

Modélisation OLAP pour Entrepôt de données

Enoncé

Une grande entreprise veut rassembler dans un **entrepôt de données** des informations sur les ventes du jour afin de dresser des tableaux de bord sur les ventes. **L'ED à modéliser doit pouvoir fournir le chiffre d'affaires des ventes d'un produit, par date, client, et vendeur, ainsi que toutes les sommes possibles de chiffre d'affaires.**

Une vente correspond à un produit et un seul,

Les produits sont regroupés par famille de produits.

La vente est effectuée par l'un des vendeurs du service de vente spécialisé dans le produit.

La semaine de vente est le numéro de semaine dans l'année.

1. Quelle est la table des faits ?
2. Combien de dimensions ont été retenues ? Quelles sont-elles ?
3. Quelles sont les hiérarchies des dimensions ? Dessinez-les.
4. Faites une représentation du cube OLAP.
5. Quels sont les niveaux de grains de ce cube ?
6. Quelles sont les opérations applicables sur les cubes OLAP en général ?

1. Table des Faits

La table principale de la base de données « entrepôt de données » sera alors la suivante :

| |
|---|
| Table VENTE |
| Date Code produit Code vendeur Code client |
| Montant de la vente |

Figure 1 : Table VENTE

En grisé clair, apparaît la **clé multiple** de l'enregistrement, constitué de 4 éléments :

date, code produit, code vendeur, code client

En grisé foncé, figure la variable à mesurer, appelée **indicateur** :

montant de la vente

Cette table **VENTE** pourrait suffire à faire des sommes de chiffre d'affaires, si l'on se contentait des codes sur les éditions. C'est la table fondamentale d'un ED. On l'appelle : **table de faits**

2 .Les dimensions

En fait, la plupart du temps, chacun des éléments de la clé multiple de la table de faits renvoie à un certain nombre d'attributs. Ici, par exemple, le code produit sera utilement complété par :

libellé du produit

code famille de produit

libellé famille

(nombreuses informations complémentaires possibles)

On fait alors une jointure entre la table de faits et une table dite table de dimension, selon le schéma suivant :

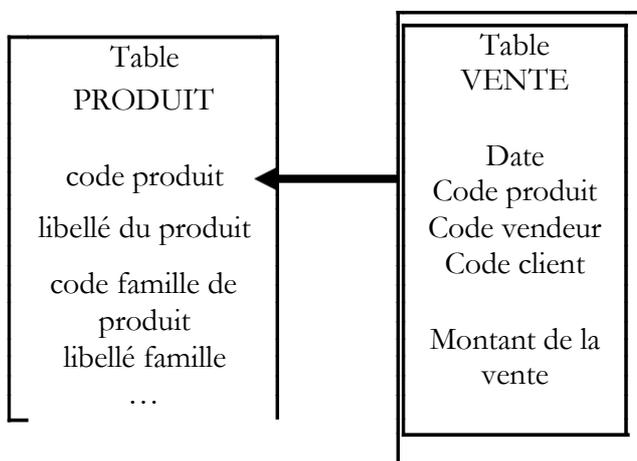


Figure 2 : Table de faits + 1 table de dimension

La table de dimension 'PRODUIT' n'a pas d'indicateur, elle comporte seulement des attributs du produit. La clé « code produit » correspond à l'élément « code produit » de la clé multiple de 'VENTE'.

De la même manière, les autres éléments de la clé multiple renvoient en général chacun à une table de dimension

3, Schéma fait_dimensions

le schéma des tables est le suivant :

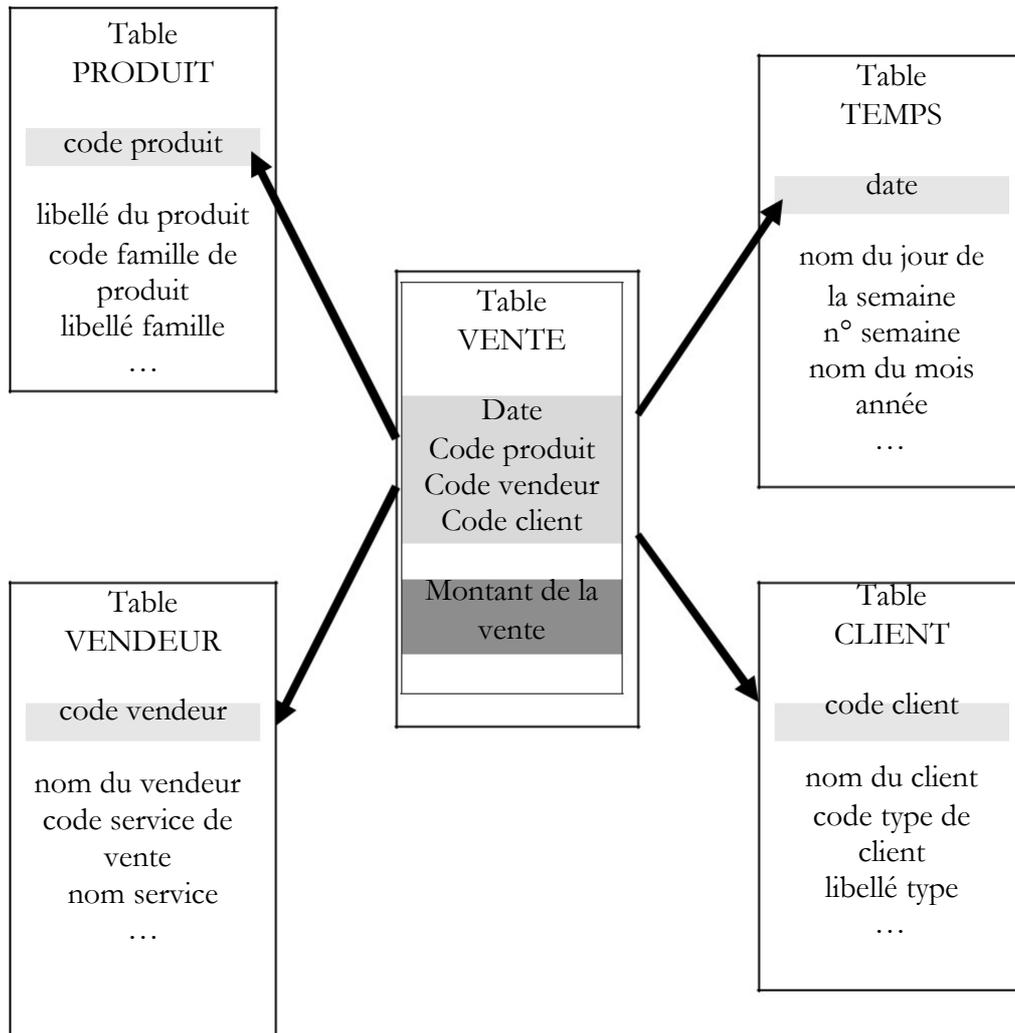


Figure : Table de faits + 4 tables de dimensions

Ce schéma, avec au centre la table de faits, et autour les tables de dimension jointes, s'appelle : **schéma en étoile**

Ce schéma est caractéristique de la **modélisation dimensionnelle** (du nom des dimensions) la plupart du temps mise en œuvre dans la conception d'un Entrepôt.

3.Cube OLAP *Online Analytical Processing*

Un schéma en étoile peut également être représenté sous forme de **cube de données**

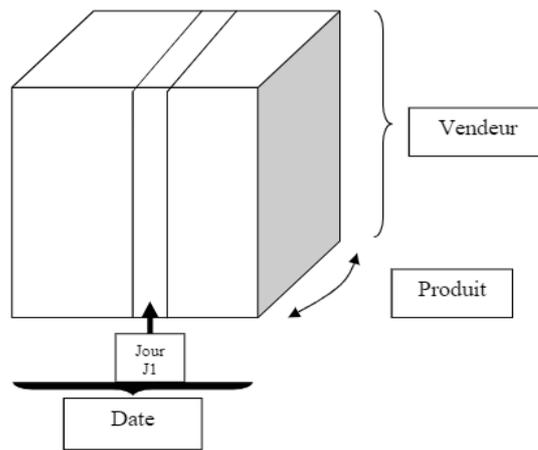


Figure : cube de données à 3 dimensions

Lorsque le nombre de dimensions est de 3, la base de données peut être représentée par un **cube**. Lorsqu'il est supérieur à 3 c'est un **hypercube**. Pour simplifier, on parle dans tous les cas d'un **cube de données**.

3,1. Coupes

Dans l'exemple traité, et représenté par le schéma en étoile (Figure 3), le cube de données est un hypercube à 4 dimensions : **produit, client, vendeur, date**.

Graphiquement, on peut dessiner en perspective 4 types de cubes à 3 dimensions :

- | | |
|------------------------------------|---------------------------------|
| A. client, vendeur, date | (pour chaque valeur de produit) |
| B. produit, vendeur, date | (pour chaque valeur de client) |
| C. produit, client, date | (pour chaque valeur de vendeur) |
| D. produit, client, vendeur | (pour chaque valeur de date) |

Dans chaque cube, l'élément de base est l'**indicateur** « montant de la vente ».

Tracé de cube D :

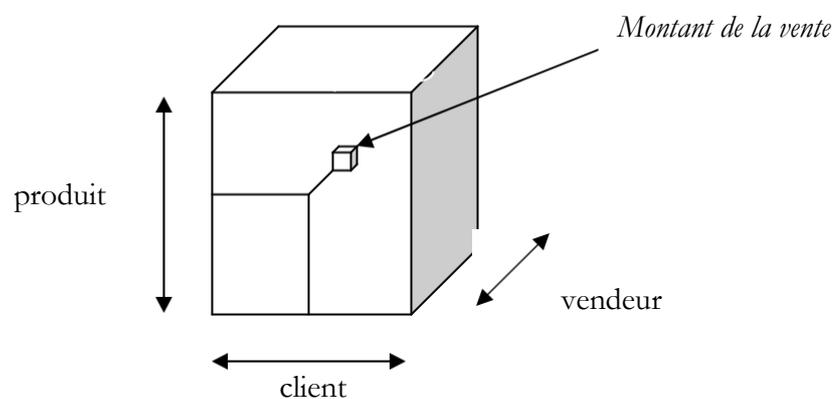


Figure 6 : un des types de cubes à 3 dimensions

On peut tracer autant de cubes D qu'il y a de valeurs pour la variable **date**.

Un cube D représente une **coupe** de l'hypercube à 4 dimensions, selon une valeur de la variable **date**.

De même, on peut faire des **coupes** du cube D pour toutes les valeurs de **produit**, par exemple. On obtient alors autant de tableaux à 2 dimensions (**client, vendeur**) qu'il y a de valeurs à **produit**.

A partir de D, on peut faire 3 types de tableaux à 2 dimensions :

(**client, vendeur**), (**produit, vendeur**), (**client, produit**)

A partir de l'ensemble A, B, C, D, on peut faire en plus les 3 coupes qui gardent **date**

(**client, date**), (**produit, date**), (**vendeur, date**),

donc en tout 6 types de tableaux à 2 dimensions.

5. Niveau de grain

Chaque fait est collecté selon un **niveau de grain** ou un niveau de granularité qui dicte le niveau de détail ou de profondeur jusqu'au quel on peut aller dans les analyses de données.

Pour le cube OLAP précédent, on peut choisir une analyse par jour, par semaine...ou par année. Les produits peuvent aussi être catégorisés en niveau par catégorie Meuble peut regrouper les grains chaises, armoires....etc,

6. Opérations sur le cube:

- Rotate: sélectionne le pivot d'analyse en faisant tourner le cube;
- Slice: extrait une tranche du cube, c'est-à-dire, un sous-ensemble des valeurs du cube correspondant à une certaine valeur d'une des dimensions non-découpées;
- Dice: provient de l'application de l'opération slice sur plus d'une dimension;
- Drill-down: permet de détailler les données en descendant le niveau hiérarchique d'une dimension;
- Roll-up (drill-up): permet d'agréger les données en augmentant le niveau hiérarchique d'une dimension;

Référence:

J.-F. Desnos, Entrepôt de données – Introduction, IHS / Master ICA, 2020