

GlusterFS : stockage et sauvegardes en HA

Pierre-Emmanuel GUÉRIN

GeM - UMR6183

CargoDay 5 - 19 novembre 2015

GlusterFS : Présentation

- GlusterFS est un système de fichiers distribué open source et scalable jusqu'à plusieurs Po.
- Les serveurs possèdent des briques de stockage qui, une fois agrégées, forment un volume utilisé par les clients.
- Chaque serveur exécute le daemon glusterfsd pour gérer le stockage.
- Les clients communiquent avec le volume via le mécanisme FUSE (Filesystem in Userspace).
- Deux protocoles de communication peuvent être utilisés (TCP ou RDMA).

GlusterFS : Présentation

Les fonctionnalités sont implémentées par des traducteurs qui sont des librairies partagées.

Exemples de Translators/Traducteurs

- Duplication et réplication par fichier
- Partage de charge par fichier
- Gestion des pannes
- Ordonnancement et cache disque
- Quotas

GlusterFS : Présentation

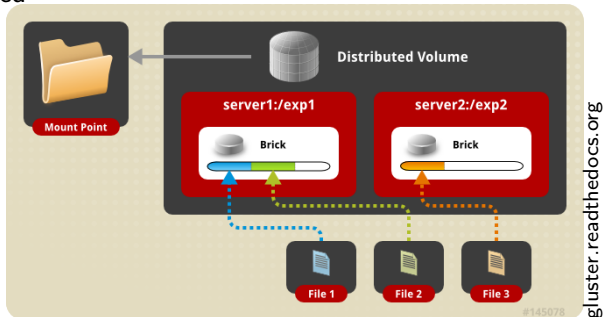
Modes de stockage

- distributed
- replicated
- distributed replicated
- striped
- distributed striped

GlusterFS : Présentation

Modes de stockage

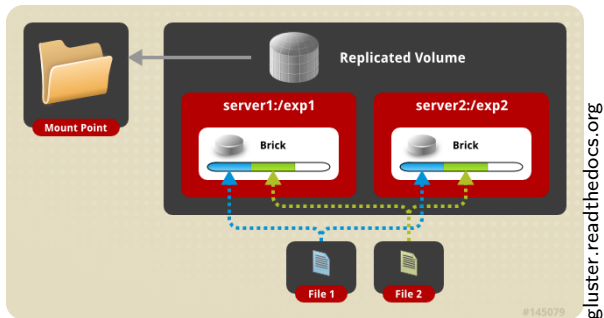
- distributed



GlusterFS : Présentation

Modes de stockage

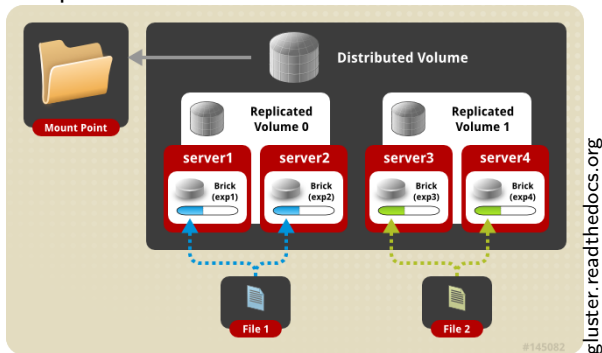
- replicated



GlusterFS : Présentation

Modes de stockage

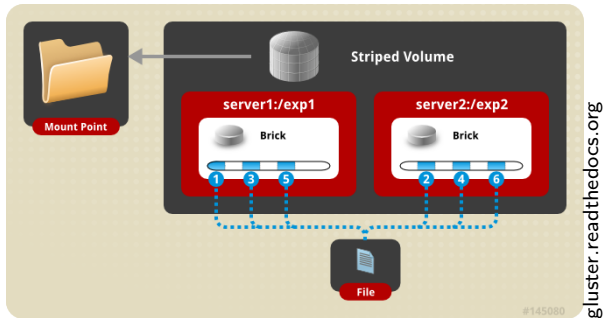
- distributed replicated



GlusterFS : Présentation

Modes de stockage

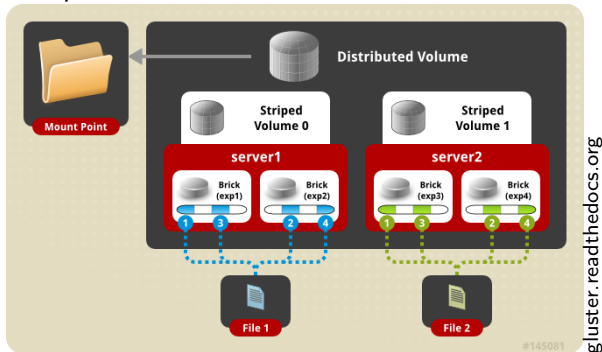
- striped



GlusterFS : Présentation

Modes de stockage

- distributed striped



GlusterFS : Installation

Les paquets de base

- glusterfs
- glusterfs-api
- glusterfs-cli
- glusterfs-client-xlators
- glusterfs-fuse
- glusterfs-libs
- glusterfs-server

GlusterFS : Configuration

Quelques commandes suffisent pour configurer un stockage répliqué

Préparation du stockage sur chaque serveur

```
server*# mkfs.xfs /dev/sdb1  
server*# mkdir /virtual  
server*# vi /etc/fstab  
On ajoute le montage au démarrage
```

Commandes gluster

```
server1# gluster peer probe server2  
  
server1# gluster volume create virtual transport tcp  
replica 2 server1:/virtual server2:/virtual force  
  
server1# gluster volume start virtual
```

GlusterFS : Configuration

Informations sur le volume créé

```
server1# gluster volume info
Volume Name: virtual
Type: Replicate
Volume ID: b75e4eee-9be8-4360-81c9-678315224298
Status: Started
Number of Bricks: 1 x 2 = 2
Transport-type: tcp
Bricks:
Brick1: server1:/virtual
Brick2: server2:/virtual
```

GlusterFS : Configuration

Il est possible d'ajouter des options pour le volume, comme la taille du cache, la gestion du quorum...

Exemple d'options

```
server1# gluster volume set virtual  
performance.cache-size 1GB
```

```
server1# gluster volume set virtual  
cluster.quorum-type auto
```

On retrouve les options appliquées au volume avec la commande précédente : `gluster volume info`

GlusterFS : clients

Les clients peuvent utiliser le volume avec différentes méthodes

- GlusterFS
- NFS
- SMB

Montages manuels

On installe bien sûr le client `glusterfs-fuse`

```
client# mount -t glusterfs server1:/virtual /data  
ou
```

```
client# mount -t nfs -o vers=3 server1:/virtual /data
```

Stocker : Choix

Pourquoi choisir GlusterFS ?

- Plateforme existante entièrement sous CentOS pour les serveurs et majoritairement sous Fedora pour les clients
- Tests de GlusterFS prometteurs
- Utilisation de DRBD avant avec 2 noeuds (peu évolutif)
- Besoin de pouvoir évoluer facilement
- Besoin de virtualiser des serveurs utilisant un stockage fiable
- J'ai choisi la plateforme de virtualisation oVirt (Famille RedHat) et GlusterFS embarqué

Stocker : oVirt

Le stockage GlusterFS est intégré à une plateforme de virtualisation gérée avec oVirt.

- 2 DELL PowerEdge R720
- 2 disques SSD RAID1 pour le système
- 6 disques SAS RAID5 pour le stockage

Aujourd'hui 18 VMs sur environ 7 To.

Pas besoin de montage, les briques communiquent directement.
Les données utilisateurs sont elles aussi virtualisées.

Stocker : Évolution

Prochainement, 2 nouveaux serveurs pour doubler le stockage, et utiliser le mode Répliqué-Distribué de GlusterFS.

Résultats attendus :

- Gain en performance
- Réduction du risque d'interruption de service

Sauvegarder : Choix

Pourquoi utiliser GlusterFS pour les sauvegardes ?

- Besoin de centraliser les sauvegardes
- Besoin de pouvoir faire évoluer facilement
- Bilan positif après 1 an d'utilisation de GlusterFS avec oVirt

Sauvegarder : Existant

Un espace de stockage conséquent à moindre coût

- 2 DELL PowerEdge R520
- 8 disques SAS RAID6 pour le stockage

Environ 20 To utiles répliqués

Utilisation de l'espace via des montages avec le client Natif, ou via NFS.

Conclusion

Avantages

- Scalable
- HA
- Récupération après crash
- Pas besoin d'arbitre pour gérer le split-brain
- Pas de serveur de meta-data

Inconvénients

- Évolution avec des briques de taille identique à celles de base
- Durée de résolution d'un split-brain
- Mode distribué, un serveur perdu, ses données inaccessibles

Merci de votre attention.