

# CHAPITRE 1: ENTREPÔTS DE DONNÉES

- Dr BERGHIDA Meryem, 2019/2020, Université de Biskra

## 1. Introduction : Notion de décision

- Tous les jours, les individus, les groupes et les entreprises sont confrontés à prendre des décisions.
  - Au niveau individuel : quelle filière suivre après le tronc commun
  - Au niveau organisationnel : quel produit lancer ou arrêter ?
- Difficulté de prise de décisions
  - Certaines décisions sont faciles : si moyenne  $> 10$  → accepté, sinon refusé
  - D'autres sont difficiles : définir la moyenne de rachat

## 1. Introduction : Notion de décision

- Définition de la décision :
- Selon Larousse, décider c'est **déterminer ce qu'on doit faire**.  
Décision = **Action de décider**
- Selon Mintzberg : une décision peut être définie comme  
« **l'engagement dans une action, ou l'intention explicite d'agir** ».

## 2. Contexte de prise de décision

- Qui prend la décision ? **Responsables, Chef d'entreprise...**
- Pourquoi ? **Gain, Avantage concurrentiel, compétitivité**
- Sur la base de quoi ? **Données détaillées et agrégées, passées ou prédictives**
- Quand ? **Pas trop tôt / surtout pas trop tard**
- Comment ? **En appliquant des modèles, en se basant sur des outils**
- A l'AIDE de qui ? **Informaticiens (nous)**

### 3. Outils d'aide à la décision

- Programmation linéaire (gestion de stock, ...)
- Théorie des jeux
- SIAD (système interactifs d'aide à la décision)
- Systèmes multi-agents
- Entrepôts de données

## 4. Problèmes liés aux données pour la prise de décision

- Données transactionnelles (I, U, D) → non historisées
- Modèles de données inadapté à l'analyse (normalisation, non redondance) → requêtes d'analyse difficiles d'écriture et temps de réponses prohibitifs
- Multiples sources de données (BdD, fichiers, ...) → incohérences, risques d'erreurs

## 4. Problèmes liés aux données pour la prise de décision

- Données distribuées : accès non uniforme aux données, problème de disponibilité de source externes
- Données détaillées : les décideurs ont besoin de détails, mais plus important que cela ; de résumés...

## 5. Solutions

- Historiser les données : pas de suppression
- Modéliser les données à des fins d'analyse : priorité données aux performances pas à la normalisation
- Intégrer les données en une seule sources : éliminer redondance, incohérence, etc...

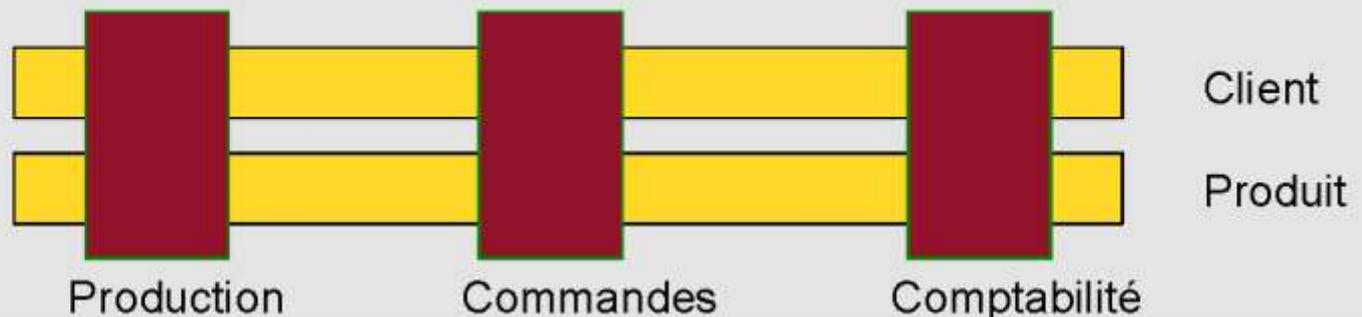
Techniquement parlant : entreposer les données

## 6. Définition d'un entrepôt de données (data warehouse)

- Selon Inmon (1996): «Un entrepôt de données est une collection de données orientées sujet, intégrées, non volatiles et historisées, organisées pour le support d'un processus d'aide à la décision ».
- A data warehouse is a « subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management's decision- making »

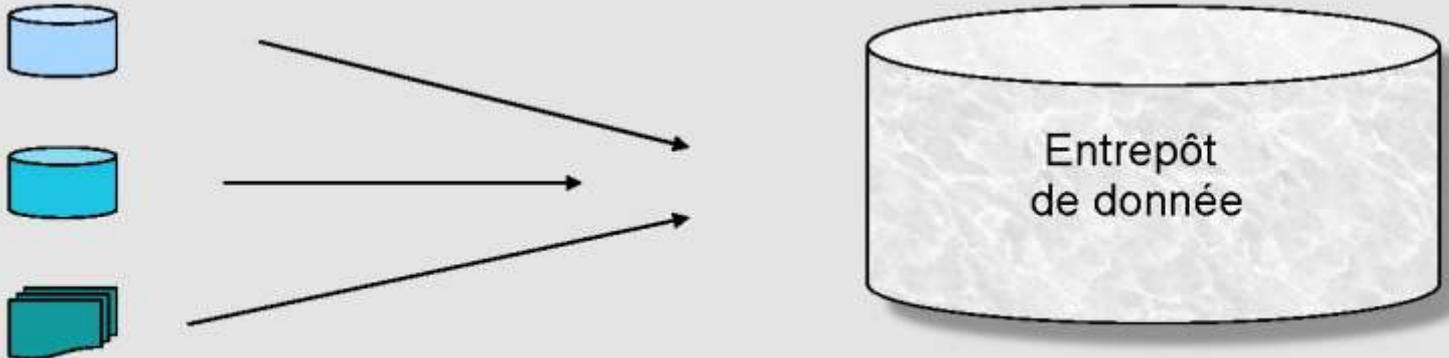
## 6. Définition d'un entrepôt de données (data warehouse)

- Orientée sujets :
  - Données organisées autour des grands sujets (thématiques): client, produit, vente
  - Modèles de données orientés sujet et non orientés transaction
  - Vision globale et synthétique au lieu de vision détaillées des données



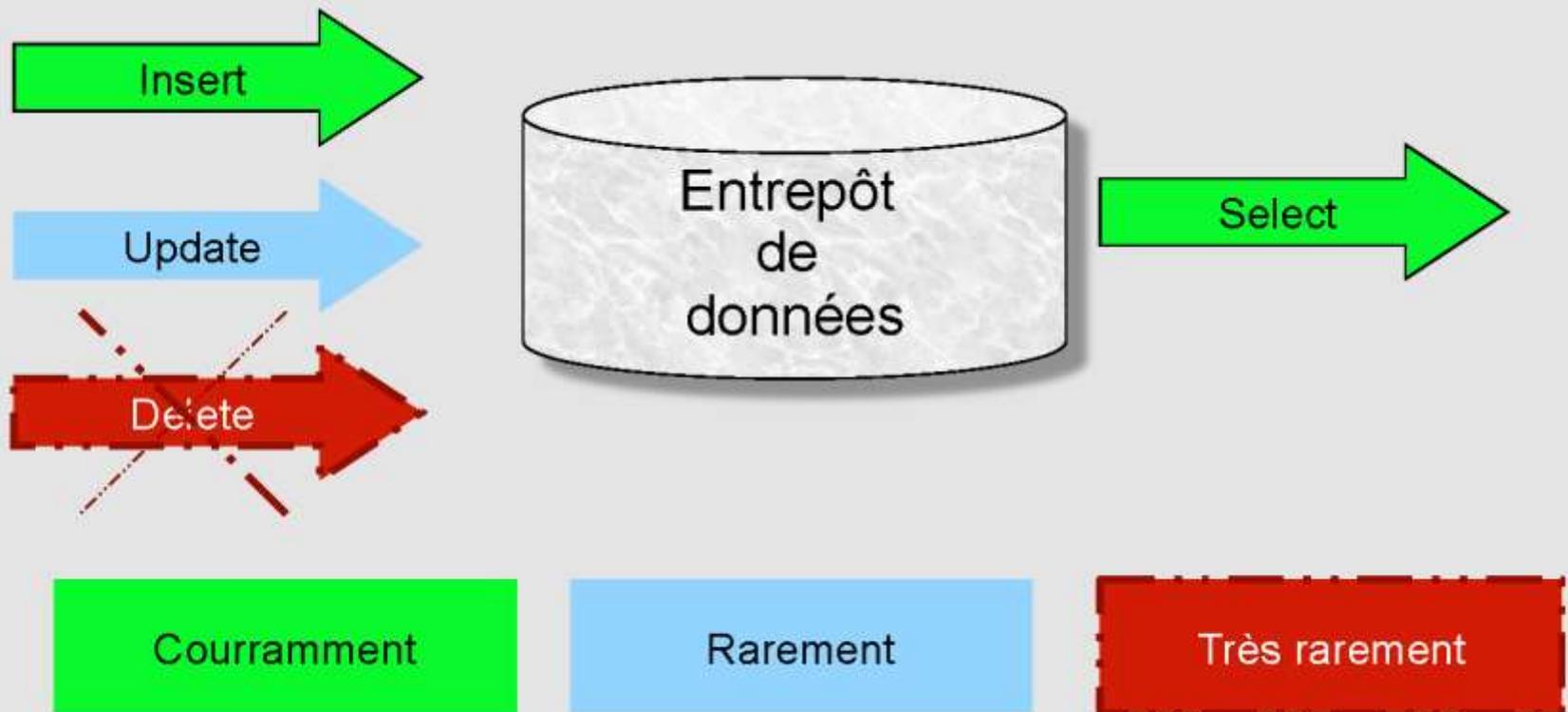
## 6. Définition d'un entrepôt de données (data warehouse)

- Intégrées :
  - Les données sont physiquement copiées après nettoyage, consolidation...
  - Impossible dans certains cas : raisons légales, données à changement rapide



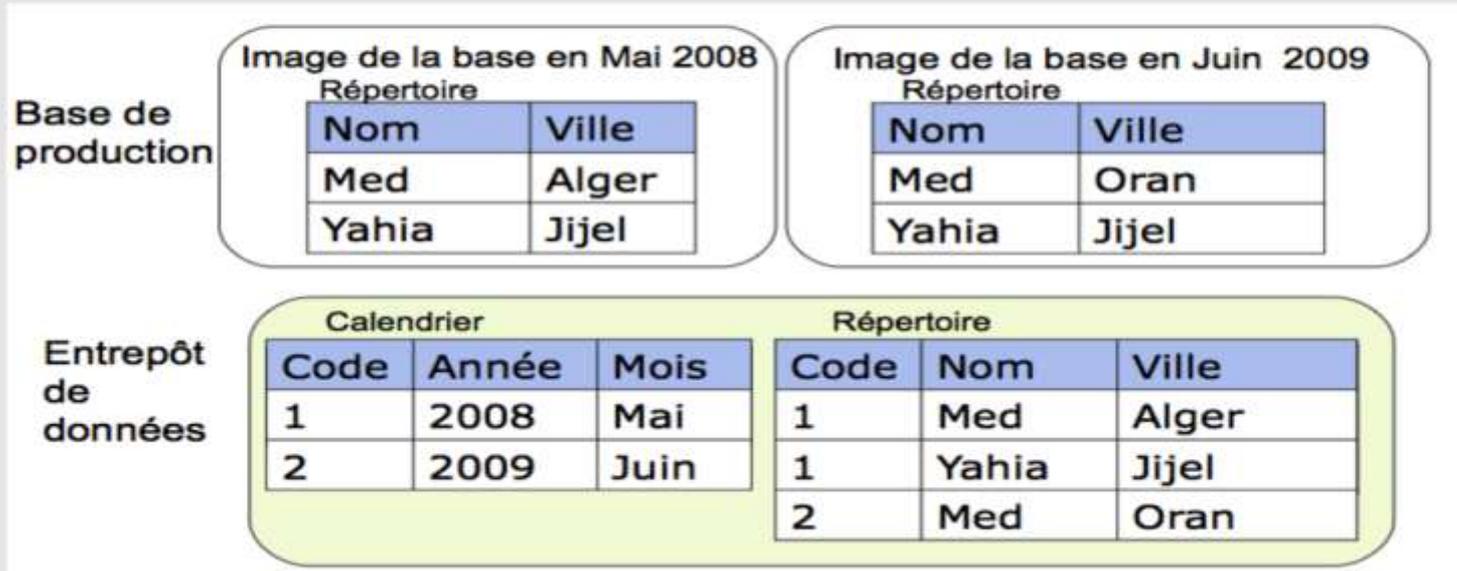
## 6. Définition d'un entrepôt de données (data warehouse)

- Non volatiles :

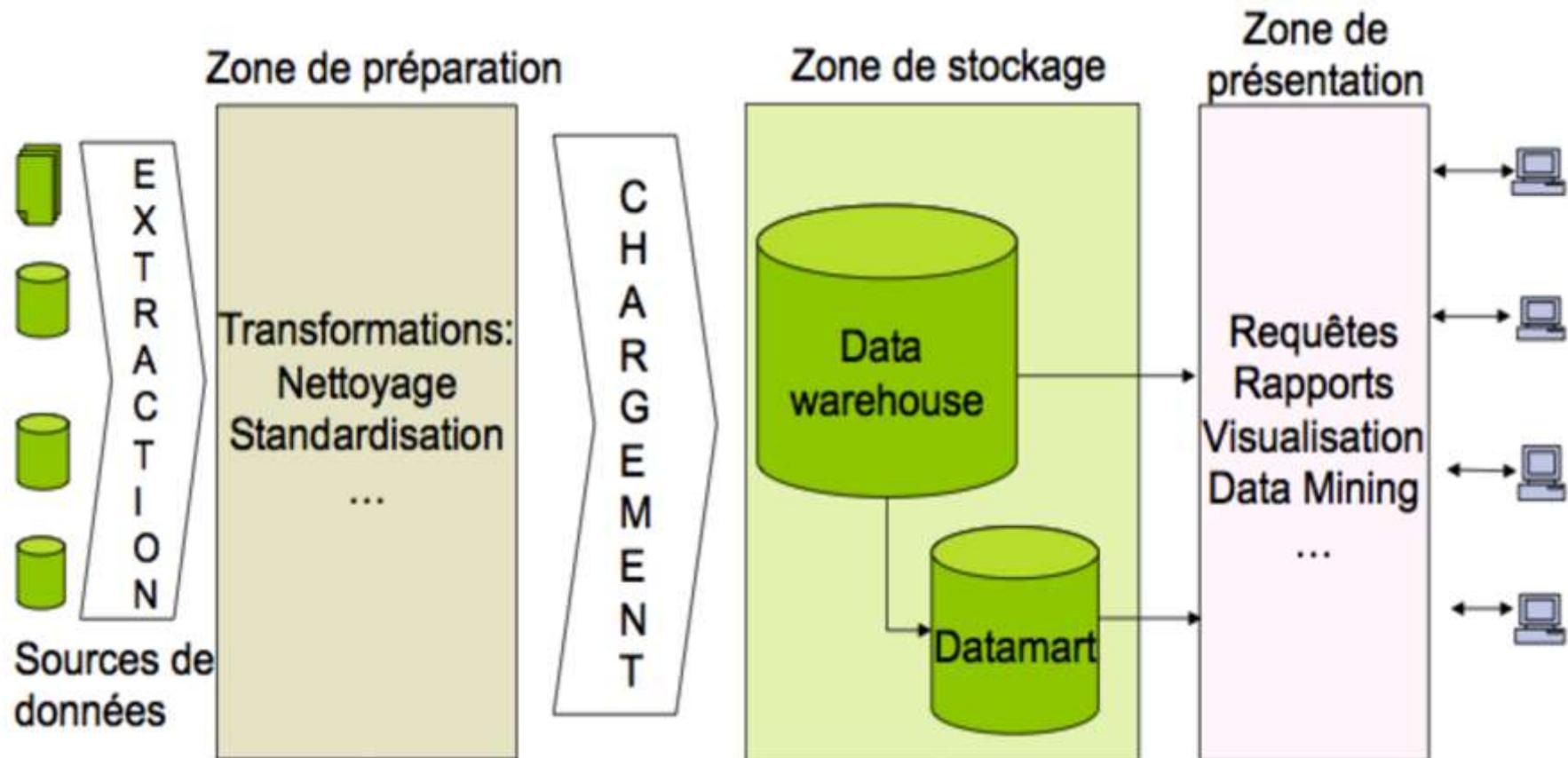


## 6. Définition d'un entrepôt de données (data warehouse)

- Historisées:
  - Le temps est un aspect primordial
  - Les changements sur les données (et schéma) sont sauvegardés -> grande volumétrie



## 6. Architecture de référence d'un entrepôt de données.



## 6. Architecture de référence d'un entrepôt de données.

- Sources de données :



- Bases de données opérationnelles



- Fichier plats (XML, texte, ...)

- Redondance, incohérence

Exemples

| Genre |
|-------|
| Homme |
| Femme |

Personne

| Sexe |
|------|
| 1    |
| 2    |

Employés

| Nature |
|--------|
| H      |
| F      |

Pers

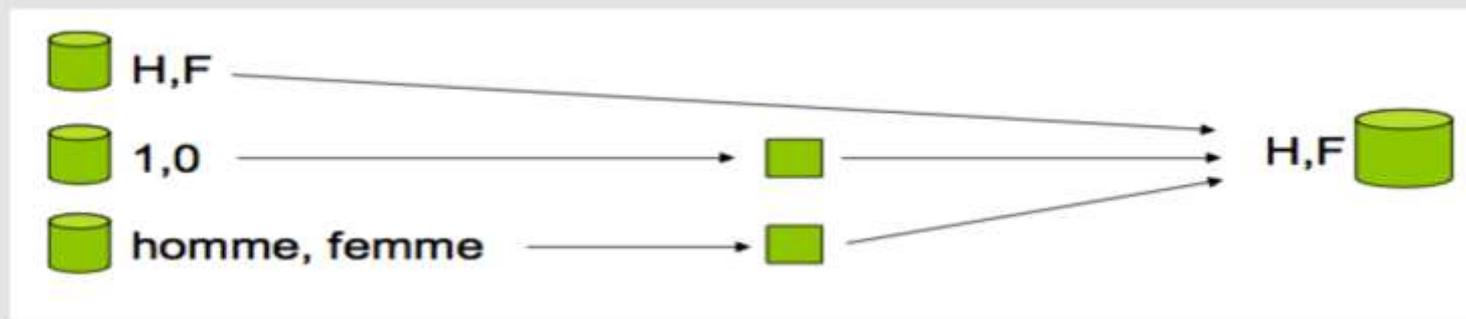
- Sources internes / sources externes

## 6. Architecture de référence d'un entrepôt de données.

- Zone de préparation:

- *Extraction* : lecture des données des sources
- *Transformation* : Élimination des redondance, unification de longueurs, de codifications, etc....
- *Chargement* : insertion à partir des bases de données opérationnelles vers l'entrepôt

- Exemple :



## 6. Architecture de référence d'un entrepôt de données.

- Zone de stockage

- Entrepôt de données
  - Transversal
  - De téraoctets aux pétaoctets
- Data marts (magasins de données)
  - Concerne un département
  - De gigaoctets aux téraoctet
- Cubes
  - Extrait de l'entrepôt ou du datamart
  - De mégaoctets aux gigaoctets

## 6. Architecture de référence d'un entrepôt de données.

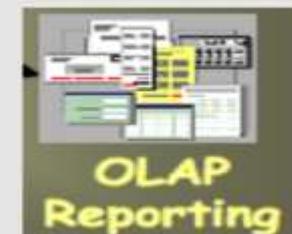
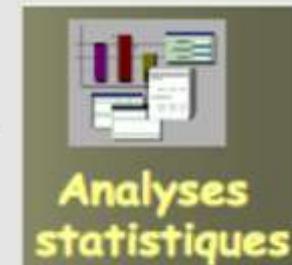
- Zone de présentations

- Outils d'analyse

- Outil de requêtage

- Fouille de données

- OLAP / reporting



## 7. Caractéristiques d'un entrepôt de données

- N'est pas un produit mais une **solution** : ne s'achète pas mais se construit.
- Offre une vue **agrégée** et permet de descendre aux données détaillées
- Pose le problème de **performance** à cause du **grand volume** de données

## 7. Différences entre un ED et une base de données opérationnelle (OLAP Vs OLTP)

| Caractéristique     | Bases de données opérationnelles                    | Entrepôts de données                         |
|---------------------|---|--|
| Buts                | Gestion de données opérationnelles                  | Aide à la décision                           |
| Nbr d'utilisateurs  | Milliers  | Centaines                                    |
| Nature des requêtes | Requêtes prédéfinies                                | Requêtes Ad Hoc                              |
| Accès aux données   | Centaines d'enregistrements , en lecture / écriture | Millions d'enregistrements, en lecture seule |

## 7. Différences entre un ED et une base de données opérationnelle (OLAP Vs OLTP)

| Caractéristique       | Bases de données opérationnelles          | Entrepôts de données            |
|-----------------------|---|---------------------------------|
| Données               | Détaillées, numériques et alphanumériques | Plutôt numériques               |
| Intégration           | Orientée application                      | Orientée sujet                  |
| Couverture temporelle | Données courantes seulement               | Donnée courantes et historiques |

## 7. Différences entre un ED et une base de données opérationnelle (OLAP Vs OLTP)

| Caractéristique   | Bases de données opérationnelles            | Entrepôts de données              |
|-------------------|---|-----------------------------------|
| Mise à jour       | Continue                                    | Périodique                        |
| Modèle de données | Normalisé                                   | Dénormalisé,<br>Multidimensionnel |
| Optimisation      | Accès transactionnel à une partie de la BdD | Accès analytique à toute la BdD   |