

# Business Intelligence

Med. AMNAI  
Master spécialisé  
BIG DATA Cloud Computing

# Plan

## 1 Introduction

# Plan

- 1 Introduction
- 2 Les entrepôts de données (Datawarehouse)

# Plan

- 1 Introduction
- 2 Les entrepôts de données (Datawarehouse)
- 3 Les magasins de données (Datamarts)

# Plan

- 1 Introduction
- 2 Les entrepôts de données (Datawarehouse)
- 3 Les magasins de données (Datamarts)
- 4 Architecture d'un Datawarehouse

# BI tente de répondre aux questions

- **Analyse descriptive** : que s'est-il passé ?
- **Analyse diagnostique** : pourquoi est-il arrivé ?
- **Analyse prédictive** : que se passera-t-il si nous ne faisons rien et que les tendances se poursuivent ?
- **Analyse prescriptive** : que devons-nous faire ? Que se passera-t-il si nous faisons X ?

# Contexte des Datawarehouse

- **Besoin** : prise de décisions stratégiques ;
- **Pourquoi** : besoin de réactivité ;
- **Qui** : les décideurs (non informaticiens) ;
- **Comment** : répondre aux demandes d'analyse des données, dégager des informations qualitatives nouvelles.

# Les données utilisables par les décideurs

- **Données opérationnelles (de production) :**
  - Bases de données (Oracle, SQL Server) ;
  - Fichiers, . . . . ;
  - Paye, gestion des RH, gestion des commandes, . . . ;
- **Caractéristiques de ces données :**
  - **Distribuées** : systèmes éparpillés ;
  - **Hétérogènes** : systèmes et structures de données différents ;
  - **Détaillées** : organisation des données selon les processus fonctionnels, données surabondantes pour l'analyse ;
  - **Peu/pas adaptées à l'analyse** : les requêtes lourdes peuvent bloquer le système transactionnel ;



# Problématique

- **Comment répondre aux demandes des décideurs ?**
  - En donnant un accès rapide et simple à l'information stratégique ;
  - En donnant du sens aux données ;
- ==» Mettre en place un système d'information dédié aux applications décisionnelles :

DataWarehouse (Entrepôts de Données)

## Domaines d'utilisation des DW

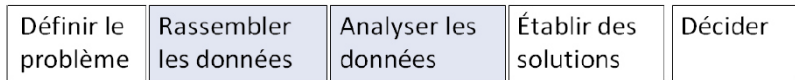
- **Banque** : Risques d'un prêt, prime plus précise ;
- **Santé** : Épidémiologie, Risque alimentaire ;
- **Commerce** : Ciblé la clientèle, Déterminer des promotions ;
- **Logistique** : Adéquation demande/production ;
- **Assurance** : Risque lié à un contrat d'assurance (voiture) ;

## Quelques métiers du décisionnel

- **Strategic Performance Management :**
  - Déterminer et contrôler les indicateurs clé de la performance de l'entreprise.
- **Finance Intelligente :**
  - Planifier, analyser et diffuser l'information financière. Mesurer et gérer les risques.
- **Human Capital Management (gestion de la relation avec les employés) :**
  - Aligner les stratégies RH, les processus et les technologies.
- **Customer Relationship Management (gestion de la relation client) :**
  - Améliorer la connaissance client, identifier et prévoir la rentabilité client, accroître l'efficacité du marketing client.
- **Supplier Relationship Management (gestion de la relation fournisseur) :**
  - Classifier et évaluer l'ensemble des fournisseurs. Planifier et

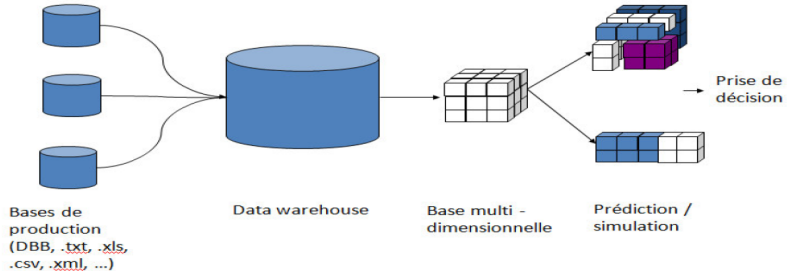
# Processus de prise de décision

Champs d'application des  
systèmes décisionnels



Processus de prise d'une décision

# Processus de prise de décision



# Définition d'un Datawarehouse (DW)

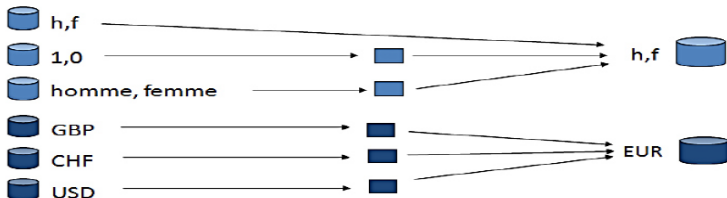
- W. H. Inmon (1996) : " Le Datawarehouse est une collection de données orientées sujet, intégrées, non volatiles et historisées, organisées pour le support d'un processus d'aide à la décision "
- Base de données dans laquelle sont déposées après nettoyage et homogénéisation les informations en provenance des différents systèmes de production de l'entreprise (OLTP).

# Données orientées sujet

- un DW rassemble et organise des **données associées** aux différentes structures fonctionnelles de l'entreprise, pertinentes pour un **sujet** ou **thème** et nécessaires aux besoins d'analyse.
- Ne tiens pas compte de l'organisation fonctionnelle des données.

# Données intégrées

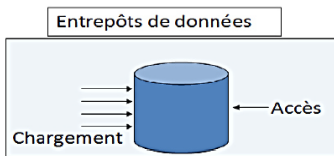
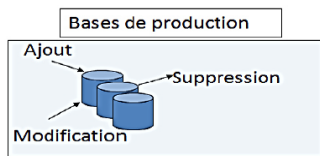
- Les données résultent de l'intégration de données provenant (**flux de données** ou "**mapping**") de différentes sources pouvant être hétérogènes.
- **Normalisation** des données.
- Définition d'un **référentiel unique**.





## Données non volatiles

- Les données du DW sont essentiellement utilisées en interrogation (consultation) et ne peuvent pas être modifiées (sauf certain cas de rafraîchissement).
- **Traçabilité** des informations et des décisions prises.
- **Copie** des données de production.



# Historisées (Données datées)

- les données d'un DW représentent l'activité d'une entreprise durant une certaine période (plusieurs années) permettant d'analyser les variations d'une donnée dans le temps.

# Bases de données et les Datawarehouse

- Ils ont des **objectifs** différents et font des **traitements** différents :
  - Ils **stockent** des données différentes.
  - Ils font l'objet de **requêtes** différentes.
- SGBD et DW ont besoin d'une **organisation** différente des données.

# Bases de données

- Les SGBD sont des systèmes dont le mode de travail est transactionnel (**OLTP** : On-Line Transaction Processing).
- Ils permettent d'insérer, modifier, interroger des informations rapidement, efficacement et en sécurité.

## Objectifs principaux

- Sélectionner, ajouter, mettre à jour et supprimer des Tuples.
- Ces opérations sont accessibles à de nombreux utilisateurs.

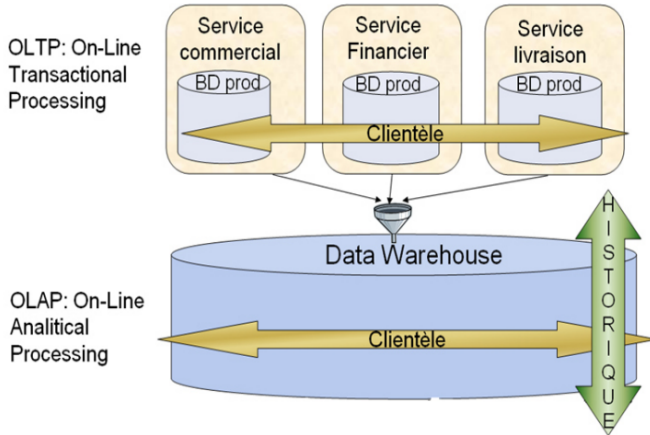
# Datawarehouse

- Un datawarehouse est un système conçu pour l'aide à la prise de décision (Mode de travail : **OLAP** On-Line Analytical Processing).
- La plupart du temps sont utilisés en lecture.

## Objectifs principaux

- 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 les informations facilement et rapidement.

# Alimentation des Datawarehouse



# Comparaison détaillé

OLTP	OLAP (DW)
Orienté transaction	Orienté analyse
Orienté application	Orienté sujet
Données courantes	Données historisées
Données détaillées	Données agrégées
Données évolutives	Données statiques
Utilisateurs nombreux, administrateurs/opérationnels	Utilisateurs peu nombreux, manager
Temps d'exécution: court	Temps d'exécution: long

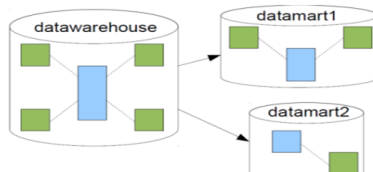
## Comparaison détaillé (suite)

OLTP	OLAP (DW)
Utilisées de façon répétitive	Utilisées de façon aléatoire
Réalisation des opérations au jour le jour	Cycle de vie différent
Uniques(pas de redondance en théorie)	Peuvent être redondantes
Petite quantité de données utilisées par un traitement	Grande quantité de données utilisée par les traitements



# Notion de Datamart

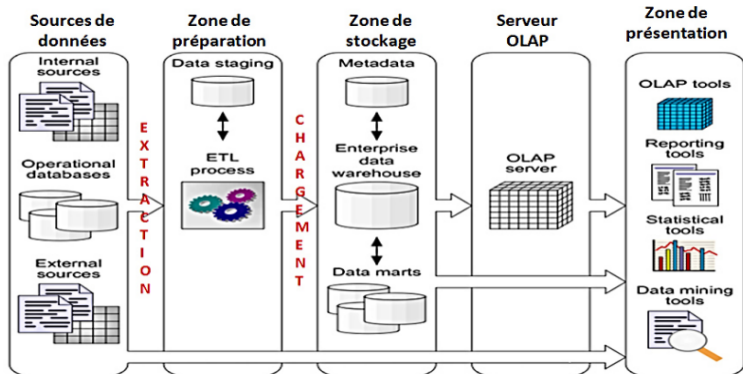
- **Sous-ensemble** d'un "Datawarehouse".
- Destiné à répondre aux **besoins** d'un secteur ou d'une fonction particulière de l'entreprise (Sujet) :
  - pour une **classe** d'utilisateurs ;
  - pour un **besoin** d'analyse spécifique.
- Point de vue spécifique selon des **critères métiers**.
  - comportement de clientèle.



## Intérêt d'un Datamart

- Nouvel environnement structuré et formaté en fonction des besoins d'un **métier** ou d'un **usage particulier**.
- Moins de données que DW :
  - Plus **facile à comprendre**, à manipuler ;
  - Amélioration des temps de réponse.
- Utilisateurs plus ciblés : DM plus facile à définir.

# Architecture générale



# Flux de données

- **Flux entrant :**
  - **Extraction** : multi-source, hétérogène ;
  - **Transformation** : filtrer, trier, homogénéiser, nettoyer ;
  - **Chargement** : insertion des données dans l'entrepôt.
- **Flux sortant :**
  - Mise à disposition des données aux utilisateurs (reporting).

## Différentes zones de l'architecture

- **Zone de préparation (Staging area) :**
  - Zone temporaire de stockage des données extraites ;
  - Réalisation des transformations avant l'insertion dans le DW (Nettoyage, Normalisation)
  - Données souvent détruites après chargement dans le DW
- **Zone de stockage (DW, DM) :**
  - On y transfère les données nettoyées.
  - Stockage permanent des données.
- **Zone de présentation**
  - Donne accès aux données contenues dans le DW ;
  - Peut contenir des outils d'analyse programmés (Rapports, Requêtes, Algorithmes DATAMINING,...)

# Processus de construction et exploitation d'un DW

- **Construction de la BD décisionnelle**
  - Modélisation conceptuelle des données multiformes et multi-sources ;
  - Conception de l'entrepôt de données ;
  - Alimentation de l'entrepôt (extraire, nettoyer, transformer, charger) ;
  - Stockage physique des données.
- **Sélection des données à analyser**
  - Besoins d'analyse de l'utilisateur ;
  - Data marts (Magasins de données) ;
  - Cubes multidimensionnels ;
  - Tableaux ou tables bidimensionnels ;
- **Analyse des données**