

Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Mariano Ezequiel Alvarez
Jefe de Departamento de Servicios de Red
Universidad Nacional de Quilmes
mariano@unq.edu.ar



Universidad
Nacional
de Quilmes

Esta obra está licenciada bajo Atribución-NoComercial-CompartirDerivadasIgual 2.5 Argentina de Creative Commons

Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Objetivos

Lograr implementación de un cluster de procesamiento basado en Proxmox, procurando minimizar la cantidad de cortes de servicio y downtime. Para esto se definieron dos líneas de trabajo:

- Infraestructura física.
- Infraestructura de servidores y servicios.

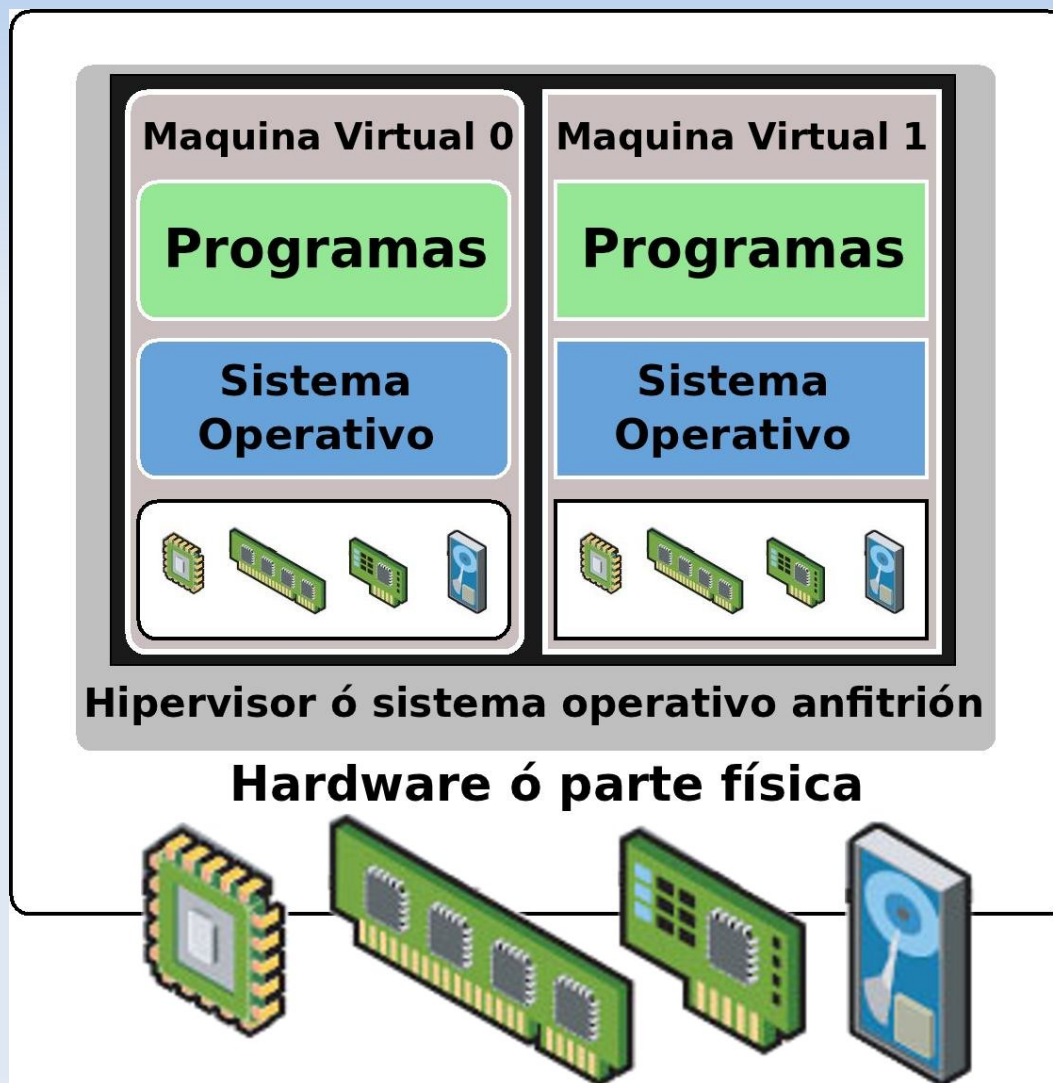
Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Principales ventajas

- Administración Web centralizada en el Master.
- Consola para visualizar todas las máquinas virtuales.
- Permite realizar migración de máquinas virtuales entre equipos físicos.
- Sincronización de plantillas OpenVZ e imágenes ISO en todos los Nodos.
- Es el paso inicial en el camino para lograr redundancia de Servidores y/o Alta Disponibilidad (HA).

Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Diagrama ilustrativo



Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Consideraciones sobre el Cluster

- Gestión centralizada web, incluyendo la consola VNC seguro
- Soporte para múltiples fuentes de autenticación (locales, MS ADS, LDAP, ...)
- Gestión de permisos basados en papel para todos los objetos (VM's, almacenes, nodos, etc.)
- Crea grupos de varios maestros (sin single master)
- Migración de máquinas virtuales entre hosts físicos

Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Requisitos

- Todos los nodos deben estar en la misma red que usa IP Multicast para la comunicación entre nodos.
- La fecha y hora tienen que estar sincronizadas.
- Se utiliza túnel SSH en el puerto 22 entre los nodos.
- El tráfico de la consola VNC está asegurada a través de SSL, utilizando los puertos entre 5900 y 5999.
- Para implementar alta disponibilidad (HA), por quórum confiable debe tener al menos 3 nodos activos en todo momento, o utilizar un qdisk.

Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Alta Disponibilidad

La versión actual de Proxmox VE (5.1-43) posee funcionalidades nativas para tener máquinas virtuales redundantes y alta disponibilidad.

No obstante, para nuestra implementación, aplicamos más conceptos en la conexión de cada nodo, como bonding, link aggregation y redundancia de switches, para mejorar anchos de banda y disponibilidad, que las veremos a continuación:

Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Etapa 1 (nodos independientes)

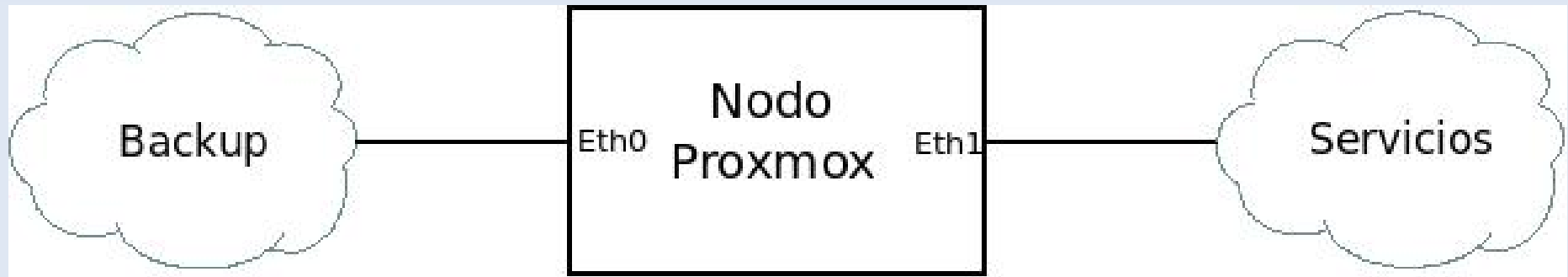
Inicialmente se contaba con:

- Servidores físicos HP.
- Proxmox como Hipervisor de nivel 1
- Corriendo VMs locales (discos virtuales creados localmente) en DAS (Direct Attached Storage).
- 2 interfaces de red a 1Gbps:
 - Red de backup.
 - Red de servicios.
- Administración web distribuída.

Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Etapa 1 (nodos independientes)

Nodo Proxmox



Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Etapa 1 (nodos independientes)

Desventajas:

- Administración web descentralizada.
- Sin redundancia de conexiones de red.
- Sin soporte para Live Migration.

Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Etapa 2 (nodos independientes + NAS)

Se incorporan 2 nodos (servidores) de almacenamiento:

- HP StoreVirtual 4330 con 7,2TB totales brutos

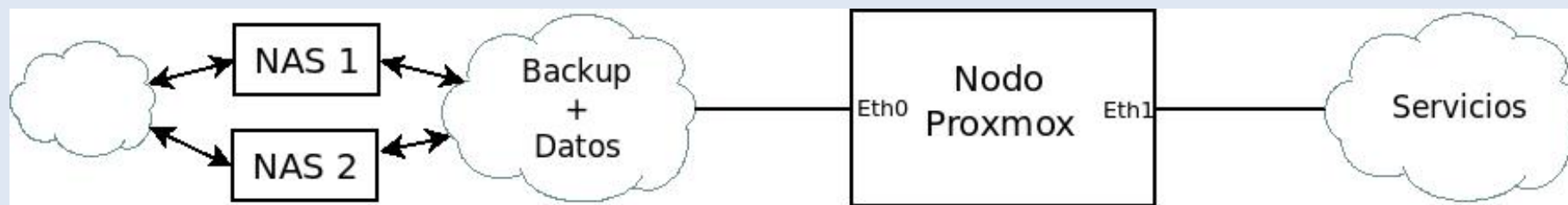


Que implementan RAID local (en cada nodo) y RAID por red (entre nodos). Debido a esto, fue necesario diseñar e implementar una red de datos dedicada, denominada SAN (Storage Area Network). Para realizarlo, se empleó la red existente de backups (aplicada para realizar el backup de datos de servidores).

Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Etapa 2 (nodos independientes + NAS)

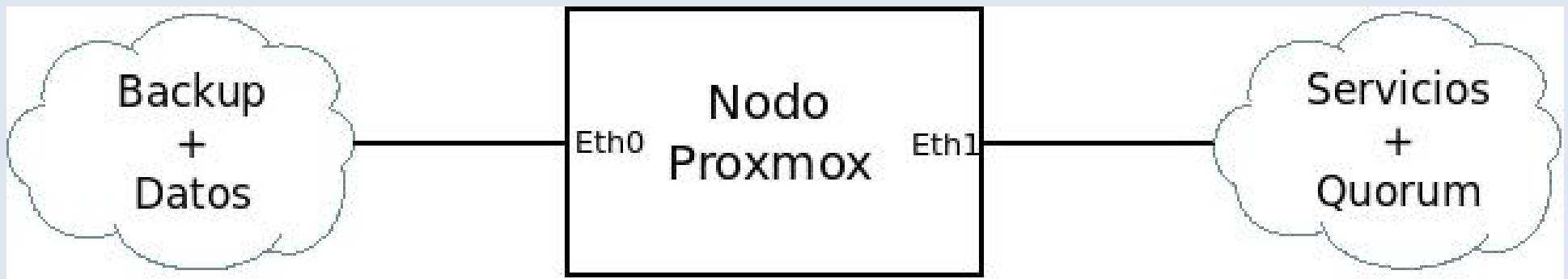
Nodo Proxmox + NAS



Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Etapa 3 (Cluster + NAS)

Luego de varias pruebas de laboratorio, comenzamos la integración de nodos para conformar el Cluster de Procesamiento, integrando los 8 nodos independientes uno por uno al cluster, manteniendo las 2 tarjetas de 1Gbps por nodo, es decir, una para backup/datos y la otra para servicios/quorum.



Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Etapa 3 (Cluster + NAS)

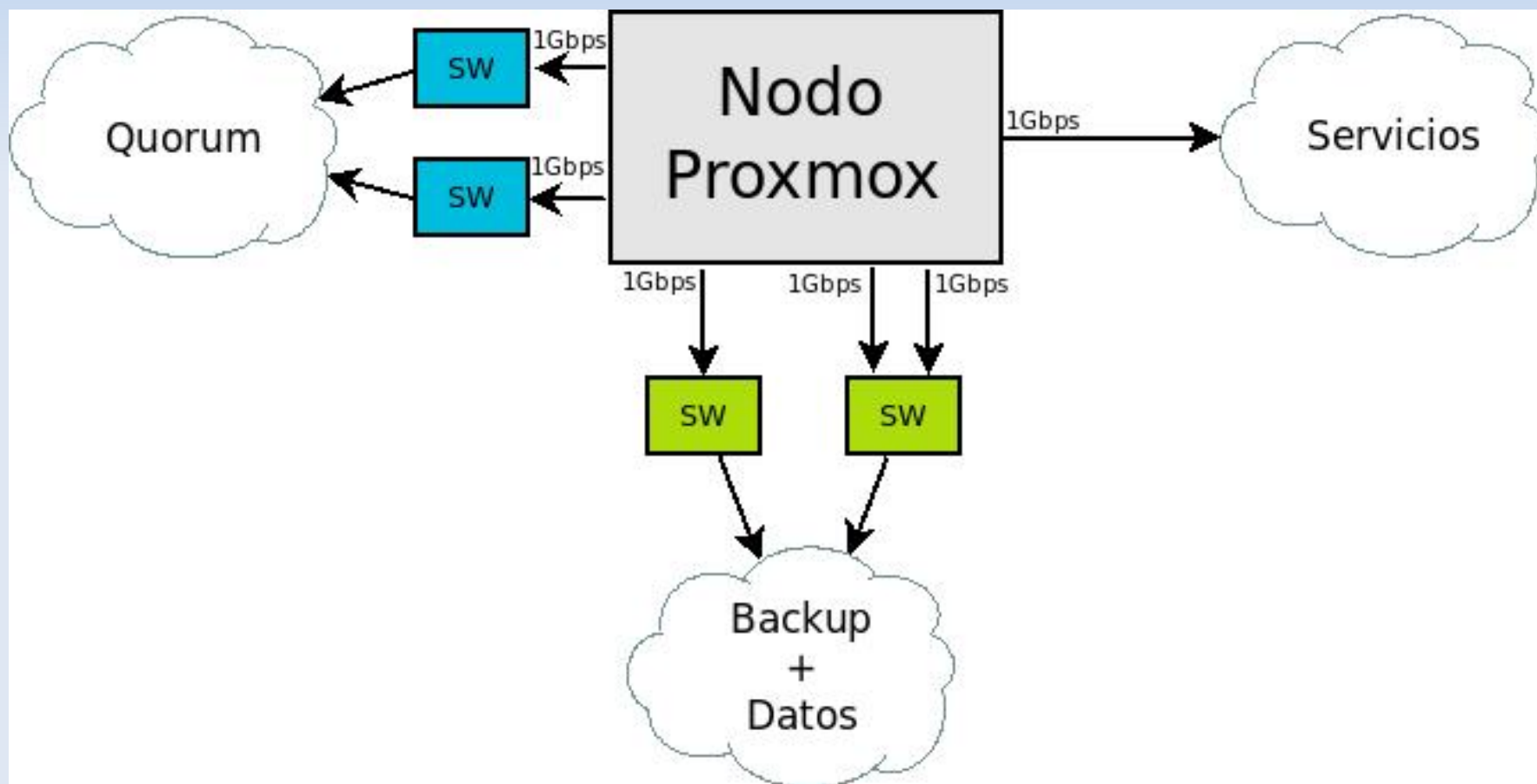
Problemas encontrados:

- Se producían salidas de servicio no programada en dos nodos que, inexplicamente, se "salían" del Cluster, arrastrando sus VMs corriendo localmente, provocando la salida de servicios de varios VMs como correo electrónico, web, dns.
- Comenzamos a investigar y concluimos en separar las redes y emplear 6 tarjetas de red por nodo, aplicando bonding, link aggregation y redundancia de switchs para mejorar anchos de banda y disponibilidad y lograr así la estabilidad general del Cluster.

Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Etapa 3 (Cluster + NAS)

Nodo cluster Proxmox



Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Cluster Proxmox + NAS

Ventajas:

- Administración web centralizada.
- Live Migration de VMs.
- Separación total de red de servicio, red de quorum y red de datos/backup.
- Mayor ancho de banda contra NAS.
- Redundancia de enlaces de red.

Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Cluster Proxmox + NAS

Desventajas:

- Infraestructura física y lógica de red:
 - Compleja.
 - Costosa.
- Almacenamiento:
 - Equipos costosos.

Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Consola Web centralizada (Home)

The screenshot displays the Proxmox VE Web Console interface. At the top, it shows the Proxmox logo and version (5.1-43), a search bar, and user information (logged in as 'mariano@pam'). Navigation buttons for 'Documentation', 'Create VM', 'Create CT', and 'Logout' are visible. The main content area is divided into a left sidebar for navigation, a central menu for the selected 'Datacenter', and a main table of resources.

Server View

- Datacenter
 - heimdall
 - hiperion
 - odin
 - persefone
 - perseo
 - prometeo
 - thor
 - tyr
 - tytan
 - AdmRedes
 - DirDesarrollo
 - Firewalls
 - ITT

Datacenter

- Search
- Summary
- Options
- Storage
- Backup
- Replication
- Permissions
 - Users
 - Groups
 - Pools
 - Roles
- Authentication
- HA
- Firewall
- Support

Type	Description ↓	Disk size	Memory us...	CPU usage
storage	vg0-perseo (perseo)	4.37 TiB		
node	tytan	94.37 GiB	26.0 %	1.7% of 32...
storage	tyr-backups (tyr)	2.42 TiB		
node	tyr	94.37 GiB	49.4 %	20.9% of 1...
node	thor	94.37 GiB	16.2 %	0.9% of 8C...
storage	storage-nfs-hiperion (tytan)	4.87 TiB		
storage	storage-nfs-hiperion (tyr)	4.87 TiB		
storage	storage-nfs-hiperion (thor)	4.87 TiB		
storage	storage-nfs-hiperion (prometeo)	4.87 TiB		
storage	storage-nfs-hiperion (perseo)	4.87 TiB		
storage	storage-nfs-hiperion (persefone)	4.87 TiB		
storage	storage-nfs-hiperion (odin)	4.87 TiB		
storage	storage-nfs-hiperion (hiperion)	4.87 TiB		
storage	storage-nfs-hiperion (heimdall)	4.87 TiB		
node	prometeo	94.37 GiB	65.3 %	18.9% of 1

Tasks Cluster log

Start Time ↓	End Time	Node	User name	Description	Status
Apr 20 05:47:04	Apr 20 05:47:11	heimdall	root@pam	Update package database	Error: command not found

Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Consola Web central (Resumen)

The screenshot displays the Proxmox Web Console interface. At the top, it shows the Proxmox logo and version (5.1-43), a search bar, and user information (logged in as 'mariano@pam'). The main content area is divided into several sections:

- Health:** Shows the cluster status as 'Online' with a green checkmark. It lists 9 nodes online and 0 offline. The cluster name is 'pmx-cluster-unq' and Quorate is 'Yes'.
- Guests:** Displays the status of Virtual Machines and LXC Containers. There are 99 running VMs and 22 stopped VMs. There are 2 running LXC containers and 0 stopped LXC containers.
- Resources:** Shows three gauges for CPU (17% of 176 CPU(s)), Memory (56% of 376.97 GiB), and Storage (75% of 32.10 TiB).

On the left, there is a 'Server View' sidebar with a tree structure of the datacenter, including nodes like 'heimdall', 'hiperion', 'odin', etc., and various services like 'AdmRedes', 'DirDesarrollo', 'Firewalls', and 'ITT'. A central navigation menu includes options like 'Summary', 'Options', 'Storage', 'Backup', 'Replication', 'Permissions', 'Users', 'Groups', 'Pools', 'Roles', 'Authentication', 'HA', 'Firewall', and 'Support'.

At the bottom, there is a 'Tasks' section with a 'Cluster log' table. The table has columns for Start Time, End Time, Node, User name, Description, and Status. A single task is visible: 'Update package database' performed by 'root@pam' on node 'heimdall' at 05:47:11.

Start Time ↓	End Time	Node	User name	Description	Status
Apr 20 05:47:11	Apr 20 05:47:11	heimdall	root@pam	Update package database	Error: command not set - und

Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Almacenamiento NAS

PROXMOX Virtual Environment 5.1-43

You are logged in as 'mariano@pam'

Documentation Create VM Create CT Logout

Server View

Datacenter

Search Summary Options Storage Backup Replication Permissions Users Groups Pools Roles Authentication HA Firewall Support

ID ↑	Type	Content	Path/Target	Shared	Enabled
VG0-prome...	LVM	Disk image, Container		No	No
VG1-prome...	LVM	Disk image, Container		No	No
cluster-nas...	LVM	Disk image, Container		Yes	Yes
hiperion-dir...	Dire...	VZDump backup file	/mnt/backups	No	Yes
local	Dire...	VZDump backup fil...	/var/lib/vz	No	Yes
lv-perseo	Dire...	VZDump backup fil...	/mnt/lv-perseo	No	Yes
nas-storage	iSCSI		iqn.2003-10.com.lefthandnetworks:storagenasunq:41:cluster-proxmox	Yes	Yes
storage-nfs...	NFS	VZDump backup fil...	/mnt/pve/storage-nfs-hiperion	Yes	Yes
thor-directo...	Dire...	VZDump backup file	/var/lib/vz/NFS	No	No
tyr-backups	Dire...	VZDump backup file	/mnt/backups	No	Yes
vg0-perseo	LVM	Disk image, Container		No	Yes

Tasks Cluster log

Start Time ↓	End Time	Node	User name	Description	Status
Apr 20 05:47:04	Apr 20 05:47:11	heimdall	root@pam	Update package database	Error command not found

Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Almacenamiento DAS

The screenshot displays the Proxmox VE 5.1-43 interface. The top navigation bar includes the Proxmox logo, the version number, a search bar, and the user 'mariano@pam'. On the left, a 'Server View' sidebar shows a tree of nodes, with 'tytan' selected. The main content area shows the 'Node tytan' summary, including system information, package versions, and a CPU usage graph.

Node 'tytan' Summary

tytan (Uptime: 72 days 23:02:29)

CPU usage	1.55% of 32 CPU(s)	IO delay	0.17%
Load average	0.96, 1.19, 1.26		
RAM usage	26.02% (20.47 GiB of 78.66 GiB)	KSM sharing	0 B
HD space(root)	5.93% (5.59 GiB of 94.37 GiB)	SWAP usage	0.69% (436.03 MiB of 62.00 GiB)

System Information:

- CPU(s): 32 x AMD Opteron(tm) Processor 6136 (4 Sockets)
- Kernel Version: Linux 4.13.13-5-pve #1 SMP PVE 4.13.13-38 (Fri, 26 Jan 2018 10:47:09 +0100)
- PVE Manager Version: pve-manager/5.1-43/dbd08029

CPU usage graph

The graph shows CPU usage percentage over time from 2018-04-11 to 2018-04-19. The usage fluctuates between approximately 1% and 5%.

Start Time ↓	End Time	Node	User name	Description	Status
Apr 20 12:48:17	Apr 20 12:49:28	pamafone	mariano@pam	Shell	OK

Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Cluster Proxmox + NAS

¿Consultas?

Implementación de cluster de virtualización con alta disponibilidad basada en Proxmox

Gracias

Mariano Alvarez (mariano@unq.edu.ar)
Jefe de Departamento de Servicios de Red

Dirección de Servicios de Comunicación
Universidad Nacional de Quilmes



Universidad
Nacional
de Quilmes