

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

DOMA: Data Organization Management Access evolution

Intro



- ▶ Pourquoi DOMA ?
- ▶ Contexte et organisation
- ▶ Vue globale des activités
- ▶ Organisation de la journée

▶ Contexte :

- A l'horizon 2024-2025, de nombreuses collaborations scientifiques seront confrontées à des besoins en stockage qui seront hors de portée des ressources financières envisageables.
- Ce constat est partagé par un nombre significatif de domaines scientifiques (HEP, astroparticule, sciences du vivant, sciences de l'univers,...)
- Les sites/services qui devront fournir ce/ces service(s) de stockages aux collaborations seront en nombre restreint et contraints de rationaliser le coût de tels services.

▶ En conséquence des efforts doivent être réalisés.

- Pour faire plus avec des moyens financiers contraints, il est nécessaire de consolider nos moyens de stockage.
- Il n'existe pas de solution sur étagère qui soit satisfaisante.
- Il est nécessaire de s'intéresser à tous les aspects d'un service de stockage.
 - Techniques.
 - Opérationnels.
 - Fonctionnels.

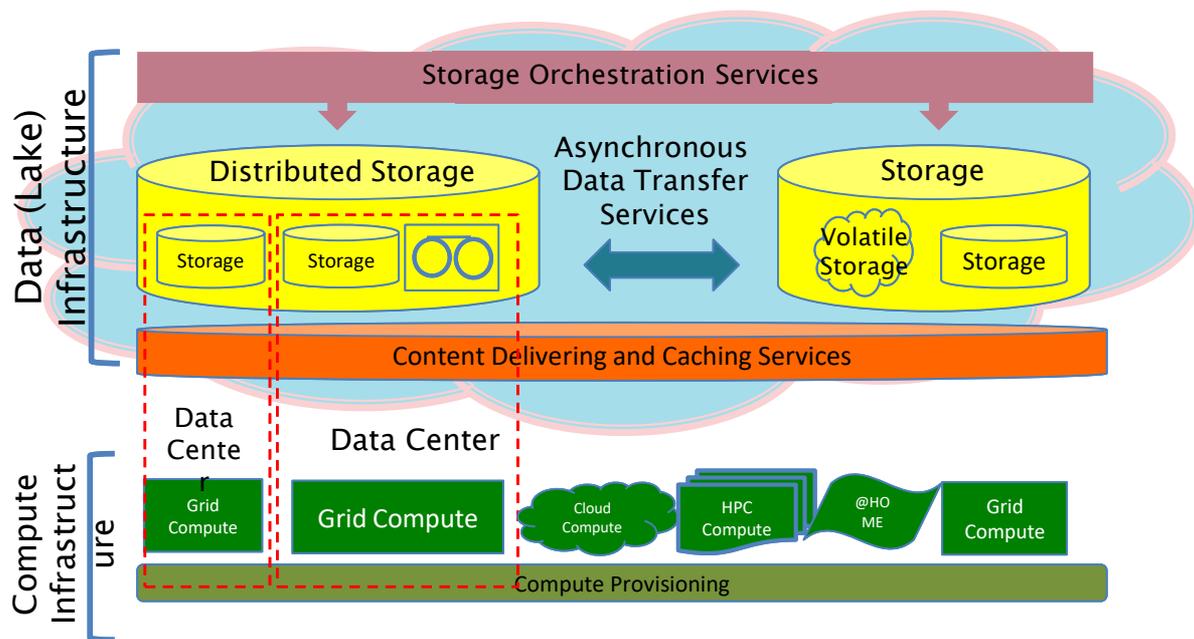
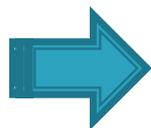
▶ Avant de définir toutes les caractéristiques de ce que doit être le service de stockage, il est indispensable de mettre en commun les connaissances, le savoir, les procédures et les expérimentations de chacun (du plus grand nombre).

- ▶ D'où la création d'une organisation « chapeau » qui regroupe toutes les initiatives qui viseront à avancer dans ce sens.
 - DOMA : Organisation internationale intégrant les différents aspects que pourrait prendre l'accès aux données dans les années à venir.
 - Les objectifs sont :
 - Suivre les avancées et développements dans les différentes activités identifiées.
 - Etre un forum de partage d'informations.
 - Veiller à l'interopérabilité des différentes solutions de stockage.
- ▶ DOMA n'est pas un projet financé, même si à différentes échelles de petits moyens sont affectés aux activités.
- ▶ Les activités de DOMA recouvrent différents aspects (purement technique, fonctionnel, financier, utilisateur,...).
- ▶ DOMA ambitionne de couvrir globalement la question du service de stockage scientifique.

▶ DOMA en un slide

- <https://twiki.cern.ch/twiki/bin/view/LCG/DomaActivities>
- “A set of R&D activities evaluating components and techniques to build a common data cloud”

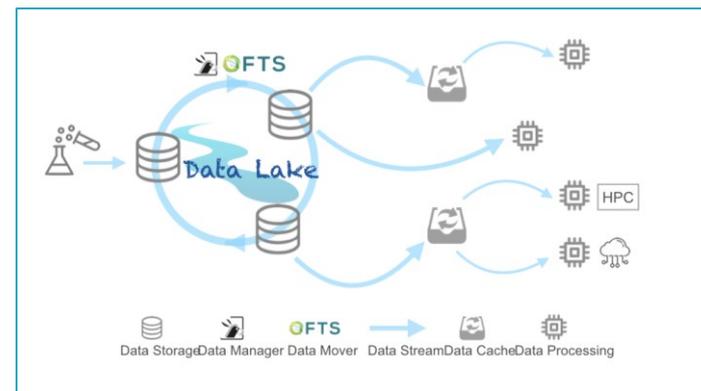
ACTIVITÉS DOMA



- ▶ C'est la composante française des activités qui s'inscrivent dans la thématique DOMA.
- ▶ Comme pour DOMA, elles sont diverses et multi approches.
- ▶ Il est souhaitable, notamment pour les fournisseurs de ressources, que le plus large champ des problématiques et approches soient prises en compte.
 - Aujourd'hui les expériences du CERN drivent fortement les activités mais.....
 - Les inputs d'autres familles d'utilisateurs, d'autres modèles de gestion des données et d'autres fournisseurs de service sont primordiaux.

▶ DATA ACCESS

- Activités autour de l'accès à la donnée.
- C'est ici que sont proposés/évalués des modèles d'accès à la donnée.
 - Fédérations de stockage
 - Fédération des File Systèmes
 - Stratégies de caches
 - Optimisation dans l'usage des ressources
 - Tape (stockage de masse)
 - Network
 -



Strawman Modèle

- Sont impliqués
 - Les fournisseurs de services et les « utilisateurs » à travers leurs modèles de data management.

▶ Fonctionnalités

- Validation et disponibilités des nouvelles fonctionnalités fournies par les piles logicielles impliquées au niveau du stockage.
- **Third Party Copy**
 - Testbeds pour valider les aspects fonctionnels et de passage à l'échelle
 - Déploiement en production à venir.
- Service d'ordonnancement des transferts.
- Service de Data Management.
-

- Sont impliqués les fournisseurs de services et les développeurs.

▶ Qualité de Service (QOS)

- Changer la façon dont est défini le besoin du stockage pour passer à une définition basée sur l'expression du besoin plutôt que sur des métriques quantitatives.
- Travail essentiellement de définition.
 - Ex : Un stream de donnée de type « COLDDATA »
 - Un niveau de disponibilité
 - Un niveau de fiabilité
 - Un coût
 - Une performance
 -

▶ Autres activités

- **Modèle de coûts**, architecture réseau, authentification,...

- ▶ Présentation de communautés hors CERN
 - INSERM.
 - Universitaire.
 - LSST (Astro)
 - CNES
- ▶ Retour sur ce qui a été réalisé en 2019
- ▶ Sessions ouvertes pour fixer des objectifs, trouver des points d'intérêts communs, de nouveaux sujets, avoir des prises de contacts,.....
- ▶ Ne pas hésiter à poser des questions, à rentrer dans des discussions techniques,.... On est là pour voir ce qu'on peut faire ensemble et en apprendre/découvrir le plus