

Comment le stockage de l'ADN pourrait réduire les centres de données à seulement quelques particules

15 Experts font part de leur réflexion :

Le volume de données que nous produisons est en surrégime, 2,5 quintillions d'octets de données chaque jour – et cela a un impact sur la taille des centres de données, le budget et la rapidité avec laquelle les données peuvent être déplacées. Cette croissance est aujourd'hui un défi majeur pour de nombreux fournisseurs de stockage de données : les besoins en capacité de stockage et en longévité continuent de croître sans aucun signe de ralentissement en vue. Bienvenue dans le stockage sur ADN, une technologie promettant de stocker des millions et des millions de gigaoctets en quelques particules, pendant des milliers d'années.

Cela semble trop beau pour être vrai ? Nous avons réuni un groupe d'experts, de fournisseurs, de chercheurs et d'analystes pour répondre aux questions ci-dessous :

- Stockage DNA : ce que c'est vraiment, ainsi que son potentiel pour l'archivage à long terme
- -Science ou fiction : les équipes informatiques mettront-elles un jour en œuvre cette technologie ou s'agit-il d'un fantasme ?
- -Une révolution du stockage : pourquoi le changement est nécessaire.
- -Le stockage de l'ADN est-il sur le point de faire une différence significative?
- -Et les avantages dont il se vante : du stockage à long terme à grande échelle à la gestion des risques

Parmi eux

Entreprises

- Daniel Chadash, membre du conseil d'administration, **DNA Data Storage Alliance**
- Thomas Ybert, PDG et cofondateur, **DNA Script**
- Sabine Sykora, PhD, Scientifique d'application et Business Developer, **Kilobaser**
- Sergei Serdyuk, vice-président de la gestion des produits, **NAKIVO**
- Curtis Anderson, architecte logiciel, **Panasas**
- Giorgio Regni, directeur technique, **Scality**

Analystes

- Scott Sinclair, directeur de la pratique, **Groupe de stratégie d'entreprise**
- Randy Kerns, stratège principal, **Groupe évaluateur**
- Alexander Harrowell, analyste principal, **Omdia**

Universitaires

- Thomas Heinis, professeur en gestion des données, **Imperial College London**
- Natalio Krasnagor, professeur, **Université de Newcastle**

Futurama : pas si loin ?

Si vous pensez que les données que nous produisons actuellement sont hors de l'échelle, pensez où nous en serons dans dix, 20 ou même 30 ans. Les DSI sont susceptibles de jongler avec des volumes de données beaucoup plus importants que leurs homologues actuels.

Thomas Ybert, PDG et cofondateur de DNA Script, un leader mondial de la synthèse enzymatique de l'ADN (EDS), explique : « Les technologies émergentes, telles que les voitures autonomes et l'intelligence artificielle, augmenteront encore le besoin d'exigences de stockage de données sans précédent. La croissance anticipée des besoins de stockage de données ne peut pas être satisfaite par les technologies actuelles à forte intensité de ressources.

Et qu'advient-il de ces données s'il n'y a nulle part où les stocker ? Je cède la parole à Curtis Anderson,

Architecte logiciel chez Panasas, le moteur de données pour l'innovation : « Nous sommes actuellement, la plupart du temps, en mesure de suivre ... par un élagage prudent des données de faible valeur et la croissance naturelle de la densité des lecteurs individuels, ainsi que la croissance du nombre de ces lecteurs. Nous aurons presque certainement besoin d'un élagage de plus en plus profond des données de faible valeur pour garder une longueur d'avance sur cette explosion. Si le stockage basé sur l'ADN tient ses promesses, nous serons en mesure d'arrêter l'élagage et de simplement stocker tout ce qu'il faut. »

Sabine Sykora, PhD. Application Scientist et Business Developer chez Kilobaser, une équipe engagée dans la démocratisation de la synthèse de l'ADN et de l'ARN, souligne que nous devons prendre en compte plus que la capacité : « Malheureusement, tous les entrepôts d'archives actuels sont confrontés à des limitations fondamentales. Les lecteurs de bandes, par exemple, souffrent d'obsolescence des supports. Les données stockées sur bande doivent être constamment migrées pour tenir compte des pannes d'équipement et des mises à niveau technologiques, ce qui fait des lecteurs de bandes un système de stockage très inefficace qui produit des quantités considérables de déchets électroniques. La technologie de stockage existante devient beaucoup trop coûteuse tout en ayant un impact négatif considérable sur l'environnement. »

Nos experts continuent de revenir sur le domaine de l'environnement et les organisations devront examiner cet impact plus attentivement à l'avenir. **Thomas Heinis**, professeur en gestion des données à l'Imperial College de Londres, confirme : « Les centres de données contribuent aujourd'hui de manière considérable aux émissions de CO2... L'ADN en tant que milieu de stockage n'a pas besoin d'être refroidi et ne nécessite pas d'électricité.

Il ne peut sûrement pas vraiment vouloir dire l'ADN, les éléments constitutifs de la vie? En fait, c'est exactement ce qu'il veut dire: des données numériques stockées dans de l'ADN synthétique et, parce que ces données sont dites récupérables pendant des millénaires, elles sont parfaites pour archiver de grands ensembles de données pendant de longues périodes.

Ne considérez pas cela comme un vœu pieux. Le stockage de l'ADN est déjà utilisé, selon Giorgio Regni, CTO de Scalify, qui unifie la gestion des données, de la périphérie au cœur en passant par le cloud. « Les utilisateurs finaux qui bénéficient déjà du stockage de l'ADN comprennent les chercheurs qui ont besoin de stocker de gros volumes de données, telles que des séquences génomiques », explique-t-il. « À l'avenir, le stockage de l'ADN pourrait être utilisé par les entreprises et les particuliers pour stocker des données pendant de longues périodes . »

Sykora de Kilobaser s'ajoute : « L'application du stockage de l'ADN est déjà en cours de réalisation pour le stockage à long terme... essentiellement la sauvegarde de la sauvegarde. À partir de cette sauvegarde, tous ceux qui stockent des informations sur l'ADN bénéficient car il existe un support alternatif stocké en cours d'utilisation. Cependant, il n'y a actuellement que quelques entreprises qui peuvent fournir un stockage d'ADN avec leur propre codage et infrastructure. Cela limite considérablement le nombre d'utilisateurs finaux à l'heure actuelle. »

Bernard Peultier est vice-président de l'innovation chez Atempo, fournisseur de protection des données. Il est d'accord avec l'affirmation de Sykora selon laquelle le stockage de l'ADN est à long terme. « Vous ne stockez pas les données que vous souhaitez conserver pendant quelques jours ou quelques mois dans l'ADN. L'ADN fait partie de ce que j'appelle le stockage alternatif à long terme. Lorsque vous archivez à long terme, tout archiviste vous le dira, vous devez stocker des informations dans différents types de stockage pour garantir leur durabilité. Ne stockez pas de copies sur le même type de stockage, tout simplement parce que s'il y a une détérioration imprévue dans le temps de laguerre, eh bien, vous avez des problèmes. »

Daniel Chadash siège au conseil d'administration de la DNA Storage Alliance, qui aide à répondre à la demande croissante de stockage d'archives en utilisant l'ADN comme support de stockage. Il explique où ce type de stockage est susceptible d'être adopté: « Les premiers utilisateurs qui en bénéficieront seront probablement dans les industries de la préservation numérique, des médias et du divertissement, des soins de santé et de la recherche scientifique avancée qui doit préserver les données pendant des décennies, voire pour toujours. » Mais tout secteur qui devrait stocker des données toujours croissantes à long terme pourrait être dans le cadre.

Sergei Serdyuk, vice-président de la gestion des produits chez NAKIVO, une société de logiciels à croissance rapide pour la protection des environnements physiques, virtuels et cloud, ajoute: « Alors que le besoin de stockage à haute densité et nécessitant peu de maintenance ne cesse de croître, de plus en plus d'entreprises de divers secteurs investissent dans son développement. Bien que le stockage durable et économe en énergie des données aura un impact positif sur de nombreuses industries, les organisations du secteur de la gestion des données seront probablement les premières à en bénéficier une fois que la technologie passera du laboratoire à la production. »

Donc, oui à l'archivage des données, mais le stockage de l'ADN promet beaucoup plus, selon le professeur d'informatique et de biologie synthétique de l'Université de Newcastle, Natalio Krasnagor. « Je crois fermement que les applications les plus intéressantes n'ont pas encore été découvertes et que celles-ci commenceront à émerger au cours des prochaines années. Mon laboratoire se concentre sur les structures de données ADN plutôt que sur le stockage d'archives. Nous pensons que ce qu'il faut, c'est la capacité non seulement de déchirer de grandes quantités de données, mais aussi de manipuler, trier, rechercher et, plus généralement, traiter ces données de manière uniforme, et de le faire là où les données sont produites, stockées et consommées.

Un mot d'avertissement cependant, de la part d'Alexander Harrowell, analyste principal en informatique avancée pour l'IA de la société de recherche Omdia: « Les principaux avantages sont que le stockage est très efficace, beaucoup d'informations sont emballées dans un petit volume d'ADN et il est stable sans alimentation électrique à long terme. Les

inconvenientsont que l'accès est extrêmement lent et nécessite des équipements, des produits chimiques et des compétences beaucoup plus rares que ceux nécessaires à la gestion d'un lecteur de bande.

Cette lente récupération est en partie ce qui, selon nos experts, limitera l'adoption. Lorsque vous réduisez des données habituellement stockées dans des centres de données qui peuvent rivaliser avec une petite ville en taille et en consommation d'énergie, pour les insérer dans un tube à essai, il y aura un compromis. Pour l'instant du moins.

Mais pourquoi faut-il autant de temps pour récupérer les données stockées dans l'ADN ? Regni de Scalit expose les détails: « Le stockage de l'ADN fonctionne en codant des données numériques dans des séquences d'ADN, qui sont ensuite synthétisées et stockées. Les données peuvent être récupérées en séquençant l'ADN et en décodant les données. L'ADN encapsulé avec du sel reste stable pendant des décennies à température ambiante et devrait durer beaucoup plus longtemps dans des environnements contrôlés comme un centre de données. L'ADN ne nécessite pas de maintenance, et les données stockées dans l'ADN peuvent facilement être copiées à faible coût.

Ybert de DNA Script développe plus loin : « L'ADN est composé de quatre bases qui peuvent être utilisées pour coder des informations. Les données codées devraient être synthétisées sous forme d'ADN à l'aide des technologies d'impression d'acides nucléiques disponibles. DNA Script utilise la synthèse enzymatique de l'ADN (EDS), qui imprime des acides nucléiques sans les produits toxiques utilisés par les méthodes conventionnelles d'impression de chimie phosphoramidite. L'EDS est un procédé chimique « propre » qui ne nécessite pas les conditions environnementales strictes, l'élimination des déchets dangereux ou l'empreinte physique comme le processus conventionnel.

Une autre étape positive vers un stockage de données plus écologique ? Peut-être, mais il y a encore du chemin à parcourir avant de réaliser les avantages potentiels de cette technologie, selon la majorité de nos experts.

Science-fiction – ou réalité ?

Bien que des travaux soient en cours pour intégrer le stockage de l'ADN dans le courant dominant – la collaboration de DNA Script avec Harvard pour fabriquer de l'ADN sur une puce semi-conductrice n'est qu'un exemple – il pourrait s'écouler un certain temps avant qu'il ne devienne un aliment de base dans votre centre de données. Les pierres d'achoppement ? Temps et coût.

Sykora de Kilobaser: « Pour rendre le stockage de l'ADN largement disponible, les deux processus – écriture et lecture – doivent être déplacés du laboratoire vers le bureau. Pour réaliser cette transition, l'écriture et la lecture doivent devenir plus faciles, moins chères et plus rapides. Elle poursuit : « L'Union européenne a décidé de financer plusieurs projets de recherche visant à rendre la synthèse de l'ADN moins chère et plus rapide. Le résultat de ces projets sera de nouvelles technologies qui rapprocheront le stockage de l'ADN d'une application à grande échelle. »

Alessia Marelli est CTO chez DNAalgo, une équipe de vétérans de l'industrie du stockage qui développe le stockage de l'ADN. Elle explique davantage: « Il y a beaucoup d'entreprises qui travaillent sur le stockage de données ADN en ce moment. La DNA Data Storage

Alliance fait un excellent travail en essayant de consolider les feuilles de route et les normes dans ce secteur. Les sociétés de biotechnologie travaillent sur la synthèse et le séquençage pour le rendre plus abordable en termes de coûts et de vitesse. D'autres entreprises comme DNAalgo travaillent sur l'encodage et le décodage pour rendre le stockage des données ADN fiable. Je pense que dans les années à venir, beaucoup de progrès vont arriver, aboutissant à un véritable produit commercial dans cinq à dix ans. »

Heinis d'ICL soutient le calendrier de Marelli : « Nous pensons qu'il sera disponible dans environ cinq ans. La synthèse de l'ADN est trop coûteuse et nous devons trouver des moyens de réduire les coûts, en adaptant également la technologie au stockage des données, c'est-à-dire en nous éloignant du cas d'utilisation des sciences de la vie.

Chadash de la DNA Storage Alliance est également d'accord: « Sur la base des annonces publiques, nous supposons qu'au moins d'ici la fin de cette décennie (2030), sinon plus anciennement, les solutions de stockage d'ADN seront largement disponibles et à un prix compétitif. À l'heure actuelle, de nombreuses entreprises, pour la plupart membres de l'alliance, travaillent dur pour le rendre plus abordable et évolutif. Le travail est axé sur l'écriture de l'ADN, pour améliorer les technologies actuelles et commercialiser de nouvelles technologies, telles que la synthèse enzymatique et la lecture de l'ADN, de nouvelles entreprises émergent avec de nouvelles technologies et des entreprises existantes lancent de nouveaux produits qui réduisent le prix de la lecture.

Mais tout le monde n'est pas de cet avis. Scott Sinclair, directeur de la pratique Cloud, Infrastructure et DevOps chez Enterprise Strategy Group, est dubitatif. « Comptez-moi comme un sceptique », dit-il. « Jusqu'à ce que le format devienne moins cher et plus rapide à exploiter, il sera difficile d'en tirer parti efficacement. Je ne dis pas que cela n'arrivera pas, le stockage de l'ADN a juste besoin de plusieurs innovations pour être dans le domaine de la considération. »

Harrowell d'Omdia est pragmatique : « Il existe des projets de démonstration qui fonctionnent réellement, si vous êtes prêt à dépenser beaucoup d'argent par écriture et à attendre longtemps pour vos données. »

Mais chaque technologie commence quelque part, rappelle **Chadash à propos de la Data Storage Alliance** : « Certains ont affirmé qu'Internet ne réussirait pas et resterait un projet scientifique entre universités, et certains doutaient de notre besoin d'écrans plats ou de disques durs flash. »

Il poursuit : « Nous avons déjà dépassé la phase de science-fiction car quelques entreprises, dont Twist Bioscience et Microsoft, ont déjà montré que la technologie fonctionne. Le défi consiste maintenant à la mettre à l'échelle et à la commercialiser, comme pour toute autre nouvelle technologie. »

Quelle est la prochaine étape?

Lorsqu'on lui demande si le stockage de l'ADN sortira de sa zone de confort archivistique, Ybert de DNA Script est optimiste : « Nous verrons... Nous ne prévoyons pas que le stockage des données changera rapidement du jour au lendemain. Nous prévoyons une coexistence entre les données « chaudes » fréquemment consultées et stockées sur les systèmes traditionnels et les données « froides », qui sont archivées, rarement consultées et ne

nécessitent pas un accès immédiat. Nous avons vu des estimations selon lesquelles 70% des données stockées sont considérées comme archivistiques.

Heinis d'ICL est d'accord : « L'archivage est clairement le fruit le plus facile à atteindre car il s'appuie sur la longévité et la durabilité de l'ADN. Toutefois, pour qu'il se rapproche, le processus doit être considérablement réduit et plus efficace. C'est un long chemin, mais c'est possible. Bien que l'accent soit mis sur le stockage d'archives pour l'instant, il existe un potentiel pour d'autres applications, peut-être grâce à la parallélisation. **Randy**

Kerns, conseiller principal chez Evaluator Group, explique : « Le parallélisme pourrait être dans le séquençage, ce qui accélérera le processus, lui permettant de fonctionner à un niveau supérieur aux attentes en matière de stockage d'archives. En outre, les opérations parallèles contre l'ADN pour permettre d'autres utilisations sont effectuées avec des opérations sur plusieurs éléments de données en même temps, similaires aux arguments en faveur du stockage de calcul avec des SSD. Un exemple peut être l'implémentation d'une fonction de recherche qui pourrait fonctionner en parallèle.

Ainsi, bien qu'il ne soit pas complètement de la science-fiction, le stockage de l'ADN a encore du chemin à parcourir avant de devenir une caractéristique standard de la plupart des centres de données. Actuellement à ses débuts, avec des coûts élevés et de longs délais de récupération, la technologie est prometteuse. Et c'est peut-être le style le plus futuriste auquel nous ayons accès en ce moment. Lorsqu'on nous demande ce qui va suivre, donnons le dernier mot à Anderson de Panasas: « Je ne suis pas encore sûr, je vais devoir demander à mon assistant IA.