

Compte-rendu de la réunion du groupe de travail Stockage

Date : 14h, le 2 mai 2013

Lieu : Salle de réunion, EOST

Participants :

Alain SARTOUT (IRMA), Guillaume SEITH (IGBMC), Rémy FRITZ (IGBMC), Marc SCHAMING (IPGS), Fabien MULLER (IPCMS), Xavier ROSE (IPGS), Michel RIGENBACH (DI / HPC), Alain HEINRICH (DI), Jérôme PANSANEL (IPHC), Alexis PALATICKY (IRMA), Nicole MAHR (LHYGES), Jean-Yves HANGOUËT (Observatoire Astronomique)

Excusés :

Yasmina RAMRANI (LIVE), Emmanuel BLINDAUER (IUT Robert Schuman)

1 Tour de table

Un tour de table est réalisé. L'objectif est que chaque participant se présente, ainsi que son infrastructure et ce qu'il attend du groupe de travail.

Alain SARTOUT présente l'infrastructure de l'IRMA. Elle est composée d'une solution mutualisée comportant des baies Iomega. Il est prévu l'achat d'une solution de 100 To. Il est intéressé par le partage de connaissance autour du stockage, ainsi que par la mise en place d'une solution de stockage mutualisée sur Strasbourg. Les discussions avec les différentes personnes présentes mettent en évidence les problèmes de financement, d'hébergement et de fonctionnement de ce type de solution. Ces problèmes ne sont pas insurmontables, mais doivent être pris en compte si nous souhaitons proposer une solution viable.

Ensuite, Guillaume REITH présente le stockage à l'IGBMC. 60 To en NAS et 30 To en SAN sont disponibles sur des baies EMC répliquées de manière synchrone entre deux bâtiments. À partir de cet été, 130 To en NAS et 30 To en SAN seront disponibles sur des baies DELL MD3660 répliquées de manière asynchrone entre les deux bâtiments avec des têtes NAS NX3610. Le logiciel HSM est utilisé pour l'archivage avec la solution openArchive (connue sous le nom de Grau Archive).

Puis, Rémy FRITZ est également intervenu. Il est très intéressé par la partie stockage pour les expériences.

Marc SCHAMING gère le stockage des données pour la Géophysique. Ces données doivent être gardées longtemps. Il essaie de trouver la bonne solution entre disques et bandes. Il recherche une solution haute-disponibilité pour le BCSF (Bureau Central Sismologique Français).

Nicole MAHR fait part que le stockage et l'archivage sont des problèmes au sein du laboratoire et restent toujours à l'étude. Elle utilise actuellement un système de sauvegarde simple (sauvegarde incrémentale journalière + complète chaque mois) redondant et dupliqué physiquement pour une volumétrie actuelle ~20 To. Elle détaille également les problèmes de financement auxquelles elle est confrontée.

Ensuite, Fabien MULLER présente le stockage à l'IPCMS. Il se compose de baies Infortrend (Transtec) de 60 TB avec une répllication à l'IPHC par rsync. Les sauvegardes sont effectuées sur des bandes LTO5. Il a également une baie pour le calcul. Il souhaite s'équiper avec une baie supplémentaire certifiée VMWare, car il rencontre des problèmes actuellement avec la sauvegarde

des machines virtuelles. Une sauvegarde des données est réalisée à l'IPHC sur une baie Infortrend. L'outil de sauvegarde utilisé est TimeNavigator.

Xavier ROSE détaille le stockage à l'IPGS (calcul et administratif). Le stockage pour le calcul est principalement du scratch (40 To), alors que le stockage pour l'administratif est sur une baie QNAP. Il recherche une solution pour la sauvegarde des données (actuellement sur disque), ainsi que celle des plate-formes mobiles (alternative à DropBox pour les chercheurs en déplacement).

Michel RIGENBACH présente le stockage associé au pôle HPC. Le stockage est basé sur la technologie GPFS (IBM) et est composé de disques de 3 To. C'est une solution moyen de gamme car ils avaient un budget ~160k€ pour une volumétrie de 300 To. Les baies sont reliées entre elles par Fibre Channel et connectées aux serveurs de calcul par InfiniBand. Pas de backup.

Alain HEINRICH décrit l'infrastructure de la DI. Elle se compose d'une baie NetApp 3140 d'une volumétrie de 200 To. La sauvegarde est faite sur une baie Transtec avec le logiciel TimeNavigator. Pas d'évolution prévue, mais la baie doit être remplacée d'ici 2 ans.

Jérôme PANSANEL présente le stockage de l'IPHC. La partie /home (Linux) est sauvegardée avec le logiciel Netvault sur bande. La sauvegarde des postes et serveurs Windows est effectuée avec une autre solution. Le stockage de grille est effectué sur des baies SUN X4500 / X4540 et HP MDS 600 (volumétrie de 1,1 Po). Il n'y a pas de sauvegarde des données de grille. Pour les baies SUN (hors garantie), une maintenance par Osiatis est actuellement étudiée.

Jean-Yves HANGOUET a déjà présenté sa futur architecture lors de la dernière réunion X/Stra. Dans le cadre de l'hébergement des données CDS, il doit regrouper le stockage sur une infrastructure fiable et avec une volumétrie qui évolue de manière conséquente. Il a opté pour une solution DELL avec des baies iSCSI (10 Gb) et DAS. Il y aura une réplication entre les baies (site distant). L'infrastructure est prévue pour pouvoir doubler la volumétrie sur 5 ans. Il est prévu également de mettre un PRA en place. Pas d'intérêt pour le tiering.

2 Organisation du groupe de travail

Il a été décidé d'organiser 3 réunions de ce groupe de travail par an, les échanges se faisant principalement par mail.

3 Livre blanc

Lors de cette réunion, de nombreuses questions ont été soulevées nécessitant étude, réflexion, essais car les réponses tant techniques que politiques et financières ne sont ni évidentes, ni imminentes :

- La mutualisation du stockage pour une haute disponibilité : avantages et inconvénients au sein d'un laboratoire, d'un regroupement de laboratoires, sur un plan local ou sur un plan national ?
- Comment tenir compte de l'évolution de la volumétrie ?
- Nécessité de stocker toutes les données ?
- Comment préserver le stockage sur un long terme ?
- Quels sont les services proposés par une mutualisation ?

Jérôme PANSANEL propose de rédiger un livre blanc sur la problématique du stockage dans les laboratoires de la région. Il pourrait être décliné en deux versions, une plutôt technique et complète pour les personnes sur le terrain et une seconde plus brève (moins de cinq pages) pour les directeurs

de laboratoire.

4 Projet plate-forme de stockage

Fabien MULLER évoque la possibilité de réaliser un dossier de financement / projet pour mettre en place une plate-forme commune de stockage. Alain SARTOUT, soulignant l'intérêt d'une telle plate-forme, montre toutefois que les problèmes soulevés sont nombreux.

Clôture de la réunion : 16h30.