

# Informatique décisionnelle

DUT STID 2ème année

F.-X. Jollois

## Sources intéressantes :

- <http://wikissimo.aim.asso.fr/>
- [http://fr.wikipedia.org/wiki/Syst%C3%A8me\\_d%27information](http://fr.wikipedia.org/wiki/Syst%C3%A8me_d%27information)
- [http://fr.wikipedia.org/wiki/Informatique\\_d%C3%A9cisionnelle](http://fr.wikipedia.org/wiki/Informatique_d%C3%A9cisionnelle)

## Livres intéressants :

- *Concevoir et déployer un data warehouse*, R. Kimball, L. Reeves, M. Ross et W. Thornthwaite, Eyrolles

# Définition d'un SI

- Ensemble organisé de ressources qui permet de collecter, stocker, traiter et diffuser de l'information ;
- Composé de deux parties :
  - Technologique : **logiciels**, de **matériels** et des **réseaux de télécommunication**,
  - Sociale : personnels, structure, ... (dont nous parlerons pas) ;
- Ne se résume donc pas à un unique SGBD

*Un SI est un réseau complexe de relations structurées où interviennent hommes, machines et procédures qui a pour but d'engendrer des flux ordonnés d'informations pertinentes provenant de différentes sources et destinées à servir de base aux décisions (Hugues Angot).*

# Composition

Dans un système d'information d'une grande entreprise, on trouve :

- **ERP** (pour *Enterprise Resource Planning*, **PGI** pour *Progiciel de Gestion Intégré*), intégrant les systèmes informatisés transactionnels ;
- des **progiciels métiers**, couvrant aussi le front-office, le middle, et/ou le back-office, et qui ne sont pas de conception maison, mais ont été bati par un éditeur spécialisé sur un métier et dont les modes de fonctionnement logiciels correspondent aux meilleurs pratiques constatées à un moment donné ;
- des systèmes **spécifiques** (non standards, de conception *maison*, développés sur mesure, que l'on ne trouve pas sur le marché, ...), où l'on rencontrera davantage d'applications dans les domaines du calcul de coûts, de la facturation, de l'aide à la production, ou de fonctions annexes.

# Pourquoi un SI ?

- Besoin de faciliter la gestion de l'entreprise au quotidien (rapidité de traitement, erreurs minimisées, simplicité à prévoir des programmes, ...);
- Besoin d'intégrer tous les éléments dans un système permettant l'unicité de chaque information (pas de redondance, données mises à jour en temps réel - par exemple, dans la gestion de stock);
- Limitation des supports non connectés (papiers, fichiers personnels, ...) pour l'intégration de toutes les données dans un même système ;
- Rendre possible l'interaction directe des clients (particuliers ou entreprises) avec les données de l'entreprise (catalogue en ligne, commande en ligne, échange d'informations entre entreprises, ...).

Un SI est fait pour que l'entreprise fonctionne au mieux, mais elle a besoin de faire des analyses pour s'améliorer, d'où la nécessité d'un **SI Décisionnel**.

# Trois concepts importants

- **Données** : mesure simple collectée par un capteur ou renseignée par une personne.

*Par exemple, le client 123456 a retiré 200€ au DAB n°321 le 01/09/2015*

- **Information** : calcul ou agrégation de données.

*La moyenne journalière des retraits au DAB n°321 est de 500€*

- **Connaissance** : résultat d'une réflexion sur la base des informations disponibles, et permettant à l'entreprise de mieux comprendre son fonctionnement et éventuellement de l'améliorer.

*Le DAB n°321 est très peu utilisé, et d'autres sont présents à proximité, il serait donc économique de le supprimer*

# Définition

L'informatique décisionnelle (ou **Business Intelligence - BI**) désigne les moyens, les outils et les méthodes qui permettent de collecter, consolider, modéliser et restituer les données d'une entreprise en vue d'offrir une aide à la décision.

Elle cherche à :

- Mesurer des indicateurs (ou mesures, **faits**, métriques)
- Restituer ceux-ci selon des axes d'analyse (ou **dimensions**)

La réalisation technique est appelée **Entrepôt de données** (ou **Data Warehouse - DW**), global pour l'entreprise, et composé de plusieurs **Magasin de données** (ou **Data Mart - DM**) plus spécifiques.

# Objectifs du data warehouse

- Rendre accessible les informations de l'entreprise (en terme de compréhension et de navigation simple et rapide) ;
- Rendre cohérente l'information d'une entreprise (qualité de l'information) ;
- Constituer une source d'information souple et adaptable (en prévoyant les modifications à venir : nouveaux produits, nouveaux processuss, ...) ;
- Représenter un bastion sécurisé qui protège le capital information (sécurité, visibilité) ;
- Constituer la base décisionnelle de l'entreprise (dans le cadre de l'aide à la décision, ce pour quoi le système est fait à la base).

Ce n'est pas une *alternative* au SI, mais bien un *complément*.



# SI vs SID

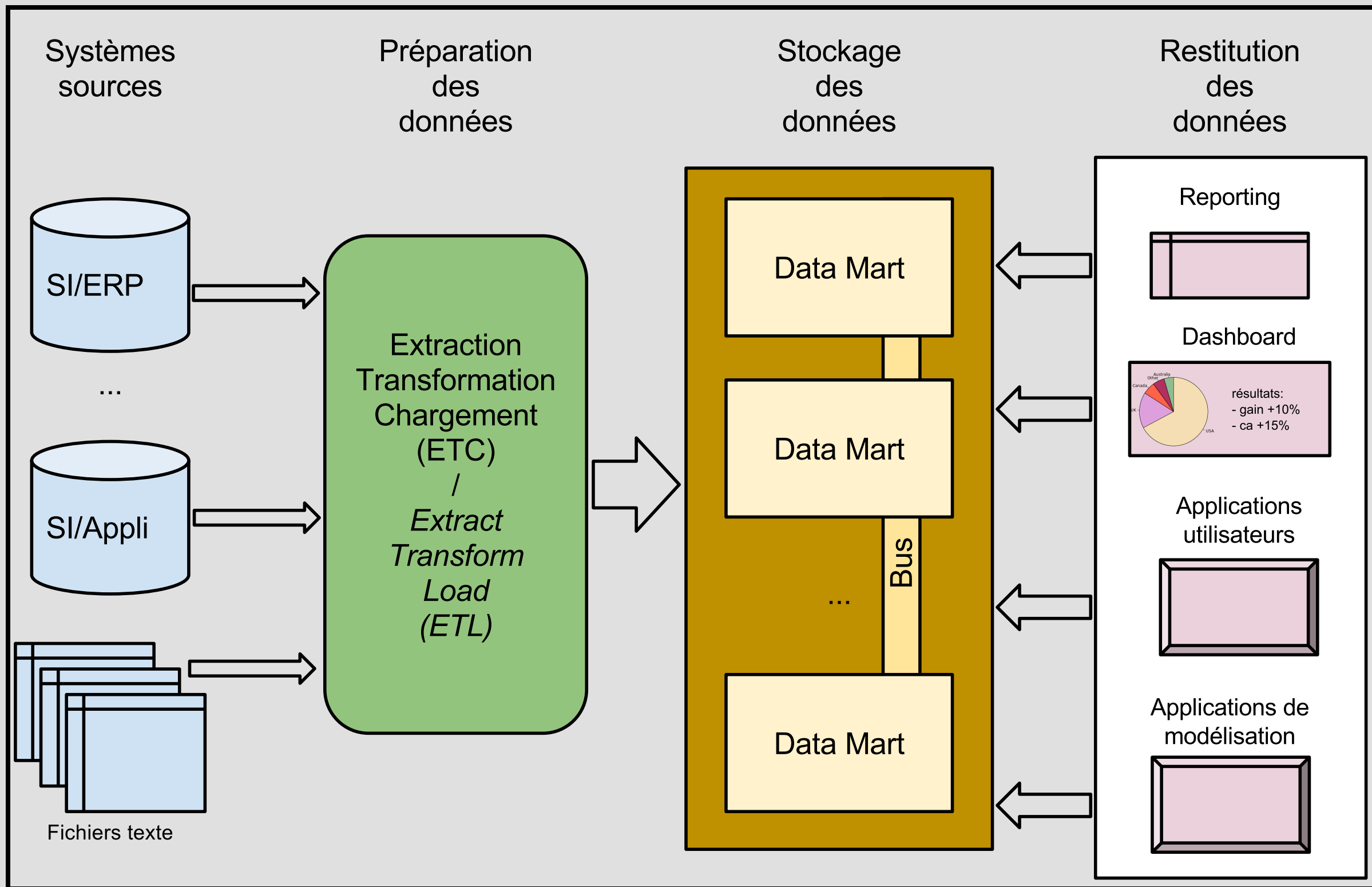
2 concepts différents :

- **OLTP** : On-Line Transaction Process
- **OLAP** : On-Line Analytical Process

<b>OLTP</b>	<b>OLAP</b>
Informatique opérationnelle	Informatique décisionnelle
Opérations journalières	Support décisionnel
Orienté applications	Orienté sujet
Usage répétitif	Usage occasionnel
Requête simple, courte	Requête complexe
Sur peu de données	Sur des millions de données
Beaucoup d'utilisateurs	Très peu d'utilisateurs

# Éléments importants d'un SID

- Préparation des données : **Extract/Tranform/Load (ETL)**
  - *Extraction* : récupération des différentes données sources
  - *Transformation* : calcul d'agrégats, nettoyage, consolidation, ...
  - *Chargement* : intégration des données dans les différents data marts
  - Zone réservée et inaccessible par les utilisateurs
  - Automatisation totale et lancement régulier
- Stockages des données :
  - Coeur du DW
  - Data mart : spécifique à un processus de l'entreprise
  - Bus permettant la liaison entre data marts
- Restitutions des données :
  - Génération automatique d'états
  - Création de reporting ad-hoc
  - Outil d'accès aux données (**OLAP**)
  - Requêtage pour analyse statistique ou autre



# Préparation des données (ETL)

## Extract

- Problèmes réguliers :
  - Hétérogénéité des sources
  - Accès aux différentes applications
  - Bases de données réparties sur différents sites
  - Différents SGBD, fichiers à plat, services web, ...
  - N'extraire que les données nouvelles dans les bases de données sources
- A prévoir :
  - Rafraichissement régulier des données
  - Marquage des données par la date de dernière modification
  - Utilisation d'un journal des transactions
  - Extensibilité de l'outil d'extraction
  - Intégrité du système à garantir (en cas de problème, panne, ...)

# Préparation des données (ETL)

## Transform

- Comment faire :
  - Filtrage parfois nécessaire
  - Nettoyage, correction orthographique, résolution de conflits de domaine, éléments manquants, conversion à des formats standard
  - Tris, traitements séquentiels, synchronisation des clés, ...
  - Modélisation en 3NF (trop coûteuse) ou **dimensionnelle**
- Problèmes usuels :
  - Connaissance précise des schémas de toutes les bases
  - Combinaison à partir de multiples sources
  - Recherche de données incohérentes
  - Conformité aux règles de références classiques à garantir
- Modélisation dimensionnelle à privilégier (simplification, rapidité, économie de temps)

# Préparation des données (ETL)

## Chargement

- Réservation d'un espace mémoire temporel dédié
  - Estimation de la taille à prévoir
- Transfert des données vers les magasins de données
  - Création si besoin des tables dimensionnelles et de faits
  - Insertion sinon
  - Historisation si mise à jour
- Indexation des données pour optimiser les requêtes futures
- Avertissement des utilisateurs de la mise à jour du magasin de données
- Précision sur les changements éventuels dans les dimensions et dans les faits

# Stockage des données

- Stockage dans une base de données le plus souvent relationnelle
- Modélisation dite **dimensionnelle** (ou *multi-dimensionnelle*)
  - Différente de la modélisation entité/relation
  - Basée sur la notion de **fait** et de **dimension**
  - Lisible par un non-informaticien
- Nombre restreint de tables par rapport à la BD original
- Utilisation massive de méta-données
  - Données sur les données
  - Structure des tables
  - Définition des faits, des agrégats
  - Hiérarchies d'attributs
  - Planning des opérations
- Requêtage SQL simplifié

# Restitution des données

- **Reporting :**
  - on parle aussi d'*état*
  - Tableaux complets d'informations, souvent agrégées, et détaillant une activité selon des axes définis
- **Dashboard :**
  - on parle aussi de *tableaux de bord*
  - Ensemble de graphiques et de tableaux sur des agrégations d'informations
- **Scorecard :**
  - ou bien *carte de performance*
  - Version plus agrégée du dashboard, sur des indicateurs dits de performance

Dans la pratique, on parle globalement de **Reporting** ou de **Tableaux de bord** pour désigner le tout.

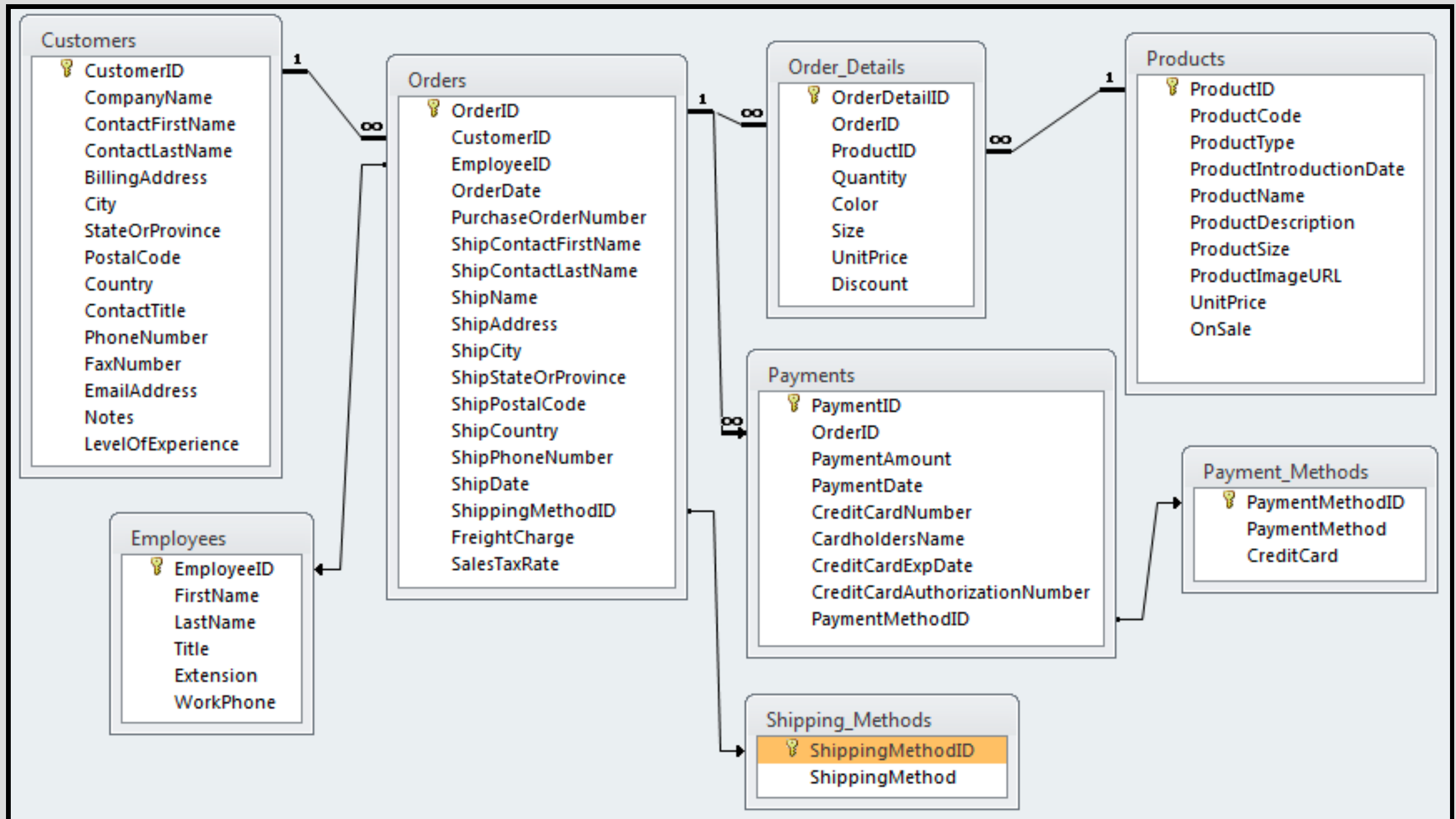


# Logiciels

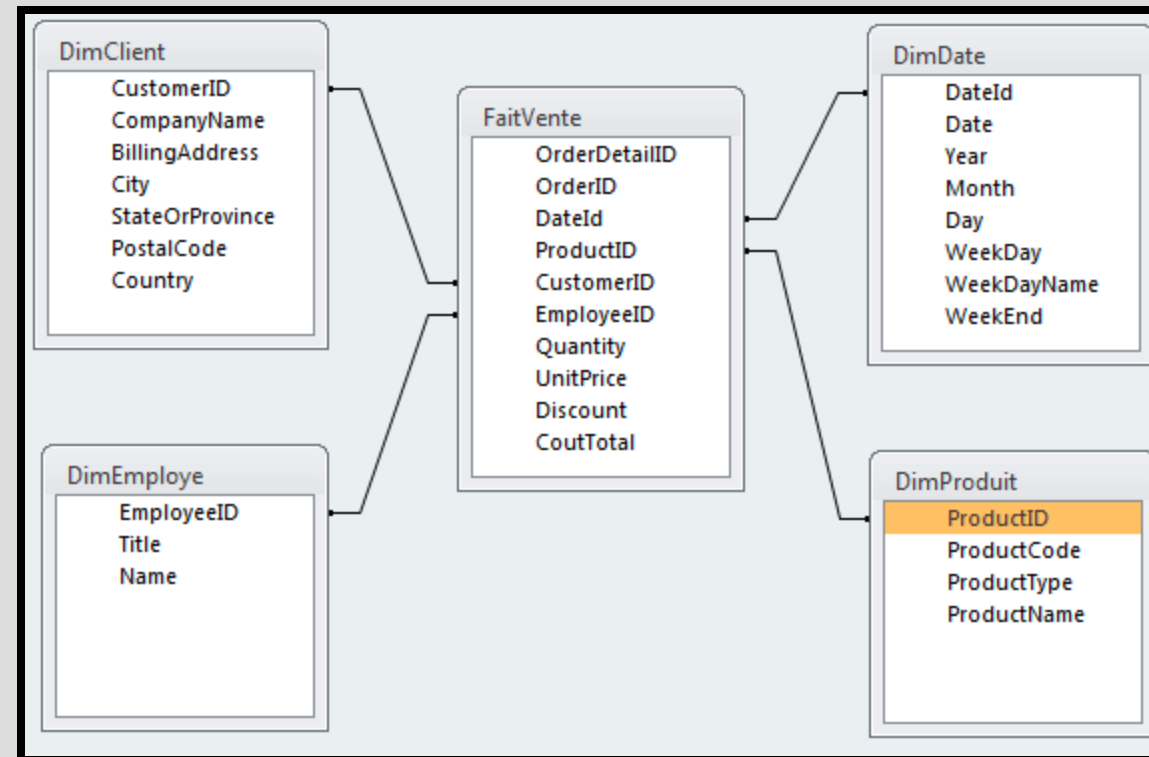
Liste non exhaustive, quelques logiciels parmi l'offre nombreuse.

	Payant	Open source
<i>ETL</i>	Microsoft SSIS	Talend
	Oracle Data Integrator	Pentaho Data Integration
<i>DW</i>	Microsoft SQL DataWarehouse	Pentaho
	Microsoft SQL Server	SpagoBI
	Oracle	JasperSoft
<i>Reporting</i>	Tableau Software	BIRT
	QlikView	JasperReports
	Microsoft SSAS	
	Business Object	

# Example : Adventures Works



# Exemple : Adventures Works



Processus modélisé ici : vente d'un produit à un client par un employé à une date

# Demande

CA journalier moyen des clients français, pour les ventes par des représentants, avec paiements par chèque et envoi via Colissimo.

A noter :

- Utilisation des presque toutes les tables
- Donc beaucoup de jointures
- Calcul du CA à partir de `Order_Details`
- Calcul du jour de la semaine à partir de la date `OrderDate`

# A partir de la BD source

```
SELECT CompanyName, JourSemaine, AVG(CA) AS CAmoyen
  FROM Customers C, Employes E, Payment_Methods PM,
    Shipping_Methods SM, Orders O, Payments P,
    (SELECT OrderID,
      SUM(UnitPrice * Quantity * (1 - Discount)) AS CA
      FROM Order_Details
      GROUP BY OrderID) CA,
    (SELECT OrderID, DAY(OrderDate) AS JourSemaine
      FROM Orders) J
 WHERE C.CustomerID = O.CustomerID
    AND O.EmployeeID = E.EmployeeID
    AND O.ShippingMethodID = SM.ShippingMethodID
    AND O.PaymentID = P.PaymentID
    AND P.PaymentMethodID = PM.MethodPaymentID
    AND O.OrderID = CA.OrderID
    AND O.OrderID = J.OrderID
    AND C.Country = "France" AND E.Title = "Représentant"
    AND SM.ShippingMethod = "Colissimo"
```

# A partir du data mart

```
SELECT CompanyName, WeekDay, AVG(CA) AS CAmoyen
  FROM FaitVente V, DimClient C, DimEmploye E, DimDate D,
      (SELECT OrderID,
          SUM(UnitPrice * Quantity * (1 - Discount)) AS CA
        FROM FraisVente
        GROUP BY OrderID) CA
 WHERE V.CustomerID = C.CustomerID
    AND V.EmployeeID = E.EmployeeID
    AND V.DateId = D.DateID
    AND V.OrderID = CA.OrderID
    AND C.Country = "France"
    AND E.Title = "Représentant"
    AND V.ShippingMethod = "Colissimo"
    AND V.PaymentMethod = "Chèque"
 GROUP BY CompanyName, WeekDay;
```

Requête seulement simplifiée et plus rapide à exécuter. Mais certaines demandes ne peuvent quasiment pas avoir de réponses sur la base de la BD source.

# Points importants

- Se concentrer sur les exigences et les objectifs
- Obtenir le soutien des responsables du côté utilisateur
- Choisir un mode de développement itératif plutôt que global
- Prévoir au mieux les coûts de chaque étape (en temps, en homme/jour, en dépenses)
- Préférer la commodité d'utilisation et la performance des requêtes plutôt que la facilité de développement
- Rechercher au maximum des solutions simples
- Respecter une architecture commune et cohérente
- Ne pas se limiter à une granularité trop importante
- Préparer et prendre en compte l'évolution future du système opérationnel et de l'environnement en général
- S'assurer de l'acceptation du système par les utilisateurs

# Conclusion

Les SID sont absolument nécessaires dans le cadre d'une entreprise pour :

- Avoir une **historisation** complète de ses données ;
- Avoir une **vision globale** de son fonctionnement ;
- Permettre de répondre à toute demande d'**analyse** pour améliorer ses processus, via des méthodes statistiques et/ou de data-mining ;
- Prendre les bonnes **décisions** sur la base de données qualifiées, consolidées et agrégées.



# Conclusion

## Impératifs et perspectives

- Avancées constantes de la technologie
- Sécurité et confidentialité
- Sauvegarde et sûreté du système (destructions physiques, piratage, ...)
- Tendances culturelles et politiques
  - Management par les chiffres
  - Utilisation croissante d'indicateurs sophistiqués
  - Sous-traitance risquée

Bien présenter les  
bonnes données