

Introduction à l'Informatique Décisionnelle - Business Intelligence (7)



Bernard ESPINASSE
Professeur à Aix-Marseille Université (AMU)
Ecole Polytechnique Universitaire de Marseille



Septembre 2013

- Emergence de l'Informatique Décisionnelle/Business Intelligence
- Problématique de l'ID
- Les outils de l'ID
- Un exemple de l'ID dans l'entreprise : la relation client

Sommaire

1. Emergence de l'Informatique Décisionnelle (ID) / Business Intelligence (BI)
2. Des SI opérationnels aux SI décisionnels
3. Les outils de l'ID/BI
4. Un exemple de l'ID/BI dans l'entreprise : la relation client (CRM)

1 – Emergence de l'Informatique Décisionnelle (ID) / Business Intelligence (BI)

- Contextes économique et informationnel actuel
- Pilotage de l'entreprise et ID
- Historique de l'ID
- Problématique de l'ID
- Domaines d'application de l'ID

Contextes économique et informationnel actuel

Contexte économique :

- Mondialisation de l'économie, ouverture de nouveaux marchés
- Concurrence toujours plus accrue
- Besoin d'informations pour prises de décisions de plus en plus rapides

Contexte informationnel :

- Décentralisation des données vers les utilisateurs
- Difficulté d'accès à l'information qui est en trop grande quantité
- Un enjeu stratégique d'entreprise
- Les informations, une source de revenu et de compétitivité

Contexte informatique :

- Puissance de calcul croissante
- Capacité de stockage croissante
- Bases de données de plus en plus importantes
- SGBD de plus en plus performants (parallélisme, ...)
- Ouverture sur le Web, ...

Pilotage de l'entreprise et informatique

- Le **pilotage** d'une entreprise dépend de ses **objectifs stratégiques**
- **Ce pilotage** doit prendre en **considération** :
 - Une organisation de plus en plus **orientée clients**
 - Des **cycles conception/fabrication** de plus en plus **courts**
 - De **nouveaux canaux de distribution** notamment les ventes en ligne sur le Web
 - L'exigence d'**internationalisation**
 - ...
- **Dans ce contexte l'entreprise se doit** :
 - d'**anticiper les besoins des clients**,
 - de **contrôler** l'intégrité et la qualité des **flux de gestion**
 - d'évaluer la **performance** des différentes entités la composant
 - ...

Outils informatiques supportant le pilotage des entreprises

Outils d'analyse :

- pour **constituer** et **mettre à jour** à partir de diverses sources des réservoirs de grande quantité de données *historisées* et *multidimensionnelles*, ...
- pour **extraire** selon divers *critères* des sous-ensembles de données de tels réservoirs
- pour **analyser** ces données selon *différents axes* (OLAP), d'**identifier** des **tendances**, des **corrélations**, faire de la **prévision** (Data Mining)

Outils de veille stratégique, de recherche d'information (RI) :

- issus de « **l'intelligence économique** » (Competitive Intelligence), ces outils ont pour finalité de **recueillir des informations sur le marché et la concurrence**
- **collecter** sur le Web d'importante quantité de données, les **filtrer** et en **extraire** les informations *pertinentes* (Web Mining) pour les analyser ensuite
=> Outils relevant de l'Informatique Décisionnelle (ID) ou Business Intelligence (BI),

Informatique Décisionnelle (ID)/Business Intelligence (BI)

- **L'Informatique Décisionnelle (ID)**, en anglais **Business Intelligence (BI)**, est l'informatique à l'usage des décideurs et des dirigeants des entreprises
- Les systèmes de ID/BI sont utilisés par les décideurs pour obtenir **une connaissance approfondie de l'entreprise** et de définir et de **soutenir leurs stratégies d'affaires**, par exemple :
 - d'acquérir un avantage concurrentiel,
 - d'améliorer la performance de l'entreprise,
 - de répondre plus rapidement aux changements,
 - d'augmenter la rentabilité, et
 - d'une façon générale la création de valeur ajoutée de l'entreprise.

Historique de l'informatique décisionnelle (1)

Années 70-90 : débuts de l'Informatique Décisionnelle :

- Développement d'**outils d'édition de rapports**, de statistiques, exploitant les BD opérationnelles
- Développement de **petits systèmes d'aide à la décision à base de tableurs** (simulation budgétaire, ...)
- Développement de **systèmes experts** (IA) systèmes à base de règles, conçus par extraction de la connaissance d'un ou plusieurs experts : intérêts et résultats limités
- Développement de **systèmes spécifiques d'aide à la décision** : basés sur des techniques de Recherche Opérationnelle (RO), la simulation, l'optimisation, ...

=> Systèmes en général mal intégrés au SI opérationnel

Historique de l'informatique décisionnelle (2)

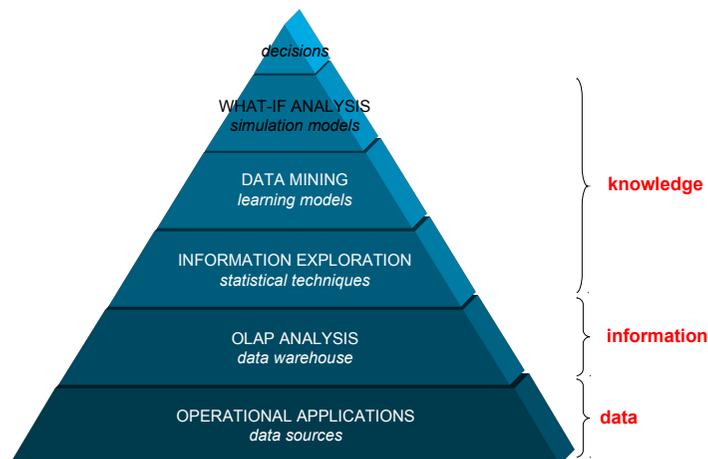
Années 90 - 2000 : essor de l'informatique décisionnelle :

- Technologie informatique permettant le développement d'**Entrepôts de Données (Data Warehouse)**
- **Nouveaux algorithmes :**
 - souvent **issus des statistiques** et de l'**IA**, permettant d'extraire des informations à partir de données brutes
 - permettant l'extraction *d'informations nouvelles* ou *cachées*, de *connaissances* à partir de données
 - regroupés dans des logiciels de **Fouille de données (Data Mining)**
- **Données issues du Web : Recherche d'Information (RI) et Fouille de données sur le Web (« Web Mining »)**

L'ID/BI moderne (après 90)

- La ID/BI est un ensemble de **processus**, d'**outils** et de **technologies** pour transformer les **données** de l'entreprise en **informations exactes** et *à jour* pour soutenir des processus décisionnels:
 - **Data Warehousing (DW)**
 - **On-Line Analytical Processing (OLAP)**
 - **Data Mining (DM) et visualisation de données (VIS)**
 - **Analyse des décisions (what-if)**
 - **Customer Relationship Management (CRM)**
- La ID/BI peut être considérée comme l'**opposé de l'Intelligence Artificielle (IA)** :
 - les *systèmes d'IA prennent des décisions pour les utilisateurs*
 - les *systèmes de BI permettent aux utilisateurs de prendre les bonnes décisions sur la base des données disponibles*
 - mais de nombreuses techniques de BI ont leurs racines dans l'IA.

La pyramide de ID/BI moderne



Exemples de requêtes relevant de l'ID/BI

- **Q1:** Le 11 Octobre 2013, trouver les 5 produits les plus vendus pour chaque sous-catégorie de produits qui représente plus de 20% des ventes dans sa catégorie de produits
- **Q2:** En date du 15 Décembre 2012, déterminer la priorité d'expédition (shipping priority) et de revenu brut potentiel (potential gross revenue) des commandes qui ont les 10 plus grandes recettes brutes (largest gross revenues) parmi les commandes qui n'avaient pas encore été expédiées. On ne considère seulement que les commandes du segment du marché du livre.

L'ID/BI est cruciale et en pleine croissance (1)

- Selon Meta Group : **DW seul = 15 billion de \$ en 2000**
- Selon Palo Alto Management Group : **BI = \$ 113 billion de \$ en 2002**
- **Le Web rend l'ID/ BI encore plus nécessaire :**
 - Les clients ne sont *pas* «physiquement» dans le magasin
 - Les clients peuvent **changer à d'autres magasins plus facilement**
 - Comment **connaître ses clients** :
 - Analyser les « **Web log** » pour analyser le comportement des clients sur le site
 - **Combiner ces données Web** avec les **données traditionnelles** des clients
- « **Internet sans fil** » ajoute à cela :
 - Les **clients sont toujours "en ligne"**
 - La **position de la clientèle est connue**
 - **Combiner la position et la connaissance sur le client** => très utile

L'ID/BI est cruciale et en pleine croissance (2)

Selon une étude du Gartner Group en 2009 :

- Les entreprises rechercheront de plus en plus des **cadres qui auront en charge l'ID/BI** et la gestion de la performance pour contribuer à la **transformation** et **l'amélioration** de leurs affaires
- **Jusqu'en 2012**, par d'outils de ID/BI **plus de 35% des 5000 principales entreprises mondiales ont régulièrement échoué dans la prise de décisions cruciales** concernant leurs marchés et leurs affaires
- **En 2010**, **20% des entreprises auront une application analytique** spécifique de ID/BI
- **En 2009**, la **prise de décision collaborative va s'imposer** et nécessiter la **combinaison d'outils de ID/BI** et de **logiciels sociaux**

Pour S. Chaudhuri, U. Dayal, V. Narasayya [CACM 2011]:

- « *Aujourd'hui, il est difficile de trouver une entreprise prospère qui n'a pas d'effet de levier technologie ID/BI pour leurs affaires* »

Problèmes clés de l'ID/BI

- **Modèles de BD opérationnelles complexes et inutilisables :**
 - souvent **difficiles à comprendre**
 - ne concerne **pas un objectif unique** d'affaire
- **Données des BD opérationnelles :**
 - **identiques** dans **différentes BD**
 - **même concept souvent défini différemment**
 - **adaptées pour les systèmes opérationnels** (comptabilité, facturation, ...), pas pour l'analyse des fonctions d'affaires
 - **de qualité mauvaise** : données manquantes, données imprécises, ...
 - **volatiles** :
 - elles sont supprimées périodiquement dans les systèmes opérationnels (6 mois)
 - la modification des données au fil du temps - aucune information historique

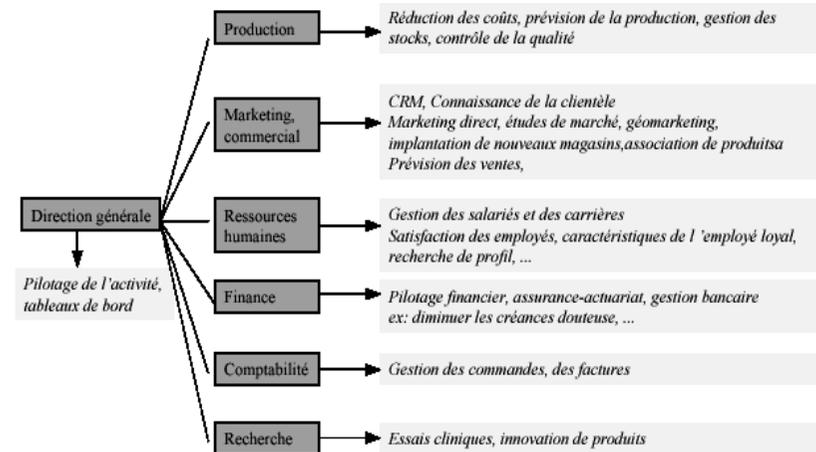
Informatique décisionnelle (ID) : nouveaux termes

Termes anglais	Termes français
Business Intelligence (BI)	<i>Informatique Décisionnelle (ID)</i>
Decision Support Systems (DSS)	<i>Systèmes d'aide à la décision (SIAD)</i>
Competitive Intelligence (CI)	<i>Intelligence Economique (IE)</i>
Data Warehouse (DW)	<i>Entrepôt de données (ED)</i>
On-Line Analytical Processing (OLAP)	<i>Analyse en ligne de données</i>
Knowledge Discovery in databases (KDD)	<i>Extraction de Connaissances dans les données</i>
Data Mining (DM)	<i>fouille de données, orpaillage</i>
Customer Relationship Management (CRM)	<i>Gestion de la Relation Client</i>

Domaines d'application de l'ID/BI

- **Domaines concernés :**
 - commercial, financier, transport, télécommunications, santé, services, ...
- **Applications :**
 - gestion de la relation client
 - gestion de commandes, de stocks
 - prévisions de ventes
 - définition de profil utilisateur
 - analyse de transactions bancaires
 - détection de fraudes
 - ...

L'ID/BI dans l'entreprise



2 – Des SI opérationnels aux SI décisionnels

- **Systèmes d'information opérationnels : traitement en ligne des données (OLTP)**
- **Nouveaux besoins : du traitement en ligne des données à leur analyse en ligne (OLAP)**
- **Des SI Opérationnels aux SI Décisionnels**

Systèmes d'information opérationnels : OLTP

- **Supportent en général une ou plusieurs grandes fonctions de l'entreprise** (production, marketing, commercial, ressources humaines, finance, comptabilité, recherche, ...)
- Parfois intégrés dans un **ERP**, ils s'appuient sur des **SGBD traditionnels** (Oracle, DB2, ...) pour gérer des **BD « opérationnelles »** ou de « **production** » (Mega-Giga octets)
- Permettent des processus de **traitement en ligne des données – OLTP (On line Transactionnal Processing)** : Interactifs, Concurrents, Nombreux, Répétitifs, Structurés, Simples
- **Ces processus OLTP concernent :**
 - La mise à jour de données
 - Un nombre restreint d'enregistrements
 - Des données précises et à jour

Exemple : un supermarché ENREGISTRANT ses ventes

Nouveaux besoins (1)

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 :

- *Quels sont les résultats des ventes par gamme de produit et par région pour l'année dernière ?*
- *Quelle est l'évolution des chiffres d'affaires par type de magasin et par période ?*
- *Comment qualifier les acheteurs de mon produit X ?*

Nouvelles attentes des SI :

- Considérer des **quantités de données historisées de plus en plus importantes** (Tera, Penta octets), **organisées selon différentes dimensions** (temps, espace géographique, gammes de produit, ...)
- Passer du **traitement en ligne des données (OLTP) à l'analyse en ligne de ces données (On Line Analysis Processing - OLAP)** selon différentes dimensions **pour procéder à des analyses** de ces données pour construire des indicateurs indispensables au pilotage de l'entreprise

Nouveaux besoins (2)

L'informatique opérationnelle ne peut satisfaire ces besoins :

- **BD opérationnelles trop complexes** pour être appréhendées facilement par tout utilisateur décideur
- **SI opérationnel ne peut être interrompu** pour répondre à des questions nécessitant des calculs importants
- **Processus d'entreposage des données mal adapté**
- **Analyse en ligne des données très limitée**

=> recours à l'ID pour l'élaboration de SI Décisionnels

3 – Les outils de l'informatique décisionnelle

- **Entrepôt de données**
- **Analyse en ligne (OLAP)**
- **Fouille de données**
- **Architecture multi-niveaux**

Nouvelle technologie informatique

Entrepôt de données (Data Warehouse) :

- Il récolte, **stocke et gère efficace des gros volumes de données** pour la prise de décision
- Il assure un **regroupement homogène et exploitable** de données hétérogènes, très nombreuses et distribuées

Analyse en ligne des données OLAP (On Line Analytical processing) :

- données **organisées** selon plusieurs **axes d'analyse** et selon différents **niveaux de détail**
- traitements permettant **interactivement** de **changer de points de vue**, de **niveau de détail**, et effectuer les opérations OLTP classiques

Fouille de données (Data Mining):

- **extraction automatique de connaissances** (propriétés cachées) dans de grands volumes de données :
- connaissances **valides, nouvelles, compréhensibles, pertinentes et implicites**

Entrepôt de données (Data Warehouse)

Définition de Inmon (1992) :

« une **collection de données thématiques, intégrées, non volatiles et historisées, organisées** pour le support d'un processus d'aide à la décision»

Données :

- **thématiques** : données pertinentes pour un **sujet** ou **thème** et nécessaire aux **besoins d'analyse**
- **intégrées** : données résultant de l'**intégration de données** provenant de **différentes sources** pouvant être **hétérogènes**
- **historisées** : données représentent l'**activité d'une entreprise durant une certaine période** (plusieurs années)
- **non-volatiles** : données **essentiellement utilisées en interrogation** (consultation) et ne peuvent pas être modifiées

OLAP (On Line Analytical processing)

Caractéristiques des processus d'analyse en ligne – OLAP :

- Interactifs
- Concurrents
- Peu nombreux
- Non prévisibles
- Complexes

Les processus OLAP concernent :

- L'exploitation des données
- Un nombre d'enregistrement très important
- Des données consolidées et synthétiques
- Exploration et analyse de données historiques
- ...

Exemple : un supermarché ANALYSANT l'ensemble de ses ventes

OLTP versus OLAP

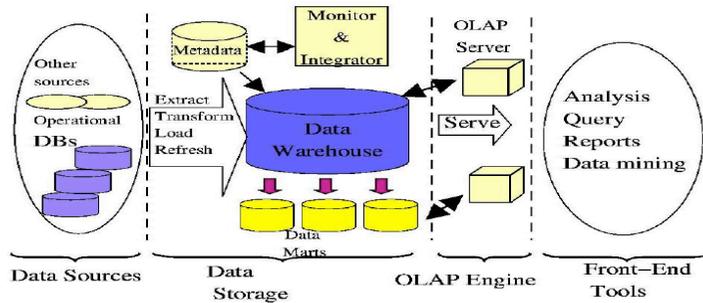
	Caractéristiques	OLTP	OLAP
Conception	Orientation Conception	Transaction Entité-Relation	Analyse Star/snowflake
Données	Granularité	Détail	Résumées, agrégées
	Nature Actualisation	Relationnelle Actualisées, mises à jour	Multidimensionnelle Historisées, recalculées
	Taille	100 Mo/Go	100 Go/To
Traitements	Unité de travail	Transaction simple	Requête complexe
	Accès	Lecture/écriture	Lecture
	Nb de tuples accédés	Dizaines	Millions
	Métrique	Débit de transactions	Temps de réponse
Utilisateurs	Utilisateur	Agent opérationnel	Analyste/décideur
	Nombre d'utilisateurs	Milliers	Centaines

Fouille de données (Data Mining)

- **recherche de la connaissance** sous forme de modèles de comportement caché dans les données : **Extraction de connaissance à partir de données** (Knowledge Discovery in Database - KDD)
- domaine jeune à l'intersection de l'**Intelligence artificielle (IA)**, les **statistique**, l'**analyse de données**, et les **BD**
- **techniques de fouille traditionnelles** : issues des **statistiques** et de l'**analyse de données**
- **nouvelles techniques de fouille** : l'induction d'arbre de décision, les algorithmes génétiques, les réseaux de neurones, ...
- tendance à une **intégration croissante de ces techniques de fouille** dans les **entrepôts de données** (notamment dans l'offre commerciale)

Architecture multi-niveau

0	1	2	3
Sources externes	Entrepôt	OLAP	Front-ends
BD	SGBDR	BD Relationnelle	tableur
Fichiers HTML		BD	Data Mining (fouille de données)
Fichiers plats		Multidimensionnelle	Outils statistiques
...		Combinaison des 2	...

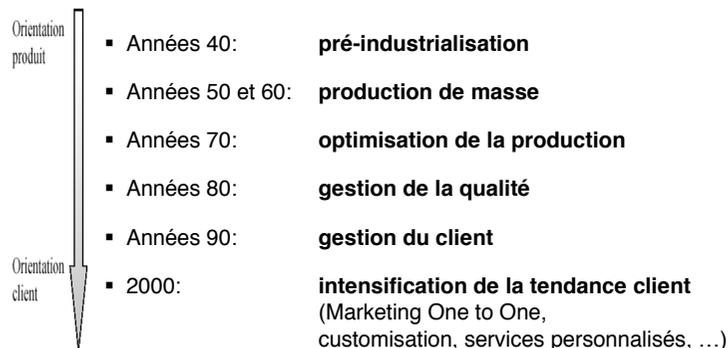


4 – Un exemple de l'ID dans l'entreprise : la relation client

- Contexte économique général
- la Gestion de la Relation Client
- ID/BI pour la Gestion de la Relation Client

Contexte économique général (1)

Evolution historique générale :



Contexte économique général (2)

Dans un climat de concurrence mondiale tendu :

- Conquérir un nouveau client coûte 5 fois plus cher que de fidéliser un client existant
- 5% d'amélioration de la fidélité des clients entraîne une augmentation des profits de 10 à 15%
- Tous les clients ne sont pas égaux : 30% des clients génèrent 70% du CA
- Le Client attend un service personnalisé, sur-mesure
- La personnalisation est une source de profit

La Gestion de la Relation Client (1)

- **Gestion de la relation client** (Customer Relationship Management – CRM) : démarche émergente dans les domaines du **marketing** et du **commercial**
 - **Afin de répondre aux questions :**
 - *Quels sont les besoins et les attentes des clients? Comment y répondre?*
 - *Quels sont les clients prêts à acheter de nouveaux produits?*
 - *Quels sont les clients les plus profitables, fidèles et pourquoi?*
 - *Quels sont les clients mécontents, et pourquoi?*
 - **Besoin de mieux connaître et comprendre ses clients pour :**
 - Réduire les coûts (clients infidèles)
 - Comment augmenter les profits ?
 - Comment fidéliser une clientèle?
 - Comment identifier les nouvelles opportunités?
 - Les concurrents s'attaquent à certains de nos clients?
- => Une réponse par l'informatique décisionnelle**

La Gestion de la Relation Client (2)

Définition du CRM :

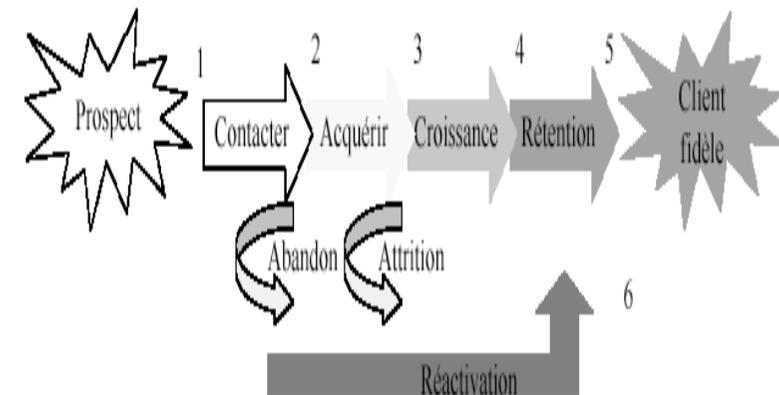
- capacité à **identifier**, à **acquérir** et à **fidéliser** les meilleurs clients dans le **but d'augmenter le chiffre d'affaires et les bénéfices**.
- capacité à **bâtir une relation profitable sur le long terme avec les meilleurs clients** en capitalisant sur l'ensemble des points de contacts
- Principales dimensions de la Gestion de la Relation Client :
 - **stratégique**,
 - **temporelle** et
 - **géographique**

La Gestion de la Relation Client (3)

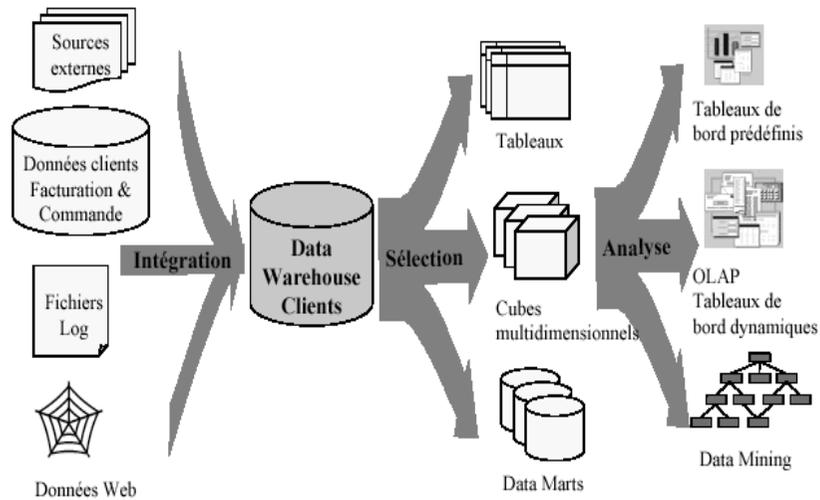
Enjeux du CRM :

- Connaître ses clients rentables et se concentrer sur les clients à fort potentiel
- Augmenter la fidélité des clients
- Comprendre le comportement de ses clients
- Augmenter le panier moyen
- Diminuer les coûts des programmes marketing avec un meilleur ciblage des meilleurs clients
- Adapter les produits au marché
- Aider les commerciaux à vendre
- Rentabiliser la BD Clients
- Mesurer les résultats des actions CRM
- ...

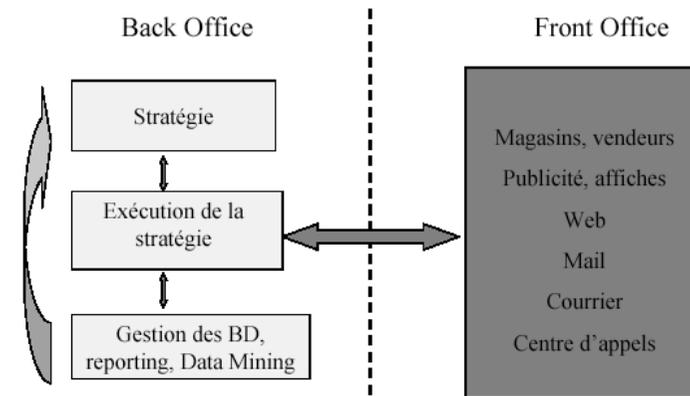
La Gestion de la Relation Client (4)



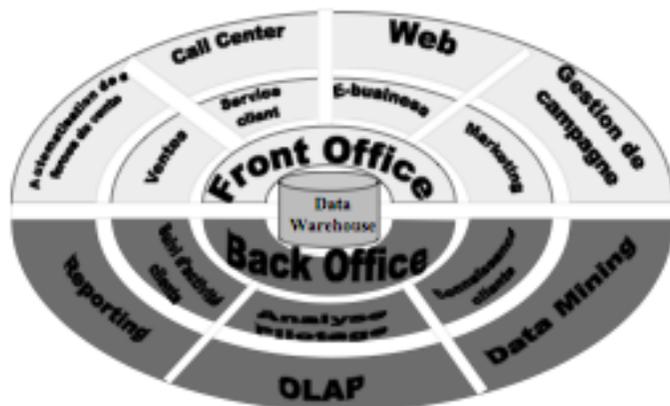
Informatique Décisionnelle pour la Gestion de la Relation Client



Informatique Décisionnelle pour la Gestion de la Relation Client



Informatique Décisionnelle pour la Gestion de la Relation Client



Gestion de la Relation Client et Web

Coût du contact Client sur le Web très faible :

- **Web** = un point de contact important
- En 2001, 1/3 des contacts clients provenaient d'Internet
- En 2003, Internet et les applications de commerce électronique génèrent 60% du volume de contact des centres d'appels

Intégration du Web dans le CRM :

- e-CRM
- Web Mining
-