**Modélisation des ED : Concepts de base**

**Modélisation des BD** : Entité et relation.

**Modélisation des DW** : dimension et mesure.

-**les dimensions** : les points de vues depuis lesquels les mesures peuvent etre observées/ une table qui contient les axes d'analyse selon lesquels on veut étudier des données observables.

-**les mesures** : valeurs numériques que l'on compare. C'est le resultat d'une operation d'agrégation des données

-**les propriétes des mésures** :

- Fait additif : additionable suivant toutes les dimensions. ex qantité vendue - chiffre d'affaire.

- Fait semi-additif : additionable selon certain les dimensions. ex niveau de stock - nombre de transactions.

- Fait non-additif : non-additionable, recalculer. ex MxCa pour l'ensemble des magasins.

**-Les faits :**

-**Un fait** represente un sujet d'analyse, la valeur d'une mesure, mesurée ou calculée, selon un membre de chacune des dimensions.

-**Les mesures** sont stockées dans les tables de faits.

-**La table de fait** contient les données observables(les faits)/les mesures, les clés vers les tables de dimensions, les dimensions dégénérées(sans attribut).

-Dans un **entrepot de données** les "Faits" sont normalement numériques puisqu'elle sont d'ordre quantitatif.

-**Caractéristiques d'une table de faits** :

-contient les valeurs numérique de ce qu'on désir.

-contient les clés étrangères.s

-contient un nombre de colonnes réduit.

-contient plus d'enregistrement qu'une table de dimensions

**-Caractéristiques des info de la table de faits** :

-elles sont numériques et sont utilisées pour faire des SUM, AVG

-les données sont additives/semi-additives

-les mesures ds la table doivent référer et avoir un lien direct aux clés de dimension.

-**une dimension peut être définie comme** un thème ou un axe selon lequel les données serons analysées. C’est un axe d’analyse

-**une dimension contient** des membres organisées en hiérarchie. ( temps: année, semestre, mois, jour).

-**structure de base d'une table de dimension** :

=>Clés de substitution(cléprimaire)/Clés d’affaires(clénatuturelle), attributs de la dim , Clés spéciales (pour la gestion de l'historique de la dimension).

------------------------------------------------------------------------------------------------------------------------------------------------

-**Caractéristiques d'une dimension**:

-contient le detail sur les faits.

-contient les info descriptives des valeurs numériques de la table de faits.

-les attributs sont souvent utilisés comme "lignes" et "colonne" dans un rapport ou résultat de requéte.

=>**Clé de substitution** : clé non-intelligente utilisée dans un DW pour remplacer et compléter la clé artificielle du sys opérationnel afin de rendre un elemnet unique dans la dimension.

=>**Clé naturelle** est en generale composée de plusieurs colonnes.

-**Dans un systeme opérationnel on utilise une clé artificielle** afin d'identifier d'une facon unique un element de l'entité.

**=>La clé de substitution ne doit pas etre confondue avec la clé artificielle attribuée par le sys opérationnel.**

**Fonctionnalité des clés de substitution** :

- Remplacer la clé naturelle : la cle de substitution remplace la clé artificielle

- Completer l'information : la clé de substit est utilisée dans l’entrepot de données seulement

**Avantages des clés de substitution :**

-Performance : accélère l’accès aux données du moment ou l’on va utiliser un index numèrique

-Indépendance du sys source : on ne peut garantir que la clé d’affaire ne change pas dans les systèmes sources

-Historique des changement et granularité infinie : gérer la cle de substitution pour garder l’historique des changements de la dimension selon certains critères

-----------------------------------------------------------------------------------------------------------------------------

**Table d'une BD multidimensionnelle :**

**Les dimensions: Clés spéciales:**

-Date effective: date du creation de l'enregistrement.

-Date retrait : date ou l'enregistrement a été retiré.

-Indicateur effectif: "O" si l'enregistrement actif, "N" sinon.

(facultatif) :

-Type de l'evolution (SCD)

-Valeur avant le changement

----------------------------------------------------------------------------------------------------------------------------------------------

**Evolution des dimensions :**

**-Lente** : gestion de la solution

-> Ecrasement de l'ancienne valeur: (correction des info erronées)

-facile a mettre en œuvre (avantage)

-perte de la trace les val antérieures des attributs. (inconvenient)

-perte de la cause de l'evolution ds les faits mesurés. (inconvenient)

-> Versionnement ( ajout d'un nouvel enregistrement, utilisation clé primaire)

-permet de suivre l'evolution des attributs (avantage)

-permet de segmenter la table de faits en fonction de l'historique. (avantage)

-accroit le volume de la table. (inconvenient)

-> Valeur d'origine/courante ( ajout d'un nouvel attribut) :

-avoir deux visions simultanées des données. (avantage)

-voir les données comme si le changement n'avait pas eu lieu (avantage)

-inadapté pour suivre plsr valeurs d'attributs intermédiaires. (inconvenient)

**-Rapide :**

-> Subit des changement très fréquents dont on veut préserver l'historique

-> Solution : isoler les attributs qui changent rapidement.

--------------------------------------------------------------------------------

**Modélisation des ED : Modèle d'un DW**

**-**Modèle en étoile : table de faits au centre du schéma

-Modèle en flocon de neige

-Modèle en constellation/galaxy

Graphical user interface

Description automatically generatedText

Description automatically generated-------------------------------------------------------

A screenshot of a computer

Description automatically generated with medium confidence

--------------------------------------

**Démarche de conception :**

Processus Métier

étape 1 : -choisir le processus à modéliser.

étape 2 : -choisir le grain de faits, données granulaire.

Fait à observer

-décider de ce que represente une ligne de la table de faits

-niveau de detail: transaction individuelles,..

étape 3 : -identifier les dimensions. (typiquement le temps, le client...). Dimensions

étape 4 : -identifier les mesures de faits ( de préférence additives).

Mesures

---------------------------------------------------------------------------------

**Analyse Multidimensionnelles**

Le Data Warehouse est le socle indispensable pour obtenir les réponses aux questions essentielles à la prise de décision et au pilotage de l’entreprise.

Quitter un modèle de base de données opérationnelle et se baser sur un modèle de base de données décisionnelle revient à : transformer les éléments du modèle de base de données opérationnelle en dimensions et fait

**BD Multidimensionnelles : Cube**

-Un cube OLAP est une structure de données multidimensionnelle stockant les faits comme des mesures indexées par plusieurs dimensions.

-Chaque cellule d’un cube représente la mesure ou valeur quantitative d’un fait sur le croisement de plusieurs dimensions.

-L’intérêt d’un cube OLAP est d’offrir à l’utilisateur la capacité de faire des analyses multidimensionnelles ou des agrégations par axe de dimension dans l’espace.

**BD Multidimensionnelles = super-tableur.**

**BD multidimensionnelle, Opérateurs :**

-Les opérateurs appliqués sur un cube sont algébriques.

-Le résultat est un autre cube.

-Les opérateurs permettent :

-Des extractions Slicing : Prendre une tranche du cube.

-Des extractions Dicing : Extraire un sous-cube.

-Des changements de granularité d’une dimension :

– Roll-up (agrégation d’une dimension -> Résumé)

– Drill-down (informations plus détaillées).

---------------------------------------------------------------------------------

**Technologies OLAP**

• **ROLAP** - Relationnel OLAP : OLAP sur du relationnel

• **MOLAP** - Multidimensionnel OLAP : OLAP sur un DW dimensionnel

• **HOLAP** - Hybride OLAP : Mélange des deux

**=>ROLAP** :

-Utilisent un SGBD relationnel classique avec des adaptations spécifiques à l’OLAP

-organisées en schémas en forme d'étoiles ou en flocon de neige.

-Peuvent conduire à des temps de réponses élevés.

-s’appuie sur la maturité de la tech relationnele

-N’effectuent pas des requêtes OLAP avec des calculs complexes.

-Il réduit le nombre de jointures à réaliser pour l’exécution d’une requête.

-lent et peu performant mais sans limites de taille.

-Exemples : IBM DB2 , Oracle, SQL Server, Informix, ...

**=>MOLAP** (OLAP multidimensionnel)

-Utilisent un SGBD multidimensionnel.

-Les données sont stockées directement ds une structure de Cube multidimensionnel dont : les structures sont optimisées + l’ accès est rapide en lecture/ écriture.

-MOLAP necessite le pre-calcul, sur tous les niveaux de hiérarchies des dimensions.

-Très rapide et performant mais avec des limitations de taille

-Exemples : Essbase, SAS OLAP Server, …

**=>HOLAP** (HOLAP permet d’avoir des DW de taille importante tout en ayant des temps de réponse satisfaisants)

-Les données multidimensionnelles sont stockées et traitées en se basant sur le SGBD Relationnel et le SGBD multidimensionnel, afin d'éviter les problèmes des systèmes MOLAP et ROLAP.

---------------------------------------------------------------

-L’approche HOLAP consiste à

• Utiliser les tables comme structure permanente de stockage des données

• Manipuler les informations du DW avec un moteur ROLAP

• Exploiter les Data Marts selon une approche multidimensionnelle avec un

système MOLAP.

-Combinent les points positifs des systèmes ROLAP et MOLAP :

• stocker les données détaillées du DW dans un SGBD Relationnel

• stocker les données agrégées, souvent des magasins de données (Data Marts) du DW dans un SGBD Multidimensionnel

-HOLAP permet d’avoir des DW de taille importante tout en ayant des temps de réponse satisfaisants.

-Exemples : Oracle, IBM DB2 OLAP Server …

---------------------------------------------------------------------------------

**Présentation du langage MDX**

MDX = MultiDimensioal eXpression

langage de requetes OLAP pour les BD multidimensionnelles.

permet de :

- naviguer ds la BD multidim

- interroger et manipuler les données stockées ds des cubes OLAP

- definir des requetes sur tous leurs objets.

-langage complexe et puissant

générer des requetes plus compacte que les requetes SQL

-> select axis1 ON COLUMNS, axis2 ON ROWS, axis3 ON AXIS(0).

from

where (slicer)

Membre = une instance d'un niveau d'une dimension.

Tuple = suite de membres entre () séparés par une virgule.

Cellule = le tuple permet d'identifier les cellules ds un cube.

Mesure = ds un tuple les mesures sont traitées comme une dimension particuliére.

Set = ensemble ordonné de tuples définit sur une meme dimension.(le mot apres ON)

Table

Description automatically generated--------------------------------------------------------------------------------------------------------------------

pour définir un axe et présenter sur l'axe tous les membres d'une dimension :

=> <dimension name>.MEMBERS

pour voir apparaitre tous les membres d'une dimension a un certain niveau :

=> <dimension name><level name>.MEMBERS

Ds un slice on peut avoir plsr membres, mais ils doivent appartenir a des dimension différentes.

// ou -- => commentaire en fin de ligne.

/\* \*/ => commentaire sur plsr lignes

Text

Description automatically generated with medium confidence

Graphical user interface, text

Description automatically generatedText

Description automatically generated