

# Mise en place d'un cluster garage

## Garage

Afin de sauvegarder les données de Rhizome, un cluster [Garage](#) est mis en place.

Ce cluster comprend des noeuds appartenant à Rhizome, mais également à ses membres, afin de permettre :

- Une grande diversité quant aux zones
- La possibilité pour les membres de sauvegarder pour eux-mêmes certaines données

Voir la [Spécifications du cluster Garage](#) (authentification requise) pour la spécification complète du cluster, incluant les noeuds des membres. La spécification propre à Rhizome est donnée ci-dessous.

Cette solution permet de s'assurer de la réplication des données dans plusieurs zones en cas de perte d'une machine. C'est pour cela que l'on peut aussi se permettre d'utiliser les machines appartenant à des membres, possiblement moins fiable, mais tant que le travail est fait sérieusement et que le cluster est bien monitoré, cela ne pose pas de problème.

## Restic

[Restic](#) est une très bonne solution de backups incrémentaux, supportant un grand nombre de backends, notamment des buckets S3, fournis par garage.

Ainsi, on peut se servir de restic pour créer un backup sur un des buckets du cluster Garage. Restic se chargera alors du backup incrémental et de son chiffrement, et garage répartira les *blobs* de données entre les serveurs du cluster.

# Sécurité des communications

On distingue deux types de communications pour le backup Garage / Restic :

- Les communications entre les noeuds du cluster, sont chiffrées par garage en utilisant la clef partagée entre les noeuds. Il n'y a donc rien à faire de ce point de vue là.
- Les communications entre les clients et les noeuds via l'API S3 **ne sont pas chiffrées par garage**. En effet, Garage n'expose qu'un *endpoint* http pour les buckets : il est donc nécessaire de mettre en place un *reverse proxy* pour accéder à l'API S3.

Pour les neuds Rhizome, afin de simplifier la mise en place, on sépare donc en deux catégories de noeuds:

- Les **noeuds de stockage** qui sont les noeuds qui possèdent les données, n'exposent pas l'API S3 (uniquement le port pour les communications Garage).
- Les **noeuds gateway** (voir [la documentation garage](#)) sont les points d'accès vers le cluster : ils ne stockent pas de données mais exposent l'API S3 derrière un *reverse proxy*.

L'idée est d'essayer de garder la gateway la plus stable possible et qu'elle ne change jamais, pour avoir tout le temps le même point d'accès même en cas de changement de la topologie du cluster.

## Noeuds Rhizome

Au moment du montage de ce cluster, Rhizome dispose d'espace libre sur Chanterelle et sur Coulemelle. Chanterelle dispose d'un RAID1 de 1To complètement inutilisé (car elle ne fait que du routage) et Coulemelle d'un RAID1 de 2To sur lequel 1.6To sont libres. Par ailleurs, Ultrôn dispose de plus de 4To de stockage, mais on souhaite garder cette machine pour la virtualisation, du moins pour le moment.

D'après la [documentation de garage](#), il n'est pas recommandé d'utiliser ext4 comme système de fichier pour les data, car il est limité sur le nombre d'inodes, ce qui pourrait poser problème si l'on venait à avoir beaucoup de petits objets. Il est plutôt recommandé d'utiliser XFS à ces fins.

Comme on ne souhaite / peut pas reformatter les disques de Coulemelle et Chanterelle (déjà en cours d'utilisation), on va créer une machine virtuelle qemu sur chacune. Ces machines sont des Debian 12, dont la configuration est donnée ci dessous.

## g1.garage.rhizome-fai.net

La machine virtuelle est située sur Chanterelle et dispose d'un disque virtuel sur le SSD de Chanterelle, avec une partition faisant office de SWAP et une autre montée sur `/`. Un autre disque

virtuel de 600Go sur le HDD de Chanterelle est monté sur le répertoire `/var/lib/garage/data`. Il est destiné au stockage pour Garage.

Le service systemd `garage` est créé en s'inspirant de [la documentation](#).

Attention, conformément à [cette issue](#), garage a un bug lorsqu'il est lancé en dynamic user: il n'est pas capable de voir qu'un plus gros disque est monté, et c'est donc la capacité du ssd qui est montrée. En attendant que le bug soit fixé, l'option DynamicUser a été désactivée

La configuration `/etc/garage.toml` est la suivante :

```
metadata_dir = "/var/lib/garage/meta"
data_dir = "/var/lib/garage/data"
db_engine = "lmbd"

replication_mode = "3"
compression_level = 2

rpc_bind_addr = "0.0.0.0:3901"
rpc_public_addr = "80.67.182.87:3901"
rpc_secret = "<...>"

[s3_api]
s3_region = "garage"
api_bind_addr = "127.0.0.1:3900"
```

L'idée de cette configuration est de permettre la communication **via IPv4** (pas d'IPv6 à Rhizome) par les tunnels RPC entre les noeuds du cluster, mais de ne pas permettre l'accès direct à l'API S3 (non chiffrée par défaut).

Malgré une configuration "valide" (pas d'erreur au démarrage du serveur) sans la dernière ligne, il semble nécessaire de spécifier tout de même cette bind address sans quoi Garage ne fonctionnera pas correctement (voir [cette issue](#))

## g2.garage.rhizome-fai.net

La machine virtuelle g2 est située sur Coulemelle. Sa configuration est la même que celle de g1 (à l'adresse IP près).

# gw.garage.rhizome-fai.net

gw est la gateway garage mise en place à Rhizome. Il s'agit d'un noeud ne disposant pas d'espace de stockage de données. Contrairement aux autres noeuds, il ne s'agit pas d'une machine virtuelle, mais d'un conteneur nixos sur Coulemelle.

En revanche, contrairement aux autres noeuds, son API S3 est exposée derrière un reverse proxy via le conteneur `web`. Voir la configuration Nixos de Rhizome pour plus de détails. Ce noeud n'a pas d'IP publique direct, mais un forwarding de ports est mis en place depuis Coulemelle pour le joindre.

---

Révision #7

Créé 9 août 2023 10:34:23 par huetremy

Mis à jour 19 août 2023 09:33:34 par huetremy