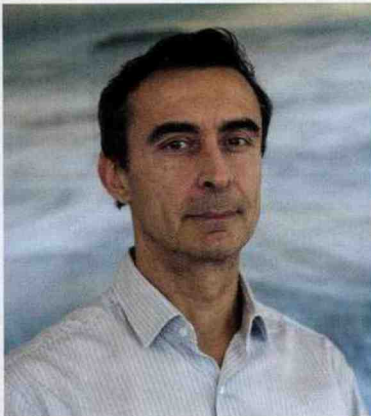


HPC : des solutions françaises, souveraines et prometteuses

Pascal Potier est vice-président exécutif d'Atempo. Dans cet entretien, il nous explique les défis majeurs que permet d'adresser le HPC (High Performance Computing) et nous présente la solution « Agnostique » Miria d'Atempo. Entretien.



Pascal Potier

comme l'aéronautique ou l'automobile. Avec les récentes innovations technologiques et numériques autour de l'IA (intelligence artificielle) et du ML (machine learning), l'usage d'un environnement HPC s'est démocratisé et a pris davantage d'importance. Il est aujourd'hui présent dans plusieurs secteurs (santé, finance, industrie pétrolière, etc.). Le calcul se fait chez l'ensemble des Cloud Providers type Amazon Web Services (AWS), Google ou OVHcloud. Désormais, le défi majeur est de fournir un accès rapide à ces données et d'être en mesure de gérer des volumes très importants.

Aujourd'hui, le grand challenge est la manipulation de larges volumes de données, ainsi que l'accessibilité et la disponibilité de ces données. Qu'en est-il concrètement ?

Auparavant, les données produites étaient plutôt réservées à un nombre restreint de chercheurs de l'entité chargée de l'environnement HPC. Alors qu'aujourd'hui, ces données sont rendues accessibles à une plus grande communauté scientifique, européenne ou mondiale, il est plutôt question de modèles de type Data Hub. On retrouve donc ces modèles dans les domaines de la génétique, de la finance, de l'industrie automobile, etc.

Quels sont les principaux enjeux auxquels répond le HPC (High Performance Computing) ?

Le principal enjeu consiste à collecter ou agréger des données massives et à exécuter des opérations complexes de manière fiable et rapide, donc à accélérer la recherche ou la prise de décision. S'appuyant sur des infrastructures composées de plusieurs milliers de processeurs (CPU ou Central Processing Unit) et sur des unités de stockage de très grande capacité, les résultats de ces calculs sont conservés pour être réutilisés ou analysés ultérieurement.

Traditionnellement, le HPC, ou calcul haute performance, était réservé à certains secteurs de la recherche scientifique, notamment aux départements militaires ou météorologiques de nombreux pays, et à certaines industries



Donc, depuis quelques années, nous assistons à l'émergence de Data Hub, qui consistent à centraliser les données issues de calculs HPC et à les partager au sein d'une communauté de chercheurs, en prenant soin de séparer les données privées ou confidentielles d'une entreprise des données publiques.

Le partage et le déplacement de ces gigantesques volumes de données ne peut se faire que si les infrastructures réseaux et de stockage ont été correctement prises en compte lors du design.

Quid des enjeux du partage de ces données et de leur déplacement ?

Une fois que les données sont générées et cloisonnées, une grande partie est mise à disposition de la communauté et conservée. La difficulté actuelle consiste à pouvoir stocker ces données sur des périodes de plus en plus longues. Alors qu'auparavant, ces données étaient conservées sur des périodes allant de trois à cinq ans, il est aujourd'hui habituel de les stocker sur 15, 20 ans ou plus. Pour ce faire, il faut s'appuyer sur des infrastructures adéquates. Pour des raisons de coûts, une grande partie de ces données produites sont conservées sur des librairies de bandes, soit à proximité de l'environnement HPC, soit dans des infrastructures de Cloud Archive Storage telles que AWS Glacier, Azure Cold Archive, ou sur la solution PCA d'OVHcloud.

Pour répondre à ces enjeux, Atempo met à disposition l'offre « agnostique » de Miria. De quoi s'agit-il ? Quels en sont les bénéfices ?

Le partage de ces très grands volumes de données nécessite un outil capable de déplacer ou dupliquer facilement tout ou partie des données d'un laboratoire de recherche vers un autre ou vers une entité de recherche d'un autre pays, tout en permettant une hétérogénéité entre le stockage source et le stockage de destination.

C'est le réel atout de Miria d'être totalement agnostique. Concrètement, Miria communique avec tous les systèmes de stockage connus du marché, permettant de s'affranchir des contraintes liées aux marques de constructeurs, grâce à des connecteurs sources et destinations compatibles avec l'entièreté des protocoles existants à ce jour. La solution logicielle Miria permet ainsi de couvrir la majorité des besoins observés.



Ces capacités nous permettent de collaborer de plus en plus avec les services de stockage de données à long terme de grandes entreprises afin qu'elles intègrent la brique fonctionnelle Miria pour le déplacement et la conservation de ces données.

Pour conclure, les véritables atouts de notre solution sont à la fois son caractère agnostique et ses capacités de performance et de scalabilité. Quand il est question de manipuler des volumes colossaux de données, il est nécessaire de disposer d'une infrastructure réseau performante, et d'un logiciel, capable de déplacer ces données à très grande vitesse. C'est exactement ce que permet la solution Miria, capable de saturer n'importe quel type de réseau haut débit, à 10 Go, 100 Go ou plusieurs attachements 100 Go/s. La solution logicielle Miria est une solution scale-out, par l'ajout de data movers pour augmenter le nombre de workloads. Si l'objectif est de déplacer les données à très grande vitesse d'un laboratoire vers un autre, ou d'un continent vers un autre, alors Miria répond parfaitement à ces besoins de performance et de capacité, exigés par le HPC.

Pour finir, quels sont les axes de développement que vous allez privilégier pour les prochaines années ?

Pour les prochaines années, nous allons multiplier nos partenariats technologiques, notamment avec les fournisseurs et hébergeurs de données. Ensuite, en tant qu'acteur français, l'un des grands sujets sur lesquels nous

sommes tous mobilisés est celui de la souveraineté des données. Pour ce faire, nous nous appuyons sur des partenaires de renom comme OVHcloud avec qui nous avons signé, début 2021, un accord majeur. Comme Atempo, l'entreprise est engagée sur le sujet brûlant de la souveraineté des données. Nous sommes d'ailleurs convaincus que l'Europe doit se mobiliser davantage et faire le choix de solutions souveraines. Dans ce cadre, Atempo est membre d'Hexatrust, et fait partie du consortium GAIA-X qui travaille actuellement au développement de Data Hub Européens. ✕

EN BREF

Atempo est un éditeur de logiciels français et le leader européen dans le domaine de la Data protection et du Data management. Reconnu sur les cinq continents pour son expertise de la gestion des grands volumes de données dans les secteurs pétroliers (banque et finance, sciences de la vie et de la terre, industries, divertissement et média, etc.), Atempo sécurise les données de ses clients en se concentrant sur ses principales missions :

- la protection des données essentielles,
- la restauration instantanée en cas de besoin,
- la conservation à long terme des données et de leur intégrité,
- la migration rapide et sécurisée des données entre stockages.