**Armazenamento Proxmox**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Directory (Diretoria)** | **LVM (Logical Volume Manager)** | **LVM-Thin (Thin Provisioning)** | **ZFS** | **Ceph** |
| **Descrição** | Armazena os dados numa pasta normal dentro de um file system (ext4, xfs). | Cria volumes lógicos dentro de um grupo de volume (VG), aloca espaço fixo. | Similar ao LVM, mas permite alocação dinâmica. | Sistema de ficheiros com gestão de volume integrado | Armazenamento distribuído para clusters, utilizado em ambientes Proxmox HA. |
| **Vantagens** | Fácil de configurar  Suporte para backups | Melhor desempenho que diretorias | Melhora uso de espaço  Snapshots rápidos e eficientes  Ideal para VMs | Snapshots eficientes  RAID integrado  Verificação de integridade automática | Redundância automática  Boa escalabilidade |
| **Desvantagens** | Sem suporte nativo para snapshots | Sem snapshots | Risco de sobrecarga se o espaço físico acabar  Mais lento que LVM normal | Alto consumo de RAM  Menor desempenho em algumas situações | Requer múltiplos nós  Complexo de configurar |
| **Thin Provisioning** | Suporta thin provisioning no formato de imagem QEMU (qcow2). | Não suporta thin provisioning. | Suporta thin provisioning. | Suporta thin provisioning. | Suporta thin provisioning. |
| **Snapshots** | Não suporta snapshots | Não suporta snapshots. | Suporta Snapshots | Suporta Snapshots | Suporta Snapshots |



|  |  |  |
| --- | --- | --- |
| **Característica** | **Ceph RBD (RADOS Block Device)** | **CephFS (Ceph File System)** |
| Tipo de armazenamento | Block Storage | File Storage |
| Uso principal | Armazenar discos de VMs e containers | Armazenar ficheiros, ISOs, backups |
| Desempenho | Alto desempenho para VMs | Mais lento para VMs (acesso via file system) |
| Snapshots | Suportado | Não tem suporte nativo |
| Thin Provisioning | Sim (permite alocação dinâmica de espaço) | Sim, mas menos eficiente para VMs |
| Live Migration | Sim, rápido e sem cópias | Não recomendado para VMs |
| Alta disponibilidade (HA) | Sim, distribuído entre os nós | Sim, mas mais indicado para ficheiros |
| Uso no Proxmox | Armazena discos de VMs de forma otimizada | Armazena ISOs, backups e ficheiros partilhados |

**Ainda não testámos o Ceph**

* pve-root(local criado por omissão)  
  O pve-root é um volume lógico (LV) do LVM dentro do grupo de volume (VG) chamado pve.
* vgs-> lista todos os grupos de volumes com os seus atributos
* lvs-> lista todos os volumes lógicos (LVs) no sistema com os seus atributos

|  |  |  |  |
| --- | --- | --- | --- |
| **#PV** | | |  | | --- | | Número de **Physical Volumes (PV)**, ou seja, quantos discos ou partições estão alocados nesse VG. | |
| **#LV** | |  | | --- | | Número de **Logical Volumes (LVs)** dentro do VG. Cada LV é como uma "partição" dentro do LVM. | |
| **#SN** | |  | | --- | | Número de **snapshots** criados dentro do VG. | |

**Comandos usados:**

* lvextend -l +100%FREE /dev/pve/root -> expandir o disco
* resize2fs /dev/pve/root

**Criar uma partição**

* fdisk /dev/sdx
* n → Nova partição
* primária (p)
* último setor (ex:+100G)
* w(salvar)

**Criar um Volume Group (VG)**

* vgcreate nomeVolume /dev/sdx

**Criar um LVM-Thin dentro do VG**

* lvcreate --name nomePool --size 50G --type thin-pool nomeVolume

**Criar Logical Volumes**

* lvcreate -l 100%FREE --name testDirect vg\_storage
* mkfs.ext4 /dev/nomeVolume/lv\_nome -> diretoria

**Editar repositórios Proxmox**

* nano /etc/apt/sources.list.d/pve-enterprise.list
* nano /etc/apt/sources.list.d/ceph.list

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **IPs** | **Disco Principal (sda)** | **Disco Secundário (sdb)** | **RAM** |
| **PC1** | 172.79.79.2 | 250GB | 250GB | 8GB |
| **PC2** | 172.79.79.3 | 500GB | 500GB | 8GB |
| **PC3** | 172.79.79.4 | 500GB | 250GB | 8GB |

**Especificações dos servidores:**

|  |  |  |  |
| --- | --- | --- | --- |
| Característica | ZFS | Ceph | NFS |
| Live Migration | Não (Apenas com replicação e desligamento da VM) | Sim (Storage distribuído) | Sim (Armazenamento compartilhado) |
| Alta Disponibilidade (HA) | Não (Cada nó tem seu próprio storage) | Sim (Storage acessível por todos os nós) | Sim (Se o servidor NFS tiver redundância) |
| Thin Provisioning | Sim, mas precisa ativar sparse=1 | Sim (Automaticamente thin provisioned) | Sim (Automaticamente thin provisioned) |
| Snapshots Rápidos | Sim (Integrado no ZFS) | Sim, mas com impacto na performance | Sim, mas pode ser mais lento que ZFS |
| Redundância de Dados | Sim (via RAID-Z ou Mirror) | Sim (Os dados são replicados entre os nós) | Não (Depende do storage físico) |
| Escalabilidade | Limitado ao armazenamento local | Alta (Facilmente escalável adicionando nós) | Média (Depende do servidor NFS) |
| Requisitos de RAM | 8GB mínimo, recomendado 16GB+ | 16GB mínimo, recomendado 32GB+ | Baixo (4GB é suficiente) |
| Requisitos de Rede | 1Gbps suficiente | 10Gbps recomendado para bom desempenho | 1Gbps pode funcionar, mas 10Gbps recomendado |
| Complexidade de Configuração | Média | Alta | Fácil |

**8/04**

Uma imagem com captura de ecrã, Teclas

Os conteúdos gerados por IA poderão estar incorretos.

|  |  |
| --- | --- |
| **Campo** | **Explicação** |
| **ID** | Identificador do OSD. Tens 3: 0, 1 e 2. |
| **CLASS** | Tipo de dispositivo. Neste caso, todos são hdd (discos rígidos). |
| **WEIGHT** | Peso do OSD no algoritmo CRUSH. Indica quantos dados ele deve receber. |
| **REWEIGHT** | Reajuste manual do peso. Está a 1.00000 para todos (valor padrão). |
| **SIZE** | Tamanho total físico do disco. Tens dois discos de 233 GiB e um de 466 GiB. |
| **RAW USE** | Espaço total físico utilizado. Todos usam cerca de 21 GiB. |
| **DATA** | Dados reais armazenados (sem metadados/omap). |
| **OMAP** | Objetos armazenados no OMAP (útil para RADOS/cephfs). Pouco espaço aqui. |
| **META** | Metadados usados pelo Ceph. Varia entre 680 MiB e 1.2 GiB. |
| **AVAIL** | Espaço disponível. Os de 233 GiB têm ~212 GiB livres; o de 466 GiB tem ~445 GiB. |
| **%USE** | Percentagem de utilização de cada OSD. |
| **VAR** | Variação de uso comparado com a média (ajuda no balanceamento). |
| **PGS** | Placement Groups atribuídos ao OSD. |
| **STATUS** | Estado do OSD |

O **volume de bloco virtual** é um dispositivo que foi associado ao sistema operativo. Ou seja, a imagem RBD associada é um **volume de disco lógico**, armazenado no Ceph, mas que se comporta como um **dispositivo de bloco local** dentro do sistema.

**# Cria uma imagem RBD com 100 GiB**

rbd create Storage1/test-image1 --size 100G

**# Isto torna a imagem visível como um dispositivo de bloco**  
rbd map Storage1/test-image1

**# Formatar**  
mkfs.ext4 /dev/rbd0

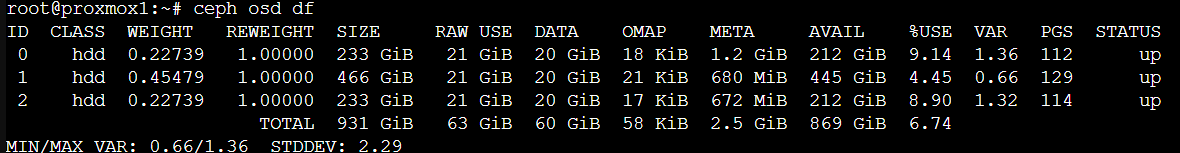
**# Montar o volume**  
mkdir /mnt/test-rbd

mount /dev/rbd0 /mnt/test-rbd  
  
**# Usar /dev/zero para escrever dados reais:**  
dd if=/dev/zero of=/mnt/rbd-test/fill.img bs=1G count=100 status=progresso

**Teste storage**

**Objetivo:** Verificar o comportamento do Ceph quando os discos atingem a capacidade máxima.

**Antes do teste**

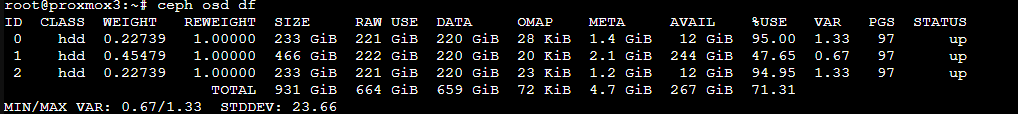


Uma imagem com captura de ecrã, texto, software, ecrã

Os conteúdos gerados por IA poderão estar incorretos.

**Conclusão:** À medida em que os dados sendo armazenados nos OSDs, verificamos que o tamanho total dos mesmos diminui simultaneamente entre os 3 OSDs e com este resultado confirmámos que está a ser armazenado uma réplica em cada.

**Depois do teste**



**Uma imagem com texto, captura de ecrã, Gráfico, file

Os conteúdos gerados por IA poderão estar incorretos.**

**Teste disco**

**Objetivo:** Verificar o comportamento do ceph quando um disco falha. **Uma imagem com texto, captura de ecrã, software, Software de multimédia

Os conteúdos gerados por IA poderão estar incorretos.**

**Uma imagem com texto, captura de ecrã, Software de multimédia, Software gráfico

Os conteúdos gerados por IA poderão estar incorretos.Conclusão:** Ao removermos o cabo de alimentação do disco, o OSD do proxmox3 ficou down/in e passado uns minutos passa para down/out. O OSD está inativo e foi removido logicamente do cluster, o que significa que o Ceph vai começar a redistribuir os dados para outros OSDs e não vai usar mais esse OSD para novos dados.

Como podemos ver na imagem abaixo o armazenamento total passou para os 434.99GB, porque agora tem dois ODS em funcionamento. Portanto só consegue garantir duas réplicas dos dados.

Depois conectar o disco com a máquina em funcionamento, a partir da saída do comando “dmesg | grep -i sd”, podemos observar que o disco sdb (250 GB) foi reconhecido e apresentou um problema de I/O (entrada/saída) durante a escrita e erros relacionados ao Ceph.

Solução: reiniciámos a máquina, o OSD passado uns minutos voltou ao estado up/in. Podemos observar que o Ceph ainda não tinha processado completamente, mostrando valores do armazenamento elevados, mas depois foram normalizados.

Uma imagem com captura de ecrã, texto, Software de multimédia, Software gráfico

Os conteúdos gerados por IA poderão estar incorretos.

Valores dos discos depois de ser reiniciado

Uma imagem com captura de ecrã, texto

Os conteúdos gerados por IA poderão estar incorretos.

Valores normalizados do ODS2 inserido

Uma imagem com texto, captura de ecrã, Tipo de letra

Os conteúdos gerados por IA poderão estar incorretos.

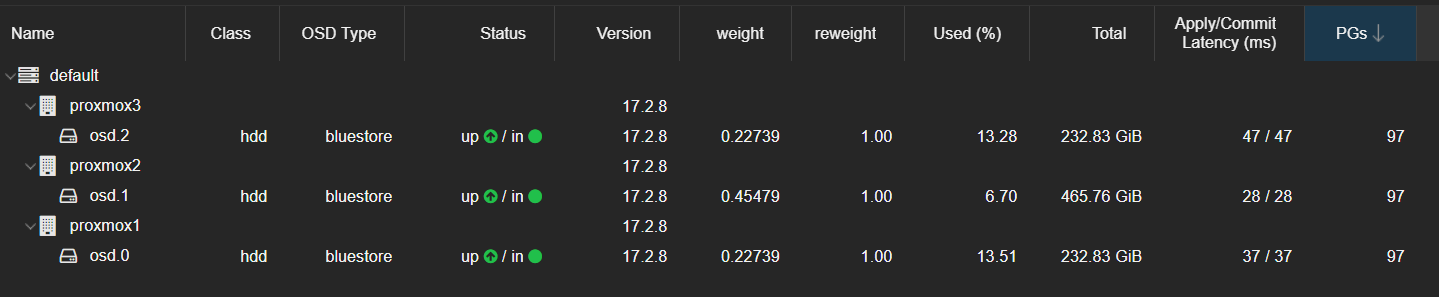
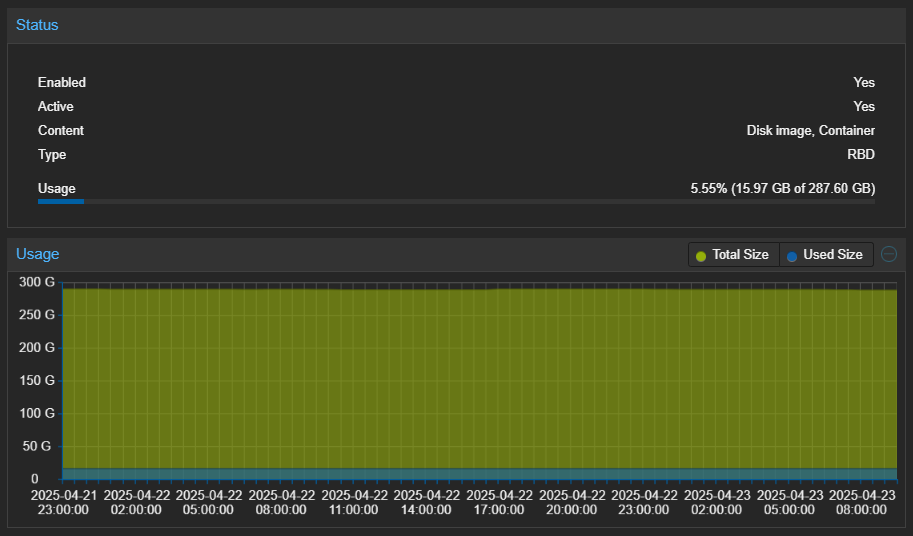
Uma imagem com captura de ecrã, Software de multimédia, Software gráfico

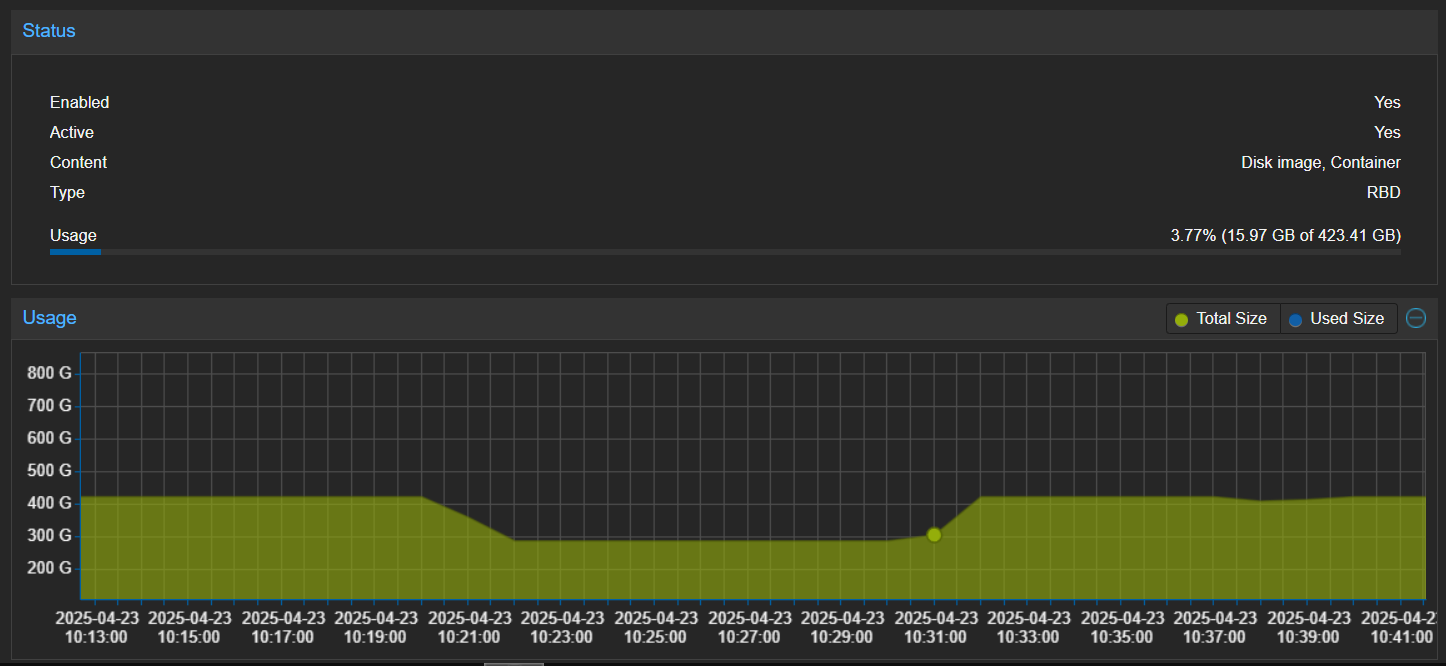
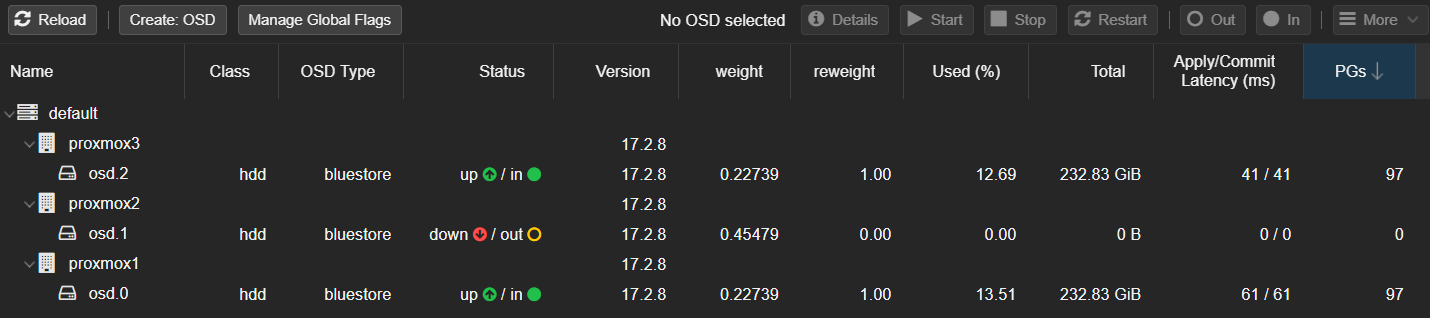
Os conteúdos gerados por IA poderão estar incorretos.

**Remover 2 discos**

**Objetivo:** Verificar o comportamento do ceph quando dois discos falham.

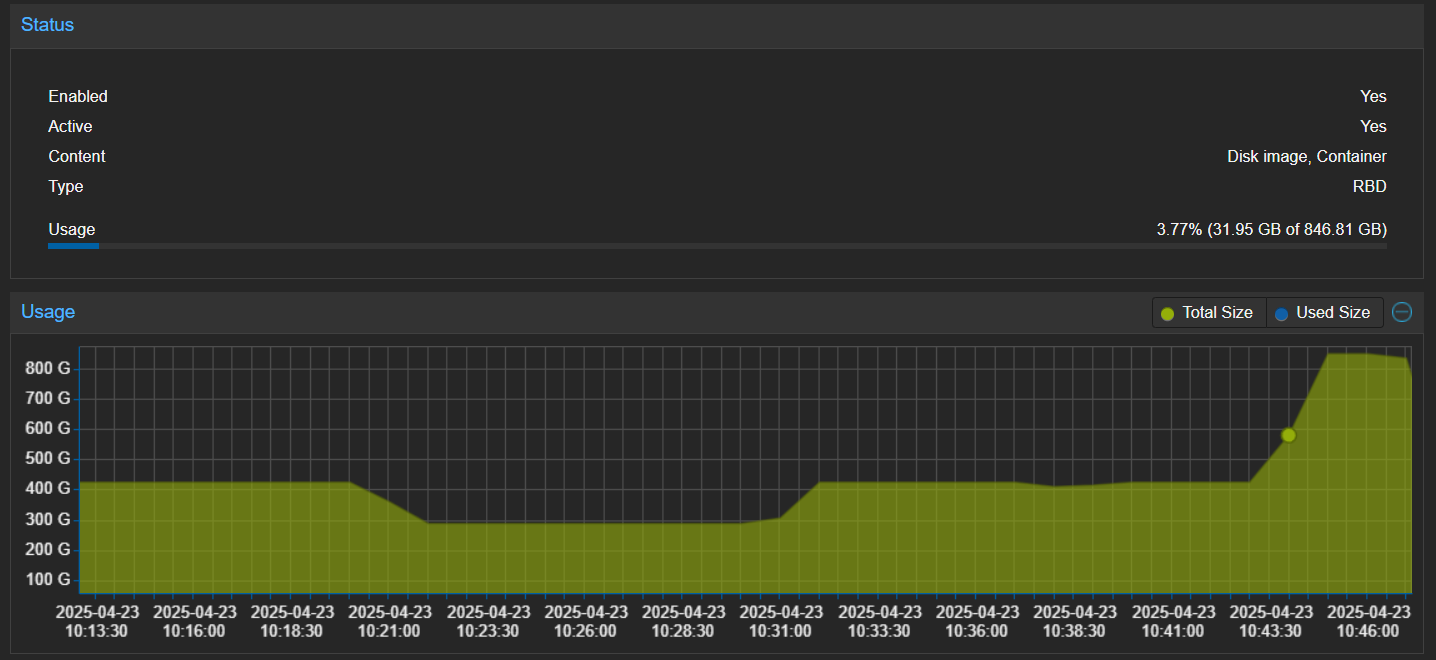
Armazenamento antes da remoção dos discos.



Armazenamento depois da remoção de um disco.  
  


Após a remoção do segundo OSD, o sistema passou a funcionar apenas com um OSD ativo. Esta situação resultou na perda da capacidade de escrita no armazenamento Ceph, uma vez que o parâmetro min\_size dos pools estava configurado com o valor padrão de 2. Este valor define o número mínimo de réplicas necessário para permitir operações de escrita. Com apenas um OSD disponível, o Ceph bloqueou automaticamente as escritas de forma a garantir a integridade dos dados, permitindo apenas leitura.

Armazenamento depois da remoção de dois discos.



Para validar o impacto da degradação do cluster Ceph, foi testada a escrita tanto no volume CephFS como no armazenamento local. A tentativa de escrita no Ceph (/mnt/pve/cephfs/teste1.txt) ficou bloqueada, sem produzir erro imediato, indicando que o sistema estava em espera por recursos de I/O. Isto é consistente com a política de integridade de dados do Ceph, que impede operações de escrita quando não é possível cumprir o número mínimo de réplicas (min\_size=2). Em contraste, a escrita no armazenamento local (/home/teste) foi realizada com sucesso e sem qualquer atraso, para compravar que o problema se restringia exclusivamente ao cluster Ceph.

Escrita no Ceph

Uma imagem com texto, captura de ecrã, Tipo de letra, software

Os conteúdos gerados por IA poderão estar incorretos.

Escrita no Local  
Uma imagem com texto, eletrónica, captura de ecrã, ecrã

Os conteúdos gerados por IA poderão estar incorretos.

Como consequência, os serviços do Proxmox que dependem de volumes montados em CephFS ou RBD ficaram bloqueados em operações de I/O. Isto afetou diretamente os processos internos do Proxmox, resultando na indisponibilidade da interface gráfica (GUI) e na impossibilidade de iniciar sessão, mesmo com as credenciais corretas. Esta limitação foi observada em todos os nós do cluster.

Uma imagem com captura de ecrã, Software de multimédia, software, Software gráfico

Os conteúdos gerados por IA poderão estar incorretos.

Uma imagem com captura de ecrã, texto, software, Software de multimédia

Os conteúdos gerados por IA poderão estar incorretos.

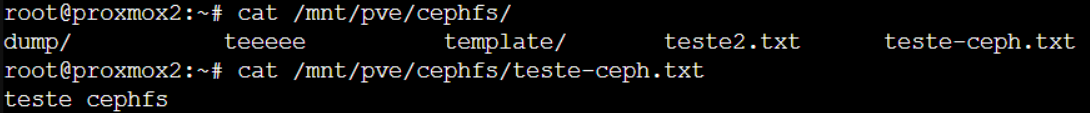
**Segundo teste de Remover 2 Discos**

**Objetivo:** Verificar o comportamento do ceph e GUI quando dois discos falham.

Durante os testes, verificou-se que a interface gráfica (GUI) do Proxmox ficou extremamente lenta, alguns dados não carregavam e que os containers/VMs não conseguiam ser iniciados. Esta situação ocorreu porque os volumes dos containers estavam armazenados na pool *Storage1*, e o cluster encontrava-se a funcionar apenas com um OSD ativo. Mesmo com a configuração min\_size = 1, o Ceph pode bloquear operações de escrita e leitura de forma a proteger a integridade dos dados, o que leva ao bloqueio de serviços críticos do Proxmox que dependem desses volumes.

Uma imagem com captura de ecrã, file, Tipo de letra

Os conteúdos gerados por IA poderão estar incorretos.



Após este teste, voltámos a ligar os discos e o Proxmox voltou a reconhecer o dispositivo, mas o daemon correspondente (ceph-osd@0) não arrancou automaticamente. Isto deveu-se ao facto de o serviço ter entrado num estado de falha, o que impediu novos arranques automáticos por parte do systemd. Apesar de o hardware estar funcional, o OSD manteve-se inativo, sendo necessária intervenção manual para o reiniciar. O serviço ceph-osd@0.service não conseguiu iniciar corretamente porque o *keyring\** necessário para a autenticação não estava presente, impedindo o arranque automático.

Uma imagem com texto, captura de ecrã, Tipo de letra

Os conteúdos gerados por IA poderão estar incorretos.

\* ficheiro que contém as chaves de autenticação (secret keys) usadas pelos diferentes componentes do cluster (OSDs, MONs, MGRs, clientes, etc.) para se autenticarem entre si

**Adicionar um novo disco**

**Objetivo:** verificar se o nó reconhecia um novo disco e adicioná-lo ao ceph sem necessidade de reiniciar o nó.

Para isso, foi inserido um novo disco numa máquina já ligada. O sistema reconheceu corretamente o novo disco e foi possível adicioná-lo como um novo OSD no cluster Ceph, sem necessidade de reiniciar o nó. O tamanho do Storage1, no Ceph, foi atualizado automaticamente após a adição do novo OSD, refletindo o aumento da capacidade total de armazenamento disponível no cluster.

Uma imagem com captura de ecrã, texto, software

Os conteúdos gerados por IA poderão estar incorretos.

**Migração do Sistema Operativo Proxmox VE para um Novo Disco**

**Objetivo:** verificar a possibilidade de transferir o Proxmox VE de um disco com problemas para um novo disco vazio, garantindo a preservação de toda a configuração do sistema sem necessidade de reinstalação, e avaliando se o novo disco permite o arranque correto e funcionamento normal do sistema após a migração.  
  
**dd if=/dev/sda of=/dev/sdc bs=64K conv=noerror,sync status=progresso**

Uma imagem com texto, eletrónica, computador, Dispositivo de saída

Os conteúdos gerados por IA poderão estar incorretos.

Após a execução do comando dd, o novo disco /dev/sdc passou a conter uma cópia integral do disco original, incluindo a tabela de partições, o sistema operativo e todos os dados do Proxmox VE. De seguida, desligou-se a máquina, foi removido fisicamente o disco com falhas (/dev/sda) e mantido apenas o novo disco. Ao reiniciar o sistema, o Proxmox arrancou corretamente a partir do novo disco, sem necessidade de intervenção adicional. Todos os serviços, configurações e máquinas virtuais mantiveram-se funcionais, validando assim o sucesso do processo de migração completa do sistema operativo para um novo disco.

down/in: O OSD está inativo (não responde), mas ainda está dentro do cluster e é considerado para armazenamento de dados.

down/out: O OSD está inativo e foi removido logicamente do cluster, o que significa que o Ceph vai começar a redistribuir os dados para outros OSDs e não vai usar mais esse OSD para novos dados.

**Adicionar um OSD**

Objetivo: Verificar o comportamento e integridade do Ceph quando se adiciona um novo OSD.

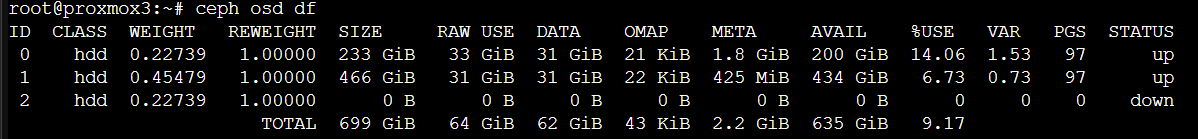
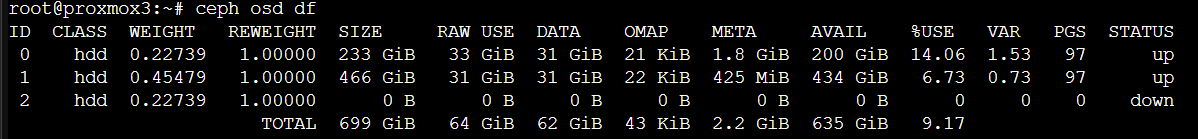
Neste ponto, o cluster Ceph tinha apenas dois OSDs ativos. A capacidade total era de 699 GiB, com 9.17% de utilização. A redundância estava comprometida, já que o número de réplicas era limitado à disponibilidade de apenas dois nós. O novo disco ainda não tinha sido reconhecido nem adicionado ao cluster, razão pela qual o osd.2 ainda não aparece na listagem.

**Hora início - 15:18:26**

**Hora fim – 16:36:52**

Estado do Cluster Antes da Criação do OSD

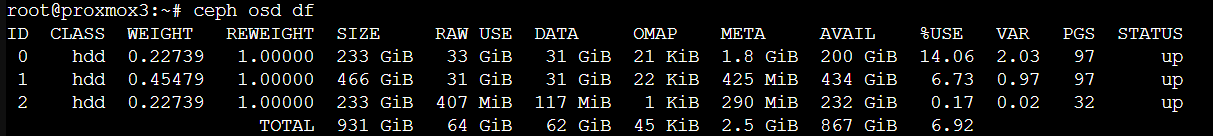
* **Armazenamento total:** 699 GiB
* **Uso:** 64 GiB (9.17%)
* **Número de OSDs ativos:** 2



Após a adição do novo OSD (osd.2), o disco foi reconhecido e adicionado ao cluster com sucesso. No entanto, neste momento, o processo de redistribuição de dados (backfill) ainda se encontrava em fase inicial. É visível que o novo OSD apresenta valores de utilização muito baixos e não está ainda totalmente equilibrado em relação aos restantes. O número de objetos degradados ainda é elevado, indicando que a recuperação está em curso.

Estado do Cluster Após a Criação do OSD

* **Armazenamento total:** 931 GiB
* **Novo OSD (ID 2):** up/in, 233 GiB
* **Espaço disponível:** 867 GiB
* **Rebalanceamento iniciado automaticamente**
* **Objetos degradados:** 29.523% (7314 / 24774)
* **PGs degradadas/undersized:** 58



Uma imagem com texto, captura de ecrã, Tipo de letra, número

Os conteúdos gerados por IA poderão estar incorretos.

Como evidenciado nos gráficos de utilização de CPU, I/O e tráfego de rede, houve uma redução significativa da atividade do sistema por volta das 16:37, momento que coincide com o fim do processo de rebalanceamento do cluster. A utilização da CPU estabiliza com uma queda acentuada, o atraso de I/O (IO delay) reduz significativamente, e o tráfego de rede (netin e netout) reduz-se drasticamente. Estes indicadores confirmam que o Ceph concluiu o processo de redistribuição dos dados e que o cluster regressou ao estado normal de operação.

Uma imagem com captura de ecrã, Software de multimédia, Software gráfico, software

Os conteúdos gerados por IA poderão estar incorretos.

Uma imagem com captura de ecrã, Software de multimédia, Software gráfico, software

Os conteúdos gerados por IA poderão estar incorretos.

Estado Final do Cluster:

* 3 OSDs ativos (up) e integrados (in)
* Todas as PGs em estado active+clean
* Objetos degradados: 0
* Espaço total disponível: 931 GiB

**Uma imagem com captura de ecrã, Software de multimédia, software

Os conteúdos gerados por IA poderão estar incorretos.  
Uma imagem com captura de ecrã, preto

Os conteúdos gerados por IA poderão estar incorretos.**

O processo de rebalanceamento foi concluído com sucesso. Todos os três OSDs encontram-se ativos (up) e integrados no cluster (in). As Placement Groups (PGs) estão em estado active+clean, o que confirma que a replicação dos dados foi totalmente restaurada. A distribuição de armazenamento revela-se equilibrada entre os OSDs, e o estado global do cluster é HEALTH\_OK.

**Conclusão:**

A adição do novo OSD foi bem-sucedida e não foi necessário reiniciar o node. O Ceph iniciou automaticamente o processo de redistribuição, restaurando a redundância e aumentando a capacidade disponível de 699 GiB para 931 GiB. A estabilidade e integridade do cluster foram restabelecidas, cumprindo-se os objetivos do teste.

Gráfico com a comparação de desempenho entre 2 e 3 réplicas. A largura de banda e o número total de escritas descem ligeiramente com 3 réplicas.

O gráfico mostra o tempo até escrever 150 ficheiros de 4 MB cada um no Ceph, comparando duas configurações de replicação (2 réplicas e 3 réplicas). A análise limita-se aos primeiros 150 ficheiros gerados com o comando:

**rados bench -p testpool1 240 write --no-cleanup -t 4**

que utiliza quatro threads de escrita durante 240 s.

Em teoria, se cada thread conseguisse concluir um ficheiro por segundo (4 ficheiros/s no total, ou 16 MB/s), os 150 ficheiros seriam enviados em apenas 37,5 s. Na prática, no entanto, a sobrecarga do Ceph de replicação e confirmação de escrita adiciona latência:

* **2 réplicas (linha vermelha):** cerca de **57 s** para os 150 ficheiros, equivalendo a uma média de **2,7 ficheiros/s** (≈10,8 MB/s).
* **3 réplicas (linha laranja):** aproximadamente **58 s**, ou **2,6 ficheiros/s** (≈10,4 MB/s).

A diferença de **1 s** traduz-se numa penalização de **1–3 %** ao passarmos de 2 para 3 réplicas. Ambas as curvas mantêm um crescimento quase linear, o que indica que o Ceph suporta bem a carga de 4 threads concorrentes, sofrendo apenas variações pontuais de latência (visíveis na grelha de 5 s/5 ficheiros) que se estabilizam à medida que o sistema “enche” o buffer de escrita.

**Teste de Escrita em Cluster Ceph com Replicação 2/2  
  
Objetivo**

O objetivo deste teste é avaliar o desempenho do sistema de armazenamento distribuído Ceph durante uma operação de escrita intensiva. Pretende-se observar o impacto desta operação nas métricas de utilização de CPU, latência dos OSDs, tráfego de rede entre os nós e utilização efetiva do espaço no pool. Este tipo de análise é essencial para validar a estabilidade e escalabilidade do sistema sob carga.

**Configuração do Teste**

O teste foi realizado com o comando rados bench, utilizando o modo de escrita (write), com duração de 5 minutos (300 segundos) e 4 threads de execução:

**rados bench -p testpool2 300 write -t 4 --no-cleanup**

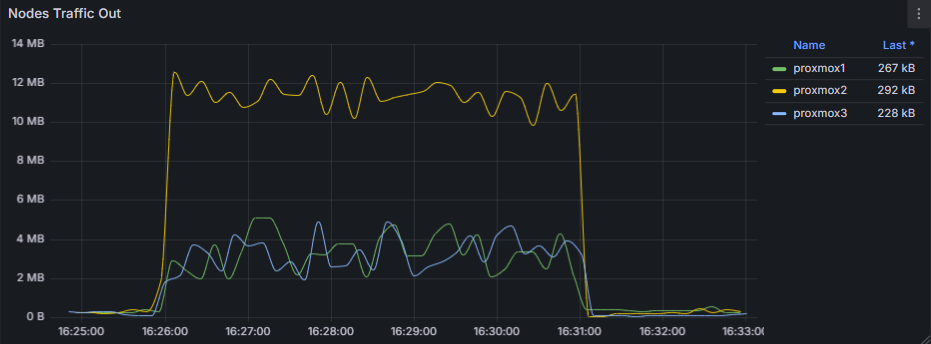
* Pool utilizado: testpool2
* Fator de replicação: 2/2 (cada dado é escrito em 2 réplicas)
* Nó de execução: proxmox2
* Hora do teste: 16:26 – 16:31

**Análise dos Resultados**

Durante o período de execução do teste, foram analisadas diversas métricas de desempenho recolhidas via monitorização em tempo real. Abaixo descreve-se o comportamento observado em cada um dos gráficos:

**Uma imagem com captura de ecrã, texto, Software gráfico, Software de multimédia

Os conteúdos gerados por IA poderão estar incorretos.**Observou-se um aumento de carga nos três nós. O proxmox3 foi o que apresentou maior uso de CPU (até 25%), indicando um envolvimento significativo no processo de escrita e replicação. O proxmox2, que iniciou o teste, registou também subida de carga, enquanto o proxmox1 teve um impacto mais moderado. **Uma imagem com captura de ecrã, diagrama, Software gráfico, Software de multimédia

Os conteúdos gerados por IA poderão estar incorretos.** As latências dos OSDs aumentaram significativamente durante o teste, com valores a ultrapassar os 300 ms em *osd.0*. Isto deveu-se à pressão de escrita e replicação no sistema, sendo um comportamento esperado em operações intensas.  
****

Verificou-se um tráfego elevado de saída no *proxmox2*, superior a 12 MB/s, refletindo a escrita ativa para os outros nós do cluster. O *proxmox1* e o *proxmox3* também participaram, recebendo dados para replicação.

**Uma imagem com captura de ecrã, texto, Tipo de letra, Gráfico

Os conteúdos gerados por IA poderão estar incorretos.**

Os três nós mostraram tráfego de entrada entre 6 MB/s e 8 MB/s, correspondendo ao recebimento dos dados replicados para manter a integridade das cópias no pool testpool2.

**Uma imagem com captura de ecrã, Gráfico

Os conteúdos gerados por IA poderão estar incorretos.**  
O gráfico de Writes mostra a escrita real efetuada nos OSDs do Ceph durante o benchmark. A taxa de escrita oscilou entre 2 MiB/s e 12 MiB/s, como é típico de rados bench, que escreve em blocos.

**Uma imagem com texto, captura de ecrã, software, Software de multimédia

Os conteúdos gerados por IA podem estar incorretos.**

O espaço total utilizado no pool testpool2 foi de 5.98 GiB, valor compatível com a replicação de 2.99 GiB de dados (x2, devido ao fator 2/2). Isto confirma que o teste decorreu corretamente e que a replicação está ativa.

A utilização de 5,98 GiB apresentada na imagem confirma os valores obtidos através dos cálculos teóricos. Considerando que o teste rados bench executou 766 escritas de 4.194.304 bytes cada, o total de dados gerados foi de aproximadamente 2,99 GiB. Com a política de replicação configurada no pool testpool2 (2/2), cada bloco de dados é armazenado duas vezes, resultando num total de 5,98 GiB utilizados, valor registado na interface do Ceph. Esta correspondência valida tanto a execução do teste como o correto funcionamento do mecanismo de replicação.

Total de escritas: 766

Tamanho de cada escrita: 4.194.304 bytes

Total de dados escritos:

766 × 4.194.304 = 3.211.306.624 bytes ≈ 2,99 GiB

Tamanho total utilizado (replicação 2/2):

2,99 GiB × 2 = 5,98 GiB

**3 cópias – teste3 – pool “testpool3” – Write - 4 threads**rados bench -p testpool3 300 write -t 4 --no-cleanup

**Hora – 11:45 – 11:50**

O comando foi feito no proxmox2

**Uma imagem com texto, captura de ecrã, diagrama

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, diagrama, texto, file

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com texto, captura de ecrã, Tipo de letra, Gráfico

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com texto, captura de ecrã, Tipo de letra, file

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, Gráfico, file

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com texto, captura de ecrã, software

Os conteúdos gerados por IA podem estar incorretos.**

**Total writes**: 469

**Write size**: 4.194.304 bytes

**Total bytes** = 469 x 4.194.304 = 1.967.128.576 bytes = 1,832 GiB

**Total Used** = 1,832 x 3 = 5,496 GiB

**2 cópias – teste4 – pool “testpool2” – Read - 4 threads**rados bench -p testpool2 300 read -t 4 --no-cleanup

**Hora – 13:20 – 13:25**

O comando foi feito no proxmox2

**Uma imagem com captura de ecrã, texto, diagrama

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com texto, captura de ecrã, file, Tipo de letra

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com texto, captura de ecrã, file, Tipo de letra

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, file, Gráfico, Tipo de letra

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, texto, diagrama, Gráfico

Os conteúdos gerados por IA podem estar incorretos.**

**Read size** = 4.194.304 bytes = 4 MiB

**Total reads made** = 1.496

**Total** = 4.194.304 × 1.496 = 6.270.197.504 bytes ÷ 1.073.741.824 ≈ 5,84 GiB

**3 cópias – teste5 – pool “testpool3” – Read - 4 threads**rados bench -p testpool3 300 read -t 4 --no-cleanup

**Hora – 13:39 – 13:44**

O comando foi feito no proxmox2

Uma imagem com texto, captura de ecrã, diagrama, Software gráfico

Os conteúdos gerados por IA podem estar incorretos.

Uma imagem com texto, captura de ecrã, Gráfico, file

Os conteúdos gerados por IA podem estar incorretos.

**Uma imagem com texto, captura de ecrã, diagrama, Gráfico

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com texto, captura de ecrã, diagrama, Gráfico

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, file, Gráfico

Os conteúdos gerados por IA podem estar incorretos.**

**Read size** = 4.194.304 bytes = 4 MiB

**Total reads made** = 1.355

**Total** = 4.194.304 × 1.355 = 5.683.281.920 bytes ÷ 1.073.741.824 ≈ 5,29 GiB

**2 cópias – teste6 – pool “pool2RBD” – Write - 4 threads**Criar a imagem: rbd create pool2RDB/testimg2x --size 10240  
Montar a imagem: rbd map pool2RBD/testimg2xfio --name=seqwrite --filename=/dev/rbd0 --rw=write --bs=4M --iodepth=1 --numjobs=4 --direct=1 --runtime=300 --time\_based --group\_reporting

**Hora – 14:09 – 14:14**

O comando foi feito no proxmox2

**Uma imagem com texto, captura de ecrã, diagrama, Software gráfico

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com texto, captura de ecrã, Tipo de letra

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com texto, captura de ecrã, Tipo de letra

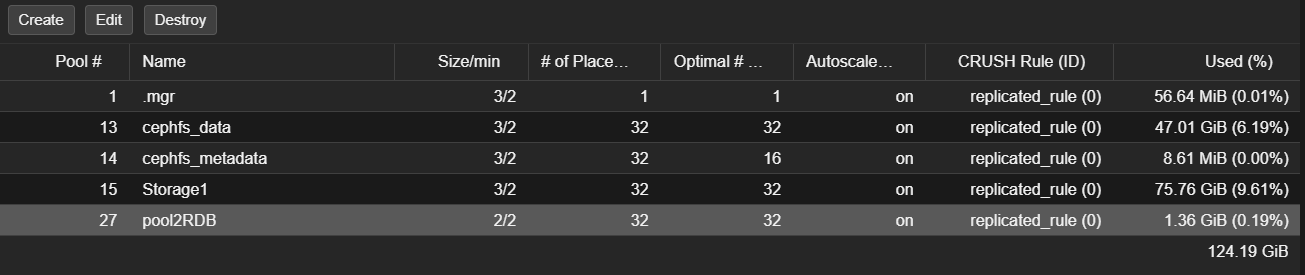
Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com texto, captura de ecrã, diagrama

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, texto, Gráfico

Os conteúdos gerados por IA podem estar incorretos.**

****

**3 cópias – teste7 – pool “pool3RBD” – Write - 4 threads**Criar a imagem: rbd create pool3RBD/testimg3x --size 10240  
Montar a imagem: rbd map pool3RBD/testimg3xfio --name=seqwrite --filename=/dev/rbd1 --rw=write --bs=4M --iodepth=1 --numjobs=4 --direct=1 --runtime=300 --time\_based --group\_reporting

**Hora – 14:39 – 14:44**

O comando foi feito no proxmox2  
Uma imagem com texto, captura de ecrã, diagrama, Software gráfico

Os conteúdos gerados por IA podem estar incorretos.

**Uma imagem com texto, captura de ecrã, Tipo de letra, file

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com texto, captura de ecrã, Tipo de letra, diagrama

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com texto, captura de ecrã, Tipo de letra, diagrama

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com texto, captura de ecrã

Os conteúdos gerados por IA podem estar incorretos.**

**2 cópias- aqui os testes foram feitos nos 3 OSDs**

Proxmox1-Proxmox2-Proxmox3

Quando o teste é feito no Promox1 e Proxmox3 há mais prioridade em ler os dados do Proxmox2 (OSD com mais peso).  
  
Lido - 1.36 Gib

**rbd bench pool2RDB/testimg2x --io-type read --io-size 4M --io-threads 4**

**Uma imagem com texto, captura de ecrã, Software gráfico, Software de multimédia

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, texto, Gráfico, file

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, texto, Gráfico, diagrama

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com texto, captura de ecrã, diagrama, Gráfico

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, file, Gráfico, diagrama

Os conteúdos gerados por IA podem estar incorretos.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Métrica** | **proxmox1** | **proxmox2** | **proxmox3** |
| **Tempo total (elapsed)** | **21 s** | **7 s** | **18 s** |
| **Operações realizadas (ops)** | **256** | **256** | **256** |
| **Operações por segundo** | **12,07 ops/s** | **34,60 ops/s** | **13,70 ops/s** |
| **Throughput (bytes/sec)** | **48 MiB/s** | **138 MiB/s** | **55 MiB/s** |

**3 cópias- aqui os testes foram feitos nos 3 OSDs**

Proxmox1-Proxmox2-Proxmox3

Quando o teste é feito no Promox1 e Proxmox3 há mais prioridade em ler os dados do Proxmox2 (OSD com mais peso).  
  
Lido - 1.08 Gib

**rbd bench pool3RBD/testimg3x --io-type read --io-size 4M --io-threads 4**

**Uma imagem com texto, captura de ecrã, diagrama, Gráfico

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, texto, file, Gráfico

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, diagrama, texto, file

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, texto, diagrama

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com file, captura de ecrã, diagrama, Gráfico

Os conteúdos gerados por IA podem estar incorretos.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Métrica** | **proxmox1** | **proxmox2** | **proxmox3** |
| **Tempo total (elapsed)** | **15 s** | **12 s** | **19 s** |
| **Operações realizadas (ops)** | **256** | **256** | **256** |
| **Operações por segundo** | **16,31 ops/s** | **20,80 ops/s** | **13,47 ops/s** |
| **Throughput (bytes/sec)** | **65 MiB/s** | **83 MiB/s** | **54 MiB/s** |

**2 cópias – CephFS – Write**

**16:35 - 16:40**

**fio --name=cephfs-write \ --directory=/mnt/pve/testcephfs \ --rw=write \ --bs=4M \ --numjobs=4 \ --runtime=300 \ --time\_based \ --ioengine=libaio \ --iodepth=4 \ --group\_reporting \ --direct=1**

**Uma imagem com texto, captura de ecrã, diagrama, mapa

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, texto, Gráfico, file

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, texto, Gráfico, file

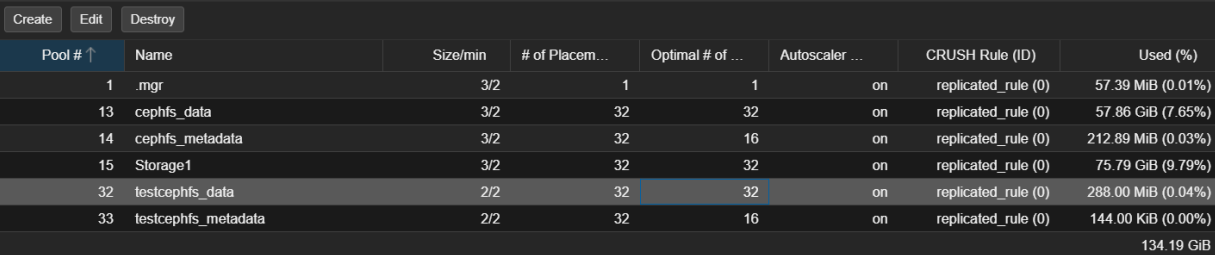
Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com texto, captura de ecrã, diagrama, Tipo de letra

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, texto, diagrama, file

Os conteúdos gerados por IA podem estar incorretos.**

****

**Uma imagem com texto, captura de ecrã, software, Tipo de letra

Os conteúdos gerados por IA podem estar incorretos.**  
Duração real do teste: ~305 segundos

Taxa média de escrita: bw= 9830,4 KiB/s = ~ 9.6 MiB/s

Volume total escrito: io=2928 KiB/s = ~ 2.92GiB

Total escrito: 2.92GiB x 2 = 5.84GiB

**3 cópias – CephFS – Write**

**15:56 - 16:01**

**fio --name=cephfs-write \ --directory=/mnt/pve/cephfs \ --rw=write \ --bs=4M \ --numjobs=4 \ --runtime=300 \ --time\_based \ --ioengine=libaio \ --iodepth=4 \ --group\_reporting \ --direct=1**

**Uma imagem com texto, captura de ecrã, diagrama, Gráfico

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, texto, Gráfico, file

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com texto, captura de ecrã, Gráfico, file

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, texto, diagrama, file

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, texto, Tipo de letra

Os conteúdos gerados por IA podem estar incorretos.**

Duração real do teste: ~305,5 segundos

Taxa média de escrita: bw=6475KiB/s = ~6,3 MiB/s

Volume total escrito: io=1932MiB = ~1,89 GiB

Total escrito: 1,89 GiB x 3 = 5.67GiB

**Uma imagem com texto, captura de ecrã, Tipo de letra

Os conteúdos gerados por IA podem estar incorretos.**

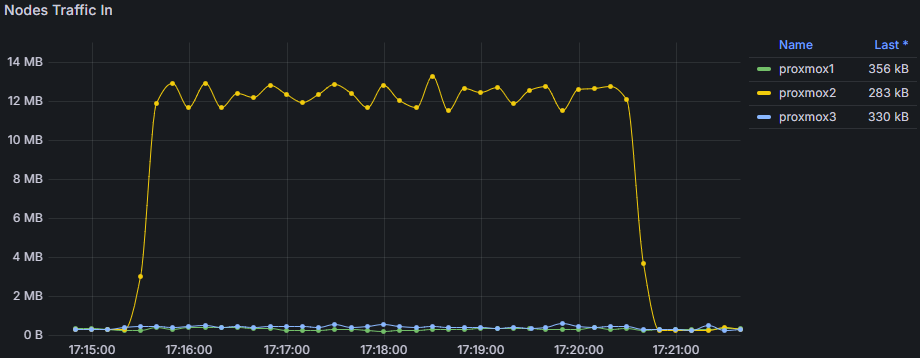
**Uma imagem com texto, captura de ecrã, software, Software de multimédia

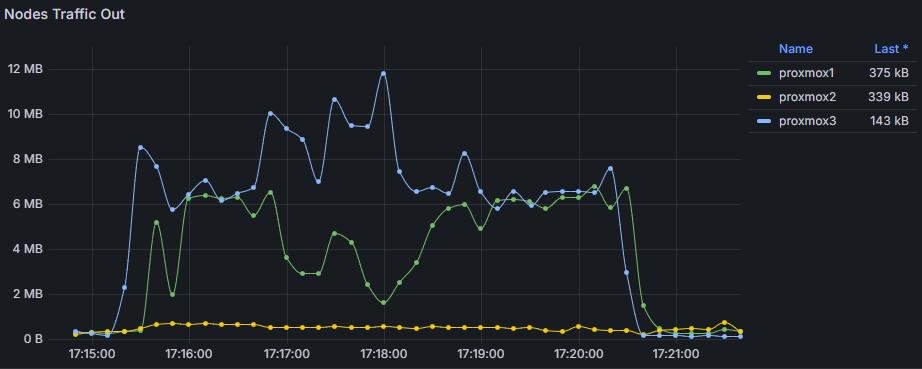
Os conteúdos gerados por IA podem estar incorretos.**

**3 cópias –   
fio --name=read-cephfs \ --rw=read \ --bs=4M \ --numjobs=4 \ --runtime=300 \ --time\_based \ --ioengine=libaio \ --iodepth=4 \ --group\_reporting \ --direct=1 \ --directory=/mnt/pve/cephfs**

**17:15 -   
  
Uma imagem com texto, captura de ecrã, diagrama, Software gráfico

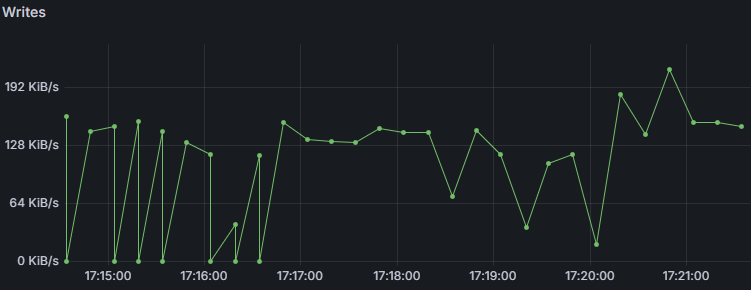
Os conteúdos gerados por IA podem estar incorretos.**

****

****

**Uma imagem com captura de ecrã, texto, diagrama

Os conteúdos gerados por IA podem estar incorretos.**

****

**2 cópias – fio --name=read-testcephfs \ --rw=read \ --bs=4M \ --numjobs=4 \ --runtime=300 \ --time\_based \ --ioengine=libaio \ --iodepth=4 \ --group\_reporting \ --direct=1 \**

**17:26 – 17:32**

**Uma imagem com captura de ecrã, texto, Software gráfico, diagrama

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, texto, Gráfico, file

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, Gráfico, file, diagrama

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com texto, captura de ecrã, diagrama

Os conteúdos gerados por IA podem estar incorretos.**

**Uma imagem com captura de ecrã, file, Gráfico, diagrama

Os conteúdos gerados por IA podem estar incorretos.**

**Teste “Falha de 1 disco (OSD)”**

Objetivo: Verificar o comportamento do ceph quando um disco falha.

Métricas a medir:

* Tempo para deteção da falha (osd down, osd out)
* Tempo total de recuperação (backfill) (ceph -s e logs)
* Número de PGs em estado degraded/remapped (ceph pg stat)
* I/O Recovery Throughput (por OSD) — ceph osd perf, ceph tell osd.\* bench
* Latência para clientes (VMs) durante a recuperação
* Uso de CPU e Rede durante o processo
* Impacto nas VMs ativas: congelamentos, perdas de I/O, aumentos de ping/latência

**Teste “Remover 2 discos (OSDs)”**

Objetivo: Verificar resiliência a múltiplas falhas e comportamento do Ceph.

Métricas a medir:

* Quorum do cluster e perda de réplicas (ceph quorum\_status)
* Número de PGs em estado incomplete, down, undersized, degraded
* Estado dos dados/VMs: continuam disponíveis? só leitura?
* Tempo de recuperação parcial (se possível)
* Logs de falha nos OSDs perdidos
* Latência e falhas de I/O nos clientes
* Percentagem de objetos não replicados corretamente

**Teste “Adicionar novo disco”**

Objetivo: Ver se o nó reconhece um novo disco e o adiciona ao Ceph sem reinício.

Métricas a medir:

* Tempo até o disco ser reconhecido no SO (lsblk, dmesg, journalctl)
* Tempo para o OSD ser criado e ativado (ceph-volume lvm create, ceph osd tree)
* Estado do cluster durante o rebalanceamento (ceph -s)
* Throughput de backfill (MB/s)
* Impacto na latência das VMs
* Custo em rede e CPU durante o rebalanceamento
* Tempo até estado HEALTH\_OK
* Comportamento esperado sem reinício? Verificação via systemctl status ceph\*, logs