

Système d'information Décisionnel

Introduction

ECOLE NATIONALE POLYTECHNIQUE D'ORAN

Département Mathématiques et Informatique

Filière IMSI : 4^{ème} année ingénieur

M. SABRI

Introduction

- Objectifs :
 - Transformer un système d'information qui avait une vocation de production en un SI décisionnel :
 - Transformation des données de production en informations stratégiques
- Exemple de requêtes décisionnelles :
 - ☐ Catégorie socioprofessionnelle des meilleurs clients de chaque région
 - ☐ Evolution de la part de marché d'un produit particulier
 - ☐ Nombre d'employé de l'entreprise par classe d'âge, par sexe, par grade
 - ☐ Quel est le profil des employés les plus performants?

Systemes transactionnels vs Systemes décisionnels

Systemes transactionnels :

- Les outils traditionnels de gestion et d'exploitation des données sont du type transactionnel ou OLTP (On-Line Transaction Processing)
- ☐ L'exploitation de données tournée vers la saisie, le stockage, la mise à jour, la sécurité et l'intégrité des données.
- ☐ Le système transactionnel est développé pour gérer les transactions quotidiennes
- ☐ Ces bases de données supportent habituellement des applications particulières
- ☐ Le contenu est fait de données actuelles, pas d'archives
- ☐ Les données sont très détaillées (détails de chacune des transactions)
- ☐ La mise à jour s'effectue par de nouvelles transactions
- ☐ Très souvent plusieurs de ces systèmes existent indépendamment les uns des autres

Systemes transactionnels vs Systemes décisionnels

Opérations dans les systèmes transactionnels :

- ☐ Ajout
- ☐ Effacement
- ☐ Mise à jour des enregistrements
- ☐ Requêtes simples
- ☐ Interrogations et modifications fréquentes des données par de nombreux utilisateurs

Systemes transactionnels vs Systemes décisionnels

Systemes décisionnels :

- ☐ Le terme décisionnel « Business Intelligence » couvre l'ensemble des technologies permettant en bout de chaîne d'apporter une aide à la décision.
- ☐ SI capable d'agrèger les données internes ou externes et de les transformer en information servant à une prise de décision rapide.
- SI capable de répondre à certains types de questions:
 - ☐ Quelles sont les ventes du produit X pendant le trimestre A de l'année B dans la région C ?
 - ☐ Comment se comporte le produit X par rapport au produit Y?
 - ☐ Quel type de client peut acheter le produit X?
 - ☐ Est-ce qu'une baisse de prix de 10% par rapport à la concurrence ferait redémarrer les ventes du produit X ?

Systemes transactionnels vs Systemes décisionnels

Systemes décisionnels :

- Ces exemples mettent en évidence les faits suivants:
 - ☐ Les questions doivent pouvoir être formulées dans le langage de l'utilisateur en fonction de son secteur d'activité:
 - ✓ ☐ *Service marketing, Service économique, service relation clients...*
 - ☐ La prévision des interrogations est difficile car elles sont du ressort de l'utilisateur.
 - ☐ Les questions vont varier selon les réponses obtenues:
 - ✓ ☐ *Si le produit X s'est vendu moins bien que l'année précédente, il va être utile de comprendre les raisons: Détailler les ventes par région par type de magasin,...*
 - ☐ Des questions ouvertes vont nécessiter la mise en place de méthodes d'extraction d'informations

Systemes transactionnels vs Systemes décisionnels

Données :

✓ *Points de ventes, géographiques, démographiques, ...*

Informations :

✓ *I vit dans R, I est âgé de A, ...*

Connaissances :

✓ *Dans X%, le produit Y est vendu en même temps que le produit Z, ...*

Décisions

✓ *Lancer la promotion de Y et Z dans R auprès des clients plus âgés que A, ...*

Applications transactionnelles vs Applications décisionnelles

Les applications transactionnelles sont constituées de traitements factuels de type OLTP (On Line Transaction Processing)

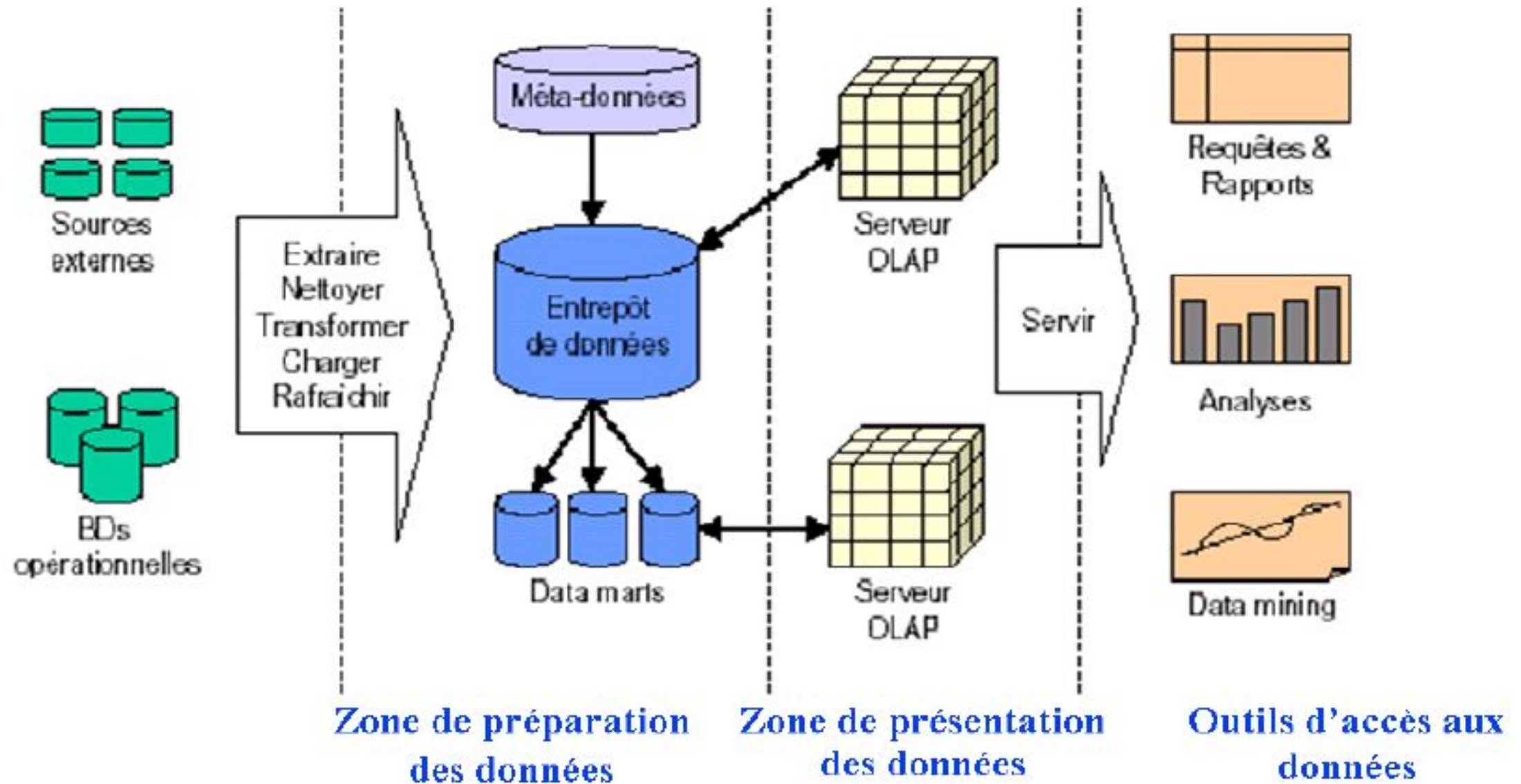
Les applications d'aide à la décision sont constituées de traitements ensembliste de type
OLAP: On Line Analytical Processing

Applications transactionnelles vs Applications décisionnelles

Incompatibilités des deux activités

- Les deux activités (OLTP & OLAP) ne peuvent co-exister sur des données dans le même système d'information: leurs objectifs de performance sont exactement opposés:
 - ❑ Les requêtes complexes et lourdes dégradent les performances des systèmes transactionnels,
 - ❑ Les données temporelles sont réparties entre données actuelles et données archivées, rendant la vue historique des données très difficile ou impossible,
- ❑ Le support efficace d'une activité OLAP nécessite la constitution d'un système d'information propre: Le Datawarehouse

Naissance du Datawarehouse



Datawarehouse : Définition

« Collection de données orientées sujets, intégrées, non volatiles et historisées, organisées pour le support du processus d'aide à la décision »

Base de données dans laquelle sont déposées après nettoyage et homogénéisation les informations en provenance des différents systèmes de production de l'entreprise OLTP

Caractéristiques des données d'un DW

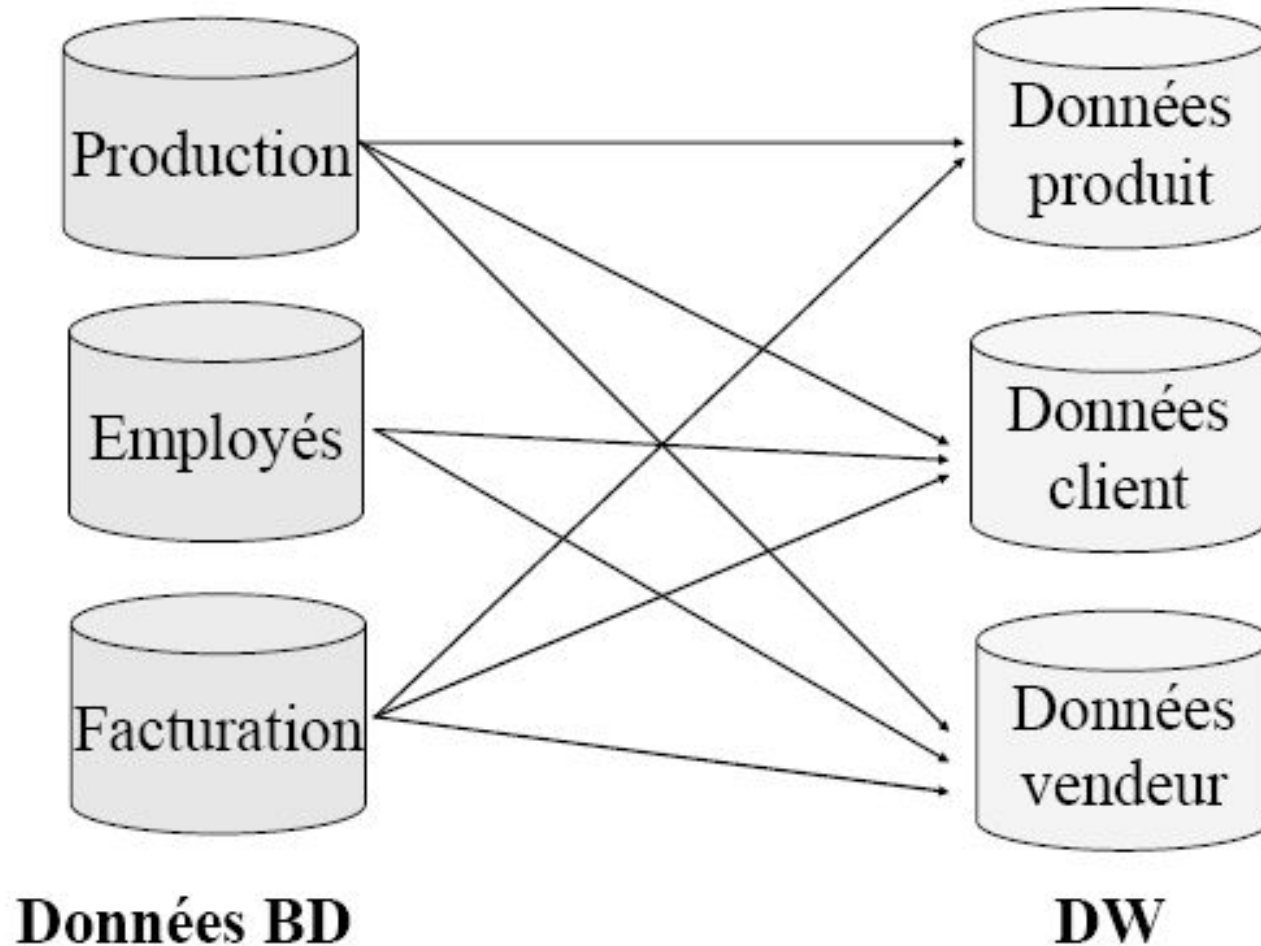
Orientées sujet :

- Organisées autour de sujets majeurs et des métiers de l'entreprise
- Données pour l'analyse et la modélisation en vue de l'aide à la décision, et non pas pour les opérations et transactions journalières
- Vue synthétique des données selon les sujets intéressant les décideurs

L'approche des données orientées sujet permet d'organiser l'entrepôt de données autour des sujets majeurs et des métiers de l'organisme et le développer via une incrémentation sujet par sujet. Cette orientation permet d'éviter la redondance des données concernées par plusieurs sujets. Les données, réorganisées autour de sujets, permettent à chaque décideur de disposer d'une vue sur les informations qui lui sont pertinentes, et qui peuvent influencer dans ses décisions

Caractéristiques des données d'un DW

Orientées sujet :



Caractéristiques des données d'un DW

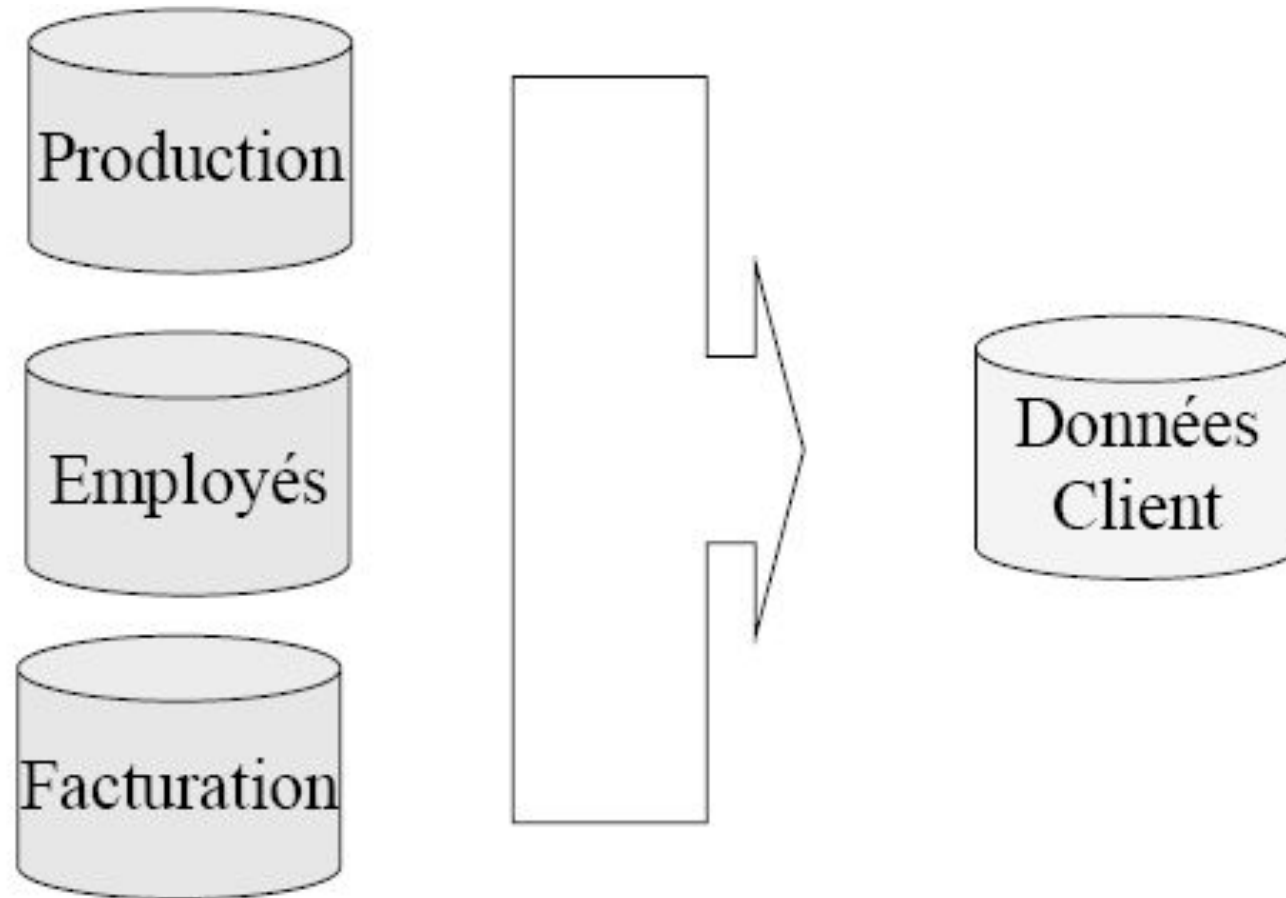
Intégrées :

- Construit en intégrant des sources de données multiples et hétérogènes
 - ✓ BDD relationnelles, fichiers, enregistrements de transactions
- Les données doivent être mises en forme et unifiées afin d'avoir un état cohérent
- Phase la plus complexe (60 à 90 % de la charge totale d'un projet DW)

Les données de l'entrepôt proviennent de plusieurs sources opérationnelles ou transactionnelles déjà existantes, et éventuellement hétérogènes (peuvent avoir des formats différents, des attributs différents, des descriptions différentes, etc.). Pour avoir une image physique unique des données dans l'entrepôt de données, les données doivent être intégrées c'est-à-dire converties, reformatées, résumées, etc.

Caractéristiques des données d'un DW

Intégrées :



Caractéristiques des données d'un DW

Historisées :

- Fournies par les sources opérationnelles
- Matière première pour l'analyse
- Stockage de l'historique des données, pas de mise à jour
- Un référentiel temps doit être associé aux données

La conservation de l'évolution des données dans le temps est fondamentale pour un système décisionnel. Elle consiste à s'appuyer sur les résultats passés pour la prise de décision et mesurer leurs effets ; autrement dit, conserver les données permet de mieux appréhender le présent et d'anticiper le futur

Caractéristiques des données d'un DW

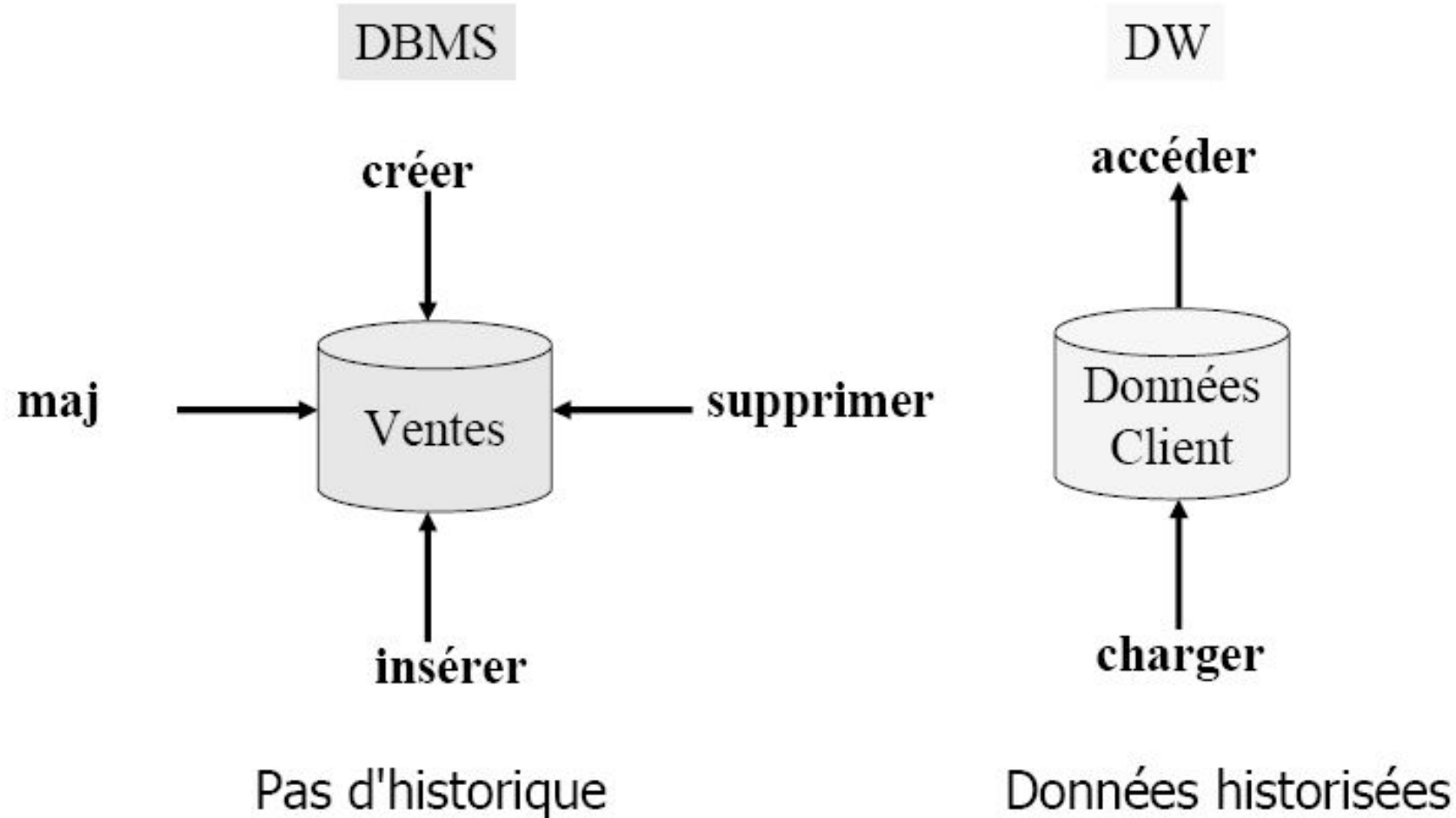
Non volatiles :

- Conséquence de l'historisation
- Une même requête effectuée à intervalle de temps, en précisant la date référence de l'information donnera le même résultat
- Stockage indépendant des BDD opérationnelles
- Pas de mises à jour des données dans le DW

La non-volatilité consiste à historiser les données tout en gardant les versions de mise à jour durant leurs existences. Il s'agit toujours d'une opération d'insertion et non de mise à jour. Les modifications ne sont autorisées que pour des cas particuliers (correction d'erreurs...etc.). Ainsi, dès lors qu'une donnée a été qualifiée pour être introduite dans l'entrepôt, elle ne peut ni être modifiée, ni supprimée. Elle devient de fait partie prenante de l'historique.

Caractéristiques des données d'un DW

Non volatiles :



Les données

- Volume très important
- Données dispersées, souvent difficiles d'accès
- Peu ou mal intégrées
- Complexes
- Non structurées pour les applications décisionnelles

Requêtes (1)

- BD-OLTP représentent les données sous forme aplatie: relation, données normalisées

produit	région	vente	date	vendeur
écrou	Est	50	01012004	X
écrou	Ouest	60	12122003	X
écrou	Centre	110	01112003	Y
vis	Est	70	01042004	Y
vis	Ouest	80	10022004	Z
vis	Centre	90	29032004	Y
boulon	Est	120	05052004	X
boulon	Ouest	10	24042004	Z
boulon	Centre	20	11022004	Y
joint	Est	50	01032004	X
joint	Ouest	40	01102003	Y
joint	Centre	70	01012003	Z

produit	prix	fournisseur
écrou	44	CC
vis	2	DD
boulon	3	VV
joint	1	BB

fournisseur	ville
...	...

Requêtes (2)

- OLTP: Requêtes simples « qui, quoi »
 - par ex. les ventes de X.
 - jointures: les ventes de X à quel prix de quel fournisseur,
- OLAP: besoin de données agrégées, synthétisées
 - nombre de ventes par vendeur, par région, par mois,
 - nombre de ventes par vendeur, par fournisseur, par mois,
 - ...
- SQL: Possibilité d'agréger les données (group by) mais très coûteux (parcourir toutes les tables) et il faut recalculer à chaque utilisation
- Sur plusieurs tables (ex : somme des ventes par fournisseur), nécessité de faire des jointures souvent coûteuses

Différences BD – DW

Caractéristiques	OLTP	OLAP
Applications	production	aide à la décision
Utilisateurs	un département professionnel IT	transversal (entreprise) décideur non IT
Données	normalisées, non agrégées	dénormalisées, agrégées
Requêtes	simples, nombreuses, régulières, prévisibles, répétitives	complexes, peu nombreuses, irrégulières, non prévisibles
Nb tuples invoqués par requête (moyenne)	dizaines	millions
Taille données	100 MB à 1 GB	1 GB à 1 TB
Ancienneté des données	récente, mises à jour	historique

Nécessité d'une structure multi-dimensionnelle

- **Les BD relationnelles ne sont pas adaptées à l'OLAP car :**
 - Pas les mêmes objectifs
 - Pas les mêmes données:
 - Les données nécessaires à l'OLAP sont multi-dimensionnelles (i.e. ventes par vendeur, par date, par ville...). Les tables en représentent une vue aplatie.
 - Pas les mêmes traitements et requêtes:
 - Non seulement perte de performances mais aussi nécessité pour les utilisateurs de savoir comment trouver les liens entre les tables pour recréer la vue multi-dimensionnelle.
- **Il est donc nécessaire de disposer d'une structure de stockage adaptée à l'OLAP, i.e. permettant de**
 - représenter les données dans plusieurs dimensions,
 - manipuler les données facilement et efficacement.

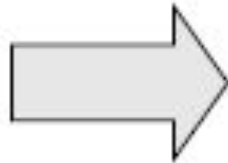
Séparation BD et ED

- **Les DW vont être physiquement séparés des BD, pour des raisons de :**
 - Performance : systèmes de production ne sont pas organisés pour pouvoir répondre efficacement aux requêtes des systèmes d'aide à la décision. Même les requêtes simples peuvent dégrader sérieusement les performances.
 - Données différentes:
 - Données historisées : aide à la décision nécessite des données sur une longue durée, non conservée dans les BD
 - Données agrégées
 - Qualité des données : sources différentes qui utilisent souvent des noms, formats, codes et mesures différents devant être uniformisés

DW-OLAP

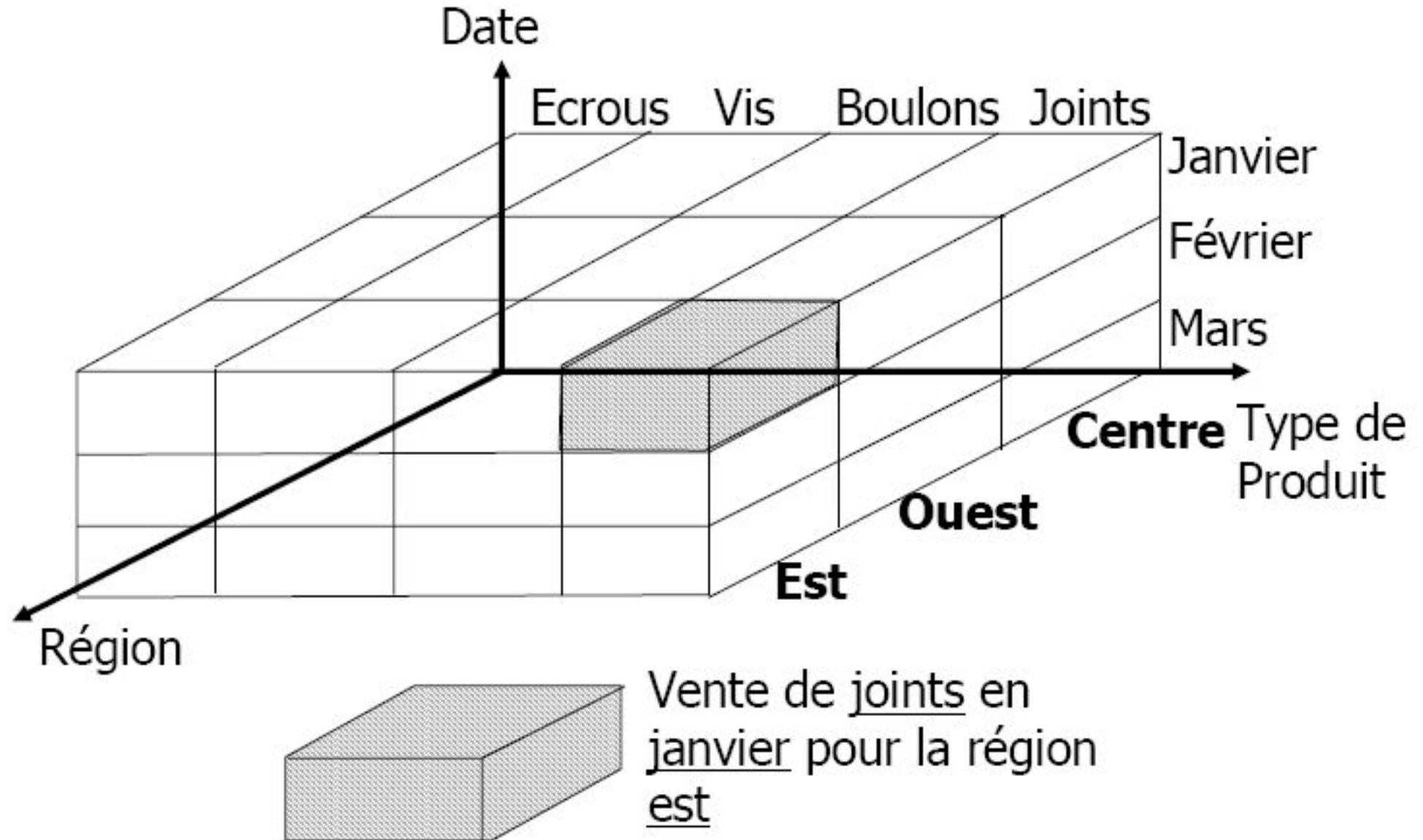
- DW-OLAP :
 - représentation des données sous forme multidimensionnelle : « Cube »

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



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

Cube: représentation des données sous forme multidimensionnelle



Exemple 2

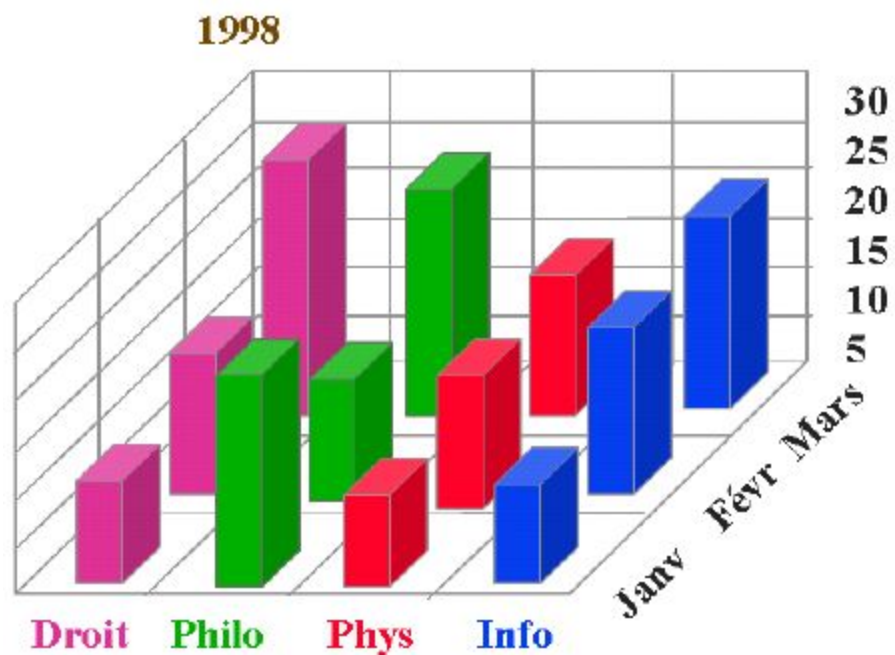
DuréeMoy	Départ.	Mois	Année
5	Info	Janv	1998
5	Phys	Janv	1998
18	Philo	Janv	1998
7	Droit	Janv	1998
12	Info	Févr	1998
8	Phys	Févr	1998
9	Philo	Févr	1998
15	Droit	Févr	1998
18	Info	Mars	1998
12	Phys	Mars	1998
22	Philo	Mars	1998
25	Droit	Mars	1998

Tableau simple

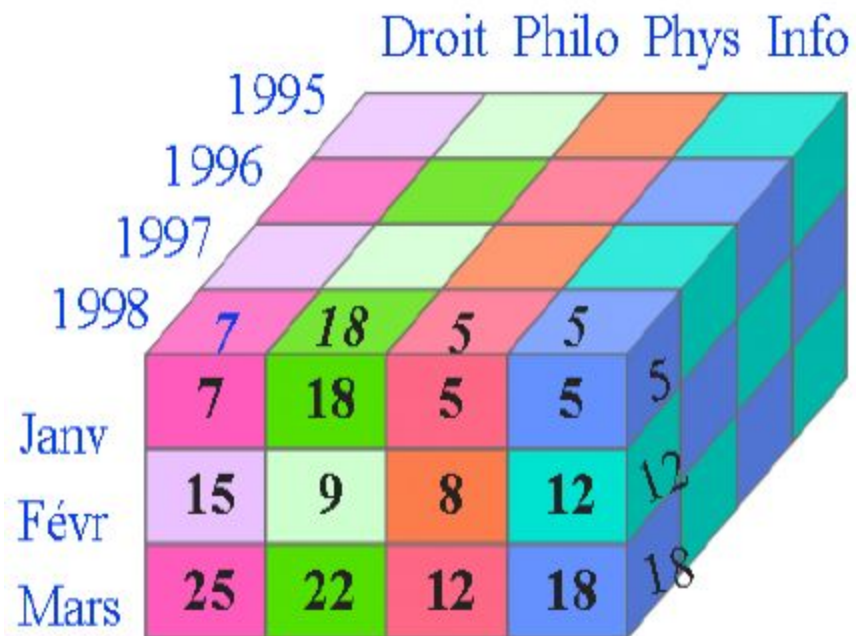
1998	Janv	Févr	Mars
Info	5	12	18
Phys	5	8	12
Philo	18	9	22
Droit	7	15	25

Tableau croisé

Exemple 2



Graphique 3D



Hyper cube:

Notion de Datamart (Magasin de données)

- **Définition :**

- C'est un sous-ensemble de données dérivées du DW ciblé sur un sujet unique

- **Caractéristiques:**

- Orienté vers un sujet unique
 - Ex: comportement de la clientèle
 - Données fortement agrégées
 - Le DW joue le rôle de source et d'historique pour le Datamart
 - Organisation multidimensionnelle (cubique)
 - Dont l'une des dimensions indique souvent le temps
 - Lien dynamique avec le DW
 - Association entre valeur agrégée et valeur détaillée
 - Interfaces simples et conviviales

Les domaines d'application du décisionnel

- La gestion de la relation client (CRM) est l'un des premiers champs d'application de la Business Intelligence.
- □ Le contrôle de gestion pour l'analyse des coûts, l'analyse de la rentabilité, l'élaboration budgétaire, les indicateurs de performance...
- La direction marketing pour le ciblage, le pilotage de gamme, les applications de géomarketing, de fidélisation clients...
- La direction commerciale pour le pilotage des réseaux, les prévisions des ventes, l'optimisation des territoires...
- Les ressources humaines pour la gestion des carrières,
- La direction de la production pour l'analyse qualité, la prévision des stocks, la gestion des flux, la fiabilité industrielle...
- La direction générale pour les tableaux de bord, indicateurs de pilotage, gestion d'alertes...

Exemple: un DW dans les télécoms

❑ Sujets :

- Suivi du marché: lignes installées/ désinstallées, services et options choisis, répartition géographique, répartition entre public et différents secteurs d'organisations
- Comportement de la clientèle
- Comportement du réseau

❑ Historique

- 5 ans pour le suivi du marché
- 1 an pour le comportement de la clientèle
- 1 mois pour le comportement du réseau

❑❑ Sources

- Fichiers clients élaborés par les agences
- Fichiers de facturation

❑❑ Requêtes

- Comportement clientèle
- Nombre moyen d'heures par client, par mois et par région
- Durée moyenne d'une communication urbaine par ville
- Durée moyenne d'une communication internationale

Système d'information Décisionnel

Modélisation et Conception d'un DW

ECOLE NATIONALE POLYTECHNIQUE D'ORAN

Département Mathématiques et Informatique

Filière IMSI : 4^{ème} année ingénieur

M. SABRI

Construction d'un Datawarehouse

❑Caractéristiques :

- Le Datawarehouse est différent des bases de données de production:
 - Les besoins pour lesquels on veut le construire sont différents
 - Il contient des informations historisées, organisées selon les métiers de l'entreprise pour le processus d'aide à décision
- ❑ Le Datawarehouse n'est pas un produit ou un logiciel mais un environnement, qui se bâtit et ne s'achète pas.

Construction d'un Datawarehouse

❑ Phases de construction d'un DW:

- Il y'a trois parties interdépendante qui relève la construction d'un Datawarehouse:
 - L'étude préalable qui va définir les objectifs, la démarche à suivre, le retour sur investissement,...
 - L'étude du modèle de données qui représente le DW conceptuellement et logiquement
 - L'étude de l'alimentation du Datawarehouse

Etude préalable

❑ Etude des besoins:

- Définir les objectifs du DW
- Déterminer le contenu du DW et son organisation, d'après:
 - Les résultats attendus par les utilisateurs,
 - Les requêtes qu'ils formuleront,
 - Les projets qui ont été définie
- Recenser les données nécessaires à un bon fonctionnement du DW:
 - Recenser les données disponibles dans les bases de production
 - Identifier les données supplémentaires requises
- Choisir les dimensions
 - Typiquement: le temps, le client, le produit, le magasin...
- Choisir les mesures de fait
 - De préférences de quantités numériques additives
- Choisir la granularité des faits
 - Niveau de détails des dimensions

Modélisation

□ Niveau conceptuel:

- Un DW est basé sur une modélisation multidimensionnelle qui représente les données dans un cube
- Un cube permet de voir les données suivant plusieurs dimensions:
 - Tables de dimensions
 - La table des faits contient les mesures et les clés des dimensions

Modélisation

□ Niveau logique:

- Plusieurs schémas types sont proposés pour représenter un DW:
 - □ Schéma en étoile;
 - □ Schéma en flocon;
 - □ Schéma en constellation;

Alimentation

- L'alimentation est la procédure qui permet de transférer des données du système opérationnel vers le DW
- La conception de cette opération est une tâche complexe
 - 60 à 90 % de la charge totale d'un projet DW
- Elle doit être faite en collaboration avec l'administrateur des bases de productions
- Il est nécessaire de déterminer:
 - Quelles données seront chargées
 - Les transformations et les vérifications nécessaires
 - La périodicité et le moment de transferts des données

Système d'information Décisionnel

Modèles en étoile et en flocon

ECOLE NATIONALE POLYTECHNIQUE D'ORAN

Département Mathématiques et Informatique

Filière IMSI : 4^{ème} année ingénieur

M. SABRI

Modélisation en étoile ou en flocons

- Modélisation conceptuelle BD : entité et relation
- Modélisation de DW : dimension et mesure
- Les mesures sont les valeurs numériques que l'on compare (ex : montant_ventes, qte_vendue)
 - Ces valeurs sont le résultat d'une opération d'agrégation des données
- Les dimensions sont les points de vues depuis lesquels les mesures peuvent être observées :
 - Ex : date, localisation, produit, etc.
 - Elles sont stockées dans les tables de dimensions

OLAP

- « Il s'agit d'une catégorie de logiciels axés sur l'exploration et l'analyse rapide des données selon une approche multidimensionnelle à plusieurs niveaux d'agrégation ».
- OLAP vise à assister l'utilisateur dans son analyse en lui facilitant l'exploration de ses données et en lui donnant la possibilité de le faire rapidement.
 - L'utilisateur n'a pas à maîtriser des langages d'interrogation et des interfaces complexes
 - L'utilisateur interroge directement les données, en interagissant avec celles-ci

Modèle conceptuel

- Approche multidimensionnelle
 - Souvent représentée par une structure à plusieurs dimensions
 - Une dimension est un attribut ou un ensemble d'attributs:
 - Temps
 - Géographie
 - Produits
 - Clients
 - Les cellules contiennent des données agrégées appelées Faits ou Indicateurs:
 - Nombre d'unités vendues
 - Chiffre d'Affaire
 - Coût
 - Représentations:
 - Relations,
 - Cube de données,
 - hyper cube de données

Définitions

- Principe de base : ce sont les analyses des indicateurs qui intéressent l'utilisateur
- Le modèle multidimensionnel contient 2 types d'attributs : les *dimensions* et les *mesures*
- Les mesures sont les valeurs numériques que l'on compare (ex : montant ventes, qté vendue)
 - Ces valeurs sont le résultat d'une opération d'agrégation des données
- Les dimensions sont les points de vues depuis lesquels les mesures peuvent être observées:
 - Ex : date, région, type de produit, etc.

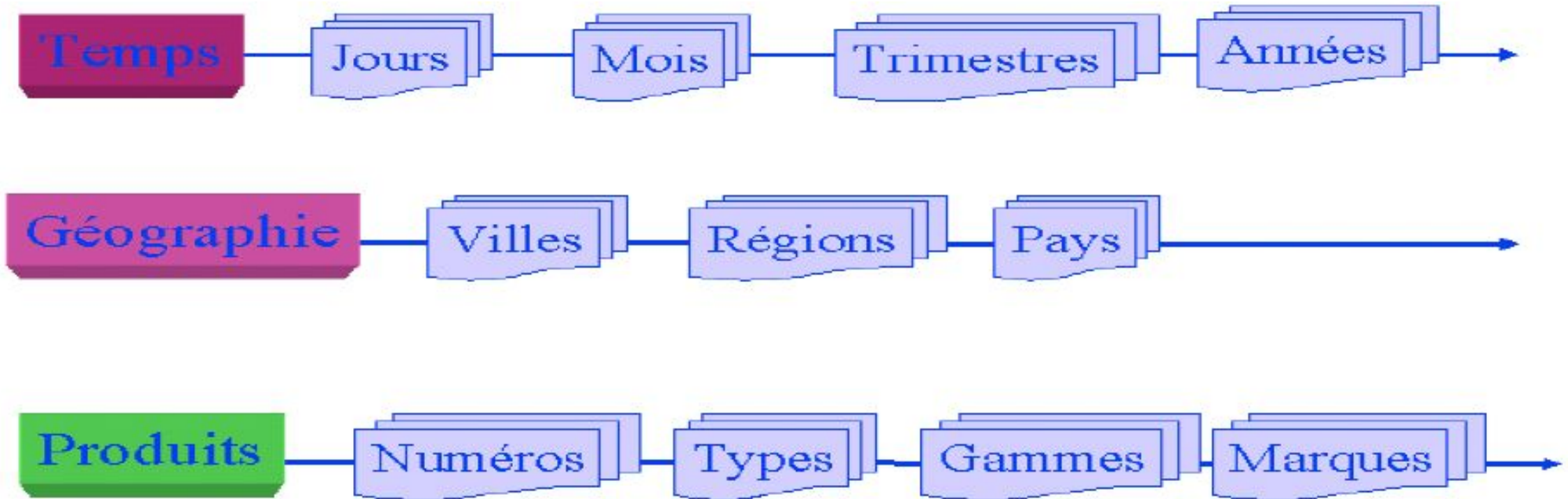
Dimension (1)

- Dimension = liste d'éléments
- Dimension contient des membres organisés en hiérarchie :
 - Chacun des membres appartient à un niveau hiérarchique (ou niveau de granularité) particulier
- Granularité d'une dimension : nombre de niveaux hiérarchiques

Temps :

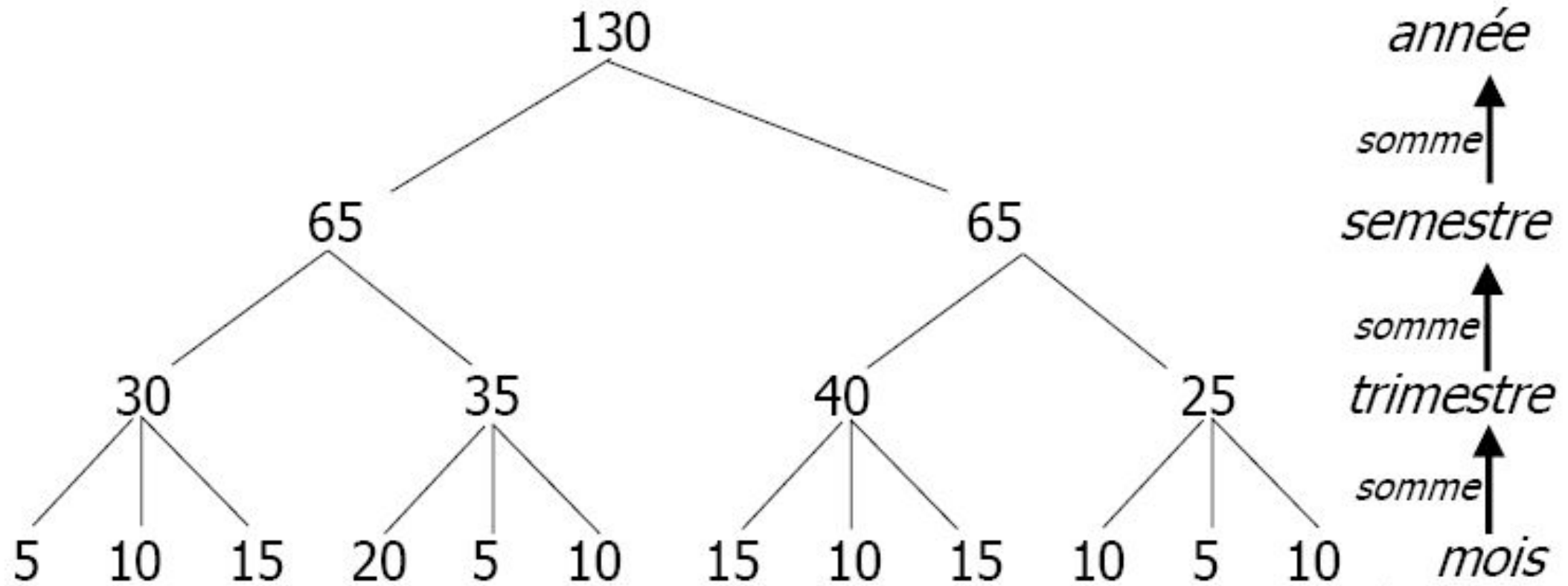
année – semestre – trimestre - mois

Granularité des dimensions



Dimension (2)

- Les axes de dimension doivent fournir des règles de calcul d'agrégat pour chaque mesure:



Les dimensions

- Une dimension peut être définie comme :
 - un thème, ou un axe (attributs), selon lequel les données seront analysées
 - Ex : Temps, Découpage administratif, Produits
- Une dimension contient des membres organisés en hiérarchie :
 - Chacun des membres appartient à un niveau hiérarchique (ou niveau de granularité) particulier
 - Ex : pour la dimension Temps: année – semestre – mois – jour

Les mesures

- Une mesure est un élément de donnée sur lequel portent les analyses, en fonction des différentes dimensions
 - Ex : coût des travaux, nombre d'accidents, ventes

Les faits

- Un fait représente la valeur d'une mesure, mesurée ou calculée, selon un membre de chacune des dimensions
 - Exemple : «250 000 DA» est un fait qui exprime la valeur de la mesure « coût des travaux » pour le membre « 2002 » du niveau année de la dimension « temps » et le membre « Oran » du niveau « ville » de la dimension « découpage administratif »

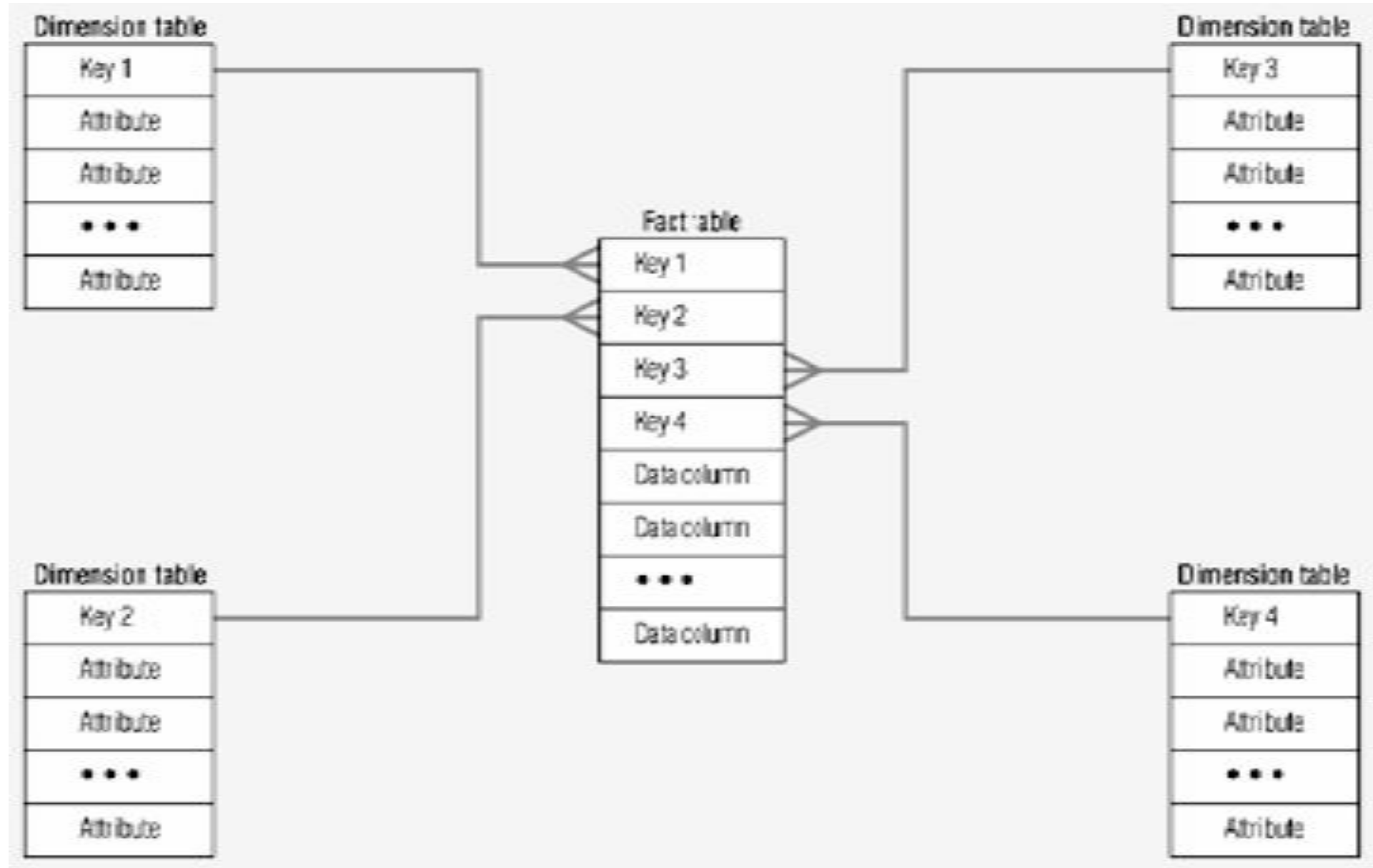
Les tables des faits

- Les mesures sont stockées dans les tables de faits
 - Table de fait contient les valeurs des mesures et les clés vers les tables de dimensions

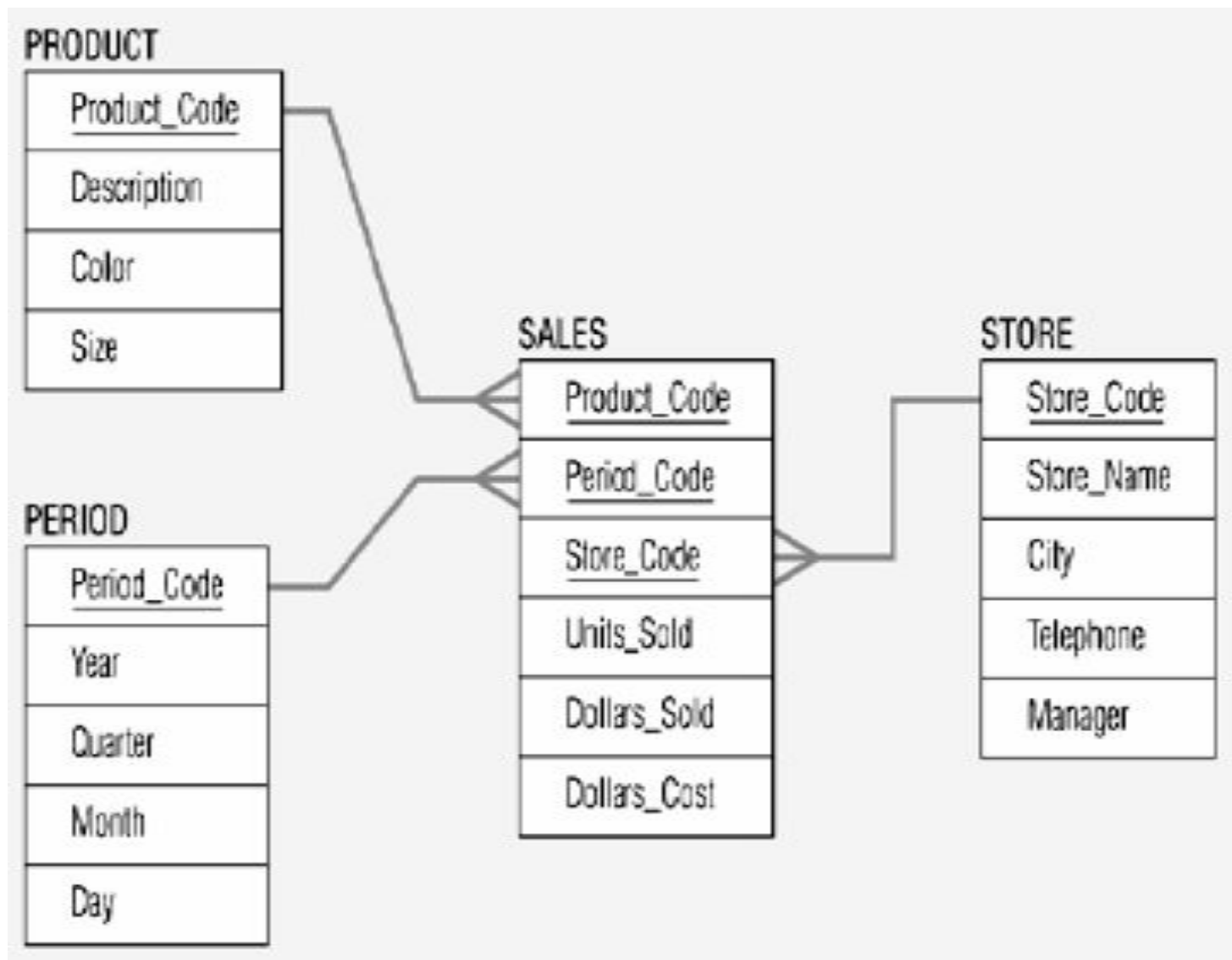
Le modèle en étoile

- Une (ou plusieurs) table(s) de faits comprenant une ou plusieurs mesures.
- Plusieurs tables de dimension dénormalisées: descripteurs des dimensions.
- Les tables de dimension n'ont pas de lien entre elles
- Avantages :
 - Facilité de navigation
 - Performances : nombre de jointures limité
 - Gestion des agrégats
- Inconvénients :
 - Redondances dans les dimensions
 - Alimentation complexe.

Le modèle en étoile



Example



Product

<u>Product _Code</u>	Description	Color	Size
100	Sweater	Blue	40
110	Shoes	Brown	10 1/2
125	Gloves	Tan	M
...			

Period

<u>Period _Code</u>	Year	Quarter	Month
001	1999	1	4
002	1999	1	5
003	1999	1	6
...			

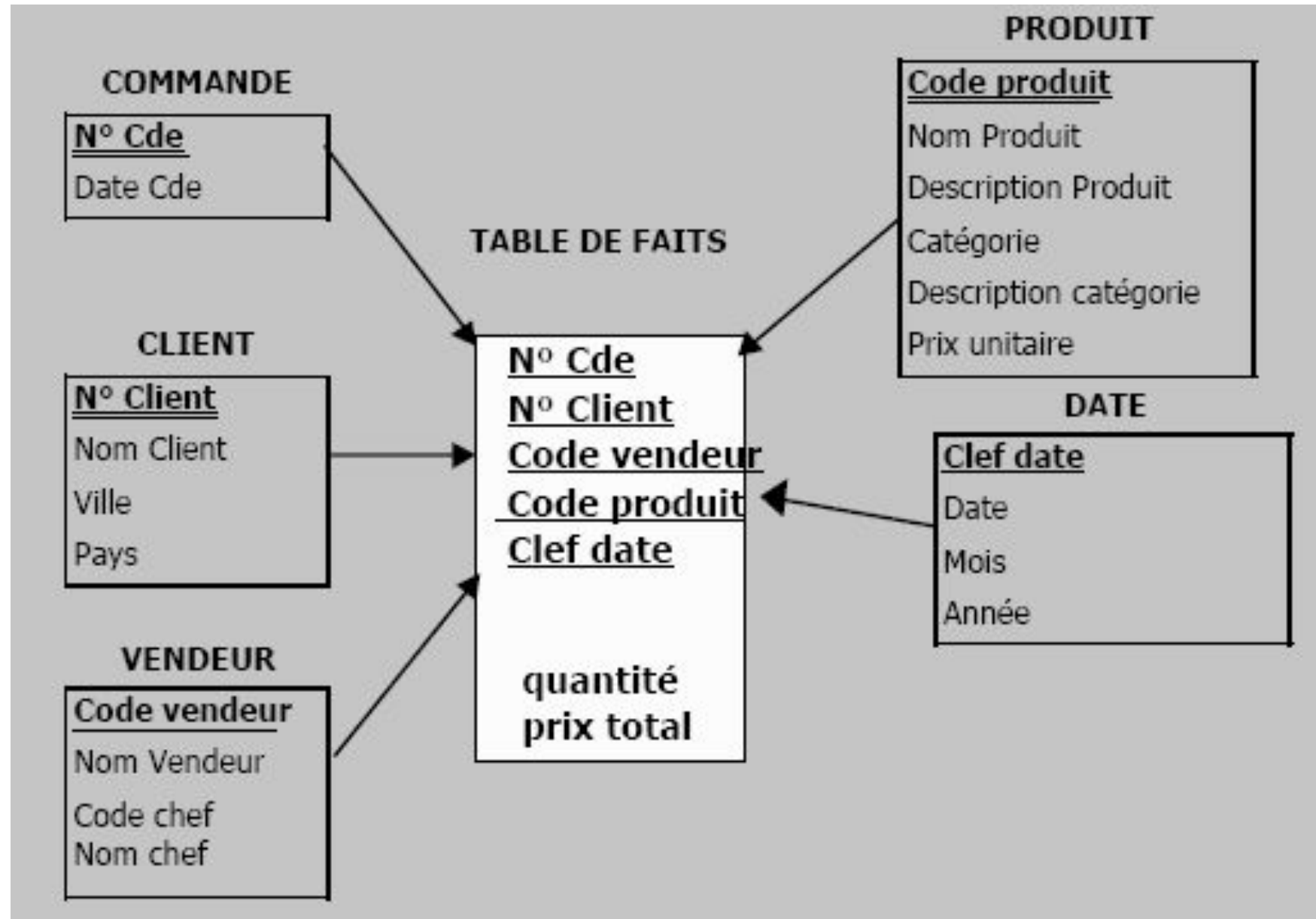
Sales

<u>Product _Code</u>	<u>Period _Code</u>	<u>Store _Code</u>	Units _Sold	Dollars _Sold	Dollars _Cost
110	002	S1	30	1500	1200
125	003	S2	50	1000	600
100	001	S1	40	1600	1000
110	002	S3	40	2000	1200
100	003	S2	30	1200	750
...					

Store

<u>Store _Code</u>	Store _Name	City	Telephone	Manager
S1	Jan's	San Antonio	683-192-1400	Burgess
S2	Bill's	Portland	943-681-2135	Thomas
S3	Ed's	Boulder	417-196-8037	Perry
...				

Exemple

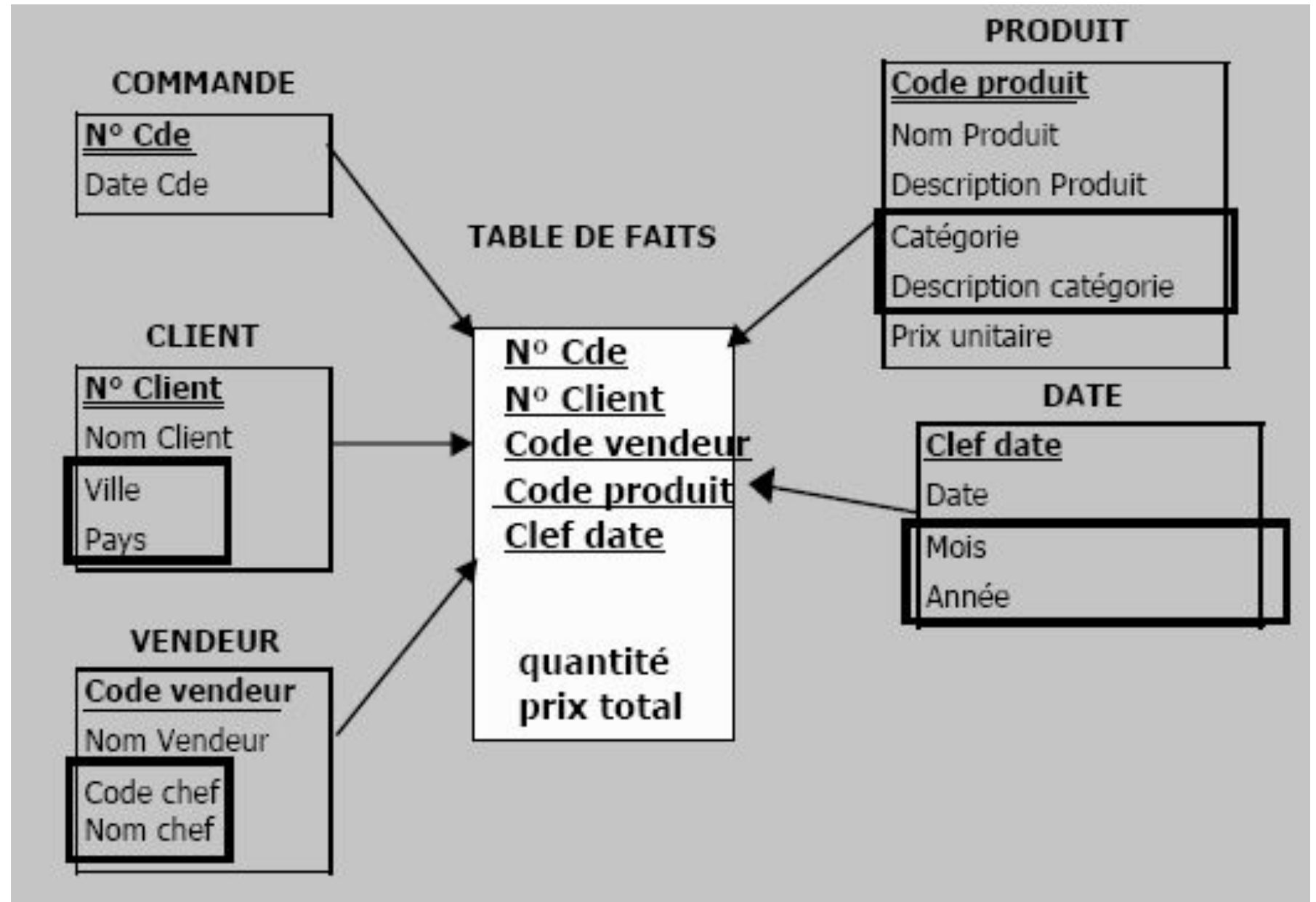


Le modèle en flocon

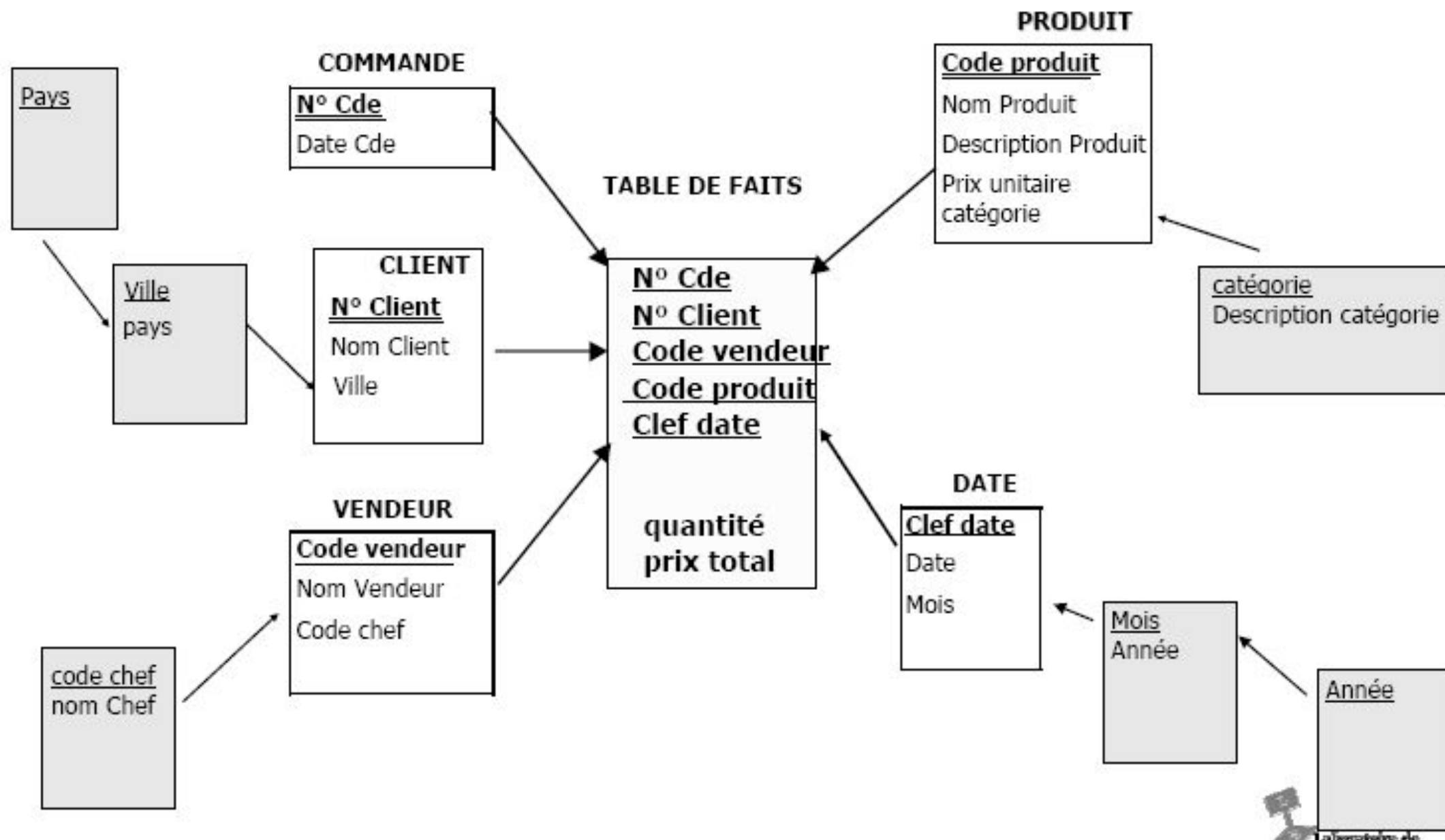
- Le schéma en flocon est dérivé du schéma en étoile où les tables de dimensions sont normalisées (la table des faits reste inchangée).
- Avec ce schéma, chacune des dimensions est décomposée selon sa (ou ses) hiérarchie(s).
- Exemple : Commune, Département, Région, Pays, Continent

Client	Continent	Pays	Region	Département	Commune
Pepone	Europe	France	RhôneAlpes	Rhône	Lyon1
Testut	Europe	France	RhôneAlpes	Rhône	Lyon2
Soinin	Europe	France	RhôneAlpes	Rhône	Lyon3
Vepont	Europe	France	Ile de France	Paris	Paris1
Martin	Europe	France	Ile de France	Paris	Paris2
Elvert	Europe	France	Ile de France	Yvelines	Versailles

Le modèle en flocon



Le modèle en flocon



Le modèle en flocon de neige

- Modèle en étoile + normalisation des dimensions
- Lorsque les tables sont trop volumineuses
- Avantages :
 - réduction du volume,
- Inconvénients :
 - navigation difficile,
 - nombreuses jointures.

Le modèle en constellation

- La modélisation en constellation consiste à fusionner plusieurs modèles en étoile qui utilisent des dimensions communes.
- Un modèle en constellation comprend donc plusieurs tables de faits et des tables de dimensions communes ou non à ces tables de faits.

Le modèle en constellation

