



# STOCKAGE | Gérer l'inflation des données numériques

Le stockage n'est plus réservé aux grands comptes. De nouvelles technologies le rendent accessibles aux PME. Mode d'emploi.

**F**ichiers CAO 3D au bureau d'études qui se multiplient; photographies ou vidéos numériques prises lors de chaque nouvelle visite de site; pièces attachées aux e-mails; documents dématérialisés... sans parler des bases de données de l'ERP dupliquées plusieurs fois ou de l'avalanche de données de traçabilité produits par la logistique: les disques durs des serveurs des entreprises débordent. «En 2008, le volume de données de type fichiers a dépassé celui des bases de données», affirme Philippe Nicolas, le président du comité régional français du SNIA (Storage Networking Industry Association). «Personne ne veut se limiter. Ces données sont la valeur de l'entreprise», observe de son côté Philippe Blanc, consultant chez Atos Origin.

## CENTRALISER LES INFORMATIONS

Face à cette inflation galopante, la première réaction consiste souvent à multiplier les disques durs, voire à acheter de nouveaux serveurs. Cette solution a ses limites. Pas tant en termes de prix (le coût d'un serveur, assez bas, est aujourd'hui peu dissuasif) qu'en termes d'efficacité et de sécurité. Or, personne n'a le temps ni les ressources de trier manuellement les données à conserver de celles à archiver ou à supprimer. «Le volume des images que nous utilisons pour calibrer nos logiciels d'optimisation photographiques embarqués double tous les ans. Nous n'avons pas le temps de faire le ménage», confirme Frédéric Boisseau, le directeur informatique

## Overdose

➤ **281 exaoctets** (milliards de Go) de données numériques ont été produites dans le monde en 2007, dont 70 % par des particuliers.

➤ **En 2011**, le volume du monde numérique atteindra 1800 exaoctets. Or, la capacité de stockage ne sera pas suffisante pour permettre un accès en temps réel à toutes ces informations. Il va falloir trier et archiver.

SOURCE IDC/EMC

de la start-up francilienne DXo Labs. La solution? Le stockage. Il consiste à centraliser les données sur des disques durs, de plus ou moins grande capacité, non plus répartis dans chaque serveur, mais rassemblés dans une ou plusieurs baies. Centralisée, l'information est plus facile à sécuriser et à retrouver. Suivant la nature des éléments, on distingue deux méthodes de stockage: le mode bloc (ou NAS) bien adapté aux bases de données; et le mode fichiers (ou SAN) pour les données non structurées, bureautiques ou multimédias.

La bonne nouvelle: des technologies récentes rendent plus accessibles ces deux méthodes de stockage aux PME. Auparavant, il fallait installer un réseau en fibre optique (fiber channel) dédié entre les serveurs et les baies de stockage pour garantir un débit d'accès aux données suffisant afin de ne pas ralentir les applications. Depuis 2004, le standard iSCSI permet de connecter une baie de stockage via un

réseau IP de l'entreprise, avec des performances acceptables. Il autorise des transmissions de données de l'ordre de 1Go/seconde. C'est quatre fois moins que sur les réseaux de fibres optiques (4Go/s), mais suffisant pour le stockage d'e-mails ou de fichiers non structurés, comme les documents bureautiques ou multimédias. «L'idéal est de créer un réseau IP réservé au stockage», rappelle Jean-Pierre Potez, expert réseau chez EMC. Reste à choisir le (ou les) bon type de disques durs sur lesquels stocker ses données. Le panachage est de plus en plus facile, voire recommandé.

## HIÉRARCHISER ET SÉCURISER LES DONNÉES

Pour optimiser les coûts (1 To vaut environ 1000 euros aujourd'hui), l'une des astuces consiste à hiérarchiser son stockage en fonction de la criticité des informations. En moyenne, seules 20 % des informations sont consultées fréquemment, 40 % rarement et le reste jamais. Et au-delà de 30 jours, une information n'est généralement plus demandée. «Sur des systèmes de production, après 60 jours, les données seront même considérées comme toxiques», prévient Hu Yoshida, le directeur technique d'Hitachi Data Systems.

En clair, il faut stocker intelligemment. Inutile donc de tout vouloir enregistrer sur des disques fiber channel, très rapides, mais plutôt chers et pas forcément économes en énergie. Une bonne partie peut généralement être migrée sur des disques SAS ou SATA (lire ci-contre). Les logiciels utilisés pour classer les données en fonction

## LES ACRONYMES À CONNAÎTRE

**FIBER CHANNEL = RÉSEAU FIBRE OPTIQUE** Protocole de stockage haut débit (vitesse d'accès 4 Go/s) mais aussi nom du réseau dédié et des disques les plus rapides (15 000 tr/min)

**ISCSI = INTERNET SCSI** Protocole de connexion SCSI (Small Computer System Interface) sur réseau internet classique ou TCP/IP, (vitesse d'accès 1 Go/s)

**NAS = NETWORK ATTACHED STORAGE**  
Type de stockage en réseau en mode fichiers

**SAN = STORAGE AREA NETWORK**  
Type de stockage en réseau en mode bloc

**SAS = SERIAL ATTACHED SCSI** Protocole d'accès et type de disques de stockage moyen débit (3 Go/s et 10 000 tr/min)

**SATA = SERIAL ATA** Protocole et type de disques de stockage bas débit (150 Mo/s, 7 500 tr/min)



## LAROCHE

Siège Chablis (Yonne)

### Activité

Viticulture, hôtellerie et restauration sur quatre domaines en France, au Chili et en Afrique du Sud

Effectif 200 salariés

Chiffre d'affaires 2007  
33 millions d'euros

Volume de données  
30 Go d'e-mails et 160 Go  
de base documentaire  
multimédia

Prévoyant. Franck Hauer, le directeur informatique de Laroche. Une sauvegarde par serveur? C'est fini! Les données que les salariés de l'entreprise doivent conserver sont automatiquement acheminées dans une baie de stockage d'une capacité de 600 Go.

des dates de création de fichiers, de dernière modification et de dernière consultation, sont issus de la famille des logiciels de gestion du cycle de vie des données ou ILM (Information Cycle Management). Certains ont été spécifiquement développés pour les messages électroniques, type Microsoft Exchange, pour optimiser leur archivage en toute transparence pour les utilisateurs. «Attention, ces logiciels ajoutent du code, sont assez intrusifs et leur mise en œuvre délicate», prévient Alain Clément, chef de produit stockage chez HP.

Pour mieux ranger, il faut aussi ne pas stocker plusieurs fois la même information. «Les technologies de déduplication, basées sur la notion d'empreinte numérique et issues du monde de la cryptographie, sont la nouvelle "killer application" du stockage», explique Philippe Nicolas, le président du comité régional français du SNIA. En environnement bureau-tique, elles permettent de réduire de

## LAROCHE CENTRALISE SES DOCUMENTS

Chez Laroche, le poids des photos, étiquettes et autres documents bureautiques, est passé de 400 Mo en 1995 à 160 Go aujourd'hui. «La croissance est exponentielle, observe Franck Hauer, le directeur informatique. Pour autant, nous ne voulons pas limiter les utilisateurs. Ces fichiers constituent la mémoire de l'entreprise.» Or, si les utilisateurs ont déjà l'habitude de sauvegarder leurs documents, non sur leur PC mais sur des serveurs centraux, la gestion de ces derniers devenait un casse-tête. «Nous avions un point de sauvegarde par serveur, mais les données étaient dispersées.» Laroche a décidé d'y mettre un peu d'ordre.

Depuis un an, toutes les données sont sauvegardées automatiquement dans une baie de stockage d'une capacité de 600 Go... pour voir venir. Les e-mails sont traités à part. Pour des problèmes d'architecture (le mode bloc de la messagerie Exchange de Microsoft est incompatible avec le mode fichier choisi pour le stockage), ils ne sont pas encore stockés dans la baie, mais un logiciel permet de les dédoubler avant un archivage sur un serveur dédié. Par souci de cohérence, toutes les briques logicielles et matérielles de la solution, qui a coûté 60 000 euros, ont été fournies par le même fournisseur, EMC.

20 à 40 % les volumes utiles. Ces ratios font néanmoins débat, les modes de calcul variant fortement d'un éditeur à un autre. Un groupe de travail du SNIA tente d'y mettre un peu d'ordre et prépare un livre blanc à paraître fin 2008. Attention, le

«dédoublage» est gourmand en ressources de calcul et impossible à réaliser au vol, durant l'enregistrement des données. Il est plus adapté à l'étape de la sauvegarde. Passer au stockage permet donc de sécuriser plus facile- Suite page 50 ▶



50

## HOMMES ET COMPÉTENCES SAVOIR-FAIRE

► Suite de la page 49 ment ses données, en commençant, bien sûr, par effectuer des copies de sauvegarde. Des mécanismes automatiques de réplication sont généralement inclus dans l'intelligence des baies. On parle de technologie Raid, qui va d'une échelle 1 (simple copie conforme donnée, donc occupant le double de mémoire) à 6 (par un mécanisme d'optimisation, elle n'occupera que 1/8 de place supplémentaire). La fréquence de ces copies dépend de leur criticité et de la technologie utilisée. Les mécanismes les plus sophistiqués permettent de ne copier que les données ayant été modifiées. « Un système complémentaire de sauvegarde établit une photographie (snapshot) des données, sans les copier en intégralité. Elle suffit à reconstituer les données en cas de sinistre », explique Cyril Van Agt, responsable avant-vente grands comptes chez Net App. Pour plus de sécurité, il est aussi possible de crypter les données. Dans ce cas, les logiciels de déduplication ne sont plus opérants. Une sauvegarde en ligne, pour récupérer les données en cas de sinistre, viendra si possible compléter le dispositif de sécurité dans le cadre d'un plan de reprise d'activité.

### PRÉVOIR LES MONTÉES EN CHARGE

Pour choisir une solution, il ne faut pas se concentrer uniquement sur le prix, mais aussi sur l'efficacité du stockage. Pour l'optimiser, il est également possible de virtualiser les baies, au même titre que les serveurs, afin d'ajouter une couche logicielle qui déconnecte les applications du matériel les portant. « La virtualisation d'un stockage en réseau permet de travailler sur des containers logiques (et non physiques), que l'on peut déplacer et redimensionner à la volée pour libérer de la place », explique Cyril Van Agt. Elle facilite ainsi le déplacement des données sur différents disques en fonction de leur criticité, ou « tiers storage management », comme évoqué plus haut. Elle permet aussi d'optimiser le taux d'utilisation de 40 à 60 %, voire de 70 % si on y adjoint des technologies de « thin provisioning », une méthode d'allocation fine des ressources disponibles de stockage en fonc-

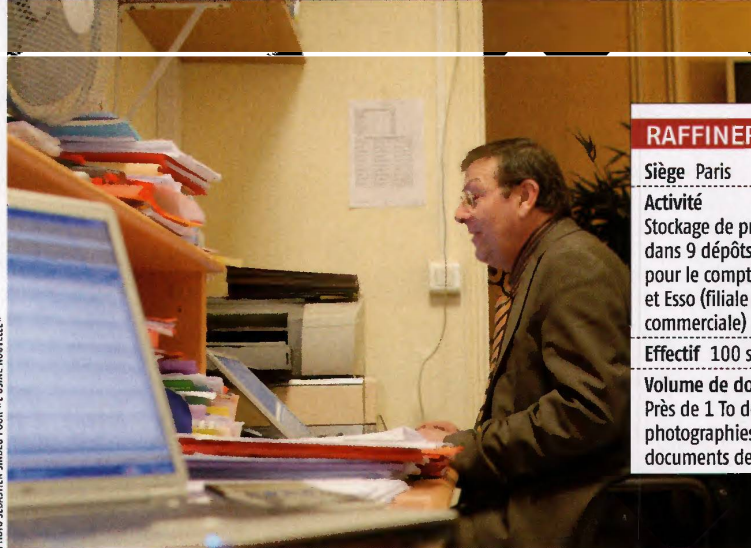


PHOTO SÉBASTIEN SINDÈR POUR « L'USINE NOUVELLE »

### RAFFINERIE DU MIDI

**Siège** Paris  
**Activité** Stockage de produits pétroliers dans 9 dépôts en France, pour le compte de Total, Shell et Esso (filiale conjointe non commerciale)  
**Effectif** 100 salariés  
**Volume de données** Près de 1 To de plans, cartes, photographies, plannings et documents de maintenance

En pointe. Pour José Nunez Alvarez, le responsable informatique de la Raffinerie du Midi, « la virtualisation permet de se détacher du matériel ».

## LA RAFFINERIE DU MIDI OPTÉ POUR LA VIRTUALISATION

En pleine réorganisation, avec la fermeture de onze sites sur vingt et le réaménagement des dépôts d'hydrocarbures restants, la Raffinerie du Midi avait un besoin criant de partage des données relatives aux sites. Or l'entreprise a déjà accumulé près de 1 To de photographies, plans et documents de maintenances. Détaché de Total il y a deux ans pour rénover l'informatique de la structure, José Nunez Alvarez a opté pour une technologie de pointe, la virtualisation, tant pour les sept serveurs (avec le logiciel VMware), que pour les trois baies de stockage HP, avec une solution logicielle indépendante signée Datamonitor. Le

logiciel permet de redimensionner à la volée les espaces de stockage des deux baies installées au siège, même en cours d'exploitation. C'est aussi lui qui gère la réplication vers une troisième baie, hébergée chez Neuf Cegetel. « La virtualisation permet de se détacher du matériel », explique le directeur informatique. D'ailleurs, il envisage déjà de troquer ses disques SATA, trop lents, contre des disques SAS, plus rapides. Ce qu'il pourra faire sans changer de baie et sans bloquer les utilisateurs. Pour encore plus de sécurité, dans le cadre d'un plan de reprise d'activité, une sauvegarde est aussi effectuée en ligne, avec la technologie d'Evault.

tion des besoins. Toutes ces nouvelles technologies ont bien sûr un coût. « Pour 1 euro dépensé en capacité de stockage, il faut compter 3 à 5 euros de logiciel pour l'optimiser », évalue Philippe Nicolas, du SNIA. Un ratio applicable uniquement pour les gros volumes. « Pour nous, la virtualisation reste trop chère », reconnaît Frédéric Boisseau, le responsable informatique de DXo Labs.

Dans tous les cas, une solution de stockage se construit sur mesure, généralement avec un prestataire intégrateur. Il aidera à évaluer la volumétrie utile et préconisera la meilleure solution de montée en charge. Les nouvelles baies sont généralement

prévues pour une montée en gamme, d'abord en ajoutant des disques, ensuite en changeant le module électronique de gestion. En moyenne, leur durée de vie est estimée à cinq ans.

Reste un point noir : la pérennité des méthodes de stockage. Pour l'instant, elles sont toutes plus propriétaires les unes que les autres. Conscients du problème, les industriels, rassemblés au sein de la SNIA, y réfléchissent. Ils préparent même un standard, le XAM (pour eXtensible Access Method). Ses spécifications devraient être publiées début 2009. Un pas vers l'archivage centenaire. On y reviendra. ■

AURÉLIE BARBAUX