



Groupe de travail CEPH

Jl 2024

23-26 sept. 2024

Les Balcons du Mont-Blanc

<https://resinfo.org/les-groupes-de-travail-11/groupe-de-travail-ceph/>

Vocation du GT CEPH

- Fournir des conseils, de la documentation, des formations pour présenter le fonctionnement de CEPH
- Développer l'expertise des ASR en charge de CEPH au niveau de RBD, CephFS ou au niveau du stockage objets avec le support du protocole S3 ou Swift
- Intégrer le stockage CEPH dans les infrastructures de Cloud Computing (OpenStack, Kubernetes) , Applications (Proxmox, Owncloud, Notebook...)
- Organiser des évènements et des retours d'expériences sur l'utilisation de CEPH dans l'enseignement et la recherche

Comité de pilotage : Sébastien GEIGER, David DELAVENNAT, Patrick Begou, Marco Leoni

Conseils

- Liste de diffusion CEPH RENATER
 - Inscrits: 265; messages: 2022:47, 2023:205, 2024:112
- Conseils pour le dimensionnement d'une infra CEPH
- Prestataires (formations, accompagnement de projets)
- Participation aux projets : France-grilles, VirtualData, SCIGNE, BIGEST, IPCMS
- Café technique CEPH :
 - Visioconférences avec les membres de la liste CEPH RENATER
 - Une moyenne de 3 sessions par an avec environ 35 participants
 - Présentation technique
 - Une revue des questions échangées sur la liste Ceph
 - Conseils pour les prochains achats et matériel retenu
 - <https://resinfo.org/les-groupes-de-travail-11/groupe-de-travail-ceph/article/cafes-techniques-ceph>

Formations et présentations techniques

- Présentations techniques
 - 2024: Projet européen FAIR-EASE : Implementing and using CEPH as an S3 compatible object store
 - 2023: Après-midi RESPIRE - CEPH
 - 2022: AuDACES-ARAMIS : Journées Proxmox VE - CEPH
 - 2018: JTech Aspects techniques de CEPH, Applications de CEPH
- ANF CEPH
 - 2022: <https://indico.mathrice.fr/event/326/overview>
 - 2017: <https://groupes.renater.fr/wiki/ceph/public/formation2017>
- Formation CEPH 2020 (3 jours à Strasbourg)
 - https://resinfo.org/IMG/pdf/ceph.formation_2020.pdf

Questionnaire sur les configurations et les outils utilisés par les admin Ceph (2023)

- 24 questions et 37 réponses
- Depuis combien de temps administrez-vous un ou plusieurs clusters Ceph: 5ans et plus:16, ≥ 2 et < 5 : 16, ≥ 1 et < 2 :3, < 1 :2
- Combien de Clusters Ceph avez-vous en production : 72 clusters
- Quelle taille brute a votre infrastructure Ceph : < 0.5 Po: 13, $\geq .5$ et < 1 : 8, ≥ 1 et < 5 : 13, ≥ 5 et < 20 : 3, ≥ 20 : 0
- Quel type de services de stockage ou de protocole réseau utilisez-vous sur votre infrastructure Ceph : RBD: 32, CEPHFS : 22, S3: 6
- https://resinfo.org/IMG/pdf/ceph.2023_enquete.pdf

Utilisation de CEPH en 2024

- Usage
 - Cloud : Openstack, Kubernetes
 - RBD : IRODS, Proxmox
 - CephFS : stockage partagé labo (type NFS)
 - S3 : S3, applicatifs
- IN2P3 (7 labos + cc, ~ 38 Po, ~ 13 admins)
 - CC5 * 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
- CEA
 - IRFU 3 * 3 Po (RBD,CEPH-FS) 2 admin

Événements

- CEPH : A Journey to 1 TiB/s
<https://ceph.io/en/news/blog/2024/ceph-a-journey-to-1tibps/>
- 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/>

Evolutions 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>

Réflexions en cours

- Guide d'utilisation pour débuter avec CEPH
- Intégrer les évolutions des matériels (architecture, stockage, réseaux)
- CRUSH Maps : comment déterminer les emplacements de stockage
- Répliquions multi-sites (PCA / PRA)
- S3 et les points d'accès multi-régions
- S3 et le Stockage Glacier
 - Règles de déplacement des données en Lecture ou en Ecriture
 - Stockage sur bande magnétique et accès via S3
- Stockage et traitements des données
 - Accès via JupyterLab, Rstudio, ...
 - Utilisation de l'IA en augmentation
 - Où déposer les données UMR, Cloud, Data Lake, Ocean storage