

UNE ARCHITECTURE MIXTE POUR
DÉPASSER LES LIMITES DE SQL SERVER

MICROSOFT S'INVITE SUR DES ENTREPÔTS DE DONNÉES



BERNARD FITOUSSI,
DIRECTEUR FRANCE
DATALLEGRO.

Après des débuts dans le conseil chez KPMG Peat Marwick Consultants, Bernard Fitoussi s'est spécialisé depuis une quinzaine d'années dans l'informatique décisionnelle, et en particulier le *datawarehouse* : tout d'abord chez Teradata, puis avec Netezza. Il est depuis 2005 le directeur France de Datallegro.

Quand le géant du logiciel Microsoft présente son projet d'acquiescer un fournisseur de solutions intégrées pour la gestion d'entrepôts de données Datallegro, c'est légitimement que l'on s'interroge sur sa stratégie. La réponse est certainement dans la vision technologique de cette jeune société, qui remplace le *datawarehouse* par des grilles d'*appliances* décisionnelles.

CIO Microsoft a annoncé son projet d'acquisition de Datallegro, dont vous êtes le représentant en France. Les technologies sont pourtant bien éloignées...

BERNARD FITOUSSI : Microsoft n'a pas seulement acheté une technologie, mais également une vision et un savoir-faire dans le domaine du *datawarehouse*. Datallegro devient ainsi le centre de compétence de Microsoft en la matière. Notre technologie (Linux/Ingres) continuera d'être disponible via nos partenaires, dont Bull en France.

NDLR : L'offre actuelle de Datallegro est basée sur des standards : serveurs Dell sur plate-forme multicœur Intel, stockage EMC sur disques, commutateurs Infiniband à 20 Gbps de Cisco et base de données Open Source Ingres sous Linux. Grâce à la technologie alternative d'*appliances* décisionnelles en grid de Datallegro, les entrepôts de données vont s'ouvrir au duo Windows et SQL Server.

Le seul risque dans cette acquisition concernera alors les clients de cette technologie actuelle lorsqu'ils seront incités

à migrer vers la version Windows/SQL Server, lorsqu'elle sera disponible. Ce risque, s'il existe, est néanmoins parfaitement identifié et sera limité, puisqu'en particulier les fonctionnalités et les interfaces du produit seront les mêmes. Et les équipes qui développent le produit 100 % Microsoft sont celles-là mêmes qui ont développé le produit actuel.

CIO Bases de données, entrepôts, décisionnel, où en sont les *datawarehouses* ?

B. F. : Au début, on parlait de SIAD (Système d'aide à la décision ou DSS), ou encore de EIS (Executive information system). De ces vocables ont émergé la notion de décisionnel, aujourd'hui solidement installée, et celle de *datawarehouse*. C'est dire qu'on n'a pas attendu les années 2000 pour utiliser les données disponibles, pour les transformer en informations et enfin en prise de décision.

Mais voilà, stocker des données puis les "faire parler" a abouti très vite à des besoins d'outils spécialisés. On ne traite pas en effet de la même manière des approches transactionnelles (transactions courtes impliquant peu de données) et des approches décisionnelles (requêtes plus longues, volumétries plus ou moins importantes). Il semble qu'il y ait aujourd'hui un consensus : le stockage des données doit rester neutre par rapport aux applications décisionnelles, ou, en d'autres termes, le *datawarehouse* concerne l'infrastructure tandis que la *business intelligence* (BI) est l'affaire des utilisateurs.

Ceci étant, certains points clés de l'infrastructure ont longtemps posé problème : la volumétrie, les performances de restitution, les temps de chargement, le nombre d'accès simultanés, la complexité des requêtes. C'était l'occasion pour certains, comme Teradata, de se poser en champion du *datawarehouse*, peu de concurrents pouvant, avant 2002/2003, rivaliser avec lui sur la plupart des points ci-dessus, en justifiant cette absence de concurrents par des coûts réputés élevés. Mais depuis 2002, avec l'apparition des *appliances* décisionnelles, l'équation a changé, même si certains mythes ont la vie dure.

CIO Pour autant, les "gros" sont toujours là...

B. F. : Et les gros font toujours peur... Il s'agit des *datawarehouses* bien sûr! Les pratiques héritées de temps où l'octet coûtait très cher perdurent curieusement. Nous sommes pourtant parfaitement habitués aujourd'hui à trouver sur nos portables des processeurs de plus en plus rapides et des disques de plus en plus gros. Le téraoctet est même disponible sur les rayons de la grande distribution. Mais dans le monde professionnel, quelques To pour son *datawarehouse* en font frémir plus d'un! Alors on continue souvent de restreindre les besoins fonctionnels pour ne pas franchir un seuil psychologique. Pourquoi les entreprises américaines ont-elles depuis longtemps franchi cette barrière auto-imposée, et pas les européennes? Mystère!

CIO L'entrepôt centralisé n'est-il pas encore présenté comme le modèle idéal?

B. F. : C'est tout du moins ce que prônent depuis très longtemps les IBM ou Teradata, rejoints aujourd'hui par HP. Intellectuellement, c'est probablement vrai : une boîte, plutôt grosse, plutôt noire, plutôt (très!) chère, censée assurer simultanément les services de chargement, respecter les fenêtres d'exploitation, restituer les données à la vitesse de la lumière, fabriquer des agrégats ou *datamarts* à la volée (souvent appelée EDW pour *Entreprise datawarehouse*). Oui mais voilà, dans la vie de tous les jours, ce modèle n'existe pas pour au moins deux raisons de bon sens : les urgences fonctionnelles provoquent l'apparition généralement mal contrôlée de *datamarts*, contre lesquels le *datawarehouse* centralisé ne peut rivaliser, ni en délai de mise en œuvre, ni en coût; le manque patent de puissance nécessaire à la simultanéité des services de base

(alimentation, stockage, agrégation, restitution, sauvegarde...). À moins de disposer d'une plate-forme démesurée, donc à coût complètement prohibitif, l'utilisation de l'un de ces services pénalise nécessairement les autres.

Depuis très longtemps, les entreprises doivent composer avec le modèle centralisé pour certains traitements, et avec le modèle décentralisé pour d'autres. L'intégrité référentielle globale des données entre EDW et *datamarts* n'est pas assurée, ni d'ailleurs la prolifération des *datamarts*, plus souvent subie que maîtrisée.

CIO D'où l'intérêt du *datawarehouse* décentralisé face au centralisé?

B. F. : En réalité, l'apparition des *appliances* décisionnelles n'a pas nécessairement résolu ce problème de choix d'architecture entre modes centralisé et décentralisé. À quelques rares exceptions, les acteurs de cette tendance se cantonnent soit dans le centralisé (IBM, HP, Teradata) soit dans le décentralisé/*datamarts* (Netezza, Kognitio, Dataupia...), laissant les questions fondamentales de l'intégrité référentielle

et du ratio coût/performance sans réponse satisfaisante. Pourtant, une réponse est apportée par une grille d'*appliances* spécialisées plutôt qu'un monolithe. Grâce à des composants technologiques élémentaires totalement standard, il est possible d'être à l'aise dans les deux mondes, depuis les plus petits *datamarts* jusqu'aux plus gros EDW. Un réseau de plusieurs *appliances*, chacune étant configurée pour une fonction précise (alimentation, agrégation, stockage principal ou *hub*, restitution, propagation de données, sauvegarde...), est une réponse

appropriée. Elle est rendue possible grâce aux mécanismes de réplication très rapides d'une part, et à la gestion des méta-données, d'autre part. Des références opérationnelles de premier plan existent déjà. Dans ces conditions, les modèles centralisé et décentralisé peuvent cohabiter harmonieusement dans une architecture mixte où chaque département utilisateur, et bien sûr l'informatique centrale, restent maîtres de leurs budgets respectifs et de leurs choix technologiques et métiers. Les applications sont nombreuses : PRA (plan de reprise d'activité), multitempérature des données en fonction de leur âge, ou encore un modèle EDW enfin opérationnel, doté de très hautes performances. ■

PROPOS RECUEILLIS PAR YVES GRANDMONTAGNE



MICROSOFT VEUT ÉTENDRE SON OFFRE DE PLATES-FORMES DE DONNÉES "MISSION CRITIQUE".

MICROSOFT VISE LES GROS VOLUMES

Microsoft a annoncé son projet d'acquiescer Datallegro et ses solutions alternatives d'*appliances* décisionnelles en *grid* pour équiper des fermes de données qui peuvent atteindre jusqu'à plusieurs centaines de téra octets (au-delà des capacités de la base de données maison SQL Server...). L'éditeur entend profiter de la croissance exponentielle du marché du stockage de données et combiner cette solution avec le futur SQL Server version 2008. Son annonce précise sa volonté d'étendre son offre de plates-formes de données "mission critique" pour les utilisateurs disposant de gros volumes de données.



<http://www.cio-online.com/>

CIO n°38 - septembre 2008

L'informatique poussée au vert

Dossier : les DSI aux premières lignes du Green IT - Exclusif : Renault en route vers une DSI "world-class" - Projet : Les députés satisfaits d'une année avec des PC sous Linux - International : Top 10 des causes d'échecs des projets SOA - Fournisseurs : **Microsoft s'invite sur des entrepôts de données** - Tribune libre : Le modèle Google face à la vie privée est un florilège de mauvaises pratiques.