



Data-NG
(Ceph@P210)



Présentation

LabEx P2IO regroupe 8 laboratoires de physique répartis entre le plateau (CEA/IRFU, Polytechnique/LLR) et la vallée (CSNSM, IAS, IMNC, IPNO, LAL, LPT).

Avec des collaborations de longues dates (GRIF [IRFU, LLR, IPNO, LAL], Agata - [CSNSM, IPNO], ...)

En 2013, une rénovation des salles machines de P2IO (Polytechnique, Salle Virtualdata vallée). L'exploitation de la salle vallée est un travail coordonné des 6 laboratoires P2IO de la vallée.

En 2016, investissement dans des coeurs de réseau 100 Gb/s pour anticiper les besoins à venir.

Genèse du projet

- Laboratoires P2IO ayant l'habitude de gérer de grosses volumétries
 - 7 Pos de stockage répartis sur les différents sites
 - De plus en plus de mal à répondre aux exigences des expériences
 - coût des reconstructions RAID
 - performance d'accès aux données (hot file, ...)
- Laboratoire ayant une forte expérience d'une exploitation commune
 - Salle Vallee gérée par 6 laboratoires
 - GRIF, 10 ans d'une expérience commune entre 4 des 8 sites P2IO
 - Agata pour l'exploitation d'un Ceph pour l'acquisition de donnée

Des systèmes à bout de souffle

- Système RAID peu performant et critique lors de reconstruction
 - Coût au niveau des performances en utilisation nominal
 - Coût des reconstructions et vieillissement homogène des disques (risque de perte de disque lors des reconstructions)
 - Aucune résilience ou à coût exorbitant
 - Explosion de la bande passante/To
- Besoin croissant des expériences
 - En volumétrie
 - En performances (analyse de données)
 - En stabilité (résilience)

Des expériences éparses

Plusieurs sites ont déjà des infrastructures Ceph en production ou en test

- Cloud@VD: ~ 1Po pour l'infrastructure OpenStack
- Agata: ~ 300To de stockage pour l'acquisition d'expérience
- Cluster Promox pour la résilience des services

Bilan des expériences

Les infrastructures Ceph profitent de la multiplicité

- Mutualiser les infrastructures de stockage entre expérience pour profiter du nombre
- Une expérience, un laboratoire ne peut pas, seul, mettre en place une infrastructure de stockage de ce type
- Ne pas multiplier les infrastructures
 - Coût humain
 - Atteindre la masse critique pour l'infrastructure

Objectif du projet

- Proposer une infrastructure de référence pour les besoins en stockage des expériences P2IO
 - Basé sur un noyau dur de compétence humaine réparti dans les différents laboratoires
 - Basé sur une infrastructure de taille représentative (1Po) mais extensible sans impact sur les performances
- Infrastructure de stockage performante et résiliente.
 - Performante : liaison 100G entre les sites
 - Résiliente : inductions électriques différentes, réplication sur différents sites, répartition géographique.
- Répondre aux besoins futurs des expériences.
 - Limiter l'impact de la perte d'un disque (Reconstruction RAID)
 - Coût maîtrisé

Nos engagements pour les expériences

Une infrastructure de stockage qui ne gère pas les données et services

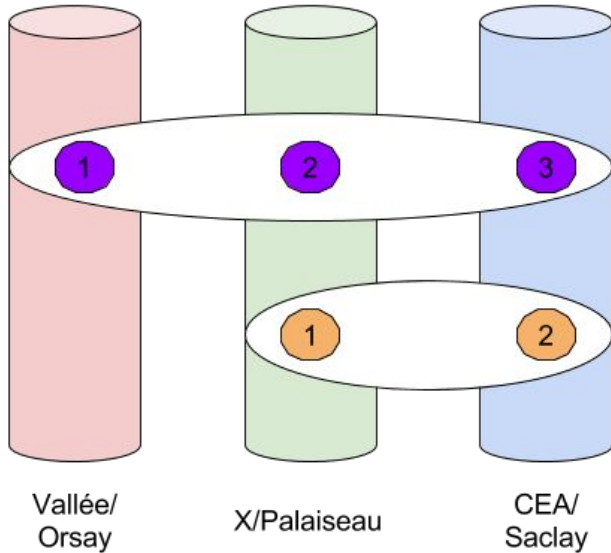
- Nous gérons en commun l'infrastructure de stockage (OSD monitor) mais pas les services associés (RADOS GW, montage RDB)
- Nous ne sommes pas responsables des données mais de l'infrastructure de stockage
 - Pas de sauvegarde des données

Cas d'utilisation

- Système de stockage pour le Cloud@VD
- Stockage de données distribuées pour l'analyse des données d'expériences
- Acquisition de données pour les plateformes P2IO locales
- Infrastructure de stockage grille pour les expériences du LHC

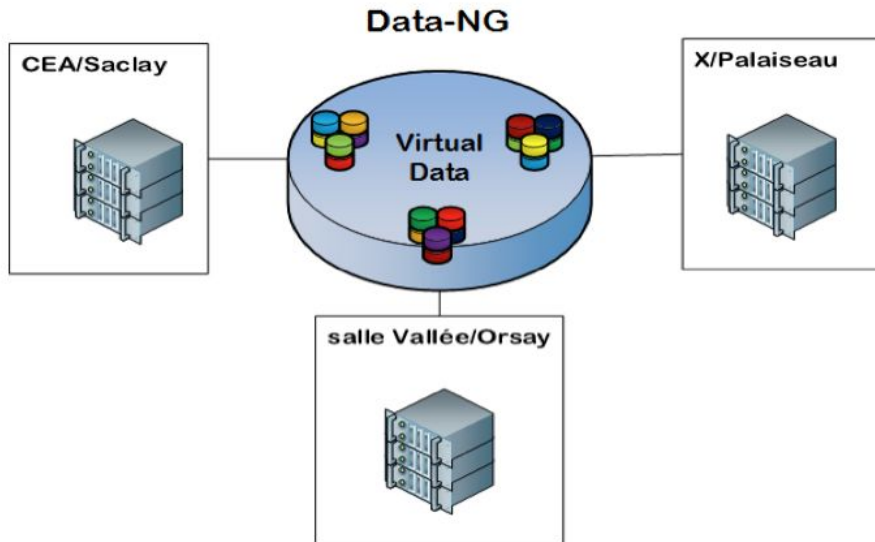
Un besoin en phase avec les expériences

- Une politique de réplication adaptée aux besoins des expériences
 - Un financement adapté aux besoins



Une expérience peut décider de mettre en place une politique de 3 réplicas alors qu'une autre peut décider de n'avoir que 2 réplicas

Infrastructure matérielle



Une infrastructure matérielle répartie de façon homogène sur l'ensemble des 3 sites.

- Éviter une dissymétrie structurelle
- Profiter de la répartition géographique pour augmenter la résilience

Une expérience humaine

Cluster de stockage gérée en synergie par différents laboratoires

- Etablir les procédures de mise en oeuvre et d'exploitation de l'infrastructure
- Coordination des activités de mise à jour
 - Processus de validation

Planning de mise en place

- 2018: Réseau dédié LLR - LAL - IRFU
- Fin 2018: Mise en place de l'infrastructure de test sur le réseau actuel
- Début 2019: Utilisation de l'infrastructure 100 Gb/s

Merci