

**Lectura:** Lectura 3.3:  
Almacenamiento de objetos con  
Amazon S3  
10 minutos

**Video:** Elija el servicio de  
almacenamiento adecuado  
6 minutos

**Lectura:** Lectura 3.4: Elija el  
servicio de almacenamiento  
adecuado  
25 minutos

**Video:** Introducción al ejercicio 5  
2 minutos

**Lectura:** Ejercicio 5: Cree un  
depósito de Amazon S3  
30 minutos

**Cuadro de aviso de la discusión:**  
Discussion - Serverless  
10 minutos

**Video:** Demostración Creación de  
un depósito de Amazon S3  
9 minutos

Bases de datos en AWS

Semana 3 Ejercicio y Evaluación

## Lectura 3.3: Almacenamiento de objetos con Amazon S3

### ¿QUÉ ES AMAZON S3?

A diferencia de Amazon EBS, Amazon S3 es una solución de almacenamiento independiente que no está vinculada a la informática. Le permite recuperar sus datos desde cualquier lugar en la web. Si alguna vez usó un servicio de almacenamiento en línea para hacer una copia de seguridad de los datos de su máquina local, lo más probable es que haya usado un servicio similar a Amazon S3. La gran diferencia entre esos servicios de almacenamiento en línea y Amazon S3 es el tipo de almacenamiento.

Amazon S3 es un *servicio de almacenamiento de objetos*. El almacenamiento de objetos almacena datos en una estructura plana, utilizando identificadores únicos para buscar objetos cuando se solicitan. Un objeto es simplemente un archivo combinado con metadatos y puede almacenar tantos de estos objetos como desee. Todas estas características del almacenamiento de objetos también son características de Amazon S3.

### COMPRENDA LOS CONCEPTOS DE AMAZON S3

En Amazon S3, debe almacenar sus objetos en contenedores llamados **cubos**. No puede cargar ningún objeto, ni siquiera una sola foto, en S3 sin crear primero un cubo. Cuando crea un depósito, elige, como mínimo, dos cosas: el nombre del depósito y la región de AWS en la que desea que resida el depósito.

La primera parte es elegir la **región** en la que desea que resida el depósito. Por lo general, esta será una región que haya usado para otros recursos, como su computación. Cuando elige una región para su depósito, todos los objetos que coloca dentro de ese depósito se almacenan de forma redundante en varios dispositivos, en varias zonas de disponibilidad. Este nivel de redundancia está diseñado para proporcionar a los clientes de Amazon S3 una durabilidad del 99,99999999 % y una disponibilidad del 99,99 % para los objetos durante un año determinado.

La segunda parte es elegir un **nombre de depósito** que debe ser único en todas las cuentas de AWS. AWS le impide elegir un nombre de depósito que ya haya elegido otra persona en otra cuenta de AWS. Una vez que elige un nombre, ese nombre es suyo y nadie más puede reclamarlo a menos que elimine ese depósito, que luego libera el nombre para que otros lo usen.

AWS utiliza este nombre como parte del identificador del objeto. En S3, cada objeto se identifica mediante una URL, que se ve así:

Después de `http://`, verá el nombre del depósito. En este ejemplo, el depósito se llama *doc*. Luego, el identificador usa el nombre del servicio, *s3* y especifica el proveedor de servicios *amazonaws*. Después de eso, tiene una carpeta implícita dentro del depósito llamada *2006-03-01* y el objeto dentro de la carpeta llamado *AmazonS3.html*. El nombre del objeto a menudo se conoce como *el nombre de la clave*.

Tenga en cuenta que puede tener carpetas dentro de los cubos para ayudarlo a organizar los objetos. Sin embargo, recuerde que no existe una jerarquía de archivos real que admita esto en el back-end. En cambio, es una estructura plana donde todos los archivos y carpetas viven en el mismo nivel. El uso de cubos y carpetas implica una jerarquía, lo que lo hace fácil de entender para el ojo humano.

### CASOS DE USO DE S3

Amazon S3 es uno de los servicios de almacenamiento más utilizados, con muchos más casos de uso de los que caben en una pantalla. La siguiente lista resume algunas de las formas más comunes en las que puede utilizar Amazon S3.

- **Copia de seguridad y almacenamiento:** S3 es un lugar natural para realizar copias de seguridad de archivos porque es muy redundante. Como se mencionó en la última unidad, AWS almacena sus instantáneas de EBS en S3 para aprovechar su alta disponibilidad.
- **Alojamiento de medios:** debido a que puede almacenar objetos ilimitados, y cada objeto individual puede tener hasta 5 TB, S3 es una ubicación ideal para alojar cargas de videos, fotos o música.
- **Entrega de software:** puede usar S3 para alojar sus aplicaciones de software que los clientes pueden descargar.
- **Lagos de datos:** S3 es una base óptima para un lago de datos debido a su escalabilidad prácticamente ilimitada. Puede aumentar el almacenamiento de gigabytes a petabytes de contenido, pagando solo por lo que usa.
- **Sitios web estáticos:** puede configurar su depósito para alojar un sitio web estático de HTML, CSS y secuencias de comandos del lado del cliente.
- **Contenido estático:** debido al escalado ilimitado, la compatibilidad con archivos grandes y el hecho de que puede acceder a cualquier objeto a través de la web en cualquier momento, S3 es el lugar perfecto para almacenar contenido estático.

### ELIGE LA OPCIÓN DE CONECTIVIDAD CORRECTA PARA TUS RECURSOS

Todo en Amazon S3 es privado de forma predeterminada. Esto significa que solo el usuario o la cuenta de AWS que creó ese recurso puede ver todos los recursos de S3, como depósitos, carpetas y objetos. Para empezar, todos los recursos de Amazon S3 son privados y están protegidos.

Si decide que desea que todos los usuarios de Internet vean sus fotos, puede optar por hacer públicos sus cubos, carpetas y objetos. Tenga en cuenta que un recurso público significa que todo el mundo en Internet puede verlo. La mayoría de las veces, no desea que sus permisos sean de todo o nada. Por lo general, desea ser más granular sobre la forma en que brinda acceso a sus recursos.

Para ser más específicos acerca de quién puede hacer qué con sus recursos de S3, Amazon S3 ofrece dos funciones

principales de administración de acceso: políticas de IAM y políticas de depósito de S3.

## COMPRENDER LAS POLÍTICAS DE IAM

Anteriormente, aprendió sobre la creación y el uso de políticas de IAM y ahora puede aplicar esto a Amazon S3. Cuando las políticas de IAM se adjuntan a usuarios, grupos y roles de IAM, las políticas definen qué acciones pueden realizar. Las políticas de IAM no están vinculadas a ningún servicio de AWS y se pueden usar para definir el acceso a casi cualquier acción de AWS. Debe usar políticas de IAM para depósitos privados cuando:

- Tiene muchos cubos con diferentes requisitos de permisos. En lugar de definir muchas políticas de depósito de S3 diferentes, puede usar políticas de IAM en su lugar.
- Desea que todas las políticas estén en una ubicación centralizada. El uso de políticas de IAM le permite administrar toda la información de políticas en una ubicación.

## COMPRENDA LAS POLÍTICAS DE DEPÓSITO DE S3

Las políticas de depósito de S3 son similares a las políticas de IAM, ya que ambas se definen con el mismo lenguaje de políticas en formato JSON. La diferencia es que las políticas de IAM se adjuntan a usuarios, grupos y roles, mientras que las políticas de depósito de S3 solo se adjuntan a depósitos. Las políticas de depósito de S3 especifican qué acciones se permiten o deniegan en el depósito.

Por ejemplo, si tiene un depósito llamado `employeebucket`, puede adjuntarle una política de depósito de S3 que permita que otra cuenta de AWS coloque objetos en ese depósito.

O si desea permitir que los espectadores anónimos lean los objetos en el depósito de empleados, puede aplicar una política a ese depósito que permita a cualquier persona leer objetos en el depósito usando "Efecto": *Permitir* en la "Acción": `"s3:GetObject"`.

Este es un ejemplo de cómo se vería esa política de depósito de S3.

```
{
  "Versión": "2012-10-17",
  "Declaración": {
    {
      "Sid": "Lectura pública",
      "Efecto": "Permitir",
      "Principal": "*",
      "Acción": ["s3:ObtenerObjeto"],
      "Recurso": ["arn:aws:s3:::employeebucket/*"]
    }
  ]
}
```

Las políticas de depósito de S3 solo se pueden colocar en depósitos y no se pueden usar para carpetas u objetos. Sin embargo, la política que se coloca en el depósito se aplica a todos los objetos de ese depósito. Debe usar las políticas de depósito de S3 cuando:

- Necesita una forma sencilla de acceder a S3 entre cuentas, sin utilizar roles de IAM.
- Sus políticas de IAM superan el límite de tamaño definido. Las políticas de depósito de S3 tienen un límite de tamaño mayor.

## CIFRAR S3

Amazon S3 refuerza el cifrado en tránsito (mientras viaja hacia y desde Amazon S3) y en reposo. Para proteger los datos en reposo, puede utilizar:

- **Cifrado del lado del servidor:** esto le permite a Amazon S3 cifrar su objeto antes de guardarlo en discos en sus centros de datos y luego descifrarlo cuando descarga los objetos.
- **Cifrado del lado del cliente:** cifre sus datos del lado del cliente y cargue los datos cifrados en Amazon S3. En este caso, usted administra el proceso de encriptación, las claves de encriptación y todas las herramientas relacionadas.

Para cifrar en tránsito, puede utilizar el cifrado del lado del cliente o la Capa de sockets seguros (SSL).

## USAR VERSIONES PARA PRESERVAR OBJETOS

Como sabe, Amazon S3 identifica objetos en parte mediante el nombre del objeto. Por ejemplo, cuando carga la foto de un empleado en S3, puede nombrar el objeto `empleado.jpg` y almacenarlo en una carpeta llamada `empleados`. Si no utiliza el control de versiones de Amazon S3, cada vez que carga un objeto denominado `employee.jpg` en la carpeta de empleados, se sobrescribe el archivo original. Esto puede ser un problema por varias razones.

- `empleado.jpg` es un nombre común para un objeto de foto de empleado. Es posible que usted u otra persona que tenga acceso a ese depósito no haya tenido la intención de sobrescribirlo y, ahora que lo ha hecho, ya no tiene acceso al archivo original.
- Es posible que desee conservar diferentes versiones de `employee.jpg`. Sin control de versiones, si desea crear una nueva versión de `employee.jpg`, debe cargar el objeto y elegir un nombre diferente para él. Tener varios objetos, todos con ligeras diferencias en las variaciones de nombres, puede causar confusión y desorden en su depósito.

¿Entonces, Qué haces? ¿Usas el control de versiones de S3! El control de versiones le permite mantener varias versiones de un solo objeto en el mismo depósito. Esto le permite conservar las versiones antiguas de un objeto sin tener que usar construcciones de nombres diferentes, en caso de que necesite recuperarse de eliminaciones accidentales, sobrescrituras accidentales o incluso fallas de la aplicación. Veamos cómo funciona esto.

Si habilita el control de versiones para un depósito, Amazon S3 genera automáticamente un ID de versión único para el objeto que se almacena. En un cubo, por ejemplo, puede tener dos objetos con la misma clave, pero diferentes ID de versión, como employeephoto.gif (versión 111111) y employeephoto.gif (versión 121212). Los cubos habilitados para el control de versiones le permiten recuperar objetos de una eliminación o sobrescritura accidentales.

- Eliminar un objeto no elimina el objeto de forma permanente. En su lugar, Amazon S3 coloca un marcador en el objeto que muestra que intentó eliminarlo. Si desea restaurar el objeto, puede eliminar este marcador y restablecer el objeto.
- Si sobrescribe un objeto, se genera una nueva versión del objeto en el depósito. Todavía tiene acceso a versiones anteriores del objeto.

#### COMPRENSIÓN DE LOS ESTADOS DE VERSIÓN

Los cubos pueden estar en uno de tres estados.

- Sin versión (predeterminado): ningún objeto nuevo o existente en el depósito tiene una versión.
- Control de versiones habilitado: Esto habilita el control de versiones para todos los objetos en el depósito.
- Control de versiones suspendido: Esto suspende el control de versiones para nuevos objetos. Todos los objetos nuevos en el depósito no tendrán una versión. Sin embargo, todos los objetos existentes mantienen sus versiones de objeto.

El estado de control de versiones se aplica a todos los objetos de ese depósito. Tenga en cuenta que se incurre en costos de almacenamiento para todos los objetos en su depósito y todas las versiones de esos objetos. Para reducir su factura de S3, es posible que desee eliminar las versiones anteriores de sus objetos que ya no están en uso.

#### ¿CUÁLES SON LAS CLASES DE ALMACENAMIENTO DE AMAZON S3?

Cuando carga un objeto en Amazon S3 y no especifica la clase de almacenamiento, lo está cargando en la clase de almacenamiento predeterminada, a menudo denominada almacenamiento estándar. Cuando aprendió sobre Amazon S3 en unidades anteriores, ¡estaba aprendiendo sobre la clase de almacenamiento estándar sin siquiera saberlo! Las clases de almacenamiento de S3 le permiten cambiar su nivel de almacenamiento a medida que cambian las características de sus datos. Por ejemplo, si ahora accede a sus fotos antiguas con poca frecuencia, es posible que desee cambiar la clase de almacenamiento en la que se almacenan esas fotos para ahorrar costos. Hay seis clases de almacenamiento S3.

1. **Amazon S3 Standard:** se considera almacenamiento de propósito general para aplicaciones en la nube, sitios web dinámicos, distribución de contenido, aplicaciones móviles y de juegos, y análisis de big data.
2. **Amazon S3 Intelligent-Tiering:** este nivel es útil si sus datos tienen patrones de acceso desconocidos o cambiantes. S3 Intelligent-Tiering almacena objetos en dos niveles, un nivel de acceso frecuente y un nivel de acceso poco frecuente. Amazon S3 monitorea los patrones de acceso de sus datos y mueve automáticamente sus datos al nivel de almacenamiento más rentable según la frecuencia de acceso.
3. **Amazon S3 Standard-Infrequent Access (S3 Standard-IA):** S3 Standard-IA es para datos a los que se accede con menos frecuencia, pero que requieren un acceso rápido cuando es necesario. S3 Standard-IA ofrece la alta durabilidad, el alto rendimiento y la baja latencia de S3 Standard, con un precio de almacenamiento por GB y una tarifa de recuperación por GB bajos. Este nivel de almacenamiento es ideal si desea almacenar copias de seguridad a largo plazo, archivos de recuperación ante desastres, etc.
4. **Amazon S3 One Zone-Accesso poco frecuente (S3 One Zone-IA):** a diferencia de otras clases de almacenamiento de S3 que almacenan datos en un mínimo de tres zonas de disponibilidad (AZ), S3 One Zone-IA almacena datos en una sola AZ y cuesta un 20 % menos que S3 Estándar-IA. S3 One Zone-IA es ideal para los clientes que desean una opción de menor costo para los datos a los que se accede con poca frecuencia, pero que no requieren la disponibilidad y la resistencia de S3 Standard o S3 Standard-IA. Es una buena opción para almacenar copias de seguridad secundarias de datos locales o datos fácilmente recreables.
5. **Amazon S3 Glacier Instant Retrieval:** Amazon S3 Glacier Instant Retrieval es una clase de almacenamiento de archivo que ofrece el almacenamiento de menor costo para datos de larga duración a los que rara vez se accede y que requieren recuperación en milisegundos.
6. **Recuperación flexible de Amazon S3 Glacier:** S3 Glacier Flexible Retrieval ofrece almacenamiento de bajo costo, hasta un 10 % más bajo (que S3 Glacier Instant Retrieval), para datos de archivo a los que se accede 1 o 2 veces al año y se recuperan de forma asíncrona.
7. **Amazon S3 Glacier Deep Archive:** S3 Glacier Deep Archive es la clase de almacenamiento de menor costo de Amazon S3 y admite la retención a largo plazo y la preservación digital de datos a los que se puede acceder una o dos veces al año. Está diseñado para clientes, en particular aquellos en industrias altamente reguladas, como los sectores de servicios financieros, atención médica y público, que retienen conjuntos de datos durante 7 a 10 años o más para cumplir con los requisitos de cumplimiento normativo.
8. **Amazon S3 Outposts:** Amazon S3 en Outposts ofrece almacenamiento de objetos a su entorno local de AWS Outposts.

#### AUTOMATICE LAS TRANSICIONES DE NIVELES CON LA GESTIÓN DEL CICLO DE VIDA DE LOS OBJETOS

Si sigue cambiando manualmente sus objetos, como las fotos de sus empleados, de un nivel de almacenamiento a otro, es posible que desee considerar la automatización de este proceso mediante una política de ciclo de vida. Cuando define una configuración de política de ciclo de vida para un objeto o grupo de objetos, puede optar por automatizar dos acciones: acciones de transición y caducidad.

- **Las acciones de transición** se utilizan para definir cuándo debe realizar la transición de sus objetos a otra clase de almacenamiento.
- **Las acciones de caducidad** definen cuándo caducan los objetos y deben eliminarse de forma permanente.

Por ejemplo, puede optar por hacer la transición de objetos a la clase de almacenamiento S3 Standard-IA 30 días después de haberlos creado, o archivar objetos en la clase de almacenamiento S3 Glacier un año después de crearlos.

Los siguientes casos de uso son buenos candidatos para la gestión del ciclo de vida.

- **Registros periódicos:** si carga registros periódicos en un depósito, es posible que su aplicación los necesite durante una semana o un mes. Después de eso, es posible que desee eliminarlos.
- **Datos que cambian en la frecuencia de acceso:** algunos documentos se acceden con frecuencia durante un

período de tiempo limitado. Después de eso, se accede a ellos con poca frecuencia. En algún momento, es posible que no necesite acceder a ellos en tiempo real, pero su organización o las regulaciones pueden requerir que los archive durante un período específico. Después de eso, puede eliminarlos.

---