



Fred Moore
Président
Horison.com

*Une nouvelle classe de services et de solutions de
stockage objets avancés pour les données froides*

ACTIVESCALE COLD STORAGE
CONÇU POUR LES ARCHIVES A LONG TERME
D'ENTREPRISE, WEBSCALE et HYPERSCALE

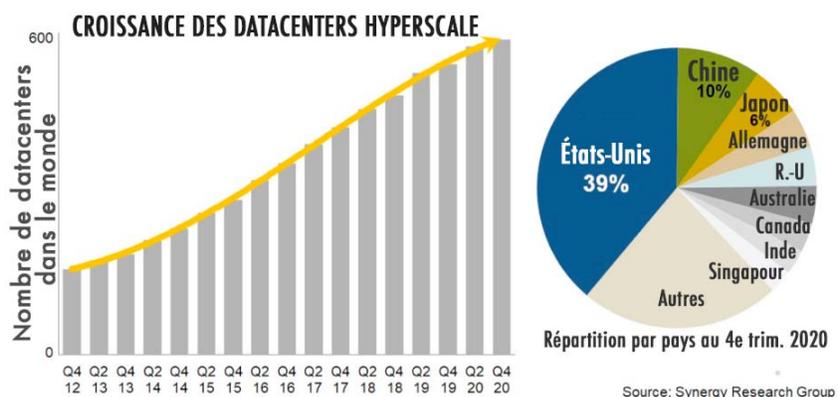
Introduction

Les données numériques occupent aujourd’hui une place prépondérante dans tous les aspects de nos vies et devraient jouer un rôle encore plus important à l’avenir. Leur croissance exponentielle est donc inexorable. Ces trois prochaines années, le volume total de données produit dépassera celui généré au cours des 30 dernières décennies. La plupart de ces données s’accumulent sous forme d’archives et de données froides en attendant que leur valeur potentielle soit libérée. Cette rapide transformation s’est accélérée avec la pandémie de Covid-19 et l’essor du télétravail qui a engendré une forte dépendance vis-à-vis des applications Internet et des communications vidéo dans des environnements de travail entièrement numériques. Près de 80 % des données sont archivées ou froides. Elles sont en majorité stockées dans un format de fichier traditionnel, limité en termes d’extensibilité, ainsi que dans le format objet relativement nouveau et en rapide expansion, qui garantit une évolutivité illimitée. Cette croissance phénoménale est due à l’usage d’applications d’ancienne génération, mais aussi de nouvelles applications basées sur Internet dans des domaines tels que la surveillance, l’Edge et le Cloud Computing, le calcul haute performance (HPC), l’Internet des objets (IoT), les médias sociaux ou encore le divertissement.

Bien plus qu’une mégatendance, le stockage des données peut-être considéré comme une mesure des progrès de l’humanité à tous les niveaux de la société, dans les entreprises publiques et privées, et dans chaque aspect de l’économie. Les besoins croissants des grandes entreprises, des infrastructures webscale et des datacenters hyperscale (HSDC) en termes de capacité hautement évolutive et à moindre coût appelle la création de nouvelles technologies et architectures de stockage. Les archives massives d’aujourd’hui nécessiteront par conséquent une solution beaucoup plus intelligente, capable d’exploiter les fonctions avancées des logiciels de gestion intelligente des données et des équipements matériels « scale-out » haute disponibilité. Le recours à une stratégie d’archivage numérique performante est devenu indispensable à la croissance continue des entreprises, des infrastructures webscale et des HSDC. Heureusement, une *nouvelle classe de stockage objets pour les données froides* a vu le jour.

Quels sont les défis du stockage à froid pour les grandes entreprises et les HSDC ?

Sans surprise, les difficultés rencontrées par la plupart des HSDC découlent en général de leur taille. La croissance inévitable des fermes de disques qui grèvent les budgets, surchargent les datacenters, créent des problèmes énergétiques et augmentent l’empreinte carbone oblige les HSDC et les grandes entreprises à migrer leurs archives et données froides vers des solutions de stockage sur bande plus économiques. Les services d’archivage dans le Cloud sont relativement peu coûteux, mais les frais liés à la restitution et au transfert (bande passante) des données Cloud s’additionnent très vite à mesure que les quantités de données transférées augmentent. Les HSDC, les grandes entreprises et



les fournisseurs de Cloud sont devenus l'épicentre des stratégies d'archivage modernes. Amazon Glacier, Amazon Glacier Deep Archive et Microsoft Azure sont des exemples de services Cloud hyperscale adaptés à l'archivage des données froides. La dernière étude de [Synergy Research Group](#) dénombre près de 600 datacenters hyperscale dans le monde, soit deux fois plus qu'il y a cinq ans. Quatre des plus grands HSDC hébergent 60 datacenters ou plus à l'échelle mondiale. Le rapport Cisco Global Cloud Index indique par ailleurs que les HSDC devraient prendre en charge 65 % de l'ensemble des données stockées dans les datacenters en 2021. Selon différentes estimations du secteur, la capacité de stockage déployée s'élèvera à un peu plus de 8 Zo d'ici fin 2021. Dans ces conditions, les datacenters d'entreprise et hyperscale seront bientôt confrontés à des défis majeurs en matière de stockage à froid.

À quel moment les données deviennent-elles froides ?

Les données archivées et froides sont principalement non structurées. Il s'agit de la catégorie de données la plus importante et celle qui connaît la croissance la plus rapide. La plupart des types de données sont archivés 90 à 120 jours après leur date de création. Plus les données restent statiques et intactes, plus elles se refroidissent. La rétention de données pendant 100 ans, voire plus, est devenue monnaie courante. Certaines sont archivées dès leur création, puis attendent des années avant d'être consultées ou analysées. Elles viennent ainsi grossir les rangs des données froides, qui sont en grande partie stockées sur des disques durs. Conserver des données inutilisées sur des disques durs rotatifs durant de longues périodes est une stratégie de stockage à la fois coûteuse et énergivore. La solution de stockage la plus économique pour l'archivage des données consiste à utiliser des bibliothèques robotisées de grande capacité, facilement extensibles à un ou plusieurs exaoctets, aussi bien en local que dans un Cloud hybride ou sur des sites distants. De récentes études sur la consommation d'énergie montrent que le stockage sur bande produit jusqu'à 87 % moins de carbone que des systèmes sur disque dur de même capacité.

Les grandes entreprises et les datacenters hyperscale misent sur le stockage objets pour les données froides

Avec la croissance des données, les formats de stockage traditionnels sont devenus coûteux et difficiles à gérer. Rendu populaire par le service Amazon S3, le stockage objets s'est imposé comme le format standard de facto pour les services d'archivage basés sur le Cloud. Le stockage objets est né du besoin de stocker et de restituer une quantité quelconque de données depuis n'importe quel lieu via Internet. C'est la solution idéale pour les fournisseurs de Cloud et de nombreux HSDC. Avec le stockage objets, les blocs de données qui composent un « objet », ainsi que ses métadonnées, sont regroupés. Le stockage traditionnel en mode fichier peut héberger des millions de fichiers, mais finit par plafonner.

Seule architecture du secteur capable d'évoluer à l'infini, le stockage objets peut rapidement passer à des centaines de pétaoctets dans un seul namespace. Le taux de croissance du stockage objets a dépassé celui du stockage conventionnel en modes fichier et bloc. Il enregistre une hausse annuelle de plus de 35 % et cette tendance devrait se poursuivre.

disque. Dans le cas de systèmes soumis à des charges élevées de traitement parallèle, cette approche permet de stocker davantage d'objets simultanément et de réduire le temps de latence global lors du processus d'écriture.

Les nouveaux services de stockage objets de Quantum apportent l'expérience du Cloud dans les environnements sur site

Outre des équipements matériels déployables sur site, Quantum a également introduit une nouvelle gamme innovante de services de stockage objets à l'échelle du pétaoctet pour les archives actives et les données froides. Reposant sur la gamme de solutions ActiveScale, ces nouveaux services sont spécialement conçus pour les grandes entreprises, les agences gouvernementales, les fournisseurs de services Cloud et les organismes de recherche. Avec deux classes de services et aucun frais d'accès aux données, ces services entièrement gérés doivent être sérieusement envisagés par les entreprises à la recherche d'une solution sur site très économique pour l'archivage actif et le stockage des données froides.

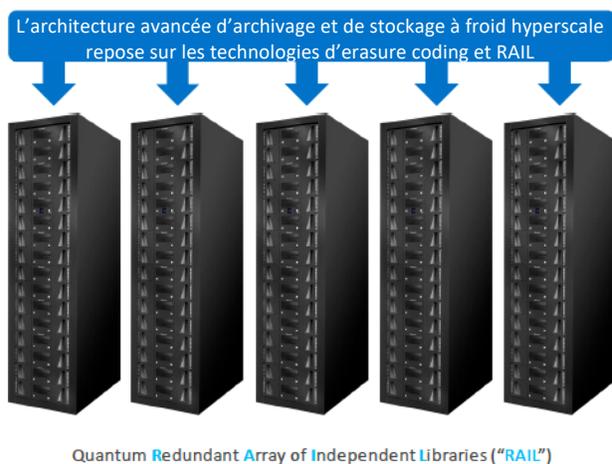
Services de stockage objets	Pour les données actives	Pour les données froides
Délai de restitution des données	Inférieur à la milliseconde	Généralement inférieur à 5 minutes
Mode d'accès aux données	Classe S3 Standard	Classe S3 Glacier
Durabilité	Jusqu'à 99,9999999999999999 %	Jusqu'à 99,9999999999999999 %
Engagements minimaux	1 pétaoctet de capacité Durée de trois ans	4 pétaoctets de capacité Durée de trois ans
Prix catalogue	0,015 \$ par Go/mois Pas de frais d'accès aux données	0,003 \$ par Go/mois Pas de frais d'accès aux données

Les configurations ActiveScale GeoSpread assurent un très haut niveau de disponibilité

Les serveurs et systèmes de stockage des HSDC sont souvent hébergés dans des datacenters géographiquement dispersés. Connue sous le nom de « redondance géographique », cette approche garantit un plus haut niveau de disponibilité et assure la résilience de l'activité en cas de sinistre, comme des incendies, des inondations, des ouragans et autres événements susceptibles d'entraîner l'arrêt d'un datacenter. Composés de disques durs et/ou de bandes, les systèmes ActiveScale GeoSpread s'appuient sur l'architecture RAIL (Redundant Arrays of Independent Libraries) développée par Quantum. Ils peuvent être déployés dans trois datacenters géographiquement dispersés (3-Geo), créant ainsi des Data Lakes de stockage objets capables de tolérer des pannes de composants et de sites. Les systèmes GeoSpread utilisent la technologie [d'erasure coding sans rendement](#) pour distribuer les données sans interruption de service entre les nœuds et les zones géographiques. Il est possible d'accéder en parallèle aux données des objets stockés sur les nœuds, ce qui augmente la capacité de bande passante globale pour le transfert des données.

Comme un système 3-Geo doit permettre de restaurer les objets en cas d'indisponibilité d'un site, pas plus d'un tiers des fragments sont stockés au sein d'un même datacenter. Lorsque la fonction de répartition hiérarchique DDP est activée pour chaque niveau, le système répartit équitablement les fragments sur 18 disques dans trois datacenters. On parle de « 18/8 » pour désigner la politique de stockage avec erasure coding associée à cette configuration. Les objets sont répartis sur 18 disques, avec une sécurité des disques s'élevant à 8. Les objets peuvent être décodés à partir de n'importe quel sous-ensemble de 10 fragments codés. Il est possible de mettre en miroir les données et applications critiques dans un seul Cloud, dans un Cloud hybride ou dans les Clouds de différents fournisseurs à des fins de redondance. En cas de panne, le datacenter sur site peut basculer sur n'importe lequel des Clouds en miroir pour maintenir la disponibilité, puis opérer un retour arrière une fois l'incident résolu.

L'architecture avancée d'archivage et de stockage à froid hyperscale repose sur les technologies d'erasure coding et RAIL



Tout comme la méthode RAID (Redundant Arrays of Independent Tape), l'architecture **RAIL** répartit les données sur plusieurs cartouches, mais dans différentes bibliothèques. Grâce aux fonctions avancées d'erasure coding, l'architecture RAIL offre une meilleure résilience géographique et une très grande disponibilité en cas de panne totale d'un datacenter ou de catastrophe naturelle. RAIL fournit une architecture de stockage durable, haute disponibilité, sécurisée et à faible coût pour les archives et les données froides de grande envergure dans les datacenters hyperscale.

La création de plusieurs copies des données archivées et froides avec une configuration RAID pour les HSDC et les datacenters à grande échelle cède aujourd'hui la place à des architectures plus économiques reposant sur les technologies d'erasure coding et RAIL. Face à l'augmentation de la capacité des disques durs, le temps de reconstruction de ces derniers à l'aide du RAID traditionnel est devenu trop excessif. Il faut parfois compter plus d'une semaine pour reconstruire un disque dur défaillant. Les risques de perte de données sont en outre plus élevés. ActiveScale prend en charge les fonctions avancées d'erasure coding (sur disque dur) et, désormais, les configurations RAIL (sur bande) dans lesquelles les données sont réparties sur trois sites via la technologie de [déplacement dynamique des données](#) (DDP, Dynamic Data Placement) d'ActiveScale pour garantir un très haut niveau de disponibilité.

Même en cas de panne totale d'un datacenter, la configuration à trois sites assure une disponibilité permanente des données et la continuité des opérations. La technologie de [réparation dynamique des données](#) (DDR, Dynamic Data Repair) d'ActiveScale procède à des audits d'intégrité des données et utilise des fonctions de réparation automatisées pour remédier à la dégradation potentielle des médias, tout en envoyant des alertes à la couche de gestion des systèmes. Des agents sont installés sur les nœuds pour

une gestion hors bande. Ils effectuent un contrôle d'intégrité et une réparation en arrière-plan de manière continue. Les tâches de supervision du système portent sur des paramètres environnementaux (température, ventilateur), avec des statistiques sur l'état des disques qui fournissent des niveaux supplémentaires de disponibilité.

Avantages de la géorépartition ActiveScale avec des architectures de stockage sur bande & RAIL

Objets immuables	Technologie WORM et chiffrement pour les archives de stockage objets sur bande.
Sécurité	L'isolement physique des bandes protège les données contre les cybercriminels et les attaques par ransomware.
Portabilité	Les médias LTO peuvent être facilement transportés en cas de sinistre, alors que le déplacement physique des disques durs est plus compliqué.
Énergie, émissions de CO2	Les bandes sont nettement moins énergivores et produisent beaucoup moins de CO2 que les disques durs. Une plus faible consommation d'énergie réduit les émissions de carbone.
Écosystème Cloud et bande	Les applications Cloud natives peuvent écrire sur les bandes et lire les données qui y sont stockées sans logiciel spécialisé ou propriétaire.
Durabilité et haute disponibilité des archives	La fiabilité des bandes LTO mesurée en taux d'erreurs de bits (BER = 1×10^{19}) dépasse celle des disques durs (BER = 1×10^{16}). Toutes les bandes modernes ont une durée de vie supérieure à 30 ans. Erasure coding avancé avec RAIL.
Géorépartition	Les technologies RAIL, DDP et DDR fournissent des services de disponibilité avancés pour les environnements hyperscale.

Le système ActiveScale Cold Storage peut évoluer de quelques pétaoctets à plusieurs exaoctets d'objets, tout en offrant la disponibilité, la durabilité et les performances nécessaires à ces cas d'utilisation hyperscale. Les architectures de stockage sur bande modernes qui utilisent la technologie RAIL se distinguent par l'évolutivité la plus économique, le coût total de possession et l'empreinte carbone les plus bas, une fiabilité hors pair, le débit le plus rapide et une protection contre les cyberattaques grâce à l'isolement physique. La valeur combinée de ces avantages devrait contribuer à accroître l'usage de la bande à l'avenir et la hisser au rang de technologie de stockage privilégiée, car les besoins en données froides et archivées à grande échelle ne montrent aucun signe de ralentissement.

Conclusion

L'arrivée du logiciel intelligent [ActiveScale](#) Cold Storage de Quantum, compatible S3, marque le début d'une nouvelle classe de stockage objets capable de gérer les données froides, de les stocker, et d'évoluer de quelques pétaoctets à plusieurs exaoctets. Les grandes entreprises, ainsi que les sociétés webscale et hyperscale, sont confrontées aux mêmes problèmes de stockage à froid et devraient sérieusement envisager l'adoption d'ActiveScale, une solution très économique dotée d'une capacité illimitée. Les datacenters à grande échelle sont ceux qui enregistrent aujourd'hui la croissance la plus rapide. Alors

que les besoins en stockage poursuivent leur envolée, l’empreinte carbone, la consommation d’énergie et le développement durable sont au cœur des préoccupations. Comme aucun ralentissement n’est prévu, il est essentiel de repenser les stratégies d’archivage et de stockage à froid pour endiguer la croissance massive des données de manière plus économique et libérer leur valeur inexploitée. Votre réussite passera par votre capacité à connecter, analyser, protéger et exploiter toutes vos données, où qu’elles se trouvent. Une architecture de stockage optimale doit réduire la complexité et permettre aux entreprises de valoriser au maximum leurs données froides sans voir leurs budgets s’envoler au même rythme que ces dernières. La nouvelle classe de stockage objets fournie par la solution ActiveScale de Quantum pour le stockage à froid est prête à relever ce défi.

À propos du sponsor

Quantum®

Avec sa technologie, ses logiciels et ses services, Quantum fournit les solutions dont les entreprises ont besoin pour rendre la vidéo et autres données non structurées plus intelligentes. De cette manière, les données sont au service de l’entreprise et non l’inverse. Forte de plus de 40 ans d’innovation, la société Quantum offre une plate-forme de bout en bout qui inclut des fonctionnalités uniques d’orchestration, de protection et d’enrichissement des données tout au long de leur cycle de vie, fournissant ainsi des renseignements plus riches et des informations exploitables. Des entreprises de premier plan, dans des secteurs aussi divers que les services Cloud, le divertissement, l’administration publique, la recherche, l’éducation, les transports et l’informatique d’entreprise, font confiance à Quantum pour tirer pleinement parti de leurs données, et travailler de manière plus efficace, sûre et intelligente. La société Quantum est cotée au Nasdaq (QMCO) et à l’indice Russell 2000®. Pour plus d’informations, rendez-vous sur www.quantum.com/fr.

À propos de l’auteur



Horison Information Strategies est une société de conseil et d’analyse spécialisée dans le stockage de données. Outre l’organisation de séminaires sectoriels et le développement de stratégies de marché, elle publie des rapports de synthèse, des livres blancs, et des rapports d’études consacrés aux technologies de stockage actuelles et futures. Horison identifie les tendances émergentes et de rupture en matière de stockage des données, ainsi que les opportunités de croissance pour les utilisateurs finaux, les fournisseurs du secteur du stockage et les start-up.

© Horison Information Strategies, Boulder, CO. Tous droits réservés.

WP00268F-v01 Octobre 2021