

## Normalisation de tables de faits

- Table de faits d'un schéma dimensionnel est naturellement très normalisé et compacte
- Un nombre très élevé de dimensions:
  - Signe que certaines dimensions ne sont pas complètement indépendantes
  - **Donc il faut les combiner** en une seule dimension.

Représenter les éléments d'une hiérarchie sous forme de dimensions séparées dans la table de faits est contraire aux principes de la modélisation dimensionnelle

53

## Concepts avancés sur les faits et les dimensions

- Additivité des faits
- Grain des tables de faits
- Tables de faits sans faits
- Dimension dégénérée
- Dimension Douteuse
- Grandes Dimensions
- Dimension Fourre-tout
- Evolution des dimensions

54

## Additivité des faits

- Faits additifs.
- Faits semi-additifs : additifs pour certaines dimensions (quantité en stock, solde du compte bancaire)
- Non additifs :
  - fonctions d'agrégations : moyenne, ratios
  - Prix unitaire

55

## Additivité des faits

- **Faits-semi-additifs:**
  - Les niveaux de stocks ne sont pas additifs sur les dates
  - Les niveaux de stocks sont additifs sur les magasins, sur les produits
  - Utiliser la moyenne

Toutes les mesures enregistrant un niveau à un moment donné (niveau de stock, soldes de compte et mesures d'intensité telles que des températures) sont intrinsèquement non additives sur les dimensions dates et éventuellement sur d'autres dimensions.

De telles mesures peuvent être agrégées utilement dans le temps (moyenne des différentes périodes).

**!! Stock moyen pour un groupe de 3 produits dans 4 magasins à 7 dates différentes . SQL AVG divise le total par  $84=4*3*7$  alors qu'il faut faire une division par 7  
Il faut faire le cumul puis diviser par 7**

56

## Grain des tables de faits

- ▶ Trois types fondamentaux de tables de faits:
  - Transactionnel
  - Instantané périodique
  - Instantané récapitulatif
- ▶ Souvent besoin de deux tables de faits complémentaires pour disposer d'une image complète

57

## Grain des tables de faits

### Les tables de faits des instantanés récapitulatifs

- ▶ Comportent généralement des dates représentant les principales étapes du processus.
- ▶ **!!** Une table de faits ayant plusieurs dates n'est pas nécessairement une table de faits instantané récapitulatif.
- ▶ Ce qui distingue la table d'instantané récapitulatif est la modification des lignes existantes par de nouveaux faits le concernant.

Exemple: dans une entreprise fabricant sur des commandes, le client passe une commande → qui entre dans un carnet de commande → jusqu'à sa mise en fabrication → les produits fabriqués sont placés en stock → ils sont expédiés aux clients et facturés (exécution de commande en forme de pipeline)  
dans la table de fait les clés: date de commande, date de mise en carnet de commande, date de mise en fabrication, date de mise en stock produits finis, date de livraison effective, date d'arrivée, date de facture

58

## Grain des tables de faits

Caractéristiques	Grain de transaction	Grain de l'instantané périodique	Grain de l'instantané récapitulatif
Période de temps représentée	Point dans le temps	Intervalle réguliers prévisibles	Durée indéterminée, généralement courte
Grain	Une ligne par événement	Une ligne par période	Une ligne pour la vie de l'entité
Chargement de la table de faits	Insertion	Insertion	Insertion et mise à jour
Mise à jour des lignes de fait	Pas de mise à jour	Pas de mise à jour	Mise à jour pour chaque activité
Dimension date	Date de la transaction	Date de la fin de la période	Dates multiples pour des étapes standard
Faits	Activité de la transaction	Performance dans un intervalle prédéfini	Performance sur une durée de vie limitée

59

## Tables de faits sans faits

- ▶ Tables de faits sans faits intéressantes pour décrire les événements et les couvertures
  - Exemple: événement présence d'un étudiant à un cours: dimensions : étudiant, professeur, cours, date et table des faits: suivi présences (analyse du processus assiduité) quels sont les cours les plus suivi? Quels sont les cours les plus régulièrement présent. Champs présence=1
  - Couverture: ventes promotionnelles par point de vente et par période. La table de fait vente n'enregistre que les produits qui sont vendu en promotion et non tous les produits en promotion que ce soit vendus ou non vendus.
  - **Les tables de faits sans faits constituent la méthode la plus appropriée pour enregistrer des événements lorsqu'il n'existe aucune mesure numérique naturelle associée aux événements. Elles permettent également de garantir la bonne couverture des événements**

60

## Dimension dégénérée

### Dimension dégénérée:

- La dimension correspondante est vide
- Les dimensions dégénérées font souvent partie de la clé primaire de la table de faits
- Les numéros d'identification opérationnels donnent souvent naissance à des dimensions dégénérées (numéros de commande, numéros de facture, numéro de transaction)

clés de dimension qui ne correspondent pas à des tables de dimensions dans des tables de faits dont le grain est le document lui-même ou une ligne du document

61

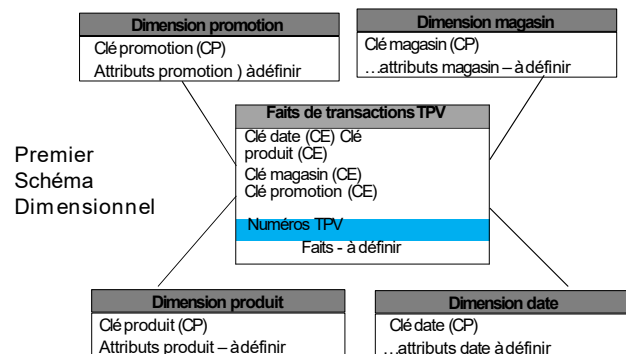
## Dimension dégénérée

### Dimension dégénérée:

- Réservées à des identificateurs de transactions opérationnelles  
(exemple: num de commande, le numéro de commande est utile pour regrouper les articles d'une même commande, dimension article, numéro commande clé dans la table de fait transaction commande client mais pas de création de table commande)

62

## Dimension dégénérée



Quels produits se vendent, dans quels magasins, quels jours, dans quelles conditions de promotions? **Modéliser les ventes au niveau des TPV (terminaux points de vente)**

63

## Grandes Dimensions

### Nombreux membres

### réduire la taille des tables

- dimension Produits (300.000)
- dimension Clients (10.000.000)

### Solutions

- L'appel du Flocon de Neige
  - tables de dimension secondaires (déportées) associées à une table de dimension
  - Faible gain de place et Navigation compromise
- Mini Dimensions
  - Mini dimensions démographiques pour les clients (clé\_zone\_démogh, Clé\_catégorie\_démogh, nom\_catégorie, décompte\_catégorie, pourcent\_catégorie, classement\_catégorie), clé\_zone\_démogh clé étrangère dans dimension client, dimension client 10 000, dimension démographie 100

64

## Dimension Douteuse

### ► Dimension Client Douteuse

- Dimension Client dans laquelle la même personne peut apparaître de nombreuses fois
- orthographes légèrement différentes
- attributs différents

65

## Dimensions fourre-tout

- Présence d'un nombre d'indicateurs pouvant prendre des valeurs restreintes:
  - Laisser les indicateurs tels qu'ils sont dans la table de faits: gonfler la taille de la table
  - Faire de chaque indicateur sa propre dimension: nombre élevé de dimension (de 5 à 25 dimensions)
  - Supprimer tous les indicateurs de la conception : question de l'utilité de ces éléments

Une dimension fourre-tout sert à grouper commodément des indicateurs et des attributs de faible cardinalité tout en les plaçant dans un cadre dimensionnel utile

66

## Evolution des dimensions

### ► Dimensions à évolution lente :

- Les attributs sont relativement stables mais se modifient un peu avec le temps (slowly changing dimensions Ralph Kimball en 1994)
- Il faut spécifier une approche de gestion de changement c-à-d réaction à toute modification dans le monde opérationnel
- Technique 1: écrasement de la valeur précédente
  - Remplacer ancienne valeur par nouvelle valeur dans la table dimensions
  - Simple, rapide
  - Un fait est associé seulement à la valeur courante d'une colonne de dimension
  - Mais ne conserve aucune trace des valeurs antérieures des attributs (exemple : déplacement d'un article d'un rayon à un autre (un jeu du rayon éducatif au rayon jeux stratégie)

Problème: aucune information sur l'effet du déplacement du produit d'un rayon à l'autre sur les ventes de l'article

67

## Evolution des dimension

### ► Dimensions à évolution lente (suite) :

- Type 2: ajout d'une ligne de dimension
  - Permet de suivre correctement les attributs des dimensions à évolution lente
  - Très puissante
  - Ligne de dimension supplémentaire permet de segmenter la table de faits en fonction de l'historique
  - Il faut aussi inclure une date d'effet ou une date d'expiration sur une ligne de chargement de type 2 (attribut date effet/expiration dans la table de dimension rien à voir avec la dimension date)
  - La solution de type 2 impose l'utilisation de clés artificielles déjà recommandé (l'article qui a subi un changement de rayon aura deux lignes dans la table dimension produit)
  - Inconvénient : faire croître le volume de la table

68