

## Fédération de Stockage : retour sur 2019

C. Adam-Bourdarios, F. Chollet, M. Gougerot, S. Jézéquel,  
P. Seraphin (LAPP)

S. Crépe-Renaudin, C. Gondrand (LPSC)

J-C. Chevaleyre (LPC), E. Knoops (CPPM)

**21 Janvier 2020**

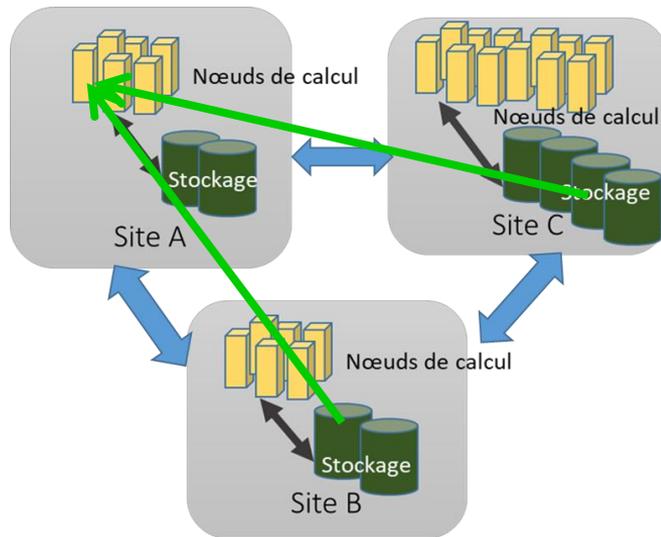
## Evolution des sites Tier 2 de WLCG (type mésocentre) dans le cadre de la R&D DOMA – HL-LHC computing

- Volonté conjointe équipes (physiciens et informaticiens) de participer à la R&D DOMA
- Objectifs liés à la **gestion des données et du stockage** :
  - Maintenir la composante Stockage et Données des Tier 2 à l'horizon HL-LHC :
    - Poursuivre la croissance du stockage à budget constant ou à coût maîtrisé
  - Développer les compétences « Gestion de données » acquises et la proximité avec l'expérience ATLAS
  - Chercher à optimiser drastiquement l'utilisation du stockage et réduire les coûts opérationnels
  - Envisager de nouvelles infrastructures de stockage : fédérée, multi-sites... Data-lake compatible
  - Tirer parti de la proximité régionale entre différents sites ATLAS : Annecy (LAPP), Grenoble (LPSC), Lyon (CC-IN2P3), LPC (Clermont), CPPM (Marseille)
  - Travailler dans une optique non exclusivement LHC qui permette à terme de répondre aux besoins d'autres projets : CTA, LSST, DUNE...

**en collaboration avec le CC-IN2P3 Tier1**

Problématiques : Gestion de données, Services de cache, Fédération des sources de données

## 1. Optimiser l'utilisation du stockage disque ; Réduire la duplication des données de l'expérience ATLAS



– En 2018 : 10 à 15 % du stockage persistant de chaque site héberge des **copies temporaires (durée de vie ~ 15 jours)**

– Réplication gérée par Rucio, l'orchestrateur de « data flows » d'ATLAS



Duplication par transfert asynchrone (FTS)



Accès local (*Analysis Download Direct IO, Production Download, Production Upload*)

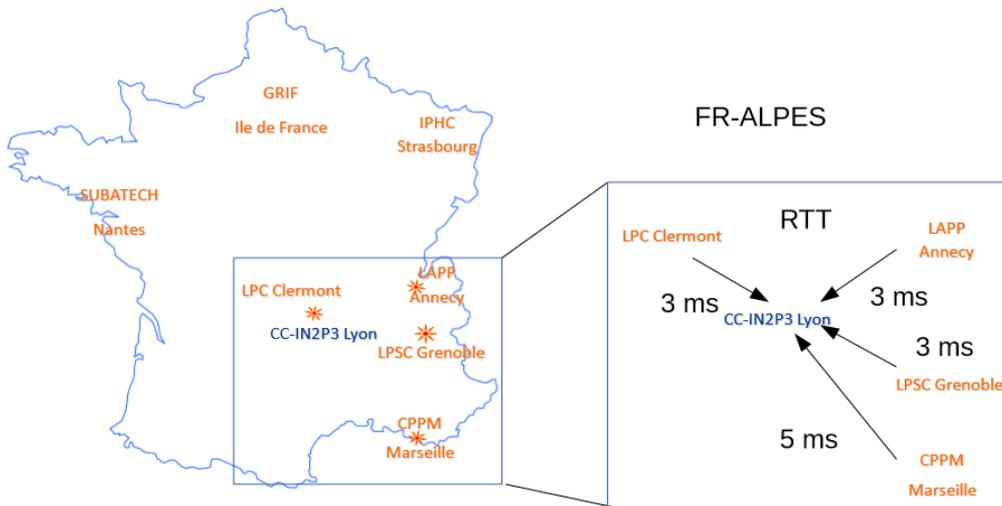
- Activation des accès distants en lecture à une échelle régionale en remplacement de transfert asynchrone entre sites
- Les nœuds de calcul de chaque site accèdent à distance aux données stockées sur disque dans les autres sites

## 2. Optimiser la mise en œuvre du stockage disque

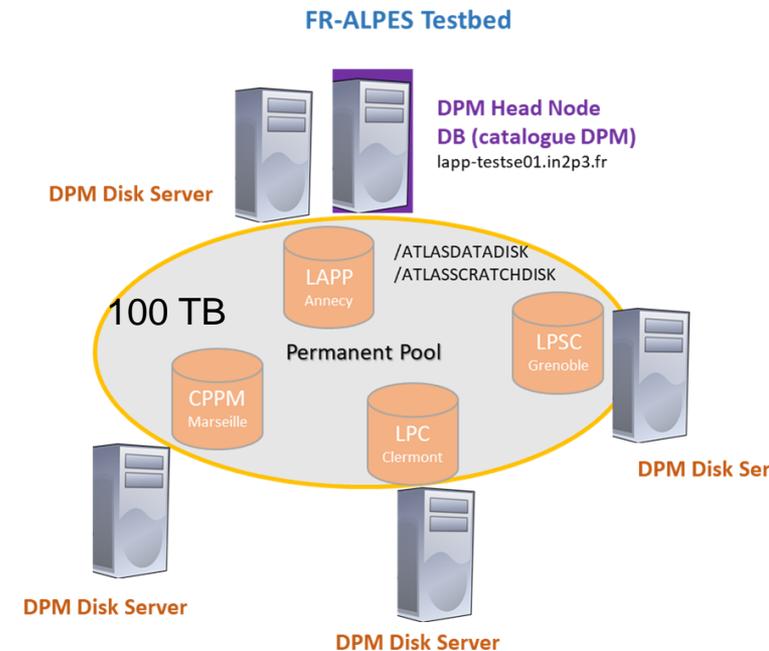
### Fédération de stockage entre sites « proches » et intégration dans ATLAS

Motivations : Gains potentiels (expertise, efficacité, visibilité...)

- Expertise DPM (techno. spécifique) : versions + récentes de la techno. DPM, anticipation de l'évolution à long terme
- Administration partagée : un seul headnode (serveurs de disques distribués)
- Un seul service et un seul point de contact pour ATLAS
- Meilleure visibilité au sein de WLCG :  
actuellement 6 PB stockage Tier 2 (Région Sud-Est)



Interconnexion LHCONe : 10 GB/s



## Deux axes de R&D au sein de DOMA-FR

1. [ATLAS Data Access](#) : Optimisation des accès aux données dans la configuration actuelle des sites
2. [DPM Volatile pools & Federated storage](#) : FR-ALPES

## ATLAS Data Access : Optimisation des accès aux données dans la configuration actuelle des sites

- Configuration des queues Panda pour autoriser des accès distants : changements dans AGIS aisés, réversibles
- Pas d'impact notable sur les jobs (notamment avec xrootd jusqu'à 20 ms de RTT)
- Impact de la saturation des interconnexions réseau à 10 Gb/s par les transferts de données asynchrones (FTS)
- Différents problèmes et bugs identifiés et corrigés dans les outils ATLAS; monitoring pas encore adapté pour des mesures précises d'efficacité CPU permettant de conclure
- Arrêt pour maintenance du stockage : amélioration de la gestion des arrêts (fail-over avec maintien du calcul actif) – tentative infructueuse en 2019
- Des pistes pour aller plus loin...

## DPM Federated storage : Banc de test FR-ALPES

- 02/2019 : LAPP (Annecy) et LPSC (Grenoble)
- 07/2019 : intégration LPC (Clermont-Ferrand) et CPPM (Marseille)
- Banc de test opérationnel (100 TB), totalement intégré au sein d'ATLAS
  - Intégration dans le système d'information ATLAS (AGIS)
  - Disponibilité (SAM/ETF), Fonctionnement/stress sur des jobs calibrés (HammerCloud)
  - Monitoring Transferts (FTS) et des jobs (Panda)
- Synergie entre sites pour l'administration de l'infrastructure
  - Utilisation du testbed pour préparer la migration à DOME des stockages DPM de production

– Premiers résultats encourageants

– Présentation à CHEP 2019 :

<https://cds.cern.ch/record/2693858>

*Implementation and performances of a DPM federated storage and integration within the ATLAS environment*

– Premier jalon vers un datalake régional

Et aussi, en collaboration avec GRIF, contribution à la rédaction du [DPM White Paper](#) *Merci à Andrea Sartirana !*

*Production HITS → AOD : 1000 events  
LAPP\_ALPES... = accès FR-ALPES  
Autres = accès SE local*



FR-ALPES : Fédération de stockage pour DOMA et ESCAPE \*

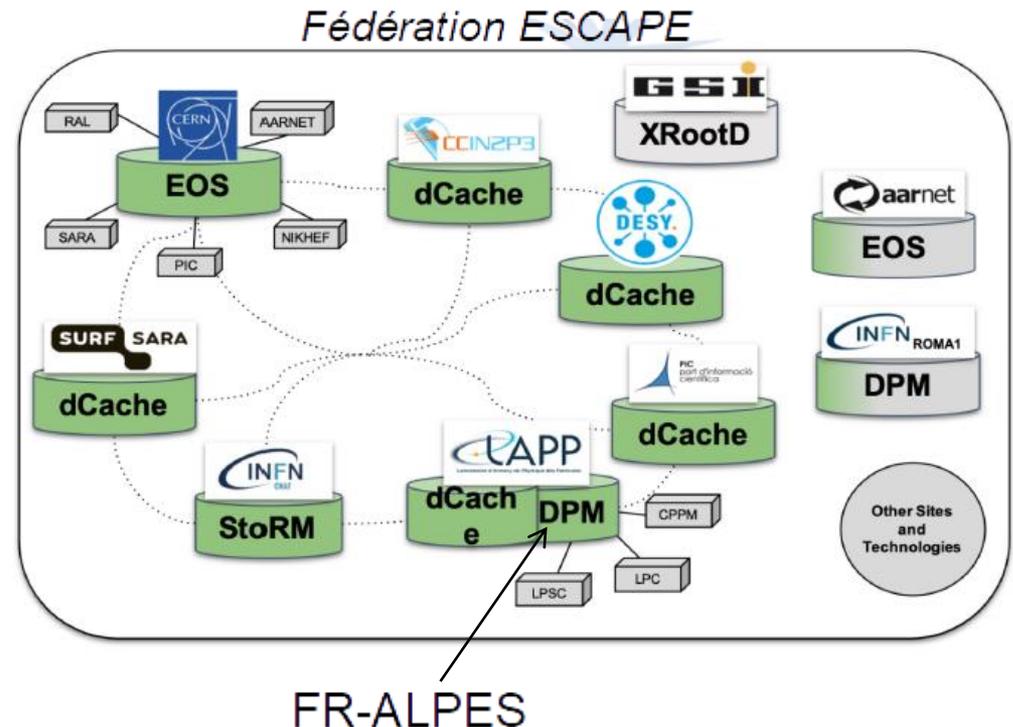
Le tout début d'une évolution à plus long terme... annonçant ce que pourrait être une architecture type Data-Lake

- Services de stockage moins nombreux mais plus importants
- Infrastructure intégrant une partie de l'intelligence et de l'orchestration
- Introduction de QOS : classes de service (latence des accès, niveau de fiabilité, de persistance, coût etc..)
- Sécurité de la donnée : réévaluation de la redondance locale et globale

Concrètement :  
Juste le support de la VO  
ESCAPE aujourd'hui

...  
À suivre en 2020 !

\* European Science Cluster of Astronomy  
& Particule physics ESFRI research  
Infrastructures



# Activités 2020

## Sujets d'intérêts

- Problèmes rencontrés / risques identifiés
  - Saturation des réseaux géographiques
    - Vision de l'activité WAN des sites FR-ALPES dans le monitoring
    - Augmentation BW : N\*10 Gb/s
    - Caching
  - Risque de limitations à long terme de la techno. DPM (suites du white paper)
    - Volatile pools : pas de use-cases ATLAS évident, limitations connues (pas de read-ahead), à revoir peut être
- FR-ALPES : Positionner le headnode DPM au CC-IN2P3 (infra fournie par le CC, installée et opérée par les admins FR-ALPES)
- Amélioration Gestion de l'arrêt du stockage dans les sites (Tier 2)
- Activités 2020 :
  - Expérimentation XCACHE
  - dans le cadre ESCAPE
    - Nouveaux modes d'authentification : token, IAA
    - Suivi QOS Stockage
    - Identification éventuelles limitations de DPM
- Encore + de collaboration
  - Avec GRIF : Etude CEPH en back-end du stockage grille (collab. GRIF)
  - Avec le CC-IN2P3 : Accès au CC dans le cadre de la modélisation Data-Lake