

Les différentes modalités de stockage et leurs usages dans un laboratoire de mathématiques

Sylvain Maurin

15 Novembre 2023

Au menu de cette présentation

Éléments de vocabulaire du stockage

Les grandeurs objectives et mesurables

Médias

Typologie

Focus

Impossibilités

Durabilité

Latence

Coûts

Stockages de laboratoire

Mon opinion

Usages

Conclusion

J'émettrai ici des opinions, vous avez le droit de critiquer mon objectivité

...mais ce sera pendant les questions SVP



Dimensions apparentes

- ▶ Coût en EUR (acquisition / location)
- ▶ Espace (brut/net)
- ▶ Débit (séquentiel, aléatoire/indexé)

Dimensions cachées

- ▶ Coût RH

$$F(t) = \cup(\textit{apprentissage}, \textit{acquisition}, \textit{deploiement}, \textit{maintenance})$$

- ▶ Fiabilité, durabilité, accessibilité,
- ▶ Maintenabilité, soutenabilité,
- ▶ Coût de sortie (export)
- ▶ Latence, gigue, taux de foisonnement

EcolInfo ?

Les moyens de stockages de masse

- ▶ Magnétiques (bandes, disques)
- ▶ Mémoires à semi-conducteurs (NVDIMM, SSD)

Typologie des stockages

- ▶ Direct Attached Storage
- ▶ Network Attached Storage
- ▶ Storage Area Network
- ▶ Object-based Storage (stockage en ligne)

Exemple d'impossibilité : l'impasse du stockage distribué à garantir l'ACIDité d'une transaction

"Le théorème CAP ou CDP"¹

aussi connu sous le nom de théorème de Brewer, dit qu'il est impossible sur un système informatique de calcul distribué de garantir en même temps (c'est-à-dire de manière synchrone) les trois contraintes suivantes :

- ▶ *Cohérence (Consistency en anglais) : tous les nœuds du système voient exactement les mêmes données au même moment ;*
- ▶ *Disponibilité (Availability en anglais) : garantie que toutes les requêtes reçoivent une réponse ;*
- ▶ *Tolérance au partitionnement (Partition Tolerance en anglais) : aucune panne moins importante qu'une coupure totale du réseau ne doit empêcher le système de répondre correctement (ou encore : en cas de morcellement en sous-réseaux, chacun doit pouvoir fonctionner de manière autonome).*

Espérance de vie d'un stockage

- ▶ Les stockages locaux
JMathrice mars 2022 IMT
- ▶ Les box (owncloud,nextcloud, seafile, ...) et le maintien des systèmes de fichier
- ▶ Sauvegardes (import/export)
- ▶ FAIR ?

A l'inverse de la fréquence

- ▶ LAN, MAN et attendre ses acquittements dans l'ACID
- ▶ Dans les nuages, une grosse transaction
- ▶ Oubliés dans les bibliothèques, des wagons de livres indexés

Une après-midi avec David Delavennat au printemps 2023 pour du S3 PLM

- ▶ Dans un cloud Privé
 - ▶ S3 multi-DC 252€ / To / an
 - ▶ S3 Uni-DC 120€ /To / an
 - ▶ S3 Sur bande 12€ / To / an
- ▶ On premium SAN/NAS 100€ / To / an
- ▶ Open source et MatInfo5
 - ▶ Distribué 40€ / To / an
 - ▶ Silo dupliqué 20€ / To / an

Le calcul

Apprenez à chercher ailleurs !

Le reste

Des caractéristiques "bureautiques" pour le parc
et des utilisateurs sages, comprenant le mathricien !

Nomadisme

- ▶ Des portables majoritaires dans les parcs informatiques
- ▶ Le télétravail

Délégation institutionnelle

- ▶ De la gestion des stockages
- ▶ Des applications et des modalités d'accès
- ▶ Sensibilité (au sens SSI) raisonnable

Ergonomie

- ▶ Autonomie du chercheur
- ▶ Facilité pour l'ASR

Support dans la communauté

- ▶ liste stockage, GT(CEPH,ZFS)
- ▶ JMathrices

Propositions PLM

Migration d'une plateforme matérielle répartie vers les centres GRICAD et VirtualData

- ▶ Avant : Disque (ssh), mail
- ▶ En migration : PLM Box, gitlab
- ▶ Dans le cloud plmshift/plmapps + gitlab

A venir, S3...