Lâ€™ADN synthétique, une révolution à venir pour le stockage de données **Internet**

Posté par : JulieM

Publié le : 22/3/2023 13:00:00

Le PEPR exploratoire MolecuArXiv, piloté par le CNRS, développe de nouveaux dispositifs de stockage de données sur ADN. Son aboutissement marquera une sensationnelle innovation, répondant à de multiples points de blocage déjà pressants. Ce programme et équipement prioritaire de recherche est certainement un des plus importants à soutenir à ce jour.

Une perspective prometteuse

Personne n'est passé à côté des travaux de recherche scientifique sur les possibilités qu'offre l'ADN synthétique en matière de stockage de l'information. Le CNRS publie ses avancées régulièrement et il y a de quoi chanter tel le coq au petit matin si l'on en juge par les progrès manifestes des chercheurs en la matière.

Il sera passionnant de suivre les résultats des tests prochaines applications concrètes. Des partenariats avec des instituts français comme l'INA, la BNF ou des institutions européennes permettront de s'extraire des essais en laboratoire et de travailler avec la matière pour laquelle l'ADN synthétique est conçu, c'est-à-dire de véritables et massifs volumes de données.

Parce qu'il est très prometteur, l'ADN en tant que solution de stockage mérite à lui seul ce coup de cœur. Il devrait contribuer à résoudre un grand nombre de casse-têtes qui se présentent déjà : la saturation du Cloud, la trop grande consommation de foncier notamment urbain des datacenters, leur contribution à l'épuisement des ressources énergétiques, mais également la problématique de la durée de vie des supports de stockage.

Si les grands fournisseurs de Cloud s'efforcent d'atteindre une forme de neutralité carbone en achetant notamment suffisamment d'énergie verte et les entreprises, en privilégiant la mutualisation des ressources, la balance écologique restera déséquilibrée compte tenu de la croissance perpétuelle des données à stocker.

Les promesses de l'ADN synthétique reposent sur la promesse d'une potentielle réduction de la consommation énergétique 2500 fois inférieure à un SSD, jusqu'à ne rien consommer en mode archive froide sans accès. Il nécessite 4 fois moins d'espace nécessaire que les disques ou les bandes et sa pérennité, à l'échelle humaine, est illimitée.

De quoi conserver toute la mémoire de l'humanité à venir.

Couvrir le besoin de stockage froid à long terme

Nous n'y sommes toutefois pas. Le brin d'ADN n'est pas encore à la portée de l'entreprise et de ses archivistes, les scientifiques cherchant pour l'heure à accélérer les vitesses d'écriture et obtenir des codes suffisamment robustes pour empêcher le temps et ses aléas d'altérer la donnée.

En l'état actuel de la technologie, le respect de l'intégrité de la donnée n'est pas garanti, qu'il s'agisse de la donnée lue telle qu'elle a été écrite comme de la

donnée écrite telle qu'on souhaite qu'elle le soit. Les baies de stockage disposent depuis toujours de mécanismes de cette nature, ce que les algorithmes ADN doivent offrir également pour être vraiment exploitables.

Par ailleurs, le stockage ADN n'est pas pensé pour la donnée en production. La donnée transactionnelle n'est pas sa finalité (en tout cas immédiate). La longue durée de conservation qu'il promet le destine au stockage froid. La durée de vie d'un SSD à base de mémoire flash est estimée à 10 ans en moyenne, dans des conditions de stockage idéales, sachant qu'une mémoire flash en outre doit être régulièrement rafraîchie.

Même « punition » avec les bandes magnétiques, que la robotique de bande doit lire régulièrement (tous les 2 ou 3 ans) pour en conserver le magnétisme.

Câ€™est un processus qui reste coûteux, en temps et en énergie.

La promesse de l'ADN est bien de s'affranchir de ce risque de perte de données sur de très longues périodes. Si cette perspective n'est pas encore pleinement confirmée, tous les travaux reposent aujourd'hui sur la Russie initiale du séquençage du génome de mammoths découverts dans le permafrost sibérien, vieux de 4 000 ans.

En attendant l'ADN de stockage

Pressées de toutes parts d'accroître la prise d'initiatives environnementales, tout en étant encouragées à collecter et exploiter leurs données, à mieux les partager, à apprendre les monétiser, les entreprises sont prises en étau entre des injonctions contradictoires.

Aujourd'hui, seules 2 entreprises sur 10 sont considérées comme datacentric, c'est-à-dire ayant véritablement modifié leur façon de travailler et d'innover par le prisme d'une donnée vivante traversant tout leur écosystème. Pour une grande majorité n'ayant pas atteint cet état encore, tout l'enjeu est de conserver en attendant de savoir traiter.

Or le stockage finit toujours par peser lourd financièrement, conduisant à des arbitrages qui pourraient se révéler fâcheux. C'est bien pourquoi le disque dur (via le stockage objet) et la bande n'ont pas dit leur dernier mot, ces deux médias affichant toujours la facture la moins élevée, malgré tout.

La mémoire flash, certes plus chère à l'achat, permet une nette diminution des besoins énergétiques, consomme moins d'espace physique et minimise le dégagement calorifique. Compte tenu de son prix, tous les vendeurs de baies de stockage dignes de ce nom ont en outre introduit plusieurs dispositifs d'efficacité de la donnée par réduction de la volumétrie consommée (compression, duplication, recherche de similarité, etc.), minimisant le besoin d'espace de stockage et d'énergie, d'autant plus dopés par la puissance des nouveaux processeurs.

En revanche, d'autres dispositifs de stockage, étudiés depuis longtemps, restent moins prometteurs. On songera notamment à la MRAM (des charges magnétiques en lieu et place des charges électriques) qui n'aboutit pas réellement.

Qu'on le veuille ou non, volumétrie et performance restent aujourd'hui les mots clés du stockage. Les entreprises vivent une forme d'urgence que ne connaît pas la recherche exploratoire. Nous ne pouvons que soutenir les travaux entamés et appeler de nos vœux plus de coordination et de structuration autour de ce projet au niveau européen.

Rappelons que dâautres alliances existent, notamment nord-amÃ©ricaines, trÃ¨s avancÃ©es Ã©galement avec dâautres techniques ADN, et susceptibles de damer le pion Ã la recherche franÃ§aise, quand bien mÃªme les deux approches seraient complÃ©mentaires, dicit Patrick DUFOUR, Directeur StratÃ©gie et alliances.