Engineer

Lagring

att spara data, eller medier som bär dem, för senare läsning

Vad det är

Datalagring, oftast kallad bara lagring, är den del av ett system där data finns mellan användningarna, på medier som behåller dem genom en omstart och ett strömavbrott. Ordlistan hos NIST ger ordet två betydelser, båda hämtade från SP 800-88r2: att data bevaras så att de kan hämtas igen, och den hårdvara som data kan föras in i och hämtas ur.

Varje sort vilar på ett fysiskt medium: en hårddisk med snurrande skivor, en SSD, ett SD-kort, ett band. Programvaran ser ett av tre lager ovanpå det:

- **Blocklagring** visar en enhet som en följd av block med fast storlek, adresserade med nummer. En diskpartition, en RAID-array och en virtuell disk i molnet ser alla ut så för systemkärnan, och ett filsystem eller en databasmotor avgör vad blocken betyder.
- **Fillagring** lägger ett filsystem över blocken: namn, kataloger, storlekar, ägare och behörigheter, och en fil kan ändras några byte i taget. Delat över NFS eller SMB betjänar ett och samma träd många maskiner.
- **Objektlagring** håller hela objekt, vart och ett en följd av byte plus metadata, under ett namn i en behållare, och nås genom ett API över HTTP i stället för genom en montering. Google Cloud Storage lagrar data som objekt i behållare som kallas buckets, varje bucket hör till ett projekt, och ett objekt är en oföränderlig bit data som består av en fil i vilket format som helst.

De tre skiljer sig åt i vad en ändring kostar. En blockenhet skriver om ett block där det ligger. Ett filsystem redigerar en del av en fil och flyttar den med ett namnbyte. Ett objektlager ersätter hela objektet: Googles sida om objekt utesluter tillägg i slutet och avkortning, och en ersättning är atomär, så den som läser får den gamla versionen tills den nya uppladdningen är klar. Beroende på verktyget kan ett namnbyte innebära en kopia under det nya namnet och en radering av originalet; Cloud Storage stöder också atomära namnbyten för objekt

inom en bucket. I en bucket med platt namnrymd är mappar en bild som verktygen ritar upp utifrån snedstreck i objektens namn; strängen `folder1/` är en del av namnet, och bucketen lagrar ingen mapp alls.

Samma fil ser ut så här i varje lager:

```
# block: kärnans bild av en disk som numrerade block
lsblk /dev/sda

# fil: ett filsystem monterat över blocken, med namn och kataloger
df -h /srv

# objekt: ett helt objekt som skickas till en bucket över ett API
gcloud storage cp readings.csv.gz gs://example-bucket/2024/01/
```

Använd begreppet i ett beslut

Använd ”Vad är lagring?” i ett konkret beslut. Vi kan knyta begreppet till systemet ni bygger eller köper. Säkerhetskopior som testas i stället för att tas för givna, och en skriftlig plan för den morgon då en server inte startar igen.

Se våra tjänster Hör av er →

Granska antagandena bakom ordet

Använd förklaringen och det länkade arbetet för att formulera frågor till projektet. Vi kan undersöka antagandena tillsammans innan ett verktyg väljs. Automatiserade 100 kritiska säkerhetskopior — 100 automatiserade kritiska säkerhetskopior av databaserna

Automatiserade 100 kritiska säkerhetskopior Hör av er →

Varför det spelar roll

Mediet avgör hur data går förlorade. En hårddisk, en SSD och ett SD-kort går sönder vart och ett på sitt sätt, och en räddningsmetod som passar det ena är fel verktyg för nästa. Redundans inuti en och samma låda, som ett speglat par eller en RAID-array, håller data läsbara efter att en enskild disk har dött, men den skriver varje radering till varje disk lika troget som den skriver data.

Hållbarhet och tillgänglighet är två skilda löften. Hållbarhet är sannolikheten att lagrade data överlever; tillgänglighet är sannolikheten att de går att läsa vid ett givet ögonblick. Cloud Storage anger en årlig hållbarhet på 99,999999999 % för varje lagringsklass, medan den typiska tillgänglighet per månad som tjänsten anger går från 99,9 % till över 99,99 %, beroende på klass och plats. En lagringsklass är metadata på varje objekt som bestämmer dess tillgänglighet och prismodell, och nya objekt ärver standardklassen i sin bucket. De kallare klasserna byter ett lägre pris för att behålla data mot något lägre tillgänglighet, en minsta

lagringstid på 30, 90 eller 365 dagar och en avgift för att läsa tillbaka data; Google riktar dem mot data som läses en gång i månaden eller mer sällan, högst en gång per kvartal och mindre än en gång per år.

Platsen är en del av datan. En kopia på samma disk, samma maskin eller samma molnkonto delar öde med den saken, och en kopia någon annanstans gör det inte. Platsen bestämmer också hastighet och kostnad: Google noterar att om data hålls i samma region som den beräkning som använder dem får dataintensivt arbete bäst prestanda, och nätverksavgifterna kan bli lägre.

Hur programvaran når lagringen formar resten. Restic, ett program för säkerhetskopiering, definierar sitt repo som helt enkelt en katalog som innehåller en uppsättning underkataloger och filer, och samma katalog kan ligga på en lokal disk, på en server som nås över SFTP eller i ett objektlager som Google Cloud Storage eller Amazon S3. Innehållet förblir detsamma, medan latensen, sätten att gå sönder och den person som styr kontot alla ändras.

Så ser det ut i praktiken

Nedifrån och upp i stacken berörde grundarens tidigare arbete det fysiska mediet som går sönder och objektlagret som flyttar hela filer. Företagets egen implementering av säkerhetskopiering beskriver ett automatiserat upplägg för en kopia hos en annan leverantör som skydd mot att en värddator går förlorad; att aktivera den lagringen och verifiera en återställning i drift är separata steg.

Mediet: dataräddningen omfattade i stort sett alla vanliga lagringstyper, från SD-kort och snurrande hårddiskar till SSD:er, RAID-arrayer, externa diskar och Mac-system. Folk kom in med lagringsmedier som hade gått sönder eller skadats, med foton, affärsdokument, den enda kopian av något som betydde något. Varje medium går sönder och lämnar ifrån sig sina data på sitt eget sätt: en SSD är ett annat problem än en disk med roterande skivor, en RAID-array är ännu ett annat problem, och ett Mac-filsystem har sina egna egenheter. Att veta vilket angreppssätt som passar just det mediet betydde lika mycket som någon enskild teknik, och kunderna fick tillbaka kritiska data från enheter som de redan hade avskrivit som förlorade.

Objektlagret: i sex ETL/ELT-pipelines som matade Google BigQuery stod en bucket i Google Cloud Storage mellan källan och analysdatalagret. Kurerade dataset och tabeller exporterades till mellanformat, komprimerades med gzip

och överfördes till bucketen, och därifrån läste verktyget bq och Python-skript in dem i BigQuery. Hela komprimerade filer gick in och hela filer kom ut, det mönster som objektlagring är byggd för.

Kopian någon annanstans ingår i företagets implementering av säkerhetskopiering och återställningstest. En Ansible-roll installerar restic-skript och systemd-enheter för en databasdump där det är konfigurerat, en krypterad, deduplicerad ögonblicksbild via SFTP till lagring hos en annan leverantör, gallring enligt bevarandepolicyn och en integritetskontroll. Ett separat månatligt arbetsflöde återställer en känd fil till ett tillfälligt arbetsutrymme och jämför dess byte, så att vägen tillbaka testas och inte bara skrivningen. Säkerhetskopiering och återställning har separata signaler för lyckade körningar, och timerenheterna kräver ett konfigurerat repo. Implementeringsbeskrivningen daterad 2026-10-27 anger att timerenheterna var inaktiverade utan ett repo och redovisar ingen genomförd återställning i drift. Den dokumenterar automatiseringen och vad som måste aktiveras; den bevisar inte att företaget redan kan förlora värddatorn och återställa sina data.

Var det går fel

Det mesta besväret med lagring kommer från en kopia, en kontroll eller en kostnad som såg bra ut fram till den dag den behövdes.

En kopia delar öde med originalet. En ögonblicksbild i samma konto, en säkerhetskopia på samma disk, en replik i samma byggnad: var och en överlever de små felen och går under med det stora. Lista de enskilda händelser som skulle kunna slå ut båda, och flytta en kopia tills listan är kort.

Redundans räknas som en säkerhetskopia. En spegel kopierar varje skrivning till varje disk, raderingar och korrupta data inräknade. Den är svaret på en död disk; en separat kopia med en egen historik är svaret på ett misstag.

Ett objektlager behandlas som ett filsystem. Kod som är skriven för filer räknar med billiga namnbyten och kataloglistningar. Vissa verktyg genomför namnbyten för objekt som en kopia plus en radering, trots att Cloud Storage också stöder atomära namnbyten inom en bucket. Google noterar att djupt nästlade, katalogliknande strukturer i en platt namnrymd saknar listningsprestandan hos ett riktigt filsystem. Lägg upp objektnamnen efter hur de kommer att läsas, och håll arbete som redigerar filer på plats kvar i fil- eller blocklagring.

Kontrollen läser indexet och hoppar över datan. Som standard verifierar kommandot `check` i restic repots struktur och lämnar de packfiler som håller de säkerhetskopierade data olästa, eftersom det skulle innebära att ladda ned varenda en. `--read-data` läser alla, och `--read-data-subset=1/5` till och med `5/5` täcker repot på fem körningar. En kontroll som bara listar filer bevisar att namnen finns, och säger lite om innehållet bakom dem.

Data som läses ofta ligger i en kall klass. En fil som läses varje dag från en klass som är prissatt för läsningar en gång i månaden betalar en avgift vid varje läsning, så matcha klassen mot det läsmönster som datan faktiskt har.

Läs mer

- NIST glossary: storage — ordets två betydelser så som NIST SP 800-88r2 definierar dem.
- About Cloud Storage objects — objektens data och metadata, regler för namn, oföränderlighet och den platta namnrymden bakom simulerade mappar.
- Cloud Storage: Storage classes — hållbarhet, tillgänglighet, minsta lagringstid och avgifter för hämtning för varje klass.
- restic: Preparing a new repository — vad ett repo är, och de lokala backends, SFTP-backends och objektlager som det kan ligga på.
- restic: Working with repositories — att kontrollera ett repos struktur och läsa tillbaka de data det håller.
- Cloud Storage: Flytta objekt — Den atomära namnbytesoperationen inom en bucket, till skillnad från verktyg som simulerar en flytt genom att kopiera och radera.



Ordlista

Välj utifrån era faktiska förutsättningar

Ta med driftsättningsvägen, driftsriskerna och dem som har jour. Vi kan avgränsa en förändring med kontroller, en väg tillbaka och tydligt ansvar.

Se vad det kostar Hör av er →

Gör begreppen praktiska

Fråga hur ”Vad är lagring?” gäller i er organisation. Ett konkret exempel från ert nuvarande system räcker för att börja samtalet.

Se våra tjänster

welcome@engineer.company +45 50 10 61 84 Köpenhamn, Danmark

Ta med dig den här sidan

Visitkort

Omsorgsfullt skapad med kärlek av D.

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

Upphovsrätt © 2025 – nu · Engineer ApS · Alla rättigheter förbehållna.

Utskrivet från https://engineer.company/sv/glossary/storage/