NUTANIX

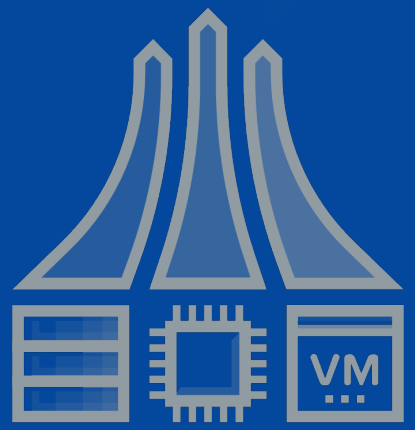
AHV + NVIDIA VGPU INTEGRATION

Malcolm Crossley – AHV GPU Architect

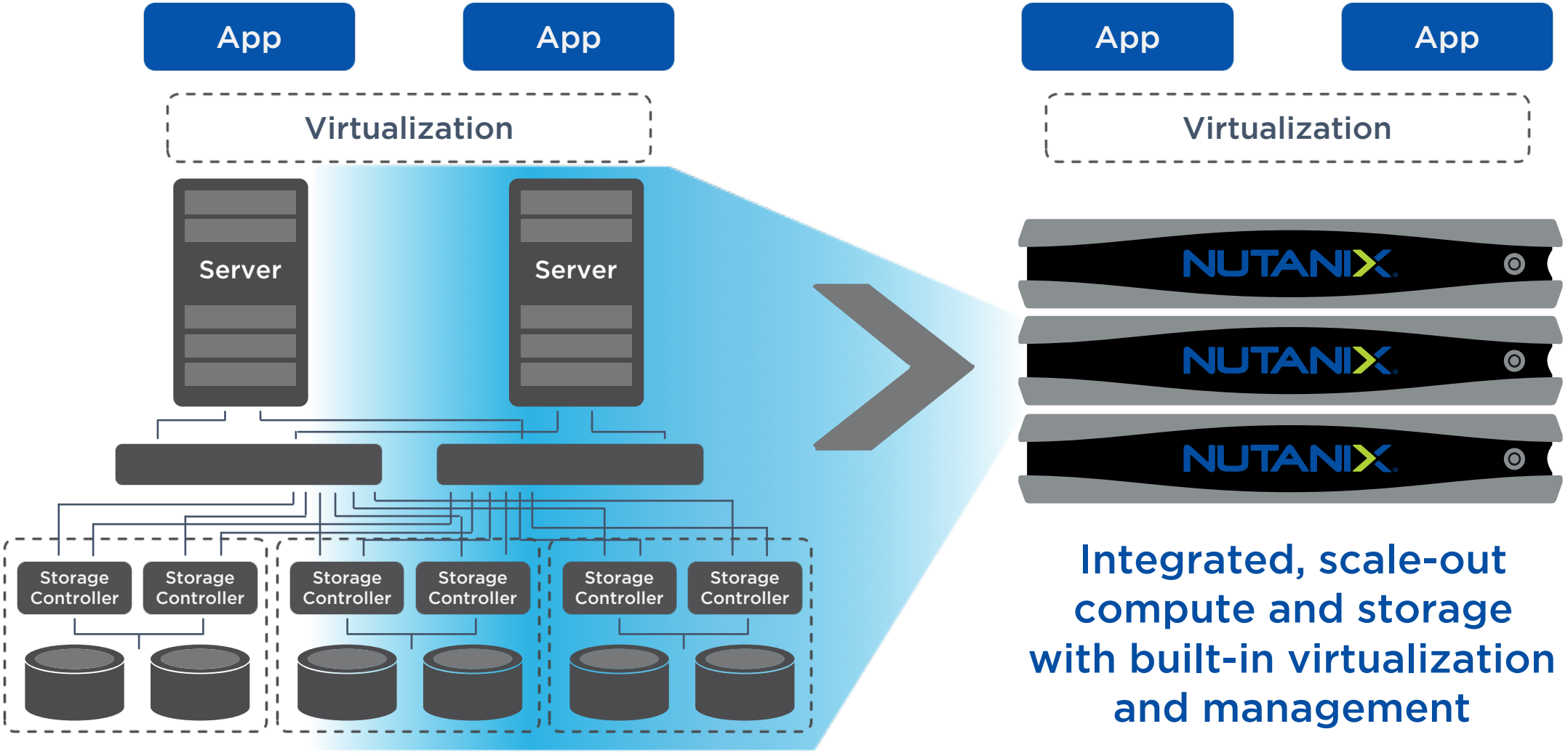
NUTANIX

Nutanix AHV

The hypervisor built for the Enterprise Cloud

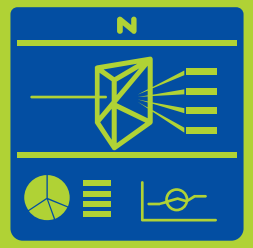


Nutanix: Built-in Virtualization

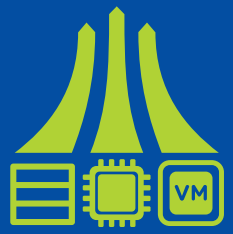


Integrated, scale-out compute and storage with built-in virtualization and management

Nutanix Platform



Nutanix Prism



Nutanix Acropolis

Comprehensive management solution that radically simplifies datacenter operations

Turnkey infrastructure platform that converges compute, storage and virtualization to run any application, at any scale

Nutanix Platform

Prism

- Infrastructure / VM Management
- Operational Insights
- Capacity Planning

Acropolis

- App Mobility Fabric
- Distributed Storage Fabric
- AHV (built-in Virtualization)

Citrix App and Desktop Virtualization

Starting with Citrix XD/XA 7.9 … AHV is integrated with MCS

Nutanix AHV is

CITRIX ready™

Citrix Infrastructure

CITRIX®

XenApp & XenDesktop
NetScaler VPX
ShareFile
And more …

Nutanix AHV

NUTANIX

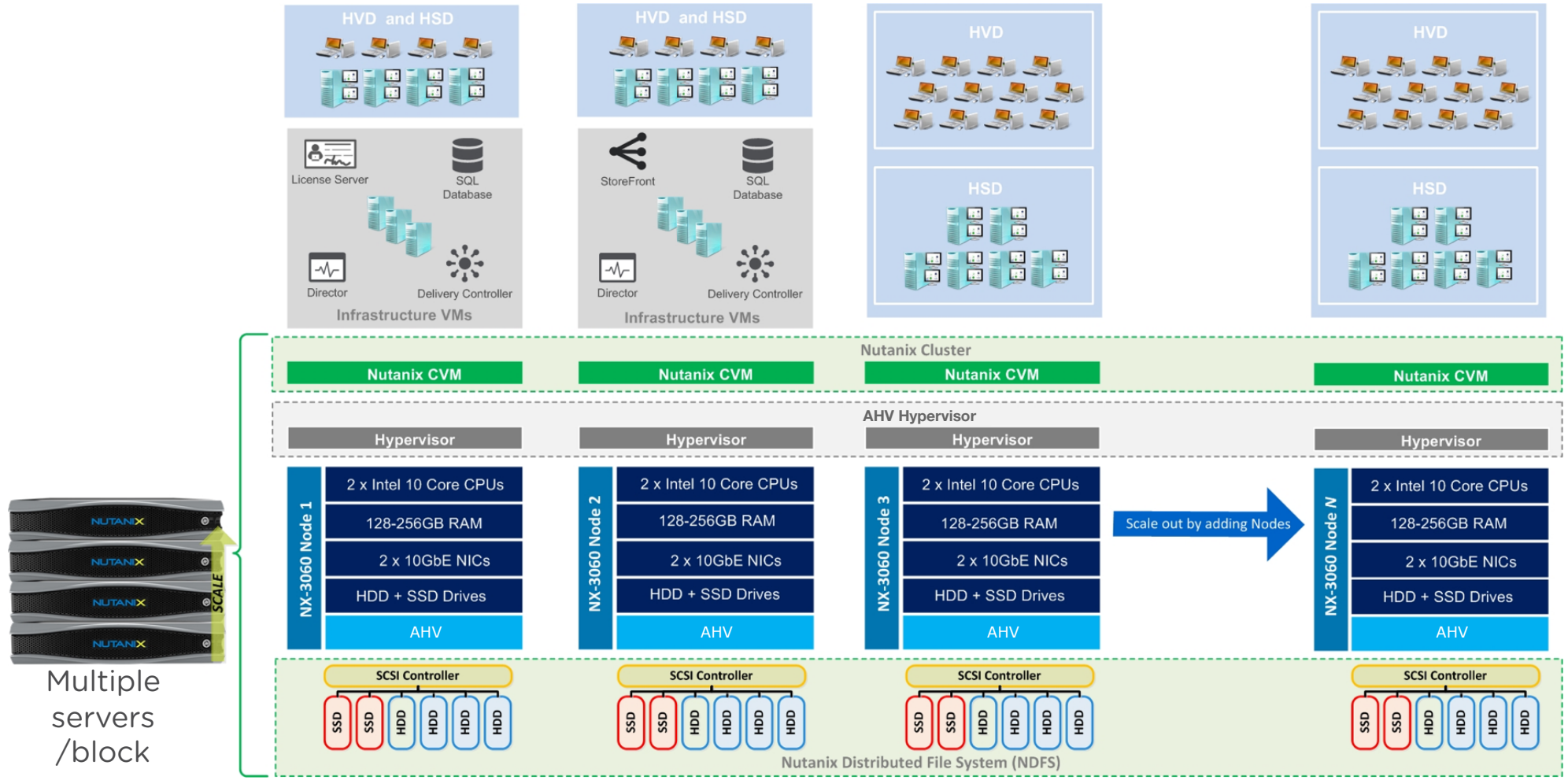
End user desktops

VM VM VM VM VM
VM VM VM VM VM

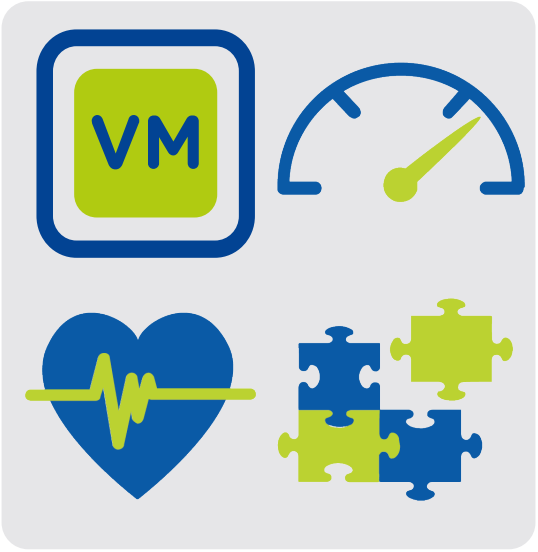
Nutanix AHV

NUTANIX

Complete XenDesktop Solution



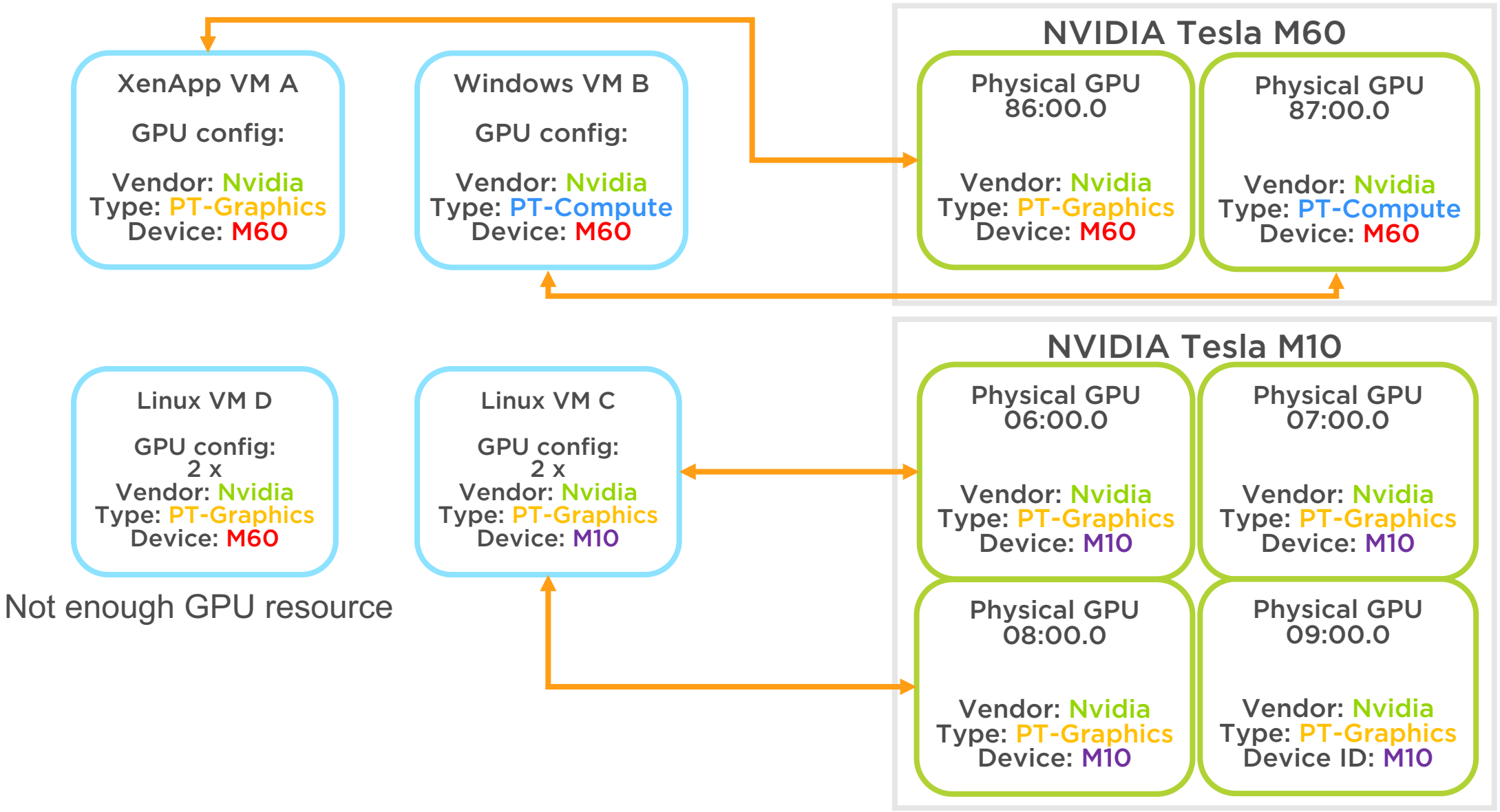
All the Features for Enterprise Applications



- Automated Resource Scheduling & VM Placement
- Performance Hot-Spot Avoidance (aka DRS)
- VM High Availability
- Live Migration
- VM-centric storage w/Snapshots and Clones
- Dynamic IP Address Management
- GPU Resource Management

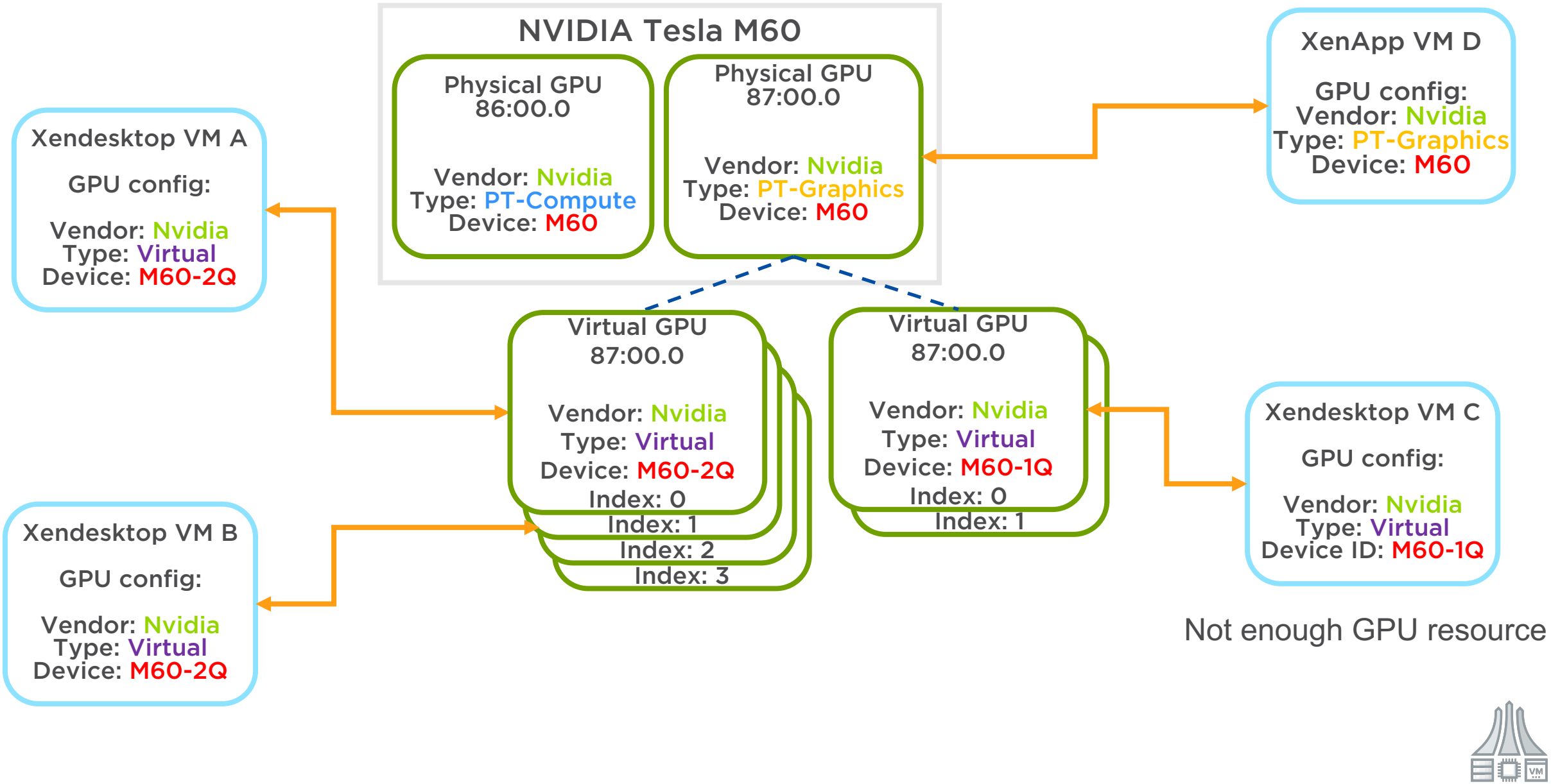
- Host Profiles
- Virtual Networking
- 1 click Upgrades
- Host Maintenance Mode
- Simplified Scaling (Automatic CPU Compatibility)
- Image Service
- Cross-Hypervisor Migration, DR & Backup

Acropolis GPU resource management concepts

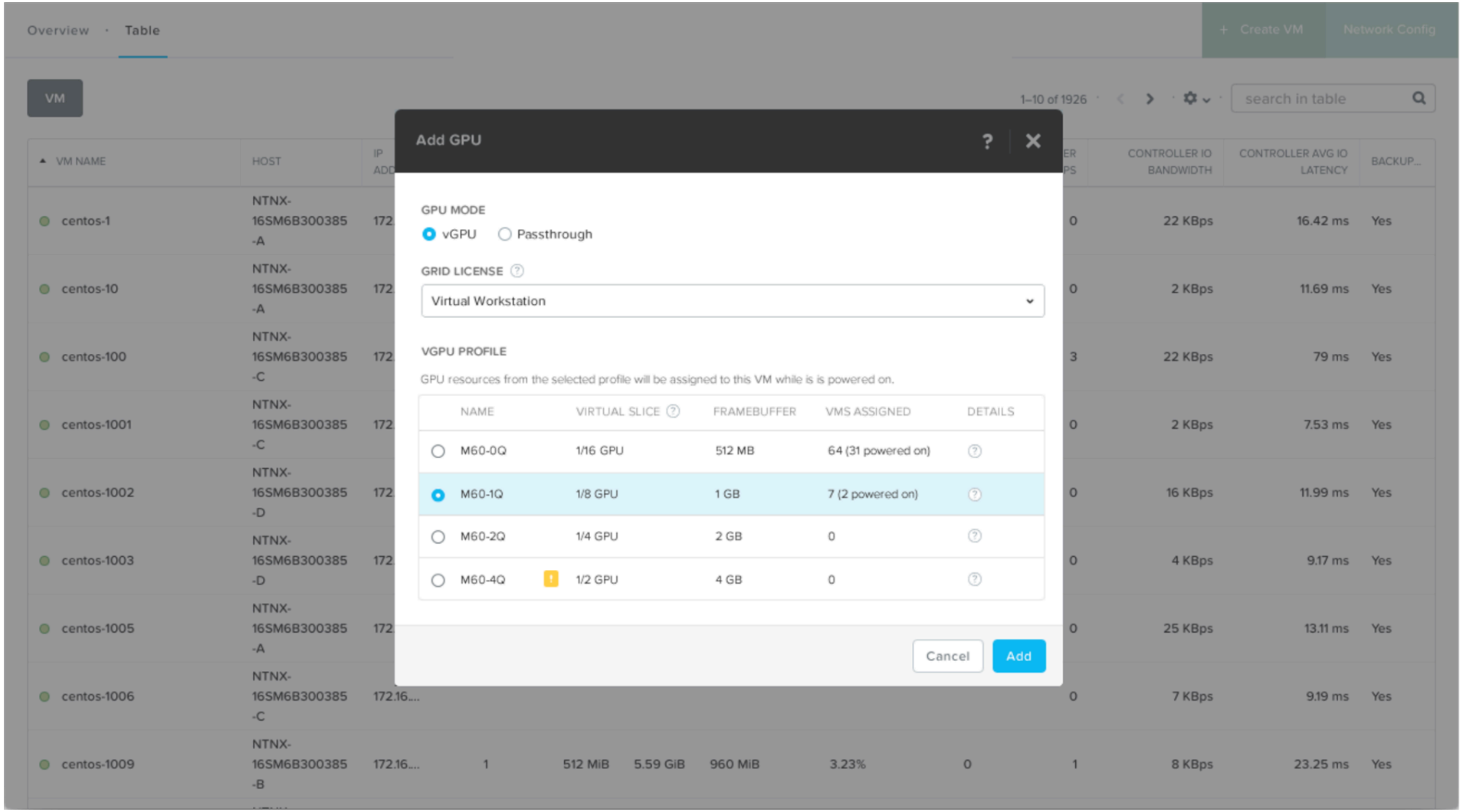


Not enough GPU resource

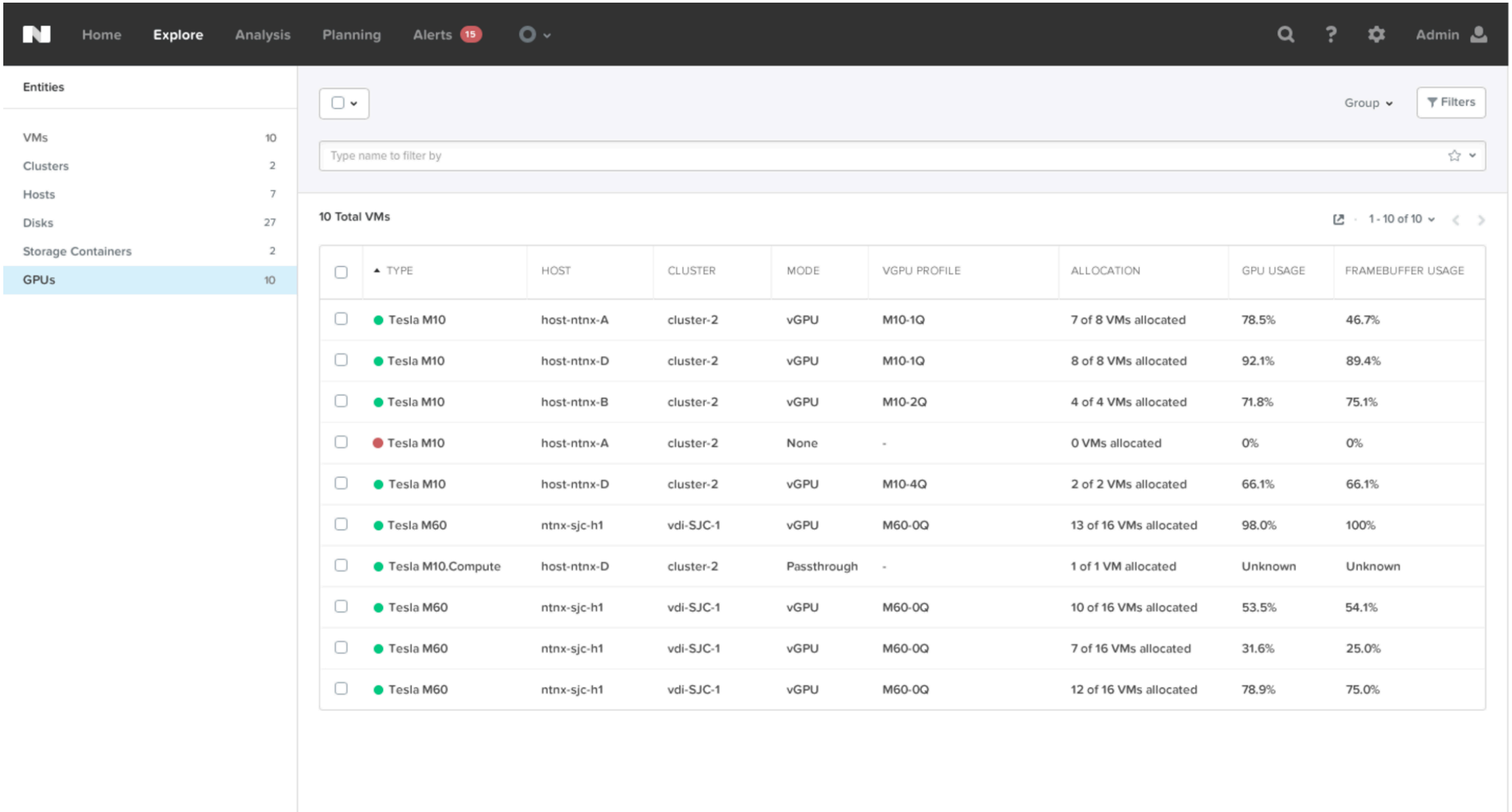
Acropolis GPU resource management concepts



AHV VM GPU resource configuration

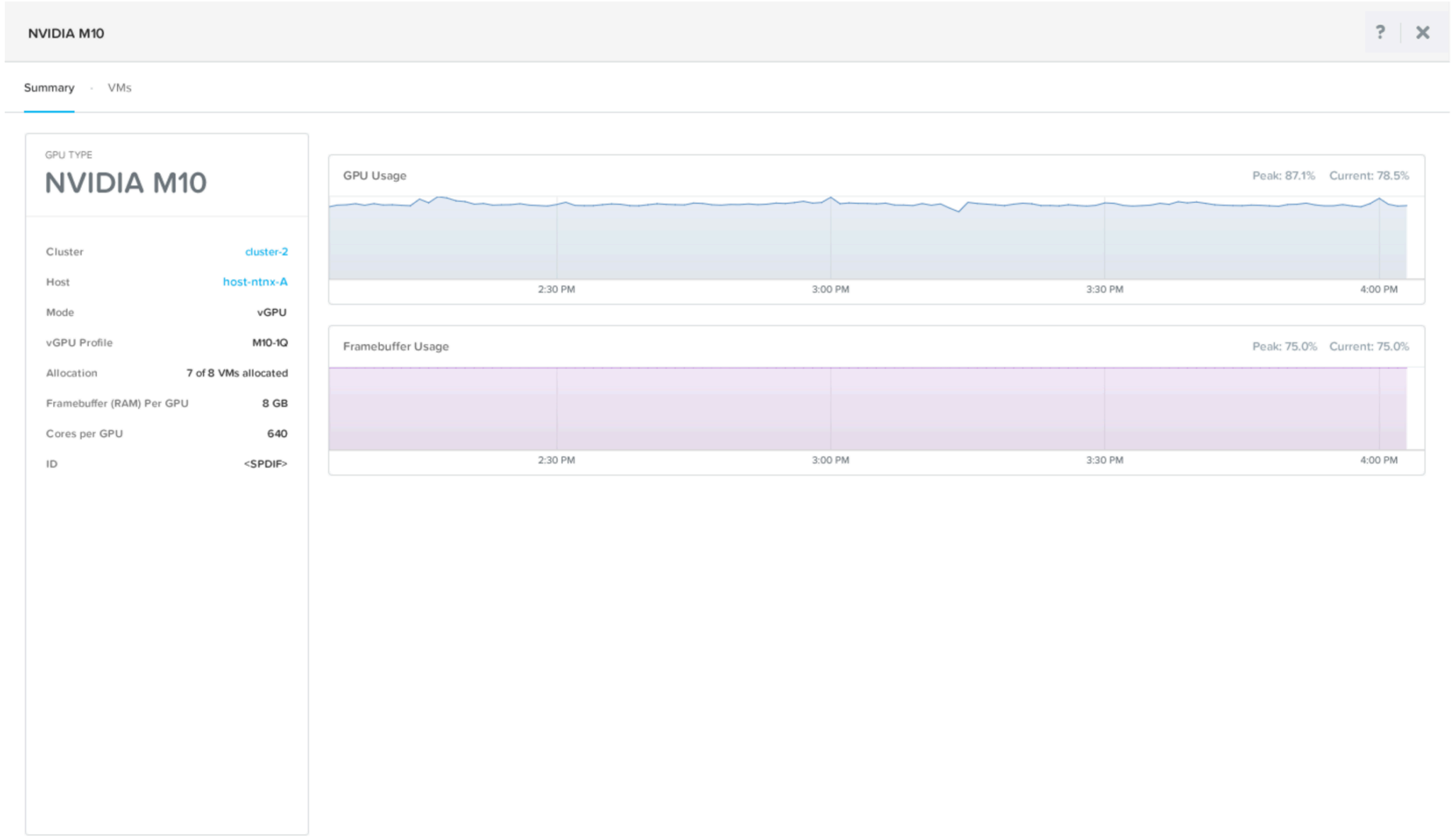


Prism UI - Physical GPU overview

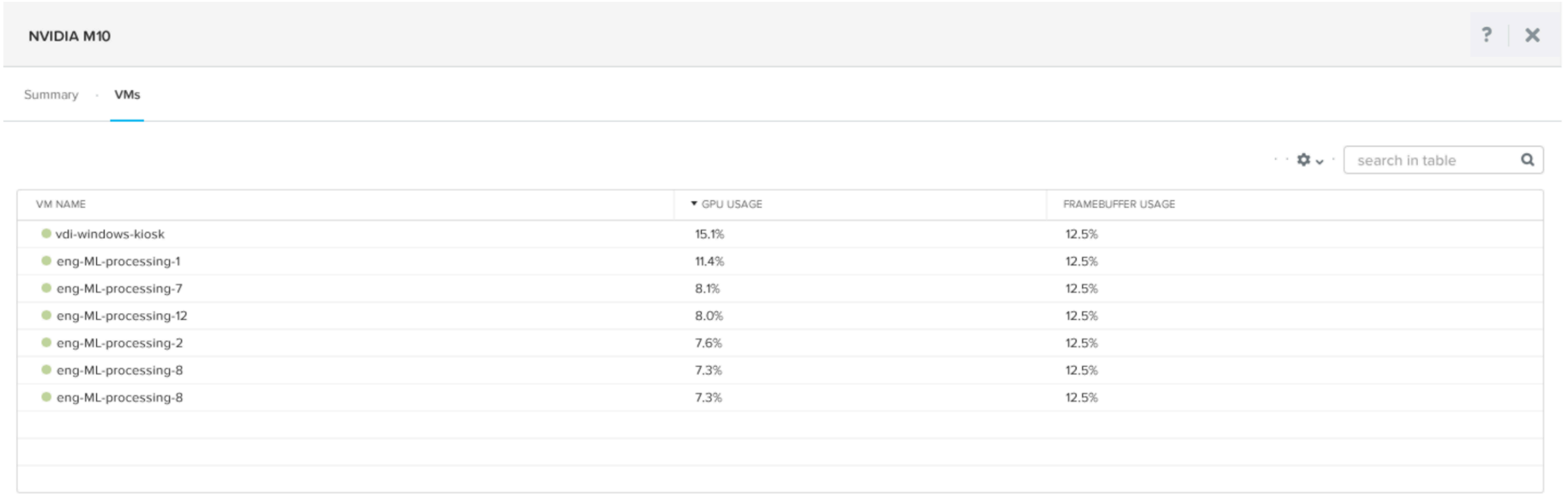


TYPE	HOST	CLUSTER	MODE	VGPU PROFILE	ALLOCATION	GPU USAGE	FRAMEBUFFER USAGE
Tesla M10	host-ntnx-A	cluster-2	vGPU	M10-1Q	7 of 8 VMs allocated	78.5%	46.7%
Tesla M10	host-ntnx-D	cluster-2	vGPU	M10-1Q	8 of 8 VMs allocated	92.1%	89.4%
Tesla M10	host-ntnx-B	cluster-2	vGPU	M10-2Q	4 of 4 VMs allocated	71.8%	75.1%
Tesla M10	host-ntnx-A	cluster-2	None	-	0 VMs allocated	0%	0%
Tesla M10	host-ntnx-D	cluster-2	vGPU	M10-4Q	2 of 2 VMs allocated	66.1%	66.1%
Tesla M60	ntnx-sjc-h1	vdi-SJC-1	vGPU	M60-0Q	13 of 16 VMs allocated	98.0%	100%
Tesla M10.Compute	host-ntnx-D	cluster-2	Passthrough	-	1 of 1 VM allocated	Unknown	Unknown
Tesla M60	ntnx-sjc-h1	vdi-SJC-1	vGPU	M60-0Q	10 of 16 VMs allocated	53.5%	54.1%
Tesla M60	ntnx-sjc-h1	vdi-SJC-1	vGPU	M60-0Q	7 of 16 VMs allocated	31.6%	25.0%
Tesla M60	ntnx-sjc-h1	vdi-SJC-1	vGPU	M60-0Q	12 of 16 VMs allocated	78.9%	75.0%

Prism UI - Physical GPU metrics



Prism UI - Virtual GPU metrics



VM NAME	GPU USAGE	FRAMEBUFFER USAGE
vdi-windows-kiosk	15.1%	12.5%
eng-ML-processing-1	11.4%	12.5%
eng-ML-processing-7	8.1%	12.5%
eng-ML-processing-12	8.0%	12.5%
eng-ML-processing-2	7.6%	12.5%
eng-ML-processing-8	7.3%	12.5%
eng-ML-processing-8	7.3%	12.5%

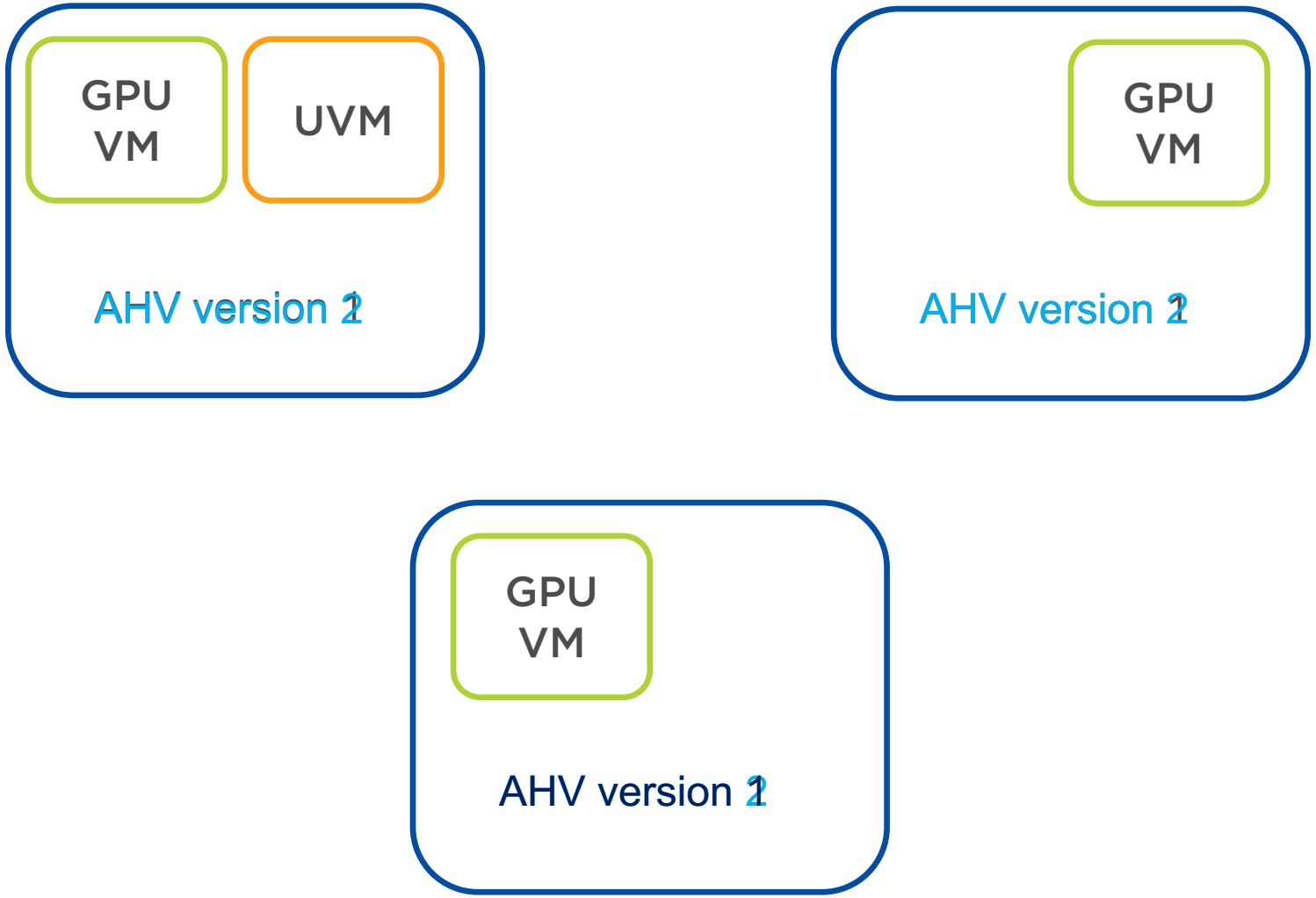
Powerful REST API for all GPU resource information

hosts

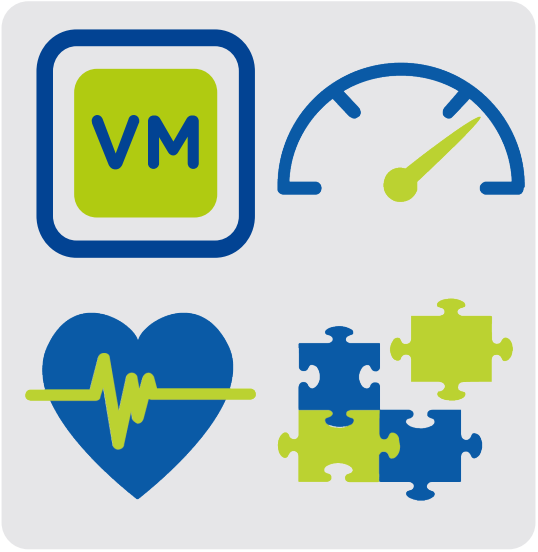
Show/Hide | List Operations | Expand Operations

GET	/hosts/gpu_profiles	Get the list of virtual GPU profiles for all physical Hosts.
GET	/hosts/host_gpus	Get the list of GPUs for all physical Host.
GET	/hosts/{uuid}/host_gpus	Get the list of GPUs for a physical Host.
GET	/hosts/{uuid}/host_gpus/{gpu_uuid}	Get a particular GPU for a physical Host.
GET	/hosts/{uuid}/host_gpus/{gpu_uuid}/stats/	Get stats for a particular GPU for a physical Host.

Acropolis 1-click operations and GPU resources



Possible future developments



- Prism App for Nvidia Dashboard
- Best effort High Availability for VMs with GPU resources

- Nutanix Guest Tools assist with Nvidia guest driver upgrades, timed to match Nvidia host driver upgrades

Questions?

NUTANIX®

Your Enterprise Cloud Platform