Engineer

Almacenamiento de datos

Guardar datos para leerlos luego, o medios que los conservan

Qué es

El almacenamiento de datos, normalmente abreviado como almacenamiento, es la capa de un sistema donde los datos permanecen entre usos, en medios que los conservan tras un reinicio o un corte de corriente. El glosario del NIST da al término dos sentidos, ambos tomados de SP 800-88r2: la conservación de datos de forma que puedan recuperarse y el hardware en el que se pueden introducir datos y del que se pueden recuperar.

Toda modalidad se apoya en un medio físico: un disco duro giratorio, una unidad de estado sólido, una tarjeta SD o una cinta. El software ve una de estas tres capas por encima:

- **Almacenamiento en bloques:** presenta un dispositivo como una secuencia de bloques de tamaño fijo identificados por un número. Una partición de disco, un conjunto RAID y un disco virtual en la nube tienen ese aspecto para el núcleo; un sistema de archivos o un motor de base de datos decide qué significan los bloques.
- **Almacenamiento de archivos:** coloca un sistema de archivos sobre esos bloques, con nombres, directorios, tamaños, propietarios y permisos, y permite modificar un archivo unos pocos bytes cada vez. Compartido mediante NFS o SMB, un árbol sirve a muchas máquinas.
- **Almacenamiento de objetos:** conserva objetos completos, cada uno una secuencia de bytes más metadatos, con un nombre dentro de un contenedor, y se accede a él mediante una API HTTP en lugar de un montaje. Google Cloud Storage guarda datos como objetos en contenedores llamados buckets; cada bucket pertenece a un proyecto y cada objeto contiene un archivo de cualquier formato como datos que no se pueden modificar una vez escritos.

Las tres modalidades se distinguen por el coste de un cambio. Un dispositivo de bloques reescribe un bloque en su ubicación. Un sistema de archivos modifica parte de un archivo y lo mueve mediante un cambio de nombre. Un almacén de objetos sustituye el objeto entero: la página de Google sobre objetos indica que no se pueden añadir datos al final de un objeto ni truncarlo, y una sustitución es atómica, por lo que los lectores reciben la versión antigua hasta que termina la nueva subida. Según la herramienta, cambiar un nombre puede significar copiar con el nombre nuevo y borrar el original; Cloud Storage también admite cambios de nombre atómicos de objetos dentro de un bucket. En un bucket con espacio de nombres plano, las carpetas son una representación que las herramientas construyen a partir de las barras de los nombres de objeto; la cadena `folder1/` forma parte del nombre y el bucket no almacena ninguna carpeta.

El mismo archivo tiene este aspecto en cada capa:

```
# block: the kernel's view of a disk as numbered blocks
lsblk /dev/sda

# file: a filesystem mounted over those blocks, with names and directories
df -h /srv

# object: a whole object sent to a bucket over an API
gcloud storage cp readings.csv.gz gs://example-bucket/2024/01/
```

Utilice el concepto en una decisión

Aplique «¿Qué es el almacenamiento de datos?» a una decisión real. Podemos ayudarle a relacionar el concepto con el sistema que está construyendo o comprando. Copias de seguridad que se prueban en vez de darse por buenas y un plan escrito para la mañana en que un servidor no vuelva a encenderse.

Explore nuestros servicios Póngase en contacto →

Compruebe los supuestos detrás del término

Utilice la explicación y los trabajos enlazados para formular preguntas sobre su proyecto. Podemos examinar los supuestos con usted antes de elegir una herramienta. 100 copias críticas automatizadas con Barman — 100 copias críticas automatizadas en las bases de datos

100 copias críticas automatizadas con Barman Póngase en contacto →

Por qué importa

El medio determina cómo se pierden los datos. Un disco duro, un SSD y una tarjeta SD fallan de formas distintas, y un método de recuperación adecuado para uno es una herramienta equivocada para otro. La redundancia dentro de

una caja, como un par de discos en espejo o un conjunto RAID, mantiene los datos legibles cuando falla un disco, pero reproduce cada borrado en todos los discos con la misma fidelidad con la que escribe los datos.

Durabilidad y disponibilidad son promesas distintas. La durabilidad es la probabilidad de que los datos almacenados sobrevivan; la disponibilidad, la de poder leerlos en un momento determinado. Cloud Storage indica una durabilidad anual del 99,999999999 % para todas las clases de almacenamiento, mientras que su disponibilidad mensual típica va del 99,9 % a más del 99,99 %, según la clase y la ubicación. Una clase de almacenamiento es un metadato de cada objeto que establece su disponibilidad y su modelo de precios, y los objetos nuevos heredan la clase predeterminada del bucket. Las clases más frías reducen el precio de conservar datos a cambio de una disponibilidad algo menor, un periodo mínimo de almacenamiento de 30, 90 o 365 días y un cargo por volver a leer los datos; Google las orienta a datos que se leen una vez al mes o menos, como máximo una vez por trimestre y menos de una vez al año.

La ubicación forma parte de los datos. Una copia en el mismo disco, la misma máquina o la misma cuenta en la nube comparte su suerte; una copia en otro lugar no. La ubicación también determina velocidad y coste: Google señala que mantener los datos en la misma región que los recursos de cómputo que los utilizan maximiza el rendimiento del trabajo intensivo en datos y puede reducir los cargos de red.

La forma en que el software accede al almacenamiento condiciona lo demás. Restic, un programa de copias de seguridad, define su repositorio simplemente como un directorio con un conjunto de subdirectorios y archivos; ese mismo directorio puede estar en un disco local, en un servidor al que se accede mediante SFTP o en un almacén de objetos como Google Cloud Storage o Amazon S3. Los bytes son los mismos, pero cambian la latencia, las formas de fallo y quién controla la cuenta.

Cómo se aplica en la práctica

El trabajo anterior de la fundadora abarcó distintas capas de la pila: desde el medio físico que falla hasta el almacén de objetos que mueve archivos completos. La implementación de copias de seguridad de la propia empresa define la automatización para conservar una copia en otro proveedor como protección ante la pérdida de un servidor; activar ese almacenamiento y verificar una restauración real son pasos distintos.

El medio: la recuperación de datos abarcó prácticamente todos los tipos habituales de almacenamiento, desde tarjetas SD y discos duros giratorios hasta SSD, conjuntos RAID, unidades externas y sistemas Mac. Las personas llevaban medios averiados o dañados que contenían fotografías, archivos de negocio o la única copia de algo importante. Cada medio falla y permite recuperar sus datos de forma distinta: un SSD plantea un problema diferente al de una unidad de platos, un conjunto RAID plantea otro y un sistema de archivos Mac tiene sus propias particularidades. Conocer el enfoque adecuado para el medio concreto importaba tanto como cualquier técnica individual, y los clientes recuperaron datos críticos de dispositivos que ya daban por perdidos.

El almacén de objetos: en seis procesos ETL/ELT que alimentaban Google BigQuery, un bucket de Google Cloud Storage estaba entre el origen y el almacén de datos. Los conjuntos de datos y las tablas seleccionados y preparados se exportaban a formatos intermedios, se comprimían con gzip y se transferían al bucket; desde allí, la herramienta bq CLI y scripts de Python los cargaban en BigQuery. Entraban archivos comprimidos completos y salían archivos completos: el patrón para el que está diseñado el almacenamiento de objetos.

La copia en otro lugar forma parte de la implementación de copias de seguridad y pruebas de restauración de la empresa. Un rol de Ansible instala scripts de restic y unidades de systemd para un volcado de base de datos cuando está configurado, una instantánea cifrada y deduplicada enviada por SFTP al almacenamiento de otro proveedor, la eliminación de copias según la política de retención y una comprobación de integridad. Un flujo mensual independiente restaura un archivo conocido en una ubicación temporal y compara sus bytes, comprobando la ruta de lectura de los datos almacenados, no solo la de escritura. La copia y la restauración tienen señales de éxito separadas, y los temporizadores requieren un repositorio configurado. El registro de implementación de fecha 2026-10-27 indica que esos temporizadores estaban desactivados sin un repositorio y no documenta ninguna restauración real completada. Documenta la automatización y lo que debe activarse; no demuestra que la empresa ya pueda perder el servidor y recuperar sus datos.

Dónde surgen los problemas

La mayoría de los problemas de almacenamiento proceden de una copia, una comprobación o un coste que parecían adecuados hasta el día en que hicieron falta.

La copia comparte el destino del original. Una instantánea en la misma cuenta, una copia de seguridad en el mismo disco o una réplica en el mismo edificio resisten los fallos pequeños, pero sucumben ante uno de gran alcance. Enumere los sucesos individuales que podrían destruir ambas copias y traslade una hasta que la lista sea corta.

La redundancia se cuenta como copia de seguridad. Un espejo reproduce cada escritura en todos los discos, incluidos los borrados y la corrupción. Resuelve la avería de un disco; una copia independiente con historial propio resuelve una equivocación.

Se trata un almacén de objetos como un sistema de archivos. El código escrito para archivos espera cambios de nombre y listados de directorios poco costosos. Algunas herramientas implementan los cambios de nombre de objetos como una copia seguida de un borrado, aunque Cloud Storage también admite cambios de nombre atómicos dentro de un bucket. Google señala que las estructuras profundamente anidadas y semejantes a directorios en un espacio de nombres plano no ofrecen el rendimiento de listado de un sistema de archivos nativo. Organice los nombres de objeto según cómo se vayan a leer y mantenga en almacenamiento de archivos o de bloques el trabajo que modifica archivos en su ubicación.

La comprobación lee el índice y omite los datos. Por defecto, la orden `check` de restic verifica la estructura del repositorio y deja sin leer los archivos de paquetes que contienen los datos respaldados, porque leerlos exige descargarlos todos. `--read-data` los lee todos, y desde `--read-data-subset=1/5` hasta `5/5` se recorre el repositorio en cinco ejecuciones. Una comprobación que solo enumera archivos demuestra que los nombres existen y poco sobre los bytes que hay detrás.

Una clase fría contiene datos calientes. Un archivo que se lee a diario desde una clase cuyo precio está pensado para lecturas mensuales genera un cargo en cada lectura; adapte la clase al patrón real de lectura de los datos.

Lecturas adicionales

- NIST glossary: storage — los dos sentidos del término según los define NIST SP 800-88r2.

- About Cloud Storage objects — datos y metadatos de objetos, reglas de nombres, inmutabilidad y espacio de nombres plano tras las carpetas simuladas.
- Cloud Storage: Storage classes — durabilidad, disponibilidad, periodos mínimos de almacenamiento y cargos por lectura de cada clase.
- restic: Preparing a new repository — qué es un repositorio y los backends locales, SFTP y de objetos donde puede residir.
- restic: Working with repositories — comprobar la estructura de un repositorio y volver a leer los datos que contiene.
- Cloud Storage: traslado de objetos — la operación de cambio de nombre atómico dentro de un bucket, distinta de las herramientas que simulan un traslado mediante copia y borrado.

Glosario

Elija según sus restricciones reales

Describa la ruta de despliegue, los riesgos operativos y quiénes se encargan de las guardias. Podemos definir el alcance de un cambio con comprobaciones, una vía de recuperación y responsabilidades claras.

Vea cuánto cuesta Póngase en contacto →

Haga práctica la terminología

Pregunte cómo se aplica «¿Qué es el almacenamiento de datos?» a su organización. Basta un ejemplo concreto de su sistema actual para iniciar la conversación.

Explore nuestros servicios

welcome@engineer.company +45 50 10 61 84 Copenhague, Dinamarca

Llévese esta página

Tarjeta de contacto

Creado con esmero y cariño por D.

CVR 45969703 · Tagensvej 105C, 3. th, 2200 København N

Derechos de autor © 2025 – actualidad · Engineer ApS · Todos los derechos reservados.

Impreso desde https://engineer.company/es/glossary/storage/