



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

DOMA-FR FdR

Novembre 2018



- ▶ Elle doit décrire :
 - Les objectifs que l'on se donne de tests/validations qui rentrent dans le cadre des R&D sur le stockage ?
 - Ou/et simplement les sujets d'intérêts identifiés ?
 - Les approches et technologies envisagées ?
 - Les moyens ?
- ▶ Doit on tenter d'y écrire aujourd'hui les objectifs pour les trois ans ?
 - Si non, doit on la réactualiser de façon continue ou bien à fréquence régulière (annuelle) ?
- ▶ Prise en compte des éventuels nouveaux "partenaires" et de leurs centres d'intérêts
- ▶ Aujourd'hui on a :
https://atrium.in2p3.fr/nuxeo/nxfile/default/2cc4ec8c-6db5-49c6-bc59-db5aac0e499b/blobholder:0/DOMA_FR.docx

- ▶ 4 axes de travail y sont explicités :
- ▶ Tests sur les accès aux données ATLAS
 - Essayer de s'affranchir dans la limite du possible (cad sans impact sur les performances d'accès aux données), de cette activité de pré-positionnement de données qui mobilise des ressources de stockage,
 - Permettre l'accès direct aux données pour les sites proches (faible latence réseau),
 - Estimer le seuil à partir duquel l'accès direct à distance pénalise les I/O et nécessite la mise en place de cache au niveau du stockage des sites.
 - Autres ?
 - Ouvrir cela à d'autres applicatifs ?
 - Des tests avec CMS sont mentionnés dans l'actuelle version du document.
 - En expliciter d'autres?

- ▶ Optimisation de la mise en œuvre du stockage au niveau sites :
 - DPM
 - CEPH
 - Doit on se limiter à ces solutions techniques (quid de dcache, EOS,...) ?

 - Notion de banc de test
 - De tests de nouvelles fonctionnalités
 - Doit on être plus précis ?

 - Etant donné l'importance des activités qui commencent sur le sujet, doit on expliciter plus la partie R&D sur l'usage des caches ?

- ▶ **Bon usage des solutions de stockage de masse**
 - Définir quelles seront les solutions techniques disponibles pour faire du stockage sur bande et quantifier leurs performances intrinsèques.
 - Doit on expliciter les solutions sur lesquelles on pourrait se positioner ?
 - Valider que les usages imaginés par les expériences, notamment l'approche carrousel mise en avant par l'expériences ATLAS, sont compatibles avec les solutions techniques.
 - Ouvrir à d'autres "workflows"/experiences

► Coût et performances

- Le premier axe d'investigation, qui vise à baisser les coûts de la gestion des données, est la possibilité d'utilisation plus intensive et généralisée du stockage sur bande.
- Le deuxième axe de recherche s'appuie sur le constat de ce groupe de travail que nous ne disposons actuellement pas de modèle d'utilisation du réseau suffisamment complet.

- ▶ **Les résultats attendus des différents travaux sont les suivants :**
 - Evaluation de nouveaux cas d'utilisation, montée en expertises et contribution à l'évolution de l'outil de gestion de données Rucio,
 - Evaluation coût-performances des mécanismes de mise en cache et des accès distants aux données dans la perspective de futurs services de stockage distribué,
 - Evaluation des fonctionnalités du middleware DPM permettant la consolidation du stockage disque et la fédération entre sites,
 - Etudes techniques basées sur la technologie open-source Ceph,
 - Analyse des enjeux liés à l'interopérabilité avec les standards communs des futurs services de stockage type Data-Lake,
 - Etude de solutions techniques de stockage de masse et validation de nouveaux cas d'utilisation (carroussel par ex.) nécessitant des performances d'accès supérieures à ce qui est requis pour l'archivage,
 - Etudes coût-performances, estimation du gain potentiel associé aux différentes évolutions envisagées,
 - Suivi de l'utilisation des ressources Réseau (LAN, WAN), évaluation des besoins futurs, établissement d'une feuille de route en concertation avec les experts IN2P3 et Renater

- ▶ On se donne de échéances pour le premier semestre 2019 ?
- ▶ Alors quoi ?
- ▶ Alors qui ?