

## **Partie1 : Introduction Générale Bi**

### **C'est quoi l'Informatique décisionnelle ?**

La Business Intelligence (BI) désigne les moyens, outils et méthodes qui permettent de collecter, consolider, modéliser et restituer les données, matérielles ou immatérielles, d'une entreprise en vue d'offrir une aide à la décision et de permettre à un décideur d'avoir une vue d'ensemble de l'activité traitée.

### **C'est quoi L'Intérêt de l'informatique décisionnelle ?**

Les entreprises utilisent une myriade d'applications, et d'outils pour gérer leur activité quotidienne.

Si chacune de ces applications permet de stocker, analyser, ou modifier certains types de données, ces dernières ne sont pas nécessairement compatibles entre elles.

Plus encore, chaque service, équipe ou département peut utiliser un panel d'applications, parfois différents des autres entités de l'entreprise.

La volumétrie de données collectées ou créées, l'absence d'uniformisation et la multiplicité des applications utilisées rendent difficile l'exploitation et l'analyse globale des données par les décisionnaires de l'entreprise, et c'est là que la BI intervient.

### **Les besoins ?**

**1<sup>er</sup> besoin : Historisation des données**

**2<sup>ème</sup> besoin : Centralisation des données**

**3<sup>ème</sup> besoin : analyse des données**

Pour prendre de « bonnes décisions », on doit pouvoir :

Accéder en temps réel aux données de l'entreprise,

Traiter ces données, Extraire l'information pertinente de ces données, par exemple pour savoir :

1. Quels sont les résultats des ventes par gamme de produits et par région de l'année dernière ? !

2. Quelle est l'évolution des chiffres d'affaires par type de magasin et par période ?
3. Comment qualifier les acheteurs de mon produit X ?

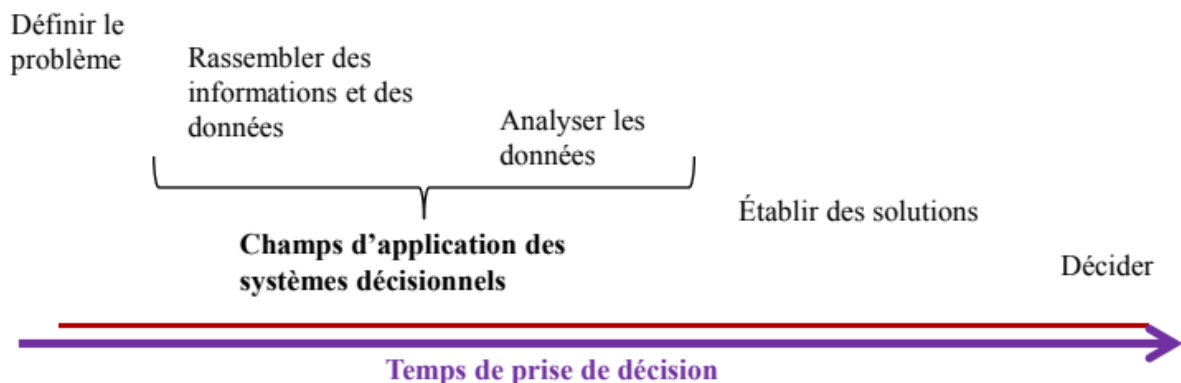
## Processus de prise de décision

### Définition Décision

Une décision est le résultat d'un processus comportant le choix entre plusieurs solutions en vue d'atteindre un objectif.

### Définition information

l'information est une image des objets et des faits ; elle les représente, elle corrige ou confirme l'idée qu'on se faisait. L'information contient une valeur de surprise, en ce sens qu'elle apporte une connaissance que le destinataire ne possédait pas ou qu'il ne pouvait pas prévoir» (Davis, Olson, Ajenstat et Peaucelle p. 116).



## Quels sont les objectifs d'un système décisionnel ?

- Transformer un système d'information qui avait une vocation de production en un système d'information décisionnel.
  - ➔ Transformation des données de production en informations stratégiques.
- Les données doivent être :
  - Extraites
  - Groupées et organisées
  - Corrélées
  - Transformées (résumé, agrégation)

## **Définition : Système d'aide à la décision**

- C'est un système qui utilise les données transitant par un système d'information, données de production le plus souvent, en informations susceptibles d'être exploitées à des fins décisionnelles.
- Il se compose d'une famille d'outils informatiques assurant le fonctionnement de la chaîne de traitement de l'information, se comportant essentiellement de quatre phases :

### **1. Alimentation**

La phase d'alimentation consiste à détecter, sélectionner, extraire, transformer et charger dans un data warehouse (DWH, entrepôt de données) l'ensemble des données brutes issues des différentes sources de stockage de l'information (bases de données, fichiers plats, applications métier, etc.).

Cette phase est généralement réalisée grâce à un outil d'ETL (Extract, Transform, Load). Grâce à des connecteurs, l'ETL peut extraire un grand nombre de données de différents types, puis grâce à des transformateurs manipuler ces données pour les agréger et les rendre cohérentes entre elles. Nous verrons plus loin l'ensemble des fonctionnalités et des avantages des ETL.

### **2. Modélisation**

Une fois les données centralisées, la phase de modélisation consiste à stocker et structurer les données dans un espace unifié (le data warehouse) pour qu'elles soient disponibles pour un usage décisionnel. Cette phase est également réalisée grâce aux outils d'ETL via des connecteurs qui permettent l'écriture dans le data warehouse.

Les données peuvent à nouveau être filtrées et transformées pour assurer la cohérence de l'ensemble dans le data warehouse. Enfin, lors de cette phase les données stockées peuvent être prétraitées via des calculs ou des agrégations pour faciliter leur accès aux outils d'analyse.

### **3. Restitution**

La phase de restitution vise à mettre les données à la disposition des utilisateurs en prenant en compte leur profil et leur besoin métier. L'accès direct au data warehouse n'est pas autorisé puisque l'objectif est de segmenter et de diffuser les données collectées pour qu'elles soient cohérentes par rapport au profil de l'utilisateur et qu'elles soient simples à exploiter.

Lors de cette phase, de nouveaux calculs de données peuvent être effectués pour répondre aux besoins spécifiques des utilisateurs. Les outils de la phase de restitution sont multiples. Il peut s'agir d'outils de reporting, de portails d'accès à des tableaux de bords, d'outils de navigation dans des cubes OPAL (ou hypercubes) ou encore des outils de statistique

#### **4. Analyse**

Dans la phase d'analyse, les utilisateurs finaux vont analyser les informations qui leur sont fournies. Habituellement, les données sont modélisées par des représentations basées sur des requêtes pour construire des tableaux de bord ou des rapports via des outils d'analyse décisionnelle (Power BI, Tableau, Qlikview, etc.).

L'objectif de cette phase est d'assister au mieux l'utilisateur pour qu'il puisse analyser les informations mises à sa disposition et prendre des décisions. Cela passe notamment par le contrôle d'accès aux rapports, la prise en charge des requêtes et la visualisation des résultats.

#### **Opérationnel VS décisionnel ?**

Les systèmes informatiques peuvent se subdiviser en deux :

- Le système transactionnel OLTP (On-Line Transaction Processing)
- Le système analytique OLAP. (On-Line Analytical Processing)

#### **Les systèmes « opérationnels » ou « de gestion », également appelés systèmes OLTP (*on-line transaction processing*)**

Sont dédiés aux métiers de l'entreprise pour les assister dans leurs tâches de gestion quotidiennes et donc directement opérationnels.

La tendance est à l'utilisation de P.G.I. (progiciels de gestion intégrée) qui regroupent tous les logiciels de gestion de l'entreprise : finances, ressources humaines, logistique, ventes, etc. en un unique progiciel paramétrable aux règles de l'entreprise, organisé autour d'une base de données, réduisant ainsi les coûts de communications entre applications.

### **Les systèmes « décisionnels », également appelés OLAP (on-line analytical processing)**

Sont dédiés au management de l'entreprise pour l'aider au pilotage de l'activité, et donc indirectement opérationnels.

Ils offrent au décideur une vision transversale de l'entreprise.

La tendance pour réaliser un système décisionnel est à la mise en place **d'un entrepôt de données**.

Bien que les systèmes d'informations OLTP et OLAP aient le point commun de regrouper les données de l'entreprise dans un S.G.B.D. (système de gestion de bases de données) et d'en fournir l'accès aux utilisateurs, ils présentent de profondes différences, présentées ci-dessous.

- Les systèmes OLTP servent, en général, de source de données pour les systèmes OLAP qui sont quant à eux, source d'analyse des données qui vont permettre d'aboutir à la décision.
- Vous comprendrez alors qu'OLTP et OLAP ont des objectifs opposés et ont un stockage de données différent faisant l'objet de requêtes différentes.

## Opérationnel VS décisionnel

### Opérationnel

- Gère les données de l'entreprise
- Processus de mise à jour transactionnel
- Modèle relationnel
- Données normalisées
- Beaucoup d'utilisateurs
- Non historiées (MAJ)
- Beaucoup de tables

### Décisionnel

- Analyse les données de l'entreprise
- Processus d'analyse des données
- Autre modèle
- Données dé normalisées
- Peu d'utilisateur
- Le temps est fondamental
- Peu de tables mais à grande taille

## Quelques solutions BI

- **Microsoft BI**
- **OBIEE**
- **IBM cognos**
- **SAP BW**
- **QlikView**



- **Pentaho**
- **Talend**
- **Spago BI**
- **JasperSoft**

**Open source ☺**

## **Partie2 : Entrepôt de données (Datawarehouse)**

### **C'est quoi un Entrepôt de données (Datawarehouse) ?**

Un entrepôt de données (ou datawarehouse en anglais) est une base de données qui contient un ensemble de données utilisés dans le contexte de l'analyse décisionnelle et la prise de décision.

C'est une représentation centralisée de toutes les données d'une entreprise.

Mode de travail : OLAP On-Line Analytical Processing

### **Quels sont ses principaux objectifs ?**

Les objectifs principaux sont :

- 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.

### **Quels sont les Caractéristiques d'un entrepôt de données ?**

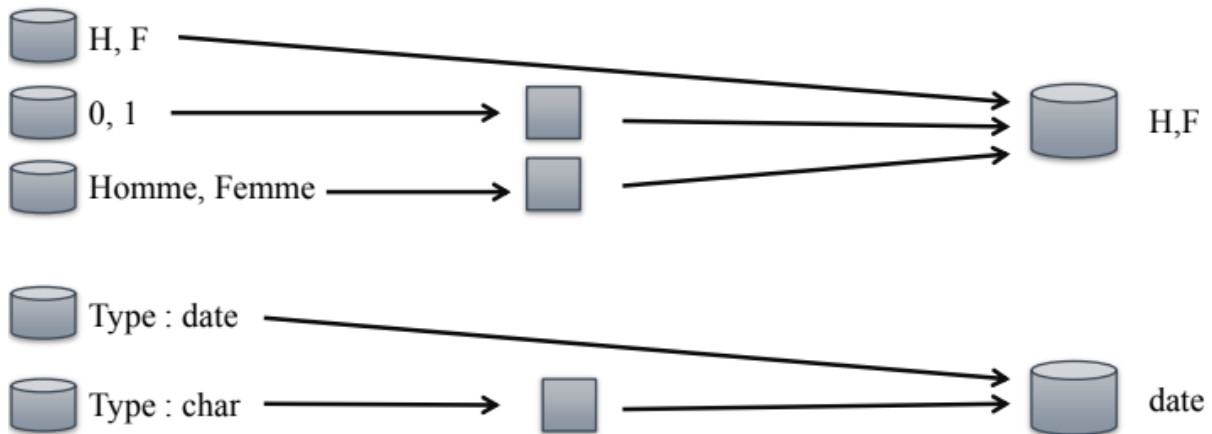
- **Orienté sujet**

Les données sont organisées par thème afin de pouvoir réaliser des analyses sur les sujets et analyser le processus dans le temps

- **Intégré**

Les données proviennent de plusieurs sources différentes. Avant d'être intégrées au sein du datawarehouse elles doivent être mise en forme et unifiées afin d'en assurer la cohérence. Cela nécessite une forte normalisation.

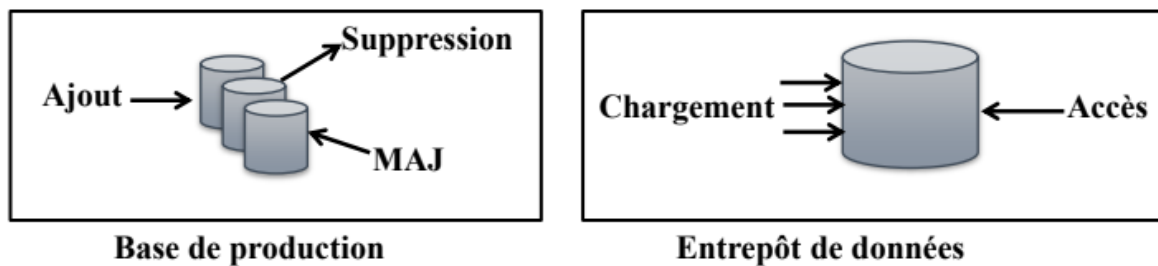
- **Intégré** : Définition d'un référentiel unique.



- **Non volatile**

Un datawarehouse doit conserver la traçabilité des informations et des décisions prises. Les données ne sont ni modifiées ni supprimées. Une requête émise sur les mêmes données à plusieurs mois d'intervalles doit donner le même résultat.

Traçabilité des informations et des décisions prises.



- **Historisé**

L'historisation est nécessaire pour suivre dans le temps l'évolution des différentes valeurs des indicateurs à analyser.

### Pourquoi ne pas réutiliser les SGBD ?

- Les SGBD sont des systèmes dont le mode de travail est transactionnel (OLTP On-Line Transaction Processing).
- Permet d'insérer, modifier, interroger des informations rapidement, efficacement, en sécurité.



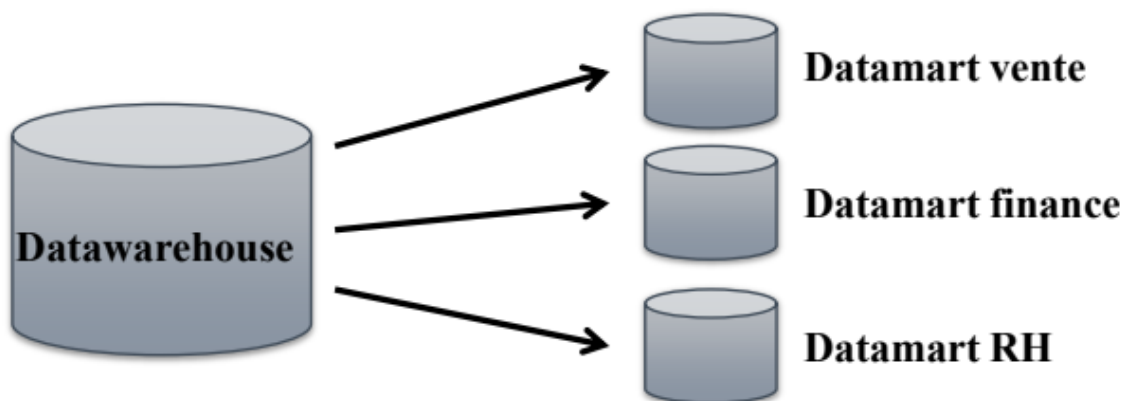
- Deux objectifs principaux :
    - Sélectionner, ajouter, mettre à jour et supprimer des tuples.
    - Ces opérations doivent pouvoir être effectuées très rapidement, et par de nombreux utilisateurs simultanément.
- ➔ En conclusion, Les systèmes OLTP sont mal adaptés à l'analyse de données.

### Quels sont les magasins de données (Datamart) ?

Un magasin de données (ou datamart en anglais) est un sous-ensemble d'un entrepôt de données.

Il contient des données ciblées et regroupées pour répondre à des besoins d'un secteur particulier de l'entreprise.

Il est plus facile à comprendre, à manipuler ainsi le temps de réponse est plus réduit par rapport à un entrepôt de donnée.



### Entrepôt de données et Magasin de données

#### Les entrepôts de données (DataWarehouse):

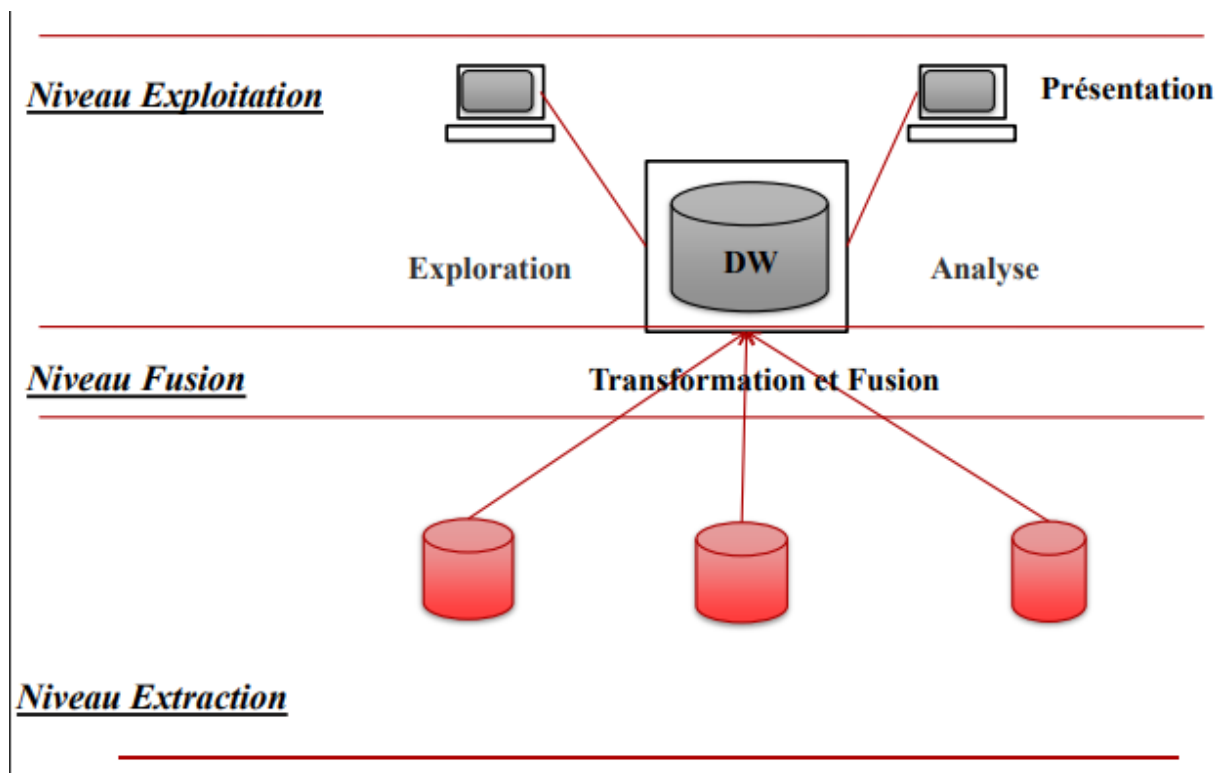
- Nécessitent de puissantes machines pour gérer de très grandes bases de données contenant des données de détail historiques
- Est le lieu de stockage centralisé d'un extrait des bases de production.

- 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

#### **Architecture fonctionnelle d'un entrepôt**



- Niveau extraction

Extraction de données des BD opérationnelles (SGBD traditionnel en OLTP) et de l'extérieur :

- Approche « push » : détection instantanée des mises à jour sur les BD opérationnelles pour intégration dans l'ED
- Approche « pull » : détection périodique des mises à jour des BD opérationnelles pour intégration dans l'ED
  - Niveau fusion

Intégration, chargement et stockage des données dans la BD entrepôt organisée par sujets, Rafraîchissement au fur et à mesure des mises à jour

- Niveau exploitation

Rapports, tableaux de bords, visualisation graphiques diverses.

Analyse et l'exploration des données entreposées (OLAP), Requêtes complexes pour analyse de tendance, extrapolation, découverte de connaissance, ... (Fouille de données)

### Modélisation d'un entrepôt de données

- Un fait :

C'est une ligne, dans une table de faits

Représente la valeur d'une mesure, mesurée ou calculée, selon un membre de chacune des dimensions

Table de Faits
Il s'est passé quelque chose
Il s'est passé autre chose
Il s'est passé quelque chose d'autre

- Les dimensions donnent le contexte du fait

Table de Faits		
Quand	Où	
Hier	Ici	Il s'est passé quelque chose
Hier	Là bas	Il s'est passé autre chose
Aujourd'hui	Ici	Il s'est passé quelque chose d'autre

- Les mesures :

Est un élément de donnée sur lequel portent les analyses, en fonction des différentes dimensions

Donnent les valeurs numériques du fait

Exemple : coût des travaux, nombre d'accidents, ventes

Table de Faits : Ventes				
Date	Magasin	Produit	Combien	Prix Unitaire
Hier	AG	Jouet	3	10
Hier	BG	Saucisson	2	2,5
Aujourd'hui	AG	Parapluie	5	5

### Table de fait

C'est la table centrale du modèle dimensionnel, elle contient les valeurs numériques de ce qu'on désire mesurer.

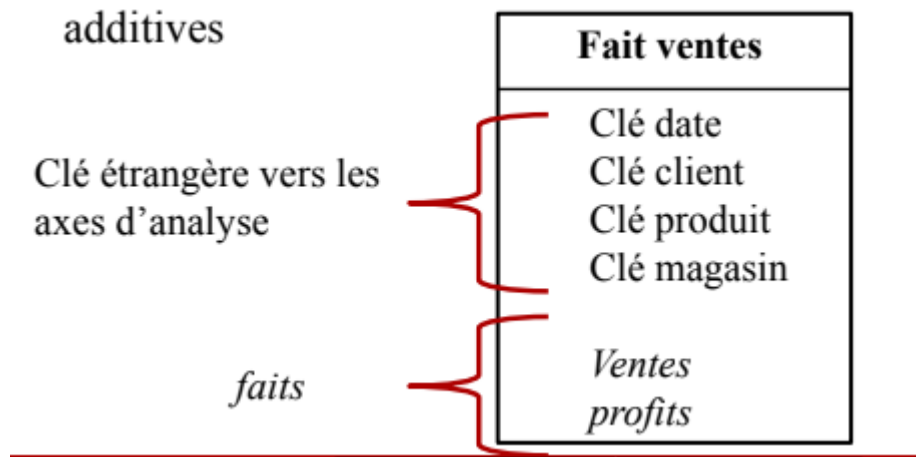
Elle contient l'information à analyser (par exemple les ventes)

Les informations dans une table de fait sont caractérisées par :

- Elles sont numériques afin d'appliquer des fonctions multi lignes

(SUM, COUNT, AVG ...).

- Les données doivent être additives ou semi-additives ou non additives



#### Type de fait :

- Additif : additionnable suivant toutes les dimensions

☐ Chiffre d'affaire, bénéfice ...

- Semi additif : additionnable suivant certaines dimensions

☐ Solde d'un compte bancaire : pas de sens d'additionner suivant l'axe date.

- Non additif : non additionnable quel que soit la dimension

☐ Prix unitaire : l'addition sur n'importe quelle dimension n'a pas de sens.

#### **Table de dimension**

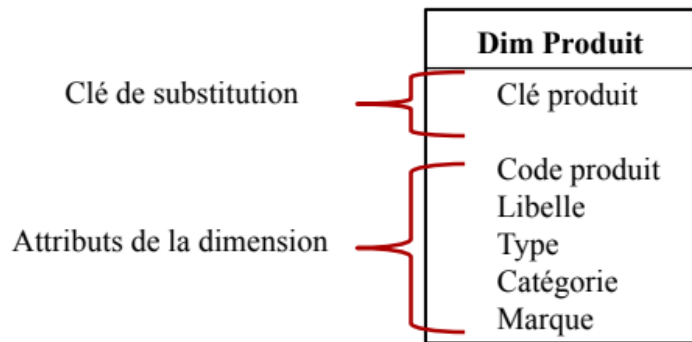
C'est la table qui représente les axes d'analyse, elle contient les détails sur Fait

Elle contient les informations sur les dimensions d'analyse (par exemple le lieu, le temps, la description du produit).

☐ Dimension = axes d'analyse

Une dimension contient des membres organisés en hiérarchie :

- Chacun des membres appartient à un niveau hiérarchique (ou niveau de granularité) particulier
- Ex : pour la dimension Temps : année – semestre – mois – jour



### Dimension Temps

- Commune à l'ensemble des tables de fait.
- Très importante pour la traçabilité des informations

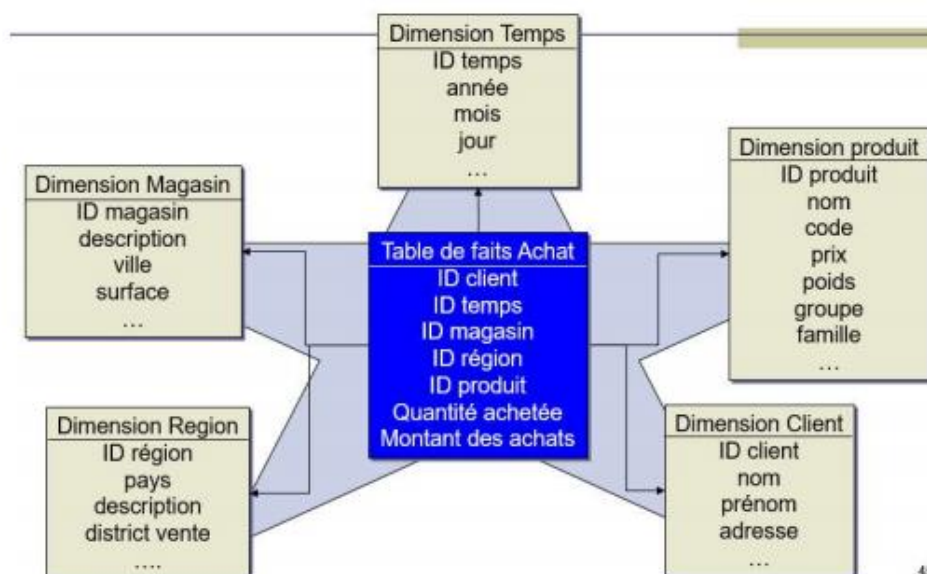
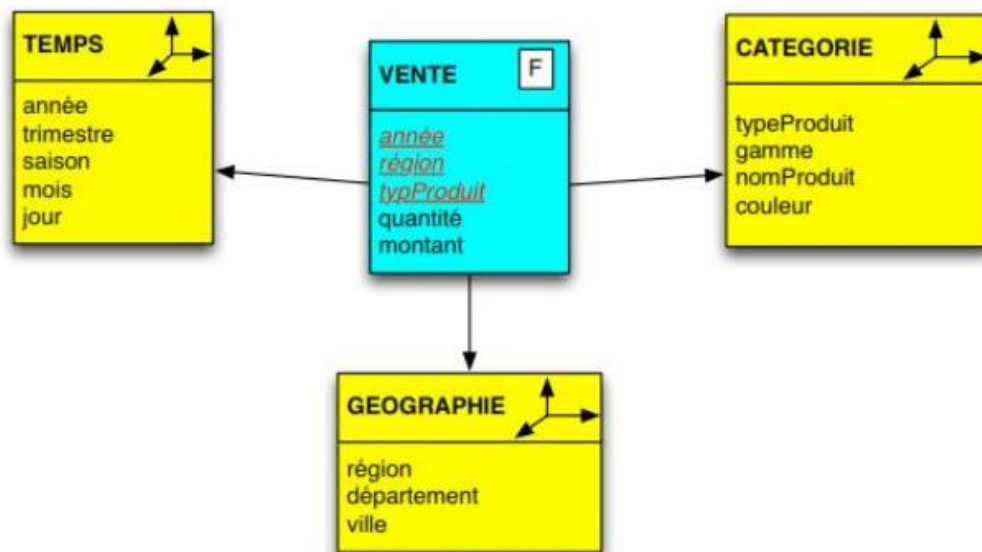
Dim Temps
Clé temps
Jour
Mois
Année
Trimestre
Semestre
...
...

### **Types de Modélisation :**

Il existe deux schémas possibles pour la modélisation de l'entrepôt :

### Le Modèle en étoile

- Une (ou plusieurs) table(s) de faits comprenant une (ou plusieurs) mesures.
- Plusieurs tables de dimension dénormalisées
- Les tables de dimension n'ont pas de lien entre elles

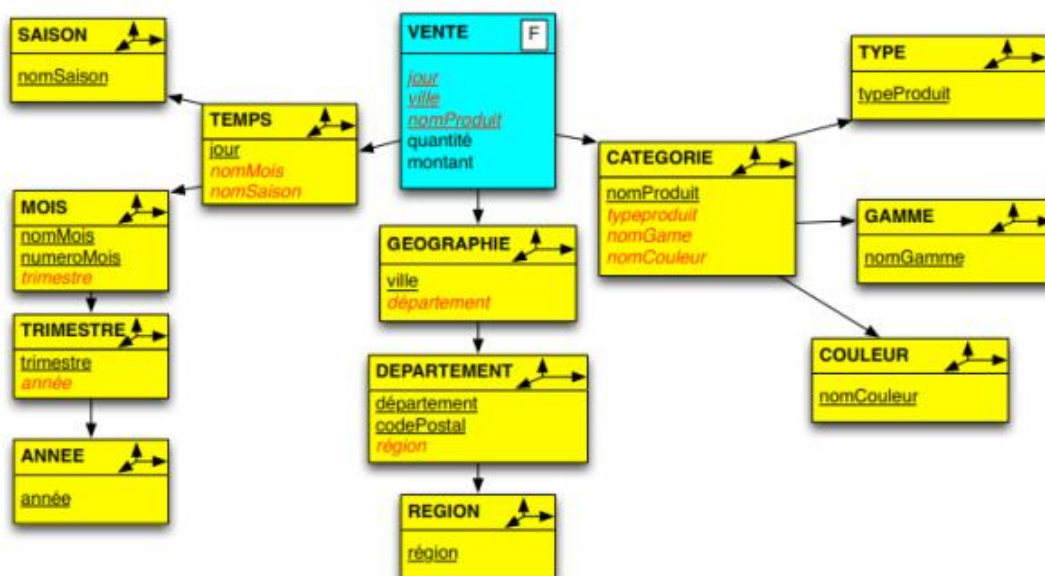


Avantages :

- Facilité de navigation
- Alimentation facile.
- Performances : nombre de jointures limité ; gestion des données creuses.
- Gestion des agrégats

Inconvénients :

- Toutes les dimensions ne concernent pas les mesures
- Redondances dans les dimensions

**Modèle en flocon (Snowflake Schema)**

Modèle en flocon = modèle en étoile + normalisation des dimensions.

Avantages :

- Formaliser une hiérarchie au sein d'une dimension
- Maintenance des tables de dimension simplifiée
- Réduction de la redondance

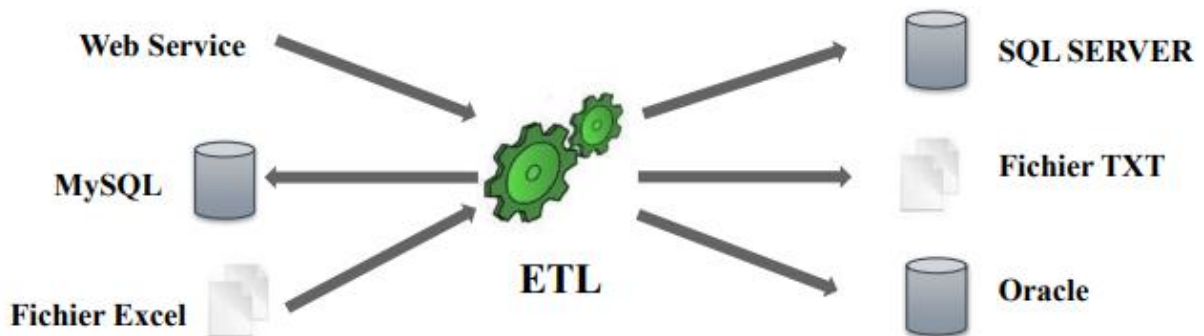


Inconvénients :

- Navigation coûteuses
- Nombreuses jointures

**Partie3 : Notions BI****ETL : Extract-Transform-Load**

Processus permettant de collecter des données en provenance de sources multiples pour ensuite les convertir dans un format adapté à une Data Warehouse et les y transférer.

**ODS (Operational Data Store) et SA (Staging Area)**

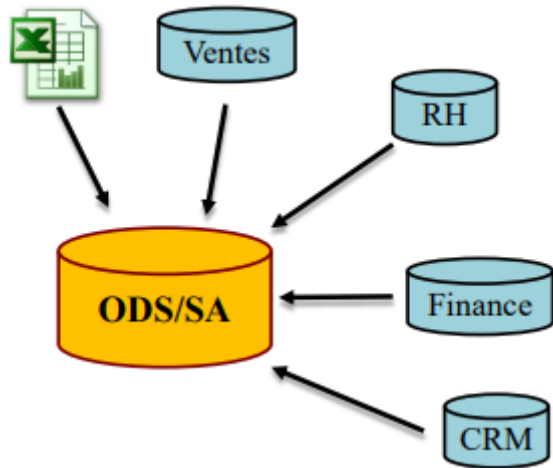
C'est une base de données conçue pour centraliser les données issues de sources hétérogènes. Le modèle d'une base des données (ODS ou SA) est un modèle relationnel classique identique au modèle du système de production.

Leurs différences :Staging Area :

Les données sont détruites directement après avoir été chargées dans le DataWarehouse

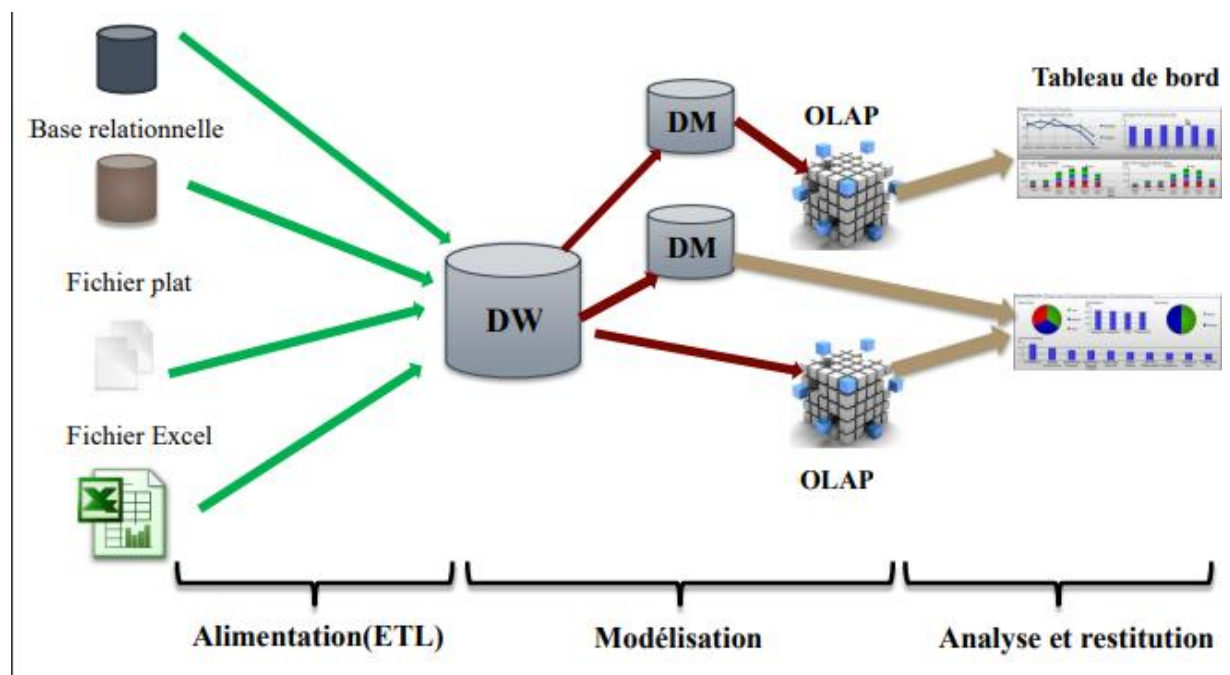
ODS :

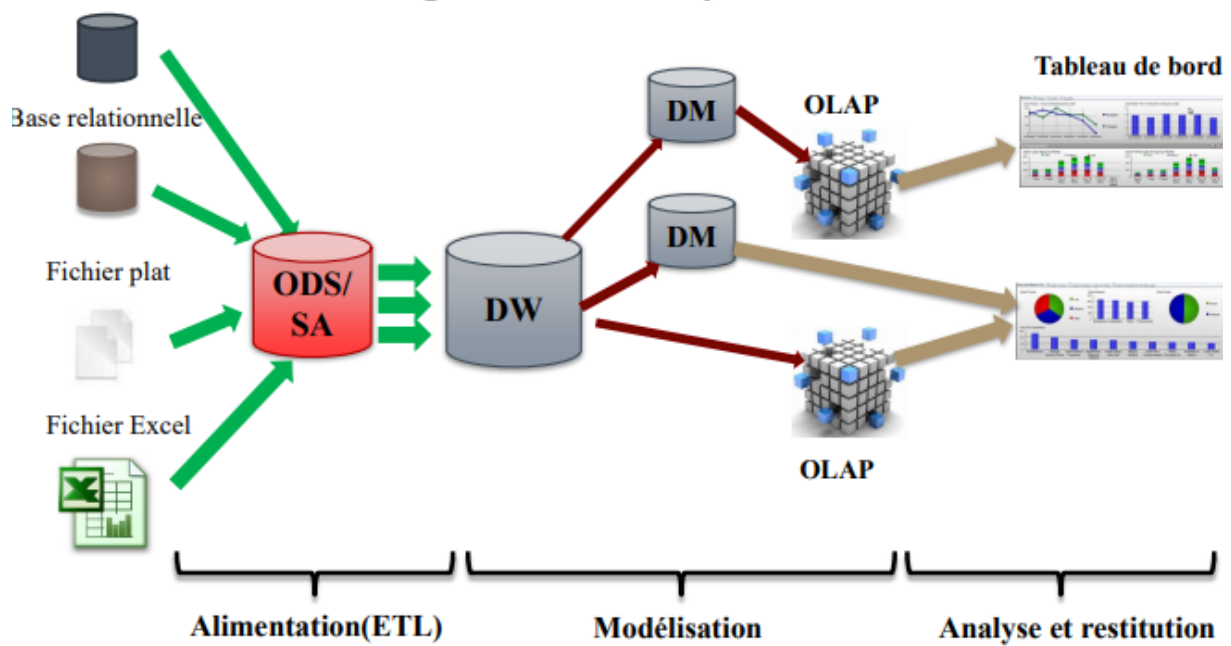
Les données auront quand même une durée de vie plus longue.



L'ODS répond plus à une problématique de reporting immédiat dans le sens où l'ODS sera mis à jour plus souvent que le datawarehouse : on pourrait dire que l'ODS pourrait être alimenté toutes les semaines et le Datawarehouse une fois par mois.

### Architecture globale d'un système décisionnel





#### Partie4 : Cube OLAP

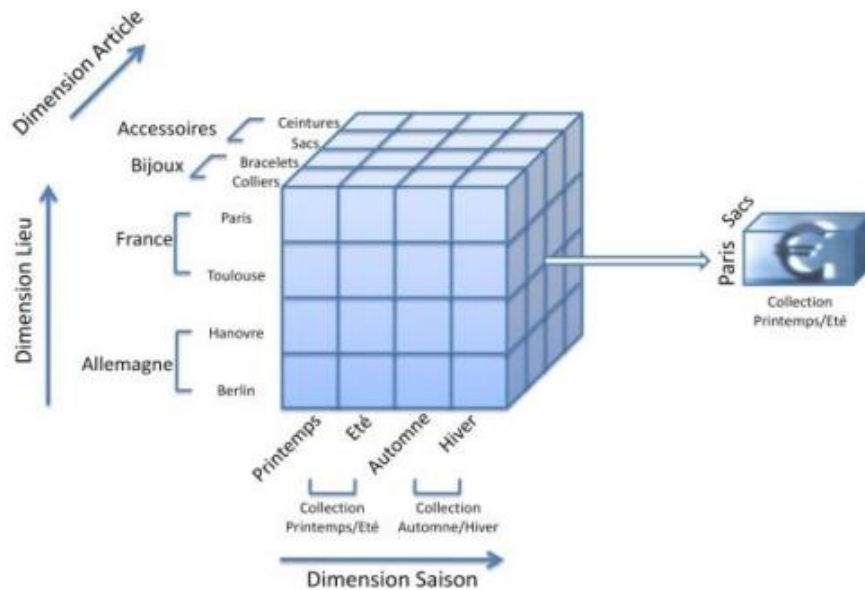
Acronyme : On-line Analytical Processing

C'est une représentation abstraite d'informations multidimensionnelles exclusivement numérique.

Cette structure est prévue à des fins d'analyses interactives par une ou plusieurs personnes (souvent ni informaticiens ni statisticiens) du métier que ces données sont censées représenter.

##### •les avantages :

- Obtenir des informations déjà agrégées selon les besoins d'utilisateur.
- Simplicité et rapidité d'accès.
- Capacité à manipuler les données agrégées selon différentes dimensions.



- Temps de réponse très court.
- Pas jointure

## Opérations OLAP

3 catégories d'opérations élémentaires :

- **Restructuration** : concerne la représentation, permet un changement de points de vue selon différentes dimensions : opérations liées à la structure, manipulation et visualisation du cube :

- Rotate/pivot
- Switch

- **Granularité** : concerne un changement de niveau de détail : opérations liées au niveau de granularité des données :

- roll-up
- drill-down

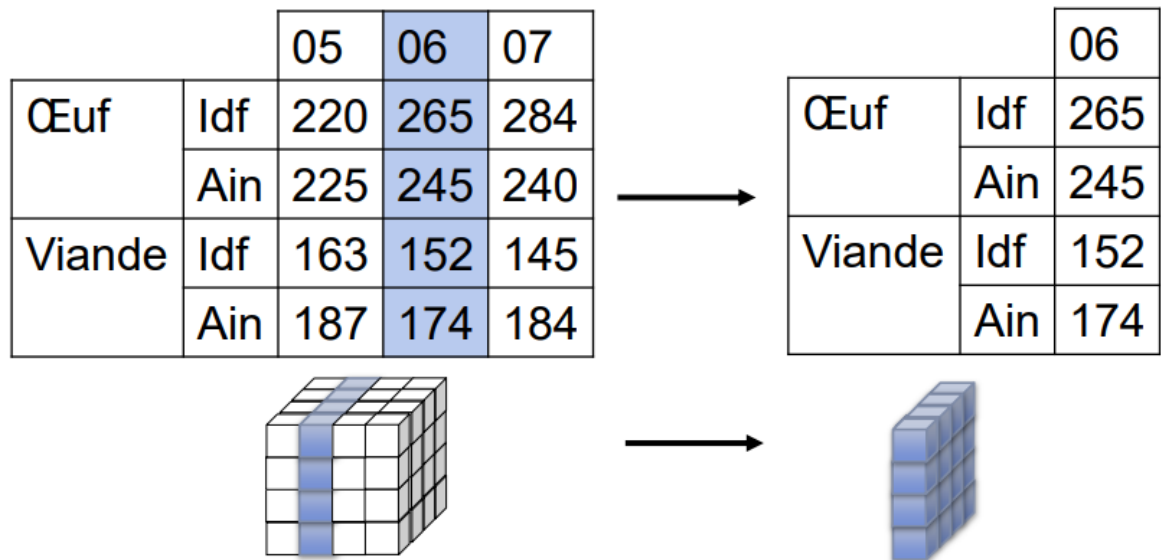
- **Ensembliste** : concerne l'extraction et l'OLTP classique :

- slice, dice
- Selection
- Projection

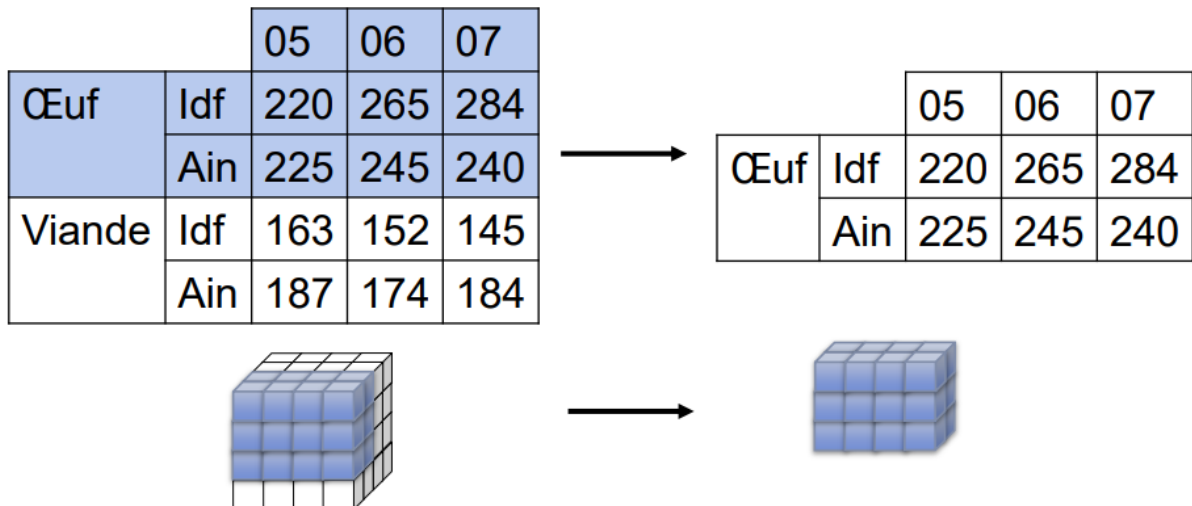
- jointure (drill-across)

## Manipulation d'un cube OLAP

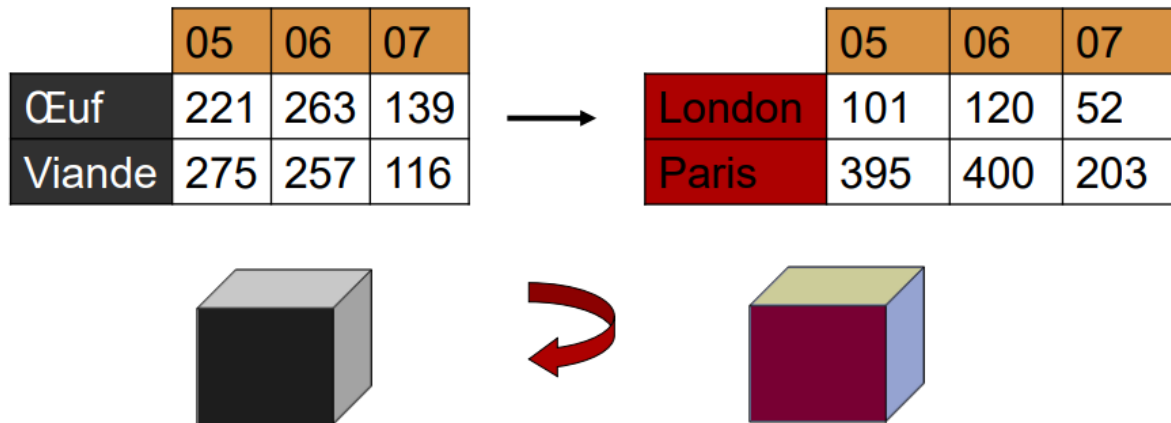
**SLICING:** Sélection de tranches du cube par des prédicats selon une dimension



**DICING:** extraction d'un sous-cube



**Roll-Up:** représenter les données du cube à un niveau de granularité supérieur conformément à la hiérarchie définie sur la dimension.



- **Drill down** : c'est la possibilité de « zoomer » sur une dimension (par exemple d'éclater les années en 4 trimestres pour avoir une vision plus fine, ou de passer du pays aux différentes régions)
- **Roll up** : c'est l'opération inverse qui permet d'« agréger » les composantes de l'un des axes (par exemple de regrouper les mois en trimestre, ou de totaliser les différentes régions pour avoir le total par pays)

### Types de Cube

- Relational OLAP (ROLAP)
  - Données sont stockées dans un SGBD relationnel
  - Un moteur OLAP permet de simuler le comportement d'un SGBD multi-dimensionnel
- Multidimensional OLAP (MOLAP)
  - Structure de stockage en cube
  - Accès direct aux données dans le cube
- Hybrid OLAP (HOLAP)
  - Données stockées dans SGBD relationnel (données de base) + structure de stockage en cube (données agrégées)