Engineer

Stockage de données

la conservation de données pour relecture, ou leurs supports

Ce que c’est

Le stockage de données, généralement abrégé en stockage, est la couche d’un système où les données résident entre deux utilisations, sur des supports qui les conservent après un redémarrage ou une coupure de courant. Le glossaire du NIST donne au mot deux sens, tous deux tirés de SP 800-88r2 : la conservation de données permettant leur récupération et le matériel sur lequel des données peuvent être inscrites puis relues.

Chaque forme repose sur un support physique : un disque dur à plateaux, un SSD, une carte SD, une bande. Le logiciel voit au-dessus de ce support l’une des trois couches suivantes :

- **Le stockage en blocs** présente un périphérique comme une suite de blocs de taille fixe, adressés par numéro. Une partition de disque, un ensemble RAID et un disque virtuel cloud se présentent tous ainsi au noyau ; un système de fichiers ou un moteur de base de données décide du sens de ces blocs.
- **Le stockage de fichiers** place un système de fichiers au-dessus de ces blocs : noms, répertoires, tailles, propriétaires et permissions. Un fichier peut être modifié quelques octets à la fois. Partagée par NFS ou SMB, une même arborescence dessert plusieurs machines.
- **Le stockage d’objets** conserve des objets entiers, chacun constitué d’une suite d’octets et de métadonnées, sous un nom dans un conteneur. Il est accessible par une API HTTP plutôt que par un montage. Google Cloud Storage conserve les données sous forme d’objets dans des conteneurs appelés buckets ; chaque bucket appartient à un projet, et chaque objet contient un fichier de n’importe quel format sous forme de données immuables après écriture.

Les trois approches divergent sur le coût d'une modification. Un périphérique de blocs réécrit un bloc sur place. Un système de fichiers modifie une partie d'un fichier et le déplace par renommage. Un stockage d'objets remplace l'objet entier : la page de Google sur les objets exclut l'ajout de données à la fin d'un objet et sa troncature, et le remplacement est atomique. Les lecteurs reçoivent donc l'ancienne version jusqu'à la fin du nouveau téléversement. Selon l'outil, renommer peut consister à copier sous le nouveau nom puis à supprimer l'original ; Cloud Storage prend aussi en charge le renommage atomique d'objets dans un même bucket. Dans un bucket à espace de noms plat, les dossiers sont une représentation créée par les outils à partir des barres obliques des noms d'objets : la chaîne `folder1/` fait partie du nom, et le bucket ne contient aucun dossier.

Le même fichier se présente ainsi à chaque couche :

```
# 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/
```

Utilisez le concept pour prendre une décision

Utilisez « Qu'est-ce que le stockage de données ? » pour éclairer une décision réelle. Nous pouvons vous aider à relier ce concept au système que vous construisez ou achetez. Des sauvegardes testées plutôt que supposées fiables, et un plan écrit pour le matin où un serveur ne redémarre pas.

Découvrir nos prestations Contactez-nous →

Vérifiez les hypothèses derrière le terme

Appuyez-vous sur l'explication et les réalisations liées pour formuler les questions de votre projet. Nous pouvons examiner les hypothèses avec vous avant le choix d'un outil. 100 sauvegardes critiques automatisées : Barman — 100 sauvegardes critiques automatisées dans les bases

100 sauvegardes critiques automatisées : Barman Contactez-nous →

Pourquoi c'est important

Le support détermine comment les données se perdent. Un disque dur, un SSD et une carte SD tombent chacun en panne à leur manière, et une méthode de récupération adaptée à l'un est le mauvais outil pour l'autre. La redondance dans un seul boîtier, comme une paire en miroir ou un ensemble RAID, garde les données lisibles après la panne d'un disque, mais reproduit chaque suppression sur tous les disques aussi fidèlement qu'elle y écrit les données.

Durabilité et disponibilité sont deux promesses distinctes. La durabilité est la probabilité que les données conservées survivent ; la disponibilité, celle de pouvoir les lire à un moment donné. Cloud Storage annonce une durabilité annuelle de 99,999999999 % pour chaque classe de stockage, tandis que la disponibilité mensuelle typique indiquée va de 99,9 % à plus de 99,99 %, selon la classe et l'emplacement. Une classe de stockage est une métadonnée de chaque objet qui fixe sa disponibilité et son modèle tarifaire, et les nouveaux objets héritent de la classe par défaut du bucket. Les classes plus froides réduisent le coût de conservation, en contrepartie d'une disponibilité légèrement inférieure, d'une durée minimale de stockage de 30, 90 ou 365 jours et de frais de lecture. Google les destine respectivement aux données lues une fois par mois ou moins, au plus une fois par trimestre et moins d'une fois par an.

L'emplacement fait partie des données. Une copie sur le même disque, la même machine ou le même compte cloud partage leur sort ; une copie ailleurs, non. L'emplacement détermine aussi la vitesse et le coût : Google indique que conserver les données dans la même région que les ressources de calcul qui les utilisent maximise les performances des traitements intensifs en données et peut réduire les frais réseau.

La manière dont un logiciel accède au stockage détermine le reste. Restic, un programme de sauvegarde, définit son dépôt comme un simple répertoire contenant des sous-répertoires et des fichiers. Ce même répertoire peut se trouver sur un disque local, sur un serveur atteint par SFTP ou dans un stockage d'objets tel que Google Cloud Storage ou Amazon S3. Les octets restent identiques tandis que la latence, les modes de panne et la personne qui contrôle le compte changent.

À quoi cela ressemble en pratique

Si l'on considère la pile de bas en haut, les travaux antérieurs de la fondatrice ont porté sur le support physique qui tombe en panne et sur le stockage d'objets qui déplace des fichiers entiers. L'implémentation de sauvegarde propre à l'entreprise prévoit l'automatisation d'une copie chez un autre fournisseur pour se protéger de la perte d'un hôte ; activer ce stockage et vérifier une restauration réelle sont des étapes distinctes.

Le support : la récupération de données a couvert pratiquement tous les types de supports courants, des cartes SD et disques durs à plateaux aux SSD, ensembles RAID, disques externes et systèmes Mac. Des personnes apportaient des

supports en panne ou endommagés contenant des photos, des fichiers professionnels, l'unique copie de quelque chose d'important. Chaque support tombe en panne et restitue ses données différemment : un SSD pose un autre problème qu'un disque à plateaux, un ensemble RAID en pose encore un autre et un système de fichiers Mac a ses particularités. Connaître la bonne approche pour le support précis comptait autant que maîtriser une technique donnée, et les clients ont récupéré des données essentielles sur des appareils qu'ils considéraient déjà comme perdus.

Le stockage d'objets : dans six pipelines ETL/ELT alimentant Google BigQuery, un bucket Google Cloud Storage se trouvait entre la source et l'entrepôt. Des jeux de données et tables préparés étaient exportés dans des formats intermédiaires, compressés avec gzip et transférés dans le bucket. Depuis celui-ci, la CLI bq et des scripts Python les chargeaient dans BigQuery. Des fichiers compressés entiers entraient et des fichiers entiers ressortaient : c'est le mode d'utilisation pour lequel le stockage d'objets est conçu.

La copie ailleurs fait partie de l'implémentation de sauvegarde et de test de restauration de l'entreprise. Un rôle Ansible installe des scripts restic et des unités systemd pour produire un dump de base de données lorsque celui-ci est configuré, envoyer par SFTP un instantané chiffré et dédupliqué vers un stockage chez un autre fournisseur, élaguer les sauvegardes selon la politique de conservation et vérifier l'intégrité. Un workflow mensuel séparé restaure un fichier connu dans un emplacement temporaire et compare ses octets, testant le chemin de relecture plutôt que la seule écriture. Sauvegarde et restauration ont des signaux de réussite distincts, et les timers exigent un dépôt configuré. Le compte rendu d'implémentation daté du 2026-10-27 indique que ces timers étaient désactivés faute de dépôt et ne consigne aucune restauration réelle achevée. Il documente l'automatisation et ce qui doit être activé ; il ne prouve pas que l'entreprise peut déjà perdre son hôte et récupérer ses données.

Où les problèmes surviennent

La plupart des problèmes de stockage viennent d'une copie, d'un contrôle ou d'un coût qui semblaient convenir jusqu'au jour où ils sont devenus déterminants.

Une copie partage le sort de l'original. Un instantané dans le même compte, une sauvegarde sur le même disque, une réplique dans le même bâtiment : chacun résiste aux petites pannes, mais succombe à un incident de grande

ampleur. Énumérer les événements qui pourraient, à eux seuls, détruire les deux et déplacer une copie jusqu'à ce que la liste soit courte.

La redondance est comptée comme une sauvegarde. Un miroir copie chaque écriture sur chaque disque, suppressions et corruptions comprises. Il répond à la panne d'un disque ; une copie séparée, avec son propre historique, répond à une erreur.

Un stockage d'objets est traité comme un système de fichiers. Le code écrit pour des fichiers s'attend à des opérations de renommage et d'énumération de répertoires peu coûteuses. Certains outils réalisent le renommage d'objets par une copie suivie d'une suppression, alors que Cloud Storage prend aussi en charge le renommage atomique dans un même bucket. Google indique que, dans un espace de noms plat, des structures profondément imbriquées ressemblant à des répertoires n'offrent pas les performances d'énumération d'un système de fichiers natif. Organiser les noms d'objets selon la manière dont ils seront lus, et garder les traitements qui modifient les fichiers sur place sur un stockage de fichiers ou de blocs.

Le contrôle lit l'index et saute les données. Par défaut, la commande `check` de restic vérifie la structure du dépôt et ne lit pas les fichiers pack qui contiennent les données sauvegardées, car les lire impose de les télécharger tous. `--read-data` les lit intégralement, et les sous-ensembles de `--read-data-subset=1/5` à `5/5` couvrent le dépôt en cinq exécutions. Un contrôle qui se borne à lister les fichiers prouve l'existence des noms et peu de chose sur les octets qu'ils désignent.

Une classe froide contient des données chaudes. Lire chaque jour un fichier dans une classe tarifée pour des lectures mensuelles entraîne des frais à chaque lecture : choisir la classe en fonction de la fréquence réelle de lecture.

Pour aller plus loin

- NIST glossary : storage — les deux sens du mot tels que les définit NIST SP 800-88r2.
- About Cloud Storage objects — données et métadonnées des objets, règles de nommage, immutabilité et espace de noms plat derrière les dossiers simulés.
- Cloud Storage : Storage classes — durabilité, disponibilité, durées minimales de conservation et frais de récupération propres à chaque classe.

- restic: Preparing a new repository — ce qu'est un dépôt et les backends locaux, SFTP ou de stockage d'objets qui peuvent l'héberger.
- restic: Working with repositories — contrôler la structure d'un dépôt et relire les données qu'il contient.
- Cloud Storage: Move objects — l'opération de renommage atomique dans un même bucket, distincte des outils qui simulent un déplacement par copie puis suppression.

Glossaire

Choisissez selon vos contraintes réelles

Présentez le parcours de déploiement, les risques opérationnels et les personnes d'astreinte. Nous pouvons définir un changement avec des vérifications, une procédure de reprise et une responsabilité claire.

Consulter les tarifs Contactez-nous →

Rendez la terminologie concrète

Demandez comment « Qu'est-ce que le stockage de données ? » s'applique à votre organisation. Un exemple concret tiré de votre système actuel suffit pour engager la discussion.

Découvrir nos prestations

welcome@engineer.company +45 50 10 61 84 Copenhague, Danemark

Emportez cette page avec vous

Fiche de contact

Créé avec soin et amour par D.

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

Droit d'auteur © 2025 – aujourd'hui · Engineer ApS · Tous droits réservés.

Imprimé depuis https://engineer.company/fr/glossary/storage/