

L'IN2P3 (CNRS) protège ses données de recherche critiques stockées sur bandes avec la solution StorSentry™ d'Hi-Stor.

Le Centre de Calcul de l'IN2P3/CNRS (CC-IN2P3) a choisi et mis en œuvre StorSentry™ au printemps 2009 pour anticiper les risques de perte de données et fiabiliser l'exploitation de sa robotique de stockage bandes.

Le Centre de Calcul de l'IN2P3 (CC-IN2P3), est une unité du CNRS en partenariat avec l'Institut de recherches sur les lois fondamentales de l'Univers (IRFU) du Commissariat à l'Energie Atomique (CEA). Situé à Lyon – Villeurbanne, le CC-IN2P3 est spécialisé dans la fourniture de services informatiques nécessaires à l'analyse et à l'interprétation des processus fondamentaux de la physique subatomique. Le CC-IN2P3 est notamment l'un des centres de ressources du LHC, le grand collisionneur de hadrons installé au CERN (près de Genève).

Le caractère indéterministe de la physique subatomique requiert en effet l'analyse statistique de milliards d'interactions entre particules afin d'étudier les processus fondamentaux. Face à cette problématique, le CC-IN2P3 a développé une expertise dans plusieurs domaines, et en particulier dans la conception et la mise en œuvre d'une architecture pour le stockage massif de données (bases de données, réseau de stockage, disques sécurisés, **bandothèque automatisée**, accès hautes performances).

Nous avons rencontré Frédéric Azevedo, coresponsable de l'exploitation de l'infrastructure de stockage sur bandes, qui a participé au choix de la solution StorSentry™, et est en charge de son utilisation au sein de l'exploitation informatique du CC-IN2P3.

La robotique bandes mise en place au CC-IN2P3 par Sun correspond aux enjeux de volumétrie d'un grand centre de recherche. Le périmètre actuel est d'environ 90 lecteurs et 20 000 bandes (LTO4 et T10K avec 3 librairies SL 8500 sous ACSLS, avec HPSS et TSM) pour une capacité de plus de 7 pétaoctets. Il sera amené à évoluer dans les 5 ans selon une croissance moyenne de la capacité de stockage d'au moins 1 pétaoctet par an.

Le taux d'activité de la robotique est très élevé (environ 20 000 montages par semaine) du fait des nombreux chercheurs accédant aux informations en permanence.

Les bandes contiennent à la fois les données brutes d'acquisition (les plus critiques pour les chercheurs en cas de perte) mais aussi les résultats de calcul et de simulation (qui représentent des heures de calcul et de travail des chercheurs) dont le coût de reconstitution en cas de perte peut être très élevé. La perte de ces données est extrêmement pénalisante pour les utilisateurs du CC-IN2P3 et se traduit par des coûts de reconstitution et des temps d'interruption d'accès de plusieurs mois.

En conséquence, le CC-IN2P3 se doit d'assurer la qualité optimum de son infrastructure de stockage bandes pour satisfaire pleinement aux exigences de ses utilisateurs.

Frédéric Azevedo nous a rappelé les enjeux principaux ayant amené à la recherche d'une solution et au choix de StorSentry™ :

"En recherchant une solution de contrôle de la qualité et de la performance de notre robotique de stockage sur bandes pour le CC-IN2P3, nous voulions :

- Minimiser les risques de perte de données en anticipant les problèmes de dysfonctionnements des lecteurs et des cartouches, et donc éviter ainsi les interruptions de service pour nos utilisateurs;*
- Optimiser l'utilisation de notre infrastructure par une meilleure analyse de ses performances, afin de mieux répartir les charges et réduire les coûts."*

Le CC-IN2P3 avait déjà étudié une première fois la solution StorSentry™ en 2004, mais elle ne répondait pas alors complètement aux exigences d'évolutivité liées au contexte à fort volume. Puis, du fait de ses évolutions récentes en termes d'architecture et d'ouverture, et sur la recommandation de son contact technique chez Sun, le CC-IN2P3 a décidé de l'envisager de nouveau à partir d'octobre 2008 et de la tester sur son site début 2009. Le CC-IN2P3 a aussi étudié et testé des solutions concurrentes qui se sont montrées nettement moins fonctionnelles que StorSentry™ pour répondre efficacement aux besoins immédiats.

StorSentry™ a finalement été retenu et mis en place au printemps 2009 par le CC-IN2P3 pour les principales raisons évoquées par le coresponsable de l'exploitation de l'infrastructure de stockage sur bandes :

"StorSentry™ est la solution la plus aboutie pour nous permettre de contrôler immédiatement notre robotique de stockage sur bandes. La qualité des diagnostics et des recommandations faits par le produit nous permet d'avoir la réactivité requise en cas de problème détecté et fiabilise encore plus l'exploitation informatique au quotidien. Par ailleurs, la richesse des informations recueillies nous permet d'envisager des analyses approfondies en utilisant la base de données de StorSentry™ grâce à l'accès direct SQL en lecture qui est proposé. Enfin, l'ouverture de l'architecture de la solution (API et SNMP) nous donne le potentiel requis pour aller plus loin dans l'automatisation des actions à mener pour faire évoluer l'infrastructure (par exemple, le remplacement de bandes ou de lecteurs)."

Après 6 mois de fonctionnement de StorSentry™ au CC-IN2P3, il confirme ainsi que la solution apporte bien les bénéfices attendus par le CC-IN2P3, et notamment le principal qui est une meilleure anticipation des problèmes de cartouches et de lecteurs afin d'intervenir de façon préventive. Il a pu le constater lui-même au quotidien en faisant remplacer des lecteurs défectueux (ce qui a été validé avec Sun) et également des cartouches.

L'équipe en charge de l'exploitation quotidienne de la robotique est tout à fait satisfaite des services et du support des équipes techniques d'Hi-Stor avec lesquelles elle est en contact régulier. Elle prévoit une utilisation future du produit plus étendue, en utilisant les capacités d'accès SQL à la base de données pour développer ses propres analyses sur le comportement de l'infrastructure. Elle travaille également sur la mise en place de procédures automatiques communes avec Sun sur la base des recommandations de StorSentry™. Enfin, le CC-IN2P3 envisage d'utiliser la capacité offerte par StorSentry de mise en place d'agents de collecte de données dans le flux (actuellement le CC-IN2P3 a mis en œuvre le mode de collecte au niveau du SAN avec le module « Agent Less » de StorSentry™) pour les données de sauvegarde gérées sous TSM.

Frédéric Azevedo indique en conclusion : « *StorSentry™ d'Hi-Stor répond bien aux enjeux d'un centre de calcul comme le nôtre, doté d'une très grande robotique de stockage bandes, grâce à sa pro activité qui nous permet d'anticiper sur les défaillances des composants et apporte la fiabilité requise pour notre exploitation quotidienne.* »