

# Utilisation avancée d'outils de reporting

Amira BEN HADID – Doctorante et ingénieure en informatique

Novembre 2023

# Déroulement du module

- 2h CM ~ 1 séance
- 2h TD ~ 1 séance
- 16h TP ~ 8 séances

Pour l'évaluation:

- 1 évaluation intermédiaire
- 1 examen final

# L'Informatique Décisionnelle (Business Intelligence) ?



# Définition de l'Informatique Décisionnelle

"Terme générique qui englobe les applications, l'infrastructure, les outils et les meilleures pratiques permettant l'accès et l'analyse de l'information afin d'améliorer et d'optimiser les décisions et les performances."



# Mise en Situation

- Le chiffre d'affaires a baissé en octobre. Pour y remédier, des décisions sont à prendre.
- Pour prendre les bonnes décisions, il faut savoir:
  - ☐ Pourquoi le CA a-t-il baissé ?
  - ☐ Comment a-t-il baissé ?
  - ☐ Dans quelles gammes de produits ?
  - ☐ Dans quels pays, quelles régions ?
  - ☐ Dans quels segments de distribution ?
  - ☐ N'avait-on pas une baisse semblable en octobre chaque année ?

# Pourquoi Construire un Système Décisionnel ?

- Un des actifs les plus importants des sociétés, c'est leur **capital d'informations** qu'elles collectent au jour le jour
- Généralement, la plupart de ces informations sont **inaccessibles**, ou réparties dans une multitude de systèmes
- Besoin de:
  - ☐ Avoir une vision globale homogène et cohérente des informations manipulées par les différents départements
  - ☐ Accéder directement à l'information nécessaire
  - ☐ Donner un sens unique aux données de gestion (marge, CA, ...)

# Pourquoi Construire un Système Décisionnel ?

- Le Système d'Information Décisionnel (**SID**) résulte d'un processus qui consiste à:
  - ☐ Extraire les données à partir des systèmes opérationnels et d'autres sources externes à l'entreprise
  - ☐ Les transformer en information de pilotage
  - ☐ Les rendre accessibles aux utilisateurs
- La Base Décisionnelle est aujourd'hui reconnue comme un actif stratégique par beaucoup d'entreprises

# La Chaîne Décisionnelle

- Les 5 grandes étapes :





# La Chaîne Décisionnelle



- Pour mettre en place une plate-forme décisionnelle, la première étape est donc la planification de ce projet
- Un tel projet nécessite une administration solide et une spécification des besoins clairs
- Exemple : les ressources humaines
  - ☐ Un responsable peut voir le salaire des personnes de son équipe
  - ☐ Mais ne peut pas voir celui de son chef
  - ➔ Nécessité d'une stratégie de sécurité rigoureuse

# La Chaîne Décisionnelle



- **ETL** : Extract, Transform, Load
    - ☐ **Extraction des données** à partir d'une ou plusieurs sources de données: fichiers texte, Excel, base de donnée, ...
    - ☐ **Transformation** des données agrégées
    - ☐ **Chargement** des données dans la banque de données de destination (datawarehouse)
- ➔ La phase d'ETL est incontournable car elle conditionne et influence la qualité du projet par la suite

# La Chaîne Décisionnelle



- Plusieurs manières de stocker la donnée dans un datawarehouse
- Chacune ayant ses avantages et ses inconvénients
- L'administrateur des bases de données décisionnelles pourra notamment choisir entre plusieurs modèles (en étoile, en flocon, ...)

# La Chaîne Décisionnelle



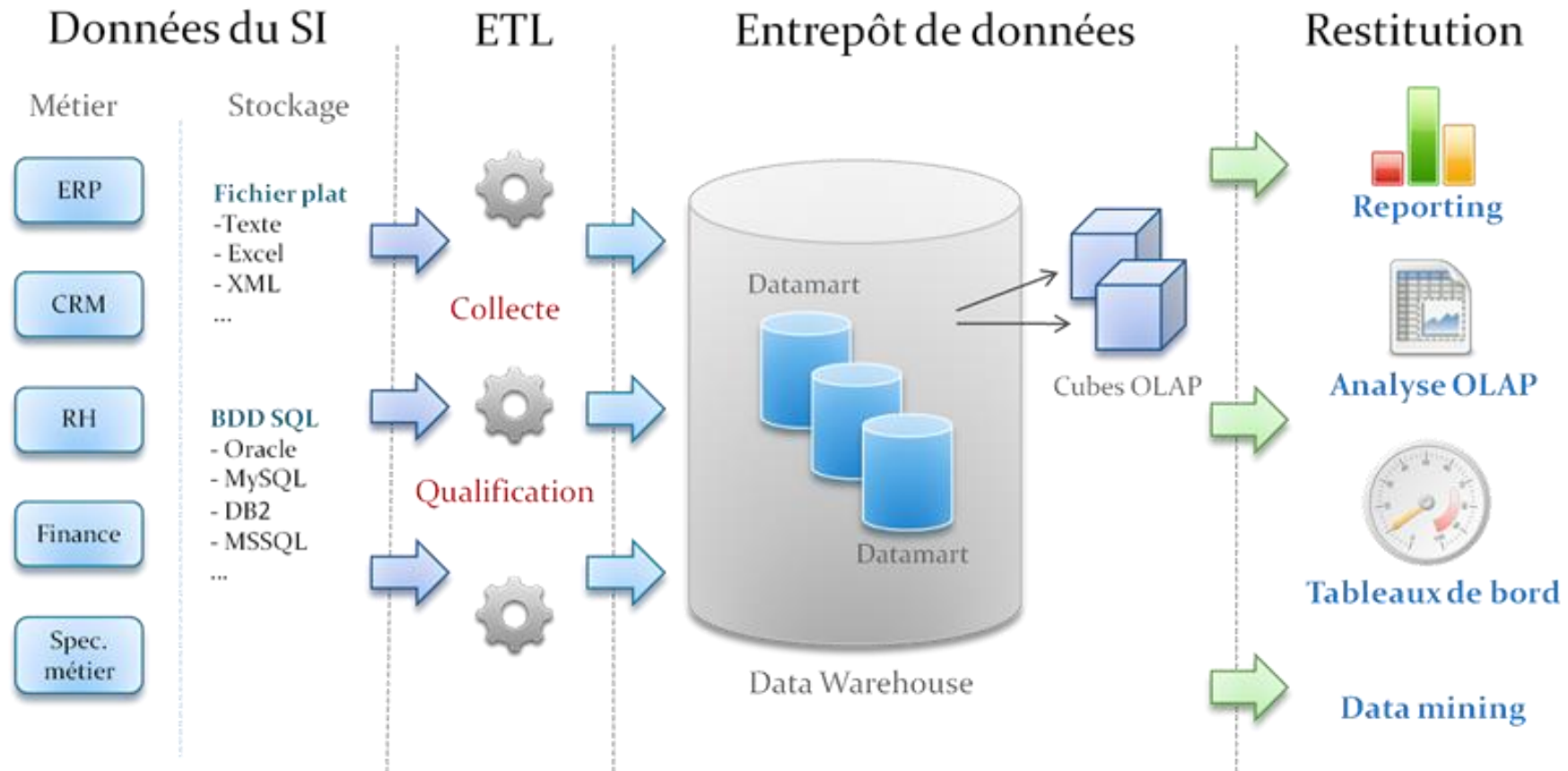
- Regroupement de l'ensemble des techniques de statistique, de Data Mining et de recherche opérationnelle
- Demande souvent des compétences statistiques avancées

# La Chaîne Décisionnelle



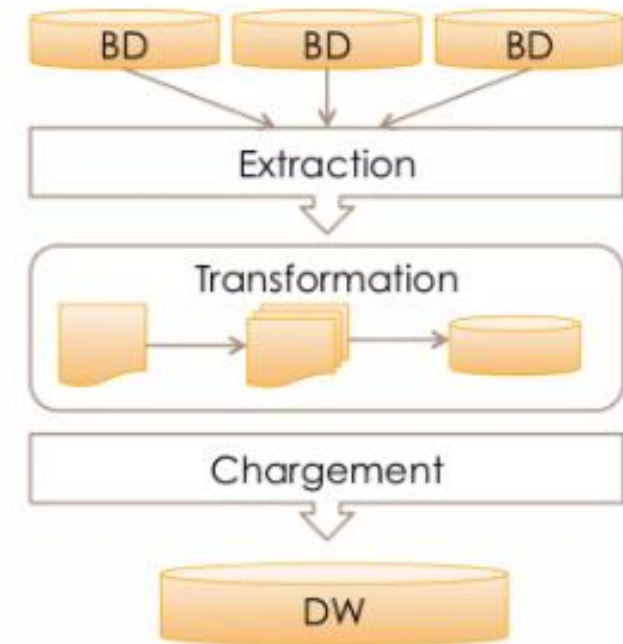
- La génération de tableau de bord
- C'est la partie que voient la plupart des utilisateurs
- Ce sont généralement des interfaces intuitives permettant à un utilisateur, en fonction de ses droits, de consulter des rapports, des tableaux de bord, de les annoter, voire de les créer lui-même

# Architecture SID



# ETL : Extract, Transform, Load

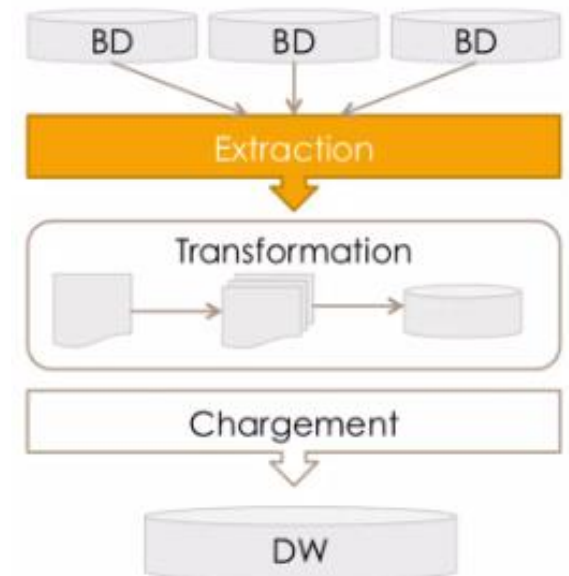
- Extraire les données sources
- Transformer les données sources
- Charger ces données dans le DW



# ETL : Extract, Transform, Load

## Extraction

- Extraction des données de leur environnement d'origine (base de données relationnelles, fichiers XML, ...)
- Besoin d'outils spécifiques pour accéder aux bases de production (requêtes sur des BD hétérogènes)
- Besoin d'une technique appropriée pour n'extraire que les données nécessaires

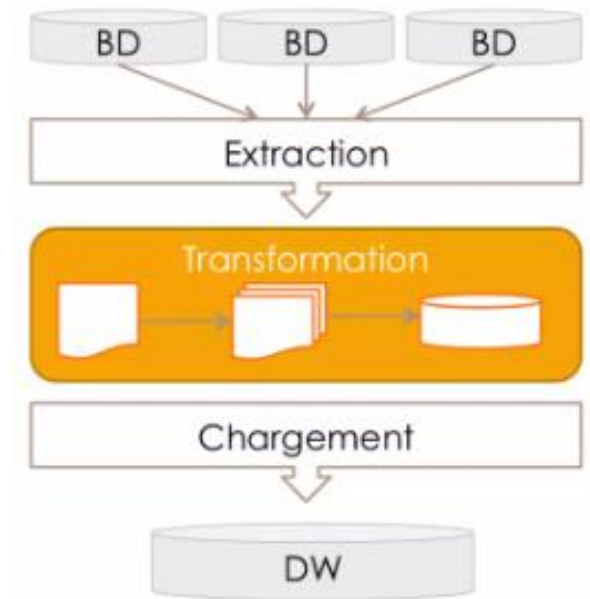




# ETL : Extract, Transform, Load

## Transformation

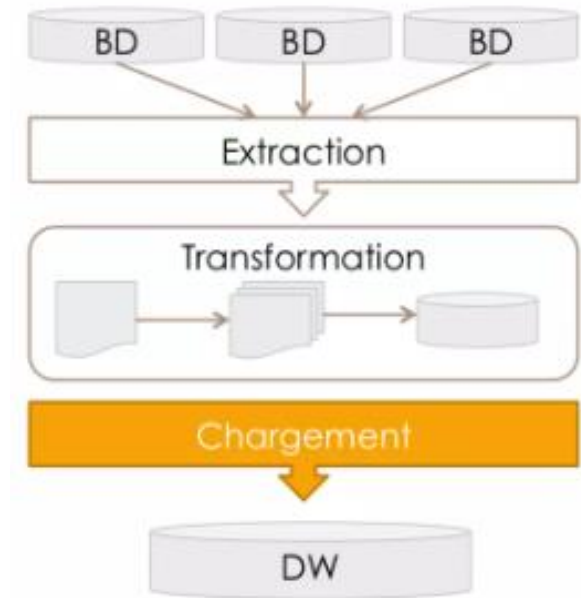
- Homogénéisation du vocabulaire, structures, valeurs
- Suppression et fusion des redondances
- Suppression des données incohérentes
- Transformation des données dans un format cible



# ETL : Extract, Transform, Load

## Chargement

- Mise en place des procédures de chargement (nocturne ? Manuel ? ) et de restauration (en cas de problème)
- Prise en compte de la notion de **granularité**



# ETL : Extract, Transform, Load

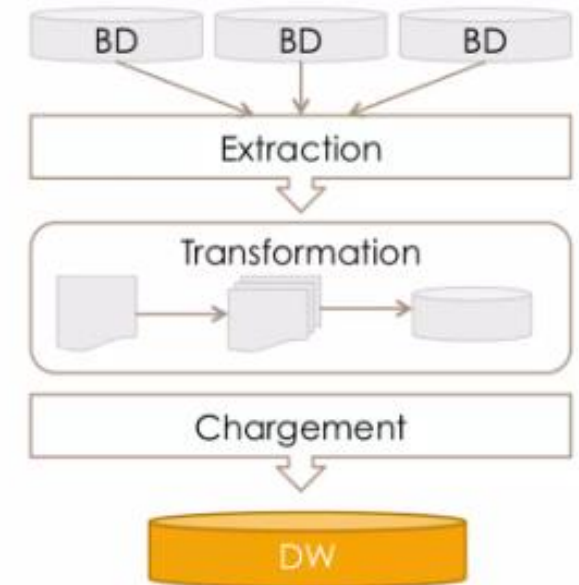
## La granularité

- La granularité en ETL désigne le niveau de détail des données traitées, allant de très détaillé (haute granularité) à résumé (basse granularité).
- **Impact sur les Analyses** : La granularité choisie affecte directement le type et la précision des analyses possibles. Une granularité plus élevée permet des analyses plus fines, tandis qu'une granularité plus basse se limite à des vues d'ensemble et des tendances.
- **Considérations de Performance et Stockage** : Une granularité élevée peut augmenter les besoins en stockage et impacter les performances, alors qu'une granularité plus basse réduit ces besoins mais peut limiter la profondeur de l'analyse.
- **Choix Stratégique** : Le niveau de granularité en ETL doit être choisi en fonction des objectifs d'affaires, des besoins analytiques et des capacités techniques de l'organisation.

# Entrepôts de données

## Définition

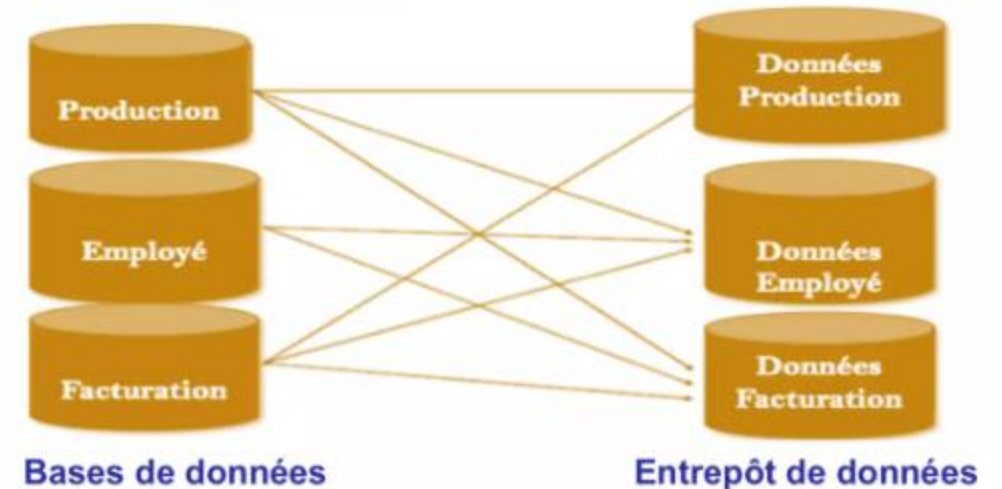
Le Data Warehouse est une collection de données **orientées sujet, intégrées, non volatiles** et **historisées**, organisées pour la prise de décision



# Entrepôts de données

## Données Orientées Sujet

- les données dans un Data Warehouse sont organisées et structurées autour de sujets clés pour l'entreprise, plutôt que par applications ou sources de données spécifiques.
- Par exemple, un Data Warehouse peut avoir des données organisées autour de sujets tels que les ventes, les clients, les produits, etc. Chaque sujet représente un domaine d'analyse pertinent pour la prise de décision d'affaires.



# Entrepôts de données

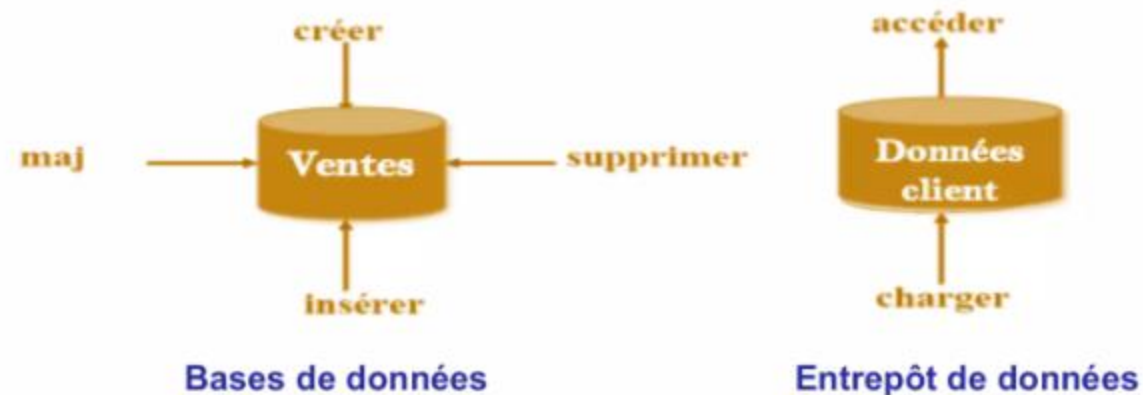
## Données Intégrées

- L'intégration fait référence au processus de rassemblement de données provenant de différentes sources dans un format cohérent et unifié au sein du Data Warehouse.
- L'objectif est de s'assurer que les données dans le Data Warehouse sont cohérentes et peuvent être utilisées de manière fiable pour l'analyse et la prise de décision.

# Entrepôts de données

## Données Historisées et Non Volatiles

- Non Volatiles : une donnée entrée dans un entrepôt l'est pour de bon et n'a pas vocation à être supprimée
- Historisées : les données sont datées, et un historique est conservé



# Systemes Operationnels vs Systemes Decisionnels

## OLTP vs OLAP

- Deux approches clés de la gestion de données.
- OLTP (Online Transaction Processing):
  - Dédiés aux métiers de l'entreprise pour les assister dans leurs tâches de gestion quotidienne
  - Utilisation des ERP pour la gestion des données
  - Utilisateurs : Agents opérationnels, nombreux et concurrents
  - Données :
    - Détaillées
    - Récentes
    - Réparties
    - Non homogènes



# Systemes Operationnels vs Systemes Decisionnels

## OLTP vs OLAP

- OLAP (On-Line Analytical Processing):
  - ❖ Dédiés à la gestion de l'entreprise pour l'aider au pilotage de l'activité pour une vision transversale de l'entreprise
  - ❖ Utilisation des entrepôts de données
  - ❖ Utilisateurs : Décideurs, peu nombreux et non concurrents
  - ❖ Données :
    - Globalisées
    - Historiques
    - Centralisées
    - Intégrées

# Systemes Operationnels vs Systemes Decisionnels

## OLTP vs OLAP

### **OLTP : On-Line Transactional Processing**

- Systeme destiné à offrir le moyen à une application d'utiliser de façon transactionnelle un serveur de base de données.
- Ensemble de logiciels que l'utilisateur peut employer de façon interactive pour accéder aux données de la manière la plus rapide et simple possible.
- Exemple : Le 15/01/2012 à 13h12, le client X a retiré 500dt du compte Y

### **OLAP : On-Line Analytical Processing**

- Catégorie de technologie logicielle permettant aux analystes, managers et décideurs d'accéder de manière rapide, consistante et interactive à une large variété d'information, transformée pour refléter la dimension réelle d'une entreprise.
- Exemple : Quel est le volume des ventes par produit et par région durant le deuxième trimestre de 2012?

# Systemes Operationnels vs Systemes Decisionnels

## OLTP vs OLAP

	OLTP	OLAP
<b>Conception</b>	<ul style="list-style-type: none"><li>- Orientée application (Application de production, de facturation...)</li><li>- Structure statique (E/R)</li></ul>	<ul style="list-style-type: none"><li>- Orientée sujet (Client, produit, vendeur...)</li><li>- Structure évolutive (en étoile, en flocon)</li></ul>
<b>Données</b>	<ul style="list-style-type: none"><li>- Détaillées, non agrégées</li><li>- Récentes, mises à jour</li><li>- Accessibles de façon individuelle</li><li>- Normalisées</li></ul>	<ul style="list-style-type: none"><li>- Résumées, recalculées, agrégées</li><li>- Historiques</li><li>- Accessibles de façon ensembliste</li><li>- Dénormalisées</li></ul>
<b>Vue</b>	<ul style="list-style-type: none"><li>- Relationnelle</li></ul>	<ul style="list-style-type: none"><li>- Multidimensionnelle</li></ul>
<b>Requêtes/ Utilisation</b>	<ul style="list-style-type: none"><li>- Simples, nombreuses, régulières, prévisibles, répétitives</li><li>- Sensibles aux performances (réponses immédiates)</li><li>- Accès à beaucoup de données</li></ul>	<ul style="list-style-type: none"><li>- Complexes, peu nombreuses, irrégulières, non prévisibles</li><li>- Non sensibles aux performances (réponses moins rapides)</li><li>- Accès à beaucoup d'informations</li></ul>
<b>Utilisateurs</b>	<ul style="list-style-type: none"><li>- Agents opérationnels</li><li>- Nombreux (par milliers)</li><li>- Concurrents</li></ul>	<ul style="list-style-type: none"><li>- Managers / Analystes</li><li>- Peu (par dizaines, centaines)</li><li>- Non concurrents</li></ul>
<b>Accès</b>	<ul style="list-style-type: none"><li>- Lecture / Écriture</li></ul>	<ul style="list-style-type: none"><li>- Lecture</li></ul>
<b>Taille de la base</b>	<ul style="list-style-type: none"><li>- 100 MB à 1 GB</li></ul>	<ul style="list-style-type: none"><li>- 100 GB à 1 TB</li></ul>

# Modélisation de DW

- Nous avons vu que l'OLAP est orienté vers l'analyse complexe pour soutenir la prise de décision stratégique, exploitant des données historiques et centralisées.
- Pour que ces analyses soient efficaces, les données doivent être structurées de manière à faciliter les interrogations multidimensionnelles et complexes.
- Les modèles en étoile et en flocon organisent les données pour les analyses OLAP, liant tables de faits et dimensions pour des requêtes rapides.
- Ils sont essentiels à la construction **des cubes OLAP**, outils clés pour l'analyse décisionnelle multidimensionnelle.

# Modélisation de DW

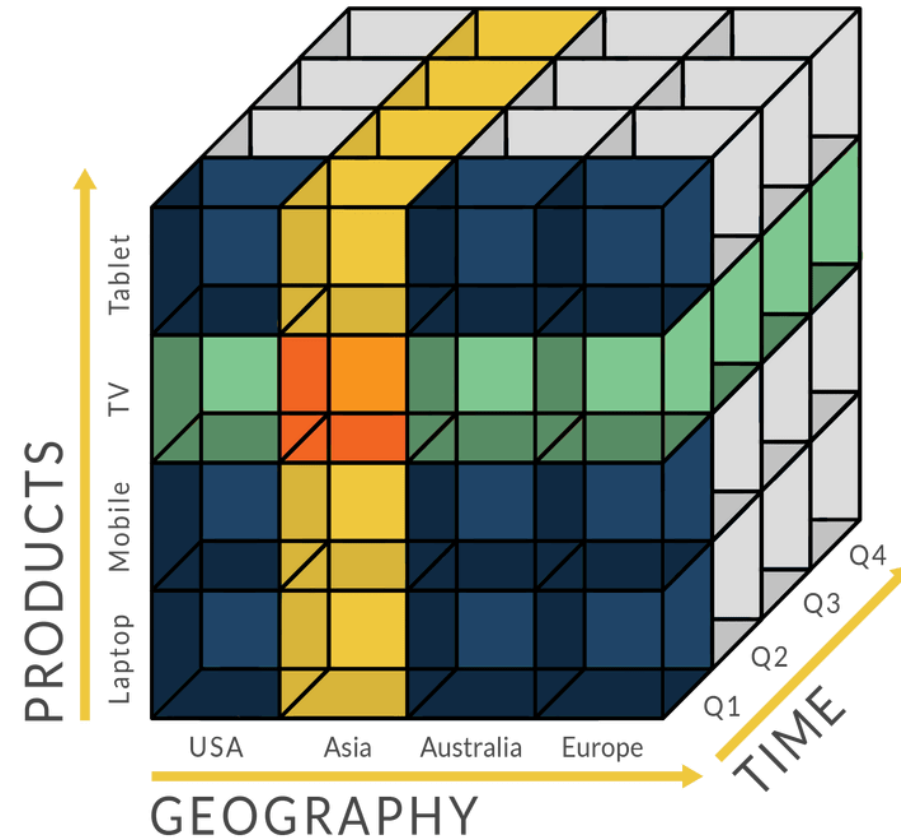
## Cube OLAP

- Un cube OLAP est une structure de données multidimensionnelle utilisée dans le cadre de l'analyse décisionnelle pour permettre une exploration et une analyse rapides et intuitives des données.
- Il organise les données en dimensions et mesures, les dimensions représentant les différentes perspectives d'analyse (comme le temps, la géographie, les produits) et les mesures représentant les valeurs quantitatives (comme les ventes, les coûts).
- Les utilisateurs peuvent effectuer des opérations comme le pivotement, le tranchage, le forage et le regroupement pour examiner les données sous divers angles et extraire des informations pertinentes pour la prise de décision.

## Modélisation de DW

# Cube OLAP

Le nombre de PC vendus en Europe durant les trois premiers mois de l'année.?



# Modélisation de DW

## Notions de Base

- La modélisation multi-dimensionnelle est une méthode de conception logique qui vise à présenter les données sous une forme standardisée, intuitive, qui permet des accès hautement performants
- Permet de considérer un sujet analysé comme point dans un espace à plusieurs dimensions
- Les données sont organisées de manière à mettre en évidence :
  - Le sujet  $\Rightarrow$  **Le Fait**
  - Les perspectives de l'analyse  $\Rightarrow$  **Les Dimensions**

# Modélisation de DW

## Les Faits

- Un fait
  - Sujet d'analyse
  - Grain de mesure de l'activité
  - Résultat d'une opération d'agrégation des données
  - Valeur d'une mesure, telle que : Chiffre d'affaires, ventes, gain, nombre de transactions
  - En général, une valeur numérique
- Une table des faits
  - Clé composite référençant les clés primaires des tables de dimension
  - Contient les valeurs des mesures et des clefs vers les tables des dimensions
  - Plusieurs tables de fait dans un datawarehouse
  - A en général plusieurs lignes et peu de colonnes



# Modélisation de DW

## Les Faits

**Fait** : Montant des ventes, chaque jour pour chaque produit dans chaque magasin



# Modélisation de DW

## Les Dimensions

- Une Dimension
  - Thème ou axe selon lequel les données sont analysées
  - En général sous forme textuelle
  - Parfois discrète (ensemble de valeurs)
- Une table des dimensions
  - Représente le point d'entrée de l'entrepôt de données
  - Contient une clef primaire unique qui correspond à l'un des composants de la clef multiple de la table des faits
  - A en général plusieurs colonnes et peu de lignes

# Modélisation de DW

## Les Dimensions

Produit
<b>Clé Produit</b> Description produit Description marque Description catégorie Description type emballage Taille emballage Poids Unité de mesure du poids Type de stockage Type de durée rayon Largeur sur étagère Hauteur sur étagère Profondeur sur étagère

# Modélisation de DW

## Modèle en étoile

- Une (ou plusieurs) table(s) de faits comprenant une ou plusieurs mesures
- Plusieurs tables de dimension dé-normalisées : descripteurs des dimensions.
- Les tables de dimension n'ont pas de lien entre elles.
- Avantages
  - Facilité de navigation.
  - Performances : nombre de jointures limité; gestion des données creuses.
  - Gestion des agrégats
- Inconvénients
  - Redondances dans les dimensions.
  - Alimentation complexe..

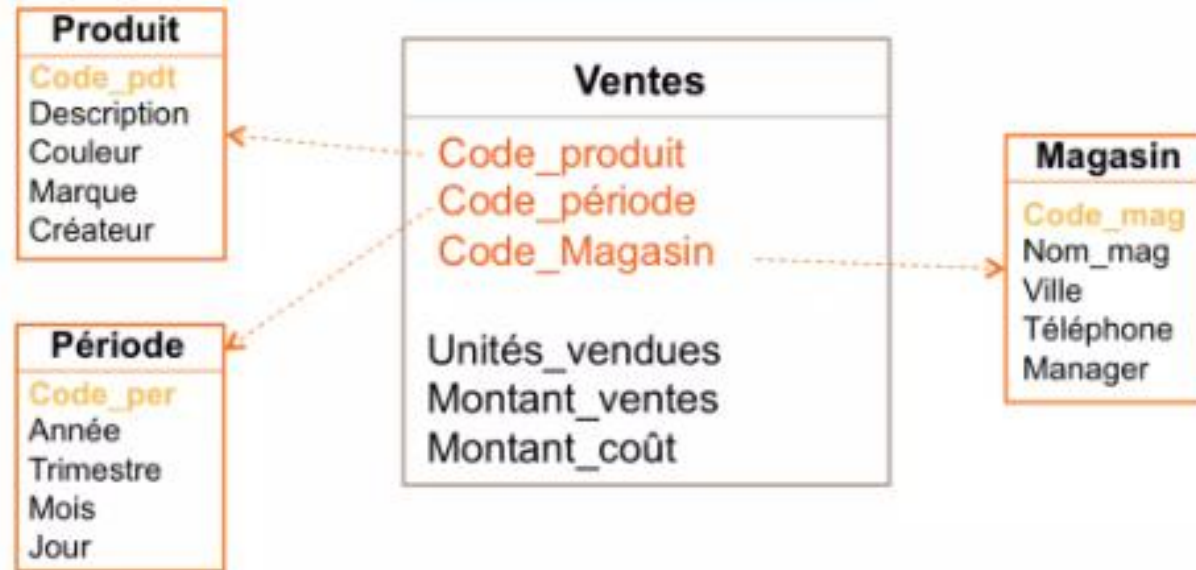
# Modélisation de DW

## Les Tables De Dimensions

- **Table de Dimension Normalisée :**
  - Une table normalisée est conçue pour minimiser la redondance des données en séparant les données en plusieurs tables reliées par des relations. (table date, mois, ...)
- **Table de Dimension Dénormalisée :**
  - Une table dénormalisée, en revanche, consolide les données en une seule table, ce qui peut introduire de la redondance mais améliore les performances des requêtes en réduisant le nombre de jointures nécessaires.

# Modélisation de DW

Modèle en étoile



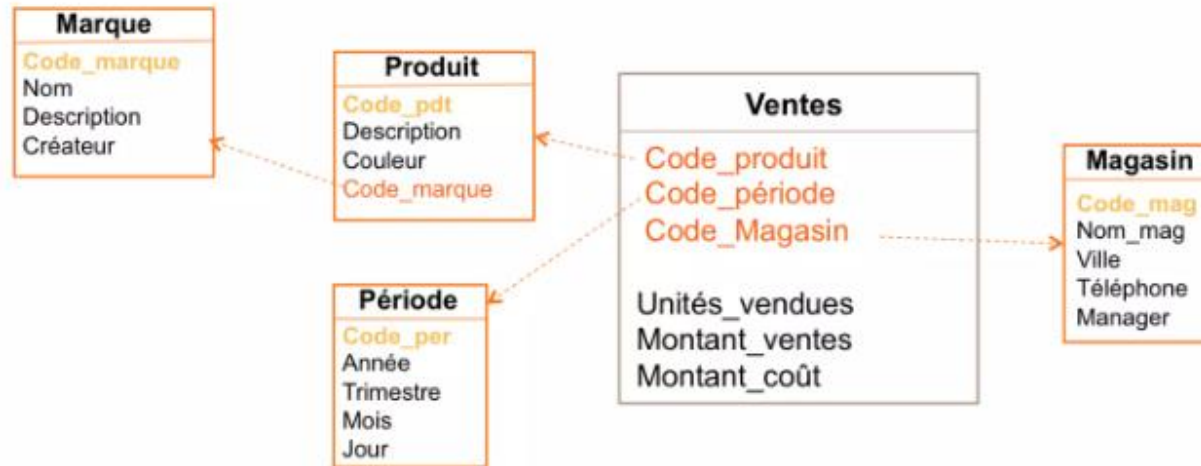
# Modélisation de DW

## Modèle en Flocon

- Dérivé du schéma en étoile où les tables de dimensions sont normalisées
- La table des faits reste inchangée
- Chacune des dimensions est décomposée selon sa (ou ses) hiérarchie(s)
- Exemple : Commune, Département, Région, Pays, Continent
- Utilisé lorsque les tables sont très volumineuses
  
- Avantages
  - Réduction du volume
  - Permettre des analyses par pallier (drill down) sur la dimension hiérarchisée
  
- Inconvénients
  - Navigation difficile
  - Nombreuses jointures

# Modélisation de DW

## Modèle en Flocon





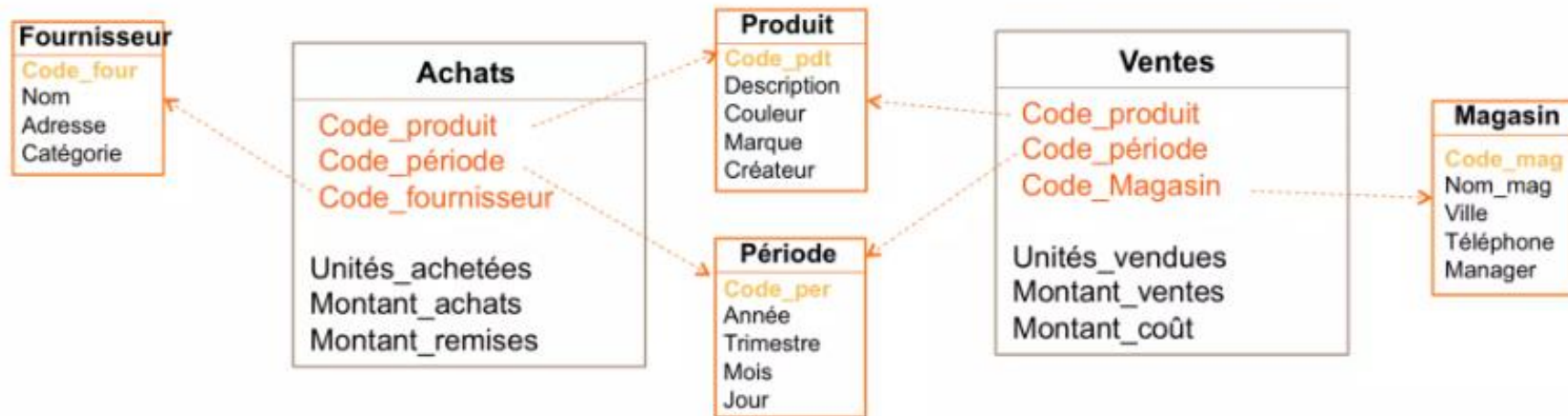
# Modélisation de DW

## Modèle en Constellation

- Fusionner plusieurs modèles en étoile qui utilisent des dimensions communes
- Un modèle en constellation comprend donc :
  - Plusieurs tables des faits
  - Des tables de dimension communes ou non à ces tables de faits

# Modélisation de DW

## Modèle en Constellation



# L'outil de Reporting



Power BI

**Des questions ?**

**Merci :)**