

Intégration de l'analyse multidimensionnelle dans la gestion d'un réseau d'auscultation géométrique

■ Aïcha DERKAOUI - Boualem GHEZALI

Le présent article illustre l'intérêt de l'outil statistique d'analyse multidimensionnelle des données issues de la mise en place d'un canevas géodésique destiné aux études d'auscultation d'ouvrages d'art ou de sites industriels.

La gestion des données variant dans le temps et d'un volume important, repose sur le concept DataWarehouse. Il s'agit d'une collection de données orientées sujet, intégrées, non volatiles, historisées, et organisées comme support d'un processus d'aide à la décision. L'outil d'analyse multidimensionnelle utilisé est le système OLAP (Online Analytical Processing) qui a l'avantage d'analyser l'information facilement et rapidement. L'application porte sur la gestion et l'analyse du réseau d'auscultation de six bacs de stockage du complexe GP 1/Z de la Sonatrach.

La structuration et l'exploitation d'une quantité importante de données dynamiques rendent nécessaire, voire indispensable, de disposer d'outils d'intégration permettant l'accessibilité, la manipulation et l'analyse multidimensionnelle des informations. A cette fin, le concept d'entrepôt de données proposé est la collection d'informations orientées sujet, intégrées, non volatiles et historisées, et organisées comme support d'un processus d'aide à la décision. (W.H. Inmon 1995, Prism. Volume 1). L'analyse de cette définition point par point montre l'intérêt de cette technique par rapport à la gestion des données.

Si les DataWarehouses ou les DataMarts sont les lieux de stockage des données analytiques, le système OLAP permet aux applications clientes d'y accéder efficacement et présente de nombreux avantages pour les utilisateurs analytiques, notamment :

- un modèle de données multidimensionnel intuitif qui permet de sélectionner, de parcourir et de maîtriser plus facilement les données ;

MOTS-CLÉS

Réseau géodésique, auscultation, multidimensionnel, DataWarehouse, OLAP

- un langage de requête analytique puissant qui permet d'explorer les relations complexes entre les données de l'entreprise ;
- le pré-calcul des données fréquemment interrogées permettant d'accélérer le temps de réponse pour les requêtes concernées.

Le travail présenté à travers cet article consiste à l'intégration de l'analyse multidimensionnelle dans la gestion d'un réseau d'auscultation qui comporte trois grandes étapes.

Inventaire des données

L'application porte sur le réseau d'auscultation géométrique du site industriel représenté par six bacs aériens de stockage de gaz liquéfié du complexe GP1/Z d'Arzew (Sonatrach). L'opération a été étalée sur deux campagnes d'observations spatiales (GPS) et terrestres, espacées de deux ans.

Présentation du réseau d'auscultation

Le réseau d'auscultation géométrique des bacs de stockage est composé d'un (figure 1) :

- **Réseau de base** (référence) ou réseau de garde, constitué de deux points géodésiques (BD09 et BC12) supposés stables, qui sert d'ossature pour tous les travaux ultérieurs réalisés dans le cadre de l'auscultation géométrique des bacs.
- **Réseau d'auscultation** appuyé sur le réseau de base constitué de 24 bornes bien réparties.



Figure 1. Schéma de la configuration du canevas d'auscultation.

- **Réseau de cibles** distribuées de façon homogène (maillage approprié) et matérialisées sur l'ouvrage et sur son voisinage immédiat. Il est composé de 23 cibles collées sur le couronnement de chaque bac et de deux cibles supplémentaires placées sur la plate-forme extérieure.

■ Observations et traitements

Plusieurs opérations d'observations et de traitements sont réalisées pour le suivi de l'évolution des bacs et de leurs voisinages :

- **Opération 0 (époque T₀)** : observation GPS et ajustement des coordonnées tridimensionnelles des points des réseaux d'auscultation et de cibles. Les positions de ces points par rapport au réseau de base constituent la configuration de référence.
- **Opération i (époque T_i)** : détermination de la position des points des réseaux d'auscultation et de cibles à partir d'une nouvelle session d'observations et de traitements pour quantifier les déplacements des bacs et de leurs voisinages par rapport à la configuration de référence. La nouvelle configuration ainsi déterminée (opération i) sert de configuration de référence pour l'opération d'auscultation suivante (i+1). La fréquence d'intervention est définie en fonction de l'exploitation et varie selon l'amplitude des déformations, la vitesse de déplacement et par suite d'événements naturels (séisme, glissement de terrain...).

La méthodologie développée pour la mise en place du canevas d'auscultation est basée sur l'exploitation de mesures de :

- la phase des signaux GPS en modes statique et statique rapide,
- nivellement de précision,
- cheminements planimétriques,
- rattachements par intersections et rayonnements,
- intersections spatiales.

Les coordonnées des points de l'ensemble du réseau de cibles pour les deux phases, espacées d'une année, constituent les données de base pour l'évaluation des mouvements reposant sur la détermination des vecteurs de déplacement et des tenseurs de déformation. Les interprétations géodésiques s'effectuent suivant les tendances des déformations et des déplacements.

Modélisation des données

La modélisation des données repose sur le concept d'entrepôt de données permettant d'intégrer et d'exploiter les données jugées pertinentes.

Un entrepôt de données est orienté sujet et doit permettre d'analyser des données suivant plusieurs dimensions, à la différence des bases de données relationnelles classiques. Il permet d'effectuer des traitements sur des données en prenant en compte plus de deux axes : les données ne sont pas modélisées sous forme tabulaire mais sous forme d'hyper-cubes (ou cube de données).

Les données utilisées, issues de deux opérations d'auscultation, pour la validation du modèle, sont celles du réseau de cibles.

Un sujet est défini par un ensemble de mesures et un ensemble de dimensions. Dans le cas étudié, les mesures du point correspondent aux coordonnées cartésiennes (X, Y, Z) et l'étude comme l'analyse reposent sur leurs valeurs numériques. Les dimensions sont : la période, le lieu et le type de coordonnées. Ce sont les points de vue depuis lesquels les mesures peuvent être calculées.

La dimension est la liste d'éléments organisés de façon hiérarchique : la granularité d'une dimension est son nombre de niveaux hiérarchiques. Pour la dimension période la hiérarchie est la suivante : année, mois et jour, soit trois niveaux. Les axes des dimensions doivent fournir des règles de calculs d'agrégat pour chaque mesure (figure 2).

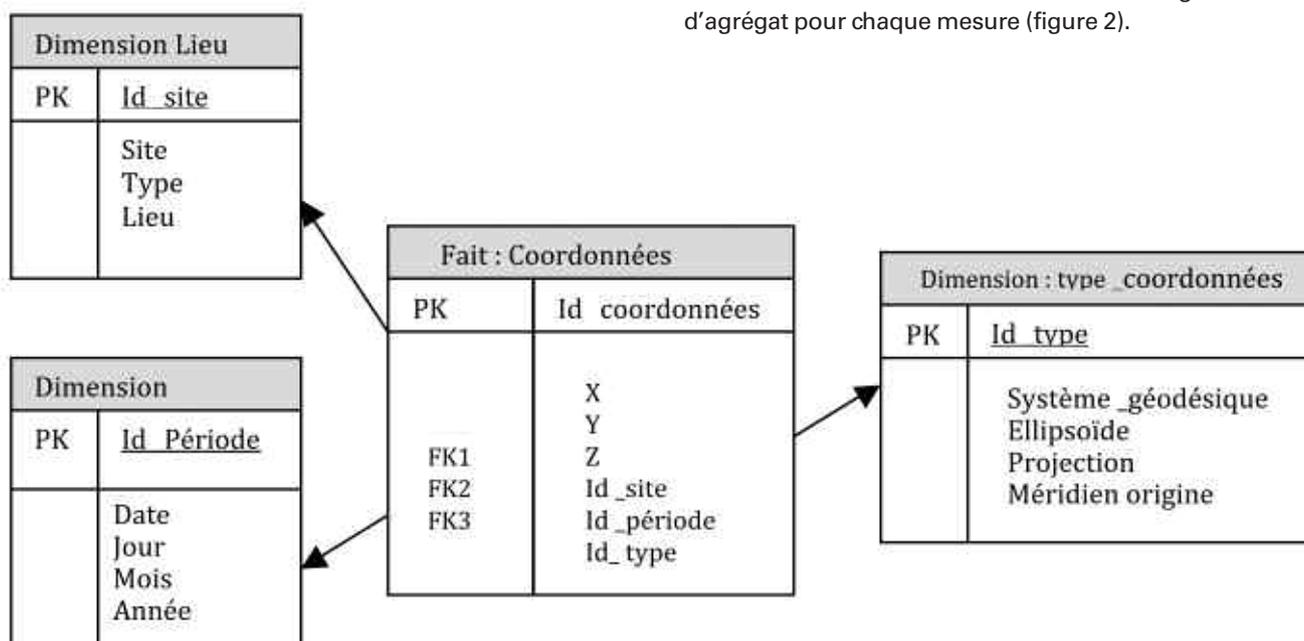
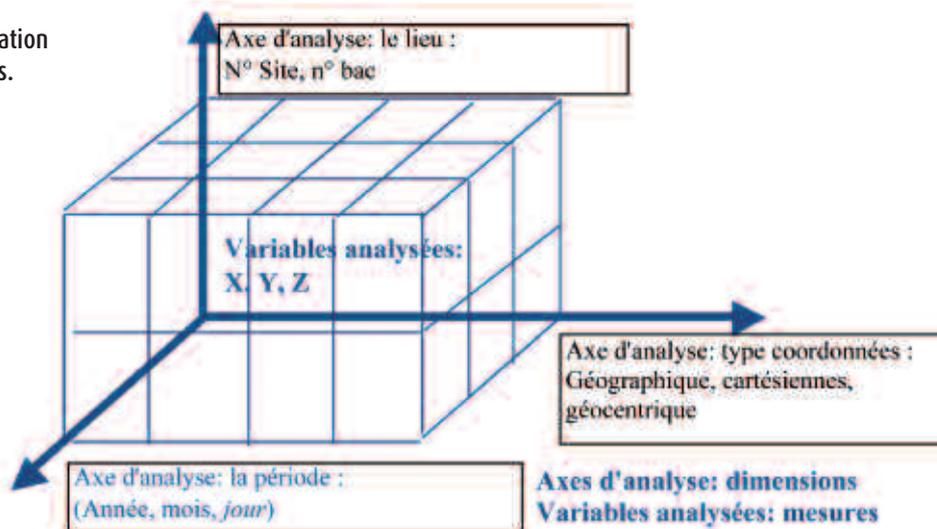


Figure 2. Modélisation de l'entrepôt à l'aide d'un schéma en étoile.

Figure 3. Représentation du cube de données.



Ces différents mécanismes permettent de placer les données dans des matrices multidimensionnelles de forme cubique (Figure 3). Les données sont interrogées directement et facilement sur n'importe quelle combinaison de dimensions, sans utiliser de requêtes trop complexes.

Passer d'une hiérarchie de dimension à une autre est réalisée facilement dans un cube de données par la technique de pivot, de rotation. Par cette technique, le cube peut être pivoté pour afficher différentes orientations des axes.

Les opérations slice et dice permettent respectivement d'extraire une tranche du cube et un sous-cube selon des prédicats sur les dimensions. Ces opérations sont illustrées respectivement par les figures 4 et 5.

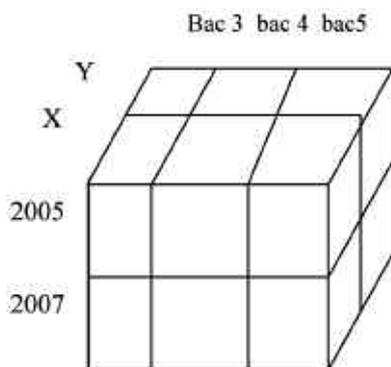


Figure 4. Opération dice, extraction d'un sous-cube.

	Bac 3	Bac 4	Bac 4	Bac 5	Bac 7	Bac 8
X						
Y						
Z						

Figure 5. Opération slice Coupe sur 2005.

D'autres opérations importantes sont possibles sur les bases de données multidimensionnelles : le roll-up et le drill-down. Ces deux opérations permettent de naviguer dans les hiérarchies de dimensions.

- **Le roll-up** : permet d'agréger les données suivant une dimension.
- **Le drill-down** : permet de faire le contraire, c'est-à-dire de détailler les données.

Analyse de données

Il existe deux types d'analyse de données : le datamining (ou fouille de données) et l'analyse multidimensionnelle (OLAP).

La fouille de données (datamining), appelé aussi Extraction de Connaissance (ECD), permet de mettre en évidence des corrélations éventuelles dans un volume important de données afin de dégager des tendances. Elle s'appuie sur des techniques d'intelligence artificielles ou sur des techniques statistiques afin de mettre des liens cachés entre les données et ainsi de prévoir des tendances (figure 6).

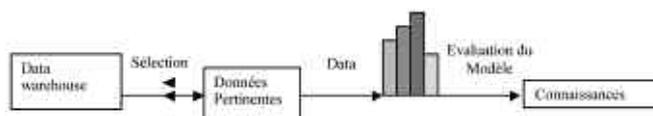


Figure 6. Processus d'extraction de connaissance.

Cette technique permet d'extraire les informations nécessaires (non triviales, implicites, non connues précédemment et potentiellement utiles).

Le résultat du modèle de mining dans le cube constitue une nouvelle dimension de ce cube, c'est une dimension virtuelle. Les données relatives au modèle de mining sont alors consultables à travers le cube virtuel regroupant le cube de départ et cette nouvelle dimension

Online Analytical processing (OLAP), c'est le terme qui désigne les bases de données multidimensionnelles destinées à l'analyse et il s'oppose au terme OLTP qui désigne les systèmes transactionnels.

Date	Point	X	Y	Z
2009-01-01	P1	10.000	20.000	30.000
2009-01-02	P2	10.000	20.000	30.000
2009-01-03	P3	10.000	20.000	30.000
2009-01-04	P4	10.000	20.000	30.000
2009-01-05	P5	10.000	20.000	30.000
2009-01-06	P6	10.000	20.000	30.000
2009-01-07	P7	10.000	20.000	30.000
2009-01-08	P8	10.000	20.000	30.000
2009-01-09	P9	10.000	20.000	30.000
2009-01-10	P10	10.000	20.000	30.000
2009-01-11	P11	10.000	20.000	30.000
2009-01-12	P12	10.000	20.000	30.000
2009-01-13	P13	10.000	20.000	30.000
2009-01-14	P14	10.000	20.000	30.000
2009-01-15	P15	10.000	20.000	30.000
2009-01-16	P16	10.000	20.000	30.000
2009-01-17	P17	10.000	20.000	30.000
2009-01-18	P18	10.000	20.000	30.000
2009-01-19	P19	10.000	20.000	30.000
2009-01-20	P20	10.000	20.000	30.000
2009-01-21	P21	10.000	20.000	30.000
2009-01-22	P22	10.000	20.000	30.000
2009-01-23	P23	10.000	20.000	30.000
2009-01-24	P24	10.000	20.000	30.000
2009-01-25	P25	10.000	20.000	30.000
2009-01-26	P26	10.000	20.000	30.000
2009-01-27	P27	10.000	20.000	30.000
2009-01-28	P28	10.000	20.000	30.000
2009-01-29	P29	10.000	20.000	30.000
2009-01-30	P30	10.000	20.000	30.000

Figure 7. Extrait de la table des déplacements.

Le concept OLAP est appliqué au modèle virtuel de représentation de données (le Cube). Il existe ensuite plusieurs déclinaisons qui permettent d'adapter le stockage des données sur différents types de bases de données pour implémenter le concept OLAP :

R-Olap : technique de modélisation et de stockage des données basée sur une structure relationnelle.

M-Olap : un OLAP optimisé pour l'analyse multidimensionnelle.

H-Olap : est un hybride entre le R-olap et le M-Olap.

Le choix du M-Olap dans ce cadre est dû à ce que la base M-Olap est l'application physique du concept Olap qui convient le mieux à l'étude. Il s'agit réellement d'une structure multidimensionnelle et les bases M-Olap sont rapides et performantes et proposent des fonctionnalités particulièrement élevées.

La possibilité de création des membres calculés ; définis à l'aide d'une combinaison de données de cube, d'opérateurs arithmétiques, de nombres et de fonctions, dépendant des membres existants dans le cube ou aussi d'autres membres calculés définis, nous permet de faire des calculs assez facilement tel le calcul des déplacements en X, Y et Z et aussi de savoir sur quelle direction le déplacement est le plus important. La figure 7 illustre les données relatives aux déplacements.

Le cube Olap donne accès à des fonctions d'extraction de l'information (visualisation, analyse et traitement), et à des fonctions de requêtes en langage MDX.

Le langage MDX similaire au langage SQL pour les SGBD est un autre moyen d'interrogation du cube et permet d'exécuter des requêtes MDX simples ou complexes.

Le langage MDX offre la possibilité d'exécuter facilement des requêtes statistiques pour le calcul de la covariance et de la corrélation. En d'autres termes, il permet le calcul de la corrélation entre les données ou encore même d'ajouter des fonctions définies par l'utilisateur lui-même.

Conclusion

Les entrepôts de données et leurs outils deviennent indispensables pour la structuration, l'exploitation et l'analyse multidimensionnelle d'une quantité importante de données dynamiques.

Le but de l'étude est la mise en place d'un processus permettant l'analyse et l'exploration du cube de données. L'entrepôt de données a été schématisé par un modèle en étoile. L'exploration de cet entrepôt a été effectuée par des opérateurs d'analyse OLAP.

L'application a été enrichie par la conception de cet entrepôt de données avec des outils offrant des possibilités d'analyse et d'exploration des cubes de données avancées et non supportées par les systèmes SGBD classiques. ●

Contacts

Aicha DERKAOUI, Boualem GHEZALI

Centre des Techniques Spatiales

Division de Géodésie Spatiale

BP 13, 31200 Arzew, Algérie.

derkaouia@hotmail.com

Références

BEDARD Y. (2002) *Introduction aux systèmes SOLAP*. Cours sujet spécial. Centre de recherche en géomatique : Université de Laval.

BEDARD Y & PROUX M-J. (2004) *Comparaison de l'approche transactionnelle des SIG avec l'approche multidimensionnelle pour l'analyse de données spatio-temporelles*. Article : (colloque géomatique 2004).

CARON P-Y. (1998) *Etude du potentiel d'OLAP pour supporter l'analyse spatio-temporelle*. Mémoire de maîtrise. Département des sciences géomatiques, faculté de foresterie et géomatique : université LAVAL.

KONNEN J. (2004) *Couplage d'un ERP et d'un datawarehouse avec un SIG : le mariage d'un monde multidimensionnel avec la cartographie*. Mémoire présenté en vue de l'obtention du diplôme post grade en informatique et organisation : Université de LAUSANNE.

[Microsoft Corporation, 2000] : Microsoft SQL Server 2000 : Analysis Services Didacticiel.

VERHAEGEN B. (2006) *Requêtes OLAP sur une base de données XML native*. Mémoire de licence en informatique : Université Libre de Bruxelles. Belgique.

ZIANI S. (2000) *Auscultation topographique par technique GPS*. Mémoire d'ingénieur d'état en sciences géodésiques et travaux topographiques : Centre National des Techniques Spatiales. Algérie

ABSTRACT

Key words: Geodetic network, auscultation, multi-dimensional, DataWarehouse, OLAP.

The present paper illustrates the statistical tool interest of multi-dimensional data analysis data issued from the setting up of geodetic canvas intended to auscultation studies of art or industrial sites.

The management of data varying in the time and important volume, rests on the DataWarehouse concept. It is about data collection oriented topic, integrated, non volatile, historized, and organized like support of a help process to the decision. The tool of multi-dimensional analysis used is the OLAP system (Online Analytical Processing) that has the advantage to analyze information easily and quickly.

The application is about the management and the network auscultation analysis of six storage tanks of the Sonatrach complex GP1/Z.