Engineer

Datenspeicherung

Daten zum späteren Lesen aufbewahren oder die Datenträger

Was es ist

Datenspeicherung, oft kurz Speicherung genannt, ist die Schicht eines Systems, in der Daten zwischen ihren Nutzungen liegen: auf Medien, die sie über einen Neustart und einen Stromausfall hinweg erhalten. Das NIST-Glossar nennt zwei Bedeutungen aus SP 800-88r2: die abrufbare Aufbewahrung von Daten und die Hardware, auf die Daten geschrieben und von der sie wieder abgerufen werden können.

Jede Art beruht auf einem physischen Medium: einer rotierenden Festplatte, einer SSD, einer SD-Karte oder einem Band. Die Software sieht darüber eine von drei Schichten:

- **Blockspeicher** stellt ein Gerät als Folge gleich großer, über Nummern adressierter Blöcke dar. Eine Festplattenpartition, ein RAID-Verbund und eine virtuelle Cloud-Festplatte sehen für den Kernel so aus. Ein Dateisystem oder eine Datenbankengine entscheidet, was die Blöcke bedeuten.
- **Dateispeicher** legt ein Dateisystem über diese Blöcke: Namen, Verzeichnisse, Größen, Eigentümer und Berechtigungen. Dabei lassen sich jeweils wenige Bytes einer Datei ändern. Über NFS oder SMB geteilt, bedient ein Verzeichnisbaum viele Rechner.
- **Objektspeicher** hält vollständige Objekte, jeweils eine Bytefolge mit Metadaten, unter einem Namen in einem Container. Der Zugriff erfolgt über eine HTTP-API statt über einen Mount. Google Cloud Storage speichert Daten als Objekte in Containern namens Buckets. Jeder Bucket gehört zu einem Projekt, und jedes Objekt enthält eine Datei beliebigen Formats als nach dem Schreiben unveränderliche Daten.

Die drei unterscheiden sich darin, was eine Änderung kostet. Ein Blockgerät schreibt einen Block an seiner bisherigen Stelle neu. Ein Dateisystem ändert einen Teil einer Datei und verschiebt sie durch Umbenennen. Ein Objektspeicher

ersetzt das ganze Objekt: Googles Objektseite schließt das Anhängen von Daten an ein Objekt und das Kürzen seines Inhalts aus. Ein Austausch ist atomar, sodass Leser bis zum Abschluss des neuen Uploads die alte Version erhalten. Je nach Werkzeug kann Umbenennen eine Kopie unter dem neuen Namen und das Löschen des Originals bedeuten; Cloud Storage unterstützt auch atomare Objektumbenennungen innerhalb eines Buckets. In einem Bucket mit flachem Namensraum sind Ordner ein von den Werkzeugen aus Schrägstrichen in Objektnamen erzeugtes Bild. Die Zeichenfolge `folder1/` ist Teil des Namens; der Bucket speichert überhaupt keinen Ordner.

Dieselbe Datei sieht auf jeder Schicht so aus:

```
# 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/
```

Das Konzept für eine Entscheidung nutzen

Nutzen Sie „Was ist Datenspeicherung?“ für eine konkrete Entscheidung. Wir können helfen, das Konzept mit dem System in Verbindung zu bringen, das Sie entwickeln oder einkaufen. Sicherungen, deren Funktion geprüft und nicht vorausgesetzt wird, und ein schriftlicher Plan für den Morgen, an dem ein Server nicht wieder startet.

Unsere Leistungen entdecken Kontakt aufnehmen →

Die Annahmen hinter dem Begriff prüfen

Nutzen Sie die Erklärung und die verlinkten Arbeiten, um Fragen für Ihr Projekt zu formulieren. Wir können die Annahmen gemeinsam mit Ihnen prüfen, bevor ein Werkzeug ausgewählt wird. 100 wichtige Sicherungen mit Barman automatisiert — 100 automatisierte geschäftskritische Datenbanksicherungen

100 wichtige Sicherungen mit Barman automatisiert Kontakt aufnehmen →

Warum es wichtig ist

Das Medium bestimmt, wie Daten verloren gehen. Eine Festplatte, eine SSD und eine SD-Karte fallen jeweils anders aus. Ein passendes Wiederherstellungsverfahren für das eine ist für das nächste das falsche Werkzeug. Redundanz innerhalb eines Geräts, etwa ein Spiegelpaar oder ein RAID-Verbund, hält Daten nach dem Ausfall einer einzelnen Platte lesbar. Sie schreibt aber jede Löschung ebenso zuverlässig auf jede Platte wie die Daten selbst.

Haltbarkeit und Verfügbarkeit sind getrennte Zusagen. Haltbarkeit bezeichnet die Wahrscheinlichkeit, dass gespeicherte Daten erhalten bleiben; Verfügbarkeit die Wahrscheinlichkeit, sie zu einem bestimmten Zeitpunkt lesen zu können. Cloud Storage nennt für jede Speicherklasse eine jährliche Haltbarkeit von 99,999999999 %. Die angegebene typische monatliche Verfügbarkeit reicht dagegen je nach Klasse und Standort von 99,9 % bis über 99,99 %. Die Speicherklasse ist ein Metadatum jedes Objekts, das Verfügbarkeit und Preismodell festlegt. Neue Objekte übernehmen die Vorgabe ihres Buckets. Kältere Klassen tauschen niedrigere Aufbewahrungskosten gegen etwas geringere Verfügbarkeit, eine Mindestspeicherdauer von 30, 90 oder 365 Tagen und eine Gebühr für das Zurücklesen der Daten. Google sieht sie für Daten vor, die höchstens einmal im Monat, höchstens einmal im Quartal beziehungsweise seltener als einmal im Jahr gelesen werden.

Der Ort gehört zu den Daten. Eine Kopie auf demselben Datenträger, demselben Rechner oder im selben Cloud-Konto teilt dessen Schicksal; eine Kopie anderswo nicht. Der Ort bestimmt auch Geschwindigkeit und Kosten. Google weist darauf hin, dass Daten in derselben Region wie die sie verarbeitenden Rechenressourcen die Leistung datenintensiver Arbeit maximieren und Netzwerkgebühren senken können.

Wie Software auf den Speicher zugreift, prägt den Rest. Das Sicherungsprogramm Restic definiert sein Repository schlicht als Verzeichnis mit einer Reihe von Unterverzeichnissen und Dateien. Dasselbe Verzeichnis kann auf einer lokalen Platte, auf einem über SFTP erreichbaren Server oder in einem Objektspeicher wie Google Cloud Storage oder Amazon S3 liegen. Die Bytes bleiben gleich, während Latenz, Ausfallarten und die Person, die das Konto kontrolliert, wechseln.

Wie es in der Praxis aussieht

Von unten nach oben im Speicheraufbau betrachtet, umfasste die frühere Arbeit der Gründerin das physische Medium, das ausfällt, und den Objektspeicher, der vollständige Dateien bewegt. Die eigene Backup-Implementierung des Unternehmens beschreibt die Automatisierung einer Kopie bei einem anderen Anbieter als Schutz gegen den Verlust eines Hosts. Diesen Speicher zu aktivieren und eine tatsächliche Wiederherstellung zu prüfen sind gesonderte Schritte.

Das Medium: Die Datenrettung umfasste so gut wie jeden gängigen Speichertyp, von SD-Karten und rotierenden Festplatten über SSDs und RAID-Verbünde bis zu externen Laufwerken und Mac-Systemen. Menschen brachten ausgefallene oder beschädigte Medien mit Fotos, Geschäftsdateien oder der einzigen Kopie von etwas Wichtigem. Jedes Medium fällt anders aus und gibt seine Daten anders preis: Eine SSD ist ein anderes Problem als ein Plattenlaufwerk, ein RAID-Verbund wieder ein anderes, und ein Mac-Dateisystem hat eigene Besonderheiten. Den passenden Ansatz für das konkrete Medium zu kennen war ebenso wichtig wie eine einzelne Technik. Kunden erhielten kritische Daten von Geräten zurück, die sie bereits als verloren abgeschrieben hatten.

Der Objektspeicher: Bei sechs ETL/ELT-Pipelines für Google BigQuery stand ein Google-Cloud-Storage-Bucket zwischen Quelle und Data Warehouse. Aufbereitete Datensätze und Tabellen wurden in Zwischenformate exportiert, mit gzip komprimiert und in den Bucket übertragen. Von dort luden die bq CLI und Python-Skripte sie in BigQuery. Vollständige komprimierte Dateien gingen hinein und vollständige Dateien kamen heraus: das Muster, für das Objektspeicher gebaut sind.

Die Kopie anderswo gehört zur Implementierung von Backups und Wiederherstellungstests des Unternehmens. Eine Ansible-Rolle installiert restic-Skripte und systemd-Units für einen gegebenenfalls konfigurierten Datenbankdump, einen verschlüsselten und deduplizierten Snapshot per SFTP zu einem anderen Speicheranbieter, das Ausdünnen nach Aufbewahrungsregeln und eine Integritätsprüfung. Ein separater monatlicher Ablauf stellt eine bekannte Datei an einem temporären Speicherort wieder her und vergleicht ihre Bytes. Er testet damit das Zurücklesen und nicht nur das Schreiben. Backup und Wiederherstellung haben getrennte Erfolgssignale, und die Timer benötigen ein konfiguriertes Repository. Der Implementierungsnachweis mit Datum 2026-10-27 hält fest, dass die Timer ohne Repository deaktiviert waren, und verzeichnet keine abgeschlossene tatsächliche Wiederherstellung. Er dokumentiert die Automatisierung und was noch aktiviert werden muss. Er beweist nicht, dass das Unternehmen bereits den Host verlieren und seine Daten wiederherstellen kann.

Wo es schiefgeht

Die meisten Speicherprobleme entstehen durch eine Kopie, eine Prüfung oder Kosten, die bis zu dem Tag unproblematisch wirkten, an dem es darauf ankam.

Eine Kopie teilt das Schicksal des Originals. Ein Snapshot im selben Konto, ein Backup auf derselben Platte, ein Replikat im selben Gebäude: Sie überstehen jeweils kleinere Störungen und fallen beim großen Ausfall gemeinsam aus. Listen Sie einzelne Ereignisse auf, die beide zerstören könnten, und verlagern Sie eine Kopie, bis die Liste kurz ist.

Redundanz wird als Backup gezählt. Eine Spiegelung kopiert jeden Schreibvorgang auf jede Platte, einschließlich Löschungen und beschädigter Daten. Sie hilft gegen eine ausgefallene Platte. Eine separate Kopie mit eigener Historie hilft bei einem Fehler.

Ein Objektspeicher wird wie ein Dateisystem behandelt. Für Dateien geschriebener Code erwartet günstige Umbenennungen und Verzeichnislisten. Manche Werkzeuge setzen Objektumbenennungen als Kopieren und Löschen um, obwohl Cloud Storage auch atomare Umbenennungen innerhalb eines Buckets unterstützt. Google weist darauf hin, dass tief verschachtelte, verzeichnisartige Strukturen in einem flachen Namensraum nicht die Auflistungsleistung eines nativen Dateisystems haben. Richten Sie Objektnamen danach aus, wie sie gelesen werden, und verwenden Sie für Aufgaben, die Dateien an Ort und Stelle ändern, Datei- oder Blockspeicher.

Die Prüfung liest den Index und überspringt die Daten. Restics Befehl `check` prüft standardmäßig die Repository-Struktur und lässt die Packdateien mit den gesicherten Daten ungelesen, weil dafür jede einzelne heruntergeladen werden müsste. `--read-data` liest alle. Mit `--read-data-subset=1/5` bis `5/5` wird das Repository in fünf Läufen abgedeckt. Eine Prüfung, die nur Dateien auflistet, beweist, dass Namen vorhanden sind, sagt aber wenig über die Bytes dahinter aus.

Eine kalte Klasse hält heiße Daten. Eine täglich gelesene Datei in einer für monatliche Abrufe bepreisten Klasse verursacht bei jedem Lesen eine Gebühr. Wählen Sie die Klasse deshalb passend zum tatsächlichen Lesemuster.

Weiterführende Lektüre

- NIST glossary: storage — die beiden Bedeutungen des Wortes nach NIST SP 800-88r2.
- About Cloud Storage objects — Objektdaten und Metadaten, Namensregeln, Unveränderlichkeit und der flache Namensraum hinter simulierten Ordnern.

- Cloud Storage: Storage classes — Haltbarkeit, Verfügbarkeit, Mindestspeicherdauer und Abrufgebühren jeder Klasse.
- restic: Preparing a new repository — was ein Repository ist und welche lokalen, SFTP- und Objektspeicher-Backends es aufnehmen können.
- restic: Working with repositories — die Repository-Struktur prüfen und die enthaltenen Daten zurücklesen.
- Cloud Storage: Objekte verschieben — atomare Umbenennung innerhalb eines Buckets, abgegrenzt von Werkzeugen, die eine Verschiebung durch Kopieren und Löschen nachbilden.

Glossar

Für Ihre tatsächlichen Rahmenbedingungen auswählen

Beschreiben Sie den Bereitstellungsweg, die Betriebsrisiken und die Bereitschaftsverantwortlichen. Wir können eine Änderung mit Prüfungen, einem Wiederherstellungsweg und klaren Zuständigkeiten abgrenzen.

Preise ansehen Kontakt aufnehmen →

Die Begriffe praktisch nutzbar machen

Fragen Sie, wie „Was ist Datenspeicherung?“ auf Ihre Organisation zutrifft. Ein konkretes Beispiel aus Ihrem bestehenden System reicht aus, um das Gespräch zu beginnen.

Unsere Leistungen entdecken

welcome@engineer.company +45 50 10 61 84 Kopenhagen, Dänemark

Nehmen Sie diese Seite mit

Kontaktkarte

Mit Sorgfalt und Liebe gestaltet von D.

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

Copyright © 2025 – heute · Engineer ApS · Alle Rechte vorbehalten.

Gedruckt von https://engineer.company/de/glossary/storage/