

# U7-Tema7.5-Servicios de almacenamiento en AWS. Amazon S3

Fundamentos de Computación en la Nube (1ºDAM/1ºDAW)

Álvaro García Sánchez



<b>Introducción a Amazon S3 (Amazon Simple Storage Service,)</b>	<b>2</b>
<b>Características</b>	<b>2</b>
<b>Elementos de un objeto</b>	<b>3</b>
<b>Buckets</b>	<b>3</b>
<b>Acceso a los datos</b>	<b>5</b>
<b>Usos principales</b>	<b>6</b>
<b>S3 Select</b>	<b>6</b>
<b>Seguridad</b>	<b>7</b>
<b>Clases de almacenamiento de Amazon S3</b>	<b>7</b>
<b>Política de ciclo de vida</b>	<b>10</b>
<b>Costes</b>	<b>11</b>
<b>Diferencias entre EBS y S3</b>	<b>12</b>
<b>Referencias</b>	<b>12</b>

# Introducción a *Amazon S3* (Amazon Simple Storage Service,)

**Amazon Simple Storage Service**, más conocido como **S3** (<https://aws.amazon.com/es/s3/>) es un servicio de **almacenamiento persistente de objetos** creado para almacenar y recuperar **cualquier cantidad de datos desde cualquier lugar** mediante una **URL**. Es usado por sitios web, aplicaciones móviles, aplicaciones corporativas, y datos de sensores, dispositivos de Internet de las cosas (IoT), análisis de Big Data, ...

## Características

- S3 es un servicio de almacenamiento **a nivel de objetos**.
- Los **datos se almacenan por lo tanto como objetos** dentro de recursos conocidos como **buckets**. Los objetos pueden ser prácticamente cualquier archivo de datos: texto, imágenes, videos, logs, ..., puede incluso almacenar instantáneas de bases de datos como objetos.
- Es una solución administrada de almacenamiento en la nube que se diseñó para ofrecer un escalado sin problemas, ofreciendo **99,99% de disponibilidad (4 nueves) y 99,99999999% (11 nueves) de durabilidad**.
- Además de poder almacenar prácticamente todos los objetos que deseemos (capacidad de almacenamiento casi ilimitado) dentro de un bucket (cada **objeto puede ocupar hasta un máximo de 5TB**),
- Proporciona **acceso de baja latencia** a los datos
  - A través de Internet u otra red mediante HTTP y/o HTTPS.
  - De forma privada mediante un punto de enlace de una nube virtual privada (VPC).
- De **forma predeterminada, no se comparte ninguno de los datos** públicamente.
- Se puede **cifrar los datos en tránsito** y elegir habilitar el **cifrado del lado del servidor de los objetos**.
- Puede acceder a Amazon S3 por medio de la **Consola de administración de AWS** en la Web de forma **programática con la API** y los **SDK**, o con soluciones de terceros, que utilizan la API o los SDK.
- Los **nombres de los buckets son universales** y deben ser **únicos entre todos los nombres de buckets existentes en Amazon S3**.
- S3 permite realizar **operaciones de escritura, lectura y eliminación de los objetos** almacenados en el bucket.
- De forma predeterminada, en Amazon S3 **los datos se almacenan de forma redundante en varias instalaciones y en diferentes dispositivos de cada instalación**.

- Los datos que almacenamos en S3 no están asociados a ningún servidor en particular (aunque los buckets se asocian a regiones, los archivos se dice que están almacenados de forma global), con lo que no necesitamos administrar ningún tipo de servidor.
- Contiene billones de objetos y, con regularidad, tiene picos de millones de solicitudes por segundo, por ende, es un **servicio de alto rendimiento, con una latencia de primer byte que se mide en milisegundos** para la mayoría de las clases de almacenamiento.
- Incluye **notificaciones de eventos** que le permiten configurar notificaciones automáticas cuando ocurren ciertos eventos, como cuando se carga un objeto a un bucket o se lo borra de un bucket en particular. Esas notificaciones se le pueden enviar o se pueden usar para activar otros procesos, como las funciones de AWS Lambda.



Fuente: <https://aprendiendoaws.com/>

## Elementos de un objeto

Cada objeto que almacenamos en S3 tiene:

- **Clave:** nombre que se le asigna al objeto y que se utiliza para recuperarlo.
- **ID de versión:** se puede mantener un **histórico de cambios mediante el versionado de los archivos**, de manera que cuando se actualice un objeto, en vez de sustituirlo, se crea una nueva versión manteniendo un histórico.
- **Valor:** contenido real que se almacena.
- **Metadatos:** pares clave-valor, que podemos definir nosotros como usuarios.
- **Subrecursos:** que utiliza AWS para almacenar información adicional.

## Buckets

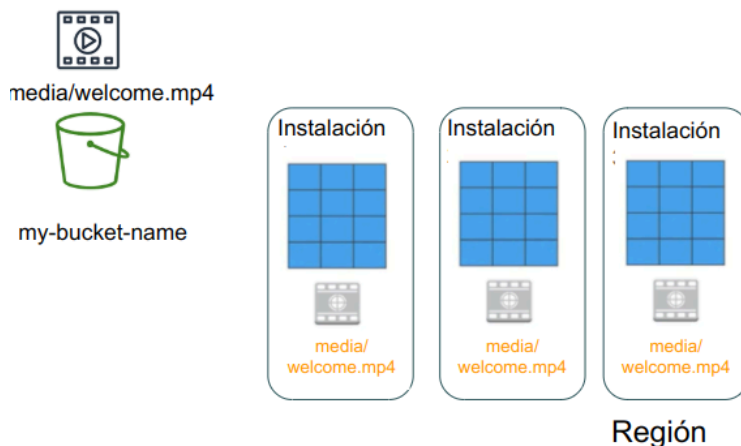
Amazon S3 almacena los datos en **buckets**, los cuales son los bloques básicos donde se estructura la información, actuando como **contenedores lógicos de objetos**.

Los buckets son esencialmente el prefijo de un conjunto de archivos y, como tales,

- **Deben tener un nombre único en todo Amazon S3 a nivel mundial.**
- Es posible **tener uno o más buckets en una cuenta.**
- Se puede controlar el acceso a cada bucket mediante mecanismos de **control de acceso** que pueden aplicarse tanto a objetos individuales como a los buckets, es decir, quién puede crear, eliminar y enumerar objetos en el bucket.
- Se puede **elegir la región geográfica donde Amazon S3 almacenará el bucket y su contenido.** Para cargar los datos (como fotos, vídeos o documentos), primero hay que crear un bucket en una región de AWS y, a continuación, cargar casi cualquier cantidad de objetos en el bucket (los objetos pueden ocupar hasta 5TB).
- Para almacenar un objeto en S3, hay que **cargarlo en un bucket.** Es posible establecer **permisos sobre los datos y cualquier metadato.**
- El servicio S3 permite **almacenar un histórico con todas las versiones de un objeto**, y poder conservar, recuperar y restaurar todas las versiones de los objetos almacenados en un bucket. Para ello, el primer paso es habilitar el **control de versiones para el bucket.**
- Es posible **obtener registros de acceso al bucket y a sus objetos**

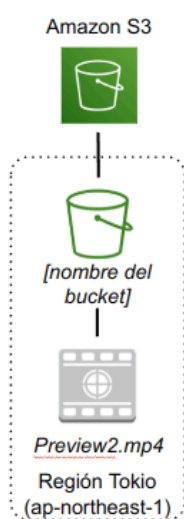
Cuando se crea un bucket en S3, este **se asocia a una región de AWS específica.**

Cuando se almacenan datos en el bucket, **estos se almacenan de forma redundante en varias instalaciones de AWS dentro de la región seleccionada.** S3 está diseñado para almacenar los datos de forma duradera, incluso en el caso de producirse una pérdida de datos simultánea en dos instalaciones de AWS.



Fuente: Curso AWS Cloud Foundations

La **URL de un bucket** se estructura como se muestra en la imagen. Se pueden usar dos estilos distintos de URL para referirse a los buckets.



Para cargar sus datos:

1. Cree un **bucket** en una región de AWS.
2. Cargue casi cualquier cantidad de **objetos** en el bucket.

Punto de enlace de la URL de tipo ruta del bucket:

<https://s3.ap-northeast-1.amazonaws.com/bucket-name>  
Código de región      Nombre del bucket

Punto de enlace de la URL de tipo alojamiento virtual del bucket:

<https://bucket-name.s3-ap-northeast-1.amazonaws.com>  
Nombre del bucket      Código de región

Fuente: Curso AWS Cloud Foundations

- s3-website guion (-) región - <http://bucket-name.s3-website-Region.amazonaws.com>
- s3-web punto (.) Región - <http://bucket-name.s3-website.Region.amazonaws.com>

Amazon S3 administra de **forma automática el almacenamiento que hay detrás del bucket, incluso cuando los datos aumentan**. El almacenamiento de datos aumentará según las necesidades de la aplicación.

Amazon S3 también es escalable, lo que permite manejar un alto volumen de solicitudes. No es necesario que se aprovisione el almacenamiento o el rendimiento, y solo se cobrará por lo que utilice.

## Acceso a los datos

Es posible acceder a Amazon S3 con diferentes opciones.

- Con la consola web de AWS.
- La Interfaz de la línea de comandos de AWS (AWS CLI)
- AWS SDK.
- De forma directa mediante puntos de enlace de REST.

Los puntos de enlace admiten el acceso de HTTP o HTTPS. Como se ha explicado anteriormente para admitir este tipo de acceso según la URL, los **nombres de los buckets de Amazon S3 deben ser únicos de forma global** y deben cumplir con los requisitos del Domain Name Server (DNS, servidor de nombres de dominio).

## Usos principales

La flexibilidad para almacenar una cantidad prácticamente ilimitada de datos y para acceder a ellos desde cualquier lugar convierte a S3 en un servicio adecuado para distintos casos:

- **Como ubicación para cualquier dato de aplicación**, ya sea nuestra propia aplicación hospedada *on-premise*, como las aplicaciones de EC2 o mediante servidores en otros hostings. Esta característica puede resultar útil para los archivos multimedia generados por el usuario, los registros del servidor u otros archivos que su aplicación deba almacenar en una ubicación común. Además, como el contenido se puede obtener de manera directa a través de Internet, podemos delegar la entrega de contenido de nuestra aplicación y permitir que los clientes la consigan ellos mismos.
- **Alojamiento web estático**. S3 puede entregar el contenido estático de un sitio web, que incluye HTML, CSS, JavaScript y otros archivos.
- **Alojamiento multimedia**: cree una infraestructura redundante, escalable y de alta disponibilidad que aloje cargas y descargas de videos, fotos o música
- **Almacenar copias de seguridad de datos**. Para una disponibilidad y capacidad de recuperación de desastres incluso mejores, S3 puede hasta configurarse para admitir la replicación entre regiones, de modo que los datos ubicados en un bucket de S3 en una región puedan replicarse de forma automática en otra región de S3.
- **Área de almacenamiento provisional para big data**.
- **Entrega de software**: alojar las aplicaciones de software para que los clientes puedan descargarlas.

## S3 Select

**Amazon S3 Select** permite utilizar **instrucciones SQL sencillas** para filtrar el contenido de los objetos de Amazon S3 y recuperar exactamente el subconjunto de datos que necesitamos. Si utilizamos S3 Select para filtrar los datos, podemos reducir la cantidad de datos que Amazon transfiere, lo que reduce también los costes y la latencia para recuperarlos:

[https://docs.aws.amazon.com/es\\_es/AmazonS3/latest/userguide/selecting-content-from-objects.html](https://docs.aws.amazon.com/es_es/AmazonS3/latest/userguide/selecting-content-from-objects.html).

Admite los formatos CSV, JSON o Apache Parquet, ya sea en crudo o comprimidos con GZIP o BZIP2 (solo para objetos CSV y JSON), así como objetos cifrados del lado del servidor.

El uso de S3 Select tiene un coste de 0,002\$ por cada GB escaneado, y 0,0007\$ por cada GB devuelto tras ejecutar la consulta.

## Seguridad

**Por defecto, todos los buckets de S3 son privados** y solo pueden acceder los usuarios a los que se les concede acceso explícitamente.

Hay disponibles varias herramientas y posibilidades para controlar el acceso a los buckets u objetos de S3:

- **Uso del bloqueo de acceso público**, la cual anula de manera sencilla cualquier otra política o permisos de objetos.
- **Escritura de políticas IAM** que indiquen los usuarios o roles que pueden acceder a buckets y objetos específicos.
- **Escritura de políticas de bucket** que definan el acceso a buckets u objetos específicos, ya sea para conceder acceso entre cuentas de AWS o para conceder acceso público o anónimo a los datos de Amazon S3. Se utiliza cuando el usuario o el sistema no pueden autenticarse mediante IAM.
- **Creación de puntos de acceso de S3**, los cuales son nombres de alojamiento únicos que aplican permisos y controles de red definidos para las solicitudes que se realizan a través de él. Es posible crear puntos de acceso individualizados con nombres y permisos personalizados para cada aplicación que accede a los datos.
- **Configuración de listas de control de acceso (ACL)**. Se utilizan con menos frecuencia (las ACL preceden a la IAM). Si se utilizan ACL, se recomienda no establecer un acceso muy permisivo.

## Clases de almacenamiento de Amazon S3

Amazon S3 ofrece una variedad de clases de almacenamiento en el nivel de objeto diseñadas para distintos casos de uso. Estas clases incluyen:

- **Amazon S3 Standard:**
  - Amazon S3 Standard está diseñado con el fin de proporcionar almacenamiento de **objetos de alta durabilidad, alta disponibilidad y alto rendimiento para los datos a los que se accede con frecuencia**.
  - Debido a que ofrece **baja latencia y alto rendimiento**, Amazon S3 Standard es adecuado para una amplia variedad de casos prácticos, como **las aplicaciones en la nube, los sitios web dinámicos, la distribución de contenido, las aplicaciones para dispositivos móviles y videojuegos y el análisis de big data**.

- **Amazon S3 Intelligent-Tiering**

- La clase de almacenamiento de Amazon S3 Intelligent-Tiering está diseñada para **optimizar los costos mediante la migración automática de los datos a la capa de acceso más rentable**, sin que esto perjudique el rendimiento ni se produzca una sobrecarga operativa.
- Por un pequeño cargo mensual de supervisión y automatización por objeto, Amazon S3 **supervisa los patrones de acceso** de los objetos de Amazon S3 Intelligent-Tiering y **desplaza a la capa de acceso poco frecuente los objetos a los que no se ha accedido durante 30 días consecutivos**. Si se accede a un objeto en la capa de acceso poco frecuente, este se desplaza automáticamente a la capa de acceso frecuente.
- No hay tarifas por recuperación si se usa la clase de almacenamiento de Amazon S3 Intelligent-Tiering, ni tampoco se cobran tarifas adicionales cuando los objetos se trasladan de una capa de acceso a otra.
- Funciona bien para **datos de larga duración con patrones de acceso desconocidos o impredecibles**.

- **Amazon S3 Standard – Acceso poco frecuente (Amazon S3 Standard-IA):**

- La clase de almacenamiento de Amazon S3 Standard-IA se utiliza **para datos a los que se accede con menos frecuencia, pero que requieren un acceso rápido cuando son necesarios**.
- Está diseñado para ofrecer la alta durabilidad, el alto rendimiento y la baja latencia de Amazon S3 Estándar, con precios de almacenamiento por GB y de recuperación por GB económicos.
- Esta combinación de alto rendimiento y bajo costo hace que Amazon S3 Standard-IA sea la opción ideal para el **almacenamiento y el respaldo a largo plazo, al igual que como almacén de datos para los archivos de recuperación ante desastres**.

- **Amazon S3 One Zone – Acceso poco frecuente (Amazon S3 One Zone-IA):**

- Amazon S3 One Zone-IA es para datos **a los que se accede con menos frecuencia, pero que requieren un acceso rápido cuando son necesarios**. A diferencia de las demás clases de almacenamiento de Amazon S3, que almacenan los datos en un mínimo de tres zonas de disponibilidad, Amazon S3 One Zone-IA **lo hace en una sola zona de disponibilidad** y cuesta menos que Amazon S3 Standard-IA.
- Es apto para los clientes que desean una **opción de menor costo para los datos a los que se accede con poca frecuencia, pero que no necesitan la disponibilidad ni la resiliencia de Amazon S3 Standard o Amazon S3 Standard-IA**.
- Es una buena opción para **almacenar copias de seguridad secundarias** de los datos que se encuentran en las instalaciones o de los datos que se pueden



volver a crear fácilmente. También puede utilizarlo **como almacenamiento rentable para los datos que se replican desde otra región de AWS** mediante la replicación entre regiones de Amazon S3.

- **Amazon S3 Glacier:**

- Amazon S3 Glacier es una clase de almacenamiento **segura, duradera y de bajo costo para el archivado de datos**. Puede almacenar con fiabilidad cualquier cantidad de datos a precios competitivos o más económicos que las soluciones en las instalaciones. Para mantener los costos bajos, pero adecuados para distintas necesidades.
- Amazon S3 Glacier ofrece tres opciones de recuperación que van **desde unos minutos hasta horas**. Puede cargar objetos a Amazon S3 Glacier de manera directa o utilizar las políticas de ciclo de vida de Amazon S3 para transferir datos entre cualquiera de las clases de almacenamiento de Amazon S3 para datos activos (Amazon S3 Standard, Amazon S3 Intelligent-Tiering, Amazon S3 Standard-IA y Amazon S3 One Zone-IA) y Amazon S3 Glacier.

- **Amazon S3 Glacier Deep Archive:**

- Amazon S3 Glacier Deep Archive es la clase de almacenamiento **de menor costo de Amazon S3**. 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 que retienen conjuntos de datos **de 7 a 10 años (o más)** para cumplir con los requisitos de conformidad normativa, como los clientes de sectores altamente regulados, como servicios financieros, sanidad y sectores públicos.
- También se puede utilizar para casos prácticos de respaldos y recuperación ante desastres. Es una alternativa rentable y fácil de administrar en comparación con los sistemas de cinta magnética, ya sea que los sistemas de cinta sean bibliotecas en las instalaciones o servicios fuera de las instalaciones.
- Complementa Amazon S3 Glacier y también está diseñado para proporcionar **once nueves de durabilidad**. Todos los objetos almacenados en Amazon S3 Glacier Deep Archive se reproducen y **almacenan en al menos tres zonas de disponibilidad dispersas geográficamente y estos objetos se pueden restaurar en un plazo de 12 horas**.



Fuente: <https://aprendiendoaws.com/>

Para obtener información sobre las clases de almacenamiento de Amazon S3, consulta <https://docs.aws.amazon.com/AmazonS3/latest/dev/storage-class-intro.html>.

## Política de ciclo de vida

Una política de ciclo de vida define qué va a pasar con los datos partiendo de su almacenamiento masivo en S3 estándar, pasando a uso poco frecuente y seguidamente a Glacier y finalmente para su eliminación, en base a plazos o métricas y reduciendo costes de forma automática.



Fuente: <https://aitor-medrano.github.io/iabd/cloud/aws.html>

Para ello, se puede monitorizar un bucket completo, un prefijo o una etiqueta de objeto, de manera que podamos evaluar los patrones de acceso y ajustar la política de ciclo de vida.

## Costes

Los costes específicos **varían en función de la región y de las solicitudes específicas que se realizan**. Como **regla general, sólo paga por las transferencias que traspasen los límites de una región**, lo que significa que **no tiene que pagar por las transferencias que ingresen a Amazon S3 o salgan de él y se dirijan a ubicaciones perimetrales de Amazon CloudFront dentro de esa misma región**.

- Solo se **paga por lo que se utiliza**, lo que incluye
  - GB por mes.
  - Transferencias de datos salientes **a otras regiones**.
  - Solicitudes PUT, COPY, POST, LIST y GET.
- **No paga** por lo siguiente:
  - Transferencias ENTRANTES a Amazon S3.
  - Transferencias SALIENTES desde Amazon S3 a Amazon CloudFront o Amazon EC2 **dentro de la misma región**

Para calcular los costes de S3 hay que tener en cuenta:

- **1. Clase de almacenamiento y cantidad almacenada:**
  - El almacenamiento estándar está diseñado para proporcionar 99,999.999.999% (11 nueves) de durabilidad y 99,99% (4 nueves) de disponibilidad. Por ejemplo, los primeros 50 TB/mes cuestan 0,023\$ por GB.
  - El almacenamiento Estándar - Acceso poco frecuente ofrece la misma durabilidad de 99,999.999.999% (11 nueves) de S3, pero con 99,9% (3 nueves) de disponibilidad en un año concreto. Su precio parte desde los 0,0125\$ por GB.
  - Y si elegimos el almacenamiento poco frecuente pero en una única zona, el precio pasa a ser de 0,01\$ por GB.
  - Si fuese a la capa Glacier, con una opción de recuperación de 1 minuto a 12 horas el precio baja a 0,004\$ por GB.
  - Finalmente, con Glacier Deep Archive (archivos que se recuperan 1 o 2 veces al año con plazos de recuperación de 12 horas) baja hasta 0,000.99\$ por GB
- **2. Solicitudes:**
  - Se consideran la cantidad y el tipo de solicitudes.
    - GET: recupera un objeto de Amazon S3. Debe tener acceso de LECTURA para utilizar esta operación.
    - PUT: agrega un objeto a un bucket. Debe contar con permisos de ESCRITURA en buckets para agregar un objeto.
    - COPY: crea una copia de un objeto que ya está almacenado en Amazon S3. Una operación COPY equivale a realizar una solicitud GET y luego una PUT.

- Las solicitudes GET generan cargos (0,000.4\$ por cada 1.000 solicitudes) a tasas diferentes de las de otras solicitudes, como PUT y COPY (0,005\$ cada 1.000 solicitudes).
- **3. Transferencia de datos:** se considera la cantidad de datos transferidos fuera de la región de S3, los datos salientes, siendo los primeros 100 GB gratuitos y luego comienza a facturar a 0,09\$ por GB. La transferencia entrante de datos es gratuita.

La información actualizada y detallada se encuentra disponible en <https://aws.amazon.com/es/s3/pricing/>.

## Diferencias entre EBS y S3

- EBS solo se puede utilizar cuando se conecta a una instancia EC2 y se puede acceder a Amazon S3 por sí solo.
- EBS no puede contener tantos datos como S3.
- EBS solo se puede adjuntar a una instancia EC2, mientras que varias instancias EC2 pueden acceder a los datos de un bucket de S3.
- S3 experimenta más retrasos que Amazon EBS al escribir datos.

Es el usuario o el diseñador de la aplicación quien debe decidir si el almacenamiento de Amazon S3 o de Amazon EBS es el más apropiado para una aplicación determinada.

## Referencias

- Curso Academy Cloud Foundation de Amazon Web Services.
- Aitor Medrano. Especialización en Inteligencia Artificial y Big Data (IABD) <https://aitor-medrano.github.io/iabd/>. Licencia CC BY-NC-SA 4.0.
- OpenAI. (2025). Chat GPT.
- <https://aprendiendoaws.com/>