1. Lab: Proxmox HA

Modul Cloud Computing (cloud)

Simon Wächter

2019

Inhalt

[1 Einleitung 2](#_Toc22390159)

[1.1 Hinweis 2](#_Toc22390160)

[1.2 Homelab 2](#_Toc22390161)

[1.3 Schritt 1: Proxmox installieren 4](#_Toc22390162)

[1.4 Schritt 2: Cluster konfigurieren 5](#_Toc22390163)

[1.5 Schritt 3: Erstellen einer VM auf dem internen Storage (Debian1) 6](#_Toc22390164)

[1.5.1 Schritt 4: Erstellen einer VM auf einem NFS Storage (Debian2) 6](#_Toc22390165)

[1.6 Schritt 5: Ressourcen Pools mit Ballooning 7](#_Toc22390166)

[1.7 Schritt 6: verschieben von Debian2 via HA 7](#_Toc22390167)

[1.8 Schritt 7: Abschlussscreenshot 8](#_Toc22390168)

[1.9 Frage 1 zur Hochverfügbarkeit 8](#_Toc22390169)

[1.10 Frage 2 und 3 zur Hochverfügbarkeit 9](#_Toc22390170)

[1.11 Fazit 10](#_Toc22390171)

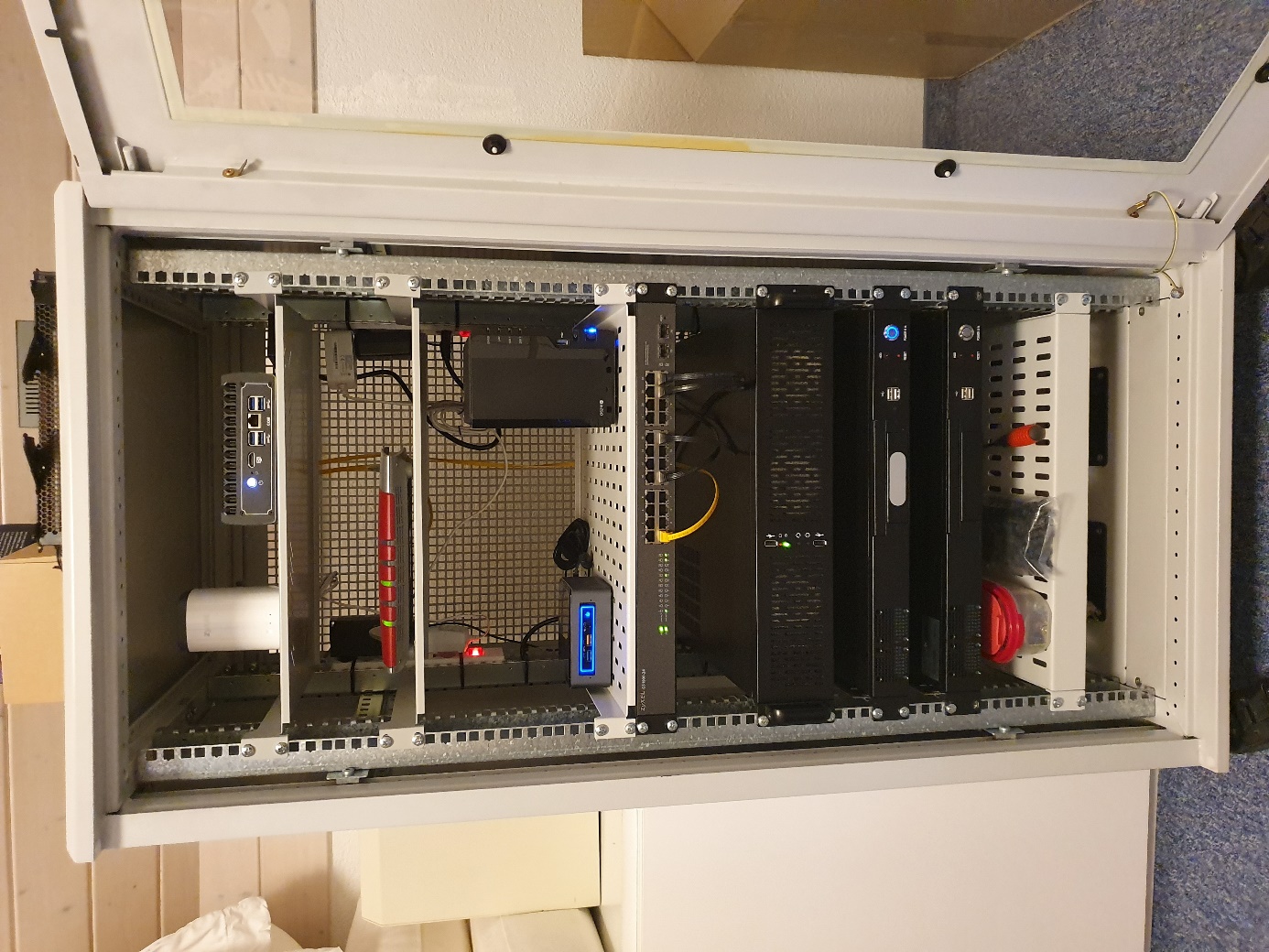
# Einleitung

## Hinweis

Da ich beruflich und privat in den Ferien war, konnte ich das erste Lab nicht an der Schule ausführen. Da ich aber über ein privates Homelab verfüge, konnte ich das Lab in abgeänderter Form nachholen (Andere IP/DNS Adressen etc.).

## Homelab

Blick auf mein Homelab samt kurzer Erklärung:



Erklärung:

* Weisser Tower (Oben links): Hauptinternetleitung über eine SIM Karte und 4G LTE+ im Bridge Modus (Einer der seltenen 4G LTE+ Router mit sauberem Bridge Mode). Das Internet ist auf dem Lande leider so langsam, das mobiles Internet schneller als Kupferkabel ist
* Schwarzer Box mit Kühlrippen (Oben rechts): Hauptrouter mit pfSense und mehreren Netzwerkkarten und Netzwerken (4 Netzwerke: DMZ, Homelab, Wasteland und Privat)
* Router WLAN Router: WLAN Router für Privates Netzwerk. Dort steht mein Deskop PC und Drucker
* Intel NUC (Mitte links): VMware vSphere Controller, der die drei ESXi Server verwaltet und steuert
* Synology NAS (Mitte rechts): Samba Server + Backup der VMs
* Managed Switch: Switch, welcher die VLANs an alle Systemen bringt
* Schwarzer 2U Server: 1. ESXi Server, Hauptsystem (Auf diesem wird das Lab ausgeführt)
* Schwarzer 1U Server, blau leuchtend: 2. ESXi Server
* Schwarzer 1U Server, nicht blau leuchtend: 3. ESXi Server, momentan Netzteilschaden

Warum erwähne ich das Setup überhaupt: Ich habe eine VMware User Group Lizenz und kann somit Features wie HA und Live Migration vom VMware verwenden. Mich nimmt es wunder, wie sich das Open Source Projekt Proxmox mit dem VMware Stack vergleichen lässt.

Meine Gedanken:

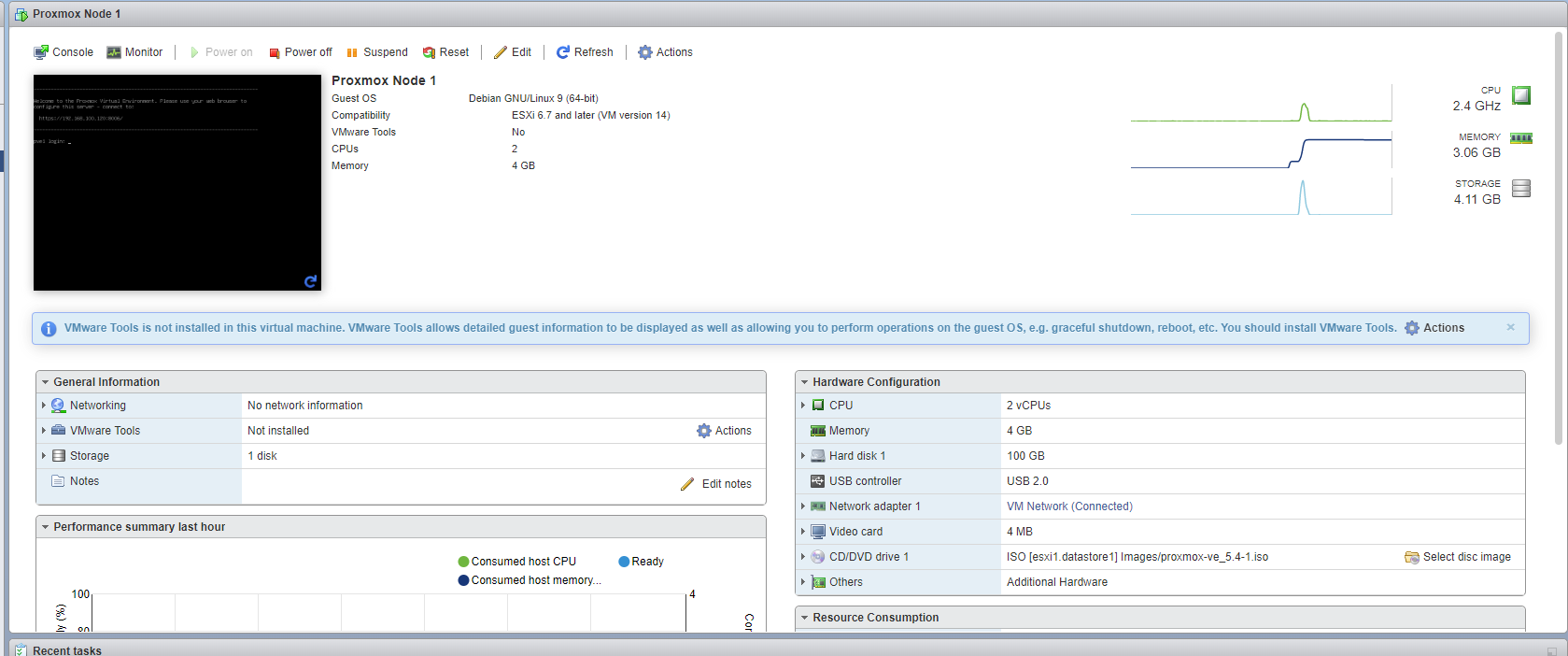
* Alle meine ESXi Server haben ihren Storage in den Server integriert, also kein iSCSI/NFS SAN/NAS Anbindung. Das Setup ist zwar recht einfach, doch ist eine Migration viel langsamer, als wenn das vSphere Center/Controller einfach die VM einer anderen Node via Live migration übergeben kann
* VMware Cluster brauchen ein vSphere Center, um das Cluster zu managen. Von dem was ich in der Labanleitung gelesen habe, managt sich ein Proxmox Cluster via eines Crash & Fail toleranten Konsensprotokoll. Heisst: Es wird keine dedizierte Kontrollnode benötigt, welche immer irgendwo laufen muss (Fun Fact: Splitbrain Situationen mit mehreren vSphere Centern können ganz schnell in einem riesen Chaos enden, wenn HA/Live Migration aktiviert sind)
* Nested Virtualisierung eines Hypervisors wird zusätzliche CPU Parameter benötigen, um die CPU voll durchreichen zu können (Fun Fact: ESXi kann man 4 Mal nesten, dann funktioniert der Netzwerkcontroller nicht mehr)

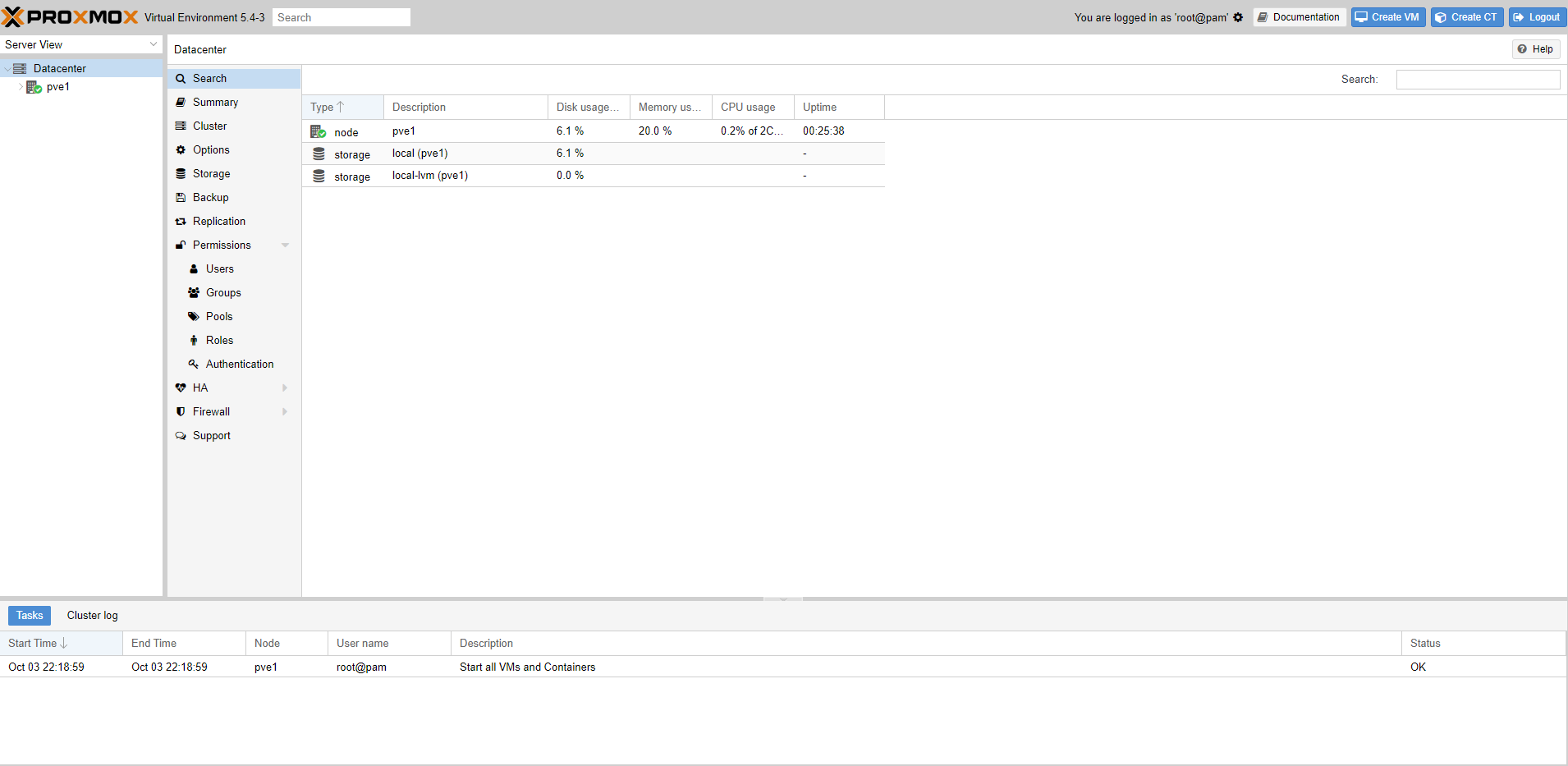
## Schritt 1: Proxmox installieren

Die Installation von Proxmox ist eigentlich selbsterklärend und wird hier nicht weiter erklärt. Für die verschachtelte Virtualisierung via ESXi müssen folgende Punkte beachtet werden (Video): <https://www.reddit.com/r/Proxmox/comments/9ws6c1/installing_proxmox_ve_52_nested_in_vmware_esxi/>

VM Konfiguration:

* CPU: 2 vCPU
  + Expose hardware assisted virtualization to the guest OS: Aktivieren
  + Expose IOMMU to the guest OS: Aktivieren
  + Enable virtualized CPU performance counters: Aktivieren
* RAM: 4 GB
* HDD: 100 GB Thin, Storage in Server
* Proxmox Version: 5.4 und nicht 5.2





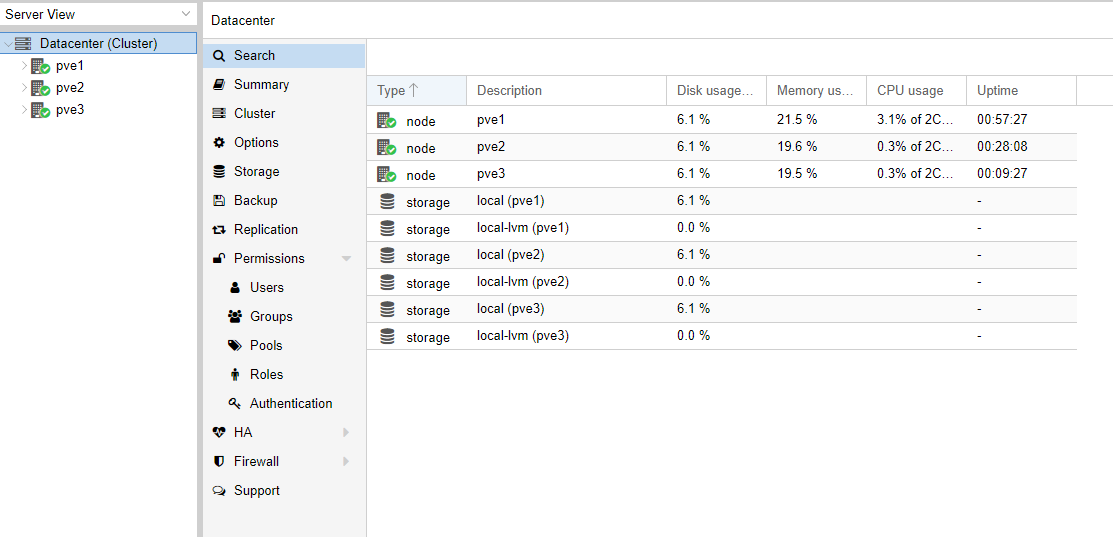
Systeme:

* NFS Server: 192.168.100.20 mit /volume1/PublicShare
* ESXi Server: 192.168.100.30
* Proxmox Node 1: 192.168.100.120 oder pve1.swaechter.ch
* Proxmox Node 2: 192.168.100.121 oder pve2.swaechter.ch
* Proxmox Node 3: 192.168.100.122 oder pve3.swaechter.ch

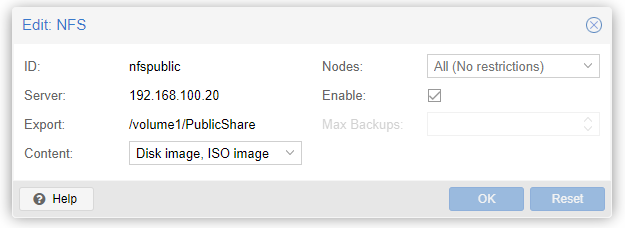
## Schritt 2: Cluster konfigurieren

Die Clustererstellung ist auch sehr einfach erklärt: <https://daniel-ziegler.com/proxmox-ve/computer/netzwerk/linux/2018/04/27/Proxmox-Virtual-Environment-Clustering/>

Das Erstellen und Joinen des Clusters via Terminal ist wirklich sehr angenehm und bestätigt meine Annahme, dass das Cluster über ein Quorum-Protokoll verfügt.



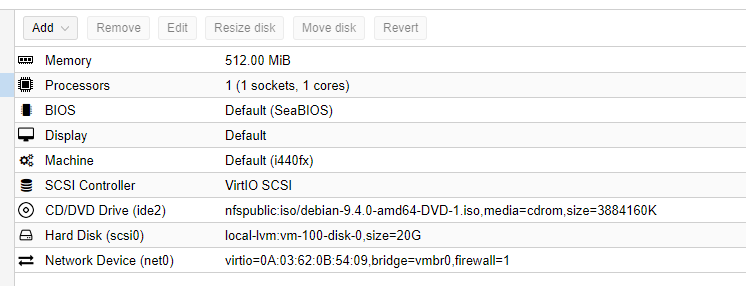
Danach muss noch der NFS Share hinzugefügt werden (Das Debian wurde via SCP schnell auf den NFS Share geschoben):



## Schritt 3: Erstellen einer VM auf dem internen Storage (Debian1)

Das Erstellen der VM hat sich auch als sehr einfach erwiesen. Folgende Fallstricke und Probleme sind aufgetreten:

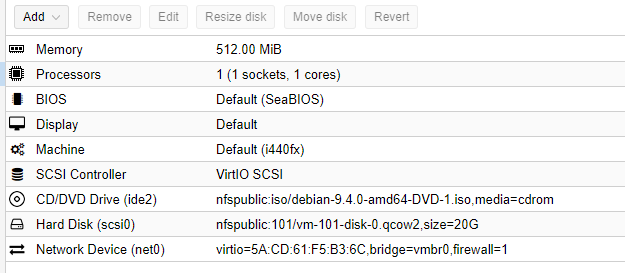
* Fallstrick: ISO Templates müssen nach template/iso im NFS Share liegen, da sie sonst in beim Erstellen der VM nicht auftauchen: <https://forum.proxmox.com/threads/nfs-isos-not-showing-in-ui.35873/>
* Fallstrick: Die VNC Webkonsole kann gewisse Sonderzeichen nicht durchreichen. Eventuell einfach einen lokal installierten Linux VNC Client verwenden. Zudem hängt sie sich bei einem Reboot der VM auf, sprich muss neu geöffnet werden
* Problem: Meine 3 Proxmox VMs verfügen über nur über je einen Netzwerkadapter ohne konfigurierten DHCP in proxmox, sodass die Debian Installer kein DHCP beziehen konnten (Und in einem solchen Fall sollte man kein Debian Netinstaller verwenden…)



### Schritt 4: Erstellen einer VM auf einem NFS Storage (Debian2)

Eigentlich wie bei Schritt 3, man muss einfach folgende Sachen beachten:

* Der konfigurierte NFS Share muss über das Content-Feature «Disk image» verfügen. Dies habe ich in weiser Voraussicht gleich aktiviert
* Als Storage natürlich auch den NFS Share auswählen



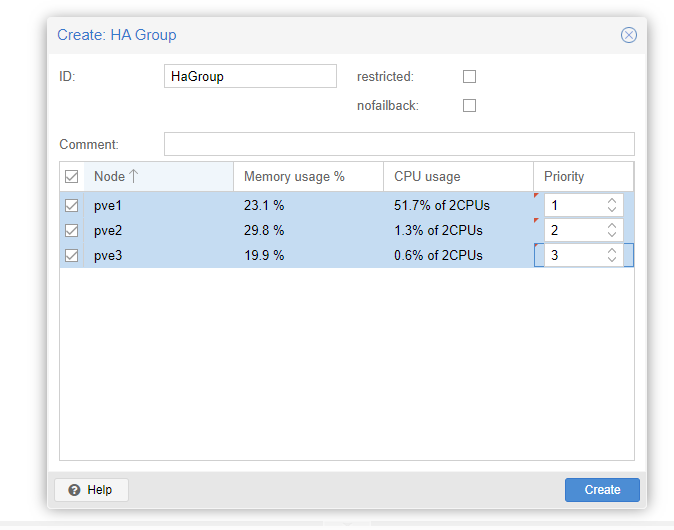
## Schritt 5: Ressourcen Pools mit Ballooning

Sehr gut beschrieben: <https://www.computerweekly.com/de/tipp/Proxmox-Konfiguration-Ressourcen-Pools-fuer-effizientere-virtuelle-Umgebungen>

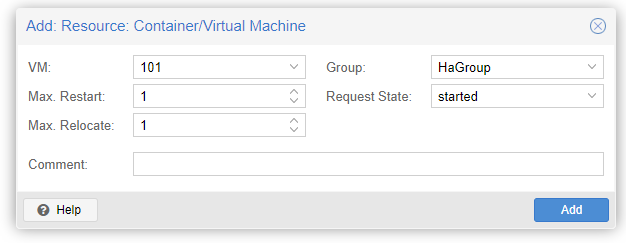
Aus Zeitgründen habe ich die Treiber nicht installieren können. Konzeptionell ist aber alles klar. Sozusagen wie die VMware Guest Tools

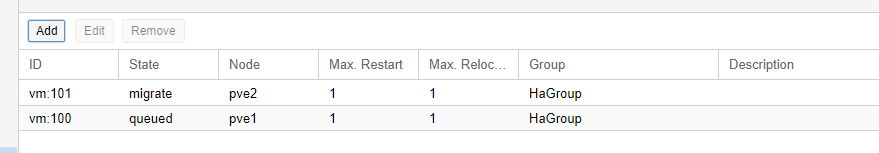
## Schritt 6: verschieben von Debian2 via HA

Bevor die VM als HA-VM deklariert werden kann, muss eine HA-Group erstellt werden: <https://www.thomas-krenn.com/de/wiki/Proxmox_HA_Cluster_erstellen>



Nach dem Erstellen der HA-Gruppe müssen dieser Gruppe auch die VMs zugewiesen werden:



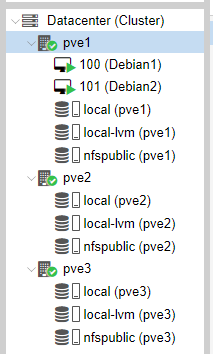


* Fallstrick: Prioritäten sind aufsteigend, nicht absteigend. Die Zahl 1 ist also nicht die 1. Priorität, sondern die niedrigste 3. Priorität in unserem Falle
* Fun Fact: Da vm101 zu dem Zeitpunkt schon auf pve2 war, wurde sie automatisch auf pve3 zurückmigriert, darum der State «migrate» 🡪 Wegen dem falschen Anwenden eben auf pve3 und nicht auf pve1. Lösung: Die Prioritäten aus der HA-Group so umkehren, dass pve3 Priorität 3, pve2 Priorität 2 und pve3 Priorität 1 hat. Kehrt man um, wird vm101 wieder migriert

Alles in allem wieder sehr übersichtlich und selbsterklärend. Warum kann VMware so komplex sein?!

## Schritt 7: Abschlussscreenshot

Sowohl Debian1 (Interner Storage) als auch Debian2 (NFS Share) wurden auf pve1 erstellt, und mehrmals nach pve2 und pve3 migriert. Dadurch, dass pve1 aber die HA Priorität 3 (Höchste) hat, wurde nach dem Erstellen der HA Gruppe wieder alles dorthin zurückmigriert:

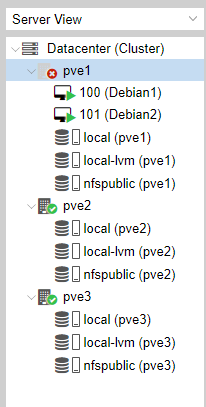


## Frage 1 zur Hochverfügbarkeit

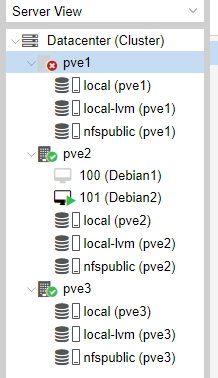
Unterschied der Migration einer VM auf einer internen Disk und einer Disk auf einem NFS Share: Bei einer laufenden VM auf einer internen Disk wird sehr wahrscheinlich ein Copy on Write System angewandt, um die VM Stück für Stück zu migrieren und dann der neuen Node die Kontrolle zu übergeben. Wird zu viel IO generiert, sodass das Delta nicht abgebaut werden kann, wird die VM wahrscheinlich auch mit künstlichen Schlafzyklen/NOPs gedrosselt. Bei einer VM auf einem NFS Share übergibt die bisherige Node der neuen Node einfach die Kontrolle, also keine Datenübertragung

## Frage 2 und 3 zur Hochverfügbarkeit

Ich habe via VMware Interface die Hauptnode pve1 heruntergefahren und auch direkt stoppen lassen. Wechselt man auf das Webinterface der pve2 Node erhält man zunächst folgendes Bild:



Nach kurzer Zeit ändert sich das Bild:



Für mich heisst das konkret:

* Bei einer auf einem iSCSI oder NFS Target/Share, also hier Debian2, übernimmt das Cluster nach einigen Sekunden die Kontrolle über die VM und startet sie neu, respektive lässt sie weiterlaufen (Ob ein Restart erfolgen musste, konnte ich leider nicht eruieren). Bei einem graceful Shutdown von pve1 kann pve1 die VM aber sicher an andere Nodes übergeben und es muss meiner Meinung nach kein Restart erfolgen
* ~~Was aber witzig ist: Auch Debian1, also die VM auf der Disk von pve1, wird auf pve2 gestartet. Konkret muss das heissen, dass das HA Feature die Disks mit gewissen Quality of Services auf anderen Nodes repliziert (@Gwerder: Stimmt das oder interpretiere ich da?)~~ Ok die VM wird zwar migriert, aber kann irgendwie nicht starten (Kurz nach dem zweiten Screenshot wurde sie als rot/fehlerhaft markiert). Heisst also, es gibt kein Kopieren der Disk, was mich eben auch verwundert hat. Will man HA einer VM, muss man einen externen Storage Driver wie iSCSI oder NFS verwenden (oder wie in VMware vSAN)

## Fazit

Ich muss ehrlich sein: Ich bin sehr positiv überrascht von Proxmox: Ich habe mir das Produkt vor circa 2.5 Jahren nebst oVirt und VMware mit ESXi/vSphere Center mal rasch angeschaut und es als «Bastelplatform» abgetan (An dieser Stelle sei mal die schlechte Dokumentation und nichtexistente Community von oVirt zu erwähnen…ah einfach RHEV lizenzieren, ok!). Aber Proxmox hat mich überrascht: Mit minimalem Googleaufwand konnte man sich rasch alles zusammensuchen und aufsetzen. Alles war dort, wo man es irgendwie erwartet hat – die Fallstricke waren minimal.

Würde ich beruflich nicht VMware ESXi und Microsoft HyperV verwenden, würde ich einen Wechsel in Betracht ziehen:

* Das Produkt ist freie Software und wirkt schlank (Ja ich meine dich ESXi/vSphere Webinterface: <https://www.reddit.com/r/vmware/comments/6fryzl/i_physically_cant_do_my_fucking_job_with_the_web/>, es ist einfach nur noch peinlich und zum heilen). Alles befindet sich dort, wo man es erwartet. Keine Probleme, sehr «polished»
* Supportlizenzen sind verhältnismässig sehr preiswert
* Ein Cluster benötigt keinen dedizierten Manager, das es über ein Konsensprotokoll kommuniziert (Das vereinfacht Splitbrain-Situationen und Ausfälle von Nodes enorm. Beispiel: Was passiert, wenn man den ESXi Server mit dem virtualisierten vSphere Center drauf in den Wartungsmodus verfrachtet?)
* HA-Gruppen sind sehr einfach und übersichtlich zu konfigurieren. Die Konfiguration ist so übersichtlich und intuitiv im Vergleich zu VMware
* Unter der Haube werden gängige Linux/Unix Technologien und Befehle verwendet. Liest man die Wikieinträge, weiss man gleich, wie Proxmox unter der Haube funktioniert (libvirt, LVM etc.)
* Support für Cloud-Init zum einfachen Bootstrappen von Linux Images (Reiner PXE ist immer etwas…nervig)
* Support für LXC Container («Docker Container» the correct Linuxish way + Docker als Firma hat immer noch kein deckendes Businessmodell)

Kurzum: Zu empfehlen. Bin gespannt auf den Ceph Storage!