

## NATIONAL INSTITUTE FOR ASTROPHYSICS

INAF

ISTITUTO NAZIONALE DI ASTROFISICA

### Rapporti Tecnici INAF INAF Technical Reports

Number	228
Publication Year	2023
Acceptance in OA@INAF	2023-01-19T10:44:46Z
Title	Migrazione del cluster di macchine virtuali di OAS da OVS a Proxmox VE
Authors	TACCHINI, ALESSANDRO
Affiliation of first author	OAS Bologna
Handle	http://hdl.handle.net/20.500.12386/32924, https://doi.org/10.20371/INAF/TechRep/228

# Migrazione del cluster di macchine virtuali di OAS da OVS a Proxmox VE

Autore: Alessandro Tacchini

## Introduzione

Nel nostro istituto, l'OAS-INAF di Bologna, si è reso necessario il passaggio del cluster di virtualizzazione realizzato con Oracle Virtual Server ad un nuovo cluster realizzato con Proxmox Virtual Environment Da molti anni oramai alcuni servizi informatici, di IASFBO prima ed OAS ora, vengono implementati attraverso server realizzati con macchine virtuali.

I vantaggi sono molteplici: flessibilità, robustezza, velocità nel deployment di una macchina.

Il primo sistema è stato creato utilizzando macchine dismesse dai vari progetti di calcolo e raccolte in un cluster.

Nel tempo, tuttavia, l'aumento dei servizi offerti e l'usura degli apparati hanno portato ad una obsolescenza che ha reso il sistema inutilizzabile. Da qui la necessità di passare ad un sistema nuovo.

In questo documento si illustrerà il passaggio dal vecchio sistema al nuovo riportando anche l'elenco dei comandi usati.

## Vecchio Cluster Virtual Machine

È un cluster formato da sette macchine, più una di controllo, non omogenee su cui è installato Oracle VM Server.

Un gruppo di macchine, per la precisione tre, ha 8 Core AMD e 16GB di RAM, due macchine hanno 8 Core Intel e 32 GB di RAM (queste cinque macchine sono contenute all'interno di una Enclosure Dell) ed infine due macchine gemelle hanno 32 Core AMD e 128GB di RAM.



Nome nodo	CPU Core	RAM (GB)			
nodo01	Intel 8	32			
nodo02	Intel 8	32			
nodo03	Intel 8	16			
nodo05	Intel 8	16			
nodo06	Intel 8	16			
vs1	AMD 32	128			
vs2	AMD 32	128			

Lo storage è altrettanto eterogeneo essendo composto da un server Supermicro riciclato da altri progetti di ricerca che esporta via iSCSI circa 6 TB di spazio disco; un server Supermicro che esporta circa 8TB sempre via iSCSI; infine un server SAN IBM Storewize V5000 che esporta via iSCSI 10 TB.

Da notare che OVS gestisce lo storage integrando le fonti esterne ed interne e permette la creazione di repository che vengono poi usati dalle macchine virtuali per lo storage.



Invece in Proxmox VE ogni fonte esterna di storage va agganciata ad ogni nodo del cluster.

Le risorse Hardware quindi sono di dimensioni variabili e questo comporta un sostanziale disequilibrio del cluster che però non incide sulle prestazioni desiderate.

Il software di Oracle realizza un'infrastruttura robusta che offre alta affidabilità ed alta disponibilità.

Le macchine virtuali ospitate sul sistema offrono servizi di due tipi: servizi generali forniti alla struttura e risorse informatiche fornite ai ricercatori nell'ambito, ma non solo, di progetti di ricerca.

I servizi generali erogati sono del tipo: DNS, DHCP, Owncloud, posta elettronica<sup>1</sup>, LDAP, RADIUS/EDUROAM, server di licenze<sup>2</sup>, gestione delle presenze in caso di emergenze.

L'utilizzo di queste risorse da parte degli utenti può essere di tipo personale, ad esempio utilizzando una macchina virtuale come workstation contenente software scientifico o CAD; oppure può essere legato a progetti di ricerca, alcuni dei quali sono ad esempio: CTA, AGILE<sup>3</sup>, THESEUS, EUCLID.

Gli aspetti negativi di OVS (Oracle Virtual Server) sono l'assenza di un sistema di backup integrato e la gestione interna dai file, ad esempio i file che contengono i dischi virtuali sono in formato Raw e sono ubicati in sottodirectory contrassegnate solo da codici numerici. I file stessi sono contrassegnati da codici alfanumerici rendendo difficile l'associazione con la relativa macchina virtuale.

Con una procedura macchinosa è possibile effettuare un disaster recovery avendo a disposizione una copia dei file di immagine dei dischi virtuali e tutte le informazioni relative alle varie macchine virtuali.

<sup>1</sup> Zimbra

<sup>&</sup>lt;sup>2</sup> Per esempio IDL

<sup>&</sup>lt;sup>3</sup> Pipeline di Agile

## Migrazione ad un cluster Proxmox temporaneo

Poichè il vecchio sistema OVS era diventato talmente instabile da comprometterne l'uso e minarne l'integrità e poiché l'hardware del nuovo cluster Proxmox non era ancora a disposizione si è resa necessaria una azione tempestiva di migrazione ad un cluster temporaneo al fine di mettere al sicuro i dati.

Si è proceduto nel modo seguente: si è scorporata la macchina vs1<sup>4</sup> dal cluster OVS spostando tutte le macchine virtuali e tutti i metadati sui nodi rimanenti, successivamente si è installato Proxmox VE 6.3-3 su un cluster di 4 macchine comprendente la macchina vs1 e tre blade server forniti in prestito da un progetto di ricerca e già presenti nel Centro di Calcolo.

Questi tre nodi sono stati chiamati oasproxmox1, oasproxmox2, oasproxmox3.

Ciascuno di questi nodi ha le seguenti caratteristiche: 16 Core; 2 dischi da 250GB in configurazione zmirror; 24GB di RAM.

Tutti i nodi fanno uso di una rete ethernet a 1GB/s con indirizzi sulla rete nascosta di istituto.

La parte più impegnativa è stata quella di identificare, per ogni macchina virtuale presente su OVS, i file di configurazione ed i codici identificativi dei dischi virtuali, e successivamente di identificare la loro posizione fisica. In pratica è stato trovato il percorso assoluto per ogni file associato ad un disco virtuale.

Questi file sono molto importanti, essi rappresentano l'equivalente del disco fisico di una macchina reale in cui sono contenuti tutti i dati. Costituiscono l'essenza della macchina virtuale.

In OVS tali file hanno l'estensione .img che è del tutto equivalente al comune formato RAW.

<sup>&</sup>lt;sup>4</sup> evicted

Nome	RAM	CPU/C ore	Dischi
LDAP	512MB	1	03deef9a65234aacb3d0f8fc8c76bdf1.img
Lucy	512MB	1	f4846e3cc97a45758e3dc281cfa547aa.img sistema 0004fb0000120000202f3416dd76fa23.img home 0004fb0000120000e9b4efa424a96656.img log
Theseus	4096MB	2	0004fb000012000093d2acb7178ea3c3.img
hangar_web	2048MB	2	db644324a8f34e8a9fb4fc9cf87e1d52.img
hermes	2048MB	2	0004fb00001200004e7c7afe23b68974.img
redsox	4096MB	2	0004fb00001200000135698d14986868.img
Agileref	32246MB	8	0004fb00001200000524245246ffab36.img sistema
ctasymposium	2048MB	2	0004fb00001200002a09ec084074f457.img
villaserver	8196MB	4	80b6898fc0f54ae2b28fa0c360d5cf00.img

Esempio tabella caratteristiche macchine virtuali

Come già accennato è stata redatta una lista che identifica tutti i file .img appartenenti a ciascuna macchina virtuale.

Questa operazione molto impegnativa ha consentito la migrazione seriale delle macchine virtuali da OVS a Proxmox VE nel più breve tempo possibile.

Le macchine virtuali sono state migrate una alla volta, per la grande dimensione dei file da copiare e soprattutto perché i servizi erogati da un particolare server sono rimasti interrotti per tutta l'operazione.

Innanzi tutto si è spenta la macchina virtuale da migrare in funzione su OVS.

Lo spegnimento è stata una operazione indispensabile per evitare il rischio di avere uno stato difettoso del sistema operativo della macchina virtuale.

Successivamente si sono copiati tutti i file .img associati alla macchina dal nodo OVS che li conteneva al nodo di Proxmox che li avrebbe importati.

Poi sul suddetto nodo di Proxmox è stata creata una macchina virtuale con le stesse caratteristiche fisiche della macchina esportata da OVS, in particolare è stato creato anche un disco (o più) delle dimensioni approssimative di quello esportato da OVS.

Tale disco è stato poi dissociato dalla nuova macchina virtuale ma ne è stato annotato il nome.

Poi è stato convertito il disco importato da OVS su Proxmox dal formato RAW al formato QCOW2. Questa operazione non strettamente necessaria è stata eseguita per fornire una migliore compatibilità col nuovo sistema.

Infine l'operazione più delicata, l'importazione del disco QCOW2 nella nuova macchina virtuale col nome del disco che era stato creato e poi dissociato.

In alcuni casi<sup>5</sup>, all'accensione della nuova macchina virtuale (copia di quella esportata da OVS) il sistema operativo Linux ha segnalato la presenza di hardware sconosciuto e questo ha creato un blocco del processo di boot segnalando un errore in initramfs<sup>6</sup>. In questo caso è stato necessario usare il tool dracut per creare un nuovo initramfs corretto.

L'installazione e le opzioni di configurazione di Proxmox usati per questo primo cluster sono molto simili a quelle usate per il sistema finale quindi per ora verrà omessa una loro descrizione accurata rimandando l'argomento al paragrafo relativo al cluster di Proxmox finale. Verranno però indicati i comandi di importazione di una macchina virtuale su un nodo del cluster Proxmox.

ci si sposta nella cartella di importazione sul nodo dove è stata creata la nuova macchina virtuale cd /rpool/data/

si importa il file xxxx.img

<sup>&</sup>lt;sup>5</sup> Per i sistemi operativi Centos 6 e Centos 5

<sup>&</sup>lt;sup>6</sup> initramfs viene usato in fase di accensione della macchina per caricare il kernel e contiene, fra le altre cose, la descrizione dei device fisici

rsync -P tacchini@moon.hide.bo.iasf:/home/tacchini/xxxx.img .

si converte nel formato qcow2 qemu-img convert -f raw -O qcow2 xxxx.img xxxx.qcow2

*si importa nella macchina virtuale* qm importdisk 104 xxxx.qcow2 oasnfs1

Per quel che riguarda l'ultimo comando, il comando di importazione vero e proprio, si può notare che <u>104</u> è l'id della macchina virtuale nuova e <u>oasnfs1</u> è lo storage in cui è stato creato.

Il comando è necessario perché il file system di Proxmox è basato su un database (database-driven) quindi non si può semplicemente copiare. Terminata l'importazione si può cancellare il file del disco virtuale che si era creato insieme alla nuova macchina virtuale e che non può più essere usato.

## PROXMOX VE

Proxmox Virtual Environment è un hypervisor, ovvero un sistema di virtualizzazione, basato su Linux che può essere installato sia su sistemi Debian che bare metal<sup>7</sup>.

Con un hypervisor è possibile risparmiare energia e ridurre i costi legati all'IT, inoltre si introduce una grande flessibilità di utilizzo e la possibilità di effettuare backup di macchine intere.

È un software open source quindi non c'è un costo per installarlo ed usarlo, ma c'è per avere supporto e per poter accedere agli aggiornamenti<sup>8</sup>.

Un server ospite su cui viene installato Proxmox viene chiamato nodo ed è possibile installarlo in un cluster di più nodi<sup>9</sup> in modo da avere Alta Disponibilità<sup>10</sup>.

Le sue caratteristiche principali sono le seguenti:

<sup>9</sup> almeno 3 nodi

<sup>&</sup>lt;sup>7</sup> installato direttamente sull'hardware senza avere già un sistema operativo

<sup>&</sup>lt;sup>8</sup> sono disponibili comunque gli aggiornamenti Linux

<sup>&</sup>lt;sup>10</sup> HA High Availability

#### Macchine virtuali di tipo KVM

La tecnologia Kernel-based Virtual Machine è molto efficiente in quanto opera come un modulo del kernel Linux della macchina ospite. Questo permette delle performance quasi uguali al sistema puro nel caso si abbiano CPU con supporto nativo alla virtualizzazione tipo le tecnologie Intel VT-x o AMD-V.

Ogni macchina virtuale emula il proprio hardware: CPU, scheda grafica, scheda di rete, ecc.

Le macchine virtuali così ottenute sono delle instance separate le une dalle altre.

#### Linux Containers (LXC)

Un tipo di virtualizzazione che opera a livello di sistema operativo e non di kernel.

È una tecnologia leggera ed agile in quanto condivide il kernel col sistema ospite.

Usata soprattutto per creare instance legate ad una applicazione. Per esempio un server web creato in un container è più isolato rispetto

al sistema ospite e questo fornisce maggiore sicurezza.

#### Sistema di Management Centralizzato

Il sistema di management di Proxmox VE è accessibile interamente da interfaccia web, sebbene esista anche l'interfaccia CLI<sup>11</sup>, da cui è possibile realizzare tutte le operazioni più importanti.

Per esempio è possibile creare e gestire le macchine virtuali ed i container, migrare le macchine da un nodo ad un altro, definire le attività di alta disponibilità, aggiungere e gestire diversi tipi di storage, aggiornare un nodo ospite, effettuare backup o live snapshot di macchine virtuali.

Realizza un design multi master, significa che è possibile gestire tutto il cluster collegandosi ad uno qualunque dei nodi che lo compongono.

<sup>&</sup>lt;sup>11</sup> Command Line Interface

#### Cluster HA<sup>12</sup>

Un cluster di più nodi di Proxmox VE può realizzare una infrastruttura ad alta disponibilità in maniera nativa e semplice da configurare. È sufficiente indicare nelle caratteristiche descrittive di una macchina virtuale l'opzione di Alta disponibilità desiderata, ad esempio si può migrare la macchina ad un altro nodo in caso di guasto su quello che la ospita.

#### Networking

Proxmox VE virtualizza l'infrastruttura di rete presentata alle macchine virtuali su ogni nodo.

La simulazione è completa e consente di creare reti virtuali anche molto complesse.

Implementa un modello di networking di tipo bridged, il che significa che vengono realizzati via software degli switch uguali a quelli fisici e per collegarsi al mondo esterno questi vengono associati alle interfacce fisiche del nodo ospite.

#### Firewall

Viene realizzato un servizio completo di Firewall in modo distribuito tra i nodi, in questo modo è possibile inserire regole a livello di singolo nodo ma anche di singola Macchine Virtuale/ Container.

#### Backup

Sono supportati in modo nativo sia backup a freddo che snapshot di macchine virtuali e container.

Si possono programmare i backup in modo automatico e gli storage supportati sono di tutti i tipi (anche se per i live snapshot è consigliato usare Ceph).

Esiste anche un prodotto dedicato alla gestione dei backup in generale che è facilmente integrabile con Proxmox VE.

<sup>&</sup>lt;sup>12</sup> High Availability

La funzione di live snapshot è cruciale per quelle macchine mission critical che devono garantire un downtime sotto il 5%.

#### Storage

La gestione dello storage è molto flessibile, le macchine virtuali possono risiedere sullo storage locale di un nodo oppure su uno storage condiviso, per esempio un NFS visibile da tutti i nodi.

A livello locale viene supportato, oltre che la normale struttura a directory del filesystem, anche LVM<sup>13</sup> e ZFS<sup>14</sup>.

L'utilizzo di storage in locale ha la limitazione di non permettere la migrazione in tempo reale delle macchine virtuali, quindi di fatto impedisce l'Alta Disponibilità.

Per sviluppare in pieno il potenziale di questo hypervisor è consigliato usare uno storage condiviso.

Le tecnologie di storage condiviso supportate sono: iSCSI, NFS, SAMBA, Ceph RBD<sup>15</sup>. Ed i file system condivisi GlusterFS e CephFS. Soprattutto Ceph come file system si è rivelato uno strumento molto potente per gestire i dati delle macchine virtuali.

È un file system distribuito e replicato che sfrutta i dispositivi fisici che stanno sui singoli nodi ma non solo, può utilizzare anche macchine dedicate allo storage e presentarlo a tutti i nodi del cluster.

È flessibile, sicuro e scalabile.

## INSTALLAZIONE NUOVO CLUSTER PROXMOX

Sono state acquistate 3 macchine per realizzare un cluster di Proxmox VE che possa garantire Alta Disponibilità e che consenta l'utilizzo di CephFS.

È stato acquistato anche uno switch Netgear XSM4324CS switch 24 porte 10 Gbs per realizzare una sottorete privata di collegamento tra i nodi.

<sup>&</sup>lt;sup>13</sup> Logical Volume Manager

<sup>&</sup>lt;sup>14</sup> Zettabyte file system

<sup>&</sup>lt;sup>15</sup> Ceph Rados Block Device

Le caratteristiche per ogni nodo sono le seguenti:

- DELL POWEREDGE R740xd;
- Doppio processore Intel Xeon Gold 6238R 2.2G, 28C/56T, 10.4GT/s, 38.5 M Cache per un totale di 56 Core;
- 128 GB di RAM;
- 2 dischi ssd da ....per il sistema operativo;
- 3 dischi Intel s4610 ssd 1,92TB per realizzare il sistema Ceph;
- Una scheda di rete INTEL X550T2 doppia porta a 10Gbs;

Lo schema seguente raffigura il sistema realizzato:



#### Installazione di Proxmox VE

Il software nella versione 6.4-8 è stato installato mediante una chiavetta usb eseguibile su ognuno dei 3 nodi identici.

Sono stati scelti i due dischi SSD da 480 GB per il sistema ed è stata scelta l'opzione di configurazione zmirror

	Hard	disk options	_ 8 X
k	Filesystem	zfs (RAID1)	
	Note: ZFS is not com controllers, for detail	patible with hardware RAID Is see the documentation.	
	Disk Setup	Advanced Options	
Har	ddisk 0 do not u	se 🔻	
Har	ddisk 1 /dev/sdb (4	480Gib, QEMU HARDDISK) 👻	
Har	ddisk 2 /dev/sdc (4	180Gib, QEMU HARDDISK) 👻	
			ОК

Lo z mirror è un raid 1 creato da ZFS, quindi ha tutte le opzioni disponibili di ZFS (sono state lasciate le impostazioni di default). Successivamente si sono impostate la lingua, la zona geografica ed il layout di tastiera. Poi la password di amministratore ed un indirizzo email di riferimento, l'email è obbligatoria per procedere.

L'ultimo passo è la configurazione della rete. Sebbene si consigliato di separare la rete che collega verso l'esterno, che usa l'interfaccia in modalità bridge, la rete che realizza il cluster (usa Corosync e necessita di bassa latenza) e la rete che collega lo storage (nel nostro caso Ceph), si sono lasciate le prime due su un unico collegamento fisico perchè

tutto sommato Corosync richiede un collegamento stabile ma poca banda e viene utilizzato uno switch fisico dedicato a questi collegamenti. Questo switch è collegato poi al centro stella del Centro di Calcolo.

<b>XPROXMO</b> Management	Proxmox VE Installer
Please verify the displayed network configuration. You will need a valid netwo configuration to access the management interface after installing. After you have finished, press the Next bu You will be shown a list of the options that chose during the previous steps.	<ul> <li>IP address (CIDR): Set the main IP address and netmask for your server in CIDR notation.</li> <li>Gateway: IP address of your gateway or firewall.</li> <li>DNS Server: IP address of your DNS server.</li> </ul>
Management Interface:	ens18 - 7e:5b:c3:c2:19:88 (virtio_net) 🕶
Hostname (FQDN):	pve.iasfbo.inaf.it
IP Address (CIDR)	192.168.100.2 / 24
Gateway:	192.168.176.253
DNS Server:	192.167.166.17
Abort	Previous Next

Terminata l'installazione su ogni nodo si può accedere, mediante browser, all'interfaccia grafica di controllo.

$\leftarrow$ $\rightarrow$ C $\textcircled{a}$	🗘 🖧 ≓ htt	ps://vs3.hide. <b>bo.iasf</b> :8006	/#v1:0:18:4:5:5		\$ ₩
🔁 Zimbra Web Client Sign   🐰	ogin to Webmin 📁 I	NAF OAS Bologna - O	🖬 pfSense - Login  xvs1 - Proxmox Virtual E 🗙 vs3 - Proxmox Virtua	al E 🔥 LDAP Account Manager 🛛 Amministrazione Zimbra 🍺 Servi	zi 😈 UniFi Network 🛛 🚿 🗅 Altri segnalil
	wironment 6.4-8 Search				Documentation Create VM Create CT 🛓 root@pam
Server View		Datacenter			@ Hel
Datacenter (OASCNR)		C Secret			
> 🛃 vs3			Health		
VS4		Summary			
107 (Zimbra)		Gluster	Status	Nodes	Ceph
112 (Workstation-calcolo)		(g) Ceph			
120 (baculatest)		Options		✓ Online 3	
450 (WEB-OAS-USERS-Cr	entOS7)	Storage		× Offline 0	
Ceph_sis_datifast (vs4)		🖺 Backup	Cluster: OASCNR, Quorate: Yes		HEALTH_OK
Cleant (vs4)		ca Replication			
I local (VS4)		Permissions			
as-proxmox-ibm (vs4)		. Users	Guests		
svmstore (vs4)		O ADI Talana			
> ស vs5		A API TOKENS	Virtual Machines		LXC Container
SASTRI-m-ICT		e Groups	Running	6	
Macchine-Galcolo     Be Reel CAS		Pools	O Stopped	17 Unning Stopped	0
• 1 001 01 ID		Roles	O remplated		
		a, Authentication			
		😻 HA 🛛			
		ACME			
		Firewall			
		Lui Metric Server	Resources		
		G Support	CPU	Memory	Storage
			27/	201	0404
			0%	8%	31%
			0%	<b>8%</b>	31%
Tasks Cluster log			0% of 336 CPU(s)	8% 103.71 GB of 1.29 TB	31% 5.70 TB of 18.15 TB
Tasks Cluster log Start Time J End Time	Node	User name	0% of 336 CPU(a) Description	8% 103.71 GB of 1.29 TB	31% 5.70 TE of 18.15 TE Status
Tasks     Cluster log       Start Time ↓     End Time       Jan 13 04:58:45     Jan 13 04:51	Node 58:49 vz3	User name root@pam	0% of 336 CPU(a) Description Update package distatase	8% 103.71 GB of 1.29 TB	31% 5.70 TIB of 18.15 TIB Status OK
Tasks         Cluster log           Start Time J.         End Time           Jan 13 04:58:45         Jan 13 04:51           Jan 13 03:54:57         Jan 13 03:54	Node 58:49 vs3 55:00 vs4	User name root@pam root@pam	0% el 336 CPU(s) Description Update perchage database Update perchage database	8% 103.71 GB of 1.29 TB	31% 5.70 TB of 18.15 TB Status OK
Tasks         Cluster log           Start Time J         End Time           Jan 13 04:58:45         Jan 13 04:58:45           Jan 13 03:55:7         Jan 13 03:35:43           Jan 13 03:35:43         Jan 13 03:35:43	Node 58:49 vs3 55:00 vs4 55:46 vs5	User name root@pam root@pam root@pam	0% of 336 CPU(s) Description Update peakage database Update peakage database	8% 103.71 GB of 1.29 TB	31% 5.70 TE of 18.15 TB Status ОК ОК
Tasks         Cluster log           Start Time J         End Time           Jan 13 04:58:45         Jan 13 03:54:57           Jan 13 03:54:57         Jan 13 03:54:30           Jan 13 03:54:30         Jan 13 03:31           Jan 12 10:21:03         Jan 12 10:33	Node           58:49         vs3           55:00         vs4           55:46         vs5           55:02         vs3	User name root@pam root@pam root@pam root@pam	0% of 336 CPU(a) Description Update package database Update package database Update package database Shell	8% 103.71 GB of 1.29 TB	31% 5.70 TB of 18.15 TB Язана ОК ОК ОК

In fase di installazione ci si è assicurati di abilitare la porta seriale in modo da avere disponibile anche xterm.js come console (può essere utile per il copia/incolla dei comandi).

$\leftarrow \rightarrow \mathbf{C}$	0	A ≈ https://	vs3.hide. <b>bo.iasf</b> :8006	/#v1:0:=node%2Fvs4:4:5::::::5	☆		⊚ ≡
Zimbra Web Client	t Sign 🐰 Login to We	ebmin 🥵 INAF	OAS Bologna - O	🛿 pfSense - Login 💥 vs1 - Proxmox Virtual E 💥 vs3 - Proxmox Virtual E 🍌 LDAP Account Manager 🔁 Amministrazione Zimbra 🍺 Servizi 🛄 UniFi Network		>> 🗅 Alt	tri segnalibri
× PRO×MO	Xirtual Environment 6	6.4-8 Search		a Documentation 📮	Create VM	Create CT	root@pam 🗸
Server View		~	Node 'vs4'	C Reboot (c) Shutdown	>_ Shell 🗸	Bulk Actions 🗸	@ Help
vers not		Node Vs4* Q Search Sammary Notes Shell Q System Hotwork Certificates Q DNS Hosts	CPU usage         0.10% of 12 CPU(a)         0 ID delay           © CPU usage         0.10% of 12 CPU(a)         0 ID delay           Image         0.11% (3.00 GB of 440.27 GB)         KSM sharing           Image         7.18% (3.00 GB of 10.42 GB)         2 SWAP usage           CPU(a)         112 x Intel(3) Xeon[3] C         112 x Intel(3) Xeon[3] C	Shell of NVC	i Buk Actons ↓	0.00% 0 B N/A	
<ul> <li>■ oas-proximo;</li> <li>■ overstore (v</li> <li>&gt; vs5</li> <li>&gt; ASTRI-m-ICT</li> <li>&gt; Macchine-Cabio</li> <li>&gt; Pool-CAS</li> </ul>	k-ibm (vs4) (sc4)		O Time     E Syslog     O Tome     Syslog     O Updates     Firowall     Disks     UM     LVM     LVM     Directory     ZFS     Geph     Papelication	CPU Urago	19-1 (Tue, 01 J pve-	PU usage IIO o	delay
Tasks Cluster log			Task History Subscription	0.03 - 0.022-01-13 2023-01-13 202	23-01-13 2023 2:15:00 12:1	-01-13 2023-01-13 9:00 12:23:00 • Load ave	3 2023-01-1 12:27:00 Hage
Gluster tog							
Start Time 🤳	End Time	Node	User name	Description		Status	
Jan 13 04:58:45	Jan 13 04:58:49	vs3	root@pam	Update package database		ок	
Jan 13 03:54:57	Jan 13 03:55:00	vs4	root@pam	Update package database		OK	
Jan 13 03:35:43	Jan 13 03:35:46	vs5	root@pam	Update package database		OK	
Jan 12 18:21:03	Jan 12 18:45:02	vs3	root@pam	Shell		ок	
Jan 12 18:15:19	Jan 12 18:15:20	vs5	root@pam	VM 901 - Stop		ок	

Per la creazione di una macchina virtuale ci sono alcune accortezze da tenere presente. Per lo storage normalmente si usa SCSI o VirtIO Block, se si dovesse installare una macchina Windows è opportuno impostare IDE (per via dell'installer di Windows) e poi eventualmente cambiare dopo. Si può scegliere già in questa fase se effettuare un backup regolare del disco.

Create: Virtual M	lachine				$\otimes$
General OS	System Hard Disk	CPU	Memory Network	Confirm	
Bus/Device: SCSI Controller: Storage: Disk size (GiB): Format:	SCSI V 0 IDE SATA VirtIO Block SCSI Raw disk image (raw)	<ul> <li></li> <li><th>Cache: Defa</th><th>ault (No cache)</th><th>~</th></li></ul>	Cache: Defa	ault (No cache)	~
SSD emulation:			Backup:		
IO thread:			Skip replication:		
Read limit (MB/s):	unlimited	$\hat{}$	Read max burst (MB):	default	0
Write limit (MB/s):	unlimited	$\hat{}$	Write max burst (MB):	default	0
Read limit (ops/s):	unlimited	0	Read max burst (ops):	default	0
Write limit (ops/s):	unlimited	0	Write max burst (ops):	default	0
Help			Ad	dvanced 🖂 Back	Next

Per quel che riguarda la CPU andrebbero sempre scelti almeno due Core e tenuta sempre abilitata l'opzione NUMA<sup>16</sup>

Create: Vir	tual M	achine					$\otimes$
General	OS	Syster	n Hard Disk	CPU	Memory Net	work Confirm	
Sockets:		1		0	Туре:	Default (kvm64)	$\sim$
Cores:	[	2		$\bigcirc$	Total cores:	2	
VCPUs:		2		0	CPU units:	1024	$\bigcirc$
CPU limit:		unlimited		$\hat{\mathbf{v}}$	Enable NUMA:		
Extra CPU F	lags:						
Default	-00	+	md-clear	Required to	o let the guest OS k	know if MDS is mitigated correctly	
Default	-00	+	pcid	Meltdown Intel CPUs	fix cost reduction or	Westmere, Sandy-, and IvyBridge	
Default	-00	+	spec-ctrl	Allows imp	roved Spectre mitig	ation with Intel CPUs	
Default	-00	+	ssbd	Protection	for "Speculative Sto	ore Bypass" for Intel models	
Default	-00	+ 0	ibpb	Allows imp	roved Spectre mitig	ation with AMD CPUs	
Default	- 00	+	virt-ssbd	Basis for "	Speculative Store B	ypass" protection for AMD models	
Help						Advanced 🖂 Back Ne	ext

<sup>16</sup> Non Uniform Memory Access

Nel caso ci siano problemi di installazione, per sistemi operativi ostici o perché si ha un cluster Proxmox molto disomogeneo, andrebbe settato il campo Type con la famiglia di processori fisici presenti sul nodo di Proxmox.

Create: Vir	tual Machine					(S) Isage
General	OS Syste	em Hard Disk	CPU	Memory Netwo	rk Confirm	10490
Sockets:	1		0	Туре:	Default (kvm64)	✓ Linux
Cores:	2		$\bigcirc$	Total cores:	Model	Vendor
					athlon	AMD
VCPUs:	2		0	CPU units:	EPYC	AMD
CPU limit:	unlimite	d	0	Enable NUMA:	EPYC-IBPB	AMD
<b>F</b>	-,		-		EPYC-Rome	AMD
Extra CPU F	lags:	md alaar	Deguired to		Opteron_G1	AMD
Delault	- ())) +	ma-clear	Required to	let the guest OS kh	Opteron_G2	AMD
Default	+	pcid	Meltdown fi	ix cost reduction on \	Opteron_G3	AMD
Dofault		ence etri	Allows imp	roved Spectre mitigat	Opteron_G4	AMD
Delault	- ())) +	spec-cili	Allows Imp	oved Spectre mitigat	Opteron_G5	AMD
Default	+	ssbd	Protection f	for "Speculative Store	phenom	AMD
Default	+	ibpb	Allows impl	roved Spectre mitigat		Intel
Default	- () () +	virt-ssbd	Basis for "S	Speculative Store Byp	ass" protection for AMD models	3 2023-01-13 2023-01
🚱 Help					Advanced 🗹 🛛 Back 🛛 Nex	t 13:04:00 13:08:

Il backup si può effettuare manualmente, usando una delle tre modalità: macchina ferma, snapshot, o macchina in sospensione; oppure si può schedulare attraverso l'apposita voce in Datacenter (le risorse/attività configurate in Datacenter riguardano il cluster intero).

$\leftrightarrow$ $\rightarrow$ C $\textcircled{a}$	0	🗛 🛋 https:/	/vs3.hide. <b>bo.iasf</b> :8006/	/#v1:0:18	4:=hosts:=	content	Backup:::21:	21				1	3	⊚ ≡
Z Zimbra Web Clien	it Sign 🐣 Login to W	ebmin 📁 INAF	OAS Bologna - O	pfSense	- Login  🗙	vsl - Pr	oxmox Virtua	l E X vs3 -	Proxmox Vi	rtual E 🔥 L	DAP Account Manager 🛛 Amministrazione Zimbra 🌔 S	iervizi 过 UniFi Network	>> 🗅 Altri	segnalibri
	X Virtual Environment 6	5.4-8 Search										Documentation Create VM	🕤 Create CT 🔒 ro	iot@pam 🗸
Server View		~	Datacoptor											<ul> <li>Help</li> </ul>
■         Datasetter (0A9           ■         0 (10) (Preser           100 (10) (Preser         100 (Preser           110 (17) (10) (Preser         110 (17) (17) (17) (17) (17) (17) (17) (17)	CNR)  CNR)  res  res  AP  AP  AP  AP  AP  AP  AP  AP  AP  A		Q. Search Summary C. Search Ceph Ceph Coptions Songe Backup Demissions Coptions Backup Demissions Copti	Add	Remove d Node - All -	Eat	Job Detail Day of week Sunday Saturday	Pun now	Star 21:00 00:00	Storage cephs evmstore	Selection 59.1502.503.530,15001,15003,15050,15254 110	Some guests are not control of the second	ered by any backup Job	Show
Taske Chuster Inc														_
Start Time	End Time	Node	Liser name	Donorinti	20								Status	
Jap 13 14:02:22	Log 12 14:02:24	14006	root@nam	Eraco da	to.								OK	
Jan 13 14:03:22	Jan 13 14:03:24	vso vs3	root@pam	Erase da	ta								OK	
Jan 13 14:03:12	Jan 13 14:03:12	vs3	root@pam	Frase da	ta								OK	
Jan 13 14:03:06	Jan 13 14:03:06	vs3	root@pam	Erase da	ta								ок	
Jan 13 13:50:42	Jan 13 13:50:43	vs3	root@pam	VM/CT 9	01 - Console								Error: Failed to run vi	ncproxy.
100716	10100110				20110014									

La configurazione del backup è abbastanza complicata da riassumere ma, per brevità, è sufficiente prestare attenzione allo storage che verrà usato (per evitare di saturarlo) ed alla modalità di backup: *snapshot* per macchine che devono restare accese e *stop* per macchine che possono essere spente e per cui l'integrità del backup è importante. Proxmox consente una gestione dei permessi molto estesa, si articola su utenti, gruppi, e sui ruoli che si possono avere sui pool di risorse (come lo storage o le macchine virtuali intere). Si configura sempre da Datacenter.

← → C ⋒	0	A = https:	//vs3.hide. <b>bo.iasf</b> :8006/	/#v1:0:18:4:=hosts:=contentBackup:::21::6		☆	
7 Zimbra Web Clie	nt Sian 🐴 Login to V	Vebmin 赌 INAI	E OAS Bologna - O.	nfSense - Login 🕱 vs1 - Proxmox Virtual E 🕱 vs3 - Proxmo	x Virtual E A LDAP Account Manager 2 Amministrazione Zim	ara 🔹 Servizi 👖 UniFi Network	>> C Altri sennalibri
	Virtual Environment	6,4-8 Search				Documentation Create VM	Create CT 🔒 root@pam 🗸
Server View		~	Datacenter				@ Help
Server View	SCNR) arcso) LDAP LDAP S2mbanew) s2mbanew) s2mbanew) s2mbanew) s2mbanew) s2mbanew) s2mbanew) s2mbanew) s2mbanew) s2mbanew) s3) bmorrMACS) constant(s3) (vs3) vs3) sabate vs3) sabate sab		Datacenter Q. Seen,	Add         Permove           Path         1           (pool/ASTR-rh-GT         0           (pool/ASTR-rh-GT)         0           (	UserGroupAPI Tolen atrime@pwe mjpol@pwe cochrigPywe @cAS	Rolo PVEVMAdmin PVEVMAdmin PVEVMUser Administrator PVEDatastoreUser	Propag true true true true
103 (Agiler 113 (testor 114 (Calco	ref2) entos7) olo-Vergani)						
Start Time	Fod Time	Nada	Llear name	Description			Statua
Jan 12 14/02/02	Lng time	NODE	user name	Eress dete			31dius
Jan 13 14:03:22	Jan 13 14:03:24	vs3	root@pam	Erase data			л. Л
Jap 13 14:03:17	Jan 12 14:03:23	voJ	root@pam	Erase data			OK
Jap 13 14:03:12	Jap 13 14:03:02	vs0 v#3	root@pam	Liaon uala			OK
Jan 13 14:03:00	Jan 12 12:50:42	100	root@pam	Linese Gata			Error: Enilod to run unenreau
Jan 13 13:30:42	aan 1a 1a:au:43	V83	roodepam	VINCOL DOL - COUSOIR			anon, railed to run vncproxy.

L'alta disponibilità su Proxmox (HA<sup>17</sup>) è basata su Corosync e richiede almeno 3 nodi per funzionare. Il quorum per un cluster di 3 nodi come il nostro è di 2.

C'è un meccanismo di fencing integrato (Softdog Kernel o Watchdog Kernel) per gestire i guasti ai nodi ma è possibile usare anche meccanismi esterni.

Anche l'HA si configura da Datacenter.

#### Ceph

Ceph è un file system condiviso sviluppato da Red Hat ed è opensource. Questo sistema è costituito da uno storage distribuito a oggetti che mette a disposizione degli utenti un object storage, un block storage (RBD) ed un File System Posix.

<sup>&</sup>lt;sup>17</sup> High Availability

Ceph è dotato dai seguenti componenti:

- Monitor: per il controllo del cluster. Necessita di almeno 2GB di RAM. Ogni nodo è stato settato come monitor.
- OSD (Object Storage Device): sono i singoli hard disk. Per ottimizzare le prestazioni si è deciso di usare dischi SSD di classe datacenter e tutti uguali.
- Manager: insieme ai monitor monitorizza e gestisce il cluster. Nel nostro sistema è vs3-
- MDS (Meta Data Server): contiene i metadati di CephFS.
- Gateway: una interfaccia l'object storage e le applicazioni.

Alcune proprietà importanti di cui è dotato intrinsecamente Ceph sono lo Thin Provisioning, la Compressione dei dati, Snapshot, Copy on Write, Load Balancing adattivo, No Single Point of Failure (solamente con 3 o più nodi).

Prima di installare Cep si sono seguite alcune raccomandazioni.

Avere un cluster di almeno 3 nodi per realizzare la No Single Point of Failure.

Avere una sottorete dedicata ai dati realizzata con uno switch fisico a 10Gbs.

Usare dischi veloci evitando i dischi meccanici.

Dimensionare la RAM di ogni nodo in modo da soddisfare il requisito di 3-5 GB di RAM per TB (nel nostro caso significa almeno 30 GB di RAM a nodo).

Dimensionare la CPU in base alla formula: 2 GHz per ssd / velocità in GHz del singolo Core.

Non usare RAID ma presentare direttamente i dischi a Ceph.

Per l'installazione vera e propria si procede per ogni nodo perché non c'è un controllo centralizzato su Datacenter.

Si può procedere via CLI oppure via interfaccia grafica.

Si è scelta quest'ultima modalità.

Dalla voce Ceph si sceglie Start Installation, si sceglie la modalità advanced e si prosegue inserendo la subnet dedicata per Ceph ed impostando il nodo stesso come monitor. Si arriva alla situazione seguente:

Node 'vs4'							ť	C Reboot	() Shutdown	>_ Shell   v	🚦 Bulk Actions 🗸	Help	
ot System -	Monitor												
≓ Network	► Start	Stop	C Restart C	reate Destroy	Syslog								
Certificates	Name ↑	Name ↑ Host Status Address Version										Quorum	
DNS	mon.vs3 vs3 running 192.168.75.103:6789/0							15.2.13					
Hosts	mon.vs4		vs4	running		192.168.75.104:6789/0	15.2.13						
⊘ Time	mon.vs5 vs5 running 192.168.75.105:6789/0											Yes	
Syslog													
2 Updates													
♥ Firewall													
🖨 Disks 👻													
LVM													
LVM-Thin													
Directory	Managar												
II ZFS	Manager												
n Ceph -	Start	Stop	C Restart C	Destroy	Syslog								
Configuration	Name 🕆		Host	Status		Address		Version					
C Monitor	mgr.vs3		vs3	active		192.168.75.103		15.2.13					
🖨 OSD													
CephFS													
A Pools													
i≣ Log													
13 Replication													
i≣ Task History													
Subscription													
~													

A questo punto vanno inseriti gli osd per ogni nodo. Scegliere Ceph - - OSD - - Create OSD. Scegliere il disco e Bluestore per OSD type. La situazione finale:

Node 'vs4'									C	Reboot 🖒 Shutdov	vn >_ Shell	V Bulk	Actions $\lor$	Help
<b>o</b> <sup>®</sup> System →	Create	: OSD Mana	ige Global Flags						osd.8:	▶ Start Stop	C Restart	O Out	• In   =	E More 🗸
	Name	Class	OSD Type	Status	Version	weight	reweight	Used (%)	Total	Apply/Commit Latency (ms)				
Certificates	√ = default													
DNS	vs5				15.2.13									
Hosts	🖂 osd.8	ssd	bluestore	up 📀 / in 🔵	15.2.13	1.7466	1.00	18.23	1.75 TiB	2/2				
<ul> <li>Time</li> </ul>	🖂 osd.7	ssd	bluestore	up 🛇 / in 🔵	15.2.13	1.7466	1.00	20.45	1.75 TiB	1/1				
i≣ Syslog	⊖ osd.6	ssd	bluestore	up 🛇 / in 🔵	15.2.13	1.7466	1.00	20.14	1.75 TiB	1/1				
2 Updates	⊂ osd 5	eed	bluestore	un 🔿 / in 🗖	15.2.13	1 7466	1.00	19.19	1 75 TiB	1/1				
Firewall	⊖ osd.4	ssd	bluestore	up 🖓 / in 🔵	15.2.13	1.7466	1.00	21.28	1.75 TiB	2/2				
🖨 Disks 👻	⊖ osd.3	ssd	bluestore	up 🔾 / in 🔵	15.2.13	1.7466	1.00	18.34	1.75 TiB	1/1				
LVM	vs3				15.2.13									
LVM-Thin	🖂 osd.2	ssd	bluestore	up 📀 / in 🔵	15.2.13	1.7466	1.00	18.68	1.75 TiB	1/1				
Directory	🖂 osd.1	ssd	bluestore	up 🔾 / in 🔵	15.2.13	1.7466	1.00	19.78	1.75 TiB	1/1				
II ZFS	🖂 osd.0	ssd	bluestore	up 🔾 / in 🔵	15.2.13	1.7466	1.00	20.35	1.75 TiB	1/1				
🌘 Ceph 🚽														
Configuration														
D Monitor														
⊖ OSD														
CephFS														
A Pools														
i≣ Log														
+3 Replication														
I Task History														
Subscription														
- Caboan pilon														

Potrebbe essere necessario reinizializzare dei dischi già usati. Si usano i comandi:

dd if=/dev/zero of=/dev/sdx bs=1M count=200 ceph-zap /dev/sdx

Per avere lo storage disponibile su nodi è necessario creare un pool Da un qualsiasi nodo scegliere Ceph - - Pools - - Create, impostare il nome del pool, settare la size max e la size min (in pratica size max andrebbe settata pari al numero dei odei mentre la size min è la soglia minima di funzionamento, per 3 nodi andrebbe messa a 1 ma si è scelto comunque 2 per non rischiare perdita di dati e perché, data la struttura del cluster, è improbabile un malfunzionamento), sceglie la crush rule (c'è solo un'opzione), impostare il numero di Placement Group (PG) ( in genere il valore è 32 ma viene messo a disposizione un calcolatore online <u>https://ceph.com/pgcalc</u>), infine con Add Storage si aggiunge il pool come storage in Datacenter (è da Datacenter che si gestisce qualunque storage).

Il PG serve a ridurre la quantità di metadati associata agli oggetti sostanzialmente raggruppandoli in gruppi, è necessario un bilanciamento in quanto un alto numero di PG migliora la granularità ma aumenta il peso computazionale.

Si è deciso di sfruttare Ceph sia come File System che come RBD<sup>18</sup>. Il File System è stato chiamato CephFS e, per semplificare, viene utilizzato come un NFS molto performante (molto veloce e moto robusto), opera a livello di file.

<sup>&</sup>lt;sup>18</sup> Radon Block Device

Lo storage RBD è stato chiamato Ceph\_sis\_datifast ed anch'esso è condiviso tra tutti i nodi ma a livelli di Block Device, viene utilizzato prevalentemente per i dischi delle macchine virtuali.

Node 'vs4'					් Reboot ර් Sh	utdown >_ Shell   ~	Bulk Actions 🗸 🔞 Help
of: System	Create Edit Destroy						
≓ Network	Name	Size/min	# of Placement Groups	Optimal # of PGs	Autoscale Mode	CRUSH Rule (ID)	Used (%)
<ul> <li>Certificates</li> </ul>	device_health_metrics	3/2	1	1	on	replicated_rule (0)	94.41 MiB (0.00%)
O DNS	Ceph_sis_datifast	3/2	512	256	on	replicated_rule (0)	2.80 TiB (19.48%)
Hosts	cephfs_data	3/2	32	32	on	replicated_rule (0)	273.44 GiB (2.25%)
O Time	cephfs_metadata	3/2	32	16	on	replicated_rule (0)	152.11 MiB (0.00%)
Svelog							3.07 TiB
E Firewall							
Pirewall     P							
LVM-Inin							
Directory							
∎ ZFS							
🍘 Ceph 🤍 🤟							
Configuration							
🖵 Monitor							
🖨 OSD							
CephFS							
A Pools							
i≣ Log							
Replication							
Task History							
Subscription							

#### Per vedere lo stato di Ceph:

Server (w) i kokol   I Gabeer (k)   I (k) (ML DAP)   I (k) (ML DAP) <tr< th=""><th>XPROXMOX Virtual Environment 6.4-8 Search</th><th></th><th></th><th></th><th></th><th></th><th></th><th>@ Docum</th><th>tentation Create VM</th><th>🕤 Create CT</th><th>root@pam 🧹</th></tr<>	XPROXMOX Virtual Environment 6.4-8 Search							@ Docum	tentation Create VM	🕤 Create CT	root@pam 🧹
Image: Section (SADSR)       C. Section (SADSR)	Server View	Node 'vs3'						් Reboot ර	Shutdown >_ Shell   ~	Bulk Actions	V 😡 Help
I Source   I Sou	University Datacenter (OASCNR)	O Same	Health			s	Status				
I Di Presco       I Sammiy       I Sammiy       I Di Samiy       I Di Di Samiy       I Di Samiy       I Di Samiy <th>ം 🔂 vs3</th> <th>Q Search</th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th>	ം 🔂 vs3	Q Search									
Image: Notes       Image: Notes       State	100 (Presenze)	Summary	Status	Severity Summary			OSDs			PGs	
10 (spectramed) <ul> <li>Instance</li> <li>Insta</li></ul>	101 (IPA-LDAP)	Notes		No Warnings/Errors			In Out		active+clean:		577
In 10 (doncenterbasew)       If Sprint         In 200, If End TJ AUKS)       If End TJ AUKS         In 200, If End TJ AUKS)       If End TJ AUKS         In 201, If End TJ AUKS       If End TJ AUKS         In 201, If End TJ AUKS       If End TJ AUKS         In 201, If End TJ AUKS       If End TJ AUKS         In 201, If End TJ AUKS       If End TJ AUKS         In 201, If End TJ AUKS       If End TJ AUKS         In 201, If End TJ AUKS       If End TJ AUKS <t< td=""><td>109 (zimbranew)</td><td>&gt;_ Shell</td><td></td><td>no namigo croro</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>	109 (zimbranew)	>_ Shell		no namigo croro							
II 11 (dondZimbaroped)       II 12 (dondZimbaroped)         II 12 (dondZimbaroped)       II Control         II 13 (dondZimbaroped)       II Control         II 14 13 (dond       II Control         II 14 14 (dond       II Control         II 14 14 (dond       II Control         II 14 15 (dond)       II Control	110 (clonezimbranew)	😋 System 🗸									
I 100 (Instrumentation)       I Contrictates       I EALTH_CK         I 100 (Instrumentation)       I DBS         I Contrictates       I Hosts         I Contrictates	111 (clone2zimbranew)	≓ Network					U DOMINI U U				
<ul> <li>Discle (Transford AudroSa)</li> <li>Discle (Transford AudroSa)</li> <li>Gepti Ass_dafiad (util)</li> <li>Gepti Ass_dafiad (util)</li></ul>	115 (Idapnuovoprod)	<ul> <li>Certificates</li> </ul>	HEALTH_OK				Total: 9				
□ code, sta, statistic (nds)       □ fines         □ code (vds)       □ Time         □ code (vds)       □ Volatistic         □ Disk       □ Volatistic         □ Disk       □ Code (vds)         □ Code (vds)       □ Code (vds) <td< td=""><td>15003 (TBMICT7-MARKST) 15050 (TBMICT7-MATCS)</td><td>ONS DNS</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></td<>	15003 (TBMICT7-MARKST) 15050 (TBMICT7-MATCS)	ONS DNS									
Image: Image	Ceph_sis_datifast (vs3)	Hosts									
I clead (wS)       I gade         I clead (wS)<	cephfs (vs3)	⊘ Time									
2 Uddate       2 Uddate         0 Disk       0 Disk         0 Disk       0 Disk         0 Matchine-Catedo       0 Disk         0 Caph Version:       152.13         Performance       0 Caph Version:         0 Caph Version:       0 Caph Version: <tr< td=""><td>E local (vs3)</td><td>E Syslog</td><td></td><td></td><td></td><td>s</td><td>Services</td><td></td><td></td><td></td><td></td></tr<>	E local (vs3)	E Syslog				s	Services				
Ownstarts (vd.)     O     Freed     Image: matrix (vd.)     Monitors     Managers     Meta Data Servers       9 545     0 10%     0 10%     0 10%     0 0	as-proxmox-ibm (vs3)	C Updates									
Image: Section Sect	sovmstore (vs3)	♥ Firewall					Monitors	Ма	nagers	Meta Data Se	rvers
Image: Second	> 🛃 vs4	🖨 Disks 🗸					vs3: 🗸 vs4: 🗸 vs5: 🗸	v	s3: 🗸	vs3: 🗸	
LNAThine Calcido     Madedine Calcido     Nordination	> vs5	LVM									
<ul> <li>Red OAS</li> <li>Capit Variabit</li> <li>Capit Variabit</li></ul>	Macchine-Calcolo	C LVM-Thin	0.111.1.000								
E 275 Performance Coph	Pool-OAS	Dimetery	Cepit Version: 10.2.13								
Image: Solution of Coche     Coche       Coche     Coche       Monitor     Coche       Coche     Solution       Solution     Solution		<ul> <li>Directory</li> <li>350</li> </ul>	D (								
Image: Configuration     Image: Configuration     340 B/s       Image: Configuration     Image: Configuration     340 B/s       Image: Configuration     Image: Configuration     156.73 KB/s       Image: Configuration     Image: Configuration     Image: Configuration       Image: Configuration		■ ZPS	Performance								
© Configuration C Monitor G OSD © CopFS Usage Writes: E Log E Replication III Taak History © Subscription S CopS: Reads: 0 0 0 0 0 0 0 0 0 0 0 0 0		(ng) Ceph 🔍									
G Motor G OSD Ceph7S Usage Writes: E Log B Replication E Task Hatory C Subscription 3.06 TE or 15.72 TB CPS: Read3: COPS: N		Configuration			Reads:						340 B/s
G GOB CephS Usage Writes: 156.73 KB/s E Log Canaditation E Tank Heaton E Subscription 20% tops: Subscription 3.08 T6 of 15.72 T8 CPS: Virites: 15.73 KB/s 156.73 KB/s		Monitor									
Copris Usage Writes: 156.73 KB/s      Pools     Iso 1     Pools     Subscription     Subscrindin     Subscription     S		⊖ OSD									
A Pools     Unset     158.73 MB/s       III Log     158.73 MB/s     158.73 MB/s       III Task Hatory     20%     10PS: Reads:     0       3.08 TB of 15.72 TB     10PS: Writes:     0		CephFS	10	ones						~	
E tog     S Replation      E Tak History      O     Subscription      O     Subscription      O     Subscription      O     Subscription      O     Subscription      Subscrite      Subsc		A Pools		Jugo	Writes:					15	6.73 KiB/s
20% toP5: Subscription     20% toP5: Subscription     3:00 TE of 15.72 TE     10P5: Writes:     10P5:		I Log									
Subscription     S		n Replication									
© Subscription Reads: 0 3.08 TB of 15.72 TB IOPS: 0 Writes: 0		Task History	2	0%	IOPS:						
3.08 TB of 15.72 TB		Subscription	-	070	Reads:						0
KOPS: Writes:			3.08 TiB	of 15.72 TiB							
IOPS: 11					10.00						
					Writes:						11

Una nota importante è che se si deve intervenire su uno degli OSD (per esempio per sostituirlo in casi di guasto) oppure se si deve riavviare un nodo è necessario settare il/gli OSD con il flag noout per evitare rischi di malfunzionamenti.

Node 'vs3'											C	Reboot 🕐 Shutdow	vn >_ Shell   < 🚦 Bulk Actions < 🕜 Help
Q Search	Create: OSD Manage Global Flags										SD selected	► Start Stop	
Summary							weight						
>_ Shell	√≣ de	efault											
🚓 System 📼	×.	vs5		hluesters		15.2.13	1 7400	1.00		10.00	1 75 T.D	0.70	
≓ Network		osd.7	ssa	bluestore	up 😡 / in 🖝	15.2.13	1.7466	1.00		20.46	1.75 TiB	2/2	
Certificates		osd.6	ssd	bluestore	up 🕢 / in 🔵	15.2.13	1.7466	1.00		20.13	1.75 TiB	2/2	
ONS	~	vs4				15.2.13							
Hosts	1.6	Manage Gl	bal OSD Flags						$\otimes$	19.15	1.75 TiB	2/2	
<ul> <li>Time</li> </ul>										21.31	1.75 TiB	2/2	
I Syslog		Enable	Name	Description						18.35	1.75 TiB	2/2	
C Updates	×		nobackfill	Backfilling of PC	às is suspended.					18.69	1 75 TiB	2/2	
Firewall			nodeep-scrub	Deep Scrubbing	is disabled.					19.78	1.75 TiB	2/2	
🖨 Disks 👘			nodown	OSD failure rep	orts are being ignore	d, such that the m	ionitors will not	mark OSDs do	-	20.35	1.75 TiB	2/2	
LVM			noin	OSDs triat were	tomatically be marked o	ut will not be man	votigured interv	an mey start.	- 1				
LVM-Thin			norebalance	Rebalancing of	PGs is suspended.	su out anter the ot	ingureu interv	ai.	- 1				
Directory			norecover	Recovery of PG	s is suspended.								
III ZFS			noscrub	Scrubbing is dis	abled.								
@ Ceph			notieragent	Cache tiering ad	tivity is suspended.								
A Configuration			noup	OSDs are not a	lowed to start.				_				
Monitor			pause	Pauses read an	d writes.				- 1				
A OSD		Help						Apply					
CephFS								_					
- Pools													
i log													
* Beplication													
Task History													
Subscription													
Groupsonption													

Per concludere un'ultima osservazione, è risultato fondamentale aver pianificato in anticipo sia l'obiettivo di funzionamento che il dimensionamento dell'hardware.

Ha semplificato molto la messa in opera di tutto il sistema.