# ZFS met RAID-Z als alternatief voor klassieke RAID-oplossingen

Jonas De Moor

Toegepaste Informatica - Systeem- en Netwerkbeheer
Hogeschool Gent

*jonas.demoor.v3741@student.hogent.be*

15 juni 2017

# Inhoud

# Motivatie voor het voeren van dit onderzoek

- RAID5 'write hole'
- Relatie tussen BTRFS en ZFS
- ZFS On Linux (cf. Ubuntu 16.04 LTS)
- Interesses: Linux en Unix

# Onderzoeksvragen

- Wat zijn de grootste verschillen tussen een klassieke RAID-oplossing en ZFS RAID-Z?
- Hoe is de architectuur van ZFS opgebouwd en op welke manieren tracht het oplossingen te vinden voor de problemen die zich voordoen bij andere bestandssystemen en RAID-opstellingen?
- Hoe staat het met data-integriteit en performantie[1] bij ZFS onder verschillende workloads en toepassingen?

[1] Met 'performantie' wordt het aantal I/O's per seconde en de globale CPU-belasting bedoeld.

# Opbouw van het onderzoek

Twee grote onderdelen:

1. Theoretisch gedeelte
   - Inleiding tot RAID-niveaus
   - Architectuur en ontwerpprincipes van ZFS
   - Interne datastructuren en transactiemodel
2. Praktisch gedeelte
   - Storage Pools & VDEV's
   - Datasets
   - Performantie & Betrouwbaarheid

# Gehanteerde methodiek

- Phoronix Benchmark: performantietesten op fysieke machine
  - FIO (Flexible I/O Tester): IOPS
  - FS-Mark: bestandssysteemoperaties
  - PostMark: simulatie van webserver/mailserver
  - SQLite: databankoperaties
- Virtuele Machine: betrouwbaarheidstesten
  - Wegvallen van een schijf (array van drie schijven)
  - Dataverlies door gebruikersfout
  - Bescherming tegen datacorruptie

# Gehanteerde methodiek

| Specificaties | |
|---|---|
| **Fabrikant** | HP |
| **Model** | HP Pavilion Elite HPE-310be |
| **CPU** | Intel Core i5 650 @ 3.2 GHz (2 Cores; 4 Threads) |
| **Geheugen** | 10GB DDR3 @ 1333MHz |
| **GPU** | AMD Radeon HD 5570 |
| **Interne schijven** | SAMSUNG HD103SJ (1TB)<br>WDC WD1002FAEX-0 (1TB)<br>WDC WD5000AZRX-0 (500GB) |
| **Externe schijf** | WD Elements 1078 (1TB) |
| **RAID Controller** | Intel Corporation SATA RAID Controller |

Tabel: Specificaties van het fysieke systeem dat gebruikt werd doorheen de bachelorproef (data verkregen via `lshw`)

# Gehanteerde methodiek

| Specificaties Virtuele Machine | |
|---|---|
| **OS** | Fedora Server 25 |
| **CPU** | 4x Host CPU (Intel Core i7-4712HQ CPU @ 2.30GHz) |
| **Geheugen** | 8GB |
| **OS-schijf** | 20GB (`/dev/sda`; SATA non-hot-pluggable) |
| **Zpool schijven** | 40GB (`/dev/sdb`; SATA hot-pluggable)<br>40GB (`/dev/sdc`; SATA hot-pluggable)<br>40GB (`/dev/sdd`; SATA hot-pluggable) |
| **NIC's** | VirtualBox NAT-adapter (10.0.2.15/24)<br>VirtualBox Host-only Adapter (192.168.56.10/24) |

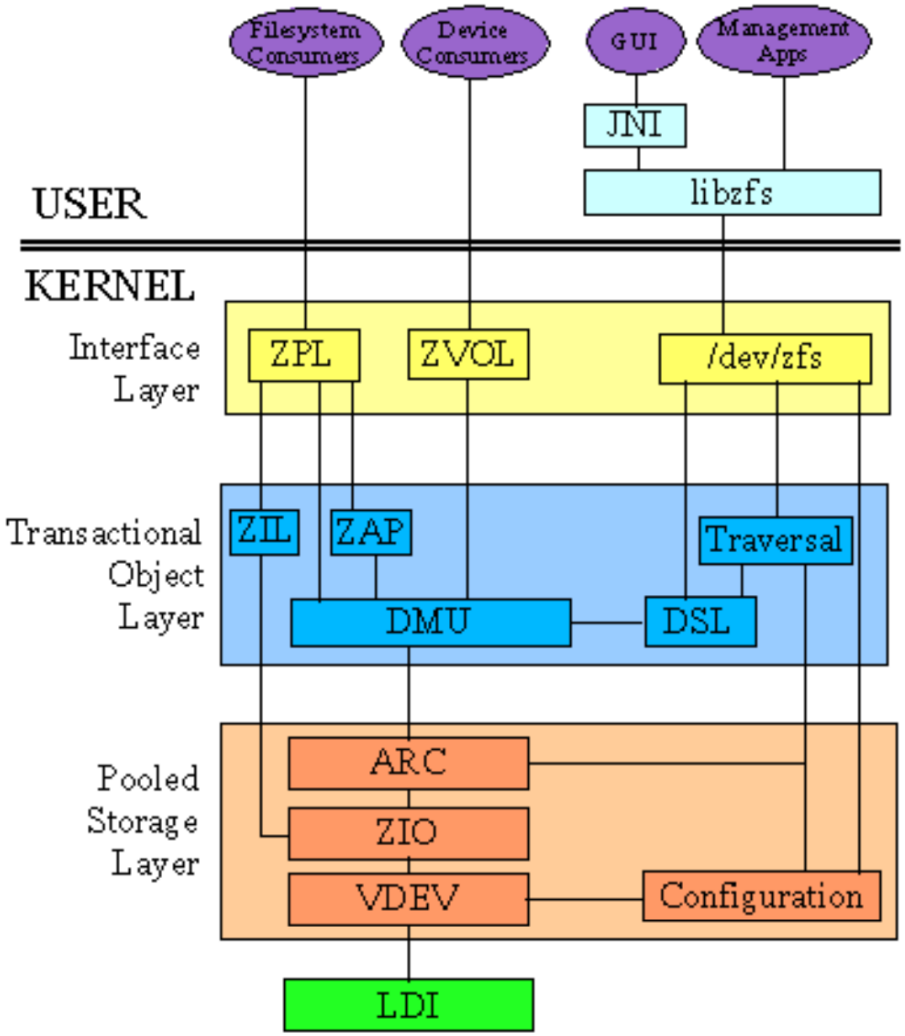
Tabel: Specificaties van de virtuele machine die gebruikt werd voor de betrouwbaarheidstesten

# Inhoud

# ZFS: een kort overzicht

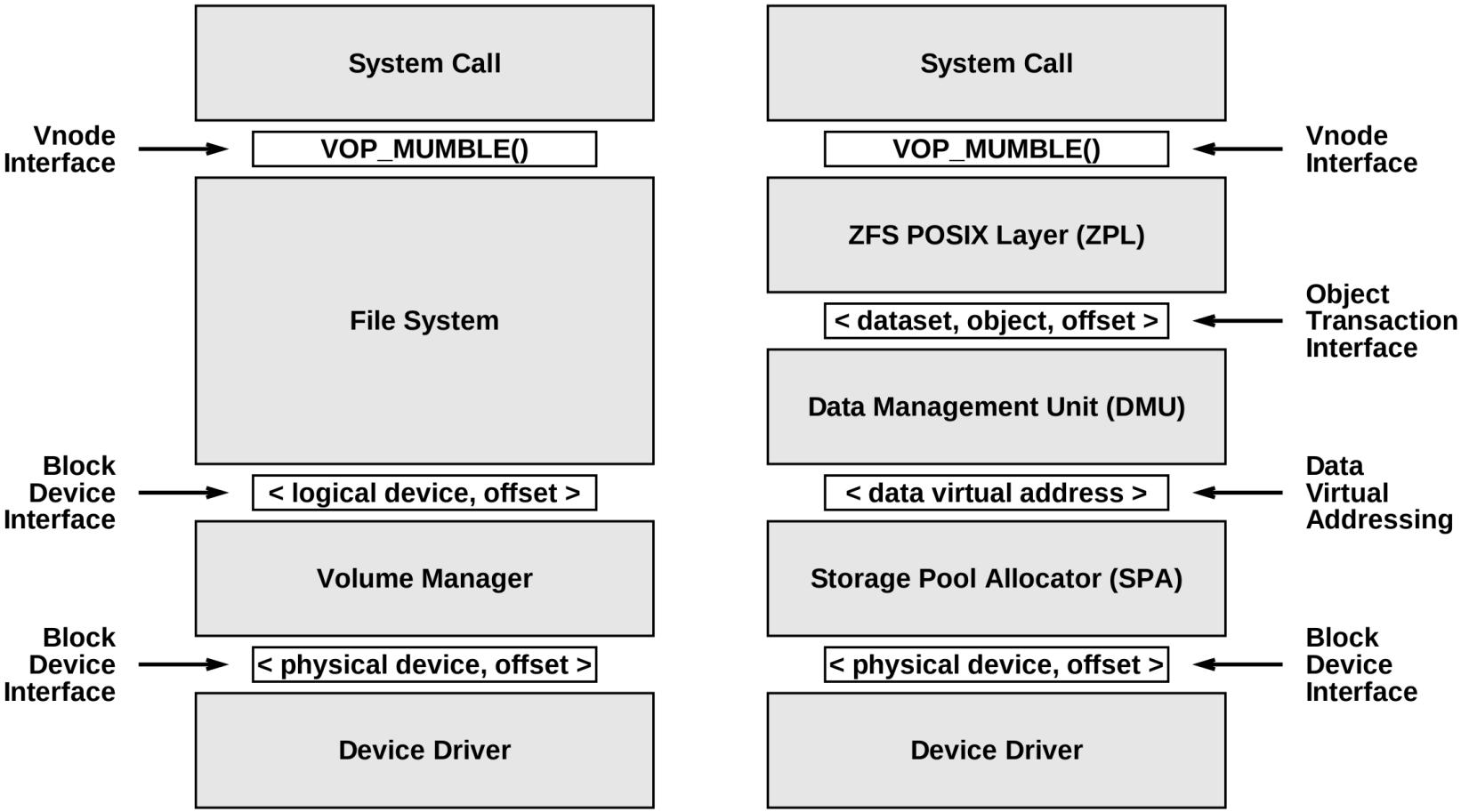
- Copy-On-Write bestandssysteem
- Ontwikkeld door Sun Microsystems (begin jaren 2000)
- Oorspronkelijk onderdeel van Solaris
- Nu: verdere ontwikkeling via OpenZFS (en Oracle)
- Ondertussen ook beschikbaar op BSD en Linux (ZFS on Linux)
- Beschikt over RAID-Z (softwarematige RAID)

# Architectuur van ZFS




Figuur: Een overzicht van de verschillende componenten van ZFS (Kendi, Onbekend)

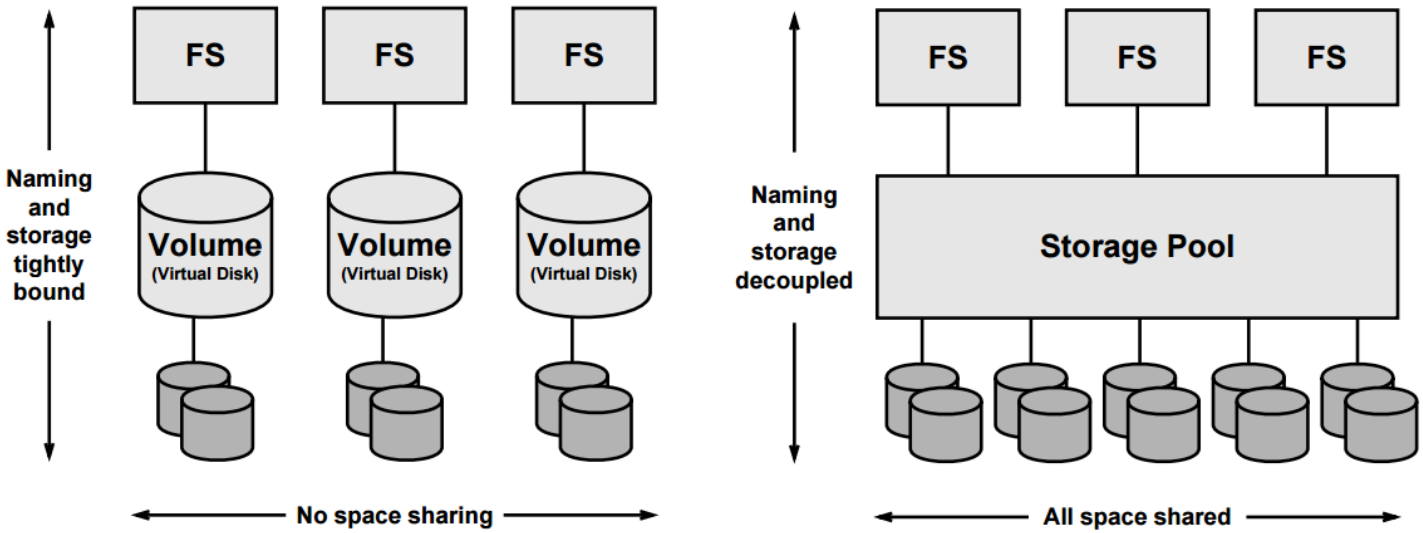
# Architectuur van ZFS




Figuur: Vergelijking tussen een 'traditionele' storage stack (links) en de ZFS storage stack (rechts) (Bonwick e.a., 2002)

# Storage Pools

- Abstractie voor fysieke apparaten → gegroepeerd in VDEV's
- Dynamische allocatie van opslagruimte
- Schijven kunnen worden toegevoegd zonder downtime[2]

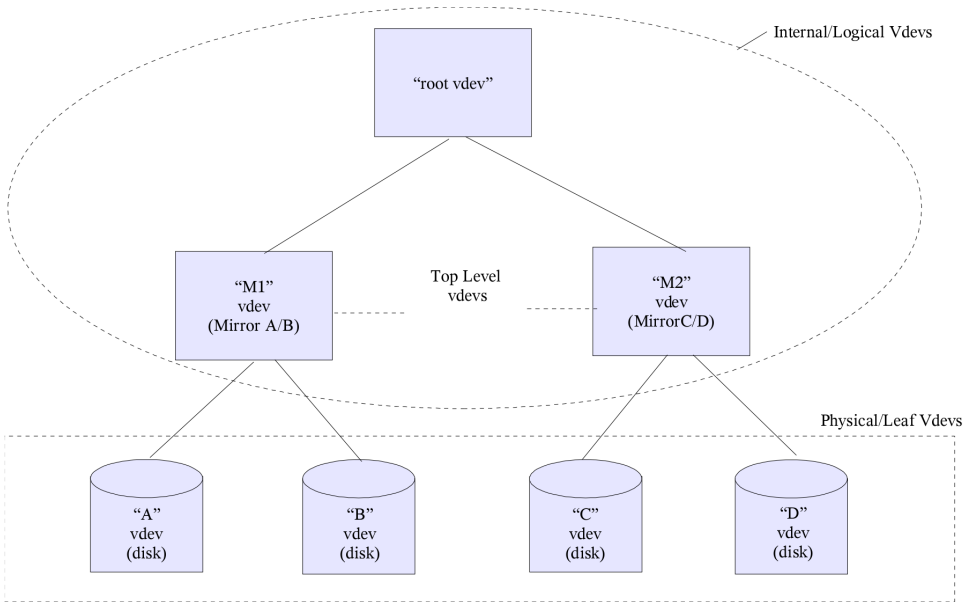



Figuur: Illustratie van ZFS pooled storage (rechts) t.o.v.volume-based storage (links) (Bonwick e.a., 2002)

[2] Afhankelijk van de situatie

# VDEV's: Virtual Devices

- Bouwstenen van storage pools
- RAID-niveaus binnen ZFS:
  - Stripes, Mirrors, RAID-Z, etc.
- Speciale VDEV's:
  - SLOG, L2ARC




Figuur: Conceptuele voorstelling van VDEV's in een boomstructuur (Sun Microsystems, 2006)

# Voorbeeld: zpool met een RAID-Z VDEV

```
$ zpool create storage raidz1 /dev/sda /dev/sdb /dev/sdc
$ zpool status
  pool: storage
  state: ONLINE
  scan: none requested
  config:

NAME          STATE     READ WRITE CKSUM
storage       ONLINE       0     0     0
  raidz1-0    ONLINE       0     0     0
    sda       ONLINE       0     0     0
    sdb       ONLINE       0     0     0
    sdc       ONLINE       0     0     0

  errors: No known data errors
```

# Inhoud

# Benchmarks

FIO-benchmark: aantal IOPS (Invoer/Uitvoer-bewerkingen per seconde)

# Referenties I


- Bonwick, J. e.a. (2002, Unknown). *The Zettabyte Filesystem*. Verkregen van http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.184.3704&rep=rep1&type=pdf
- Kendi, C. (Onbekend). ZFS: Enhancing the Open Source Storage System (and the Kernel). Verkregen van https://www.blackhat.com/presentations/bh-dc-10/Kendi_Christian/Blackhat-DC-2010-Kendi-Enhancing-ZFS-slides.pdf
- Sun Microsystems. (2006). *ZFS on-disk specification*. Verkregen van http://www.giis.co.in/Zfs_ondiskformat.pdf

# Zijn er nog vragen?