

# Business Intelligence & Data Warehouse

Résumé du Chapitre 3 : Conception d'un Data Warehouse

Professeur : Mme HILAL Imane

Année Universitaire 2025-2026 – 3ème année

## 1 La Modélisation Multidimensionnelle

### 1.1 Du modèle relationnel au modèle multidimensionnel

La conception d'un Data Warehouse repose sur la modélisation multidimensionnelle, qui diffère du modèle relationnel Entité/Association traditionnel.

**Passage :** Schéma relationnel (Entité/Relation) → Schéma Multidimensionnel (Fait/Dimension)

### 1.2 Composants de la modélisation multidimensionnelle

#### 1.2.1 Table de Fait

**Définition :** La table de faits contient les mesures et les indicateurs quantitatifs qui permettent le bon suivi des activités et processus.

**Caractéristiques :**

- Collections de mesures associées à un processus métier spécifique
- Les mesures sont stockées dans les colonnes
- Contient les clés étrangères vers les tables de dimension

**Exemple de table de fait :**

Table de Fait Vente
Clé période
Clé client
Clé produit
Clé magasin
Quantité
Chiffre_Affaires

#### 1.2.2 Granularité

**Définition :** Le grain détermine le niveau de détail de la mesure de la table de faits.

**Principe important :**

- Le plus petit niveau de granularité est meilleur pour l'analyse car il permet l'accès à l'information détaillée
- Toutefois, il n'est pas optimal pour le stockage ni la performance des analyses
- Compromis entre détail et performance

### 1.2.3 Table de Dimension

**Définition :** Les tables de dimension représentent les différents axes d'analyses des indicateurs de la table de fait. Elles contiennent des attributs (champs) avec lesquels les mesures sont calculées dans la table de faits.

**Caractéristiques :**

- Regroupent un nombre d'attributs permettant de caractériser la dimension
- S'identifient par la même clé

**Composants d'une dimension :**

1. **Clé non significative** : Établit un lien avec les lignes de la table de faits
2. **Clé significative** : Reprise d'une source de données opérationnelle ou externe
3. **Attributs** : Caractéristiques descriptives de la dimension

**Exemple de table de dimension :**

Dimension Produit	Type
Clé produit	Clé non significative
ID produit	Clé significative
Nom produit	Attribut
Groupe	Attribut
Marque	Attribut
Unité	Attribut
Conditionnement	Attribut

### 1.2.4 Hiérarchie

**Définition :** Les hiérarchies sont représentées par des relations de groupement entre les attributs d'une dimension.

**Exemple :** Date → Semaine → Mois → Trimestre → Année

## 1.3 Slowly Changing Dimension (SCD)

Les **Dimensions à Variation Lente** gèrent les changements dans les attributs des dimensions au fil du temps.

### 1.3.1 Type 1 : Remplacement

- Remplacer les anciennes données par les nouvelles
- **Perte des données historiques**
- Utilisation : Quand l'historique n'est pas important

### 1.3.2 Type 2 : Nouvelle ligne (Approche la plus courante)

- Créer une nouvelle ligne de table de dimension chaque fois que l'objet de dimension change
- Toutes les caractéristiques de la dimension sont conservées au moment du changement
- **Préservation complète de l'historique**

### 1.3.3 Type 3 : Champs multiples

- Pour chaque attribut changeant, créer un champ de valeur courante et plusieurs champs d'ancienne valeur
- Conservation partielle de l'historique (nombre limité de valeurs)

## 2 Le Schéma en Étoile

### 2.1 Définition

Un schéma en étoile contient une seule table centrale, appelée table de faits, entourée de plusieurs tables appelées dimensions.

### 2.2 Structure

**Composants :**

- Une (ou plusieurs) table(s) de faits au centre
- Plusieurs tables de dimension dénormalisées autour
- Les tables de dimension ne sont jamais reliées entre elles

### 2.3 Exemple : Schéma Vente

**Structure simplifiée :**

- **Fait Vente** : Clé période, Clé client, Clé produit, Clé magasin, Quantité, Chiffre d'affaires
- **Dimension Période** : Clé période, ID période, Date, Jour, Semaine
- **Dimension Client** : Clé client, ID client, Nom, Adresse, Sexe, Année naissance, Région
- **Dimension Produit** : Clé produit, ID produit, Nom, Groupe, Marque, Unité, Conditionnement
- **Dimension Magasin** : Clé magasin, ID magasin, Type, Nom, Région

### 2.4 Analyse avec le schéma en étoile

**Exemple de requête :**

- Group by Commune (depuis Dimension Magasin)
- Group by Marque (depuis Dimension Produit)
- Sum(Chiffre d'affaires) depuis la table de Fait

**Résultat :** Total du chiffre d'affaires groupé par commune et par marque

### 2.5 Avantages et Inconvénients

Avantages	Inconvénients
Facilité de navigation	Redondance et duplication des données dans les dimensions
Nombre de jointures limité	Saturation du stockage
Alimentation simple	Performance dégradée des requêtes
Compréhension intuitive	Espace disque important

## 3 Le Schéma en Flocon de Neige

### 3.1 Définition

Le schéma Snowflake (en flocon) est dérivé du schéma en étoile où les tables de dimensions sont normalisées.

### 3.2 Caractéristiques

- La table des faits reste inchangée
- Les dimensions sont décomposées/normalisées selon une hiérarchie
- Élimination de la redondance par la normalisation

### 3.3 Exemple : Transformation du schéma Vente

**Décomposition de la Dimension Produit :**

- Table Produit : Clé produit, ID produit, Nom produit, Unité, Conditionnement, Clé marque
- Table Marque : Clé marque, ID marque, Nom marque

**Décomposition de la Dimension Magasin :**

- Table Magasin : Clé magasin, ID magasin, Type, Nom, Clé commune
- Table Commune : Clé commune, ID commune, Nom commune

### 3.4 Avantages et Inconvénients

Avantages	Inconvénients
Amélioration des performances de requête	Maintenance complexe
Réduction du stockage sur disque	Augmentation des hiérarchies entre dimensions
Réduction de l'assemblage des tables	Complexité des requêtes d'analyse
Élimination de la redondance	Plus de jointures nécessaires

### 3.5 Passage du schéma étoile au flocon

**Formule :**

$$\text{Modèle en flocons de neige} = \text{Modèle en étoile} + \text{Normalisation des dimensions}$$

## 4 La Normalisation

### 4.1 Principe

La normalisation est le processus qui permet d'éviter les données redondantes dans les bases de données. Cela implique de restructurer les tables pour atteindre successivement des formes plus élevées de normalisation.

### 4.2 Hiérarchie des formes normales

$$1\text{NF} \subset 2\text{NF} \subset 3\text{NF}$$

- 1ère Forme Normale (1NF) est considérée comme le plus faible
- 2ème FN est plus forte que 1ère FN
- 3ème FN est plus forte que 2ème FN
- Chaque niveau supérieur est un sous-ensemble du niveau inférieur

### 4.3 Première Forme Normale (1NF)

**Règle :** Une relation est en 1NF si et seulement si tout attribut contient une valeur atomique.

**Exemple :**

- Avant : LIVRE(ID\_Livre, Auteurs)
- Après 1NF : LIVRE(ID\_Livre, Auteur)

**Principe :** Un seul auteur par ligne (valeur atomique)

### 4.4 Deuxième Forme Normale (2NF)

**Règle :** Les attributs non clé dépendent de toute la clé et non d'une partie de la clé.

**Conditions :**

1. La table doit être en 1NF
2. Tous les attributs non clés de la table doivent être fonctionnellement dépendants de la clé primaire entière

**Principe :** Élimination des dépendances partielles

### 4.5 Troisième Forme Normale (3NF)

**Règle :** Chaque attribut de la relation ne dépend que de la clé et pas d'un autre attribut de la relation.

**Conditions :**

1. La table doit être en 2NF
2. Aucun attribut ne dépend de manière transitoire de la clé primaire

**Principe :** Élimination des dépendances transitives

### 4.6 Dimension particulière : Date/Temps

La dimension Date/Temps est une dimension cruciale pour l'analyse.

**Avant normalisation :**

Dimension Date
Clé date
FullDate
DayOfWeek
IsHoliday
Fiscal Week
Calendar Week
Fiscal Month
Calendar Month
...

**Après normalisation (3NF) :**

Hiérarchie décomposée :

- Date → Fiscal Week / Calendar Week
- Week → Fiscal Month / Calendar Month
- Month → Fiscal Quarter / Calendar Quarter
- Quarter → Fiscal Year / Calendar Year

## 5 Le Schéma en Constellation

### 5.1 Définition

Le schéma en constellation fusionne plusieurs modèles en étoile qui utilisent des dimensions communes.

### 5.2 Caractéristiques

- Contient plusieurs tables de faits
- Des dimensions communes ou pas
- Permet de réutiliser les dimensions entre plusieurs faits
- Modélisation plus complexe mais plus flexible

### 5.3 Exemple : Schéma Vente et Livraison

#### Fait Vente :

- Clé période, Clé client, Clé produit, Clé magasin
- Mesures : Quantité, Chiffre d'affaires

#### Fait Livraison :

- Clé période, Clé produit, Clé fournisseur
- Mesures : Quantité, Volume

#### Dimensions partagées :

- Dimension Période (partagée par Vente et Livraison)
- Dimension Produit (partagée par Vente et Livraison)

#### Dimensions spécifiques :

- Dimension Client (pour Vente uniquement)
- Dimension Magasin (pour Vente uniquement)
- Dimension Fournisseur (pour Livraison uniquement)

## 6 Comparaison des Schémas

Critère	Étoile	Flocon	Constellation
Complexité	Simple	Moyenne	Élevée
Redondance	Élevée	Faible	Variable
Performance requêtes	Excellente	Moyenne	Variable
Espace disque	Important	Réduit	Variable
Nombre de jointures	Faible	Élevé	Variable
Facilité de navigation	Excellente	Moyenne	Complexé
Nombre de tables faits	1	1	Multiple