



AWS Certified Data Engineer



Two dark blue rectangular call-to-action buttons. The top button features a play icon inside a video camera frame and the text "Vídeo curso" followed by a right-pointing arrow. The bottom button features a checklist icon with a pen and the text "Exámenes de práctica" followed by a right-pointing arrow.



Joan Amengual

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 de AWS Certified Data Engineer - 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!

Tabla de contenidos

- [Introducción al curso](#)
- [Identity and Access Management \(IAM\)](#)
- [Almacenamiento de EC2](#)
- [Computación](#)
- [Gestión y gobierno](#)
- [S3 \(Simple Storage Service\)](#)
- [S3 Avanzado](#)
- [Almacenamiento avanzado en AWS](#)
- [Introducción a las bases de datos de AWS](#)
- [Amazon RDS](#)
- [Amazon Aurora](#)

Tabla de contenidos

- [Amazon DynamoDB](#)
- [Amazon Redshift](#)
- [Otras bases de datos](#)
- [Route 53](#)
- [CloudFront](#)
- [Docker en AWS](#)
- [AWS Glue](#)
- [Integración de aplicaciones](#)
- [Amazon Kinesis](#)
- [Amazon Athena](#)

Tabla de contenidos

- [Apache Spark](#)
- [Amazon EMR](#)
- [Amazon QuickSight](#)
- [Amazon AppFlow](#)
- [Amazon MWAA](#)
- [Amazon MSK](#)
- [Amazon OpenSearch Service](#)
- [Preparación del examen de certificación](#)
- [¡Enhorabuena!](#)



AWS Certified Data Engineer

www.blockstellart.com

Todos los derechos reservados © BLOCKSTELLART www.blockstellart.com

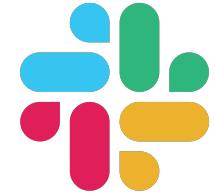
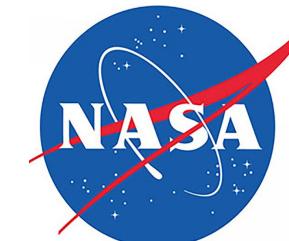
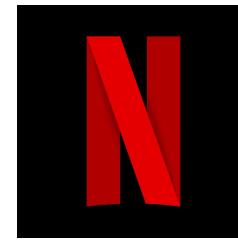
¡Es un placer tenerte en esta formación!

- Vamos a empezar la formación sobre Amazon Web Services (AWS)
- Los conocimientos básicos de IT son útiles, pero se explicará todo
- Se van a cubrir **más de 50 servicios de AWS**
 - Tómate tu tiempo, no es una carrera
- **Aprender haciendo - ¡técnica clave de aprendizaje!**
Este curso combina la teoría y la práctica



¿Qué es AWS?

- AWS 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
 - Netflix, Airbnb, NASA, Twitch, Samsung, Slack...



Sobre el instructor

Joan Amengual

- Ingeniero Full Stack en una empresa tecnológica en Silicon Valley, USA
- He trabajado con AWS, Azure y Google Cloud varios años en diversas empresas para la migración y el escalado de servicios en el Cloud
- Premiado como Joven Talento en Ingeniería
- Puedes encontrarme en:
 - LinkedIn: <https://www.linkedin.com/in/joanamengual7>
 - Frecuentemente hago publicaciones interesantes sobre el Cloud



Comunidad - Cloud Computing



LinkedIn



Discord



Telegram

<https://blockstellart.com/comunidades>

Identity and Access Management (IAM)

Usuarios y grupos en IAM

- 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
 - Cada usuario de IAM puede tener permisos individuales asignados directamente. Esto permite un control granular del acceso a los recursos de AWS.
- Los **grupos** sólo contienen usuarios, no otros grupos
 - Agrupar usuarios con requisitos de acceso similares facilita la administración de permisos y mejora la seguridad al reducir la posibilidad de errores en la asignación de permisos.
- Los usuarios no tienen que pertenecer a un grupo, y el usuario puede pertenecer a varios grupos

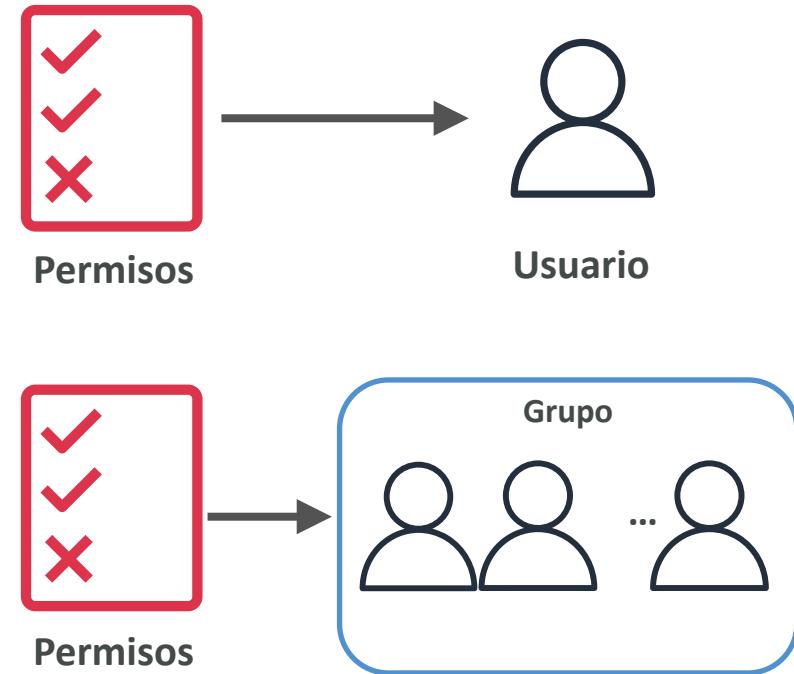
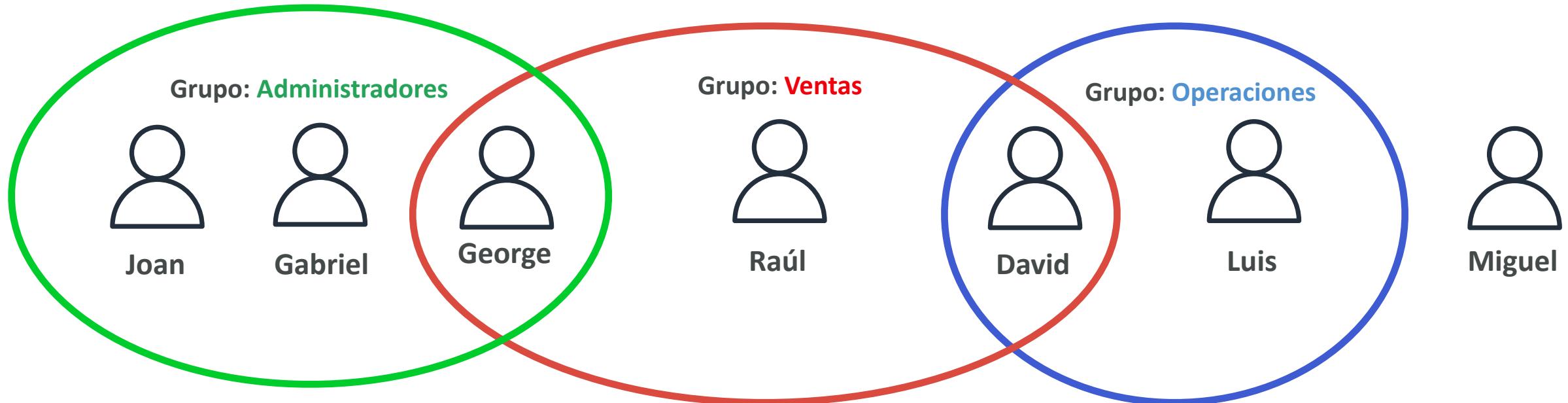


Diagrama de usuarios y grupos de IAM



- **George** y **David** pertenecen a dos grupos al mismo tiempo
- Miguel no pertenece a ningún grupo (no es una acción obligatoria)
 - Es común esta práctica para asignar permisos específicos individualizados

Buenas prácticas - Usuarios y grupos en IAM

-  Regularmente **revisar y eliminar usuarios IAM que ya no necesitan acceso**, para minimizar el riesgo de acceso no autorizado
-  Utilizar **etiquetas (tags) para organizar usuarios y grupos** por departamento, función u otros criterios, facilitando la gestión y el seguimiento de accesos
-  Asigna a los usuarios y grupos solo los permisos necesarios para realizar sus tareas. **Evita otorgar permisos excesivos que puedan llevar a riesgos de seguridad**

IAM - Políticas para definir permisos y privilegios

- 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

| Políticas (1186) Información | | | | |
|----------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|------------------|-------------|------------------------------------------------|
| C Acciones ▾ Eliminar Crear política | | | | |
| Una política es un objeto en AWS que define permisos. | | | | |
| Filtrar por Tipo | | | | |
| <input type="text"/> Buscar Todos los... ▾ | | | | |
| < 1 2 3 4 5 6 7 ... 60 > ⚙️ | | | | |
| Nombre de la política ▲ Tipo ▼ Usad... ▼ Descripción | | | | |
| <input type="radio"/> |  AccessAnalyzerSer... | Administrada ... | Política... | Allow Access Analyzer to analyze resource ... |
| <input type="radio"/> |  AdministratorAccess | Administrada ... | Política... | Provides full access to AWS services and re... |
| <input type="radio"/> |  AdministratorAcce... | Administrada ... | Ninguno | Grants account administrative permissions ... |

Ejemplo - Definición de política en IAM

- Las políticas en AWS IAM establecen **acciones permitidas en servicios específicos**
- Si una acción no está explícitamente permitida por una política, se considera prohibida
- Las políticas pueden tener **diferentes versiones**
 - Permite gestionar y revisar los cambios en las políticas a lo largo del tiempo, manteniendo versiones anteriores para referencia o revertir a una configuración anterior si es necesario

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

Asignación de políticas en IAM

Herencia de políticas

- Los nuevos usuarios **heredan automáticamente permisos al ser añadidos a un grupo** con políticas predefinidas
- Facilita la administración al cambiar permisos en un solo lugar que afecta a todos los usuarios del grupo

Políticas en línea

- Se adjuntan directamente a un único recurso o usuario en el sistema IAM
- Permiten permisos detallados y personalizados **para un caso de uso específico**



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 (*statement*) constan de:
 - **Sid**: un identificador para la declaración (opcional)
 - **Effect**: si la sentencia permite o deniega el acceso (Allow y Deny)
 - **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": "PolicyID12345",  
  "Statement": [  
    {  
      "Sid": "1",  
      "Effect": "Allow",  
      "Principal": {  
        "AWS": ["arn:aws:iam::123456789012:user/MyUser"]  
      },  
      "Action": [  
        "s3:GetObject",  
        "s3:PutObject",  
        "s3:DeleteObject"  
      ],  
      "Resource": [  
        "arn:aws:s3:::mybucket/*",  
        "arn:aws:s3:::mybucket2/*"  
      ]  
    }  
  ]  
}
```

Autenticación Multifactor (MFA)

- **MFA es un procedimiento en el que se solicita a los usuarios durante el proceso de inicio de sesión una forma adicional de identificación**
 - Por ejemplo: un código en el teléfono móvil o una lectura de la huella digital
- **Principal beneficio de MFA:** si una contraseña es robada o hackeada, la cuenta no se ve comprometida
- **Quieres proteger tu cuenta root y los usuarios de IAM**



Autenticación Multifactor (MFA)

- El funcionamiento de MFA se basa en exigir uno o varios de los siguientes métodos de autenticación:



Algo que conoces

Normalmente una contraseña



Algo que tienes

Como un dispositivo de confianza que no se puede duplicar con facilidad



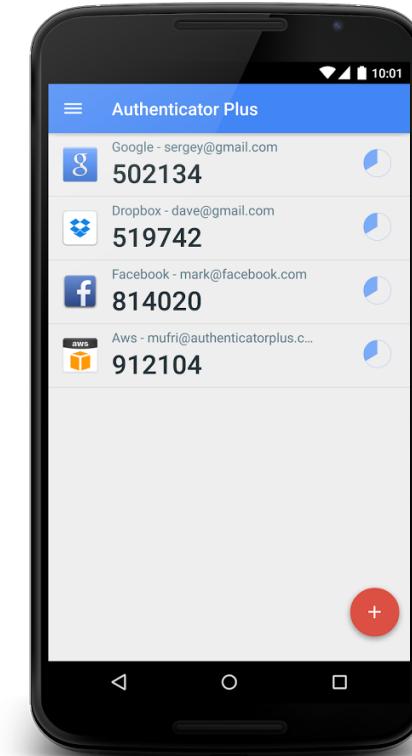
Algo que forma parte de ti

Información biométrica como una huella digital o una detección de rostro

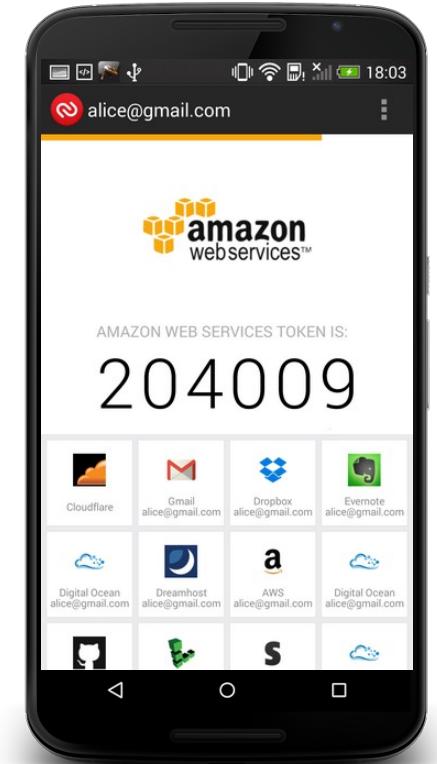
Opciones de dispositivos MFA en AWS

Dispositivo virtual MFA

- Un dispositivo virtual MFA es una **aplicación de software que se instala en un dispositivo inteligente**, como un teléfono móvil o una tableta, y se utiliza para **generar códigos de autenticación de dos factores (2FA)**
- No está diseñado únicamente para AWS, puede usarse para otras aplicaciones



Autenticador de Google
(sólo en el teléfono)



Authy
(multi-dispositivo)

Opciones de dispositivos MFA en AWS

Clave de seguridad U2F

- Este dispositivo genera un **código de seguridad único y temporal**, que el usuario debe ingresar para acceder a una cuenta o sistema protegido
- Estos dispositivos aumentan la seguridad al **requerir algo físico que el usuario debe tener en su posesión, además de su contraseña habitual**



Yubico - YubiKey 5 NFC



Yubico YubiKey 5C

Opciones de dispositivos MFA en AWS

Dispositivo MFA por hardware

- Dispositivo físico que **genera tokens de seguridad** para el acceso a AWS
- Al requerir **tanto una contraseña como el código generado por este dispositivo**, ayuda a asegurar que solo los usuarios autorizados puedan acceder a los recursos de AWS



SafeNet IDProve 110 6-digit OTP Token for
Use with Amazon Web Services Only

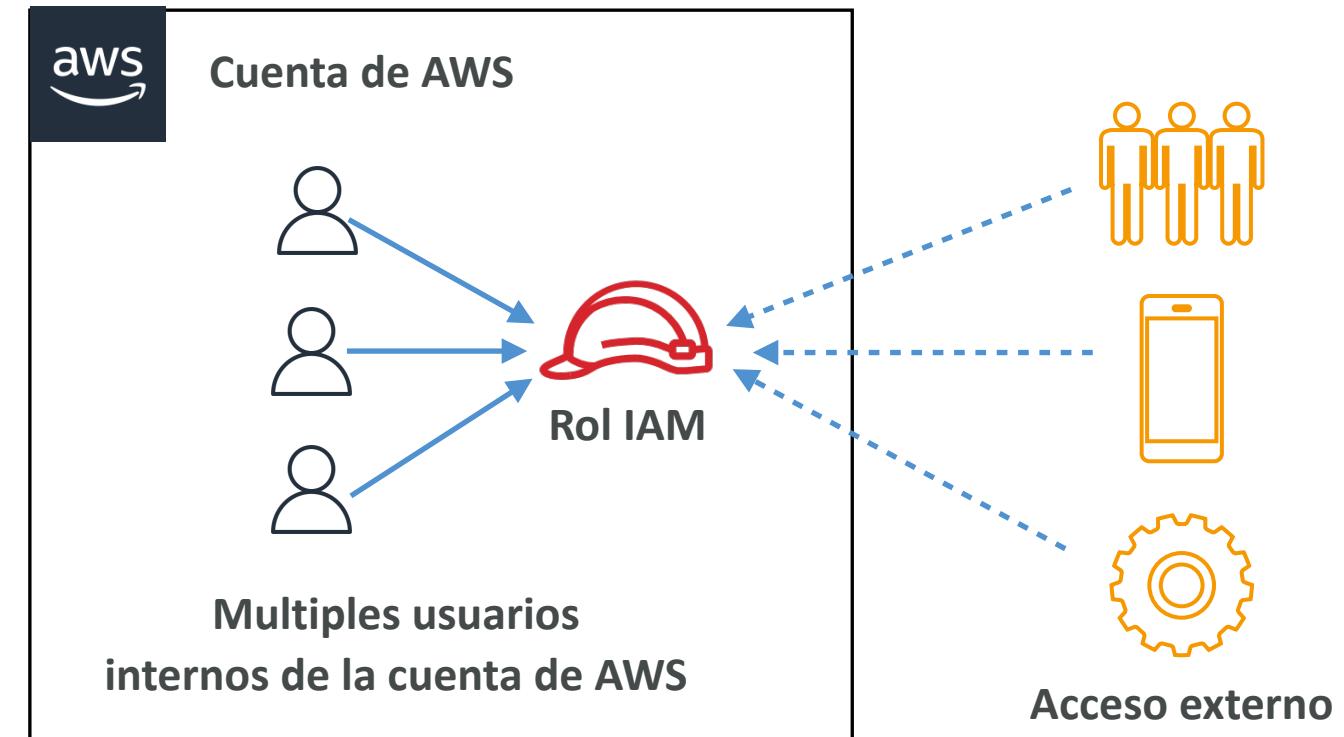
Visión general de los roles IAM

- Los roles IAM se definen como identidades temporales para otorgar permisos específicos, sin credenciales permanentes
 - Facilitan la asignación segura y temporal de permisos
 - Control granular sobre accesos mediante políticas actualizables
 - Permiten acceso mediante identidades externas sin crear usuarios IAM
- Usuarios vs. Roles de IAM:
 - **Usuarios IAM:** Accesos permanentes, personales o de servicio con credenciales estáticas
 - **Roles de IAM:** Accesos temporales sin asignación directa, para flexibilidad y seguridad en accesos cruzados y federación



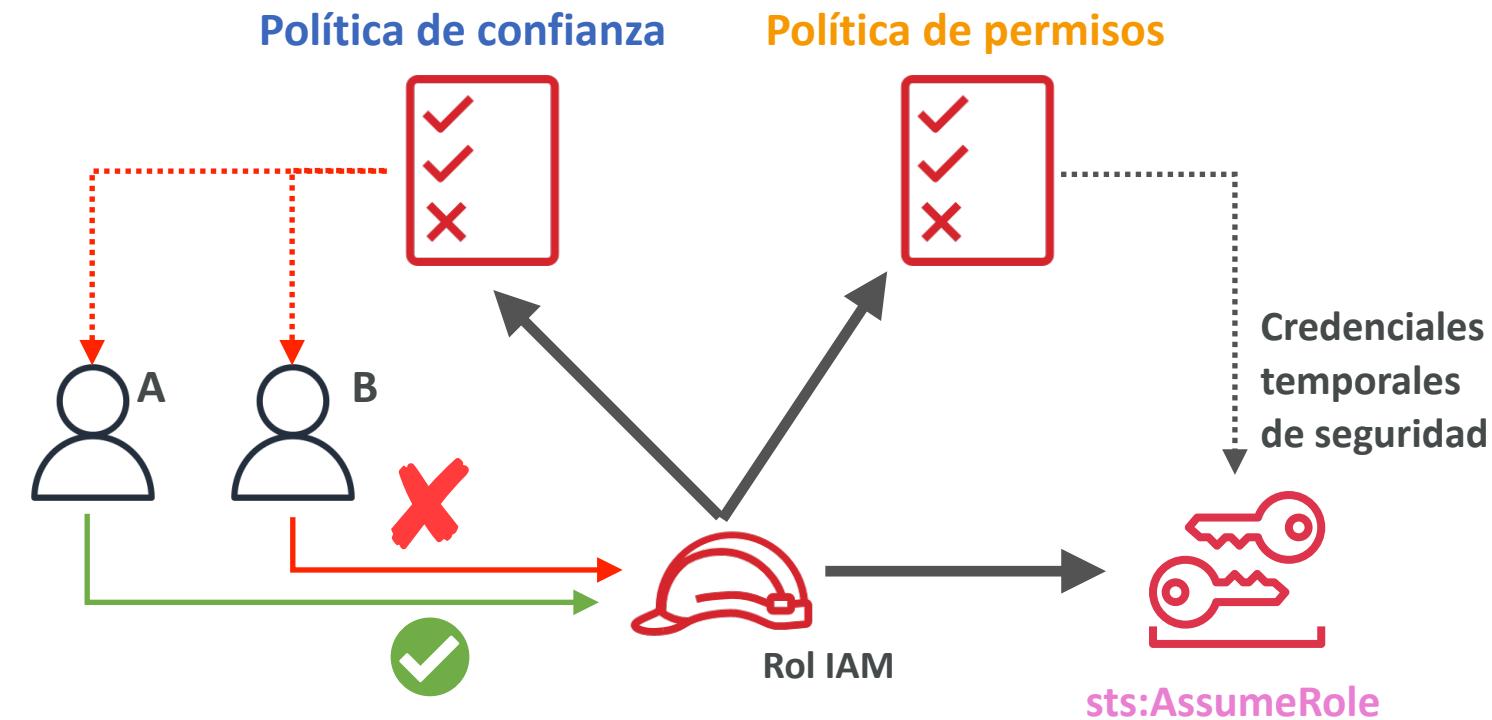
Detalles de los roles IAM

- Un rol de IAM...
 - puede ser asumido por cualquier usuario autorizado dentro de la misma cuenta de AWS y usuarios / dispositivos externos.
 - es crucial para operaciones entre cuentas de AWS, permitiendo que usuarios de una cuenta accedan a recursos de otra cuenta de forma controlada.
 - son ideales para la integración con servicios de identidad externos, permitiendo a usuarios no-AWS asumir roles temporalmente para acceder a recursos de AWS.



Funcionamiento de los roles IAM

- La **política de confianza** determina **quién puede asumir el rol**. Y la **política de permisos** especifica las **acciones permitidas**
- El proceso de asumir un rol incluye la autenticación de la identidad y la autorización para usar el rol con base en las políticas establecidas
- Utilizar **sts:AssumeRole** para obtener credenciales temporales refuerza una seguridad dinámica y reduce el riesgo de compromiso de credenciales permanentes.



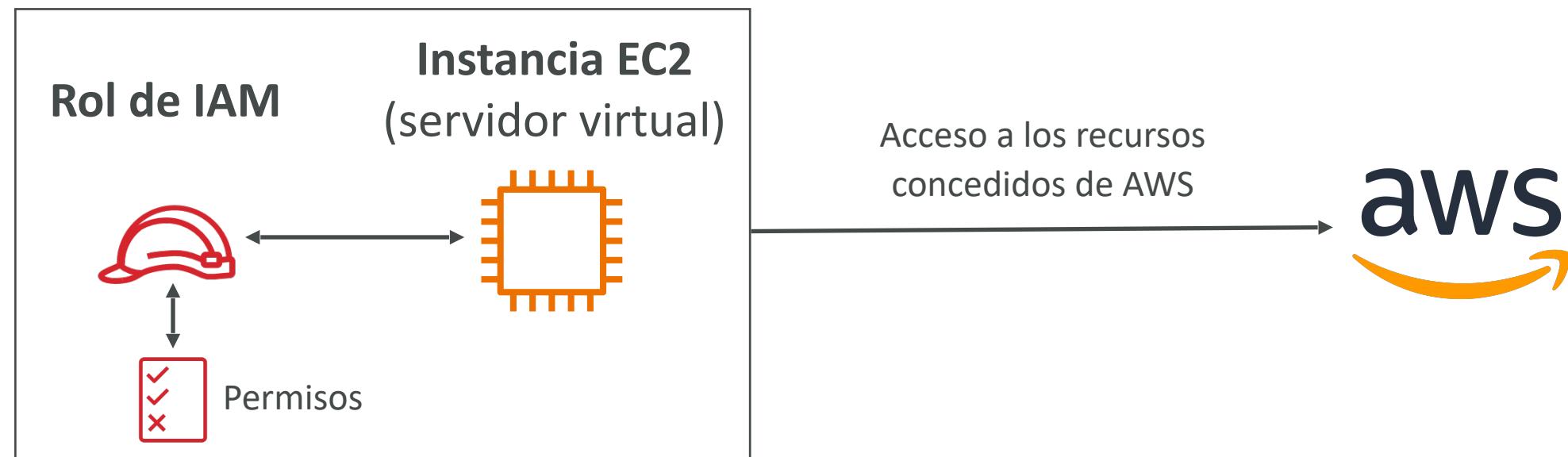
¿Los roles sólo pueden aplicarse a usuarios?

- Los roles también pueden aplicarse a servicios de AWS
 - Algún servicio de AWS tendrá que realizar acciones en tu nombre
 - Para ello, podemos asignar **permisos** a los servicios de AWS con **roles IAM**
-
-  **RECOMENDACIÓN:**
 - Asigna únicamente los permisos necesarios para las tareas a realizar



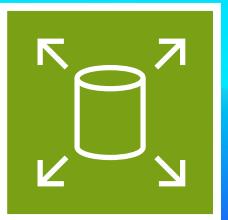
Caso práctico: Roles aplicados a Instancias EC2

- **Caso práctico:** Una empresa utiliza roles de IAM para permitir que sus instancias EC2 desplieguen aplicaciones automáticamente, asegurándose de que cada instancia solo tenga acceso a los recursos que necesita para su función específica



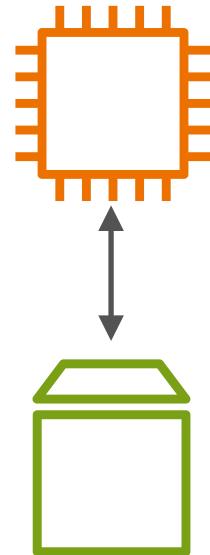
Almacenamiento de EC2

¿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
- Están vinculados a **una zona de disponibilidad específica**
- Piensa en ellos como una "memoria USB de red"

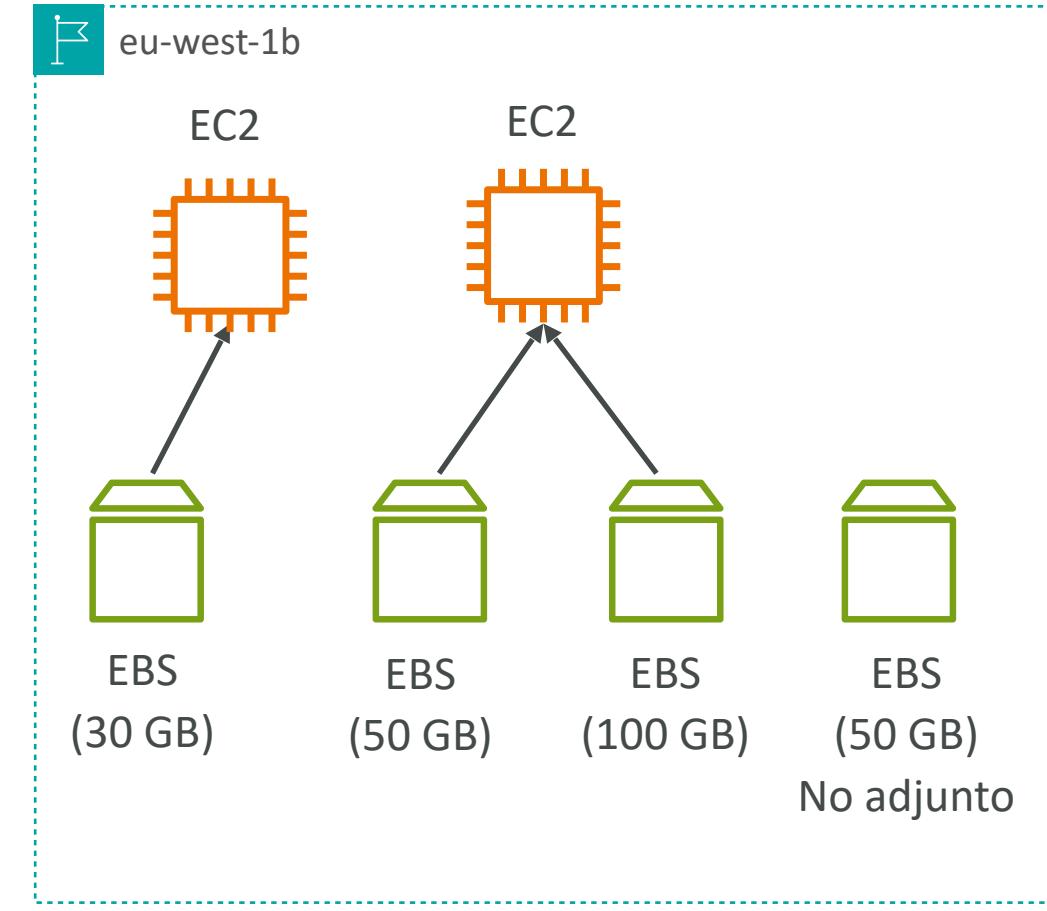
Instancia EC2
(servidor virtual)



Volumen EBS

Detalles importantes de los volúmenes 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 eu-west-1a no puede adjuntarse a eu-west-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



Configuración de volúmenes EBS

- El volumen EBS se puede configurar **durante la creación de la instancia EC2**
- Los parámetros a configurar son:
 - Tamaño (GiB), tipo de volumen, IOPS, Eliminar volumen al terminar instancia EC2, cifrado, clave KMS, rendimiento
 - Por defecto, se elimina el volumen EBS root / raíz (atributo habilitado)

▼ Storage (volumes) [Información](#) [Simple](#) [Ocultar detalles](#)

Volumenes de EBS

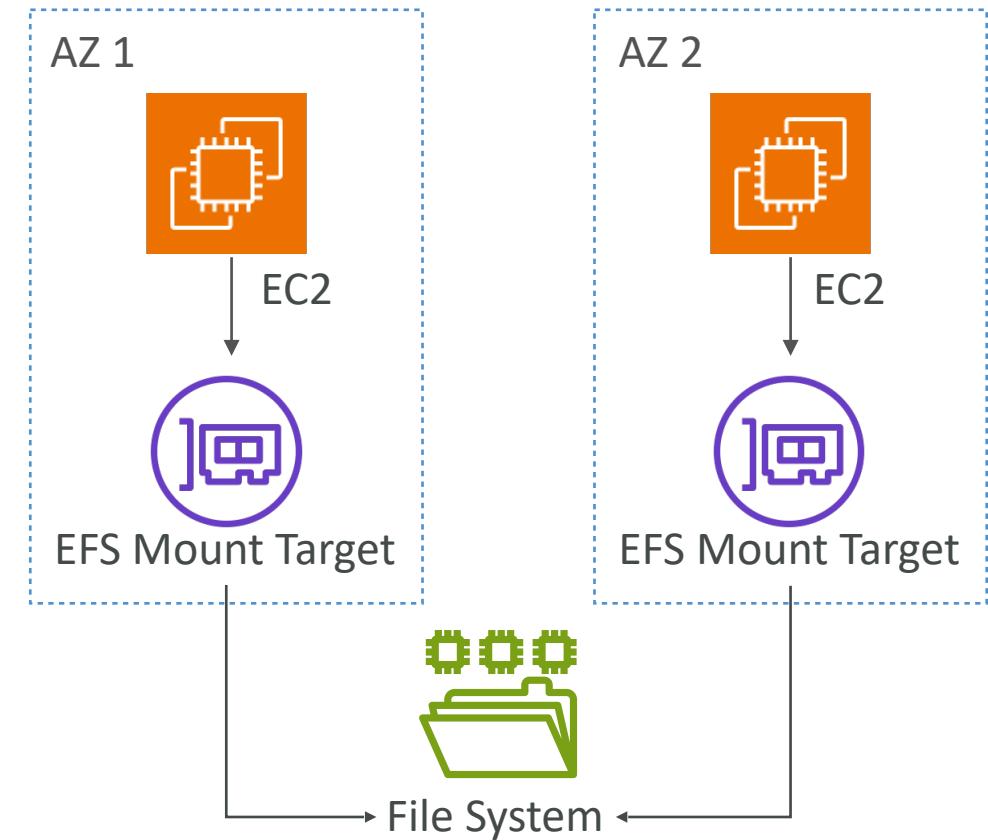
▼ Volumen 1 (Raíz de AMI)

| | | |
|-----------------------------------------------------------|--------------------------------------------------------------------------------------|-------------------------------------------------------------------|
| Tipo de almacenamiento Información EBS | Nombre del dispositivo - <i>obligatorio</i> Información /dev/xvda | Instantánea Información snap-015ca71e26658eb10 |
| Tamaño (GiB) Información 8 | Tipo de volumen Información gp3 | IOPS Información 3000 |
| Eliminar cuando termine Información Sí | Cifrado Información No cifrado | Clave de KMS Información Seleccionar |
| Rendimiento Información 125 | Las claves de KMS solo se aplican cuando se establece el cifrado en este volumen. | |

Free tier eligible customers can get up to 30 GB of EBS General Purpose (SSD) or Magnetic storage [X](#)

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, más caro, pago por uso
- Amazon EFS está **diseñado para ampliarse a petabytes** según la demanda sin interrumpir las aplicaciones



Clases de almacenamiento de Amazon EFS

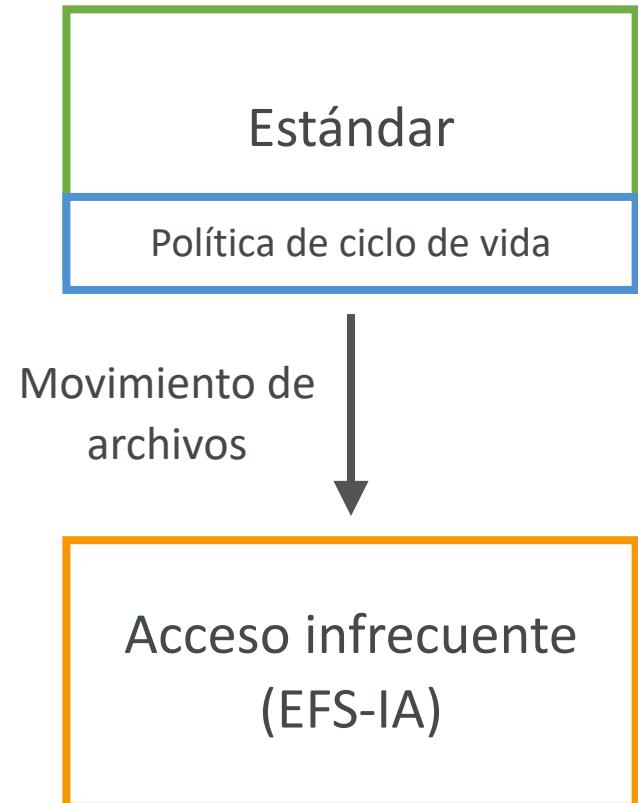
- **Niveles de almacenamiento (función de gestión del ciclo de vida: mover el archivo después de N días)**

- 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

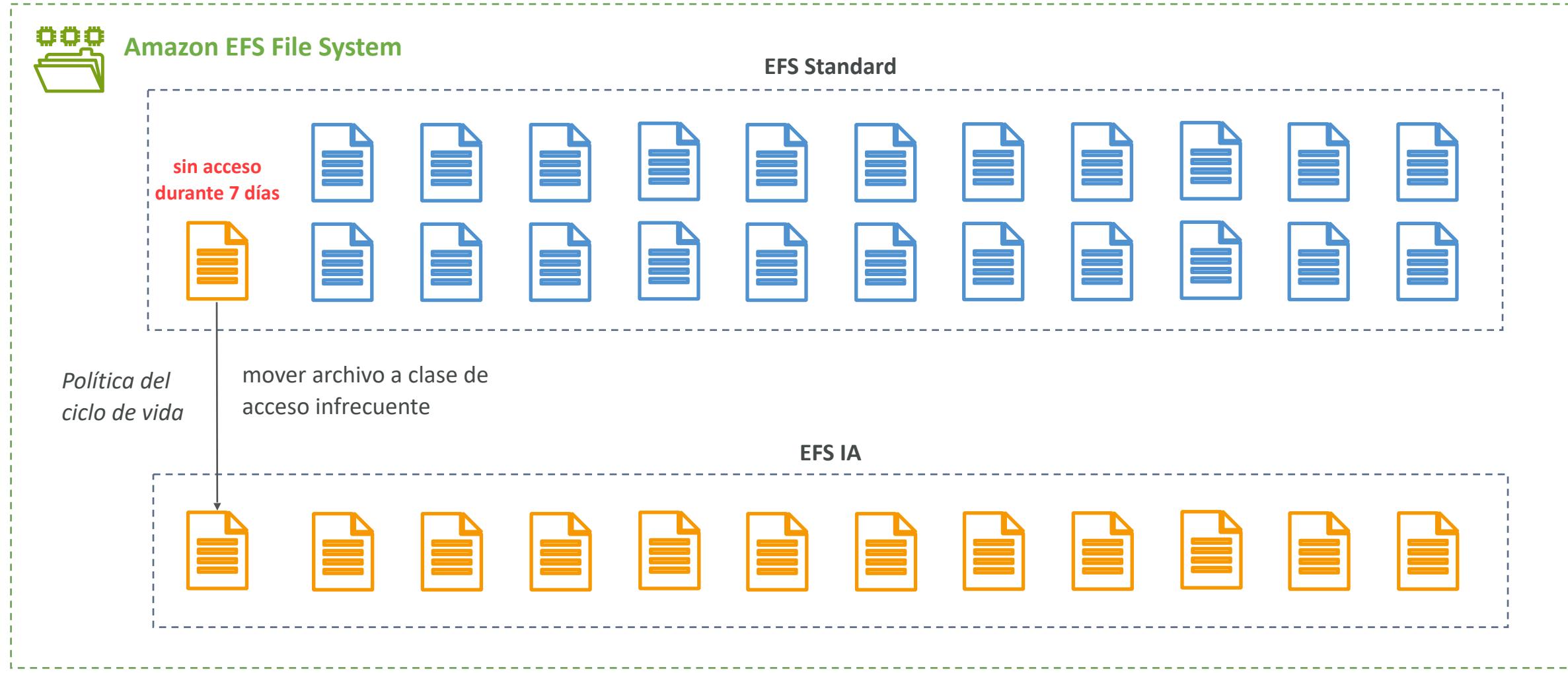
- **Disponibilidad y durabilidad**

- 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

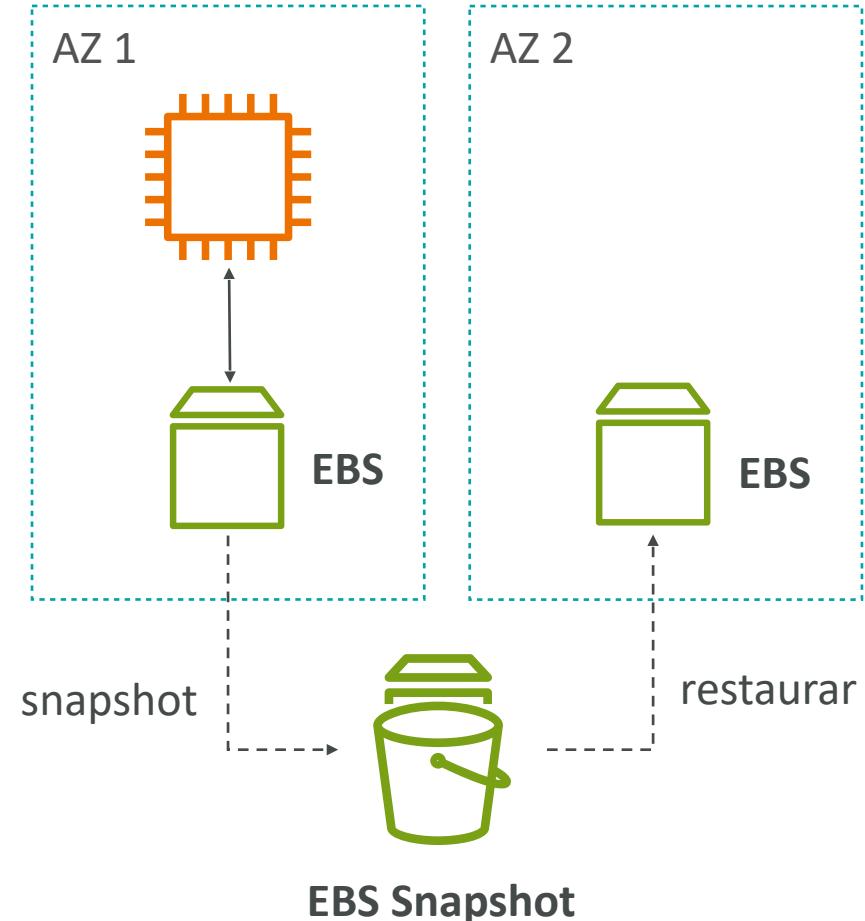


Clases de almacenamiento de Amazon EFS



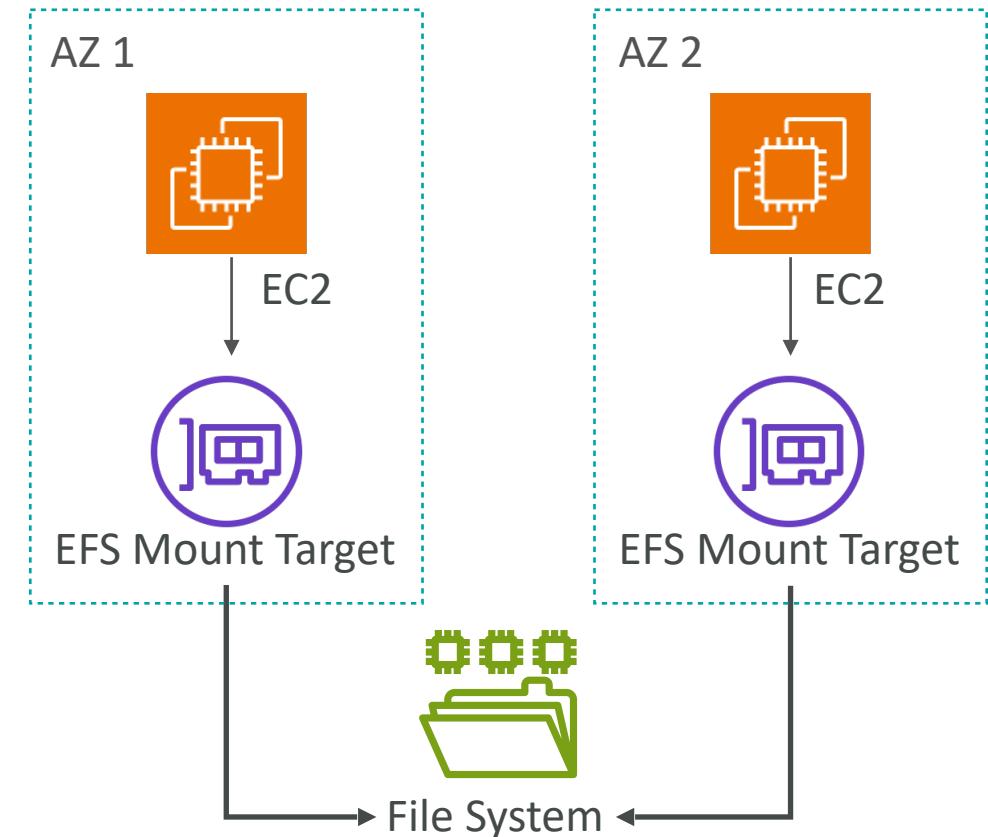
EBS vs EFS – Elastic Block Storage

- Los volúmenes EBS sólo pueden adjuntarse a una instancia a la vez o usar EBS Multi-Attach y están bloqueados a nivel de AZ
 - **Multi-Attach es compatible exclusivamente en volúmenes de SSD de IOPS aprovisionadas (io1 y io2 Block Express)**
- Para migrar un volumen EBS a través de la AZ:
 - Haz una snapshot
 - Restaura la snapshot en otra AZ
- 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

- EFS está diseñado para ser altamente disponible y duradero
- Permite el acceso concurrente a miles de instancias de EC2
- El almacenamiento en EFS **escala automáticamente hacia arriba o hacia abajo** a medida que se agregan o eliminan archivos
- EFS elimina la necesidad de administrar la infraestructura de almacenamiento subyacente y es más fácil de configurar y administrar en comparación con EBS
- Sólo para instancias Linux (POSIX)
- **EFS tiene un precio más elevado que EBS**





Computación

www.blockstellart.com

Todos los derechos reservados © BLOCKSTELLART www.blockstellart.com

Instancias EC2 en el Big Data

- EC2 permite escalar recursos de computación de manera elástica según la demanda de procesamiento de Big Data
 - **Instancias spot:** Ideales para tareas que pueden tolerar interrupciones. Ofrecen bajo costo y son adecuadas para procesos con puntos de control frecuentes, como en aprendizaje automático (ML)
 - **Instancias reservadas:** Recomendadas para clusters y bases de datos de largo plazo (más de un año) debido a su estabilidad y costo predecible
 - **Instancias bajo demanda:** Utilizadas para las cargas de trabajo que no encajan en las categorías anteriores, proporcionando flexibilidad sin compromisos a largo plazo
- EC2 se integra sin problemas con otros servicios de AWS como Amazon S3 para almacenamiento, Amazon RDS y DynamoDB para bases de datos, y Amazon EMR para procesamiento de datos distribuidos



Procesadores AWS Graviton

- AWS Graviton es una **familia de procesadores** diseñada para ofrecer el mejor precio y rendimiento para sus cargas de trabajo en la nube que se ejecutan en Amazon EC2
 - Propósito general: M7, T4
 - Optimizado para cómputo: C7, C6
 - Optimizado para memoria: R7, X2
 - Optimizado para almacenamiento: Ix4, Is4
 - Cómputo acelerado (inferencia ML): G5
- **Mejor rendimiento por mejor precio**
- Opción para muchos servicios de ingeniería de datos:
 - MSK, RDS, MemoryDB, ElastiCache, OpenSearch, EMR, Lambda, Fargate



Fork Media logra reducir los costos un 40 % con bases de datos que se ejecutan en procesadores AWS Graviton2



Datadog utilizó AWS Graviton para ofrecer más valor y mantener los costos constantes

Gestión y gobierno

Por qué es importante la monitorización

- Nuestras aplicaciones se despliegan, y a nuestros usuarios no les importa cómo lo hicimos...
- ¡A nuestros usuarios sólo les importa que la aplicación funcione!
 - Latencia de la aplicación: ¿aumentará con el tiempo?
 - Caídas de la aplicación: la experiencia del cliente no debe degradarse
 - Que los usuarios se pongan en contacto con el departamento de IT o se quejen no es un buen resultado
- Supervisión interna:
 - ¿Podemos prevenir los problemas antes de que ocurran?
 - Rendimiento y coste
 - Tendencias (patrones de escalado)
 - Aprendizaje y mejora



Monitorización en AWS

- **AWS CloudWatch:**

- Métricas: Recopila y realiza un seguimiento de las métricas clave
- Logs: Recopila, monitoriza, analiza y almacena logs
- Eventos: Envía notificaciones cuando ocurran determinados eventos
- Alarmas: Reacciona en tiempo real ante métricas / eventos



- **AWS X-Ray:**

- Solución de problemas de rendimiento y errores de la aplicación
- Rastreo distribuido de microservicios

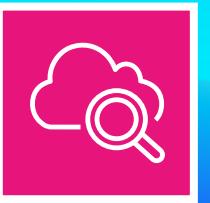


- **AWS CloudTrail:**

- Monitorización interna de las llamadas a la API que se realizan
- Auditoría de los cambios realizados por tus usuarios



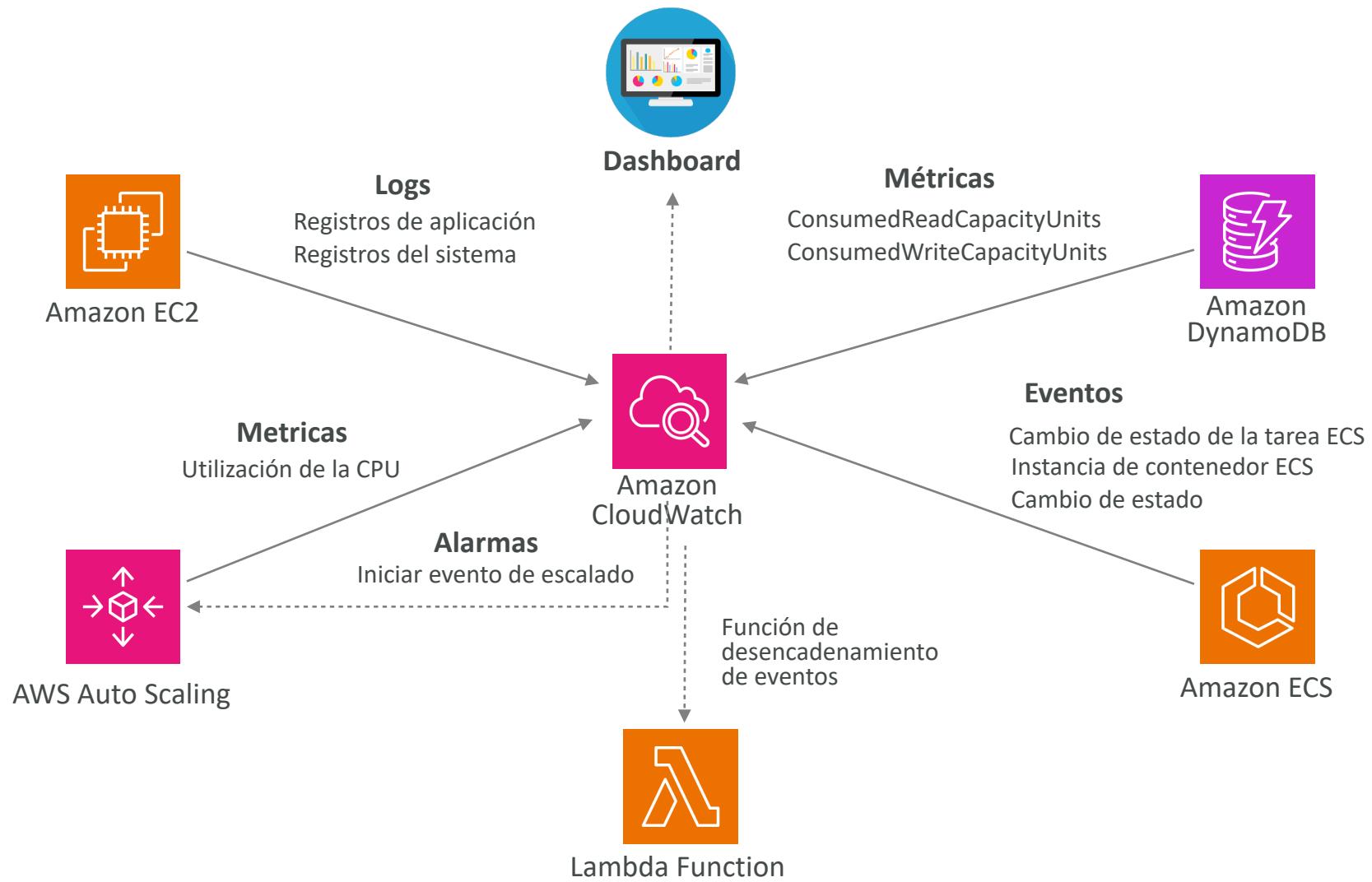
Amazon CloudWatch



- **CloudWatch proporciona monitoreo en tiempo real de recursos y aplicaciones de AWS**
- CloudWatch proporciona métricas para todos los servicios de AWS
- Algunas métricas destacables de CloudWatch son:
 - CPUUtilization
 - NetworkIn/NetworkOut
 - DiskReadOps/DiskWriteOps
 - Latency
- Las métricas tienen **marcas de tiempo**
- Puedes crear **dashboards de CloudWatch** con las métricas procedentes de CloudWatch

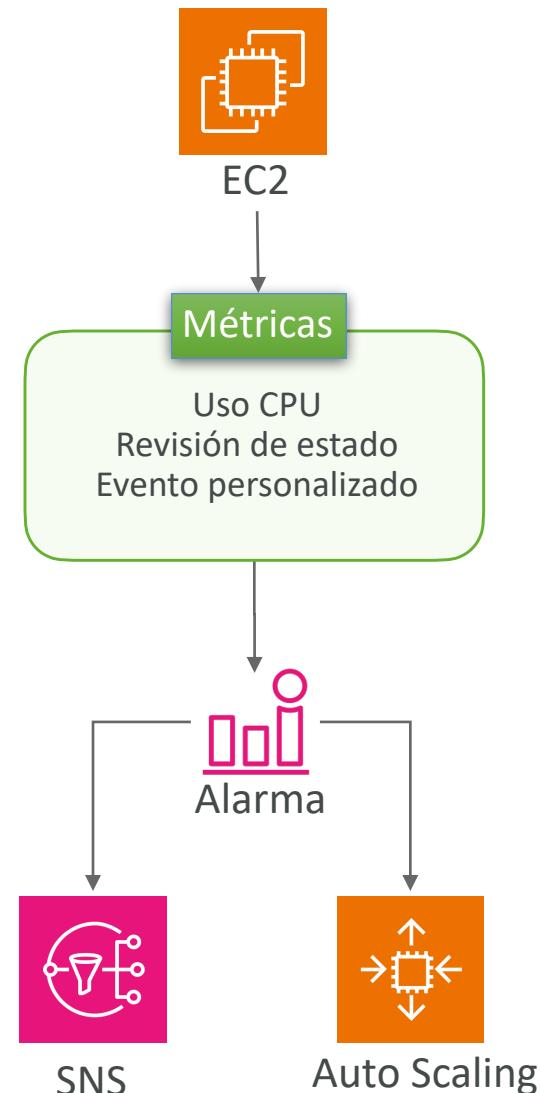


Visión general de Amazon CloudWatch



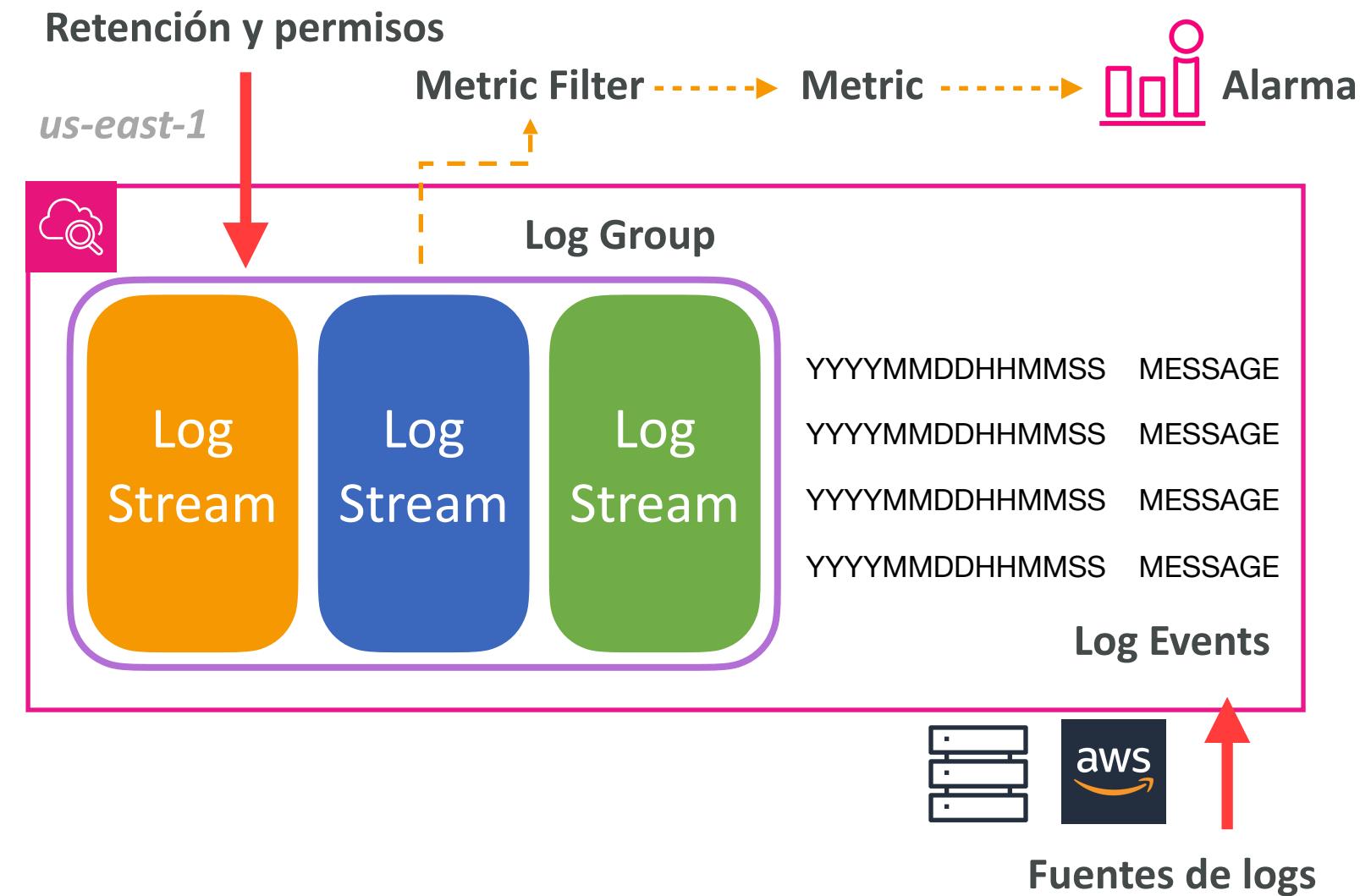
Monitorización detallada de EC2

- Las métricas de las instancias EC2 tienen métricas "cada 5 minutos"
- Con la monitorización detallada (por un coste), obtienes datos "cada 1 minuto"
- ¡Utiliza la monitorización detallada si quieres escalar más rápido tu ASG!
- La capa gratuita de AWS nos permite disponer de 10 métricas de monitorización detallada
- Nota: el uso de memoria de EC2 no se envía por defecto (debe enviarse desde dentro de la instancia como métrica personalizada)



Amazon CloudWatch Logs

- CloudWatch Logs permite **centralizar los logs de aplicaciones y servicios en AWS**, facilitando su gestión y análisis en un único lugar.
- **Monitorización de logs en tiempo real**, permitiendo identificar y responder rápidamente a los problemas
- CloudWatch Logs puede recoger logs de aplicaciones, contenedores, funciones, instancias, consultas DNS...
- Retención de logs de CloudWatch ajustable



Registros de CloudWatch (Logs)

- Puede definir políticas de expiración de logs (nunca expiran, 30 días, etc..)
- **CloudWatch Logs puede enviar logs a destinos como:**

- Amazon S3 (exportaciones)
- Flujos de datos de Kinesis
- Kinesis Data Firehose
- AWS Lambda
- Etc...



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



- API Gateway



- CloudTrail basado en filtro

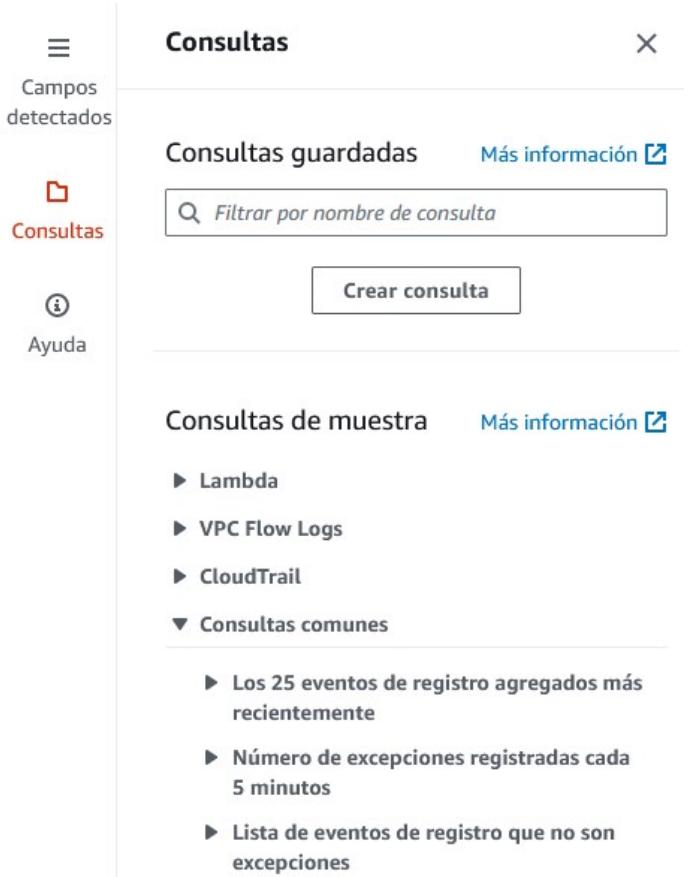


- Route 53 (DNS)



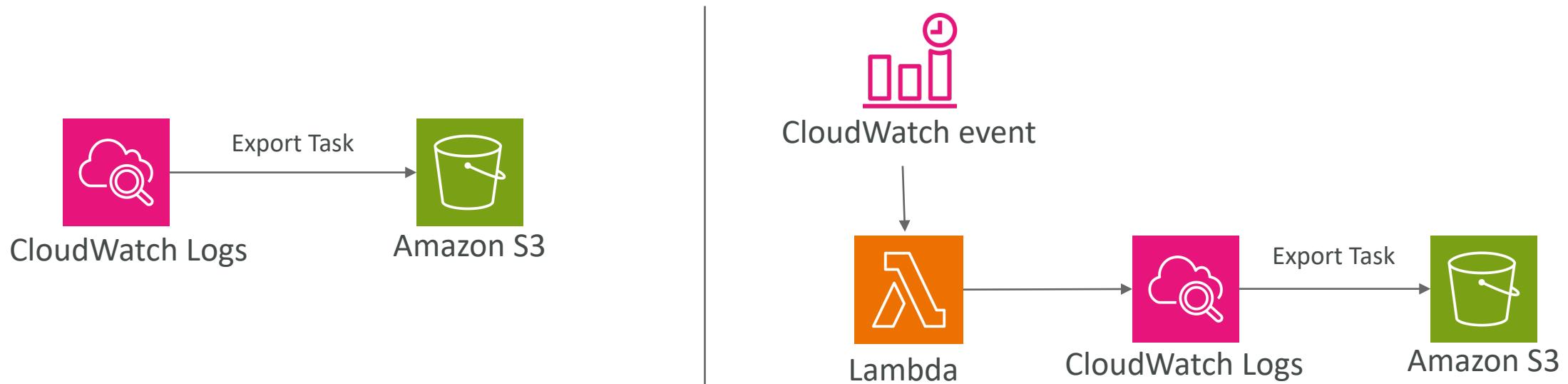
Filtro de métricas e información de CloudWatch Logs

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



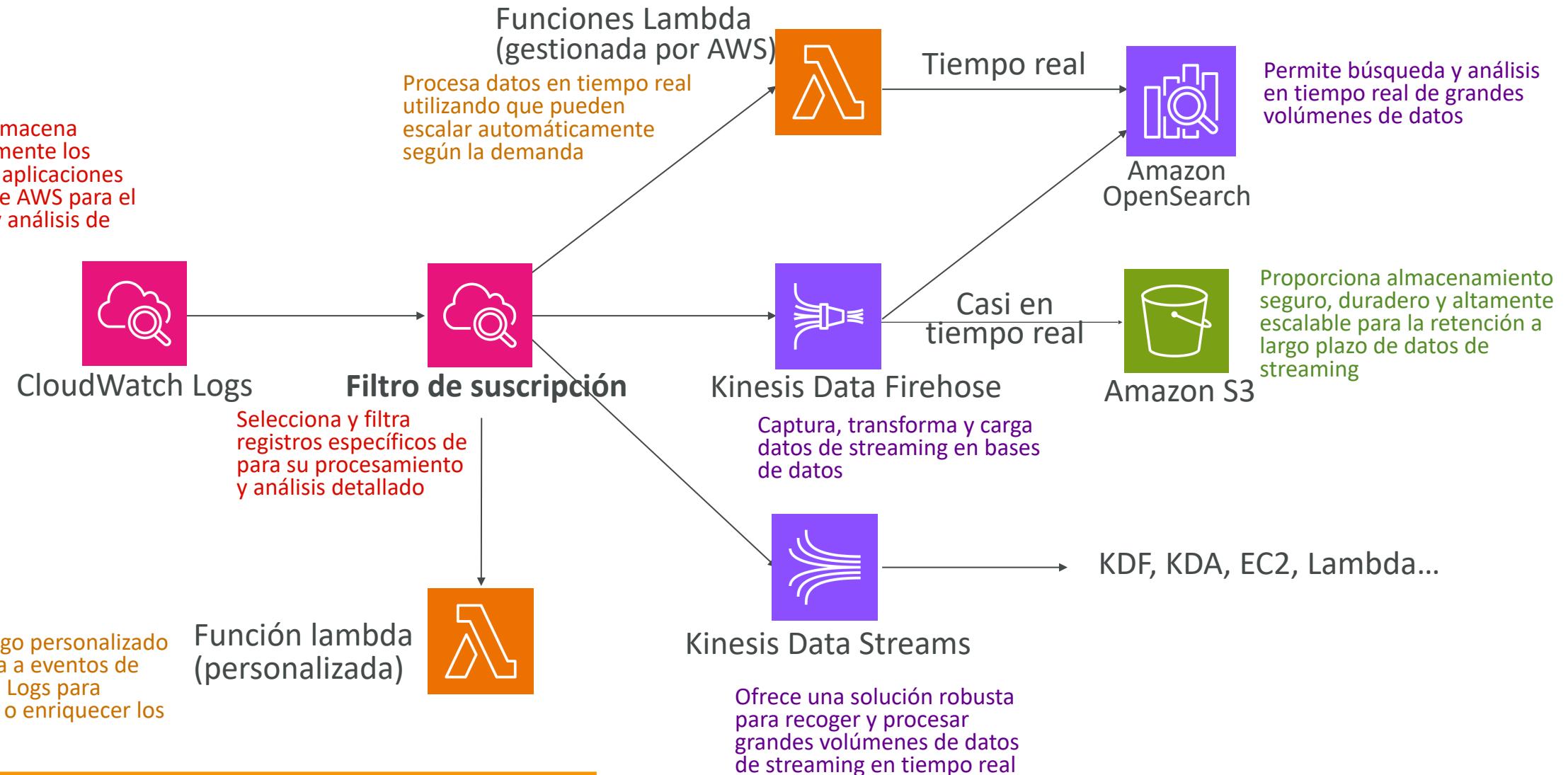
CloudWatch Logs - Exportación a S3

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

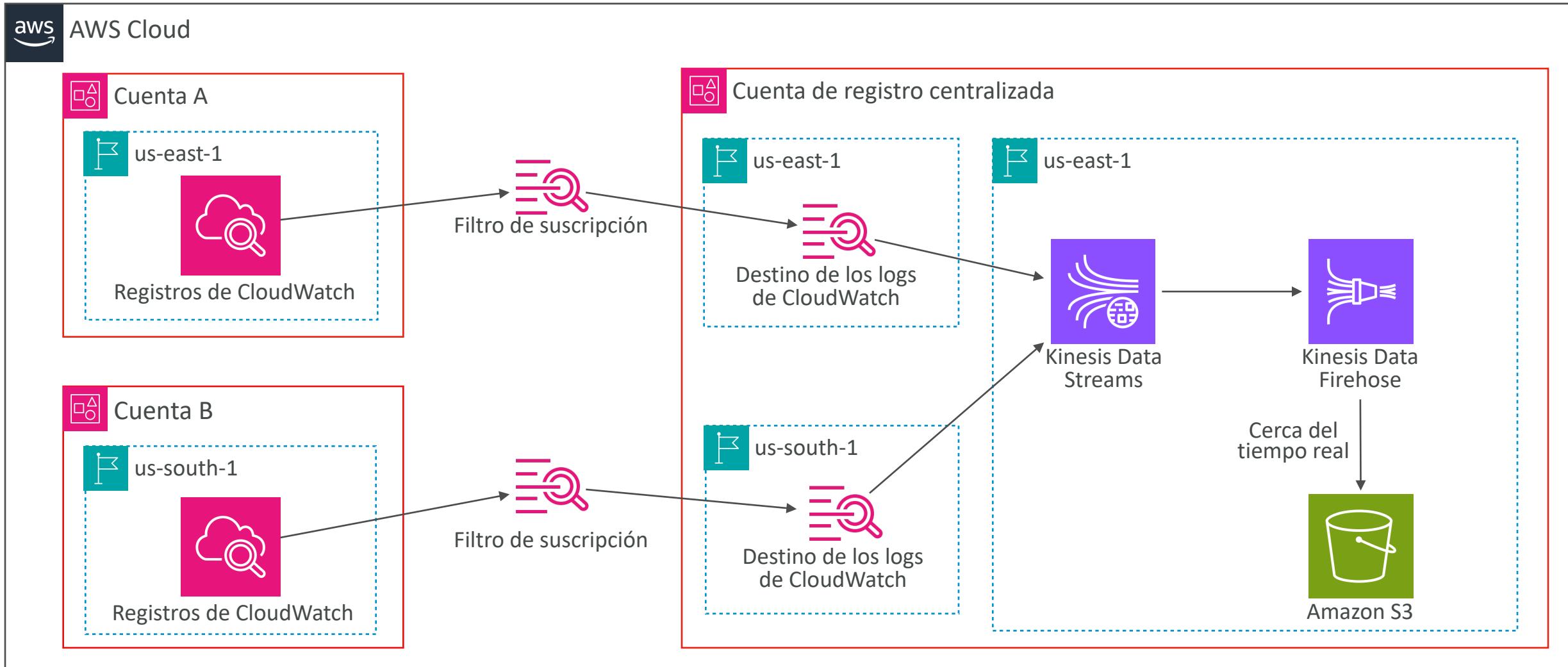


Suscripciones a CloudWatch Logs

Registra y almacena automáticamente los registros de aplicaciones y recursos de AWS para el monitoreo y análisis de problemas

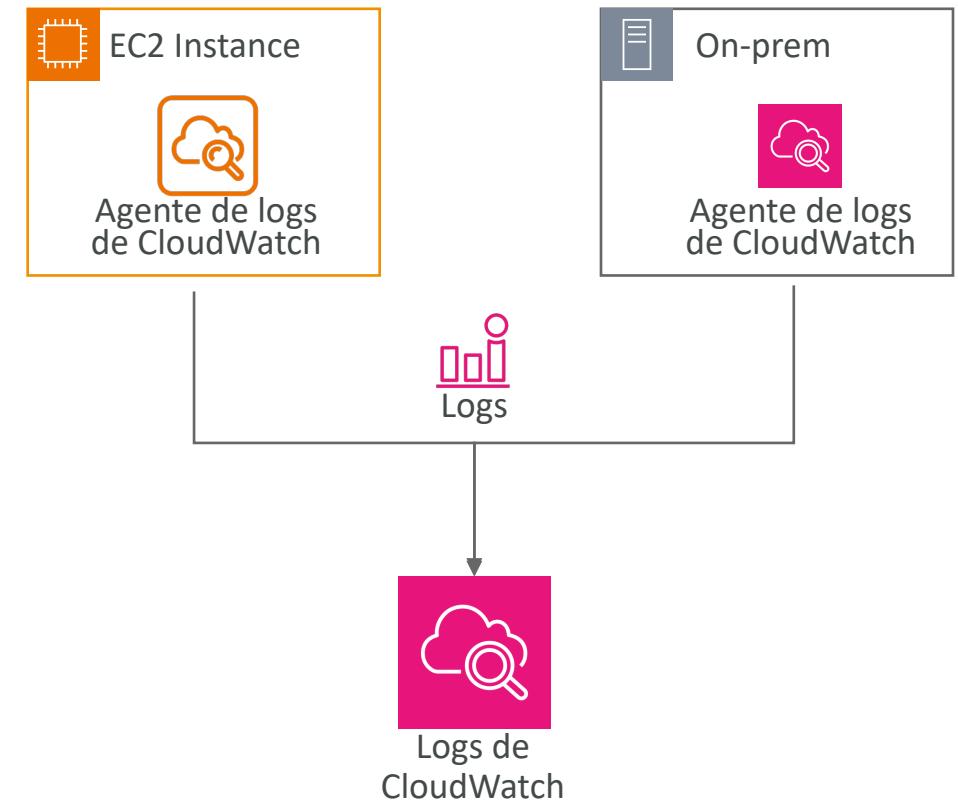


Agregación de CloudWatch Logs Multi-Cuenta y Multi-Región



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



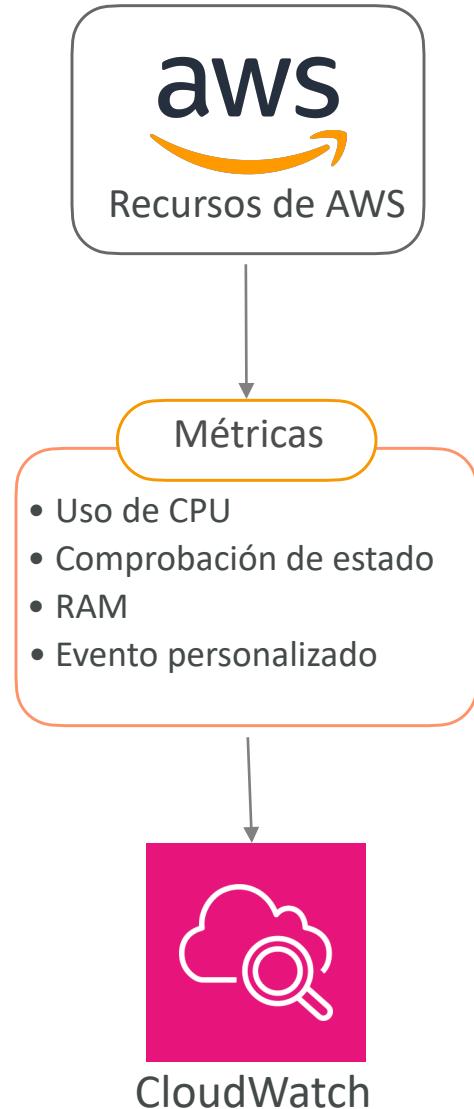
Agente CloudWatch Logs y Agente Unificado

- 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



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)



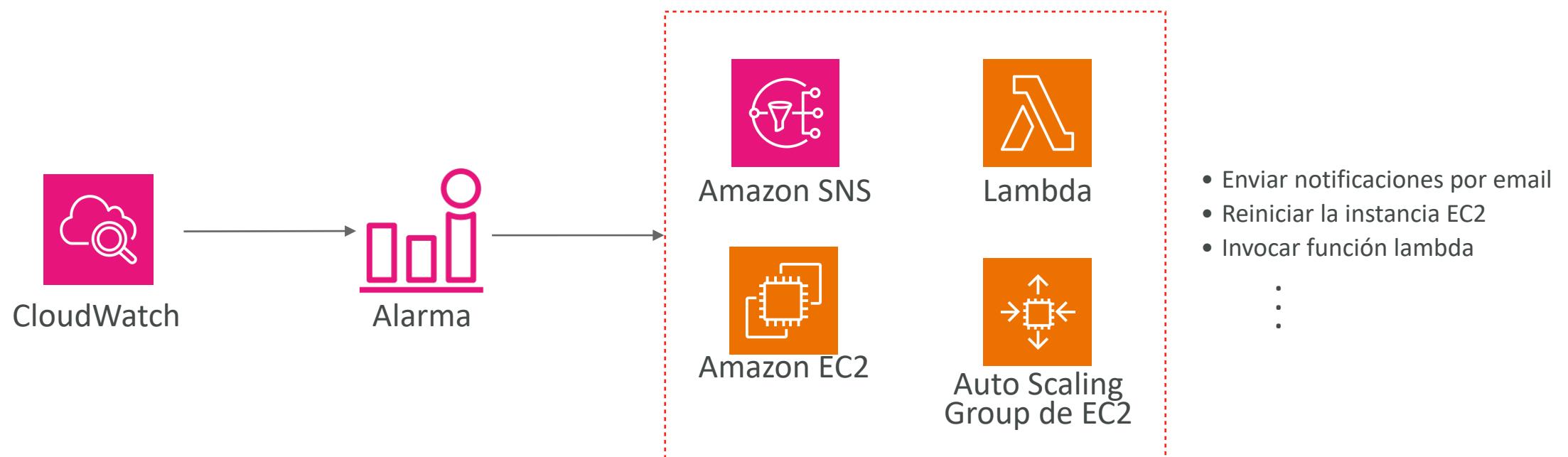
Amazon CloudWatch Alarms



- Configura alertas para **detectar anomalías en el comportamiento de tus recursos AWS**, permitiendo una respuesta rápida
- Automatiza respuestas para resolver problemas o escalar recursos en función del estado de las alarmas
- Establece **umbrales** específicos para las métricas monitorizadas, adecuados a las necesidades y patrones de uso de tus aplicaciones (ej: 70%)
- Ejemplo: crear **una alarma de facturación** en la métrica de facturación de CloudWatch
- Estados de la alarma:
 - OK, INSUFFICIENT_DATA, ALARM

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 EventBridge



- **EventBridge facilita la conexión de aplicaciones con datos provenientes de diversas fuentes de AWS y aplicaciones de software como servicio (SaaS)** externas, permitiendo una integración sin problemas y la automatización de flujos de trabajo
- Los usuarios pueden definir y enviar sus propios eventos a través de buses de eventos de EventBridge, lo que permite crear aplicaciones altamente personalizadas y reactivas basadas en eventos específicos de negocio.
 - **Bus de eventos predeterminado** - generado por servicios de AWS (CloudWatch Events)
 - **Bus de eventos de socios** - recibe eventos de servicios SaaS o aplicaciones (Zendesk, DataDog, Segment, Auth0...)
 - **Buses de eventos personalizados** - para tus propias aplicaciones

Amazon EventBridge - Registro de esquemas

- EventBridge puede analizar los eventos de tu bus e inferir el **esquema**
 - *Los esquemas son estructuras de datos que definen la forma y el formato de los eventos
- 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

aws.codepipeline@CodePipelineActionExecutionStateChange

Schema details

| Schema name | Last modified | Schema ARN |
|---------------------------------------------------------|---------------------------|-------------|
| aws.codepipeline@CodePipelineActionExecutionStateChange | Dec 1, 2019, 12:11 AM GMT | - |
| Schema registry | Number of versions | Schema type |
| aws.events | 1 | OpenAPI 3.0 |

Description

Schema for event type CodePipelineActionExecutionStateChange, published by AWS service aws.codepipeline

Version 1 Created on Dec 1, 2019, 12:11 AM GMT

Action ▾ Download code bindings

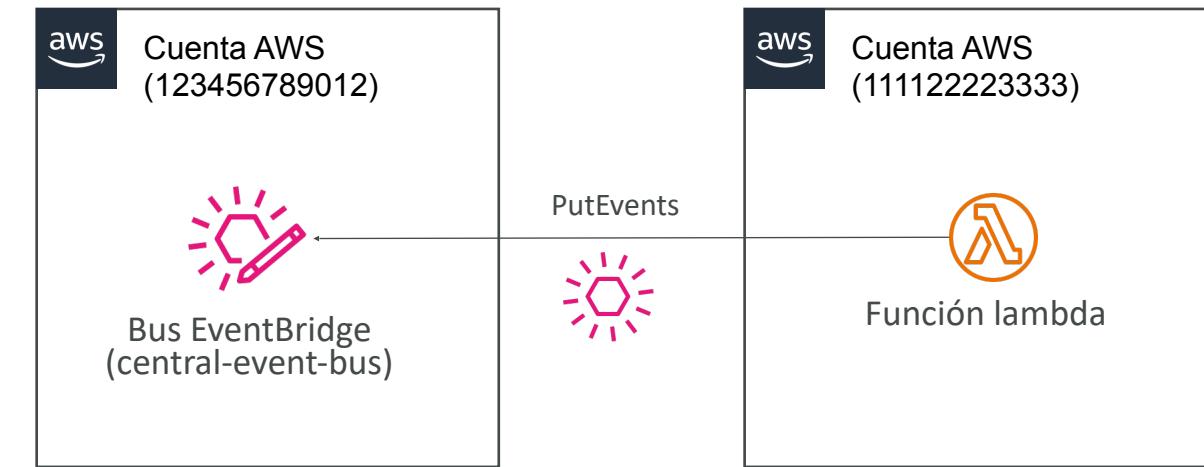
```
1 {
2   "openapi": "3.0.0",
3   "info": {
4     "version": "1.0.0",
5     "title": "CodePipelineActionExecutionStateChange"
6   },
7   "paths": {},
8   "components": {
9     "schemas": {
10       "AWSEvent": {
```

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 de AWS en una única cuenta AWS o región de AWS

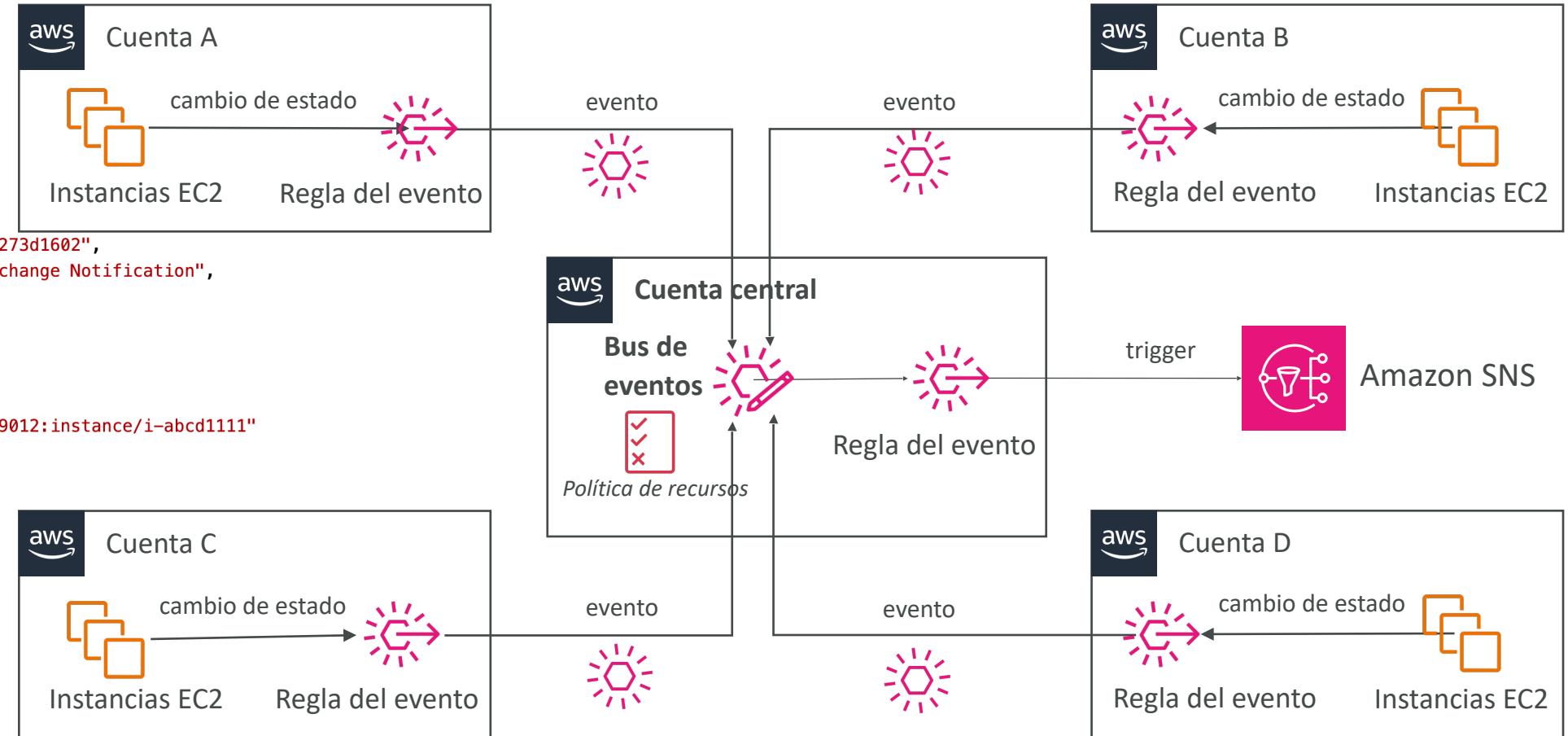
```
{  
    "Version": "2012-10-17",  
    "Statement": [  
        {  
            "Effect": "Allow"  
            "Action": "events:PutEvents"  
            "Principal": {"AWS": "1112223333"}  
            "Resource": "arn:aws:events:us-  
east-1_123456789012:event-bus/central-event-bus"  
        }  
    ]  
}
```

Permitir **eventos** desde otra cuenta AWS



EventBridge - Agregación multicuenta

```
        "id": "7bf73129-1428-4cd3-a780-95db273d1602",
        "detail-type": "EC2 Instance State-change Notification",
        "source": "aws.ec2",
        "account": "123456789012",
        "time": "2021-11-11T21:29:54Z",
        "region": "us-east-1",
        "resources": [
            "arn:aws:ec2:us-east-1:123456789012:instance/i-abcd1111"
        ],
        "detail": {
            "instance-id": "i-abcd1111",
            "state": "pending"
        }
    }
```



Caso práctico real de Amazon EventBridge

- **Caso Práctico: Monitoreo de salud en tiempo real**

- Una empresa de tecnología utiliza AWS para hospedar su infraestructura de aplicaciones. Quieren monitorear la salud de sus servicios en tiempo real para responder rápidamente a cualquier incidente

- **Implementación con EventBridge:**

- **Detección de incidentes**

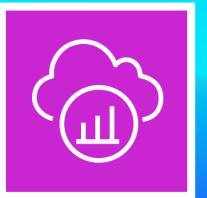
- Los servicios como Amazon EC2 y Amazon RDS envían automáticamente alertas de estado a EventBridge cuando detectan problemas como sobrecarga de CPU o fallos de conexión

- **Automatización de respuestas**

- EventBridge recibe estas alertas y dispara funciones AWS Lambda preconfiguradas para realizar acciones inmediatas, como escalar recursos o reiniciar instancias

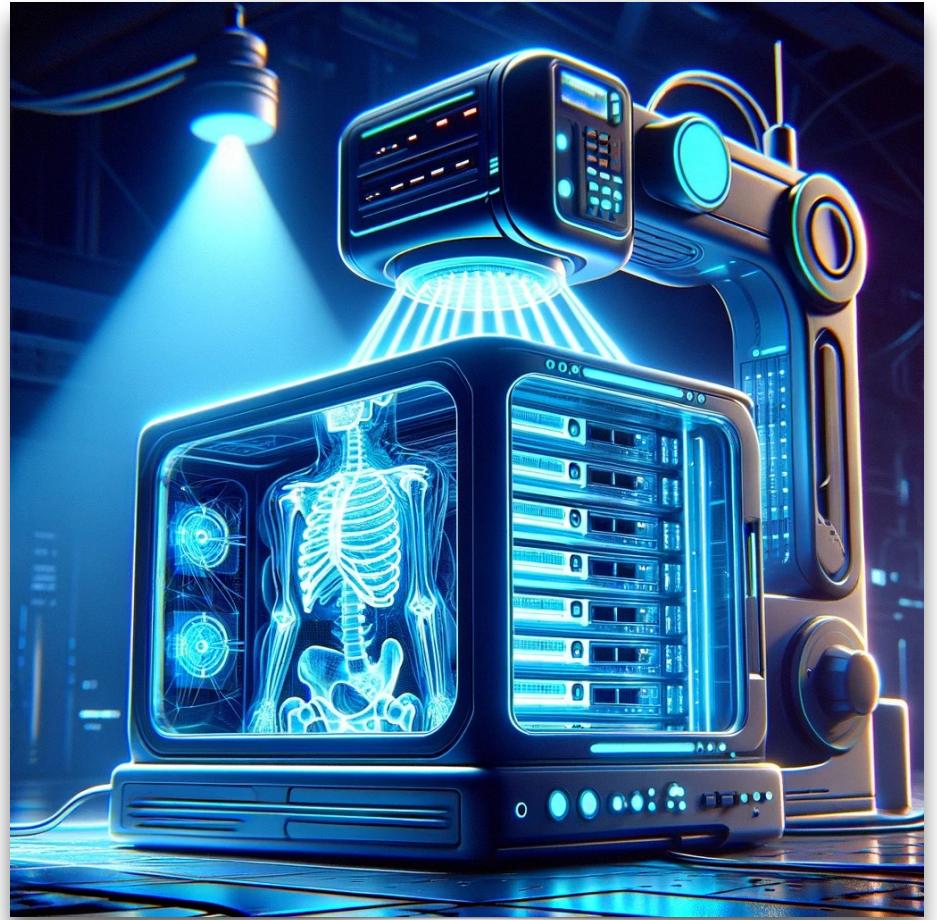
- **Notificación a equipos**

- Además, EventBridge puede enviar notificaciones a un canal de Slack o Microsoft Teams para informar al equipo de TI sobre el incidente y las acciones tomadas



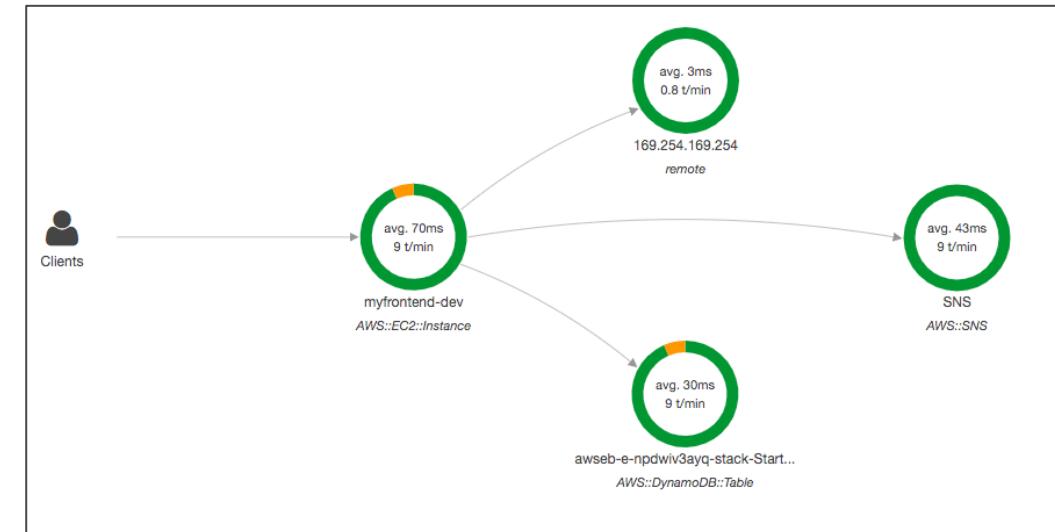
AWS X-Ray

- Depurar en producción, a la antigua:
 - Test localmente
 - Añade logs por todas partes
 - Vuelve a desplegar en producción
- Los formatos de logs difieren entre las aplicaciones que utilizan CloudWatch y el análisis es difícil
- Depuración: monolito "fácil", servicios distribuidos "difícil"
- No hay vistas comunes de toda tu arquitectura
- Entra... ¡AWS X-Ray!



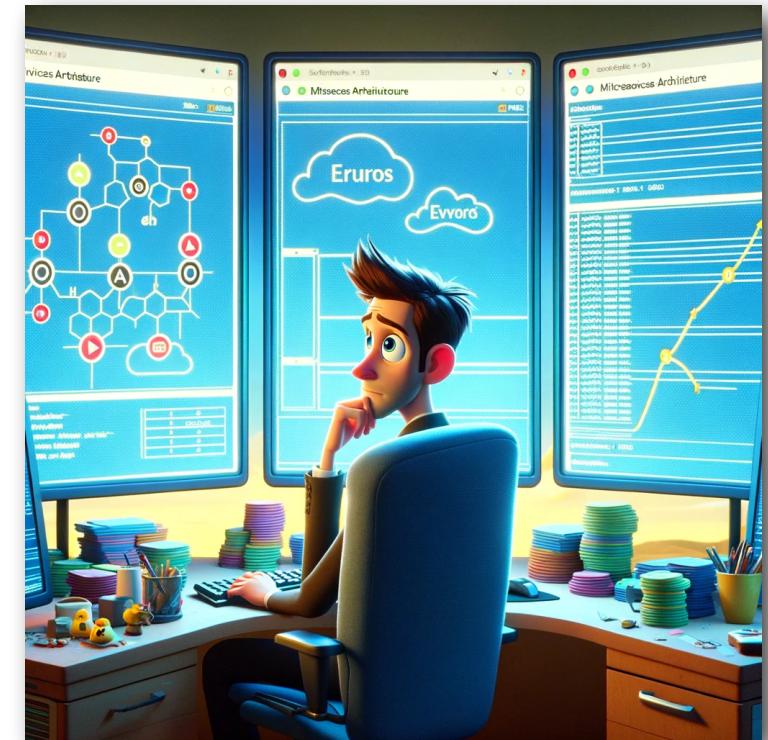
La magia de X-Ray

- El servicio X-Ray recoge datos de todos los diferentes servicios
- El mapa de servicios se calcula a partir de todos los segmentos y trazas
- X-Ray es gráfico, por lo que incluso personas no técnicas pueden ayudar a solucionar problemas

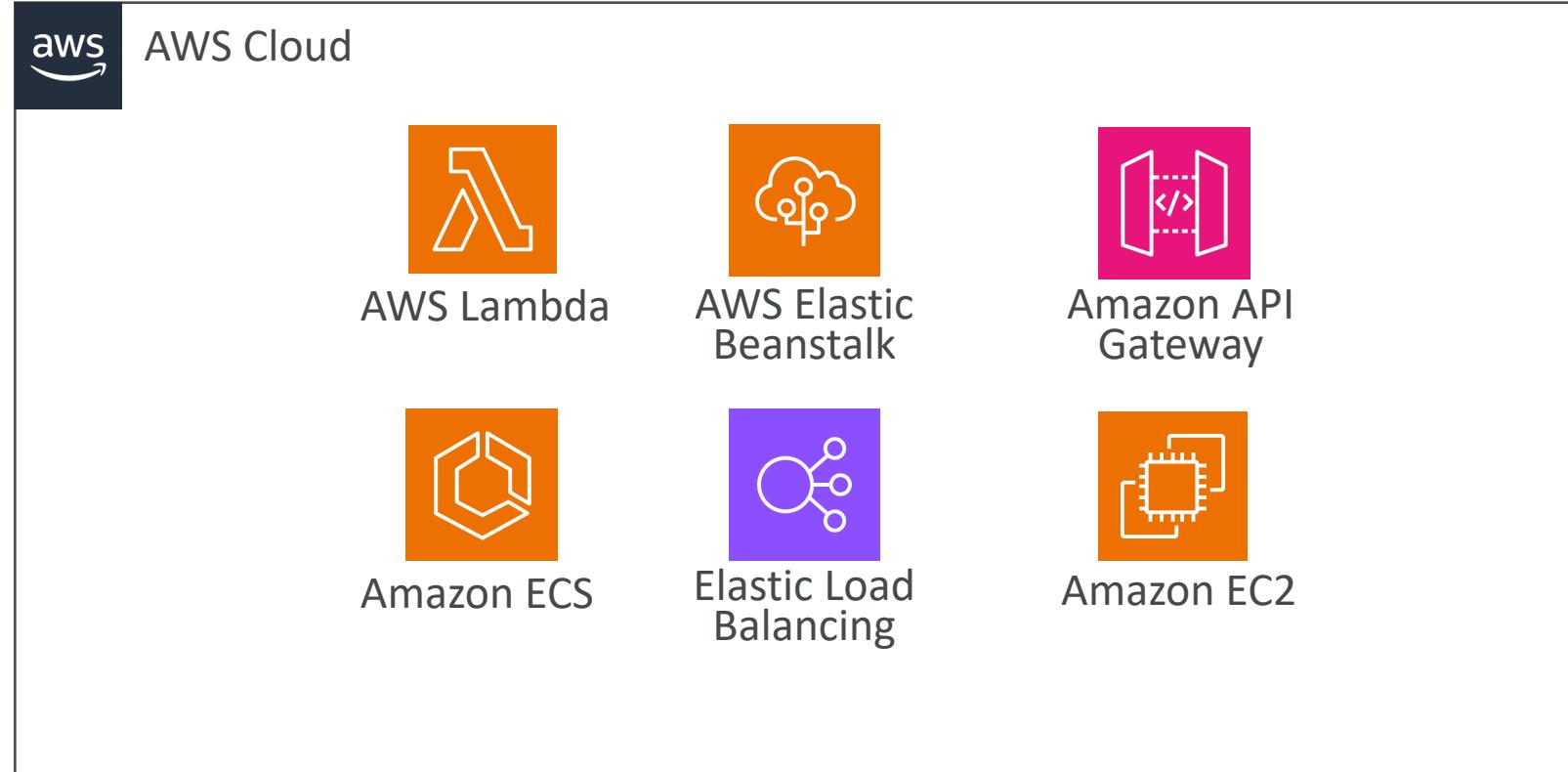


Ventajas de AWS X-Ray

- Resolución de problemas de rendimiento (cuellos de botella)
- Comprender las dependencias en una arquitectura de microservicios
- Localizar problemas de servicio
- Revisar el comportamiento de las peticiones
- Encontrar errores y excepciones
- ¿Cumplimos el SLA de tiempo?
- ¿Dónde estoy estrangulado?
- Identifica a los usuarios afectados



Compatibilidad con X-Ray



AWS X-Ray aprovecha el rastreo

- El rastreo es una forma integral de seguir una "petición"
- Cada componente que se ocupa de la petición añade su propio "seguimiento"
- El seguimiento se compone de segmentos (+ subsegmentos)
- Se pueden añadir anotaciones a las trazas para proporcionar información adicional
- Posibilidad de rastrear:
 - Cada petición
 - Muestra de petición (en %, por ejemplo, o en tasa por minuto)
- Seguridad X-Ray:
 - IAM para autorización
 - KMS para cifrado en reposo

AWS X-Ray

¿Cómo activarlo?

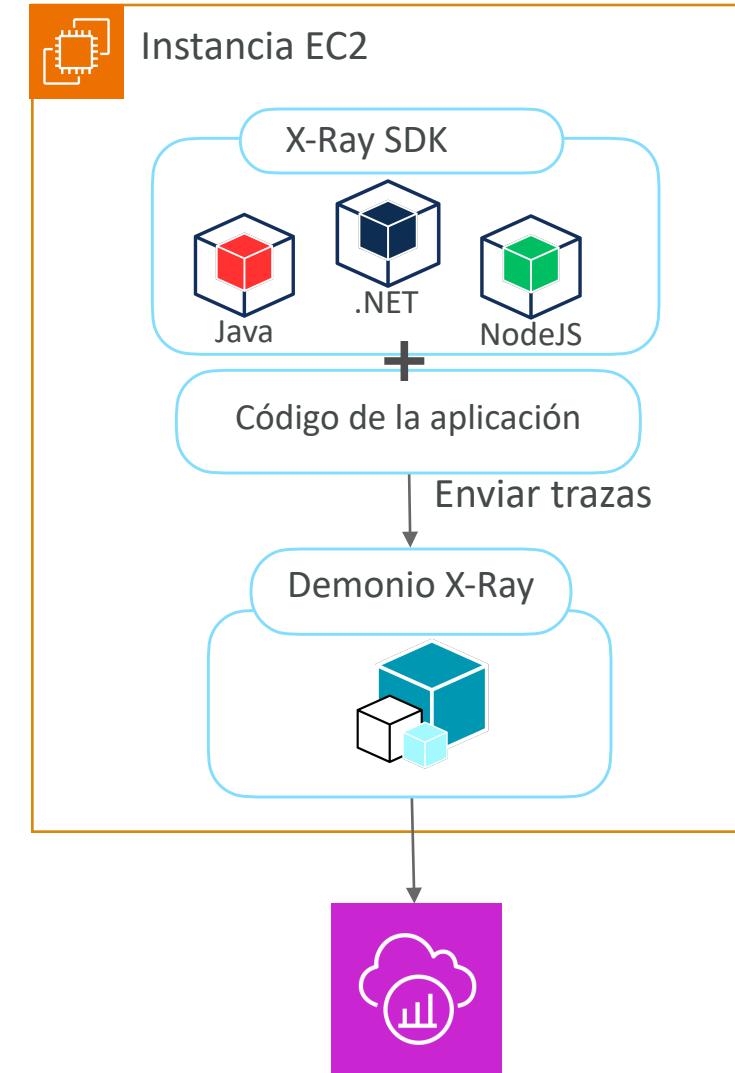
1) Tu código (Java, Python, Go, Node.js, .NET) debe importar el SDK de AWS X-Ray

- Se necesita muy poca modificación del código
- A continuación, el SDK de la aplicación capturará
 - Llamadas a servicios de AWS
 - Peticiones HTTP / HTTPS
 - Llamadas a bases de datos (MySQL, PostgreSQL, DynamoDB)
 - Llamadas a colas (SQS)

2) Instala el demonio (daemon) X-Ray o activa la integración de X-Ray en AWS

El demonio X-Ray funciona como un interceptor de paquetes UDP de bajo nivel (Linux / Windows / Mac...)

AWS Lambda / otros servicios de AWS ya ejecutan el demonio X-Ray por ti
Cada aplicación debe tener los derechos IAM para escribir datos en X-Ray

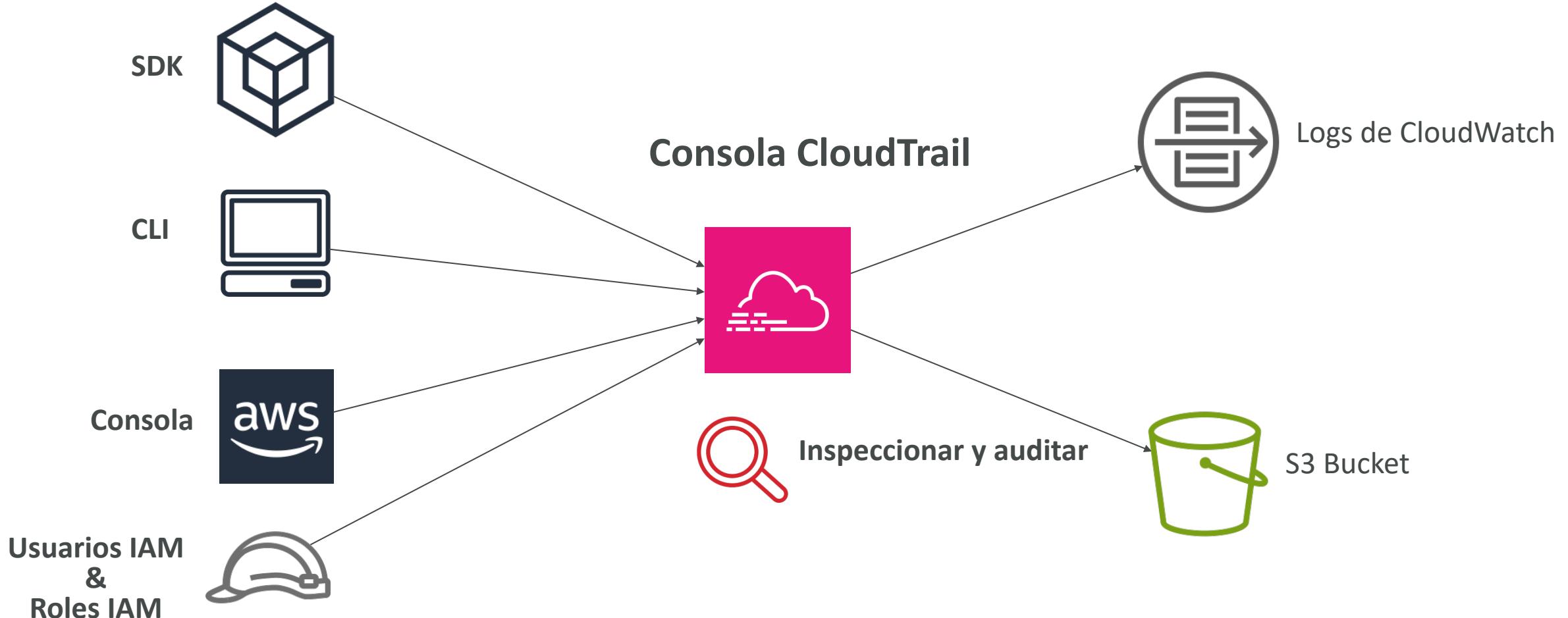


AWS CloudTrail



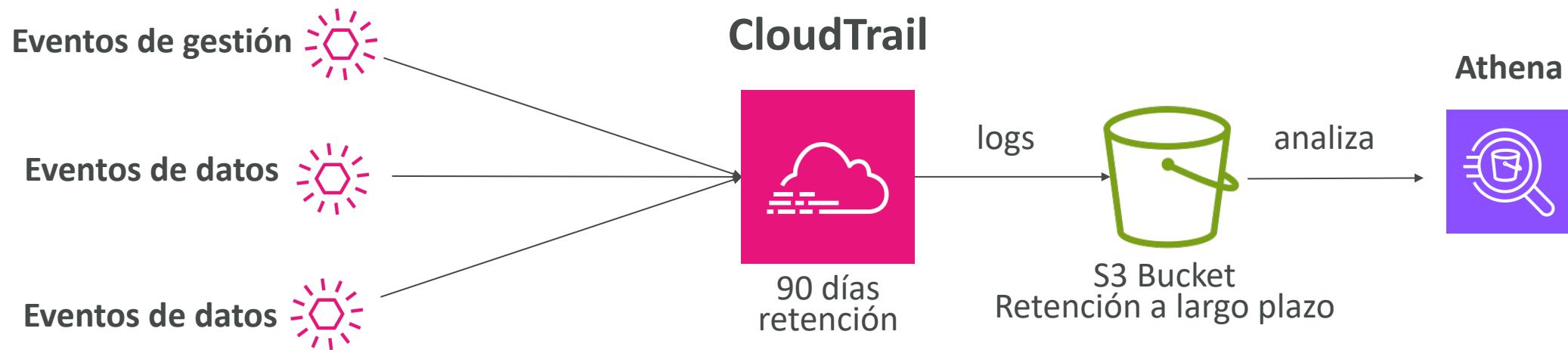
- **Proporciona gobernanza, normativa y auditoría para tu cuenta de AWS**
- CloudTrail está activado por defecto
- Obtén **un historial de eventos / llamadas a la API realizadas dentro de tu Cuenta de AWS** por:
 - Consola
 - SDK
 - CLI
 - Servicios de AWS
- Puedes poner logs de CloudTrail en CloudWatch 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



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

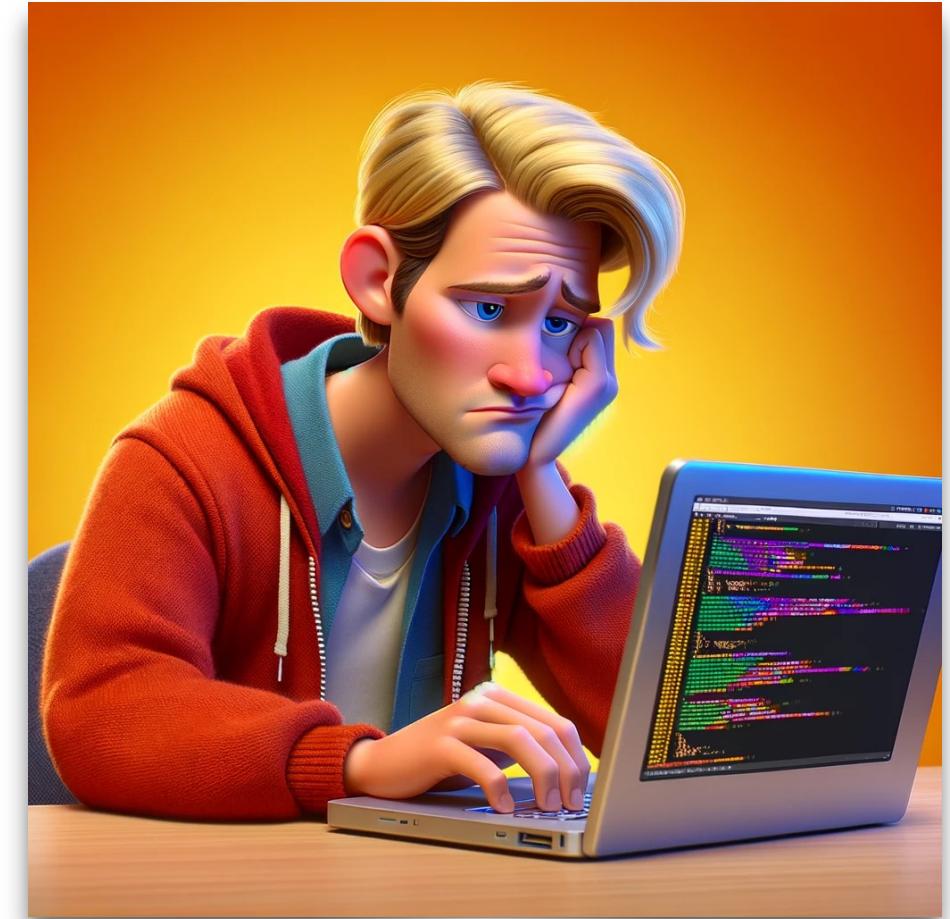


CloudTrail vs CloudWatch vs X-Ray

- CloudTrail:
 - Auditoría de llamadas a la API realizadas por usuarios / servicios / consola de AWS
 - Útil para detectar llamadas no autorizadas o la causa raíz de los cambios
- CloudWatch:
 - CloudWatch Metrics a lo largo del tiempo para monitorización
 - CloudWatch Logs para almacenar logs de aplicaciones
 - CloudWatch Alarms para enviar notificaciones en caso de métricas inesperadas
- X-Ray:
 - Análisis de trazas automatizado y visualización de mapas de servicios
 - Análisis de latencia, errores y fallos
 - Seguimiento de peticiones en sistemas distribuidos

Infraestructura como código (IaC)

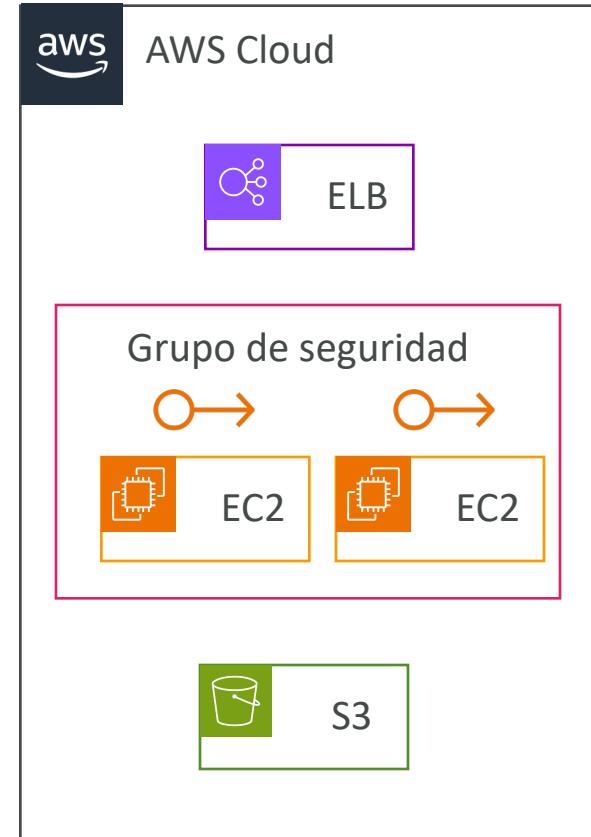
- Actualmente, hemos estado haciendo mucho trabajo manual...
- Todo este trabajo manual será muy difícil de reproducir:
 - En otra región
 - En otra cuenta de AWS
 - Dentro de la misma región si se borrara todo
- **¿No sería genial que toda nuestra infraestructura fuera... código? 🤔**
- Ese código se desplegaría y crearía / actualizaría / eliminaría nuestra infraestructura



¿Qué es CloudFormation?

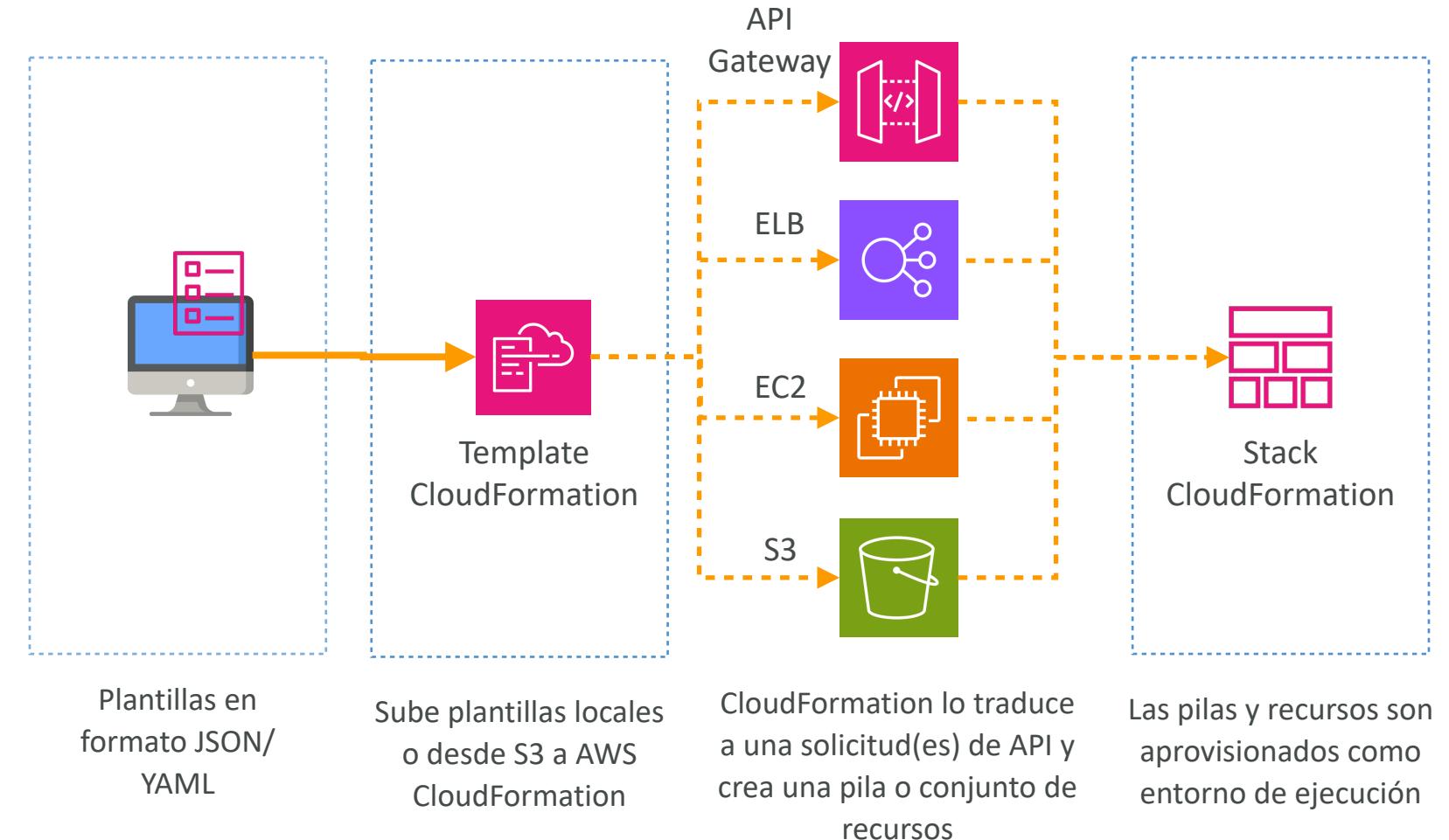


- CloudFormation es una forma declarativa de esbozar tu infraestructura de AWS, para cualquier recurso (la mayoría de ellos están soportados)
- Por ejemplo, dentro de una plantilla de CloudFormation, dices:
 - Quiero un grupo de seguridad
 - Quiero dos máquinas EC2 que utilicen este grupo de seguridad
 - Quiero dos IPs elásticas para estas máquinas EC2
 - 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**



Visión general de CloudFormation

- El uso de CloudFormation es **totalmente gratis**. Sólo pagas los recursos aprovisionados
- Se encuentra disponible en todas las regiones de AWS



Ventajas de AWS CloudFormation (I/2)

- **Infraestructura como código (IaC)**

- No se crean recursos manualmente, lo que es excelente para el control
- El código puede controlarse mediante versiones, por ejemplo, utilizando git
- Los cambios en la infraestructura se revisan a través del código



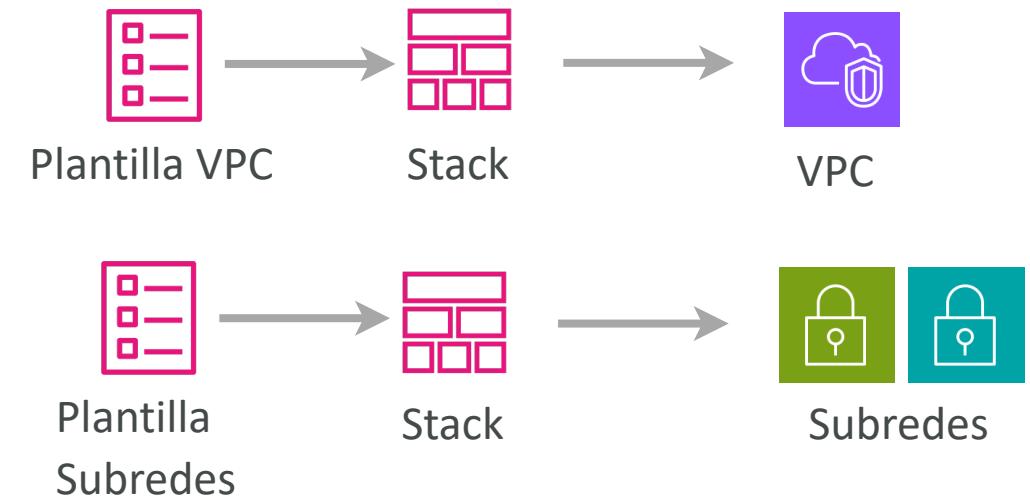
- **Coste**

- Cada recurso dentro de la pila se etiqueta con un identificador para que puedas ver fácilmente cuánto te cuesta un stack
- Puedes estimar los costes de tus recursos utilizando la plantilla CloudFormation
- El uso de CloudFormation no genera costo, solo los recursos creados por 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



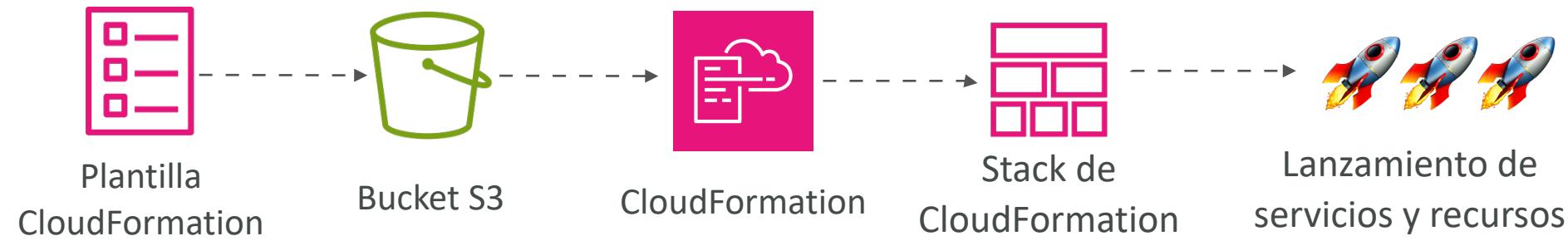
Ventajas de AWS CloudFormation (2/2)

- **Productividad**
 - Capacidad para destruir y volver a crear una infraestructura en el Cloud sobre la marcha
 - Generación automatizada de diagramas para tus plantillas
- **Separación de intereses:** crea muchos stacks para muchas aplicaciones, y muchas capas. Ej:
 - Stack 1:VPC
 - Stack 2: Subredes
 - Stack 3: Aplicaciones
- **No reinventar la rueda**
 - Aprovecha las plantillas existentes en la web
 - Aprovecha la documentación



¿Cómo funciona CloudFormation?

- Las plantillas pueden cargarse en un bucket S3 y luego referenciarse en CloudFormation (No es obligatorio usar S3)
- Para actualizar una plantilla, no podemos editar las anteriores. Tenemos que volver a subir una nueva versión de la plantilla a AWS
- Los stacks se identifican con un nombre
- Al borrar un stack se borran todos y cada uno de los artefactos creados por CloudFormation



Bloques de CloudFormation

Componentes de las plantillas:

- **AWSTemplateFormatVersion** - Versión del formato
- **Descripción** - Comentarios sobre la plantilla
- **Recursos (OBLIGATORIO)** - Recursos AWS declarados
- **Parámetros** - Las entradas dinámicas de tu plantilla
- **Mappings** - las variables estáticas de tu plantilla
- **Salidas** - Referencias a lo que se ha creado
- **Condicionales** - Lista de condiciones para realizar la creación de recursos
- **Metadatos** - Metadata opcional para incluir objetos JSON o YAML arbitrarios que proporcionan detalles

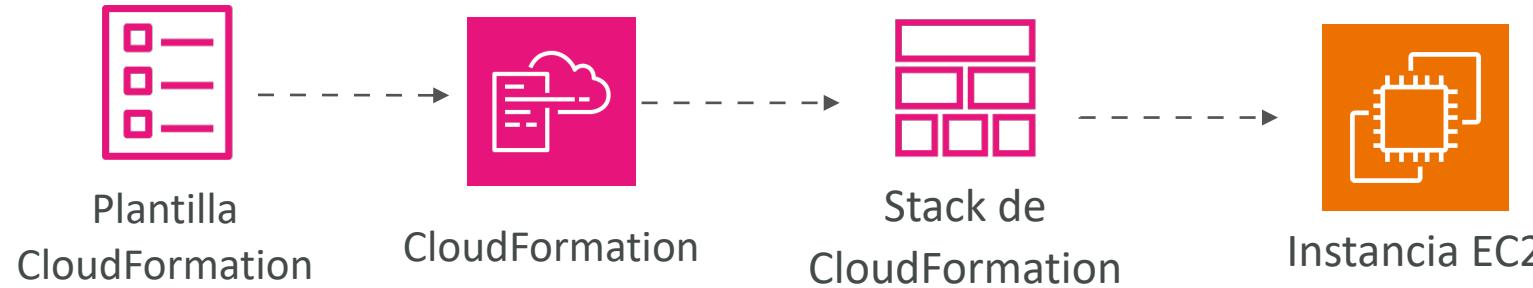
Plantillas de ayuda:

- **Referencias**
- **Funciones**

```
AWSTemplateFormatVersion:  
"fecha de la version"  
Description:  
  String  
Resources:  
  conjunto de recursos  
Parameters:  
  conjunto de parametros  
Rules:  
  conjunto de reglas  
Mappings:  
  conjunto de asignaciones  
Conditions:  
  asignación de condicionales  
Transform:  
  conjunto de transformaciones  
Outputs:  
  conjunto de salidas  
Metadata:  
  metadatos de la plantilla
```

Ejemplo introductorio de CloudFormation

- Vamos a crear una instancia EC2 sencilla con CloudFormation
- Luego vamos a crearla para añadirle una IP elástica
- Y vamos a añadirle dos grupos de seguridad
- Por ahora, olvídate de la sintaxis del código de la plantilla
- ¡Veremos cómo en un abrir y cerrar de ojos somos capaces de empezar a trabajar con CloudFormation!



S3 (Simple Storage Service)

Visión general de Amazon S3



- Amazon S3 es una **plataforma global de almacenamiento - regional**
- Servicio **público**, con capacidad de almacenar infinita información
 - Películas, Audio, Fotos, Texto, Datasets muy grandes, ...
- S3 se define como uno de los principales bloques de construcción de AWS
- Muchos sitios web utilizan Amazon S3 como columna vertebral
- Muchos servicios de AWS utilizan Amazon S3 como una integración también
- Componentes a destacar de S3: **Buckets** y **Objetos**

Objetos de Amazon S3

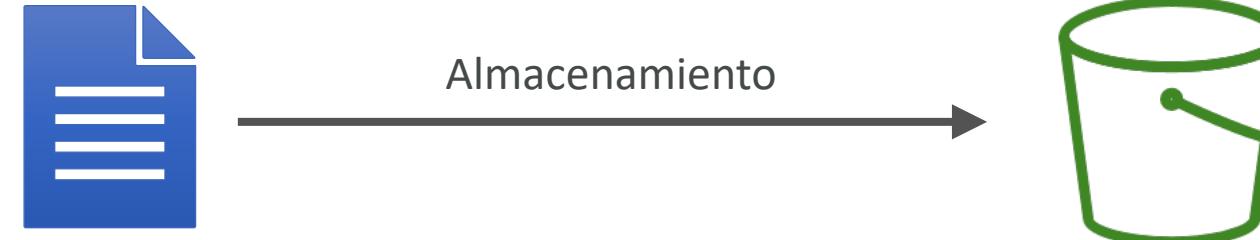


Visión general de Amazon S3 - Buckets

- 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)**
- S3 parece un servicio global pero los buckets se crean en una región



Amazon S3 - Objetos (1/2)



Objeto: mi-archivo.txt

Bucket: mi-bucket

- Los objetos (archivos) tienen una clave (key)
- La **clave** es la ruta **COMPLETA**:
 - s3://mi-bucket/**mi-archivo.txt**
 - s3://mi-bucket/**mi_carpeta1/otra_carpeta/mi_archivo.txt**

Amazon S3 - Objetos (2/2)

- **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 ("/")
- **El tamaño máximo del objeto es de 5TB (5000GB)**
- Si se suben más de 5GB, se debe usar "carga de varias partes"

The screenshot shows the Amazon S3 console interface for the bucket 'jamengual.com'. The top navigation bar is highlighted with a red box. Below it, the bucket name 'jamengual.com' is shown with a 'Información' link. A horizontal menu bar includes 'Objetos' (which is underlined), 'Propiedades', 'Permisos', 'Métricas', and 'Administración'. The 'Objetos' tab is active, displaying a list of two items:

| <input type="checkbox"/> | Nombre | Tipo |
|--------------------------|------------|---------|
| <input type="checkbox"/> | index.html | html |
| <input type="checkbox"/> | src/ | Carpeta |

Below the list are several action buttons: 'Copiar URI de S3', 'Copiar URL', and 'Descargar'. There is also a search bar labeled 'Buscar objetos por prefijo'.

Ejemplo - Claves de objetos

Información general sobre el objeto

| | |
|------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Propietario | URI DE S3 |
| 70246fdb32b07d4654b0976600e3b958b714f461a57cacd342fa277f03cb064e | <input type="button"/> s3://jamengual.com/src/imagen-demo.png |
| Región de AWS | Nombre de recurso de Amazon (ARN) |
| Europa (París) eu-west-3 | <input type="button"/> arn:aws:s3:::jamengual.com/src/imagen-demo.png |
| Última modificación | Etiqueta de entidad (Etag) |
| 17 Dec 2023 2:28:07 PM CET | <input type="button"/> ef40c0e31aeba7542fa82983921f0eaf |
| Tamaño | URL del objeto |
| 782.6 KB | <input type="button"/> https://s3.eu-west-3.amazonaws.com/jamengual.com/src/imagen-demo.png |
| Tipo | |
| png | |
| Clave | |
| <input type="button"/> src/imagen-demo.png | |

Casos de uso de Amazon S3

- Copia de seguridad y almacenamiento
- Recuperación de desastres
- Gestión de datos para aplicaciones IoT
- Almacenamiento en el Cloud híbrido
- Alojamiento de aplicaciones
- Alojamiento de medios
- Uso para análisis big data
- Entrega de software
- Sitio web estático



J.P.Morgan

Políticas de bucket S3

- **Resource:** el bucket, el objeto, el punto de acceso o el trabajo de Amazon S3 al que se aplica la política
- **Effect:** el efecto obtenido cuando el usuario solicita la acción específica, que puede ser permitir o denegar
- **Actions:** conjunto de operaciones para cada recurso
- **Entidad principal:** la cuenta o el usuario con permiso de acceso a las acciones
- **Condition:** condiciones para cuando surte efecto una política



```
{  
  "Version": "2012-10-17",  
  "Id": "ExamplePolicy01",  
  "Statement": [  
    {  
      "Sid": "ExampleStatement01",  
      "Effect": "Allow",  
      "Principal": {  
        "AWS": "arn:aws:iam::123456789012:user/Dave"  
      },  
      "Action": [  
        "s3:GetObject",  
        "s3:GetBucketLocation",  
        "s3>ListBucket"  
      ],  
      "Resource": [  
        "arn:aws:s3:::awsexamplebucket1/*",  
        "arn:aws:s3:::awsexamplebucket1"  
      ]  
    }  
  ]  
}
```

Bloquear el acceso público al almacenamiento de Amazon S3

- De forma predeterminada, los buckets, puntos de acceso y objetos nuevos **no permiten el acceso público**
- Estos ajustes se crearon para evitar la filtración de datos de la empresa
- Sin embargo, los usuarios pueden modificar las políticas de bucket, las políticas de punto de acceso o los permisos de objeto para **permitir el acceso público**



Configuración de bloqueo de acceso público para este bucket

Se concede acceso público a los buckets y objetos a través de listas de control de acceso (ACL), políticas de bucket, políticas de puntos de acceso o todas las anteriores. A fin de garantizar que se bloquee el acceso público a todos sus buckets y objetos, active Bloquear todo el acceso público. Esta configuración se aplica exclusivamente a este bucket y a sus puntos de acceso. AWS recomienda activar Bloquear todo el acceso público, pero, antes de aplicar cualquiera de estos ajustes, asegúrese de que las aplicaciones funcionarán correctamente sin acceso público. Si necesita cierto nivel de acceso público a los buckets u objetos, puede personalizar la configuración individual a continuación para adaptarla a sus casos de uso de almacenamiento específicos. [Más información](#)

Bloquear todo el acceso público
Activar esta configuración equivale a activar las cuatro opciones que aparecen a continuación. Cada uno de los siguientes ajustes son independientes entre sí.

Bloquear el acceso público a buckets y objetos concedido a través de nuevas listas de control de acceso (ACL)
S3 bloqueará los permisos de acceso público aplicados a objetos o buckets agregados recientemente, y evitará la creación de nuevas ACL de acceso público para buckets y objetos existentes. Esta configuración no cambia los permisos existentes que permiten acceso público a los recursos de S3 mediante ACL.

Bloquear el acceso público a buckets y objetos concedido a través de cualquier lista de control de acceso (ACL)
S3 ignorará todas las ACL que conceden acceso público a buckets y objetos.

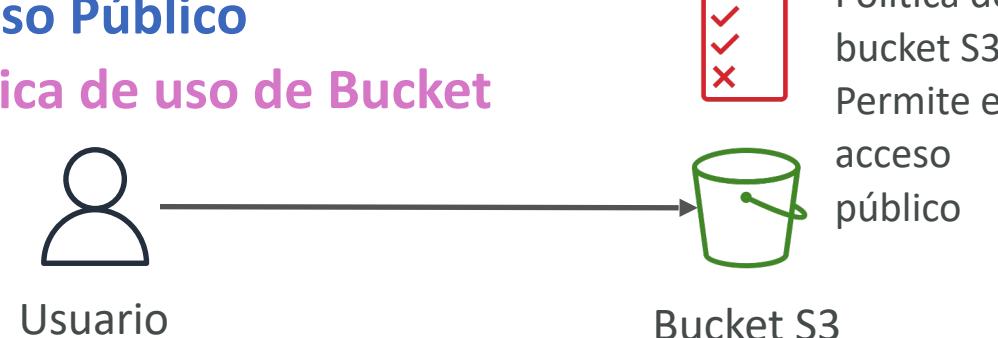
Bloquear el acceso público a buckets y objetos concedido a través de políticas de bucket y puntos de acceso públicas nuevas
S3 bloqueará las nuevas políticas de buckets y puntos de acceso que concedan acceso público a buckets y objetos. Esta configuración no afecta a las políticas ya existentes que permiten acceso público a los recursos de S3.

Bloquear el acceso público y entre cuentas a buckets y objetos concedido a través de cualquier política de bucket y puntos de acceso pública
S3 ignorará el acceso público y entre cuentas en el caso de buckets o puntos de acceso que tengan políticas que concedan acceso público a buckets y objetos.

Ejemplos de seguridad en Amazon S3

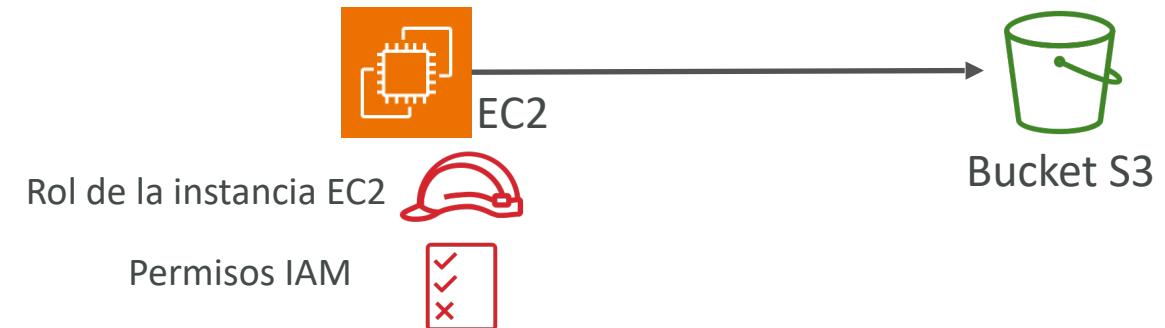
Acceso Público

Política de uso de Bucket



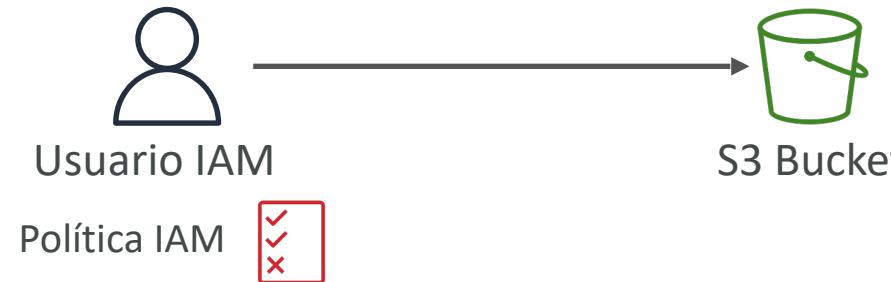
Acceso a instancias EC2

Usar roles IAM



Acceso del usuario a S3

Permisos IAM



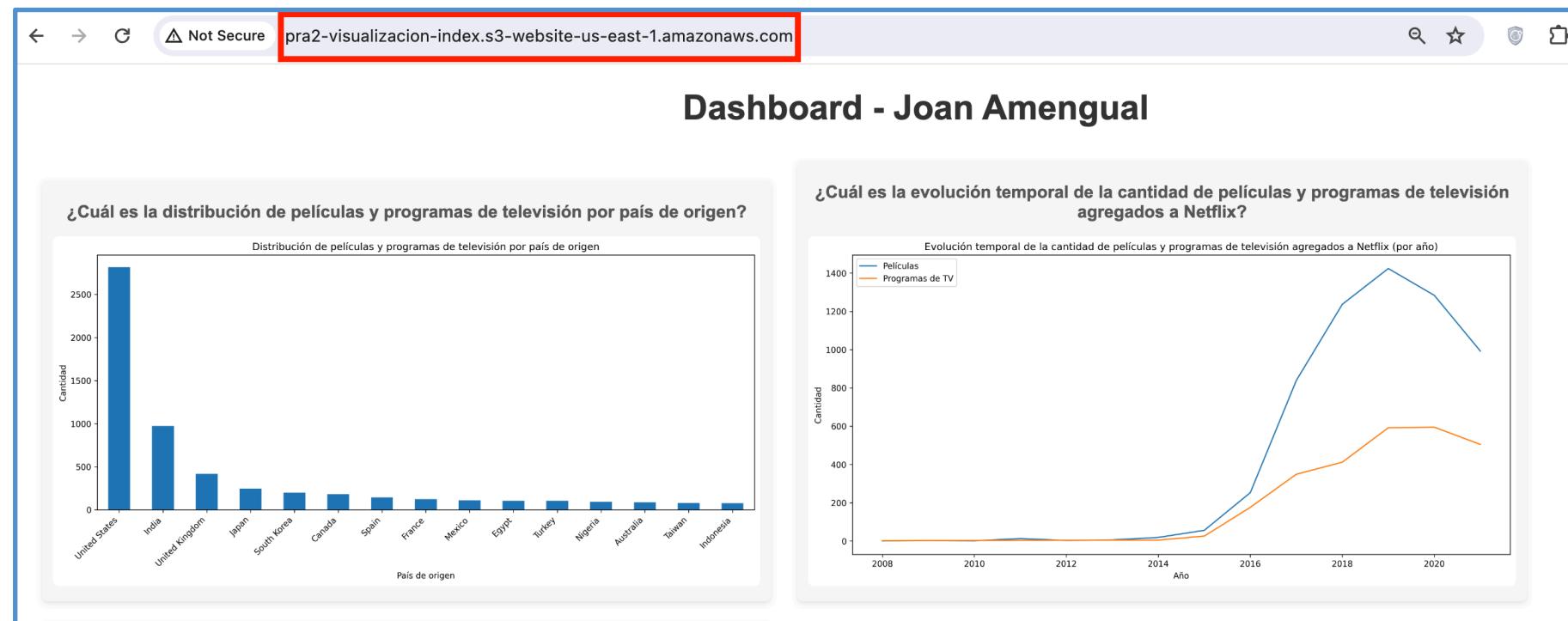
Acceso entre cuentas

Política de Bucket



Alojamiento de sitios web estáticos en S3

- **S3 puede alojar sitios web estáticos y tenerlos accesibles en Internet**
- La URL del sitio web cambiará según la región



¡No te olvides de la política de bucket!

Si obtienes un **error 403 Prohibido**, debes asegurarte de que la política del bucket permite las lecturas públicas

Política de bucket

EditarEliminar

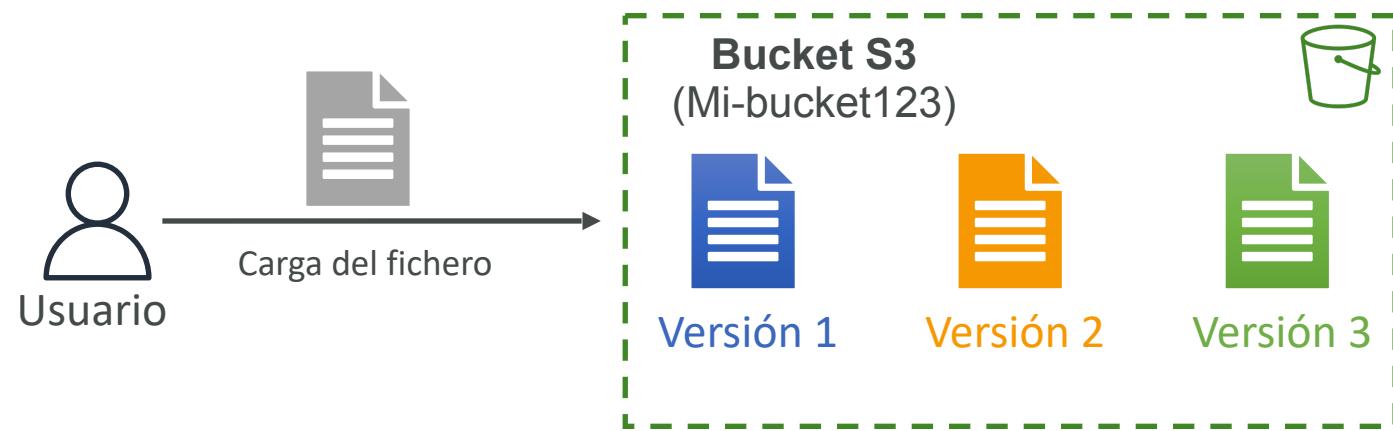
La política del bucket, escrita en JSON, proporciona acceso a los objetos almacenados en el bucket. Las políticas de bucket no se aplican a los objetos que pertenecen a otras cuentas. [Más información](#)

```
{  
    "Version": "2012-10-17",  
    "Statement": [  
        {  
            "Sid": "PublicRead",  
            "Effect": "Allow",  
            "Principal": "*",  
            "Action": "s3:GetObject",  
            "Resource": "arn:aws:s3:::pra2-visualizacion-index/*"  
        }  
    ]  
}
```

Copiar

Amazon S3 - Versionado

- Amazon S3 permite versionar nuestros archivos
- Esta característica se activa a nivel de bucket



- Cualquier archivo que no esté versionado antes de activar el versionado tendrá la versión "nula"
- Suspender el versionado no elimina las versiones anteriores

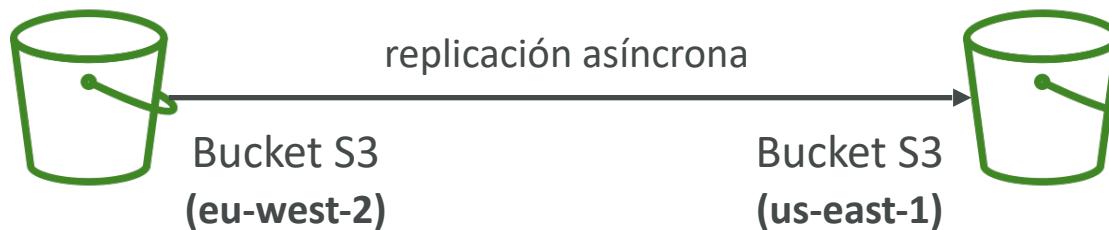
¿Por qué es una buena práctica versionar tus buckets?

- Protege contra los borrados involuntarios
- Facilidad para volver a la versión anterior



Replicación de información en buckets S3

- Podemos replicar la información en diferentes buckets S3
- ⚠ IMPORTANTE: Debes activar el control de versiones en los buckets de origen y destino
- Los buckets pueden estar en diferentes cuentas de AWS
- La copia es **asíncrona**
- Debes dar los **permisos IAM** adecuados a S3
- Después de activar la replicación, **sólo se replican los objetos nuevos**
- Existen diferentes tipos de replicación:
 - **Replicación entre regiones (CRR)**: normativa, acceso de menor latencia, replicación entre cuentas
 - **Replicación en la misma región (SRR)**: agregación de logs, replicación en vivo entre cuentas de producción y de prueba



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 de Amazon S3

- Amazon S3 ofrece una **variedad de clases de almacenamiento** entre las cuales puedes elegir en función de:
 - Requisitos de rendimiento
 - Acceso a los datos
 - Resiliencia
 - Costos de sus cargas de trabajo
- Las clases de almacenamiento de S3 se crearon específicamente para brindar el **menor costo posible de almacenamiento para los diferentes patrones de acceso**
- Se puede pasar de una clase a otra **manualmente** o utilizando las **configuraciones del ciclo de vida de S3**

S3 Standard

S3 Express One Zone

S3 Standard-IA

S3 One Zone-IA

S3 Glacier Instant Retrieval

S3 Glacier Flexible Retrieval

S3 Glacier Deep Archive

S3 Intelligent Tiering

Clases de almacenamiento (1/3)



S3 Standard



S3 Express One Zone



S3 Standard-IA



S3 One Zone-IA

- Datos activos, frecuentemente accedidos
- Acceso en milisegundos
- > 3 AZ

- Acceso ultrarrápido a datos
- Selección precisa de AZ
- 1 AZ
- Para cargas de trabajo intensivas en datos

- Datos accedidos infrecuentemente
- Acceso en milisegundos
- > 3 AZ
- Tarifa por recuperación por GB
- Duración mínima de almacenamiento
- Tamaño mínimo de objeto
- Mín. duración de almacenamiento: **30 días**

- Datos recreables, menos accedidos
- Acceso en milisegundos
- 1 AZ
- Tarifa por recuperación por GB
- Duración mínima de almacenamiento
- Tamaño mínimo de objeto
- Mín. duración de almacenamiento: **30 días**

*Precios de Norte de Virginia

0,023 USD por GB

0,16 USD por GB

0,0125 USD por GB

0,01 USD por GB

Clases de almacenamiento (2/3)



S3 Glacier Instant Retrieval



S3 Glacier Flexible Retrieval



S3 Glacier Deep Archive

- Almacenamiento de archivos de bajo costo
- Recuperación en milisegundos para datos raramente accesibles
- Mín. duración de almacenamiento: **90 días**

- Recuperación de datos asincrónica
- > 3 AZ
- Tarifa por recuperación por GB
- Duración mínima de almacenamiento
- Mín. duración de almacenamiento: **90 días**

- Datos archivados a largo plazo
- > 3 AZ
- Tarifa por recuperación por GB
- Duración mínima de almacenamiento
- Mín. duración de almacenamiento: **180 días**

0,004 USD por GB

0,0036 USD por GB

0,00099 USD por GB

Clases de almacenamiento (3/3)

Reduce costos moviendo automáticamente datos a niveles de precio más bajo según su uso.

- **Costos de migración:** Pequeño **cargo mensual por monitoreo**; no hay costos de recuperación.
- **Ahorro:** Desde 2018, ha ahorrado a los clientes mil millones de dólares frente a S3 Standard.
- **Uso:** Ideal para **patrones de acceso impredecibles**, aplicable a cualquier carga de trabajo.
- **Sin costos extras:** No hay gastos de operación, cargos de ciclo de vida o mínimos de almacenamiento.
- **Tamaño de objeto:** Objetos **menores de 128 KB** no son elegibles para la automatización y se tarifican al nivel frecuente.



S3 Intelligent-Tiering

0,0025 USD por 1000 objetos
+ coste por clase usada



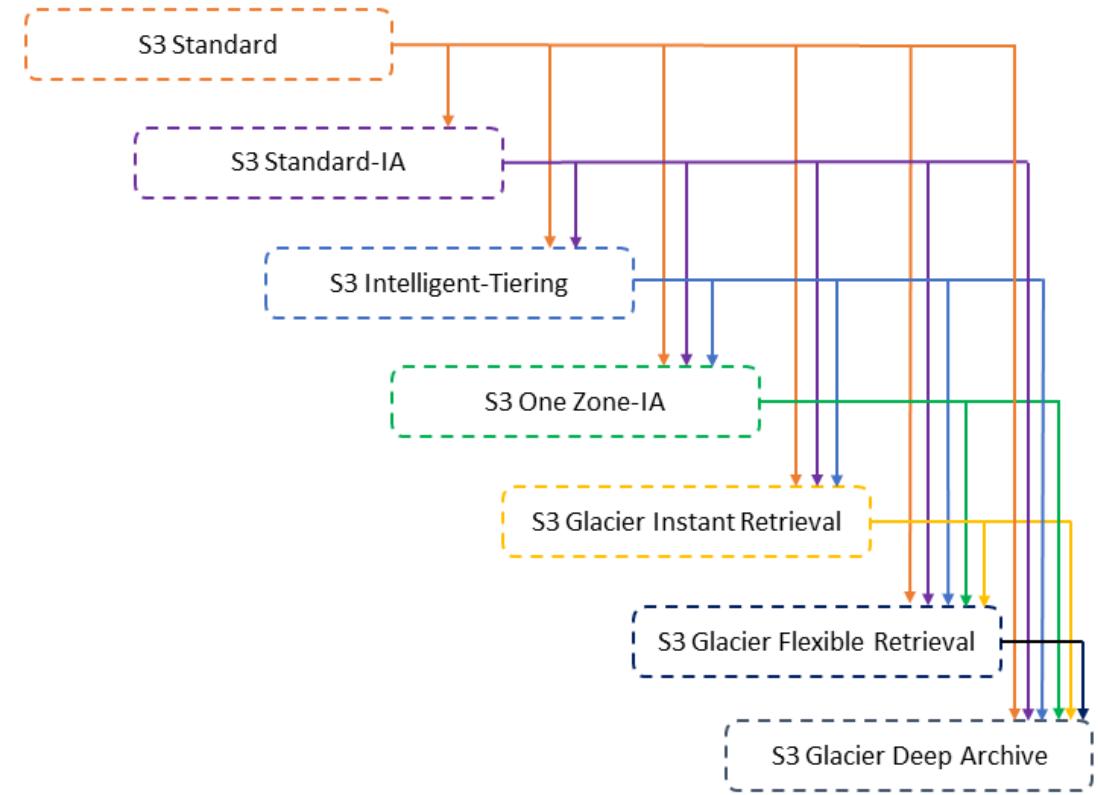
S3 Avanzado

www.blockstellart.com

Todos los derechos reservados © BLOCKSTELLART www.blockstellart.com

Movimiento de objetos entre diferentes clases de almacenamiento de Amazon S3

- En una configuración del ciclo de vida de S3, puedes definir **reglas para pasar objetos de una clase de almacenamiento a otra** para ahorrar costos de almacenamiento
- Para los objetos a los que se accede con poca frecuencia, muévelos a S3 Standard-IA
- Para los objetos a los que no necesitas acceder rápidamente, muévelos a S3 Glacier
- Amazon S3 admite un **modelo en cascada** para las transiciones entre clases de almacenamiento



https://docs.aws.amazon.com/es_es/AmazonS3/latest/userguide/lifecycle-transition-general-considerations.html

Reglas del ciclo de vida de Amazon S3

- Movimiento de los objetos entre diferentes clases de almacenamiento
 - Mover los objetos a la clase S3 Standard-IA 60 días después de su creación
 - Mover a S3 Glacier Flexible Retrieval para archivar después de 180 días
- Configura los objetos para que se eliminen después de X tiempo
 - **Se puede utilizar para eliminar versiones antiguas de archivos (si el versionado está activado)**
- Se pueden crear reglas...
 - para un determinado prefijo (ejemplo: `s3://mibucket/fotos/*`)
 - para determinados objetos con etiquetas (ejemplo: `Fase:Test`)

Acciones de la regla del ciclo de vida
Elija las acciones que desea que realice esta regla. Se aplican tarifas por solicitud. [Obtenga más información](#) o consulte [los precios de Amazon S3](#).

Transferir las versiones actuales de los objetos entre las clases de almacenamiento
 Transferir las versiones desactualizadas de los objetos entre las clases de almacenamiento
 Hacer que venzan las versiones actuales de los objetos
 Eliminar permanentemente las versiones no actuales de los objetos
 Eliminar los marcadores de eliminación de objetos vencidos o las cargas multipartes incompletas

Estas acciones no se admiten cuando se filtra por etiquetas de objetos o tamaño de objeto.

Realizar la transición de las versiones actuales de los objetos entre las clases de almacenamiento
Elija transiciones para transferir las versiones actuales de los objetos entre clases de almacenamiento en función del escenario de uso y los requisitos de acceso de rendimiento. Estas transiciones comienzan a partir del momento en que se crean los objetos y se aplican consecutivamente. [Más información](#)

| Elegir transiciones de clase de almacenamiento | Días después de la creación del objeto | Eliminar |
|------------------------------------------------|----------------------------------------|-----------------------------------------|
| Estándar - Acceso poco frecuente | 60 | <input type="button" value="Eliminar"/> |
| Recuperación instantánea de Glacier | 180 | <input type="button" value="Eliminar"/> |

Agregar transición

Ejercicio I - Reglas del ciclo de vida

- **Imagina que tienes una aplicación en AWS que procesa y almacena videos de capacitación en S3. Cada video se transcodifica en varios formatos para diferentes dispositivos y solo se necesitan durante 90 días, ya que después se actualizan con nuevas versiones. Los videos originales deben estar disponibles de inmediato durante esos 90 días, pero después pueden almacenarse de forma más económica si su acceso es menos frecuente. ¿Cuál sería tu enfoque de diseño?**
- Los videos originales podrían almacenarse en S3 con la clase de almacenamiento Estándar, y se usaría una política de ciclo de vida para moverlos a S3 Glacier o Glacier Deep Archive tras 90 días.
- Los formatos transcodificados, que son recreables, se almacenarían en una clase de almacenamiento S3 IA (Infrequent Access) con una política de ciclo de vida para eliminarlos automáticamente después de los 90 días.

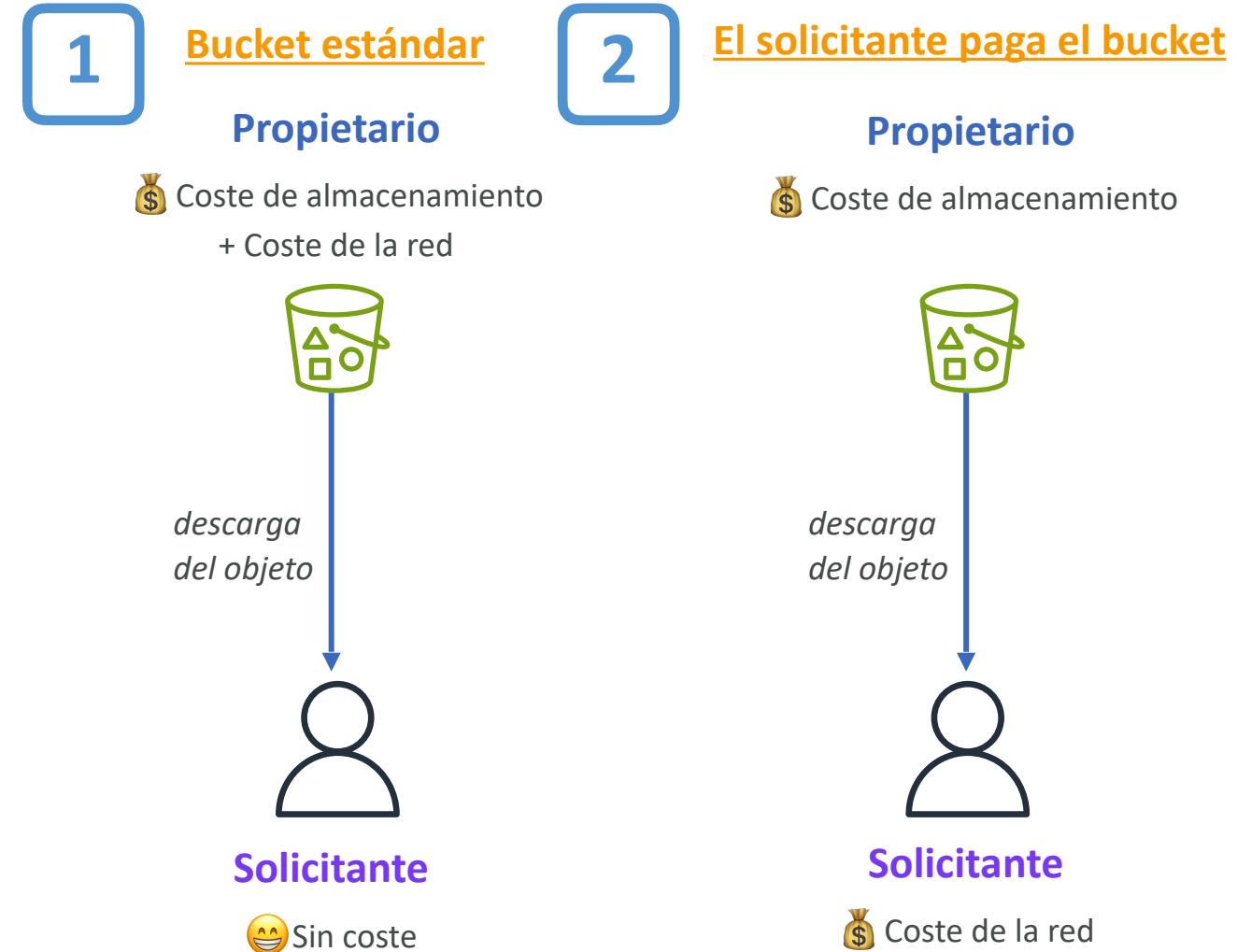
Ejemplo 2 - Reglas del ciclo de vida

- Supongamos que estás diseñando un sistema de copias de seguridad para una base de datos en AWS, donde los snapshots se crean diariamente y deben ser accesibles inmediatamente por un período de 45 días. Sin embargo, después de ese periodo, estos pueden ser archivados, aunque con la posibilidad de recuperación dentro de un día en situaciones de emergencia que podrían suceder ocasionalmente. Los snapshots después de 45 días no necesitarán ser recuperados más que una vez al año. ¿Cómo estructurarías esta solución?
- Implementa el versionado en S3 para los snapshots de la base de datos, asegurando que se mantengan accesibles durante el periodo de 45 días en la clase de almacenamiento Estándar.
- Despues de 45 días, mueve los snapshots a S3 Standard-IA, y tras un año, a S3 Glacier o Glacier Deep Archive, con políticas de ciclo de vida para la transición y recuperación dentro de 24 horas en caso de ser necesarios.

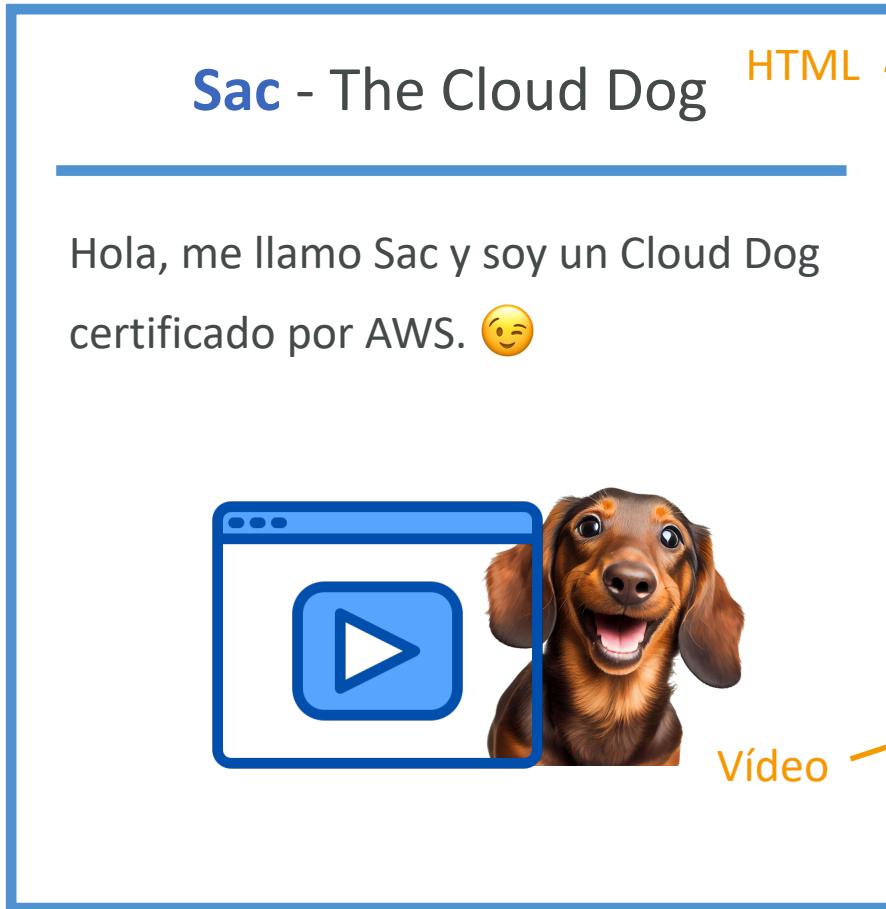
Buckets S3 de pago por solicitante

- Normalmente, los propietarios de los buckets S3 pagan todos los costes de almacenamiento y transferencia de datos de Amazon S3 asociados a su bucket
- No obstante, hay una opción donde **los solicitantes realizan el pago del coste de petición y descarga de datos**
- Es útil cuando quieras compartir grandes conjuntos de datos con otras cuentas

⚠ Si habilitas los pagos por solicitante en un bucket, no se permite el acceso anónimo a ese bucket.



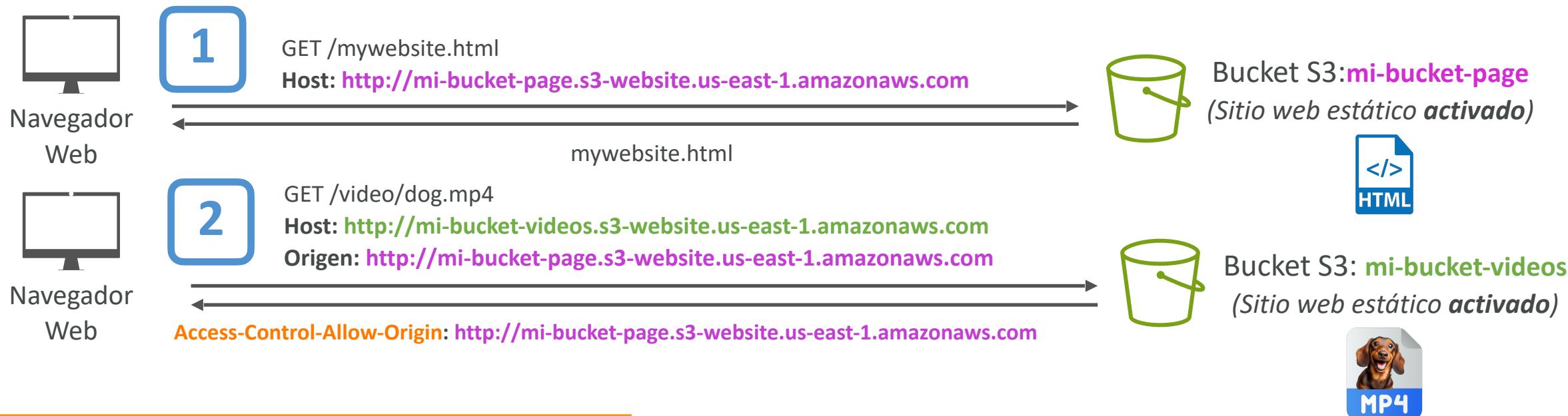
Contextualización del uso de CORS



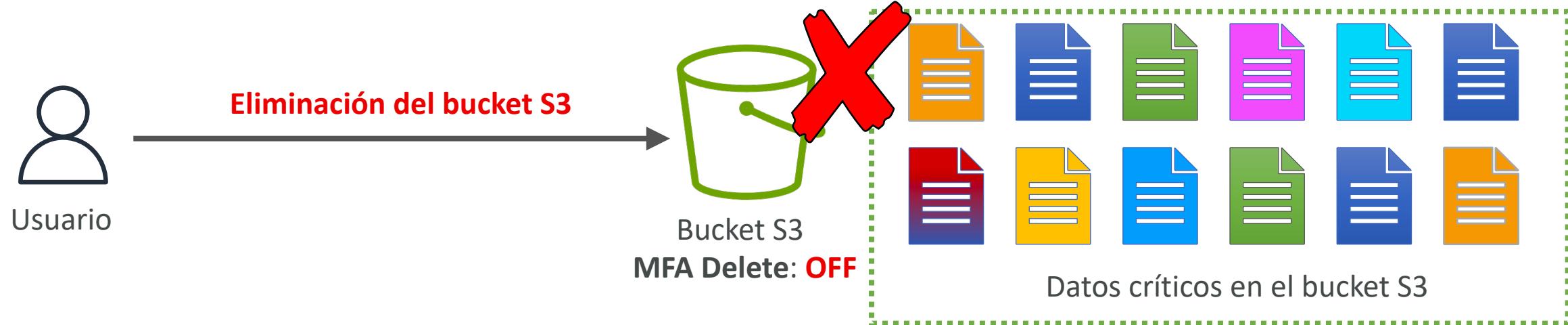
Uso compartido de recursos entre orígenes (CORS)

- Mecanismo **basado en el navegador web** para permitir peticiones a otros orígenes mientras se visita el origen principal
- Las peticiones no se cumplirán a menos que el otro origen permita las peticiones, utilizando **cabeceras CORS**

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

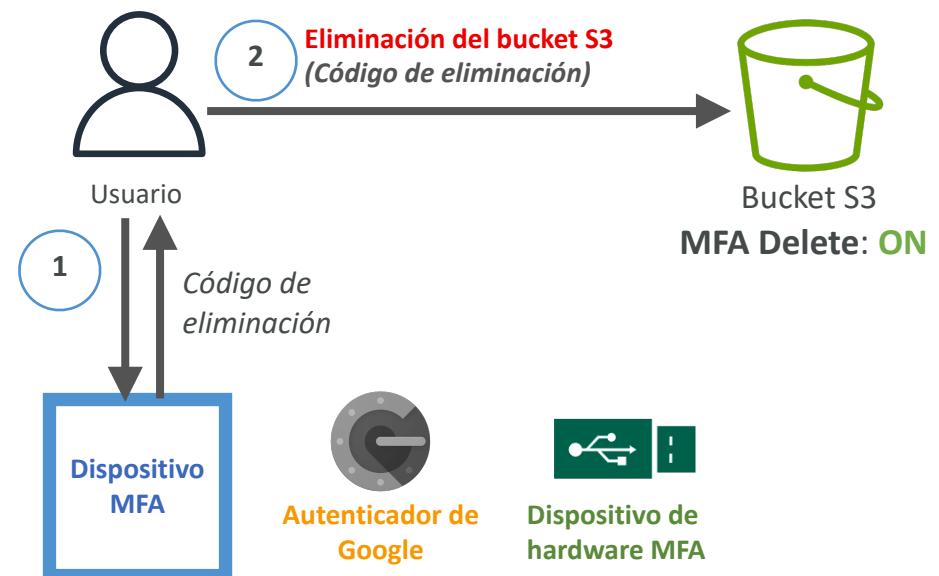


MFA Delete - Evita perder información valiosa



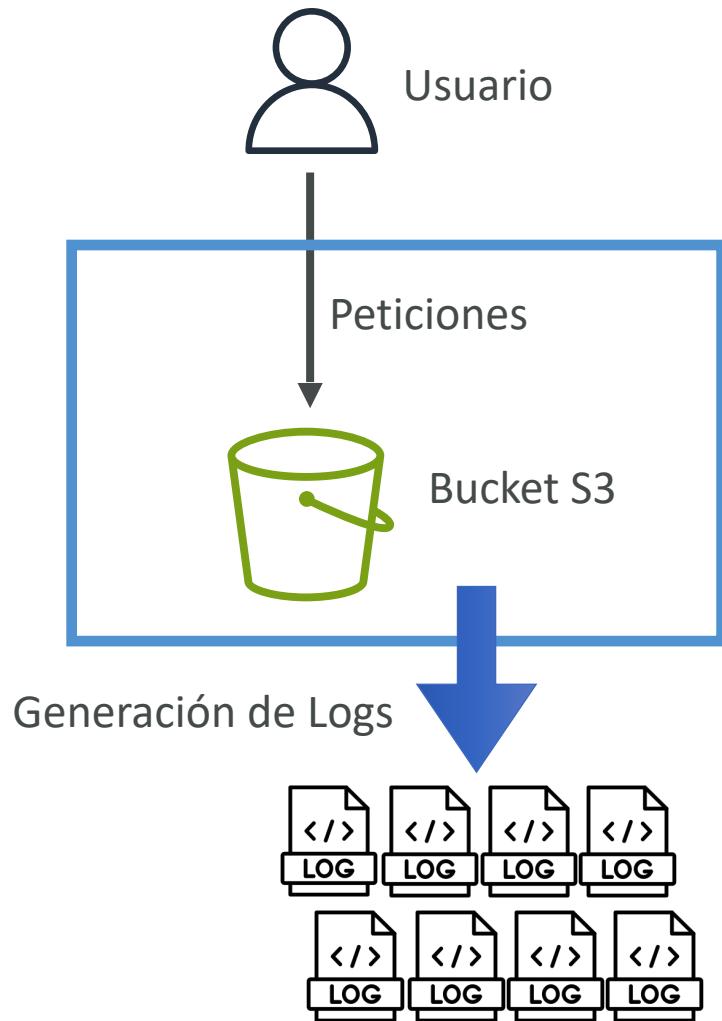
¿Cómo podemos evitar el borrado de información crítica en buckets S3?

- **MFA** obliga a los usuarios a generar un código en un dispositivo 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
- Para utilizar MFA Delete, **el control de versiones debe estar activado** en el bucket S3
- **Sólo el propietario del bucket (cuenta root) puede activar/desactivar MFA Delete**



Logs de acceso de Amazon S3

- 💡 Es crucial registrar todos los **accesos a los buckets de S3** para fines de auditoría
 - Esto permitirá mantener un **seguimiento detallado de todas las operaciones** que se realizan en tus recursos
- Cualquier petición realizada a S3, desde cualquier cuenta, autorizada o denegada, se registrará en **otro** bucket de S3
- El bucket de logs de destino debe estar en la misma región de AWS



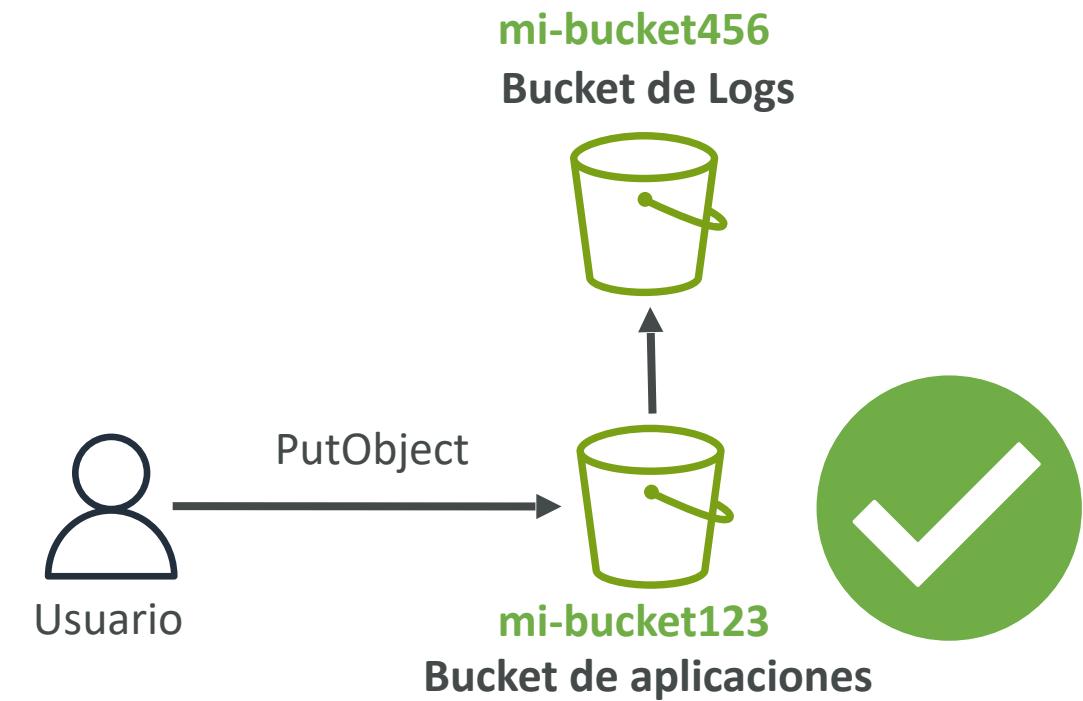
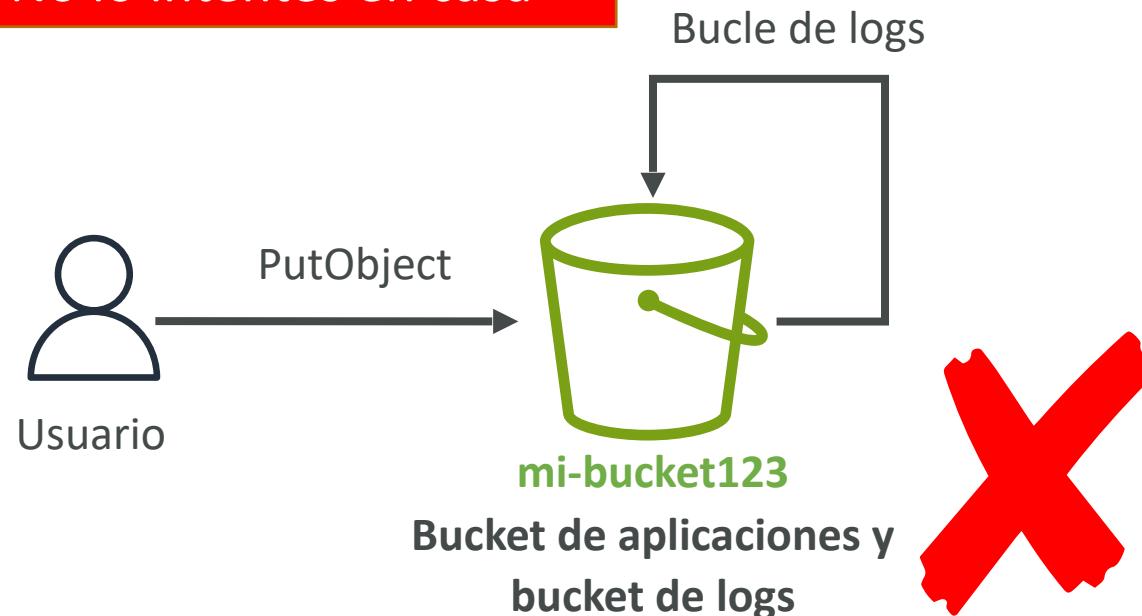


S3 Access Logs



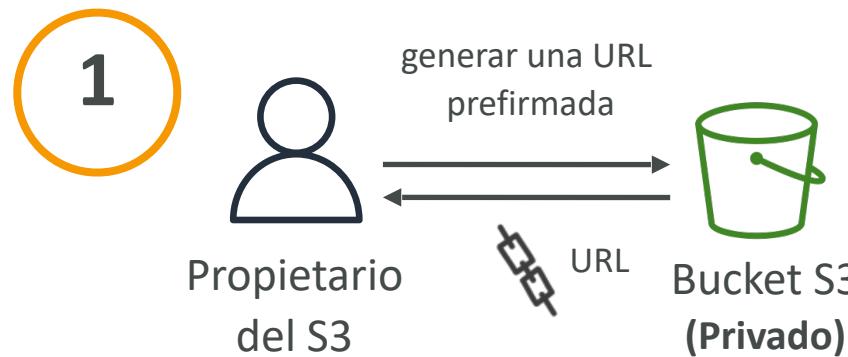
- No configures tu bucket de logs para que sea el bucket monitorizado
- Se creará un bucle de logs, y **tu bucket crecerá exponencialmente**

No lo intentes en casa

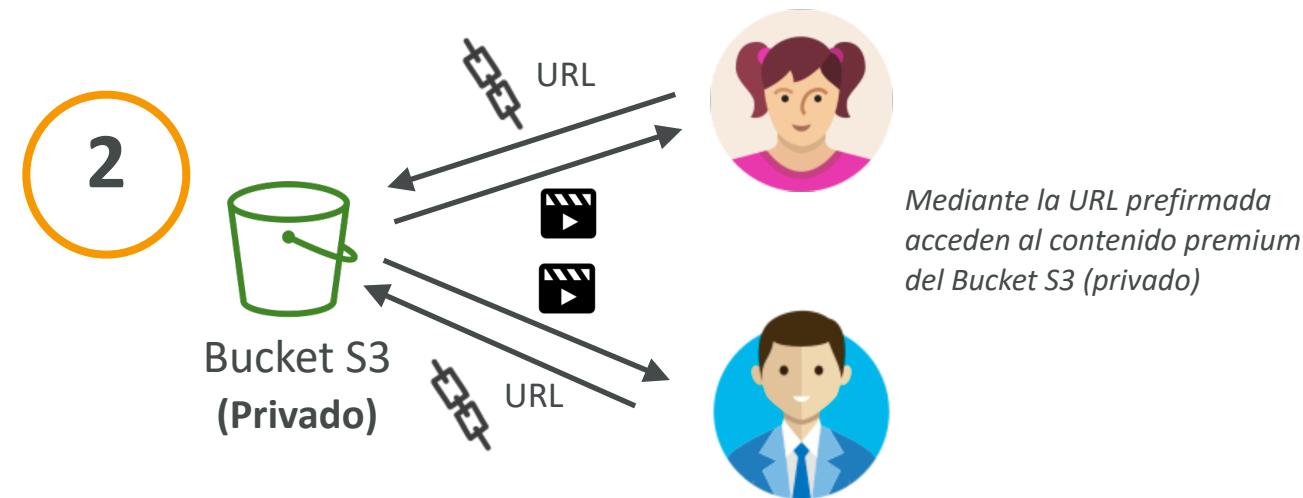


URLs prefirmadas en Amazon S3

- Otorgar **acceso limitado en el tiempo a objetos en Amazon S3** sin necesidad de actualizar su política de buckets
- Generar URLs prefirmadas usando la consola de **S3, la CLI de AWS o el SDK**
- **Tiempo de caducidad de la URL:**
 - **Consola de S3:** 1 minuto - 12 horas
 - **AWS CLI / SDK:** hasta en 7 días



- **Casos de uso de URLs prefirmadas:**
 - Permite que sólo los usuarios que han iniciado sesión descarguen un vídeo premium de un bucket de S3
 - Permitir temporalmente que un usuario suba un archivo a una ubicación precisa en tu bucket de S3



Bloqueo de información con S3 GlacierVault Lock

- Adoptar un **modelo WORM** (Write Once Read Many)
- Crea una “política de bloqueo de bóveda” (**Vault Lock Policy**)
- Bloquea la política para futuras ediciones (ya no se puede modificar ni borrar)
- Los objetos almacenados en el bucket S3 Glacier no podrán ser borrados posteriormente
- Caso de uso común: Útil para el cumplimiento de la normativa y la retención de datos



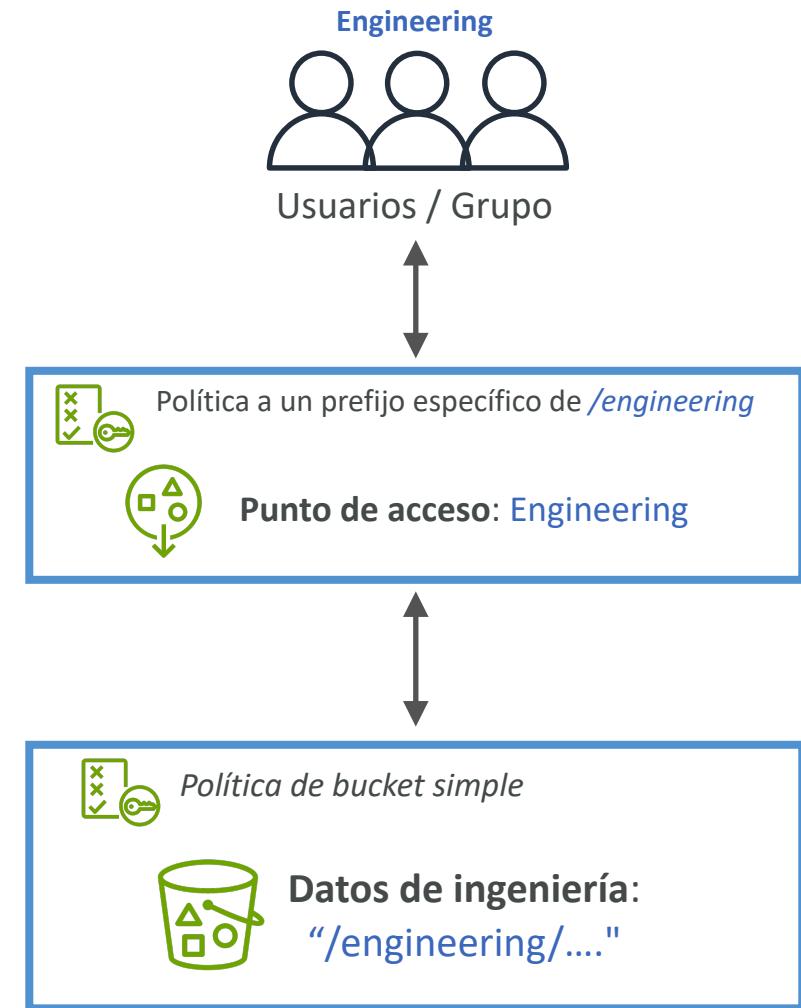
Bloqueo de objetos con S3 Object Lock

- Permite bloquear el borrado de una versión del objeto durante un tiempo determinado
- El tiempo de bloqueo puede ser modificado posteriormente
- Mediante el permiso IAM **s3:PutObjectLegalHold** no es necesario aplicar un período de retención (es indefinido hasta que eliminas el permiso IAM)

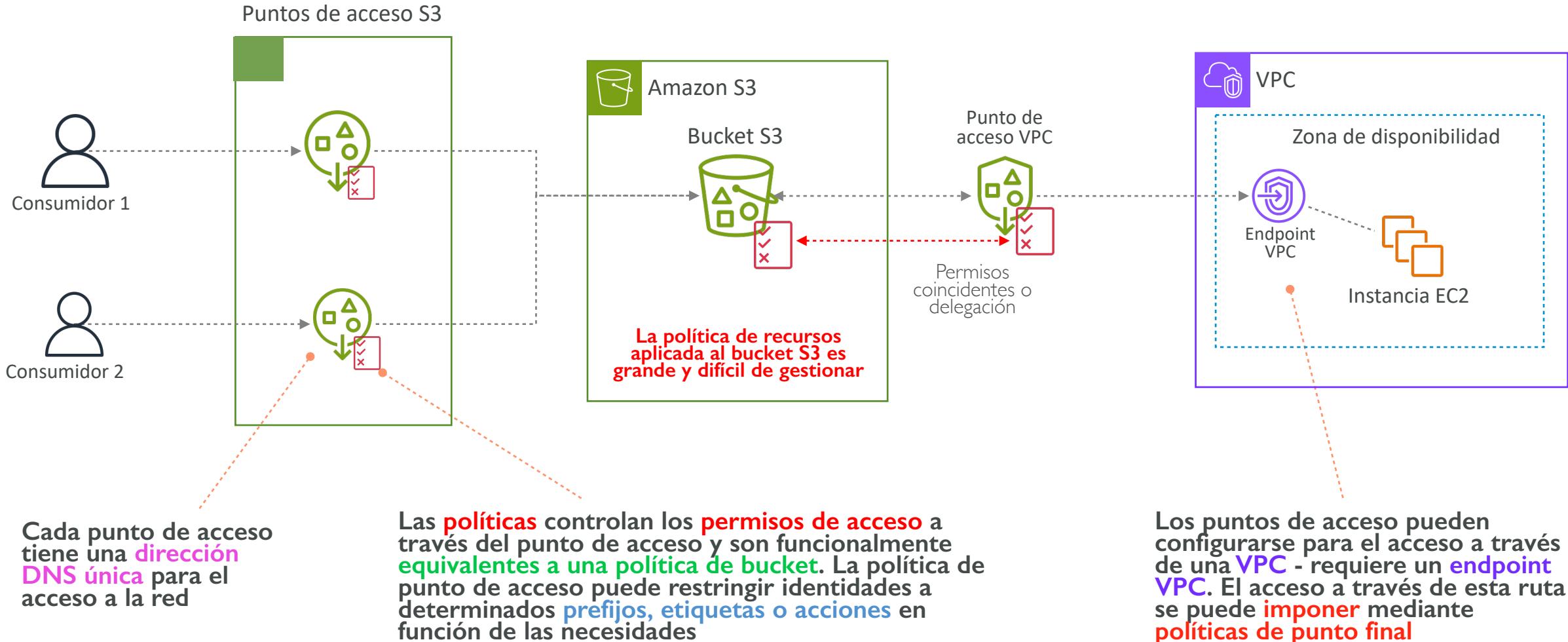


Puntos de acceso (AP) de Amazon S3 (I/2)

- Los puntos de acceso están diseñados para **proporcionar acceso granular a los buckets de S3**
 - Permitiendo a los usuarios crear políticas de acceso específicas para diferentes aplicaciones o usuarios, mejorando la seguridad y la gestión del acceso
- Los usuarios pueden acceder a datos a través de un **ARN de punto de acceso o un alias de punto de acceso**, facilitando la integración con diferentes servicios y aplicaciones
- Cada punto de acceso está configurado con su **propia política de acceso**, adaptada a casos de uso o aplicaciones específicas, y puede admitir desde un solo usuario hasta grupos de usuarios o aplicaciones

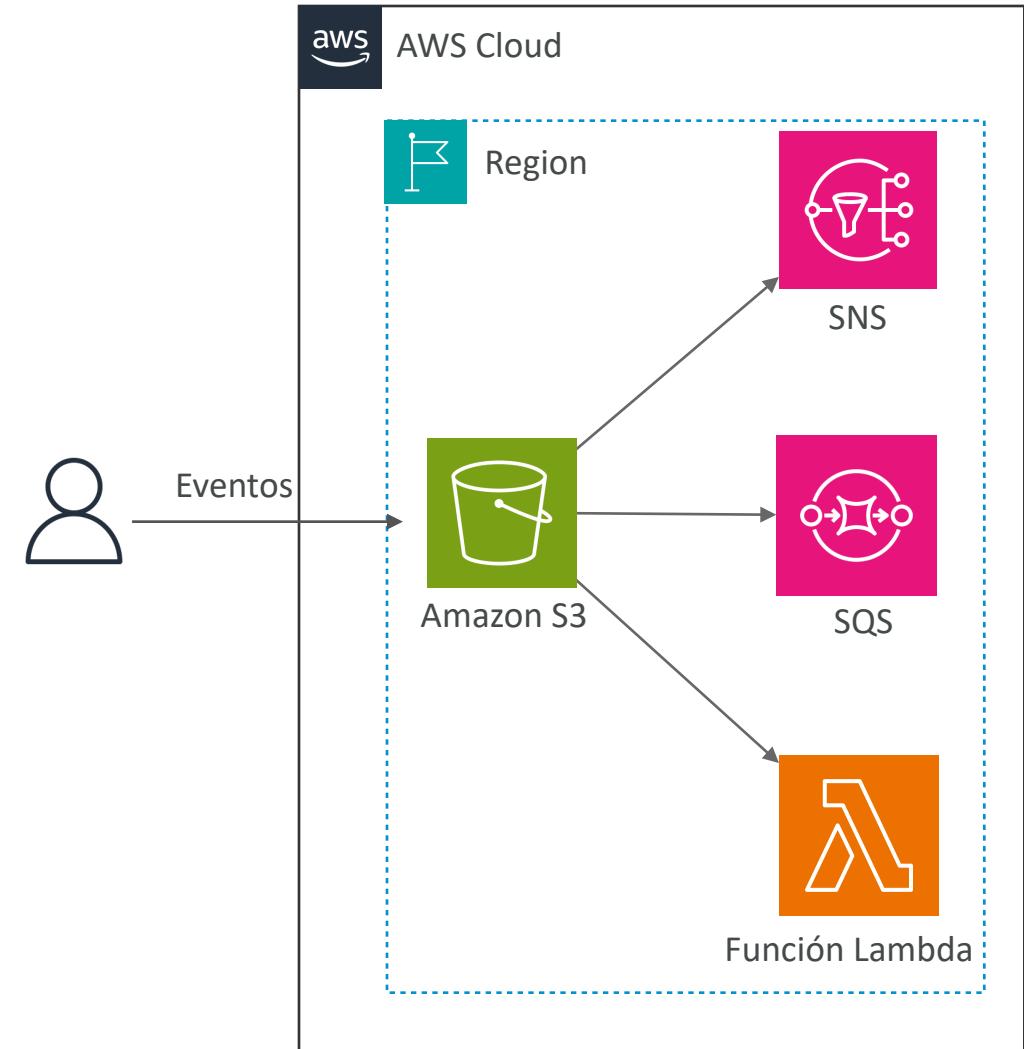


Puntos de acceso (AP) de Amazon S3 (2/2)

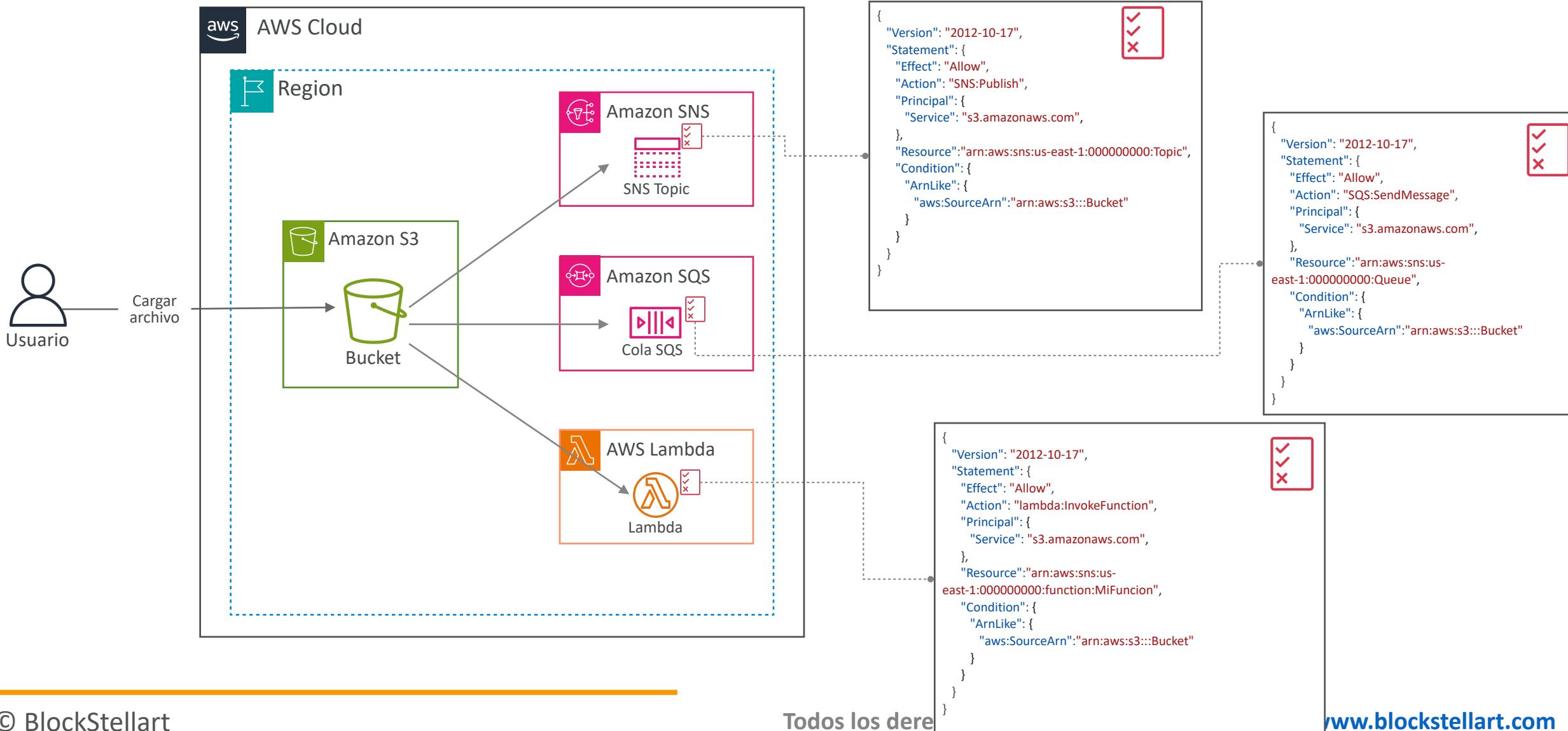


Notificaciones de eventos S3

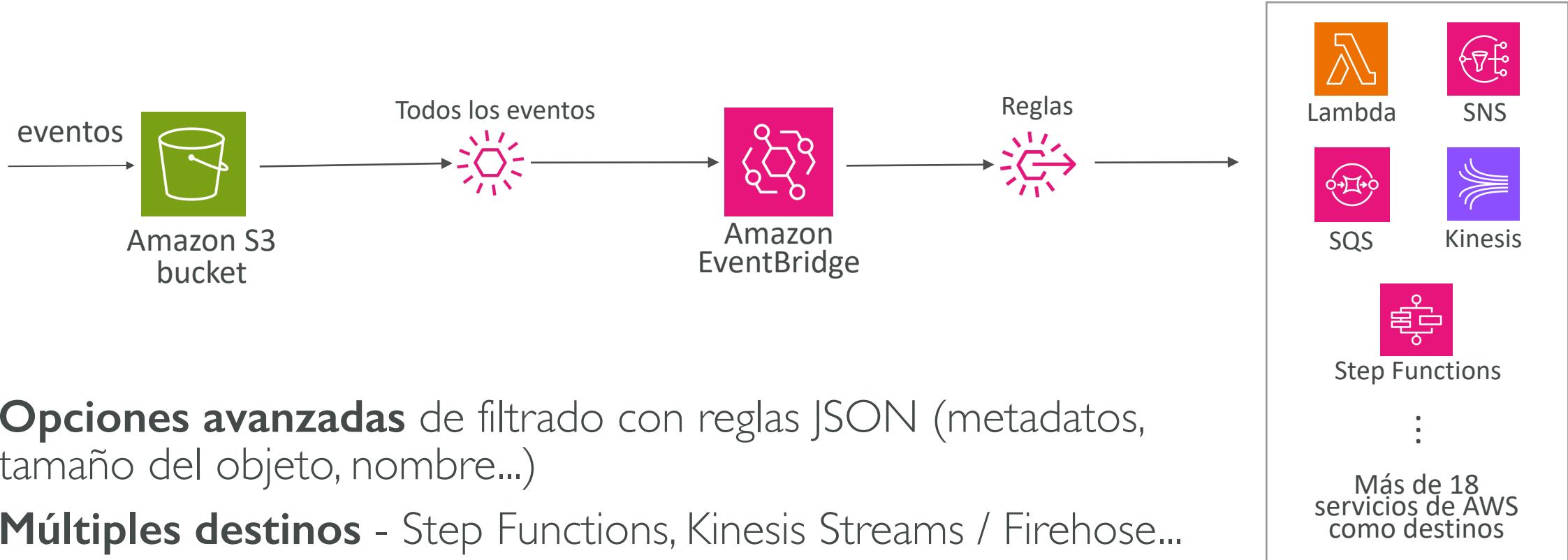
- S3 puede generar eventos como los siguientes:
 - S3:ObjectCreated, S3:ObjectRemoved, S3:ObjectRestore, S3:Replication...
- Posibilidad de filtrar el nombre del objeto (*.jpg)
- **Se pueden crear tantos "eventos S3" como se desee**
- Las notificaciones de eventos S3 suelen entregar los eventos en segundos,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



Notificaciones de eventos S3 - Permisos IAM



Notificaciones de eventos S3 con EventBridge

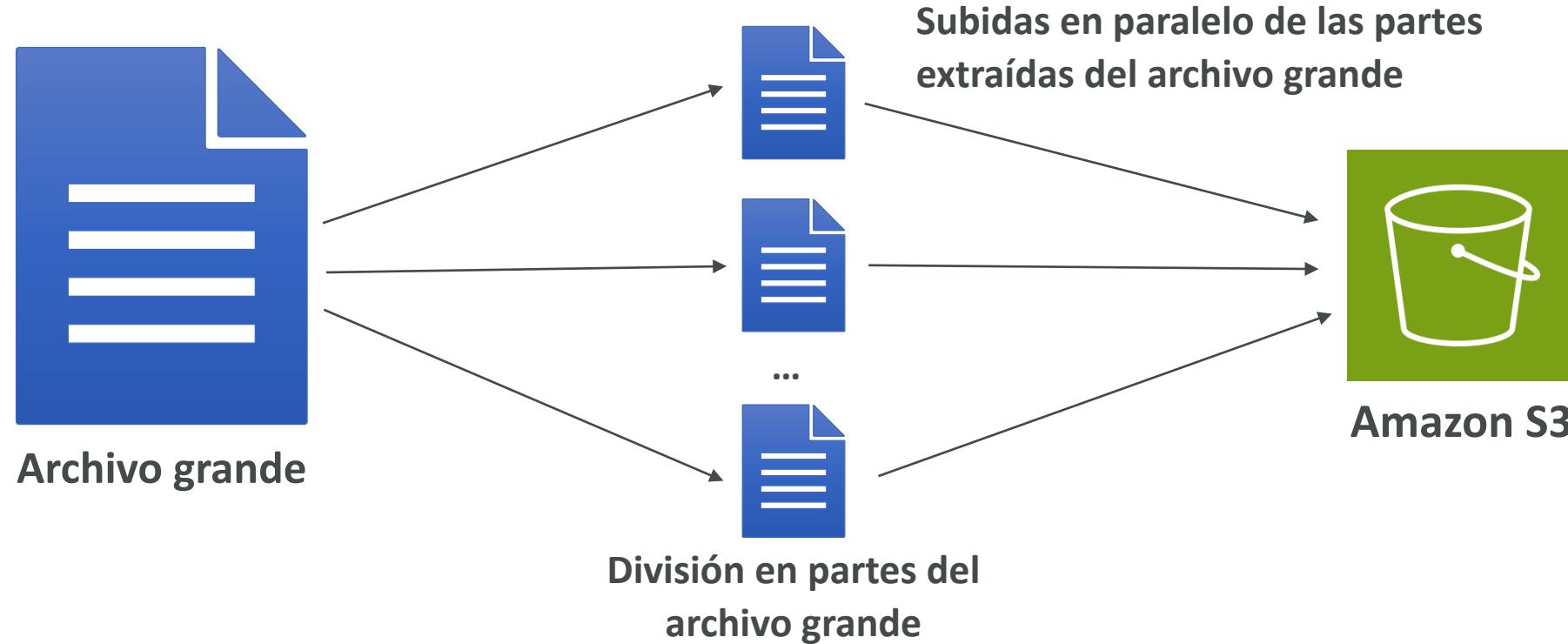


- **Opciones avanzadas** de filtrado con reglas JSON (metadatos, tamaño del objeto, nombre...)
- **Múltiples destinos** - Step Functions, Kinesis Streams / Firehose...
- **Capacidades de EventBridge** - Repetición de eventos, entrega fiable

Rendimiento S3 (1/2)

Carga multiparte

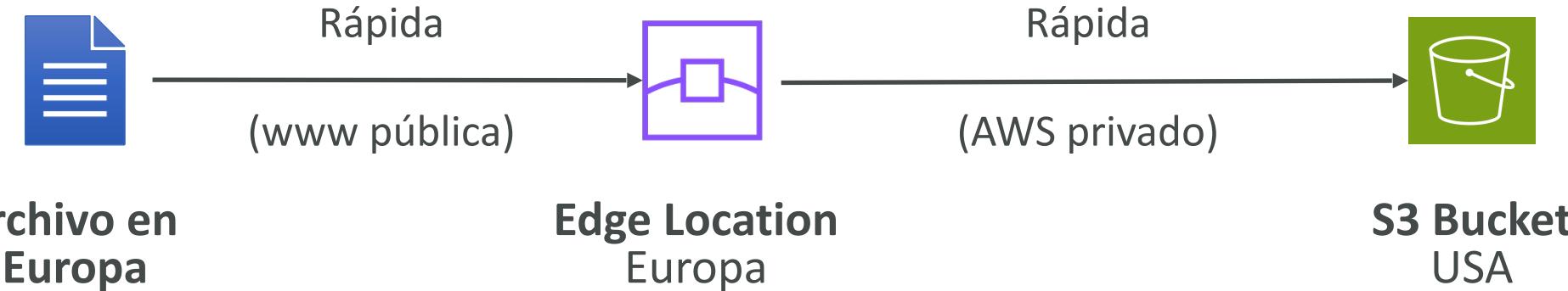
- Recomendado para archivos > 100MB, obligatorio para archivos > 5GB
- Puede ayudar a parallelizar las subidas (acelerar las transferencias)



Rendimiento S3 (2/2)

Aceleración de transferencia en S3

- Utiliza la red global de AWS para **minimizar la latencia** de transferencia de datos
- Aprovecha la optimización de ruta para usar el ancho de banda disponible de forma más eficiente
- Se **beneficia de la red de entrega de contenido (CDN)** de AWS para acelerar la entrega de archivos
- Utiliza protocolos seguros para proteger los datos durante la transferencia
- **Reduce los costos operativos** al disminuir el tiempo de transferencia y potencialmente el tráfico saliente
- Ideal para la migración de grandes volúmenes de datos a la nube de AWS
- Compatible con la carga multiparte



Amazon S3 Select y Amazon S3 Glacier Select

- **Amazon S3 Select** permite utilizar instrucciones SQL para **filtrar el contenido de los objetos de S3 y recuperar exactamente el subconjunto de datos que necesitas**
- **Amazon S3 Glacier Select** permite ejecutar consultas directamente en los datos almacenados en **Amazon Glacier** y recupera solo los datos necesarios de los archivos para utilizarlos en el análisis
- Menos transferencia de red, menos coste de CPU en el lado del cliente
- Puedes filtrar por filas y columnas (simples sentencias SQL)



Ejemplos: Utilizar Amazon S3 Select para buscar y recuperar únicamente los registros que coincidan con una cuenta determinada, o solo los datos de facturación para un cliente específico

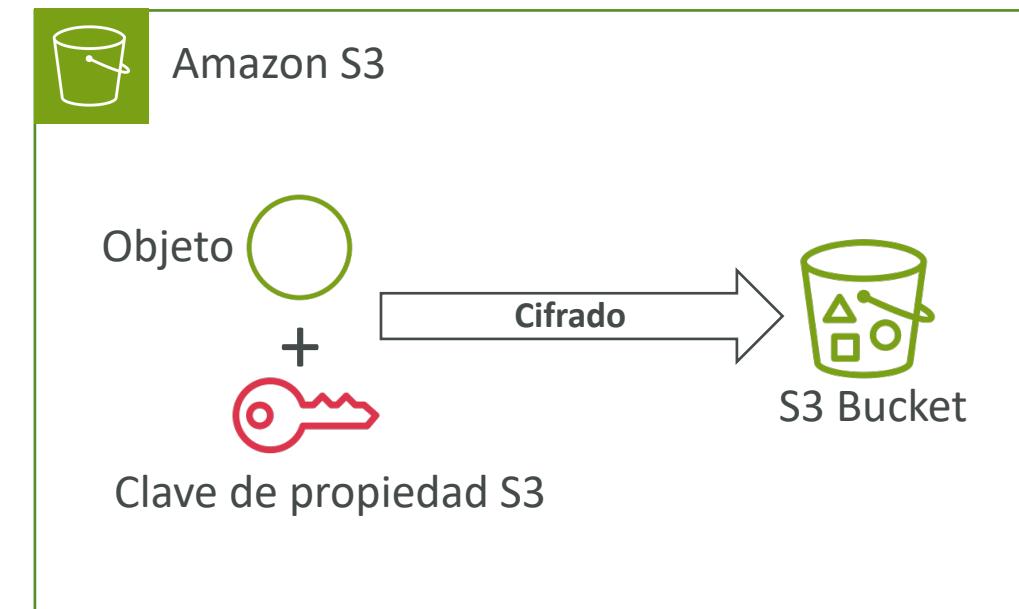
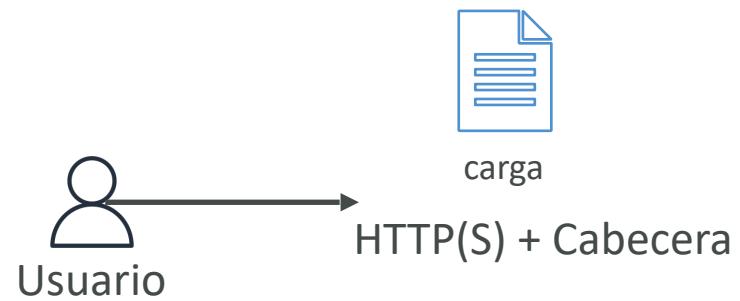
Amazon S3 - Cifrado de objetos

- Puedes cifrar objetos en buckets de S3 utilizando uno de los 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 exámenes de certificación de AWS



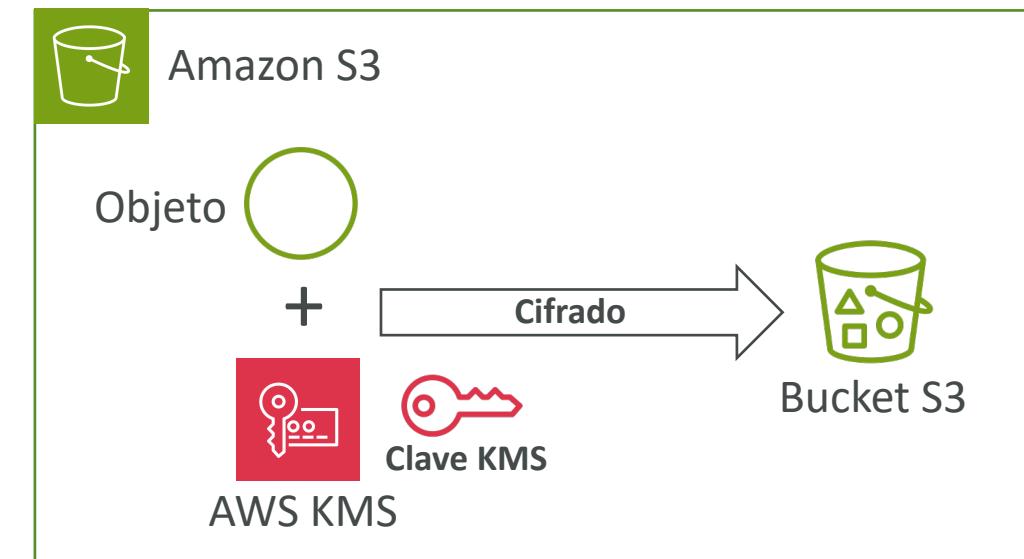
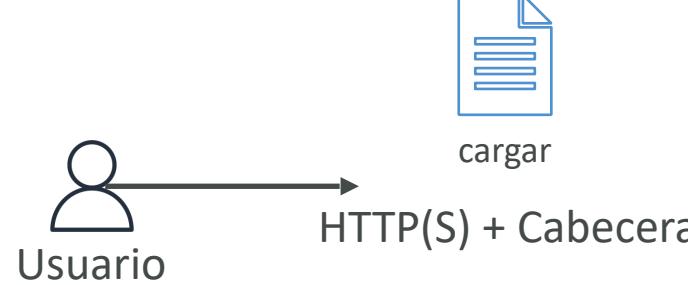
Cifrado de Amazon S3 - SSE-S3

- 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**



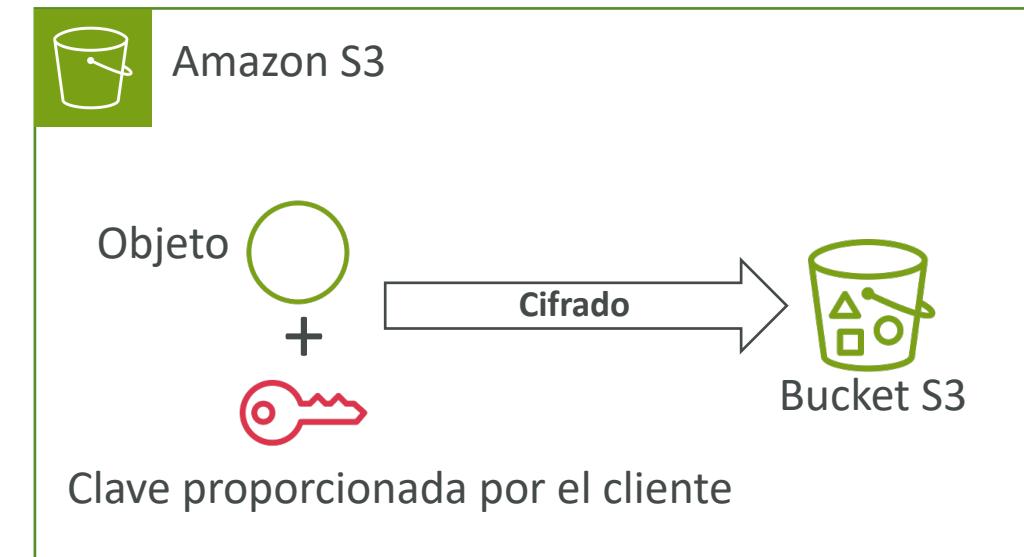
Cifrado de Amazon S3 - SSE-KMS

- 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"**



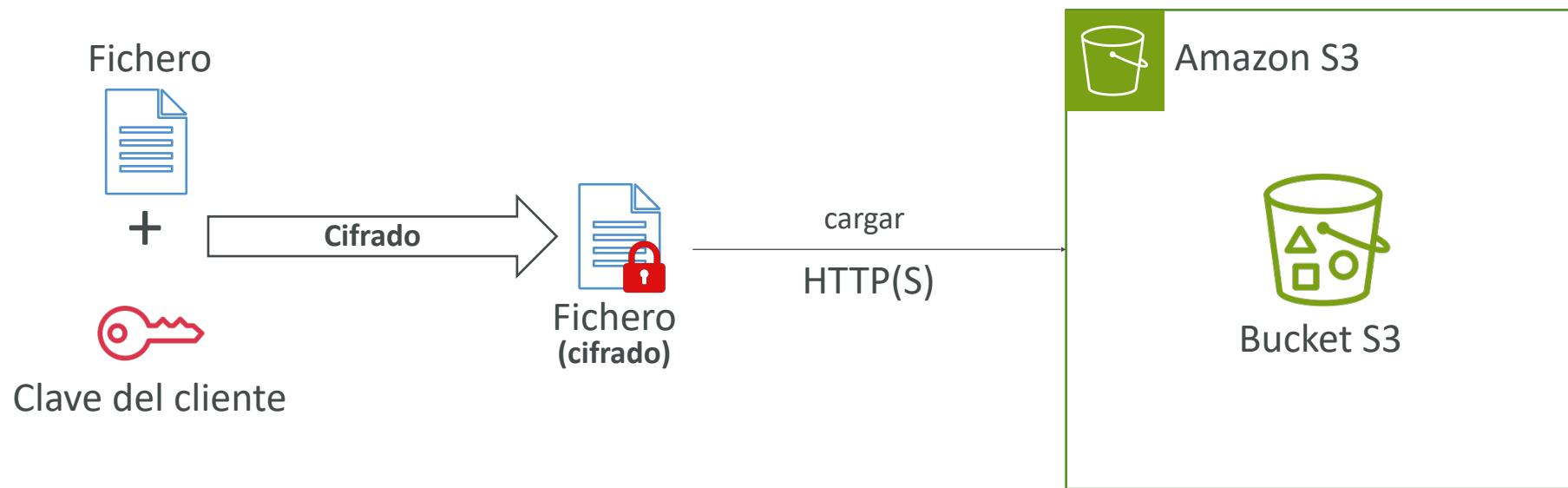
Cifrado de Amazon S3 - SSE-C

- 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



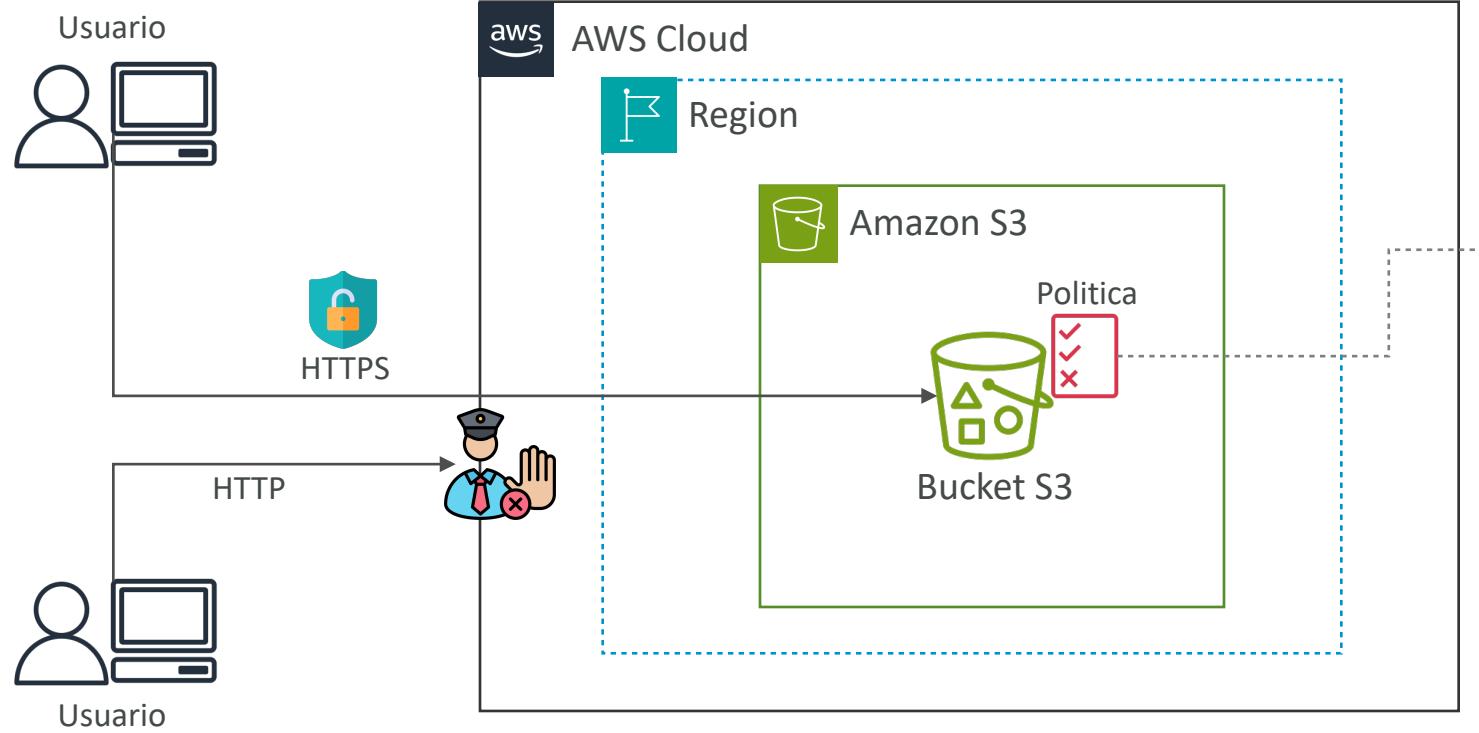
Amazon S3 - Cifrado en tránsito (SSL/TLS)

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



Forzado del cifrado en tránsito en Amazon S3

Importancia del cifrado en tránsito para la protección contra ataques de tipo "man-in-the-middle"



```
{  
  "Version": "2012-10-17",  
  "Statement": [  
    {  
      "Effect": "Deny",  
      "Action": "s3:GetObject",  
      "Principal": "*",  
      "Resource": "arn:aws:s3:::bucket/*",  
      "Condition": {  
        "Bool": {  
          "aws:SecureTransport": "false"  
        }  
      }  
    }  
  ]  
}
```

Uso de **HTTPS** para proteger la integridad y la confidencialidad de los datos entre el cliente y el bucket S3

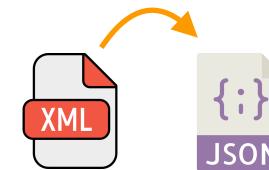
Implementación de políticas de seguridad en S3 para garantizar el cifrado en tránsito

Amazon S3 Object 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:

- Convertir entre formatos de datos, como convertir XML a JSON



- Extraer partes específicas de archivos, como ciertas columnas de un archivo CSV, antes de enviarlos al usuario final



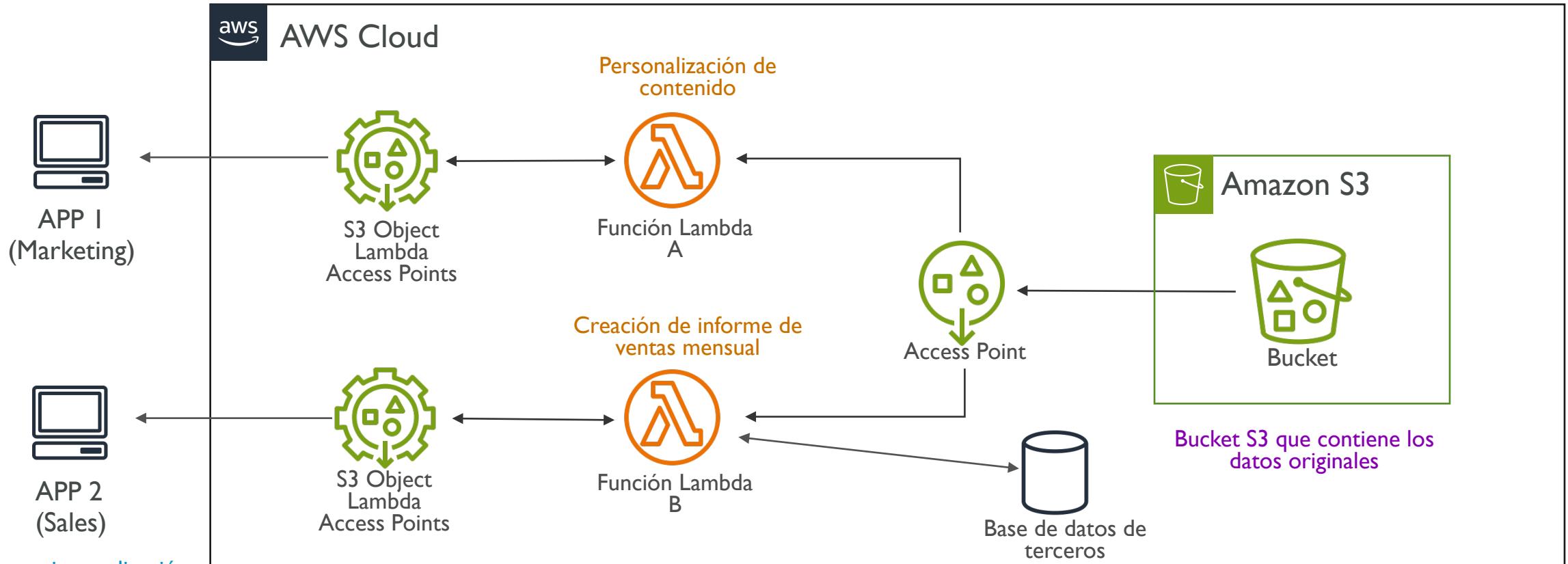
- Automáticamente ocultar o eliminar información confidencial, como números de tarjetas de crédito o datos personales



- Redimensionar y poner marcas de agua a las imágenes



Amazon S3 Object Lambda



Un usuario o aplicación realiza una solicitud para acceder a un objeto almacenado en un bucket de S3 utilizando un punto de acceso de Lambda de objetos S3

Estos puntos de acceso controlan las solicitudes de cada una de las aplicaciones que piden datos de S3

En algunos casos se puede necesitar consultar o actualizar datos en una base de datos externa para completar la solicitud



Almacenamiento avanzado en AWS

www.blockstellart.com

Todos los derechos reservados © BLOCKSTELLART www.blockstellart.com

Familia AWS Snow

- Dispositivos portátiles de alta seguridad que permiten la **recopilación** de datos, el **procesamiento** de datos y el **movimiento** de datos desde y hacia AWS



Migración de datos



Edge computing

Familia AWS Snow

Migración de datos

Snowcone



Snowball Edge



Snowmobile



Edge computing

Snowcone



Snowball Edge



Migraciones de datos con AWS Snow

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!

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

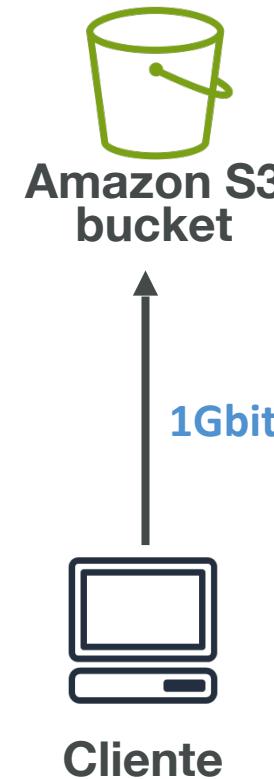
Datos a enviar

Velocidad de transmisión

| Datos a enviar | 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 |

Diagramas de almacenamiento de datos

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.** Es la alternativa a mover datos a través de la red (y pagar tarifas de red)
- Casos de uso: migraciones al Cloud de grandes volúmenes de datos, recuperación ante desastres
- Proporciona almacenamiento de bloques y almacenamiento de objetos compatible con Amazon S3
- Dos variantes para distintos requerimientos:
 - **Optimizado para almacenamiento:**
 - 80 TB de capacidad de HDD para volumen de bloques y almacenamiento de objetos compatible con S3
 - **Optimizado para cómputo:**
 - 42 TB de capacidad HDD o 28 TB de capacidad NVMe para volumen de bloques y almacenamiento de objetos compatible con S3
- Cifrado de datos con claves de 256 bits gestionadas por AWS KMS y cumplimiento con los estándares HIPAA y FedRAMP





AWS Snowcone

AWS Snowcone es un dispositivo pequeño (4,5 libras - 2,1 kg), resistente y seguro que ofrece computación de borde, almacenamiento y transferencia de datos en cualquier momento y lugar en entornos austeros con conectividad escasa o nula.

- Dispositivo utilizado para computación de borde, almacenamiento y transferencia de datos
- Utiliza AWS Snowcone en entornos con limitaciones de espacio
- Debes proporcionar tu propia batería / cables
- Disponible en versiones de:
 - **HDD - 8 TB de almacenamiento utilizable**
 - **SSD - 14 TB de almacenamiento utilizable**
- Se puede enviar a AWS sin conexión, o conectarlo a internet





AWS Snowmobile

- Transfiere exabytes de datos (1 EB = 1.000 PB = 1.000.000 TBs)
- Cada Snowmobile tiene 100 PB de capacidad (utiliza varias en paralelo)
- **Mejor que la Snowball si transfieres más de 10 PB**
- Alta seguridad:
 - Temperatura controlada, GPS, videovigilancia 24/7
 - Viene equipado con un sistema de cifrado end-to-end para asegurar los datos durante el tránsito.
 - Amazon proporciona guardias de seguridad para el contenedor durante el transporte, si el cliente lo solicita.



Matriz de comparación de características



Snowcone



Snowball Edge



Snowmobile

| | AWS SNOWCONE | AWS SNOWBALL EDGE STORAGE OPTIMIZED 80 TB | AWS SNOWBALL EDGE STORAGE OPTIMIZED 210 TB | AWS SNOWBALL EDGE COMPUTE OPTIMIZED | AWS SNOWMOBILE |
|------------------------------------------------|-----------------------------------------------------------|----------------------------------------------|-----------------------------------------------|-------------------------------------------|--------------------------------|
| Almacenamiento HDD utilizable | 8 TB | Unidad de disco duro de 80 TB | N/D | N/D | 100 PB |
| Almacenamiento SSD disponible | 14 TB | 1 TB | NVMe de 210 TB | 28 TB | No |
| CPU virtuales útiles | 2 CPU virtuales | 40 vCPU | 104 vCPU | 104 vCPU | N/D |
| Memoria útil | 4 GB | 80 GB | 416 GB | 416 GB | N/D |
| Tamaño del dispositivo | 9 in x 6 in x 3 in 227 mm x 148,6 mm x 82,65 mm | 548 mm x 320 mm x 501 mm | 548 mm x 320 mm x 501 mm | 548 mm x 320 mm x 501 mm | Contenedor de envío de 13,71 m |
| Peso del dispositivo | 2,1 kg (4,5 lb) | 22,3 kg (49,7 lb) | 22,3 kg (49,7 lb) | 22,3 kg (49,7 lb) | N/D |
| Almacenamiento de clústeres | No | No | No | Sí, de 3 a 16 nodos | N/D |
| Cifrado de 256 bits | Sí | Sí | Sí | Sí | Sí |
| Conformidad con los requisitos de HIPAA | No | Sí | Sí | Sí | Sí |

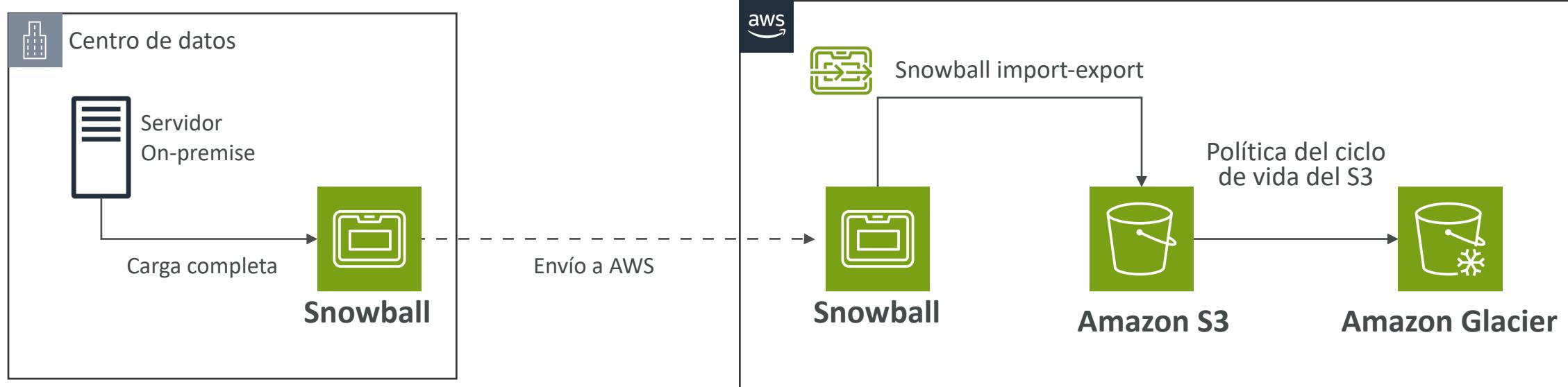
Proceso de uso de los dispositivos AWS Snow

1. **Solicitud de Snowball:** Accede a la consola de AWS y solicita un dispositivo Snowball Edge adecuado para tus necesidades de cómputo o almacenamiento.
2. **Instalación de cliente o AWS OpsHub:** Descarga e instala el cliente de Snowball o AWS OpsHub para gestionar el dispositivo Snowball Edge.
3. **Conexión y transferencia:** Conecta el dispositivo Snowball Edge a tu red y comienza la transferencia de datos desde tus servidores utilizando el cliente o AWS OpsHub.
4. **Empaque y devolución:** Una vez finalizada la transferencia, prepara el dispositivo Snowball Edge para su devolución siguiendo las instrucciones proporcionadas.
5. **Carga de datos en S3:** AWS recibirá el dispositivo, conectará los datos al bucket de S3 especificado durante la configuración.
6. **Borrado seguro de datos:** AWS se asegurará de que todos los datos en el dispositivo Snowball Edge sean borrados de forma segura.
7. **Confirmación y reporte:** Recibirás una notificación cuando tus datos estén seguros en S3, junto con un reporte de auditoría del proceso de transferencia.
8. **Monitoreo y gestión:** Utiliza la consola de AWS para monitorear el progreso y administrar tus recursos de AWS asociados con la transferencia.

Envío de datos de AWS Snowball a un S3 Glacier

Diseño de arquitectura

- **Snowball no puede importar datos directamente a S3 Glacier**
- Debes utilizar primero Amazon S3, en combinación con una política de ciclo de vida de S3
- Ideal para **datos que requieren retención a largo plazo** por regulaciones de cumplimiento



Visión general de Amazon FSx



Amazon FSx es un servicio de almacenamiento de archivos completamente administrado que facilita la implementación, ejecución y escalado de sistemas de archivos ricos en funciones y de alto rendimiento en la nube



FSx para Windows
File Server



FSx para Lustre



FSx para
ONTAP de NetApp

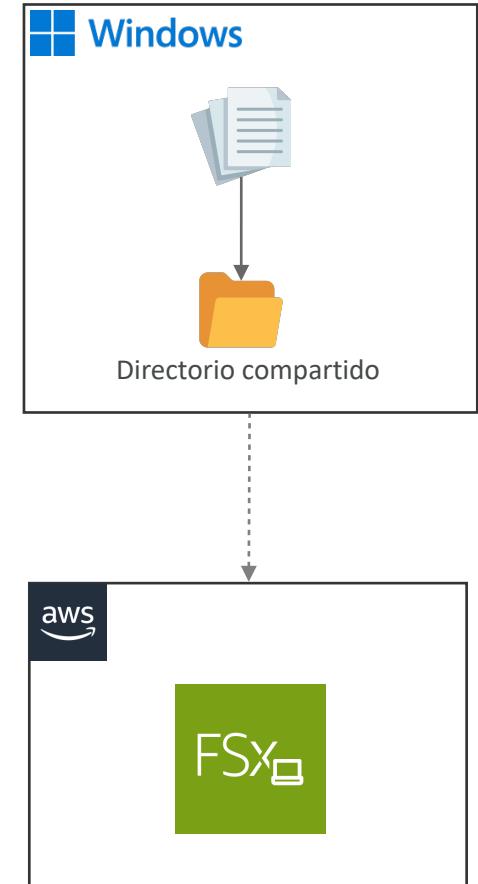


FSx para
OpenZFS

Amazon FSx para Windows



- **FSx para Windows** es una unidad compartida del sistema de archivos de **Windows** totalmente gestionada
- Soporta el protocolo SMB y el NTFS de Windows
- **Se puede montar en instancias EC2 de Linux**
- 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

FSx

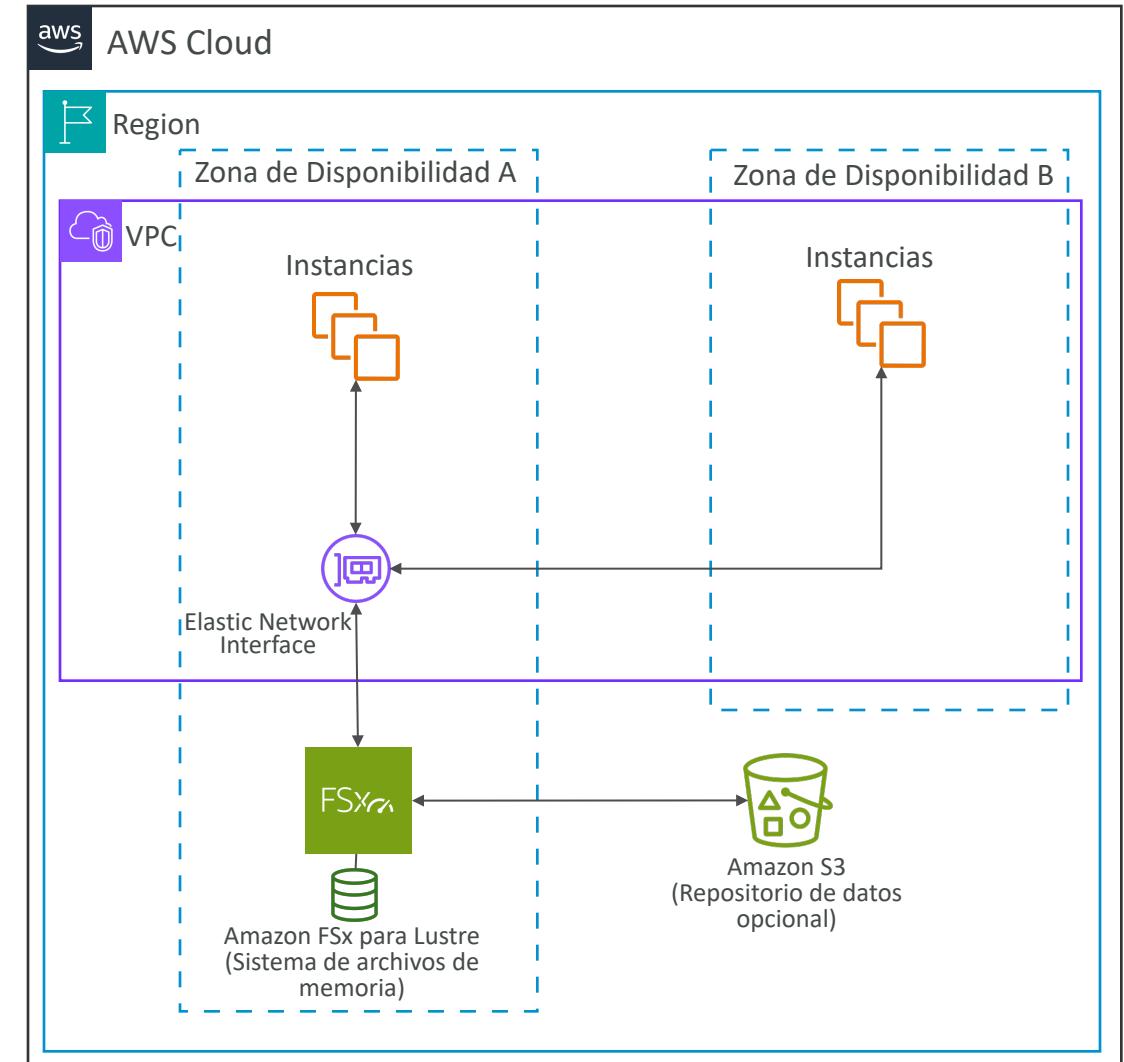
- Lustre es un tipo de sistema de archivos distribuido en paralelo, para la informática a gran escala
- El nombre Lustre deriva de "Linux" y "Cluster"
- Escala hasta 100s GB/s, millones de IOPS, latencias sub-ms
- Opciones de almacenamiento:
 - **SSD** - baja latencia, cargas de trabajo intensivas en 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 de Amazon S3 como un sistema de archivos (a través de FSx)
 - Puede escribir la salida de los cálculos de vuelta a Amazon S3 (a través de FSx)
- Puede utilizarse desde servidores locales (VPN o Direct Connect)

FSx Lustre - Opciones de despliegue del sistema de archivos (1/2)

Sistema de archivos en memoria

Ideal para cargas de trabajo temporales, como procesamiento de datos, análisis y simulaciones.

- **Alto rendimiento:** Diseñado para cargas de trabajo intensivas en I/O
- **Almacenamiento temporal:** Los datos no se conservan a largo plazo
- **Baja latencia:** Acceso rápido a los datos
- **No es duradero:** No se realiza una copia de seguridad de los datos en Amazon S3
- **Uso recomendado:** Experimentación, procesamiento de lotes y tareas temporales

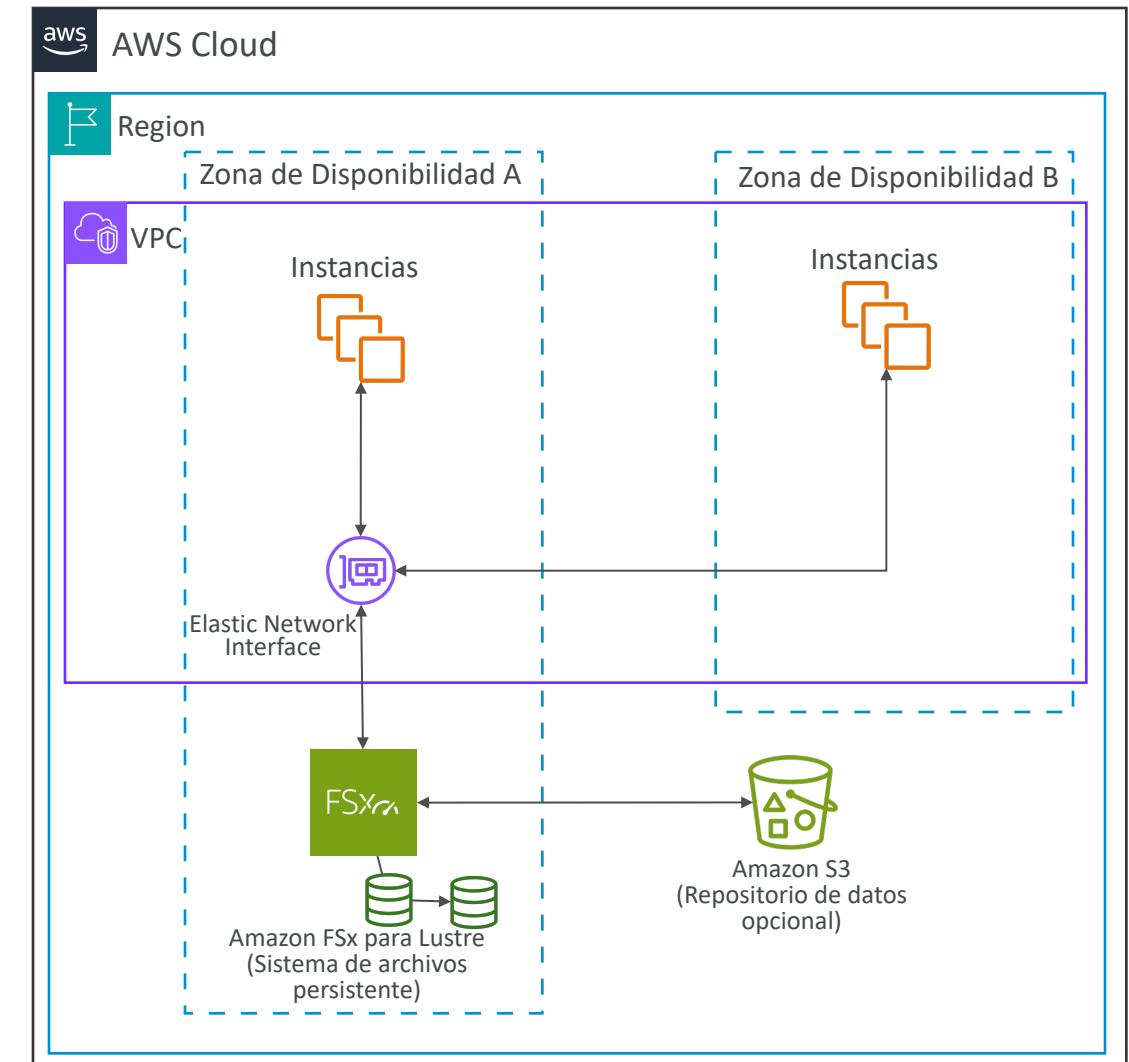


FSx Lustre - Opciones de despliegue del sistema de archivos (2/2)

Sistema de archivos persistente

Adequado para cargas de trabajo que requieren almacenamiento a largo plazo y alta durabilidad.

- **Alto rendimiento:** Similar al modo scratch, pero con la ventaja de la persistencia
- **Almacenamiento duradero:** Los datos se conservan incluso después de que se detenga la instancia
- **Copia de seguridad en Amazon S3:** Los datos se vinculan a un repositorio de datos de Amazon S3 para almacenamiento a largo plazo
- **Mayor latencia que en memoria:** Debido a la persistencia y la copia de seguridad
- **Uso recomendado:** Almacenamiento a largo plazo, cargas de trabajo continuas y aplicaciones críticas



Amazon FSx para NetApp ONTAP

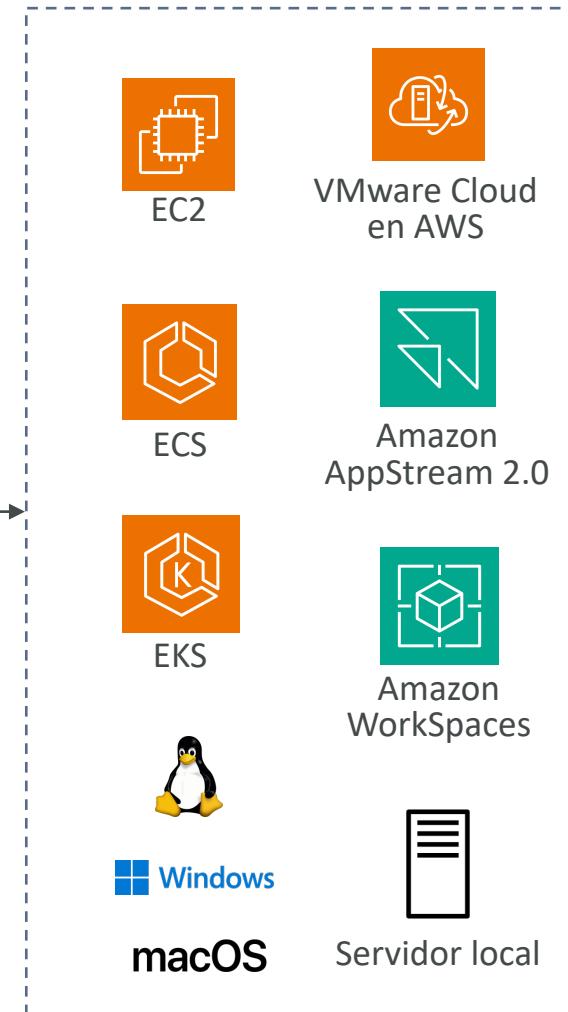


- NetApp ONTAP gestionado en AWS
- **Sistema de archivos compatible con el protocolo NFS, SMB, iSCSI**
- Funciona con:
 - Linux
 - Windows
 - MacOS
 - VMware Cloud en AWS
 - Amazon Workspaces y AppStream 2.0
 - Amazon EC2, ECS y EKS
- El almacenamiento se reduce o crece automáticamente
- Snapshots, replicación,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

Amazon FSx para
NetApp ONTAP FS



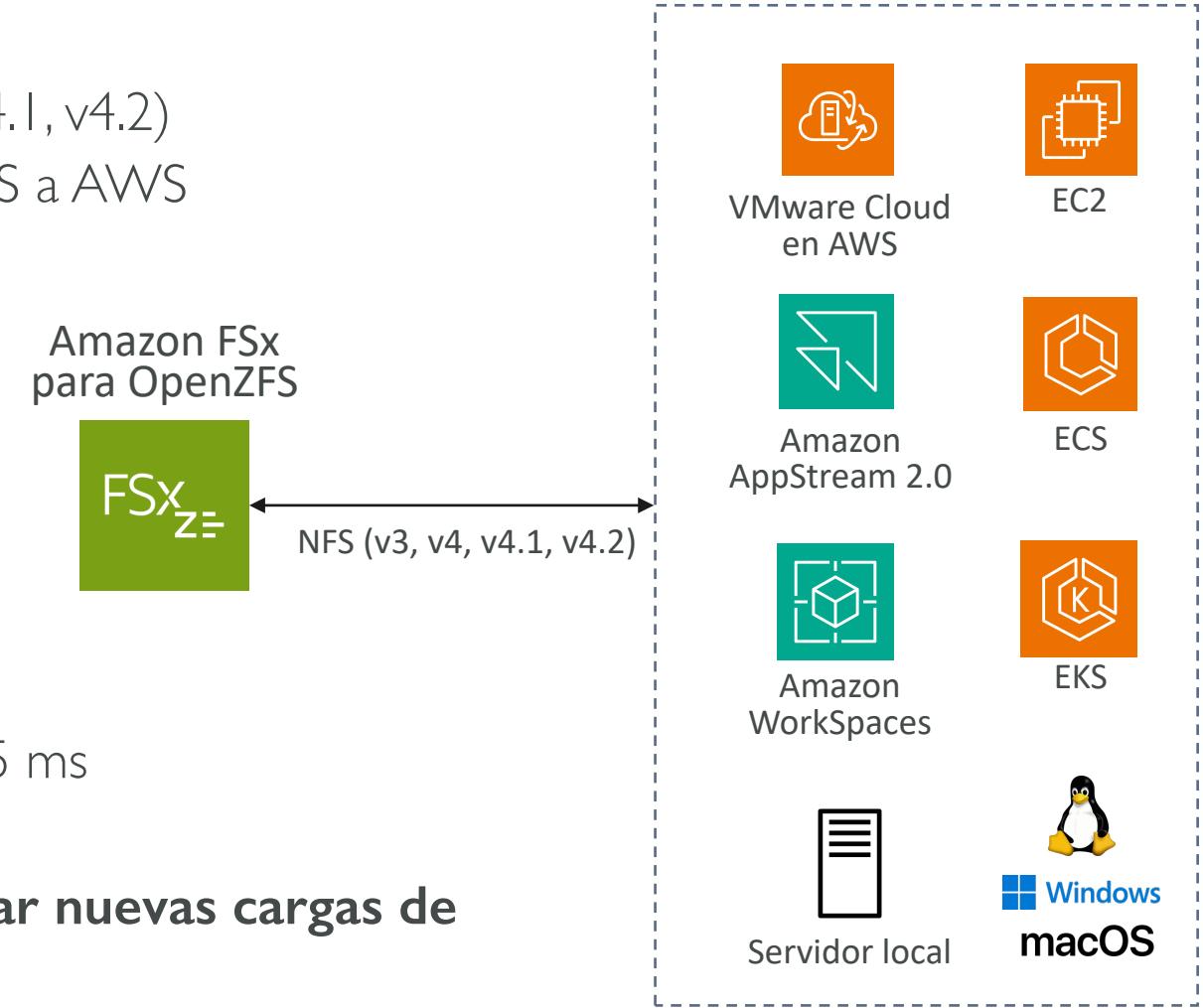
NFS, SMB, iSCSI



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)**

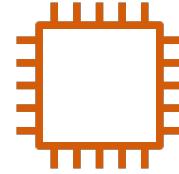


Opciones nativas de almacenamiento de AWS

Bloque



Amazon EBS



Almacén de
instancias EC2

Fichero



Amazon EFS



Amazon FSx

Objeto



Amazon S3



Amazon Glacier

Cloud híbrido para el almacenamiento

- AWS está impulsando el Cloud híbrido
 - Cloud + instalaciones locales
- ¿Por qué? - 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 de AWS, así que...
 - **¿cómo expones los datos de S3 en las instalaciones?** 🤔
 - **Usando AWS Storage Gateway!!**



AWS Storage Gateway



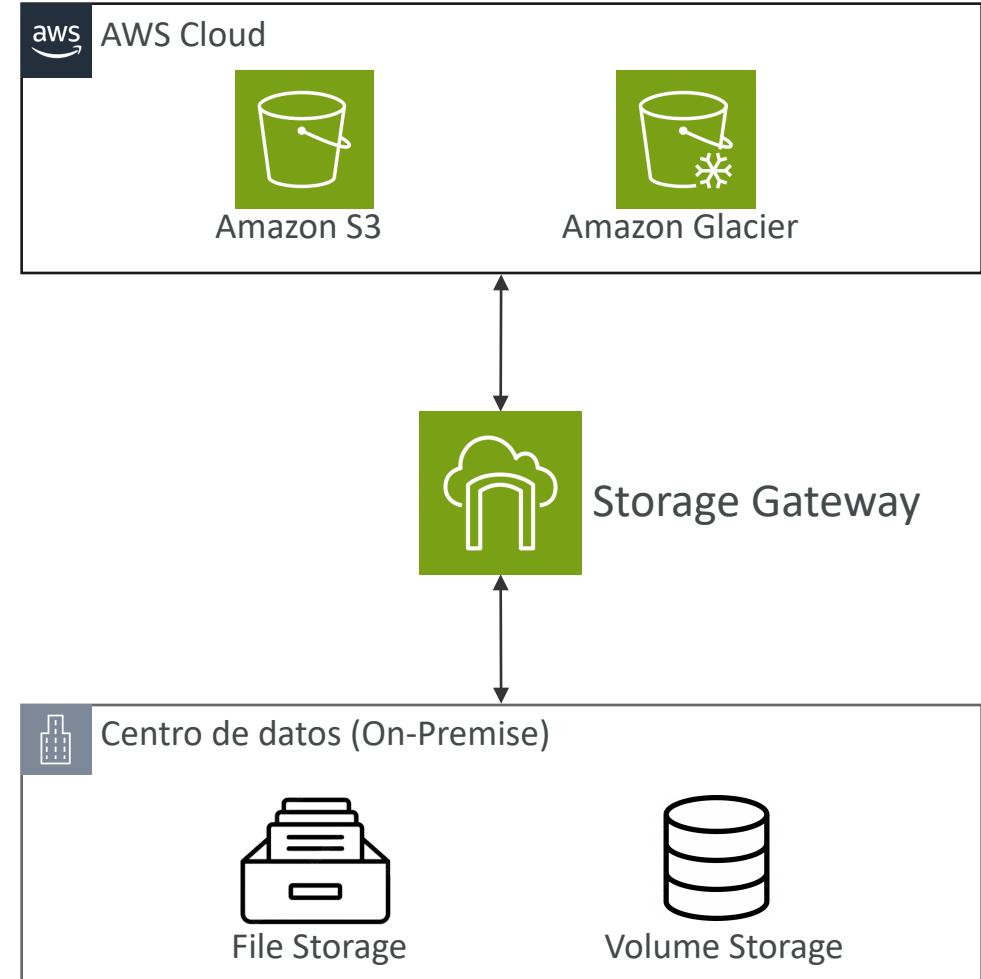
- **Puente entre los datos locales y los del Cloud**

- Tipos de Storage Gateway:

- **S3 File Gateway**
- **FSx File Gateway**
- **Volume Gateway**
- **Tape Gateway**

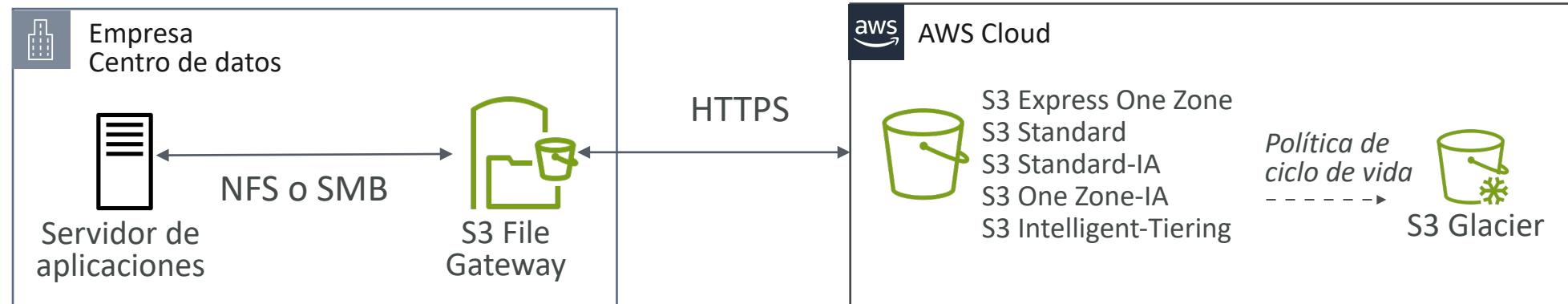
- **Casos de uso:**

- Recuperación de desastres
- Copia de seguridad y restauración
- Almacenamiento por niveles
- Caché local y acceso a archivos de baja latencia



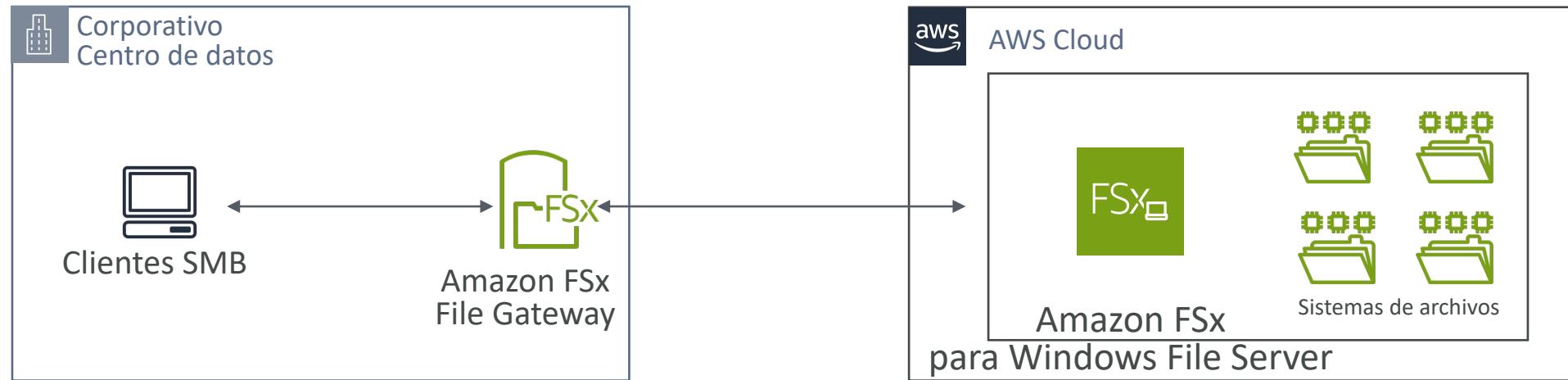
Amazon S3 File Gateway

- 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**
- **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



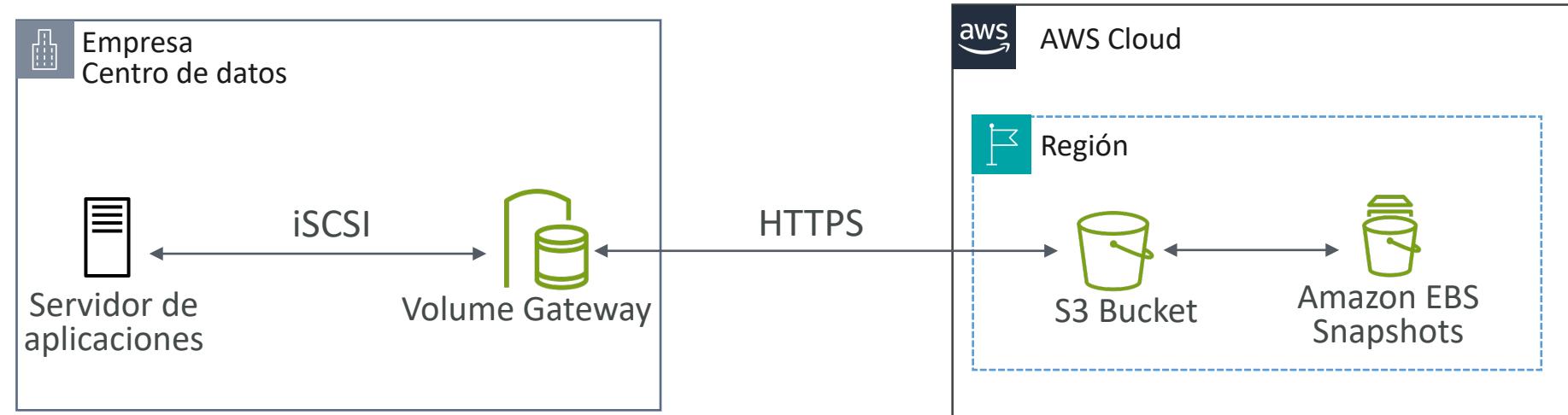
Amazon FSx File Gateway

- 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



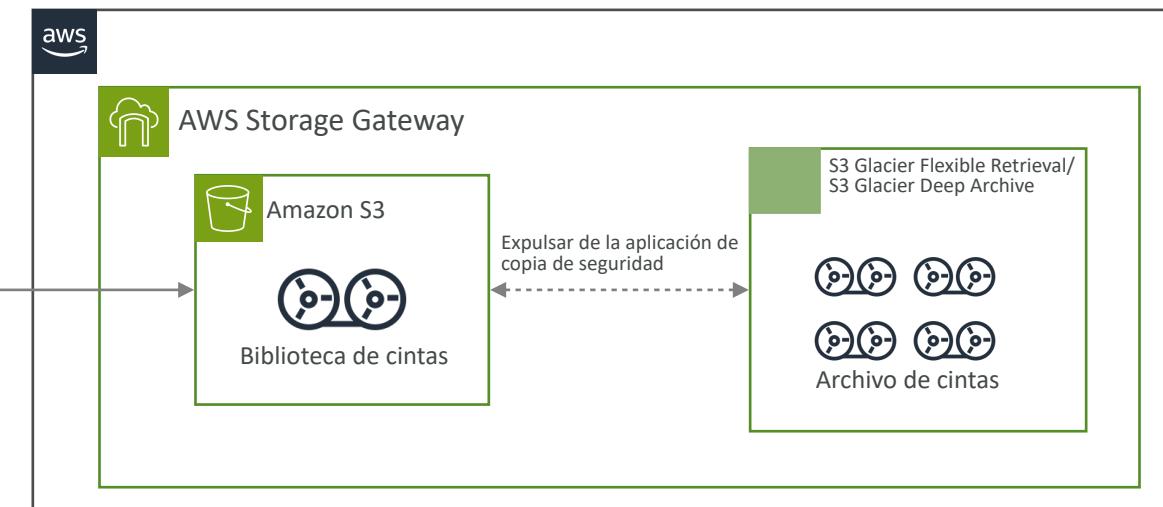
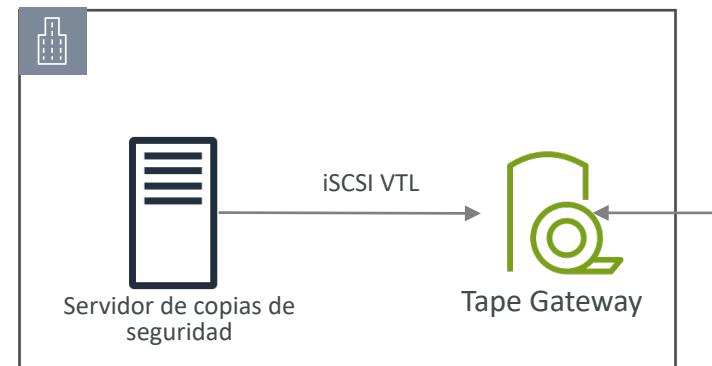
Volume Gateway (Puerta de enlace de volumen)

- Almacenamiento en bloque respaldado por Amazon S3
- Respaldado por Snapshots de EBS que pueden ayudar a restaurar los volúmenes locales
- Volume Gateway dispone de dos modos de funcionamiento:
 - **Modo caché**: acceso de baja latencia a los datos más recientes
 - **Modo almacenado**: todo el conjunto de datos está en las instalaciones, copias de seguridad programadas en S3



Tape Gateway

- 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
- Realiza copias de seguridad de los datos utilizando los procesos existentes basados en cintas
- Funciona con los principales proveedores de software de copia de seguridad



Dispositivo AWS Storage Gateway

- Utilizar AWS Storage Gateway significa que necesitas virtualización en sitio
- Puedes comprar un **dispositivo de hardware Storage Gateway** en amazon.com
- Tiene los recursos necesarios de CPU, memoria, red y caché SSD

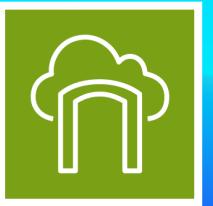
AWS Storage Gateway pre-loaded on a Dell EMC PowerEdge server

by Dell

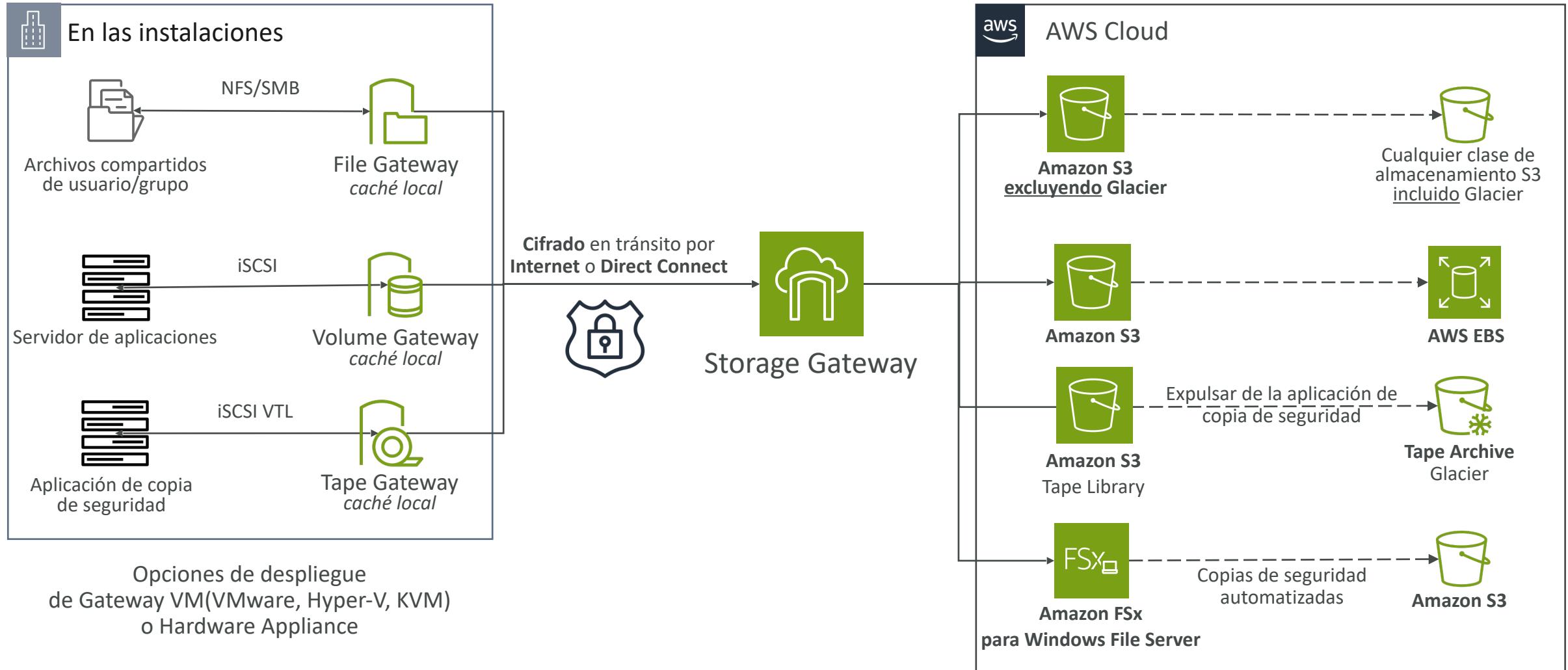
[Be the first to review this item](#)

Price: \$12,250.50

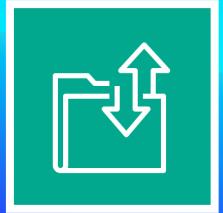




AWS Storage Gateway



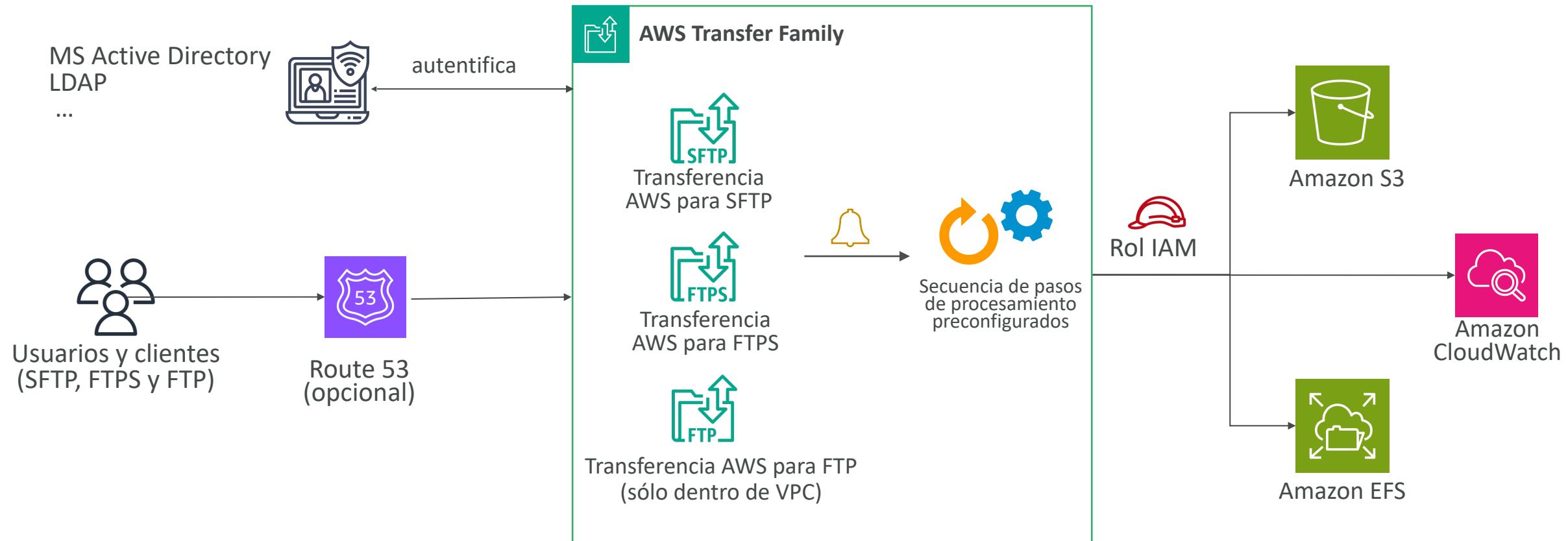
AWS Transfer Family



- Un servicio totalmente gestionado para la transferencia de archivos hacia y desde Amazon S3 o Amazon EFS
- Escala de forma segura las transferencias de archivos recurrentes de empresa a empresa a los servicios de almacenamiento de AWS mediante los protocolos **SFTP, FTPS, FTP y AS2**
- Infraestructura administrada, escalable, fiable, altamente disponible (multi-AZ)
- Paga por endpoint aprovisionado por hora + transferencias de datos en GB
- Casos de uso: compartir archivos, conjuntos de datos públicos, CRM, ...



AWS Transfer Family



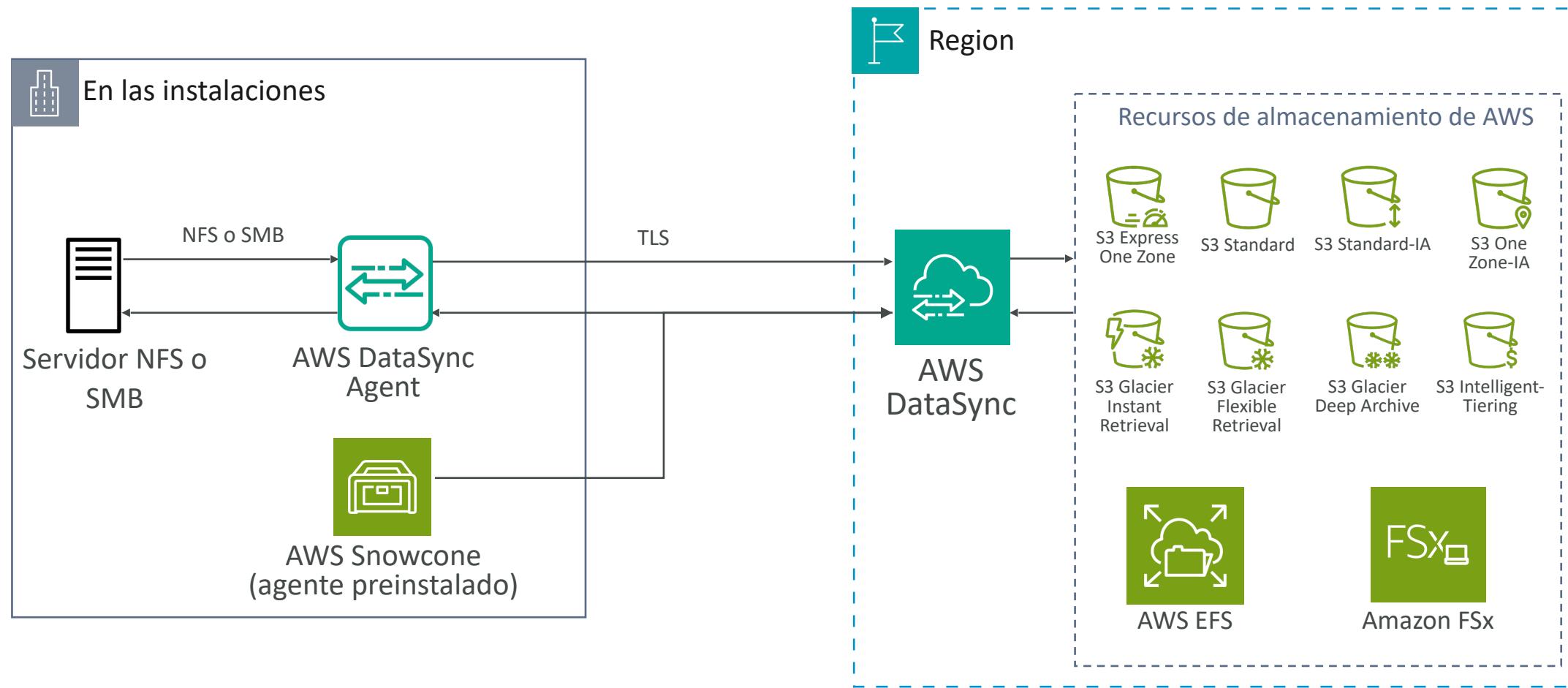
AWS DataSync



- 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...)

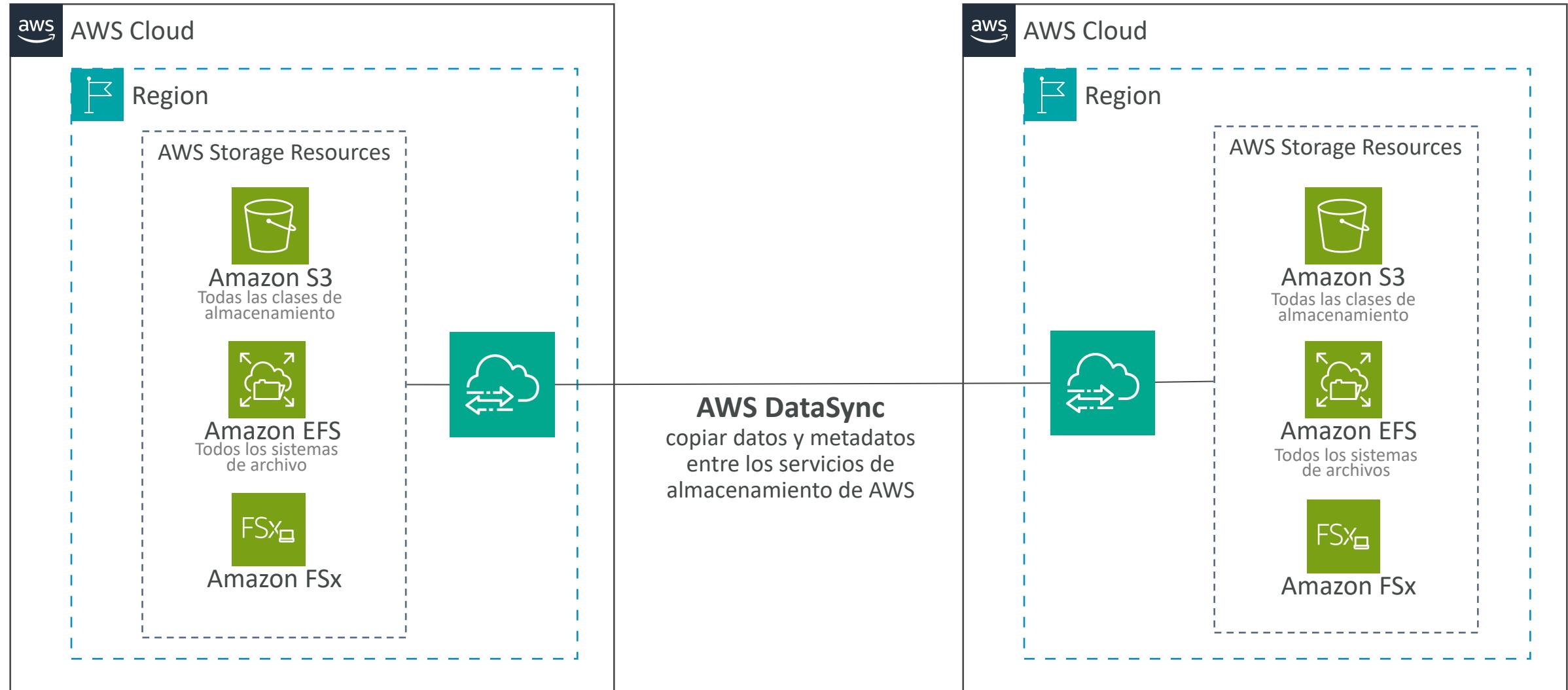
AWS DataSync

Transferencia entre instalaciones locales y AWS



AWS DataSync

Transferencia entre servicios de almacenamiento de AWS



Comparación de opciones 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 y FSx File Gateway, Volume Gateway (caché y almacenado), Tape Gateway
- **Familia de transferencia:** Interfaz FTP, FTPS, SFTP sobre Amazon S3 o Amazon EFS
- **DataSync** 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

Introducción a las bases de datos de AWS

Bases de datos relacionales vs. no relacionales

Bases de datos relacionales (SQL)

- Una base de datos relacional (o base de datos SQL) almacena los datos en formato tabular con filas y columnas
- Las columnas contienen atributos de datos, mientras que en las filas hay valores de dato

Bases de datos no relacionales (NoSQL)

- Utilizan diversos modelos de datos para acceder a estos y administrarlos
- Están optimizadas específicamente para aplicaciones que requieren grandes volúmenes de datos, baja latencia y modelos de datos flexibles

<https://aws.amazon.com/es/compare/the-difference-between-relational-and-non-relational-databases/>

Bases de datos relacionales vs. no relacionales

RELACIONALES (SQL)



Amazon
RDS



Amazon
Aurora

NO RELACIONALES (NoSQL)



Amazon
DynamoDB



Amazon
DocumentDB



Amazon
Timestream



Amazon
Keyspaces

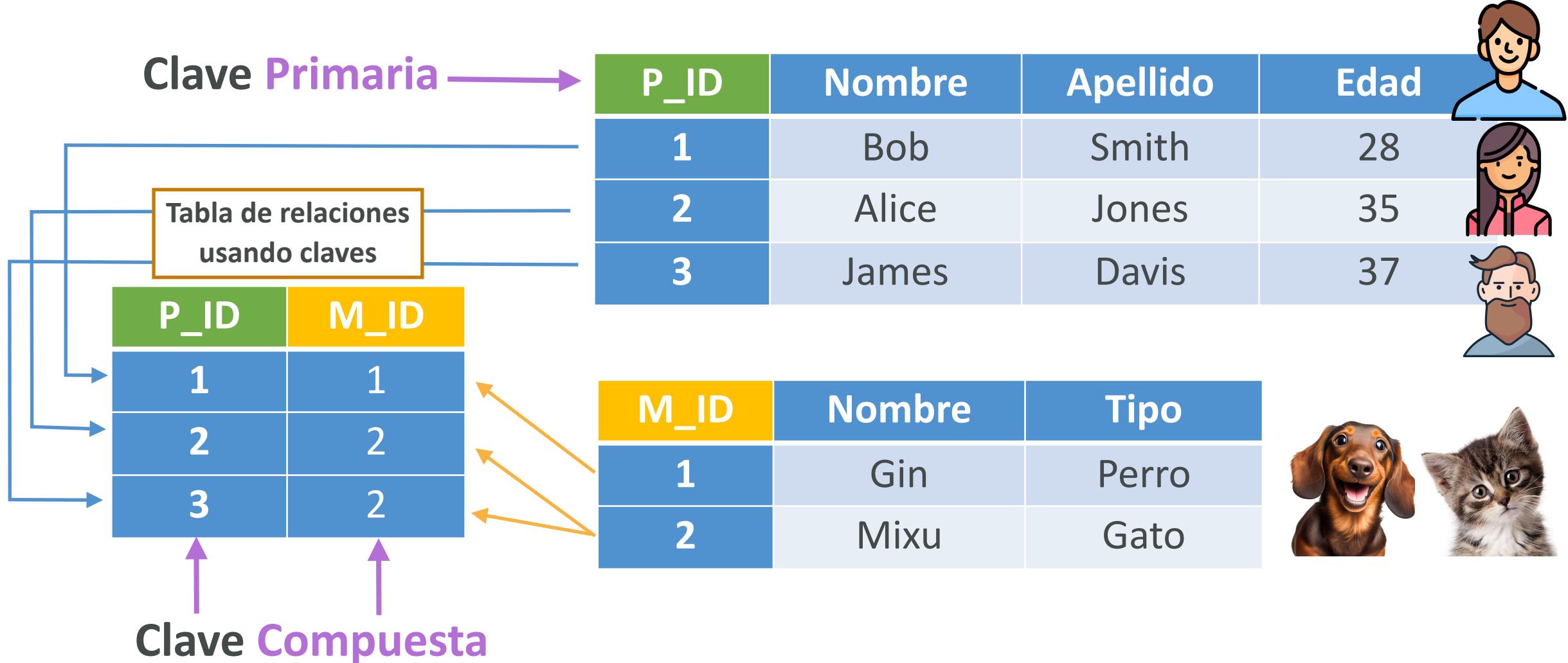


Amazon
Neptune



Amazon
ElastiCache

Bases de datos Relacionales (SQL)



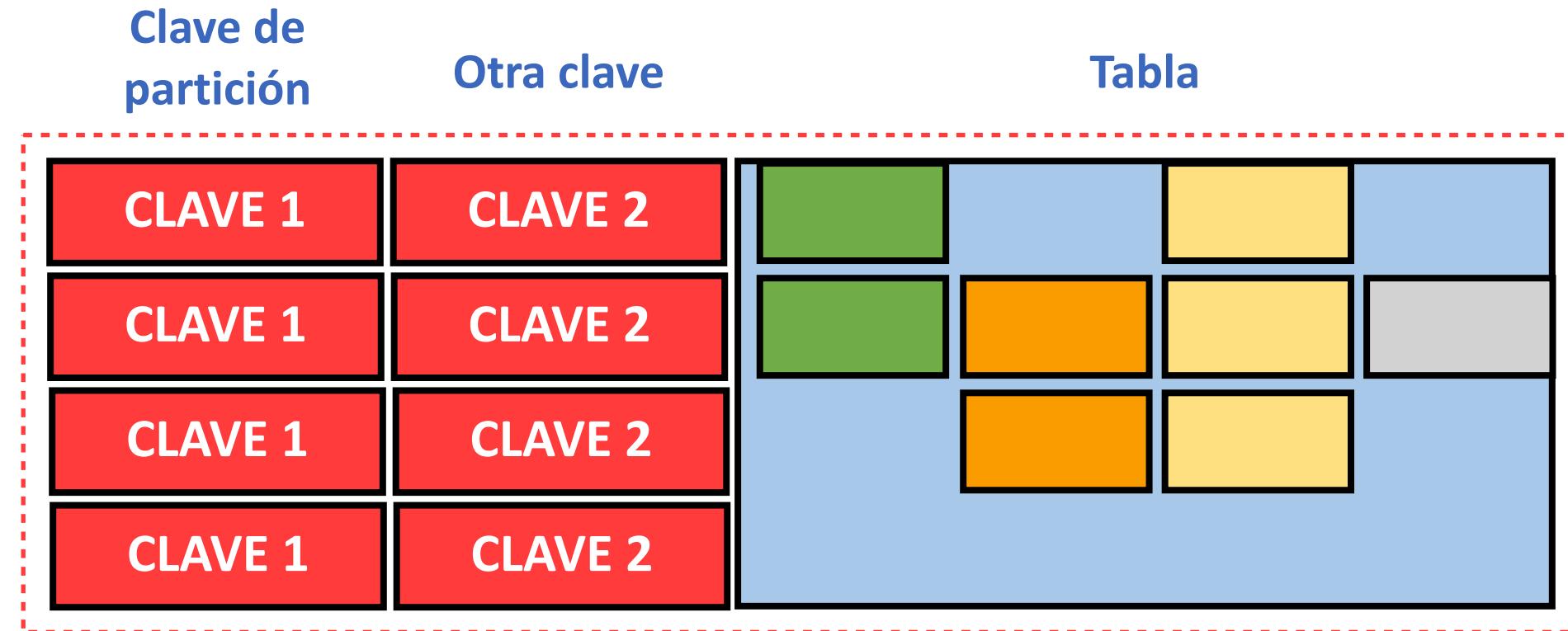
Bases de datos No Relacionales (NoSQL)

Clave - Valor

| | Clave | Valor |
|---------------|------------------|----------|
| No esquema | 2025-03-18 13:00 | 15 |
| No estructura | 2025-03-18 14:00 | 30 |
| Escalable | | |
| Muy rápido | 2025-03-18 15:00 | 0 |
| | Tiempo | Personas |

Bases de datos No Relacionales (NoSQL)

Almacenamiento de columna amplia



Bases de datos No Relacionales (NoSQL)

Documento

```
<publicacion id="p5001">
  <autorId>u10001</autorId>
  <texto>Mi primera publicación</texto>
  <fecha>2024-01-11</fecha>
  <likes>150</likes>
</publicacion>
```

Perfil de usuario en un sistema de redes sociales

```
{
  "_id": "u10001",
  "nombre": "Juan Pérez",
  "email": "juanperez@example.com",
  "publicacionFavorita": "p5001"
}
```

Publicación en el sistema de redes sociales

Son ideales para aplicaciones que requieren **esquemas flexibles** y
escalabilidad horizontal

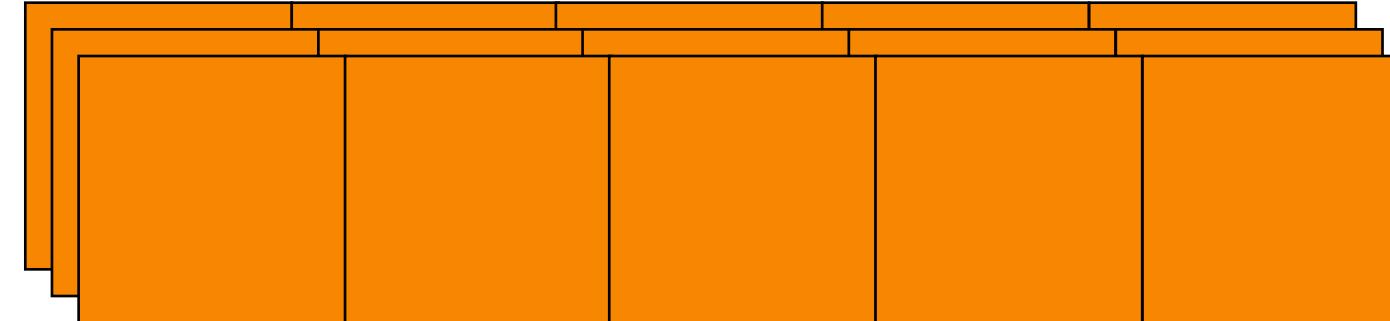
Bases de datos No Relacionales (NoSQL)

Filas (Almacén de filas - MySQL)

FILA

| ID | Producto | Color | Tamaño | Precio |
|----|----------|-------|--------|--------|
| | | | | |

Almacenado
como
filas

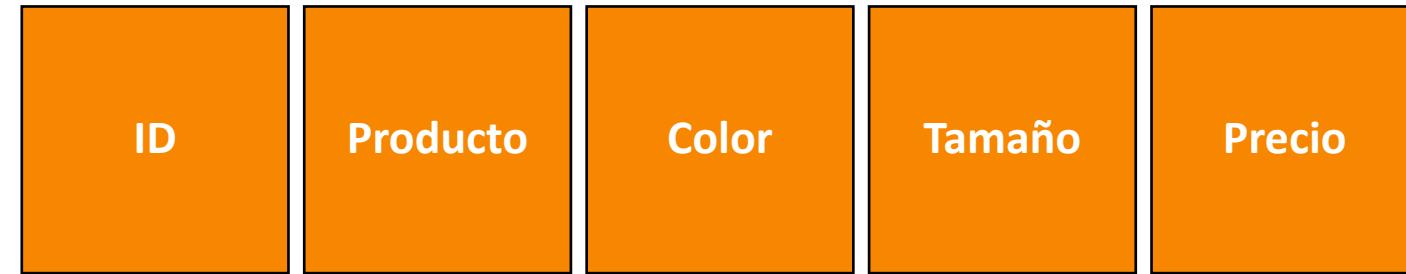


Ideal si operas con filas añadiendo, actualizando, borrando

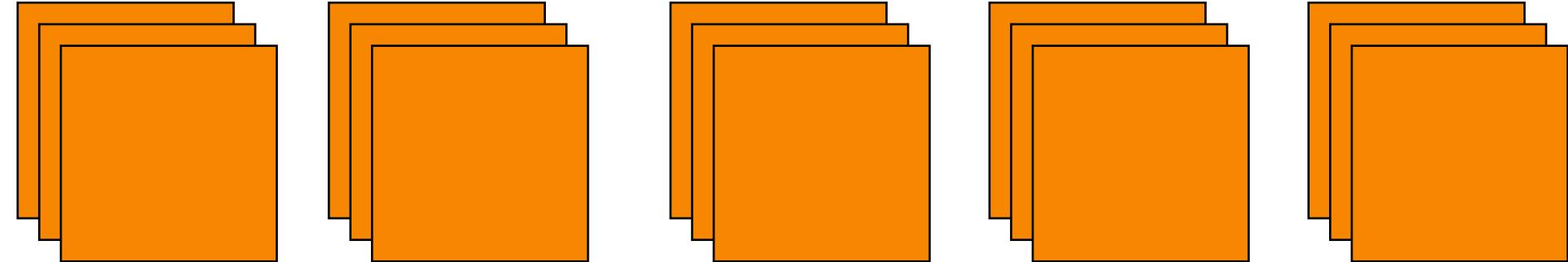
Bases de datos No Relacionales (NoSQL)

Columnas (Almacén de columnas - Redshift)

FILA



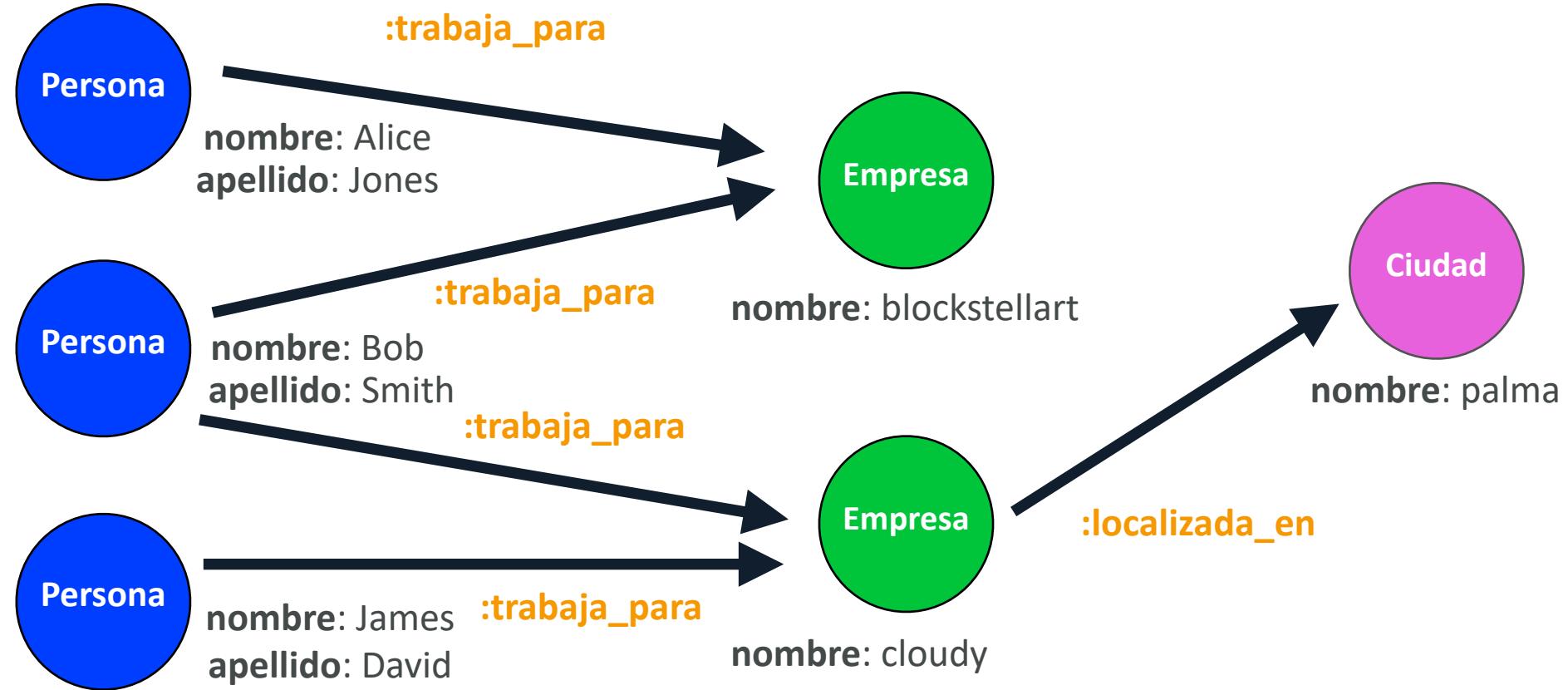
**Almacenado
como
columnas**



Las columnas se almacenan juntas. Ideal cuando se necesitan todos los valores de un atributo concreto (tamaño)

Bases de datos No Relacionales (NoSQL)

Grafo





Amazon RDS

www.blockstellart.com

Todos los derechos reservados © BLOCKSTELLART www.blockstellart.com

Visión general de Amazon RDS



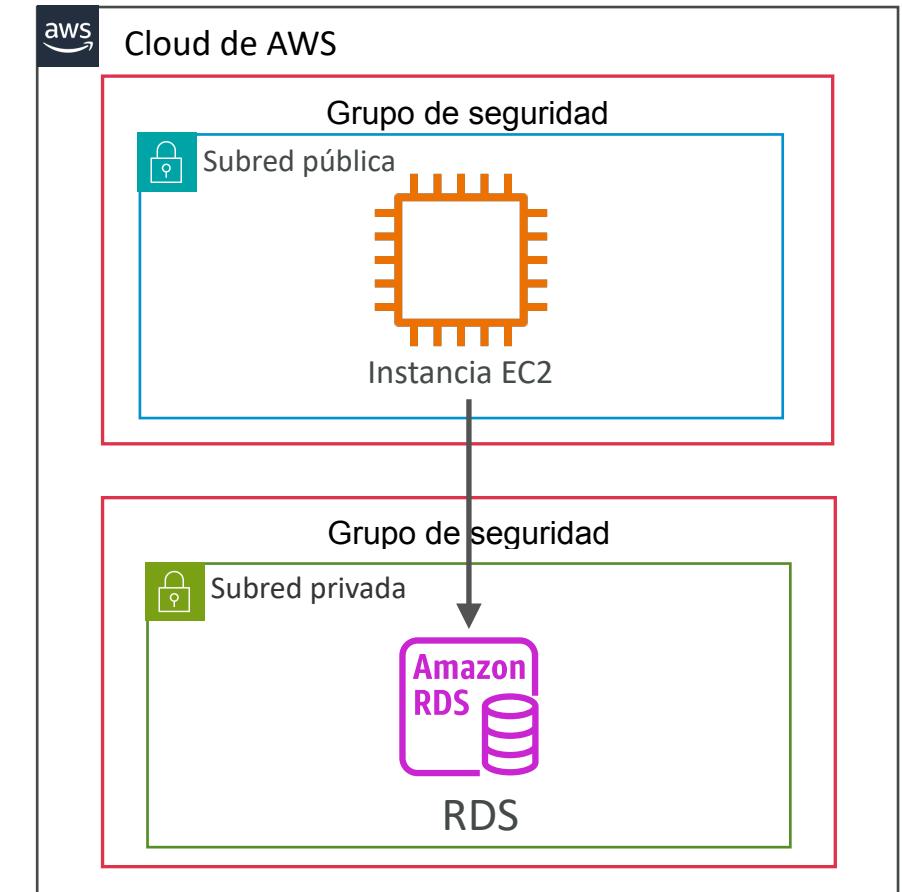
- **RDS = Servicio de bases de datos relacionales**
- Es un servicio de bases de datos gestionado para que las bases de datos utilicen SQL como lenguaje de consulta
- Amazon RDS admite actualmente los siguientes motores:
 - MariaDB
 - Microsoft SQL Server
 - MySQL
 - Oracle
 - PostgreSQL

¿RDS o base de datos en EC2?

- **RDS es un servicio gestionado. Lo que significa...**

- Aprovisionamiento automatizado
- Parcheo del sistema operativo
- Copias de seguridad continuas y restauración a una fecha determinada (Point in Time Restore)
- Dashboards de monitorización
- Réplicas de lectura para mejorar el rendimiento de lectura
- Configuración multi AZ para recuperación de desastres (Disaster Recovery)
- Capacidad de escalado (vertical y horizontal)
- Almacenamiento respaldado por EBS

- **PERO no puedes acceder por SSH a tus instancias**



Autoescalado de almacenamiento de RDS

- Evita **escalar manualmente** el almacenamiento de tu base de datos
- 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 en la base de datos escala automáticamente
- Tienes que establecer el **umbral máximo de almacenamiento** (límite máximo de almacenamiento de la base de datos)
- Útil para aplicaciones con **cargas de trabajo imprevisibles**
- Soporta todos los motores de bases de datos RDS (MariaDB, MySQL, PostgreSQL, SQL Server,...)



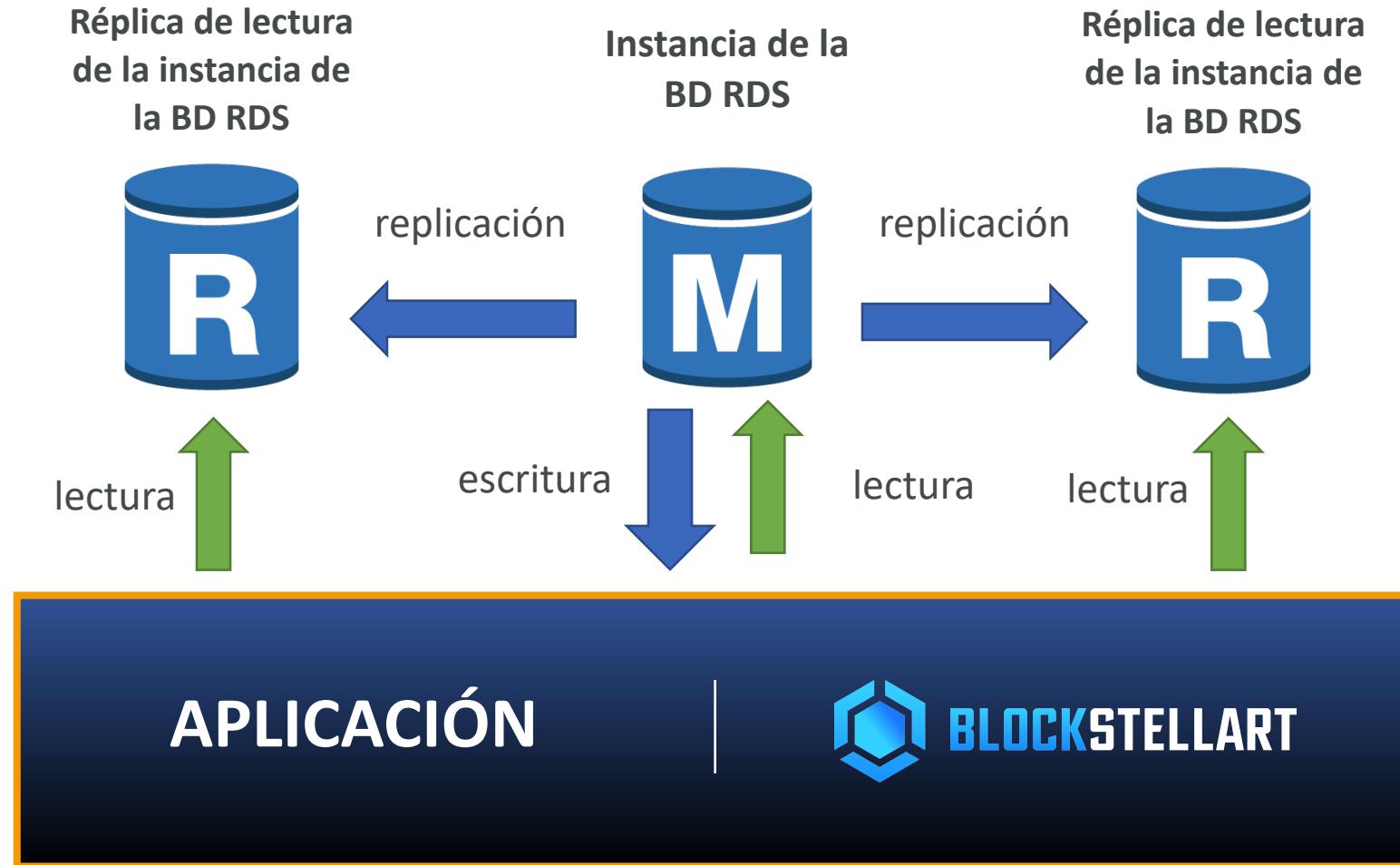
Introducción a las réplicas de Amazon RDS

- Una réplica es una copia de una base de datos principal que se sincroniza continuamente con ella. Las réplicas se usan principalmente para dos propósitos: **escalabilidad de lectura** y **alta disponibilidad**

- **Mejora del rendimiento de lectura:** Escalan eficientemente las operaciones de lectura, distribuyendo la carga entre la instancia principal y las réplicas
- **Alta disponibilidad:** En caso de fallo, una réplica puede convertirse rápidamente en la instancia principal, reduciendo el tiempo de inactividad
- **Desacoplamiento de cargas de trabajo:** Separan las operaciones de lectura y escritura para optimizar el rendimiento
- **Facilidad de implementación:** Integración sencilla con varias bases de datos soportadas por RDS. Las aplicaciones deben actualizar la cadena de conexión para aprovechar las réplicas de lectura



Réplicas de lectura de Amazon RDS

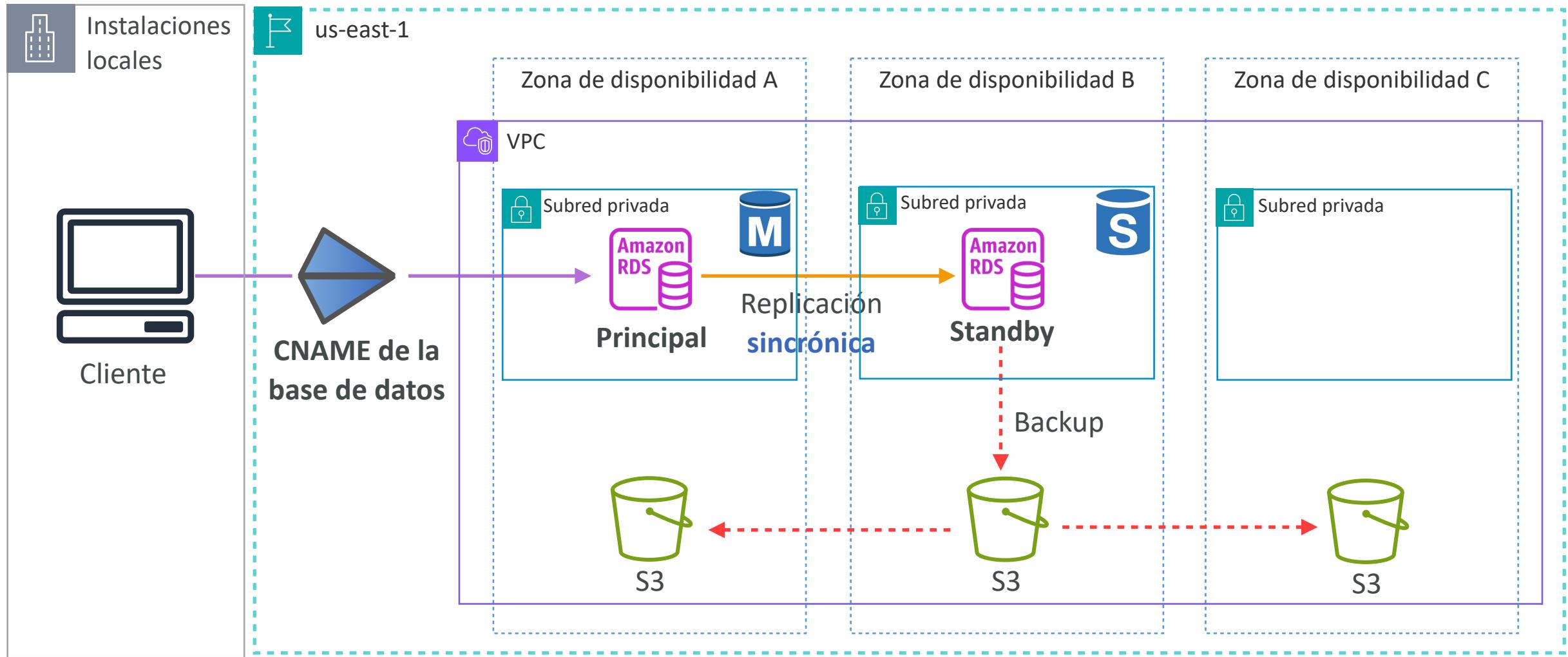


RDS Multi AZ

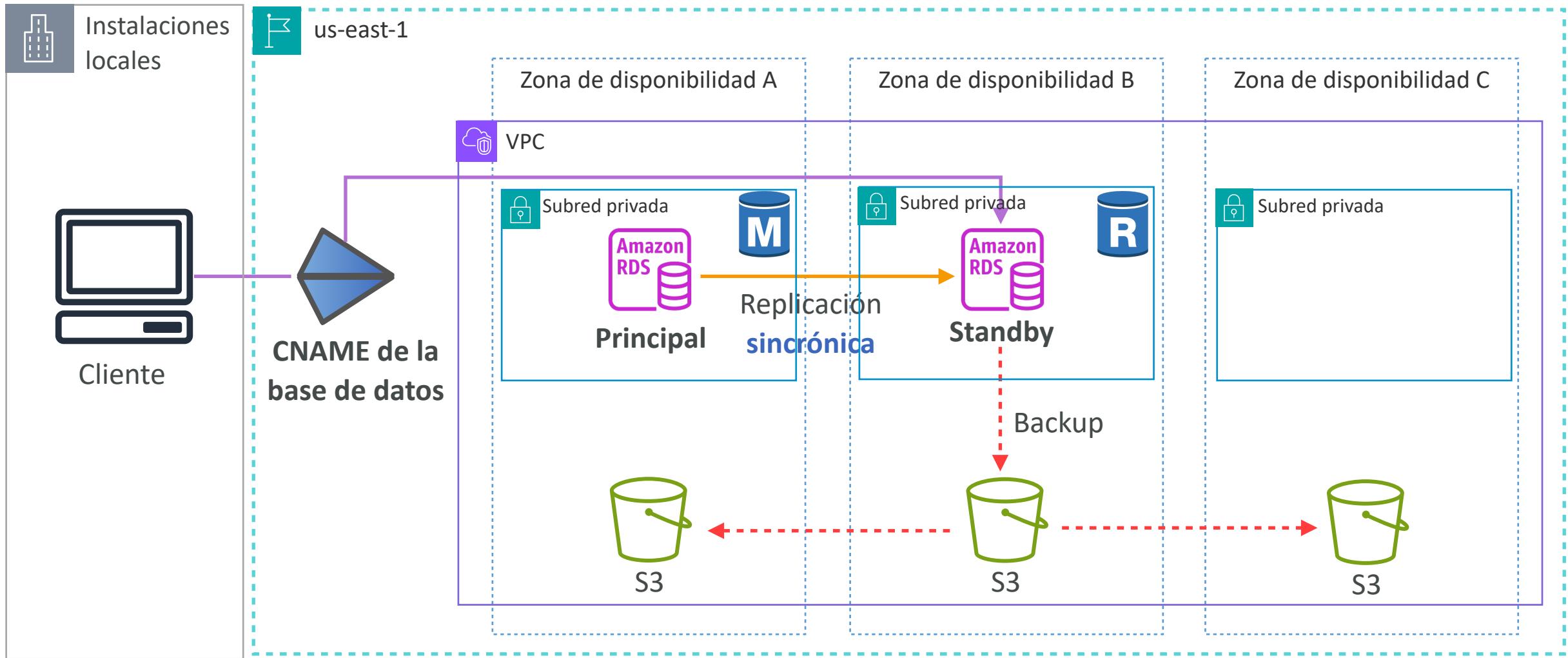
RDS Multi AZ - **Con una instancia en espera**

RDS Multi AZ - **Con dos instancias legibles**

RDS Multi AZ - Con una instancia en espera



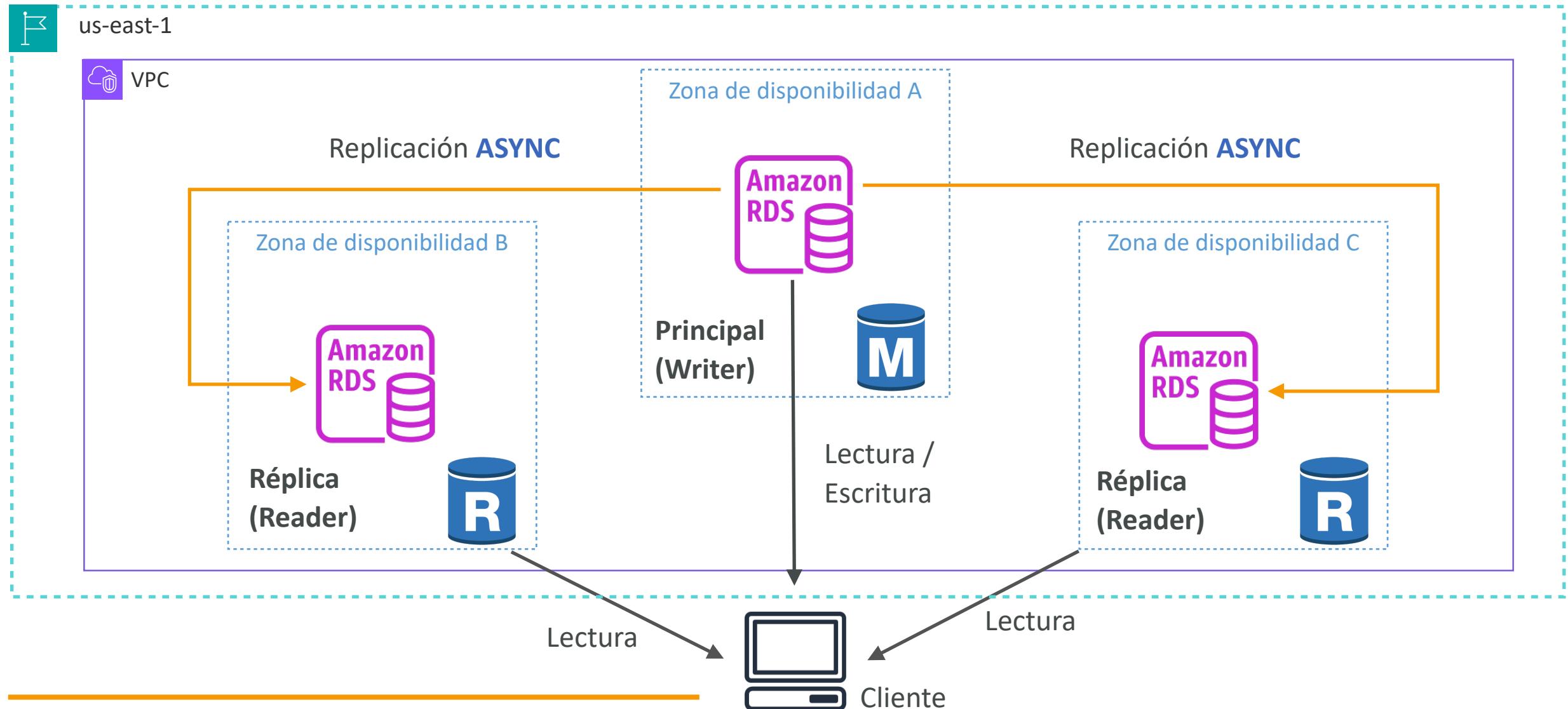
RDS Multi AZ - Con una instancia en espera



RDS Multi AZ - Con una instancia en espera

- Los datos se replican de la instancia primaria a la instancia de reserva (standby) de **forma sincrónica**
- Este servicio **no es parte de la capa gratuita** (hay un costo adicional por la réplica)
- **Solo se permite una réplica de reserva (StandBy)**
- La réplica de reserva no se puede utilizar para lecturas o escrituras
- El tiempo de failover es de 60 a 120 segundos
- Solo se permite **en la misma región**, pero en diferentes zonas de disponibilidad (AZ) dentro de esa región
- Las copias de seguridad (backups) se toman desde la instancia de reserva para mejorar el rendimiento
- Un fallo en la zona de disponibilidad, un fallo primario, un failover manual, cambios en el tipo de instancia y parches de software son algunos de los escenarios cubiertos

RDS Multi AZ - Con dos instancias en espera legibles



RDS Multi AZ - Con dos instancias en espera legibles

- Un escritor y dos instancias de base de datos lectoras en distintas zonas de disponibilidad (AZs)
- Escrituras rápidas al almacenamiento local => se transfieren a EBS
- Los lectores pueden ser utilizados para lecturas, permitiendo cierta escalabilidad de lectura
- La replicación se realiza a través de registros de transacciones, lo que resulta más eficiente
- El failover es más rápido, aproximadamente 35 segundos más la aplicación del registro de transacción
- Las escrituras se "confirman" cuando un lector ha confirmado

Coste de la red (misma región)



Región (A)



VPC

Misma región: **Gratis**

Zona de disponibilidad B



Réplica
(Reader)



Zona de disponibilidad A



Principal
(Writer)



Lectura /
Escritura

Misma región: **Gratis**

Zona de disponibilidad C



Réplica
(Reader)



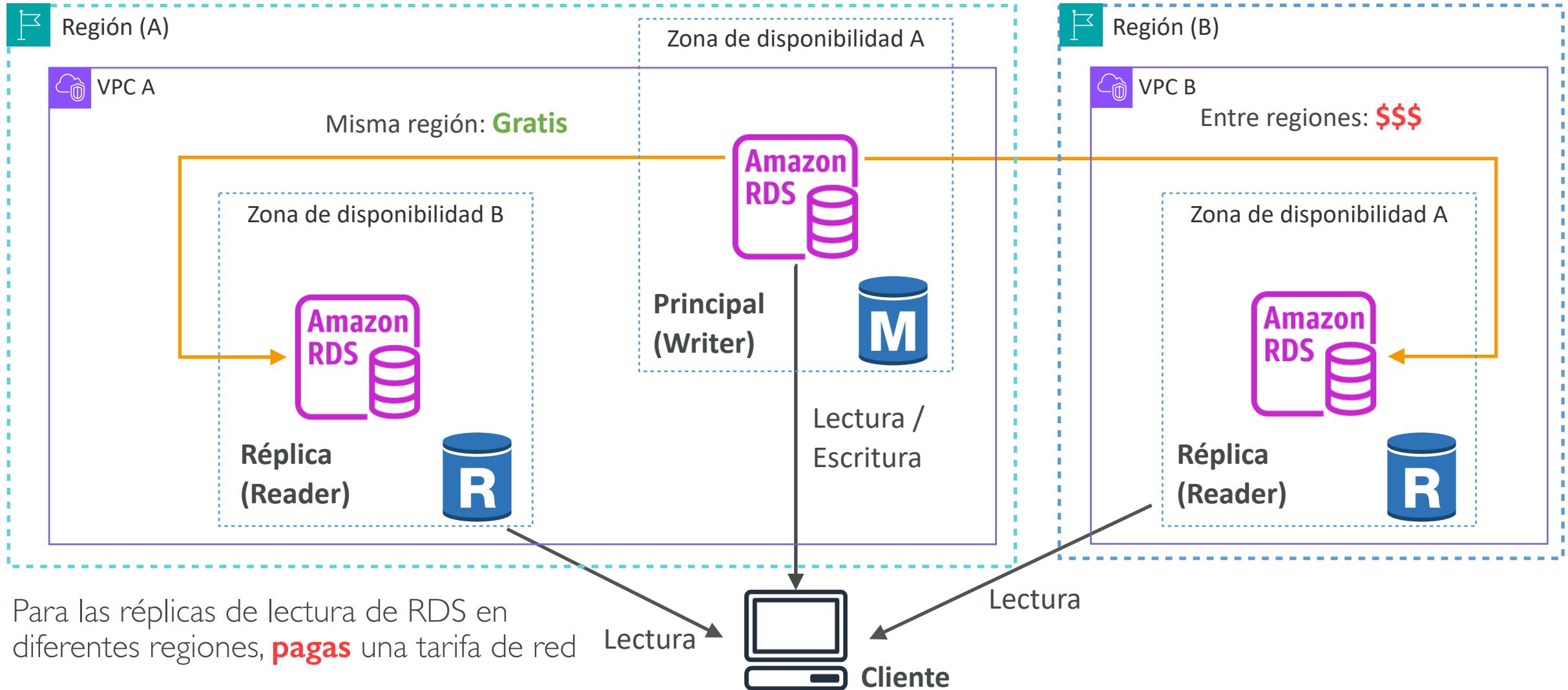
Lectura

Lectura

Cliente

Para las réplicas de lectura de RDS dentro de la misma región, **no pagas** tarifa de red

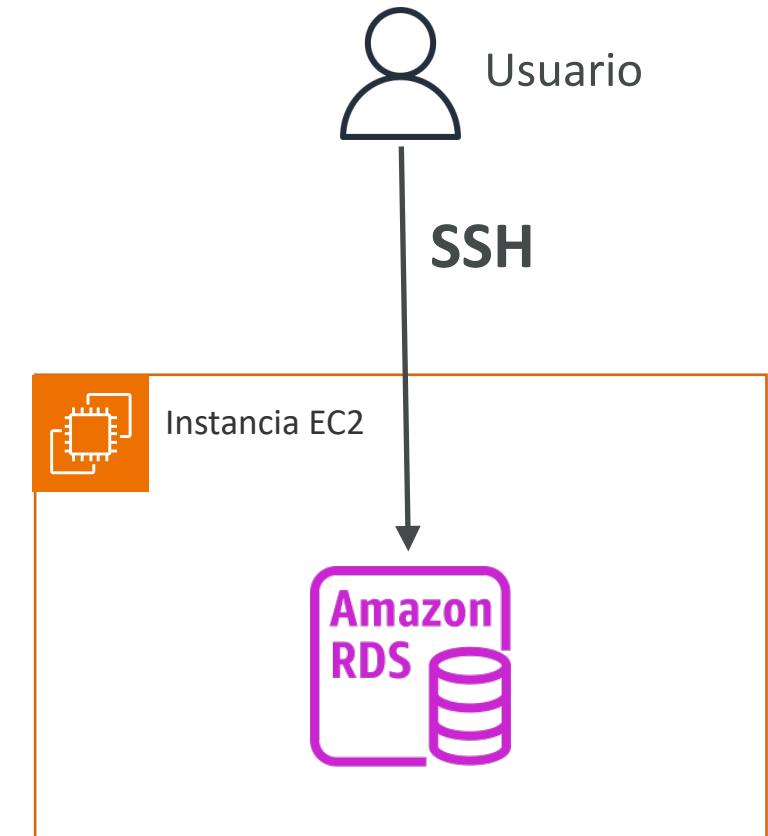
Coste de la red (diferentes regiones)



RDS Personalizado / Custom RDS

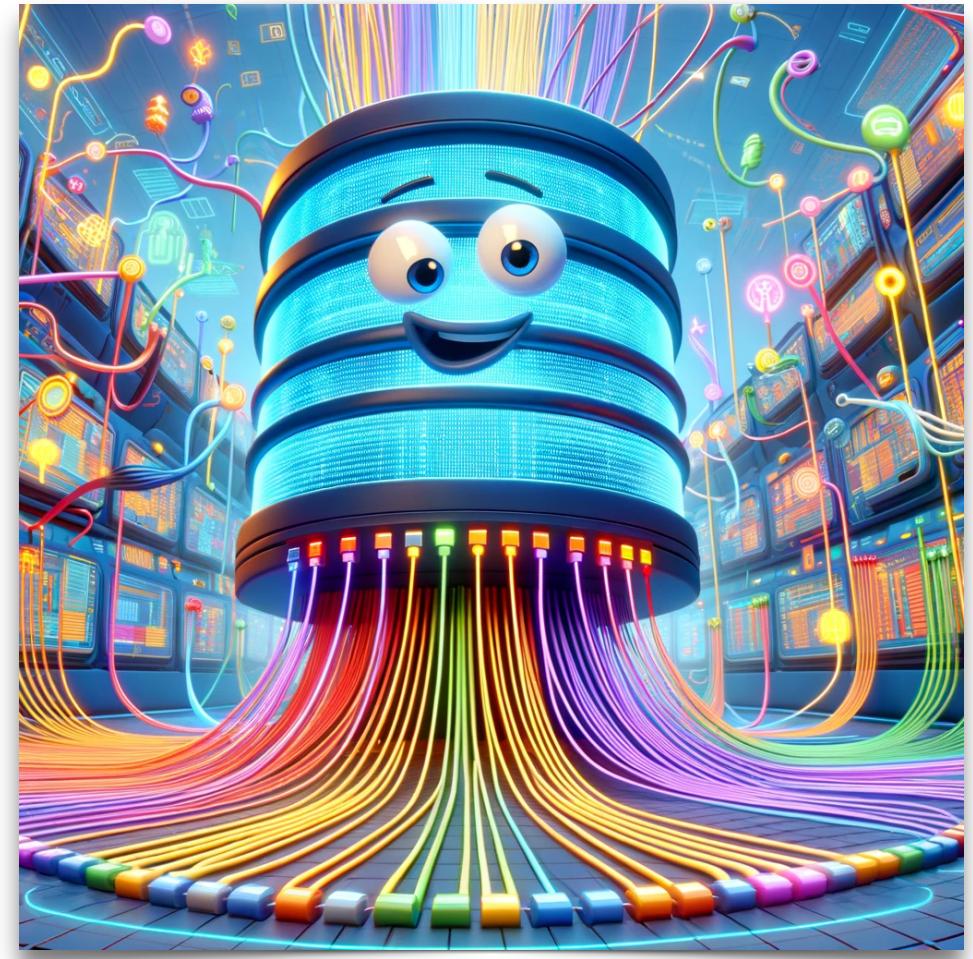
- **Amazon RDS** = Si necesitas que toda la base de datos y el sistema operativo estén completamente administrados por AWS
 - **Amazon RDS Custom** = Si necesitas derechos administrativos sobre la base de datos y el sistema operativo subyacente para que las aplicaciones dependientes estén disponibles
-

- Base de datos gestionada de **Oracle y Microsoft SQL Server** con personalización del sistema operativo y de la base de datos
- 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 Custom:
 - RDS: toda la base de datos y el SO serán gestionados por AWS
 - RDS Custom: acceso administrativo completo al SO subyacente y a la base de datos

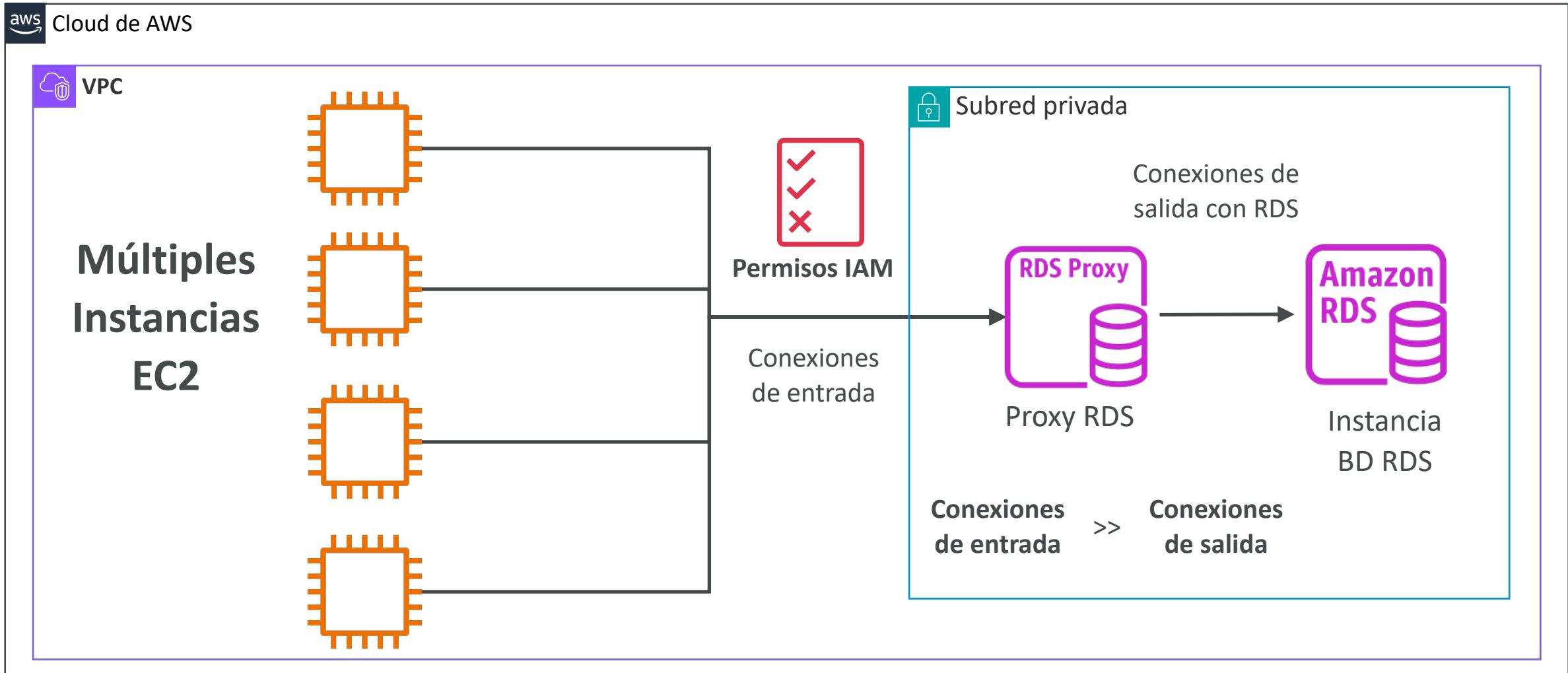


Proxy de Amazon RDS

- 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)**
- Reduce el tiempo de conmutación por error de RDS y Aurora hasta en un 66%
- No se requieren cambios de código para la mayoría de las aplicaciones
- **Aplica los permisos IAM para la base de datos**
- **El proxy RDS nunca es accesible al público (debe accederse desde la VPC)**
- **Soporta RDS (MySQL, PostgreSQL, MariaDB) y Aurora (MySQL, PostgreSQL)**



Proxy de Amazon RDS



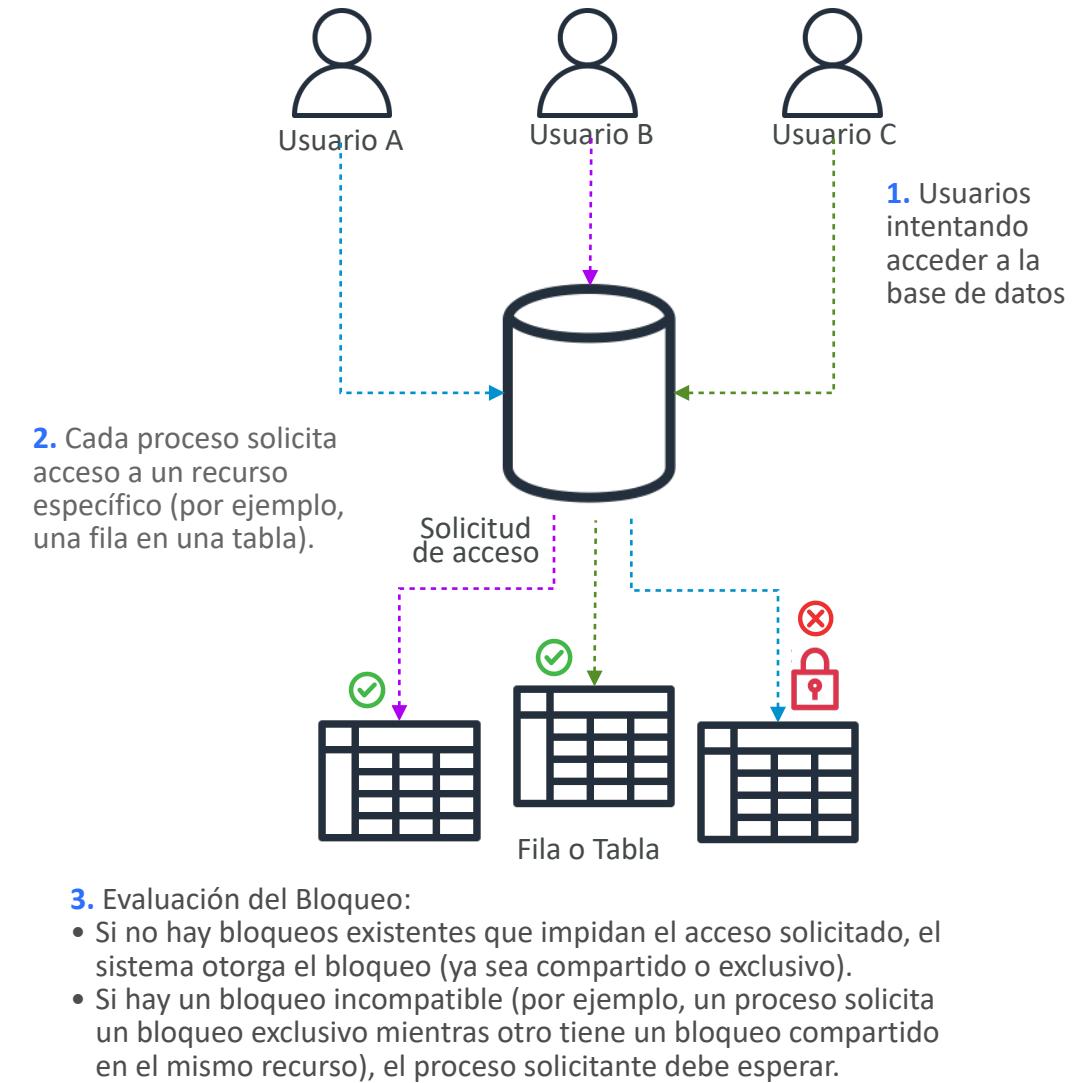
Uso del comando LOCK

  Las bases de datos relacionales implícitamente **"bloquean"** tablas o filas para evitar que escriban en ellas al mismo tiempo, o que se lea mientras un proceso de escritura está en curso

Tipos de bloqueos:

1. **Bloqueos compartidos (S LOCK)**: Permite a múltiples usuarios leer un elemento bloqueado, pero impide que cualquiera lo modifique mientras el bloqueo esté activo. Pueden ser mantenidos por múltiples transacciones. (*FOR SHARE*)

2. **Bloqueos exclusivos (X LOCK)**: Impide que cualquier otro usuario o proceso lea o escriba en el elemento bloqueado hasta que el bloqueo sea liberado. Solo una transacción puede mantener un bloqueo exclusivo. (*FOR UPDATE*)



Ejemplos - MySQL

Bloquear una tabla completa:



`LOCK TABLES employees WRITE;`

Bloquea la tabla completa 'employees' para operaciones de escritura

Para Liberar el bloqueo



`UNLOCK TABLES`

Nota: Redshift también tiene un comando LOCK para el mismo propósito.

Bloqueo compartido (permite lecturas, previene otras escrituras durante esta transacción):



`SELECT * FROM employees WHERE department = 'Finance' FOR SHARE;`

Bloqueo exclusivo (previene todas las lecturas y escrituras durante esta transacción):



`SELECT * FROM employees WHERE employee_id = 123 FOR UPDATE;`

Asegúrate de que las transacciones y los bloqueos estén completos, o podrías terminar con un "bloqueo mutuo" (deadlock)

Directrices operativas de Amazon RDS



Monitoreo

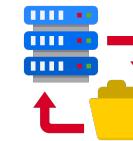
Uso CloudWatch para monitorear memoria, CPU, almacenamiento, y latencia de réplica.



Performance

Una I/O insuficiente hará que la recuperación después de un fallo sea lenta.

1. Migra a una instancia de DB con más I/O.
2. Mueve a General Purpose o IOPS provisionadas.



Backups

Realiza copias de seguridad automáticas durante los momentos de baja escritura en IOPS.



Multi-AZ & Conmutación por error

1. Establece el TTL en DNS para tus instancias de DB a 30 segundos o menos desde tus aplicaciones.
2. Prueba de conmutación por error antes de que lo necesites.



Gestión de recursos

Provee suficiente RAM para incluir todo tu conjunto de trabajo.

1. Si tu métrica de ReadIOPS es pequeña y estable, estás bien.



Protección de datos

Los límites de tasa en Amazon API Gateway pueden usarse para proteger tu base de datos.

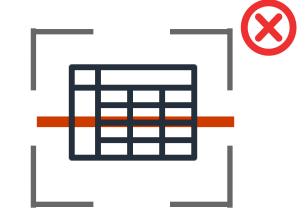
Optimización de consultas en RDS

Utiliza índices para acelerar las sentencias **SELECT**

- Utiliza planes **EXPLAIN** para identificar los índices que necesitas.



Evita los escaneos completos de tablas



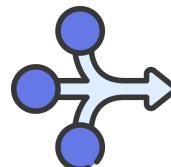
Utiliza **ANALYZETABLE** periódicamente



Optimizaciones específicas del motor



Simplifica las cláusulas **WHERE**



Ajustes específicos para bases de datos



MariaDB



Mantén las tablas bien por debajo de 16TB, idealmente menos de 100GB.



Asegúrate de tener suficiente RAM para sostener los índices de las tablas usadas activamente.



Intenta tener menos de 10,000 tablas.



Usa InnoDB como motor de almacenamiento.



Al cargar datos, desactiva las copias de seguridad de la base de datos y Multi-AZ. Ajusta varios parámetros de la base de datos como `maintenance_work_mem`, `max_wal_size`, `checkpoint_timeout`. Desactiva `synchronous_commit`, activa `autovacuum` y asegura que las tablas estén registradas.



Utiliza autovacuum.

Ajustes específicos para bases de datos



Utiliza los Eventos de BD de RDS para monitorear failovers.



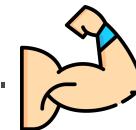
No habilites el modo de recuperación simple, modo fuera de línea, o modo solo lectura (esto rompe Multi-AZ).



Implementa en todas las Zonas de Disponibilidad (AZ's).

ORACLE

Oracle es una bestia propia.





Amazon Aurora

www.blockstellart.com

Todos los derechos reservados © BLOCKSTELLART www.blockstellart.com

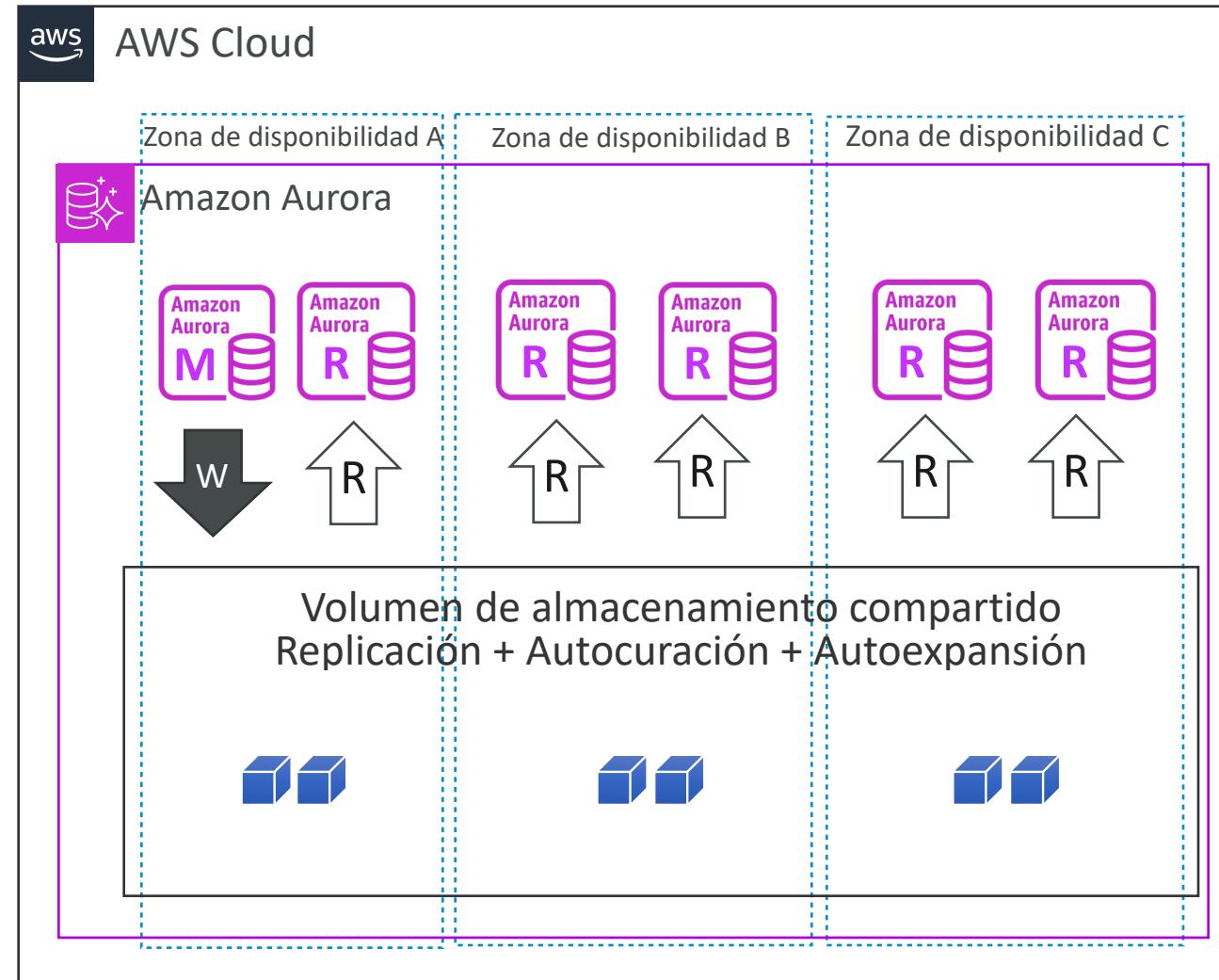


Visión general de Amazon Aurora

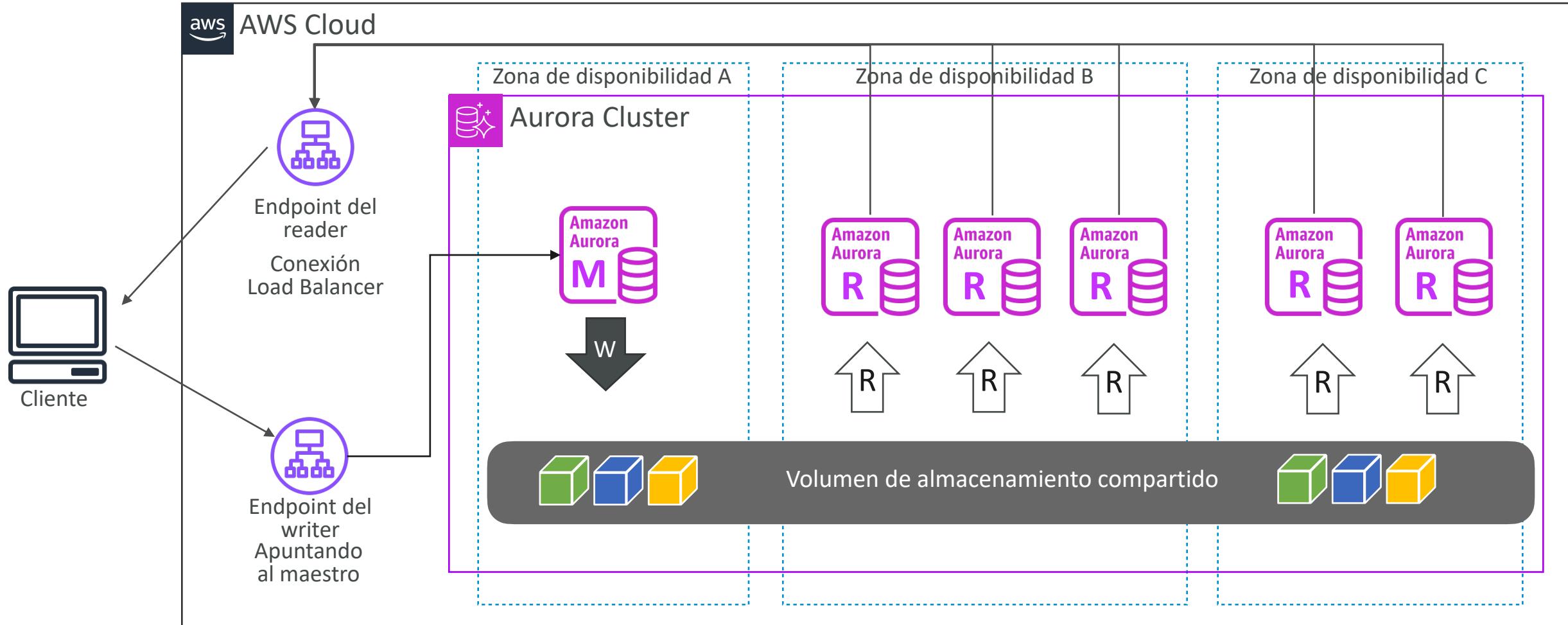
- Aurora es una tecnología propietaria de AWS (no de código abierto)
- Tanto Postgres como MySQL se soportan como base de datos de Aurora (eso significa que tus controladores funcionarán como si Aurora fuera una base de datos Postgres o MySQL)
- Aurora está "optimizada para el Cloud de AWS" y afirma que mejora 5 veces el rendimiento de MySQL en RDS, y más de 3 veces el rendimiento de Postgres en RDS
- El almacenamiento de Aurora crece automáticamente en incrementos de 10 GB, hasta 128 TB.
- Aurora puede usar réplicas de BD con un proceso de replicación muy rápido (retardo de réplica inferior a 10 ms)
- La conmutación por error en Aurora es instantánea
- Aurora cuesta más que RDS (un 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)
 - Master + Réplicas de lectura
- Recuperación automática del master en menos de 30 segundos



Cluster de BD Aurora

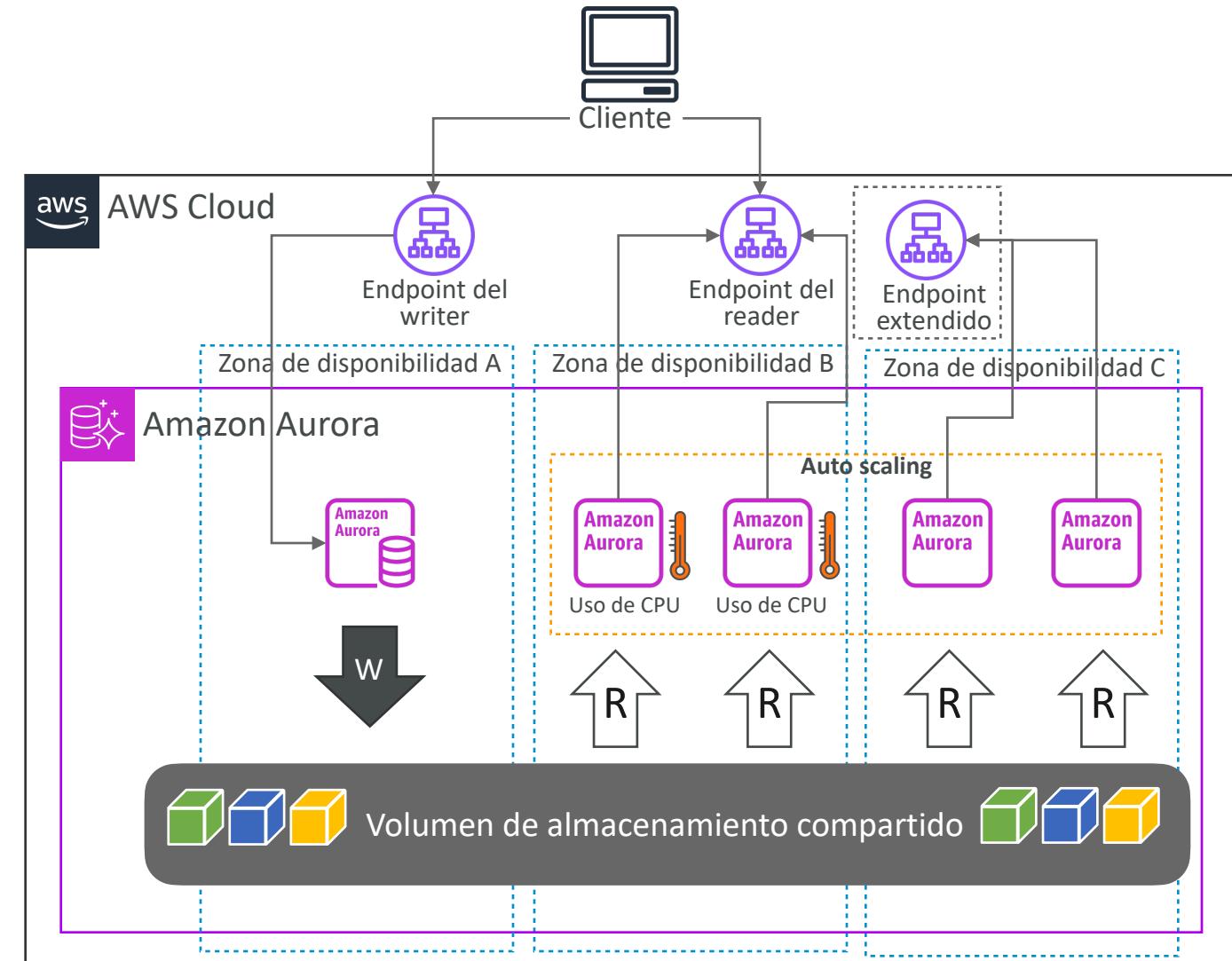


Características de Aurora

- 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
- Restaura los datos en cualquier momento sin usar copias de seguridad

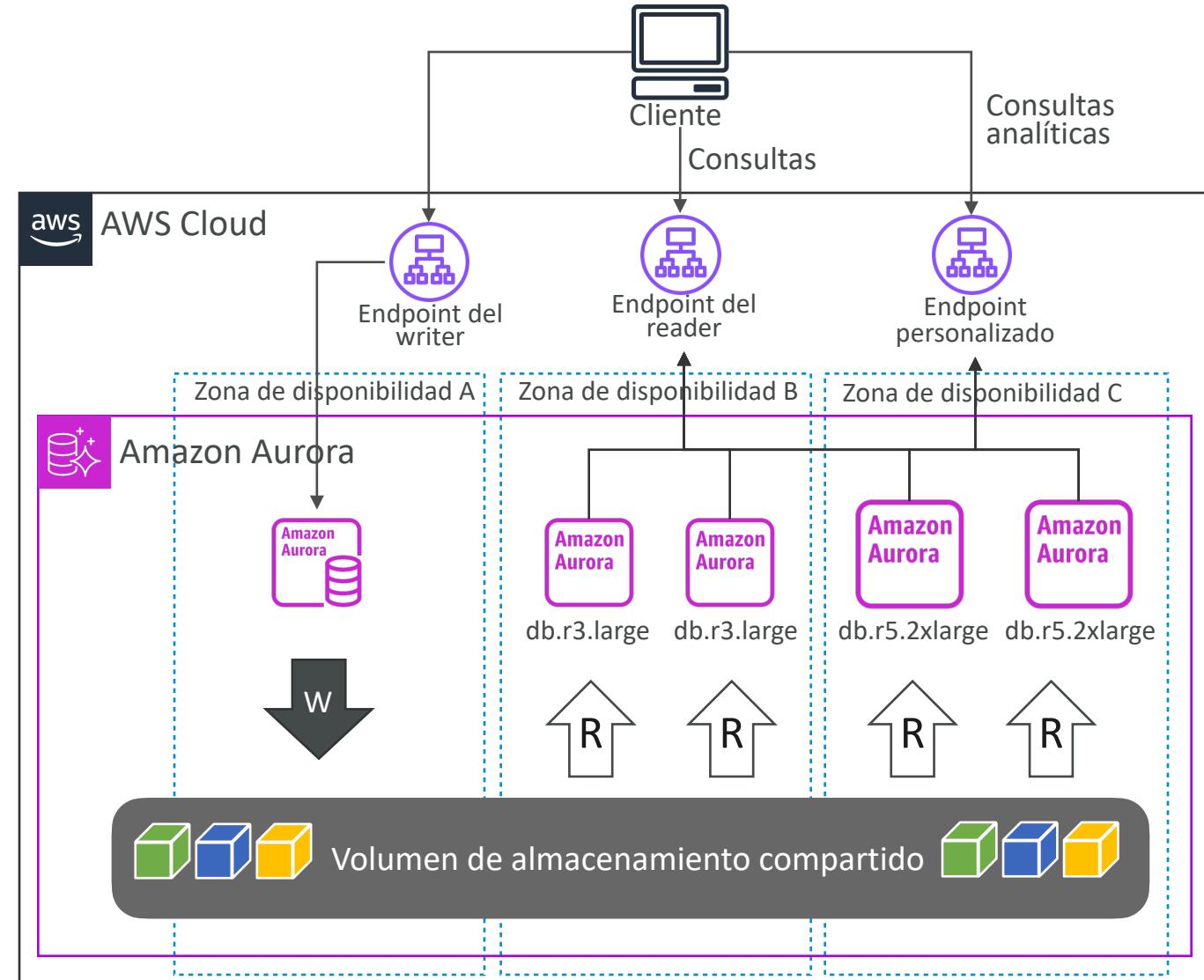
Aurora - Auto Scaling de réplicas

- Ajusta automáticamente el número de réplicas de lectura en respuesta a cambios en la carga de trabajo
- Aurora distribuye automáticamente las solicitudes de lectura entre las réplicas de lectura activas, optimizando el rendimiento y reduciendo la latencia



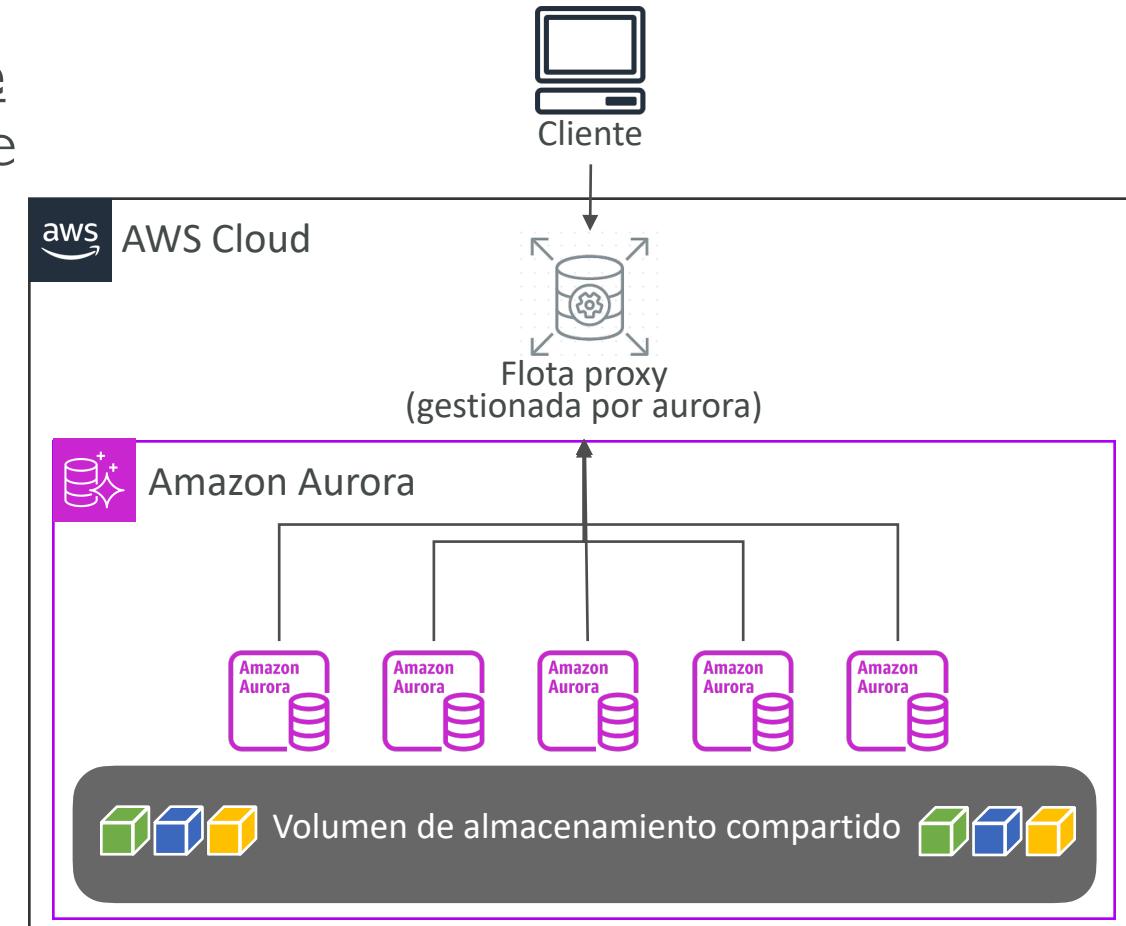
Aurora - Endpoints personalizados

- Permiten configurar puntos de enlace personalizados para acceder a instancias de Aurora, facilitando la alta disponibilidad y la tolerancia a fallos
- Puedes dirigir el tráfico de lectura a réplicas específicas y el tráfico de escritura a la instancia principal, optimizando el rendimiento de la base de datos
- Ofrecen la flexibilidad de crear diferentes puntos de enlace para distintos tipos de operaciones (lectura, escritura, etc.), según las necesidades de la aplicación



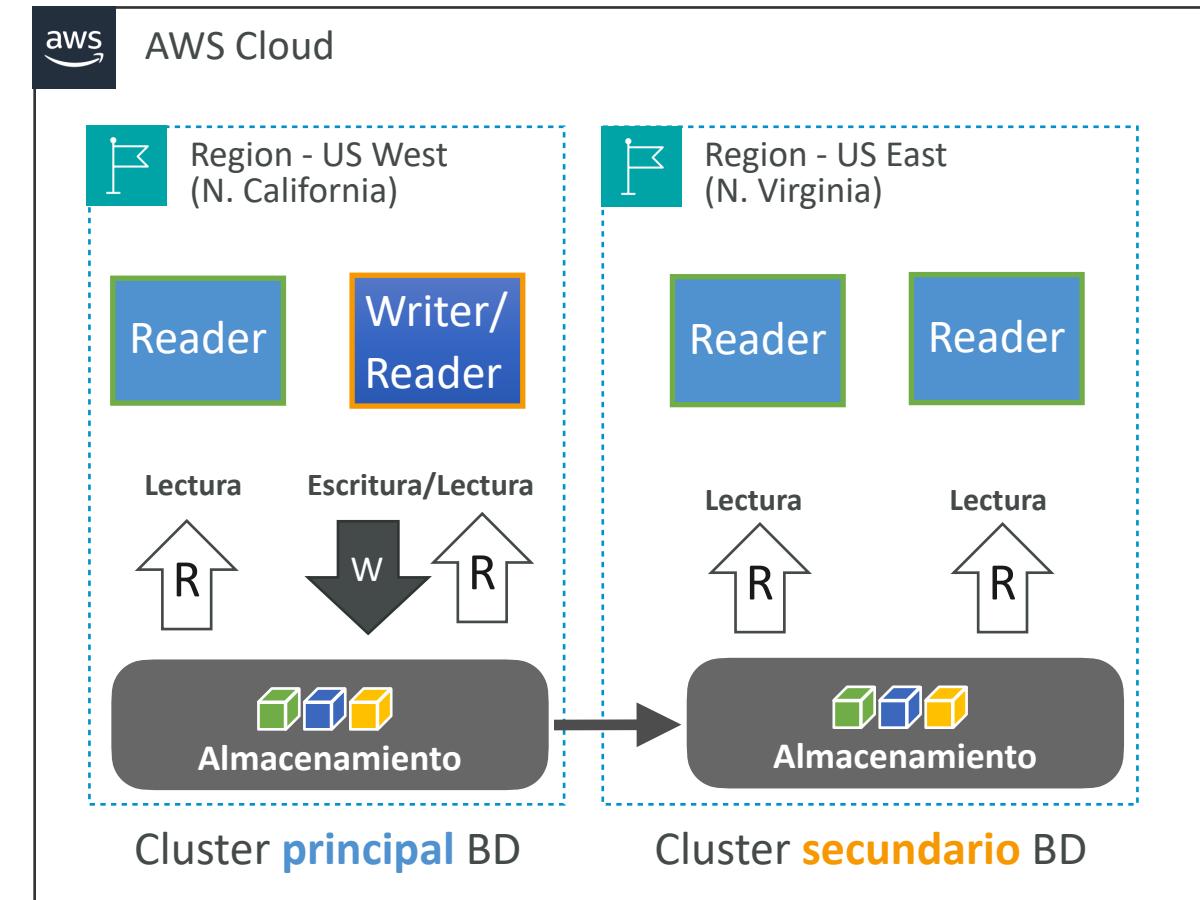
Aurora - Sin servidor (serverless)

- Ajusta automáticamente la capacidad de la base de datos **en función de la demanda de la aplicación**, eliminando la necesidad de gestión manual
- **Sólo pagas por los recursos que realmente utilizas**, lo que puede reducir significativamente los costos en comparación con las instancias tradicionales
- Puede iniciarse, detenerse y escalarse automáticamente **según los patrones de uso** de la aplicación, optimizando la utilización de recursos
- Elimina la necesidad de gestionar la infraestructura subyacente



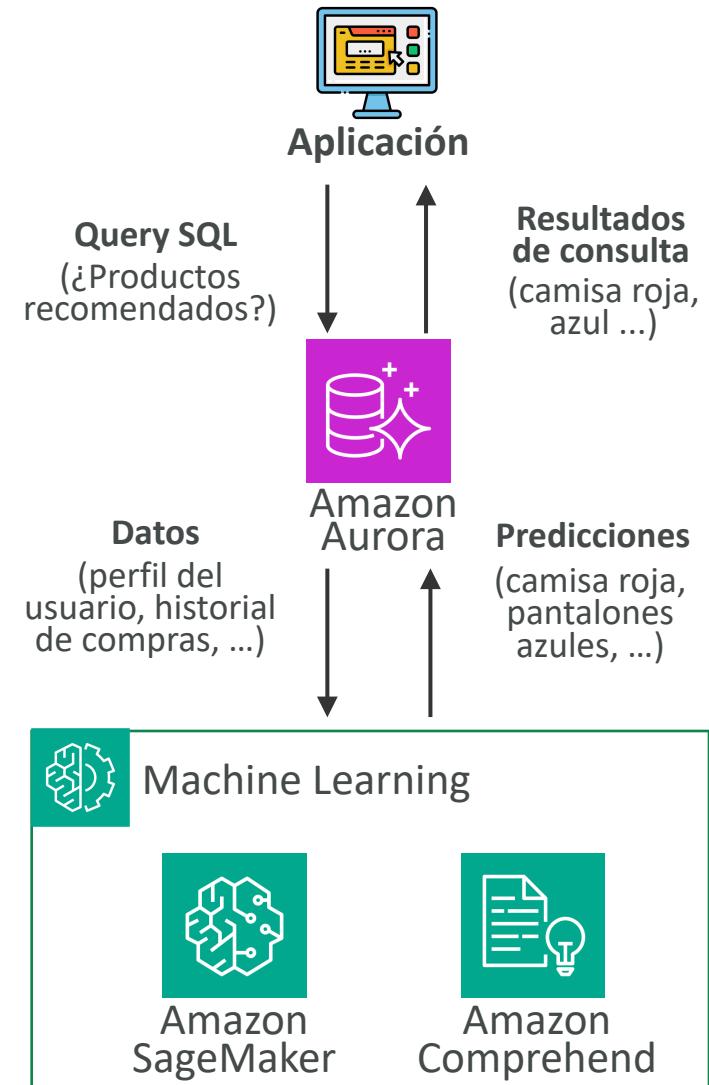
Aurora - Bases de datos globales

- Las bases de datos globales de Amazon Aurora abarcan múltiples regiones de AWS, lo que permite lecturas globales de baja latencia y proporcionan una recuperación rápida de cualquier interrupción
- **Clúster principal** = 1 Región Principal (lectura/escritura)
- **Clústeres secundarios** = Regiones secundarias (solo lectura), la latencia de replicación es de menos de 1 segundo
- Réplicas de lectura por región secundaria
- Ayuda a disminuir la latencia
- Promover otra región (para recuperación ante desastres) tiene un RTO de menos de 1 minuto
- La replicación típica entre regiones toma menos de 1 segundo



Aurora - Machine Learning

- Aurora permite utilizar SQL para agregar predicciones basadas en ML sin necesidad de aprender nuevas herramientas ni tener experiencia previa en ML
- Integración optimizada y segura con los servicios de ML de AWS, como SageMaker, Comprehend y Bedrock
- Ideal para aplicaciones en tiempo real y de baja latencia, como detección de fraudes y recomendaciones de productos
- Gestión segura de datos con cifrado de extremo a extremo y control de acceso mediante AWS IAM
- Casos de uso: Recomendaciones personalizadas de productos ,servicios al cliente mejorados, etc.



Aurora - Copias de seguridad

- Amazon Aurora realiza **copias de seguridad automáticas del cluster de la base de datos** sin impactar el rendimiento de las operaciones de la base de datos
- Permiten la recuperación punto a punto, pudiendo restaurar la base de datos a cualquier segundo dentro del periodo de retención, usualmente **hasta los últimos 35 días**
- Aurora almacena solo los cambios en los bloques de datos desde la última copia de seguridad, optimizando el uso del espacio y reduciendo los costos



RDS y Aurora - Opciones de restauración

- Restaurar una copia de seguridad de RDS / Aurora o una instantánea crea una nueva base de datos
- Restauración de base de datos RDS MySQL desde S3
 - Crea una copia de seguridad de tu base de datos local
 - Guárdala en Amazon S3 (almacenamiento de objetos)
 - Restaura el archivo de copia de seguridad en una nueva instancia de RDS que ejecute MySQL
- Restauración de cluster Aurora MySQL desde S3
 - Crea una copia de seguridad de tu base de datos local usando Percona XtraBackup
 - Almacena el archivo de copia de seguridad en Amazon S3
 - Restaura el archivo de copia de seguridad en un nuevo clúster de Aurora que ejecute MySQL



Aurora - Clonación de base de datos

- **Crea un nuevo clúster de Aurora a partir de uno existente**
- Más rápido que hacer una snapshot y restaurar
- Utiliza el protocolo de copia - escritura
 - Inicialmente, el nuevo clúster de DB utiliza el mismo volumen de datos que el clúster DB original (rápido y eficiente — no se necesita copia)
 - Cuando se realizan actualizaciones en los datos del nuevo clúster DB, se asigna almacenamiento adicional y los datos se copian para separarlos
- Útil para crear una base de datos de "pruebas" a partir de una base de datos de "producción" sin impactar la base de datos de producción

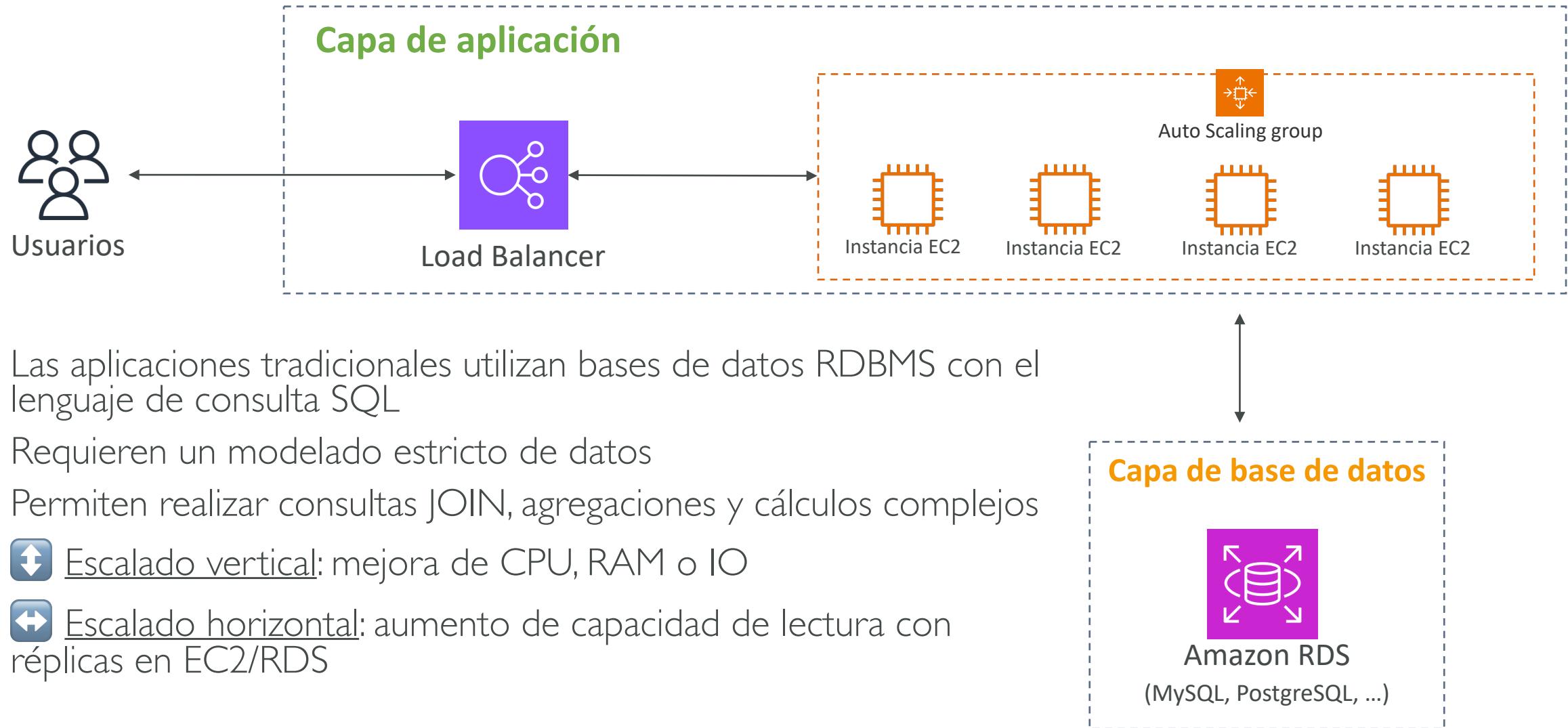


RDS y Aurora - Seguridad

- **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 BD (en lugar de username/password)
- **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

Amazon DynamoDB

Diseño tradicional de las arquitecturas



Bases de datos NoSQL

- **Bases de datos NoSQL**

- No son relacionales y están distribuidas
- Ejemplos: MongoDB, DynamoDB, Cassandra, Redis
- No soportan consultas "JOIN"
- Todos los datos necesarios para una consulta suelen estar en una fila
- No realizan agregaciones como "SUM", "AVG"
- Escalan horizontalmente

- **Comparación con SQL**

- No hay "correcto o incorrecto" entre NoSQL y SQL
- Requieren modelar los datos de forma diferente
- Es necesario pensar en las consultas de los usuarios de manera distinta



DynamoDB



redis

Visión general de DynamoDB



- Totalmente gestionada y altamente disponible con replicación MultiAZ
- Base de datos NoSQL (no relacional)
- Escala para cargas de trabajo masivas
- Soporta millones de solicitudes por segundo, trillones de filas y cientos de TB de almacenamiento
- Rendimiento rápido y consistente
- Integrada con IAM para seguridad, autorización y administración
- Clases de tablas de acceso estándar y acceso poco frecuente (IA)

zoom

Zoom Video Communications, Inc. administró el incremento de 10 millones a 300 millones de participantes diarios en reuniones



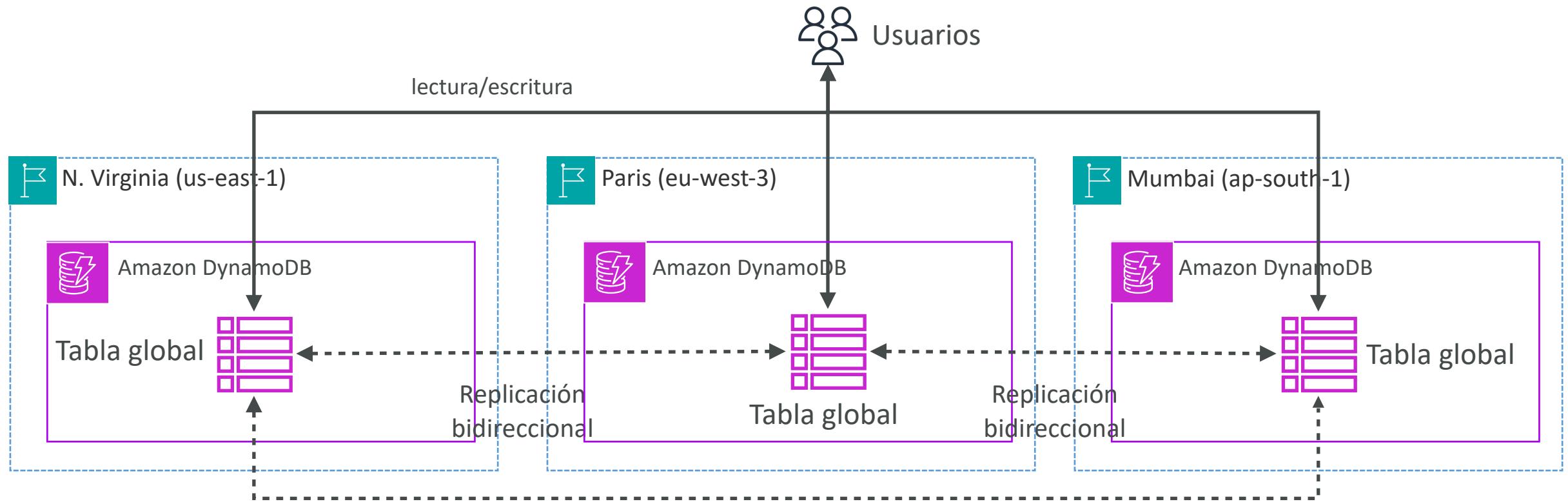
Snap Inc. redujo la mediana de latencia en un 20 % con Amazon DynamoDB



Dropbox se ahorró millones de dólares en costes de expansión con un nuevo sistema de almacenamiento

Tablas globales de DynamoDB

- Haz que una tabla de DynamoDB sea accesible con **baja latencia** en varias regiones
- Replicación **activa-activa (lectura/escritura)** en cualquier región de AWS



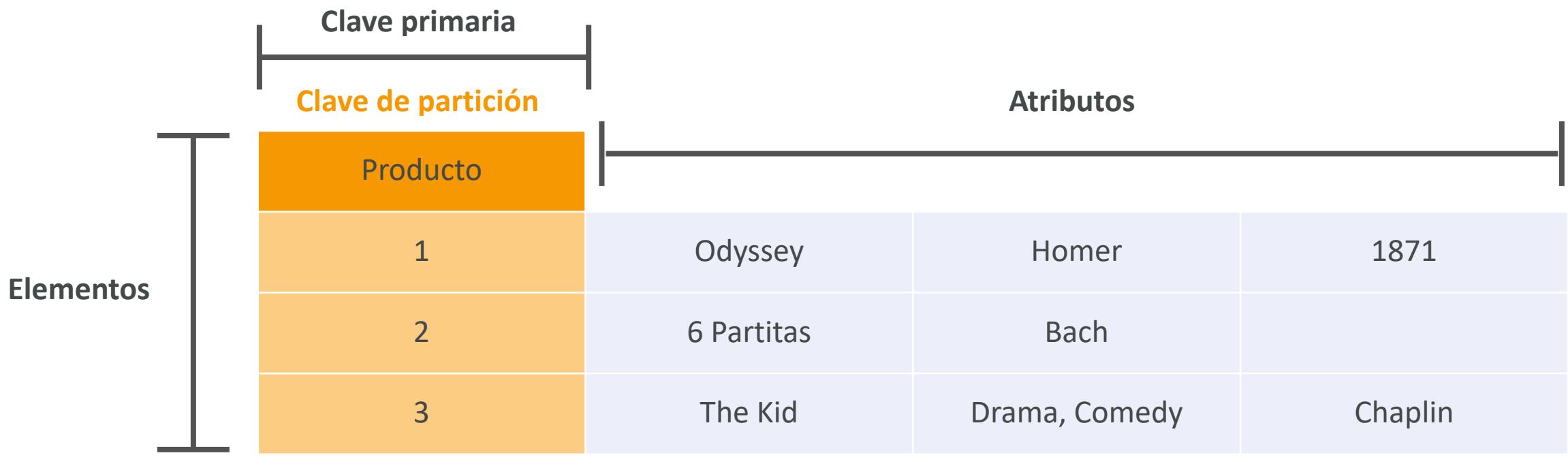
DynamoDB - Conceptos básicos

- DynamoDB está formado por **Tablas**
- Almacenan datos organizados en filas, cada una con una **estructura de clave-valor**
- Cada tabla tiene una **Clave Primaria** (se define al crear la tabla)
 - Esencial para cada tabla, establece la unicidad de cada elemento
- Cada tabla puede tener un número infinito de elementos (= filas)
- Cada elemento tiene **atributos**
 - Los atributos pueden ser nulos, permitiendo flexibilidad en los datos almacenados
- El tamaño máximo de un elemento es de **400 KB**

DynamoDB - Claves primarias

- **Opción 1: Clave de partición (HASH)**

- La clave de partición debe ser única para cada elemento
- La clave de partición debe ser "diversa" para que los datos estén distribuidos



DynamoDB - Claves primarias

- **Opción 2: Clave de partición + Clave de ordenación / clasificación (HASH + RANGO)**
 - La combinación debe ser única para cada elemento
 - Los datos se agrupan por clave de partición

The diagram illustrates the structure of a primary key in DynamoDB. It shows a hierarchical key definition at the top, followed by a table structure below.

Clave primaria:

- Clave de partición:** Represented by the first column in the table.
- Clave de ordenación:** Represented by the second column in the table.

Elementos: The rows in the table represent individual items.

Misma clave de partición (Same partition key): Items 1 and 2 share the same partition key value (1).

Diferente clave de ordenación (Different sort key): Item 1 has a Book ID, while Item 2 has an Album ID.

Atributos: The remaining columns represent attributes for each item.

| | Clave de partición | Clave de ordenación | Atributos | | |
|-----------|--------------------|---------------------|---------------|---------------|---------|
| Elementos | Producto | Tipo | | | |
| | 1 | Book ID | Odyssey | Homer | 1871 |
| | 2 | Album ID | 6 Partitas | Bach | |
| | 2 | Album ID: Track ID | Partita No. 1 | | |
| | 3 | Movie ID | The Kid | Drama, Comedy | Chaplin |

DynamoDB - Claves de partición (Ejercicio)

- ¿Cuál es la mejor clave de partición para maximizar la distribución de los datos?
 - movie_id
 - producer_name
 - leader_actor_name
 - movie_language
 - movie_budget
 - movie_genre
- “**movie_id**” tiene la cardinalidad más alta, por lo que es un buen candidato
- “**movie_language**” no admite muchos valores y puede estar sesgado hacia el inglés, por lo que no es una buena opción para la clave de partición



DynamoDB

Modos de capacidad 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/disminuyen 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

Modos de capacidad R/W - Aprovisionado

- La tabla debe tener **aprovisionadas** unidades de capacidad de lectura y escritura
 - **Unidades de Capacidad de Lectura (RCU)** - rendimiento para lecturas
 - **Unidades de Capacidad de Escritura (WCU)** - rendimiento de escritura
- Opción para configurar el **escalado automático** del rendimiento
- El rendimiento puede superarse temporalmente utilizando la "**Capacidad de ráfaga**" (**Burst Capacity**)
- Si se ha consumido la capacidad de ráfaga, obtendrás una "**ProvisionedThroughputExceededException**"
- En ese caso, se aconseja hacer un reintento de **retroceso exponencial**

Modos de capacidad R/W - Bajo demanda

- Las lecturas/escrituras **aumentan/disminuyen automáticamente** con tus cargas de trabajo
- No necesitas planificar la capacidad (WCU / RCU)
- ⚡⚠ **2,5 veces más caro** que la capacidad aprovisionada (¡con cuidado!)
- WCU y RCU ilimitadas, sin acelerador, más caras
- Te cobran por las lecturas/escrituras que utilizas en términos de RRU y WRU
- **Unidades de petición de lectura (RRU)** - rendimiento de las lecturas (igual que RCU)
- **Unidades de petición de escritura (WRU)** - rendimiento de las escrituras (igual que WCU)
- Casos de uso:
 - Cargas de trabajo desconocidas
 - Tráfico de aplicación impredecible, ...

DynamoDB – Unidades de capacidad de escritura (WCUs)

- Una Unidad de Capacidad de Escritura (WCU) representa una escritura por segundo para un elemento de hasta **1 KB** de tamaño
- 1 WCU permite realizar una escritura por segundo para un ítem de hasta 1 KB
- Si los elementos son mayores de 1 KB, se consumen más WCUs
- Para ítems que excedan 1 KB, el número de WCUs necesarias **aumenta en proporción al tamaño del ítem**. Por ejemplo, un ítem de 2 KB requerirá 2 WCUs por escritura por segundo
- El **tamaño del ítem se redondea al próximo KB** completo para calcular las WCUs. Un ítem de 1.5 KB se tratará como un ítem de 2 KB



DynamoDB – Unidades de capacidad de escritura (WCUs)

- Escribimos 20 elementos por segundo, con un tamaño de elemento de 3 KB

SOLUCIÓN

$$\text{Necesitamos } 20 * \left(\frac{3 \text{ KB}}{1 \text{ KB}} \right) = 60 \text{ WCUs}$$

DynamoDB – Unidades de capacidad de escritura (WCUs)

- **Escribimos 6 elementos por segundo, con un tamaño de elemento de 6,5 KB**

SOLUCIÓN

Necesitamos $6 * \left(\frac{7 \text{ KB}}{1 \text{ KB}}\right) = 42 \text{ WCUs}$ (6,5 se redondea al KB superior)

DynamoDB – Unidades de capacidad de escritura (WCUs)

- **Escribimos 180 elementos por minuto, con un tamaño de elemento de 2 KB**

SOLUCIÓN

$$\text{Necesitamos } \left(\frac{180}{60} \right) * \left(\frac{2 \text{ KB}}{1 \text{ KB}} \right) = 6 \text{ WCUs}$$

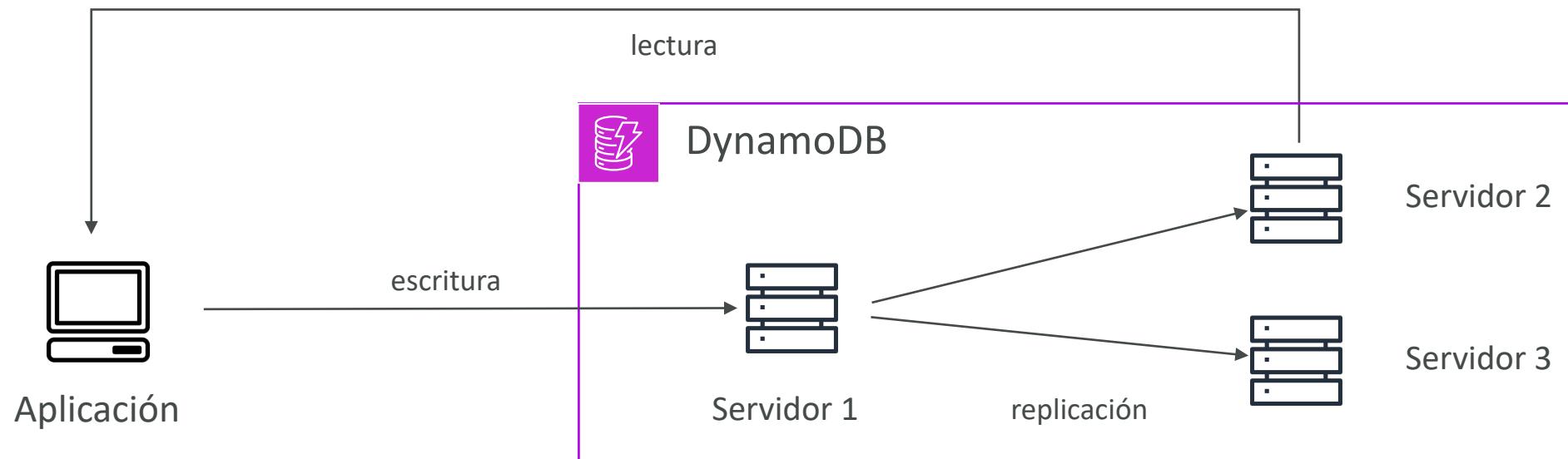
Lectura eventualmente consistente vs. Lectura fuertemente consistente

Lectura eventualmente consistente (por defecto)

- Si leemos justo después de una escritura, es posible que obtengamos algunos datos obsoletos debido a la replicación

Lectura fuertemente consistente

- Si leemos justo después de una escritura, obtendremos los datos correctos
- Pon el parámetro "**ConsistentRead**" a **True** en las llamadas a la API (GetItem, BatchGetItem, Query, Scan)
- Consumir el doble de RCU



DynamoDB – Unidades de capacidad de lectura (RCUs)

- Una Unidad de Capacidad de Lectura (RCU) representa **una lectura fuertemente consistente** por segundo, o **dos lecturas eventualmente consistentes** por segundo, para un elemento de hasta **4 KB** de tamaño
- Para ítems que superen los 4 KB, se necesitarán múltiples RCUs para una sola lectura, dependiendo del tamaño del ítem.



DynamoDB – Unidades de capacidad de lectura (RCUs)

- **16 Lecturas eventualmente consistentes por segundo, con un tamaño de elemento de 12 KB**

SOLUCIÓN

Necesitamos $\left(\frac{16}{2}\right) * \left(\frac{12\ KB}{4\ KB}\right) = 24\ RCUs$

DynamoDB – Unidades de capacidad de lectura (RCUs)

- 10 lecturas fuertemente consistentes por segundo, con un tamaño de elemento de 4 KB

SOLUCIÓN

Necesitamos

$$10 * \left(\frac{4 \text{ KB}}{4 \text{ KB}} \right) = 10 \text{ RCUs}$$

DynamoDB – Unidades de capacidad de lectura (RCUs)

- **10 lecturas fuertemente consistentes por segundo, con un tamaño de elemento de 6 KB**

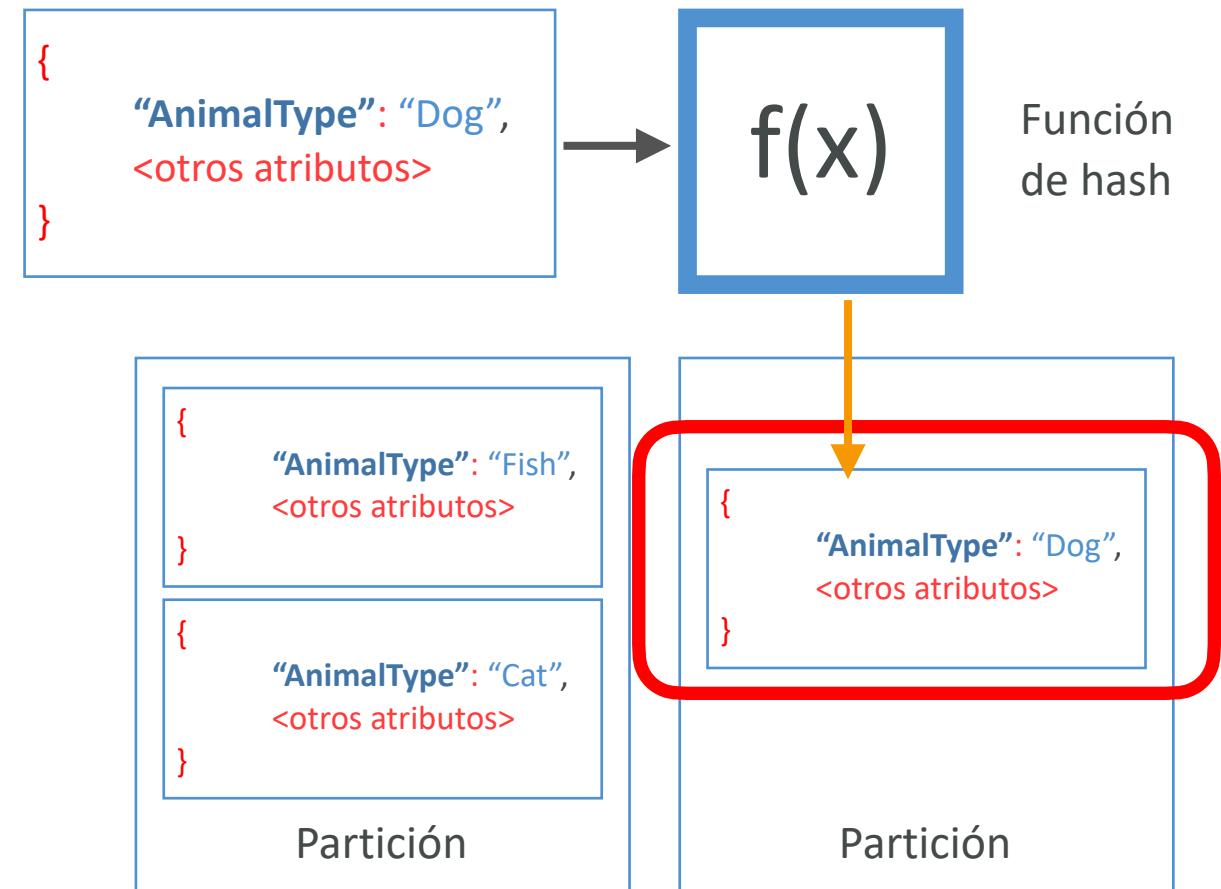
SOLUCIÓN

$$\text{Necesitamos } 10 * \left(\frac{8 \text{ KB}}{4 \text{ KB}} \right) = 20 \text{ RCUs}$$

- Debemos redondear 6 KB a 8 KB

DynamoDB - Particiones y distribuciones

- Los datos se almacenan en particiones
- Las particiones se diseñan para distribuir las operaciones de manera uniforme entre ellas, evitando cuellos de botella y mejorando el rendimiento global
- Para agregar un elemento, DynamoDB aplica una función hash
- El resultado de la función hash determina en qué partición se almacenará el elemento
- Los elementos no se almacenan de forma ordenada; su ubicación depende del valor hash de su clave de partición



DynamoDB - Particiones y distribuciones

- Para calcular el número de particiones:

$$\bullet \ #particiones_{capacidad} = \left(\frac{RCUs_{Total}}{3000} \right) + \left(\frac{WCUs_{Total}}{1000} \right)$$

$$\bullet \ #particiones_{tamaño} = \frac{TamañoTotal}{10GB}$$

$$\bullet \ #particiones = \text{ceil}(\max(\#particiones_{capacidad}, \#particiones_{tamaño}))$$

- **Las WCUs y RCUs se reparten uniformemente por las particiones**

DynamoDB – Throttling (Estrangulamiento)

- Si superamos las RCU o WCU aprovisionadas, obtenemos la siguiente excepción:
“ProvisionedThroughputExceededException”
- 🤔 ¿Cuáles pueden ser los motivos?
 - **Claves calientes** - una clave de partición se lee demasiadas veces (ej: elemento popular)
 - **Particiones calientes**
 - **Elementos muy grandes**, recuerda que RCU y WCU dependen del tamaño de los elementos
- ✅ ¿Cómo se puede solventar el problema?:
 - **Retroceso exponencial** cuando se encuentra una excepción (disponible en el SDK)
 - **Distribuye las claves de partición** tanto como sea posible
 - Si hay problemas de RCU, podemos **utilizar el Acelerador de DynamoDB (DAX)**

DynamoDB - Escritura de datos

- **PutItem**

- Crea o sustituye un elemento con la misma clave primaria
- Consume Unidades de Capacidad de Escritura (WCU)



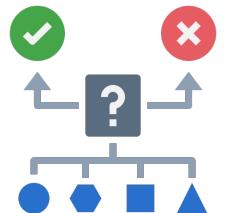
- **UpdateItem**

- Modifica atributos de un elemento o añade uno nuevo si no existe
- Útil para contadores atómicos que incrementan sin condiciones



- **Conditional Writes**

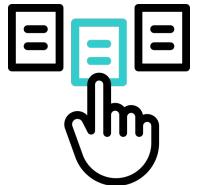
- Realiza escrituras condicionales que solo proceden si se cumplen ciertas condiciones, sin impactar el rendimiento



DynamoDB - Lectura de datos

- **GetItem**

- Lectura basada en clave primaria
- La clave primaria puede ser **HASH** o **HASH+RANGE**
- Lectura eventualmente consistente (por defecto)
- Opción de utilizar lecturas fuertemente consistentes (puede tardar más)
- Se puede especificar una **ProjectionExpression** para recuperar sólo determinados atributos



DynamoDB - Lectura de datos (Consulta)

- La consulta devuelve elementos basados en:
- **KeyConditionExpression:**
 - Clave de partición:
 - Utiliza el operador '=' (obligatorio)
 - Clave de ordenación:
 - Soporta '=', '<', '<=' , '>' , '>=' , 'BETWEEN' , 'BEGINS WITH' (opcional)
- **FilterExpression**:
 - Aplica filtros después de recuperar datos, solo para atributos no clave
- Resultados:
 - Devuelve hasta el número de elementos especificados en Limit o hasta 1 MB de datos.



DynamoDB - Lectura de datos (Escaneado)

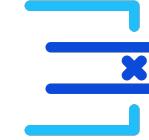
- Escanea toda la tabla, después filtra (ineficiente)
- **Hasta 1 MB** de datos, usa paginación para más
- Controla el impacto con Limit o reduciendo el tamaño de resultados
- Mejora la velocidad usando múltiples "workers" para escanear segmentos simultáneamente, lo que aumenta el RCU usado pero no el impacto de los escaneos
- Usa **ProjectionExpression** y **FilterExpression** para eficiencia sin afectar RCU



DynamoDB - Eliminación de datos

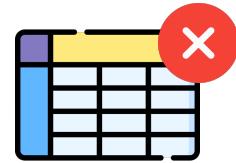
- **DeleteItem**

- Borrar un elemento individual
- Posibilidad de realizar una eliminación condicional



- **DeleteTable**

- Borrar una tabla entera y todos sus elementos
- Eliminación mucho más rápida que llamar a **DeleteItem** en todos los ítems



DynamoDB - Operaciones por lotes

- Te permite ahorrar en latencia reduciendo el número de llamadas a la API
- Las operaciones se realizan en paralelo para una mayor eficiencia
- Parte de un lote puede fallar, en cuyo caso hay que volver a intentarlo con los elementos fallidos

• **BatchWriteItem**

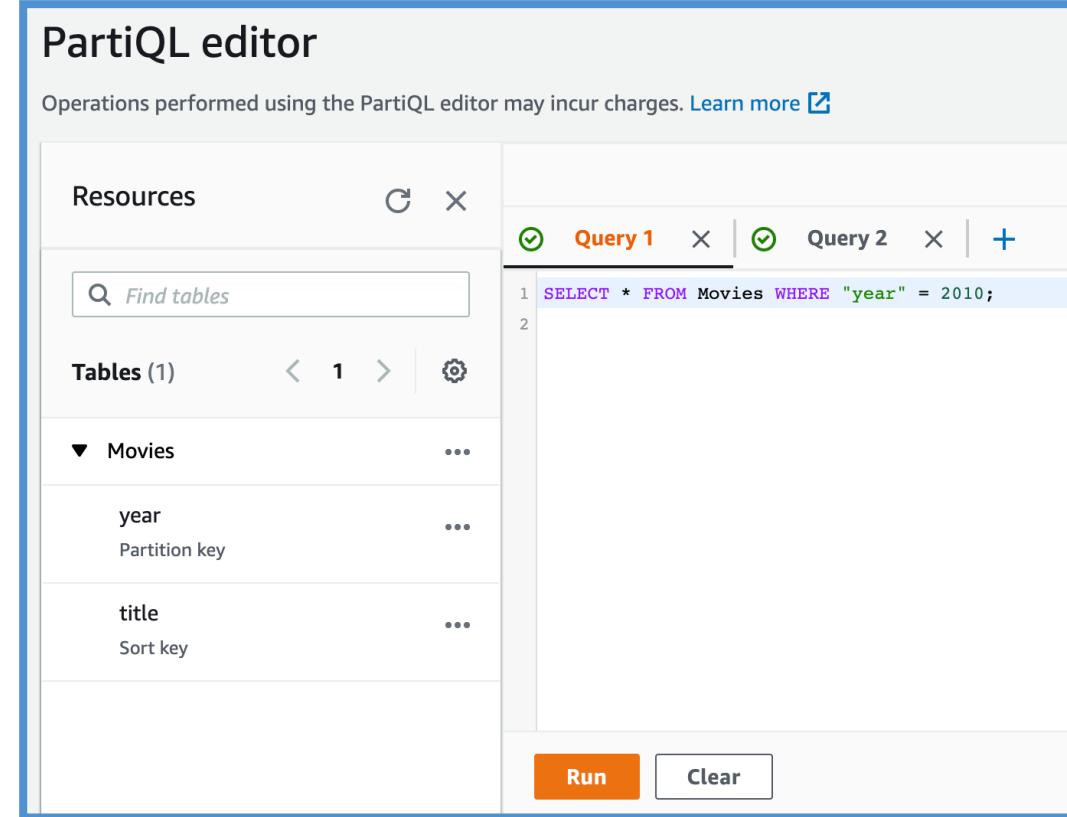
- Hasta 25 **PutItem** y/o **DeleteItem** en una misma llamada
- Hasta 16 MB de datos escritos, hasta 400 KB de datos por elemento
- No se pueden actualizar los elementos (utiliza **UpdateItem**)
- UnprocessedItems para operaciones de escritura fallidas (retroceso exponencial o añadir WCU)

• **BatchGetItem**

- Devuelve elementos de una o varias tablas
- Hasta 100 elementos, hasta 16 MB de datos
- Los elementos se recuperan en paralelo para minimizar la latencia
- UnprocessedKeys para operaciones de lectura fallidas (retroceso exponencial o añadir RCU)

DynamoDB – PartiQL

- Facilita las consultas en **DynamoDB usando la sintaxis SQL**, soporta datos estructurados, semiestructurados y anidados
- Optimiza el rendimiento y minimiza la latencia de las consultas
- Se conecta con otros servicios de AWS para análisis avanzados
- Ejecuta consultas PartiQL desde:
 - Consola de administración de AWS
 - NoSQL Workbench para DynamoDB
 - API de DynamoDB
 - AWS CLI
 - AWS SDK



The screenshot shows the PartiQL editor interface. On the left, there's a sidebar titled "Resources" with a search bar labeled "Find tables". Below it, there's a list of "Tables (1)" with one item named "Movies". Under "Movies", there are two columns: "year" (Partition key) and "title" (Sort key). On the right, there's a main area titled "PartiQL editor" containing two tabs: "Query 1" and "Query 2". The "Query 1" tab is active and contains the following SQL-like query:

```
1 SELECT * FROM Movies WHERE "year" = 2010;
```

Below the queries are "Run" and "Clear" buttons.

DynamoDB - Escrituras condicionales

- Son operaciones que solo se ejecutan si cumplen con una serie de condiciones preestablecidas sobre los atributos de los datos
- Utilizan una sintaxis especial para definir las condiciones que deben cumplirse antes de que la operación de escritura pueda proceder, como *attribute_exists* o *attribute_not_exists*
- **Nota:** Las expresiones de filtro filtran los resultados de las consultas de lectura, mientras que las expresiones de condición son para las operaciones de escritura

```
params = {}  
    'ConditionExpression': (  
        'attribute_not_exists(expirationDate) AND '  
        'attribute_exists(address) AND '  
        'attribute_type(details, :typeMap) AND '  
        'quantity > :minQ AND '  
        'NOT contains(historicalIds, :id) AND '  
        'begins_with(details.address.street, :streetPrefix) AND '  
        'details.address.zipCode = :zipCode AND '  
        '(:val IN attribute_list) AND '  
        'ProductCategory IN (:cat1, :cat2) AND '  
        'price BETWEEN :low AND :high AND '  
        'size(details.description) <= :maxSize AND '  
        '(revenue - cost) > :profit'  
    )  
}
```

Escrituras condicionales - Ejemplo en actualizar elemento

Actualizaciones condicionales

```
aws dynamodb update-item \  
  --table-name ProductCatalog \  
  --key '{"Id": {"N": "456"} }' \  
  --update-expression "SET Price = Price - :discount" \  
  --condition-expression "Price > :limit" \  
  --expression-attribute-values file://values.json
```

Valores iniciales

```
{  
  "Id": {"N": "456"},  
  "ProductCategory": {"S": "Sporting Goods"},  
  "Price": {"N": "650"}  
}
```

values.json

```
{  
  ":discount": {"N": "75"},  
  ":limit": {"N": "500"}  
}
```

- Si el valor inicial de Price es **650**, la operación UpdateItem reduce el Price a **575**.
- Si ejecuta la operación UpdateItem de nuevo, el valor de Price se reduce a **500**.
- Si se ejecuta una tercera vez, la expresión de condición se evalúa en **false** y la actualización no se lleva a cabo.

Escrituras condicionales - Eliminar elemento

- **attribute_not_exists**

- Sólo tiene éxito si el atributo aún no existe (sin valor)

```
aws dynamodb delete-item \  
--table-name ProductCatalog \  
--key '{"Id": {"N": "456"}}' \  
--condition-expression "attribute_not_exists(Price)"
```

- **attribute_exists**

- Opuesto a attribute_not_exists

```
aws dynamodb delete-item \  
--table-name ProductCatalog \  
--key '{"Id": {"N": "456"}}' \  
--condition-expression "attribute_exists(ProductReviews.OneStar)"
```

Escrituras condicionales - Condición compleja

Condición avanzada

```
aws dynamodb delete-item \
  --table-name ProductCatalog \
  --key '{"Id":{"N":"456"}}' \
  --condition-expression "(ProductCategory IN (:cat1, :cat2)) and (Price between :lo and :hi)" \
  --expression-attribute-values file://values.json
```

Valores iniciales

```
{  
  "Id": {"N": "456"},  
  "ProductCategory": {"S": "Sporting Goods"},  
  "Price": {"N": "650"}  
}
```

- En este ejemplo, la comparación de ProductCategory se evalúa en true, pero la comparación de Price se evalúa en false. Esto hace que la expresión de condición se evalúe en false y, por consiguiente, la operación DeleteItem no se lleva a cabo.

values.json

```
{  
  ":cat1": {"S": "Sporting Goods"},  
  ":cat2": {"S": "Gardening Supplies"},  
  ":lo": {"N": "500"},  
  ":hi": {"N": "600"}  
}
```

Escrituras condicionales - Comprobación del valor inicial de cadena

- **begins_with** – comprobar si el prefijo coincide
- **contains** – comprobar si una string está contenida en otra string

Comprobación del valor inicial de cadena

```
aws dynamodb delete-item \  
  --table-name ProductCatalog \  
  --key '{"Id": {"N": "456"}}' \  
  --condition-expression "begins_with(Pictures.FrontView, :v_sub)" \  
  --expression-attribute-values file://expression-attribute-values.json
```

expression-attribute-values.json

```
{  
  ":v_sub": {"S": "http://"}  
}
```

DynamoDB - Índices secundarios

- Algunas aplicaciones solo necesitan consultar datos mediante la clave principal
- Sin embargo, puede haber situaciones en las que una **clave de ordenación alternativa** sería útil
- Para que la aplicación disponga de diversas claves de ordenación entre las que elegir, puedes **crear uno o varios índices secundarios** en una tabla de DynamoDB
- DynamoDB admite dos tipos de índices secundarios:
 - **Índice secundario local**: se considera "local" en el sentido de que el ámbito de todas sus particiones se corresponde con una partición de la tabla base que tiene el mismo valor de clave de partición
 - **Índice secundario global**: se considera "global" porque las consultas que se realizan en el índice pueden abarcar todos los datos de la tabla base y todas las particiones

DynamoDB - Índice Secundario Local (LSI)

- **Clave de ordenación alternativa** para tu tabla (misma **clave de partición** que tabla base)
- La clave de ordenación consiste en un atributo escalar (String, Number o Binary)

| Clave primaria | | Atributos | | |
|--------------------|---------------------|-----------------------|-------|--------|
| Clave de partición | Clave de ordenación | LSI | | |
| User_ID | Game_ID | Game_TS | Score | Result |
| 1234 | 1001 | "2025-02-15T18:18:07" | 95 | Win |
| 6789 | 2002 | "2025-07-20T20:02:32" | | Lose |

- Hasta 5 Índices Secundarios Locales (LSI) por tabla
- **Deben definirse al crear la tabla**
- **Proyecciones de atributos** - pueden contener algunos o todos los atributos de la tabla base
(KEYS_ONLY, INCLUDE, ALL)

DynamoDB - Índice Secundario Global (GSI)

- **Clave primaria alternativa (HASH o HASH+RANGE) de la tabla base**
- Aceleran las consultas sobre atributos no clave, mejorando la eficiencia y velocidad de acceso a los datos
- **Proyecciones de atributos** - algunos o todos los atributos de la tabla base (**ALL, KEYS_ONLY, INCLUDE**)
- Es necesario asignar RCU's y WCU's adecuadas para mantener un rendimiento óptimo del índice.
- **Los GSIs no están restringidos por la clave primaria de la tabla base, permitiendo crear índices completamente independientes para consultas especializadas**
- Facilitan el acceso a grandes volúmenes de datos de manera eficiente

| Clave de partición | Clave de ordenación | Atributos |
|--------------------|---------------------|-----------------------|
| User_ID | Game_ID | Game_TS |
| 1234 | 1001 | "2025-02-15T18:18:07" |
| 6789 | 2002 | "2025-07-20T20:02:32" |

TABLE (consulta por "User_ID")

| Clave de partición | Clave de ordenación | Atributos |
|--------------------|-----------------------|-----------|
| Game_ID | Game_TS | User_ID |
| 1001 | "2025-02-15T18:18:07" | 1234 |
| 2002 | "2025-07-20T20:02:32" | 6789 |

INDEX GSI (consulta por "Game_ID")

DynamoDB - Índices y estrangulamiento

- **Índice Secundario Local (LSI)**

- Utiliza las WCUs y RCUs de la tabla principal
- Sin consideraciones especiales de estrangulamiento

- **Índice Secundario Global (GSI)**

- **⚠ Si se estrangulan las escrituras en el GSI, ¡se estrangulará la tabla principal!**

- *Aunque las WCU de las tablas principales estén bien
- Recomendaciones:
 - Elige con cuidado tu clave de partición GSI
 - Asigna cuidadosamente la capacidad de tu WCU



DynamoDB Accelerator - DAX

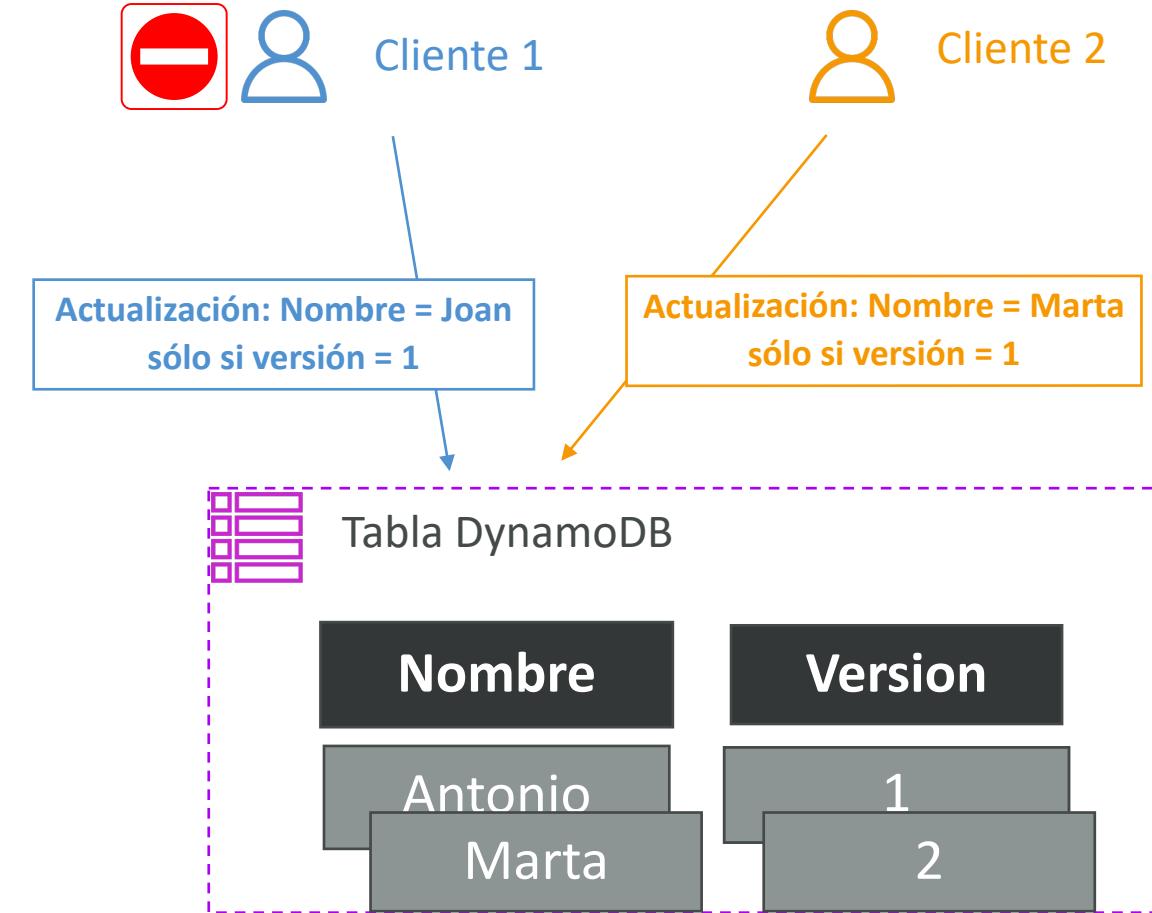
- **Caché en memoria** totalmente gestionada para DynamoDB
- **Mejora del rendimiento x 10** - latencia de un milisegundo a microsegundos - al acceder a tus tablas de DynamoDB



- Seguridad, alta escalabilidad y alta disponibilidad
- Diferencia con ElastiCache: **DAX sólo se utiliza y se integra con DynamoDB**, mientras que ElastiCache puede utilizarse para otras bases de datos

DynamoDB - Bloqueo optimista

- Estrategia para asegurarse de que el elemento del lado del cliente que se va a actualizar (o eliminar) sea el mismo que figura en Amazon DynamoDB
- Las escrituras en la base de datos se **protegen contra posibles sobrescrituras** de otros y viceversa
- Con el bloqueo optimista, cada elemento tiene un atributo que actúa como número de versión
- Si recuperas un elemento de una tabla, la aplicación registra el número de versión de ese elemento



DynamoDB Streams

- Captura una **secuencia en orden cronológico de las modificaciones** (inserciones, actualizaciones y borrados) de los elementos en una tabla de DynamoDB
- Permite acceder a los datos de los cambios en los últimos **24 horas**, incluyendo el contenido de los elementos modificados
- Las aplicaciones pueden obtener acceso a este registro y ver los elementos de datos tal y como se encontraban antes y después de la modificación
- Facilita la creación de arquitecturas impulsadas por eventos,
- **Casos de uso:**
 - Reaccionar a los cambios en tiempo real (correo electrónico de bienvenida a los usuarios)
 - Sincronizar datos en tiempo real entre tablas de DynamoDB en diferentes regiones
 - Analíticas, activación de funciones Lambda, auditoría de cambios, etc...