

CONCEPTOS BÁSICOS DE AWS

CONTENIDO

INFRAESTRUCTURA GLOBAL DE AWS.....	2
➤ REGIONES	2
➤ ZONAS DE DISPONIBILIDAD.....	4
➤ UBICACIONES DE BORDE (EDGES LOCATIONS) O UBICACIONES PERIMETRALES	5
ALMACENAMIENTO.....	6
FTP (Protocolo de Transferencia de Archivos)	6
¿Qué es el FTP?	6
Funcionamiento del FTP.....	6
SERVICIOS DE ALMACENAMIENTO EN AWS.....	7
AMAZON S3 - SIMPLE STORAGE SERVICE	7
¿Cómo funciona Amazon S3?.....	8
Clases de almacenamiento de Amazon S3.....	9
S3 Standard	9
S3 Intelligent-Tiering	10
Clases de almacenamiento para objetos a los que se obtiene acceso con poca frecuencia	10
S3 Standard - IA	11
S3 One Zone-IA.....	11
Clases de almacenamiento para el archivado de objetos.....	12
S3 Glacier.....	12
S3 Glacier Deep Archive	12
S3 Outposts	12
AMAZON EBS - ELASTIC BLOCK STORE.....	13
Comparación entre Amazon S3 y Amazon EBS	15
AMAZON EFS - ELASTIC FILE SYSTEM	16
Comparación entre Amazon EBS y Amazon EFS	17
CONCLUSIONES SOBRE LOS SERVICIOS DE ALMACENAMIENTO	18

INFRAESTRUCTURA GLOBAL DE AWS

AWS tiene más de un millón de clientes en 190 países de todo el mundo y atiende a estos clientes a través de su infraestructura global, que consta de 105 zonas de disponibilidad en 33 regiones geográficas y más de 600 POP (Points of Presence - Puntos de Presencia) de CloudFront.

AWS tiene el concepto de una región, que es una ubicación física en todo el mundo donde agrupa los centros de datos. Llama a cada grupo de centros de datos lógicos “zona de disponibilidad”. Cada región de AWS consta de un mínimo de tres zonas de acceso aisladas y físicamente separadas dentro de un área geográfica. A diferencia de otros proveedores de nube, que a menudo definen una región como un solo centro de datos, el diseño múltiple de zonas de disponibilidad de cada región de AWS ofrece ventajas para los clientes. Cada zona de disponibilidad tiene alimentación, refrigeración y seguridad física independientes y está conectada a través de redes redundantes de latencia ultra baja.

Los clientes de AWS centrados en la alta disponibilidad pueden diseñar sus aplicaciones para que se ejecuten en múltiples zonas de disponibilidad y lograr una mayor tolerancia a errores. Las regiones de infraestructura de AWS cumplen con los niveles más altos de seguridad, cumplimiento y protección de datos.

Una región es un área donde AWS brinda servicios en la nube, cuyo propósito es permitir que los usuarios accedan a lugares cercanos y reducir la latencia de la red. Por lo general, varias zonas de disponibilidad en una ciudad forman una región.

A continuación, vamos a ver a detalle cada componente de la infraestructura global de AWS:

➤ REGIONES

Una región de AWS es un área geográfica independiente. Cada región contiene múltiples zonas de disponibilidad. Por ejemplo, la Región Este de EE. UU. (Norte de Virginia) y la Región de Oeste de EE. UU. (Oregón) son dos regiones distintas en AWS. Cada región está diseñada para ser resistente a fallas, lo que significa que, si ocurre un problema en una zona de disponibilidad, las otras zonas seguirán funcionando sin interrupciones.

En AWS una región es una ubicación física en todo el mundo donde agrupamos los centros de datos. Cada región de AWS consta de un mínimo de tres zonas de disponibilidad aisladas y físicamente separadas dentro de un área geográfica.

AWS proporciona una presencia global más extensa que cualquier otro proveedor de nube, y, para respaldar su presencia global y garantizar que los clientes reciban

servicios en todo el mundo, AWS abre nuevas regiones rápidamente. AWS mantiene múltiples regiones geográficas, incluidas las regiones de América del Norte, América del Sur, Europa, China, Asia-Pacífico, Sudáfrica y Medio Oriente.

Dividir los recursos en distintas regiones nos permite:

- Crear nuestra infraestructura lo más cerca posible a nuestros clientes para que puedan acceder a nuestros servicios con una latencia más baja.
- Cumplir con las normativas legales para los proyectos que deben ubicarse dentro de fronteras nacionales.
- Aislar grupos de recursos entre ellos para tener una mejor postura de seguridad.

Al determinar la región adecuada para los servicios, los datos y las aplicaciones, debemos tener en cuenta los siguientes factores:

1. **Cumplimiento con los requisitos legales y de gobernanza de datos:** Dependiendo de la empresa y de la ubicación, es posible que tengamos que ejecutar los datos desde áreas específicas. Por ejemplo, si la empresa requiere que todos los datos residan dentro de los límites del Reino Unido, se elegiría la región de Londres.

No todas las empresas tienen regulaciones de datos en función de la ubicación, por lo que es posible que deban centrarse más en los otros factores.

2. **Proximidad a los clientes:** Seleccionar una región cercana a los clientes nos ayuda a enviarles contenido más rápido. Por ejemplo, si una empresa tiene sede en Washington DC y muchos de los clientes viven en Singapur, podemos considerar la posibilidad de ejecutar la infraestructura en la región del Norte de Virginia para estar cerca de la sede de la empresa y ejecutar las aplicaciones desde la región de Singapur.
3. **Servicios disponibles dentro de una región:** A veces, es posible que la región más cercana no tenga todas las características que deseamos ofrecer a los clientes. AWS innova con frecuencia mediante la creación de nuevos servicios y la ampliación de las características de los servicios existentes. Sin embargo, para que los nuevos servicios estén disponibles en todo el mundo, a veces es necesario que AWS desarrolle hardware físico de una región a la vez.

Imaginemos que nuestros desarrolladores quieren crear una aplicación que utiliza Amazon Braket (la plataforma de computación cuántica de AWS), a hoy, Amazon Braket aún no está disponible en todas las regiones de AWS del mundo, por lo que los desarrolladores tendrían que ejecutarlo en una de las regiones que ya lo ofrecen.

4. **Precios:** Supongamos que estamos considerando ejecutar aplicaciones tanto en Estados Unidos como en Brasil. Debido a la estructura tributaria de Brasil, podría costar un 50 % más ejecutar la misma carga de trabajo en la región de São Paulo en comparación con la región de Oregón (hay varios factores que determinan los precios en las regiones).

El costo de los servicios puede variar de una región a otra.

➤ ZONAS DE DISPONIBILIDAD

Dentro de cada región hay zonas de disponibilidad (AZ). Una AZ consta de dos a seis centros de datos, con fuentes de alimentación redundantes y conectividad de red, estos datacenters no comparten ni el hardware ni la electricidad con otra zona de disponibilidad lo cual permite asegurar que, si alguna AZ falla no interfiera con el funcionamiento de otra, cumpliendo esto con las premisas de alta disponibilidad y tolerancia a fallas.

Una zona de disponibilidad es una ubicación física dentro de una región. Estas zonas están aisladas y separadas para garantizar la alta disponibilidad y la tolerancia a fallas. Por ejemplo, en la Región de Oregón, hay cuatro zonas de disponibilidad. Las aplicaciones y servicios pueden distribuirse en diferentes zonas de disponibilidad para evitar la interrupción completa en caso de problemas en una de las zonas.

Una zona de disponibilidad (AZ) es uno o más centros de datos con alimentación, redes y conectividad redundantes en una región de AWS. Las zonas de disponibilidad permiten que los clientes operen bases de datos y aplicaciones de producción con un nivel de disponibilidad, tolerancia a errores y escalabilidad mayor que el que ofrecería un centro de datos único.

Todas las zonas de disponibilidad en una región de AWS están interconectadas con redes de alto ancho de banda y baja latencia, a través de una fibra metropolitana exclusiva totalmente redundante que proporciona una red de alto rendimiento y baja latencia entre las zonas de disponibilidad. Todo el tráfico entre las AZ está cifrado. El rendimiento de la red es suficiente como para llevar a cabo la replicación sincrónica entre las zonas de disponibilidad. Las AZ facilitan la partición de las aplicaciones para una alta disponibilidad. Si una aplicación se divide en AZ, las empresas estarán mejor aisladas y protegidas de problemas como cortes de energía, rayos, tornados, terremotos, etc. Las AZ están físicamente separadas entre sí por una distancia significativa de muchos kilómetros, aunque todas están dentro de un rango de 100 km (60 millas) de separación.

La latencia entre las zonas de disponibilidad dentro de una región es inferior a un solo dígito, como resultado, podemos replicar de forma sincrónica los datos entre las zonas de disponibilidad.

Una zona de disponibilidad es un centro de datos único o un grupo de centros de datos dentro de una región. Las zonas de disponibilidad están ubicadas a decenas de kilómetros de distancia entre sí. Esto es lo suficientemente cercano para tener una latencia baja (el tiempo que transcurre entre el momento en que se solicita y se recibe el contenido) entre las zonas de disponibilidad. Sin embargo, si algún desastre ocurre en una parte de la región, están lo suficientemente separadas para reducir la posibilidad de que varias zonas de disponibilidad se vean afectadas.

Una buena práctica recomendada por AWS es siempre ejecutar en al menos dos zonas de disponibilidad en una región. Eso significa implementar de manera redundante la infraestructura en dos zonas de disponibilidad diferentes.

➤ UBICACIONES DE BORDE (EDGES LOCATIONS) O UBICACIONES PERIMETRALES

AWS además de sus regiones y zonas de disponibilidad, proporciona ubicaciones de borde que desempeñan un papel crucial en la entrega de contenido y la mejora de la latencia para los usuarios finales.

Las ubicaciones de borde alojan una red de entrega de contenido (CDN: Content delivery network) llamada Amazon CloudFront. Esta red se utiliza para entregar contenido a los usuarios finales de manera rápida y eficiente. Las solicitudes de contenido se dirigen automáticamente a la ubicación de borde más cercana, lo que mejora la velocidad de entrega.

Una ubicación de borde es básicamente una configuración pequeña en diferentes ubicaciones que proporciona conectividad de baja latencia al proporcionar contenido estático para que esté disponible desde la ubicación más cercana de la solicitud. Lo que sucede es que, en lugar de obtener la información de la fuente, simplemente se enruta a la ubicación de borde más cercana y entrega la información, lo que reduce la latencia. De esta forma el tiempo de acceso es menor y la respuesta es más rápida.

Las ubicaciones de borde de AWS son puntos estratégicos en la red de entrega de contenido Amazon CloudFront. Estas ubicaciones permiten almacenar en caché contenido como imágenes, videos y archivos estáticos, lo que mejora la velocidad de carga y reduce la latencia para los usuarios finales.

Cuando se utiliza la red global de ubicaciones de borde, los clientes tienen acceso a una entrega de contenido más rápida, las ubicaciones de borde suelen estar ubicadas en zonas muy pobladas.

En América del Sur, las ubicaciones de borde de AWS incluyen:

Río de Janeiro, Brasil
São Paulo, Brasil
Bogotá, Colombia
Buenos Aires, Argentina
Santiago, Chile

Estas ubicaciones de borde permiten el almacenamiento en caché de contenido y contribuyen a una experiencia de usuario más rápida y eficiente al acceder a recursos alojados en la nube de AWS

En resumen, Amazon CloudFront utiliza estas ubicaciones de borde para distribuir contenido de manera eficiente a través de una red global de servidores. Cuando los usuarios solicitan contenido, se direccionan automáticamente a la ubicación de borde más cercana, lo que garantiza una experiencia de usuario más rápida y eficiente al acceder a recursos alojados en la nube de AWS.

ALMACENAMIENTO

FTP (Protocolo de Transferencia de Archivos)

El FTP o protocolo de transferencia de archivos, es un conjunto de normas que permite transferir archivos directamente entre dispositivos. Aunque ha estado vigente durante más de 50 años, actualmente se utiliza menos debido a alternativas como el almacenamiento en la nube.

¿Qué es el FTP? El FTP es un protocolo que facilita la transferencia de archivos entre distintos sistemas en línea. Puede utilizarse dentro de una red interna de ordenadores o en línea entre servidores web.

Funcionamiento del FTP: Imagina dos ordenadores, A y B. El ordenador A se conecta directamente al ordenador B y puede ver los archivos disponibles para compartir. Luego, puede descargarlos directamente en su propio equipo. Es un método rápido y directo para compartir archivos sin intermediarios.

La nube ha reemplazado en gran medida al FTP porque es un protocolo antiguo y presenta varios problemas. A medida que avanzamos hacia soluciones más modernas, el FTP se ha vuelto menos adecuado para las necesidades actuales de transferencia de archivos.

Algunos de los problemas con el FTP incluyen:

Falta de cifrado: El FTP original no admite cifrado, lo que significa que la comunicación entre el cliente y el servidor está en texto sin cifrar.

Problemas de firewall: Los firewalls modernos bloquean el tráfico que no es HTTPS por defecto, lo que requiere ajustes en las reglas del firewall para permitir el tráfico FTP.

Codificación de caracteres: El FTP no negocia la codificación de caracteres, lo que puede ser un problema para empresas con relaciones comerciales internacionales.

Modo de transferencia: El FTP define modos ASCII y binario para la transferencia de archivos, lo que puede causar problemas de compatibilidad entre sistemas.

Metadatos de archivos: El FTP ignora los metadatos de los archivos, como la fecha de creación y modificación

La nube ofrece una alternativa más segura y versátil para la transferencia de archivos en comparación con el FTP tradicional.

Los servicios de almacenamiento en la nube sincronizan archivos y los almacenan en la nube, además, ofrecen APIs para integración con otros sistemas.

SERVICIOS DE ALMACENAMIENTO EN AWS

AMAZON S3 - SIMPLE STORAGE SERVICE

Amazon Simple Storage Service (Servicio de almacenamiento simple) o Amazon S3, es un servicio de almacenamiento de **objetos** creado para almacenar y recuperar cualquier volumen de datos desde cualquier ubicación. Es un servicio de almacenamiento sencillo que ofrece excelente durabilidad, disponibilidad, rendimiento, seguridad y escalabilidad prácticamente ilimitada a costos muy reducidos.

Amazon S3 proporciona una sencilla interfaz de servicio web que podemos utilizar para almacenar y recuperar el volumen de datos que deseemos, cuando lo necesitemos y desde cualquier ubicación. Es un servicio con un alto nivel de escalabilidad y solo pagamos por lo que usamos.

Podemos cargar cualquier tipo de archivo en Amazon S3, como imágenes, videos, archivos de texto, etc. Por ejemplo, podemos utilizar Amazon S3 para almacenar archivos de copia de seguridad, archivos multimedia de un sitio web o documentos archivados.

Amazon S3 resulta útil para almacenar múltiples tipos de archivos porque es un almacenamiento de datos que permite conservar y recuperar una cantidad prácticamente ilimitada de datos a cualquier escala.

También se ha diseñado para ser muy flexible, nos permite enfocarnos en innovar, en lugar de invertir tiempo en averiguar cómo almacenar nuestros datos.

Es un servicio que proporciona almacenamiento a nivel de objeto, Amazon S3 almacena los datos como objetos en buckets.

¿Cómo funciona Amazon S3?

Amazon S3 almacena datos como objetos en depósitos llamados buckets, podemos almacenar prácticamente todo tipo de datos en cualquier formato: Un objeto consiste en un archivo y los buckets son los contenedores de los objetos.

En otras palabras, los datos se almacenan como objetos, pero en lugar de almacenarlos en un directorio de archivos, los almacenamos en lo que llamamos buckets. Pensemos en un archivo guardado en un disco duro, eso es un objeto; ahora pensemos en un directorio de archivos, es un bucket. El tamaño máximo de objeto que podemos cargar es de cinco terabytes.

Para almacenar un objeto en Amazon S3, cargamos el archivo que deseamos almacenar en un bucket. Cuando cargamos el archivo, podemos:

- Establecer permisos en el objeto para controlar la visibilidad y el acceso a él (quién puede crear, eliminar y listar los objetos en el bucket).
- Utilizar la función de control de versiones de Amazon S3 para realizar un seguimiento de los cambios en sus objetos a lo largo del tiempo, como un rastro documental.
- Protegerlos ante una eliminación accidental.
- Elegir la región geográfica donde Amazon S3 almacenará el bucket y su contenido.

El volumen total de datos y el número de objetos que podemos almacenar en Amazon S3 son ilimitados:

Número de Buckets: Por defecto, podemos crear hasta 100 buckets en cada una de nuestras cuentas de AWS. Si necesitamos más, podemos aumentar la cuota de buckets a un máximo de 1000 buckets mediante una solicitud de aumento de cuota.

Tamaño de Objetos: Los objetos individuales en Amazon S3 pueden variar en tamaño desde un mínimo de 0 bytes hasta un máximo de 5 terabytes. El objeto más grande que se puede cargar en una sola operación PUT (método de transferencia de archivos y protocolos web) es de 5 gigabytes.

Amazon S3 ofrece una durabilidad de datos del 99,999999999 % (11 nueves) de los objetos durante un periodo de un año. Este nivel de durabilidad corresponde a una pérdida anual media prevista del 0,000000001% de los objetos. Por ejemplo, si almacenamos 10.000 objetos con Amazon S3, podríamos esperar incurrir en una pérdida promedio de un objeto cada 10.000.000 años.

Clases de almacenamiento de Amazon S3

Amazon S3 ofrece una variedad de clases de almacenamiento entre las cuales podemos elegir en función de los requisitos de acceso a los datos, la resiliencia y los costos de las 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. Las clases de almacenamiento de S3 son ideales prácticamente para cualquier caso de uso, incluidos los que cuentan con necesidades de rendimiento demandantes, patrones de acceso desconocidos o cambiantes o almacenamiento de archivos.

Con Amazon S3, solo pagamos por lo que utilizamos. Podemos elegir entre varias clases de almacenamiento para seleccionar una opción que se ajuste a las necesidades de negocio y costos. Al seleccionar una clase de almacenamiento de Amazon S3, debemos tener en cuenta estos dos factores: Con qué frecuencia pensamos recuperar los datos y qué tanta disponibilidad deben tener los datos.

Vamos a ver a detalle las clases de almacenamiento de Amazon S3:

S3 Standard: Para Uso general -> Ofrece almacenamiento de objetos de alta durabilidad y rendimiento para datos a los que se obtiene acceso con frecuencia, almacenando datos en un mínimo de tres zonas de disponibilidad. Dada su baja latencia y alto nivel de procesamiento, el tipo S3 Estándar es apropiado para una amplia variedad de casos de uso, como aplicaciones en la nube, sitios web, distribución de contenido, aplicaciones para dispositivos móviles y videojuegos, y el análisis de datos.

Amazon S3 Standard tiene un costo más alto que las otras clases de almacenamiento destinadas a datos de acceso poco frecuente y al almacenamiento en archivos.

Características principales:

- Baja latencia y alto nivel de procesamiento.
- Diseñado para ofrecer un nivel de durabilidad de 99,999999999% de los objetos en varias zonas de disponibilidad.
- Resistente a los eventos que afectan a una zona de disponibilidad completa.
- Admite SSL para datos en tránsito y cifrado de datos en reposo.
- Administración del ciclo de vida de S3 para la migración automática de objetos a otras clases de almacenamiento S3.

También podemos usar las políticas de ciclo de vida de S3 para trasladar automáticamente objetos entre tipos de almacenamiento.

S3 Intelligent-Tiering: (Niveles inteligentes): Para Acceso desconocido -> Es el primer almacenamiento en la nube que reduce automáticamente los costos de almacenamiento a nivel pormenorizado de objeto al mover de manera automática los datos al nivel de acceso más rentable según la frecuencia de acceso, sin afectar el rendimiento, las tarifas de recuperación ni los gastos operativos generales.

Podemos utilizar S3 Intelligent-Tiering como la clase de almacenamiento predeterminada para prácticamente cualquier carga de trabajo, en especial lagos de datos, análisis de datos, aplicaciones nuevas y contenido generado por el usuario.

Es ideal para datos con patrones de acceso desconocidos o cambiantes.

Por un pequeño cargo mensual de monitoreo y automatización de objetos, S3 Intelligent-Tiering monitorea los patrones de acceso y traslada de forma automática los objetos a los que no se accedió a los niveles de acceso de menor costo.

S3 Intelligent-Tiering almacena objetos automáticamente en tres niveles de acceso: Un nivel optimizado para el acceso frecuente en el estándar, un nivel con un costo 40 % menor optimizado para el acceso poco frecuente (glacier) y otro nivel con un costo 68 % menor optimizado para los datos a los que se accede en pocas ocasiones (glacier deep archive).

En la clase de almacenamiento de S3 Intelligent-Tiering, Amazon S3 supervisa los patrones de acceso de los objetos y traslada los objetos a los que no se accedió durante 30 días consecutivos al nivel de acceso poco frecuente (glacier) y, después de 90 días sin acceso, al nivel de acceso instantáneo al archivo (glacier deep archive), finalmente, si se accede a un objeto en el nivel de acceso poco frecuente, Amazon S3 lo traslada automáticamente al nivel de acceso frecuente, S3 Standard.

Características principales:

- El nivel de acceso poco frecuente ahorra hasta un 40 % en costos de almacenamiento.
- El nivel de acceso archivo ahorra hasta un 68 % en costos de almacenamiento.
- Está diseñada para ofrecer un nivel de durabilidad del 99,999999999% de los objetos en varias zonas de disponibilidad.
- Presenta un pequeño cargo mensual por monitoreo y ubicación automática en el nivel adecuado.
- No tiene gastos operativos generales, cargos de ciclo de vida, cargos de recuperación ni duración mínima del almacenamiento.

Clases de almacenamiento para objetos a los que se obtiene acceso con poca frecuencia:

Las clases de almacenamiento S3 Standard-IA y S3 One Zone-IA están diseñadas para datos de larga duración y acceso poco frecuente. IA significa acceso poco frecuente, infrequent access. Los objetos S3 Standard-IA y S3 One Zone-IA están disponibles para acceder en milisegundos (de modo similar a la clase de almacenamiento S3 Standard). Amazon S3 cobra una tarifa de recuperación para estos objetos, por lo que son más adecuados para los datos a los que se obtiene acceso con poca frecuencia.

Por ejemplo, podemos elegir las clases S3 Standard-IA y S3 One Zone-IA para lo siguiente:

- Para almacenar copias de seguridad.
- Para datos más antiguos a los que se obtiene acceso con poca frecuencia, pero que todavía necesitan acceso en milisegundos. Por ejemplo, cuando cargamos datos, podemos elegir la clase de almacenamiento S3 Standard y utilizar la configuración del ciclo de vida para indicar a Amazon S3 que realice la transición de los objetos a la clase S3 Standard-IA o S3 One Zone-IA.

Estas clases de almacenamiento se diferencian en lo siguiente:

S3 Standard - IA: Ideal para datos a los que se accede con poca frecuencia, pero requiere una alta disponibilidad cuando es necesario -> Similar a Amazon S3 Standard, pero tiene un precio de almacenamiento más bajo y un precio de recuperación más alto.

Amazon S3 almacena los datos de los objetos de forma redundante en varias zonas de disponibilidad separadas geográficamente (de forma similar a la clase de almacenamiento S3 Standard). Los objetos S3 Standard-IA resisten a la pérdida de una zona de disponibilidad. Esta clase de almacenamiento ofrece mayor disponibilidad y resistencia que la clase S3 One Zone-IA.

Tanto en Amazon S3 Standard como en Amazon S3 Standard - Acceso poco frecuente, los datos se almacenan en un mínimo de tres zonas de disponibilidad. Amazon S3 Standard IA proporciona el mismo nivel de disponibilidad que Amazon S3 Standard, pero con un precio de almacenamiento más bajo y un precio de recuperación más alto.

S3 One Zone-IA: Amazon S3 almacena los datos de los objetos en una única zona de disponibilidad, lo que resulta más económico que la clase S3 Standard-IA. Sin embargo, los datos no son resistentes a la pérdida física de la zona de disponibilidad como consecuencia de desastres, como terremotos e inundaciones. La clase de almacenamiento S3 One Zone-IA es tan duradera como la clase S3 Standard-IA, pero tiene menor disponibilidad y resistencia.

En comparación con S3 Standard y S3 Standard-IA, que almacenan datos en un mínimo de tres zonas de disponibilidad, S3 One Zone-IA almacena datos en una única zona de disponibilidad. Esto hace que sea una buena clase de almacenamiento a tener en cuenta si se cumplen las siguientes condiciones:

- Deseamos ahorrar costos de almacenamiento.
- Podemos reproducir fácilmente los datos en caso de que se produzca un error en la zona de disponibilidad.

La recomendación es la siguiente:

- **S3 Standard-IA:** Utilizar esta clase de almacenamiento para la copia principal o única de los datos que no se puedan volver a crear.
- **S3 One Zone-IA:** Utilizar esta clase de almacenamiento si podemos volver a crear los datos cuando la zona de disponibilidad produce un error y para las réplicas de objetos cuando se configura la replicación entre regiones de S3.

Clases de almacenamiento para el archivado de objetos:

Las clases de almacenamiento S3 Glacier y S3 Glacier Deep Archive están diseñadas para archivar datos de forma económica. Estas clases de almacenamiento ofrecen la misma durabilidad y resistencia que las clases de almacenamiento S3 Standard y S3 Standard-IA. Estas clases de almacenamiento se diferencian en lo siguiente:

S3 Glacier: Se utiliza para archivar datos a los que se accede con poca frecuencia y requieren recuperación en milisegundos. Los datos almacenados en la clase de almacenamiento S3 Glacier ofrecen un ahorro de costos en comparación con la clase de almacenamiento S3 Standard, pero con la misma latencia y rendimiento.

Ofrece escalabilidad prácticamente ilimitada y está diseñada para lograr un 99,999999999% (11 nueves) de durabilidad de datos. S3 Glacier ofrece opciones para acceder más rápidamente a los datos de archivos y tiene el menor costo en almacenamiento de archivos en la nube.

S3 Glacier Deep Archive: Se usa para archivar datos a los que se necesita obtener acceso en contadas ocasiones. Los datos almacenados en la clase S3 Glacier Deep Archive tienen un periodo mínimo de almacenamiento de 180 días y un tiempo predeterminado de recuperación de 12 horas.

S3 Glacier Deep Archive es la opción de almacenamiento más económica de AWS. Los costos de almacenamiento de S3 Glacier Deep Archive son menores que los de la clase de almacenamiento S3 Glacier. Se pueden reducir los costos de recuperación de S3 Glacier Deep Archive si se utiliza la recuperación en bloque, que devuelve los datos en menos de 48 horas. Todos los objetos de esta clase de almacenamiento se replican y almacenan en al menos tres zonas de disponibilidad dispersadas geográficamente.

Las clases de almacenamiento de Amazon S3 Glacier se crearon específicamente para el archivado de datos y ofrecen el mayor rendimiento, la mayor flexibilidad de recuperación y el menor costo de almacenamiento de archivos en la nube.

S3 Outposts: Amazon S3 en Outposts ofrece almacenamiento de objetos a su entorno local de AWS Outposts. Mediante el uso de las características y las API de S3 disponibles en las regiones de AWS actualmente, S3 en Outposts facilita el almacenamiento y la recuperación de datos en el Outpost, así como la seguridad de los datos, el control del acceso, el etiquetado y la generación de informes. S3 en Outposts le proporciona una clase de almacenamiento Amazon S3 única, denominada «OUTPOSTS», que utiliza las API de S3 y

está diseñada para almacenar datos de manera duradera y redundante en múltiples dispositivos y servidores de su Outposts. La clase de almacenamiento de S3 Outposts es ideal para cargas de trabajo con requisitos de residencia de datos locales y para satisfacer las necesidades de rendimiento exigentes al mantener los datos cerca de las aplicaciones locales.

Amazon S3 Outposts está diseñado para almacenar datos de forma duradera y redundante a través de varios dispositivos y servidores en sus Outposts. Funciona bien para las cargas de trabajo con requisitos de residencia de datos locales que deben satisfacer las necesidades de rendimiento demandantes y mantener los datos cerca de las aplicaciones en las instalaciones.

Características:

- Compatibilidad y administración de buckets de S3 Object a través de S3 SDK.
- Diseñado para un almacenamiento de datos duradero y redundante en su Outposts.
- Cifrado utilizando SSE-S3 y SSE-C.
- Autenticación y autorización utilizando IAM y puntos de acceso de S3.
- Transferencia de datos a regiones de AWS utilizando AWS DataSync.
- Acciones de vencimiento del ciclo de vida de S3.

AMAZON EBS - ELASTIC BLOCK STORE

Amazon Elastic Block Store (Amazon EBS) es un servicio de almacenamiento en **bloque** fácil de usar, escalable y de alto rendimiento diseñado para Amazon Elastic Compute Cloud (Amazon EC2).

Con EBS podemos crear un disco duro virtual, llamado volumen de EBS, que se puede adjuntar a las instancias de EC2.

Los volúmenes de almacenamiento a nivel de bloque se comportan como discos duros físicos.

Estos son unidades independientes de los volúmenes de almacén de instancia local y no están vinculados directamente al host, que EC2 está ejecutando. Eso significa que los datos que escribe en un volumen de EBS pueden persistir entre la detención y los lanzamientos de una instancia de EC2.

Las instancias de Amazon EC2 son servidores virtuales, si iniciamos una instancia desde un estado detenido, la instancia podría iniciarse en otro host, donde el volumen del almacén de instancias usado anteriormente ya no existe, por lo que AWS recomienda usar los volúmenes de EBS, que vienen en diferentes tamaños y tipos.

¿Cómo funciona EBS?

Es definir el tamaño, el tipo y las configuraciones del volumen que necesitamos. Aprovisionar el volumen y, a continuación, adjuntarlo a una instancia de EC2. Desde ahí podemos configurar la aplicación para escribir en el volumen y listo. Si se detiene y, a continuación, lanzamos la instancia de EC2, los datos del volumen permanecen.

Debido a que el caso práctico de los volúmenes de EBS es tener un disco duro que sea persistente, en el que las aplicaciones puedan escribir, tal vez sea importante respaldar dichos datos. EBS permite realizar respaldos progresivos de los datos denominados instantáneas. Es muy importante tomar instantáneas regulares de los volúmenes de EBS, de esta manera, si alguna vez una unidad se daña, no se han perdido los datos y podemos restaurarlos desde una instantánea.

Casos de Uso:

Creación de Volúmenes: Con EBS, podemos crear volúmenes de almacenamiento y enlazarlos a una instancia de EC2. Estos volúmenes actúan como discos virtuales que se pueden adjuntar a las instancias EC2.

Sistema de Archivos o Bases de Datos: Una vez que hemos atado un volumen a una instancia EC2, podemos crear un sistema de archivos sobre ese volumen o incluso ejecutar una base de datos en él. Es decir, podemos utilizar estos volúmenes para cualquier otro tipo de uso que requiera un dispositivo de bloque.

SAN en la nube: Podemos crear nuestra SAN (Storage Area Network) en la nube para aplicaciones de uso intensivo de E/S. Esto es especialmente útil para migrar cargas de trabajo de SAN desde instalaciones locales a la nube.

Bases de Datos Relacionales o NoSQL: EBS es ideal para ejecutar bases de datos relacionales como SAP HANA, Oracle, Microsoft SQL Server, PostgreSQL, MySQL, Cassandra y MongoDB.

Motores de Análisis de Macrodatos: Es posible cambiar fácilmente el tamaño de los clústeres para motores de análisis de big data como Hadoop y Spark utilizando EBS.

Características clave de Amazon EBS:

- **Escalabilidad:** Podemos ajustar el tamaño de los volúmenes según nuestras necesidades.
- **Alto rendimiento:** Ofrece velocidades de E/S rápidas para cargas de trabajo exigentes.
- **Facilidad de uso:** Se integra sin problemas con las instancias de Amazon EC2.

- **Replicación y alta disponibilidad:** Los volúmenes almacenados en EBS se replican automáticamente dentro de una zona de disponibilidad, lo que garantiza protección frente a errores y desastres.

En resumen, Amazon EBS es una herramienta esencial para gestionar el almacenamiento persistente en la nube, especialmente cuando se trabaja con instancias de Amazon EC2. EBS proporciona volúmenes de almacenamiento de bloques persistentes, perfectos para una variedad de casos de uso, desde bases de datos hasta aplicaciones empresariales y desarrollo.

Comparación entre Amazon S3 y Amazon EBS

Para determinar correctamente la clase de almacenamiento a elegir, debemos entender la diferencia entre el almacenamiento de objetos y el almacenamiento en bloques.

El almacenamiento de objetos trata cualquier archivo como un objeto completo y discreto. Esto es ideal para documentos, imágenes, archivos de video que se cargan y se consumen como objetos enteros. Pero cada vez que se modifica el objeto debemos volver a cargar el archivo completo, es decir, no hay actualizaciones diferenciales, mientras que el almacenamiento en bloques descompone esos archivos en piezas o bloques de componentes pequeños.

Para entender mejor la diferencia, veamos algunos ejemplos:

Supongamos que estamos ejecutando un sitio web de análisis de fotos donde los usuarios suben una foto de ellos mismos y la aplicación encuentra los animales que se parecen a ellos. Posiblemente tengamos millones de imágenes de animales que deben ser indexadas y quizás sean vistas por miles de personas a la vez. Este es el caso práctico perfecto para S3.

Amazon S3 ya está habilitado para la web. Cada objeto ya tiene una URL de la que podemos controlar los derechos de acceso para definir quién puede ver o administrar imágenes. Está distribuido regionalmente, lo que significa que ya tiene 11 nueves durabilidad, así que no hay que preocuparnos por la estrategia de respaldo porque S3 es la estrategia de respaldo. Además, el ahorro de costo es sustancial al sobrepasar la misma carga de almacenamiento en EBS.

Con la ventaja adicional de trabajar sin servidor, no se necesitamos instancias de Amazon EC2.

Ahora, supongamos que tenemos un archivo de video de 80 gigabytes en el que estamos haciendo correcciones de edición.

Eso significa que, para este archivo de 80 gigabytes, cuando editamos una escena de la película y guardamos ese cambio, el motor solo actualiza los bloques donde se encuentran esos bites.

Si estamos realizando un montón de micro ediciones, EBS es el caso práctico perfecto, mientras que, si utilizamos S3 cada vez que guardemos los cambios, el sistema tendría que cargar los 80 gigabytes, todo el proceso, cada vez.

Esto significa que, si estamos usando objetos completos o solo cambios ocasionales, S3 es la mejor opción, pero si nos encontramos realizando funciones complejas de lectura, escritura y cambios, entonces, sin duda, EBS es la mejor opción.

La mejor alternativa de almacenamiento depende de la carga de trabajo particular.

Cada servicio es el servicio adecuado para necesidades específicas. Una vez que entendemos la necesidad, sabremos qué servicio es el mejor.

En resumen,

EBS:

- Almacenamiento de bloques
- Con un peso en tamaños de hasta 16 terabytes
- Con una capacidad única para sobrevivir hasta la finalización de sus instancias de Amazon EC2
- Es de estado sólido SSD de forma predeterminada
- Tiene opciones de disco mecánico HDD

S3:

- Almacenamiento de objetos
- Con un peso de almacenamiento ilimitado
- Con objetos individuales de hasta 5000 gigabytes de tamaño
- Se especializa en escribir una vez/leer muchas
- Tiene hasta once nueves (99.999999999) de durabilidad

AMAZON EFS - ELASTIC FILE SYSTEM

Amazon Elastic File System (EFS) es un sistema de **archivos** administrado y escalable que se utiliza con los servicios de la nube de AWS y los recursos en las instalaciones. A medida que agregamos y eliminamos archivos, Amazon EFS crece y se reduce automáticamente, sin necesidad de administración o aprovisionamiento. EFS puede escalar a petabytes según la demanda sin interrumpir las aplicaciones.

Es muy común que las empresas compartan sistemas de archivos entre múltiples aplicaciones. Por ejemplo, podemos tener varios servidores que ejecutan análisis de grandes cantidades de datos almacenados en un sistema de archivos compartido. Estos datos tradicionalmente se alojan en las instalaciones.

Con las soluciones de almacenamiento de archivos en las instalaciones, tenemos que asegurar que pueda mantenerse al día con la cantidad de datos que estamos almacenando, asegurarnos de realizar respaldos y de que los datos se almacenen de forma redundante, así como de administrar todos los servidores que alojan esos datos.

Con AWS, no tenemos que preocuparnos por comprar todo ese hardware y mantener todo el sistema de archivos funcionando desde un punto de vista operativo. Con Amazon EFS, podemos mantener todos los sistemas de archivos existentes y con el beneficio de que AWS hace todo el trabajo pesado del escalado y la replicación.

Amazon EFS permite tener varias instancias que pueden acceder a los datos de EFS al mismo tiempo. Se escala y reduce verticalmente según sea necesario, sin tener que hacer nada para que ese escalado ocurra.

Todas estas características parecen muy buenas, ¿verdad? Pero Amazon EBS también permite almacenar archivos a los que podemos acceder desde instancias de EC2. Entonces, ¿cuál es exactamente la diferencia entre EBS y EFS? La respuesta es muy sencilla: Los volúmenes de Amazon EBS se adjuntan a la instancia de EC2 y constituyen un recurso a nivel de disponibilidad. Para juntar otro EC2 a EBS, necesita estar en la misma zona de disponibilidad. Podemos guardar archivos ahí. También podemos ejecutar una base de datos o almacenar aplicaciones. Es un disco duro. Si nosotros aprovisionamos un volumen de EBS de dos terabytes y lo llenamos, no se escala automáticamente para proporcionar más almacenamiento. Así que eso es EBS.

Mientras que Amazon EFS puede tener varias instancias leyendo y escribiendo al mismo tiempo. Pero no es solo un disco duro en blanco en el que podemos escribir. Es un verdadero sistema de archivos para Linux. También es un recurso regional. Lo que significa que cualquier instancia de EC2 de la región puede escribir en el sistema de archivos de EFS. A medida que escribimos más datos en EFS, se escala automáticamente. No es necesario aprovisionar más volúmenes.

Comparación entre Amazon EBS y Amazon EFS

Un volumen de Amazon EBS almacena los datos en una única zona de disponibilidad. Para adjuntar una instancia de Amazon EC2 a un volumen de EBS, tanto la instancia de Amazon EC2 como el volumen de EBS deben encontrarse en la misma zona de disponibilidad.

Amazon EFS es un servicio regional. Almacena datos en varias zonas de disponibilidad y entre ellas.

El almacenamiento duplicado permite acceder a los datos simultáneamente desde todas las zonas de disponibilidad de la región en la que se ubica un sistema de archivos. Además, los servidores en las instalaciones pueden acceder a Amazon EFS mediante AWS Direct Connect.

CONCLUSIONES SOBRE LOS SERVICIOS DE ALMACENAMIENTO

- La nube ofrece una alternativa más segura y versátil para la transferencia de archivos en comparación con el FTP tradicional. Considera adoptar soluciones en la nube para tus necesidades de transferencia de archivos.
- Amazon S3 es para el almacenamiento de objetos, Amazon EBS para el almacenamiento en bloques y Amazon EFS para el almacenamiento de archivos.
- Podemos decir que S3 es el servicio más barato, siendo configurable y accesible desde cualquier sitio. EFS es mucho más caro que EBS, pero EBS solo puede ser accesible desde una instancia Amazon EC2, pudiendo generar un problema para algunos tipos de aplicaciones. Sin embargo, Amazon EFS es más potente cuando se trata de sistemas de archivos compartidos y escalables.
- Es importante evaluar las necesidades específicas para tomar la decisión correcta consiguiendo el mejor equilibrio entre las necesidades, el coste y el rendimiento deseado.
- Hoy en día hay aplicaciones donde se pueden utilizar las tres opciones en función de la modularidad que tenga, una opción no es excluyente de otra.