



Entrepôts de données

Données, modélisation,
intégration et analyse

Amri Manel
Experte en TIC

Contexte (1)

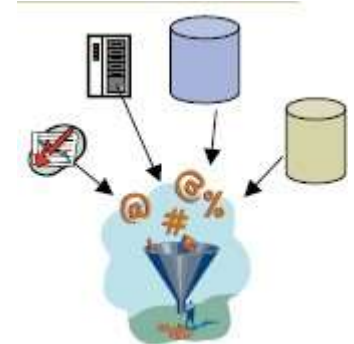
- Besoin :
 - Prise de décisions stratégiques et tactiques
 - Réactivité
- Qui :
 - les décideurs (non informaticiens, non statisticiens)
- Comment :
 - Répondre aux demandes d'analyse de données
 - Dégager des informations qualitatives nouvelles



Contexte (2)

- Type de données : données opérationnelles (de production)

- Bases de données, Fichiers, Paye, Gestion RH, ...



- Caractéristiques des données :

- Distribuées : systèmes éparpillés

- Hétérogènes : systèmes et structures de données différents

- Détaillées : organisation de données selon les processus fonctionnels et données trop abondantes pour l'analyse

- Peu/pas adaptées à l'analyse : des requêtes lourdes peuvent bloquer le système transactionnel

- Volatiles : pas d'historisation systématique



Problématique (1)

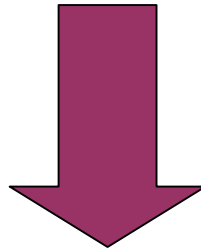
Nous avons donc :

- Une grande masse de données
 - Distribuées
 - Hétérogènes
 - Très détaillées
- à traiter
 - Synthétiser / résumer
 - Visualiser
 - Analyser
- pour une utilisation par des
 - Experts / analystes d'un métier
 - Non informaticiens
 - Non statisticiens

Problématique (2)

- Comment répondre aux besoins de décideurs afin d'améliorer les performances décisionnelles de l'entreprise?
 - En donnant un accès rapide et simple à l'information stratégique
 - En donnant du sens aux données
 - En donnant une vision transversale des données de l'entreprise (intégration de différentes bases de données)
 - En extrayant, groupant, organisant, corrélant et transformant (résumé, agrégation) les données

Problématique (3)



Mettre en place un SI dédié aux applications décisionnelles : un entrepôt de données (*datawarehouse*)

- Transformer des données de production en informations stratégiques

données
run the business

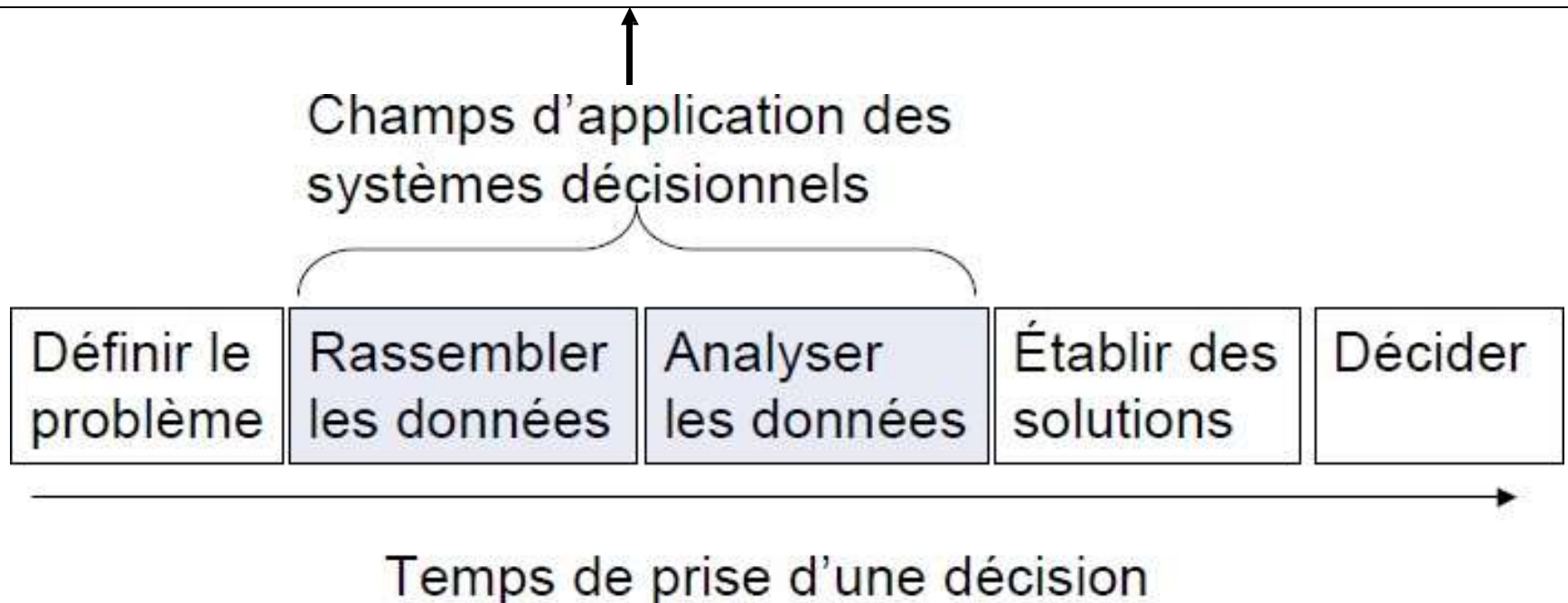
→ informations
manage the business

Le processus de prise de décision (1)

Business Intelligence (BI)

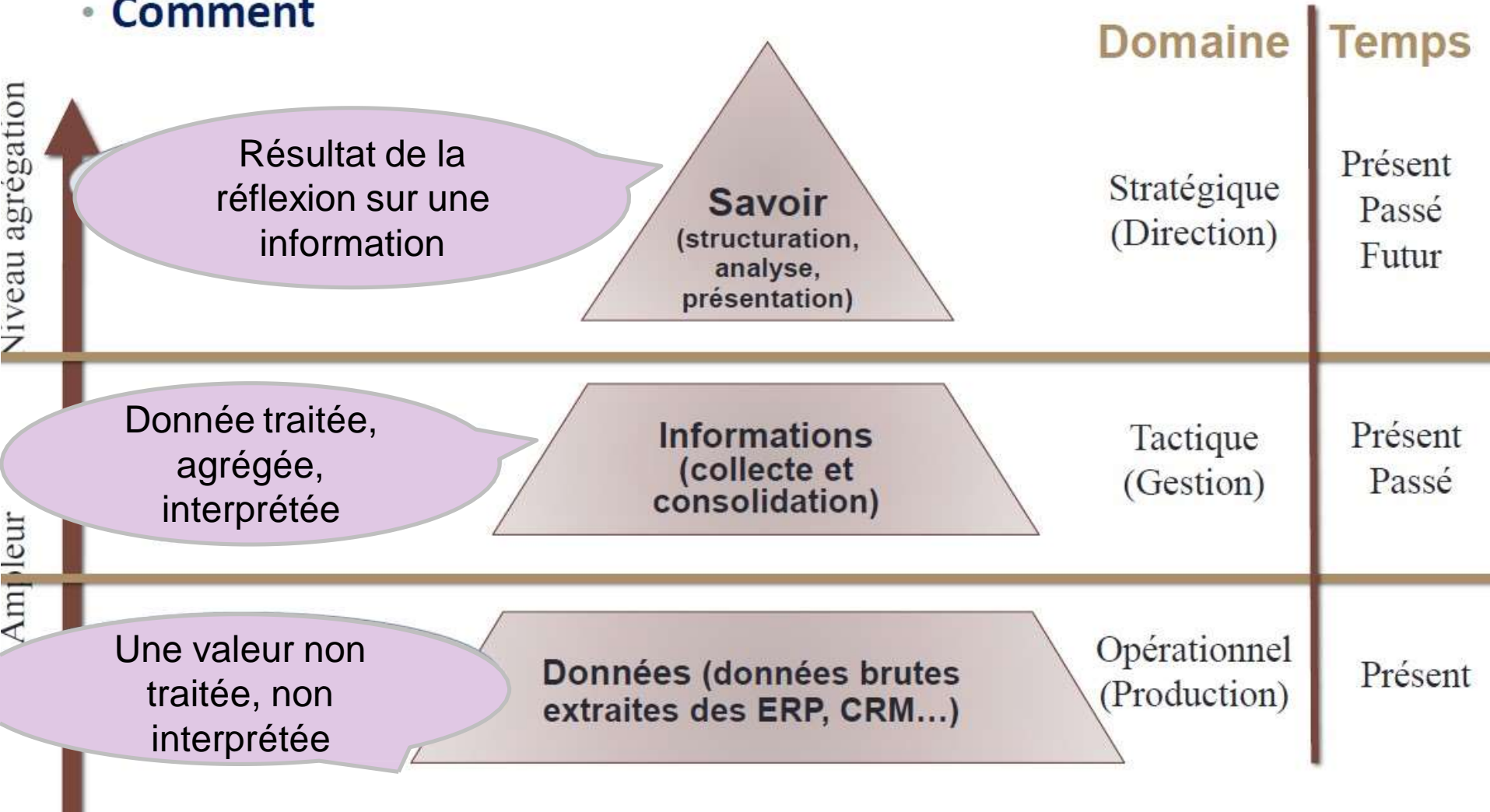
Moyens, outils et méthodes qui permettent à un décideur

- d'avoir une vue d'ensemble de l'activité traitée
- de trouver l'**information** pertinente et complète pour prendre rapidement la meilleure décision



Données, Informations, Connaissances (1)

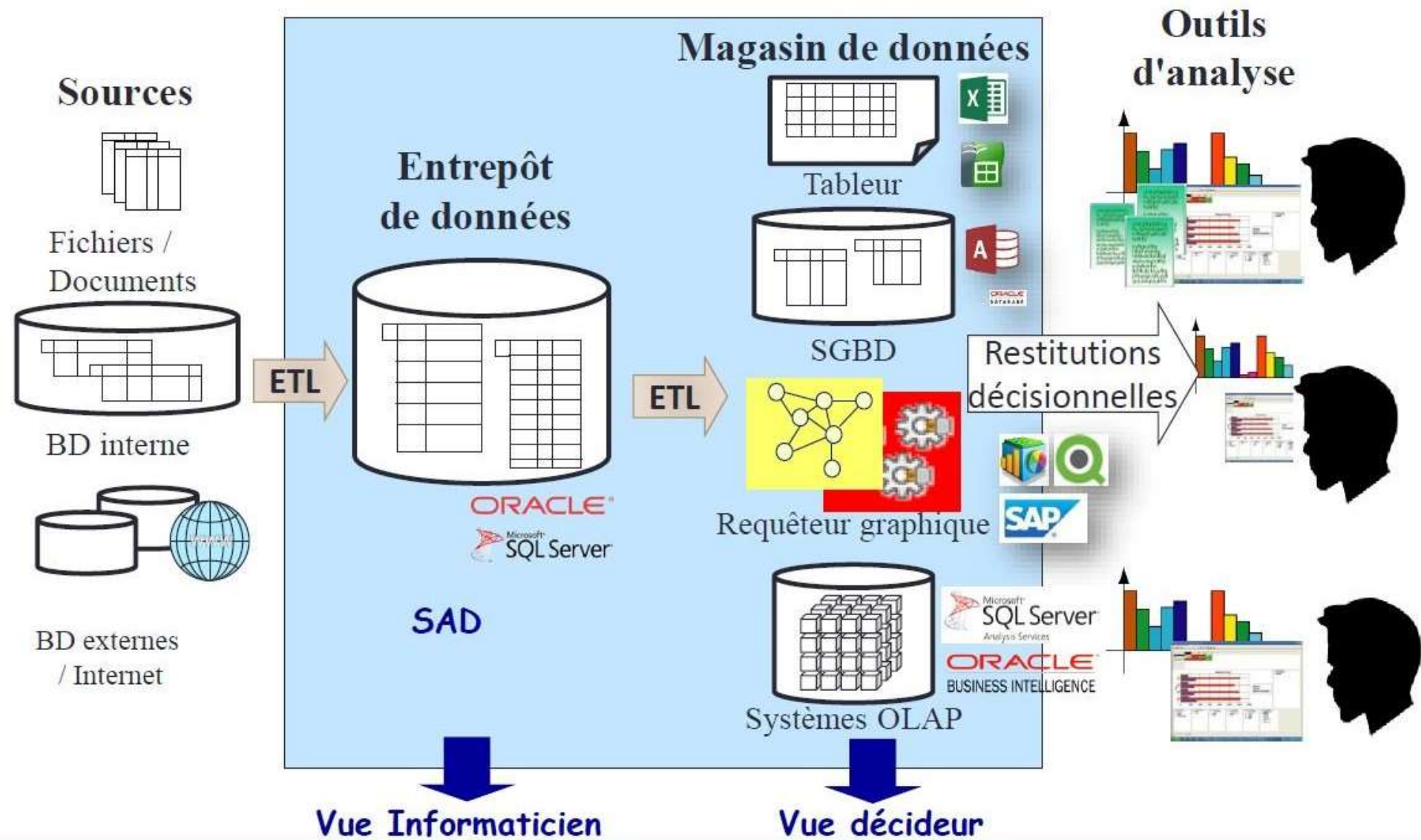
• Comment





Le stockage (DW/DM)

Stockage données décisionnelles



L'entrepôt : Définition

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

W.H. Inmon (1996)

- C'est une BD à des fins d'analyse !!

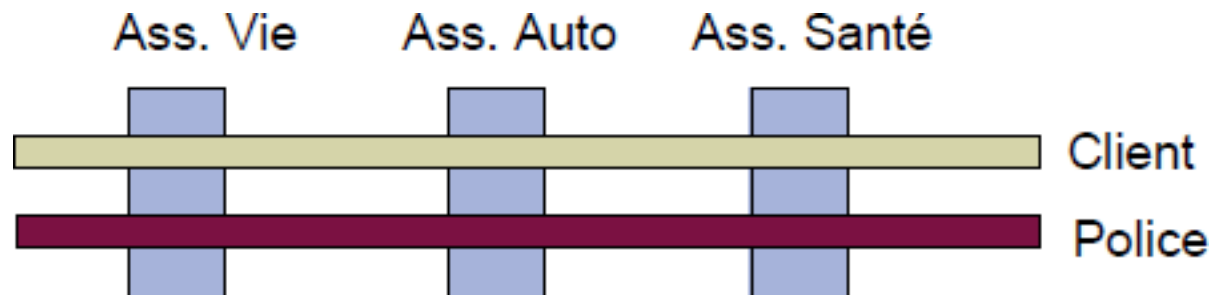


L'entrepôt

- Objectif : Préparation des données décisionnelles
- Principe : Lieu de stockage centralisé d'un extrait des sources pertinent pour les décideurs, historisé, non volatile, disponible pour l'interrogation décisionnelle, organisé selon un modèle informatique facilitant la gestion des données

Caractéristiques d'un DW (1)

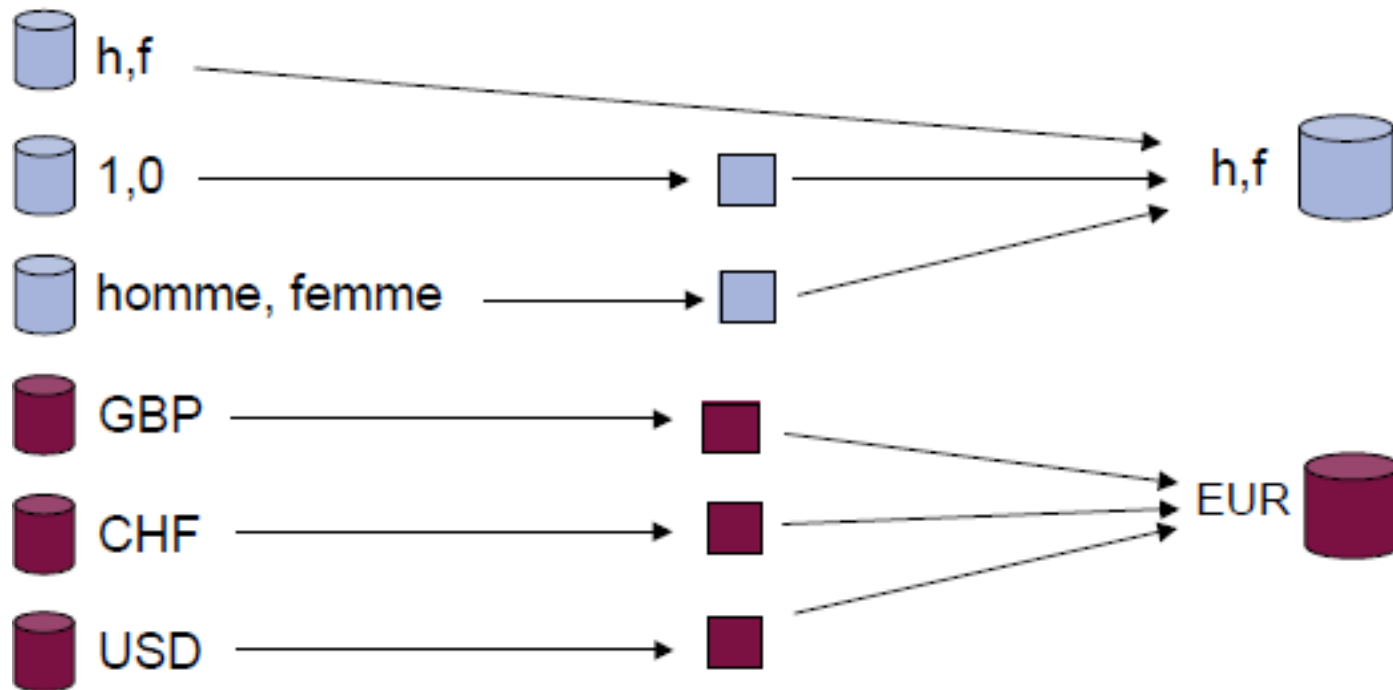
- Données orientées sujet
 - Regroupe les informations des différents métiers
 - Ne tiens pas compte de l'organisation fonctionnelle des données



Caractéristiques d'un DW (2)

■ Données intégrées

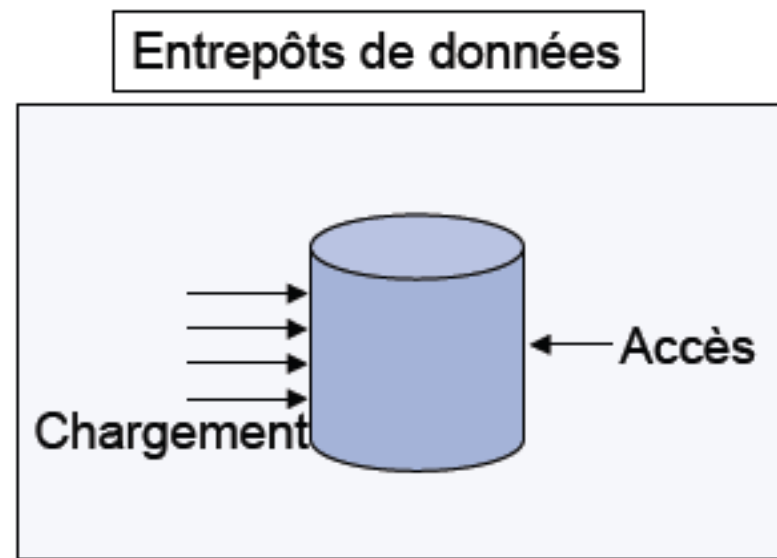
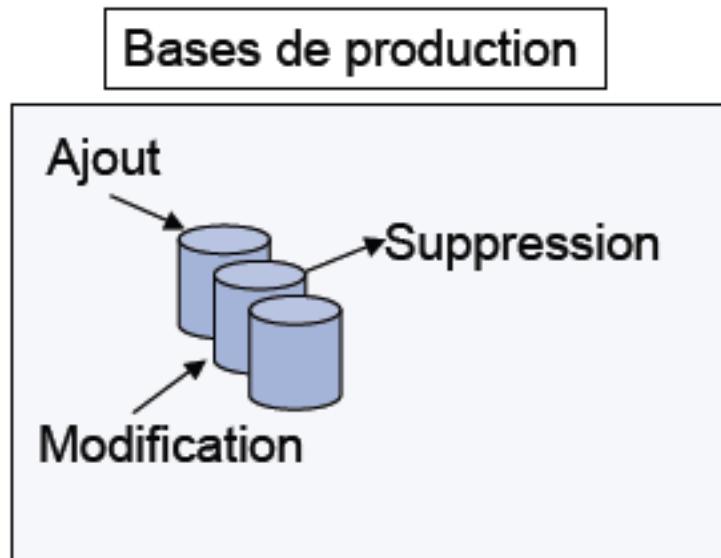
- Normalisation des données
- Définition d'un référentiel unique



Caractéristiques d'un DW (3)

■ Données non volatiles

- Traçabilité des informations et des décisions prises
- Copie des données de production



Caractéristiques d'un DW (4)

■ Données historisées / datées

- Les données persistent dans le temps
- Mise en place d'un référentiel temps

Base de
production

Image de la base en Mai 2005

Répertoire

Nom	Ville
Dupont	Paris
Durand	Lyon

Image de la base en Juillet 2006

Répertoire

Nom	Ville
Dupont	Marseille
Durand	Lyon

Entrepôt
de
données

Calendrier

Code	Année	Mois
1	2005	Mai
2	2006	Juillet

Répertoire

Code	Année	Mois
1	Dupont	Paris
1	Durand	Lyon
2	Dupont	Marseille

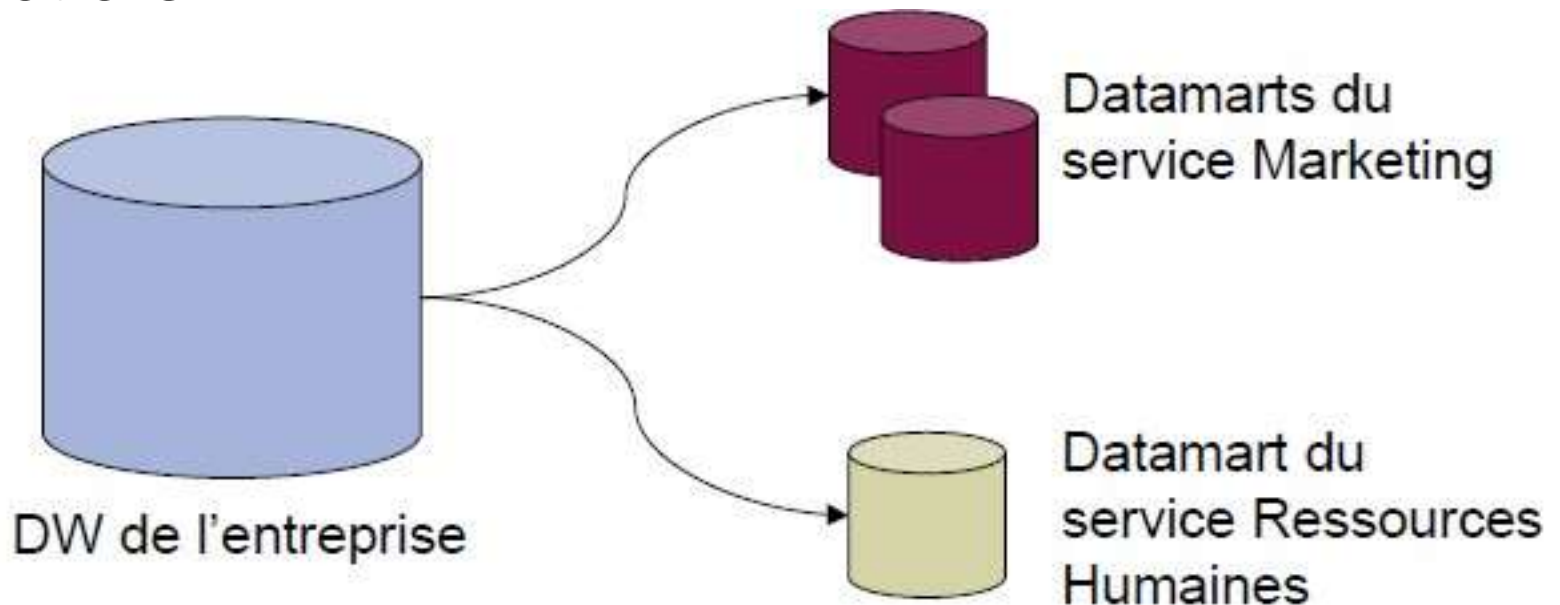
Caractéristiques d'un DW (5)

- Inconvénient :

De par sa taille, le DW est rarement utilisé directement par les décideurs car il contient plus que nécessaire pour une classe de décideurs

Le datamart

- Sous-ensemble d'un entrepôt de données
- Destiné à répondre aux besoins d'un secteur ou d'une fonction particulière de l'entreprise
- Point de vue spécifique selon des critères métiers



Le datamart

- Objectif : Présentation des données décisionnelles
- Principe :
 - Extrait de l'entrepôt de données
 - Adapté aux besoins d'une classe de décideurs
 - Organisé selon un modèle informatique adapté aux outils décisionnels

Pourquoi pas un SGBD ? (1)

■ Fonctions d'un SGBD :

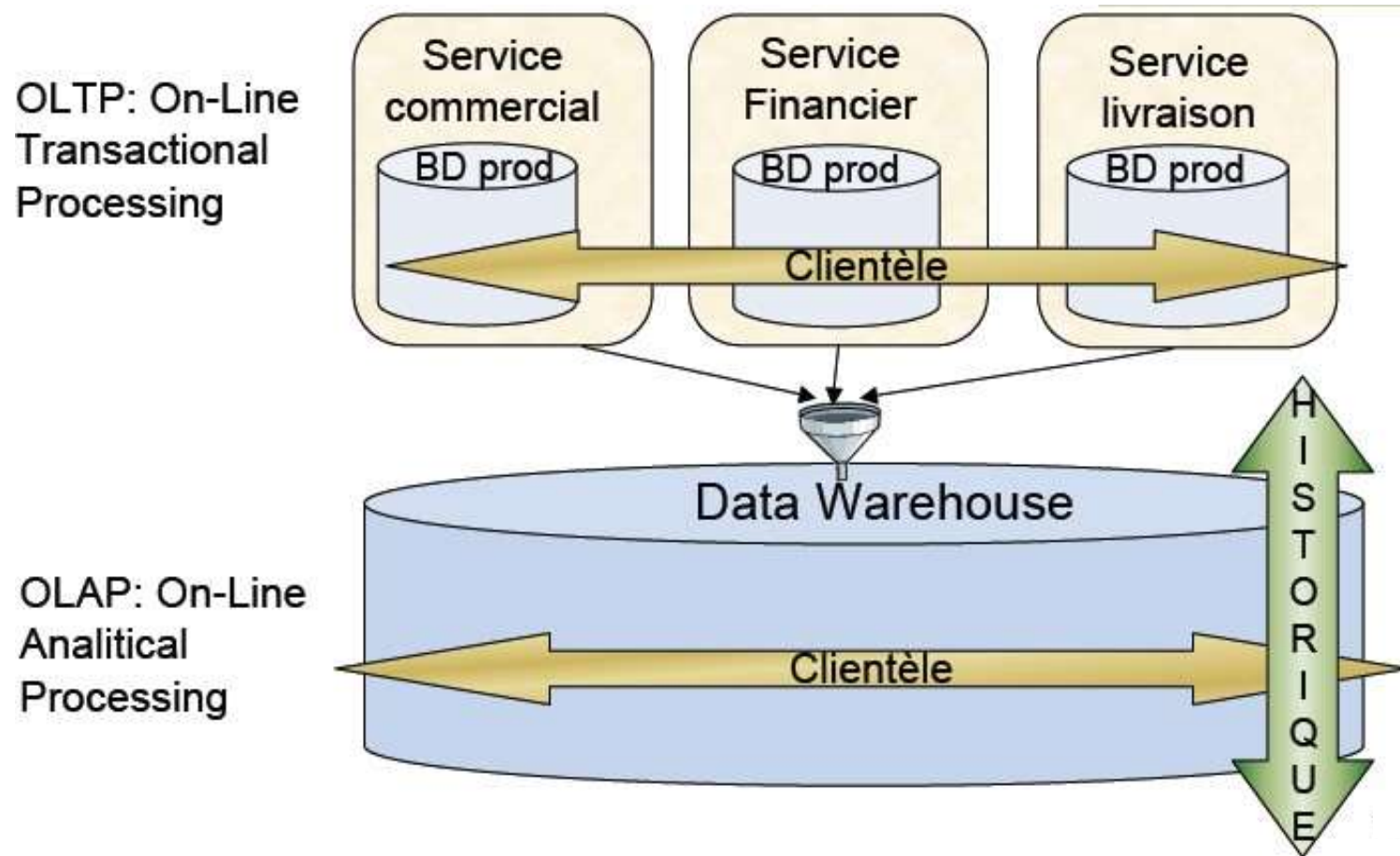
- Systèmes transactionnels (OLTP)
- Permettre d'insérer, modifier, interroger rapidement, efficacement et en sécurité les données de la base
- Sélectionner, ajouter, mettre à jour, supprimer des tuples
- Répondre à de nombreux utilisateurs simultanément

Pourquoi pas un SGBD ? (2)

■ Fonctions d'un DW :

- Systèmes pour l'aide à la prise de décision (OLAP)
- Regrouper, organiser des informations provenant de sources diverses
- Intégrer et stocker les données pour une vue orientée métier
- Retrouver et analyser l'information rapidement et facilement

Pourquoi pas un SGBD ? (4)



Pourquoi pas un SGBD ? (3)

	OLTP	DW
Utilisateurs	Nombreux Employés	Peu Analystes
Données	Alphanumériques Détaillées / atomiques Orientées application Dynamiques	Numériques Résumées / agrégées Orientées sujet Statiques
Requêtes	Prédéfinies	« one-use »
Accès	Peu de données (courantes)	Beaucoup d'informations (historisées)
But	Dépend de l'application	Prise de décision
Temps d'exécution	Court	Long
Mises à jour	Très souvent	Périodiquement