



Réunion de travail à destination des administrateurs gérant une infrastructure CEPH

Jl 2024

23-26 sept. 2024

Les Balcons du Mont-Blanc

Événements

- HEPiX Workshop 2024

- Aperçus et perspectives du stockage

<https://indico.cern.ch/event/1377701/contributions/5896909/>

- L'avenir des disques durs

<https://indico.cern.ch/event/1377701/contributions/5896847/>

- CEPH version

Name	Initial Latest	Release	End of life (estimated)
Quincy	2022-04-19	17.2.7	2024-06-01
Reef	2023-08-07	18.2.4	2025-08-01
Squid	2024-10-??	19.2.0	2026-10-??

- Cephalocon

- 4-5 December 2024 Cern

<https://events.linuxfoundation.org/cephalocon/>

Ceph Squid: À propos de la nouvelle version

- Améliorations apportées à BlueStore pour réduire la latence et les besoins en CPU pour les charges de travail gourmandes en snapshots.
- BlueStore utilise par défaut la compression RocksDB pour augmenter les performances et une utilisation réduite de l'espace disque.
- Améliorations apportées à CRUSH pour les pools en Erasure coding et moins de déplacement de données lors des recovery
- Les performances de lecture et la stabilité sont nettement améliorées avec Crimson. Il prend désormais le scrub, et le « partial recovery ». Améliorations des IOPS avec des CPU multi-cœurs
- RHEL9 Conteneur centos-stream 9

<https://www.linuxfoundation.org/press/introducing-ceph-squid-the-future-of-storage-today>

Évolutions CEPH

- **Crimson - Seastore**
 - 4K IOPS : NVMe 1.8M , SDD 20k, HDD 200 , CPU : nombre de transistors augmente, fréquence et puissance de dissipation qui stagne, nombre de cœur qui augmente
 - CEPH initialement créé pour utiliser des HDD
 - Crimson : vise à minimiser les overheads par IO : single thread par core, pas de changement de contexte
 - Seastore : FS OSD optimisé pour les disques modernes (SSD, NVMe), Dynamic sharding, tiering, bluestore pour les HDD
 - CEPH Squid : Tech preview , Ceph T-release : EC support, PG splits and merges, Performance
 - <https://www.youtube.com/watch?v=QjFliCekAlo> <https://ceph.io/en/news/crimson/>
- **Next Generation Erasure Coding**
 - Optimiser les performances des IO pour les pools en EC
 - Utilisation Parity-delta-write pour limiter la ré-écritures de tous les chunks
 - IOPS Données 64k à 256k HDD ~200 (constant) , SSD et NVME IO divisé par la taille des datas => utiliser des chunks de tailles variables (objet <1MB : 4k sinon 256k)
 - CEPH T-release : Tech preview <https://www.youtube.com/watch?v=bwcntmYP7ic>

CEPH à l'IN2P3

- Utilisation
 - Cloud : Openstack, Kubernetes
 - RBD : IRODS, Proxmox
 - CephFS : stockage partagé labo (NFS)
 - S3 : applicatifs
- IN2P3 (7 labos + cc, ~ 38 Po, ~ 13 admins)
 - CC 5 * 20 Po (Cloud, CEPH-FS, S3) 2 admins
 - CPPM 1 * 0.7 Po (Cloud, CEPH-FS) 1 admins
 - IJCLAB 4 * 12 Po (Cloud, CEPH-FS, S3) 4 admins
 - IPHC 1 * 1.5 Po (Cloud, RBD, CEPH-FS) 2 admins
 - LAPP 2 * 0.8 Po (RBD, CEPH-FS) 2 admins
 - LPCA co-admin université
 - LPNHE 1 * 2 Po (Cloud, CEPH-FS) 1 admin
 - LPSC 1 * 1 Po (?) 1 admin

GT CEPH thématique

- Ansible
 - host
 - ceph-ansible
 - cephadm-ansible
- Stockage
 - Gestion des caches
 - Configuration BlueStore et block.wal block.db SSD NVME
 - LVM-cache
 - openCAS <https://youtu.be/yfkwI2EhHPk>
 - Gestion de la OSD RAM
- Réseaux
 - LACP ou réseaux publique et privé
 - 10Gbs,25Gbs,100Gbs,...

GT CEPH thématique

- CRUSHMAP
 - astuces multi-sites
- Documentation et rédaction articles
 - Guide d'utilisation de ceph
 - Présenter des articles blog ceph
 - <https://ceph.io/en/news/blog/2024/ceph-a-journey-to-1tibps/>
 - <https://ceph.io/en/news/blog/2024/auto-tiering-ceph-object-storage-part-1/>
 - <https://ceph.io/en/news/blog/2024/auto-tiering-ceph-object-storage-part-2/>
- S3 et le Stockage Glacier
 - Règles de déplacement des données en Lecture ou en Ecriture
 - Stockage sur bande magnétique et accès via S3
 - https://ceph.io/assets/pdfs/user_dev_meeting_2024_01_18_joel_davidow.pdf
- GT-CEPH
 - qu'attendez-vous du GT-CEPH ?
 - Retour d'expérience, autres ?