

# Plan du Chapitre

- Problématique des données dans un système décisionnel
- Architecture fonctionnelle d'un système décisionnel
- Les entrepôts de données

# Problématique des données dans un système décisionnel



## CHAPITRE 2: ENTREPÔTS DE DONNÉES – DATA WAREHOUSE

# Systemes Opérationnels vs. Décisionnels

## Systemes opérationnels

- Appelés **OLTP** (**O**n-**L**ine **T**ransaction **P**rocessing) ou systèmes de gestion
- Dédiés aux métiers de l'entreprise pour les assister dans leurs tâches de gestion quotidiennes
- Utilisation des **PGI** (ou **ERP**) pour la gestion des données

## Systemes décisionnels

- Appelés **OLAP** (**O**n-**L**ine **A**nalytical **P**rocessing)
- Dédiés à la gestion de l'entreprise pour l'aider au pilotage de l'activité pour une vision transversale de l'entreprise
- Utilisation des **Entrepôts de données**

# Données Opérationnelles vs. Décisionnelles

## Données Opérationnelles

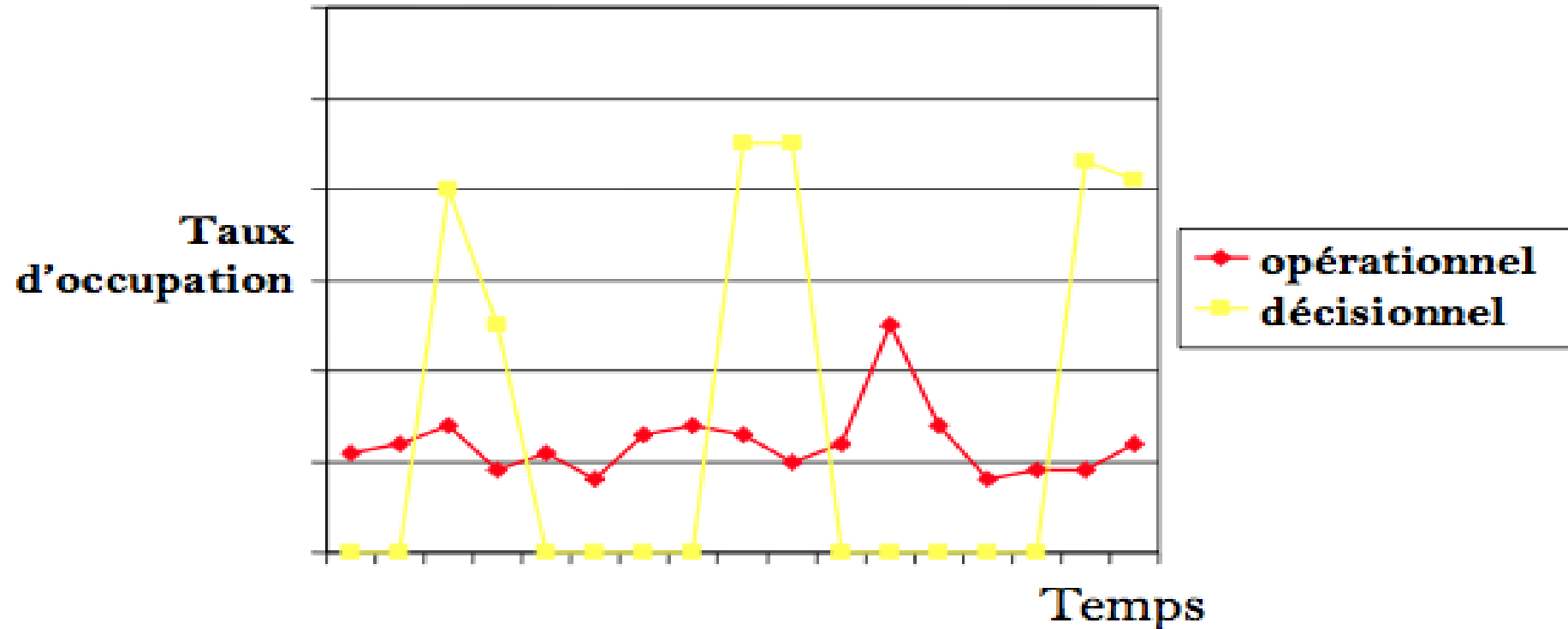
- Données détaillées
- Données récentes
- Structure compréhensible et optimisée pour l'informaticien
- Données réparties et non homogènes
- **Utilisateurs:** Agents opérationnels, nombreux, et concurrents

## Données Décisionnelles

- Données globalisées
- Données historiques
- Structure compréhensible par le décideur
- Données centralisées, intégrées
- **Utilisateurs:** Décideurs, analystes, peu nombreux et non concurrents

# Charge du Serveur de Données

Source: H.E.C. Liège - U.E.R. Systèmes d'Information (F. Fontaine)



# Solutions

- **Stockage de données** : Data Warehouse
  - Base de données unique - vocabulaire unique
  - Contenu adapté aux besoins des décideurs
    - ✓ Structure multidimensionnelle spéciale
    - ✓ Niveau de détail bien étudié
    - ✓ Données historiques
- **Interrogation des données** :
  - Outil interactif, convivial
  - Outil offrant des fonctions d'analyse
    - ✓ Tri des données
    - ✓ Roll-up, Drill-down
    - ✓ Calcul et comparaison
    - ✓ Analyse des tendances, relations et exceptions
    - ✓ Simulation

# OLTP vs. OLAP

## ➤ **OLTP : On Line Transaction Processing**

- Système destiné à offrir le moyen à une application d'utiliser de façon transactionnelle un serveur de base de données. C'est un ensemble logiciel que l'utilisateur peut employer de façon interactive pour accéder aux données de la manière la plus rapide et simple possible.
- *Exemple* : Le 15/01/2012 à 13h12, le client X a retiré 500dt du compte Y

## ➤ **OLAP : On Line Analytical Processing**

- Catégorie de technologie logicielle permettant aux analystes, managers et décideurs d'accéder de manière rapide, consistante et interactive à une large variété d'information, transformée pour refléter la dimension réelle d'une entreprise.
- *Exemple* : Quel est le volume des ventes par produit et par région durant le deuxième trimestre de 2012?

# OLTP vs. OLAP

8

	OLTP	OLAP
<b>Conception</b>	<ul style="list-style-type: none"><li>- Orientée application (Application de production, de facturation...)</li><li>- Structure statique (E/R)</li></ul>	<ul style="list-style-type: none"><li>- Orientée sujet (Client, produit, vendeur...)</li><li>- Structure évolutive (en étoile, en flocon)</li></ul>
<b>Données</b>	<ul style="list-style-type: none"><li>- Détaillées, non agrégées</li><li>- Récentes, mises à jour</li><li>- Accessibles de façon individuelle</li><li>- Normalisées</li></ul>	<ul style="list-style-type: none"><li>- Résumées, recalculées, agrégées</li><li>- Historiques</li><li>- Accessibles de façon ensembliste</li><li>- Dénormalisées</li></ul>
<b>Vue</b>	<ul style="list-style-type: none"><li>- Relationnelle</li></ul>	<ul style="list-style-type: none"><li>- Multidimensionnelle</li></ul>
<b>Requêtes/Utilisation</b>	<ul style="list-style-type: none"><li>- Simples, nombreuses, régulières, prévisibles, répétitives</li><li>- Sensibles aux performances (réponses immédiates)</li><li>- Accès à beaucoup de données</li></ul>	<ul style="list-style-type: none"><li>- Complexes, peu nombreuses, irrégulières, non prévisibles</li><li>- Non sensibles aux performances (réponses moins rapides)</li><li>- Accès à beaucoup d'informations</li></ul>
<b>Utilisateurs</b>	<ul style="list-style-type: none"><li>- Agents opérationnels</li><li>- Nombreux (par milliers)</li><li>- Concurrents</li></ul>	<ul style="list-style-type: none"><li>- Managers / Analystes</li><li>- Peu (par dizaines, centaines)</li><li>- Non concurrents</li></ul>
<b>Accès</b>	<ul style="list-style-type: none"><li>- Lecture / Écriture</li></ul>	<ul style="list-style-type: none"><li>- Lecture</li></ul>
<b>Taille de la base</b>	<ul style="list-style-type: none"><li>- 100 MB à 1 GB</li></ul>	<ul style="list-style-type: none"><li>- 100 GB à 1 TB</li></ul>



# Technologies

**Stockage et Gestion  
Efficace des Gros  
Volumes**



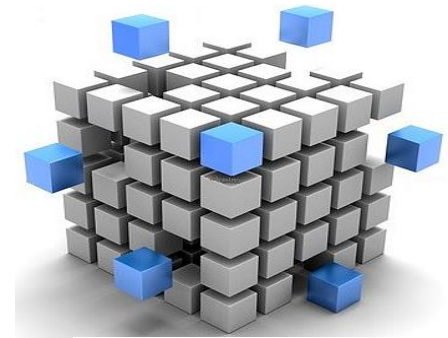
**Entrepôt de Données  
(Data Warehouse)**



**Traitements  
Complexes sur ces  
Volumes**



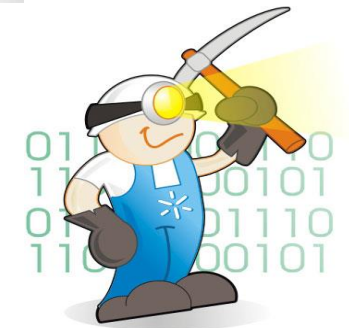
**Serveur OLAP**



**Application Cliente**

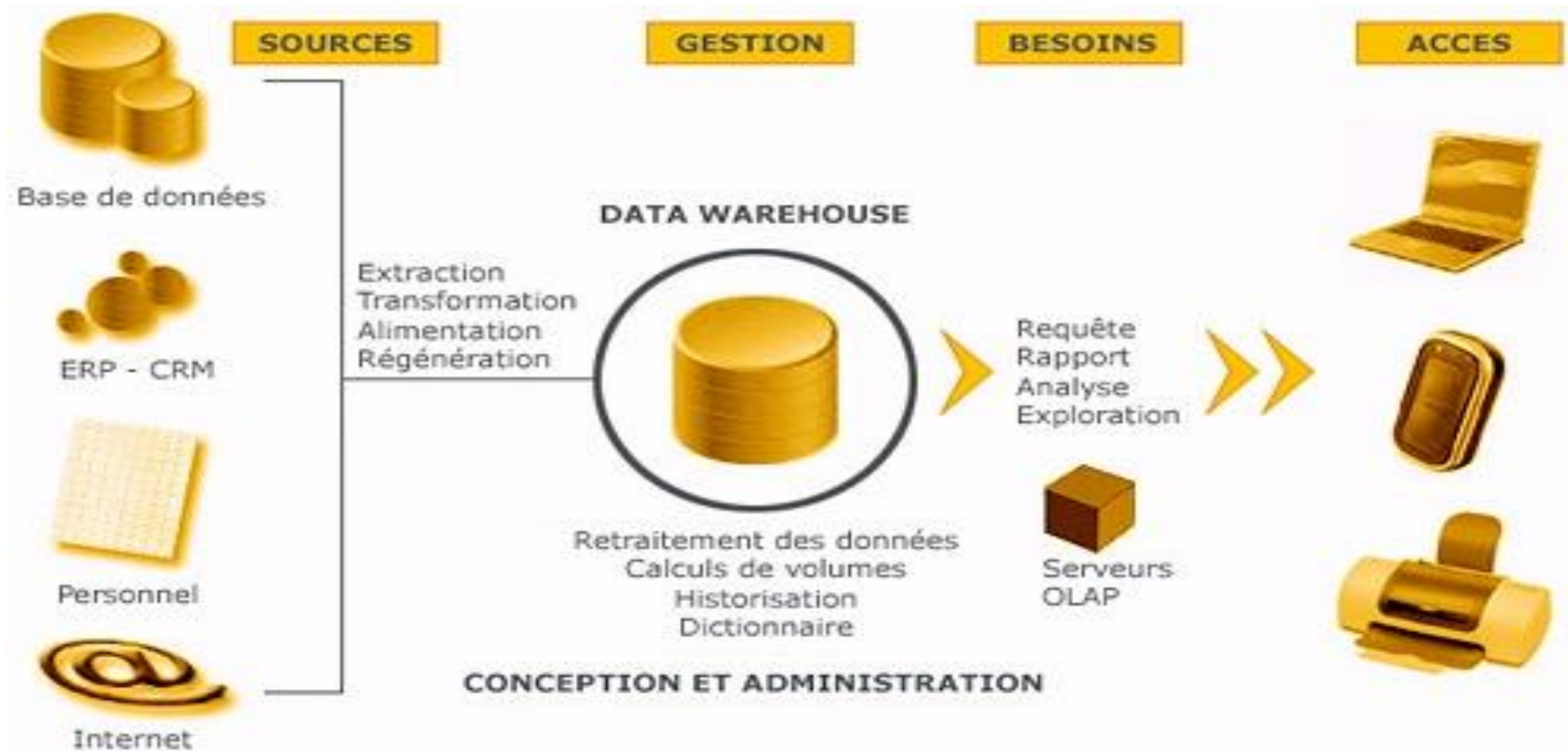


**Fouille de Données  
(Data Mining)**



# Structure d'un Système Décisionnel

10



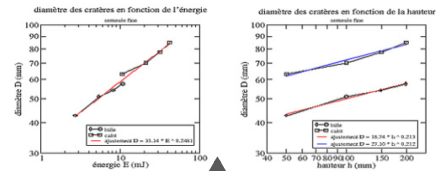
# Structures de Stockage de Données dans un Système Décisionnel



## CHAPITRE 2: ENTREPÔTS DE DONNÉES – DATA WAREHOUSE

# Structures de Stockage de Données dans un Système Décisionnel

Bille 5.5g				Cabo 21.5g				
Diamètre D (mm)				Diamètre D (mm)				
Energie (mJ)	Mes 1	Mes 2	Moyenne	Energie (mJ)	Mes 1	Mes 2	Moyenne	Rapport des masses
2.70	40	45	42.5	10.55	60	66	63	4.33
5.40	50	52	51	21.09	75	85	78	3.55
8.10	55	53	54	31.64	75	80	77.5	4.24
10.79	58	57	57.5	42.18	80	90	85	4.78



**Data Mining**  
(Extraction de données)

**Data Marts**  
(Magasins de données)

**Data Warehouse**  
(Entrepôt de données)

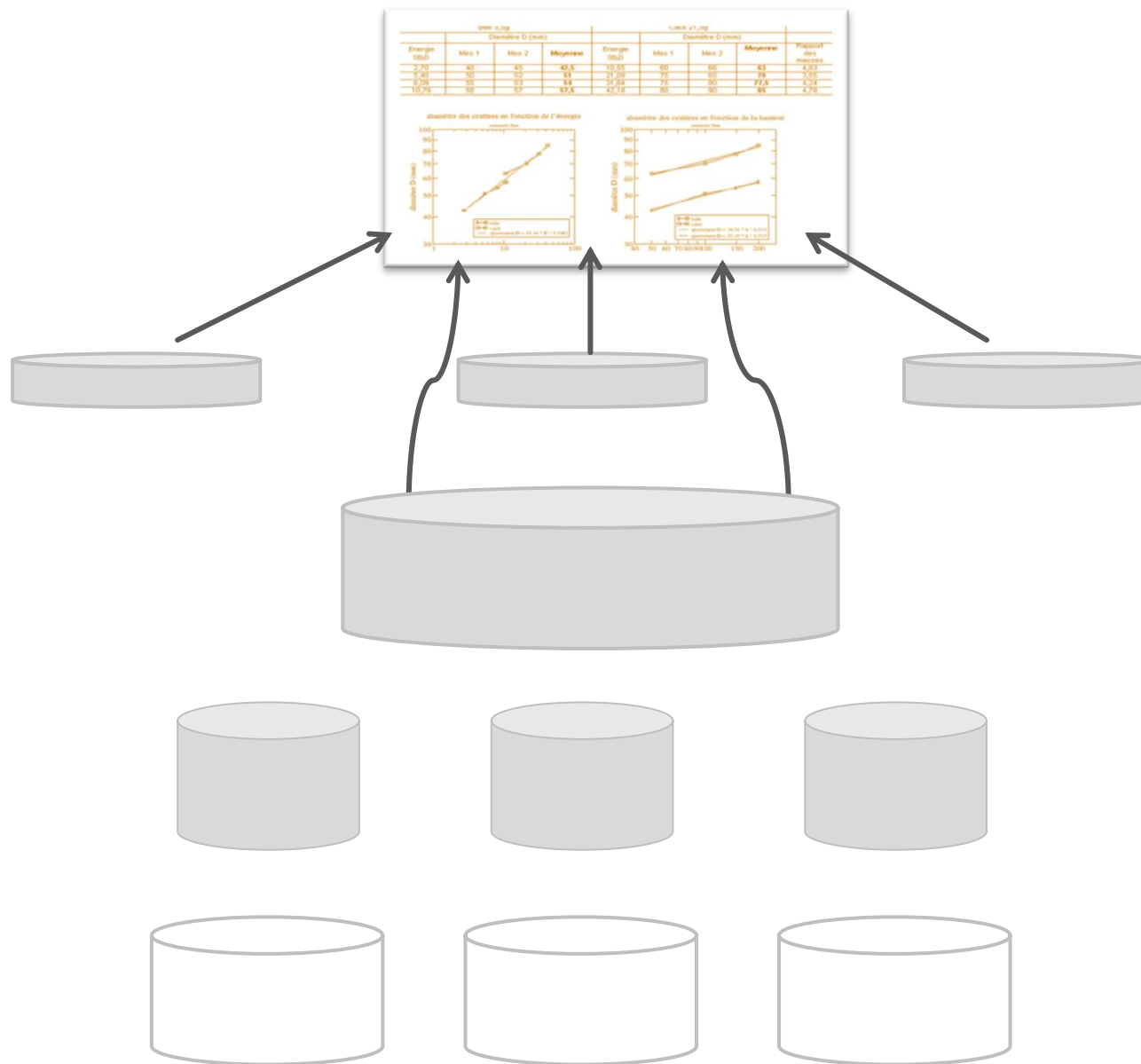
**ODS**  
(Operational Data Store)

**Systèmes  
Opérationnels**

Synthèse

Alimentation  
Transformation

Extraction



# Définitions

## ➤ **ODS: Operational Data Store**

- Collection de données orientées sujet, volatiles, organisées pour le support d'un processus de décision ponctuel, en support à une activité opérationnelle particulière
- Donne la vision immédiate et intégrée de l'état d'un ou plusieurs systèmes opérants

## ➤ **Data Warehouse**

- Entrepôt de données spécifique au monde décisionnel, destiné principalement à analyser les leviers business potentiels
- Collection de données orientées sujet, intégrées, non volatiles et historisées, organisées pour le support d'un processus d'aide à la décision

## ➤ **Data Mart**

- Magasin de données orienté sujet, non volatile, mis à la disposition des utilisateurs dans un contexte décisionnel décentralisé, ciblé pour un usage particulier

## ➤ **Data Mining**

- Ensemble d'outils, méthodes et technologies d'analyse mises en œuvre pour définir des tendances, pour segmenter l'information ou pour établir des corrélations entre les données

# Data Warehouse vs. ODS

- Data Warehouse
  - Intégration des données **hors ligne**
- ODS
  - Intégration des données **en ligne**
  - Sauvegarde des données récentes
  - Utilisé quand les données sont dispersées sur plusieurs supports de stockage, et on a besoin de les rassembler
- *Exemple:*
  - On veut avoir une vue unique sur un patient qu'on pourra modifier en ligne
  - Les données de ce patient sont disposées dans plusieurs bases de données (liste des hospitalisations, liste des diagnostics, liste des achats pharmaceutiques...)
  - ODS peut être utilisé pour extraire ces données et les afficher

# Data Warehouse vs. Data Mart

## ➤ Data Warehouse

- Dépôt de données au niveau entreprise
- Combinaison de plusieurs Data Marts
- Contient toutes les mesures et dimensions nécessaires
- Assure l'intégrité de ces mêmes dimensions à travers tous les Data Marts

## ➤ Data Mart

- Ensemble de dimensions et mesures limitées
- Utilisées pour des thèmes métier spécifiques
- Construites à partir des données des entrepôts

## ➤ Exemple

- Dans une entreprise, il existe un seul entrepôt de données mais plusieurs magasins de données : Finance, Vente...



# Les Entrepôts de Données



## CHAPITRE 2:

### ENTREPÔTS DE DONNÉES – DATA WAREHOUSE

# Pourquoi ne pas utiliser un SGBD?

## ➤ SGBD et DW

- Ont des objectifs différents et font des traitement différents
- Stockent des données différentes
- Font l'objet de requêtes différentes
- Besoin d'une organisation différente des données
- Doivent être physiquement séparés

## ■ SGBD

- ✓ Mode de travail transactionnel (OLTP)
- ✓ Permettent d'insérer, modifier, interroger des informations rapidement, efficacement et en sécurité

### **Objectifs:**

1. Sélectionner, ajouter, mettre à jour et supprimer des tuples
2. Opérations rapides, faites par plusieurs utilisateurs simultanément

# Entrepôt de Données : Objectifs

- Regrouper, organiser des informations provenant de sources diverses
- Les intégrer et les stocker pour donner à l'utilisateur une vue orientée métier
- Retrouver et analyser l'information selon plusieurs critères
- Transformer un système d'information qui avait une vocation de production en un SI décisionnel
- Doit contenir des informations cohérentes
- Les données doivent pouvoir être séparées et combinées au moyen de toutes les mesures possibles de l'activité
- Le DW ne contient pas uniquement des données, mais aussi un ensemble d'outils de requêtes, d'analyse et de présentation de l'information.

# Entrepôt de Données : Définition

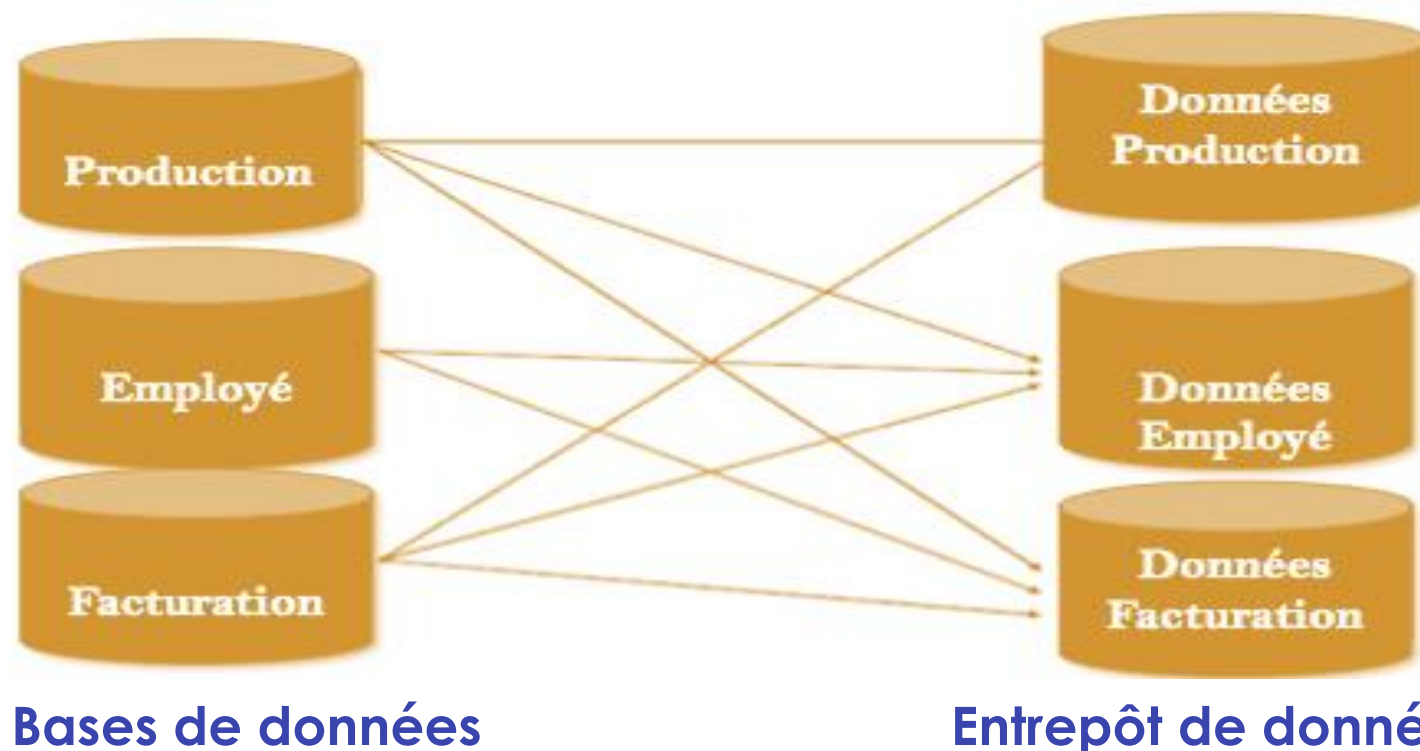
*(Bill Inmon)*

Le Data Warehouse est une collection de données **orientées sujet, intégrées, non volatiles et historisées**, organisées pour la prise de décision.

# Caractéristiques d'un DW

## Données Orientées Sujet

- Le DW est orienté **sujets**, ce qui signifie que les données collectées doivent être orientées *métier*, et donc triées par thème



# Caractéristiques d'un DW

## Données Intégrées

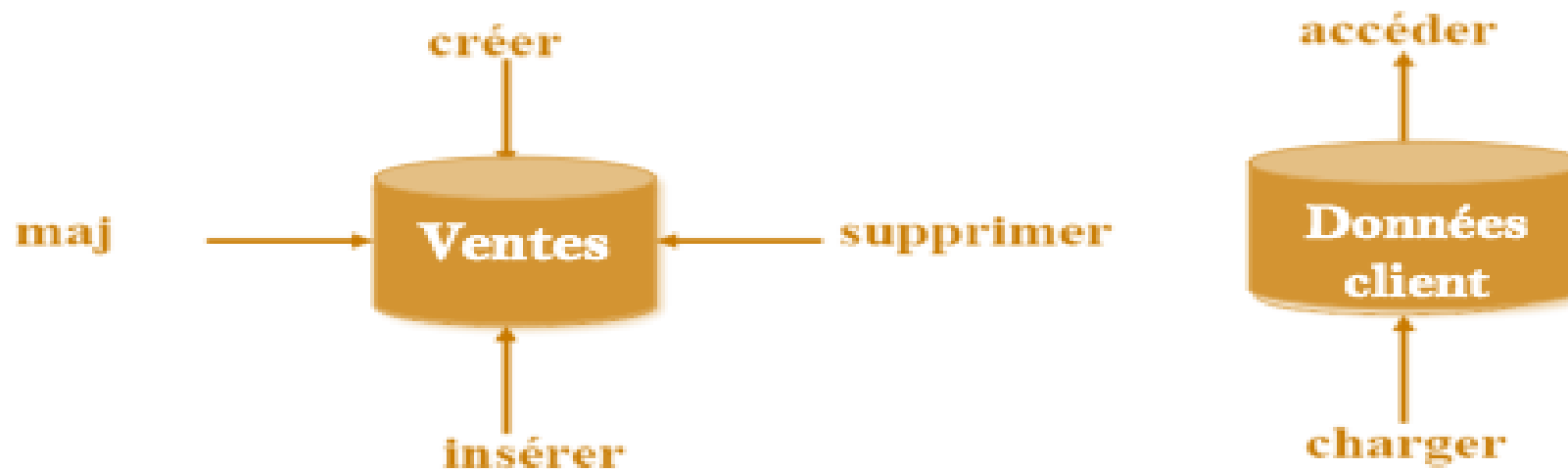
- Le DW est composé de **données intégrées**, c'est à dire qu'un « nettoyage » préalable des données est nécessaire dans un souci de rationalisation et de normalisation



# Caractéristiques d'un DW

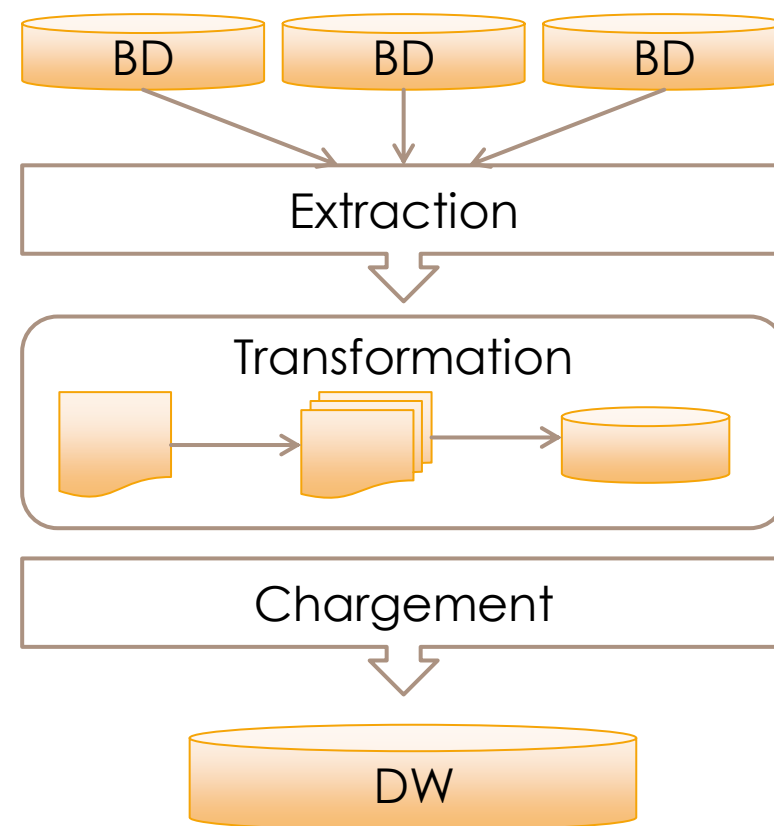
## Données Historisées et Non Volatiles

- Les données du DW sont **non volatiles**, ce qui signifie qu'une donnée entrée dans l'entrepôt l'est pour de bon et n'a pas vocation à être supprimée
- Les données du DW doivent être **historisées**, donc datées



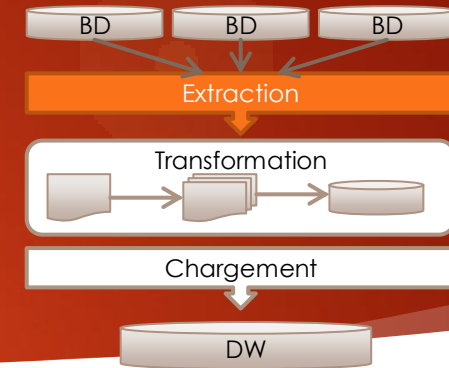
# Data Warehouse : ETL

- Outils d'alimentation pour
- Extraire
  - Transformer
  - Charger dans un DW
- } les données sources





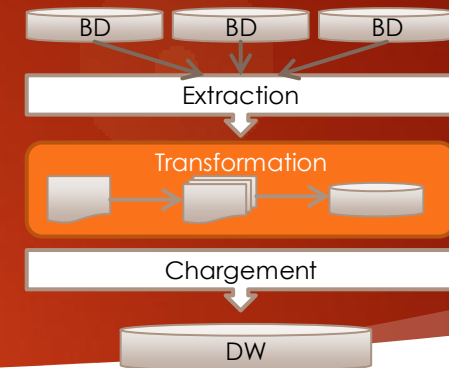
# ETL : Extraction



- Extraction des données de leur environnement d'origine (base de données relationnelles, fichiers plats...)
- Besoin d'outils spécifiques pour accéder aux bases de production (requêtes sur des BD hétérogènes)
- Besoin d'une technique appropriée pour n'extraire que les données nécessaires
  - Données créées ou modifiées depuis la dernière opération d'extraction
- Attention:
  - L'extraction ne doit pas perturber l'activité de production

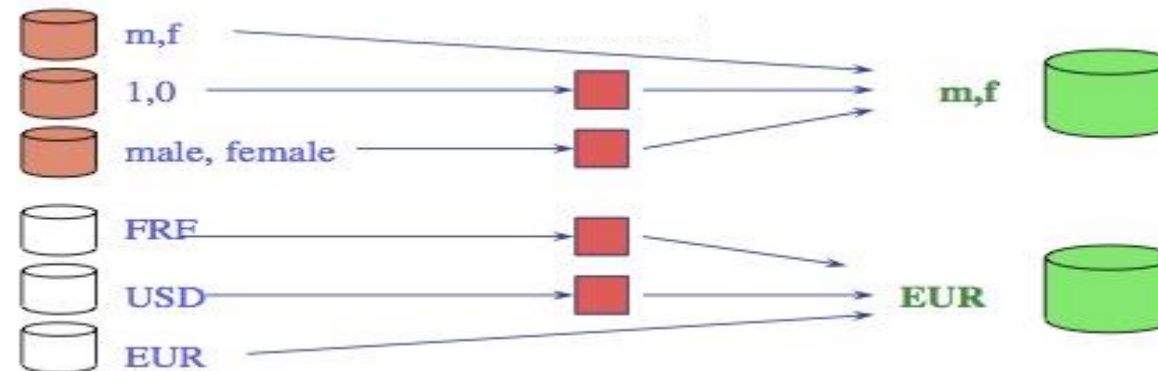
# ETL : Transformation

26

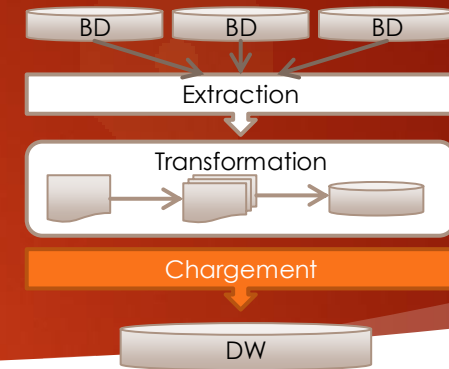


## ➤ Intégration des données

- Homogénéisation du vocabulaire, structures, valeurs
- Suppression et fusion des redondances
- Épuration des données (suppression des données incohérentes)
- Transformation des données dans un format cible



# ETL : Chargement



27

- **Le chargement** des données résultantes dans les différentes applications décisionnelles : Data Warehouse ou Enterprise Data Warehouse, Data Marts, applications OLAP (Online Analytical Processing) ou “cubes”, etc

# ETL (résumé)

- Les processus ETL (Extraction, Transformation et Chargement) sont les composants les plus critiques - et les plus importants – pour l'alimentation d'un data warehouse ou d'un système décisionnel
  - **L'extraction** de données des applications et des bases de données de production (ERP, CRM, SGBDR, fichiers, etc.)
  - **La transformation** de ces données pour les réconcilier entre les différentes sources, pour effectuer des calculs ou du découpage de texte, pour les enrichir avec des données externes et aussi pour respecter le format requis par les système cibles (Troisième Forme Normale, Schéma en Etoile, Dimensions à Evolution Lente, etc.)
  - **Le chargement** des données résultantes dans les différentes applications décisionnelles : Data Warehouse ou Enterprise Data Warehouse, Data Marts, applications OLAP (Online Analytical Processing) ou "cubes", etc.

# ETL

## ➤ Propriétaire

- Anatella
- [DataStudio \(Data\)](#)
- [Feature Manipulation Engine](#) (FME)
- [Hurence](#) avec un ETL natif [Hadoop](#)
- [IBM InfoSphere DataStage3](#)
- [Informatica PowerCenter4](#)
- MapReport
- [Microsoft SQL Server](#) Integration Services (SSIS)
- [OpenText Genio](#)
- [Oracle](#) Data Integrator (Sunopsis)
- [Oxio Data Intelligence solution ETL](#)
- SAP Data Services
- SAS Data Integration Studio
- [Stambia](#)
- [STATISTICA ETL \(StatSoft\)](#)
- SynchroDB

## ➤ Open-source

- [Talend6](#)
- [Logstash](#) (souvent utilisé avec Elasticsearch)
- [Pentaho Data Integration](#)
- [CloverETL](#)

# Bibliographie

## Supports de Cours

- Karima Tekaya – « Informatique Décisionnelle » - *INSAT*
- Fatma Baklouti – « Les entrepôts de données (Data Warehouses) » - *INSAT*
- Hweichao Lu – « Online Analytical Processing (OLAP) » - *CS157B-02 Spring 2007*
- Didier Donsez – « Systèmes d'information décisionnels (Data Warehouse / Data Mining) » - *Université Joseph Fourier*
- Jacky Akoka et Isabelle Comyn-Wattiau – « Le Data Warehouse et les Systèmes Multidimensionnels » - *Université de Versailles*

## Livres

- Ralph Kimball - « Concevoir et déployer un Data Warehouse » - *Editions Eyrolles, 2000*