

Cluster i Alta disponibilitat

Cluster i HA

Muntatge de cluster en varios hipervisors i alta disponibilitat

Alfredo Rafael Vicente Boix i Javier Estellés Dasi

05-05-2024



CEFIRE
FORMACIÓ PROFESSIONAL
ENSENYANCES ARTÍSTIQUES
I ESPORTIVES

Virtualització als centres educatius
amb LliureX i Proxmox

Continguts

1	Introducció	3
2	Muntar clúster en PROXMOX	3
3	Muntatge d'alta disponibilitat	11
3.1	NAS i màquina virtual	11
4	Alta disponibilitat	15
5	Consideracions finals	20



Este documento está sujeto a una licencia creative commons que permite su difusión y uso comercial reconociendo siempre la autoría de su creador. Este documento se encuentra para ser modificado en el siguiente repositorio de github: <https://github.com/arvicenteboix/lliurexproxmox>

1 Introducció

En aquesta última unitat, explicarem com muntar un clúster en PROXMOX. És un procediment molt senzill, però s'han de tenir en compte les següents consideracions:

- En muntar un clúster pot tenir diferents màquines entre els hipervisors del clúster.
- Si no muntem un sistema d'alta disponibilitat, no té massa sentit.
- Per a muntar un sistema d'alta disponibilitat cal muntar un CEPH (no s'explica en aquest curs) o tenir una cabina externa.



Les cabines externes són sistemes molt fiables i difícilment fallen, ja que només serveixen dades. Tot el processament es faria al sistema PROXMOX.



Un sistema d'alta disponibilitat permetrà que quan una màquina deixe de funcionar immediatament altre hipervisor se n'adonarà que alguna cosa està passant i la posarà en marxa, sense necessitat que cap persona intervinga.

2 Muntar clúster en PROXMOX

Ja que no disposam de tres ordinadors, es pot muntar un sistema d'alta disponibilitat en Virtualbox per a veure el seu funcionament. La màquina utilitzada és un Ryzen 5 amb 8 GB de RAM i un NAS. S'ha muntat el següent sistema:

Màquina	Característiques
Proxmox 1	Proxmox al Virtualbox
Proxmox 2	Proxmox al Virtualbox
Proxmox 3	Proxmox al Virtualbox
TrueNAS	Muntat a una màquina externa



La quantitat mínima d'hipervisors per a poder fer funcionar un sistema d'alta disponibilitat és de 3. Es necessita per a fer votació i saber qui està viu, qui guanya amb dos vots (ell i altre) està viu, això significa tenir quòrum, qui tinga només un vot està mort (significaria que no té accés a la xarxa). Per a ser justos podem dir que, en realitat podem tindre 2 hipervisors i una màquina que solament vote per a fer quòrum, en eixe cas instal·laríem el corosync en aquesta màquina que podria ser perfectament una raspberry pi, per exemple, però no anem a tractar-ho en este curs.

El primer que cal fer és tenir 3 servidors instal·lats amb PROXMOX. En aquest cas hem muntat els 3 hipervisors amb una única targeta de xarxa. Si es muntara el model de centre tots 3 hipervisors, haurien de tenir la mateixa configuració de xarxa a excepció de la seu IP, obviament.

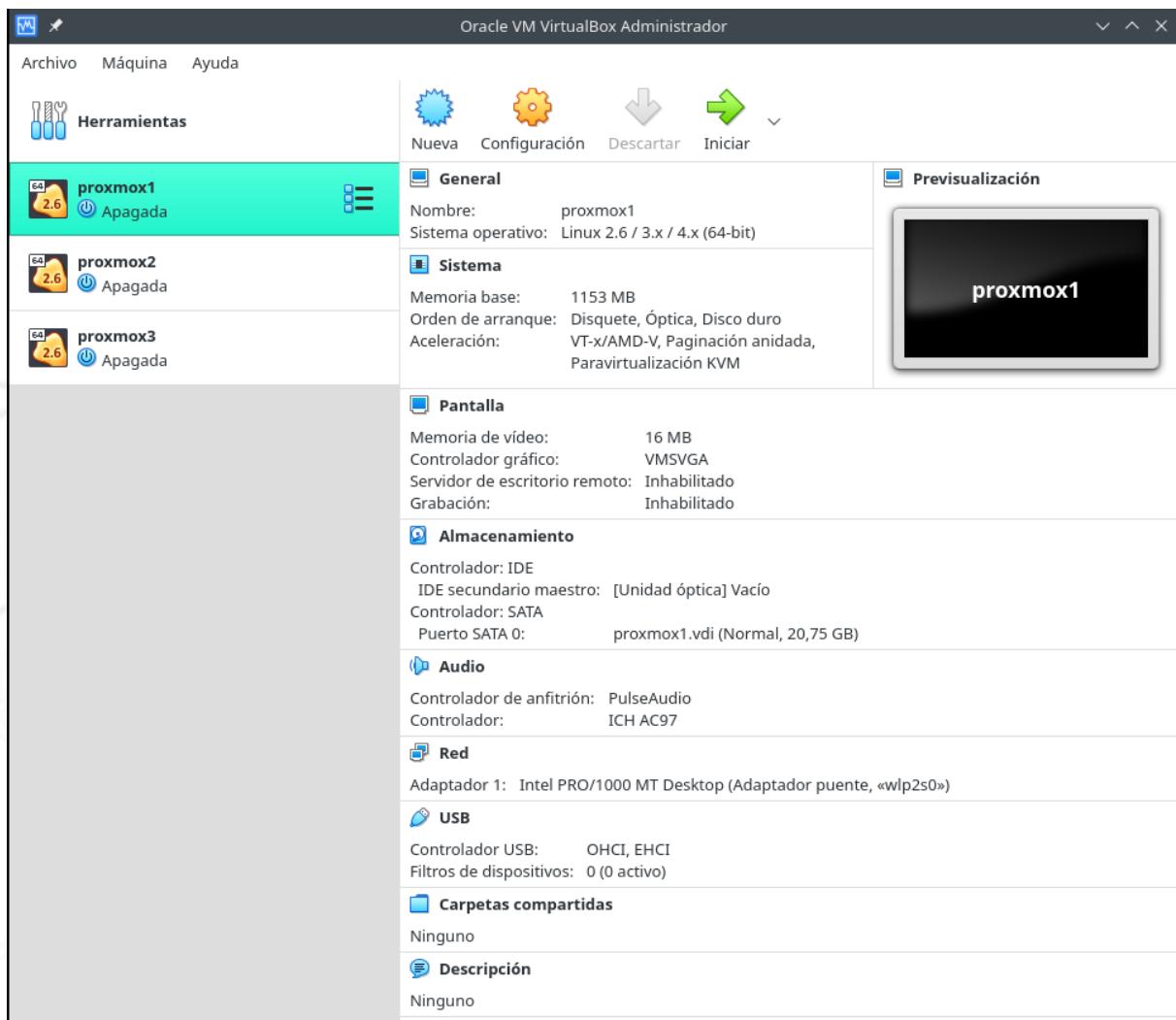


Figura 1: Màquines virtuals al Virtualbox

::tip Recordeu que és necessari que cada PROXMOX tinga un hostname diferent. Si el nom del node coincidix cal canviar-lo. Es deu modificar a l'arxiu /etc/hostname, i a /etc/hosts. :::

Una vegada tenim els 3 hipervisors funcionant. Farem clic sobre el datacenter i anem **Cluster** i clickem a **Create cluster**.

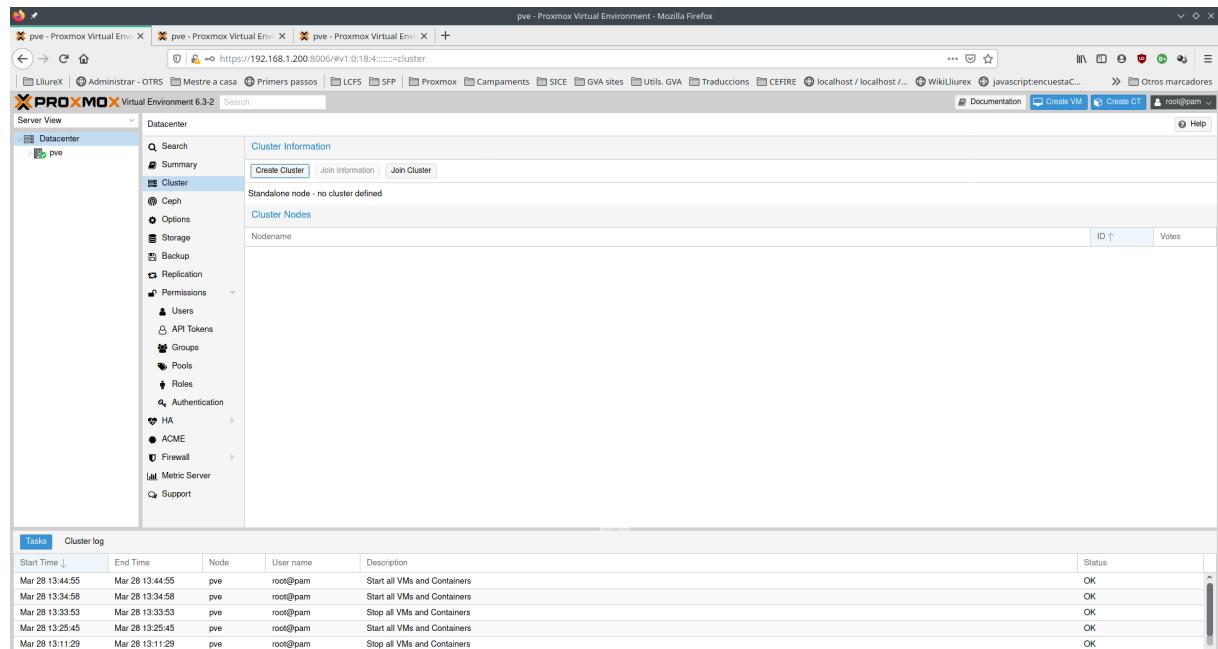


Figura 2: Creació de cluster

Una vegada polsem a **Create cluster** ens apareixerà la següent finestra on li donarem el nom al clúster que vulguem:

Cluster Name:

Cluster Network: Link: 0

Add Multiple links are used as failover, lower numbers have higher priority.

Figura 3: Nom del cluster

Quan posem a **Create** ens apareixerà la següent finestra:

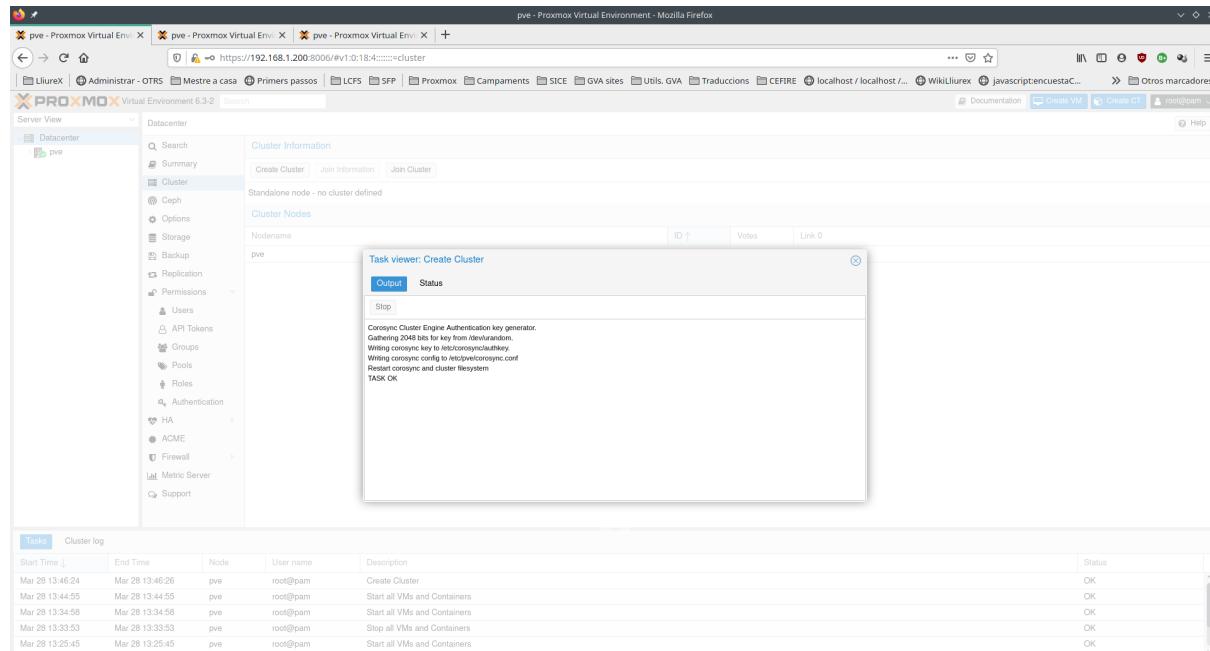
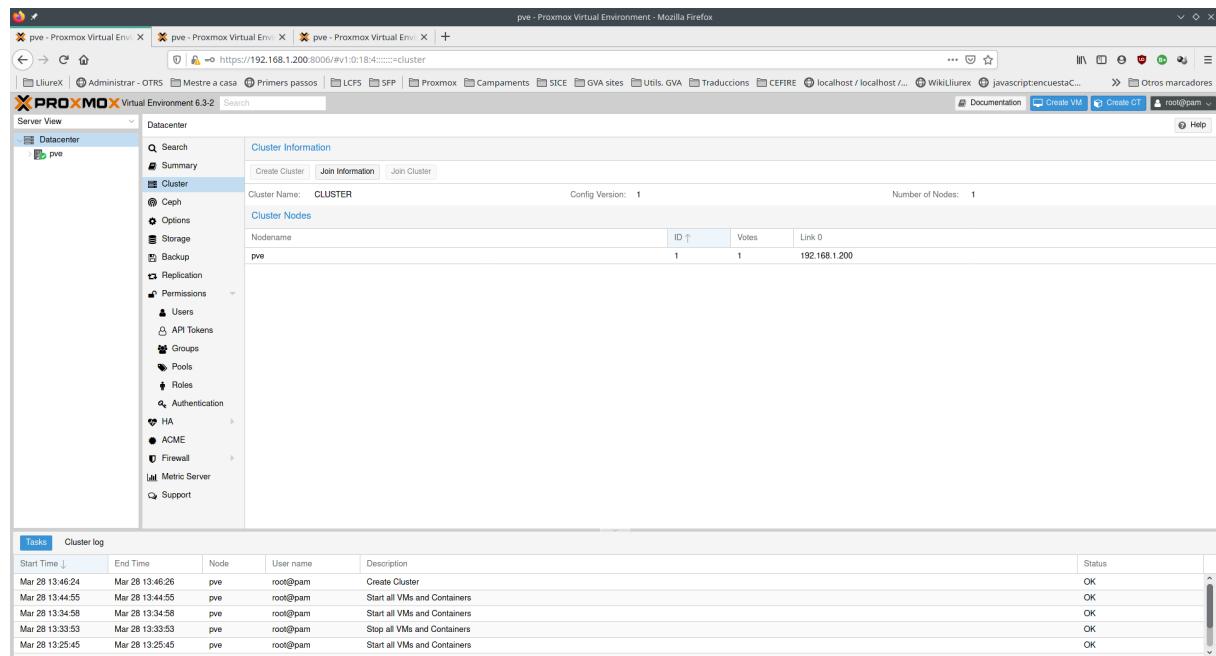
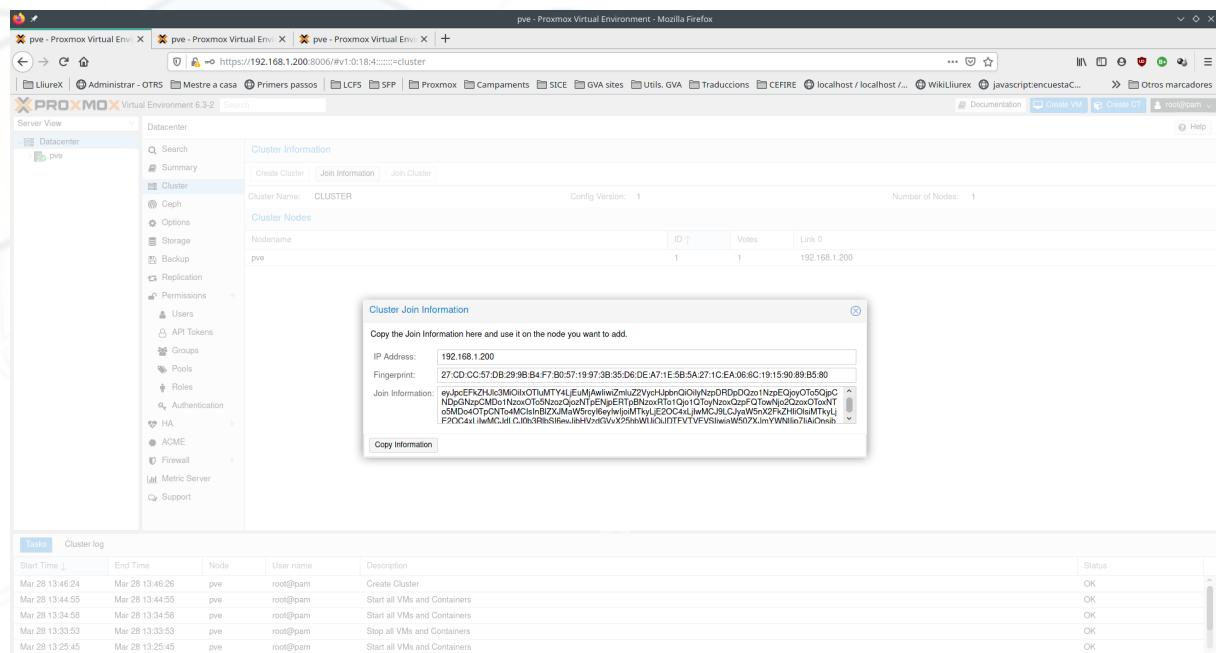


Figura 4: Creació de cluster amb èxit

Com podem veure en aquests moments només tindrem un hipervisor al nostre clúster. Si volem afegir més hipervisors farem clic sobre **Join information**:

**Figura 5:** Membres del cluster

I li donarem a **Copy information**. Aquests són els paràmetres que utilitzarà PROXMOX per a afegir els altres hipervisors al clúster.

**Figura 6:** Informació a passar a altres hipervisors

Canviem d'hipervisor i anem novament a Clúster en el nostre Datacenter:

Start Time	End Time	Node	User name	Description	Status
Mar 28 13:45:16	Mar 28 13:45:16	pve	root@pam	Start all VMs and Containers	OK
Mar 28 13:35:27	Mar 28 13:35:27	pve	root@pam	Start all VMs and Containers	OK
Mar 28 13:33:50	Mar 28 13:33:50	pve	root@pam	Stop all VMs and Containers	OK
Mar 28 13:25:51	Mar 28 13:25:51	pve	root@pam	Start all VMs and Containers	OK
Mar 28 13:25:17	Mar 28 13:25:17	pve	root@pam	Stop all VMs and Containers	OK

Figura 7: Unir-se a un clúster

Polsarem a **Join cluster** i pegarem la informació copiada del primer hipervisor:

Start Time	End Time	Node	User name	Description	Status
Mar 28 15:22:35	Mar 28 15:22:35	pve1	root@pam	Start all VMs and Containers	OK
Mar 28 15:21:59	Mar 28 15:21:59	pve1	root@pam	Stop all VMs and Containers	OK
Mar 28 15:20:51	Mar 28 15:20:51	pve1	root@pam	Join Cluster	Error: Check if node may join cluster
Mar 28 15:19:37	Mar 28 15:19:37	pve1	root@pam	Start all VMs and Containers	OK
Mar 28 15:19:00	Mar 28 15:19:01	pve1	root@pam	Stop all VMs and Containers	OK

Figura 8: Pegar informació

Al clicar sobre **Join cluster** ens apareixerà la següent informació:

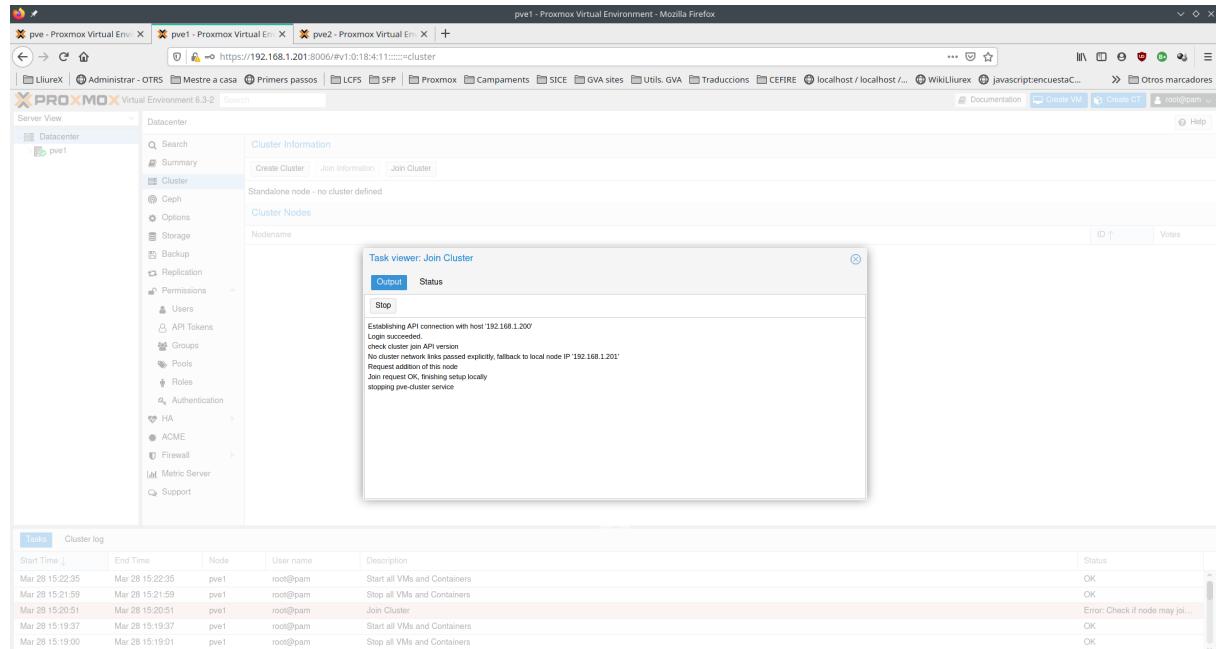


Figura 9: Unió al clúster amb èxit

Aquest procés es repetirà per a altre hipervisor:

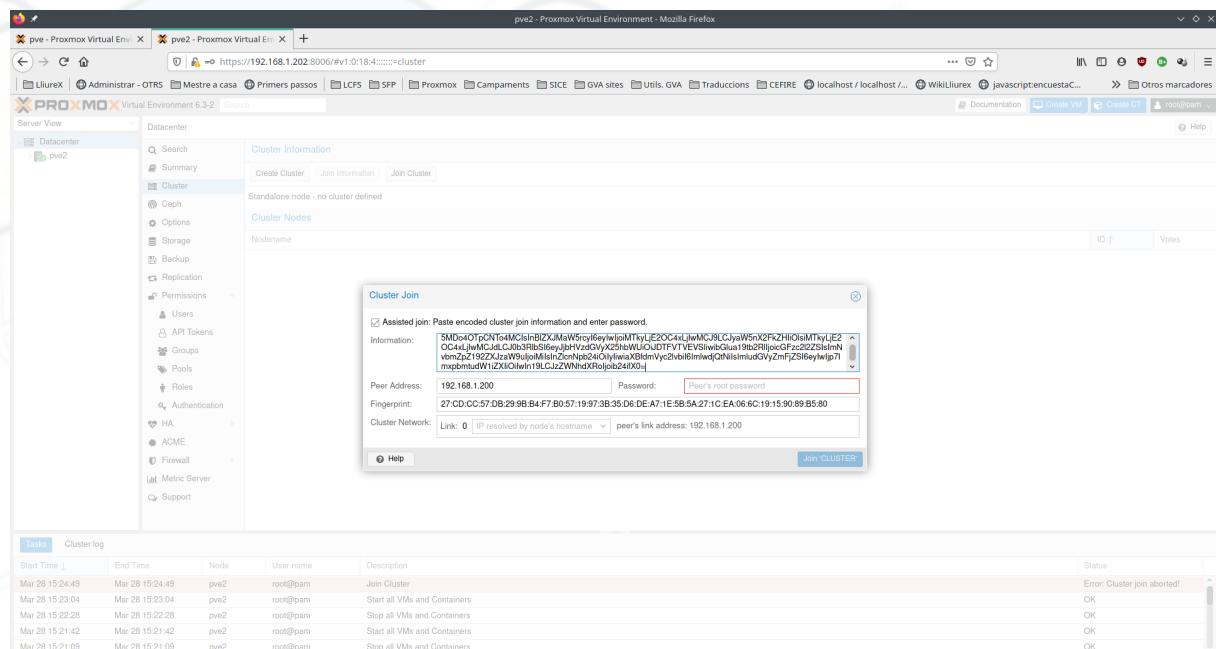
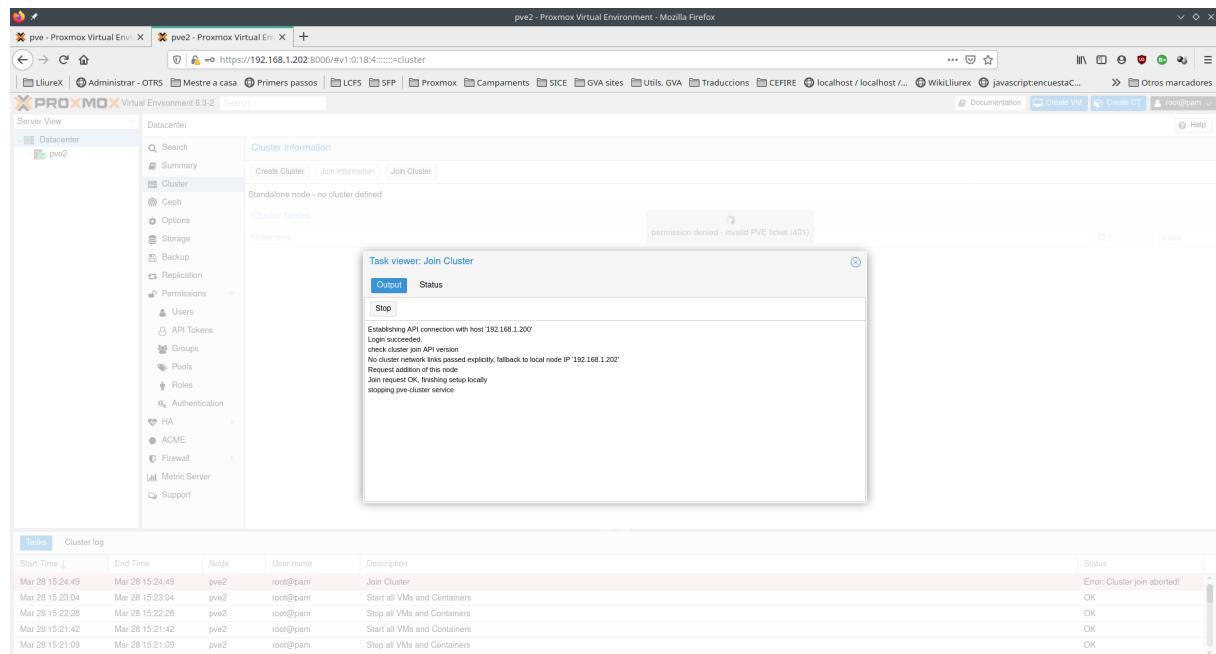
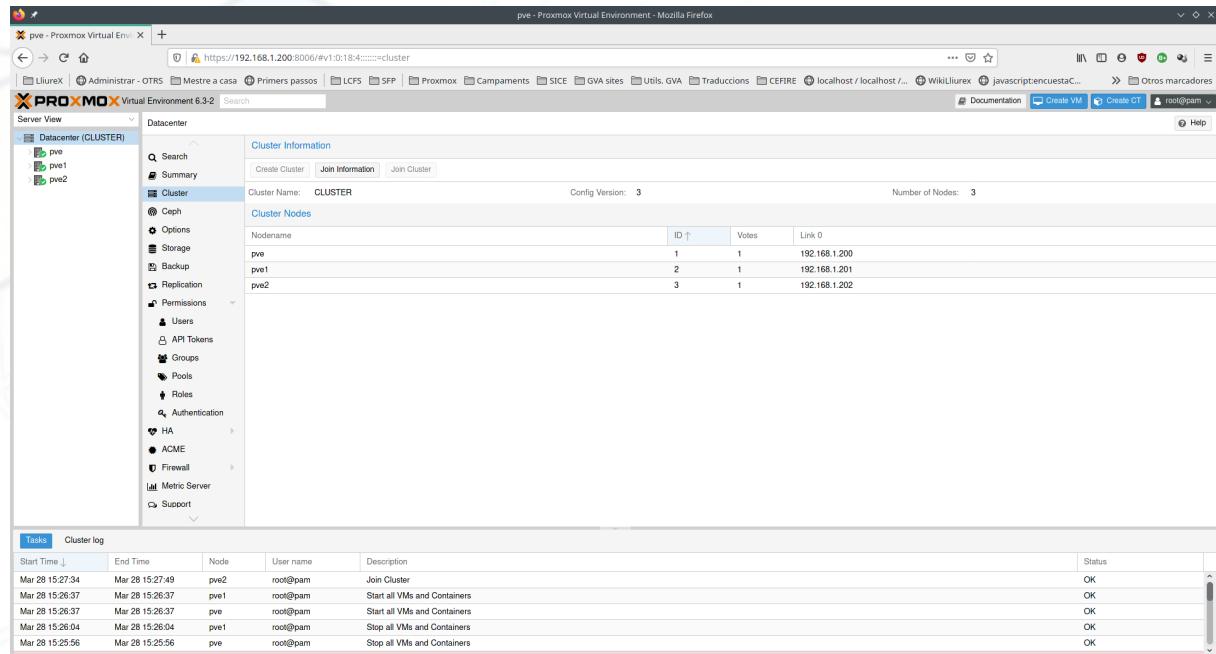


Figura 10: Tercer hipervisor

**Figura 11:** Unió al clúster amb èxit

Ara podrem veure com tenim els tres hipervisors al mateix clúster:

**Figura 12:** Relació d'hipervisors al clúster

En aquests moments ja tenim el nostre clúster muntat.

3 Muntatge d'alta disponibilitat

Per a realitzar el nostre muntatge d'alta disponibilitat utilitzarem una cabina externa. Per a fer les proves s'ha utilitzat la distro Tiny Core, la versió de 16 Mb. Es tracta d'una distro que a penes consumeix recursos i es poden fer proves amb ella en virtualització anidada.



Si es decidix fer el muntatge d'alta disponibilitat amb una cabina, cal disposar d'una cabina amb 4 targetes de xarxa per a muntar un bond, o d'una targeta de 10Gb que obligaria a tenir un switch que suportara 10 Gb.

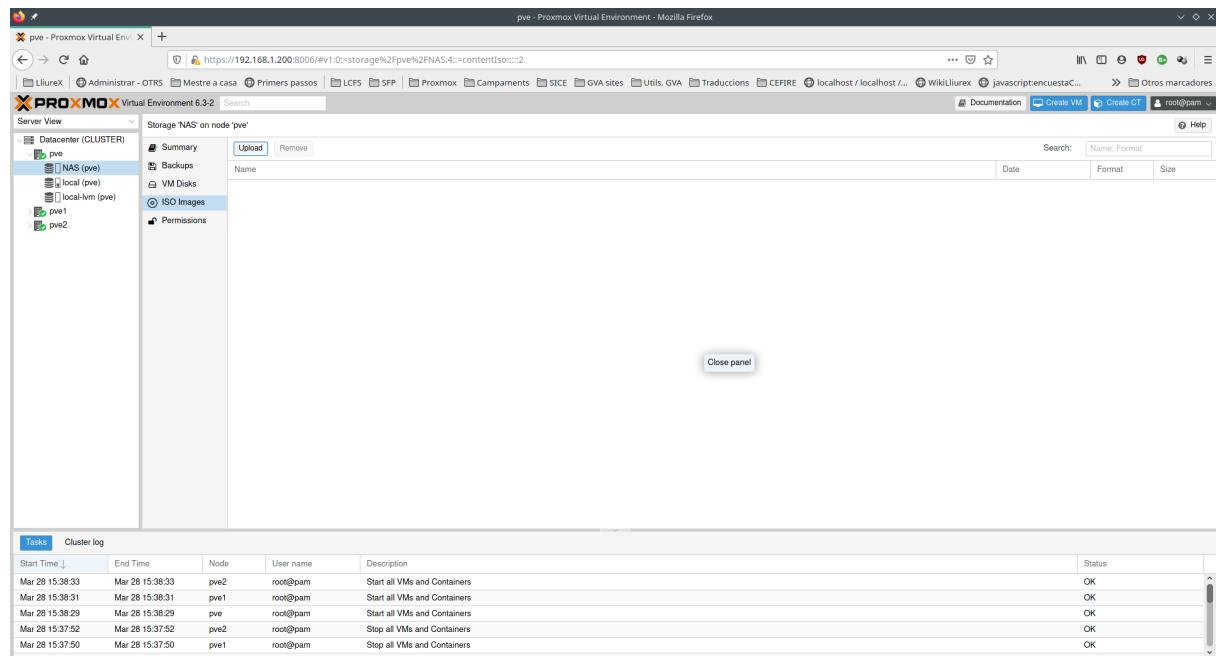
3.1 NAS i màquina virtual

En primer lloc, afegim el nostre NAS al Proxmox. Quan l'afegeim al datacenter serà visible per part de tots els hipervisors.

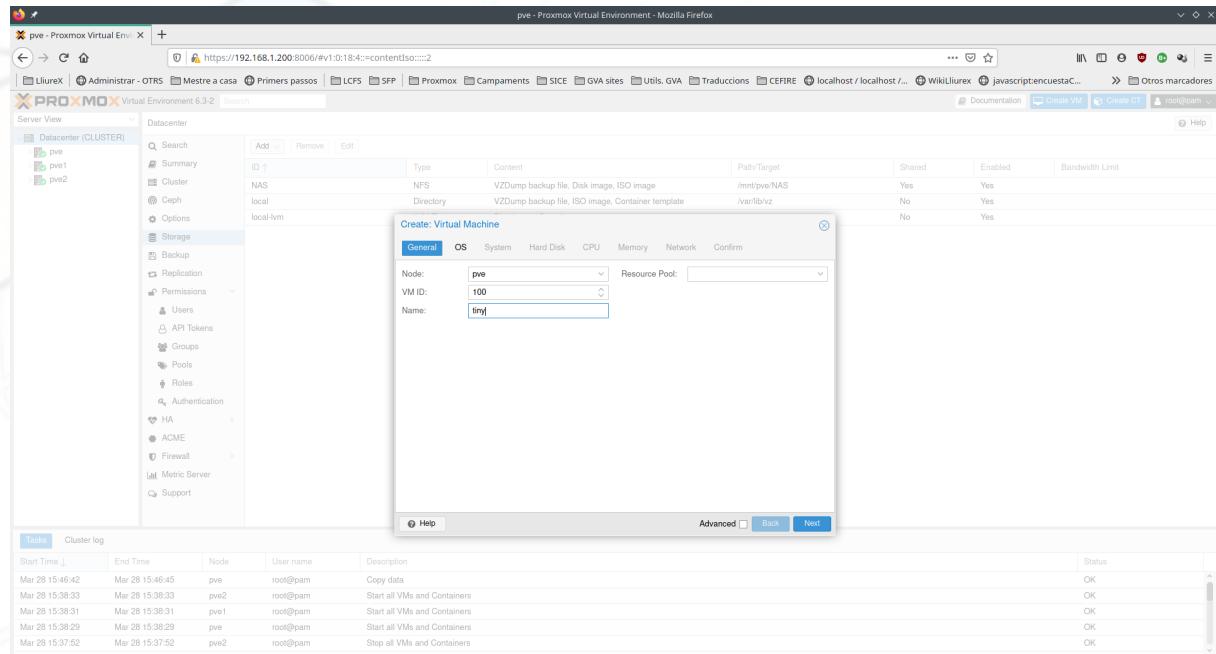
The screenshot shows the Proxmox VE web interface. On the left, the 'Storage' section of the sidebar is selected. In the main area, there's a table of storage volumes. One row is highlighted: 'local-lvm' (Type: LVM-Thin, Content: Disk image, Container). Below the table, a modal dialog titled 'Add: NFS' is open, showing fields for 'ID' (NAS), 'Server' (192.168.1.99), 'Export' (/mnt/disk/proxmoxnas), and 'Content' (Disk image, ISO image). At the bottom of the dialog are 'Advanced' and 'Add' buttons. The background shows a cluster log table with several entries from March 28, 2023, and a status bar at the bottom right indicating 'OK' for all nodes.

Figura 13: Afegeir NAS

Pugem la nostra ISO (no tenim el perquè fer-ho al NAS)

**Figura 14:** Pujar ISO

I creem la nostra màquina virtual:

**Figura 15:** Creació de màquina virtual

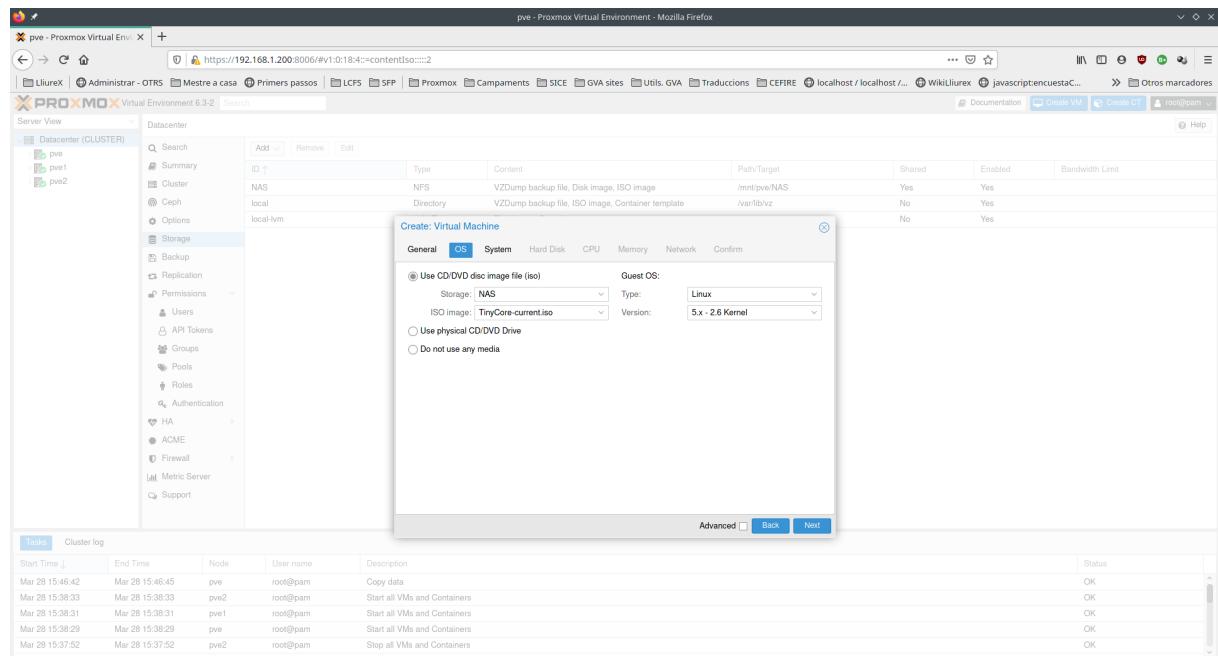


Figura 16: Creació de màquina virtual

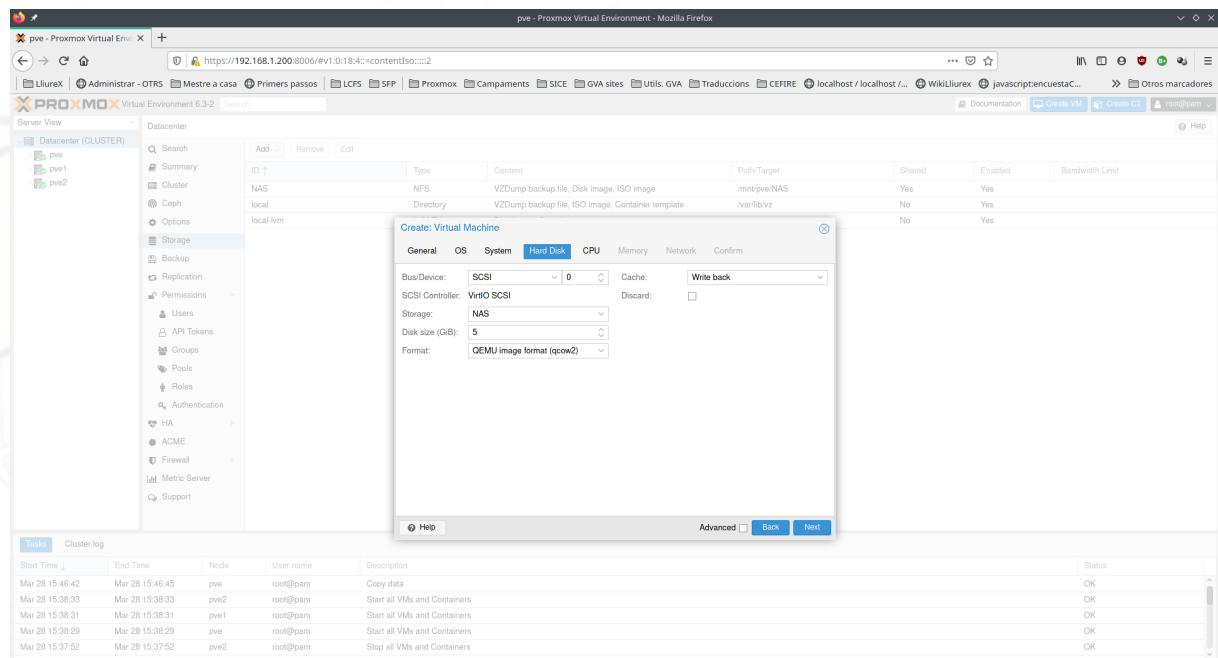


Figura 17: Creació de màquina virtual

No cal donar molta memòria a aquesta màquina:

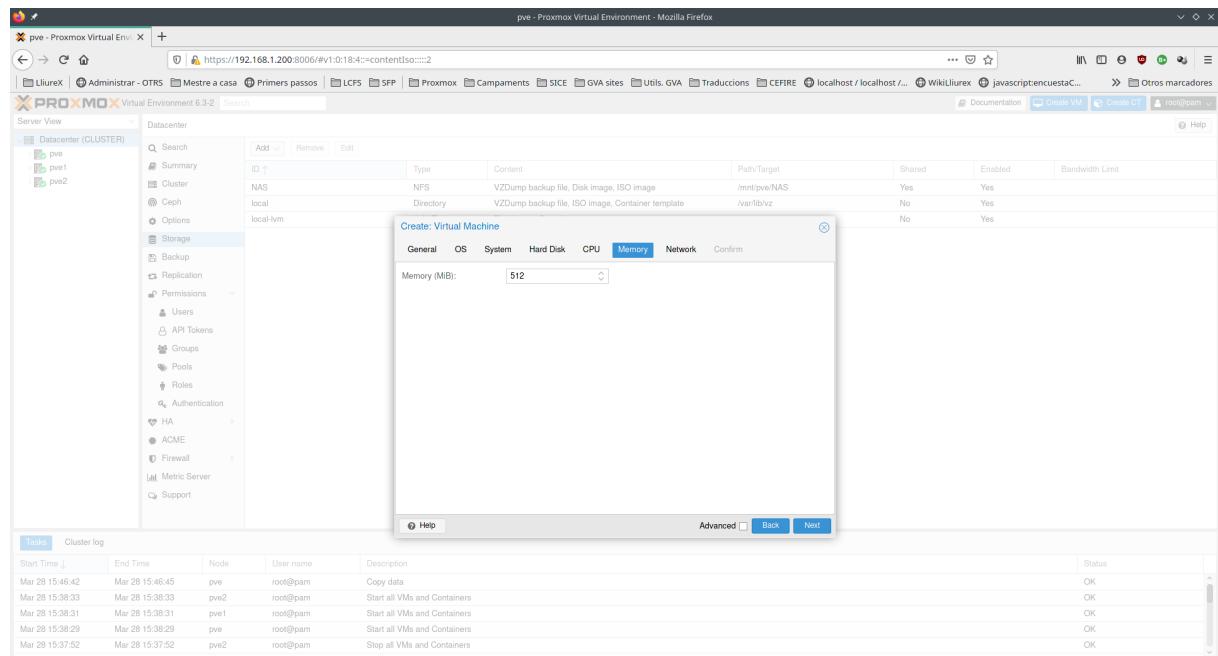


Figura 18: Creació de màquina virtual

I arranquem la màquina:

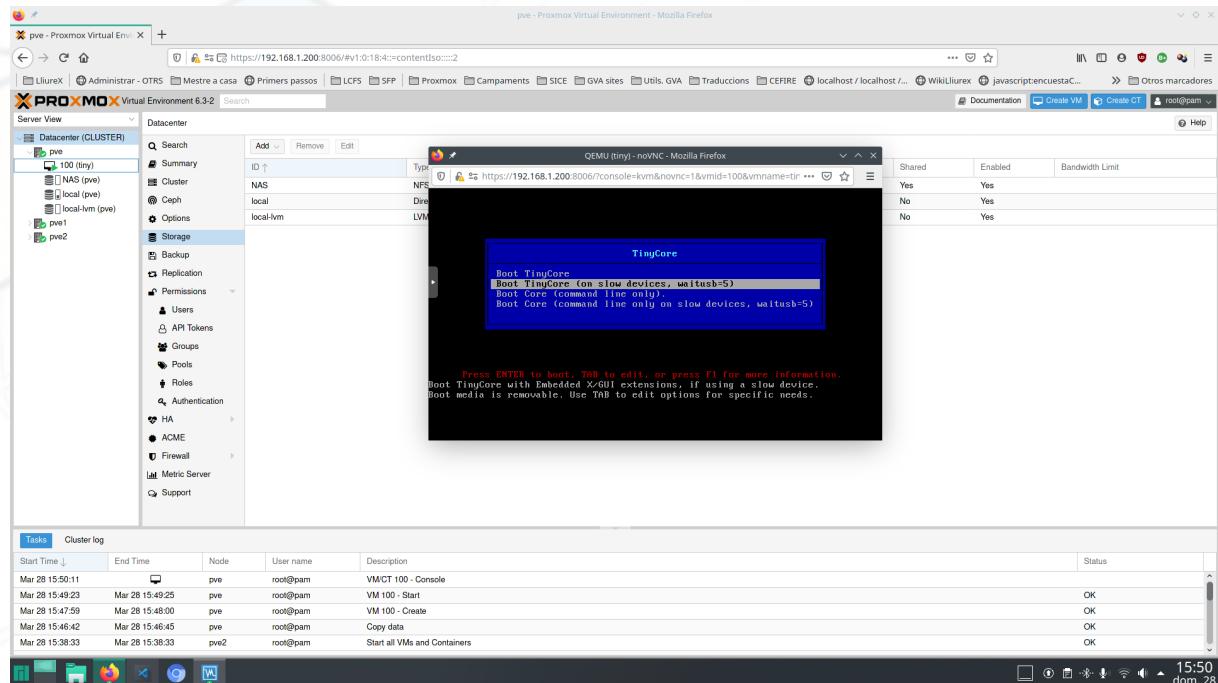


Figura 19: Màquina en funcionament

La màquina arranca de seguida i funciona molt fluida:

4 Alta disponibilitat

Una vegada tenim la màquina funcionant anirem a Datacenter i després HA i clicarem sobre **Add**.

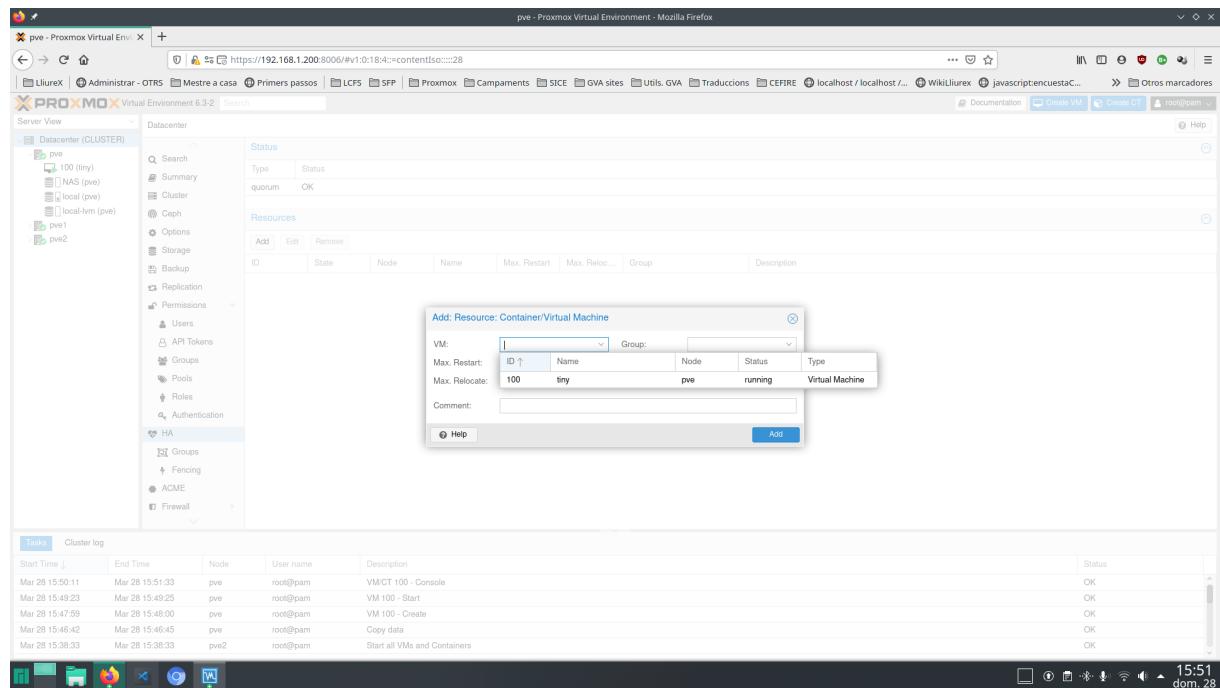
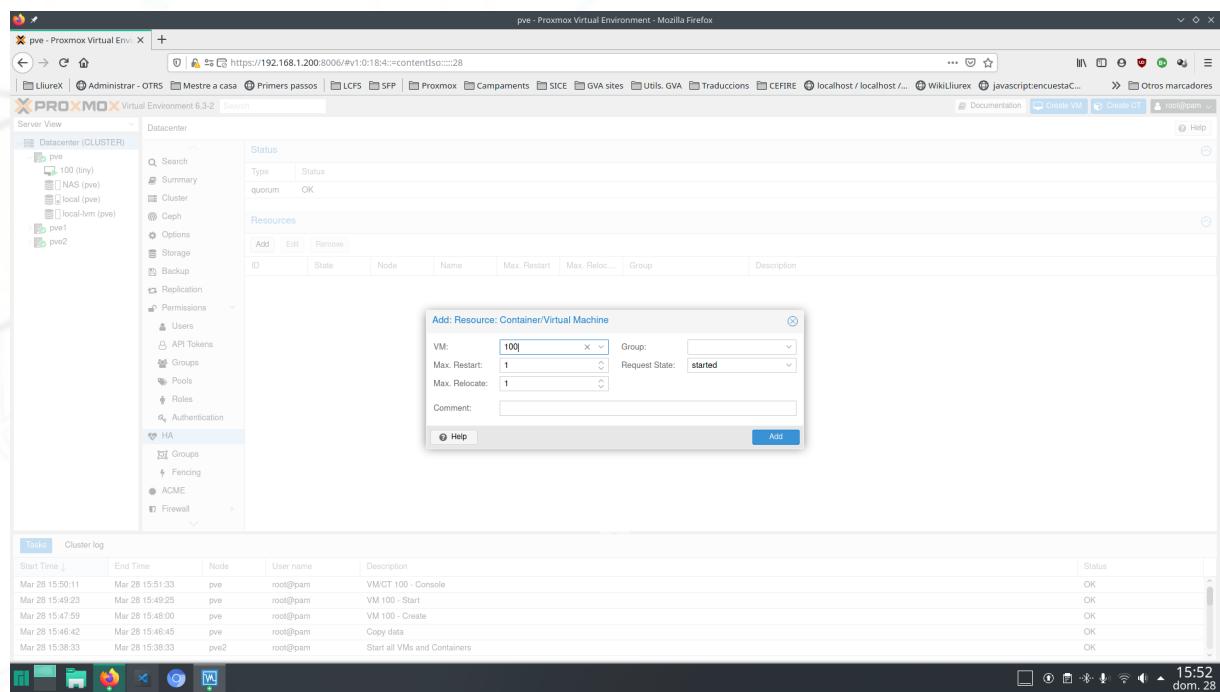
The screenshot shows the PVE web interface with the following details:

- Server View:** Shows the Datacenter (CLUSTER) with nodes pve, pve1, and pve2.
- Datacenter:** Status shows quorum OK. Resources table is empty.
- HA:** Groups, Fencing, ACME, Firewall sections are visible.
- Tasks:** Cluster log table shows the following entries:

Start Time	End Time	Node	User name	Description	Status
Mar 28 15:50:11	Mar 28 15:51:33	pve	root@pam	VMCT 100 - Console	OK
Mar 28 15:49:23	Mar 28 15:49:25	pve	root@pam	VM 100 - Start	OK
Mar 28 15:47:59	Mar 28 15:48:00	pve	root@pam	VM 100 - Create	OK
Mar 28 15:46:42	Mar 28 15:46:45	pve	root@pam	Copy data	OK
Mar 28 15:38:33	Mar 28 15:38:33	pve2	root@pam	Start all VMs and Containers	OK

Figura 20: Alta disponibilitat

Ens apareixerà la següent finestra on haurem d'escollar la màquina que volem tenir funcionant en tot moment:

**Figura 21:** Escollir màquina**Figura 22:** Escollir màquina

En **Request state** cal tenir estat started.

I ja tenim el nostre sistema muntat:

Type	Status
quorum	OK
master	pve (active, Sun Mar 28 15:52:21 2021)
lrm	pve (active, Sun Mar 28 15:52:24 2021)
lrm	pve1 (idle, Sun Mar 28 15:52:24 2021)
lrm	pve2 (idle, Sun Mar 28 15:52:23 2021)

ID	State	Node	Name	Max. Restart	Max. Reloc...	Group	Description
vm:100	starting	pve	tiny	1		1	

Start Time	End Time	Node	User name	Description	Status
Mar 28 15:50:11	Mar 28 15:51:33	pve	root@pam	VM/CT 100 - Console	OK
Mar 28 15:49:23	Mar 28 15:49:25	pve	root@pam	VM 100 - Start	OK
Mar 28 15:47:59	Mar 28 15:48:00	pve	root@pam	VM 100 - Create	OK
Mar 28 15:46:42	Mar 28 15:46:45	pve	root@pam	Copy data	OK
Mar 28 15:38:33	Mar 28 15:38:33	pve2	root@pam	Start all VMs and Containers	OK

Figura 23: Llistat de màquines amb Alta disponibilitat

Amb Virtualbox podem provar d'apagar una màquina per a veure què passa:

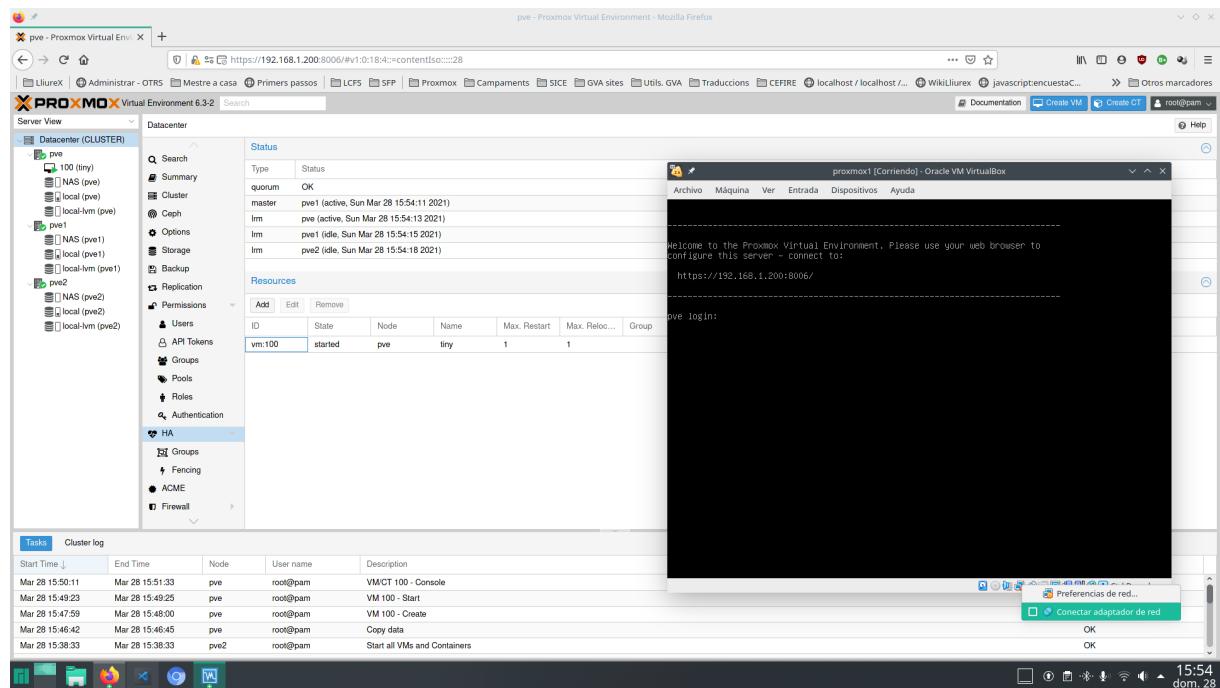


Figura 24: Prova de pèrdua d'un hipervisor

Al cap del temps podem veure com el nostre hipervisor ja no està en marxa i com la màquina ha canviat d'hipervisor i ara està funcionant a l'hipervisor 2:

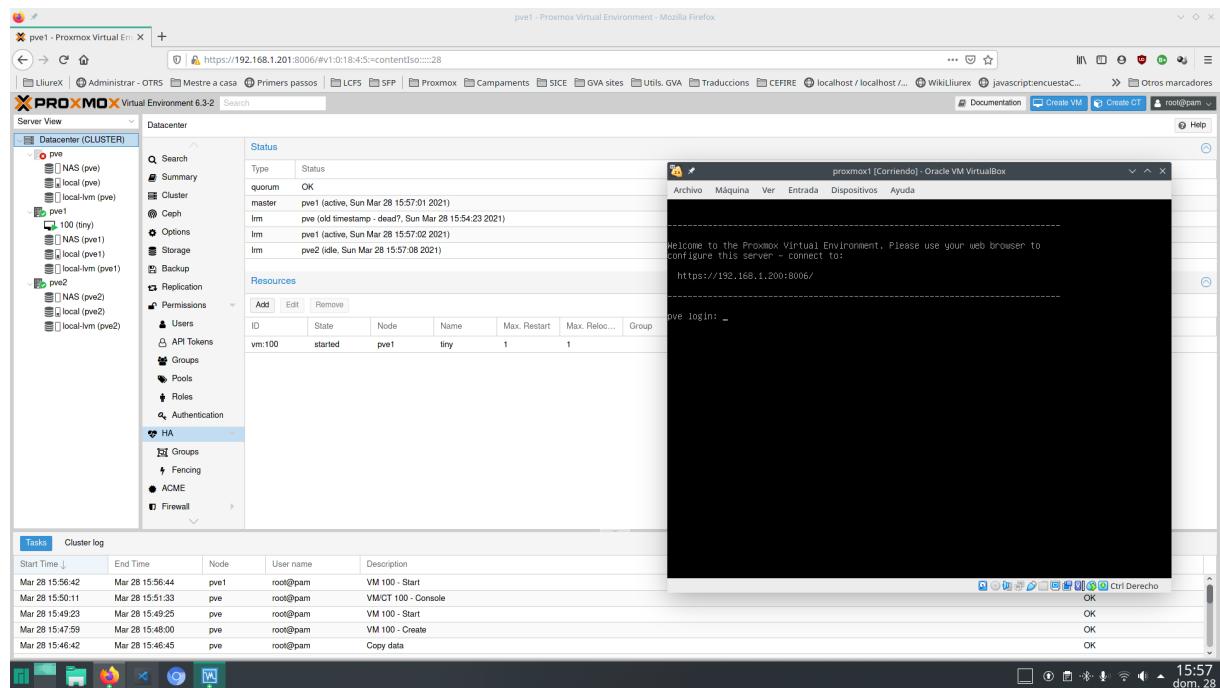


Figura 25: Mànquina que ha passat automàticament a altre hipervisor

Podem veure que la màquina està funcionant:

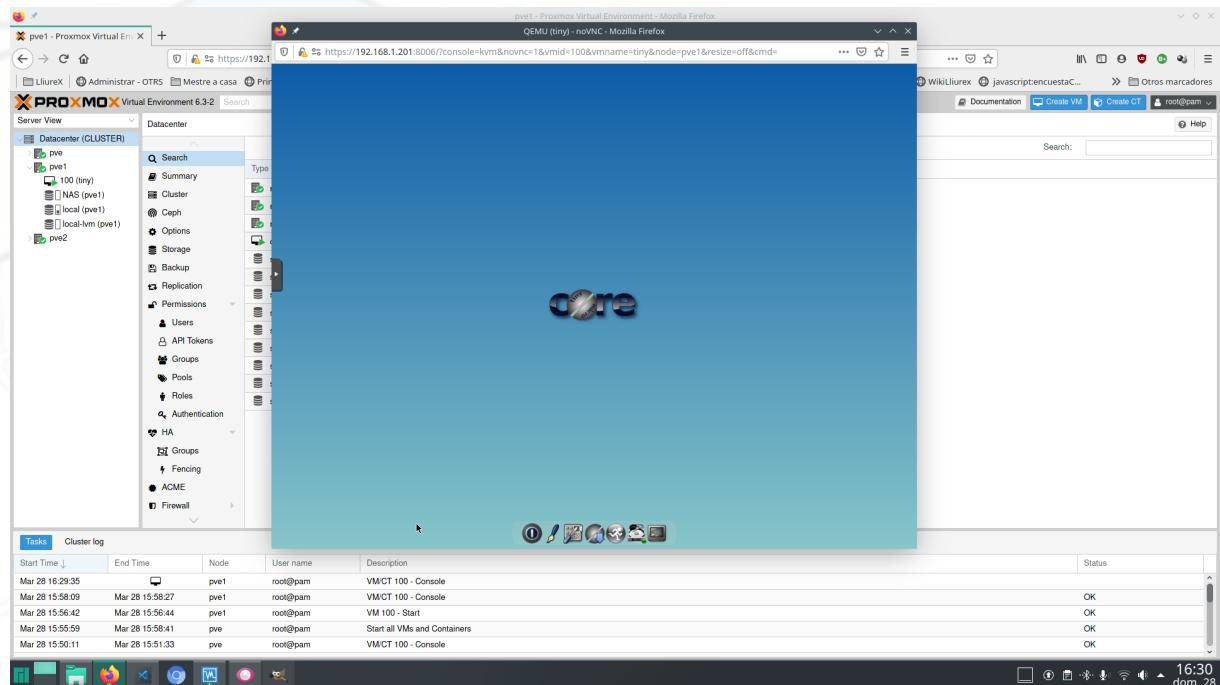


Figura 26: Tiny Core funcionant a altre hipervisor



No espereu que la màquina es pose en marxa de seguida, el procés pot trigar uns 5 minuts i si hi han clients semilleugers funcionant, probablement deixaran de funcionar correctament. En aquest procés pot haver-hi certa pèrdua d'informació ja que no es recupera l'estat de la màquina.

5 Consideracions finals

Es pot tractar de tenir un sistema on hi haja una mínima pèrdua d'informació amb dos hipervisors i si no disposes de NAS. Es podrien tenir dues màquines duplicades amb dos hipervisors i utilitzar la replicació.

Es podria tenir el /net a un disc virtual a banda i que aquest anara replicant-se cada 30 min entre els dos hipervisors, si un d'ells caiguera. Es podria posar en marxa l'altra màquina a l'altre hipervisor i, com a molt, s'haurien perdut 30 minuts d'informació.

Podrien configurar-ho dins de la màquina virtual a l'apartat de **Replication**.

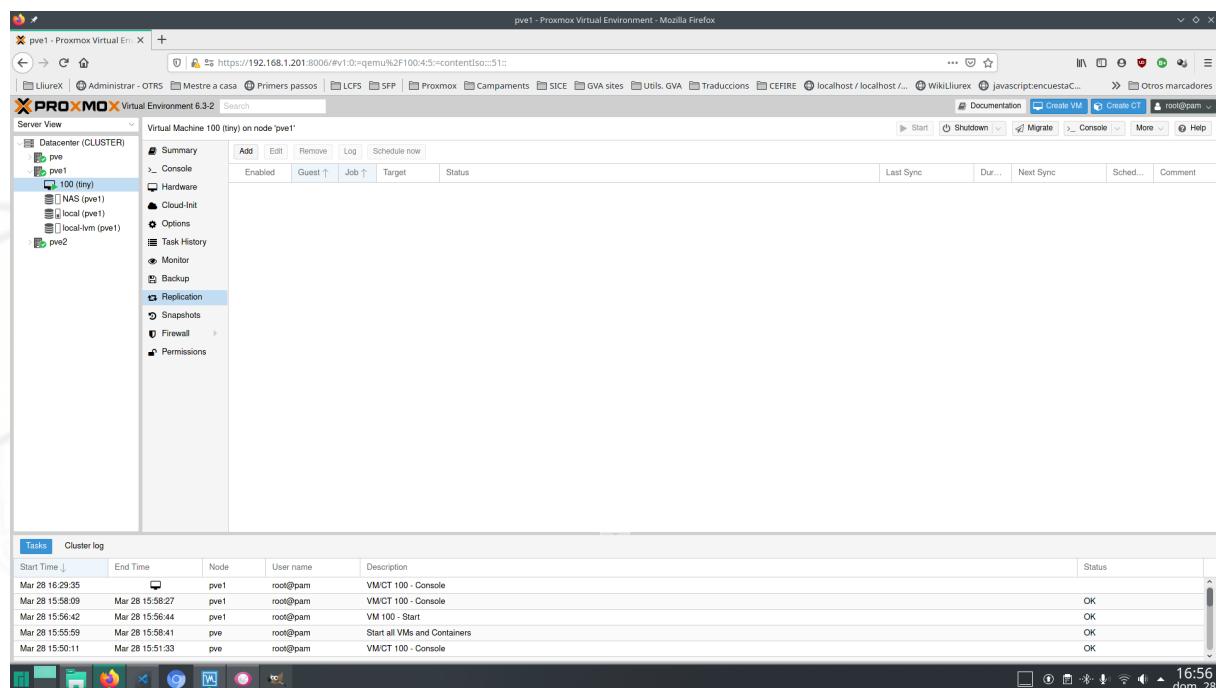


Figura 27: Replicació de discs

Es podrien fer moltes més coses, ja que PROXMOX permet automatitzar moltes tasques i tenir un munt de serveis. Amb aquest curs hem tractat de donar-vos unes pinzellades sobre el que es pot fer i com es podria utilitzat als centres educatius.