

# Stockage et ILM : pour des données autogérées

Dopé par un beau succès outre-Atlantique, ce concept débarque en Europe. Jon William Toïgo, célèbre consultant et auteur de nombreux ouvrages sur le stockage, nous a livré ses réflexions. Selon lui, toute solution d'ILM devrait reposer sur deux caractéristiques fondamentales: -d'une part, une méthode fournissant automatiquement les ressources de stockage nécessaires aux applications qui en font la demande (dans une optique pas plus pas moins, donc pour un encombrement et un délai d'accès parfaitement circonstanciés), -d'autre part, un ratio d'optimisation du coût d'exploitation des plates-formes de stockage qui permette, par simple analyse statistique, de déplacer les données les moins utilisées vers des supports de stockage bon marché.

Jusque là, rien de très original, sauf qu'il n'existe pas pour l'heure de standard en matière d'ILM, pas plus que de stratégie réellement adaptée à l'infrastructure informatique actuelle. Premier espoir pour une rapide évolution de cette situation: Windows 'Longhorn', le futur système d'exploitation de Microsoft, successeur de XP. En effet, celui-ci remplacera les systèmes de fichiers par une base de données SQL dans laquelle lesdits fichiers sont en fait traités comme des objets binaires. Ce modèle offre l'avantage de doter les données d'un en-tête d'autodescription (cf. infra). Il en est de même sur la version 10g d'Oracle et sur la prochaine mouture de DB2 Open Edition d'IBM. La dérive hors des architectures ILM propriétaires est donc tout juste amorcée. Il n'empêche que, pour le moment, il convient d'être encore prudent lorsque l'on veut gérer correctement ses données sur la durée et qu'il est souvent préférable de demeurer dans l'expectative face à des fournisseurs qui clament à qui mieux mieux qu'ils sont les plus beaux et les plus forts en matière de gestion du mûrissement des données, à moins d'avoir de bonnes raisons pour planifier une infrastructure en fonction des produits d'un seul fournisseur (le fameux syndrome de La Boétie : *«parce que c'était IBM, parce que c'était moi!»*). Or, précise Jon W. Toïgo, il n'est pas d'ILM sans infrastructure pensée pour en tirer le meilleur parti. **La déferlante iSCSI** Dans ce cadre d'ailleurs, les responsables du stockage devraient prêter un peu plus d'attention à la déferlante iSCSI. L'avantage c'est ici d'avoir une seule et même infrastructure pour les applications et les données (le réseau IP, même si on le sépare en deux «domaines», l'un réservé aux progiciels et aux transactions, l'autre au stockage proprement dit, une infrastructure qui s'appuiera sur un standard incontournable en matière de réseau, Ethernet, en version 10 Gbps. Cette conjonction au niveau de l'interconnexion des ressources devrait d'ailleurs reléguer à plus ou moins long terme le Fibre Channel dans un marché de niche, s'il faut en croire l'expert américain. Car, comme le fait remarquer Franck Didi, fondateur d'InfineData, société de conseil spécialisée dans le stockage : *«Le réseau SAN dédié n'est pas toujours la solution idéale. Un serveur NAS milieu ou haut de gamme connecté aux librairies de bandes soulage lui aussi le réseau d'entreprise»*. **Décrivez, décrivez, il en restera toujours quelque chose** L'auto-description des données, quant à elle, est un préalable indispensable pour une gestion «bien tempérée» (merci Jean-Sébastien !) du cycle de vie de l'information. Ceci pour une raison simple. Faire naître une donnée en la caractérisant d'emblée selon son type, son importance, sa durée probable de vie, c'est lui offrir dès le départ un «acte de naissance» qui lui donne un véritable état civil. Ce n'est toutefois pas suffisant pour aller au-delà d'un simple placement des données sur tel ou tel support en

fonction de sa maturité (d'autant que, le plus souvent, les données de plus de six mois sont considérées comme obsolètes dans 80 % des cas). L'arrivée du numérique comme source de preuve dans les transactions commerciales bouleverse la donne et oblige à aller plus loin, notamment en obligeant (dixent quelques arrêts récents, dont l'affaire Enron) toute entité concernée par un litige à fournir les informations requises dans les plus brefs délais. D'où l'idée d'aller encore un pas plus loin en proposant une gestion des données qui s'apparente largement à celle d'autres ressources. Encore faudrait-il ne pas perdre de temps. Autrement dit, ce devrait être aux applications de remplir (automatiquement) cet état civil dès la naissance de la donnée, idem pour la mise à jour de la fiche de renseignements au fur et à mesure de l'évolution (et de l'exploitation) de l'information. C'est ce à quoi s'emploie une nouvelle technologie de stockage, baptisée **DRM (Data Resource Management)**, laquelle considère toute information comme une entité vivante et dont l'auscultation doit tenir compte de ses inter-relations avec les applications, les utilisateurs (tant internes qu'externes), l'infrastructure d'information, etc. Tout un programme, dont hélas les programmes justement ne tiennent pas encore compte. A suivre.