GT Proxmox

Stockage : ZFS, CEPH, LVM, iSCSI, NFS

Sylvain MAURIN

8 Octobre 2024

Rappels

Proxmox

Quel stockage ?

Questions ?

Processus

Une instance

- Contexte d'exécution
- Ressources : RAM, CPU
- Des services d'échange : Accès aux fichiers, sémaphores, IPC...

Fichiers

Fichier

Tout est fichier

- Droits d'accès
- Deux modes d'accès : La pile (accès séquentiel), l'adressage (accès indexé)

Système de fichiers

Système de fichiers

- Un moyen de retrouver l'information dans un volume indexé
- Des propriétés et des limitations
- Une interface unifiée pour les processus pour échanger de l'information

Périphériques

Périphériques

Tout est fichier

- Des points d'entrée dans le système de fichiers
- Deux modes d'accès : La pile (accès séquentiel), l'adressage (accès indexé)
- Des entrées et des sorties

Carte des périphériques

Device Mapper

- Modèle par couche
- Des fonctionnalités (cache, chiffrement, RAID [striped, mirror], instantané, multipath, thin)

Conteneurs

Historique

- Les registres d'états héritiers des anneaux Multics
- chroot
- Namespaces

Rien que des processus

QEMU

Quick Emulator

- Processeur
- Bus
- Contrôleurs
- Périphériques

Rien que des processus

KVM

Kernel-based Virtual Machine

- Processeur avec Intel VT-x ou AMD-V
- Bus/contrôleurs passthrough (VT-c/d)

Toujours que des processus !

KVM/QEMU

Kernel-based Virtual Machine

- Émulation de contrôleurs
- Fichiers d'images

Émulation d'un contrôleur

Un service de passe-plat éduqué

- Bus de communication matériel (SCSI, SAS, ATA, PATA, SATA, USB)
- Émulation de contrôleurs (HBA)

Un contrôleur permet de faire transiter des informations en lecture ou écriture vers un périphérique

QEMU QCOW

Disk image file formats

- Allocation à l'utilisation
- Instantanés
- Compression
- Chiffrement
- Journalisation *Reste une sémantique de fichiers appelés par un processus du point de vue de l'hyperviseur*

QEMU QCOW

Disk image file formats

Un fichier QCOW

- Repose sur un système de fichiers et en hérite les limitations
- Contient les données d'une émulation de périphérique
- Des boîtes à outils de gestion QEMU
- Réside sur l'hyperviseur : binddir, BTRFS, ZFS

Système de fichiers réseau

1 serveur / N clients

- CIFS
- NFS

Pas de différence de sémantique d'accès au système de fichiers pour les applications avec un FS local.

Système de fichiers réseau

N serveurs / N clients

- Gluster
- CephFS

Pas de différence de sémantique d'accès au système de fichiers pour les applications avec un FS local, possibilité de supprimer le SPOF.
L'emplacement réel de la donnée est défini par le client sur la base d'un algorithme déterministe basé sur la géométrie du stockage, problème à l'écriture simultanée d'un même élément avec la gestion d'un verrou unique distribué.

Un service de passe-plat éduqué

Service de passe-plat

Un processus de l'hyperviseur peut lire et écrire sur le système de fichiers de l'hyperviseur :

- Dans un fichier régulier, les données résultant de l'émulation d'un périphérique (Rq. Les deux sont à accès indexés)
- Dans un fichier spécial présent dans le système de fichiers de l'hyperviseur */dev/...* et pointant vers un périphérique en mode bloc apparaissant dans la cartographie des périphériques.

storage.cfg

Fichier de configuration du stockage

- `<PVEtype> <STORAGE_ID>`
  - `<property> <value>`

storage.cfg

Table: Types de stockage disponibles

Description	PVEtype	Level	Shared	Snapshots
Directory	dir	file	no	no
NFS	nfs	file	yes	no
CIFS	cifs	file	yes	no
GlusterFS	glusterfs	file	yes	no
CephFS	cephfs	file	yes	yes
ZFS (local)	zfspool	both	no	yes
LVM	lvm	block	no	no
LVM-thin	lvmthin	block	no	yes
iSCSI/kernel	iscsi	block	yes	no
iSCSI/libiscsi	iscsidirect	block	yes	no
Ceph/RBD	rbd	block	yes	yes
ZFS over iSCSI	zfs	block	yes	yes

storage.cfg

Fichier de configuration du stockage

Propriétés communes :

- `<disable> # storage_check_enabled`
- `<shared> # ! atomic lock by vmid`
- `<nodes> <cluster_nodes_id> # allowed`
- `<content> <content_type>  images, rootdir, vztmpl, backup, iso, snippets`
- `<prune-backups> <keep-(all|hourly|daily|weekly|monthly|yearly)=NN>`
- `<bandwidth> <NN> # ratelimit_bps`

Propriétés communes des backends de stockage de fichiers :

- `<format> <file_format> # raw|vmdk|qcow2`
- `<preallocation> <mode> # off|metadata|falloc|full for raw and qcow2 images`

storage.cfg

Fichier de configuration du stockage

Propriétés du plugin (alias PVEtype) :

- Chemin : `/usr/share/perl5/PVE/Storage/[Custom/]<STORAGE_ID>Plugin`
- Options du plugin
  - `sub properties # backend plugin configuration properties`
  - `sub option ...`

storage.cfg

Fichier de configuration du stockage

Exemple : dir

```
dir: local
        disable
        path /var/lib/vz
        content vztmpl,backup,iso
        shared 0
```

/etc/pve/storage.cfg

Fichier de configuration du stockage

Exemple : lvmthin

```
lvmthin: local-lvm
        disable
        thinpool data
        vgname pve
        content images,rootdir
```

storage.cfg

Fichier de configuration du stockage

Exemple : Ceph

```
rbd: cephinfo
        content images,rootdir
        krbd 1
        pool cephinfo
```

storage.cfg

Fichier de configuration du stockage

Exemple : nfs

```
nfs: proxmoxsave
        export /data/proxmoxsave
        path /mnt/pve/proxmoxsave
        server disque-math.univ-lyon1.fr
        content iso,backup,snippets,vztmpl,images,rootdir
        max-protected-backups 4
        prune-backups keep-daily=2,keep-last=2,keep-monthly=2,keep-weekly=2,keep-
```

VMID.conf

Attachement d'une ressource

- Configuration d'une VM
  `/etc/pve/nodes/<node_name>/<lxc|qemu-server>/<vmid>.conf`
  - `<vm_storage_bus_ID>: <storage_backend>:vm-<vmid>-disk-<NN>,[options storage bus]`
- Exemple :
  - `scsi0: cephinfo: vm-500-disk-0,discard=on,size=3G,ssd=1`
  - `rootfs: cephinfo: vm-104-disk-0,acl=0,size=10G`

LXC

Savoir s'accrocher

Un conteneur est un contexte d'exécution d'une ou plusieurs applications sur un noyau Linux : avant de le lancer, il faut donc préparer son environnement. Exemple : Un périphérique doit être monté dans le système de fichiers du noyau partagé pour être rendu visible aux applications d'un conteneur.

- CONTAINER HOOKS
- Sous conditions, autorisation à des accès directs à des fichiers spéciaux

LXC

pct

Exemple : Un périphérique doit être monté dans le système de fichiers du noyau partagé pour être rendu visible aux applications d'un conteneur.

- CONTAINER HOOKS
- Sous conditions, autorisation à des accès directs à des fichiers spéciaux

QEMU-KVM

qm - QEMU/KVM Virtual Machine Manager

- Lit `<VMID>.conf`
- Attache une ressource de stockage Proxmox

dir

Un accès direct sur le FS d'un hyperviseur

- Pas de migration
- Permet d'utiliser tous les mécanismes Linux de montage dans le FS de l'hyperviseur

NFS/CIFS

Partage réseau

- Permet les migrations
- Réutilisation d'infrastructure et de technologies présentes
- SPOF

GlusterFS

Grille de stockage

- Permet les migrations
- Stockage des images sous forme de fichier, y compris sur les noeuds de stockage
- Performance, complexité

CephFS

Un FS sur un stockage objet

- Permet les migrations
- Stockage des images sur des objets Ceph
- Performance, intégration dans Proxmox VE

LVM/LVM-thin

LVM hyperviseur

- Pas de migration
- Permet d'utiliser tous les mécanismes Linux de cartographie des périphériques

iSCSI

Partage de block

Comme pour NFS/CIFS

- Permet les migrations
- Réutilisation d'infrastructure et de technologies présentes
- SPOF !

Ceph

Stockage objet en grille

- Permet les migrations
- Stockage sur des objets Ceph
- Performance, intégration dans Proxmox VE

ZFS

ZFS (Local)

- Des réplications
- Snapshots, clones

ZFS over iSCSI

- Permet les migrations
- Snapshots, clones

Virtio-9p

À retrouver dans le forum

proxmox-storage-host-to-vm-folder-passthrough-with-9p

- `args:`
  `-virtfs local,`
  `id=faststore9p,path=/rpool/faststore,`
  `security_model=passthrough,`
  $\texttt{mount}_{t}ag = faststore9p$
- `Un FS de l'hyperviseur dans le FS des VM`

DRBD

- Permet les migrations
- Snapshots, clones
- Scalabilité, maintenabilité