



Roghanna, Roghanna

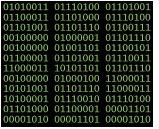
(Choices, Choices)

```
01010011 01110100 01101001
01100011 01101000 01110100
01101001 01101110 01100111
00100000 01000001 01101110
00100000 01001101 01100101
01100001 01101001 01110011
11000011 10101101 01101110
00100000 01000100 11000011
10101001 01101110 11000011
10100001 01110010 01110100
01101000 01100001 00001101
00001010 00001101 00001010
```

Harmen de Ruiter & Martijn van Toor

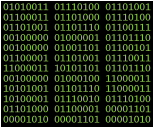
Stichting An Meaisín Dénártha

Programma



- Even Voorstellen
- Voor wie?
- Virtualisatie
- Proxmox VE
- KVM
- ZFS
- Demo Promox VE
- Vragen

Even Voorstellen



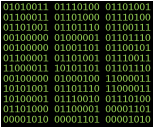
- Harmen de Ruiter
- Martijn van Toor
- Stichting An Meaisín Dénártha

Harmen de Ruiter



- Sr. Systeembeheerder
- Sinds 1996 werkzaam in de ICT
- Eerste kennismaking met Linux/OSS in 1996
- Bekend met Windows, Linux, AS/400, VMS, FreeBSD, etc, etc, etc.
- Bedenker en ex-voorzitter van Hostingvereniging Soleus
- Bedenker en voorzitter van Stichting An Measín Dénártha

Martijn van Toor



- Programmeur / Scrummaster /
Systeembeheerder light
- Sinds ongeveer 2003 werkzaam in de ICT
- Werkzaam in de Goede Doelen Branche
- Verantwoordelijk ontwikkelaar van
ProjectConnect, het eerste pakket
waarin goede doelen met hun partners
cloudbased kunnen samenwerken.
- Bestuurslid van Stichting An Measín
Dénártha

Stichting An Meaisín Dénártha



- Stichting
- Opgericht op 17 januari 2016
- Eigen hardware bij Datacenter Arnhem
- (VPS) hosting en diensten
- Kennisdeling
- Samenwerking
- Maar bovenal, een community!

Vragen?



Graag aan het einde. :)

Voor wie?



- Iedereen met de behoefte aan
 - Mailserver
 - Webserver
 - Centrale betrouwbare opslag bestanden
 - Backups
 - Spielerei (denk aan meerder OSen of distro's, test-vm's)
 - Kortom: Iedereen!

Waarom voor iedereen?

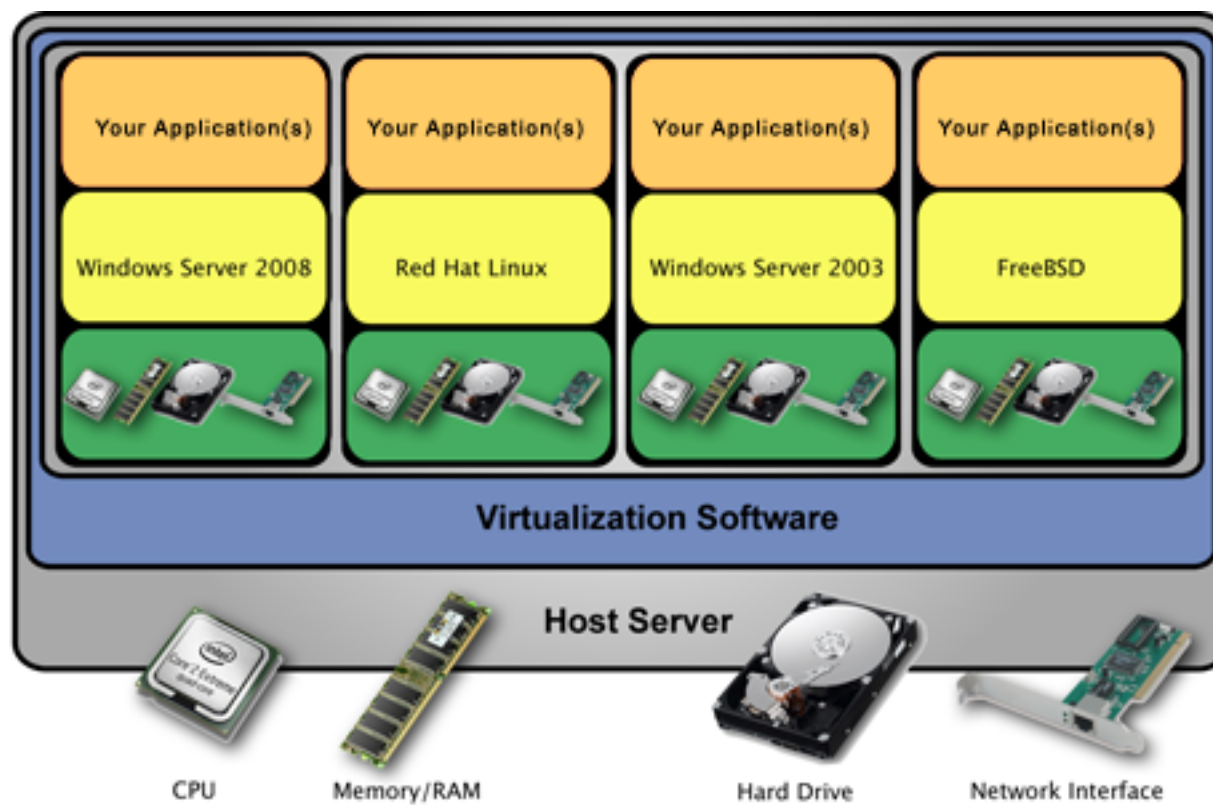


Je met virtualisatie eenvoudig met
meerdere systemen en operating systems
kunt testen.

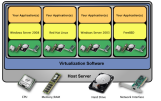
Je efficiënter gebruik kunt maken van je
hardware.

virtualisatie leerzaam en leuk is!

virtualisatie

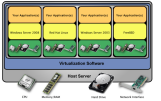


Voordelen (1/2)



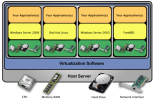
- Je kan makkelijk meerdere systemen op 1 fysieke machine draaien.
- Efficiënt gebruik van over het algemeen overbemeten hardware
- Flexibiliteit door eenvoudig toewijzen van resources, zoals CPU, memory, disk, netwerkkaarten etc.

Voordelen (2/2)

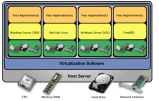


- Eenvoudig backups maken van complete systemen die elders opnieuw op te bouwen zijn
- Eenvoudig clonen van systemen
- Hogere betrouwbaarheid eenvoudig te realiseren door middel van clustering, ofwel een HA (High Availability) omgeving
- Maakt het „appliance“ principe makkelijker

Nade1en



- Je verliest iets aan performance door dat de virtualiseren; 2-3 procent.
- Complexere setup dan een enkele machine



Overwegingen

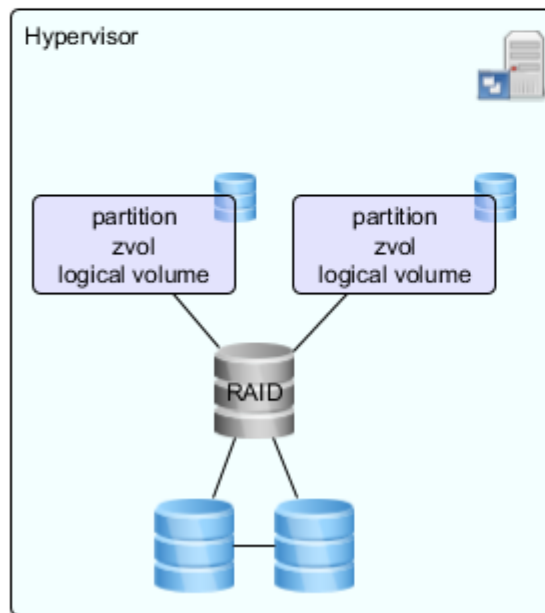
- Niet alles wil / kan je virtualiseren; denk aan hardware voor tapedrives ed.
- Grafische applicaties
- Het kan zijn dat een bepaalde taak dusdanig zwaar is, dat het goedkoper / efficiënter is om hem op een dedicated machine te installeren, en niet te virtualiseren
- 1 VM kan de hele node 'plat' trekken

virtualisatie - Opslag

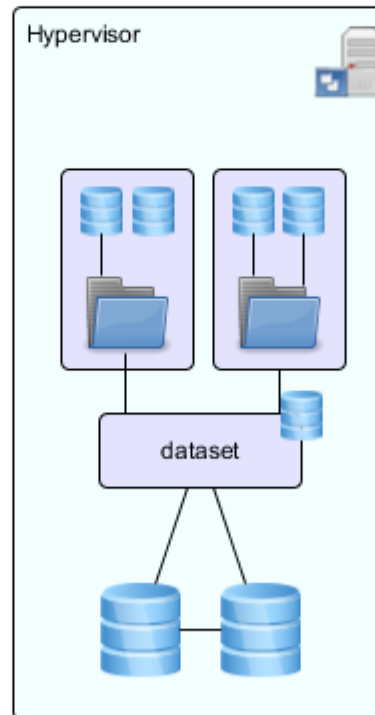


virtualisatie - Opslag

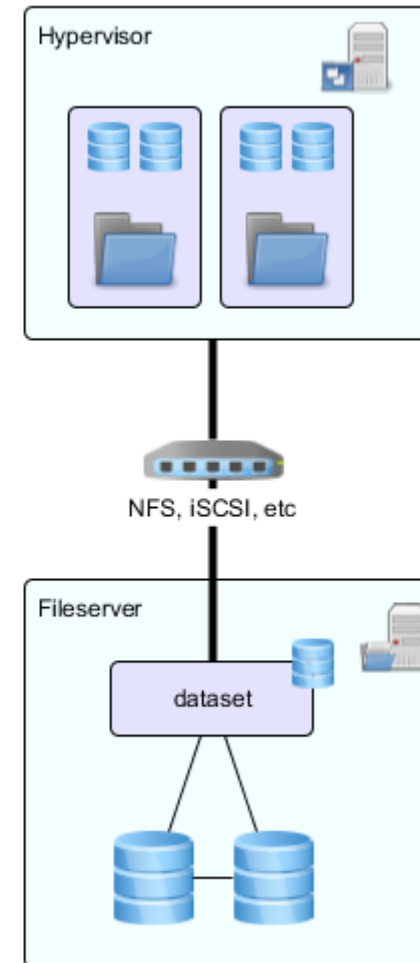
Overview



- Lokale opslag



- Lokale opslag
- Images



- Centrale opslag
- Images

Lokaal



- Voordelen
 - Lage latency, maximale doorvoersnelheid
- Nadelen
 - HA setup kan, maar alleen dmv replicatie (DRBD)

Image + Lokaa1



- Voordelen
 - flexibeler in het verplaatsen van vm's
 - mogelijkheid tot migratie naar centrale storage
- Nadelen
 - performanceverlies door het gebruik van images

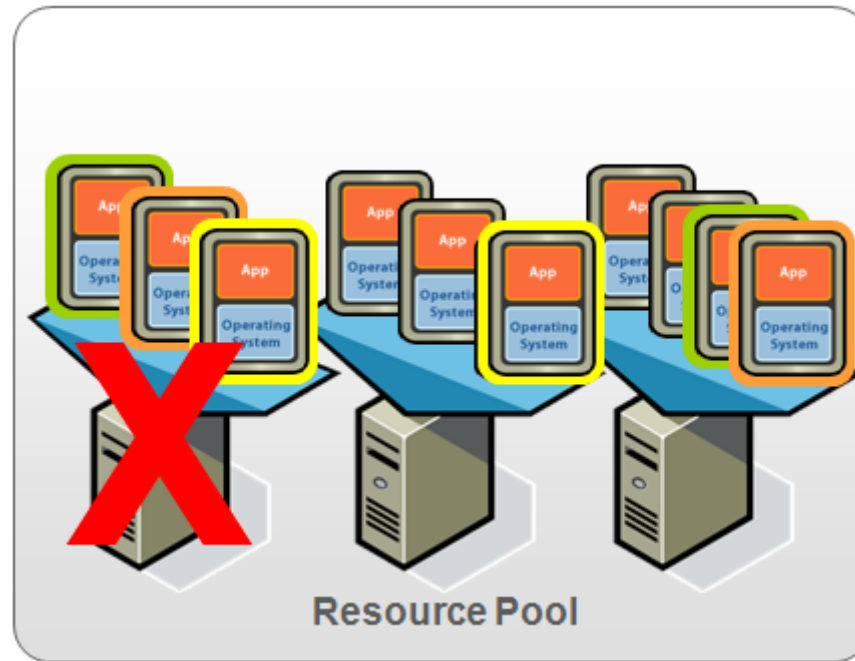
Image + Transport + Remote



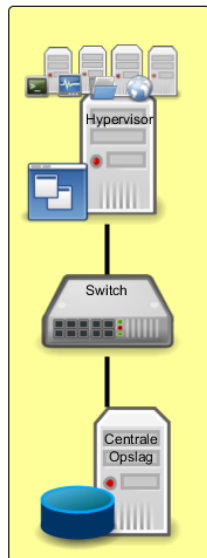
Zo doen de grote jongens het!

- Voordelen
 - HA setup op hypervisors mogelijk
 - Schaalbaar
- Nadelen
 - Trager (beperking door netwerk)
 - Netwerk componenten (switches) nodig
 - Voor meer snelheid, duurdere (L2 managed) switches nodig of 10Ge kaarten + switches (duur!)

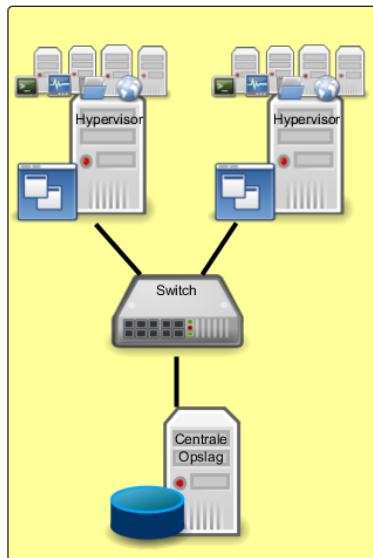
High Availability & Schaalbaarheid



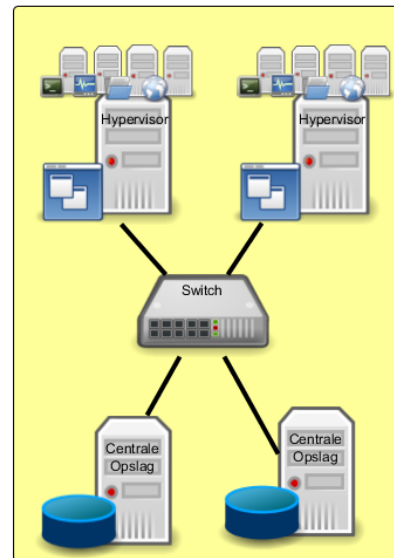
HA & Schaalbaarheid - Overview



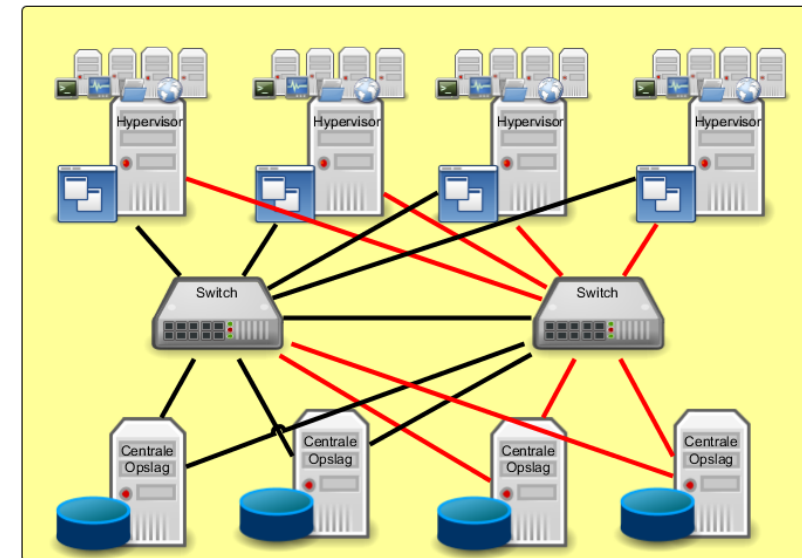
- Enkele Hypervisors
- Enkele centrale opslag



- Meerdere Hypervisors
- Enkele switch
- Enkele centrale opslag

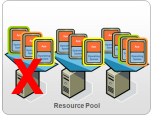


- Meerdere Hypervisors
- Enkele switch
- Meervoudige Centrale opslag



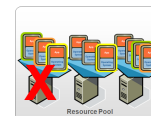
- Meerdere Hypervisors
- Meerdere switches
- Meervoudige Centrale opslag

Meerdere hypervisors



- Eenvoudig vm's van de ene hypervisor naar de andere migreren zonder downtime
- Andere hypervisors in het clusters kunnen dit nagenoeg naadloos overnemen
- Downtime beperkt of niet bestaand

Meerdere opslag systemen



- Eenvoudig vm's van het ene opslagsysteem naar de andere migreren zonder downtime
- Andere hypervisors in het clusters kunnen dit nagenoeg naadloos overnemen
- Downtime beperkt of niet bestaand

Meerdere hypervisors + Meerdere opslag Systemen



- Best of both :-)
- Maar tegen hoge kosten

Schaalbaarheid



Van klein ...



- 1x HP Microserver Gen8
 - CPU: Intel Celeron G1610T
 - Memory: 16GByte DDR3
 - Disks: 2-4x HDD
 - Prijs: ~800 EUR



Via iets groter...



- 1x HP Microserver Gen8
 - CPU: Intel Celeron G1610T
 - Memory: 16GByte DDR3
 - Disk: 1x SSD
- 1x HP Microserver Gen8
 - CPU: Intel Celeron G1610T
 - Memory: 16GByte DDR3
 - Disk: 2-4x HDD
 - Prijs (totaal): ~1500 EUR



Tot nog wat groter...



- HP Storageworks P4500 G2 "LeftHand"
- CPU: 2x Intel Xeon E5620 @ 2.40GHz
- Memory: 96GByte DDR3 (max. 144GByte)
- Disk Controller: LSI SAS 9207-8i
- Disks: 2x 128G SSD + 4x 5TByte HDD (max. 12 disks)



Tot extreem groot!



- Tianhe-2
- CPU: 32.000 Intel Xeon E5-2692 + 48.000 Xeon Phi 31S1P
- Memory: 1,375 TiB
- Storage: 12,4 PB
- Prijs: \$390 million



Platform



Commerciëel vs OSS



- Commerciëel
 - VMWare (ESXi, Workstation)
 - Microsoft Hyper-V
 - Citrix XenServer
- OSS
 - VirtualBox
 - XenProject
 - KVM
 - Proxmox (KVM + LXC)



Proxmox VE



PROXMOX

Waarom Proxmox VE



- Eenvoudig in installatie en gebruik
- Enterprise features
 - HA op storage & hypervisors
- Prijs
- 100% OSS
- KVM + LXC
- vele andere features

Voordelen



- KVM & LXC
- HTML5 GUI (inclusief console)
- Linux (debian) onderlaag
- Out-of-the-box oplossing (werkt meteen en zonder toevoegingen)
- Authenticatie tegen diverse backends
- Commerciële (betaalde) support mogelijk
- Clustering van hypervisors mogelijk

Nade1en



- Minder „enterprise“ dan bijvoorbeeld VMWare of XenServer (minder geschikt voor echt grote omgevingen)

Features - Filesystems



- Filesystem support
 - ZFS
 - NFS
 - iSCSI
 - Etc.

Features – Image formats



- Ondersteuning voor diverse formaten images.
 - QCOW2, Native formaat van KVM
 - VMDK, Native formaat van VMware
 - RAW

Features – Virtualisatie



- Full-virtualisatie (KVM)
 - Alle operating systems
- Paravirtualisatie (KVM)
 - KVM-aware operating systems
- Containervirtualisatie (LXC)
 - Linux only!

Features – Guest OSsen



- Draait (dankzij KVM) onder andere
 - Linux
 - FreeBSD
 - FreeDOS
 - Windows
 - ... en nog een hele hoop andere operating systems

Features – Overig



- Live migratie (inclusief storage, en het veranderen van image-formaat)
- Snapshots en backups
- Role-based access
- Resource pools (vooral voor functiescheiding)
- KVM word gesupport / gebouwd door RedHat (virtIO drivers)
- Native ZFS ondersteuning (eigen kernel)

Storage



RAID?



- Redundant
- Array
- of
- Independent (Inexpensive, orig.)
- Disks

RAID vs Enkele disks



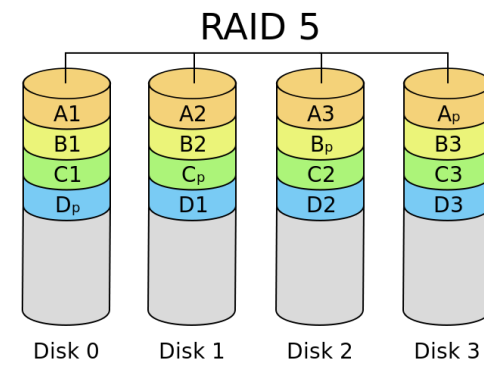
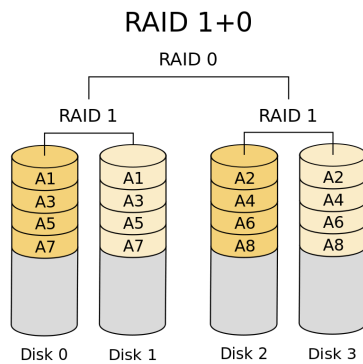
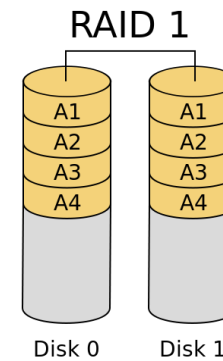
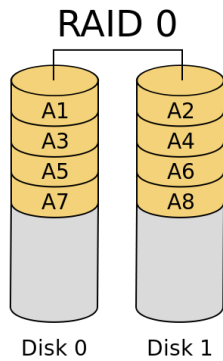
- Voordelen van RAID (vs 1 disk)
 - Verminderde kans op dataverlies
 - Meer snelheid
- Nadelen van RAID (vs 1 disk)
 - Overhead (meer disks nodig, kostbaarder)

RAID is geen backup!



- RAID is geen backup!
- Het beschermt tegen bepaalde vormen van hardware falen, maar is niet onfeilbaar daarin.
- Dataverlies door malware, virussen of user-error is met RAID niet anders

RAID Levels



Hardware RAID – Kenmerken



- Wat is het
 - Berekeningen worden door gespecialiseerde hardware controllers uitgevoerd met eigen geheugen
- Wat heb je er voor nodig
 - Hardware RAID controller
 - Meerdere harddisks (minimaal 2, aantal afhankelijk van setup)

Hardware RAID – Voordelen



- Minder CPU belasting (al is dat met moderne CPU's verwaarloosbaar)
 - LET OP: NIET sneller, dat het de CPU in theorie minder belast, wil niet zeggen dat het ook sneller is omdat de RAID controller zelf alle berekeningen moet doen en daarmee dus juist trager kan zijn dan een moderne CPU
- OS onafhankelijk, in principe is de data toegankelijk voor een willekeurig OS

Hardware RAID – Nadeln



- Duur (hardware RAID controllers zijn duurder dan hun HBA equivalent, zeker met ingewikkelder RAID levels)
- Weinig flexibel (disks / arrays kunnen niet zomaar worden overgezet naar een ander type of RAID controller van een andere fabrikant)

Hardware RAID – Opmerkingen



- Ook de grote jongens (NetApp, doen met WAFL al jaren software RAID, hardware RAID verdwijnt meer en meer)

Software RAID – Kenmerken



- Wat is het
 - Berekeningen worden door software in CPU van het OS uitgevoerd
- Wat heb je er voor nodig
 - willekeurige disk poorten (SATA, SAS, on-board of HBA)
 - Eventueel een HBA (bij meerdere disks of als de snelheid omhoog moet)

Software RAID – Voor- en nadelen



- Voordelen
 - Flexibiliteit (disks kunnen zo over naar een andere installatie, compleet andere machine, andere disk controllers, etc, zonder dataverlies)
- Nadelen
 - Kost enige rekenkracht van de CPU
 - OS afhankelijk, je kunt een Linux RAIDset niet aan een Windows machine hangen

ZFS

Zetabyte
File
System

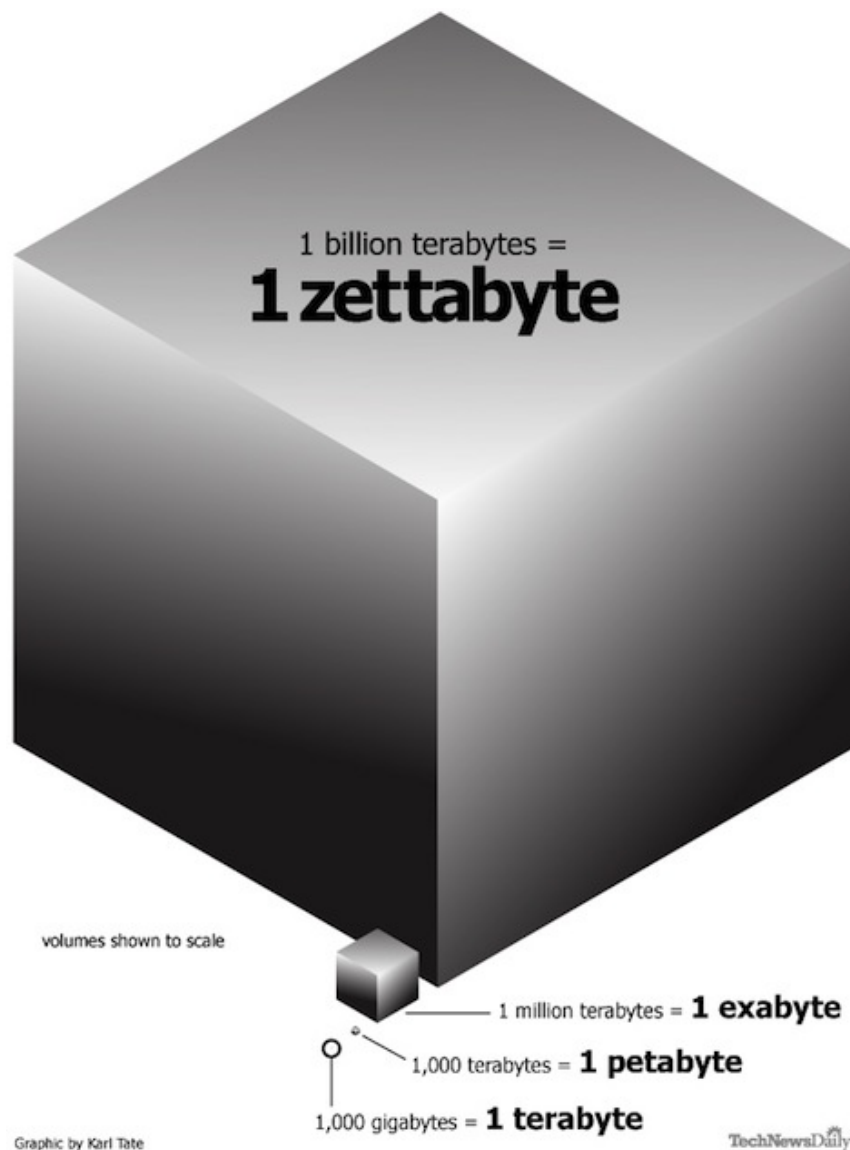


Open**ZFS**

Zetabyte?

- 1ZB = 10 maanden lang ál het internetverkeer
- 125 miljoen jaar van je favoriete TV show
- De chinese muur, volledig gebouwd van iPads
- 2x het hele internet in 2009

1ZB (schaal)





256 quadrillion ZB

- ZFS kan 256 quadrillion zettabytes opslaan (zeggen ze)
- 256,000,000,000,000,000 ZB dus...



ZFS – Voordelen (1/2)

- md, LVM en ext4 laag vervallen (geïntegreerd)
- Alle writes zijn “atomic”
- Snel, snel, snel, extreem snel
- Bitrot word gezien en gecorrigeerd
- RAID equivalenten mogelijk, inclusief RAID levels die niet bestaan (RAIDZ-3 == RAID7, 3 disks uitval mogelijk)



ZFS – Voordelen (1/2)

- Rebuilds bij uitval zijn sneller (ALLEEN de data die aanwezig is word gerebuild)
- ZFS send & ZFS receive
- Snapshots



ZFS – Nadeln

- RAIDsets (zpool's) kunnen niet qua aantal disks worden vergroot (!)
 - Disks kunnen wel worden vervangen door grotere exemplaren, na het vervangen van het laatste exemplaar neemt de capaciteit toe

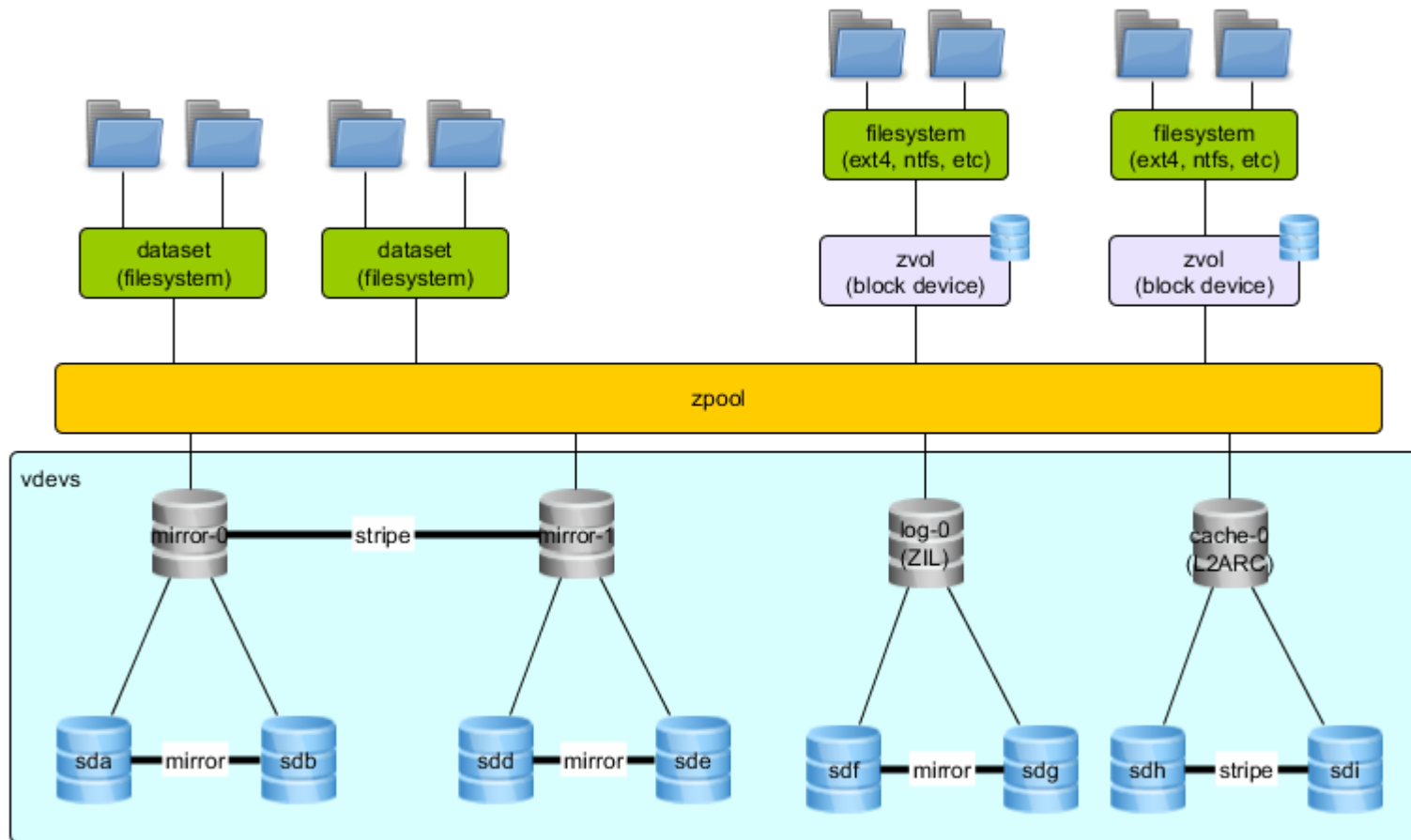


ZFS – Kenmerken

- Compleet anders denken (andere terminologie)
- ZFS wil graag **direct** bij zijn disks, dus een HBA in zgn. IT mode, ABSOLUUT GEEN RAID controller (ook geen P410 met per disk een RAID0 array)
- Relatief nieuw op Linux (stable in 2013)

ZFS – Zooming in...

ZFS - Terminologie





ZFS – Terminologie

- VDEV
 - Metadevice
- RAIDZ
- Zpool
- Dataset
- ZVOL's
 - A ZVOL is a "ZFS volume" that has been exported to the system as a block device.

ZFS – Snelheid (Caching)

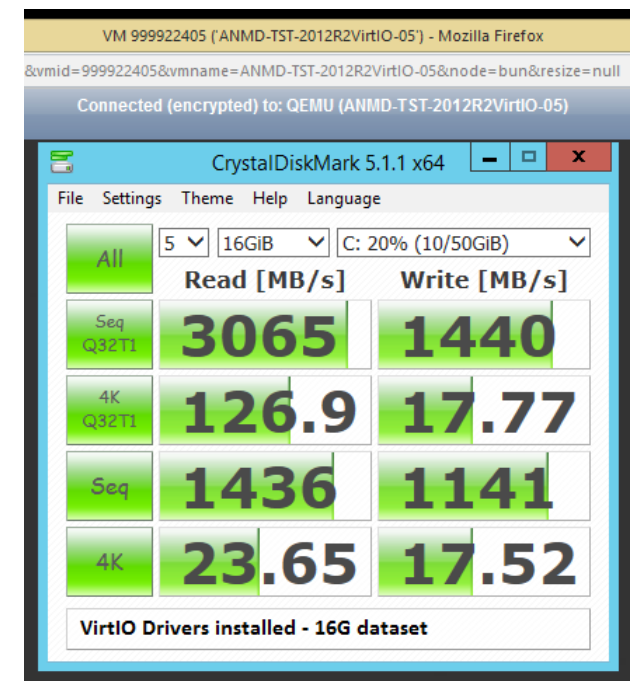
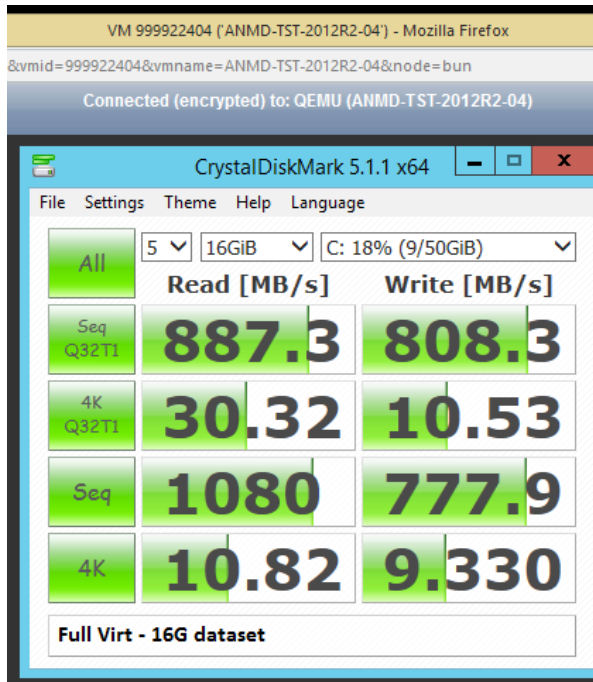
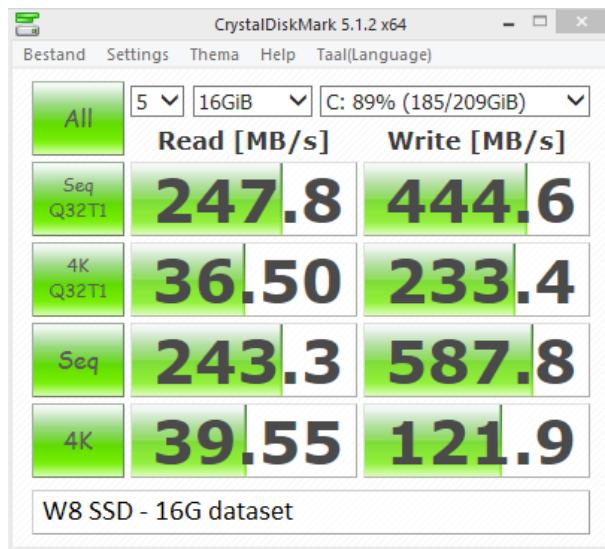
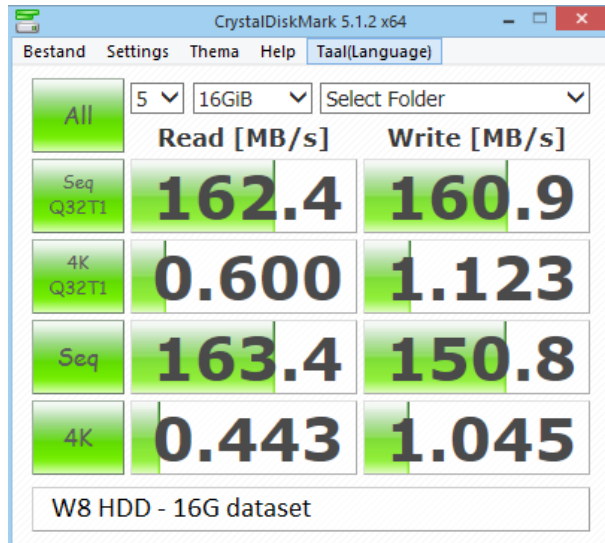
- Caching
 - Read + write caching mogelijk op losse disks
 - Later toe te voegen zonder downtime
 - TIP: Niet direct met caching op aparte disks beginnen, pas later meten als eea in productie is en DAN kiezen welke vormen van caching nodig zijn

ZFS – Snelheid (Compression)



- Compression
 - Door dit aan te zetten, gaat de performance omhoog.
 - Het comprimeren kost minder tijd dan het kost om de extra bytes weg te schrijven in het geval van traditionele harde schijven.
 - Het scheelt diskruimte

ZFS – Snelheid (benchmark)



ZFS – Bitrot

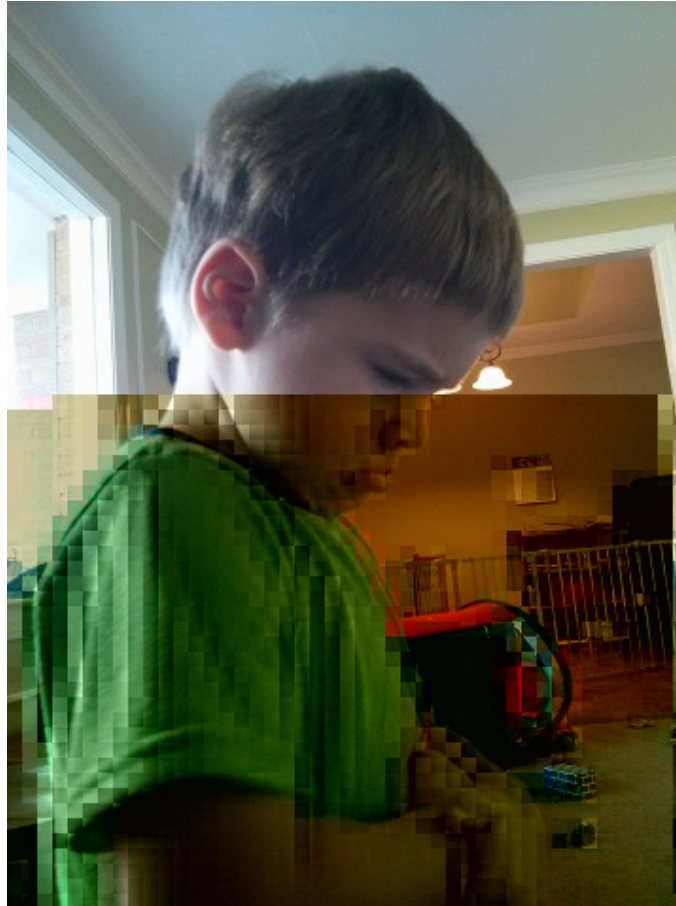
- Voorbeelden afkomstig van ArsTechnica
„Bitrot and atomic COWs: Inside “next-gen” filesystems“

<http://arstechnica.com/information-technology/2014/01/bitrot-and-atomic-cows-inside-next-gen-filesystems/3/>

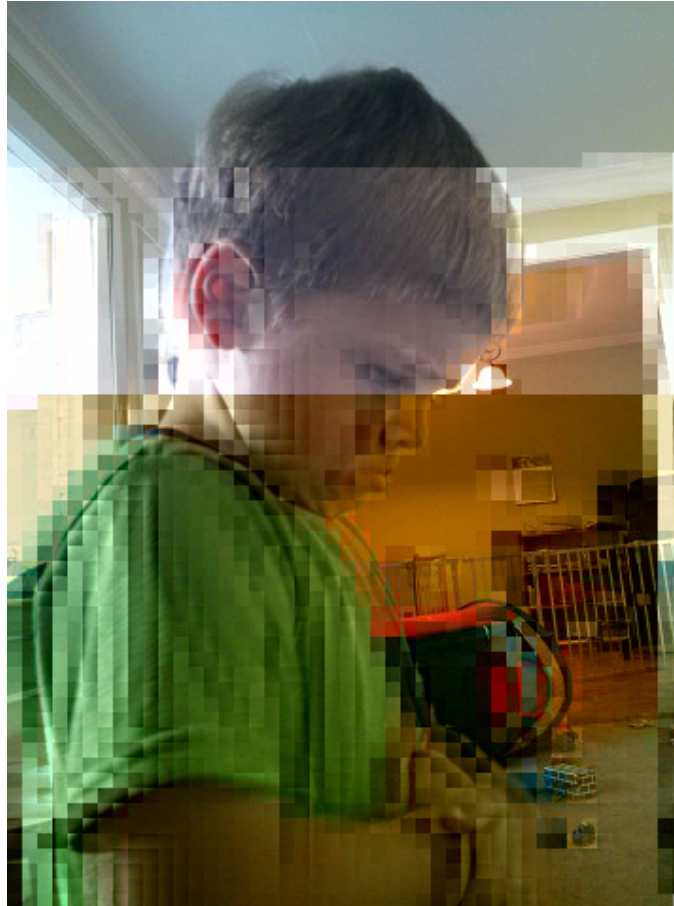
Bitrot - Original



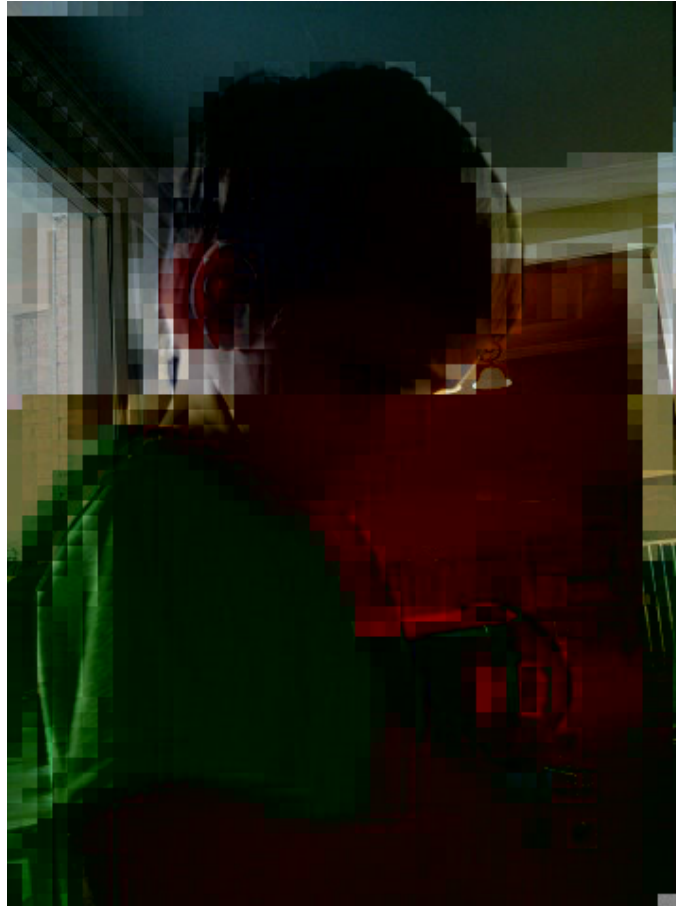
Bitrot – 1 bit flipped



Bitrot – 2 bits flipped



Bitrot – 3 bits flipped



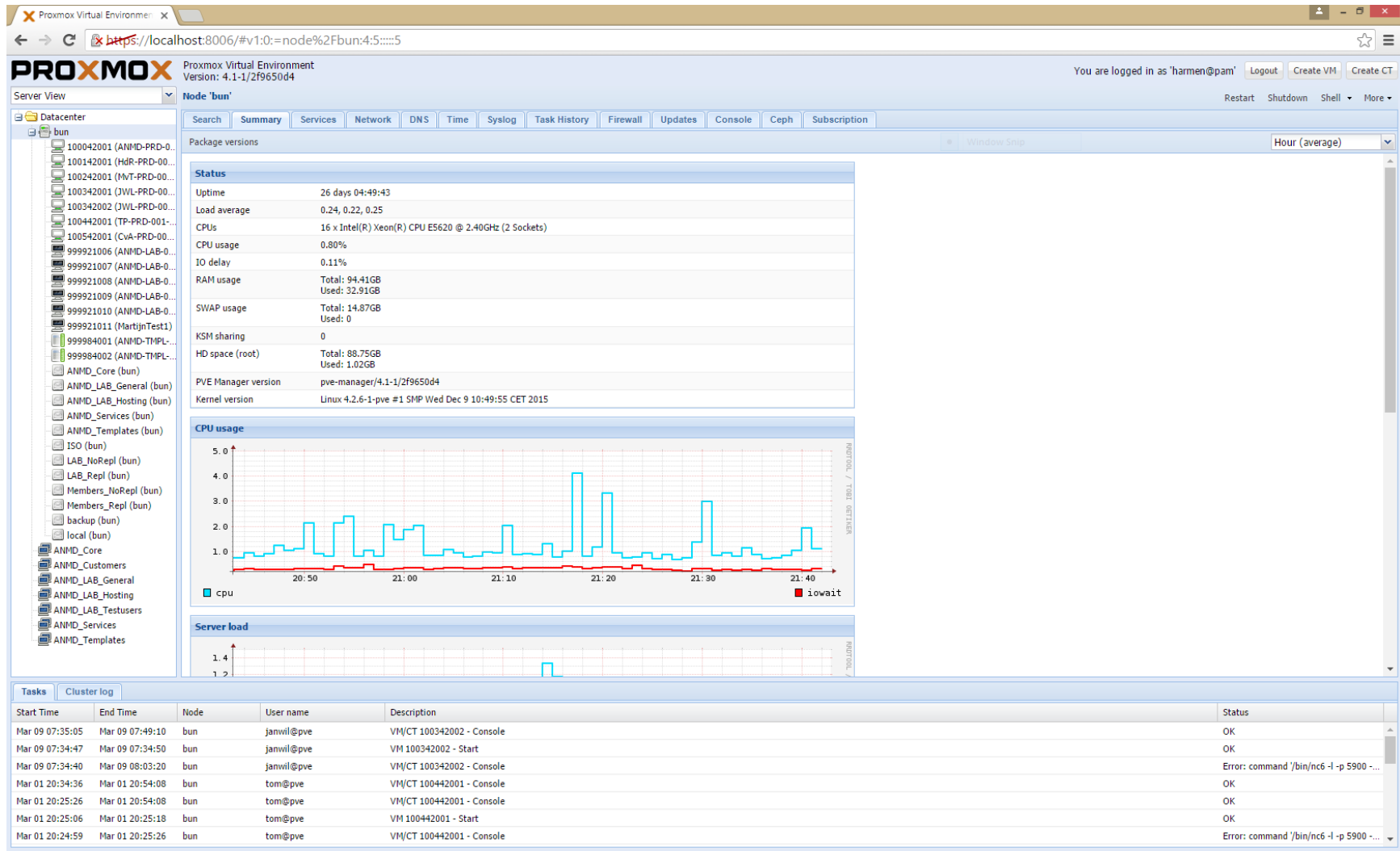
Bitrot – ZFS gecorrigeerd



ZFS – Keuzes

- Striped mirror (RAID10 equivalent)
 - Snellere rebuild tov RAIDZ
 - Minder stress op de disks dan tijdens een RAIDZ rebuild
 - Meer aaneengesloten disk capaciteit dan enkel mirroring
 - Sneller dan enkel mirroring of striping (RAID0 of RAID1 equivalent)

Demo time!





Vragen?

Links

- Stichting An Meaisín Dénártha
 - <http://anmd.org>
 - <https://www.facebook.com/anmeaisindenartha/>
 - <https://www.facebook.com/groups/anmeaisindenartha/>
- RAID.edu
 - <http://www.acnc.com/raid>
- Aaron Toponce on ZFS
 - <https://pthree.org/2012/04/17/install-zfs-on-debian-gnulinux/>
- ArsTechnica on BitRot
 - <http://arstechnica.com/information-technology/2014/01/bitrot-and-atomic-cows-inside-next-gen-filesystems/3/>