Jonathan A nderson

AWS

Certified Solutions Architect

Associate

The ultimate guide for the

SAA-C03

Imagen que contiene lego, juguete

Descripción generada automáticamenteE X A M

# Chapter 2

**Easy Storage Service from Amazon (Amazon S3) and Storage for Amazon Glacier**

Este capítulo le ayudará a comprender Amazon Simple Storage Service (Amazon S3) y Amazon Glacier, los principales servicios de almacenamiento de objetos disponibles en AWS.

Amazon S3 ofrece almacenamiento en la nube estable, fiable y extremadamente escalable para desarrolladores y equipos de TI. Con una interfaz de servicio web sencilla, Amazon S3 es un almacenamiento de objetos fácil de usar que se puede utilizar para almacenar y recuperar cualquier cantidad de datos, en cualquier lugar de Internet. Amazon S3 también le permite pagar solo por el almacenamiento que utiliza, lo que elimina la planificación de la capacidad y las restricciones de capacidad asociadas con el almacenamiento tradicional.

Amazon S3 fue uno de los primeros servicios introducidos por AWS. Sirve como uno de los servicios web más importantes: casi todas las aplicaciones se ejecutan en AWS utilizando Amazon S3, ya sea directa o indirectamente. Amazon S3 se puede utilizar solo o en combinación con otros servicios de AWS, al igual que con muchos otros servicios en la nube de AWS, ya que proporciona una amplia gama de integración. Por ejemplo, para las instantáneas de Amazon Kinesis y Amazon Elastic MapReduce (Amazon EMR), Amazon S3 sirve como almacenamiento de destino duradero y se utiliza como almacenamiento de instantáneas. para Amazon Elastic Block Store (Amazon EBS) y Amazon Relational Database Service (Amazon RDS). Además, se utiliza como plataforma de almacenamiento de datos o mecanismo de carga para Amazon Redshift y Amazon DynamoDB, entre muchas otras funciones. Debido a la flexibilidad de Amazon S3 y su alto nivel de integración y usabilidad, es importante comprender el servicio en detalle.

Los escenarios de uso comunes para el almacenamiento de Amazon S3 son:

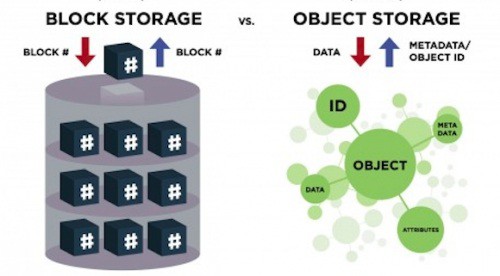
* 1. Copia de seguridad y archivado de datos en las instalaciones o en la nube
  2. Almacenamiento y distribución de contenido, medios y software
  3. Análisis de Big Data
  4. Alojamiento de sitios web estáticos
  5. Alojamiento de aplicaciones móviles e internet nativas en la nube
  6. Recuperación ante desastres

Para admitir estas y más opciones de uso, Amazon S3 ofrece una gama de clases de almacenamiento diseñadas para una variedad de escenarios de uso común: uso general, acceso poco frecuente y archivado. Para ayudar a administrar los datos a lo largo de su ciclo de vida, Amazon S3 ofrece políticas de ciclo de vida configurables. Mediante el uso de estas directivas de ciclo de vida, puede migrar automáticamente los datos a la clase de almacenamiento más adecuada, sin cambiar el código de la aplicación. Para controlar quién puede acceder a sus datos, Amazon S3 ofrece un conjunto completo de permisos, controles de acceso y opciones de cifrado.

Amazon Glacier es otro servicio de almacenamiento en la nube vinculado a Amazon S3 pero optimizado para el archivado de datos y la copia de seguridad a largo plazo a un costo extremadamente bajo. Amazon Glacier es adecuado para "datos fríos", es decir, datos que no se solicitan con frecuencia y para los que es aceptable un tiempo de recuperación de tres a cinco horas. Amazon Glacier se puede utilizar como clase de almacenamiento de Amazon S3 (consulte Clases de almacenamiento y Administración del ciclo de vida de objetos en la sección Características avanzadas de Amazon S3) y como un servicio de almacenamiento de archivado independiente (consulte la sección Amazon Glacier).

**Traditional Block and File Storage versus Object Storage**

En los entornos informáticos tradicionales, dominan dos tipos de almacenamiento: el almacenamiento masivo y el almacenamiento de archivos. El almacenamiento en bloques opera a un nivel inferior, el nivel de dispositivo de almacenamiento en bruto, y administra los datos como un conjunto de bloques numerados de tamaño fijo. El almacenamiento de archivos funciona a un nivel superior (nivel del sistema operativo) y manejar datos como una jerarquía de archivos y carpetas con nombre. El almacenamiento de archivos y bloques a menudo es accesible en red en forma de una red de área de almacenamiento (SAN) para almacenamiento en bloque, utilizando protocolos como iSCSI o Fiber Channel, o como un servidor de archivos de almacenamiento conectado en red (NAS) o "archivador" para el almacenamiento de archivos, utilizando protocolos como Common Internet File System (CIFS) o Network File System (NFS). Ya sea directamente conectado o conectado a la red, masivo o archivo, este tipo de almacenamiento está estrechamente relacionado con el servidor y el sistema operativo que utiliza el almacenamiento.



Guardar objetos de Amazon S3 es otra forma. Amazon S3 se utiliza para el almacenamiento de objetos en la nube. En lugar de estar estrechamente vinculado a un servidor, el almacenamiento de Amazon S3 es independiente del servidor y accesible desde Internet. En lugar de administrar datos como bloques o archivos con protocolos SCSI, CIFS o NFS, los datos se administran como objetos mediante una interfaz de programa de aplicación (API) basada en verbos HTTP estándar.

Tanto los datos como los metadatos se pueden encontrar en cualquier objeto de Amazon S3. Los objetos residen en contenedores denominados buckets y una clave específica especificada por el usuario (nombre de archivo) identifica cada objeto. Los buckets son carpetas planas simples sin jerarquía en el sistema de archivos. Puede poseer varios cubos, pero no puede poseer un subgacharón. Un número infinito de cosas pueden llevar cualquier cubo.

Un objeto de Amazon S3 es fácil de ilustrar, como un archivo y la clave como nombre de archivo. Sin embargo, tenga en cuenta que Amazon S3 no es un sistema de archivos convencional y varía sustancialmente. OBTIENE un objeto o Coloca un objeto en Amazon S3, entonces, puede trabajar en todo el objeto a la vez en lugar de actualizar incrementalmente partes del objeto como lo haría con un archivo. No puede "montar" un bucket, "abrir" un objeto, instalar un sistema operativo en Amazon S3 o ejecutar una base de datos en él.

En lugar de un sistema de archivos, Amazon S3 es un almacenamiento de objetos muy duradero y altamente escalable, optimizado para la lectura y diseñado con un conjunto de características intencionalmente minimalista. Proporciona una abstracción de almacenamiento de archivos simple y robusta que lo libera de muchos de los detalles subyacentes que normalmente encontraría en el almacenamiento tradicional.

No tiene que preocuparse por la durabilidad o la replicación de datos entre zonas de disponibilidad: los objetos de Amazon S3 se replican automáticamente en varios dispositivos en varias instalaciones de la misma región. Lo mismo ocurre con la escalabilidad: si la velocidad de sus solicitudes aumenta constantemente, Amazon S3 distribuye automáticamente buckets para admitir velocidades de solicitud muy rápidas y también permite el acceso simultáneo de muchos clientes.

AWS proporciona opciones para un almacenamiento de archivos o bloques convencional además del almacenamiento de Amazon S3. El servicio Amazon EBS proporciona almacenamiento a nivel de bloque para instancias de Amazon Elastic Compute Cloud (Amazon EC2). Amazon Elastic File System (AWS EFS) proporciona almacenamiento de archivos compartido en red (almacenamiento NAS) mediante el protocolo NFS v4.

**Amazon Simple Storage Service (Amazon S3) Basics**

Ahora que tiene la comprensión básica de algunas de las principales diferencias entre dos almacenamientos simples en la nube, exploremos los conceptos básicos de Amazon S3 con más detalle.

## Buckets

Un bucket es un contenedor (carpeta web) almacenado en Amazon S3 para objetos (archivos). Un bucket normalmente consta de un objeto de Amazon S3. Para Amazon S3, los buckets son el espacio de nombres de nivel superior y los nombres de bucket siempre son globales. Esto implica que, al igual que los nombres de dominio del Sistema de nombres de dominio (DNS), y no solo dentro de su cuenta, los nombres de su bucket deben ser idénticos a todas las cuentas de AWS. Se pueden usar hasta 63 números, letras minúsculas, guiones y puntos para crear un bucket de Nombres. Puede crear y usar múltiples compartimentos y también puede tener hasta 100 por cuenta estándar.

Se recomienda encarecidamente utilizar nombres de bucket que contengan su nombre de dominio y que también cumplan con las reglas de nombres DNS. Esto garantiza que los nombres de su bucket sean suyos, se puedan usar en todas las regiones y también puedan alojar sitios web estáticos.

## AWS Regions

Aunque el espacio de nombres del bucket de Amazon S3 es global, cada bucket de Amazon S3 se crea en una región específica de su elección. Esto le permite monitorear dónde almacena sus datos. Para reducir la latencia, puede crear y usar buckets que estén cerca de una colección específica de usuarios finales o clientes, o en una región determinada para abordar problemas de ubicación y soberanía de datos, o lejos de sus instalaciones principales para habilitar la recuperación de cumplimiento ante desastres y las necesidades de cumplimiento. También puede determinar la ubicación de sus datos; los datos de un bucket de Amazon S3 se pueden almacenar en esa región a menos que los copie explícitamente en otro bucket de una región diferente.

**Objects**

En los contenedores de Amazon S3, los objetos son las entidades o archivos almacenados. En cualquier formato, un objeto puede almacenar casi cualquier tipo de datos. Los tamaños de los elementos pueden variar de 0 a 5 TB bytes, y se puede encontrar un número infinito de objetos en un solo bucket. Esto implica que Amazon S3 puede guardar una cantidad casi infinita de datos.

Además, un objeto se compone de datos (el archivo en sí) y metadatos (los datos del archivo). La sección de datos de un objeto de Amazon S3 es invisible para Amazon S3. Esto significa que los datos de un objeto se pueden tratar como un único flujo de bytes. Amazon S3 no sabe ni se preocupa por el tipo de datos que está almacenando, y el servicio no funciona de manera diferente para los datos de texto y los datos binarios.

Una lista de pares nombre/valor que se pueden utilizar para identificar el objeto es similar a los metadatos asociados a un objeto de Amazon S3. Hay dos tipos de metadatos: metadatos del dispositivo y metadatos del usuario. Amazon S3 genera y utiliza metadatos de dispositivos e incluye elementos como la fecha de la última actualización, el tamaño del objeto, el resumen MD5 y una forma de contenido HTTP. Los metadatos de usuario son opcionales y se pueden definir cuando se crea un objeto. Puede etiquetar los datos con atributos que sean relevantes para usted mediante metadatos personalizados.

## Keys

Un identificador único denominado clave identifica cada objeto almacenado en un bucket de S3. La clave se puede considerar como un nombre de archivo. Una clave puede contener hasta 1024 bytes de caracteres Unicode UTF-8, incluidas barras diagonales integradas, barras diagonales inversas, puntos y guiones.

Las llaves deben ser únicas dentro del mismo compartimento, pero los diferentes compartimentos pueden contener una llave similar. La combinación de bucket, clave e ID de versión opcional se puede utilizar para identificar de forma exclusiva un objeto de Amazon S3.

## Object URL

Amazon S3 es un almacenamiento para web y cada objeto de Amazon S3 se puede abordar mediante una URL única que es el punto de enlace de los servicios web, el nombre del bucket, la clave y el objeto. Por ejemplo, con la URL:

* 1. HTTP:/mybucket.s3.amazonaws.com / jack.doc Mybucket es el nombre del bucket de S3 y el nombre principal o de archivo es jack.doc. Si se crea un objeto diferente, como:
  2. HTTP:/mybucket.s3.amazonaws.com / fee / fi / fog / smoke / jack.doc.com entonces, el nombre del bucket sigue siendo mybucket pero ahora la clave o nombre de archivo es string.
  3. fresco / fi / fo / fum / jack.doc. Una clave puede contener caracteres delimitadores, como barras diagonales o barras diagonales inversas, que pueden ayudarle a nombrar y organizar lógicamente sus objetos de Amazon S3, pero para Amazon S3, implica solo un largo nombre de clave en un espacio de nombres plano. No existe una jerarquía real de archivos y carpetas. Para obtener más información, consulte el tema "Prefijos y separadores" en la sección "Características avanzadas de Amazon S3" que se muestra a continuación.

Para mayor comodidad, la consola de Amazon S3 y la característica Prefijo y separador le permiten navegar por un bucket de Amazon S3 como si estuvieran organizados en la jerarquía de carpetas. Tenga en cuenta, sin embargo, que un bucket con claves no organizadas es un espacio de nombres único y plano.

## Operations of Amazon S3

La API de Amazon S3, con solo un puñado de ediciones comunes, es fácil de usar. Ellos deben incluír:

* 1. Creación/eliminación de un bucket
  2. Escribir una entidad
  3. Lectura de un objeto
  4. Eliminar un elemento
  5. Lista de llaves dentro de un compartimento

**REST Interface**

Una API REST (Representational State Transfer) es la interfaz nativa de Amazon S3. Con la interfaz REST, puede utilizar solicitudes HTTP o HTTPS normales para crear y eliminar buckets, claves de lista y objetos de lectura y escritura. De forma predeterminada, REST asigna "verbos" HTTP (métodos HTTP) a operaciones CRUD típicas (Crear, Leer, Actualizar, Eliminar). Create es HTTP PUT (y a veces POST); la lectura es HTTP GET; delete es HTTP DELETE y la actualización es HTTP POST.

Para garantizar la calidad y garantizar que sus solicitudes y datos estén seguros, utilice siempre HTTPS para las solicitudes de API de Amazon S3.

Los usuarios no aplican explícitamente la interfaz REST en la mayoría de las situaciones, pero se comunican con Amazon S3 a través de una de las interfaces de nivel superior disponibles. Estos incluyen los SDK de AWS (bibliotecas contenedoras) para iOS, Android, JavaScript, Java, .NET, Node.js, PHP, Python, Ruby, Go y C ++, la línea de comandos de interfaz (CLI) y la administración de AWS. Consola.

Amazon S3 originalmente admitía una API de Protocolo simple de acceso a objetos (SOAP) además de la API de REST, pero debe aplicar la API de REST. El antiguo extremo HTTPS sigue estando disponible, pero no se admiten las nuevas características.

**Durability and Availability**

La durabilidad y la disponibilidad de datos están relacionadas, pero son conceptos ligeramente diferentes. La sostenibilidad responde a la pregunta: "¿Mis datos seguirán estando disponibles en el futuro?" La disponibilidad responde a la pregunta: "¿Puedo acceder a mis datos ahora?" Amazon S3 está diseñado para entregar sus datos con un alto grado de confiabilidad y disponibilidad.

En un año determinado, Amazon S3 Standard Storage se crea para una longevidad del 99,9999999999% y un 99,99 % de disponibilidad de artículos. Por ejemplo, si almacena 10.000 objetos de Amazon S3, en promedio, puede esperar la pérdida de un objeto cada 10.000.000 de años. Al almacenar datos automáticamente de forma redundante en varios dispositivos en varias instalaciones y en la misma área, Amazon S3 puede lograr una alta durabilidad. Está diseñado para soportar la pérdida simultánea de datos en dos instalaciones sin perder datos de usuario. Amazon S3 ofrece una infraestructura de almacenamiento altamente duradera diseñada para el almacenamiento de datos críticos y primarios.

Supongamos que desea almacenar datos derivados no críticos o fácilmente reproducibles (como miniaturas de imágenes) que no requieren este alto nivel de durabilidad, puede usar Redundancia reducida Almacenamiento (RRS) por una tarifa baja. RRS ofrece una durabilidad del 99,99 % con un costo de almacenamiento más bajo que el almacenamiento tradicional de Amazon S3.

Aunque el almacenamiento de Amazon S3 ofrece una durabilidad muy alta a nivel de infraestructura, se recomienda proporcionar protección contra la eliminación accidental o la sobrescritura de datos a nivel de usuario aprovechando las características de monitoreo adicionales que incluyen el control de versiones. replicación interregional y eliminación de MFA.

**Data Consistency**

Amazon S3 es, en última instancia, un sistema cohesivo. A medida que los datos se replican automáticamente en varios servidores y ubicaciones de una región, los cambios de datos pueden tardar más tiempo en propagarse a todas las demás ubicaciones. Por lo tanto, en esta situación, la información que lea inmediatamente después de una actualización puede dar lugar a datos obsoletos.

Esto no es un problema en el uso de PUTs para objetos nuevos: en este caso, Amazon S3 ofrece coherencia de lectura después de escritura. Sin embargo, para utilizar PUTs para objetos existentes (sobrescribir objeto a una clave existente) y para objetos DELETE, Amazon S3 proporciona consistencia final.

Cualquier consistencia significa que si coloca nuevos datos en una clave existente, un GET posterior puede devolver los datos antiguos. Del mismo modo, si ELIMINA un objeto, otro GET para ese objeto aún puede leer el objeto eliminado. De cualquier manera, las actualizaciones de clave única son atómicas: para lecturas consistentes hasta el final, puede obtener datos nuevos o datos antiguos, pero nunca una combinación inconsistente de datos.

## Access Control

Amazon S3 está protegido de forma predeterminada; al crear un bucket o un objeto en Amazon S3, solo usted puede tener acceso a él. Para permitirle conceder a otros accesos controlados, Amazon S3 proporciona controles de acceso de grano grueso (listas de control de acceso [ACL] de Amazon S3) y controles de acceso detallados (políticas de bucket de Amazon S3, políticas de AWS Identity and Access Management (IAM) y Autenticación de cadena de consulta).

Las ACL de Amazon S3 le permiten conceder dichos permisos de grano grueso a nivel de objeto o bucket: LECTURA, ESCRITURA o CONTROL TOTAL. Las ACL son un antiguo sistema de control de acceso producido antes de que existiera IAM. Hoy en día, las ACL se utilizan mejor para un número limitado de casos de uso, como habilitar el registro de buckets o crear un bucket que aloje un sitio web estático y legible globalmente.

En lugar de un IAM primario, están conectados al recurso de bucket. Incluyen en la política, una referencia explícita al IAM principal. Este principio se puede asociar a otra cuenta de AWS para que pueda asignar acceso de varias cuentas a los recursos de Amazon S3 mediante la política de bucket de Amazon S3.

Con una política de bucket de Amazon S3, puede decidir quién puede acceder al bucket, desde dónde (por bloque de enrutamiento entre dominios sin clases [CIDR] o dirección IP) y en un momento determinado del día.

Por último, las políticas de IAM se pueden asociar directamente con los IAM principales que conceden acceso a un bucket de Amazon S3, del mismo modo que pueden ofrecer cualquier servicio de AWS y acceso a los recursos. Claramente, solo puede delegar políticas de IAM a los clientes de cuentas de AWS que usted controle.