

# Gestion des données

## Cours 5 - Information décisionnelle

Olivier Schwander  
<olivier.schwander@sorbonne-universite.fr>

Sorbonne Université

2021-2022

# Business intelligence

## ETL: Extract - Transform - Load

- ▶ Récupérer les données là où elles sont
- ▶ Transformer les données si besoin
- ▶ Stocker les données de façon exploitable

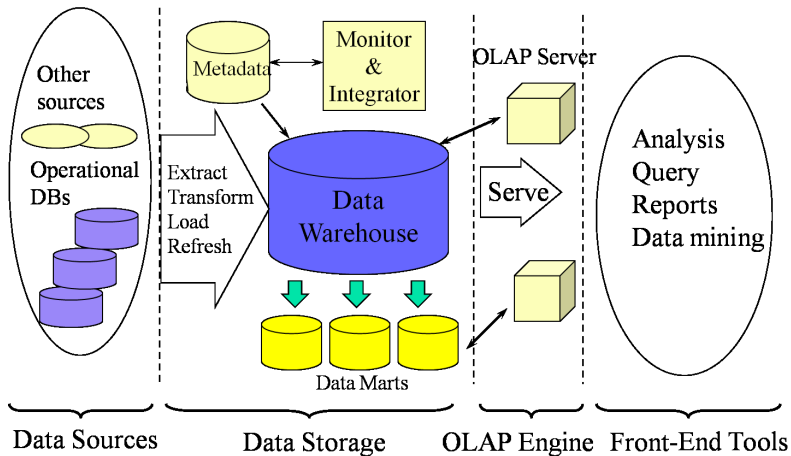
## Data Warehouse

- ▶ Stockage des données
- ▶ Toute l'histoire de l'entreprise
- ▶ Stable dans le temps

## OLAP: Online Analytical Processing

- ▶ Données en grande dimension
- ▶ Visualisation, structuration
- ▶ Pas forcément de traitement statistique compliqué

# Processus



# Stockages

## Base de données opérationnelle

- ▶ Fonctionnement normal de l'entreprise
- ▶ Pas forcément un historique très grand
- ▶ Peut changer dans le temps

## Datawarehouse

- ▶ Stockage pour le BI
- ▶ Archivage sur toute l'histoire de l'entreprise
- ▶ Format stable dans le temps

## Datamart

- ▶ Vue métier
- ▶ À destination du décideur

# Entrepôt de données

## Datawarehouse

Le terme entrepôt de données (ou base de données décisionnelle, ou encore data warehouse) désigne une base de données utilisée pour collecter, ordonner, journaliser et stocker des informations provenant de base de données opérationnelles et fournir ainsi un socle à l'aide à la décision en entreprise.

## Datamart

Un DataMart (littéralement en anglais magasin de données) est un sous-ensemble d'un DataWarehouse destiné à fournir des données aux utilisateurs, et souvent spécialisé vers un groupe ou un type d'affaire.

# Flux de données et données statiques

## Données statiques

- ▶ Image à un instant donné de l'état de l'entreprise
- ▶ Rapports d'activité, bilans, inventaire

## Flux de données

- ▶ Mise à jour en temps réel
- ▶ Comptes rendus quotidiens, commandes, livraisons

## Datawarehouse

- ▶ Archivage sur toute l'histoire de l'entreprise
- ▶ Format stable dans le temps

# Objectifs

## Stockage de toute l'information de l'entreprise

- ▶ **Collecter**: récupérer l'information
- ▶ **Ordonner**: structurer l'information
- ▶ **Journaliser**: conserver l'historique de l'information

# Comparaison

	Base opérationnelle	Entrepôt de donnée
Rôle	Fonctionnement quotidien	Prise de décision
Contenu	Information utile à un moment précis	Historique des données
Utilisateurs	Techniciens, vendeurs	Gestionnaires, analystes
Usage	Requêtes simples, prévisibles	Requêtes complexes, spécifiques
Conception	Performances et disponibilité	Flexibilité et facilité d'accès
Requêtes	Mise à jour fréquentes, sur peu de lignes	Mise à jour périodiques, requêtes



# Données orientées sujets

## Production

- ▶ Données organisées par processus fonctionnels

## Aide à la décision

- ▶ Données organisées par thème
- ▶ À travers plusieurs fonctions, services, départements, etc

# Modèle relationnel

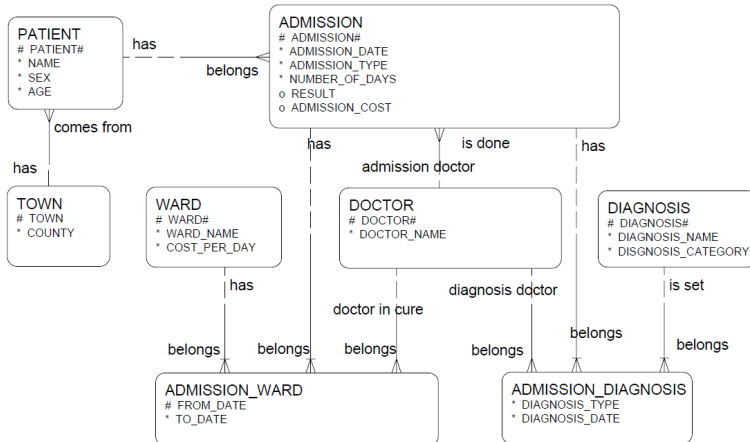
## Avantages

- ▶ Normalisé, garanties de sûreté
- ▶ Requêtes efficaces

## Inconvénients

- ▶ Requêtes peu adaptées pour le BI
- ▶ Beaucoup de relations
- ▶ Trop complexe pour les utilisateurs de BI

# Modèle relationnel



# Modèle dimensionnel

## Analyse du besoin

- ▶ Que veut l'utilisateur ?
- ▶ Sur quoi l'analyse porte-t-elle ?
- ▶ Quels sont les objets à étudier ? Selon quels critères ?

## Contenu de l'entrepôt

- ▶ **Faits:** évènements élémentaires
- ▶ **Dimensions:** critères utilisés

# Modèle dimensionnel

## Faits

- ▶ Événements élémentaires intéressants l'entreprise
- ▶ Décrits par des *mesures* organisées selon des *dimensions*

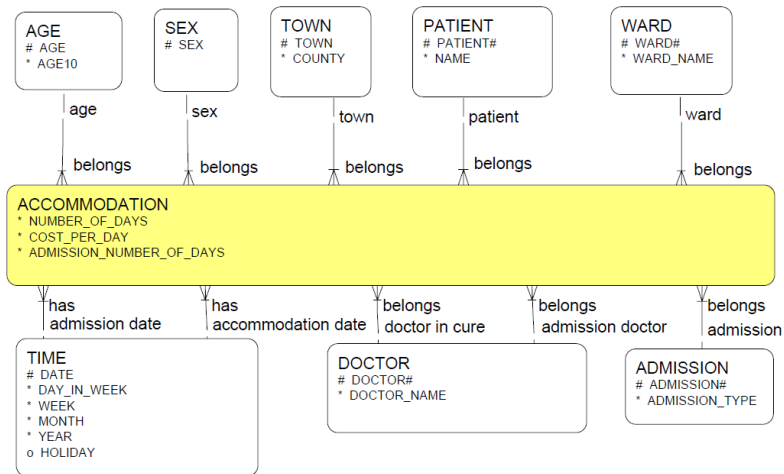
## Dimensions

- ▶ Axes utilisés pour l'analyse

## Mesures

- ▶ Caractéristiques

# Modèle dimensionnel



## Modèle en étoile

## Exemple: contexte

Plusieurs magasins qui vendent des produits

### Produit

- ▶ Nom, marque, couleur, taille, fabricant

### Client

- ▶ Nom, prenom, adresse
- ▶ Achète un produit dans une certaine quantité pour un certain montant

### Magasin

- ▶ Responsable, région

# Exemple: questions

## Questions posées par le service marketing

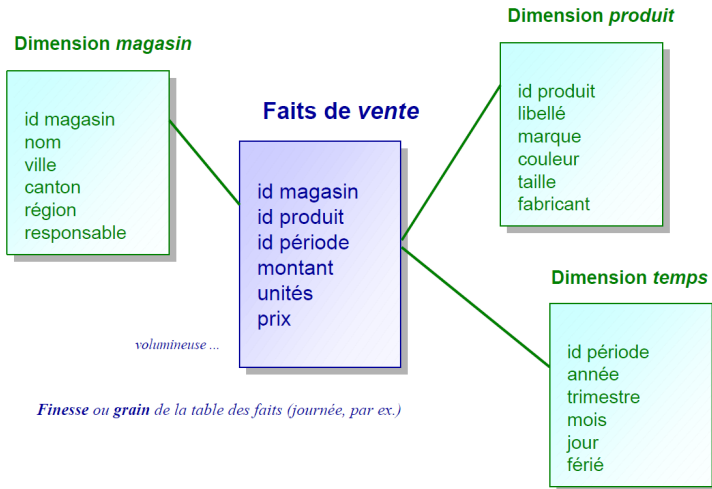
- ▶ Ventes globales dans le temps
- ▶ Ventes par magasin
- ▶ Ventes par produit et par magasin

## Exercice

- ▶ Dessiner le schéma de l'entrepôt de données, en suivant le modèle dimensionnel
- ▶ Avec un exemple



## Exemple: schéma



## Modèle en étoile

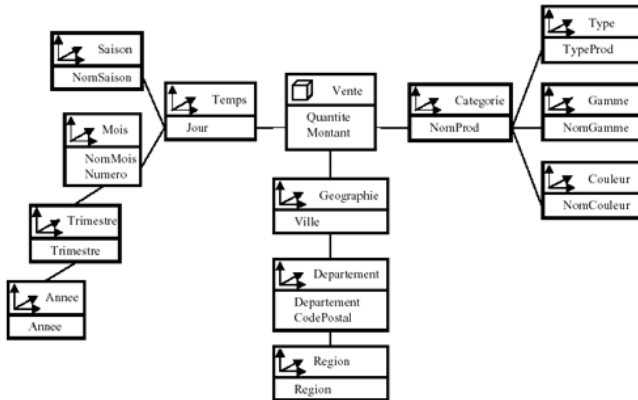
## Exemple: plus de questions

- ▶ Ventes par jours
- ▶ Ventes par mois
- ▶ Ventes par saisons
- ▶ Ventes par couleur
- ▶ Ventes par catégories de produits
- ▶ Ventes par régions

### Exercice

- ▶ Comment adapter le schéma ?

## Exemple: schéma



Modèle en flocon

# Dimensions hiérarchiques

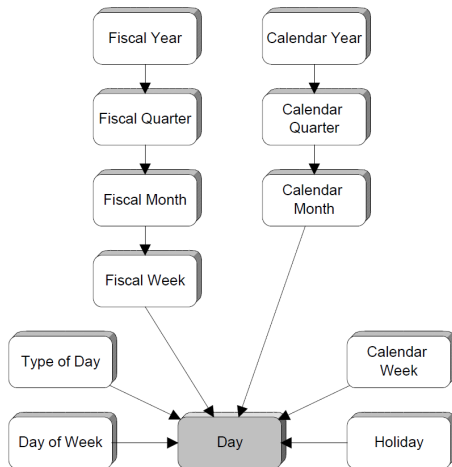
## Fait de vente

- ▶ Date
- ▶ Autres informations

## Une date ?

- ▶ Jour, mois, année
- ▶ Jour de la semaine
- ▶ Week-end, pas week-end
- ▶ Vacances pas vacances
- ▶ Année civile, scolaire, fiscale

# Dimensions hiérarchiques



# Stabilité et flexibilité dans le temps

## Objectif de l'entrepôt

- ▶ Stocker toutes les données sur toute l'histoire de l'entreprise
- ▶ De façon organisée

## Un produit change de nom

- ▶ On veut se souvenir que ça reste le même produit

# Dimensions à évolution lente

## Définition

Changement de la description d'un objet dans la base

## Exemples

- ▶ Nom d'un produit
- ▶ Adresse d'un client
- ▶ Fournisseur d'un produit

# Solutions

## Type 0

- ▶ Ne rien faire: l'application au-dessus s'en charge
- ▶ Dangereux

## Type 1

- ▶ Écraser l'ancienne valeur
- ▶ Perte d'une partie de l'historique

## Type 2

- ▶ Numéros de version ou dates de début et de fin
- ▶ Changements dans le passé délicats



# Solutions

## Type 3

- ▶ Colonnes supplémentaires pour stocker les anciennes valeurs
- ▶ Historique limité

## Type 4

- ▶ Tables supplémentaires pour stocker les anciennes versions
- ▶ Plus de tables

# Exemples

## Donnée de base

Key	Code	Name	State
123	ABC	Acme Supply Co	CA

Changement de l'origine du produit

## Type 1

Key	Code	Name	State
123	ABC	Acme Supply Co	IL

# Exemples

## Type 2

Versions:

Key	Code	Name	State	Version
123	ABC	Acme Supply Co	CA	0
124	ABC	Acme Supply Co	IL	1

Dates:

Key	Code	Name	State	StartDate	EndDate
123	ABC	Acme Supply Co	CA	01-Jan-2000	21-Dec-2004
124	ABC	Acme Supply Co	IL	22-Dec-2004	NULL

# Exemples

## Type 3

Key	Code	Name	OriginalState	CurrentState	EffectiveDate
123	ABC	Acme Supply Co	CA	IL	22-Dec-2004

## Type 4

Supplier:

Key	Code	Name	State
124	ABC	Acme Supply Co	IL

History:

Key	Code	Name	State	CreateDate
123	ABC	Acme Supply Co	CA	14-June-2003
124	ABC	Acme Supply Co	IL	22-Dec-2004

# Analyse en ligne

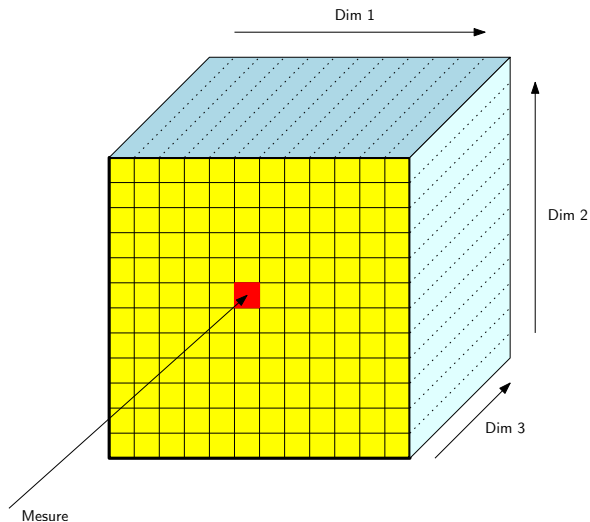
## Tableaux multidimensionnels

- ▶ Hypercube

## Cases

- ▶ Décrites par des dimensions
- ▶ Contenant une ou plusieurs mesures

# Hypercube



# OLAP: Online Analytical Processing

## Online

- ▶ Résultat obtenus immédiatement

## Transversal

- ▶ Plusieurs départements
- ▶ Plusieurs activités

## Prise de décision

- ▶ Opposé du système opérationnel

# OLAP

## Opérations génériques

- ▶ Par des non-experts
- ▶ Peu importe la nature des données

## But

- ▶ Calcul et mise à jour des hypercubes

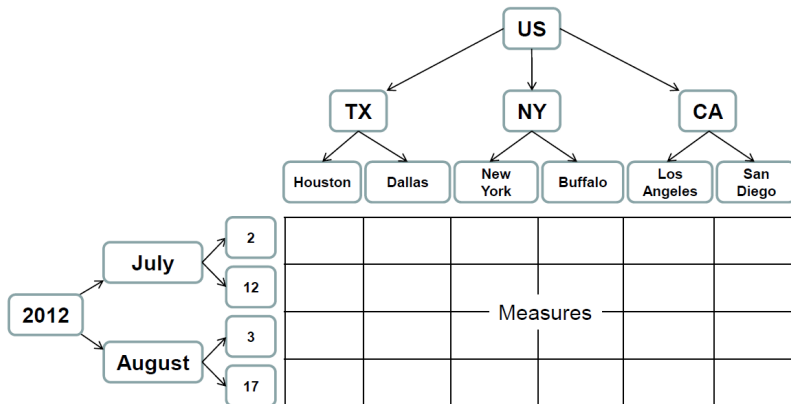
## Requêtes dans l'hypercube

- ▶ Langage de requête



# Hiérarchies

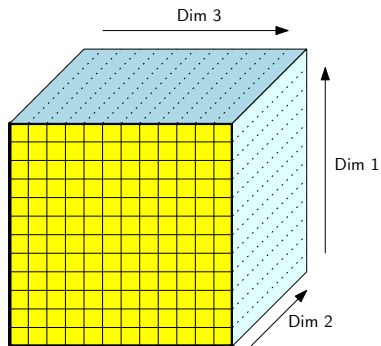
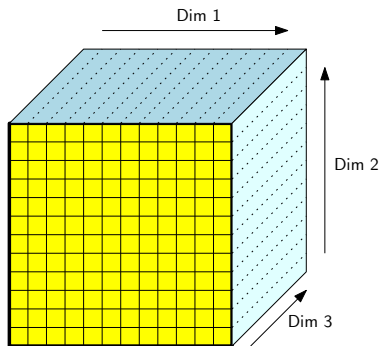
Différents niveaux de détails possibles.



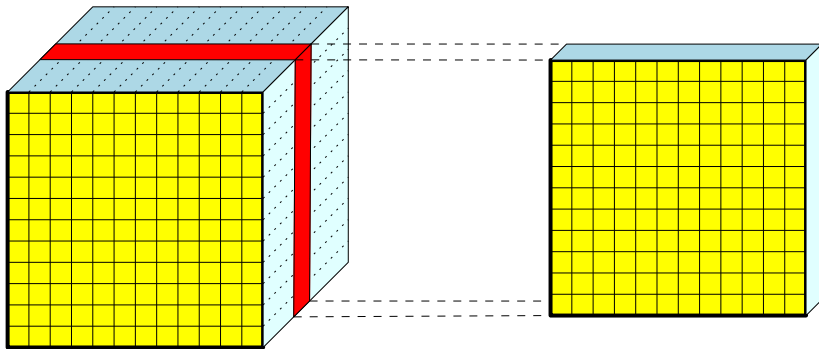
# Manipulation du cube

- ▶ Rotate: sélection des dimensions
- ▶ Slicing: extraction d'une tranche
- ▶ Scoping: extraction d'un bloc de données
- ▶ Drill-up: zoom arrière
- ▶ Drill-down: zoom avant
- ▶ Drill-through: mouvement dans une dimension

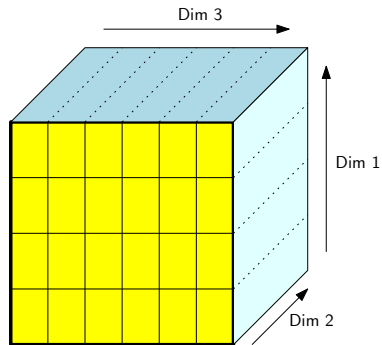
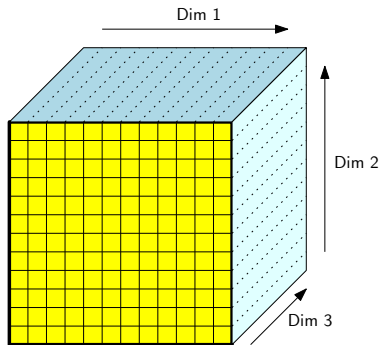
# Rotate



# Slicing

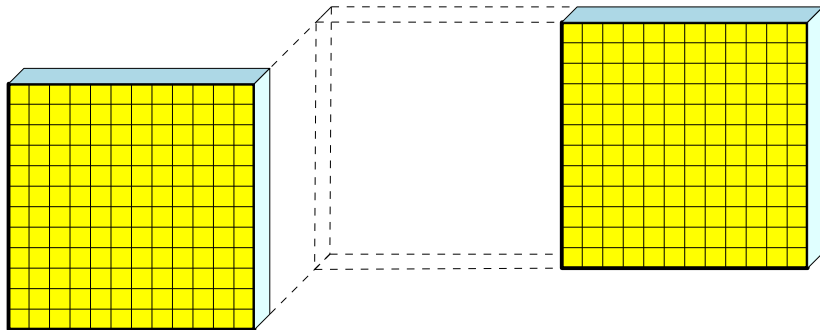


## Drill-up, drill-down



- ▶ Drill-up: plus de détails
- ▶ Drill-down: moins de détails

# Drill-through



## Mouvement selon une dimension

- ▶ Valeurs proches
- ▶ Proche selon quel critère ?
- ▶ Ordre numérique, alphabétique, proximité géographique, data-mining

# Requêtes sur le cube

## MDX: multidimensional expressions

- ▶ Proche du SQL et des tableurs
- ▶ Traduction des demandes vers le stockage concret

## Exemple

```
SELECT  
{ [Measures].[Store Sales] } ON COLUMNS,  
{ [Date].[2002], [Date].[2003] } ON ROWS  
FROM Sales  
WHERE ( [Store].[USA].[CA] )
```