

Histoires d'un sysadmin perfectionniste sous pression

2 avril 2022 - Frédéric Zind - JDLL (Lyon)







**Découvrez openZFS : un
stockage fiable, puissant et
accessible.**





Qui suis-je ?

-  2000: technicien en mécanique
-  2018: apprenti charmeur de serpent
-  2019: dev django
-  2020: Soigneur de *pool* ZFS (sys admin)

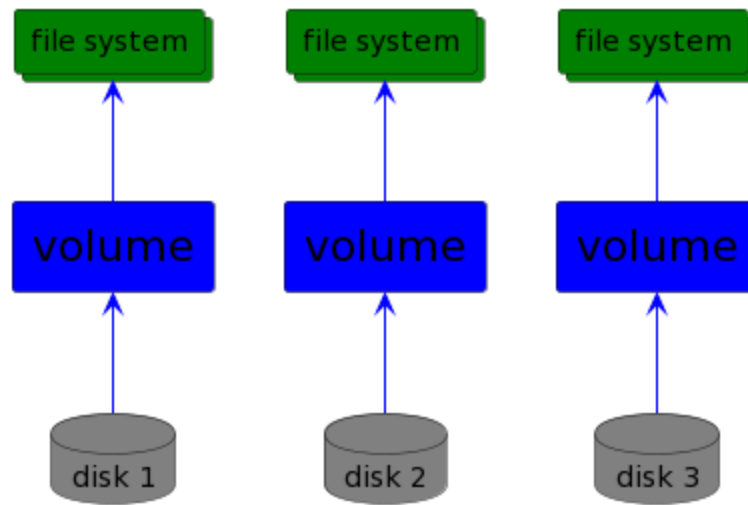


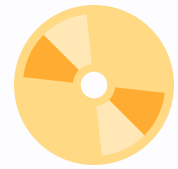
C'est quoi ZFS

- Gestionnaire de volume ET système de fichiers
- Stockage en *pool*
- *Copy-On-Write*
- Usage agressif de *cache* (la RAM)
- Administration simple

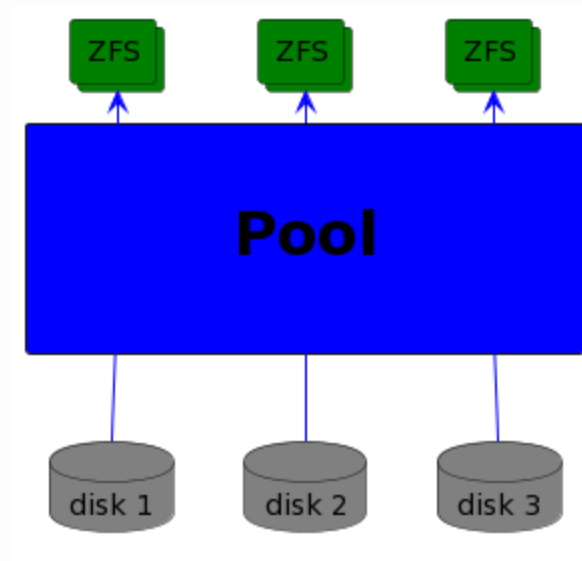


Gestion traditionnelle





Gestion par pool





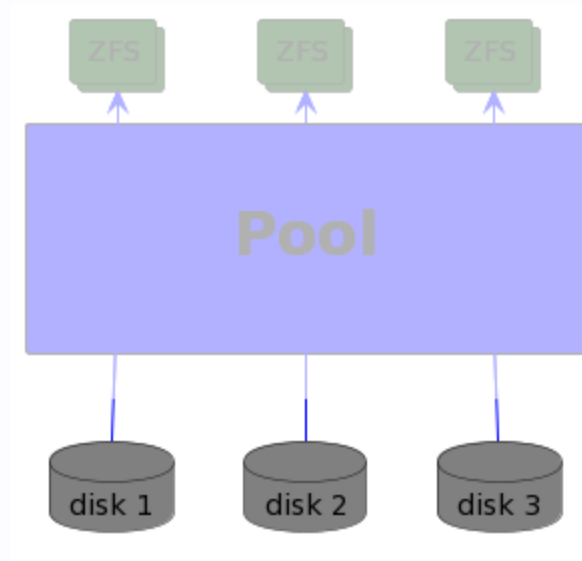
Historique

- 2001 : Naissance chez Sun
- 2005 : Le code source de ZFS est publié
- 2008 : ZFS est publié dans FreeBSD 7.0
- 2010 : Rachat Oracle arrête contributions ZFS
- 2010 : Illumos successeur d'OpenSolaris
- 2013 : Naissance OpenZFS
- 2020 : ZFSv2 Fusion du code FreeBSD/Linux



Quelques concepts

vdev

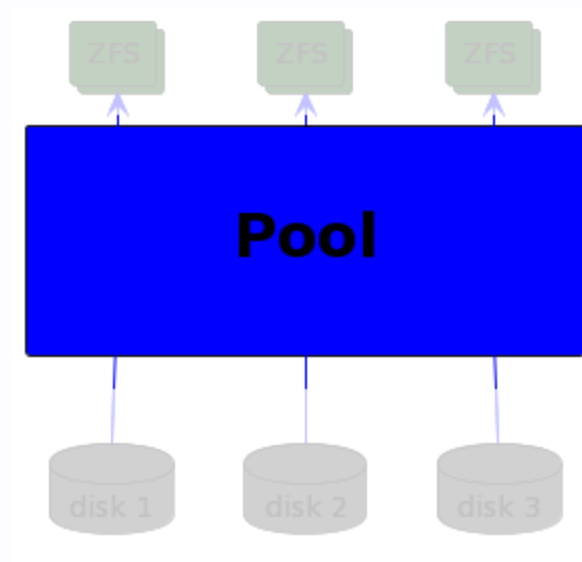


vdev

- Miroir
- *RAID-Z*
- *spare* (chaud ou froid)
- Log (ZIL)
- Cache (L2ARC)



Pool

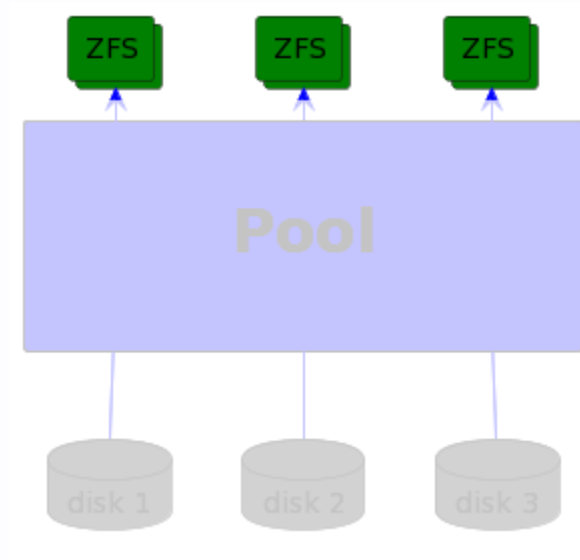


Pool

- Gère les disques
- Peut s'agrandir +++
- Maintenance préventive
- Contient des *datasets*



Dataset



Dataset

- File system, snapshot, clone, ...
- $< 2^{48}$ datasets / pool
- Gigogne/arborescent avec héritage
- Propriétés
 - Reservation / Quota (dataset/reference)
 - Compression, deduplication
 - ACLs, Prop. personnalisée avec `:`, etc.

Cache

- ARC
- L2ARC



Modèle transactionnel

- *Copy-On-Write*
 - Toujours cohérent: pas de FSCK, jamais
- Snapshots
- Send / receive
 - Expédition de snapshots
 - Unidirectionnel
 - Re-démarrable



Administration simple

- Administration a chaud/online
- 2 commandes:
 - `zpool` : *pool*
 - `zfs` : *dataset*
- Délégation

Communauté

- code base unique: FreeBSD / Linux
 - macOS©®
 - Windows©®

Nota bene

- ZFS ♥ RAM
- Choix des vdevs: IOPS **ou** stockage
- snapshots != sauvegardes
- La compression coute moins cher que la déduplication
- Ce n'est pas parce que c'est possible qu'il faut le faire
- Pas de magie !



**Faites gaffe quand
même...**

Gandi - Postmortem: September 30 storage incident

“ 30/09/2020 @ 05:38 UTC, one of our storage units went down.

30/09/2020 @ 11:52 UTC we managed to bring the storage unit back online. ”

➔ Erreur humaine: HDD -> ZIL (SSD)

LTT - Our data is GONE... Again



➔ Erreurs humaines



Références & merci

- Matt Ahrens & George Wilson pour: OpenZFS Basics at SCALE16x, March 2018
- Ubuntu — An overview of ZFS concepts
- FreeBSD Handbook — The Z File System (ZFS)
- Things Nobody Told You About ZFS
- *PU.storage team*

Merci !

!? Questions, remarques,
réclamations, etc.



<http://pro.zind.fr>