



"La protection de contenus numériques pendant de très longues périodes", par Joe Jurneke : cette étude décrit les différents aspects des opérations d'archivage et les risques encourus par la protection de ces données. L'archivage basé sur les technologies actuelles, sur de très longues périodes sera abordé tout au long de cette étude.

Avec la migration massive d'informations du domaine analogique vers les mécanismes, toujours en pleine évolution, des domaines numériques, une question importante reste posée: combien de temps les données vont survivre au temps et vont pouvoir être relues ? Les résultats de cette étude mettent en cause des moyens, des performances et les limitations connues des différentes technologies d'aujourd'hui.

> Technologies en compétition :

Aujourd'hui, il existe 4 technologies qui sont en compétition pour archiver des données.

On y trouve la bande magnétique, les disques magnétiques, les disques optiques, la Ram non volatile, solid state logic SSL (Flash ou équivalent). Deux, parmi elles, sont compétitives en terme de prix (€/Go), les disques magnétiques et la bande magnétique. Les disques optiques ont des capacités inférieures pour des coûts €/Go supérieurs avec des vitesses de transfert plus lentes que les disques et bandes magnétiques. Les mémoires SSL sont bien trop chers pour des applications d'archivage.

Nous limiterons notre discussion aux technologies sur disques et sur bandes magnétiques :

Il y a 4 applications fondamentales pour ces disques et ces bandes : le stockage "online" des données, "near-the-line", les back up, et les archives offline.

. En applications Online, les moyens de stockage (disques ou bandes magnétiques) sont opérationnels pendant de longues heures sans interruption.

. En mode Near-the-line, les accès aux données sont moins fréquents que dans le mode online, on peut utiliser ici les disques magnétiques mécaniques, des bibliothèques de bandes ou tout simplement des lots de disques durs bon marché.

. Les applications de backup sont de plus en plus fréquemment utilisées pour palier aux situations désastreuses de recouvrement des données, c'est pour cela que les solutions utilisées ici ont migré, récemment, de la bande aux disques durs.

. Les applications d'archivage en particulier, trouve des solutions où les données sont écrites sur des média dans le but de les conserver très longtemps et de les consulter plus tard, dans le futur. Les données archivées le sont souvent dans des placards ou sur des étagères, mais peuvent aussi être conservées en modes online ou near the line.

> Les applications qui en demandent le plus :

Les applications qui en demandent le plus sont celles des archivages de longues durées. Les utilisateurs souhaitent retrouver leurs archives de longues années après la création de leur archivage. La perte des données est souvent, dans ce cas, intolérable et très ennuyeuse. Les données vidéo sont des exemples types de l'archivage pendant de très longues périodes. Dans cet environnement, des copies additionnelles des données ne sont pas disponibles. Pire, la perte de données devient définitive. La solution de multiples copies pourrait amoindrir ce problème, mais dans de nombreuses circonstances, elle ne pourrait garantir une

sécurité cohérente.

La vérification de l'intégrité des données ainsi que leur qualité sont les points les plus importants dans la création d'archivage de qualité. La vérification entraîne celle de l'intégrité des données stockées sur les média après une période de stockage. Nous en parlerons plus tard, mais elle est plus importante lorsque l'archivage est effectué sur un disque que sur des bandes magnétiques. Les personnels en charge de l'archivage sont plus sensibilisés à la qualité de l'enregistrement de l'archive.

Les deux technologies que nous examinons ici, diffèrent totalement dans leurs modes de vérification :

. Dans le cas des bandes magnétiques la vérification se fait au moment de l'écriture (read-verified-write) (lecture-vérification-écriture) sans impact sur les taux de transfert.

. Dans le cas des disques, une lecture après l'écriture est nécessaire afin de vérifier l'intégrité des données, ce qui réduit les taux de transfert par un facteur minimum de deux.

> Problèmes mécaniques :

En terme de durabilité, de nombreux facteurs jouent sur la stabilité magnétiques des enregistrements. Le stockage et l'environnement de travail jouent une rôle critique dans la stabilité des données enregistrées. Les températures et les taux d'humidité devront être respectés selon les recommandations des données constructeurs d'utilisation des media magnétiques. Ces recommandations s'appliquent aussi bien aux disques durs qu'aux bandes magnétiques. Le NML, pour National Media Lab en 1995 a publié un rapport sur les contenus enregistrés sur des média à technologies magnétiques. Le NML recommande une température constante de 20° C à une humidité relative de 40%. Depuis cette époque, la densité aériale (nombre de données stockées par pouce² des surfaces magnétiques) a largement augmenté. Les considérations environnementales aujourd'hui sont encore plus importantes qu'il y a quinze ans.²

De fait, AMIA, (Association of Moving Image Archivists) recommande en 2007 un environnement maintenu à 7° C à un taux d'humidité de 25%.² Le transport et l'utilisation

mécaniques peuvent jouer un rôle crucial dans la perte de données. La chute d'un disque au sol réduit la probabilité de retrouver ses données, voire à tout perdre. Basé sur la publication des coûts de récupération de données, le prix pour récupérer les contenus d'un disque est supérieur à 3000€. La réussite n'est pas garantie.

Échapper une cartouche sur le sol entraînera peut-être une casse physique de la boîte extérieure. Avec la bande, dans les cas extrêmes de chute, la plus grande partie des données pourra être récupérée, voire l'intégralité. La majorité des fabricants de cartouches à bande magnétique annonce que leurs cartouches ont passé des tests rigoureux en ce qui concerne l'évaluation de la solidité à la chute.

Les deux technologies sont relativement sensibles aux chocs sévères, mais les bandes ont démontrés une robustesse supérieure à cet égard.

L'organisation des Error Correction Code (ECC) entre disque et bande est relativement différente. Les disques durs organisent les ECC en les mixant avec les données qu'ils protègent dans les mêmes track/secteur. Avec les bandes, les ECC sont organisées de manière multi-dimensionnelle. Les syndromes ECC sont répartis sur toute la surface de la bande, en bas et au travers les multiples pistes dans le but de protéger les données des défauts et des accidents. En plus, dans de nombreux cas, les données sont réparties de manière multi-dimensionnelle pour renforcer la robustesse contre les défauts et accidents. C'est pour cela qu'il est bien plus facile de récupérer des données en erreur lorsqu'elles sont stockées sur des bandes que celles qui le sont, sur des disques.

En considérant que les données environnementales et les risques de chocs mécaniques sont sous contrôle, quels sont les autres points qui restent à prendre en compte ?

Les disques et les lecteurs enregistreurs de cartouches possèdent des mécanismes complexes. Les deux sont des outils sophistiqués contenant des technologies nombreuses. Lorsqu'on enregistre des archives sur disques durs, l'intégralité du disque est sollicité (électronique, media, les têtes, les moteurs,...) Dans un environnement d'archive sur bande, seul le volume de la bande est concerné par l'archive. Les têtes, les mécanismes, les moteurs ne le sont pas. Il y a bien moins de risque de dégradation des électroniques dans le temps. Moins de complexité, assure une meilleure tenue dans le temps.

> Taux de probabilité de pannes des disques durs :

la probabilité des pannes de disque dur s'étage en trois étapes dans le temps. Mortalité infantile (casse en début de vie), panne durant la vie disque, mort du disque. L'utilisation de disques pour l'archivage et spécialement en les utilisant, comme les K7, stockés sur des étagères, entraîne des risques de probabilité de panne de mort infantile associé à des problèmes, souvent très graves, pour les archives.

En considérant des études publiées par Google 17, CMU 6 et les autres 5 & 7, les disques souffrent de panne de début de vie significatives. Avec la croissance des capacités des disques durs, les taux de pannes tendent à augmenter proportionnellement à ces évolutions de volume. Le contenu de ces études sont basées sur une très large population qui excède les 100 000 unités. Dans le cas des applications d'archivage, les pannes de début de vie des disques durs sont les plus ennuyeuses. La plus part des archivages sur HDD sont réalisées sur des disques durs neufs qui sont defacto exposé de plein droit à ce taux de casse de jeunesse.

Les fabricants de disques durs spécifient des mesure de MTBF (Mean time Between failure) à la mesure de leur solidité. Les résultats des tests effectués ne prennent pas en compte les défauts comme « No defect found » ou encore comme « Cannot Replicate problem » sur les estimations de taux de panne. Les écrits suggèrent que cela résulterait à des estimations plus qu'optimistes de 30 à 40 % (6&18). Le MTBF ne peut être une donnée métrique valable dans le cas de l'archivage dans la détermination des performances.

Du point de vue de l'utilisateur, tout disque retiré de son service, pour quelque raison, constitue une panne grave. Les nombreuses heures passées à créer le volume enregistré deviennent, en cas de panne quelconque, une perte importante de temps et d'argent.

Bien que les 1 million d'heures d'utilisation du MTBF paraissent énormes, de nombreux disques rendent l'âme dans les 60 jours de leur première utilisation plutôt que dans leur période pour lesquels on les a destiné. Certains rapports conseillent d'utiliser pendant au moins 3 mois (ou 2000 heures) les disques durs avant de les mettre réellement en service. D'autres données suggèrent que les disques haut de gamme ne sont pas meilleurs que ceux destinés aux

consommateurs. Les données de CMU et de Google suggèrent que les pannes sont affectées par les longues mises en utilisation, par les activités opérationnelles, par l'âge, par l'environnement d'utilisation, et par le fournisseur/ fabricant . Le rapport du CMU dit que le risque de mortalité infantile des disques durs persiste jusqu'aux 200 heures d'utilisation. Cette mortalité prématurée des disques s'accroît pour les disques conservés sur les étagères dans le but d'archiver des contenus. Les archivistes prudents auront soin d'avoir réalisé un duplicata de leurs archives. Mais même cette approche est risquée.

Lorsqu'une société achète des équipements de sauvegarde elle le fera généralement en quantité. Les pannes de disque dur ne sont pas systématiques. Des facteurs multiples de pannes affectent aussi le vieillissement des disques. (la marque, la qualité des assemblages, les nouveautés technologiques, etc...). les nouveaux équipements faisant l'objet d'un achat groupé proviennent en général des mêmes lots de fabrication, construits avec les mêmes pièces détachées, assemblés avec les mêmes procédures.

Les taux de pannes et les casses mécaniques pour un lot spécifique sont en général similaires, ce sont les rapports qui le disent. Si un disque a une panne précoce, il est sage de penser que les autres disques achetés en même temps pourront aussi tomber en panne précocement. Les problèmes liés à l'utilisation de disques similaires en lot et en date d'achat impliquent la stratégie de multiplier le nombre de copies de sauvegarde, ce qui n'est certainement pas une solution appropriée. Si un disque tombe en panne rapidement, il y a une possibilité distincte que les autres disques d'un même âge et du même fabricant tombent aussi en panne rapidement ce qui remettrait en question l'intégrité des archives concernées.

> Stabilité thermique des particules magnétiques :

Les particules magnétiques sur disque ou sur bande sont sujettes à dégradation dans le temps. A quelle vitesse agit cette dégradation c'est ce qui concerne les archivistes. La température a un impact important sur la stabilité des particules magnétiques. Des températures trop élevées réduisent la durée de vie prévue (voit les rapports NML (1) et AMIA (2))

Les média magnétiques sont opposés à deux forces antagonistes qui s'affrontent. La force de

blocage qui établit et essaie de maintenir l'état de la magnétisation, elle est liée aux propriétés magnétiques des média et par le volume de particules magnétiques utilisées (Le grain). La force de destabilisation qui est liée à la température. Lorsque la force de blocage et celle de destabilisation sont égales, les particules peuvent changer d'état magnétique. La magnétisation des particules s'amointrit d'une manière exponentielle avec le temps. Elle s'amointrit progressivement de la magnétisation totale vers le niveau zéro.

Pour quantifier ces phénomènes, une formule de mérite a été établie pour en mesurer la stabilité thermique. La formule est KuV/kT . Ku étant la constante de l'isotropie de la particule. V étant le volume de la particule et k est la constante de Boltzman et T la température en degré Kelvin. Lorsque le ratio est supérieur à 65, la stabilité thermique peut faire espérer une durée de vie de 30 années. Les résultats de cette formule sont exponentiels. Une durée de vie de cinq années aura un KuV/kT d'environ 45.

Les chiffres montrent qu'un disque dur magnétisé (média perpendiculaires ou en pattern) ont un coefficient KuV/kT compris entre 45 et 60. (12, 13, 14) Les bandes magnétiques ont un ratio largement supérieur à 100 pour des particules MP conventionnelles. Cela prouve que la bande destinée à l'archivage possède une stabilité thermique bien plus stable.

La conservation longue des bandes :

Il y a encore des clients qui utilisent des bandes ouvertes à 7 ou 9 pistes NRZI ou même des bandes digitales 1600bpi PE/6250 GCR qui ont été enregistrées dans les années 1950 jusque dans les années 80s. Des services de maintenance et de réparation des bandes sont disponibles encore aujourd'hui. En plus, il existe au moins une société spécialisée dans la conversions des vieilles archives stockées sur des bandes anciennes vers des bandes à cartouches modernes.

La vie des archives sur bande est largement dépendante de l'environnement de stockage. En suivant les conseils et recommandations pour l'environnement de stockage des archives il est possible d'obtenir 30 années d'utilisation grâce aux technologies modernes. Tous les appareils à bandes utilisés par les entreprises et professionnels depuis 40 ans ont servi pour l'archivage des données. Les contenus archivés sur des cartouches au format LTO à bandes avec des

formulations en particules métalliques sont donnés par les fabricants FujiFilm, Maxell, Sony & TDK pour être utilisables pendant une durée de vie de trente années.

Des annonces récentes de Fujifilm, IBM et Oracle ont démontré la faisabilité d'au moins trois futures générations pour des bandes à base de ferrite de Barium destinées au stockage des données. Tout au long de ces démonstrations, il a été confirmé que la durée de vie de la ferrite de barium était de 30 ans. Ces annonces ont été avalisées par les fabricants de lecteurs enregistreurs de bandes LTO (8).

Les bandes jouent un rôle crucial dans tous les aspects de l'intégrité des données. Les technologies LTO ont été élaborées par des entreprises de renom comme HP, IBM et Quantum. Les LTO occupent 90% du marché de la bande à moyenne contenance. Depuis la sortie des cartouches LTO, il y a 10 ans, 3,5 millions de lecteurs enregistreurs et plus de 150 millions de cartouches ont été vendus. (10, 11, 19, 22, 23)

Une vraie technologie d'archivage pour les données digitales (pour plus de 100 ans) n'existe pas encore. Le problème, aujourd'hui, est de choisir un média qui aie la meilleure protection possible, pour la période la plus longue.

Les disques durs n'ont jamais été créés pour offrir un moyen d'archivage offline. Ils ont été fabriqués pour stocker des données online. Et comme le recommande une étude (4) il faudrait mettre en chauffe pendant 1000 heures (ou trois mois) les disques durs ce qui reviendrait trop cher et serait une perte de temps.

Par contre les bandes magnétiques ont toujours été utilisées pour des backup et des sauvegardes dans le temps. Ce média des archives n'est pas sensible aux pannes précoces ni aux pannes de mécanismes. En plus sa stabilité thermique est bien supérieure, ce qui signifie que les périodes d'utilisation des bandes stockées sur étagères est largement supérieures à celles des disques durs.

Si l'on compare la compatibilité intrinsèque d'archivage sur bandes à celle sur disques, la

simplicité gagnera toujours. La bande sera du point de vue de l'utilisateur plus facile à utiliser.

Il y a de nombreuses technologies de bandes magnétiques différentes sur le marché. Certaines sont supérieures aux autres. Il faut choisir celles les plus répandues. Les formats qui ont la plus grande percée sur le marché. Il faut abandonner celles dont les ventes chutent.

La technologie LTO avec 90% du marché des bandes à moyenne contenance et avec l'association de fabricants majeurs comme HP, IBM et Quantum, occupe la majorité de l'archivage.

La chute d'une cartouche est habituellement moins grave que la chute d'un disque dur. Les bandes sont plus facilement réparables. Pendant les réparations, les données ne sont pas exposées à des pertes comme elles le sont avec des disques durs, spécialement lors de pannes des têtes ou de l'assemblage des disques.

Un certain nombre de problèmes ont été exposés dans cette étude sur les solutions d'archivage de données. La panne précoce est un problème souvent négligé par ceux qui archivent sur des disques durs. Elle peut significativement contribuer à la perte de données contenus sur ces disques rangés en conteneurs d'archives sur des étagères. De plus, le coefficient thermique inférieur des disques durs face à celui des bandes peut engendrer des pertes de données dans le temps.

En conclusion, les bandes magnétiques représentent le meilleur moyen d'archivage offline de données gardées sur étagères.