Proxmox Best Practice Setup für ZFS und / oder Ceph im Cluster für Einsteiger (Stand Dezember 2024)

Proxmox Cluster mit ZFS und Ceph – Best Practices für Einsteiger

Praxiswissen aus rund 150 Proxmox- und ZFS-Systemen

Der Verfasser dieses Artikels, **Christian-Peter Zengel**, verfügt über rund **15 Jahre Praxiserfahrung mit Proxmox VE und ZFS** und betreibt aktuell etwa **150 Systeme mit Proxmox und ZFS**.

Die Erfahrungen reichen von Standalone-Installationen bis hin zu Proxmox-Clustern mit bis zu zehn Hosts. Der Schwerpunkt liegt klar auf **ZFS, Replikation, Backup, Monitoring und praxistauglicher Hochverfügbarkeit**.

Hinweis: Dieser Artikel versteht sich nicht als Ceph-Expertenleitfaden. Ceph wird hier vor allem zum Funktionsvergleich und als Einstieg behandelt. Für viele kleine und mittlere Installationen bis etwa zehn Hosts ist ZFS mit Replikation in unserer Praxis häufig die einfachere, günstigere und besser beherrschbare Lösung.

Proxmox Best Practice Kurs mit ZFS und Ceph

Diese Dokumentation basiert auf einem Onlinekurs von **cloudistboese.de**.

Proxmox Best Practice Cluster mit ZFS und Ceph für Einsteiger

Der Kurs vermittelt einen schnellen, praxisnahen Einstieg in den Aufbau eines Proxmox-Clusters mit ZFS, einfacher Ceph-Grundlage, Backup, Snapshots, Monitoring und Hochverfügbarkeit.

Aktuelle Kurse und Wiederholungen auf cloudistboese.de ansehen

Für wen eignet sich diese Anleitung?

Der folgende Artikel richtet sich besonders an Einsteiger, Administratoren und IT-Dienstleister, die mit mehr als einem Proxmox-Host starten möchten und eine robuste, nachvollziehbare Cluster-Grundlage suchen.

- Proxmox VE mit ZFS installieren
- Proxmox Cluster mit mehreren Nodes aufbauen
- ZFS-Replikation verstehen und kontrollieren
- VMs und LXCs performant anlegen
- einfache Ceph-Grundlagen kennenlernen
- Proxmox Backup Server integrieren
- Snapshots, Replikation und Monitoring verbessern
- Hochverfügbarkeit sinnvoll einordnen

Nach diesem Kurs können Sie einen kleinen Proxmox-Cluster sicherer planen, installieren und betreiben.

ZFS oder Ceph im Proxmox Cluster?

ZFS ist für kleine und mittlere Proxmox-Installationen eine sehr starke Grundlage. Durch Prüfsummen, Snapshots, Replikation und einfache Wiederherstellung schützt ZFS zuverlässig vor vielen typischen Problemen wie defekten Datenträgern, versehentlichen Änderungen und

manchen Folgen von Ransomware-Angriffen.

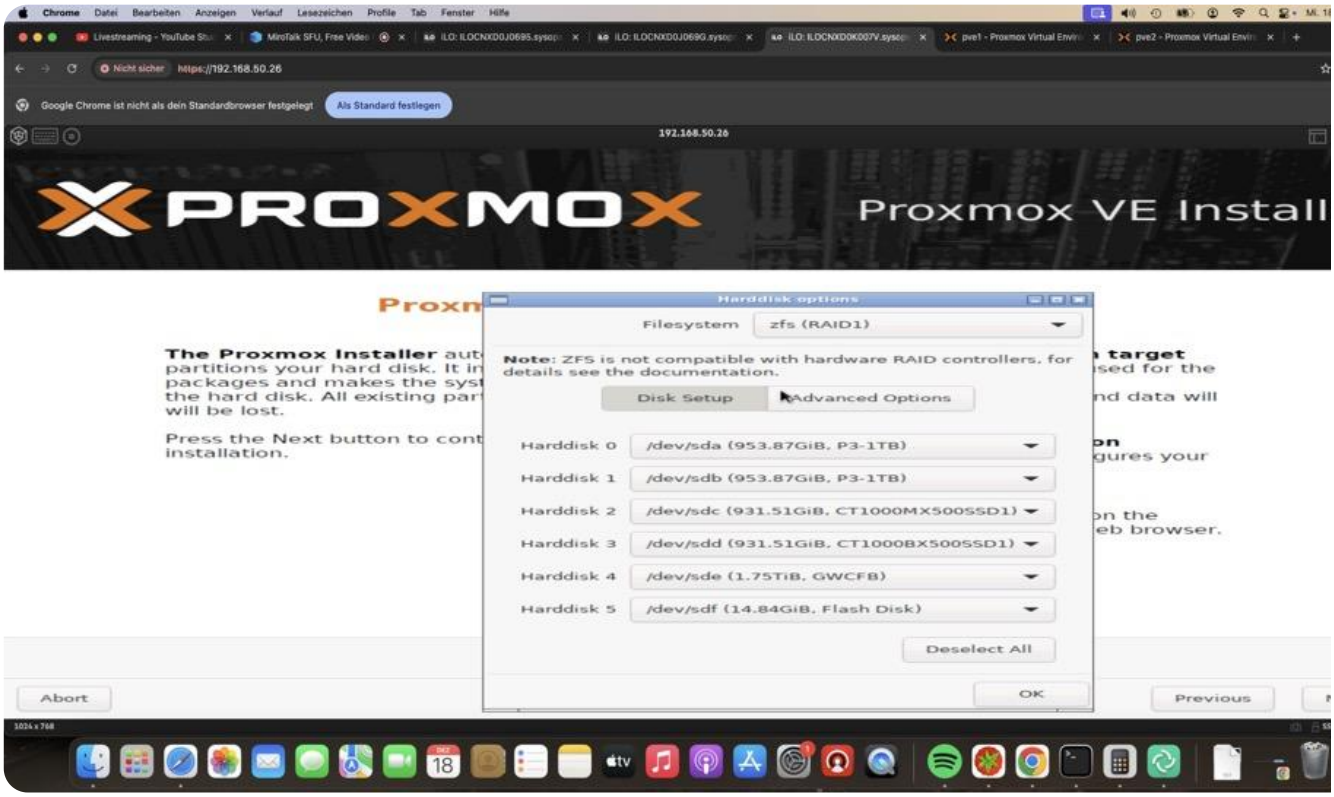
ZFS ersetzt jedoch kein Backup. Gegen kompromittierte Root-Zugänge, Brand, Diebstahl, vollständigen Standortausfall oder absichtlich gelöschte Snapshots hilft nur ein sauberes Backup- und Replikationskonzept.

Ceph bietet echten verteilten Storage und hohe Skalierbarkeit, benötigt aber deutlich mehr Hardware, Netzwerkplanung und Betriebserfahrung. Für viele Umgebungen bis etwa zehn Hosts ist ZFS mit Replikation oft wirtschaftlicher und einfacher zu betreiben.

Installation von Proxmox VE auf ZFS

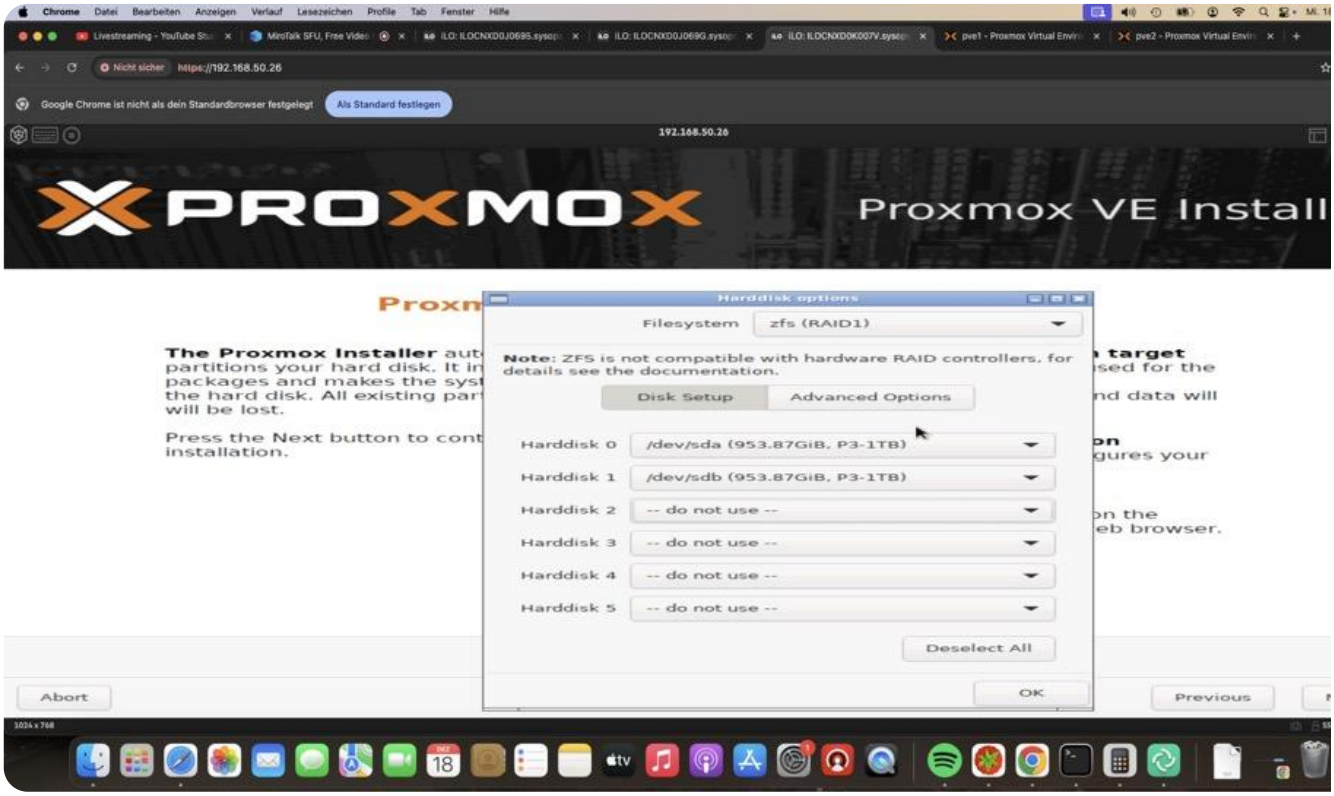
Für den Workshop wird ein HPE Microserver Gen10 Plus V2 verwendet. Empfehlenswert ist eine Ausstattung mit ECC-RAM und direkt ansprechbaren Datenträgern.

HPE Microserver Gen10 Plus V2 ansehen



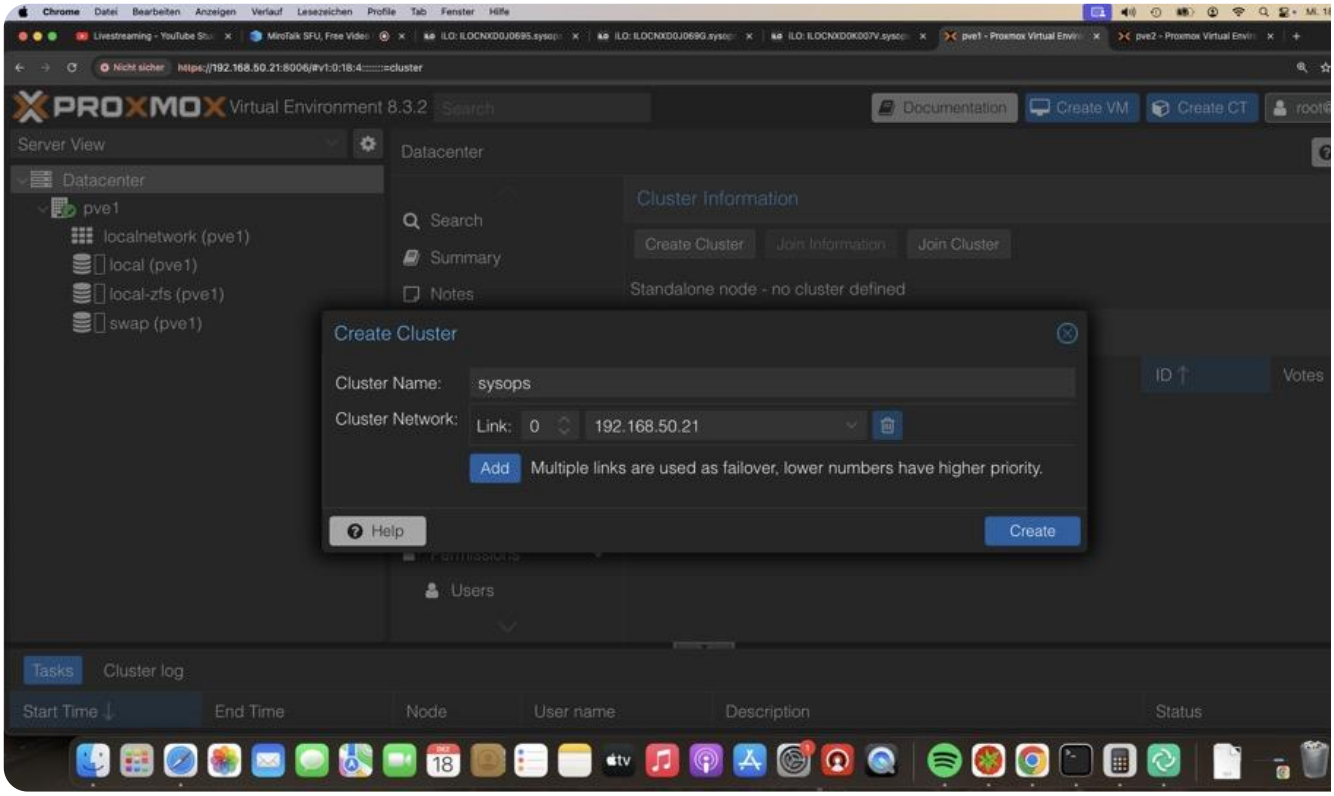
Wir installieren Proxmox VE bevorzugt auf **ZFS Mirror, RAID10 oder RAIDZ**. Hardware-RAID-Controller sollten für ZFS vermieden werden. Empfehlenswert sind Controller im HBA- oder IT-Mode, damit ZFS direkten Zugriff auf die Datenträger erhält.

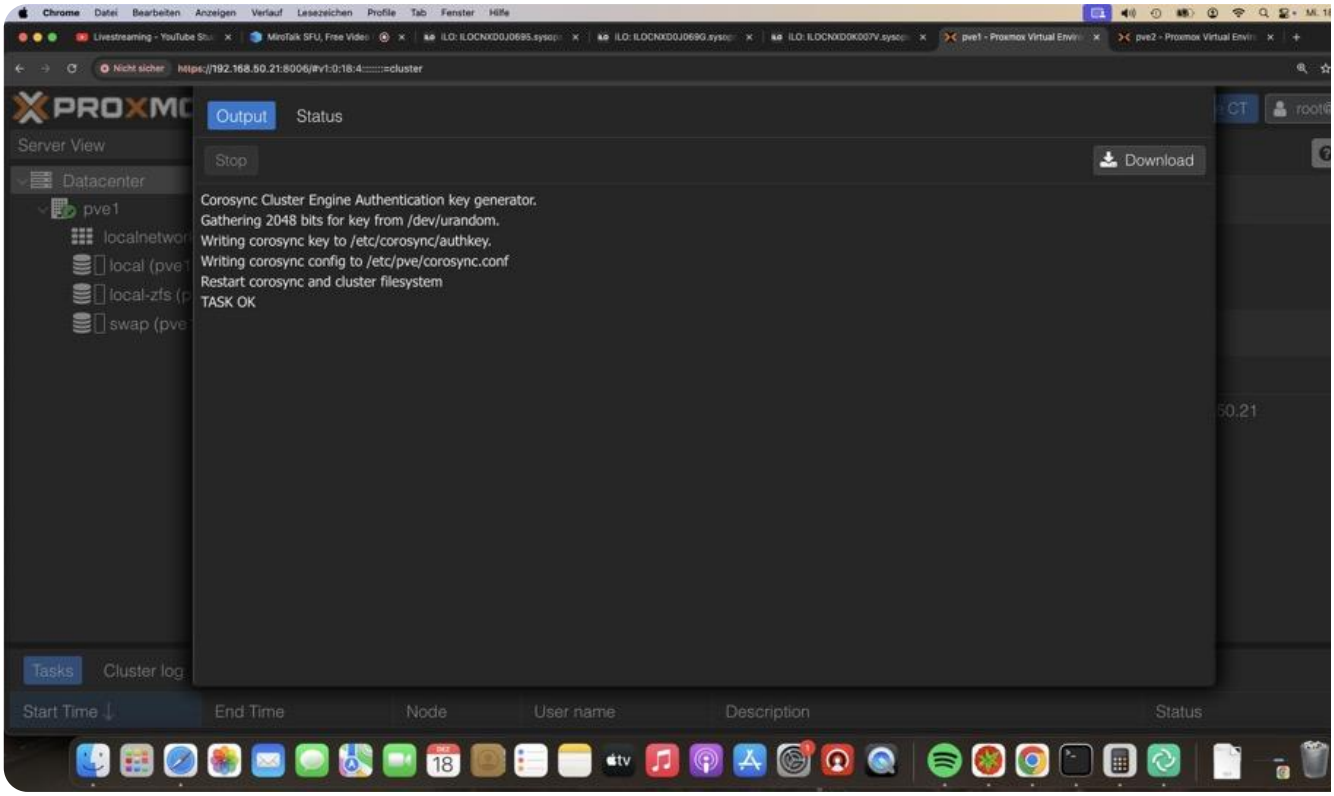
> Für Cluster sollten Hostnamen, IP-Adressen, Netzwerke und DNS-Einträge vor der Installation sauber geplant werden. Nachträgliche Änderungen verursachen häufig unnötige Zusatzarbeit.

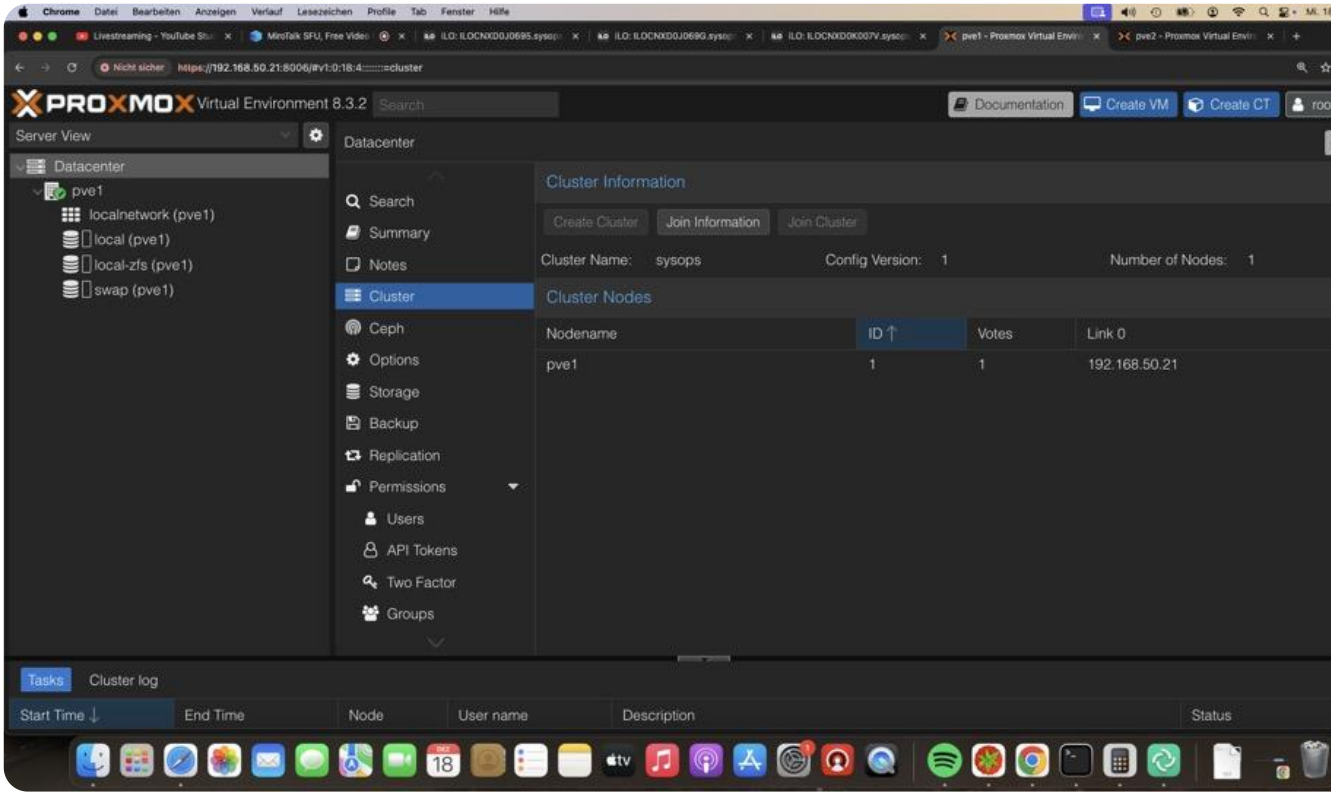


Proxmox Cluster mit drei Nodes erstellen

Ein Proxmox-Cluster kann sehr schnell erzeugt werden. Auch ein einzelner Node kann zunächst problemlos allein laufen. Für echte Cluster-Funktionen und Quorum sollten jedoch mindestens drei Nodes oder ein sinnvoller QDevice-Aufbau geplant werden.

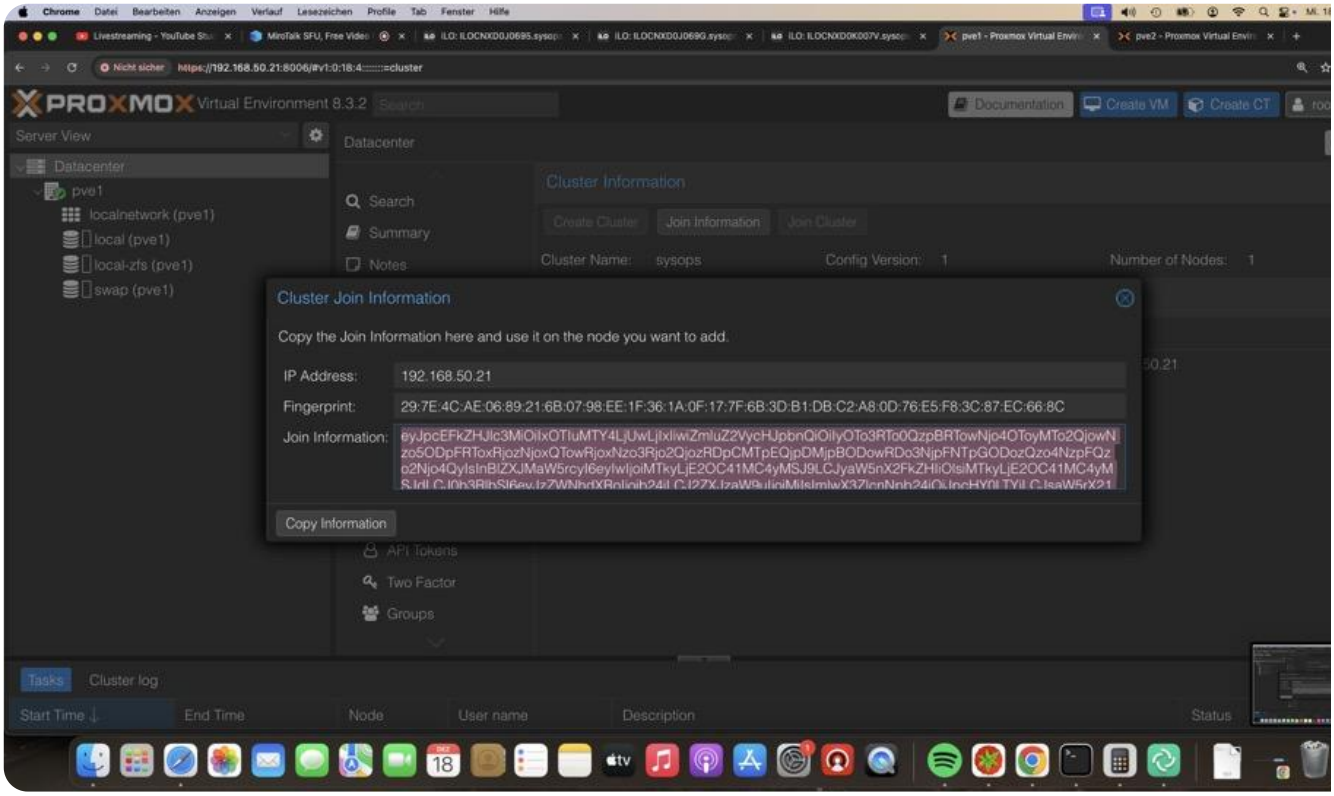

PROXMOX Virtual Environment 8.3.2
Documentation
Create VM
Create CT
Server View
Datacenter
pve1
localnetwork (pve1)
local (pve1)
local-zfs (pve1)
swap (pve1)
Search
Summary
Notes
Cluster Information
Create Cluster
Join Information
Join Cluster
Standalone node - no cluster defined
Create Cluster
Cluster Name:
sysops
Cluster Network:
Link: 0
192.168.50.21
Add
Multiple links are used as failover, lower numbers have higher priority.
Help
Create
Users
Tasks
Cluster log
Start Time
End Time
Node
User name
Description
Status

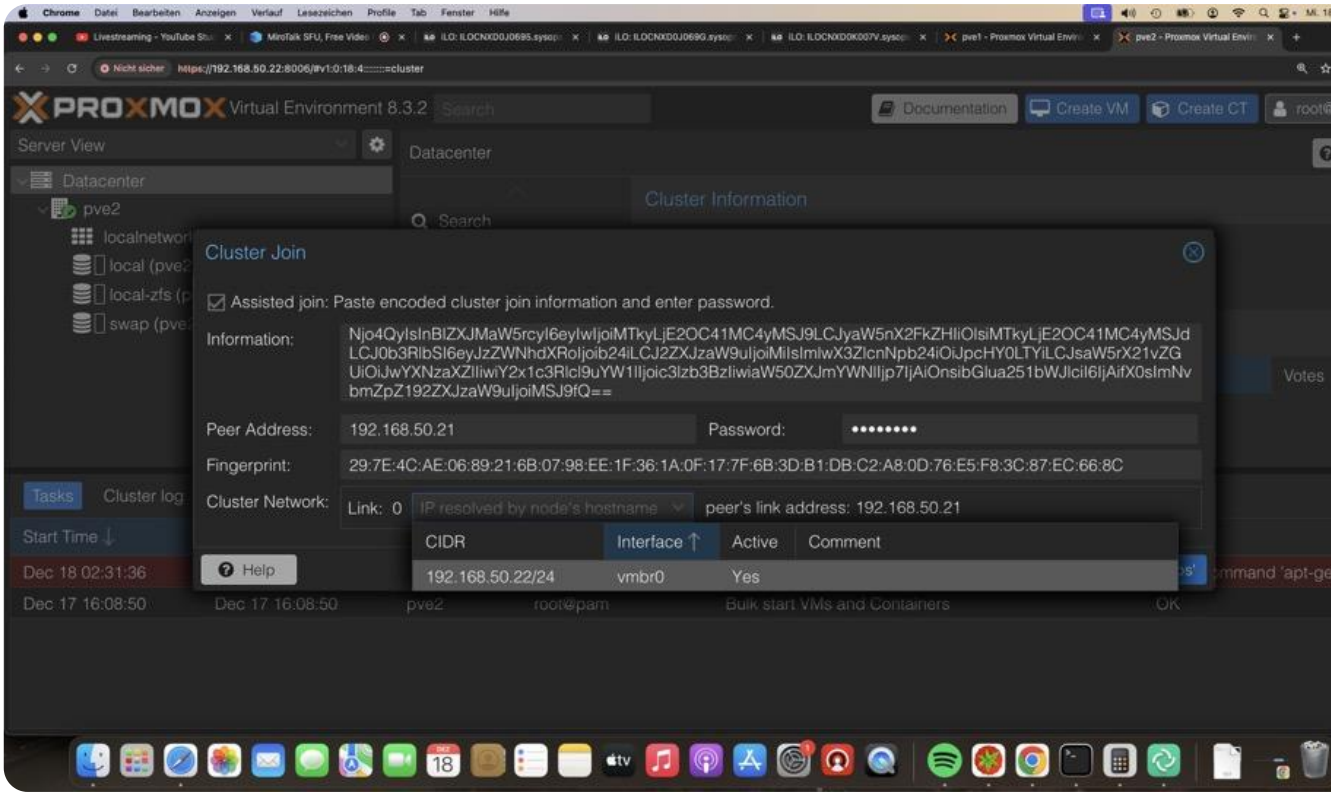

Output
Status
Stop
Download
Corosync Cluster Engine Authentication key generator.
Gathering 2048 bits for key from /dev/urandom.
Writing corosync key to /etc/corosync/authkey.
Writing corosync config to /etc/pve/corosync.conf
Restart corosync and cluster filesystem
TASK OK
Server View
Datacenter
pve1
Tasks
Cluster log
Start Time
End Time
Node
User name
Description
Status

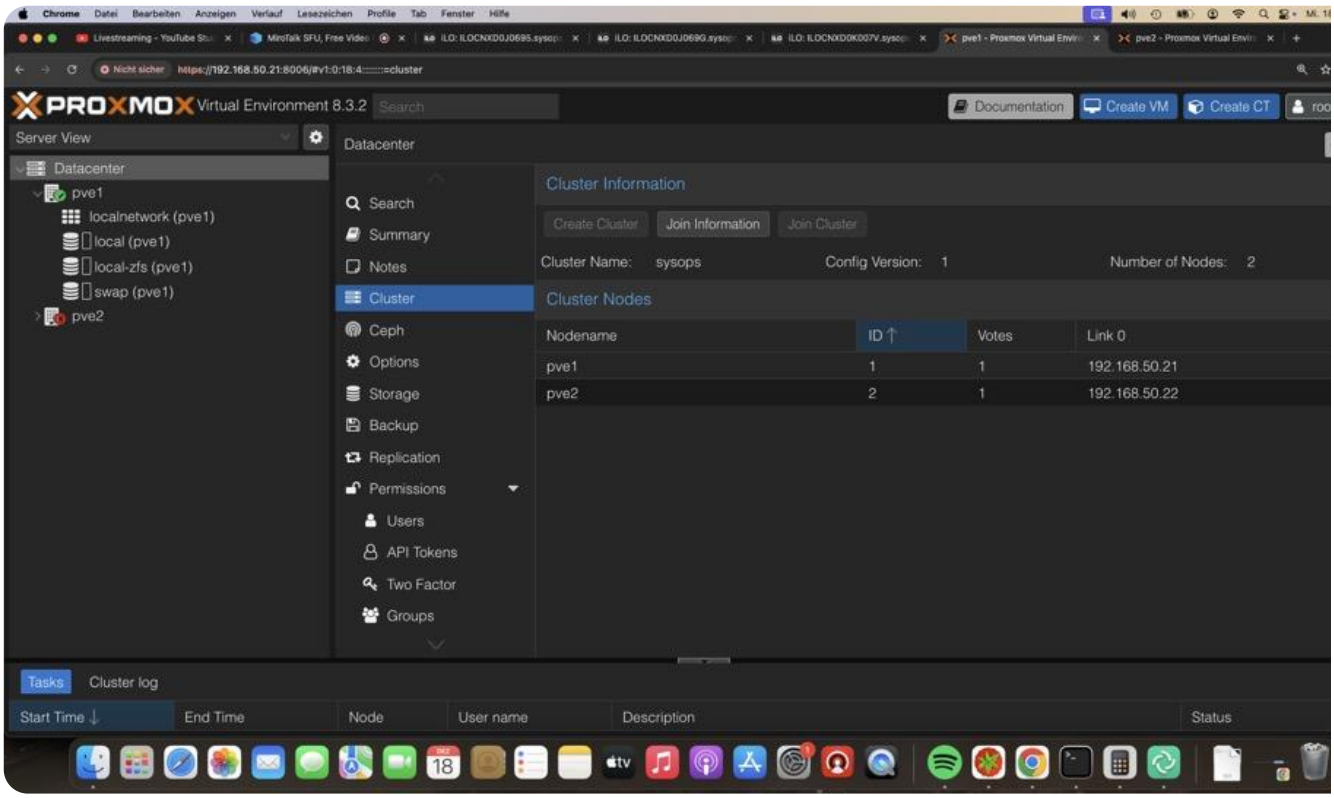


Weitere Proxmox Nodes hinzufügen

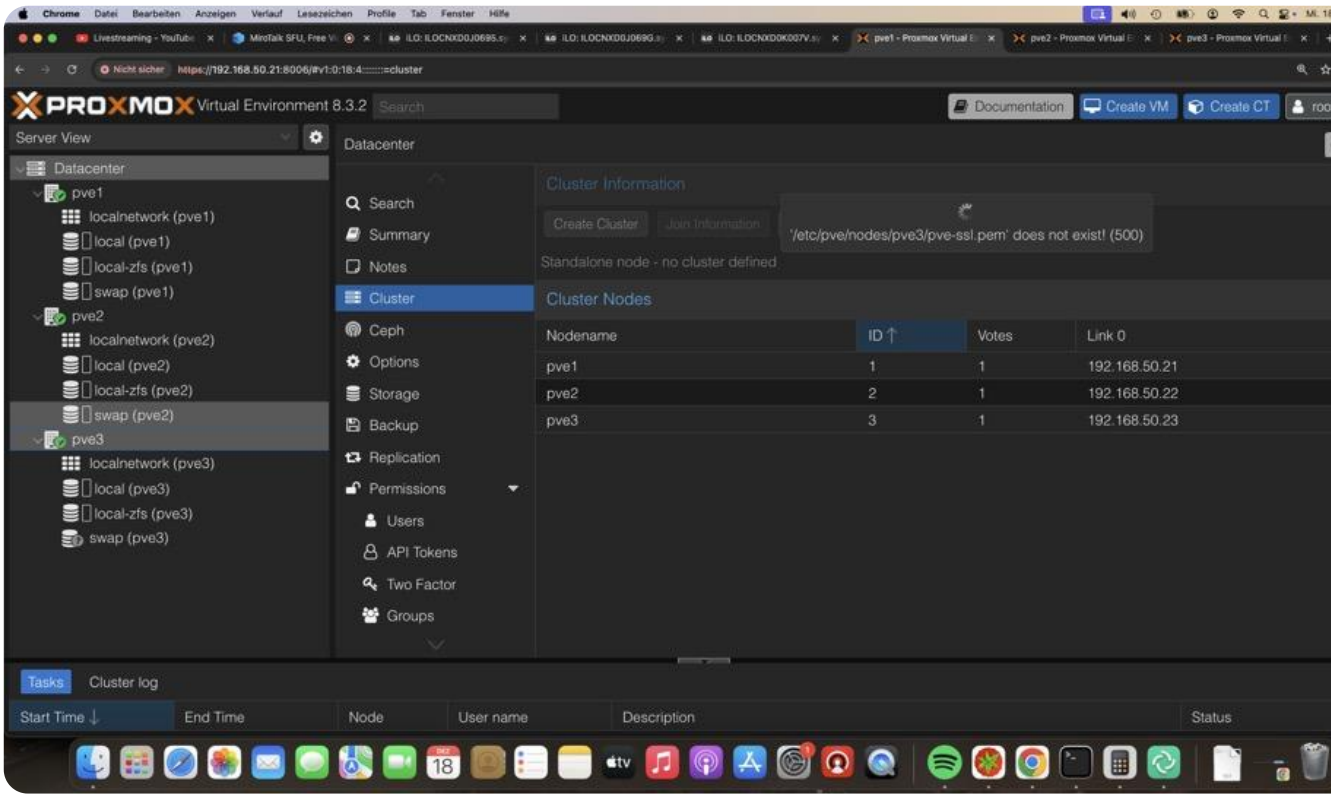
Weitere Nodes werden über die Join-Informationen des ersten Cluster-Nodes hinzugefügt. Auf dem neuen Node werden die Informationen unter **Join Cluster** eingefügt. Zusätzlich werden Root-Passwort und Cluster-Netzwerk angegeben.


PROXMOX Virtual Environment 8.3.2
Server View
Datacenter
pve1
localnetwork (pve1)
local (pve1)
local-zfs (pve1)
swap (pve1)
Cluster Information
Cluster Join Information
Copy the Join Information here and use it on the node you want to add.
IP Address: 192.168.50.21
Fingerprint: 29:7E:4C:AE:06:89:21:6B:07:98:EE:1F:36:1A:0F:17:7F:6B:3D:B1:DB:C2:A8:0D:76:E5:F8:3C:87:EC:66:8C
Join Information:
Copy Information
Tasks
Cluster log


PROXMOX Virtual Environment 8.3.2
Server View
Datacenter
pve2
Cluster Information
Cluster Join
Assisted join: Paste encoded cluster join information and enter password.
Information:
Peer Address: 192.168.50.21
Password:
Fingerprint: 29:7E:4C:AE:06:89:21:6B:07:98:EE:1F:36:1A:0F:17:7F:6B:3D:B1:DB:C2:A8:0D:76:E5:F8:3C:87:EC:66:8C
Cluster Network: Link: 0
peer's link address: 192.168.50.21
CIDR
Interface
Active
Comment
192.168.50.22/24
vmbr0
Yes
Help
Tasks
Cluster log



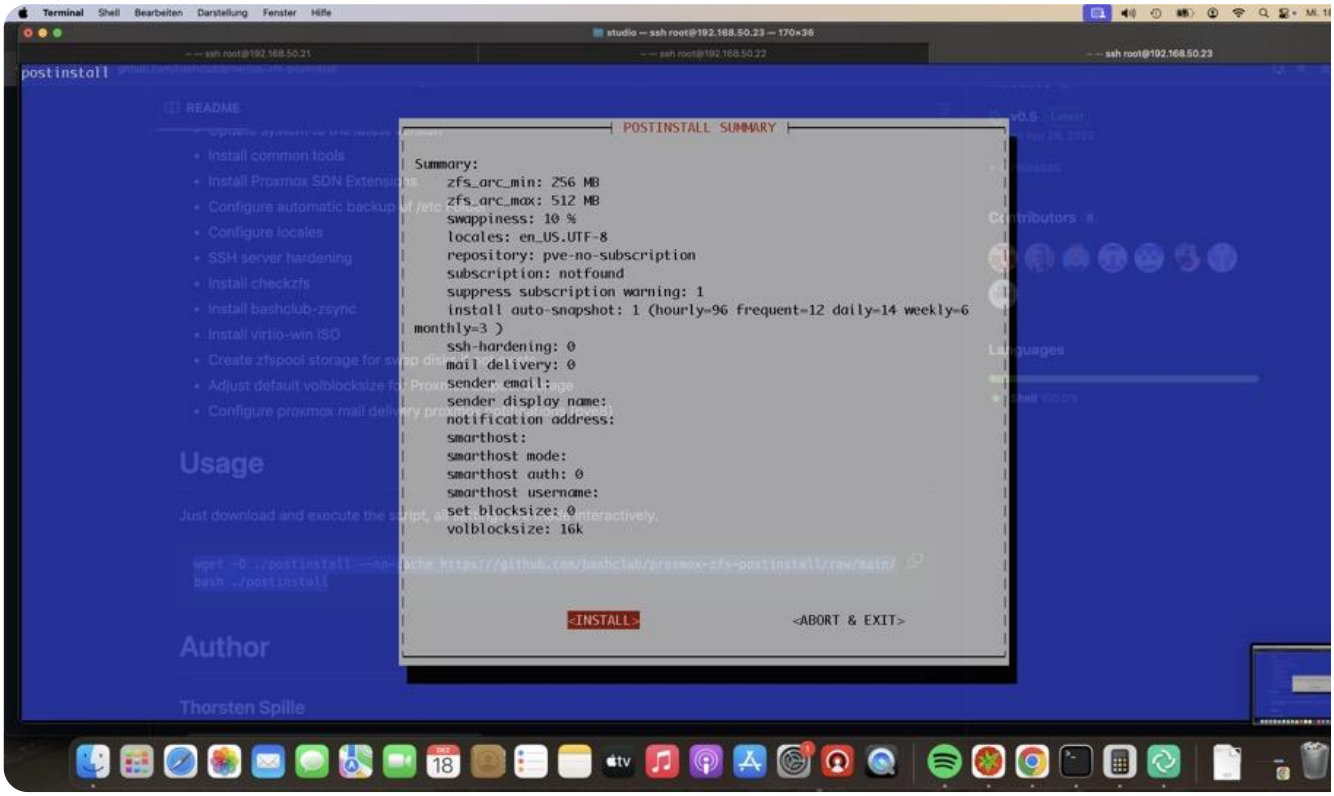
> Während des Cluster-Beitritts kann der neue Node kurzzeitig die Verbindung zum Browser verlieren, da Schlüssel und Zertifikate des Clusters übernommen werden. Danach sollte die Weboberfläche neu geladen werden.



Postinstaller für Proxmox VE

Unser **Proxmox ZFS Postinstaller vom Bashclub** hilft bei vielen wiederkehrenden Aufgaben nach der Installation.

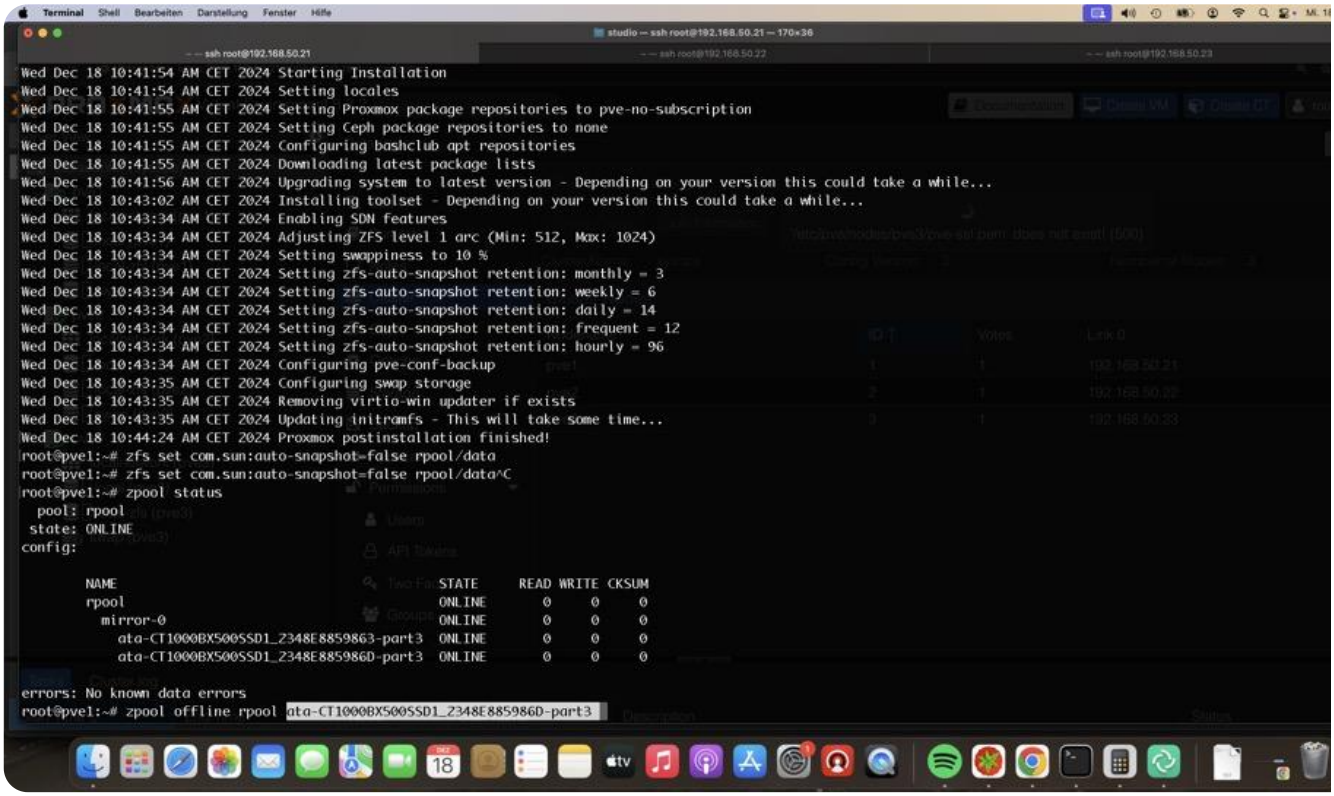
Für Cluster-Setups sollten die Optionen bewusst gewählt werden. Im Workshop verzichten wir beispielsweise auf Autosnapshots für `rpool/data` und auf SSH-Hardening, soweit dies aus Kompatibilitätsgründen sinnvoll ist.

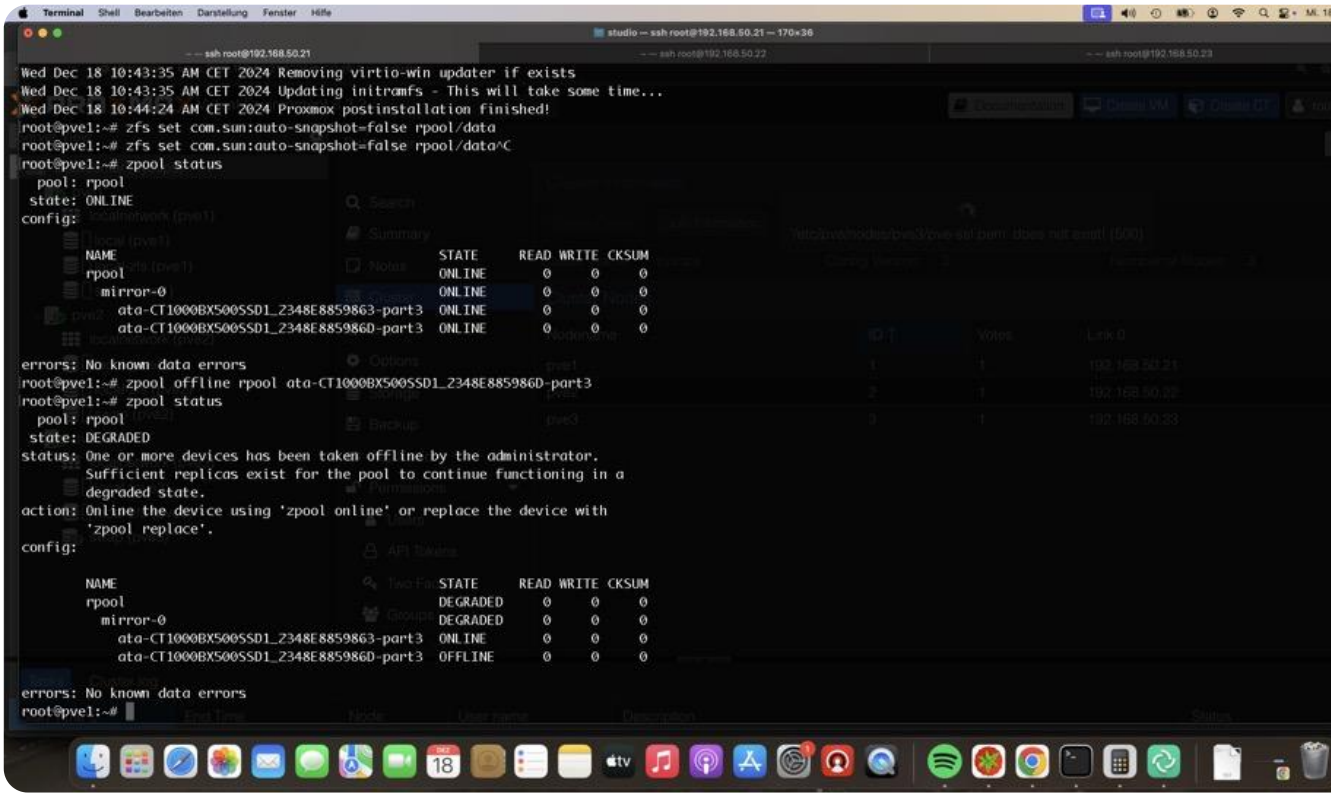


ZFS RAID demonstrieren und überwachen

ZFS-Software-RAID lässt sich im Fehlerfall sehr schnell analysieren und wiederherstellen. Der direkte Zugriff auf die Datenträger ermöglicht ein gutes Monitoring und kann Probleme frühzeitig sichtbar machen.

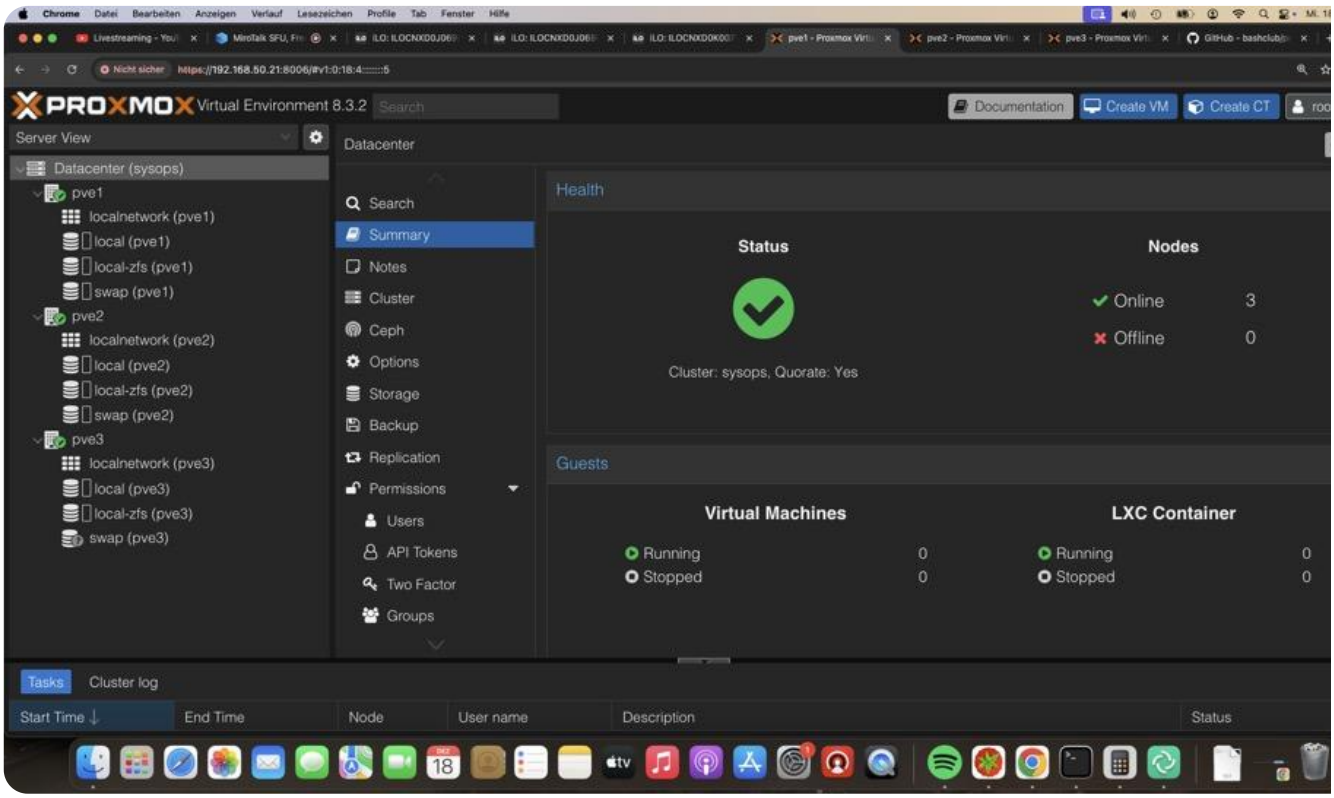
Im Workshop wird beispielhaft eine Platte vor einem Reboot offline gesetzt. Das ist eine Demonstration und sollte auf produktiven Systemen nicht unüberlegt nachgestellt werden.

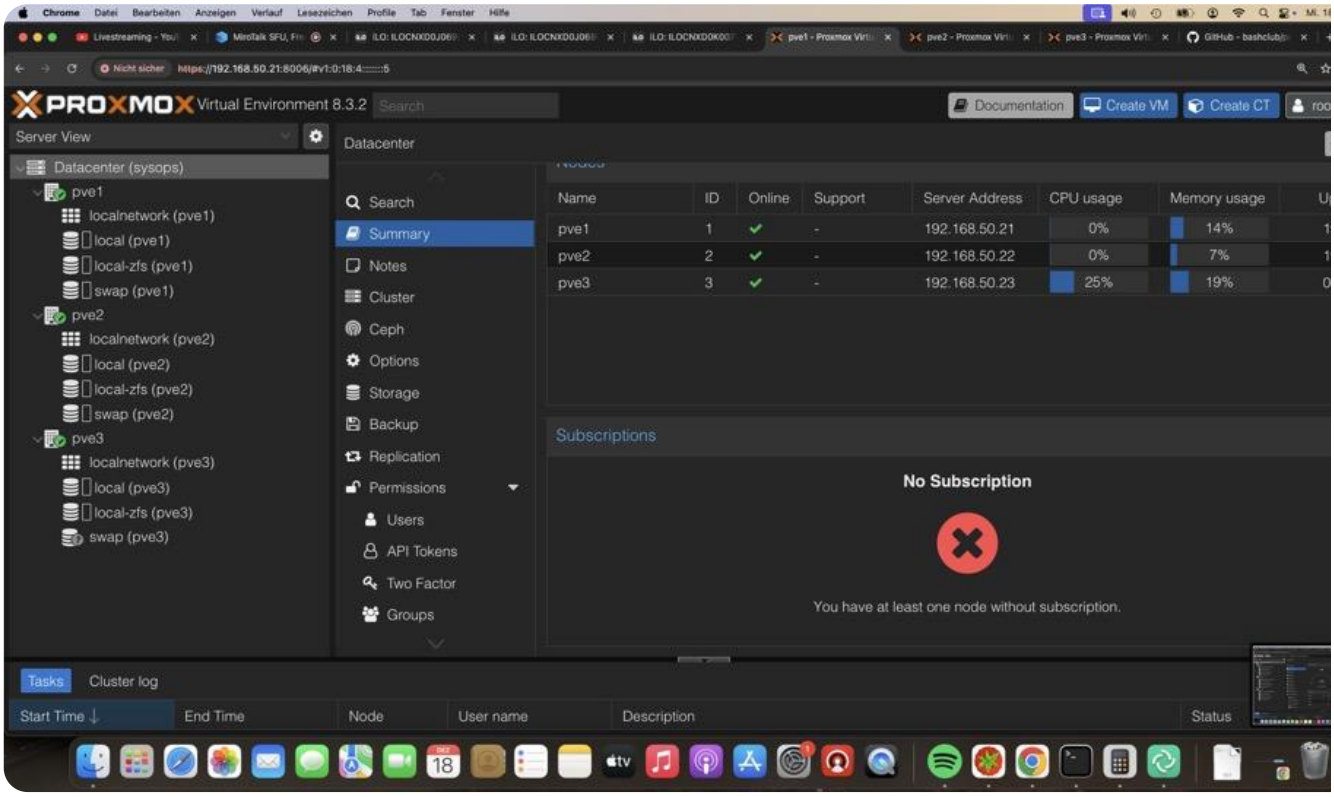




> Das Proxmox-Dashboard zeigt nicht alle Storage-Probleme so deutlich, wie es für den produktiven Betrieb wünschenswert wäre. Für professionelle

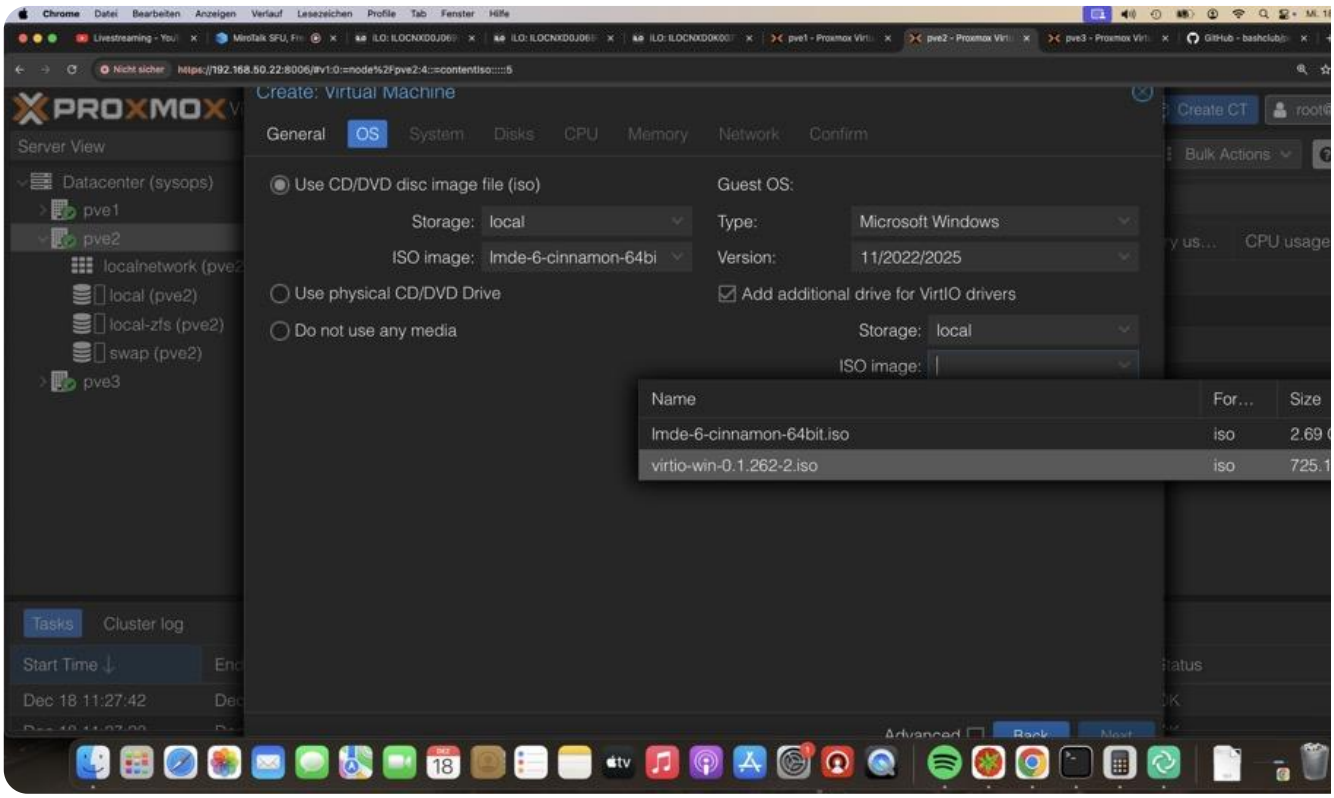
> Umgebungen empfehlen wir zusätzliches Monitoring, zum Beispiel mit Checkmk oder CV4PVE.

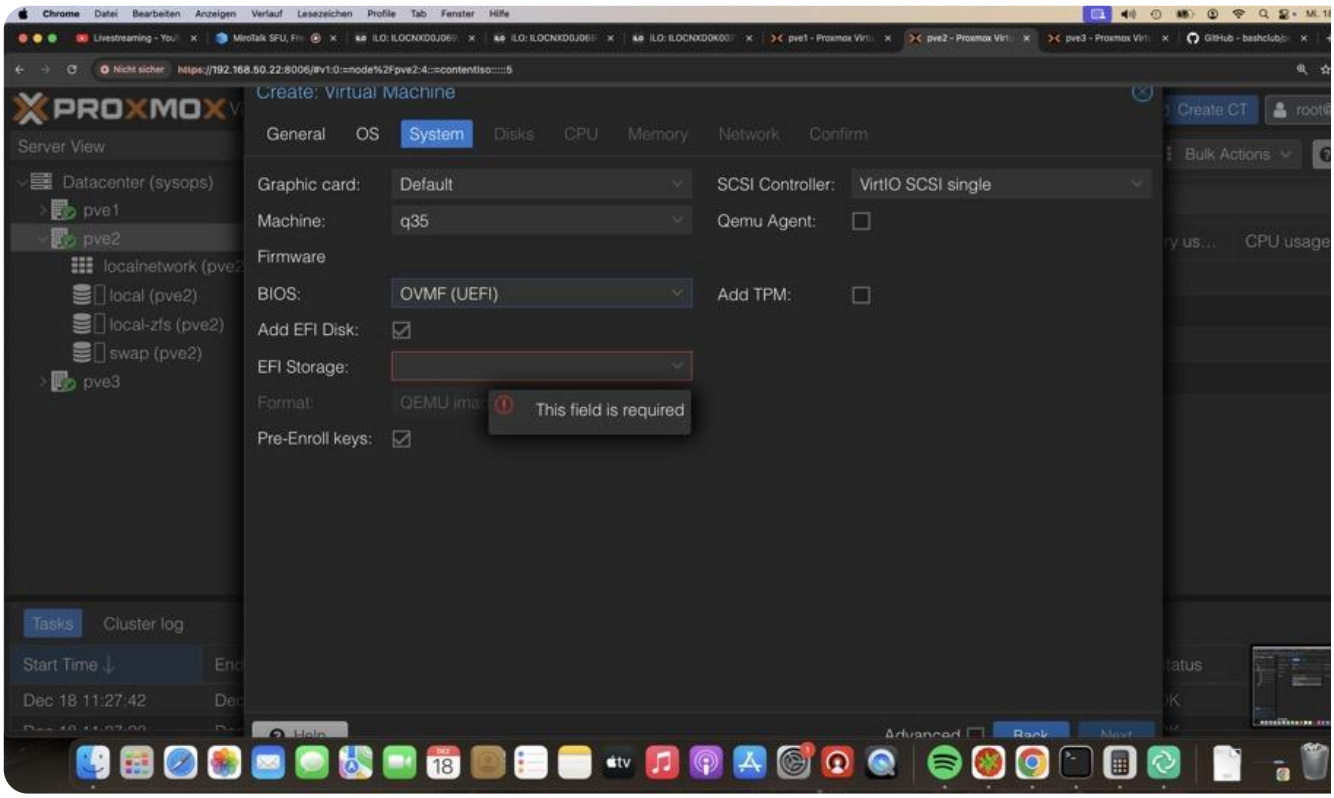




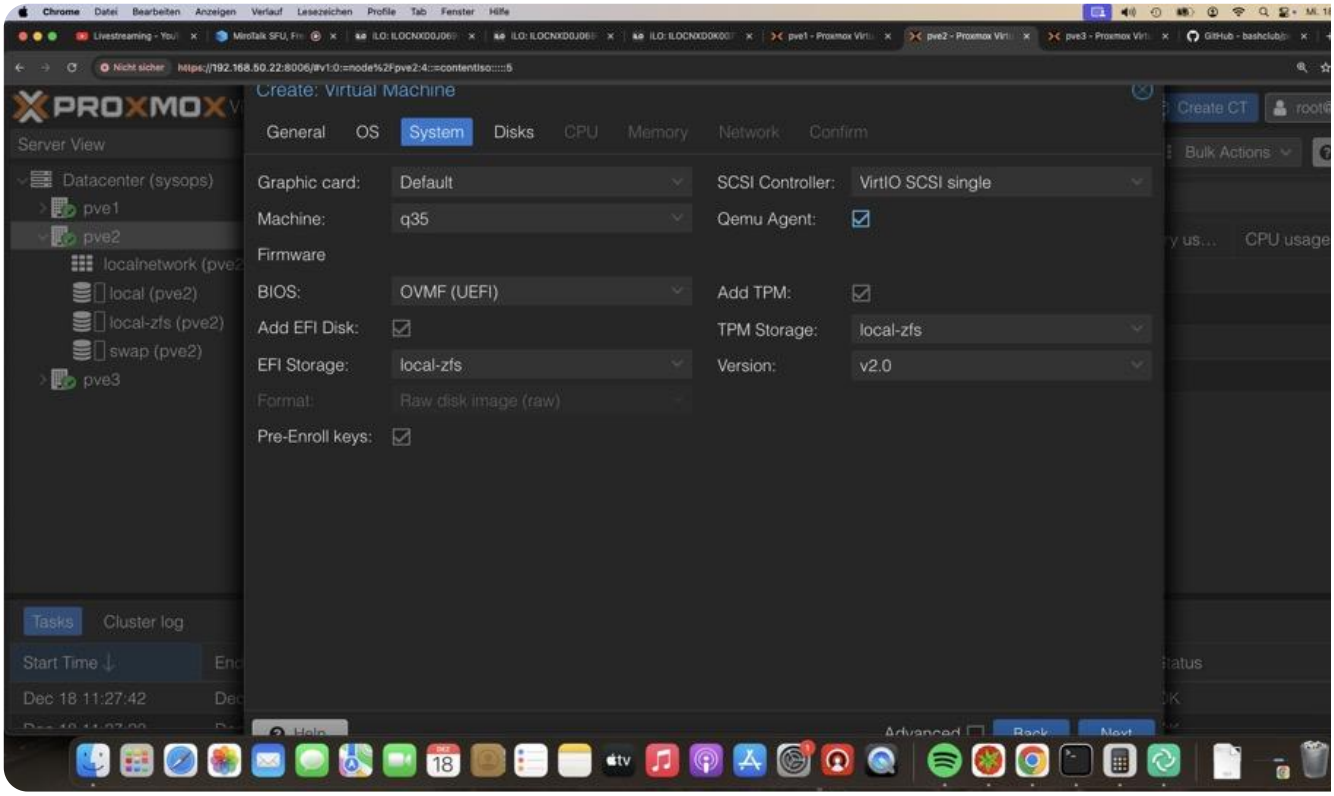
VMs und LXCs unter Proxmox richtig anlegen

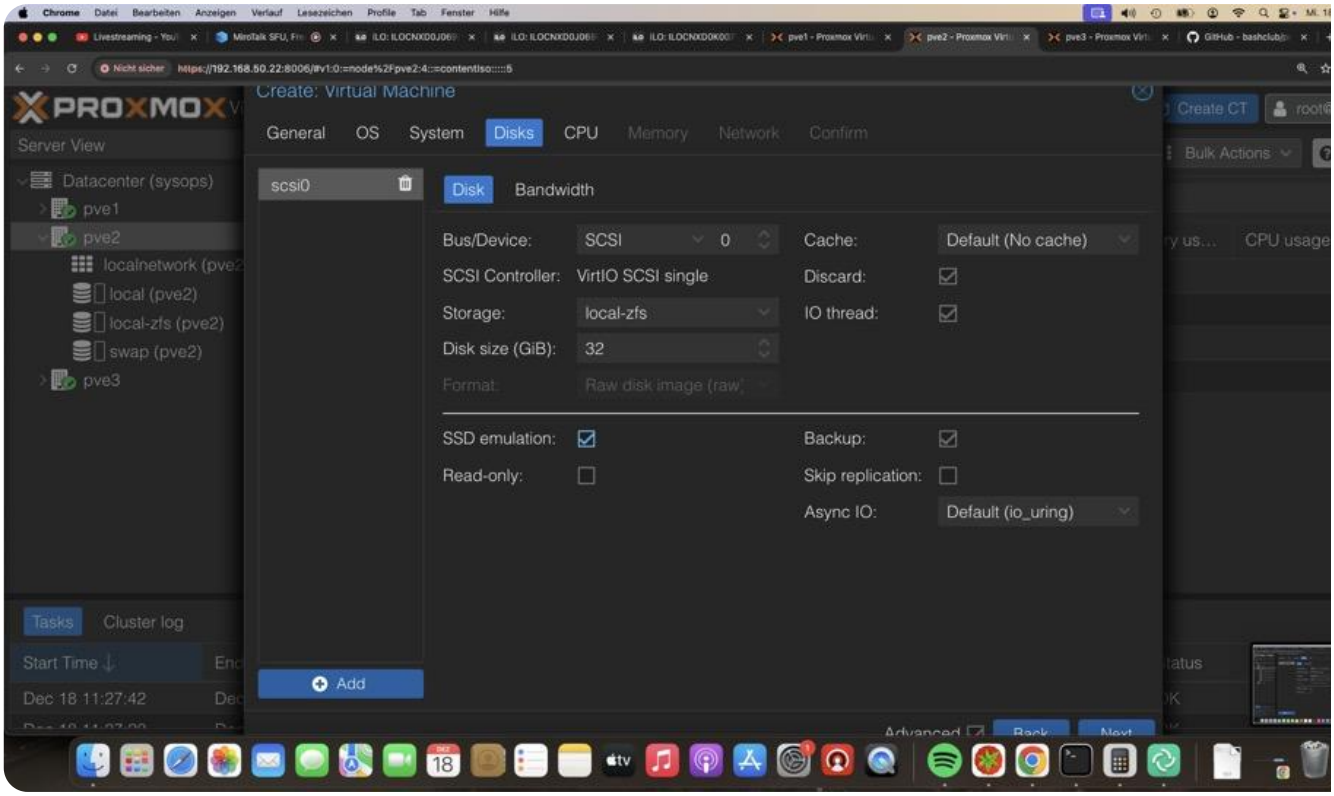
Für moderne virtuelle Maschinen sollten heute in der Regel **q35** und **OVMF/UEFI** verwendet werden. Ältere Einstellungen wie i440FX und klassisches BIOS sind für neue Systeme meist nicht mehr empfehlenswert.

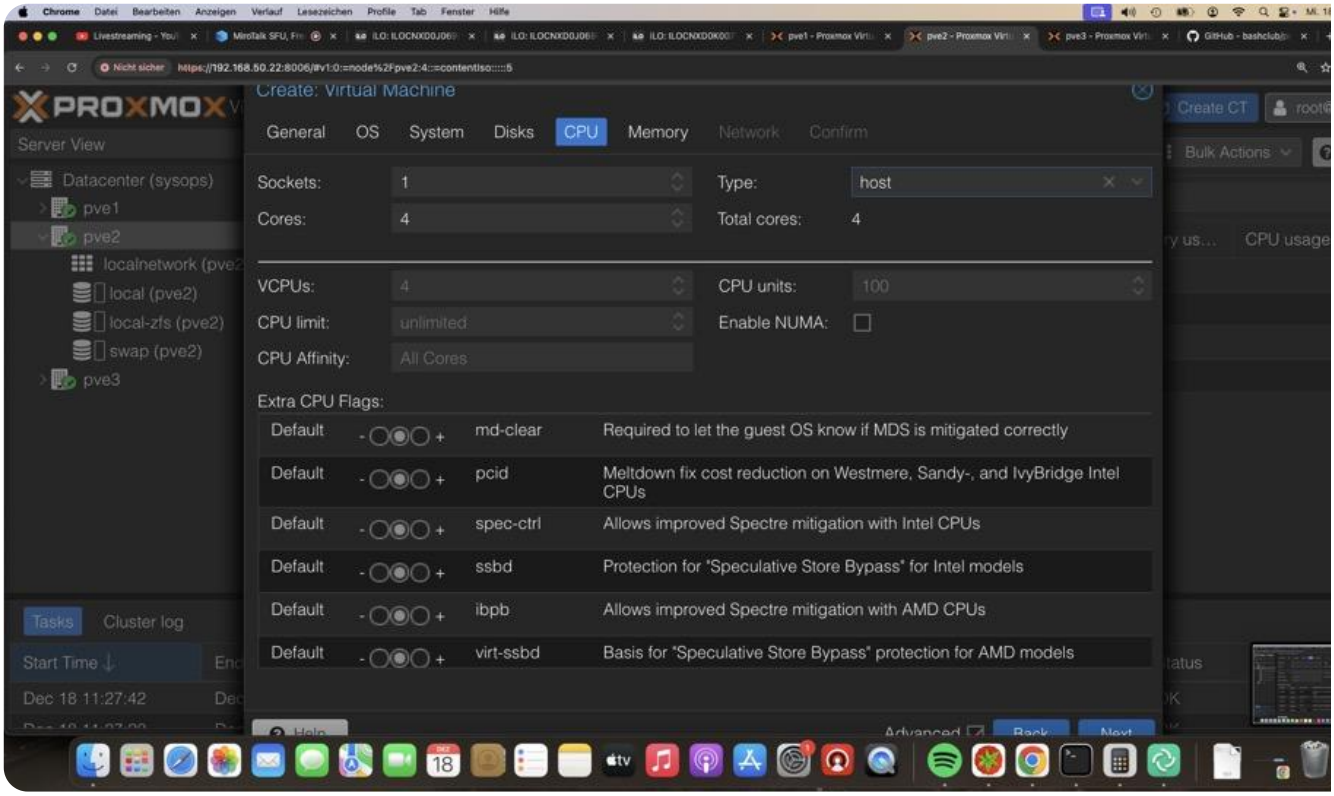


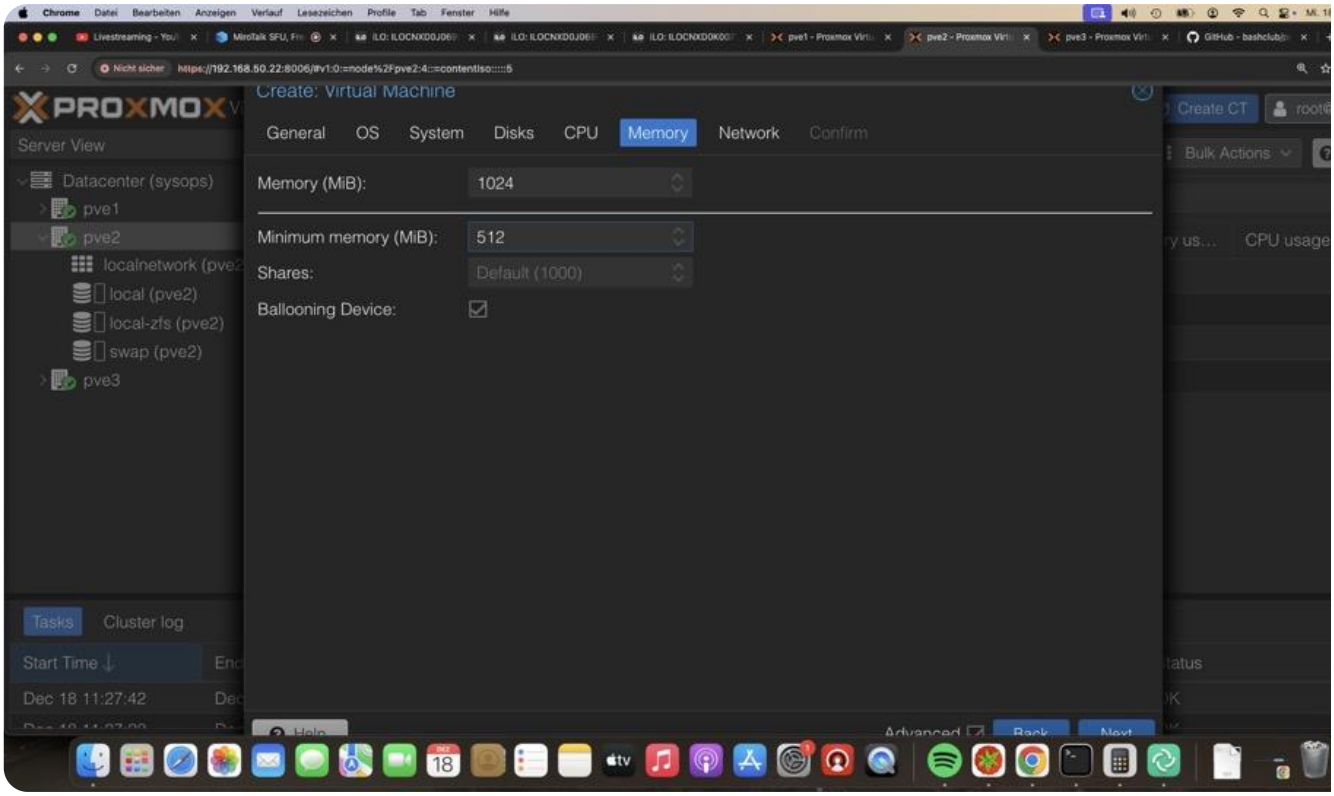


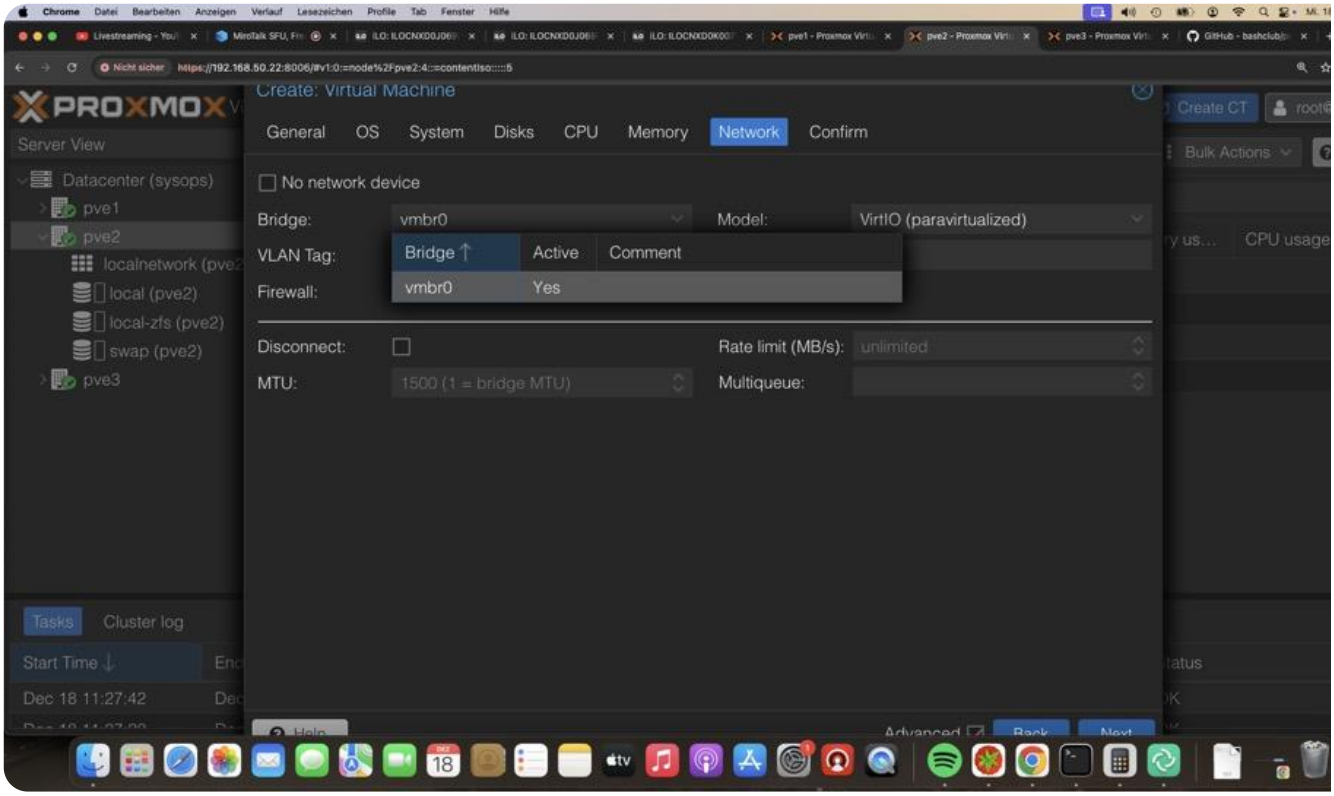
- **VirtIO SCSI Single:** gute Performance und eigener Prozess pro Disk
- **QEMU Guest Agent:** wichtig für Shutdown, Freeze, IP-Erkennung und Backups
- **Discard und SSD Emulation:** bessere Rückgabe gelöschter Blöcke an den Host
- **IO Thread:** bessere Entkopplung von I/O-Last
- **CPU Typ Host:** ideal bei identischer Hardware, vorsichtig bei gemischten CPUs im Cluster
- **Ballooning:** ermöglicht dynamischere RAM-Nutzung, benötigt aber passende Treiber


Create: Virtual Machine
General OS System Disks CPU Memory Network Confirm
Graphic card: Default
Machine: q35
Firmware
BIOS: OVMF (UEFI)
Add EFI Disk:
EFI Storage: local-zfs
Format: Raw disk image (raw)
Pre-Enroll keys:
SCSI Controller: VirtIO SCSI single
Qemu Agent:
Add TPM:
TPM Storage: local-zfs
Version: v2.0
Server View
Datacenter (sysops)
pve1
pve2
localnetwork (pve2)
local (pve2)
local-zfs (pve2)
swap (pve2)
pve3
Tasks Cluster log
Start Time
Dec 18 11:27:42


Create: Virtual Machine
General OS System Disks CPU Memory Network Confirm
scsi0
Disk Bandwidth
Bus/Device: SCSI 0
SCSI Controller: VirtIO SCSI single
Storage: local-zfs
Disk size (GiB): 32
Format: Raw disk image (raw)
Cache: Default (No cache)
Discard:
IO thread:
SSD emulation:
Read-only:
Backup:
Skip replication:
Async IO: Default (io_uring)
Add
Server View
Datacenter (sysops)
pve1
pve2
localnetwork (pve2)
local (pve2)
local-zfs (pve2)
swap (pve2)
pve3
Tasks Cluster log
Start Time
Dec 18 11:27:42


Create: Virtual Machine
General OS System Disks CPU Memory Network Confirm
Sockets: 1
Cores: 4
Type: host
Total cores: 4
VCPUs: 4
CPU limit: unlimited
CPU Affinity: All Cores
CPU units: 100
Enable NUMA:
Extra CPU Flags:
Default md-clear Required to let the guest OS know if MDS is mitigated correctly
Default pcid Meltdown fix cost reduction on Westmere, Sandy-, and IvyBridge Intel CPUs
Default spec-ctrl Allows improved Spectre mitigation with Intel CPUs
Default ssbd Protection for "Speculative Store Bypass" for Intel models
Default ibpb Allows improved Spectre mitigation with AMD CPUs
Default virt-ssbd Basis for "Speculative Store Bypass" protection for AMD models

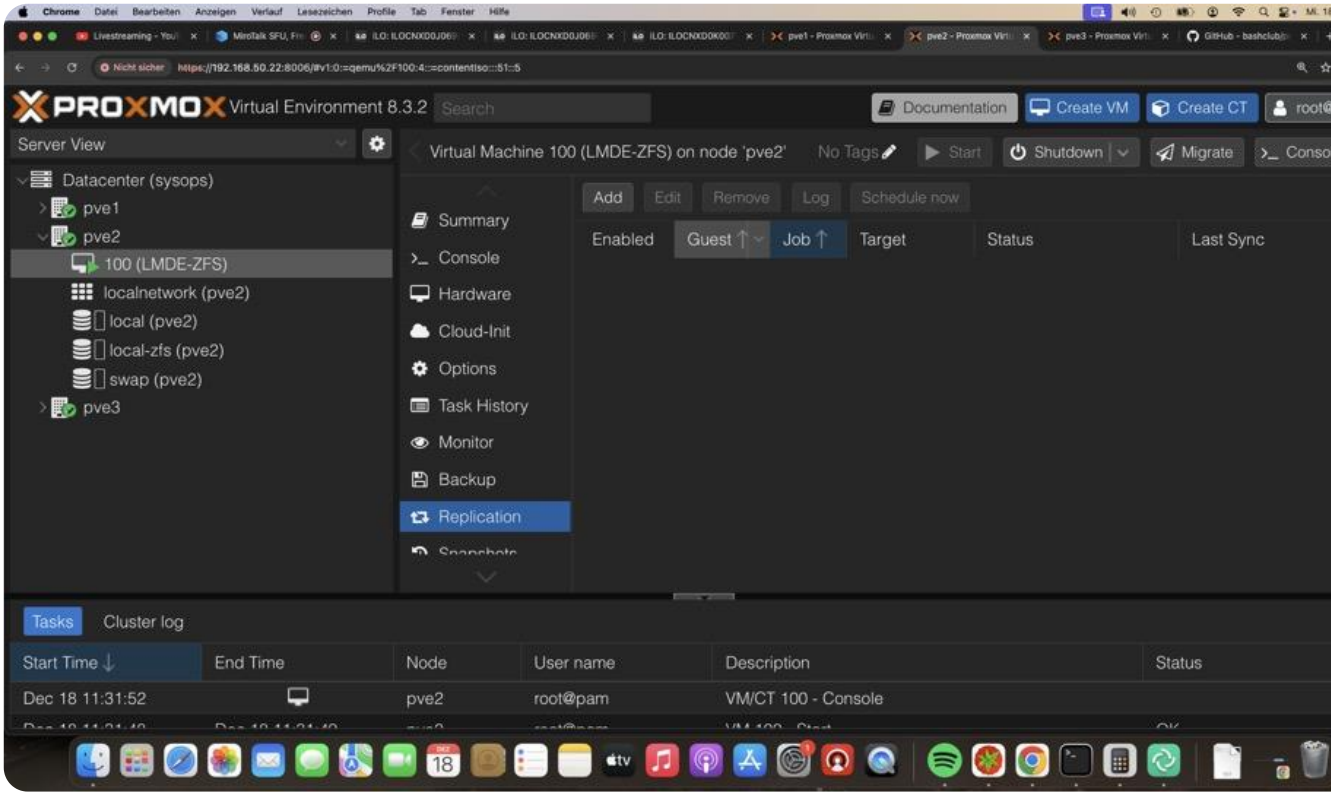

Create: Virtual Machine
General OS System Disks CPU Memory Network Confirm
Memory (MiB): 1024
Minimum memory (MiB): 512
Shares: Default (1000)
Ballooning Device:

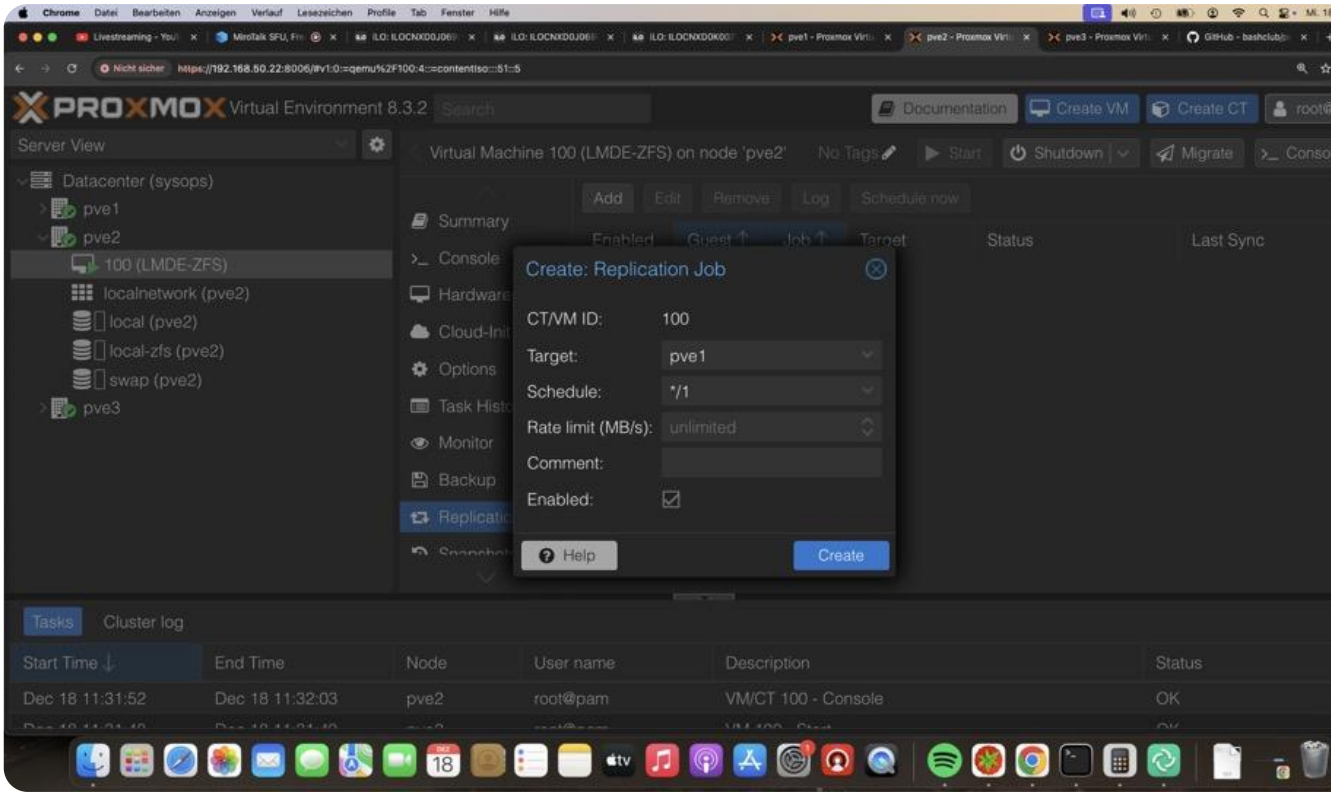


ZFS-Replikation in Proxmox VE

Die eingebaute ZFS-Replikation von Proxmox VE ist für viele kleine und mittlere Cluster eine sehr gute Grundlage. Bei vorhandener Replikation liegt der größte Teil der VM bereits auf dem Zielsystem und muss nur inkrementell ergänzt werden.

Ohne vorherige Replikation muss die komplette VM oder der komplette Container übertragen werden. Je nach Zustand erfolgt die Übertragung über KVM-Mechanismen oder ZFS. Dabei kann es Unterschiede beim Erhalt von Snapshots geben.





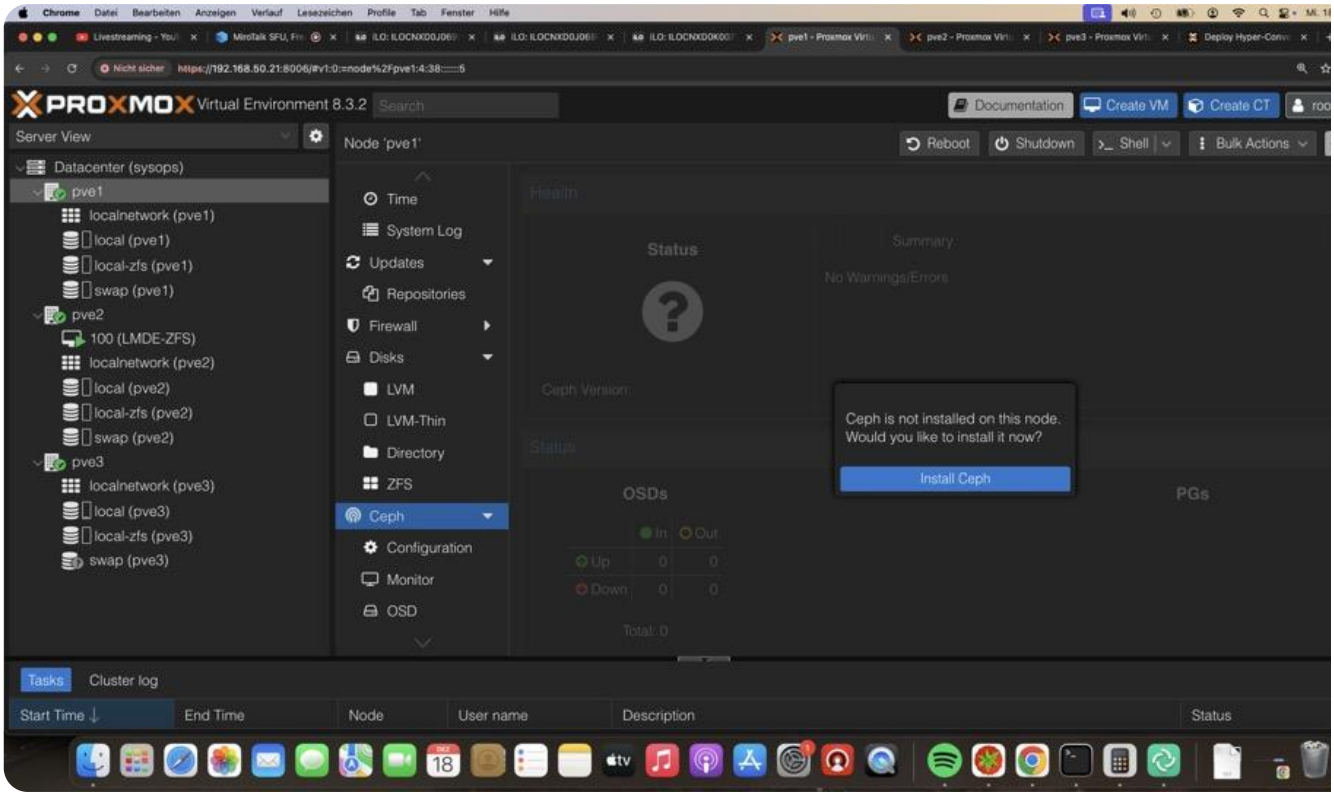
> Replikationsintervalle von 15 bis 60 Minuten sind für viele Umgebungen ein guter Kompromiss. Eine Minute ist technisch möglich, sollte aber nicht pauschal für jede VM verwendet werden.

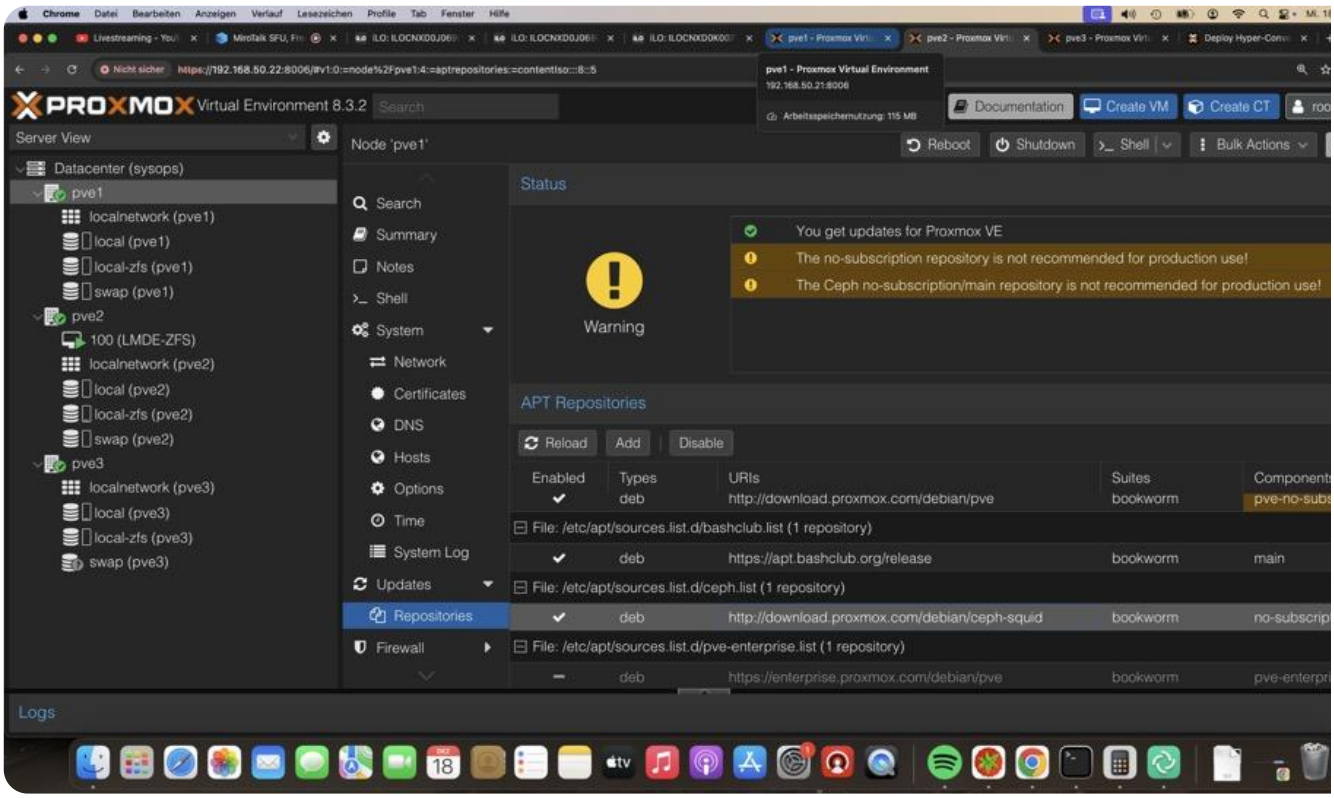
Ceph-Grundlagen im Proxmox Cluster

Ceph ermöglicht verteilten Storage innerhalb eines Proxmox-Clusters. Dadurch können VMs auf gemeinsamem Storage liegen und bei passender Architektur sehr flexibel zwischen Nodes betrieben werden.

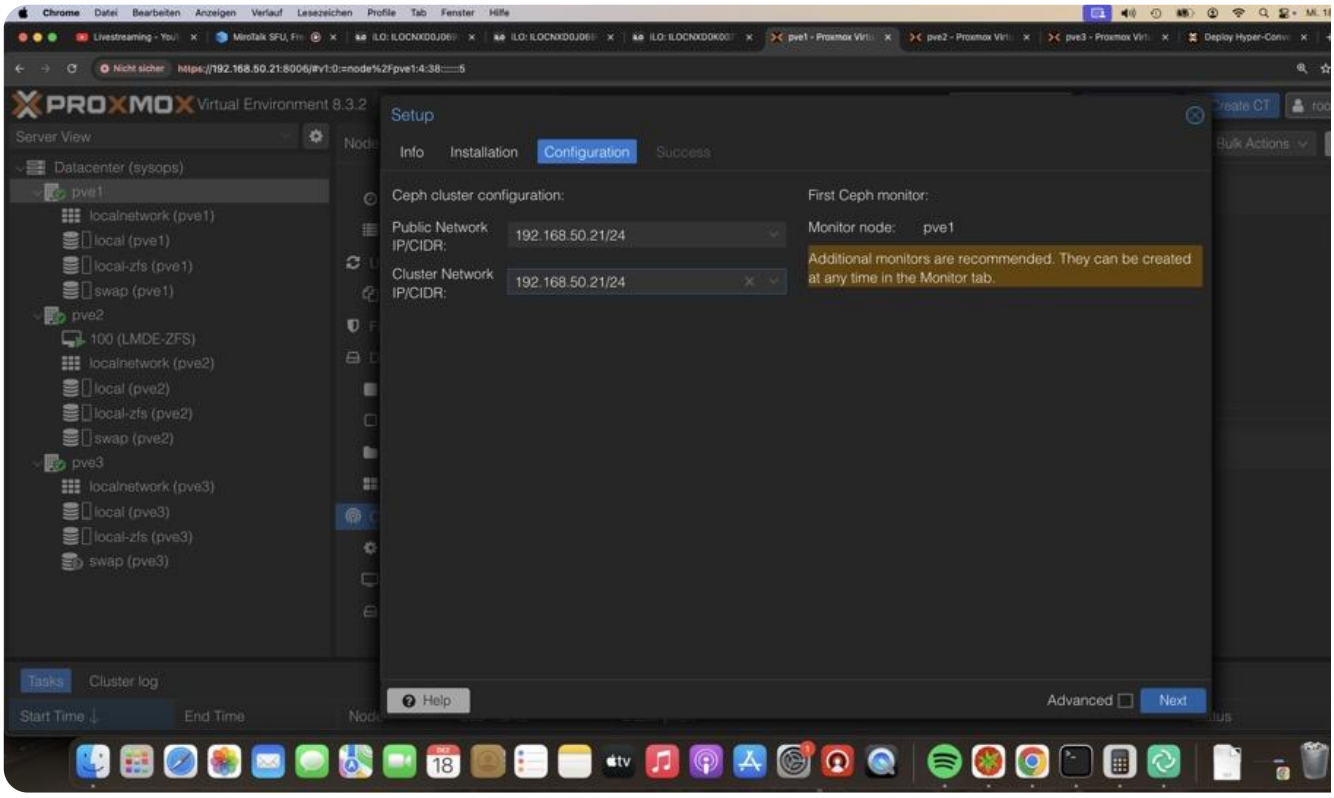
Ceph benötigt jedoch sorgfältige Planung, geeignete Netzwerke, passende Hardware und ausreichende Redundanz.

> Für produktive Ceph-Cluster empfehlen wir mindestens eine Proxmox-Subscription und erfahrene Planung. Ein fehlerhaft aufgebauter Ceph-Cluster kann selbst zum kritischen Ausfallrisiko werden.





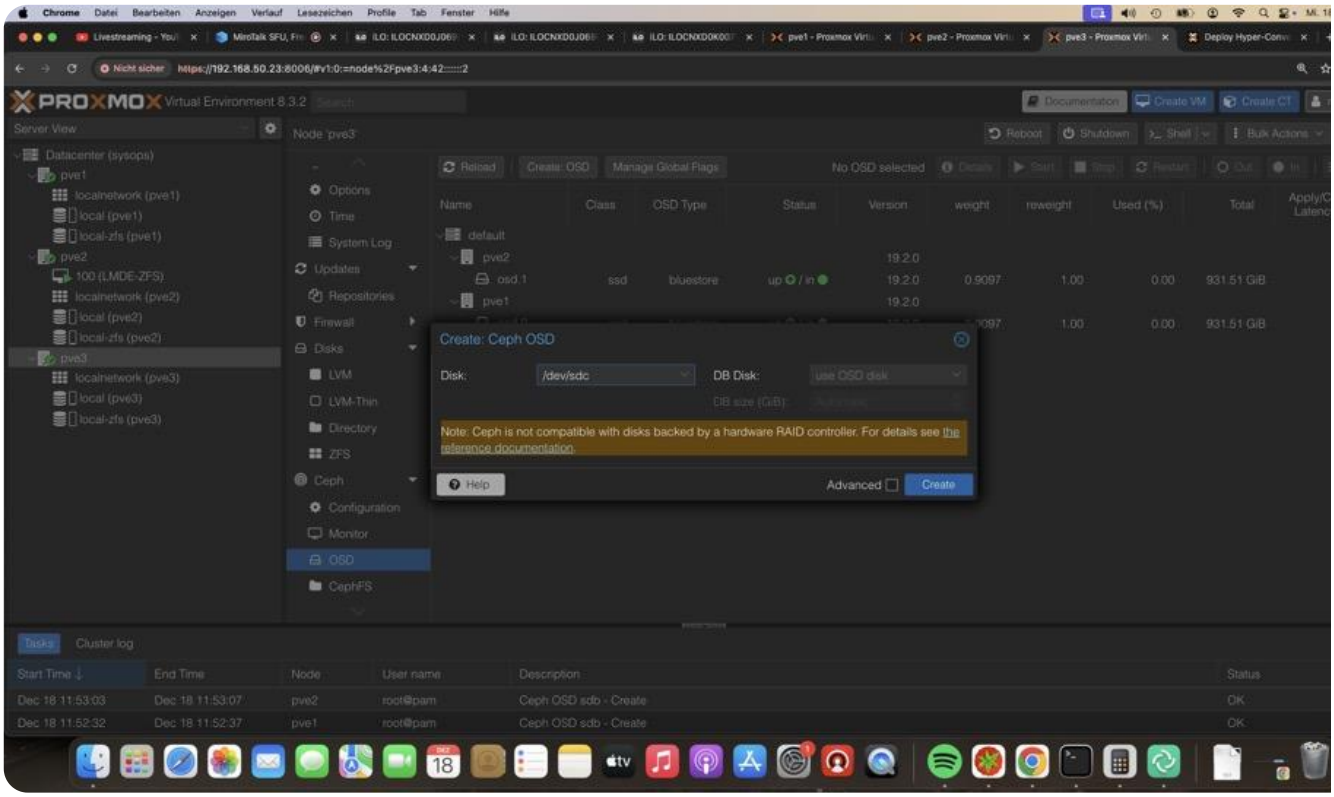
Die Auswahl der Netzwerke für Cluster, Ceph, Redundanz und weitere Segmentierungen sollte durch erfahrene Fachleute erfolgen. Für produktive Ceph-Umgebungen sollte das normale LAN nicht einfach mitverwendet werden.

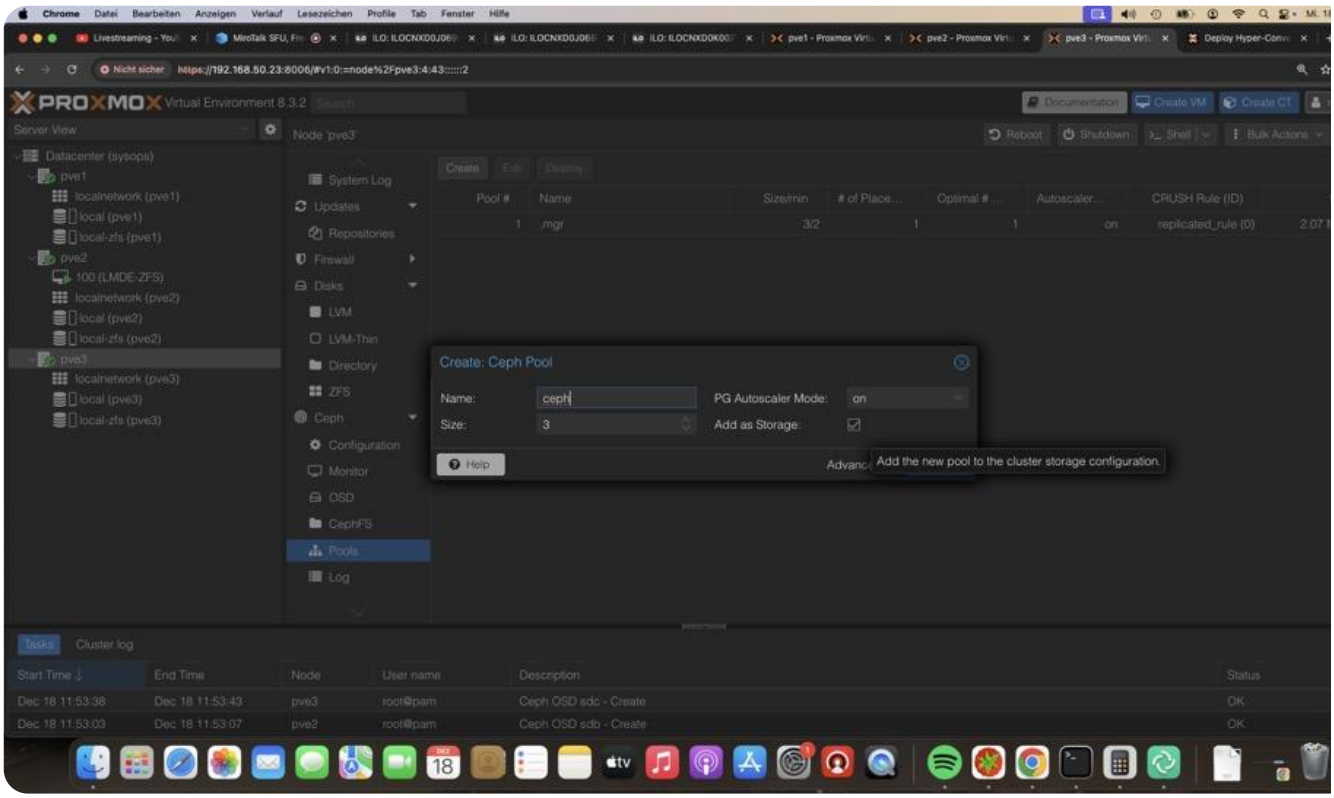


Ceph OSDs und Pools einrichten

Jeder Host benötigt geeignete Datenträger als OSDs. Für ein einfaches Testsetup reichen drei OSDs, produktiv sollte Ceph jedoch deutlich sorgfältiger und redundant geplant werden.

Hardware-RAID-Controller sollten auch für Ceph nicht verwendet werden. Ceph benötigt direkten Zugriff auf die Datenträger.



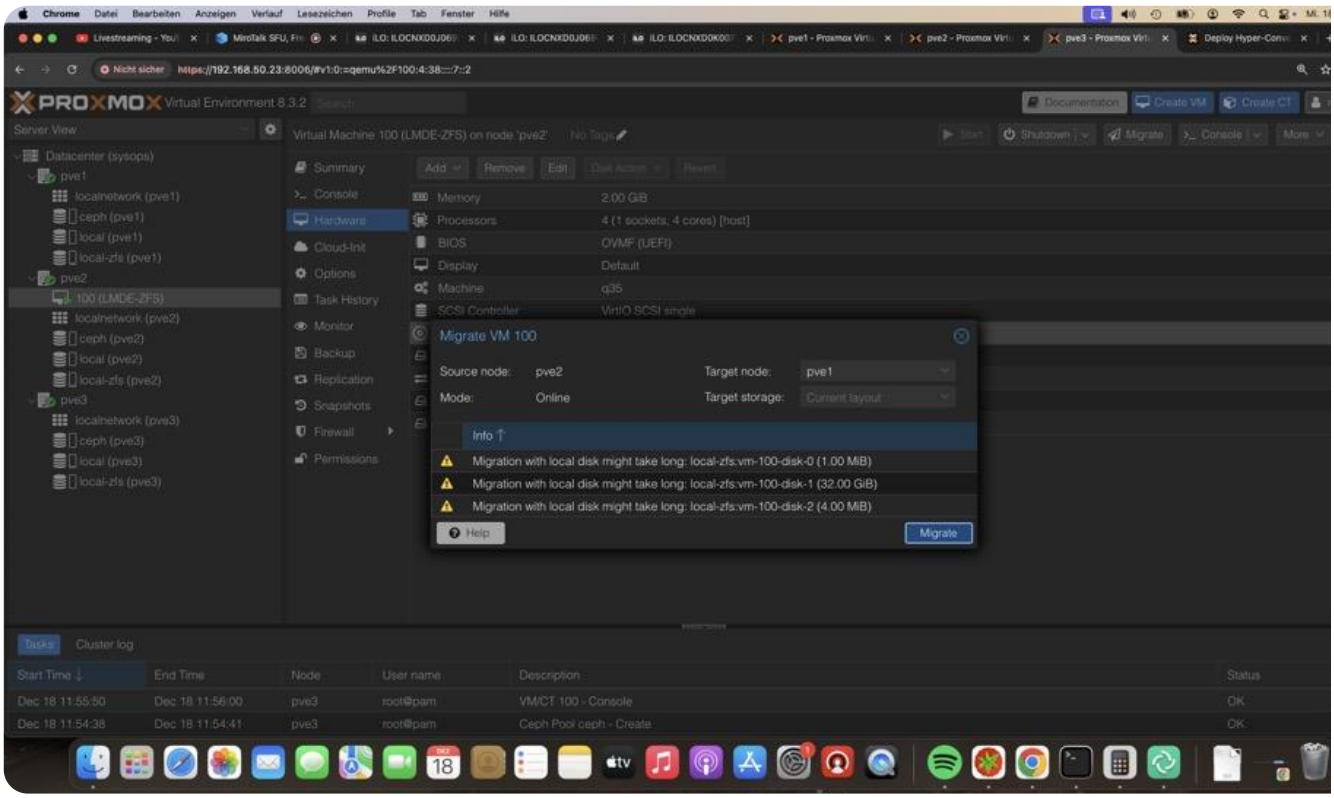


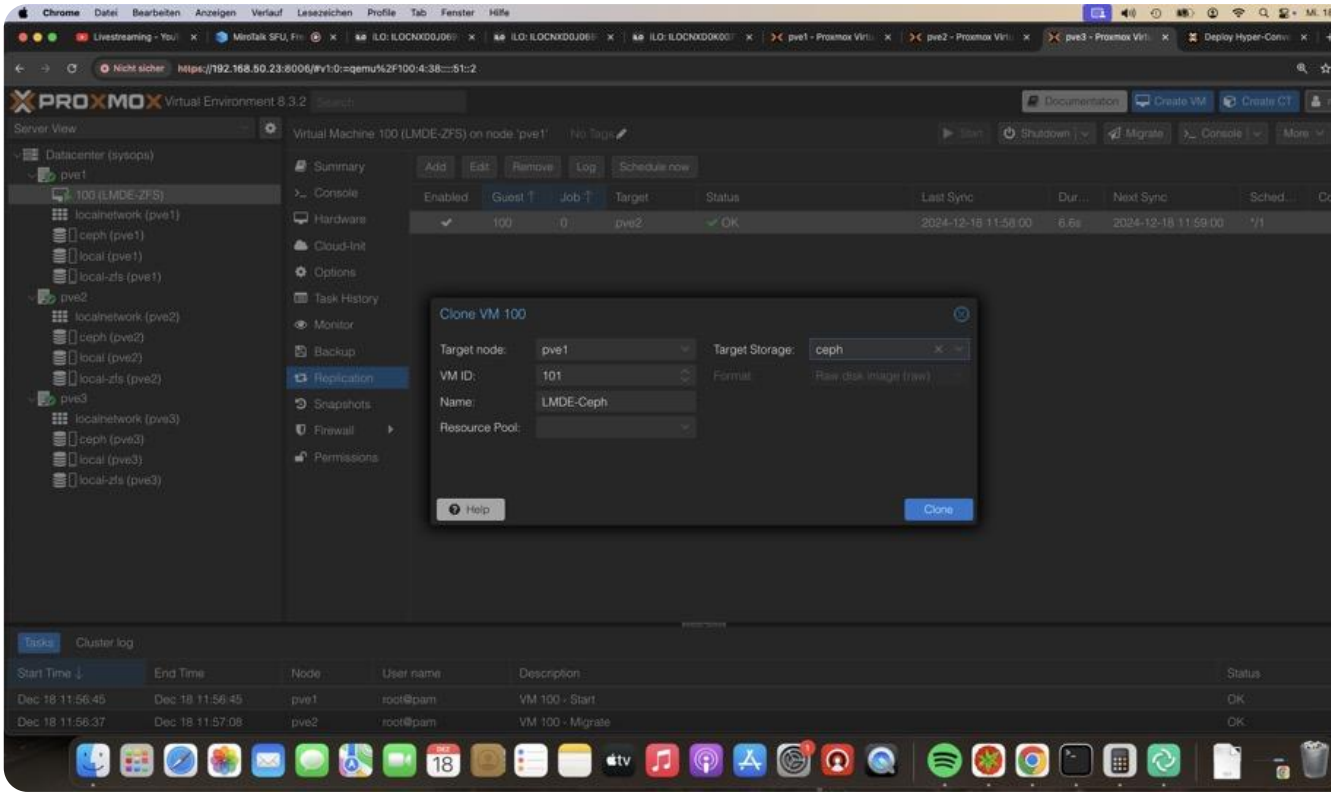
Monitor- und Manager-Rollen sollten auf mehrere Nodes verteilt werden. Andernfalls kann bereits ein Reboot eines einzelnen Nodes zu Problemen führen.

Migration zwischen Nodes

Die Migration laufender VMs oder gestoppter LXCs ist sowohl mit ZFS als auch mit Ceph möglich. Der wesentliche Unterschied liegt im Storage-Konzept.

- **Ceph:** Speicher liegt verteilt im Cluster
- **ZFS:** Speicher muss auf dem Zielhost vorhanden sein oder übertragen werden
- **ZFS mit Replikation:** nur die Änderungen müssen ergänzt werden
- **ZFS ohne Replikation:** vollständige Übertragung notwendig

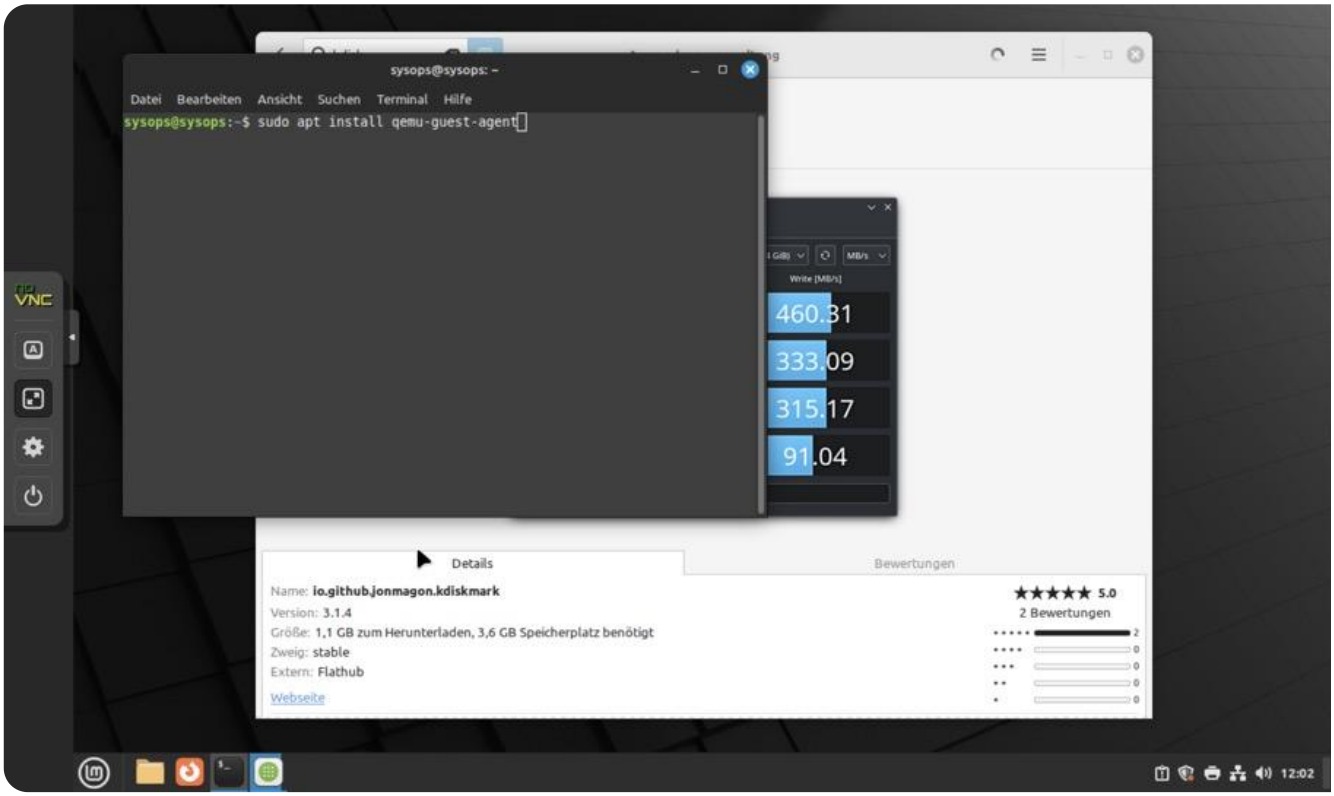




QEMU Guest Agent installieren

Der **QEMU Guest Agent** sollte in virtuellen Maschinen immer installiert und aktiviert werden. Er ermöglicht geordnetes Herunterfahren, bessere Backup-Konsistenz, IP-Erkennung

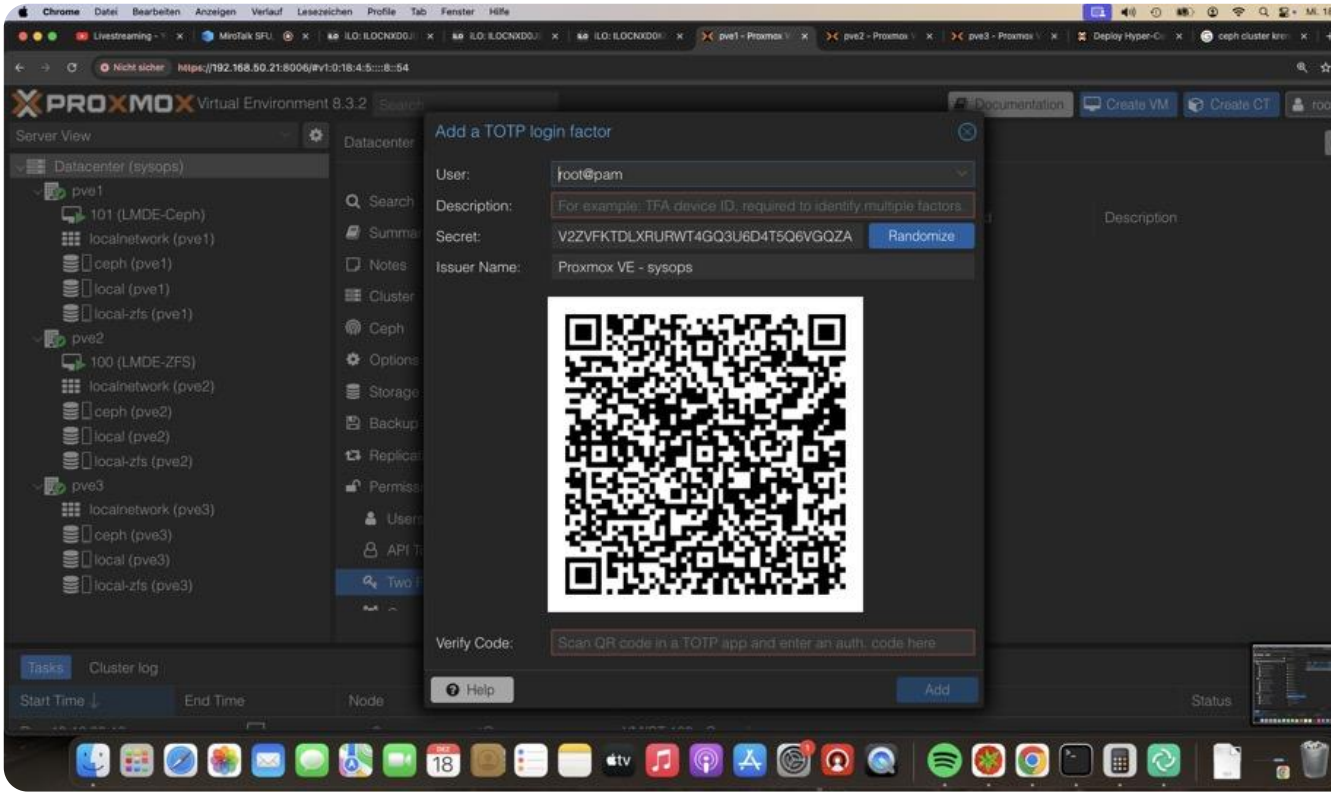
und weitere Verwaltungsfunktionen.

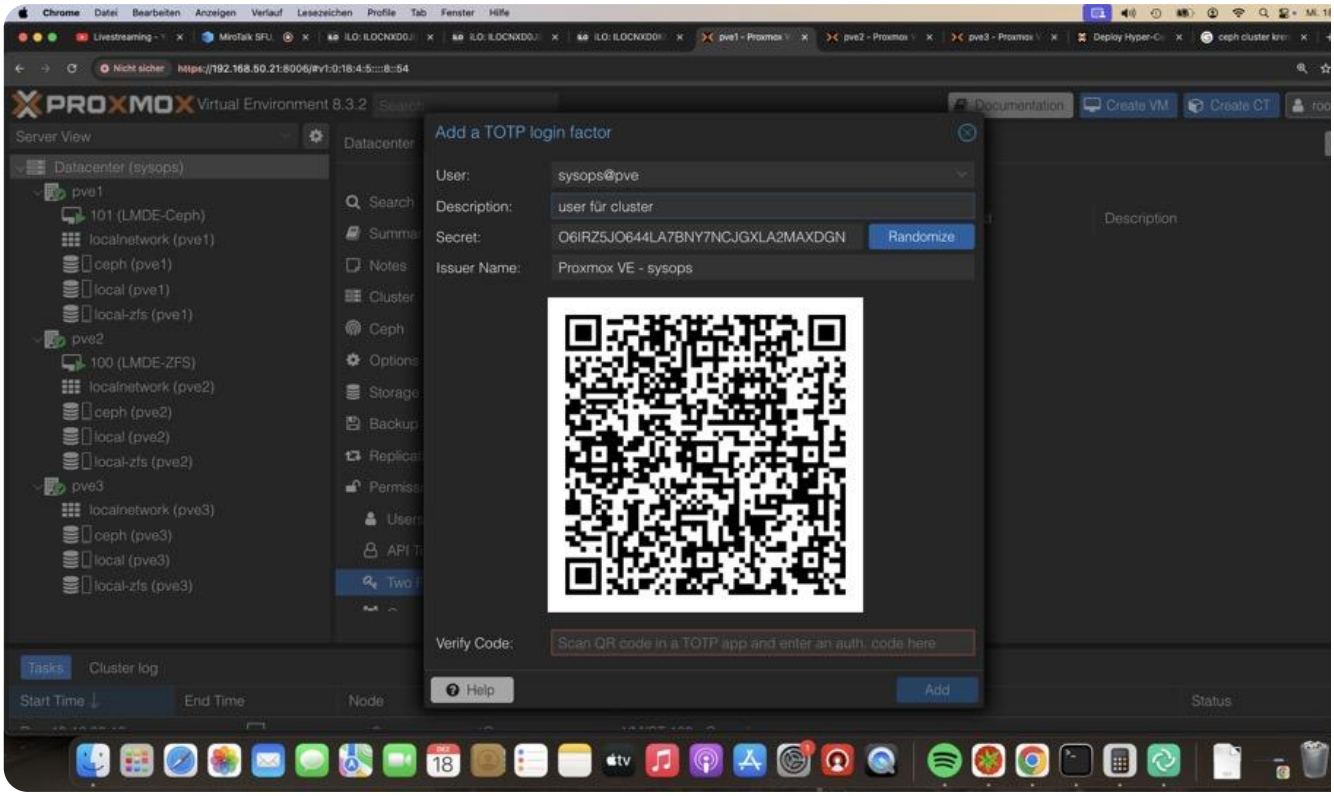


Proxmox Sicherheit: 2FA, SSH und Netzwerksegmentierung

Zwei-Faktor-Authentifizierung aktivieren

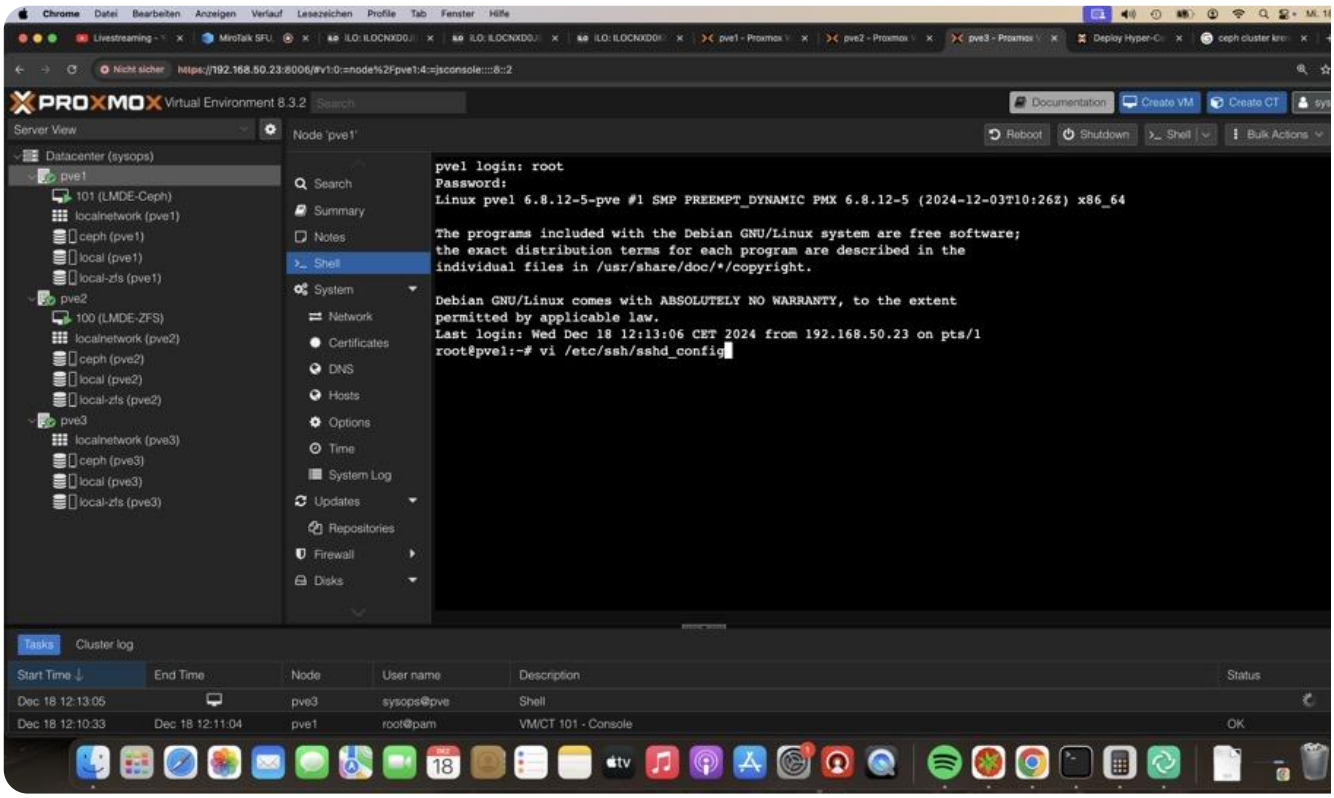
Root- und Proxmox-Benutzer sollten mit Zwei-Faktor-Authentifizierung abgesichert werden.





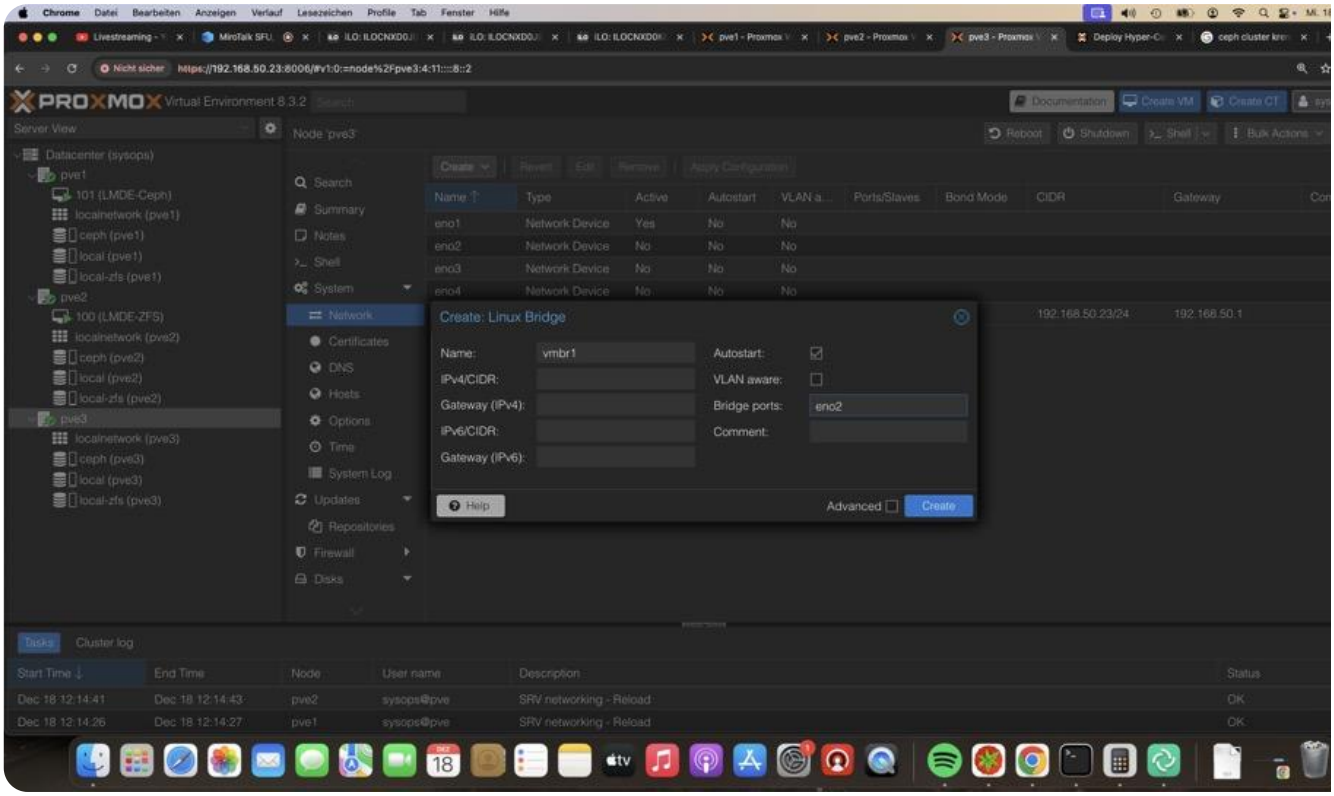
SSH gegen Passwort-Login absichern

SSH-Zugänge sollten nach Möglichkeit nur mit Schlüsseln und ohne Passwort-Login betrieben werden.



Verwaltungsnetz trennen

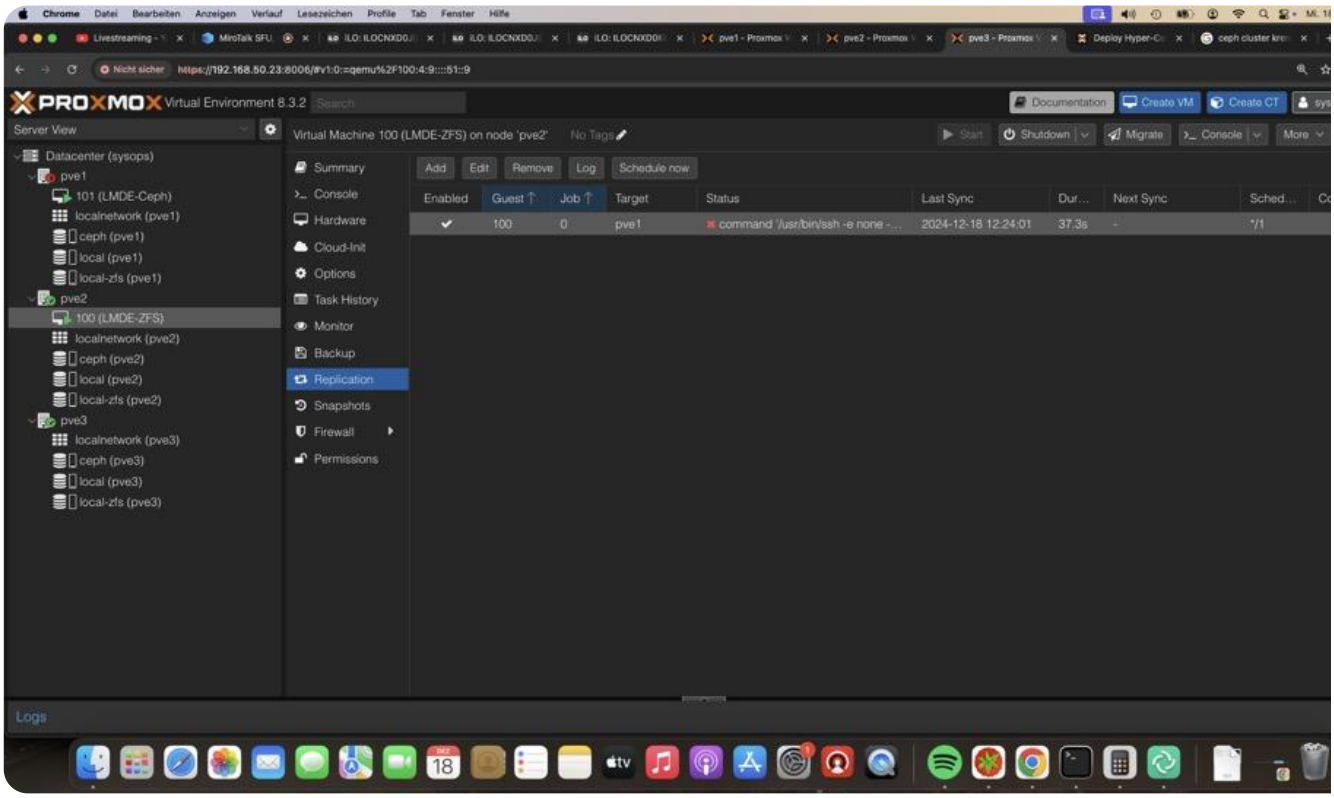
Wenn die Ressourcen vorhanden sind, sollten Anwender- und Gast-VMs keinen direkten Zugriff auf Proxmox VE, Switches oder Firewalls erhalten.

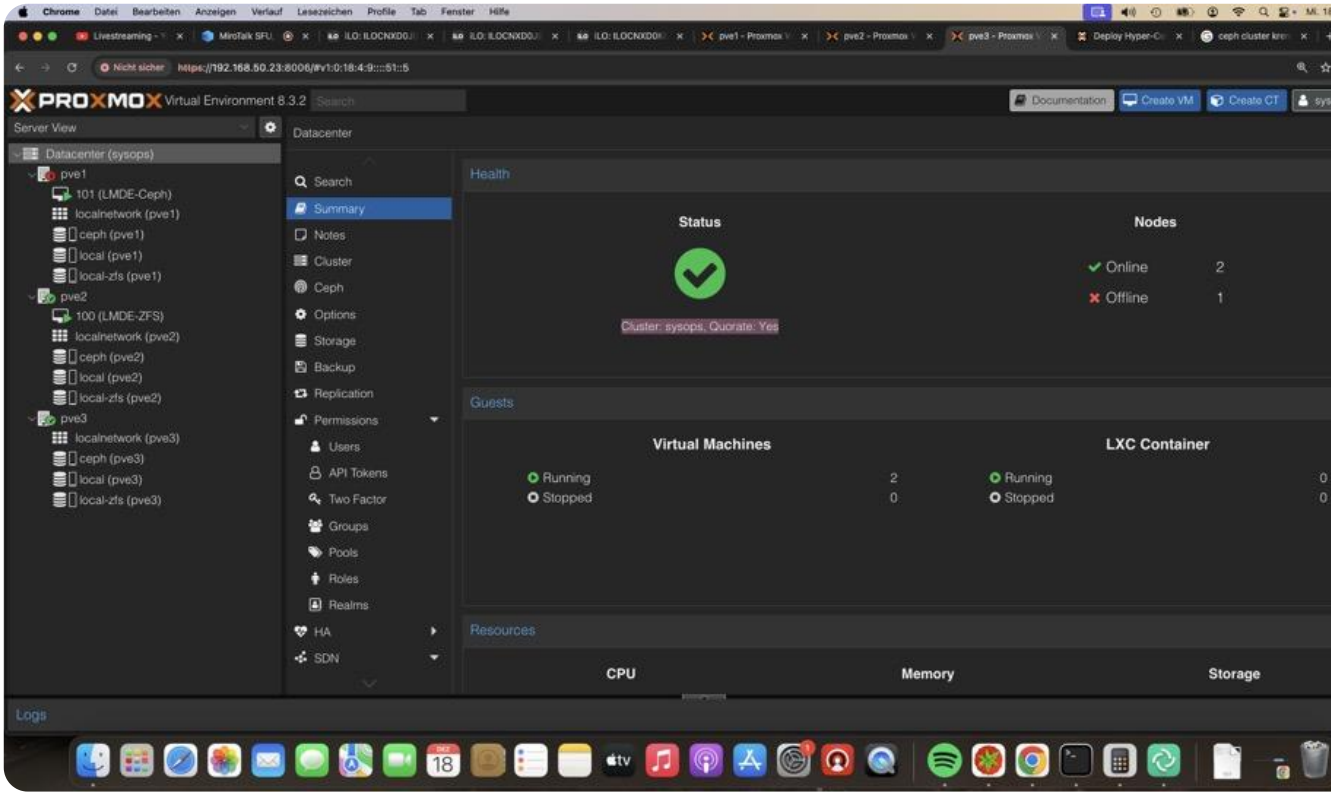


Replikationen zuverlässig prüfen

Ein Blick auf ZFS allein zeigt nicht zuverlässig, welche VM gerade auf welchem Host aktiv ist. Deshalb müssen Replikationen kontrolliert und nachvollziehbar überwacht werden.

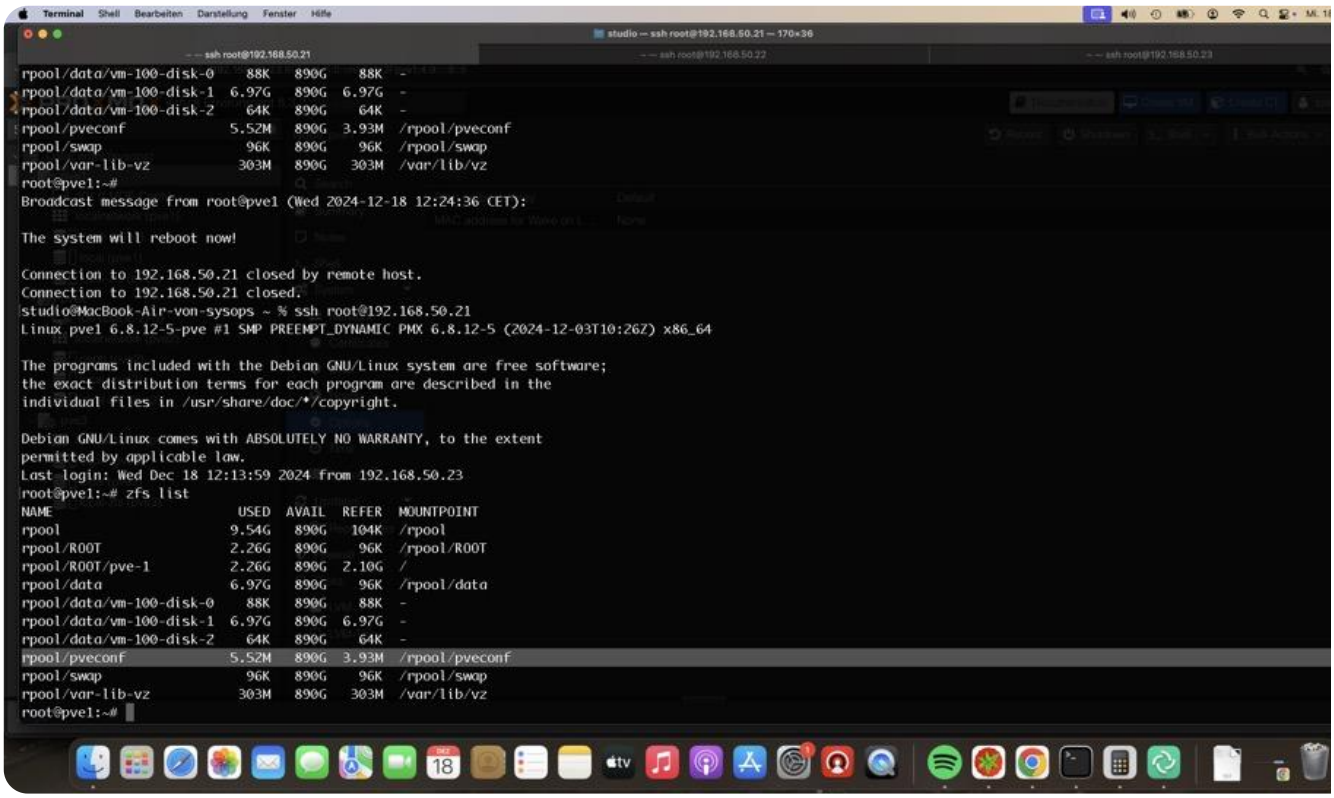
Das Proxmox Dashboard zeigt defekte Platten, RAID-Probleme und Replikationsfehler nicht immer so zentral, wie es für größere Umgebungen wünschenswert wäre. Deshalb sollten Checkmk, CV4PVE oder eigene Prüfwerkzeuge eingesetzt werden.





Proxmox Konfiguration sichern

Die Proxmox-Konfiguration liegt in einer Cluster-Datenbank und wird unter `/etc/pve` eingebunden. Unser Postinstaller kann diese Konfiguration regelmäßig nach `/rpool/pveconf` sichern und mit ZFS-Autosnapshots historisieren.



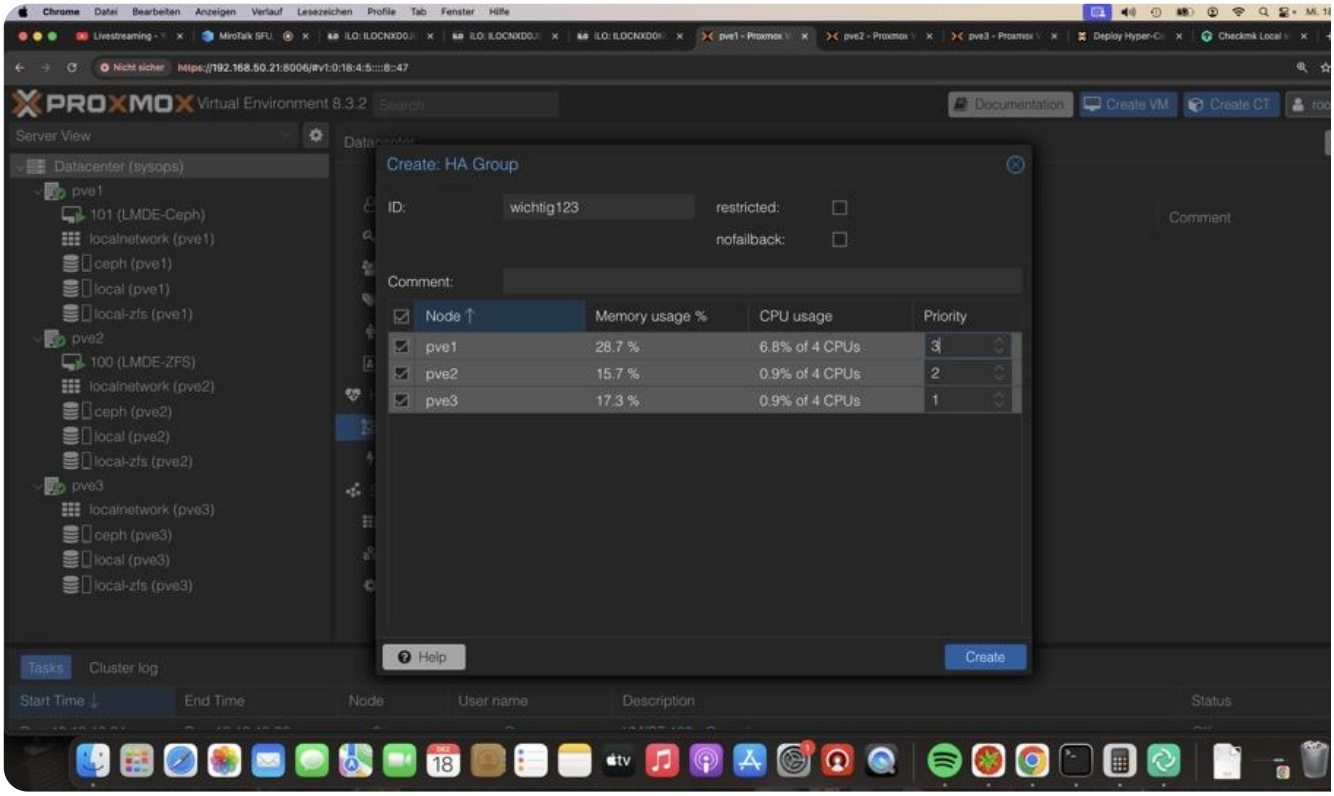
Hochverfügbarkeit mit Proxmox HA

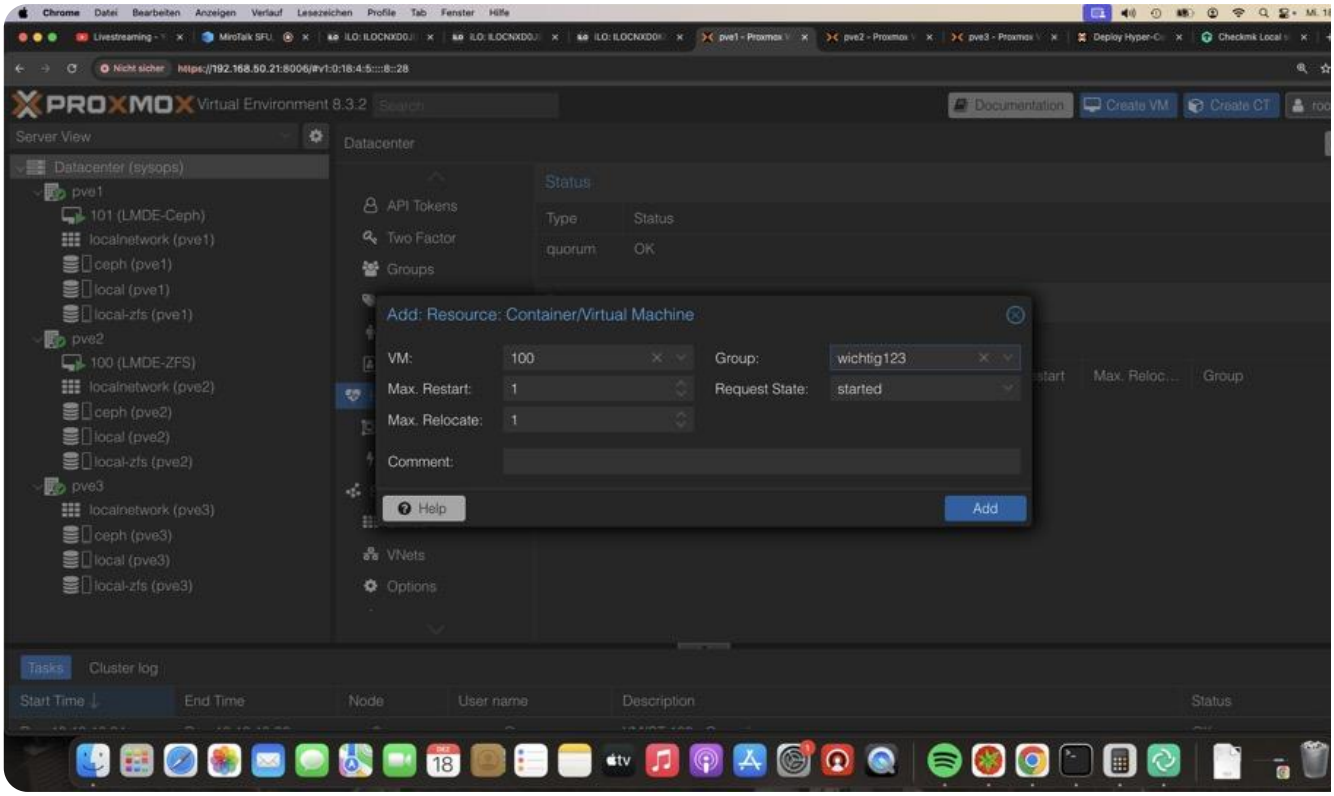
Die Proxmox HA-Engine sorgt dafür, dass VMs oder LXCs beim Ausfall eines Hosts auf einem anderen Host gestartet werden können.

Wichtig ist die Unterscheidung:

- **Ceph oder SAN:** Storage ist gemeinsam verfügbar
- **ZFS-Replikation:** es kann die Replikationslücke verloren gehen
- **HA mit ZFS:** gut für Dienste mit tolerierbarem Datenverlustfenster

Bei ZFS hängt das mögliche Datenverlustfenster vom Replikationsintervall ab. Daher eignet sich HA mit ZFS besonders für Systeme wie Firewalls, Telefonanlagen oder Dienste, bei denen ein kurzer Rückstand vertretbar ist.



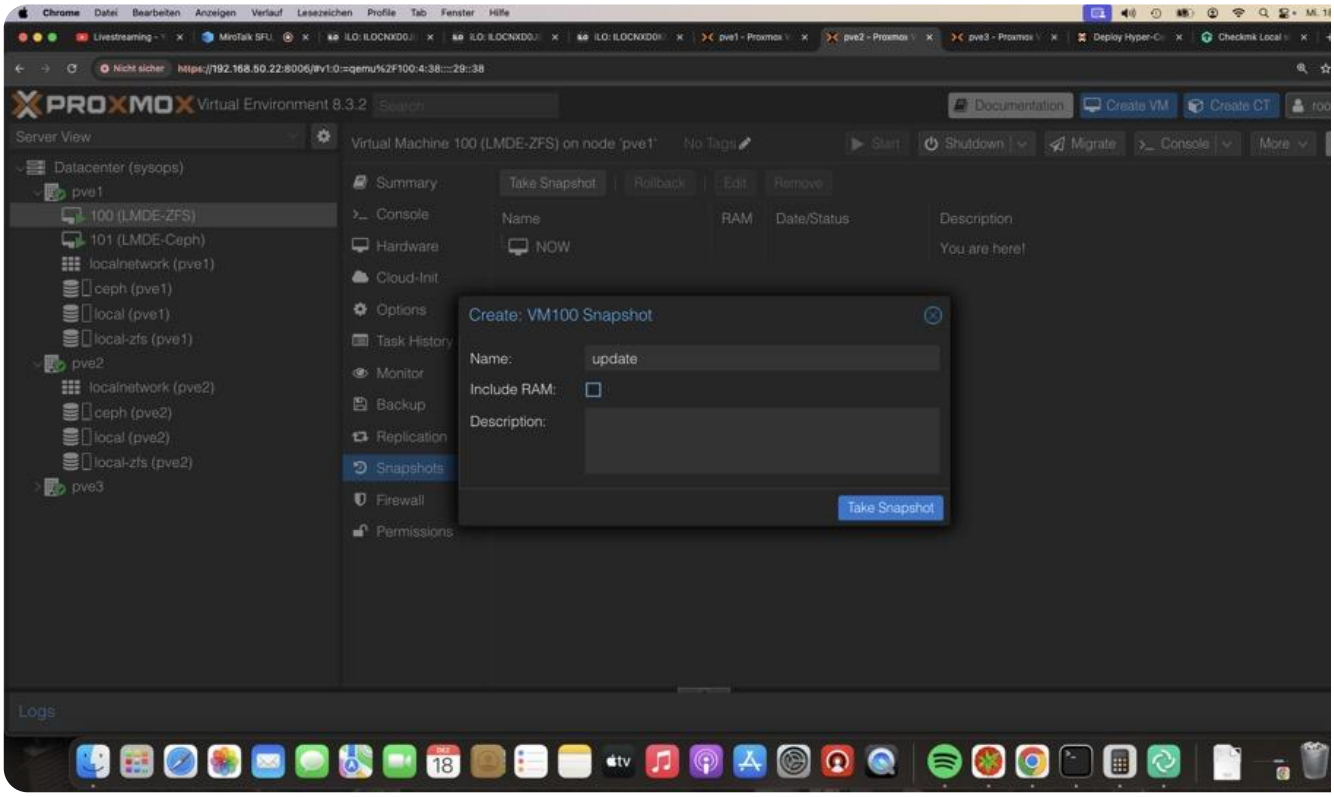


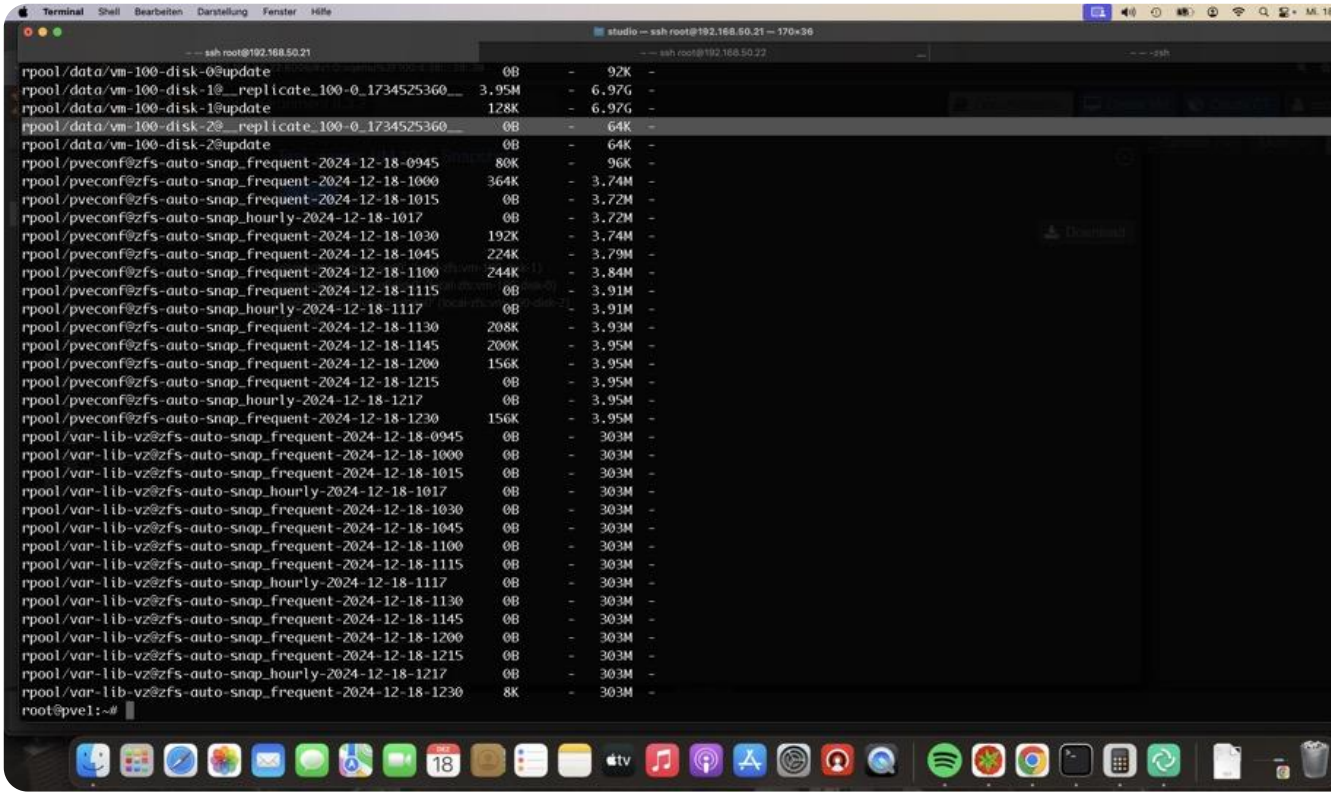
> Wird ein Node regulär neu gestartet, sollten VMs vorher bewusst migriert werden. Ein Reboot ersetzt keine geplante Wartung.

Snapshots manuell und automatisch nutzen

Manuelle Snapshots über die Proxmox GUI oder API werden abhängig vom darunterliegenden Storage erzeugt – beispielsweise auf ZFS, Ceph oder anderen Storage-Backends.

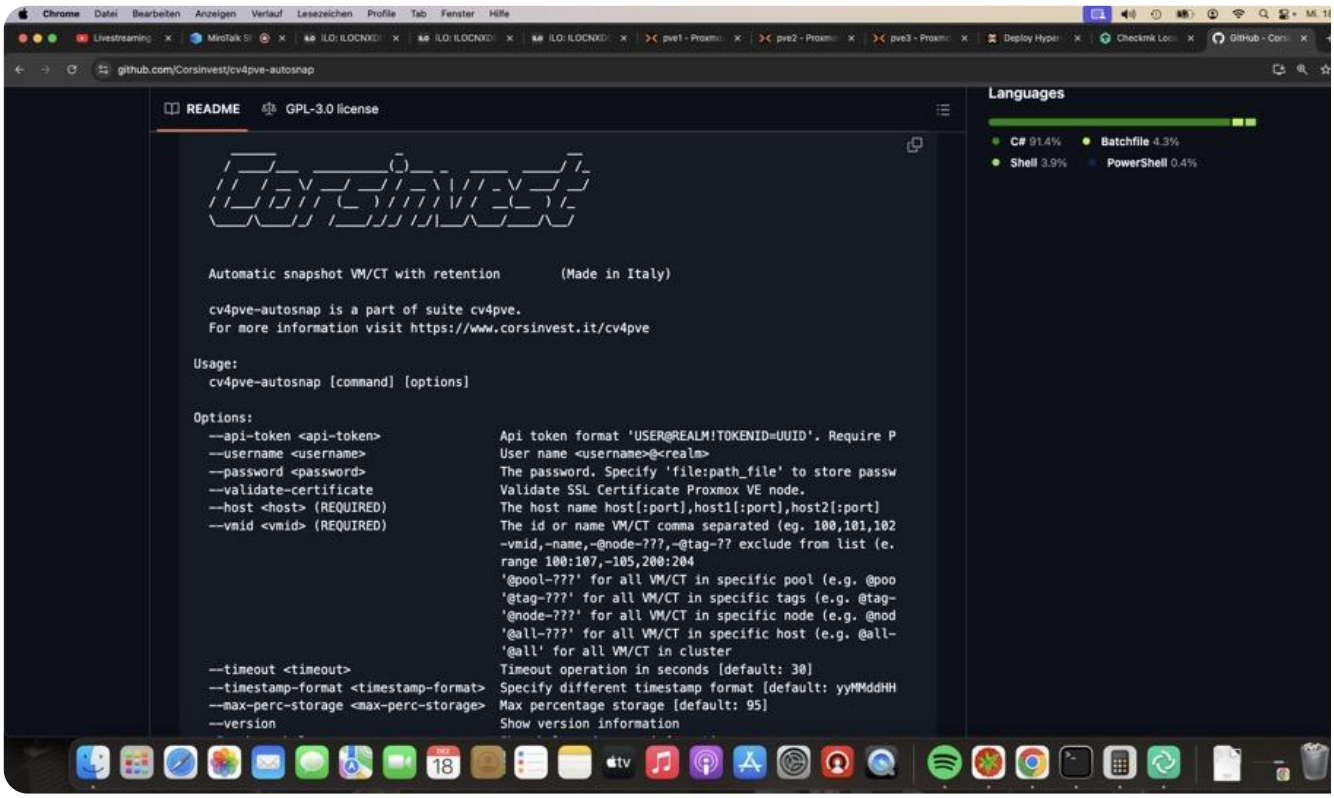
Bei ZFS werden Snapshot-Zustände mitrepliziert und in der VM-Konfiguration sichtbar.

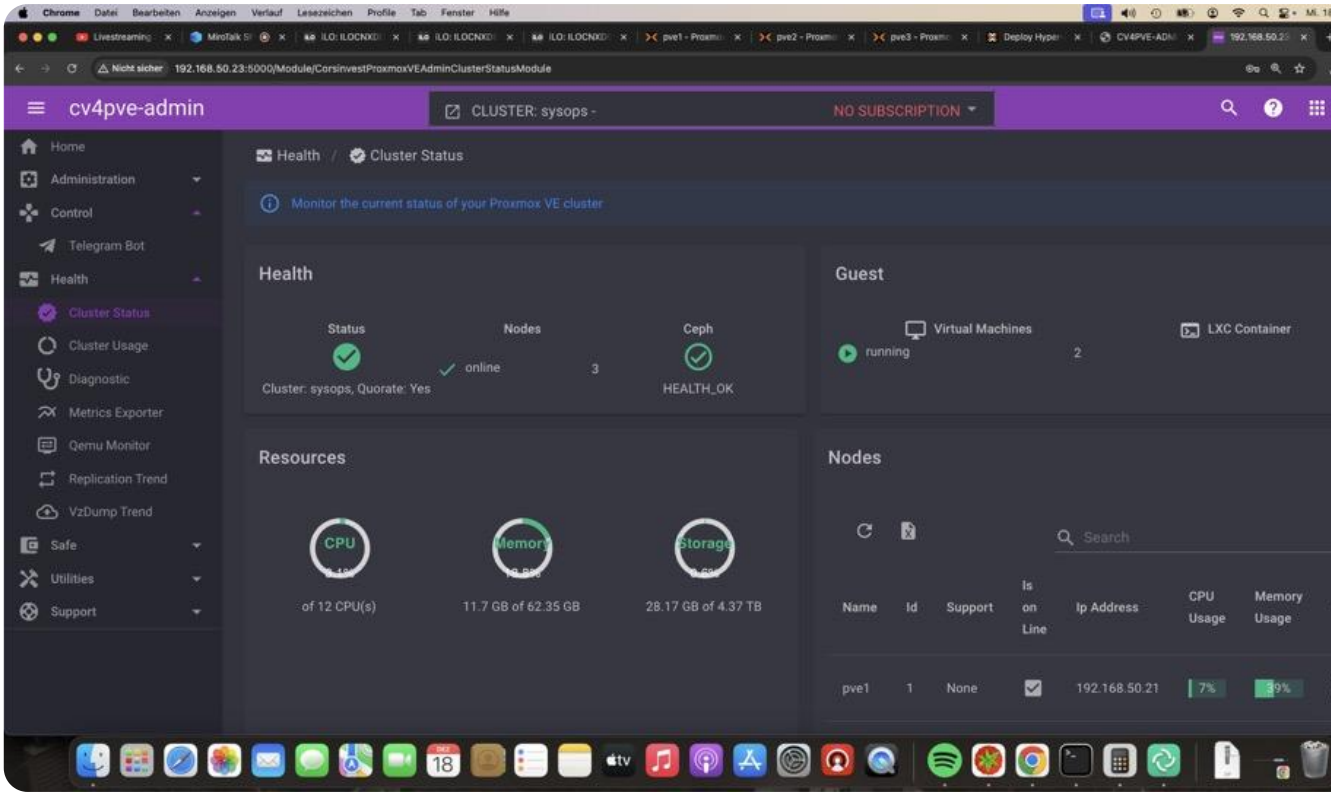




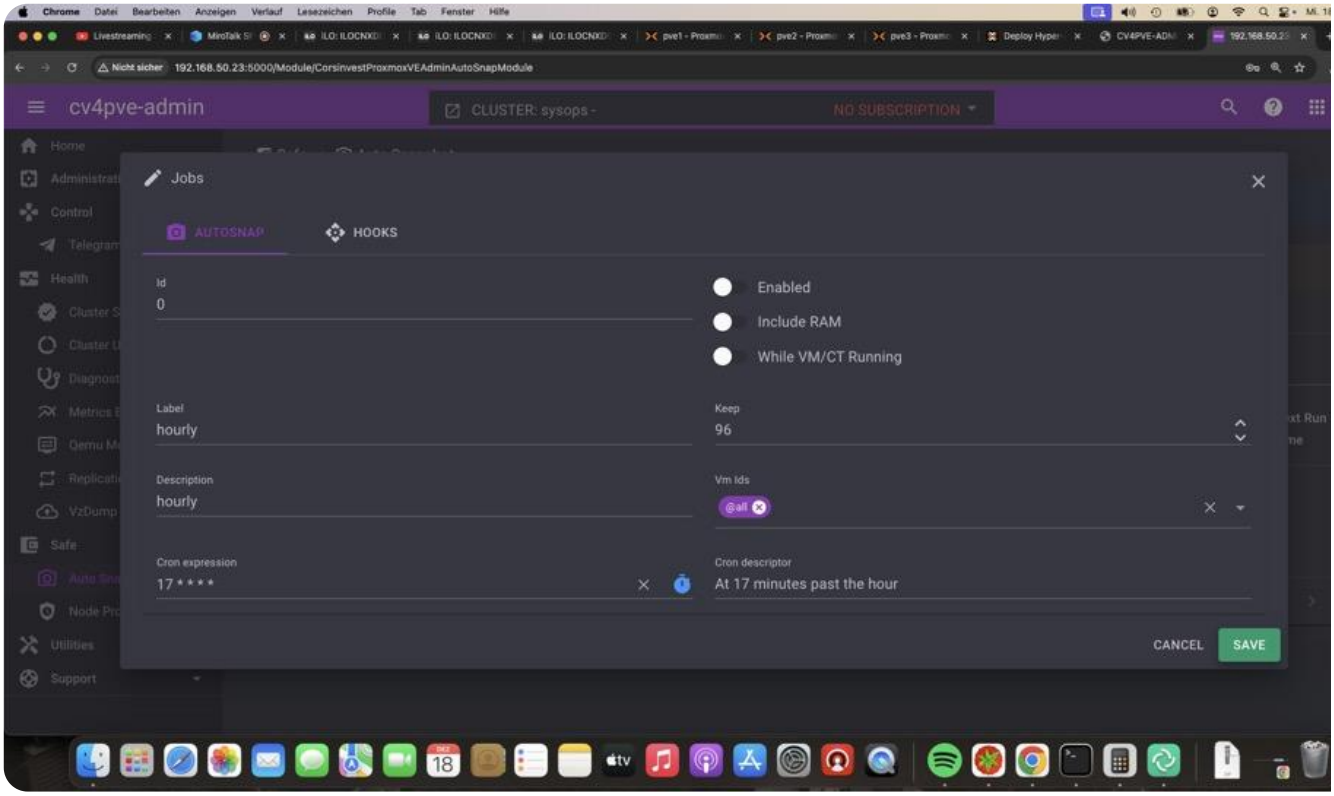
CV4PVE-Admin für Snapshots, Monitoring und Replikationen

CV4PVE-Admin kann eine sinnvolle Ergänzung zu Proxmox VE sein. Es bietet bessere Übersicht über Replikationen, Snapshots und bestimmte Monitoring-Funktionen.





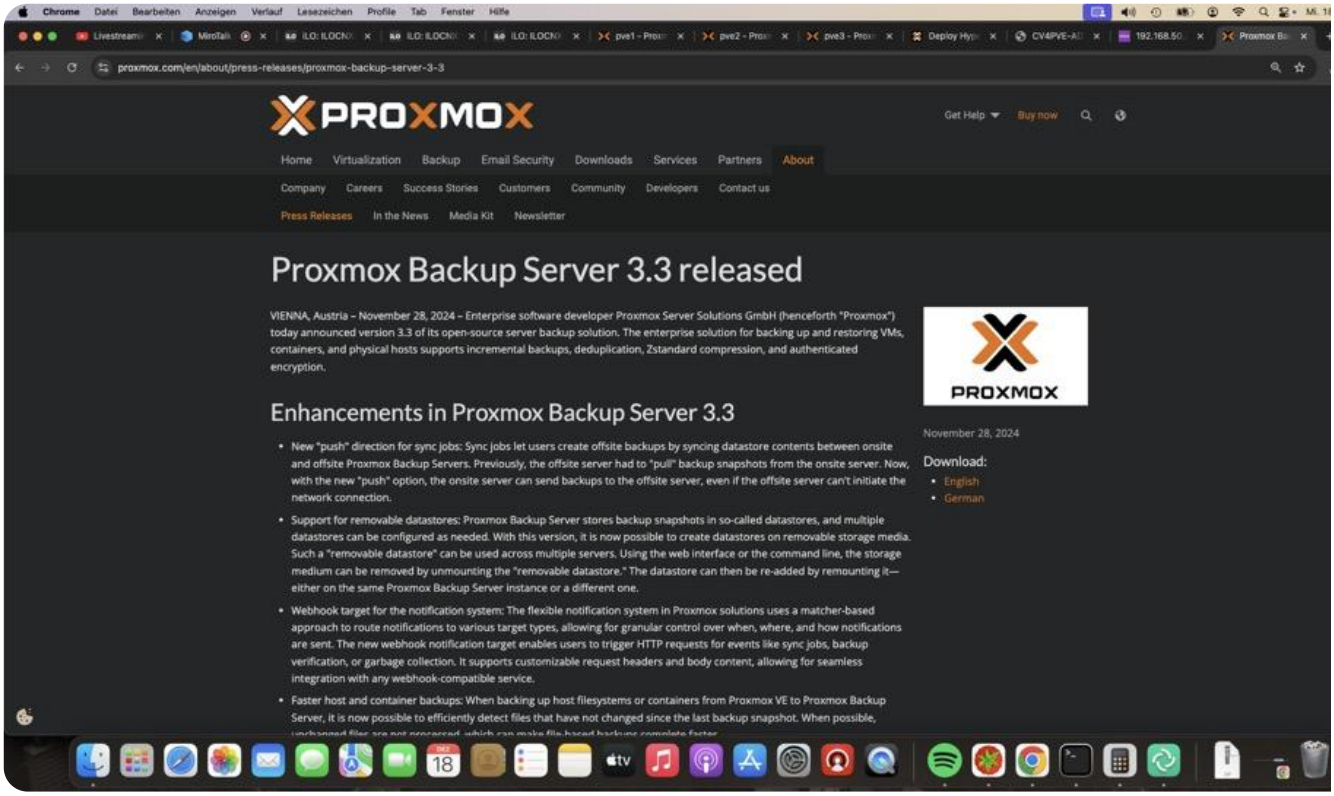
Automatische Snapshots können dort definiert werden. Wichtig ist, dass das System zuverlässig läuft und überwacht wird.

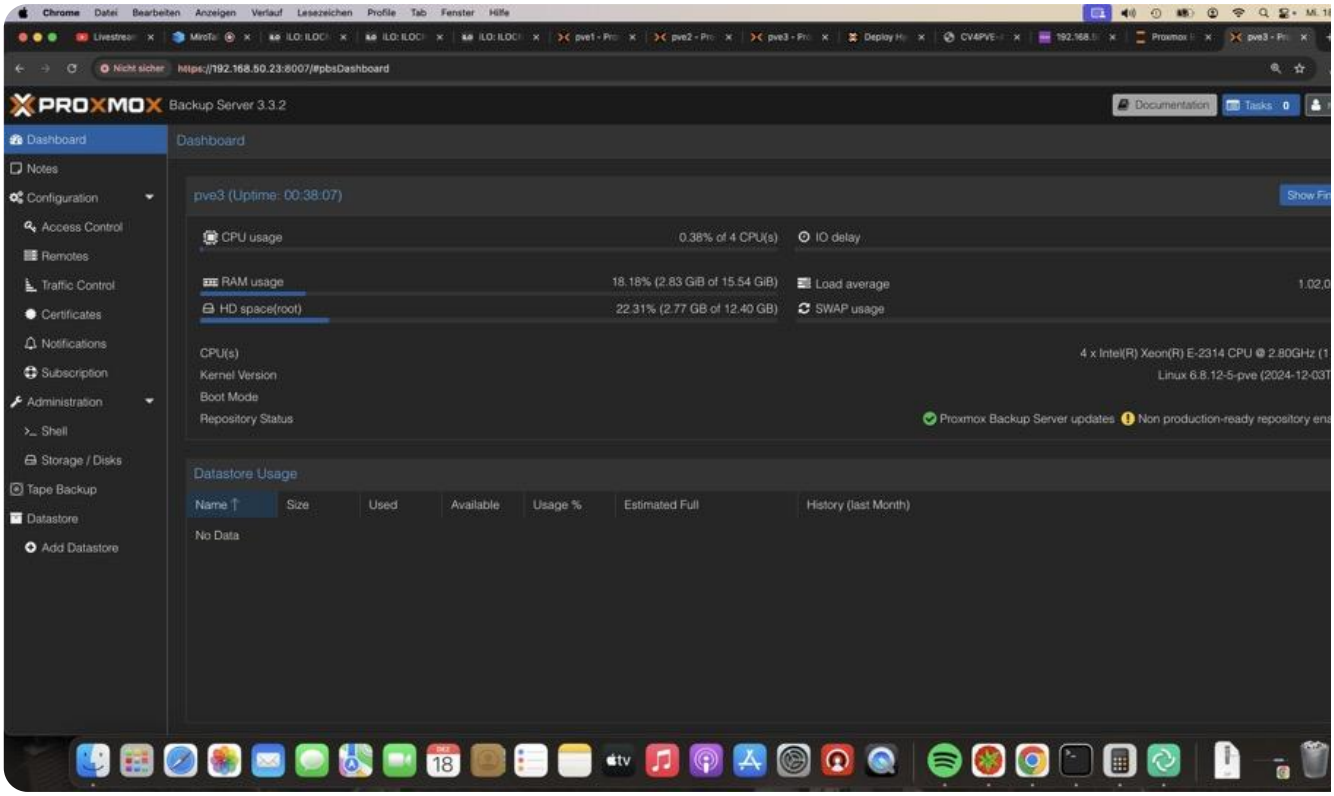


Proxmox Backup Server einrichten

Proxmox VE kann klassische Backups erstellen. Der **Proxmox Backup Server** ergänzt inkrementelle, deduplizierte Backups, zentrale Verwaltung, Verify Jobs, Prune Jobs und Garbage Collection.

PBS kann auf eigener Hardware, in einer VM oder in bestimmten Szenarien auch als LXC betrieben werden. Für produktive Backups ist ein dediziertes, abgesichertes System vorzuziehen.



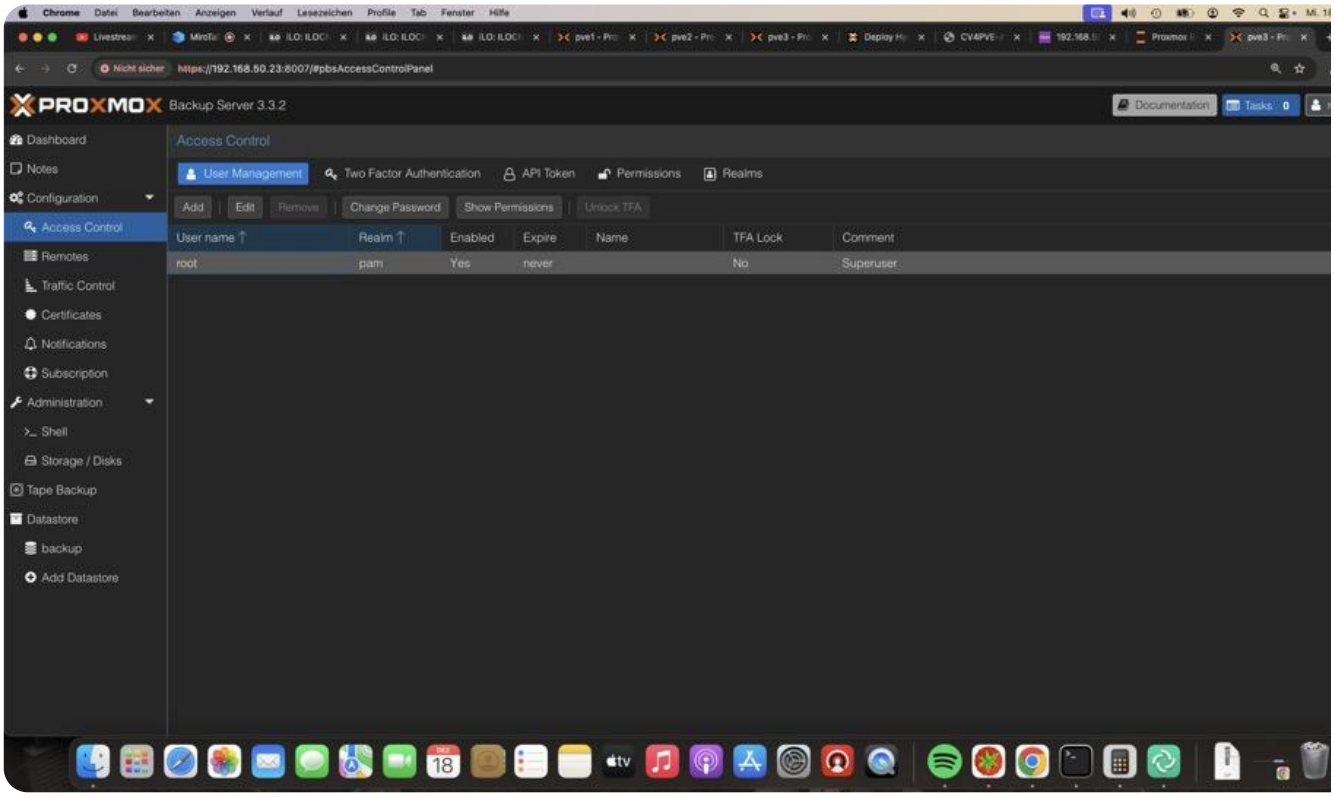


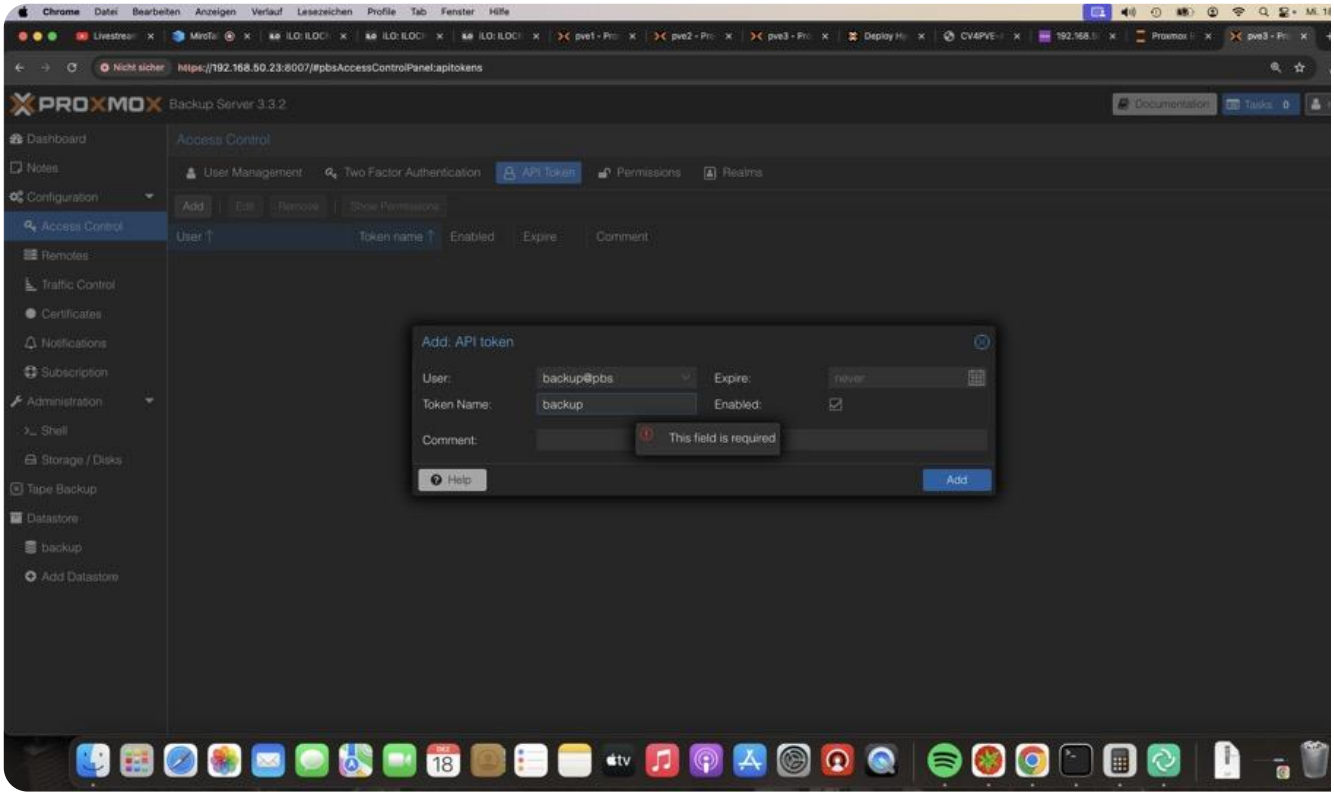
Für den Datastore empfehlen sich lokale Dateisysteme oder geeignete Storage-Anbindungen. SMB verursacht in vielen Umgebungen zusätzlichen Overhead und ist für PBS meist nicht die beste Wahl.

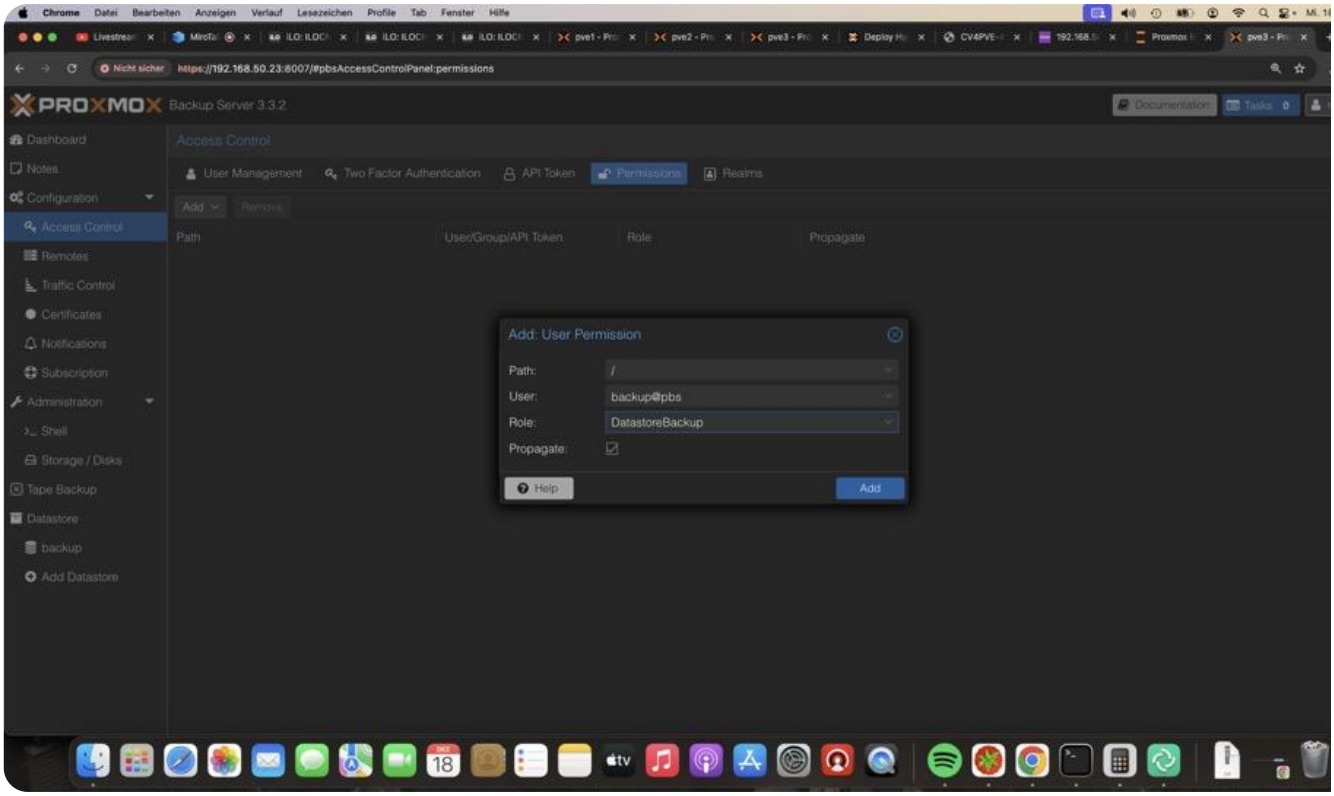
Backup-Benutzer, Token und Rechte

Für Backups sollte ein eigener Backup-Benutzer mit Token verwendet werden. Root-Zugänge und Backup-Zugänge sollten mit Zwei-Faktor-Authentifizierung abgesichert werden.

Der Benutzer oder Token benötigt die Berechtigung **Datastore Backup**.



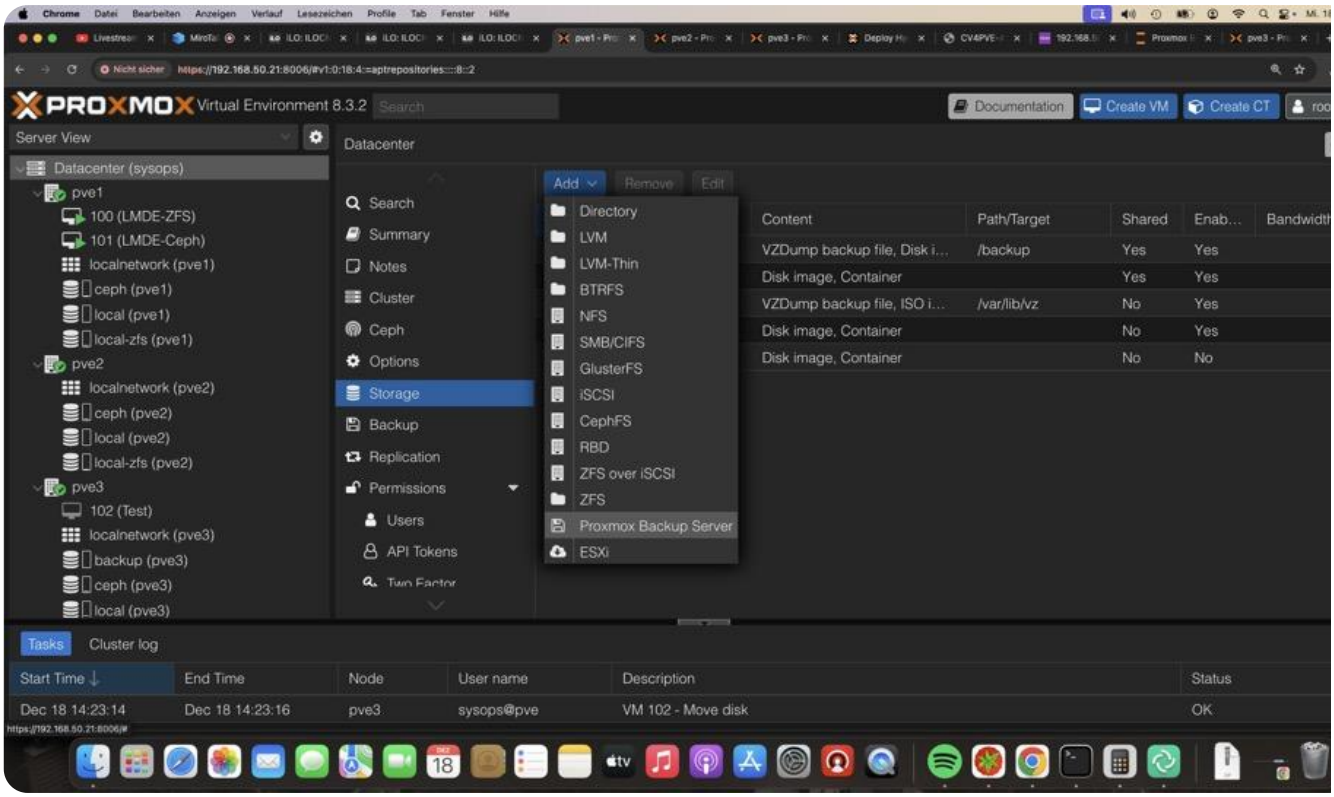


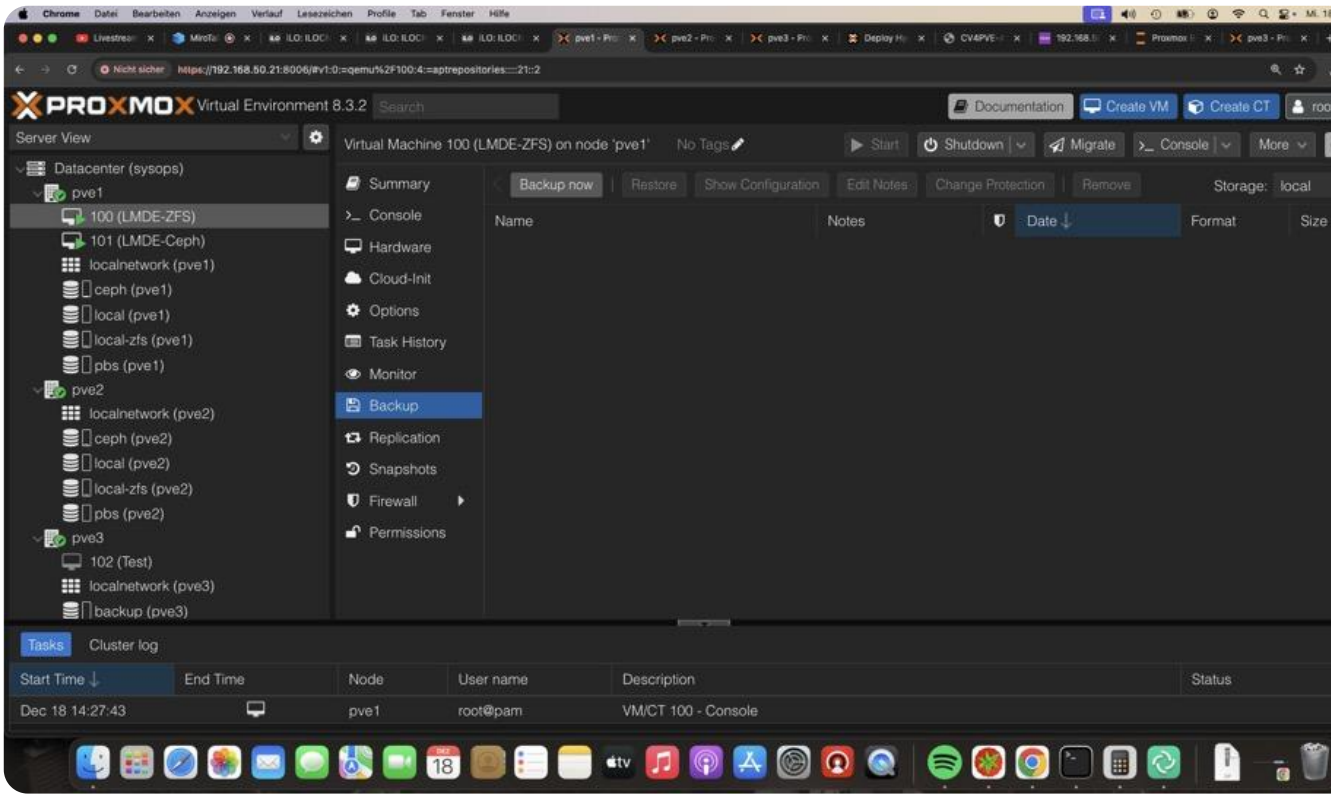


Proxmox Backup Server in Proxmox VE einbinden

Für die Einbindung des Proxmox Backup Servers in Proxmox VE werden mehrere Angaben benötigt:

- Name des Datastores in Proxmox VE
- Serveradresse des PBS
- Datastore-Name im PBS
- Benutzer oder Token
- Token Secret
- Fingerprint des PBS-Zertifikats



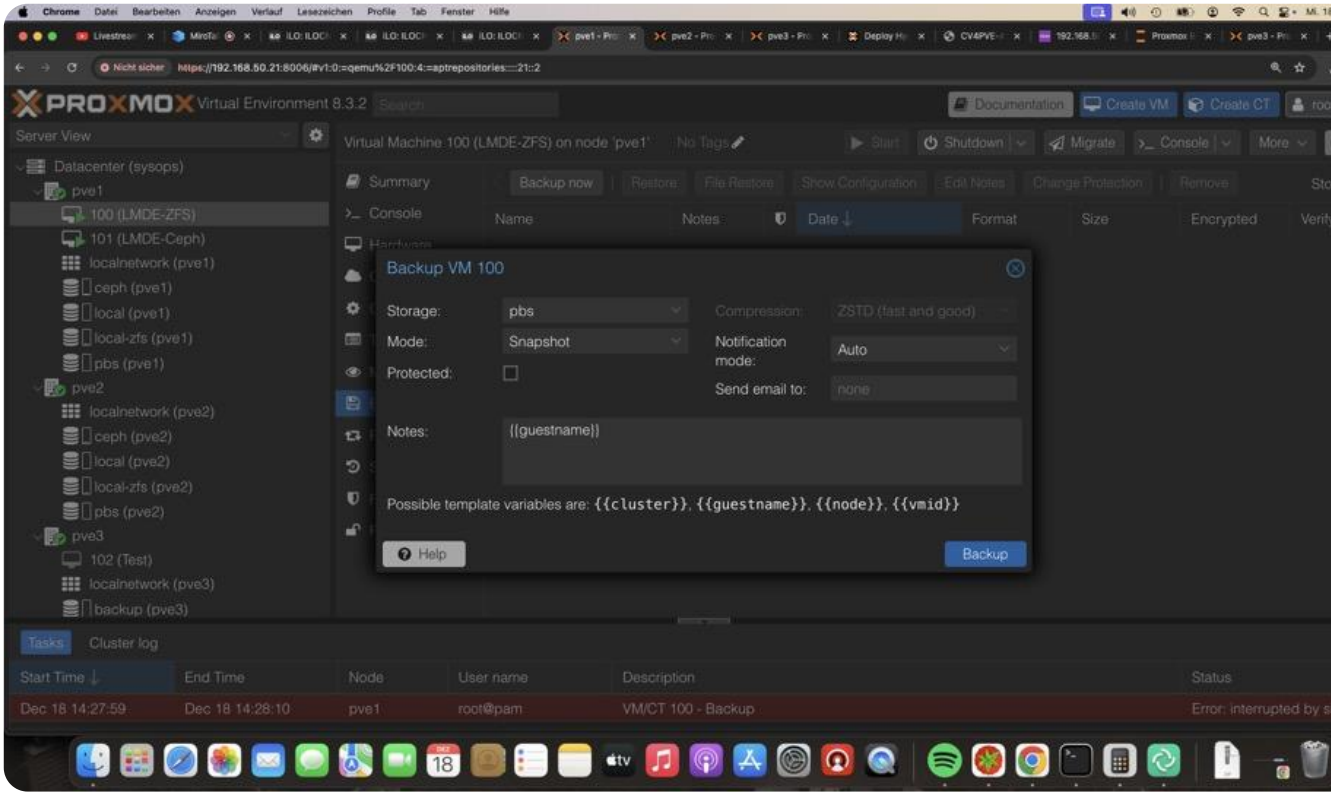


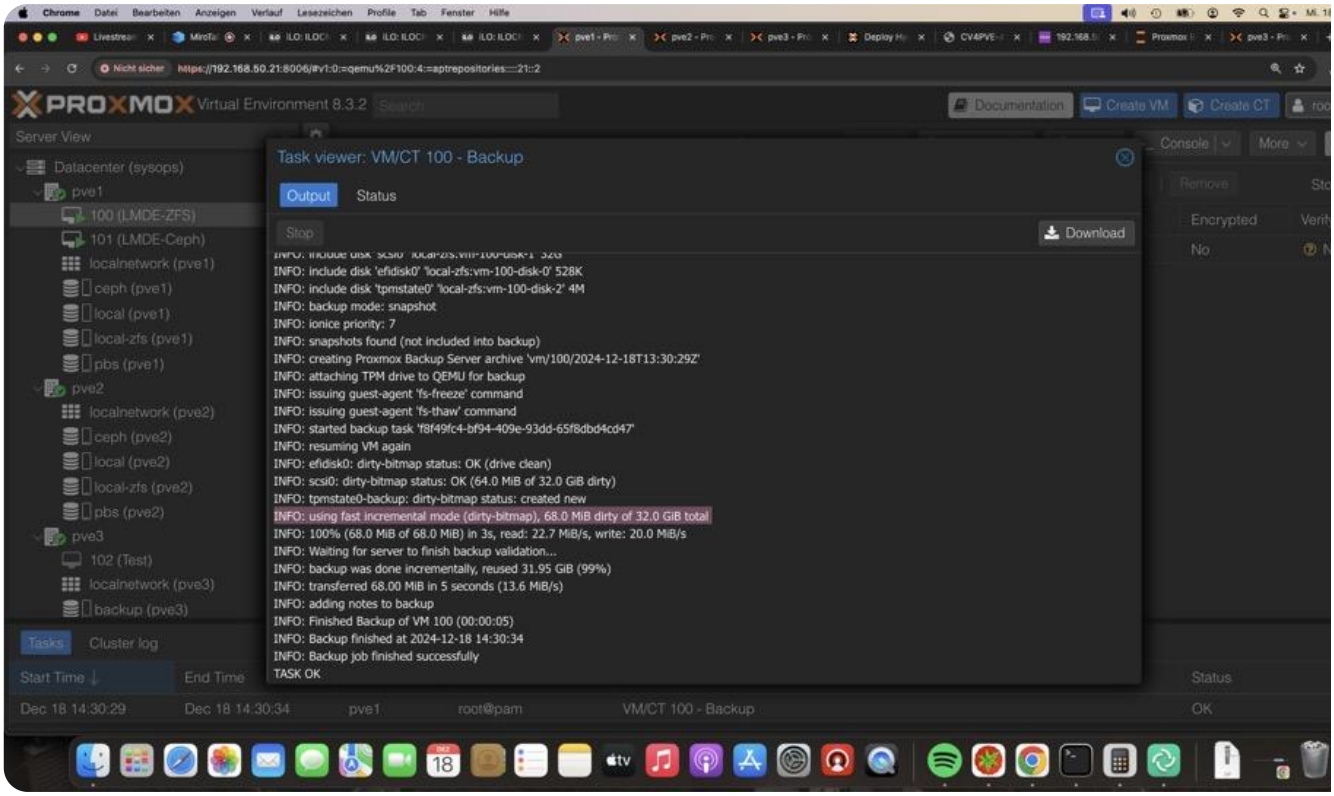
> Das lokale Bootvolume von Proxmox VE sollte nicht als Backup-Ziel verwendet werden.

Backups, Prune, Garbage Collection und Verify

Das erste Backup einer VM ist in der Regel ein vollständiger Durchlauf. Danach werden mit Dirty Bitmap und Deduplizierung nur noch geänderte Daten effizient übertragen und gespeichert.

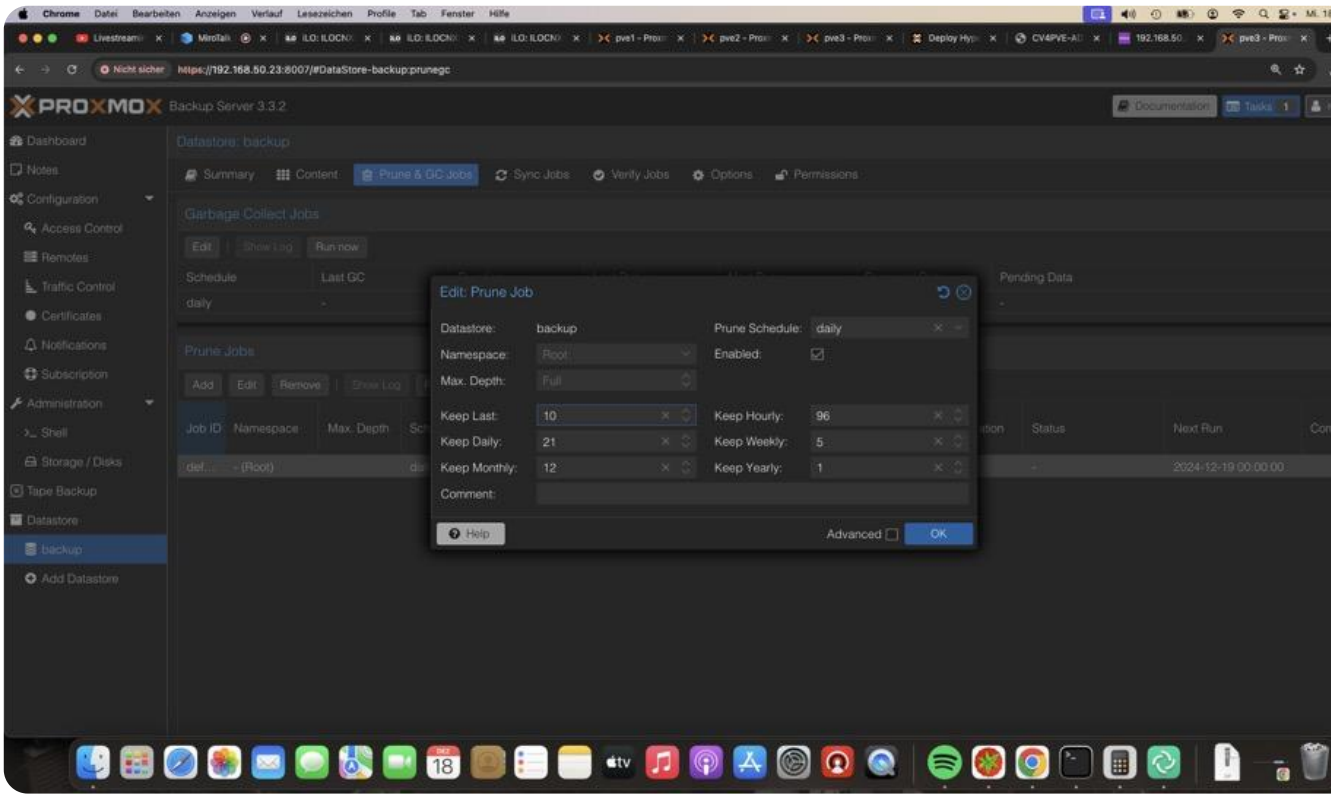
Nach einem Stop der VM oder Reboot des Hosts kann wieder ein vollständiger Scan notwendig werden. Durch Chunk-Deduplizierung werden jedoch nur neue oder geänderte Chunks gespeichert.





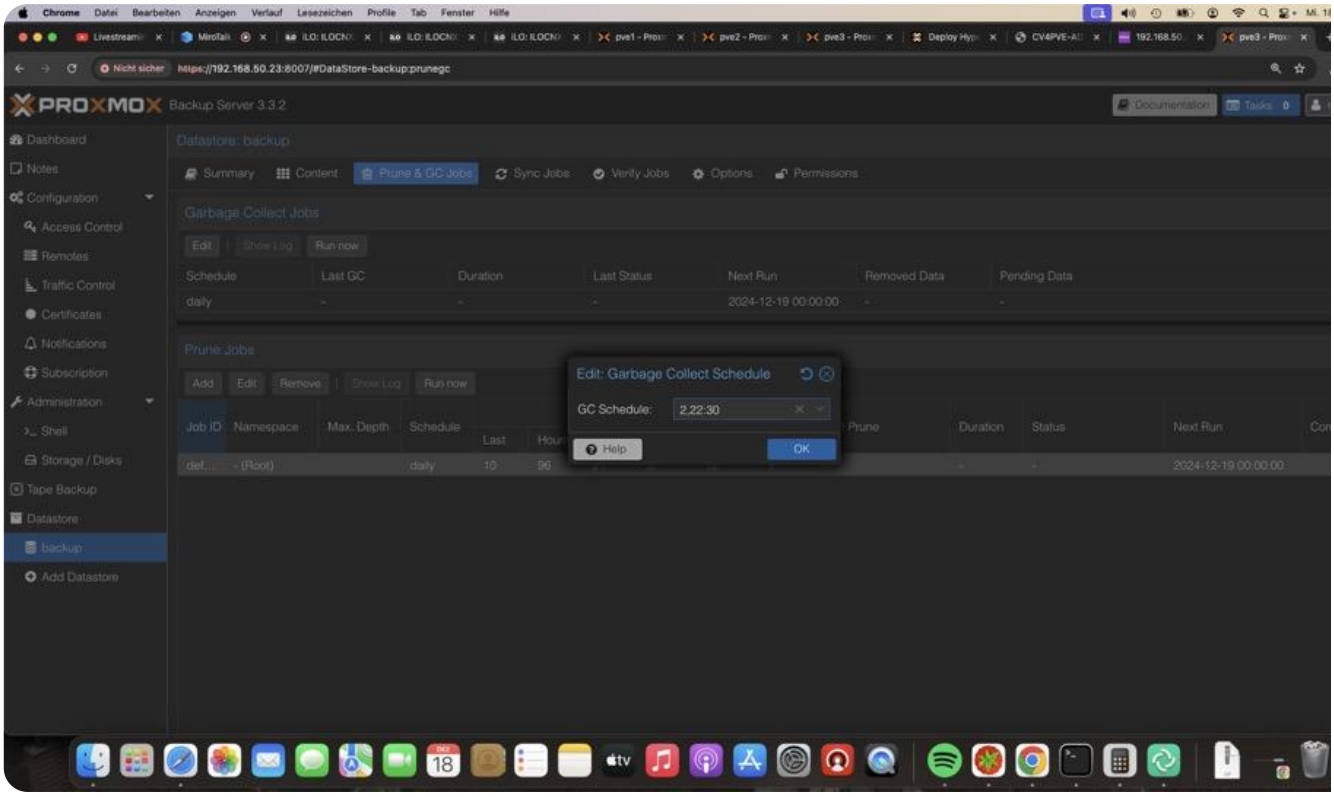
Prune Jobs

Prune Jobs definieren Aufbewahrungszeiten. Der Proxmox-Simulator hilft bei der Prüfung der Regeln.



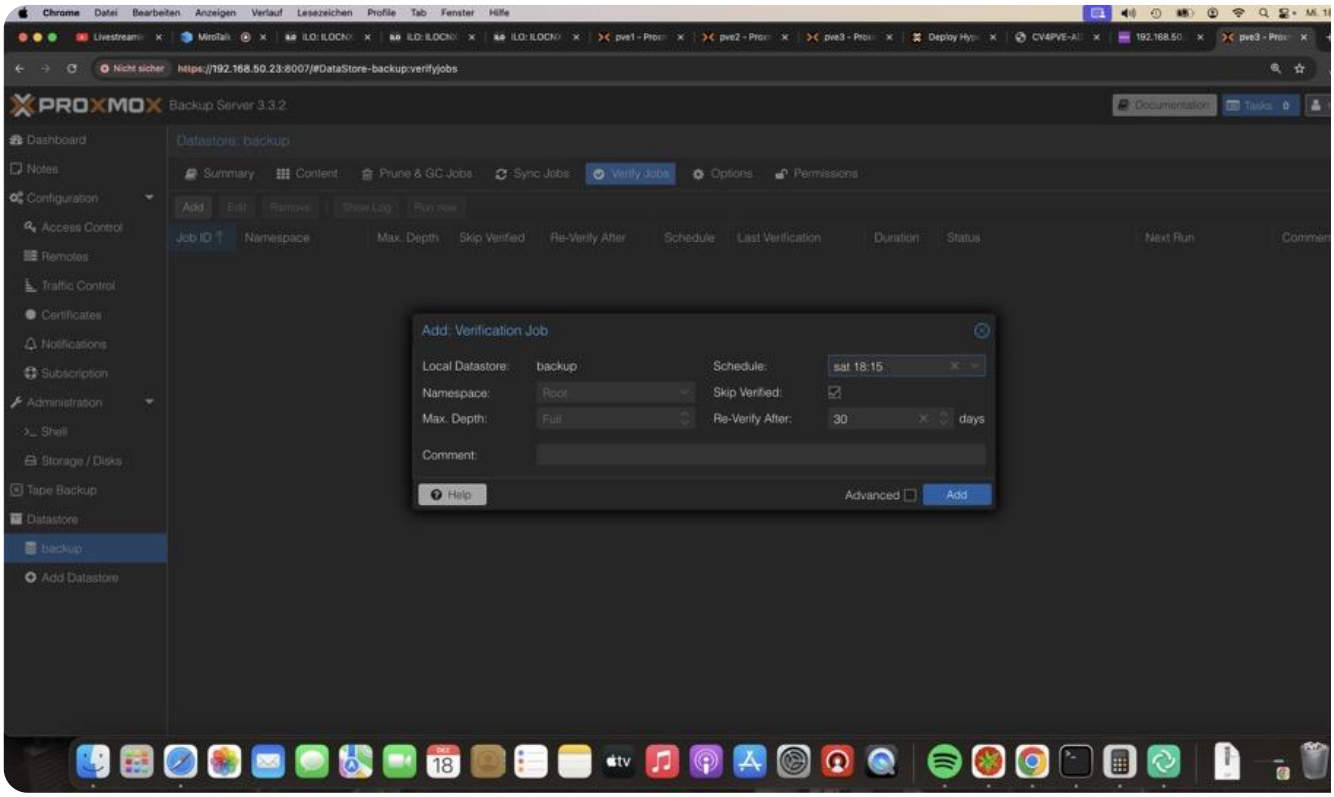
Garbage Collection

Die Garbage Collection räumt nicht mehr benötigte Chunks auf. Dieser Vorgang kann Last erzeugen und sollte sinnvoll geplant werden.



Verify Jobs

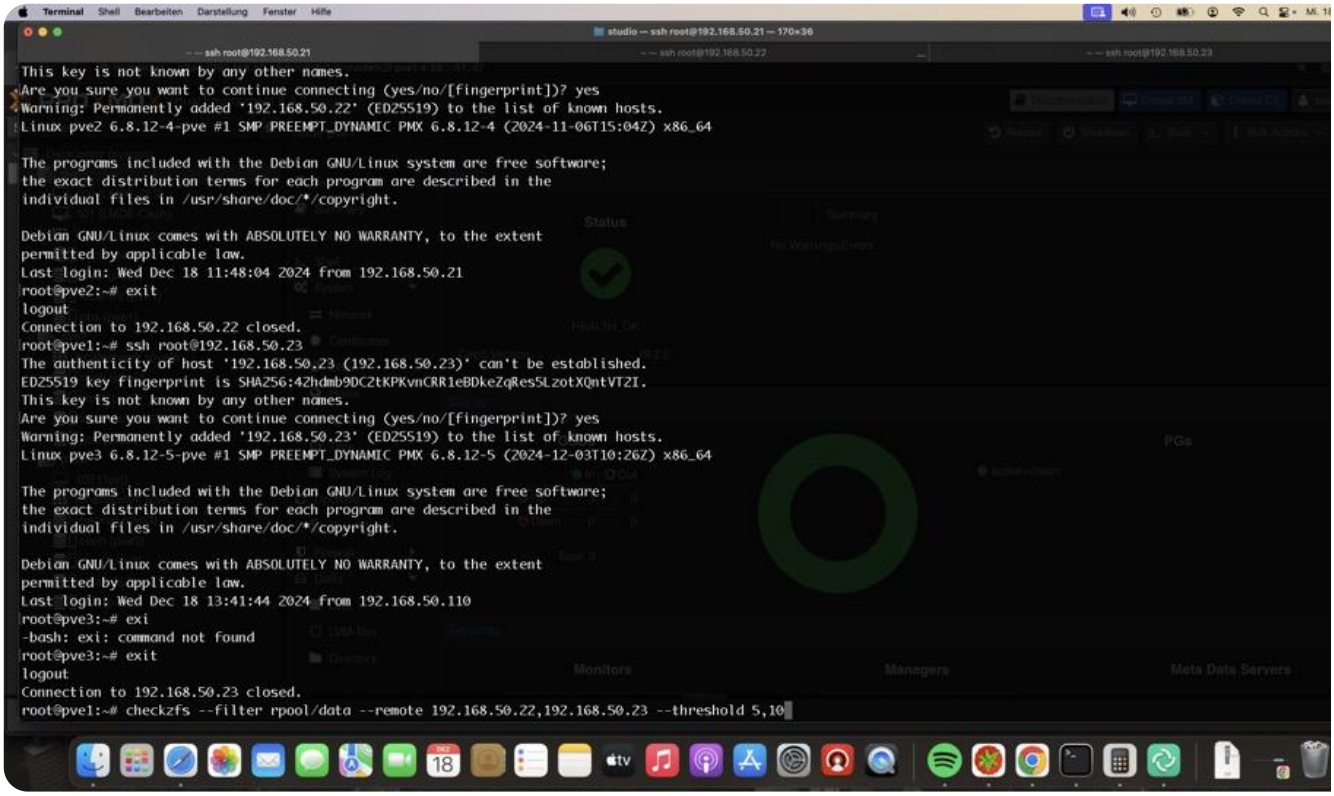
Verify Jobs prüfen regelmäßig die Integrität der Backups. Der Erfolg wird sowohl im PBS als auch in Proxmox VE angezeigt.

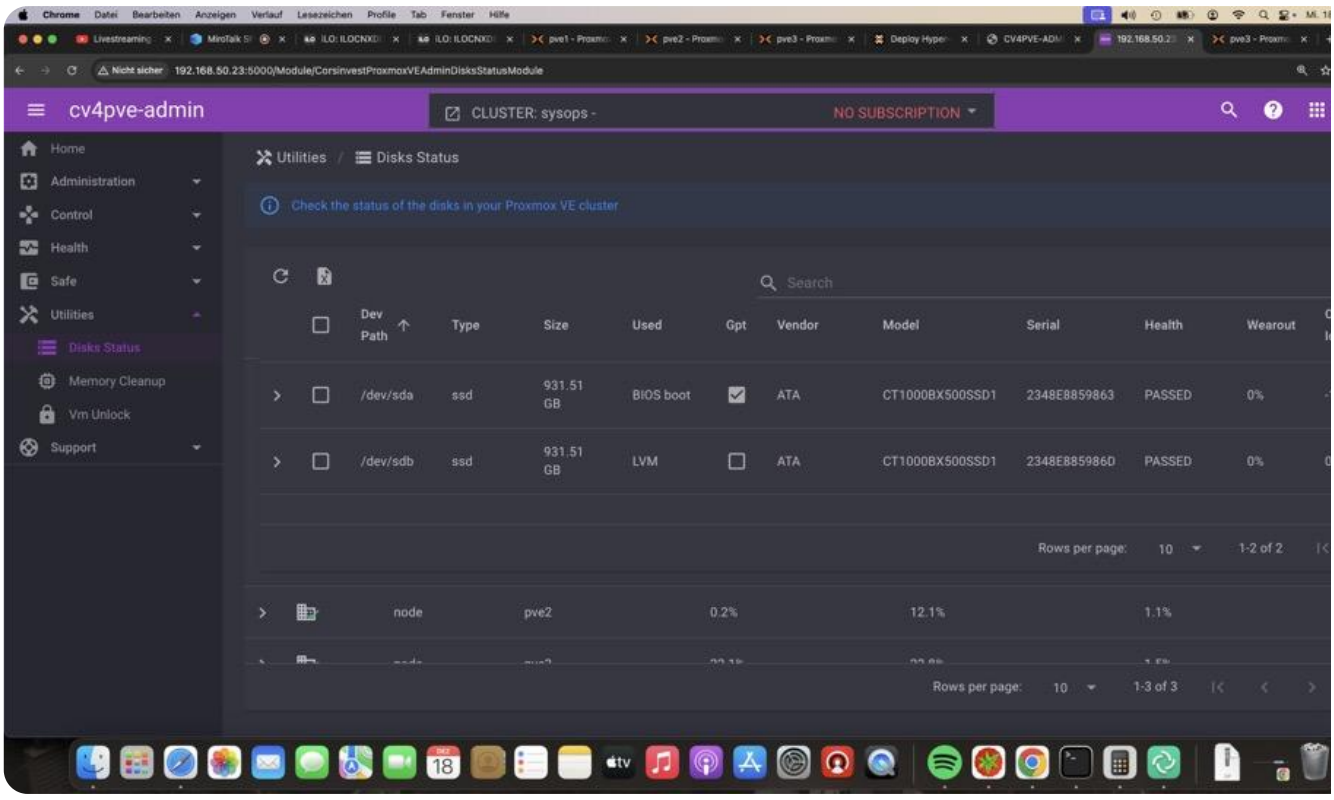


Optional: Replikationskontrolle mit checkzfs

Mit **checkzfs** vom Bashclub können ZFS-Replikationen zusätzlich kontrolliert werden. Da nicht immer sofort erkennbar ist, welcher Host welche VM aktiv betreibt, kann ein Vergleich der Datastores hilfreich sein.

Der Check sollte nicht während laufender Replikationen ausgeführt werden, da dies zu Fehlmeldungen führen kann.





Fazit: Proxmox Cluster mit ZFS bewusst planen

Für viele kleine und mittlere Umgebungen ist ein Proxmox-Cluster mit ZFS, Replikation, Proxmox Backup Server und gutem Monitoring eine sehr leistungsfähige und wirtschaftliche Lösung.

Ceph ist besonders interessant, wenn echter verteilter Storage, hohe Skalierung und passende Hardware vorhanden sind. Für kleinere Installationen ist ZFS jedoch häufig einfacher zu verstehen, schneller einzurichten und leichter zuverlässig zu betreiben.

- ZFS für kleine und mittlere Cluster bevorzugt prüfen
- Ceph nur mit passender Hardware und Erfahrung produktiv einsetzen
- Backups nie durch Snapshots ersetzen
- Replikationen regelmäßig kontrollieren
- Monitoring zwingend einplanen
- 2FA und SSH-Hardening aktivieren
- Proxmox Backup Server sauber absichern

Ressourcen

- **Aktueller sysops Prospekt**
- **aow.de – Linkliste zu allen Projekten**
- **cloudistboese.de – Schulungsportal für Proxmox, ZFS und Open Source**
- **sysops.tv auf YouTube**
- **Bashclub Proxmox ZFS Postinstaller auf GitHub**

Proxmox Schulung oder Beratung anfragen

Sie möchten Proxmox VE mit ZFS, Replikation, Backup und Monitoring professionell einsetzen? Die sysops GmbH unterstützt bei Planung, Installation, Migration, Schulung und Betrieb.

info@sysops.de
+49 (0) 6021 2125-0

Vielen Dank für Ihre Aufmerksamkeit.
Christian-Peter Zengel – sysops.tv

Update: Ceph-Vernetzung mit Proxmox VE 9 und Network Fabrics

Seit **Proxmox VE 9** lässt sich die Ceph-Vernetzung deutlich einfacher abbilden. Mit den neuen **SDN Network Fabrics** können geroutete Underlay-Netzwerke direkt über die Proxmox-Oberfläche erstellt werden.

Damit werden insbesondere **Full-Mesh-Netzwerke für Ceph** übersichtlicher und wartbarer. Proxmox beschreibt selbst, dass Fabrics ab PVE 9 über den SDN-Stack in der GUI erstellt werden können und sich für Full-Mesh-Ceph-Setups eignen. [oai_citation:0‡Proxmox VE](https://pve.proxmox.com/wiki/Full_Mesh_Network_for_Ceph_Server?utm_source=chatgpt.com)

Für neue Ceph-Installationen sollte daher geprüft werden, ob die Vernetzung über **Datacenter → SDN → Fabrics** sinnvoller ist als eine manuelle statische Netzwerkkonfiguration.

Version #18
Erstellt: 2024-12-19 10:26:44 UTC von Christian Zengel (sysops GmbH)
Zuletzt aktualisiert: 2026-07-08 11:17:12 UTC von Christian Zengel (sysops GmbH)