

*DOMA et LCG-France

Le groupe de travail Data Organisation, Management and Access a été créé suite aux discussions lors du HSF/WLCG workshop de Naples. Les premières discussions sur ce sujet dans LCG-France ont démarré avant la création de ce groupe et après la sortie du draft du WLCG Strategy paper qui reprend beaucoup des idées du HSF Community Paper et des discussions du workshop: <https://indico.in2p3.fr/event/16844>. La coordination LCG-France suggérait d'aborder les R&D DOMA dans les perspectives de notre expérience avec GRIF, et des possibilités de fédération et de R&D sur le caching. Les sites étaient invités à réfléchir à leur participation possible aux R&D dans le cadre de DOMA.

La réunion "kick-off" de DOMA a eu lieu le 4 et 5 juin: <https://indico.cern.ch/event/729930/> et a été résumée lors des journées LCG-France <https://indico.in2p3.fr/event/17582/contributions/62875/attachments/48880/61892/DOMA-v2.pdf>

Le résumé a été suivi d'une discussion dans laquelle certains sites ont exprimé soit un intérêt soit une volonté de démarrer des R&D :

- GRIF a toujours eu la volonté d'aller vers un site distribué plus unifié, il y a eu des actions récentes comme l'unification du batch system entre sites. GRIF a un projet CEPH distribué (non LCG) qui est conditionné à une mise à niveau des réseaux. Une fois l'expérience acquise, le but est de construire un stockage distribué pour le site de grille sur cette base, soit en l'utilisant comme block device pour DPM (par ex), soit en l'utilisant directement comme dans RAL-ECHO (solution préférée à priori). Cela donnerait de l'expérience sur le stockage distribué à l'échelle régionale, ce qui rentre dans le cadre de R&D data lake de DOMA.
 - un des points critiques est l'impact de la latence
 - une discussion s'engage autour des caches (xrootd, DPM volatile cache): voir ce que fait RAL mais aussi CERN-Wigner
 - le LAPP (Philippe) indique que DPM propose maintenant des pools volatiles.
- Le LAPP et le LPSC ont commencé une R&D dans le cadre de ATLAS dans laquelle les spacetokens des deux sites sont associés aux CE des deux sites, forçant implicitement une partie des jobs en accès distants sur l'autre site.
- Le LPSC serait éventuellement intéressé par d'autres sujets (les caches) mais cela n'a pas été rediscuté/décidé en interne (au moins au moment de la discussion dans les journées)
- Le CC-IN2P3 pourrait participer, potentiellement dans les aspects liés aux bandes / QoS / tape carousel, mais la concertation interne n'a pas encore eu lieu
 - Catherine mentionne que dans les scénarios envisagés les gros sites de stockage comme les Tiers-1 joueront certainement un rôle central et que des pistes de collaboration CC-IN2P3/Tier-2 seraient intéressantes, comme les caches par exemple.
- Outre les propositions des sites, la discussion porte sur divers sujets:
 - Il y aurait des choses à regarder du côté INDIGO-DATA CLOUD par ex pour l'utilisation de caches (la déclinaison "stockage" est XDC, mais voir aussi EOSC, EUDAT etc) - LCG-France organisera une réunion sur XdC
 - La discussion réitère l'importance du monitoring pour comprendre ce qui se passe (niveau site, réseau, expérience)
 - Sauf à utiliser des testbeds, tout cela doit se faire en bonne intelligence avec les expériences.
 - Les propositions actuelles ont une certaine coloration ATLAS, il serait bon que l'on regarde ce qui peut intéresser les autres expériences. Dans le passé il

y avait eu une proposition de regarder les accès distants pour ALICE (avec éventuellement regroupement du stockage de sites) mais cela avait été découragé par ALICE Computing sur des sites en production. Peut-être que leur position a changé avec la situation très tendue que vit ALICE au Run2 et les perspective du Run3. Concernant LHCb, Stefan nous confirme que, vu le modèle de calcul, le caching n'apporte rien.

- Catherine ajoute qu'il ne faut pas perdre de vue que le but ultime de DOMA est de diminuer les coûts et rappelle que 5 personnes de LCG-France sont impliquées dans le groupe de travail "Syst. performances and cost modeling", très complémentaire des R&Ds que les sites se proposent d'investiguer.

Se pose le problème des moyens financiers et humains. Une petite discussion est engagée et Volker indique que cela pourrait être porté comme un projet technique IN2P3 et bénéficier d'un support financier. Comme les discussions internes n'ont pas eu lieu plusieurs sites et comme certains autres sites pourraient vouloir lancer des R&D suite à cette discussion, la proposition de LCG-France est qu'elle soit tenue au courant des évolutions durant le mois de juillet et si une masse critique est atteinte qu'une demande projet IN2P3 soit préparée pour la rentrée.