

# Introduction à l'Informatique Décisionnelle et aux entrepôts de données



Bernard ESPINASSE  
Professeur à Aix-Marseille Université (AMU)  
Ecole Polytechnique Universitaire de Marseille



Septembre 2021

- **Émergence de l'Informatique Décisionnelle**
- **Les entrepôts de données**
- **Modélisation et implantation d'un entrepôt de données**
- **Exploitation d'un entrepôt de données (OLAP)**
- **Domaines d'application des entrepôts de données**

## Plan

- 1. Émergence de l'Informatique Décisionnelle (ID)**
  - Pilotage de l'entreprise et informatique décisionnelle (BI)
  - De l'OLTP à l'OLAP
  - De la Business Intelligence à la Business Analytics
  - Un exemple de l'ID/BI dans l'entreprise : la relation client
- 2. Les entrepôts de données**
  - Définition d'un entrepôt de données
  - De l'entrepôt à l'aide à la décision
  - Entrepôts de données Versus Magasins de données
  - Architecture fonctionnelle d'un entrepôt de données
- 3. Modélisation et implantation d'un entrepôt de données**
  - Modélisation multidimensionnelle
  - De la table au cube
  - Stratégies d'implantation : ROLAP, MOLAP et HOLAP
  - Schéma en étoile, en flocon et en constellation
- 4. Exploitation d'un entrepôt de données**
  - Problématique de l'OLAP, OLAP versus OLTP
  - Les opérations élémentaires et langages de l'OLAP
  - Reporting, tableaux de bords et visualisation autour de l'OLAP.
- 5. Domaines d'application des entrepôts de données**

## 1 – Émergence de l'Informatique Décisionnelle (ID)

- **Pilotage de l'entreprise et informatique**
- **Définition de l'ID (*Business Intelligence - BI*)**
- **Domaines d'application de l'ID dans l'entreprise**
- **De la Business Intelligence à la Business Analytics**
- **Un exemple de l'ID/BI dans l'entreprise : la relation client**

## Pilotage de l'entreprise et informatique

- Le **pilotage** d'une entreprise dépend de ses **objectifs stratégiques**
- **Ce pilotage** doit prendre en **considération** :
  - une organisation de plus en plus **orientée clients**,
  - des **cycles conception/fabrication** de plus en plus **courts**,
  - de **nouveaux canaux de distribution** comme les ventes en ligne sur le Web,
  - l'exigence d'**internationalisation**,
  - ...
- **Dans ce contexte l'entreprise se doit** :
  - d'**anticiper les besoins des clients**,
  - de **contrôler** l'intégrité et la qualité des **flux de gestion**,
  - d'évaluer la **performance** des différentes entités la composant,
  - ...

## Outils informatiques supportant le pilotage des entreprises

### Outils d'entreposage et d'analyse :

- pour **constituer** et **mettre à jour** à partir de diverses sources des « **réservoirs** » de grande quantités de données **historisées** et **multidimensionnelles**, ...
- pour en **extraire** selon divers **critères** des sous-ensembles de données,
- pour les **analyser** selon **différents axes** (OLAP), d'**identifier** des **tendances**, des **corrélations**, faire de la **prévision** (Data Mining).

### Outils de veille stratégique :

- rattachés à « **l'intelligence économique** » (Competitive Intelligence)
- pour la **collecte** sur le Web d'importante quantité de données, leur **filtrage** et en **extraire** les informations **pertinentes** (Web Mining) pour les analyser ensuite

=> Outils relevant de **l'Informatique Décisionnelle (ID) ou Business Intelligence (BI)**,

## Informatique Décisionnelle (ID/BI) : définition

- **L'Informatique Décisionnelle (ID)**, en anglais **Business Intelligence (BI)**, est l'informatique à l'usage des décideurs et dirigeants des entreprises

En **management**, elle permet une **connaissance approfondie de l'entreprise** et la **définition** et le **soutien** de **stratégies d'affaires**, par exemple :

- d'acquérir un avantage concurrentiel,
- d'améliorer la performance de l'entreprise,
- de répondre plus rapidement aux changements,
- d'augmenter la rentabilité, et
- d'une façon générale la création de valeur ajoutée de l'entreprise.

- Mais les techniques de ID/BI sont utilisées aussi dans **d'autres domaines que le management** : santé, sciences, énergie, télécommunication, ...

## L'ID/BI est cruciale et d'actualité

### • Le Web rend l'ID/ BI encore plus nécessaire :

- les clients ne sont **pas «physiquement» dans le magasin**
- les clients peuvent **changer à d'autres magasins plus facilement**
- comment **connaître ses clients** :
  - **analyser les « Web log »** pour comprendre le comportement des clients sur le site
  - **combiner ces données Web** avec les **données traditionnelles** des clients

### • « Internet sans fil » ajoute à cela :

- les **clients sont toujours "en ligne"**
- la **position de la clientèle est connue**
- **combiner la position et la connaissance sur le client** => très utile

⇒ Les entreprises sont toujours à la recherche de **cadres pour développer l'ID/BI** pour **l'amélioration** de leurs performances et leur **transformation**

## Systèmes d'information opérationnels : OLTP

- Permettent des processus de **traitement en ligne des données – OLTP (On line Transactional Processing)** : Interactifs, Concurrents, Nombreux, Répétitifs, Structurés, Simples
- **Supportent en général une ou plusieurs grandes fonctions de l'entreprise** (production, marketing, commercial, RH, finance, comptabilité, recherche, ...)
- Exploite des **BD Relationnelles** (Oracle, DB2, ...), parfois intégrés dans un **ERP**, gérer des **BD « opérationnelles »** ou de « **production** » (Mega-Giga octets)
- **Ces processus OLTP concernent** :
  - la mise à jour de données
  - un nombre restreint d'enregistrements
  - des données précises et à jour

⇒ Exemple : un supermarché **ENREGISTRANT** ses ventes

## Limites des BD opérationnelles

- **Modèles de BD opérationnelles complexes et inutilisables :**
  - souvent **difficiles à comprendre**
  - ne concerne **pas un objectif unique** d'affaire
- **Données des BD opérationnelles :**
  - **identiques** dans **différentes BD**
  - **même concept** souvent **défini différemment**
  - **adaptées pour les systèmes opérationnels** (comptabilité, facturation, ...), pas pour l'analyse des fonctions d'affaires
  - **volatiles :**
    - supprimées périodiquement dans les systèmes opérationnels (6 mois)
    - modifiées au fil du temps - aucune information historique

## Nouveaux attentes des SI

- Considérer des **quantités de données HISTORISEES de plus en plus importantes** (Tera, Penta octets), **organisées selon différentes dimensions** (temps, espace géographique, gammes de produit, ...) stockées dans des **ENTREPOTS DE DONNEES**
- Passer du **TRAITEMENT transactionnel en ligne des données (OLTP)** à **l'ANALYSE EN LIGNE** (*On Line Analysis Processing - OLAP*) de ces entrepôts

Pour prendre de « **bonnes décisions** », il faut accéder en **temps réel** à ces données, les analyser pour en extraire l'information pertinente, par exemple pour savoir :

- *Quels sont les résultats des ventes par gamme de produit et par région pour l'année dernière ?*
- *Quelle est l'évolution des chiffres d'affaires par type de magasin et par période ?*
- *Comment qualifier les acheteurs de mon produit X ?*
- ...

⇒ **Informatique Décisionnelle (Business Intelligence)**

## Nouvelle technologie informatique de l'ID/BI

### Entrepôt de données (Data Warehouse) :

- il récolte, **stocke et gère efficace des gros volumes de données** pour la prise de décision,
- les données y sont **organisées dans des regroupements homogènes** selon plusieurs **axes d'analyse** et différents **niveaux de détail** (d'agrégation).

### Analyse en ligne des données OLAP (On Line Analytical processing) :

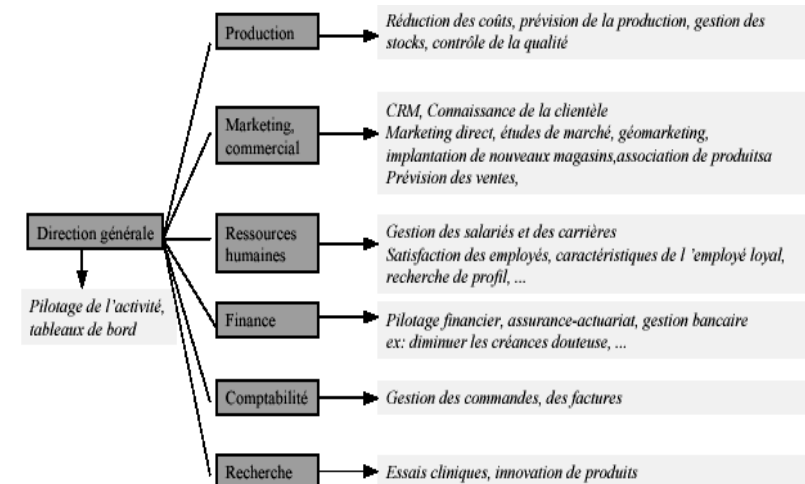
- exploitation d'entrepôts de données permettant **interactivement** de conduire des analyses par **changement de points de vue, de niveau de détail ...**

### Fouille de données (Data Mining):

- **extraction automatique de connaissances** (propriétés cachées) dans de grands volumes de données,
- par des **techniques traditionnelles** issues des *statistiques* et de *l'analyse de données*,
- par des **techniques plus récentes** comme *l'induction d'arbre de décision, les algorithmes génétiques, les réseaux de neurones, ... (Intelligence Artificielle)*.

⇒ **Tendance à une intégration croissante des techniques de fouille dans les entrepôts de données (offre commerciale)**

## L'ID/BI dans l'entreprise : domaines concernés



## De la Business Intelligence à la Business Analytics (1)

### Business Intelligence (BI) :

- s'intéresse à ce qui s'est passé
  - basée sur les *entrepôts de données, l'analyse en ligne (OLAP), le reporting, la surveillance et l'alerte automatisées, les tableaux de bord ...*
- ⇒ répondre à des questions telles que : que s'est-il passé, combien, à quelle fréquence, où se situe le problème et quelles sont les actions nécessaires.

### Business Analytics (BA) :

- s'intéresse aux raisons pour lesquelles cela s'est produit et si cela se reproduira, en s'appuyant sur la *Business Intelligence*
  - basée sur l'analyse statistique et quantitative, l'exploration de données, la modélisation prédictive ...
- ⇒ répondre à des questions telles que : pourquoi cela se produit ?, que se passe-t-il si ces tendances se poursuivent ?, que se passera-t-il ensuite ? (prédictions) et quel est le meilleur résultat possible ? (optimisation).

## Exemple de l'ID dans l'entreprise : la relation client (1)

### Contexte économique général

#### Dans un climat de concurrence mondiale tendu :

- Conquérir un nouveau client coûte 5 fois plus cher que de fidéliser un client existant
- 5% d'amélioration de la fidélité des clients entraîne une augmentation des profits de 10 à 15%
- Tous les clients ne sont pas égaux : 30% des clients génèrent 70% du CA
- Le Client attend un service personnalisé, sur-mesure
- La personnalisation est une source de profit

⇒ Gestion de la relation client  
(Customer Relationship Management – CRM)

## Exemple de l'ID dans l'entreprise : la relation client (2)

### Définition du CRM :

- capacité à identifier, à acquérir et à fidéliser les meilleurs clients dans le but d'augmenter le chiffre d'affaires et les bénéfices.
- capacité à bâtir une relation profitable sur le long terme avec les meilleurs clients en capitalisant sur l'ensemble des points de contacts

### Mieux connaître et comprendre ses clients pour :

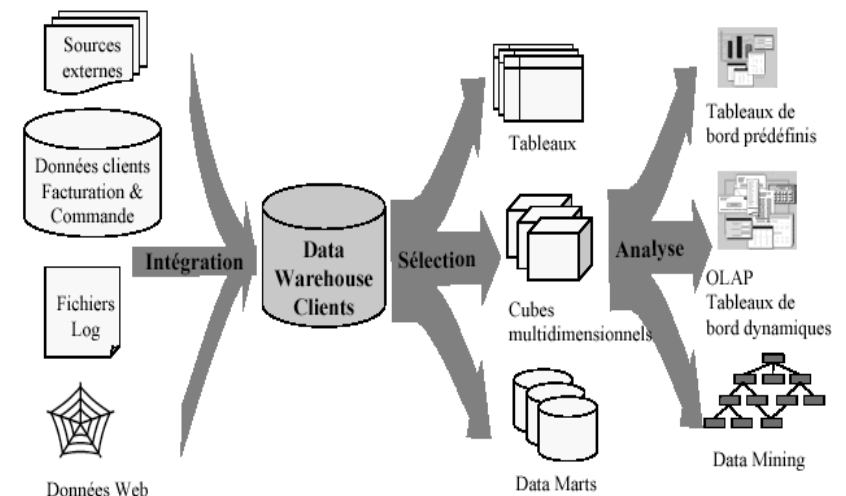
- Réduire les coûts (clients infidèles)
- Comment fidéliser une clientèle ?
- Comment augmenter les profits ?
- Comment identifier les nouvelles opportunités ?

### ... Doit permettre par exemple de répondre aux questions :

- Quels sont les besoins et les attentes des clients? Comment y répondre?
- Quels sont les clients prêts à acheter de nouveaux produits?
- Quels sont les clients les plus profitables, fidèles et pourquoi?
- Quels sont les clients mécontents, et pourquoi?
- ...

⇒ Informatique décisionnelle (BI)

## Exemple de l'ID dans l'entreprise : la relation client (3)



## 2 – Les entrepôts de données

- Définition d'un entrepôt de données
- De l'entrepôt à l'aide à la décision
- Entrepôts de données Versus Magasins de données
- Architecture fonctionnelle d'un entrepôt de données

## Définition d'un entrepôt de données (Data Warehouse)

### Définition de Inmon (1992) :

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

### Données :

- **Thématique ou orientées sujet** : un ED rassemble et organise des données issues de l'entreprise et de son environnement, pertinentes pour un sujet particulier à analyser
- **Intégrées** : les données résultent de l'intégration de données provenant de différentes sources pouvant être hétérogènes
- **Historisées** : les données représentent l'activité d'une entreprise durant une certaine période (plusieurs années) permettant de d'analyser les variations d'une donnée dans le temps
- **Non-volatiles** : les données de l'ED sont essentiellement utilisées en interrogation (consultation) et ne peuvent pas être modifiées (sauf certain cas de rafraîchissement).

## De l'entrepôt à l'aide à la décision (1)

**Entreposage des données** : avant d'être chargées dans l'entrepôt, les données sélectionnées doivent être :

- **extraites des sources** (internes : BD opérationnelles, externes : BD et fichiers notamment issus du Web)
- **soigneusement épurées** afin d'éliminer des erreurs et réconcilier les différentes sémantiques associées aux sources)

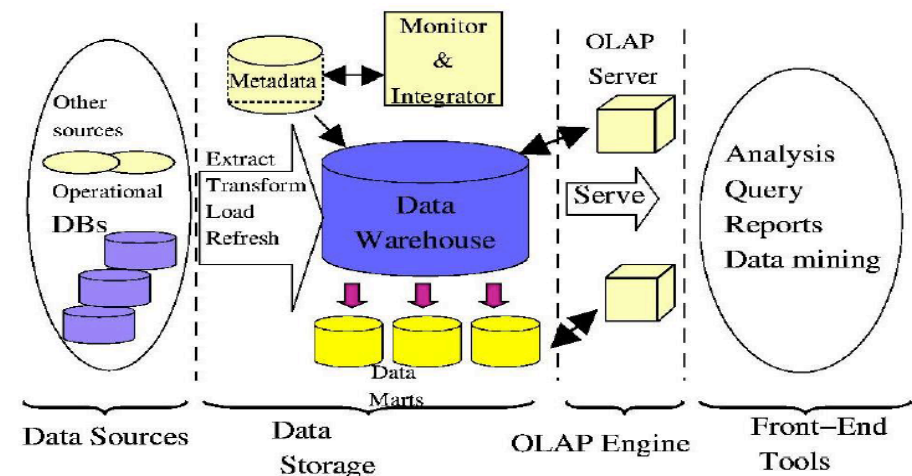


### Exploitation des données de l'ED : systèmes décisionnels

- A partir des données d'un ED **diverses analyses** peuvent être faites, notamment par des techniques « On-Line Analytical Processing » (**OLAP**) ou de **fouille de données** (Data Mining) et de **visualisation**.

⇒ Généralement les informations et connaissances obtenues par exploitation d'un ED ont un **impact direct sur les bénéfices de l'entreprise** (augmentation des ventes par un marketing plus ciblé, amélioration de la rotation des stocks, ...)

## De l'entrepôt à l'aide à la décision (2)



## Entrepôt de données VS Bases de données opérationnelles

	BD opérationnelles	Entrepôt de données
<b>Niveau de détail des informations</b>	▪ Très détaillé	▪ Données agrégées, métadonnées
<b>Homogénéité des informations</b>	▪ Informations homogènes	▪ Information pas nécessairement homogènes, ▪ intégration de données souvent nécessaire
<b>Fonctions de l'entreprise concernées par les données</b>	▪ Données organisées par processus fonctionnel	▪ Données orientées sujet
<b>Comparaison de données sur plusieurs années</b>	▪ Non : Archivage ou mise à jour des données	▪ Oui : Données non volatiles, données historisées
<b>Opérations réalisées sur les données</b>	▪ Consultation, mais surtout mise à jour et ajout de données	▪ Consultation de données uniquement

## Entrepôts de données Versus Magasins de données

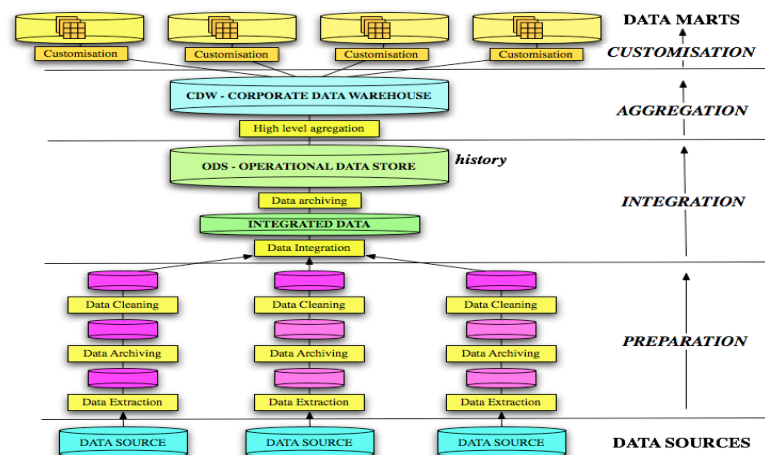
### L'entrepôt de données - ED (Data Warehouse - DW) :

- nécessitent de **puissantes machines** pour gérer de **très grandes bases de données** contenant des **données de détail historisées**
- lieu de **stockage centralisé** d'un extrait des bases de production.
- l'**organisation des données** est faite selon un modèle **facilitant la gestion efficace des données et leur historisation**.

### Les magasins de données – MD (Data Marts - DM) :

- **petits entrepôts** nécessitant une **infrastructure plus légère** et sont mis en œuvre plus **rapidement** (6 mois environs)
- **conçus pour l'aide à la décision** à partir de **données extraites d'un ED** plus conséquent ou de BD sources existantes
- les **données extraites sont adaptées pour l'aide à la décision** (pour classe de décideurs, usage particulier, recherche de corrélation, logiciel de statistiques,...)
- l'**organisation des données** est faite selon un **modèle facilitant les traitements décisionnels**

## Entrepôts et magasins de données (3)



- **ODS Operational Data Store** : regroupe les données intégrées récupérées des sources
- **CDW Corporate Data Warehouse** : regroupe les vues agrégées

## Architecture fonctionnelle d'un entrepôt de données

### Niveau extraction (Préparation) : Extraction de données des BD opérationnelles (SGBD traditionnel en OLTP) et de l'extérieur :

- **approche « push »** : détection **instantanée** des mises à jour sur les BD opérationnelles pour intégration dans l'ED
- **approche « pull »** : détection **périodique** des mises à jour des BD opérationnelles pour intégration dans l'ED

### Niveau integration :

- **Intégration, chargement et stockage** des données dans la BD entrepôt organisée par sujets
- **Rafraîchissement** au fur et à mesure des mises à jour

### Niveau exploitation :

- **Customisation** : Data Marts (Magasins de données) ...
- **Rapports, tableaux de bords, visualisation graphiques diverses, ...**
- **Analyse et l'exploration** des données entreposées (**OLAP**)
- **Fouille de données** (Data Mining) pour découverte de connaissances, ...

### 3 – Modélisation et implantation d'un entrepôt de données (ED)

- Modélisation multidimensionnelle
- De la table au cube
- Stratégies d'implantation
- Schéma en étoile, en flocon et en constellation

### Modélisation multidimensionnelle (1)

- Les analyses décisionnelles (**OLAP**) sont directement reliées à une **modélisation de l'information spécifique** :
  - **proche de la perception qu'en a l'analyste**
  - basée sur une **vision multidimensionnelle des données**
- **Modélisation « multidimensionnelle »** :
  - considère un sujet analysé comme un **point dans un espace à plusieurs dimensions**
  - les **données y sont organisées** de façon à **mettre en évidence le sujet analysé** et les différentes **perspectives de l'analyse**.

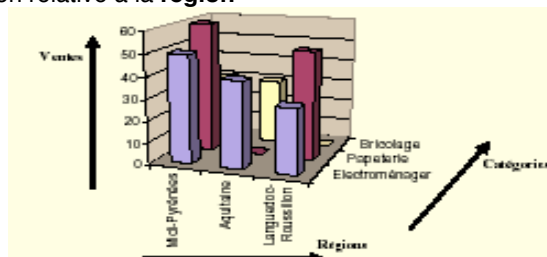
### Modélisation multidimensionnelle (2)

Soit les données relatives aux ventes de 1999 d'une entreprise de distribution :

Catégories des produits	Régions	Montant des ventes
Electroménager	Midi-Pyrénées	50
Electroménager	Aquitaine	40
Electroménager	Languedoc-Roussillon	30
Papeterie	Midi-Pyrénées	60
Papeterie	Languedoc-Roussillon	50
Bricolage	Midi-Pyrénées	30
Bricolage	Aquitaine	30

Différentes perspectives pour observer ces données :

- une dimension relative à la **catégorie des produits**
- une dimension relative à la **région**



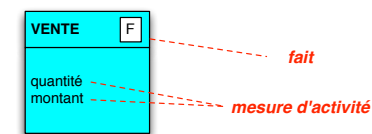
### Modélisation multidimensionnelle : concept de fait

**Un fait :**

- modélise le **sujet** de l'analyse
- est formé de **mesures** correspondant aux informations de l'activité analysée.
- ces mesures sont **numériques** et généralement **valorisées de façon continue**, on peut les **additionner**, les **dénombrer** ou bien **calculer** le minimum, le maximum ou la moyenne.

**Exemple :** le fait de « Vente » peut être constitué des mesures d'activités suivantes :

- quantité de produits vendus et
- montant total des ventes



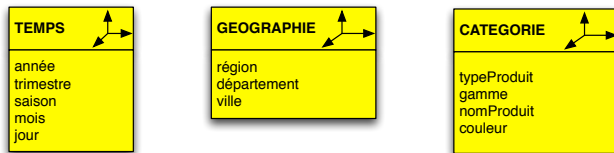
## Modélisation multidimensionnelle : concept de dimension

Le sujet analysé, le fait, est **analysé** suivant **différentes perspectives** ou **axes** caractérisant ses mesures de l'activité : on parle de **dimensions**.

### Une dimension :

- modélise un **axe d'analyse**
- se compose de **paramètres** correspondant aux informations faisant varier les mesures de l'activité.

**Ex:** Dans l'exemple précédent, le fait « **Vente** » peut être analysé suivant différentes perspectives correspondant à trois dimensions : la dimension **Temps**, la dimension **Géographie** et la dimension **Catégorie** :



## Modélisation multidimensionnelle : hiérarchie de dimension

- Les faits sont **analysés** selon les dimensions qui les caractérisent
- Nécessaire de définir pour **chaque dimension** ses **différents niveaux hiérarchiques de détail** (d'agrégation),
- Les **hiérarchies** de dimensions définissent des **niveaux de détail de l'analyse** sur les dimensions

### Ex:

- **Dimension « temps »** : H1 : jour → mois → trimestre → année ; H2 : jour → mois → trimestre → année ; H3 : jour → mois → saison → année ;
- **Dimension « géographie »** : ville → département → région (chaque ville appartient à un département qui est situé dans une région)
- **Dimension « catégorie »** : couleur → nomProduit → gamme → typeProduit (chaque produit appartient à une gamme de produit qui appartient à un type de produit)

## De la table ... au cube

### De la table ...

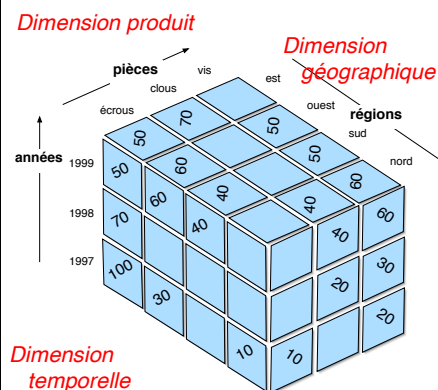
Table Ventes :

VENTES	pièces	Régions	Années	quantités
	écrous	est	1999	50
	clous	est	1997	100
	vis	ouest	1998	50
	...	...	...	...
	écrous	est	total	220
	...	...	...	...
	écrous	total	total	390
	...	...	...	...
	total	total	total	1200

(pièce, région, année) → quantité

### ... au cube

Cube Ventes :



## Stratégies d'implantation d'un ED

### 3 stratégies :

#### 1 - Usage d'un SGBD Relationnel (systèmes ROLAP)

- les **SGBDR** représentant **plus de 80% des SGBD** : ils sont principalement envisagés pour le développement d'ED mais doivent être adaptés
- Ils doivent cependant **être adaptés** car ils n'ont pas les caractéristiques adéquates pour répondre aux besoins des ED.

#### 2 - Usage d'un SGBD Multidimensionnel (systèmes MOLAP)

- Un SGBD Multidimensionnel (SGBDM) est un **SGBD capable de stocker et traiter des données multidimensionnelles**
- A ce jour **pas encore de cadre technologique commun** pour le développement de tels systèmes : chaque produit est spécifique

#### 3 - Usage d'un SGBD Hybride (systèmes HOLAP)

Tire profit des avantages des technologies ROLAP et MOLAP :

- un **ROLAP** pour stocker, gérer les **données détaillées ET**
- un **MOLAP** pour stocker, gérer les **données agrégées**



## Schéma d'un entrepôt de données

Niveau logique « ROLAP » :

3 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 en étoile (1)

- **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

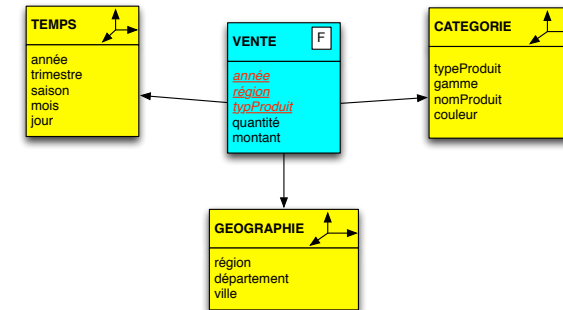
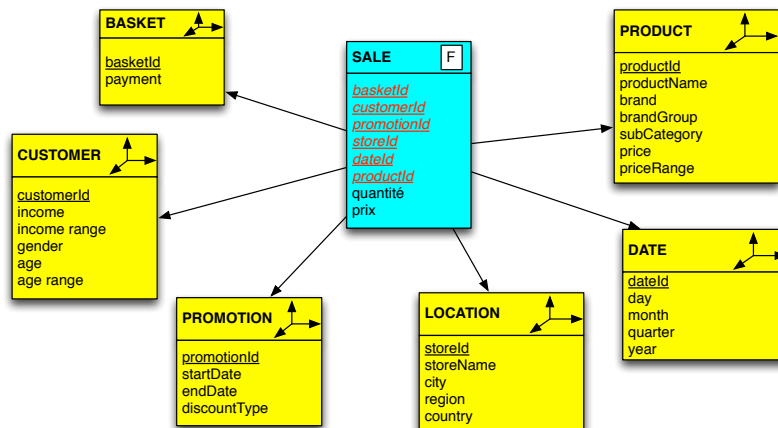


Schéma en étoile modélisant les analyses des quantités et des montants des médicaments dans les pharmacies selon 3 dimensions : le temps, la catégorie et la situation géographique

- Table de **faits** : **Vente**
- Tables de **dimension** : **Temps, Catégorie, Géographie**

## Schéma en étoile (2)

Ex 2 : Ventes d'articles dans un supermarché



## Schéma en étoile (3)

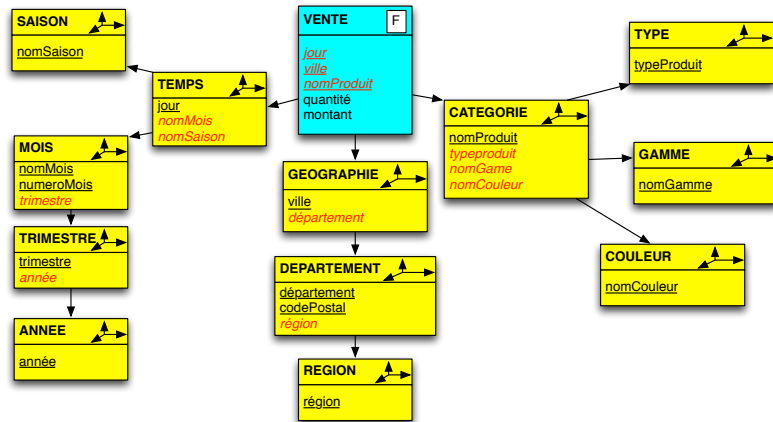
Associé à Ex 2 :

- **un fait** :
  - il a été acheté **3 exemplaires à 1 euro** (SALE)
  - du **produit pid3**
  - par le **client cid1**
  - à la **date did3**
  - dans le **magasin mid2** (store)
  - dans le **chariot cid8** (basket)
  - correspondant à la **promotion prid1**
- **un élément de la dimension location** :
  - store id **mid2**
  - store name **rondpoint**
  - city **blois**
  - region **centre**
  - country **France**

## Schéma en flocon

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

Ex 3: Vente de médicaments dans des pharmacies

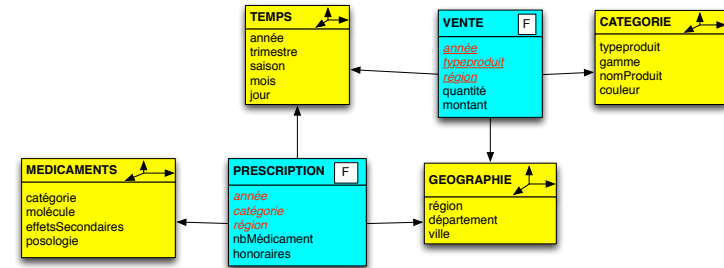


Chaque dimension du schéma en étoile précédent est **dénormalisée**

## 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
- les dimensions **Temps** et **Géographie** sont partagées par les faits PRESCRIPTION et VENTE.

## 4 – Exploitation d'un entrepôt de données (ED)

- OLAP, OLAP versus OLTP
- Les opérations élémentaires de l'OLAP et langages pour l'OLAP
- Réalisation de rapports divers (*Reporting*)
- Réalisation de tableaux de bords à partir d'un ED (*Dashboards*)
- Visualisations autour d'un ED (*visualizations*)

## OLAP (On Line Analytical Processing)

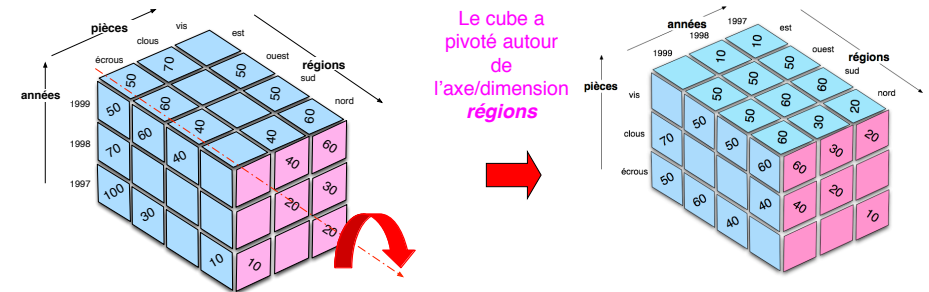
- Apparue dans les années 90 dans les entreprises
- Façon la plus naturelle d'exploiter un ED du fait de son organisation multidimensionnelle - cube
- Permettent de réaliser des synthèses, des analyses et de la consolidation dynamique de données multidimensionnelles organisés en cubes
  - Exemple d'analyse OLAP : un supermarché ANALYSANT l'ensemble de ses ventes
- 3 types d'opérations OLAP élémentaires liées à des transformations du cube :
  - Restructuration** : permet un changement de points de vue selon différentes dimensions : opérations liées à la structure, manipulation et visualisation du cube : **Rotate/pivot, Switch, Split, nest, push, pull**
  - Granularité** : concerne un changement de niveau de détail : opérations liées au niveau de granularité des données : **roll-up, drill-down**
  - Ensemble** : concerne l'extraction et l'OLTP classique : **slice, dice, selection, projection, jointure (drill-across)**

## OLTP versus OLAP

	Caractéristiques	OLTP	OLAP
<b>Conception</b>	Orientation Conception	Transaction Entité-Relation	Analyse Star/snowflake
<b>Données</b>	Granularité	Détail	Résumées, agrégées
	Nature	Relationnelle	Multidimensionnelle
	Actualisation	Actualisées, mises à jour	Historisées, recalculées
	Taille fichiers	< 100 Mo/Go	> 100 Go/To
<b>Traitements</b>	Unité de travail	Transaction simple	Requête complexe
	Accès	Lecture/écriture	Lecture
	Nb de tuples accédés	Dizaines	Millions
	Métrique	Débit de transactions	Temps de réponse
<b>Utilisateurs</b>	Utilisateur	Agent opérationnel	Analyste/décideur
	Nombre d'utilisateurs	Milliers	Centaines

## Exemple d'opérations OLAP de restructuration : rotate/pivot

**Rotate/pivot** : effectue au cube une rotation autour d'un de ses 3 axes passant par le centre de 2 faces opposées, de façon à présenter un ensemble de faces différent (sélection de faces)



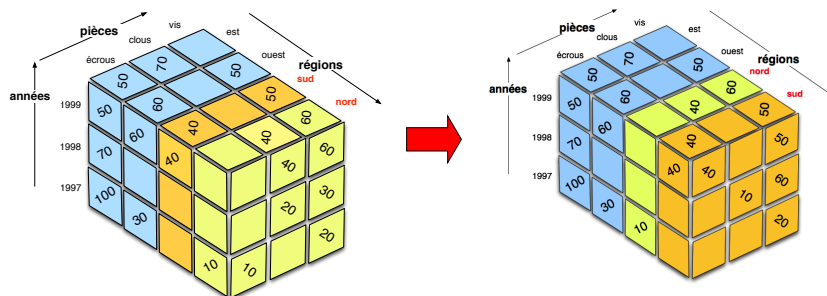
la visualisation résultante est souvent 2D :

	vis	1999	1998	1997
nord		60	30	20
vis		40	20	10
clous				
écrous				

	vis	1999	1998	1997
est		10	10	
ouest		50	50	50
sud		50	60	60
nord		60	30	20

## Exemple d'opérations OLAP de restructuration : switch

**Switch ou permutation** : consiste à interchanger la position des membres d'une dimension :



Ici sont interchangés les membres nord et sud de la dimension régions

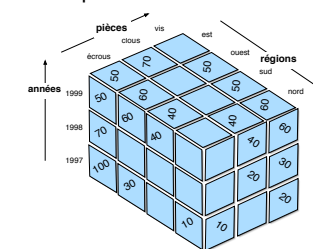
la visualisation résultante est souvent 2D :

	1999	1998	1997
nord	60	30	20
vis	40	20	10
clous			
écrous			

	1999	1998	1997
sud	50	60	60
vis	50	10	10
clous	40	20	
écrous			

## Exemple d'opérations OLAP de restructuration : split

**Split ou division** : consiste à présenter chaque tranche du cube et de passer de sa présentation tridimensionnelle à sa présentation sous la forme d'un ensemble de tables.



ici un split(region) du cube Ventes conduit aux 4 tables suivantes :

ventes est	1999	1998	1997
écrous	50	70	100
vis		10	10
clous	70	70	100

ventes ouest	1999	1998	1997
écrous		10	30
vis	50	50	50
clous		10	40

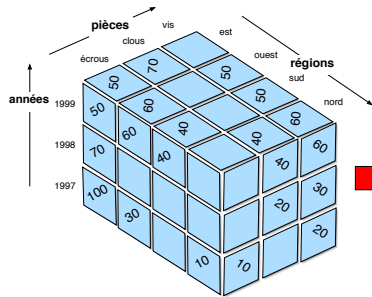
ventes sud	1999	1998	1997
écrous	40	20	
vis	50	60	60
clous		10	

ventes nord	1999	1998	1997
écrous			10
vis	60	30	20
clous	40	20	

## Exemple d'opérations OLAP de restructuration : nest

**Nest ou l'emboîtement:** permet d'imbriquer des membres à partir du cube, de grouper sur une même représentation bi-dimensionnelle toutes les informations (mesures et membres) d'un cube quelque soit le nombre de ses dimensions.

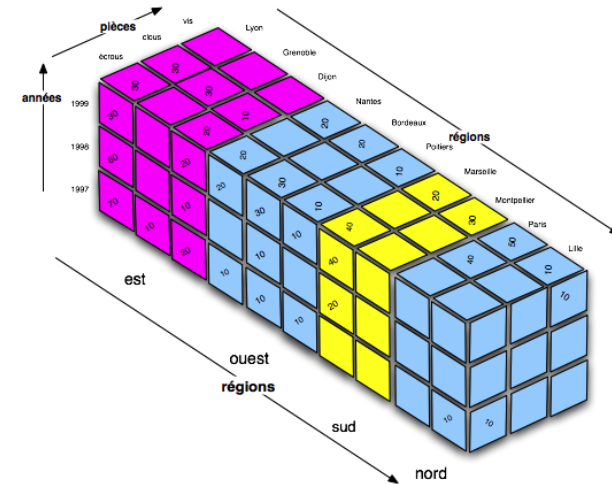
nest(pièces, région) :



ventes nest		1999	1998	1997
écrous	est	50	70	100
	ouest		10	30
	nord			10
	sud	40	20	
vis	est		10	10
	ouest	50	50	50
	nord	60	30	20
	sud	50	60	60
clous	est	70	70	100
	ouest		10	40
	nord	40	20	
	sud		10	

## Exemple d'opérations OLAP de de granularité : drill-down

Drill-down du niveau des régions au niveau villes : Drill-down(regions) :



## Langages pour l'OLAP

2 langages possibles pour faire de l'OLAP :

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

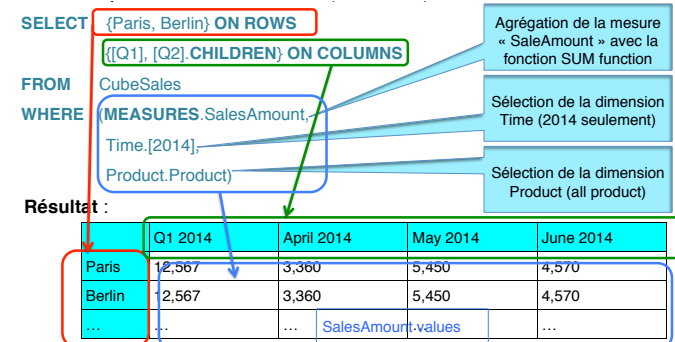
- Nouvelles fonctions SQL d'agrégation: *Rank, N\_tile, ...*
- Nouvelles fonctions de la clause GROUP BY :
  - *ROLLUP* equivalent to "control breaks"
  - *CUBE* equivalent to "cross tabulation"
  - *GROUPING SETS* equivalent to multiple GROUP BYs
- Fenêtre glissante : *WINDOWS/OVER/PARTITION, ...*

### 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

## Exemple de requête OLAP en MDX :

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



## OLAP et reporting

- Pour des utilisateurs qui ont besoin d'un accès régulier à des informations d'une manière presque statique

Ex: les hôpitaux doivent envoyer des rapports mensuels à des agences nationales

- Un rapport est défini par une requête (plusieurs requêtes) et une mise en page (diagrammes, histogrammes, etc)
- Les rapports peuvent être exécutés automatiquement ou manuellement

PROGRIE2	SICRIE2	SDOP	BUDGET 2010	PREVIO	REALE	DIFFEREN
3770	28,4%			272	272	
	-1,00			0,00%	0,00%	
3780	2,5%			202	202	
	0,50%			0,00%	0,00%	
1.198	19,00%	I		1.000	1.000	
	0,00%			0,00%	0,00%	
44.058	6,50%			44.050	44.050	
	1,40%			3,00%	3,00%	
2.202	0,5%			2.200	2.200	
	0,00%			0,00%	0,00%	
8.040	10,9%			8.040	8.040	
	0,00%			0,00%	0,00%	
3.27	4,8%			3.27	3.27	
	0,00%			0,00%	0,00%	
6.05	8,7%			6.05	6.05	
	0,00%			0,00%	0,00%	
601.000	27,7%			601.000	601.000	
	0,00%			0,00%	0,00%	
3.040	4,3%			3.040	3.040	
	0,00%			0,00%	0,00%	
31.1.027	31,4%	I		31.1.027	31.1.027	
	0,00%			0,00%	0,00%	
1.039.480	14,8%			1.039.480	1.039.480	
	0,44%			0,44%	0,44%	
56.000	0,8%			56.000	56.000	
	0,00%			0,00%	0,00%	
70.000	1,0%			70.000	70.000	
	0,00%			0,00%	0,00%	
4.000	0,6%			4.000	4.000	
	0,00%			0,00%	0,00%	
133.000	1,8%			133.000	133.000	
	0,00%			0,00%	0,00%	
3.000	0,0%			3.000	3.000	
	0,00%			0,00%	0,00%	
1.000	0,1%			1.000	1.000	
	0,00%			0,00%	0,00%	
100.000	1,4%			100.000	100.000	
	0,00%			0,00%	0,00%	
500	0,1%			500	500	
	0,00%			0,00%	0,00%	
1000	0,1%			1000	1000	
	0,00%			0,00%	0,00%	
500	0,1%			500	500	
	0,00%			0,00%	0,00%	

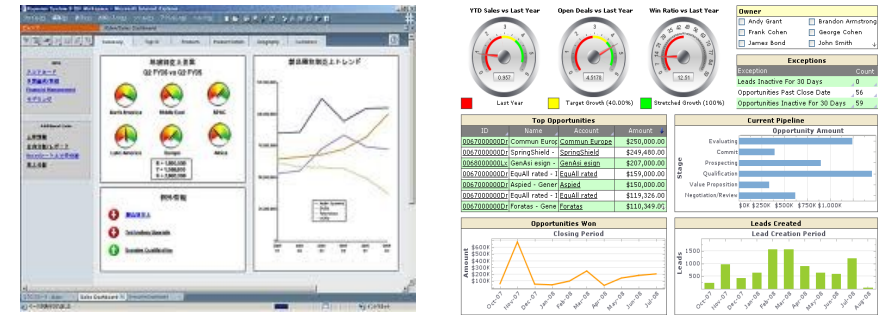
Totale consumi		484.411	419.551	0%	0
costi personale (non da pianificare)	30,0%	994.634			
unità personale	30,0%	5,75	7,00	0%	0,00
presenza media	31,0%	6,18			
Tasso utilizzo letti	31,0%	54,10%	51,31%	0%	0,00%
degenza media	31,0%	3,33	3,16	0%	0,00
lasso sp	30,0%	70,76%	87,23%	0%	0,00%
% ricoveri di 1 giorno	31,0%	7,84%	2,00%	50%	0,00%
peso medio org	30,0%	0,86	0,81		
mobilità provinciale passiva	30,0%	151.280	0	0%	0
mobilità provinciale attiva	30,0%	880.010	0		
mobilità intraruck	30,0%	30.775	0		
servizio trasporti n° pazienti	31,0%	32	0	0%	0
servizio trasporti €	31,0%	1.560	0		
Summa Gewichtung				100%	

## OLAP et tableaux de bords (Dashboards)

- Affichent une quantité limitée d'informations dans un format graphique facile à lire
- Fréquemment utilisé par les cadres supérieurs qui ont besoin d'un rapide aperçu des changements les plus importants

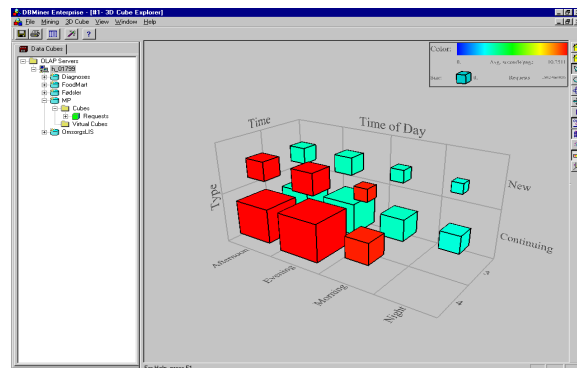
Ex : un aperçu en temps réel d'évolutions

- Pas vraiment utile pour une analyse complexe et détaillée



## OLAP et visualisation de données

- Facilitent l'analyse et l'interprétation de données
- convertissent des données complexes en images, graphiques en 2 et 3 dimensions, voire en animations
- Sont de plus en plus intégrées dans les ED



## 5 – Exemple de domaines d'application des entrepôts de données

- Domaine bancaire
- Domaine de la grande distribution
- Domaine des télécommunications
- Domaines de l'assurance et de la pharmacie
- Domaine de la santé, ...

## Domaine bancaire : un des premiers utilisateurs des ED

- Pour une banque, il est important de pouvoir **regrouper les informations relatives à un client** afin de répondre à ses demandes de crédit par exemple
- **Des mailing ciblés doivent aussi être rapidement élaborés** à partir de toutes les informations disponibles sur un client lors de la commercialisation d'un nouveau produit
- **L'utilisation de cartes de crédit nécessite des contrôles à posteriori**, par exemple pour la recherche de fraudes : la mémorisation des mouvements peut rendre de grands services
- **Les échanges d'actions et de conseils de courtages sont facilités** par une mémorisation de l'histoire et une exploitation par des outils décisionnels avancés par exemple pour déterminer des tendances de marchés

## Domaine de la grande distribution : fortement demandeur d'ED

- Intéressant de **regrouper les informations de ventes** pour déterminer les **produits à succès**, mieux suivre les **modes**, détecter les **habitudes d'achats**, les **préférences des clients** par **secteur géographique**
- La **fouille de données** (Data Mining) a permis de développer des techniques sophistiquées d'exploitation de données qui aident à **mettre en évidence les règles de consommation**
- **Explorer le panier de la ménagère** est devenu un exercice d'école : il s'agit de trouver à partir de l'enregistrement des transactions quelles sont les habitudes d'achats, plus précisément quels sont les produits achetés en même temps

### Apports constatés dans la grande distribution :

- **augmentation des ventes grâce à un meilleur marketing**
- **amélioration des taux de rotation de stocks**
- **élimination des produits obsolètes**
- **réduction des rabais, remises, ristournes**
- **meilleure négociation des achats**

## Domaine très concurrentiel des télécommunications : utilise beaucoup les ED

- **Grande masse de données** concernant les abonnés et les appels est **enregistrée**
- **Plusieurs mois de description détaillée des appels** comprenant, pour chaque appel appelant, appelé, heure et durée sont disponibles chez les opérateurs

En respectant les lois de sécurité et liberté, que peut-on faire de telles données ?

**Couplées ou non avec des informations comptables**, l'exploitation de ces données regroupées en ED par des techniques d'analyse et d'exploration permet :

- **D'analyser le trafic**,
- **De mieux cerner les besoins des clients**,
- **De classer les clients** par catégories,
- **De comprendre pourquoi certains changent d'opérateurs** et mieux répondre à leurs besoins

## Domaines de l'assurance et de la pharmacie : très friands de techniques décisionnelles

- L'exercice de base de **l'assureur** est de **déterminer le facteur de risque d'un assuré**
- Celui d'un **producteur pharmaceutique** est de **détecter l'impact d'un médicament**
- Plus généralement, le **suivi des informations relatives à la liaison produit-client** sur un ED est souvent synonyme de **gains importants** : meilleure connaissance des produits, détection des défauts, meilleure connaissance des clients, détection de rejets, ciblage du marketing, etc
- ...