

OLAP PAR L'EXEMPLE

La vente de
chaussures

NEGRE Elsa
Université Paris-Dauphine

2013-2014

Source : Bernard Lupin

CONTEXTE

- La société *Au bon pied*,
- installée à Bonneuil,
- désire construire une base de données
- pour suivre l'évolution de ses ventes de chaussures,
- par modèle et par mois.

■ 7 étapes :

1. Les propriétés de la table VENTE
2. Les 3 dimensions du NOMBRE
3. Les hiérarchies de la dimension TEMPS
4. Les calculs à la volée
5. Les attributs
6. Les calculs temporels
7. Les données éparses

ETAPE 1 :

LES PROPRIÉTÉS DE LA TABLE *VENTE* (1)

- But : construire progressivement une base de données tout à fait classique, avec des tables constituées de colonnes.
- Une table **Vente** :

Mois	Modèle	Nombre	Total HT
Janvier 2000	Escarpin	5	1 700 €
Janvier 2000	Ville	300	65 000 €
.....			

- Dans la réalité : les différents types de chaussure sont stockés dans une table **modèle**, et on ne met dans la table **vente** que la clé correspondante.

ETAPE 1 :

LES PROPRIÉTÉS DE LA TABLE *VENTE* (2)

■ Reportings possibles :

- nombre de paires de chaussure par mois et par modèle

Modèle \ Mois	Janvier 2000
Escarpin	5
Ville	300

- Valeur HT des paires de chaussure par mois et par modèle

Modèle \ Mois	Janvier 2000
Escarpin	1 700 €
Ville	65 000 €

ETAPE 1 :

LES PROPRIÉTÉS DE LA TABLE *VENTE* (3)

- La société s'agrandit et dispose maintenant de plusieurs commerces.

Mois	Modèle	Magasin	Nombre	Total HT
Avril 2000	Botte	Champigny	10	1 500 €
Avril 2000	Tennis	Paris Bastille	850	260 000 €
.....				

- Ici aussi, une base relationnelle disposera d'une table **magasin**, avec un nom et une clé, plus d'autres caractéristiques telles que l'adresse.

ETAPE 1 : LES PROPRIÉTÉS DE LA TABLE *VENTE* (4)

- L'analyse s'avère déjà un peu plus complexe... la représentation tabulaire ne possède que deux dimensions.
- Si l'on veut étudier les performances d'un magasin, il faut faire une sélection sur celui-ci.
- Mais si on s'intéresse ensuite à ce qui s'est passé en février sur tous les magasins, il faut repartir des données de départ et faire une autre sélection.

ETAPE 1 : LES PROPRIÉTÉS DE LA TABLE VENTE (5)

- 6 types de reportings possibles :
 - Pour chaque magasin ou pour le total des magasins, les deux reportings précédents (nombre de paires de chaussures ou valeur HT avec les modèles en ligne et les mois en colonne),
 - Pour chaque modèle de chaussure ou pour le total, deux reportings avec les mois en ligne et les magasins en colonne,
 - Pour chaque mois ou pour l'année, deux reportings avec les magasins en ligne et les modèles en colonne.
- Ces six peuvent encore être multipliés par deux si l'on fait pivoter les axes des abscisses et des ordonnées.

ETAPE 1 :

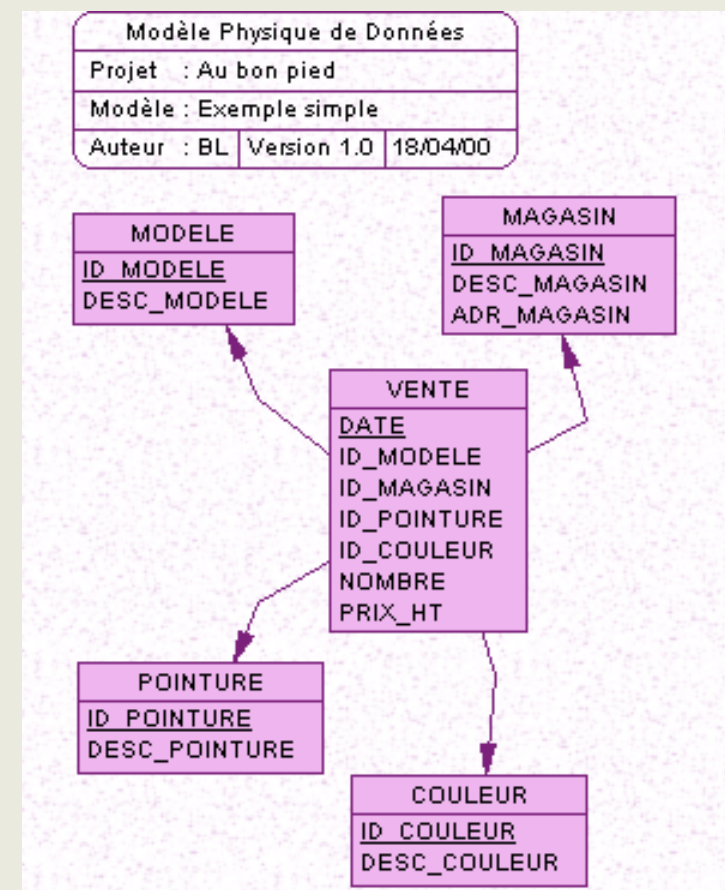
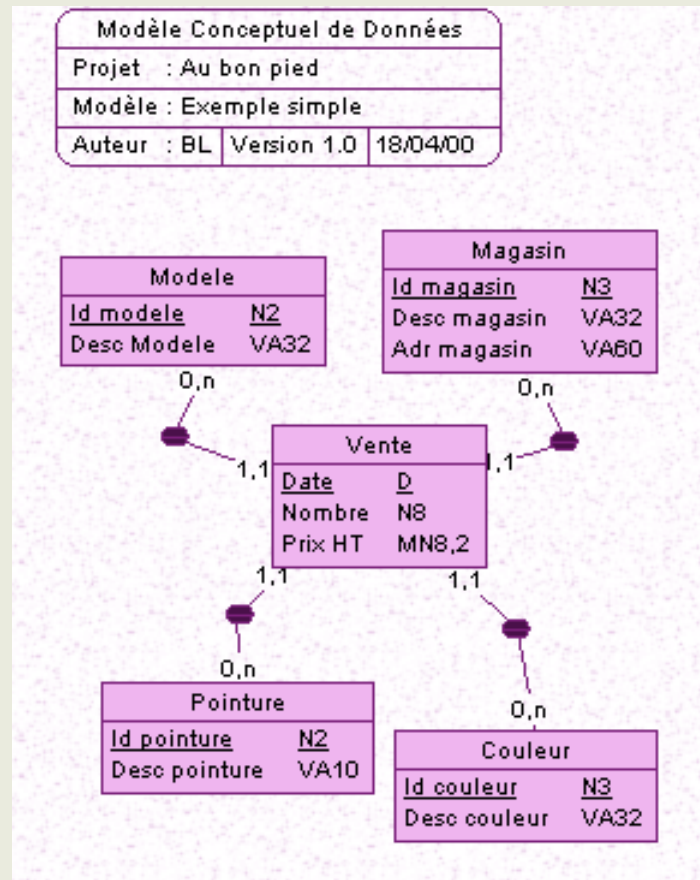
LES PROPRIÉTÉS DE LA TABLE VENTE (6)

- La société désire aussi étudier la répartition de ses ventes suivant d'autres critères, comme
 - Genre (Homme/Femme/Enfant),
 - Pointure,
 - Couleur,
 - ...
- Le tableau de reporting est devenu très important, à la fois en nombre de colonnes et de lignes.
- Pour analyser ces données, il faut faire des sommes sur des regroupements, avec des temps de traitement pas toujours envoûtants, il faut donc gérer des index, etc.

ETAPE 1 : LES PROPRIÉTÉS DE LA TABLE *VENTE* (7)

- Structure de données relationnelle construite : schéma en étoile
 - centre de l'étoile = table vente,
 - branches = tables magasin, modèle, couleur, etc.
= axes d'analyses

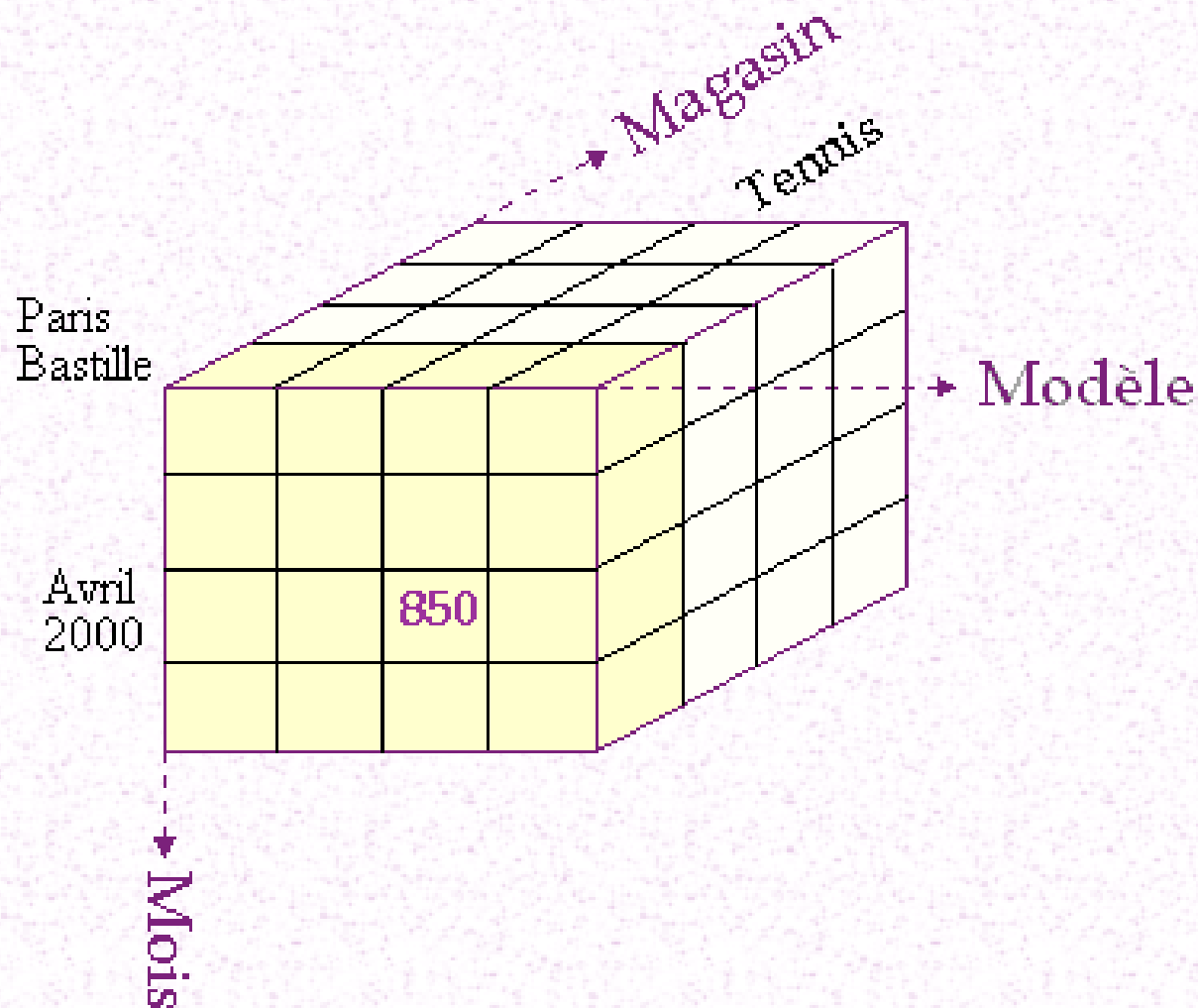
ETAPE 1 : LES PROPRIÉTÉS DE LA TABLE VENTE (8)



ETAPE 2 : LES TROIS DIMENSIONS DU *NOMBRE* (1)

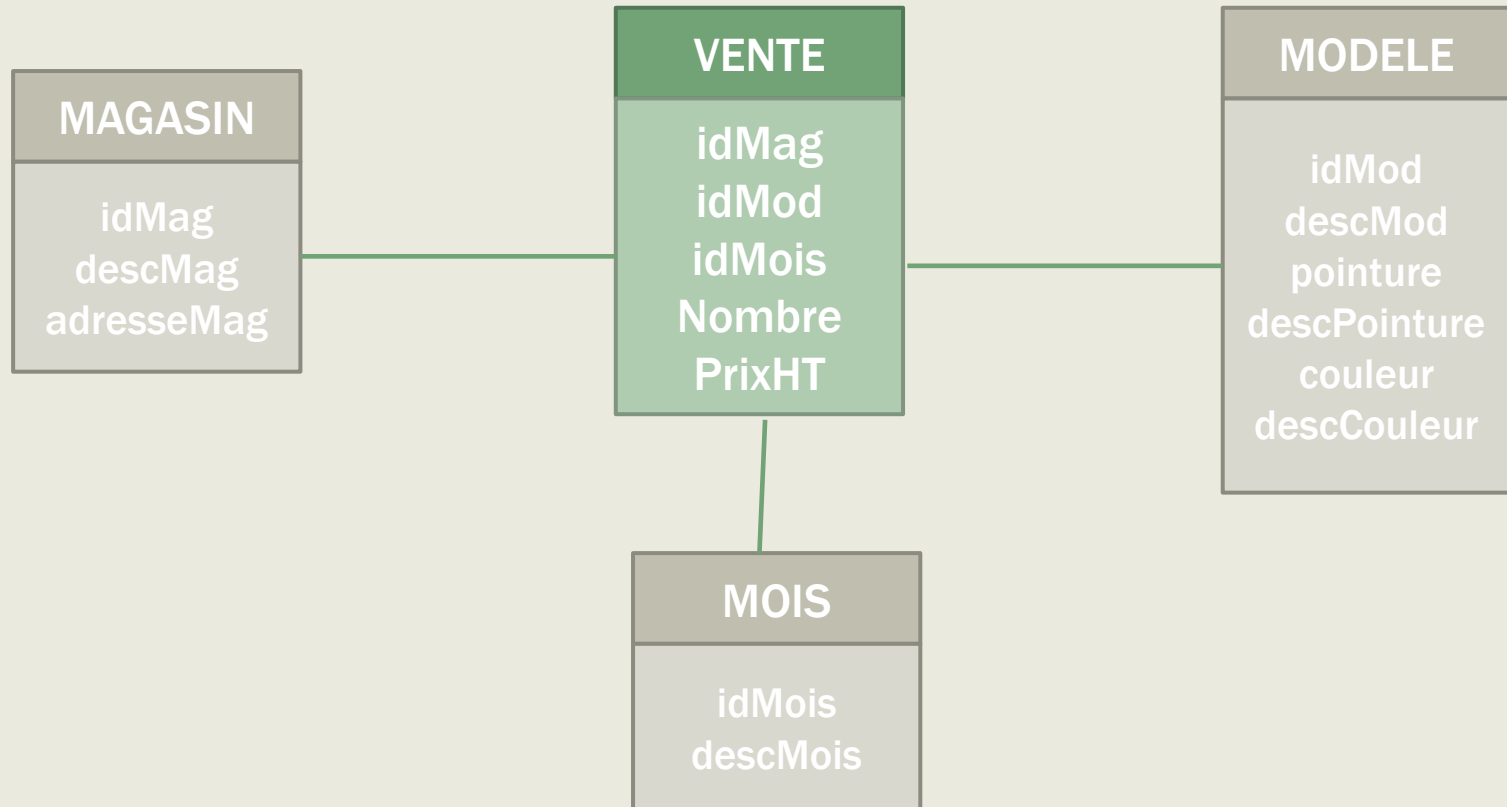
- la société *Au bon pied* désire donc suivre l'évolution de ses ventes de chaussures, par exemple en fonction des mois, modèle et magasin.
 - -> nos dimensions
- Les deux indicateurs suivis par les responsables de la société sont le nombre de chaussures vendues et les prix de vente hors taxe.
 - -> nos mesures
- Notre base de données est devenue multidimensionnelle.

ETAPE 2 : LES TROIS DIMENSIONS DU *NOMBRE* (2)



ETAPE 2 : LES TROIS DIMENSIONS DU *NOMBRE* (3)

■ Schéma en étoile

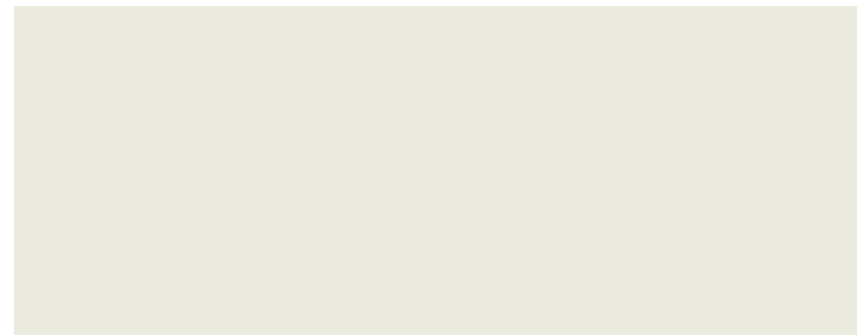


ETAPE 3 : LES HIÉRARCHIES SUR LA DIMENSION *TEMPS* (1)

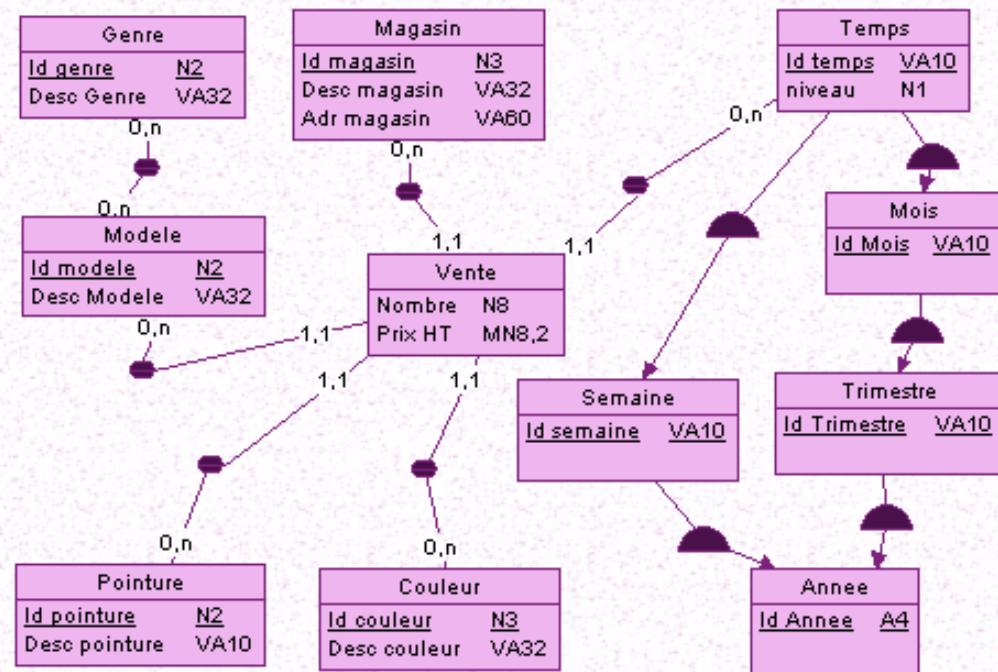
- L'étude des ventes mois par mois est certes utile, mais elle reste restrictive.
- Nous allons donc rebaptiser la dimension mois, et l'appeler par exemple *Temps*.
- Les positions de la dimension Temps peuvent être
 - des mois,
 - mais aussi des jours,
 - des trimestres
 - ou des années.
- Ainsi, les responsables des ventes pourront analyser les résultats des différents magasins sur des niveaux plus ou moins précis.

ETAPE 3 : LES HIÉRARCHIES SUR LA DIMENSION *TEMPS* (2)

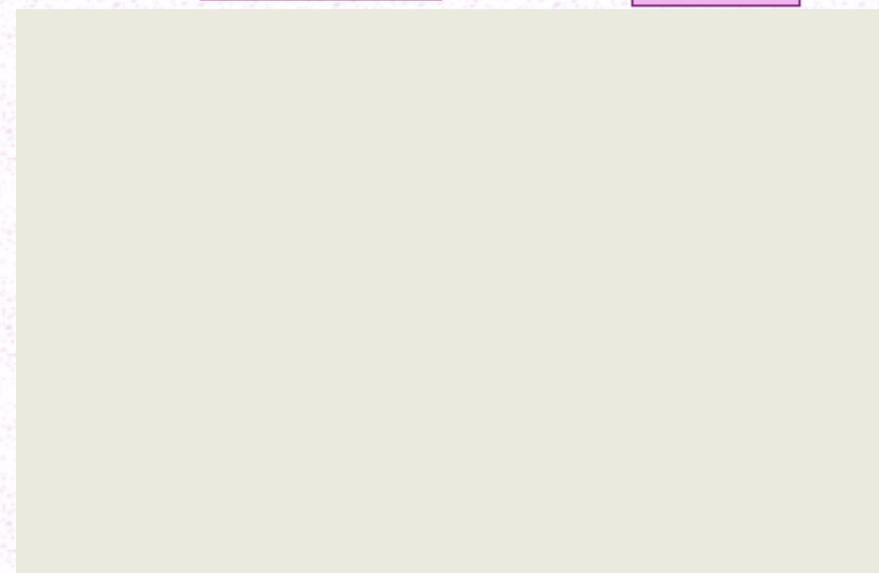
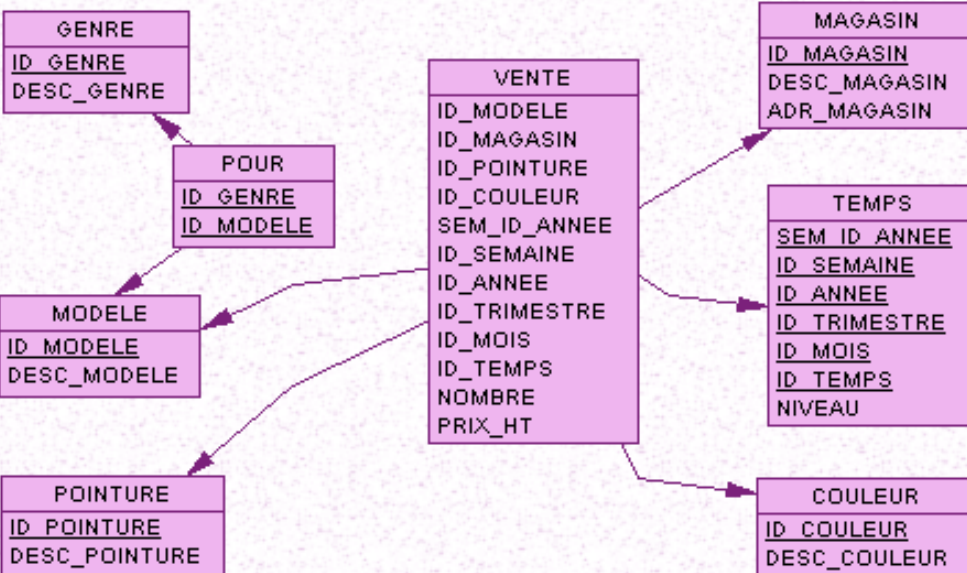
- Pour la dimension Magasin, nous avons une hiérarchie à 5 niveaux : idMag (descMag), ville, département, région et pays.
- Pour la dimension Temps, nous avons 2 hiérarchies :
 - La 1ère à 5 niveaux : idTps, jour, mois (nomMois), trimestre, année.
 - On pourra alors tout naturellement rattacher le 18/05/00 à mai 2000, puis à 2ème trimestre 2000, puis à 2000.
 - La 2ème à 4 niveaux : idTps, jour, semaine et année.
 - Ici, le 18/05/00 est rattaché à semaine 19 2000, puis à 2000.
- Pour la dimension Modèle, nous avons 3 hiérarchies :
 - La 1ère à 2 niveaux : idMod (descMod) et genre
 - La 2ème à 2 niveaux : idMod (descMod) et couleur
 - La 3ème à 2 niveaux : idMod (descMod) et pointure
- Le nouveau modèle relationnel associé est un "schéma en flocons".



Modèle Conceptuel de Données		
Projet : Au bon pied		
Modèle : Exemple simple		
Auteur : BL	Version 1.0	02/06/00



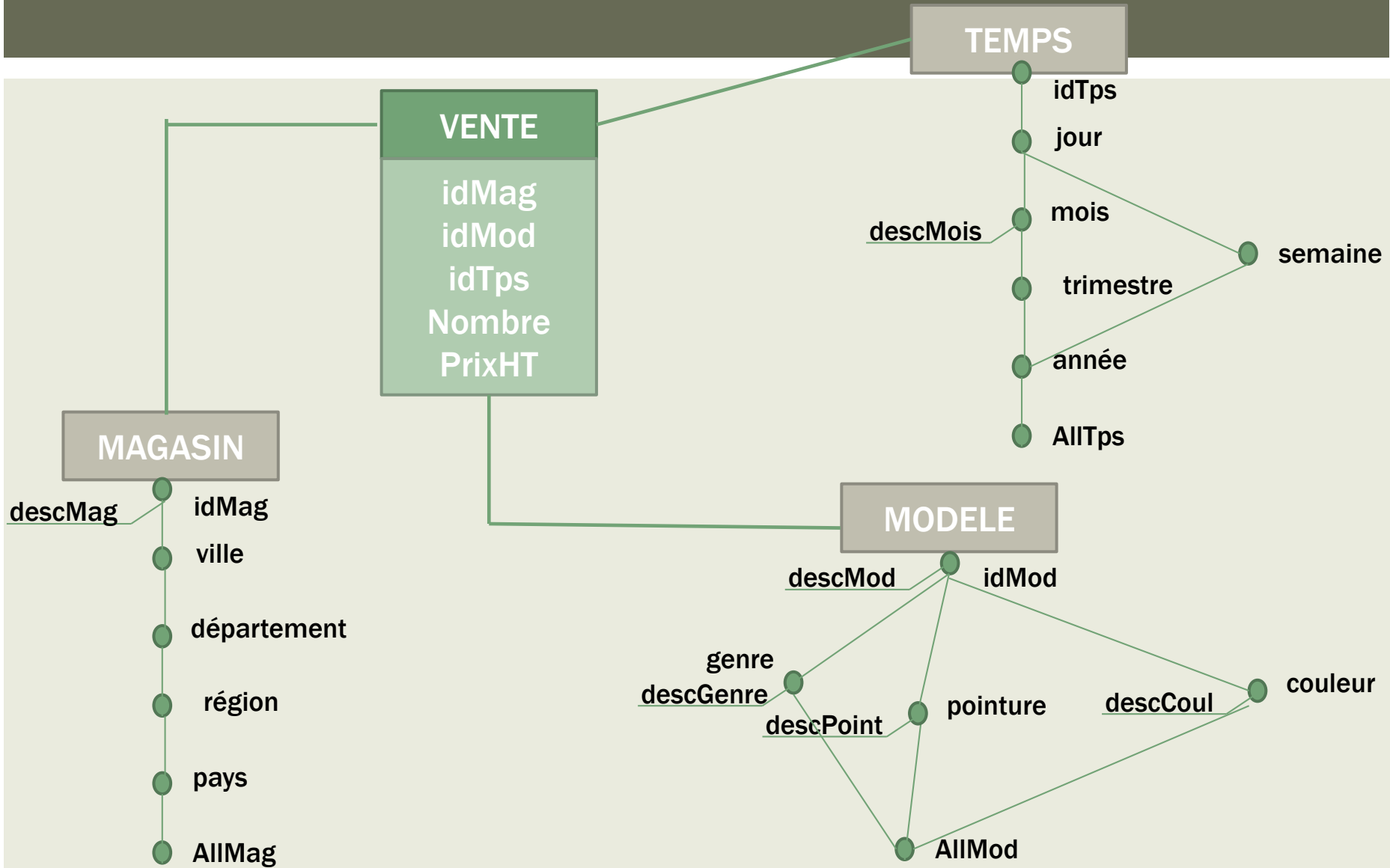
Modèle Physique de Données		
Projet : Au bon pied		
Modèle : Exemple simple		
Auteur : BL	Version 1.0	02/06/00



ETAPE 3 : LES HIÉRARCHIES SUR LA DIMENSION *TEMPS* (4)

- Les mesures Nombre et Total HT sont maintenant dimensionnées sur Temps, Modèle et Magasin.
- Chaque magasin fournit ses données quotidiennement, et la base de données OLAP se charge de consolider ces données sur les différentes hiérarchies.
- Lorsque cette consolidation consiste à effectuer des sommes, comme c'est le cas ici, des fonctions intégrées permettent de faire cette opération très rapidement et très simplement.

ETAPE 3 (5)



ETAPE 4 :

LES CALCULS À LA VOLÉE (1)

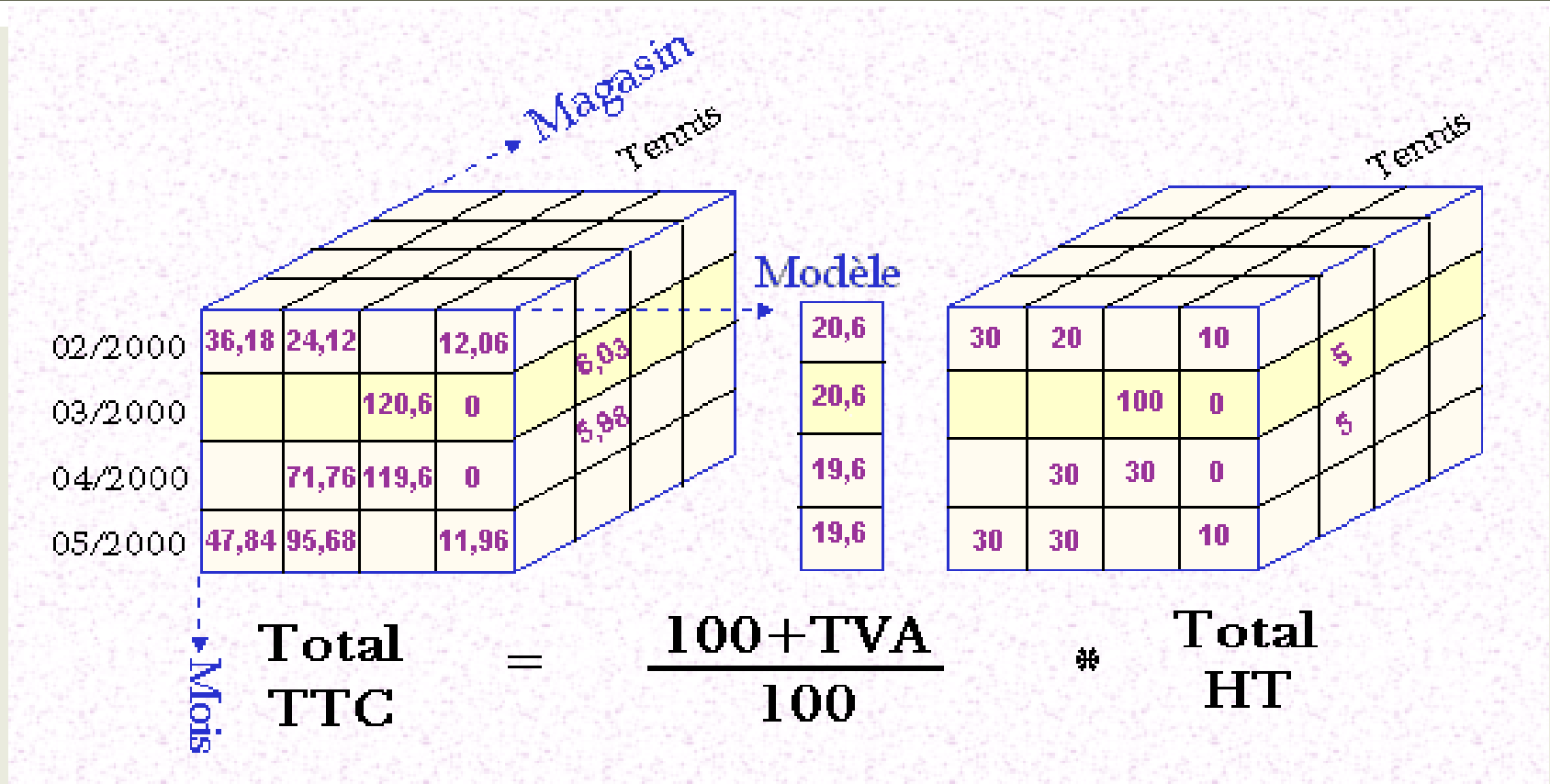
- Notre base OLAP dispose pour l'instant de deux mesures :
 - nombre de chaussures vendues
 - prix total hors taxe.
- Ces mesures sont dimensionnées par Temps, Modèle et Magasin.
- A partir de ces deux mesures, nous pouvons en créer d'autres, qui cette fois ne seront pas stockées dans la base de données mais calculées dynamiquement lorsqu'un utilisateur désire y accéder.

ETAPE 4 :

LES CALCULS À LA VOLÉE (2)

- Du point de vue de l'utilisateur, il n'y a aucune différence entre une mesure stockée et une mesure calculée.
- *Prix moyen* est dimensionnée suivant Temps, Modèle et Magasin et est de type décimal.
 - $\text{Prix moyen} = \text{Total HT} / \text{Nombre}$
- Il n'est pas nécessaire que tous les éléments du calcul soient dimensionnés de la même façon.
 - $\text{Total TTC} = \text{Total HT} * 1,196$
- Comme ce taux change (trop) souvent, il est possible de créer une nouvelle variable, *TVA*, dimensionnée seulement sur le temps.
 - Cette variable indiquera le taux de TVA à appliquer pour chaque position de la dimension Temps.
 - $\text{Total TTC} = \text{Total HT} * (100 + \text{TVA}) / 100$

ETAPE 4 : LES CALCULS À LA VOLÉE (3)



- De même, si le taux de TVA est différent suivant les produits vendus, alors les dimensions de la variable TVA doivent être Temps et Produit. Graphiquement, cette variable ne sera plus représentée par une ligne mais par un plan.

ETAPE 5 :

LES ATTRIBUTS (1)

- Savoir combien de chaussures ont été vendues, c'est bien, mais trouver celles qui se vendent le mieux, voilà qui va permettre à la société *Au bon pied* de prospérer.
- Les différents magasins vont devoir fournir des données plus précises sur leur ventes, en indiquant chaque jour le nombre de modèles vendus dans chaque référence disponible.
- Nous allons donc remplacer la dimension *Modèle* par la dimension *Référence*.
 - le magasin Paris Bastille a vendu 7 paires de chaussures de référence 215431324 le 18 Août 1998.

ETAPE 5 : LES ATTRIBUTS (2)

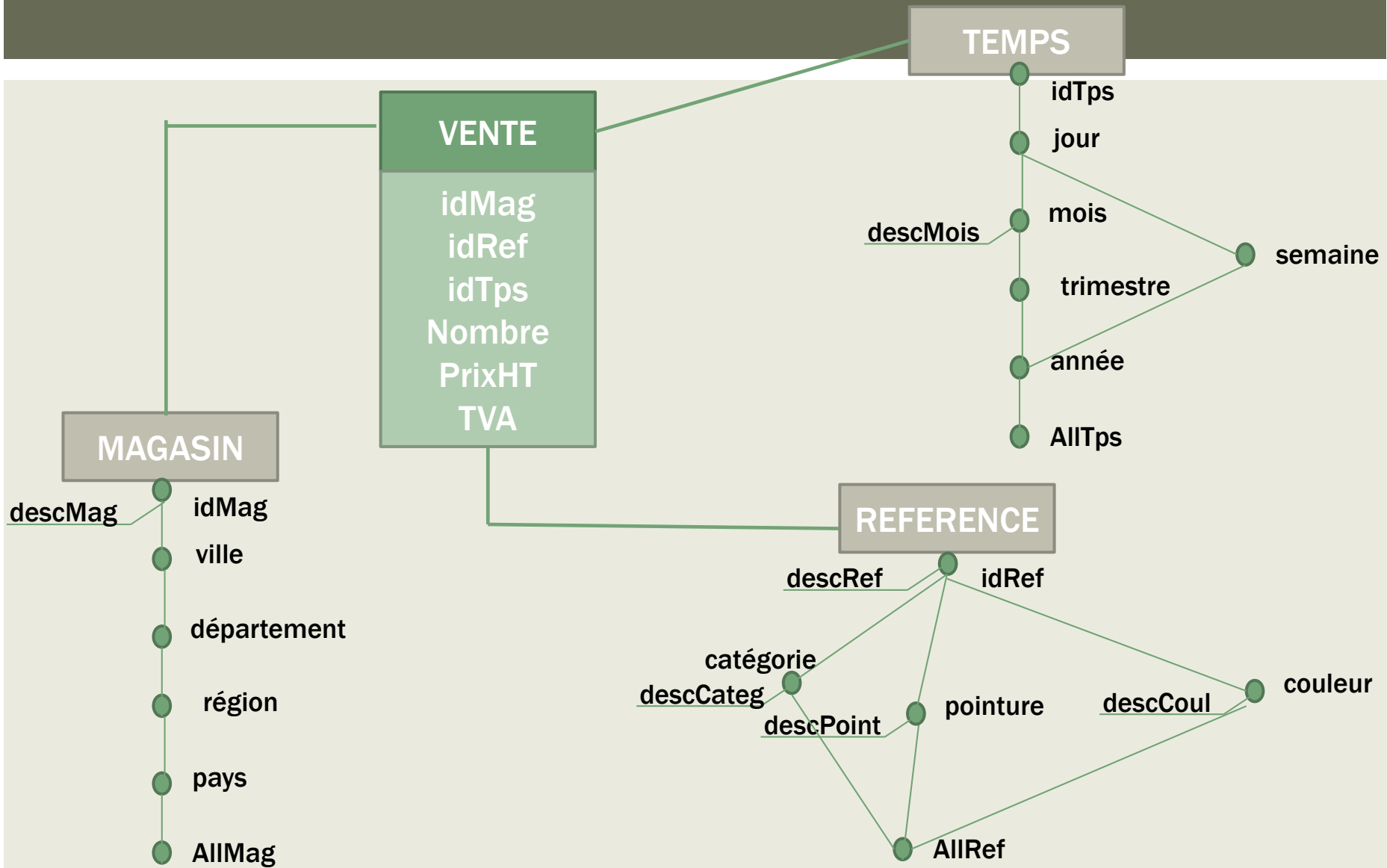
- Par contre, pour se rendre compte de l'état du marché, c'est plus difficile.
- Il suffit de définir des attributs pour chaque référence de chaussure.
 - Couleur (bleu, blanc, rouge),
 - Matière (cuir, toile, synthétique),
 - Catégorie (homme, femme, enfant).
- Si les références sont différentes pour chaque pointure d'un modèle, alors Pointure peut aussi être un attribut, sinon ce sera une nouvelle dimension.

ETAPE 5 :

LES ATTRIBUTS (3)

- notre base OLAP permet maintenant de répondre à des questions plus complexe comme :
 - Quelle couleur de chaussure se vend le plus en Août 1998 ?
 - Combien le magasin Paris Bastille a-t-il vendu de chaussures femme en cuir en 1998 ?
 - Quelle est la part de chaussures homme, femme et enfants chez *Au bon pied* ?

ETAPE 5 (4)



ETAPE 6 :

LES CALCULS TEMPORELS (1)

- Les exemples de calculs proposés dans les étapes précédentes ne suffisent plus à la société *Au bon pied*.
- Pour mieux cibler les actions à entreprendre, il faut en effet pouvoir étudier l'évolution des données déjà archivées, et si possible prévoir leur évolution dans le futur.
- Les bases OLAP sont tout à fait adaptées pour cela, elles disposent toujours de fonctions plus ou moins évoluées permettant de manipuler des informations temporelles.

ETAPE 6 :

LES CALCULS TEMPORELS (2)

- Par l'écriture de petits programmes simples, nous pouvons maintenant définir des formules répondant aux questions de ce type :
 - Quelle est l'évolution de mon bénéfice net par rapport au même mois de l'année précédente.
 - Quelle est l'évolution de mon chiffre d'affaire par rapport à la moyenne des trois derniers mois.
 - Quelle sera la tendance des 12 prochains mois suivant telle ou telle règle d'extrapolation.
- Pour répondre à des questions de ce type à l'aide d'une base relationnelle, il faudrait certainement écrire des programmes très complexes, ayant tellement de traitements et de données à analyser que l'utilisateur n'obtiendrait peut-être jamais sa réponse...

ETAPE 7 :

LES DONNÉES ÉPARSES (1)

- Ca y est ! La société *Au bon pied*, grâce à sa gestion rigoureuse, est maintenant très bien implantée, elle dispose en outre d'un choix de chaussures absolument impressionnant.
 - Pas moins de 12000 références sont disponibles,
 - et plus de 500 commerces ont ouvert leurs portes à travers le pays.
 - Chacun d'eux envoie quotidiennement ses chiffres de vente au siège de la société, qui est resté à Bonneuil.
 - Au bout d'un an, on obtient donc $12000 \times 500 \times 365$, soit plus de 2 milliard de cellules potentielles pour la variable "Nombre".
- Cependant, chaque magasin ne propose pas l'ensemble du catalogue à la clientèle.
 - En moyenne, environ 600 références sont disponibles dans chaque échoppe, qui n'est ouverte que 250 jours par an. I
 - I n'y aura donc réellement que $600 \times 500 \times 250$ cellules remplies, c'est à dire 75 millions de cellules (25 fois moins que la taille potentielle).

ETAPE 7 :

LES DONNÉES ÉPARSES (2)

- Les données ne sont pas réparties de manière aléatoire au sein des variables définies dans notre base OLAP.
 - Par exemple, le magasin "Paris Bastille" ne vend jamais de chaussure référencée "23154". Quelque soit le jour de l'année, la variable "Nombre" sera donc vide pour ce magasin et cette référence.
- Toutes les dimensions d'une variable n'ont donc pas la même importance.
- Afin d'optimiser la gestion de l'espace disque et l'accès aux données, il convient donc d'indiquer au moteur OLAP quelles sont les caractéristiques d'une variable (dense, éparse, ... -> dimension composite, dimension conjointe, ...).