

# Le Data Warehouse et les Systèmes Multidimensionnels

1

## 1. Définition d'un Datawarehouse (DW)

- Le **Datawarehouse** est une collection de données orientées sujet, intégrées, non volatiles et historisées, organisées pour le support d'un processus d'aide à la décision (Inmon, 94).

2

# 1. Définition d'un Datawarehouse

## 1.1 Données orientées sujet

- Données structurées par thèmes (sujets majeurs de l'entreprise) et non suivant les processus fonctionnels.
- Le sujet est transversal aux structures fonctionnelles et organisationnelles de l'entreprise. On peut accéder aux données utiles sur un sujet.
- L'intégration des différents sujets se fait dans une structure unique.

# 1. Définition d'un Datawarehouse

## 1.1 Données orientées sujet

- Il n'y a pas de duplication des informations communes à plusieurs sujets.
- La base de données est construite selon les thèmes qui touchent aux métiers de l'entreprise (clients, produits, risques, rentabilité, ...).
- Les données de base sont toutefois issues des Systèmes d'Information Opérationnels (SIO).

# 1. Définition d'un Datawarehouse

## 1. 2. Données intégrées

- Les données, issues de différentes applications de production, peuvent exister sous toutes formes différentes.
- Il faut les intégrer afin de les homogénéiser et de leur donner un sens unique, compréhensible par tous les utilisateurs.
- Elle doivent posséder un codage et une description unique.

# 1. Définition d'un Datawarehouse

## 1. 2 Données intégrées

- La phase d'intégration est longue et pose souvent des problèmes de qualification sémantique des données à intégrer (synonymie, homonymie, etc...).
- Ce problème est amplifié lorsque des données externes sont à intégrer avec les données du SIO.

# 1. Définition d'un Datawarehouse

## 1.3 Données non-volatiles

- Une information est considérée volatile quand les données sont régulièrement mises à jour comme dans les Systèmes d'Information Opérationnels.
- Dans un SIO, les requêtes portent sur les données actuelles. Il est difficile de retrouver un ancien résultat.
- Dans un DW, il est nécessaire de conserver l'historique de la donnée. Ainsi, une même requête effectuée à deux mois d'intervalle en spécifiant la date de référence de la donnée, donnera le même résultat.

Copyright J. Akoka - I. Comyn-Wattiau

7

# 1. Définition d'un Datawarehouse

## 1.4 Données historisées

- Dans un SIO, les transactions se font en temps réel, et les données sont mises à jour constamment. L'historique des valeurs de ces données sont conservées car elles sont inutiles.
- Dans un DW, la donnée n'est jamais mise à jour.
- Les données du DW s'ajoutent aux données déjà engrangées.

Copyright J. Akoka - I. Comyn-Wattiau

8

# 1. Définition d'un Datawarehouse

## 1. 4 Données historisées

- Le DW stocke donc l'historique des valeurs que la donnée aura prises au cours du temps.
- Un référentiel de temps est alors associé à la donnée afin d'être capable d'identifier une valeur particulière dans le temps.
- Les utilisateurs possèdent un accès aux données courantes ainsi qu'à des données historisées.

# 1. Définition d'un Datawarehouse

## 1. 5 Support d'un processus d'aide à la décision

- Un DW est un système d'information dédié aux applications décisionnelles dont les principales contraintes sont :
  - des requêtes complexes à plusieurs niveaux d'agrégation
  - la nécessité de disposer d'informations synthétiques (« reporting » de gestion, analyse des ventes, gestion de la masse salariale, etc)
  - le stockage des données sous une forme multi-dimensionnelle
  - des mises à jour périodiques

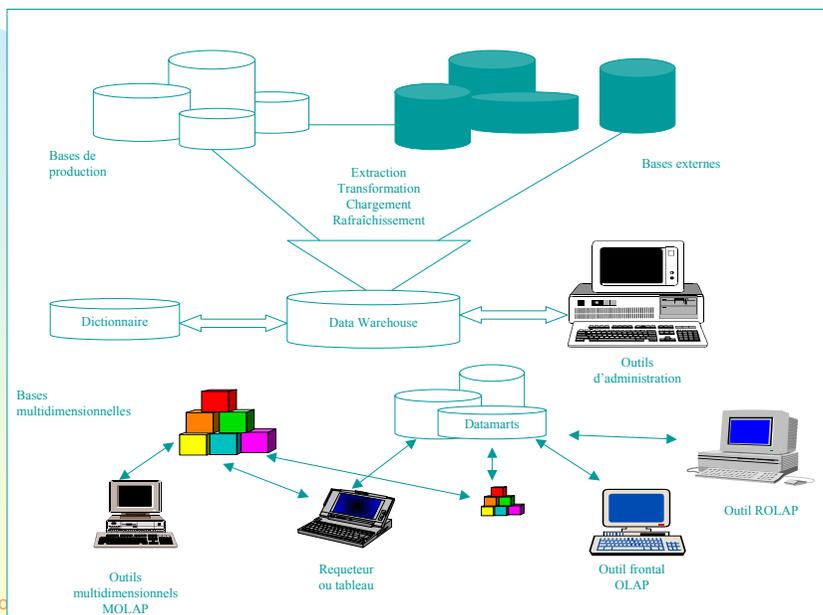
## 2. Objectifs d'un Datawarehouse

- permet le développement d'applications décisionnelles et de pilotage de l'entreprise et de ses processus
- joue un rôle de référentiel pour l'entreprise puisqu'il permet de fédérer des données souvent éparpillées dans différentes bases de données
- offre une vision globale et orientée métiers de toutes les données que manipule l'entreprise
- permet de faire face aux changements du marché et de l'entreprise
- offre une information compréhensible, utile et rapide

Copyright J. Akoka - I. Comyn-Wattiau

11

## 3. Architecture d'un Datawarehouse



Co

12

## 3. Architecture d'un Datawarehouse

### 3. 1 Les Bases de Données

- Bases de production de l'entreprise
- Bases créées par les utilisateurs
- Bases de données externes à l'entreprise (Nielsen, INSEE, ...) qui nécessitent leur identification, leur rapatriement et leur intégration.

## 3. Architecture d'un Datawarehouse

### 3. 2 Opérations sur les données

#### EXTRACTION

- Extraire les données de leur environnement d'origine (bases de données relationnelles, fichiers plats, ...).
- Utiliser une technique appropriée pour n'extraire que les données nécessaires : données créées ou modifiées depuis la dernière opération d'extraction.

## 3. Architecture d'un Datawarehouse

### 3.2 Opérations sur les données

#### TRANSFORMATION

- Une même donnée peut avoir une structure ou une valeur différente en fonction de la base (production, externe, utilisateurs) dont elle provient.
- On peut être confronté à des redondances (un même client peut apparaître avec différents attributs et propriétés selon la source consultée).

## 3. Architecture d'un Datawarehouse

### 3.2 Opérations sur les données

#### TRANSFORMATION

- Il faut supprimer certaines données aberrantes qui risqueraient de fausser les analyses.
- Il faut donc épurer et transformer les données.

## 3. Architecture d'un Datawarehouse

### 3.2 Opérations sur les données

#### CHARGEMENT/RAFRAICHISSEMENT

- Effectuer sur les données des opérations de calcul et d'agrégation.
- Remplacer certaines bases si aucune solution d'extraction satisfaisante n'est possible.

## 3. Architecture d'un Datawarehouse

### 3.2 Opérations sur les données

#### CHARGEMENT/RAFRAICHISSEMENT

- Mettre en place des procédures de chargement (nocturnes?) et de restauration (en cas de problème).
- Si la disponibilité du système ne peut être interrompue, envisager la mise en place de systèmes redondants.

## 3. Architecture d'un Datawarehouse

### 3.2 Opérations sur les données

#### LES OUTILS

- On peut automatiser tout ou partie des opérations décrites.
- Des outils sont disponibles : Extract d'ETI, Genio de Leonard 's Logic, ...
- Le développement d'outils spécifiques est envisageable mais risque d'alourdir les tâches.

## 3. Architecture d'un Datawarehouse

### 3.3 Dictionnaire de Données

- Le dictionnaire de données regroupe les méta-données.
- Une méta-donnée représente une donnée sur les données. Il s'agit de l'ensemble des informations qui permettent de qualifier une donnée, notamment par sa provenance, sa qualité, etc...
- les méta-données permettent de préciser de quelle table provient la donnée, à quelles dates et heures elle en a été extraite, l'état de la base à cet instant, etc...

## 3. Architecture d'un Datawarehouse

### 3.3 Dictionnaire de Données

- Une méta-donnée permet de « remonter la chaîne » et de reconstituer l'ensemble d'événements et données qui ont servi à obtenir l'information associée.
- Le dictionnaire de données contient toutes les informations permettant d'exploiter les données.
- C'est un référentiel destiné aux utilisateurs et à l'administrateur du DW.
- A ce jour, il n'existe pas de normes en ce qui concerne la structure et la gestion des dictionnaires de données. Chaque outil propose sa solution et son approche.

Copyright J. Akoka - I. Comyn-Wattiau

21

## 3. Architecture d'un Datawarehouse

### 3.4 Les Data-Marts

- Un data-mart est un DW focalisé sur un sujet particulier, souvent au niveau départemental ou métier.
- C'est donc un mini DW lié à un métier particulier de l'entreprise (finance, commercial, ...).
- Un DW est souvent volumineux (plusieurs centaines de Go voire quelques To) avec des performances inappropriées (temps de réponse trop longs). Un Data-mart, quant à lui, comporte moins de 50 Go, ce qui permet des performances acceptables.
- La création d'un data-mart peut être un moyen de débiter un projet de DW (projet pilote).

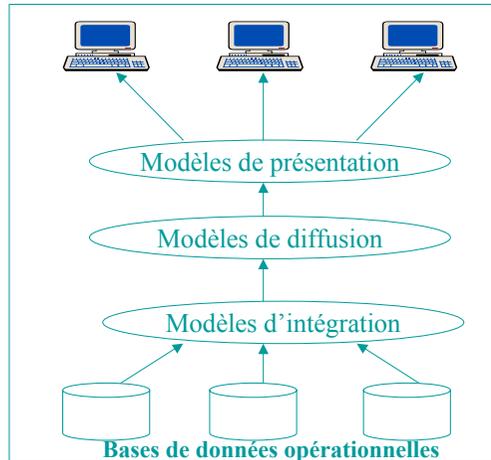
Copyright J. Akoka - I. Comyn-Wattiau

22

## 3. Architecture d'un Datawarehouse

### 3. 5 Les bases multidimensionnelles et les outils OLAP

#### 3.5.1 Les modèles de données



Copyright J. Akoka - I. Comyn-Wattiau

23

## 3. Architecture d'un Datawarehouse

### 3. 5 Les bases multidimensionnelles et les outils OLAP

#### 3.5.1 Les modèles de données

- Le modèle d'intégration unifie les données opérationnelles.
- Le modèle de diffusion représente le modèle conceptuel des données. Il correspond aux bases multidimensionnelles (serveur OLAP).
- Le modèle de présentation est un complément au modèle conceptuel. C'est à travers ce modèle que l'utilisateur voit les données. Il correspond à différents outils physiques : les tableurs, les requêteurs, les outils clients OLAP, etc...

Copyright J. Akoka - I. Comyn-Wattiau

24

## 3. Architecture d'un Datawarehouse

### 3.5 Les bases multidimensionnelles et les outils OLAP

#### 3.5.2 Les outils OLAP (On-Line Analytical Processing)

- OLAP caractérise l'architecture nécessaire à la mise en place d'un système d'information décisionnel.
- OLAP s'oppose à OLTP (On-Line Transactional Processing) qui caractérise les SIO.

## 3. Architecture d'un Datawarehouse

### 3.5 Les bases multidimensionnelles et les outils OLAP

#### 3.5.2 Les outils OLAP (On-Line Analytical Processing)

- OLAP constitue l'ensemble des outils multidimensionnels nécessaires à l'accès, le stockage et la manipulation des données utiles pour un SID ou pour un EIS.
- OLAP désigne les outils d'analyse s'appuyant sur les bases de données multidimensionnelles.

## 3. Architecture d'un Datawarehouse

### 3. 5 Les bases multidimensionnelles et les outils OLAP

#### 3.5.3 Les 12 règles de E.F. CODD (1993)

**Vue multidimensionnelle** : Les données sont structurées en dimensions métiers.

**Transparence** : L'utilisateur doit pouvoir utiliser les logiciels habituels (tableurs, ...) sans percevoir la présence d'un outil OLAP.

**Accessibilité** : L'outil doit se charger d'accéder aux données stockées dans n'importe quel type de bases de données (interne + externe) et le faire simultanément.

**Performance continue dans les restitutions** : A mesure que le nombre de dimensions ou la taille de la base augmente, l'utilisateur ne doit pas subir de baisse sensible de performance.

## 3. Architecture d'un Datawarehouse

### 3. 5 Les bases multidimensionnelles et les outils OLAP

#### 3.5.3 Les 12 règles de E.F. CODD (1993)

**Architecture client-serveur** : Tout produit OLAP doit fonctionner en mode C/S avec une répartition des traitements.

**Dimension générique** : Chaque dimension (avec l'analyse) doit être équivalent aux autres à la fois dans sa structure et dans ses capacités opérationnelles. Une seule structure logique dans l'ensemble des dimensions.

**Gestion dynamique des matrices creuses** : OLAP doit gérer les cellules non renseignées de manière optimale.

**Support multi-utilisateurs** : OLAP doit assurer un accès simultané aux données, gérer l'intégrité et la sécurité de ces données.

## 3. Architecture d'un Datawarehouse

### 3. 5 Les bases multidimensionnelles et les outils OLAP

#### 3.5.3 Les 12 règles de E.F. CODD (1993)

**Opérations entre les dimensions** : OLAP doit gérer des calculs associés entre les dimensions sans faire appel à l'utilisateur pour définir le contenu de ces calculs

**Manipulation intuitive** : Minimiser le recours à des menus ou les allers et retours avec l'interface utilisateur

**Flexibilité des restitutions** : convivialité des états de gestion ou des états de sortie - ergonomie

**Nombre de dimensions et niveaux de hiérarchie illimité** : l'outil doit gérer au moins quinze dimensions et ne pas limiter le nombre de niveaux hiérarchiques.

## 3. Architecture d'un Datawarehouse

### 3. 5 Les bases multidimensionnelles et les outils OLAP

#### 3.5.4 Fast Analysis of Shared Multidimensional Information (FASMI)

**Analyse** : fournir des possibilités d'analyse (statistiques et autres)

**Rapide** : l'essentiel des réponses doit être rendu dans un délai de moins de cinq secondes

**Information** : accéder à l'ensemble des données indépendamment de leur localisation

**Multidimensionnelle** : fournir une vue conceptuelle multidimensionnelle

**Partagée** : être accessible à un grand nombre d'utilisateurs et ne pas limiter le nombre de niveaux hiérarchiques.

## 3. Architecture d'un Datawarehouse

### 3. 5 Les bases multidimensionnelles et les outils OLAP

#### 3.5.5 *Les outils relationnels OLAP*

**Outils relationnels** : requêteurs, infocentres, jointures complexes  
exemple : Business Objects

**Hypercubes relationnels** : les données sont stockées dans une BD relationnelle, mais avec une structure adaptée aux données multidimensionnelles  
exemple : SGBD relationnels

**OLAP relationnel (ROLAP)** : ces outils utilisent directement le modèle relationnel. Au travers des méta-données, ils permettent de transformer l'analyse multidimensionnelle en requêtes SQL : distinguent les axes d'analyse et les faits à observer (modèles en étoile ou en flocon)

## 3. Architecture d'un Datawarehouse

### 3. 5 Les bases multidimensionnelles et les outils OLAP

#### 3.5.6 *Intégration Infocentre Hypercube*

- Principe proche de l'OLAP relationnel
- Intégration d'un outil d'infocentre et d'un outil d'analyse multidimensionnelle dans une même interface située sur le poste client
- L'outil d'infocentre assure la gestion d'un référentiel commun, la sélection des données et leur valorisation
- L'outil multidimensionnel assure la création d'un hypercube, l'implémentation des fonctionnalités OLAP (consolidation, zoom avant, glisser-déplacer, gestion des seuils, etc.)

## 3. Architecture d'un Datawarehouse

### 3. 5 Les bases multidimensionnelles et les outils OLAP

#### 3.5.7 Les outils multidimensionnels MOLAP

- Les BD multidimensionnelles sont propriétaires (pas de standard)
- Les données sont dynamiquement structurées et compressées (optimisation de l'espace disque)
- Les données sont organisées en dimensions et hiérarchies
- Les formules de calcul sont généralement complexes
- Les temps de réponse sont constants

## 3. Architecture d'un Datawarehouse

### 3. 5 Les bases multidimensionnelles et les outils OLAP

#### 3.5.7 Les outils multidimensionnels MOLAP

- La constitution de la base se fait selon le processus suivant
  - extraction des données provenant des SGBD ou fichiers
  - décomposition des données en dimensions, attributs et variables
  - calcul des consolidations
  - chargement de l'hypercube selon la structure dimensionnelle choisie
- L'interrogation de la base possède les caractéristiques suivantes :
  - interface graphique (drill down, slice and dice, etc)
  - gestion des seuils et des alertes (codage couleurs)
  - temps de réponse court et constant
  - SQL non implémenté
  - Exemple : Oracle Express

## 3. Architecture d'un Datawarehouse

### 3. 6 Les limites du multidimensionnel

- Format et langage propriétaire
- Structure figée
- Accès au détail difficile
- Peu d 'outils disponibles
- Outils d 'administration insuffisants
- Difficulté de réaliser des sélections sur un hypercube
- Pas de standard ni pour la structure physique ni pour l 'interrogation
- Manque de souplesse et absence de gestion de méta-données

## 3. Architecture d'un Datawarehouse

### 3. 7 Conclusion

- Un marché florissant
  - nombreux outils (ROLAP,MOLAP,..)
  - concentration du nombre d 'éditeurs de logiciels
- Nécessité de méthodologie de conception
  - démarche
  - modélisation conceptuelle et logique
  - implication des utilisateurs
- Un avenir réel
  - l 'informatique opérationnelle est mature
  - la demande des utilisateurs est importante
  - la technologie est disponible.