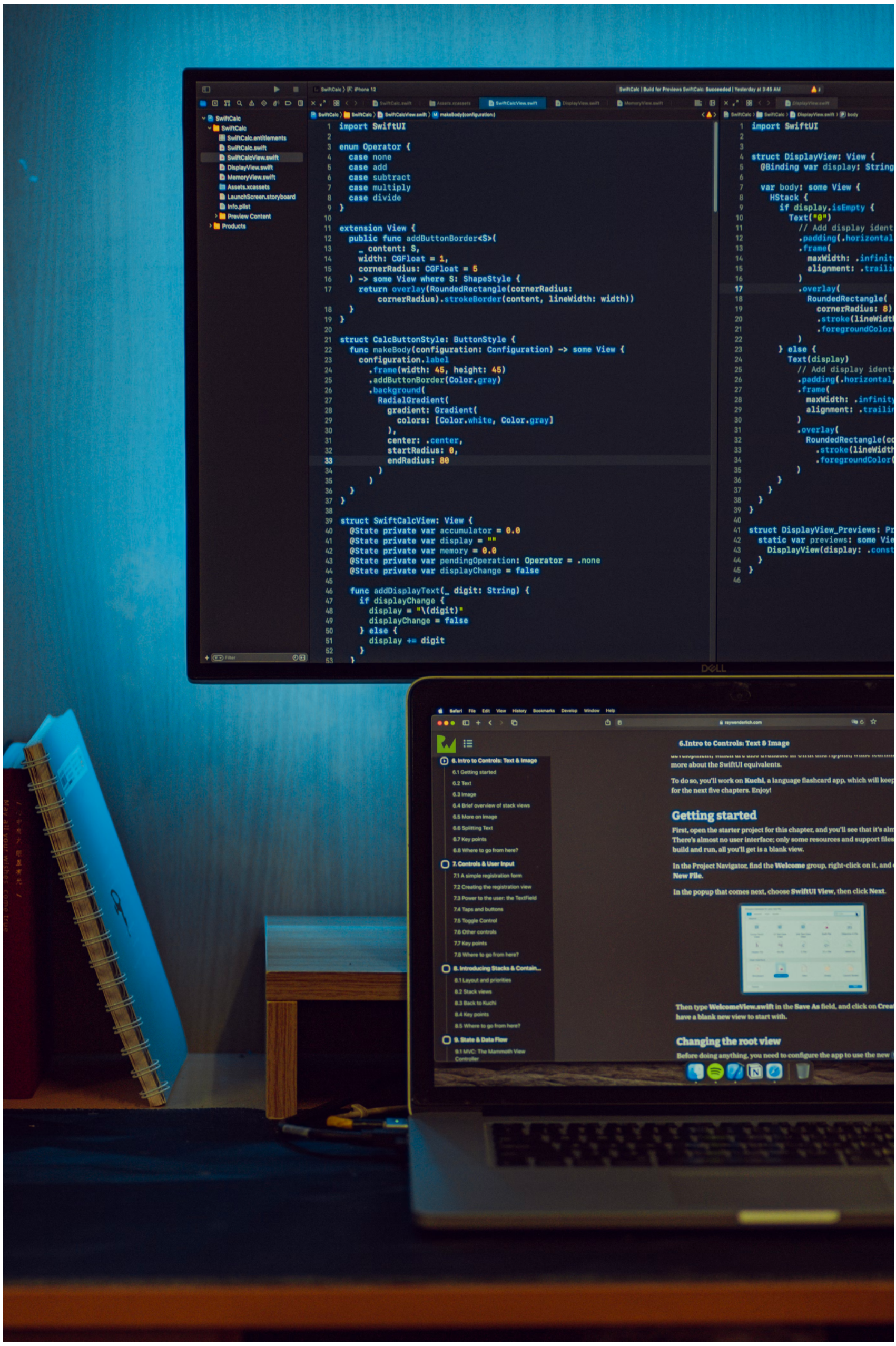
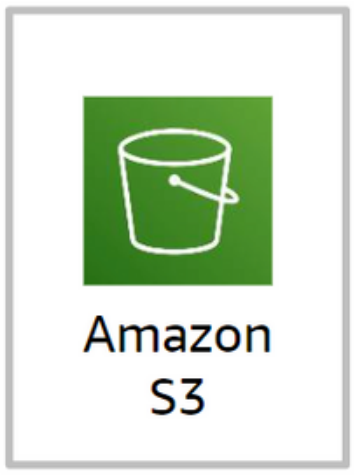
Lección 2: Almacenamiento por Objeto con AWS S3



Amazon S3



Un servicio de almacenamiento de objetos:

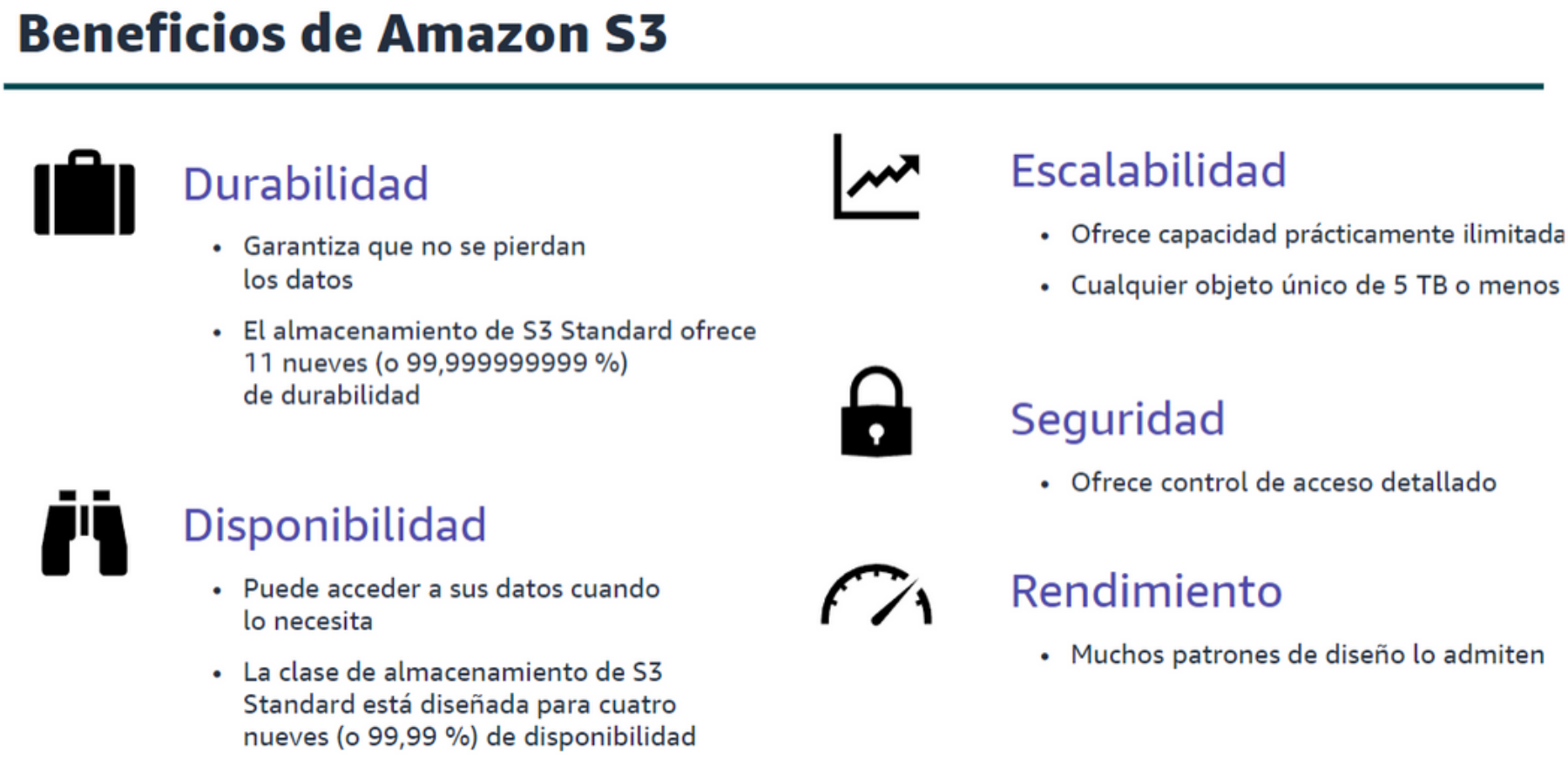
- Almacena cantidades masivas (ilimitadas) de datos no estructurados
- Los archivos de datos se almacenan como objetos en un bucket que usted define
- 5 TB es el tamaño máximo de archivo de un solo objeto
- Todos los objetos tienen una URL única accesible globalmente por REST (espacio de nombres universal)
- Todos los objetos tienen una clave, ID de versión, valor, metadatos y subrecursos

Amazon S3 es un servicio de almacenamiento de objetos. Le permite almacenar cantidades casi ilimitadas de datos. Los archivos de datos se almacenan como objetos. Coloca los objetos en un bucket, que usted define. Cada bucket debe tener un nombre que sea globalmente único entre las regiones. Esto significa que el nombre del bucket tiene que ser único entre todas las cuentas de clientes de AWS.

Los objetos que almacena pueden variar de tamaño de 0bytes a 5TB. Aunque los objetos individuales no pueden superar los 5TB, puede almacenar la cantidad total de datos que necesite.Cada objeto tiene cinco características consistentes. Primero, tiene una clave, que es el nombre que le asigna a un objeto. La clave del objeto se usa para recuperar el objeto. En la Consola de administración de AWS, se puede crear un directorio dentro de un bucket y cargar un objeto en ese directorio. Sin embargo, en realidad, Amazon S3 no conoce los directorios, por lo que el valor de clave incluye la ruta completa a raíz del bucket. Los objetos también incluyen un ID de versión. En un bucket, una clave y un ID de versión identifican un objeto de forma única, aprenderá más sobre el control de versiones, más adelante en esta unidad.

El valor del objeto es el contenido real que almacena. Puede ser una secuencia de bytes. Los valores del objeto son inmutables, es decir que después de que carga un objeto, no puede modificar el valor. Si desea modificar el objeto, debe hacer un cambio fuera de Amazon S3 y volver a cargar el objeto. Los objetos también incluyen metadatos, que es un conjunto de pares de nombre-valor que se puede usar para almacenar información sobre el objeto.

También se permite asignar metadatos, los que se denomina metadatos definidos por el usuario, a sus objetos en Amazon S3. Amazon S3 también asigna metadatos del sistema a estos objetos, que utiliza para administrar objetos. Por último, Amazon S3 también usa subrecursos para almacenar información adicional específica del objeto.



Amazon S3 ofrece muchas funciones que lo hacen un componente importante de muchas soluciones creadas en AWS. En primer lugar, ofrece durabilidad, que describe el promedio de la pérdida anual prevista de objetos. Once nueves de durabilidad significa que cada año, hay un 0,000000001 por ciento de probabilidad de perder un objeto. Por ejemplo, si almacena 10000 objetos con Amazon S3, puede esperar una pérdida de un solo objeto cada 10000000 años, en promedio.

Amazon S3 almacena de forma redundante sus objetos en varios dispositivos en múltiples instalaciones en la región de Amazon S3 que designe. Amazon S3 está diseñado para sostener errores de dispositivos simultáneos detectando con rapidez y reparando cualquier redundancia perdida. Amazon S3 también verifica con regularidad la integridad de sus datos mediante sumas de verificación. Amazon S3 también ofrece cuatro nueves (o 99,99 por ciento) de disponibilidad. La disponibilidad se refiere a su capacidad de acceder a los datos con rapidez, cuando lo desee. También ofrece una capacidad casi ilimitada para almacenar sus datos, por lo que es escalable. Amazon S3 tiene una configuración de seguridad robusta. Ofrece muchas maneras de controlar el acceso a los datos que almacena y también le permite cifrar los datos.

Por último, Amazon S3 tiene un alto rendimiento, con una latencia hasta el primer byte que se mide en milisegundos para la mayoría de las clases de almacenamiento. Para obtener más información sobre los patrones de diseño de rendimiento de S3, consulte la documentación de Amazon S3. Los enfoques comunes incluyen usar el almacenamiento en caché para contenido de acceso frecuente; lógica configurable de reintentos y tiempo de espera para objetos que reciben un significativo tráfico de solicitudes en un período breve; y escalado horizontal y paralelización de solicitudes para obtener un alto rendimiento en toda la red.

Patrones de uso comunes de Amazon S3

Amazon S3

¿Qué problemas puede resolver con Amazon S3?
Ahora examinará algunos casos prácticos.

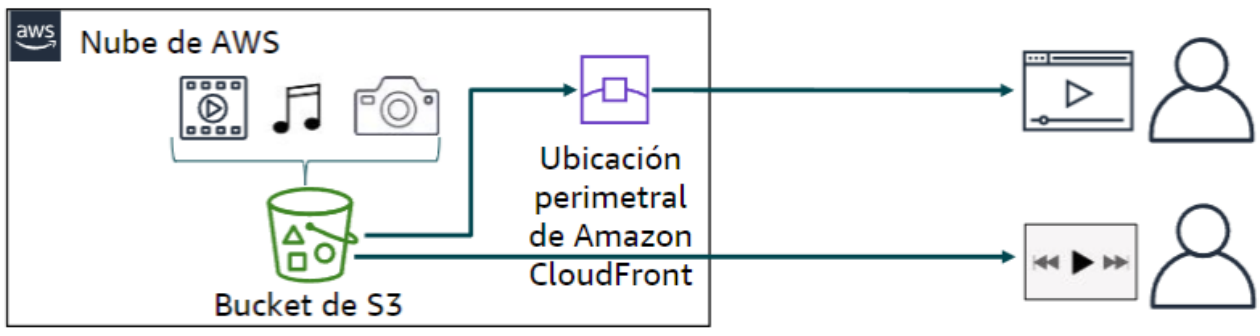
Ahora que conoce muchas funciones de Amazon S3, ¿cómo puede usar estas funciones para atender sus necesidades? En esta lección de la unidad, conocerá cuatro casos prácticos comunes que usan Amazon S3 como una parte esencial de una solución de arquitectura robusta.

Caso práctico 1 de Amazon S3: Almacenar y distribuir contenido y multimedia web

Cree una infraestructura redundante, escalable y de alta disponibilidad que aloje cargas y descargas de videos, fotos o música.

https://<bucket-name>.s3.amazonaws.com

https://<bucket-name>.s3.amazonaws.com/video.mp4



Una situación de uso común de Amazon S3 es utilizarlo para el alojamiento de medios. En este caso práctico, Amazon S3 se usa para almacenar y distribuir videos, fotos, archivos de música y otros medios. Este contenido se puede entregar directamente desde Amazon S3 porque cada objeto en Amazon S3 tiene una URL HTTP única.

Como alternativa, Amazon S3 puede actuar como un almacén de origen para una red de entrega de contenido (CDN), como Amazon CloudFront. La elasticidad de Amazon S3 lo hace ideal para alojar contenido web que necesita ancho de banda para dar respuesta a los picos extremos de demanda. Además, como no necesita aprovisionar almacenamiento para Amazon S3, funciona bien para sitios web de rápido crecimiento que alojan contenido de uso intensivo de datos generado por el usuario, como sitios de uso compartido de videos o fotos.

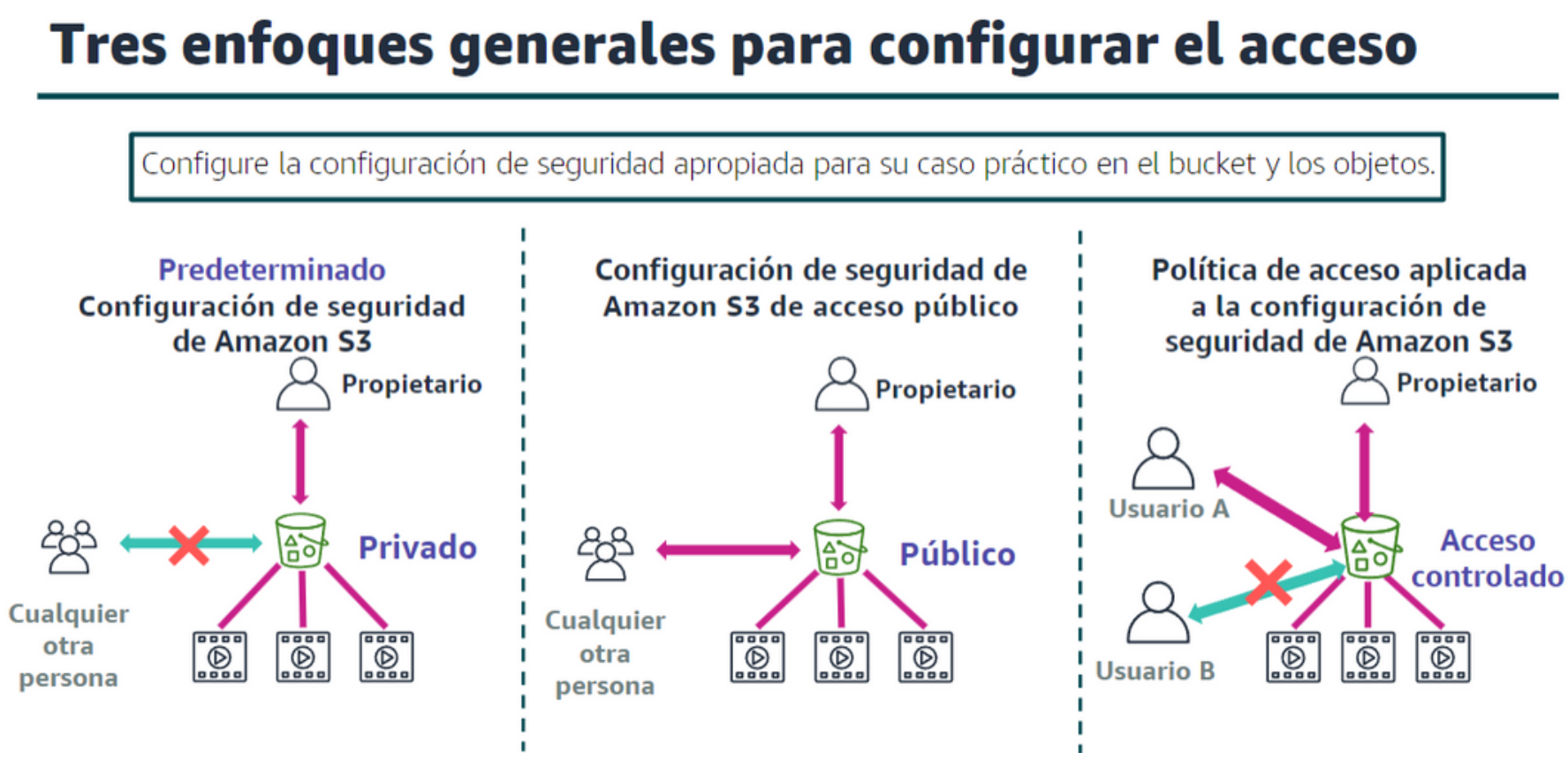
Protección de buckets y objetos de Amazon S3

- Los buckets y objetos que se crean recientemente son privados y están protegidos de forma predeterminada.
- Cuándo los casos prácticos deben compartir datos de Amazon S3:
  - Administre y controle el acceso a los datos.
  - Siga el principio de mínimo privilegio
- Herramientas y opciones para controlar el acceso a los datos de Amazon S3:
  - Función de bloqueo de acceso público: está habilitada en buckets nuevos de forma predeterminada y es fácil de administrar
  - Políticas de IAM: una buena opción cuando el usuario puede autenticarse mediante IAM
  - Políticas de bucket: puede definir el acceso a un objeto o bucket específico
  - Listas de control de acceso (ACLs): Un mecanismo de control de acceso heredado
  - Puntos de acceso S3: Puede configurar el acceso con nombres y permisos específicos para cada aplicación.
  - URL prefirmadas: puede conceder acceso limitado a objetos con URL temporales.
  - Comprobación de permisos de bucket de AWS Trusted Advisor: una función gratuita

De forma predeterminada, todos los buckets de S3 son privados y solo pueden acceder a ellos los usuarios a los que se ha concedido acceso explícitamente. Es esencial que administre y controle el acceso a los datos de Amazon S3. AWS ofrece muchas herramientas y opciones para controlar el acceso a sus buckets de S3 u objetos, herramientas tales como:

- Bloqueo del acceso público de Amazon S3. Esta configuración anula otras políticas o permisos de objetos. Habilite el bloqueo del acceso público para todos los buckets que no desee que sean de acceso público. Esta función ofrece un método sencillo de evitar la exposición no intencional de datos de Amazon S3.
- Escribir políticas de AWS Identity and Access Management (AWS IAM) que especifiquen los usuarios o roles que pueden acceder a buckets y objetos específicos.
- Escribir políticas de bucket que definan el acceso a bucket sus objetos específicos. Esta opción se usa habitualmente cuando el usuario o sistema no puede autenticarse mediante IAM. Se pueden configurar políticas de bucket para otorgar acceso entre cuentas de AWS o para otorgar acceso público o anónimo a los datos de Amazon S3. Si se usan políticas de bucket, deben escribirse con cuidado y probarse completamente. Puede especificar un enunciado de denegación en una política de bucket para restringir el acceso. El acceso se restringirá incluso si los usuarios tienen permisos que se otorgan en una política basada en identidad adjunta a los usuarios.

- Crear puntos de acceso de S3. Los puntos de acceso son nombres de host únicos que aplican permisos y controles de red distintos para solicitudes que se realizan a través de esos puntos. Los clientes con conjuntos de datos compartidos pueden escalar el acceso para muchas aplicaciones creando puntos de acceso individualizados con nombres y permisos personalizados para cada aplicación. Configurar listas de control de acceso (ACL) en sus buckets y objetos. Las ACL se usan con menos frecuencia (las ACL son anteriores a IAM). Si usa ACL, no configure un acceso que sea demasiado abierto o permisivo.
- AWS Trusted Advisor ofrece una función de comprobación de permisos de bucket. Es una herramienta útil para descubrir si cualquiera de los buckets en su cuenta tiene permisos que otorgan acceso global.



Hay tres enfoques generales diferentes para configurar el acceso a los objetos en un bucket de S3. La parte izquierda de la imágen, muestra la configuración de seguridad predeterminada de Amazon S3. De forma predeterminada, los buckets de Amazon S3 y los objetos almacenados en ellos son privados (están protegidos).

Las únicas entidades con acceso a un bucket recién creado, sin modificar, son el administrador de la cuenta y el usuario raíz de la cuenta de AWS. El propietario del recurso puede otorgar permisos de acceso específicos a otros, pero a quienes no se les hayan concedido esos permisos no tendrán acceso. En la mitad de la imágen se puede notar una situación en la que se ha desactivado la configuración de seguridad de S3 y cualquiera puede acceder públicamente a los objetos almacenados en el bucket.

¡Precaución!

Usar un bucket de Amazon S3 para alojar un sitio web estático es un ejemplo de cómo configurar una arquitectura de AWS rápidamente. Sin embargo, para la mayoría de los casos prácticos de Amazon S3, no es recomendable otorgar acceso público a Amazon S3. En la mayoría de los casos prácticos, no se requiere acceso público.

Con más frecuencia, Amazon S3 se usa para almacenar datos utilizados por una aplicación que se ejecuta fuera de Amazon S3, o para respaldar información confidencial. Para estos casos prácticos comunes, nunca se debe otorgar acceso público a los buckets que contienen los datos.

A la derecha la imagen muestra un caso en que se configuró Amazon S3 para proporcionar acceso controlado. Al Usuario A se le otorgó acceso a los objetos en el bucket, pero al Usuario B se le denegó el acceso. Las situaciones de acceso controlado son frecuentes. Las puede configurar el propietario del bucket mediante una o más de las herramientas o las opciones para controlar el acceso a los datos de Amazon S3 que se analizaron anteriormente en este módulo.

Cuando su objetivo es proteger los datos digitales, el cifrado de datos es una herramienta esencial. El cifrado de datos toma los datos que son legibles y los codifica. Los datos cifrados son ilegibles para cualquiera que no tenga acceso a la clave secreta que se puede usar para decodificarlos. Por lo tanto, si un atacante obtiene acceso a sus datos, no puede comprenderlos. Tiene dos opciones principales para cifrar los datos almacenados en Amazon S3.

Cuando configura la opción de cifrado Predeterminado de un bucket, se habilita el cifrado del lado del servidor. Con esta función, Amazon S3 cifra su objeto antes de guardarlo en el disco. Luego, Amazon S3 los descifrará cuando se descargue el objeto. El cifrado del cliente es la otra opción. Cuando usa este enfoque, cifra los datos en el lado del cliente antes de cargarlo en Amazon S3. En este caso, administrar el proceso de cifrado, las claves de cifrado y las herramientas relacionadas. Al igual que el cifrado del lado del servidor, el cifrado del cliente puede reducir el riesgo cifrando los datos con una clave que se almacena en un mecanismo diferente que el mecanismo que almacena los datos en sí.

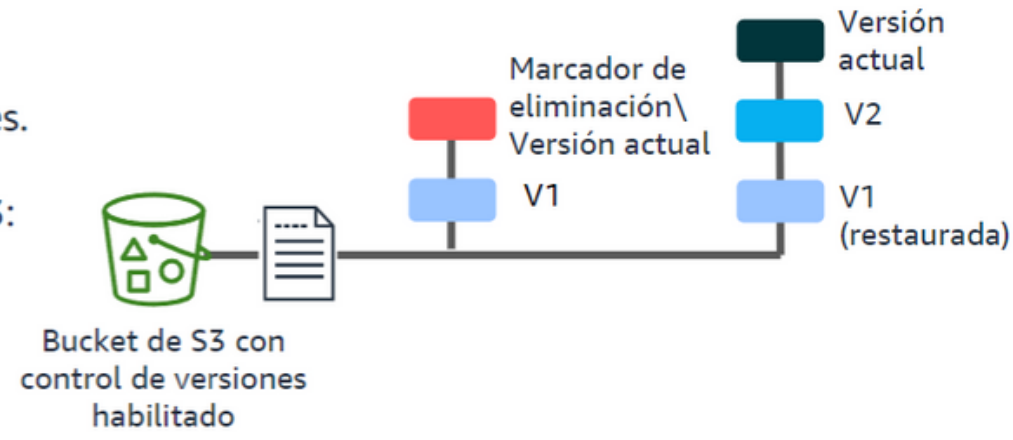
Un segundo caso práctico de Amazon S3 es utilizar el servicio para alojar un sitio web estático. En un sitio web estático, las páginas web individuales incluyen contenido estático. También pueden incluir scripts del lado del cliente. Por el contrario, un sitio web dinámico depende del procesamiento del lado del servidor, que podría implicar consultas de la base de datos que se ejecutan en respuesta a los scripts del lado del servidor, como PHP, JSP o ASP.NET. Amazon S3 no admite el scripting del lado del servidor.

Sin embargo, AWS ofrece otros servicios que le permiten alojar sitios web dinámicos. Para alojar un sitio web estático, configure un bucket de S3 para alojamiento de sitios web. Luego, cargue el contenido de su sitio web en el bucket. El ejemplo muestra que el sitio estático podría consistir en archivos HTML, imágenes, videos y scripts del cliente en formatos como JavaScript. Con este enfoque, no necesita ejecutar una máquina virtual que aloje un servidor web. De hecho, no necesita ejecutar un servidor. Sin embargo, igualmente puede alojar un sitio web. Amazon S3 proporciona una solución de bajo costo para alojamiento web que incluye alto rendimiento, escalabilidad y disponibilidad.

Práctica recomendada de Amazon S3: control de versiones

- Protege contra sobrescrituras y eliminaciones accidentales sin penalización de rendimiento
- Genera una nueva versión con cada carga
- Permite recuperar fácilmente objetos eliminados o restaurar versiones anteriores.
- Tres estados posibles de una bucket de S3:

1. *Predeterminado*: control de versiones no habilitado
2. Control de versiones habilitado
3. Control de versiones suspendido



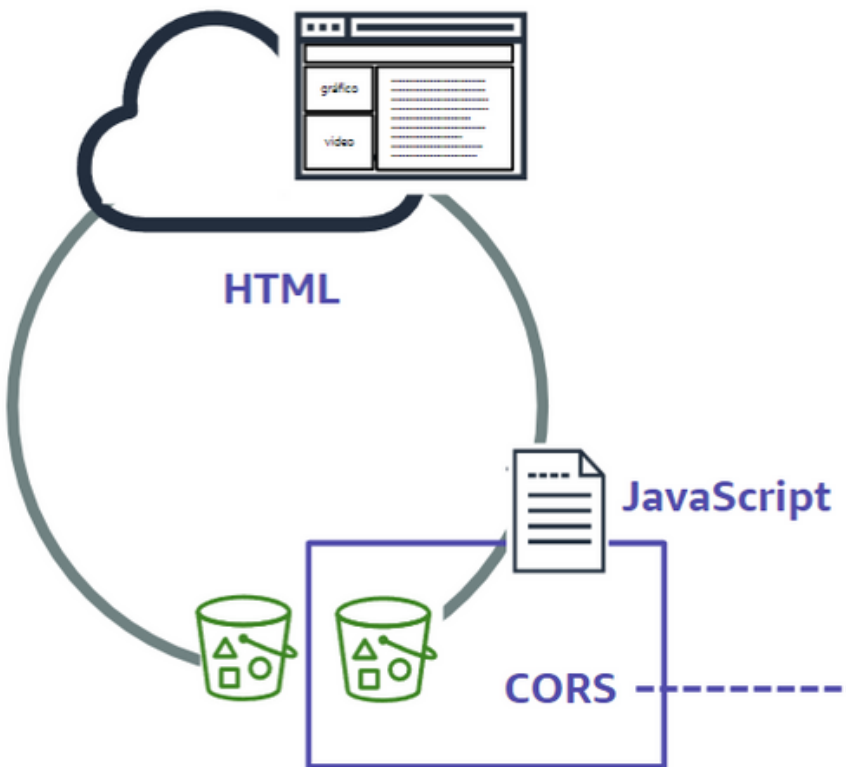
Amazon S3 ofrece a los clientes una infraestructura de almacenamiento muy segura y duradera. El control de versiones ofrece un nivel adicional de protección. Proporciona una manera de recuperar los datos si falla una aplicación o cuando los clientes sobrescriben o borran objetos por accidente. El control de versiones es un método para conservar diversas variantes de un objeto en el mismo bucket. Puede utilizar el Control de Versiones para conservar, recuperar y restaurar todas las versiones de los objetos almacenados en un bucket de S3.

Si elimina un objeto, en lugar de eliminarlo de forma permanente, Amazon S3 inserta un marcador de eliminación, que se convierte en la versión actual del objeto. Siempre puede restaurar la versión anterior.
Sobrescribir un objeto y crear una nueva versión del objeto en el bucket. Siempre puede restaurar la versión anterior.

Los buckets pueden estar en uno de tres estados: (I) sin control de versiones (el valor predeterminado), (II) control de versiones habilitado o (III) control de versiones suspendido. Después de habilitar el control de versiones para un bucket, no puede cambiarlo a un estado sin control de versiones. Sin embargo, puede suspender el control de versiones para ese bucket.

Ahora, el instructor podría elegir **demostrar el control de versiones de Amazon S3 con la Consola de administración de AWS.**

Compatibilidad con el uso compartido de recursos entre orígenes (CORS)



Ejemplo de configuración CORS

```
<CORSConfiguration>
 <CORSRule>

<AllowedOrigin>http://www.example.com</AllowedOrigin>
 <AllowedMethod>PUT</AllowedMethod>
 <AllowedMethod>POST</AllowedMethod>
 <AllowedMethod>DELETE</AllowedMethod>
 <AllowedHeader>*</AllowedHeader>
 <MaxAgeSeconds>3000</MaxAgeSeconds>
 <ExposeHeader>x-amz-server-side-
encryption</ExposeHeader>
 <ExposeHeader>x-amz-request-id</ExposeHeader>
 <ExposeHeader>x-amz-id-2</ExposeHeader>
 </CORSRule>
</CORSConfiguration>
```

El uso compartido de recursos entre orígenes (CORS) define una forma para que las aplicaciones web del cliente que, están cargadas en un dominio, interactúen con los recursos en un dominio distinto. Con el respaldo de CORS, puede crear aplicaciones web del lado del cliente completas con Amazon S3. Asimismo, puede permitir un acceso entre orígenes a sus recursos de Amazon S3 de manera selectiva. Para configurar el bucket de modo que permita solicitudes entre orígenes, cree una configuración CORS. Una configuración CORS es un documento XML con reglas que identifican:

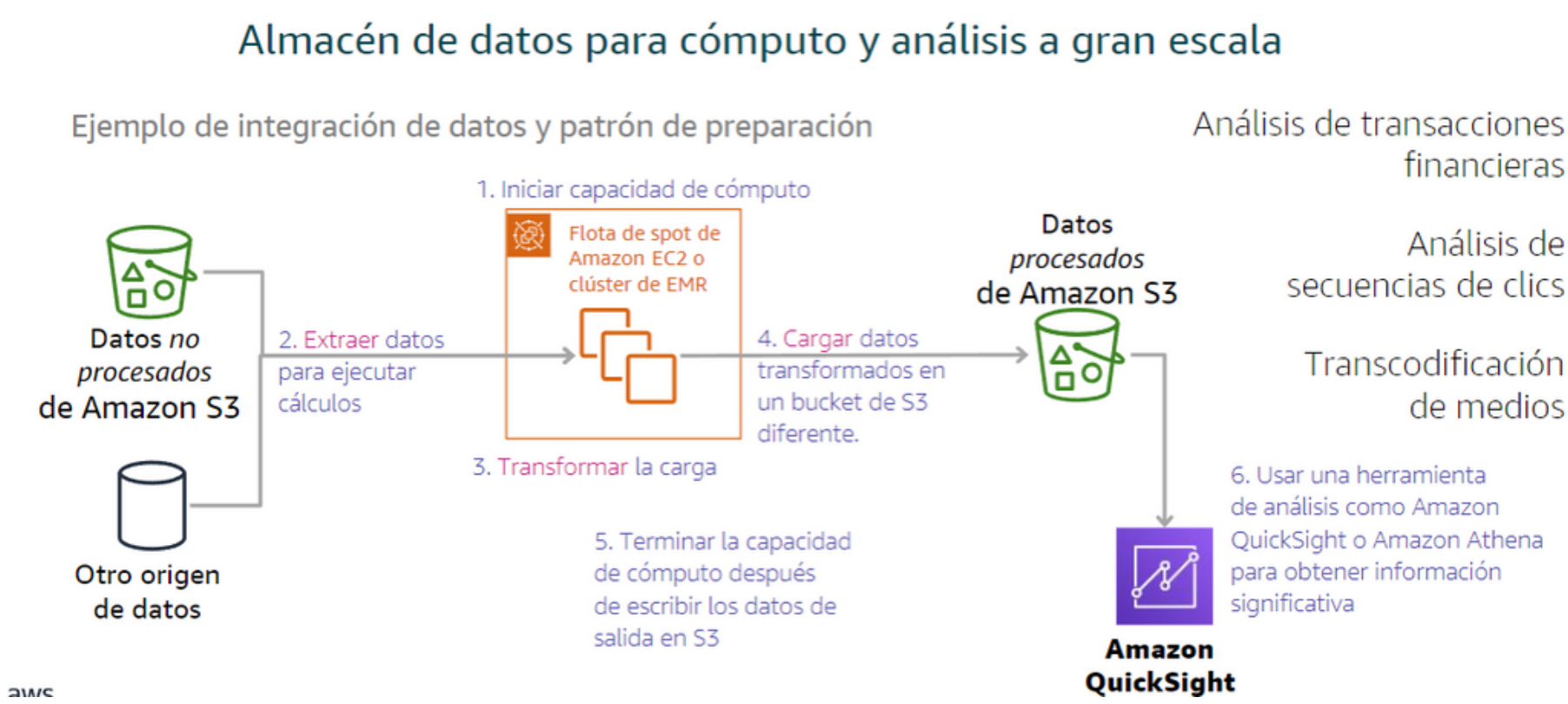
- Los orígenes a los que les permitirá acceder a su bucket.
- Las operaciones (métodos HTTP) que admitirán a cada origen: En este ejemplo, se permiten las solicitudes PUT, POST y DELETE desde el origen http://www.example.com, que se podría configurar con Amazon Route 53 para que sea otro bucket de S3.
- Otra información específica de operaciones.

Aquí se completa la unidad 2 con el laboratorio guiado: **Alojamiento de un sitio web estático.**

Laboratorio guiado: tareas

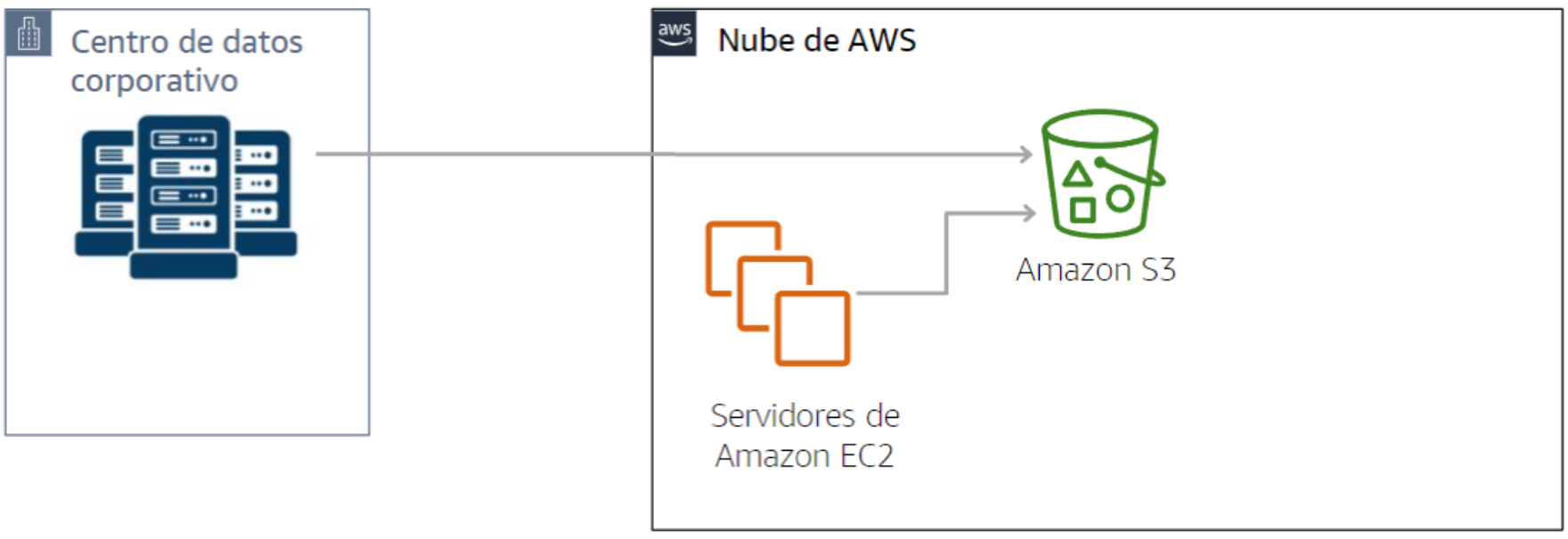
1. Crear un bucket en Amazon S3
2. Cargar contenido en el bucket
3. Habilitar el acceso a los objetos
4. Actualizar el sitio web

Caso práctico 3 de Amazon S3: Almacén de datos para cómputo y análisis



También puede usar Amazon S3 como un almacén de datos para cómputo o análisis a gran escala, como análisis de transacciones financieras, análisis de secuencias de clics y transcodificación de medios. Amazon S3 puede admitir estas cargas de trabajo gracias a su capacidad de escalado horizontal, que posibilita la ejecución de múltiples transacciones simultáneas.En este ejemplo, una flota de spot de Amazon Elastic Compute Cloud (Amazon EC2) se inicia cuando el precio de oferta de las instancias de spot es bajo o cuando se inicia un clúster de Amazon EMR. De todas formas, después de que la capacidad de cómputo está disponible, se extraen los datos sin procesar de Amazon S3 y de otro origen de datos. Los datos se ejecutan mediante algoritmos de cómputo que los integran y transforman. Los datos procesados resultantes se cargan en un bucket de Amazon S3 diferente. Ahora que se han procesado los datos, se termina la capacidad de cómputo para ahorrar costos. Por último, se podría usar una herramienta de análisis, como Amazon QuickSight, para obtener información significativa de los datos procesados. Este es tan solo una situación de ejemplo de cómo Amazon S3 puede tener un papel fundamental en el almacenamiento de datos en una arquitectura de soluciones de análisis de gran escala.

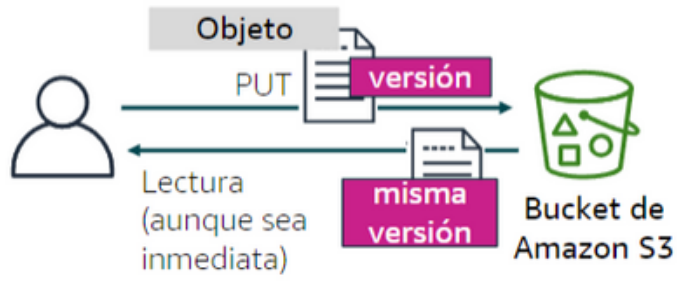
Caso práctico 4 de Amazon S3: Respaldar y archivar los datos críticos



En el cuarto y último caso práctico que se analiza en esta unidad, Amazon S3 se usa como una solución de respaldo de datos. Debido a su naturaleza muy escalable y de larga duración, Amazon S3 funciona bien como herramienta de respaldo y archivado de datos. En la situación, los datos se respaldan desde un centro de datos corporativo en las instalaciones y de una gran cantidad de servidores de Amazon EC2. Estos servidores ejecutan aplicaciones que generan datos. Además, puede mover datos a largo plazo desde el almacenamiento estándar de Amazon S3 hacia Amazon Simple Storage Service Glacier. Este proceso se analizará con más detalle, posteriormente, en esta unidad. Otra opción de Amazon S3 que puede configurar en sus buckets, para lograr niveles aún mayores de durabilidad, es la replicación entre regiones. En la replicación entre regiones, los objetos que se cargan en un bucket en una región se copiará automáticamente en otros buckets de S3 en otras regiones.

Modelo de consistencia de datos de Amazon S3

- Amazon S3 tiene una **consistencia alta** para todos los *objetos* nuevos y existentes en todas las regiones
  - Ofrece consistencia de lectura tras escritura para todas las operaciones en objetos GET, LIST y PUT en buckets de S3
  - El modelo de consistencia ofrece una ventaja para las cargas de trabajo de big data
- Las configuraciones de bucket tienen un modelo eventualmente consistente



Muchos clientes desarrollan aplicaciones de análisis de big data que usan Amazon S3 para el almacenamiento de objetos. Estas aplicaciones suelen requerir acceso a un objeto inmediatamente después de una operación de escritura. Antes de diciembre de 2020, Amazon S3 proporciona consistencia eventual para sobrescribir las operaciones PUT y DELETE en todas las regiones. Sin embargo, Amazon S3 ahora tiene consistencia alta para todos los objetos de S3 nuevos y existentes en todas las regiones de AWS. Amazon S3 logra una alta disponibilidad replicando los datos en varios servidores en centros de datos de AWS. Si una solicitud PUT se realiza correctamente, los datos se guardan de forma segura. Cualquier operación de lectura (GET o LIST) que se inicie después de una respuesta PUT devolverá los datos escritos por la operación PUT. Esta consistencia alta de lectura después de la escritura existe automáticamente para todas las aplicaciones, sin cambios en el rendimiento o la disponibilidad.

La consistencia alta simplifica la migración de cargas de trabajo de análisis en las instalaciones eliminando la necesidad de hacer cambios para admitir aplicaciones. Tampoco es necesaria una infraestructura adicional, como S3 Guard, para proporcionar una consistencia alta.

Aunque los objetos tienen una consistencia alta, las configuraciones de bucket de Amazon S3 tienen un modelo de consistencia eventual. Por ejemplo, si elimina un bucket y de inmediato hace una lista de todos los buckets, el bucket eliminado igualmente podría aparecer en la lista. Sin embargo, en poco tiempo, si vuelve a ejecutar el comando de lista de buckets, el bucket eliminado ya no aparecerá en los resultados de la lista de buckets.