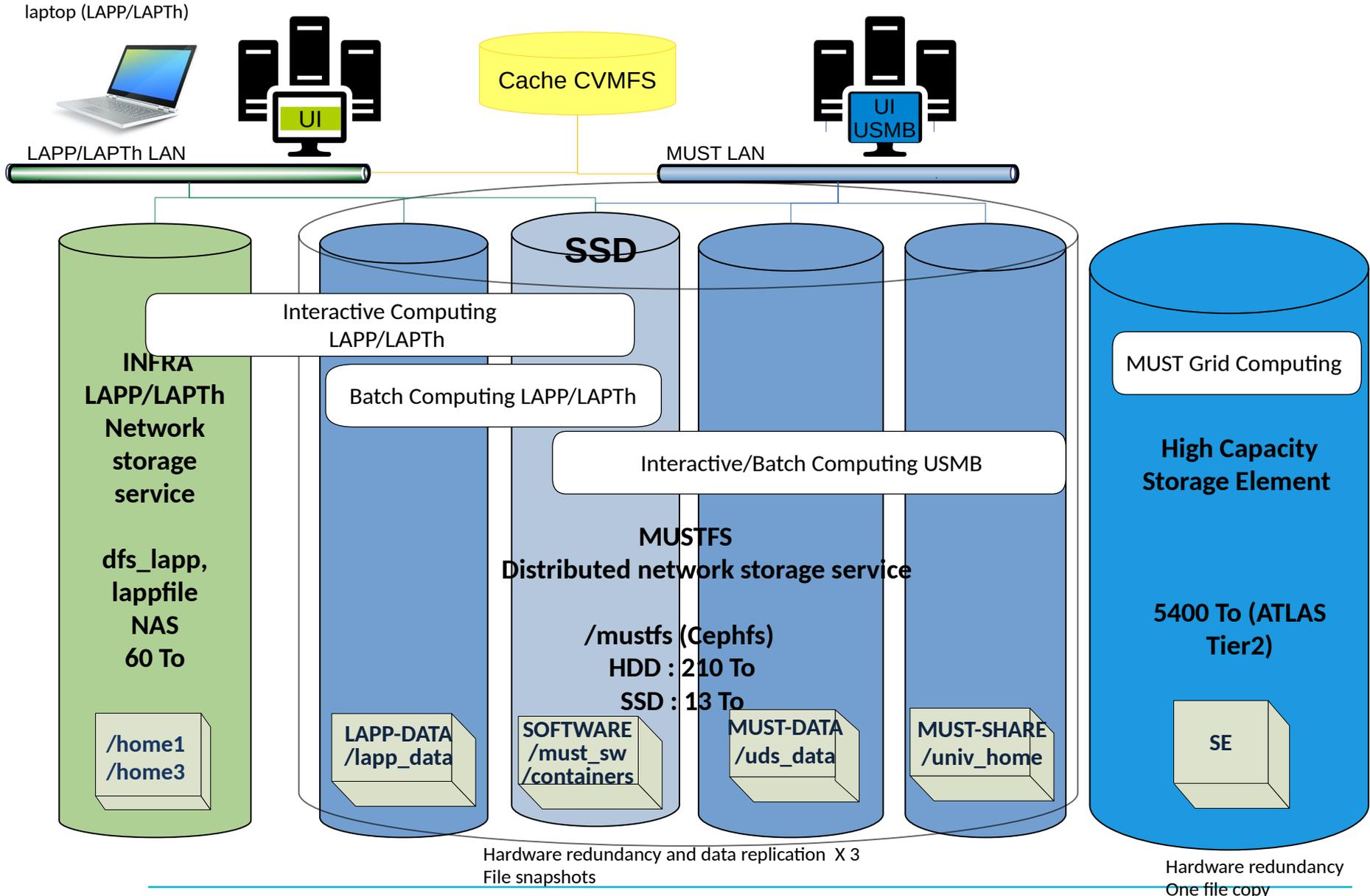


# Amélioration du service de stockage distribué MUSTFS

Réunion utilisateurs, 25 octobre 2022

- Le stockage dans MUST
- Qu'est-que MustFS
- MustFS avant et après
- Comparaison des performances
- Snapshots
- Bonnes pratiques
- Conclusion



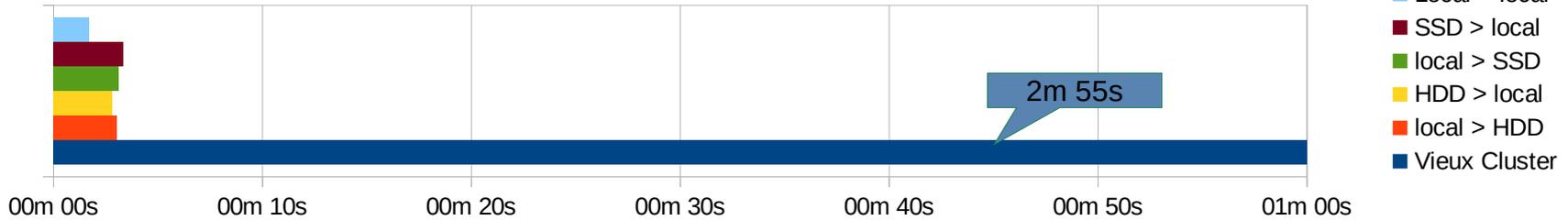
Storage areas	Purpose / Usages	Visibility	Access	Allocation	Data Protection
HOME directory <i>USMB users only</i>	Personal data	<i>MUST Shared area between USMB interactive servers and computes nodes</i>	MUSTFS File system	<i>10 GB per user not accessible by other users, but access rights may be modified</i>	Hardware redundancy and file replication File snapshots
TMPDIR	Transient data storage shared among all users	Local to any interactive or compute node	MUSTFS File system	~ 30 GB per job	No back-up
GROUP data area  /lapp_data /uds_data	Large data volume accessed regularly by colleagues of the same UNIX group	MUST Shared area between USMB interactive servers and computes nodes	MUSTFS File system	1 TB per group Additional storage capacity available under conditions and upon request at support-must	Hardware redundancy and file replication File snapshots
CVMFS	Remote storage used for software distribution	Read-only cache mounted in the universal namespace /cvmfs	Cache, seen as read-only file system	N/A	N/A
DPM Storage Element	High capacity storage dedicated to large-scale or grid computing	MUST Sites involved in LHC and EGI grid computing	Specific protocols	N/A	Duplication in other sites

- Utilise la fonctionnalité de système de fichiers de Ceph
- Ceph : stockage distribué, c'est-à-dire, réparti sur plusieurs serveurs : 9 pour MustFS (bientôt 12)
- Stockage partagé entre toutes les UIs et les WNs
- Partagé en zones permettant :
  - Montage différencié sur les clients
  - Classes de stockage (Matériel différent) suivant les besoins en performance
  - Politique de snapshot adaptée
- Respecte la norme POSIX (gère correctement les droits Unix)
- Gère les quotas (espace occupé ou nombre de fichiers)

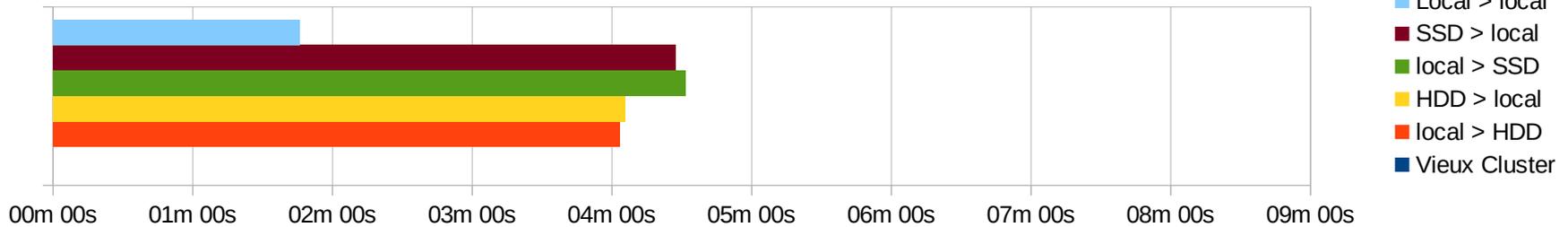
- Suite à un gros crash en 2020, il a été décidé de revoir complètement l'architecture de MustFS

	Avant	Après
<b>Mémoire</b>	96 Go par nœud	128 Go par nœud
<b>Réseau</b>	2 * 10 Gb/s en simple attachement	4 * 25 Gb/s en double attachement sur 2 répartiteurs réseau
<b>Utilisation de SSD</b>	- Métadonnées du FS	- Métadonnées du FS - Données de certaines zones - Métadonnées et journaux Ceph des disques HDD
<b>Redondance des données</b>	- Erasure Coding (~ RAID sur plusieurs serveurs)	- Réplication (3 versions)
<b>Sécurité des données</b>		- Snapshots - Backup non garanti
<b>Serveur de métadonnées</b>	1	4

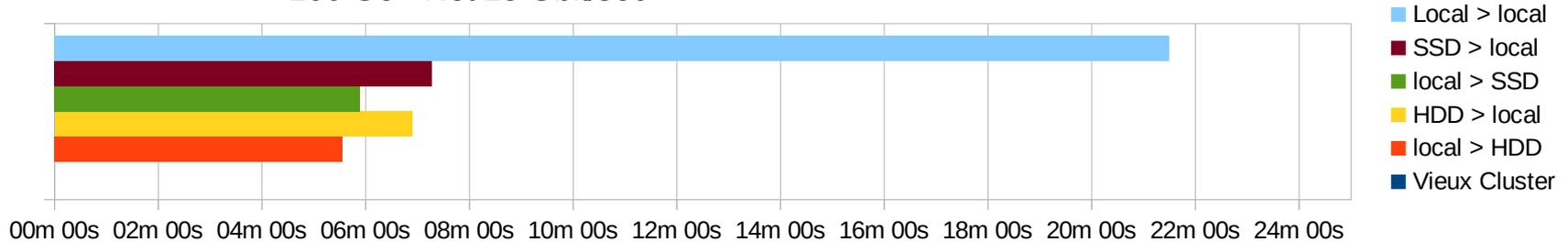
22 Mo 3055 fichiers 338 dir - Net 25 Gbit/sec



12 Go / 77955 fichiers / 8091 dir - Net 25 Gbit/sec



100 Go - Net 25 Gbit/sec



Activation conda (sur disque HDD) : avant ~6 min, après ~300 ms

- Copies instantanées de fichier (versions récentes) permettant un retour arrière très rapide en cas de problème
- Général :
  - 1 snapshot/jour, rétention 7 jours
- /univ\_home :
  - 1 snapshot/heure, rétention 24 heures
  - 1 snapshot/jour, rétention 7 jours

Pour y accéder :

- Répertoire « .snap », invisible même avec un « ls -a »

- Éviter les ls -l
- Éviter les petits fichiers, sinon obligation, stocker dans un fichier « tar » puis le « détarer » en disque local (dans /var/tmp »)
- Ne surtout pas faire de « hardlink »

## Environnement Conda

- × Milliers de petits fichiers
- × Utilise par défaut des « hardlink »
- ✓ Utiliser des environnements par défaut
- ✓ Utiliser les options « allow\_softlinks » et « always\_softlink »

La nouvelle infrastructure est plus résistante et beaucoup plus performante.

Néanmoins, pour conserver de bonnes performances, il est nécessaire de bien respecter les bonnes pratiques