



مخبر هندسة
الأنظمة المعقدة

LABORATOIRE D'INGÉNIERIE
DES SYSTÈMES COMPLEXES



Administration des bases de données avancées

Master I: Ingénierie des Logiciels Complexes (ILC)

Dr Kamilia MENGHOUR

Laboratoire d'Ingénierie des Systèmes Complexes

Université Badji Mokhtar-Annaba

K_menghour@yahoo.fr

Année Universitaire : 2024- 2025

Chapitre 2

Les modèles avancés de bases de données

Plan du cours

Chapitre 2 : *Les modèles avancés de bases de données (20%)*

- Les bases de données orientées objet.
- Bases de données actives.
- Bases de données déductives.
- Bases de données géographiques.
- Les entrepôts de données.

Introduction

Le système d'information décisionnel est un ensemble de données organisées de façon spécifiques, facilement accessibles et appropriées à la prise de décision.

Les systèmes de gestion sont dédiés aux métiers et management de l'entreprise.

Pourquoi un entrepôt de données? Améliorer les performances décisionnelles de l'entreprise

Comment? En répondant aux demandes d'analyse des décideurs

Exemples :

- ✓ **Clientèle:** Qui sont mes clients? Pourquoi sont-ils mes clients? Comment les conserver ou les faire revenir (préférence d'achat, habitudes, ...)? Ces clients sont-ils vraiment intéressants pour moi?
- ✓ **Marketing,** actions commerciales: où placer ce produit dans les rayons? Comment cibler plus précisément le mailing concernant ce produit?

Objectifs pour les DWs

L'objectif du data warehouse est de permettre des requêtes sur de grands ensembles des données, la plupart du temps sous forme d'agrégats (GROUP BY) afin d'en obtenir une vision synthétique (propre à la prise de décision).

Les DWs permettent:

- ✓ Accessibilité des informations
- ✓ Cohérence des informations
- ✓ Adaptation au changement
- ✓ Présentation des informations en temps opportun
- ✓ Sécurité des informations
- ✓ Fiabilité des informations

L'entrepôt de données ou Data warehouse

« une collection de données **orientées sujet, intégrées, non volatiles et historisées, organisées** pour le support d'un processus d'aide à la décision»

Inmon (1996)

Définition d'un entrepôt de données (data warehouse)

Orientées sujets : le DW se concentre sur les besoins analytiques des différents secteurs d'une organisation.

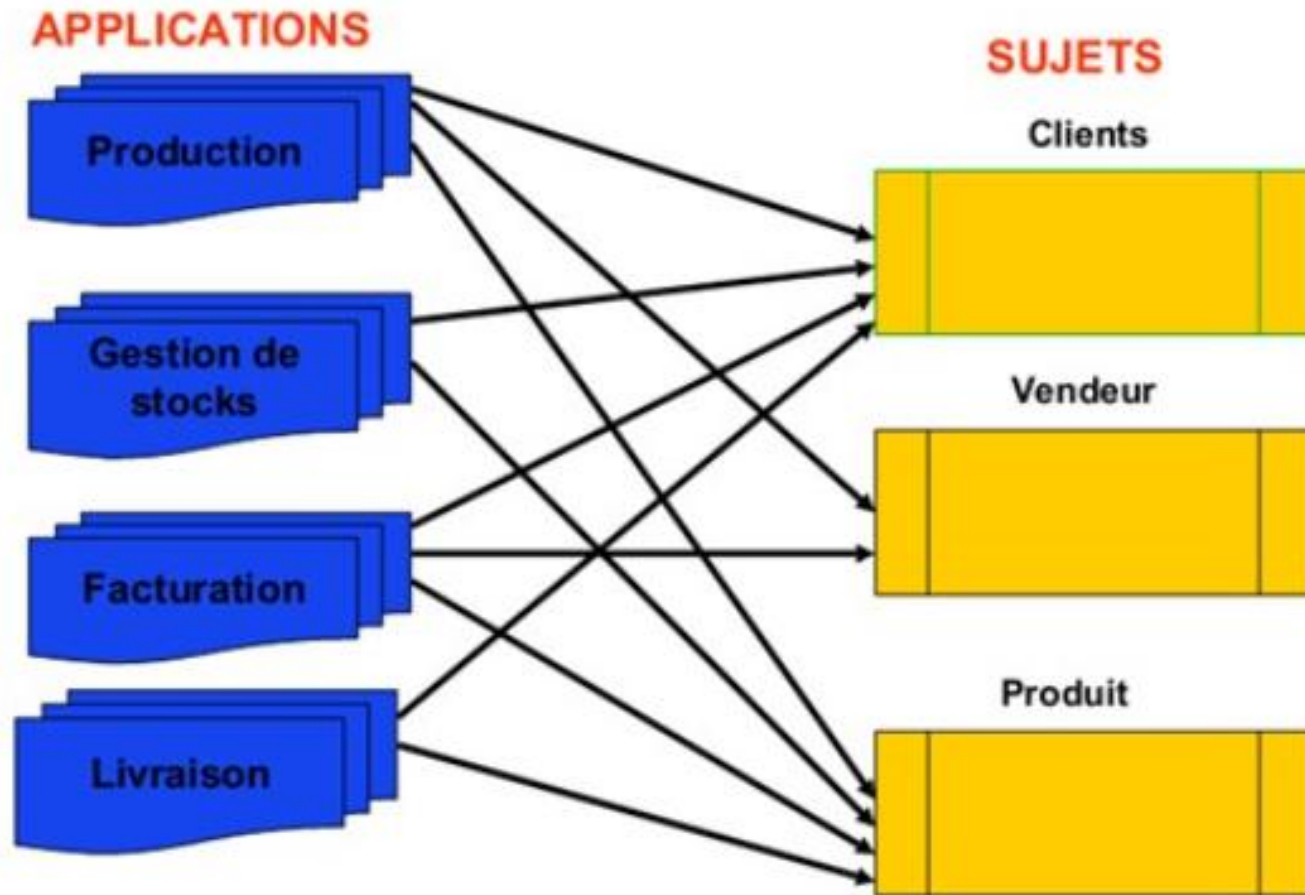
- Organisé autour d'un sujet bien précis
- S'intéresse à la modélisation et l'analyse des données pour aider les décideurs, non pas pour des activités quotidiennes ou traitement transactionnel
- Fournit une vue simple concernant un sujet particulier en excluant les données qui ne servent pas à la prise de décision

Exemple: Pour une compagnie d'assurance :

les applications peuvent concerner le *traitement de l'automobile, de la vie, de la santé et les accidents* alors que les principaux sujets peuvent être *les clients, les polices d'assurance, les primes et les*

Définition d'un entrepôt de données (data warehouse)

Exemple:



Définition d'un entrepôt de données (data warehouse)

Données Intégrées :

- Les données dans un DWH sont chargées de différentes sources contenant des données sur différents formats.
- Les données doivent être vérifiées, triées et transformées dans un format unifié afin de faciliter et accélérer l'accès.
- Résoudre les problèmes dus aux différences de définition et de contenu des données telles que
 - ✓ les différences de format et de codification des données,
 - ✓ les synonymes (champs avec des noms différents mais les mêmes données),
 - ✓ homonymes (champs avec le même nom mais différentes significations),
 - ✓ la multiplicité des occurrences.

Définition d'un entrepôt de données (data warehouse)

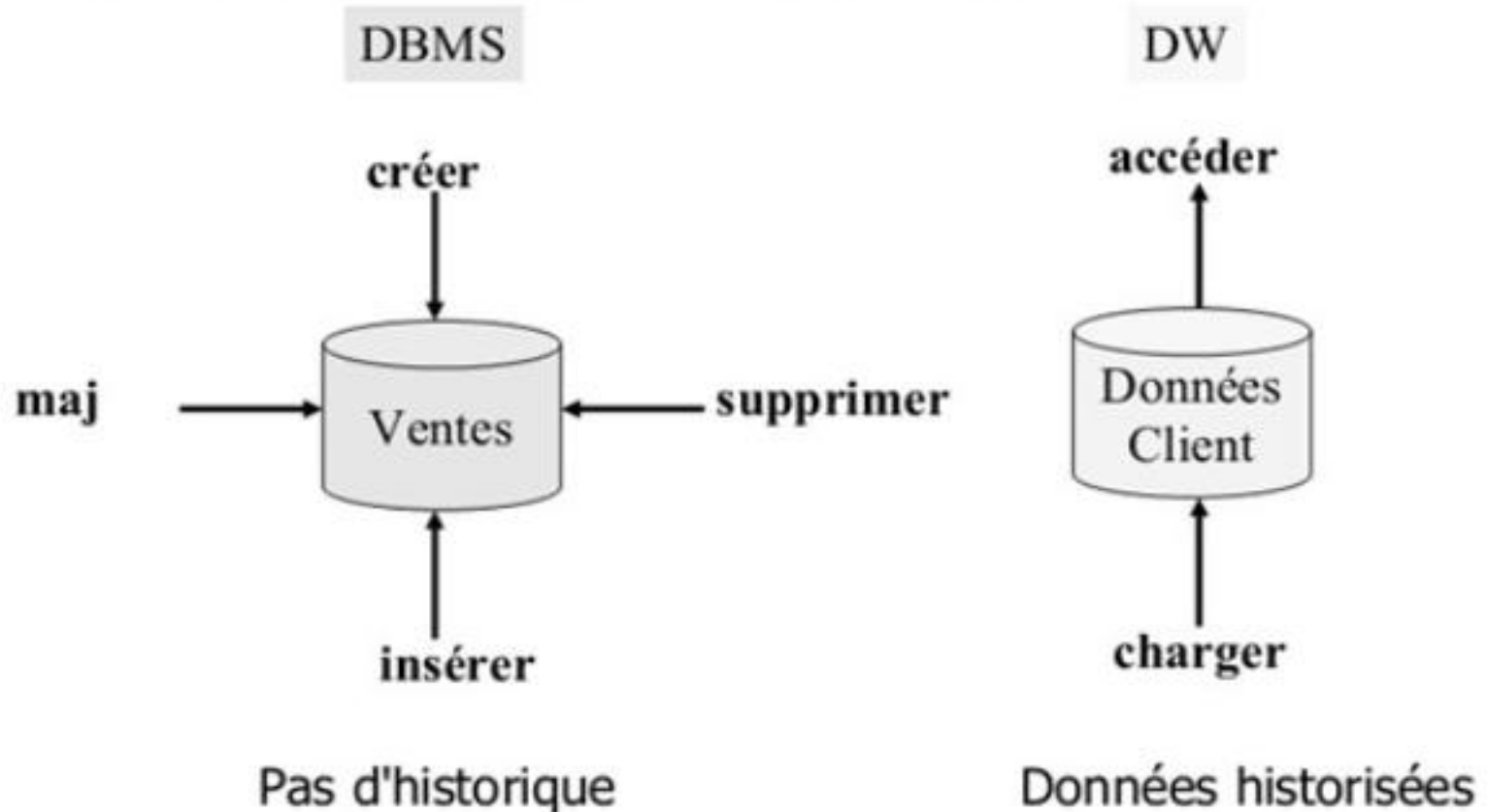
Non volatiles :

- les données sont
 - ✓ stables
 - ✓ en lecture seule: non modifiables et non supprimables
- Afin de conserver la traçabilité des informations et des décisions prises, les informations stockées au sein du DW ne doivent pas disparaître...
- Le seul changement provient du chargement de nouvelles données

Définition d'un entrepôt de données (data warehouse)

Non volatiles :

Non volatiles : traçabilité ⇒ non suppression

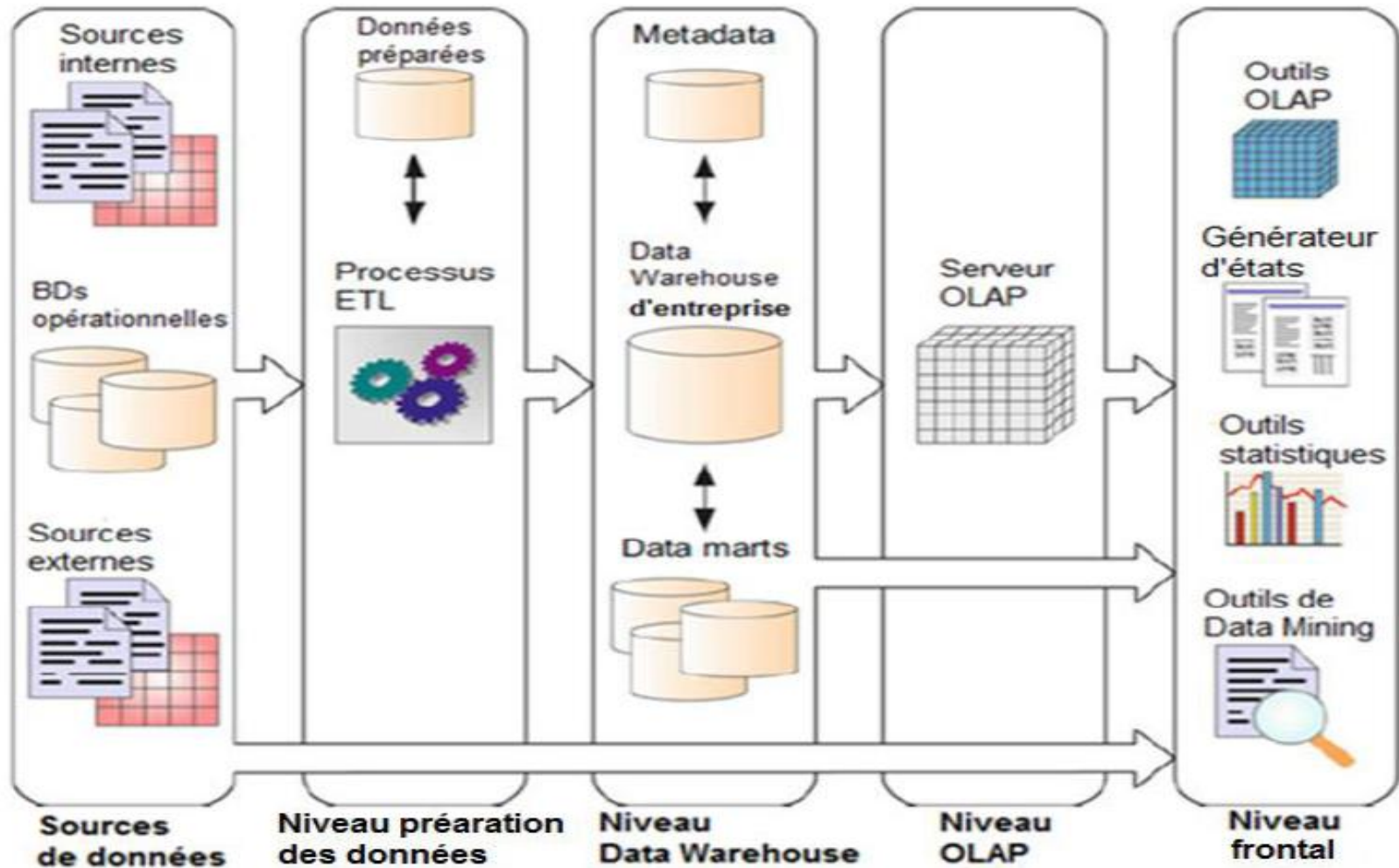


Définition d'un entrepôt de données (data warehouse)

Données Historiées :

- et donc datées :
- avec une conservation de l'historique et de son évolution
- pour permettre les analyses comparatives (par exemple, d'une année sur l'autre, etc.).
- Dans un DW, un référentiel de temps est nécessaire : c'est l'axe temps ou l'axe période.

Architecture des DWHs



Architecture typique d'un DW (Vaisman 2014)

Le niveau préparation des données

A ce niveau, les données sont extraites à partir des BDs opérationnelles qui peuvent être des BDRs, BDOOs, BDROs ou BDDs et d'autres sources de données (internes ou externes de l'organisation).

Ces données sont ensuite intégrées dans le data Warehouse.

La préparation des données est réalisée par les outils (ETL) : Extraction, Transformation et Loading (chargement).

Une base de données intermédiaire, généralement relationnelle (BDR), est dédiée à l'exécution de l'extraction et la transformation des données avant leur chargement dans le Data Warehouse. De telle base de données est appelée opérationnel.

Le niveau préparation des données

- a) Extraction:* à partir de multiple sources de données hétérogènes, internes ou externes à l'organisation. Ces sources peuvent être opérationnelles ou des fichiers avec différents formats.
- b) Transformation:* plusieurs transformations peuvent être appliquées telles que le nettoyage (conversion en format standardisé et suppression des erreurs et des incohérences dans les données telles que les erreurs d'orthographe et les conflits de domaine), combinaison des données provenant de plusieurs sources et suppression des doublons.
- c) Loading (chargement):* Les données transformées sont alors chargées dans le DW. Cela inclut également la propagation des mises à jour des sources de données vers le DW à une fréquence qui peut être, selon la politique de l'organisation, mensuelle, plusieurs fois par jour ou presque en temps réel.

Le niveau Data Warehouse

A ce niveau, il y a trois composantes importantes à savoir :

Data Warehouse d'entreprise: il est centralisé et couvre toute l'organisation

Datamart: par contre, le Datamart est spécifique à une fonction ou un département dans l'organisation. Le Datamart est chargé par des données (déjà nettoyées et intégrées) du Data Warehouse d'entreprise.

Metadata: les métadonnées sont définies comme des « données sur les données » ou « les données relatives aux données ».

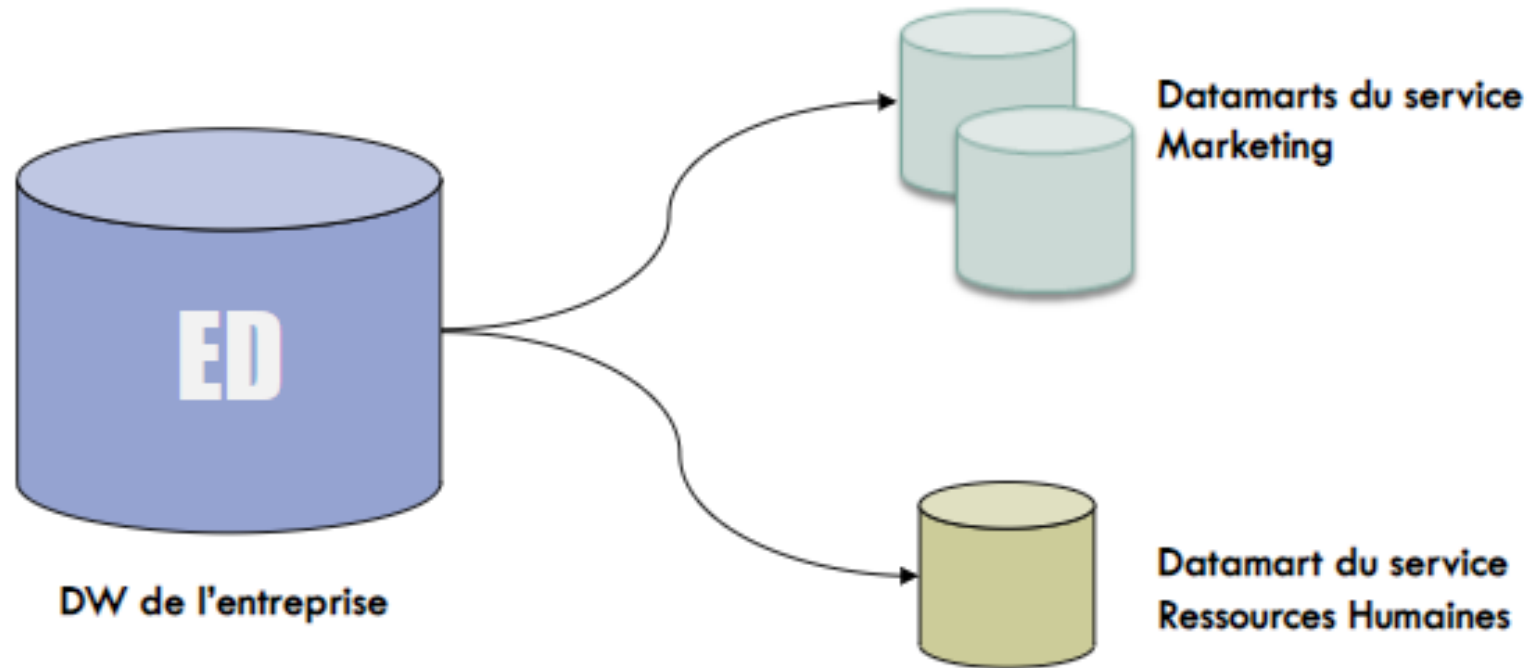
Data warehouse et data mart

Un **data warehouse** et un **data mart** se distinguent par la partie qu'il recouvre :

- **Le data warehouse** recouvre l'ensemble des données et problématiques d'analyse visées par l'entreprise.
- **Le data mart** recouvre une partie des données destinée à répondre aux besoins et problématiques liées à un secteur ou d'une fonction particulière de l'entreprise
- d'un Point de vue spécifique selon des critères métiers
- Datamarts du service Marketing

Un **data mart** est un sous-ensemble du data warehouse de l'entreprise, obtenu par extraction et agrégation des données de celui-ci.

Data warehouse et data mart



Pourquoi les data marts

Les data marts sont destinés à pré-agrégéer des données disponibles de façon plus détaillée dans les **data warehouse**, afin de traiter plus facilement certaines questions spécifiques, critiques, etc.

- ✓ un nouvel environnement structuré et formaté en fonction des besoins d'un métier ou d'un usage particulier
- ✓ Moins de données que DW
- ✓ Plus facile à comprendre, à manipuler
- ✓ Amélioration des temps de réponse
- ✓ Utilisateurs plus ciblés: DM plus facile à définir

Architecture des DWHs

Le niveau OLAP : Ce niveau est composé d'un serveur OLAP fournissant aux utilisateurs professionnels des données multidimensionnelles à partir du *Data Warehouse* ou des *Datamarts*.

Le niveau frontal : Afin de permettre aux clients d'exploiter les données du DW, le niveau interface leur fournit les outils suivants :

- **Outils OLAP** : facilitent la formulation des requêtes complexes impliquant de grandes quantités de données.
- **Générateurs d'états** : produisent des rapports en utilisant des requêtes prédéfinies.
- **Outils statistiques** : utilisent des méthodes statistiques dans l'analyse et la visualisation des données.
- **Outils Data Mining** : permettent d'analyser les données dans le but d'établir un modèle afin d'aider à la prédiction.

Base de données VS entrepôt de données

Pourquoi dissocier une BD d'un DW?

- ✓ Les objectifs de performances dans les BD ne sont pas les mêmes que ceux dans les DWs
 - *BD : requêtes simples, méthodes d'accès et d'indexation*
 - *DW: requêtes OLAP souvent complexes !!!*
- ✓ La nécessité de combiner des données provenant de diverses sources, d'effectuer des agrégations dans un DW et d'offrir des vues multidimensionnelles
- ✓ Les données d'un DW sont souvent non volatiles et ont donc une plus longue durée de vie que celles d'une BD.

Récapitulatif

Caractéristiques	Bases de données	Entrepôts de données
Opération	Gestion courante	Support à la décision
Modèle	Surtout entité-relation	Étoile, flocon de neige
Normalisation	Plus fréquente	Rare
Données	Actuelles, brutes	Historiques, agrégées
Mise à jour	Immédiate	Plutôt différée
Consolidation	Très faible	Très élevée
Perception	Bidimensionnelle	Multidimensionnelle
Opérations	Lecture et écriture	Lecture et rafraîchissement
Taille	Des giga-octets	Plutôt des téra-octets

Modèle multidimensionnel

Le modèle multidimensionnel est bien adapté aux requêtes telles que :

- quel est le montant total des recettes enregistrées l'an dernier par état et par catégorie de produit ?
- quelles commandes maximisent les recettes ?
- Utiliser les langages traditionnels tels que SQL pour exprimer ces types de requêtes est une tâche difficile.
- L'exécution de telles requêtes sur des BDs opérationnels prendra un temps de réponse inacceptable.
- les SGBDs opérationnels ne peuvent assurer une analyse de données basée sur l'exécution de requêtes complexes nécessitant la jointure de beaucoup de tables et l'agrégation de larges volumes de données incluant aussi leurs historisation.

Modèle multidimensionnel

Les systèmes OLAP sont des systèmes orientés vers l'analyse de données dans le but d'aide à la décision.

Ils se concentrent sur les requêtes analytiques qui exigent l'agrégation, beaucoup de temps, la lecture transversale des enregistrements des tables et de nouvelles techniques d'indexation.

Le Data Warehouse est défini dans le but de répondre aux besoins à des solutions qui peuvent supporter ces techniques.

Data Warehouse et OLAP se basent sur le modèle multidimensionnel qui visualise les données dans un espace de n-dimensions appelé les données cube.

Les données cube sont définies par les faits et les dimensions.

Modèle multidimensionnel: Cube de données

Un cube est une structure de données multidimensionnelles définie dans le but de capturer et analyser des données.

Un cube généralise la feuille de calcul d'un tableur avec une dimension supérieure à deux.

Un cube peut avoir une dimension supérieure à trois contrairement à ce que son nom peut signifier. Pour cette raison le mot hypercube est parfois utilisé à la place de cube.

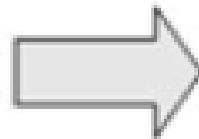
Une collection de cubes constitue une base de données multidimensionnelle ou un Data Warehouse multidimensionnel.

Un data cube, ex: ventes, permet de voir les données selon plusieurs dimensions

Modèle multidimensionnel: Cube de données

Table relationnelle

produit	région	vente
écrou	Est	50
écrou	Ouest	60
écrou	Centre	110
vis	Est	70
vis	Ouest	80
vis	Centre	90
boulon	Est	120
boulon	Ouest	10
boulon	Centre	20
joint	Est	50
joint	Ouest	40
joint	Centre	70

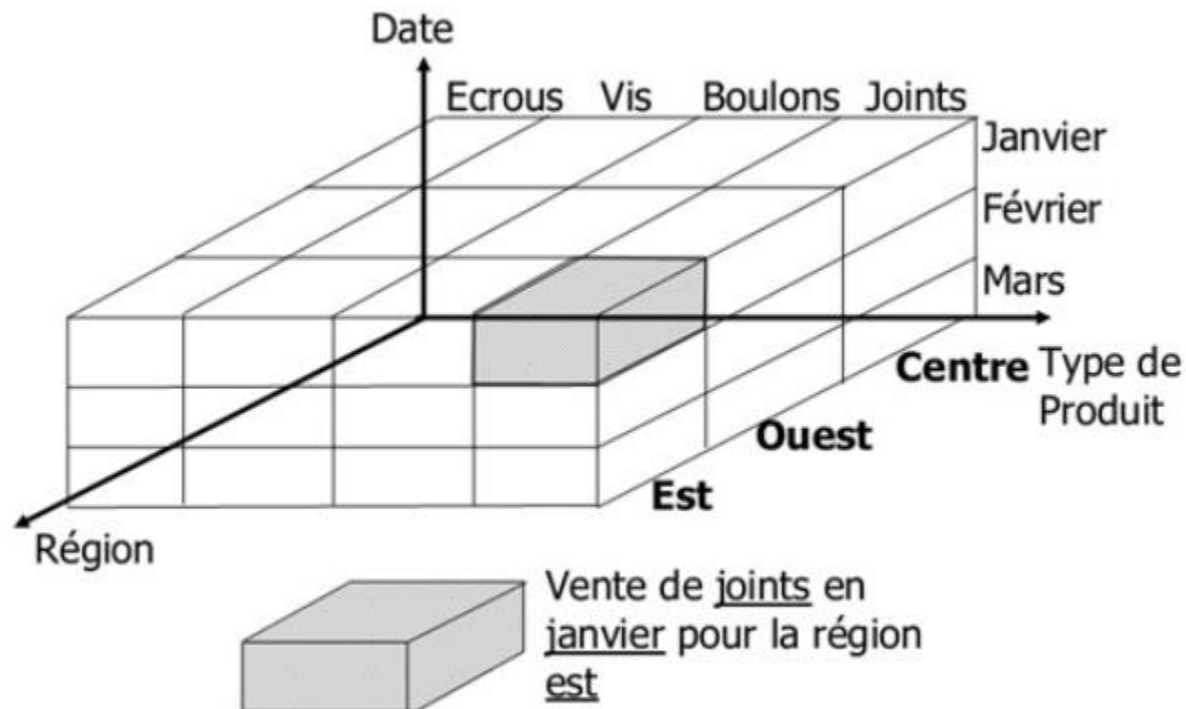


Représentation multidimensionnelle

	Est	Ouest	Centre
écrous	50	60	110
vis	70	80	90
boulons	120	10	20
joints	50	40	70

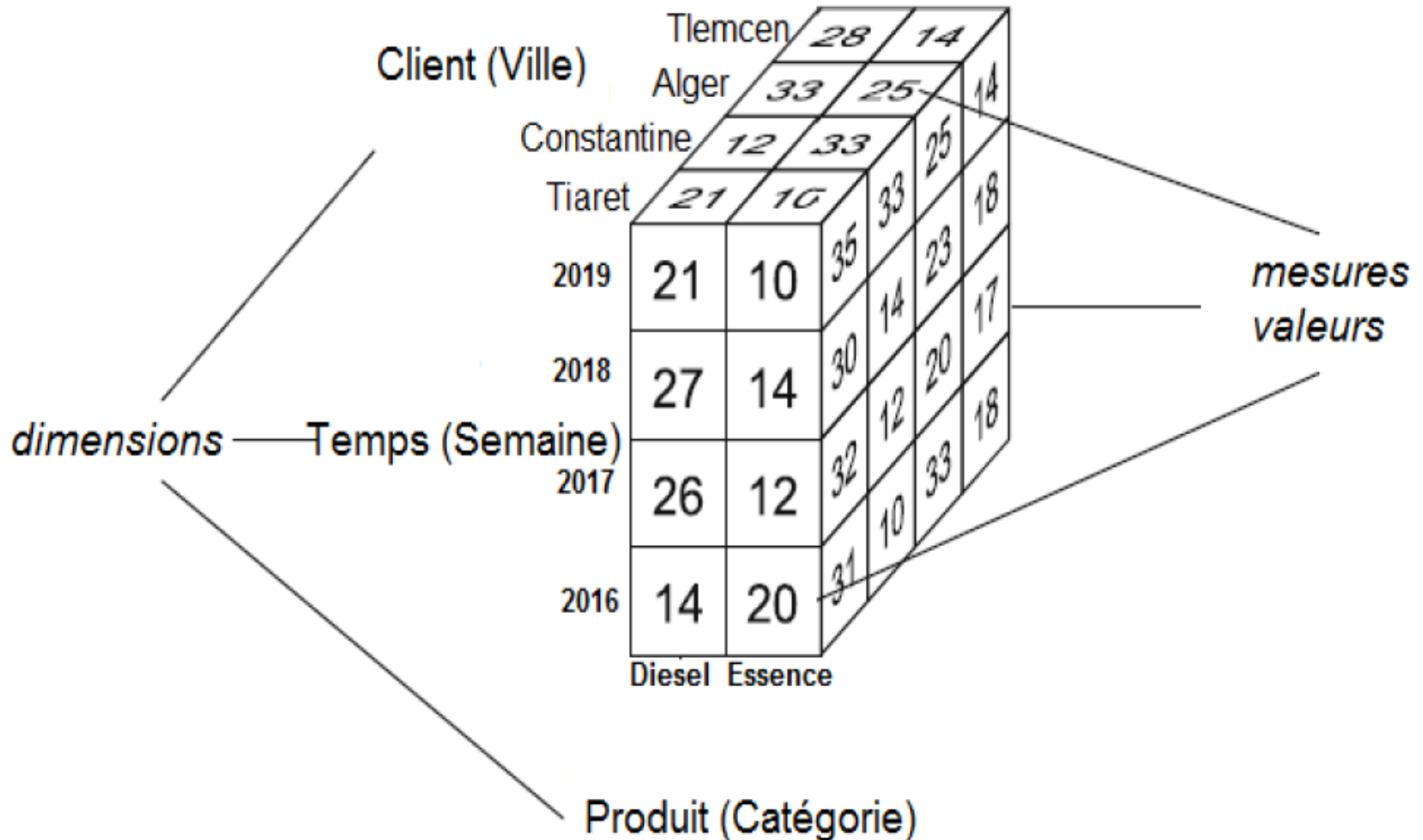
Modèle multidimensionnel: Cube de données

Représentation multidimensionnelle: cube de données



Modèle multidimensionnel: Cube de données

Exemple: Un cube en trois dimensions pour les données de vente avec les dimensions : *Produit*, *Temps* et *Client*.



Modèle multidimensionnel: Cube de données

Un cube est composé de cellules où chacune d'elles représente une intersection de dimensions.

Une cellule quand elle est non vide est appelée un **fait** et quand elle est vide cela signifie qu'il n'y a aucune information à enregistrer pour les valeurs de dimensions données.

Dans l'exemple: il y a une seule mesure à savoir le nombre de ventes pour chaque fait qui représente une combinaison d'une année, de type d'énergie de voiture, et de ville.

Par exemple: *le total de ventes dans la ville de 'Tiaret' en 2019 est 31 en additionnant les deux nombres 21 et 10.*

Modèle multidimensionnel: Dimensions

Les dimensions sont utilisées pour la sélection des données et leur l'agrégation.

Le niveau de la dimension est la granularité (le niveau du détail) à laquelle (auquel) les mesures représentent les dimensions.

Dans l'exemple précédant:

les ventes sont agrégées aux niveaux *Catégorie, année et Ville* respectivement pour les dimensions *Produit, Temps et Client*.

Les instances d'une dimension sont appelées membres.

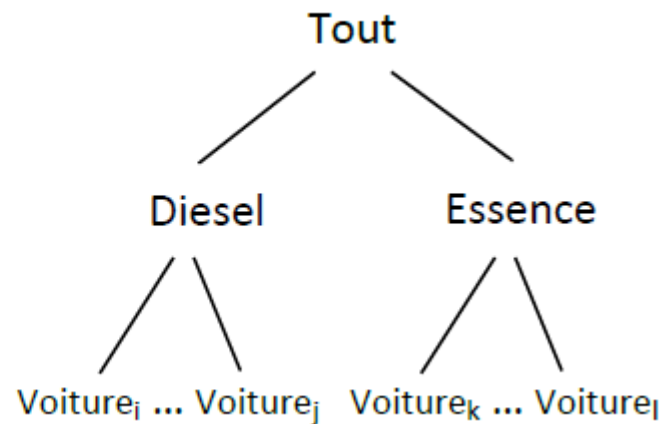
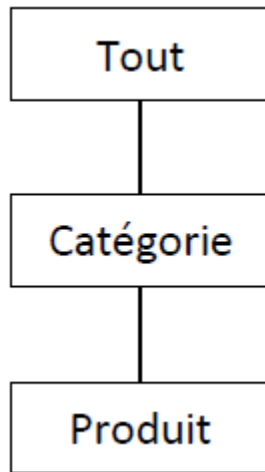
Par exemple, Diesel et Essence sont des membres de la dimension Produit au niveau Catégorie.

Les dimensions ont aussi des attributs associés qui les décrivent. Par exemple, la dimension *Produit* doit avoir les attributs *NumProduit* et *Prix* qui ne sont pas visible sur la figure précédente.

Modèle multidimensionnel: Dimensions

Les dimensions sont organisées hiérarchiquement en niveaux de détail.

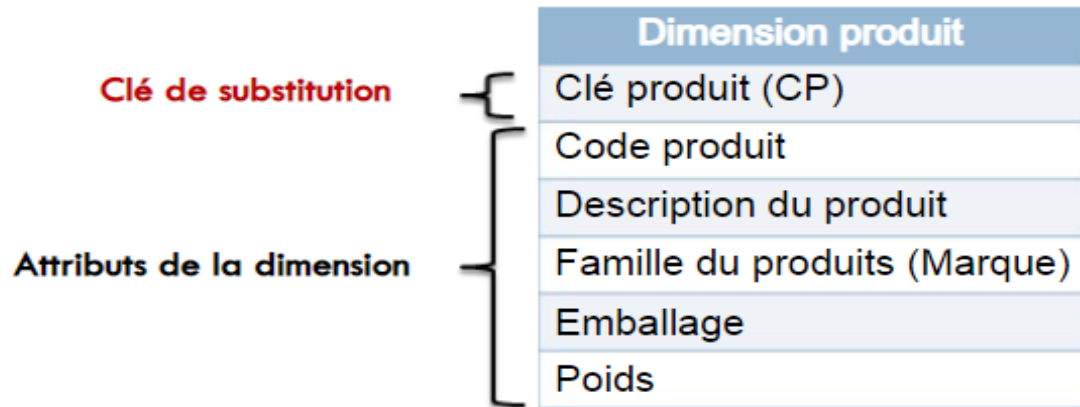
Par exemple, pour la dimension *Produit*, la granularité la plus faible est *Produit*, qui est agrégée en *Catégorie*



La hiérarchie de la dimension est définie dans un Metadata au niveau du cube ou au niveau du modèle multidimensionnel, si la dimension peut être partagée.

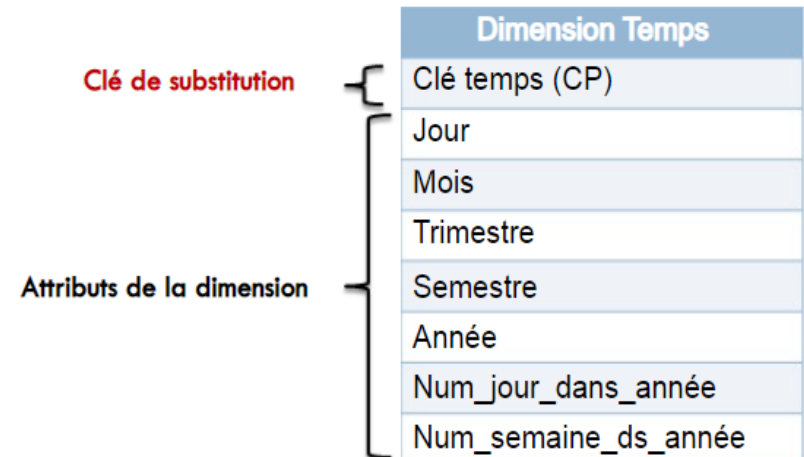
Modèle multidimensionnel: Dimensions

Table de dimension :Axe d'analyse selon lequel vont être étudiées les données observables (faits) Contient le détail sur les faits



La dimension Temps:

- Commune à l'ensemble des DW;
- Reliée à toute table de faits

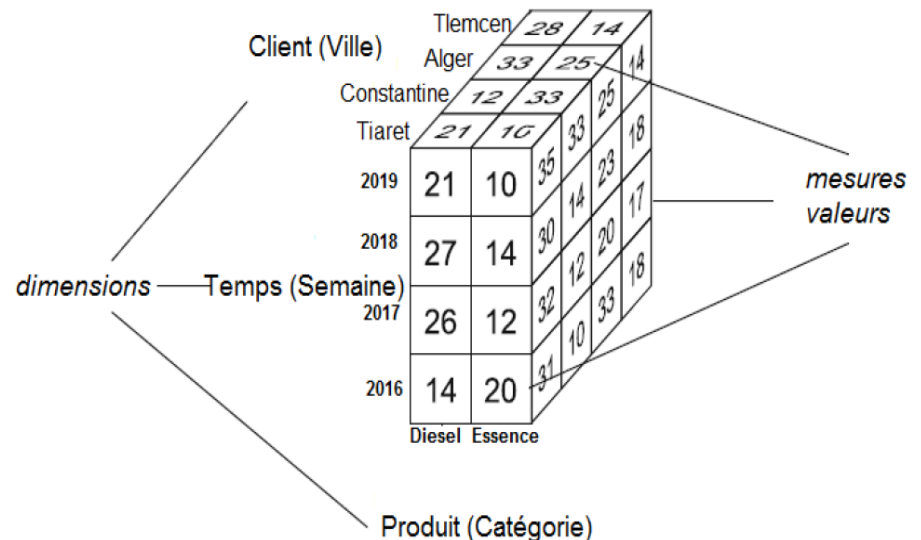


Modèle multidimensionnel: Faits

Les faits sont les objets des sujets à être analysés dans le but de comprendre leurs comportements tels que les ventes dans l'exemple précédant.

Les faits sont définis par leur combinaison de valeurs des dimensions. Donc, les faits existent dans les cellules non vides.

Dans l'exemple, chaque nombre affiché dans la cellule représente le nombre de voitures vendues par catégorie, par année et par ville du client.

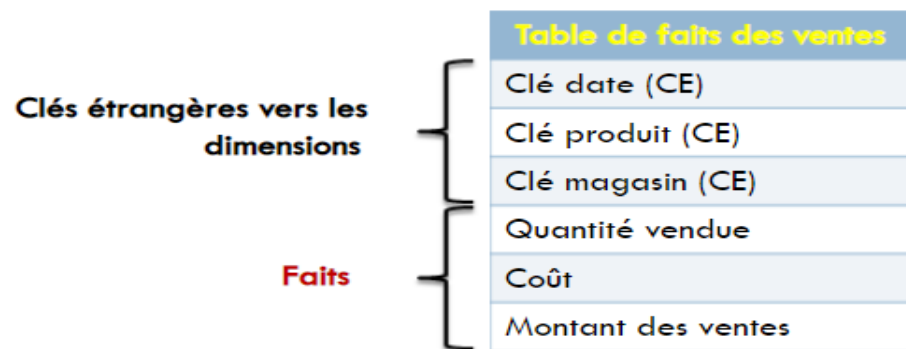


Modèle multidimensionnel: Faits

Table des faits

Table principale du modèle dimensionnel

Contient les données observables (les faits) sur le sujet étudié selon divers axes d'analyse (les dimensions)



Faits: Ce que l'on souhaite mesurer: Quantités vendues, montant des ventes... Contient les clés étrangères des axes d'analyse (dimension) Date, produit, magasin

Modèle multidimensionnel: Faits

Trois types de faits: Additif, Semi additif et Non additif

Additif: additionnable suivant toutes les dimensions

- ✓ Quantités vendues, chiffre d'affaire
- ✓ Peut être le résultat d'un calcul:
 - Bénéfice = montant vente – coût

Semi additif: additionnable suivant certaines dimensions

- ✓ Solde d'un compte bancaire:
 - Pas de sens d'additionner sur les dates car cela représente des instantanés d'un niveau
 - Σ sur les comptes: on connaît ce que nous possédons en banque

Non additif: fait non additionnable quelque soit la dimension

- ✓ Prix unitaire: l'addition sur n'importe quelle dimension donne un nombre dépourvu de sens .

Stratégies d'implantation d'un DW

On-line Analytical Processing (OLAP): analyse interactive des données dans un environnement DW. Trois stratégies :

Utilisation d'un SGBD Relationnel (systèmes ROLAP)

- ✓ SGBDR : Nécessite des adaptations pour répondre aux besoins des DWs
- ✓ Stockage des données dans un SGBDR
- ✓ Utilisation d'un middle-ware pour implémenter les opérations spécifiques de l'OLAP

Utilisation d'un SGBD Multidimensionnel (systèmes MOLAP)

- ✓ SGBD capable de stocker et traiter des données multidimensionnelles
- ✓ Basé sur un stockage par tableau
- ✓ Indexation rapide des données calculées



Utilisation d'un SGBD Hybride (systemes HOLAP)

- ✓ Tirer profit des avantages des technologies ROLAP et MOLAP :
- ✓ un ROLAP pour stocker et gérer les données détaillées
- ✓ un MOLAP pour stocker et gérer les données agrégées

Schéma d'un DW

Niveau logique « ROLAP » :

Trois grands types de schémas :

- schéma en étoile (star schema) 
- schéma en flocon (snowflake schema) 
- schéma en constellation (fact constellation)

Le schéma en étoile est souvent utilisé pour l'implantation physique

Schéma d'un DW: schéma en étoile

✓ table des faits :

- normalisée,
- de taille très importante,
- avec de nombreux champs

✓ tables de dimensions :

- dimensions de l'analyse,
- taille peu importante,
- avec peu de champs



Ex 1 : *Vente de médicaments dans des pharmacies*

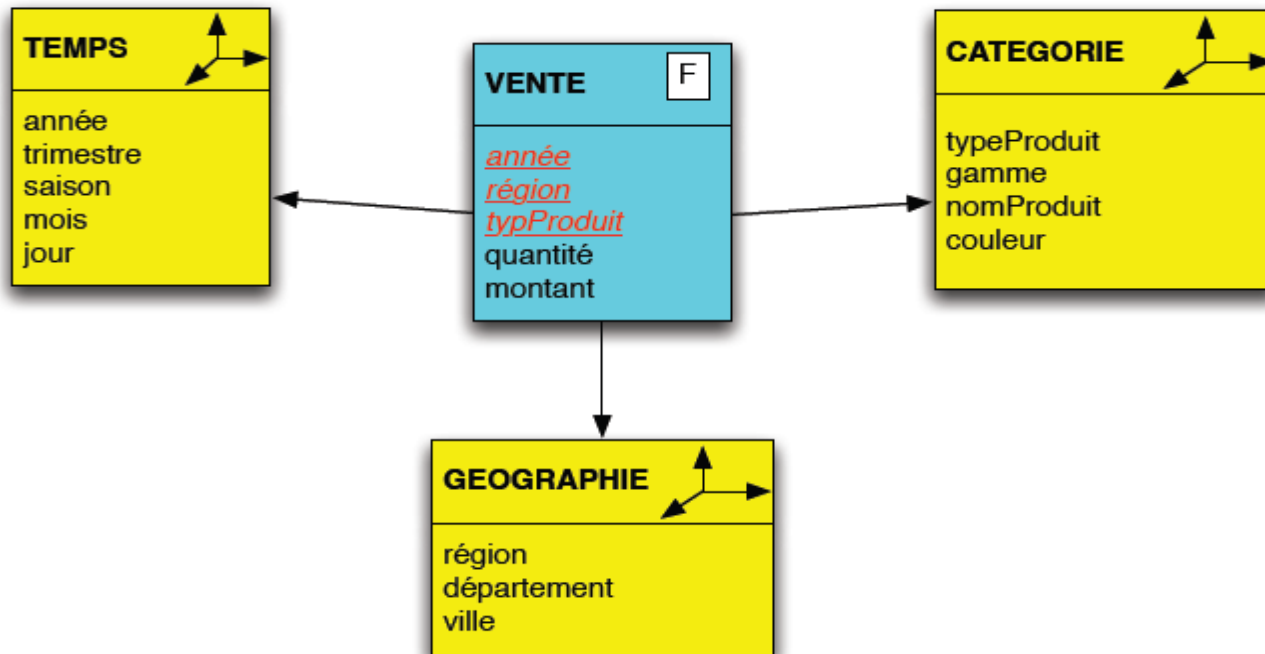


Schéma d'un DW: schéma en étoile

Ex 2 : Ventes d'articles dans un supermarché

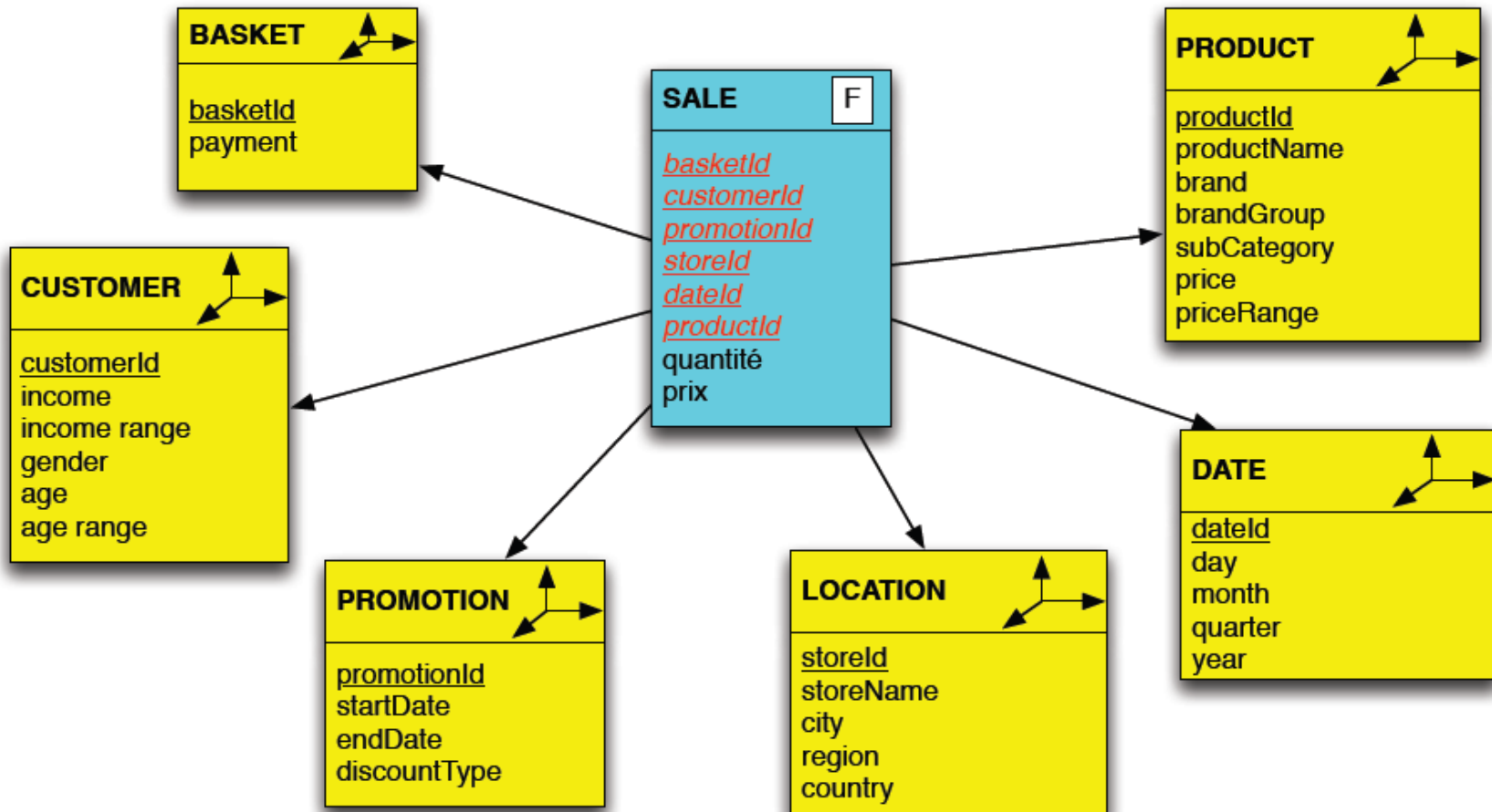


Schéma d'un DW: schéma en étoile

✓ Avantages



- Simple
- Le plus utilisé !!!

✓ Inconvénients

- Possibilité de redondance car les tables de dimension ne sont pas nécessairement normalisées.
- Taille de dimensions plus grosse

Schéma d'un DW: schéma en flocon de neige

Evolution du schéma en étoile avec une décomposition des tables de dimensions du modèle en étoile selon leurs hiérarchies



- ✓ Variante du modèle en étoile.
- ✓ Les tables de dimensions sont normalisées
- ✓ Réduction de la redondance mais exécution parfois plus lente des requêtes (jointure de tables).
- ✓ Modèle adopté par Oracle!!
- ✓ Modèle mixte
 - Seules certaines tables sont normalisées Taille de dimensions plus grosse

Schéma d'un DW: schéma en flocon de neige

Ex 3: Vente de médicaments dans des pharmacies

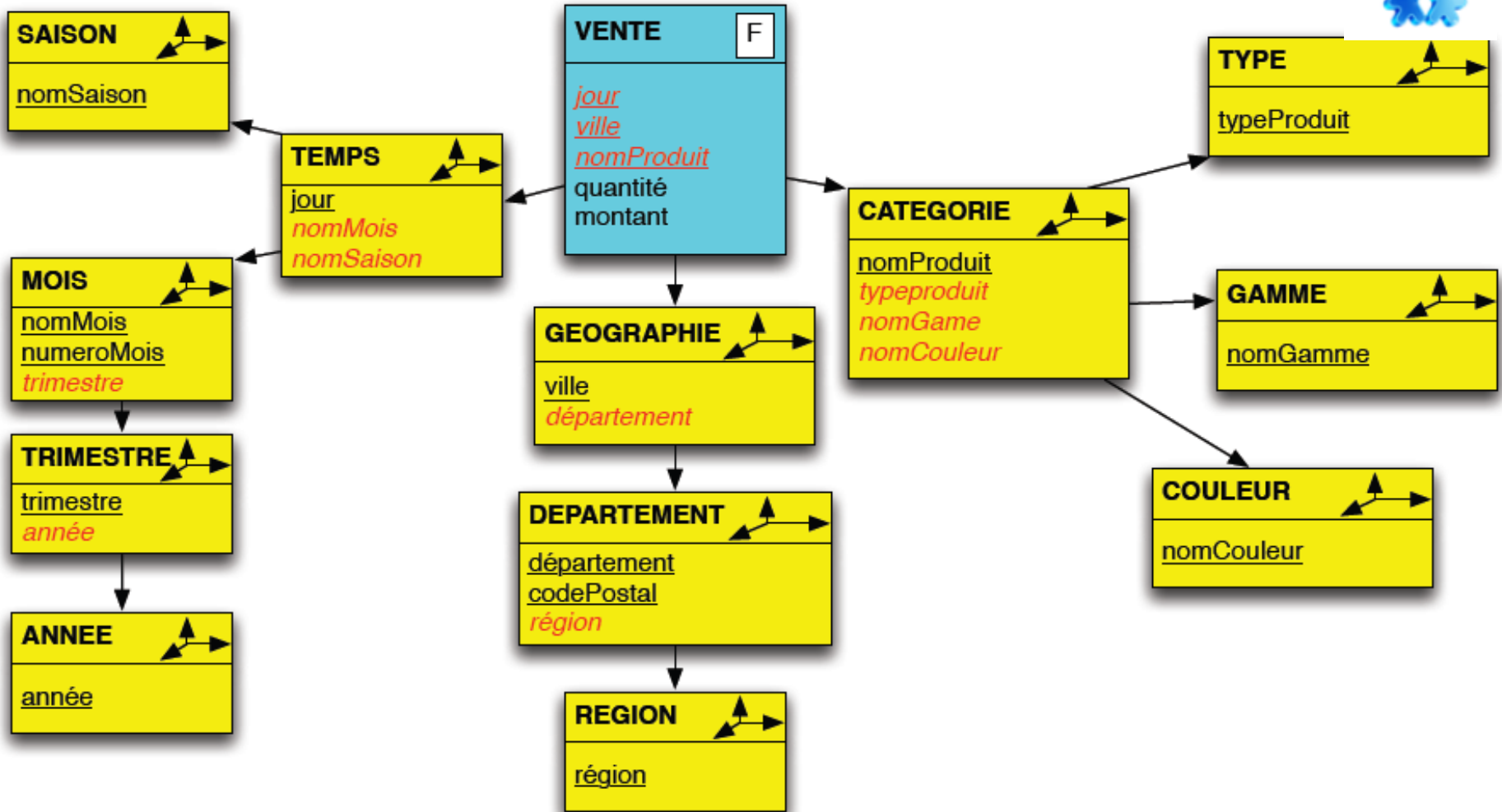
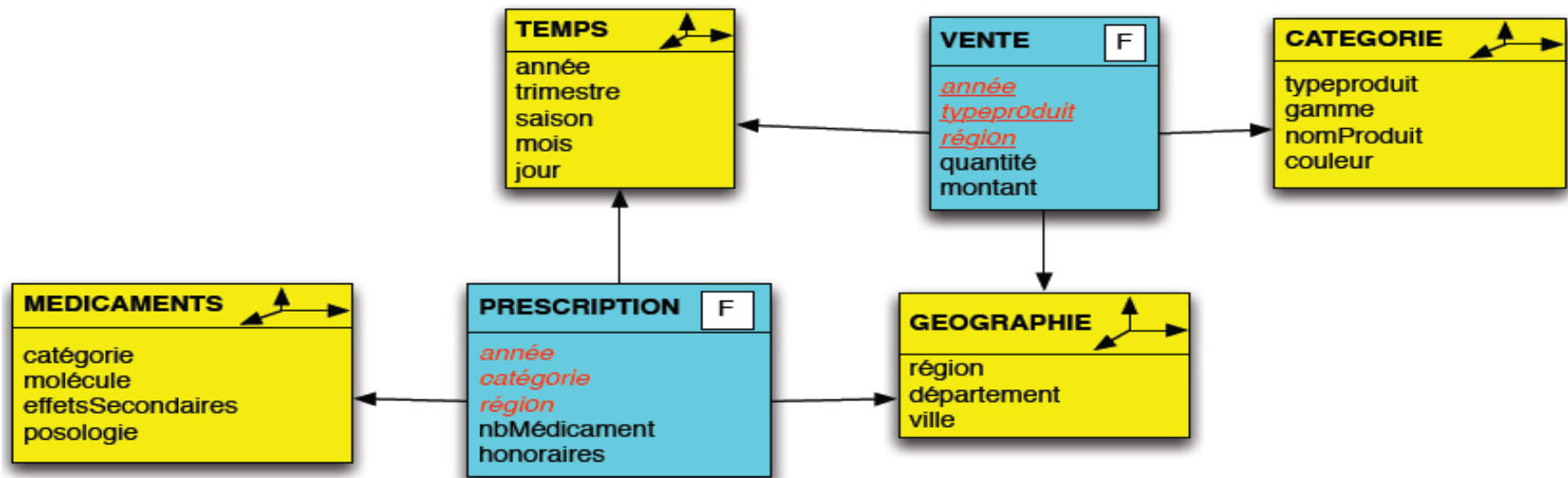


Schéma d'un DW: schéma en constellation

- ✓ Fusionne plusieurs modèles en étoile qui utilisent des dimensions communes.
- ✓ comprend en conséquence plusieurs faits et des dimensions communes ou non

Ex : Vente de médicaments dans des pharmacies



Une constellation est constituée de **2 schémas en étoile** :

- l'un correspond aux VENTES effectuées dans les pharmacies et
- l'autre analyse les PRESCRIPTIONS des médecins.

Manipulation des données multidimensionnelles

Le Modèle MOLAP: Multidimensional On-Line Analytical Processing

- ✓ Utiliser un système multidimensionnel « pur » qui gère les structures multidimensionnelles natives (les cubes).
- ✓ Accès direct aux données dans le cube.
- ✓ Plus difficile à mettre en place
- ✓ Formats souvent propriétaires
- ✓ Conçu exclusivement pour l'analyse multidimensionnelle.
- ✓ Exemples de moteurs MOLAP:
 - Microsoft Analysis Services
 - Hyperion

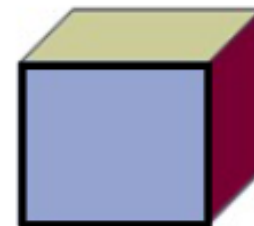
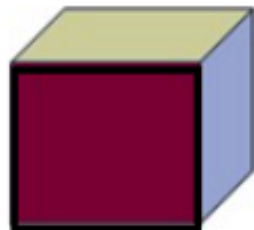
Opérations OLAP

- Opération agissant sur la structure
 - ▣ Rotation (**rotate**): présenter une autre face du cube

	05	06	07
Œuf	221	263	139
Viande	275	257	116



	05	06	07
Idf	101	120	52
Ain	395	400	203



Manipulation des données multidimensionnelles

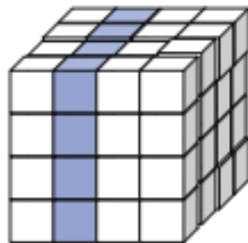
Opérations OLAP

■ Opération agissant sur la structure

- ▣ Tranchage (**slicing**): consiste à ne travailler que sur une tranche du cube. Une des dimensions est alors réduite à une seule valeur

		05	06	07
Œuf	Idf	220	265	284
	Ain	225	245	240
Viande	Idf	163	152	145
	Ain	187	174	184

		06
Œuf	Idf	265
	Ain	245
Viande	Idf	152
	Ain	174



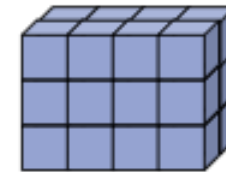
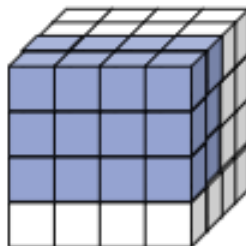
Opérations OLAP

■ Opération agissant sur la structure

- ▣ Extraction d'un bloc de données (**dicing**): ne travailler que sous un sous-cube

		05	06	07
Œuf	Idf	220	265	284
	Ain	225	245	240
Viande	Idf	163	152	145
	Ain	187	174	184

		05	06	07
Œuf	Idf	220	265	284
	Ain	225	245	240

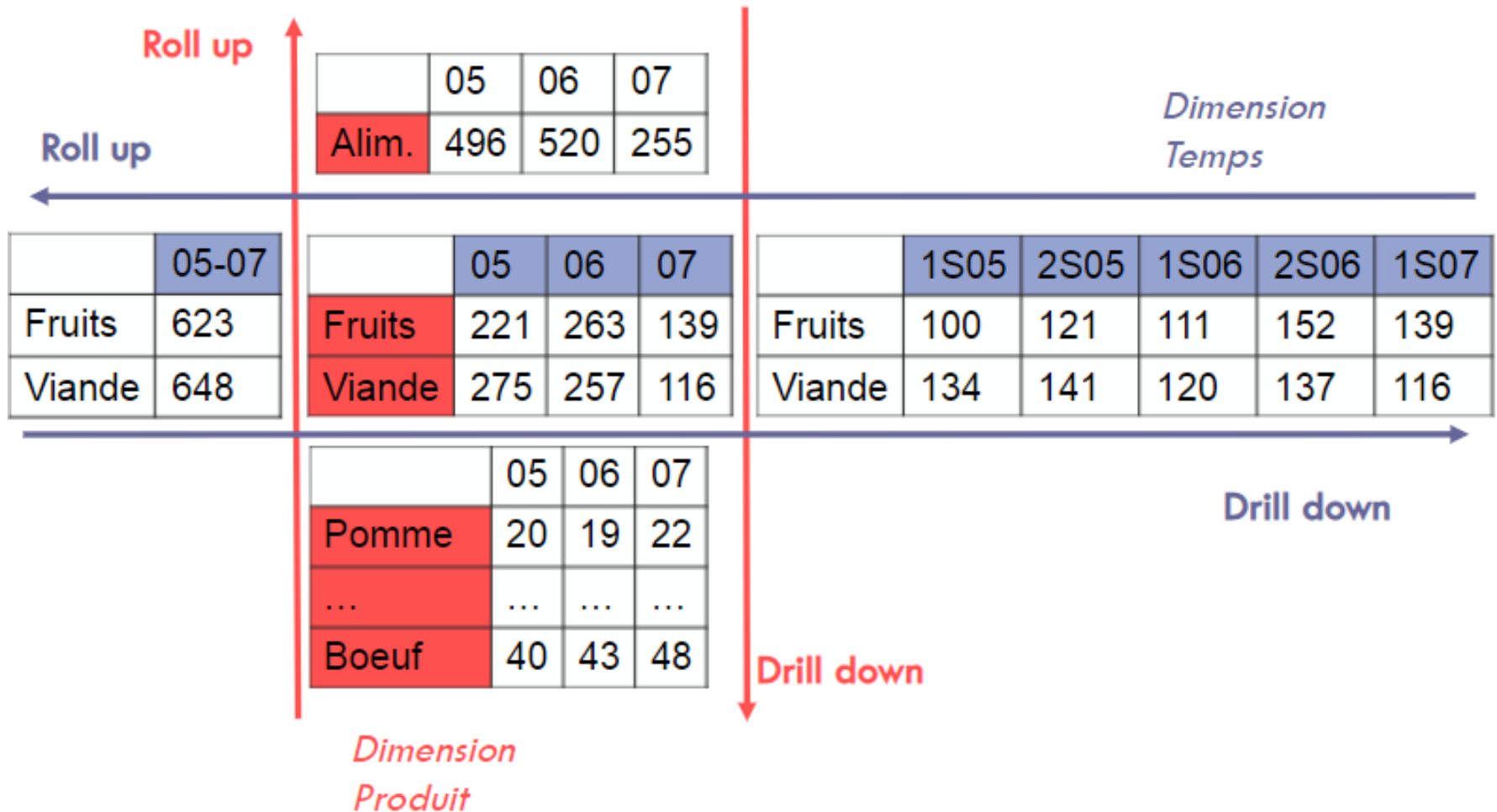


Opérations OLAP

- Opérations agissantes sur la granularité
 - ▣ Forage vers le haut (**roll-up**): « *dézoomer* »
 - Obtenir un niveau de granularité supérieur
 - Utilisation de fonctions d'agrégation
 - ▣ Forage vers le bas (**drill-down**): « *zoomer* »
 - Obtenir un niveau de granularité inférieur
 - Données plus détaillées

Manipulation des données multidimensionnelles

Opérations OLAP



Langages pour faire l'OLAP

Deux langages possibles pour faire de l'OLAP :

1. *SQL étendu (Extensions de SQL-3 / SQL-99 pour OLAP)* :

- ✓ Nouvelles fonctions SQL d'agrégation: *Rank, N_tile, ...*
- ✓ Nouvelles fonctions de la clause GROUP BY :
 - *ROLLUP* équivalent to “control breaks”
 - *CUBE* équivalent to “cross tabulation”
 - *GROUPING SETS* équivalent to multiple GROUP BYs
- ✓ Fenêtre glissante : *WINDOWS / OVER / PARTITION, ...*

2. *MDX (Multi Dimensional eXpression)* :

• **langage de requêtes inventé pour faire de l'OLAP par**

Mosha Pasumansky (Microsoft)

- ✓ disponible dans la plupart des serveurs OLAP
- ✓ plus puissant que SQL pour faire de l'OLAP

Langages pour faire l'OLAP

Exemple d'une requête OLAP en MDX:

```
SELECT {Paris, Berlin} ON ROWS  
        {[Q1], [Q2].CHILDREN} ON COLUMNS  
FROM   CubeSales  
WHERE  (MEASURES.SalesAmount,  
         Time.[2014],  
         Product.Product)
```

SELECT {Paris, Berlin} **ON ROWS**

{[Q1], [Q2].CHILDREN} **ON COLUMNS**

FROM CubeSales

WHERE (MEASURES.SalesAmount,
 Time.[2014],
 Product.Product)

Agrégation de la mesure
« SalesAmount » avec la
fonction SUM

Sélection de la dimension
Time (2014 seulement)

Sélection de la dimension
Product (all product)

Résultat :

	Q1 2014	April 2014	May 2014	June 2014
Paris	12,567	3,360	5,450	4,570
Berlin	12,567	3,360	5,450	4,570
...	SalesAmount.values	...

*Merci de votre
attention*

Des Questions



K_menghour@yahoo.fr