

# ENTREPÔT DE DONNÉES

INES SLIMENE

# PLAN

- Objectifs
- Définition d'un DW
- Caractéristiques des DW
- Data Warehouse et Datamart
- Architecture du DW
- Construction et d'exploitation d'un DW

# OBJECTIFS

- regrouper, organiser des informations provenant de sources diverses,
  - ◆ les intégrer et les stocker pour donner à l'utilisateur une vue orientée métier (sujet),
  - ◆ retrouver et analyser l'information facilement et rapidement.
- La simple logique de production (produire pour répondre à une demande) ne suffit plus pour pérenniser l'activité d'une entreprise.
- Pour faire face aux nouveaux enjeux, l'entreprise doit collecter, traiter, analyser les informations de son environnement pour anticiper.
- Problème : l'information produite par l'entreprise est surabondante, non organisée et éparpillée dans de multiples systèmes opérationnels hétérogènes et peut provenir de toutes les places de marchés (mondialisation des échanges).
- Solution : Rassembler et homogénéiser les données.

# DÉFINITION D'UN DW

- **Ralph Kimball**

« un entrepôt de données est constitué peu à peu par les datamart de l'entreprise regroupant le niveau d'agrégation et d'historisation au sein d'une même base ».

- **W. H. Inmon (1996):**

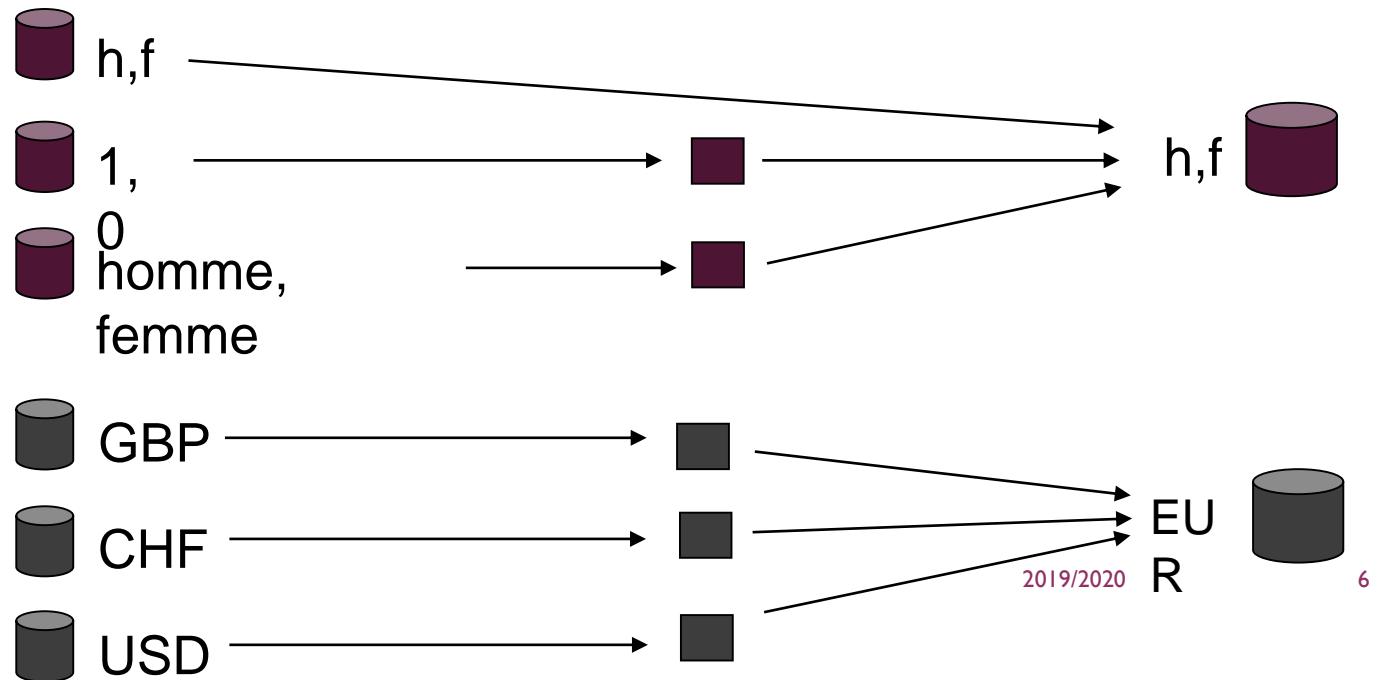
« Le data Warehouse est une collection de données orientées sujet, intégrées, non volatiles et historisées, organisées pour le support d'un processus d'aide à la décision »

## LES 4 CARACTÉRISTIQUES DES DW: DONNÉES ORIENTÉES SUJET

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

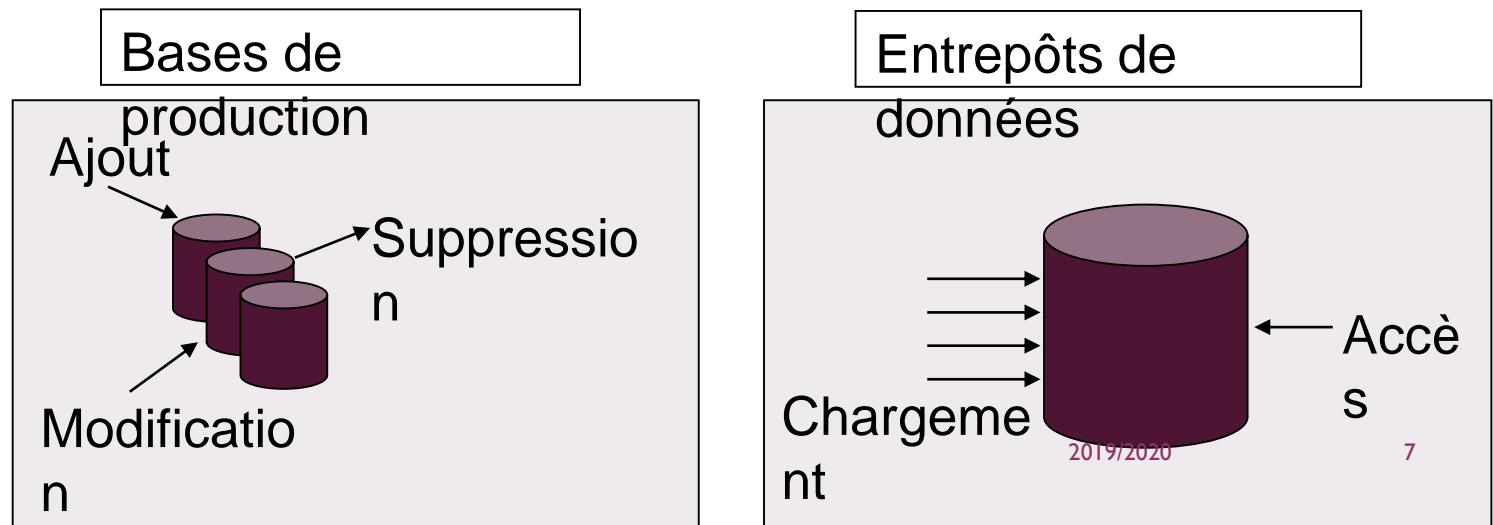
# LES 4 CARACTÉRISTIQUES DES DW : DONNÉES INTÉGRÉES

- Normalisation des données
- Définition d'un référentiel unique
- Les données doivent être variées, triées et transformées dans un format unique pour faciliter et accélérer l'accès.



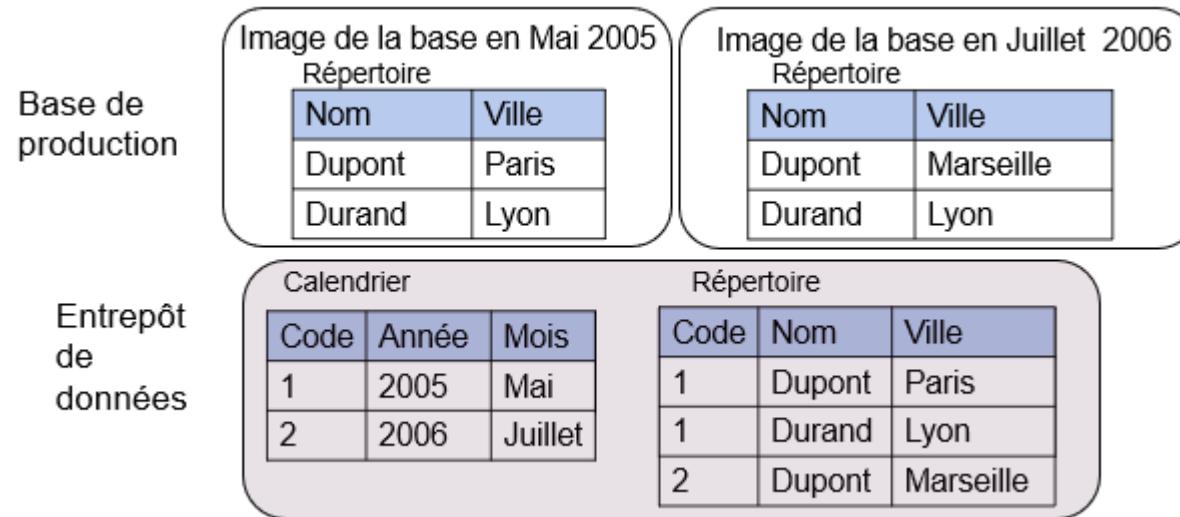
# LES 4 CARACTÉRISTIQUES DES DW : DONNÉES NON VOLATILES

- Afin de conserver la traçabilité des informations et des décisions prises, les informations stockées au sein du Data Warehouse ne doivent pas disparaître...
- Copie des données de production :
  - Stables
  - en lecture seule
  - non modifiables



# LES 4 CARACTÉRISTIQUES DES DW : DONNÉES HISTORIÉES

- Données datées :
- avec une conservation de l'historique et de son évolution pour permettre les analyses comparatives (par exemple, d'une année à l'autre, etc.).
- Dans un Data Warehouse, un référentiel de temps est nécessaire : c'est l'axe temps ou l'axe période.



# DATA WAREHOUSE ET DATAMART

## Les entrepôts de données :

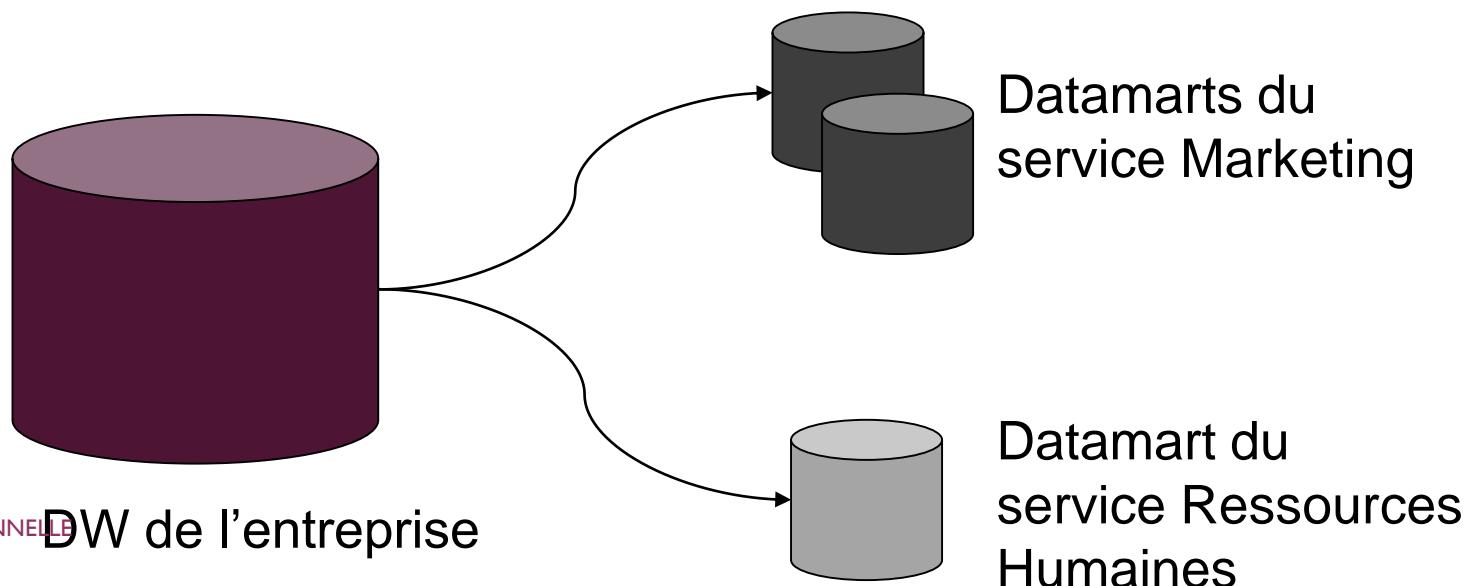
- Nécessitent de puissantes machines pour gérer de très grandes bases de données contenant des données de détail historiées
- Est le lieu de stockage centralisé d'un extrait des bases de production.
- L'organisation des données est faite selon un modèle facilitant la gestion efficace des données et leur historisation.

## Les magasins de données (Data Marts) :

- Sont de petits entrepôts nécessitant une infrastructure plus légère et sont mis en œuvre plus rapidement
- Les données extraites sont adaptées pour l'aide à la décision (pour classe de décideurs, usage particulier, recherche de corrélation, logiciel de statistiques,...)
- L'organisation des données est faite selon un modèle facilitant les traitements décisionnels

# DATAMART

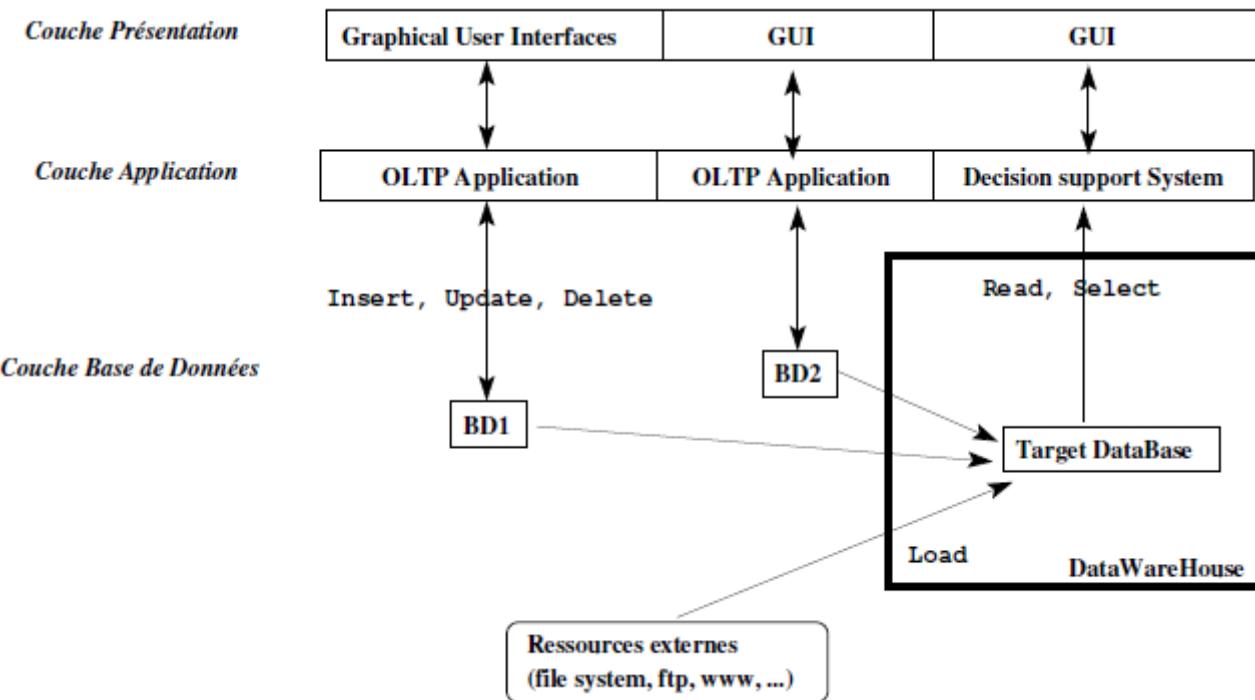
- Sous-ensemble d'un entrepôt de données
- Destiné à répondre aux besoins d'un secteur ou d'une fonction particulière de l'entreprise
- Point de vue spécifique selon des critères métiers



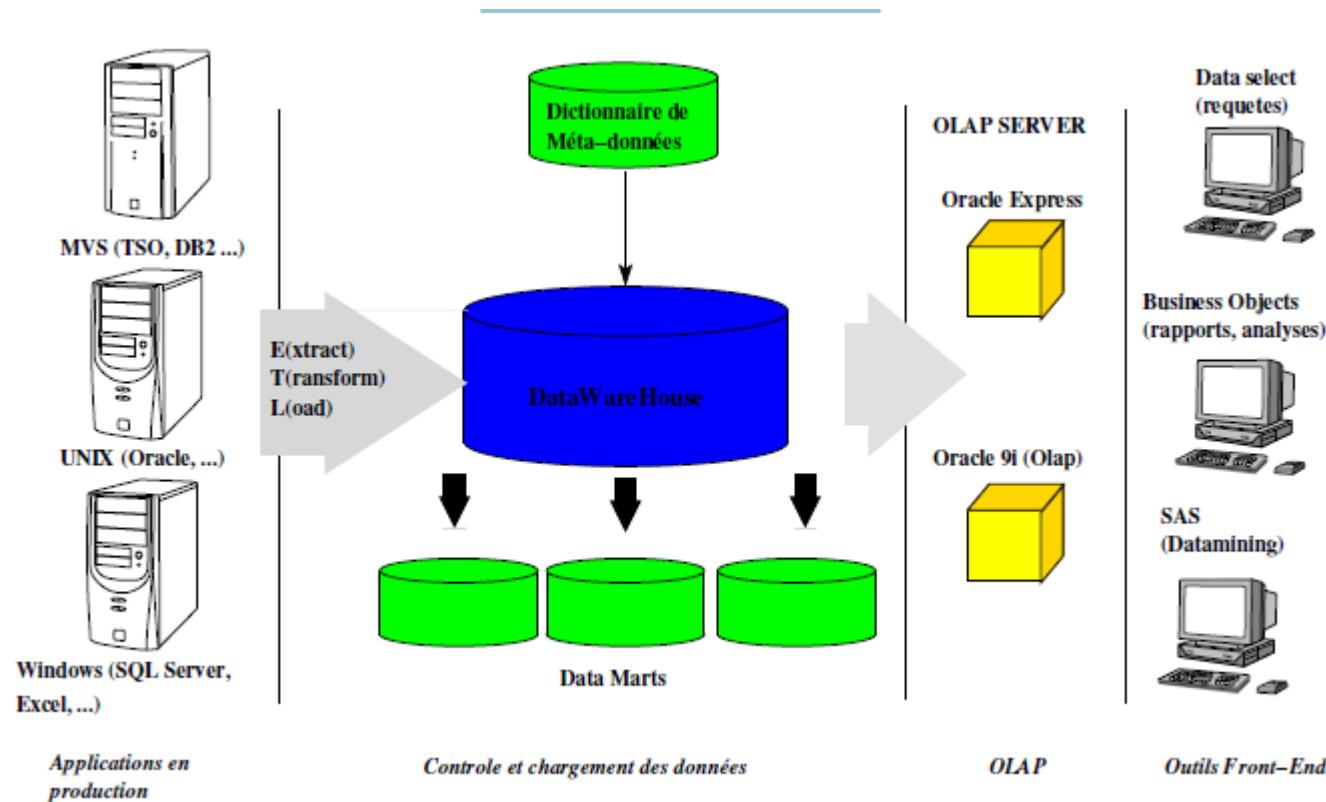
# INTÉRÊT DES DATAMART

- 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 D'UN SYSTÈME OPÉRATIONNEL



# ARCHITECTURE DU DW : ARCHITECTURE MULTI-TIERS



# ARCHITECTURE DU DW

- Zone de préparation (Staging area)
  - Zone temporaire de stockage des données extraites
  - collecte des données sources, qualification et organisation avant de pouvoir les déverser dans le DW
  - Données souvent détruites après chargement dans le DW
- ODS (Operational Data Store)
  - un espace de stockage persistant dans lequel seront stockées les données sources à un niveau détaillé.
  - le socle des données du Data Warehouse.

# ARCHITECTURE DU 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...

# CONSTRUCTION ET EXPLOITATION D'UN DW

- **Étude des besoins:**
  - Définir les objectifs du DW
  - Déterminer le contenu du DW et son organisation, d'après:
    - Les résultats attendus par les décideurs
    - Les requêtes qu'ils formuleront
    - Les projets qui ont été définis
  - Recenser les données nécessaires à un bon fonctionnement du DW:
    - Recenser les données disponibles dans les bases de production
    - Identifier les données supplémentaires requises
  - Choisir les dimensions Typiquement: le temps, le client, le produit, le magasin...
  - Choisir les mesures de fait : Des valeurs numériques additives
  - Choisir la granularité des faits : Niveau de détails des dimensions

# CONSTRUCTION ET EXPLOITATION D'UN DW

Processus en 3 phases :

## I. Construction de la BD décisionnelle

1. Modélisation conceptuelle des données multiformes et multi-sources
2. Conception de l'entrepôt de données
3. Alimentation de l'entrepôt (extraire, nettoyer, transformer, charger)
4. Stockage physique des données

# CONSTRUCTION ET EXPLOITATION D'UN DW

1. Sélection des données à analyser
  1. Besoins d'analyse de l'utilisateur
  2. Datamarts (Magasins de données)
  3. Cubes multidimensionnels
  4. Tableaux ou tables bidimensionnels
2. Analyse des données
  1. Statistiques et reporting, OLAP, Data Mining

# MODÉLISATION

# PLAN

- Modélisation Entité-Relation
- Normalisation dans les BDR
- Limites de la Modélisation E/R
- Les approches de modélisation d'un DW
- Conception logique d'un DW
  - Table de faits
  - Table de dimensions
- Modèles de DW

# MODÉLISATION ENTITÉ-RELATION

- Discipline permettant d'éclairer les relations microscopiques entre les données
- Supprimer la redondance des données
- Simplifier le traitement des transactions
- Aider le concepteur dans la répartition des propriétés entre les entités
- Principes
  - Notion d'identifiant
  - Dépendance fonctionnelle
  - Décomposition
  - Formes normales

# NORMALISATION DANS LES BDR

- Forme normale:
  - Type de relation particulier entre les entités
  - Permet d'éviter les anomalies transactionnelles dues à une mauvaise modélisation des données
  - Permet de vérifier la robustesse de la conception des modèles de données pour éviter les problèmes de redondance et de mise à jour du contexte
- Dans le modèle OLTP, il existe 8 formes normales
- Elles s'emboitent les unes dans les autres
- Le respect d'une FN de niveau supérieur implique le respect des FN des niveaux inférieurs

# LIMITES DE LA MODÉLISATION E/R

- Modèle complexe
- Plusieurs tables et jointures mises en œuvre
- Risque de dégradation des performances
- Pas de compréhension pour l'utilisateur
- Données historiques difficilement représentées
- Contraire aux objectifs du DW

# LES APPROCHES DE MODÉLISATION D'UN DW

- **Bill Inmon** : On ne fait rien tant que tout n'est pas désigné. Le Datawarehouse doit être exhaustif!
- **Ralph Kimball** : Que chacun construise ce qu'il veut, on intégrera ce qu'il faudra quand il faudra!

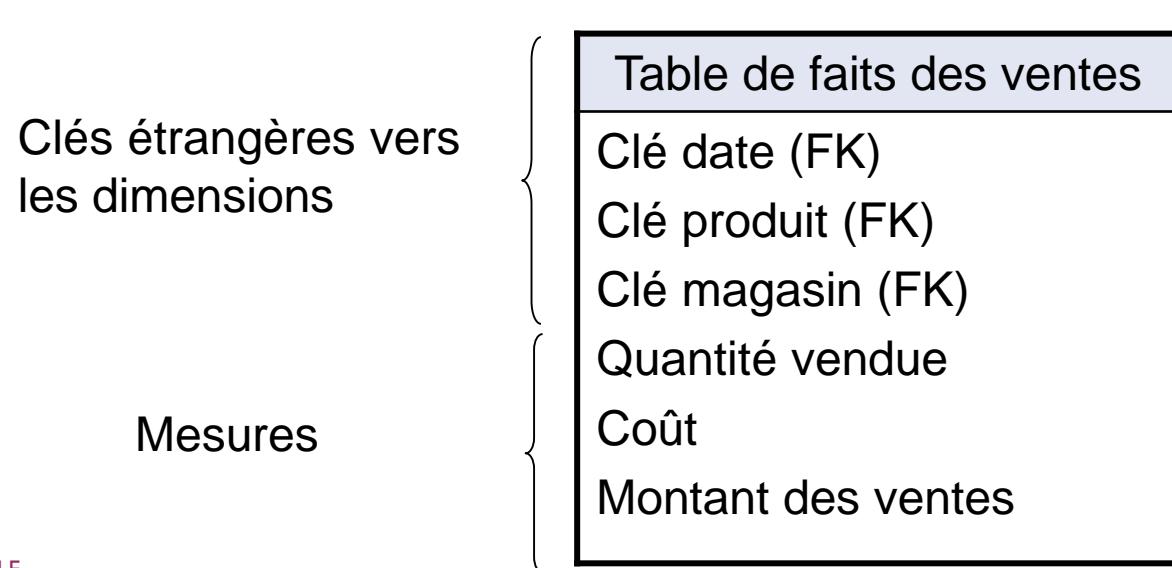
	<b>Kimball</b>	<b>Inmon</b>
<b>Processus</b>	Bottom-Up	Top-Down
<b>Organisation</b>	Datamarts	Datawarehouse
<b>Schématisation</b>	Etoile	Flocon

# CONCEPTION LOGIQUE D'UN DW

- Principe : Ne pas normaliser au maximum
- Définition des objets
  - Table de faits : contient l'information à analyser (par exemple les ventes)
  - Table de dimensions : contiennent les informations sur les dimensions d'analyse (par exemple le lieu, le temps, la description du produit).
  - Les mesures sont les valeurs numériques que l'on compare (ex : montant\_ventes, qte\_vendue) Ces valeurs sont le résultat d'une opération d'agrégation des données.
    - Les mesures sont stockées dans les tables de faits
- Définition des relations entre objets
- Choix d'un modèle de conception (Modèle en étoile, Modèle en flocon..)

# TABLE DE FAITS

- Table principale du modèle dimensionnel
- Contient les données observables (les faits) sur le sujet étudié selon divers axes d'analyse (les dimensions)



# TABLE DE FAITS (SUITE)

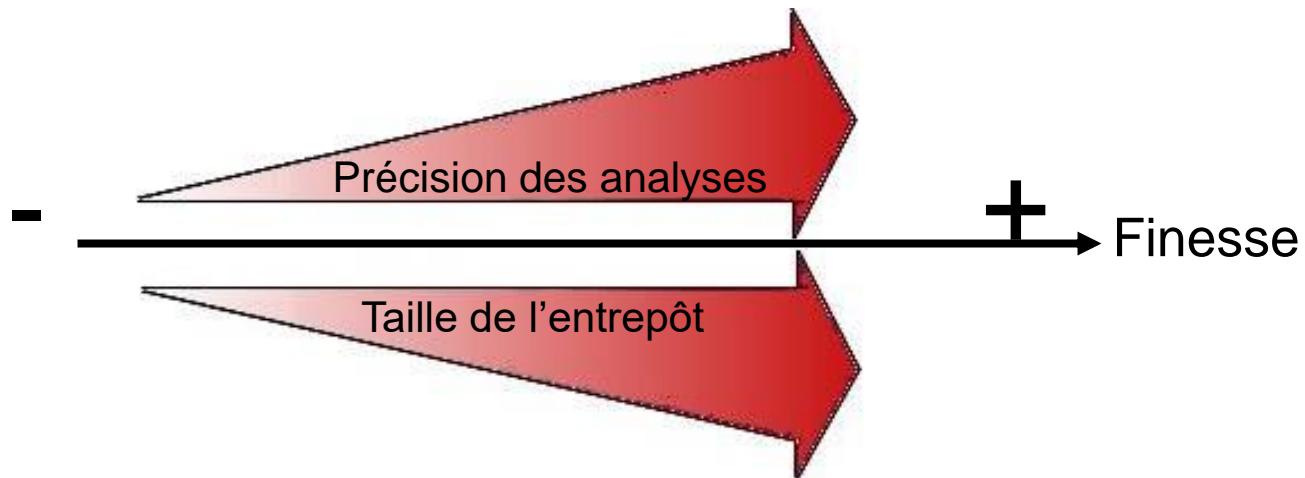
- Fait:
  - Ce que l'on souhaite mesurer
    - Quantités vendues, montant des ventes...
  - Contient les clés étrangères des axes d'analyse (dimension)
    - Date, produit, magasin
  - Trois types de faits:
    - Additif
    - Semi additif
    - Non additif

## TABLE DE FAITS (SUITE)

- Additif: additionnable suivant toutes les dimensions
  - Quantités vendues, chiffre d'affaire
  - Peut être le résultat d'un calcul:
    - Bénéfice = montant vente - coût
- Semi additif: additionnable suivant certaines dimensions
  - Solde d'un compte bancaire:
    - Pas de sens d'additionner sur les dates car cela représente des instantanés d'un niveau
    - $\Sigma$  sur les comptes: on connaît ce que nous possédons en banque
- Non additif: fait non additionnable quelque soit la dimension
  - Prix unitaire: l'addition sur n'importe quelle dimension donne un nombre dépourvu de sens

# GRANULARITÉ DE LA TABLE DE FAITS

- Répondre à la question :
  - Que représente un enregistrement de la table de faits?
- La granularité définit le niveau de détails de la table de faits:
  - Exemple: une ligne de commande par produit, par client et par jour

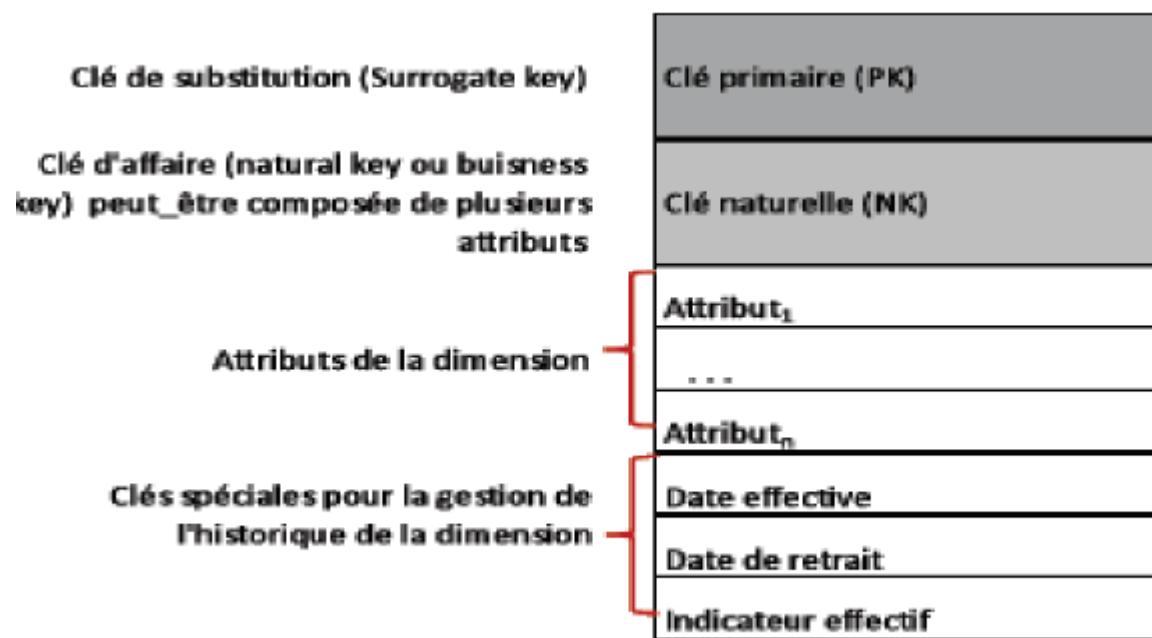


# DIMENSIONS

- Une table de dimension contient le détail sur les faits
- Une table de dimension contient les informations descriptives des valeurs numériques de la table des faits
- Vu que les données dans la table de dimensions sont normalisées, elle contient un plus grand nombre de colonnes
- Une table de dimensions contient en général beaucoup moins d'enregistrements qu'une table des faits
- Les attributs d'une table de dimensions sont souvent utilisés comme «Tête de lignes » et «Tête de colonnes » dans un rapport ou résultat de requête.

# TABLE DE DIMENSION

- Axe d'analyse selon lequel vont être étudiées les données observables (faits)
- Contient le détail sur les faits



# TABLE DE DIMENSION (SUITE)

- Commune à l'ensemble du DW
- Reliée à toute table de faits

Dimension Temps
Clé temps (PK)
Jour
Mois
Trimestre
Semestre
Année
Num_jour_dans_année
Num_semaine.ds_année

# GRANULARITÉ D'UNE DIMENSION

- Une dimension contient des membres organisés en hiérarchie :
  - Chacun des membres appartient à un niveau hiérarchique (ou niveau de granularité) particulier
  - Granularité d'une dimension : nombre de niveaux hiérarchiques
  - Temps :
    - année – semestre – trimestre - mois

# CLÉ DE SUBSTITUTION

- Une Clé de substitution (Surrogate Key) est une clé non intelligente utilisée afin de substituer la clé naturelle (Business Key) qui provient des systèmes opérationnels.
- La clé naturelle est en général composée de plusieurs colonnes
- Dans un système opérationnel, on utilise une clé artificielle afin d'identifier d'une façon unique un élément de l'entité :(client\_id Pour l'entité Client, emp\_id pour l'entité Employé)
- La clé de substitution ne doit pas être confondue avec la clé artificielle attribuée par les systèmes opérationnels.
- La clé de substitution est alors utilisée dans un entrepôt de données pour remplacer et compléter la clé artificielle du système opérationnel afin de rendre un élément unique dans la dimension

# CLÉ DE SUBSTITUTION

- **Remplacer la clé artificielle ou naturelle:**
- Une clé de substitution remplace la clé artificielle en termes d'utilisation, ce n'est plus la clé naturelle qui sera utilisées pour faire les jointures avec les tables des faits ou les autres tables de dimension
- **Compléter l'information:**
- La clé de substitution n'a aucun sens en terme d'affaire, elle est utilisée dans le DW seulement.
- La clé artificielle ou naturelle dans la dimension est toujours nécessaire pour pouvoir faire la correspondance entre l'élément de dimension (un client par exemple) dans le DW et l'élément de la table des clients dans le système opérationnel.

# CLÉ DE SUBSTITUTION

- Performance : Accélère l'accès aux données du moment où l'on va utiliser un index numérique vu que le type de données de la clé de substitution est numérique.
- Indépendance du système source : on ne peut garantir que la clé d'affaire ne change pas dans les systèmes sources.
- Historique des changements et granularité infinie: si l'on désire garder l'historique des changements de la dimension selon certains critères, on doit gérer la clé de substitution. On retrouve facilement avec plusieurs enregistrements de la même clé d'affaire dans la dimension.

# ÉVOLUTION DES DIMENSIONS

- Dimensions à évolution lente
- Dimensions à évolution rapide

# ÉVOLUTION DES DIMENSIONS

- Dimensions à évolution lente
  - Un client peut se marier, avoir des enfants...
  - Un produit peut changer de noms ou de formulation:
    - « Raider » en « Twix »
    - « yaourt à la vanille » en « yaourt saveur vanille »
  - Gestion de la situation, 3 solutions:
    - Écrasement de l'ancienne valeur
    - Versionnement
    - Valeur d'origine / valeur courante

# DIMENSIONS À ÉVOLUTION LENTE (1/3)

- Écrasement de l'ancienne valeur :
  - Correction des informations erronées
- Avantage:
  - Facile à mettre en œuvre
- Inconvénients:
  - Perte de la trace des valeurs antérieures des attributs
  - Perte de la cause de l'évolution dans les faits mesurés

Clé produit	Description du produit	Groupe de produits
12345	Intelli-Kids	<del>Logiciel</del> Jeux éducatifs

## DIMENSIONS À ÉVOLUTION LENTE (2/3)

- Ajout d'un nouvel enregistrement:
  - Utilisation d'une clé de substitution
- Avantages:
  - Permet de suivre l'évolution des attributs
  - Permet de segmenter la table de faits en fonction de l'historique
- Inconvénient:
  - Accroît le volume de la table

Clé produit	Description du produit	Groupe de produits
12345	Intelli-Kids	Logiciel
25963	Intelli-Kids	Jeux éducatifs

## DIMENSIONS À ÉVOLUTION LENTE (3/3)

- Ajout d'un nouvel attribut:
  - Valeur origine/valeur courante
- Avantages:
  - Avoir deux visions simultanées des données :
    - Voir les données récentes avec l'ancien attribut
    - Voir les données anciennes avec le nouvel attribut
  - Voir les données comme si le changement n'avait pas eu lieu
- Inconvénient:
  - Inadapté pour suivre plusieurs valeurs d'attributs intermédiaires

Clé produit	Description du produit	Groupe de produits	Nouveau groupe de produits
12345	Intelli-Kids	Logiciel	Jeux éducatifs

# ÉVOLUTION DES DIMENSIONS

- Dimensions à évolution rapide
  - Subit des changements très fréquents (tous les mois) dont on veut préserver l'historique
  - Solution: isoler les attributs qui changent rapidement

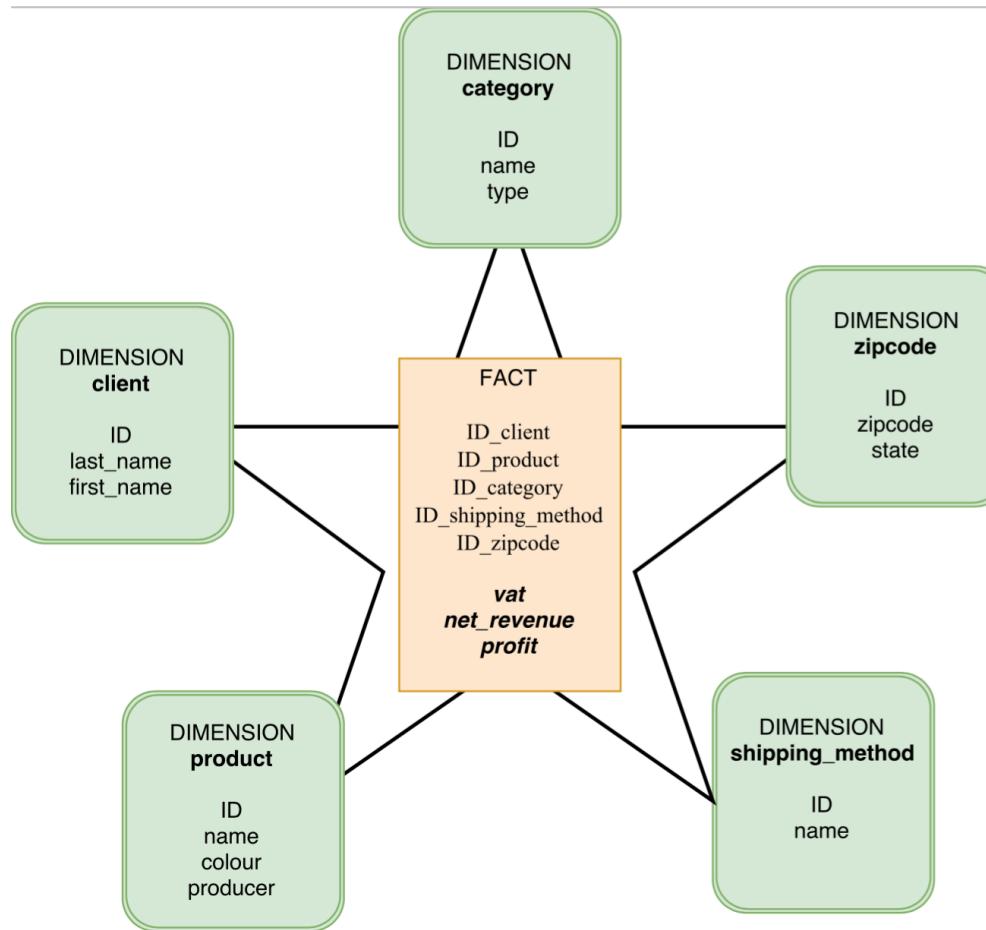
# LES TYPES DE MODÈLES

- Modèle en étoile
- Modèle en flocon
- Modèle en constellation

# MODÈLE EN ÉTOILE

- Une table de fait centrale et des dimensions
- Les dimensions n'ont pas de liaison entre elles
- Avantages:
  - Facilité de navigation
  - Nombre de jointures limité
  - Alimentation facile.
- Inconvénients:
  - Redondance dans les dimensions
  - Toutes les dimensions ne concernent pas les mesures

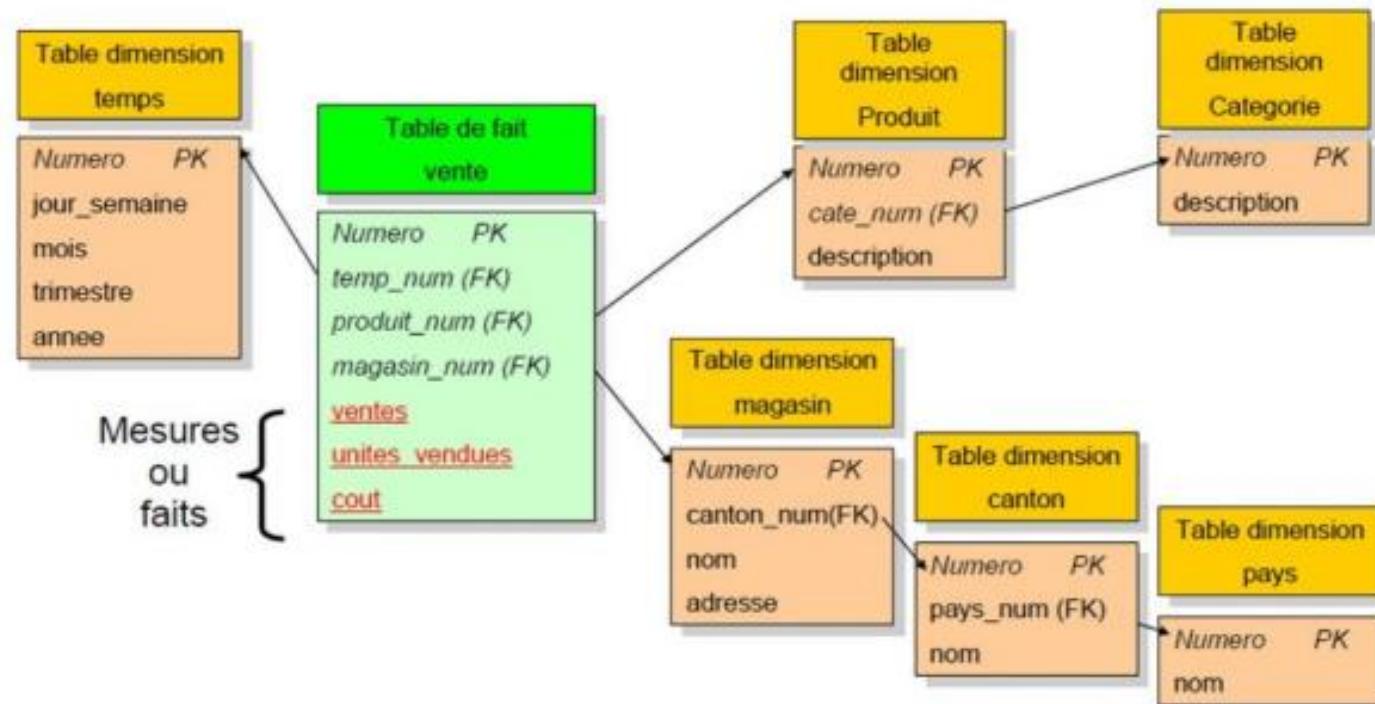
# MODÈLE EN ÉTOILE



# MODÈLE EN FLOCON

- Une table de fait et des dimensions décomposées en sous hiérarchies (Exemple : Commune, Département, Région, Pays, Continent)
- On a un seul niveau hiérarchique dans une table de dimension
- La table de dimension de niveau hiérarchique le plus bas est reliée à la table de fait. On dit qu'elle a la granularité la plus fine
- Avantages:
  - Normalisation des dimensions
  - Économie d'espace disque
- Inconvénients:
  - Modèle plus complexe (jointure)
  - Requêtes moins performantes

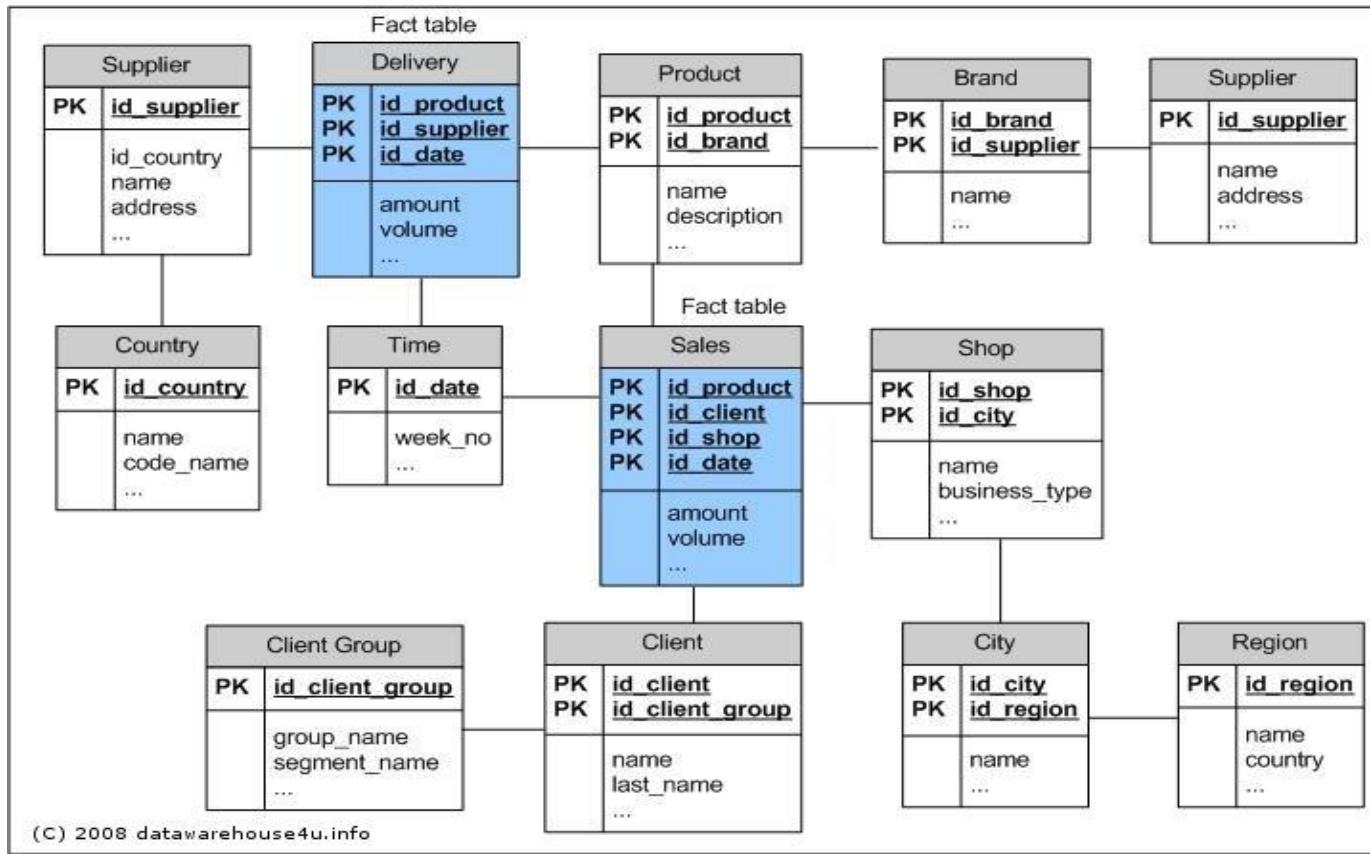
# MODÈLE EN FLOCON



# MODÈLE EN CONSTELLATION

- Fusion de plusieurs modèles en étoile qui utilisent des dimensions communes
- Enregistrement de plusieurs faits avec des dimensions communes ou non
- En général, on a un schéma de constellation de faits pour l'entrepôt
- une étoile de la constellation pour un magasin de données (Data Mart)
- Exemple : Vente de médicaments dans des pharmacies
- une constellation est constituée de 2 schémas en étoile :
  - Schéma en étoile 1 :VENTE effectuées dans les pharmacies
  - Schéma en étoile 2 :analyse des PRESCRIPTION des médecins
- Dimensions Temps et Géographie partagées par les faits PRESCRIPTION et VENTE

# MODÈLE EN CONSTELLATION



# ETL

# PLAN

- Implémentation physique des DW
- Processus d'alimentation d'un DW
- Définition d'un ETL
- Etapes d'alimentation

# IMPLÉMENTATION PHYSIQUE DES DW

Phase I :

- Il faut assurer la migration :
  - des entités vers des tables
  - des relations vers des clés étrangères
  - des attributs vers des colonnes
  - des identifiants uniques vers des clés primaires

# IMPLÉMENTATION PHYSIQUE DES DW

## Phase 2 :

- Il faut créer un ensemble de structures parmi les suivantes :
  - les tablespaces
  - les tables et les tables partitionnées
  - les vues
  - Les contraintes d'intégrités
  - les dimensions, ...
- Et pour améliorer les performances
  - les index et les index partitionnés
  - les vues matérialisées

# ARCHITECTURE FONCTIONNELLE DU DW

- Les données d'un entrepôt se structurent suivant
- un axe synthétique : établissement d'une hiérarchie d'agrégation incluant les données détaillées : les évènements les plus récents
  - les données agrégées : synthèse des données détaillées
  - les données fortement agrégées : synthèse à un niveau supérieur des données agrégées
- un axe historique
  - incluant les données détaillées historiées représentant les évènements passés
  - Stockage des métadonnées : informations concernant les données du DW (provenances, structures, méthodes utilisées pour l'agrégation, ...)

# DICTIONNAIRE ET MÉTADONNÉES

- Dictionnaire contenant des informations (métadonnées) sur :
  - toutes les données du DW
  - chaque étape de la construction du DW
- le passage d'un niveau de données à un autre (exploitation du DW)
- Rôle : définition, fabrication, stockage, accès et présentation des données

# DICTIONNAIRE ET MÉTADONNÉES

Toutes les informations nécessaires pour la construction et l'administration de l'entrepôt

- informations présentes dans l'entrepôt
- données source, données dérivées, dimensions, hiérarchies contraintes d'intégrité schéma de l'entrepôt indexes, partitions requêtes redénies ...
- informations d'administration
- règles de nettoyage, transformation, extraction
- politique de rafraîchissement
- sécurité
- monitoring, statistiques
- traçage des données

# DICTIONNAIRE ET MÉTADONNÉES

- Chaque composant de l'entrepôt fournit des métadonnées
- doit connaître celles des autres composants
- doit savoir où ces métadonnées sont situées
- Une BD est dédiée aux métadonnées

# PROCESSUS D'ALIMENTATION D'UN DW

- Apres avoir conçu le modèle des données, comment alimenter l'entrepôt ?
- Faut-il ramener toutes les données sous le même format ?
- Si oui, quel format choisir et pourquoi ?
- Sinon, comment faire pour interroger toutes ces différentes structures ?
- Quel(s) langage(s) d'interrogation va-t-on utiliser ?
- Quelle architecture utiliser ?
- problématique de l'ETL (Extracting, Transforming and Loading)

# DÉFINITION D'UN ETL

- Offre un environnement de développement
- Offre des outils de gestion des opérations et de maintenance
- Permet de découvrir, analyser et extraire les données à partir de sources hétérogènes
- Permet de nettoyer et standardiser les données
- Permet de charger les données dans un entrepôt

# ETAPES D'ALIMENTATION

4 étapes :

1. Sélection des données sources
  2. Extraction des données
  3. Nettoyage et Transformation
  4. Chargement
- Etapes 1 et 2 : Jusqu'a 80 % du temps de développement d'un entrepôt

# EXTRACTION

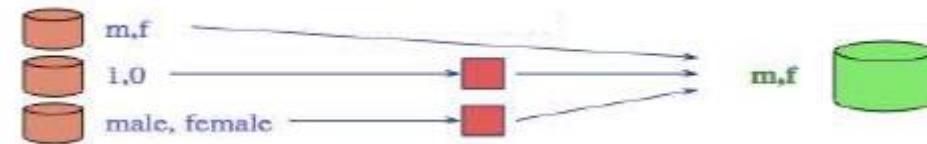
- Extraire des données des systèmes de production
- Dialoguer avec différentes sources:
  - Base de données,
  - Fichiers,
  - Bases propriétaires
- Utilise divers connecteurs :
  - ODBC,
  - SQL natif,
  - Fichiers plats

# EXTRACTION

- Elle ne doit pas perturber les activités de production.
- Il faut faire attention aux données cycliques. Celles qu'on doit calculer à chaque période, pour pouvoir les prendre en considération.
- L'extraction peut se faire en interne selon l'horloge interne ou par un planificateur ou par la détection d'une donnée cible (du DW); ou en externe par des planificateurs externes.
- Les données extraites doivent être marquées par “ *horodatage* ” afin qu’elles puissent être pistées.

# TRANSFORMATION

- Intégration des données
  - Homogénéisation du vocabulaire, structures, valeurs
  - Suppression et fusion des redondances
  - Épuration des données (suppression des données incohérentes)
  - Transformation des données dans un format cible
- Rendre cohérentes les données des différentes sources
  - Transformer, nettoyer, trier, unifier les données
  - Exemple: unifier le format des dates (MM/JJ/AA  $\Rightarrow$  JJ/MM/AA)



# TRANSFORMATION

- Les données doivent être filtrées afin d'éliminer les données aberrantes: données sans valeurs ou avec des valeurs manquantes.
- Souvent dans les bases de production, certaines données sont sémantiquement fausses.
- Savoir corriger les données pour les doter d'un vrai sens sémantique.
- Pour avoir une alimentation de qualité, il faut avoir une bonne connaissance des données à entreposer et des règles qui les régissent.
- On peut dédoubler des données pour gagner au niveau de la cohérence.

# NETTOYAGE ET TRANSFORMATION DES DONNÉES

- Objectifs du nettoyage :
  - Résolution des problèmes de consistance des données au sein de chaque source
- Remarques :
  - une centaine de type d'inconsistances ont été répertoriées
  - 5 à 30 % des données des BD commerciales sont erronées

# TYPES D'INCONSISTANCES

- Présence de données fausses dès leur saisie
- Fautes de frappe
- différents formats dans une même colonne
- texte masquant de l'information (e.g., "N/A")
- valeur nulle
- incompatibilité entre la valeur et la description de la colonne
- duplication d'information
- Persistance de données obsolètes

# CHARGEMENT

- Opérations de calcul et d'agrégation des données
- Mise en place de procédures de chargement (nocturne?) et de restauration (en cas de problème)
  - Envisager la mise en place de systèmes redondants si la disponibilité du système ne peut être interrompue
- Prise en compte de la notion de **granularité**
  - Conservation des données détaillées
  - Possibilité d'agrégation des données pour la synthèse
- Utilisation de connecteurs:
  - ODBC, SQL natif, Fichiers plats

# CHARGEMENT

- Objectif : Stockage des données nettoyées et préparées dans l'ODS ou le SA
- Opération :
  - risquant d'être assez longue
  - plutôt mécanique
  - la moins complexe
- Mais il est nécessaire de définir et mettre en place :
  - des stratégies pour assurer de bonnes conditions à sa réalisation
  - une politique de rafraîchissement

# MAINTENANCE DES DW

- Quand et comment assurer les mises a jour (la maintenance) d'un entrepôt ?
- Quelles anomalies peuvent être causées par la maintenance ?
- A quel niveau pourrait-on automatiser cette maintenance ?
- Comment mesure et assurer la performance et quel critère choisir ?
- La maintenance ou l'auto-maintenance pourra-t-elle à elle seule garantir les performances ?

# MAINTENANCE DES DW : REFRESHING

3 stratégies :

## 1. Reconstruction périodique

- la plus simple
- la plus longue
- elle suppose une longue période d'indisponibilité

## 2. Mise à jour périodique

- volume de données concerné plus petit
- algorithmes plus complexes que pour une reconstruction

## 3. Mise à jour instantanée

- nécessite de nombreuses communications

# PAR RECONSTRUCTION

- Rafraîchissement périodique et de manière incrémentale
- Prise en compte des changements des sources
- Suppression des données anciennes

# DÉTECTION DES CHANGEMENTS

- Dépend des sources
- Triggers utilisés pour déclencher la mise à jour
- Exploitation des logs des changements
- Extraction des changements pertinents par requêtes
- Comparaison de différentes images de la source

# MÉTHODOLOGIE: 9 ÉTAPES DE KIMBALL

1. Choisir le sujet
2. Choisir la granularité des faits
3. Identifier et adapter les dimensions
4. Choisir les faits
5. Stocker les pré-calculs
6. Établir les tables de dimensions
7. Choisir la durée de la base
8. Suivre les dimensions lentement évolutives
9. Décider des requêtes prioritaires, des modes de requêtes