

REPORTAGE: Congrès 'storage' SNWE à Cannes (fin)

Comment faire comprendre aux utilisateurs d'espaces de stockage à géométrie exponentielle qu'acheter et acheter toujours du disque ou de la bande à l'infini n'est pas la seule solution?...

Réfléchir avant d'agir D'habitude on stocke tout en n'importe quoi en priant Dieu de ne pas être obligé de devoir restaurer ce qui, malgré des tableaux de bord dignes de l'informatique des années 80 (tant les outils, bien que puissants, sont peu conviviaux !) n'est pas à la portée de chacun sans un minimum d'organisation. D'où des masses de données inutiles et pourtant encombrant les bandothèques. En réponse à cette explosion combinatoire, SNWE a fait apparaître **deux grandes tendances** : – d'une part, un modèle de **hiérarchisation du stockage** qui répartit les données en fonction de leur utilisation en stockage primaire (généralement, sur disques), stockage secondaire (sur bande) et archivage (sur support optique). Ceci permet de réduire considérablement le prix du Go stocké et surtout cette approche apprend aux utilisateurs à donner un poids à leurs données; – d'autre part, une **intelligence logicielle désormais intégrée** dans le matériel au niveau des switches (ou des directors, dans les architectures nécessitant une certaine redondance). Cet apport logiciel allégeant d'autant le fonctionnement des serveurs et, l'examen des données étant effectué à la volée, la bande passante ne s'en trouve pas affectée. En bref, c'est vers la consolidation des plates-formes de stockage et vers plus d'intelligence embarquée (tout en proposant une meilleure intégration des outils et des périphériques de stockage) que l'on va. Mais la sauvegarde-restauration des grandes bases de données de l'entreprise nécessite encore d'être pensée et repensée pour enfin aboutir à une solution qui soit efficace. Il reste que l'ergonomie des outils devrait être plus soignée: les paradigmes pour l'heure employés ne dépassent guère la feuille Excel un peu sophistiquée. À quand un véritable GUI stockage avec un code couleur pour indiquer la saturation des serveurs? A quand surtout une programmation intuitive des sauvegardes et des chemins de données (et des types de données) qui puisse être effectuée par des personnes n'ayant pas reçu une formation (si l'on compte le nombre de tableaux de bord actuellement disponibles sur les programmes) digne de celle d'un pilote d'Airbus? Alors que les éditeurs prônent tous plus de simplicité, ils devraient commencer par intégrer plus d'ergonomie ! **Stockage de données : une histoire belge** Que penseriez-vous de quelqu'un qui dirait effectuer le back-up de 300 serveurs sous Windows, Linux et Solaris (pour un volume de données de plus de 2,7 Teraoctets) en moins d'une heure ? La réponse aurait de fortes chances d'évoquer l'asile ou l'histoire de la sardine bloquant le port de Marseille. Pourtant, c'est la prouesse que réalise quotidiennement Dedigate, un fournisseur de services d'hébergement pour les ASP, les designers Web, les intégrateurs système et les entreprises spécialisées dans l'e-business. Pour ce faire, elle utilise la technique de back-up intelligent DC-Protect développée par une petite SSII belge: DataCenter Technologies. L'immense intérêt de ce système s'appuyant sur des agents de copie et des filers NAS, c'est que le volume global du back-up sur une période de trois mois (durée requise de conservation de tous les fichiers demandée par Dedigate) ne dépasse pas les 2 tera-octets (compression oblige) et que, de plus, une restauration intégrale d'un système s'effectue également en une heure. La sécurité des données est, elle aussi, garantie: toutes les données sont encryptées lorsqu'elles traversent le réseau, les métadonnées de backup étant, quant à elles, protégées par un mot de passe. La beauté de ce

« *pico-backup* » (toutefois, uniquement réservé à la sauvegarde de fichiers et non de bases de données) tient à son mode de fonctionnement. Fondé sur un examen du contenu des fichiers au moment de l'opération de backup, DC-Protect sauvegarde les fichiers en fonction de leur contenu et non pas uniquement en fonction de leur nom ou de leur date de modification, ce qui évite l'encombrement de l'espace de stockage par une foultitude de copies d'un même fichier, considéré comme original parce que renommé ou présent sur une station de travail différente. Le principe est très astucieux. Chaque fichier reçoit une empreinte unique (codée sur 128 bits) résultant de l'examen de son contenu. Deux contenus n'ayant qu'une chance sur un google (1 suivi de 100 zéros) de recevoir une empreinte identique, l'empreinte est donc unique et valide sur l'ensemble du réseau. Toute modification faite au fichier engendrera une adaptation de l'empreinte. Les fichiers étant ainsi identifiés sur la base de leur contenu, ceux qui ont déjà été sauvegardés n'ont plus besoin d'être physiquement réexpédiés sur le réseau. D'où un certain nombre d'avantages qui vont faire rêver plus d'un responsable d'exploitation. En effet, cette solution réduit d'un facteur 100 les besoins en espace de stockage (idem de la bande passante), le temps de backup étant 100 fois plus rapide. Pas mal, non? Mais, ce n'est pas tout. Comme le souligne Philippe Jansens, le directeur administratif de DataCenter Technologies : « *Cette approche orientée contenu va bien plus qu'une simple fonction de backup-restauration. Elle permet par exemple d'établir un inventaire des fichiers au niveau organisationnel, ce qui s'avère fort utile lorsqu'on veut faire de la gestion de versions, de la rétention (conservation des données effacées), de l'archivage, etc. Les technologies classiques ne peuvent offrir facilement de tels avantages, puisque leur backup est effectué au niveau du bloc de données et non pas à celui du fichier.* » Il est rare qu'un produit nous enthousiasme réellement. C'est ici le cas. Pour plus d'infos www.datatechnologies.com