dentro de sus límites de permisos, por ejemplo crear nuevos usuarios IAM
- Permitir que los desarrolladores se autoasignen políticas y gestionen sus propios permisos, asegurándose al mismo tiempo de que no puedan "escalar" sus privilegios (= hacerse admin)
- Útil para restringir a un usuario concreto (en lugar de a toda una cuenta mediante Organizaciones y SCP)

Lógica de evaluación de la política IAM



*A session principal is either a role session or an IAM federated user session.

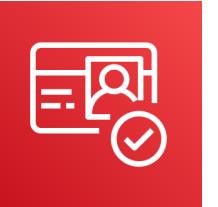
https://docs.aws.amazon.com/IAM/latest/UserGuide/reference_policies_evaluation-logic.html

Ejemplo de Política IAM

- ¿Puedes ejecutar sqs>CreateQueue?
- ¿Puedes ejecutar sqs>DeleteQueue?
- ¿Puedes ejecutar ec2:DescribeInstances?

```
Version: "2012-10-17",
Statement: [
  {
    Action: "sts:AssumeRole",
    Effect: "Allow",
    Resource: "*"
  },
  {
    Action: [
      "sts:AssumeRole"
    ],
    Effect: "Allow",
    Resource: "*"
  }
]
```

Amazon Cognito



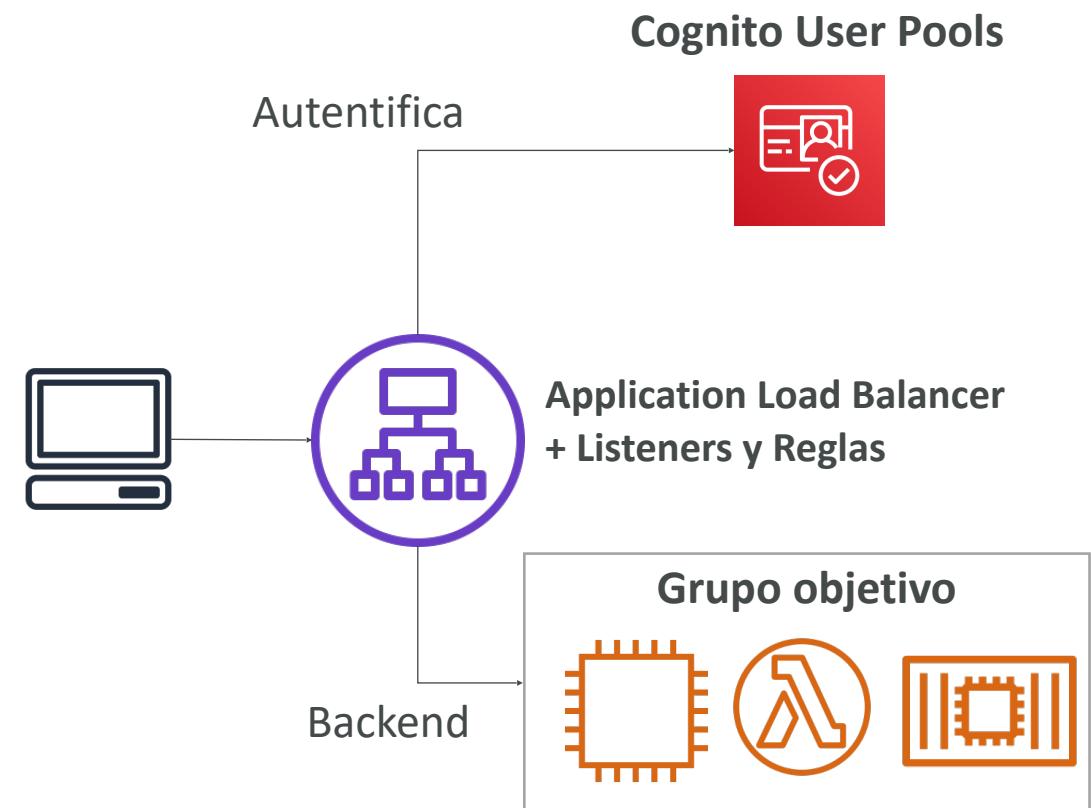
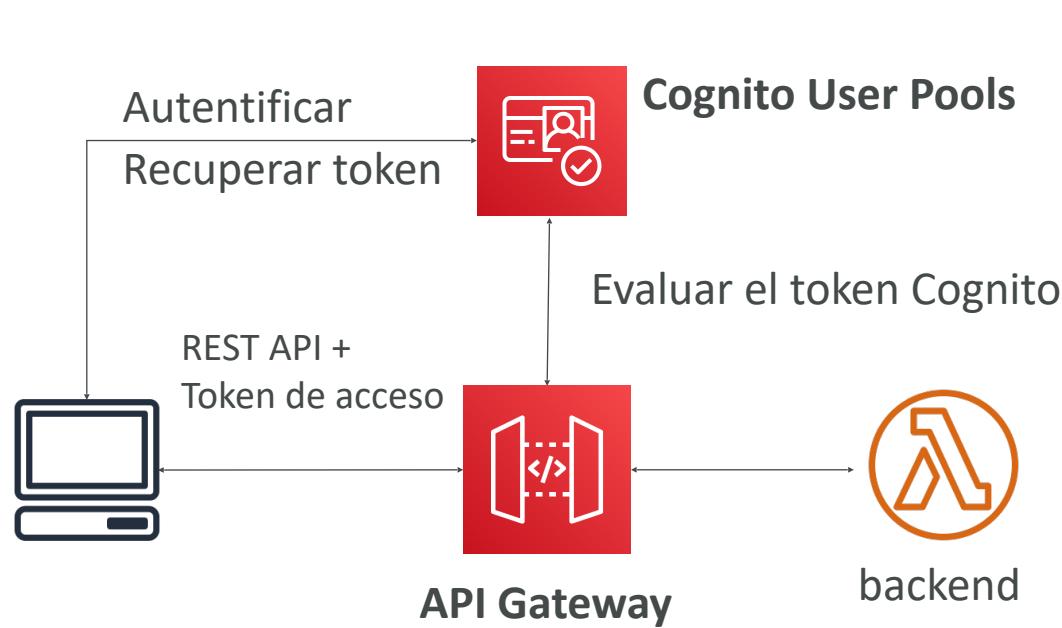
- Dar a los usuarios una identidad para interactuar con nuestra aplicación web o móvil
- **Grupos de usuarios Cognito (Cognito User Pools):**
 - Funcionalidad de inicio de sesión para usuarios de aplicaciones
 - Integración con API Gateway y Application Load Balancer
- **Cognito Identity Pools (Identidad Federada):**
 - Proporciona credenciales de AWS a los usuarios para que puedan acceder directamente a los recursos de AWS
 - Integrar con Cognito User Pools como proveedor de identidades
- **Cognito vs IAM:** "cientos de usuarios", "usuarios móviles", "autenticar con SAML"

Cognito User Pools (CUP) - Características del usuario

- **Crea una base de datos de usuarios sin servidor para tus aplicaciones web y móviles**
- Inicio de sesión sencillo: Combinación de nombre de usuario (o correo electrónico) / contraseña
- Restablecimiento de contraseña
- Verificación de correo electrónico y número de teléfono
- Autenticación multifactor (MFA)
- Identidades federadas: usuarios de Facebook, Google, SAML...

Cognito User Pools (CUP) - Integraciones

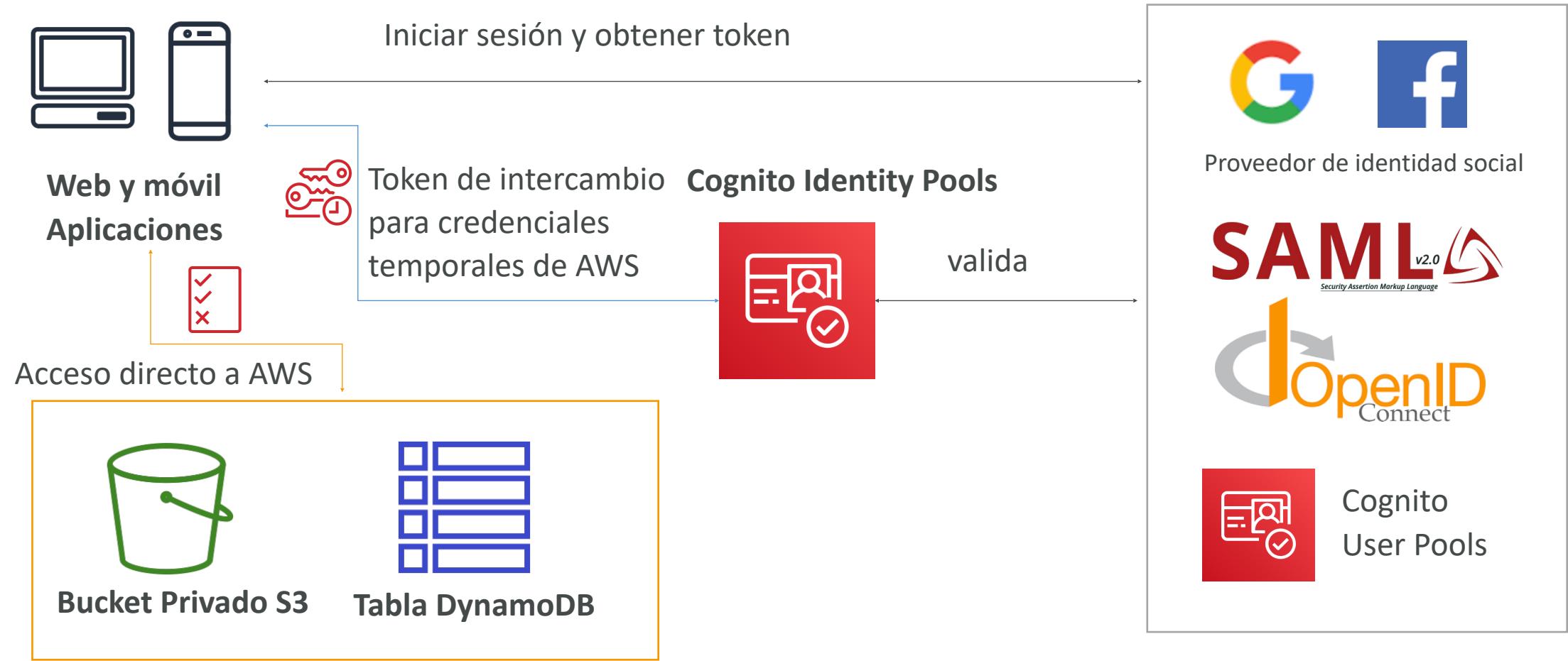
- CUP se integra con API Gateway y Application Load Balancer



Cognito Identity Pools (Identidades Federadas)

- **Obtener identidades para "usuarios" para que obtengan credenciales temporales de AWS**
- El origen de los usuarios puede ser Cognito User Pools, inicios de sesión de terceros, etc.
- **Los usuarios pueden acceder a los servicios de AWS directamente o a través de API Gateway**
 - Las políticas IAM aplicadas a las credenciales se definen en Cognito
 - Se pueden personalizar en función del user_id para un control más preciso.
 - **Roles IAM por defecto** para usuarios autenticados e invitados

Cognito Identity Pools - Diagrama



Cognito Identity Pools

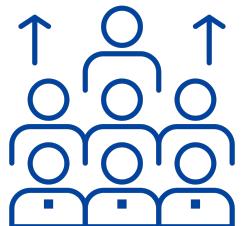
Seguridad a nivel de fila en DynamoDB

```
{  
    "Version": "2012-10-17",  
    "Statement": [  
        {  
            "Effect": "Allow",  
            "Action": [  
                "dynamodb:GetItem", "dynamodb:BatchGetItem", "dynamodb:Query",  
                "dynamodb:PutItem", "dynamodb:UpdateItem", "dynamodb:DeleteItem",  
                "dynamodb:BatchWriteItem"  
            ],  
            "Resource": [  
                "arn:aws:dynamodb:us-west-2:123456789012:table/MyTable"  
            ],  
            "Condition": {  
                "ForAllValues:StringEquals": {  
                    "dynamodb:LeadingKeys": [  
                        "${cognito-identity.amazonaws.com:sub}"  
                    ]  
                }  
            }  
        }  
    ]  
}
```

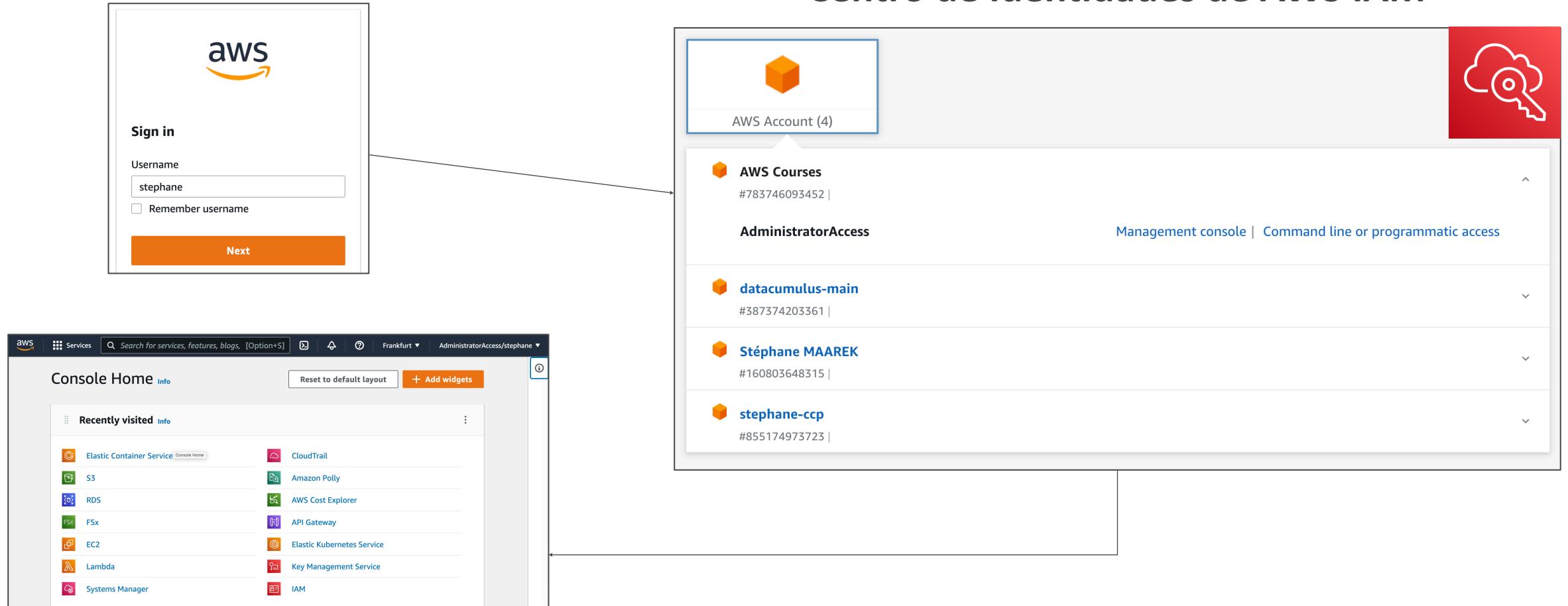
Centro de Identidad de AWS IAM (sucesor de AWS Single Sign-On)



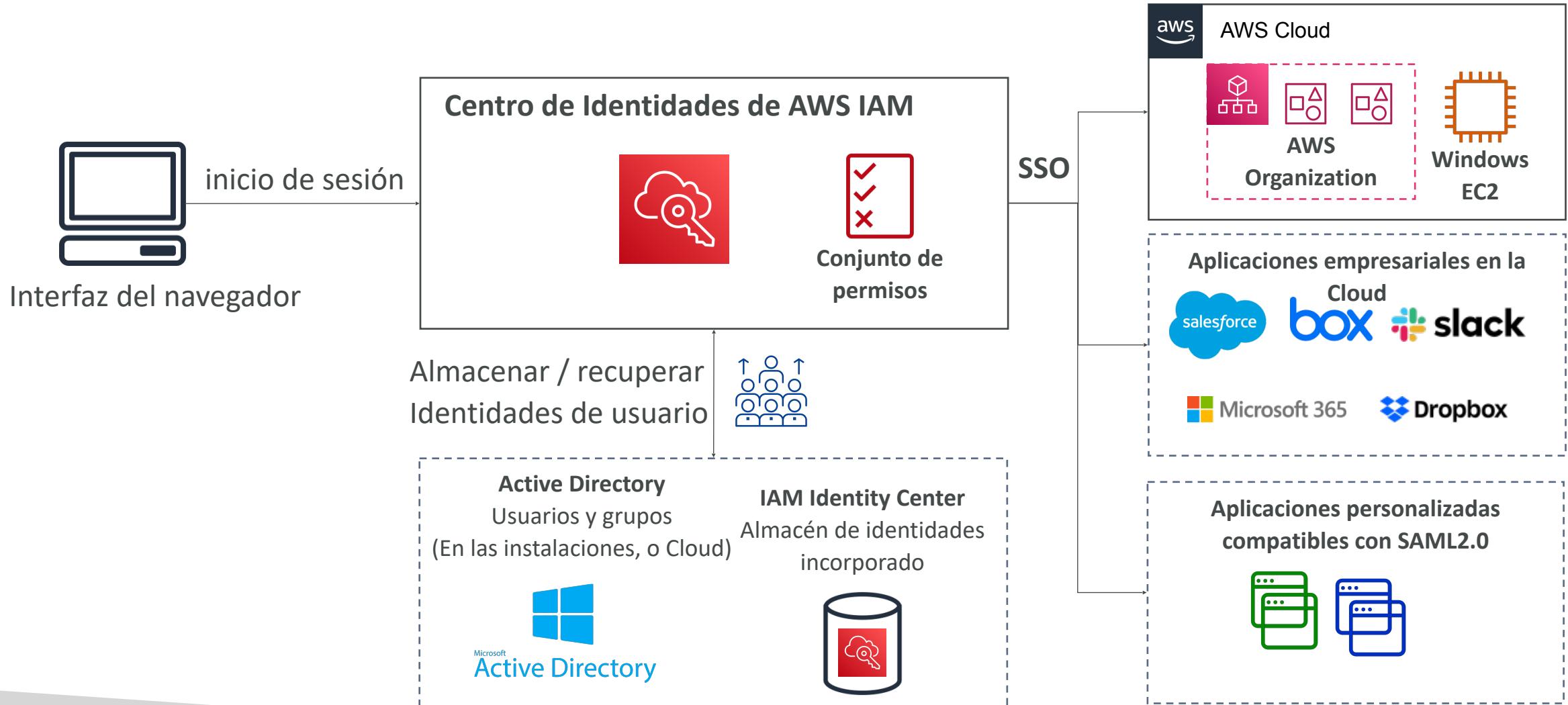
- Un inicio de sesión (inicio de sesión único) para todas tus
 - **cuentas de AWS en AWS Organizations**
 - Aplicaciones empresariales en el Cloud (por ejemplo, Salesforce, Box, Microsoft 365, ...)
 - Aplicaciones habilitadas para SAML2.0
 - Instancias de Windows EC2
- Proveedores de identidad
 - Almacén de identidades incorporado en el Centro de Identidades de IAM
 - De terceros: Active Directory (AD), OneLogin, Okta...



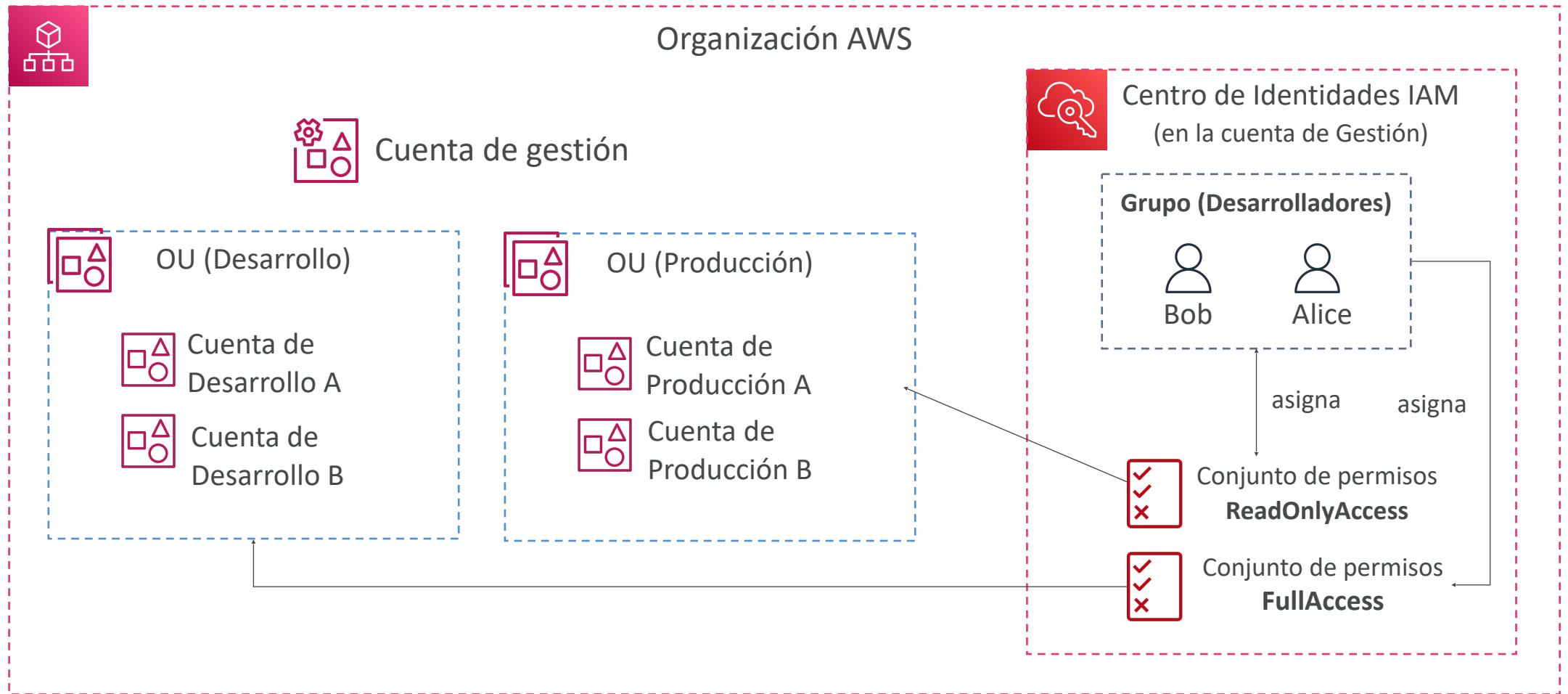
Centro de Identidades de AWS IAM - Flujo de inicio de sesión



Centro de Identidades de AWS IAM



Centro de Identidad IAM



Centro de Identidades de AWS IAM

Permisos y asignaciones en detalle



• Permisos multicuenta

- Gestiona el acceso a través de las cuentas de AWS en tu Organización AWS
- Conjuntos de permisos - una colección de una o más políticas IAM asignadas a usuarios y grupos para definir el acceso a AWS

• Asignaciones de aplicaciones

- Acceso SSO a muchas aplicaciones empresariales SAML 2.0 (Salesforce, Box, Microsoft 365, ...)
- Proporciona las URL, certificados y metadatos necesarios

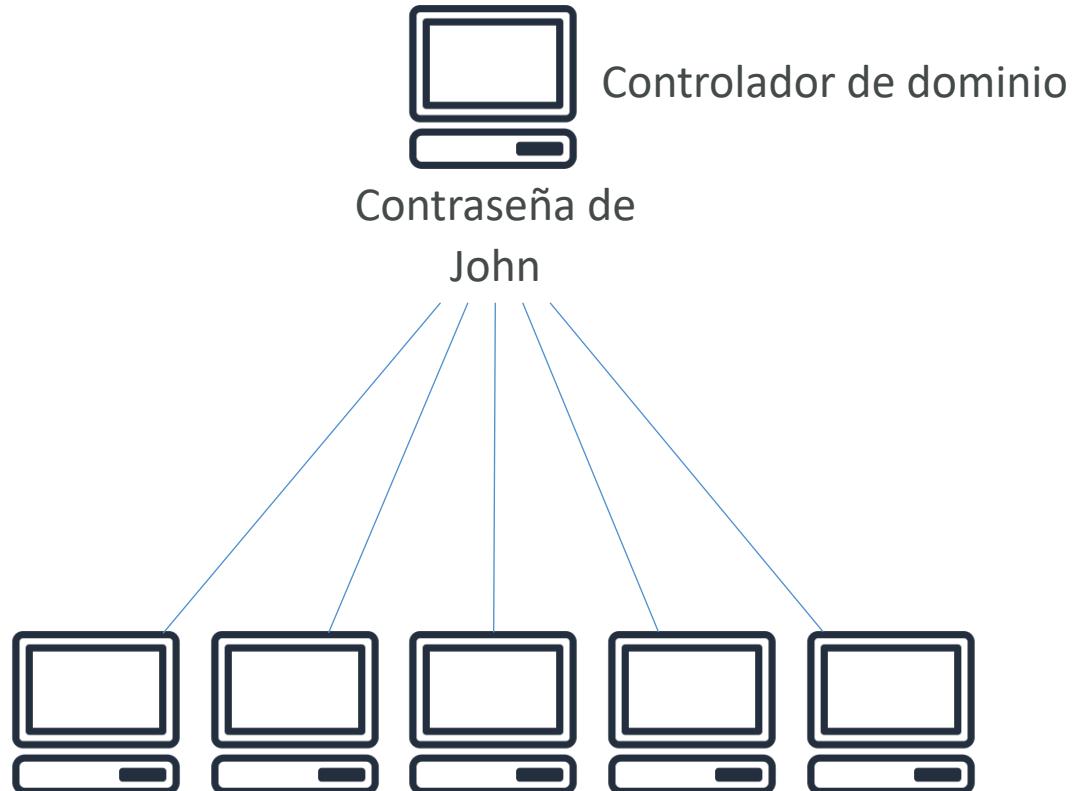
• Control de acceso basado en atributos (ABAC)

- Permisos detallados basados en los atributos de los usuarios almacenados en el Almacén de Identidades del Centro de Identidades IAM
- Ejemplo: centro de costes, cargo, configuración regional, ...
- Caso práctico: Define los permisos una vez, y luego modifica el acceso a AWS cambiando los atributos



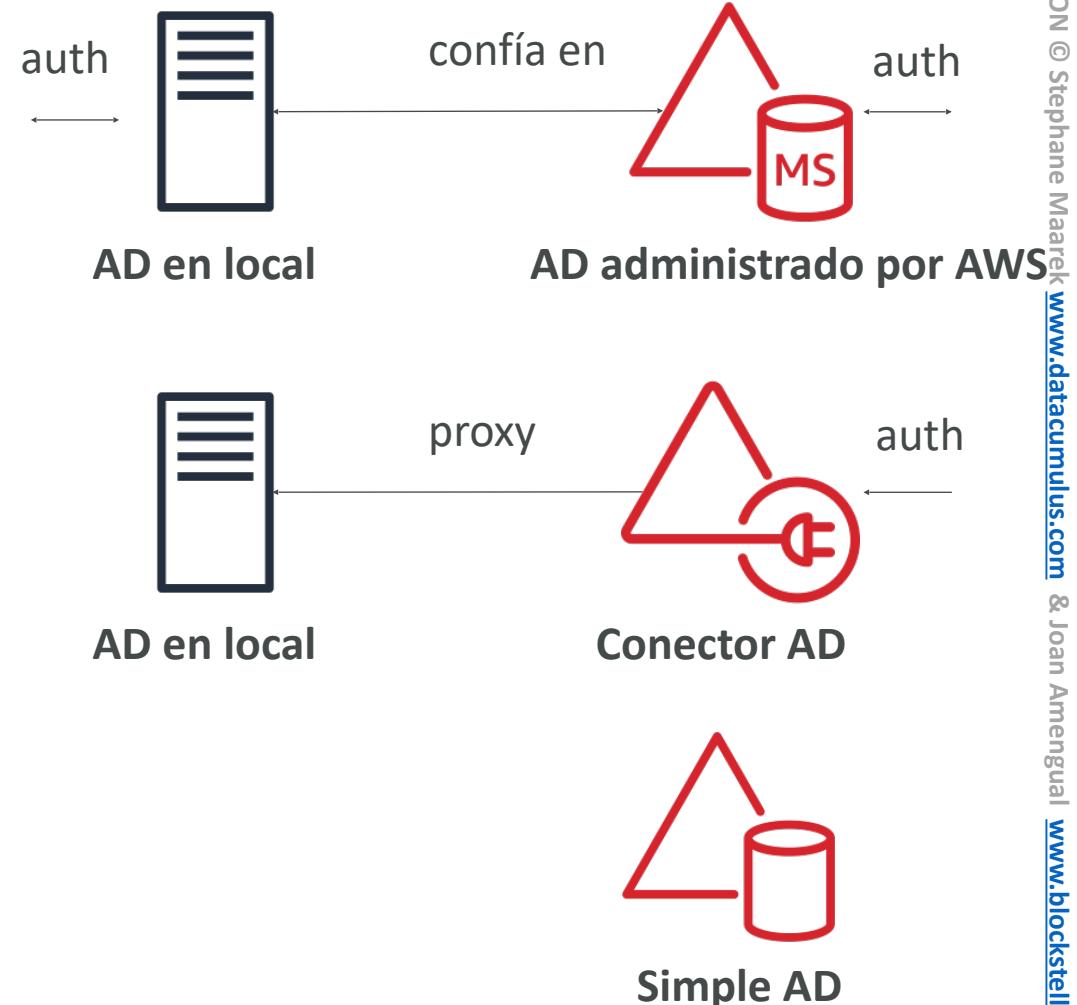
¿Qué es Microsoft Active Directory (AD)?

- Se encuentra en cualquier Servidor Windows con Servicios de Dominio AD
- Base de datos de **objetos**: Cuentas de usuario, ordenadores, impresoras, archivos compartidos, grupos de seguridad
- Gestión centralizada de la seguridad, crear cuenta, asignar permisos
- Los objetos se organizan en **árboles**
- Un grupo de árboles es un **bosque**



Servicios de directorio de AWS

- **Microsoft AD administrado por AWS**
 - Crea tu propio AD en AWS, administra usuarios localmente, soporta MFA
 - Establece conexiones de "confianza" con tu AD local
- **Conektor AD**
 - Directory Gateway (proxy) para redirigir al AD local, soporta MFA
 - Los usuarios se gestionan en el AD local
- **AD simple**
 - Directorio gestionado compatible con AD en AWS
 - No se puede unir con AD local



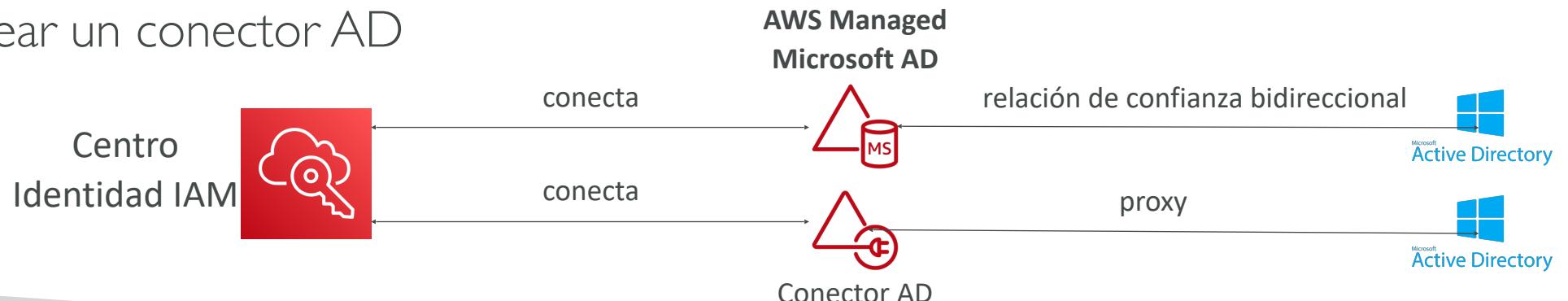
Centro de Identidades IAM - Configuración del Directorio Activo

- **Conectarse a un Microsoft AD (Servicio de directorio) administrado por AWS**
 - La integración está fuera de la caja



- **Conectarse a un directorio autogestionado**

- Crear una relación de confianza bidireccional utilizando Microsoft AD administrado por AWS
- Crear un conector AD



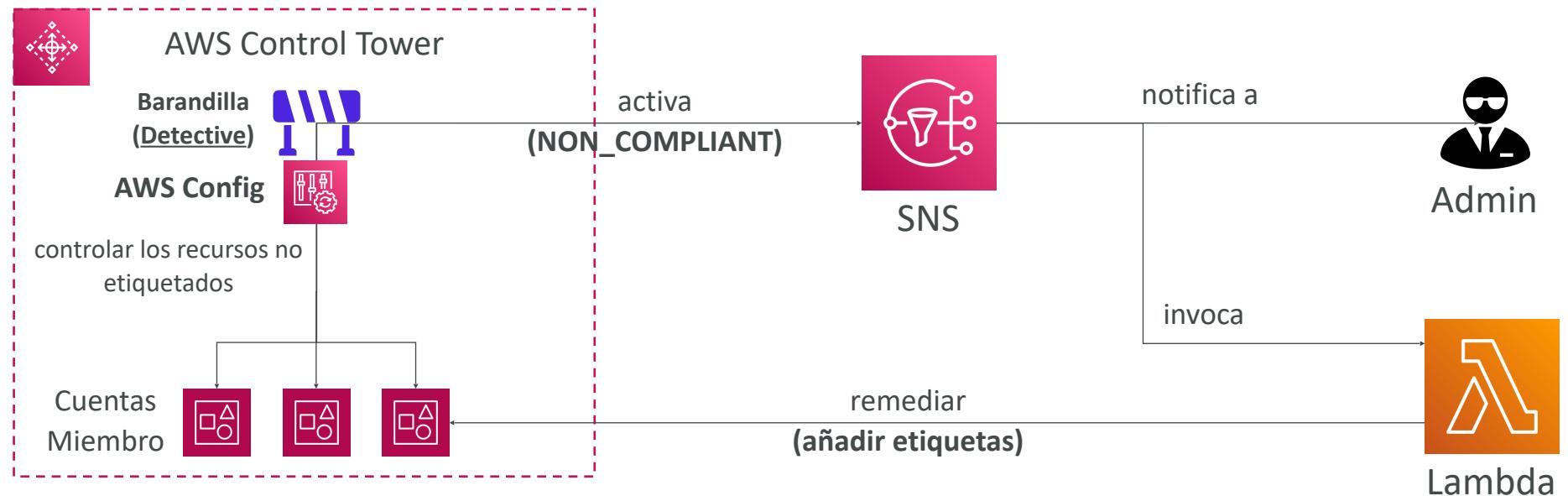
AWS Control Tower



- Forma sencilla de **configurar y gobernar un entorno AWS multicuenta seguro** y conforme a las mejores prácticas
- AWS Control Tower utiliza AWS Organizations para crear cuentas
- Ventajas:
 - Automatiza la configuración de tu entorno con unos pocos clics
 - Automatiza la gestión continua de las políticas mediante *guardrails*
 - Detecta las infracciones de las políticas y corrígelas
 - Controla la normativa mediante un dashboards interactivo

AWS Control Tower - Guardarraíles / Barandillas

- Proporciona gobernanza continua para tu Entorno de AWS Control Tower (Cuentas de AWS)
- **Guardrail preventivo - utilizando SCP** (por ejemplo, restringir regiones en todas tus cuentas)
- **Guardrail Detectivo - utilizando AWS Config** (por ejemplo, identificar recursos no etiquetados)

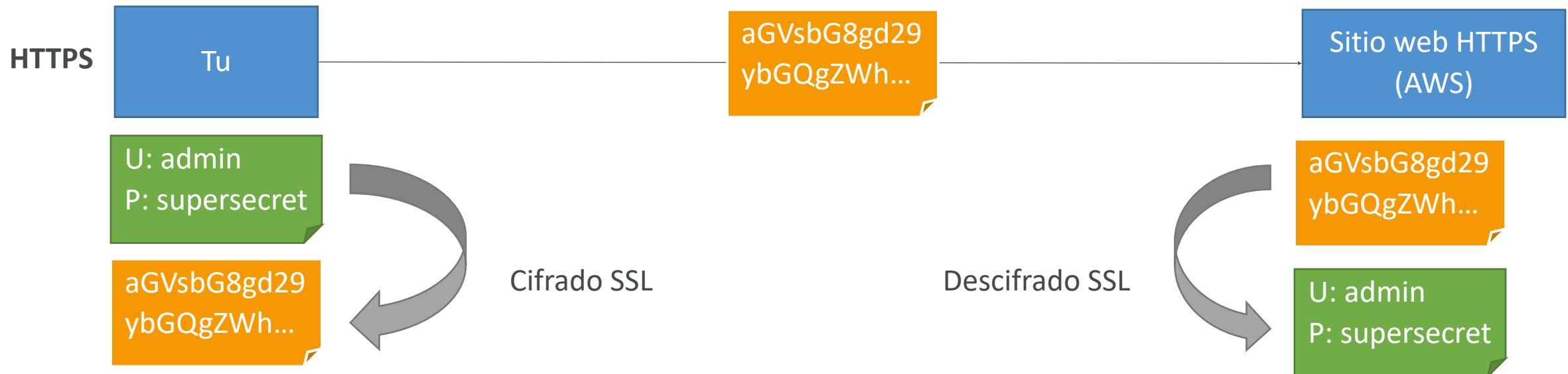


Seguridad y cifrado de AWS

KMS, Encryption SDK, SSM Parameter Store

¿Por qué cifrado? Cifrado en vuelo (SSL)

- Los datos se cifran antes de enviarlos y se descifran después de recibirlos
- Los certificados SSL ayudan al cifrado (HTTPS)
- El cifrado en vuelo garantiza que no pueda producirse un ataque MITM (man in the middle attack)



¿Por qué cifrado?

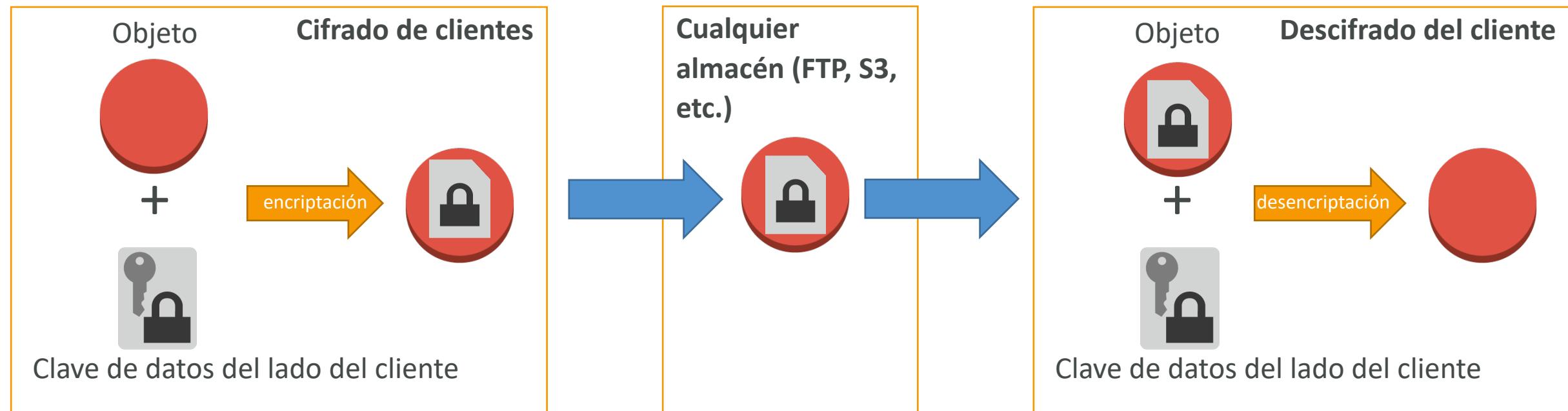
Cifrado del lado del servidor en reposo

- Los datos se cifran después de ser recibidos por el servidor
- Los datos se descifran antes de ser enviados
- Se almacenan cifrados gracias a una clave (normalmente una clave de datos)
- Las claves de cifrado/descifrado deben gestionarse en algún lugar y el servidor debe tener acceso a ellas



¿Por qué cifrado? Cifrado del lado del cliente

- Los datos son cifrados por el cliente y nunca descifrados por el servidor
- Los datos serán descifrados por un cliente receptor
- El servidor no debería poder descifrar los datos



AWS KMS (Servicio de administración de claves)



- Cada vez que oigas "cifrado" para un servicio de AWS, lo más probable es que se trate de KMS
- AWS gestiona las claves de cifrado por nosotros
- Totalmente integrado con IAM para la autorización
- Forma sencilla de controlar el acceso a tus datos
- Capaz de auditar el uso de claves KMS mediante CloudTrail
- Perfectamente integrado en la mayoría de los servicios de AWS (EBS, S3, RDS, SSM...)
- **Nunca jamás almacenes tus secretos en texto plano, ¡especialmente en tu código!**
 - El cifrado de claves KMS también está disponible a través de llamadas a la API (SDK, CLI)
 - Los secretos cifrados pueden almacenarse en el código / variables de Entorno

Tipos de claves KMS

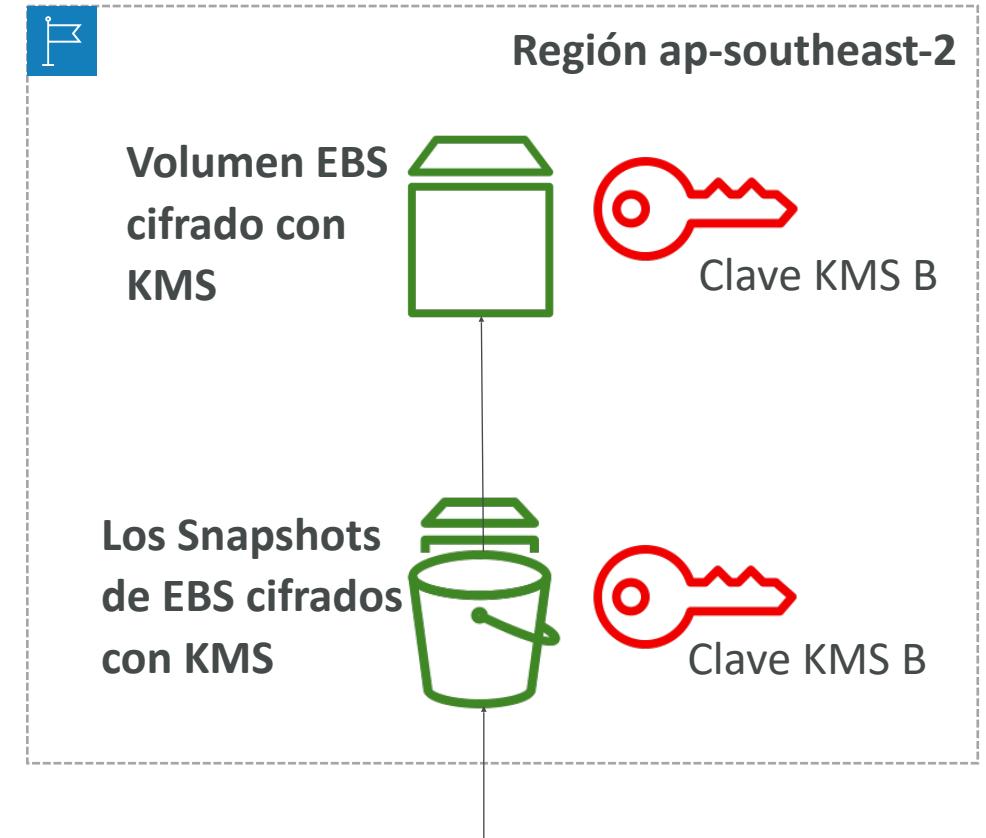
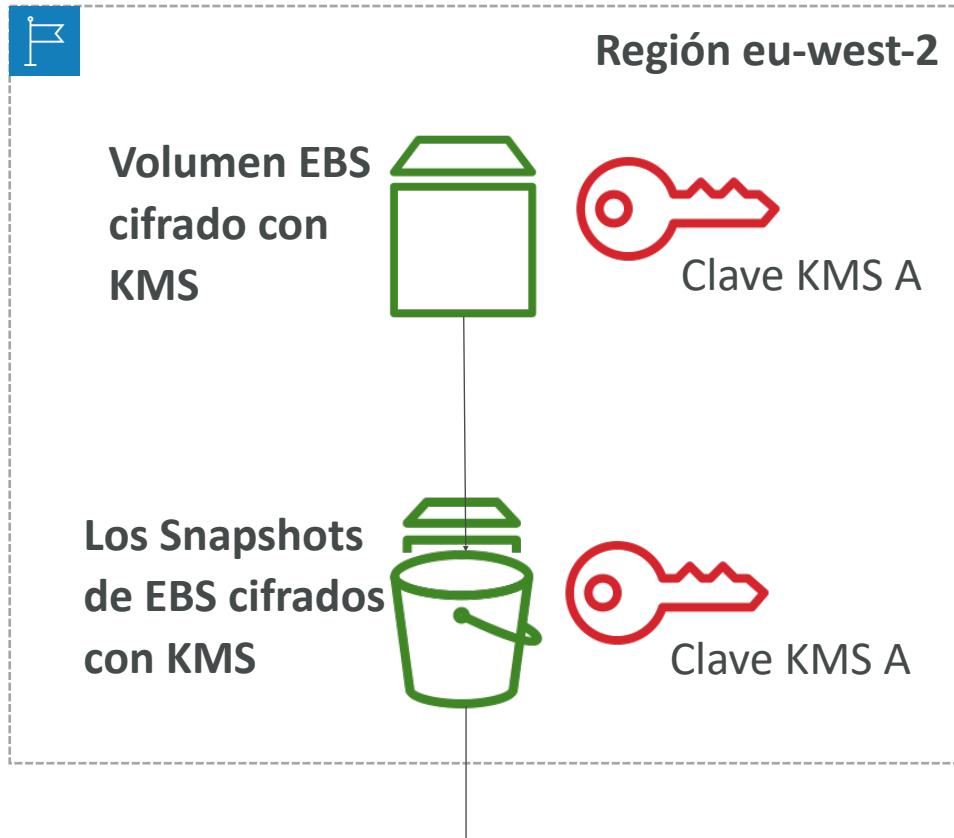
- **Claves KMS es el nuevo nombre de Clave Maestra de Cliente KMS**
- **Simétrica (claves AES-256)**
 - Clave de cifrado única que se utiliza para cifrar y descifrar
 - Los servicios de AWS integrados con KMS utilizan CMK Simétricas
 - Nunca tienes acceso a la clave KMS sin cifrar (debes llamar a la API KMS para utilizarla)
- **Asimétrica (pares de claves RSA y ECC)**
 - Par de claves pública (cifrar) y privada (descifrar)
 - Se utiliza para operaciones de Cifrar/Descifrar o Firmar/Verificar
 - La clave pública se puede descargar, pero no puedes acceder a la Clave Privada sin cifrar

AWS KMS (Servicio de administración de claves)



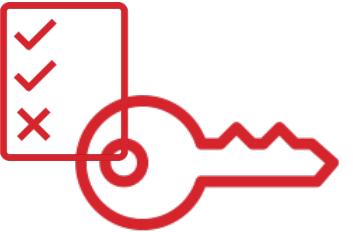
- Tres tipos de Claves KMS:
 - Clave administrada por AWS: **gratis** (aws/nombre-servicio, ejemplo: aws/rds o aws/ebs)
 - Claves gestionadas por el cliente (CMK) creadas en KMS: **I \$ / mes**
 - Claves gestionadas por el cliente importadas (deben ser claves simétricas de 256 bits): **I \$ / mes**
 - + pago por llamada API a KMS (0,03 \$ / 10000 llamadas)
- C
- Rotación automática de claves:
 - Clave KMS gestionada por AWS: automática cada 1 año
 - Clave KMS gestionada por el cliente: (debe estar activada) automática cada 1 año
 - Clave KMS importada: sólo es posible la rotación manual mediante alias

Copiar Snapshots entre regiones



Reencryptación KMS con clave B KMS

Políticas clave KMS



- Controlar el acceso a las claves KMS, "similar" a las políticas de bucket S3
- Diferencia: no puedes controlar el acceso sin ellas
- **Política de claves KMS por defecto:**
 - Se crea si no proporcionas una Política de Claves KMS específica
 - Acceso completo a la clave para el usuario root = toda la cuenta de AWS
- **Política de claves KMS personalizada:**
 - Define los usuarios y roles que pueden acceder a la clave KMS
 - Define quién puede administrar la clave
 - Útil para el acceso entre cuentas de tu clave KMS

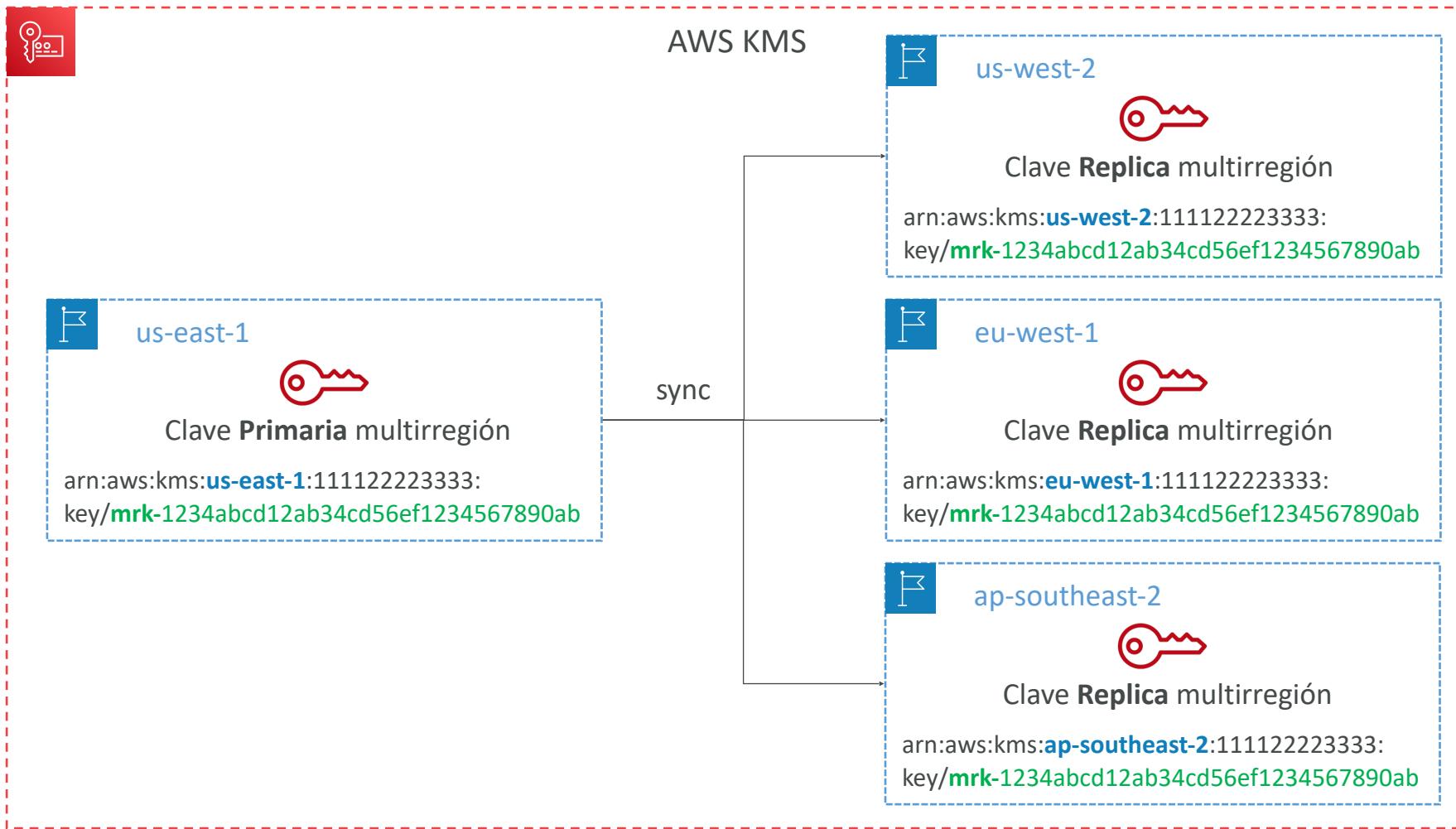
Copiar Snapshots entre cuentas

1. Crea una Snapshot, cifrada con tu propia Clave KMS (Clave Gestionada por el Cliente)
2. **Adjunta una Política de Clave KMS para autorizar el acceso entre cuentas**
3. Comparte la Snapshot cifrada
4. (en destino) Crea una copia de la Snapshot, cífrala con una CMK en tu cuenta
5. Crea un volumen a partir de la Snapshot

```
{  
  "Sid": "Allow use of the key with destination account",  
  "Effect": "Allow",  
  "Principal": {  
    "AWS": "arn:aws:iam::TARGET-ACCOUNT-ID:role/ROLENAMESPACE"  
  },  
  "Action": [  
    "kms:Decrypt",  
    "kms>CreateGrant"  
  ],  
  "Resource": "*",  
  "Condition": {  
    "StringEquals": {  
      "kms:ViaService": "ec2.REGION.amazonaws.com",  
      "kms:CallerAccount": "TARGET-ACCOUNT-ID"  
    }  
  }  
}
```

Política de claves KMS

Claves multirregión KMS



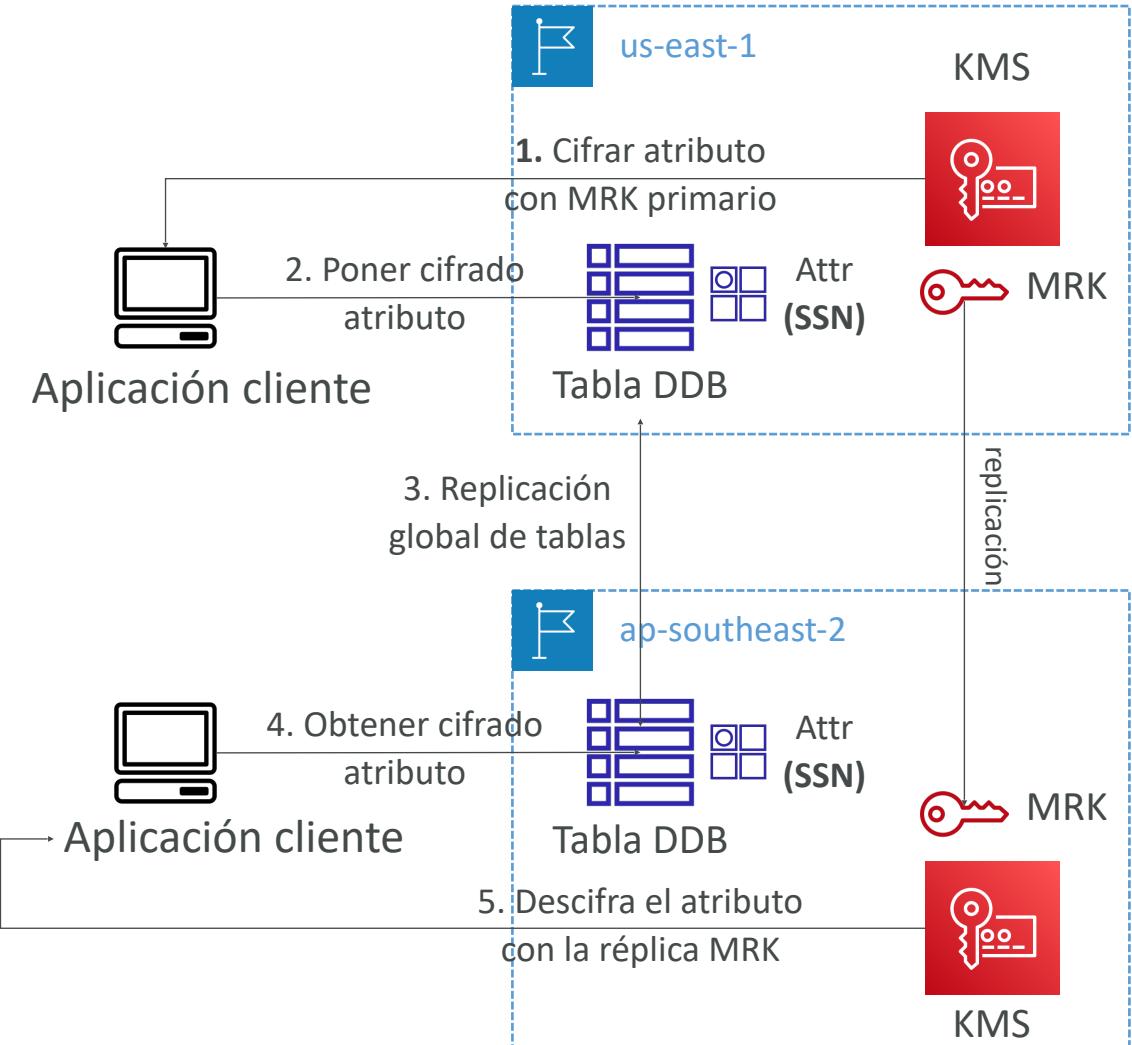
Claves multirregión KMS



- Llaves KMS idénticas en diferentes Regiones AWS que pueden utilizarse indistintamente
- Las claves multirregión tienen el mismo ID de clave, material de clave, rotación automática...
- Cifrado en una Región y descifrado en otras Regiones
- No es necesario volver a cifrar ni hacer llamadas a la API entre regiones
- Los KMS Multi-Región NO son globales (Primario + Rélicas)
- Cada clave Multi-Región se gestiona de forma **independiente**
- **Casos de uso:** cifrado global del lado del cliente, cifrado en DynamoDB Global, Aurora Global

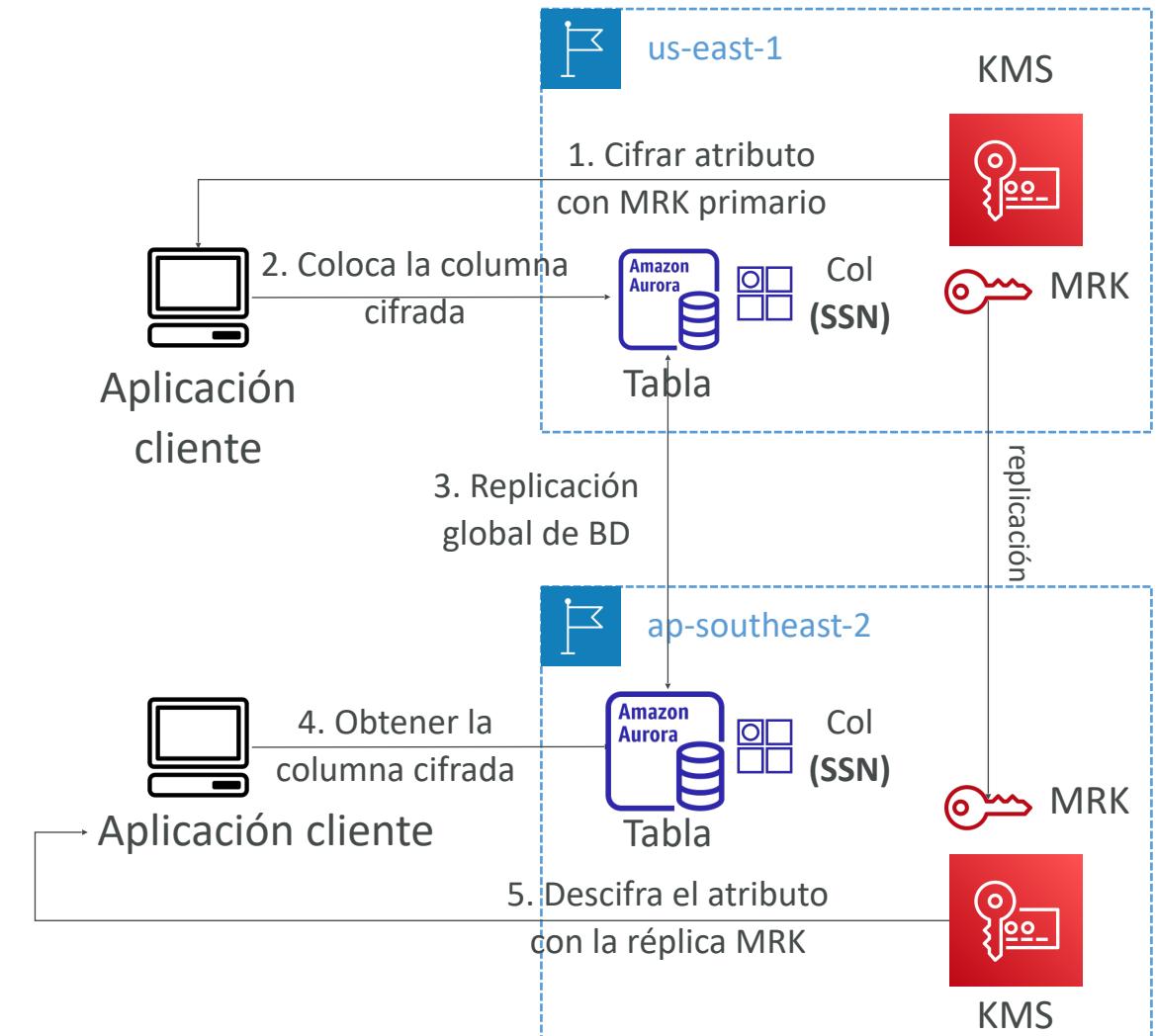
Cifrado del lado del cliente de tablas globales DynamoDB y claves multirregión KMS

- Podemos cifrar atributos específicos del lado del cliente en nuestra tabla DynamoDB utilizando el **Cliente de Cifrado de Amazon DynamoDB**
- En combinación con las Tablas Globales, los datos cifrados del lado del cliente se replican en otras regiones
- Si utilizamos una clave multirregión, replicada en la misma región que la tabla DynamoDB Global, los clientes de estas regiones pueden utilizar llamadas API de baja latencia al KMS de su región para descifrar los datos del lado del cliente
- Utilizando el cifrado del lado del cliente podemos proteger campos específicos y garantizar sólo el descifrado si el cliente tiene acceso a una clave API



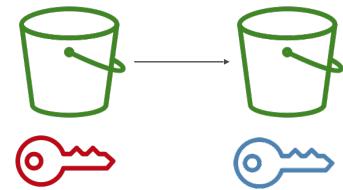
Cifrado del lado del cliente de las claves globales Aurora y KMS multirregión

- Podemos cifrar atributos específicos del lado del cliente en nuestra tabla Aurora utilizando el **SDK de cifrado de AWS**
- En combinación con las Tablas Globales Aurora, los datos cifrados del lado del cliente se replican a otras regiones
- Si utilizamos una clave multirregión, replicada en la misma región que la base de datos global de Aurora, los clientes de estas regiones pueden utilizar llamadas API de baja latencia al KMS de su región para descifrar los datos del lado del cliente.
- Usando el cifrado del lado del cliente podemos proteger campos específicos y garantizar sólo el descifrado si el cliente tiene acceso a una clave API, **podemos proteger campos específicos incluso de los administradores de la base de datos**



Replicación S3

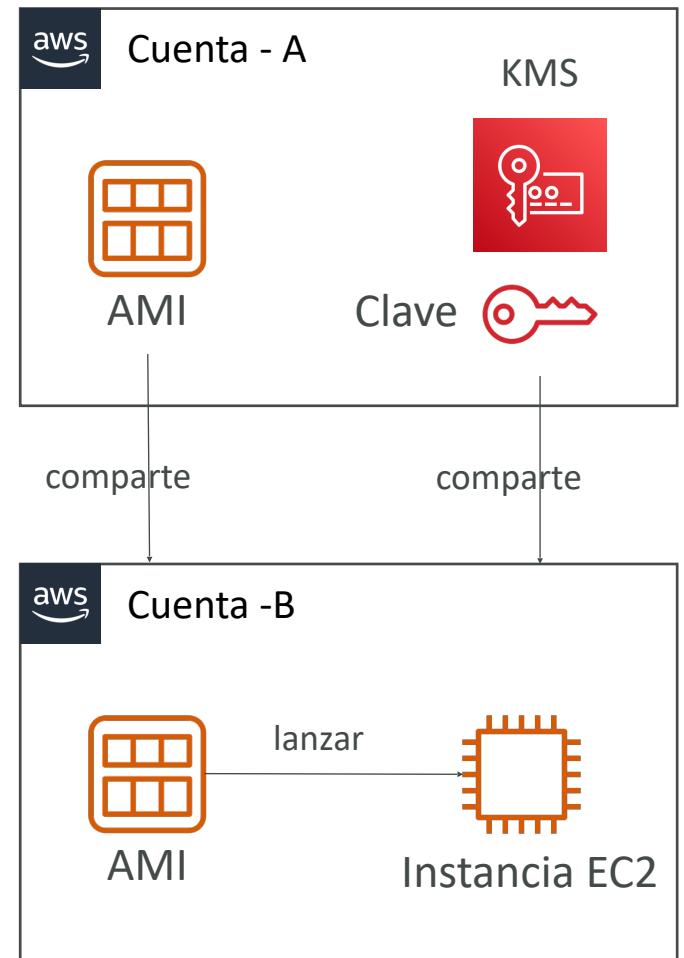
Consideraciones sobre el cifrado



- **Los objetos no cifrados y los cifrados con SSE-S3 se replican por defecto**
- Los objetos cifrados con SSE-C (clave proporcionada por el cliente) nunca se replican
- **Para los objetos cifrados con SSE-KMS,** debes activar la opción
 - Especificar qué clave KMS debe cifrar los objetos dentro del bucket de destino
 - Adaptar la Política de Claves KMS para la clave de destino
 - Un rol IAM con kms:Decrypt para la clave KMS de origen y kms:Encrypt para la clave KMS de destino
 - Es posible que obtengas errores de KMS, en cuyo caso puedes solicitar un aumento de las cuotas de servicio
- **Puedes utilizar Claves KMS de AWS multirregión, pero actualmente Amazon S3 las trata como claves independientes (el objeto seguirá siendo descifrado y luego cifrado)**

Proceso de compartición de AMI cifrada mediante KMS

1. La AMI de la cuenta de origen está cifrada con la clave KMS de la cuenta de origen
2. Debe modificar el atributo de la imagen para añadir un **Permiso de Lanzamiento** que corresponda a la cuenta AWS de destino especificada
3. Debes compartir las Claves KMS utilizadas para cifrar la Snapshot a la que hace referencia la AMI con la cuenta / Rol IAM de destino
4. El rol/usuario IAM de la cuenta de destino debe tener permisos para DescribeKey, ReEncrypted, CreateGrant, Decrypt
5. Al lanzar una instancia EC2 desde la AMI, opcionalmente la cuenta de destino puede especificar una nueva clave KMS en su propia cuenta para volver a cifrar los volúmenes



SSM Parameter Store

Almacén de parámetros SSM

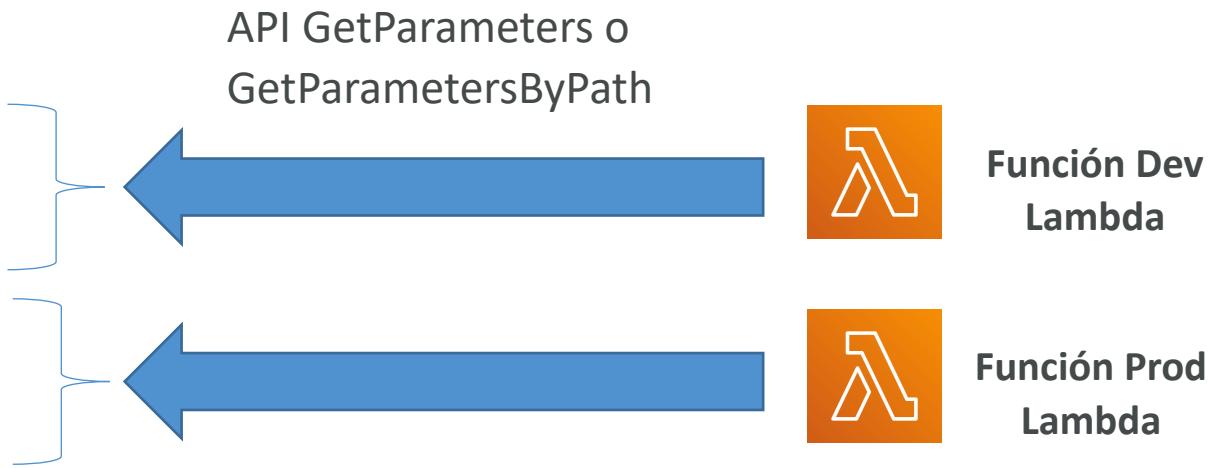


- Almacenamiento seguro de la configuración y los secretos
- Cifrado sin fisuras opcional mediante KMS
- SDK sin servidor, escalable, duradero y sencillo
- Seguimiento de versiones de configuraciones / secretos
- Seguridad mediante IAM
- Notificaciones con Amazon EventBridge
- Integración con CloudFormation



Jerarquía del almacén de parámetros SSM

- /mi-departamento/
 - mi-app/
 - dev/
 - db-url
 - db-contraseña
 - prod/
 - db-url
 - db-contraseña
 - otra-app/
 - /otro-departamento/
 - /aws/reference/secretsmanager/secret_ID_in_Secrets_Manager
 - /aws/service/ami-amazon-linux-latest/amzn2-ami-hvm-x86_64-gp2 (público)



Niveles de parámetros estándar y avanzado

	Estándar	Avanzado
Número total de parámetros permitidos (por cuenta AWS y Región)	10,000	100,000
Tamaño máximo del valor de un parámetro	4 KB	8 KB
Políticas de parámetros disponibles	No	Si
Coste	Sin coste adicional	Se aplican cargos
Precios de almacenamiento	Gratis	0,05 \$ por parámetro avanzado al mes

Políticas de parámetros (para parámetros avanzados)

- Permite asignar un TTL a un parámetro (fecha de caducidad) para forzar la actualización o eliminación de datos sensibles como contraseñas
- Puede asignar varias políticas a la vez

Expiration (para eliminar un parámetro)

```
{  
  "Type": "Expiration",  
  "Version": "1.0",  
  "Attributes": {  
    "Timestamp": "2020-12-02T21:34:33.000Z"  
  }  
}
```

ExpirationNotification (EventBridge)

```
{  
  "Type": "ExpirationNotification",  
  "Version": "1.0",  
  "Attributes": {  
    "Before": "15",  
    "Unit": "Days"  
  }  
}
```

NoChangeNotification (EventBridge)

```
{  
  "Type": "NoChangeNotification",  
  "Version": "1.0",  
  "Attributes": {  
    "After": "20",  
    "Unit": "Days"  
  }  
}
```

AWS Secrets Manager



- Servicio más nuevo, pensado para almacenar secretos
- Capacidad para forzar la **rotación de secretos** cada X días
- Automatizar la generación de secretos en la rotación (utiliza Lambda)
- Integración con **Amazon RDS** (MySQL, PostgreSQL, Aurora)
- Los secretos se cifran mediante KMS
- Pensado principalmente para la integración con RDS

AWS Secrets Manager - Secretos multirregión

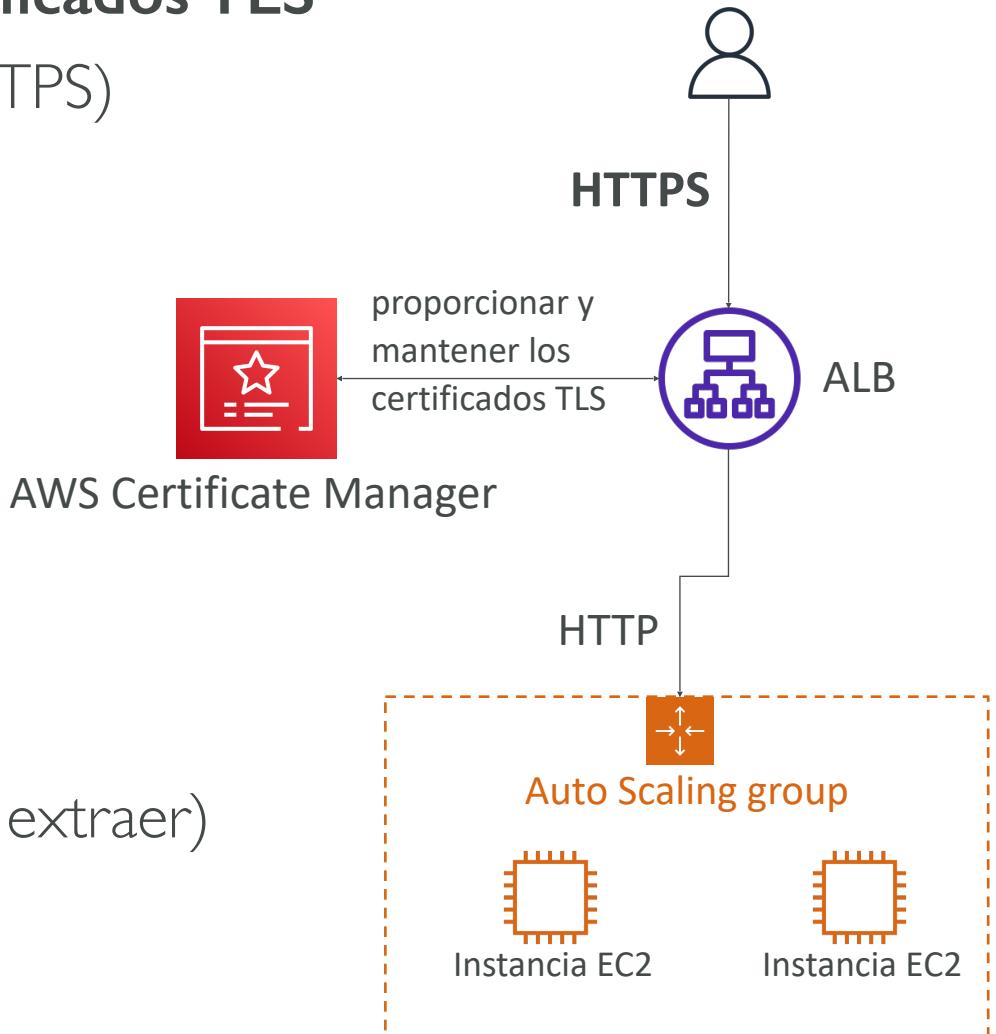
- Replica Secretos en varias regiones de AWS
- AWS Secrets Manager mantiene las réplicas de lectura sincronizadas con el Secreto principal
- Posibilidad de promover un Secreto de réplica de lectura a Secreto independiente
- Casos de uso: aplicaciones multirregión, estrategias de recuperación de desastres, bases de datos multirregión...



AWS Certificate Manager (ACM)



- Aprovisiona, gestiona y despliega fácilmente **Certificados TLS**
- Proporciona cifrado en vuelo para sitios web (HTTPS)
- Soporta certificados TLS públicos y privados
- Gratuito para certificados TLS públicos
- Renovación automática de certificados TLS
- Integraciones con (carga certificados TLS en)
 - Elastic Load Balancers (CLB, ALB, NLB)
 - Distribuciones CloudFront
 - APIs en API Gateway
- No se puede utilizar ACM con EC2 (no se puede extraer)



ACM - Petición de certificados públicos

1. **Lista de nombres de dominio que se incluirán en el certificado**

- Nombre de dominio completo (FQDN): corp.ejemplo.com
- Dominio comodín: *.ejemplo.com

2. **Selecciona el método de validación: Validación DNS o por correo electrónico**

- La validación por DNS es preferible por motivos de automatización
- La validación por correo electrónico enviará correos electrónicos a las direcciones de contacto
- La validación por DNS utilizará un registro CNAME en la configuración DNS (por ejemplo, Route 53)

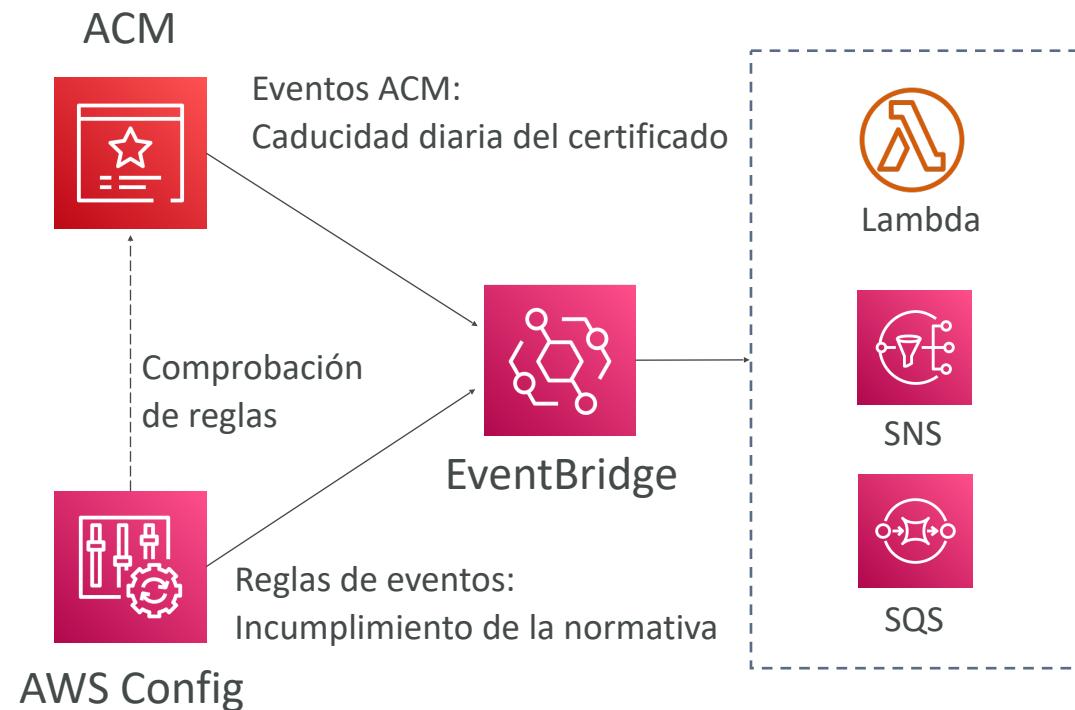
3. **Tardará unas horas en verificarse**

4. **El Certificado Público se inscribirá para su renovación automática**

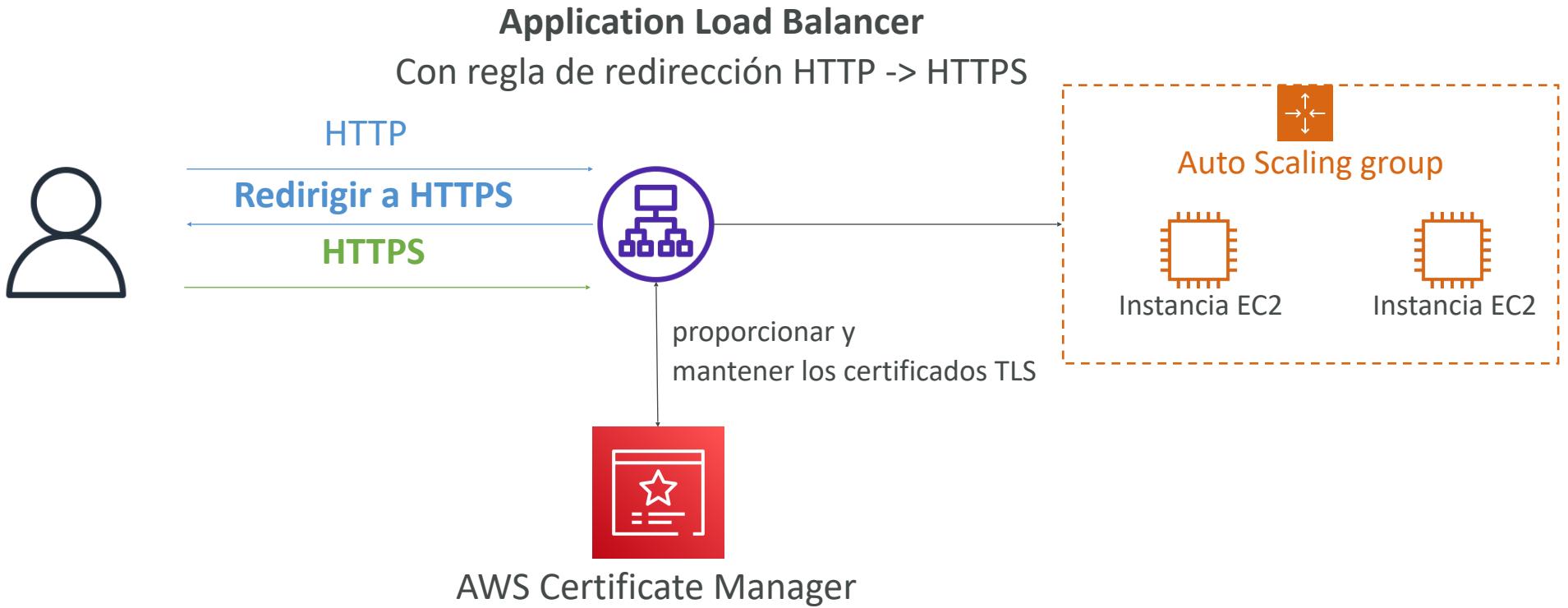
- ACM renueva automáticamente los certificados generados por ACM 60 días antes de su caducidad

ACM - Importar certificados públicos

- Opción de generar el certificado fuera de ACM y luego importarlo
- **No hay renovación automática**, debes importar un nuevo certificado antes de que caduque
- **ACM envía eventos de caducidad diarios** a partir de 45 días antes de la caducidad
 - Se puede configurar el número de días
 - Los eventos aparecen en EventBridge
- **AWS Config** tiene una regla gestionada llamada *acm-certificate-expiration-check* para comprobar si los certificados caducan (número de días configurable)



ACM - Integración con ALB

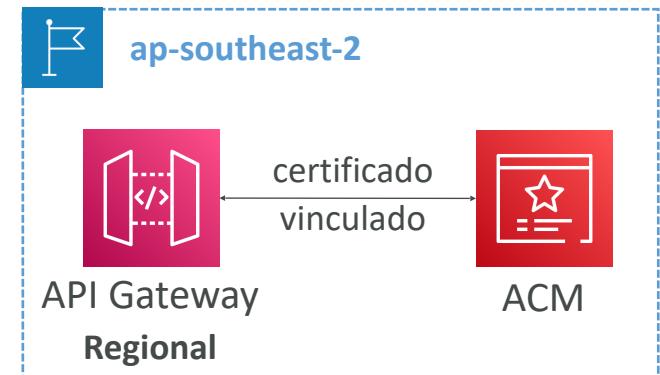
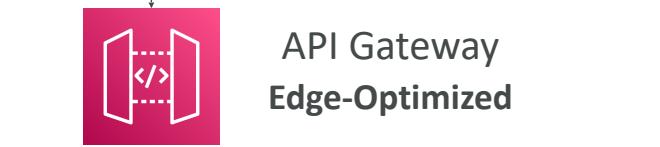
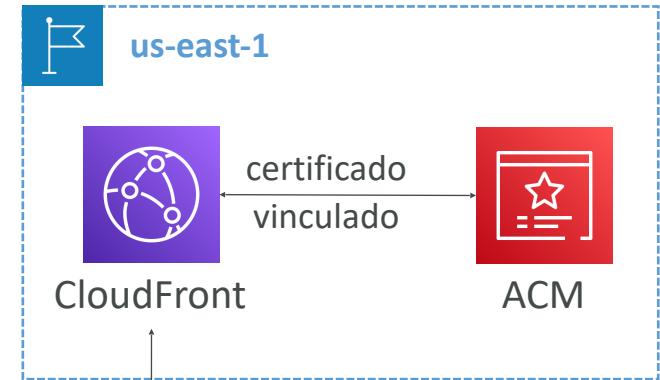


API Gateway - Tipos de endpoint

- **Optimizado para bordes (por defecto):** Para clientes globales
 - Las peticiones se enrutan a través de las Edge Locations de CloudFront (mejora la latencia)
 - La API Gateway sigue viviendo en una sola región
- **Regional:**
 - Para clientes dentro de la misma región
 - Podría combinarse manualmente con CloudFront (más control sobre las estrategias de almacenamiento en caché y la distribución)
- **Privada:**
 - Sólo se puede acceder desde tu VPC utilizando un endpoint VPC de interfaz (ENI)
 - Utiliza una política de recursos para definir el acceso

ACM - Integración con API Gateway

- Crear un **nombre de dominio personalizado** en API Gateway
- **Optimizado para el borde (por defecto):** Para clientes globales
 - Las peticiones se enrutan a través de las Edge Locations de CloudFront (mejora la latencia)
 - La API Gateway sigue viviendo en una sola región
 - **El Certificado TLS debe estar en la misma región que CloudFront, en us-east-1**
 - A continuación, configura el registro CNAME o (mejor) A-Alias en Route 53
- Regional:
 - Para clientes dentro de la misma región
 - **El Certificado TLS debe importarse en API Gateway, en la misma región que API Gateway**
 - A continuación, establece un registro CNAME o (mejor) A-Alias en Route 53



AWS WAF - Firewall de aplicaciones web



- Protege tus aplicaciones web de los exploits web habituales (Capa 7)
- **La Capa 7 es HTTP** (frente a la Capa 4 que es TCP/UDP)
- Despliega en
 - **Application Load Balancer**
 - **Gateway API**
 - **CloudFront**
 - **API GraphQL de AppSync**
 - **Grupo de usuarios Cognito (Cognito User Pool)**

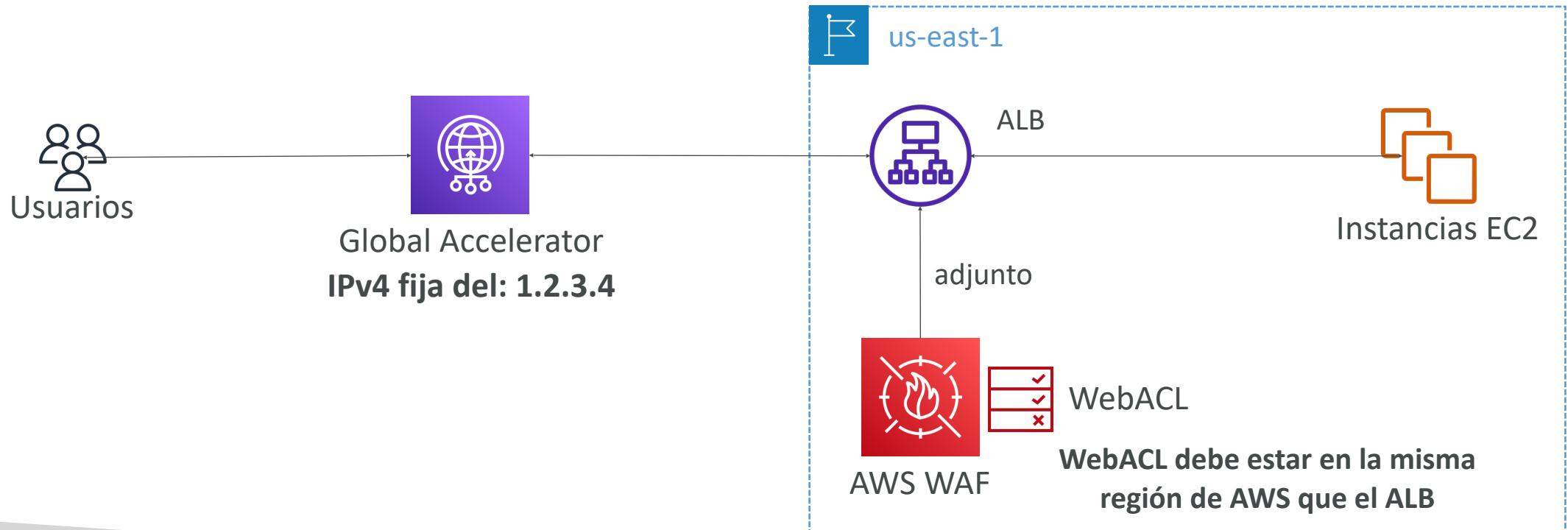
AWS WAF - Firewall de aplicaciones web



- Define Reglas ACL (Lista de Control de Acceso Web):
 - **Conjunto de IP hasta 10.000 direcciones IP** - utiliza varias Reglas para más IPs
 - Cabeceras HTTP, cuerpo HTTP o Strings URI Protege de ataques comunes:
inyección SQL y **Cross-Site Scripting (XSS)**
 - Restricciones de tamaño, **geo-match (bloquear países)**
 - **Reglas basadas en la tasa** (para contar las ocurrencias de eventos) - **para protección DDoS**
- Las ACL web son regionales, excepto CloudFront
- Un grupo de reglas es **un conjunto reutilizable de reglas que puedes añadir a una ACL web**

WAF - IP fija al utilizar WAF con un Load Balancer

- WAF no soporta el Network Load Balancer (Capa 4)
- Podemos utilizar Global Accelerator para IP fija y WAF en el ALB





AWS Shield: protección contra ataques DDoS

- **DDoS:** Denegación de Servicio Distribuida - muchas peticiones al mismo tiempo
- **AWS Shield Estándar:**
 - Servicio gratuito que se activa para todos los clientes de AWS
 - Proporciona protección contra ataques como SYN/UDP Floods, ataques de Reflexión y otros ataques de capa 3/capa 4
- **AWS Shield Avanzado:**
 - Servicio opcional de mitigación de DDoS (3.000 \$ al mes por organización)
 - Protege contra ataques más sofisticados en [Amazon EC2](#), [Elastic Load Balancing \(ELB\)](#), [Amazon CloudFront](#), [AWS Global Accelerator](#) y [Route 53](#)
 - Acceso 24/7 al equipo de respuesta DDoS de AWS (DRP)
 - Protégete contra las tarifas más altas durante los picos de uso debidos a DDoS
 - Shield Avanzado crea una mitigación automática avanzada de DDoS en la capa de aplicación, evalúa y despliega automáticamente reglas de AWS WAF para mitigar los ataques de la capa 7

AWS Firewall Manager



- **Gestionar reglas en todas las cuentas de una AWS Organizations**
- Política de seguridad: conjunto común de reglas de seguridad
 - Reglas WAF (Application Load Balancer, API Gateways, CloudFront)
 - AWS Shield Avanzado (ALB, CLB, NLB, Elastic IP, CloudFront)
 - Grupos de seguridad para EC2, Application Load Balancer y recursos ENI en VPC
 - AWS Network Firewall (Nivel VPC)
 - Resolver Firewall DNS de Amazon Route 53
- Las políticas se crean a nivel de región
- **Las reglas se aplican a los nuevos recursos a medida que se crean (bueno para la normativa) en todas y futuras cuentas de tu organización**

WAF vs. Firewall Manager vs. Shield



AWS WAF



AWS Firewall Manager



AWS Shield

- **WAF, Shield y Firewall Manager se utilizan juntos para una protección integral**
- Define tus reglas ACL Web en WAF
- Para una protección granular de tus recursos, WAF solo es la opción correcta
- Si quieres utilizar AWS WAF en todas las cuentas, acelerar la configuración de WAF, automatizar la protección de nuevos recursos, utiliza Firewall Manager con AWS WAF
- Shield Advanced añade funciones adicionales a AWS WAF, como el soporte dedicado del Shield Response Team (SRT) y la elaboración de informes avanzados.
- Si eres propenso a frecuentes ataques DDoS, considera la compra de Shield Advanced

Mejores prácticas de AWS para la resiliencia

Mitigación DDoS de Edge Locations (BP1, BP3)

- **BP1 - CloudFront**

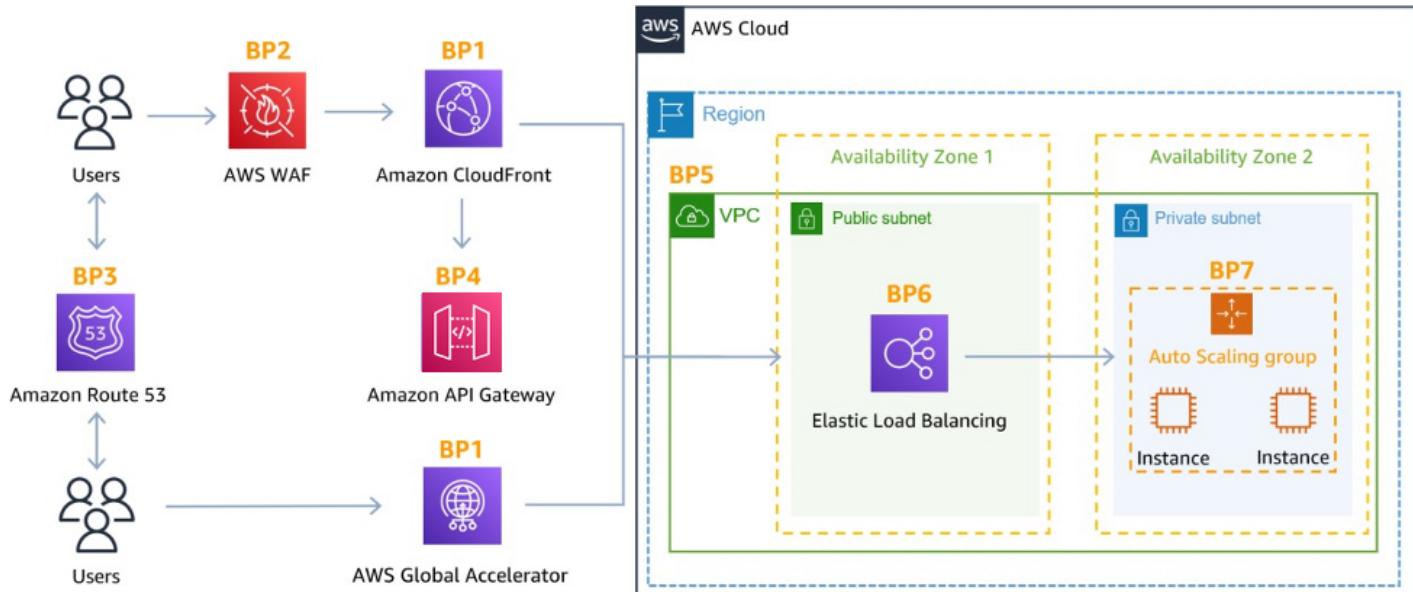
- Entrega de aplicaciones web en el borde
- Protege de los ataques DDoS comunes (inundaciones SYN...)

- **BP1 - Global Accelerator**

- Accede a tu aplicación desde el borde
- Integración con Shield para protección DDoS
- Útil si tu backend no es compatible con CloudFront

- **BP3 - Route 53**

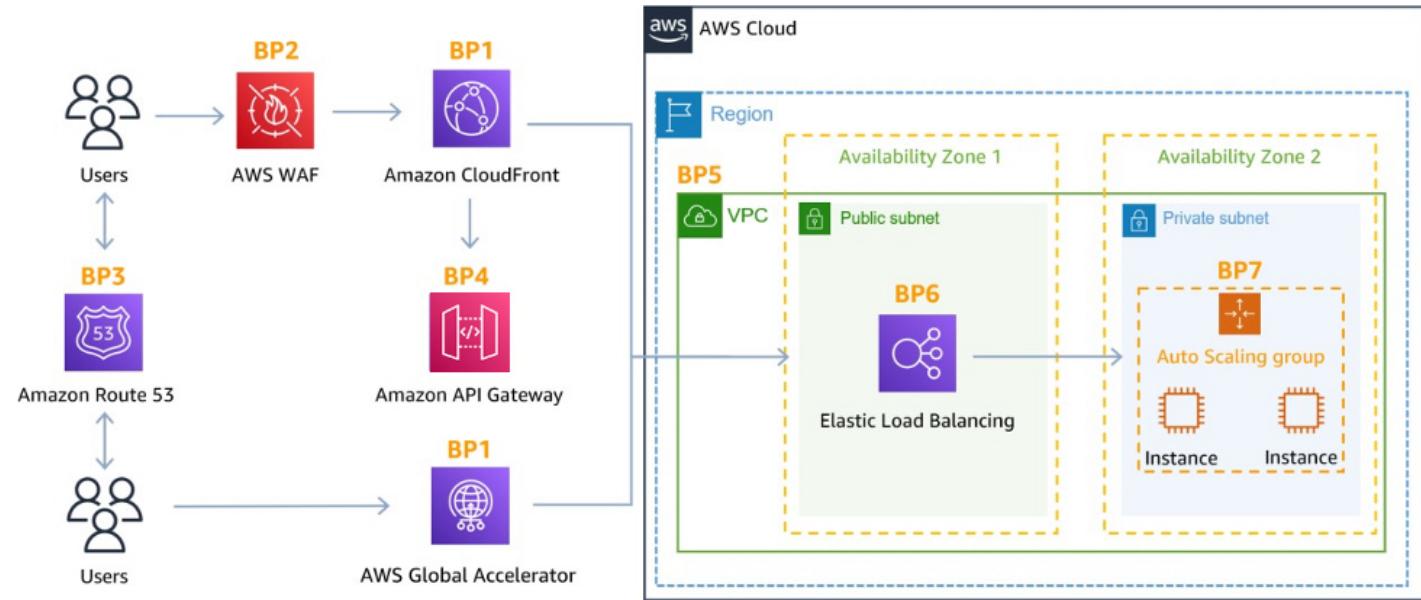
- Resolución de nombres de dominio en el borde
- Mecanismo de protección DDoS



Mejores prácticas de AWS para la resiliencia DDoS

Mejores prácticas para la mitigación de DDoS

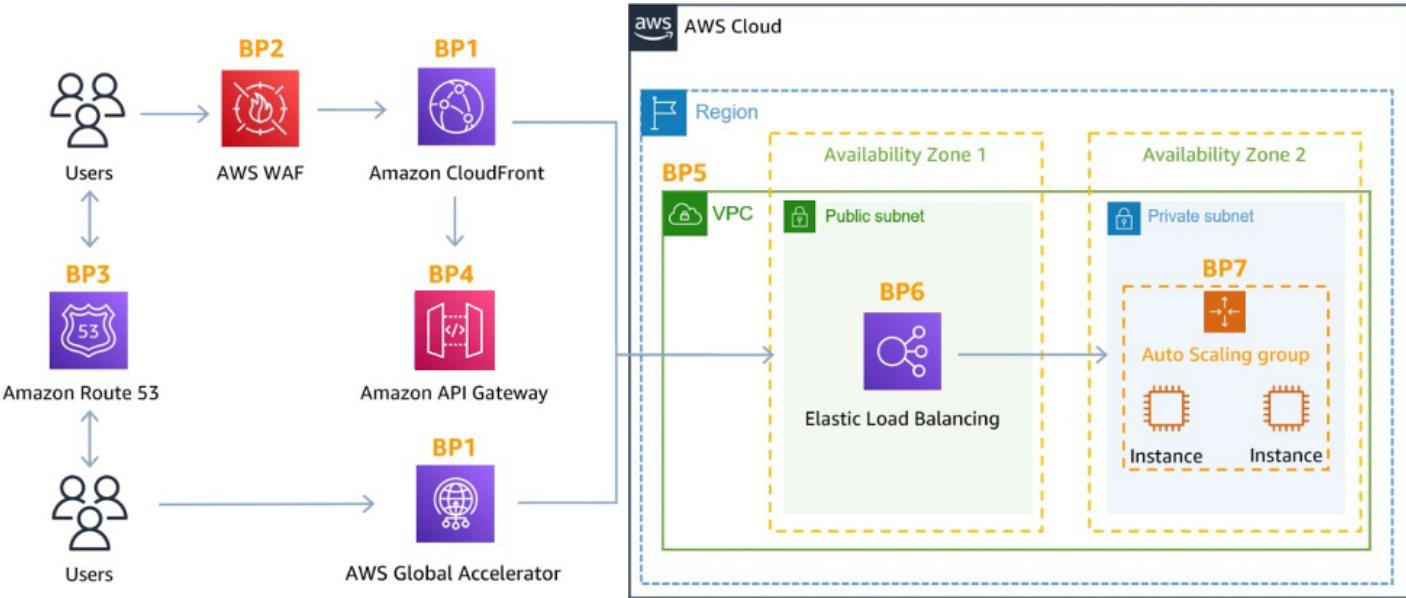
- **Defensa de la capa de infraestructura (BP1, BP3, BP6)**
 - Protege Amazon EC2 contra el tráfico elevado
 - Esto incluye el uso de Global Accelerator, Route 53, CloudFront, Elastic Load Balancing
- **Amazon EC2 con Autoescalado (BP7)**
 - Ayuda a escalar en caso de aumentos repentinos de tráfico, incluyendo una multitud repentina o un ataque DDoS
- **Elastic Load Balancing (BP6)**
 - Elastic Load Balancing escala con los aumentos de tráfico y distribuirá el tráfico a muchas instancias EC2



Mejores prácticas de AWS para la resiliencia

Defensa de la capa de aplicación

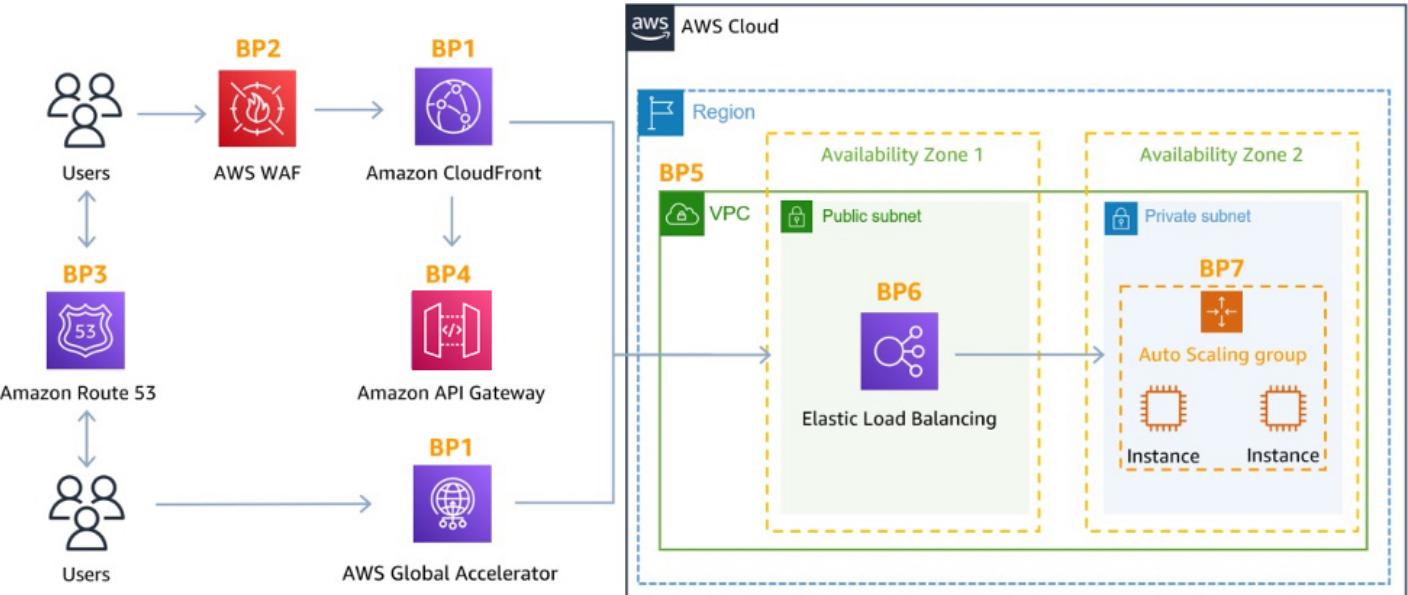
- **Detecta y filtra peticiones web maliciosas (BP1, BP2)**
 - CloudFront almacena en caché el contenido estático y lo sirve desde Edge Locations, protegiendo tu backend
 - AWS WAF se utiliza sobre CloudFront y Application Load Balancer para filtrar y bloquear peticiones basadas en firmas de peticiones
 - Las reglas basadas en la tasa de WAF pueden bloquear automáticamente las IP de los malos actores
 - Utiliza reglas gestionadas en WAF para bloquear ataques basados en la reputación de la IP, o bloquear IPs anónimas
 - CloudFront puede bloquear geografías específicas
- **Shield Avanzado (BP1, BP2, BP6)**
 - La mitigación automática de DDoS en la capa de aplicación de Shield Advanced crea, evalúa y despliega automáticamente reglas WAF de AWS para mitigar los ataques de la capa 7



Mejores prácticas de AWS para la resiliencia

Reducción de la superficie de ataque

- **Ofuscar los recursos de AWS (BP1, BP4, BP6)**
 - Uso de CloudFront, API Gateway, Elastic Load Balancing para ocultar tus recursos backend (funciones Lambda, instancias EC2)
- **Grupos de Seguridad y ACLs de Red (BP5)**
 - Utiliza grupos de seguridad y NACLs para filtrar el tráfico basado en IP específicas a nivel de subred o ENI
 - Las IP elásticas están protegidas por AWS Shield Advanced
- **Proteger los endpoints API (BP4)**
 - Ocultar EC2, Lambda, en otro lugar
 - Modo optimizado para Edge, o CloudFront + modo regional (más control para DDoS)
 - WAF + API Gateway: límites de ráfagas, filtrado de cabeceras, uso de claves API

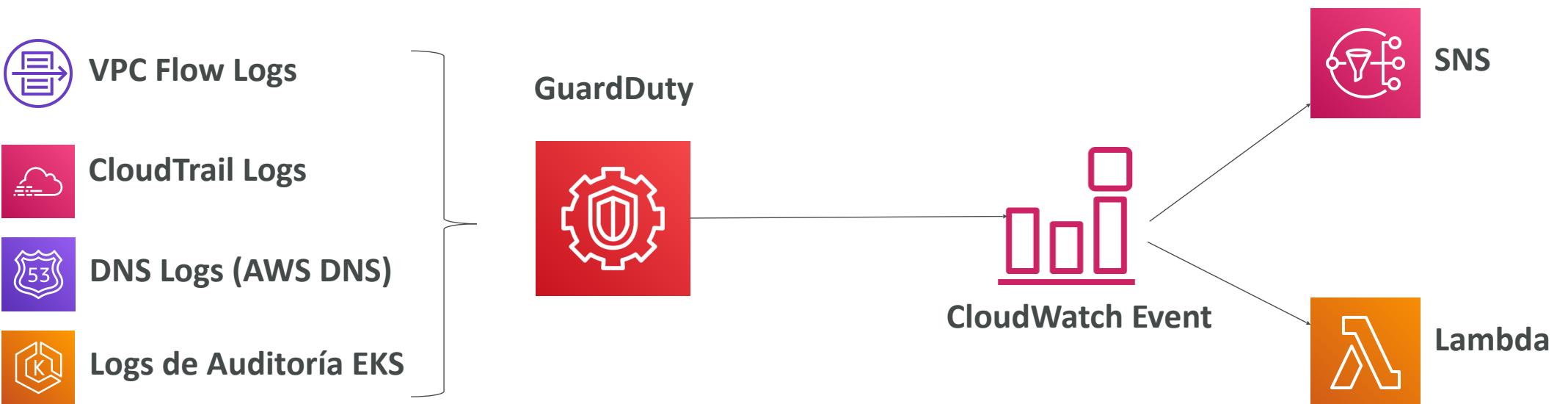




Amazon GuardDuty

- Descubrimiento inteligente de amenazas para proteger la cuenta de AWS
- Utiliza algoritmos de Machine Learning, detección de anomalías, datos de terceros
- Se activa con un clic (30 días de prueba), sin necesidad de instalar software
- Los datos de entrada incluyen:
 - **Logs de Eventos de CloudTrail** - llamadas inusuales a la API, despliegues no autorizados
 - **Eventos de gestión de CloudTrail** - crear subred VPC, crear rastro, ...
 - **Eventos de Datos S3 de CloudTrail** - obtener objeto, listar objetos, borrar objeto, ...
 - **VPC Flow Logs** - tráfico interno inusual, dirección IP inusual
 - **DNS Logs** - instancias EC2 comprometidas que envían datos codificados en consultas DNS
 - **Kubernetes Audit logs** - actividades sospechosas y posibles compromisos de Cluster EKS
- Puedes configurar **CloudWatch Events** para recibir notificaciones en caso de hallazgos
- Las reglas de CloudWatch Events pueden dirigirse a AWS Lambda o SNS
- **Puede proteger contra ataques a criptomonedas (tiene un "hallazgo" dedicado a ello)**

Amazon GuardDuty



Amazon Inspector

- Evaluaciones de seguridad automatizadas
- Para instancias EC2
 - Aprovechando el agente **AWS System Manager (SSM)**
 - Analiza contra **accesibilidad no intencionada a la red**
 - Analiza el **SO en ejecución** frente a **vulnerabilidades conocidas**
- Para contenedores enviados a Amazon ECR
 - Evaluación de contenedores a medida que se envían
- Para Funciones Lambda
 - Identifica vulnerabilidades de software en funciones de código y dependencias
 - Evaluaciones de como son desplegadas las funciones
- Informes e integración con AWS Security Hub
- Envía los resultados a Amazon Event Bridge



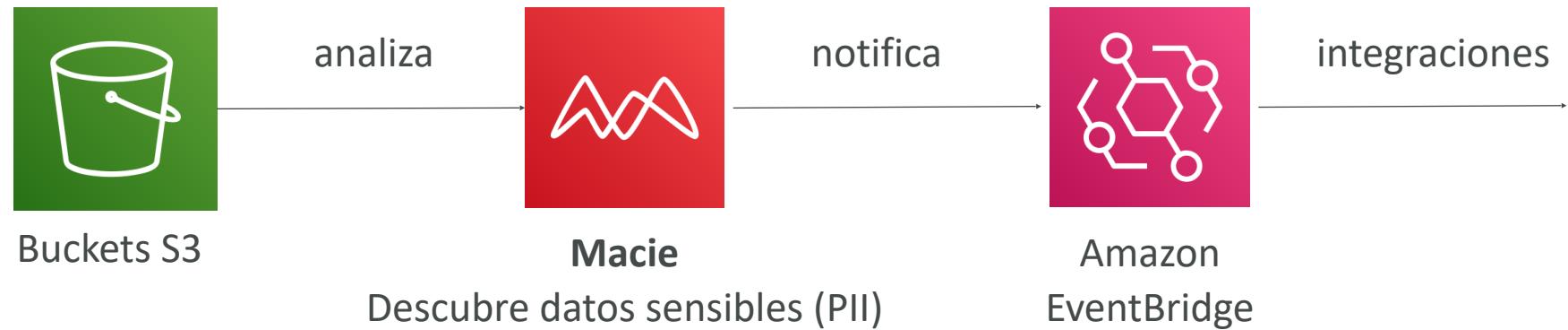
¿Qué evalúa Amazon Inspector?

- **Recuerda: sólo para instancias EC2 e infraestructura de contenedores**
- Escaneo continuo de la infraestructura, sólo cuando sea necesario
- Se asocia una puntuación de riesgo a todas las vulnerabilidades para priorizarlas

Amazon Macie

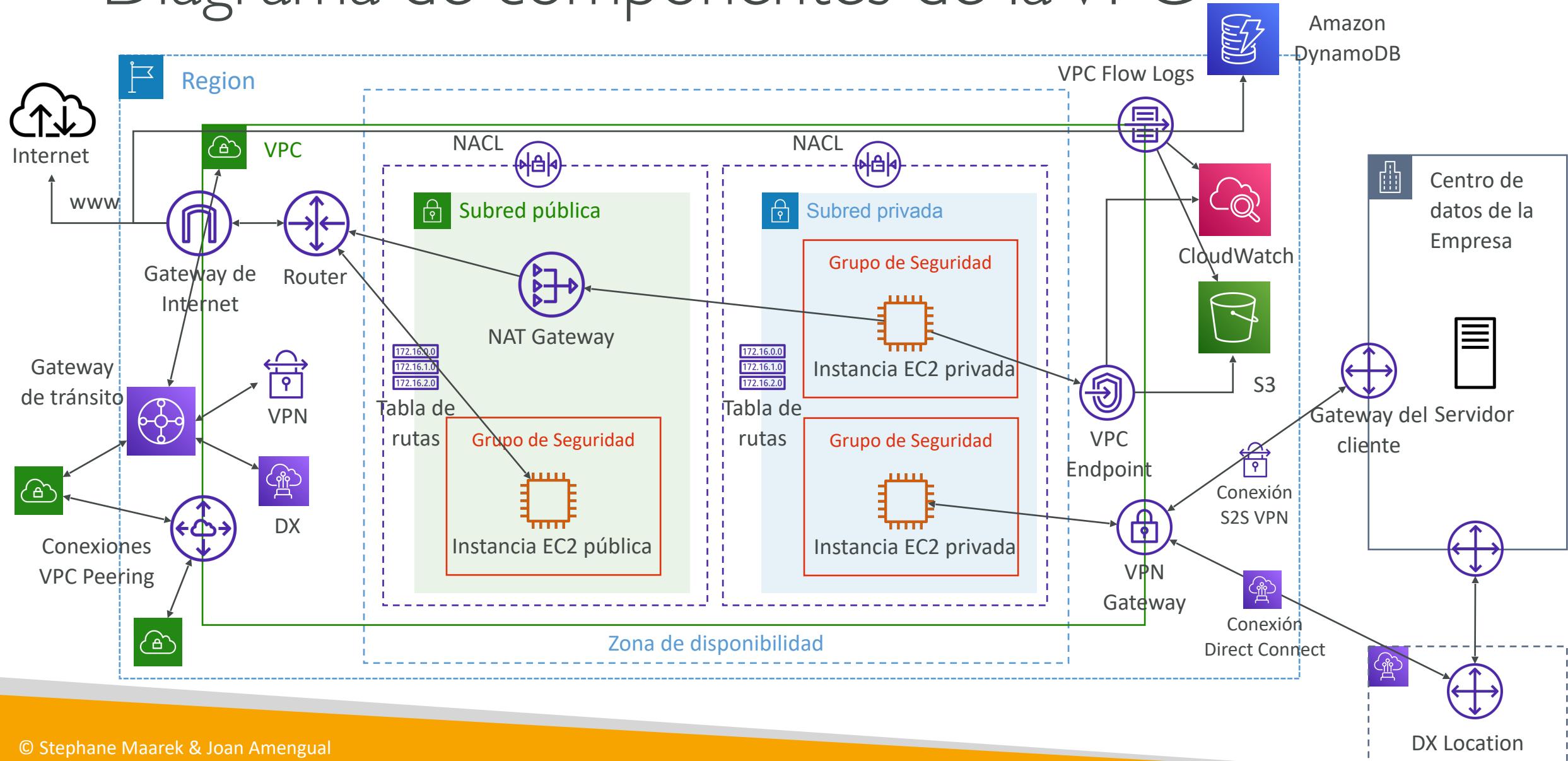


- Amazon Macie es un servicio de seguridad y privacidad de datos totalmente gestionado que utiliza el **Machine Learning y la concordancia de patrones para descubrir y proteger tus datos sensibles en AWS**.
- Macie te ayuda a identificar y alertarte sobre **datos sensibles, como la información de identificación personal (PII)**



Virtual Private Cloud (VPC)

Diagrama de componentes de la VPC



Entender CIDR - IPv4

- **Enrutamiento entre dominios sin clase** - un método para asignar direcciones IP
- Se utiliza en las **reglas de los grupos de seguridad** y en la red de AWS en general

IP version	Type	Protocol	Port range	Source	Description
IPv4	SSH	TCP	22	122.149.196.85/32	-
IPv4	HTTP	TCP	80	0.0.0.0/0	-

- Ayudan a definir un rango de direcciones IP:
 - Hemos visto WW.XX.YY.ZZ/32 => una IP
 - Hemos visto 0.0.0.0/0 => todas las IPs
 - Pero podemos definir: 192.168.0.0/26 => 192.168.0.0 - 192.168.0.63 (64 direcciones IP)

Entender CIDR - IPv4

- Un CIDR consta de dos componentes
- **IP base**
 - Representa una IP contenida en el rango (XX.XX.XX.XX)
 - Ejemplo: 10.0.0.0, 192.168.0.0, ...
- **Máscara de subred**
 - Define cuántos bits pueden cambiar en la IP
 - Ejemplo: /0, /24, /32
 - Puede adoptar dos formas:
 - /8 \Leftrightarrow 255.0.0.0
 - /16 \Leftrightarrow 255.255.0.0
 - /24 \Leftrightarrow 255.255.255.0
 - /32 \Leftrightarrow 255.255.255.255

Entender CIDR - Máscara de subred

- La máscara de subred básicamente permite que parte de la IP subyacente obtenga valores siguientes adicionales de la IP base

192	·	168	·	0	·	0	/32 => permite 1 IP (2^0)	→ 192.168.0.0
192	·	168	·	0	·	0	/31 => permite 2 IP (2^1)	→ 192.168.0.0 → 192.168.0.1
192	·	168	·	0	·	0	/30 => permite 4 IP (2^2)	→ 192.168.0.0 → 192.168.0.3
192	·	168	·	0	·	0	/29 => permite 8 IP (2^3)	→ 192.168.0.0 → 192.168.0.7
192	·	168	·	0	·	0	/28 => permite 16 IP (2^4)	→ 192.168.0.0 → 192.168.0.15
192	·	168	·	0	·	0	/27 => permite 32 IP (2^5)	→ 192.168.0.0 → 192.168.0.31
192	·	168	·	0	·	0	/26 => permite 64 IP (2^6)	→ 192.168.0.0 → 192.168.0.63
192	·	168	·	0	·	0	/25 => permite 128 IP (2^7)	→ 192.168.0.0 → 192.168.0.127
192	·	168	·	0	·	0	/24 => permite 256 IP (2^8)	→ 192.168.0.0 → 192.168.0.255
...								
192	·	168	·	0	·	0	/16 => permite 65,536 IP (2^{16})	→ 192.168.0.0 → 192.168.255.255
...								
192	·	168	·	0	·	0	/0 => permite todas IPs	→ 0.0.0.0 → 255.255.255.255



Nota rápida

Octetos

1st · 2nd · 3rd · 4th

- /32 – ningún octeto puede cambiar
- /24 – el último octeto puede cambiar
- /16 – los 2 últimos octetos pueden cambiar
- /8 – los 3 últimos octetos pueden cambiar
- /0 – todos los octetos pueden cambiar

Entender CIDR - Pequeño ejercicio

- $192.168.0.0/24 = \dots ?$
 - $192.168.0.0 - 192.168.0.255$ (256 IPs)
- $192.168.0.0/16 = \dots ?$
 - $192.168.0.0 - 192.168.255.255$ (65,536 IPs)
- $134.56.78.123/32 = \dots ?$
 - Solo $134.56.78.123$
- $0.0.0.0/0$
 - Todas las IPs!
- En caso de duda, utiliza este sitio web:
 - <https://www.ipaddressguide.com/cidr>

IP pública frente a IP privada (IPv4)

- La Autoridad de Asignación de Números de Internet (IANA) estableció ciertos bloques de direcciones IPv4 para uso de direcciones privadas (LAN) y públicas (Internet)
- La **IP privada** sólo puede permitir determinados valores:
 - 10.0.0 – 10.255.255.255 (10.0.0/8) ← en grandes redes
 - 172.16.0.0 – 172.31.255.255 (172.16.0/12) ← VPC por defecto de AWS en ese rango
 - 192.168.0.0 – 192.168.255.255 (192.168.0/16) ← por ejemplo, redes domésticas
- El resto de direcciones IP de Internet son Públicas

Visión general de la VPC por defecto

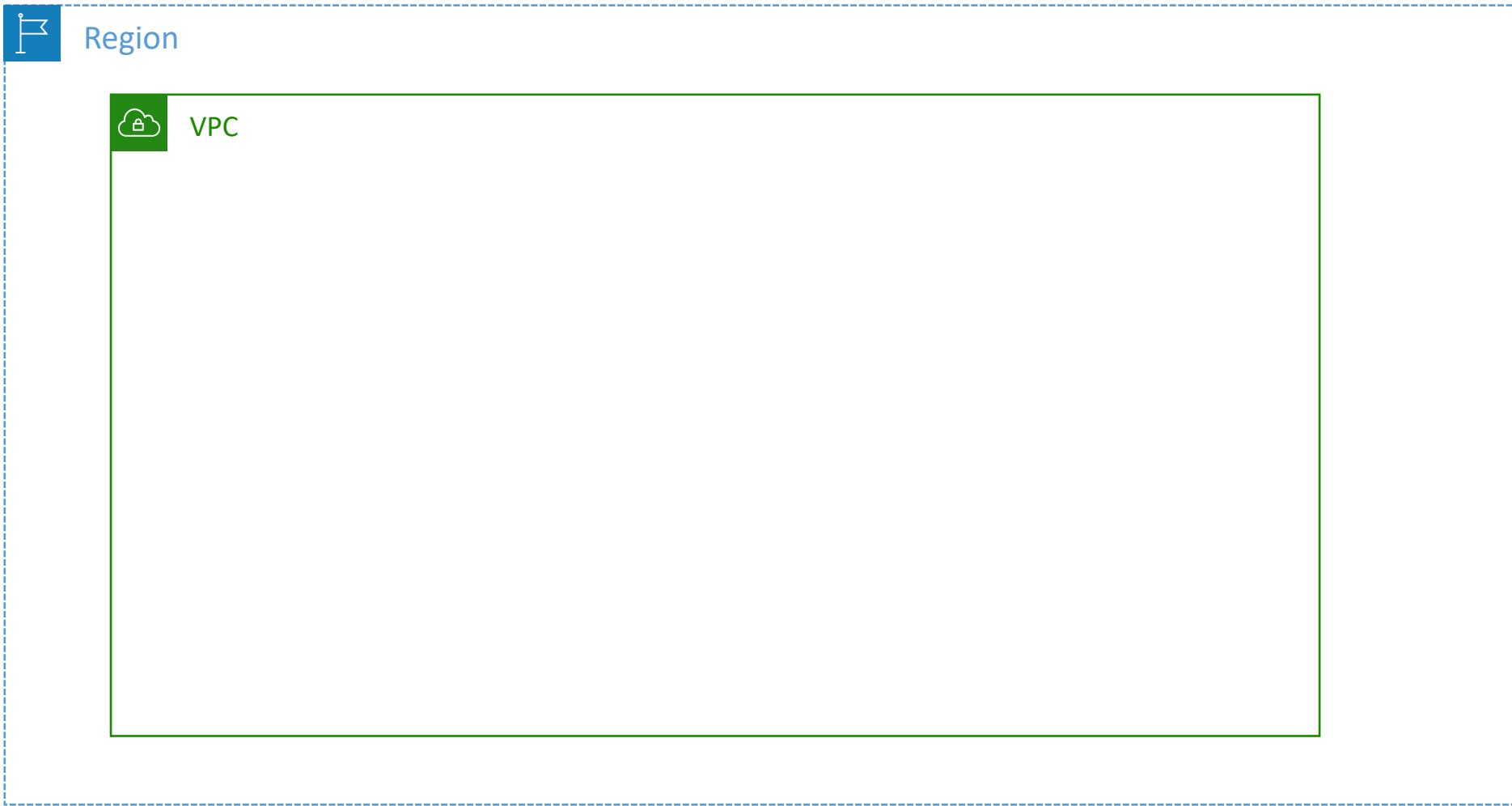
- Todas las cuentas nuevas de AWS tienen una VPC por defecto
- Las nuevas instancias EC2 se lanzan en la VPC por defecto si no se especifica ninguna subred
- La VPC predeterminada tiene conectividad a Internet y todas las instancias EC2 dentro de ella tienen direcciones IPv4 públicas
- También obtenemos un nombre DNS IPv4 público y otro privado

VPC en AWS - IPv4

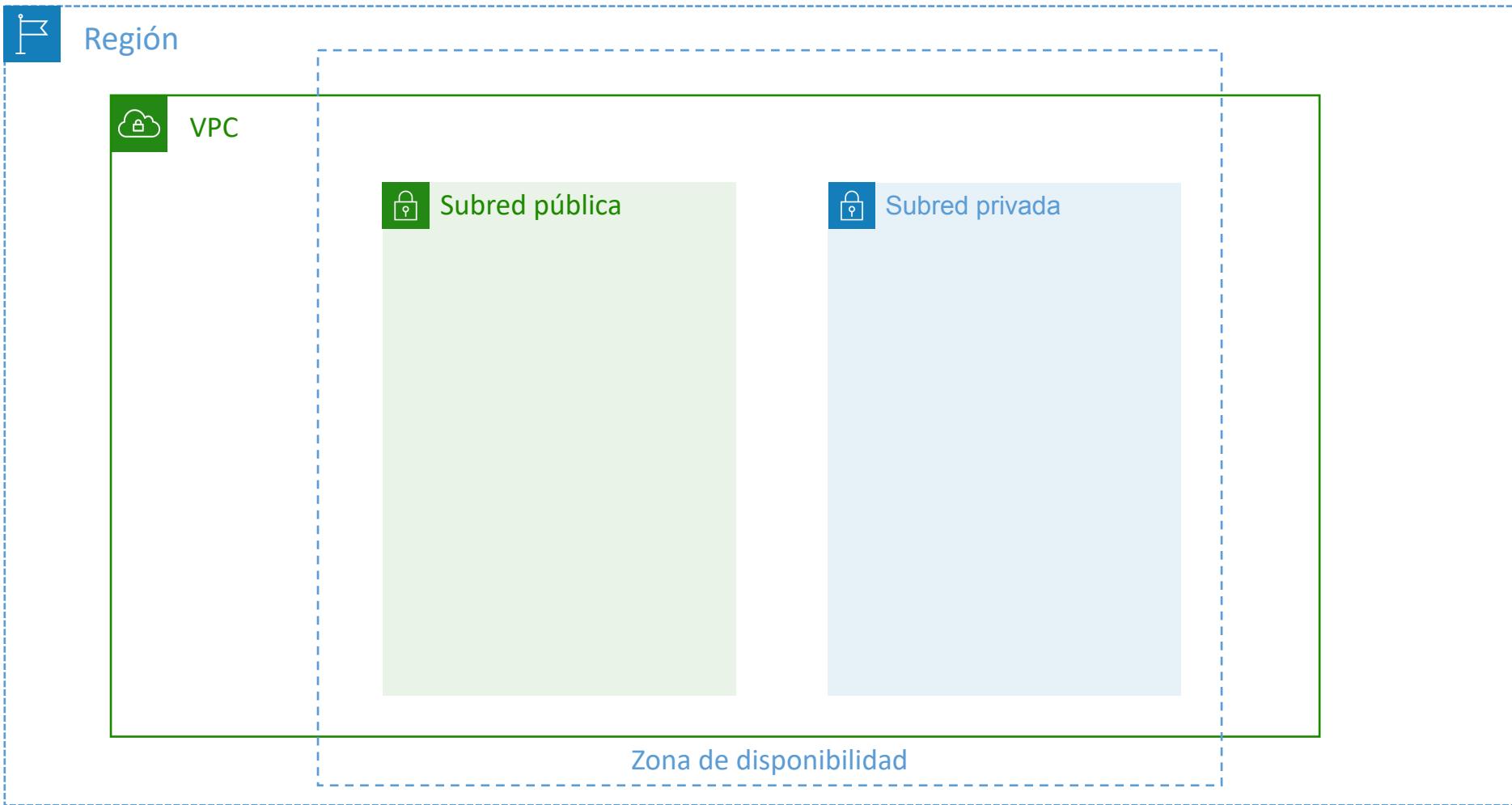


- **VPC = Virtual Private Cloud (nube privada virtual)**
- Puedes tener varias VPC en una región de AWS (máx. 5 por región - límite suave)
- CIDR máx. CIDR por VPC es 5, por cada CIDR:
 - **El tamaño mínimo es /28 (16 direcciones IP)**
 - **El tamaño máximo es /16 (65536 direcciones IP)**
- Como la VPC es privada, sólo se permiten los rangos IPv4 Privados:
 - 10.0.0.0 - 10.255.255.255 (10.0.0.0/8)
 - 172.16.0.0 - 172.31.255.255 (172.16.0.0/12)
 - 192.168.0.0 - 192.168.255.255 (192.168.0.0/16)
- **El CIDR de tu VPC NO debe solaparse con tus otras redes (por ejemplo, la corporativa)**

Estado de la práctica



Añadir subredes

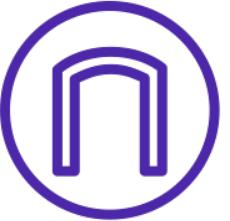


VPC - Subred (IPv4)



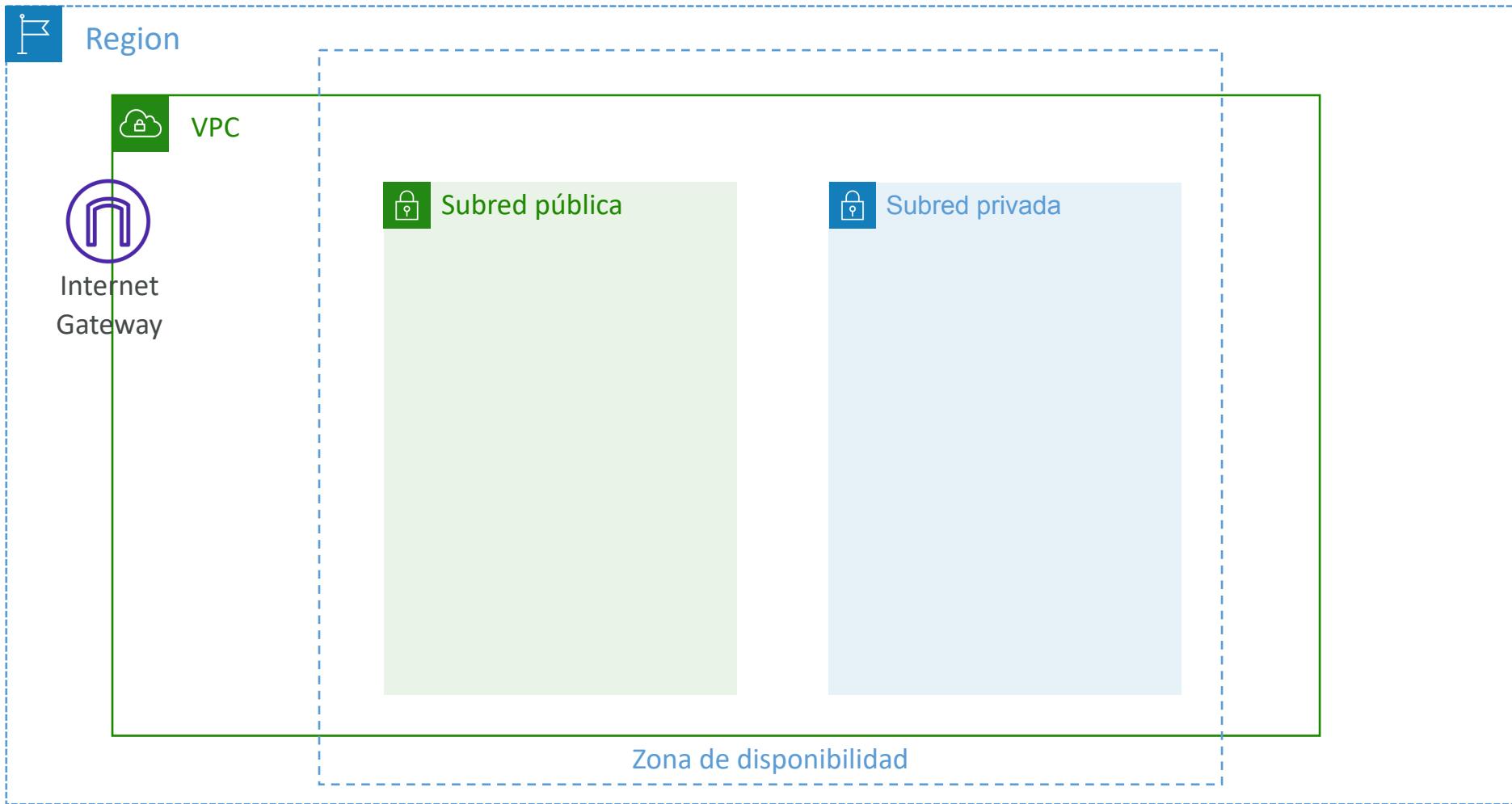
- AWS reserva **5 direcciones IP (4 primeras y 1 última)** en cada subred
- Estas 5 direcciones IP no están disponibles para su uso y no se pueden asignar a una instancia EC2
- Ejemplo: si el bloque CIDR es 10.0.0.0/24, las direcciones IP reservadas son:
 - **10.0.0.0** - Dirección de red
 - **10.0.0.1** - reservada por AWS para el router de la VPC
 - **10.0.0.2** - reservada por AWS para la asignación al DNS proporcionado por Amazon
 - **10.0.0.3** - reservada por AWS para uso futuro
 - **10.0.0.255** - Dirección de difusión de red. AWS no soporta broadcast en una VPC, por lo que la dirección está reservada
- **Consejo de examen**, si necesitas 29 direcciones IP para instancias EC2:
 - No puedes elegir una subred de tamaño /27 (32 direcciones IP, $32 - 5 = 27 < 29$)
 - Tienes que elegir una subred de tamaño /26 (64 direcciones IP, $64 - 5 = 59 > 29$)

Gateway de Internet (IGW)

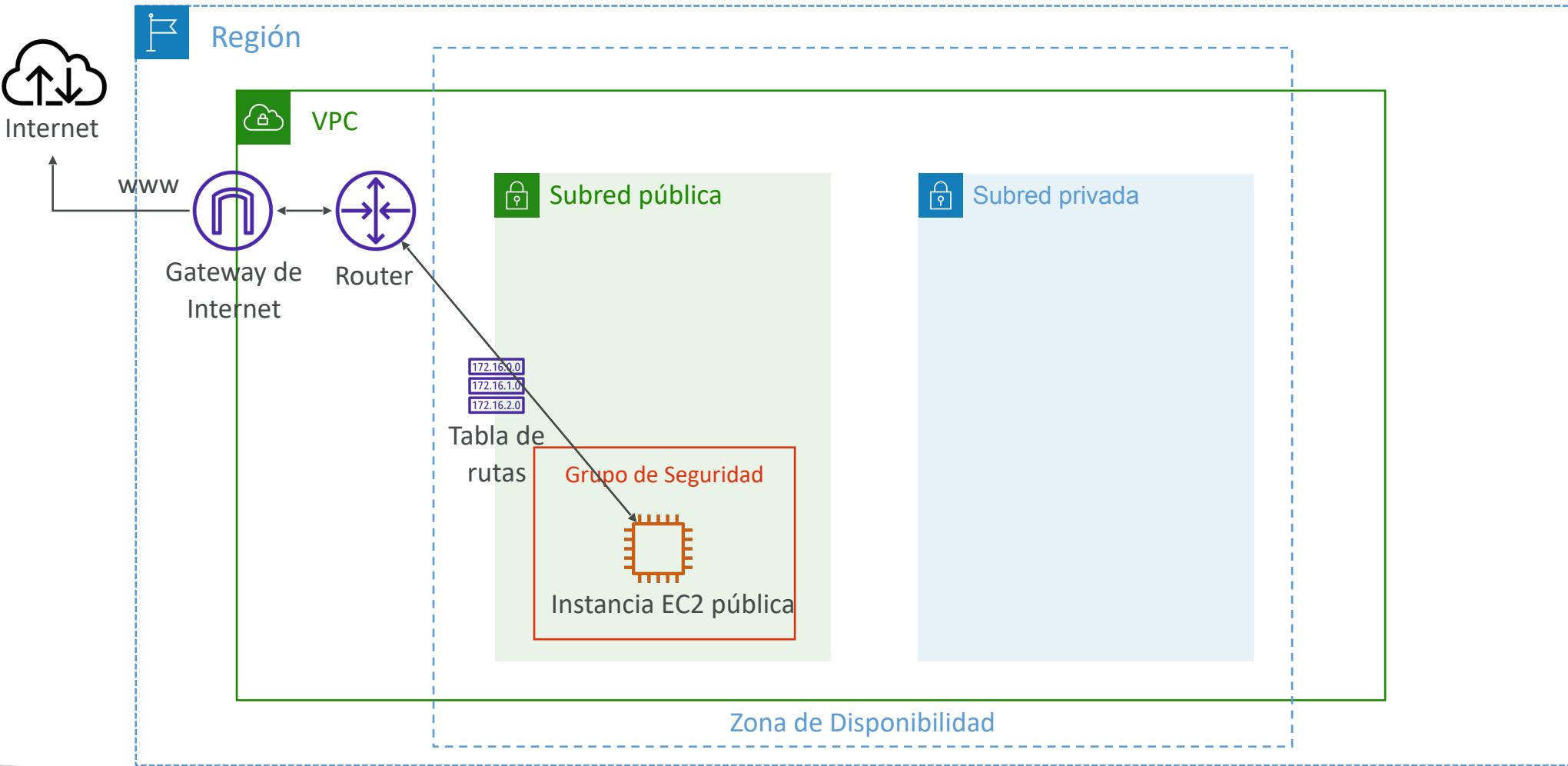


- Permite que los recursos (por ejemplo, instancias EC2) de una VPC se conecten a Internet
 - Se escala horizontalmente y tiene alta disponibilidad y redundancia
 - Debe crearse por separado de una VPC
 - Una VPC sólo puede conectarse a una IGW y viceversa
-
- Las Puertas de enlace (Gateway) de Internet por sí solas no permiten el acceso a Internet...
 - ¡Las tablas de rutas también deben editarse!

Añadir Gateway de Internet

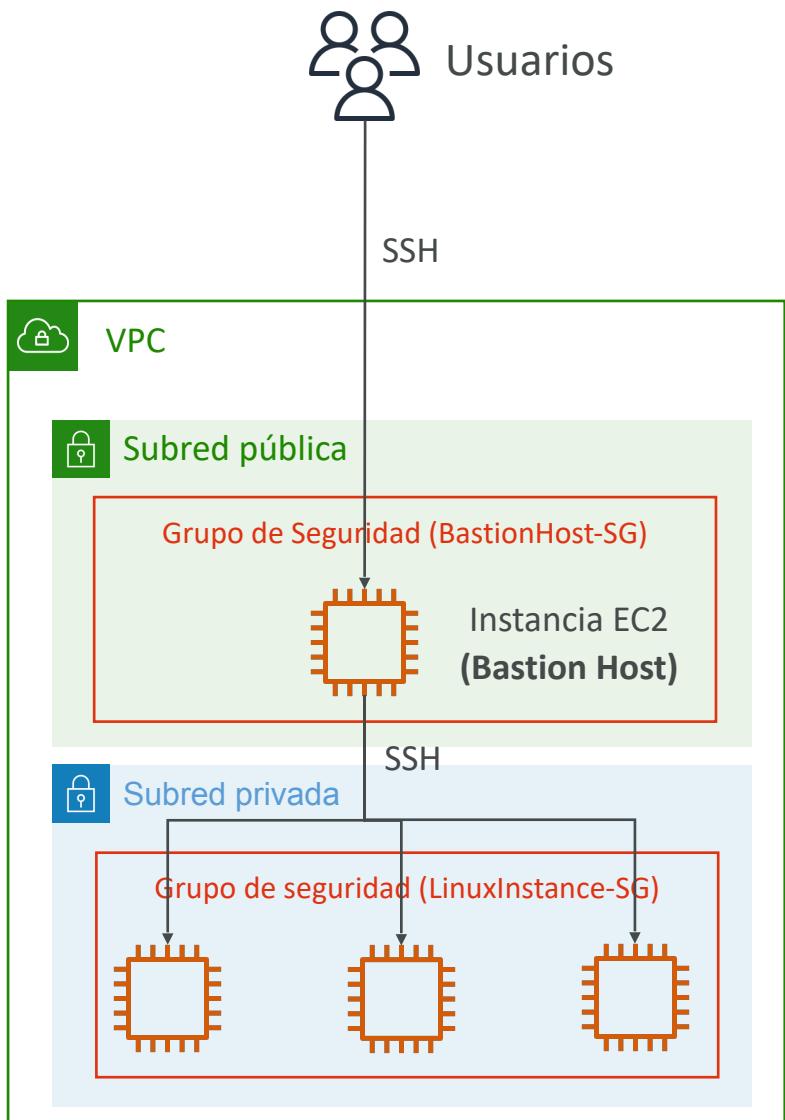


Editar tablas de rutas



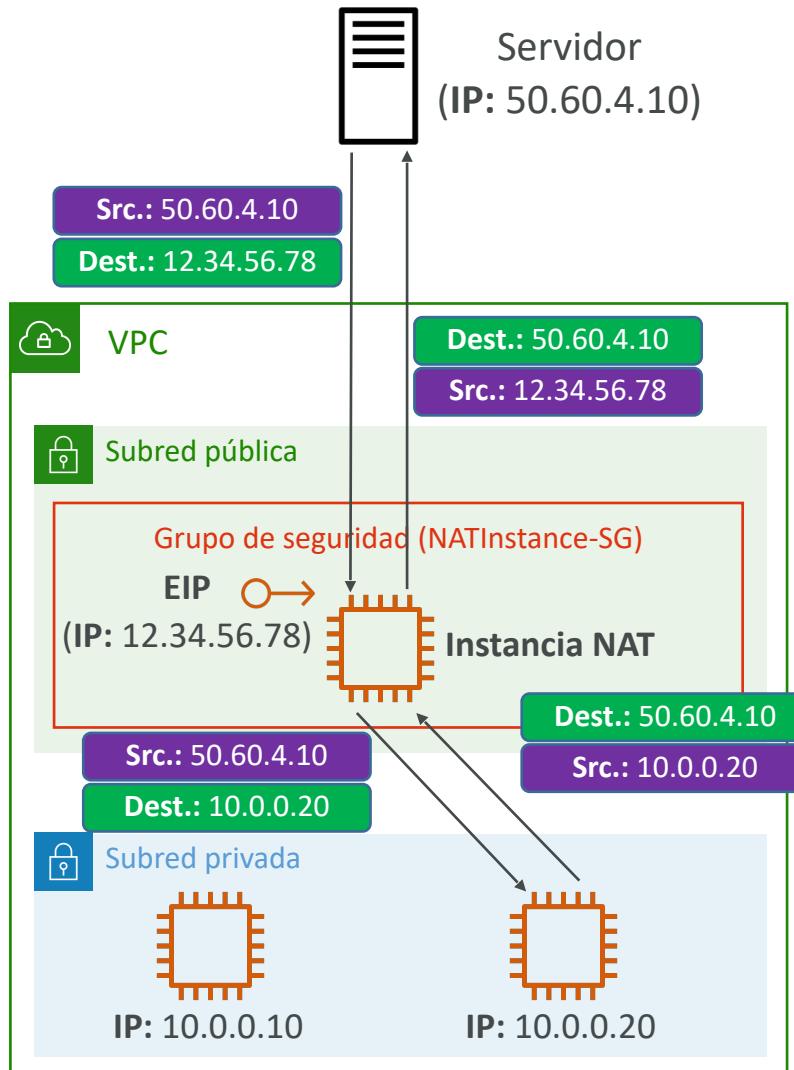
Bastion Host (Host Bastión)

- Podemos utilizar un bastion host para acceder mediante SSH a nuestras instancias EC2 privadas
- El Bastion Host está en la subred pública, que a su vez está conectada a todas las demás subredes privadas
- **El grupo de seguridad del Bastion Host debe permitir** la entrada desde Internet en el puerto 22 desde un CIDR restringido, por ejemplo el CIDR público de tu empresa
- **El grupo de seguridad de las instancias EC2** debe permitir el grupo de seguridad del Bastion Host, o la IP privada del Bastion Host

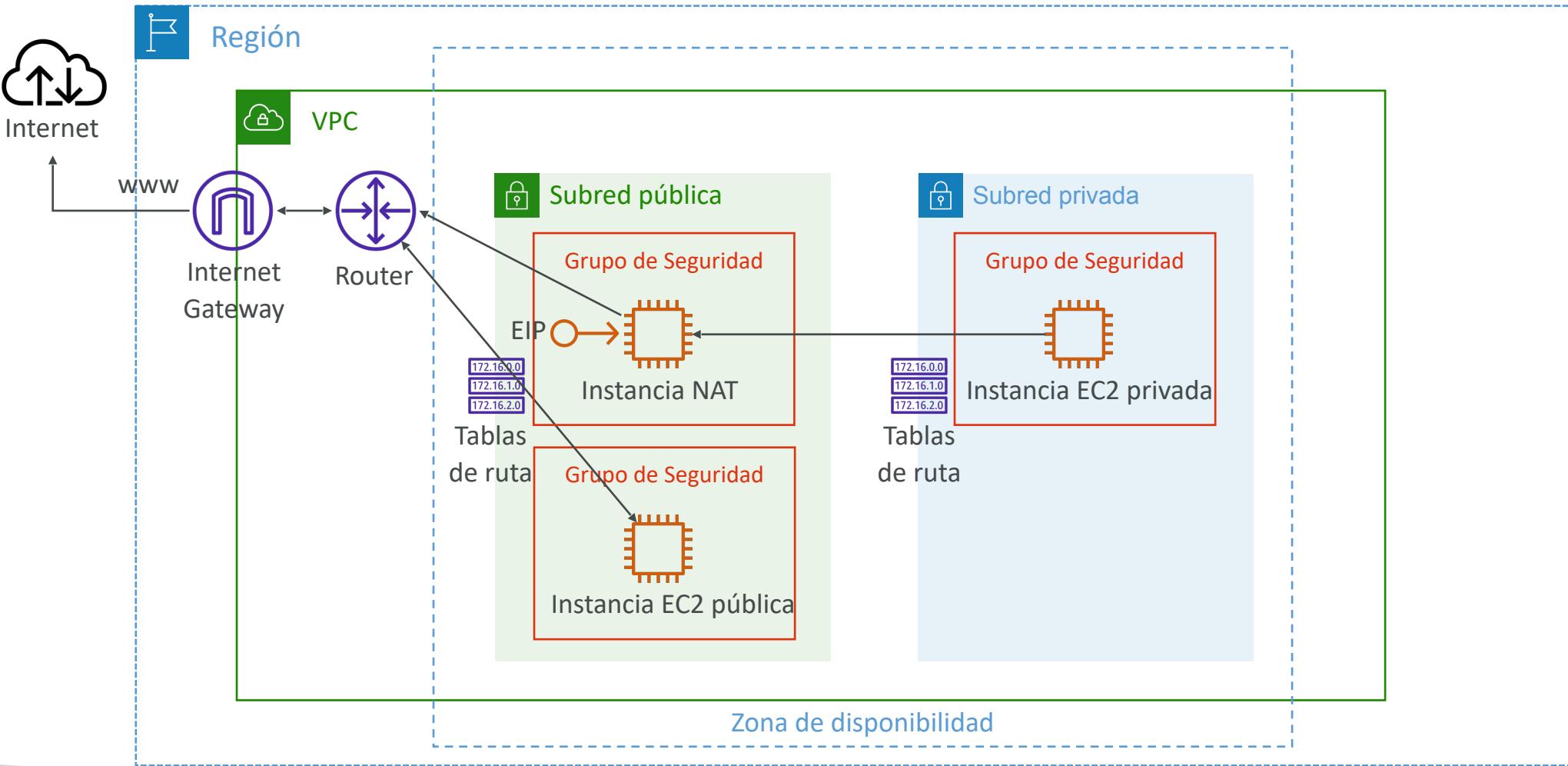


Instancia NAT (**obsoleta**, pero aún en el examen)

- **NAT = Traducción de direcciones de red**
- Permite a las instancias EC2 en subredes privadas conectarse a Internet
- Debe lanzarse en una subred pública
- Debe desactivar la configuración de EC2:
Comprobación origen / destino
- Debe tener IP elástica asociada
- Las tablas de ruta deben estar configuradas para dirigir el tráfico de subredes privadas a la Instancia NAT

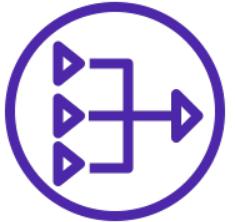


Instancia NAT



Instancia NAT - Comentarios

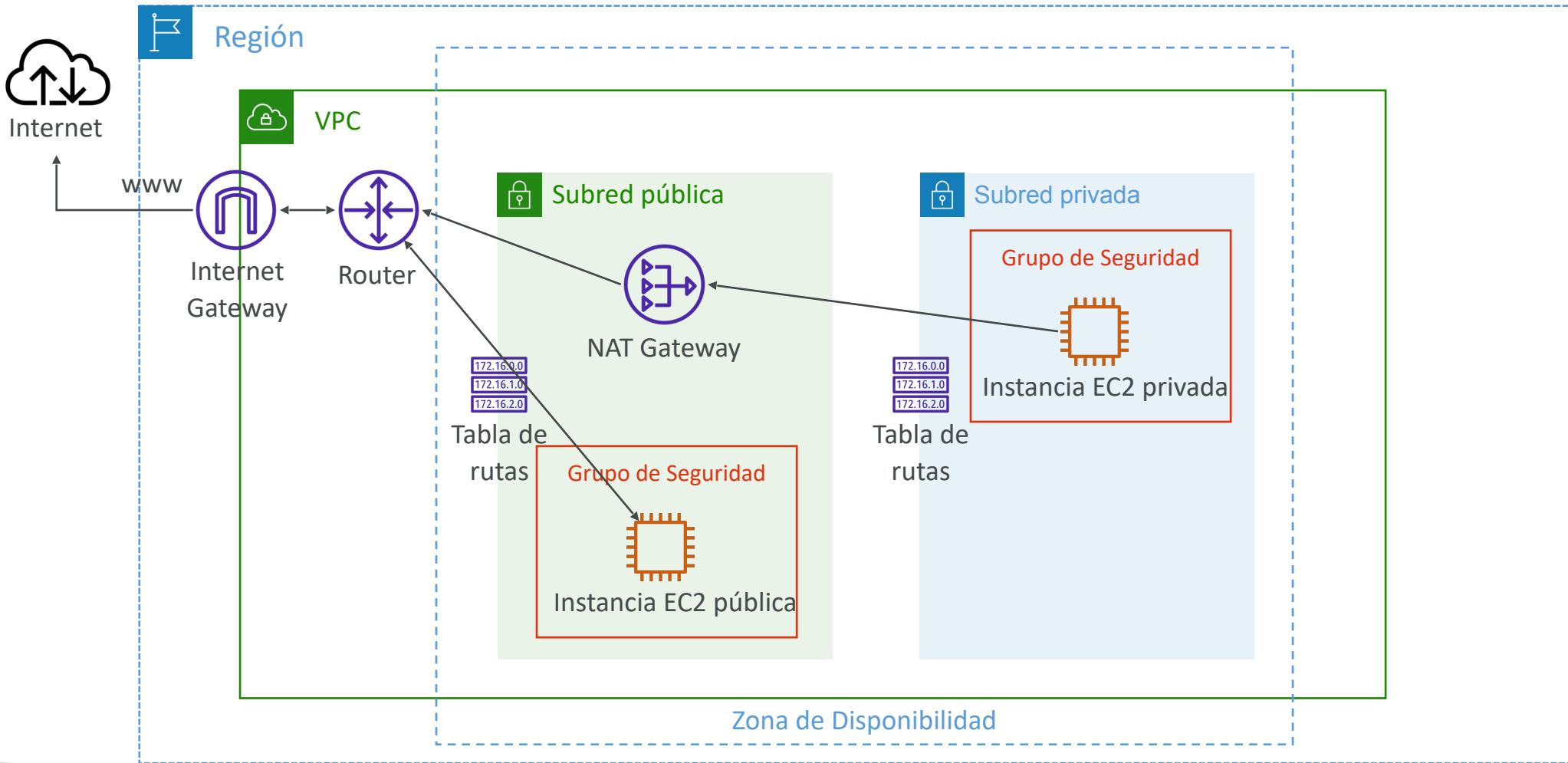
- Ya está disponible la AMI preconfigurada de Amazon Linux
 - El soporte estándar finalizó el 31 de diciembre de 2020
- No es una configuración de alta disponibilidad / resiliencia inmediata
 - Necesitas crear un ASG en multi-AZ + script de datos de usuario resiliente
- El ancho de banda del tráfico de Internet depende del tipo de instancia EC2
- Debes gestionar los Grupos de Seguridad y las reglas
 - Entrante:
 - Permitir tráfico HTTP / HTTPS procedente de subredes privadas
 - Permitir SSH desde tu red doméstica (el acceso se proporciona a través de Internet Gateway)
 - Saliente:
 - Permitir el tráfico HTTP / HTTPS hacia Internet



Gateways NAT

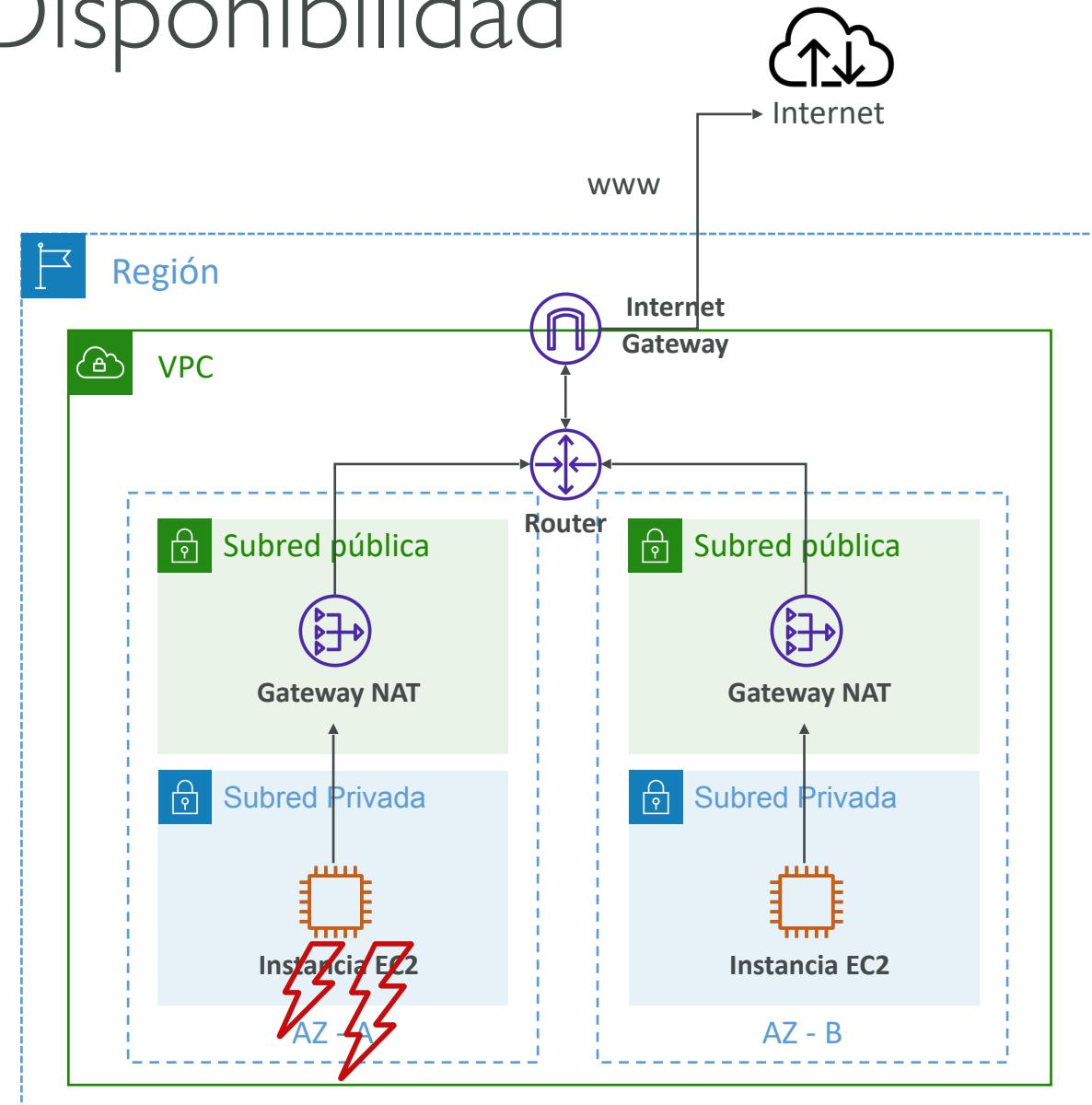
- NAT gestionado por AWS, mayor ancho de banda, alta disponibilidad, sin administración
- Paga por hora de uso y ancho de banda
- NATGW se crea en una Zona de Disponibilidad específica, utiliza una IP Elástica
- No puede ser utilizada por una instancia EC2 en la misma subred (sólo desde otras subredes)
- Requiere una IGW (subred privada => NATGW => IGW)
- 5 Gbps de ancho de banda con escalado automático hasta 45 Gbps
- No se necesitan / gestionan Grupos de Seguridad

Gateways NAT



Gateways NAT con Alta Disponibilidad

- Los Gateways NAT son resilientes dentro de una única Zona de Disponibilidad
- Debes crear varios **Gateways NAT en varias AZ** para la tolerancia a fallos
- No es necesaria la comutación por error entre zonas de disponibilidad, porque si una zona de disponibilidad se cae, no necesita NAT.



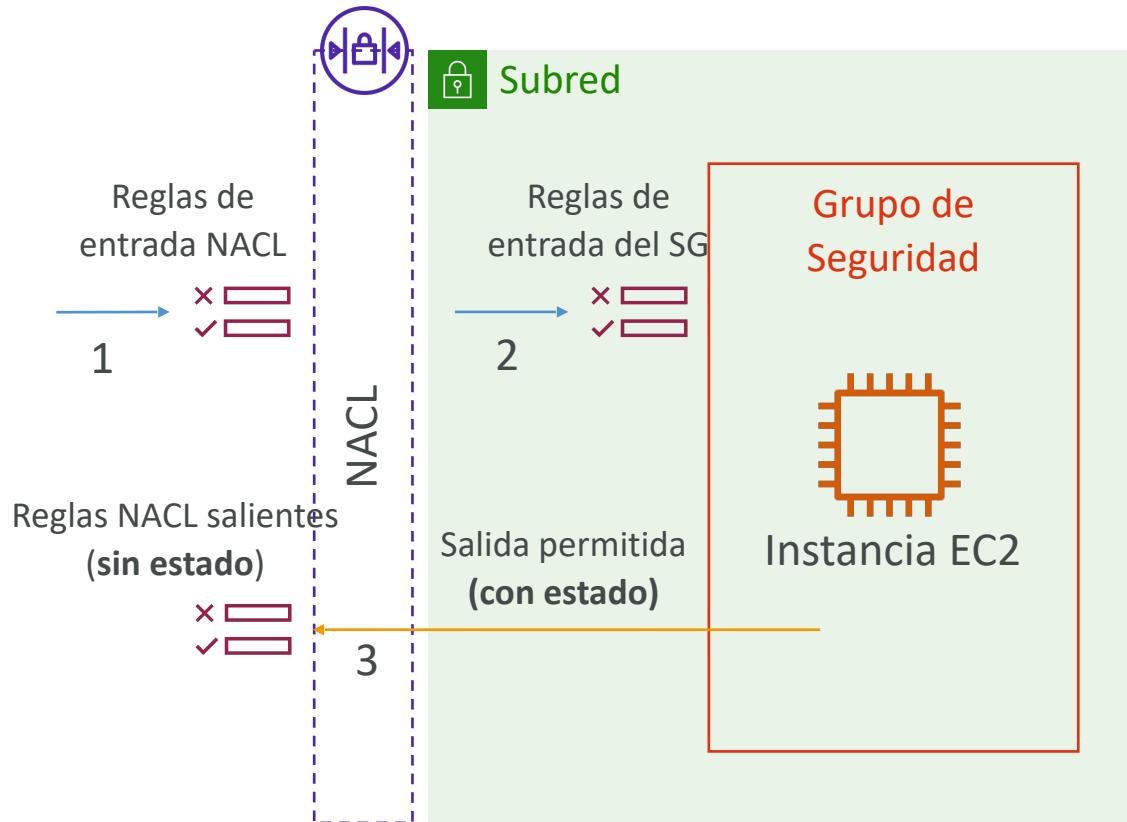
Los Gateways NAT vs. Instancia NAT

	Gateway NAT	Instancia NAT
Disponibilidad	Alta disponibilidad dentro de AZ (crear en otra AZ)	Utiliza un script para gestionar la conmutación por error entre instancias
Ancho de banda	Hasta 45 Gbps	Depende del tipo de instancia EC2
Mantenimiento	Gestionado por AWS	Gestionados por ti (por ejemplo, software, parches del SO, ...)
Coste	Por hora y cantidad de datos transferidos	Por hora, tipo y tamaño de instancia EC2, + red \$
Publica IPv4	✓	✓
Privada IPv4	✓	✓
Grupos de seguridad	✗	✓
¿Utilizar como Host Bastion?	✗	✓

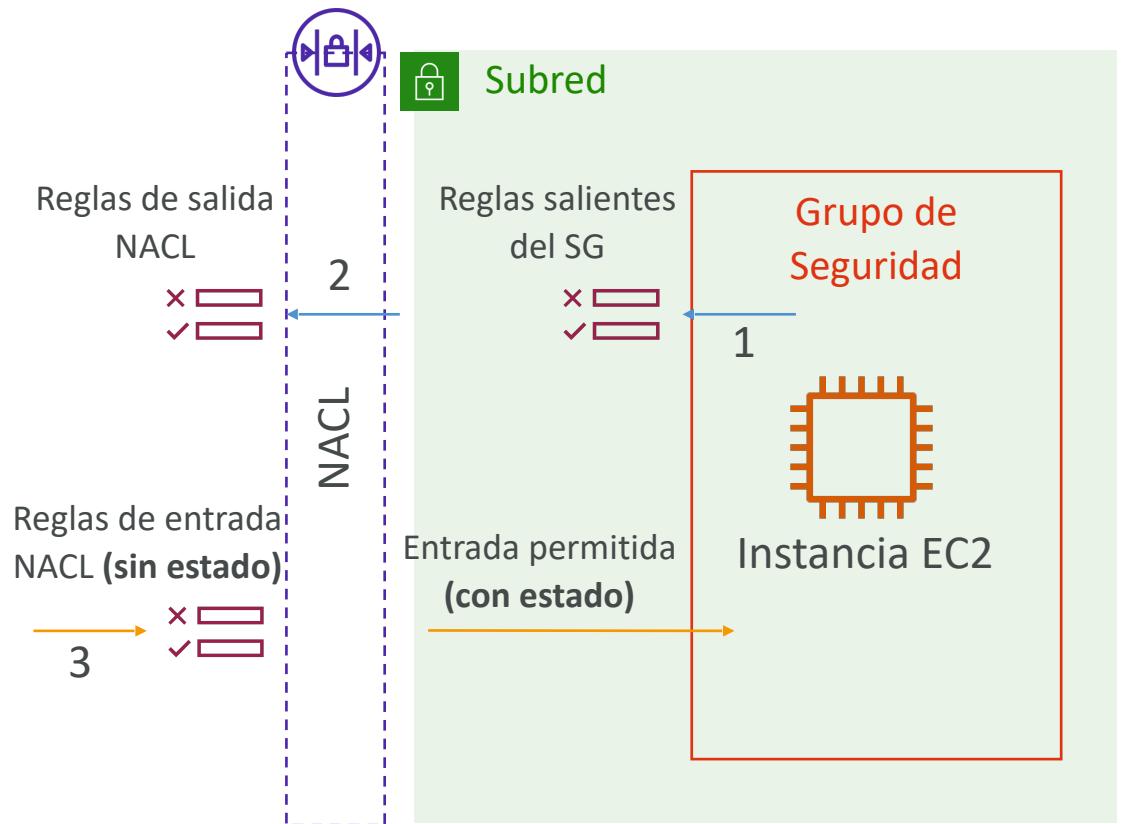
Más en: <https://docs.aws.amazon.com/vpc/latest/userguide/vpc-nat-comparison.html>

Grupos de seguridad y NACL

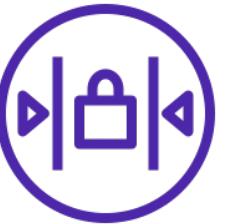
Petición entrante



Petición de salida

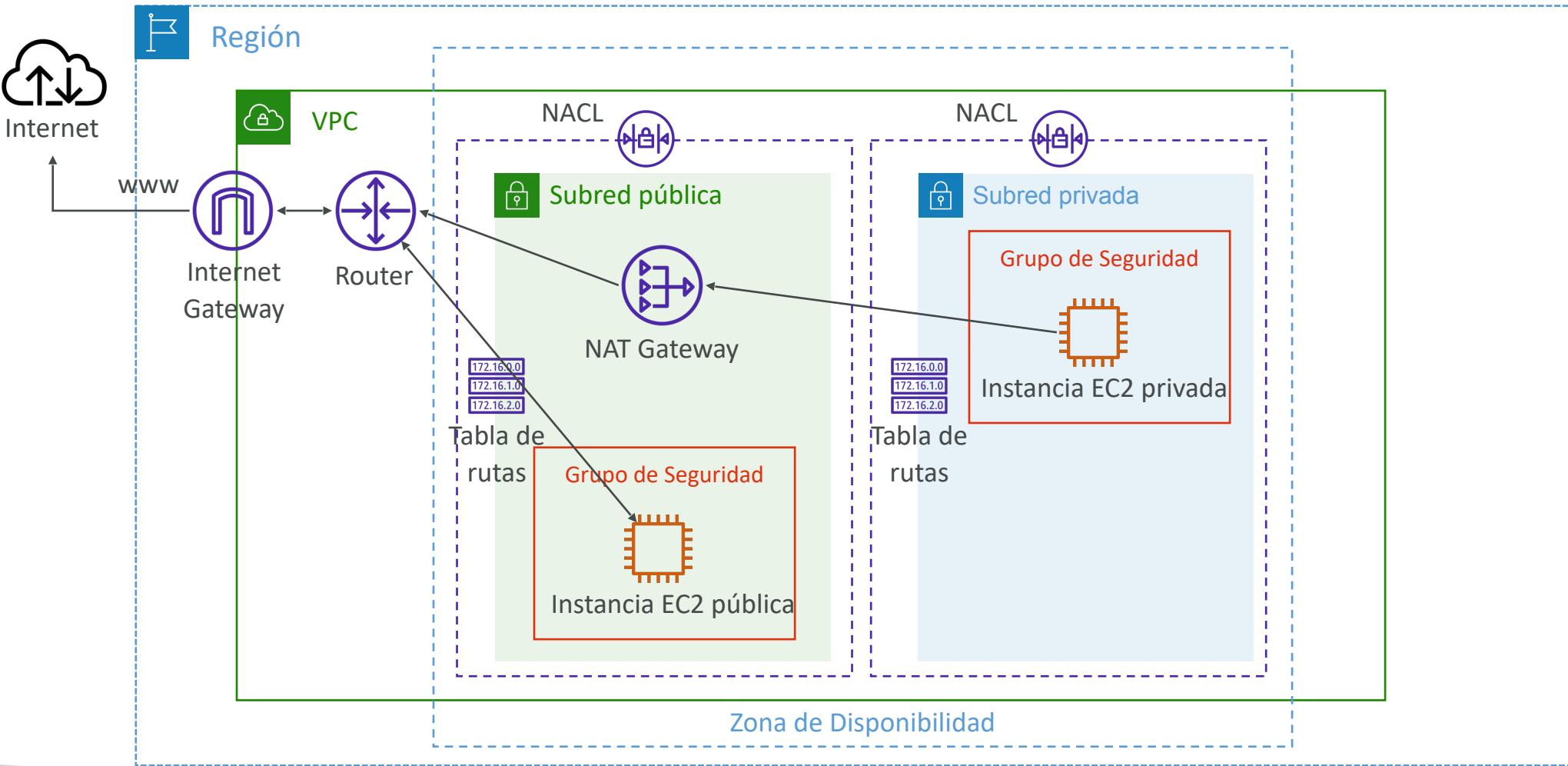


Lista de control de acceso a la red (NACL)



- Las NACL son como un firewall que controla el tráfico desde y hacia las **subredes**
- **Una NACL por subred**, a las subredes nuevas se les asigna la **NACL por defecto**
- Tú defines las **Reglas NACL**:
 - Las reglas tienen un número (1-32766), mayor precedencia con un número menor
 - La primera coincidencia de reglas determinará la decisión
 - Ejemplo: si defines #100 PERMITIR 10.0.0.10/32 y #200 DENEGAR 10.0.0.10/32, se permitirá la dirección IP porque 100 tiene mayor precedencia que 200
 - La última regla es un asterisco (*) y deniega una petición en caso de que no coincida ninguna regla
 - AWS recomienda añadir reglas en incrementos de 100
- Las NACL recién creadas lo denegarán todo
- Las NACL son una buena forma de bloquear una dirección IP concreta a nivel de subred

NACLs



NACL por defecto

- Acepta todo lo que entra/sale con las subredes a las que está asociado
- **NO** modifiques la NACL por defecto, en su lugar crea NACLs personalizadas



NACL por defecto para una VPC que soporta IPv4

Reglas de entrada

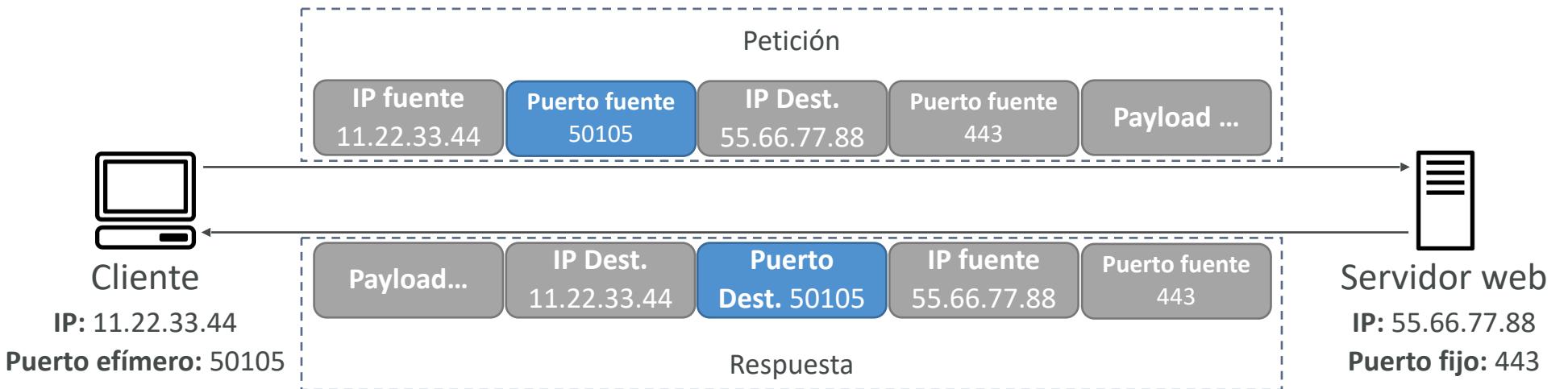
Regla #	Tipo	Protocolo	Alcance del puerto	Destino	Permitir/Denegar
100	Todo el tráfico IPv4	Todos	Todos	0.0.0.0/0	PERMITIR
*	Todo el tráfico IPv4	Todos	Todos	0.0.0.0/0	DENEGAR

Reglas de salida

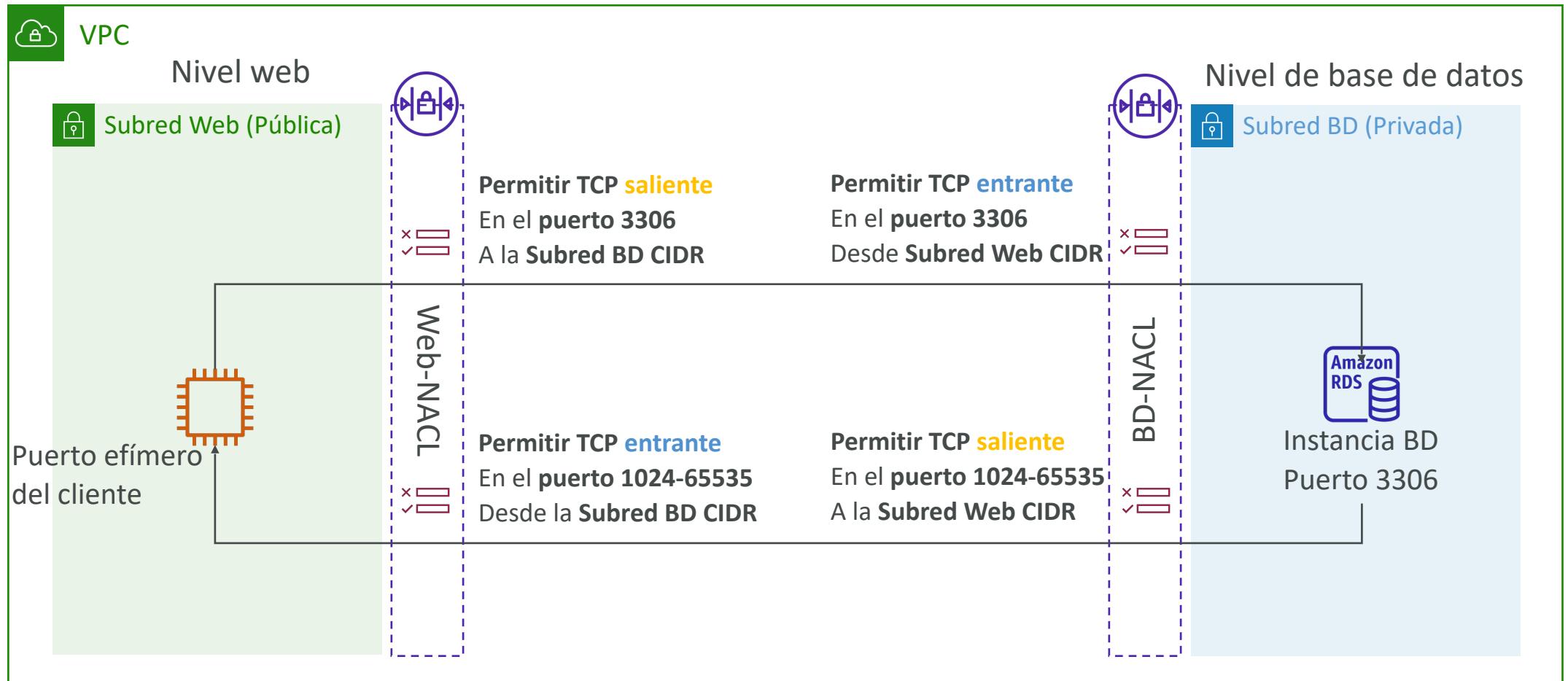
Regla #	Tipo	Protocolo	Alcance del puerto	Destino	Permitir/Denegar
100	Todo el tráfico IPv4	Todos	Todos	0.0.0.0/0	PERMITIR
*	Todo el tráfico IPv4	Todos	Todos	0.0.0.0/0	DENEGAR

Puertos efímeros

- Para que dos endpoints cualesquiera establezcan una conexión, deben utilizar puertos
- Los clientes se conectan a **un puerto definido**, y esperan una respuesta en un **puerto efímero**
- Los distintos Sistemas Operativos utilizan distintos rangos de puertos, ejemplos:
 - IANA y MS Windows 10 → 49152 - 65535
 - Muchos núcleos Linux → 32768 - 60999

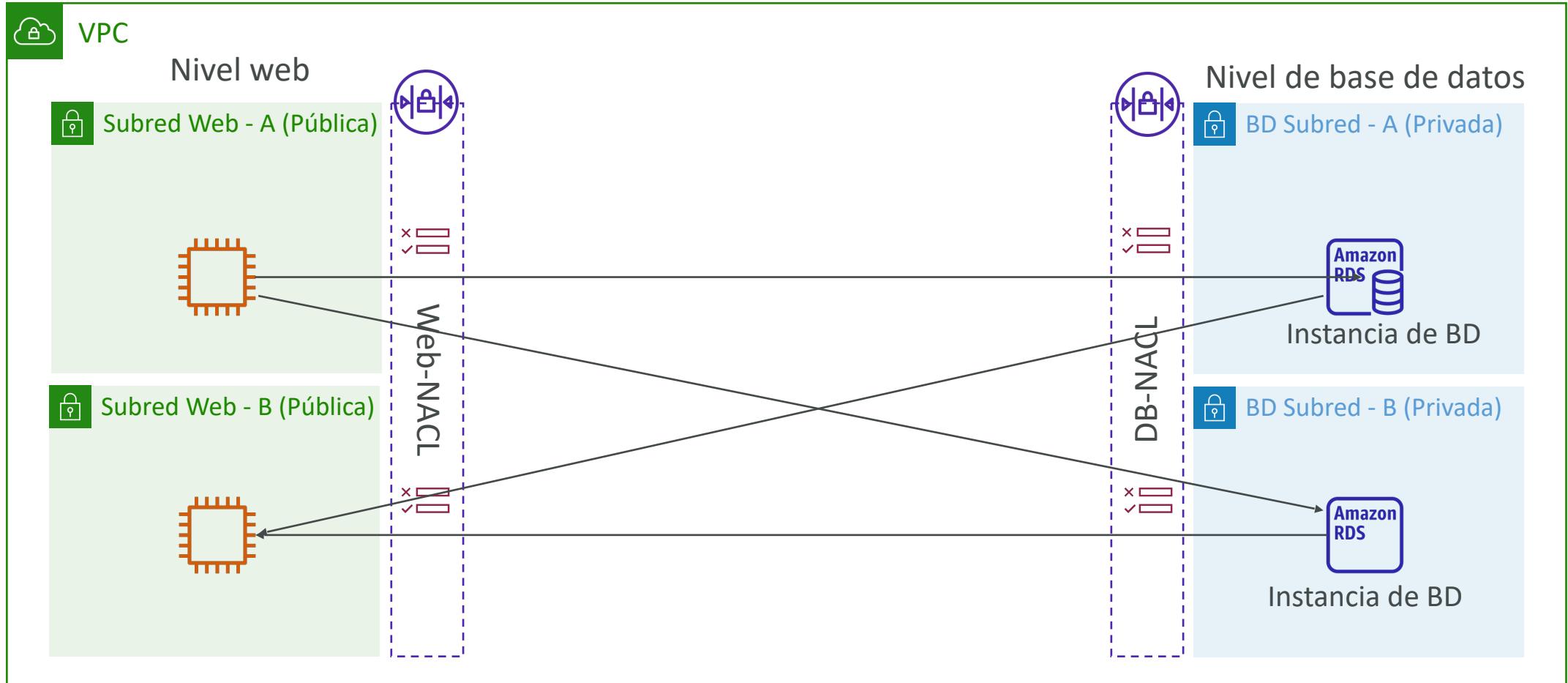


NACL con puertos efímeros



<https://docs.aws.amazon.com/vpc/latest/userguide/vpc-network-acls.html#nacl-ephemeral-ports>

Crear reglas NACL para cada subred de destino CIDR



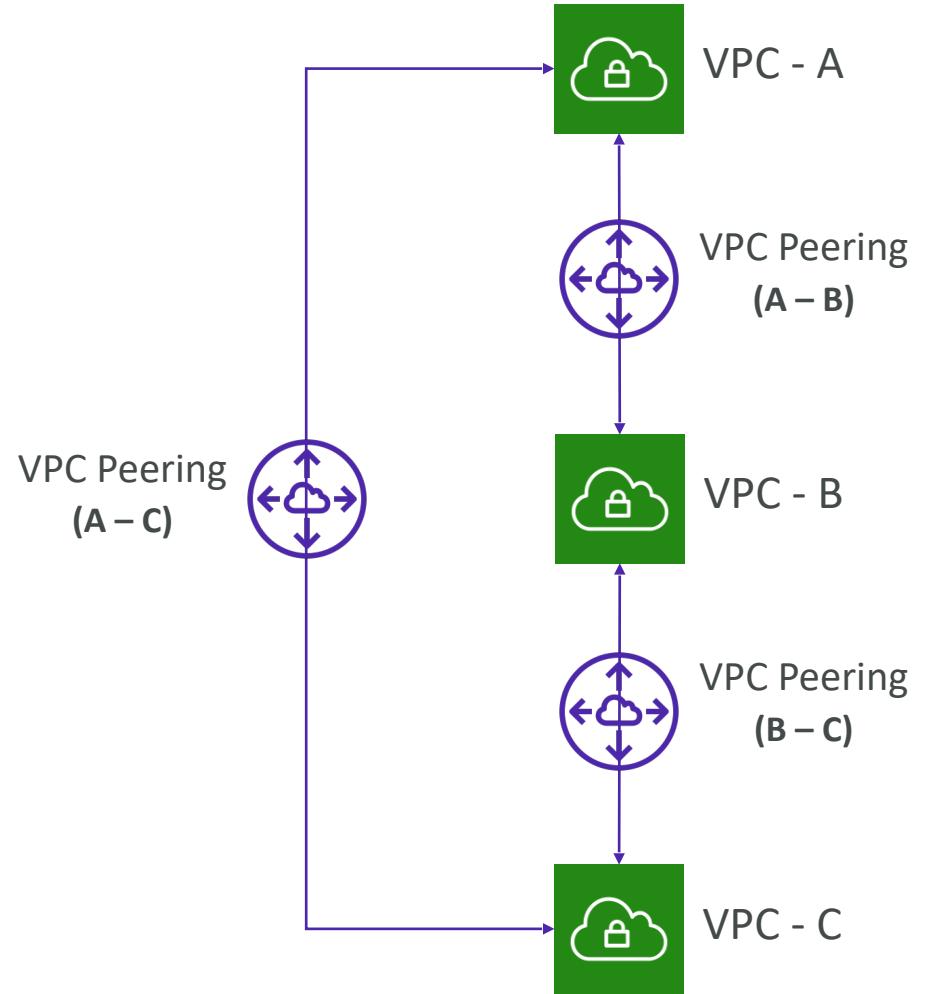
Grupo de Seguridad vs. NACLs

Grupo de Seguridad	NACL
Funciona a nivel de instancia	Funciona a nivel de subred
Soporta sólo reglas de permiso	Soporta reglas de permiso y reglas de denegación
Estado completo: el tráfico de retorno se permite automáticamente, independientemente de cualquier regla	Sin estado: el tráfico de retorno debe estar explícitamente permitido por reglas (piensa en puertos efímeros)
Todas las reglas se evalúan antes de decidir si se permite el tráfico	Las reglas se evalúan en orden (de menor a mayor) al decidir si se permite el tráfico, la primera coincidencia gana
Se aplica a una instancia EC2 cuando alguien lo especifica	Se aplica automáticamente a todas las instancias EC2 de la subred a la que está asociado

NACL Ejemplos: <https://docs.aws.amazon.com/vpc/latest/userguide/vpc-network-acls.html>

VPC Peering

- Conecta de forma privada dos VPC utilizando la red de AWS
- Haz que se comporten como si estuvieran en la misma red
- No deben tener CIDRs solapados
- La VPC Peering connection **NO es transitiva** (debe establecerse para cada VPC que necesite comunicarse entre sí)
- **Debes actualizar las tablas de rutas en las subredes de cada VPC para garantizar que las instancias EC2 puedan comunicarse entre sí**



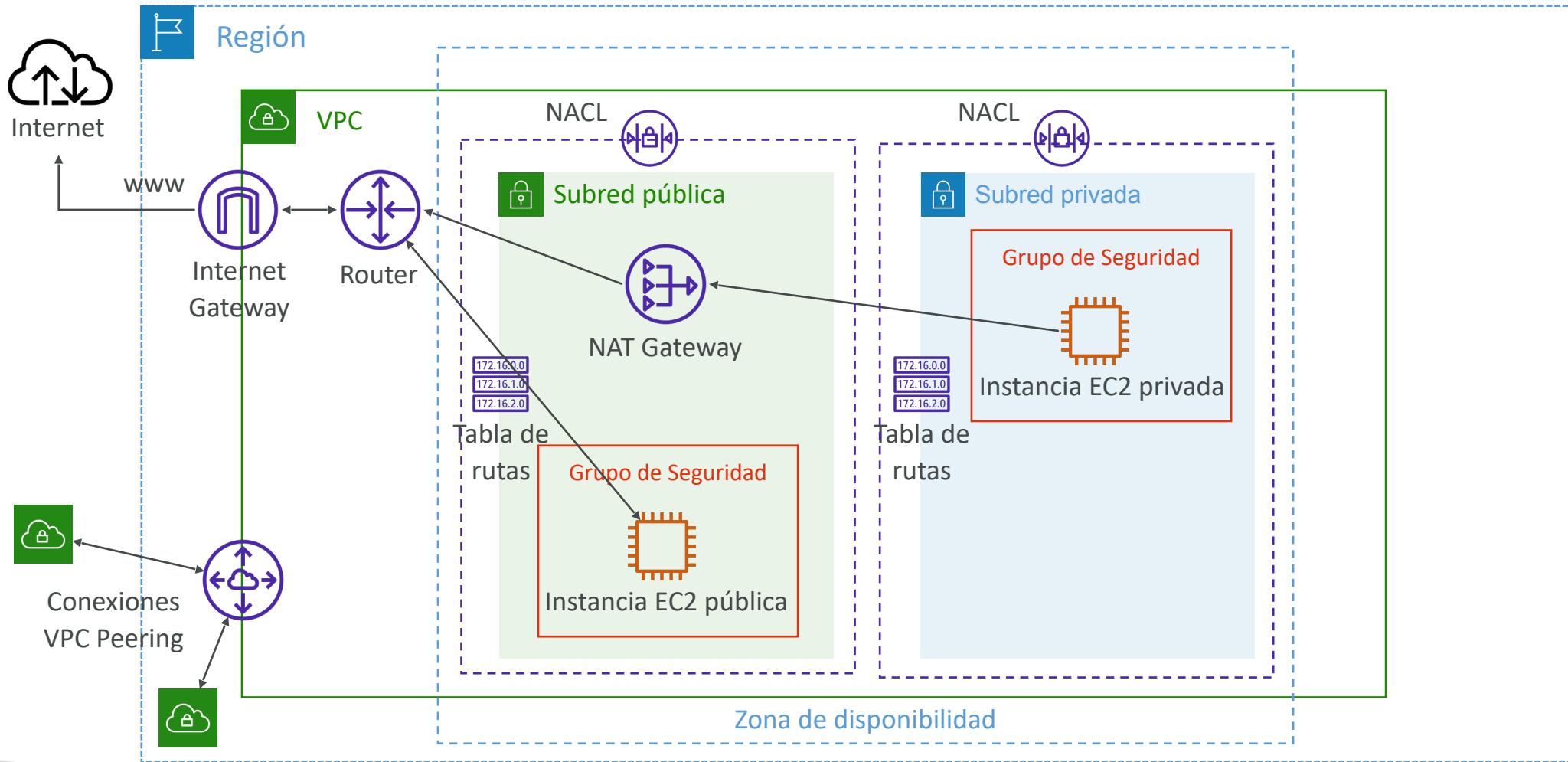
VPC Peering - Es bueno saber que...

- Puedes crear VPC Peering connection entre VPCs en **diferentes cuentas/regiones de AWS**
- Puedes hacer referencia a un grupo de seguridad en una VPC peered (funciona entre cuentas - misma región)

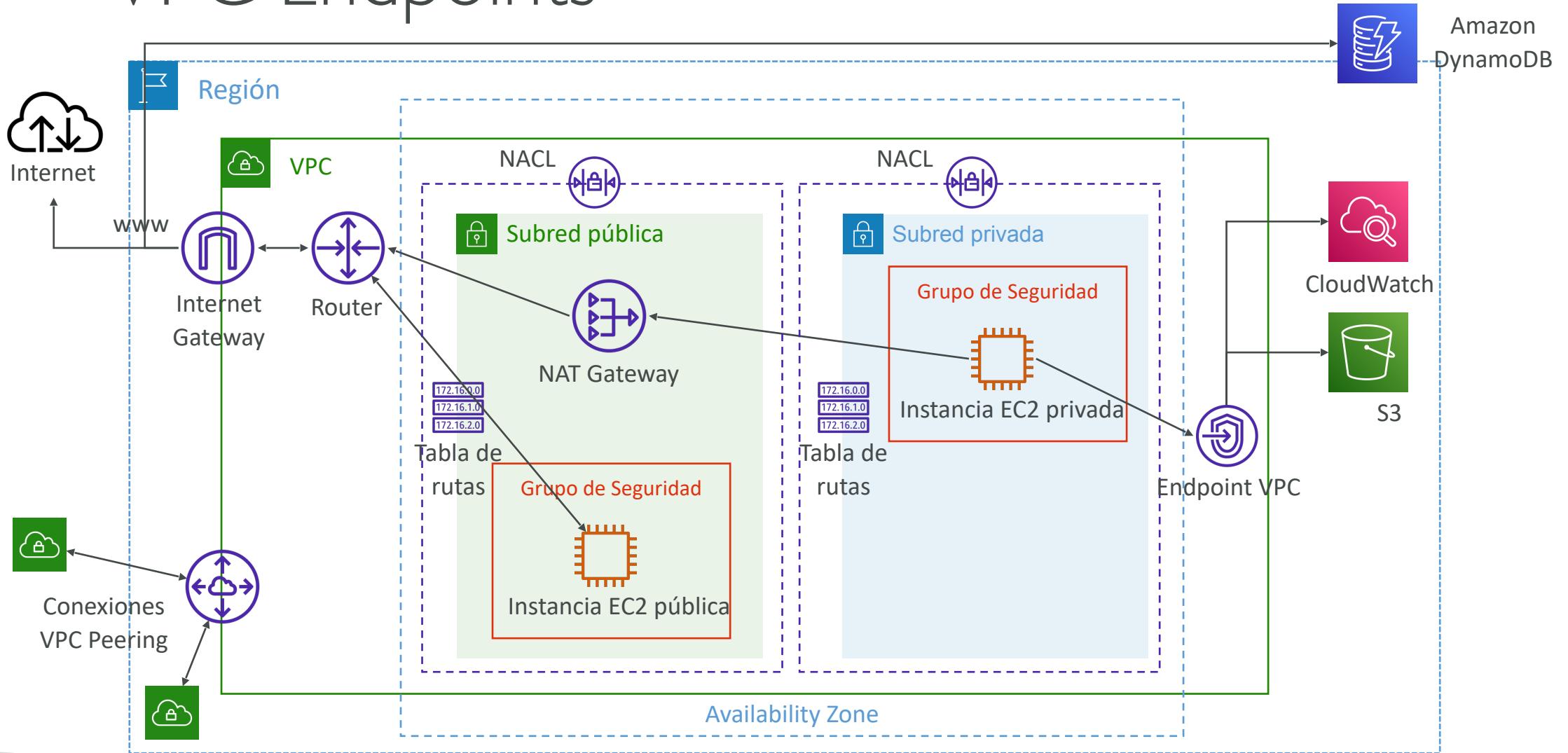
Type	Protocol	Port range	Source
HTTP	TCP	80	sg-04991f9af3473b939 / default
HTTP	TCP	80	[REDACTED] / sg-027ad1f7865d4be76

ID de cuenta

VPC Peering



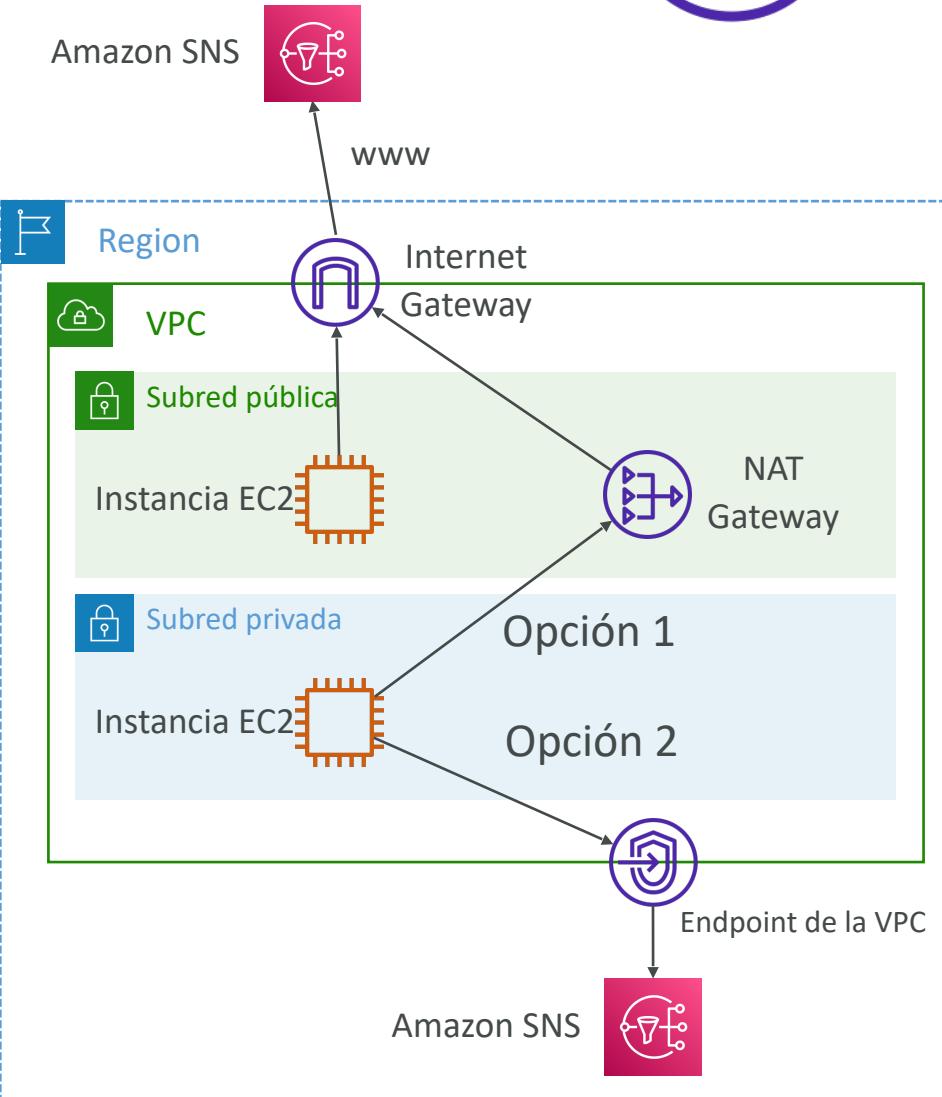
VPC Endpoints



Endpoints de la VPC (AWS PrivateLink)



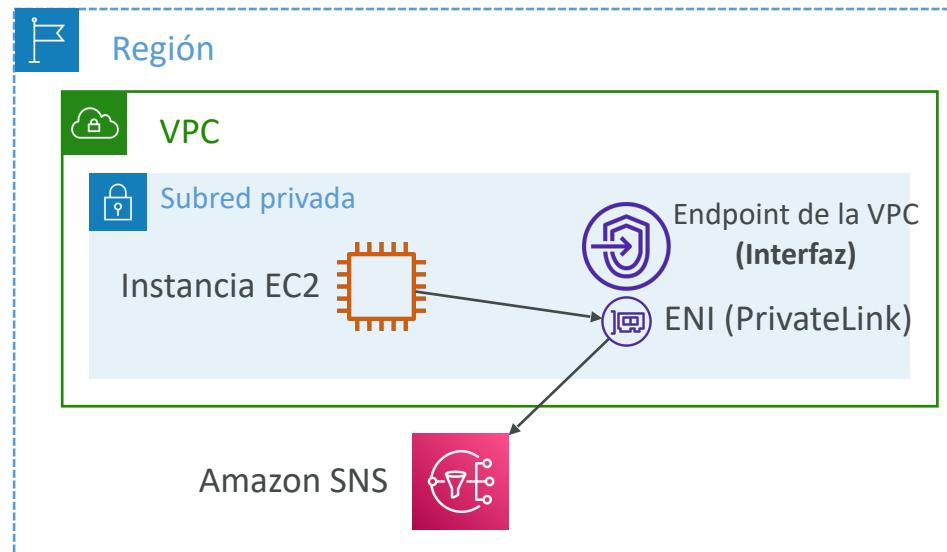
- Cada servicio de AWS está expuesto públicamente (URL pública)
- Los endpoints de la VPC (con tecnología AWS PrivateLink) te permiten conectarte a los servicios de AWS mediante una **red privada** en lugar de utilizar la Internet pública
- Son redundantes y escalan horizontalmente
- Eliminan la necesidad de IGW, NATGW, ... para acceder a los servicios de AWS
- En caso de problemas
 - Comprueba la Resolución de Ajustes DNS en tu VPC
 - Comprueba las tablas de rutas



Tipos de endpoints

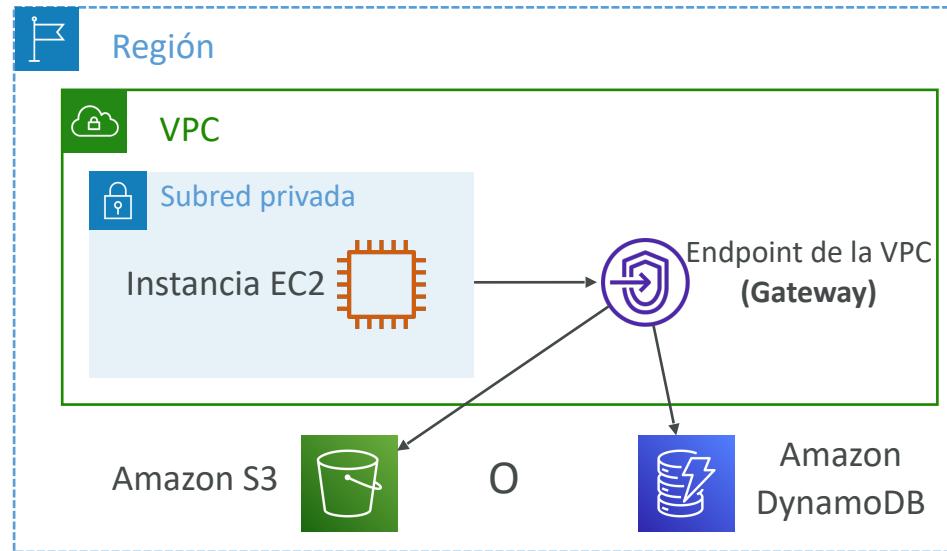
- **Endpoints de interfaz (alimentados por PrivateLink)**

- Proporciona una ENI (dirección IP privada) como punto de entrada (debe adjuntar un Grupo de Seguridad)
- Soporta la mayoría de los servicios de AWS
- \$ por hora + \$ por GB de datos procesados



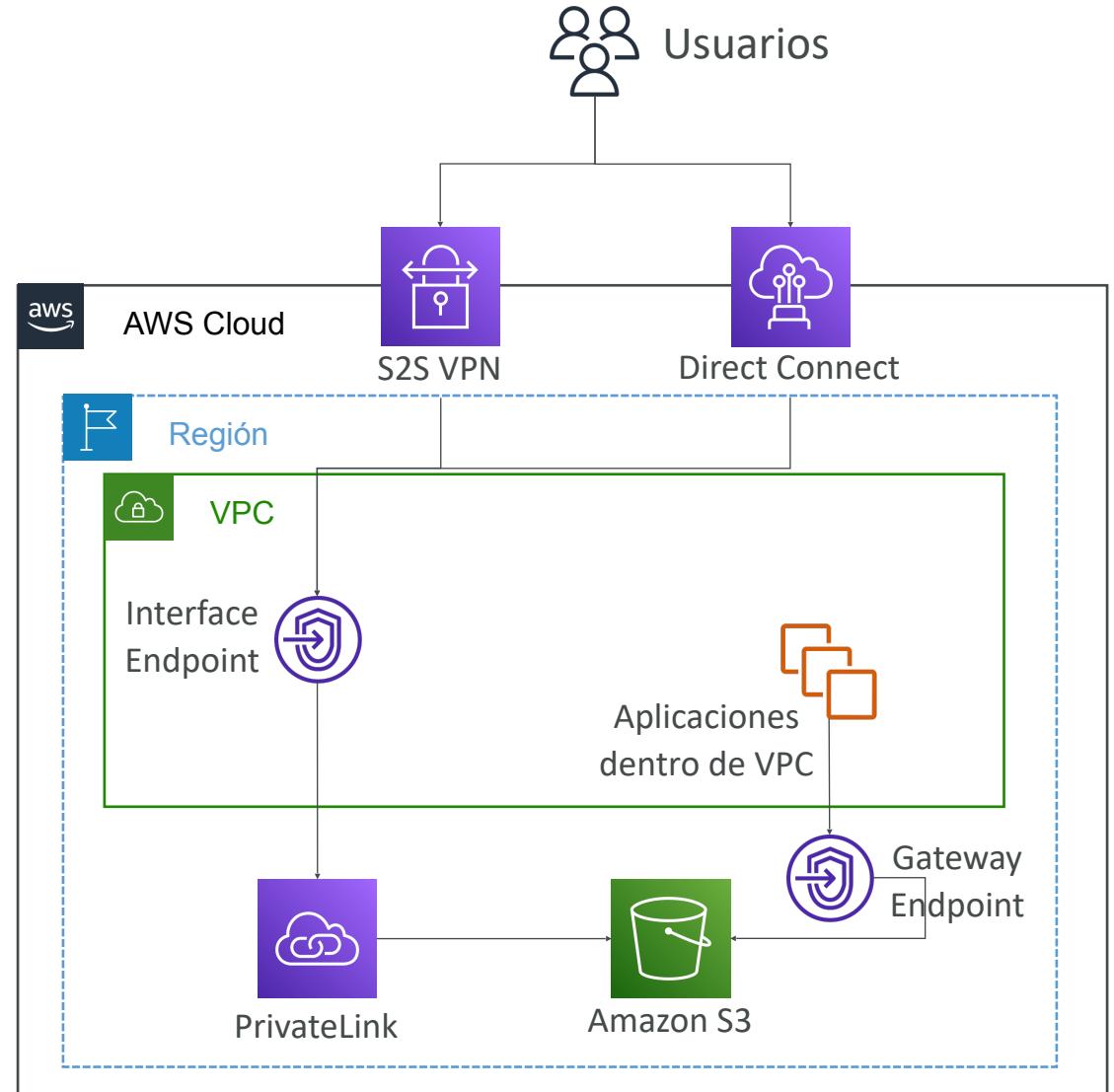
- **Gateway Endpoints**

- Proporciona un Gateway y debe utilizarse como destino en una tabla de rutas (no utiliza grupos de seguridad)
- Soporta tanto S3 como DynamoDB
- Gratis



¿Gateway o Interface Endpoint para S3?

- Lo más probable es que se prefiera Gateway todo el tiempo en el examen
- Coste: gratis para Gateway, \$ para interface endpoint
- Interfaz endpoint es preferible si el acceso se requiere desde las instalaciones (VPN Site to Site o Direct Connect), una VPC diferente o una región diferente

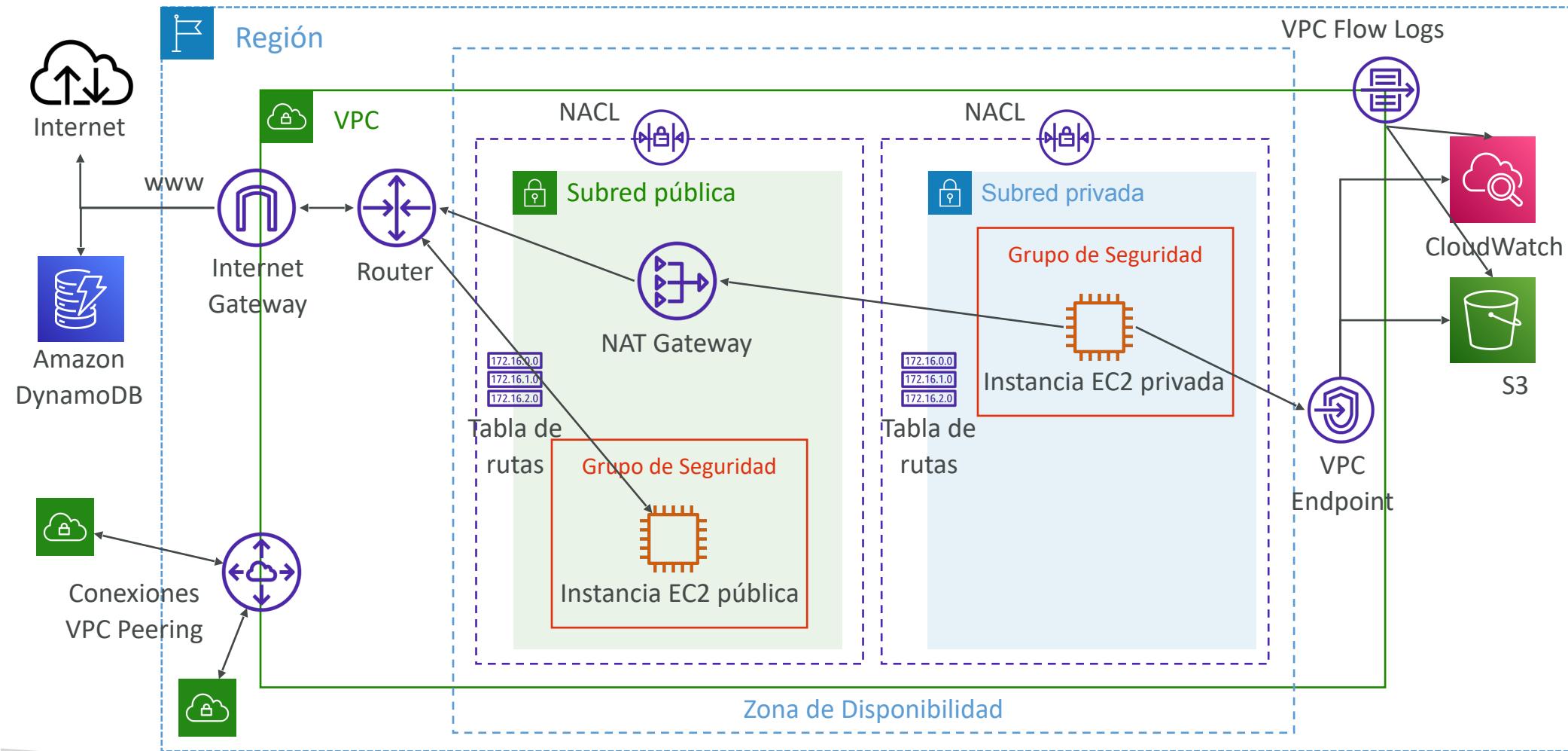




Logs de flujo de la VPC

- Captura información sobre el tráfico IP que entra en tus interfaces:
 - Logs de flujo de VPC
 - Logs de flujo de subred
 - Logs de flujo de Elastic Network Interface (ENI)
- Ayuda a supervisar y solucionar problemas de conectividad
- Los datos de los logs de flujo pueden ir a S3 / CloudWatch Logs
- Captura también información de red de las interfaces gestionadas por AWS ELB, RDS, ElastiCache, Redshift, WorkSpaces, NATGW, Transit Gateway...

Logs de flujo de la VPC



Sintaxis de los logs de flujo de la VPC

versión	Id de interfaz	dstaddr	dstport	paquetes	inicio	acción
2	123456789010	eni-1235b8ca123456789	172.31.16.139	172.31.16.21	20641	22 6 20 4249 1418530010 1418530070 ACCEPT OK
2	123456789010	eni-1235b8ca123456789	172.31.9.69	172.31.9.12	49761	3389 6 20 4249 1418530010 1418530070 REJECT OK
Id de cuenta	srcaddr	srcport	protocolo	bytes	fin	logs-estado

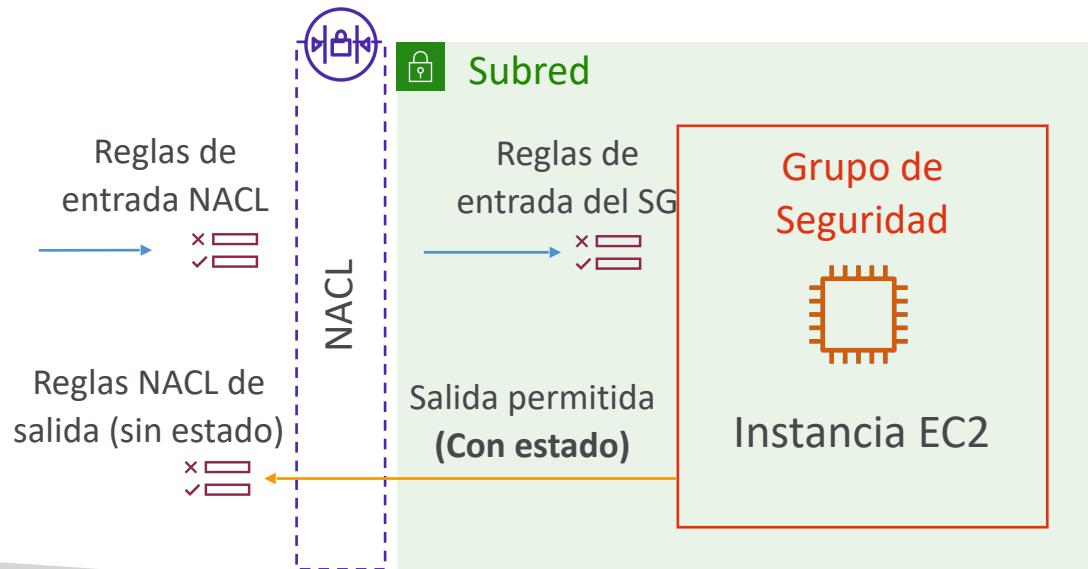
- **srcaddr & dstaddr** - ayudan a identificar la IP problemática
- **srcport & dstport** - ayudan a identificar los puertos problemáticos
- **Acción** - éxito o fracaso de la petición debido al Grupo de Seguridad / NACL
- Puede utilizarse para analizar patrones de uso o comportamientos maliciosos
- **Consulta los logs de flujo de la VPC utilizando Athena en S3 o CloudWatch Logs Insights**
- Ejemplos de logs de flujo:
 - <https://docs.aws.amazon.com/vpc/latest/userguide/flow-logs-records-examples.html>

VPC Flow Logs - Solucionar problemas de SG y NACL

Mira el campo “ACCIÓN”

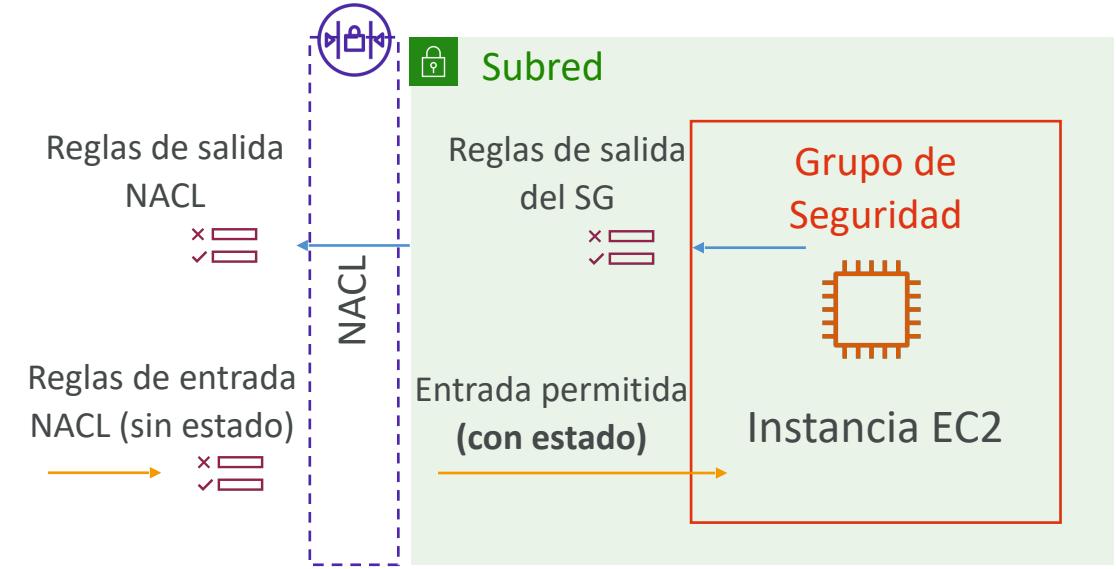
Peticiones entrantes

- RECHAZO Entrante => NACL o SG
- Entrada ACCEPT, Salida REJECT => NACL



Peticiones salientes

- RECHAZO Saliente => NACL o SG
- Salida ACCEPT, Entrada REJECT => NACL



VPC Flow Logs - Arquitecturas



VPC Flow Logs



CloudWatch Logs



CloudWatch Contributors
Insights

Top-10 de direcciones IP



VPC Flow Logs



CloudWatch Logs



Filtro de métricas
SSH, RDP...



CloudWatch Alarm



Alerta



Amazon SNS



VPC Flow Logs



Bucket S3

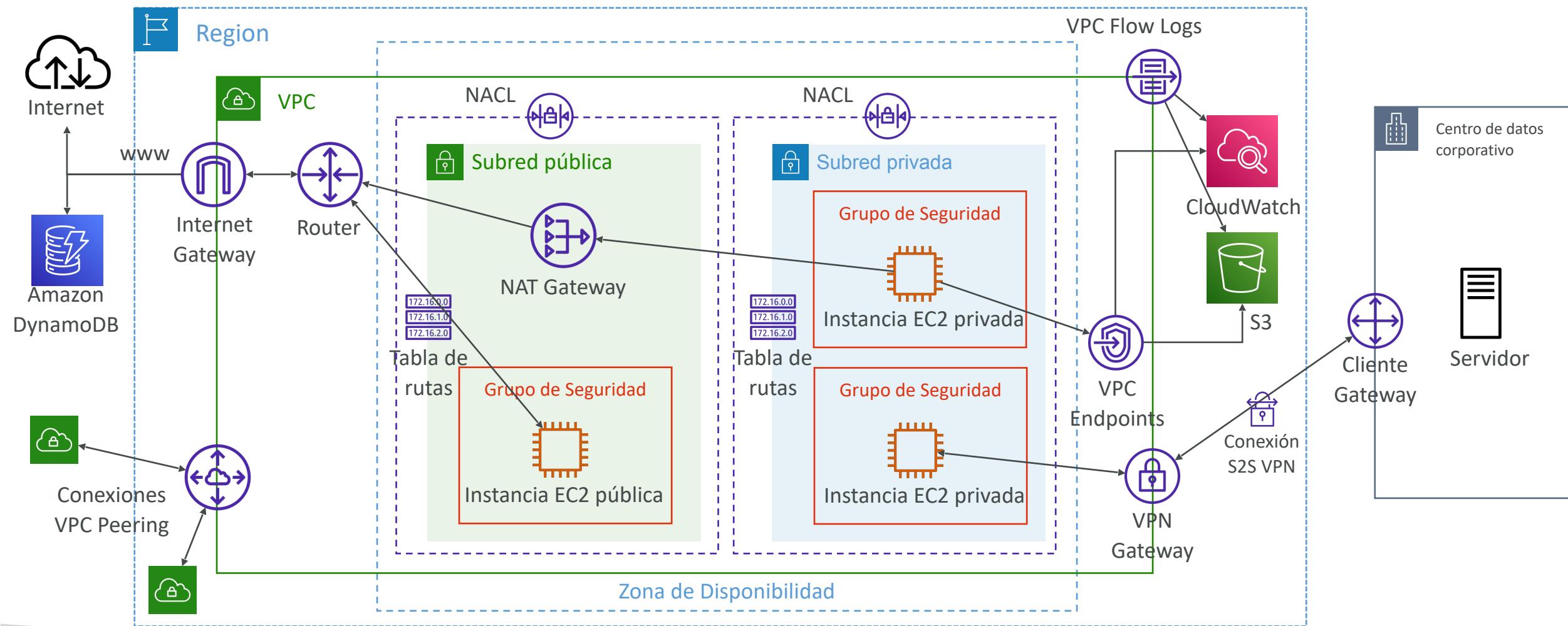


Amazon
Athena



Amazon
QuickSight

VPN Site-to-Site de AWS (Sitio a Sitio)



VPN Site-to-Site de AWS (Sitio a Sitio)



- **Virtual Private Gateway / Gateway Privado Virtual (VGW)**

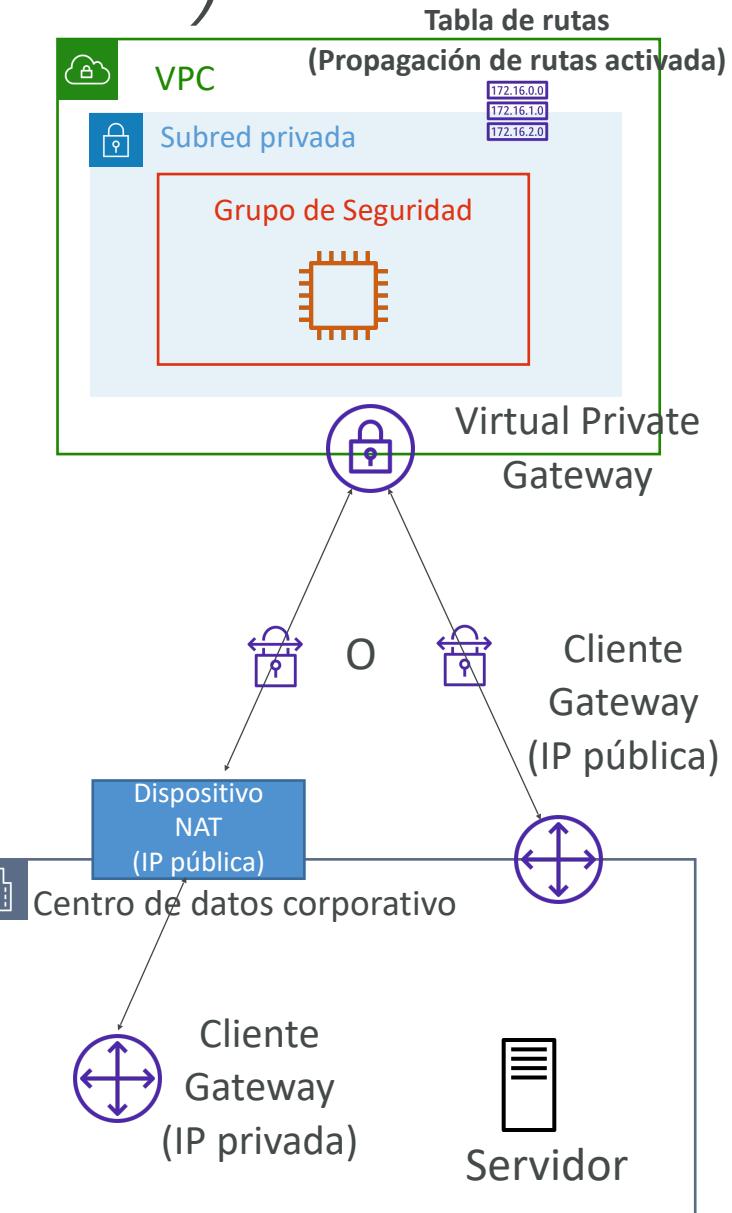
- Concentrador VPN en el lado AWS de la conexión VPN
- La VGW se crea y se adjunta a la VPC desde la que quieras crear la conexión VPN Site-to-Site (Sitio a Sitio)
- Posibilidad de personalizar el ASN (Número de Sistema Autónomo)

- **Gateway del cliente (CGW)**

- Aplicación de software o dispositivo físico en el lado del cliente de la conexión VPN
- <https://docs.aws.amazon.com/vpn/latest/s2svpn/your-cgw.html#DevicesTested>

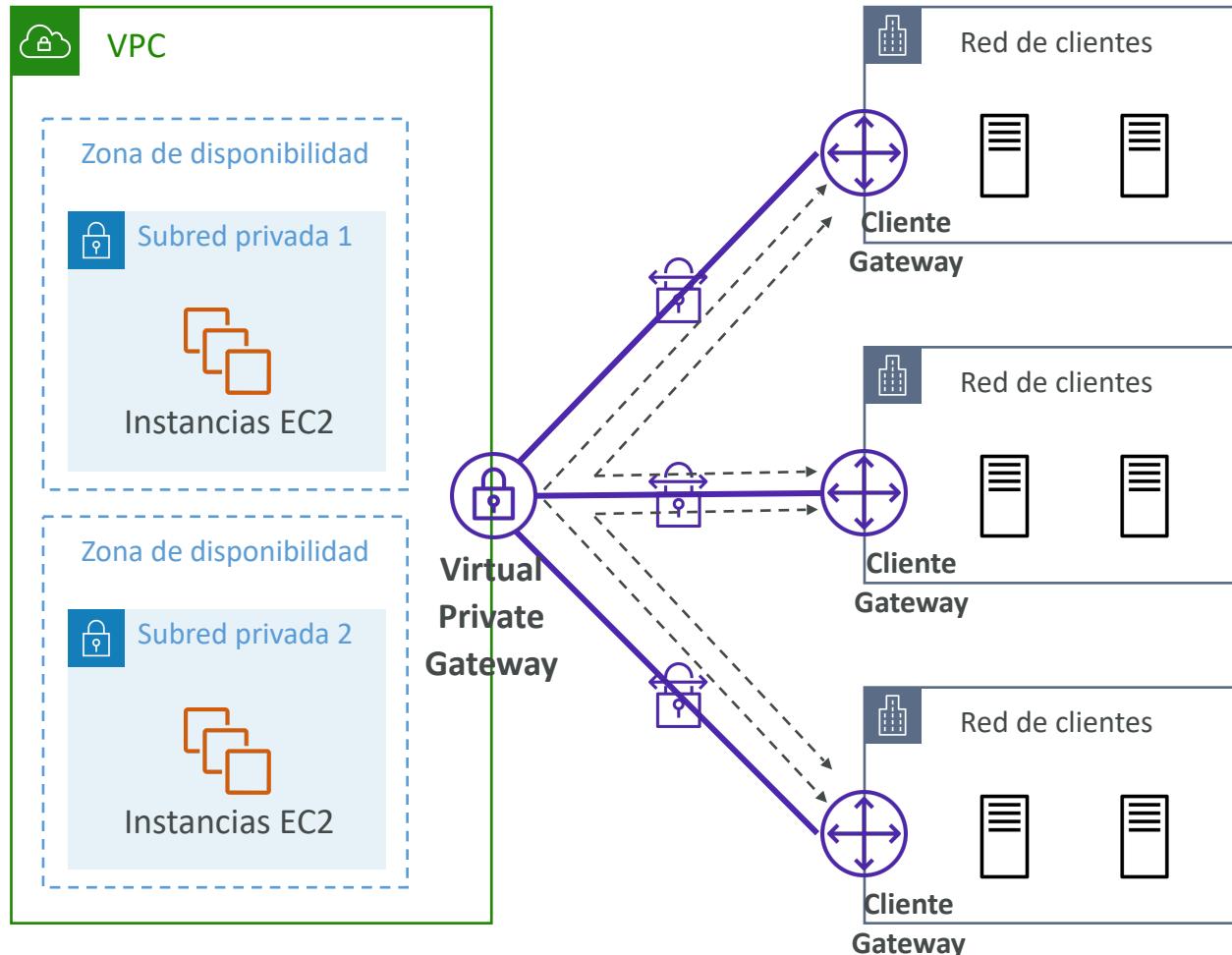
Conexiones VPN Site-to-Site (Sitio a Sitio)

- **Dispositivo Gateway del cliente (en las instalaciones)**
 - **¿Qué dirección IP utilizar?**
 - Dirección IP pública enrutable por Internet para tu dispositivo Gateway del cliente
 - Si está detrás de un dispositivo NAT habilitado para atravesar NAT (NAT-T), utiliza la dirección IP pública del dispositivo NAT
- **Paso importante:** activa la **propagación de rutas** para la Puerta de enlace virtual en la tabla de rutas asociada a tus subredes.
- Si necesitas hacer ping a tus instancias EC2 desde el local, asegúrate de añadir el protocolo ICMP en la entrada de tus grupos de seguridad



VPN de AWS CloudHub

- Proporciona una comunicación segura entre varias sedes, si tienes varias conexiones VPN
- Modelo hub-and-spoke de bajo coste para la conectividad de red primaria o secundaria entre distintas sedes (sólo VPN)
- Es una conexión VPN, así que va por la Internet pública
- Para configurarla, conecta varias conexiones VPN en la misma VGW, establece un enrutamiento dinámico y configura tablas de rutas

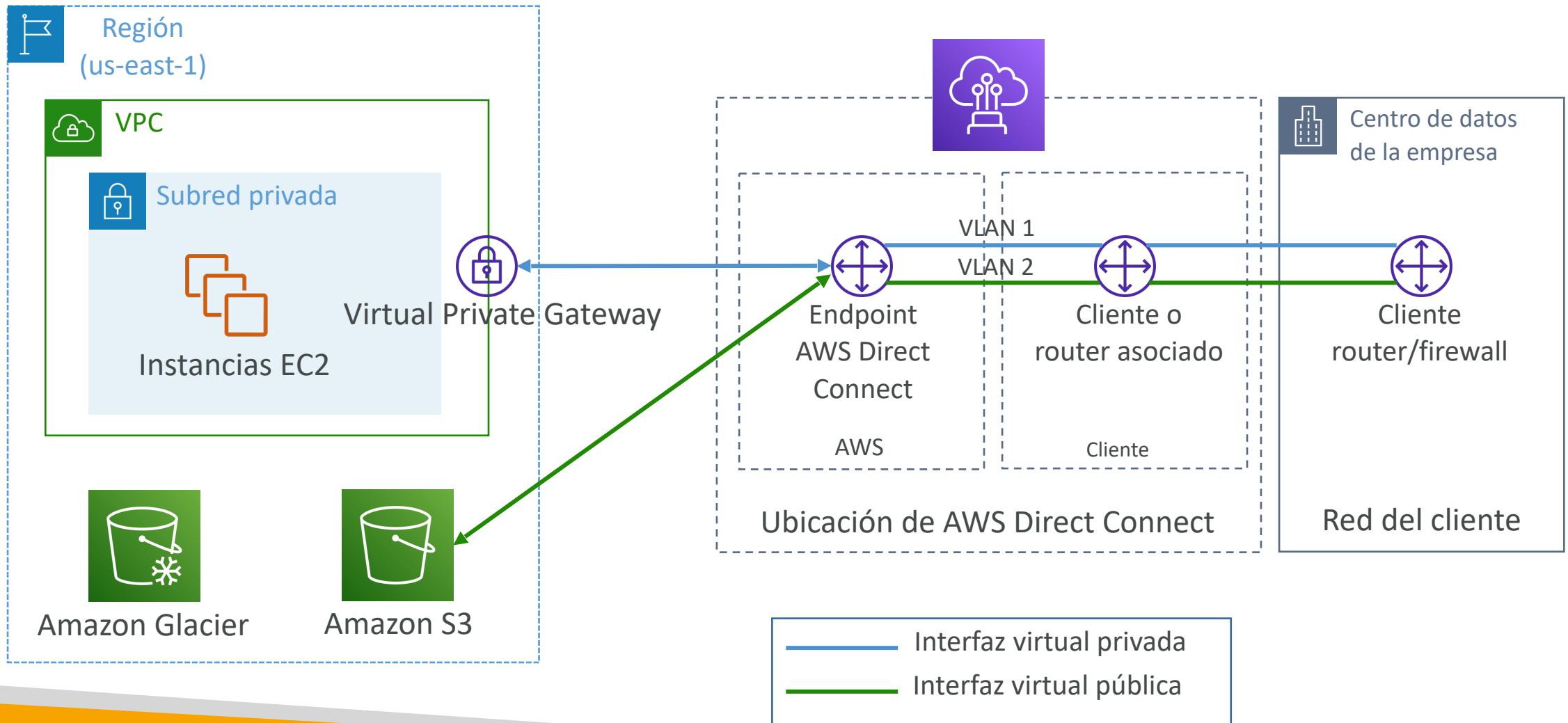




Direct Connect (DX)

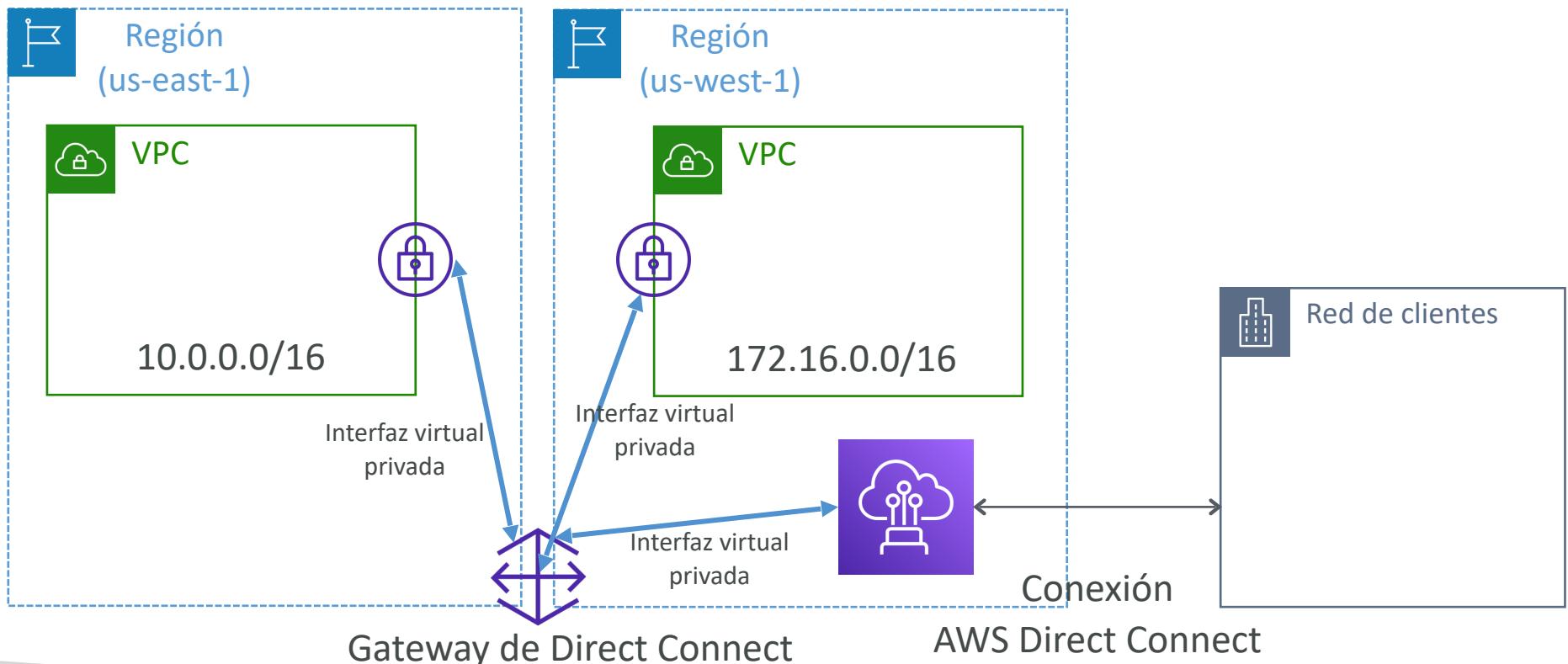
- Proporciona una conexión **privada** dedicada desde una red remota a tu VPC
- La conexión dedicada debe configurarse entre tu DC y las ubicaciones de AWS Direct Connect
- Necesitas configurar una Virtual Private Gateway en tu VPC
- Accede a recursos públicos (S3) y privados (EC2) en la misma conexión
- Casos de uso:
 - Aumentar el rendimiento del ancho de banda - trabajar con grandes conjuntos de datos - menor coste
 - Experiencia de red más consistente - aplicaciones que utilizan alimentación de datos en tiempo real
 - Entornos híbridos (on prem + cloud)
- Soporta tanto IPv4 como IPv6

Diagrama de Direct Connect



Gateway de Direct Connect

- Si quieres configurar una Direct Connect a una o más VPC en muchas regiones diferentes (misma cuenta), debes utilizar un Gateway de Direct Connect

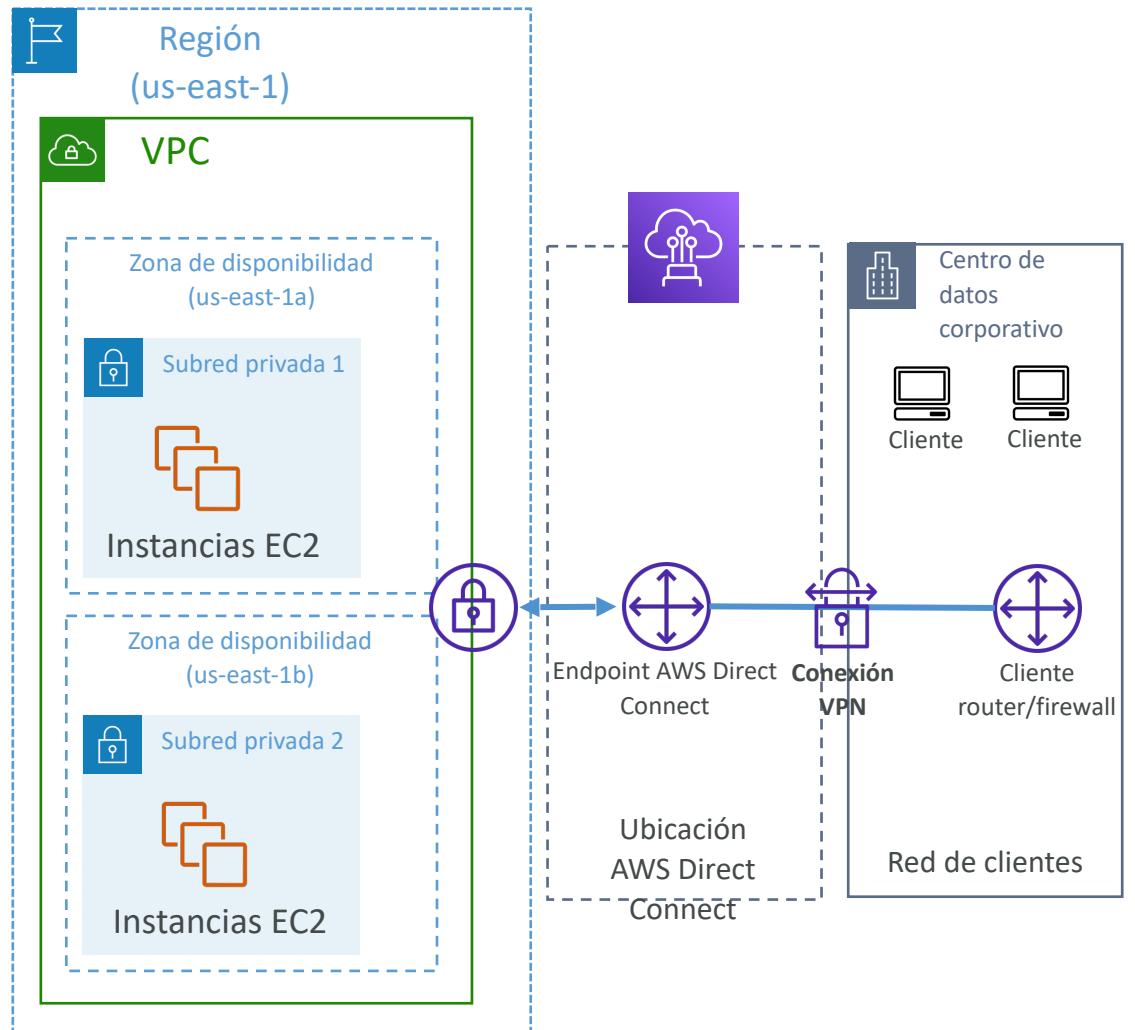


Direct Connect - Tipos de conexión

- **Conexiones dedicadas:** Capacidad de 1 Gbps, 10 Gbps y 100 Gbps
 - Puerto ethernet físico dedicado a un cliente
 - Primero se hace la petición a AWS y luego la completan los socios de AWS Direct Connect
- **Conexiones alojadas:** 50 Mbps, 500 Mbps, a 10 Gbps
 - Las peticiones de conexión se realizan a través de los socios de AWS Direct Connect
 - Se puede **añadir o eliminar capacidad bajo demanda**
 - 1, 2, 5, 10 Gbps disponibles en socios selectos de AWS Direct Connect
- Los plazos suelen ser superiores a 1 mes para establecer una nueva conexión

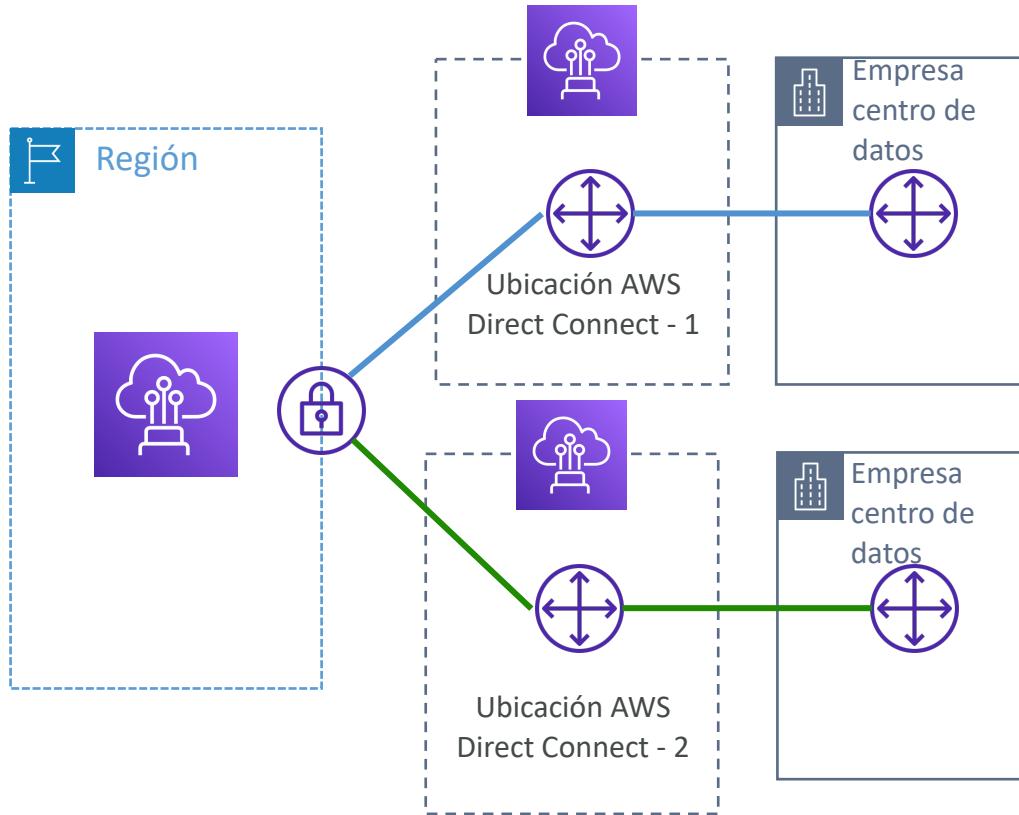
Direct Connect - Cifrado

- Los datos en tránsito no están cifrados, pero son privados
- AWS Direct Connect + VPN proporciona una conexión privada cifrada mediante IPsec
- Bueno para un nivel extra de seguridad, pero algo más complejo de implantar



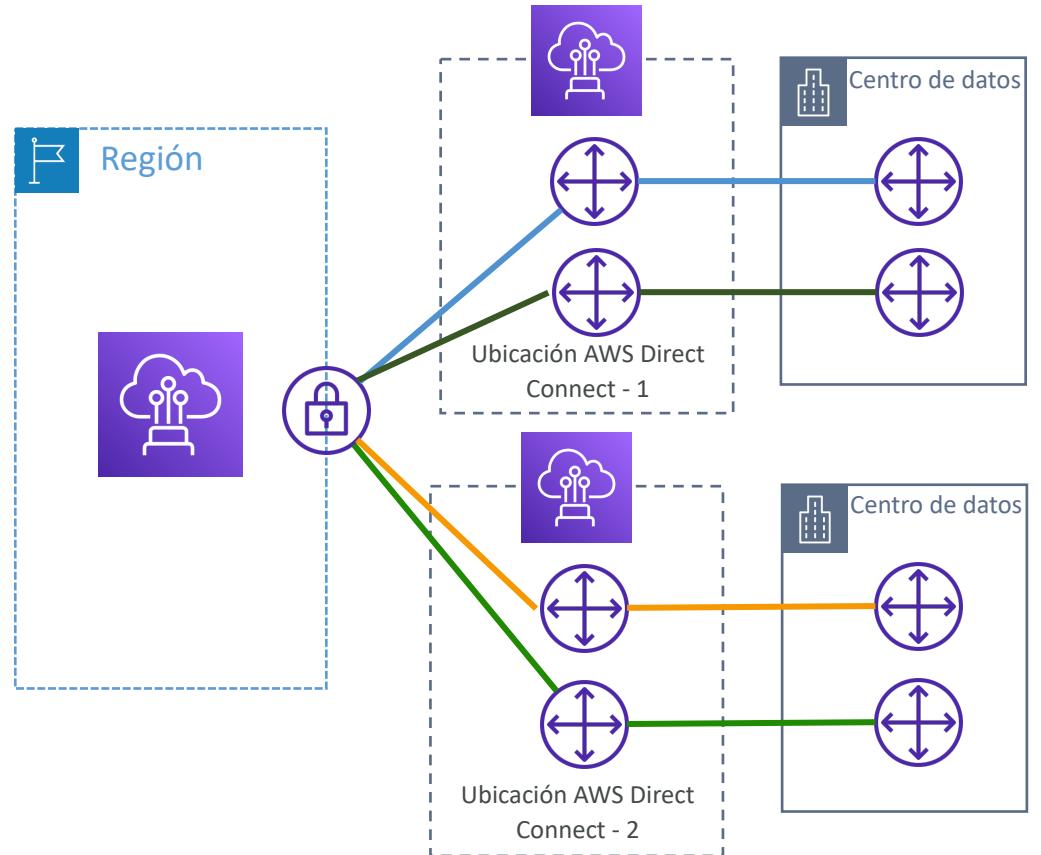
Direct Connect - Resiliencia

Alta resiliencia para cargas de trabajo críticas



Una conexión en varios lugares

Máxima resiliencia para cargas de trabajo críticas



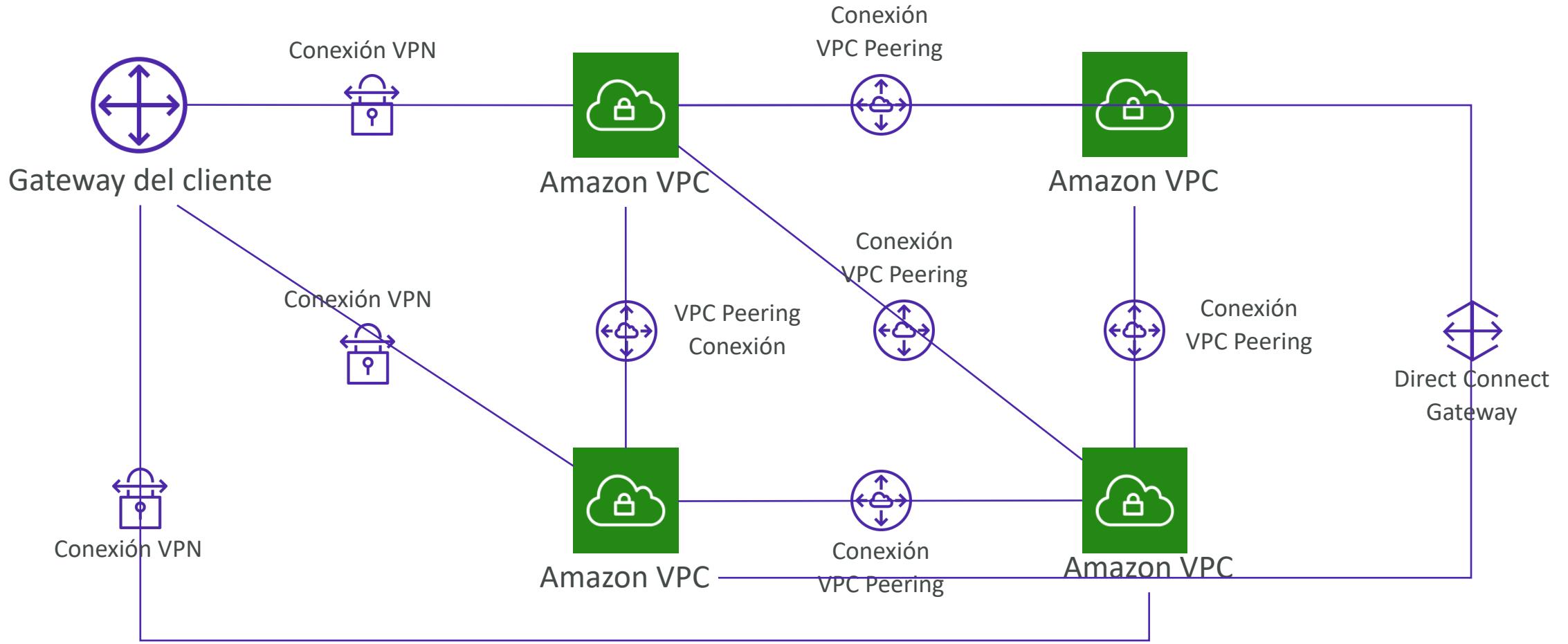
La máxima resiliencia se consigue mediante conexiones separadas que terminan en dispositivos separados en más de una ubicación.

Conexión VPN Site-to-Site (Sitio a Sitio) como copia de seguridad

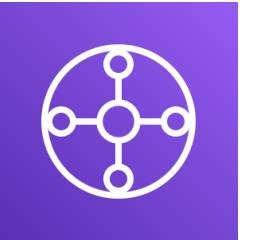
- En caso de que Direct Connect falle, puedes configurar una conexión de reserva de Direct Connect (cara), o una conexión VPN Site-to-Site (Sitio a Sitio)



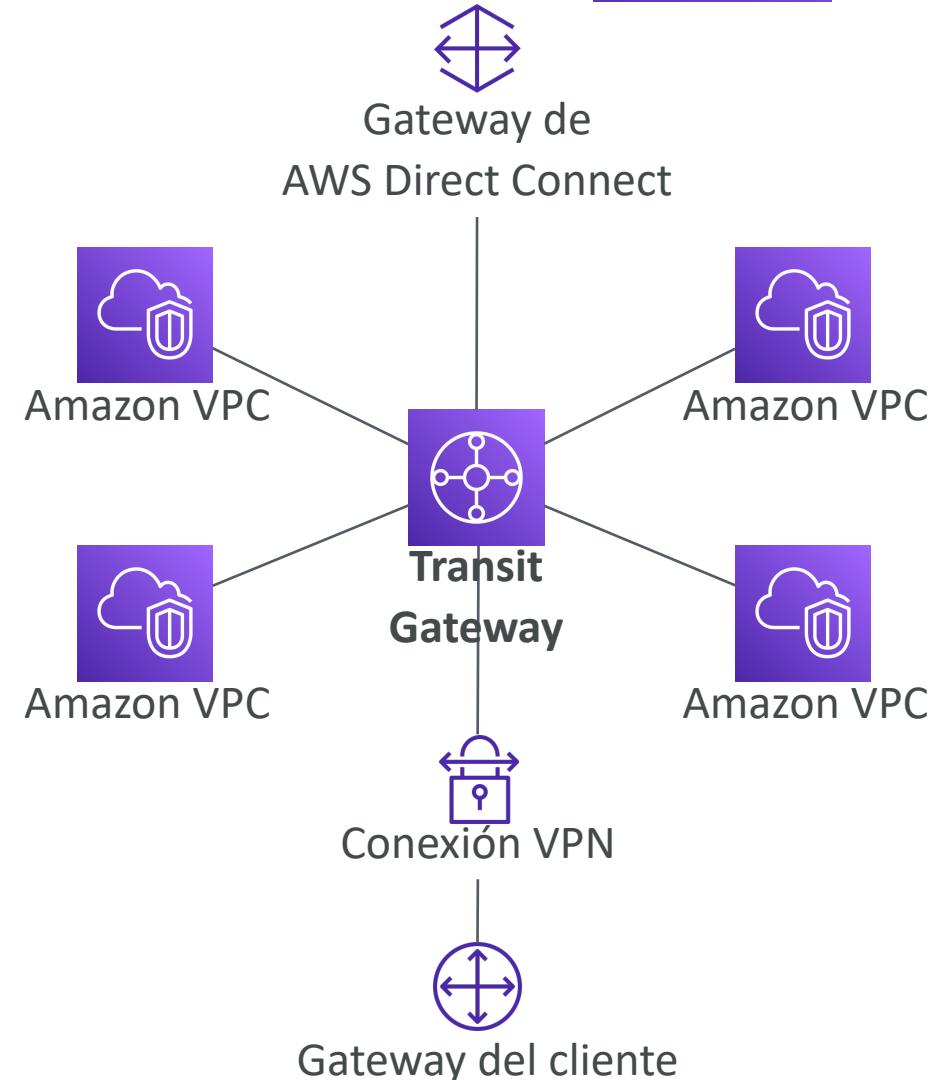
Las topologías de red pueden complicarse



Gateway de tránsito (Transit Gateway)

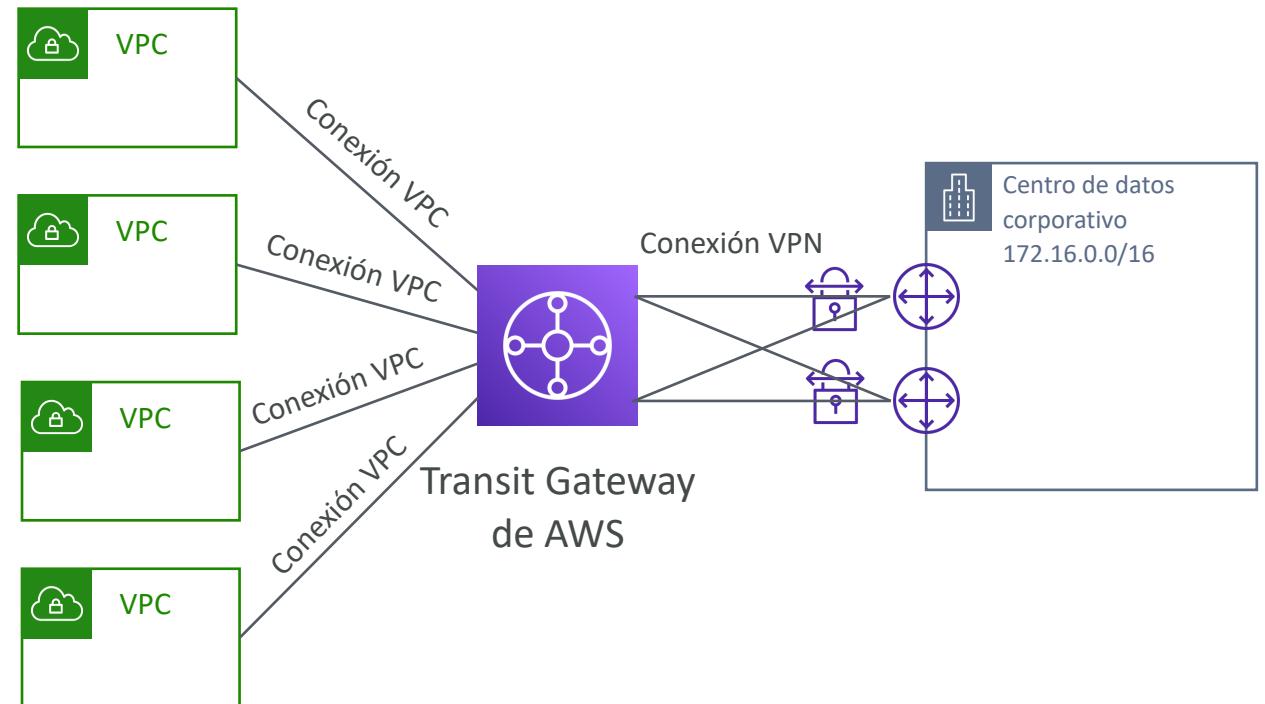


- Para tener peering transitivo entre miles de VPC y en las instalaciones, conexión hub-and-spoke (estrella)
- Recurso regional, puede funcionar en toda la región
- Comparte entre cuentas utilizando el Gestor de Acceso a Recursos (RAM)
- Puedes interconectar Gateways de tránsito entre regiones
- Tablas de rutas: limita qué VPC puede hablar con otra VPC
- Funciona con Direct Connect Gateway, conexiones VPN
- Soporta **IP Multicast** (no soportada por ningún otro servicio de AWS)



Gateway de tránsito:VPN Site-to-Site ECMP

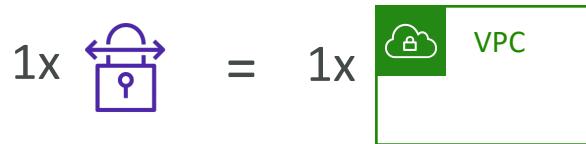
- **ECMP = Enrutamiento multirouteo de igual coste**
- Estrategia de enrutamiento que permite reenviar un paquete por múltiples rutas óptimas
- Caso práctico: crear varias conexiones VPN Site-to-Site (Sitio a Sitio) **para aumentar el ancho de banda de tu conexión a AWS**



Transit Gateway: rendimiento con ECMP



VPN a Virtual Private Gateway



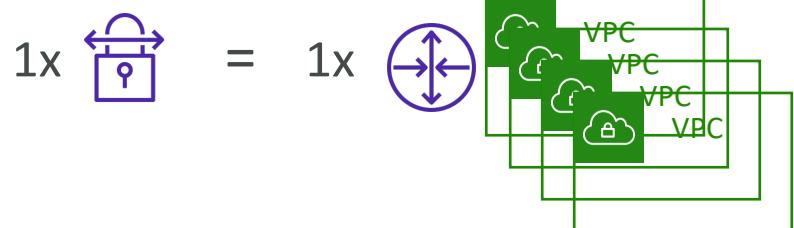
1x = 1.25 Gbps



Conexión VPN
(2 túneles)



VPN a Transit Gateway



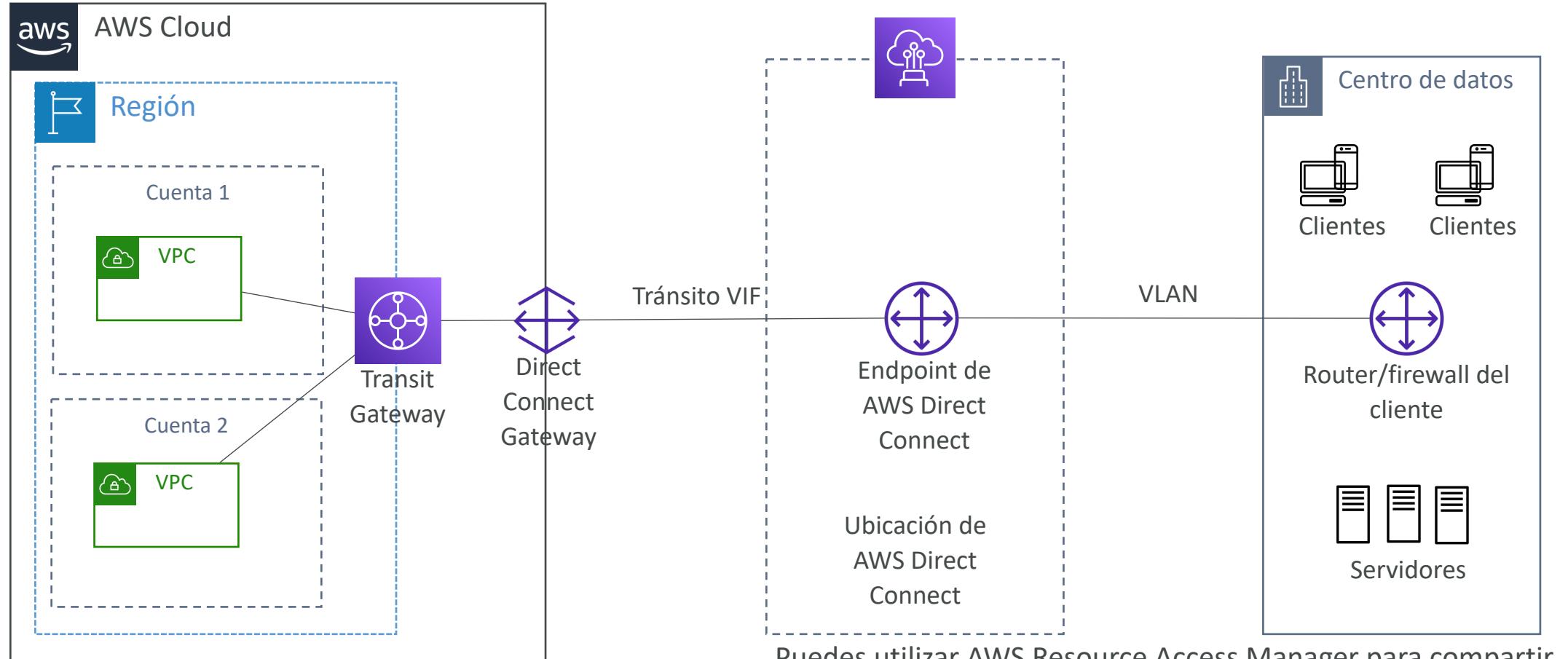
1x = 2,5 Gbps (ECMP) - 2 túneles utilizados

2x = 5.0 Gbps (ECMP)

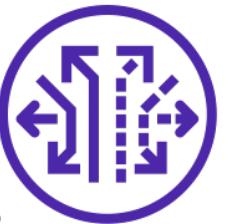
3x = 7.5 Gbps (ECMP)

+\$\$ por GB de TGW
procesados

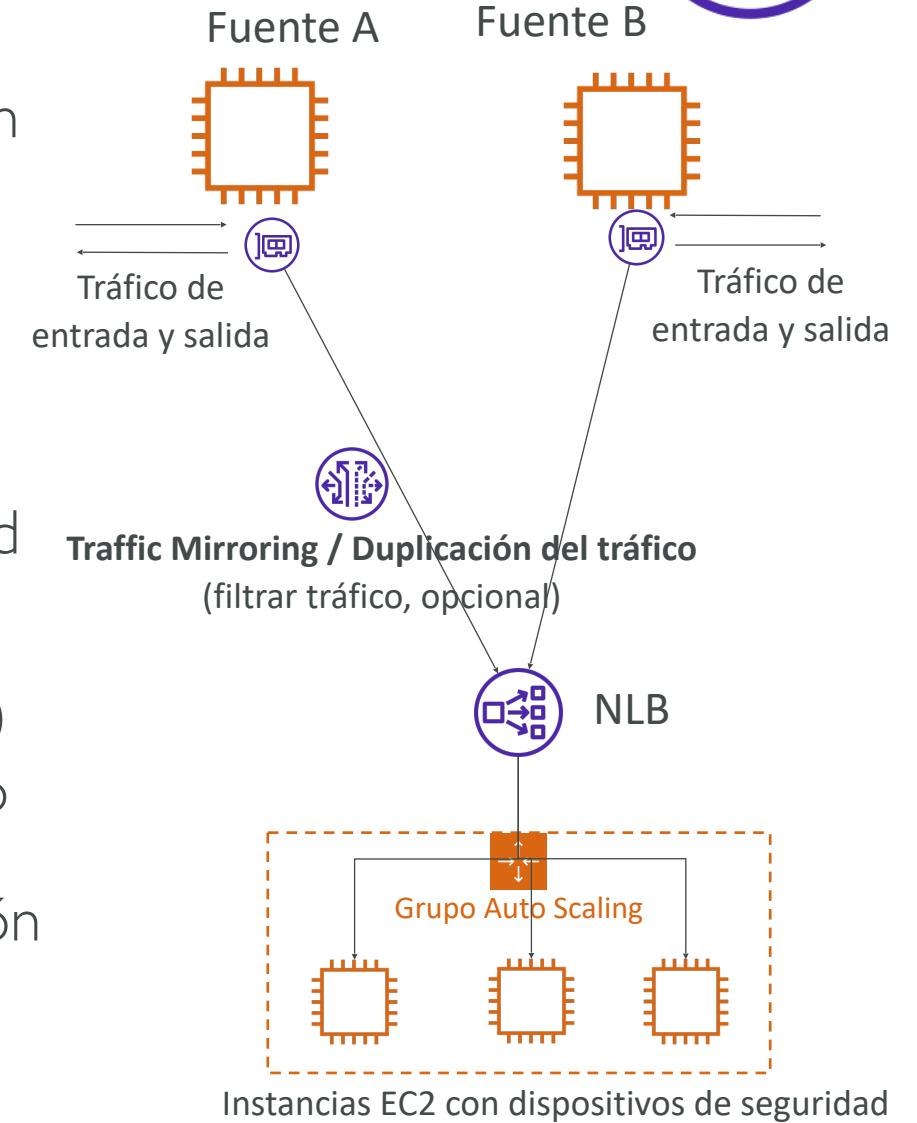
Transit Gateway - Comparte Direct Connect entre varias cuentas



VPC - Traffic Mirroring



- Te permite capturar e inspeccionar el tráfico de red en tu VPC
- Enrutar el tráfico a los dispositivos de seguridad que gestionas
- Captura el tráfico
 - **Desde (Fuente)** - ENIs
 - **Hacia (Objetivos)** - una ENI o un Network Load Balancer
- Captura todos los paquetes o captura los paquetes que te interesen (opcionalmente, trunca los paquetes)
- El origen y el destino pueden estar en la misma VPC o en distintas VPC (VPC Peering)
- Casos de uso: inspección de contenidos, monitorización de amenazas, resolución de problemas, ...

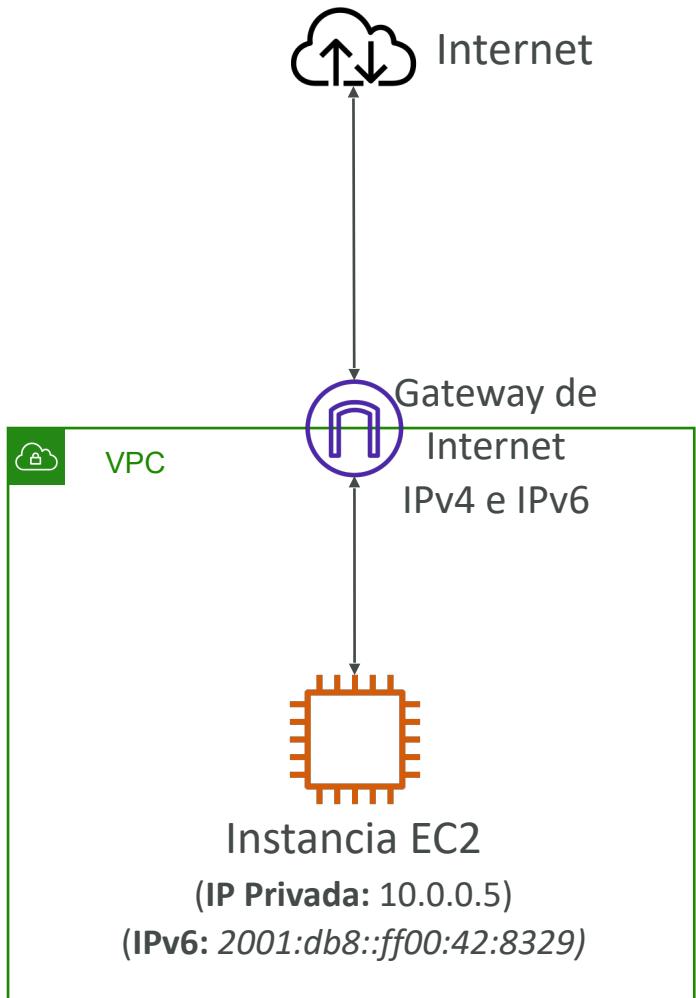


¿Qué es IPv6?

- IPv4 se diseñó para proporcionar 4.300 millones de direcciones (se agotarán pronto)
- IPv6 es el sucesor de IPv4
- IPv6 está diseñado para proporcionar **3.4×10^{38}** direcciones IP únicas
- Cada dirección IPv6 es pública y enrutable por Internet (no hay rango privado)
- Formato → x.x.x.x.x.x.x (x es hexadecimal, el rango puede ser de 0000 a ffff)
- Ejemplos:
 - 2001:db8:3333:4444:5555:6666:7777:8888
 - 2001:db8:3333:4444:cccc:dddd:eeee:ffff
 - :: → los 8 segmentos son cero
 - 2001:db8:: → los 6 últimos segmentos son cero
 - ::1234:5678 → los 6 primeros segmentos son cero
 - 2001:db8::1234:5678 → los 4 segmentos centrales son cero

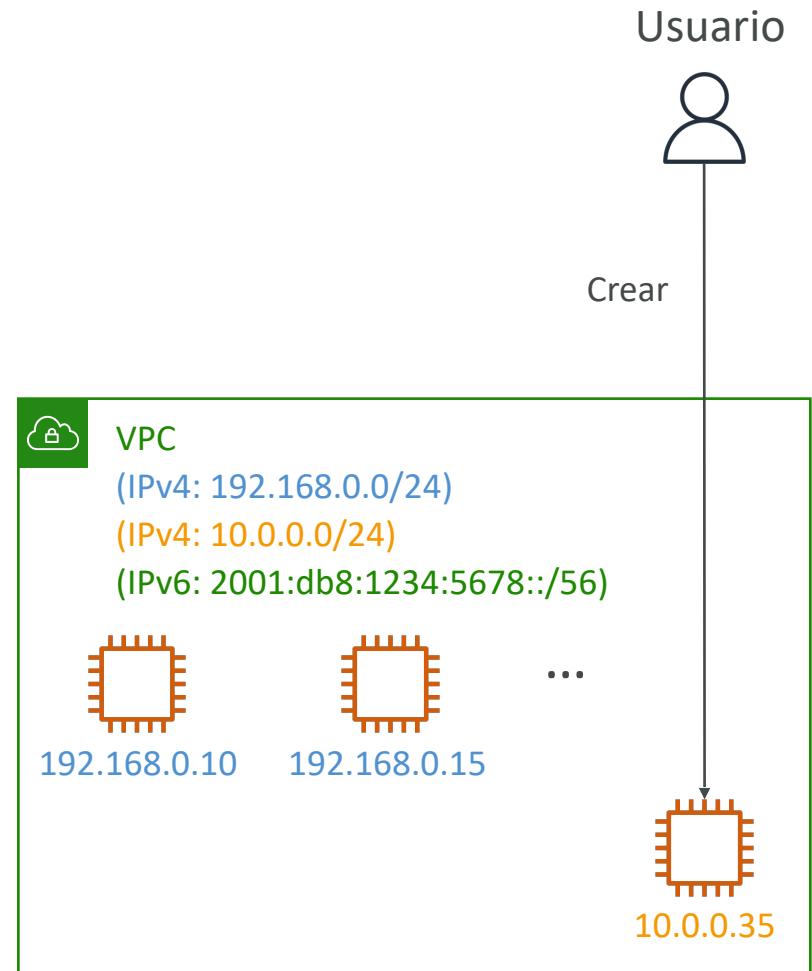
IPv6 en VPC

- **No se puede deshabilitar IPv4 para tu VPC y subredes**
- Puedes habilitar IPv6 (son direcciones IP públicas) para funcionar en modo dual stack
- Tus instancias EC2 tendrán al menos una IPv4 interna privada y una IPv6 pública
- Pueden comunicarse utilizando IPv4 o IPv6 a Internet a través de un Gateway de Internet

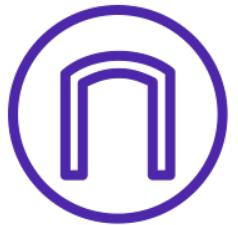


Solución de problemas de IPv6

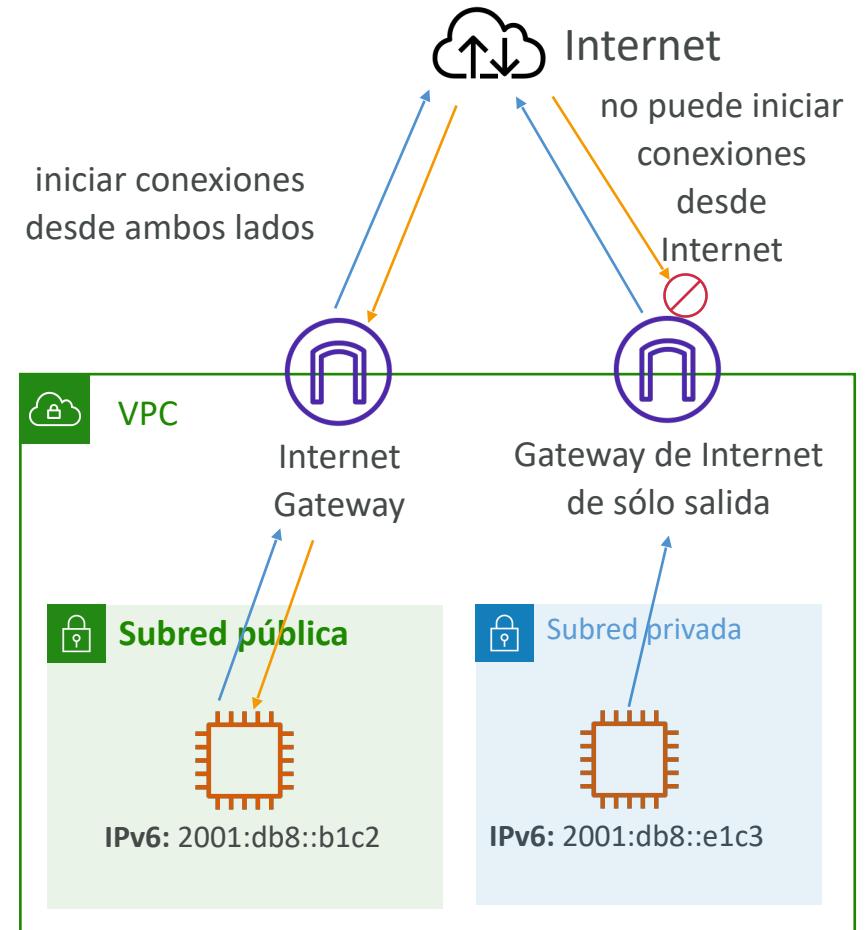
- **No se puede deshabilitar IPv4 para tu VPC y subredes**
- Por tanto, si no puedes lanzar una instancia EC2 en tu subred
 - no es porque no pueda adquirir una IPv6 (el espacio es muy grande)
 - Es porque no hay IPv4 disponibles en tu subred
- **Solución:** crea un nuevo CIDR IPv4 en tu subred



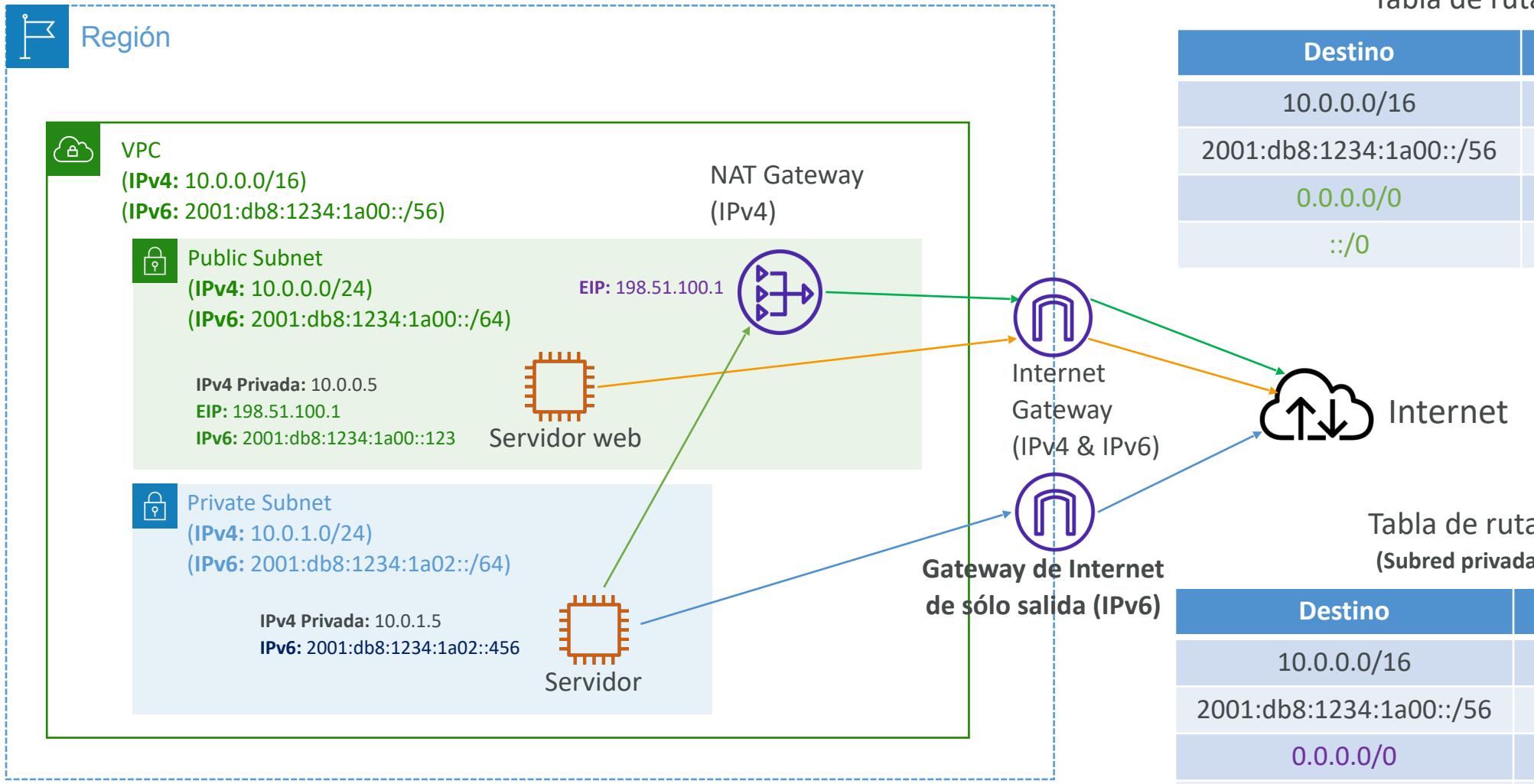
Gateway de Internet sólo de salida



- **Se utiliza sólo para IPv6**
- (similar a Los Gateways NAT pero para IPv6)
- Permite a las instancias de tu VPC conexiones salientes a través de IPv6, al tiempo que impide que Internet inicie una conexión IPv6 con tus instancias
- **Debes actualizar las Tablas de Rutas**



Enrutamiento IPv6



Resumen de la sección VPC (I/3)

- **CIDR** - Rango IP
- **VPC** - Virtual Private Cloud => definimos una lista de CIDR IPv4 & IPv6
- **Subredes** - vinculadas a una AZ, definimos un CIDR
- **Gateway de Internet** - a nivel de VPC, proporciona acceso a Internet IPv4 e IPv6
- **Tablas de rutas** - deben editarse para añadir rutas desde subredes al IGW,VPC Peering connection,VPC endpoint, ...
- **Bastion Host** - instancia EC2 pública a la que acceder mediante SSH, que tiene conectividad SSH a instancias EC2 en subredes privadas.
- **Instancias NAT** - da acceso a Internet a instancias EC2 en subredes privadas. Antiguo, debe configurarse en una subred pública, desactivar la bandera de comprobación Origen / Destino
- **Gateways NAT** - gestionados por AWS, proporcionan acceso escalable a Internet a instancias EC2 privadas, sólo IPv4
- **DNS privado + Route 53** - habilita la Resolución DNS + Nombres de host DNS (VPC)

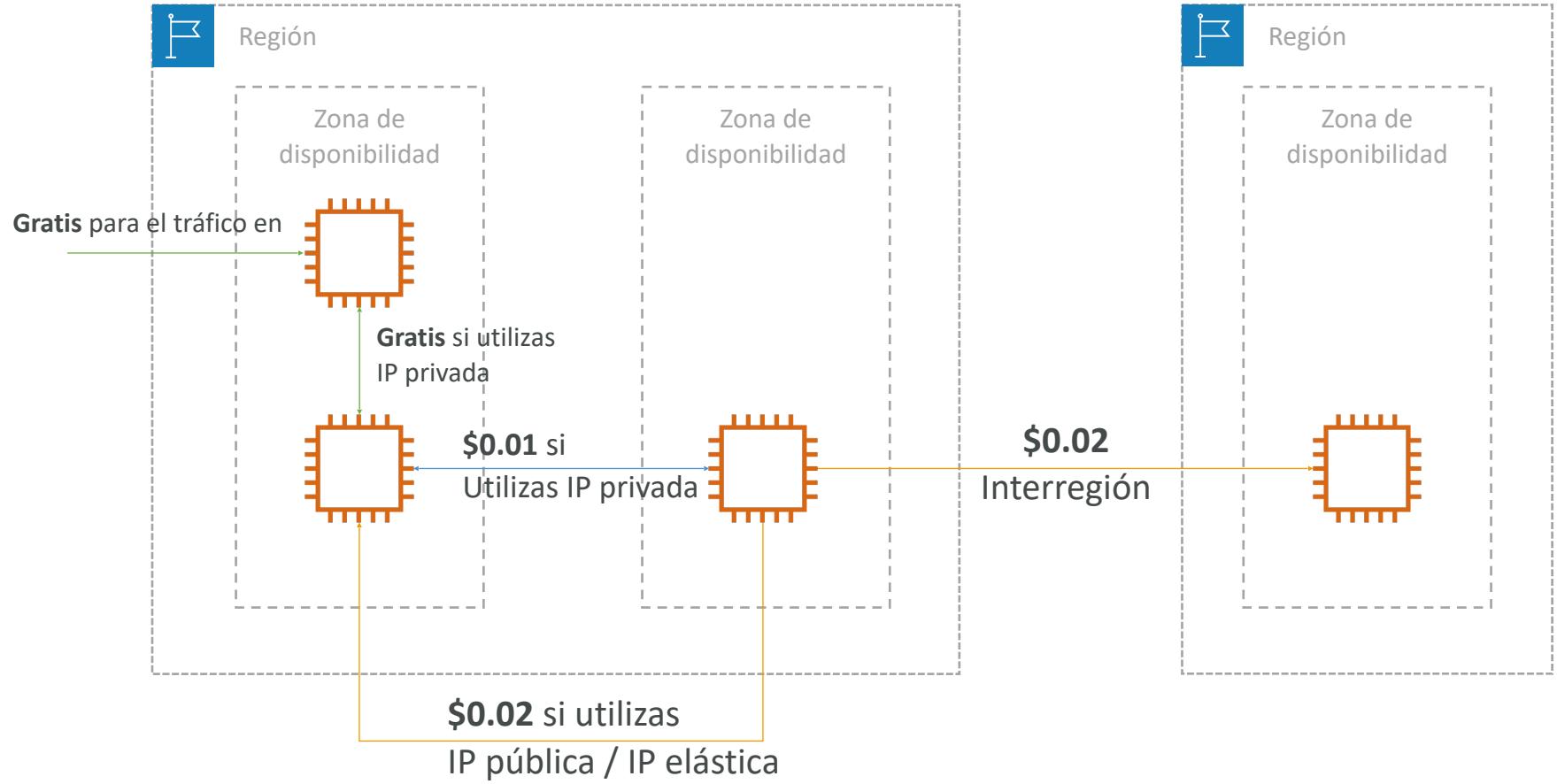
Resumen de la sección VPC (2/3)

- **NACL** - sin estado, reglas de subred para entrantes y salientes, no olvides los Puertos Efímeros
- **Grupos de Seguridad** - con estado, operan a nivel de instancia EC2
- **Analizador de alcanzabilidad** - realiza pruebas de conectividad de red entre recursos AWS
- **VPC Peering** - conecta dos VPC con CIDR no solapados, no transitivos
- **VPC Endpoints** - proporcionan acceso privado a los servicios de AWS (S3, DynamoDB, CloudFormation, SSM) dentro de una VPC
- **Logs de flujo VPC** - puede configurarse a nivel de VPC / subred / ENI, para tráfico de ACEPTACION y RECHAZO, ayuda a identificar ataques, analizar utilizando Athena o CloudWatch Logs Insights
- **VPN Site-to-Site** - configura una Customer Gateway en DC, una Virtual Private Gateway en VPC y una VPN Site-to-Site a través de Internet pública.
- **AWS VPN CloudHub** - modelo VPN hub-and-spoke para conectar tus sitios

Resumen de la sección VPC (3/3)

- **Direct Connect** - configura una Virtual Private Gateway en la VPC y establece una conexión privada directa con una ubicación de AWS Direct Connect
- **Direct Connect Gateway** - configura una Direct Connect a muchas VPC en diferentes regiones de AWS
- **Servicios AWS PrivateLink / VPC Endpoint:**
 - Conecta servicios de forma privada desde tu VPC de servicios a la VPC de clientes
 - No necesita VPC Peering, Internet pública, Gateways NAT, Tablas de Ruta
 - Debe utilizarse con Network Load Balancer y ENI
- **ClassicLink** - conecta instancias EC2-Classic EC2 de forma privada a tu VPC
- **Transit Gateway** - conexiones peering transitivas para VPC, VPN y DX
- **Traffic Mirroring / Duplicación del tráfico** - copia el tráfico de red de las ENI para su posterior análisis
- **Gateway de Internet sólo de salida** - como un NAT Gateways, pero para IPv6

Costes de red en AWS por GB - Simplificado

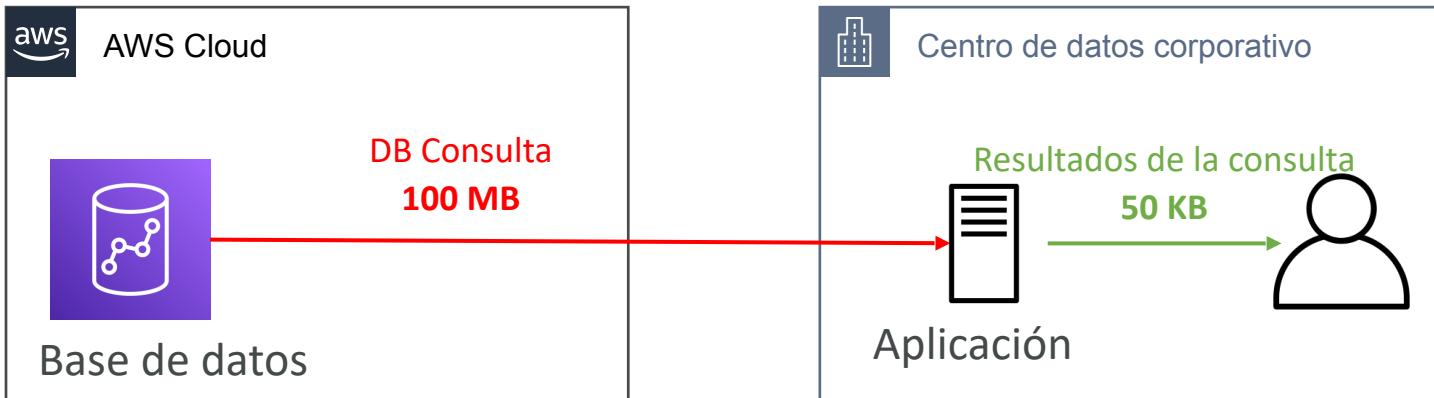


- Utiliza IP Privada en lugar de IP Pública para un buen ahorro y un mejor rendimiento de la red
- Utiliza la misma AZ para un ahorro máximo (a costa de una alta disponibilidad))

Minimizar el coste de la red de tráfico de salida

- Tráfico de salida: tráfico saliente (de AWS al exterior)
- Tráfico de entrada: tráfico entrante - del exterior a AWS (normalmente gratis)
- Intenta mantener la mayor cantidad de tráfico de Internet dentro de AWS para minimizar los costes
- **Las ubicaciones de Direct Connect situadas en la misma región de AWS tienen un coste inferior para la red de salida**

El coste de salida es elevado



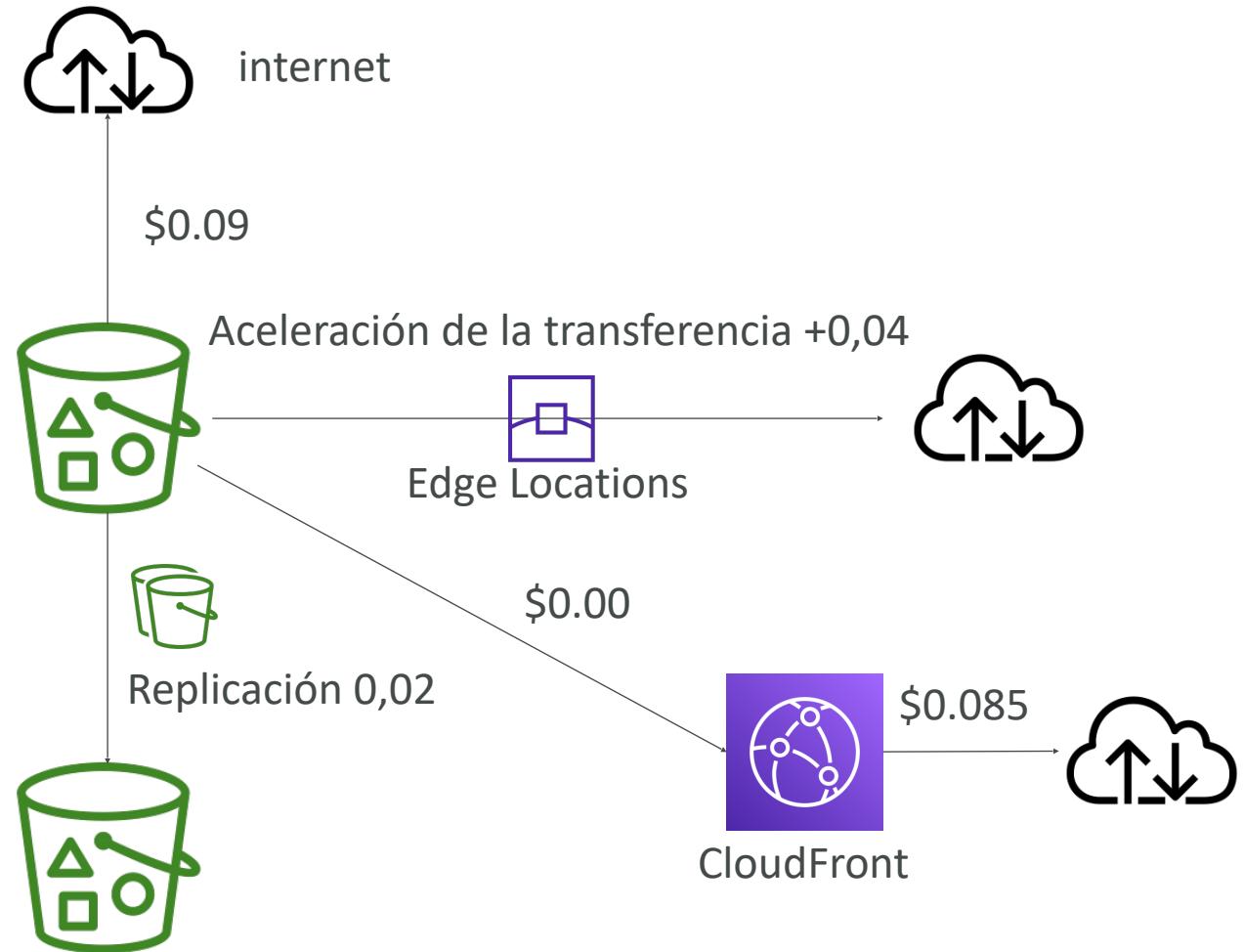
Se minimiza el coste de salida



Precios de transferencia de datos S3

Análisis para EE.UU.

- **Entrada a S3:** gratis
- **S3 a Internet:** 0,09 \$ por GB
- **Aceleración de transferencias de S3:**
 - Tiempos de transferencia más rápidos (entre un 50 y un 500% mejores)
 - Coste adicional sobre el Precio de Transferencia de Datos: +0,04 a 0,08 \$ por GB
- **S3 a CloudFront:** 0,00 \$ por GB
- **CloudFront a Internet:** 0,085 \$ por GB (ligeramente más barato que S3)
 - Capacidad de almacenamiento en caché (menor latencia)
 - Reduce los costes asociados al precio de las peticiones de S3 (7 veces más barato con CloudFront)
- **Replicación entre regiones de S3:** 0,02 \$ por GB



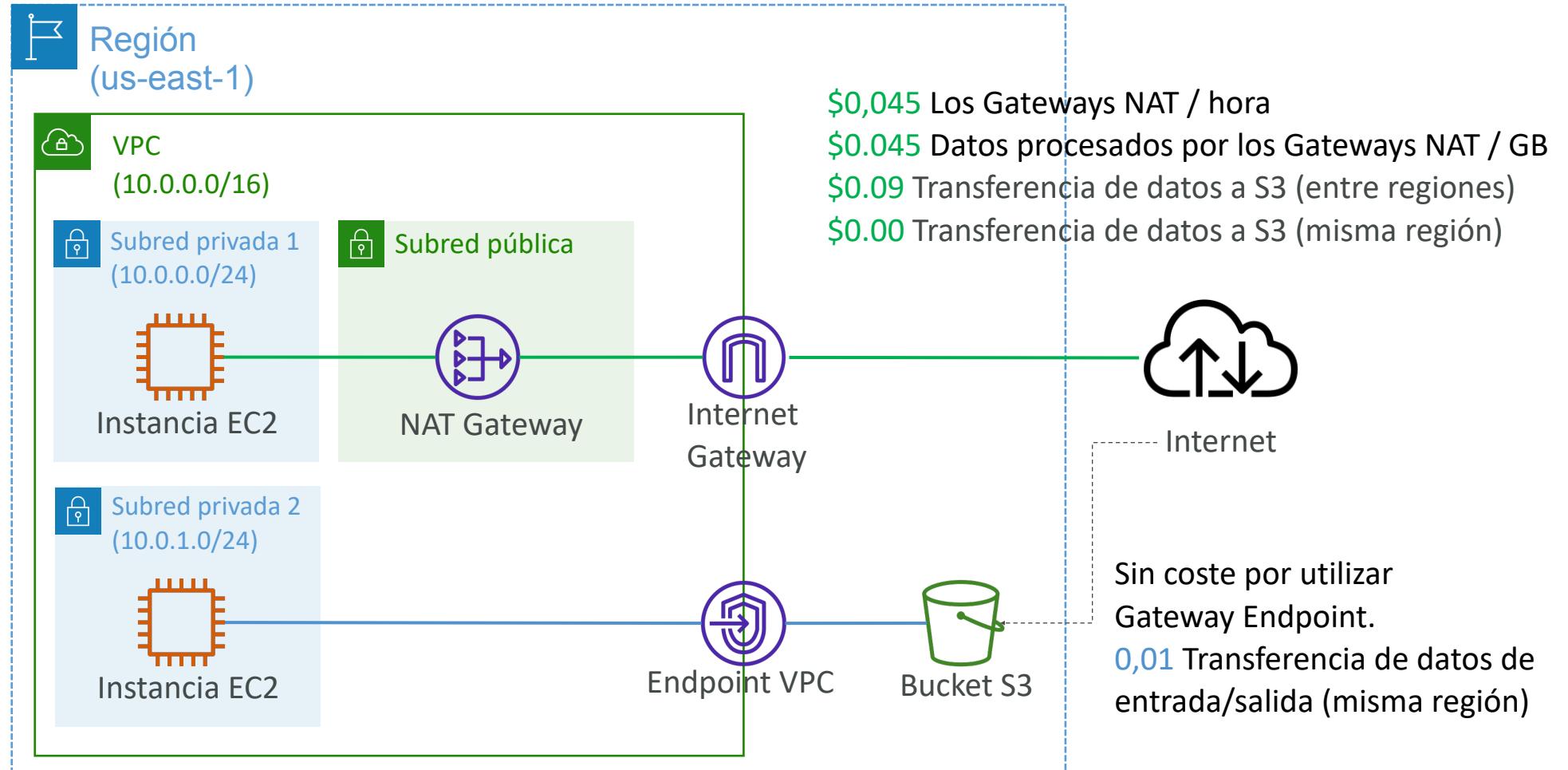
Precios Gateways NAT vs Gateway VPC Endpoint

Tabla de rutas de la subred 1

Destino	Objetivo
10.0.0.0/16	Local
0.0.0.0/0	igw-id

Tabla de rutas de la subred 2

Destino	Objetivo
10.0.0.0/16	Local
pl-id for Amazon S3	vpce-id

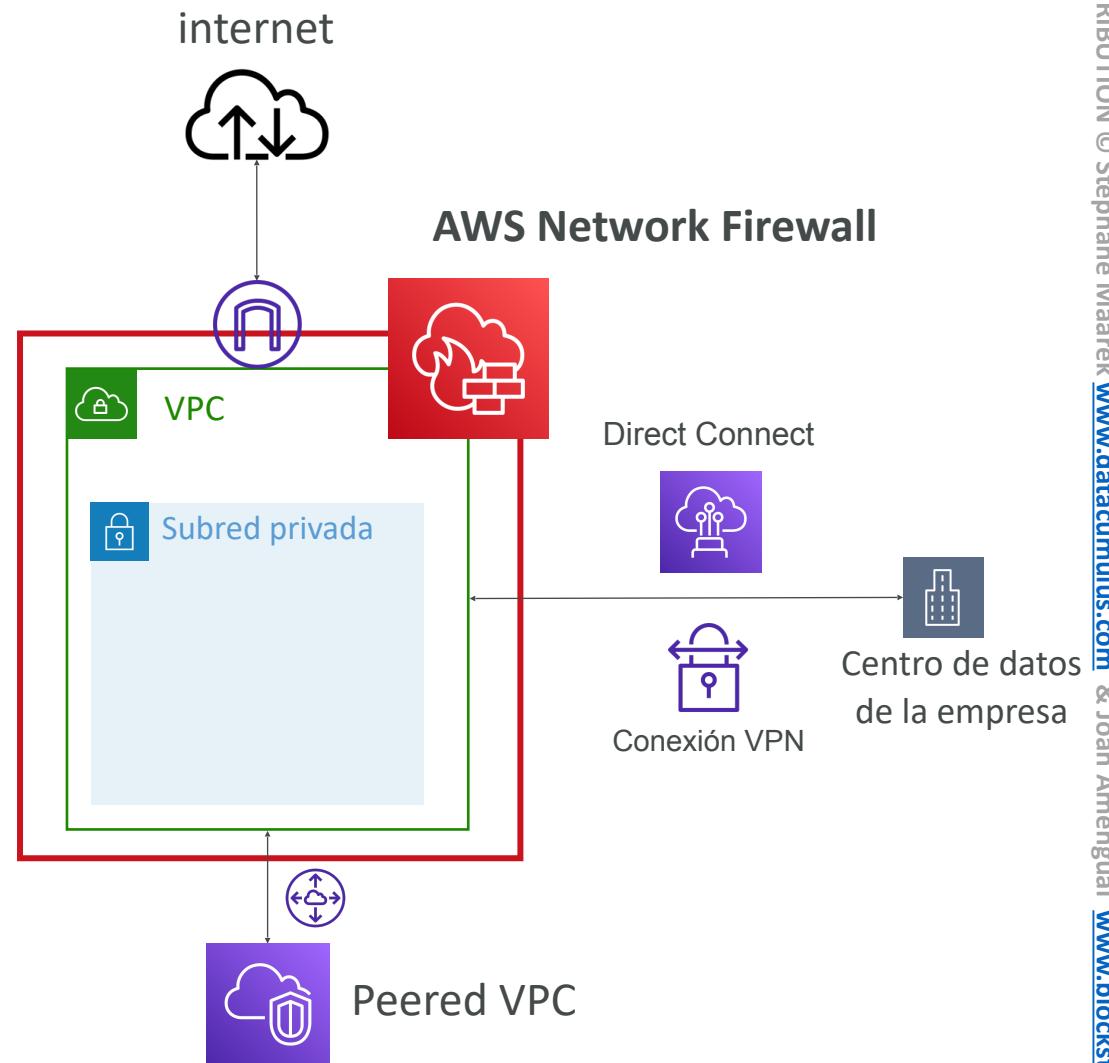


Protección de redes en AWS

- Para proteger la red en AWS, hemos visto
 - Listas de control de acceso a la red (NACL)
 - Grupos de seguridad Amazon VPC
 - AWS WAF (protección contra peticiones maliciosas)
 - AWS Shield y AWS Shield Avanzado
 - AWS Firewall Manager (para gestionarlos entre cuentas)
- Pero ¿y si queremos proteger de forma sofisticada toda nuestra VPC?

Firewall de red de AWS

- Protege toda tu Amazon VPC
- Protección de Capa 3 a Capa 7
- En cualquier dirección, puedes inspeccionar
 - Tráfico de VPC a VPC
 - Saliente a Internet
 - Entrante desde Internet
 - Hacia / desde Direct Connect y VPN Site-to-Site (Sitio a Sitio)
- Internamente, el AWS Network Firewall utiliza el AWS Gateway Load Balancer
- Las reglas se pueden gestionar de forma centralizada entre cuentas mediante AWS Firewall Manager para aplicarlas a muchas VPCs





Firewall de red - Controles al detalle

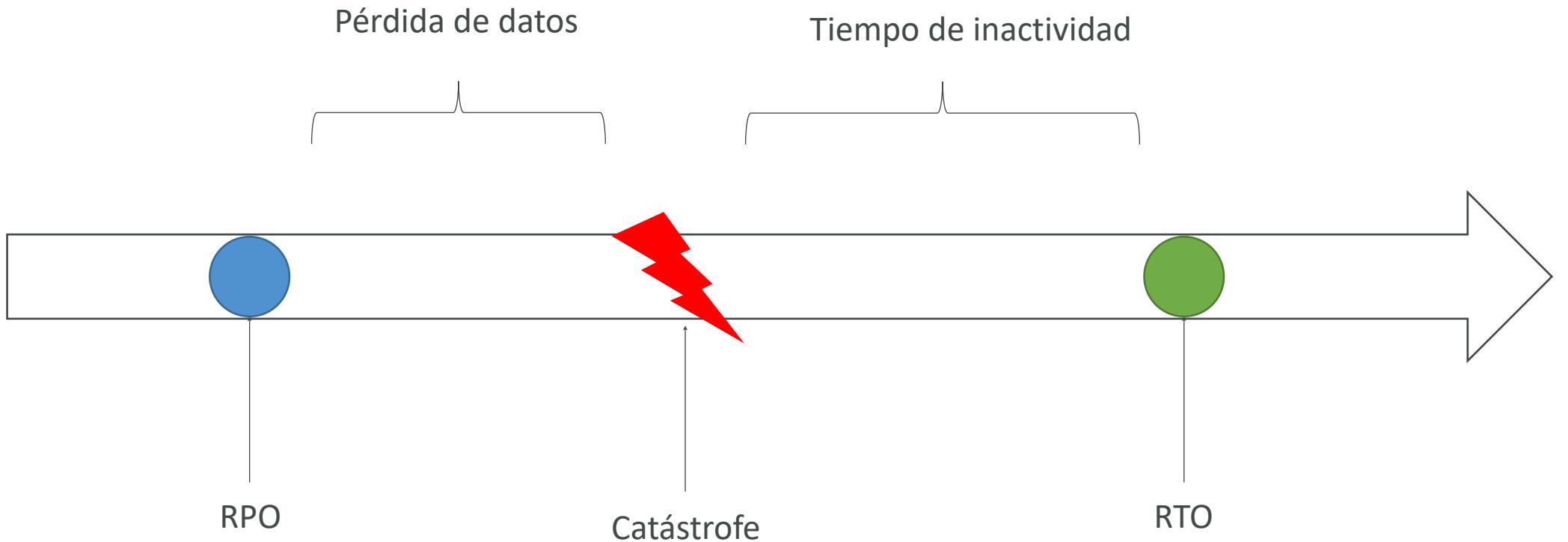
- Soporta 1000s de reglas
 - IP y puerto - ejemplo: filtrado de 10.000s de IPs
 - Protocolo - ejemplo: bloquear el protocolo SMB para comunicaciones salientes
 - Grupos de reglas de listas de dominios Stateful: sólo permite el tráfico saliente a *.mycorp.com o a repositorios de software de terceros
 - Concordancia general de patrones mediante regex
- **Filtrado de tráfico: Permitir, descartar o alertar del tráfico que coincide con las reglas**
- **Inspección de flujo activo** para proteger contra amenazas de red con funciones de prevención de intrusiones (como Gateway Load Balancer, pero todo gestionado por AWS)
- Envía logs de las coincidencias de las reglas a Amazon S3, CloudWatch Logs, Kinesis Data Firehose

Recuperación ante desastres y migraciones

Visión general de la recuperación tras catástrofes

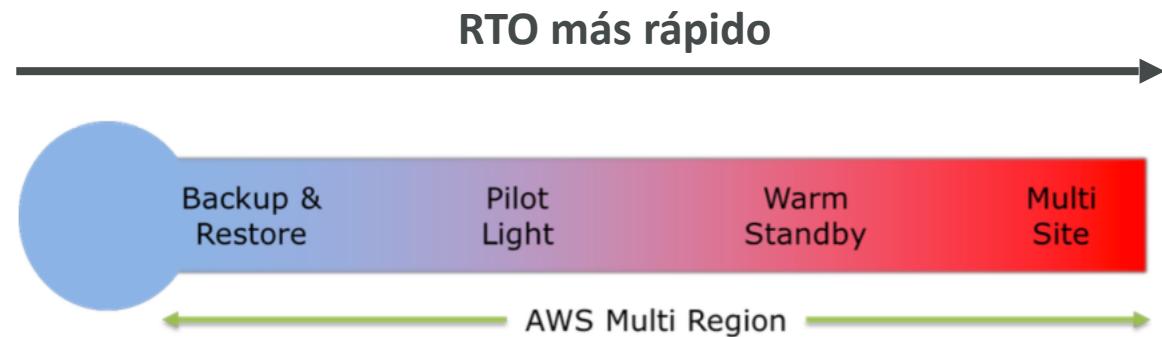
- Cualquier acontecimiento que tenga un impacto negativo en la continuidad de la actividad o en las finanzas de una empresa es un desastre
- La recuperación ante catástrofes (RD) consiste en prepararse y recuperarse de una catástrofe
- ¿Qué tipo de recuperación ante desastres?
 - En las instalaciones => En las instalaciones: DR tradicional, y muy cara
 - En las instalaciones => Cloud de AWS: recuperación híbrida
 - Región A de AWS Cloud => Región B de AWS Cloud
- Necesidad de definir dos términos
 - RPO: Objetivo de Punto de Recuperación
 - RTO: Objetivo de Tiempo de Recuperación

RPO y RTO

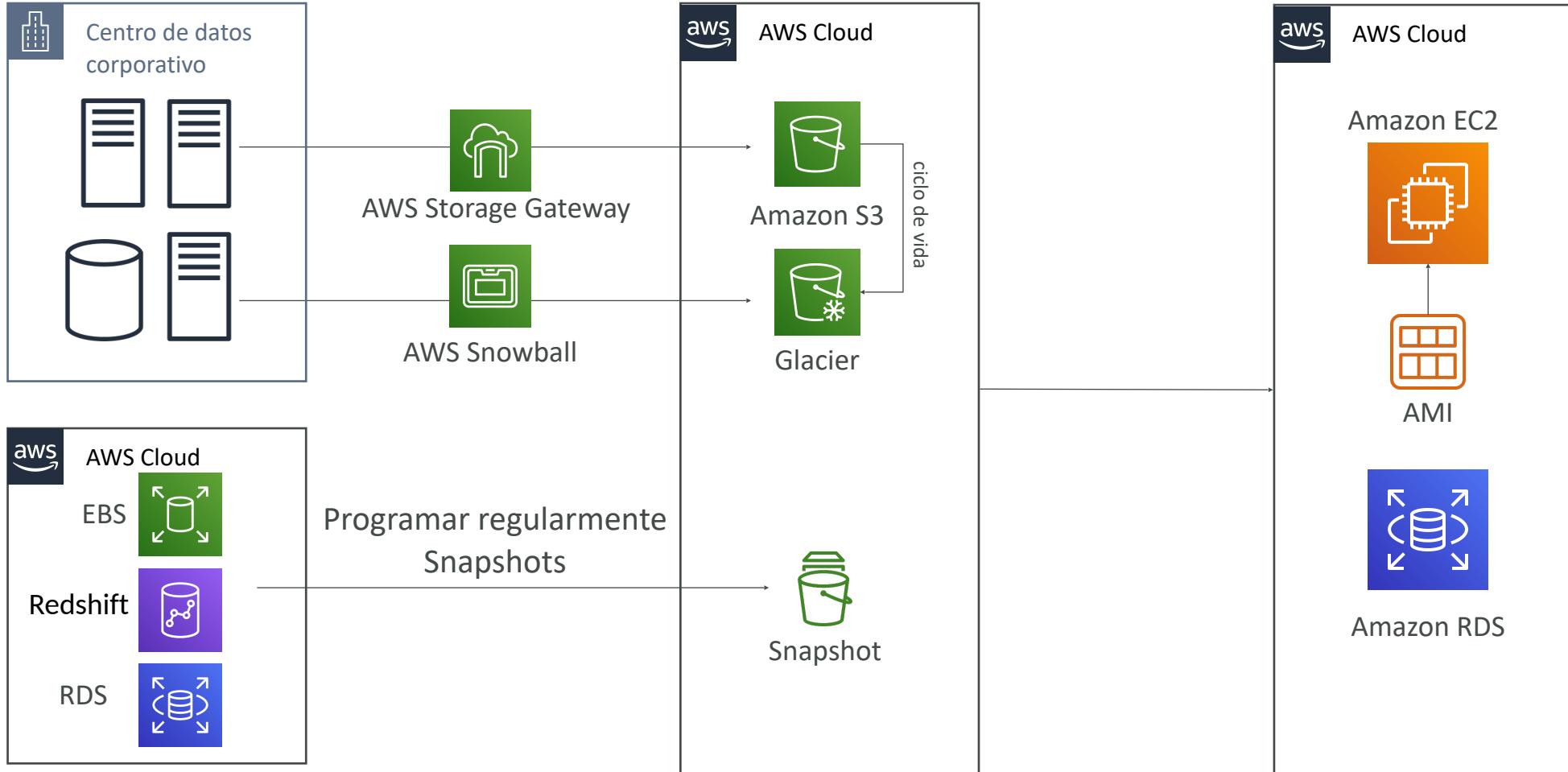


Estrategias de recuperación en caso de catástrofe

- Copia de seguridad y restauración
- Luz piloto
- Espera caliente
- Enfoque Hot Site / Multi Site

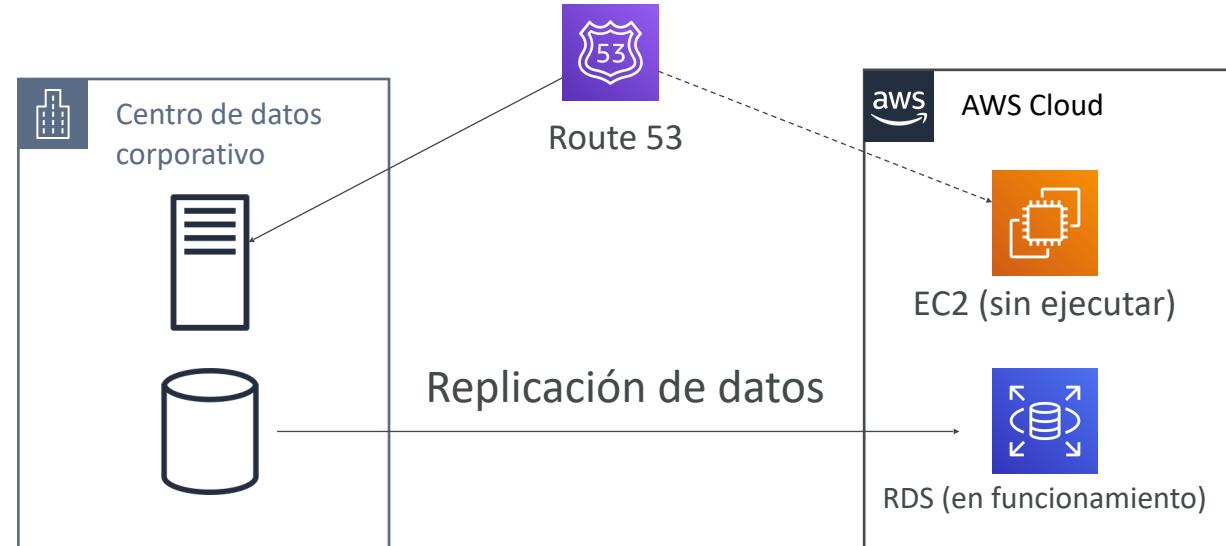


Copia de seguridad y restauración (RPO alto)



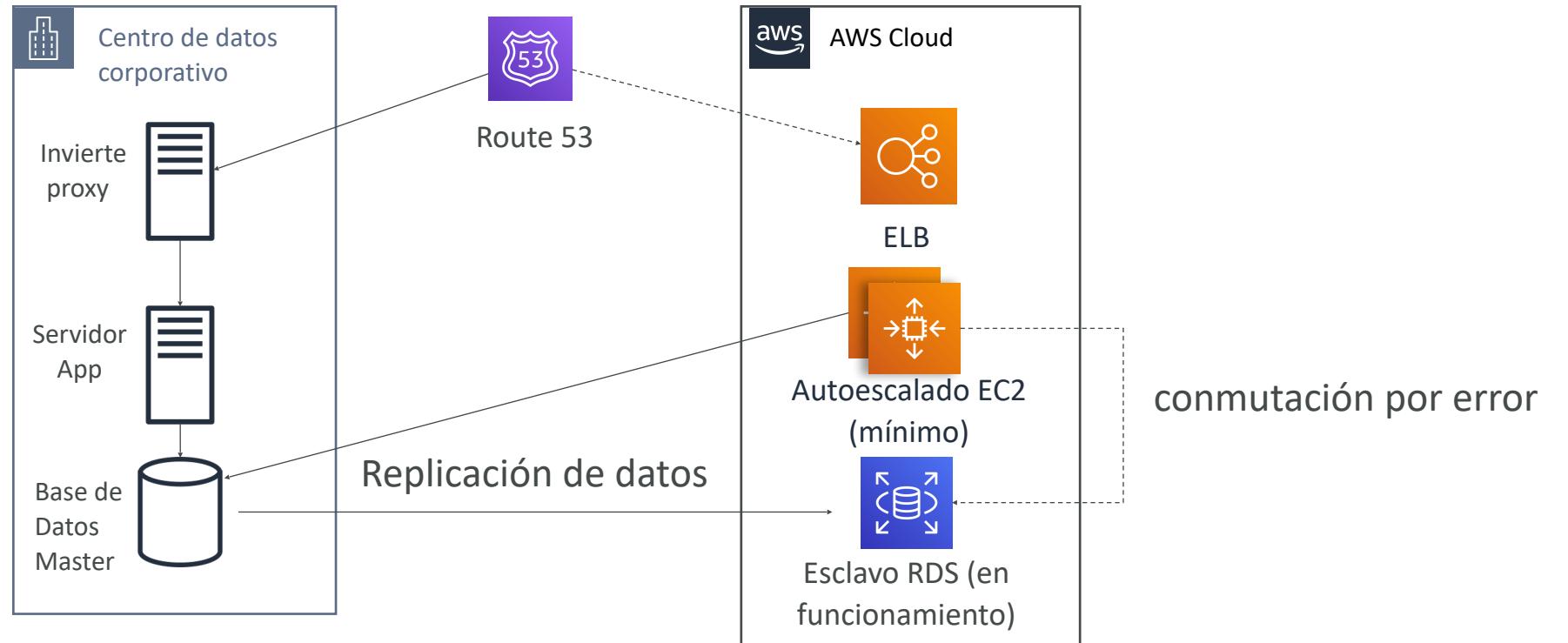
Recuperación en caso de catástrofe - Pilot Light

- Una pequeña versión de la aplicación se ejecuta siempre en el Cloud
- Útil para el núcleo crítico (piloto)
- Muy similar a Copia de Seguridad y Restauración (Backup & Restore)
- Más rápido que Copia de Seguridad y Restauración, ya que los sistemas críticos ya están en marcha



Espera caliente (Warm Standby)

- El sistema completo está en funcionamiento, pero con un tamaño mínimo
- En caso de desastre, podemos escalar a la carga de producción



Enfoque multisitio/en caliente

- RTO muy bajo (minutos o segundos) - muy caro
- La escala de producción completa se ejecuta en AWS y en las instalaciones



Todas las regiones AWS (Multi)



Consejos para la recuperación en caso de catástrofe

• Copias de seguridad

- Los Snapshots de EBS, copias de seguridad automáticas de RDS / Snapshots, etc...
- Envíos regulares a S3 / S3 IA / Glacier, Política de ciclo de vida, Replicación entre regiones
- Desde las instalaciones: Snowball o Storage Gateway

• Alta disponibilidad

- Utiliza Route53 para migrar DNS de una región a otra
- RDS Multi-AZ, ElastiCache Multi-AZ, EFS, S3
- VPN Site-to-Site como recuperación de Direct Connect

• Replicación

- Replicación RDS (entre regiones), AWS Aurora + bases de datos globales
- Replicación de bases de datos desde las instalaciones a RDS
- Gateway de almacenamiento

• Automatización

- CloudFormation / Elastic Beanstalk para volver a crear un entorno completamente nuevo
- Recuperar / Reiniciar instancias EC2 con CloudWatch si fallan las alarmas
- Funciones AWS Lambda para automatizaciones personalizadas

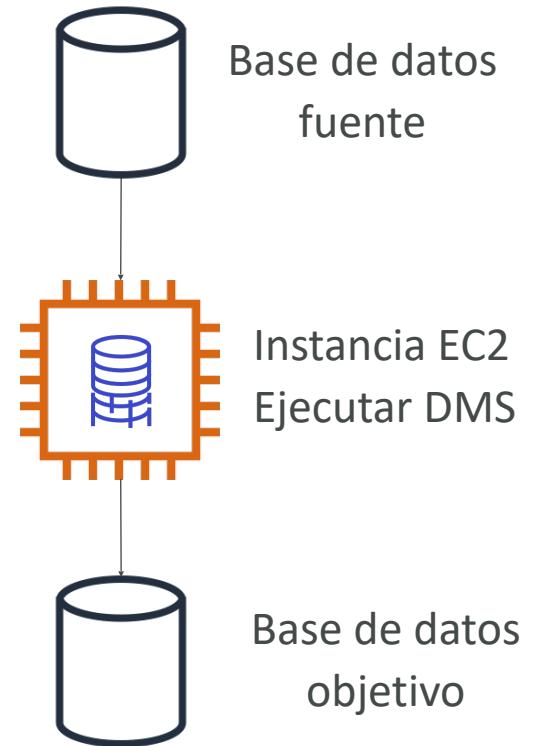
• Caos

- Netflix tiene un "ejército de simios" que termina EC2 aleatoriamente



DMS - Servicio de Migración de Bases de Datos

- Migra bases de datos a AWS de forma rápida y segura, resiliente y autorreparable
- La base de datos de origen sigue disponible durante la migración
- Soporta:
 - Migraciones homogéneas: por ejemplo, de Oracle a Oracle
 - Migraciones heterogéneas: ex Microsoft SQL Server a Aurora
- Replicación continua de datos mediante CDC
- Debes crear una instancia EC2 para realizar las tareas de replicación



Fuentes y objetivos de DMS

FUENTES:

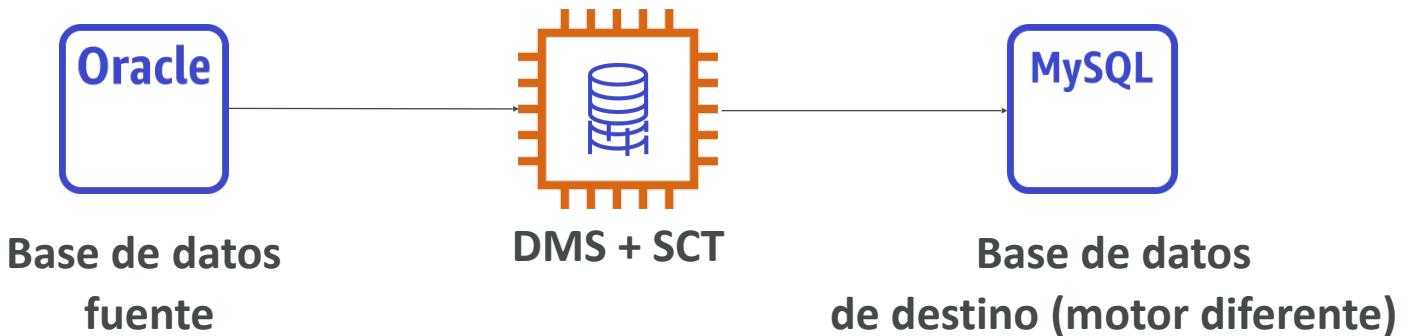
- Bases de datos locales e instancias EC2: Oracle, MS SQL Server, MySQL, MariaDB, PostgreSQL, MongoDB, SAP, DB2
- Azure: Base de datos Azure SQL
- Amazon RDS: todos, incluido Aurora
- Amazon S3

OBJETIVOS:

- Bases de datos locales e instancias EC2: Oracle, MS SQL Server, MySQL, MariaDB, PostgreSQL, SAP
- Amazon RDS
- Amazon Redshift
- Amazon DynamoDB
- Amazon S3
- Servicio ElasticSearch
- Kinesis Data Streams
- DocumentDB

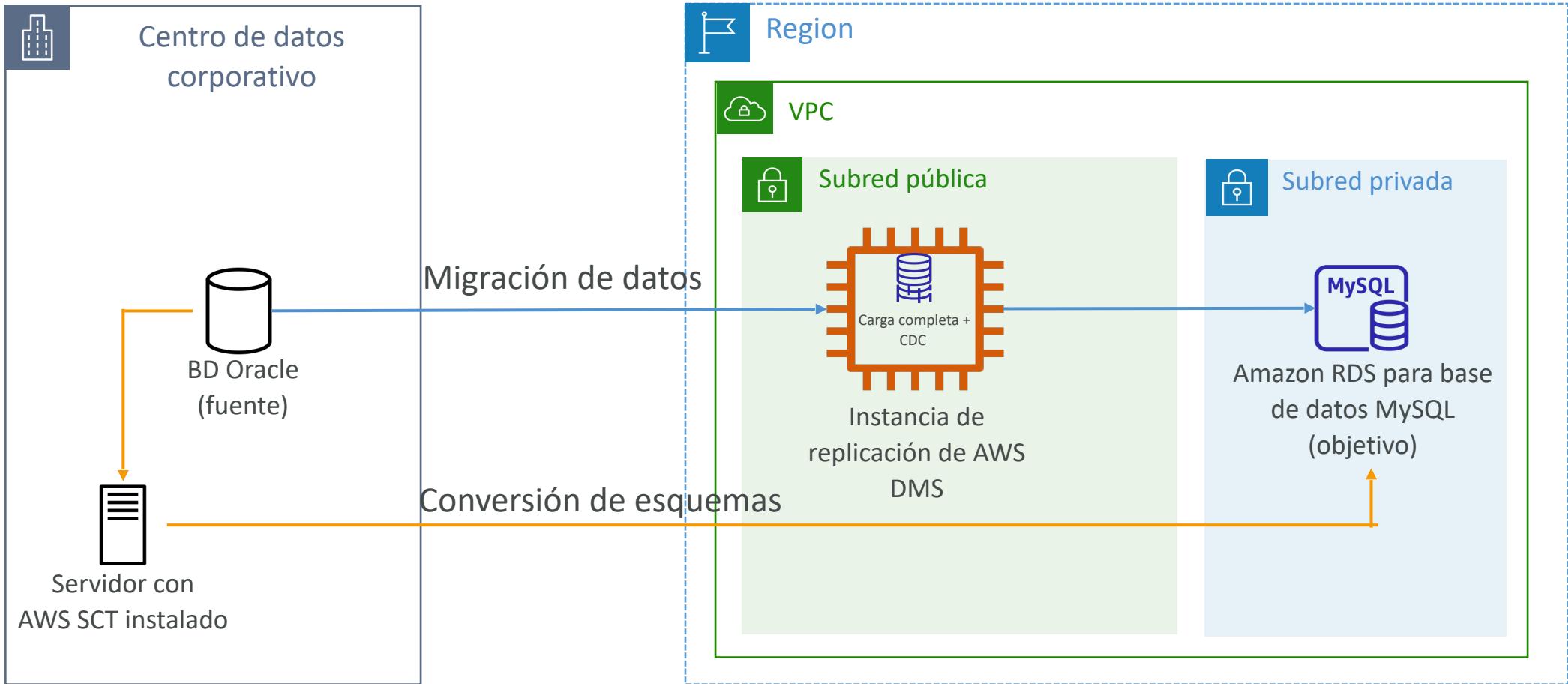
Herramienta de conversión de esquemas de AWS (SCT)

- Convierte el esquema de tu base de datos de un motor a otro
- Ejemplo OLTP: (SQL Server u Oracle) a MySQL, PostgreSQL, Aurora
- Ejemplo OLAP: (Teradata u Oracle) a Amazon Redshift
- Prefiere instancias de cálculo intensivo para optimizar las conversiones de datos



- **No necesitas usar SCT si estás migrando el mismo motor de BD**
 - Ej: PostgreSQL local => PostgreSQL RDS
 - El motor de base de datos sigue siendo PostgreSQL (RDS es la plataforma)

DMS - Replicación continua



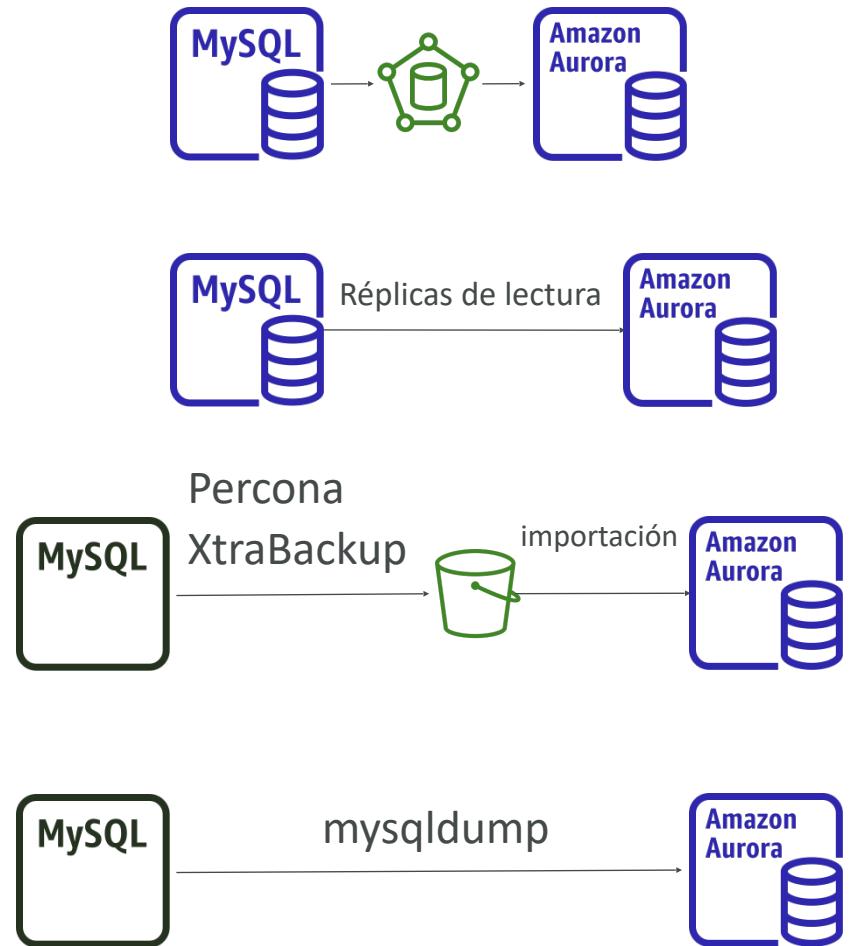
Migraciones RDS y Aurora MySQL

• RDS MySQL a Aurora MySQL

- Opción 1: Snapshots de BD de MySQL RDS restaurados como BD Aurora MySQL
- Opción 2: Crea una réplica de lectura Aurora a partir de tu RDS MySQL, y cuando el retardo de replicación sea 0, promuévela como su propio Cluster de BD (puede llevar tiempo y costar \$)

• MySQL externo a MySQL Aurora

- Opción 1:
 - Utiliza Percona XtraBackup para crear una copia de seguridad de archivos en Amazon S3
 - Crea una BD Aurora MySQL desde Amazon S3
- Opción 2:
 - Crear una BD MySQL de Aurora
 - Utiliza la utilidad mysqldump para migrar MySQL a Aurora (más lento que el método S3)
- Utiliza DMS si ambas bases de datos están en funcionamiento



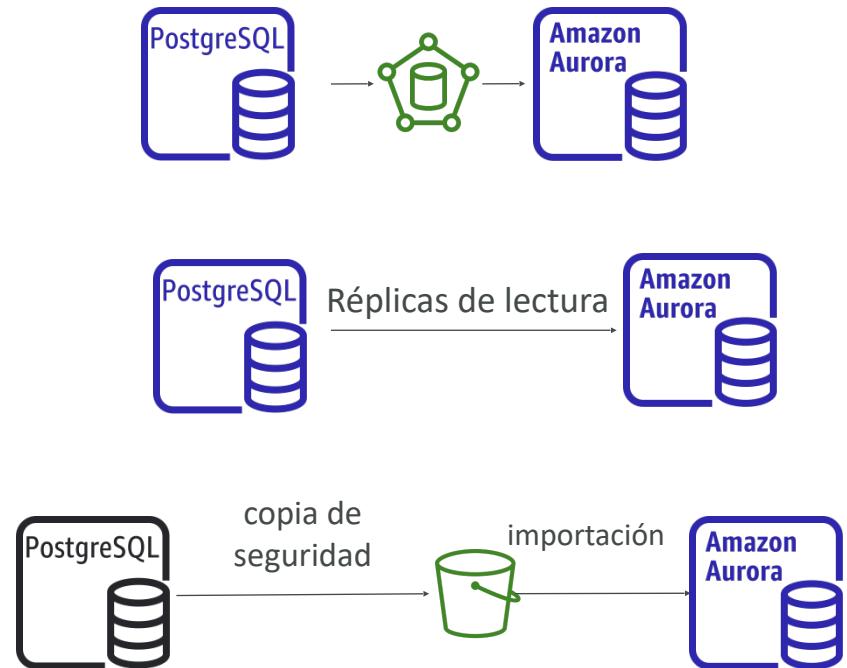
Migraciones RDS y Aurora PostgreSQL

• RDS PostgreSQL a Aurora PostgreSQL

- Opción 1: Snapshots de BD de PostgreSQL RDS restauradas como BD Aurora PostgreSQL
- Opción 2: Crea una réplica de lectura Aurora a partir de tu PostgreSQL RDS, y cuando el retardo de replicación sea 0, promuévela como su propio Cluster de BD (puede llevar tiempo y costar \$)

• PostgreSQL externo a PostgreSQL de Aurora

- Crea una copia de seguridad y ponla en Amazon S3
 - Impórtala utilizando la extensión aws_s3 de Aurora
-
- Utiliza DMS si ambas bases de datos están en funcionamiento



Estrategia en las instalaciones con AWS

- **Posibilidad de descargar Amazon Linux 2 AMI como VM (formato .iso)**
 - VMWare, KVM, VirtualBox (Oracle VM), Microsoft Hyper-V
- **Importación / Exportación de VM**
 - Migrar aplicaciones existentes a EC2
 - Crea una estrategia de repositorio de DR para tus VM locales
 - Puedes volver a exportar las máquinas virtuales de EC2 a las locales
- **Servicio de descubrimiento de aplicaciones de AWS**
 - Recopila información sobre tus servidores locales para planificar una migración
 - Utilización de servidores y asignaciones de dependencias
 - Realiza un seguimiento con AWS Migration Hub
- **Servicio de Migración de Bases de Datos de AWS (DMS)**
 - Replica On-premise => AWS , AWS => AWS, AWS => On-premise
 - Funciona con varias tecnologías de bases de datos (Oracle, MySQL, DynamoDB, etc.)
- **Servicio de migración de servidores de AWS (SMS)**
 - Replicación incremental de servidores activos locales a AWS

AWS Backup



- Servicio totalmente gestionado
- Administra y automatiza centralmente las copias de seguridad en todos los servicios de AWS
- Sin necesidad de crear scripts personalizados ni procesos manuales
- Servicios soportados:
 - Amazon EC2 / Amazon EBS
 - Amazon S3
 - Amazon RDS (todos los motores de BD) / Amazon Aurora / Amazon DynamoDB
 - Amazon DocumentDB / Amazon Neptune
 - Amazon EFS / Amazon FSx (Lustre y Servidor de archivos de Windows)
 - AWS Storage Gateway (Volume Gateway)
- Soporta backups entre regiones
- Soporta backups entre cuentas

AWS Backup



- Soporta PITR para los servicios soportados
- Copias de seguridad bajo demanda y programadas
- Políticas de copia de seguridad basadas en etiquetas
- Creas políticas de copia de seguridad conocidas como **Planes de copia de seguridad**
 - Frecuencia de la copia de seguridad (cada 12 horas, diaria, semanal, mensual, expresión cron)
 - Ventana de copia de seguridad
 - Transición al almacenamiento en frío (Nunca, Días, Semanas, Meses, Años)
 - Periodo de retención (Siempre, Días, Semanas, Meses, Años)

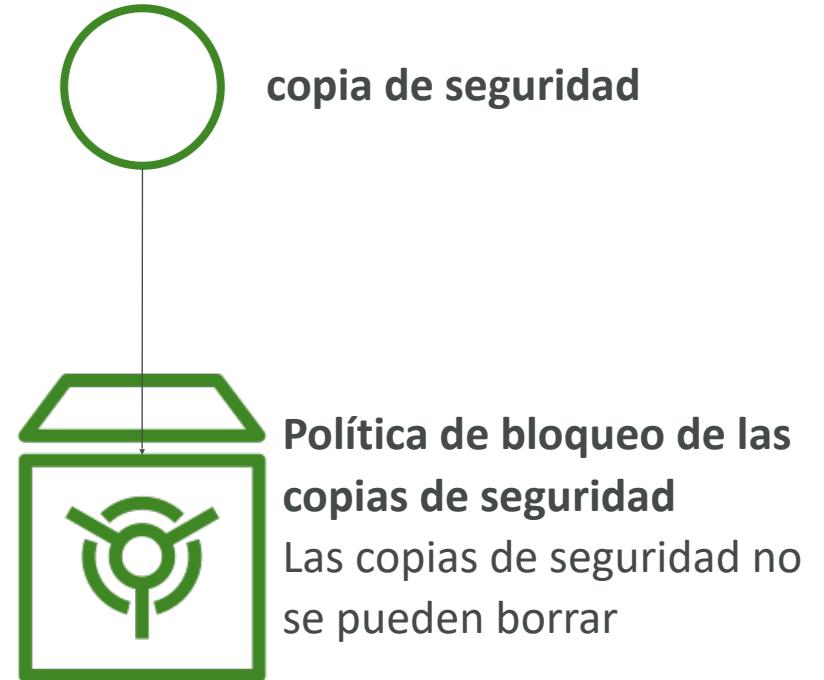
AWS Backup



AWS Backup Vault Lock

Bloqueo de las copias de seguridad

- Aplica un estado WORM (Write Once Read Many) a todos los backups que almacenes en tu bóveda de backups de AWS
- Capa adicional de defensa para proteger tus copias de seguridad contra:
 - Operaciones de borrado inadvertidas o malintencionadas
 - Actualizaciones que acorten o alteren los periodos de retención
- Ni siquiera el usuario root puede borrar copias de seguridad cuando está activado

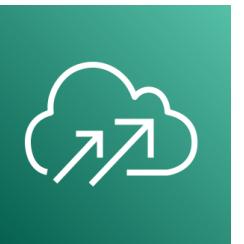


AWS Application Discovery Service

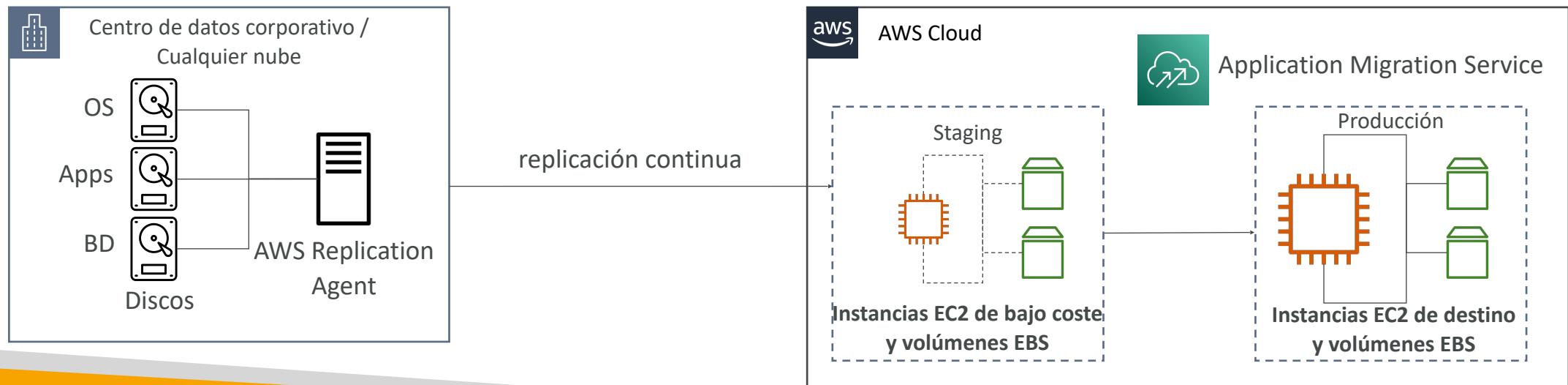


- Planificar los proyectos de migración recopilando información sobre los centros de datos locales
- Los datos de utilización de los servidores y la asignación de dependencias son importantes para las migraciones
- Descubrimiento sin agente (conector de descubrimiento sin agente de AWS)
 - Inventario de máquinas virtuales, configuración e historial de rendimiento, como el uso de la CPU, la memoria y el disco
- Descubrimiento basado en agentes (AWS Application Discovery Agent)
 - Configuración del sistema, rendimiento del sistema, procesos en ejecución y detalles de las conexiones de red entre sistemas
- Los datos resultantes pueden verse en el AWS Migration Hub

AWS Application Migration Service (MGN)



- La "evolución AWS" de CloudEndure Migration, que sustituye al Servicio de Migración de Servidores de AWS (SMS)
- Solución Lift-and-shift que simplifica la migración de aplicaciones a AWS
- Convierte tus servidores físicos, virtuales y basados en la nube para que se ejecuten de forma nativa en AWS
- Soporta una amplia gama de plataformas, sistemas operativos y bases de datos



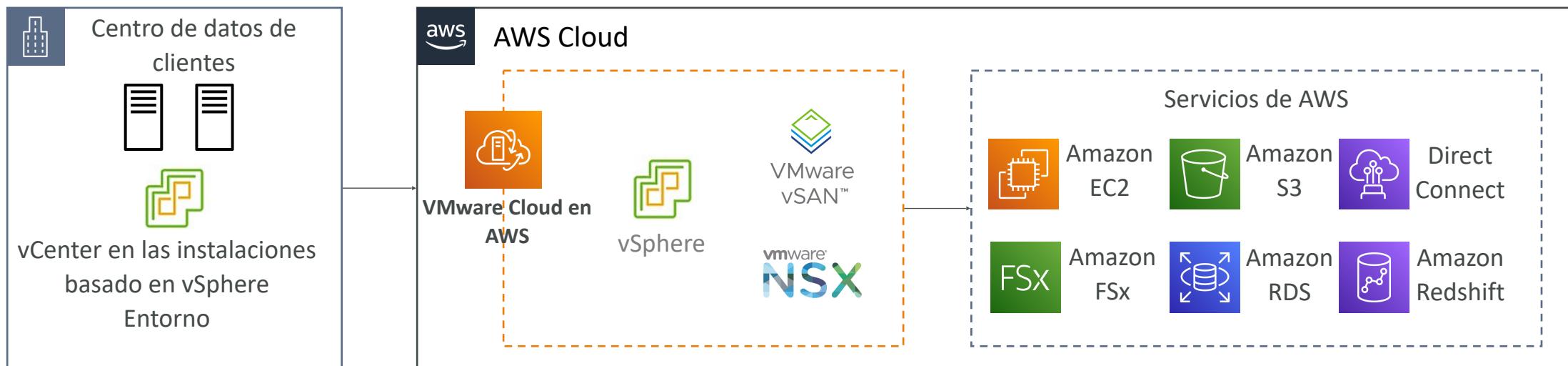
Transferir grandes cantidades de datos a AWS

- Ejemplo: transferir 200 TB de datos en el Cloud. Tenemos una conexión a Internet de 100 Mbps.
- **A través de Internet / VPN Site-to-Site (Sitio a Sitio):**
 - Configuración inmediata
 - Tardará $200(\text{TB}) * 1000(\text{GB}) * 1000(\text{MB}) * 8(\text{Mb}) / 100 \text{ Mbps} = 16.000.000\text{s} = 185\text{d}$
- **Sobre Direct Connect | Gbps:**
 - Mucho tiempo para la configuración única (más de un mes)
 - Llevará $200(\text{TB}) * 1000(\text{GB}) * 8(\text{Gb}) / 1 \text{ Gbps} = 1.600.000\text{s} = 18,5\text{d}$
- **Sobre Snowball:**
 - Llevará de 2 a 3 Snowballs en paralelo
 - La transferencia de extremo a extremo tarda aproximadamente 1 semana
 - Puede combinarse con DMS
- **Para replicación / transferencias en curso:** VPN Site-to-Site o DX con DMS o DataSync

VMware Cloud en AWS

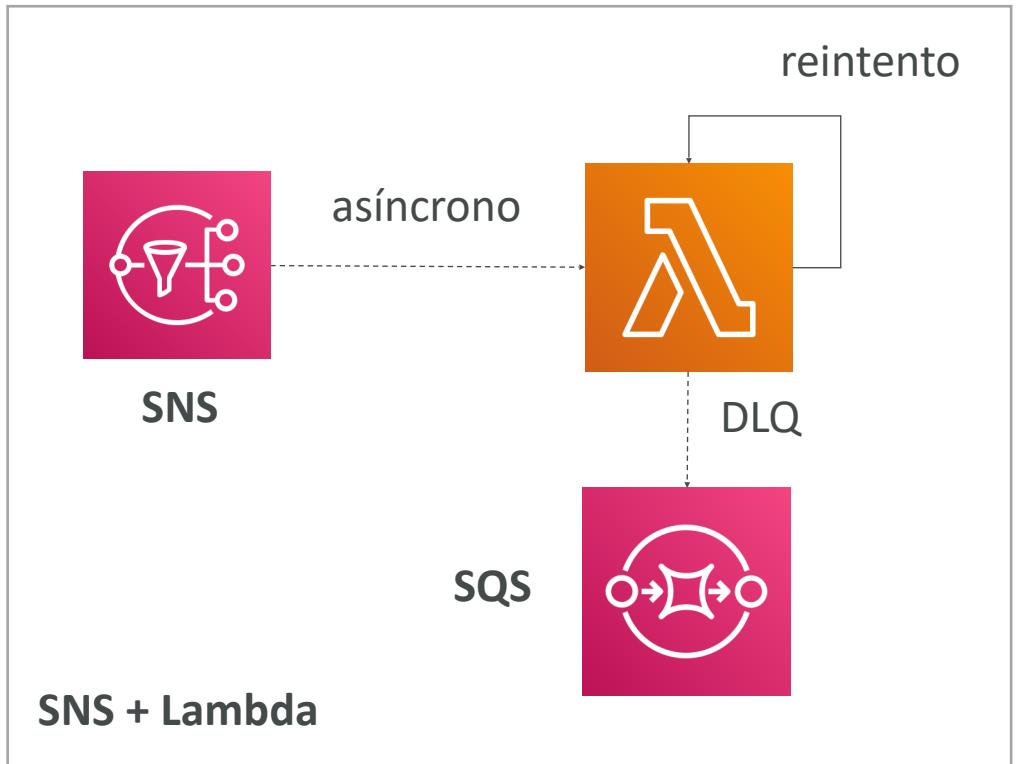
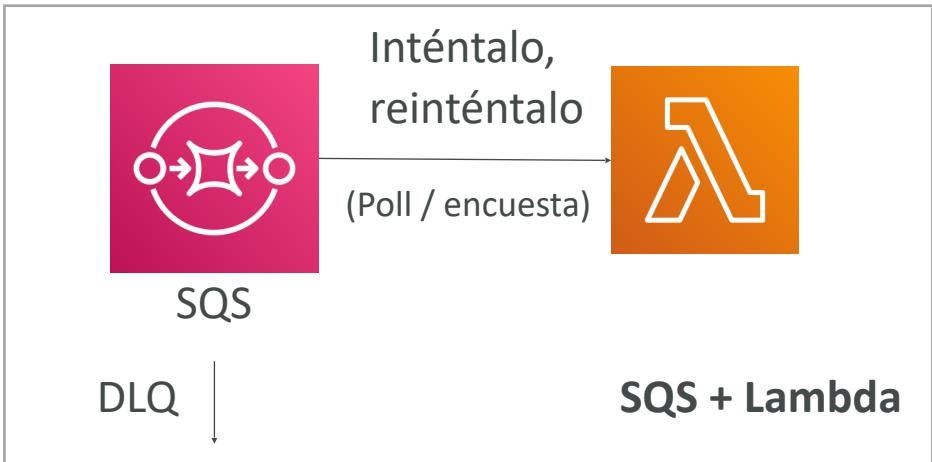


- Algunos clientes utilizan VMware Cloud para gestionar su Centro de Datos local
- Quieren ampliar la capacidad del Centro de Datos a AWS, pero siguen utilizando el software VMware Cloud
- ...Entra en VMware Cloud en AWS
- Casos prácticos
 - Migra tus cargas de trabajo basadas en VMware vSphere a AWS
 - Ejecuta tus cargas de trabajo de producción en entornos de nube privada, pública e híbrida basados en VMware vSphere
 - Tener una estrategia de recuperación de desastres

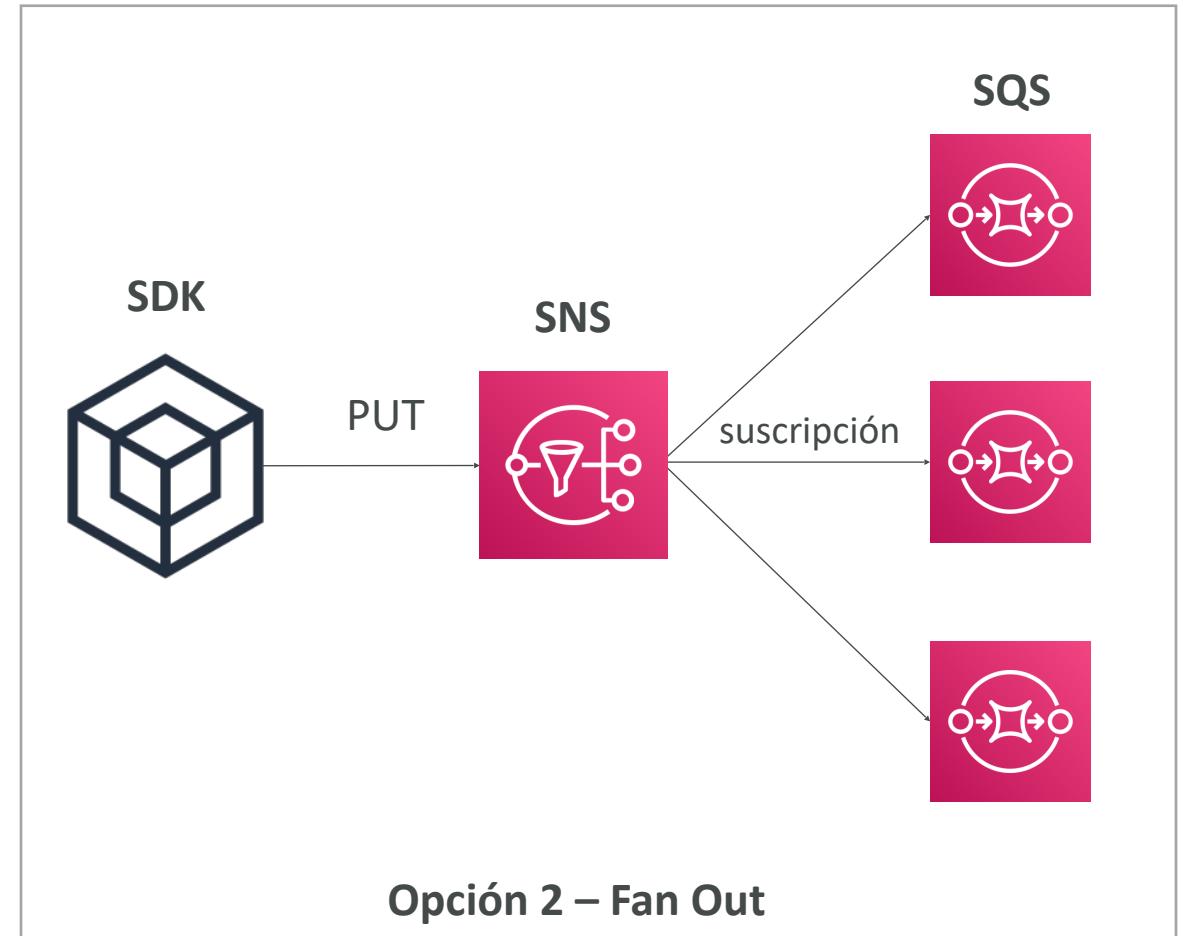
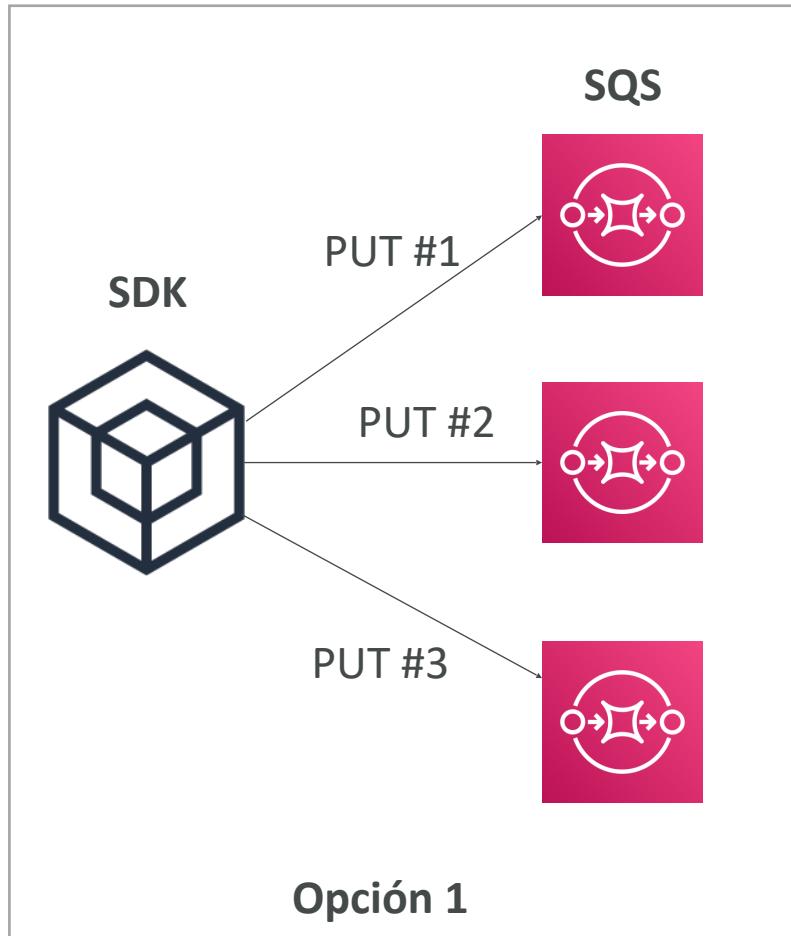


Discusiones adicionales sobre más Arquitecturas de Soluciones

Lambda, SNS & SQS

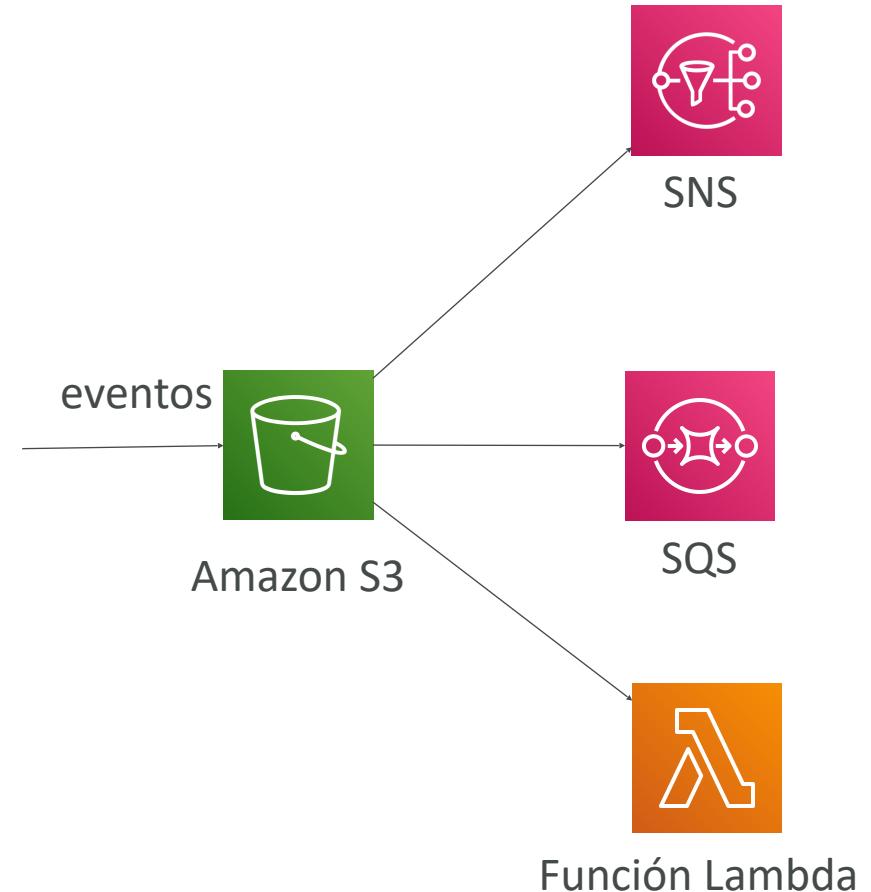


Patrón Fan Out: entrega a varios SQS

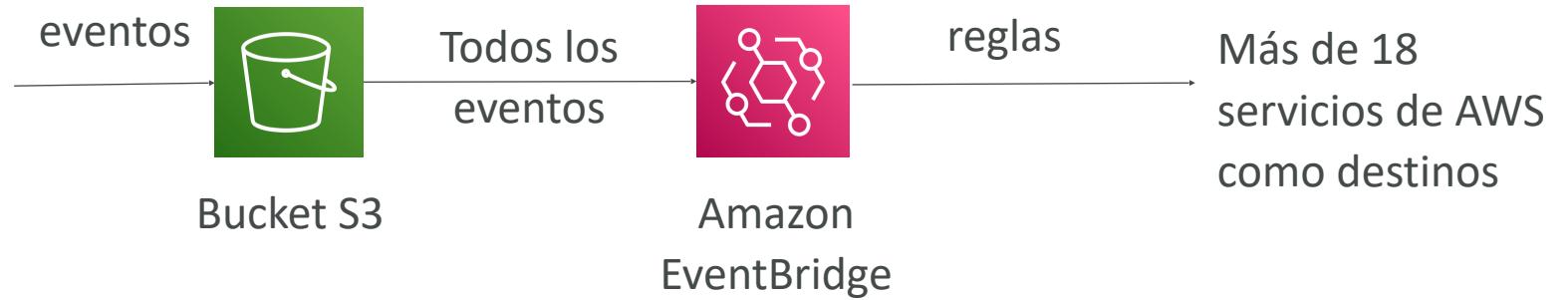


Notificaciones de eventos S3

- S3:ObjectCreado, S3:ObjectEliminado, S3:ObjectRestaurado, S3:Replicación...
- Posibilidad de filtrar el nombre del objeto (*.jpg)
- Caso de uso: generar miniaturas de imágenes subidas a S3
- **Se pueden crear tantos "eventos S3" como se desee**
- Las notificaciones de eventos S3 suelen entregar los eventos en segundos, pero a veces pueden tardar un minuto o más

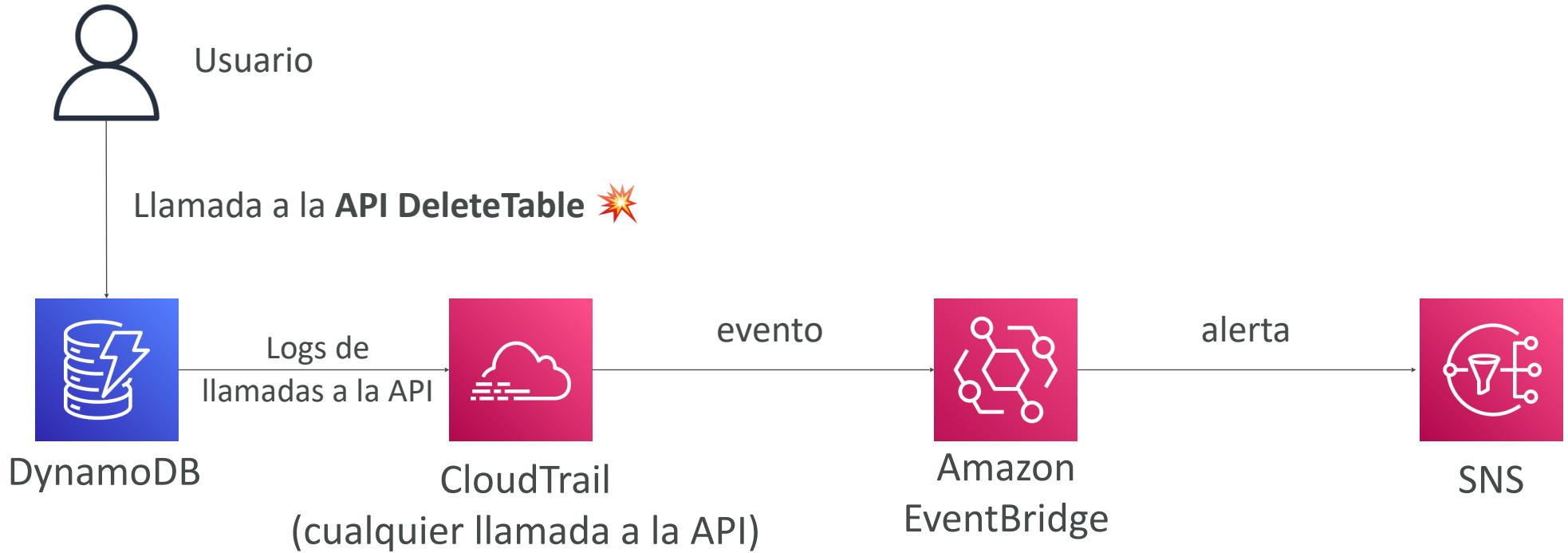


Notificaciones de eventos S3 con Amazon EventBridge



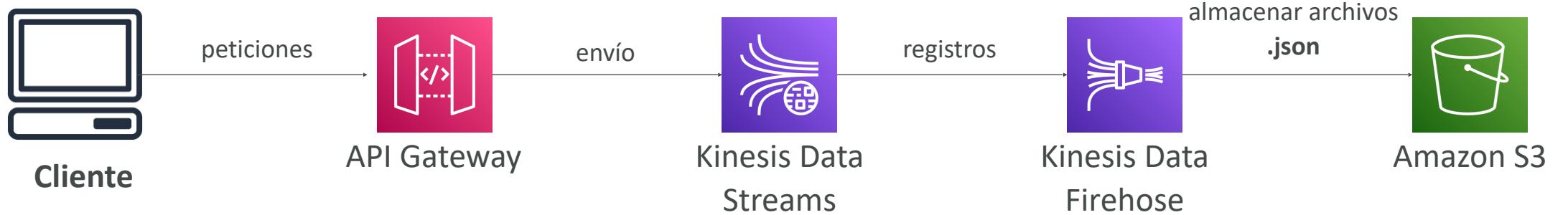
- **Opciones avanzadas de filtrado** con reglas JSON (metadatos, tamaño del objeto, nombre...)
- **Múltiples destinos** - Funciones ex Step, Kinesis Streams / Firehose...
- **Capacidades EventBridge** - Archivar, repetición de eventos, entrega fiable

Amazon EventBridge - Interceptar llamadas a la API

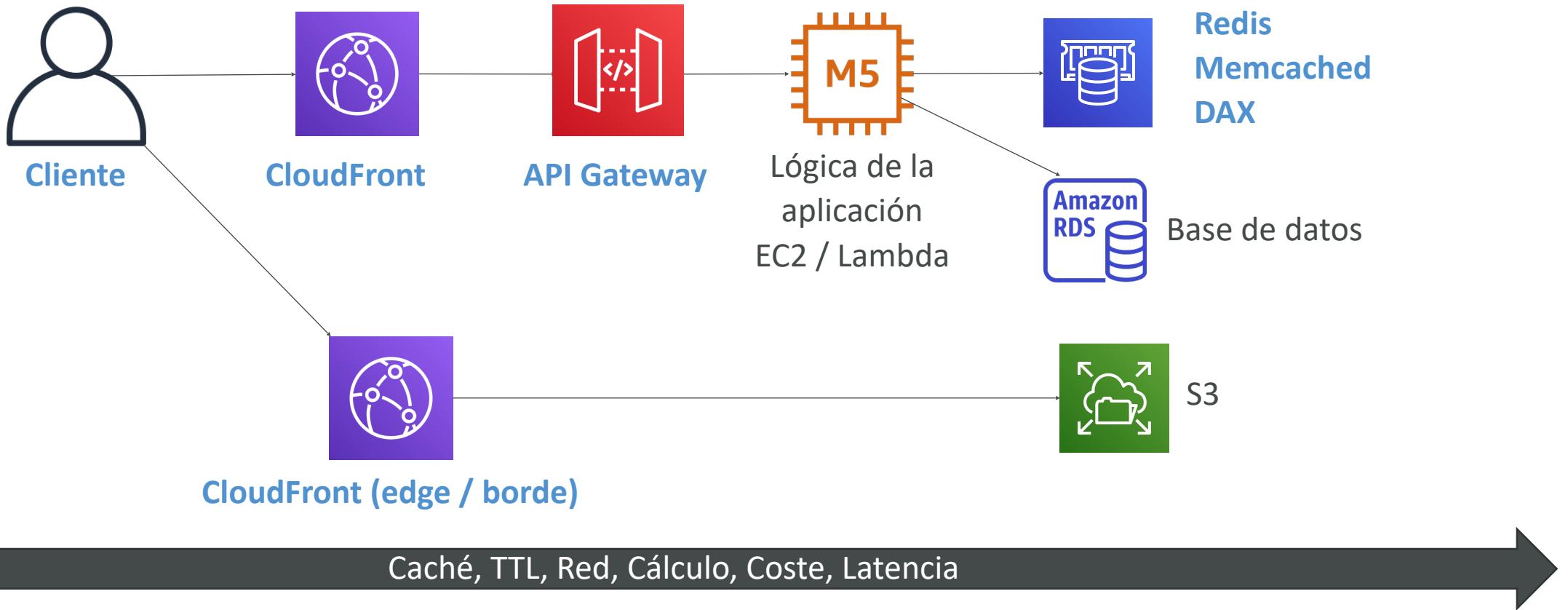


API Gateway - Integración de servicios

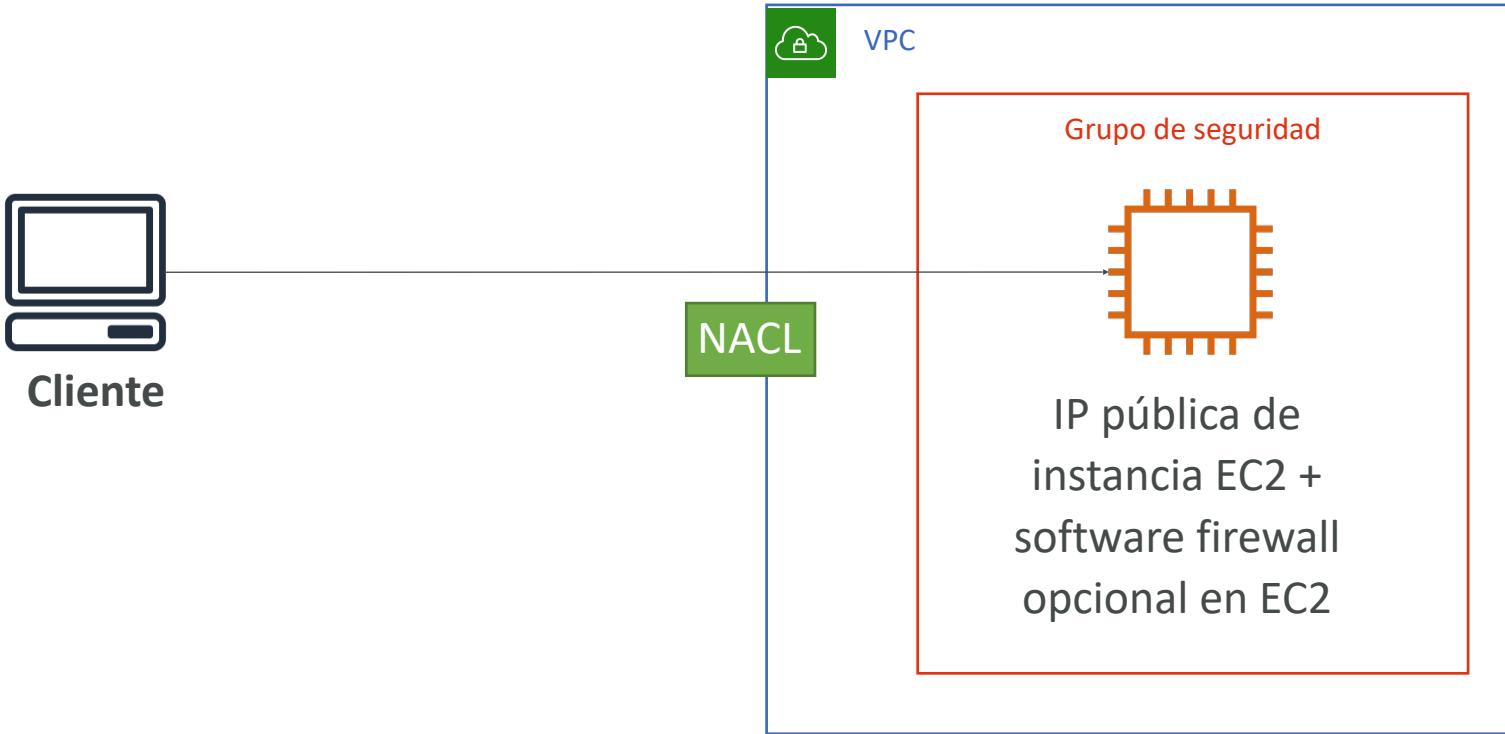
Ejemplo: Kinesis Data Streams



Estrategias de almacenamiento en caché



Bloquear una dirección IP



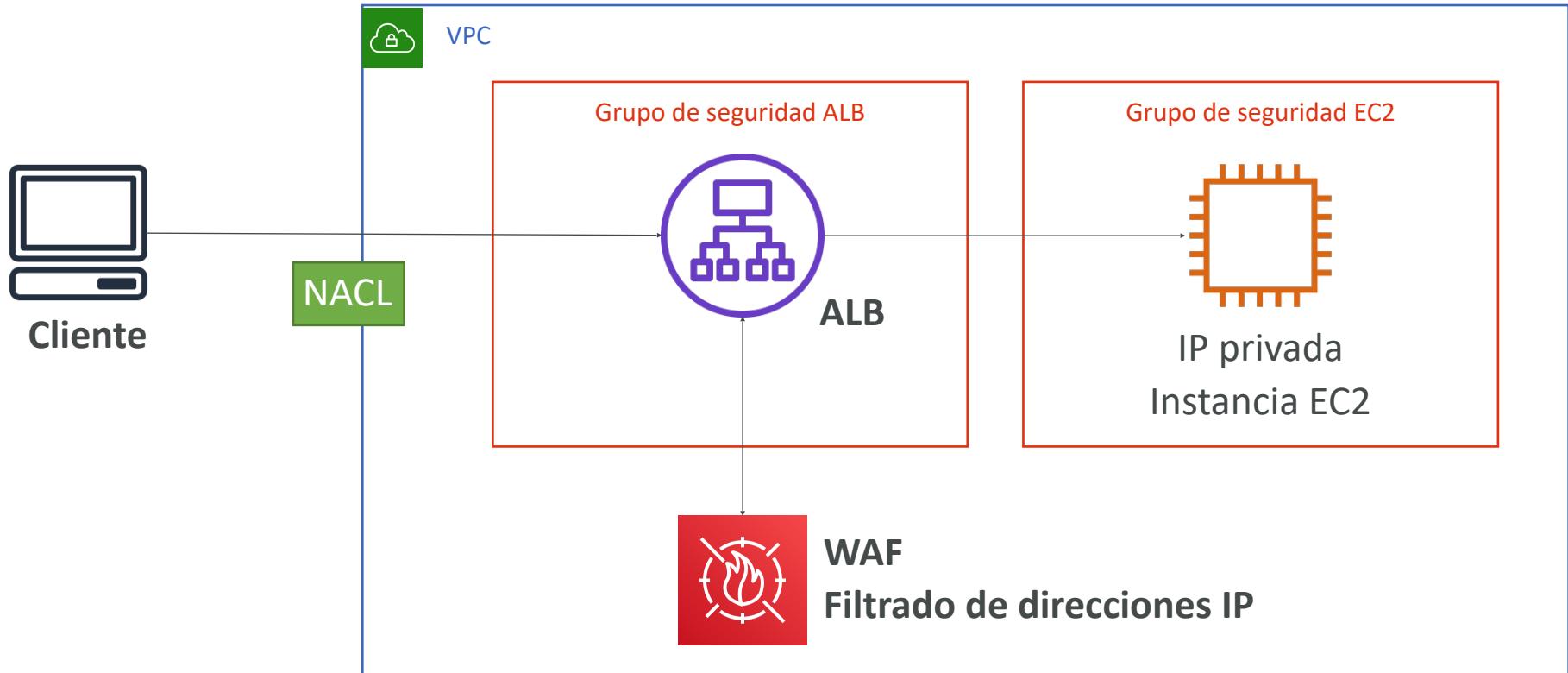
Bloquear una dirección IP - con un ALB



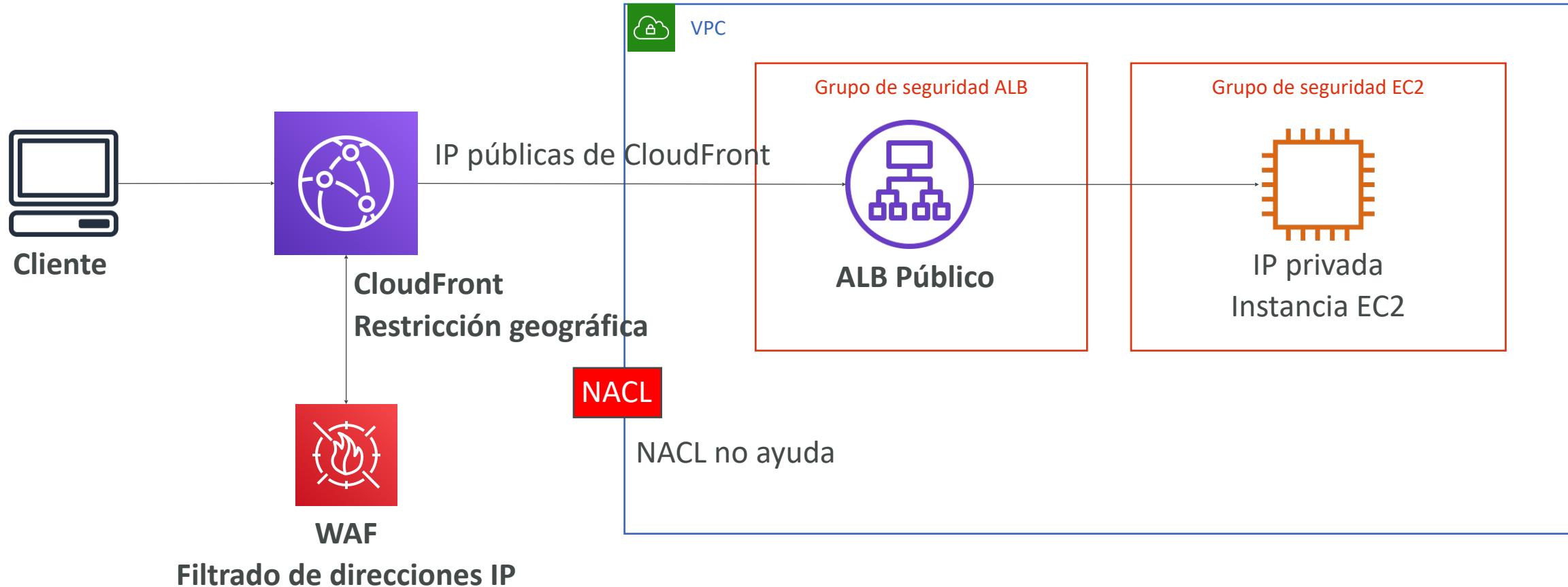
Bloquear una dirección IP - con un NLB



Bloquear una dirección IP - ALB + WAF



Bloquear una dirección IP - ALB, CloudFront WAF



High Performance Computing (HPC)

Computación de Alto Rendimiento

- El Cloud es el lugar perfecto para realizar HPC
- Puedes crear un número muy elevado de recursos en muy poco tiempo
- Puedes acelerar el tiempo de obtención de resultados añadiendo más recursos
- Puedes pagar sólo por los sistemas que hayas utilizado
- Realiza química computacional, modelización de riesgos financieros, predicción meteorológica, Machine Learning, aprendizaje profundo, conducción autónoma
- ¿Qué servicios ayudan a realizar HPC?

Gestión y transferencia de datos

- **AWS Direct Connect:**

- Mueve GB/s de datos al Cloud, a través de una red privada segura

- **Snowball y Snowmobile:**

- Mueve PB de datos al Cloud

- **AWS DataSync:**

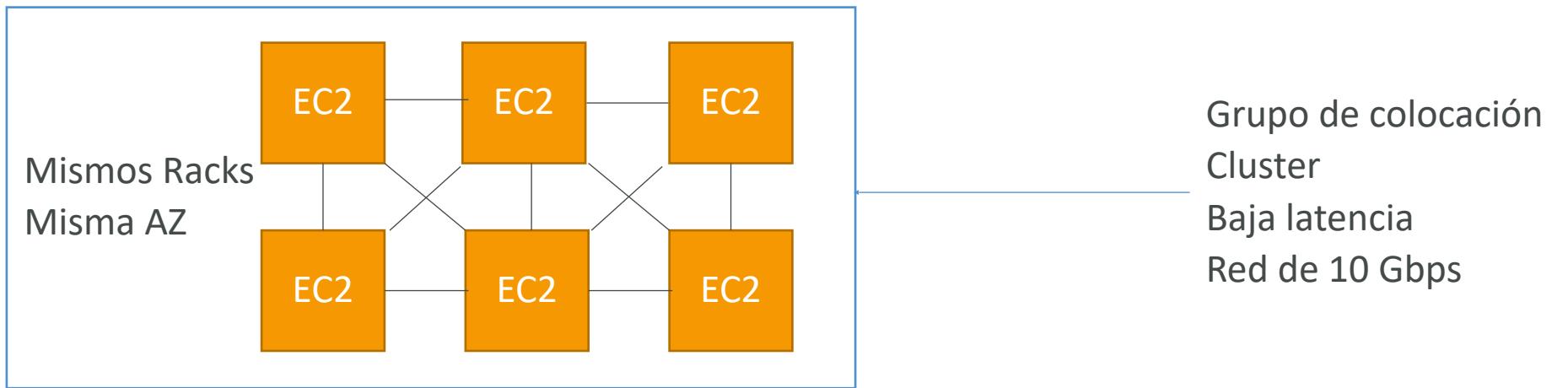
- Mueve grandes cantidades de datos entre las instalaciones y S3, EFS, FSx para Windows

Informática y Redes

- **Instancias EC2:**

- Optimizadas para CPU, optimizadas para GPU
- Instancias de Spot / Flotas de Spot para ahorrar costes + Autoescalado

- **Grupos de Colocación EC2:** Cluster para un buen rendimiento de la red



Informática y Redes

- Redes mejoradas EC2 (SR-IOV)
 - Mayor ancho de banda, mayor PPS (paquetes por segundo), menor latencia
 - Opción 1: **Elastic Network Adapter (ENA)** hasta 100 Gbps
 - Opción 2: Intel 82599 VF hasta 10 Gbps - LEGACY
- **Adaptador Elastic Fabric (EFA)**
 - ENA mejorado para **HPC**, sólo funciona para **Linux**
 - Excelente para comunicaciones entre nodos, **cargas de trabajo estrechamente acopladas**
 - Aprovecha el estándar Message Passing Interface (MPI)
 - Evita el sistema operativo Linux subyacente para proporcionar un transporte fiable y de baja latencia

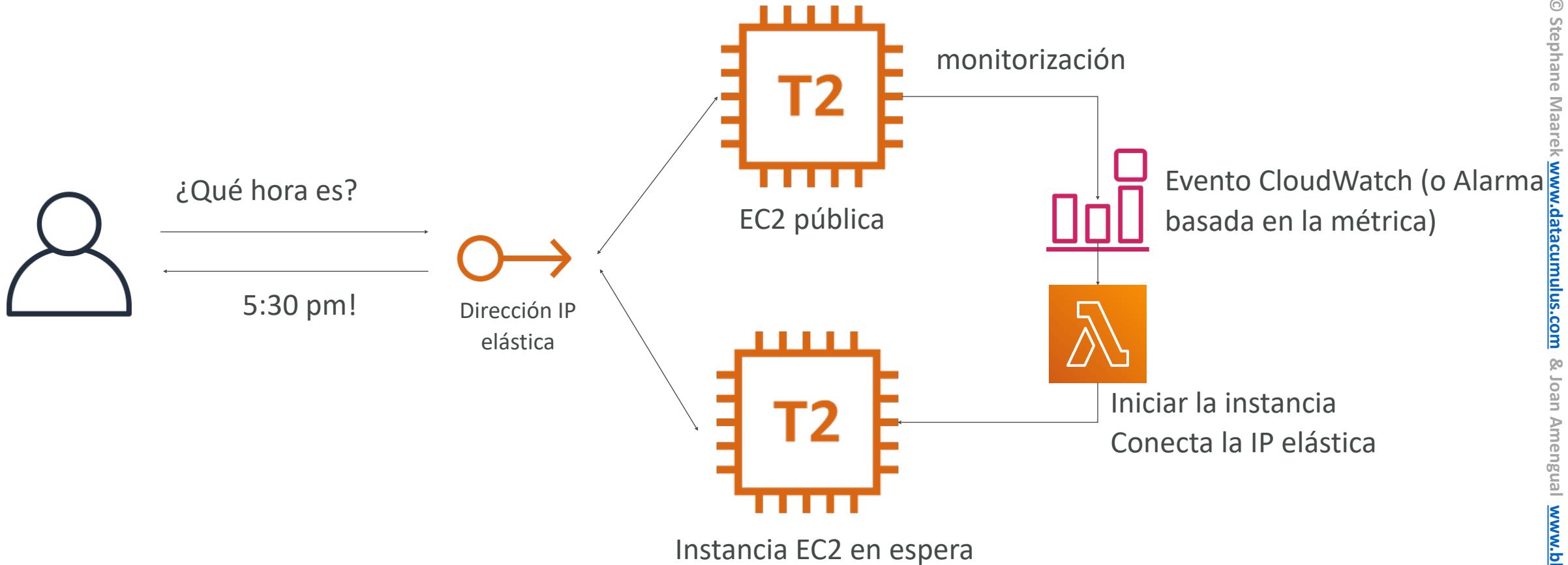
Almacenamiento

- Almacenamiento conectado a instancia:
 - **EBS:** escala hasta 256.000 IOPS con io2 Block Express
 - **Instance Store:** escala a millones de IOPS, vinculado a la instancia EC2, baja latencia
- Almacenamiento en red:
 - **Amazon S3:** blob grande, no es un sistema de archivos
 - **Amazon EFS:** escala las IOPS en función del tamaño total, o utiliza IOPS provisionadas
 - **Amazon FSx para Lustre:**
 - Sistema de archivos distribuidos optimizado para HPC, millones de IOPS
 - Respaldado por S3

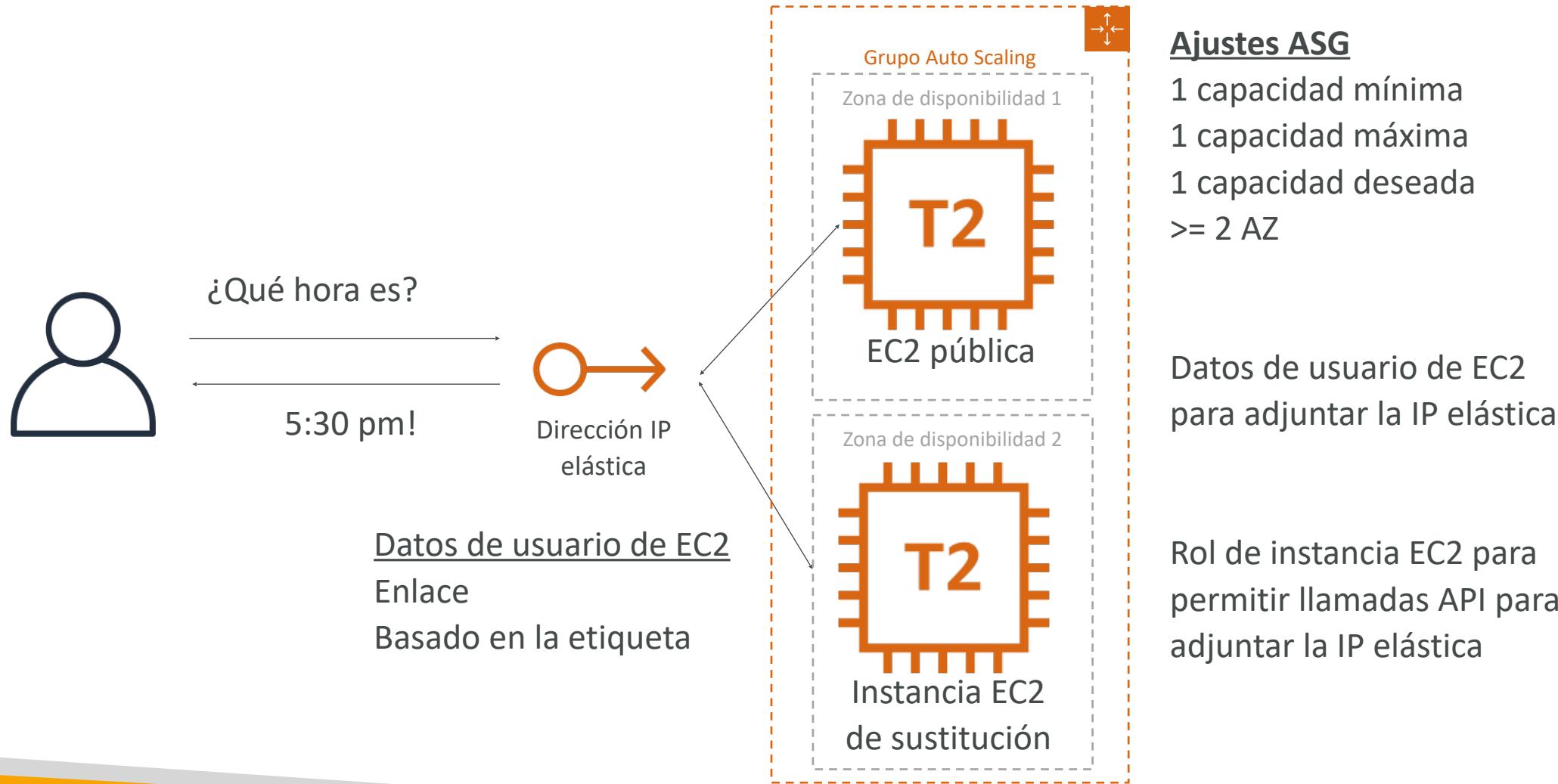
Automatización y orquestación

- AWS Batch
 - **AWS Batch** soporta trabajos paralelos multinodo, lo que te permite ejecutar trabajos únicos que abarcan varias instancias **EC2**.
 - Programa fácilmente los trabajos y lanza las instancias EC2 correspondientes
- AWS ParallelCluster
 - Herramienta de gestión de Cluster de código abierto para implementar HPC en AWS
 - Configurar con archivos de texto
 - Automatiza la creación de VPC, Subred, tipo de Cluster y tipos de instancia
 - **Posibilidad de activar EFA en el Cluster (mejora el rendimiento de la red)**

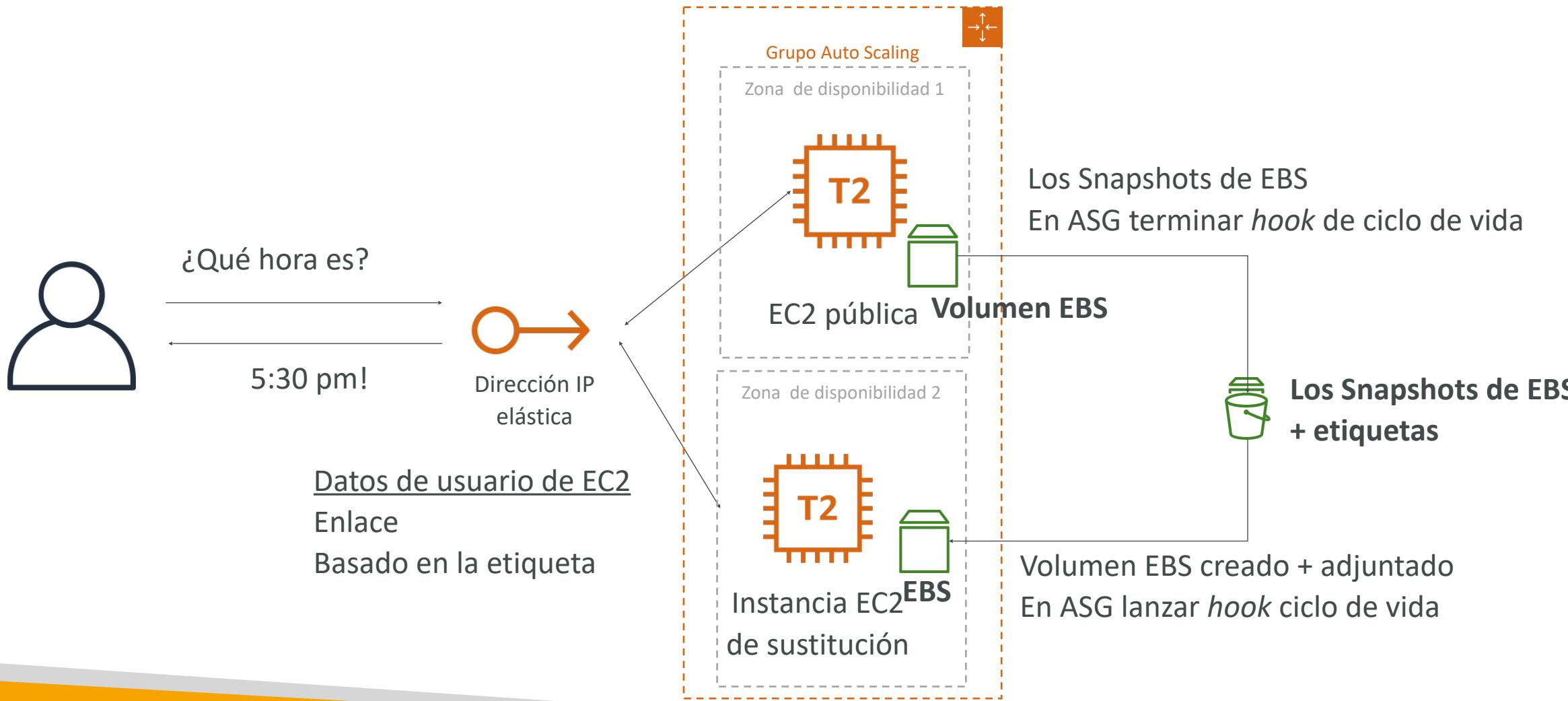
Crear una instancia EC2 de alta disponibilidad



Creación de una instancia EC2 de alta disponibilidad con un Auto Scaling Group



Crear una instancia EC2 de alta disponibilidad con ASG + EBS



Otros servicios

Visión general de los Servicios que podrían surgir en algunas preguntas



Qué es CloudFormation

- CloudFormation es una forma declarativa de esbozar tu infraestructura de AWS, para cualquier recurso (la mayoría de ellos son compatibles).
- Por ejemplo, dentro de una plantilla de CloudFormation, dices
 - Quiero un grupo de seguridad
 - Quiero dos instancias EC2 que utilicen este grupo de seguridad
 - Quiero un bucket S3
 - Quiero un load balancer (ELB) delante de estas máquinas
- Entonces CloudFormation los crea por ti, en el **orden correcto**, con la **configuración exacta** que especifiques

Ventajas de AWS CloudFormation (1/2)

- **Infraestructura como código**

- No se crean recursos manualmente, lo que es excelente para el control
- Los cambios en la infraestructura se revisan a través del código

- Coste

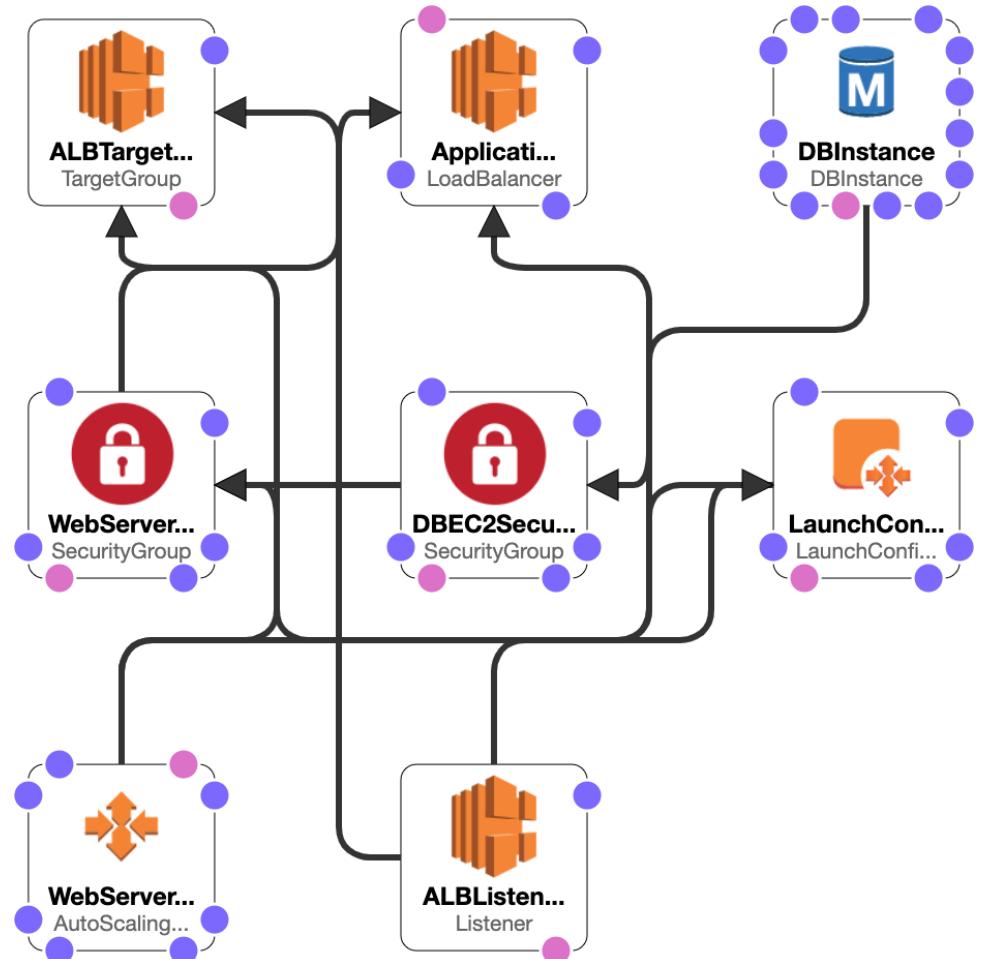
- Cada recurso dentro de la pila está etiquetado con un identificador para que puedas ver fácilmente cuánto te cuesta una pila
- Puedes estimar los costes de tus recursos utilizando la plantilla de CloudFormation
- Estrategia de ahorro: En Dev, podrías automatizar la eliminación de plantillas a las 5 de la tarde y volver a crearlas a las 8 de la mañana, de forma segura

Ventajas de AWS CloudFormation (2/2)

- Productividad
 - Posibilidad de destruir y volver a crear una infraestructura en el Cloud sobre la marcha
 - Generación automatizada de diagramas para tus plantillas
 - Programación declarativa (no es necesario averiguar el orden y la orquestación)
- No vuelvas a inventar la rueda
 - Aprovecha las plantillas existentes en la web
 - Aprovecha la documentación
- **Soporta (casi) todos los recursos de AWS:**
 - Todo lo que veremos en este curso es compatible
 - Puedes utilizar "recursos personalizados" para los recursos que no son compatibles

Stack Designer de CloudFormation

- Ejemplo: Stack de CloudFormation para WordPress
- Podemos ver todos los **recursos**
- Podemos ver las **relaciones** entre los componentes



Amazon Simple Email Service (Amazon SES)



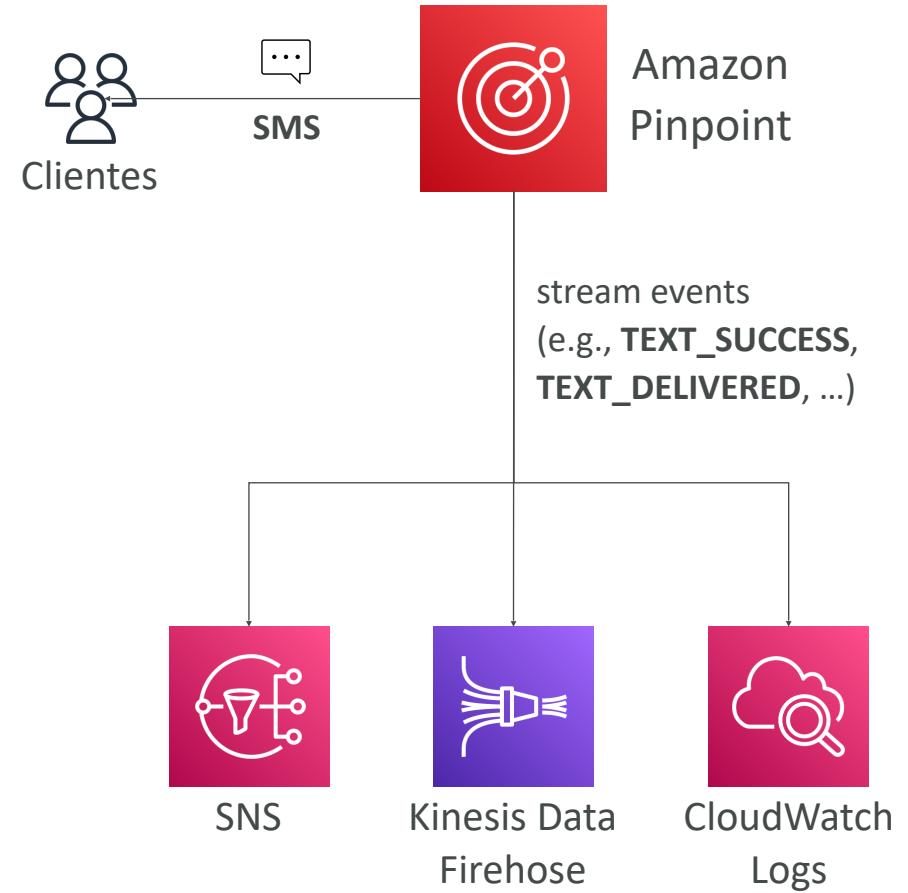
- **Servicio totalmente gestionado para enviar correos electrónicos de forma segura, global y a escala**
- Permite correos electrónicos entrantes/salientes
- Dashboards de reputación, perspectivas de rendimiento, información antispam
- Proporciona estadísticas como entregas de correos electrónicos, rebotes, resultados del bucle de retroalimentación, correos electrónicos abiertos
- Soporta DomainKeys Identified Mail (DKIM) y Sender Policy Framework (SPF)
- Despliegue de IP flexible: IP compartida, dedicada y propiedad del cliente
- Envía correos electrónicos con tu aplicación utilizando la consola de AWS, las API o SMTP
- Casos de uso: comunicaciones transaccionales, de marketing y masivas por correo electrónico



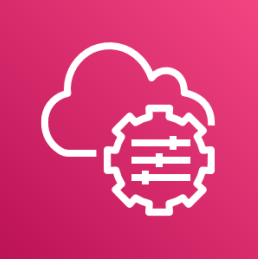
Amazon Pinpoint



- Servicio escalable de comunicaciones de marketing **bidireccional (saliente/entrante)**
- Soporta correo electrónico, SMS, push, voz y mensajería in-app
- Posibilidad de segmentar y personalizar los mensajes con el contenido adecuado para los clientes
- Posibilidad de recibir respuestas
- Escala a miles de millones de mensajes al día
- Casos de uso: realiza campañas enviando mensajes SMS de marketing, masivos y transaccionales
- **Frente a Amazon SNS o Amazon SES**
 - En SNS y SES gestionas la audiencia, el contenido y el calendario de entrega de cada mensaje
 - En Amazon Pinpoint, creas plantillas de mensajes, horarios de entrega, segmentos altamente segmentados y campañas completas



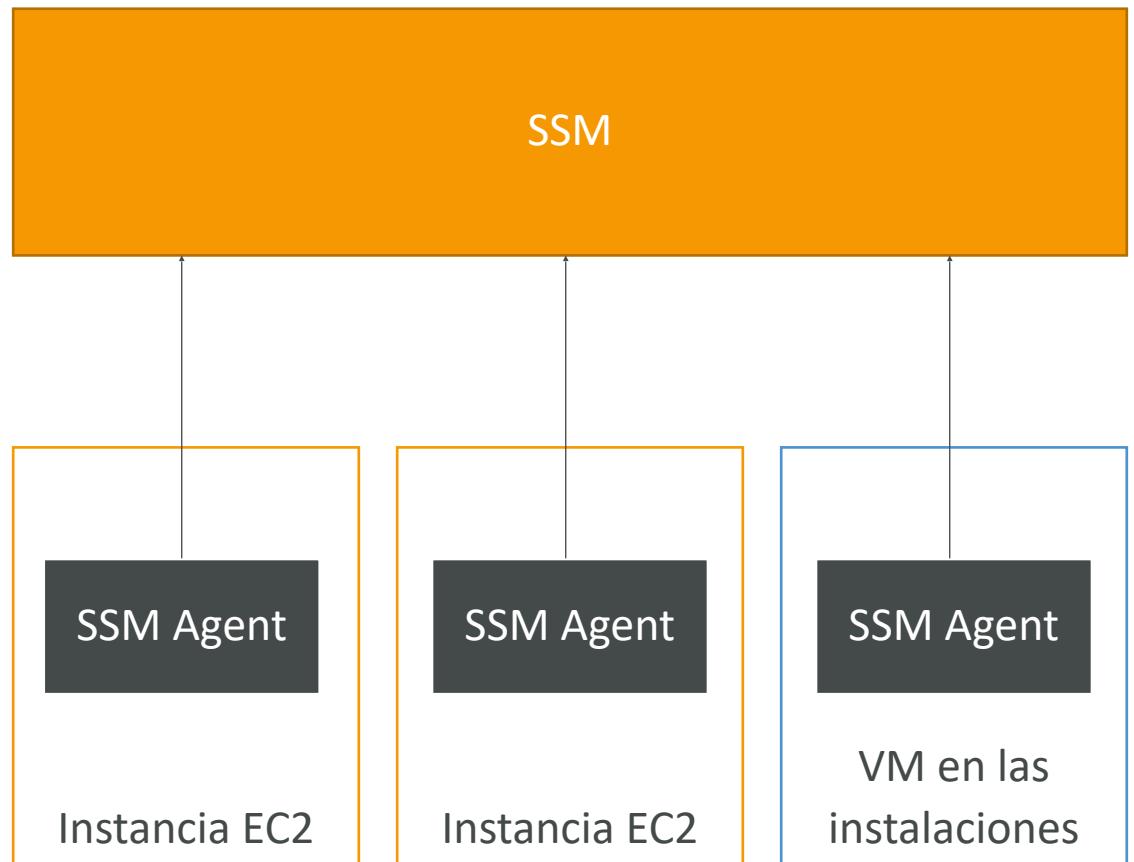
AWS Systems Manager (SSM)



- Te ayuda a gestionar tus sistemas **EC2 y On-Premises** a escala
- Otro servicio **híbrido** de AWS
- Obtén información operativa sobre el estado de tu infraestructura
- Conjunto de más de 10 productos
- Las características más importantes son:
 - **Automatización de parches para mejorar la normativa**
 - **Ejecuta comandos en toda una flota de servidores**
 - Almacena la configuración de los parámetros con el almacén de parámetros SSM
- Funciona tanto para el sistema operativo Windows como para el Linux

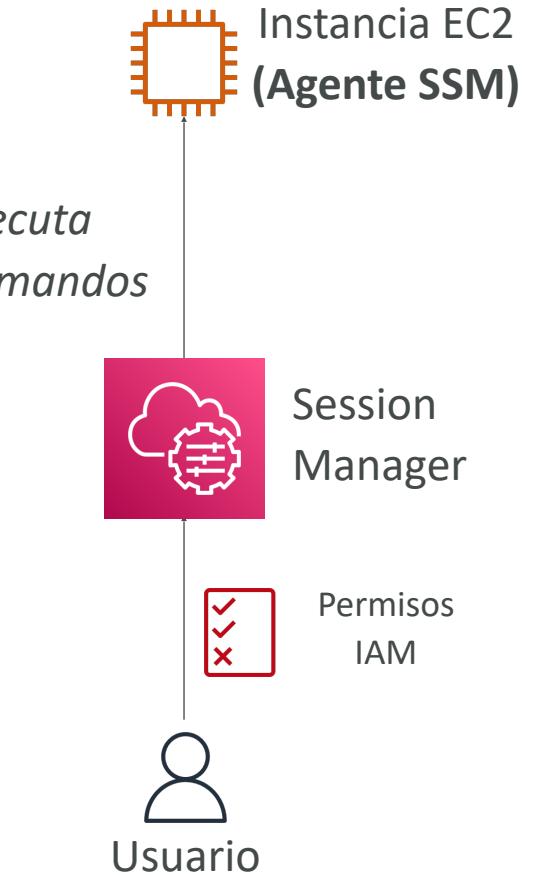
Cómo funciona Systems Manager

- Necesitamos instalar el agente SSM en los sistemas que controlamos
- Se instala por defecto en las AMI de Amazon Linux y en algunas AMI de Ubuntu
- Si una instancia no puede ser controlada con SSM, probablemente se trate de un problema con el agente SSM
- Gracias al agente SSM, podemos **ejecutar comandos, parchear y configurar** nuestros servidores



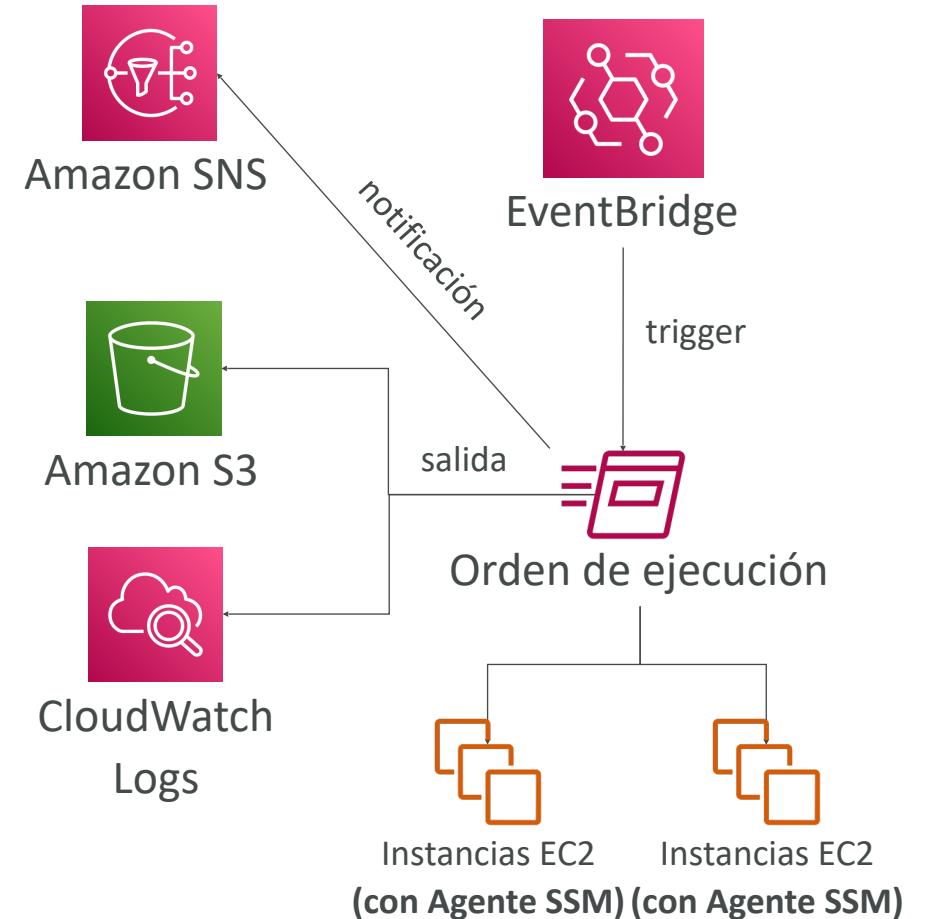
Systems Manager – SSM Session Manager

- Te permite iniciar un shell seguro en tus servidores EC2 y locales
- **No se necesita acceso SSH ni claves SSH**
- **No se necesita el puerto 22 (mayor seguridad)**
- Es compatible con Linux, macOS y Windows
- Envía los datos de registro de la sesión a S3 o a CloudWatch Logs



Systems Manager – Orden de ejecución

- Ejecutar un documento (= script) o simplemente ejecutar un comando
- Ejecuta el comando en varias instancias (utilizando grupos de recursos)
- No necesitas SSH
- La salida del comando puede mostrarse en la consola de AWS, enviarse al bucket de S3 o a los logs de CloudWatch
- Envía notificaciones a SNS sobre el estado del comando (En curso, Correcto, Fallido, ...)
- Integrado con IAM y CloudTrail
- Se puede invocar mediante EventBridge



Systems Manager – Gestor de parches



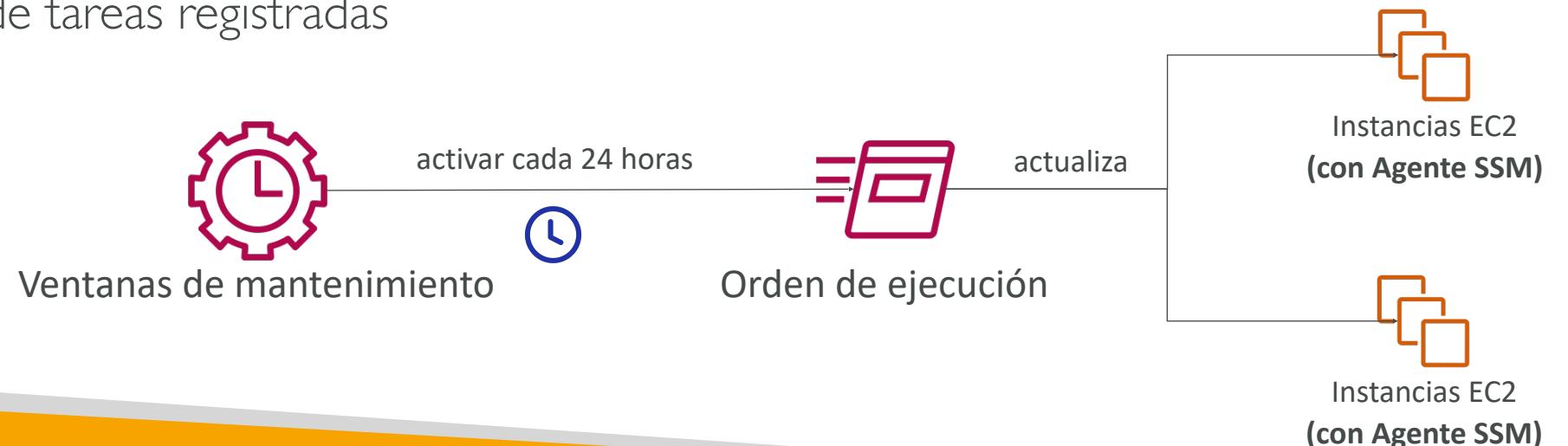
- Automatiza el proceso de aplicación de parches a las instancias gestionadas
- Actualizaciones del SO, actualizaciones de aplicaciones, actualizaciones de seguridad
- Soporta instancias EC2 y servidores locales
- Soporta Linux, macOS y Windows
- Parchea bajo demanda o de forma programada mediante **las ventanas de mantenimiento**
- Escanea instancias y genera informes de normativa de parches (parches que faltan)



Systems Manager - Ventanas de mantenimiento

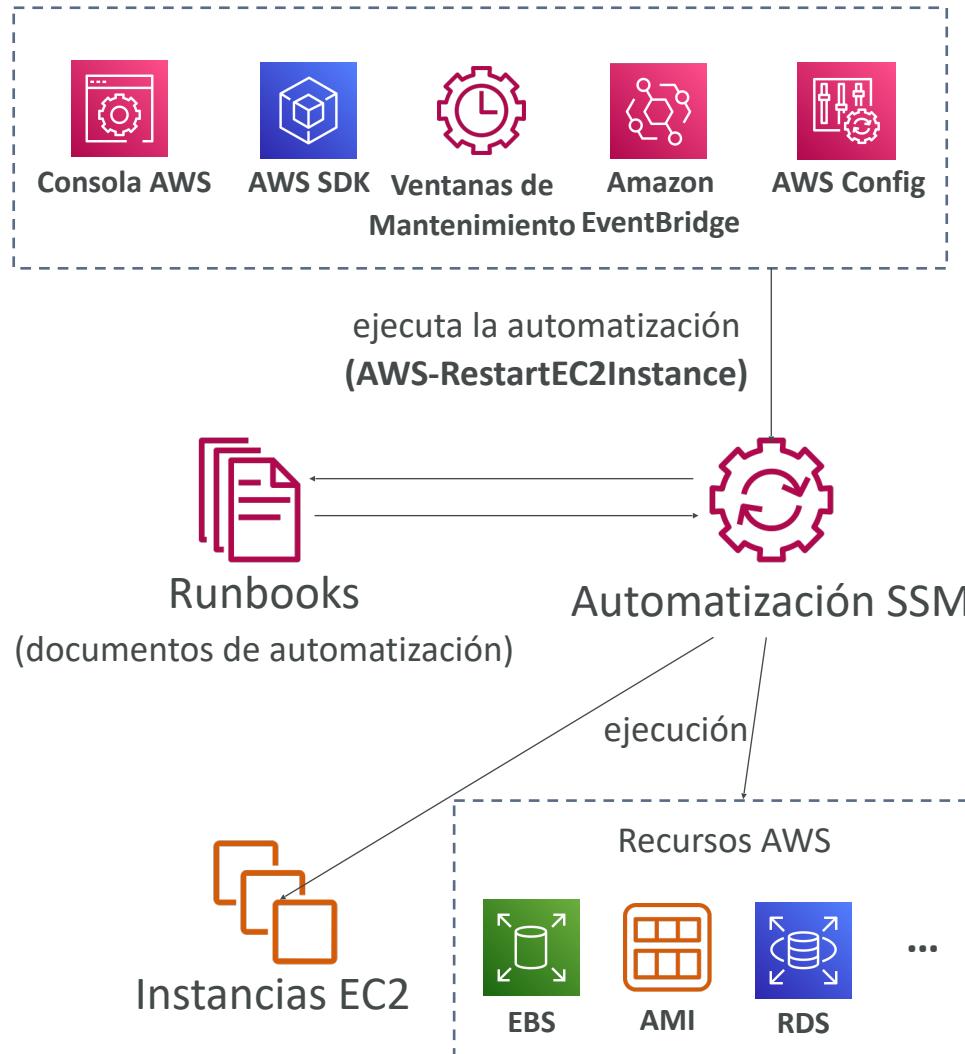


- Define un calendario para saber cuándo realizar acciones en tus instancias
- Por ejemplo Parcheado del SO, actualización de controladores, instalación de software, ...
- La ventana de mantenimiento contiene
 - Programa
 - Duración
 - Conjunto de instancias registradas
 - Conjunto de tareas registradas



Systems Manager - Automatización

- Simplifica las tareas comunes de mantenimiento e implementación de instancias EC2 y otros recursos de AWS
- Ejemplos: reiniciar instancias, crear una AMI, Snapshot de EBS
- **Automation Runbook** - Documentos SSM para definir acciones preformadas en tus instancias EC2 o recursos AWS (predefinidas o personalizadas)
- Pueden activarse mediante:
 - Manualmente mediante la consola de AWS, AWS CLI o SDK
 - Amazon EventBridge
 - De forma programada mediante Ventanas de Mantenimiento
 - Mediante AWS Config para remediaciones de reglas

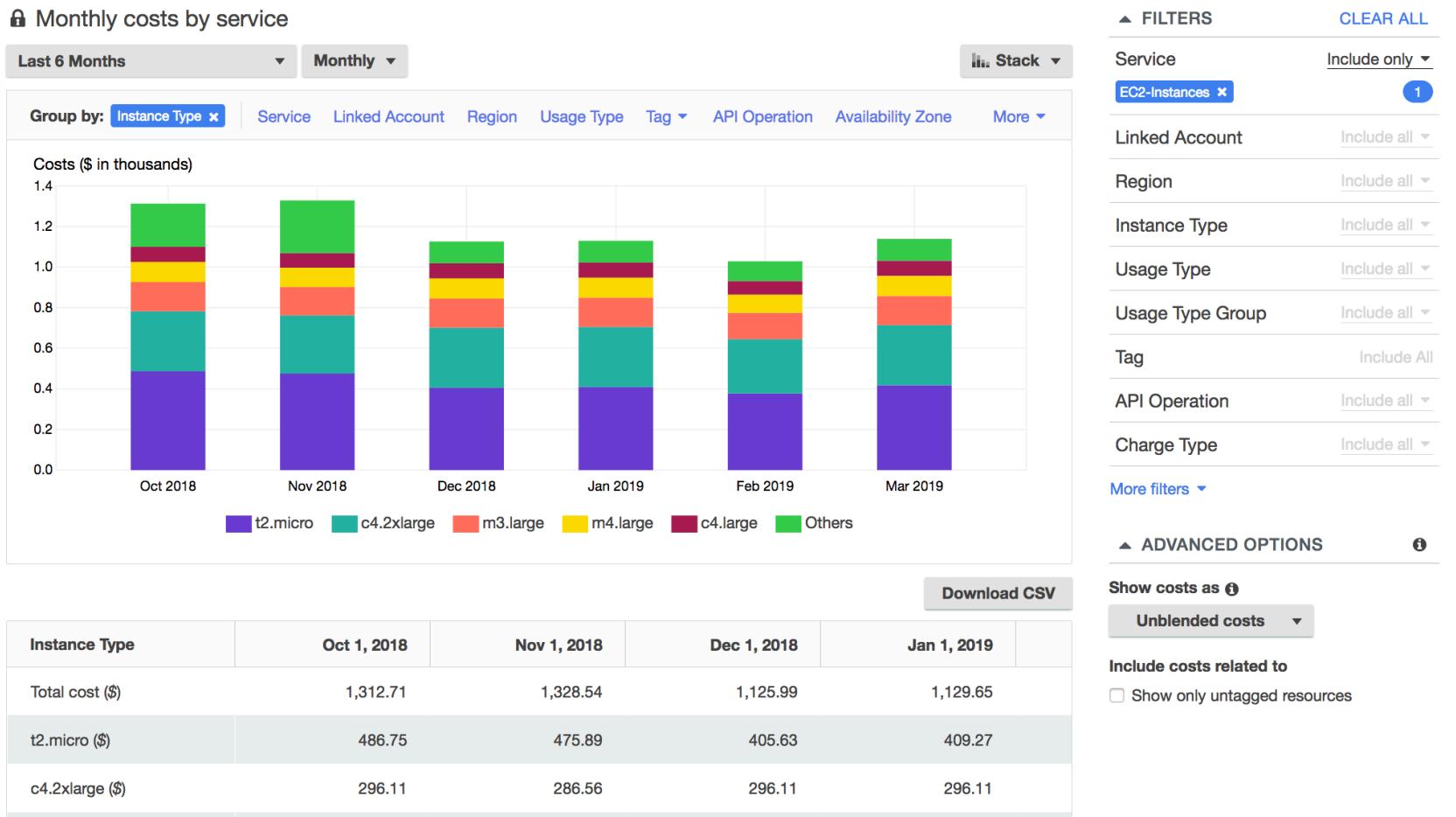


Cost Explorer

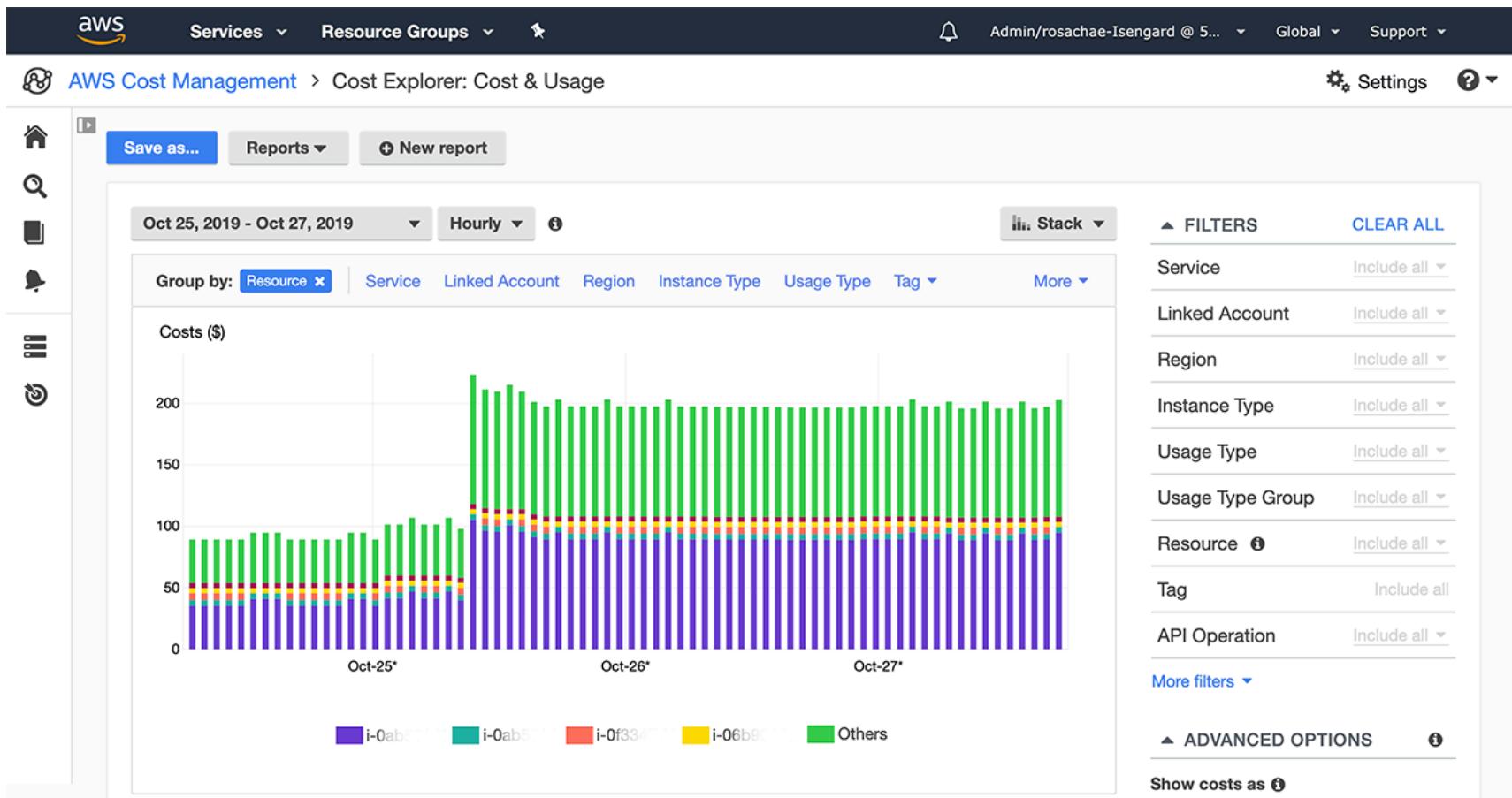


- Visualiza, entiende y gestiona tus costes y uso de AWS a lo largo del tiempo
- Crea informes personalizados que analicen los datos de costes y uso.
- Analiza tus datos a alto nivel: costes totales y uso en todas las cuentas
- O con granularidad mensual, por horas, a nivel de recursos
- Elige un **Plan de Ahorro** óptimo (para reducir los precios de tu factura)
- **Prevé el uso hasta 12 meses basándote en el uso anterior**

Cost Explorer – Coste mensual por servicio de AWS



Cost Explorer– Nivel de horas y recursos



Cost Explorer - Plan de ahorro

Alternativa a las instancias reservadas

Recommendation options

Savings Plans type <input checked="" type="radio"/> Compute <input type="radio"/> EC2 Instance	Savings Plans term <input type="radio"/> 1-year <input checked="" type="radio"/> 3-year	Payment option <input checked="" type="radio"/> All upfront <input type="radio"/> Partial upfront <input type="radio"/> No upfront	Based on the past <input type="radio"/> 7 days <input type="radio"/> 30 days <input checked="" type="radio"/> 60 days
--	---	---	--

Recommendation: Purchase a Compute Savings Plan at a commitment of \$2.40/hour

You could save an estimated **\$1,173** monthly by purchasing the recommended Compute Savings Plan.

Based on your past **60 days** of usage, we recommend purchasing a Savings Plan with a commitment of **\$2.40/hour** for a **3-year term**. With this commitment, we project that you could save an average of **\$1.61/hour** - representing a **40%** savings compared to On-Demand. To account for variable usage patterns, this recommendation maximizes your savings by leaving an average **\$0.04/hour** of On-Demand spend.

Before recommended purchase	After recommended purchase (based on your past 60 days of usage)
Monthly On-Demand spend <small> ⓘ</small> \$2,955 (\$4.05/hour) Based on your On-Demand spend over the past 60 days	Estimated monthly spend <small> ⓘ</small> \$1,782 (\$2.44/hour) Your recommended \$2.40/hour Savings Plans commitment + an average \$0.04/hour of On-Demand spend Estimated monthly savings <small> ⓘ</small> \$1,173 (\$1.61/hour) 40% monthly savings over On-Demand \$2,955 - \$1,782 = \$1,173

This recommendation examines your usage over the past 60 days (including your existing Savings Plans and EC2 Reserved Instances) and calculates what your costs would have been had you purchased the recommended Savings Plans. See applicable rates for Savings Plans [here](#). To generate this recommendation, AWS simulates your bill for different commitment amounts and recommends the commitment amount that provides the greatest estimated savings. [Learn more](#)

Recommended Compute Savings Plans

[Download CSV](#) [Add selected Savings Plan\(s\) to cart](#)

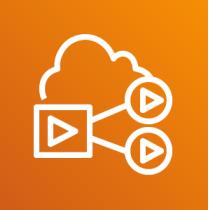
x	Term	Payment option	Recommended commitment	Estimated hourly savings
<input checked="" type="checkbox"/>	3-year	All upfront	\$2.40/hour	\$1.61 (40%)

*Average hourly spend and minimum hourly spend based on your current on-demand spend for the given instance family.

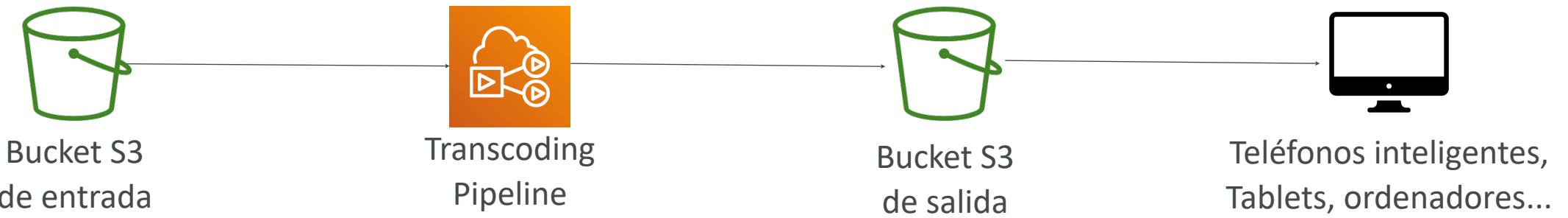
Cost Explorer – Previsión de uso



Amazon Elastic Transcoder



- Elastic Transcoder se utiliza para **convertir los archivos multimedia almacenados en S3 en archivos multimedia en los formatos requeridos por los dispositivos de reproducción de los consumidores (teléfonos, etc.)**
- Ventajas:
 - Fácil de usar
 - Altamente escalable - puede manejar grandes volúmenes de archivos multimedia y archivos de gran tamaño
 - Rentable: modelo de precios basado en la duración
 - Totalmente gestionado y seguro, paga por lo que usas

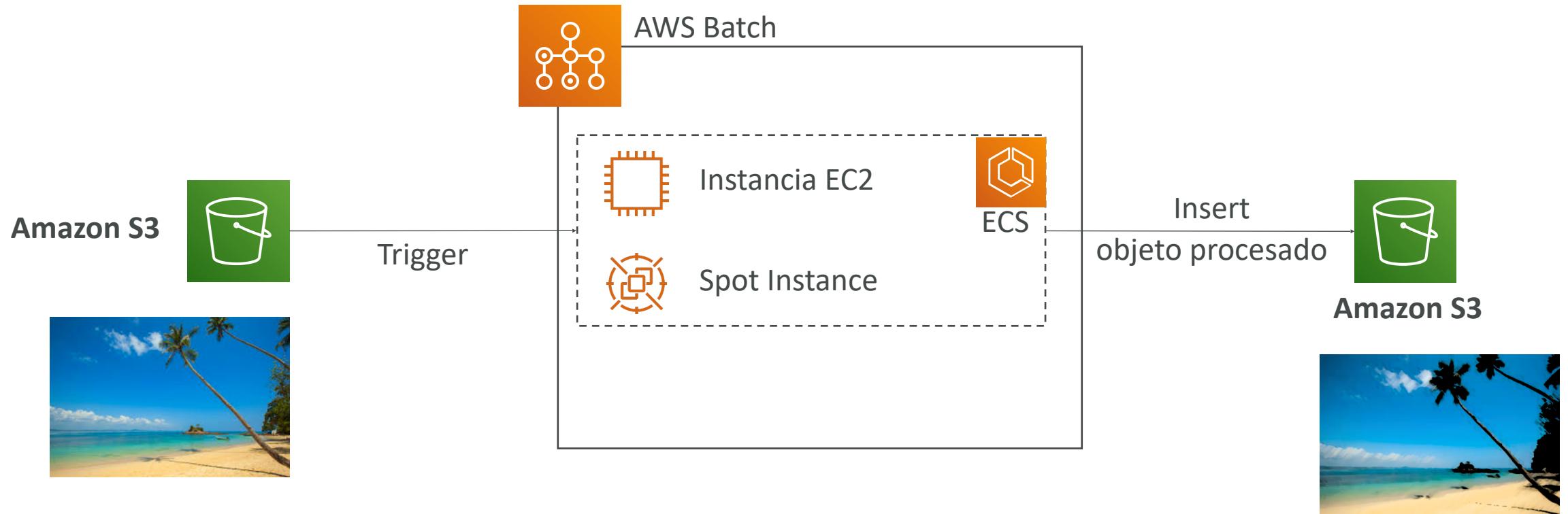


AWS Batch



- **Procesamiento por lotes** totalmente gestionado a **cualquier escala**
- Ejecuta eficientemente 100.000 trabajos de computación por lotes en AWS
- Un trabajo "por lotes" es un trabajo con un inicio y un final (en contraposición a uno continuo)
- Batch lanzará dinámicamente instancias **EC2 o Spot Instances**
- AWS Batch proporciona la cantidad adecuada de computación / memoria
- Tú envías o programas los trabajos por lotes y AWS Batch se encarga del resto
- Los trabajos por lotes se definen como **imágenes Docker** y se **ejecutan en ECS**
- Útil para optimizar los costes y centrarse menos en la infraestructura

AWS Batch – Ejemplo simplificado



Batch vs Lambda

- Lambda:
 - Límite de tiempo
 - Tiempos de ejecución limitados
 - Espacio de disco temporal limitado
 - Serverless
- Por lotes:
 - Sin límite de tiempo
 - Cualquier tiempo de ejecución siempre que esté empaquetado como imagen Docker
 - Depende de EBS / almacén de instancias para el espacio en disco
 - Depende de EC2 (puede ser gestionado por AWS)



Amazon AppFlow



- Servicio de integración totalmente gestionado que te permite transferir datos de forma segura entre aplicaciones de **software como servicio (SaaS) y AWS**
- Fuentes: **Salesforce**, SAP, Zendesk, Slack y ServiceNow
- Destinos: Servicios AWS como **Amazon S3, Amazon Redshift** o no AWS como SnowFlake y Salesforce
- Frecuencia: programada, en respuesta a eventos o bajo demanda
- Capacidades de transformación de datos como filtrado y validación
- Cifrado a través de Internet público o privado a través de AWS PrivateLink
- No pierdas tiempo escribiendo las integraciones y aprovecha las API inmediatamente

White Papers y Arquitecturas

Well Architected Framework, Recuperación en caso de desastre, etc.

Visión general de la sección

- Whitepaper de Well Architected Framework
- Herramienta Well Architected
- Asesor de confianza de AWS
- Recursos de arquitecturas de referencia (para el mundo real)
- Whitepaper sobre recuperación de desastres en AWS

Well Architected Framework (Marco de buena arquitectura)

Principios generales

- <https://aws.amazon.com/architecture/well-architected>
- Deja de adivinar tus necesidades de capacidad
- Prueba sistemas a escala de producción
- Automatiza para facilitar la experimentación arquitectónica
- Permite arquitecturas evolutivas
 - Diseña en función de los requisitos cambiantes
- Impulsa las arquitecturas utilizando datos
- Mejorar mediante días de juego
 - Simular aplicaciones para días de venta flash

Well Architected Framework

6 Pilares

- 1) Excelencia operativa
 - 2) Seguridad
 - 3) Fiabilidad
 - 4) Eficiencia del rendimiento
 - 5) Optimización de costes
 - 6) Sostenibilidad
-
- No son algo a equilibrar; ni compensaciones, son una sinergia

AWS Well-Architected Tool



- Herramienta gratuita para **revisar tus arquitecturas** según los 6 pilares de Well-Architected Framework y adoptar las **mejores prácticas de arquitectura**.
- ¿Cómo funciona?
 - Selecciona tu carga de trabajo y responde a las preguntas
 - Revisa tus respuestas comparándolas con los 6 pilares
 - Obtén asesoramiento: obtén vídeos y documentación, genera un informe, ve los resultados en un dashboards
- Echemos un vistazo <https://console.aws.amazon.com/wellarchitected>

A screenshot of the AWS Well-Architected Tool interface. The top navigation bar includes links for 'Well-Architected Tool' and 'Workloads'. Below the navigation is a search bar labeled 'Search by workload name'. A toolbar with buttons for 'Generate report', 'View details', 'Edit', 'Delete', and 'Define workload' is visible. A pagination control shows page 1 of 1. The main content area displays a table titled 'Workloads' with the following data:

Name	Overall status	High risks	Medium risks	Improvement status	Last updated
Internal Employee Portal	Answered	13	2	None	Nov 24, 2018 3:40 PM UTC-8
Mobile app - Android	Answered	9	1	None	Nov 24, 2018 3:43 PM UTC-8
Mobile app - iOS	Answered	0	1	None	Nov 24, 2018 3:49 PM UTC-8
Retail Website- EU	Unanswered	0	0	None	Nov 24, 2018 3:52 PM UTC-8
Retail Website- North America	Unanswered	0	0	None	Nov 24, 2018 3:19 PM UTC-8

© 2008 - 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved. [Privacy Policy](#) [Terms of Use](#)

<https://aws.amazon.com/blogs/aws/new-aws-well-architected-tool-review-workloads-against-best-practices/>

Trusted Advisor



- Sin necesidad de instalar nada - evaluación de alto nivel de la cuenta de AWS
- Analiza tus cuentas de AWS y proporciona recomendaciones en 5 categorías
 - **Optimización de costes**
 - **Rendimiento**
 - **Seguridad**
 - **Tolerancia a los fallos**
 - **Límites del servicio**

Checks

- ▶ ✓ **Amazon EBS Public Snapshots**

Checks the permission settings for your Amazon Elastic Block Store snapshots. 0 EBS snapshots are marked as public.
- ▶ ✓ **Amazon RDS Public Snapshots**

Checks the permission settings for your Amazon Relational Database Service snapshots. 0 RDS snapshots are marked as public.
- ▶ ✓ **IAM Use**

This check is intended to discourage the use of root access keys. At least one IAM user has been created for this account.

Trusted Advisor – Planes de soporte

7 CORE CHECKS (7 CONTROLES BÁSICOS)

Plan de soporte: Basic y Developer

- Permisos de buckets S3
- Security Groups – Puertos específicos sin restricciones
- Uso de IAM (un usuario IAM como mínimo)
- MFA en la cuenta root
- EBS Public Snapshots
- RDS Public Snapshots
- Service Quotas

FULL CHECKS (CONTROLES COMPLETOS)

Plan de soporte: Business y Enterprise

- Comprobaciones completas disponibles en las 5 categorías
- Posibilidad de establecer alarmas de CloudWatch cuando se alcanzan los límites
- Acceso programado mediante la AWS Support API

Más ejemplos de arquitecturas

- Hemos explorado los patrones de arquitecturas más importantes:
 - Clásico: EC2, ELB, RDS, ElastiCache, etc...
 - Sin servidor: S3, Lambda, DynamoDB, CloudFront, API Gateway, etc...
- Si quieres ver más arquitecturas de AWS:
 - <https://aws.amazon.com/architecture/>
 - <https://aws.amazon.com/solutions/>

Repaso del examen y consejos

Punto de control del estado de aprendizaje

- Veamos hasta dónde hemos llegado en nuestro viaje de aprendizaje
- <https://aws.amazon.com/certification/certified-solutions-architect-associate/>

La práctica hace al maestro

- Si eres nuevo en AWS, practica un poco gracias a este curso antes de lanzarte al examen
 - El examen recomienda que tengas uno o más años de experiencia práctica en AWS
 - ¡La práctica hace al maestro!
-
- Si te sientes abrumado por la cantidad de conocimientos que acabas de aprender, repásalos una vez más

Proceder por eliminación

- La mayoría de las preguntas se basarán en situaciones hipotéticas.
 - Para todas las preguntas, descarta las respuestas que sepas con certeza que son incorrectas
 - Para las respuestas restantes, entiende cuál tiene más sentido
-
- Hay muy pocas preguntas trampa
 - No le des demasiadas vueltas
 - Si una solución parece factible pero muy complicada, probablemente sea errónea

Hojea los Whitepapers de AWS

- Puedes leer algunos Whitepapers de AWS aquí:
 - Arquitectura para el Cloud: Mejores prácticas de AWS
 - Marco bien diseñado de AWS / AWS Well-architected framework
 - Recuperación de desastres de AWS (<https://aws.amazon.com/disaster-recovery/>)
- En general, hemos explorado todos los conceptos más importantes del curso
- ¡Nunca está de más echar un vistazo a los Whitepapers que te parezcan interesantes!

Lee las FAQ de cada servicio

- FAQ = Preguntas más frecuentes
- Ejemplo: <https://aws.amazon.com/vpc/faqs/>
- Las FAQ cubren muchas de las preguntas que se hacen en el examen
- Ayudan a confirmar tu comprensión de un servicio

Entra en la Comunidad AWS

- Ayudar y debatir con otras personas en las preguntas y respuestas del curso
 - Revisa las preguntas formuladas por otras personas en las preguntas y respuestas
 - Haz el examen práctico de esta sección
-
- Lee foros en línea
 - Lee blogs online
 - Asiste a reuniones locales y debate con otros ingenieros de AWS

¿Cómo funcionará el examen?

- Tendrás que inscribirte en línea en: <https://www.aws.training/>
- La tasa del examen es de 150 USD
- Proporciona un documento de identidad (DNI, Pasaporte, los detalles están en los correos electrónicos que te enviamos...)
- No se permiten notas, ni bolígrafo, ni hablar
- Se harán 65 preguntas en 130 minutos
- Utiliza la función "Marcar" para marcar las preguntas que quieras volver a revisar
- Al final puedes revisar opcionalmente todas las preguntas / respuestas

- Para aprobar necesitas una puntuación mínima de 720 sobre 1000
- Sabrás en un plazo de 5 días si has aprobado / suspendido los exámenes (la mayoría de las veces menos)
- Sabrás la puntuación global unos días después (notificación por correo electrónico)
- No sabrás qué respuestas eran correctas / incorrectas
- Si suspendes, puedes volver a hacer el examen 14 días después

¡Enhорabuena!

¡Enhorabuena!

- ¡Enhorabuena por haber terminado el curso!
- Espero que apruebes el examen sin problemas ☺.
- Si aún no lo has hecho, ¡me encantaría que me dieras tu opinión!
- Si has aprobado, me alegrará saber que te he ayudado
 - Públícalo en Preguntas y Respuestas para ayudar y motivar a otros estudiantes. ¡Comparte tus consejos!
 - ¡Públícalo en LinkedIn y etiquétame!
- En general, espero que hayas aprendido a utilizar AWS y que seas un Arquitecto de Soluciones AWS tremendamente bueno