

Proxmox à UT1 retour d'expérience



Journée Capitoul du 15 décembre 2016

Philippe ORTH & Virginie GIROU

(Direction du Système d'Information)



Du modèle physique au IAAS

Ludovic
KVM

Virtualization

Philippe O.
PROXMOX + CEPH

Clusters

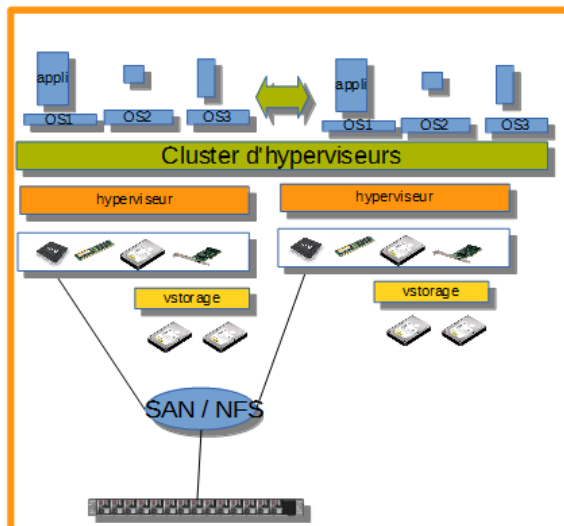
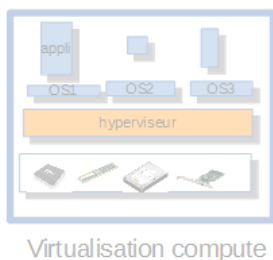
Philippe S.
OPENSTACK + CEPH

IAAS

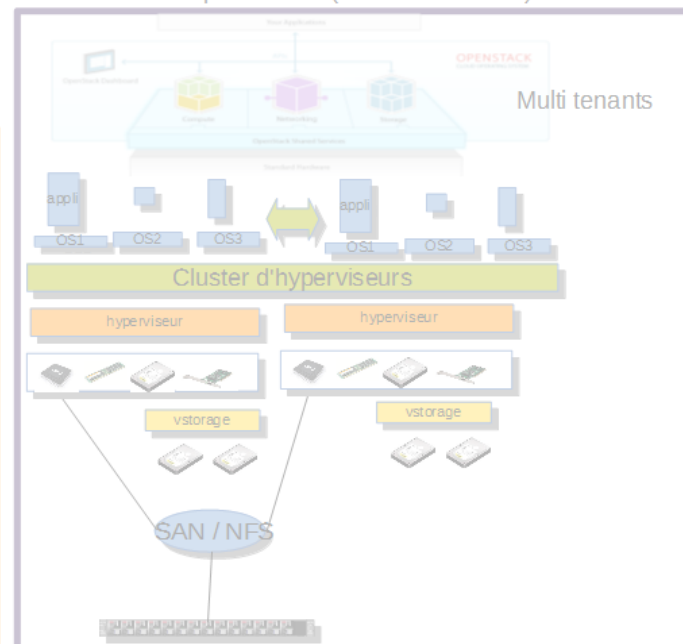
Compute : PROXMOX, Vmware VSPHERE
Storage : SAN, vSAN, CEPH
Network : Openvswitch (L2 + L3 + service)

Compute : KVM, ESXi
Storage : SAN, vSAN, CEPH
Network : vswitch (L2)

Compute :
KVM
HyperV
Vmware ESXi



Virtualisation
compute (cluster)
± storage
± network L2
→ contexte LAN



Virtualisation
compute (cluster)
± storage
± network L2 + L3
→ contexte LAN, WAN

- A. Pourquoi Proxmox ?
- B. L'infra de « production » (Proxmox + SAN)
- C. L'infra « système » (Proxmox + CEPH)



A - Pourquoi PROXMOX ?

- Un peu d'histoire à UT1:
 - ❑ ~2005 : 1^{ère} infra de virtualisation : VMWare ESX (3 nœuds)
 - ❑ 2007 : arrivée de XEN
 - ❑ 2013 : remplacement de XEN par Proxmox VE à base de KVM
- CF [présentation de F. Soulier \(capitoul du 16/10/2012\)](#)
- Aujourd'hui :
 - ❑ VmWare toujours là (3 nœuds) plus d'évolution
 - ❑ Proxmox :
 - ❑ plus de 160 VM sur 2 infras
 - ❑ Quelques projets dans les cartons...

A - Pourquoi PROXMOX ?

■ Proxmox c'est :

- ❑ Une solution « clé en main » pour virtualiser (V4.4 13/12/2016)
 - ❑ Basé sur Linux debian
 - ❑ KVM (émulation matérielle) et LXC (isolation) (remplace OpenVz)
 - ❑ Interface d'administration graphique intuitive (pas de java ☺)
 - ❑ La console des VM utilise spice ou html5 (toujours pas de java ☺)
- ❑ Un support applicatif réactif (en anglais ou en allemand)
- ❑ Un système de licence simplissime

■ Apports par rapport à KVM « seul »

- ❑ Gestion native mode cluster, multi-maître
- ❑ Gestion native des backups de VM
- ❑ Interface graphique : « vmotion » (VM et storage)
- ❑ Intégration du stockage CEPH

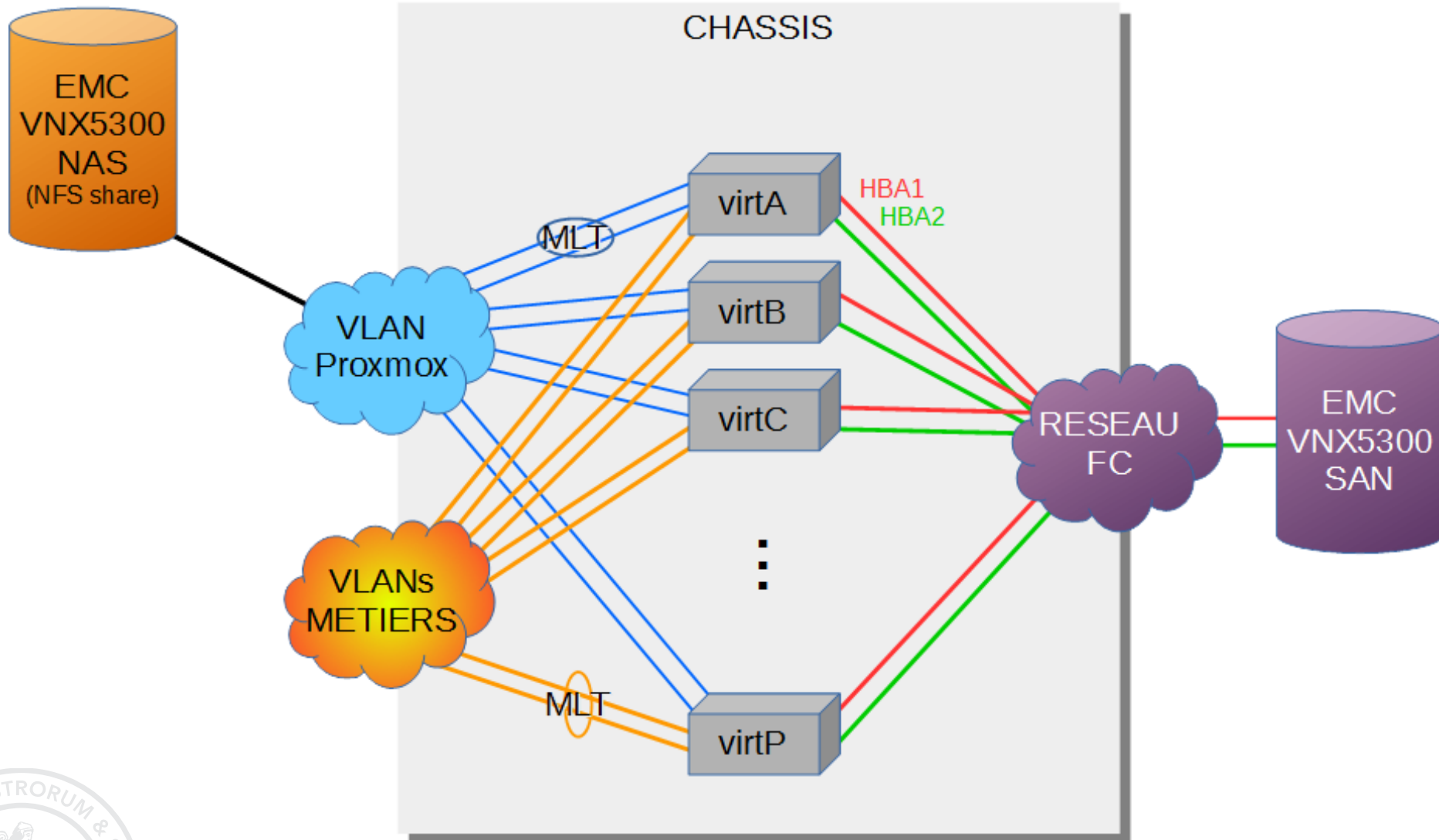


B - Le cluster de production

- Plus de 150 VM serveurs
- Un cluster de 16 nœuds
 - 1 Chassis Blade DELL M1000e
 - 16 lames DELL M630 (32 Cœurs et 256Go Ram par lame)
 - 2 commutateurs internes « 10G/40G ethernet » DELL MXL
 - 2 commutateurs internes « 16G Fiber Channel » BROCADE M6505
 - 32 cartes SD 32Go pour la partition système (!)
 - 16 licences proxmox
- 1 baie EMC VNX 5300 (SAN+NAS)
 - 1 volume en mode bloc (production)
 - 2 volumes NFS
 - Backup
 - Préproduction, tests, formations, etc (QCOW2)



B - Schéma architecture production



B - Interface administration Proxmox

PROXMOX Virtual Environment 4.3-12/6894c9d9 You are logged in as 'root@pam' [Help](#) [Create VM](#) [Create CT](#) [Logout](#)

Server View ▼

- Datacenter
 - virtf
 - virtg
 - virth
 - virti
 - virtj
 - virtk
 - virtl
 - virtm
 - virtn
 - virtu

Datacenter Search

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

Type ↑	Description	Disk usage...	Memory us...	CPU usage	Uptime
node	virtf	51.8 %	36.0 %	5.4% of 32CPUs	7 days 17:20:42
node	virtg	47.8 %	38.0 %	5.5% of 32CPUs	7 days 17:59:20
node	virth	52.0 %	71.8 %	9.4% of 32CPUs	7 days 17:57:30
node	virti	52.3 %	0.9 %	0.1% of 32CPUs	7 days 16:57:24
node	virtj	51.6 %	0.6 %	0.1% of 32CPUs	7 days 13:40:43
node	virtk	52.4 %	21.2 %	1.5% of 32CPUs	7 days 16:30:53
node	virtl	52.3 %	0.9 %	0.1% of 32CPUs	7 days 13:09:29
node	virtm	56.8 %	46.4 %	1.5% of 32CPUs	7 days 17:51:48
node	virtn	57.9 %	38.5 %	3.6% of 32CPUs	7 days 15:53:28
node	virtu	57.2 %	0.9 %	0.1% of 32CPUs	7 days 13:47:31
qemu	202 (redmine)		79.6 %	5.4% of 1CPU	7 days 16:19:37
qemu	207 (cas-test)		49.4 %	5.5% of 1CPU	7 days 16:19:48
qemu	218 (squidguard)		87.0 %	3.8% of 2CPUs	7 days 16:20:05
qemu	224 (ent-mysql-test)		35.7 %	4.9% of 1CPU	7 days 16:20:15
qemu	230 (jerry)		91.4 %	2.8% of 4CPUs	7 days 16:20:45
qemu	235 (ents1)		79.9 %	5.4% of 1CPU	7 days 16:20:57
qemu	236 (ents3)		80.0 %	6.0% of 1CPU	7 days 16:21:11
qemu	242 (esup-proxy-1)		68.3 %	6.0% of 1CPU	7 days 16:21:21
qemu	253 (ade-prod)				-
qemu	256 (ksup-prod)		92.2 %	15.2% of 2CPUs	7 days 16:21:53
qemu	261 (opera)		82.9 %	2.3% of 1CPU	7 days 16:22:09
qemu	264 (mimir)		62.0 %	1.6% of 1CPU	7 days 16:22:22
qemu	274 (idunn2)		93.9 %	0.4% of 8CPUs	7 days 16:23:04

Tasks Cluster log

Start Time ↓	End Time	Node	User name	Description	Status
Dec 15 03:30:03	Dec 15 03:30:21	virtn	root@pam	Update package database	OK
Dec 15 03:14:02	Dec 15 03:14:13	virti	root@pam	Update package database	OK
Dec 15 03:09:02	Dec 15 03:09:11	virtk	root@pam	Update package database	OK
Dec 15 02:28:02	Dec 15 02:28:18	virtu	root@pam	Update package database	OK
Dec 14 22:04:20	Dec 14 22:04:20	virtg	root@pam	VM 303 - Snapshot	Error: snapshot name 'J3' alr...

C - L'infra « système »

■ But

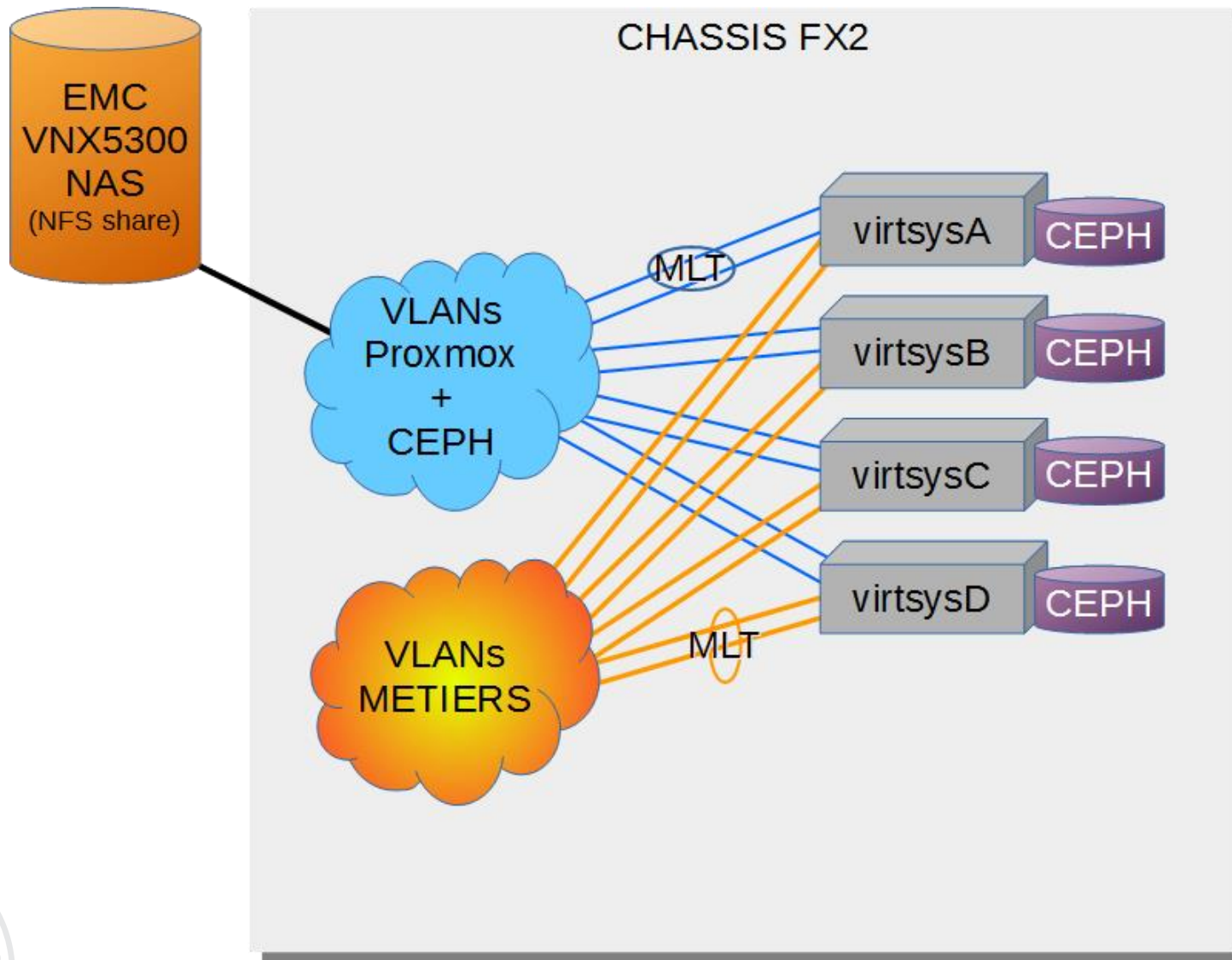
- ❑ Hébergement des VMs primaires (dns, dhcp, parefeu, etc)
- ❑ Etre indépendant de tout autre système (démarrage en 1^{er})

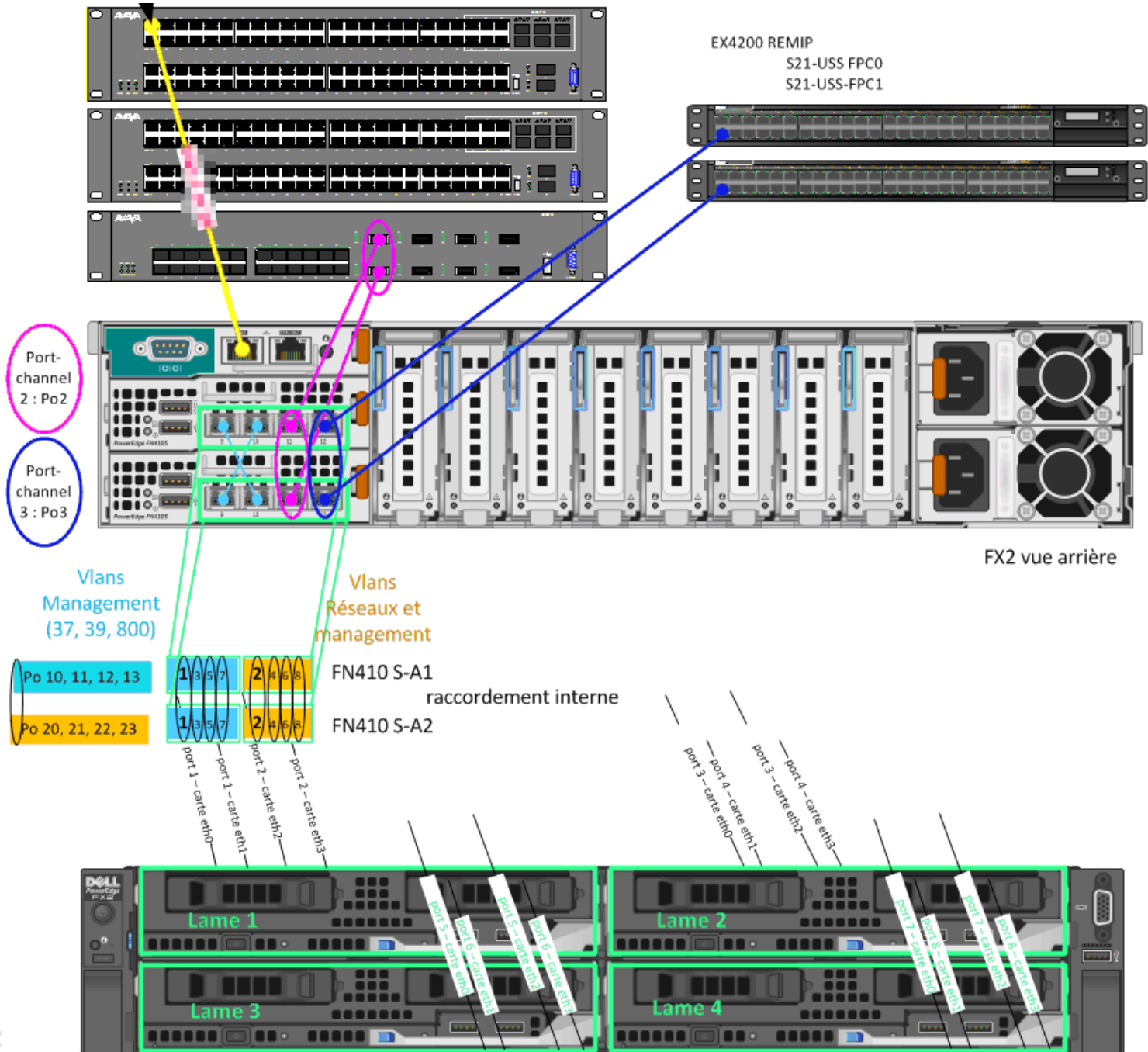
■ Chassis DELL FX2

- ❑ 4 lames FC630 avec pour chacune :
 - ❑ CPU : 48 cœurs (Xeon E5-2650 v4)
 - ❑ RAM : 128Go
 - ❑ Disques internes :
 - 2x 480Go ssd
 - 2x 800Go ssd
- ❑ 2 commutateurs 10G FN410s
 - ❑ 8 ports internes
 - ❑ 4 ports externes







C - L'infra « système » - schéma





Server View

Datacenter

- >  virtsysa
- >  virtsysb
- >  virtsysc
- >  virtsysd

Datacenter

Search

Summary

Options

Storage

Backup

Permissions

Users

Groups

Pools

Roles

Authentication

HA

Firewall

Support

Datacenter Health

Status



Cluster: cluster, Quorate: Yes

Nodes

✓ Online

4

✗ Offline

0

Ceph



HEALTH_OK

Guests

Virtual Machines

● Running

11

○ Stopped

0

LXC Container

● Running

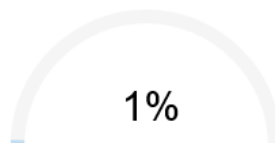
0

○ Stopped

0

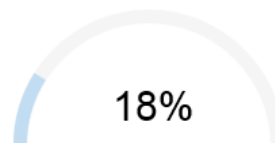
Cluster Resources

CPU



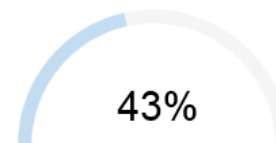
of 192 CPU(s)

Memory



90.12 GiB of 503.22 GiB

Storage



5.46 TiB of 12.69 TiB

Nodes



Name	ID	Online	Support	Server Address	CPU Usage	Memory Usage	Uptime
virtsysa	1	✓	Basic	10.16.4.1	1%	28%	12 days 14...
virtsysb	2	✓	Basic	10.16.4.2	3%	23%	00:02:54
virtsysc	3	✓	Basic	10.16.4.3	0%	10%	12 days 11:...
virtsysd	4	✓	Basic	10.16.4.4	0%	10%	12 days 11:...

C - L'infra « système »

PROXMOX Virtual Environment 4.4-1/eb2d6f1e Search You are logged in as 'root@pam' Help Create VM Create CT Logout

Server View Virtual Machine 906 () on node 'virtsysa'

Start Shutdown Reset Remove Migrate Console Help

Connected (encrypted) to: QEMU ()

Summary Console Hardware Options Task History Monitor Backup Snapshots Firewall Permissions

Appuyez sur CTRL+ALT+SUPPR pour ouvrir une session

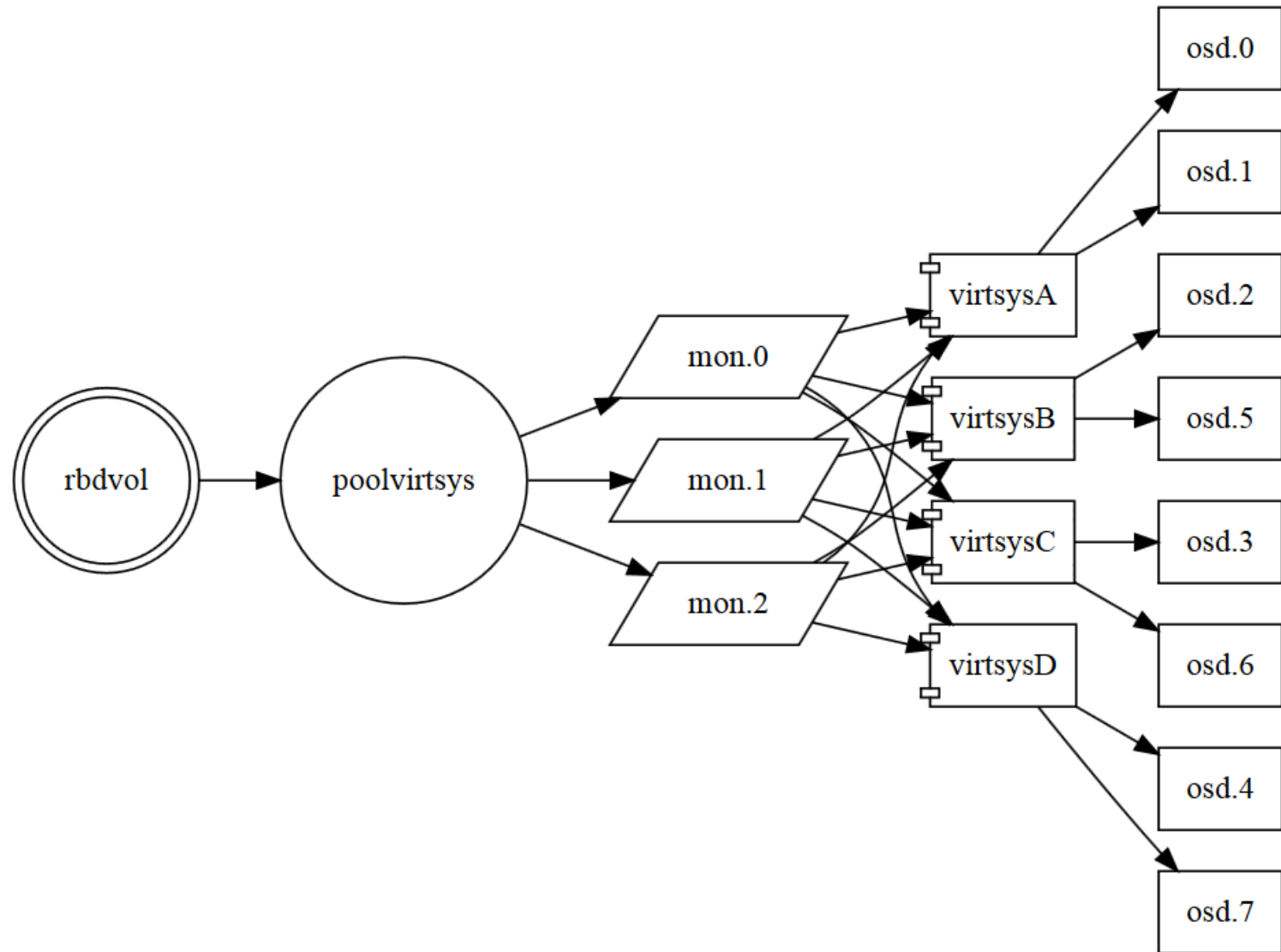
Windows Server 2008 R2 Standard

Tasks Cluster log

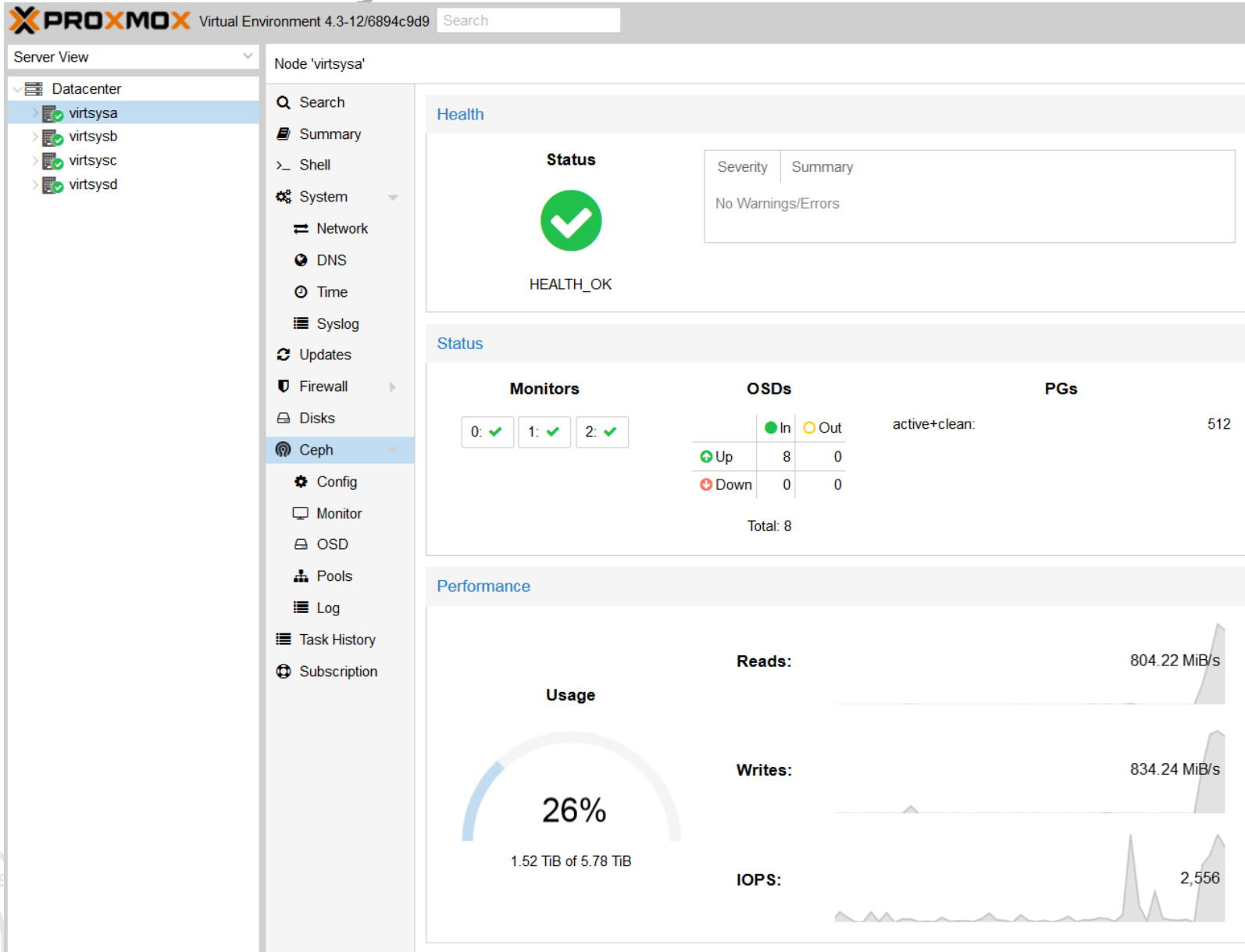
Start Time ↓	End Time	Node	User name	Description	Status
Dec 15 05:03:00		virtsysb	root@pam	VM/CT 906 - Console	

16/12/2016

C - L'infra « système » - CEPH



C - L'infra « système » - tableau de bord CEPH



C - L'infra « système » - tableau de bord CEPH

Node 'virtsysa'

Name ↑	Type	Status	weight	reweight	Used		Latency (ms)	
					%	Total	Apply	Commit
default	root							
virtsysa	host							
osd.0	osd	up / in	0.719986	1	26.04	739.85 GiB	4	3
osd.1	osd	up / in	0.719986	1	27.16	739.85 GiB	10	8
virtsysb	host							
osd.2	osd	up / in	0.719986	1	26.40	739.85 GiB	6	4
osd.5	osd	up / in	0.719986	1	25.46	739.85 GiB	1	0
virtsysc	host							
osd.3	osd	up / in	0.719986	1	25.39	739.85 GiB	2	1
osd.6	osd	up / in	0.719986	1	24.83	739.85 GiB	7	5
virtsysd	host							
osd.4	osd	up / in	0.719986	1	27.01	739.85 GiB	10	9
osd.7	osd	up / in	0.719986	1	24.97	739.85 GiB	2	1

C - L'infra « système » - tableau de bord CEPH

Node 'virtsysa'

Restar

Search	2016-12-15 04:52:06.892628 mon.0 10.97.1.1:6789/0 11479 : cluster [INF] pgmap v3427591: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 28173 B/s wr, 6 op/s
Summary	2016-12-15 04:52:07.895199 mon.0 10.97.1.1:6789/0 11480 : cluster [INF] pgmap v3427592: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 217 kB/s wr, 37 op/s
Shell	2016-12-15 04:52:10.889578 mon.0 10.97.1.1:6789/0 11481 : cluster [INF] pgmap v3427593: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 120 kB/s wr, 21 op/s
System	2016-12-15 04:52:11.893132 mon.0 10.97.1.1:6789/0 11482 : cluster [INF] pgmap v3427594: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 30739 B/s wr, 8 op/s
Network	2016-12-15 04:52:12.895050 mon.0 10.97.1.1:6789/0 11483 : cluster [INF] pgmap v3427595: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 159 kB/s wr, 17 op/s
DNS	2016-12-15 04:52:15.889477 mon.0 10.97.1.1:6789/0 11484 : cluster [INF] pgmap v3427596: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 80935 B/s wr, 6 op/s
Time	2016-12-15 04:52:16.893442 mon.0 10.97.1.1:6789/0 11485 : cluster [INF] pgmap v3427597: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 100175 B/s wr, 16 op/s
Syslog	2016-12-15 04:52:17.897013 mon.0 10.97.1.1:6789/0 11486 : cluster [INF] pgmap v3427598: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 644 kB/s wr, 121 op/s
Updates	2016-12-15 04:52:20.891981 mon.0 10.97.1.1:6789/0 11487 : cluster [INF] pgmap v3427599: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 416 kB/s wr, 76 op/s
Firewall	2016-12-15 04:52:21.895300 mon.0 10.97.1.1:6789/0 11488 : cluster [INF] pgmap v3427600: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 194 kB/s wr, 38 op/s
Disks	2016-12-15 04:52:22.897241 mon.0 10.97.1.1:6789/0 11489 : cluster [INF] pgmap v3427601: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 239 kB/s wr, 29 op/s
Ceph	2016-12-15 04:52:25.892052 mon.0 10.97.1.1:6789/0 11490 : cluster [INF] pgmap v3427602: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 212 kB/s wr, 22 op/s
Config	2016-12-15 04:52:26.895418 mon.0 10.97.1.1:6789/0 11491 : cluster [INF] pgmap v3427603: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 137 kB/s wr, 28 op/s
Monitor	2016-12-15 04:52:27.899678 mon.0 10.97.1.1:6789/0 11492 : cluster [INF] pgmap v3427604: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 524 kB/s wr, 90 op/s
OSD	2016-12-15 04:52:30.891280 mon.0 10.97.1.1:6789/0 11493 : cluster [INF] pgmap v3427605: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 243 kB/s wr, 33 op/s
Pools	2016-12-15 04:52:31.894484 mon.0 10.97.1.1:6789/0 11494 : cluster [INF] pgmap v3427606: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 105 kB/s wr, 3 op/s
Log	2016-12-15 04:52:32.896874 mon.0 10.97.1.1:6789/0 11495 : cluster [INF] pgmap v3427607: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 538 kB/s wr, 27 op/s
Task History	2016-12-15 04:52:35.892386 mon.0 10.97.1.1:6789/0 11496 : cluster [INF] pgmap v3427608: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 212 kB/s wr, 16 op/s
Subscription	2016-12-15 04:52:36.895730 mon.0 10.97.1.1:6789/0 11497 : cluster [INF] pgmap v3427609: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 2133 kB/s wr, 8 op/s
	2016-12-15 04:52:37.898051 mon.0 10.97.1.1:6789/0 11498 : cluster [INF] pgmap v3427610: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 4991 kB/s wr, 27 op/s
	2016-12-15 04:52:40.892619 mon.0 10.97.1.1:6789/0 11499 : cluster [INF] pgmap v3427611: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 420 kB/s wr, 11 op/s
	2016-12-15 04:52:41.896439 mon.0 10.97.1.1:6789/0 11500 : cluster [INF] pgmap v3427612: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 11283 kB/s wr, 12 op/s
	2016-12-15 04:52:42.900848 mon.0 10.97.1.1:6789/0 11501 : cluster [INF] pgmap v3427613: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 74713 kB/s wr, 82 op/s
	2016-12-15 04:52:45.894219 mon.0 10.97.1.1:6789/0 11502 : cluster [INF] pgmap v3427614: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 41779 kB/s wr, 47 op/s
	2016-12-15 04:52:46.895627 mon.0 10.97.1.1:6789/0 11503 : cluster [INF] pgmap v3427615: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 19686 kB/s wr, 22 op/s
	2016-12-15 04:52:47.899905 mon.0 10.97.1.1:6789/0 11504 : cluster [INF] pgmap v3427616: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 8327 kB/s wr, 29 op/s
	2016-12-15 04:52:50.894236 mon.0 10.97.1.1:6789/0 11505 : cluster [INF] pgmap v3427617: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 106 kB/s wr, 17 op/s
	2016-12-15 04:52:51.897529 mon.0 10.97.1.1:6789/0 11506 : cluster [INF] pgmap v3427618: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 783 kB/s wr, 12 op/s
	2016-12-15 04:52:52.901330 mon.0 10.97.1.1:6789/0 11507 : cluster [INF] pgmap v3427619: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 1977 kB/s wr, 32 op/s
	2016-12-15 04:52:55.895625 mon.0 10.97.1.1:6789/0 11508 : cluster [INF] pgmap v3427620: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 12834 kB/s wr, 35 op/s
	2016-12-15 04:52:56.898992 mon.0 10.97.1.1:6789/0 11509 : cluster [INF] pgmap v3427621: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 38935 B/s rd, 22347 kB/s wr, 43 op/s
	2016-12-15 04:52:57.902636 mon.0 10.97.1.1:6789/0 11510 : cluster [INF] pgmap v3427622: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 85772 B/s rd, 89266 kB/s wr, 138 op/s
	2016-12-15 04:53:00.895819 mon.0 10.97.1.1:6789/0 11511 : cluster [INF] pgmap v3427623: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 4097 B/s rd, 35059 kB/s wr, 53 op/s
	2016-12-15 04:53:01.899666 mon.0 10.97.1.1:6789/0 11512 : cluster [INF] pgmap v3427624: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 181 kB/s wr, 7 op/s
	2016-12-15 04:53:02.901780 mon.0 10.97.1.1:6789/0 11513 : cluster [INF] pgmap v3427625: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 914 kB/s wr, 28 op/s
	2016-12-15 04:53:05.895747 mon.0 10.97.1.1:6789/0 11514 : cluster [INF] pgmap v3427626: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 319 kB/s wr, 12 op/s
	2016-12-15 04:53:06.899377 mon.0 10.97.1.1:6789/0 11515 : cluster [INF] pgmap v3427627: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 651 kB/s wr, 3 op/s
	2016-12-15 04:53:07.901567 mon.0 10.97.1.1:6789/0 11516 : cluster [INF] pgmap v3427628: 512 pgs: 512 active+clean; 512 GB data, 1533 GB used, 4385 GB / 5918 GB avail; 1461 kB/s wr, 21 op/s



C - L'infra « système » - test Perf CEPH

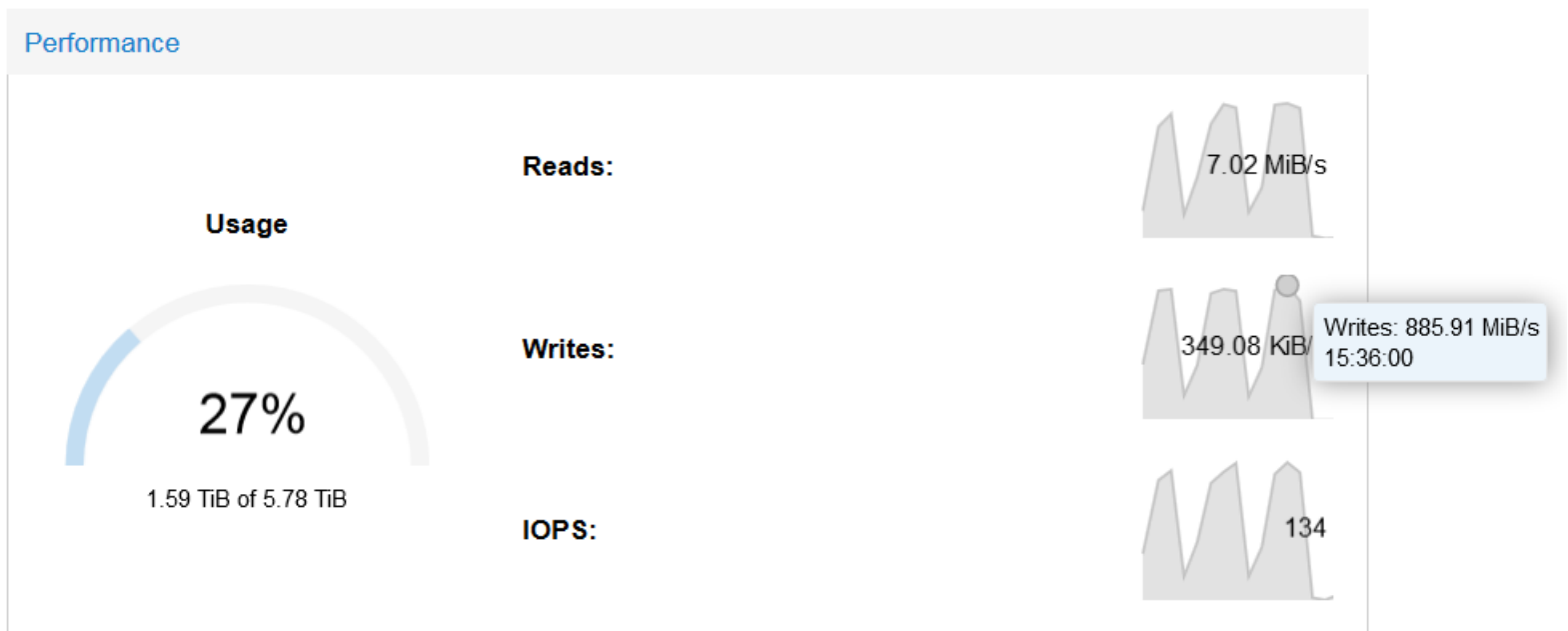
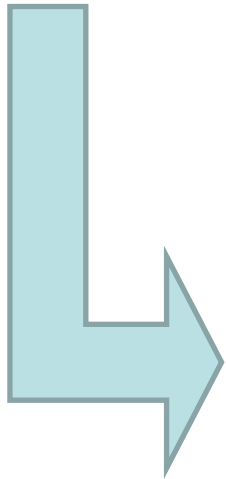
Clone VM 901

Target node:	virtsysb	Mode:	Full Clone
VM ID:	998	Snapshot:	current
Name:	test-perf-ceph	Resource Pool:	
		Target Storage:	rbdvoll
		Format:	Raw disk image (raw)

Clone

Test de clone d'une VM de 30Go

Durée : 1mn 10s (430Mo/s moyenne)



C – L'Infra « système » - CEPH – reboot

Server View

Datacenter

virtsysa

backup (virtsysa)

local (virtsysa)

local-lvm (virtsysa)

rbdvol (virtsysa)

virtsysb

901

902

903

904

905

906

907

909

910

backup (virtsysb)

local (virtsysb)

local-lvm (virtsysb)

rbdvol (virtsysb)

virtsysc

912

backup (virtsysc)

local (virtsysc)

local-lvm (virtsysc)

rbdvol (virtsysc)

virtsysd

Node 'virtsysb'

Search

Summary

Shell

System

Network

DNS

Time

Syslog

Updates

Firewall

Disks

Ceph

Config

Monitor

OSD

Pools

Log

Task History

Subscription

Health

Status

!

HEALTH_WARN

Severity	Summary
!	394 pgs degraded
!	58 pgs stuck unclean
!	394 pgs undersized
!	recovery 101540/395490 objects degraded (25.674%)
!	2/8 in osds are down
!	1 mons down, quorum 1,2 1,2

Status

Monitors

0: ✗ 1: ✓ 2: ✓

OSDs

	In	Out
Up	6	0
Down	2	0
Total:	8	

PGs

active+clean:	118
active+undersized+degraded:	394

Performance

Usage

26%

1.50 TiB of 5.78 TiB

Reads:

0 B/s

Writes:

152.37 KiB/s

IOPS:

25

Et ensuite ?

■ Aujourd'hui :

- ❑ VmWare toujours là (3 nœuds) plus d'évolution
 - ❑ Migration des VMs vers Proxmox et/ou Cloud UFTMIP
- ❑ Proxmox
 - ❑ plus de 160 VM sur 2 infras

■ Demain : quelques projets dans les cartons...

- ❑ Infra de virtualisation dédiés aux enseignants (UFR info)
- ❑ Mise à dispo de VM pour les étudiants
- ❑ Architecture OPEN STACK locale et/ou répartie

