Engineer

Datalagring

å lagre data, eller mediet som holder dem, for senere lesing

Hva det er

Datalagring, til daglig ofte bare kalt lagring, er det laget i et system der data blir liggende mellom hver gang de brukes, på medier som tar vare på dem gjennom en omstart og et strømbrudd. NIST-ordlisten gir ordet to betydninger, begge hentet fra SP 800-88r2: at data oppbevares slik at de kan hentes fram igjen, og maskinvaren som data kan skrives inn på og leses ut fra.

Hver form for lagring hviler på et fysisk medium: en harddisk med roterende plater, en SSD, et SD-kort, et bånd. Programvaren ser ett av tre lag over det:

- **Blokklagring** viser en enhet som en rekke blokker av fast størrelse, adressert med nummer. En diskpartisjon, et RAID-oppsett og en virtuell disk i skyen ser alle slik ut for systemkjernen, og et filsystem eller en databasemotor avgjør hva blokkene betyr.
- **Fillagring** legger et filsystem over de blokkene: navn, mapper, størrelser, eiere og rettigheter, og en fil kan endres noen få byte om gangen. Delt over NFS eller SMB kan ett tre betjene mange maskiner.
- **Objektlagring** holder hele objekter, hvert av dem en rekke byte med metadata, under et navn i en beholder, og nås gjennom et HTTP-API i stedet for å monteres. Google Cloud Storage lagrer data som objekter i beholdere som kalles bøtter (buckets), hver bøtte hører til et prosjekt, og et objekt er en uforanderlig bit data som består av en fil i et hvilket som helst format.

De tre går hver sin vei når det gjelder hva en endring koster. En blokkenhet skriver om én blokk der den står. Et filsystem endrer en del av en fil og flytter den med et navnebytte. Et objektlager erstatter hele objektet: Googles side om objekter utelukker både å legge til på slutten og å kutte et objekt kortere, og en erstatning er atomisk, så lesere får den gamle versjonen helt til den nye opplastingen er ferdig. Avhengig av verktøyet kan et navnebytte bety en kopi under det nye navnet og en sletting av originalen; Cloud Storage støtter også

atomiske navnebytter for objekter innenfor én bøtte. I en bøtte med flatt navnerom er mapper et bilde verktøyene tegner ut fra skråstrekene i objektnavnene. Strengen `folder1/` er en del av navnet, og bøtta lagrer ingen mappe i det hele tatt.

Den samme filen ser slik ut på hvert av lagene:

```
# 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/
```

Bruk begrepet i en beslutning

Bruk «Hva er datalagring?» i en konkret beslutning. Vi kan knytte begrepet til systemet dere bygger eller kjøper. Sikkerhetskopier som er testet i stedet for tatt for gitt, og en skriftlig plan for morgenen da en server ikke kommer opp igjen.

Se tjenestene våre Ta kontakt →

Undersøk antakelsene bak ordet

Bruk forklaringen og det tilknyttede arbeidet til å formulere spørsmål for prosjektet. Vi kan undersøke antakelsene sammen før et verktøy velges. Automatiserte 100 kritiske sikkerhetskopier — 100 kritiske sikkerhetskopier automatisert på tvers av databasene

Automatiserte 100 kritiske sikkerhetskopier Ta kontakt →

Hvorfor det er viktig

Mediet avgjør hvordan data går tapt. En harddisk, en SSD og et SD-kort svikter hver på sin måte, og en gjenopprettingsmetode som passer for det ene, er feil verktøy for det neste. Redundans inne i én boks, som et speilet diskpar eller et RAID-oppsett, holder data lesbare etter at én disk dør, men den skriver hver sletting til hver disk like trofast som den skriver dataene.

Holdbarhet og tilgjengelighet er to forskjellige løfter. Holdbarhet er sjansen for at lagrede data overlever, og tilgjengelighet er sjansen for at de kan leses i et gitt øyeblikk. Cloud Storage oppgir en årlig holdbarhet på 99,999999999 % for hver lagringsklasse, mens den typiske månedlige tilgjengeligheten tjenesten gir, går fra 99,9 % til over 99,99 %, avhengig av klasse og plassering. En lagringsklasse er metadata på hvert objekt som bestemmer tilgjengeligheten og prismodellen, og nye objekter arver standardklassen til bøtta si. De kaldere klassene gir en lavere

pris for å oppbevare data og tar betalt på andre måter: litt lavere tilgjengelighet, en minste lagringstid på 30, 90 eller 365 dager og en avgift for å lese dataene tilbake. Google retter dem mot data som leses én gang i måneden eller sjeldnere, høyst én gang i kvartalet og sjeldnere enn én gang i året.

Plassering er en del av dataene. En kopi på den samme disken, den samme maskinen eller i den samme skykontoen deler skjebne med den disken, maskinen eller kontoen, og en kopi et annet sted gjør det ikke. Plasseringen styrer også fart og kostnad: Google påpeker at data som ligger i samme region som datakraften som bruker dem, gir best ytelse for dataintensivt arbeid og kan gi lavere nettverkskostnader.

Måten programvaren når lagringen på, former resten. Restic, et program for sikkerhetskopiering, definerer sikkerhetskopirepoet sitt som ganske enkelt en mappe med et sett undermapper og filer, og den samme mappen kan ligge på en lokal disk, på en server som nås over SFTP, eller i et objektlager som Google Cloud Storage eller Amazon S3. Bytene er de samme, mens latensen, måtene ting svikter på og hvem som styrer kontoen, alle endrer seg.

Slik ser det ut i praksis

Nedenfra og opp i stakken berørte grunnleggerens tidligere arbeid det fysiske mediet som svikter, og objektlageret som flytter hele filer. Selskapets egen implementering av sikkerhetskopiering beskriver et automatisert opplegg for en kopi hos en annen leverandør som beskyttelse mot tapet av en vert; å aktivere den lagringen og verifisere en gjenoppretting i drift er separate steg.

Mediet: datagjenoppretting ble utført på så å si alle vanlige lagringstyper, fra SD-kort og roterende harddisker til SSD-er, RAID-oppsett, eksterne disker og Mac-systemer. Folk kom med lagringsmedier som hadde sviktet eller tatt skade, og som holdt bilder, forretningsfiler eller den eneste kopien av noe som betydde noe. Hvert medium svikter og gir fra seg dataene sine på sin egen måte: en SSD er et annet problem enn en disk med plater, et RAID-oppsett er et annet problem igjen, og et filsystem på en Mac har sine egne særheter. Å kjenne riktig framgangsmåte for akkurat det mediet betydde like mye som noen enkelt teknikk, og kundene fikk tilbake kritiske data fra enheter de allerede hadde avskrevet.

Objektlageret: i seks ETL/ELT-pipeliner som matet Google BigQuery, stod en bøtte i Google Cloud Storage mellom kilden og datavarehuset. Kuraterte datasett og tabeller ble eksportert til mellomformater, komprimert med gzip og overført til bøtta, og derfra lastet bq CLI og Python-skript dem inn i BigQuery. Hele komprimerte filer gikk inn, og hele filer kom ut, akkurat det mønsteret objektlagring er bygd for.

Kopien et annet sted inngår i selskapets implementering av sikkerhetskopiering og gjenopprettingstester. En Ansible-rolle installerer restic-skript og systemd-enheter for en databasedump der dette er konfigurert, et kryptert, deduplisert øyeblikksbilde over SFTP til lagring hos en annen leverandør, opprydding etter oppbevaringsregelen og en integritetssjekk. En separat månedlig arbeidsflyt gjenoppretter en kjent fil til et midlertidig arbeidsområde og sammenligner bytene. Dermed testes tilbakeføringen av data, ikke bare skrivingen. Sikkerhetskopiering og gjenoppretting har separate signaler ved vellykket kjøring, og timerne krever et konfigurert sikkerhetskopirepo. Implementeringsdokumentasjonen datert 2026-10-27 sier at disse timerne var deaktivert uten et repo, og dokumenterer ingen fullført gjenoppretting i drift. Den dokumenterer automatiseringen og hva som må aktiveres; den beviser ikke at selskapet allerede kan miste verten og gjenopprette dataene sine.

Hvor det går galt

Lagring svikter sjelden der noen ser etter. Som regel er det en kopi, en sjekk eller en kostnad som virket i orden fram til dagen den ble satt på prøve.

En kopi som deler skjebne med originalen. Et øyeblikksbilde i den samme kontoen, en sikkerhetskopi på den samme disken, en replika i den samme bygningen: hver av dem overlever de små feilene og går under med den store. Skriv opp hver enkelthendelse som kan ta med seg begge, og flytt én kopi til listen er kort.

Redundans som telles som en sikkerhetskopi. Et speil kopierer hver skriving til hver disk, slettinger og korrupte data inkludert. Det redder deg fra en død disk. En egen kopi med sin egen historikk er det som redder deg fra en feil.

Et objektlager som behandles som et filsystem. Kode skrevet for filer regner med billige navnebytter og raske mappelister. Noen verktøy utfører navnebytter for objekter som en kopi pluss en sletting, selv om Cloud Storage også støtter atomiske navnebytter innenfor én bøtte. Google påpeker at dypt nestede,

mappelignende strukturer i et flatt navnerom mangler listeytelsen til et ekte filsystem. Legg opp objektnavnene etter måten de skal leses på, og hold arbeid som endrer filer der de står, på fil- eller blokklagring.

En sjekk som leser indeksen og hopper over dataene. Som standard sjekker restic-kommandoen `check` strukturen i sikkerhetskopirepoet og lar pakkefilene som holder de sikkerhetskopierte dataene, ligge uleste, fordi det å lese dem betyr å laste ned hver eneste en. `--read-data` leser alle, og `--read-data-subset=1/5` til og med `5/5` dekker dem alle over fem kjøringer. En sjekk som bare lister filer, beviser at navnene finnes, og sier lite om bytene bak dem.

En kald klasse som holder varme data. En fil som leses daglig fra en klasse priset for månedlige lesinger, betaler en avgift hver gang den leses, så velg klassen etter lesemønsteret dataene faktisk har.

Les mer

- NIST glossary: storage — de to betydningene av ordet slik NIST SP 800-88r2 definerer dem.
- About Cloud Storage objects — data og metadata i et objekt, regler for navn, uforanderlighet og det flate navnerommet bak simulerte mapper.
- Cloud Storage: Storage classes — holdbarhet, tilgjengelighet, minste lagringstid og avgifter for å hente data for hver klasse.
- restic: Preparing a new repository — hva et sikkerhetskopirepo er, og de lokale, SFTP-baserte og objektbaserte lagrene det kan ligge på.
- restic: Working with repositories — å sjekke strukturen i et sikkerhetskopirepo og lese tilbake dataene det holder.
- Cloud Storage: Flytt objekter — Atomisk navnebytte innenfor én bøtte, til forskjell fra verktøy som simulerer flytting ved å kopiere og slette.

Ordliste

Velg ut fra deres faktiske rammer

Ta med utrullingsløpet, driftsrisikoene og dem som har vakt. Vi kan avgrense en endring med kontroller, en vei tilbake og tydelig ansvar.

Se hva det koster Ta kontakt →

Gjør begrepene praktiske

Spør hvordan «Hva er datalagring?» gjelder i organisasjonen deres. Et konkret eksempel fra dagens system er nok til å starte samtalen.

Se tjenestene våre

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

Ta med deg denne siden

Visittkort

Laget med kjærlighet av D.

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

Opphavsrett © 2025 – nå · Engineer ApS · Alle rettigheter forbeholdt.

Skrevet ut fra https://engineer.company/nb/glossary/storage/