



Modélisation OLAP (Online Analytical Processing)

Ziane.S



Qu'est-ce que l'OLAP ?

L'OLAP est un système utilisé pour l'**analyse multidimensionnelle des données**, principalement dans le but de **soutenir la prise de décision** en entreprise.

Définition du Data Warehouse selon Bill 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. »

 Explication des concepts clés :

 **Orientées sujet**

Les données sont organisées autour des **thèmes majeurs de l'entreprise**

 *Exemple : ventes, clients, produits, temps*

 **Historisées**

Le Data Warehouse **conserve l'historique** des données

 *Suivi de l'évolution des KPIs dans le temps*

 *Exemple : l'évolution des ventes mensuelles sur 5 ans*

Définition du Data Warehouse selon Bill Inmon (1996)

Intégrées:

Les données proviennent de **sources multiples** et sont **unifiées**

→ *Exemple : ERP, CRM, fichiers Excel, bases SQL*

✓ Format standardisé

✓ Référentiel commun (ex. : même code client dans toutes les sources)

Non volatiles:

Les données **ne sont pas modifiées ni supprimées** une fois intégrées

→ *Permet de garantir la stabilité des analyses dans le temps*

Définition du Data Warehouse selon Bill Inmon (1996)

 **But principal : Aide à la décision**

Grâce à ces caractéristiques, le DW permet :

-  **Analyses avancées**
-  **Prévisions**
-  **Exploration de tendances**
-  **Optimisation stratégique**

Base de données transactionnelle VS Entrepôt de données

 OLTP – Online Transaction Processing

 Systèmes transactionnels

-  Gèrent les **transactions courantes** (ex : commandes, paiements)
-  Contiennent des **données en temps réel**
-  Priorité à la **vitesse et à la fiabilité**
-  Utilisé par les **opérations quotidiennes**
-  Données très détaillées et souvent normalisées
- Exemple : systèmes de caisse, gestion des stocks

Base de données transactionnelle VS Entrepôt de données

 OLAP – Online Analytical Processing

 Entrepôt de données (Data Warehouse)

-  Sert à **analyser** les données historiques
-  Supporte l'**aide à la décision**
-  Données **agrégées, historisées** et organisées par thèmes
-  Utilisé par les **analystes, managers, décideurs**
-  Axé sur la **performance analytique**
- Exemple : analyse des tendances de vente, tableaux de bord de performance

Base de données transactionnelle VS Entrepôt de données

Relation entre OLTP et OLAP

 Les systèmes **OLTP** fournissent les **données sources**

 Les systèmes **OLAP** permettent leur **analyse approfondie**

 On peut diviser les systèmes informatiques en deux catégories principales :

✓ **OLTP (opérationnels)** pour les transactions

✓ **OLAP (analytiques)** pour la prise de décision

Modélisation conceptuelle dans les systèmes décisionnels (Data Warehouse)

Deux types de modélisation :

1.  **Modélisation entité-association**
 - Utilisée dans les bases de données traditionnelles (transactionnelles)
 -  Représente les entités (clients, produits...) et leurs relations

2.  **Modélisation dimensionnelle** (*utilisée en Data Warehouse*)
 - Basée sur **faits – mesures – dimensions**

Modélisation conceptuelle dans les systèmes décisionnels (Data Warehouse)

 **Composants de la modélisation dimensionnelle :**

 **Mesures**

- Valeurs **numériques** analysées

 Exemples : **montant des ventes, quantité vendue, revenu brut** (Sont les **indicateurs** de performance)

 **Dimensions**

- Façons d'**observer et regrouper** les mesures
-  **Exemples** : date, produit, localisation, client
- Stockées dans les **tables de dimensions**

Modélisation conceptuelle dans les systèmes décisionnels (Data Warehouse)

Faits

- Contiennent les **valeurs mesurées ou calculées**
- Reliés à plusieurs **dimensions**
- Exemples : chiffre d'affaires pour un **produit**, vendu à une **date**, dans un **magasin** spécifique

Structure typique :

Modèle en **étoile** :

- Table centrale = table de **faits**
- Autour = tables de **dimensions**



Structure du Data Warehouse : Tables de faits & Tables de dimensions



Pourquoi ces deux types de tables ?

Dans un Data Warehouse, la **modélisation dimensionnelle** repose sur deux éléments fondamentaux :



Table de faits (au centre)

- Contient les **indicateurs chiffrés** à analyser
-  Exemples : **montant des ventes, quantité vendue, bénéfice brut**
- Reliée aux tables de dimensions via des **clés étrangères** 
-  Sert de **point central** pour l'analyse



Structure du Data Warehouse : Tables de faits & Tables de dimensions



Tables de dimensions (en périphérie)

- Fournissent le **contexte descriptif** des faits
-  Exemples : nom du produit, catégorie, client, région, date
- Permettent de **filtrer, regrouper, et analyser** les données
- Rendent les analyses **plus lisibles et flexibles**



Structure du Data Warehouse : Tables de faits & Tables de dimensions



Architecture BI classique



Modèle en étoile (Star Schema)

- Une table de faits au centre reliée directement aux dimensions
-  Simple et rapide pour les requêtes



Modèle en flocon de neige (Snowflake Schema)

- Dimensions **normalisées** (plusieurs niveaux)
-  Optimisé pour **intégrité des données** et réutilisation



La Table de Faits dans un Data Warehouse

🎯 Définition et Rôle

La **table de faits** représente les **valeurs mesurées ou calculées** selon différentes dimensions.

- 📌 Située **au centre** du schéma de données
- 📈 Contient les **indicateurs de performance clés (KPI)**
- 📄 Chaque ligne représente une **mesure quantitative** (ex : ventes, coût)
- 🔗 Reliée aux tables de dimensions via une **clé multiple** (concaténation des clés des dimensions)



La Table de Faits dans un Data Warehouse

Type de fait	Description	Exemples
+ Additif	Peut être additionné sur toutes les dimensions	Quantité vendue, chiffre d'affaires
÷ Semi-additif	Peut être additionné selon certaines dimensions	Niveau de stock, solde, nombre de transactions
⊘ Non-additif	Ne peut pas être additionné	Ratio, marge brute

À retenir

La **table de faits** est le cœur analytique du Data Warehouse. Elle permet de mesurer l'activité de l'entreprise et de **suivre les performances** sur différents axes (temps, produit, région...).



La Table de Faits dans un Data Warehouse



Example: Conversion Rate

Let's say you manage two websites

:

Website	Visitors	Purchases	Conversion Rate
A	1,000	100	10%
B	2,000	100	5%

Now, can we say the total conversion rate is:

10% + 5% = 15%? ❌ Wrong!



La Table de Faits dans un Data Warehouse

✓ The correct way:

We need to calculate the total conversion rate **based on actual numbers**:

- **Total visitors = 1,000 + 2,000 = 3,000**
- **Total purchases = 100 + 100 = 200**
- **$\frac{1}{3} \frac{2}{4}$ True conversion rate = 200 / 3,000 = 6.67%**

Why can't we add percentages?

Because **each rate depends on a different base number** (visitors in this case).

Adding them gives a **misleading** result.

- **Non-Additive Facts** like **conversion rates** or **percentages can't be summed**.
- You should **recalculate them** or **use averages** if appropriate.

Let me know if you'd like this in a slide or need more examples!

Schémas de Modélisation d'un Data Warehouse (DW)

Définition

Un **schéma** est une **description logique** de l'ensemble de la base de données.

Dans les bases relationnelles classiques

- Utilisation du **modèle entité-relation (E/R)**
 - ◆ Entités, attributs, relations, cardinalités...

Dans les entrepôts de données (Data Warehouse)

- Utilisation de **modèles multidimensionnels** :
 - ◆ Axés sur l'analyse et la prise de décision
 - ◆ Organisés autour des **faits** et **dimensions**

Schémas de Modélisation d'un Data Warehouse (DW)

★ Types de schémas courants :

1.  **Schéma en Étoile (Star Schema)**
 - Une **table de faits** centrale
 - Reliée directement à des **tables de dimensions** simples
2.  **Schéma en Flocon (Snowflake Schema)**
 - Dimensions **normalisées** en plusieurs niveaux
 - Améliore l'**intégrité des données**
3.  **Schéma en Constellation (Galaxy Schema)**
 - Plusieurs **tables de faits** partagent des dimensions communes
 - Utilisé dans des systèmes complexes (multi-sujets)

 Le choix du schéma dépend de la **complexité des données**, de la **performance requise**, et de la **lisibilité attendue** pour les utilisateurs métier.

★ Modèle en Étoile (Star Schema)

Définition Générale

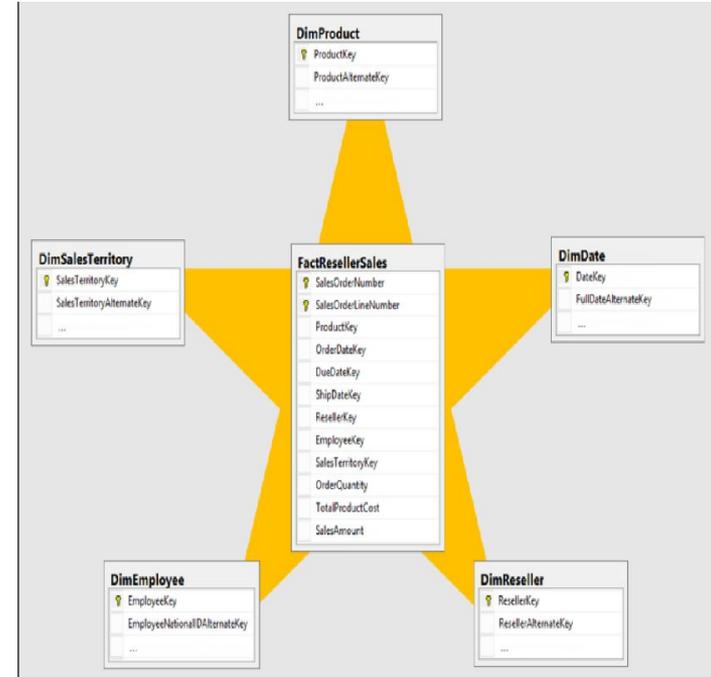
Le **modèle en étoile** est une **approche de modélisation** des bases de données, largement utilisée dans les **entrepôts de données (Data Warehouses)** pour :

- ◆ Simplifier l'analyse
- ◆ Optimiser la prise de décision

Structure du Modèle en Étoile

- ◆ Une **table de faits** centrale ()
 - Contient les **mesures quantitatives** (ex : ventes, revenus)
- ◆ Plusieurs **tables de dimensions** périphériques ()
 - Donnent le **contexte** aux faits (ex : date, client, produit)

 Chaque dimension est **directement connectée** à la table de faits, formant une **structure en étoile**.



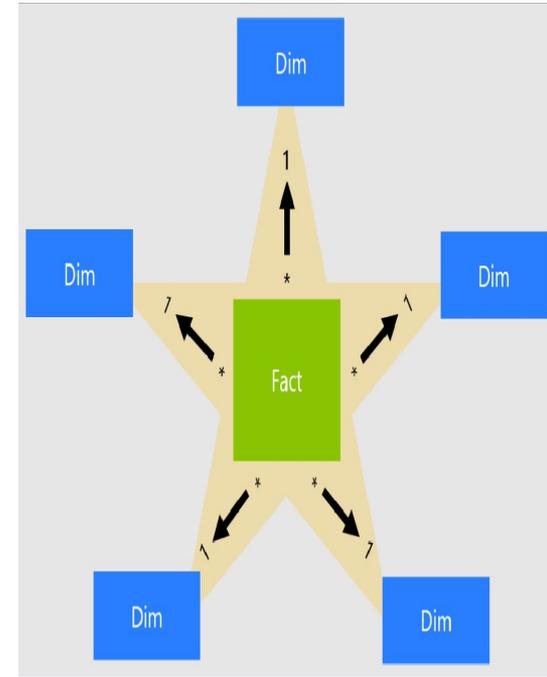
★ Modèle en Étoile (Star Schema)

✓ Avantages

- Lecture rapide des données
- Simplicité de compréhension pour les utilisateurs métier
- Performance améliorée pour les requêtes analytiques

🧠 Exemple

- 📄 Fait : Montant des ventes
- 📅 Dimension Date
- 👤 Dimension Client
- 📦 Dimension Produit





1. Table des Faits (Fact Table)

La **table des faits** est la **table principale** de l'**entrepôt de données (data warehouse)**.

Elle contient les **données quantitatives à analyser** appelées **indicateurs**, ainsi que les **liens vers les dimensions**.

🔑 Clé multiple (clé composite)

Chaque ligne (enregistrement) de la table des faits est **identifiée par une clé composée de plusieurs éléments** :

-  **Date**
-  **Code Produit**
-  **Code Vendeur**
-  **Code Client**



➔ Ces **quatre éléments** forment ensemble la **clé primaire composite** (grisé clair dans le schéma).



1. Table des Faits (Fact Table)



Indicateur mesuré

La variable que l'on veut analyser est appelée **indicateur**, ou **mesure**.

Dans ce cas, c'est :



Montant de la vente



C'est la **valeur numérique** que l'on souhaite **agrég**er (par produit, par jour, par vendeur, etc.).

Cette valeur est généralement en **grisé foncé** dans le schéma.



À retenir :

- La **table des faits** ne contient **pas beaucoup de texte**, mais **des chiffres** et des **clés étrangères**.
- Elle est **liée** à plusieurs **tables de dimensions** (client, produit, vendeur, date).
- Elle est au **centre** du schéma en étoile (**modèle en étoile**).
- Elle permet de **faire des analyses rapides et puissantes** grâce aux outils OLAP (Tableaux de bord, rapports...).

2. Table des Dimensions

2. Les Dimensions

Dans un entrepôt de données, les **dimensions** sont des **tables descriptives** qui apportent **du contexte** aux données chiffrées de la **table des faits**.

Lien entre faits et dimensions :

La **table des faits** contient des **clés de référence** (comme `code_produit`, `code_client`, etc.).

Mais ces codes seuls n'ont **pas de sens** pour l'utilisateur final.

 C'est pourquoi, on les **relie** à des **tables de dimensions** qui donnent plus d'informations.

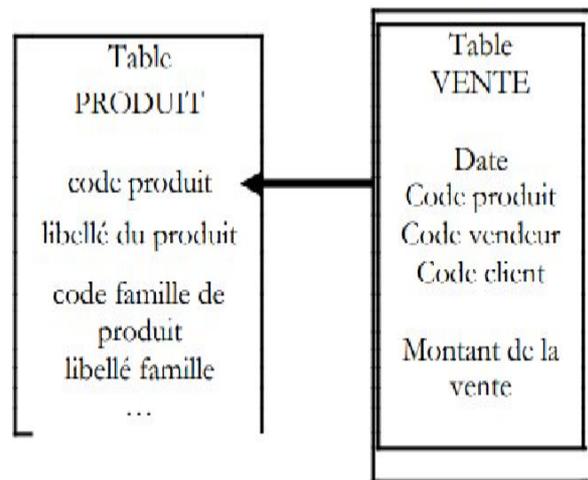


Figure 2 : Table de faits + 1 table de dimension

2. Table des Dimensions

Exemple : Dimension Produit

Le champ `code_produit` dans la table des faits sera relié à une **table de dimension produit** contenant :

Champ	Rôle
<code>code_produit</code>	 Identifiant du produit (clé primaire)
<code>libellé_produit</code>	 Nom du produit
<code>code_famille_produit</code>	 Identifiant de la famille
<code>libellé_famille_produit</code>	 Nom de la famille

2. Table des Dimensions

La jointure (relation)

On crée alors une **relation (jointure)** entre :

-  **table des faits** (par ex. `code_produit`)
et
-  **table de dimension** (avec `code_produit` en clé primaire)

 Ceci permet de **remplacer un simple code opaque** par **des informations utiles et lisibles** dans les analyses

2. Table des Dimensions

Pourquoi les dimensions sont importantes ?

-  Elles permettent de **filtrer, grouper et trier** les données dans les rapports.
-  Elles facilitent les analyses comme :
"Quel est le chiffre d'affaires par **famille de produit** ?"
"Quel produit a généré le plus de ventes ?"
-  Elles définissent les **axes d'analyse** dans un cube OLAP.

2. Table des Dimensions

La table de dimension PRODUIT n'a pas d'indicateur, elle comporte seulement des attributs du produit. La clé « code produit » correspond à l'élément « code produit » de la clé multiple de VENTE.

De la même manière, les autres éléments de la clé multiple renvoient en général chacun à une table de dimension 3, Schéma fait_dimensions le schéma des tables est le suivant :

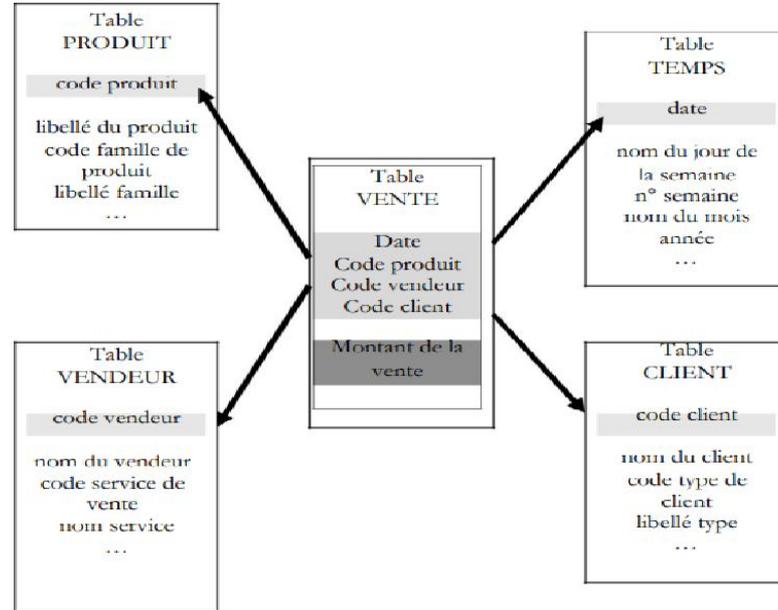


Figure : Table de faits + 4 tables de dimensions

Cube de données (OLAP Cube)

Un schéma en étoile peut également être représenté sous forme de cube de données

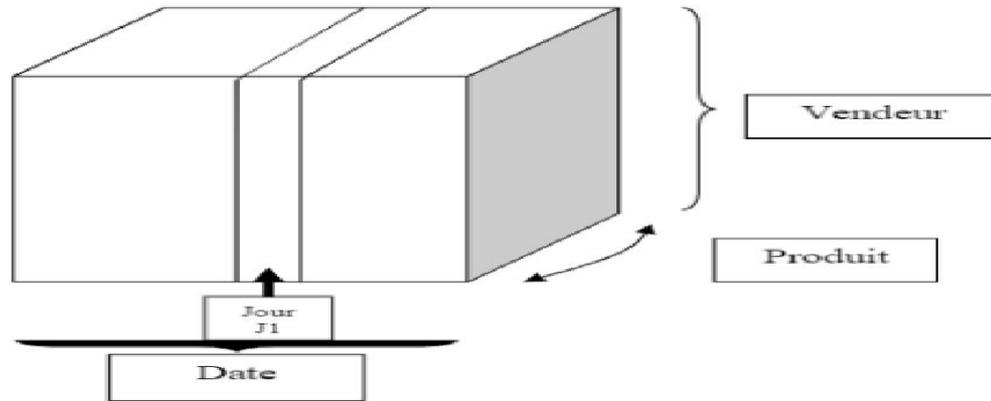


Figure : cube de données à 3 dimensions

Cube de données (OLAP Cube)

Quand parle-t-on de cube de données ?

- Si on a **3 dimensions** → on peut **le représenter graphiquement comme un cube** en 3D.
- Si on a **plus de 3 dimensions** → on parle alors d'un **hypercube** (même s'il est difficile à représenter visuellement).

 **Mais dans la pratique**, on continue à utiliser le mot **cube** pour simplifier, quel que soit le nombre de dimensions.

Cube de données (OLAP Cube)

Exemple traité : Hypercube à 4 dimensions

D'après le **schéma en étoile**, notre projet contient **4 dimensions** :

1.  Produit
2.  Client
3.  Vendeur
4.  Date

 Cela forme un **hypercube à 4 dimensions**.

Cube de données (OLAP Cube)

Comment représenter cela graphiquement ?

Puisqu'on ne peut pas représenter 4 dimensions en 3D sur papier, on **fixe une valeur dans une dimension**, et on dessine un **cube à 3 dimensions** avec les 3 autres.

Voici **4 vues possibles** du cube :

Cube	Dimensions affichées	Dimension "fixée"
A	Client  – Vendeur  – Date 	Produit 
B	Produit  – Vendeur  – Date 	Client 
C	Produit  – Client  – Date 	Vendeur 
D	Produit  – Client  – Vendeur 	Date 

Cube de données (OLAP Cube)

Indicateur dans chaque case du cube

Dans chaque cube, l'**élément de base** est toujours une **valeur de mesure** :

 **Montant de la vente** ()

Chaque **cellule du cube** représente :

 **Le chiffre d'affaires d'un produit, vendu par un vendeur, à un client, à une date donnée** (ou selon la perspective choisie).

Tracé du Cube D (Produit, Client, Vendeur)

Ce cube montre :

- Axe X : **Produit**
- Axe Y : **Client**
- Axe Z : **Vendeur**
- Et pour **chaque valeur de Date**, on a un cube différent.

 Chaque cube correspond à **une date spécifique** (ex : semaine 12 de l'année),
et on peut y lire les montants des ventes selon les produits, clients, et vendeurs.

Cube de données (OLAP Cube)

À retenir :

-  Un **cube OLAP** peut contenir 3 dimensions (représentable visuellement).
-  Un **hypercube** contient plus de 3 dimensions.
-  **On parle toujours de "cube de données"**, peu importe le nombre réel de dimensions.
-  Pour l'analyse, on **projette** des vues 3D du cube en **fixant une dimension**.

⚙️ Opérations OLAP (Online Analytical Processing)

Les opérations OLAP permettent de **manipuler, explorer et analyser** les données stockées dans un **cube de données**.

Elles sont essentielles dans les outils de **Business Intelligence** pour créer des **analyses multidimensionnelles**.

📁 Roll-up

👉 **Regrouper** les données à un niveau plus **général** dans la hiérarchie.

📌 Exemple : Jour → Mois → Trimestre → Année

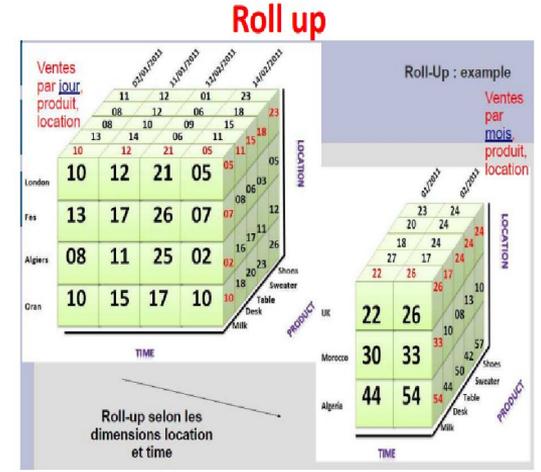
📊 Affiche le total des ventes **par trimestre** au lieu de par jour.

📁 Drill-down

👉 Explorer les données à un niveau **plus détaillé**.

📌 Exemple : Année → Mois → Jour

🔍 Permet de **zoomer** dans les données pour voir les détails.



⚙️ Opérations OLAP (Online Analytical Processing)

🍰 Slice

👉 **Filtrer** le cube sur **une seule valeur** d'une dimension.

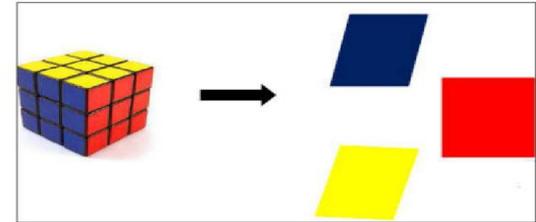
📌 Exemple : Afficher uniquement les données du **mois de janvier**.

🎲 Dice

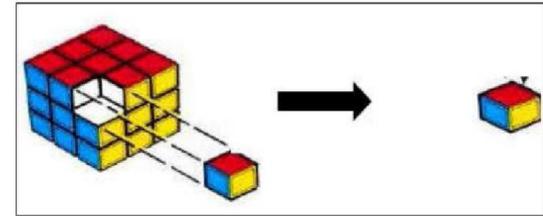
👉 **Filtrer** le cube sur **plusieurs valeurs** ou plages de plusieurs dimensions.

📌 Exemple : Produits = [A, B], Régions = [Nord, Sud], Mois = [janvier, février]

Slice (projection)



Dice (Selection)



🔧 Opérations OLAP (Online Analytical Processing)

🔄 Pivot / Rotate

👉 **Changer l'orientation** des dimensions dans le cube.

📌 Exemple : Mettre "Produit" en lignes au lieu de colonnes.

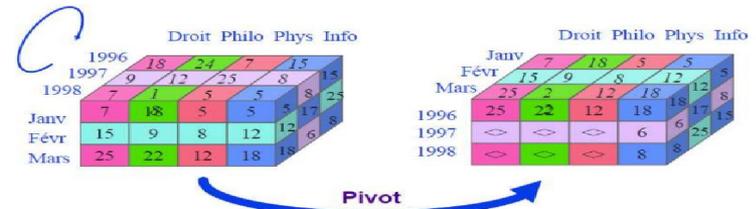
🎯 Aide à **représenter les données sous un nouvel angle**.

🔄 Switch

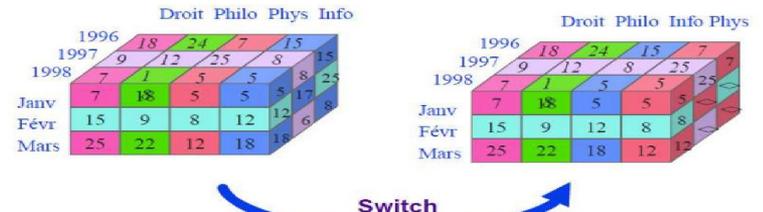
👉 **Échanger deux dimensions** dans le cube.

📌 Exemple : Mettre "Client" à la place de "Produit" dans l'axe principal.

Pivot: Rotation par rapport à l'un des axes de dimensions



Switch: Permutation de valeurs de dimensions



⚙️ Opérations OLAP (Online Analytical Processing)

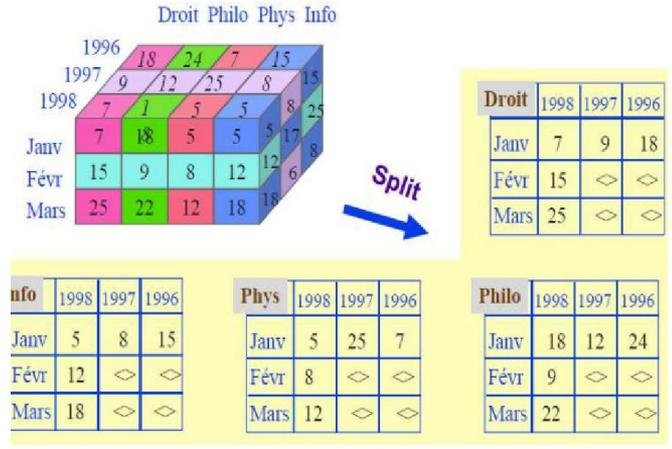
✂️ Split

- 👉 **Diviser** une dimension ou une mesure en **sous-groupes**.
- 📌 Exemple : Séparer les ventes nationales et internationales.

📦 Nest

- 👉 **Imbriquer** une dimension **dans une autre** pour avoir une vue hiérarchique.
- 📌 Exemple : Région → Ville → Magasin

Split (Décomposition)



Opérations OLAP (Online Analytical Processing)

Select

 Extraire les **lignes du cube** qui répondent à une **condition spécifique**.

 Exemple : Ventes > 10 000 €, ou Client = "Entreprise X"

Project

 Ne sélectionner que **certaines colonnes ou dimensions** du cube.

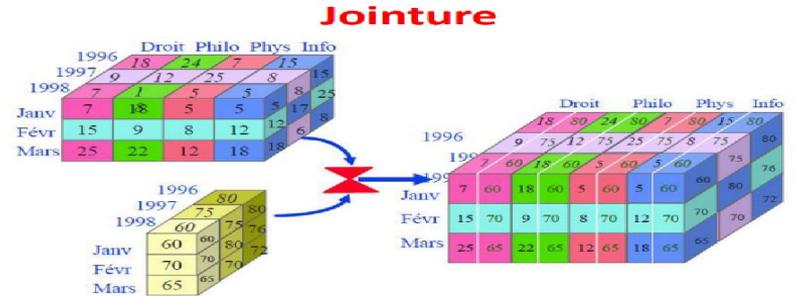
 Exemple : Afficher uniquement la **date** et le **montant des ventes**

🔧 Opérations OLAP (Online Analytical Processing)

🔗 Jointure (Join)

👉 Combiner des données provenant de **plusieurs tables ou cubes**, selon des clés communes.

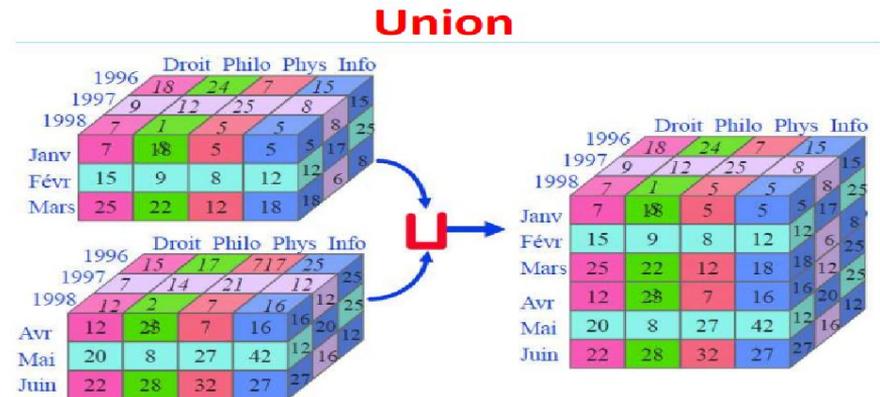
📌 Exemple : Joindre les données de vente avec les données de stock.



⚙️ Opérations ensemblistes

👉 Appliquer des **opérations mathématiques d'ensemble** :

- **Union** : toutes les valeurs des deux jeux de données
- **Intersection** : valeurs communes aux deux jeux
- **Différence** : valeurs présentes dans un jeu mais pas dans l'autre



Thank you