

Extrait du Axège, aide à la décision et ERP - Axège Santé

<http://www.axege.com>

Les bases de données vectorielles au service du décisionnel

- Thème techno -

Date de mise en ligne : vendredi 13 novembre 2009

Axège, aide à la décision et ERP - Axège Santé

La base de données vectorielles, représente une **nouvelle génération** de gestionnaire de base de données sur le marché des SGBD. Destinées essentiellement aux **systèmes décisionnels** et aux datamarts, les bases de données vectorielles sont des structures qui répondent aux difficultés que peuvent rencontrer les bases de données relationnelles traditionnelles.

Avantages significatifs pour l'utilisateur :

- ✿ **Performances nettement supérieures** aux bases relationnelles classiques lors de la lecture d'un grand volume d'information (de l'ordre de 1000 fois sur les requêtes complexes).
- ✿ **Stockage compressé** des données telle qu'elles existent dans le monde réel en encodant les données de manière individuelle **évitant ainsi toute redondance**
- ✿ Infiniment extensible et centrée sur l'utilisateur (**meilleure ergonomie** et création des bases associée facilitée).

Écrits selon les standards du marché (SQL), les SGBDV (systèmes de gestion de base de données vectorielles) permettent donc une **meilleure prise en main** par les utilisateurs. L'architecture est complètement ouverte et multi-plateforme (Windows, Linux, AIX) avec des postes « clients allégés » permettant la **gestion de gros volumes** de données en un **temps record**.

Sources :

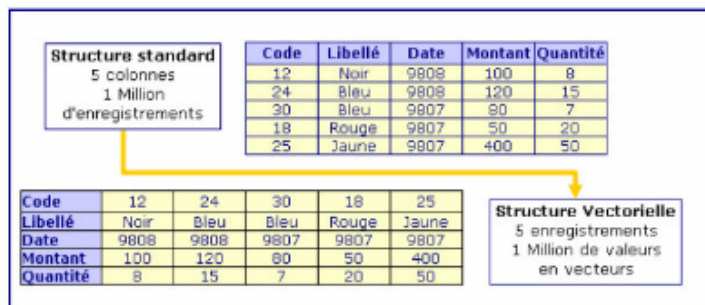
Les bases de données multidimensionnelles :

<http://www.journaldunet.com/developpeur/tutoriel/sql/061012-sgbd-multidimensionnel.shtml>

Découverte de l'architecture fonctionnelle des SGBDV :

Une base de données vectorielles va stocker les données par colonne et non par ligne comme dans une table classique (de type fichier ou table relationnelle qui représente chaque enregistrement par une valeur par colonne). Cette orientation permet d'ajouter des colonnes plus facilement aux tables (les lignes n'ont pas besoin d'être redimensionnées). Elle permet de plus une compression par colonne, efficace lorsque les données de la colonne se ressemblent.

Une table vectorielle est donc constituée par la création d'un enregistrement par colonne. Chaque enregistrement va décrire la totalité des valeurs d'une colonne pour la table initiale. L'enregistrement "nom" contiendra toutes les valeurs "nom" du fichier.



Cette structure physique des tables ne change rien à la vision et l'utilisation par l'utilisateur, qui continue à travailler sur un fichier "client" ou "ligne de factures" conceptuellement identiques [1].

Sources :

The Vector Database : <http://www.vectordatabase.com/index.html>

Les SGBDV : <http://keydatasoft.com/sqbdv.aspx>

Une technologie sur la voie de la banalisation :

Avec l'amélioration des processeurs (64bits) et la baisse du coût de la mémoire vive, une autre révolution est en train de s'opérer dans le monde du décisionnel. En effet, le pilotage de l'entreprise est désormais possible de manière quasi-instantanée avec le chargement des données décisionnelles directement dans la mémoire des serveurs. Les actions coûteuses en temps que sont les lectures mécaniques des disques n'existent plus avec ces technologies. Cela explique en bonne partie la rapidité de ces SGBD.

Certains outils sont dorés et déjà capable d'utiliser ces solutions. C'est le cas par exemple de MonetDB et de QlikView.

MonetDB [2] :

MonetDB est un système de gestion de base de données vectorielles Open Source. Il a été conçu pour offrir de hautes performances sur des requêtes complexes ainsi que sur de grandes bases de données, par exemple combinant texte, tableaux avec des centaines de colonnes et de plusieurs millions de lignes. MonetDB a été appliqué avec succès dans des applications de haute performance pour l'extraction de données comme OLAP, les SIG, XML Query.

Source : MonetDB : <http://en.wikipedia.org/wiki/MonetDB>

Qlikview [3] :

Qlikview quant à lui est une solution « packagée » commerciale incluant l'alimentation, le stockage et un outil qui génère la restitution des données (avec interface graphique). QlikView base donc son innovation sur l'intégration, la facilité d'utilisation et le rendu de tous les indicateurs de manière graphique sur un portail privé permettant ainsi l'analyse en « temps réel » des données et les prises de décisions.

Source : Qlikview : www.qliktech.com



Conclusion :

Ainsi, le monde du décisionnel informatique est en complète refonte depuis quelques années afin d'aboutir à des solutions de pilotage en temps réel, plus adapté aux marché fluctuant d'aujourd'hui. L'analytique en mémoire se place également comme un complément essentiel aux systèmes de données traditionnels.

Les SGBDV sont **très performants pour la lecture et l'extraction** des données mais supportent encore l'inconvénient d'être assez **long lors de la modification ou de la création** des enregistrements (puisque le système doit vérifier l'ensemble des colonnes pour faire les changements). C'est pour cela qu'Axège s'oriente vers l'**utilisation des deux systèmes** jugés par nous **complémentaires**. D'un coté, nous n'abandonnons pas les bases de données classiques qui constituent le **socle** des applications et de l'autre, nous étudions la mise en place des SGBDV pour les analyses de données dynamiques (cube) permettant des **résultats immédiats**.

[1] Une requête se décompose classiquement en deux phases : sélection d'enregistrements puis extraction et éventuelle agrégation. En environnement vectoriel, seules les colonnes présentes dans les requêtes sont lues. Aucun cube n'est préalablement défini. Le nombre de colonnes initiales du fichier n'intervient donc pas dans les performances.

[2] Développé à l'Institut national de recherche en Mathématiques et Informatique des Pays-Bas, MonetDB s'appuie sur la mémoire vive et sur le processeur pour représenter les données internes. La pagination à la demande permet de s'écarter des SGBD de conception traditionnelle, impliquant une gestion complexe des données et de grands magasins en mémoire limitée.

MonetDB introduit des innovations dans toutes les couches d'un SGBD pour permettre un modèle de stockage basées sur la fragmentation verticale. Son processeur vectorisées lui confère dans l'exécution de la requête un avantage en terme de vitesse très largement supérieur à un SGBDR typique. Il dispose également d'un auto-index et d'une optimisation des requêtes temps pour une architecture logicielle modulaire. La

famille MonetDB se compose de :

- ▶ MonetDB / SQL : la solution de base de données relationnelle.
- ▶ MonetDB / Xquery : la solution de base de données XML.
- ▶ MonetDB serveur : le serveur multi-base de données modèle.

[3] Plutôt qu'une technologie de cube Olap multidimensionnelle traditionnelle, Qliktech a implémenté dans son logiciel Qlikview le stockage vectoriel des données. Alors que le cube Olap met l'accent sur le préformatage et le précalcul des données afin d'optimiser les temps d'accès disque, le stockage vectoriel en RAM stocke des pointeurs vers les données redondantes et optimise ainsi la volumétrie pour monter en mémoire vive les données collectées par son ETL (Extract Transfert Load). On peut stocker un très grand volume de données dans un espace mémoire de quelques gigaoctets, et chaque attribut est potentiellement un axe d'analyse. La vitesse de l'accès en RAM est suffisante pour assurer des temps de réponse quasi instantanés.