



## **Le téraoctet toxique**

**Arrêtons de stocker, ça déborde !**

*Le stockage en masse et mal géré de l'information  
peut pénaliser les performances de l'entreprise*

---

## Sommaire

---

- 2 *Le revers de la médaille*
- 5 *Une prolifération devenue encombrante*
- 7 *Les données : on les aime trop !*
- 9 *IBM-ILM : la solution intégrale*
- 10 *Un changement culturel nécessaire*

### **En route vers le zettaoctet**

**Alors qu'elles amassent les informations à un rythme accéléré, les entreprises s'exposent à des hausses de coûts et à une perte d'efficacité qui menacent leur aptitude à conduire leurs activités. La réponse ne réside pas seulement dans de nouvelles approches technologiques et des services informatiques, mais aussi dans un changement de comportement chez les individus et les organisations.**

#### **Le revers de la médaille**

Cette année, les constructeurs électroniques fabriqueront plus de transistors (au moins 26 milliards de millions) que les cultivateurs ne produiront de grains de riz à l'échelle mondiale. Intégrés sur les puces qui équipent les ordinateurs personnels, les téléphones portables et quantité d'autres appareils, ces "briques de base" des technologies de l'information (TI) coûteront à peu près autant qu'un caractère imprimé dans un journal.

Ces chiffres reflètent la loi de Moore, formulée pour la première fois en 1965 par Gordon Moore, le fondateur d'Intel.\*\* Selon lui, chaque nouvelle génération de puces serait deux fois plus performante que la précédente, ce qui se traduirait par une augmentation exponentielle des performances assortie d'une baisse en proportion du coût de la puissance informatique. Moore avait raison, et sa prédiction d'un doublement des performances tous les 18-24 mois se vérifie encore aujourd'hui.

À première vue, l'idée de gagner constamment en puissance avec des coûts toujours moindres semble bénéfique, et favorise le développement de "l'informatique diffuse" (pervasive computing) pour d'innombrables applications très souhaitables. Mais il y a un inconvénient, et cet inconvénient devient très difficile à ignorer. Car les systèmes informatiques ne se contentent pas de traiter les données : ils les stockent, et en quantités toujours plus considérables.

On estime que dans seulement quatre ans, la masse d'informations dans le monde doublera toutes les 11 heures. La croissance des données numériques est si rapide que le vocabulaire employé pour décrire les volumes a dû évoluer au même rythme. Il y a dix ans, les utilisateurs et les responsables informatiques professionnels s'exprimaient en kilooctets et en mégaoctets. Aujourd'hui, les écoliers ont accès à des ordinateurs portables d'une capacité de stockage de plusieurs dizaines de gigaoctets, tandis que les responsables réseau doivent penser en téraoctets (1000 gigaoctets) et en pétaoctets (1000 téraoctets). Viennent ensuite l'exaoctet, le zettaoctet et le yottaoctet - chacun étant mille fois plus grand que le précédent.

---

**En bref**

---

Certains observateurs ont pu comparer la situation actuelle à la révolution industrielle, qui a marqué le passage de la production artisanale au travail à la chaîne, avec l'obtention de gains de productivité massifs. À la différence qu'aujourd'hui la production de masse ne concerne pas des casseroles ou des voitures, mais des bits d'information, par milliards et milliers de milliards.

La tendance est à une hausse constante. Les appareils équipés de processeurs touchent un marché de masse qui se développe à un rythme accéléré. Les gens sont de plus en plus habitués à capturer et à stocker des images fixes et animées, de la musique et d'autres contenus de divertissement. L'existence même des entreprises, des administrations et des autres organisations repose sur les réseaux et sur les bases de données, au point que la préservation du stock d'informations peut être plus importante que la perte temporaire d'un siège commercial ou d'un site de production.

Le phénomène commence au niveau le plus élémentaire, avec les systèmes de messagerie des entreprises, qui génèrent de grosses quantités de données. Le courrier électronique professionnel (qui ne véhicule pas que des messages essentiels) devrait connaître une croissance annuelle de 25 % à 30 %. Et, quel que soit l'intérêt d'un e-mail, son poids sur le système est amplifié par des pratiques telles que l'adressage multiple et l'adjonction de pièces jointes volumineuses (fichiers texte, audio, voire vidéo).

Aucune organisation n'est immunisée. L'explosion des données touche tous les secteurs, de l'industrie aux services financiers.

L'aérospatiale, en particulier, connaît une transformation des méthodes de conception et de construction des avions et de leurs principaux sous-systèmes. Alors qu'auparavant un grand projet aéronautique reposait sur des milliers de plans papier créés par de talentueux dessinateurs industriels, les informations relatives à un nouveau projet sont 100 % électroniques et générées par de puissants logiciels de conception et de fabrication assistées par ordinateur (CFAO). Cette approche représente un progrès majeur, car elle accélère considérablement le développement des prototypes et contribue à l'élimination précoce des erreurs, en amont du montage final, avant qu'elles ne deviennent très coûteuses. Mais elle produit en même temps des montagnes de données.

Une autre technique, l'analyse numérique en mécanique des fluides (CFD - Computational fluid dynamics), s'est aujourd'hui banalisée dans le secteur de l'aérospatiale. Elle consiste à simuler des choses comme l'écoulement de l'air autour de l'appareil et dans les moteurs à des vitesses supersoniques à l'aide d'ordinateurs très puissants. Si la CFD évite de nombreux tests coûteux et risqués, elle engendre quantités de fichiers qui se chiffrent en téraoctets.

***L'explosion des données touche tous les domaines de l'activité commerciale.***

---

**En bref**

---

L'analyse quantitative - la modélisation mathématique du comportement futur d'un titre déterminé, d'une opération complexe ou d'un marché entier - est typique de la problématique "données" du secteur des services financiers. L'historique des cours est indispensable à ce type d'analyse, et il n'est pas rare d'utiliser à cet effet les 20 dernières années. À l'origine, les analystes examinaient les données journalières (cours d'ouverture et de clôture, et volumes journaliers), qui se chiffraient en gigaoctets. Aujourd'hui, ils doivent étudier les cours et les volumes de chaque transaction d'un titre particulier sur plusieurs années, avec des jeux de données qui atteignent plusieurs téraoctets.

Les programmes de santé de certains pays visant à numériser des populations entières vont générer des masses d'informations encore plus considérables. Les services de santé publique ont entrepris de numériser les dossiers des patients, y compris les résultats d'examens de diagnostic tels que radiographies et IRM (DMP - Le dossier médical personnel).

***La génération de données augmente à un rythme exponentiel.***

Entre l'adolescent muni de son lecteur MP3 et de son portable-appareil photo, le cadre d'entreprise qui envoie des mémos par e-mail, le médecin qui télécharge les clichés comparatifs d'une tumeur et le policier qui contrôle une plaque minéralogique, les citoyens des sociétés industrialisées utilisent et créent des données à un rythme qui s'accélère de façon exponentielle.

La plupart de ces informations sont stockées - sur les ordinateurs portables et les téléphones intelligents des particuliers, sur les serveurs des entreprises, sur des sites d'archivage et dans des entrepôts de données. Et il est toujours facile, jusqu'à présent, de trouver l'espace physique nécessaire : la technologie du stockage progresse aussi vite que celle du traitement, et les prix chutent au moins aussi rapidement.

Mais la disponibilité immédiate d'un espace pour entreposer les données en fin de vie utile masque un problème qui pourrait, dans la décennie à venir, entamer sérieusement la rentabilité des entreprises et l'efficacité des services de santé, des forces de police et de sécurité, des administrations locales et nationales, et de nombreux autres types d'organisations.

---

**En bref**

---

**Une prolifération devenue néfaste**

Savoir, c'est pouvoir - mais seulement si l'on peut extraire l'information rapidement et efficacement d'une masse de données toujours plus importante. Les entreprises sont confrontées à des stocks d'informations qui grossissent au point de devenir ingérables et nocifs pour la bonne marche de l'activité parce qu'ils grèvent les performances et la rentabilité.

Devant l'amoncellement des données, le réflexe est de recourir à des supports de stockage moins coûteux, en grandes quantités. Mais le fait de tout "déverser" dans un réservoir apparemment illimité présente trois inconvénients majeurs :

1. *La récupération rapide des informations est de plus en plus difficile.*
2. *Il faut toujours plus de personnel pour gérer des "dépôts" de données de plus en plus anarchiques.*
3. *Les réseaux et les performances des applications sont ralentis par le surcroît de trafic généré par des utilisateurs qui peinent à trouver les informations dont ils ont besoin.*

***La solution classique de stockage des données n'est pas aussi économique qu'elle en a l'air.***

Face à ces inconvénients de la culture du «tout garder», les organisations commencent à s'interroger sur le coût réel d'un traitement du problème à coup de disques durs... et se rendent compte que cette approche n'est pas aussi économique qu'elle n'y paraît. La facture d'énergie n'a plus rien de négligeable, et l'obligation probable de contrôler les émissions de CO<sup>2</sup> pourrait se traduire par une toute nouvelle source de coûts à l'avenir.

Il y a enfin ceux qui pensent que l'accumulation des données s'emballé tellement qu'elle finira par dépasser la technologie du stockage telle que nous la concevons aujourd'hui et qu'il n'y aura jamais assez de disques pour y répondre.

Nous n'en sommes pas encore là. Mais les entreprises commencent à s'apercevoir que les problèmes de la récupération de l'information ne nuisent pas seulement à l'efficacité ; ils représentent aussi un coût.

---

**En bref**

---

L'e-mail est apparu comme l'une des premières sources du problème. Considéré au départ comme rien de plus qu'un outil de communication souple et rapide, le courrier électronique capte aujourd'hui, selon les estimations, jusqu'à 75 % du patrimoine intellectuel de l'entreprise. Des documents e-mail interviennent dans les trois quarts des litiges entre sociétés. L'utilisation intensive de la messagerie en a fait dans nombre d'entreprises le référentiel de documents numéro un - ce que la législation entérine en exigeant la conservation à long terme des messages.

***Le stockage et la récupération du courrier électronique ne sont qu'une part des problèmes auxquels sont confrontées les entreprises.***

Les entreprises découvrent à leurs dépens la nécessité de prendre très au sérieux le stockage du courrier électronique. Cinq banques américaines ont été récemment condamnées à une amende de 1,25 million de dollars chacune pour avoir été dans l'incapacité de retrouver les e-mails qui leur étaient demandés. Une entreprise de Fortune 500 a dû dépenser 750 000 \$ pour extraire des e-mails de documents d'archive dans le cadre d'une assignation. Et un laboratoire pharmaceutique a été contraint de mobiliser d'importantes ressources pour effectuer des recherches dans 30 millions de messages pour les besoins d'une action en justice.

L'importance que les législateurs accordent à la conservation de l'information a tout l'air de devoir perdurer. Si l'on ajoute la mise en œuvre de services de communication mobile haut débit pour les consommateurs, la disponibilité d'un accès Internet sans fil à l'échelle nationale et une connectivité réseau sans fil à très haut débit dans les foyers et les bureaux, cette obligation réglementaire ne fera qu'accentuer le besoin inflationniste de données que nous connaissons aujourd'hui.

Les entreprises et les organisations sont confrontées à un choix de plus en plus pressant quant à la réponse qu'il convient d'apporter à ce défi redoutable. Elles peuvent continuer à stocker massivement leurs données dans des "décharges" toujours plus vastes et incohérentes, et payer le prix fort chaque fois qu'il leur faut récupérer des contenus de valeur. Ou elles peuvent s'attaquer au problème et déterminer comment gérer activement les informations tout au long de leur cycle de vie, en éliminant très tôt la masse des données éphémères pour ne conserver que les contenus susceptibles de receler une valeur à long terme.

---

**En bref**

---

***La gestion du cycle de vie des informations (ILM) offre une solution au problème du stockage massif des données.***

**Les données : on les aime trop !**

De même que le monde entier reconnaît la nécessité de mieux gérer la planète et ses ressources, le secteur informatique a reconnu le caractère “non durable” de l’approche actuelle en matière de création et de stockage des données. Sa réponse, qu’il considère seulement comme une partie de la solution, est la gestion du cycle de vie des informations (ILM - Information lifecycle management).

Les principes de l’ILM ont été définis par la Storage Networking Industry Association (SNIA), qui compte parmi ses membres IBM et d’autres grands acteurs du marché informatique. Il s’agit d’un processus qui consiste à gérer l’information de bout en bout - de sa conception à sa mise au rebut - selon sa valeur intrinsèque pour l’entreprise et d’une façon qui assure une utilisation maximale des ressources de stockage tout en réduisant les coûts de récupération.

En d’autres termes, l’ILM est une déclaration de guerre au stockage massif et indiscipliné des données. Cette approche est conçue pour éliminer le plus tôt possible les informations de faible valeur et conserver les autres sur des unités de stockage à long terme gérées de façon active qui assurent un accès rapide et économique. Une solution ILM est constituée de matériel et de logiciel, mais le meilleur point de départ (même s’il y en a d’autres) consiste à développer les premiers filtres, les pratiques et les règles qui vont déterminer la valeur économique, l’origine et le destin des différents types de données véhiculées par le réseau de l’entreprise.

Certains prestataires sont sans doute capables d’assister l’entreprise dans le développement des règles à conduire, mais le meilleur juge pour déterminer quelles données stocker et pendant combien de temps, ce doit être l’entreprise elle-même. A priori, le candidat idéal pour conduire cette mission est le directeur informatique, car c’est bien lui, après tout, qui se débat avec ce problème depuis des années.

---

**En bref**

---

La personne chargée de mettre en place les règles de conservation des données doit parfaitement connaître l'entreprise et avoir une bonne idée de ses orientations stratégiques. L'établissement de ces règles relève moins de considérations techniques que de la nature même de l'entreprise, ses priorités aujourd'hui et celles à venir. Les responsables informatiques qui souhaitent recentrer leur rôle sur une problématique opérationnelle et qui reçoivent un soutien adéquat de leur direction auront plaisir à relever le défi.

Parfois, les hauts dirigeants d'entreprise n'ont pas perçu la menace que représente le stockage massif des données. Et s'ils en ont conscience, cette menace est toujours assimilée à un problème informatique, relevant du DSI, plutôt qu'une question requérant des décisions stratégiques au plus haut niveau. Même tendance chez les employés : ce qui se passe sur le réseau est l'affaire des informaticiens, pas la leur. Et dans les entreprises où chacun possède un ordinateur portable, les employés sont bien trop absorbés par leur travail pour appliquer les règles rigoureuses d'une gestion responsable de l'information.

***Les responsables informatiques, avec le soutien de la direction, peuvent jouer un rôle crucial dans la mise en place des règles qui permettront de résoudre le problème.***

L'approche ILM fournit au département informatique des outils de base pour résoudre le problème. Selon le fournisseur, cette boîte à outils contiendra non seulement du matériel et du logiciel, mais aussi une assistance au développement de nouvelles pratiques et de nouvelles règles, des exemples de bonnes pratiques, et un support décisionnel destiné aux managers qui doivent faire le tri entre les informations à conserver et les informations à éliminer. Surtout, l'ILM rendra le processus si facile pour le personnel et pour l'entreprise que personne n'hésitera à le mettre en œuvre.



---

**En bref**

---

***Une évolution des comportements est nécessaire.***

En fin de compte, les individus et les organisations ne changent leur comportement que si cela va clairement dans leur propre intérêt. Ces dernières années, les écologistes ont réussi à sensibiliser la population aux multiples dangers d'une consommation effrénée des ressources naturelles. S'il reste encore du chemin à faire dans certains pays, ils ont cependant tiré la sonnette d'alarme sur le changement climatique, banalisé le recyclage des déchets dans de nombreuses sociétés et impulsé un nouvel élan à l'industrie des biocarburants. Qui fait - ou devrait faire - campagne en faveur d'une approche "plus verte" de la génération et du stockage des données ?

La SNIA et ses membres sont en première ligne pour promouvoir la gestion du cycle de vie des informations (ILM), ainsi que les produits et les services fondés sur cette discipline. Certains experts réputés ont sans aucun doute un rôle à jouer pour convaincre les esprits.

**IBM et ILM : la solution intégrale**

De par l'étendue de ses offres - matériel, logiciels, services et conseil -, IBM est mieux placé que quiconque pour aider ses clients à mettre en œuvre une solution intégrale de gestion du cycle de vie des informations. D'autres approches ILM, en revanche, continuent à privilégier le stockage des données par rapport aux pratiques de travail, à l'établissement de règles, aux architectures et au support à long terme qui forment les ingrédients indispensables d'une solution complète.

***IBM aide les entreprises à mettre en œuvre des stratégies ILM efficaces.***

En tant que fournisseur de services d'externalisation informatique, IBM Global Technology Services développe actuellement des stratégies ILM complètes pour plusieurs grandes organisations dans le monde - et a amorcé le déploiement d'une solution ILM globale et intégrée répondant à ses propres besoins. IBM Global Technology Services met à profit les connaissances acquises pour aider les grandes et moyennes entreprises à définir et à mettre en œuvre leurs stratégies - en commençant souvent avec des projets modestes offrant des retours optimaux.

---

## En bref

---

### *Quelques succès récents d'IBM.*

- Un hôpital américain a vu ses coûts de stockage divisés par deux tandis que sa capacité de stockage augmentait de 500 % après l'installation d'un système disques IBM.
- Une compagnie d'énergie de Fortune 500 a réduit de moitié son coût de revient grâce à l'accroissement du taux d'utilisation consécutif à la mise en œuvre d'une solution de stockage multiniveau IBM.
- Aux États-Unis, un service de police a réduit de plus de 80 % le coût par suspect identifié après avoir déployé une application d'identification IBM à l'échelle nationale.
- Une banque régionale américaine avait des problèmes avec son système de stockage des données : son stock doublait tous les 12-18 mois et le taux d'utilisation des disques était catastrophique (28 %). L'adjonction d'un logiciel IBM de virtualisation du stockage (SVC) a porté le taux d'utilisation à 80 % et considérablement réduit les coûts.
- Un grand assureur du secteur de la santé a fait appel à la technologie IBM pour consolider son stockage, multipliant par trois les performances et raccourcissant de 50 % la durée des sauvegardes.
- Une multinationale du secteur de l'alimentation a stocké 15 mois de données (l'équivalent de trois millions de documents) dans une solution d'archivage IBM. Les processus de gestion ont été accélérés de 20-25 % - avec des économies immédiates de 70 000 \$, et d'autres à venir.

### **Un changement culturel nécessaire**

Il est tentant de voir l'accumulation des données comme un désastre programmé, un changement d'état subit et violent débouchant sur une défaillance des systèmes et un arrêt de l'activité. Un tel sort peut guetter des entreprises qui n'auraient pas pris la réelle mesure du problème. Mais l'issue la plus probable, dans le cas général, est une perte de vitesse constante et inéluctable.

Le boulet qui leste les entreprises et les autres organisations dans le monde finira bientôt par pénaliser la rentabilité ou l'activité, à tel point qu'on ne saurait ignorer le problème plus longtemps.

---

**En bref**

---

***Les véritables solutions ILM permettent de résoudre le problème tant au niveau humain qu'au niveau de l'entreprise.***

Le département informatique pense détenir une partie de la solution avec la gestion du cycle de vie des informations (ILM) - un ensemble de principes conçus pour assurer une gestion active, à la baisse, du volume des informations transférées sur des ressources de stockage, et pour veiller à ce que ces informations puissent être récupérées rapidement et de façon économique, à court et à long terme. L'ILM relève du simple bon sens, mais il reste beaucoup à faire avant qu'elle ne soit réellement mise en œuvre.

En premier lieu, les fournisseurs informatiques doivent mettre de l'ordre dans leurs offres. Trop de fournisseurs, très loquaces sur la notion de cycle de vie de l'information qui sous-tend l'ILM, continuent à ne rien proposer d'autre que du stockage et encore du stockage. Les véritables solutions ILM permettent de résoudre le problème tant au niveau humain qu'au niveau de l'entreprise, tout en fournissant les outils logiciels et matériels.

Ensuite, il faut redoubler d'efforts pour sensibiliser tous les acteurs de l'entreprise. La direction n'a pas encore pris conscience du problème et des menaces qui pèsent sur la rentabilité. Soit les responsables informatiques ne sont pas assez proches de leur direction, ou bien ils sont trop occupés pour accorder à la stratégie de stockage toute l'attention qu'elle mérite. Quant aux employés, ils considèrent le stockage comme une ressource illimitée, et son accès est mis au rang d'un droit fondamental.

Soixante ans après l'avènement de l'ère de l'information, et tout juste un quart de siècle après l'entrée massive des ordinateurs dans notre vie professionnelle et privée, on observe un fait déconcertant. Nous avons l'habitude de considérer l'information comme une source de pouvoir, comme un atout compétitif essentiel. Mais elle pourrait aujourd'hui se transformer en faiblesse, s'accumulant à un tel rythme qu'elle risque de paralyser l'activité économique.

Il est temps pour les directions informatiques et pour les entreprises de s'engager dans la mise en œuvre de l'ILM dans sa forme la plus complète.



### En savoir plus

Pour tout savoir sur les solutions d'infrastructure proposées par IBM Global Technology Services, visiter :

**ibm.com**/services/fr

Découvrez le site IBM dédié aux Directeurs informatiques :

**ibm.com**/fr

### IBM France

Tour Descartes  
La Défense 5  
2 avenue Gambetta  
92400 Courbevoie

Adresse Internet d'IBM : **ibm.com**/fr

IBM et le logo IBM sont des marques d'International Business Machines Corporation aux États-Unis et dans d'autres pays.

\*\* Intel est une marque d'Intel Corporation ou de ses filiales aux États-Unis et dans d'autres pays.

Les autres noms de société, de produit et de service peuvent appartenir à des tiers.

Le fait que des produits ou des services IBM soient mentionnés dans le présent document ne signifie pas qu'IBM ait l'intention de les commercialiser dans tous les pays où elle exerce une activité. La copie ou le téléchargement, sans l'accord écrit d'IBM, des images contenues dans ce document sont expressément interdites.

La présente publication est à usage purement informatif.

© Copyright IBM Corporation 2006  
Tous droits réservés.

### Contributeurs

*Paul Coles – Architecte en solutions de stockage, Paul Coles possède une grande expérience dans la conception et le déploiement de solutions pour le compte de nombreux clients en Europe.*

*Tony Cox – Tony Cox a acquis une grande expérience des services d'infrastructure dans le domaine de la gestion des systèmes et du stockage en tant qu'expert commercial et consultant. Il aide les clients à résoudre leurs problèmes de gestion de l'information et du stockage.*

*Chris Mackey – Consultant en infrastructure, Chris Mackey a travaillé dans de nombreux secteurs d'activité dans les domaines du delivery, de l'infrastructure et des processus.*

*Simon Richardson – Simon Richardson est Service Delivery Manager chez IBM. Il possède une vaste expérience notamment en tant que consultant en infrastructures de stockage.*