

## Cours Entrepôts des données

Enseignante: Salma DAMMAK

2<sup>ème</sup> année Licence GSI

Année Universitaire: 2020/2021

## Partie 1: Introduction des Entrepôts de données



### Les systèmes d'information (Rappel)

- Le véhicule de la communication dans l'entreprise



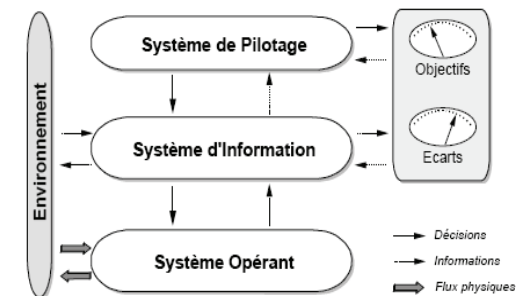
- Restitution d'une information à la bonne personne et au bon moment sous le format approprié.

- Principales fonctions d'un SI: Saisie, Mémorisation, Traitement, Communication

Codrs entrepôt de données

3

### Les systèmes d'information (Rappel)



Modèle systémique des organisations [Gouarné 1997]

➡ Système de couplage entre le système opérant et le système de pilotage

## Les systèmes d'information (Rappel)

- Une entreprise dispose de données provenant soit de son système opérationnel soit de l'extérieur
  - Surabondance de données
  - Les données représentent une **mine d'informations**
- L'entreprise a besoin d'informations pertinentes pour faire face à
  - Concurrence
  - Evolution du marché, fluctuation de la consommation



Cours entrepôt de données

5

## Les systèmes d'information (Rappel)

### ■ But

- Améliorer les performances décisionnelles de l'entreprise en répondant aux demandes d'analyse des décideurs non informaticiens et non statisticiens.

### ■ Exemples

- Analyse Clientèle
- Analyse Chiffre d'affaire
- Marketing

### ■ Solution

- Passer d'un système d'information qui avait une vocation de production à **un SI décisionnel** dont la vocation de **pilotage** devient majeure.

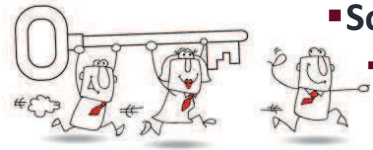


Cours entrepôt de données

6

## Les systèmes d'information (Rappel)

### ■ Solution



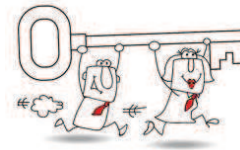
- Passer d'un système d'information qui avait une vocation de production à **un SI décisionnel** dont la vocation de **pilotage** devient majeure.
- L'entreprise ne doit pas seulement avoir une vue verticale de ses métiers (Syst. de gestion) mais une vue transversale (Syst. décisionnel).

Cours entrepôt de données

7

## Les systèmes d'information (Rappel)

### ■ Solution



- la BD orientée Transaction de la BD orientée Aide à la Décision
- Naissance du concept Entrepôt de Données = Data Warehouse

Cours entrepôt de données

8

## Les entrepôts de données

« Un entrepôt de données est une collection de données **orientées sujet, intégrées, non volatiles, historisées**, organisées pour le support d'un processus d'aide à la décision »

(Bill Inmon, 1996)

Cours entrepôt de données

9

## Les entrepôts de données

### ■ Orientés sujet:

Les données sont organisées par sujet (ex: clients, produits, ventes, etc.).

### ■ Intégrés:

Les données, qui proviennent de diverses sources hétérogènes, sont consolidées et intégrées dans l'entrepôt.

### ■ Historiques:

Les données ont très souvent une composante temporelle (ex: date et heure d'une transaction).

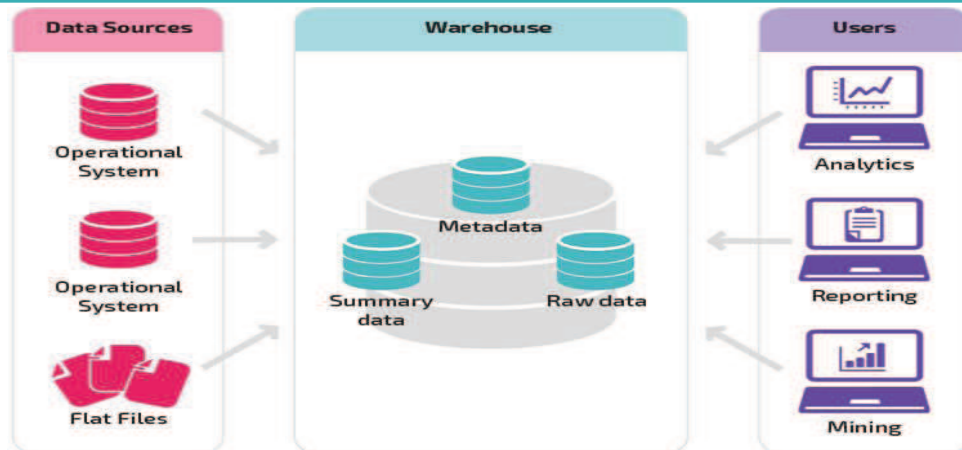
### ■ Non-volatiles:

Une fois insérées dans l'entrepôt, les données ne sont jamais modifiées ou effacées; elle sont conservées pour des analyses futures.

Cours entrepôt de données

10

## Les entrepôts de données



Cours entrepôt de données

11

## Les entrepôts de données

■ Intégration des données à partir de sources multiples et Hétérogènes

■ Répondre aux requêtes du système d'aide à la décision.

**On-Line Analytical Processing OLAP**

■ Plus efficace en coût d'exécution des requêtes

Cours entrepôt de données

12

## Les entrepôts de données

- **OLTP:** consultent et mettent à jour les données des BD opérationnelles.
  - Exécution en temps réel des transactions, pour l'enregistrement des opérations quotidiennes : inventaires, commandes, paye, comptabilité
- **OLAP:** lisent les données seulement pour avoir de nouvelles informations à partir des données sources.
  - Rapports, Etats, Tableaux de Bord, Graphiques, Synthèses, Groupement,...
  - Traitement efficace des requêtes d'analyse pour la prise de décision.

Cours entrepôt de données

13

## Les entrepôts de données

- **Désavantage de partager la même BD entre OLAP et OLTP**
  - OLAP ne peut utiliser que les données actuellement stockées dans les BD
  - Les analyses historiques sont souvent impossibles à cause des opérations de mises à jour qui changent les données historiques
  - Des opérations de verrouillage des données (Locking operations)
  - Des problèmes de performance
    - les requêtes analytiques demandent l'accès à de très grands nombre de tuples.

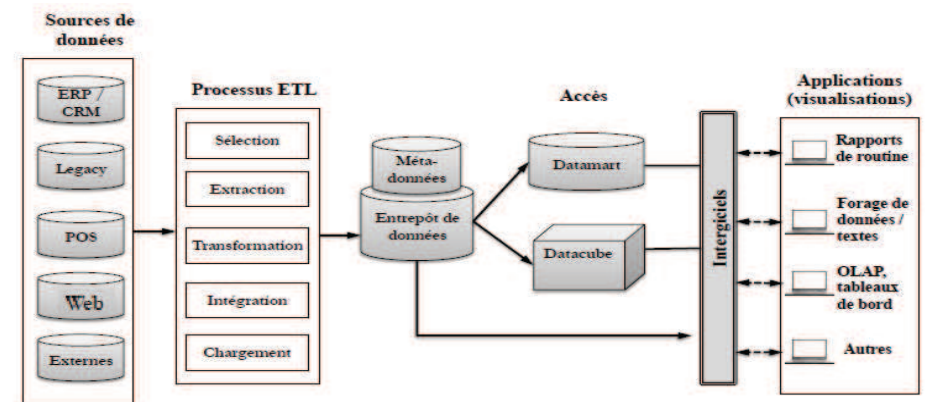
Cours entrepôt de données

14

## Les entrepôts de données

Critères	OLTP	OLAP
Espace requis	Relativement petit si les données historiques sont archivées	Grand due aux données historiques et aux données d'agrégation
Données	exhaustives, courantes, dynamiques , non volumineuses	Résumées, historiques, statiques, volumineuses
Modélisation de la BD	Un grand nombre de tables hautement normalisées	Moins de tables, pas de normalisation (schémas en étoile ou flocon)
Requêtes	Simple Mise à jour, Interrogation	Complexe Interrogation
Sauvegarde et récupération	Fait de façon régulière	Fait de façon irrégulière
Ressources requises	Simple DBMS relationnel	DBMS spécialisé multi-processeurs et à grande capacité
Vue	Relationnelle	Multidimensionnelle
Nombre d'utilisateurs simultanés	Grand	Petit

## Architecture des entrepôts de données



Cours entrepôt de données

16

## Architecture des entrepôts de données

### ■ Les sources de données

- Enterprise Resource Planning (ERP):
  - Gèrent les processus opérationnels d'une entreprise (ex: ressources humaines, finances, distribution, approvisionnement, etc.).
- Customer Relationship Management (CRM):
  - Gèrent les interactions d'une entreprise avec ses clients (ex: marketing, ventes, après-vente, assistance technique, etc.).
- Systèmes legacy:
  - Matériels et logiciels obsolètes mais difficilement remplaçables.

17

## Architecture des entrepôts de données

### ■ Les sources de données:

- Point of sale (POS):
  - Matériels et logiciels utilisés dans les caisses de sorties d'un magasin.
- WEB:
  - Clickstreams (fichier log), structure site web.
- Externes:
  - Ex: données concurrentielles achetées, données démographiques

Cours entrepôt de données

18

## Architecture des entrepôts de données

### ■ Les sources de données:

- Sources diverses et disparates (ex: BD, fichier texte, etc.);
- Sources sur différentes plateformes et OS;
- Qualité de données douteuse et changeante dans le temps;
- Incohérence entre les différentes sources;
- Données dans un format difficilement interprétable ou ambigu.



Cours entrepôt de données

19

## Architecture des entrepôts de données

### ■ Les processus ETL: *Extract-Transform-Load*

- ETL résume généralement les données afin de réduire leur taille et d'améliorer leur performance pour des types d'analyse spécifiques.
- Partie la plus longue du développement (jusqu'à 70% du temps total).



Cours entrepôt de données

20

## Architecture des entrepôts de données

### ▪ Les processus ETL: *Extract-Transform-Load*

- **Extrait** l'information nécessaire à l'analyse
  - Identifier les données sources utiles
  - Déterminer tous les changements survenus durant une période donnée (ex: heure, jour, semaine, mois).

## Architecture des entrepôts de données

### ▪ Les processus ETL: *Extract-Transform-Load*

- **Transforme** : Consolider les données
  - Données redondantes, manquantes, incohérentes, etc.
  - Découpage, fusion, conversion, aggrégation, –etc..

## Architecture des entrepôts de données

### ▪ Les processus ETL: *Extract-Transform-Load*

- **Charge** (Load) les données dans l'ED
  - Faire les chargements en lot dans une période creuse (entrepôt de données non utilisé);
  - Avoir un plan pour évaluer la qualité des données chargées dans l'entrepôt;
  - Commencer par charger les données des tables de dimension;

## Architecture des entrepôts de données

### ▪ Les Entrepôts de données: *Data Warehouse*

- Lieu de stockage centralisé et extrait des sources.
- intègre et «historise» l'ensemble des données utiles pour les prises de décisions.
- Son organisation doit faciliter la gestion des données et la conservation des évolutions.

## Architecture des entrepôts de données

### ■ DSA: *Data Staging Area*

- Zone de préparation des données)
- Base de données intermédiaire qui stocke les données issues des systèmes de production opérationnelle dans un format proche de ces derniers.
- Un stockage tampon avant l'intégration des données dans l'ED.

## Architecture des entrepôts de données

### ■ Les Magasins de données: *Data Marts*

- Contiennent une portion du contenu de l'entrepôt de données;
- Se concentre sur un seul sujet d'analyse (ex: les ventes OU l'inventaire, mais pas les deux);
- Servent à faire des analyses simples et spécialisées (ex: fluctuations des ventes par catégorie de produits);

## Architecture des entrepôts de données

### ■ Entrepôts de données et Magasins de données

- L'approche de **Bill Inmon: Approche descendante (top-down)**
  - Conception **intégrale** de l'ED.
  - ED le dépôt centralisé de toutes les données
  - Les Data Marts des différents secteurs sont ensuite créés sur la base du modèle de l'entrepôt.

☺ Vision conceptuelle complète de l'ED.

☺ Normalisation des données, absence de redondance.

☹ Difficulté de mise en œuvre .

☹ Manque d'évolutivité.



## Architecture des entrepôts de données

### ■ Entrepôts de données et Magasins de données

- L'approche de **Ralph Kimball : Approche Ascendantes (bottom-up)**
  - Conception **incrémentale** de l'ED.
  - Le Data Warehouse est simplement une combinaison (union) de différents Data Marts qui facilite le reporting et l'analyse.

☺ Simplicité de mise en œuvre

☺ Rapide

☹ Problème d'intégration des magasins de données



## Architecture des entrepôts de données

### ▪L'interrogation

- Consiste à utiliser les données des magasins pour prendre des décisions.
- La représentation des données doit faciliter leur compréhension et leur manipulation par les décideurs non informaticiens.

## Architecture des entrepôts de données

### ▪Big Data Warehousing

- Une architecture de gestion et d'organisation des données
- Utilise à la fois des architectures d'entrepôt de données traditionnelles et des technologies Big Data modernes
- Fournir une analyse rapide sur un large éventail de types d'informations.

## Architecture des entrepôts de données

### ▪Cloud Data Warehousing

- Entrepôt à la demande
- En mode Cloud (ou DWaaS pour Data Warehousing as a Service)
- Un modèle d'externalisation dans lequel un prestataire de services configure et gère les ressources matérielles et logicielles requises par à un ED.