



High Availability and CRS



About me



- Mark Schouten, 45 years old
- Founder & CEO Tuxis
- Linux since 2001
- Proxmox Trainer since 2022



About me

- Tuxis
- Founded in 2009
- Proxmox since 2014



High Availability & CRS

- What it is:
 - Automatic recovery
 - Maintenance convenience
 - Cluster Balancing
 - Affinity Management
- What it is NOT:
 - Zero downtime
 - Service loadbalancing
 - VMWare's Fault Tolerance feature

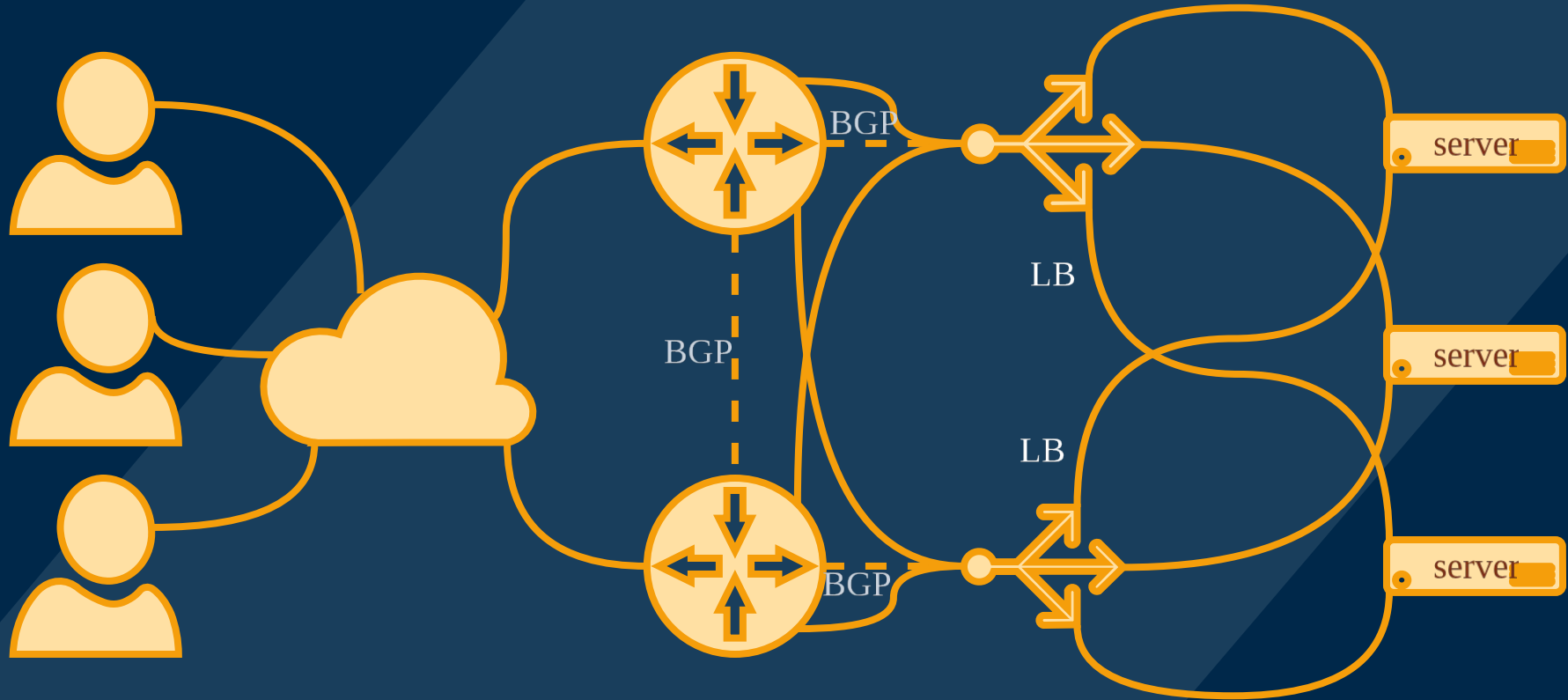


Determine your amount of 9s

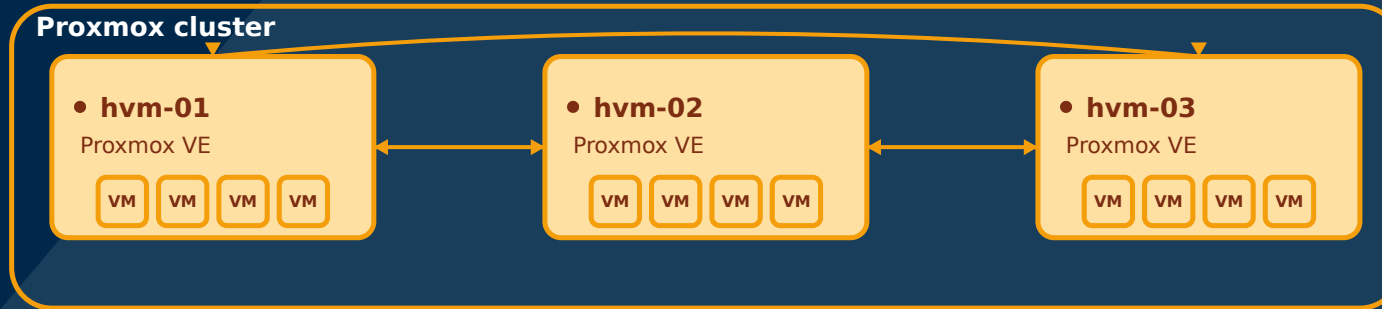
Availability (%)	Accepted downtime per year
99	3,65 days
99,9	8,76 hours
99,99	52,56 minutes
99,999	5,25 minutes
<u>Proxmox HA Cutoff Point</u>	<u>~ 2 minutes</u>
99,9999	31,5 seconds
99,99999	3,15 seconds



Need more 9's?



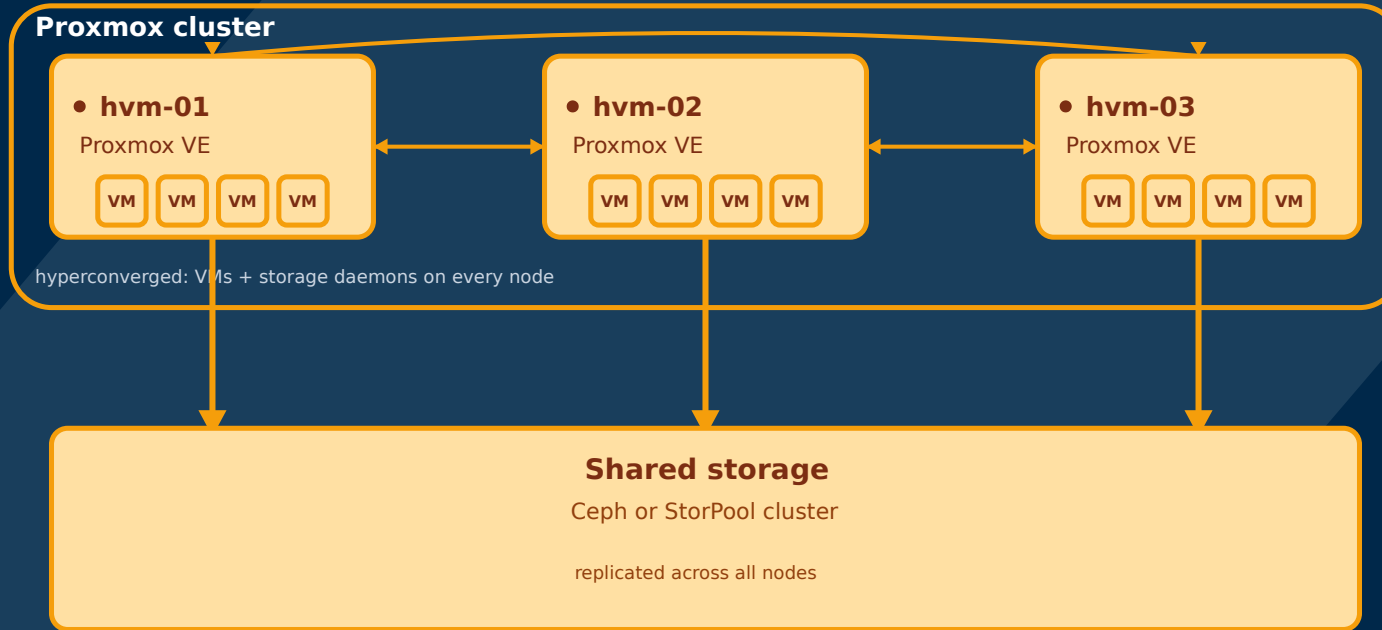
What do we need for HA?



Generated using Tuxis LLM



What do we need for HA?



Generated using Tuxis LLM



Configuring a HA resource

Virtual Machine 138 (pbs-portal) on node 'node01' `prtl_depends`

[Start](#) [Shutdown](#) [Migrate](#) [Console](#) [More](#) [Help](#)

[Summary](#) [Console](#) [Hardware](#) [Cloud-Init](#)

Week

[pbs-portal \(Uptime: 20 days 00:54:48\)](#)

[Notes](#)

- Clone
- Convert to template
- Manage HA**
- Remove



Configuring a HA resource

Edit: Resource: Container/Virtual Machine

VM: 138

Max. Restart: 1

Max. Relocate: 1

Comment: Quick and easy here!

Failback: ☒

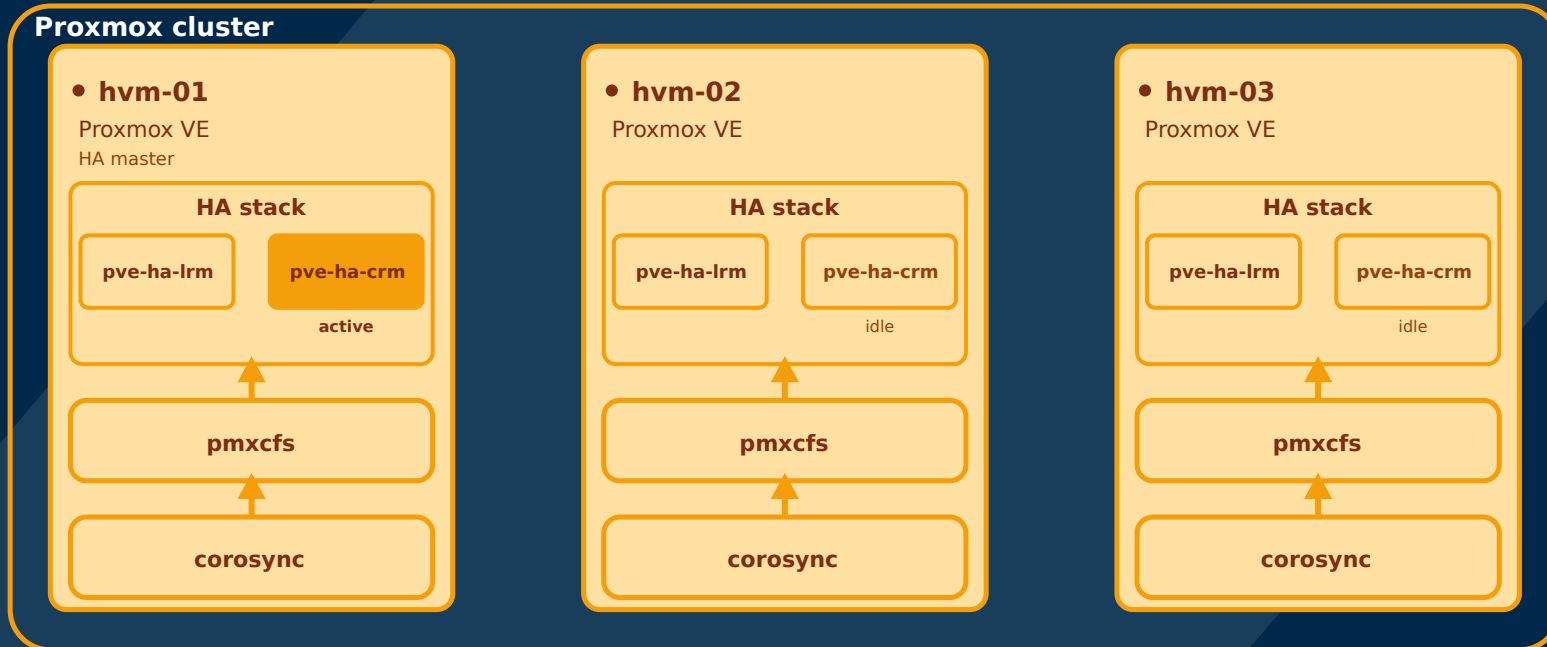
Auto-Rebalance: ☒

Request State: started

Help OK



Background operations



Generated using Tuxis LLM





Disarming & Status

PROXMOX Virtual Environment 9.2.11

Folder View 

Datacenter

Left Sidebar:

- Datacenter (Demo)
 - LXC Container
 - Network
 - Nodes
 - Resource Pool
 - Virtual Machine
 - Storage
- Search
- Summary
- Notes
- Cluster
- Ceph
- Options
- Storage
- Backup
- Replication
- Permissions
 - Users
 - API Tokens
 - Two Factor
 - Groups
 - Pools
 - Roles
 - Realms
- HA**
 - Affinity Rules
 - Fencing

HA Controls:  Arm HA  Disarm HA

Status

Type: Status

quorum: OK

fencing: armed (CRM watchdog active)

master: node01 (active, Tue Sep 15 20:53:44 2026) - dynamic load CRS (load imbalance: 2.4%)

lrm: node01 (active, watchdog active, Tue Sep 15 20:53:46 2026)

lrm: node02 (active, watchdog active, Tue Sep 15 20:53:50 2026)

lrm: node03 (active, watchdog active, Tue Sep 15 20:53:52 2026)

Resources

 Add  Edit  Remove

ID	State	Node	Name ↓	Max. Restart
vm:100026	started	node02	william-ansible-cynocsure-bak	1
vm:100027	started	node02	william-ansible-client-doo-william	1
vm:100070	started	node02	vps100070.ede01.customers.tuxis.net	1
vm:100069	started	node02	vps100069.ede01.customers.tuxis.net	1
vm:100067	started	node02	vps100067.ede01.customers.tuxis.net	1
vm:100066	started	node02	vps100066.ede01.customers.tuxis.net	1
vm:100064	started	node02	vps100064.ede01.customers.tuxis.net	1
vm:100063	started	node03	vps100063.ede01.customers.tuxis.net	1



HA Resource Options

true
true
true
true
true
true
true
true
true
true
true
true
true
true
true
true
true
true
true
true
true
true

Edit: Resource: Container/Virtual Machine

VM: 100026

Max. Restart: 1

Max. Relocate: 1

Comment:

Failback: ☒

Auto-Rebalance: ☒

Request State: started

Help

started

stopped

ignored

disabled



HA Affinity Options

- (Anti-)Affinity
 - Node based
 - Resource based

e.g. Keep your database and application server grouped on specific nodes, but keep database servers from running on the same node



HA Affinity Options

Add: HA Resource Affinity ⓧ

Enable: ☒ Affinity: ⌵

HA Resources: ⌵

Comment:

ⓘ Help Add

- Keep Together (positive)
- Keep Separate (negative)



HA Affinity Options

Add: HA Node Affinity

Enable: ☒

Strict: ☐

HA Resources:

Affinity:

Prefer Nodes (positive)

Prefer Nodes (positive)

Avoid Nodes (negative)

Comment:

☐	Node ↑	Memory usage %	CPU usage	Priority
☐	node01	43.8 %	5.2% of 64 CPUs	
☐	node02	47.1 %	5.2% of 64 CPUs	
☐	node03	49.6 %	3.9% of 64 CPUs	

?

 Help

Add

Mark Schouten / Proxday 2026

Affinity Example

app01 / app02 / app03 — all on different nodes

db01 / db02 — different nodes

Proxmox cluster

• hvm-01

Proxmox VE

db01

web01

app01

• hvm-02

Proxmox VE

db02

web02

app02

• hvm-03

Proxmox VE

app03

— affinity — must stay on the same node

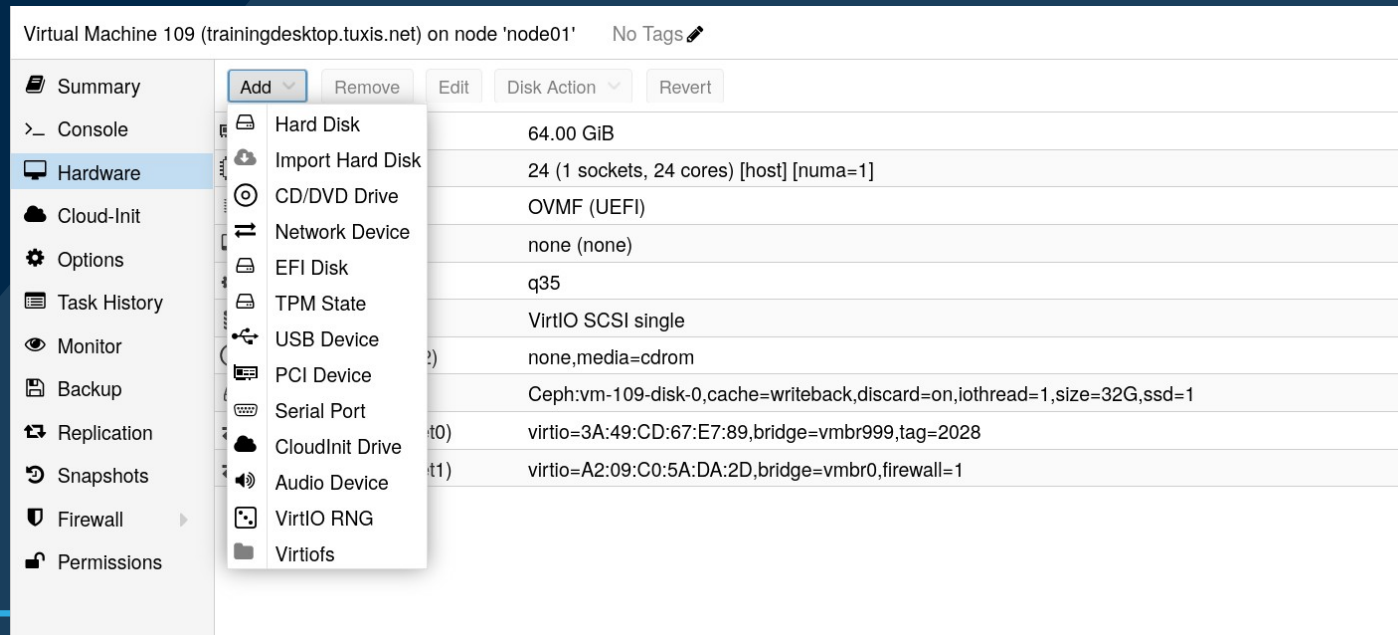
- - - anti-affinity — must not share a node

Generated using Tuxis LLM



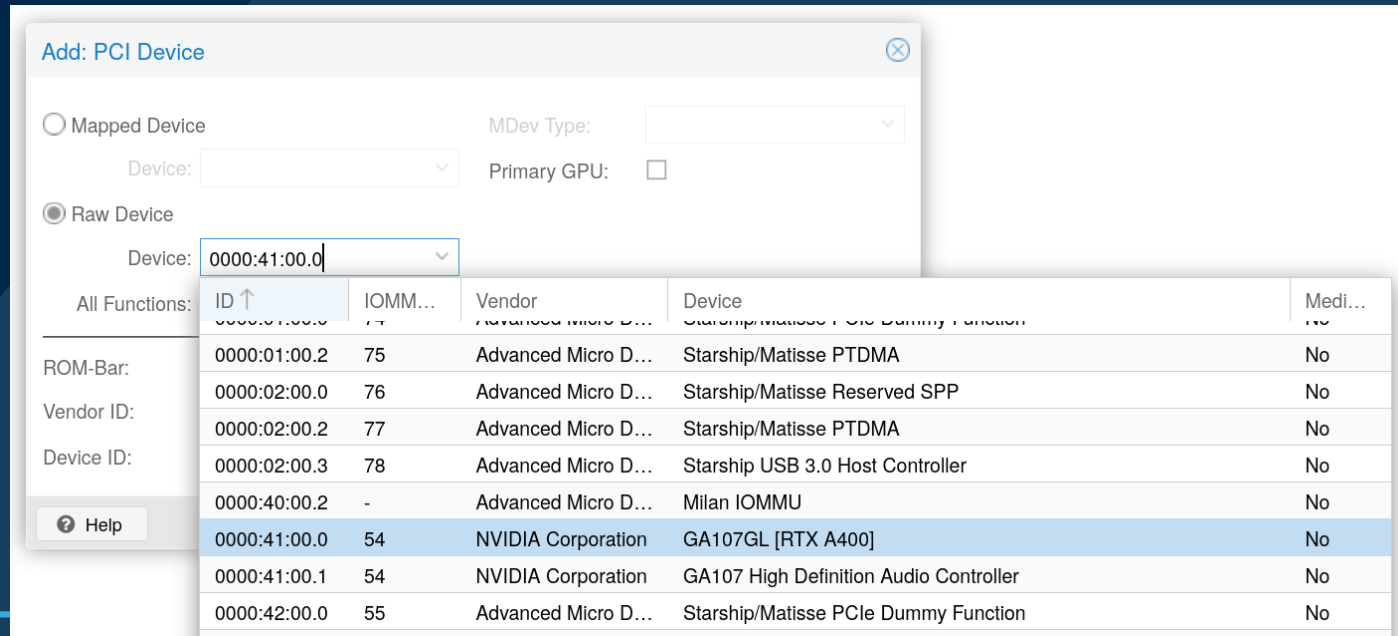
Another affinity usecase

- Proxmox Supports PCI(e) Passthrough



Another affinity usecase

- Hardware stuck to a single hypervisor



Another affinity usecase

- Or, Mapped!

The screenshot shows a dialog box titled "Add: PCI Device". The "Mapped Device" radio button is selected. The "Device" dropdown menu is open, showing a list with the following columns: ID ↑, Description, and Status. The first item in the list is "RTX_A400" with the status "Mapping matches host data".

Below the dropdown, the "Raw Device" radio button is unselected. The "All Functions" checkbox is unselected.

At the bottom, there are several fields and checkboxes:

- ROM-Bar: ☒
- PCI-Express: ☐
- Vendor ID: From Device
- Sub-Vendor ID: From Device
- Device ID: From Device
- Sub-Device ID: From Device

At the bottom right, there is an "Advanced" checkbox (checked) and an "Add" button.



Mapping devices

- Go to Datacenter → Resource mappings → Add

Create: PCI mapping

Name:

Mapping on Node:

Use with Mediated Devices: ☐

Live Migration Capable: ☐ Experimental

☐	ID ↑	IO...	Vendor	Device	Me...
☐	0000:41:00	54	NVIDIA Corporation	Pass through all functions as one device	No
☐	0000:41:00.0	54	NVIDIA Corporation	GA107GL [RTX A400]	No
☐	0000:41:00.1	54	NVIDIA Corporation	GA107 High Definition Audio Controller	No
☒	0000:42:00	55	Advanced Micro Devic...	Pass through all functions as one device	No
☐	0000:42:00.0	55	Advanced Micro Devic...	Starship/Matisse PCIe Dummy Function	No
☐	0000:42:00.2	56	Advanced Micro Devic...	Starship/Matisse PTDMA	No
☐	0000:43:00	57	Advanced Micro Devic...	Pass through all functions as one device	No
☐	0000:43:00.0	57	Advanced Micro Devic...	Starship/Matisse Reserved SPP	No
☐	0000:43:00.1	58	Advanced Micro Devic...	Starship/Matisse Cryptographic Coprocessor PSPC...	No
☐	0000:43:00.2	59	Advanced Micro Devic...	Starship/Matisse PTDMA	No

Comment



Mapping devices

- Expand this mapping

Create: PCI mapping

Name: Dummy Use with Mediated Devices: ☐

Mapping on Node: node02 Live Migration Capable: ☐ Experimental

☐	ID ↑	IO...	Vendor	Device	Me...
☐	0000:41:00.1	54	NVIDIA Corporation	GA107 High Definition Audio Controller	No
☒	0000:42:00	55	Advanced Micro Devic...	Pass through all functions as one device	No
☐	0000:42:00.0	55	Advanced Micro Devic...	Starship/Matisse PCIe Dummy Function	No
☐	0000:42:00.2	56	Advanced Micro Devic...	Starship/Matisse PTDMA	No
☐	0000:43:00	57	Advanced Micro Devic...	Pass through all functions as one device	No
☐	0000:43:00.0	57	Advanced Micro Devic...	Starship/Matisse Reserved SPP	No
☐	0000:43:00.1	58	Advanced Micro Devic...	Starship/Matisse Cryptographic Coprocessor PSPC...	No
☐	0000:43:00.2	59	Advanced Micro Devic...	Starship/Matisse PTDMA	No
☐	0000:43:00.3	60	Advanced Micro Devic...	Starship USB 3.0 Host Controller	No
☐	0000:44:00.0	61	Advanced Micro Devic...	FCH SATA Controller [AHCI mode]	No



Mapping devices

- Add your mapped device to your VM

Device (net0) virtio=3A:49:CD:67:E7:89,bridge=vibr999,tag=2028

Device (net1) virtio=A2:09:C0:5A:DA:2D,bridge=vibr0,firewall=1

Add: PCI Device

☒ Mapped Device

MDev Type:

Device:

Primary GPU: ☐

☐ Raw Device

Device:

All Functions:

ID ↑

Description

Status

Dummy

✓ Mapping matches host data

RTX_A400

✓ Mapping matches host data

ROM-Bar: ☒

PCI-Express: ☐

Vendor ID:

From Device

Sub-Vendor ID:

From Device

Device ID:

From Device

Sub-Device ID:

From Device

Help

Advanced ☒

Add



Lock down your resources

- Using Node affinity

A Resource Affinity Rules

Add Edit Remove

Enabled State Affinity

✓ ✓ neg

Add: HA Node Affinity

Enable: ☒

Strict: ☒

HA Resources: 109

Affinity: Prefer Nodes (positive)

Comment: Dummy passthrough

☐	Node ↑	Memory usage %	CPU usage	Priority
☒	node01	41.5 %	6.6% of 64 CPUs	⬆️⬆️
☒	node02	54.2 %	6.7% of 64 CPUs	⬆️⬆️
☐	node03	53.8 %	3.8% of 64 CPUs	⬆️⬆️

? Help

Add



Maintenance mode

- New in HA, maintenance mode
 - `ha-manager crm-command node-maintenance enable $hostname`
- Tries to empty your node, ONLY looks at HA Resources!



Maintenance mode

- We still prefer pmmaint ;)

PROXMOX Virtual Environment 9.2.11

Folder View

Datacenter (Demo)

- LXC Container
- Network
- Nodes
 - node01
 - node02
 - node03
- Resource Pool
- Virtual Machine
- Storage

Datacenter

Search

Summary

Notes

Cluster

Ceph

Options

Storage

Backup

Replication

Permissions

Users

Arm HA

Disarm HA

Status

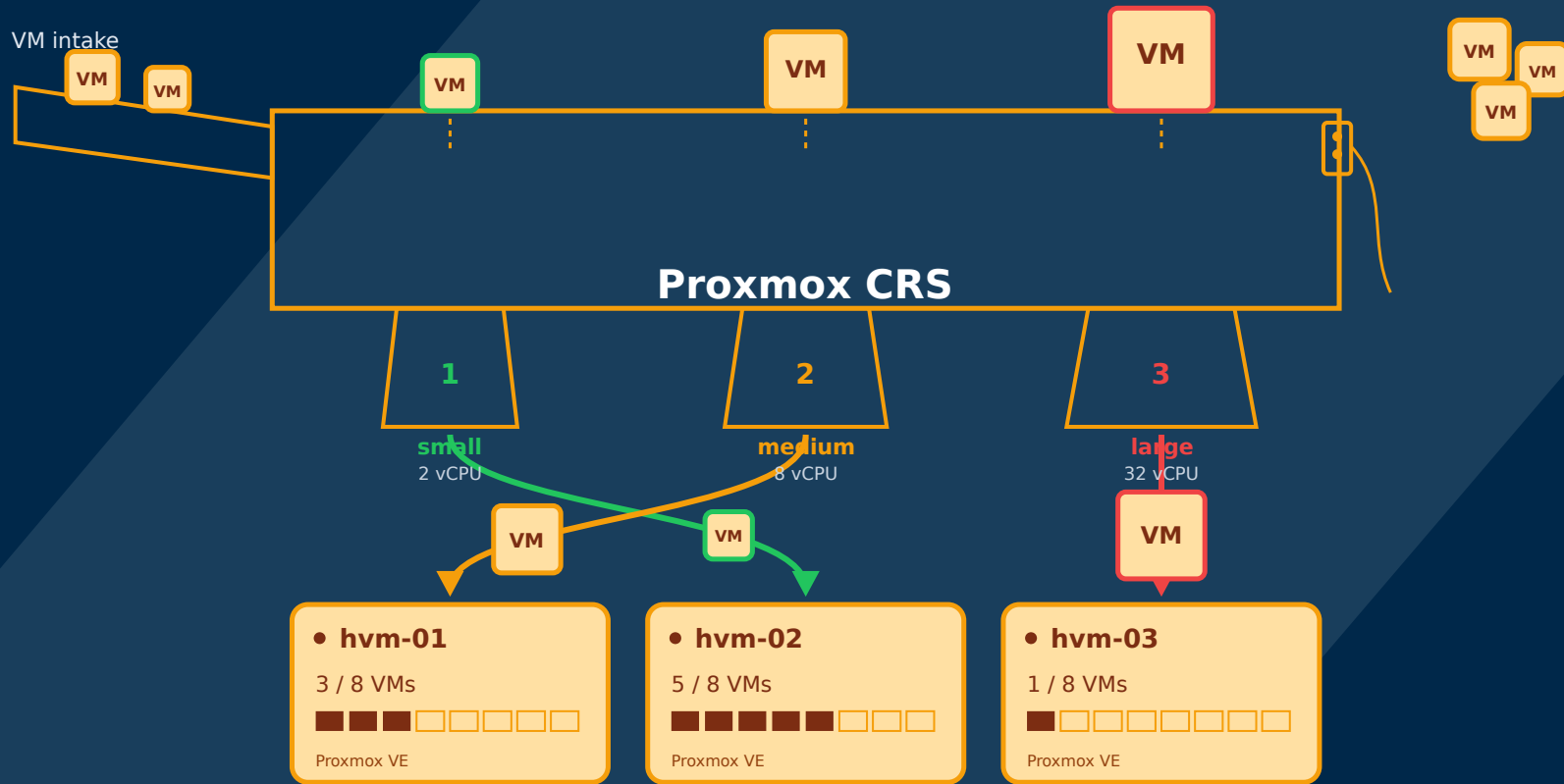
Type	Status
quorum	OK
fencing	armed (CRM watchdog active)
master	node01 (active, Thu Sep 24 12:06:53 2026) - dynamic load CRS
lrm	node01 (maintenance mode, watchdog active, Thu Sep 24 12:06:53 2026)
lrm	node02 (active, watchdog active, Thu Sep 24 12:06:55 2026)
lrm	node03 (active, watchdog active, Thu Sep 24 12:06:53 2026)

Resources

Add Edit Remove

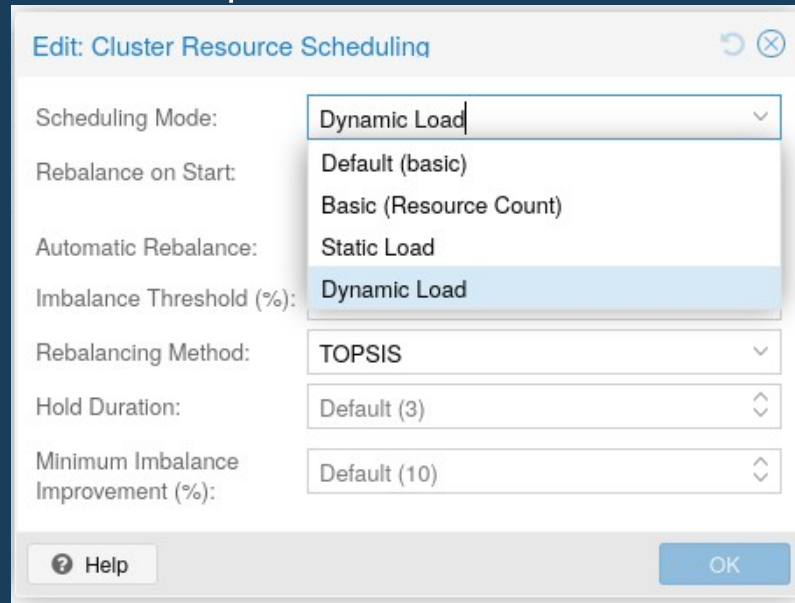


How about CRS?



How about CRS?

- Datacenter → Options → CRS



The screenshot shows a dialog box titled "Edit: Cluster Resource Scheduling". It contains several configuration options:

- Scheduling Mode:** A dropdown menu with "Dynamic Load" selected. The dropdown list is open, showing options: "Default (basic)", "Basic (Resource Count)", "Static Load", and "Dynamic Load" (highlighted).
- Rebalance on Start:** A dropdown menu with "Default (basic)" selected.
- Automatic Rebalance:** A dropdown menu with "Static Load" selected.
- Imbalance Threshold (%):** A dropdown menu with "Dynamic Load" selected.
- Rebalancing Method:** A dropdown menu with "TOPSIS" selected.
- Hold Duration:** A spinner box with "Default (3)" selected.
- Minimum Imbalance Improvement (%):** A spinner box with "Default (10)" selected.

At the bottom of the dialog, there is a "Help" button (with a question mark icon) and an "OK" button.



CRS Options

The image displays two overlapping configuration windows from the Proxmox Virtual Environment (VE) interface.

The foreground window is titled "Edit: Cluster Resource Scheduling". It contains the following settings:

- Scheduling Mode: **Dynamic Load**
- Rebalance on Start: ☒ Use CRS to select the least loaded node when starting an HA resource
- Automatic Rebalance: ☒ Automatically rebalance HA resources
- Imbalance Threshold (%): **10**
- Rebalancing Method: **TOPSIS**
- Hold Duration: **Default (3)**
- Minimum Imbalance Improvement (%): **Default (10)**

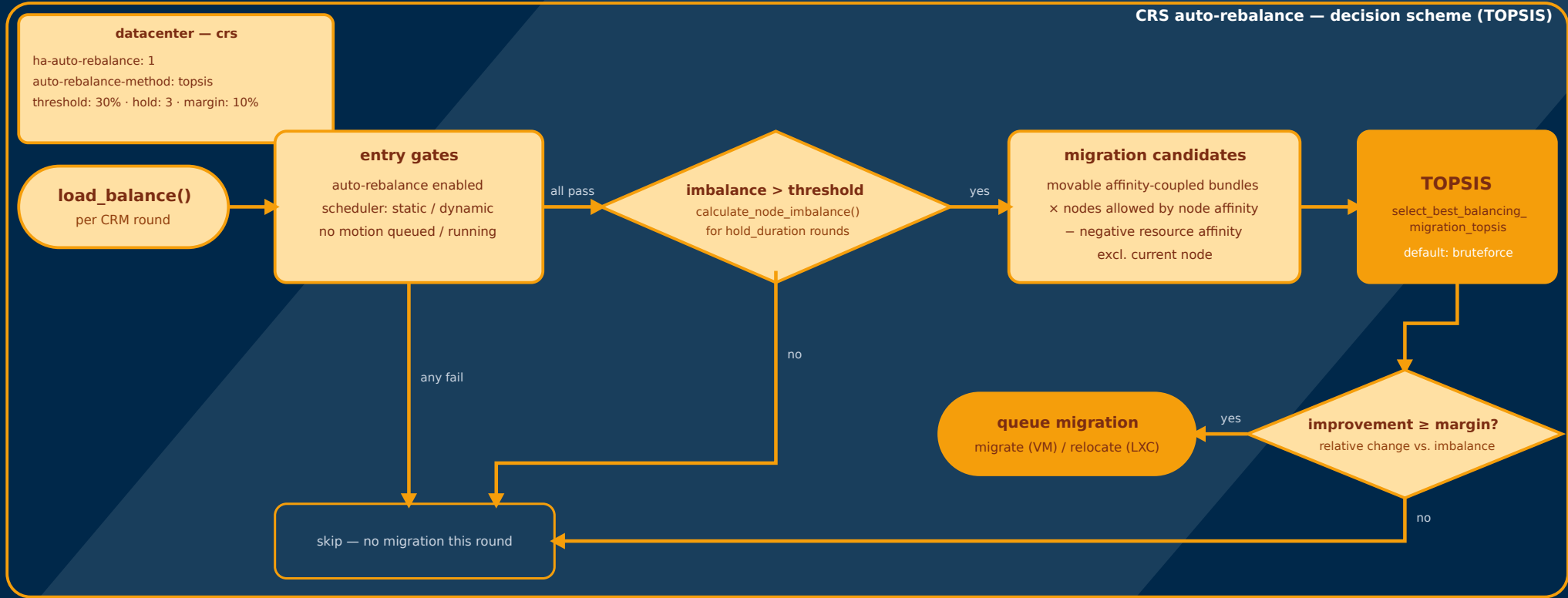
The background window is titled "Edit: Resource: Container/Virtual Machine". It shows VM ID **138**. An orange circle highlights the following options:

- Failback: ☒
- Auto-Rebalance: ☒
- Request State: **started**

Both windows have "OK" buttons at the bottom right. The foreground window also has a "Help" button at the bottom left.



CRS Active



Generated using Tuxis LLM



CRS in action



Conclusion

- HA Has seen major development in the last two years. Yet more to come: Check the Roadmap!
- CRS responds to the call of new userbase. Now, the basics are in place. Check the Roadmap!
- Tuxis will start enabling HA on most services, with HA & CRS, the cost/benefits shifted from costly to beneficial..

