Stockage

- WWID NULL & Disques sans chemins

WWID NULL & Disques sans chemins

Composant : DM-multipath (Ubuntu)
Contexte : cluster 6 nœuds GFS2 sur baie HPE Alletra MP B10000 (FC dual-fabric)

Symptôme

Après ajout de nouveaux LUN sur la baie et `createvlun` vers les hôtes, certains nœuds voient bien les disques bruts (`sd*`) via `lsblk` mais **aucune carte** `mpath` **n'est créée** pour ces LUN, et aucune entrée n'apparaît sous `/dev/mapper`. Le comportement est **incohérent d'un nœud à l'autre** : un même LUN peut être correctement multipathé sur un nœud et absent sur un autre.

> **i Point de départ typique** Les LUN historiques (déjà en production et montés) fonctionnent partout. Seuls les LUN *récemment présentés* sont concernés, et pas sur tous les nœuds de façon identique.

Diagnostic

1. Vérifier la config et l'état multipath

Sur chaque nœud, comparer le mode `find_multipaths`, le fichier des WWID mémorisés et les cartes actives :

```
multipathd show config | grep -i find_multipaths
cat /etc/multipath/wwids
multipath -ll
lsblk
```

Si un LUN est bien listé dans `/etc/multipath/wwids`, possède ses 4 chemins et n'a malgré tout aucune carte, le problème n'est **ni** la config `find_multipaths`, **ni** le fichier `wwids` : il faut regarder le WWID réellement calculé.

2. Lancer une reconstruction en observant les messages

```
multipath
```

Sortie révélatrice du problème :

```
sdq: size 1048576000, expected 4294967296. Discard
sdy: size 107374182400, expected 4294967296. Discard
...
reject: 30000000000000000 undef 3PARdata,VV
size=2.0T features='1 queue_if_no_path' hwhandler='1 alua' wp=undef
|-+- policy='service-time 0' prio=50 status=undef
| |- 0:0:0:2 sdc 8:32  undef ready running
| `- 1:0:1:2 sdl 8:176 undef ready running
`-+- policy='service-time 0' prio=10 status=undef
  |- 0:0:1:2 sdf 8:80  undef ready running
  `- 1:0:0:2 sdi 8:128 undef ready running
```

> **Signature du problème** Le WWID `30000000000000000` (tout à zéro) est attribué à **plusieurs LUN différents en même temps**. Les tailles ne correspondant pas, multipath en garde une seule et rejette / discard les autres (`expected 4294967296` = 2 To, la taille du premier LUN traité).

3. Confirmer le WWID lu par le device (non destructif)

Le WWID est dérivé de la page VPD 0x83 (SCSI Device Identification). On lit directement ce que renvoie le device :

```
/lib/udev/scsi_id -g -d /dev/sdc
/lib/udev/scsi_id -g -d /dev/sdq
/lib/udev/scsi_id -g -d /dev/sdy

# VPD brute, si sg3-utils est installe
sg_inq -p 0x83 /dev/sdc
```

Sur un nœud sain, la commande renvoie un WWID valide (`360002ac…`). Sur un nœud en échec, elle renvoie `30000000000000000` : le device retourne un identifiant nul.

Cause racine

Les LUN concernés répondent à l'INQUIRY standard (`vendor = 3PARdata,VV`) mais renvoient une page VPD 0x83 **à zéro**. multipath calcule donc le même WWID nul pour tous ces LUN, les fusionne à tort en une seule carte et rejette l'ensemble à cause du conflit de taille.

L'identifiant nul provient d'un **état SCSI périmé / incomplet côté hôte** : les LUN ont été découverts trop tôt après le `createvlun`, avant que la baie ait publié leur identité. La preuve que ce n'est pas un défaut de la baie : les LUN historiques sur le même hôte, avec la *même* `multipath.conf`, ont un WWID correct — et un autre nœud lit parfois ces mêmes LUN sans problème. Un défaut baie frapperait tous les nœuds de façon identique.

Observation	Interprétation
Disques `sd*` visibles, 4 chemins présents	Login FC, zoning et `createvlun` OK — le SAN est hors de cause
WWID `3000000000000000` partagé par plusieurs LUN	VPD 0x83 nulle : identifiant non lu côté hôte
`size … expected … Discard`	Conséquence de la collision de WWID (tailles incompatibles)
Comportement variable entre nœuds	État de scan transitoire, pas un défaut permanent de la baie

Résolution

> ⚠ **Prérequis de sécurité** À n'appliquer que sur les chemins **non montés et non mappés** (les LUN en échec). Ne jamais supprimer les chemins des LUN en service. Adapter la liste des `sd*` à la sortie `lsblk` du nœud traité, et procéder nœud par nœud.

```
# 1. Retirer les chemins perimes (exemple : 3 LUN x 4 chemins)
for d in sdc sdf sdi sdl sdq sdr sds sdt sdy sdz sdaa sdab; do
  echo 1 > /sys/block/$d/device/delete
done

# 2. Re-scanner les HBA FC pour relire une VPD fraiche
for h in /sys/class/scsi_host/host*/scan; do echo "- - -" > "$h"; done

# 3. Relister (les lettres sd* peuvent changer) et reverifier un WWID
lsblk
/lib/udev/scsi_id -g -d /dev/sdc

# 4. Purger l'entree nulle du fichier wwids
```

```
sed -i '\#^/30000000000000000/$#d' /etc/multipath/wwids

# 5. Reconstruire et verifier
multipath
multipath -ll
lsblk
```

> **Résultat attendu** À l'étape 3, le WWID redevient `360002ac…`. L'étape 5 crée alors une carte par LUN, chacune avec sa taille correcte, et le `lsblk` affiche les `mpath` attendus.

> **Si le WWID reste nul après re-scan** Le problème bascule côté baie : vérifier `showvv` et `showvlun` pour les VV concernés sur l'Alletra, et escalader à HPE. Ce cas reste peu probable si le WWID est lu correctement sur au moins un autre nœud.

Prévention

La cause de fond est l'absence de re-scan SCSI propre après le `createvlun`. Établir la routine suivante, appliquée à **tous** les nœuds du cluster en une passe (`clush`, `pdsh` ou Ansible) :

1. `createvlun` côté baie
2. Re-scan SCSI sur chaque nœud (`echo "- - -" > /sys/class/scsi_host/host*/scan`)
3. `multipath` sur chaque nœud
4. Contrôle : `multipath -ll` identique sur tous les nœuds

Références

- multipath.conf(5) — attribut `uid_attribute`, mode `find_multipaths`
- multipath(8) — modes de fonctionnement de la commande
- scsi_id(8) — lecture du WWID depuis la VPD SCSI