

Information Géographique et SGBD

chronique d'une convergence annoncée

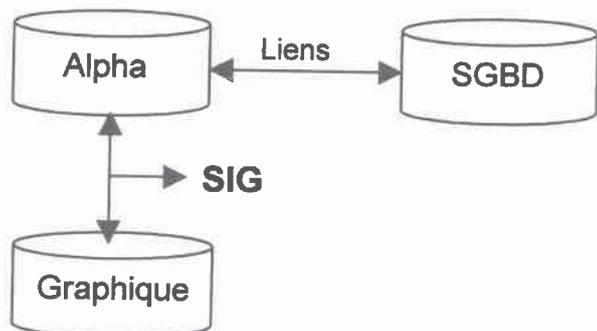
Michel Bernard (SIAGE Conseil)

Parmi les évolutions technologiques majeures dans le monde des SIG de cette fin de décennie, deux éléments émergent de façon notable. Le premier - le plus visible - est bien entendu Internet. Nous avons déjà évoqué dans ces colonnes l'impact important du World Wide Web sur l'industrie de la géomatique, et l'on a pu constater en l'espace d'une année l'évolution de ce "monde virtuel", culminant récemment en ce qui concerne l'hexagone avec la décision de la DGI de mettre les plans cadastraux en ligne sur Internet.

Le second phénomène, moins évident et moins immédiatement perceptible pour les utilisateurs, mais tout aussi important et fondamental à notre sens, est celui de la convergence des technologies géomatiques et des Systèmes de Gestion de Bases de Données. Si Internet offre d'ores et déjà la possibilité de démocratiser l'information géographique en la rendant plus accessible et plus facilement "consommable", encore faut-il que les technologies de stockage et de distribution, encore largement fondées sur des modèles propriétaires, évoluent vers une réelle standardisation. Cette évolution que l'on attendait depuis environ dix ans et qui relevait encore de la recherche ou des effets d'annonce plus ou moins fondés est aujourd'hui portée par une majorité d'éditeurs. Parmi ceux-ci, il est important de noter que figurent non seulement les acteurs traditionnels du monde SIG (Esri, Mapinfo par exemple) mais également les sociétés spécialisées dans la fourniture de SGBDR généralistes tels Oracle ou Informix.

RETOUR AUX SOURCES

Le principe même des SIG consiste à mettre en relation permanente des données à caractère spatial (géométriques) avec des informations sémantiques et à assurer la cohérence de ces relations. Pour autant, dans la quasi-totalité des outils du marché, le stockage des données est réalisé selon un modèle hybride comportant d'une part le stockage des données géométriques dans un système propriétaire, d'autre part la gestion des données descriptives dans un second système, parfois standard (dbase, oracle, access), le plus souvent lui aussi propriétaire.



Architecture classique des SIG

Bien sûr les liens ODBC (Open DataBase Connectivity) et autres outils d'accès à des bases de données distantes (Remote Data Access) ont permis de mettre en place des architectures assez ouvertes. Ces assemblages ont pourtant leurs limites, tant sur le plan de l'intégrité des données que de la gestion des transactions, des accès concurrents ou la résolution des problèmes de mise à jour et d'historisation des données.

Cette dualité se retrouve également dans les logiciels que l'on a souvent abusivement qualifiés d'orienté-objet (Géocity ou Smallworld). Un rare exemple de stockage intégré est celui du produit CCHR utilisé par APIC pour gérer à la fois les données géométriques et sémantiques, mais les bénéfices offerts par cette technique sont aujourd'hui contrés par les limites en termes d'évolution et d'ouverture vers le monde des SGBD Relationnels. Limites que l'éditeur APIC systèmes cherche lui aussi à contourner en se tournant vers de nouvelles solutions dites "SGBD Universels".

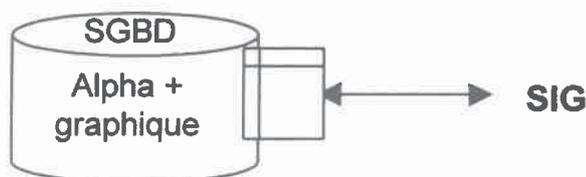
L'enjeu pour le monde de la géomatique est important car il s'agit de tendre le plus possible vers les standards du marché, afin d'éviter une marginalisation de l'information géographique numérique.

La fusion tant annoncée des données géographiques dans les systèmes d'information, que les outils Internet/Intranet peuvent contribuer à rendre réelle, ne saurait s'imaginer dans un contexte de technologies propriétaires.

ÉVOLUTION DES SYSTÈMES DE STOCKAGE : DES SGBD UNIVERSELS

Ayant mis au point des systèmes de gestion de plus en plus robustes et performants, les éditeurs de SGBDR ont pu s'intéresser à des marchés émergents et potentiellement porteurs, la géomatique étant — ce qui est rassurant — considérée comme telle. L'essentiel des recherches menées ces dernières années par les éditeurs a porté sur la mise au point de systèmes de stockage et d'extraction permettant de gérer les données à taille variable (objets (géo) graphiques, images, sons) dans des systèmes prévus à l'origine pour des données à taille fixe (numériques, alphanumériques).

Les outils qui arrivent sur le marché ont pour nom SDO/SDC (Spatial Data Cartridge) chez Oracle, Datablades chez Informix mais aussi SDE (Spatial Data Extension) pour Esri et Spatialware pour Mapinfo. Schématiquement ces offres sont en fait ce que l'on appelle généralement du "middleware" c'est à dire des couches logicielles qui viennent se connecter à des outils généralistes (les SGBD Universels) pour permettre de gérer et distribuer les données spatiales.



Leur potentiel consiste non seulement à offrir les moyens de stocker l'information géographique (points, lignes et polygones ou pixels) avec ses attributs associés mais également à en faciliter l'accès grâce à des techniques d'index spatial performantes. Cette dernière notion est particulièrement importante pour pouvoir optimiser la recherche et l'extraction de données à l'intérieur de volumes de plus en plus importants.

Parmi ces techniques d'indexation, certaines sont connues depuis longtemps (la méthode des Quadtree notamment ou les clés de Peano) et d'autres sont plus récentes (le modèle Helical Hyperspatial ou HHCODE développé par le service hydrographique de la marine canadienne et la société Oracle).

On voit donc fleurir des offres de plus en plus solides en termes de potentiel à gérer de très grosses bases de données, mais encore faut-il savoir comment accéder à ces outils et comment développer des solutions pratiques dans une vision utilisateur final ?

ÉVOLUTION DES LANGAGES D'INTERROGATION

L'un des écueils à l'utilisation des SGBDR pour la manipulation des données géographiques et plus généralement des données à taille et format variable (multimédia) a longtemps été l'absence de langages de description et de manipulation (DML) adaptés.

Dans ses versions normalisées actuelles le langage SQL (Structured Query Language), développé dans l'optique de systèmes d'information classiques, n'offre pas de possibilité de déclarer (donc de créer, mettre à jour et analyser) des données géographiques ni de les inter-

roger sur des critères spatiaux (à l'intérieur de, à distance, dessus, dessous...). La plupart des éditeurs respectent pour l'instant le niveau SQL2 qui est normalisé par l'International Standards Organisation (ISO). Mais à ce niveau, le langage ne comporte aucun opérateur spatial et ne permet pas de déclarer des objets géographiques.

Pour dépasser ces limites, des travaux importants ont été menés depuis trois ou quatre ans (notamment sous l'impulsion de la société Unisys) pour aboutir à un langage plus riche : SQL/MM (MM pour multimédia). Ces spécifications ne sont pas encore acceptées en tant que normes mais quelques éditeurs commencent déjà à les proposer. Quand on sait que la société Mapinfo (éditeur du logiciel du même nom) a racheté à Unisys sa technologie de gestion de bases de données géographiques on comprend mieux pourquoi les produits proposés sous le label "SpatialWare" offrent précisément d'exploiter ces extensions.

L'intérêt d'un langage d'interrogation spatiale normalisé est utile au moins à deux titres : d'une part pour permettre d'adresser des requêtes à différentes bases de données géographiques en étant sûr que les résultats seront cohérents, d'autre part pour faciliter l'exploitation des données géographiques dans des applications "non géographiques".

À titre d'exemple, si une application de gestion des transports scolaires peut envoyer directement à un SGBD une requête du type "trouver toutes les adresses des élèves situés à moins de 1 km d'un collège" et récupérer ces adresses pour analyser les conditions d'octroi de subventions, on peut considérer que 90 % du traitement utile du point de vue de la géographie sera réalisé sans recourir à un "SIG classique".

Cet exemple (volontairement choisi) met toutefois en évidence les limites potentielles de cette approche car si les opérateurs proposés à ce jour dans les SGBD universels savent traiter le cas simple de l'inclusion d'objets dans un polygone (à vol d'oiseau), il n'en va pas de même pour le traitement de graphe qui est nécessaire à la résolution de la question posée (les élèves ne vont pas à l'école en volant mais en marchant...).

Mais en réalité, rares sont aujourd'hui les questions qui atteignent un tel degré de complexité (et également aussi rares — ou alors complexes à utiliser — les SIG qui permettent d'y répondre !).

Pour résumer, on peut donc exécuter à l'aide d'un SGBD universel comportant des extensions spatiales, la plupart des tris et requêtes utiles sur des données géographiques, les extraire puis les exploiter dans des outils ne comportant qu'une partie (géo) graphique très limitée, voire nulle.

Bien entendu ce n'est pas la seule finalité, l'intérêt étant évidemment de pouvoir également mettre en place des systèmes d'information dans lesquelles les clients (au sens client/serveur) seront des SIG, des applications métier "géographiques" ou des outils de cartographie.

Par ailleurs, il est bien évident que l'on ne saurait se passer de systèmes purement géomatiques pour les étapes d'acquisition ou de mise à jour des données spatiales, les éditeurs de SGBD n'ayant pas vocation à développer de tels outils "périphériques".

COMMENT UTILISER CES TECHNOLOGIES ?

Pour l'instant il ne s'agit pas de prétendre que tous les utilisateurs de SIG doivent converger vers des SGBDR étendus et utiliser SQL/MM, ne serait-ce que parce qu'il s'agit de technologies en pleine maturation et également encore assez coûteuses.

Des solutions clé en mains sont toutefois disponibles chez quelques éditeurs qui commencent à proposer des solutions de connexion directe avec quelques-uns des SIG du marché (Mapinfo, Arc/Info ou Arcview, Autocad Map et Autodesk World...).

SGBDR	Produit	Editeur SIG
.ORACLE	SDO/SDC (ORACLE)	Autodesk Autres ?
ORACLE, DB2,	SDE (ESRI)	ESRI
INFORMIX	DATABLADES	Mapinfo

Quelques SGBD Universels du marché

Ces outils s'adressent dès aujourd'hui aux organisations disposant de bases de données géographiques volumineuses et qui désirent mettre en place des systèmes de gestion et de distribution ouverts vers une plus large gamme d'utilisateurs. Les collectivités sont parmi les premiers utilisateurs visés (certaines ont franchi le pas ces derniers mois) mais d'autres gros consommateurs dans des segments de marché moins classiques (télécommunications, marketing, immobilier) sont également intéressés.

ET LES FORMATS D'ÉCHANGE ?

Le passage de bases de données entre différents systèmes de gestion impose évidemment des traductions et modifications de structure. Des travaux importants ont

été menés ces dernières années — dans plusieurs pays — autour des problèmes de formats d'échange.

Cette étape de conversion est rendue nécessaire pas le fait que chaque outil utilise son propre modèle de stockage et structure les données selon des approches différentes. Dans la perspective d'un transfert classique (livraison d'une base de données), le format d'échange est une clé essentielle.

Par contre, en cas d'accès par une procédure de requête sur une base de données "universelle", il y aura bien évidemment modification du format initial mais cette opération sera totalement masquée à l'utilisateur. Dans la mesure où la structure même des bases de données se conforme à un standard (SQL/MM), un système doit pouvoir extraire des informations et les traiter dans leur de manière transparente, à l'image de ce qui se passe lorsque l'on crée une jointure entre des tables sur des bases de données différentes.

EN CONCLUSION

Les Systèmes d'Information Géographique sont encore souvent perçus — à juste titre — comme des outils trop spécialisés pour une grande majorité d'utilisateurs, et ce même avec l'arrivée d'outils bureautiques.

Pourtant, la diffusion de l'information est en train de connaître une véritable révolution gr,ce aux technologies Internet et Intranet, mais le prix à payer pour rester "en ligne" sera de passer par les standards et d'abandonner les environnements et les structures de données trop "propriétaires".

Les SGBD universels et leurs corollaires (SQL/MM) arrivent donc à point nommé pour permettre à l'information géographique d'être plus largement diffusée dans les systèmes d'information.

*Michel Bernard est consultant spécialisé en SIG à la société SIAGE Conseil :
2 rue Edmond Lautard, 34184 Montpellier Cedex 4
E-mail : mbernard@siage. fr*

Association Française de Topographie ADHEREZ

Pour le contact permanent avec la profession, la prise directe avec la science et la technique du métier. Pour se situer dans la topographie dont l'université est probante. L'une des solutions est d'adhérer à l'AFT.

L'AFT est le lieu géométrique où se rencontrent les grands organismes de la topographie, le cadastre, le Service Hydrographique, l'IGN... les grands Ecoles de la profession, l'ENSAIS, l'ENSG, l'ESTP, l'ENC... les hommes et les femmes des grandes Ecoles de la Nation, Polytechnique, Centrale... et aussi tous les ingénieurs, techniciens, hommes de terrain, qui font chaque jour le tissu expérimenté d'un métier que l'AFT a pour vocation de faire partager par tous, en promouvant la solidarité professionnelle.