

Variations dans les dimensions

par Yazid Grim ([Business Intelligen\(ce\)](#))

Date de publication : 09-01-2008

Dernière mise à jour : 09-01-2008

Cet article présente les trois principales manières de gérer le changement de données dans un Data Warehouse.

Variations dans les dimensions.

Cas pratique.

Le problème.

Les conséquences.

Solution.

Gestion de l'historique dans les entrepôts de données.

Changements de type 1.

Changements de type 2.

Changements de type 3.

Autres types de changements.

Conclusion.

Remerciements

Variations dans les dimensions.

Une des principales différences entre les OLTP (systèmes opérationnels) et les entrepôts de données est la capacité de ces derniers à capter les changements et les conserver.

En effet il est primordial, pour une bonne analyse, de savoir que le prix d'un produit a changé au courant de l'année ou qu'un représentant a changé de zone géographique au milieu de l'année. Les conséquences pourraient être fâcheuses si ces informations venaient à disparaître (et c'est souvent le cas). Voyons voir pourquoi cela.

Cas pratique.

Un cas pratique étant plus parlant que toute chose, analysons les variations dans les dimensions d'un entrepôt de données

Le problème.

Imaginons une entreprise qui vend des imprimantes et assure un soutien technique auprès des clients. Cette entreprise possède un système de gestion rigoureusement alimenté par les employés (techniciens, représentants, responsables de stocks, etc.). Ce système de gestion fait très bien son travail : consignation de l'information, récapitulatifs quotidiens, accès distant pour les nomades, facturation, etc. Mais une fois n'est pas coutume, les patrons décident de l'utiliser pour sortir des analyses de ventes sur plusieurs années, voir le profit généré par territoire, par représentant, les activités des techniciens par territoire, etc, etc.

Cela va sans dire, le système de gestion va planter à coup sur, et ajouter de la mémoire au serveur ne fera que retarder l'inévitable, mais ceci n'est pas l'objet de cet article. Le problème est ce qui se produit si l'un de ses événements (qui sont courants dans la vie d'une entreprise) se réalise :

- **Le prix d'une imprimante change au milieu de l'année.**
- **Un client déménage au milieu de l'année.**
- **Un représentant change de territoire pour servir une autre catégorie de clients.**

Les conséquences.

Voici les conséquences de tels changements dans l'ordre :

- **Pour l'imprimante qui a changé de prix** : comme les systèmes opérationnels ne prennent pas en charge la gestion de l'historique (pas complètement, et c'est normal car la gestion d'historique est un domaine à lui seul). L'ancien prix sera perdu, donc les chiffres des ventes à la fin de l'année ne se baseront que sur le nouveau prix. Donc des chiffres faux.
- **Pour le client qui déménage** : la plupart des systèmes opérationnels de ventes utilisent la localisation géographique comme une partie de la clé primaire de la table Clients. Donc si le client déménage (ou change de nom, ou fusionne avec une autre compagnie, etc.), certains voudront créer un nouveau client pour ne pas compliquer les choses. Le problème est qu'en créant un nouveau client, on "omet" toutes ses transactions précédentes ! Donc si à la fin de l'année, les patrons décident de récompenser les meilleurs clients, et que le client en question déménage en Juin. Il n'aura que six mois de transactions à son actif# Pire ! Si l'on veut retirer le pourcentage de rabais aux mauvais clients et que durant les six derniers mois le client n'a pas payé à temps, ce dernier se verra perdre ses avantages, même s'il a été très rentable les six premiers mois !!!

- **Pour le représentant qui change de territoire** : c'est le cas que je préfère. Et bien en réassignant un autre représentant à un territoire, il se verra attribuer toutes les ventes de ce territoire, même celles qui ont été faites avant son affectation. Sympa !

Solution.

La solution étant évidente : FAIRE UN ENTREPÔT DE DONNÉES !!! Qui conservera l'historique des changements pertinents et sera capable de les restituer dans les calculs, rapports et pendant les explorations de données. L'entrepôt de données est fait pour ça ! Et les systèmes opérationnels sont faits pour opérer, c'est-à-dire exécuter des tâches de production, quotidiennes, élémentaires et répétitives.

Gestion de l'historique dans les entrepôts de données.

Un entrepôt est fait de dimensions et de faits. Pour plus de détails à ce sujet je vous invite à consulter mon blog. Les faits sont des tables consignants # des faits, c'est-à-dire tout ce que l'entreprise veut mesurer : ventes, commandes, réclamations, temps de traitements, productivité, etc. Il est très rare qu'il y ait des changements dans ces tables (on ne va pas changer le montant d'une facture). L'autre élément qui compose l'entrepôt est la dimension. Ce sont des tables qui correspondent à nos axes d'analyse (avec quoi faire l'analyse). On pourrait avoir des dimensions Clients, Pays, Fournisseur, Vendeur pour faire des analyses par clients, par pays, par fournisseur, etc. Exemple : le chiffre d'affaire annuel par pays, par fournisseur et par client. Ce sont les dimensions qui changent, et ce sont les changements dans les tables de dimensions que l'on veut conserver. Voyons voir comment faire. Monsieur Ralf Kimball, père du Data Warehousing, a défini trois types de gestion de changements basiques dans un entrepôt de données.

Changements de type 1.

C'est le cas le plus simple, et pour cause, on applique un changement de type 1 lorsqu'on ne veut pas conserver l'historique d'un champ. En effet, il est rare qu'une entreprise veuille conserver les changements du numéro de téléphone d'un client, ou veuille conserver les changements dans le sexe (sans commentaire !). Pour ces champs que l'entreprise ne juge pas pertinents, un simple UPDATE dans la table de dimension fera très bien l'affaire.

Changements de type 2.

Les choses sérieuses commencent ! Les changements du prix d'un produit est une chose que tout le monde voudrait conserver. Voyons voir comment : Si vous en avez lu pas mal sur le Data Warehousing, vous avez sûrement entendu dire qu'il ne faut JAMAIS utiliser la clé primaire d'une table du système de production comme clé dans l'entrepôt de données. Et bien c'est pour pouvoir implémenter des changements de type 2 (entre autres) que cette vive recommandation est souvent formulée.

Imaginez une dimension Produit toute simple, avec une clé primaire, la clé du système de prod, l'intitulé du produit et le prix. Voici un Exemple de ce que peut contenir cette dimension :

Clé de Dimension	Clé d'entreprise	Nom du produit	Prix du produit
123	52268	Planche à pain en bois	29.99 \$

Le prix de notre planche à pain passe de 29.99 à 19.99. Première étape : modéliser le changement.

Clé de Dimension	Clé d'entreprise	Nom du produit	Prix du produit
123	52268	Planche à pain en bois	29.99 \$
...

359	52268	Planche à pain en bois	19.99 \$
-----	-------	------------------------	----------

Notre changement est maintenant consigné. Deuxième étape : comment faire pointer les ventes avant le changement du prix vers la ligne avec l'ancien prix et les ventes après changement du prix vers notre nouvelle ligne. Et bien nous avons deux possibilités :

- Ajouter un flag d'activité : les anciennes lignes étant déjà liées avec la ligne 123, il suffit d'ajouter un flag " actif " à la table, de mettre la nouvelle ligne active, et l'ancienne inactive, et modifier la requête d'insertion pour prendre en compte ce champ (lier les éléments de la table de faits avec les éléments de dimension actifs).
- Ajouter un champ de date début et un autre de date fin : solution équivalente, plus explicite aussi et surtout meilleure si l'implémentation d'un tel changement s'est fait après la mise en production de l'entrepôt. Les comparaisons des dates permettront d'utiliser la bonne ligne dans la table de dimension.

Changements de type 3.

Ce type de changement est un peu moins ouvert, et le moins utilisé. Il est utilisé lorsqu'on veut garder un historique restreint (le dernier changement, ou les deux derniers changements). L'implémentation de cette solution est simple : ajouter autant de colonnes que de changements désirés, avec les dates associées. Exemple : Notre dimension Client est de la structure suivante :

Clé de Dimension	Clé d'entreprise	Nom du produit	Prix du produit	Adresse
------------------	------------------	----------------	-----------------	---------


Si nous voulions garder l'ancienne et la nouvelle adresse, nous appliquerions un changement de type 3 comme suit :

Clé de Dimension	Clé d'entreprise	Nom du produit	Prix du produit	Adresse avant changement	Date du changement	Adresse actuelle
------------------	------------------	----------------	-----------------	--------------------------	--------------------	------------------

Avec ce type de gestion d'historique, on peut éviter l'accumulation d'informations inutiles.

Il est clair que cette solution est peu évolutive, car un niveau d'historique supplémentaire équivaut à des ajouts de colonnes dans la dimension. Mais dans plusieurs cas, cette solution peut être très utile.

Autres types de changements.

Il existe dans la littérature d'autres types de changements, (article sur  [les dimensions à variation lente sur Wikipedia](#)). Je vous invite à les consulter si vous vous sentez assez avancés pour ce faire. Il faut juste savoir que les trois types décrits plus haut sont de loin les plus utilisés dans les projets d'entrepôt de données.

Conclusion.

Nous avons vu l'importance de la sauvegarde des données historiques dans une entreprise, puis, nous avons vu comment cette sauvegarde peut être implémentée à travers les trois principaux types de changements. La chose essentielle qu'il faut garder en tête lors de la conception d'un entrepôt de données, est qu'il est le garant ou la mémoire de l'entreprise. Toute information à son poids et doit être minutieusement étudiée pour savoir son importance et le degré de son importance. Une omission peut rendre l'entrepôt inutilisable car non fiable. Faisons attention ! :)

Remerciements

Je voudrais remercier RideKick et Adrien Artero pour leurs conseils avisés et leurs relectures attentives. Merci les gars !

