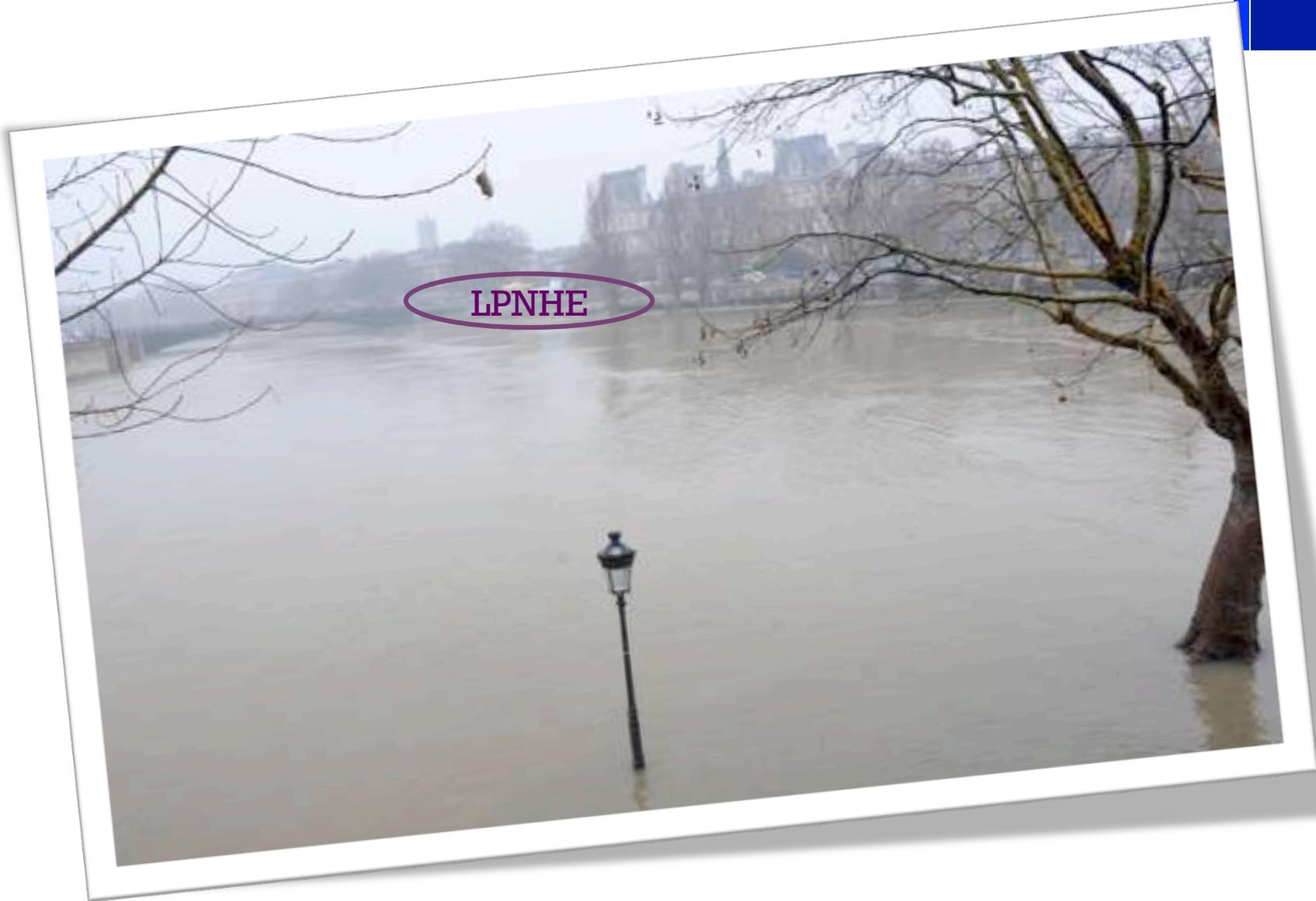


La problématique de stockage au LPNHE 3 Octobre 2017

Plan

- Présent
- Perspective
- Ceph 1^{er} pas

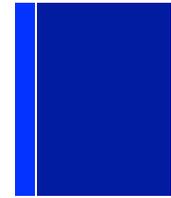
Olivier Dadoun
dadoun@in2p3.fr



LPNHE

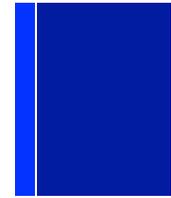


Stockage @ LPNHE – Présent -



- ◆ But pouvoir fonctionner en mode dégradé si la salle info. du RC est down: assurer l'accès aux volumes des VM et les données manip
- ◆ Redondance des données au 2eme étage
 - ✓ Plus petite salle mais suffisante
 - ✓ Réseau 10 Giga
- ◆ Profite de la bonté du CC qui nous offre les licences GPFS
 - ✓ 0(10 K€)
 - ✓ Typiquement 2x60 To (dont une manip' qui prend la moitié)

Stockage @ LPNHE – Présent -



Matériel plus ou moins récent

- ◆ 2x4 Serveurs (+1) SL6& SL7

- ✓ 2x2 serveurs de données
- ✓ 2x1 serveurs de meta données
- ✓ 2x1 serveurs CNFS (NFS env. dans cluster GPFS)
- ✓ 1 client pour effectuer des copies au CC via TSM
~8 To inspecté/nuit ~ 0.1% transféré/nuit ~7h

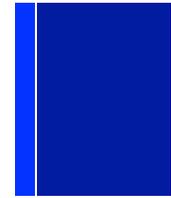


- ◆ Récupération de « baie de stockage » avec beaucoup de disques de 750 Gb en *spare* (2x45 T)

- ✓ Serveur de données slow

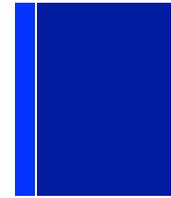


Stockage @ LPNHE – Présent -



- ◆ Volumétrie 2x70 To (« rapide ») + 2x45 To (« lent »)
 - ✓ Actuellement utilisé à 60%
- ◆ Cron qui permet de ventiler les données en fonction de leurs utilisations
 - ✓ Du stockage rapide vers le stockage lent
 - ✓ Du stockage lent vers le rapide
- ◆ Pas de problèmes notables depuis quelques années (sauf entre l'ordinateur et le siège)

Stockage @ LPNHE – Présent -



- ◆ Mais badaboum ou patatraque (même si on le sentait venir)

Bonjour Olivier,
Effectivement, le financement du droit d'utilisation et du support logiciel de GPFS ne sera assuré par le CC-IN2P3 que pour les 12 prochains mois.

- ◆ Le CC préconise quelque chose ?

Nous n'avons pas de préconisation particulière pour le remplacement de GPFS. Les prérequis du CC-IN2P3 pour un tel système de stockage sont :

- un accès fichiers (POSIX)
- la gestion des quota
- la gestion des ACL
- évolutivité en terme de capacité et de performance

En fonction de vos besoins et des critères retenus, les solutions opensource ou commerciales sont nombreuses.

Cordialement,

*Ok super Merci
à la prochaine ...*

Stockage @ LPNHE – Perspective -

◆ Super groupe de travail

✓ ~~Passer tout stockage sur EC2 d'amazon~~

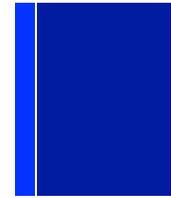
■ ~~Utiliser les économies réalisées sur la PFI~~

DFS (Open Source)	Lustre	BeeGFS	GlusterFS	MooseFS	XtreemFS	Ceph	RozoFS	LizardFS*	HDFS	EOS (CERN)	DPM	GPFS
Google rank**: \$i+filesystem (K)	163	9	108	36	1.5	266	1.2	69	477	120	115	142
Google rank*: \$i+filesystem + "data recovery" (K)	27	0.4	31	1.1	0.5	50	0.1	0.4	227	114	39	13
L2/L1 (%)	17	4	28	3	33	18	3	8	48	95	34	10
Strictly Posix	1	1	1	1	1	1	1	1	0	1	0	1
Quota	1	1	1	1	0	1	1	1	1	1	1	1
Snapshots	0	0	1	1	1	1	0	1	1	1	1	1
Data replication	0	1	1	1	1	1	1	1	1	1	1	1
OpenStack	1		1			1			1	0		
Tot	3	3	5	4	3	5	2	4	4	4	3	4
	NON	OK	OK	OK	NON	OK	NON	OK	NON	OK	NON	
				NON				NON	OK			

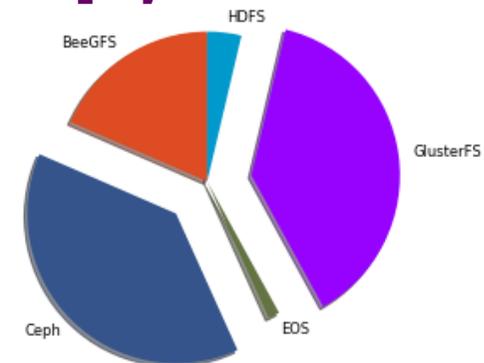
*Fork de MooseFS

**Google query with daterange:2016-09-21..2017-09-21

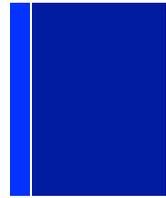
Stockage @ LPNHE – Perspective -



- ◆ 6 DFS : toujours trop (temps et man power)
- ◆ En parallèle d'études on a demandé à la liste ASR + liste stockage
Pour les différents DFS soit une croix:
 - verte, j'aime
 - jaune, aucun avis
 - rouge, beurk pas bien ou « je le sens pas celui-là »
- ◆ 2 DFS : Ceph, Gluster
 - ✓ BeeGFS facteur limitant pas de snapshot et quota payant
- ◆ A titre perso j'aimerais bien sauvé EOS ...
(actions)



Stockage @ LPNHE – Perspective -



- ◆ 2 ou 3 DFS à tester
 - ✓ Ceph, en cours (projet Cloud du labo)
 - ✓ ClusterFS
 - ✓ Éventuellement EOS (ou un autre)

- ◆ Travail qui pourrait avoir éventuellement une résonance dans le projet ComputeOps ?...

Ceph @ LPNHE

◆ Maquette :

- ✓ 3x Mon : Dell PE1950
- ✓ 6x OSD : SunFire X4540 (48x 1TB)
- ✓ Configuration : ceph-ansible

◆ Tests en cours :

- ✓ Luminous + Bluestore
- ✓ SSD : couche de stockage rapide
- ✓ SSD : journaux

cephmetrics



Ceph @ LPNHE

◆ Problèmes rencontrés :

✓ les besoins

- beaucoup de machines
- beaucoup de mémoire
- 2x réseau 10 Gbps