

Centre de Calcul
de l'Institut National de Physique Nucléaire
et de Physique des Particules

IN2P3 Site Status

HPSS-FR 2025 – Lyon - Pierre-Emmanuel Brinette

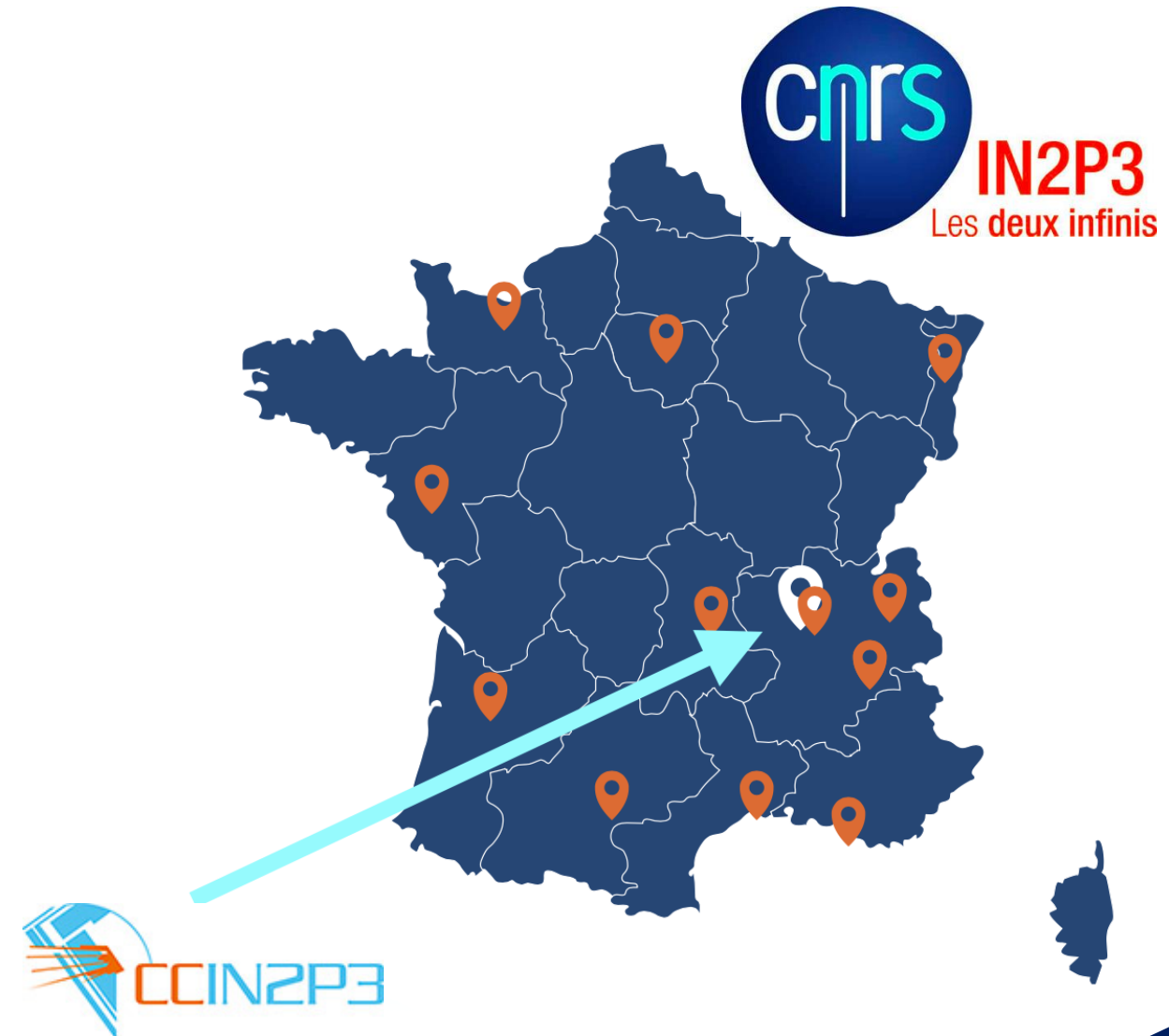
IN2P3:

French National Institute of High Energy Physics

- ~1000 researchers / 1500 support staff
- 12 laboratories over territory
- 1 Computing Center in Lyon (CC CIN2P3)

Involved in major international collaboration

- HEP (Tier 1 for LHC experiment at CERN)
- Astrophysics (Euclid, LSST Vera C. Rubin)





Resources

- 80 people (**65 IT engineers**)
- Budget : 7.3M€ (HR excluded)
 - 2.5M€ buildings running costs (incl. **1.2M€ electricity**)
 - 4M€ IT investments (incl. 2M€ for WLCG)

Facilities

- **1700 m2** over two computer rooms
- 1,4 MW total (PUE 1.4)

Computing

- ~850 servers, 55k HTC, **931 kHS23**

Storage

- Total allocated storage : **~320 PB** (62% tapes)

Networking

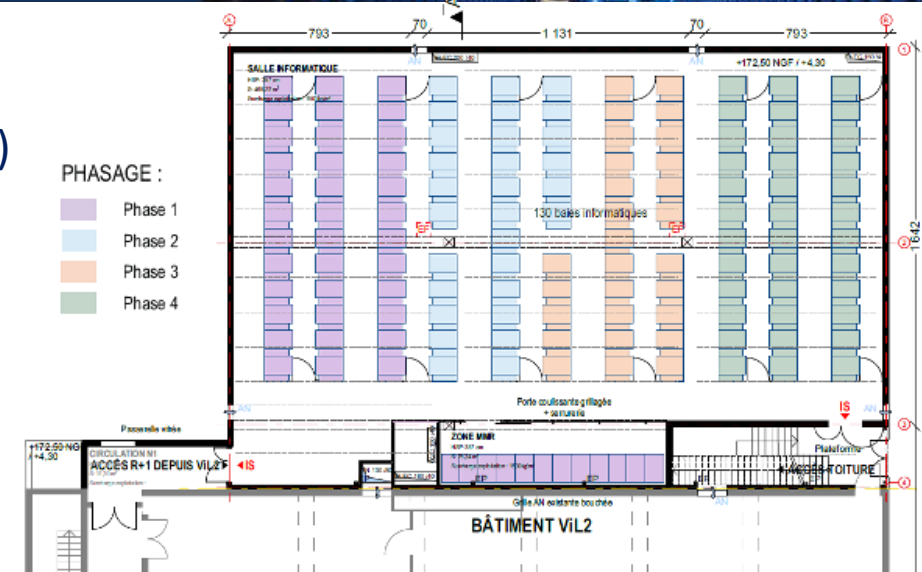
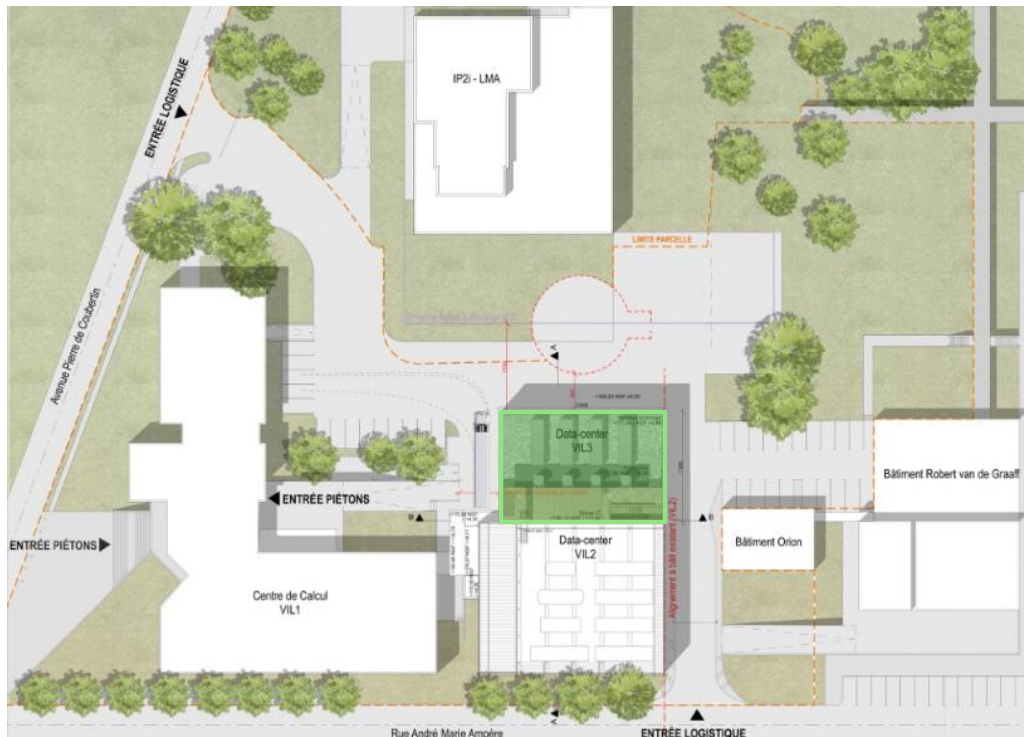
- 2x 200Gbps for WLCG (LHCOP & LHCONE)
- 1x 100 Gbps dedicated to IDRIS
- 1x 100 Gbps as backup and general purpose



100+ scientific collaborations

New computer room VIL3

- Funded through a national project with IDRIS (CNRS HPC site)
- **Target usage : research infrastructures hosting**
- **Construction from nov 24' end to 25' end**
- **2MW IT (130 racks 15kW)**
- **7,7 M€ pre-tax budget**



Euclid satellite launched on July 2023
Significant contribution to the VIS
instrument calibration

Survey started on February, 14th 2024
CC-IN2P3 is providing 25%
of data storage and processing

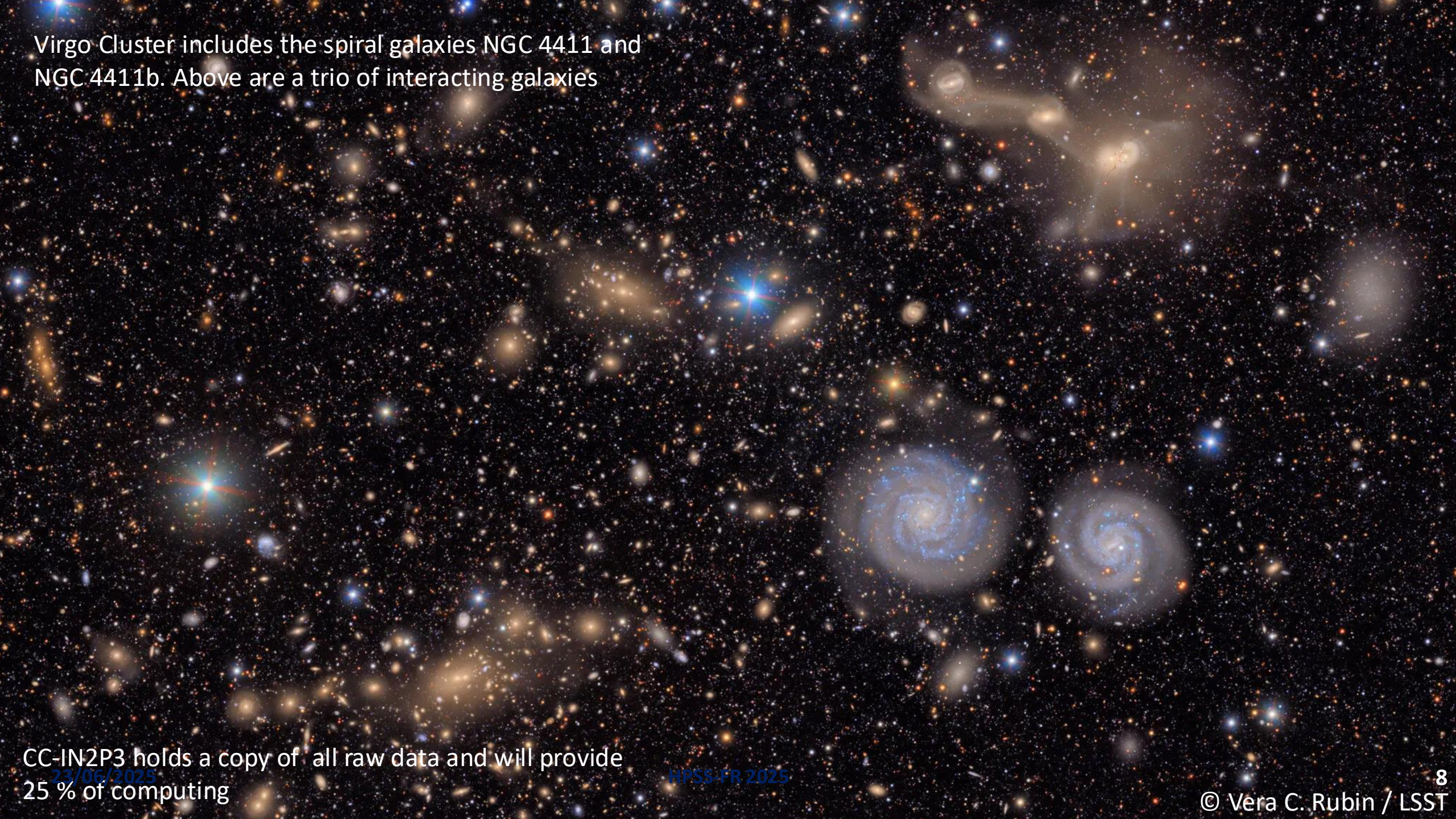
660TB in HPSS / 1.4 PB on disk
10 millions HS23 hours of CPU time used in March 2024

HPSS-FR 2025

Trifid and Lagoon nebulae, clouds of hydrogen emit a pink glow while hot stars shine blue



Virgo Cluster includes the spiral galaxies NGC 4411 and
NGC 4411b. Above are a trio of interacting galaxies



CC-IN2P3 holds a copy of all raw data and will provide
25 % of computing

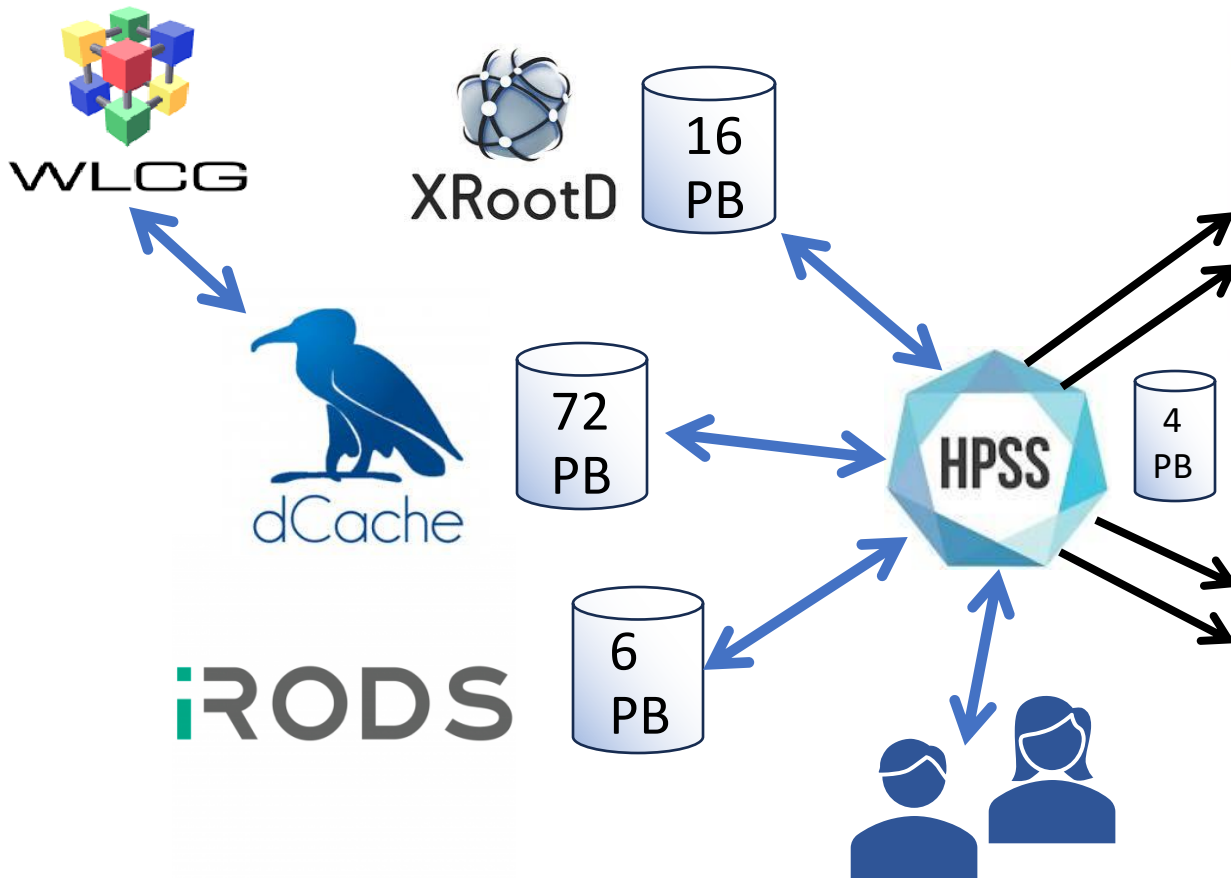
HPSS-PR 2025

© Vera C. Rubin / LSST

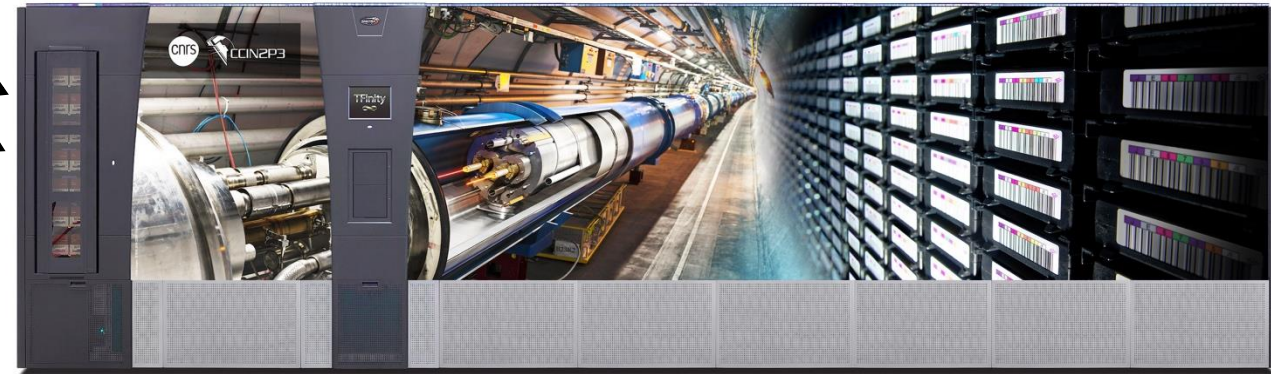


Stockage au CC-IN2P3

HPSS, monitoring an libraries



210 PB on tape (single copy)



Dedicated storage middleware ~ 100 PB disk
HPSS as tape backend

Writing : Average 150 TB/d with peak to 500 TB/d
Reading : Average 120 TB/d with peak to 500 TB/d
Archive growth : +45 PB in 2024 (+30 %/ year)



Tape Libraries

HPSS, monitoring an libraries



23/06/2025

HPSS-FR 2025

210 PB on tapes

Growth :

- +30 CAGR %

2 xTfinity

- 2x9 frames
- 2x48 Drives TS1160
- 2x6800 slots
- Fully provisioned

Might be full by end of 2025

→ Q1 2025 call for tenders for a new library

Lumos Upgrade (September 2024)

Hardware upgrade completed in one day

- RCM/LCM replaced
- Frontpanel replaced

Bluescale databases converted

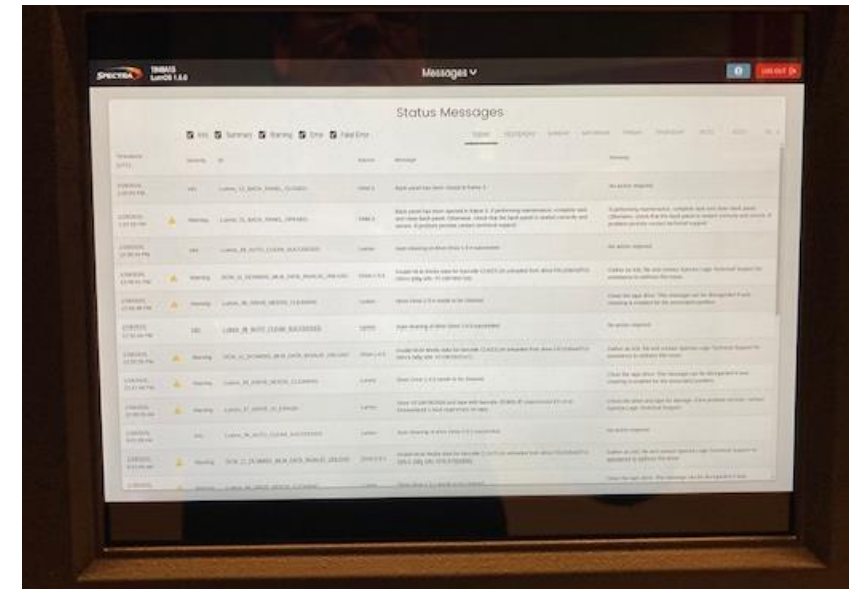
- Could take hours

Bugs in LumOS 1.4.2

Upgraded to LumOS 1.5 in November 2024

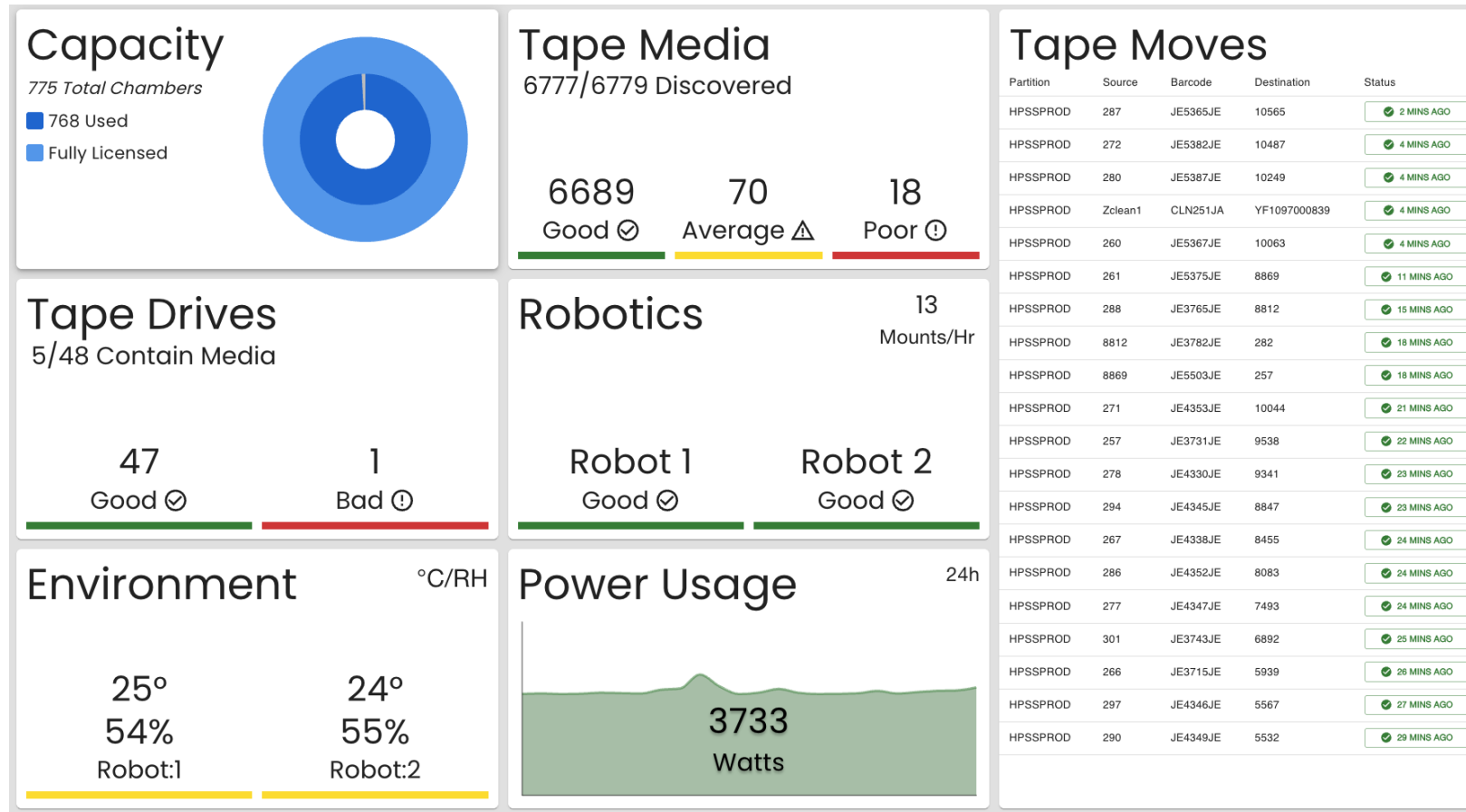
- Upgrade completed within 30 mn

Very stable since



MAJ LumOS 2.2 en juin 2025

30 mn d'indisponibilité pour la MAJ, pas de redémarrage électrique requis





HPSS Status

HPSS, monitoring an libraries

Volumétrie lors du dernier HPSS FR (Septembre 2021)

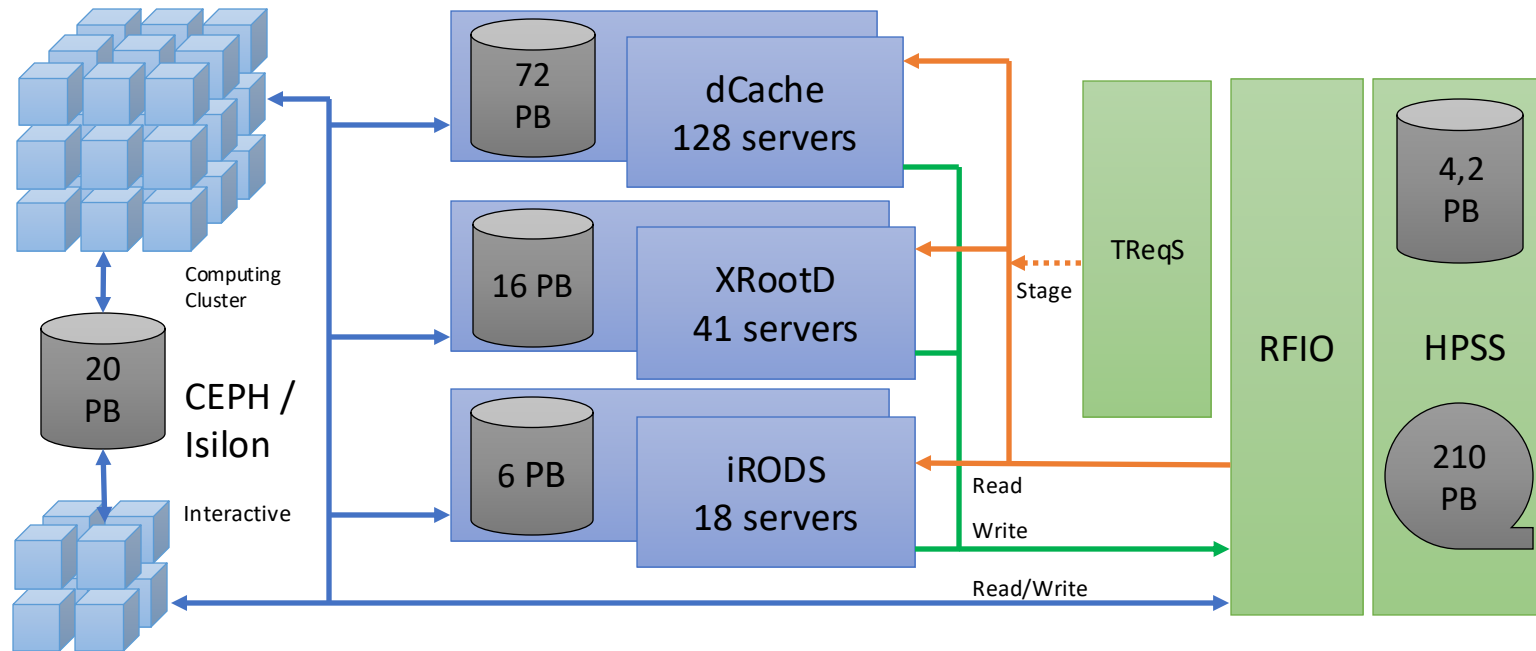
Bytes Used	100,318,879,905,094,643
------------	-------------------------

Volumétrie lors du dernier HPSS FR (juin 2025)

Bytes Used	211,878,121,236,455,656
------------	-------------------------

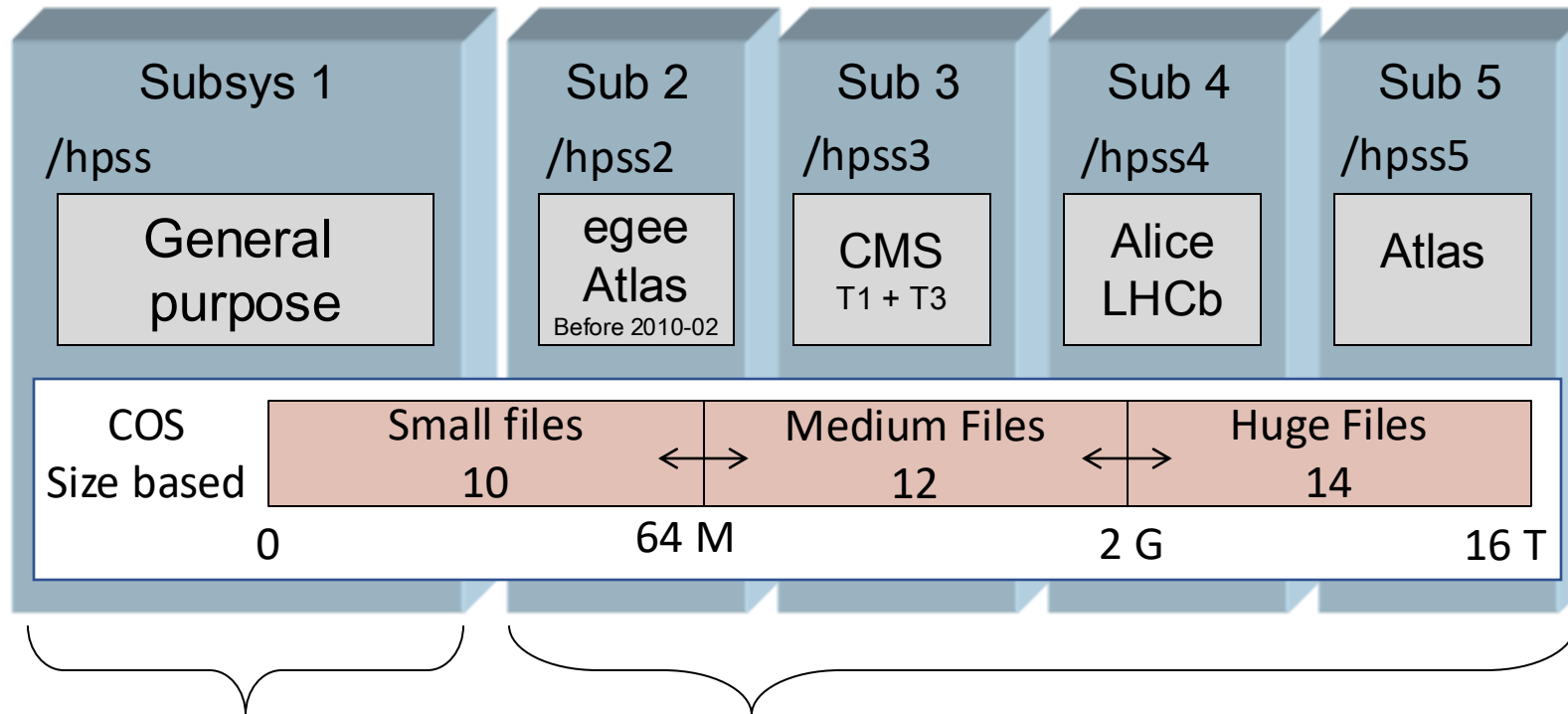
Croissance annuelle composée = 22 %

(En réalité 30% sur bande à cause de effacements
qui nécessitent des repacks)



- ▶ HPSS Interface : RFIO with HPSS extensions
- ▶ Historically, direct access from users/jobs to HPSS using RFIO
- ▶ Now, 85 % of access are performed through storage middleware
 - Reduce stress on robotic due to the large disk cache
- ▶ Read operations from storage middleware handled by TReqS
 - Limit numbers of drives used for large reading campaign
 - Optimize files recall by sorting files on the same tape to speedup read

- 5 subsystems, 3 COS Only (selected by size), 10th file families



- ▶ Small files in double copy
- ▶ + 2 archival COS
 - Double copy
 - ~ 7 PB

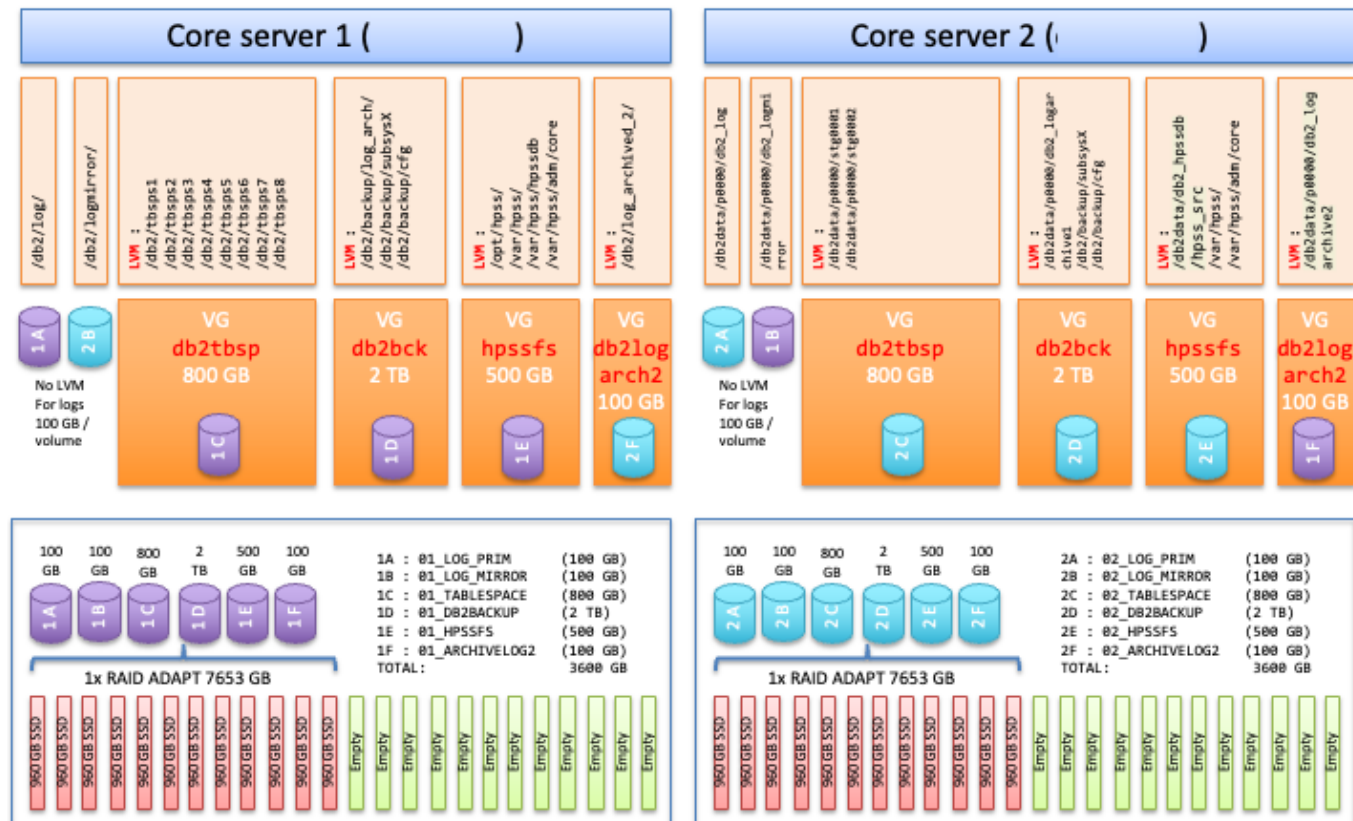
- ▶ Historical
 - 63 PB
 - ~2000 UID
 - 80 M files
- ▶ Newly created
 - 136 PB
 - 68 M files
 - for LHC Data

Core serveurs :

- Renouvelés en mars 2021
- 2x DELL R640 (1U)
 - 2 CPU Xeon Silver 4208
 - 96 Go RAM
- 2x ME4024 FC
 - 10x SSD - 960 GB
 - Adaptative RAID 7653 GB
 - Volumes "mirrors" croisés
- Fin de garantie en mars 2028

Migrations HPSS :

- Décembre 2021 : Migration HPSS 7.5.3 → 8.3 (Serveur 1 → Serveur 2)
- Septembre 2022 : Migration HPSS 8.3 → 9.3 (Serveur 2 → Serveur 1)
- Décembre 2024 : Migration HPSS 9.3 → 10.3 (Serveur 1 → Serveur 2)



- 16 disk movers (DAS) :
 - HPE ProLiant XL420 Gen10 + SAS extension
 - 20 Gbits Ethernet
 - SAS Disk in RAID
 - Total 4,2 PB
- 16 Tape movers
 - HPE ProLiant DL360 Gen10
 - 1 mover / 6 TS1160
 - 20 Gbits Ethernet
- 2 remotes PVR
 - Installed on tape movers
 - One per library
 - NO PVR on the core server
- **Fiber Channel**
 - Cout d'acquisition / maintenance élevé sur les gros SAN
 - Nouvelle stratégie:
 - 1 SAN FC 64 ports / librairie
 - Non inter connectée

Oracle T10K to TS1160 completed early 2023. Almost 3 years to complete the migration of 80 PB

Was the opportunity to reorganize the storage classes :

- Reducing the number of general purpose COS from 4 to 3
(Small < 64Mo / Medium < 2Go / Large > 2 Go).
- Remove legacy COS
- Setup double copy for small files COS

Small files migration strategy.

- Introduce new level on the existing hierarchy (not a new COS).
- Enable FAR
- For each tape volume:
 - lsvol
 - Stage the 1st file only of each aggregate (FAR would stage the others).
 - Check files in disk cache / errors / restage missing files
 - Use “Force Migrate” with BFID of each file to create the new copies on both tape SC
- 340 Volumes / 24 M files proceed in 1 month

MAJ de HPSS par puppet

- Clients, Mover, serveurs RFIO et treqs

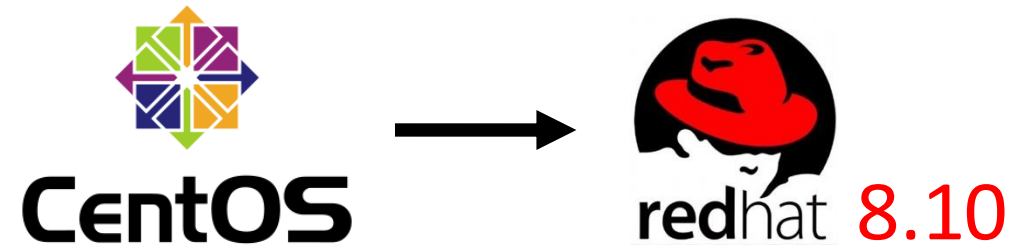
Suite à la migration, le core serveur crachait toutes les heures exactement (core dump)

- Après intervention du support (Shandrof) et analyse des logs, il s'est avéré qu'un client HPSS 9.3 continuait à se connecter
- VM des test HPSS-Fuse, oubliée dans la liste des machines
- Filesystem /hpss monté
- Echec d'exécution de puppet (dépendance RPM)
- Outils de monitoring des espace disque d'exécute 1x / heure

Problème de fuite mémoire en HPSS 10.3u8 entrainant le crash du SSM toute les 2 -3 jours ~

Migration RHEL7 / CentOS 7 devenait urgente

- Plus de MAJ de sécurité depuis Juin 2024
- Jusqu'en 2024 : Rocky Linux 8
- A partir de mi 2024 : RHEL TM
- Tentative échouée de migrer le core HPSS en RHEL 9.4
 - PB due au package KRB5 empêchant l'authentification unix sur le SSM
- Installation des cores en RHEL 8.10
 - Release en stable (plus d'évolution)
 - Fin de support 31 mai 2029



- aarch64 is an ARM 64bits RISC based architecture massively used on mobile devices.
- **Server CPU now available :**
 - Many core per socket
 - Better performance per Watt (vs x86)
 - Reduce TCO and Carbon footprint (vs x86)
- **ATLAS (Cern LHC experiment) already ported his simulation software.**
- **Some HEP Computing Center interested :**
 - Glasgow University will provide 20% of computing pledges with ARM [[1](#)]
 - BNL [[2](#)] and KIT [[3](#)] start experiments.
- **ARM might be the future of computing for HEP Experiments**

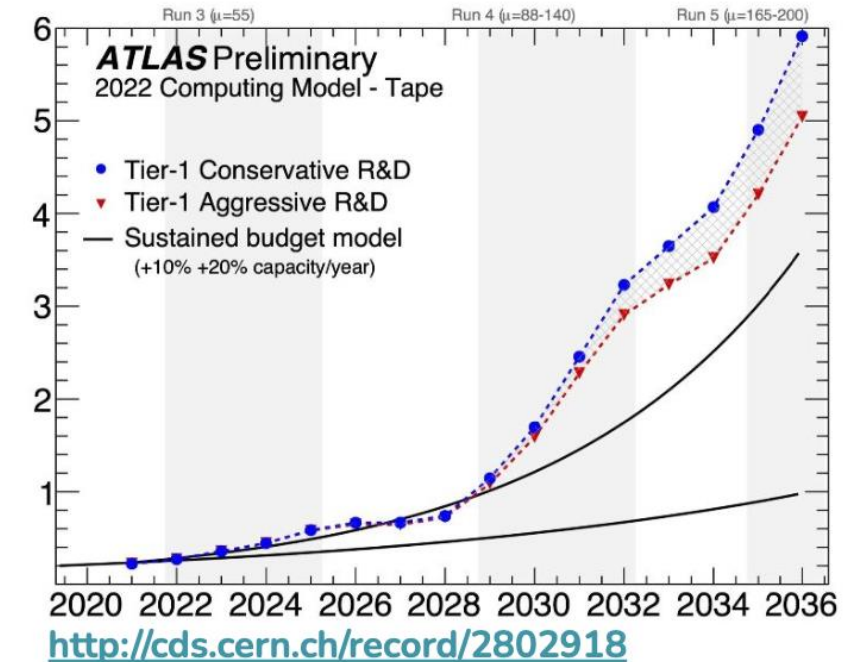


- **IN2P3 started integration**
 - Ampere Altra on HPE RL300
 - Gigabyte 2U4N-DP H262-P61 chassis (2048 cores)
 - Estimating ~20/40% gains on the TCO (procurement & energy) vs AMD Milan (7nm / 2021)
- **Software stack almost ported to RHEL9/arm**

Officially supported in HPSS 11.3 !

Futur TODO List

- **HPSS : No reason (so far) to move to another product**
 - Good knowledge of the product
 - Flat licensing
- **Main concern**
 - High Luminosity LHC -> breakdown in annual CAGR
 - Tape Carousel
 - Tapes will be used as “active” data repository for datasets
 - Optimizing read back performances for LHC-experiment
 - “Smart Writing”
 - Writing data on tape according the way users will read them back
 - Specific setup per users
 - File Family, “Per directory aggregate”
 - Full Aggregate Recall for reading



Appel d'offre publié le 5 juin 2025 pour un livraison au début d'automne

Acquisition d'une librairie de bande automatisé

- 48 lecteurs LTO10-FC
- Longueur 7,20 m maximum
- 5 ans de support

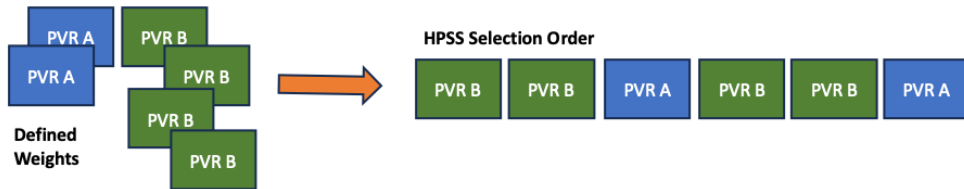
Première intégration de bandes LTO dans HPSS.

- Nouvelle tête TDS sur les LTO-10, pas d'initialisation nécessaire
- Performances similaires aux TS1160/1170 (400 Mo/s)
- Cout des médias !
- Incertitudes quand au futur de la bande Entreprise
- FC vs SAS ?
 - Bien que moins cher, le SAS plus contraignant (distance lecteurs / serveurs, Switch SAS, etc)

Nouvelle fonctionnalité HPSS 11.2 permettant d'exploiter pleinement les librairies « Balance New Tape VV Selection accross multiples PVR »

Recent Releases – HPSS 11.1 -
Balance New Tape VV Selection

- More on policy-based selection
 - Weighting scheme
 - For a given tape SC, each PVR has a weight
 - Higher weight = chosen more often
 - Lower weight = chosen less often



IBM <http://www.hpss-collaboration.org>

HPSS Futures – HPSS Collaboration Use Only

15



Premiers tests avec HPSS 11.2u1

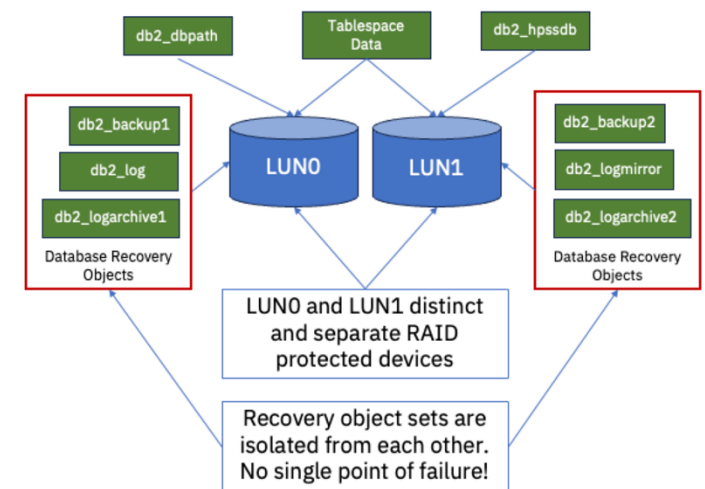
- RFIO compile et fonctionne (toujours)
- Gros changement dans la mécanique d'authentification (unix)
- Les clients s'authentifie auprès du daemon `hpss_unix_service` avant de se connecter aux COREs

Metadata Performance issue appears recently

- Some collaborations directories contains millions of file
- Issues when running hundreds of // files restore on theses directories.
 - MM0009 Now using up to **1052** database connections
 - Queries take up to 2 minutes to complete
 - Client timeout within 60s, rerun readdir immediately, triggering a new DB request
- Other use case when updating namespace :
 - Change group (chgrp -R ...) over 10 millions files / directories
 - 4 days to complete

→ Setting up DB2 partitioning might be helpful

→ Opportunity to setup the new simplified DB2 filesystem layout



Merci