

Sondage CEPH 2026

Le COPIL du GT Ceph de RESINFO à élaborer une enquête à destination des administrateurs d'infrastructure de stockage Ceph.

Le COPIL a choisi des questions destinées à mieux comprendre l'utilisation de Ceph dans le domaine de l'enseignement supérieur et de la recherche tant sur les services de stockage utilisés, les clients qui consomment ces ressources ou les bonnes pratiques pour la gestion d'une telle infrastructure.

Réponses à l'enquête : 39

Structure (39 réponses) :

Unité de recherche	#####	25 (64,1%)
Université et mésocentre	#####	8 (20,5%)
Grande École, ministère	#####	6 (15,4%)

Ceph est largement déployé à tous les échelons de l'enseignement supérieur et de la recherche. Bien que son usage soit majoritairement concentré au sein des unités de recherche (64,1 %), il conserve un ancrage solide au niveau des universités, des mésocentres, des grandes écoles et des ministères.

Domaine des unités de recherche (25 réponses)

Biologie	#####	6 (24,0%)
Physique	#####	4 (16,0%)
Nucléaire & particules	#####	4 (16,0%)
Terre & Univers	#####	3 (12,0%)
Chimie	####	2 (8,0%)
Sciences humaines & sociales	####	2 (8,0%)
Sciences informatiques	####	2 (8,0%)
Ingénierie	##	1 (4,0%)
Écologie & environnement	##	1 (4,0%)

Les domaines de la biologie, de la physique et des sciences nucléaires/particules concentrent à eux seuls plus de la moitié des usages (56 %). Ceph s'impose également de façon transversale dans d'autres disciplines (Terre & Univers, SHS, Chimie, Informatique), où les volumes de données à traiter sont en constante augmentation ou en cours de structuration.

Combien d'année d'expériences ont les administrateurs Ceph (N = 67)

Depuis moins d'un an	#####	9 (13,4%)
Entre 1 et 2 ans	#####	11 (16,4%)
Entre 2 et 5 ans	#####	12 (17,9%)
Plus de 5 ans	#####	35 (52,2%)

Plus de la moitié des administrateurs (52,2 %) possèdent une solide expertise avec plus de 5 ans d'expérience sur Ceph. Parallèlement, l'arrivée de nouveaux profils (près de 30 % des administrateurs ont moins de 2 ans d'expérience) témoigne d'un renouvellement régulier des compétences et d'un renforcement des équipes.

Nombre de cluster Ceph (N = 120)

en production	#####	100 (83,3%)
en test ou pré-prod	#####	20 (16,7%)

Avec une moyenne d'environ trois clusters par site et près de 17 % d'environnements de test ou de préproduction, les résultats traduisent une grande maturité opérationnelle ainsi qu'une forte dynamique d'adoption de Ceph au sein de la communauté.

Version de Ceph (N = 100)

Tentacle (v20,2,x)	#####	25 (25,0%)
Squid (v19,2,x)	#####	54 (54,0%)
Reef (v18,2,x)	#####	14 (14,0%)
Quincy (v16,2,x)	#	1 (1,0%)
Pacific (v16,2,x)	#	3 (3,0%)
Octopus (v15,2,x) ou plus ancien	#	3 (3,0%)

Près de 80 % des clusters s'appuient sur l'une des deux versions stables les plus récentes (Squid à 54 % et Tentacle à 25 %), ce qui montre un très bon suivi des mises à jour. Les rares clusters restés sur d'anciennes versions (Octopus ou antérieures) correspondent principalement à des infrastructures en fin de vie ou en attente de renouvellement matériel.

Taille brute de chaque'un de vos cluster Ceph (N = 101)

< 0,5 Po	#####	46 (45,5%)
De 0,5 Po à 1 Po	#####	15 (14,9%)
De 1 Po à 5 Po	#####	33 (32,7%)
De 5 Po à 10 Po	##	4 (4,0%)
De 10 Po à 20 Po	##	3 (3,0%)
Plus de 20 Po		0 (0,0%)

Les clusters de moins de 0,5 Po représentent près de la moitié du parc (45,5 %), ce qui correspond généralement aux usages d'infrastructures hyperconvergées pour la virtualisation. Les clusters de volumétrie intermédiaire (1 à 5 Po) constituent près du tiers du parc (32,7 %) et répondent aux besoins de stockage massif dédié. Enfin, bien que plus rares, certains clusters dépassent les 5, voire 10 Po, sans toutefois franchir le seuil des 20 Po.

```
# Quel système d'exploitation pour les Nodes (N = 100)
Ubuntu | ##### 16 ( 16,0%)
Debian | ##### 30 ( 30,0%)
RHEL | ## 3 ( 3,0%)
Clone de RHEL (Alma, Rocky, etc,) | ##### 15 ( 15,0%)
Proxmox (PVE) | ##### 36 ( 36,0%)
```

Proxmox VE confirme son adoption massive (36 %), illustrant le fort déploiement de Ceph au sein d'architectures hyperconvergées. Pour les déploiements sur OS générique, Debian demeure la distribution privilégiée (30 %), suivie par Ubuntu (16 %) et les dérivés de RHEL (15 %).

```
# Comment vous déployer les services (N = 103)
cephadm | ##### 53 ( 51.5%)
Ceph-ansible | # 3 ( 2.9%)
Rook | 0 ( 0.0%)
Proxmox (PVE) | ##### 36 ( 35.0%)
Dashboard Ceph | ### 6 ( 5.8%)
autre | ## 5 ( 4.9%)
```

Hors environnement Proxmox VE (35 %), le déploiement des services Ceph s'appuie principalement sur l'outil officiel cephadm (51,5 %). L'absence d'utilisation de Rook suggère que les infrastructures Kubernetes sur ces sites s'appuient principalement sur du stockage externe ou sur des offres cloud managées.

```
# Solution pour déployer les nœuds (N = 105)
Ansible | ##### 33 ( 31,4%)
Puppet | ##### 11 ( 10,5%)
Proxmox | ##### 39 ( 37,1%)
Manuel | ##### 22 ( 21,0%)
```

Le choix des outils de gestion de configuration reflète la maturité et les pratiques internes des équipes. En dehors de l'écosystème Proxmox VE (37,1 %), Ansible s'impose comme la solution privilégiée (31,4 %). Il est à noter qu'une part non négligeable des déploiements d'OS reste effectuée de manière manuelle (21 %).

```
# Technologies de stockage pour leurs OSD (N = 100)
```

Disques durs magnétiques HDD (SATA, SAS)	#####	62 (62.0%)
Disques flash SSD (SATA, SAS)	#####	22 (22.0%)
Disques très haute performance	#####	16 (16.0%)

Les disques magnétiques (HDD) restent prédominants (62 %) pour répondre aux besoins de stockage capacitif à moindre coût. L'usage des technologies Flash (SSD et NVMe), qui représente 38 % des OSD au total, est quant à lui réservé aux environnements nécessitant de hautes performances d'E/S ou de faibles latences.

Quelle(s) configuration(s) utilisez-vous pour RockDB/WAL sur les OSD (N = 166)

RockDB/WAL colocalisé avec le HDD	#####	26 (15,7%)
RockDB/WAL sur SSD et données sur HDD	#####	75 (45,2%)
RockDB/WAL colocalisé avec le SSD	#####	39 (23,5%)
RockDB/WAL sur NVMe et données sur SSD	##	9 (5,4%)
RockDB/WAL colocalisé avec le NVMe	#####	17 (10,2%)

La majorité des répondants privilégie le déport de RocksDB/WAL sur des médias rapides (SSD ou NVMe) lorsque les données sont stockées sur disques magnétiques HDD. La colocalisation reste en revanche la norme pour les pools intégralement composés de flash (100 % SSD ou 100 % NVMe).

Quel type de réseau utilisez-vous entre les nœuds du cluster Ceph (N = 109)

1Gb Ethernet	#	1 (0,9%)
2 à 10Gb Ethernet	#####	42 (38,5%)
11 à 25Gb Ethernet	#####	51 (46,8%)
26 à 40Gb Ethernet	#	1 (0,9%)
41 à 100Gb Ethernet	#####	14 (12,8%)

Les débits de 10 à 25 GbE constituent aujourd'hui le cœur des infrastructures dans l'enseignement supérieur et la recherche (près de 75 % des réseaux interconnectés), offrant un excellent compromis coût/performance. Le 100 GbE réalise une percée notable sur les plus gros clusters (14 %), tandis que le 1 GbE est en voie d'extinction.

Utilisez-vous un réseau privé pour la réplication des OSD (N = 92)

Oui	#####	60 (65,2%)
Non	#####	32 (34,8%)

Près des deux tiers des sites (65 %) isolent le flux de réplication (cluster network) du flux client (public network). Cette pratique recommandée garantit les performances d'E/S des utilisateurs lors des phases intensives de reconstruction (recovery).

```
# Quelle classe d'IP utilisez-vous pour le cluster Ceph (N = 99)
IPv4 uniquement | ##### 96 ( 97,0%)
IPv6 uniquement | # 2 ( 2,0%)
Dual Stack (IPv4 + IPv6) | # 1 ( 1,0%)
```

L'adoption d'IPv6 demeure très marginale au cœur des réseaux de stockage : 97 % des déploiements fonctionnent encore exclusivement en IPv4. Le Dual Stack et l'IPv6 pur restent à ce jour de l'ordre de l'expérimentation.

```
# Quel type de protection utilisez-vous pour vos pools Ceph (N = 113)
Réplication | ##### 73 ( 64,6%)
Erasure Code | ##### 40 ( 35,4%)
```

La réplication reste le mode de protection majoritaire (près de 65 % des usages) pour sa simplicité et ses performances en écriture. L'Erasure Coding (35 %) conserve une part significative, particulièrement plébiscité pour optimiser le coût au gigaoctet sur les volumétries capacitives.

```
# (Si Erasure Code) Quelle configuration utilisez-vous pour K (données) (N = 50)
1 | 0 ( 0,0%)
2 | #### 3 ( 6,0%)
3 | ##### 6 ( 12,0%)
4 | ##### 17 ( 34,0%)
5 | ##### 9 ( 18,0%)
6 | ##### 5 ( 10,0%)
7 | ##### 6 ( 12,0%)
8 | ##### 4 ( 8,0%)
9 | 0 ( 0,0%)
10k ou plus | 0 ( 0,0%)
```

Le profil K = 4 s'impose comme le découpage préférentiel (35 % des configurations en Erasure Code), suivi par K = 5 et K = 3. Ce paramétrage offre le meilleur ratio entre efficacité d'espace et surcharge de l'interconnexion réseau lors des reconstructions de données.

```
# (Si Erasure Code) Quelle configuration utilisez-vous pour M (parité) (N = 48)
1M | # 1 ( 2,1%)
2M | ##### 36 ( 75,0%)
3M ou plus | ##### 11 ( 22,9%)
```

Le réglage M = 2 (double parité) constitue le standard quasi absolu (75 % des réponses). Il apporte la tolérance aux pannes requise en production (perte simultanée de deux nœuds ou disques) sans imposer un surcoût de calcul excessif.

```
# Quel est le nombre d'unités (U) supplémentaires par domaine de
défaillance (N = 52)
```

0	#####	19 (36,5%)
1	#####	8 (15,4%)
2	#####	10 (19,2%)
3 à 5	#####	4 (7,7%)
5 à 10	###	2 (3,8%)
plus de 10	#####	9 (17,3%)

Plus de la moitié des clusters disposent d'au moins 1 à 2 nœuds de marge ($U \geq 1$) pour encaisser une reconstruction sans dégradation de service. Néanmoins, 19 sites fonctionnent sans réserve capacitaire immédiate ($U = 0$), ce qui exige une réactivité forte en cas d'incident matériel.

Pour quel type de client utilisez-vous votre infrastructure Ceph pour fournir du stockage (N = 203)

RBD (direct)	#####	46 (22,7%)
CephFS (direct)	#####	51 (25,1%)
S3 (direct)	#####	10 (4,9%)
iSCSI	#	1 (0,5%)
NVMe-oF		0 (0,0%)
SMB	#	2 (1,0%)
NFS	#####	11 (5,4%)
Kubernetes/OpenShift	###	6 (3,0%)
OpenStack/CloudStack/OpenNebula	#####	12 (5,9%)
Proxmox	#####	11 (5,4%)
(extérieur aux clusters Ceph)		
Proxmox hyperconvergé	#####	31 (15,3%)
Stockage pour KVM/QEMU/libvirt	#####	22 (10,8%)

Le stockage en mode bloc (RBD, porté par Proxmox VE et QEMU/KVM) et le stockage de fichiers distribué (CephFS) dominent largement le paysage. Le stockage objet (S3) poursuit son ancrage pour les besoins de recherche (10 sites), alors que les accès en passerelle de type NFS ou SMB restent marginaux.

Réalisez-vous une sur-allocation (Thin Provisioning) en fonction du type de stockage (N = 70)

RBD	#####	25 (35,7%)
CephFS	##	4 (5,7%)
S3		0 (0,0%)
Non utilisé	#####	41 (58,6%)

Le thin provisioning est essentiellement exploité sur les volumes en mode bloc (RBD), où il est très largement répandu. Il est très peu sollicité sur CephFS et inexistant sur le stockage objet S3, la majorité des administrateurs préférant réserver la capacité réelle pour éviter le surengagement.

Avez-vous activé la compression sur vos pools en fonction du type de stockage (N = 64)

RBD	###	6 (9,4%)
CephFS	#	3 (4,7%)

S3	#	1 (1,6%)
pas de compression	#####	54 (84,4%)

La compression à la volée est très peu activée (seuls 10 % des clusters l'utilisent, principalement sur le stockage bloc RBD). Les équipes privilégient le maintien des performances d'E/S et la préservation des cycles CPU sur les nœuds OSD.

Utilisez-vous un planificateur pour les snapshots en fonction du type de stockage (N = 61)

RBD	###	4 (6,6%)
CephFS	#####	16 (26,2%)
S3 (versionning)	#	2 (3,3%)
Non utilisé	#####	39 (63,9%)

La planification automatisée des snapshots est principalement déployée sur CephFS (16 sites), répondant au besoin de protection des espaces de fichiers partagés ou de calcul. Pour la majorité des infrastructures, la gestion des snapshots reste ponctuelle ou gérée au niveau applicatif (ex. sauvegardes de machines virtuelles).

Réalisez-vous une réplication asynchrone pour la reprise d'activité (N = 67)

RBD	#	3 (4,5%)
CephFS		0 (0,0%)
S3	#	3 (4,5%)
Pas de réplication asynchrone	#####	61 (91,0%)

La réplication asynchrone inter-site (RBD-mirroring ou S3 multi-site) reste très peu répandue (seuls 6 sites l'ont mise en place). La plupart des établissements ne s'appuient pas sur les mécanismes natifs de Ceph pour leur Plan de Reprise d'Activité (PRA).

Votre infrastructure est-elle multi-sites (N = 83)

Non	#####	60 (72,3%)
2 sites/allées	##	4 (4,8%)
3 sites/allées	#####	19 (22,9%)
Plus de 3 sites/allées		0 (0,0%)

Bien que les infrastructures mono-site restent prédominantes (75 %), les clusters étendus (stretched clusters) répartis sur 3 sites ou salles machines représentent 25 % des installations. Cela témoigne d'un fort besoin de haute disponibilité synchrone au niveau de la métropole ou du campus.

Quels outils de métrologie utilisez-vous pour collecter vos métriques (N = 134)

Dashboard Ceph intégré	#####	53 (39,6%)
Prometheus (externe)	#####	20 (14,9%)
Telegraf (externe)		0 (0,0%)

Proxmox et sont dashboard	#####	37 (27,6%)
InfluxDB	###	6 (4,5%)
Collectd	#	1 (0,7%)
autres	#####	16 (11,9%)
pas d'outils dédiés	#	1 (0,7%)

L'observabilité repose en grande partie sur les interfaces intégrées natives : le Dashboard Ceph et l'environnement Proxmox VE couvrent la majorité des usages. Les équipes gérant des infrastructures plus vastes ou intégrées à un SI global privilégient Prometheus et InfluxDB.

# Quelle type de support (N = 90)		
Communautaire / Auto-soutenu	#####	62 (68,9%)
Red Hat Ceph Storage (avec support éditeur)	#	1 (1,1%)
Proxmox PVE (pour ceph)	##	6 (6,7%)
Prestataire tiers	#####	21 (23,3%)

La grande majorité des infrastructures repose sur un modèle communautaire (68,9 %). Néanmoins, près de 23 % des sites font appel à des prestataires tiers pour se faire accompagner ou sécuriser leurs opérations délicates (la liste des prestataires est disponible sur le site du GT Ceph de RESINFO). Enfin, le recours direct au support éditeur (Red Hat ou Proxmox) demeure marginal.

Bilan et conclusion

L'enquête menée auprès des 39 sites de l'enseignement supérieur et de la recherche dresse le portrait d'une infrastructure Ceph mûre, moderne et parfaitement intégrée aux enjeux de stockage et de virtualisation des établissements.

Les grands enseignements de l'étude

Un socle technologique maîtrisé et à jour : Les équipes techniques font preuve d'une grande rigueur opérationnelle, illustrée par le maintien à jour des versions (près de 80 % du parc sous Squid ou Tentacle) et la généralisation des bonnes pratiques d'architecture (déport de RocksDB/WAL sur SSD/NVMe, isolation du réseau de réplication, interconnexions à 25 GbE ou plus).

Deux cas d'usage prédominants :

L'hyperconvergence de proximité : Portée par l'écosystème Proxmox VE (36 % des déploiements) pour des volumes généralement inférieurs à 0,5 Po en stockage bloc (RBD).

Le stockage capacitif pour la recherche : Orienté vers les gros volumes (1 à 5+ Po) en mode fichier (CephFS) ou objet (S3), s'appuyant sur du disque HDD optimisé par de l'Erasure Coding (principalement K=4, M=2).

Une communauté autonome et structurée : L'entraide communautaire domine largement (69 %), complété par le recours ciblé à des prestataires spécialisés (23 %). L'expertise est bien établie (>50 % des administrateurs ont plus de 5 ans d'expérience) tout en continuant d'accueillir de nouveaux profils.