

# Technologies de l'information et décision : l'essor des Datawarehouses

**Jean-Fabrice LEBRATY**

Maitre de Conférences à l'IUT de Nice  
Département Techniques de Commercialisation

---

## RÉSUMÉ

---

*L'objet de cet article est d'analyser un système, fédérant méthodologies conceptuelles, matériels et logiciels, et visant à améliorer les capacités décisionnelles d'une organisation. Après avoir insisté, sur l'importance de la prise en compte des rapports entre données, informations et connaissances et sur les contraintes en résultant, le concept de Datawarehouse est analysé dans sa nature, sa structure et son environnement. Un exemple illustrant les potentialités mais aussi les difficultés inhérentes à la mise en œuvre d'un Datawarehouse est proposé. Les perspectives et limites de ce type de système sont enfin évoquées en conclusion.*

**Mots-clés** : Technologies de l'information et de la communication, Systèmes d'information, Aide à la décision, Datawarehouse, Datamart.

## ABSTRACT

---

*The purpose of this article is to analyse a system which groups conceptual methodologies, hardware, software and which supplies decision processes. After insisting on the consequences of the process which goes from data to information and finally to knowledge, we analyse Datawarehouses on their nature, structure and environment. Then we give an example of an implementation. At the end, we examine near future prospects of Datawarehouse.*

**Key-words** : Information and communication technologies, Information systems, Decision support Systems, Datawarehouse, Datamart.

Nous nous trouvons actuellement au seuil d'une nouvelle grande évolution dans les rapports : Technologies de l'Information et de la Communication<sup>(1)</sup> (TIC) - Organisation. Depuis le début des années 1990, les différents constituants de la notion de réseau (conceptuels, technologiques et comportementaux) sont parvenus à un niveau de maturité (prix, potentialités, caractéristiques d'emploi) permettant leur utilisation à grande échelle. Deux principes fondamentaux du concept de réseau exercent une influence sur l'ensemble de l'organisation, notamment le domaine des Systèmes d'Information (SI). Le premier principe réside dans la fonctionnalité de circulation permettant de rendre disponible, en temps réel, un volume quasi illimité de données, d'informations et de connaissances pour l'ensemble des acteurs de l'organisation. Le second principe est issu de la capacité offerte aux administrateurs de réseaux de choisir avec précision le degré de centralisation ou de décentralisation des données, des applications et des moyens de traitement. Ainsi, ces profonds bouleversements dans la gestion des flux immatériels ont amené les organisations à repenser le management des processus de création d'informations et de connaissances dans une optique décisionnelle, l'attention se portant d'abord sur la gestion des données, éléments de base dans la constitution d'informations et de connaissances.

Cependant, comme sur tout marché émergent, il existe dans le

monde des réseaux une profusion d'outils dont les champs d'actions se superposent et dont, bien souvent, les caractéristiques se révèlent incompatibles. L'objectif premier de cet article est de présenter un système cohérent intégrant des méthodologies, des matériels et des outils et ayant pour but de gérer le processus qui va de l'extraction des données jusqu'à la prise de décision, en passant par la création d'informations et de connaissances. Ce système est communément appelé **DATAWAREHOUSE (DW)** que l'on peut traduire par le terme **ENTREPÔT DE DONNÉES**.

L'intérêt pour ce type de système nous est apparu au cours de recherches sur les applications d'aide à la décision fondées sur l'apprentissage. Nous avons déterminé en effet que tout apprentissage impliquait des contraintes fortes sur les caractéristiques des données disponibles dans l'entreprise. Ainsi, par exemple, est-il avéré que l'hétérogénéité des méthodologies, lieux et moyens de stockage exerce un impact négatif sur le traitement des données, compliquant ainsi la phase d'apprentissage et limitant donc l'implémentation de telles applications (Lebraty, 1997a).

Dès lors, un système cohérent tel que le Datawarehouse facilite le développement d'outils d'aide à la décision modernes, notamment quand ils font intervenir une fonction d'apprentissage. L'impact sur les processus de décision se traduit pour l'entrepreneur par un retour sur investissement mesurable. Dans une étude récente me-

(1) Nous définissons les TIC comme l'ensemble des technologies permettant d'effectuer des opérations (stockage, traitement, circulation) sur les données, les informations et les connaissances.

née aux USA<sup>(2)</sup>, le cabinet IDC estime un ROI moyen compris entre 300 % et 400 % selon les cas, le délai de récupération du capital investi variant, quant à lui, de 1 à 3 ans. L'argument du ROI, d'ailleurs habilement mis en avant par les différents promoteurs de ces solutions a conduit 40 % des entreprises américaines à démarrer un projet de Datawarehouse et 30 % à en envisager sa mise en place pour les trois prochaines années<sup>(3)</sup>. Le pourcentage cumulé monte à 95 % si l'on ne prend en compte que les 500 premières entreprises américaines<sup>(4)</sup>. En France, une étude qui vient d'être publiée<sup>(5)</sup> fournit des précisions sur la prise en compte de ce concept par les entreprises. Il en résulte que 11 % des grands comptes ont implémenté de telles applications, 9 % ont mis en œuvre une application pilote et 35 % ont un tel projet, en fait, seuls 8 % des répondants ne se prononcent pas sur le déploiement en leur sein d'un système de Datawarehouse.

Dans cet article, nous nous proposons, dans un premier temps, de préciser le concept de Datawarehouse. Dans la section II, nous analyserons avec précision l'architecture d'un système de Datawarehouse afin de pouvoir en souligner les apports au niveau décisionnel. Ensuite, nous présenterons synthétiquement un exemple d'implémentation. Enfin, nous tenterons de décrire les perspectives

d'évolution à moyen terme du Datawarehouse.

## **I. LE DATAWAREHOUSE : UN SYSTÈME INTÉGRÉ**

### **1. Définitions et principales caractéristiques d'un Datawarehouse**

Dès l'instant où un décideur travaille sur des données numérisées, la question de l'amélioration de l'efficacité et de l'efficience des moyens technologiques d'aide à la décision se pose.

Depuis 5 ans, un concept fédérateur ne cesse de s'étendre dans certaines entreprises : le Datawarehouse. Selon Inmon<sup>(6)</sup> souvent présenté comme l'un des créateurs de ce concept : *"Un système de DW organise et conserve les données nécessaires aux processus informationnels et analytiques dans une perspective de long terme. Ce système correspond à un ensemble de données orientées selon un sujet, intégrées, évoluant dans le temps et non volatiles, qui a pour but l'aide aux processus de prise de décision de gestion."* (Inmon, 1996).

Concrètement, un système de DW est constitué<sup>(7)</sup> d'un ensemble de méthodes, de matériels et de logiciels regroupés autour de trois grandes fonctions classiques dans le domaine des systèmes de gestion de bases de données, à savoir :

(2) (IDC, 1996).

(3) Statistiques citées dans (01Info 1997), p. 35.

(4) (Franco, 1997), p. 19.

(5) (IDC, 1997).

(6) Précisons tout de même que W. Inmon apparaît, certes, comme le théoricien de ce concept, mais qu'il est aussi le vice-président de PRISM SOLUTIONS un des principaux acteurs du marché du Datawarehouse.

(7) Voir le dossier paru dans (Infomag, 1996), p. 82-101.

- l'acquisition ;
- le stockage ;
- le traitement des données.

La définition d'Inmon met en lumière plusieurs aspects novateurs d'un système de Datawarehouse.

En premier lieu, un concept technologique de large champ prend en compte, pour la première fois, les rapports "données - informations". Traditionnellement dans les entreprises, les données sont uniquement stockées en fonction des applications qui les ont créées. Par exemple, les données issues des applications de comptabilité sont stockées à un endroit particulier, celles issues des applications dédiées au service commercial le sont ailleurs et sans véritable lien avec les précédentes. Ensuite, il appartient à chaque service de construire ses propres informations à partir de ses propres données. Or, parfois les informations peuvent porter sur un même thème. Ainsi, le service financier et le service commercial peuvent générer des informations sur le client. A ce stade, une confusion ne peut manquer d'apparaître entre la donnée unique et théoriquement objective (le client possède telle et telle caractéristique) et ses interprétations multiples et variables selon les décideurs (c'est un client qui paye avec du retard, donc problématique au niveau comptable, mais qui commande en grande quantité donc intéressant pour les commerciaux). La dissémination de données, à la base recouvrant les mêmes réalités, fait donc apparaître un risque pour l'organisation : la création, à terme, d'incohérences. Ces dernières résulteront

notamment des inévitables mises à jour apportées par les différents utilisateurs. Si, par exemple, le service financier décide unilatéralement de modifier le fichier client en ajoutant le champ "client à risque", quel sera, alors, le fichier client de référence ? Le service commercial sera-t-il averti de ne plus faire d'offres à tel ou tel client ? La solution à ce type de situation réside dans la structure même du DW qui implique que le stockage des données soit unique et que ce soient les applications de traitement et d'analyse qui permettent de transformer des données en informations personnalisées pour les utilisateurs. Un DW constitue donc un moyen essentiel de concilier des visions disparates et les objectifs fondamentaux de l'organisation<sup>(8)</sup>. Il permet en outre, aux décideurs de raisonner de manière cohérente, à la fois, sur des éléments opérationnels (des données très détaillées) et sur des perspectives plus stratégiques (des agrégats constitués par les données opérationnelles)<sup>(9)</sup>.

En second lieu, au niveau de la structure des données et de l'identification des entités, le DW se fonde sur les sujets en rapport avec le métier de l'entreprise (client, produit etc.) et non sur ses fonctions (comptabilité, vente, etc.). La démarche est très schématiquement la suivante : on définit d'abord la structure des données puis, on conçoit les applications qui vont se servir de ces dernières. Cette façon de faire permet de mieux appréhender la réalité environnementale, l'ensemble des données constitutives d'un sujet étant transversal aux struc-

(8) (IDC, 1996), p. 10.

(9) (IDC, 1996), p. 11

tures fonctionnelles d'une organisation. Par ailleurs, cette "orientation sujet" autorise une personnalisation des applications. Dans l'ensemble, les utilisateurs ont à la fois une vision plus précise de leur environnement et des outils mieux adaptés à leur travail. Comme nous le verrons, la phase d'analyse visant à construire une telle structure est délicate et conditionne le projet dans son ensemble.

En troisième lieu, dans un DW, les données sont accumulées et non remplacées par de nouvelles comme dans un classique processus de mise à jour. Ainsi, le montant des ventes de dentifrice d'une marque donnée, pour le mois de janvier, ne sera pas remplacé par celui du mois de février, les deux valeurs seront conservées. Par rapport aux structures traditionnelles, le DW prend en compte une dimension "temps". Ceci entraîne d'ailleurs le fait que la capacité de stockage nécessaire soit importante, elle se compte désormais en Téra Octets<sup>(10)</sup> (To). Le principal intérêt de "l'historisation" des données est que l'on voit se constituer une véritable mémoire de l'organisation où une grande partie des événements est conservée, structurée, interrogeable et manipulable à tout instant. Ce stock ouvre la voie à toute une série d'outils dont le fonctionnement est fondé sur l'expérience comme le sont par exemple, les réseaux de neurones.

Enfin, le Datawarehouse constitue un système technologique to-

talement intégré à la stratégie d'entreprise. D'une part, la mise en œuvre d'un tel système s'effectue dans le cadre d'actions stratégiques organisées comme, par exemple, le souhait d'orienter les gammes produits de l'entreprise en fonction des comportements du client en améliorant, en même temps, la coordination interne entre les différents services. D'autre part, en mettant en lumière certaines informations, ce système peut exercer une influence sur les décisions stratégiques futures. On voit donc apparaître une forte interaction entre stratégie d'entreprise et Datawarehouse.

## 2. Aide à la décision et Datawarehouse

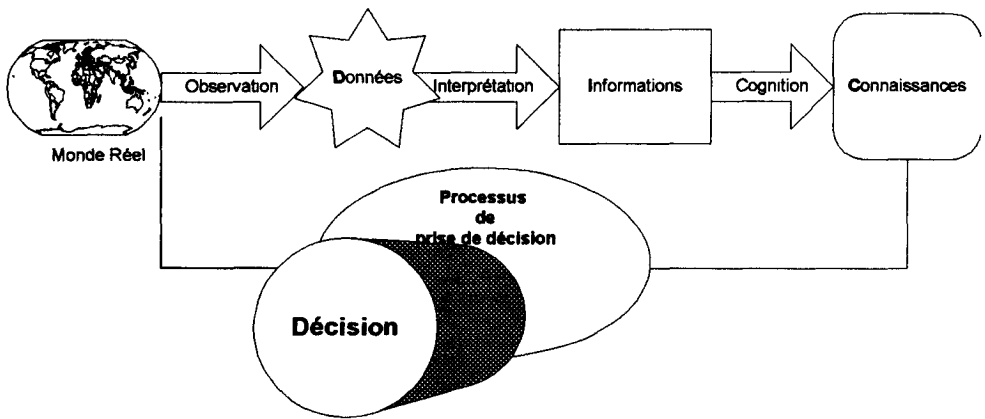
Les recherches et approches relatives aux processus de prise de décision sont nombreuses et variées. On peut, cependant, déceler certaines convergences, notamment au niveau du rôle central qu'occupent données, informations et connaissances. L'existence d'un continuum menant des données aux connaissances nous paraît une hypothèse raisonnable que l'on peut illustrer de la manière suivante<sup>(11)</sup> :

Si l'on admet une telle représentation, on mesure aussitôt l'importance des données sur la prise de décision : elles en constituent le fondement originel (a). Quelles propriétés doivent alors caractériser les données pour faciliter le déroulement du processus décisionnel<sup>(12)</sup> ?

(10) Par exemple, le DW de WALLMART aux USA est en train de passer de 7,5 To à 24 To, et l'on évoque le Péta Octet ( $10^{15}$  octets) pour l'an 2000.

(11) D'après (Lebraty, 1997b).

(12) Afin de déterminer ces contraintes, nous nous sommes inspirés de nos recherches sur l'information dans : (Lebraty, 1994) p. 31-34.



**Figure 1 : Le processus D-I-C**

Tout d'abord, au niveau de la phase d'observation, nous déterminons deux types de caractéristiques :

- **Accessibilité** : c'est-à-dire, disponibilité à tout moment en quantité suffisante pour le décideur. La cohérence dans le stockage et dans la manière dont circulent les données est alors primordiale.

- **Valorisation** : exigence de qualité (exactes et utiles), de globalité (possibilité d'avoir accès à un ensemble de données en une même transaction) et de présentation (ergonomie appropriée).

Si l'on reporte ces caractéristiques et contraintes dans un tableau, on obtient :

**Tableau 1 : Caractéristiques et contraintes des données liées à la phase d'observation**

Caractéristiques	n°	Exemples de contraintes
<b>Accessibilité</b>		
Quantité	C1	Prise en compte de l'ensemble des données existant en interne et capacité d'ouverture sur l'extérieur
	C2	Capacité d'ajouter des nouveaux champs de données dans les bases existantes
Délai de disponibilité	C3	Les données doivent être disponibles en temps réel
<b>Valeur</b>		
Qualité	C4	Exactitude : Un système ou une procédure de contrôle des données doit exister
	C5	Utilité : Les données doivent servir au décideur
Globalité	C6	L'ensemble des données concernant un sujet doit pouvoir être disponible d'un bloc
Présentation	C7	Emploi d'outil de mise en valeur des données

En poursuivant le processus menant à la décision et en nous arrêtant à la phase d'interprétation des données (création d'information) des contraintes supplémentaires apparaissent :

- il doit être possible de manipuler toutes les données nécessaires à un traitement, au travers d'une interface unique ;

- les applications destinées aux utilisateurs doivent permettre la

personnalisation des recherches. En effet, à partir des mêmes données, les utilisateurs créeront des informations qui leur seront propres. Il va ainsi s'établir des habitudes de travail. L'outil doit donc offrir des fonctions de personnalisation, en rapport avec la tâche qu'effectue l'utilisateur et en tenant compte de ses caractéristiques personnelles (visualisation des données en colonne ou en ligne, par exemple).

**Tableau 2 : Caractéristiques et contraintes des données liées à la phase d'interprétation**

Caractéristiques	n° de Contrainte	Exemple de contraintes
Standardisation des formats	C8	Les données doivent être dans un format compatible avec l'ensemble des applications de traitement
Personnalisation	C9	Les outils doivent inclure des options de personnalisation

Enfin, la phase de cognition implique certaines caractéristiques génératrices de contraintes :

- la création de connaissances via un processus fondé sur l'expérience impose de raisonner en fonction des événements passés. Il devient ainsi obligatoire de conserver un historique des données

dans le temps et un historique des décisions déjà prises ;

- devant le volume illimité des données historisées, des outils doivent permettre de faciliter la mise en valeur des corrélations cachées entre groupes de données.

**Tableau 3 : Caractéristiques et contraintes des données liées à la phase de cognition**

Caractéristiques	n° de Contrainte	Exemple de contraintes
Historisation des données	C10	Les données doivent être historisées et non uniquement mises à jour
Historisation des décisions	C11	Les décisions prises doivent être intégrées dans le système d'information
Détection des corrélations cachées	C12	Les outils doivent faciliter la découverte de formes dans l'ensemble des données

L'ensemble des contraintes décrites permet de mieux cerner le domaine des solutions possibles dans le déroulement souhaité du

processus décisionnel. Répondre à ces contraintes garantit une base solide de construction d'applications efficaces et spécifiques d'aide

à la décision ; la prise en compte de toutes ces exigences constitue donc un facteur clé de succès de l'aide à la décision.

Voyons maintenant, plus précisément en quoi un système de Datawarehouse représente justement une solution intégrée à la mise en œuvre d'applications décisionnelles de qualité.

## II. ARCHITECTURE

On a assisté, des années 70 au début des années 90, à la création de nombreux concepts et à la mise en œuvre d'une multitude de produits, tous censés améliorer les processus décisionnels. Dans les années 70, l'attention s'est portée sur la notion de système d'information. Dans ce contexte, de nombreuses méthodes de normalisations des flux de données ont été conçues et appliquées. Dans les années 80, l'utilisateur a occupé une place primordiale et l'objectif a été de lui fournir des outils applicatifs d'aide à la décision. Cependant, plusieurs raisons ont limité l'impact des concepts et produits issus des TIC sur les processus de prise de décision. Rappelons-en au moins trois :

- déconnexion entre système d'information et application décisionnelle ce qui a notamment été préjudiciable aux systèmes à base de connaissance et aux réseaux neuronaux ;

- comportement des utilisateurs face à l'informatique en général. Pour beaucoup, l'informatique apparaissait comme un outil complexe, peu convivial et non exempt d'erreurs ; d'où, la manifestation de réactions de résistance forte aux changements technologiques<sup>(13)</sup> ;

- limites techniques et financières se manifestant par des coûts de stockage élevés, des capacités de traitement limitées et un manque d'évolutivité évident lié à des architectures informatiques fondées sur des modèles propriétaires.

Pourtant, le besoin d'une informatique décisionnelle globale restait fort. Aussi, dès le début des années 80, un essai de solution a été apporté par la mise en œuvre d'infocentres. Mais, notamment pour les raisons précitées, l'infocentre ne constituait que le support de processus décisionnels ponctuels et non d'actions stratégiques<sup>(14)</sup>. La conception d'un infocentre n'était d'ailleurs en rien motivée par des objectifs stratégiques.

Aujourd'hui, le concept opérationnel de réseau, le comportement des utilisateurs et la continue volonté d'améliorer les capacités décisionnelles ont eu comme conséquence de faire évoluer le concept d'infocentre vers celui de Datawarehouse. Ce concept, comme nous allons le voir, a pour objectif l'aide à un large

(13) Le thème des difficultés de l'intégration des TIC dans une organisation est malheureusement, dans certaines entreprises ou administrations, toujours d'actualité. La recherche des causes constitue un champ d'étude dont la portée est trop souvent sous-estimée.

(14) Les données brutes provenant directement de certaines bases de production, étaient mises à jour et les utilisateurs se servaient d'applications généralistes (tableur, etc.). Ainsi, il était, par exemple, possible de déterminer le nombre de produits vendus pour le mois en cours.

champ de décisions, notamment l'aide aux décisions stratégiques.

## 1. La conception d'un Datawarehouse

Afin de mieux décrire l'architecture d'un Datawarehouse, il convient dans un premier temps de préciser les quatre étapes menant à sa mise en œuvre opérationnelle.

### a. Analyse

La première étape est centrée sur l'analyse de l'existant, sa modélisation et la prise en compte des développements futurs. Soulignons que le degré de cohérence de l'analyse conditionnera la qualité du projet dans son ensemble.

- Un projet de DW répond d'abord à un objectif de portée stratégique, tel que par exemple, l'augmentation des ventes, ou la fidélisation de la clientèle. Cet aspect est important car il confère au DW le statut d'outil au service de l'ensemble de l'organisation et non celui d'applications ponctuelles destinées aux cadres de tel ou tel service. L'orientation collective du système constitue de notre point de vue un atout face aux inévitables mécanismes de résistance au changement que les concepteurs rencontrent inévitablement.

- Un projet de DW repose ensuite sur une modélisation de l'activité de l'organisation. Les concepteurs devront alors déterminer les données pertinentes par rapport à cette modélisation. Cette tâche est extrêmement délicate. En effet, les données peuvent être présentes dans les bases de pro-

duction de l'entreprise ce qui est l'hypothèse la plus fréquente. Mais dans le cas contraire, il faut les récupérer ou les acquérir ; cette opération implique la mise en œuvre de moyens techniques supplémentaires générateurs de coûts (par exemple, le droit d'accès à certaines banques de données) et de moyens organisationnels (par exemple, la création d'un service de veille technologique). Ajoutons que certaines données peuvent ne pas être accessibles. Dans ce cas, des aménagements conceptuels seront à opérer.

Le fait que ce ne soit qu'un groupe<sup>(15)</sup> qui détermine les données pertinentes pour l'ensemble des utilisateurs ne répond qu'en partie à la contrainte C5. Il devient primordial, alors, que le groupe chargé de cette tâche ait l'aptitude la meilleure possible à formaliser les besoins existants et, ce qui est beaucoup plus difficile, à imaginer leurs évolutions. La qualité du système dans son ensemble sera largement fonction de cette aptitude à la révélation des besoins.

Un projet de DW exige enfin, la détermination des caractéristiques d'infrastructures techniques (matériel et logiciel) et les moyens organisationnels qui seront alloués. Ce dernier aspect peut donner lieu à de nombreuses études. En effet, que ce soit pour la mise en œuvre, la maintenance ou le développement d'un DW, on voit apparaître tout un ensemble de nouvelles activités (responsable architecture infrastructure, administrateur de données ou chef de projet

(15) Ce groupe peut être constitué d'utilisateurs ou de personnels extérieurs, notamment dans le cas de l'externalisation de la conception du Datawarehouse.

Datawarehouse<sup>(16)</sup> dont les impacts sur l'organisation sont loin d'être neutres.

### **b. Acquisition**

La phase d'acquisition<sup>(17)</sup> d'un DW consiste à préparer et à faire migrer les données issues des bases opérationnelles ou de bases de données externes, en vue de leur stockage.

Cette phase constitue, dans de nombreux cas, un moment très délicat. En effet, la cohérence dans l'alimentation du DW conditionnera, ensuite, le fonctionnement de l'ensemble des outils de traitement et donc influera sur l'efficacité globale du système d'aide à la décision. Cette phase est à ce point importante que, selon Inmon, 80 % des efforts déployés pour construire un DW lui sont alloués. Ce pourcentage peut expliquer le développement des Datamarts que nous évoquerons dans le II.3.

La phase d'acquisition est composée de plusieurs opérations :

- **Identification des données.** Il s'agit de repérer les données pertinentes précédemment définies. Au niveau de l'identification, il y a deux cas de figure, soit les données ne sont pas présentes dans une des bases de production. Dans ce cas, il faut mettre en place une procédure de récupération ou d'acquisition. L'extension des systèmes d'information via la normalisation engendrée par le protocole TCP/IP, d'une part augmente le volume de données récupérables, tout en facilitant, d'autre part, leur récupération (contrainte

**C1).** Dans cette optique, des passerelles vers l'Internet sont mises en œuvre (Varney, 1996a). En revanche, si les données sont déjà présentes dans l'entreprise, il faut les localiser et vérifier leur validité. Par exemple, il est fréquent que pour une même entité, plusieurs fichiers existent ; dans ce cas, quel sera le meilleur, le plus à jour, celui dont le format est le mieux standardisé ? Faut-il plutôt prendre des champs dans plusieurs fichiers ? Ainsi, dès ce niveau, un premier travail de contrôle des données doit être opéré répondant en cela à la contrainte **C4**.

- **Extraction des données.**

Cette opération est triple : il s'agit de récupérer automatiquement les données sélectionnées dans les bases opérationnelles, de prévoir un certain niveau de flexibilité selon les différents formats des données sources et selon leur volume et enfin, de veiller à ce que cette procédure d'extraction soit renouvelée périodiquement, c'est-à-dire, à chaque fois qu'un nouveau flux de données vient s'ajouter aux données déjà stockées. Notons que l'extraction de données suscite principalement des difficultés d'ordre technique tenant notamment à l'hétérogénéité des formats des sources de données.

- **Nettoyage et transformation des données.** Il s'agit, en premier lieu, de vérifier l'exactitude des données. L'exemple cité dans (Franco, 1997)<sup>(18)</sup> illustre ce travail de vérification : une société de location de véhicules, en analysant les bons résultats qu'elle obtenait dans la région de Boston,

(16) Offres d'emploi parues dans 01 Informatique le 9 janvier 1998.

(17) On emploie aussi le terme "alimenter".

(18) p. 134.

s'est rendu compte qu'en fait, les touristes étrangers louant une voiture à l'aéroport de Boston étaient considérés comme résidant dans cette ville. En effet, le système n'avait pas prévu les codes postaux étrangers mais imposait la saisie d'un tel champ, aussi les opérateurs entraient le code de Boston pour valider ce champ. Certaines données étaient donc fausses.

En second lieu, des opérations de transformation vont être réalisées. Par exemple, le regroupement de différents champs au sein d'une même entité, ou la décomposition d'un champ ("adresse") en plusieurs champs ("N°", "Nom de la rue", "Code postal", etc.). En fait, nous pouvons dégager trois types de données : les données détaillées ("prix", "poids", etc.), les données agrégées (résultant d'une opération de traitement, le chiffre d'affaires, par exemple) et les métadonnées. Cette troisième catégorie est spécifique en ce qu'elle n'est pas du même ordre, au niveau de l'abstraction, que les deux autres. Une métadonnée peut être définie comme de l'information sur une donnée détaillée ou agrégée. Elle qualifie une donnée en précisant, par exemple, sa sémantique, les règles de gestion associées, sa source ou son format (Franco, 1997)<sup>(19)</sup>. On peut dire aussi qu'une métadonnée représente tout ce que les outils technologiques "connaissent" des données et de leur environnement informatique (Darling, 1996). Les opérations de nettoyage et de transformation répondent directement à la contrainte **C4** et sont si importantes qu'elles peuvent entraîner

des modifications de la structure des bases de production, ou nécessiter une étape intermédiaire entre l'extraction et l'alimentation dans la base cible.

### **c. Stockage**

Le mode de stockage dépend étroitement du contexte du Data-warehouse. Ainsi les caractéristiques des données (nombre d'entités, de relations), la capacité d'évolutivité de la base et la nature d'utilisation prévue vont conditionner le choix d'un type de base de données.

Quelles sont alors les caractéristiques du stockage ? Tout d'abord, le stockage est centralisé et constitue l'entrepôt principal des données, lieu de modification de ces dernières. La base est donc volumineuse et pour améliorer sa disponibilité (contrainte **C3**), on a recours à la réplication de certaines parties de la base principale. Mais dans ce cas, les données répliquées sont en "lecture seule" ce qui évite les incohérences lors de mises à jour concurrentes. Ensuite, toutes les données se situant dans une même base, les problèmes d'incompatibilité entre formats disparaissent (contrainte **C8**). Enfin, dans un DW, il n'y a pas de mise à jour (contrainte **C10**), les données sont surajoutées, ce qui permet notamment de stocker une même requête à plusieurs dates et donc de faire des comparaisons.

### **d. Traitement**

C'est au cours de cette étape que les données stockées vont être transformées en informations. L'interprétation des données en

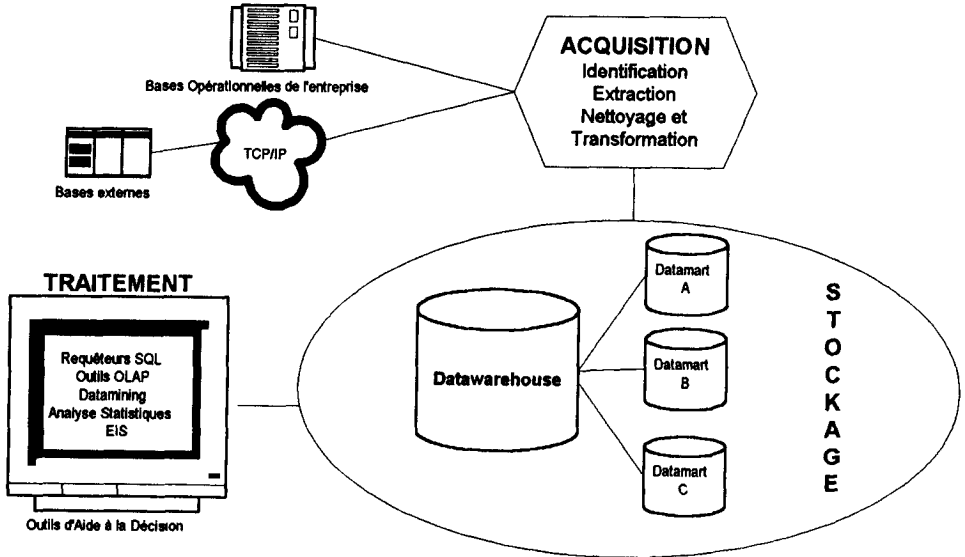
(19) p. 194.

information sera effectuée par l'utilisateur et par l'outil applicatif. Le rôle de ce dernier variant de la simple aide à la présentation jusqu'à la mise en évidence automatique d'informations pertinentes. Afin de mieux assister l'utilisateur, l'ensemble des outils offre la possibilité d'effectuer une double personnalisation, la première en

fonction de la tâche à accomplir et la seconde en fonction de l'utilisateur lui-même. Cet aspect de personnalisation répond à la contrainte C9.

**2. Architecture générale**

Le schéma suivant illustre l'agencement des différents composants d'un système de DW :



**Figure 2 : Les composants du système de Datawarehouse**

Analysons l'architecture d'un DW à trois niveaux (méthodologie, matériel et logiciel) et en fonction des quatre étapes précédemment définies (analyse, acquisition, stockage et traitement).

**a. Méthodologie**

La mise en œuvre d'un DW, outre la méthodologie d'implémentation, fait appel à de nombreuses méthodes issues du monde de l'informatique. Ces méthodes s'ap-

pliquent lors des phases d'analyse, de stockage et de traitement.

- **Phase d'analyse :** Il va s'agir de modéliser l'activité que prendra en compte le système. De nombreuses méthodes de modélisation classiques comme MERISE peuvent être utilisées. Mais la prise en compte de dimensions d'analyse comme le temps, par exemple, exige l'emploi d'autres méthodes de modélisation notamment des méthodes émergentes

comme celle dite "en étoile"<sup>(20)</sup>, ou celle en "flocons"<sup>(21)</sup> dérivée de la précédente. Ces méthodes permettent de mettre en évidence des faits (appelés aussi métriques telles que la marge ou les ventes) et des dimensions (axes d'analyse pertinents comme le temps, les clients ou les produits).

- **Phase de stockage** : La méthode utilisée ici est conditionnée par celle employée lors de la phase d'analyse. Ainsi, le stockage est fondé, actuellement, sur deux grandes approches. En premier lieu, l'approche relationnelle qui découle d'une analyse de type MERISE. Cette approche conduit à mettre en œuvre des SGBD-Relationnels (80 % du marché) dont les avantages sont nombreux, tant au niveau de la normalisation que des capacités d'évolution (C2). En second lieu, les modèles "étoile" ou "flocon" sont plutôt adaptés aux bases multidimensionnelles. Ces dernières commencent à émerger<sup>(22)</sup> du fait de leur orientation résolument décisionnelle. Notons, enfin, la faible part des SGBD Objet dont les potentialités sont grandes mais qui pâtissent de la complexité dans la modélisation et d'une offre logicielle non encore stabilisée.

- **Phase de traitement** : La phase d'interprétation des données en informations utiles à la prise de décision se fonde sur deux approches. La première, classique, a pour objet de faciliter le travail du décideur, en lui permettant notamment d'effectuer des simula-

tions. A cet effet, les bases multidimensionnelles permettent de disposer les données sous forme de cubes dans lesquels l'utilisateur peut naviguer. Dans la seconde approche, on intègre les comportements et objectifs du décideur de manière à ce que ce soit l'application qui crée l'information pertinente ou qui fasse émerger de la connaissance sur un domaine précis. Dans ce cadre, on recourt aux techniques d'intelligence artificielle (approches expertes ou neuronales et algorithmes génétiques).

## **b. Matériel**

A un niveau général, le DW ne s'adresse qu'aux entreprises disposant d'un réseau informatique important<sup>(23)</sup> comportant de nombreux serveurs aux fonctions spécialisées et plusieurs centaines de stations de travail. En revanche, un Datamart, que nous évoquons dans le point suivant, peut être installé sur des architectures plus restreintes, comme par exemple, un gros réseau local.

- **Phase d'analyse** : Les préoccupations à ce niveau sont triples. Tout d'abord, il s'agit de s'assurer que le matériel puisse supporter les applications. Ensuite, il faut prendre en compte l'évolution du système. Enfin, l'attention se portera sur la sécurité des données. Ce dernier problème est difficile, la sécurité ayant un coût dont il n'est pas évident de percevoir le retour (absence de pannes).

(20) Voir à ce sujet (Kimball, 1996) qui explique de manière détaillée la modélisation en étoile et montre son application sur différentes situations.

(21) (EDS, 1997).

(22) Les bases multidimensionnelles sont récentes, comme en témoigne un manque de standardisation dans les concepts ou les technologies.

(23) Généralement un réseau étendu fédérant plusieurs réseaux locaux.

• **Phases d'analyse, de stockage et de traitement** : La capacité physique de stockage influence la phase du même nom, mais aussi les étapes en amont (acquisition) et en aval (traitement). D'une manière générale, le DW nécessite une capacité de stockage élevée (avoisinant le To) de même qu'une puissance de traitement forte, aussi bien du côté client que du côté serveur. Il faut, par exemple, permettre à un utilisateur d'effectuer une simulation à partir de plusieurs dizaines de Go de données, et ce dans des délais restreints.

### c. Logiciel

C'est à ce niveau que l'on trouve le plus de diversité. En effet, la complexité du processus qui mène de l'extraction des données à leur transformation en informations pertinentes, limite le développement de solutions intégratives (une grosse application qui appréhende tout le processus). En revanche, on observe un grand nombre d'outils fortement spécialisés.

• **Phase d'acquisition** : Il existe<sup>(24)</sup> plus de **155** applications qui ont pour mission d'extraire, de nettoyer, de transformer ou d'alimenter la base principale. D'une part, ce grand nombre d'outils témoigne de la diversité des formats des données, des systèmes d'exploitation réseau et des capacités techniques. D'autre part, on constate que la phase d'acquisition peut se segmenter en sous

phases spécifiques interdépendantes. Il faut alors assurer la coordination des différentes applications spécialisées sur chacune des phases. Ajoutons enfin qu'il existe à ce jour **15** outils créateurs de métadonnées. Surveillant la phase d'acquisition et créant des "données sur les données", ils ajoutent un aspect qualitatif aux données stockées ce qui est déterminant dans une optique décisionnelle<sup>(25)</sup>.

• **Phase de stockage** : Il existe environ **15** grands éditeurs de bases de données. Ce nombre relativement faible témoigne du haut niveau de complexité et de sécurité inhérent à tout SGBD. Cependant, plusieurs catégories d'outils viennent compléter la ou les bases principales. Tout d'abord, dans le cas de bases distribuées sur des plates-formes différentes, des applications de "Middleware" (on en dénombre **34**) sont utilisées pour surmonter les problèmes de connectivité en les rendant transparents pour l'utilisateur<sup>(26)</sup>. Ensuite, on note toute une série d'outils facilitant, d'une part la création et le chargement de requêtes (**24**) et d'autre part l'administration du SGBD (**8**). Enfin, des outils (**5**) assistent les concepteurs de SGBD dans leur tâche de modélisation.

• **Phase de traitement** : C'est au cours de cette phase que les données sont interprétées et transformées en informations. Le sens conféré aux données, à l'aide d'un très important nombre d'outils<sup>(27)</sup> est fonction des tâches va-

(24) Au 15/01/1998 (Source Datawarehousing Information Center. Voir sites Internet).

(25) Le décideur peut juger de la fiabilité des données sur lesquelles il s'appuie, par exemple en regardant la source et la date de ces dernières.

(26) (Sing, 1998), p. 5.

(27) Ce volume se traduit au niveau du total des ventes pour 1996 des outils de traitement qui est supérieur à 1 milliard de dollars. (Infomag, 1998), p. 55.

riées à accomplir et des différents types de décideurs (dirigeants, cadres intermédiaires, par exemple). Tous ces outils ont pour objectif d'aider à la prise de décision, et l'orientation "sujet" dans le stockage des données facilite l'obten-

tion d'une vue globale d'un problème (C6), en raison de la transversalité qu'elle entraîne.

Parmi les différents outils disponibles, mentionnons les plus importants.

### **Cf. tableau 4 ci-après**

## **3. Architecture concrète**

Entre l'approche théorique du DW et sa réalisation pratique, il existe un décalage. Firestone (1997) estime que l'implémentation d'un DW est généralement trop difficile, trop chère, trop lourde de conséquences et requiert une période de développement trop longue pour de nombreuses organisations. L'idée de l'entrepôt de données demeure néanmoins séduisante. Les éditeurs de solutions informatiques ont donc mis en œuvre le concept de DATAMART (MAGASIN DE DONNÉES). Il s'agit en fait d'entrepôts de données dédiés uniquement à une activité, à une zone géographique ou à un département de l'entreprise (Varney, 1996b). On retrouve, ainsi, dans un Datamart, tous les éléments constitutifs d'un Datawarehouse, mais la limitation du champ des données à prendre en compte simplifie énormément les aspects de modélisation de l'activité ou certains problèmes techniques. De la même manière, les investissements en termes de coût, de temps de mise en place et de formation des utilisateurs sont nettement plus faibles pour un Datamart. Cette réduction du champ implique enfin une différence au niveau de la finalité de l'outil. En effet, un Datamart a alors pour objectif de répondre à une question précise telle que par exemple l'impact d'une promotion sur une

gamme particulière de produits. Les retours en termes de réponses aux questions sont donc plus facilement contrôlables et visibles.

Pour toutes ces raisons, les Datamarts ont émergé comme une alternative au Datawarehouse en réintroduisant un aspect de décentralisation. Cependant, très vite le danger de voir, dans une même entreprise, plusieurs datamarts indépendants et contenant des champs redondants, a conduit à envisager une complémentarité entre DW et Datamarts. Cette complémentarité s'exprime par le développement parallèle des deux types de systèmes. Il s'agit, alors, d'imaginer un modèle global de données qui constituera la référence pour le DW et les datamarts, ces derniers étant conçus dans l'optique d'une intégration, à moyen terme, dans un DW. Cette implémentation nécessite de nombreux "feed-back". En effet, pendant la phase de développement, l'environnement du système évolue : nouvelles données, nouveaux besoins, etc. Ces mécanismes de feed-back permettent, d'une part pour les utilisateurs, de réduire les différences entre les besoins collectifs et les besoins individuels, d'autre part, d'ajuster le modèle global de données. Le schéma suivant illustre la création en parallèle de Datamarts et d'un Datawarehouse selon un modèle de données global avec processus de feed-back :

**Tableau 4 : Logiciels d'aide à la décision utilisables dans un Datawarehouse aux USA**

Type d'outils	Principales fonctions	Offre logicielle
Outils d'interrogation <sup>(28)</sup>	Générateurs de requêtes SQL, il permettent de créer des tableaux de données à partir de SGBD-R. Ces requêtes sont sauvegardées et constituent des bases d'informations personnalisées (C9).	131
OLAP <sup>(29)</sup> (On Line Analytical Processing)	Outils d'analyse permettant, à partir de bases multidimensionnelles (MOLAP) ou relationnelles (ROLAP) <sup>(30)</sup> , d'agencer les données selon plusieurs dimensions (cubes de données). L'apport de la navigation dans des cubes de données est déterminant dans la révélation de corrélations cachées (C7 et C12).	73
EIS	Outils d'aide à la décision dont la particularité est la facilité d'utilisation. L'ergonomie de la présentation des informations est optimisée de manière à ce que l'utilisateur interprète le plus rapidement possible ce qu'il voit (C7).	59
Datamining	Outil permettant l'exploration et l'analyse de grandes quantités de données afin de découvrir des formes et des règles en utilisant des moyens automatiques ou semi-automatiques (Berry, 1997), (contrainte C12). Le Datamining a pour principal objet la classification, l'estimation, la prédiction, le groupement par similitudes, l'analyse de clusters et la description de ce qui se passe dans une base de données complexe <sup>(31)</sup> . Donc, pour l'utilisateur confronté à de grands volumes de données, il facilite l'émergence de sens.	107
Analyse documentaire	L'origine de ces outils est double : ils permettraient de prendre en compte des gisements de données inexploités dans les différents documents circulant dans l'entreprise, et ils permettraient une ouverture sur le Web (C1).	78
SIG <sup>(32)</sup>	Facilitent l'analyse des données en fonction des caractéristiques géographiques (C7).	29
Statistiques et analyse décisionnelle	Ces outils traditionnels permettent la création d'arbres de décision ou des analyses de risques, par exemple.	74
Modélisation de processus	Ces outils donnent la possibilité de représenter des processus graphiquement sous la forme d'organigrammes (C7).	24
Agents	Ces outils de recherche d'informations indépendants parcourent les SGBD à la recherche d'un événement spécifié par exemple. Ils offrent donc un caractère dynamique par rapport aux traditionnels requêteurs SQL.	27
Reporting	Ces outils créent des rapports périodiques en temps réel.	15

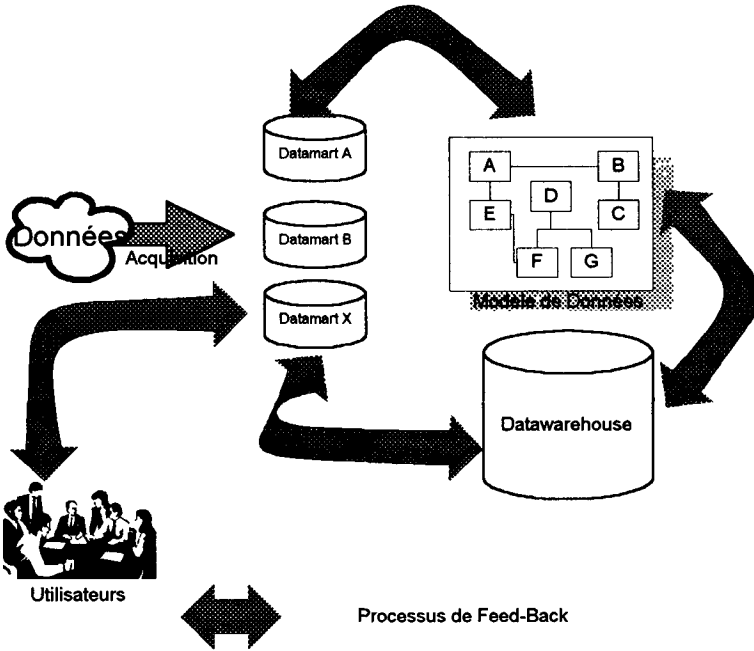
(28) Que l'on trouve aujourd'hui sous le terme de "requêteurs".

(29) OLAP est une norme définie par E.F. Codd en 1993, qui décrit les 12 critères nécessaires à une base de données pour être qualifiée de "décisionnelle" (Gray 1998), p. 383-386.

(30) La création de cubes de données, à partir de bases relationnelles, pose encore aujourd'hui de nombreux problèmes d'optimisations (Gupta, 1997).

(31) *ibid.* Chapitre 4.

(32) Systèmes d'Informations Géographiques.



**Figure 3 : Modèle parallèle avec feed-back**

Source : (Firestone 1997)

### III. MISE EN ŒUVRE

Les exemples de mises en œuvre de systèmes de DW sont extrêmement nombreux<sup>(33)</sup>. Cependant, en raison de l'enjeu stratégique de ces systèmes, les études englobant les aspects techniques et organisationnels sont rares.

Dans cette section, nous allons succinctement décrire un cas dans le domaine d'implémentation.

#### 1. Description de cas<sup>(34)</sup>

• **Problème** : Le groupe Casino, soumis à une rude concurrence de la part des autres grands distributeurs, a décidé dans le début des années 90, une reformulation de sa stratégie. En 1993, deux axes ont été tracés : d'une part une volonté d'ouverture sur l'international, et d'autre part, la décision d'optimisation des parts de marchés sur le territoire français. A cette époque, les efforts de rationalisation de Casino étaient principalement concentrés sur la fonction achats aux fournisseurs

(33) Citons par exemple le projet Safarit de la SNCF concernant 1500 utilisateurs et 200 Go de données dont l'objectif est de fédérer les données commerciales et transport pour le fret SNCF. Ou encore, le Datamart (spécialisé au niveau de la France) mis en œuvre chez McDonald's gérant 630 "restaurants" et 40 Go de données.

(34) Documents sources : "Casino : le filon de l'information" dans (Echos, 1997), p. 40, (Pinsky 1997) et sur les sites suivants : celui de Casino : <http://www.groupe-casino.fr/sommaire/som.html> et celui de la bourse de Paris : <http://www.bourse-de-paris.fr/>

et sur les livraisons aux magasins. Cette entreprise constituait, d'ailleurs, dans le domaine de la logistique, une référence établie. La visibilité du client restait cependant limitée. On se situait, alors, clairement dans une optique de flux poussés vers le client.

• **Solution organisationnelle :** En 1993, un mouvement de réorganisation interne est initié avec pour objectif d'orienter l'activité vers les ventes au client ; on passe donc d'une approche de flux poussés vers une vision de flux tirés. Pour ce faire, de nouvelles règles de gestion sont mises en œuvre (par exemple, les acheteurs de produits deviennent aussi responsables des ventes des mêmes produits). Les bénéfices escomptés de telles règles consistaient en l'augmentation de motivation du responsable et ce, malgré son contrôle croissant par la Direction. Il devenait, en effet, plus aisé de définir les responsables de denrées non vendues.

• **Solution technologique :** Pour accompagner ces actions organisationnelles, Casino décide, fin 1993, de mettre en œuvre un Datawarehouse ADAM (Aides à la Décision Achats et Marketing). Il existait déjà un système d'aide à la décision sous la forme d'un infocentre dérivé d'une base de données DB2. Cet infocentre avait pour objectif de fournir des informations concernant les ventes. Cependant, l'architecture de ce système et des contraintes d'ordre technologique limitaient fortement ses capacités de traitement et d'analyse des données, tout en réduisant sa faculté d'évolution.

Ainsi, au côté de cet ancien système, Casino a réalisé ADAM avec pour objectif d'analyser les flux de marchandises. Voyons, rapidement les caractéristiques d'ADAM.

Tout d'abord, pour la phase d'acquisition des données, les sources sont variées puisqu'il existe plusieurs bases opérationnelles fondées sur des architectures différentes. Apparemment, le groupe n'a pas fait appel à des banques de données externes. Cependant, il est précisé que ce sont les responsables marketing et achats qui ont défini leurs besoins informationnels, à charge des concepteurs d'en déduire les données pertinentes.

Ensuite, concernant le stockage, la base de données principale a une capacité de 480 Go et permet de stocker, sur une durée de 25 mois, les données concernant les produits vendus dans 202 hypermarchés, 320 supermarchés, 400 supérettes et 50 auto-services.

Enfin, au niveau du traitement, le choix s'est porté sur FOCUS, du constructeur INFORMATION BUILDERS. Cette application, largement connue<sup>(35)</sup>, permet de générer des outils d'analyse dans une optique d'aide à la décision.

Comme nous l'avons indiqué précédemment, un DW évoluant, l'ancien infocentre a été fermé récemment et l'ensemble des données qu'il traitait a été récupéré par le système de DW<sup>(36)</sup>. La prise en compte de l'activité vente montre les capacités d'évolution du système et met en lumière l'intégration croissante du DW dans le système d'information de Casino.

(35) La première version est sortie il y a plus de 15 ans.

(36) Une fois cette opération totalement terminée, le DW s'appellera ADAM (Marketing) et EVE (et ventes).

Une fois toute la chaîne du produit (de l'achat à la vente) prise en compte, les responsables prévoient d'installer des outils d'analyses plus puissants, notamment des outils de Datamining.

## 2. Commentaires

Plusieurs enseignements peuvent être tirés de ce cas :

- Il illustre l'intégration stratégique - technologie. En effet, l'efficacité du système de DW n'est rendue possible que parce qu'elle s'inscrit dans le cadre plus large d'une meilleure prise en compte de la clientèle. Cette intégration est marquée notamment par le fait que les deux personnes apparaissant comme responsables de ce projet sont le Directeur des systèmes d'information et le Directeur de l'organisation. Dans un autre cas que nous avons observé<sup>(37)</sup>, l'entreprise sous-traitant le projet de DW a rencontré de nombreuses difficultés dont l'origine résidait dans la grande autonomie accordée aux différents centres<sup>(38)</sup>. Les conséquences étaient triples. Tout d'abord, il existait une très forte hétérogénéité des matériels et applications. Ensuite, il était extrêmement complexe de définir une modélisation commune de l'activité. Enfin, il existait des blocages d'ordre financier pour savoir qui devait assumer le coût de changement de nombreuses architectures informatiques.

- La phase de conception du système a été facilitée par la longue expérience de Casino dans le domaine des bases de données. Le

point important est que le DW a été bâti sur l'activité que maîtrisait le mieux Casino, à savoir les achats. Cela se voit notamment par le fait que chaque produit possédait déjà une méthode cohérente d'identification ce qui a permis de gagner du temps dans la modélisation et l'administration des données. Ce n'est qu'ensuite que le groupe a intégré l'activité ventes. Cette démarche a permis de n'avoir à faire face qu'à une tâche complexe à la fois (soit le système technologique, soit l'activité).

- L'intérêt de disposer d'un système structuré de données réside dans le fait que l'on peut ensuite ajouter des outils (ici de Datamining) sans avoir à remettre en cause l'architecture du système. La structure ainsi créée constitue une base cohérente pour l'utilisation efficace de tout un ensemble d'outils applicatifs décisionnels.

- L'augmentation des capacités de traitement des données issues de la clientèle a conduit le groupe à lancer tout un ensemble de produits de fidélisation de cette dernière, notamment sous la forme de cartes.

- Les responsables du projet estiment qu'au niveau du coût de ADAM, l'investissement de départ, compris entre 10 et 30 millions de francs, a été rentabilisé en un an, ce qui montre une efficacité supérieure à la moyenne<sup>(39)</sup>. Mais il est permis de s'interroger sur les méthodes choisies pour mesurer le retour sur investissement et sur les techniques d'imputation

(37) Il s'agit d'un autre groupe de distribution.

(38) Autonomie profitable selon bien d'autres points de vue.

(39) L'équilibre sur investissement est atteint en moyenne en 2,31 ans, la médiane étant de 1,67 ans. L'investissement moyen est de 2,2 millions de dollars. (Franco, 1997), p. 45.

des apports de ce système par rapport aux actions de nature organisationnelle notamment.

- Des limites technologiques demeurent. En effet, le nombre d'utilisateurs semble strictement limité et certaines requêtes sont très longues à exécuter (près de 20 heures). Cependant, les évolutions dans le domaine coût/puissance rendront à terme ces limites caduques. Les véritables difficultés resteront au niveau de l'administration des données (quelles évolutions dans la structure des bases ?), dans l'administration des applications (quels nouveaux outils installer ?) et dans celle des utilisateurs (qui fera quoi ?).

En résumé, ce cas montre comment la stratégie d'un grand groupe peut être soutenue par un système technologique approprié. Sans une nouvelle orientation

stratégique, le Datawarehouse se serait révélé d'intérêt moindre. Cependant, il apparaît que la relation d'impact de la stratégie vers l'outil comportera selon toute vraisemblance un effet en retour. Il sera intéressant, dès que les applications de Datamining seront opérationnelles, d'observer l'impact réel des corrélations mises en évidence sur le fonctionnement de l'organisation.

#### IV. PERSPECTIVES

Nous insisterons dans cette section sur deux types de perspectives de développement.

En utilisant le logiciel d'analyse UMAP<sup>(40)</sup>, une étude précise nous a conduits à mettre en évidence la carte de connaissance<sup>(41)</sup> suivante :

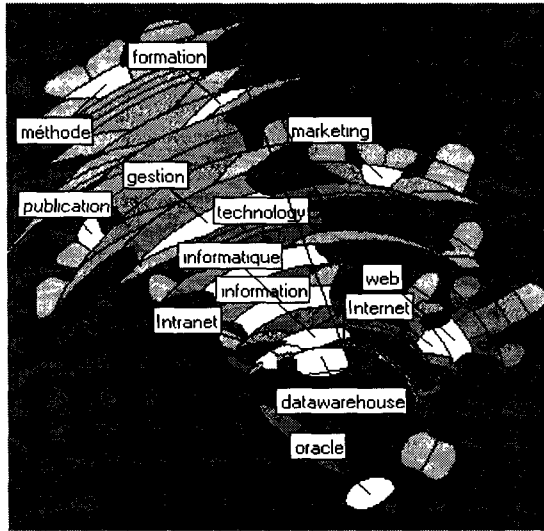
#### *Cf. figure 4 ci-contre*

On voit clairement apparaître trois axes de développement. La zone 1 qui met en évidence tous les éléments concernant le Datawarehouse traditionnel dans une organisation. La zone 2 qui montre l'intérêt d'un tel système pour le marketing. En effet, l'un des principaux objectifs de ce dernier domaine est la parfaite connaissance des besoins de la clientèle

tant actuelle que future. Enfin, dans la zone 3 on constate la complémentarité entre un système de DW et le concept Internet. Ce dernier constitue une source de données importante qu'il faut intégrer au système. Cependant, les données circulant sur Internet sont de formats différents, de périodicité incertaine et de fiabilité variable. Aussi, de nombreux ou-

(40) UMAP est un logiciel de recherche et de représentation d'informations selon le concept des arbres de connaissances (voir les ouvrages de P. Lévy notamment). La recherche a été effectuée sur le Web à partir du mot Datawarehouse sur 7 moteurs de recherche avec un nombre de documents limité à 150, une taille à 400 Ko et pendant une durée de deux jours. La carte a été tracée à partir d'une simulation de la carte principale où les mots du thesaurus non pertinents ont été supprimés. Enfin nous avons pris comme paramètre pour le niveau thématique (liens entre les mots à l'intérieur d'un texte) et pour le niveau contextuel (liens entre les textes qui partagent les mots) 50 %.

(41) Selon les concepteurs du logiciel, une carte permet de repérer les cohérences thématiques d'un ensemble de textes selon le point de vue d'un utilisateur.



**Figure 4 : Carte de connaissance  
à partir du mot Datawarehouse**

tils existants se sont ouverts à Internet, et de nombreux autres apparaissent (dans le domaine de l'analyse documentaire, par exemple). Notons, pour terminer la position d'Oracle sur la carte de connaissance. On peut interpréter cette position de deux manières. Soit il s'agit du fait qu'Oracle occupe une position dominante écrasant les autres constructeurs de SGBD, soit cela peut indiquer l'existence d'un décalage de cette entreprise par rapport aux directions d'évolution précitées du Datawarehouse.

En ce qui concerne la seconde perspective de développement, les applications de Datamining, utilisées dans le cadre d'un DW sont susceptibles de créer des connaissances à partir de données. Cependant, le système ne permet pas d'intégrer ces connaissances, et de les diffuser. En effet, si l'on considère que les données sont

stockées dans les tables des SGBD, les informations dans des requêtes SQL, par exemple, rien ne permet d'appréhender les connaissances ou de tenir compte de l'expérience emmagasinée (contrainte C11). Ainsi, le recours au stockage de règles sous forme de "questions - réponses" (FAQ) n'est pas envisagé. Par ailleurs, on note une déconnexion entre les approches Datawarehouse et Groupware. Cette déconnexion s'illustre notamment par le fait que le mot Groupware n'apparaît pas lors d'une recherche avec UMAP sur le mot DW et réciproquement. Les passerelles entre les informations et connaissances produites par le traitement des données d'un entrepôt et celles produites dans les forums de groupware, leur intégration dans un système global, constitueront peut-être le prochain défi de l'informatique décisionnelle dans les organisations.

## CONCLUSION

L'objet de cet article était de faire le point sur le système de Datawarehouse, pour au moins trois raisons<sup>(42)</sup> :

- il s'agit du premier système informatique d'aide à la décision de cette envergure tenant compte des rapports Donnée - Information ;

- le DW est un système fédérateur d'une multitude d'outils et assure la cohérence des données qu'il gère ;

- la mise en œuvre d'un tel système conduit généralement à une normalisation du système d'information de l'entreprise ;

Cependant, le DW possède plusieurs limites :

- il permet de conduire à la création de connaissances mais ne prend pas en compte leur gestion ;

- il est très difficile de mettre en œuvre dans un seul projet un DW englobant l'ensemble des activités de l'entreprise, d'où le développement des Datamarts ;

- les modèles relationnels ou multidimensionnels ne sont pas conçus pour appréhender des données complexes (images, sons etc.). Une solution résiderait dans l'intégration de l'approche objet au sein des bases relationnelles<sup>(43)</sup>,

ce qui permettrait de profiter de la puissance de cette approche tout en gardant une structure que les utilisateurs maîtrisent ;

- d'autres limites sont plus générales, ainsi est-il admis que le DW n'apporte pas de contributions particulières aux voies et moyens de la modélisation de l'activité. Il n'apporte par ailleurs rien de bien nouveau en ce qui concerne la définition de la notion de donnée en tant qu'atome élémentaire de la connaissance<sup>(44)</sup>.

En résumé, un système de DW rend compte du degré de développement et de maîtrise des TIC dans l'organisation. Il témoigne, en outre, de l'importance accordée à une gestion rigoureuse et intégrale des processus décisionnels via la gestion des données.

## BIBLIOGRAPHIE

Berry, M.J., Linoff, G. (1997), *Data-mining : techniques appliquées au marketing, à la vente et au service clients*, InterEditions 379 p.

Darling, C.B. (1996), « How to Integrate your DataWarehouse », *Datamation* Vol. 42, n° 10, mai, p. 40-51.

Echos, (1997), « Dossier Informatique », *Les Echos* 16 avril, p. 39-42.

EDS, (1997), *Modélisation des bases décisionnelles*, Etude du Cabinet Pro-méthéus, EDS.

(42) On pourrait en ajouter une quatrième : la quasi-majorité des grandes entreprises américaines ont lancé un tel projet.

(43) Comme le montre la nouvelle version (8) du SGBD de la société ORACLE qui offre une approche mixte, relationnelle - Objet.

(44) De nombreuses disciplines tentent de déterminer des méthodes visant à préciser les conditions d'une vision unifiée des données. Spiegler, par exemple, propose deux méthodes pour unifier les aspects physiques, logiques et conceptuels des données (Spiegler, 1995). On trouve aussi dans l'histoire de la recherche comptable un problème du même ordre : l'ambition de déterminer le fait comptable élémentaire que chaque utilisateur va pouvoir transformer en information comptable correspondant à sa spécificité. L'origine de ces recherches se trouve chez E. de Fagés de la Tour en 1926 avec sa conception des "infosiores" et trouve un écho très récent dans la comptabilité à parties multiples.

Firestone, J.M. (1997), *Datawarehouses and Datamarts : A Dynamic View*, Executive Information Systems Inc, White Paper n° 3, 27 mars, 11 p.

Franco, J.M. (1997), *Le Datawarehouse - Le Datamining*, Eyrolles, 203 p.

Gupta, H. et al. (1997), *Index Selection for OLAP*, White Paper - Stanford University - Department of Computer Science, 14 p.

IDC, (1996), *The Foundations of Wisdom : A Study of the Financial Impact of Data Warehousing*, International Data Corporation Special Report, White Paper, 25 p.

IDC, (1997), *Le marché français du Datawarehouse*, International Data Corporation. Etude Logiciel et service, Juillet.

Gray, P. Watson, H.J. (1998), *Decision Support in the Datawarehouse*, Prentice Hall PTR.

Inmon, B. (1996), *Building the Data Warehouse, 2<sup>nd</sup> Edition*, John Wiley & Sons, Inc., New York.

Inmon, (1997), *Corporate Information Factory*, John Wiley & Sons, Inc., New York.

Infomag, (1996), *Naviguer dans le datawarehouse*, Dossier, Informatiques Magazines n° 16, avril, p. 82-99.

Infomag, (1997), *Datamining : Comment explorer ses données*, Dossier, Informatiques Magazines n° 26, mars, p. 109-114.

Infomag, (1998), *Datamart : comment exploiter le Web*, Dossier, Informatiques Magazines n° 40, janvier, p. 52-61.

Kimball, R. (1996), *The Datawarehouse Toolkit : Practical techniques for building dimensional Datawarehouses*, John Wiley & Sons.

Lebraty, J.F. (1994), *Nouvelles Technologies de l'Information et Processus de Prise de Décision*, Thèse de Doctorat - Université de Nice Sophia-Antipolis.

Lebraty, J.F. (1997a) *Réseaux de neurones et structure informationnelle de l'organisation*, Papiers de recherches du RODIGE, juin, 23 p.

Lebraty, J.F. (1997b), *Gestion de la connaissance et nouvelles technologies de l'information*, Cahiers de recherches de l'AFC, n° 3, septembre, p. 55-68.

Pinsky, D. (1997), « Casino nourrit son Datawarehouse ». *Informatiques Magazines* n° 27, avril, p. 38-40.

Singh, H.S. (1998), *Data Warehousing, Concepts, Technologies, Implementations, and Management*, Prentice Hall PTR.

Spiegler, I. (1995), « Toward a unified view of data ». *Information Systems Management* Vol. 12, n° 2, p. 7-14.

Varney, S.E. (1996a), « Datawebs ! Link the Web to Your Legacy Data and Apps ». *Datamanion* Vol. 42, n° 7, April, p. 38-47.

Varney, S.E. (1996b), « Datamarts : Coming to an IT Mall near you ! ». *Datamanion* Vol. 42, n° 11, June, p. 44-46.

O1Info, (1997), *Le Datawarehouse*, Dossier, O1 Informatique, n° 1442, 28 février, p. 35-41.

### Sites web

<http://pwp.starnetinc.com/larryg/index.html> (*Datawarehousing Information Center, la référence en la matière*)  
<http://www.ibi.com>  
<http://www.informatiques.com>  
<http://sas.com>  
<http://direct.boulder.ibm.com/bi/>  
[http://www.kenan.com/acumate/acu\\_home.htm](http://www.kenan.com/acumate/acu_home.htm)  
<http://www-db.stanford.edu/warehousing/publications.html>  
<http://www.cs.su.oz.au/~thierry/ckdd.html>  
<http://www.olapcouncil.org>  
<http://www.cs.bam.ac.uk/~anp.TheDataMine.html>  
<http://www.oracle.com>  
<http://www.groupe-casino.fr/>  
<http://www.casino.fr>  
<http://www.bourse-de-paris.fr/>

### CD ROM

Le Datawarehouse. CD ROM édité par Business et Décision. 23, rue Cambon, 75001 Paris (Tél. : 01 43 16 44 00).