Engineer

Datalagring

at gemme data til senere brug, eller medier, der rummer dem

Hvad det er

Datalagring, i daglig tale ofte bare lagring, er det lag i et system, hvor data ligger mellem brugene, på medier, der bevarer dem gennem en genstart og et strømsvigt. NIST's ordliste giver ordet to betydninger, begge hentet fra SP 800-88r2: en bevaring af data, der lader dem hente frem igen, og det hardware, som data kan skrives til og hentes fra.

Al lagring hviler på et fysisk medie: en roterende harddisk, et SSD-drev, et SD-kort, et bånd. Software ser et af tre lag oven over det:

- **Bloklagring** viser en enhed som en række blokke af fast størrelse, adresseret med nummer. En diskpartition, et RAID-array og en virtuel disk i skyen ser alle sådan ud for kernen, og et filsystem eller en databasemotor bestemmer, hvad blokkene betyder.
- **Fillagring** lægger et filsystem over blokkene: navne, mapper, størrelser, ejere og rettigheder, og en fil kan ændres nogle få byte ad gangen. Delt over NFS eller SMB betjener ét mappetræ mange maskiner.
- **Objektlagring** gemmer hele objekter, hvert en række byte plus metadata, under et navn i en beholder og nås gennem et HTTP-API i stedet for en montering. Google Cloud Storage gemmer data som objekter i beholdere kaldet buckets, hver bucket hører til et projekt, og et objekt er et uforanderligt stykke data, der består af en fil i et vilkårligt format.

De tre skilles ved, hvad en ændring koster. En blokenhed omskriver én blok på stedet. Et filsystem retter en del af en fil og flytter den med en omdøbning. Et objektlager erstatter hele objektet: Googles side om objekter udelukker tilføjelse og afkortning, og en erstatning sker atomart, så læsere får den gamle version, indtil den nye upload er færdig. Afhængigt af værktøjet kan en omdøbning betyde en kopi under det nye navn og en sletning af originalen; Cloud Storage understøtter også atomare omdøbninger af objekter inden for én bucket. I en

bucket med fladt navnerum er mapper et billede, som værktøjerne tegner ud fra skråstreger i objektnavnene; strengen `folder1/` er en del af navnet, og bucketen gemmer slet ingen mappe.

Den samme fil ser sådan ud i hvert af de tre lag:

```
# blok: kernens syn på en disk som nummererede blokke
lsblk /dev/sda

# fil: et filsystem monteret over blokkene, med navne og mapper
df -h /srv

# objekt: et helt objekt sendt til en bucket over et API
gcloud storage cp readings.csv.gz gs://example-bucket/2024/01/
```

Brug begrebet i en beslutning

Brug »Hvad er datalagring?« i en konkret beslutning. Vi kan knytte begrebet til det system, I bygger eller køber. Sikkerhedskopier, der er testet og ikke bare antaget at virke, og en nedskrevet plan for den morgen, en server ikke starter igen.

Se vores ydelser Kontakt os →

Undersøg antagelserne bag ordet

Brug forklaringen og det tilknyttede arbejde til at formulere spørgsmål til jeres projekt. Vi kan undersøge antagelserne sammen, før et værktøj vælges. Har automatiseret 100 kritiske databackups — 100 kritiske backups af databaserne automatiseret

Har automatiseret 100 kritiske databackups Kontakt os →

Hvorfor det betyder noget

Mediet afgør, hvordan data går tabt. En harddisk, et SSD-drev og et SD-kort svigter hver på sin måde, og en gendannelsesmetode, der passer til det ene, er det forkerte værktøj til det næste. Redundans inde i én kasse, som et spejlet diskpar eller et RAID-array, holder data læsbare, når en enkelt disk dør, men den skriver hver sletning til hver disk lige så trofast som selve dataene.

Holdbarhed og tilgængelighed er to forskellige løfter. Holdbarhed er chancen for, at lagrede data overlever; tilgængelighed er chancen for, at de kan læses i et givent øjeblik. Cloud Storage angiver en årlig holdbarhed på 99,999999999 % for alle lagerklasser, mens den typiske månedlige tilgængelighed går fra 99,9 % til over 99,99 % afhængigt af klasse og placering. En lagerklasse er metadata på hvert objekt, som fastlægger dets tilgængelighed og prismodel, og nye objekter arver deres buckets standardklasse. De koldere klasser bytter en lavere pris for at opbevare data mod lidt lavere tilgængelighed, en minimumslagringsperiode på

30, 90 eller 365 dage og et gebyr for at læse dataene igen; Google sigter dem mod data, der læses én gang om måneden eller sjældnere, højst én gang i kvartalet og sjældnere end én gang om året.

Placering er en del af dataene. En kopi på samme disk, samme maskine eller samme cloudkonto deler skæbne med den, og en kopi et andet sted gør det ikke. Placering sætter også hastighed og pris: Google bemærker, at data i samme region som den beregning, der bruger dem, giver den bedste ydelse til dataintensivt arbejde og kan sænke netværksomkostningerne.

Måden, software når lageret på, former resten.
Sikkerhedskopieringsprogrammet restic beskriver sit repository som blot en mappe med et sæt undermapper og filer, og den samme mappe kan ligge på en lokal disk, på en server, der nås over SFTP, eller i et objektlager som Google Cloud Storage eller Amazon S3. Dataene er de samme, mens latens, fejlmåder og den person, der styrer kontoen, alle skifter.

Sådan ser det ud i praksis

Fra bunden af stakken og opefter omfattede grundlæggerens tidligere arbejde det fysiske medie, der svigter, og objektlageret, der flytter hele filer. Virksomhedens egen backupimplementering beskriver automatisering, der skal sørge for en kopi hos en anden udbyder som beskyttelse mod tabet af en vært; aktivering af det lager og kontrol af en reel gendannelse er særskilte trin.

Nederst ligger det fysiske medie, og datagendannelsen omfattede stort set alle almindelige lagertyper, fra SD-kort og roterende harddiske til SSD'er, RAID-arrays, eksterne drev og Mac-systemer. Folk kom med lagermedier, der var gået i stykker eller beskadiget, med billeder, forretningsfiler, den eneste kopi af noget, der betød noget. Hvert medie går i stykker og giver sine data fra sig på sin egen måde: en SSD er et andet problem end et pladedrev, et RAID-array er et andet problem igen, og et Mac-filsystem har sine egne særheder. At kende den rette tilgang til det konkrete medie betød lige så meget som nogen enkelt teknik, og kunderne fik kritiske data tilbage fra enheder, de allerede havde afskrevet som tabte.

Et lag højere ligger objektlageret. I seks ETL/ELT-pipelines, der fødte data ind i Google BigQuery, stod en Google Cloud Storage-bucket mellem kilden og datavarehuset. Kuraterede datasæt og tabeller blev eksporteret til

mellemformater, komprimeret med gzip og overført til bucketen, og derfra indlæste bq CLI og Python-scripts dem i BigQuery. Hele komprimerede filer gik ind, og hele filer kom ud, det mønster objektlagring er bygget til.

Kopien et andet sted indgår i virksomhedens implementering af sikkerhedskopiering og gendannelsestest. En Ansible-rolle installerer restic-scripts og systemd-enheder til et databasedump, hvor det er konfigureret, et krypteret, deduplikeret snapshot over SFTP til lager hos en anden udbyder, udtynding efter opbevaringspolitikken og et integritetstjek. Et separat månedligt workflow gendanner en kendt fil til et midlertidigt arbejdsområde og sammenligner dens bytes, så vejen tilbage bliver afprøvet og ikke kun skrivningen. Sikkerhedskopiering og gendannelse har hver sit successignal, og timerne kræver et konfigureret backuplager. Implementeringsbeskrivelsen dateret 2026-10-27 angiver, at timerne var deaktiveret uden et backuplager, og dokumenterer ingen gennemført gendannelse i drift. Den dokumenterer automatiseringen og det, der skal aktiveres; den beviser ikke, at virksomheden allerede kan miste værten og gendanne sine data.

Hvor det går galt

Det meste besvær med lagring kommer fra en kopi, et tjek eller en pris, der så fin ud indtil den dag, den skulle bruges.

En kopi deler skæbne med originalen. Et snapshot på samme konto, en sikkerhedskopi på samme disk, en replika i samme bygning: hver af dem overlever de små svigt og går ned med det store. Skriv de enkelte hændelser op, der kunne tage begge, og flyt én kopi, til listen er kort.

Redundans tælles som sikkerhedskopi. Et spejl kopierer hver skrivning til hver disk, sletninger og korruption medregnet. Det svarer på en død disk; en separat kopi med sin egen historik svarer på en fejltagelse.

Et objektlager bruges som filsystem. Kode skrevet til filer forventer billige omdøbninger og mappelister. Nogle værktøjer udfører omdøbninger af objekter som en kopi plus en sletning, selv om Cloud Storage også understøtter atomare omdøbninger inden for én bucket. Google bemærker, at dybt indlejrede, mappeagtige strukturer i et fladt navnerum mangler den ydelse ved listning, som et rigtigt filsystem har. Læg objektnavnene an efter den måde, de skal læses på, og hold arbejde, der retter filer på stedet, på fil- eller bloklagring.

Tjekket læser indekset og springer dataene over. Som standard verificerer restics `check` -kommando repositoryets struktur og lader de pack-filer, der rummer de sikkerhedskopierede data, være ulæste, fordi det kræver, at hver eneste af dem hentes. `--read-data` læser dem alle, og `--read-data-subset=1/5` til `5/5` dækker repositoryet på fem kørsler. Et tjek, der kun lister filer, beviser, at navnene findes, og siger kun lidt om indholdet bag dem.

En kold klasse rummer varme data. En fil, der læses dagligt fra en klasse prissat til månedlige læsninger, betaler et gebyr ved hver læsning, så afstem klassen efter det læsemønster, dataene faktisk har.

Læs mere

- NIST glossary: storage — ordets to betydninger, som NIST SP 800-88r2 definerer dem.
- About Cloud Storage objects — objektdata og metadata, navneregler, uforanderlighed og det flade navnerum bag simulerede mapper.
- Cloud Storage: Storage classes — holdbarhed, tilgængelighed, minimumslagringsperioder og gebyrer for hentning for hver klasse.
- restic: Preparing a new repository — hvad et repository er, og de lokale, SFTP- og objektlagerbackends, det kan ligge på.
- restic: Working with repositories — at tjekke et repositorys struktur og læse de data, det rummer, tilbage.
- Cloud Storage: Flyt objekter — Den atomare omdøbning inden for én bucket, til forskel fra værktøjer, der simulerer en flytning ved kopiering og sletning.

Ordliste

Vælg ud fra jeres faktiske vilkår

Medbring udrulningsforløbet, driftsrisiciene og dem, der har vagten. Vi kan afgrænse en ændring med kontroller, en vej tilbage og tydeligt ansvar.

Se, hvad det koster Kontakt os →

Gør begreberne praktiske

Spørg, hvordan »Hvad er datalagring?« gælder i jeres organisation. Et konkret eksempel fra jeres nuværende system er nok til at begynde samtalen.

Se vores ydelser

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

Tag denne side med dig

Visitkort

Lavet med kærlighed af D.

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

Copyright © 2025 – nu · Engineer ApS · Alle rettigheder forbeholdes.

Udskrevet fra https://engineer.company/da/glossary/storage/