

Le stockage objet DataCore Swarm à l'IBMP et plus généralement : comment faire tenir un océan dans une bouteille...

mercredi 8 octobre 2025 11:15 (45 minutes)

La rétention et la sécurisation des données de la recherche sur le long terme nécessitent des systèmes de plus en plus capacitifs, évolutifs et résilients. Dans cette optique, l'IBMP a investi dans la solution commerciale DataCore Swarm, qui a été déployée dans le laboratoire. Nous avons choisi cette solution du fait de sa structure objet, de sa facilité de déploiement et d'administration, ainsi que pour des raisons de coûts. Les premiers usages à ce jour concernent le stockage des grands jeux de données de bio-informatique, mais aussi le stockage des données intégrées au sein du cahier de laboratoire électronique ElabFTW déployé à l'IBMP. Dans une vue plus prospective, nous aborderons les développements logiciels à l'étude au sein du laboratoire afin de donner toute sa puissance à l'architecture objet du stockage, indépendamment de la solution S3, qui est notoirement sous-utilisée et constitue un frein à son adoption par les utilisateurs finaux.

Nous explorerons avec DataCore la problématique de la production de données qui dépasse la capacité de stockage généralement présente sur nos campus. En effet, les jeux de données grandissent exponentiellement, mais les budgets IT sont au mieux stables, s'ils ne baissent pas. Entre financements de recherche contraints et explosion des volumes, l'équation est complexe : comment maximiser la valeur scientifique tout en optimisant les investissements IT ? Comment équilibrer les besoins de performance HPC pour l'analyse active et les stratégies de stockage objet économiques pour l'archivage à long terme ? Enfin, comment éviter que la recherche d'un jeu de données spécifique ne devienne elle-même un projet de recherche ? Nous allons examiner, dans le contexte particulier de la recherche, des architectures HPC haute performance couplées avec le stockage parallèle, les stratégies de tiering intelligent vers du S3 et stockage objet, les modèles économiques hybrides on-premise, les systèmes d'indexation intelligente et l'enrichissement automatique des métadonnées (parce que "grep" a ses limites). Enfin, nous aborderons l'optimisation des coûts via la migration automatique HPC vers objet pour une rétention à long terme rentable afin de laisser l'essentiel des ressources pour la découverte scientifique.

Orateurs: M. EVRARD, Jean-Luc (IBMP); M. LBARS, Laurent (DataCore)