Engineer

Зберігання даних

утримання даних так, щоб їх можна було згодом прочитати

Що це таке

Зберігання даних, часто просто зберігання, — це шар системи, де дані перебувають між використаннями, на носіях, що переживають перезавантаження й знеструмлення. Глосарій NIST дає слову два значення, обидва взяті з SP 800-88r2: збереження даних, придатне для подальшого видобування, і апаратне забезпечення, на яке дані можна записати й з якого їх можна прочитати.

Будь-яке зберігання спирається на фізичний носій: жорсткий диск із пластинами, твердотільний накопичувач, SD-картку, стрічковий картридж. Програми бачать над ним один із трьох шарів:

- **Блочне зберігання** подає пристрій як ряд блоків фіксованого розміру, адресованих номерами. Розділ диска, масив RAID і віртуальний диск у хмарі виглядають для ядра саме так, а файлова система чи рушій бази даних вирішує, що означають блоки.
- **Файлове зберігання** кладе на ці блоки файлову систему: імена, каталоги, розміри, власників і права доступу, а файл можна змінювати по кілька байтів. Через NFS чи SMB одне дерево каталогів обслуговує багато машин.
- **Об’єктне зберігання** тримає цілі об’єкти, кожен — послідовність байтів плюс метадані, під іменем у контейнері, і до нього звертаються через HTTP API, а не через монтування. Google Cloud Storage зберігає дані як об’єкти в контейнерах, які називаються бакетами, кожен бакет належить до проєкту, а об’єкт — це незмінний фрагмент даних, що складається з файлу будь-якого формату.

Три шари розходяться в тому, скільки коштує зміна. Блочний пристрій переписує один блок на місці. Файлова система редагує частину файлу й переносить його перейменуванням. Об’єктне сховище замінює об’єкт

цілком: сторінка Google про об’єкти виключає дописування й обрізання, а заміна відбувається атомарно, тож читачі отримують стару версію, доки нове завантаження не завершиться. Залежно від інструмента перейменування може означати копію під новим іменем і видалення оригіналу; Cloud Storage також підтримує атомарне перейменування об’єктів у межах одного бакета. У бакеті з пласким простором імен папки — це картинка, яку інструменти малюють зі скісних рисок в іменах об’єктів; рядок `folder1/` є частиною імені, а окремої папки бакет не зберігає.

Той самий файл у кожному з трьох шарів:

```
# блок: ядро бачить диск як пронумеровані блоки
lsblk /dev/sda

# файл: файлова система над блоками, з іменами й каталогами
df -h /srv

# об'єкт: цілий об'єкт, надісланий у бакет через API
gcloud storage cp readings.csv.gz gs://example-bucket/2024/01/
```

Застосуйте поняття до рішення

Застосуйте «Що таке зберігання даних?» до конкретного рішення. Ми можемо пов'язати поняття із системою, яку ви створюєте або купуєте. Резервні копії, які перевірено, а не просто вважають робочими, і записаний план на той ранок, коли сервер не ввімкнеться.

Переглянути наші послуги Зв'язатися з нами →

Перевірте припущення за терміном

Використайте пояснення та пов'язані роботи, щоб сформулювати запитання для проєкту. Ми можемо перевірити припущення разом до вибору інструмента. Автоматизувала 100 критичних резервних копій — 100 критичних резервних копій автоматизовано в усіх базах

Автоматизувала 100 критичних резервних копій Зв'язатися з нами →

Чому це важливо

Носій визначає, як дані втрачаються. Жорсткий диск, SSD і SD-картка ламаються кожен по-своєму, і метод відновлення, що підходить одному, стає хибним інструментом для наступного. Надлишковість усередині одного корпусу, як-от дзеркальна пара дисків чи масив RAID, зберігає дані придатними для читання, коли помирає один диск, але записує кожне видалення на кожен диск так само сумлінно, як і самі дані.

Довговічність і доступність — дві різні обіцянки. Довговічність — це шанс, що збережені дані вціліють; доступність — шанс, що їх вдасться прочитати в певну мить. Cloud Storage заявляє річну довговічність 99,999999999% для всіх класів зберігання, тоді як типова місячна доступність становить від 99,9% до понад 99,99% залежно від класу й розташування. Клас зберігання — це метадані кожного об'єкта, що задають його доступність і модель ціноутворення, а нові об'єкти успадковують клас бакета за замовчуванням. Холодніші класи міняють нижчу ціну зберігання на трохи нижчу доступність, мінімальний строк зберігання 30, 90 або 365 днів і плату за повторне читання даних; Google призначає їх для даних, які читають раз на місяць або рідше, щонайбільше раз на квартал і рідше ніж раз на рік.

Розташування — частина даних. Копія на тому самому диску, тій самій машині чи в тому самому хмарному обліковому записі поділяє з ними долю, а копія деінде — ні. Розташування задає також швидкість і вартість: Google зазначає, що дані в тому самому регіоні, що й обчислення, які їх використовують, дають найкращу продуктивність для інтенсивної роботи з даними й можуть зменшити плату за мережу.

Спосіб, у який програма дістається до сховища, визначає решту. Програма резервного копіювання restic описує свій репозиторій як просто каталог із набором підкаталогів і файлів, і той самий каталог може лежати на локальному диску, на сервері, до якого звертаються через SFTP, або в об'єктному сховищі на кшталт Google Cloud Storage чи Amazon S3. Дані лишаються тими самими, а затримка, способи відмови й людина, яка керує обліковим записом, змінюються.

Як це виглядає на практиці

Якщо розглядати стек знизу вгору, попередня робота засновниці охоплювала фізичний носій, що виходить із ладу, та об'єктне сховище, що переміщує цілі файли. Власна реалізація резервного копіювання компанії описує автоматизацію створення копії в іншого провайдера як захист від втрати хоста; активація цього сховища й перевірка реального відновлення — окремі кроки.

На рівні фізичного носія відновлення даних охоплювало практично всі поширені типи носіїв, від SD-карток і класичних HDD до SSD, масивів RAID, зовнішніх дисків і систем Mac. Люди приносили носії, які вийшли з ладу або були пошкоджені, а на них — фотографії, бізнес-файли, єдину копію чогось

по-справжньому важливого. Кожен носій по-своєму ламається і по-своєму віддає дані: SSD — зовсім інша задача, ніж диск із пластинами, масив RAID — ще одна, а файлова система Mac має свої особливості. Знати правильний підхід до конкретного носія було не менш важливо, ніж володіти будь-якою окремою технікою, і клієнти отримували назад критично важливі дані з пристроїв, які вже подумки списали як втрачені.

Шар об'єктного сховища видно в конвеєрах даних: у шести ETL/ELT-пайплайнах, що постачали дані в Google BigQuery, між джерелом і BigQuery стояв бакет Google Cloud Storage. Впорядковані набори даних і таблиці експортували в проміжні формати, стискали gzip і передавали в бакет, а звідти bq CLI і скрипти на Python завантажували їх у BigQuery. Цілі стиснуті файли заходили, цілі файли виходили — саме для такого шаблону й створено об'єктне зберігання.

Копія в іншому місці є частиною реалізації резервного копіювання та тестового відновлення компанії. Роль Ansible встановлює скрипти restic і юніти systemd для дампа бази даних, якщо його налаштовано, надсилання зашифрованого знімка з дедуплікацією до сховища іншого провайдера через SFTP, очищення за політикою зберігання та перевірки цілісності. Окремий щомісячний процес відновлює відомий файл у тимчасове місце й порівнює його байти, перевіряючи шлях зворотного читання, а не лише запису. Резервне копіювання й відновлення мають окремі сигнали про успіх, а таймерам потрібен налаштований репозиторій. У записі про реалізацію від 2026-10-27 зазначено, що ці таймери були вимкнені за відсутності репозиторію; жодного завершеного реального відновлення не зафіксовано. Запис документує автоматизацію та те, що потрібно активувати, але не доводить, що компанія вже може втратити хост і відновити свої дані.

Де виникають проблеми

Більшість проблем зі зберіганням походить від копії, перевірки чи ціни, які здавалися надійними аж до дня, коли знадобилися.

Копія поділяє долю з оригіналом. Знімок у тому самому обліковому записі, резервна копія на тому самому диску, репліка в тій самій будівлі: кожна переживає дрібні збої й гине разом із великим. Випишіть окремі події, що могли б знищити обидві копії, і переносьте одну, доки список не стане коротким.

Надлишковість вважають резервною копією. Дзеркало копіює кожен запис на кожен диск, разом із видаленнями й пошкодженнями. Воно рятує від мертвого диска; окрема копія з власною історією рятує від помилки.

Об'єктним сховищем користуються як файловою системою. Код, написаний для файлів, очікує дешевих перейменувань і переліків каталогів. Деякі інструменти реалізують перейменування об'єктів як копію плюс видалення, хоча Cloud Storage також підтримує атомарне перейменування в межах одного бакета. Google зазначає, що глибоко вкладені структури, схожі на каталоги, у пласкому просторі імен поступаються рідній файловій системі у швидкості переліку. Добирайте імена об'єктів під те, як їх читатимуть, а роботу, що редагує файли на місці, лишіть на файловому чи блочному зберіганні.

Перевірка читає індекс і пропускає дані. За замовчуванням команда restic `check` перевіряє структуру репозиторію й лишає непрочитаними pack-файли з самими даними резервних копій, бо для цього довелося б завантажити кожен із них. `--read-data` читає їх усі, а підмножини від `--read-data-subset=1/5` до `5/5` охоплюють репозиторій за п'ять запусків. Перевірка, яка лише перелічує файли, доводить, що імена існують, і мало що каже про байти за ними.

Холодний клас тримає гарячі дані. Файл, який щодня читають із класу, розрахованого на щомісячне читання, платить за кожне читання, тож добирайте клас під те, як дані насправді читають.

Що почитати далі

- NIST glossary: storage — два значення слова так, як їх визначає NIST SP 800-88r2.
- About Cloud Storage objects — дані й метадані об'єктів, правила іменування, незмінність і пласкі простори імен за імітованими папками.
- Cloud Storage: Storage classes — довговічність, доступність, мінімальні строки зберігання й плата за отримання для кожного класу.
- restic: Preparing a new repository — що таке репозиторій і локальні, SFTP та об'єктні бекенди, на яких він може лежати.
- restic: Working with repositories — перевірка структури репозиторію й зворотне читання даних, які він містить.

- Cloud Storage: переміщення об’єктів — Атомарне перейменування в межах одного бакета, на відміну від інструментів, які імітують переміщення копіюванням і видаленням.

Словник

Обирайте за фактичними умовами

Опишіть шлях розгортання, експлуатаційні ризики та чергову команду. Ми можемо визначити зміну з перевірками, шляхом відновлення та чіткою відповідальністю.

Дивитися, скільки це коштує Зв'язатися з нами →

Зробіть термінологію практичною

Запитайте, як «Що таке зберігання даних?» стосується вашої організації. Конкретного прикладу з поточної системи достатньо для початку розмови.

Переглянути наші послуги

welcome@engineer.company +45 50 10 61 84 Копенгаген, Данія

Візьміть цю сторінку із собою

Візитівка

Дбайливо зроблено з любов'ю — D.

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

Авторське право © 2025 – дотепер · Engineer ApS · Усі права захищено.

Надруковано з https://engineer.company/uk/glossary/storage/