

Administration des espaces stockage : l'ILM

à la rescousse

L'ILM? Concept lancé par Legato avec le soutien de StorageTek, certes. Mais c'est surtout un bon moyen de régler les problèmes de conformité à la réglementation en archivant ce qui est indispensable et surtout en mettant à disposition un moyen sûr et efficace de retrouver les données recherchées au bon moment -et au bon endroit. C'est, en substance, ce qu'explique Carolyn DiCenzo, analyste du Gartner Group. D'où, par voie de conséquence, une productivité accrue dans les accès aux bases de données, lesquelles ne seront plus encombrées par des données obsolètes ou de peu d'intérêt. Idem pour la messagerie d'entreprise (ôtons-nous de la tête l'illusion qu'Outlook ou Notes sont faits pour gérer des centaines, voire des milliers d'e-mails). Idem encore pour les applications métier, lesquelles pourront plus facilement définir des niveaux de service et donc les mises à niveau nécessaires dans ce cadre (plus de processeurs, de mémoire, de stockage). L'introduction d'une solution ILM dans l'entreprise est également l'occasion où jamais de faire migrer les données « en trop » vers des supports moins onéreux et de mettre en place une métrique qui permette de définir avec précision le 'workflow' que l'on désire traiter de la sorte. Par ailleurs, le passage d'un support disque rapide et fiable à un environnement plus léger est aussi l'opportunité de calculer le retour sur investissement que l'on est en droit d'attendre d'une telle approche.

Ceci n'empêche toutefois pas de penser que l'ILM vient rajouter un niveau de complexité dans une infrastructure « data » qui n'en a pas vraiment besoin. Il est vrai qu'il n'est pas simple de savoir à l'avance comment indexer les données et quelle durée de vie leur affecter. D'où la nécessité d'une planification rigoureuse ex ante que recommandent l'ensemble des éditeurs de progiciels d'ILM. Tout cela est bel et bon, mais un hic demeure. Lorsqu'on regarde d'un peu plus près les systèmes de définition de règles d'ILM, on s'aperçoit qu'ils sont loin d'être complets et qu'ils ne font pas vraiment de différence entre un document interne, un fichier d'accompagnement, une réponse, etc. Bref, il manque encore beaucoup d'éléments pour rendre l'ILM efficace, dont notamment une indexation qui intégrerait: – la nature intrinsèque de l'objet à sauvegarder, – le niveau de risque de litige pouvant être attaché à ce document, – l'évolution de ce risque dans le temps (car certains documents anodins, à l'origine, deviennent vitaux lorsqu'une contestation surgit), de la force probante pouvant être attachée au document (s'agit-il d'un document signé, unilatéralement ou par plusieurs parties? Fait-il l'objet d'un dépôt chez un notaire? S'appuie-t-il sur des textes dérogeant au droit commun? Met-il en jeu la responsabilité de l'entreprise? A-t-il des implications pénales en cas de non respect?). La liste des questions est ici trop longue pour pouvoir être traitée ici in extenso ; mais elle permet de se faire une petite idée de ce qui manque encore à l'ILM, par exemple un « moteur heuristique » appliqué au juridique qui simplifierait bien les choses. Aussi, se prend-t-on à rêver d'un temps proche où les éditeurs auront enfin résolu ce problème et proposeront, avec leurs produits, des gabarits par métier donnant un ensemble de règles prédéfinies selon les activités, bibliothèque de *templates* que l'utilisateur n'aura plus qu'à adapter à son cas particulier. Par ailleurs, il n'est pas si simple que cela de fixer au moment même de l'écriture d'un mail sa durée de vie probable, même si un minimum de formation et quelques secondes de réflexion épargneraient bien des efforts par la suite. En effet, si un 'pop-up' nous demandait par exemple si le mail en

question concerne une affaire interne ou un contrat en cours, nous gagnerions un temps précieux dans l'indexation du document et de ses réponses. Car tout bon système d'ILM utilise une approche plus ou moins « objet » pour indexer en cascade les documents. Si un premier document est vital, les réponses à celui-ci le seront aussi, etc. **Des phases de vie de quelle durée?** Il reste aussi à se poser en termes adéquats la durée des différentes phases de vie des données. En règle générale, on s'aperçoit que 75 % des données stockées sur un disque primaire ne sont pas retouchées pendant les trois derniers mois. D'ailleurs, 90 % d'entre elles ne sont plus jamais consultées. D'où l'envie de déposer pour un temps de telles données sur des disques moins onéreux (comme les disques SATA). C'est notamment le cas lorsque les données présentent un faible taux d'activité (s'agissant des entrées/sorties), ce qui concerne plus particulièrement les contenus fixes, le 'backup' de disque à disque, le stockage de masse, la « business intelligence », le stockage tiers des e-mails anciens, les dessins de CAO/DAO et l'archivage à court terme. C'est ainsi que l'on utilise des matrices de disques ATA en tant que support primaire pour ces applications, là où auparavant on aurait employé des disques plus fiables, mais ô combien plus onéreux. Il n'empêche que mêmes ces données connaissent une croissance rapide. Et à un moment ou à un autre, il faut les archiver pour laisser de la place aux données nouvelles. Ce qui implique un timing précis pour déterminer quand basculer les données vers un support magnétique de type bande ou vers un support optique. Mais là encore, le critère juridique ou légal de l'archivage n'est pas toujours prise en compte correctement, car seules des données non modifiables et donc confiées à un support non réinscriptible (WORM, CD-ROM, DVD-ROM, voire disque dur à secteurs non modifiables comme on en voit apparaître depuis quelque temps) auront valeur probante. Le problème, c'est que chaque donnée présente un profil particulier quant à son cycle de vie et ce même profil peut encore varier avec l'évolution de l'activité de l'entreprise. La définition de classes de données est certes une première étape, mais elle ne doit en aucun cas demeurer la même une fois pour toutes, sous peine d'engendrer à terme plus de problèmes qu'elle n'en résout. Il est donc nécessaire de choisir des produits d'automatisation qui permettent d'effectuer régulièrement des analyses de feed-back surtout l'utilisation des données et donc de remonter les informations nécessaires à la modification probable de certaines procédures (ou de certaines classes de données). Dans l'absolu, il serait bon de pouvoir ensuite tout centraliser dans un seul et même outil d'administration qui puisse gérer aussi bien la réplication la migration d'un support à un autre, la protection des données pour des raisons juridiques, leur effacement en fin de vie, etc. Dans la pratique, on en est encore bien loin, notamment pour les raisons précédemment invoquées. D'autant que s'il existe des outils de capture et de stockage de certains enregistrements électroniques, il est bien rare qu'ils collaborent les uns avec les autres. Nous avons déjà énoncé le cas de la messagerie électronique, laquelle ne peut aujourd'hui être traitée de façon efficace que par une indexation manuelle de chaque message. On imagine dans ce cadre la perte de temps et donc de productivité et donc le peu de succès d'une approche manuelle auprès des entreprises: elles préfèrent fermer les yeux pour ne pas penser au pire. Il manque indéniablement, ici, un outil que l'on ne peut qu'appeler de nos vœux pour un avenir le plus proche possible. **Les outils d'administration appliqués au métier** S'agissant des outils d'administration des processus métier, la situation est certes meilleure, mais en fait guère plus enviable. Quelques outils permettent de scruter le contexte dans lequel le message a été créé et permettent d'étendre l'indexation établie par métadonnées à tous les messages similaires lors de leur archivage. Cette taxonomie est encore toutefois grossière et se résume généralement à la définition de classes simples comme stockage, business, accès, extraction qui commanderont à leur tour le type de

support sur lequel seront stockées les données correspondantes. Le problème demeure au niveau de l'extraction, puisque cette dernière dépend de la disponibilité d'un périphérique particulier et de l'application pouvant lire les données (qui est encore capable de relire, par exemple des données stockées sur un disque 8 pouces à partir de programmes CPM réalisés sur Silz16, c'est-à-dire des données qui ont moins de trente ans et qui peuvent être requises pour des raisons pénales ou médicales). D'autant que la plupart des logiciels de backup utilisent un format propriétaire pour inscrire les données et qu'il est indispensable aujourd'hui d'avoir l'application d'origine si l'on veut pouvoir lire quelque chose de compréhensible. On attend donc la moulinette universelle, le deBabelizer du stockage qui permette de relire n'importe quel format quel que soit le nombre d'années passées depuis la création des données à extraire. Avis aux éditeurs? **Il reste un... casse-tête** Bref, le casse-tête de l'ILM, malgré les belles promesses d'automatisation qu'il propose, est loin d'être résolu pour peu que l'on veuille creuser un peu les conséquences d'une automatisation trop hâtive ou mal ciblée. Les acteurs de ce marché ont donc encore des efforts à fournir s'ils veulent vraiment convaincre les entreprises de basculer vers de tels systèmes. Quant aux P.M.E. n'en parlons même pas, sans gabarits pré-packagés, sans services de conversion proposés sous forme d'abonnement par les intégrateurs, elles n'y verront qu'un gadget de plus.